SIAM/ASA J. Uncertainty Quantification
Vol. 10, No. 3, pp. 886–914

Continuum Covariance Propagation for Understanding Variance Loss in Advective Systems*

Shay Gilpin†, Tomoko Matsuo‡, and Stephen E. Cohn§

Abstract. Motivated by the spurious variance loss encountered during covariance propagation in atmospheric and other large-scale data assimilation systems, we consider the problem for state dynamics governed by the continuity and related hyperbolic partial differential equations. This loss of variance has been attributed to reduced-rank representations of the covariance matrix, as in ensemble methods for example, or else to the use of dissipative numerical methods. Through a combination of analytical work and numerical experiments, we demonstrate that significant variance loss, as well as gain, typically occurs during covariance propagation, even at full rank. The cause of this unusual behavior is a discontinuous change in the continuum covariance dynamics as correlation lengths become small, for instance in the vicinity of sharp gradients in the velocity field. This discontinuity in the covariance dynamics arises from hyperbolicity: the diagonal of the kernel of the covariance operator is a characteristic surface for advective dynamics. Our numerical experiments demonstrate that standard numerical methods for evolving the state are not adequate for propagating the covariance, because they do not capture the discontinuity in the continuum covariance dynamics as correlations lengths tend to zero. Our analytical and numerical results show that this leads to significant, spurious variance loss in certain regions and gain in others. The results suggest that developing local covariance propagation methods designed specifically to capture covariance evolution near the diagonal may prove a useful alternative to current methods of covariance propagation.

Key words. covariance propagation, variance loss, data assimilation, advective systems

AMS subject classifications. 35L65, 47B38, 65M22, 86A10, 65M75, 65M32, 65M80

DOI. 10.1137/21M1442449

* Received by the editors August 26, 2021; accepted for publication (in revised form) February 24, 2022; published electronically August 03, 2022.

https://doi.org/10.1137/21M1442449

Funding: This work was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under grant DGE-1650115. The work of the second author was supported by the NSF CAREER Program under grant AGS-1848544. The work of the third author was supported by the National Aeronautics and Space Administration (NASA) under Modeling, Analysis and Prediction (MAP) program Core funding to the Global Modeling and Assimilation Office (GMAO) at Goddard Space Flight Center (GSFC). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or of NASA.

† Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO 80303-0429 USA (shay.gilpin@colorado.edu).

‡ Ann and H. J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO 80303-0429 USA (tomoko.matsuo@colorado.edu).

§ Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (stephen.e.cohn@nasa.gov).

1. Introduction. At the heart of modern data assimilation is covariance propagation. Data assimilation techniques evolve the estimation error covariance along with the model state, either explicitly as in the Kalman filter, implicitly as in variational methods, or using a reduced-rank approximation as in ensemble schemes [22, 38, 9, 13]. To provide context for the problem addressed in this paper, we start with a stochastic model state N-vector $\boldsymbol{q}$, which is propagated discretely in data assimilation schemes from time t_{k-1} to t_k as

$$\boldsymbol{q}_k = \boldsymbol{M}_{k,k-1}\boldsymbol{q}_{k-1}, \tag{1.1}$$

where $\boldsymbol{M}_{k,k-1}$ is the deterministic $N\times N$ propagation matrix representing the model dynamics. For simplicity, we consider here the linear case with no forcing, random or otherwise. From the model state we can define the $N \times N$ symmetric positive semidefinite covariance matrix at time t_k,

$$\boldsymbol{P}_k = \mathbb{E}[(\boldsymbol{q}_k - \overline{\boldsymbol{q}}_k)(\boldsymbol{q}_k - \overline{\boldsymbol{q}}_k)^T], \tag{1.2}$$

where $\mathbb{E}[\cdot]$ is the expectation operator, $\overline{\boldsymbol{q}}_k = \mathbb{E}[\boldsymbol{q}_k]$ is the mean state, and superscript T denotes transpose. The basic equation of discrete covariance propagation behind modern data assimilation schemes then follows directly from the discrete state propagation (1.1),

$$\boldsymbol{P}_k = \boldsymbol{M}_{k,k-1}\boldsymbol{P}_{k-1}\boldsymbol{M}_{k,k-1}^T, \tag{1.3}$$

where $\boldsymbol{P}_{k-1}$ and $\boldsymbol{P}_k$ are the covariance matrices at times t_{k-1} and t_k, respectively [22, 21, chapter 6]. We omit a process noise term in (1.3), which will be discussed later.

Motivated by atmospheric data assimilation schemes used for global numerical weather prediction (NWP) [11, 23], we consider covariance propagation associated with hyperbolic partial differential equations (PDEs). Let $\boldsymbol{x} \in S_r^2$, where S_r^2 is the surface of the sphere of radius r, and take time $t \geq t_0$. To fix ideas, we first consider the continuity equation, which describes the continuum evolution of the stochastic model state $q = q(\boldsymbol{x}, t)$ as follows:

$$\begin{aligned} q_t + \boldsymbol{v}\cdot\boldsymbol{\nabla} q + (\boldsymbol{\nabla}\cdot\boldsymbol{v})q = 0,\\ q(\boldsymbol{x}, t_0) = q_0(\boldsymbol{x}). \end{aligned} \tag{1.4}$$

The subscript t denotes the time derivative unless noted otherwise. The two-dimensional velocity field $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x}, t)$ is taken to be deterministic and continuously differentiable while the initial state q_0 is stochastic with mean $\overline{q}_0$. Equation (1.4) is the statement of mass conservation when q is the density of a passively advected tracer in a thin layer of atmosphere between two isentropic surfaces, for instance [29, 17, sec. 2.5 – 2.6]. For the model state in (1.4), we can define the covariance between two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S_r^2$ as

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = \mathbb{E}[(q(\boldsymbol{x}_1, t) - \overline{q}(\boldsymbol{x}_1, t))(q(\boldsymbol{x}_2, t) - \overline{q}(\boldsymbol{x}_2, t))]. \tag{1.5}$$

The corresponding covariance evolution equation for (1.4) on $S_r^2 \times S_r^2$ and for $t \geq t_0$ becomes

$$\begin{aligned} P_t + \boldsymbol{v}_1\cdot\boldsymbol{\nabla}_1 P + \boldsymbol{v}_2\cdot\boldsymbol{\nabla}_2 P + (\boldsymbol{\nabla}_1\cdot\boldsymbol{v}_1 + \boldsymbol{\nabla}_2\cdot\boldsymbol{v}_2)P = 0,\\ P(\boldsymbol{x}_1, \boldsymbol{x}_2, t_0) = P_0(\boldsymbol{x}_1, \boldsymbol{x}_2). \end{aligned} \tag{1.6}$$

Here $\boldsymbol{\nabla}_i$ denotes the gradient with respect to $\boldsymbol{x}_i$, and $\boldsymbol{v}_i = \boldsymbol{v}(\boldsymbol{x}_i, t)$, $i = 1, 2$.

Both the model state and covariance equations given in (1.4) and (1.6) are hyperbolic PDEs, the former in two space dimensions and the latter in four. For the covariance equation, the characteristic (or trajectory) equations that describe the coordinate vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of parcels located initially at coordinates $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, respectively, both satisfy the same ordinary differential equation,

$$\begin{aligned} \frac{d\boldsymbol{x}}{dt} &= \boldsymbol{v}(\boldsymbol{x}, t), \\ \boldsymbol{x}(t_0) &= \boldsymbol{s} \end{aligned} \tag{1.7}$$

[37, following Ch. 3 notation]. The coordinate vector $\boldsymbol{x}_i$ for $i = 1, 2$ can be written as $\boldsymbol{x}_i = \boldsymbol{x}(t; \boldsymbol{s}_i)$, which is the solution to (1.7) that represents the arrival point $\boldsymbol{x}_i$ at time t of the characteristic (trajectory) departing from the point $\boldsymbol{s}_i$ at t_0. In the case that initial parameters for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are equal, $\boldsymbol{s}_1 = \boldsymbol{s}_2$, then it follows from (1.7) that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are also equal, $\boldsymbol{x}_1 = \boldsymbol{x}(t; \boldsymbol{s}_1) = \boldsymbol{x}(t; \boldsymbol{s}_2) = \boldsymbol{x}_2$. Therefore, initial covariances that start on the hyperplane $\boldsymbol{x}_1 = \boldsymbol{x}_2$ (i.e., $\boldsymbol{s}_1 = \boldsymbol{s}_2$) remain on $\boldsymbol{x}_1 = \boldsymbol{x}_2$ for all time, implying that the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane is everywhere characteristic [6, p. 3130]. We show in section 2 that as a result, there is a discontinuous change in solutions to (1.6) along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane in the limit as correlation lengths approach zero, for example in the vicinity of sharp gradients in the velocity field which can arise naturally, as seen for instance in [27].

To describe this discontinuity, suppose first that the initial covariance P_0 is continuous on $S_r^2 \times S_r^2$, and denote it by $P_0^d(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Thus, the stochastic initial state q_0 is spatially correlated, with variance $\sigma_0^2(\boldsymbol{x}) = P_0^d(\boldsymbol{x}, \boldsymbol{x})$. It follows that the solution of the covariance evolution equation (1.6) along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane corresponds to the variance $\sigma^2(\boldsymbol{x}, t) = P(\boldsymbol{x}, \boldsymbol{x}, t)$ and satisfies

$$\begin{aligned} \sigma_t^2 + \boldsymbol{v} \cdot \boldsymbol{\nabla}\, \sigma^2 + 2(\boldsymbol{\nabla} \cdot \boldsymbol{v})\sigma^2 &= 0, \\ \sigma^2(\boldsymbol{x}, t_0) = \sigma_0^2(\boldsymbol{x}) &= P_0^d(\boldsymbol{x}, \boldsymbol{x}), \end{aligned} \tag{1.8}$$

where $\sigma^2 = \sigma^2(\boldsymbol{x}, t)$ for $\boldsymbol{x} \in S_r^2$ and $t \geq t_0$, which can be derived either from (1.6) or directly from (1.4) [6, sec. 2].

Now suppose instead that the initial state q_0 is spatially uncorrelated, and denote its covariance function by $P_0^c(\boldsymbol{x}_1)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where δ is the Dirac delta and P_0^c is continuous on S_r^2. As we show in subsection 2.3, for $t \geq t_0$ the solution to (1.6) is then given by $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = P^c(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where P^c satisfies

$$\begin{aligned} P_t^c + \boldsymbol{v} \cdot \boldsymbol{\nabla}\, P^c + (\boldsymbol{\nabla} \cdot \boldsymbol{v})P^c &= 0, \\ P^c(\boldsymbol{x}, t_0) = P_0^c(\boldsymbol{x}). \end{aligned} \tag{1.9}$$

Thus, near zero correlation length, the behavior of solutions of the covariance evolution equation along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane changes abruptly from that of the variance equation (1.8) to that of (1.9), in the case of a nonzero divergent velocity field, $\boldsymbol{\nabla} \cdot \boldsymbol{v} \neq 0$. Such a discontinuous change in covariance dynamics for states governed by the continuity equation (1.4) also holds for states governed by a generalized version of (1.4), which we present in section 2.

The characteristic behavior of the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane in the continuum does not translate into discrete space for typical discretizations (1.1) of (1.4): diagonal elements of $\boldsymbol{P}_k$ in (1.3) depend on off-diagonal elements of $\boldsymbol{P}_{k-1}$, and a diagonal initial covariance matrix in (1.3) does not remain diagonal for all time. In this paper, we study the behavior of solutions of the continuum covariance evolution equation (1.6) and a generalized version (2.8) near the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane, and we contrast this with the behavior of discretizations (1.3) near the diagonal, using a combination of analytical and numerical methods. We conduct numerical experiments using a one-dimensional version of (1.4) and study the covariance and variance propagation as a function of correlation length scales of the initial covariance. We find in some cases that the variance propagated numerically according to (1.3) bears little resemblance to that of the continuum variance dynamics (1.8), exhibiting both variance loss and variance gain relative to the continuum solution. In particular, variances propagated numerically according to (1.3) tend to be better approximations of (1.9) than of (1.8) for short initial correlation lengths, quite independently of any numerical dissipation or dispersion effects. This property manifests itself as a large, spurious loss of variance in regions where the amplitude index m, defined in (3.7) and (3.8) of subsection 3.3, is greater than one, and a spurious variance gain in regions where the index m is less than one, as our analytical and numerical results illustrate. This behavior is the result of the discontinuous change in the continuum covariance dynamics in the limit as the correlation length tends to zero.

Spurious loss of variance is well known in the data assimilation literature. How covariances propagated through data assimilation schemes tend to underestimate the exact error covariances has long been noted [28, pp. 23–24]. Variance loss has been discussed primarily in the context of ensemble schemes [28, 25, 13] in recent years, where spurious variance loss can be attributed to the use of reduced-rank covariance representations [14]. Several methods have been developed to address variance loss to prevent filter divergence, such as covariance inflation [1, 31, 28, Ch. 9.2], a scale-selective generalization of covariance inflation [7, sec. 2.4.4], and methods of discrete covariance propagation that address variance loss associated with numerical discretization [33, 30]. Loss of variance is sometimes addressed through an artificial model error or process noise term added to the discrete covariance propagation in (1.3) [21, sec. 8.8–8.9]. Accurately estimating an appropriate model error/process noise term is difficult because spurious variance loss can be due to several different known and unknown sources, though it has been shown that adding a model error term can help rectify the negative impact of variance loss, for instance, by increasing ensemble spread in the case of the ensemble schemes [31, 19]. Stochastic parameterization of subgrid scale physics also helps to increase ensemble spread to prevent filter divergence [3, p. 567].

For the purpose of illuminating a root cause of variance loss, we consider in this work only the unforced covariance dynamics and omit a model error term, artificial or not. The focus of this paper is on the spurious loss and gain of variance associated with the peculiar, discontinuous limiting behavior of solutions of the continuum covariance evolution equation (1.6) and its generalization (2.8), where spurious variance gain, while not often noted in the literature, can also cause issues during data assimilation. Spurious loss and gain of variance due to this discontinuous limiting behavior is a full-rank effect. We believe ours is the first work to identify and study this effect.

The layout of this paper is as follows. In section 2, we consider a slightly generalized form of the continuity equation to study the continuum covariance propagation. In subsection 2.1 we establish the generalized problem, defining the necessary operators and associated PDEs used for the analysis. This is followed by subsection 2.2, where the continuum polar decomposition is defined for use in subsection 2.3, which derives the generalized version of (1.9) and discusses the discontinuous change in the continuum dynamics as initial correlation lengths approach zero. The analysis sections are followed by numerical experiments, where we illustrate spurious loss and gain of variance in full-rank covariance propagation through a simple one-dimensional example. Subsection 3.1 details the experimental setup of the one-dimensional problem, describing the numerical propagation methods, discretization schemes, and initial covariances. Subsection 3.2 summarizes the results from these numerical experiments, followed by subsection 3.3, which discusses additional interpretations of the numerical experiments. Section 4 contains concluding remarks, followed by Appendices A and B, which contain additional derivations.

2. Analysis.

We first study covariance propagation in the continuum. We will consider the state and covariance equations as PDEs with solutions in the Hilbert space L^2, define the associated linear operators, and use tools from functional analysis to study these equations and operators. This continuum analysis is crucial for interpreting the results of the numerical experiments given later in section 3.

2.1. Preliminaries.

Let $\Omega = S_r^2$ and take $\boldsymbol{x} \in \Omega$ and $t \geq t_0$. We will consider the generalized advection equation for the model state $q = q(\boldsymbol{x}, t)$,

$$\begin{aligned} q_t + \boldsymbol{v} \cdot \boldsymbol{\nabla} q + bq &= 0, \\ q(\boldsymbol{x}, t_0) &= q_0(\boldsymbol{x}). \end{aligned} \tag{2.1}$$

Here $b = b(\boldsymbol{x}, t)$ is a scalar, and we note that setting $b = \boldsymbol{\nabla} \cdot \boldsymbol{v}$ yields the continuity equation (1.4). From (2.1) we have

$$\frac{d}{dt} \int_\Omega q^2 d\boldsymbol{x} + \int_\Omega (2b - \boldsymbol{\nabla} \cdot \boldsymbol{v}) q^2 d\boldsymbol{x} = 0, \tag{2.2}$$

which is derived via integration by parts, where we assume that $2b - \boldsymbol{\nabla} \cdot \boldsymbol{v} \in L^\infty(\Omega)$ and $q_0 \in L^2(\Omega)$ so that (2.1) has a unique solution $q \in L^2(\Omega)$ for all time (e.g., using energy arguments applied to (2.2) similar to those presented in [37, sec. 5.3]). We write this solution as

$$q(\boldsymbol{x}, t) = (\boldsymbol{\mathcal{M}}_t q_0)(\boldsymbol{x}), \tag{2.3}$$

where $\boldsymbol{\mathcal{M}}_t \colon L^2(\Omega) \mapsto L^2(\Omega)$ is the solution operator of (2.1),

$$(\boldsymbol{\mathcal{M}}_t f)(\boldsymbol{x}) = \int_\Omega M(\boldsymbol{x}, t; \boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{2.4}$$

The subscript t on all operators denoted using the calligraphy font style, as in $\boldsymbol{\mathcal{M}}_t$ for example, indicates the operator evaluated at time t, not the time derivative. The kernel of the operator $\boldsymbol{\mathcal{M}}_t$, $M = M(\boldsymbol{x}, t; \boldsymbol{\xi})$, is the fundamental solution of (2.1),

$$\begin{aligned} M_t + \boldsymbol{v} \cdot \boldsymbol{\nabla}\, M + bM &= 0, \\ M(\boldsymbol{x}, t_0; \boldsymbol{\xi}) &= \delta(\boldsymbol{x}, \boldsymbol{\xi}), \end{aligned} \tag{2.5}$$

where the initial condition is the Dirac delta. Here, we simply view the Dirac delta as the kernel of the identity operator $\boldsymbol{\mathcal{I}}\colon L^2(\Omega) \mapsto L^2(\Omega)$,

$$(\boldsymbol{\mathcal{I}} f)(\boldsymbol{x}) = f(\boldsymbol{x}) = \int_\Omega \delta(\boldsymbol{x}, \boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{2.6}$$

Equation (2.3) is analogous to the discrete state propagation computed in data assimilation schemes. We can propagate our discrete state $\boldsymbol{q}$ in (1.1) from time t_0 to t_k,

$$\boldsymbol{q}_k = \underbrace{\boldsymbol{M}_{k,k-1}\boldsymbol{M}_{k-1,k-2}\ldots\boldsymbol{M}_{2,1}\boldsymbol{M}_{1,0}}_{\boldsymbol{M}_{k,0}}\boldsymbol{q}_0, \tag{2.7}$$

which is the discrete version of (2.3) evaluated at time $t = t_k$; the operator $\boldsymbol{\mathcal{M}}_{t_k}$ is the continuum version of the propagation matrix $\boldsymbol{M}_{k,0}$.

With the model state now defined, we can derive the corresponding covariance evolution equation for $P = P(\boldsymbol{x}_1, \boldsymbol{x}_2, t)$ with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega$ and $t \geq t_0$ [6, sec. 2 and references therein]:

$$\begin{aligned} P_t + \boldsymbol{v}_1 \cdot \boldsymbol{\nabla}_1\, P + \boldsymbol{v}_2 \cdot \boldsymbol{\nabla}_2\, P + (b_1 + b_2)P &= 0, \\ P(\boldsymbol{x}_1, \boldsymbol{x}_2, t_0) &= P_0(\boldsymbol{x}_1, \boldsymbol{x}_2), \end{aligned} \tag{2.8}$$

where, again, $\boldsymbol{\nabla}_i$ refers to the gradient with respect to $\boldsymbol{x}_i$, and $\boldsymbol{v}_i = \boldsymbol{v}(\boldsymbol{x}_i, t)$, $b_i = b(\boldsymbol{x}_i, t)$ for $i = 1, 2$.

The solution of the covariance evolution equation (2.8) can be expressed using the fundamental solution operator $\boldsymbol{\mathcal{M}}_t$ and its adjoint $\boldsymbol{\mathcal{M}}_t^*$. The adjoint fundamental solution operator is defined using the inner product over the Hilbert space $L^2(\Omega)$,

$$(\boldsymbol{\mathcal{M}}_t^* f, g)_2 = (f, \boldsymbol{\mathcal{M}}_t g)_2 \quad \forall f, g \in L^2(\Omega). \tag{2.9}$$

The adjoint operator, $\boldsymbol{\mathcal{M}}_t^*\colon L^2(\Omega) \mapsto L^2(\Omega)$, can be expressed as an integral operator whose kernel M^* is the solution to the adjoint final value problem associated with (2.1),

$$(\boldsymbol{\mathcal{M}}_t^* f)(\boldsymbol{\xi}) = \int_\Omega M^*(\boldsymbol{\xi}; \boldsymbol{x}, t) f(\boldsymbol{x}) d\boldsymbol{x}. \tag{2.10}$$

With respect to the method of characteristics, the fundamental solution operator $\boldsymbol{\mathcal{M}}_t$ propagates the solution forward along the characteristics determined by departure points, and the adjoint fundamental solution operator $\boldsymbol{\mathcal{M}}_t^*$ propagates the solution backwards along the characteristics determined by arrival points. This yields the symmetry property [8, p. 729] that at any fixed time t the kernels satisfy

$$M(\boldsymbol{x}, t; \boldsymbol{\xi}) = M^*(\boldsymbol{\xi}; \boldsymbol{x}, t). \tag{2.11}$$

Using (2.11), we can express the covariance in terms of the kernels of the fundamental solution and adjoint fundamental solution operators,

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = \int_\Omega \int_\Omega M(\boldsymbol{x}_1, t; \boldsymbol{\xi}_1) P_0(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) M^*(\boldsymbol{\xi}_2; \boldsymbol{x}_2, t) d\boldsymbol{\xi}_2 d\boldsymbol{\xi}_1, \tag{2.12}$$

or simply

$$\boldsymbol{\mathcal{P}}_t = \boldsymbol{\mathcal{M}}_t \boldsymbol{\mathcal{P}}_0 \boldsymbol{\mathcal{M}}_t^*, \tag{2.13}$$

where $\boldsymbol{\mathcal{P}}_0 \colon L^2(\Omega) \mapsto L^2(\Omega)$ is the operator whose kernel is P_0,

$$(\boldsymbol{\mathcal{P}}_0 f)(\boldsymbol{x}_1) = \int_\Omega P_0(\boldsymbol{x}_1, \boldsymbol{x}_2) f(\boldsymbol{x}_2) d\boldsymbol{x}_2, \tag{2.14}$$

and $\boldsymbol{\mathcal{P}}_t \colon L^2(\Omega) \mapsto L^2(\Omega)$ is the resulting operator at time t,

$$(\boldsymbol{\mathcal{P}}_t f)(\boldsymbol{x}_1) = \int_\Omega P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) f(\boldsymbol{x}_2) d\boldsymbol{x}_2. \tag{2.15}$$

Thus, the covariance evolution equation (2.8) is interpreted as the evolution equation for the kernel of the covariance operator $\boldsymbol{\mathcal{P}}_t$. As with the state propagation, (2.13) evaluated at time t_k is the continuum version of the discrete covariance propagation,

$$\boldsymbol{P}_k = \boldsymbol{M}_{k,0} \boldsymbol{P}_0 \boldsymbol{M}_{k,0}^T, \tag{2.16}$$

following from (1.3) and (2.7).

2.2. The polar decomposition. We next define the (left) polar decomposition of the fundamental solution operator $\boldsymbol{\mathcal{M}}_t$, which will bring to light important properties of the covariance evolution that will be discussed in subsection 2.3. The polar decomposition is a canonical form for all bounded linear operators on Hilbert spaces [36, pp. 196–198]. It is the unique decomposition $\boldsymbol{\mathcal{M}}_t = \boldsymbol{\mathcal{D}}_t \boldsymbol{\mathcal{U}}_t$, where $\boldsymbol{\mathcal{D}}_t = (\boldsymbol{\mathcal{M}}_t \boldsymbol{\mathcal{M}}_t^*)^{1/2}$ and $\boldsymbol{\mathcal{U}}_t$ is a partial isometry.

To derive the polar decomposition for $\boldsymbol{\mathcal{M}}_t$, we first decompose the fundamental solution M into

$$M(\boldsymbol{x}, t; \boldsymbol{\xi}) = d(\boldsymbol{x}, t) u(\boldsymbol{x}, t; \boldsymbol{\xi}), \tag{2.17}$$

where $d = d(\boldsymbol{x}, t)$ and $u = u(\boldsymbol{x}, t; \boldsymbol{\xi})$ satisfy the following PDEs:

$$\begin{gathered} d_t + \boldsymbol{v} \cdot \boldsymbol{\nabla}\, d + \left(b - \frac{1}{2} \boldsymbol{\nabla} \cdot \boldsymbol{v} \right) d = 0, \\ d(\boldsymbol{x}, t_0) = 1, \end{gathered} \tag{2.18}$$

$$\begin{gathered} u_t + \boldsymbol{v} \cdot \boldsymbol{\nabla}\, u + \frac{1}{2} (\boldsymbol{\nabla} \cdot \boldsymbol{v}) u = 0, \\ u(\boldsymbol{x}, t_0; \boldsymbol{\xi}) = \delta(\boldsymbol{x}, \boldsymbol{\xi}). \end{gathered} \tag{2.19}$$

The solution u of (2.19) is quadratically conservative,

$$\frac{d}{dt} \int_\Omega u^2(\boldsymbol{x}, t; \boldsymbol{\xi}) d\boldsymbol{x} = 0, \tag{2.20}$$

and therefore defines a bounded linear operator $\boldsymbol{\mathcal{U}}_t\colon L^2(\Omega) \mapsto L^2(\Omega)$ whose kernel is the solution to (2.19),

$$(\boldsymbol{\mathcal{U}}_t f)(\boldsymbol{x}) = \int_\Omega u(\boldsymbol{x}, t; \boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{2.21}$$

The operator $\boldsymbol{\mathcal{U}}_t$ is an invertible isometry and therefore unitary.

To obtain the operator $\boldsymbol{\mathcal{D}}_t$, we consider the operator $\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*$. From (2.13), the operator $\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*$ is just the covariance operator $\boldsymbol{\mathcal{P}}_t$ when the initial covariance operator $\boldsymbol{\mathcal{P}}_0$ is the identity operator (2.6); equivalently, the kernel of the operator $\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*$ is the solution to the covariance equation (2.8) when the initial covariance $P_0(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is the Dirac delta, $\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$. We show in Appendix A that the solution of (2.8) with the initial condition $P_0(\boldsymbol{x}_1, \boldsymbol{x}_2) = \delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = d^2(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where d is the solution to (2.18). In other words, according to (2.13), $\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*$ is in fact a multiplication operator. A multiplication operator $\boldsymbol{\mathcal{K}}\colon L^2(\Omega) \mapsto L^2(\Omega)$ is defined as one for which

$$(\boldsymbol{\mathcal{K}} f)(\boldsymbol{x}) = k(\boldsymbol{x}) f(\boldsymbol{x}) = \int_\Omega k(\boldsymbol{\xi})\delta(\boldsymbol{x}, \boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{2.22}$$

where $k(\boldsymbol{x}) \in L^\infty(\Omega)$ is the multiplication function. The operator $\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*\colon L^2(\Omega) \mapsto L^2(\Omega)$,

$$(\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^* f)(\boldsymbol{x}) = d^2(\boldsymbol{x}, t) f(\boldsymbol{x}), \tag{2.23}$$

is a multiplication operator, where the multiplication function satisfies the following differential equation:

$$\begin{gathered} d_t^2 + \boldsymbol{v} \cdot \boldsymbol{\nabla}\, d^2 + (2b - \boldsymbol{\nabla} \cdot \boldsymbol{v})\, d^2 = 0, \\ d^2(\boldsymbol{x}, t_0) = 1. \end{gathered} \tag{2.24}$$

As the operator $\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*$ is nonnegative, its square root exists, and we define the operator $\boldsymbol{\mathcal{D}}_t\colon L^2(\Omega) \mapsto L^2(\Omega)$ as this square root,

$$(\boldsymbol{\mathcal{D}}_t f)(\boldsymbol{x}) = ((\boldsymbol{\mathcal{M}}_t\boldsymbol{\mathcal{M}}_t^*)^{1/2} f)(\boldsymbol{x}) = d(\boldsymbol{x}, t) f(\boldsymbol{x}), \tag{2.25}$$

with the multiplication function for $\boldsymbol{\mathcal{D}}_t$ being the solution to (2.18). Note that because the operator $\boldsymbol{\mathcal{D}}_t$ is a multiplication operator with a real-valued multiplication function, it is self-adjoint.

The decomposition (2.17) of the kernel of the fundamental solution operator gives us the (left) polar decomposition of $\boldsymbol{\mathcal{M}}_t$,

$$\boldsymbol{\mathcal{M}}_t = \boldsymbol{\mathcal{D}}_t\boldsymbol{\mathcal{U}}_t, \tag{2.26}$$

with $\boldsymbol{\mathcal{D}}_t$ and $\boldsymbol{\mathcal{U}}_t$ as defined above. In the next section, we use this polar decomposition to study the continuum covariance propagation.

Remark 2.1. The polar decomposition is a canonical form for all bounded linear operators on Hilbert spaces [36, pp. 196–198] and is crucial for deriving the continuous spectrum equation (2.33) in subsection 2.3. Related to the polar decomposition is the continuum singular value

decomposition (SVD), which is a canonical form for compact linear operators on Hilbert spaces [36, Thm. VI.17]. The fundamental solution operator $\boldsymbol{\mathcal{M}}_t$ defined for the advective dynamics presented in this work, (2.1), is not compact; therefore the continuum SVD does not apply. If a diffusion term were added to (2.1), say $\kappa\boldsymbol{\nabla}^2 q$ for $\kappa > 0$ a diffusion coefficient and $\boldsymbol{\nabla}^2$ the Laplacian operator, the corresponding fundamental solution operator would become compact and a continuum SVD would exist. The discrete SVD is used widely in matrix analysis, particularly in NWP data assimilation literature [26, 5, 12, 2].

2.3. Continuum covariance propagation. The solution of the continuum covariance evolution equation (2.8) depends on the initial covariance $P_0(\boldsymbol{x}_1, \boldsymbol{x}_2)$, and we show in this section how the behavior of the solution changes as the initial correlation length tends to zero. To do so, we will interpret the covariance function $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t)$ as the kernel of the operator $\boldsymbol{\mathcal{P}}_t$ defined in (2.15), whose evolution can be written in terms of the fundamental solution operator, its adjoint, and the initial covariance, as given in (2.13).

We consider two cases for the initial covariance $P_0(\boldsymbol{x}_1, \boldsymbol{x}_2)$. First, assume the initial covariance is continuous on $\Omega \times \Omega$, and denote it as $P_0^d(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Since P_0^d is continuous, the solution to the covariance equation (2.8) is a strong solution and has a bounded L^2-norm. Therefore the corresponding covariance operator, which we will denote as $\boldsymbol{\mathcal{P}}_t^d$, is a self-adjoint Hilbert–Schmidt operator [36, pp. 210–211]. Hilbert–Schmidt operators are a subclass of the compact operators, and it follows from spectral theory that self-adjoint Hilbert–Schmidt operators only contain eigenvalues in their spectrum [20, pp. 230–232] with the possible exception of zero in the continuous spectrum. The set of eigenvalues is often referred to as the discrete spectrum; hence we can refer to the covariance operator $\boldsymbol{\mathcal{P}}_t^d$ as the discrete spectrum covariance operator.

Now, consider the case where the initial state q_0 is spatially uncorrelated, whose covariance we represent by $P_0^c(\boldsymbol{x}_1)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$, and assume that the function P_0^c is continuous on Ω. The Dirac delta in the initial covariance reduces the initial covariance operator (2.14) to a multiplication operator (2.22) with multiplication function P_0^c; denote this operator as $\boldsymbol{\mathcal{P}}_0^c$. We can see how this impacts the corresponding covariance operator, which we denote as $\boldsymbol{\mathcal{P}}_t^c$, by applying the polar decomposition (2.26) to (2.13) with $\boldsymbol{\mathcal{P}}_0 = \boldsymbol{\mathcal{P}}_0^c$,

$$\boldsymbol{\mathcal{P}}_t^c = \boldsymbol{\mathcal{D}}_t \boldsymbol{\mathcal{U}}_t \boldsymbol{\mathcal{P}}_0^c \boldsymbol{\mathcal{U}}_t^* \boldsymbol{\mathcal{D}}_t = \boldsymbol{\mathcal{D}}_t \tilde{\boldsymbol{\mathcal{P}}}_t \boldsymbol{\mathcal{D}}_t, \tag{2.27}$$

where

$$\tilde{\boldsymbol{\mathcal{P}}}_t = \boldsymbol{\mathcal{U}}_t \boldsymbol{\mathcal{P}}_0^c \boldsymbol{\mathcal{U}}_t^*. \tag{2.28}$$

From the definition of $\boldsymbol{\mathcal{U}}_t$ in (2.21), it follows that the operator $\tilde{\boldsymbol{\mathcal{P}}}_t \colon L^2(\Omega) \mapsto L^2(\Omega)$ has kernel $\tilde{P} = \tilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2, t)$ given by

$$\begin{aligned} \tilde{P}_t + \boldsymbol{v}_1 \cdot \boldsymbol{\nabla}_1 \tilde{P} + \boldsymbol{v}_2 \cdot \boldsymbol{\nabla}_2 \tilde{P} + \frac{1}{2}(\boldsymbol{\nabla}_1 \cdot \boldsymbol{v}_1 + \boldsymbol{\nabla}_2 \cdot \boldsymbol{v}_2)\tilde{P} = 0, \\ \tilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2, t_0) = P_0^c(\boldsymbol{x}_1)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2), \end{aligned} \tag{2.29}$$

whose solution is shown in Appendix A to be

$$\tilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = \tilde{P}^c(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2), \tag{2.30}$$

where

$$\begin{aligned} \tilde{P}^c_t + \boldsymbol{v} \cdot \boldsymbol{\nabla} \tilde{P}^c &= 0, \\ \tilde{P}^c(\boldsymbol{x}, t_0) &= P^c_0(\boldsymbol{x}). \end{aligned} \tag{2.31}$$

Thus, $\tilde{\boldsymbol{\mathcal{P}}}_t$ is a multiplication operator, and it follows from (2.18) and (2.31) that $\boldsymbol{\mathcal{P}}^c_t$ in (2.27) is also a multiplication operator,

$$(\boldsymbol{\mathcal{P}}^c_t f)(\boldsymbol{x}_1) = P^c(\boldsymbol{x}_1, t) f(\boldsymbol{x}_1) = \int_\Omega P^c(\boldsymbol{x}_1, t) \delta(\boldsymbol{x}_1, \boldsymbol{x}_2) f(\boldsymbol{x}_2) d\boldsymbol{x}_2, \tag{2.32}$$

where $P^c = d^2 \tilde{P}$ is the solution to the continuous spectrum equation

$$\begin{aligned} P^c_t + \boldsymbol{v} \cdot \boldsymbol{\nabla} P^c + (2b - \boldsymbol{\nabla} \cdot \boldsymbol{v}) P^c &= 0, \\ P^c(\boldsymbol{x}, t_0) &= P^c_0(\boldsymbol{x}). \end{aligned} \tag{2.33}$$

Since multiplication operators contain only a continuous spectrum [20, pp. 219, 240], we will refer to $\boldsymbol{\mathcal{P}}^c_t$ as the continuous spectrum covariance operator and to P^c as the continuous spectrum solution.

Thus, we have shown that the solution of the covariance evolution equation (2.8) for initial condition $P^c_0(\boldsymbol{x}_1)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = P^c(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$; white noise evolved under the state dynamics (2.1) remains white. Further, we can see explicitly that the dynamics of the covariance along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane are governed by the continuous spectrum equation (2.33) for spatially uncorrelated initial states, rather than by the variance equation

$$\begin{aligned} \sigma^2_t + \boldsymbol{v} \cdot \boldsymbol{\nabla} \sigma^2 + 2b\sigma^2 &= 0, \\ \sigma^2(\boldsymbol{x}, t_0) = \sigma^2_0(\boldsymbol{x}) &= P^d_0(\boldsymbol{x}, \boldsymbol{x}), \end{aligned} \tag{2.34}$$

which follows directly from (2.8) when the initial covariance is continuous on $\Omega \times \Omega$. Only in cases where the velocity field is divergence-free are the continuous spectrum equation and variance equation identical. The dynamics of the covariance along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane are governed by the variance equation (2.34) for all continuous initial covariances, independently of nonzero initial correlation lengths, but at zero correlation length the dynamics change abruptly to those of the continuous spectrum equation (2.33).

Covariances associated with spatially uncorrelated initial states, which correspond to covariances with zero initial correlation length scales, are limiting cases in our analysis as well as in practice. Atmospheric wind fields, for example, have sharp vertical correlation structures relative to long horizontal correlations that need to be represented in covariances for NWP models [35], and horizontal wind shear leads to tracer correlations that shrink in the direction perpendicular to the flow such that they are no longer spatially resolved [27]. Through careful analysis and use of the polar decomposition, we are able to derive the discontinuous change in dynamics along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane as the correlation length approaches zero for covariances governed by (2.8), which is not readily apparent when first considering the model state and covariance equations or when only considering these equations in discrete space.

3. Numerical experiments.

The purpose of the numerical experiments is to examine the evolution of the diagonal of the covariance matrix $\boldsymbol{P}_k$ during discrete covariance propagation and its relation to the discontinuous change in the continuum dynamics along the $\boldsymbol{x}_1, \boldsymbol{x}_2$-hyperplane as the initial correlation length approaches zero. To clearly illustrate the problems associated with full-rank covariance propagation, a simple example is used so that our numerical results can be compared to a known exact solution. We conduct these experiments for various types of initial covariances specified with different initial standard deviations and different correlation kernel functions with varying correlation length scales.

3.1. Experimental setup.

For these experiments, we will consider the one-dimensional version of the continuity equation (1.4) over the unit circle S_1^1. We take the velocity field to be independent of time and spatially varying:

$$v(x) = \sin(x) + 2. \tag{3.1}$$

The exact solution to the one-dimensional continuity equation with this velocity field can be obtained explicitly using the method of characteristics and is used as reference for our experiments; see Appendix B for further discussion.

The spatial domain S_1^1 is discretized on a uniform grid, $x_i = i\Delta x,\ i = 0, 1, \ldots, N-1$, where $N = 200$ and $\Delta x = \frac{2\pi}{N}$. The time discretization is given by $t_k = k\Delta t$, where the time step Δt is determined from the Courant–Friedrichs–Lewy number λ,

$$\lambda = \max_{x\in[0,2\pi]} |v(x)| \frac{\Delta t}{\Delta x} = 3\frac{\Delta t}{\Delta x} \leq 1. \tag{3.2}$$

For these experiments, we take $\lambda = 1$. We ran experiments with several other values of $\lambda < 1$ (not shown) and found it did not have a significant impact on the results.

3.1.1. Numerical covariance propagation.

We use two methods of propagation to illustrate the impact of numerical schemes on the discrete covariance propagation (1.3) and to leverage insights from continuum covariance analysis in the form of the polar decomposition. The covariance matrix is propagated discretely using either of two methods:

1. Traditional propagation: the covariance is propagated as in (1.3), where the matrix $\boldsymbol{M}_{k,k-1}$ is the finite difference discretization of the fundamental solution equation (2.5).
2. Polar decomposition propagation: the polar decomposition of the fundamental solution operator (2.26) is discretized and used in place of the matrix $\boldsymbol{M}_{k,k-1}$ in (1.3).

In the polar decomposition propagation, (2.26) is discretized as follows. The operator $\boldsymbol{\mathcal{D}}_t$ of (2.25) is a self-adjoint multiplication operator; therefore when $t = t_k$ its corresponding discretization is the diagonal matrix $\boldsymbol{D}_{k,0}$ whose diagonal elements are $(\boldsymbol{D}_{k,0})_{ii} = d(x_i, t_k)$. Here $d(x_i, t_k)$ is the solution to (2.18) evaluated on the discrete spatial grid at time t_k, which we generate using the exact solution to (2.18), as discussed in Appendix B. The discretization of the operator $\boldsymbol{\mathcal{U}}_t$ of (2.21) is done via finite differences to generate the propagation matrix $\boldsymbol{U}_{k,k-1}$ corresponding to (2.19). The matrix $\boldsymbol{U}_{k,k-1}$, which propagates the solution from time t_{k-1} to t_k, is independent of time by virtue of (3.1); therefore we will denote $\boldsymbol{U}_{k,k-1} = \boldsymbol{U}$ for simplicity. To compute the covariance matrix at time t_k, $\boldsymbol{P}_k$, we do not construct the

matrix $\boldsymbol{M}_{k,k-1}$ explicitly using the polar decomposition, but instead compute the covariance as follows:

$$\begin{aligned}\boldsymbol{P}_k &= \boldsymbol{D}_{k,0}\boldsymbol{U}_{k,k-1}\boldsymbol{U}_{k-1,k-2}\dots\boldsymbol{U}_{1,0}\boldsymbol{P}_0\boldsymbol{U}_{1,0}^T\dots\boldsymbol{U}_{k,k-1}^T\boldsymbol{D}_{k,0}\\ &= \boldsymbol{D}_{k,0}\boldsymbol{U}^k\boldsymbol{P}_0(\boldsymbol{U}^T)^k\boldsymbol{D}_{k,0}.\end{aligned} \tag{3.3}$$

The Lax–Wendroff [24] and Crank–Nicolson [10] finite difference schemes are used to generate the matrices $\boldsymbol{U}_{k,k-1}$ and $\boldsymbol{M}_{k,k-1}$ corresponding to their respective PDEs (2.19) and (2.5). Like $\boldsymbol{U}_{k,k-1}$, $\boldsymbol{M}_{k,k-1}$ is independent of time; therefore throughout the rest of the paper we will denote $\boldsymbol{M}_{k,k-1}$ simply as $\boldsymbol{M}$. We choose these two simple finite difference schemes to illustrate their contrasting behaviors, particularly when generating the matrix $\boldsymbol{U}$. According to the continuum analysis, the operator $\boldsymbol{\mathcal{U}}_t$ is unitary. The Crank–Nicolson discretization preserves the unitary property of $\boldsymbol{\mathcal{U}}_t$ because the scheme is quadratically conservative, but the Lax–Wendroff scheme does not. In the numerical results, subsections 3.2 and 3.3, propagated covariances are labeled by the finite difference scheme used to construct the matrices $\boldsymbol{M}$ and $\boldsymbol{U}$ (Lax–Wendroff or Crank–Nicolson) followed by the method of propagation (traditional or polar decomposition) as described in the beginning of this section.

3.1.2. Initial covariances. We generate four types of initial covariances from the following two different correlation kernel functions: the Gaspari–Cohn (GC) fifth-order piecewise rational function [15, eq. (4.10)], and the first order autoregressive function (FOAR) [16, eq. (23)]. Using each correlation function, we construct the initial covariance matrix with either a constant initial variance or spatially varying initial variance. Initial covariances with constant initial variance take the initial variance to be one, while initial covariances with the spatially varying initial variance take the initial variance as the square of the standard deviation

$$\sigma_0(x) = \frac{\sin(3x)}{3} + 1. \tag{3.4}$$

The GC correlation function $C_0(r(x_i, x_j), 1/2, c)$ is a compactly supported approximation to a Gaussian function, supported on the interval $0 \le r(x_i, x_j) \le 2c$, where

$$r(x_i, x_j) = 2\sin(|x_i - x_j|/2) \tag{3.5}$$

is the chordal distance between x_i and x_j on S_1^1. On the uniform spatial grid of 200 grid points for these experiments, values of $c = 1$, 0.25, and 0.05 correspond to 100, 16, and 3 grid lengths (Δx), respectively, from the peak of the correlation function to where it becomes zero, and 33, 8, and just 1 grid length, respectively, from the peak value of 1 to values less than 0.2; see Figure 1 for these examples.

The FOAR correlation function given by

$$F_L(r(x_i, x_j)) = exp(-r(x_i, x_j)/L) \tag{3.6}$$

is continuous but nondifferentiable at the origin because of its cusp-like behavior (see Figure 1). As with the GC correlation function, the chordal distance (3.5) is used to reflect periodicity of the domain. The FOAR correlation functions are nonzero on the full spatial domain. On the spatial grid, $L = 0.5$, 0.25, and 0.03 corresponds to 26, 12, and just 1 grid length, respectively, from the peak of the correlation to where it becomes less than 0.2.

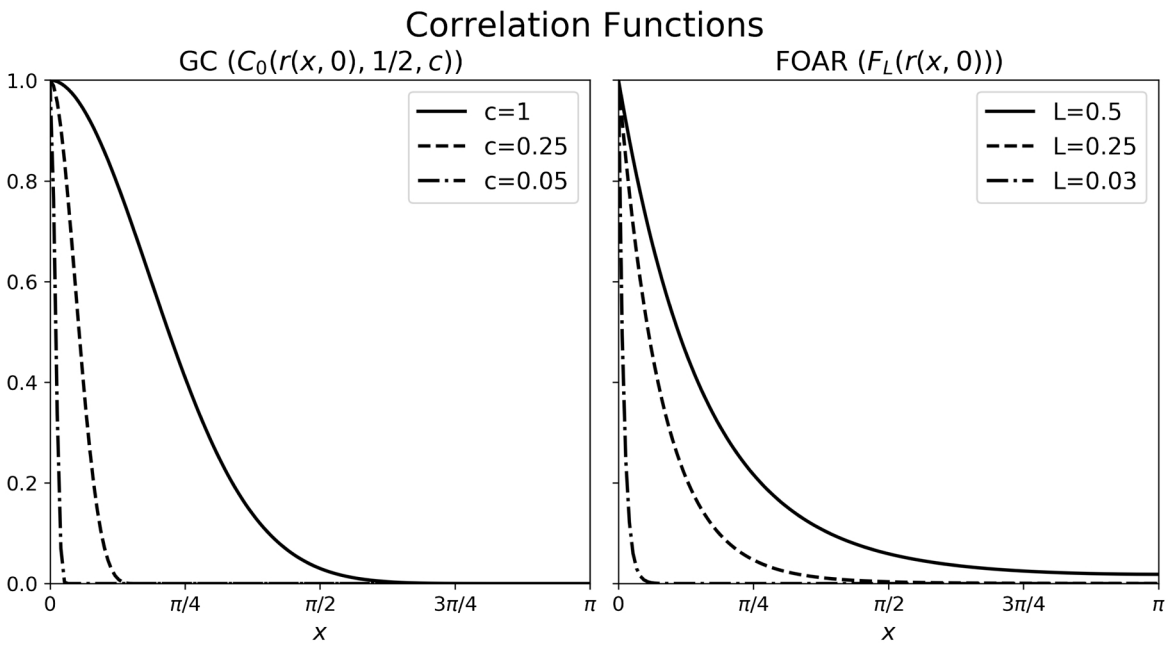

Figure 1. *Examples of the correlation functions used to generate the initial covariances for numerical propagation. The GC correlation functions (left) are functions of compact support length parameter c, where the FOAR correlation functions (right) are functions of length scale L.*

3.2. Experimental results. From the continuum analysis, one might expect the diagonal of the covariance matrix $\boldsymbol{P}_k$ to behave according to the variance equation for covariances with nonzero initial correlation lengths and according to the continuous spectrum equation for covariances with zero initial correlation length. To establish a baseline, Figure 2 illustrates the solutions to the variance equation (1D version of (1.8)) and continuous spectrum equation (1D version of (1.9)) for unit initial condition and for the spatially varying initial condition given by the square of (3.4). Considering the exact solutions first (black), we see that the dynamics of the variance solution and continuous spectrum solution are quite different due to the spatially varying velocity field (3.1). We also see that solving either the variance equation or the continuous spectrum equation directly using the Crank–Nicolson scheme (colored) produces solutions that are nearly indistinguishable from the exact solutions. Solutions to the variance and continuous spectrum equations computed using Lax-Wendroff differ very slightly from Crank-Nicolson and are not shown.

Figure 2 shows that propagating the diagonal of the covariance matrix $\boldsymbol{P}_k$ numerically, independently of the rest of the matrix, using the known dynamics of either the variance equation or the continuous spectrum equation, produces minimal discretization error. Figure 3 illustrates further that, at least for covariance matrices with relatively long initial correlation lengths, the numerically propagated full covariance itself also contains only minor discretization errors typically expected from finite difference approximations. Figure 3 shows results for the GC initial correlation supported on the full spatial domain ($c = 1$). The dissipative behavior of both Lax–Wendroff schemes is clear in the normalized spectra (left panels), whereas

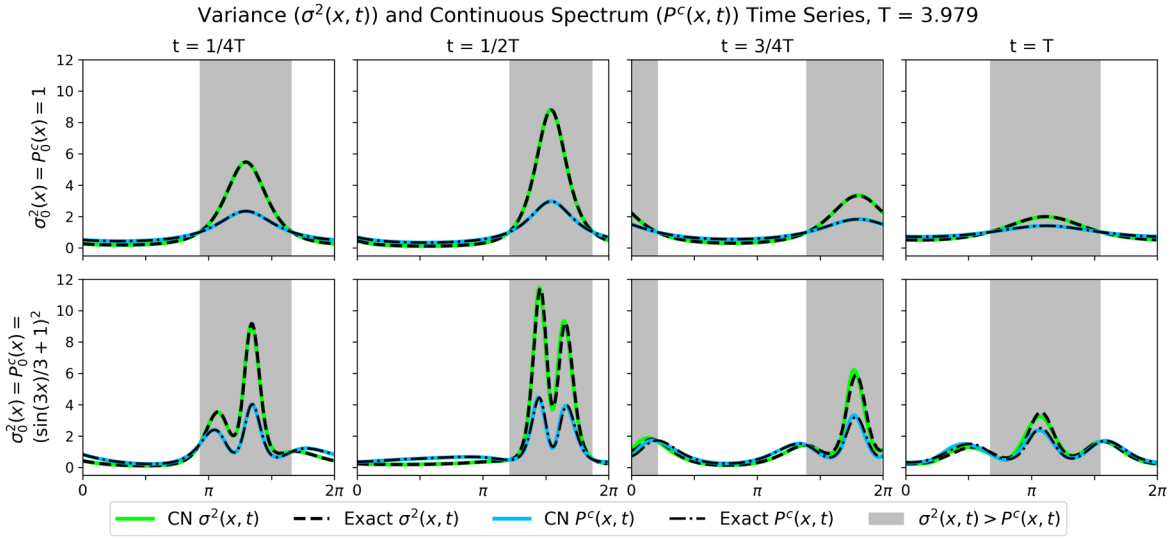

Figure 2. *Top: Solutions to the one-dimensional variance equation (1.8) and continuous spectrum equation (1.9) at various times with velocity field (3.1), for unit initial condition (top row) and spatially varying initial condition (bottom row) taken to be the square of (3.4). Time $T = 3.979$ corresponds to slightly after a full period. Exact solutions for each case are given in black dashed (variance) and black dot-dashed lines (continuous spectrum). Green curves denote the solution to the variance equation computed with Crank–Nicolson (abbreviated CN); blue curves are solutions to the continuous spectrum equation computed with Crank–Nicolson. Regions highlighted in grey correspond to regions where the exact solution to the variance equation is greater than the exact solution to the continuous spectrum equation, i.e., $\sigma^2 > P^c$; unhighlighted regions correspond to regions where $\sigma^2 < P^c$.*

both the traditional Crank–Nicolson propagation and polar decomposition propagation using the Crank–Nicolson $\boldsymbol{U}$ (hereafter referred to as Crank–Nicolson polar decomposition) capture the normalized spectra quite well. The small amounts of variance loss and gain seen in both the constant initial variance and spatially varying initial variance cases (middle panels of Figure 3) are consistent with dissipation and phase errors expected from finite differences. The polar decomposition methods (dashed) reduce the errors in the diagonals of the covariance matrices only slightly compared to the traditional methods (solid). The correlations at row 150 (right panels), which correspond to where the variance reaches a maximum as the velocity field is at a minimum, are nearly identical to the exact correlations, suggesting minimal errors in correlation propagation.

The accuracy of numerically propagated full covariances is much worse for short initial correlation lengths, increasingly so as they approach zero. We monitor the total amount of variance lost or gained over time through the trace of the covariance matrix, $Tr(\boldsymbol{P}_k)$. Figure 4 shows the trace time series for both the GC and FOAR cases with spatially varying initial variance as initial correlation lengths tend to zero (the constant initial variance case results are similar to Figure 4 and are not shown). The GC and FOAR cases in Figure 4 exhibit similar behaviors in the trace over time, even though their initial correlation structures are quite different. As the initial correlation lengths tend to zero, the amount of variance lost during propagation increases strikingly in the Lax–Wendroff schemes. The polar decomposition

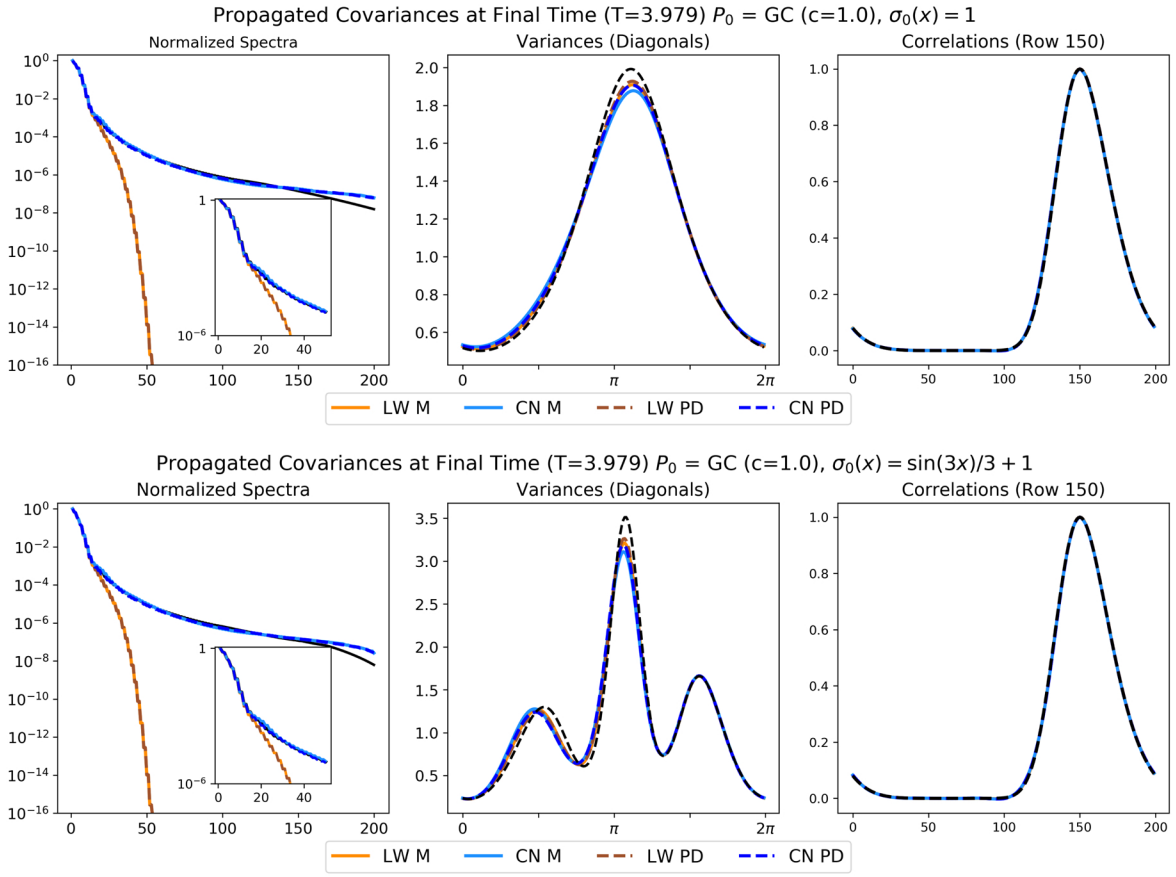

Figure 3. *Propagated covariances at the final time T (slightly after a full period) for GC initial correlations with $c = 1$ (supported on the full domain) for all four propagation methods. Top row: Constant initial variance. Bottom row: Spatially varying initial variance. The left panels correspond to the normalized spectra (relative to largest eigenvalue), the middle panels show the discrete variances (covariance matrix diagonals), and the right panels show the correlations at row 150 (location of maximum in variance in space and time). The exact normalized spectra are given in solid black lines, and the exact variances and correlations are given in black dashed lines. Crank–Nicolson (CN) and Lax–Wendroff (LW) M refer to traditional propagation using Crank–Nicolson or Lax–Wendroff (solid). CN and LW PD refer to propagation using the polar decomposition (dashed) with the matrix $\boldsymbol{U}$ constructed via the Crank–Nicolson or Lax–Wendroff scheme, respectively.*

propagation schemes (dashed) are an improvement over traditional propagation in some cases but are worse in others and generally suffer similar amounts of variance loss. We also observe that as initial correlation lengths tend to zero, the numerical schemes gradually approach their own limiting behavior at $c = L = 0$, rather than a discontinuous change in dynamics as seen in the continuum analysis.

Had we not performed the continuum analysis of section 2, one might assume that the variance loss in Figure 4 is caused simply by dissipation. However, the Crank–Nicolson scheme is not dissipative and yet produces significant variance loss. In fact, we see for the traditional Crank–Nicolson propagation (solid light blue lines) that there are regions of both variance loss and gain. We also see that for short, nonzero initial correlation lengths ($c = 0.05$, $L = 0.03$ in

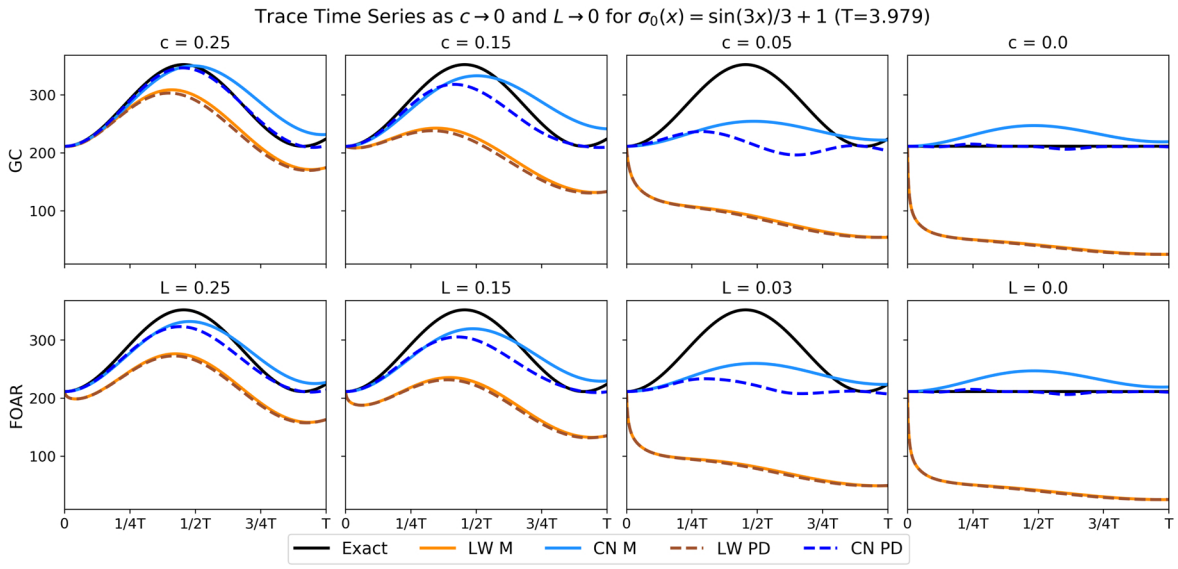

Figure 4. *Trace time series for GC (top row) and FOAR (bottom row) for the spatially varying initial variance (the square of (3.4)). Each panel corresponds to different values of c and L, decreasing from left to right towards $c = L = 0$. Refer to Figure 3 for a description of the curves. For the cases when $c = L = 0$ (rightmost panels), the exact solutions (solid black lines) are constant in time due to the fact that continuous spectrum solution P^c satisfies the continuity equation (1.9); hence its integral over space is constant.*

particular), the numerical schemes better approximate the limiting case of $c = L = 0$ than the correct behavior for $c,\ L > 0$. This suggests that inaccurate discrete diagonal propagation is particularly pronounced for short correlation lengths.

The behavior of the trace time series in Figure 4 indicates that covariance propagation itself can be a source of spurious loss and gain of variance; however, it does not indicate where exactly this manifests itself. To gain a better understanding of the source of variance loss and gain, we examine various aspects of the propagated covariance matrix for different types of initial covariances as we did in Figure 3, but now for covariances with shorter initial correlation lengths.

Figures 5 and 6 are final time snapshots (in the same format as shown in Figure 3) of propagated covariances specified using the GC correlation function with $c = 0.25$ and FOAR correlation function with $L = 0.25$, respectively, which are initially well resolved as described in subsection 3.1.2 and correspond to the mildest variance loss and gain cases shown in leftmost panels of Figure 4. The normalized spectra in both of these cases are similar to the spectra seen in Figure 3, where the Lax–Wendroff schemes are severely dissipative and both Crank–Nicolson methods approximate the exact spectrum moderately well.

The diagonals extracted from the numerically propagated covariances in Figures 5 and 6 are strikingly different from the exact solution. Though covariances with these values of c and L are well resolved initially, the extracted diagonals are smooth but wholly inaccurate. Across all four methods of propagation we see regions of both variance loss and variance gain, clearly illustrating that inaccurate discrete diagonal propagation is the problem more so than dissipation. The diagonal propagation becomes worse when the initial covariance has a

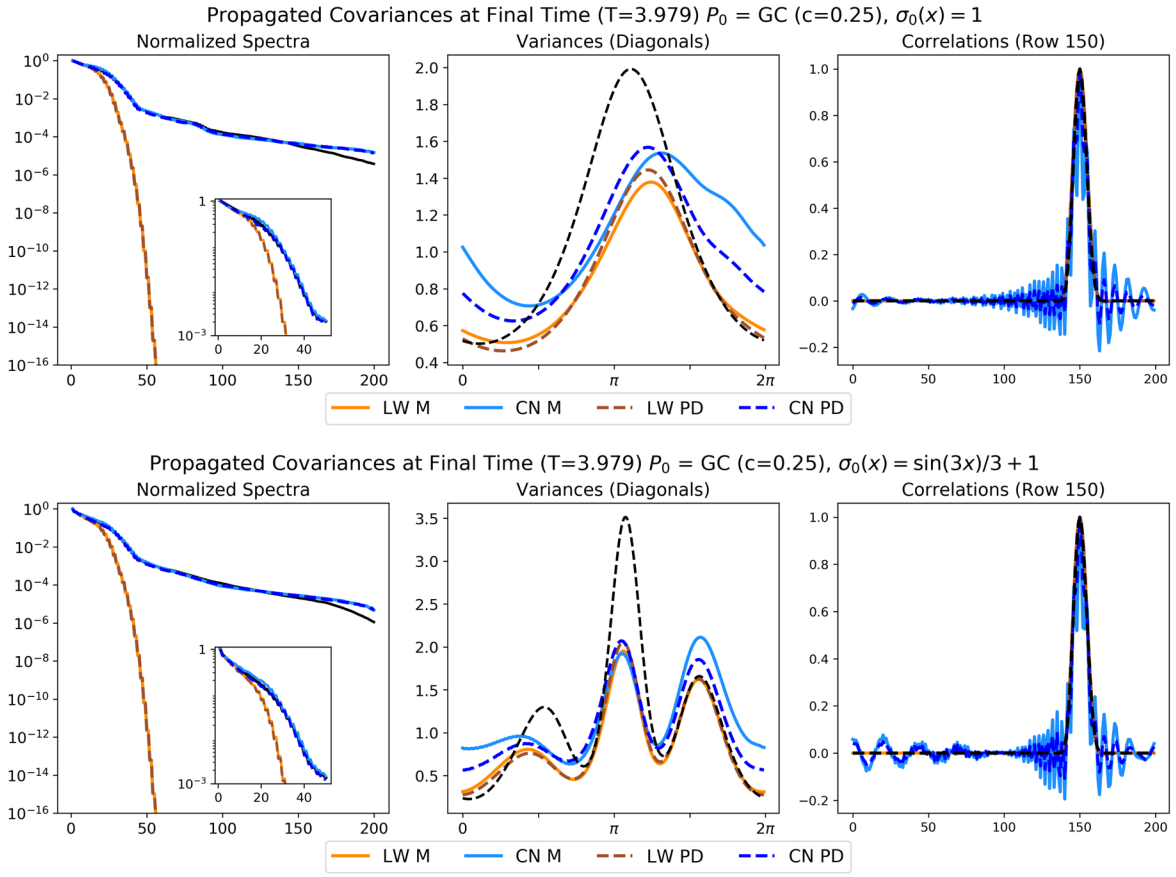

Figure 5. *Same as Figure 3 for* $c = 0.25$.

spatially varying variance (bottom rows of Figures 5 and 6), which is a more realistic situation in practice. The errors we observe in the diagonals of the covariance matrices, interestingly, are not reflected in the normalized spectra; considering the normalized spectra alone would not even hint at the problems occurring in the discrete diagonal propagation. The correlations, as expected, show dispersion off the diagonal. The oscillatory behavior of both Crank–Nicolson schemes due to numerical dispersion is expected for this finite difference scheme [4, p. 46].

In the limiting case when the initial correlation lengths become zero, we see two contrasting behaviors in the numerically propagated covariances depending on the initial variance; see Figure 7. When the initial variance is constant (i.e., the initial covariance is the identity matrix), the Crank–Nicolson polar decomposition is the only scheme that correctly captures the behavior of the exact covariance (top row of Figure 7). This is expected from the definition of the Crank–Nicolson polar decomposition in this case. Since the initial covariance is the identity matrix and the matrix $\boldsymbol{U}$ constructed using the Crank–Nicolson scheme is unitary, the Crank–Nicolson polar decomposition propagation (3.3) reduces to $\boldsymbol{P}_k = \boldsymbol{D}_{k,0}\boldsymbol{U}^k I(\boldsymbol{U}^k)^T\boldsymbol{D}_{k,0} = \boldsymbol{D}_{k,0}\boldsymbol{U}^k(\boldsymbol{U}^k)^T\boldsymbol{D}_{k,0} = \boldsymbol{D}_{k,0}^2$, which is exact since the diagonal of $\boldsymbol{D}_{k,0}$ is evaluated analytically. When the initial variance varies spatially (bottom row

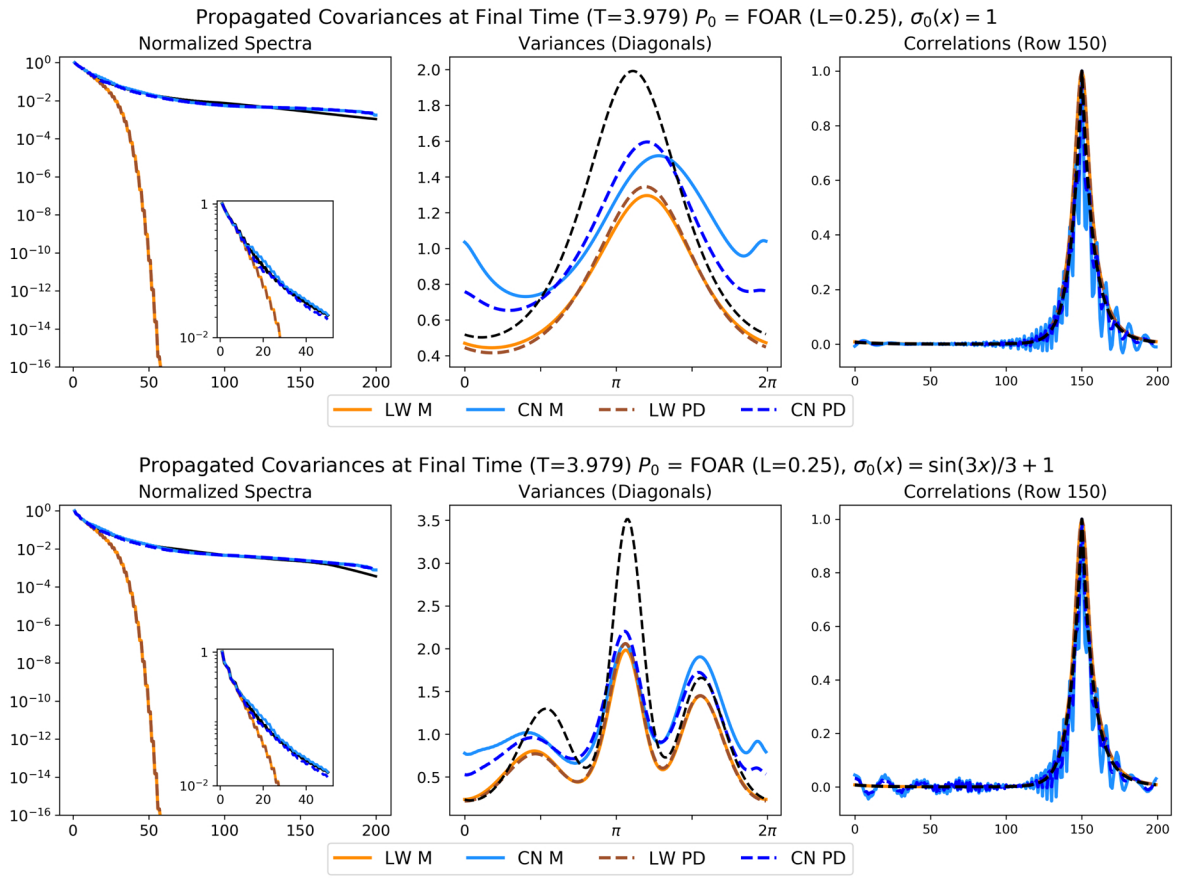

Figure 6. *Same as Figure 5 for the FOAR correlation function with* $L = 0.25$.

of Figure 7), the Crank–Nicolson polar decomposition propagation instead behaves more similarly to the Crank–Nicolson traditional propagation, and both have regions of variance loss and variance gain. Carefully comparing the diagonals extracted from the traditional and polar decomposition propagated covariances, the Crank–Nicolson polar decomposition propagation is a slight improvement over the Crank–Nicolson traditional propagation, but these differences are relatively minor compared to their absolute errors. The Lax–Wendroff schemes are substantially dissipative in all cases. We also observe that as the values of c and L decrease towards zero, the normalized spectra in Figures 3 and 5 to 7 decay more slowly and become relatively flat. This suggests that low-rank approximations would have difficulty capturing these covariances as correlation lengths shrink.

Comparing the diagonals of the covariance matrices at the final time across a series of initial correlation lengths in Figures 8 and 9 demonstrates the severity of the variance loss and gain caused by inaccurate discrete diagonal propagation and provides a closer look at the approach to a limiting behavior seen in the trace time series. Without prior knowledge of the discontinuous change in dynamics as the initial correlation length tends to zero, one might surmise from Figures 8 and 9 that the observed discrete diagonal behavior is caused

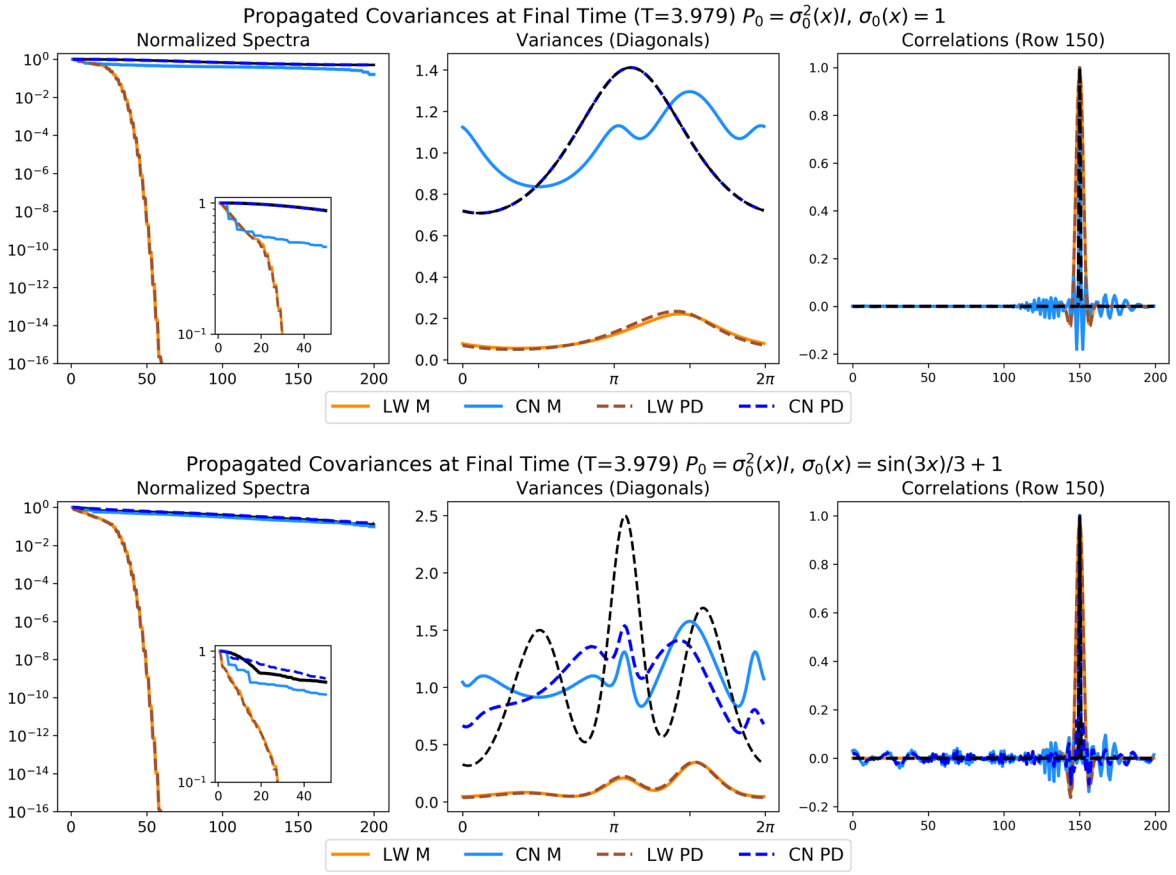

Figure 7. *Propagated covariances at the final time T (slightly after a full period) for initial covariances with zero initial correlation lengths ($c = L = 0$) for all four propagation methods. Top row: Constant initial variance (identity matrix). Bottom row: Spatially varying initial variance. The left panels correspond to the normalized spectra (relative to the largest eigenvalue), the middle panels show the discrete variances (covariance matrix diagonals), and the right panels show the correlations at row* 150 *(location of the maximum in variance in space and time). The exact normalized spectra are given in solid black lines, and the exact diagonals (solutions to the continuous spectrum equation) and correlations are given in black dashed lines. See caption of Figure* 3 *for a description of the colored curves.*

by dissipation. Knowing the continuum behavior, however, makes it clear that we are observing inaccurate discrete diagonal propagation associated with the discontinuous change in continuum dynamics. The Crank–Nicolson polar decomposition propagation for a constant initial variance is the only scheme that captures the correct diagonal behavior when the initial covariance is the identity and gradually approaches this behavior as the initial correlation length decreases. For all other cases in Figures 8 and 9, the resulting covariance matrix diagonals are smooth, but grossly incorrect, and gradually approach their own limiting behavior at $c = L = 0$ rather than changing abruptly as in the continuum case. The Lax–Wendroff schemes are severely dissipative, hardly resembling the correct dynamics, and are not shown in these and subsequent figures.

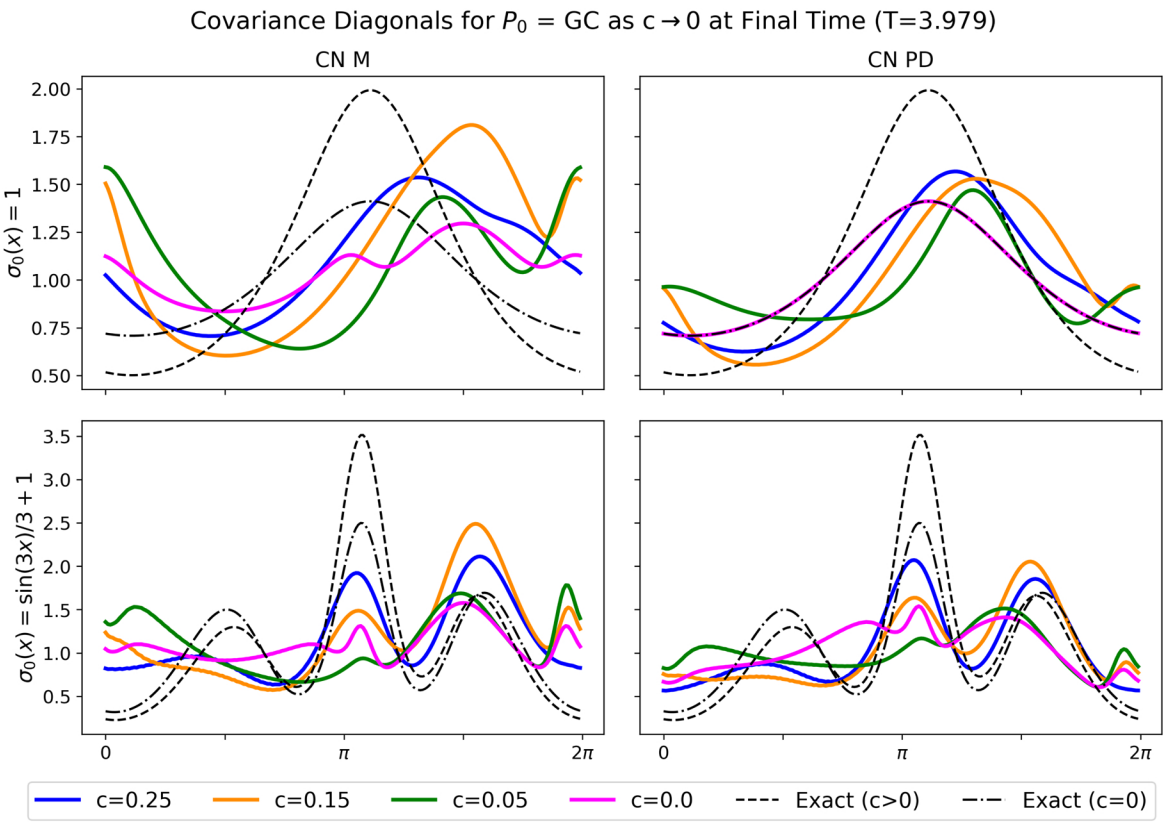

Figure 8. *Covariance diagonals extracted from Crank–Nicolson traditional and polar decomposition propagation methods at the final time T for GC initial correlations as c approaches zero. Top row: Constant initial variance. Bottom row: Spatially varying initial variance. Left column: Propagation via traditional Crank–Nicolson (CN M). Right column: Propagation via Crank–Nicolson polar decomposition (CN PD). Black curves are the exact diagonals, dashed curves are covariances with nonzero initial correlation lengths (variance solution), and dot-dashed curves are covariances with zero initial correlation length (continuous spectrum solution). These exact curves here labeled as* Exact $(c > 0)$ *and* Exact $(c = 0)$ *are the same exact curves labeled as* Exact $\sigma^2(x,t)$ *and* Exact $P^c(x,t)$ *in Figure 2. In the top right panel, the exact curve for $c = 0$ (black dot-dashed) and the CN PD curve for $c = 0$ (magenta) identically overlap.*

The errors in the discrete diagonal propagation are not limited to the final time; errors start to accumulate early on in the propagation cycle. We show this for the GC case in Figure 10, where the FOAR results are similar but are not shown. We can see clearly that rather than approximating the variance for $c > 0$ (black dashed lines), the Crank–Nicolson schemes (blue) tend to approximate the continuous spectrum ($c = 0$, brown), which is not the correct diagonal behavior for this case. Along with the diagonals extracted from the propagated covariances in Figure 10, we include the variance propagated independently by solving the one-dimensional version of (1.8) using the Crank–Nicolson scheme, as was shown in Figure 2. Including the variance solution computed using the Crank–Nicolson scheme in Figure 10 emphasizes that propagating the covariance diagonal independently using the known diagonal dynamics significantly reduces the errors in the variance compared to the diagonal extracted from the propagated covariance matrix.

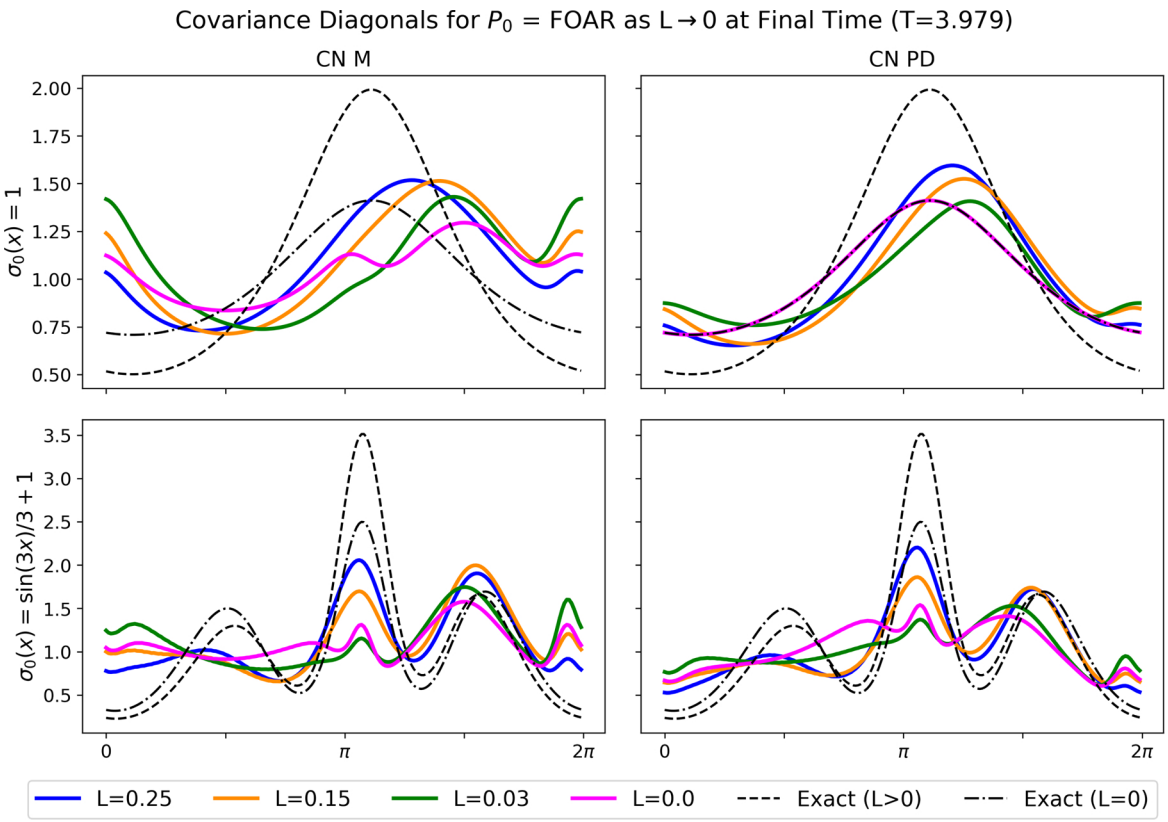

Figure 9. *Same as Figure 8 for FOAR correlation function initial covariances.*

3.3. Interpretation of the experimental results. Grey regions in Figure 10 correspond to where the exact variance solution is larger than the exact continuous spectrum solution, and these regions tend to correspond to where the numerically propagated diagonal for the short, nonzero initial correlation length (blue) exhibits variance loss. Further insight into this behavior can be gained by returning to the generalized continuum problem presented in section 2. Assuming $P_0^d(\boldsymbol{x}, \boldsymbol{x}) = P_0^c(\boldsymbol{x})$, as is the case in Figure 10, by multiplying the variance equation (2.34) by P^c and the continuous spectrum equation (2.33) by σ^2, taking the difference, and dividing by $(P^c)^2$, we find from the quotient rule that the ratio

$$\frac{\sigma^2(\boldsymbol{x}, t)}{P^c(\boldsymbol{x}, t)} = m(\boldsymbol{x}, t) \tag{3.7}$$

satisfies the continuity equation with unit initial condition,

$$\begin{aligned} m_t + \boldsymbol{v} \cdot \boldsymbol{\nabla} m + (\boldsymbol{\nabla} \cdot \boldsymbol{v}) m = 0, \\ m(\boldsymbol{x}, t_0) = 1, \end{aligned} \tag{3.8}$$

for all $\boldsymbol{x} \in \Omega$ and $t \geq t_0$. From (3.8), the ratio σ^2/P^c must be conserved, and this holds for the generalized variance and continuous spectrum equations (2.33) and (2.34)—not just

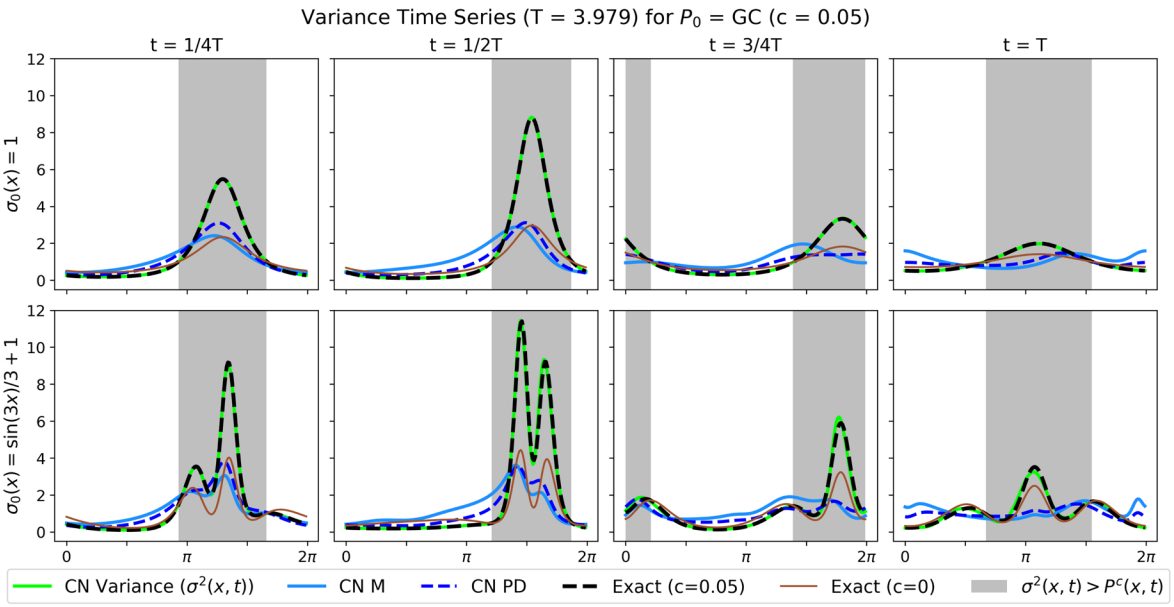

Figure 10. *Variance time series for GC case with $c = 0.05$. Top row: Constant initial variance. Bottom row: Spatially varying initial variance. Diagonals extracted from the Crank–Nicolson traditional and polar decomposition propagation are shown (solid and dashed blue lines) as well as propagating the variance by solving the variance equation (one-dimensional version of (1.8)) independently using Crank–Nicolson (green). The exact solutions to the variance and continuous spectrum equations are shown in black dashed lines and solid brown lines and are denoted as* Exact ($c = 0.05$) *and* Exact ($c = 0$), *respectively. Regions where the exact variance is larger (smaller) than the exact continuous spectrum are highlighted in grey (white).*

those presented in the numerical experiments. In regions where $m > 1$ we have $\sigma^2 > P^c$, and conversely in regions where $m < 1$ we have $\sigma^2 < P^c$. In fact, the function m of (3.8) on S_1^1 with velocity field (3.1) can be expressed explicitly using (B.7), taking one as the initial condition. Solving $m = 1$, or equivalently $v(s(x,t)) - v(x) = 0$ (following the notation of Appendix B), determines the boundaries between regions where m is less than or greater than one in S_1^1 at every time $t > t_0$.

For short initial correlation lengths, if the numerical schemes are better approximating the continuous spectrum solution P^c, regions where $m > 1$ should correspond to variance loss and regions where $m < 1$ should correspond to variance gain. Therefore, we refer to the ratio m as the *amplitude index*, which can be used to interpret the variance loss and gain observed in our numerical results. For example, regions where the index m is greater than one, i.e., $\sigma^2 > P^c$, highlighted in grey in Figure 10, generally do coincide with regions where we see variance loss in the Crank–Nicolson traditional and polar decomposition schemes. Conversely, in the unshaded regions of Figure 10 where the index m is less than one, i.e., $\sigma^2 < P^c$, the Crank–Nicolson traditional and polar decomposition schemes generally exhibit variance gain. Hence, we can exploit the amplitude index m to indicate regions of variance loss and gain.

Since the amplitude index m must be conserved according to (3.8), for a typical velocity field $\boldsymbol{v}$ there will always be regions of the spatial domain where m is greater than one and regions where m is less than one at any given time t. This implies that we should see both

loss and gain of variance at individual times; whether there will be global variance loss or variance gain depends on the details of the velocity field $\boldsymbol{v}$. In all the numerical results, Figures 3–10, we see both spurious variance loss and gain as a result of inaccurate discrete diagonal propagation. Data assimilation literature tends to focus on loss of variance because of its known negative impact on data assimilation schemes [21, sec. 8.8, 28, sec. 9.2, 1, 18, 27]. However, we see here for a general advective system that both loss and gain exist as reflected by the conservation law (3.8), cumulatively causing wholly inaccurate discrete diagonal propagation. Variance inflation, a tool often used in data assimilation practice to combat variance loss by rescaling the variance, typically by a multiplicative or additive factor, could perhaps be optimally adjusted in this context. Distinguishing between regions where the index m is greater than or less than one in general isolates the regions of variance loss and gain, respectively; thus the inflation factor could be tuned to only inflate where we expect variance loss. This would avoid using a single scale-factor that inflates the whole variance function and would prevent variance inflation in regions where we already see variance gain.

Since covariance information on and close to the diagonal may be sufficient information for many applications, local covariance evolution, where the variance and correlation lengths are propagated rather than the full matrix, may prove useful. Reference [6] first discussed local covariance evolution through continuum analysis of hyperbolic and parabolic PDEs, similar to the equations discussed here. The covariance equation in $2N$ spatial dimensions, where N is the number of spatial dimensions of the state, is reduced to a system of auxiliary PDEs in N dimensions consisting of variance and correlation length equations, which approximate the covariance locally. The parametric Kalman filter (PKF) [34, 32, 33] applies the ideas of [6] to advective-diffusive dynamics and the Burgers equation, where the PKF evolves variance and local diffusion tensor dynamics to approximate covariance matrices. Our results, together with the results of [6] and [34, 32, 33], suggest further investigation into local covariance propagation, which may help reduce computational expense and the spurious loss and gain of variance observed in full covariance propagation.

4. Conclusions and discussion. In this work, we study covariance propagation associated with random state variables governed by a generalized advection equation. We do this in an effort to understand the root causes of spurious loss of variance, which is observed in data assimilation schemes wherein the covariance is explicitly or implicitly propagated. Using a continuum analysis to guide the numerical results, new insights are gained through the detailed study of both the continuum and discrete covariance evolution. The main conclusions are as follows:

1. Continuum analysis of the state and covariance equations (2.1) and (2.8) is necessary to establish a fundamental understanding of the covariance evolution. In particular, the continuum analysis uncovers the discontinuous behavior of the dynamics along $\boldsymbol{x}_1 = \boldsymbol{x}_2$ as the correlation length approaches zero, for example in the vicinity of sharp gradients, which is an insight crucial for understanding the spurious loss and gain of variance observed in our numerical experiments.
2. Comparison of the numerical results with the continuum analysis shows that full-rank covariance propagation via (1.3) typically results in considerable spurious loss

of variance. This is due to the peculiar discontinuous behavior of the dynamics far more than to any numerical dissipation. Discrete propagation using (1.3) produces inaccurate covariance matrix diagonals, resulting in both variance loss in regions where the amplitude index m is greater than one and variance gain in regions where this index m is less than one. When propagating initial covariances with short, nonzero correlation lengths, the numerical schemes better approximate the dynamics of the zero correlation length case than those of the nonzero correlation length case.

3. Isolating the variance and propagating it independently eliminates the variance loss and gain observed during full-rank covariance propagation and yields accurate propagation because it adheres to the continuum dynamics. This result suggests further investigation into alternate methods of covariance propagation.

Though the continuum analysis performed here may not be tractable in all situations, it serves as a foundation for understanding covariance evolution and interpreting the results of our numerical experiments. Central to the continuum analysis is the polar decomposition of the fundamental solution operator (2.26), which is general in that it is a canonical form for all bounded linear operators on Hilbert spaces. Thus, we expect that much of the work presented here can be extended to general hyperbolic systems of PDEs having a quadratic energy functional, e.g., [7].

It is important to recognize that the loss of variance observed in our numerical experiments is a result of the discontinuous change in continuum covariance dynamics discussed in subsection 2.3. Even when propagating covariance matrices using a fully Lagrangian scheme, as done in [27], propagated covariances still suffer from spurious loss of variance that is not due to the numerical scheme but rather to the discontinuous change in dynamics that we've identified in this work. Covariance propagation tends to be overlooked in the data assimilation literature as a potential source of variance loss, particularly when using the same numerical method that propagates the state. Data assimilation schemes that do not propagate the covariance explicitly may experience errors similar to what we observe here because the underlying cause of these errors is the covariance dynamics—not the numerical scheme. Studying full-rank covariance propagation as in (1.3) isolates the spurious loss of variance as an issue with covariance dynamics and implies that approximations to (1.3), such as in ensemble Kalman filters, can suffer variance loss in a similar manner. The errors caused during the covariance propagation may be a neglected source of the model error or "system error" observed in data assimilation schemes [19, p. 3285].

Our work brings to light a fundamental issue associated with current approaches to numerical covariance propagation and recommends investigation into alternate methods of covariance propagation. Figures 2 and 10, which display the independent variance propagation for example, suggest that local covariance evolution may be an adequate alternative. As discussed in [6, 34, 32, 33], for applications in which information on and close to the diagonal is sufficient, evolving the variance and correlation lengths themselves may serve as an alternative to full covariance evolution.

Appendix A. Proofs of solutions with Dirac delta initial conditions. We begin by verifying that $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = d^2(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where d^2 satisfies (2.24), is the solution to the covariance evolution equation (2.8) for $P_0(\boldsymbol{x}_1, \boldsymbol{x}_2) = \delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

Proof. The function $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = d^2(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a weak (distribution) solution of (2.8) with $P_0(\boldsymbol{x}_1, \boldsymbol{x}_2) = \delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ if

$$(\mathcal{L}^*\phi, P) = 0 \tag{A.1}$$

for all test functions $\phi \in C^1(\Omega \times \Omega \times [t_0, T])$ with period $t = T$, where $\mathcal{L}^*$ is the adjoint differential operator

$$\mathcal{L}^* = -\partial_t - \boldsymbol{v}_1 \cdot \boldsymbol{\nabla}_1 - \boldsymbol{v}_2 \cdot \boldsymbol{\nabla}_2 + b_1 + b_2 - \boldsymbol{\nabla}_1 \cdot \boldsymbol{v}_1 - \boldsymbol{\nabla}_2 \cdot \boldsymbol{v}_2 \tag{A.2}$$

corresponding to the differential operator

$$\mathcal{L} = \partial_t + \boldsymbol{v}_1 \cdot \boldsymbol{\nabla}_1 + \boldsymbol{v}_2 \cdot \boldsymbol{\nabla}_2 + b_1 + b_2 \tag{A.3}$$

of (2.8). Here we denote $(\cdot, \cdot)$ as the inner product over $L^2(\Omega \times \Omega \times [t_0, T])$ and we will denote $(\cdot, \cdot)'$ as the inner product over $L^2(\Omega \times [t_0, T])$. Substituting $P = d^2(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ into the expression for $(\mathcal{L}^*\phi, P)$, applying the Dirac delta, and expanding the result yields

$$\begin{aligned}(\mathcal{L}^*\phi, P) = (&-\phi_t(\boldsymbol{x}_1, \boldsymbol{x}_1, t) - \boldsymbol{\nabla}_1 \cdot (\boldsymbol{v}_1\phi(\boldsymbol{x}_1, \boldsymbol{x}_1, t)) \\ &+ (2b_1 - \boldsymbol{\nabla}_1 \cdot \boldsymbol{v}_1)\phi(\boldsymbol{x}_1, \boldsymbol{x}_1, t), d^2(\boldsymbol{x}_1, t))',\end{aligned} \tag{A.4}$$

where we note that $\boldsymbol{\nabla}_2 \cdot \boldsymbol{v}_2 \,|_{\boldsymbol{x}_2 = \boldsymbol{x}_1} = \boldsymbol{\nabla}_1 \cdot \boldsymbol{v}_1$ and $\boldsymbol{\nabla}_1\phi(\boldsymbol{x}_1, \boldsymbol{x}_1, t) = \boldsymbol{\nabla}_1\phi(\boldsymbol{x}_1, \boldsymbol{x}_2, t)|_{\boldsymbol{x}_2 = \boldsymbol{x}_1} + \boldsymbol{\nabla}_2\phi(\boldsymbol{x}_1, \boldsymbol{x}_2, t)|_{\boldsymbol{x}_2 = \boldsymbol{x}_1}$. Using integration by parts to move derivatives off of the test function ϕ onto d^2 in (A.4) gives us

$$(\mathcal{L}^*\phi, P) = (\phi(\boldsymbol{x}_1, \boldsymbol{x}_1, t), d_t^2 + \boldsymbol{v}_1 \cdot \boldsymbol{\nabla}_1 \, d^2 + (2b_1 - \boldsymbol{\nabla}_1 \cdot \boldsymbol{v}_1)d^2)' = 0, \tag{A.5}$$

since d^2 satisfies (2.24). Thus, $P(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = d^2(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a weak solution to (2.8). ∎

To show that $\tilde{P} = \tilde{P}^c(\boldsymbol{x}_1, t)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is the solution to (2.29) for $\tilde{P}_0 = P_0^c(\boldsymbol{x}_1)\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$ follows an identical proof as given above, where we take

$$b_i \mapsto \frac{1}{2}\boldsymbol{\nabla}_i \cdot \boldsymbol{v}_i, \;\; i = 1, 2, \tag{A.6}$$

$$d^2(\boldsymbol{x}_1, t) \mapsto \tilde{P}^c(\boldsymbol{x}_1, t), \tag{A.6}$$

where $\tilde{P}^c$ satisfies (2.31).

Appendix B. Solution to the state equation. In this appendix we solve the one-dimensional version of (1.4) used in the numerical experiments with velocity field (3.1),

$$\begin{aligned} q_t + vq_x + v'q &= 0, \;\; x \in S_1^1, \\ q(x, t_0) &= q_0(x), \\ v &= \sin(x) + 2. \end{aligned} \tag{B.1}$$

To solve (B.1), we start by rewriting the equation in terms of the Lagrangian/total derivative

$$\frac{D}{Dt} \equiv \partial_t + v\partial_x. \tag{B.2}$$

By doing so, (B.1) becomes the ordinary differential equation (1-D version of (1.4))

$$\begin{aligned} \frac{Dq(x(t),t)}{Dt} &= -v'(x(t))q(x(t),t), \\ q(x,t_0) &= q_0(x). \end{aligned} \tag{B.3}$$

Associated with (B.3) are the Lagrangian trajectories, or characteristic equations, that determine how the spatial variable x evolves over time:

$$\begin{aligned} \frac{dx(t)}{dt} &= v(x(t)), \\ x(t_0) &= s. \end{aligned} \tag{B.4}$$

Here, we take s as a general initial parameter and define our solutions to (B.4) as $x(t;s)$, following the notation of [37, Ch. 3] and discussion in section 1.

Since the state equation is reduced to an ordinary differential equation in (B.3), solutions are of the form

$$q(x,t) = q_0(x(t_0))e^{\int_{t_0}^{t} -v'(x(\tau))d\tau}. \tag{B.5}$$

We can solve the integral in the exponential of (B.5) using a simple u–substitution of $u = x(t)$ along with (B.4),

$$\int_{t_0}^{t} -v'(x(\tau))d\tau = \int_{s}^{x(t)} -\frac{v'(u)}{v(u)}du = \ln\left(\frac{v(s)}{v(x(t))}\right). \tag{B.6}$$

To rewrite (B.6) in terms of (x,t), we first solve for $x(t;s)$ using the characteristic equation (B.4), then invert to determine $s = s(x,t)$, which can be done since the velocity field v is continuously differentiable [37, Ch. 3]. Therefore, substituting (B.6) and $x(t_0) = s(x,t)$ into (B.5), we have the explicit solution to (B.1),

$$q(x,t) = q_0(s(x,t))\left(\frac{\sin(s(x,t))+2}{\sin(x)+2}\right), \tag{B.7}$$

$$s(x,t) = 2\tan^{-1}\left(\frac{\sqrt{3}}{2}\tan\left(\tan^{-1}\left(\frac{2\tan(x/2)+1}{\sqrt{3}}\right) - \frac{\sqrt{3}t}{2}\right) - \frac{1}{2}\right). \tag{B.8}$$

The function $\tan^{-1}$ refers to the principle branch of inverse tangent, where $y = \tan^{-1}(x)$, $x \in \mathbb{R}$, $y \in (-\pi/2, \pi/2)$, defined by solving $x = \tan(y)$ [39, sec. 6.3.3]. Figure 11 plots (B.7) for two different initial conditions, along with solving (B.1) using Lax–Wendroff and Crank–Nicolson schemes on a uniform grid of 200 grid points and unit Courant–Friedrichs–Lewy number.

The amplitude index m of (3.8) on S_1^1 with velocity field (3.1) can be expressed explicitly using (B.7) taking one as the initial condition. The exact solution d of (2.18) on S_1^1 for $b = v_x$, as in the numerical experiments, can also be computed explicitly from (B.7) (as d^2 in this case satisfies (B.1) with unit initial condition).

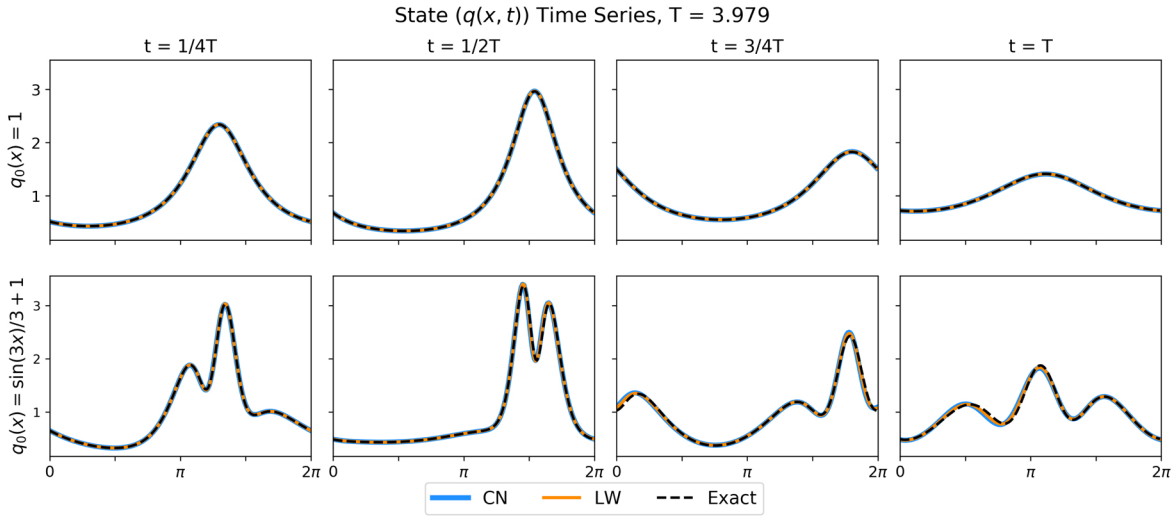

Figure 11. *Solutions to* (B.1) *for* $q_0(x) = 1$ *(top row) and* $q_0(x) = \sin(3x)/3 + 1$ *(bottom row) at four different times. The exact solution (black dashed line) is shown along with finite-difference solutions using Crank–Nicolson (blue) and Lax–Wendroff (orange). The finite difference solutions are nearly indistinguishable from the exact solution.*

Acknowledgments. The authors would like to thank the two anonymous referees for their constructive feedback and suggestions.

REFERENCES

[1] J. L. Anderson and S. L. Anderson, *A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts*, Mon. Weather Rev., 127 (1999), pp. 2741–2758, https://doi.org/10.1175/1520-0493(1999)127⟨2741:AMCIOT⟩2.0.CO;2.

[2] J. Barkmeijer, M. Van Gijzen, and F. Bouttier, *Singular vectors and estimates of the analysis-error covariance metric*, Q. J. Roy. Meteor. Soc., 124 (1998), pp. 1695–1713, https://doi.org/10.1002/qj.49712454916.

[3] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. d. l. Cámara, H. M. Christensen, M. Colangeli, D. R. B. Coleman, D. Crommelin, S. I. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Weniger, P. D. Williams, and J.-I. Yano, *Stochastic parameterization: Toward a new view of weather and climate models*, Bull. Amer. Meteorol. Soc., 98 (2017), pp. 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1.

[4] G. Beylkin and J. Keiser, *An adaptive pseudo-wavelet approach for solving nonlinear partial differential equations*, in Multiscale Wavelet Methods for Partial Differential Equations, Wavelet Anal. Appl. 6, Academic Press, San Diego, CA, 1997, pp. 137–197.

[5] R. Buizza and T. Palmer, *The singular-vector structure of the atmospheric global circulation*, J. Atmos. Sci., 52 (1995), pp. 1434–1456, https://doi.org/10.1175/1520-0469(1995)052⟨1434:TSVSOT⟩2.0.CO;2.

[6] S. E. Cohn, *Dynamics of short-term univariate forecast error covariances*, Mon. Weather Rev., 121 (1993), pp. 3123–3149, https://doi.org/10.1175/1520-0493(1993)121⟨3123:DOSTUF⟩2.0.CO;2.

[7] S. E. Cohn, *The principle of energetic consistency in data assimilation*, in Data Assimilation, Springer, Berlin, Heidelberg, 2010, pp. 137–216, https://doi.org/10.1007/978-3-540-74703-1˙7.

[8] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol.* II: *Partial Differential Equations*, Interscience Publishers, New York, 1962.
[9] P. Courtier and O. Talagrand, *Variational assimilation of meteorological observations with the adjoint vorticity equation*, II: *Numerical results*, Q. J. Roy. Meteor. Soc, 113 (1987), pp. 1329–1347, https://doi.org/10.1002/qj.49711347813.
[10] J. Crank and P. Nicolson, *A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type*, Math. Proc. Cambridge Philos. Soc., 43 (1947), pp. 50–67, https://doi.org/10.1017/S0305004100023197.
[11] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[12] M. Ehrendorfer and J. Tribbia, *Optimal prediction of forecast error covariances through singular vectors*, J. Atmos. Sci, 54 (1997), pp. 286–313, https://doi.org/10.1175/1520-0493(1999)127⟨2741:AMCIOT⟩2.0.CO;2.
[13] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res., 99 (1994), pp. 10143–10162, https://doi.org/10.1029/94JC00572.
[14] R. Furrer and T. Bengtsson, *Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants*, J. Multivariate Anal., 98 (2007), pp. 227–255, https://doi.org/10.1016/j.jmva.2006.08.003.
[15] G. Gaspari and S. E. Cohn, *Construction of correlation functions in two and three dimensions*, Q. J. Roy. Meteor. Soc., 125 (1999), pp. 723–757, https://doi.org/10.1002/qj.49712555417.
[16] G. Gaspari, S. E. Cohn, J. Guo, and S. Pawson, *Construction and application of covariance functions with variable length-fields*, Q. J. Roy. Meteor. Soc., 132 (2006), pp. 1815–1838, https://doi.org/10.1256/qj.05.08.
[17] J. Holton and G. J. Hakim, *An Introduction to Dynamic Meteorology*, 5th ed., Academic Press, New York, 2013.
[18] P. L. Houtekamer and H. L. Mitchell, *An adaptive ensemble Kalman filter*, Mon. Weather Rev., 128 (2000), pp. 416–433, https://doi.org/10.1175/1520-0493(2000)128⟨0416:AAEKF⟩2.0.CO;2.
[19] P. L. Houtekamer and H. L. Mitchell, *Ensemble Kalman filtering*, Q. J. Roy. Meteor. Soc., 131 (2005), pp. 3269–3289, https://doi.org/10.1256/qj.05.135 .
[20] J. K. Hunter and B. Nachtergaele, *Applied Analysis*, World Scientific, River Edge, NJ, 2001.
[21] A. H. Jazwinski, *Stochastic Processes and Filter Theory*, Academic Press, New York, 1970.
[22] R. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME Ser. D. J. Basic Engrg., 82 (1960), pp. 35–45, https://doi.org/10.1115/1.3662552.
[23] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK, 2003.
[24] P. Lax and B. Wendroff, *Systems of conservation laws*, Comm. Pure Appl. Math., 13 (1960), pp. 217–237, https://doi.org/10.1002/cpa.3160130205.
[25] C. E. Leith, *Theoretical skill of Monte Carlo forecasts*, Mon. Weather Rev., 102 (1974), pp. 409–418.
[26] E. Lorenz, *A study of the predictability of a* 28*-variable atmospheric model*, Tellus, 17 (1965), pp. 321–333, https://doi.org/10.3402/tellusa.v17i3.9076.
[27] P. M. Lyster, S. E. Cohn, B. Zhang, L.-P. Chang, R. Ménard, K. Olson, and R. Renka, *A Lagrangian trajectory filter for constituent data assimilation*, Q. J. Roy. Meteor. Soc., 130 (2004), pp. 2315–2334, https://doi.org/10.1256/qj.02.234.
[28] P. S. Maybeck, Stochastic Models, Estimation and Control, *Vol.* 2, Academic Press, London, 1982.
[29] R. Ménard, S. E. Cohn, L. P. Chang, and P. M. Lyster, *Assimilation of stratospheric chemical tracer observations using a Kalman filter, Part I: Formulation*, Mon. Weather Rev., 128 (2000), pp. 2654–2671, https://doi.org/10.1175/1520-0493(2000)128.
[30] R. Ménard, S. Skachko, and O. Pannekoucke, *Numerical discretization causing error variance loss and the need for inflation*, Q. J. Roy. Meteor. Soc., 147 (2021), pp. 3498–3520, https://doi.org/10.1002/qj.4139.
[31] H. L. Mitchell and P. L. Houtekamer, *An adaptive ensemble Kalman filter*, Mon. Weather. Rev., 128 (2000), pp. 416–433, https://doi.org/10.1175/1520-0493(2000)128.

[32] O. Pannekoucke, M. Bocquet, and R. Ménard, *Parametric covariance dynamics for the nonlinear diffusive Burgers equation*, Nonlinear Processes Geophys., 25 (2018), pp. 481–495, https://doi.org/10.5194/npg-25-481-2018.

[33] O. Pannekoucke, R. Ménard, M. E. Aabaribaoune, and M. Plu, *A methodology to obtain model-error covariances due to the discretization scheme from the parametric Kalman filter perspective*, Nonlinear Processes Geophys., 28 (2021), pp. 1–22, https://doi.org/10.5194/npg-28-1-2021.

[34] O. Pannekoucke, S. Ricci, S. Barthelemy, R. Ménard, and O. Thual, *Parametric Kalman Filter for chemical transport models*, Tellus A, 68 (2016), 31547–31561, https://doi.org/10.3402/tellusa.v68.31547.

[35] N. A. Phillips, *The spatial statistics of random geostrophic modes and first-guess errors*, Tellus A, 38A (1986), pp. 314–332, https://doi.org/10.1111/j.1600-0870.1986.tb00418.x.

[36] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, New York, 1972.

[37] M. Shearer and R. Levy, *Partial Differential Equations: An Introduction to Theory and Applications*, Princeton University Press, Princeton, NJ, 2015.

[38] O. Talagrand and P. Courtier, *Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory,* Q. J. Roy. Meteor. Soc., 113 (1987), pp. 1311–1328, https://doi.org/10.1002/qj.49711347812.

[39] D. Zwillinger, ed., *CRC Standard Mathematical Tables and Formulae*, 30th ed., CRC Press, Boca Raton, FL, 1996.