33 Multistep and Multi-scale Variational Data Assimilation: Spatial Variations of Analysis Error Variance

Qin Xu[1]* and Li Wei[2]

[1]NOAA/National Severe Storms Laboratory, Norman, Oklahoma
[2]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma

1. Introduction

Spectral formulations were derived (Xu et al. 2016, X16 hereafter) to efficiently compute the analysis error covariance for multistep and multi-scale variational data assimilation in which broadly distributed coarse-resolution observations are analyzed first and then locally distributed high-resolution observations are analyzed in the second step. However, the analysis error variance computed from these spectral formulations is constant and thus limited to represent only the spatially averaged error variance. When the coarse-resolution observations used in the first step of multi-step and multi-scale variational data assimilation become increasingly non-uniform and/or sparse, the error variance of the first-step analysis tends to have increasingly large spatial variations. In this case, it is necessary to overcome the limitation caused by the constant analysis error variance estimated from the spectral formulations. To this end, semi-empirical formulations are constructed in this paper to efficiently estimate the spatial variation of analysis error variance by properly combining the error variance reduction produced by analyzing each and every coarse-resolution observation as a single observation, and the estimated analysis error variance are used to further estimate the related variation in analysis error covariance. The detailed formulations are presented in the next three sections for one-dimensional cases with increased complexity and generality (from uniformly distributed observations with periodic extension to non-uniformly distributed observations without periodic extension). The improved accuracies of these formulations and their positive impacts on the two-step variational analysis are demonstrated by idealized experiments. Conclusions follow in section 5.

2. Error variance reduction produced by uniform coarse-resolution observations

When observations are optimally analyzed in terms of the Bayesian estimation (see chapter 7 of Jazwinski 1970), the background error covariance matrix $\mathbf{B}$ is updated to the analysis error covariance matrix $\mathbf{A}$ according to

$$\mathbf{A} = \mathbf{B} - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}, \qquad (1)$$

where $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{H}$ is the linearized observation operator. For a single observation, say, at x_m in the one-dimensional space of x, the inverse matrix $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ in (1) reduces to $(\sigma_b^2 + \sigma_o^2)^{-1}$, so the ith diagonal element of $\mathbf{A}$ is simply given by

$$\sigma_m^2(x_i) \equiv \sigma_b^2 - \gamma_b[\sigma_b C_b(x_i - x_m)]^2, \qquad (2)$$

where $\gamma_b = \sigma_b^2/(\sigma_b^2 + \sigma_o^2)$, σ_b^2 (or σ_o^2) is the background (or observation) error variance, $C_b(x)$ is the background error correlation function, x_i denotes the ith point in the discretized analysis space R^N, N is the number of grid points over the analysis domain. The length of the analysis domain is $D = N\Delta x$, where Δx is the analysis grid spacing and D is assumed to be much larger than the background error de-correlation length scale L.

Note that $C_b(x)$ is a continuous function of x, so (2) can be written into $\sigma_m^2(x) \equiv \sigma_b^2 - \Delta\sigma_m^2(x)$ also as a continuous function of x, where

$$\Delta\sigma_m^2(x) \equiv \gamma_b[\sigma_b C_b(x - x_m)]^2 \qquad (3)$$

is the error variance reduction produced by analyzing a single observation at $x = x_m$. The error variance reduction in (3) decreases rapidly as $|x - x_m|$ increases, and it becomes much smaller than it peak value of $\gamma_b \sigma_b^2 C_b^2$ at $x = x_m$ as $|x - x_m|$ increases to L. This implies that the error variance reduction produced by analyzing M sparsely distributed coarse-resolution observations can be estimated by properly combining the error variance reduction computed by (3) for each coarse-resolution observation as a single observation.

Assume that the M coarse-resolution observations are uniformly distributed with a resolution of $\Delta x_{co} \equiv D/M$ in the above analysis domain that can be extended periodically. In this case, the error variance reduction produced by each observation can be considered as an additional reduction to the reduction produced by its neighboring observations, and this additional reduction is always smaller than the reduction produced by the same observation but treated as a single observation. This implies that the error variance reduction produced by analyzing the M coarse-resolution observations, denoted by $\Delta\sigma_M^2(x)$, is bounded above by $\sum_m \Delta\sigma_m^2(x)$; that is,

$$\Delta\sigma_M^2(x) \leq \sum_m \Delta\sigma_m^2(x), \qquad (4)$$

where $\sum_m$ denotes the summation over m for the M observations. The equality in (4) is for the limiting case of $\Delta x_{co}/L \to \infty$ only. The inequality in (4) implies that the domain averaged value of $\sum_m \Delta\sigma_m^2(x)$ is larger than the true averaged reduction estimated by $\Delta\sigma_{be}^2 \equiv \sigma_b^2 - \sigma_e^2$, where σ_e^2 is the domain averaged analysis error variance estimated by the spectral formulation for one-dimensional cases in section 2.2 of X16.

The domain averaged value of $\sum_m \Delta\sigma_m^2(x)$ can be computed by

$$\Delta\sigma_{bs}^2 \equiv \int_D dx \sum_m \Delta\sigma_m^2(x)/D = \gamma_b \sigma_b^2 \sum_m \int_D dx C_b^2(x - x_m)/D$$
$$\approx \gamma_b \sigma_b^2 \sum_m \sum_i C_b^2(x_i - x_m)/N, \qquad (5a)$$

where $\int_D dx$ denotes the integration over the analysis domain, $\sum_i$ denotes the summation over i for the N grid points, and D

* Corresponding author address: Qin Xu, National Severe Storms Laboratory, 120 David L. Boren Blvd., Norman, OK 73072-7326; E-mail: Qin.Xu@noaa.gov

$= N\Delta x$ is used in the last step. By extending $C_b^2(x - x_m)$ with the analysis domain periodically, $\Delta\sigma_{bs}^2$ can be also estimated analytically as follows:

$$\begin{aligned} \Delta\sigma_{bs}^2 &\equiv \int_D dx \sum_m \Delta\sigma_m^2(x)/D \\ &= \gamma_b \sigma_b^2 \sum_m \sum_k \int_D dx C_b^2(x - x_m - kD)/D \\ &= \gamma_b \sigma_b^2 M \int dx C_b^2(x)/D = \gamma_b \sigma_b^2 I_1 L/\Delta x_{co}, \end{aligned} \quad (5b)$$

where $\int dx$ denotes the integration over the infinite space of x, $\sum_m \sum_k \int_D dx C_b^2(x - x_m - kD) = \sum_m \int dx C_b^2(x - x_m) = \sum_m \int dx C_b^2(x) = M \int dx C_b^2(x)$ is used in the second to last step, and $I_1 \equiv \int C_b^2(x) dx/L$ is used with $\Delta x_{co} \equiv D/M$ in the last step. For the double-Gaussian form of $C_b(x) = 0.6\exp(-x^2/2L^2) + 0.4\exp(-2x^2/L^2)]$ used in (5) of X16, we have $I_1 = (2\pi)^{1/2}(0.44/2^{1/2} + 0.48/5^{1/2})$. The analytically derived value in (5b) is very close to (slightly larger than) the numerically computed value from (5a). With the domain averaged value of $\sum_m \Delta\sigma_m^2(x)$ adjusted from $\Delta\sigma_{bs}^2$ to $\Delta\sigma_{be}^2$, $\Delta\sigma_M^2(x)$ can be estimated by

$$\Delta\sigma_M^2(x) = \sum_m \Delta\sigma_m^2(x) - \Delta\sigma_{bs}^2 + \Delta\sigma_{be}^2. \quad (6)$$

The analysis error variance, $\sigma_a^2(x)$, is then estimated by

$$\sigma_a^2(x) \approx \sigma_{a*}^2(x) \equiv \sigma_b^2 - \Delta\sigma_M^2(x). \quad (7)$$

As shown by the example in Fig. 1 (in which $D = 110.4$ km is the same as that in section 3 of X16 but M is reduced from 20 to 10 and thus $\Delta x_{co} = D/M = 11.04$ km is close to $L = 10$ km), the estimated $\sigma_{a*}^2(x)$ in (7) has nearly the same spatial variation as the benchmark $\sigma_a^2(x)$ that is computed precisely from (1), although the amplitude of spatial variation of $\sigma_{a*}^2(x)$, defined by $\max\sigma_{a*}^2(x) - \min\sigma_{a*}^2(x)$, is slightly smaller than that of the true $\sigma_a^2(x)$, defined by $\max\sigma_a^2(x) - \min\sigma_a^2(x)$.
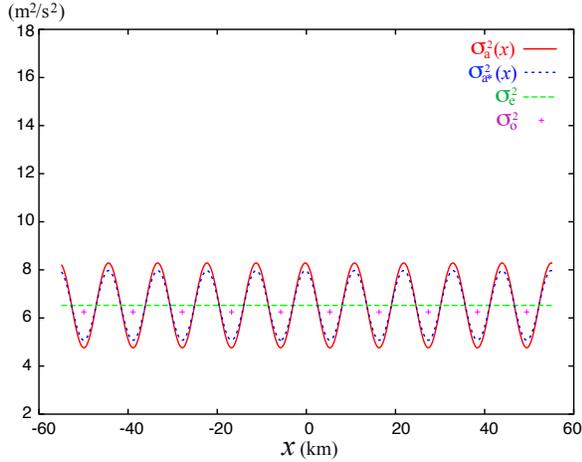
Fig. 1. Benchmark analysis error variance $\sigma_a^2(x)$ plotted by red solid curve, and estimated analysis error variance $\sigma_{a*}^2(x)$ in (7) plotted by blue dotted curve. The green dashed line shows the constant analysis error variance σ_e^2 estimated from the spectral formulation in section 2.2 of X16. The purple + signs show the observation error variance ($\sigma_o^2 = 2.5^2$ m^{2}s^{-2}) at the locations of M (= 10) uniformly distributed coarse-resolution observations with $\Delta x_{co} = D/M$ (= 11.04 km). The background error covariance $\sigma_b^2 C_b(x)$ is the same as the double-Gaussian form used in section 3 of X16; that

is, $C_b(x) = 0.6\exp(-x^2/2L^2) + 0.4\exp(-2x^2/L^2)]$ with $\sigma_b^2 = 5^2$ m^{2}s^{-2} and $L = 10$ km. The analysis domain and grid are the same as those described in section 3.1 of X16; that is, $D = N\Delta x = 110.4$ km with $N = 260$ and $\Delta x = 0.24$ km, but the number of coarse-resolution observations are reduced from $M = 20$ to 10.
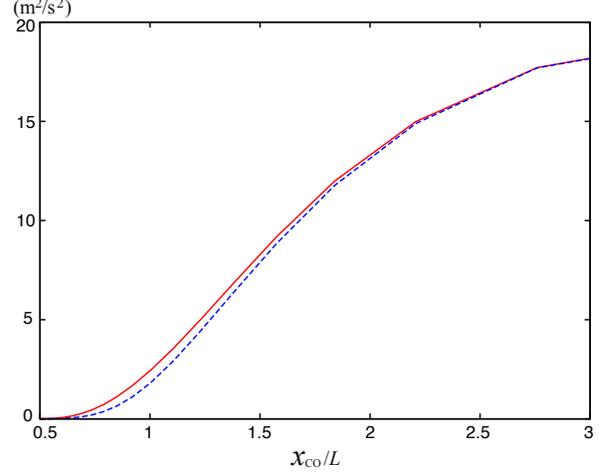
Fig. 2. Amplitude of spatial variation of benchmark $\sigma_a^2(x)$, defined by $\max\sigma_a^2(x) - \min\sigma_a^2(x)$, plotted by red solid curve as a function of $\Delta x_{co}/L$. Amplitude of spatial variation of estimated $\sigma_{a*}^2(x)$, defined by $\max\sigma_{a*}^2(x) - \min\sigma_{a*}^2(x)$, plotted by blue dotted curve as a function of $\Delta x_{co}/L$.

As shown in Fig. 2, the amplitude of spatial variation of benchmark $\sigma_a^2(x)$ decreases rapidly to virtually zero and then exactly zero (or increases monotonically toward its asymptotic upper limit of $\gamma_b \sigma_b^2 = 20$ m^{2}s^{-2}) as $\Delta x_{co}/L$ decreases to 0.5 and then to $\Delta x/L = 0.1$ (or increases toward ∞), and this decrease (or increase) of the amplitude of spatial variation of $\sigma_a^2(x)$ with $\Delta x_{co}/L$ is closely captured by the amplitude of spatial variation of the estimated $\sigma_{a*}^2(x)$ as a function of $\Delta x_{co}/L$.

Using the estimated $\sigma_{a*}(x)$ in (7), the previously estimated analysis error covariance matrix, denoted by $\mathbf{A}_e$ with its ij^{th} element given by $A_{eij} \equiv \sigma_e^2 C_a(x_i - x_j)$, from the spectral formulations in section 2.2 of X16 can be modified into $\mathbf{A}_a$, $\mathbf{A}_b$ or $\mathbf{A}_c$ with its ij^{th} element given by

$$\begin{aligned} A_{aij} &\equiv \sigma_{a*}(x_i)\sigma_{a*}(x_j)C_a(x_i - x_j), & (8a) \\ A_{bij} &\equiv \sigma_{a*}^2[(x_i + x_j)/2]C_a(x_i - x_j), & (8b) \\ \text{or} \quad A_{cij} &\equiv A_{eij} + \{\sigma_{a*}^2[(x_i + x_j)/2] - \sigma_e^2\}C_b(x_i - x_j). & (8c) \end{aligned}$$

The formulation in (8a) is conventional, as in (2.1) of Purser et al. (2003) or originally (11) of Rutherford (1972), in which the covariance is modified by applying $\sigma_{a*}(x)$ separately to each entry (indexed by i and j) of $C_a(x_i - x_j)$ to retain the self-adjointness. The expression of A_{aij} in (8a) can be viewed alternatively as A_{eij} plus a correction term of $[\sigma_{a*}(x_i)\sigma_{a*}(x_j) - \sigma_e^2]C_a(x_i - x_j)$. Ideally, a correction term should completely offset the deviation of A_{eii} from the true covariance, but the above correction term used in (8a) offsets only a part of the deviation.

2
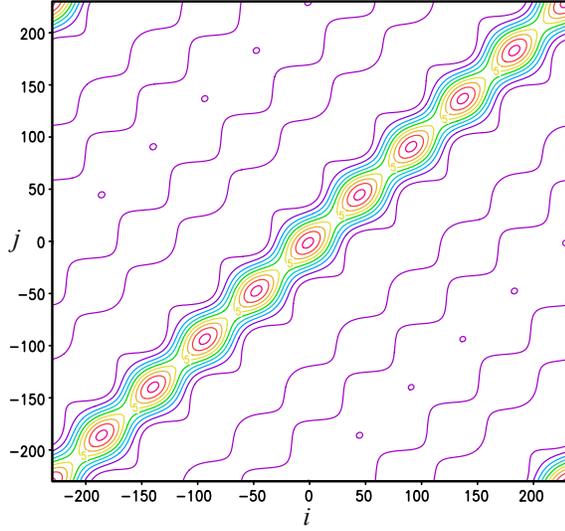
Fig. 3. Structure of benchmark **A** plotted by color contours every 1 m^{2}s^{-2} for the case in Fig. 1.

(a)

(b)

(c)

(d)
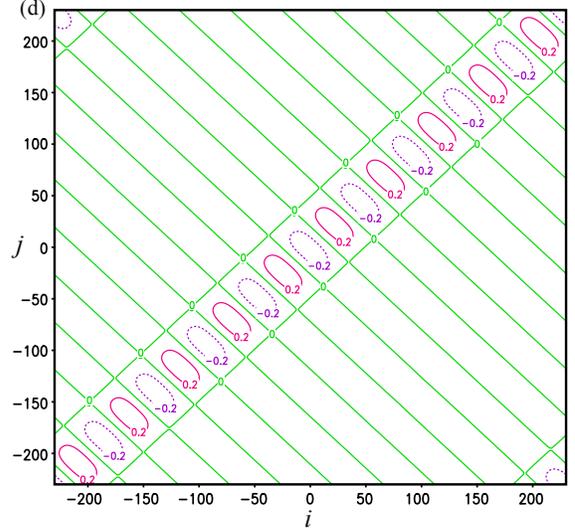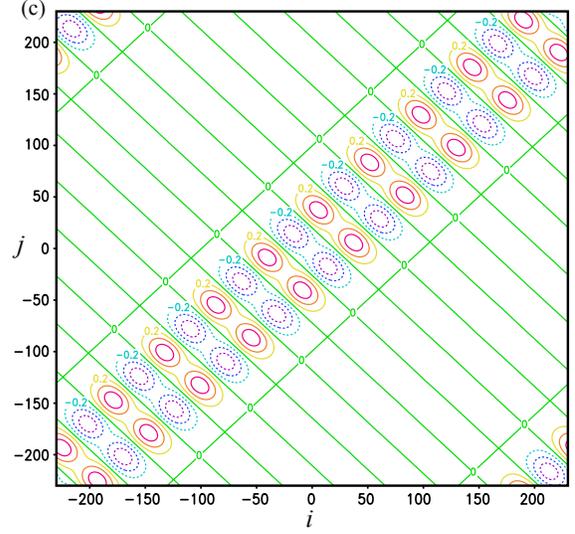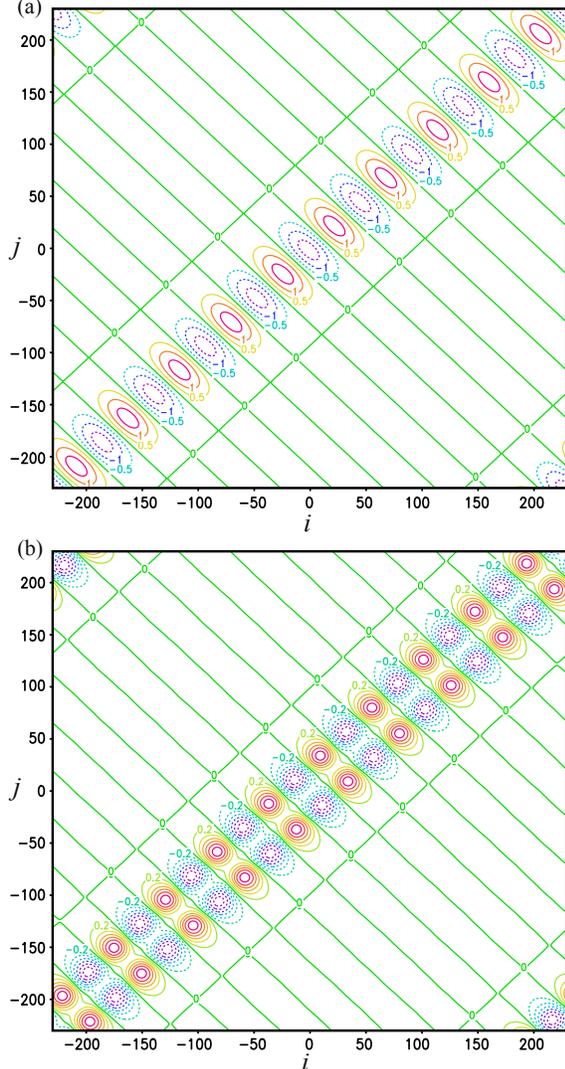
Fig. 4. (a) Deviation of **A**$_e$ from benchmark **A** in Fig 3 plotted by color contours every 0.5 m^{2}s^{-2}. Deviations of **A**$_a$, **A**$_b$ and **A**$_c$ from benchmark **A** are plotted by color contours every 0.2 m^{2}s^{-2} in panels (b), (c) and (d), respectively. Here, **A**$_e$ is the previously estimated analysis error covariance matrix with its ijth element given by $A_{eij} \equiv \sigma_e^2 C_a(x_i - x_j)$ as shown in section 2.2 of X16, while **A**$_a$, **A**$_b$ and **A**$_c$ are the newly modified estimates of **A** as shown in (8a), (8b) and (8c), respectively.

For the case in Fig. 1, the benchmark analysis error covariance matrix, denoted by **A**, is computed precisely from (1) and is plotted in Fig. 3, while the deviations of **A**$_e$, **A**$_a$, **A**$_b$ and **A**$_c$ from the benchmark **A** are shown in Figs. 4a, 4b, 4c, and 4d, respectively. As shown, the deviation becomes increasingly small when **A**$_e$ is modified successively to **A**$_a$, **A**$_b$ and **A**$_c$. Note that the aforementioned correction term used in (8a) is $C_a(x_i - x_j)$ modulated by $\sigma_{a*}(x_i)\sigma_{a*}(x_j) - \sigma_e^2$. This modulation has a chessboard structure, while the desired modulation revealed by the to-be-corrected deviation of **A**$_e$ in Fig. 4a has a banded structure (along the direction of $x_i + x_j =$ constant, perpendicular to the diagonal line). This explains

why the correction term in (8a) offsets only a part of the deviation as revealed by the deviation of $\mathbf{A}_a$ in Fig. 4b. On the other hand, the correction term used in (8b) is modulated by $\sigma_{a*}^2[(x_i + x_j)/2] - \sigma_e^2$. This modulation not only retains the self-adjointness but also has the desired banded structure, so this correction term is an improvement over that used in (8a), as shown by the deviation of $\mathbf{A}_b$ in Fig. 4c versus that of $\mathbf{A}_a$ in Fig. 4b. However, as revealed by Fig. 4c, the deviation of $\mathbf{A}_b$ still has two significant maxima (or minima) along each band on the two sides of the diagonal line of $x_i = x_j$, while the to-be-corrected deviation of $\mathbf{A}_e$ in Fig. 4a has a single maximum (or minimum) along each band. This implies that the function form of $C_a(x_i - x_j)$ is not sufficiently wide for the correction. As a further improvement, this function form is widened to $C_b(x_i - x_j)$ for the correction term in (8c), so the deviation of $\mathbf{A}_c$ in Fig. 4d is further reduced from that of $\mathbf{A}_b$ in Fig. 4c.

When an estimated $\mathbf{A}$ is used to update the background error covariance in the second step for analyzing the high-resolution observations in the nested domain, the accuracy of the second-step analysis depends not only, to a certain extent, on the number of iterations performed by the minimization algorithm but also on the accuracy of the estimated $\mathbf{A}$ over the nested domain plus its extended vicinities within the distance of $2L_a$ outside the nested domain. Here, L_a is the de-correlation length scale of $C_a(x)$ defined by $L_a^2 \equiv [-C_a(x)/d_x^2 C_a(x)]|_{x=0}$ [see (4.3.10) of Daley 1991], and L_a (= 4.45 km for the case in Figs. 1 and 3) can be easily computed as a by-product from the spectral formulation in (3) of X16. Over this extended nested domain, the relative error (RE) of the estimated $\mathbf{A}_e$ with respect to the benchmark $\mathbf{A}$ can be measured by

$$\text{RE}(\mathbf{A}_e) \equiv \|\mathbf{I}_s(\mathbf{A}_e - \mathbf{A})\mathbf{I}_s\|_F / \|\mathbf{I}_s\mathbf{A}\mathbf{I}_s\|_F, \qquad (9)$$

where $\mathbf{I}_s$ denotes the unit matrix in the subspace associated with the grid points in the extended nested domain and thus $\mathbf{I}_s(\mathbf{A}_e - \mathbf{A})\mathbf{I}_s$ (or $\mathbf{I}_s\mathbf{A}\mathbf{I}_s$) is the sub-matrix of $\mathbf{A}_e - \mathbf{A}$ (or $\mathbf{A}$) associated only with the grid points in the extended nested domain, and $\|( )\|_F$ denotes the Frobenius norm of ( ) defined by the square root of the sum of the squared absolute values of the elements of the matrix in ( ) [see (2.2-4) of Golub and Van Loan 1983]. The REs of $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ can be measured by the same form of Frobenius norm ratio as that defined for $\mathbf{A}_e$ in (9). The REs of $\mathbf{A}_e$, $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ are computed for the case in Fig. 1 and listed in the first column of Table 1. As shown by the listed values, the RE becomes increasingly small when $\mathbf{A}_e$ is modified successively to $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$, and this is consistent with and also quantifies the successively reduced deviation shown in Figs. 4a-4d.

To examine to what extent the successively improved estimate of $\mathbf{A}$ in (8) can improve the two-step analysis, idealized experiments are designed similarly to those in section 3.2 of X16, but here we have four types of two-step experiments, named TEe, TEa, TEb and TEc, in which $\mathbf{A}_e$, $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ are used, respectively, to update the background error covariance for analyzing the high-resolution innovations in the second step. The single-step experiment (in which all the innovations are analyzed together in a single step) is similar to that in X16 and is still named SE. The TEe is similar to the first type of two-step experiment (named TEA) in X16, but the TEa, TEb and TEc are new here. Besides, a set of simulated innovations is newly generated for the above five types of experiments. This new set, called the first set, consists of M (= 10) uniformly distributed coarse-resolution innovations over the analysis domain of length $D = N\Delta x = 110.4$ km (with $N = 460$ and $\Delta x = 0.24$ km as those in section 3 of X16) and M' (= 74) high-resolution innovations in the nested domain of length $D/6$ (similar to those shown by the purple × signs in Fig. 1 of X16 but generated at the grid points not covered by the coarse-resolution innovations within the nested domain). All the innovations are generated by simulated observation errors subtracting simulated background errors at observation locations. Observation errors are sampled from computer-generated uncorrelated Gaussian random numbers with $\sigma_o = 2.5$ ms^{-1} for both coarse-resolution and high-resolution observations. Background errors are sampled from computer-generated spatially-correlated Gaussian random fields [with $\sigma_b = 5$ ms^{-1} and $C_b(x)$ modeled by the double-Gaussian form in (5) of X16]. The coarse-resolution innovations are thus generated in consistency with the case in Figs. 1 and 3.

In the SE, the analysis increment is obtained [by minimizing the preconditioned cost-function similar to that in (7) of X16] with the number of iterations limited to $n = 20$, 50 or 100 before the final convergence to mimic the computationally constrained situations in operational data assimilation. In each two-step experiment, the analysis increment is obtained with the number of iterations also limited to $n = 20$, 50 or 100 before the final convergence. The accuracy of the analysis increment obtained from each experiment with each limited n is measured by its domain-averaged RMS error (called RMS error for short hereafter) with respect to the benchmark analysis increment [computed precisely from (1a) of X16]. Table 1 lists the RMS errors of the analysis increments obtained from the SE, TEe, TEa, TEb and TEc with the number of iterations increased from $n = 20$ to 50, 100 and/or the final convergence.

Table 1. Entire-domain averaged RMS errors (in ms^{-1}) for the analysis increments obtained from SE, TEe, TEa, TEb and TEc applied to the first set of innovations with periodic extension and consecutively increased n, where n is the number of iterations. All the RMS errors are evaluated with respect to the benchmark analysis increment. The relative error (RE) of the estimated analysis error covariance for updating the background error covariance in the second step of the two-step analysis is listed with the experiment name in the first column for each two-step experiment.

Experiment	$n = 20$	$n = 50$	$n = 100$	Final
SE	0.671	0.365	0.187	0.013 at $n = 481$
TEe $\text{RE}(\mathbf{A}_e) = 0.229$	0.171	0.150	0.142	0.135 at $n = 210$
TEa $\text{RE}(\mathbf{A}_a) = 0.156$	0.169	0.142	0.144	0.144 at $n = 116$
TEb $\text{RE}(\mathbf{A}_b) = 0.101$	0.147	0.098		0.090 at $n = 67$
TEc $\text{RE}(\mathbf{A}_c) = 0.042$	0.145	0.063	0.062	0.032 at $n = 176$

As shown in Table 1, the TEe outperforms SE for $n = 20$, 50 and 100 but not for n increased to the final convergence. The improved performance of TEe over SE is similar to but less significant than that of TEA over SE in Table 1 of X16.

The reduced improvement can be largely explained by the fact that the coarse-resolution innovations are generated here more sparsely than those used in section 3.2 of X16 and the deviation of $\mathbf{A}_e$ from the benchmark $\mathbf{A}$ is thus increased (as seen from Fig. 4a in comparison with Fig. 5b of X16). The TEa outperforms TEe for $n = 20$ and 50 before n increased to 100 (which is very close to the final convergence at $n = 116$ for TEa). The improvement of TEa over TEe is consistent with and can be largely explained by the improved accuracy of $\mathbf{A}_a$ [RE($\mathbf{A}_a$) = 0.156] over $\mathbf{A}_e$ [RE($\mathbf{A}_e$) = 0.229]. The TEb outperforms TEa for $n = 20$ and 50 (before the final convergence at $n = 67$). The improvement of TEb over TEa is consistent with the improved accuracy of $\mathbf{A}_b$ [RE($\mathbf{A}_b$) = 0.101] over $\mathbf{A}_a$. The TEc outperforms TEb for each listed value of n, and the improvement is consistent with the improved accuracy of $\mathbf{A}_c$ [with RE($\mathbf{A}_c$) = 0.042] over $\mathbf{A}_b$.

3. Error variance reduction by non-uniform coarse-resolution observations with periodic extension

Consider that the M coarse-resolution observations are now non-uniformly distributed in the analysis domain of length D with periodic extension. Their averaged resolution can be defined by $\Delta x_{co} \equiv D/M$. The spacing of a concerned coarse-resolution observation, say the mth observation, from its right (or left) adjacent observation can be denoted by Δx_{com+} (or Δx_{com-}). Now we can consider the following two limiting cases.

First, we consider the case of $\Delta x_{com+} \to 0$ with $\Delta x_{com-} = \Delta x_{co}$ (or $\Delta x_{com-} \to 0$ with $\Delta x_{com+} = \Delta x_{co}$). In this case, the concerned m^{th} observation collapses onto the same point with its right (or left) adjacent observation, that is, the $(m+1)^{th}$ [or $(m-1)^{th}$] observation. The two collapsed observations should be combined into one super-observation with a reduced error variance from σ_o^2 to $\sigma_o^2/2$. The error variance reduction produced by this super-observation still can be estimated by (3) but with

$$\gamma_b = \sigma_b^2/(\sigma_b^2 + \sigma_o^2/2). \qquad (10a)$$

On the other hand, without super-Obbing, the error variance reduction produced by the two collapsed observations will be over-estimated by (3) with

$$\gamma_b = 2\sigma_b^2/(\sigma_b^2 + \sigma_o^2) = \sigma_b^2/(\sigma_b^2/2 + \sigma_o^2/2). \qquad (10b)$$

By comparing (10b) with (10a), it is easy to see that this overestimation can be corrected if the error variance is inflated from σ_o^2 to $\sigma_o^2 + \sigma_b^2$ for each of the two collapsed observations.

Then, we consider the case of $\Delta x_{com+} \to 0$ and $\Delta x_{com-} \to 0$. In this case, the concerned m^{th} observation collapses with its two adjacent observations, that is, the $(m+1)^{th}$ and $(m-1)^{th}$ observations. The three collapsed observations should be combined into one super-observation with a reduced error variance from σ_o^2 to $\sigma_o^2/3$. The error variance reduction produced by this super-observation still can be estimated by (3) but with

$$\gamma_b = \sigma_b^2/(\sigma_b^2 + \sigma_o^2/3). \qquad (10c)$$

On the other hand, without super-Obbing, the error variance reduction produced by the three collapsed observations will be over-estimated by (3) with

$$\gamma_b = 3\sigma_b^2/(\sigma_b^2 + \sigma_o^2) = \sigma_b^2/(\sigma_b^2/3 + \sigma_o^2/3). \qquad (10d)$$

By comparing (10d) with (10c), it is easy to see that this overestimation can be corrected if the error variance is inflated from σ_o^2 to $\sigma_o^2 + 2\sigma_b^2$ for each of the three collapsed observations.

Based on the above analyses, when the error variance reduction produced by the mth observation is estimated by (3), the error variance should be adjusted for this observation unless $\Delta x_{com+} = \Delta x_{com-} = \Delta x_{co}$. In particular, its error variance can be adjusted from σ_o^2 to $\sigma_{om}^2 = \sigma_o^2 + \beta_m \sigma_b^2$ with β_m given empirically by

$$\beta_m = [C_b^2(\Delta x_{com+}) + C_b^2(\Delta x_{com-}) - 2C_b^2(\Delta x_{co})]/[1 - C_b^2(\Delta x_{co})]. \qquad (11a)$$

Note that $\beta_m = 2$ for $\Delta x_{com+} = \Delta x_{com-} = 0$, so the adjusted error variance is $\sigma_{om}^2 = \sigma_o^2 + 2\sigma_b^2$ which recovers the result derived from (10c)-(10d). Note also that $\beta_m = 1$ for $\Delta x_{com+} = 0$ and $\Delta x_{com-} = \Delta x_{co}$ (or $\Delta x_{com-} = 0$ and $\Delta x_{com+} = \Delta x_{co}$), so the adjusted error variance is $\sigma_{om}^2 = \sigma_o^2 + \sigma_b^2$ which recovers the result derived from (10a)-(10b). Clearly, for $\Delta x_{com-} = \Delta x_{com+} = \Delta x_{co}$, $\beta_m = 0$, so σ_o^2 is not adjusted which recovers the result for uniformly distributed coarse-resolution observations.

The above results suggest that $\gamma_b = \sigma_b^2/(\sigma_b^2 + \sigma_o^2)$ should be modified into

$$\gamma_m = \sigma_b^2/(\sigma_b^2 + \beta_m\sigma_b^2 + \sigma_o^2) \qquad (11b)$$

for the definition of $\Delta\sigma_m^2(x)$ in (3). This modification can improve the similarity of the spatial variation of $\sum_m \Delta\sigma_m^2(x)$ to that of the true error variance reduction, denoted by $\Delta\sigma_{ba}^2(x) \equiv \sigma_b^2 - \sigma_a^2(x)$, but the maximum (or minimum) of $\sum_m \Delta\sigma_m^2(x)$, denoted by $\Delta\sigma_{emx}^2$ (or $\Delta\sigma_{emn}^2$), is usually not very close to that of $\Delta\sigma_{ba}^2(x)$. The maximum (or minimum) of $\Delta\sigma_{ba}^2(x)$ can be closely estimated by $\Delta\sigma_{mx}^2$ (or $\Delta\sigma_{mn}^2$) – the maximum (or minimum) of $\Delta\sigma_M^2(x)$ computed by (6) for uniform coarse-resolution observations but with Δx_{co} decreased to Δx_{omn} (or increased to Δx_{omx}), where Δx_{omn} (or Δx_{omx}) is the minimum (or maximum) spacing between two adjacent observations among all non-uniformly distributed coarse-resolution observations in the one-dimension analysis domain. By adjusting $\Delta\sigma_{emx}^2$ to $\Delta\sigma_{mx}^2$ and $\Delta\sigma_{emn}^2$ to $\Delta\sigma_{mn}^2$, the error variance reduction can be estimated by

$$\Delta\sigma_M^2(x) = F(x) \equiv [\sum_m \Delta\sigma_m^2(x) - \Delta\sigma_{emn}^2]\rho + \Delta\sigma_{mn}^2, \qquad (12a)$$

where $\rho = [\Delta\sigma_{mx}^2 - \Delta\sigma_{mn}^2]/[\Delta\sigma_{emx}^2 - \Delta\sigma_{emn}^2]$.

The analysis error variance is then estimated by $\sigma_a^2(x) \approx \sigma_{a*}^2(x) \equiv \sigma_b^2 - \Delta\sigma_M^2(x)$ as in (7), except that $\Delta\sigma_M^2(x)$ is computed by (12a) instead of (6). As shown by the example in Fig. 5, the estimated $\sigma_{a*}^2(x)$ captures closely not only the maximum and minimum but also the spatial variation of the benchmark $\sigma_a^2(x)$ computed from (1). Using this estimated $\sigma_{a*}(x)$, the previously estimated $\mathbf{A}_e$ from the spectral formulations in X16 can be modified into $\mathbf{A}_a$, $\mathbf{A}_b$ or $\mathbf{A}_c$ with its ij^{th} element given by the same formulation as shown in

5

(8a), (8b) or (8c). For the case in Fig. 5, the benchmark **A** is plotted in Fig. 6, while the deviations of $\mathbf{A}_e$, $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ from the benchmark **A** are shown in Figs. 7a, 7b, 7c and 7d, respectively. As shown, the deviation becomes increasingly small when the estimated analysis error covariance matrix is modified successively to $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$.
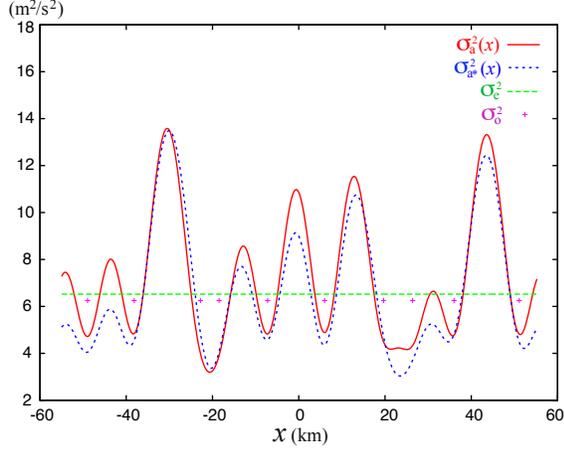
Fig. 5. As in Fig. 1 but for M (= 10) non-uniformly distributed coarse-resolution observations.
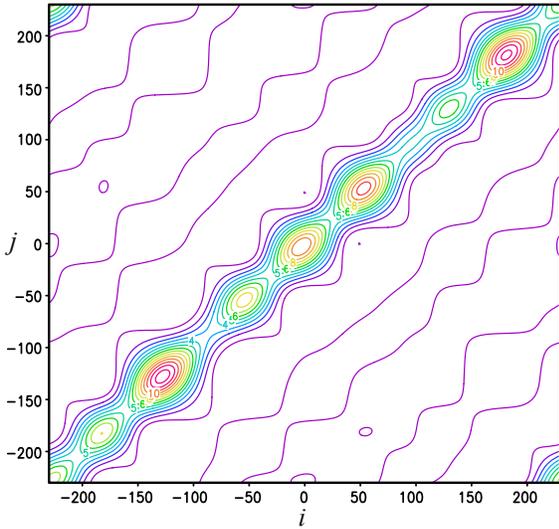
Fig. 6. As in Fig. 3 but for the case in Fig. 5.

As explained in section 2, the accuracy of the second-step analysis depends on the accuracy of the estimated **A** over the extended nested domain (that is, the nested domain plus its extended vicinities within the distance of $2L_a$ on each side outside the nested domain), while the latter can be measured by the smallness of the RE of the estimated **A** with respect to the benchmark **A**, as defined for $\mathbf{A}_e$ in (9). The REs of $\mathbf{A}_e$, $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ computed for the case in Fig. 5 are listed in the first column of Table 2. As listed, the RE becomes increasingly small when $\mathbf{A}_e$ is modified successively to $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$, which quantifies the successively reduced deviation shown in Figs. 7a-7d.
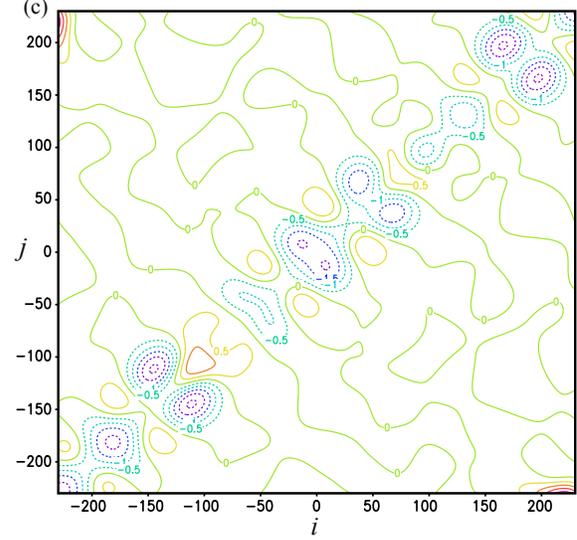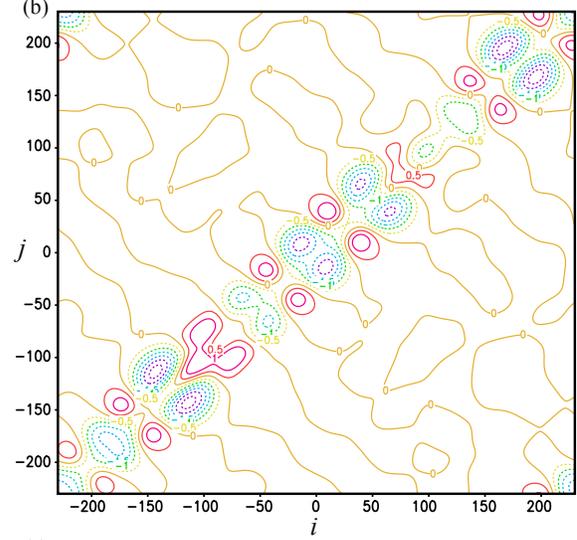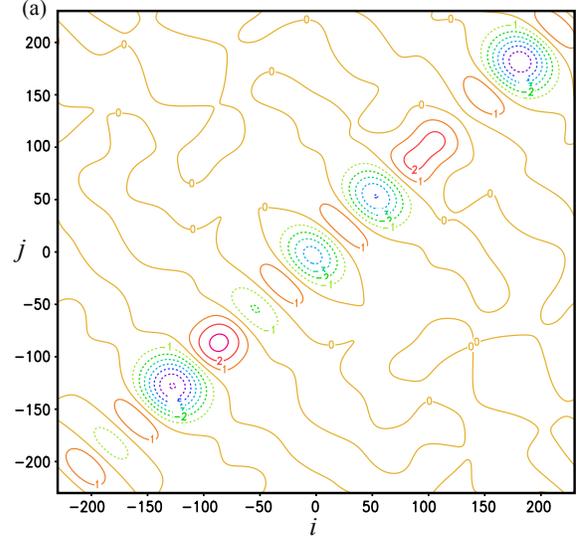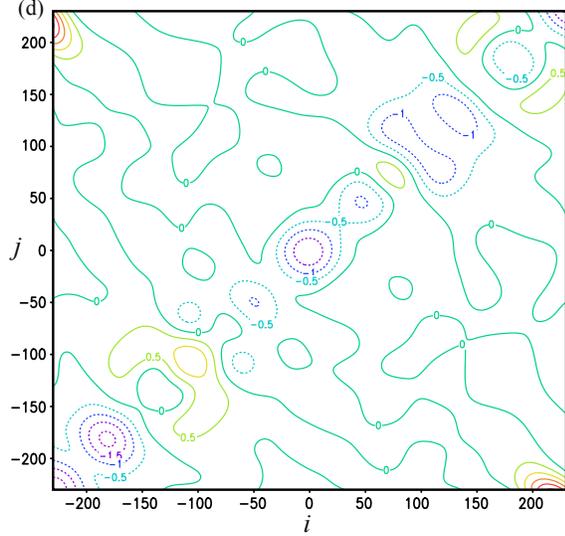
Fig. 7. As in Fig. 4 but for the case in Fig. 5.

Again, idealized experiments are designed and named similarly to those in section 2, but the first set of simulated innovation data is now replaced by the second set that consists of M (= 10) non-uniformly distributed coarse-resolution innovations over the analysis domain of length D (= $N\Delta x$ = 110.4 km) and M' (= 74) high-resolution innovations in the nested domain of length $D/6$ (generated at those grid points that are not covered by the coarse-resolution innovations within the nested domain). All the innovations in this second set are generated similarly to those described in section 2 but in consistency with the case in Figs. 5-7.

Table 2. As in Table 1 but for the second set of innovations with periodic extension.

Experiment	$n = 20$	$n = 50$	$n = 100$	Final
SE	0.711	0.334	0.276	0.018 at $n = 404$
TEe $RE(\mathbf{A}_e) = 0.355$	0.482	0.439		0.442 at $n = 76$
TEa $RE(\mathbf{A}_a) = 0.238$	0.418	0.388	0.348	0.353 at $n = 108$
TEb $RE(\mathbf{A}_b) = 0.197$	0.318	0.288	0.257	0.243 at $n = 179$
TEc $RE(\mathbf{A}_c) = 0.148$	0.213	0.151		0.155 at $n = 52$

The domain-averaged RMS errors of the analysis increments obtained from the four two-step experiments are shown in Table 2 versus those from the SE. As shown, the TEe outperforms SE for $n = 20$ but not so for $n = 50$. The improvement of TEe over SE is similar to but much less significant than that of TEA over SE in Table 2 of X16. This reduced improvement can be largely explained by the fact that the coarse-resolution innovations are generated here not only more sparsely but also more non-uniformly than those in section 3.3 of X16 and the deviation of $\mathbf{A}_e$ from the benchmark $\mathbf{A}$ becomes much larger in Fig. 7a here than that in Fig. 7b of X16. The TEa outperforms TEe for $n = 20$ and 50 but still underperforms SE for n increased to 50 and beyond. The improvement of TEA over TEe is consistent with the

improved accuracy of $\mathbf{A}_a$ [$RE(\mathbf{A}_a) = 0.238$] over $\mathbf{A}_e$ [$RE(\mathbf{A}_e) = 0.355$]. The TEb outperforms TEa for each listed value of n and also outperforms SE for n up to 100. The improvement of TEb over TEa is consistent with the improved accuracy of $\mathbf{A}_b$ [$RE(\mathbf{A}_b) = 0.197$] over $\mathbf{A}_a$. The TEc outperforms TEb for each listed value of n, and the improvement is consistent with the improved accuracy of $\mathbf{A}_c$ [$RE(\mathbf{A}_c) = 0.148$] over $\mathbf{A}_b$.

4. Error variance reduction by non-uniform coarse-resolution observations without periodic extension

Consider that the M coarse-resolution observations are still non-uniformly distributed in the one-dimensional analysis domain of length D but without periodic extension. In this case, their produced error variance reduction $\Delta\sigma_M^2(x)$ still can be estimated by (12a) except for the following three modifications.

(i) The maximum (or minimum) of $\sum_m\Delta\sigma_m^2(x)$, that is, $\Delta\sigma_{emx}^2$ (or $\Delta\sigma_{emn}^2$) should be found in the interior domain between the leftist and rightist observation points.

(ii) For the leftist (or rightist) observation that has only one adjacent observation to its right (or left) in the one-dimensional analysis domain, its error variance is still adjusted from σ_o^2 to $\sigma_{om}^2 = \sigma_o^2 + \beta_m\sigma_b^2$ but β_m is calculated by setting $C_b^2(\Delta x_{com-}) = 0$ [or $C_b^2(\Delta x_{com+}) = 0$] in (11a) for calculating γ_m in (11b).

(iii) Note from (12a) that $\sum_m\Delta\sigma_m^2(x) \to 0$ and thus $F(x) \to \Delta\sigma_{mn}^2 - \rho\Delta\sigma_{emn}^2$ as x moves outward far away from the leftist (or rightist) measurement point and thus far away from all the observations points. In this case, if $\Delta\sigma_{mn}^2 - \rho\Delta\sigma_{emn}^2 < 0$ (as for the case in this section), then $\Delta\sigma_M^2(x)$ estimated by $F(x)$ in (12a) may become unrealistically negative as x moves outward beyond the leftist (or rightist) measurement point, denoted by x_{mb}. To avoid this problem, (12a) is modified into

$$\Delta\sigma_M^2(x) = F(x_{mb}) - [F(x_{mb}) - F(x)]R_1 \text{ for } x \text{ beyond } x_{mb},$$
$$(12b)$$

where R_1 is a factor defined by

$$R_1 = \min\{1, [F(x_{mb})]/[F(x_{mb}) + \rho\Delta\sigma_{emn}^2 - \Delta\sigma_{mn}^2]\}.$$

It is easy to see from (12b) that for $\Delta\sigma_{mn}^2 - \rho\Delta\sigma_{emn}^2 < 0$ and thus $R_1 < 1$, $\Delta\sigma_M^2(x) = F(x_{mb}) - [F(x_{mb}) - F(x)]R_1 \to 0$ as $|x| \to \infty$, so the estimated $\Delta\sigma_M^2(x)$ in (12b) can never become unrealistically negative.

The analysis error variance is estimated by $\sigma_a^2(x) \approx \sigma_{a*}^2(x) \equiv \sigma_b^2 - \Delta\sigma_M^2(x)$ as in (7), except that $\Delta\sigma_M^2(x)$ is computed by (12a) [or (12b)] for x within (or beyond) the interior domain. As shown by the example in Fig. 8, the estimated $\sigma_{a*}^2(x)$ captures closely the spatial variation of the benchmark $\sigma_a^2(x)$ not only within but also beyond the interior domain. Using this estimated $\sigma_{a*}(x)$, $\mathbf{A}_e$ can be modified into $\mathbf{A}_a$, $\mathbf{A}_b$ or $\mathbf{A}_c$ with its ij^{th} element given by the same formulation as shown in (8a), (8b) or (8c). For the case in Fig. 8, the benchmark $\mathbf{A}$ (not shown) has the same interior structure (for interior grid points i and j) as that for the case with periodic extension in Fig. 6, but significant differences are seen in the following two aspects around the four corners (similar to those seen from Figs. 7a and 11a of X16): (i) The element value becomes large toward the two corners along the diagonal line (which is consistent with the increased analysis error variance toward

the two ends of the analysis domain as shown in Fig. 8 in comparison with that in Fig. 5). (ii) The element value becomes virtually zero toward the two off-diagonal corners (because there is no periodic extension). The deviations of $\mathbf{A}_e$, $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ from the benchmark $\mathbf{A}$ (not shown) are mostly similar to those in Figs. 8a, 8b, 8c, and 8d, respectively. Again, the deviation becomes increasingly small when the estimated analysis error covariance matrix is modified successively to $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$. The REs of $\mathbf{A}_e$, $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$ are listed in the first column of Table 3. As listed, the RE becomes increasingly small when $\mathbf{A}_e$ is modified successively to $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_c$, which quantifies the successively reduced deviation.
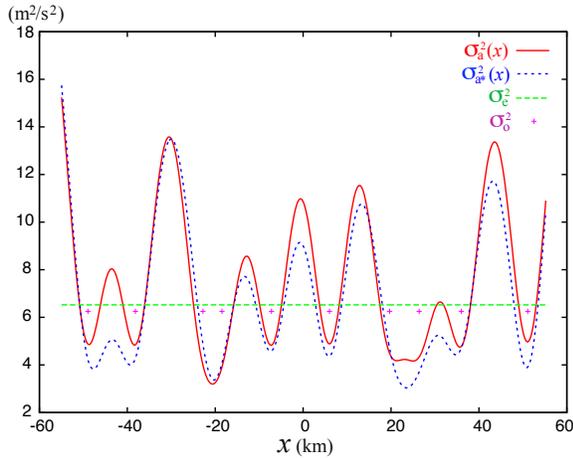
Fig. 8. As in Fig. 5 but without periodic extension.

Table 3. As in Table 2 but without periodic extension.

Experiment	$n = 20$	$n = 50$	$n = 100$	Final
SE	0.499	0.328	0.194	0.012 at $n = 451$
TEe RE($\mathbf{A}_e$) = 0.355	0.463	0.424		0.399 at $n = 73$
TEa RE($\mathbf{A}_a$) = 0.238	0.394	0.358		0.385 at $n = 54$
TEb RE($\mathbf{A}_b$) = 0.196	0.281	0.273		0.248 at $n = 77$
TEc RE($\mathbf{A}_c$) = 0.147	0.215	0.149		0.123 at $n = 77$

Idealized experiments are designed and named as those in section 3 except that there is no periodical extension. The domain-averaged RMS errors of the analysis increments obtained from the four two-step experiments are shown in Table 3 versus those from the SE. As shown, the TEe outperforms SE for $n = 20$ but not so for $n = 50$. The improvement of TEe over SE is much less significant than that of TEA over SE in Table 3 of X16, and this reduced improvement can be explained by the same fact as stated for the previous case in section 3. The TEa outperforms TEe for $n = 20$ and 50, and the improvement is consistent with the improved accuracy of $\mathbf{A}_a$ [RE($\mathbf{A}_a$) = 0.238] over $\mathbf{A}_e$ [RE($\mathbf{A}_e$) = 0.355]. The TEb outperforms TEa for each listed value of n, which is consistent with the improved accuracy of $\mathbf{A}_b$ [RE($\mathbf{A}_b$) = 0.196] over $\mathbf{A}_a$. The TEc outperforms TEb for each listed

value of n, which is consistent with the improved accuracy of $\mathbf{A}_c$ [RE($\mathbf{A}_c$) = 0.147] over $\mathbf{A}_b$.

5. Conclusions

The two-step variational method developed in X16 for analyzing observations of different spatial resolutions is improved by considering and efficiently estimating the spatial variation of analysis error variance produced by analyzing coarse-resolution observations in the first step. The constant analysis error variance computed from the spectral formulations in X16 can represent the spatial averaged value of the true analysis error variance but it cannot capture the spatial variation in the true analysis error variance. As revealed by the examples presented in this paper (see Figs. 1-2, 5 and 8 for one-dimensional cases), the true analysis error variance tends to have increasingly large spatial variations when the coarse-resolution observations become increasingly non-uniform and/or sparse, and this is especially true and serious when the separation distances between neighboring coarse-resolution observations become close to or even locally larger than the background error de-correlation length scale. In this case, the spatial variation of analysis error variance and associated spatial variation in analysis error covariance need to be considered and estimated efficiently in order to further improve the two-step analysis.

The analysis error variance can be viewed conveniently as the background error variance minus the total error variance reduction produced by analyzing all the coarse-resolution observations. To efficiently estimate the latter, semi-empirical formulations are constructed for three types of coarse-resolution observations in one-dimensional space with successively increased complexity and generality (from uniformly distributed observations with periodic extension to non-uniformly distributed observations without periodic extension).

The estimated spatially varying analysis error variance is used to modify the analysis error covariance computed from the spectral formulations of X16 in three different forms [see (8)]. The first is a conventional formulation in which the covariance is modulated by the spatially varying standard deviation separately via each entry of the covariance to retain the self-adjointness. This modulation has a chessboard structure but the desired modulation has a banded structure (along the direction perpendicular to the diagonal line) as revealed by the to-be-corrected deviation from the benchmark truth (see Fig. 4a), so the deviation is only partially reduced (see Fig. 4b). The second formulation is new, in which the modulation is realigned to capture the desired banded structure and yet still retain the self-adjointness. The deviation from the benchmark truth is thus further reduced (see Fig. 4c), but the deviation is reduced not broadly enough along each band. By properly broadening the reduction distribution in the third formulation, the deviation is much further reduced (see Fig. 4d).

The successive improvements made by the above three formulations are demonstrated for all the three types of coarse-resolution observations. The improvements are quantified by the successively reduced relative errors [REs, measured by the Frobenius norm defined in (9)] of their modified analysis error covariance matrices with respect to the benchmark truths (see REs listed in the first columns of Tables 1-3). The impacts of the improved accuracies of the

modified analysis error covariance matrices on the two-step analyses are examined with idealized experiments that are similar to but extend those in X16. As expected, the impacts are found to be mostly positive (especially for the third formulation) and largely in consistency with the improved accuracies of the modified analysis error covariance matrices (see Tables 1-3). As new improvements to the conventional formulation, the second and third formulations may also be useful in constructing covariance matrices with non-constant variances for general applications beyond this paper.

The formulations constructed in this paper for estimating the spatial variation of analysis error variance and associated spatial variation in analysis error covariance are effective for further improving the two-step variational method developed in X16, especially when the coarse-resolution observations become increasingly non-uniform and/or sparse. These formulations can be extended to two- and three-dimensional spaces together with the spectral formulations of X16 for real-data applications with the variational data assimilation system of Gao et al. (2013), in which the analyses are univariate and performed in two steps. These extensions are currently being developed.

REFERENCES

Daley, R. 1991: *Atmospheric data analysis.* Cambridge University Press, New York, 457 pp.

Gao, J., T. M. Smith, D. J. Stensrud, C. Fu, K. Calhoun, K. L. Mnaross, J. Brogden, V. Lakshmanan, Y. Wang, K. W. Thomas, K. Brewster, and M. Xue, 2013: A realtime weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Wea. Forecasting*, **28**, 727–745.

Golub, G. H., and C. F. Van Loan, 1983: *Matrix Computations.* Johns Hopkins University Press, Baltimore, 476 pp.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, San Diego, 376 pp.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. part II: spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131**, 1536–1548.

Rutherford, I., 1972: Data assimilation by statistical interpolation of forecast error fields, *J. Atmos. Sci.*, **29**, 809–815.

Xu, Q., L. Wei, J. Gao, Q. Zhao, K. Nai, and S. Liu, 2016: Multistep variational data assimilation: important issues and a spectral approach. *Tellus A*, **68**, 26 pages.