University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Lee, Wonjung, McDougall, D. and Stuart, A. M.

Article Title: Kalman filtering and smoothing for linear wave equations with model error

Year of publication: 2011

Link to published article:

http://dx.doi.org/10.1088/0266-5611/27/9/095008

Publisher statement: Published in Inverse Problems, (2011). 29(7)

Kalman filtering and smoothing for linear wave equations with model error

Wonjung Lee[1,2]**, D. McDougall**[1] **and A.M. Stuart**[1]

[1] Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

[2] Mathematical Institute, University of Oxford, Oxford, OX1 3LB, UK

E-mail:

Wonjung.Lee@maths.ox.ac.uk

D.McDougall@warwick.ac.uk

A.M.Stuart@warwick.ac.uk   http://www.warwick.ac.uk/∼masdr/

Abstract. Filtering is a widely used methodology for the incorporation of observed data into time-evolving systems. It provides an online approach to state estimation inverse problems when data is acquired sequentially. The Kalman filter plays a central role in many applications because it is exact for linear systems subject to Gaussian noise, and because it forms the basis for many approximate filters which are used in high dimensional systems. The aim of this paper is to study the effect of model error on the Kalman filter, in the context of linear wave propagation problems. A consistency result is proved when no model error is present, showing recovery of the true signal in the large data limit. This result, however, is not robust: it is also proved that arbitrarily small model error can lead to inconsistent recovery of the signal in the large data limit. If the model error is in the form of a constant shift to the velocity, the filtering and smoothing distributions only recover a partial Fourier expansion, a phenomenon related to aliasing. On the other hand, for a class of wave velocity model errors which are time-dependent, it is possible to recover the filtering distribution exactly, but not the smoothing distribution. Numerical results are presented which corroborate the theory, and also to propose a computational approach which overcomes the inconsistency in the presence of model error, by relaxing the model.

1. Introduction

Filtering is a methodology for the incorporation of data into time-evolving systems [1, 2]. It provides an online approach to state estimation inverse problems when data is acquired sequentially. In its most general form the dynamics and/or observing system are subject to noise and the objective is to compute the probability distribution of the current state, given observations up to the current time, in a sequential fashion. The Kalman filter [3] carries out this process exactly for linear dynamical systems subject to additive Gaussian noise. A key aspect of filtering in many applications is to understand

the effect of model error – the mismatch between model used to filter and the source of the data itself. In this paper we undertake a study of the effect of model error on the Kalman filter in the context of linear wave problems.

Section 2 is devoted to describing the linear wave problem of interest, and deriving the Kalman filter for it. The iterative formulae for the mean and covariance are solved and the equivalence (as measures) of the filtering distribution at different times is studied. In section 3 we study consistency of the filter, examining its behaviour in the large time limit as more and more observations are accumulated, at points which are equally spaced in time. It is shown that, in the absence of model error, the filtering distribution on the current state converges to a Dirac measure on the truth. However, for the linear advection equation, it is also shown that arbitrarily small model error, in the form of a shift to the wave velocity, destroys this property: the filtering distribution converges to a Dirac measure, but it is not centred on the truth. Thus the order of two operations, namely the successive incorporation of data and the limit of vanishing model error, cannot be switched; this means, practically, that small model error can induce order one errors in filters, even in the presence of large amounts of data. All of the results in section 3 apply to the smoothing distribution on the initial condition, as well as the filtering distribution on the current state. Section 4 concerns non-autonomous systems, and the effect of model error. We study the linear advection equation in two dimensions with time-varying wave velocity. Two forms of model error are studied: an error in the wave velocity which is integrable in time, and a white noise error. In the first case it is shown that the filtering distribution converges to a Dirac measure on the truth, whilst the smoothing distribution converges to a Dirac measure which is in error, i.e., not centred on the truth. In the second, white noise, case both the filter and smoother converge to a Dirac measure which is in error. In section 5 we present numerical results which illustrate the theoretical results. We also describe a numerical approach which overcomes the effect of model error by relaxing the model to allow the wave velocity to be learnt from the data.

We conclude the introduction with a brief review of the literature in this area. Filtering in high dimensional systems is important in a range of applications, especially within the atmospheric and geophysical sciences [4, 5, 6, 7]. Recent theoretical studies have shown how the Kalman filter can not only systematically incorporate data into a model, but also stabilize model error arising from an unstable numerical discretization [8], from an inadequate choice of parameter [9, 10], and the effect of physical model error in a particular application is discussed in [11]. The Kalman filter is used as the basis for a number of approximate filters which are employed in nonlinear and non-Gaussian problems, and the ensemble Kalman filter in particular is widely used in this context [5, 12, 13]. Although robust and widely useable, the ensemble Kalman filter does not provably reproduce the true distribution of the signal in the large ensemble limit, given data, except in the Gaussian case. For this reason it would be desirable to use the particle filter [2] on highly non-Gaussian systems. However, recent theoretical work and a range of computational experience shows that, in its current form, particle

filters will not work well in high dimensions [14, 15, 16]. As a consequence a great deal of research activity is aimed at the development of various approximation schemes within the filtering context; see [17, 18, 19, 20, 21, 22] for example. The subject of consistency of Bayesian estimators for noisily observed systems, which forms the core of our theoretical work in this paper, is an active area of research. In the infinite dimensional setting, much of this work is concerned with linear Gaussian systems, as we are here, but is primarily aimed at the perfect model scenario [23, 24, 25]. The issue of model error arising from spatial discretization, when filtering linear PDEs, is studied in [26]. The idea of relaxing the model and learning parameters in order to obtain a better fit to the data, considered in our numerical studies, is widely used in filtering (see the chapter by Künsch in [2], and the paper [27] for an application in data assimilation).

2. Kalman Filter on Function Space

2.1. Statistical model for discrete observations

The test model, which we propose here, is a class of PDEs

$$\partial_t v + \mathcal{L}v = 0, \quad \forall (x,t) \in \mathbf{T}^2 \times (0, \infty) \tag{1}$$

on a two dimensional torus. Here $\mathcal{L}$ is an anti-Hermitian operator satisfying $\mathcal{L}^* = -\mathcal{L}$ where $\mathcal{L}^*$ is the adjoint in $L^2(\mathbf{T}^2)$. Equation (1) describes a prototypical linear wave system and the advection equation with velocity c is the simplest example:

$$\partial_t v + c \cdot \nabla v = 0, \quad \forall (x,t) \in \mathbf{T}^2 \times (0, \infty). \tag{2}$$

The state estimation problem for Equation (1) requires to find the 'best' estimate of the solution $v(t)$ (shorthand for $v(\cdot, t)$) given a random initial condition v_0 and a set of noisy observations, called data. Suppose the data is collected at discrete times $t_n = n\Delta t$, then we assume that the entire $v_n = v(t_n)$ solution on the torus is observed with additive noise η_n at time t_n, and further that the η_n are independent for different n. Realizations of the noise η_n are $L^2(\mathbf{T}^2)$-valued random fields and the observations y_n are given by

$$y_n = v_n + \eta_n = e^{-t_n \mathcal{L}} v_0 + \eta_n, \quad \forall x \in \mathbf{T}^2. \tag{3}$$

Here $e^{-t\mathcal{L}}$ denotes the forward solution operator for (1) through t time units. Let $Y_N = \{y_1, \ldots, y_N\}$ be the collection of data up to time t_N, then we are interested in finding the conditional distribution $\mathbb{P}(v_n|Y_N)$ on the Hilbert space $L^2(\mathbf{T}^2)$. If $n = N$, this is called the filtering problem, if $n < N$ it is the smoothing problem and for $n > N$, the prediction problem. We here emphasize that all of the problems are equivalent for our deterministic system in that any one measure defines the other simply by a push forward under the linear map defined by Equation (1).

In general calculation of the filtering distribution $\mathbb{P}(v_n|Y_n)$ is computationally challenging when the state space for v_n is large, as it is here. A key idea is to estimate the signal v_n through sequential updates consisting of a two-step process: prediction by time evolution, and analysis through data assimilation. We first perform a one-step statistical prediction to obtain $\mathbb{P}(v_{n+1}|Y_n)$ from $\mathbb{P}(v_n|Y_n)$ through some forward operator. This is

followed by an analysis step which corrects the probability distribution on the basis of the statistical input of noisy observations of the system using Bayes rule:

$$\frac{\mathbb{P}(v_{n+1}|Y_{n+1})}{\mathbb{P}(v_{n+1}|Y_n)} \propto \mathbb{P}\left(y_{n+1}|v_{n+1}\right). \tag{4}$$

This relationship exploits the assumed independence of the observational noises η_n for different n. In our case, where the signal v_n is a function, this identity should be interpreted as providing the Radon-Nikodym derivative (density) of the measure $\mathbb{P}(dv_{n+1}|Y_{n+1})$ with respect to $\mathbb{P}(dv_{n+1}|Y_n)$ [28].

In general implementation of this scheme is non-trivial as it requires approximation of the probability distributions at each step indexed by n. In the case of infinite dimensional dynamics this may be particularly challenging. However, for linear dynamical systems such as (1), together with linear observations (3) subject to Gaussian observation noise η_n this may be achieved by use of the Kalman filter [3]. Our work in this paper begins with Theorem 2.1, which is a straightforward extension of the traditional Kalman filter theory in finite dimension to measures on an infinite dimensional function space. For reference, basic results concerning Gaussian measures required for this paper are gathered in Appendix A.

Theorem 2.1. *Let v_0 be distributed according to a Gaussian measure $\mu_0 = \mathcal{N}\left(m_0, \mathcal{C}_0\right)$ on $L^2(\mathbf{T}^2)$ and let $\{\eta_n\}_{n\in\mathbb{N}}$ be i.i.d. draws from the $L^2(\mathbf{T}^2)$-valued Gaussian measure $\mathcal{N}\left(0, \Gamma\right)$. Assume further that v_0 and $\{\eta_n\}_{n\in\mathbb{N}}$ are independent of one another, and that $\mathcal{C}_0$ and Γ are strictly positive. Then the conditional distribution $\mathbb{P}(v_n|Y_n) \equiv \mu^n$ is a Gaussian $\mathcal{N}\left(m^n, \mathcal{C}^n\right)$ with mean and covariance satisfying the recurrence relations*

$$
\begin{aligned}
m^{n+1} &= e^{-\Delta t \mathcal{L}} m^n \\
&\quad - e^{-\Delta t \mathcal{L}} \mathcal{C}^n e^{-\Delta t \mathcal{L}^*} \left(\Gamma + e^{-\Delta t \mathcal{L}} \mathcal{C}^n e^{-\Delta t \mathcal{L}^*}\right)^{-1} \left(e^{-\Delta t \mathcal{L}} m^n - y_{n+1}\right), \tag{5a}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{C}^{n+1} &= e^{-\Delta t \mathcal{L}} \mathcal{C}^n e^{-\Delta t \mathcal{L}^*} \\
&\quad - e^{-\Delta t \mathcal{L}} \mathcal{C}^n e^{-\Delta t \mathcal{L}^*} \left(\Gamma + e^{-\Delta t \mathcal{L}} \mathcal{C}^n e^{-\Delta t \mathcal{L}^*}\right)^{-1} e^{-\Delta t \mathcal{L}} \mathcal{C}^n e^{-\Delta t \mathcal{L}^*}, \tag{5b}
\end{aligned}
$$

where $m^0 = m_0$, $\mathcal{C}^0 = \mathcal{C}_0$.

Proof. Let $m_{n|N} = \mathbb{E}(v_n|Y_N)$ and $\mathcal{C}_{n|N} = \mathbb{E}\left[(v_n - m_{n|N})(v_n - m_{n|N})^*\right]$ denote the mean and covariance operator of $\mathbb{P}(v_n|Y_N)$ so that $m^n = m_{n|n}$, $\mathcal{C}^n = \mathcal{C}_{n|n}$. Now the prediction step reads

$$
\begin{aligned}
m_{n+1|n} &= \mathbb{E}(e^{-\Delta t \mathcal{L}} v_n | Y_n) = e^{-\Delta t \mathcal{L}} m_{n|n}, \\
\mathcal{C}_{n+1|n} &= \mathbb{E}\left[e^{-\Delta t \mathcal{L}}(v_n - m_{n|n})(v_n - m_{n|n})^* e^{-\Delta t \mathcal{L}^*}\right] \\
&= e^{-\Delta t \mathcal{L}} \mathcal{C}_{n|n} e^{-\Delta t \mathcal{L}^*}. \tag{6}
\end{aligned}
$$

To get the analysis step, choose $x_1 = v_{n+1}|Y_n$ and $x_2 = y_{n+1}|Y_n$ then (x_1, x_2) is jointly Gaussian with mean $(m_{n+1|n}, m_{n+1|n})$ and each components of the covariance operator for (x_1, x_2) are given by

$$
\begin{aligned}
C_{11} &= \mathbb{E}\left[(x_1 - m_{n+1|n})(x_1 - m_{n+1|n})^*\right] = \mathcal{C}_{n+1|n}, \\
C_{22} &= \mathbb{E}\left[(x_2 - m_{n+1|n})(x_2 - m_{n+1|n})^*\right] = \Gamma + \mathcal{C}_{n+1|n},
\end{aligned}
$$

$$C_{12} = \mathbb{E}\left[(x_1 - m_{n+1|n})(x_2 - m_{n+1|n})^*\right] = \mathcal{C}_{n+1|n} = C_{21}.$$

Using Lemma A.2 we obtain

$$m_{n+1|n+1} = m_{n+1|n} - \mathcal{C}_{n+1|n}\left(\Gamma + \mathcal{C}_{n+1|n}\right)^{-1}\left(m_{n+1|n} - y_{n+1}\right),$$

$$\mathcal{C}_{n+1|n+1} = \mathcal{C}_{n+1|n} - \mathcal{C}_{n+1|n}\left(\Gamma + \mathcal{C}_{n+1|n}\right)^{-1}\mathcal{C}_{n+1|n}. \tag{7}$$

Combining Equations (6) and (7) yields Equation (5). $\qquad\square$

Note that μ_n, the distribution of $v_0|Y_n$, is also Gaussian and we denote its mean and covariance by m_n and $\mathcal{C}_n$ respectively. The measure μ_n is the image of μ^n under the linear transformation $e^{t_n\mathcal{L}}$. Thus we have $m_n = e^{t_n\mathcal{L}}m^n$ and $\mathcal{C}_n = e^{t_n\mathcal{L}}\mathcal{C}^n e^{t_n\mathcal{L}^*}$.

In this paper we study a wave propagation problem for which $\mathcal{L}$ is anti-Hermitian. Furthermore we assume that both $\mathcal{C}_0$ and Γ commute with $\mathcal{L}$. Then the formulae (5) simplify to give

$$m^{n+1} = e^{-\Delta t\mathcal{L}}m^n - \mathcal{C}^n\left(\Gamma + \mathcal{C}^n\right)^{-1}\left(e^{-\Delta t\mathcal{L}}m^n - y_{n+1}\right), \tag{8a}$$

$$\mathcal{C}^{n+1} = \mathcal{C}^n - \mathcal{C}^n\left(\Gamma + \mathcal{C}^n\right)^{-1}\mathcal{C}^n. \tag{8b}$$

The following gives explicit expressions for $(m_n, \mathcal{C}_n)$ and $(m^n, \mathcal{C}^n)$, based on the Equations (8).

Corollary 2.2. *Suppose that $\mathcal{C}_0$ and Γ commute with the anti-Hermitian operator $\mathcal{L}$, then the means and covariance operators of μ_n and μ^n are given by*

$$m_n = \left(nI + \Gamma\mathcal{C}_0^{-1}\right)^{-1}\left[\Gamma\mathcal{C}_0^{-1}m_0 + \sum_{l=0}^{n-1} e^{t_{l+1}\mathcal{L}}y_{l+1}\right], \tag{9a}$$

$$\mathcal{C}_n = \left(n\Gamma^{-1} + \mathcal{C}_0^{-1}\right)^{-1}, \tag{9b}$$

and $m^n = e^{-t_n\mathcal{L}}m_n$, $\mathcal{C}^n = \mathcal{C}_n$.

Proof. Assume for induction that $\mathcal{C}^n$ is invertible. Then the identity

$$\left(\Gamma^{-1} + (\mathcal{C}^n)^{-1}\right)\left(\mathcal{C}^n - \mathcal{C}^n\left(\Gamma + \mathcal{C}^n\right)^{-1}\mathcal{C}^n\right)$$

$$= \left(\Gamma^{-1} + (\mathcal{C}^n)^{-1}\right)\left(\mathcal{C}^n - \mathcal{C}^n\left[(\mathcal{C}^n)^{-1}\left(\Gamma^{-1} + (\mathcal{C}^n)^{-1}\right)^{-1}\Gamma^{-1}\right]\mathcal{C}^n\right) = I$$

leads to $(\mathcal{C}^{n+1})^{-1} = \Gamma^{-1} + (\mathcal{C}^n)^{-1}$ from Equation (8b), and hence $\mathcal{C}^{n+1}$ is invertible. Then Equation (9b) follows by induction. By applying $e^{t_{n+1}\mathcal{L}}$ to Equation (8a) we have

$$m_{n+1} = m_n - \mathcal{C}^n\left(\Gamma + \mathcal{C}^n\right)^{-1}\left(m_n - e^{t_{n+1}\mathcal{L}}y_{n+1}\right).$$

After using $\mathcal{C}^n(\Gamma + \mathcal{C}^n)^{-1} = \left((n+1)I + \Gamma\mathcal{C}_0^{-1}\right)^{-1}$ from Equation (9b), we obtain the telescoping series

$$\left((n+1)I + \Gamma\mathcal{C}_0^{-1}\right)m_{n+1} = \left(nI + \Gamma\mathcal{C}_0^{-1}\right)m_n + e^{t_{n+1}\mathcal{L}}y_{n+1}$$

and Equation (9a) follows. The final observations follow since $m_n = e^{t_n\mathcal{L}}m^n$ and $\mathcal{C}_n = e^{t_n\mathcal{L}}\mathcal{C}^n e^{t_n\mathcal{L}^*} = \mathcal{C}^n$. $\qquad\square$

2.2. Equivalence of measures

Now suppose we have a set of specific data and they are a single realization of the observation process (3),

$$y_n(\omega) = e^{-t_n \mathcal{L}} u + \eta_n(\omega). \tag{10}$$

Here ω is an element of the probability space Ω generating the entire noise signal $\{\eta_n\}_{n \in \mathbb{N}}$. We assume that the initial condition $u \in L^2(\mathbf{T}^2)$ for the true signal $e^{-t_n \mathcal{L}} u$ is non-random and hence independent of μ_0. We insert the fixed (non-random) instances of the data (10) into the formulae for μ_0, μ_n and μ^n, and will prove all three measures are equivalent. Recall that two measures are said to be equivalent if they are mutually absolutely continuous, and singular if they are concentrated on disjoint sets [29]. Lemma A.3 (Feldman-Hajek) tells us the conditions under which two Gaussian measures are equivalent.

Before stating the theorem, we need to introduce some notation and assumptions. Let $\phi_k(x) = e^{2\pi i k \cdot x}$, where $k = (k_1, k_2) \in \mathbb{Z} \times \mathbb{Z} \equiv \mathbb{K}$, be a standard orthonormal basis for $L^2(\mathbf{T}^2)$ with respect to the standard inner product $(f, g) = \int_{\mathbf{T}^2} f \bar{g} \, dx dy$ where the upper bar denotes the complex conjugate.

Assumptions 2.3. *The operators $\mathcal{L}, \mathcal{C}_0$ and Γ are diagonalizable in the basis defined by the ϕ_k. The two covariance operators, $\mathcal{C}_0$ and Γ, have positive eigenvalues $\lambda_k > 0$ and $\gamma_k > 0$ respectively:*

$$\mathcal{C}_0 \phi_k = \lambda_k \phi_k,$$
$$\Gamma \phi_k = \gamma_k \phi_k. \tag{11}$$

The eigenvalues ℓ_k of $\mathcal{L}$ have zero real parts and $\mathcal{L}\phi_0 = 0$.

This assumption on the simultaneous diagonalization of the operators implies the commutativity of $\mathcal{C}_0$ and Γ with $\mathcal{L}$. Therefore we have Corollary 2.2 and Equations (9) can be used to study the large n behaviour of the μ_n and μ^n. We believe that it might be possible to obtain the subsequent results without Assumption 2.3, but to do so would require significantly different techniques of analysis; the simultaneously diagonalizable case allows a straightforward analysis in which the mechanisms giving rise to the results are easily understood.

Note that, since $\mathcal{C}_0$ and Γ are the covariance operators of Gaussian measures on $L^2(\mathbf{T}^2)$, it follows from Lemma A.1 that the λ_k, γ_k are summable: i.e., $\sum_{k \in \mathbb{K}} \lambda_k < \infty$, $\sum_{k \in \mathbb{K}} \gamma_k < \infty$. We define $H^s(\mathbf{T}^2)$ to be the Sobolev space of periodic functions with s weak derivatives and

$$\|\cdot\|_{H^s(\mathbf{T}^2)} \equiv \sum_{k \in \mathbb{K}^+} |k|^{2s} |(\cdot, \phi_k)|^2 + |(\cdot, \phi_0)|^2$$

where $\mathbb{K}^+ = \mathbb{K} \backslash \{(0,0)\}$ noting that this norm reduces to the usual L^2 norm when $s = 0$. We denote by $\|\cdot\|$ the standard Euclidean norm.

Theorem 2.4. *If $\sum_{k \in \mathbb{K}} \lambda_k / \gamma_k^2 < \infty$, then the Gaussian measures μ_0, μ_n and μ^n on the Hilbert space $L^2(\mathbf{T}^2)$ are equivalent $\eta - a.s.$*

Proof. We first show the equivalence between μ_0 and μ_n. For $h = \sum_{k \in \mathbb{K}} h_k \phi_k$ we get, from Equations (9b) and (11),

$$\frac{1}{c^+} \le \frac{(h, \mathcal{C}_n h)}{(h, \mathcal{C}_0 h)} = \frac{\sum_{k \in \mathbb{K}} \left(n\gamma_k^{-1} + \lambda_k^{-1}\right)^{-1} h_k^2}{\sum_{k \in \mathbb{K}} \lambda_k h_k^2} \le 1,$$

where $c^+ = \sup_{k \in \mathbb{K}} (n\lambda_k/\gamma_k + 1)$. We have $c^+ \in [1, \infty)$ because $\sum_{k \in \mathbb{K}} \lambda_k/\gamma_k^2 < \infty$ and Γ is trace-class. Then the first condition for Feldman-Hajek is satisfied by Lemma A.4.

For the second condition, take $\{\mathbf{g}_k^{l+1}\}_{k \in \mathbb{K}}$ where $l = 0, \ldots, n-1$, to be a sequence of complex-valued unit Gaussians independent except for the condition $\mathbf{g}_{-k}^l = \bar{\mathbf{g}}_k^l$. This constraint ensures that the Karhunen-Loève expansion

$$e^{t_{l+1}\mathcal{L}}\eta_{l+1} = \sum_{k \in \mathbb{K}} \sqrt{\gamma_k}\mathbf{g}_k^{l+1}\phi_k,$$

is distributed according to the real-valued Gaussian measure $\mathcal{N}(0, \Gamma)$, and is independent for different values of l. Thus

$$
\begin{aligned}
\|m_n - m_0\|_{\mathcal{C}_0}^2 &\equiv \left\|\mathcal{C}_0^{-\frac{1}{2}}(m_n - m_0)\right\|_{L^2(\mathbf{T}^2)}^2 \\
&= \sum_{k \in \mathbb{K}} \left(\frac{\lambda_k^{-1/2}}{n + \gamma_k/\lambda_k}\right)^2 \left|-n(m_0, \phi_k) + n(u, \phi_k) + \sqrt{\gamma_k}\sum_{l=0}^{n-1}\mathbf{g}_k^{l+1}\right|^2 \\
&\le C(n) \sum_{k \in \mathbb{K}} \frac{\lambda_k}{\gamma_k^2} < \infty,
\end{aligned}
$$

where

$$C(n) \equiv \sup_{k \in \mathbb{K}} \left|-n(m_0, \phi_k) + n(u, \phi_k) + \sqrt{\gamma_k}\sum_{l=0}^{n-1}\mathbf{g}_k^{l+1}\right|^2 < \infty$$

η−a.s. from the strong law of large numbers [30].

For the third condition, we need to show the set of eigenvalues of T, where

$$T\phi_k = \left(\mathcal{C}_n^{-\frac{1}{2}}\mathcal{C}_0\mathcal{C}_n^{-\frac{1}{2}} - I\right)\phi_k = n\left(\frac{\lambda_k}{\gamma_k}\right)\phi_k,$$

are square-summable. This is satisfied since $\mathcal{C}_0$ is trace-class:

$$\sum_{k \in \mathbb{K}} \frac{\lambda_k^2}{\gamma_k^2} \le \left(\sup_{k \in \mathbb{K}} \lambda_k\right) \sum_{k \in \mathbb{K}} \frac{\lambda_k}{\gamma_k^2} < \infty.$$

The equivalence between μ_0 and μ^n is immediate because m^n is the image of m_n under a unitary map $e^{-t_n\mathcal{L}}$, and $\mathcal{C}^n = \mathcal{C}_n$. □

To illustrate the conditions of the theorem, let $(-\triangle)$ denote the negative Laplacian with domain $\mathcal{D}(-\triangle) = H^2(\mathbf{T}^2)$. Assume that $\mathcal{C}_0 \propto (-\triangle + k_A I)^{-A}$ and $\Gamma \propto (-\triangle + k_B I)^{-B}$, where the conditions $A > 1$, $B > 1$ and $k_A, k_B > 0$ ensure, respectively, that the two operators are trace-class and positive-definite. Then the condition $\sum_{k \in \mathbb{K}} \lambda_k/\gamma_k^2 < \infty$ reduces to $2B \le A - 1$.

3. Measure Consistency and Time-Independent Model Error

In this section we study the large data limit of the smoother μ_n and the filter μ^n in Corollary 2.2 for large n. We study measure consistency, namely the question of whether the filtering or smoothing distribution converges to a Dirac measure on the true signal as n increases. We study situations where the data is generated as a single realization of the statistical model itself, so there is no model error, and situations where model error is present.

3.1. Data model

Suppose that the true signal, denoted $v'_n = v'(\cdot, t_n)$, can be different from the solution v_n computed via the model (1) and instead solves

$$\partial_t v' + \mathcal{L}'v' = 0, \quad v'(0) = u \tag{12}$$

for another anti-Hermitian operator $\mathcal{L}'$ on $L^2(\mathbf{T}^2)$ and fixed $u \in L^2(\mathbf{T}^2)$. We further assume possible error in the observational noise model so that the actual noise in the data is not η_n but η'_n. Then, instead of y_n given by (10), what we actually incorporate into the filter is the true data, a single realization y'_n determined by v'_n and η'_n as follows:

$$y'_n = v'_n + \eta'_n = e^{-t_n \mathcal{L}'}u + \eta'_n. \tag{13}$$

We again use $e^{-t\mathcal{L}'}$ to denote the forward solution operator, now for (12), through t time units. Note that each realization (13) is an element in the probability space Ω', which is independent of Ω. For the data $Y'_n = \{y'_1, \ldots, y'_n\}$, let μ'_n be the measure $\mathbb{P}(v_0 | Y_n = Y'_n)$. This measure is Gaussian and is determined by the mean in Equation (9a) with y_l replaced by y'_l, which we relabel as m'_n, and the covariance operator in Equation (9b) which does not depend on the data so we retain the notation $\mathcal{C}_n$. Clearly, using (13) we obtain

$$m'_n = \left(nI + \Gamma \mathcal{C}_0^{-1}\right)^{-1} \left[\Gamma \mathcal{C}_0^{-1} m_0 + \sum_{l=0}^{n-1} \left(e^{t_{l+1}(\mathcal{L}-\mathcal{L}')}u + e^{t_{l+1}\mathcal{L}}\eta'_{l+1}\right)\right], \tag{14}$$

where $m'_0 = m_0$.

   This conditioned mean m'_n differs from m_n in that $e^{t_{l+1}(\mathcal{L}-\mathcal{L}')}u$ and η'_{l+1} appear instead of u and η_{l+1}, respectively. The reader will readily generalize both the statement and proof of Theorem 2.4 to show the equivalence of μ_0, μ'_n and $(\mu')^n \equiv \mathbb{P}(v_n | Y_n = Y'_n) = \mathcal{N}((m')^n, \mathcal{C}^n)$, showing the well-definedness of these conditional distributions even with errors in both forward model and observational noise. We now study the large data limit for the filtering problem, in the idealized scenario where $\mathcal{L} = \mathcal{L}'$ (Theorem 3.2) and the more realistic scenario with model error so that $\mathcal{L} \neq \mathcal{L}'$ (Theorem 3.7). We allow possible model error in the observation noise for both theorems, so that the noises η'_n are draws from i.i.d. Gaussians $\eta'_n \sim \mathcal{N}(0, \Gamma')$ and $\Gamma'\phi_k = \gamma'_k \phi_k$. Even though γ'_k (equivalently Γ') and $\mathcal{L}'$ are not exactly known in most practical situations, their asymptotics can be predicted within a certain degree of accuracy. Therefore, without

limiting the applicability of the theory, the following are assumed in all subsequent theorems and corollaries in this paper:

Assumptions 3.1. *There are positive real numbers* $s, \kappa \in \mathbb{R}^+$ *such that:*

(i) $\sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k < \infty$, $\sum_{k \in \mathbb{K}} |k|^{2s} \gamma'_k < \infty$;

(ii) $\gamma_k / \lambda_k = O\left(|k|^{\kappa}\right)$;

(iii) $m_0, u \in H^{s+\kappa}(\mathbf{T}^2)$.

Then Assumptions 3.1(1) imply that $\eta \sim \mathcal{N}(0, \Gamma)$ and $\eta' \sim \mathcal{N}(0, \Gamma')$ are in $H^s(\mathbf{T}^2)$ since $\mathbb{E}\|\eta\|^2_{H^s(\mathbf{T}^2)} < \infty$ and $\mathbb{E}\|\eta'\|^2_{H^s(\mathbf{T}^2)} < \infty$.

3.2. Limit of the conditioned measure without model error

We first study the large data limit of the measure μ'_n without model error.

Theorem 3.2. *For the statistical model (1) and (3), suppose that the data* $Y'_n = \{y'_1, \cdots, y'_n\}$ *is created from (13) with* $\mathcal{L} = \mathcal{L}'$. *Then, as* $n \to \infty$, $\mathbb{E}(v_0|Y'_n) = m'_n \to u$ *in the sense that*

$$\|m'_n - u\|_{L^2(\Omega'; H^s(\mathbf{T}^2))} = O\left(n^{-\frac{1}{2}}\right), \tag{15a}$$

$$\|m'_n - u\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s., \tag{15b}$$

for the probability space Ω' *generating the true observation noise* $\{\eta'_n\}_{n \in \mathbb{N}}$, *and for any non-negative* $\theta < 1/2$. *Furthermore,* $\mathcal{C}_n \to 0$ *in the sense that its operator norm from* $L^2(\mathbf{T}^2)$ *to* $H^s(\mathbf{T}^2)$ *satisfies*

$$\|\mathcal{C}_n\|_{\mathbf{L}(L^2(\mathbf{T}^2); H^s(\mathbf{T}^2))} = O(n^{-1}). \tag{16}$$

Proof. From Equation (14) with $\mathcal{L}' = \mathcal{L}$, we have

$$\left(nI + \Gamma \mathcal{C}_0^{-1}\right) m'_n = \Gamma \mathcal{C}_0^{-1} m_0 + nu + \sum_{l=0}^{n-1} e^{t_{l+1}\mathcal{L}} \eta'_{l+1},$$

thus

$$\left(nI + \Gamma \mathcal{C}_0^{-1}\right) (m'_n - u) = \Gamma \mathcal{C}_0^{-1}(m_0 - u) + \sum_{l=0}^{n-1} e^{t_{l+1}\mathcal{L}} \eta'_{l+1}.$$

Take $\{(\mathbf{g}'_k)^{l+1}\}_{k \in \mathbb{K}}$ where $l = 0, \ldots, n-1$, to be an i.i.d. sequence of complex-valued unit Gaussians subject to the constraint that $(\mathbf{g}'_{-k})^l = (\bar{\mathbf{g}}'_k)^l$. Then the Karhunen-Loève expansion for $e^{t_{l+1}\mathcal{L}} \eta'_{l+1} \sim \mathcal{N}(0, \Gamma')$ is given by

$$e^{t_{l+1}\mathcal{L}} \eta'_{l+1} = \sum_{k \in \mathbb{K}} \sqrt{\gamma'_k} (\mathbf{g}'_k)^{l+1} \phi_k. \tag{17}$$

It follows that

$$\|m'_n - u\|^2_{L^2(\Omega'; H^s(\mathbf{T}^2))} = \mathbb{E}\|m'_n - u\|^2_{H^s(\mathbf{T}^2)}$$

$$= \sum_{k \in \mathbb{K}^+} |k|^{2s} \mathbb{E}|(m'_n - u, \phi_k)|^2 + \mathbb{E}|(m'_n - u, \phi_0)|^2$$

$$= \sum_{k \in \mathbb{K}^+} \frac{|k|^{2s}}{(n + \gamma_k/\lambda_k)^2} \left[\left(\frac{\gamma_k}{\lambda_k} \right)^2 |(m_0 - u, \phi_k)|^2 + \gamma_k' n \right]$$

$$+ \frac{1}{(n + \gamma_0/\lambda_0)^2} \left[\left(\frac{\gamma_0}{\lambda_0} \right)^2 |(m_0 - u, \phi_0)|^2 + \gamma_0' n \right]$$

$$\leq \sum_{k \in \mathbb{K}^+} |k|^{2s} n^{-2} \left[\left(\frac{\gamma_k}{\lambda_k} \right)^2 |(m_0 - u, \phi_k)|^2 + \gamma_k' n \right]$$

$$+ n^{-2} \left[\left(\frac{\gamma_0}{\lambda_0} \right)^2 |(m_0 - u, \phi_0)|^2 + \gamma_0' n \right]$$

$$\leq \left(C \|m_0 - u\|_{H^{s+\kappa}(\mathbf{T}^2)} \right) n^{-2} + \left(\sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k' \right) n^{-1},$$

and so we have Equation (15a). Here and throughout the paper, C is a constant that may change from line to line.

Equation (15b) follows from the Borel-Cantelli Lemma [30]: for arbitrary $\epsilon > 0$, we have

$$\sum_{n \in \mathbb{N}} \mathbb{P} \left(n^\theta \|m_n' - u\|_{H^s(\mathbf{T}^2)} > \epsilon \right) = \sum_{n \in \mathbb{N}} \mathbb{P} \left(n^{2r\theta} \|m_n' - u\|_{H^s(\mathbf{T}^2)}^{2r} > \epsilon^{2r} \right)$$

$$\leq \sum_{n \in \mathbb{N}} \frac{n^{2r\theta}}{\epsilon^{2r}} \mathbb{E} \|m_n' - u\|_{H^s(\mathbf{T}^2)}^{2r} \qquad \text{(Markov inequality)}$$

$$\leq \sum_{n \in \mathbb{N}} C n^{2r\theta} \left(\mathbb{E} \|m_n' - u\|_{H^s(\mathbf{T}^2)}^2 \right)^r \qquad \text{(Lemma A.5)}$$

$$\leq \sum_{n \in \mathbb{N}} \frac{C}{n^{r(1-2\theta)}} < \infty \qquad \text{(by (15a))}$$

and if $\theta \in (0, 1/2)$ then we can choose r such that $r(1 - 2\theta) > 1$.

Finally, for $h = \sum_{k \in \mathbb{K}} h_k \phi_k$,

$$\|\mathcal{C}_n\|_{\mathbf{L}(L^2(\mathbf{T}^2); H^s(\mathbf{T}^2))}^2 = \sup_{\|h\|_{L^2(\mathbf{T}^2)} \leq 1} \|\mathcal{C}_n h\|_{H^s(\mathbf{T}^2)}^2$$

$$= \sup_{\|h\|_{L^2(\mathbf{T}^2)} \leq 1} \sum_{k \in \mathbb{K}} |k|^{2s} \left| n\gamma_k^{-1} + \lambda_k^{-1} \right|^{-2} |h_k|^2$$

$$\leq C \sum_{k \in \mathbb{K}} |k|^{2s} \left| n\gamma_k^{-1} + \lambda_k^{-1} \right|^{-2}$$

$$\leq \frac{C}{n^2} \sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k^2.$$

and use the fact that $\sup_{k \in \mathbb{K}} \gamma_k < \infty$, since Γ is trace-class, together with Assumptions 3.1(1) to get the desired convergence rate. $\qquad \square$

Corollary 3.3. *Under the same assumptions as Theorem 3.2, if Equation (15b) holds with $s > 1$, then*

$$\|m_n' - u\|_{L^\infty(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \qquad \Omega' - a.s., \tag{18}$$

for any non-negative $\theta < 1/2$.

Proof. Equation (18) is immediate from Equation (15b) and the Sobolev embedding,

$$\|\cdot\|_{L^\infty(\mathbf{T}^2)} \leq C \|\cdot\|_{H^s(\mathbf{T}^2)},$$

when $s > d/2 = 1$ since $d = 2$ is the dimension of the domain. $\square$

Corollary 3.4. *Under the same assumptions as Theorem 3.2, as $n \to \infty$, the $\Omega' - a.s.$ weak convergence $\mu'_n \Rightarrow \delta_u$ holds in $L^2(\mathbf{T}^2)$.*

Proof. To prove this result we apply Example 3.8.15 in [31]. This shows that for weak convergence of $\mu'_n = \mathcal{N}(m'_n, \mathcal{C}_n)$ a Gaussian measure on $\mathcal{H}$ to a limit measure $\mu = \mathcal{N}(m, \mathcal{C})$ on $\mathcal{H}$, it suffices to show that $m'_n \to m$ in $\mathcal{H}$, that $\mathcal{C}_n \to \mathcal{C}$ in $\mathbf{L}(\mathcal{H}, \mathcal{H})$ and that second moments converge. Note, also, that Dirac measures, and more generally semi-definite covariance operators, are included in the definition of Gaussian. The convergence of the means and the covariance operators follows from Equations (15b) and (16) with $m = u$, $\mathcal{C} = 0$ and $\mathcal{H} = L^2(\mathbf{T}^2)$. The convergence of the second comments follows if the trace of $\mathcal{C}_n$ converges to zero. From (9b) it follows that the trace of $\mathcal{C}_n$ is bounded by n^{-1} multiplied by the trace of Γ. But Γ is trace-class as it is a covariance operator on $L^2(\mathbf{T}^2)$ and so the desired result follows. $\square$

In fact, the weak convergence in the Prokhorov metric between μ'_n and δ_u holds, and the methodology in [24] could be used to quantify its rate of convergence.

We now obtain the large data limit of the filtering distribution $(\mu')^n$ without model error from the smoothing limit (15). Recall that this measure is the Gaussian $\mathcal{N}((m')^n, \mathcal{C}_n)$.

Theorem 3.5. *Under the same assumptions as Theorem 3.2, as $n \to \infty$, $(m')^n - v'_n \to 0$ in the sense that*

$$\|(m')^n - v'_n\|_{L^2(\Omega';H^s(\mathbf{T}^2))} = O\left(n^{-\frac{1}{2}}\right), \tag{19a}$$

$$\|(m')^n - v'_n\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s., \tag{19b}$$

for any non-negative $\theta < 1/2$.

Proof. Equation (19b) follows from

$$
\begin{aligned}
\|(m')^n - v'_n\|_{H^s(\mathbf{T}^2)} &= \left\|e^{-t_n\mathcal{L}}m'_n - e^{-t_n\mathcal{L}'}u\right\|_{H^s(\mathbf{T}^2)} \\
&= \left\|e^{-t_n\mathcal{L}}(m'_n - u)\right\|_{H^s(\mathbf{T}^2)} \\
&\leq \left\|e^{-t_n\mathcal{L}}\right\|_{\mathbf{L}(H^s(\mathbf{T}^2);H^s(\mathbf{T}^2))} \|m'_n - u\|_{H^s(\mathbf{T}^2)} \\
&= \|m'_n - u\|_{H^s(\mathbf{T}^2)}.
\end{aligned}
$$

Then Equation (19a) follows from

$$
\begin{aligned}
\|(m')^n - v'_n\|^2_{L^2(\Omega';H^s(\mathbf{T}^2))} &= \mathbb{E}\|(m')^n - v'_n\|^2_{H^s(\mathbf{T}^2)} \\
&\leq \mathbb{E}\|m'_n - u\|^2_{H^s(\mathbf{T}^2)}.
\end{aligned}
$$

$\square$

The following corollary has the same as for Corollary 3.4, and so we omit it.

Corollary 3.6. *Under the same assumptions as Theorem 3.2, as $n \to \infty$, the $\Omega' - a.s.$ weak convergence $(\mu')^n - \delta_{v'_n} \Rightarrow 0$ holds in $L^2(\mathbf{T}^2)$.*

Theorem 3.5 and Corollary 3.6 are consistent with known results concerning large data limits of the finite dimensional Kalman filter shown in [32]. However, Equations (19) and (16) provide convergence rates in an infinite dimensional space and hence cannot be derived from the finite dimensional theory; mathematically this is because the rate of convergence in each Fourier mode will depend on the wavenumber k, and the infinite dimensional analysis requires this dependence to be tracked and quantified, as we do here.

3.3. Limit of the conditioned measure with time-independent model error

The previous subsection shows measure consistency results for data generated by the same PDE as that used in the filtering model. In this section, we study the consequences of using data generated by a different PDE. It is important to point out that, in view of Equation (14), the limit of m'_n is determined by the time average of $e^{t_{l+1}(\mathcal{L} - \mathcal{L}')}u$, i.e.,

$$\frac{1}{n} \sum_{l=0}^{n-1} e^{t_{l+1}(\mathcal{L} - \mathcal{L}')}u. \tag{20}$$

For general anti-Hermitian $\mathcal{L}$ and $\mathcal{L}'$, obtaining an analytic expression for the limit of the average (20), as $n \to \infty$, is very hard. Therefore in the remainder of the section we examine the case in which $\mathcal{L} = c \cdot \nabla$ and $\mathcal{L}' = c' \cdot \nabla$ with different constant wave velocities c and c', respectively. A highly nontrivial filter divergence takes place even in this simple example.

We use the notation $\mathcal{F}_{(p,q)}f \equiv \sum_{(k_1/p, k_2/q) \in \mathbb{Z} \times \mathbb{Z}}(f, \phi_k)\phi_k$ for part of the Fourier series of $f \in L^2(\mathbf{T}^2)$, and $\langle f \rangle \equiv (f, \phi_0) = \int_{\mathbf{T}^2} f(x, y) \, dx dy$ for the spatial average of f on the torus. We also denote by $\delta c \equiv c - c'$ the difference between wave velocities.

Theorem 3.7. *For the statistical model (1) and (3) with $\mathcal{L} = c \cdot \nabla$, suppose that the data $Y'_n = \{y'_1, \cdots, y'_n\}$ is created from (13) with $\mathcal{L}' = c' \cdot \nabla$ and that $\delta c \neq 0 \bmod(1, 1)$ (equivalently $\delta c \notin \mathbb{Z} \times \mathbb{Z}$). As $n \to \infty$,*

(i) if $\Delta t \, \delta c = (p'/p, q'/q) \in \mathbb{Q} \times \mathbb{Q}$ and $\gcd(p', p) = \gcd(q', q) = 1$, then $m'_n \to \mathcal{F}_{(p,q)}u$ in the sense that

$$\left\| m'_n - \mathcal{F}_{(p,q)}u \right\|_{L^2(\Omega'; H^s(\mathbf{T}^2))} = O\left(n^{-\frac{1}{2}}\right), \tag{21a}$$

$$\left\| m'_n - \mathcal{F}_{(p,q)}u \right\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s., \tag{21b}$$

for any non-negative $\theta < 1/2$;

(ii) if $\Delta t \, \delta c \in \mathbb{R} \backslash \mathbb{Q} \times \mathbb{R} \backslash \mathbb{Q}$, then $m'_n \to \langle u \rangle$ in the sense that

$$\left\| m'_n - \langle u \rangle \right\|_{L^2(\Omega'; H^s(\mathbf{T}^2))} = o\left(1\right), \tag{22a}$$

$$\left\| m'_n - \langle u \rangle \right\|_{H^s(\mathbf{T}^2)} = o\left(1\right) \quad \Omega' - a.s. \tag{22b}$$

Proof. See Appendix B. □

Remark 3.8. It is interesting that it is not the *size* of $\Delta t \, \delta c$ but its *rationality or irrationality* which determines the limit of m'_n. This result may be understood intuitively from Equation (20), which reduces to

$$\frac{1}{n} \sum_{l=0}^{n-1} u(\cdot + (l+1)\Delta t \, \delta c), \tag{23}$$

when $\mathcal{L} = c \cdot \nabla$ and $\mathcal{L}' = c' \cdot \nabla$. It is possible to guess the large n behaviour of Equation (23) using the periodicity of u and an ergodicity argument. The proof of Theorem 3.7 tells us that the prediction resulting from this heuristic is indeed correct.

The proof of Corollary 3.3 tells us that, whenever $s > 1$, the $H^s(\mathbf{T}^2)-$norm convergence in (21b) or (22b) implies the almost everywhere convergence on $\mathbf{T}^2$ with the same order. Therefore, $m'_n \to \mathcal{F}_{(p,q)}u$ or $m'_n \to \langle u \rangle$ a.e. on $\mathbf{T}^2$ from Equation (21b) or from Equation (22b), when $\Delta t \, \delta c = (p'/p, q'/q) \in \mathbb{Q} \times \mathbb{Q}$ and $\gcd(p', p) = \gcd(q', q) = 1$, or when $\Delta t \, \delta c \in \mathbb{R} \backslash \mathbb{Q} \times \mathbb{R} \backslash \mathbb{Q}$, respectively. We will not repeat the statement of the corresponding result in the subsequent theorems.

When $\Delta t \, \delta c \in \mathbb{R} \backslash \mathbb{Q} \times \mathbb{R} \backslash \mathbb{Q}$, Equation (22) is obtained using

$$\frac{1}{n} \sum_{l=0}^{n-1} e^{2\pi i (k \cdot \delta c) t_{l+1}} = o(1), \tag{24}$$

as $n \to \infty$, from the theory of ergodicity [33]. The convergence rate of Equation (22) can be improved if we have higher order convergence in Equation (24). It must be noted that in general there exists a fundamental relationship between the limits of m'_n and the left-hand side of Equation (24) for various c and c', as we will see.

Note also from Theorem 3.2 and Theorem 3.7, the limit of m'_n does not depend on the observation noise error $\Gamma \neq \Gamma'$ but does depend sensitively on the model error $\mathcal{L} \neq \mathcal{L}'$. □

Corollary 3.9. *Under the same assumptions as Theorem 3.7, as $n \to \infty$, the $\Omega' - a.s.$ weak convergence $\mu'_n \Rightarrow \delta_{\mathcal{F}_{(p,q)}u}$ or $\mu'_n \Rightarrow \delta_{\langle u \rangle}$ holds, when $\Delta t \, \delta c = (p'/p, q'/q) \in \mathbb{Q} \times \mathbb{Q}$ and $\gcd(p', p) = \gcd(q', q) = 1$, or when $\Delta t \, \delta c \in \mathbb{R} \backslash \mathbb{Q} \times \mathbb{R} \backslash \mathbb{Q}$, respectively.*

Proof. This is the same as the proof of Corollary 3.4, so we omit it. □

Remark 3.10. Our theorems show that the observation accumulation limit and the vanishing model error limit cannot be switched, i.e.,

$$\lim_{n \to \infty} \lim_{||\delta c|| \to 0} ||m'_n - u||_{H^s(\mathbf{T}^2)} = 0,$$

$$\lim_{||\delta c|| \to 0} \lim_{n \to \infty} ||m'_n - u||_{H^s(\mathbf{T}^2)} \neq 0.$$

Note the second limit is nonzero because m'_n converges either to $\mathcal{F}_{(p,q)}u$ or $\langle u \rangle$. □

We can also study the effect of model error on the filtering distribution, instead of the smoothing distribution. The following theorem extends Theorem 3.7 to study the

measure $(\mu')^n$, showing that the truth v'_n is not recovered if $\delta c \neq 0$. This result should be compared with Theorem 3.5 in the case of no model error.

Theorem 3.11. *Under the same assumptions as Theorem 3.7, as $n \to \infty$,*

(i) *if $\Delta t \, \delta c = (p'/p, q'/q) \in \mathbb{Q} \times \mathbb{Q}$ and $\gcd(p', p) = \gcd(q', q) = 1$, then $(m')^n - \mathcal{F}_{(p,q)} e^{-t_n \mathcal{L}} u \to 0$ in the sense that*

$$\left\| (m')^n - \mathcal{F}_{(p,q)} e^{-t_n \mathcal{L}} u \right\|_{L^2(\Omega'; H^s(\mathbf{T}^2))} = O\left(n^{-\frac{1}{2}} \right), \tag{25a}$$

$$\left\| (m')^n - \mathcal{F}_{(p,q)} e^{-t_n \mathcal{L}} u \right\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta} \right) \quad \Omega' - a.s., \tag{25b}$$

for any non-negative $\theta < 1/2$;

(ii) *if $\Delta t \, \delta c \in \mathbb{R} \backslash \mathbb{Q} \times \mathbb{R} \backslash \mathbb{Q}$, then $(m')^n \to \langle u \rangle$ in the sense that*

$$\left\| (m')^n - \langle u \rangle \right\|_{L^2(\Omega'; H^s(\mathbf{T}^2))} = o\left(1 \right), \tag{26a}$$

$$\left\| (m')^n - \langle u \rangle \right\|_{H^s(\mathbf{T}^2)} = o\left(1 \right) \quad \Omega' - a.s. \tag{26b}$$

Proof. This is the same as the proof of Theorem 3.5 except $\mathcal{F}_{(p,q)} e^{-t_n \mathcal{L}} u$ or $\langle u \rangle$ is used in place of v'_n, so we omit it. $\qquad \square$

Corollary 3.12. *Under the same assumptions as Theorem 3.7, as $n \to \infty$, the $\Omega' - a.s.$ weak convergence $(\mu')^n - \delta_{\mathcal{F}_{(p,q)} e^{-t_n \mathcal{L}} u} \Rightarrow 0$ or $(\mu')^n - \delta_{\langle u \rangle} \Rightarrow 0$ holds, when $\Delta t \, \delta c = (p'/p, q'/q) \in \mathbb{Q} \times \mathbb{Q}$ and $\gcd(p', p) = \gcd(q', q) = 1$, or when $\Delta t \, \delta c \in \mathbb{R} \backslash \mathbb{Q} \times \mathbb{R} \backslash \mathbb{Q}$, respectively.*

Proof. This is the same as the proof of Corollary 3.4, so we omit it. $\qquad \square$

Theorem 3.2 and Theorem 3.5 show that, in the perfect model scenario, the smoothing distribution on the initial condition and filtering distribution recover the true initial condition and true signal, respectively, even if the statistical model fails to capture the genuine covariance structure of the data. Theorem 3.7 and Theorem 3.11 show that the smoothing distribution on the initial condition, and the filtering distribution, do not converge to the truth, in the large data limit, when the model error corresponds to a constant shift in wave velocity, however small. In this case the wave velocity difference causes an advection in Equation (20) leading to recovery of only part of the Fourier expansion of u as a limit of m'_n. The next section concerns time-dependent model error in the wave velocity, and includes a situation intermediate between those considered in this section. In particular, a situation where the smoothing distribution is not recovered correctly, but the filtering distribution is.

4. Time-Dependent Model Error

In the previous section we studied model error for autonomous problems where the operators $\mathcal{L}$ and $\mathcal{L}'$ (and hence c and c') are assumed time-independent. However, our approach can be generalized to situations where both operators are time-dependent: $\mathcal{L}(t) = c(t) \cdot \nabla$ and $\mathcal{L}'(t) = c'(t) \cdot \nabla$. To this end, this section is devoted two problems

where both the statistical model and the data are generated by the non-autonomous dynamics. In the first case deterministic dynamics with $c(t) - c'(t) \to 0$ (Theorem 4.1); and in the second case where the data is generated by the non-autonomous random dynamics with $\mathcal{L}'(t; \omega') = c'(t; \omega') \cdot \nabla$ and $c'(t; \omega')$ being a random function fluctuating around c (Theorem 4.3). Here ω' denotes an element in the probability space that generates $c'(t; \omega')$ and η'_n, assumed independent.

Now the statistical model (1) becomes

$$\partial_t v + c(t) \cdot \nabla v = 0, \quad v(x, 0) = v_0(x) \tag{27}$$

Unless $c(t)$ is constant in time, the operator $e^{-t\mathcal{L}}$ is not a semigroup operator but for convenience we still employ this notation to represent the forward solution operator from time 0 to time t even for non-autonomous dynamics. Then the solution of Equation (27) is denoted by

$$v(x, t) = \left(e^{-t\mathcal{L}} v_0 \right)(x) \equiv v_0 \left(x - \int_0^t c(s) \, ds \right).$$

This will correspond to a classical solution if $v_0 \in C^1\left(\mathbf{T}^2, \mathbb{R}\right)$ and $c \in C\left(\mathbb{R}^+, \mathbb{R}^2\right)$; otherwise it will be a weak solution. Similarly, the notation $e^{-t\mathcal{L}'}$ will be used for the forward solution operator from time 0 to time t given the non-autonomous deterministic or random dynamics, i.e., Equation (12) becomes

$$\partial_t v' + c'(t; \omega') \cdot \nabla v' = 0, \quad v'(0) = u \tag{28}$$

and we define the solution of Equation (28) by

$$v'(x, t; \omega') = \left(e^{-t\mathcal{L}'} u \right)(x) \equiv u \left(x - \int_0^t c'(s; \omega') \, ds \right),$$

under the assumption that $\int_0^t c'(s; \omega') \, ds$ is well-defined. We will be particularly interested in the case where $c'(t; \omega')$ is an affine function of a Brownian white noise and then this expression corresponds to the Stratonovich solution of the PDE (28) [34]. Note the term $e^{t_{l+1}(\mathcal{L} - \mathcal{L}')} u$ in Equation (14) should be interpreted as

$$\left(e^{t_{l+1}(\mathcal{L} - \mathcal{L}')} u \right)(x) \equiv u \left(x + \int_0^t \left(c(s) - c'(s; \omega') \right) ds \right).$$

We now study the case where both $c(t)$ and $c'(t)$ are deterministic time-dependent wave velocities. We here exhibit an intermediate situation between the two previously examined cases where the smoothing distribution is not recovered correctly, but the filtering distribution is. This occurs when the wave velocity is time-dependent but converges in time to the true wave velocity, i.e., $\delta c(t) \equiv c(t) - c'(t) \to 0$, which is of interest especially in view of Remark 3.10. Let $u_\alpha \equiv u(\cdot + \alpha)$ be the translation of u by α.

Theorem 4.1. *For the statistical model (1) and (3), suppose that the data $Y'_n = \{y'_1, \cdots, y'_n\}$ is created from (13) with $\mathcal{L}(t) = c(t) \cdot \nabla$ and $\mathcal{L}'(t) = c'(t) \cdot \nabla$ where*

$\delta c(t) = c(t) - c'(t)$ *satisfies* $\int_0^t \delta c(s)\, ds = \alpha + O\left(t^{-\beta}\right)$. *Then, as* $n \to \infty$, $m'_n \to u_\alpha$ *in the sense that*

$$\|m'_n - u_\alpha\|_{L^2(\Omega';H^s(\mathbf{T}^2))} = O\left(n^{-\phi}\right), \tag{29a}$$

$$\|m'_n - u_\alpha\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s., \tag{29b}$$

for $\phi = 1/2 \wedge \beta$ *and for any non-negative* $\theta < \phi$.

Proof. See Appendix B. □

Theorem 4.2. *Under the same assumptions as Theorem 4.1, if u is Lipschitz continuous in $H^s(\mathbf{T}^2)$ where s is given in Assumptions 3.1, then $(m')^n - v'_n \to 0$ in the sense that*

$$\|(m')^n - v'_n\|_{L^2(\Omega';H^s(\mathbf{T}^2))} = O\left(n^{-\phi}\right), \tag{30a}$$

$$\|(m')^n - v'_n\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s., \tag{30b}$$

for $\phi = 1/2 \wedge \beta$ *and for any non-negative* $\theta < \phi$.

Proof. Equation (30b) follows from

$$
\begin{aligned}
\|(m')^n - v'_n\|_{H^s(\mathbf{T}^2)} &= \left\| e^{-t_n \mathcal{L}}(m'_n - u_\alpha) + \left(e^{-t_n \mathcal{L}} u_\alpha - e^{-t_n \mathcal{L}'} u \right) \right\|_{H^s(\mathbf{T}^2)} \\
&\leq \left\| e^{-t_n \mathcal{L}} \right\|_{\mathbf{L}(H^s(\mathbf{T}^2);H^s(\mathbf{T}^2))} \|m'_n - u_\alpha\|_{H^s(\mathbf{T}^2)} \\
&\quad + \left\| u\left(\cdot + \alpha - \int_0^{t_n} c(s)\, ds \right) - u\left(\cdot - \int_0^{t_n} c'(s)\, ds \right) \right\|_{H^s(\mathbf{T}^2)} \\
&\leq \|m'_n - u_\alpha\|_{H^s(\mathbf{T}^2)} + C \left\| \alpha - \int_0^{t_n} \delta c(s)\, ds \right\|,
\end{aligned}
$$

and Equation (30a) follows from

$$
\begin{aligned}
\|(m')^n - v'_n\|^2_{L^2(\Omega';H^s(\mathbf{T}^2))} &= \mathbb{E} \|(m')^n - v'_n\|^2_{H^s(\mathbf{T}^2)} \\
&= \mathbb{E} \left\| e^{-t_n \mathcal{L}}(m'_n - u_\alpha) + \left(e^{-t_n \mathcal{L}} u_\alpha - e^{-t_n \mathcal{L}'} u \right) \right\|^2_{H^s(\mathbf{T}^2)} \\
&\leq 2 \left\| e^{-t_n \mathcal{L}} \right\|^2_{\mathbf{L}(H^s(\mathbf{T}^2);H^s(\mathbf{T}^2))} \mathbb{E} \|m'_n - u_\alpha\|^2_{H^s(\mathbf{T}^2)} \\
&\quad + 2 \left\| u\left(\cdot + \alpha - \int_0^{t_n} c(s)\, ds \right) - u\left(\cdot - \int_0^{t_n} c'(s)\, ds \right) \right\|^2_{H^s(\mathbf{T}^2)}.
\end{aligned}
$$

□

Finally, we study the case where $c(t)$ is deterministic but $c'(t;\omega')$ is a random process. We here note that while the true signal solves a linear SPDE with multiplicative noise, Equation (28), the statistical model used to filter is a linear deterministic PDE, Equation (27). We study the specific case $c'(t;\omega') = c(t) - \varepsilon \dot{W}(t)$, i.e., the deterministic wave velocity is modulated by a white noise with small amplitude $\varepsilon > 0$.

Theorem 4.3. *For the statistical model (1) and (3), suppose that the data $Y'_n = \{y'_1, \cdots, y'_n\}$ is created from (13) with $\mathcal{L}(t) = c(t) \cdot \nabla$ and $\mathcal{L}'(t; \omega') = c'(t; \omega') \cdot \nabla$ where $\int_0^t c'(s; \omega') \, ds = \int_0^t c(t) \, ds - \varepsilon W(t)$ and $\varepsilon W(t)$ is the Wiener process with amplitude $\varepsilon > 0$. Then, as $n \to \infty$, $m'_n \to \langle u \rangle$ in the sense that*

$$\|m'_n - \langle u \rangle\|_{L^2(\Omega'; H^s(\mathbf{T}^2))} = O\left(n^{-\frac{1}{2}}\right), \tag{31a}$$

$$\|m'_n - \langle u \rangle\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s., \tag{31b}$$

for any non-negative $\theta < 1/2$.

Proof. See Appendix B. $\qquad\qquad\square$

We do not state the corresponding theorem on the mean of the filtering distribution as $(m')^n$ converges to the same constant $\langle u \rangle$ with the same rate shown in Equations (31).

Remark 4.4. In Theorem 4.3 the law of $\varepsilon W(t)$ converges weakly to a uniform distribution on the torus by the Lévy-Cramér continuity theorem [30]. The limit $\langle u \rangle$ is the average of u with respect to this measure. This result may be generalized to consider the case where $\int_0^t (c(s) - c'(s; \omega')) \, ds$ converges weakly to the measure ν as $t \to \infty$. Then $m'_n(\cdot) \to \int u(\cdot + y) \, d\nu(y)$ as $n \to \infty$ in the same norms as used in Theorem 4.3 but, in general, with no rate. The key fact underlying the result is

$$\frac{1}{n} \sum_{l=0}^{n-1} e^{2\pi i k \cdot \int_0^{t_{l+1}} (c(s) - c'(s; \omega')) \, ds} - \int u(\cdot + y) \, d\nu(y) = o(1) \quad \Omega' - a.s., \tag{32}$$

as $n \to \infty$, which follows from the strong law of large numbers and the Kolmogorov zero-one law [30]. Depending on the process $\int_0^t (c(s) - c'(s; \omega')) \, ds$, an algebraic convergence rate can be determined if we have higher order convergence result for the corresponding Equation (32). This viewpoint may be used to provide a common framework for all the limit theorems in this paper.

5. Numerical Illustrations

5.1. Algorithmic Setting

The purpose of this section is twofold: first to illustrate the preceding theorems with numerical experiments; and secondly to show that relaxing the statistical model can avoid some of the lack of consistency problems that the theorems highlight. All of the numerical results we describe are based on using the Equations (1), (3), with $\mathcal{L} = c \cdot \nabla$ for some constant wave velocity c, so that the underlying dynamics is given by Equation (2). The data is generated by (12), (13) with $\mathcal{L}' = c'(t) \cdot \nabla$, for a variety of choices of $c'(t)$ (possibly random), and in subsection 5.2 we illustrate Theorems 3.2, 3.7, 4.1 and 4.3. In subsection 5.3 we will also describe a numerical method in which the state of the system *and* the wave velocity are learnt by combining the data and statistical model. Since this problem is inherently non-Gaussian we adopt from the outset a Bayesian approach which coincides with the (Gaussian) filtering or smoothing approach when the

wave velocity is fixed, but is sufficiently general to also allow for the wave velocity to be part of the unknown state of the system. In both cases we apply functionspace MCMC methods [35] to sample the distribution of interest. Note, however, that the purpose of this section is *not* to determine the most efficient numerical methods, but rather to study the properties of the statistical distributions of interest.

For fixed wave velocity c the statistical model (1), (3) defines a probability distribution $\mathbb{P}(v_0, Y_n | c)$. This is a Gaussian distribution and the conditional distribution $\mathbb{P}(v_0 | Y_n, c)$ is given by the measure $\mu_n = \mathcal{N}(m_n, \mathcal{C}_n)$ studied in sections 2, 3 and 4. In our first set of numerical results, in subsection 5.2, the wave velocity is considered known. We sample $\mathbb{P}(v_0 | Y_n, c)$ using the functionspace random walk from [36]. In the second set of results, in subsection 5.3, the wave velocity is considered as an unknown constant. If we place a prior measure $\rho(c)$ on the wave velocity then we may define $\mathbb{P}(c, v_0, Y_n) = \mathbb{P}(v_0, Y_n | c)\rho(c)$. We are then interested in the conditional distribution $\mathbb{P}(c, v_0 | Y_n)$ which is non-Gaussian. We adopt a Metropolis-within-Gibbs approach [37, 38] in which we sample alternately from $\mathbb{P}(v_0 | c, Y_n)$, which we do as in subsection 5.2, and $\mathbb{P}(c | v_0, Y_n)$, which we sample using a finite dimensional Metropolis-Hastings method.

Throughout the numerical simulations we represent the solution of the wave equation on a grid of $2^5 \times 2^5$ points, and observations are also taken on this grid. The observational noise is white (uncorrelated) with variance $\sigma^2 = 10^{-4}$ at each grid point. The continuum limit of such a covariance operator does not satisfy Assumptions 3.1, but is used to illustrate the fact that the theoretical results can be generalized to such observations. Note also that the numerical results are performed with model error so that the aforementioned distributions are sampled with $Y_n = Y_n'$ from (12), (13).

5.2. Sampling the initial condition with model error

Throughout we use the wave velocity

$$c = (-0.5, -1.0), \tag{33}$$

in our statistical model. The true initial condition used to generate the data is

$$u(x_1, x_2) = \sum_{k_1, k_2 = 1}^{3} \sin(2\pi k_1 x_1) + \cos(2\pi k_2 x_2). \tag{34}$$

This function is displayed in Figure 1(a). As prior on v_0 we choose the Gaussian $\mathcal{N}(0, (-\triangle)^{-2})$ where the domain of $-\triangle$ is $H^2(\mathbf{T}^2)$ with constants removed, so that it is positive. We implement the MCMC method to sample from $\mathbb{P}(v_0 | c, Y_n = Y_n')$ for a number of different data Y_n', corresponding to different choices of $c' = c'(t, \omega')$. We calculate the empirical mean of $\mathbb{P}(v_0 | c, Y_n = Y_n')$, which approximates $\mathbb{E}(v_0 | c, Y_n = Y_n')$. The results are shown in Figures 1(b)–1(f). In all cases the Markov chain is burnt in for 10^6 iterations, and this transient part of the simulation is not used to compute draws from the conditioned measure $\mathbb{P}(v_0 | c, Y_n = Y_n')$. After the burn in we proceed to iterate a further 10^7 times and use this information to compute the corresponding moments.

The data size n is chosen sufficiently large that this distribution is approximately a Dirac measure.

In the perfect model scenario $(c = c')$, the empirical mean shown in Figure 1(b), should fully recover the true initial condition u from Theorem 3.2. Comparison with Figure 1(a) shows that this is indeed the case, illustrating Corollary 3.3. We now demonstrate the effect of model error in the form of a constant shift in the wave velocity: Figure 1(c) and Figure 1(d) show the empirical means when $c - c' = (1/2, 1/2) \in \mathbb{Q} \times \mathbb{Q}$ and $c - c' = (1/e, 1/\pi) \in \mathbb{R}\backslash\mathbb{Q} \times \mathbb{R}\backslash\mathbb{Q}$, respectively. From Theorem 3.7, the computed empirical distribution should be close to, respectively, $\mathcal{F}_{(2,2)}u$ comprising only the mode $(k_1, k_2) = (2, 2)$ from (34), or $\langle u \rangle = 0$; this is indeed the case.

If we choose $c'(t)$ satisfying $\int_0^\infty (c - c'(s)) \, ds = (1/2, 1/2)$, then Theorem 4.1 tells us that Figure 1(e) should be close to a shift of u by $(1/2, 1/2)$, and this is exactly what we observe. In this case, we know from Theorem 4.2 that although the smoother is in error, the filter should correctly recover the true v_n' for large n. To illustrate this we compute $\|\mathbb{E}(v_n|c, Y_n = Y_n') - v_n'\|_{L^2(\mathbf{T}^2)}$ as a function of n and depict it in Figure 2(a). This shows convergence to 0 as predicted. To obtain a rate of convergence, we compute the gradient of a log-log plot of Figure 2(b). We observe the rate of convergence is close to $O(n^{-2})$. Note that this is higher than the theoretical bound of $O(n^{-\phi})$, with $\phi = 1/2 \wedge \beta$, given in Equation (30b); this suggests that our convergence theorems do not have sharp rates.

Finally, we examine the random $c'(t, \omega')$ cases. Figure 1(f) shows the empirical mean when $c'(t; \omega')$ is chosen such that

$$\int_0^t (c - c'(s; \omega')) \, ds = W(t)$$

where $W(t)$ is a standard Brownian motion. Theorem 4.3 tells us that the computed empirical distribution should have mean close to $\langle u \rangle$, and this is again the case.

5.3. Sampling the wave velocity and initial condition

The objective of this subsection is to show that the problems caused by model error in the form of a constant shift to the wave velocity can be overcome by sampling c and v_0. We generate data from (12), (13) with $c' = c$ given by (33) and initial condition (34). We assume that neither the wave velocity nor the initial condition are known to us, and we attempt to recover them from given data.

The desired conditional distribution is multimodal with respect to c – recall that it is non-Gaussian – and care is required to seed the chain close to the desired value in order to avoid metastability. Although the algorithm does not have access to the true signal v_n', we do have noisy observations of it: y_n'. Thus it is natural to choose as initial $c = c^*$ for the Markov chain the value which minimizes

$$\sum_{j=1}^{n-1} \left\| \log\left(\frac{(v_{j+1}', \phi_k)}{(v_j', \phi_k)} \right) - 2\pi i k \cdot c\Delta t \right\|^2. \tag{35}$$

(a) $u(x_1, x_2)$

(b) $\delta c = (0,0)$

(c) $\delta c = (1/2, 1/2)$

(d) $\delta c = (1/e, 1/\pi)$

(e) $\int_0^\infty (c - c'(s)) \, ds = (1/2, 1/2)$

(f) $\int_0^t (c - c'(s; \omega')) \, ds = W(t)$
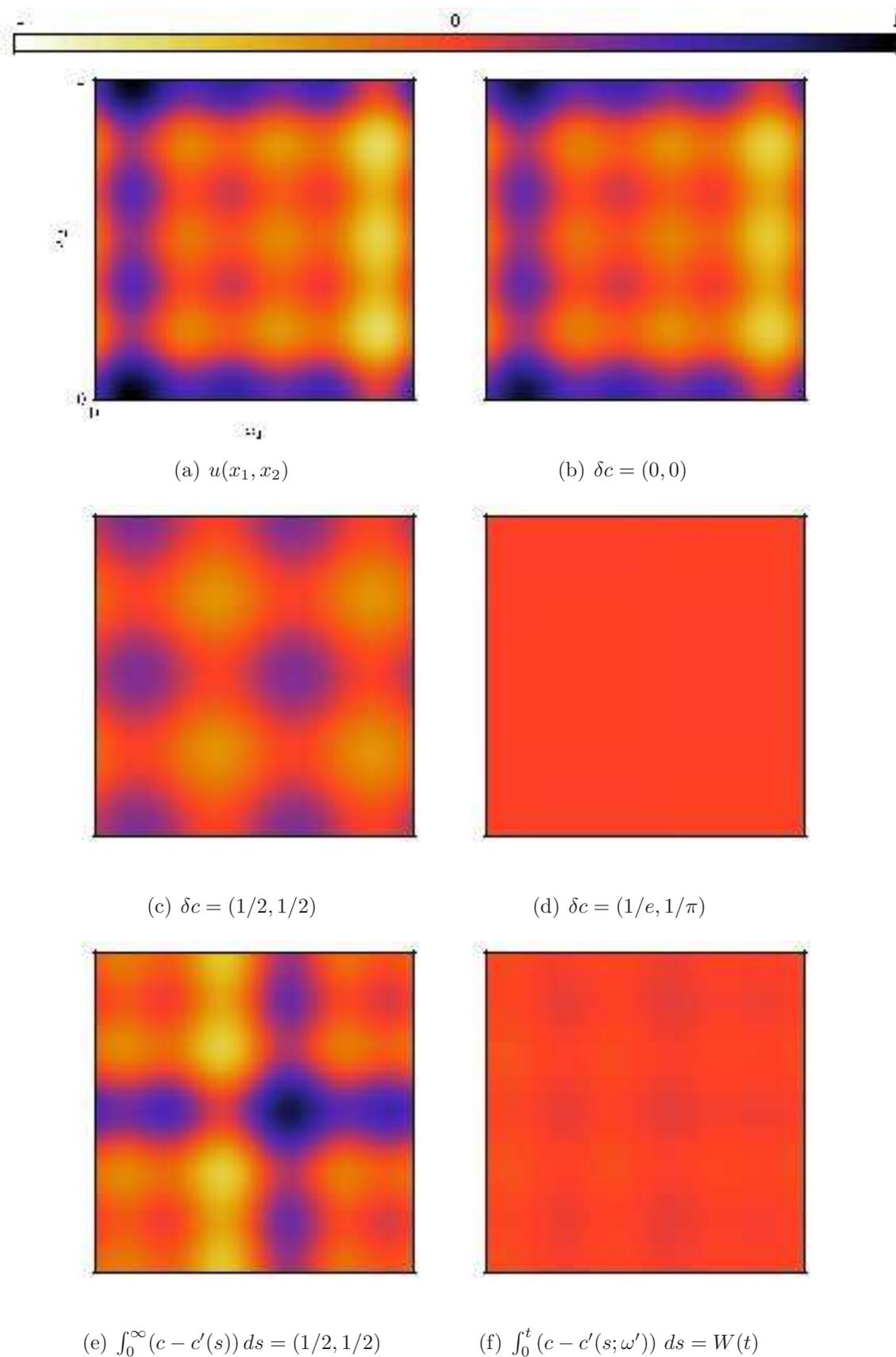
Figure 1. Figure 1(a) is the true initial condition. Figures 1(b) – 1(f) show the desired empirical mean of the smoothing $\mathbb{P}(v_0 | Y_n = Y_n')$ for $\delta c = (0,0), \delta c = (1/2, 1/2), \delta c \in \mathbb{R} \setminus \mathbb{Q} \times \mathbb{R} \setminus \mathbb{Q}, \int_0^\infty \delta c \, dt = (1/2, 1/2)$ and $\delta c = \dot{W}$ respectively.

(a)   (b)
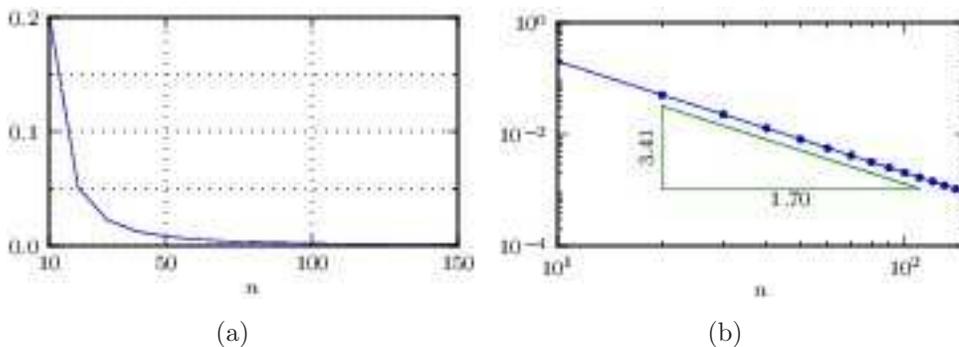
Figure 2. Plot 2(a) shows $\|\mathbb{E}(v_n|c, Y_n = Y_n') - v_n'\|_{L^2(\mathbb{T}^2)}^2$ as a function of n, when $\int_0^\infty \delta c(s)\, ds = (1/2, 1/2)$. Its log-log plot, along with a least squares fit, is depicted in Plot 2(b), demonstrating quadratic convergence.

Because of the observational noise this estimate is more accurate for small values of k and we choose $k = (1, 0)$ to estimate c_1 and $k = (0, 1)$ to estimate c_2.

Figure 3 shows the marginal distribution for c computed with four different values of the data size n, in all cases with the Markov chain seeded as in (35). The results show that the marginal wave velocity distribution $\mathbb{P}(c|Y_n)$ converges to a Dirac on the true value as the amount of data is increased. Although not shown here, the initial condition is also converging to a Dirac on the true value (33) in this limit.

We round-off this subsection by mentioning related published literature. First we mention that, in a setting similar to ours, a scheme to approximate the true wave velocity is proposed which uses parameter estimation within 3D Var for the linear advection equation with constant velocity [9], and its hybrid with the EnKF for the non-constant velocity case [10]. These methodologies deal with the problem entirely in finite dimensions but are not limited to the linear dynamics. Secondly we note that, although a constant wave velocity parameter in the linear advection equation is a useful physical idealization in some cases, it is a very rigid assumption, making the data assimilation problem with respect to this parameter quite hard; this is manifest in the large number of samples required to estimate this constant parameter. A notable, and desirable, direction in which to extend this work numerically is to consider the time-dependent wave velocity as presented in Theorems 4.1–4.3. For efficient filtering techniques to estimate time-dependent parameters, the reader is directed to [39, 40, 41, 42].

6. Conclusions

In this paper, we study an infinite dimensional state estimation problem in the presence of model error. For the statistical model of advection equation on a torus, with noisily observed functions in discrete time, the large data limit of the filter and the smoother both recover the truth in the perfect model scenario. If the actual wave velocity differs from the true wave velocity in a time-integrable fashion then the filter recovers the truth, but the smoother is in error by a constant phase shift, determined by the integral

(a) $n = 10$ (b) $n = 50$

(c) $n = 100$ (d) $n = 1000$
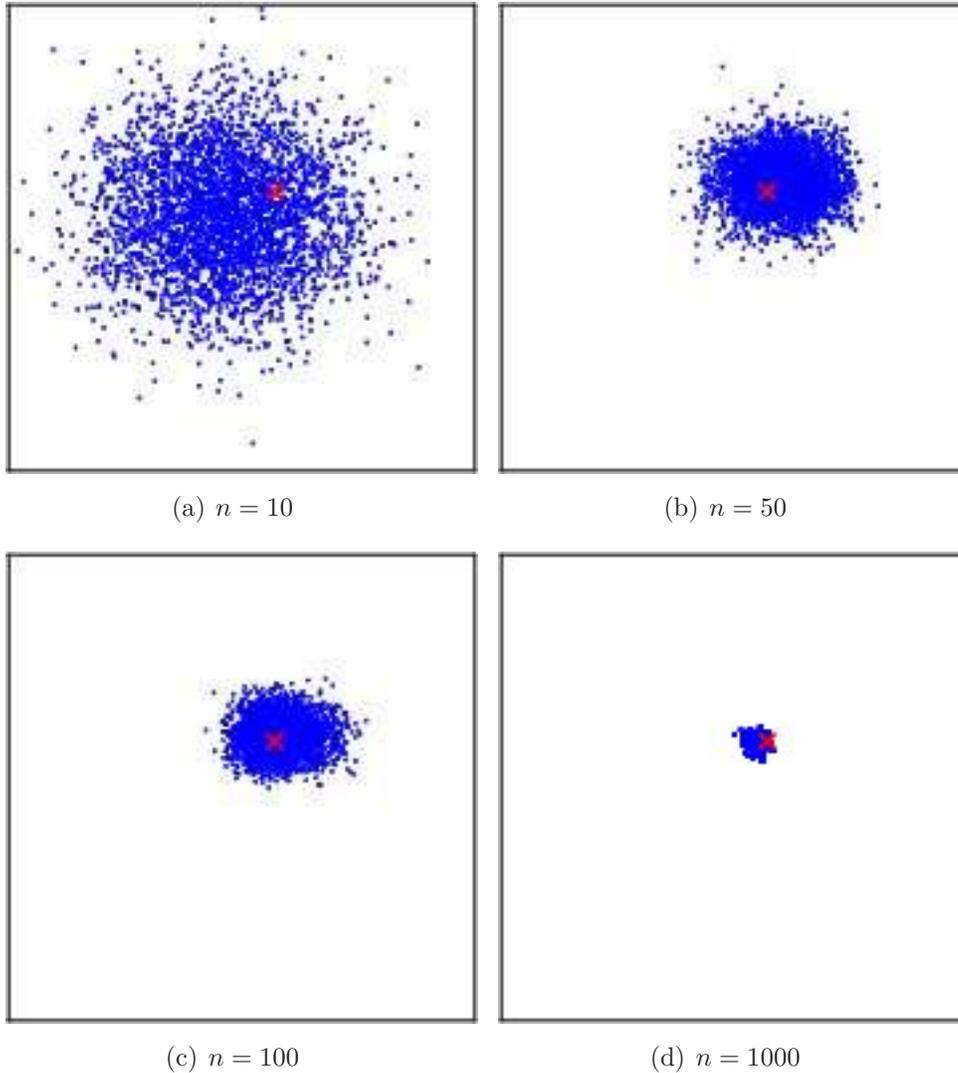
Figure 3. The marginal distribution of $\mathbb{P}(c, v_0 | Y_n)$ with respect to c are depicted on the square 1.4×10^{-4} by $\times 10^{-4}$. The red cross marks the true wave velocity $c = (-0.5, -1.0)$.

of the difference in wave velocities. When the difference in wave velocities is constant neither filtering nor smoothing recovers the truth in the large data limit. And when the difference in wave velocities is a fluctuating random field, however small, neither filtering nor smoothing recovers the truth in the large data limit.

In this paper we consider the dynamics as a hard constraint, and do not allow for the addition of mean zero Gaussian noise to the time evolution of the state. Adding such noise to the model is sometimes known as a weak constraint approach in the data assimilation community and the relative merits of hard and weak constraint approaches are widely debated; see [4, 43] for discussion and references. New techniques of analysis would be required to study the weakly constrained problem, because the inverse covariance does not evolve linearly as it does for the hard constraint problem we study here. We leave this for future study.

There are a number of other ways in which the analysis in this paper could be generalized, in order to obtain a deeper understanding of filtering methods for high dimensional systems. These include: (i) the study of dissipative model dynamics; (ii) the study of nonlinear wave propagation problems; (iii) the study of Lagrangian rather than Eulerian data. Many other generalizations are also possible. For nonlinear systems, the key computational challenge is to find filters which can be justified, either numerically or analytically, and which are computationally feasible to implement. There is already significant activity in this direction, and studying the effect of model/data mismatch will form an important part of the evaluation of these methods.

Acknowledgments

Appendix A. Basic Theorems on Gaussian Measures

Suppose the probability measure μ is defined on the Hilbert space $\mathcal{H}$. A function $m \in \mathcal{H}$ is called the mean of μ if, for all ℓ in the dual space of linear functionals on $\mathcal{H}$,

$$\ell(m) = \int_{\mathcal{H}} \ell(x)\mu(dx),$$

and a linear operator $\mathcal{C}$ is called the covariance operator if for all k, ℓ in the dual space of $\mathcal{H}$,

$$k\left(\mathcal{C}\ell\right) = \int_{\mathcal{H}} k(x - m)\ell(x - m)\mu(dx).$$

In particular, a measure μ is called Gaussian if $\mu \circ \ell^{-1} = \mathcal{N}\left(m_\ell, \sigma_\ell^2\right)$ for some $m_\ell, \sigma_\ell \in \mathbb{R}$. Since the mean and covariance operator completely determine a Gaussian measure, we denote a Gaussian measure with mean m and covariance operator $\mathcal{C}$ by $\mathcal{N}(m, \mathcal{C})$.

The following lemmas, all of which can be found in [29], summarize the properties of Gaussian measures which we require for this paper.

Lemma A.1. *If $\mathcal{N}(0, \mathcal{C})$ is a Gaussian measure on a Hilbert space $\mathcal{H}$, then $\mathcal{C}$ is a self-adjoint, positive semi-definite nuclear operator on $\mathcal{H}$. Conversely, if $m \in \mathcal{H}$ and $\mathcal{C}$ is a self-adjoint, positive semi-definite, nuclear operator on $\mathcal{H}$, then there is a Gaussian measure $\mu = \mathcal{N}(m, \mathcal{C})$ on $\mathcal{H}$.*

Lemma A.2. *Let $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ be a separable Hilbert space with projections $\Pi_i : \mathcal{H} \to \mathcal{H}_i, i = 1, 2$. For an $\mathcal{H}$-valued Gaussian random variable (x_1, x_2) with mean $m = (m_1, m_2)$ and positive-definite covariance operator $\mathcal{C}$, denote $C_{ij} = \Pi_i \mathcal{C} \Pi_j^*$. Then the conditional distribution of x_1 given x_2 is Gaussian with mean*

$$m_{1|2} = m_1 - C_{12}C_{22}^{-1}(m_2 - x_2),$$

and covariance operator

$$\mathcal{C}_{1|2} = C_{11} - C_{12}C_{22}^{-1}C_{21}.$$

Lemma A.3. (Feldman-Hajek) *Two Gaussian measures $\mu_i = \mathcal{N}(m_i, \mathcal{C}_i)$, $i = 1, 2$, on a Hilbert space $\mathcal{H}$ are either singular or equivalent. They are equivalent if and only if the following three conditions hold:*
(i) $Im\left(\mathcal{C}_1^{\frac{1}{2}}\right) = Im\left(\mathcal{C}_2^{\frac{1}{2}}\right) := E$;
(ii) $m_1 - m_2 \in E$;
(iii) the operator $T := \left(\mathcal{C}_1^{-\frac{1}{2}}\mathcal{C}_2^{\frac{1}{2}}\right)\left(\mathcal{C}_1^{-\frac{1}{2}}\mathcal{C}_2^{\frac{1}{2}}\right)^* - I$ is Hilbert-Schmidt in $\bar{E}$.

Lemma A.4. *For any two positive-definite, self-adjoint operators $\mathcal{C}_i$, $i = 1, 2$, on a Hilbert space $\mathcal{H}$, the condition $Im\left(\mathcal{C}_1^{\frac{1}{2}}\right) \subseteq Im\left(\mathcal{C}_2^{\frac{1}{2}}\right)$ holds if and only if there exists a constant $K > 0$ such that*

$$(h, \mathcal{C}_1 h) \leq K(h, \mathcal{C}_2 h), \quad \forall h \in \mathcal{H}$$

where $(\cdot, \cdot)$ denotes the inner product on $\mathcal{H}$.

Lemma A.5. *For Gaussian X on a Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$ and for any integer n, there is constant $C_n \geq 0$ such that $\mathbb{E}(\|X\|^{2n}) \leq C_n \left(\mathbb{E}\left(\|X\|^2\right)\right)^n$.*

Appendix B. Proof of Limit Theorems

In this Appendix, we will prove the Limit Theorems 3.2, 3.7, 4.1, where $\mathcal{L} = \mathcal{L}'$, $\mathcal{L} \neq \mathcal{L}'$, $\mathcal{L}(t) \neq \mathcal{L}'(t)$, and Theorem 4.3 where $\mathcal{L}(t) \neq \mathcal{L}'(t; \omega')$, respectively. In all cases, we use the notations $e^{-t\mathcal{L}}$ and $e^{-t\mathcal{L}'}$ to denote the forward solution operators through t time units (from time zero in the non-autonomous case). We denote by M the putative limit for m'_n. The identity

$$\left(nI + \Gamma\mathcal{C}_0^{-1}\right)(m'_n - M) = \Gamma\mathcal{C}_0^{-1}(m_0 - M) + \sum_{l=0}^{n-1} e^{t_{l+1}\mathcal{L}}\eta'_{l+1}$$

$$+ \sum_{l=0}^{n-1}\left(e^{t_{l+1}(\mathcal{L}-\mathcal{L}')}u - M\right), \tag{B.1}$$

obtained from Equation (14), will be used to show $m'_n \to M$. In Equation (B.1), we will choose M so that the contribution of the last term is asymptotically negligible. Define the Fourier representations

$$\mathbf{e}_n \equiv m'_n - M = \sum_{k \in \mathbb{K}} \hat{\mathbf{e}}_n(k)\phi_k,$$

$$\xi_l \equiv e^{t_{l+1}(\mathcal{L}-\mathcal{L}')}u - M = \sum_{k \in \mathbb{K}} \hat{\xi}_l(k)\phi_k.$$

Then M will be any one of u, $\mathcal{F}_{(p,q)}u$, $\langle u \rangle$ and u_α. Hence there is C_1 independent of k, l such that $\hat{\xi}_l(0) = 0$ and $\mathbb{E}|\hat{\xi}_l(k)| \leq C_1|(u, \phi_k)|$ with $C_1 < \infty$ (the expectation here is

trivial except in the case of random $\mathcal{L}'$). Using Equation (17), these Fourier coefficients satisfy the relation

$$\left(n + \frac{\gamma_k}{\lambda_k} \right) \hat{\mathbf{e}}_n(k) = \frac{\gamma_k}{\lambda_k} \hat{\mathbf{e}}_0(k) + \sqrt{\gamma_k'} \sum_{l=0}^{n-1} (\mathbf{g}_k')^{l+1} + \sum_{l=0}^{n-1} \hat{\xi}_l(k), \quad k \in \mathbb{K}.$$

In order to prove $m_n' \to M$ in $L^2(\Omega'; H^s(\mathbf{T}^2))$, we use the monotone convergence theorem to obtain the following inequalities,

$$
\begin{aligned}
n^\delta \|\mathbf{e}_n\|_{L^2(\Omega'; H^s(\mathbf{T}^2))}^2 &= n^\delta \mathbb{E} \|\mathbf{e}_n\|_{H^s(\mathbf{T}^2)}^2 \\
&= \sum_{k \in \mathbb{K}^+} |k|^{2s} n^\delta \mathbb{E} |\hat{\mathbf{e}}_n(k)|^2 + n^\delta \mathbb{E} |\hat{\mathbf{e}}_n(0)|^2 \\
&= \sum_{k \in \mathbb{K}^+} |k|^{2s} \frac{n^\delta}{(n + \gamma_k/\lambda_k)^2} \left[\left(\frac{\gamma_k}{\lambda_k} \right)^2 |\hat{\mathbf{e}}_0(k)|^2 + \gamma_k' n \right. \\
&\quad + 2\mathrm{Re} \left\{ \frac{\gamma_k}{\lambda_k} \bar{\hat{\mathbf{e}}}_0(k) \mathbb{E} \left(\sum_{l=0}^{n-1} \hat{\xi}_l(k) \right) \right\} + \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \right] \\
&\quad + \frac{n^\delta}{(n + \gamma_0/\lambda_0)^2} \left[\left(\frac{\gamma_0}{\lambda_0} \right)^2 |\hat{\mathbf{e}}_0(0)|^2 + \gamma_0' n \right] \\
&\leq \sum_{k \in \mathbb{K}^+} |k|^{2s} n^{\delta-2} \left[\left(\frac{\gamma_k}{\lambda_k} \right)^2 |\hat{\mathbf{e}}_0(k)|^2 \right. \\
&\quad + \gamma_k' n + 2C_1 \frac{\gamma_k}{\lambda_k} |\hat{\mathbf{e}}_0(k)| |(u, \phi_k)| n + \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \right] \\
&\quad + n^{\delta-2} \left[\left(\frac{\gamma_0}{\lambda_0} \right)^2 |\hat{\mathbf{e}}_0(0)|^2 + \gamma_0' n \right] \\
&\leq \sum_{k \in \mathbb{K}^+} |k|^{2s} n^{\delta-2} \left[\left(\frac{\gamma_k}{\lambda_k} \right)^2 |\hat{\mathbf{e}}_0(k)|^2 + \gamma_k' n \right. \\
&\quad + C_1 \left(\left(\frac{\gamma_k}{\lambda_k} \right)^2 |\hat{\mathbf{e}}_0(k)|^2 + |(u, \phi_k)|^2 \right) n + \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \right] \\
&\quad + n^{\delta-2} \left[\left(\frac{\gamma_0}{\lambda_0} \right)^2 |\hat{\mathbf{e}}_0(0)|^2 + \gamma_0' n \right] \\
&\leq \left(C \|m_0 - M\|_{H^{s+\kappa}(\mathbf{T}^2)} \right) n^{\delta-2} + \left(\sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k' \right) n^{\delta-1} \\
&\quad + \left(C_1 \|u\|_{H^s(\mathbf{T}^2)} \right) n^{\delta-1} + \left(\sum_{k \in \mathbb{K}^+} |k|^{2s} n^{\delta-2} \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \right) \\
&\leq C \left(\|m_0\|_{H^{s+\kappa}(\mathbf{T}^2)} + \|u\|_{H^{s+\kappa}(\mathbf{T}^2)} \right) n^{\delta-2}
\end{aligned}
$$

$$+ \left(\sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k' + C_1 \|u\|_{H^s(\mathbf{T}^2)} \right) n^{\delta-1}$$

$$+ \left(\sum_{k \in \mathbb{K}} |k|^{2s} n^{\delta-2} \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \right), \tag{B.2}$$

and here the first two terms in the last equation can be controlled by Assumptions 3.1. In order to find $\delta \in [0,1]$ such that this equation is $O(1)$ or $o(1)$,

$$\sum_{k \in \mathbb{K}} |k|^{2s} n^{\delta-2} \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \tag{B.3}$$

is the key term. This term arises from the model error, i.e., the discrepancy between the operator $\mathcal{L}$ used in the statistical model and the operator $\mathcal{L}'$ which generates the data. We analyze it, in various cases, in the subsections which follow.

In order to prove $m_n' \to M$ in $H^s(\mathbf{T}^2)$, $\Omega' - a.s.$, suppose we have

$$\frac{1}{n} \sum_{l=0}^{n-1} \hat{\xi}_l(k) \to 0 \quad \Omega' - a.s., \tag{B.4}$$

for each $k \in \mathbb{K}$. We then use the strong law of large numbers to obtain the following inequalities, which holds $\Omega' - a.s.$,

$$\begin{aligned}
\|\mathbf{e}_n\|_{H^s(\mathbf{T}^2)}^2 &= \sum_{k \in \mathbb{K}^+} |k|^{2s} |\hat{\mathbf{e}}_n(k)|^2 + |\hat{\mathbf{e}}_n(0)|^2 \\
&= \sum_{k \in \mathbb{K}^+} |k|^{2s} \left| \frac{(\gamma_k/\lambda_k)\,\hat{\mathbf{e}}_0(k) + \sqrt{\gamma_k'} \sum_{l=0}^{n-1} (\mathbf{g}_k')^{l+1} + \sum_{l=0}^{n-1} \hat{\xi}_l(k)}{n + \gamma_k/\lambda_k} \right|^2 \\
&\quad + \left| \frac{(\gamma_0/\lambda_0)\,\hat{\mathbf{e}}_0(0) + \sqrt{\gamma_0'} \sum_{l=0}^{n-1} (\mathbf{g}_0')^{l+1}}{n + \gamma_0/\lambda_0} \right|^2 \\
&\leq C \sum_{k \in \mathbb{K}} (1 + |k|^{2s}) \left(\left| \frac{(\gamma_k/\lambda_k)\,\hat{\mathbf{e}}_0(k)}{n} \right|^2 + \left| \frac{\sqrt{\gamma_k'} \sum_{l=0}^{n-1} (\mathbf{g}_k')^{l+1}}{n} \right|^2 \right. \\
&\quad \left. + \left| \frac{\sum_{l=0}^{n-1} \hat{\xi}_l(k)}{n} \right|^2 \right) \\
&\leq C \sum_{k \in \mathbb{K}} (1 + |k|^{2s}) \left(|k|^{2\kappa} |\hat{\mathbf{e}}_0(k)|^2 + \gamma_k' + |(u, \phi_k)|^2 \right) \\
&\leq C \left(\|m_0\|_{H^{s+\kappa}(\mathbf{T}^2)} + \|u\|_{H^{s+\kappa}(\mathbf{T}^2)} + \sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k' + \|u\|_{H^s(\mathbf{T}^2)} \right). \tag{B.5}
\end{aligned}$$

Therefore, using Weierstrass M-test, we have $\|\mathbf{e}_n\|_{H^s(\mathbf{T}^2)}^2 \to 0$, $\Omega' - a.s.$, once Equation (B.4) is satisfied.

Appendix B.1. Proof of Theorem 3.2

This proof is given directly after the theorem statement. For completeness we note that Equation (15a) follows from Equation (B.2) with $\delta = 1$. Once it has been established that

$$\|\cdot\|_{L^2(\Omega';H^s(\mathbf{T}^2))} = O\left(n^{-\delta/2}\right),$$

then the proof that

$$\|\cdot\|_{H^s(\mathbf{T}^2)} = o\left(n^{-\theta}\right) \quad \Omega' - a.s.,$$

for any $\theta < \delta/2$ follows from a Borel-Cantelli argument as shown in the proof of Theorem 3.2. We will not repeat this argument for the proofs of Theorems 3.7, 4.1 and 4.3.

Appendix B.2. Proof of Theorem 3.7

Appendix B.2.1. When $\mathcal{L} \neq \mathcal{L}'$ and $\Delta t\,\delta c = (p'/p, q'/q)$, choose $M = \mathcal{F}_{(p,q)}u$ then

$$\hat{\xi}_l(k) = e^{2\pi i(k\cdot\delta c)t_{l+1}}(u, \phi_k) - \delta_{(p,q)}^k(u, \phi_k)$$
$$= \left(1 - \delta_{(p,q)}^k\right)e^{2\pi i(k\cdot\delta c)t_{l+1}}(u, \phi_k),$$

where $\delta_{(p,q)}^k$ is 1 if k_1 and k_2 for $k = (k_1, k_2)$ are multiples of p and q respectively, and 0 otherwise. Using

$$\left|\sum_{l=0}^{n-1}\hat{\xi}_l(k)\right|^2 = \left(1 - \delta_{(p,q)}^k\right)^2\left|\sum_{l=0}^{n-1}e^{2\pi i(k\cdot\delta c)t_{l+1}}\right|^2|(u, \phi_k)|^2$$

$$= \left(1 - \delta_{(p,q)}^k\right)\left[\frac{\sin(n\pi(k\cdot\delta c)\Delta t)}{\sin(\pi(k\cdot\delta c)\Delta t)}\right]^2|(u, \phi_k)|^2$$

$$\leq \left[\sin^2\left(\pi\left(\frac{1}{p} + \frac{1}{q}\right)\right)\right]^{-1}|(u, \phi_k)|^2,$$

the quantity in Equation (B.3) becomes

$$\left(\left[\sin^2\left(\pi\left(\frac{1}{p} + \frac{1}{q}\right)\right)\right]^{-1}\|u\|_{H^s(\mathbf{T}^2)}\right)n^{\delta-2},$$

so that from Equation (B.2)

$$n\,\|\mathbf{e}_n\|_{L^2(\Omega';H^s(\mathbf{T}^2))}^2 = O(1).$$

Appendix B.2.2. When $\mathcal{L} \neq \mathcal{L}'$ and $\Delta t\,\delta c \in \mathbb{R}\backslash\mathbb{Q} \times \mathbb{R}\backslash\mathbb{Q}$, choose $M = \langle u \rangle$ then

$$\hat{\xi}_l(k) = e^{2\pi i(k\cdot\delta c)t_{l+1}}(u, \phi_k),$$

for $k \in \mathbb{K}^+$. It is immediate that

$$\|\mathbf{e}_n\|_{H^s(\mathbf{T}^2)} = o(1) \quad \Omega' - a.s.,$$

since

$$\frac{1}{n}\sum_{l=0}^{n-1}e^{2\pi i(k\cdot\delta c)t_{l+1}} = o(1),$$

as $n \to \infty$, by ergodicity [33] and Equation (B.4). Furthermore, when $\delta = 0$, the quantity in Equation (B.3) is bounded by

$$\sum_{k \in \mathbb{K}} |k|^{2s} \left| \frac{1}{n} \sum_{l=0}^{n-1} e^{2\pi i (k \cdot \delta c) t_{l+1}} \right|^2 |(u, \phi_k)|^2 \leq C \|u\|_{H^s(\mathbf{T}^2)},$$

and Weierstrass M-test can be used to show

$$\|\mathbf{e}_n\|_{L^2(\Omega'; H^s(\mathbf{T}^2))}^2 = o(1).$$

Appendix B.3. Proof of Theorem 4.1

When $\mathcal{L}(t) \neq \mathcal{L}'(t)$ and $\int_0^t \delta c(s) \, ds = \alpha + O(t^{-\beta})$, choose $M = u_\alpha$ then

$$\hat{\xi}_l(k) = \left(e^{2\pi i k \cdot \left(\int_0^{t_{l+1}} \delta c(s) \, ds \right)} - e^{2\pi i k \cdot \alpha} \right) (u, \phi_k),$$

and we obtain

$$\left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 = \left| \sum_{l=0}^{n-1} \left(e^{2\pi i k \cdot \left(\int_0^{t_{l+1}} \delta c(s) \, ds - \alpha \right)} - 1 \right) \right|^2 |(u, \phi_k)|^2.$$

Now utilizing that

$$\left| \sum_{l=0}^{n-1} \left(e^{i x_l} - 1 \right) \right| \leq C \sum_{l=0}^{n-1} |x_l| \leq C_2 n^{1-\beta},$$

when $x_l = O\left(l^{-\beta} \right)$ and for $\beta \in (0, 1/2]$, Equation (B.2) gives

$$\begin{aligned}
n^\delta \|\mathbf{e}_n\|_{L^2(\Omega'; H^s(\mathbf{T}^2))}^2 \leq\ & C \left(\|m_0\|_{H^{s+\kappa}(\mathbf{T}^2)} + \|u\|_{H^{s+\kappa}(\mathbf{T}^2)} \right) n^{\delta-2} \\
& + \left(\sum_{k \in \mathbb{K}} |k|^{2s} \gamma_k' + C_1 \|u\|_{H^s(\mathbf{T}^2)} \right) n^{\delta-1} \\
& + \left(C_2 \|u\|_{H^s(\mathbf{T}^2)}^2 \right) n^{\delta-2\beta} = O(1),
\end{aligned}$$

for $\delta = 1 \wedge 2\beta$.

Appendix B.4. Proof of Theorem 4.3

When $\mathcal{L}(t) \neq \mathcal{L}'(t; \omega')$ and $\int_0^t c'(s; \omega') \, ds = \int_0^t c(s) \, ds - \varepsilon W(t)$, choose $M = \langle u \rangle$ then we obtain

$$\xi_l = u \left(\cdot + \varepsilon W(t_{l+1}) \right) - \langle u \rangle,$$

and

$$\hat{\xi}_l(k) = e^{2\pi i k \cdot \varepsilon W(t_{l+1})} (u, \phi_k),$$

for $k \in \mathbb{K}^+$. Using

$$\mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 = \mathbb{E} \left(\sum_{l=0}^{n-1} \sum_{l'=0}^{n-1} e^{2\pi i k \cdot \varepsilon \left(W(t_{l+1}) - W(t_{l'+1}) \right)} |(u, \phi_k)|^2 \right)$$

$$= \sum_{l=0}^{n-1} \sum_{l'=0}^{n-1} e^{-2\pi^2 \varepsilon^2 |k|^2 \Delta t |l-l'|} |(u, \phi_k)|^2$$

$$= \left(n + 2 \sum_{l=1}^{n-1} \sum_{l'=0}^{n-1} e^{-2\pi^2 \varepsilon^2 |k|^2 \Delta t |l-l'|} \right) (u, \phi_k)|^2$$

$$= \left[n + \frac{2}{e^{2\pi^2 \varepsilon^2 |k|^2 \Delta t} - 1} \left(\frac{e^{-2\pi^2 \varepsilon^2 |k|^2 \Delta t (n-1)} - 1}{e^{2\pi^2 \varepsilon^2 |k|^2 \Delta t} - 1} + n - 1 \right) \right] |(u, \phi_k)|^2$$

$$\leq \left[n + \frac{2}{e^{2\pi^2 \varepsilon^2 \Delta t} - 1} n \right] |(u, \phi_k)|^2 = \left(\frac{e^{2\pi^2 \varepsilon^2 \Delta t} + 1}{e^{2\pi^2 \varepsilon^2 \Delta t} - 1} \right) n |(u, \phi_k)|^2,$$

we get

$$\sum_{k \in \mathbb{K}^+} |k|^{2s} n^{\delta-2} \mathbb{E} \left| \sum_{l=0}^{n-1} \hat{\xi}_l(k) \right|^2 \leq \left(\frac{e^{2\pi^2 \varepsilon^2 \Delta t} + 1}{e^{2\pi^2 \varepsilon^2 \Delta t} - 1} \right) \|u\|^2_{H^s(\mathbf{T}^2)} n^{\delta-1} = O(1),$$

for $\delta = 1$.

References

[1] Anderson B, Moore J and Barratt J 1979
[2] Doucet A, De Freitas N and Gordon N 2001 *Sequential Monte Carlo methods in practice* (Springer Verlag) ISBN 0387951466
[3] Kalman R *et al.* 1960 *Journal of basic Engineering* **82** 35–45
[4] Bennett A 2002 *Inverse modeling of the ocean and atmosphere* (Cambridge Univ Pr) ISBN 0521813735
[5] Evensen G 1994 *Journal of geophysical research* **99** 10143 ISSN 0148-0227
[6] Kalnay E 2003 *Atmospheric modeling, data assimilation, and predictability* (Cambridge Univ Pr) ISBN 0521796296
[7] van Leeuwen P 2009
[8] Majda A and Grote M 2007 *Proceedings of the National Academy of Sciences* **104** 1124
[9] Smith P, Dance S, Baines M, Nichols N and Scott T 2009 *Ocean Dynamics* **59** 697–708 ISSN 1616-7341
[10] Smith P, Dance S and Nichols N 2010
[11] Griffith A and Nichols N 2001 *J. of Flow, Turbulence and Combustion* **65** 469–488
[12] Evensen G 2009 *Data assimilation: The ensemble Kalman filter* (Springer Verlag) ISBN 3642037100
[13] Evensen G v L P J 1996 *Monthly Weather Review* **124** 85–96
[14] Snyder C, Bengtsson T, Bickel P and Anderson J 2008 *Monthly Weather Review* **136** 4629–4640 ISSN 1520-0493
[15] Bengtsson T, Bickel P and Li B 2008 *Probability and Statistics: Essays in Honor of David A. Freedman* **2** 316–334
[16] Bickel P, Li B and Bengtsson T 2008 *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh* **3** 318–329
[17] Chorin A and Krause P 2004 *Proceedings of the National Academy of Sciences of the United States of America* **101** 15013
[18] Chorin A and Tu X 2009 *Proceedings of the National Academy of Sciences* **106** 17249
[19] Chorin A and Tu X 2009 *Arxiv preprint arXiv:0910.3241*
[20] Harlim J and Majda A 2008 *Nonlinearity* **21** 1281

[21] Castronovo E, Harlim J and Majda A 2008 *Journal of Computational Physics* **227** 3678–3714 ISSN 0021-9991

[22] Majda A, Harlim J and Gershgorin B 2010 *Discrete and Continuous Dynamical Systems* **27** 441–486

[23] Jean-Pierre F and Simoni A 2010 *IDEI Working Papers*

[24] Neubauer A and Pikkarainen H 2008 *Journal of Inverse and Ill-posed Problems* **16** 601–613 ISSN 0928-0219

[25] van der Vaart A W and van Zanten J H 2008 *Annals of Statistics* **36** 1435–1463

[26] Pikkarainen H 2006 *Inverse Problems* **22** 365

[27] Smith P, Dance S, Baines M, Nichols N and Scott T 2009 *Ocean Dynamics* **59** 697–708 ISSN 1616-7341

[28] Cotter S, Dashti M, Robinson J and Stuart A 2009 *Inverse Problems* **25** 115008

[29] Da Prato G and Zabczyk J 1992 *Stochastic equations in infinite dimensions* (Cambridge Univ Pr) ISBN 0521385296

[30] Varadhan S 2001 *New York University Courant Institute of Mathematical Sciences, New York*

[31] Bogachev V 1998 *Gaussian measures* vol 62 (American Mathematical Society)

[32] Kumar P and Varaiya P 1986 *Stochastic systems: estimation, identification and adaptive control* (Prentice-Hall, Inc.)

[33] Breiman L 1968 *Reading, Mass*

[34] Flandoli F, Gubinelli M and Priola E 2010 *Inventiones Mathematicae* **180** 1–53 ISSN 0020-9910

[35] Stuart A 2010 *Acta Numerica* **19** 451–559 ISSN 1474-0508

[36] Cotter SL D M and Robinson JC S A 2011 *International Journal for Numerical Methods in Fluids*

[37] Geweke J and Tanizaki H 2001 *Computational Statistics & Data Analysis* **37** 151–170 ISSN 0167-9473

[38] Roberts G and Rosenthal J 2006 *The Annals of Applied Probability* **16** 2123–2139 ISSN 1050-5164

[39] Cohn S 1997 *Journal of the Metereological Society of Japan* **75** 147–178

[40] Dee D and Da Silva A 1998 *Quarterly Journal of the Royal Meteorological Society* **124** 269–295

[41] Baek S, Hunt B, Kalnay E, Ott E and Szunyogh I 2006 *Tellus Series A* **58** 293–306

[42] Gershgorin B, Harlim J and Majda A 2010 *Journal of Computational Physics* **229** 1–31

[43] Apte A, Jones C, Stuart A and Voss J 2008 *International Journal for Numerical Methods in Fluids* **56** 1033–1046