Assimilating irregularly spaced sparsely observed turbulent signals with hierarchical Bayesian reduced stochastic filters

Kristen A. Brown, John Harlim*

Department of Mathematics, North Carolina State University, NC 27695

Abstract

In this paper, we consider a practical filtering approach for assimilating irregularly spaced, sparsely observed turbulent signals through a hierarchical Bayesian reduced stochastic filtering framework. The proposed hierarchical Bayesian approach consists of two steps, blending a data-driven interpolation scheme and the Mean Stochastic Model (MSM) filter. We examine the potential of using the deterministic piecewise linear interpolation scheme and the ordinary kriging scheme in interpolating irregularly spaced raw data to regularly spaced processed data and the importance of dynamical constraint (through MSM) in filtering the processed data on a numerically stiff state estimation problem. In particular, we test this approach on a two-layer quasi-geostrophic model in a two-dimensional domain with a small radius of deformation to mimic ocean turbulence. Our numerical results suggest that the dynamical constraint becomes important when the observation noise variance is large. Second, we find that the filtered estimates with ordinary kriging

*Corresponding author.

Email addresses: kabrown6@ncsu.edu (Kristen A. Brown), jharlim@ncsu.edu (John Harlim)
are superior to those with linear interpolation when observation networks are not too sparse; such robust results are found from numerical simulations with many randomly simulated irregularly spaced observation networks, various observation time intervals, and observation error variances. Third, when the observation network is very sparse, we find that both the kriging and linear interpolations are comparable.

*Keywords:* hierarchical Bayesian reduced stochastic filter, Mean Stochastic Model, data assimilation, filtering interpolated data

1. Introduction

Filtering is a numerical method for finding the best statistical estimate of the true signal based on partially observed noisy measurements. The goal of this paper is to extend the reduced stochastic models for filtering turbulent complex systems [1] to account for irregularly spaced sparse observations. This problem has been studied earlier in [2] on a one-dimensional domain problem. Here, our main contributions are to extend the approach in [2] with a spatial statistical interpolation scheme (below, we will provide a probabilistic interpretation of this approach) and to generalize it to two-dimensional domain problems. As in [2], we will consider implementing the simplest reduced stochastic models in [1], the Mean Stochastic Model (MSM), whose designed was attributed to the standard approach for modeling turbulent fluctuations [3, 4, 5, 6, 7, 8] with a linear damping and white noise (replacing nonlinear terms).

The Mean Stochastic Model (MSM) consists of block diagonal Langevin equations in Fourier space where the parameters are obtained through an
off-line regression fitting to a training data set (we will review this filtering method in Section 3 below). Given special observation networks (including the case of plentiful observations [9] and the case of sparse regularly spaced observations [10]) with i.i.d. Gaussian noise, we can implement the Kalman filter in Fourier space including only correlations between appropriate wavenumbers. For such special observation networks, this cheap diagonal filter (MSM) has shown encouraging results in various applications, including: the mid-latitude barotropic flows with baroclinic instability and shorter radius of deformation with sparse observations in a numerically stiff regime, mimicking ocean turbulence [11], and the initiation of moist convectively coupled waves in the tropics as well as the Madden Julian Oscillation-like traveling waves [12]. Independently in [13], Law and Stuart assessed the performance of various data assimilation methods including the MSM (they called it the Fourier Diagonal Filter or FDF) on the 2D Navier-Stokes equations resolving $O(10^3)$ horizontal grid points; in particular, they found that in a turbulent regime the FDF performs better than 3D-VAR and ensemble Kalman filter for longer observation times and they explained that the success of FDF in this regime is due to the stability of the forward model MSM by design.

In practice, observations are not always available at regularly spaced grid points, e.g., the radiosonde measurements are sparse in the ocean and dense on the land. The conventional filtering approach [14, 15, 16] for irregularly spaced observations is to use the following canonical observation model

$$\tilde{v}_m = G(u_m) + \tilde{\sigma}_m, \quad \tilde{\sigma}_m \sim \mathcal{N}(0, r^o I),$$

where operator $G$ is an appropriate choice of interpolation function that
maps the physical space model variables \( u_m = (u_{j,m})_{j=1}^{N} \) to the physical space observation variables \( \tilde{v}_m = (\tilde{v}_{\ell,m})_{\ell=1}^{M} \); here, index \( j \) denotes the regularly spaced model grid point, index \( \ell \) denotes the irregular observation grid point, and index \( m \) denotes the discrete time step. In (1), the observation \( \tilde{v}_m \) is corrupted by an i.i.d. Gaussian noise \( \tilde{\sigma}_m^o \) with mean zero and variance \( r^o \).

Given such an irregularly spaced observation platform, the reduced filtering strategy with an analogous Fourier based observation model as described above is not attainable.

On the other hand, there are plenty of applications where data assimilation is applied on processed data where the “processing scheme” varies from simple interpolations to inversion of complicated nonlinear differential equations, (e.g., the satellite retrieval data, the sea surface temperature measurement from the TRMM - Tropical Rainfall Measuring Mission satellite - Microwave Image, is assimilated in tropical pacific hindcast experiment in [17]). In many applications, processing data is unavoidable since the raw data can be very noisy and uninformative (e.g., an important aspect of hydrological modeling is the areal estimation of precipitation given limited rain-gauge measurement). Motivated by this issue, the second author studied the effect of various interpolation techniques of noisy data on filtering a one-dimensional toy model for complex turbulent signals [2] and found that MSM applied to the processed data (with an appropriate interpolation scheme) produces more accurate filtered solutions beyond the conventional strategy, with observational model in (1), when observations are extremely irregularly spaced and very sparse.

Probabilistically, the standard approach in (1) and the reduced filtering
approach in [2] can be described as follows. Consider \( \tilde{V} \) as the random variable of the irregularly spaced observations and \( U \) as the random variable of the model state. The conventional filtering approach is an application of Bayes’ theorem to obtain the posterior statistics of the following conditional distribution

\[
P(u|\tilde{v}) \propto P(u)P(\tilde{v}|u),
\]

where \( P(u) \) denotes the prior distribution associated with the dynamics and \( P(\tilde{v}|u) \) denotes a likelihood function corresponding to the observation model in (1). The approach in [2] can be probabilistically interpreted as an application of the hierarchical Bayesian principle:

\[
P(u|\tilde{v}, v) \propto P(u)P(v|u, \tilde{v}) \propto P(u)P(v|u)P(\tilde{v}|v, u),
\]

where \( v \) is a sample from \( V \), the random variable of the interpolated observations at the regularly spaced model grid points. The hierarchical Bayesian principle in (3) can be understood as follows: The first step is to apply \( P(v|u)P(\tilde{v}|v, u) \) through a statistical interpolation technique to interpolate \( \tilde{v} \in \tilde{V} \) to \( v \in V \). The outcome can be represented by a conditional distribution \( P(v|u, \tilde{v}) \). The second step is to apply \( P(u)P(v|u, \tilde{v}) \) through a reduced stochastic filtering algorithm; here, the computational cost is significantly reduced compared to the standard Bayesian approach in (2) with the observation model in (1). This two-step process blends the purely data-driven approach in the first step and the dynamical constraint through the prior distribution in the second step. The approximate reduced filter in [2] is a special application of this two-step hierarchical Bayesian framework in which the first step is applied using standard deterministic interpolation schemes.
In this paper, we will apply the hierarchical Bayesian approach in (3) to filter irregularly spaced, sparsely observed geophysical turbulence on a two-dimensional spatial periodic domain [3, 18], in a numerically stiff regime through small radius of deformation mimicking ocean turbulence [19, 11]. In particular, we will compare the performance of two interpolation schemes on various observation networks for the first step in (3): the deterministic piecewise linear interpolation that is assessed in [2] and the spatial statistical interpolation, ordinary kriging [20]. As we mentioned above, we will only consider the simplest reduced stochastic model, MSM, for the second step throughout this paper.

The paper is organized as follows: in Section 2, we briefly review the two-layer quasi-geostrophic model [3, 18] to be used in our numerical examples. In Section 3, we will discuss step-by-step the hierarchical Bayesian approach in (3) in detail. Subsequently, we report the filtering results in Section 4 and close the paper with a short summary in Section 5.

2. Truth Model

We will consider prototype studies of oceanic turbulent dynamics utilizing the quasi-geostrophic approximation of the primitive equation [3, 21]. As a simple paradigm model, we consider a two-layer quasi-geostrophic (QG) model that is externally forced by a mean shear with streamfunctions

\[ \Psi_1 = -Uy, \quad \Psi_2 = Uy, \]  

such that it exhibits baroclinic instability; the properties of the turbulent cascade have been extensively discussed in this setting (e.g., see [3, 21] and citations in [18]).
The governing equations for the two-layer QG model with a flat bottom, rigid lid, and equal depth $H$ are given by:

$$
\left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) q_1 + J(\psi_1, q_1) + \frac{\partial \psi_1}{\partial x} (\beta + k_d^2 U) + \nu \nabla^8 q_1 = 0, \quad (5)
$$

$$
\left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) q_2 + J(\psi_2, q_2) + \frac{\partial \psi_2}{\partial x} (\beta - k_d^2 U) + \kappa \nabla^2 \psi_2 + \nu \nabla^8 q_2 = 0, \quad (6)
$$

where $\psi$ denotes the perturbed streamfunction in (4), subscript 1 corresponds to the upper layer and 2 the bottom [2, 21]. In (5)-(6), $\beta$ is the meridional gradient of the Coriolis parameter; $\kappa$ is the Ekman bottom drag coefficient; $J(\psi, q) = \psi_x q_y - \psi_y q_x$ is a Jacobian function that acts as nonlinear advection; $U$ is the zonal mean shear, selected so that the QG equations exhibit baroclinic instability with a turbulent cascade; $\nu$ is the hyperviscosity coefficient, chosen so that $\nu \nabla^8 q$ filters out the energy buildup on smaller scales when finite discretization is enforced. The perturbed QG potential vorticity $q$ is defined in each layer by

$$
q_i = \nabla^2 \psi_i + \frac{1}{2} \frac{k_d^2}{d} (\psi_{3-i} - \psi_i). \quad (7)
$$

The parameter $k_d = \sqrt{8}/L_d$ gives the wavenumber associated with radius of deformation, or Rossby radius, $L_d$ (the scale at which Earth's rotation becomes significant to the dynamics of the system).

In our numerical simulations, the true signal is generated by resolving (5)-(6) with $128 \times 64 \times 2$ Fourier modes, which correspond to $128 \times 128 \times 2$ grid points. With such resolution, the numerical integration of this turbulent system only needs slightly more than 30,000 state variables because one can always compute $\psi$ from $q$ and vice versa, via (7). This model has two important nondimensional parameters: $b = \beta(L/2\pi)^2/U_o$, where $U_o = 1$ is
the horizontal non-dimensionalized velocity scale and $L$ is the horizontal domain size in both directions (we choose $L = 2\pi$), and $F = (L/2\pi)^2/L_d^2$, which is inversely proportional to the deformation radius [11]. To imitate a turbulent jet in the ocean, we choose a short Rossby radius $L_d$, such that $F = L_d^{-2} = 40$, as in [19, 21]. Note that Equations (5)-(6) are both stiff, as $k_d^2 = 8/L_d^2 \approx 320$ is quite large. In this regime, instability occurs on a wide waveband and the horizontal wavenumbers, $k$ and $l$, associated with the maximum growth rate $\tilde{\sigma}_{\text{max}} = \text{Im}(c)k$, with wave speed $c$, belong to the set $\{(k, l) : 1 \leq k \leq 15, 0 \leq l \leq 15, 8.5 < \sqrt{k^2 + l^2} < 15.6\}$.

The large-scale components of this turbulent system are barotropic [3, 19] and this regime exhibits barotropic blocking regime transitions between zonal flow and large-scale Rossby waves (see Figure 1) [8, 11]. For the two layer model, the barotropic streamfunction is defined as an average between the two layers, $\psi^b = (\psi_1 + \psi_2)/2$. Furthermore, let $q^b$ be the average of $q_1$ and $q_2$ and define the baroclinic (small scale) terms $\psi^c = (\psi_1 - \psi_2)/2$ and $q^c = (q_1 - q_2)/2$ (see Figure 1 for the baroclinic velocity field and streamfunction).

In this paper, we consider $M$ randomly located observations of the large-scale (barotropic) streamfunction corrupted by i.i.d. Gaussian noise with mean zero and variance $r^o$ chosen to correspond to approximately 10% and 25% of the climatological barotropic streamfunction variance, $E = \text{Var}(\psi^b) \approx 69.46$. The choice of observation network here is to mimic the practical situation where observations of the small-scale (baroclinic) streamfunctions are typically not available in the ocean and observations of the large-scale (barotropic) streamfunctions are sparse. This setup makes this test problem difficult for filtering since the dynamics of the observable variable (the
barotropic streamfunction),
\[
\frac{\partial q^b}{\partial t} + J(\psi^b, q^b) + \beta \frac{\partial \psi^b}{\partial x} + \kappa \nabla^8 q^b
+ \left[ J(\psi^c, q^c) + U \frac{\partial \nabla^2 \psi^c}{\partial x} - \kappa \nabla^2 \psi^c \right] = 0,
\]
are numerically stiff. The stiffness is in resolving the baroclinic terms in the square brackets in (8).

3. Hierarchical Bayesian Analysis for Filtering Irregularly spaced sparse observations

As mentioned in the introduction, the proposed hierarchical Bayesian filtering in (3) consists of two steps: the first step is to interpolate \( \tilde{v} \) to \( v \) through an appropriate interpolation scheme and obtain \( P(v|u, \tilde{v}) \); the second step is to apply \( P(u)P(v|u, \tilde{v}) \) through a reduced stochastic filtering algorithm. In this paper, we will compare two different interpolations in the first step, a deterministic linear interpolation and a geostatistical interpolation technique called ordinary kriging.

Let \( \tilde{s}_i = (\tilde{x}_i, \tilde{y}_i), i = 1, ..., M \) denote the irregularly spaced nodes on a \([0, 2\pi] \times [0, 2\pi]\) periodic domain and let \( s_{i,j} = (x_i, y_j), i, j = 1, ..., N \) with
\[
x_i = C + \frac{2\pi i h}{N} \quad y_j = C + \frac{2\pi j h}{N},
\]
with a constant \( C \) and distance \( h \) be the regularly spaced grid points. In Sections 3.1 and 3.2, we discuss two schemes that interpolate \( Z(\tilde{s}_i) \), the noisy barotropic streamfunction,
\[
Z(\tilde{s}_i) = \psi^b(\tilde{s}_i) + \tilde{\sigma}_i^o, \quad \tilde{\sigma}_i^o \sim \mathcal{N}(0, \sigma_o),
\]
where $\psi^b = (\psi_1 + \psi_2)/2$ is a solution of (5)-(6), into $Z(s_{ij})$. Subsequently, we discuss the reduced filtering approach in Section 3.3.

3.1. Linear Interpolation

There are non-unique criteria to interpolate two-dimensional noisy data at irregularly spaced observation networks. In this particular setup, finding an optimal interpolator, even with a specific choice of interpolation function, is non trivial; this depends on how many and which observations (or data points) to use. In this paper, we choose to use three observational points to interpolate at each regular grid, for simplicity. These three are selected so that their distances from the interpolation grid point are minimal and that they form a triangle with the grid point in their interior. Since our criteria are ad-hocly chosen, we expect that the linear interpolation results in Section 4 will not be robust when the criteria are changed.

Suppose we interpolate $Z(s_{ij})$ and denote the three irregularly spaced nodes $\tilde{s}_m$, $\tilde{s}_n$, and $\tilde{s}_p$, chosen based on the above criteria. Then

$$Z(x_i, y_j) = L_m(x_i, y_j)Z(\tilde{s}_m) + L_n(x_i, y_j)Z(\tilde{s}_n) + L_p(x_i, y_j)Z(\tilde{s}_p),$$

where

$$L_m(x, y) = \frac{(\tilde{y}_n - y)(\tilde{x}_n - \tilde{x}_p) - (\tilde{x}_n - x)(\tilde{y}_n - \tilde{y}_p)}{(\tilde{y}_n - \tilde{y}_m)(\tilde{x}_n - \tilde{x}_p) - (\tilde{x}_n - \tilde{x}_m)(\tilde{y}_n - \tilde{y}_p)}$$

is a Lagrange function about three points. The other two Lagrange functions $L_n(x, y)$ and $L_p(x, y)$ are defined similarly.

Next we compute the error covariance matrix in Fourier space associated with this technique to quantify the uncertainty in this method. For an interpolated observation $z_{j,m} = Z(x_j, y_m)$, denote the three corresponding
irregular observations as $\tilde{z}_{j,m,1}$, $\tilde{z}_{j,m,2}$ and $\tilde{z}_{j,m,3}$ so that

$$z_{j,m} = L_1(x_j, y_m)\tilde{z}_{j,m,1} + L_2(x_j, y_m)\tilde{z}_{j,m,2} + L_3(x_j, y_m)\tilde{z}_{j,m,3},$$

and define $C_{j,m,s} \equiv L_s(x_j, y_m)$, to simplify the notation. Its discrete Fourier transform is given by

$$\hat{z}_{k,l} = \frac{h^2}{4\pi^2} \sum_{j=1}^{N} \sum_{m=1}^{N} z_{j,m} e^{-ikj h} e^{-ilmh},$$

$$= \frac{h^2}{4\pi^2} \sum_{j=1}^{N} \sum_{m=1}^{N} (C_{j,m,1}\tilde{z}_{j,m,1} + C_{j,m,2}\tilde{z}_{j,m,2} + C_{j,m,3}\tilde{z}_{j,m,3}) e^{-i(kj+lm)h},$$

where $Nh = 2\pi$, $x_j = j h$, $y_m = m h$. Furthermore, let $\hat{\sigma}_{k,l}$ be the $(k,l)$-Fourier component of the interpolated noise $\sigma_{j,m}$ and let $\{\tilde{\sigma}_{j,m,s}\}_{s=1,2,3}$ be the noises associated with the three closest observations that encompass $\sigma_{j,m}$. Then the cross-covariance of the noise between wavenumbers $(k,l)$ and $(k',l')$ is given by

$$\langle \hat{\sigma}_{k,l}^{*} (\hat{\sigma}_{k',l'}) \rangle = \left( \frac{h^2}{4\pi^2} \sum_{j=1}^{N} \sum_{m=1}^{N} \sum_{s=1}^{3} C_{j,m,1}\tilde{\sigma}_{j,m,s} e^{-i(kj+lm)h} \right) \times \left( \frac{h^2}{4\pi^2} \sum_{j'=1}^{N} \sum_{m'=1}^{N} \sum_{s=1}^{3} C_{j',m',s}\tilde{\sigma}_{j',m',s} e^{i(k'j'+l'm')h} \right),$$

$$= \left( \frac{h^4}{16\pi^4} \sum_{j,m=1}^{N} \sum_{s=1}^{3} C_{j,m,s}\tilde{\sigma}_{j,m,s} e^{-i(kj+lm)h} \right) \times \left( \sum_{j',m'=1}^{N} \sum_{s=1}^{3} C_{j',m',s}\tilde{\sigma}_{j',m',s} e^{i(k'j'+l'm')h} \right),$$

where $\langle \bullet \rangle$ denotes expectation with respect to the Gaussian noise distribu-
The diagonal terms (where $k = k'$ and $l = l'$), simplify to

$$\langle |\sigma_{k,l}^2| \rangle = \frac{r^o h^4}{16 \pi^4} \sum_{j,m=1}^{N} \sum_{j',m'=1}^{N} \sum_{s,s'=1}^{3} C_{j,m,s} C_{j',m',s'} e^{-i[k(j-j')+l(m-m')]} h$$

$$= \frac{r^o h^4}{16 \pi^4} \left( \sum_{j,m=1}^{N} \sum_{s=1}^{3} C_{j,m,s}^2 \right)$$

$$+ \left[ \sum_{j,m=1}^{N} \sum_{j',m'=1}^{N} \sum_{s,s'=1}^{3} C_{j,m,s} C_{j',m',s'} e^{-i[k(j-j')+l(m-m')]} \right]$$

where $r^o = \langle \tilde{\sigma}_{j,m,s} \tilde{\sigma}_{j',m',s'} \rangle$. The $C_{j,m,s}^2$ terms outside the square bracket reflect the contributions of the diagonal terms of the physical space covariance matrix while the terms inside the square bracket reflect the contributions of the off-diagonal terms of the physical space covariance.

In Figure 2, we show a linear interpolation of irregularly spaced, sparse observations, with locations represented by circles, at an arbitrary time $T = 340$. In this example and throughout this paper, we will interpolate to $N = 6$ grid points on each axis. Notice the relatively good agreement between the interpolated observations and the truth, both resolved at $N \times N = 36$ coarse grid points. The noise covariance matrix, illustrated in the top left panel of Figure 3 in physical space, is diagonally dominant with nearby nonzero correlations as expected, due to interpolation. Notice that the error covariance matrix depends on observations locations rather than the observations themselves. In Fourier space, the error covariance matrix of the interpolated observations is diagonally dominant (see the bottom left panel of Figure 3 for the real component; the imaginary component has one order magnitude smaller than the real component).
3.2. Ordinary Kriging

Generally speaking, kriging is a maximum likelihood estimator of a random field $Z$ modeled by

$$Z(s) = \mu(s) + \delta(s),$$

assuming the noises $\delta(s) \sim \mathcal{N}(0, C(s, s))$ are Gaussian and stationary processes. The key idea in kriging is to fit the observations, $\vec{Z} = [Z(\tilde{s}_1), Z(\tilde{s}_2), ..., Z(\tilde{s}_M)]^T$, to an empirically chosen isotropic parametric covariance function $C$. By isotropic, we mean $C(s_1, s_2) = C(\|s_1 - s_2\|)$; this equality holds under the stationary assumption. Note that each component in $\vec{Z}$ is exactly defined in (9).

Computationally, we employ a nonlinear optimization algorithm to obtain the parameters in $C$ on-the-fly. This is typically a low-dimensional nonlinear optimization problem; in our particular numerical experiments below, we will choose $C$ from an exponential family with two parameters. By on-the-fly, we mean that these parameters change when different sets of observations $Z$ become available at different instances. In this sense, we called kriging a data-driven interpolation scheme even when the observation locations are fixed; in contrast, the Lagrange functions for the linear interpolation are constant for fixed observation network. Various types of kriging model $\mu(s)$ differently; in our application below, we consider ordinary kriging, which assumes that $\mu(s)$ is locally constant. Mathematically, the estimator for $Z$ at location $s_{i,j}$ is given by the conditional expectation, $\mathbb{E}(Z(s_{i,j})|\vec{Z})$. The uncertainty of
this estimator is given by the conditional covariance, \( \text{Cov}(Z(s_{i,j}), Z(s'_{i',j'})|\mathbf{Z}) \).

The detailed expressions of these conditional statistics are given in the Appendix (see Eqns (21) and (24)). In the remaining of this section, we will discuss how to model the covariance function \( C \), as well as mean \( \mu \).

Let us define variogram by a functional \( 2\gamma \) that satisfies the relation

\[
2\gamma(\tilde{s}_i - \tilde{s}_j) \equiv \text{Var}(Z(\tilde{s}_i) - Z(\tilde{s}_j)) = \langle [Z(\tilde{s}_i) - Z(\tilde{s}_j)]^2 \rangle,
\]

[10], to characterize the spatial dependence of the process \( Z \). Expanding the right hand side of (10) gives

\[
2\gamma(\tilde{s}_i - \tilde{s}_j) = \text{Var}(Z(\tilde{s}_i)) + \text{Var}(Z(\tilde{s}_j)) - 2\text{Cov}(Z(\tilde{s}_i), Z(\tilde{s}_j))
\]

\[
= 2(r^o - C(\|\tilde{s}_i - \tilde{s}_j\|)),
\]

where we use \( r^o = C(0) \) and the stationary assumption. This yields a useful relation between covariance and variogram:

\[
C(\tilde{s}_i, \tilde{s}_j) = C(\|\tilde{s}_i - \tilde{s}_j\|) = r^o - \gamma(\tilde{s}_i - \tilde{s}_j).
\]

While the exact variogram \( \gamma \) is unknown, we estimate it using the irregularly spaced observations and therefore get an estimate of the covariance structure. To construct a variogram estimator, denoted \( \hat{\gamma}(r) \), we follow the description in [20]. The algorithm begins by placing observations into bins based on distance from one another. A bin consists of all pairs of observations which are within distance \( r \) of each other, where \( r \) is a relatively small distance. Then for each bin, which is defined entirely by \( r \), compute

\[
2\hat{\gamma}(r) = \frac{1}{|N(r)|} \sum_{i,j \in N(r)} (R(\tilde{s}_i) - R(\tilde{s}_j))^2,
\]

14
where $R(\tilde{s}_i) = Z(\tilde{s}_i) - \mu(\tilde{s}_i)$ represents the residual and $N(r)$ is the size of the bin. This equation defines the variogram estimator and demonstrates the fact that kriging is data-driven; subsequently, the covariance estimator in (11) depends on observations $Z(\tilde{s})$. To compute the residual $R(\tilde{s}_i)$ we need to estimate the mean, $\mu(\tilde{s}_i)$. In our numerical implementation, we estimate $\mu(\tilde{s}_i)$ with median polishing, an algorithm that estimates the mean value at an observation location by averaging the observations in the same mesh row and column.

We construct bins for several distances $r$ so that we can estimate $\gamma(\tilde{s}_i - \tilde{s}_j)$ in (11) for numerous values $\|\tilde{s}_i - \tilde{s}_j\|$. Based on the shape of the variogram estimator as a function of $r$, we choose an appropriate parametric function $\hat{\gamma}^*(r)$ to model $\hat{\gamma}(r)$. An example plot of a variogram fit at time $T = 340$ is given in Figure 4. The values of $\hat{\gamma}(r)$, depicted as circles in Figure 4, appear to follow an exponential curve. Therefore, we choose $\gamma$ from an exponential family,

$$\hat{\gamma}^*(r) = \sigma^2 \exp(-\rho r), \quad r \geq 0,$$

(12)

where $\rho, \sigma > 0$ are the two parameters to be determined. Note that there are many other parametric forms that can be used besides the exponential family [20]; the appropriate form is dictated by the data (shape of the variogram estimator). The solid line in Figure 4 gives the least squares fit of (12) with respect to parameters $\sigma$ and $\rho$ to $\hat{\gamma}(r)$ at $T = 340$. The corresponding parameters of the residual process at this instance are $\sigma = 0.7325$ and $\rho = 1.3019$. The covariance $C$ is estimated by

$$C(\tilde{s}_i, \tilde{s}_j) = r^0 - \sigma^2 \exp(-\rho \|\tilde{s}_i - \tilde{s}_j\|).$$

In Figure 2 we show a kriging result at time $T = 340$. The right panels
in Figure 3 show the associated error covariance matrix for the kriging result in Figure 2, in physical space and Fourier space. Notice that, by design, the error covariance matrix (in Fourier space) is diagonal due to the stationarity of the covariance estimator with isotropic function \( C \). The slight nonzero non-diagonal terms in the Fourier domain in Figure 3 are numerical artifacts. In contrast to the linear interpolation scheme, the error covariance matrix obtained from kriging changes in time even when the observation locations are fixed because the term \( C(\tilde{s}_i, \tilde{s}_j) \) in (11) depends on a variogram estimator that varies based on observations \( Z(\tilde{s}_j) \). In the atmospheric and oceanic data assimilation community, ordinary kriging is comparable to 3D-VAR or optimal interpolation; normally 3D-VAR assumes a constant prior error covariance matrix \( C \) (e.g., see chapter 5 of [22]) and zero correlation between observations and prior state, \( c(s_{i,j}) = 0 \).

3.3. Reduced Stochastic Filter

The second step in the hierarchical Bayesian framework in (3) is to apply the reduced stochastic filter. In this paper, we report the numerical simulations with only the simplest reduced stochastic model, the Mean Stochastic Model (MSM) [12]. Below, we will review the construction of MSM for the dynamics of the barotropic mode in (8). Following standard closure modeling for turbulent systems [3, 4, 5, 6, 7, 8], we replace the baroclinic terms and the nonlinear advective terms in (8) with stochastic noise and linear damping terms, such that each Fourier component can be written as

\[
d\hat{\psi}_{k,l}(t) = (-d_{k,l} + i\omega_{k,l})\hat{\psi}_{k,l}(t)dt + \bar{\psi}_{k,l}dt + \sigma_{k,l}dW_{k,l}(t).
\]  

(13)
In (13), $\hat{\psi}_{k,l}$ denotes the horizontal Fourier component of the barotropic streamfunction $\psi^b$; $W_{k,l}(t) = (W_1(t) + iW_2(t))/\sqrt{2}$ denotes a complex-valued Wiener process with independent standard Wiener processes, $W_1$ and $W_2$. There are four parameters to be determined on each mode $(k,l)$: damping $d$, frequency $\omega$, constant external forcing $\bar{\psi}$, and noise strength $\sigma$. Following MSM [1], we set the constant forcing equal to the time average of its associated Fourier component and fit the other three parameters statistically to the climatological energy spectrum (see Figure 5) and to the real and imaginary components of the correlation time. In particular, we use the explicit statistical solutions of the forced Ornstein-Uhlenbeck process in (13),

$$E_{k,l} = \lim_{t \to \infty} \langle (\hat{\psi}_{k,l}(t) - \bar{\psi}_{k,l})(\hat{\psi}_{k,l}(t) - \bar{\psi}_{k,l})^* \rangle = \frac{\sigma_{k,l}^2}{2d_{k,l}},$$

(14)

$$T_{k,l} - i\theta_{k,l} = \lim_{t \to \infty} \int_0^\infty \langle (\hat{\psi}_{k,l}(t + \tau) - \bar{\psi}_{k,l})E_{k,l}^{-1}(\hat{\psi}_{k,l}(t) - \bar{\psi}_{k,l})^* \rangle d\tau = \frac{1}{d_{k,l} + i\omega_{k,l}},$$

(15)

to obtain $d_{k,l}$, $\omega_{k,l}$ and $\sigma_{k,l}$ with $E_{k,l}$, $T_{k,l}$ and $\theta_{k,l}$ empirically estimated from solutions of the true model in (5)-(6). In real applications, we may not have a true model, but typically there is a training data set (e.g., reanalysis in geophysical applications).

In Figure 5, we show the empirical distribution of the climatological energy, $E_{k,l}$. The horizontal axis in Figure 5 corresponds to the barotropic modes, ordered according to empirical orthogonal functions (EOF) using the barotropic streamfunction variance; that is, from the largest to the smallest variance, averaged over a long period of time [11]. The first twelve modes are ordered as $(k,l)$ = (1,0), (0,1), (1,1), (-1,1), (0,2), (2,0), (2,1), (-2,1), (1,2), (-1,2), (-2,2) and (2,2). The large-scale zonal jet modes carry the second and
fifth largest variances and the Rossby mode (1,0) has the largest variance. The magnitude of the variances of the first two modes is on the same order, which indicates competition between two distinct regimes, Rossby waves and zonal jets (see also Figure 1). The marginal pdf of the solutions of the first two modes are shown in Figure 6 whereas the remaining marginal pdf’s have Gaussian shape (see [11]). These marginal pdfs are generated through bin counting the barotropic streamfunction, centered at 0, such that both panels in Figure 6 show a histogram \( d\psi = \hat{\psi}^b - \langle \hat{\psi}^b \rangle \), and encompass solutions of (8) up to \( T = 400 \) time units.

Harlim and Majda [11] considered regularly spaced sparse observations of the barotropic streamfunction on \( N \times N \) grid points where \( N = 6 \) (much less than the truth model, which is resolved with \( N = 128 \)). In this paper, we consider \( M \) irregularly spaced sparse observations of the barotropic streamfunction through the observation model in (9). The first step in the hierarchical Bayesian analysis in (3) produces the following conditional statistics to be filtered, \( \mathbb{E}(Z(s_{i,j})|\vec{Z}) \) and \( \text{Cov}(Z(s_{i,j})|\vec{Z}) \) at \( i, j = 1, \ldots, N \) (see Sections 3.1 and 3.2), where we use \( N = 6 \), following [11]. With this minimal resolution, we at least resolve the twelve most energetic modes (see Figure 5). In Figure 7, we show the decaying time with \( d_{k,l} \), estimated through solving (14)-(15); based on this decaying time, we simulate the observation time intervals to be \( T_{\text{obs}} = 0.01, 0.04 \) and 0.08, which are shorter than the model damping times on modes 1-12.

The discrete-time reduced filtering model for the second-step in the hier-
archival Bayesian analysis in (3) is defined as follows:

\[
\hat{\psi}_{k,l,m+1} = F_{k,l} \hat{\psi}_{k,l,m} + \tilde{f}_{k,l,m+1} + \eta_{k,l,m+1}, \tag{16}
\]

\[
\hat{\psi}_{k,l,m} = \tilde{\psi}_{k,l,m} + \epsilon_{k,l,m}. \tag{17}
\]

where subscripts \( k, l \) denote the two-dimensional wavenumbers and subscript \( m \) denotes the discrete time step with \( t_{m+1} - t_m = T_{obs} \). The parameters in (16) are given by exact solution of the forced Langevin equation in (13) with

\[
F_{k,l} = e^{(-d_{k,l} + iw_{k,l})T_{obs}},
\]

\[
\tilde{f}_{k,l,m+1} = \tilde{\psi}_{k,l,m}(1 - F_{k,l}),
\]

\[
\eta_{k,l,m+1} \sim \mathcal{N}(0, Q_{k,l}), \quad Q_{k,l} = \frac{\sigma_{k,l}^2}{2d_{k,l}} (1 - e^{-2d_{k,l}T_{obs}}).
\]

The observation model in (17) is defined with \( \hat{\psi}_{k,l,m} \) denoting the Fourier coefficients of the conditional estimate \( \mathbb{E}(Z_m(s_{i,j}) | \vec{Z}_m) \) and Gaussian noises \( \epsilon_{k,l,m}^o \sim \mathcal{N}(0, \hat{R}_{k,l}^o) \), where \( \hat{R}_{k,l}^o \) denotes the Fourier representation of the error covariance matrix \( \text{cov}(Z(s_{i,j}) | \vec{Z}) \). The linear observation model in (17) defines the connection to the dynamical equation in (16) with \( P(Z_m(s_{i,j}) | \vec{Z}_m) = P(Z_m(s_{i,j}) | \vec{Z}_m, \hat{\psi}_{k,l,m}) \), where \( Z_m(s_{i,j}) = \psi, \vec{Z}_m = \vec{v}, \) and \( \hat{\psi}_{k,l,m} = u \), in accordance to the notation in (3). The major computational reduction in the proposed filtering approach in (16)-(17) is through the tacit assumption that different modes are statistically uncorrelated (see the nearly diagonal observation error covariance matrix shown in Figure 3). This assumption only holds for appropriate interpolation schemes as investigated in detail in [2] on a simpler set-up.

The basic Kalman filter solution to (16)-(17) produces scalar estimates of the mean and covariance prior and posterior to observation \( \hat{\psi}_{k,l,m+1} \) through
the following recursive equations:

\[
\begin{align*}
\tilde{\psi}_{k,l,m+1} &= F_{k,l} \tilde{\psi}_{k,l,m} + \tilde{f}_{k,l,m+1}, \\
R_{k,l,m+1} &= F_{k,l} R_{k,l,m} F_{k,l}^* + Q_{k,l}, \\
\tilde{\psi}_{k,l,m+1}^+ &= \tilde{\psi}_{k,l,m+1} + K_{k,l,m+1} (\hat{\psi}_{k,l,m+1}^o - \tilde{\psi}_{k,l,m+1}), \\
R_{k,l,m+1}^+ &= (I - K_{k,l,m+1}) R_{k,l,m}, \\
K_{k,l,m+1} &= R_{k,l,m+1}^- (R_{k,l,m+1}^- + R_{k,l,m}^o)^{-1}.
\end{align*}
\]  

(18)

In (18), we denote the prior mean state and covariance by \(\tilde{\psi}_{k,l,m}\) and \(R_{k,l,m}^-\), and the posterior mean state and covariance by \(\tilde{\psi}_{k,l,m}^+\) and \(R_{k,l,m}^+\) respectively.

In this paper, we only apply the recursive filter on the twelve most energetic modes to resolve \(N \times N\) grid points, where \(N = 6\).

4. Filtering Results

In this section, we report the numerical results of applying the hierarchical Bayesian filtering strategy in (3) in assimilating irregularly spaced observations of barotropic streamfunction in (9) to four different cases, varying the number of observations, \(M\), and observation error variance, \(r^o\). In each of these four cases, we run the filtering experiment on 50 different randomly chosen, irregularly spaced, observation networks and compare the filtering skill of the Kalman filter recursive algorithm in (18) in assimilating the interpolated observations obtained either by the linear interpolation (see Section 3.1) or ordinary kriging (Section 3.2). In the remainder of this paper, we refer to these two methods as the filtered linear interpolation scheme and the filtered kriging scheme. To measure the filtering skill, we compute the
average Root-Mean-Square error,
\[ \mathcal{E} = \left( \frac{1}{N^2(T - T_o)} \sum_{m=T_o}^{T} \sum_{i,j=1}^{N} (\dddot{\psi}_{i,j,m}^+ - \psi_{m}^b(s_{i,j}))^2 \right)^{1/2}, \tag{19} \]
and the time average physical space pattern correlation,
\[ \mathcal{C} = \frac{1}{T - T_o} \sum_{m=T_o}^{T} \frac{(\dddot{\psi}_{m}^+)^T \psi_{m}^b}{\|\dddot{\psi}_{m}^+\|_2 \|\psi_{m}^b\|_2}, \tag{20} \]
between the mean posterior state, \( \dddot{\psi}_{i,j,m}^+ \), and the true barotropic streamfunction, \( \psi_{m}^b(s_{i,j}) \). Here \( T \) is the number of time steps and we ignore the first \( T_o \) steps. The total assimilation step is defined as \( T = T/T_{obs} \), where we set \( T = 400 \) time units. For \( T_{obs} = 0.01 \), the average is over \( T = 40,000 \) steps, ignoring the first \( T_o = 1000 \) steps; we also consider \( T_{obs} = 0.04 \), in which case \( T_o = 200 \) and \( T_{obs} = 0.08 \), in which case \( T_o = 100 \). To check the effectiveness of the second step in the proposed hierarchical Bayesian filtering scheme, we also report the estimates of the interpolated observations without implementing stochastic recursive filter (that is, only apply either the linear interpolation or kriging). In the remainder of this paper, we refer to these two methods as the unfiltered linear interpolated scheme and the unfiltered kriging scheme. For these two schemes, the RMS error and pattern correlation between \( \mathbb{E}(Z_m(s_{i,j})|\tilde{Z}_m) \) (instead of \( \dddot{\psi}_{m}^+ \)) and \( \psi_{m}^b \) are computed with (19) and (20), respectively. To demonstrate the robustness of the filtering skill, we show the averages of the RMS error in (19) and pattern correlation in (20) over 50 different randomly chosen, irregularly spaced, observation networks (see Figure 8). For diagnostic purpose, we also include the filtering skill for regularly spaced observations at \( N \times N = 36 \) model grid points for observations noise variance \( r^o = 10\%E \) and \( r^o = 25\%E \) (see Figure 9).
4.1. Case 1: $M = N^2 = 36$, $r^o = 6.9 = 10\% E$.

In this first case, we consider observation networks with $M = N^2 = 36$, where $M$ denotes the total number of observations and $N$ denotes the model resolution in each coordinate axis. Here, we consider observation noise variance of only 10% of the climatological (temporal averaged) barotropic streamfunction variance $E \approx 69.64$. In this scenario, ordinary kriging supersedes the linear interpolation scheme; see the first row of Figure 8 that displays the RMS error and pattern correlation as functions of observation time interval, $T_{obs}$. The two skill measures in (19) and (20) suggest that the estimates are approximately constant as $T_{obs}$ increases. The dynamical constraint produces improved estimates, by only 0.1 in RMS measure and 2-3% in pattern correlation measure, compared to the unfiltered estimates. The filtering skills for these methods are much worse than those corresponding to the regularly spaced observations exactly at the $N \times N = 36$ model grid points (see the much lower RMS errors and higher correlation in the top panels of Fig 9). In Figure 10 we show the true barotropic streamfunctions and the corresponding estimates at times $T = 230.72$ and $T = 363.28$ using a randomly chosen observation network (in circles). Comparing the estimates with the truth, we find that filtered kriging clearly produces a much better estimate compared to the linear interpolation scheme despite enormous scale variation.

4.2. Case 2: $M = 18 < N^2 = 36$, $r^o = 6.9$.

In this case, we consider fewer observations but subject to the same noise variance as in Case 1 above. As we expected, the sparse observation network degrades the accuracy of both methods compared to the previous case (see the higher average RMS errors and lower average pattern correlations
in the second row of Figure 8). In this case, the sparse observation networks allow degenerate Lagrange functions for linear interpolation as well as poor variogram fitting in the covariance estimation in (11). Based on the two measures in (19) and (20), averaged over 50 different observation networks, we find that, overall, both schemes are comparable. We find that the linear interpolation (implemented with the criteria discussed in Section 3.1) can sometimes provide a better skill relative to kriging, depending on the distribution of the observations. However, the high filtering skill with linear interpolation estimates in two-dimensional domain is not robust since there are non-unique criteria for constructing the interpolating Lagrange functions. From our numerical simulations, we also find that the filtered estimates are better than the unfiltered estimates independent of the observation times. Here, the RMS error and pattern correlation skill measures are consistently improved by about 0.1 and 2-3%, respectively, when the dynamical constraint is imposed. The numerical result in this regime suggests that the data-driven step is not effective since the data is too sparse.

4.3. Case 3: \( M = 49 > N^2 = 36 \), \( r^o = 6.9 \).

Now we consider more observations compared to the filtered model resolution. In this case, we find that more observations does not improve the linear interpolation by much. Comparing the first and third rows in Figure 8), we see that the average RMS error decreases from 3.7 to 3.4. For the same scenario, the kriging estimates are improved significantly in the presence of more observations (the average RMS error decreases from 3.3 to 2.6). Here, the filtering skills are still worse than those corresponding to the regularly spaced observations exactly at the \( N \times N = 36 \) model grid.
points (see Figure 9) but there is clear improvement relative to the Case 1 above. In Figure 11 we show the true barotropic streamfunctions and the corresponding estimates at times $T = 230.72$ and $T = 363.28$ (similar to those in Figure 10) using a randomly chosen observation network with more observations, $M = 49$ (in circles). Notice that both kriging and the linear interpolation are improved with more observations (compare Figures 10 and 11). Qualitatively, we can see that kriging produces much better estimates of the true field compared to the linear interpolation scheme.

4.4. Case 4: $M = N^2 = 36$, $r_o = 17.3$.

In the final case, we consider a larger observation noise variance, $r_o$ is about 25% of $E$. In this case, the average RMS errors of both schemes increase significantly compared to Case 1 with smaller observation noise variance. The poor estimates with the unfiltered linear interpolation and kriging are not so surprising since the given data is very noisy. Notice that in this case, the average RMS errors and pattern correlation for the filtered estimates are improved by about 0.2 and 4%, respectively, compared to those of the unfiltered estimates. This numerical result suggests that the dynamical constraint becomes more important when observations are noisy.

Now let us closely examine the filtering results. In Figure 12, we show the RMS errors of the filtered solutions (with linear interpolation on the top panel and kriging on the bottom panel) as functions of time for observation time interval $T_{obs} = 0.01$ and Case 4 ($M = N^2 = 36$ and $r_o = 17.3$). To understand why the errors oscillate as functions of time, we show the true streamfunction at time $T = 61$, when the RMS error is at a minimum, and at the time $T = 363.28$, when the RMS error is at a maximum (see Figures 13
and 14). As a diagnostic, we also show the estimates obtained with regularly spaced observations at the 36 model grid points (top right panel). Notice that there is a zonal jet at both instances, but most importantly, the contour scaling of $\psi^b$ varies between $\pm 6$ at time $T = 61$ and between $\pm 20$ at $T = 363.28$ and the filter estimate at time $T = 363.28$ is much more accurate compared to the filter estimate at time $T = 61$. This result suggests that the filter estimates are quite accurate when the signal-to-noise ratio (between the truth signals and the observation noises) are large (e.g., at time $T = 363.28$). On the other hand, when the signal-to-noise ration is small (for e.g., at time $T = 61$), the estimates are less accurate. In this case, one needs less noisy observations to improve the estimates.

5. Summary

In this paper, we investigate a hierarchical Bayesian approach for filtering irregularly spaced, sparse observations of geophysical turbulence. In particular, we blend a data-driven interpolation scheme (through either ordinary kriging or a deterministic piecewise linear interpolation scheme) with a reduced stochastic filtering approach (the Fourier domain Mean Stochastic Model filter). The two-step hierarchical Bayesian approach proposed here interpolates raw data at irregularly spaced locations to regularly spaced filter model grid points and then assimilates this processed data set with the basic Kalman filter scheme on a diagonal Fourier domain mean stochastic model. We find that the dynamical constraint (through MSM) becomes more important in this estimation approach when the observation noise variance is large. Second, when the observation observation network is very sparse, both
the linear interpolation and kriging produce comparable posterior estimates. Third, the two steps hierarchical Bayesian approach with kriging produces much improved filtered solutions when observation network is denser in space.

The filtered estimates with irregularly spaced observation networks are worse than the estimates with regularly spaced observation network (exactly at the filter model $N^2 = 36$ grid points). As we pointed out above, we need denser observations to improve the interpolated estimates in the first step of the proposed hierarchical Bayesian framework. For example, in the case of $T_{obs} = 0.01$ and $r^o = 17.3$ (Case 4 in Section 4.4), the RMS error for the filtered kriging estimates with the irregularly spaced observations is about 3.6, much worse than the RMS error of the estimates with regularly spaced observations, 0.96. On the other hand, we should also mention that this case was considered in [11] with regularly spaced sparse observations at exactly the filter model $N^2 = 36$ grid points; in there, they reported filtered estimates with much higher average RMS error obtained from a perfect model simulation with Local Least Squares-Ensemble Adjustment Kalman Filter (LLS-EAKF) scheme, 6.76 (see Table 3 of [11]); we ignore reporting LLS-EAKF for the irregularly spaced sparse observations case here since there is no reason to believe that LLS-EAKF, which is implemented sequentially and locally [14, 23], can produce improved filtered solutions when sparse observation location is changed from regularly to irregularly spaced. Comparing these numerical results, we conclude that the hierarchical Bayesian approach is much better compared to the standard ensemble filtering approach that directly filters at the observation locations, however, we anticipate that much improved estimates with this framework can be achieved with the strategies
discussed below.

The encouraging results in this paper are obtained despite various tacit assumptions for practical computational consideration. The first assumption is the Gaussianity of the spatial distribution of the data which may not be true for different applications. In such cases, one may want to consider statistical interpolation schemes with appropriate distributions other than Gaussian distribution. For geostatistical data, however, kriging seems to be quite successful in practice [20]. Second, note that higher order deterministic interpolation schemes will easily introduce large error correlations between different interpolated locations [2]. In there, it is found that the linear interpolation for a one-dimensional domain produces an interpolated error covariance matrix that is nearly diagonal in Fourier domain (see also Figure 4 in this manuscript for the two-dimensional case). Furthermore, it was shown that the approximate MSM filter that ignores these non-diagonal covariance terms produces better estimates compared to those that account these non-diagonal covariance terms. In contrast, kriging, by design, produces a Fourier diagonal interpolated error covariance matrix. This is because kriging assumes stationarity and uses isotropic covariance models. As a consequence, the second step in our proposed hierarchical Bayesian approach can be performed without any additional approximation when the Fourier diagonal reduced stochastic models such as MSM are used with kriging. Third, we expect that the results with linear interpolation scheme are not robust on two-dimensional domain since there are non-unique criteria for constructing Lagrange interpolating functions. Similarly, the kriging results in this paper are based on a specific implementation of ordinary kriging with covariance
model from an exponential family in (12); we suspect that improved estimates can be obtained with different classes of function or different types of kriging [20].

In Section 4, we show numerical results using the simplest reduced stochastic models in the second step, in particular, the MSM-filter model that ignores correlations between different Fourier modes. With such a poorman’s model, we find that the dynamics help improve the filtering skill by about 0.2 in terms of RMS error (or 4% in terms of correlation skill) when observation noise is 25% of the climatological variance. In the future, we will consider replacing MSM with models that account for correlations between appropriate Fourier modes such as the super-resolution stochastic filters [25] and the more flexible Gaussian closure filter [26].

Acknowledgement

The research of J.H. is partially supported by the ONR Grant N00014-11-1-0310, ONR MURI Grant N00014-12-1-0912, the NC State startup fund, and the NC State Faculty Research and Professional Development fund. K.B. is partially supported by Harlim’s NC State startup fund and the NSF funded Research for Early Graduate Student program DMS-0943855. The authors thank Emily L. Kang for bringing the kriging method to their attention. The authors also thank A.J. Majda for his comments and suggestions.

Appendix

Kriging is a statistical method of interpolation that models $Z$ at position $s$ with $Z(s) = \mu(s) + \delta(s)$, where the noises, $\delta(s)$, are assumed to be Gaus-
sian and stationary processes; ordinary kriging assumes that \( \mu(s) \) is constant locally [20]. Given observations,

\[
\vec{Z} = [Z(\tilde{s}_1), Z(\tilde{s}_2), \ldots, Z(\tilde{s}_M)]^T,
\]

the joint distribution \((Z(s_{i,j}), \vec{Z})\) is expressed by:

\[
(Z(s_{i,j}), \vec{Z}) \sim \mathcal{N}\left(\begin{pmatrix} \mu(s_{i,j}), \vec{\mu} \end{pmatrix}, \begin{bmatrix} C(s_{i,j}, s_{i,j}) & c(s_{i,j}) \\ c(s_{i,j}) & \Sigma \end{bmatrix}\right),
\]

where \( C \) denotes the covariance function and

\[
c(s_{i,j}) = [C(s_{i,j}, \tilde{s}_1), C(s_{i,j}, \tilde{s}_2), \ldots, C(s_{i,j}, \tilde{s}_M)], \quad \Sigma_{i,j} = C(\tilde{s}_i, \tilde{s}_j).
\]

In general, kriging estimates \( Z \) at \( s_{i,j} \) with the following conditional expectation,

\[
\mathbb{E}(Z(s_{i,j})|\vec{Z}) \equiv \mu(s_{i,j}) + c(s_{i,j})^T \Sigma^{-1}(\vec{Z} - \vec{\mu})
\]

\[
= c(s_{i,j})^T \Sigma^{-1} \vec{Z} + \mu(1 - c(s_{i,j})^T \Sigma^{-1} \vec{1})
\]

\[
= \vec{\lambda}^T \vec{Z} + \mu(1 - \vec{\lambda}^T \vec{1}),
\]

where we write \( \vec{\lambda} = \Sigma^{-1} c(s_{i,j}) \) and assume that locally, \( \mu(s_{i,j}) = \mu(\tilde{s}_k) = \mu; \) here, \( \mu(\tilde{s}_k) \) is a component of \( \vec{\mu} \). See Theorem 3.5 of [27] for detailed derivation of conditional expectation of multivariable Gaussian distribution.

The estimator above is also known as the maximum likelihood Gaussian estimator. Ordinary kriging imposes the following condition, \( \vec{\lambda}^T \vec{1} = 1 \), or

\[
\sum_{k=1}^{M} \lambda_{i,j,k} = 1,
\]

which eliminates the dependency of the estimator to the unknown locally constant \( \mu(s) \).
Mathematically, the ordinary kriging is defined with the following estimator:

\[
E(Z(s_{i,j})|\vec{Z}) = \sum_{k=1}^{M} \lambda_{i,j,k} Z(\tilde{s}_k)
\]

(21)

such that the mean square predicted error,

\[
MSPE(\vec{\lambda}) = E\left[\left(Z(s_{i,j}) - \sum_{k=1}^{M} \lambda_{i,j,k} Z(\tilde{s}_k)\right)^2\right],
\]

(22)

subject to the constraint,

\[
\sum_{k=1}^{M} \lambda_{i,j,k} = 1,
\]

is minimized.

Rewrite (22) as

\[
MSPE(\vec{\lambda}) = \text{Var}\left(\sum_{k=0}^{M} a_k Z_k\right)
\]

where \(a_0 = 1\) and \(a_k = -\lambda_{i,j,k} = -\lambda_k\) when \(k \neq 0\). Here \(Z_0 = Z(s_{i,j})\) and \(Z_k = Z(\tilde{s}_k)\) when \(k \neq 0\). Then

\[
MSPE(\vec{\lambda}) = C(Z_0, Z_0) - 2 \sum_{i=1}^{M} \lambda_i C(Z_i, Z_0) + \sum_{i,j=1}^{M} \lambda_i \lambda_j C(Z_i, Z_j).
\]

The constraint minimization problem in (22) is equivalent to solving

\[
\min_{\lambda, \alpha} \left\{ MSPE(\vec{\lambda}) + 2\alpha \left(\sum_{k=1}^{M} \lambda_k - 1\right) \right\} \equiv \min_{\lambda, \alpha} L(\vec{\lambda}, \alpha).
\]

Setting

\[
\frac{\partial L}{\partial \lambda_k} = 0, \quad \frac{\partial L}{\partial \alpha} = 0,
\]

30
we obtain
\[-2C(Z_0, Z_k) + 2 \sum_{j=1}^{M} \lambda_j C(Z_j, Z_k) + 2\alpha = 0\]
and
\[\sum_{k=1}^{M} \lambda_k - 1 = 0.\]
This can be written as a linear system:
\[
\Gamma_0 \bar{\lambda}_0 = \bar{\gamma}_0
\]
with
\[
\Gamma_0 = \begin{bmatrix}
C(\bar{s}_i, \bar{s}_j) & 1 & \cdots & 1 & 0 \\
1 & \cdots & 1 & 0 & 0 \\
\vdots & & & & \vdots \\
1 & \cdots & 1 & 0 & 0 \\
1 & \cdots & 1 & 0 & 0
\end{bmatrix}, \quad \bar{\lambda}_0 = \begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_M \\
\alpha
\end{bmatrix}, \quad \bar{\gamma}_0 = \begin{bmatrix}
C(s_{i,j}, \bar{s}_1) \\
\vdots \\
C(s_{i,j}, \bar{s}_n) \\
1
\end{bmatrix}.
\]
In our notation, the multipliers are thus given by
\[
\begin{bmatrix}
\lambda_{i,j,1} \\
\vdots \\
\lambda_{i,j,n} \\
\alpha
\end{bmatrix} = \begin{bmatrix}
C(\bar{s}_i, \bar{s}_j) & 1 & \cdots & 1 \\
1 & \cdots & 1 & 0 \\
\vdots & & & \vdots \\
1 & \cdots & 1 & 0
\end{bmatrix}^{-1} \begin{bmatrix}
C(s_{i,j}, \bar{s}_1) \\
\vdots \\
C(s_{i,j}, \bar{s}_n) \\
1
\end{bmatrix}.
\] (23)
The conditional variance is then given as
\[
MSPE(\bar{x}_0) = C(Z_0, Z_0) - 2 \sum_{i=1}^{M} \lambda_i C(Z_i, Z_0) + \sum_{i=1}^{M} \sum_{j=1}^{M} \lambda_i \lambda_j C(Z_i, Z_j)
\]
\[
= C(Z_0, Z_0) - 2 \sum_{i=1}^{M} \lambda_i C(Z_i, Z_0) + \sum_{i=1}^{M} \lambda_i [C(Z_i, Z_0) - \alpha]
\]
\[
= C(Z_0, Z_0) - \sum_{i=1}^{M} \lambda_i C(Z_i, Z_0) - \alpha
\]
\[
= C(Z_0, Z_0) - \bar{\lambda}_0^T \bar{\gamma}_0.
\]
31
Then

$$\text{Var}[Z_0 | \tilde{Z}] = C(Z_0, Z_0) - \tilde{\gamma}_0^T \Gamma_0^{-1} \tilde{\gamma}_0$$

and

$$\text{Var}[Z(s_{i,j}) | \tilde{Z}] = C(s_{i,j}, s_{i,j}) - [c(s_{i,j})^T, 1] \Gamma_0^{-1} [c(s_{i,j})^T, 1]^T.$$ 

Thus the generalized covariance is

$$\text{Cov}[Z(s_{i,j}), Z(s_{i',j'}) | \tilde{Z}] = C(s_{i,j}, s_{i',j'}) - [c(s_{i,j}), 1] \Gamma_0^{-1} [c(s_{i',j'})^T, 1]^T. \quad (24)$$

To apply ordinary kriging, we empirically choose a parametric model for the covariance function, $C(s, s)$, based on the residuals of the observations $\tilde{Z}$ (this is discussed in Section 3.2). Subsequently, we compute the multiplier by solving the linear problem in (23). The estimator and its error covariance are given by the conditional mean and covariance in (21) and (24), respectively.

References


[21] G. Vallis, Atmospheric and Oceanic Fluid Dynamics: Fundamentals and


Figure 1: The barotropic velocity field (arrows) and streamfunction, $\psi^b$, (contour) (top) and the baroclinic velocity field and streamfunction, $\psi^c$ (bottom) at times 100 and 200.
Figure 2: True process, compared with a linear interpolation and kriging at $T = 340$, with $r^o = 25\%E$ and $M = N^2 = 36$. The circles denote observation locations and the contours denote the barotropic streamfunction, $\psi^b$, on coarse mesh.
Figure 3: The interpolated observation error covariance matrices in physical space (top panels) and in Fourier space (bottom panels) at $T = 340$, with $r^o = 25\%E$ and $M = N^2 = 36$. For the bottom panels, we only show the real component since the imaginary component is one order magnitude smaller than the real component.
Figure 4: An exponential variogram fit at $T = 340$, with $r^o = 25\%E$ and $M = N^2 = 36$, plotted as a function of $r$. 
Figure 5: Percentage of variances of the barotropic streamfunction as a function of modes (wavenumbers), sorted from the largest to the smallest values.
Figure 6: Marginal pdfs of the barotropic streamfunction (centered around their means), in which solid lines indicate the real part and dashed lines the imaginary parts.
Figure 7: Damping time, $1/d$ of MSM.
Figure 8: The average RMS errors (first column) and pattern correlations (second column) for numerical experiments in Case 1-4. In each panel, we denote the unfiltered kriging with dashes, the filtered kriging with ‘+-’, the unfiltered linear interpolation with solid line, and the filtered linear interpolation with ‘o–’. 43
Figure 9: The average RMS errors (first column) and pattern correlations (second column) for numerical experiments with $r^o = 10\%E = 6.9$ (top panels) and $r^o = 10\%E = 17.3$ (bottom panels) obtained with regularly spaced observations at the $N \times N = 36$ model grid points.
Figure 10: The truth and the filtered estimates at times $T = 230.72$ (first column) and $T = 363.28$ (second column), for case 1 ($M = N^2 = 36$, $r^o = 6.9$ and $T_{obs} = 0.01$). In each panel, we also show the observation locations (in circles).
Figure 11: The truth and the filtered estimates at times $T = 230.72$ (first column) and $T = 363.28$ (second column), for case 3 ($M = 49, N^2 = 36, r^o = 6.9$ and $T_{obs} = 0.01$).
Figure 12: The RMS errors as functions of time for case 4 ($M = N^2 = 36, r^o = 17.3$) and $T_{obs} = 0.01$. 
Figure 13: The true field $\psi^b$ (top left) and the filtered estimates at the time when RMS is at a minimum for case 4 ($M = N^2 = 36$, $r^o = 17.3$ and $T_{obs} = 0.01$).
Figure 14: The true field $\psi^b$ (top left) and filtered estimates at the time when RMS is at a maximum for case 4 ($M = N^2 = 36$, $r^o = 17.3$ and $T_{obs} = 0.01$).