Localized Ensemble-Based Tangent Linear Models and Their Use in Propagating Hybrid Error Covariance Models

SERGEY FROLOV
University Corporation for Atmospheric Research, Monterey, California
CRAIG H. BISHOP
Naval Research Laboratory, Marine Meteorology Division, Monterey, California

(Manuscript received 3 April 2015, in final form 9 December 2015)

ABSTRACT

Hybrid error covariance models that blend climatological estimates of forecast error covariances with ensemble-based, flow-dependent forecast error covariances have led to significant reductions in forecast error when employed in 4DVAR data assimilation schemes. Tangent linear models (TLMs) designed to predict the differences between perturbed and unperturbed simulations of the weather forecast are a key component of such 4DVAR schemes. However, many forecasting centers have found that TLMs and their adjoints do not scale well computationally and are difficult to create and maintain—particularly for coupled ocean–wave–ice–atmosphere models. In this paper, the authors create ensemble-based TLMs (ETLMs) and test their ability to propagate both climatological and flow-dependent parts of hybrid error covariance models. These tests demonstrate that rank deficiency limits the utility of unlocalized ETLMs. High-rank, time-evolving, flow-adaptive localization functions are constructed and tested using recursive application of short-duration ETLMs, each of which is localized using a static localization. Since TLM operators do not need to be semi-positive definite, the authors experiment with a variety of localization approaches including step function localization. The step function localization leads to a local formulation that was found to be highly effective. In tests using simple one-dimensional models with both dispersive and nondispersive dynamics, it is shown that practical ETLM configurations were effective at propagating covariances as far as four error correlation scales.

1. Introduction

Recent results show that ensemble-variational hybrids provide theoretical (Bishop and Satterfield 2013) and practical (Kuhl et al. 2013; Buehner et al. 2010a; Zhang and Zhang 2012; Buehner et al. 2010b; Kleist and Ide 2015) benefits for four-dimensional data assimilation. Most operational implementations of the four-dimensional hybrid system (Kuhl et al. 2013; Lorenc et al. 2015) use the tangent linear model (TLM) and its adjoint (conjugate transpose) to propagate error statistics between observation and analysis times (see section 2 for a formal review and definition of such methods). Unfortunately, the computational scalability of such systems is limited by the computational scalability of the dynamics-based TLM (Lorenc 2003a) and the difficulty of maintaining accurate TLM and adjoints in coupled systems.

Several national numerical weather prediction (NWP) centers have tested an alternative four-dimensional ensemble-variational approach (4DEnVar) that does not use a TLM and adjoint operators (Lorenc 2003a; Lorenc et al. 2015; Buehner et al. 2010a; Kleist and Ide 2015). However, the performance of such adjoint-free 4DEnVar methods has been found to be worse than Hybrid 4DVAR methods. Lorenc et al. (2015) suggests that the primary cause of the relatively poor performance of the 4DEnVar method is due to suboptimal localization of the four-dimensional covariances and the inability of 4DEnVar to propagate the climatological part of the error covariance.

In this paper, we introduce the localized ensemble-based TLM (ETLM) that uses an ensemble of forward
models to construct TLM and adjoint operators. The ETLM models provide several possible advantages over the traditional TLM models. Ensemble runs used by the ETLM can be precomputed before the analysis cycle and the cost of the ensemble is already justified by purposes other than data assimilation. ETLM can be constructed from an ensemble of forward model integrations even in cases when linearized dynamics-based TLM is not available. And finally, no additional maintenance will be required for ETLM models, which will allow for more rapid development of the forward model.

Nonlocalized versions of ETLMs have been introduced by several prior studies (e.g., Pelc et al. 2012; Lermusiaux and Robinson 1999). For example, (Pelc et al. 2012) demonstrated the successful application of nonlocalized ETLMs for the case of a low-dimensional 1D ecosystem model in which the number of state variables and ensemble members was similar. However, such nonlocalized ETLMs will fail to propagate full-rank hybrid covariance matrices. Here, we attempt to extend this work to high-dimensional systems in which the number of degrees of freedom is much larger than the ensemble size. We show that in order to propagate modes of the covariance function that lie outside the ensemble subspace, elements of the ETLM must be moderated or localized.

Several aspects of ETLM localization are considered in this paper. First, a recursive localization method is introduced that constructs a four-dimensional localization functions from recursive application of short-duration ETLMs, each of which is localized using simple three-dimensional static localization. Over this short-duration period, (i) the signal of interest is unlikely to propagate far and a static localization is appropriate and (ii) the TLM matrix is likely to be sparse, hence, requiring fewer ensemble members to specify nonzero elements of the ETLM operator accurately. In this paper, we show that in high-dimensional systems, the accuracy of ETLMs is usually limited by the rank deficiency of the ETLM, rather than the statistical causes of spurious correlations. In particular, inaccuracies may arise if there are fewer ensemble members than nonzeros elements in the true TLM operator. To address this rank deficiency, we propose several solutions, including a local ETLM (LETLM), which computes parts of the TLM model on small subdomains of the global grid. For each subdomain, the ratio of unknown ETLM elements to the number of ensemble members is more favorable than on the global grid. We also show potential advantages of constructing ETLM in a spectral space. Specifically, we show that almost perfect localization can be achieved when ETLM is localized in its own eigenspace, where ETLM becomes diagonal and as few as two ensemble members are required to compute the ETLM perfectly.

To test the proposed ETLM approach we use three one-dimensional forward models with prescribed nondispersive and dispersive dynamics. The dispersive dynamics simulates dispersion characteristic of barotropic Rossby waves (Holton 2004) and baroclinic Eady-edge waves (Bishop and Heifetz 2000). We chose to test the developed methods using dispersive dynamics because waves in the atmosphere can propagate with different speeds and, hence, at longer lead times, they can generate complex, noncompactly supported covariance structures (Hakim 2005).

Section 2 of this paper reviews established methods for computation of four-dimensional hybrid error covariances and the role that TLMs and ETLMs play in these computations. Section 3a of the paper established the connection between traditional TLMs and the ETLMs. The practical aspects of ETLM computation and localization are presented in sections 3b–3e. The experiment setup is presented in section 4 and the results, in section 5. We conclude the paper with the discussion of the results and our perspective on how the developed methods can be applied in realistic high-dimensional systems.

2. Role of tangent linear models in hybrid DA

In many operational implementations of the four-dimensional hybrid DA systems the outer minimization loop is not used. In such cases, it is possible to write variational estimation equations using an equivalent Kalman gain formulation, even though the gain matrix itself is never computed explicitly. To simplify notation, but without a loss of generality, we defined these estimation equations for only two time levels: the initial time $t = 0$ and future time $t = 1$:

$$
\begin{align*}
\begin{cases}
    \mathbf{x}_t' &= \mathbf{x}_0' + \mathbf{K}[\mathbf{y} - H(\mathbf{x}_t')]
    \\
    \mathbf{K} &= \mathbf{P}_{00}^{3D \text{ hybrid}} [\mathbf{I} \mathbf{M}^T] \mathbf{H}^T \\
    \mathbf{P}_{00}^{3D \text{ hybrid}} &= \mathbf{\alpha}_{\text{clim}} \mathbf{P}_{00}^{3D \text{ clim}} + \mathbf{\alpha}_{\text{ens}} \mathbf{P}_{00}^{3D \text{ loc}} \\
    \mathbf{P}_{00}^{3D \text{ ens}} &= \mathbf{P}_{00}^{3D \text{ ens}}
\end{cases}
\end{align*}

(1)
where \( \mathbf{x}_0 \) and \( \mathbf{x}_0' \) are the first guess and the analysis, respectively, of the model state at initial time \( t = 0 \); \( y \) and \( H(\mathbf{x}') \) are the observations and the forecast of observations taken either at the initial time \( t = 0 \) and the later time \( t = 1 \); \( \mathbf{x}' = [\mathbf{x}'_0 \mathbf{x}'_1] \) is the state vector listing the variables at both observation times; \( \mathbf{P}_{00} \) is the covariances of the forecast errors \( \mathbf{e}_0 = \mathbf{x}_0' - \mathbf{x}_0 \) at the initial time \( t = 0 \); \( \mathbf{R} \) is the covariance of observation errors; \( \mathbf{M} \) and \( \mathbf{H} \) are linearizations of nonlinear forecast \( M(\cdot) \) and observation \( H(\cdot) \) operators, respectively; \( \mathbf{C}_{00}^{3D\text{loc}} \) is the ensemble localization matrix; \( \odot \) is the Schur (elementwise) matrix product; and \( \alpha_{\text{clim}} \) and \( \alpha_{\text{ens}} \) are the scalar weights on the climatological and flow-dependent parts, respectively, of the ensemble covariance. The linearized dynamical model \( \mathbf{M} \) is often referred to as the TLM operator and its conjugate transpose \( \mathbf{M}^\top \) as the adjoint operator.

Application of the TLM and adjoint operators in Eq. (1) defines a linear approximation to the temporal evolution of a (two level) four-dimensional covariance matrix \( \mathbf{P}^{4D} \):

\[
\mathbf{P}^{4D} = \alpha_{\text{clim}} \mathbf{P}^{4D\text{clim}} + \alpha_{\text{ens}} \mathbf{C}^{4D\text{loc}} \odot \mathbf{P}^{4D\text{ens}}
\]

where subscripts \( i \) and \( j \) in \( \mathbf{P}_{ij} \) refer to the time levels of the four-dimensional covariance. We use superscript 3D to highlight when the covariance is defined for a single time level \( (i = j) \). The benefit of using a linear approximation to compute covariance propagation in Eq. (2) is that we have a good working understanding of how to specify both the climatological three-dimensional covariance \( \mathbf{P}_{00}^{3D\text{clim}} \) and the three-dimensional localization \( \mathbf{C}_{00}^{3D\text{loc}} \) for the ensemble covariance \( \mathbf{P}_{00}^{4D\text{clim}} \). Unfortunately, the computational scalability of such systems is limited by the computational scalability of the dynamics-based TLM (Lorenc 2003a). The TLM and its adjoint are also hard to develop and maintain, especially in the case of coupled (e.g., ocean–atmosphere) dynamics.

Several national NWP centers tested an alternative ensemble-variational approach (4DEnVar) on Eqs. (1)–(2) that does not use a TLM and adjoint operators (Lorenc 2003a; Lorenc et al. 2015; Buehner et al. 2010a; Kleist and Ide 2015). Using our two-time level example in Eq. (1), these ensemble-variational approaches define the four-dimensional covariance \( \mathbf{P}^{4D} \) as follows:

\[
\mathbf{P}^{4D} = \alpha_{\text{clim}} \mathbf{P}^{4D\text{clim}} + \alpha_{\text{ens}} \mathbf{C}^{4D\text{loc}} \odot \mathbf{P}^{4D\text{ens}}
\]

While Eq. (3) is computationally scalable and does not require a TLM, it is unclear how to specify a four-dimensional localization \( \mathbf{C}^{4D\text{loc}} \) or the four-dimensional climatological error covariance \( \mathbf{P}^{4D\text{clim}} \). Instead, the four-dimensional climatological covariances and four-dimensional localization functions were approximated using their three-dimensional counterparts (Lorenc et al. 2015):

\[
\mathbf{P}^{4D\text{clim}} \approx [\mathbf{P}_{00}^{3D\text{clim}} \mathbf{P}_{01}^{3D\text{clim}}] \quad \mathbf{P}^{4D\text{ens}} \approx [\mathbf{P}_{00}^{3D\text{ens}} \mathbf{P}_{01}^{3D\text{ens}}]
\]

Brought to you by provisional account | Unauthenticated | Downloaded 05/26/22 06:59 AM UTC
Lorenc et al. (2015) has clearly demonstrated the inferiority of the results arising from these assumptions.

In this paper, we replace the dynamics-based TLM and adjoint of traditional 4DVAR with ensemble-based TLM and adjoint. We propose to compute the four-dimensional climatological covariance \( \mathbf{P}^{\text{4D clim}} \) as

\[
\mathbf{P}^{\text{4D clim}} \approx \mathbf{P}^{\text{3D clim}} \mathbf{M}^{\text{ens}}\mathbf{P}^{\text{3D clim}} \quad \text{and} \quad \mathbf{P}^{\text{4D clim}} \approx \mathbf{P}^{\text{3D clim}} \mathbf{M}^{\text{ens}}\mathbf{P}^{\text{3D clim}} \mathbf{M}^{\text{ens}}^T ,
\]

and the localized four-dimensional ensemble covariance \( \mathbf{P}^{\text{4D ens}} \) as

\[
\mathbf{P}^{\text{4D ens}} \approx \mathbf{P}^{\text{3D ens}} \mathbf{M}^{\text{ens}}\mathbf{P}^{\text{3D ens}} \quad \text{and} \quad \mathbf{P}^{\text{4D ens}} \approx \mathbf{P}^{\text{3D ens}} \mathbf{M}^{\text{ens}}\mathbf{P}^{\text{3D ens}} \mathbf{M}^{\text{ens}}^T .
\]

We define the ETLM \( \mathbf{M}^{\text{ens}} \) in the methods section below.

3. Methods

a. Traditional and statistical TLMs

A traditional (dynamics based) tangent linear operator \( \mathbf{M} \) [which we used in Eq. (2)] to propagate the covariance in time is defined as

\[
\delta \mathbf{x}_i = \mathbf{M}_{ij}(\mathbf{x}_i + \delta \mathbf{x}_i) - \mathbf{M}_{ij}(\mathbf{x}_i) = \mathbf{M}_{ij} \delta \mathbf{x}_i + \epsilon_i ,
\]

where \( \mathbf{x}_i = \mathbf{M}_{ij}(\mathbf{x}_i) \) is the original nonlinear model of the forward dynamics; \( \mathbf{x}_i \) is the unperturbed state at initial time \( i = 0 \), \( \delta \mathbf{x}_i \) is the initial-time perturbation; and \( \epsilon_i \) is the linear truncation error. The TLM operator \( \mathbf{M} \) is chosen to minimize the truncation errors \( \epsilon_i \) for the case where \( \delta \mathbf{x}_i \) is infinitesimally small:

\[
\lim_{\delta \mathbf{x}_i \to 0} [\mathbf{M}_{ij}(\mathbf{x}_i + \delta \mathbf{x}_i) - \mathbf{M}_{ij}(\mathbf{x}_i) - \mathbf{M}_{ij} \delta \mathbf{x}_i] = 0 .
\]

However, as Lorenc (2003b) points out, it is not possible to derive or use a true tangent model \( \mathbf{M} \) derived for infinitesimal perturbations in the context of the NWP. Hence, in practice some form of a perturbation model is optimized for finite-size perturbations—often it is tuned for perturbations around the same order of magnitude as the forecast error or analysis correction.

Alternatively, in the ensemble literature (Van der Merwe and Wan 2001; Lermusiaux and Robinson 1999), ensemble tangent linear models can be defined as ensemble regression models that allow propagation of finite-amplitude perturbations:

\[
\mathbf{M}_{ij}(\overline{\mathbf{x}}_i + \delta \mathbf{x}_i) - \mathbf{M}_{ij}(\overline{\mathbf{x}}_i) = \mathbf{M}^{\text{stat}}_{ij} \delta \mathbf{x}_i + \epsilon_i ,
\]

where \( \overline{\mathbf{x}}_i \) and \( \mathbf{M}_{ij}(\overline{\mathbf{x}}_i) \) are the mean of some ensemble of initial states and the mean of the nonlinear ensemble trajectories at final time, respectively. In the sense of minimum error variance, the best linear unbiased estimate of the statistical TLM in Eq. (9) for finite-sized perturbations can be used in statistical expectations. First, we right multiply both sides of Eq. (9) by \( \delta \mathbf{x}_i^T \) to yield the following:

\[
\langle [\mathbf{M}_{ij}(\overline{\mathbf{x}}_i + \delta \mathbf{x}_i) - \mathbf{M}_{ij}(\overline{\mathbf{x}}_i)]\delta \mathbf{x}_i^T \rangle = \mathbf{M}^{\text{stat}}_{ij} \langle \delta \mathbf{x}_i \delta \mathbf{x}_i^T \rangle + \langle \epsilon \delta \mathbf{x}_i^T \rangle ,
\]

where we assume that \( \langle \epsilon \delta \mathbf{x}_i^T \rangle = 0 \). Then, the statistical TLM is

\[
\mathbf{M}_{ij}^{\text{stat}} = \langle [\mathbf{M}_{ij}(\overline{\mathbf{x}}_i + \delta \mathbf{x}_i) - \mathbf{M}_{ij}(\overline{\mathbf{x}}_i)]\delta \mathbf{x}_i^T \rangle \langle \delta \mathbf{x}_i \delta \mathbf{x}_i^T \rangle^{-1} .
\]

By construction, \( \mathbf{M}_{ij}^{\text{stat}} \) in Eqs. (9)–(11) is optimized for the distribution of finite-size perturbations from which it is derived. Note that in the special case where the distribution of perturbations used in Eqs. (9)–(11) are infinitesimally small and the perturbations span the entire space of the model operator, then \( \mathbf{M}_{ij}^{\text{stat}} \) is identical to the traditional (dynamics based) TLM \( \mathbf{M}_{ij} \) defined by Eq. (7).

b. Unlocalized ensemble-based approximations to statistical TLMs

If a \( K \)-member ensemble forecast is available and the \( n \times K \) matrices \( \mathbf{X}_i \) and \( \overline{\mathbf{X}}_i \) denote the ensemble perturbations at two times \( i \) and \( j \), respectively (where \( i < j \)), a low-rank unlocalized ensemble approximation \( \mathbf{M}_{ij}^{\text{ens}} \) to the statistical TLM in Eq. (11) can be expressed as

\[
\mathbf{M}_{ij}^{\text{stat}} \approx \mathbf{M}_{ij}^{\text{ens}} = \mathbf{M}_{ij}^{\text{ens}} = \mathbf{X}_i (\mathbf{X}^T_i \mathbf{X}^T_i)^{-1} ,
\]

where the superscript “−1” on the last term indicates the pseudoinverse (defined below). It can be shown that an equivalent form of Eq. (12) is given by

\[
\mathbf{M}_{ij}^{\text{ens}} = \mathbf{X}_i (\mathbf{X}^T_i \mathbf{X}^T_i)^{-1} ,
\]

(13)
Computation of Eq. (13) is more concise than Eq. (12) because it requires fewer computations in cases where the number of ensemble members is smaller than the number of state variables.

The matrix products $X_i X_j^T$ and $X_j^T X_i$ in Eqs. (12) and (13) are rank deficient and often poorly conditioned (recall that at least one eigenvalue is equal to zero because the sum of the perturbations is equal to zero). Hence, we compute their inverses using matrix pseudoinverse, which can be defined using the singular value decomposition (SVD) as follows. If the SVD of the perturbation matrix $X_i$ is defined as $USV^T \leftarrow \text{svd}(X_i)$, where $U$ and $V$ are orthonormal and $S$ is a diagonal matrix, then the pseudoinverses can be written as

\[
(X_i X_j^T)^{-1} = (X_j^T X_i) = VS^{-V^T} \\
(X_j^T X_i)^{-1} = (X_i X_j^T)^{-1} = US^{-U^T}
\]

where $S^{-}$ is the diagonal matrix that has scalar inverses of the first $r$ numerically significant singular values ($s_1$ through $s_r$) and zeros for the others. In practice, we used MATLAB’s `pinv` function, which uses the following value for the threshold $r$

\[
s_r = \epsilon l \max(s),
\]

where $l$ is the number of the singular values and $\epsilon$ is the machine round-off error.

It is easy to show that in a system where the ensemble perturbations are small enough to be governed by linear dynamics, the ETLMs given by Eqs. (12) and (13) are error free for perturbations that lie within the ensemble subspace. Such perturbations take the form $\delta x_i = X_i b$ at the initial time, and hence, when the ensemble perturbations are governed by linear dynamics, the final time state of the perturbations is simply $(\delta x_i = X_i b)$. To see that this property is enforced by the ETLM $M^{M0}_{i,j}$ given by Eq. (13) note that

\[
M^{M0}_{i,j} \delta x_i = [X_j (X_i^T X_i)^{-1} X_i^T] X_i b = X_i (X_i^T X_i)^{-1} (X_i^T X_j) b = X_i b.
\]

The consequence of Eq. (16) above is that ETLM is perfect when ensemble perturbations span the entire state space and the error dynamics is linear.

c. Localized, recursive ETLMs

The ETLMs given by Eqs. (12) and (13) are error free for infinitesimal perturbations that lie within the ensemble subspace. However, this perfection is not expected for perturbations that lie outside of the ensemble subspace, such as perturbations associated with the climatological part of a hybrid forecast error covariance matrix $P^{3D}_{00}$ or localized ensemble covariance ($C^{3D \text{loc}} \otimes P^{3D \text{ens}}$). To aid the ETLM in propagating perturbations that lie outside of the ensemble subspace, we introduce ETLM localization methods that can increase the rank of the ETLM (even if at the cost of destroying the perfection of ETLM within the unlocalized ensemble subspace).

It is challenging to define a flow-adaptive, four-dimensional localization between arbitrarily space–time points. Instead, we suggest using a static localization applied recursively to a sequence of short-duration ETLM models:

\[
M^{M\gamma}_{m,0} = M^{M\gamma \text{loc}}_{m,0(m-\Delta t)} \times \cdots \times M^{M\gamma \text{loc}}_{m,0},
\]

where $M^{M\gamma \text{loc}}_{m,0}$ is a short-duration ETLM localized using one of the static localization methods described below (M1, M2, or M3) and $M^{M\gamma \text{loc}}_{m,0}$ is the resulting recursive product. To denote the type of the static localization used, we include the localization type in the name of the recursive product. For example, for the $M^{M1 \text{loc}}_{m+1,m}$, we use the following:

\[
M^{M1 \text{loc}}_{m+1,m} = \gamma C^{M1} \otimes M^{M0})_{m+1,m}.
\]

We envision that the ETLM localization will be computed for time steps $\Delta t$ that are much larger than those of a traditional TLM. In most of the experiments in this paper, ETLM localization is applied every five time steps (we envision that for weather forecasting application the time step can be as large as 1 h).

To construct the simplest localization function $M^{M1 \text{loc}}_{i+1,i}$, we assume that most of the information in the ETLM is local and input of far-away locations can be explicitly set to zero. Then, we can specify that

\[
M^{M1 \text{loc}}_{i+1,i} = \gamma C^{M1} \otimes M^{M0})_{i+1,i},
\]

where $C^{M1}$ is a static, distance-based localization function like the Gaspari–Cohn (GC) (Gaspari and Cohn 1999) or the Tukey window (TW) function described below. In most applications (where we have fewer ensemble members than the state variables), it will be highly beneficial to use the computationally concise formulation of the $M^{M0}$ ETLM in Eq. (13).

To compensate for possible attenuation or amplification associated with repeated application of the ETLM, we introduce an optional scaling factor $\gamma$. We used the TW localization function (Harris 1978) instead of a more traditional box car or GC localization function to avoid undesired attenuation of local
information (such as advection). The TW localization has a flat top region that preserves the magnitude of state variables that are in a close neighborhood of the grid point (cf. shapes of localization function in Fig. 1a). Also, due to smooth tapering, TW has fewer negative eigenvalues as compared to the box car localization (cf. the eigenvalue spectra in Fig. 1b). Since TLMs are not semipositive definite matrices, the importance of positive definiteness for the localization function applied to an ETLM is unclear. The formula for the TW localization is given by the following:

\[
TW(\Delta x, \alpha, L) = \begin{cases} 
1 & \text{if } \Delta x \leq \alpha L \\
1 + \cos \left[ \frac{2\pi}{\alpha} \left( \frac{\Delta x}{2L} - 1 + \frac{\alpha}{2} \right) \right] & \alpha L < \Delta x \leq L \\
0 & \Delta x > L
\end{cases}
\]

(20)

where \(\Delta x\) is the distance between two localized variables, \(L\) is the width of the window where localization function decays to zero, and \(\alpha\) is the ratio that determines the distance over which the localization function transitions from a value of 1 to a value of 0.

ETLM \(M^M_{j+1,j}\) in Eq. (19) is based on the computationally concise formulation of the \(M^M_{j+1,j}\) ETLM in Eq. (13) that does not allow for the matrix product \((X_i^T X_i)\) to be localized before it is inverted. To allow for localization of the inverted matrix product, we introduce M2 localization based on the \(M^M_{j+1,j}\) formulation in Eq. (12):

\[
M^M_{M2,loc,i+1,j} = \left[ C^{M2} \odot \left( X_i^T X_i \right) \right] \left[ C^{M2} \odot \left( X_i^T X_i \right) \right]^{-1}.
\]

(21)

We found that it was beneficial to use the same localization \(C^{M2}\) for both the nominator and the denominator in Eq. (21). Because \(C^{M2}\) is applied to semipositive definite covariances matrix \((X_i^T X_i)\), we chose to use GC-based localization for \(C^{M2}\) instead of TW. Notice that \(M^M_{M2,loc,i+1,j}\) requires inverse of the matrix \([C^{M2} \odot \left( X_i^T X_i \right)]^{-1}\), which is the size of the state space \((n \times n)\). For such an approach to be computationally feasible, rapid methods for computing \([C^{M2} \odot \left( X_i^T X_i \right)]^{-1}\) will be necessary. We present one possible computational solution based on patch decomposition of the domain (section 3e). However, we defer a detailed examination of the rapidity with which this could be done in large models to later work. Here, our objective is just to indicate the potential utility of the localized ETLM given by Eqs. (21) and (22).

The next ETLM formulation that we consider combines aspects of both the \(M^M_{M2,loc}\) and \(M^M_{M1,loc}\) localizations:

\[
M^M_{M3,loc,i+1,j} = C^{M3} \odot M^{M2,loc}_{i+1,j}.
\]

(22)
where $\mathbf{C}^\text{M3}$ is the TW-based localization and $\mathbf{M}^\text{M2,loc}_{i,i}$ is defined in Eq. (21).

The values of parameters for localization functions $\mathbf{C}^\text{loc}$, $\mathbf{C}^\text{M1}$, $\mathbf{C}^\text{M2}$, and $\mathbf{C}^\text{M3}$ used in our experiments are listed in Table 1.

d. Localization of ETLM in spectral spaces

In the previous sections, we implicitly assumed that the matrix of perturbations $\mathbf{X}$ and the corresponding ETLMs were defined in the space of the gridded model. However, it might be more effective to localize the ETLM in a spectral space that is a linear transformation of the gridded perturbations:

$$\mathbf{X}^{\text{spectral}} = \Phi^{-1}\mathbf{X}^{\text{gridded}} \quad \text{and} \quad \mathbf{X}^{\text{gridded}} = \Phi\mathbf{X}^{\text{spectral}}, \quad (23)$$

where $\Phi^{-1}$ transforms the state from gridded space to spectral space and $\Phi$ is its inverse. Using spectral transform in Eq. (23), we can define spectral version of $\mathbf{M}^\text{M3}$ ETLM as follows:

$$\{ \mathbf{M}^\text{M1, spectral}_{m,n} = [\gamma \mathbf{C}^\text{M1, spectral} \odot (\mathbf{M}^\text{M0, spectral}_{m,n,m-1})] \times \cdots \times [\gamma \mathbf{C}^\text{M1, spectral} \odot (\mathbf{M}^\text{M0, spectral}_{1,0})] \} \times \mathbf{X}^{\text{spectral}}_{i}^{-1} \times (\mathbf{X}^{\text{spectral}}_{i})^T. \quad (24)$$

It is trivial to derive similar spectral-space definitions for the $\mathbf{M}^\text{M2}$ and $\mathbf{M}^\text{M3}$ ETLMs.

One spectral space that is especially amenable for ETLM localization is the eigen basis of the ETLM model:

$$\mathbf{M}_{ij} = \mathbf{E}\mathbf{\Sigma}\mathbf{E}^{-1}, \quad (25)$$

where $\Phi = \mathbf{E}$ is the complete set of (complex) eigenvectors of the ETLM and the diagonal matrix $\mathbf{\Sigma}$ lists the (complex) eigenvalues. In this spectral space ($\Phi = \mathbf{E}$) the operation of the ETLM is reduced to multiplication of each of the spectral components of the perturbation by a complex scalar:

$$\delta \mathbf{x}^{\text{spectral}}_{j} = \mathbf{E}^{-1}\mathbf{M}_{ij}\delta \mathbf{x}_{j} = \mathbf{E}^{-1}\mathbf{M}_{ij}\mathbf{E}\delta \mathbf{x}^{\text{spectral}}_{i} = \mathbf{\Sigma}\delta \mathbf{x}^{\text{spectral}}_{j} \quad (26)$$

and, hence, the eigenspace-localized ETLM is precisely diagonal. For the dynamics considered in this paper, Fourier basis provides the eigenspace of the ETLM model. It is likely that for large-scale models, heuristic basis can be found that will result in highly sparse (but not diagonal) representation of the ETLM.

To localize the eigenvector ETLM, we used the same TW and GC functions as for the gridpoint ETLM, but we specified localization distance in terms of wavenumbers (row 4 of Table 1). For example, the $\mathbf{C}^\text{M3}$ localization function for initial covariance decayed to zero for any two wavenumbers that were 29 wavenumbers apart.

e. Local ETLM computation through patch decomposition

To reduce the rank-deficiency of the ETLM estimation problem, we introduce the local version of the ETLM computation (LETLM). For the LETLM, we divide the computational domain into a complete set of overlapping influence volumes (e.g., see our one-dimensional circular domain in Fig. 2). Each influence volume consists of a central patch region that contains the variables to be updated (unique to each volume) and an outer halo region that contains additional variables likely to influence the evolution of variables within the patch. For each influence volume, we will build an individual ETLM operator (LETLM) that will have a more favorable ratio (than a global ETLM) between the number of available ensemble members and the number of nonzero elements in the LETLM operator. Each influence volume is analogous to the observation volume terminology used in the local ensemble transform Kalman filter (Hunt et al. 2007) and in the consistent hybrid ensemble Kalman filter (Bishop et al. 2015).

The halo region of the influence volume is designed to be large enough to contain the regional information required to predict evolution of the state within the patch region (we assume that this influence can be localized for one nonrecursive step of the ETLM model). For example, compare the LETLM in Fig. 3a with the global solutions in Figs. 3b and 3c. The LETLM elements within the patch (shown with the white rectangle in Fig. 3a) are fully resolved while the bottom five grid points in the halo region clearly show unresolved processes associated with advection of information from outside of the influence volume.

To develop a formal notation for the LETLM, we first define a selection operator $\mathbf{S}_p$ for each influence volume $p$ that maps the global state vector $\mathbf{x}$ to the $p$th local influence volume state vector $\mathbf{x}(p)$ as follows:
TABLE 1. Parameters of the ETLM used in presented experiments. Here GC stands for Gaspari–Cohn localization and TW stands for Tukey window localization. The localization distance $L$ (in units of grid points, g.p.) and for the row 4 in experiment (a). The spacial decorrelation length scale for initial ensemble members is never specified directly; instead the spectral parameter is specified using formulas in Table 2. However, we list the equivalent grid-space decorrelation scales (rightmost column) for reference.

<table>
<thead>
<tr>
<th>Line No.</th>
<th>Section of the manuscript</th>
<th>Figure No.</th>
<th>Domain size (grid points, n)</th>
<th>Localization type</th>
<th>$d$ (spectral space)</th>
<th>Correlation length (grid points)</th>
<th>Ensemble initial condition (grid space, $n$)</th>
<th>Correlation type</th>
<th>$C_{0,0}$</th>
<th>$C_{0,1}$</th>
<th>$C_{1,0}$</th>
<th>$C_{1,1}$</th>
<th>$C_{2,0}$</th>
<th>$C_{2,1}$</th>
<th>$C_{3,0}$</th>
<th>$C_{3,1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S1</td>
<td>49</td>
<td>Grid space GC 8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Ensemble initial condition (Table 2)</td>
<td>Correlation type</td>
<td>$C_{0,0}$</td>
<td>$C_{0,1}$</td>
<td>$C_{1,0}$</td>
<td>$C_{1,1}$</td>
<td>$C_{2,0}$</td>
<td>$C_{2,1}$</td>
<td>$C_{3,0}$</td>
<td>$C_{3,1}$</td>
</tr>
<tr>
<td>2</td>
<td>S1</td>
<td>49</td>
<td>Grid space GC 14</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Ensemble initial condition (Table 2)</td>
<td>Correlation type</td>
<td>$C_{0,0}$</td>
<td>$C_{0,1}$</td>
<td>$C_{1,0}$</td>
<td>$C_{1,1}$</td>
<td>$C_{2,0}$</td>
<td>$C_{2,1}$</td>
<td>$C_{3,0}$</td>
<td>$C_{3,1}$</td>
</tr>
<tr>
<td>3</td>
<td>S1</td>
<td>49</td>
<td>Grid space GC 29</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Ensemble initial condition (Table 2)</td>
<td>Correlation type</td>
<td>$C_{0,0}$</td>
<td>$C_{0,1}$</td>
<td>$C_{1,0}$</td>
<td>$C_{1,1}$</td>
<td>$C_{2,0}$</td>
<td>$C_{2,1}$</td>
<td>$C_{3,0}$</td>
<td>$C_{3,1}$</td>
</tr>
<tr>
<td>4</td>
<td>S1</td>
<td>49</td>
<td>Grid space GC 57</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Ensemble initial condition (Table 2)</td>
<td>Correlation type</td>
<td>$C_{0,0}$</td>
<td>$C_{0,1}$</td>
<td>$C_{1,0}$</td>
<td>$C_{1,1}$</td>
<td>$C_{2,0}$</td>
<td>$C_{2,1}$</td>
<td>$C_{3,0}$</td>
<td>$C_{3,1}$</td>
</tr>
<tr>
<td>5</td>
<td>S1</td>
<td>49</td>
<td>Grid space GC 29</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Ensemble initial condition (Table 2)</td>
<td>Correlation type</td>
<td>$C_{0,0}$</td>
<td>$C_{0,1}$</td>
<td>$C_{1,0}$</td>
<td>$C_{1,1}$</td>
<td>$C_{2,0}$</td>
<td>$C_{2,1}$</td>
<td>$C_{3,0}$</td>
<td>$C_{3,1}$</td>
</tr>
</tbody>
</table>

where $S_p^{\text{halo}}$ and $S_p^{\text{patch}}$ selects just the halo and the patch points, respectively, of the influence volume. We restrict selection operators to be orthonormal matrices, where $S_p(S_p)^T = I$.

For each influence volume $p$, we will define an LETLM operator $[M_{m(m-1)}(p)]$ that updates points within the patch region $[x_m^{\text{patch}}(p)]$ of this influence volume based on all the points (patch and halo) of this influence volume at the previous time step $[x_{m-1}(p) = S_m^{\text{halo}} x_{m-1}]$:

$$x_m^{\text{patch}}(p) = M_{m(m-1)}(p) S_m^{\text{halo}} x_{m-1}.$$  \hspace{1cm} (28)

Each $M_{m(m-1)}(p)$ can be computed by either $M_0^{\text{LETLM}}, M_1^{\text{LETLM}}, M_2^{\text{LETLM}},$ or $M_3^{\text{LETLM}}$ ETLMs as follows:

$$M_{m(m-1)}^{\text{LETLM}}(p) = (S_p^{\text{patch}} C_{m-1}^{\text{LETLM}}(p) S_p^{\text{patch}})^T,$$  \hspace{1cm} (29)

or

$$M_{m(m-1)}^{\text{LETLM}}(p) = (S_p^{\text{patch}} C_{m-1}^{\text{LETLM}}(p) S_p^{\text{patch}})^T,$$  \hspace{1cm} (30)

$$M_{m(m-1)}^{\text{LETLM}}(p) = (S_p^{\text{patch}} C_{m-1}^{\text{LETLM}}(p) S_p^{\text{patch}})^T.$$  \hspace{1cm} (31)

$$M_{m(m-1)}^{\text{LETLM}}(p) = (S_p^{\text{patch}} C_{m-1}^{\text{LETLM}}(p) S_p^{\text{patch}})^T.$$  \hspace{1cm} (32)

Notice, that because LETLM operator imposes an implicit localization (by excluding points outside of the influence volume), application of the $M_{m(m-1)}^{\text{LETLM}}(p)$ ETLM operator results in a localized ETLM.

To assemble solutions for individual influence volumes in Eq. (28) into a global solution, we will employ the following summation operator:

$$x_m = \sum_{p=1}^{n_{\text{patch}}} [(S_p^{\text{patch}})^T M_{m(m-1)}^{\text{LETLM}}(p) (S_p x_{m-1})].$$  \hspace{1cm} (33)

Where the $(S_p^{\text{patch}})^T$ operator embeds the local patch solution $x_m^{\text{patch}}(p)$ onto a global grid. In the special case, when the size of the patch corresponds to an individual grid point, the operator $(S_p^{\text{patch}})^T M_{m(m-1)}^{\text{LETLM}}(p) (S_p x_{m-1})$ computes individual rows of the TLM matrix.

It is important to recognize that Eq. (33) can be implemented efficiently on massively parallel computers by computing each element of the summation operator.
\[ \sum_{\text{patch}} \text{in Eq. (33)} \text{ independently of each other. Hence, Eq. (33) is like the local ensemble transform Kalman filter because it updates the individual patches of the grid using only the observations within a certain localization radius. The independent patch update in Eq. (33) makes ETLM computation also compatible with the computational decomposition of the consistent hybrid ensemble filter.} \]

4. Experiment setup

a. Dynamical model

To test the developed ETLM methods, we used three dynamical models that describe evolution of a scalar quantity on a one-dimensional periodic domain. The dynamics of these models were defined in the spectral space of Fourier coefficients:

**FIG. 2.** Patch decomposition of the circular domain.

**FIG. 3.** Example of (a) an LETLM patch compared to (b) the global ETLM solution and, to (c) ETLM computed with a very large ensemble. ETLM is computed for the domain with 399 points and nondispersive dynamics (only the first 50 are plotted. LETLM in (a) is computed for the first 50 points. Need for the halo region is clearly visible at the bottom of the patch region in (a). The white horizontal lines in (a) delineate the patch region from the halo regions.
where \( \omega(p) \) are the wavenumber-dependent speeds specified in Table 2, \( n \) is the number of independent sinusoidal wave functions, and \( p \) is the index of the wavenumber. Because \( \omega(p) \) was wavenumber dependent, it allowed each Fourier frequency to propagate with different speeds and, hence, for dispersion. However, because the spectral model in Eq. (34) was diagonal, there was no exchange of energy between different frequencies (i.e., the model was linear).

The conversion between physical (gridded space) and spectral (Fourier space) representations were defined using Fourier basis \( \Phi \):

\[
\begin{align*}
x_{i+1}^{\text{spectral}} &= M_{i+1,i}^{\text{spectral}} x_i \\
x_{i}^{\text{gridded}} &= \text{real}(\Phi x_i^{\text{spectral}}) \\
M_{i+1,i}^{\text{gridded}} &= \text{real}(\Phi M_{i+1,i}^{\text{spectral}} \Phi^H) \\
\Phi &= \text{ifft}(\sqrt{n!}),
\end{align*}
\tag{35}
\]

where \( \text{ifft}(\cdot) \) is the inverse Fourier transformation. Note that because of the orthonormal properties of the Fourier basis, \( \Phi^{-1} = \Phi^H \) where \( \Phi^H \) is the Hermitian transpose of \( \Phi \).

To create random ensemble perturbations with known grid-space covariance \( P \), we first defined a true forecast error covariance matrix of the following form:

\[
P = \Phi \Lambda \Phi^H,
\tag{36}
\]

where \( \Lambda \) is a diagonal matrix specifying variances as a function of wavenumber. To create a random sample perturbation \( x^{\text{sample \_gridded}} \) we then used:

\[
x^{\text{sample \_gridded}} = \Phi \Lambda^{1/2} a^{\text{sample \_spectral}},
\tag{37}
\]

where \( a^{\text{sample \_spectral}} \) is a random vector with each element drawn from a random normal distribution with mean zero and variance 1 [i.e., \( a^{\text{sample \_gridded}} \sim N(0, I) \)]. It can be readily shown that the covariance of perturbations generated using Eq. (37) is \( P \). The prescribed initial variances \( \Lambda \) are specified for each model in Table 2. The specific width \( d \) of the power spectrum that was used for each of the experiments described in the results section 5 (and the implied correlation scale for this width) is listed in Table 1.

We used three different models for the wavenumber-dependent phase speed \( \omega(p) \) (Table 2). The non-dispersive model was equivalent to uniform advection of the original initial condition on a periodic domain (Fig. 4IA, line 1 in Table 2). The Rossby dispersion relationship propagated different wavenumbers with different speeds, which caused a compact initial condition in the gridded space to disperse with time (Fig. 4IB, line 2 in Table 2). In the Eady-edge wave case, the original wave package remained somewhat localized in space but the shape of the wave changed dramatically from time to time (Fig. 4IC, line 3 in Table 2). The Rossby model was designed to simulate dispersion of barotropic Rossby waves (Holton 2004) and the Eady-edge model was designed to simulate the dispersion of the baroclinic Eady-edge waves (Bishop and Heifetz 2000).

All of the models in Table 2 are linear in spectral and physical (grid) space and, hence, their TLM models are equivalent to the forward operator and can be computed analytically. A column of the TLM operator shows how a unit perturbation at one grid point maps to the rest of the domain at a later time. A column for each of the three grid-space TLM models used in this study is presented in Fig. 4Ia–c. For a short time window, \( \tau = 5 \) TLM models for each of the dispersion relationships are
compactly supported (not shown). However, for a longer time window \( (t = 75) \), only the TLM for the non-dispersive case (Fig. 4IIa) remains compactly supported, while the TLMs for the Rossby and Eady-edge models are no longer supported (Figs. 4IIb and 4IIc).

### b. Error metrics

To measure the accuracy of the propagated covariance, we use the normalized mean square error (NMSE):

\[
\text{NMSE} = \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} [P_{m,0}^{\text{true}}(k,l) - P_{m,0}^{\text{ETLM}}(k,l)]^2}{\sum_{k=1}^{n} \sum_{l=1}^{n} [P_{m,0}^{\text{true}}(k,l)]^2},
\]

where \( P_{m,0}^{\text{true}} \) and \( P_{m,0}^{\text{ETLM}} \) are the climatological covariances propagated with the perfect model and the ETLM:

\[
P_{m,0}^{\text{true}} = M_{m,0}^{\text{true}} P_{00}^{3D \text{ clim}},
\]

\[
P_{m,0}^{\text{ETLM}} = M_{m,0}^{\text{ETLM}} P_{00}^{3D \text{ clim}}.
\]

(b) Error metrics

For the propagated covariance, we use the normalized mean square error (NMSE):

\[
\text{NMSE} = \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} [P_{m,0}^{\text{true}}(k,l) - P_{m,0}^{\text{ETLM}}(k,l)]^2}{\sum_{k=1}^{n} \sum_{l=1}^{n} [P_{m,0}^{\text{true}}(k,l)]^2}.
\]

where \( P_{m,0}^{\text{true}} \) and \( P_{m,0}^{\text{ETLM}} \) are the climatological covariances propagated with the perfect model and the ETLM:

\[
P_{m,0}^{\text{true}} = M_{m,0}^{\text{true}} P_{00}^{3D \text{ clim}},
\]

\[
P_{m,0}^{\text{ETLM}} = M_{m,0}^{\text{ETLM}} P_{00}^{3D \text{ clim}}.
\]

or the localized ensemble covariance localized at the initial time and propagated with the perfect model and the ETLM:

\[
P_{m,0}^{\text{true}} = M_{m,0}^{\text{true}} (C_{00}^{3D \text{ ens}} \odot P_{00}^{3D \text{ ens}}),
\]

\[
P_{m,0}^{\text{ETLM}} = M_{m,0}^{\text{ETLM}} (C_{00}^{3D \text{ ens}} \odot P_{00}^{3D \text{ ens}}).
\]
NMSE, essentially measures the L2 norm of the difference in all elements of the covariance advected with the ETLM $\mathbf{P}_{m,0}^{\text{ETLM}}$ as compared to the covariance advected with the perfect linear model $\mathbf{P}_{m,0}^{\text{true}}$. NMSE scores are normalized by the mean square of all the elements in the true covariance. The NMSE score of 1 means that the approximation error is as large as the entire signal of the true covariance matrix.

5. Results

a. Rank deficiency of nonlocalized ETLM

We found that when we used a full-rank ETLM, spurious temporal correlations were not an obstacle for propagating initial perturbations accurately. In fact, when a single column of a nonlocalized, four-dimensional covariance $\mathbf{P}_{5,0}^{\text{ens}}$ is plotted in Fig. 5a (red...
it clearly shows sampling noise when compared to a similar covariance computed from a very large four-dimensional ensemble (black line). (The red line in Fig. 5awas computed using 50 members, while the black line was computed using 10 000 members.) However, further investigation showed that most of the undesired noise was due to ensemble sampling error at the initial time $t = 0$ and not due to the temporal noise in the ETLM. To show this, we compared the raw four-dimensional covariance (Fig. 5a) with the four-dimensional covariance that was produced by applying unlocalized ETLM $M_{M0}$ to localized initial time covariance ($C_{0,0} P_{ens0,0}$) (shown with the red line in Fig. 5b). When we localized the noisy covariance at the initial time (cf. green lines in Figs. 5a,b), we were able to propagate the initial-time structure through time perfectly. Essentially, the ETLM-propagated red line in Fig. 5b is almost exactly on top of the blue line, which was propagated with the perfect TLM.

While the preliminary results shown in Fig. 5b are encouraging, they are too optimistic for realistic applications in large computational domains. In fact, if non-localized (low rank) ETLM is applied to a full-rank initial time covariance the resulting four-dimensional covariance:

$$P_{m,m} = M_{m,0} P_{0,0} (M_{m,0})^T,$$

is confined to the low-dimensional ensemble subspace and, as a result, can only use the number of degrees of freedom in the ensemble to model all of the complexities in the error field of a high-dimensional model.

Figure 5c shows that, in fact, when we increase the size of the model domain from $n = 50$ to $n = 250$, but keep the ensemble size to 50, the resulting four-dimensional covariance $P_{5,0}$ becomes noisy again (cf. red lines in Figs. 5b,c). This increase in noise can only be explained by the rank deficiency of the four-dimensional covariance $P_{5,0} = M_{5,0} (C_{3D ens0,0} P_{3D ens0,0})$ produced by the rank deficient ETLM. From results in Figs. 5a and 5b, we know that 50 ensemble members are sufficient to capture the spatial translation of the initial covariance function of this width. However, because the domain in Fig. 5c is 5 times longer, the 50 ensemble members can no longer form a complete orthogonal basis, which leads to rank deficiency and noise.

b. Localization in the original grid space

1) ETLMs for propagation of the 3D climatological covariance

Figure 6 shows that ETLMs can successfully propagate high-rank climatological covariance in time. Compared to a reference solution of a static three-dimensional covariance ($P_{3D clim0}$ shown with the black line), climatological covariance propagated using ETLMs have significantly lower NMSE (colored lines). For example, when the initial covariance is propagated one decorrelation length scale, the NMSE error of using a three-dimensional covariance is around 1.5. In
contrast, the NMSE of the covariance propagated with $\mathbf{M}^{M1}$ ETLM (blue line) is as low as 0.07 for the non-dispersive case and 0.08 for the Rossby dispersion case.

Figure 6 shows that, among the three dispersion models, the M3 localization (red line) has consistently the lowest NMSE scores among all localized global ETLMs, with the more computationally efficient M1 localization being a close second. After two-to-three decorrelation scales, the nonlocalized ETLM $\mathbf{M}^{M0}$ (green line) becomes more accurate than either of the localized ETLMs. However, as we discussed in section 5a, the low error of the nonlocalized $\mathbf{M}^{M0}$ ETLM is likely to be an artifact of this low-dimensional benchmark problem and is unlikely to remain as low for high-dimensional weather forecast models.

The comparisons between the three dispersion models (presented in Fig. 6) show that grid-space-localized ETLMs are most accurate for nondispersive dynamics (Fig. 6a) and are less accurate for the Rossby and Eady-edge dispersion dynamics (Figs. 6b and 6c). To highlight these differences we plotted one column of the propagated climatological covariance valid at time sufficient to shift the initial condition by one correlation length (Fig. 7a) and three correlation lengths (Fig. 7b). For illustration purposes, we only show Rossby dispersion model as it had the most exaggerated representation of the error growing modes. Early in the temporal evolution (Fig. 7a), the initial covariance structure was dispersed only slightly and all of the ETLMs captured the advected column of the initial covariance well. Comparisons between unlocalized ($\mathbf{M}^{M0}$) and localized ($\mathbf{M}^{M1}$, $\mathbf{M}^{M2}$, and $\mathbf{M}^{M3}$) ETLMs show that the unlocalized solution has low-level noise distributed throughout the domain, while localized ETLMs produce noise-free covariance away from the main covariance structure.

At longer lead times (three correlation length scales in Fig. 7b), the initial condition is both advected and dispersed (solid black line). The nonlocalized $\mathbf{M}^{M0}$ ETLM captures the dispersion of the initial condition well; however, at the cost of low-level noise throughout the domain. Unfortunately, as we discussed in section 5a, this perfection is unlikely to be maintained for realistic high-rank weather prediction models. The localized ETLMs ($\mathbf{M}^{M1}$, $\mathbf{M}^{M2}$, and $\mathbf{M}^{M3}$) capture the main signal well, but fail to capture the details of the dispersed waves away from the main signal.

2) **ETLM-BASED PROPAGATION OF THE ENSEMBLE COVARIANCE LOCALIZED AT THE INITIAL TIME**

Figure 8 shows that ETLMs are capable of propagating ensemble-based covariances localized at the initial time. The localized ensemble covariances propagated with different ETLM models (colored lines in Fig. 8) have lower NMSE scores than either the nonlocalized four-dimensional covariance (dashed line) or the four-dimensional covariance localized using static localization (line with open circles). Results in Fig. 8 confirm that recursive application of the localized ETLM essentially constructs a four-dimensional localization function. Snapshots of the four-dimensional covariance in Fig. 9 further illustrate this capability. For example, compare the evolution of the covariance with M1 localization [panels labeled $\mathbf{M}^{M1}(\mathbf{C}_c^{P3D\text{ ens}})$] with the application of a static localization function (panels labeled $\mathbf{C}_c^{P3D\text{ ens}}$). The M1 localization was clearly more effective at both retaining the advected signal and at reducing the noise away from the location of the main signal [cf. panels for unlocalized $\mathbf{M}^{M0}$ ETLM labeled $\mathbf{M}^{M0}(\mathbf{C}_c^{P3D\text{ ens}})$].

c. **Localization in the eigenvector space**

Figure 10 shows that ETLMs localized in the eigenvector space can capture dispersive dynamics at longer lead times (farther than one-to-two correlation length scales). When the same number of ensemble members was used (30) as in the case of the grid-space-localized ETLMs (Fig. 10a), the NMSE errors were lower for the eigenspace-localized ETLM than the ETLM localized in the grid space (cf. Figs. 10a and 6b). If we take the climatological covariance, M1 localization, and Rossby dispersion dynamics as an example, the NMSE error at four correlation length scales was 0.23 for the eigenspace-localized ETLM and was 0.47 for the grid-space ETLM. Even more remarkably, the M2 and M3 localizations were effectively perfect (NMSE $\sim 10^{-10}$).

The benefits of the localization in the eigenspace of the ETLM case were even clearer when we constructed the ETLMs using only two ensemble members (Fig. 10b). In this case, M2 and M3 localization remain still very effective. The fidelity of the eigenspace-localized ETLM is apparent when we plot the covariance structures at very long lead times, see Fig. 11. Even at these long lead times the M2 and M3 localizations capture the dispersion relationships perfectly with as few as two ensemble members (Fig. 11b).

To understand why eigenspace localization was more effective it is important to recall that in the eigenspace, the true TLM models for the considered benchmark problems can be expressed as complex diagonal matrices [Eq. (26)]. The diagonal elements in these matrices have all the information necessary to encode the dispersion relationships in the eigen (Fourier) basis. Localization of these diagonal matrices becomes trivial as shown in Fig. 12 (cf. Fig. 12a for the true TLM and...
FIG. 7. Four-dimensional covariances valid at (a) $t = 10$ (one decorrelation scale) and (b) $t = 30$ (three decorrelation scale) for $M^x P_{\text{clim}}$ for $x = 0, 1, 2,$ and $3$. Only plots for the Rossby dispersion model are shown. The dashed black line is the climatological covariance at initial time computed with 10,000 ensemble members. The black line is the four-dimensional covariance propagated to the future time with a perfect model. The red line is the four-dimensional covariance propagated to the future time with an ETLM. Covariances are shown with gridpoint 50. ETLMs are computed using 30 ensemble members. The correlation length scale of the ensemble was 10 grid points.
FIG. 8. As in Fig. 6, but for flow-dependent error covariance (true covariance computed using 10,000 members). Approximations are based on a raw nonlocalized ensemble of 30 members (line with dots), raw ensemble localized using static localization (circles), and a three-dimensional climatological covariance localized at the initial time propagated with ETLMs (colored lines). NMSEs were computed using Eqs. (38) and (40).

Fig. 12b for the ETLM). Figure 12 also shows the benefit of inverting the localized matrix \( \mathbf{C}^{-1} \) used by the M2 and M3 localizations (Eqs. 21 and 22) as compared to inverting nonlocalized matrix \( (\mathbf{X}^T \mathbf{X})^{-1} \) used by the M0 and M1 localization in Eqs. (13) and (19) (cf. Figs. 12c and 12d).

Figure 12 also highlights an interesting feature of the ETLM models. The dynamic TLM model (Fig. 12a) is a diagonal matrix with nonzero numbers present along the entire diagonal (for all wavenumbers). In contrast, the \( \mathbf{M}^{00} \) ETLM defined using a very large number of ensemble members (Fig. 12b) and the \( \mathbf{M}^{03} \) ETLM defined using two members (Fig. 12d) have a part of the diagonal that contain zeros (for wavenumbers with absolute value greater than \( \sim 65 \)). It turns out that the variance of the initial ensemble at these wave frequencies was at, or lower, than the machine precision. In essence, we designed the ensemble to be smoother than these wavenumbers.

d. LETLM results

Our computational results with various patch and halo sizes showed that patch decomposition can result in considerably lower NMSE errors for both static and flow-dependent covariances. The results for the climatological error covariance are plotted in Fig. 13. The NMSE errors were considerably lower for \( \mathbf{M}^{00} \) and \( \mathbf{M}^{01} \) LETLM for all of the dynamic models. (Nondispersive dynamics is shown in Figs. 13a–c and Rossby dynamics is shown in Figs. 13d–f. Eady-edge results were similar to the Rossby results and are not shown.) However, for dispersive dynamics, it was important to use large halo regions to avoid fatal corruption of the ETLM results. For example, in the case of the Rossby dynamics (Figs. 13d–f) and the patch size of 1, we needed to use a halo size of greater than 30 points to ensure that the \( \mathbf{M}^{00} \) and \( \mathbf{M}^{01} \) LETLM results remain stable. The need for the balance between the size of the halo and the speed of the propagating waves is analogous to the Courant–Friedrichs–Lewy condition in the traditional numerical analysis, where the propagating wave should not cross more than one grid cell in a single time step. In the case of the \( \mathbf{M}^{03} \) LETLM, the NMSE error was the same as the NMSE error of the global solution (Figs. 13c and 13f). The accuracy of the \( \mathbf{M}^{03} \) LETLM was not very sensitive to the selection of the patch and halo size. We attribute this insensitivity of the \( \mathbf{M}^{03} \) LETLM to the size of the patch to the fact that global \( \mathbf{M}^{03} \) ETLM uses full rank, localized matrices in its computations.

Data in Fig. 13 show that when the size of the halo and patch region were selected properly, the smaller patches produced lower NMSE errors for \( \mathbf{M}^{00} \) and \( \mathbf{M}^{01} \).
(a) Four-dimensional covariances valid at t=10

Fig. 9. As in Fig. 7, but for $P^{4D}_{ens}$, $C^{loc} \otimes P^{4D}_{ens}$, and $M^{Mx}(C^{loc} \otimes P^{4D}_{ens})$ for $x = 0, 1, 2, 3$. The dashed black line is the sample ensemble covariance of 30 members localized at the initial time.

(b) Four-dimensional covariances valid at t=30
LETLMs. This finding is consistent with our hypothesis that smaller size of each $M^p$ problem results in fewer nonzero elements that needed to be estimated and, hence, better conditioning of the estimation problem for a given size of the ensemble. This conclusion is further supported by lower $M^0$ and $M^1$ LETLM errors (colored lines in Figs. 13a,b and 13d,e) as compared to the global $M^0$ and $M^1$ ETLM errors (dashed black lines in Figs. 13a,b and 13d,e).

To isolate the impact of the implicit localization due to patch decomposition from the impact of the M1 localization of each patch, we compared the NMSE errors for the global $M^0$ and $M^1$ ETLMs (solid lines in Fig. 14) with the errors of the $M^0$ and $M^1$ LETLMs (lines with circles in Fig. 14). We found that most of the reduction in the NMSE error can be attributed to patch decomposition and increase in the rank of each LETLM solution (cf. the solid green line for the global $M^0$ and the green line with circles for local $M^0$). The M1 localization of each individual LETLM provided an additional small positive benefit (cf. the green and red lines with circles for $M^0$ and $M^1$ LETLM, respectively).

Our estimate of the computational cost of the $M^1$ and $M^3$ LETLMs showed that, for the domain size used in this study, the computational cost of $M^1$ LETLM was comparable to the cost of the global solution (Fig. 15a).

However, for the $M^3$ ETLMs, the LETLM algorithm considerably reduced the computational cost (by an order of magnitude or more) because the size of the matrix inverse in Eq. (32) was reduced from $n^{state}$ to $n^{patch}$. Recall that the cost of matrix inverse scales as a cube of the size of the matrix. We did not present the cost of the $M^2$ LETLM because it was very similar to the cost of the $M^3$ LETLM.

6. Summary and conclusions

A new approach has been proposed for propagating hybrid error covariances through time without the use of a traditional dynamics-based TLM. The approach is based on using localized ETLM operators. It has been demonstrated that it is possible to construct four-dimensionally localized ETLM operators as a recursive product of short-duration ETLM operators, each of which is localized using a simple static localization.

Our results indicate that it was the rank deficiency rather than covariance sampling errors (associated with the small ensemble size) that lead to ETLM errors. To remedy the rank deficiency, we propose four different models for localization of ETLMs. The M1 localization increases the rank of the ETLM by enforcing the local
FIG. 11. As in Fig. 7, but covariances valid at $t = 40$ (four decorrelation scales) for Rossby dispersion and eigenspace-localized ETLMs. The black line is the four-dimensional covariance propagated to time $t = 40$ with a perfect model. The red line is the four-dimensional covariance propagated to time $t = 40$ with an ETLM.
nature of the TLM operator through distance-based attenuation of ETLM elements. In contrast, the M2 and the M3 localizations create high-rank ETLMs by localizing individual submatrices in the ETLM definition. And finally, the LETLM solves the rank deficiency problem by computing ETLM on a collection of local subdomains of the global grid; for each such subdomain, fewer ensemble members are required to compute the LETLM operator accurately.

We demonstrated that when \( \mathbf{M}^{M2} \) and \( \mathbf{M}^{M3} \) ETLMs are constructed in the space of the eigenvectors of the dynamics operator, very few ensemble members are required to construct a highly accurate global ETLM. These results clearly support the hypothesis that it is rank deficiency and not the sampling noise that limits the fidelity of the ETLM. Practical applications of the eigenspace localization will be more challenging in global three-dimensional models where it would be impractical to compute eigenvectors of the ETLM. However, it is likely that it might be possible to specify heuristic basis where ETLMs can look very sparse and, hence, will require fewer ensemble members to specify an accurate ETLM.

An alternative to eigenspace localization is to compute ETLMs in the original grid space using global \( \mathbf{M}^{M1} \) ETLM or using the LETLM algorithm. Without patch decomposition, the simple and computationally efficient \( \mathbf{M}^{M1} \) ETLM was effective at propagating covariances up to one-to-two decorrelation scales, even in cases of dispersive dynamics. However, with patch decomposition, we were able to maintain the same level of errors up to four decorrelation scales in the Rossby wave dispersion model. For the nondispersive model, the LETLM perfectly predicted the propagation of perturbations.

In our opinion, the LETLM algorithm presents a very attractive avenue for estimating TLMs for realistic, high-dimensional models of the geophysical flow. In most geophysical models, the future state of the model grid point explicitly depends (through a finite-difference or a finite-element approximation) on the previous state of the model in a vicinity of this point (the “computational stencil” in the numerical modeling terminology). Hence, over a single step of a numerical model, the LETLM can be constructed perfectly when the number of ensemble members is greater than or equal to the
number of grid points in the computational stencil. (A demonstration of this property is not explicitly shown in this paper but will appear in our follow-up papers.) In practice, it may be computationally advantageous to construct our short-duration ETLM operators [in the recursive Eq. (17)] for periods longer than a single time step of the model (e.g., over a 1-hour interval in atmospheric models). Over these longer intervals, the volume of the stencil of the model expands with the cube of time; hence, the number of grid points in the stencil can quickly outpace the number of available ensemble members. However, for such longer intervals it might still be possible to construct an accurate ETLM approximation if the number of ensemble members is comparable to the number of degrees of freedom that determine the future state of a single grid point. Further

![Graph](image1)

**Fig. 13.** NMSE for global and local ETLM: (a)–(c) nondispersive dynamics and (d)–(f) Rossby dynamics: (left to right) for $M^M_k$ for $x = 0, 1,$ and 3. NMSE for global ETLM solutions are shown in black dashed lines and NMSE for LETLM are shown in color. ETLMs were run to $t = 40$ at a $t = 5$ recursive time step on a 399-point periodic grid. The NMSE were calculated for climatological covariance using Eqs. (38) and (39). NMSE errors are shown off scale when ETLM results diverged [e.g., in (d) and (e) for a patch size of 1 and a halo size of 30].

![Graph](image2)

**FIG. 14.** NMSE for Rossby dynamics. ETLMs were run to $t = 40$ at a $t = 5$ recursive time step on a 399-point periodic grid. For LETLMs, the patch size was 1 grid point and the halo size was 60 points. The NMSE were calculated for climatological covariance using Eqs. (38) and (39).
research for specific applications will be required to determine the best practical balance between the length of the ETLM time step and the number of ensemble members.

Results displayed in this paper were obtained for the simplistic models of linear wave advection. In our future work we will investigate if these positive results are relevant for realistic models of oceanic and atmospheric dynamics. Specifically, we hope that ETLM computations can improve the performance of current 4DEnVar algorithms (Buehner et al. 2010a; Lorenc et al. 2015) by propagating the static part of the hybrid background error covariance matrix. This advantage might be especially relevant for coupled ocean–atmosphere DA, where traditional TLMs are often not available.

Acknowledgments. This work was supported by the Office of Naval Research Award N0001412WX20323. We are grateful to Douglas Allen from NRL-DC for providing comments on the early version of this manuscript. We are grateful to Chris Snyder and two other anonymous reviewers for their insightful comments that helped to improve the quality of this manuscript.

REFERENCES


