Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 793–804
Printed in Singapore. All rights reserved

TELLUS

Measuring information content from observations for data assimilation: spectral formulations and their implications to observational data compression

By QIN XU*, *NOAA/National Severe Storms Laboratory, Norman, OK 73072-7326, USA*

(Manuscript received 18 January 2011; in final form 22 March 2011)

ABSTRACT

The previous singular-value formulations for measuring information content from observations are transformed into spectral forms in the wavenumber space for univariate analyses of uniformly distributed observations. The transformed spectral formulations exhibit the following advantages over their counterpart singular-value formulations: (i) The information contents from densely distributed observations can be calculated very efficiently even if the background and observation space dimensions become both too large to compute by using the singular-value formulations. (ii) The information contents and their asymptotic properties can be analysed explicitly for each wavenumber. (iii) Super-observations can be not only constructed by a truncated spectral expansion of the original observations with zero or minimum loss of information but also explicitly related to the original observations in the physical space. The spectral formulations reveal that (i) uniformly thinning densely distributed observations will always cause a loss of information and (ii) compressing densely distributed observations into properly coarsened super-observations by local averaging may cause no loss of information under certain circumstances.

1. Introduction

Remotely sensed observations collected by rapid scans from satellites and ground-based radars are characterized by their dense spatial distributions. Over their covered areas, these observations are often much denser than the model background and analysis grids used in data assimilations for numerical weather predictions. Since the spatial densities of these observations are far in excess of the resolution of the analysis, there can be a significant degree of resolution redundancy in terms of information content extractable from these observations by the data assimilation system. Such observation resolution redundancies were demonstrated by the illustrative examples with phased-array radar observations in Xu (2007) and Xu et al. (2009).

Redundant observation resolutions not only impose unnecessary computational burdens on a data analysis system but can also cause the cost function used in the analysis to be illconditioned. To reduce or eliminate information redundancy from observations for data assimilation, it is necessary to compress densely distributed observations into fewer super-observations with minimum possible loss of information. To address this issue, Purser et al. (2000) proposed some general methods for constructing super-observations by restricting the observation operator to a local volume (or area), and the Shannon entropy difference was the proposed information measure to quantify and compare the information contents from the original observations and super-observations. Xu (2007) further examined the above issue without restricting the observation operator to a local volume and proposed to use the relative entropy in addition to the Shannon entropy difference to measure the information contents from the original observations and super-observations. By applying the singular-value decomposition (SVD) to the scaled observation operator $\mathbf{M} \equiv \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$, Xu (2007) converted the general formulations of relative entropy and Shannon entropy difference into singular-value forms for measuring information content from observations. Facilitated by the singular-value formulations, the information redundancy from densely distributed observations can be explicitly quantified in connection with the rank deficiency of $\mathbf{M}$, and super-observations can be constructed with minimum or zero loss of information. However, since the constructed super-observations and their covariance are tied up with the SVD of $\mathbf{M}$, their connections to the original observations cannot be seen explicitly in the physical space. Besides, the SVD of $\mathbf{M}$ will become too expensive or impractical to compute if the background space and observation space are both very large, as often seen in real data assimilation (Parrish and Derber, 1992; Gao et al., 2004).

*Correspondence.
e-mail: Qin.Xu@noaa.gov
DOI: 10.1111/j.1600-0870.2011.00524.x

This paper aims to overcome the above shortcomings and limitations by transforming the singular-value formulations into wavenumber space for situations in which the observations are locally uniform and the observation and background error covariances or correlations (i.e. the covariances scaled by their respective variances) are locally homogeneous over the analysis area. In particular, with the spectral formulations derived in this paper, the information contents from either densely distributed original observations or their compressed super-observations can be computed efficiently and analysed easily for each wavenumber (as will be seen in Sections 2.2 and 3), and the super-observations can be explicitly related to the original observations in the physical space (as will be seen in Section 4.1). The paper is organized as follows. The spectral formulations are derived in the next section after a brief review of the previously derived singular-value formulations, and then used to examine the asymptotic properties of their measured information contents partitioned for each wavenumber in the limit of perfect background or perfect observations in Section 3. The implications of the spectral formulations to observational data compression are examined in Section 4. Conclusions follow in Section 5.

2. Spectral formulations of information measures for uniformly distributed observations

2.1. Review and generalization of singular-value formulations

As shown in Xu (2007, referred to as X07 hereafter) and Xu et al. (2009), the information extracted from observations by an analysis can be measured by the relative entropy defined by $R(p, q) \equiv \langle \ln(p/q) \rangle_p$, where $q = q(\mathbf{v})$ is the background probability density function (pdf), $p = p(\mathbf{v})$ is the analysis pdf, $\mathbf{v} \in R^N$ denotes the random state vector in the model space R^N, and $\langle () \rangle_p \equiv \int d\mathbf{v} p(\mathbf{v}) ()$ denotes the expectation of () based on $p(\mathbf{v})$. For a linear (or nearly linear) Bayesian optimal analysis with Gaussian pdfs, the relative entropy is given by

$$R(p, q) = J_b + \mathrm{SD} - DFS/2. \tag{1}$$

Here, $J_b = (\mathbf{a} - \mathbf{b})^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{a} - \mathbf{b})/2$ is the signal part of the relative entropy which has the same form as the background term in the cost-function of the optimal analysis, $\mathbf{a} \in R^N$ is the analysis mean, $\mathbf{b} \in R^N$ is the background mean, $\mathbf{B}$ is the background covariance matrix and $()^{\mathrm{T}}$ denotes the transpose of (). The last two terms in (1) are the dispersion part of the relative entropy. The first term in the dispersion part is identical to the Shannon entropy difference defined by

$$\mathrm{SD} \equiv \langle \ln p \rangle_p - \langle \ln q \rangle_q = \ln[\mathrm{Det}(\mathbf{B}^{1/2} \mathbf{A}^{-1} \mathbf{B}^{1/2})]/2, \tag{2}$$

where $S(q) = -\langle \ln q \rangle_q$, is the Shannon entropy of $q(\mathbf{v})$, Det() denotes the determinant of (), $\mathbf{A}$ is the analysis covariance matrix and $\mathbf{B}^{1/2}$ denotes the square root of $\mathbf{B}$ (satisfying $\mathbf{B}^{1/2} = \mathbf{B}^{\mathrm{T}/2}$). The second term in the dispersion part is related to the degrees of freedom for signal defined by

$$DFS \equiv \langle 2J_b \rangle = \mathrm{Tr}\left(\mathbf{I}_N - \mathbf{B}^{-1/2} \mathbf{A} \mathbf{B}^{-1/2}\right), \tag{3}$$

where $\langle () \rangle \equiv \int\int d\mathbf{v} d\mathbf{w} p(\mathbf{v}, \mathbf{w}) ()$ denotes the expectation of () based on the joint pdf $p(\mathbf{v}, \mathbf{w}) = q(\mathbf{v}) p(\mathbf{w}|\mathbf{v})$ [see (8)–(9) of Xu et al., 2009], $\mathbf{w} \in R^M$ is the observation vector, Tr() denotes the trace of (), and $\mathbf{I}_N$ denotes the identity matrix in R^N. As explained in section 2.4.2 of Rodgers (2000), *DFS* measures how many of degrees of freedom of an observation are related to signal (versus noise).

For a Bayesian optimal analysis, $\mathbf{a} - \mathbf{b} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{d}$ and $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ or, equivalently, $\mathbf{A} = \mathbf{B} - \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}$, where $\mathbf{d} = \mathbf{w} - H(\mathbf{b})$ is the innovation vector, $H()$ is the observation operator, $\mathbf{H}$ is the tangent-linearization of $H()$ at $\mathbf{v} = \mathbf{b}$, and $\mathbf{R}$ is the observation covariance matrix. These results can be rewritten into $\mathbf{a} - \mathbf{b} = \mathbf{B}^{1/2}\mathbf{M}^{\mathrm{T}}(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}})^{-1}\mathbf{R}^{-1/2}\mathbf{d}$, $\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2} = \mathbf{I}_N + \mathbf{M}^{\mathrm{T}}\mathbf{M}$ and $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} = \mathbf{I}_N - \mathbf{M}^{\mathrm{T}}(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}})^{-1}\mathbf{M}$, where $\mathbf{M} \equiv \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$ is the scaled observation operator and $\mathbf{I}_M$ is the identity matrix in the observation space R^M. Substituting the above results into J_b and (2)–(3) gives

$$J_b = \mathbf{d}^{\mathrm{T}}\mathbf{R}^{-1/2}\left(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}}\right)^{-1}\mathbf{M}\mathbf{M}^{\mathrm{T}}\left(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}}\right)^{-1}\mathbf{R}^{-1/2}\mathbf{d}/2, \tag{4}$$

$$\mathrm{SD} = \mathrm{Tr}[\ln\left(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}}\right)]/2, \tag{5}$$

$$DFS = \mathrm{Tr}\left[\mathbf{M}^{\mathrm{T}}\left(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}}\right)^{-1}\mathbf{M}\right], \tag{6}$$

where the identities $\ln[\mathrm{Det}(\mathbf{I}_N + \mathbf{M}^{\mathrm{T}}\mathbf{M})] = \ln[\mathrm{Det}(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}})] = \mathrm{Tr}[\ln(\mathbf{I}_M + \mathbf{M}\mathbf{M}^{\mathrm{T}})]$ are used (see e.g. http://en.wikipedia.org/wiki/Determinant) in the derivation of (5). The singular value decomposition (SVD) of $\mathbf{M}$ can be expressed in general by $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}}$, where $()^{\mathrm{H}}$ denotes the Hermitian transport of (). Substituting this SVD into (4)–(6) gives

$$J_b = \sum_{i=1}^{r} |d_i'|^2 |\lambda_i|^2 (1 + |\lambda_i|^2)^{-2}/2, \tag{7}$$

$$\mathrm{SD} = \sum_{i=1}^{r} [\ln(1 + |\lambda_i|^2)]/2, \tag{8}$$

$$DFS = \sum_{i=1}^{r} |\lambda_i|^2 (1 + |\lambda_i|^2)^{-1}, \tag{9}$$

where $r = \mathrm{rank}(\mathbf{M})$, d_i' is the ith element of $\mathbf{d}' = \mathbf{U}^{\mathrm{H}}\mathbf{R}^{-1/2}\mathbf{d}$, and λ_i is the ith singular value, that is, the ith element of the diagonal matrix $\mathbf{\Lambda}$. The singular-value formulations derived in (4)–(9) are similar to those in X07 and Xu et al. (2009), but the diagonal matrix $\mathbf{\Lambda}$ and associated unitary matrices $\mathbf{U}$ and $\mathbf{V}$ can be complex, so the formulations are more general than those introduced in X07. This generalization is necessary for deriving the spectral formulations in the subsequent subsections.

2.2. *Spectral formulations for one-dimensional observations*

The formulations in (7)–(9) are based on the SVD of the scaled observation operator **M**. For operational data assimilations, the observation space and background space are both very large, so the SVD of **M** is too expensive to compute. However, for the purpose of examining observation resolution redundancy, the background variable can be restricted to (or transformed into) the same variable as the observed, so the analysis is univariate with a linear observation operator **H** that involves only spatial interpolations. In this case, if the observations are uniformly distributed and the error covariances (or correlations) for the observation and background fields (or observation and background fields scaled by their respective variances) are homogeneous over the analysis area (as often assumed in variational data assimilation), then the singular values of **M** and associated right singular vectors (in **U**) can be computed efficiently in the spectral space. This approach will be examined for one-dimensional observations first and then extended for observations in multidimensional spaces in the next subsection.

Assume that there are M observations uniformly distributed at $x_o = x_m = m\Delta x_o$ for $m = M_-, \ldots M_+$ over the range of $M_-\Delta x_o \leq x_o \leq M_+\Delta x_o$, where x_o is the coordinate for the one-dimensional observations, Δx_o is the observation spacing (which is also the range represented by each observation), $M_- \equiv \text{Int}[(1-M)/2]$, $M_+ \equiv \text{Int}[M/2]$, and Int[] denotes the integer part of []. It is easy to see that $M_- = (1-M)/2$ and $M_+ = (M-1)/2$ for an odd M, and $M_- = 1 - M/2$ and $M_+ = M/2$ for an even M, so $(M_+ - M_- + 1)\Delta x_o = M\Delta x_o = D$, where D is the length of the domain and the domain is centred at $x_o = 0$ for odd M or at $x_o = \Delta x_o/2$ for even M. The model background grid (which is also taken to be the analysis grid) has N grid points at $x = x_n = n\Delta x$ for $n = N_-, \ldots N_+$ over the range of $N_-\Delta x \leq x \leq N_+\Delta x$, where x is the coordinate for the background grid, Δx is the background grid spacing, $N_- \equiv \text{Int}[(1-N)/2]$ and $N_+ \equiv \text{Int}[N/2]$. It is also easy to see that $(N_+ - N_- + 1)\Delta x = N\Delta x = D = M\Delta x_o$. The background grid covers the same domain D as the observations, but the background coordinate origin (at $x = 0$) is not generally collocated with the observation coordinate origin (at $x_o = 0$). In the background coordinate, the location of the observation coordinate origin is at $x = \delta x \equiv (N_+ + 1/2)\Delta x - (M_+ + 1/2)\Delta x_o$. In general, $\delta x \neq 0$ unless $M = N$ (and thus $\Delta x_o = \Delta x$).

Denote by $d_m = d(x_m)$ the innovation (observation minus interpolated background) at the mth observation point over the domain D in the x_o-coordinate, and denote by $b_n = b(x_n)$ the background at the nth grid point over the same domain D but in the x-coordinate. Assume that the above domain D can be expanded periodically, so $d_{m\pm M} = d_m$ for any m and $b_{n\pm N} = b_n$ for any n. There are M observation points and N grid points over each periodic domain. The background field $b(x_n)$ for $n = N_-, \ldots N_+$ can be represented by vector $\mathbf{b} \in R^N$ and transformed to a complex vector $\mathbf{s} = \mathbf{F}_N\mathbf{b} \in C^N$ in the model-spectral space, where $\mathbf{F}_N$ denotes the $N \times N$ matrix of the normalized discrete Fourier transform (DFT) (see e.g. http://en.wikipedia.org/wiki/Discrete_Fourier_transform). The ith element in the nth column of $\mathbf{F}_N$ is

$$(\mathbf{F}_N)_{in} = N^{-1/2}\exp(-jk_i x_n) = N^{-1/2}\exp(-j2\pi in/N) \quad \text{for } n \text{ and } i = N_-, \ldots N_+,$$
$$(\mathbf{F}_N)_{in} = N^{-1/2}\exp(-jk_i x_n) = N^{-1/2}\exp(-j2\pi in/N) \quad \text{for } n \text{ and } i = N_-, \ldots N_+, \qquad (10)$$

where $j \equiv (-1)^{1/2}$ is the imaginary unit, $k_i = i\Delta k$ and $\Delta k \equiv 2\pi/D$ is the minimum resolvable wavenumber associated with the length D of the periodic domain. The ith element of $\mathbf{s}$ can be denoted by $s_i = s(k_i)$ for $i = N_-, \ldots N_+$.

Note that the DFT can be extended to all integers i (beyond $N_- \leq i \leq N_+$) and the resulting infinite sequence is a periodic extension of the DFT with period N. With this extension, we have $s_{i+N} = s_i$ or, equivalently, $s(k_i + k_N) = s(k_i)$, where $k_N \equiv N\Delta k = 2\pi/\Delta x$ is the maximum wavenumber determined by the background grid resolution. Since b_n is real, s_j obeys the symmetry, that is, $s_{-i} = s_i{}^*$, where $()^*$ denotes the complex conjugate of (). This means that the DFT output $\mathbf{s}$ for real input $\mathbf{b}$ is half redundant. Note that the zero-wavenumber element s_0 is purely real and, for even N, the element $s_{N+} = s_{N/2}$ associated with the Nyquist-wavenumber $k_{N/2} = \pi/\Delta x$ is also real, so there are exactly N non-redundant real elements in the complex output $\mathbf{s}$.

Similarly, the innovations d_m for $m = M_-, \ldots M_+$ can be represented by vector $\mathbf{d} \in R^M$ and transformed to a complex vector $\mathbf{c} = \mathbf{F}_M\mathbf{d} \in C^M$ in the observation-spectral space. The ith element of $\mathbf{c}$ can be denoted by $c_i = c(k_i)$ for $i = M_-, \ldots M_+$ over the main period, and c_i can be extended periodically by $c_{i+M} = c_i$ or, equivalently, $c(k_i + k_M) = c(k_i)$, where $k_M \equiv M\Delta k = 2\pi/\Delta x_o$ is the maximum wavenumber determined by the observation resolution. Again, since d_m is real, c_i obeys the symmetry, that is, $c_{-i} = c_i{}^*$, so $\mathbf{c}$ is also half redundant. One can verify that $\mathbf{F}_N$ and $\mathbf{F}_M$ are unitary matrices, that is,

$$\mathbf{F}_N^{\mathrm{H}} = \mathbf{F}_N^{1} \quad \text{and} \quad \mathbf{F}_M^{\mathrm{H}} = \mathbf{F}_M^{-1}. \qquad (11)$$

With the above periodic expansion, the random background error, denoted by $\varepsilon(x_n)$ and represented by vector $\boldsymbol{\varepsilon}$ can be transformed to $\mathbf{e} = \mathbf{F}_N\boldsymbol{\varepsilon}$. The background covariance matrix is given by $\mathbf{B} = \langle\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\rangle$ and can be represented by $\mathbf{S} \equiv \langle\mathbf{e}\mathbf{e}^{\mathrm{H}}\rangle = \mathbf{F}_N\langle\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\rangle\mathrm{F}_N^{\mathrm{H}} = \mathbf{F}_N\mathbf{B}\mathbf{F}_N^{\mathrm{H}}$ in the model-spectral space. The background covariance function, defined by $B(x, x') \equiv \langle\varepsilon(x)\varepsilon(x')\rangle$, is an even and periodic function of $x - x'$ satisfying $B(x, x') = B(x - x') = B(|x - x' \pm D|)$. Here, $B(x, x')$ is the continuous counterpart of $\mathbf{B}$, so the nth element in n'th column of $\mathbf{B}$ is given by $B_{nn'} = B(x_n - x_{n'}) = B(x_{n''})$ where $x_{n''} = x_n - x_{n'} = n''\Delta x$ and

$n'' = n - n'$. The ith element in i'th column of $\mathbf{S}$ is then given by

$$\begin{aligned} S_{ii'} &= N^{-1} \sum_{n=N_-}^{N_+} \sum_{n'=N_-}^{N_+} B(x_n - x'_n) \exp(-jk_i x_n) \exp(k_{i'} x_{n'}) \\ &= N^{-1} \sum_{n=N_-}^{N_+} \exp[-j(k_i - k_{i'})x_n] \sum_{n''=N_-}^{N_+} B(x_{n''}) \exp(-k_{i'} x_{n''}) \\ &= \delta_{ii'} S(k'_i) = \delta_{ii'} S(k_i), \end{aligned} \tag{12}$$

where $\delta_{ii'}$ is the Kronecker delta, $S(k_i) = \sum_{n=N_-}^{N_+} B(x_n) \exp(-jk_i x_n)$ is the power spectrum of $\boldsymbol{\varepsilon}$ computed by the non-normalized conventional DFT of $B(x_n)$. In the derivation of (12), $B(x_{n''}) = B(|x_{n''} \pm N\Delta x|)$ is used in the second step, and $N^{-1} \sum_{n=N_-}^{N_+} \exp[-j(k_i - k_{i'})x_n] = N^{-1} \sum_{n=N_-}^{N_+} \exp[-j\pi(i - i')n/N] = \delta_{ii'}$ for $N_- \le i \le N_+$ and $N_- \le i' \le N_+$ is used in the last step. As shown in (12), $\mathbf{S}$ is a diagonal matrix and its diagonal elements are given by $S(k_i)$. As the power spectrum of $\boldsymbol{\varepsilon}$, $S(k_i)$ is a real, non-negative and even function of k_i. The symmetry $S(k_i) = S(-k_i)$ implies that $\mathbf{S}$ is about half redundant. The non-redundant elements in $\mathbf{S}$ are given by

$$S(k_i) = 2 \sum_{n=0}^{N_+} B(x_n) q_n \cos(k_i x_n) \quad \text{for } i = 0, \ldots N_+, \tag{13}$$

where $q_n = 1$ for $1 \le n < N_q$ and $q_n = 1/2$ for $n = 0$ and $n = N_q \equiv \text{Int}[(N+1)/2]$. For even (or odd) N, (13) gives the type-1 (or type-5) discrete cosine transform of $B(x_n)$, called DCT-1 (or DCT-5) as in Martucci (1994).

Similarly, $\mathbf{R}$ can be represented by $\mathbf{C} = \mathbf{F}_M \mathbf{R} \mathbf{F}_M^{\text{H}}$ in the observation-spectral space and $\mathbf{C}$ is a diagonal matrix with its diagonal elements given by the power spectrum, denoted by $C(k_i)$, of the random observation error. Again, $C(k_i)$ is also a real, non-negative and even periodic function of k_i, so $\mathbf{C}$ is about half redundant. The non-redundant elements in $\mathbf{C}$ are given by:

$$C(k_i) = 2 \sum_{m=0}^{M_+} R(x_m) q_m \cos(k_i x_m) \quad \text{for } i = 0, \ldots M_+, \tag{14}$$

where $q_m = 1$ for $1 \le m < M_q$ and $q_m = 1/2$ for $m = 0$ and $m = M_q \equiv \text{Int}[(M+1)/2]$. For even (or odd) M, (14) gives the type-1 (or type-5) discrete cosine transform of $R(x_m)$. Since $\mathbf{F}_N$ and $\mathbf{F}_M$ are unitary matrices, $\mathbf{S} = \mathbf{F}_N \mathbf{B} \mathbf{F}_N^{\text{H}}$ and $\mathbf{C} = \mathbf{F}_M \mathbf{R} \mathbf{F}_M^{\text{H}}$ lead to

$$\mathbf{B} = \mathbf{F}_N^{\text{H}} \mathbf{S} \mathbf{F}_N \quad \text{and} \quad \mathbf{R} = \mathbf{F}_M^{\text{H}} \mathbf{C} \mathbf{F}_M, \tag{15}$$

$$\mathbf{B}^{1/2} = \mathbf{F}_N^{\text{H}} \mathbf{S}^{1/2} \mathbf{F}_N \quad \text{and} \quad \mathbf{R}^{1/2} = \mathbf{F}_M^{\text{H}} \mathbf{C}^{1/2} \mathbf{F}_M, \tag{16}$$

where $\mathbf{B}^{1/2}$ and $\mathbf{R}^{1/2}$ are the square roots of $\mathbf{B}$ and $\mathbf{R}$, respectively.

The observation operator $\mathbf{H}$ projects $\mathbf{b}$ (as well as $\boldsymbol{\varepsilon}$) from the model space R^N to the observation space R^M by an interpolation scheme in the physical space. When the model and observation spaces are transformed into their respective spectral spaces, the observation operator $\mathbf{H}$ is transformed to $\mathbf{F}_M^{-1} \mathbf{H} \mathbf{F}_N^{-1} = \mathbf{F}_M^{\text{H}} \mathbf{H} \mathbf{F}_N^{\text{H}}$, and this transformed operator projects $\mathbf{s}$ ($= \mathbf{F}_N \mathbf{b}$) from the model-spectral space C^N to the observation-spectral space C^M. When $N < M$, this projection expands $\mathbf{s}$ by adding zero-value elements $s(k_i) = 0$ for $M_- \le i < N_-$ and $N_+ < i \le M_+$, which gives an ideal interpolation in the physical space according to the well-known zero-padding theorem (see e.g. Smith, 2007; section 7.2.7). The trigonometric interpolation polynomial is not unique due to the periodicity of $s(k_i)$. The ideal trigonometric interpolation polynomial is unique since it is constructed from the non-redundant Fourier coefficients associated with the smallest possible wavenumbers. This ideal interpolation consists of sinusoids whose frequencies have the smallest possible magnitudes (and therefore minimizes the mean-square slope defined by $\int_D |db(x)/dx|^2 dx$), and it ensures the interpolation polynomial to remain real for real input $\mathbf{b}$. With this ideal interpolation, the continuous counterpart of $\mathbf{b}$ is constructed from $\mathbf{s}$ ($= \mathbf{F}_N \mathbf{b}$) in the following form:

$$\begin{aligned} b(x) &= N^{-1/2} \sum_{i=N_-}^{N_+} s_i \exp(jk_i x) \\ &= 2N^{-1/2} \sum_{i=0}^{N_+} [\text{Re}(s_i) q_i \cos(jk_i x) + \text{Im}(s_i) \sin(jk_i x)], \end{aligned} \tag{17}$$

where Re() and Im() denote the real and imaginary parts of (), respectively, $q_i = 1$ for $1 \le i < N_q$ and $q_i = 1/2$ for $i = 0$ and $i = N_q$, and the symmetry $s_{-i} = s_i^*$ is used. Note that $\text{Im}(s_i) = 0$ for $i = 0$ and $i = N_+$ for even N, so the number of non-zero real Fourier coefficients in the summation is always N [since it is given by $1 + 2N_+ = N$ for odd N and by $2 + 2(N_+ - 1) = N$ for even N], which is exactly the number of non-redundant real elements in $\mathbf{s}$.

By using (17), the mth element of $\mathbf{H}\mathbf{b} = \mathbf{H}\mathbf{F}_N^{\text{H}} \mathbf{s}$ can be interpolated from $b(x)$ at $x = x_m + \delta x$ (that is, the mth observation point at $x_o = x_m = m\Delta x_o$ in the x_o-coordinate) to yield $(\mathbf{H}\mathbf{b})_m = b(x_m) = N^{-1/2} \sum_{i=N_-}^{N_+} s_i e_i \exp(jk_i x_m)]$ where $e_i \equiv \exp(jk_i \delta x)$. This result shows that the ith column of $\mathbf{H}\mathbf{F}_N^{\text{H}}$ consists of M elements with its mth element given by $N^{-1/2} e_i \exp(jk_i x_m)$. Note that the ith column of $\mathbf{F}_M^{\text{H}} \mathbf{E}_M$ also consists of M elements but with its mth element given by $M^{-1/2} e_i \exp(jk_i x_m)$, where $\mathbf{E}_M \equiv \text{diag}(e_{M-}, \ldots e_{M+})$ is the diagonal matrix composed of e_i for $M_- \le i < M_+$. Thus, as long as i is in the range of $\mu_- \le i \le \mu_+$, the ith column of $\mathbf{H}\mathbf{F}_N^{\text{H}}$ is the same as the ith column of $\mathbf{F}_M^{\text{H}} \mathbf{E}_M$ but multiplied by $\beta^{1/2}$, where $\beta \equiv M/N = \Delta x/\Delta x_o$, $\mu_- \equiv \text{Int}[(1-\mu)/2]$, $\mu_+ \equiv \text{Int}[\mu/2]$ and $\mu \equiv \min(N, M)$. When $N \le M$, $N_- = \mu_- \le i \le \mu_+ = N_+$ is satisfied by all the columns (indexed by i from N_- and N_+) in $\mathbf{H}\mathbf{F}_N^{\text{H}}$, so the above relationship gives $\mathbf{H}\mathbf{F}_N^{\text{H}} = \beta^{1/2} \mathbf{F}_M^{\text{H}} \mathbf{E}_M \mathbf{I}_{MN}$ or, equivalently,

$$\mathbf{F}_M \mathbf{H} \mathbf{F}_N^{\text{H}} = \beta^{1/2} \mathbf{E}_M \mathbf{I}_{MN} \quad \text{for } N \le M, \tag{18}$$

where $\mathbf{I}_{MN}$ is a $M \times N$ matrix with its ii'th element is given by $(\mathbf{I}_{MN})_{ii'} = \delta_{ii'}$ for $M_- \le i \le M_+$ and $N_- \le i' \le N_+$. Clearly, for $N \le M$, $\mathbf{I}_{MN}$ contains only one non-zero submatrix $\mathbf{I}_\mu$ over the central μ rows (for $\mu_- \le m \le \mu_+$), where $\mathbf{I}_\mu$ is the identity matrix in R^μ. In this case, $\mathbf{I}_{MN}$ is a zero-padding operator that

projects a vector from R^N to R^M by adding zero elements on the two ends of the vector, while $\mathbf{I}_{NM} = \mathbf{I}_{MN}^{\mathrm{T}}$ is a truncation operator that projects a vector from R^M to R^N by truncating the extra elements on the two ends of the vector. When $N = M$, $\mathbf{H}$, $\mathbf{E}_M$ and $\mathbf{I}_{MN}$ all reduce to $\mathbf{I}_N = \mathbf{I}_M$ in (18).

When $N > M$, the right-hand side of (18) gives only the central M columns of $\mathbf{F}_M\mathbf{H}\mathbf{F}_N^{\mathrm{H}}$ in the range between M_- and M_+. For $i < M_-$ or $i > M_+$, the ith column of $\mathbf{H}\mathbf{F}_N^{\mathrm{H}}$ goes beyond the column range of $\mathbf{F}_M^{\mathrm{H}}\mathbf{E}_M$. In this case, the periodicity of $\exp(jk_ix_m) = \exp(jk_{i\pm M}x_m)$ and $e_i = e_{i\pm M}$ can be used to relate the ith column of $\mathbf{H}\mathbf{F}_N^{\mathrm{H}}$ to the i'th column of $\mathbf{F}_M^{\mathrm{H}}\mathbf{E}_M$, where $i' = i - lM$ is still in the range between M_- and M_+, $l = \mathrm{Int}[(i - M_+)/M] < 0$ for $i < M_-$, and $l = \mathrm{Int}[(i - M_-)/M] > 0$ for $i > M_+$. This gives $\mathbf{H}\mathbf{F}_N^{\mathrm{H}} = \beta^{1/2}\mathbf{F}_M^{\mathrm{H}}\mathbf{E}_M\mathbf{P}_{MN}$ or, equivalently,

$$\mathbf{F}_M\mathbf{H}\mathbf{F}_N^{\mathrm{H}} = \beta^{1/2}\mathbf{E}_M\mathbf{P}_{MN}, \tag{19}$$

where $\mathbf{P}_{MN} \equiv (\mathbf{P}_{L-}, \ldots \mathbf{P}_l, \ldots \mathbf{P}_{L+})$ is a $M\times N$ matrix, $\mathbf{P}_l = \mathbf{I}_M$ for $L_- < l < L_+$, $L_- \equiv \mathrm{Int}[(N_- - M_+)/M]$, $L_+ \equiv \mathrm{Int}[(N_+ - M_-)/M]$, $\mathbf{P}_{L-}$ is a $M\times l_-$ matrix containing only one non-zero submatrix $\mathbf{I}_{l-}$ over the lowest l_- rows with $l_- \equiv ML_- + M_+ - N_-$, and $\mathbf{P}_{L+}$ is a $M\times l_+$ matrix containing only one non-zero submatrix $\mathbf{I}_{l+}$ over the highest l_+ rows with $l_+ \equiv ML_+ + N_+ - M_-$. By redefining $\mathbf{P}_0 \equiv \mathbf{I}_{M\mu}$ in $\mathbf{P}_{MN}$, (18) is combined into (19). When $N \le M$, $\mathbf{P}_{MN}$ reduces to $\mathbf{P}_0 = \mathbf{I}_{M\mu} = \mathbf{I}_{MN}$ and (19) reduces to (18). When $N > M$, $\mathbf{P}_{MN}$ is a collapsing operator as it maps a vector from R^N to R^M by shifting (by Nl positions) the vector elements outside the range of $M_- \le i \le M_+$ (on the two ends of the vector in R^N) inward and collapsing them into the range of $M_- \le i \le M_+$.

Substituting (16) and (19) into $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$ gives

$$\mathbf{M} = \mathbf{F}_M^{\mathrm{H}}\mathbf{C}^{-1/2}\mathbf{F}_M\mathbf{H}\mathbf{F}_N^{\mathrm{H}}\mathbf{S}^{1/2}\mathbf{F}_N = \mathbf{F}_M^{\mathrm{H}}\mathbf{\Gamma}_{MN}\mathbf{F}_N, \tag{20}$$

where $\mathbf{\Gamma}_{MN} \equiv \beta^{1/2}\mathbf{C}^{-1/2}\mathbf{E}_M\mathbf{P}_{MN}\mathbf{S}^{1/2}$. When $N \le M$, $\mathbf{P}_{MN}$ reduces to $\mathbf{I}_{M\mu}$. In this case, as shown by (18) and (20), $\mathbf{\Gamma}_{MN}$ is a diagonal matrix composed of the (complex) singular values of $\mathbf{M}$, while $\mathbf{F}_M^{\mathrm{H}}$ and $\mathbf{F}_N$ are the unitary matrices composed of the left and right singular vectors of $\mathbf{M}$, respectively. When $N > M$, $\mathbf{P}_{MN}$ is no longer a diagonal matrix, so the SVD of $\mathbf{M}$ can no longer be obtained from (20). However, by using (11) and (20), the squared absolute singular values of $\mathbf{M}$ can be obtained from the following eigenvalue decomposition of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$

$$\mathbf{M}\mathbf{M}^{\mathrm{T}} = \mathbf{F}_M^{\mathrm{H}}\mathbf{\Gamma}_{MN}\mathbf{\Gamma}_{MN}^{H}\mathbf{F}_M = \mathbf{F}_M^{\mathrm{H}}\beta\mathbf{\Pi}\mathbf{C}^{-1}\mathbf{F}_M, \tag{21}$$

where $\mathbf{\Pi} = \mathbf{E}_M\mathbf{P}_{MN}\mathbf{S}\mathbf{P}_{MN}^{\mathrm{T}}\mathbf{E}_M{}^{\mathrm{H}} = \mathbf{P}_{MN}\mathbf{S}\mathbf{P}_{MN}^{\mathrm{T}} = \sum_{l=L_-}^{L_+}\mathbf{P}_l\mathbf{S}_l\mathbf{P}_l{}^{\mathrm{T}}$ is a diagonal matrix in R^μ, $\mathbf{E}_M\mathbf{E}_M^{\mathrm{H}} = \mathbf{I}_M$ is used, and the diagonal matrix $\mathbf{S}$ is partitioned into $\mathrm{diag}(\mathbf{S}_{L-}, \ldots \mathbf{S}_{L+})$ in consistency with $\mathbf{P}_{MN} = (\mathbf{P}_{L-}, \ldots \mathbf{P}_{L+})$. When $N \le M$, $L_- = L_+ = 0$ and $\mathbf{S} = \mathbf{S}_0 = \mathrm{diag}(S_{\mu-}, \ldots S_{\mu+})$. When $N > M$, L_- and L_+ are at least not both zero, $\mathbf{S}_l = \mathrm{diag}(S_{lM+M_-}, \ldots S_{lM+M_+})$ for $L_- < l < L_+$, $\mathbf{S}_l = \mathrm{diag}(S_{N_-}, \ldots S_{lM+M_+})$ for $l = L_-$, and $\mathbf{S}_l = \mathrm{diag}(S_{lM+M_-}, \ldots S_{N_+})$ for $l = L_+$.

As we can see from (21), $\beta\mathbf{\Pi}\mathbf{C}^{-1}$ is a diagonal matrix composed of the eigenvalues of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$, while $\mathbf{F}_M^{\mathrm{H}}$ is the eigen-matrix of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ and therefore is still composed of the left singular vectors of $\mathbf{M}$ in consistency with (20). Thus, $\beta\mathbf{\Pi}\mathbf{C}^{-1}$ and $\mathbf{F}_M^{\mathrm{H}}$ are the spectral representations of $\mathbf{\Lambda}^2$ and $\mathbf{U}$, respectively, for the SVD of $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}}$ (see Section 2.1). Substituting (21) into (4)–(6) gives

$$\begin{aligned} J_b &= \sum_{i=\mu_-}^{\mu_+}\left(|c_i|^2/C_i\right)|\gamma_i|^2\left(1+|\gamma_i|^2\right)^{-2}/2 \\ &= \sum_{i=0}^{\mu_+} q_i\left(|c_i|^2/C_i\right)|\gamma_i|^2/\left(1+|\gamma_i|^2\right)^2 \end{aligned} \tag{22}$$

$$SD = \sum_{i=\mu_-}^{\mu_+}\left[\ln\left(1+|\gamma_i|^2\right)\right]/2 = \sum_{i=0}^{\mu_+} q_i\left[\ln\left(1+|\gamma_i|^2\right)\right] \tag{23}$$

$$DFS = \sum_{i=\mu_-}^{\mu_+}|\gamma_i|^2/\left(1+|\gamma_i|^2\right) = 2\sum_{i=0}^{\mu_+} q_i|\gamma_i|^2/\left(1+|\gamma_i|^2\right), \tag{24}$$

where $q_i = 1$ for $1 \le i < \mu_q$ and $q_i = 1/2$ for $i = 0$ and $i = \mu_q \equiv \min(N_q, M_q)$, $c_i = c(k_i)$ is the ith element of $\mathbf{c} = \mathbf{F}_M\mathbf{d}$, $C_i = C(k_i)$ is the ith diagonal element of $\mathbf{C}$, $|\gamma_i|^2 = \beta\Pi_i/C_i$, Π_i is the ith diagonal element of $\mathbf{\Pi}$, $\mathbf{F}_M\mathbf{R}^{-1/2}\mathbf{d} = \mathbf{F}_M\mathbf{C}^{-1/2}\mathbf{F}_M^{\mathrm{H}}\mathbf{F}_M\mathbf{d} = \mathbf{C}^{-1/2}\mathbf{c}$ [see (11) and (16)] is used in the derivation of (22), and the symmetry of $c_{-i} = c_i^*$, $C_{-i} = C_i$ and $S_{-i} = S_i$ [see (13) and (14)] are used in the second step of each equation. According to (21), the ith diagonal element of $\mathbf{\Pi}$ is given by

$$\Pi_i = \sum_{l=L_-}^{L_+} S(k_i + lk_M) \quad \text{for } \mu_- \le i \le \mu_+ \quad \text{and } N_- \le i + lM \le N_+. \tag{25}$$

When $N > M$, Π_i is the ith component of the background error spectrum collapsed into the range of the observation error spectrum, we may call Π_i the collapsed background error spectrum. When $N \le M$, $L_- = L_+ = 0$ and (25) reduces to $\Pi_i = S_i$.

By comparing (22)–(24) with (7)–(9), one can see that $|c_i|^2/C_i$ and $|\gamma_i|^2 = \beta\Pi_i/C_i$ are the spectral representations of $d'_i{}^2$ and $|\lambda_i|^2$, respectively, except that the index i is counted from μ_- to μ_+ (or, equivalently from 0 to μ_+ by considering the symmetry with respect to $i = 0$) rather than from 1 to $r \le \mu$ $(= \mu_+ - \mu_- + 1)$ as in (7)–(9). Note that $\beta\Pi_i$ is the background variance for wavenumber k_i in the observation-spectral space [see (36)], so $|\gamma_i|^2 = \beta\Pi_i/C_i$ is the ratio between the background and observation variances for wavenumber k_i in the observation-spectral space. Thus, in the observation-spectral space, the information contents can be computed efficiently by (22)–(24) for the univariate analysis under the condition stated at the beginning of this subsection.

2.3. Spectral formulations for two-dimensional observations

The spectral formulations derived above for one-dimensional observations can be extended for two-dimensional observations, and this extension will be presented in this subsection. A similar extension can be done for three-dimensional observations, and this will be briefly described at the end of this subsection.

To facilitate the extension, we assume that the two-dimensional observations are spaced uniformly by Δx_o and Δy_o, over the ranges of D_x and D_y along the locally selected x and y coordinates, respectively. The background grid covers the same range as the observations in each direction and spaced by Δx and Δy along the local x and y coordinates, respectively. The innovation and background fields can be then denoted by $d(m_x\Delta x_o, m_y\Delta y_o)$ and $b(n_x\Delta x, n_y\Delta y)$, respectively, where $(m_x\Delta x_o, m_y\Delta y_o)$ is the (m_x, m_y)th observation point and $(n_x\Delta x, n_y\Delta y)$ is the (n_x, n_y)th grid point. The observation points are counted from $m_x = M_{x-} \equiv \text{Int}[(1 - M_x)/2]$ to $M_{x+} \equiv \text{Int}[M_x/2]$ along the x_o-coordinate and from $m_y = M_{y-} \equiv \text{Int}[(1 - M_y)/2]$ to $M_{y+} \equiv \text{Int}[M_y/2]$ along the y_o-coordinate, while the background grid points are counted from $n_x = N_{x-} \equiv \text{Int}[(1 - N_x)/2]$ to $N_{x+} \equiv \text{Int}[N_x/2]$ along the x-coordinate and from $n_y = N_{y-} \equiv \text{Int}[(1 - N_y)/2]$ to $N_{y+} \equiv \text{Int}[N_y/2]$ along the y-coordinate. The analysis domain, denoted by $\Omega = \{(x, y)|\, N_{x-}\Delta x \le x \le N_{x+}\Delta x, N_{y-}\Delta y \le y \le N_{y+}\Delta y\}$, is thus covered by $M = M_x \times M_y$ observation points and $N = N_x \times N_y$ background grid points, where $M_x\Delta x_o = D_x = N_x\Delta x$ and $M_y\Delta y_o = D_y = N_y\Delta y$. Over Ω, the background field can be represented by $\mathbf{b} \in R^N$ and transformed to $\mathbf{s} = \mathbf{F}_N\mathbf{b} \in R^N$ in the model-spectral space, where $R^N = R^{Nx} \otimes R^{Ny}$, $\otimes$ denotes the tensor product of two vector spaces or two matrices (see e.g. http://en.wikipedia.org/ wiki/Tensor_product),

$$\mathbf{F}_N = \mathbf{F}_{Nx} \otimes \mathbf{F}_{Ny}, \tag{26}$$

and $\mathbf{F}_{Nx}$ (or $\mathbf{F}_{Ny}$) is the normalized DFT matrix for the transform in the x (or y) space constructed in the same way as for the one-dimensional case in (10). Similarly, the observation innovation can be represented by $\mathbf{d} \in R^M$ and transformed to $\mathbf{c} = \mathbf{F}_M\mathbf{d} \in R^M$ in the observation-spectral space, where $R^M = R^{Mx} \otimes R^{My}$ and $\mathbf{F}_M = \mathbf{F}_{Mx} \otimes \mathbf{F}_{My}$. One can verify that $\mathbf{F}_N$ and $\mathbf{F}_M$ are still unitary matrices and thus satisfy (11).

The state vector of the random background error $\varepsilon(n_x\Delta x, n_y\Delta y)$ over Ω can be still denoted by $\boldsymbol{\varepsilon} \in R^N$. By applying $\mathbf{F}_N$ to $\boldsymbol{\varepsilon}$, the spectral representation of the background covariance matrix can be derived by $\mathbf{S} = \mathbf{F}_N\langle\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\rangle\mathbf{F}_N^{\mathrm{H}} = \mathbf{F}_N\mathbf{B}\mathbf{F}_N^{\mathrm{H}}$. Since the background covariance is assumed to be locally homogeneous over Ω, its DFT over Ω can be analysed along each direction of (x, y) in the same way as in (12). With this extension, one can verify that $\mathbf{S}$ is a diagonal matrix in $R^N = R^{Nx} \otimes R^{Ny}$ and its (i, j)th diagonal element is the value of the power spectrum of $\boldsymbol{\varepsilon}$ at $(k_x, k_y) = (i\Delta k_x, j\Delta k_y)$, where $\Delta k_x = 2\pi/D_x$ and $\Delta k_y = 2\pi/D_y$ are the minimum resolvable wavenumbers in the x and y directions, respectively, and the double indices (i, j) are used to count the diagonal elements of $\mathbf{S}$ in $R^{Nx} \otimes R^{Ny}$. The power spectrum, denoted by $S_{ij} = S(i\Delta k_x, j\Delta k_y)$, is computed by applying the conventional DFT, along x and y, to the discretized background covariance function over Ω. Similarly, $\mathbf{R}$ can be represented by $\mathbf{C} = \mathbf{F}_M\mathbf{R}\mathbf{F}_M^{\mathrm{H}}$ in the observation-spectral space and $\mathbf{C}$ is a diagonal matrix with its diagonal elements given by the power spectrum, denoted by $C_{ij} = C(i\Delta k_x, j\Delta k_y)$, of the random observation error. With the above extensions, the non-symmetric square root matrices $\mathbf{B}^{1/2}$ and $\mathbf{R}^{1/2}$ can be constructed in the same way as in (16) except that $\mathbf{F}_N = \mathbf{F}_{Nx} \otimes \mathbf{F}_{Ny}$ and $\mathbf{F}_M = \mathbf{F}_{Mx} \otimes \mathbf{F}_{My}$.

For the above two-dimensional observations, the observation operator $\mathbf{H}$ is extended to $\mathbf{H} = \mathbf{H}_x \otimes \mathbf{H}_y$, where $\mathbf{H}_x$ (or $\mathbf{H}_y$) is the observation operator for one-dimensional interpolation along x (or y) and the interpolation is the same as described in the previous subsection. The spectral representations of $\mathbf{H}_x$ and $\mathbf{H}_y$ can be derived in the same way as in (19). Substituting these spectral representations into $\mathbf{F}_M\mathbf{H}\mathbf{F}_N^{\mathrm{H}} = \mathbf{F}_{Mx}\mathbf{H}_x\mathbf{F}_{Nx}^{\mathrm{H}} \otimes \mathbf{F}_{My}\mathbf{H}_y\mathbf{F}_{Ny}^{\mathrm{H}}$ gives

$$\mathbf{F}_M\mathbf{H}\mathbf{F}_N^{\mathrm{H}} = (\beta_x\beta_y)^{1/2}\mathbf{E}_{Mx}\mathbf{P}_{MxNx} \otimes \mathbf{E}_{Mx}\mathbf{P}_{MyNy}, \tag{27}$$

where $\beta_x = M_x/N_x = \Delta x/\Delta x_o$, $\beta_y = M_y/N_y = \Delta y/\Delta y_o$, $\mathbf{E}_{Mx}$ and $\mathbf{E}_{My}$ are defined as $\mathbf{E}_M$ in (18) but with M replaced by M_x and M_y, respectively, $\mathbf{P}_{MxNx}$ and $\mathbf{P}_{MyNy}$ are defined as $\mathbf{P}_{MN}$ in (19) but with M and N replaced, respectively, by M_x and N_x in $\mathbf{P}_{MxNx}$ and by M_y and N_y in $\mathbf{P}_{MyNy}$. Substituting the extended (16) and (27) into $\mathbf{M} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$ gives the same decomposition as in (20) except that $\boldsymbol{\Gamma}_{MN}$ is extended to $\boldsymbol{\Gamma}_{MN} = \boldsymbol{\Gamma}_{MxNx} \otimes \boldsymbol{\Gamma}_{MyNy}$. By using $(\mathbf{E}_{Mx} \otimes \mathbf{E}_{My})(\mathbf{E}_{Mx} \otimes \mathbf{E}_{My})^{\mathrm{H}} = \mathbf{I}_{Mx} \otimes \mathbf{I}_{My}$, the eigenvalue decomposition of $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ can be then derived in the same way as in (21) to give

$$\mathbf{M}\mathbf{M}^{\mathrm{T}} = \beta_x\beta_y\mathbf{F}_M^{\mathrm{H}}\boldsymbol{\Pi}\mathbf{C}^{-1}\mathbf{F}_M, \tag{28}$$

where $\boldsymbol{\Pi} = \sum_{l=L_{x-}}^{L_{x+}}\sum_{l'=L_{y-}}^{L_{y+}}(\mathbf{P}_l^x \otimes \mathbf{P}_{l'}^y)\mathbf{S}_{ll'}(\mathbf{P}_l^x \otimes \mathbf{P}_{l'}^y)^{\mathrm{T}}$ is a diagonal matrix in $R^{\mu x} \otimes R^{\mu y}$, $L_{x-} \equiv \text{Int}[(N_{x-} - M_{x+})/M_x]$, $L_{x+} \equiv \text{Int}[(N_{x+} - M_{x-})/M_x]$, $L_{y-} \equiv \text{Int}[(N_{y-} - M_{y+})/M_y]$, $L_{y+} \equiv \text{Int}[(N_{y+} - M_{y-})/M_y]$, $\mu_x \equiv \min(N_x, M_x)$, $\mu_y \equiv \min(N_y, M_y)$, $\mathbf{S}_{ll'}$ is the ll'th diagonal matrix partitioned from $\mathbf{S}$ in consistency with $\mathbf{P}_{MxNx} \otimes \mathbf{P}_{MyNy} = (\mathbf{P}^x{}_{Lx-}, \ldots \mathbf{P}^x{}_{Lx+}) \otimes (\mathbf{P}^y{}_{Ly-}, \ldots \mathbf{P}^y{}_{Ly+})$ and the partition is done in each (x or y) direction in the same way as described for $\mathbf{S}_l$ in (21).

Substituting (28) into (4)–(6) gives

$$\begin{aligned} J_b &= \sum_{\mathrm{i}=\mu_{x-}}^{\mu_{x+}}\sum_{\mathrm{j}=\mu_{y-}}^{\mu_{y+}}\left(|c_{ij}|^2/C_{ij}\right)|\gamma_{ij}|^2\left(1 + |\gamma_{ij}|^2\right)^{-2}/2 \\ &= 2\sum_{\mathrm{i}=0}^{\mu_{x+}}\sum_{j=0}^{\mu_{y+}} q_i q_j\left(|c_{ij}|^2/C_{ij}\right)|\gamma_{ij}|^2\left(1 + |\gamma_{ij}|^2\right)^{-2}, \end{aligned} \tag{29}$$

$$\begin{aligned} SD &= \sum_{\mathrm{i}=\mu_{x-}}^{\mu_{x+}}\sum_{j=\mu_{y-}}^{\mu_{y+}}\left[\ln\left(1 + |\gamma_{ij}|^2\right)\right]/2 \\ &= 2\sum_{i=0}^{\mu_{x+}}\sum_{j=0}^{\mu_{y+}} q_i q_j\left[\ln\left(1 + |\gamma_{ij}|^2\right)\right], \end{aligned} \tag{30}$$

$$DFS = \sum_{i=\mu_{x-}}^{\mu_{x+}} \sum_{j=\mu_{y-}}^{\mu_{y+}} |\gamma_{ij}|^2 \left(1+|\gamma_{ij}|^2\right)^{-1} = 4 \sum_{i=0}^{\mu_{x+}} \sum_{j=0}^{\mu_{y+}} q_i q_j |\gamma_{ij}|^2 \left(1+|\gamma_{ij}|^2\right)^{-1}, \quad (31)$$

where $q_i = 1$ for $1 \le i < \mu_{xq}$ but $q_i = 1/2$ for $i = 0$ and $i = \mu_{xq} \equiv \min\{\text{Int}[(N_x + 1)/2], \text{Int}[(M_x + 1)/2]\}$, $q_j = 1$ for $1 \le j < \mu_{yq}$ but $q_j = 1/2$ for $j = 0$ and $j = \mu_{yq} \equiv \min\{\text{Int}[(N_y + 1)/2], \text{Int}[(M_y + 1)/2]\}$, $c_{ij} = c(i\Delta k_x, j\Delta k_y)$ is the ijth element of $\mathbf{c} = \mathbf{F}_M\mathbf{d}$, $|\gamma_{ij}|^2 = \beta_x\beta_y\Pi_{ij}/C_{ij}$, $C_{ij} = C(i\Delta k_x, j\Delta k_y)$ is the ijth diagonal element of $\mathbf{C}$, and Π_{ij} is the ijth diagonal element of $\mathbf{\Pi}$. From (21), it is easy to see that the ijth diagonal element of $\mathbf{\Pi}$ is given by

$$\Pi_{ij} = \sum_{l=L_{x-}}^{L_{x+}} \sum_{l=L_{y-}}^{L_{y+}} S(i\Delta k_x + lk_{Mx},\, j\Delta k_y + l'k_{My}) \quad \text{for } \mu_{x-} \le i \le \mu_{x+}, \quad \mu_{y-} \le j \le \mu_{y+}, \quad N_{x-} \le i + lM_x \le N_{x+} \quad \text{and} \quad N_{y-} \le j + l'M_y \le N_{y+}, \quad (32)$$

where $k_{Mx} = M_x\Delta k_x$, $k_{My} = M_y\Delta k_y$, $\mu_{x-} \equiv \text{Int}[(1 - \mu_x)/2]$, $\mu_{x+} \equiv \text{Int}[\mu_x/2]$, $\mu_{y-} \equiv \text{Int}[(1 - \mu_y)/2]$, and $\mu_{y+} \equiv \text{Int}[\mu_y/2]$. When $N_x \le M_x$ and $N_y \le M_y$, (32) reduces to $\Pi_{ij} = S_{ij}$.

The spectral formulations can be similarly extended for three-dimensional observations that are distributed uniformly along each direction in a local Cartesian coordinate system over the analysis volume. In this case, the double summations in (29)–(32) will be replaced by their extended triple summations if the observation and background covariances (or correlations) are homogeneous, at least along each direction of the local Cartesian coordinate system. Since the required homogeneity conditions are too restrictive to be satisfied even approximately by real three-dimensional observations and background fields, the three-dimensional extension is not further explored in this paper. The required homogeneity conditions, however, can be satisfied approximately by two-dimensional radar observations (on each conical surface of radar scans) or satellite observations (over a satellite-scanned horizontal area), while the background covariance is nearly or locally homogeneous in the horizontal at each vertical level as often assumed in variational data assimilation (Daley, 1991; Parrish and Derber, 1992; Gao et al., 2004; Xu et al., 2010). The spectral formulations derived for two-dimensional observations in this subsection will be applied to real radar data compression in a follow-up paper.

3. Wavenumber-partitioned information contents and their asymptotic properties

The spectral formulations derived in Sections 2.2 and 2.3 reveal that the information contents can be partitioned and measured independently for each wavenumber in the observation-spectral space. For given wavenumber $k_i = i\Delta k$ (with $\mu_- \le i \le \mu_+$), the partitioned terms from (22)–(24) are

$$J_{bi} = \left(|c_i|^2/C_i\right)|\gamma_i|^2/\left(1+|\gamma_i|^2\right)^2 = \left[|c_i|^2/(\beta\Pi_i)\right]|\gamma_i|^4/\left(1+|\gamma_i|^2\right)^2, \quad (33)$$

$$SD_i = \ln\left(1+|\gamma_i|^2\right), \quad (34)$$

$$DFS_i = 2|\gamma_i|^2/\left(1+|\gamma_i|^2\right), \quad (35)$$

respectively, where $|\gamma_i|^2 = \beta\Pi_i/C_i$ is used in the derivation of the second expression in (33).

The first expression in (33) shows that $J_{bi} \propto |\gamma_i|^2/(1+|\gamma_i|^2)^2$ for given c_i and C_i. Note that $|\gamma_i|^2/(1+|\gamma_i|^2)^2 \to 0$ as $|\gamma_i|^2 \to 0$ or ∞. This implies that the signal part of the information content J_{bi} will approach zero as $\beta\Pi_i \to 0$ or ∞ for given $|c_i|$ and C_i. As explained in (22), c_i is the ith element of $\mathbf{c}$ (for $i = \mu_-, \ldots \mu_+$) and $\mathbf{c} = \mathbf{F}_M\mathbf{d}$ is the innovation vector in the observation-spectral space, so c_i is the innovation for wavenumber k_i. From (11)–(12) and (19), it is easy to see that $\mathbf{c} = \mathbf{F}_M(\boldsymbol{\varepsilon}^o - \mathbf{H}\boldsymbol{\varepsilon}) = \mathbf{F}_M\boldsymbol{\varepsilon}^o - \mathbf{F}_M\mathbf{H}\mathbf{F}_N{}^H\mathbf{e} = \mathbf{F}_M\boldsymbol{\varepsilon}^o - \beta^{1/2}\mathbf{E}_M\mathbf{P}_{MN}\mathbf{e}$, where $\boldsymbol{\varepsilon}^o$ and $\boldsymbol{\varepsilon}$ are the state vectors of random observation and background errors, respectively, and $\mathbf{e} = \mathbf{F}_N\boldsymbol{\varepsilon}$. This expression of $\mathbf{c}$ leads to

$$\langle \mathbf{c}\mathbf{c}^H \rangle = \mathbf{F}_M\mathbf{R}\mathbf{F}_M^H + \mathbf{F}_M\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{F}_M^H = \mathbf{C} + \beta\mathbf{\Pi}, \quad (36)$$

where $\mathbf{F}_M\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{F}_M^H = \mathbf{F}_M\mathbf{H}\mathbf{F}_N^H\mathbf{F}_N\mathbf{B}\mathbf{F}_N^H\mathbf{F}_N\mathbf{H}^T\mathbf{F}_M^H = \beta\mathbf{E}_M\mathbf{P}_{MN}\mathbf{S}\mathbf{P}_{MN}^T\mathbf{E}_M^H = \beta\mathbf{\Pi}$ is used [see (15)–(21)]. As shown in (36), the innovation covariance $\langle \mathbf{c}\mathbf{c}^H \rangle$ is the sum of the observation covariance $\mathbf{C}$ and background covariance $\beta\mathbf{\Pi}$, and this relationship is well known but is presented here in the observation-spectral space. According to (15) and (21), $\mathbf{C}$ and $\mathbf{\Pi}$ are diagonal matrices in R^μ, so $\langle \mathbf{c}\mathbf{c}^H \rangle$ is also a diagonal matrix in R^μ according to (36) and the ith diagonal element of $\langle \mathbf{c}\mathbf{c}^H \rangle$ is given by $\langle |c_i|^2 \rangle = C_i + \beta\Pi_i$ (for $i = \mu_-, \ldots \mu_+$). The innovation variance is thus the sum of the observation error variance and background error variance for each wavenumber in the observation-spectral space.

In the limit of $\beta\Pi_i \to 0$, the background state becomes perfectly accurate (with no error), so no additional information is needed or can be gained from the observations for wavenumber k_i. This explains why $J_{bi} \to 0$ as $\beta\Pi_i \to 0$ for given c_i and C_i. In the limit of $\beta\Pi_i \to \infty$, the background error becomes infinitely large, so $\langle |c_i|^2 \rangle = C_i + \beta\Pi_i \propto \beta\Pi_i \to \infty$ according to (36). In this case, if $|c_i|$ is finite for an individual realization in this limit, then the implied observation error for this realization must be infinitely large and the signal part of information content gained from the observations for this realization will become infinitely small for the ith wavenumber. This explains why $J_{bi} \to 0$ as $\beta\Pi_i \to \infty$ for given (finite) $|c_i|$ and C_i. Now, if $|c_i|^2$ is not fixed but becomes proportional to $\beta\Pi_i$ (as it should be statistically)

in the limit of $\beta\Pi_i \to \infty$ with fixed C_i, then $J_{bi} \to |\gamma_i|^4/(1 + |\gamma_i|^2)^2 \to 1$ as $|\gamma_i|^2 \to \infty$. This result is consistent with the asymptotic limit of (35), that is, $DFS_i \to 2$ for $|\gamma_i|^2 \to \infty$, since $\langle 2J_{bi}\rangle = DFS_i$ according to (3).

In addition to the above properties, it is also notable that $|\gamma_i|^2/(1 + |\gamma_i|^2)^2$ reaches the maximum value of $\frac{1}{4}$ when $|\gamma_i|^2 = 1$. This implies that the signal part of the information content J_{bi} will reach the maximum value of $(1/4)|c_i|^2/C_i$ if $\beta\Pi_i = C_i$ or, equivalently, $\Delta x\Pi_i = \Delta x_o C_i$ for given $|c_i|$ and C_i. Note that the signal part J_{bi} measures the information gained from the observations that improves the mean state (for wavenumber k_i) in the pdf updated by the optimal analysis (see section 5.1 of X07). Thus, the above result implies physically that the mean state can be improved most significantly for wavenumber k_i with given $|c_i|$ and C_i if the background spacing weighted background error variance for wavenumber k_i in the model-spectral space (that is, $\Delta x\Pi_i$) equals to the observation spacing weighted observation error variance for wavenumber k_i in the observation-spectral space (that is, $\Delta x_o C_i$).

The last expression in (33) shows that $J_{bi} \propto |\gamma_i|^4/(1 + |\gamma_i|^2)^2$ is a monotonic increasing function of $|\gamma_i|^2$ for given $|c_i|$ and $\beta\Pi_i$, and this function approaches the upper limit of 1 as $|\gamma_i|^2 \to \infty$. This implies that the signal part J_{bi} of the information content increases monotonically toward $|c_i|^2/(\beta\Pi_i)$ as $C_i \to 0$ and thus $|\gamma_i|^2 \to \infty$ for wavenumber k_i with given $|c_i|$ and $\beta\Pi_i$. Again, according to (36), the limit of $C_i \to 0$ implies that $\langle |c_i|^2\rangle = C_i + \beta\Pi_i \to \beta\Pi_i$. Applying this property to the above asymptotic limit gives $\langle J_{bi}\rangle \to 1$ as $C_i \to 0$ for fixed $\beta\Pi_i$. This result is also consistent with the asymptotic limit of (35), that is, $DFS_i = \langle 2J_{bi}\rangle \to 2$ for $C_i \to 0$ and thus $|\gamma_i|^2 \to \infty$.

It is easy to see from (34) and (35) that SD_i, DFS_i and $SD_i - DSF_i/2$ all become zero as $|\gamma_i|^2 \to 0$. Physically, this means that no information content can be gained from the observations for the dispersion part of wavenumber k_i when either the background state becomes perfectly accurate (with no error) or the observation error becomes infinitely large for wavenumber k_i. Note also from (34) and (35) that SD_i and $SD_i - DSF_i/2$ both increase boundlessly as $|\gamma_i|^2 \to \infty$. Physically, this means that the dispersion part of the information content (measured by either the Shannon entropy difference SD_i or the dispersion part of the relative entropy $SD_i - DSF_i/2$) will become boundlessly large for wavenumber k_i if the background error becomes infinitely large or the observation error becomes infinitely small for wavenumber k_i.

The above asymptotic properties are obtained from the spectral formulations in (22)–(24) for one-dimensional observations. Similar asymptotic properties can be obtained for each wavenumber vector (k_i, k_j) from (29)–(31) for two-dimensional observations. For real-data applications, these asymptotic properties represent the behaviours of wavenumber-partitioned information contents in situations of $|\gamma_i|^2 \ll 1$ and $|\gamma_i|^2 \gg 1$ for one-dimensional observations and $|\gamma_{ij}|^2 \ll 1$ and $|\gamma_{ij}|^2 \gg 1$ for two-dimensional observations.

4. Observation resolution redundancy and data compression—Super-obbing

4.1. Resolution redundancy revealed by spectral formulations

In variational data assimilation (Daley, 1991; Parrish and Derber, 1992; Gao et al., 2004; Xu et al., 2010), the background error covariances are often parameterized by smooth analytical correlation functions, and their associated power spectra are virtually band-limited in the wavenumber spectral space (Hollingsworth and Lönnberg, 1986; Xu and Wei, 2001; Xu et al., 2007). Thus, there is an integer $\nu < N_+$ such that $S_i = 0$ (or negligibly small) for $i > \nu$. If the observations are dense with $M \geq N$, then $\nu < \mu_+ = N_+ \leq M_+$. In this case, Π_i reduces to S_i [see (25)], $\mathbf{P}_{MN}$ reduces to $\mathbf{I}_{MN}$ [see (18)–(19)], the rank of $\mathbf{M}$, denoted by $r \equiv \text{rank}(\mathbf{M})$, reduces from μ to $1 + 2\nu$ (or even to 2ν if the negative wavenumber for $i = -\nu$ is also truncated) and the range of the first summations in (22)–(24) reduces from (μ_-, μ_+) to (r_-, r_+), where $r_- \equiv \text{Int}[(1 - r)/2]$ and $r_+ \equiv \text{Int}[r/2] = \nu$.

In the above case (with $\gamma_i \neq 0$ only for $r_- \leq i \leq r_+$), only the central r components of $\mathbf{c}$ can contribute to the signal part in (22). Similar to (4.2)–(4.4) of X07, the analysis increment and covariance matrix can be rewritten into the following forms

$$\begin{aligned}\mathbf{a} - \mathbf{b} &= \mathbf{B}^{1/2}\mathbf{M}^{\mathrm{T}}\left(\mathbf{M}\mathbf{M}^{\mathrm{T}} + \mathbf{I}_M\right)^{-1}\mathbf{R}^{-1/2}\mathbf{d} \\ &= \mathbf{B}^{1/2}\mathbf{F}_N^{\mathrm{H}}\mathbf{I}_{Nr}\boldsymbol{\Gamma}_r^{\mathrm{H}}\left(\boldsymbol{\Gamma}_r\boldsymbol{\Gamma}_r^{H} + \mathbf{I}_r\right)^{-1}\mathbf{C}_r^{-1/2}\mathbf{c}_r, \end{aligned} \tag{37}$$

$$\begin{aligned}\mathbf{A} &= \mathbf{B} - \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B} \\ &= \mathbf{B} - \mathbf{B}^{1/2}\mathbf{F}_N^{\mathrm{H}}\mathbf{I}_{Nr}\boldsymbol{\Gamma}_{\mathrm{r}}^{\mathrm{H}}\left(\boldsymbol{\Gamma}_r\boldsymbol{\Gamma}_r^{H} + \mathbf{I}_r\right)^{-1}\boldsymbol{\Gamma}_{\mathrm{r}}\mathbf{I}_{\mathrm{rN}}\mathbf{F}_N\mathbf{B}^{1/2}, \end{aligned} \tag{38}$$

where $\boldsymbol{\Gamma}_r = \mathbf{I}_{rM}\boldsymbol{\Gamma}_{MN}\mathbf{I}_{Nr} = \beta^{1/2}\mathbf{C}_r{}^{-1/2}\mathbf{E}_r\mathbf{S}_r{}^{1/2}$, $\mathbf{S}_r = \mathbf{I}_{rN}\mathbf{S}\mathbf{I}_{Nr}$, $\mathbf{E}_r = \mathbf{I}_{rM}\mathbf{E}_M\mathbf{I}_{Mr}$, $\mathbf{C}_r = \mathbf{I}_{rM}\mathbf{C}\mathbf{I}_{Mr}$, $\mathbf{c}_r = \mathbf{I}_{rM}\mathbf{c}$, and (20) is used. Here, $\mathbf{I}_{Nr}$ and $\mathbf{I}_{Mr}$ are $N\times r$ and $M\times r$ matrices, defined in same way as the $M\times N$ matrix $\mathbf{I}_{MN}$ in (18), so they are zero-padding operators for $r < \mu$, while $\mathbf{I}_{rN} = \mathbf{I}_{Nr}^{\mathrm{T}}$ and $\mathbf{I}_{rM} = \mathbf{I}_{Mr}^{\mathrm{T}}$ are truncation operators. Thus, $\mathbf{c}_r = \mathbf{I}_{rM}\mathbf{c}$ is a truncated vector composed of the central r components of $\mathbf{c}$ for $r_- \leq i \leq r_+$ in association with the r non-zero diagonal components of $\mathbf{S}$ where $\boldsymbol{\Gamma}_r$ and $\mathbf{C}_r$ are truncated diagonal matrices composed of the central r diagonal components of $\boldsymbol{\Gamma}_{MN}$ and $\mathbf{C}$, respectively. As shown by (37)–(38), the above truncation does not degrade or change the analysis and its covariance matrix, so it should cause no information loss. The reason is simply because $\gamma_i \neq 0$ only for $r_- \leq i \leq r_+$ in (22)–(24) or (33)–(35). This result implies that there is a resolution redundancy for the densely distributed observations when $M \geq N > r$, so the M original observations can be compressed into r super-observations as shown above. Such a compression is called super-Obbing (Purser et al., 2000).

Note that $\mathbf{I}_{rM}\mathbf{F}_M\mathbf{H} = \beta^{1/2}\mathbf{E}_r\mathbf{I}_{rN}\mathbf{F}_N$ for $M \geq N$ according to (18), so $\mathbf{c}_r = \mathbf{I}_{rM}\mathbf{F}_M\mathbf{w} - \mathbf{I}_{rM}\mathbf{F}_M\mathbf{H}\mathbf{b} = \mathbf{I}_{rM}\mathbf{F}_M\mathbf{w} - \beta^{1/2}\mathbf{E}_r\mathbf{I}_{rN}\mathbf{F}_N\mathbf{b}$. The continuous counterpart of the innovation $\mathbf{d}$ can be

constructed from $\mathbf{c}_r$ in the same way as that for the background field in (17). By using the same analysis as in (17)–(18), one can verify that $\mathbf{d}_r = (r/M)^{1/2}\mathbf{F}_r{}^{\mathrm{H}}\mathbf{c}_r$ is the super-innovation vector generated by interpolating the above continuous innovation on r uniformly spaced super-observation points over the periodic domain D. Thus, applying $(r/M)^{1/2}\mathbf{F}_r{}^{\mathrm{H}}$ to $\mathbf{c}_r = \mathbf{I}_{rM}\mathbf{F}_M\mathbf{w} - \beta^{1/2}\mathbf{E}_r\mathbf{I}_{rN}\mathbf{F}_N\mathbf{b}$ gives

$$\mathbf{d}_r = \mathbf{H}_{rM}\mathbf{w} - \mathbf{H}_{rN}\mathbf{b}, \tag{39}$$

where $\mathbf{H}_{rM} = (r/M)^{1/2}\mathbf{F}_r{}^{\mathrm{H}}\mathbf{I}_{rM}\mathbf{F}_M$ is the super-Obbing operator and $\mathbf{H}_{rN} = (r/N)^{1/2}\mathbf{F}_r{}^{\mathrm{H}}\mathbf{E}_r\mathbf{I}_{rN}\mathbf{F}_N$ is the super-observation operator. The super-observation covariance matrix is then given by $\mathbf{R}_r = \mathbf{H}_{rM}\mathbf{R}\mathbf{H}_{rM}{}^{\mathrm{H}}$. The smth element of $\mathbf{H}_{rM}$ is given by

$$(\mathbf{H}_{rM})_{sm} = H_{rM}(x_s - x_m) \equiv \mathrm{M}^{-1}\sum_{i=r_-}^{r_+}\exp[jk_i(x_s - x_m)]$$
$$\text{for } r_- \le s \le r_+ \quad \text{and} \quad M_- \le m \le M_+, \tag{40}$$

where $x_m = m\Delta x_o$, $x_s = s\Delta x_s$, and $\Delta x_s = (M/r)\Delta x_o$ is the super-observation resolution. Note that $H_{rM}(x) = k_M{}^{-1}\sum_{i=r_-}^{r_+}\exp(jk_ix)\Delta k \rightarrow 2k_M^{-1}\int_0^{k_r/2}\cos(kx)dk = (k_r/k_M)\mathrm{sinc}(k_rx/2) = (\Delta x_o/\Delta x_r)\mathrm{sinc}(\pi x/\Delta x_r)$ as $\Delta k = 2\pi/D \rightarrow 0$ (or $M \rightarrow \infty$) for fixed $k_M = M\Delta k = 2\pi/\Delta x_o$ and $k_r = r\Delta k = 2\pi/\Delta x_r$, where $\Delta x_r = D/r$ and $\mathrm{sinc}() \equiv \sin()/()$ is the sinc function of (). Thus, $H_{rM}(x)$ can be approximated by $(\Delta x_r/\Delta x_o)\mathrm{sinc}(\pi x/\Delta x_r)$, when M is very large. Similarly, the snth element of $\mathbf{H}_{rN}$ is given by $H_{rN}(x_s - x_n)$ where $x_s = s\Delta x_s$ and $x_n = n\Delta x$, and $H_{rN}(x) \equiv N^{-1}\sum_{i=r_-}^{r_+}\exp(jk_ix)$ can be approximated by $(\Delta x/\Delta x_r)\mathrm{sinc}(\pi x/\Delta x_r)$ when N is very large.

Substituting (40) into (39) and $\mathbf{R}_r = \mathbf{H}_{rM}\mathbf{R}\mathbf{H}_{rM}^{\mathrm{H}}$ gives the following super-Obbing formulations explicitly

$$d_r(x_s) = \sum_{m=M_-}^{M_+} H_{rM}(x_s - x_m)w(x_m) - \sum_{n=N_-}^{N_+} H_{rN}(x_s - x_n)b(x_n), \tag{41}$$

$$R_r(x_s - x_{s'}) = \sum_{m=M_-}^{M_+}\sum_{m=M_-}^{M_+} H_{rM}(x_s - x_m)R(x_m - x_{m'})H_{rM}(x_{m'} - x_{s'}), \tag{42}$$

where $d_r(x_s)$ is the sth component of $\mathbf{d}_r$ for $r_- \le s \le r_+$, and $R_r(x_s - x_{s'})$ is the ss'th component of $\mathbf{R}_r$ for $r_- \le s \le r_+$ and $r_- \le s' \le r_+$. By using the identity $M^{-1}\sum_{m=M_-}^{M_+}\exp(jk_ix_m) = \delta_{i0}$ for $M_- \le i \le M_+$, one can verify that $\sum_{m=M_-}^{M_+} H_{rM}(x_s - x_m) = 1$ and $\sum_{n=N_-}^{N_+} H_{rN}(x_s - x_n) = 1$. If observations errors are spatially uncorrelated, then $R(x_m - x_{m'}) = \sigma_o^2\delta_{mm'}$ where σ_o^2 denotes the observation error variance. Substituting this and (40) into (42) gives $R_r(x_s - x_{s'}) = \sigma_o^2 H_{rM}(x_s - x_{s'})$, where $M^{-1}\sum_{m=M_-}^{M_+}\exp[j(k_{i'} - k_i)x_m] = \delta_{ii'}$ for $M_- \le i \le M_+$ and $M_- \le i' \le M_+$ is used.

Note from (41) that the first summation on the right-hand side is the sth super-observation and the second summation is its associated background, and their difference gives the sth super-innovation $d_r(x_s)$ on the left-hand side. As shown by the first summation, each super-observation is a weighted sum of all the original observations and each weight depends sensitively on the precise location of its associated original observation. By using (41)–(42), the original M observations are compressed into r super-observations with no information loss. Similar super-Obbing formulations can be derived for the two-dimensional observations considered in Section 2.3. These super-Obbing formulations, however, require the exact conditions assumed by the derivations of the spectral formulations in Sections 2.2 and 2.3; that is, the observations must be uniformly spaced and their errors (as well as the background errors) must be bias-free and homogeneous. Since these conditions are often not precisely satisfied in real data assimilation, the above super-Obbing formulations cannot be directly and precisely used without information loss. Nevertheless, these super-Obbing formulations clearly reveal and precisely quantify the resolution redundancy for densely and uniformly distributed observations.

4.2. *Information loss caused by uniform thinning*

When observation errors are spatially uncorrelated, $\gamma_i^2 = \beta\Pi_i/\sigma_o^2$ in (22)–(24). For $M > N$ and thus $\beta = \Delta x/\Delta x_o > 1$, the upper limit of the summations in (22)–(24) is $\mu_+ = N_+$ independent of M and $\Pi_i = S_i$ in (25). Assume that $\beta = M/N = \Delta x/\Delta x_o$ is an integer (>1), so the observation resolution can be coarsened from Δx_o to $\Delta x_s = \Delta x = \beta\Delta x_o$ by thinning uniformly (to retaining one observation over each Δx), where Δx_s denotes the thinned observation resolution. This thinning does not affect Π_i ($= S_i$), but $|\gamma_i|^2 = \beta S_i/\sigma_o^2$ is reduced to $|\gamma_i|^2 = \beta_s S_i/\sigma_o^2 = S_i/\sigma_o^2$ where $\beta_s \equiv \Delta x/\Delta x_s = 1$. According to the analysis in Section 3 for fixed $\Pi_i = S_i$, the reduction of $|\gamma_i|^2$ indicates that the above thinning reduces the information contents measured by (33)–(35) and (22)–(24).

When observation errors are spatially correlated, $\mathbf{R}$ is no longer diagonal and $C_i = C(k_i)$ is an even function of i satisfying $\sum_{i=M_-}^{M_+} C_i = M\sigma_o^2$ according to the inverse of (14). Again, if the M observations are thinned uniformly into N observations with the observation resolution coarsened from Δx_o to $\Delta x_s = \Delta x$, then the thinned-observation vector is $\mathbf{T}_{NM}\mathbf{w}$ and the thinned-observation covariance matrix is $\mathbf{T}_{NM}\mathbf{R}\mathbf{T}_{NM}{}^{\mathrm{T}}$. Here, $\mathbf{T}_{NM}$ is the thinning operator defined by a $N \times M$ matrix with its smth element given by $(\mathbf{T}_{NM})_{sm} = \delta_{s'm}$ with $s' = s\beta$ for $N_- \le s \le N_+$ and $M_- \le m \le M_+$. By using the same analysis as in (18)–(19), one can verify that $\mathbf{F}_N\mathbf{T}_{NM}\mathbf{F}_M{}^{\mathrm{H}} = \beta^{-1/2}\mathbf{P}_{NM}$, where $\mathbf{P}_{NM}$ is defined as $\mathbf{P}_{MN}$ in (19) but with M and N interchanged. In the spectral space, the thinned-observation vector is $\mathbf{F}_N\mathbf{T}_{NM}\mathbf{F}_M{}^{\mathrm{H}}\mathbf{F}_M\mathbf{w} = \beta^{-1/2}\mathbf{P}_{NM}\mathbf{F}_M\mathbf{w}$ and the thinned-observation covariance matrix is $\mathbf{C}_N = \mathbf{F}_N\mathbf{T}_{NM}\mathbf{R}\mathbf{T}_{NM}{}^{\mathrm{T}}\mathbf{F}_N{}^{\mathrm{T}} = \mathbf{P}_{NM}\mathbf{C}\mathbf{P}_{NM}{}^{\mathrm{T}}/\beta$. One can verify that $\mathbf{C}_N$ is a diagonal matrix of size N and its ith diagonal element is given by $\sum_{l=\beta_-}^{\beta_+} C_{i+lN}/\beta$ for $N_- \le i \le N_+$, where $\beta_- \equiv \mathrm{Int}[(1-\beta)/2]$, $\beta_+ \equiv \mathrm{Int}[\beta/2]$, and the periodicity of

$C_i = C_{i\pm M}$ is used. Clearly, the above thinning still does not affect Π_i $(= S_i)$, but $\gamma_i^2 = \beta S_i/C_i$ is reduced to $\gamma_i^2 = \beta_s S_i/(\sum_{l=\beta_-}^{\beta_+} C_{i+lN}/\beta) = \beta S_i/(\sum_{l=\beta_-}^{\beta_+} C_{i+lN})$ since $\beta_s = 1$ and $\sum_{l=\beta_-}^{\beta_+} C_{i+lN} \geq C_i$ for $N_- \leq i \leq N_+$. Thus, when observation errors are correlated, thinning still reduces the information contents.

Note that $\beta_+ - \beta_- + 1 = \beta$ and $\mathbf{P}_{NM}\mathbf{P}_{NM}{}^{\mathrm{T}}/\beta = \mathbf{I}_N$. Thus, if $\mathbf{C} = \sigma_o^2\mathbf{I}_M$, then $\mathbf{C}_N = \mathbf{P}_{NM}\mathbf{C}\mathbf{P}_{NM}{}^{\mathrm{T}}/\beta = \mathbf{I}_N/\sigma_o^2$ and $\beta S_i/(\sum_{l=\beta_-}^{\beta_+} C_{i+lN}) = S_i/\sigma_o^2$. This recovers the result for uncorrelated observation errors. If observation errors are locally positively correlated, then C_i decreases as $|i|$ increases from 0 to M_+. In this case, $(\sum_{l=\beta_-}^{\beta_+} C_{i+lN})/C_i < \beta$, so $\Delta_s|\gamma_i|^2/[\beta S_i/C_i] = 1 - C_i/(\sum_{l=\beta_-}^{\beta_+} C_{i+lN}) < \Delta|\gamma_i|^2/(\beta S_i/\sigma_o^2) = 1 - 1/\beta$, where $\Delta_s|\gamma_i|^2 = \beta S_i/C_i - \beta S_i/(\sum_{l=\beta_-}^{\beta_+} C_{i+lN})$ and $\Delta|\gamma_i|^2 = (\beta S_i/\sigma_o^2 - S_i/\sigma_o^2)$ denote the reductions of $|\gamma_i|^2$ caused by the same thinning with correlated and uncorrelated observation errors, respectively. This result indicates that the information loss caused by the same thinning is smaller for observations with locally positively correlated errors than with uncorrelated errors. The above analysis and result (for $\Delta x_s = \Delta x$) can be extended and applied to other uniform thinning (with $\Delta x_s \neq \Delta x$).

4.3. Super-Obbing by local averaging

If observation errors are not spatially correlated and the observation resolution is coarsened from Δx_o to $\Delta x_s = \Delta x$ not by thinning but by local averaging over each Δx, then the error variance of the super-observations produced by the local averaging can be estimated by

$$\sigma_s^2 = \left\langle \left(\sum_{m=1}^{\beta} \varepsilon_m^{\circ}/\beta \right)^2 \right\rangle = \sigma_{\circ}^2/\beta, \tag{43}$$

where $\varepsilon^{\mathrm{o}}{}_m$ denotes the observation error at the mth observation point within each Δx, and $\beta = \Delta x/\Delta x_o$ is assumed again to be an integer larger that 1. When these super-observations are used in place the original observations, β and σ_o^2 are replaced by $\beta_s \equiv \Delta x/\Delta x_s = 1$ and σ_s^2, respectively, in (22)–(24), but the resulting $|\gamma_i|^2 = \beta_s S_i/\sigma_s^2 = \beta S_i/\sigma_o^2$ still has the same value as that obtained from the original observations. This implies that coarsening the observation resolution from Δx_o to Δx by the local averaging (over each Δx) does not affect the dispersion part of the information entropy (measured by SD, DSF or $SD - DSF/2$) and thus cause no information loss to the dispersion part. The signal part of the relative entropy, however, is more or less affected by the local averaging, because it depends on the individual realization of the innovation and the super-innovation $\mathbf{d}_s = \mathbf{w}_s - \mathbf{b}$ is not ensured to yield exactly the same value of J_b as the original innovation $\mathbf{d} = \mathbf{w} - \mathbf{H}\mathbf{b}$.

The above results are obtained for observations with spatially uncorrelated errors. When the observation errors are correlated between neighbouring observation points, as seen for radar observations (X07), the above local averaging can greatly reduce the error correlation (between neighbouring super-observation points) for its produced super-observations, but the information content extractable from the super-observations will be no longer generally be exactly the same as that from the original observations.

4.4. Remarks

(i) An observation error includes both measurement error and sampling error. The sampling error is an error of representativeness. This error can be caused by the limited spatial resolution in the calculation of the observation operator and it depends on the subgrid turbulent structures within the model effective resolution volume over which the true state is averaged. The model effective resolution, denoted by Δx_e, can be significantly coarser than the model grid resolution Δx in the horizontal direction (Frehlich, 2008). This effective resolution Δx_e $(> \Delta x)$ also limits the finest horizontal scale resolvable by the background error covariance, so the background error power spectrum $S(k_i)$ is essentially band-limited and should become zero (or set to zero) for $|k_i| > k_e$, where $k_e = \pi/\Delta x_e$ is the effective Nyquist-wavenumber. In this case, the super-observation resolution can be further coarsened to Δx_e (by averaging over each Δx_e) with no information loss. This property can be derived similarly as that in Section 4.3, except that the band-limited condition ($S_i = 0$ for $|k_i| > k_e$) is used here to ensure that Π_i still has the reduced form of S_i in (25) (although the number of the super-observations becomes smaller than N).

(ii) As mentioned in Section 4.1, the background error covariance used in variational data assimilation is often parameterized by simple and smooth analytical functions to capture the dominant error correlation patterns rather than resolve any fine structures. Because of this, the finest horizontal scale resolvable by the background error covariance, denoted by Δx_c [which is equivalent to $\Delta x_s = (M/r)\Delta x_o$ defined in (40)], can be coarser or much coarser than the model effective resolution Δx_e. The background error power spectrum $S(k_i)$ is thus further band-limited and can be set to zero for $|k_i| > k_c = \pi/\Delta x_c$. In this case, the super-observation resolution can be further coarsened to Δx_c by averaging over each Δx_c and Π_i still has the reduced form of S_i in (25), but the sampling error in the original observations will increase due to the fact that the true state is now averaged over Δx_c instead of Δx_e $(< \Delta x_c)$. Formally, the super-observation error variance can be estimated by $\sigma_o^2 \Delta x_o/\Delta x_c$ similarly to that in (43), but an increased value of σ_o^2 should be used to account for the increased sampling error. Because of this, further coarsening the super-observation resolution from Δx_e to Δx_c (by averaging over each Δx_c) can cause information loss, and the loss will increase as Δx_c becomes increasingly larger than Δx_e.

(iii) In this paper, the observation errors are assumed to be purely random with zero mean and homogenous covariance. When observations are not free of bias error with non-zero mean (or contain significant bias errors), the formulations

derived in Section 4.1 become inaccurate (or invalid). The local averaging considered in Section 4.3, however, may still be used for super-Obbing, but the bias error is not reduced by the averaging and must be considered together with the reduced random error estimated in (43). In this case, the information content extracted from the original observations cannot be accurately estimated by the spectral formulations derived in this paper. If the random error is inflated by adding the bias error and used in the formulations derived in this paper to consider approximately the effect of the bias error, then it is easy to see that the super-Obbing by local averaging proposed in Section 4.3 will cause some information loss, because the bias error is not reduced by the averaging.

5. Conclusions

In this paper, the singular-value formulations derived in X07 for measuring information content from observations are revisited and transformed into spectral forms in the wavenumber space for univariate analyses in which the observations are uniformly distributed and the observation and background error covariances (or correlations) are locally homogeneous over the analysis area. The transformed spectral formulations have precise correspondences to their counterpart singular-value formulations except that the index is counted with the increase of the wavenumber rather than with the decrease of the singular value [see (22)–(24) versus (7)–(9)]. In particular, for one-dimensional observations, if the ratio between the background error power spectrum S_i [or collapsed background error power spectrum Π_i if $N > M$, as shown in (25)] and observation error power spectrum C_i multiplied by their resolution ratio β ($= \Delta x/\Delta x_o$), that is, $|\gamma_i|^2$ ($= \beta\Pi_i/C_i$) is a decreasing function of the absolute wavenumber (indexed by $|i|$), then $|\gamma_i|$ for $i > 0$ (or $i \leq 0$) is precisely the absolute value of the $(2i)$th [or $(1 + 2i)$th] singular value in the singular-value formulations [see (7)–(9)]. Since the above power spectra can be computed efficiently by the Fast Fourier Transform or even derived analytically, the information contents from densely distributed observations (or compressed super-observations) can be easily calculated even if the background and observation spaces are both extremely large and too large to be computed by the singular-value formulations. This is an obvious advantage of the spectral formulations over their counterpart singular-value formulations.

Another advantage of the spectral formulations is that the information contents and especially their asymptotic properties can be analysed explicitly for each wavenumber in the limit of either $\gamma_i^2 \to 0$ or $\gamma_i^2 \to \infty$, and this is demonstrated in Section 3. As mentioned in the introduction, super-observations and their error covariance constructed by truncated singular-value expansions with zero or minimum loss of information are tied up with the SVD of the scaled observation operator $\mathbf{M}$ ($\equiv \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$), so their connections to the original observations are neither explicit nor transparent in the physical space. As shown in Section 4.1, the very same super-observations and their error covariances can be not only equivalently constructed by truncated spectral expansions but also explicitly related to the original observations and original observation error covariances in the physical space. This is the third advantage of the spectral formulations, although constructing super-observations by truncated spectral expansions has very limited utility (if any) without modification and/or without information loss (see the comment at the end of Section 4.1).

Facilitated by the spectral formulations, the analyses in Sections 4.2 and 4.3 reveal that uniformly thinning densely distributed original observations will always cause a loss of information, but compressing densely distributed observations into super-observations by local averaging with the observation resolution coarsened to the background grid resolution will cause no loss of information if the original observations are unbiased and their errors are not spatially correlated. As remarked in Section 4.4, compressing observations, called super-Obbing (Purser et al., 2000), by local averaging will also cause no loss of information if the super-observation resolution is further coarsened to the effective background resolution which can be significantly coarser than the background grid resolution (Frehlich, 2008). However, there can be a loss of information if the further coarsened super-observation resolution becomes resolvable by the background error covariance. The idea of super-Obbing by local averaging or weighted local averaging is not new, and efficient super-Obbing techniques were developed based on this idea and used for densely distributed radar observations (see e.g. Purser et al., 2000, Liu et al., 2005; Lu et al., 2011, section 3.3). Since real radar observations are not exactly uniformly distributed especially in the presence of data holes, the information loss caused by super-Obbing cannot be evaluated by the spectral formulations without approximation, but it can be accurately evaluated by the singular-value formulations if the observation or background space dimension is not very large.

Practical applications of the spectral formulations derived in this paper are limited by the assumed and required uniform observation distribution. Nevertheless, when densely distributed radar observations are treated roughly as uniformly distributed, the spectral formulations may still be used approximately to assess and compare the information losses (if any) caused by super-Obbing with successively coarsened resolutions and thus to find the optimally coarsened resolution for super-Obbing that causes no loss or no significant loss of information. Applications of the spectral formulations in this direction will be explored with real radar observations in a follow-up study.

6. Acknowledgments

The author is thankful to the anonymous reviewers for their comments and suggestions that improved the presentation of the results. The research work was supported by the ONR Grants

N000140410312 and N000141010778 to CIMMS, the University of Oklahoma.

References

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, New York, 457.

Frehlich, R. 2008. Atmospheric turbulence component of the innovation error covariance. *Quart. J. Roy. Meteor. Soc.* **134**, 931–940.

Gao, J., Xue, M., Brewster, K. and Droegemeier, K. K. 2004. A three-dimensional variational data assimilation method with recursive filter for single-Doppler radar. *J. Atmos. Oceanic. Technol.* **21**, 457–469.

Hollingsworth, A. and Lönnberg, P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. *Tellus* **38A**, 111–136.

Liu, S., Xue, M., Gao, J. and Parrish, D. 2005. Analysis and impact of super-obbed Doppler radial velocity in the NCEP grid-point statistical interpolation (GSI) analysis system. In: *Proceedings of the 17th Conference on Numerical Weather Prediction*. Amer. Meteor. Soc., Washington, DC, 13A.4.

Lu, H., Xu, Q., Yao, M. and Gao, S. 2011. Time-Expanded sampling for ensemble-based filters: assimilation experiments with real radar observations. *Adv. Atmos. Sci.* **28**, doi:10.1007/S00376-011-0185-5.

Martucci, S. A. 1994. Symmetric convolution and the discrete sine and cosine transforms. *IEEE Trans. Sig. Processing* SP-42, 1038–1051.

Parrish, D. and Derber, J. 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Purser, R. J., Parrish, D. F. and Masutani, M. 2000. Meteorological observational data compression; An alternative to conventional "super-Obbing". *Office Note 430*, National Centers for Environmental Prediction, Camp Springs, MD, 12 pp. Available at: http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html#2000 (last accessed 20 Apr 2011).

Rodgers, C. D. 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing Co. Ltd., London, 256.

Smith, J. O. 2007. *Mathematics of the Discrete Fourier Transform (DFT): with Audio Applications*. BookSurge Publishing. Available at: http://www.booksurge.com, 299 pp (last accessed 20 Apr 2011).

Xu, Q. 2007. Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. *Tellus* **59A**, 198–209.

Xu, Q., Nai, K. and Wei, L. 2007. An innovation method for estimating radar radial-velocity observation error and background wind error covariances. *Quart. J. Roy. Meteor. Soc.* **133**, 407–415.

Xu, Q. and Wei, L. 2001. Estimation of three-dimensional error covariances. Part II: analysis of wind innovation vectors. *Mon. Wea. Rev.* **129**, 2939–2954.

Xu, Q., Wei, L., Gu, W., Gong, J. and Zhao, Q. 2010. A 3.5-dimensional variational method for Doppler radar data assimilation and its application to phased-array radar observations. *Adv. Meteor.* **2010**, 14, doi:10.1155/2010/797265.

Xu, Q., Wei, L. and Healy, S. 2009. Measuring information content from observations for data assimilations: connection between different measures and application to radar scan design. *Tellus* **61A**, 144–153.