Spectral characteristics of background error covariance and multiscale data assimilation

Zhijin Li[1,*,†], Xiaoping Cheng[2], William I. Gustafson Jr.[3] and Andrew M. Vogelmann[4]

[1]*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA*
[2]*The Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California, USA*
[3]*Pacific Northwest National Laboratory, Richland, Washington, USA*
[4]*Brookhaven National Laboratory, Upton, New York, USA*

SUMMARY

The spatial resolutions of numerical atmospheric and oceanic circulation models have steadily increased over the past decades. Horizontal grid spacing down to the order of 1 km is now often used to resolve cloud systems in the atmosphere and sub-mesoscale circulation systems in the ocean. These fine resolution models encompass a wide range of temporal and spatial scales, across which dynamical and statistical properties vary. In particular, dynamic flow systems at small scales can be spatially localized and temporarily intermittent. Difficulties of current data assimilation algorithms for such fine resolution models are numerically and theoretically examined. An analysis shows that the background error correlation length scale is larger than 75 km for streamfunctions and is larger than 25 km for water vapor mixing ratios, even for a 2-km resolution model. A theoretical analysis suggests that such correlation length scales prevent the currently used data assimilation schemes from constraining spatial scales smaller than 150 km for streamfunctions and 50 km for water vapor mixing ratios. These results highlight the need to fundamentally modify currently used data assimilation algorithms for assimilating high-resolution observations into the aforementioned fine resolution models. Within the framework of four-dimensional variational data assimilation, a multiscale methodology based on scale decomposition is suggested and challenges are discussed. Copyright © 2016 John Wiley & Sons, Ltd.

1. INTRODUCTION

The continuing growth in available computational power allows numerical atmospheric and oceanic circulation models to increase steadily in spatial resolution. Horizontal grid spacings down to the order of 1 km are now often used in atmospheric and oceanic models. In atmospheric applications, this resolution directly resolves cloud systems and thus improves representation of mesoscale cloud and precipitation systems as well as their interactions with large-scale circulations [1, 2]. In oceanic applications, this resolution resolves sub-mesoscale systems that are important to the dynamics of ocean circulation, particularly for energy cascade and dissipation, restratification of vertical density structures, and material transports [3–5]. Parallel to this, as radar and remote sensing technologies have advanced, the number, type, and quality of observations have increased. The expansion of radar networks and increase in the number and quality of satellite data products afford high-resolution observations.

---

*Correspondence to: Zhijin Li, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.
†E-mail: Zhijin.Li@jpl.nasa.gov

Assimilation of high-resolution observations into fine resolution models has been intensively investigated in recent years. In these investigations, data assimilation algorithms use the ensemble Kalman filter (EnKF) [6–8], three-dimensional variational data assimilation (3DVAR)[9–11], and four-dimensional variational data assimilation (4DVAR) [12, 13], where 3DVAR is still primarily used in operational centers for regional fine resolution models. All of these algorithms are formulated using the framework of minimum error variance or maximum likelihood estimation (or maximum posterior estimation in statistics), known as optimal estimation [14]. These algorithms are equivalent to one another when the optimal estimation problems are accurately solved [15–17].

However, the implementation of the aforementioned algorithms for fine resolution models has encountered fundamental difficulties. Optimal estimation depends on a background field and its solution hinges on the associated error covariance, that is, the background error covariance. Studies have suggested that a sequence of data assimilations should be applied for a set of decreasing correlation length scales so that small-scale components of the model solution may be constrained by assimilating high-resolution observations. Such a sequence of data assimilations has been applied in both ensemble Kalman filter [6, 7] and 3DVAR [10, 11]. Correlation length scale is a parameter central to the characterization of background error covariance [18].

Recently, we argued that the optimal estimation formulation for data assimilation is inherently ineffective when applied to fine resolution models, and that the ineffectiveness arises from the filtering properties [19]. The filtering properties are dictated by the background error correlation length scale. A long correlation length scale imposes strong filtering on small scales [20]. In [19], this ineffectiveness is addressed within the framework of 3DVAR. We will show here that a similar ineffectiveness occurs in 4DVAR.

By definition, background error covariance is a statistical quantity in the ensemble sense. A variety of intense, small-scale system types are spatially localized and temporally intermittent which account for only a limited portion of the total background error covariance. In this study, we present an analysis of background error correlation length scales using the 'National Meteorological Center (NMC) method' proposed in Parrish and Derber [21]. The results show that the background error correlation length scales are so large that optimal estimation algorithms are unable to constrain small scales that are well resolved in models.

To mitigate this ineffectiveness, it has been suggested that the small-scale component should be separated from the large-scale component and estimated separately, which can be achieved by decomposing a cost function for distinct scales [19]. We refer to an approach that uses decomposed cost functions as multiscale data assimilation (MS-DA), in which the data assimilation algorithm estimates the distinct scales separately. Here, we present a decomposition of the 4DVAR cost function that is similar to that for 3DVAR.

The outline of this paper is as follows. Section 2 presents a brief description of the formulation of basic data assimilation within the framework of 4DVAR. In section 3, background error correlation horizontal length scales are estimated. Section 4 addresses the spectral characteristics of background error correlations, and illustrates the relation to prescribed correlation length scales. Section 5 derives an optimal estimation solution based on an idealized two-dimensional (2-D) problem to illustrate filtering properties of data assimilation. In section 6, a decomposition of the 4DVAR cost function is derived and difficulties in data assimilation for the small-scale component are addressed. Finally, the summary and discussion are given in section 7.

2. BASIC FORMULATION

To proceed, we describe the basic incremental 4DVAR scheme. Incremental 4DVAR seeks a solution that minimizes the cost function

$$J(\delta \boldsymbol{x}_0) = \frac{1}{2} \delta \boldsymbol{x}_o^T \boldsymbol{B}^{-1} \delta \boldsymbol{x}_0 + \frac{1}{2} \sum_{i=o}^{n} (\boldsymbol{H}_i \delta \boldsymbol{x}_i - \boldsymbol{d}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{H}_i \delta \boldsymbol{x}_i - \boldsymbol{d}_i), \qquad (1)$$

where $\delta \boldsymbol{x}_i = \boldsymbol{x}_i - \boldsymbol{x}_i^b$ ($i = 0, 2, \cdots, n$) is the increment between the state, $\boldsymbol{x}_i$, and the background/forecast state, $\boldsymbol{x}_i^b$, at time t_i from a nonlinear forward model. The increment δx_i is governed by

$$\delta \boldsymbol{x}_i = \boldsymbol{L}(t_i, t_0)\delta \boldsymbol{x}_0, \tag{2}$$

where $\boldsymbol{L}(t_i, t_0)$ is a propagation operator, or simply propagator, which propagates the increment in time from the initial time t_0 to a later time t_i through a tangent linear model. We note that $\boldsymbol{L}(t_i, t_i) = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. The propagator is a linearized approximation to a nonlinear propagator. This linearized approximation is known as incremental approximation. The innovation vector is given at each time step t_i by $\boldsymbol{d}_i = \boldsymbol{y}_i^o - \boldsymbol{H}_i \boldsymbol{x}_i^b$, where $\boldsymbol{y}_i^o$ is the observation vector, $\boldsymbol{H}_i$ the observation operator, and $\boldsymbol{R}_i$ the error covariance associated with the observation $\boldsymbol{y}_i^o$. For more details on incremental 4DVAR, the reader is referred to Courtier et al. [23].

In (1), $\boldsymbol{B}$ is the error covariance associated with the initial background state $\boldsymbol{x}_0^b$. $\boldsymbol{B}$ is defined here because its properties are central to the later discussion. By definition, the background error is

$$\boldsymbol{\varepsilon} = \boldsymbol{x}_0^b - \boldsymbol{x}_0^t, \tag{3}$$

where $\boldsymbol{x}_0^t$ is the unknown true state at the initial time. The background error covariance can be written as

$$\boldsymbol{B} = \langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \rangle, \tag{4}$$

where $\langle \cdot \rangle$ denotes the ensemble mean or mathematical expectation, and the superscript 'T' stands for transpose.

Following [24], the cost function can be written in the compact form

$$J(\delta \boldsymbol{x}_0) = \frac{1}{2}\delta \boldsymbol{x}_o^T \boldsymbol{B}^{-1} \delta \boldsymbol{x}_0 + \frac{1}{2}\left(\boldsymbol{G}\delta \boldsymbol{x}_0 - \boldsymbol{d}\right)^T \boldsymbol{R}^{-1}(\boldsymbol{G}\delta \boldsymbol{x}_0 - \boldsymbol{d}), \tag{5}$$

where $\boldsymbol{d}$ is a vector that consists of $\boldsymbol{d}_i$ at all the time steps and has the form $\boldsymbol{d} = \left(\boldsymbol{d}_1^T, \boldsymbol{d}_2^T, \cdots, \boldsymbol{d}_n^T\right)^T$. Correspondingly, $\boldsymbol{R}$ is a block diagonal matrix, with each block being $\boldsymbol{R}_i$. $\boldsymbol{G}$ is a matrix that has the form $\boldsymbol{G} = [(\boldsymbol{H}_0 \boldsymbol{L}(t_0, t_0))^T, (\boldsymbol{H}_1 \boldsymbol{L}(t_0, t_1))^T, \cdots, (\boldsymbol{H}_n \boldsymbol{L}(t_0, t_n))^T]^T$.

Letting $\frac{\partial J(\delta \boldsymbol{x}_0)}{\partial \delta \boldsymbol{x}_0} = 0$, we obtain the following solution that minimizes 5

$$\delta \boldsymbol{x}_0^a = \boldsymbol{x}_0^a - \boldsymbol{x}_0^b = \boldsymbol{B}\boldsymbol{G}^T\left(\boldsymbol{G}\boldsymbol{B}\boldsymbol{G}^T + \boldsymbol{R}\right)^{-1}\boldsymbol{d}, \tag{6}$$

which is known as the analysis increment, and $\boldsymbol{x}_0^a$ is the optimal estimate. We can see that $\boldsymbol{B}$ on the left hand side of the analysis increment acts as a spatial smoothing operator, which will be further illustrated using the spectral analyses in sections 4 and 5.

We note that the cost function (5) and solution (6) reduce to those for incremental 3DVAR when the number of the time steps n is set to be zero. Hence, we can use (5) and (6) to discuss properties of both 3DVAR and 4DVAR.

3. BACKGROUND ERROR CORRELATION LENGTH-SCALE ANALYSES

The background error covariance $\boldsymbol{B}$ can be decomposed into

$$\boldsymbol{B} = \boldsymbol{\Sigma}\boldsymbol{C}\boldsymbol{\Sigma}, \tag{7}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix whose elements are the background error standard deviation associated with $\boldsymbol{x}_0^b$, and $\boldsymbol{C}$ is the correlation matrix whose elements consist of the spatial correlations.

In typical atmospheric and oceanic applications, the dimensions of the state vector are larger than 10^6. Thus, because of the large dimensions, the correlation matrix cannot be given element-wise and, to make it computationally feasible to solve, it must be approximately reduced. There are two approaches to do so. One approach is to reduce the rank or dimensions of the correlation matrix. For this approach, there are ensemble based methods [25–27], which are currently extensively used [28], reduced-order methods based on Empirical Orthogonal functions (EOFs) [29], or similarly Proper Orthogonal Decomposition (POD) methods [30–32]. The other approach is to

parameterize the correlation matrix. Here, we address the spectral representation of the correlations using parameterized correlations.

3.1. Correlation length scale

Correlation length scale is central to the parameterization of correlation. For a smooth and isotropic correlation function c at the origin, the correlation length scale is defined in [20] as

$$L_D = \sqrt{-\frac{1}{\nabla^2 c(0)}}$$

in one dimension and

$$L_D = \sqrt{-\frac{2}{\nabla^2 c(0)}}$$

in two dimensions, where ∇^2 is the Laplace operator. By specifying a correlation length scale, a correlation function can be constructed [18] and, hence, this length scale primarily characterizes the correlation. We will show how this correlation length scale dictates the spectral characteristics of the correlation and the filtering properties of optimal estimation.

To illustrate, a frequently used parameterized correlation function is a Gaussian function, $c(r) = e^{-\frac{r^2}{2L_D^2}}$, where r is the distance between grid points, and L_D is the correlation length scale. In this case, the correlation scale is the spatial distance over which the correlation decreases from 1 to $e^{-\frac{1}{2}}$ (e is the base of the natural logarithm). The second-order autoregressive function (SOAR) is also often used as correlation function [20].

3.2. Correlation length scale analysis

In the introduction, we argue that the background error correlation length scale is generally large even for a fine-resolution model, because the small scales account only for a limited amount of background error due to the localized and intermittent nature of small-scale systems. Here, we present an analysis of correlation length scales to illustrate this point.

In operational 3DVAR/4DVAR, the most widely used method for estimating the background error covariance $\mathbf{B}$ is the 'NMC method' detailed in [21]. In this method, the differences, say, between 24 h and 48 h forecasts valid for the same time are used as surrogates of the background error to estimate $\mathbf{B}$. Recently, a $\mathbf{B}$ so estimated is often combined with that estimated using ensemble methodologies, known as ensemble hybrid methods [28, 33]. For $\mathbf{B}$ estimated from the ensemble-based method, the correlation length scale is generally overestimated. We here use only the NMC method for our analysis.

The data assimilation and forecasting system that is used to produce forecasts used here have been described in Li et al. [34] and Feng et al. [35]. The Weather Research and Forecast (WRF) model (version 3.7) is employed using triple-nested domains. The domains are centered roughly at 36.6 N 97.5 W, which is the location of the Southern Great Plains (SGP) Site of the Atmospheric Radiation Measurement (ARM) Climate Research Facility [36, 37]. The grid spacings of the nested domains are 18 km, 6 km, and 2 km, going from the outer to inner domain. There are 45 vertical levels with the model top at 100 hPa. We set up these model domains to leverage the unique and dense observations from the ARM Facility. In particular, the inner 2-km domain approximately encompasses the area of the ARM observing network (Figure 1). A mutliscale 3DVAR (MS-3DVAR) scheme is implemented based on the Community Gridpoint Statistical Interpolation (GSI) system (http://www.dtcenter.org). The data assimilation system is executed every 6 h at 00, 06, 12, and 18 UTC. The observations that are assimilated are summarized in Li et al. [34], which includes conventional and satellite observations processed and distributed by the National Centers
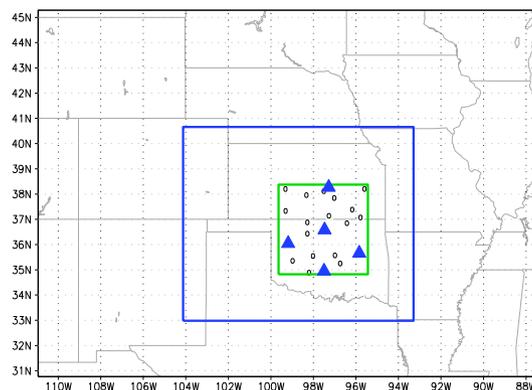
Figure 1. The Weather Research and Forecast triple nested domains with grid spacings of 18, 6, and 2 km going from the outer to inner domain. The blue triangles indicate the locations of the Atmospheric Radiation Measurement sonde launches and the circles indicate the locations of Surface Meteorological Observation Systems. The 2-km resolution domain encompasses the enhanced Atmospheric Radiation Measurement observing network.

for Environmental Prediction (NCEP) [22] along with those obtained from the ARM observing network.

To estimate correlation length scales, twice daily for two months in June and July 2007, we generated the differences between the 12 and 24 h forecasts that are valid at the same time. From these differences, spatial correlations are first calculated. The Gaussian function is then fit to the calculated correlations to estimate correlation length scales. Specifically, the correlation length scale is computed as $L_D = \sqrt{-\frac{r^2}{2 ln c(r)}}$, where $c(r)$ is a correlation coefficient and r is a distance between the two points. Because the computation is not reliable for $c(r)$ approaching 1 and 0, we compute L_D for $0.4 \leqslant c(r) \leqslant 0.95$.

Figure 2 shows the calculated correlations and correlation length scales at an altitude of about 1.8 km, which is located in the upper boundary layer. Figure 2a1–a3 present for the three model domains the streamfunction correlations between the ARM Southern Great Plains site and all the model grid points. The correlation length scales are as large as 240, 120, and 75 km, respectively, for grid spacings of 18, 6, and 2 km (Figure 2c1–c3). The correlation length scales are variable dependent where, for example, the correlation length scale of moisture is considerably smaller than that of the streamfunction and velocity potential. Figure 2b1–b3 present the correlations of water vapor mixing ratio, where the correlation length scales are about 80 km, 40 km, and 25 km for the three domains (Figure 2c1–c3). We note that the correlation length scales do not decrease linearly with decreasing model grid spacing. The calculated correlation length scales in Figure 2 are only for one location, but they represent the model domain averages well. Also, these length scales do not vary much with altitude from 1000 to 3000 m.

4. SPECTRAL REPRESENTATION OF ERROR CORRELATION FOR TWO-DIMENSIONAL PROBLEMS

The filtering properties of data assimilation are basically the manifestations of the spectral properties of the background error correlation. For simplicity, we use a horizontal 2-D discrete problem to illustrate the spectral properties of the background error correlation using bi-Fourier transforms for a limited area.

4.1. Bi-Fourier transforms of physical fields

A spectral expansion can be applied to a limited-area [38]. For biperiodic lateral boundary conditions, Fourier transforms are naturally applicable. Practical methods have been developed
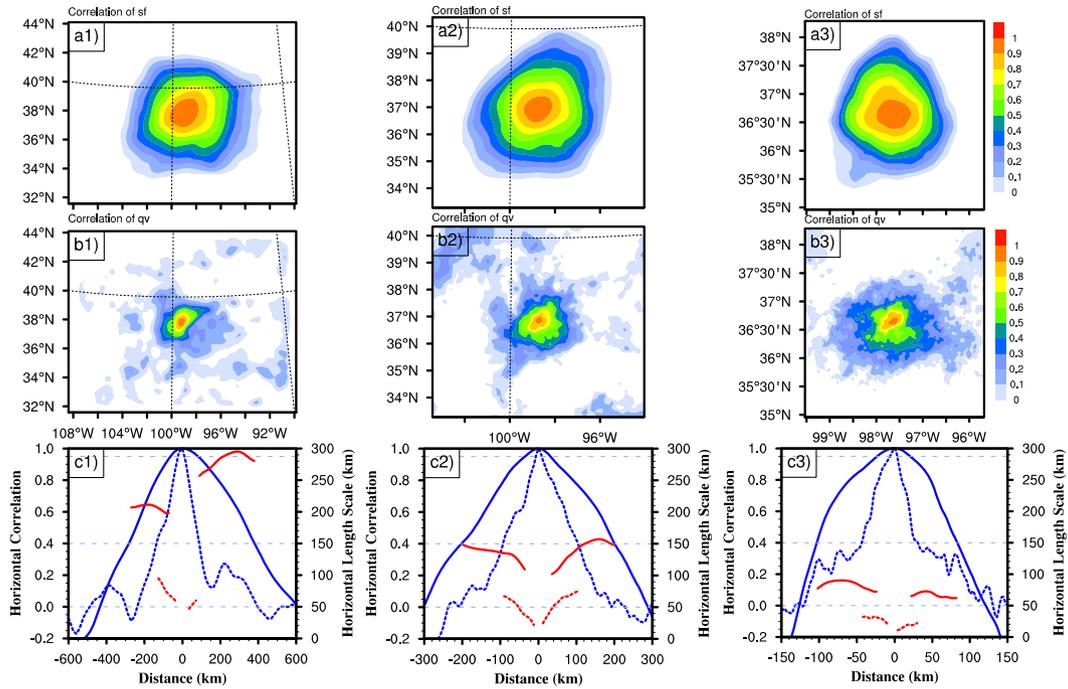
Figure 2. Horizontal correlations and correlation length scales for three model domains at a height of about 1.8 km. The three columns from left to right are for the 18, 6, and 2 km grid spacing domains. The correlations for the streamfunctions between the location of the Atmospheric Radiation Measurement Southern Great Plains site and all model grid points are in the top row (a1–a3), and the correlations for water vapor mixing ratio are in the middle row (b1–b3). The correlation profiles across the Atmospheric Radiation Measurement Southern Great Plains site, from south to north, and the corresponding estimated correlation scales are given in the bottom row (c1–c3). In c1–c3, the blue curves are horizontal correlations, and the red curves are the estimated horizontal correlation length scales for correlations between 0.4 and 0.95. Solid curves are for the streamfunction, and dotted curves are for water vapor mixing ratio.

for a limited area with non-periodic lateral boundary conditions; for example, an extension zone can be used to make the field biperiodic [39].

In the data assimilation formulation as presented in section 2, the state variable is expressed as a column vector. For a limited area as shown in Figure 3, a discrete 2-D field can be written as a column vector of length $n_x n_y$ as follows

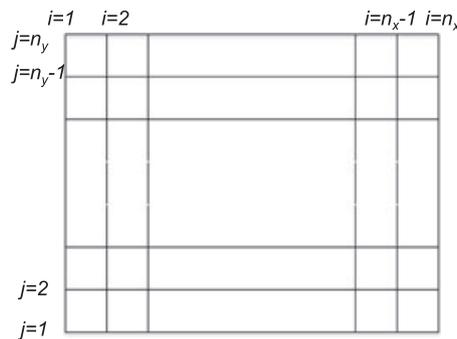$$\boldsymbol{f} = \left(\boldsymbol{f}_{x1}^T, \boldsymbol{f}_{x2}^T, \cdots, \boldsymbol{f}_{xn_y}^T\right)^T,$$

Figure 3. The two-dimensional grid. There are n_x grid points in the *x*-direction, and n_y grid points in the *y*-direction.

where $\boldsymbol{f}_{xj}$ $(j = 1, 2, \cdots, n_y)$ is a column vector of length n_x that encompasses the grid points along the j-th row. We also denote $\boldsymbol{f}_{yi}$ $(i = 1, 2, \cdots, n_x)$ as a column vector of length n_x that encompasses the grid points along the i-th column.

With the definitions given above, the 1-D Fourier transforms can be written as

$$\boldsymbol{f}_{sj} = \boldsymbol{F}_x \boldsymbol{f}_{xj} \text{ and } \boldsymbol{f}_{si} = \boldsymbol{F}_y \boldsymbol{f}_{yi} \tag{8}$$

for the x-direction and y-direction. Here the Fourier transform matrix $\boldsymbol{F}_x$ is an $n_x \times n_x$ matrix, and the Fourier transform matrix $\boldsymbol{F}_y$ an $n_y \times n_y$ matrix.

From the 1-D Fourier transforms for the x-directions and y-directions, we obtain the bi-Fourier expansion for a 2-D field in the form

$$\boldsymbol{f}_s = (\boldsymbol{F}_x \otimes \boldsymbol{F}_y) \boldsymbol{f}, \tag{9}$$

where $\boldsymbol{f}_s$ is an $n_x \times n_y$ vector. The $m \times n$-th element of $\boldsymbol{f}_s$, $\boldsymbol{f}_{s(m,n)}$, is the Fourier coefficient at a wavenumber of m in x-direction and a wavenumber of n in y-direction. Here $\otimes$ denotes a tensor product, also known as Kronecker product (Appendix B).

By definition, the covariance in spectral space is $\boldsymbol{f}_s \boldsymbol{f}_s^T$. Following (9), we have

$$\langle \boldsymbol{f}_s \boldsymbol{f}_s^T \rangle = (\boldsymbol{F}_x \otimes \boldsymbol{F}_y) \langle \boldsymbol{f} \boldsymbol{f}^T \rangle (\boldsymbol{F}_x \otimes \boldsymbol{F}_y)^T. \tag{10}$$

4.2. Spectral representation of the background error correlation

The spectral properties of the background error correlation can be related to the classic Wiener–Khintchine theorem [40]. The theorem states that the autocorrelation function of a wide-sense-stationary random process has a spectral decomposition given by the power spectrum of that process. It means that the power spectrum of a process is equivalent to the autocorrelation function. For one-dimensional (1-D), infinite domains that have homogenous and isotropic spatial correlations, the Wiener–Khintchine theorem implies that the specification of the spatial correlation function of a random field is equivalent to the specification of its power spectral density (PSD).

The expansion of meteorological fields in terms of spherical harmonics has been used in operational global variational assimilation schemes, such as in the Spectral Statistical Interpolation (SSI) system at NCEP (former the NMC) [21, 22] and the 3DVAR system at ECMWF [41]. Assuming homogeneous and isotropic conditions, in spectral space the background error covariance is simply the variance for each wavenumber, that is, the PSD.

For homogeneous background errors, the spectral representation of the covariance matrix $\boldsymbol{B}$ is equivalent to that of the correlation matrix $\boldsymbol{C}$. We here consider a particular context where $\boldsymbol{C}$ is separable in the x-directions and y-directions. In this context, the 2-D correlation can be expressed as

$$\boldsymbol{C} = \boldsymbol{C}_x \otimes \boldsymbol{C}_y, \tag{11}$$

where $\boldsymbol{C}_x$ and $\boldsymbol{C}_y$ are 1-D correlations in the x-directions and y-directions. The expression in (11) has been used in the construction of the background error correlation [42, 43].

Following (10) and (11), the correlation matrix in the spectral form is

$$\boldsymbol{S} = (\boldsymbol{F}_x \otimes \boldsymbol{F}_y)(\boldsymbol{C}_x \otimes \boldsymbol{C}_y)(\boldsymbol{F}_x \otimes \boldsymbol{F}_y)^T. \tag{12}$$

Using the Kronecker product relations (B2) and (B3), we obtain

$$\boldsymbol{S} = (\boldsymbol{F}_x \boldsymbol{C}_x \boldsymbol{F}_x^T) \otimes (\boldsymbol{F}_y \boldsymbol{C}_y \boldsymbol{F}_y^T) = \boldsymbol{S}_x \otimes \boldsymbol{S}_y, \tag{13}$$

where $\boldsymbol{S}_x$ and $\boldsymbol{S}_y$ are the spectral transformations of the 1-D correlation matrices $\boldsymbol{C}_x$ and $\boldsymbol{C}_y$, respectively.

For isotropic correlations, $\boldsymbol{S}_x$ and $\boldsymbol{S}_y$ are diagonal and their diagonal elements are the PSD of each wavenumber in the x-directions and y-directions. Following (13), we determine from the Kronecker

product property (Appendix B, equation B.5) that S is diagonal, and its diagonal elements consist of the 2-D PSD. The equivalent to (13) has been obtained for a global domain in Boer [44], where spherical harmonics are used for Fourier transforms. We use the equality (13) to examine the relationship between correlation scale and PSD.

To illustrate, we assume that the correlation function is a Gaussian function and the diagonal elements of S_x and S_y are calculated using (A.7). In Figure 4a and b, the PSD is given as a function of the wavenumber in the x-directions and y-directions for correlation length scales of 50 and 25 km.

A comparison between Figure 4a and b reveals that the PSD at small scales declines more quickly for the larger correlation scales. For a correlation length scale of 50 km, the PSD decreases to be less than 10^{-2} at a wavelength of 100 km and to 10^{-5} at a wavelength of 70 km. In contrast, for a correlation length scale of 25 km, the PSD decreases to be less than 10^{-2} at a wavelength of 50 km and to 10^{-5} at a wavelength of 35 km. Overall, the PSD decreases to 10^{-2} at a wavelength that is twice the correlation length scale. We note that the PSD is constant for uncorrelated background errors and, thus, the decrease of PSD with spatial scale arises solely from the correlation.

The rapid decrease of PSD with decrease in spatial scale is responsible for the inability of data assimilation to correct small-scale background errors. To illustrate, we examine the background term of the cost function, the first term in (1). The ratio r_{nc} between the components for wavenumbers larger than a chosen wavenumber n_c and those smaller than or equal to n_c is given by

$$r_{n_c} = \sum_{n>n_c}\sum_{m>n_c}PSD_{mn}/\sum_{n\leqslant n_c}\sum_{m\leqslant n_c}PSD_{mn}, \tag{14}$$

where PSD_{mn} denotes the PSD for wavenumber m in the x-direction and n in the y-direction. When n_c is given such that the corresponding wavelength is larger than twice the correlation length scale, r_{n_c} becomes virtually zero, and the small-scale components cannot be accounted for in the background term.
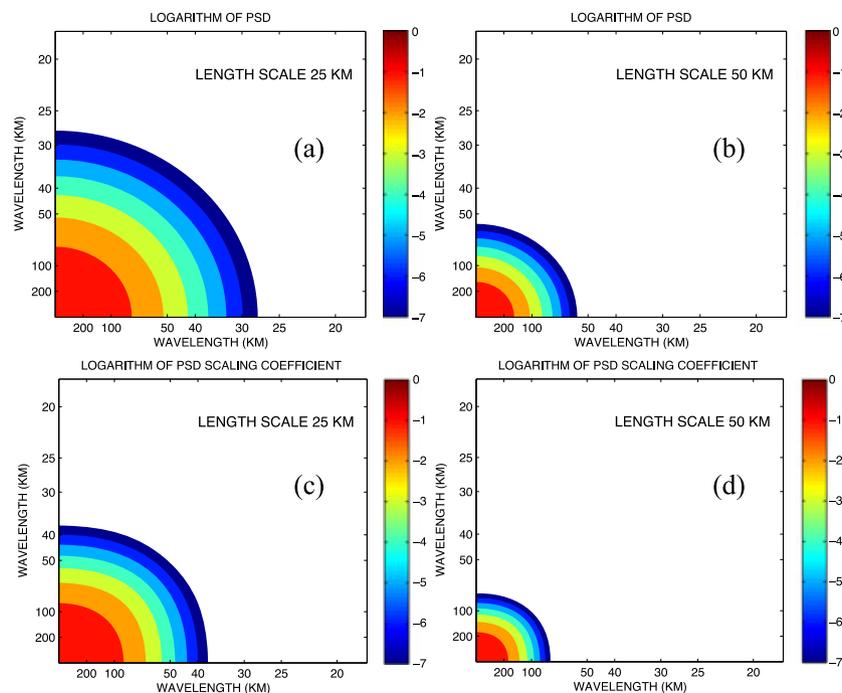

Figure 4. Power spectral density (PSD) of the background error and the scaling coefficients, which scale the PSD of the innovation to produce the PSD of the analysis increment of each wavenumber. For correlation length scales of 50 and 25 km, the PSDs are given in (a) and (b) and the scaling coefficients in (c) and (d). The PSDs and scaling coefficients are both standardized.

5. FILTERING PROPERTIES

To further illustrate the inability of the data assimilation to correct small-scale background errors, we examine analytical solutions. We analyze a 2-D problem without the vertical and time dimensions. In this analysis, we assume again that the background errors are homogeneous and isotropic, and that there is one observation at each grid point. Under these assumptions, the solution in (6) reduces to

$$x_0^a = x_0^b + B(B + R)^{-1}d \tag{15}$$

Applying the bi-Fourier transform to both sides of (15) yields

$$(S_x \otimes S_y)x_0^a = (S_x \otimes S_y)x_0^b + (S_x \otimes S_y)B(S_x \otimes S_y)(S_x \otimes S_y)^{-1}(B + R)^{-1}(S_x \otimes S_y)^{-1}(S_x \otimes S_y)d, \tag{16}$$

In (16), we used the relationship $(F_x \otimes F_y)(F_x \otimes F_y)^{-1} = I$. Using the Kronecker product property (B.4) and (13), we obtain from (16)

$$s_0^a = s_0^b + (S_x \otimes S_y)\left((S_x \otimes S_y) + \frac{\sigma^{o2}}{\sigma^{b2}}I\right)^{-1}s_d, \tag{17}$$

where $s_0^a = (S_x \otimes S_y)x_0^a$ contains the spectral coefficients of the analysis, $s_0^b = (S_x \otimes S_y)x_0^b$ contains the spectral coefficients of the background field, and $s_d = (S_x \otimes S_y)d$ contains the spectral coefficients of the innovation. Further, σ^{b2} and σ^{o2} are the background and observation error variances, respectively.

In (17), the matrix that multiplies s_d is

$$K = (S_x \otimes S_y)\left((S_x \otimes S_y) + \frac{\sigma^{o2}}{\sigma^{b2}}I\right)^{-1}. \tag{18}$$

It is the Kalman gain [45] that, here, is a diagonal matrix. The elements of this matrix are the scaling coefficients that scale the spectral coefficients of the innovation to yield the spectral coefficients of the analysis of each wavenumber.

From the gain matrix K, we can quantitatively illustrate the filtering properties associated with the background error correlations given in section 4. Here we specify $\frac{\sigma^{o2}}{\sigma^{b2}} = 0.25$, that is, the magnitude of the observational error is half the background error. Figure 4c and b present the scaling coefficients of each wavenumber for correlation length scales of 50 and 25 km. For a correlation length scale of 50 km, the scale coefficient approaches 10^{-4} at a wavelength of 100 km. For a correlation length scale of 25 km, the scale coefficient approaches 10^{-4} at a wavelength of 50 km. We note that $\frac{\sigma^{o2}}{\sigma^{b2}} \approx 1$ is generally used in operational applications and, thus, the scaling coefficients are even smaller. From such scaling coefficients, we conclude that the component of the analysis increment is virtually zero at least for scales less than twice the background error correlation length scale.

6. MULTISCALE DATA ASSIMILATION METHODOLOGIES

In section 3, the analyses show that the correlation length scales are as large as 75 km for the streamfunction and 25 km for moisture mixing ratios, even for a model with a grid spacing of 2 km. Because data assimilation cannot correct the background errors for scales less than at least twice the correlation length scale, current data assimilation methodologies are, thus, ineffective at constraining small-scales.

To mitigate this issue, the large and small scales should be decomposed and estimated separately, as mentioned before. In 3DVAR, this is achieved by decomposing the cost function for distinct scales [19]. We call the resulting scheme MS-3DVAR. Following MS-3DVAR, we here decompose the 4DVAR cost function (5).

To proceed, we decompose the increment for two distinct scales as

$$\delta x_0 = F_L \delta x_0 + F_S \delta x_0 = \delta x_{0L} + \delta x_{0S}, \tag{19}$$

where δx_{0L} and δx_{0S} denote the large-scale and small-scale components of δx_0, respectively, and the two bi-Fourier expansion operators satisfy $F_L + F_S = I$.

Correspondingly, the decomposition of the background error can be written as

$$\varepsilon = F_L \varepsilon + F_S \varepsilon = \varepsilon_L + \varepsilon_S, \tag{20}$$

where ε_L and ε_S are the large-scale and small-scale components of the background error, respectively. From (20), we have

$$B = \left\langle \varepsilon_L \varepsilon_L^T \right\rangle + \left\langle \varepsilon_S \varepsilon_S^T \right\rangle = B_L + B_S, \tag{21}$$

where B_L and B_S are the error covariances associated with x_{0L}^b and x_{0S}^b. To obtain (21), we have assumed that the large-scale and small-scale background errors are uncorrelated.

With the decompositions in (20) and (21), the cost function in (5) can be decomposed into two cost functions

$$J(\delta x_{0L}) = \frac{1}{2} \delta x_{oL}^T B_L^{-1} \delta x_{0L} + \frac{1}{2} \left(G \delta x_{0L} - d \right)^T \left(G B_S G^T + R \right)^{-1} \left(G \delta x_{0L} - d \right), \tag{22}$$

$$J(\delta x_{0S}) = \frac{1}{2} \delta x_{oS}^T B_S^{-1} \delta x_{0S} + \frac{1}{2} \left(G \delta x_{0S} - d \right)^T \left(G B_L G^T + R \right)^{-1} \left(G \delta x_{0S} - d \right). \tag{23}$$

The decomposed cost functions (22) and (23) can then be used to estimate the large-scale and small-scale components separately, and the resulting algorithm is a multiscale 4DVAR (MS-4DVAR).

The cost function (22) actually represents a basic incremental 4DVAR [23], or 3DVAR if no time dimension is considered, that is, $n = 0$ in (1). In (22), there is an additional component, $G B_S G^T$, in the observation error covariance. This component results from representativeness errors, referred to as multi-scale representativeness errors [19].

The solution obtained by minimizing (23) is the small-scale analysis increment. In (23), B_S generally has small variances, but its correlation length scales could be large. This is because B_S involves an ensemble and/or domain mean. Even in the case where there are intense small-scale activities, B_S may still have limited variances and large correlation length scales, due to the temporally intermittent and spatially localized characteristics of small-scale systems. In practice, an inflation and localization of B_S is needed.

We note that the innovation d in (22) and (23) is not decomposed. Some high-resolution observations can be decomposed and they can then be assimilated. In this case, the multi-scale representativeness errors are removed [19].

7. SUMMARY AND DISCUSSION

The currently used data assimilation algorithms are primarily formulated within the framework of optimal estimation theory [14, 46]. In these algorithms, a background state is used, and a data assimilation algorithm can be described as correcting the background error in terms of maximum likelihood or minimum error variance estimates. The associated background error covariance functions in three primary ways: it filters out observation errors, spreads observational information in space and time, and incorporates dynamic balance for propagating observational

information among state variables. These three functions of the background error covariance play a key role in the data assimilation.

However, as numerical model resolutions increase to resolve cloud systems in the atmosphere or to resolve sub-mesoscale systems in the ocean, fundamental difficulties arise from the use of the background error covariance. One major difficulty is that its filtering effect precludes optimal estimation algorithms from constraining small scales using high-resolution observations. We analyzed this difficulty in 3DVAR in [19] and here in 4DVAR.

The background error covariance is mainly characterized by its correlation length scale. We analyzed correlation length scales using the 'NMC' method [21]. The results show that correlation length scales are as large as 75 km for the streamfunction and are larger than 25 km for moisture mixing ratios, even though the model has a grid spacing of 2 km.

Such large correlation scales suggest that data assimilation algorithms for 3DVAR and 4DVAR cannot correct background errors at small scales. Specifically, background errors cannot be corrected at scales less than at least twice the correlation length scales. In terms of the estimated correlation scales, data assimilation algorithms do not correct background errors for scales less than 150 km in the streamfunction fields and 50 km in the moisture fields, even though fine resolution observations are available to assimilate. Of course, the estimated correlation scales are determined by multiple factors, such as the model used, effectiveness of data assimilation, and particular observations assimilated.

The inability of the currently used data assimilation algorithms to constrain small scales suggests the need to adjust or even reformulate them. We propose decomposing the cost functions for distinct spatial scales and implement data assimilation separately for each distinct spatial scale. In [19], we derive a decomposition of the 3DVAR cost function for large and small scales. Here the decomposition has been extended to the 4DVAR cost function. For the large-scale component, we can simply use existing data assimilation algorithms. For the small-scale component, however, we argue that the maximum likelihood or minimum error variance estimates may not be suitable because the small scales account for a very limited amount of the error variance.

We have implemented a MS-3DVAR for oceanic applications [47]. In this MS-3DVAR, the data assimilation for the small-scale component follows the framework of conventional 3DVAR, but the correlation length scale is empirically reduced. For MS-4DVAR, it is straightforward to use a basic 4DVAR system for the large-scale component. For the small-scale component, it is also possible to use a basic 4DVAR system, but major adjustments are necessary.

For small-scales, the Rossby number is larger than one so nonlinear advection is important; also, an energetic small-scale system is often associated with convection that is nonlinear. Thus, small-scale systems are more nonlinear and non-Gaussian than large-scale systems. The nonlinear and non-Gaussian issue could be alleviated to some degree in MS-DA because the increment for small scales is smaller in magnitude than the total increment for all scales. While modifying a currently used data assimilation algorithm such as MS-DA to constrain small-scales should continue to be attractive in a practical sense, we note that non-Gaussian and nonlinear data assimilation algorithms such as particle filters [48–50] are worth exploring.

APPENDIX A. SPECTRAL TRANSFORMATION OF ONE-DIMENSIONAL COVARIANCE

Consider a 1-D problem in the x-direction. The background error has the Fourier expansion

$$\varepsilon(x) = \frac{1}{L_x} \sum_{m=-M}^{M} \varepsilon_m e^{i2\pi mx/L_x}, \tag{A.1}$$

where L_x is the length of the 1-D interval, and the Fourier coefficient at wavenumber m is

$$\varepsilon_m = \int_0^{L_x} \varepsilon(x) e^{-i2\pi mx/L_x} dx. \tag{A.2}$$

From (A.1) and (A.2), we can have

$$\int_0^{L_x} \varepsilon^2(x)dx = \frac{1}{L_x}\sum_{m=-M}^{M}\varepsilon_m\varepsilon_m^*. \tag{A.3}$$

The equality (A.3) is the well-known Parseval's theorem [40].

The covariance between the Fourier coefficients at wavenumbers m and n can be written as

$$\langle\varepsilon_m\rangle\varepsilon_n^* = \int_0^{L_x}\int_0^{L_x} \langle\varepsilon(x_1)\varepsilon(x_2)\rangle e^{-i2\pi mx_1/L_x}e^{i2\pi nx_2/L_x}dx_1dx_2 \tag{A.4}$$

For homogeneous and isotropic background errors, the covariance has the symmetric property

$$\langle\varepsilon(x_1)\varepsilon(x_2)\rangle = c(x_1 - x_2) = c(x_2 - x_1) \tag{A.5}$$

Using the symmetry (A.5), we can obtain from (A.4)

$$\langle\varepsilon_m\varepsilon_n^*\rangle = \begin{cases} \int_0^{L_x}(L_x - x)c(x)e^{-i2\pi x/L_x}d\tau, & m = n \\ 0, & m \neq n \end{cases} \tag{A.6}$$

(A.6) shows that the spectra at distinct wavenumbers are not correlated.

From (A.3), the spectral power density is $s_m = \frac{1}{L_x}\langle\varepsilon_m\varepsilon_m^*\rangle$. When the correlation length scale L_D is much smaller than L_x, that is, $L_D/L_x \ll 1$, we can obtain from (A.6)

$$s_m \approx \int_0^{L_x} c(\tau)e^{-i2\pi m\tau/L_x}d\tau \tag{A.7}$$

(A.7) is simply the manifestation of the classic Wiener–Khintchine theorem.

APPENDIX B. PROPERTIES OF KRONECKER PRODUCT

Let A be an $m \times n$ matrix and B a $p \times q$ matrix. The Kronecker product (or tensor product) of A and B is then defined as the matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}, \tag{B.1}$$

which is an $mp \times nq$ matrix. The same definition holds if A and B are complex-valued matrices.

A comprehensive discussion on properties of the Kronecker product operator can be found in Graham [51]. The relations that have been used are as follows:

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \tag{B.2}$$

$$(A \otimes B)^T = A^T \otimes B^T, \tag{B.3}$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}, \tag{B.4}$$

$$A \otimes B \text{ is diagonal if } A \text{ and } B \text{ are diagonal.} \tag{B.5}$$

In these relations, we assume that the dimensions of the matrix are appropriately defined for all matrix operations.

REFERENCES

1. Randall DA, Khairoutdinov MF, Arakawa A, Grabowski WW. Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society* 2003; **84**:1547–1564.
2. Guichard F, Coauthors. Modeling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Quart. J. Roy. Meteor. Soc.* 2004; **131**:3139–3172.
3. Lapeyre G, Klein P. Impact of the small-scale elongated filaments on the oceanic vertical pump. *Journal of Marine Research* 2006; **64**:835–851.
4. Capet X, McWilliams JC, Molemaker MJ, Shchepetkin AF. Mesoscale to submesoscale transition in the California Current System. Part I: Flow structure, eddy flux, and observational tests. *Journal of Physical Oceanography* 2008; **38**:29–43.
5. Mahadevan A, Tandon A, Ferrari R. Rapid changes in mixed layer stratification driven by submesoscale instabilities and winds. *Journal of Geophysical Research* 2010; **115**:C03017.
6. Zhang F, Weng Y, Sippel JA, Meng Z, Bishop CH. Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review* 2009; **137**:2105–2125.
7. Zhang F, Weng Y, Gamache JF, Marks FD. Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations. *Geophysical Research Letters* 2011 L15810; **38**.
8. Aksoy A, Dowell DC, Snyder C. A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Monthly Weather Review* 2009; **137**:1805–1824.
9. Brousseau P, Berre L, Bouttier F, Desroziers G. Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. *Q.J.R. Meteorol. Soc.* 2011; **137**:409–422.
10. Xie Y, Koch S, McGinley J, Albers S, Bieringer PE, Wolfson M, Chan M. A space–time multiscale analysis system: A sequential variational analysis approach. *Monthly Weather Review* 2011; **139**:1224–1240.
11. Gao J, Smith TM, Stensrud DJ, Fu C, Calhoun K, Mnaross KL, Brogden J, Lakshmanan V, Wang Y, Thomas KW, Brewster K, Xue M. A realtime weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Wea. Forecasting* 2013; **28**:727–745.
12. Kawabata T, Kuroda T, Seko H, Saito K. A Cloud-Resolving 4DVAR Assimilation Experiment for a Local Heavy Rainfall Event in the Tokyo Metropolitan Area. *Monthly Weather Review* 2011; **139**:1911–1931.
13. Sun J, Wang H. Radar data assimilation with WRF 4D-Var. Part II: Comparison with 3D-Var for a squall line over the U.S. Great Plains. *Monthly Weather Review* 2013; **141**:2245–2264.
14. Cohn SE. Estimation theory for data assimilation problems: Basic conceptual framework and some open questions. *Journal of the Meteorological Society of Japan* 1997; **75**(1B):257–288.
15. Ménard R, Daley R. The application of Kalman smoother theory to the estimatiòn of 4DVAR error statistics. *Tellus* 1996; **48A**:221–237.
16. Li Z, Navon IM. Optimality of variational data assimilation and its relationship with the Kalman filter and Kalman smoother. *Quart. J. Roy. Meteorol. Soc.* 2001; **127**:661–684.
17. Fisher M, Leutbecher M, Kelly GA. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society* 2005; **131**:3235–3246.
18. Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.* 1999; **125**:723–757.
19. Li Z, McWilliams JC, Ide K, Fararra JD. A Multi-Scale Data Assimilation Scheme: Formulation and Illustration. *Monthly Weather Review* 2015; **143**:3804–3822.
20. Daley, R., *Atmospheric data assimilation*. Cambridge atmospheric and space science series, Cambridge University Press, Cambridge, USA, 1991, pp. 472.
21. Parrish DF, Derber JC. The National Meteorological Center's spectral-interpolation system. *Monthly Weather Review* 1992; **120**:1747–1763.

22. Kleist DT, Parrish DF, Derber JC, Treadon R, Wu W-S, Lord S. 2009: Introduction of the GSI into the NCEP Global Data Assimialtion System. *Wea. Forecasting* 2009; **24**:1691–1705.
23. Courtier P, Thépaut J-N, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* 1994; **120**:1367–1388.
24. Courtier P. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society* 1997; **123**:2449–2461.
25. Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research* 1995; **99**(C5):10143–10162.
26. Nino-Ruiz ED, Sandu A. Ensemble Kalman filter implementations based on shrinkage covariance matrix estimation. *Ocean Dynamics* 2015; **65**:1423–1439.
27. Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil D, Yorke JA. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* 2004; **56**:415–428.
28. Lorenc AC, Bowler NE, Clayton AM, Pring SR, David F. Comparison of Hybrid-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP. *Monthly Weather Review* 2015; **143**:212–229.
29. Brasseur P. Jacques Verron. *The SEEK filter method for data assimilation in oceanography: a synthesis. Ocean Dynamics* 2006; **56**:650–661.
30. Daley R. Richard Ménard, Spectral Characteristics of Kalman Filter Systems for Atmospheric Data Assimilation. *Monthly Weather Review* 1993; **121**:1554–1565.
31. Robert C, Durbiano S, Blayo E, Verron J, Blum J, Le Dimet FX. A reduced-order strategy for 4D-VAR data assimilation. *Journal of Marine Systems* 2005; **57**:70–82.
32. Ştefănescu R, Sandu A, Navon IM. POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation. *Journal of Computational Physics* 2015; **295**:569–595.
33. Wang X, Barker DM, Snyder C, Hamill TM. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Monthly Weather Review* 2008; **136**:5116–5131.
34. Li Z, Feng S, Liu Y, Lin W, Zhang M, Toto T, Vogelmann AM, Endo S. Development of Fine Resolution Analysis and Expanded Properties of Large-Scale Forcing. Part I: Methodology and Evaluations. *Journal. Geophysics. Research. Atmospheric* 2014; **120**:1–13.
35. Feng S, Li Z, Liu Y, Lin W, Zhang M, Vogelmann AM, Toto T, Endo S. Developmen of Fine Resolution Analysis and Expanded Properties of Large-Scale Forcing. Part II, Scale-Awareness and Application to Single Column Models. *Journal of Geophysical Research, [Atmospheres]* 2014; **120**(2014):667–677.
36. Stokes GM, Schwartz SE. The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation testbed. *Bulletin of the American Meteorological Society* 1994; **75**:1201–1221.
37. Mather JH, Voyles JW. The ARM Climate Research Facility: A Review of Structure and Capabilities. *Bulletin of the American Meteorological Society* 2013; **94**:377–392.
38. Haugen J-E, Machenhauer B. A spectral limited-area model formulation with time-dependent boundary conditions applied to the shallow-water equations. *Monthly Weather Review* 1993; **121**:2631–2636.
39. Berre L. Estimation of synoptic and mesoscale forecast error covariances in a limited area model. *Monthly Weather Review* 2000; **128**:644–667.
40. Chatfield C. *The Analysis of Time Series—An Introduction (sixth ed.)*. Chapman and Hall: London, 2003; 352.
41. Courtier P, Andersson E, Heckley W, Pailleux J, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F, Fisher M. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). *I: Formulation. Q. J. R. Meteorol. Society* 1998; **124**:1783–1807.
42. Li Z, Chao Y, McWilliams JC, Ide K. A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System. *Journal of Atmospheric and Oceanic Technology* 2008; **25**:2074–2090.
43. Li Z, Chao Y, McWilliams JC, Ide K. A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System: Implementation and basic experiments. *Journal Geophysicae. Research* 2008 C05002; **113**.
44. Boer GJ. Homogeneous and isotropic turbulence on sphere. *Journal of Atmospheric Science* 1983; **40**:154–163.
45. Jazwinski AH. *Stochastic processes and filtering theory*. Academic Press: New York, USA, 1970.
46. Lorenc AC. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society* 1986; **112**:1177–1194.
47. Li Z, McWilliams JC, Ide K, Fararra JD. Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme. *Ocean Dynamics* 2015; **65**:1001–1015.
48. Snyder C, Bengtsson T, Bickel P, Anderson J. Obstacles to high-Dimensional particle filtering. *Monthly Weather Review* 2008; **136**:4629–4640.
49. van Leeuwen PJ. Particle Filtering in Geophysical Systems. *Monthly Weather Review* 2010; **137**:4089–4114.
50. van Leeuwen PJ. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Q.J.R. Meteorology. Association.* 2010; **136**:1991–1999.
51. Graham A. *Kronecker Product and Matrix Calculus with Applications*. Ellis Horwood Ltd., 1981; 160.