On the Choice of Momentum Control Variables and Covariance Modeling for Mesoscale Data Assimilation

QIN XU

NOAA/National Severe Storms Laboratory, Norman, Oklahoma

ABSTRACT

For mesoscale variational data assimilation with high-resolution observations, there has been an issue concerning the choice of momentum control variables and related covariance modeling. This paper addresses the theoretical aspect of this issue. First, relationships between background error covariance functions for differently chosen momentum control variables are derived, and different choices of momentum control variables are proven to be theoretically equivalent in the sense that they lead to the same optimally analyzed incremental wind field in the limit of infinitely high spatial resolution provided their error covariance functions satisfy the derived relationships. It is then shown that when the velocity potential χ and streamfunction ψ are used as momentum control variables with their background error autocovariance functions modeled by single-Gaussian functions, the derived velocity autocovariance functions contain significant negative sidelobes. These negative sidelobes can represent background wind error structures associated with baroclinic waves on the synoptic scale but become unrepresentative on the mesoscale. To reduce or remove these negative sidelobes for mesoscale variational data assimilation, Gaussian functions are used with two types of modifications to model the velocity covariance functions in consistency with the assumed homogeneity and isotropy in variational data assimilation. In this case, the random (χ, ψ) background error fields have no classically valid homogeneous and isotropic covariance functions, but generalized (χ, ψ) covariance functions can be derived from the modified velocity covariance functions for choosing (χ, ψ) as momentum control variables. Mathematical properties of generalized covariance functions are explored with physical interpretations. Their important implications are discussed for mesoscale data assimilation.

1. Introduction

The streamfunction ψ and velocity potential χ have been widely used as momentum control variables in operational three-dimensional variational data assimilation (3DVar) systems (Lorenc et al. 2000; Wu et al. 2002). A well-recognized advantage for using (χ, ψ) as momentum control variables in 3DVar is that ψ can be linearly related to the mass field via the geostrophy on the synoptic scale, and this linear relationship can be built into the background error covariance. In addition, the χ field can be prescribed with a relatively small error variance to suppress the associated divergent increments, while the ψ field (or mass field) can be split into geostrophically balanced and unbalance parts with a relatively small error variance prescribed to the unbalance part to suppress the unbalanced increments. The aforementioned advantages, however, diminish as the horizontal scale of the analyzed features reduces toward

the mesoscale. Moreover, using (χ, ψ) as momentum control variables with their background error autocovariance modeled by single-Gaussian functions in the mesoscale range can generate spurious structures in the analyzed incremental wind field outside and around areas covered by patched high-resolution observations (such as those from Doppler radars) as a result of the presence of significant negative sidelobes in the velocity autocovariance functions [derived from the (χ, ψ) Gaussian autocovariance functions]. To avoid this problem, Gaussian functions were used with simplifications to model the velocity covariance functions in the recently developed radar wind analysis system (Xu et al. 2015).

Xie and MacDonald (2012) performed spectral and numerical analyses to examine various possible analysis errors that could be potentially introduced by the use of (χ, ψ), versus the use of the horizontal velocity components (u, v) and the use of the vorticity and divergence, as momentum control variables in 3DVar. In particular, their analysis of a single velocity observation indicated that using (χ, ψ) as momentum control variables in

Corresponding author: Qin Xu, qin.xu@noaa.gov

3DVar can generate erroneous negative sidelobes around the observation point because (u, v) are derivatives of (χ, ψ). However, since the background error covariance was modeled (implicitly by a single-Gaussian function or similar type) through applications of a simple recursive filter in their numerical experiments, the presence and absence of negative sidelobes were examined only in terms of choices of momentum control variables independent of covariance modeling. Thus, whether and how the concerned negative sidelobes can be reduced or eliminated by modifying or generalizing the conventional covariance modeling were not studied but need to be explored (as shown later in this paper).

Sun et al. (2016) performed real-data experiments to examine the impacts of different choices of momentum control variables on limited-area high-resolution data assimilation and subsequent convective precipitation forecasting. Their experiments showed that using (χ, ψ), in comparison with using (u, v), as momentum control variables in high-resolution radar data assimilation can result in degraded analysis and forecast. Velocity components were also used favorably as momentum control variables with their background error covariance modeled by single Gaussians and mimicked by recursive filters in the 3DVar system developed by Gao et al. (2004, 2013) for real-time radar data assimilation. Since the aforementioned studies advocated the use of (u, v) in place of (χ, ψ) as momentum control variables for high-resolution variational data assimilation, there has been an issue on the choice of momentum control variables and related covariance modeling. This issue can be important for mesoscale variational data assimilation as well as multiscale variational data assimilation (Xie et al. 2011; Li et al. 2015; Xu et al. 2016). This paper aims to address the theoretical aspect of this issue.

The next section derives the relationships between background error covariance functions for differently chosen momentum control variables and proves that different choices of momentum control variables are theoretically equivalent in the sense that they lead to the same optimally analyzed incremental wind field in the limit of infinitely high spatial resolution provided their error covariance functions satisfy the derived relationships. Section 3 presents typical examples of negative sidelobes in the velocity autocovariance functions derived from single-Gaussian (χ, ψ) autocovariance functions and explains physically why negative sidelobes can represent background wind error structures on the synoptic scale but not on the mesoscale. Section 4 shows how Gaussian functions can be modified to model the isotropic canonical-form velocity autocovariance functions with negative sidelobes reduced or removed for mesoscale variational data assimilation, and explains

how and why (χ, ψ) covariance functions derived by integrating the modified velocity covariance become classically invalid but can still be used as generalized covariance functions for (χ, ψ) chosen as momentum control variables. Mathematical properties of generalized covariance functions are further explored with physical interpretations in appendixes D–F. Conclusions follow in section 5, where the theoretical findings are summarized with their implications remarked for mesoscale data assimilation.

2. Equivalence between different choices of momentum control variables

a. Relationships between covariance functions for different momentum control variables

The covariance of background wind errors at two points, say, $\mathbf{z}_i \equiv (x_i, y_i, \eta_i)^{\mathrm{T}}$ and $\mathbf{z}_j \equiv (x_j, y_j, \eta_j)^{\mathrm{T}}$, in the three-dimensional space (where the vertical coordinate η is as in the concerned data assimilation system) is a 2×2 tensor defined by

$$\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \mathbf{v}_i \mathbf{v}_j^{\mathrm{T}} \rangle, \qquad (1)$$

where $\langle (\cdot) \rangle$ denotes the statistical mean of $(\cdot)$, $\mathbf{v}_i \equiv \mathbf{v}(\mathbf{z}_i)$ [or $\mathbf{v}_j \equiv \mathbf{v}(\mathbf{z}_j)$] is the value of $\mathbf{v}$ at $\mathbf{z} = \mathbf{z}_i$ (or $\mathbf{z}_j$), $\mathbf{v} \equiv (u, v)^{\mathrm{T}}$ denotes the background wind error as a random vector field in $\mathbf{z}$ with $\langle \mathbf{v} \rangle = 0$ (or with $\langle \mathbf{v} \rangle$ subtracted if $\langle \mathbf{v} \rangle \neq 0$), and $(\cdot)^{\mathrm{T}}$ denotes the transpose of $(\cdot)$. The velocity potential χ and streamfunction ψ of $\mathbf{v}$ are two random scalar fields defined by $\mathbf{v} = \nabla \chi + \mathbf{k} \times \nabla \psi$, where $\nabla \equiv (\partial_x, \partial_y)^{\mathrm{T}}$ is the horizontal gradient operator, and $\mathbf{k}$ is the unit vector along the vertical coordinate so $\mathbf{k} \times \nabla = (-\partial_y, \partial_x)^{\mathrm{T}}$ is the horizontal gradient operator rotated counterclockwise by 90°.

The χ and ψ composed vector field, $\mathbf{c} \equiv (\chi, \psi)^{\mathrm{T}}$, is related to $\mathbf{v}$ by $\mathbf{v} = (\nabla, \mathbf{k} \times \nabla)\mathbf{c}$. Substituting this into (1) gives

$$\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_i, \mathbf{z}_j) = \langle [(\nabla_i, \mathbf{k} \times \nabla_i)\mathbf{c}_i][(\nabla_j, \mathbf{k} \times \nabla_j)\mathbf{c}_j]^{\mathrm{T}} \rangle$$
$$= (\nabla_i, \mathbf{k} \times \nabla_i)[(\nabla_j, \mathbf{k} \times \nabla_j)\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_j, \mathbf{z}_i)]^{\mathrm{T}}, \qquad (2)$$

where ∇_i (or ∇_j) denotes ∇ applied to $\mathbf{z}_i$ (or $\mathbf{z}_j$), $\mathbf{c}_i \equiv \mathbf{c}(\mathbf{z}_i)$, $\mathbf{c}_j \equiv \mathbf{c}(\mathbf{z}_j)$, and $\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \mathbf{c}_i \mathbf{c}_j^{\mathrm{T}} \rangle$ $[= \langle \mathbf{c}_j \mathbf{c}_i^{\mathrm{T}} \rangle^{\mathrm{T}} \equiv \mathbf{C}_{\mathbf{cc}}(\mathbf{z}_j, \mathbf{z}_i)^{\mathrm{T}}]$ is the 2×2 covariance tensor of $\mathbf{c} \equiv (\chi, \psi)^{\mathrm{T}}$ with its two diagonal (or off diagonal) elements defined by $C_{\chi\chi}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \chi_i \chi_j \rangle$ and $C_{\psi\psi}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \psi_i \psi_j \rangle$ [or $C_{\chi\psi}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \chi_i \psi_j \rangle$ and $C_{\psi\chi}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \psi_i \chi_j \rangle$].

The divergence and vorticity of $\mathbf{v}$ are random scalar fields that are defined by $\delta \equiv \nabla \cdot \mathbf{v} = \nabla^2 \chi$ and $\zeta \equiv \mathbf{k} \cdot \nabla \times \mathbf{v} = \nabla^2 \psi$, respectively, where $\nabla^2 \equiv \nabla \cdot \nabla = \partial_x^2 + \partial_y^2$ is the horizontal Laplace operator. Their composed vector

field, $\mathbf{q} \equiv (\delta, \zeta)^{\mathrm{T}}$, is related to $\mathbf{c}$ by $\mathbf{q} = \nabla^2 \mathbf{c}$. The covariance tensor of $\mathbf{q}$ is thus related to that of $\mathbf{c}$ by

$$\mathbf{C_{qq}}(\mathbf{z}_i, \mathbf{z}_j) \equiv \langle \mathrm{q}_i \mathrm{q}_j^{\mathrm{T}} \rangle = \nabla_i^2 \nabla_j^2 \mathbf{C_{cc}}(\mathbf{z}_i, \mathbf{z}_j), \qquad (3)$$

where $\mathbf{q}_i \equiv \mathbf{q}(\mathbf{z}_i)$, $\mathbf{q}_j \equiv \mathbf{q}(\mathbf{z}_j)$, and ∇_i^2 (or ∇_j^2) denotes ∇^2 applied to $\mathbf{z}_i$ (or $\mathbf{z}_j$).

b. Equivalence between different choices of momentum control variables

Using the relationships derived in the previous section, we can prove the theoretical equivalence between the following three different choices of momentum control variables: (i) $\mathbf{v} \equiv (u, v)^{\mathrm{T}}$ in choice 1, (ii) $\mathbf{c} \equiv (\chi, \psi)^{\mathrm{T}}$ in choice 2, and (iii) $\mathbf{q} \equiv (\delta, \zeta)^{\mathrm{T}}$ in choice 3. Here, by theoretical equivalence we mean the equivalence in the continuous limit (i.e., the limit of infinitely high spatial resolution) between optimally analyzed incremental wind fields obtained by using different choices of momentum control variables with their background error covariance tensor functions satisfying the relationships derived in section 2a. Also, to simplify the notations, the symbols $\mathbf{v}$, $\mathbf{c}$, and $\mathbf{q}$ used in the previous subsection to denote the random-vector error fields are reused here for the control-variable vector fields.

The cost function for the minimization problem formulated with each choice of control variables in a discretized domain of $\mathbf{z}$ has the following general incremental form:

$$J = \Delta\mathbf{a}^{\mathrm{T}}\mathbf{B}^{-1}\Delta\mathbf{a} + (\mathbf{H}\Delta\mathbf{a} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\Delta\mathbf{a} - \mathbf{d}), \quad (4a)$$

where $\Delta\mathbf{a} = \mathbf{a} - \mathbf{b}$ denotes the state vector of analysis increment expressed by the chosen control variable, $\mathbf{a}$ (or $\mathbf{b}$) denotes the state vector of analysis (or background), $\mathbf{B}$ denotes the background error covariance matrix constructed by the error covariance tensor function of the chosen control variables in the discretized space of $(\mathbf{z}_i, \mathbf{z}_j)$, $\mathbf{R}$ denotes the observation error covariance matrix, $\mathbf{d} = \mathbf{y} - H(\mathbf{b})$ denotes the innovation vector (i.e., observation minus background in the observation space), $\mathbf{y}$ is the observation vector, H denotes the (nonlinear) observation operator that maps the model space to the observation space, and $\mathbf{H}$ is the linearized H around $\mathbf{b}$. Since the wind observations considered in this study are linearly related to the model's velocity state vector (or velocity field in the continuous limit), H reduces to $\mathbf{H}$ in (4a) for 3DVar. [By extending H to include the forward model operator, the cost function in (4a) can be used for perfect-model 4DVar. In this case, H is nonlinear in general and J in (4a) is an approximation of its original nonincremental form, but the approximation can be improved via outer-loop

iterations by updating $\mathbf{b}$ to $\mathbf{b} + \Delta\mathbf{a}$ once $\Delta\mathbf{a}$ is obtained from the previous outer loop (see section 3b of Lorenc 2003). Thus, although H becomes nonlinear, (4a) still can be used with a given $\mathbf{b}$ to prove the theoretical equivalence between the different choices of momentum control variables, because $\mathbf{b}$ will be updated also equivalently in the subsequent outer loop.]

The cost function in (4a) can be transformed into its dual form (Courtier 1997) in the observation space:

$$J = \mathbf{p}^{\mathrm{T}}\mathbf{HBH}^{\mathrm{T}}\mathbf{p} + (\mathbf{HBH}^{\mathrm{T}}\mathbf{p} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HBH}^{\mathrm{T}}\mathbf{p} - \mathbf{d}), \quad (4b)$$

where $\mathbf{p}$ is the transformed control vector related to $\Delta\mathbf{a}$ by $\Delta\mathbf{a} = \mathbf{BH}^{\mathrm{T}}\mathbf{p} = \mathbf{p}^{\mathrm{T}}\mathbf{BH}$. The minimizer of J in the space of $\mathbf{p}$ is given by the solution of $\nabla_{\mathbf{p}}J = 2\mathbf{HBH}^{\mathrm{T}}\mathbf{R}^{-1}[(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})\mathbf{p} - \mathbf{d}] = 0$ or, equivalently, $(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})\mathbf{p} = \mathbf{d}$. Thus, the minimizer of J in the space of $\Delta\mathbf{a}$ is given by

$$\Delta\mathbf{a} = \mathbf{BH}^{\mathrm{T}}\mathbf{p}, \qquad (5a)$$

with $\mathbf{p}$ solved from

$$(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})\mathbf{p} = \mathbf{d}. \qquad (5b)$$

In (5b), $\mathbf{R}$ and $\mathbf{d}$ are independent of the choice of control variables. As shown in appendix A, $\mathbf{HBH}^{\mathrm{T}}$ also becomes independent of the choice of control variables in the continuous limit and so is the solution of $\mathbf{p}$ obtained from (5a) [also see (A10)]. When $\mathbf{p}$ is transformed back to give $\Delta\mathbf{a} = \mathbf{BH}^{\mathrm{T}}\mathbf{p}$ for the solutions in the control variable spaces of $\mathbf{v}$, $\mathbf{c}$ and $\mathbf{q}$, these solutions satisfy the required relationships of $\mathbf{v} = (\nabla, \mathbf{k} \times \nabla)\mathbf{c}$ and $\mathbf{q} = \nabla^2\mathbf{c}$ in the continuous limit [see (A16) and (A22) in appendix A]. This proves the theoretical equivalence between different choices of control variables.

The above proof assumes implicitly that $\mathbf{C_{cc}}(\mathbf{z}, \mathbf{z}')$ is a differentiable and classically valid covariance tensor function (i.e., to not only satisfy the required symmetry and positive semidefiniteness but also approach zero as $|\mathbf{z}' - \mathbf{z}|$ increases unboundedly). This ensures $\mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}')$ and $\mathbf{C_{qq}}(\mathbf{z}, \mathbf{z}')$ are also classically valid covariance tensor functions, as they are derived from $\mathbf{C_{cc}}(\mathbf{z}, \mathbf{z}')$ via differentiations in (2) and (3), respectively. The inverse, however, does not hold true. In particular, the classical validity of $\mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}')$ [or $\mathbf{C_{qq}}(\mathbf{z}, \mathbf{z}')$] cannot ensure the classical validity of $\mathbf{C_{cc}}(\mathbf{z}, \mathbf{z}')$ [or $\mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}')$], as the latter is derived from the former by inverting (2) [or (3)]. This problem will be revisited later in section 4c.

c. Relationships between homogeneous and isotropic covariance functions

In variational data assimilation, the random vector field of background wind error $\mathbf{v}$ (or $\mathbf{v}$ normalized by its standard deviation) is assumed to be homogeneous and

isotropic in $\mathbf{x} \equiv (x, y)^{\mathrm{T}}$. In this case, $\mathbf{C_{vv}}$ defined in (1) can be transformed into the following canonical form:

$$\langle (l_i, t_i)^{\mathrm{T}}(l_j, t_j) \rangle = \mathbf{R}_\alpha \mathbf{C_{vv}} \mathbf{R}_\alpha^{\mathrm{T}}, \qquad (6)$$

where $(l_i, \ t_i)^{\mathrm{T}} \equiv \mathbf{R}_\alpha \mathbf{v}_i$, $(l_j, \ t_j)^{\mathrm{T}} \equiv \mathbf{R}_\alpha \mathbf{v}_j$, $\mathbf{R}_\alpha \equiv [(\cos\alpha, -\sin\alpha)^{\mathrm{T}}, (\sin\alpha, \cos\alpha)^{\mathrm{T}}]$ is the rotational matrix that rotates the x axis to the direction of $\mathbf{r} \equiv \mathbf{x}_j - \mathbf{x}_i$; and $\alpha \equiv \tan^{-1}[(y_j - y_i)/(x_j - x_i)]$ is the angle of the rotation, measured positively counterclockwise. In the above locally rotated coordinate system, $\mathbf{v}_i$ (or $\mathbf{v}_j$) is transformed to $(l_i, t_i)^{\mathrm{T}}$ [or $(l_j, t_j)^{\mathrm{T}}$] with the l component along $\mathbf{r}$ and the t component perpendicular to $\mathbf{r}$ with positive to the left (see Fig. 1). Since $(l_i, t_i)^{\mathrm{T}}$ and $(l_j, t_j)^{\mathrm{T}}$ are invariant in the locally rotated coordinate system, their constructed canonical form of the covariance tensor in (6) is invariant with respect to translations and rotations of the system of $\mathbf{x}_i$ and $\mathbf{x}_j$ under the above assumed homogeneity and isotropy. This implies that the covariance tensor $\langle (l_i, t_i)^{\mathrm{T}}(l_j, t_j) \rangle$ is a function of (r, η_i, η_j) independent of α, where $r \equiv |\mathbf{r}|$. In this case, $\mathbf{C_{cc}}$ and $\mathbf{C_{qq}}$ reduce to their respective homogeneous and isotropic forms that are also functions of (r, η_i, η_j) independent of α, and the reduced homogeneous and isotropic form of (2) recovers (A.1) of Xu and Wei (2001, hereafter XW).

For the canonical form in (6) to be invariant not only to horizontal translations and rotations but also to mirror reflections of the system of horizontal points $\mathbf{x}_i$ and $\mathbf{x}_j$, it is necessary to assume $C_{lt} = C_{tl} = 0$ or, equivalently, $C_{\chi\psi} = C_{\psi\chi} = 0$ [see (2.6) of XW], where $C_{lt} \equiv \langle l_i t_j \rangle$ and $C_{tl} \equiv \langle t_i l_j \rangle$. This assumption is commonly used in variational data assimilation (see section 5.2 of Daley 1991), and the validity of this assumption is supported by the smallness of the estimated $\psi - \chi$ correlation (see Fig. 20 of Hollingsworth and Lonnberg 1986, hereafter HL86) and the smallness of estimated $C_{lt} = C_{tl}$ (see Fig. 3c of XW) from radiosonde innovations (observations minus background values at observation points). With this assumption, $\mathbf{C_{cc}}$ reduces $(C_{\chi\chi}, C_{\psi\psi})^{\mathrm{diag}}$, and (6) reduces to $(C_{ll}, C_{tt})^{\mathrm{diag}} = \mathbf{R}_\alpha \mathbf{C_{vv}} \mathbf{R}_\alpha^{\mathrm{T}}$, or, equivalently,

$$\mathbf{C_{vv}} = \mathbf{R}_\alpha^{\mathrm{T}} (C_{ll}, C_{tt})^{\mathrm{diag}} \mathbf{R}_\alpha, \qquad (7)$$

where $C_{ll} \equiv \langle l_i l_j \rangle$, $C_{tt} \equiv \langle t_i t_j \rangle$, and $(\cdot, \cdot)^{\mathrm{diag}}$ denotes a 2×2 diagonal matrix composed by two diagonal elements inside of $(\cdot, \cdot)$. The two off-diagonal terms of $\mathbf{C_{vv}}$ in (7) become zero only when $C_{ll} = C_{tt}$.

Under the assumed condition of $C_{\chi\psi} = C_{\psi\chi} = 0$, $\mathbf{v}$ can be partitioned into two uncorrelated parts; that is, a rotational part defined by $\mathbf{v}_r \equiv (-\partial_y \psi, \partial_x \psi)^{\mathrm{T}}$ and a divergent part defined by $\mathbf{v}_d \equiv (\partial_x \chi, \partial_y \chi)^{\mathrm{T}}$. Substituting the partitioned $\mathbf{v} = \mathbf{v}_d + \mathbf{v}_r$ into (1) gives $\mathbf{C_{vv}} = \mathbf{C}_d + \mathbf{C}_r$, where $\mathbf{C}_d \equiv \langle \mathbf{v}_{di} \mathbf{v}_{dj}^{\mathrm{T}} \rangle$ and $\mathbf{C}_r \equiv \langle \mathbf{v}_{ri} \mathbf{v}_{rj}^{\mathrm{T}} \rangle$ are the covariance
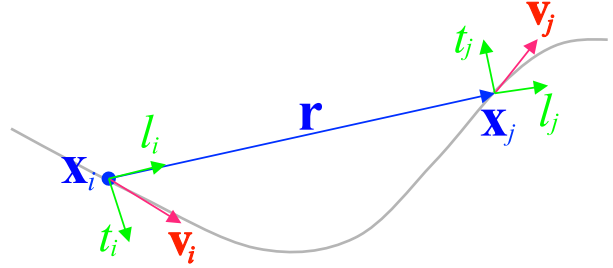

FIG. 1. Velocity errors $\mathbf{v}_i$ and $\mathbf{v}_j$ (red arrows) at points $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively, over the distance of $\mathbf{r} = \mathbf{x}_j - \mathbf{x}_i$ (blue arrow), and their l components (green arrows labeled as l_i and l_j) and t components (green arrows labeled as t_i and t_j). The gray curve plots schematically the background streamline (or geopotential height contour) on a mid-troposphere pressure surface associated with a typical midlatitude baroclinic wave on the synoptic scale. It is assumed that the background winds associated with the baroclinic wave are overpredicted, so the background wind errors largely follow the same directions of their associated true winds. In this case, as shown by the two green arrows (denoted as t_i and t_j), the t components of the background wind errors are negatively correlated when the separation distance $r = |\mathbf{x}_j - \mathbf{x}_i|$ is at and around a half wavelength of the baroclinic wave.

tensors of $\mathbf{v}_d$ and $\mathbf{v}_r$, respectively. The canonical forms of $\mathbf{C}_d$ and $\mathbf{C}_r$ can be derived similarly, also as functions of (r, η_i, η_j) independent of α, by $(C_{ll}^d, C_{tt}^d)^{\mathrm{diag}} = \mathbf{R}_\alpha \mathbf{C}_d \mathbf{R}_\alpha^{\mathrm{T}}$ and $(C_{ll}^r, C_{tt}^r)^{\mathrm{diag}} = \mathbf{R}_\alpha \mathbf{C}_r \mathbf{R}_\alpha^{\mathrm{T}}$, respectively. Their sum gives $(C_{ll}, C_{tt})^{\mathrm{diag}} = (C_{ll}^d, C_{tt}^d)^{\mathrm{diag}} + (C_{ll}^r, C_{tt}^r)^{\mathrm{diag}}$. As shown in (2.14) of XW, these two canonical forms are related to $C_{\chi\chi}$ and $C_{\psi\psi}$, respectively, by

$$(C_{ll}^d, C_{tt}^d)^{\mathrm{diag}} = -(\partial_r^2, r^{-1}\partial_r)^{\mathrm{diag}} C_{\chi\chi}, \qquad (8\mathrm{a})$$

$$(C_{ll}^r, C_{tt}^r)^{\mathrm{diag}} = -(r^{-1}\partial_r, \partial_r^2)^{\mathrm{diag}} C_{\psi\psi}. \qquad (8\mathrm{b})$$

Also, under the condition of $C_{\chi\psi} = C_{\psi\chi} = 0$ with the assumed homogeneity and isotropy, (3) reduces to

$$(C_{\delta\delta}, C_{\zeta\zeta})^{\mathrm{diag}} = (\partial_r^2 + r^{-1}\partial_r)^2 (C_{\chi\chi}, C_{\psi\psi})^{\mathrm{diag}}. \qquad (9)$$

Only homogeneous and isotropic covariance functions will be considered in the remaining sections, where they will be treated as functions of r only with (η_i, η_j) dropped from the independent variables (r, η_i, η_j) to simplify the notations.

3. Negative sidelobes in velocity autocovariance functions

a. Negative sidelobes in derived velocity autocovariance functions

When $C_{\chi\chi}$ and $C_{\psi\psi}$ are modeled by Gaussian functions, C_{ll} and C_{tt} contain negative sidelobes and so do the two diagonal components of $\mathbf{C_{vv}}$. To show this, we set

$$C_{\chi\chi} = \sigma_\chi^2 G(r/L_\chi), \tag{10a}$$

$$C_{\psi\psi} = \sigma_\psi^2 G(r/L_\psi), \tag{10b}$$

where $G(\cdot) \equiv \exp[-(\cdot)^2/2]$, σ_χ^2 (or σ_ψ^2) denotes the error variance of χ (or ψ), and L_χ (or L_ψ) denotes the decorrelation length of $C_{\chi\chi}$ (or $C_{\psi\psi}$). Substituting (10) into the traces of (8a) and (8b) gives

$$C_d \equiv \mathrm{Tr}(\mathbf{C}_d) = C_{ll}^d + C_{tt}^d = -(r^{-1}\partial_r + \partial_r^2)C_{\chi\chi}$$
$$= 2\sigma_d^2[1 - r^2/(4L_d^2)]G(r/\sqrt{2}L_d), \tag{11a}$$

$$C_r \equiv \mathrm{Tr}(\mathbf{C}_r) = C_{ll}^r + C_{tt}^r = -(r^{-1}\partial_r + \partial_r^2)C_{\psi\psi}$$
$$= 2\sigma_r^2[1 - r^2/(4L_r^2)]G(r/\sqrt{2}L_r), \tag{11b}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of $(\cdot)$, $\sigma_d^2 = \sigma_\chi^2/L_\chi^2$ (or $\sigma_r^2 = \sigma_\psi^2/L_\psi^2$) is the error variance for each vector component of $\mathbf{v}_d$ (or $\mathbf{v}_r$), and

$$L_d \equiv [2C_d(r)|_{r=0}]^{1/2}[-\nabla^2 C_d(r)|_{r=0}]^{-1/2} = L_\chi/\sqrt{2}$$

and

$$L_r \equiv [2C_r(r)|_{r=0}]^{1/2}[-\nabla^2 C_r(r)|_{r=0}]^{-1/2} = L_\psi/\sqrt{2}$$

are the decorrelation lengths of C_d and C_r, respectively. Here, the decorrelation length is defined based on the local curvature of the covariance function at $r = 0$ according to (4.3.12) of Daley (1991), and $\nabla^2 = \partial_r^2 + r^{-1}\partial_r$ is used for isotropic covariance functions.

Substituting (10) into the sum of (8a) and (8b) gives

$$C_{ll} = C_{ll}^d + C_{ll}^r = -\partial_r^2 C_{\chi\chi} - r^{-1}\partial_r C_{\psi\psi}$$
$$= \sigma_d^2(1 - r^2/2L_d^2)G(r/\sqrt{2}L_d) + \sigma_r^2 G(r/\sqrt{2}L_r), \tag{12a}$$

$$C_{tt} = C_{tt}^d + C_{tt}^r = -\partial_r^2 C_{\psi\psi} - r^{-1}\partial_r C_{\chi\chi}$$
$$= \sigma_r^2(1 - r^2/2L_r^2)G(r/\sqrt{2}L_r) + \sigma_d^2 G(r/\sqrt{2}L_d). \tag{12b}$$

From (12), it is easy to see that C_{ll} and C_{tt} switch to each other as (σ_d^2, L_d) and (σ_r^2, L_r) switch, so there is a duality between C_{ll} and C_{tt} associated with the switch between (σ_d^2, L_d) and (σ_r^2, L_r).

When $L_\chi = L_\psi = L$ and thus $L_d = L_r = L/\sqrt{2}$ and $\sigma_d^2/\sigma_r^2 = \sigma_\chi^2/\sigma_\psi^2$, (12) reduces to

$$C_{ll} = \sigma^2[1 - (r/L)^2(1 + \beta)^{-1}]G(r/L), \tag{13a}$$

$$C_{tt} = \sigma^2[1 - (r/L)^2\beta(1 + \beta)^{-1}]G(r/L), \tag{13b}$$

where $\beta \equiv \sigma_r^2/\sigma_d^2$ and $\sigma^2 \equiv \sigma_d^2 + \sigma_r^2 = (1 + \beta)\sigma_d^2 = (1 + \beta^{-1})\sigma_r^2$ is the error variance for each vector component of $\mathbf{v}$. The aforementioned duality between C_{ll} and C_{tt} is now associated only with the switch between σ_d^2 and σ_r^2 or, equivalently, the inversion of β. This duality is seen directly from (13). Owing to this duality, C_{ll} computed from (13a) for $\beta = 4, 2, 1, \frac{1}{2},$ and $\frac{1}{4}$ are the same as C_{tt} computed from (13b) for $\beta = \frac{1}{4}, \frac{1}{2}, 1, 2,$ and 4, respectively, so they can be plotted compactly by the same set of five curves, as shown in Fig. 2.

Negative sidelobes contained in C_{ll}, C_{tt}, C_{uu}, and C_{vv} can be illustrated as typical examples in the following three cases.

Case 1. For $L_d = L_r = L/\sqrt{2}$ and $\beta = 1$, we have $C_d = C_r = C_{ll} = C_{tt} = \sigma^2[1 - (r/L)^2/2]G(r/L)$ with $\sigma^2 = 2\sigma_r^2 = 2\sigma_d^2$ according to (11) and (13), so $C_{uu} = C_{vv} = C_{ll} = C_{tt}$ and $C_{uv} = C_{vu} = 0$ according to (7). In this case, C_{uu} $(=C_{vv})$ becomes isotropic and has a distinct negative sidelobe over the range of $r > 1.4L$, as shown by the thick blue curve in Fig. 2.

Case 2. For $L_d = L_r = L/\sqrt{2}$ and $\beta = 2$, we have $C_r = 2C_d$ and $C_{ll} = \sigma^2[1 - (r/L)^2/3]G(r/L) \neq C_{tt} = \sigma^2[1 - 2(r/L)^2/3]G(r/L)$ with $\sigma^2 = 3\sigma_r^2/2 = 3\sigma_d^2$, so $C_{uu} \neq C_{vv}$ and $C_{uv} = C_{vu} \neq 0$. In this case, $C_{ll}(r)$ has a relatively shallow negative sidelobe over the range of $r > 1.75L$ as shown by the thin red curve in Fig. 2, but $C_{tt}(r)$ has a relatively deep negative sidelobe over the range of $r > 1.22L$ as shown by the thick red curve in Fig. 2. Also, the four components of $\mathbf{C}_{vv}$ are functions of $\mathbf{r}$ and thus nonisotropic, and $C_{vv}(\mathbf{r})$ is the same as 90°-rotated $C_{uu}(\mathbf{r})$. As shown by the color contours in Figs. 3a and 3b, $C_{uu}(\mathbf{r})$ contains two negative sidelobes and $C_{uv}(\mathbf{r})$ $[=C_{uv}(-\mathbf{r}) = C_{vu}(\mathbf{r})]$ is nonzero everywhere except along the two coordinate lines (i.e., $y_j = y_i$ and $x_j = x_i$).

Case 3. For $L_d = L_r = L/\sqrt{2}$ and $\beta = 1/2$, C_{ll} (or C_{tt}) is the same as C_{tt} (or C_{ll}) in case 2 but with $\sigma^2 = 3\sigma_r^2 = 3\sigma_d^2/2$ as a result of the aforementioned duality between C_{ll} and C_{tt} associated with the inversion of β, so C_{uu} (or C_{vv}) is the same as C_{vv} (or C_{uu}) in case 2 and C_{uv} $(=C_{vu})$ is the same as $-C_{uv}$ $(=-C_{vu})$ in case 2 according to (7).

b. Representativeness of negative sidelobe on synoptic scale

Negative sidelobes in C_{tt} on the synoptic scale may represent background wind error structures associated with not only the recirculations arising from the near geostrophy and prevailing nondivergence (see Fig. 7a of HL86) but also the dominant baroclinic waves on the synoptic scale. As shown schematically in Fig. 1, when the background winds associated with a baroclinic wave
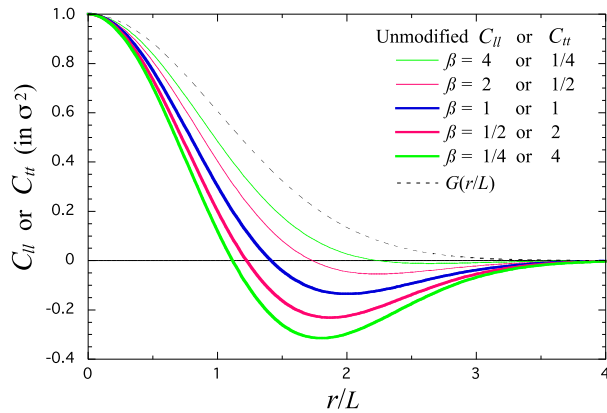
FIG. 2. $C_{ll}(r)$ computed from (13a) for $\beta = 4, 2, 1, \frac{1}{2}$, and $\frac{1}{4}$ [or $C_{tt}(r)$ computed from (13b) for $\beta = \frac{1}{4}, \frac{1}{2}, 1, 2$, and 4] plotted by thin green, thin red, thick blue, thick red, and thick green curves, respectively. The black dashed curve plots the Gaussian function $G(r/L) \equiv \exp[-(r/L)^2/2]$ defined in (10).

are overpredicted, the background wind errors largely follow the same directions of their associated true winds and their t components (shown by the two green arrows denoted by t_i and t_j) are negatively correlated at and around a half wavelength of the baroclinic wave. Conversely, if the background winds are underpredicted, then the background wind errors will be largely opposite of the directions of their associated true winds, and both of the green arrows in Fig. 1 will reverse their directions but remain negatively correlated. Furthermore, if the predicted baroclinic wave has a phase error, then the t component of the background (or true) wind associated with the baroclinic wave can be represented loosely by a sinusoidal function with the same wavelength as the baroclinic wave but with (or without) a phase shift along the direction of the wave propagation. The difference of these two sinusoidal functions is also a sinusoidal function of the same wavelength but represents the t component of the background wind error, so the t components are negatively correlated at and around a half wavelength of the baroclinic wave, regardless of the sign of the phase error.

According to the background wind error variances estimated in the partitioned form of $\sigma^2 = \sigma_r^2 + \sigma_d^2$ from radiosonde innovations in Fig. 10 of HL86 for the ECMWF global data assimilation system and in Fig. 6 of XW for the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS), σ_r is about twice as large as σ_d on the synoptic scale. Also, according to the estimated L_r and L_d in Fig. 12 of XW, L_r is smaller than L_d but larger than $L_d/2$ at every vertical level and becomes very close to L_d at the vertical levels of 900, 500, and 200 hPa. Based on the above estimates, we may assume $\beta = 4$ with $L_\chi = L_\psi = L$ and thus obtain
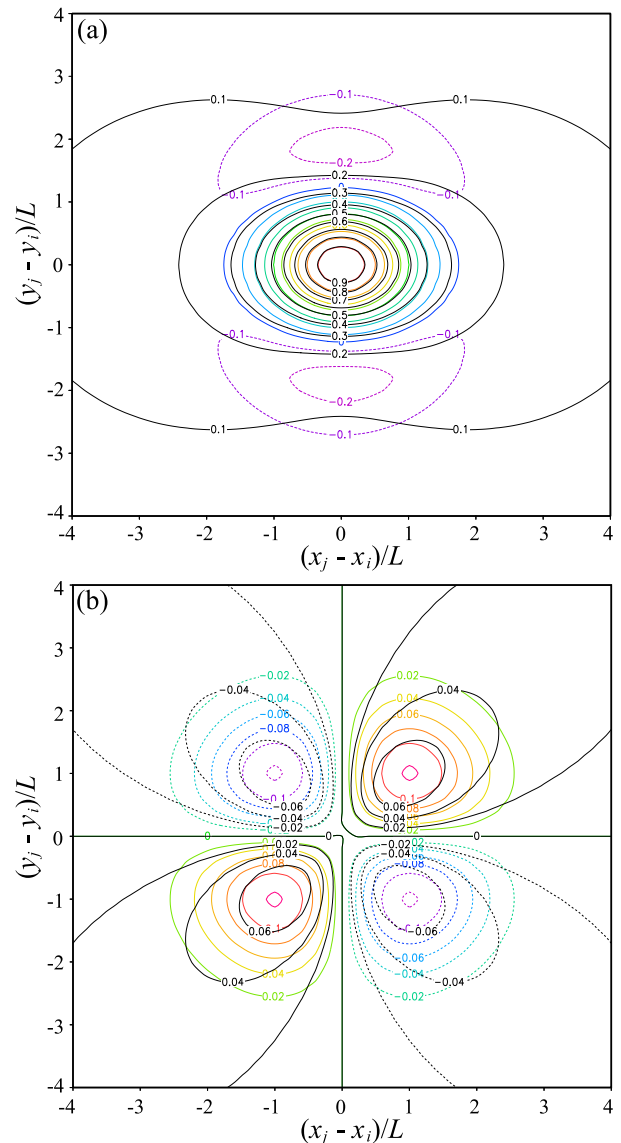

FIG. 3. (a) Unmodified $C_{uu}(\mathbf{r})$ plotted by color contours and type-2 modified $C_{uu}(\mathbf{r})$ plotted by black contours. (b) As in (a), but for $C_{vv}(\mathbf{r})$. The unmodified $C_{uu}(\mathbf{r})$ and $C_{uv}(\mathbf{r})$ are computed from (7) with C_{ll} and C_{tt} given by (13) for $\beta = 2$ and $L_d = L_r = L/\sqrt{2}$. The modified $C_{uu}(\mathbf{r})$ and $C_{uv}(\mathbf{r})$ are also computed from (7) but with C_{ll} and C_{tt} given by (21) for $\beta = 2$ and $L_d = L_r = L/\sqrt{3}$. The contours are plotted every $0.1\sigma^2$ but labeled with σ^2 omitted.

$C_{ll} = \sigma^2[1 - (r/L)^2/5]G(r/L)$ and $C_{tt} = \sigma^2[1 - 4(r/L)^2/5]G(r/L)$ with $\sigma^2 = 5\sigma_r^2/4 = 5\sigma_d^2$ from (13). In this case, $C_{tt}(r)/\sigma^2$ has a deep negative sidelobe over the range of $r > 1.12L$ as shown by the thick green curve in Fig. 2, but $C_{ll}(r)/\sigma^2$ has only a very shallow negative sidelobe over the range of $r > 2.2L$ as shown by the thin green curve in Fig. 2. We can see similar features (but with virtually no negative sidelobe in C_{ll} and a relative shallow negative sidelobe in C_{tt}) from the estimated C_{ll} and C_{tt} in Figs. 5 and 6 of

HL86 and Figs. 3a and 3b of XW. However, since $C_{\chi\chi}$ and $C_{\psi\psi}$ are modeled by single-Gaussian functions in (10), the negative sidelobe of the above derived C_{tt}/σ^2 (shown by the thick green curve in Fig. 2) is about twice as deep as that of the estimated $C_{tt}(r)/\sigma^2$ [with $C_{tt}(r)|_{r=0} = \sigma^2$] in Fig. 6 of HL86 or Fig. 3b of XW. Excessively negative sidelobes of this type (caused in C_{tt} by using single-Gaussian functions to model $C_{\psi\psi}$ and $C_{\chi\chi}$) can be reduced if $C_{\chi\chi}$ and $C_{\psi\psi}$ are modeled by fat-tailed Gaussian-weighted mixtures of Gaussians or multiple Gaussians (see Fig. 1 of Purser et al. 2003b).

c. Unrepresentativeness of negative sidelobe on mesoscale

For mesoscale data assimilation, the background wind error covariance functions can be estimated from radar radial-velocity innovations (Xu et al. 2007a, hereafter X07a; Xu et al. 2007b, hereafter X07b). As an example, Fig. 4 shows the structures of $C_+(r)/C_+(0)$, $C_r(r)/C_+(0)$, and $C_d(r)/C_+(0)$ estimated from radar radial velocity innovations, where $C_+(r) \equiv \mathrm{Tr}(\mathbf{C_{vv}}) = C_{ll}(r) + C_{tt}(r) = C_r(r) + C_d(r)$ and $C_+(0) \equiv C_+(r)|_{r=0} = 2\sigma^2$. [Here, Fig. 4 is produced by using the same method as described in X07b, but the innovation data are generated from radial wind observations collected for a rainy-weather case on 23 May 2005. In this case, radial velocity observations cover more areas between radars than in the clear-weather case considered in X07b, so $C_+(r)/C_+(0)$ is more reliably estimated without the unrepresentative depression (around $r = 100$ km as shown in Fig. 5 of X07b) caused by the data-void gap (from $r = 40$ to 100 km as shown in Fig. 1 of X07b) in the clear-weather case.] As shown in Fig. 4, the estimated $C_+(r)$ and its partitioned $C_d(r)$ and $C_r(r)$ contain no negative sidelobe, and $C_r(r) \approx C_d(r)$. Thus, although the condition of $\beta = 1$ assumed in case 1 of section 3a can be roughly supported by the result of $C_r(r) \approx C_d(r)$ in Fig. 4, the significantly negative sidelobe in C_{ll} ($= C_{tt} = C_{uu} = C_{vv} = C_+/2 = C_d = C_r$) for $\beta = 1$ in case 1 is not supported by the results in Fig. 4 and thus is unrepresentative on the mesoscale. Moreover, there is no physical reason in general to expect such a significantly negative correlation on the mesoscale.

As mentioned in the introduction, negative sidelobes on the mesoscale can generate spurious structures in the analyzed incremental wind field outside and around areas covered by patched high-resolution radar radial velocity observations. As exemplified by case 1 in section 3a, when (χ, ψ) are used as momentum control variables with $C_{\chi\chi}$ and $C_{\psi\psi}$ modeled by Gaussian functions, especially single Gaussians, negative sidelobes are generated inevitably and implicitly in C_{uu} and C_{vv}. These negative sidelobes may provide a theoretical explanation for the degraded analysis produced by the use of (ψ, χ), in
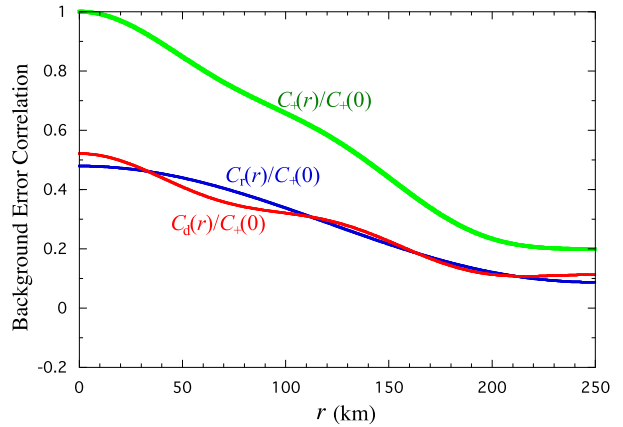

FIG. 4. $C_+(r)/C_+(0)$, $C_r(r)/C_+(0)$, and $C_d(r)/C_+(0)$ estimated from radar radial wind innovations. The radial wind innovations are produced in the same way and from the same data sources as those described in X07b but sampled in the vertical layer $z = 0.8 \pm 0.2$ km under a rainy-weather condition on 23 May 2005.

comparison with the use of (u, v), as momentum control variables in the real-data experiments of Sun et al. (2016) in which $(C_{\chi\chi}, C_{\psi\psi})$ and (C_{uu}, C_{vv}) were modeled by single Gaussians through applications of recursive filter (Purser et al. 2003a) with reduced horizontal decorrelation lengths for (C_{uu}, C_{vv}).

The result of $C_r(r) \approx C_d(r)$ in Fig. 4 was obtained for a particular case. Based on this result alone, we may assume $C_r(r) = C_d(r)$ and thus model $C_{uu} = C_{vv}$ by $\sigma^2 G(r/L)$ with $C_{uv} = C_{vu} = 0$ to eliminate the negative sidelobes illustrated in case 1 of section 3a. In general, however, we need to consider $C_r(r) \neq C_d(r)$ for mesoscale variational data assimilation, especially if the assimilation is performed with high-resolution observations in the second step (in which the background error decorrelation length is reduced to the mesoscale range) after coarse-resolution observations are analyzed on the synoptic scale in the first step (Xie et al. 2011; Li et al. 2015; Xu et al. 2016). In this case, as shown in Fig. 3, C_{uu} and C_{vv} are no longer isotropic, so their contained negative sidelobes can no longer be removed, as in case 1, in consistence with the assumed isotropy in 3DVar, by using isotropic Gaussian functions directly to model them with $C_{uv} = C_{vu} = 0$. However, C_{ll} and C_{tt} remain to be isotropic, so Gaussian functions can be used with some modifications to model them or their partitioned recombinations to reduce or remove negative sidelobes not only in C_{ll} and C_{tt} but also in C_{uu} and C_{vv}, as shown in the next section.

4. Modified velocity covariance functions

In this section two types of modified velocity covariance functions are formulated to reduce or remove

negative sidelobes concerned in the previous section. The detailed formulations are presented in the following subsections.

a. Type-1 modified covariance functions with negative sidelobes reduced

The results in (13) can be recombined into

$$C_+ \equiv C_{ll} + C_{tt} = 2\sigma^2[1 - (r/L)^2/2]G(r/L), \quad (14a)$$

$$C_- \equiv C_{ll} - C_{tt} = 2a\sigma^2(r/L)^2 G(r/L), \quad (14b)$$

where $2a \equiv (\beta - 1)/(\beta + 1)$ and a is confined within $(-1/2, 1/2)$ because $\beta \equiv \sigma_r^2/\sigma_d^2$ is confined within $(0, \infty)$. When the value of β is inverted, a changes sign and so does C_- in (14b). This property is tied up with the duality between C_{ll} and C_{tt} associated with the inversion of β, as noted earlier in (13). The normalized function form of C_+ is the same as that for C_{ll} ($= C_{tt}$) in case 1 and thus also has a negative sidelobe as shown by the solid blue curve in Fig. 2. This negative sidelobe can be removed by using a Gaussian function to model C_+ with its decorrelation length denoted by L_+, while C_- is still modeled by (14b) but with L redenoted by L_-, so (14) is modified into

$$C_+ \equiv C_{ll} + C_{tt} = 2\sigma^2 G(r/L_+), \quad (15a)$$

$$C_- \equiv C_{ll} - C_{tt} = 2a\sigma^2(r/L_-)^2 G(r/L_-). \quad (15b)$$

By changing the notation L_+ to L and L_- to $\sqrt{b}L$, (15) recovers (3.2) of Xu and Gong (2003, hereafter XG). According to (3.7) of XG, L_r and L_d are related to L_+ and L_- by

$$L_r^2 = (1 + 2a)/(L_+^{-2} + 4aL_-^{-2}), \quad (16a)$$

$$L_d^2 = (1 - 2a)/(L_+^{-2} - 4aL_-^{-2}). \quad (16b)$$

Note that the left-hand side of (16a) [or (16b)] is nonnegative and so is the right-hand side, and $1 \pm 2a$ is also nonnegative because $2a$ is confined within $(-1, 1)$ as explained in (14). Thus, $L_+^{-2} \pm 4aL_-^{-2}$ must be nonnegative or, equivalently, L_-^2/L_+^2 cannot be smaller than $\min(0, 4a)$. Furthermore, the autocorrelation functions $C_{ll} = (C_+ + C_-)/2$ and $C_{tt} = (C_+ - C_-)/2$ derived from (15) must be positive semidefinite or, equivalently, their power spectra must be nonnegative. As shown in appendix B, this leads to the following additional constraints on $|a|$ and L_-^2/L_+^2:

$$|a| \leq a_m = e/8 \, (\approx 0.33398), \quad (17a)$$

$$2|a_m/a|[1 - (1 - |a/a_m|)^{1/2}] \leq L_-^2/L_+^2$$
$$\leq 2|a_m/a|[1 + (1 - |a/a_m|)^{1/2}], \quad (17b)$$

where e (≈ 2.71828) is the base of the natural logarithm. By substituting the Taylor expansion of $(1 - |a/a_m|)^{1/2}$ around $|a/a_m| = 0$ into (17b), one can verify that the admissible range of L_-^2/L_+^2 increases toward $(1, \infty)$ as $|a| \to 0$. As $|a|$ increases from 0 to its upper bound a_m in (17a), the admissible range of L_-^2/L_+^2 in (17b) shrinks from $(1, \infty)$ to a single point at $L_-^2/L_+^2 = 2$.

Mathematically, the covariance functions in (15) can be constructed with a and L_-^2/L_+^2 specified within their respective admissible ranges in (17), but the two physical parameters L_r and L_d become implicit and cannot be freely specified. Physically, it is convenient to specify $\beta \equiv \sigma_r^2/\sigma_d^2$, L_r, and L_d first, and then obtain a, L_+, and L_-, and this is facilitated by the following relationships derived from (16a) and (16b):

$$L_r^2/L_d^2 = \beta(L_-^2/L_+^2 - 4a)/(L_-^2/L_+^2 + 4a)$$
$$= \beta[1 - 8a/(L_-^2/L_+^2 + 4a)], \quad (18a)$$

$$L_+^2 = (\beta + 1)L_r^2/(\beta L_r^2/L_d^2 + 1), \quad (18b)$$

$$L_-^2 = 2(\beta - 1)L_r^2/(\beta L_r^2/L_d^2 - 1), \quad (18c)$$

where $2a \equiv (\beta - 1)/(\beta + 1)$ is used. Here, β is confined by

$$1/\beta_m \leq \beta \leq \beta_m, \quad (19)$$

where $\beta_m = (1 + 2a_m)/(1 - 2a_m) \approx 5.02$ and (17a) is used.

If $\beta = 1$ and thus $a = 0$, then (18a) gives $L_r/L_d = 1$ and (18b) gives $L_+ = L_r$ for any given $L_r = L_d$, but (18c) becomes redundant because $a = 0$ and $C_- = 0$ in (15b) and L_- is no longer needed. If $\beta \neq 1$, then L_r^2/L_d^2 is confined by

$$\beta[1 - 8a/(\lambda_{mn} + 4a)] \leq L_r^2/L_d^2 \leq \beta[1 - 8a/(\lambda_{mx} + 4a)]$$
$$\text{for} \quad \beta > 1 \quad (\text{with } a > 0),$$

$$\beta[1 + 8|a|/(\lambda_{mx} - 4|a|)] \leq L_r^2/L_d^2$$
$$\leq \beta[1 + 8|a|/(\lambda_{mn} - 4|a|)] \quad \text{for} \quad \beta < 1 \quad (\text{with } a < 0),$$

where $\lambda_{mn} = 2|a_m/a|[1 - (1 - |a/a_m|)^{1/2}]$, $\lambda_{mx} = 2|a_m/a|[1 + (1 - |a/a_m|)^{1/2}]$, and (17) and (18) are used in the derivation. These conditions lead to $L_r^2/L_d^2 < \beta$ for $\beta > 1$ and $L_r^2/L_d^2 > \beta$ for $\beta < 1$, which ensure $L_-^2 > 0$ in (17b). If $L_r = L_d = L/\sqrt{2}$ as assumed in (13), then (18b) gives $L_+ = L/\sqrt{2}$ and (18c) gives $L_- = L$ for any admissible β in (19).

Modified $C_{ll} = (C_+ + C_-)/2$ and $C_{tt} = (C_+ - C_-)/2$ are constructed from C_+ and C_- in (15) with $L_r = L_d = L/\sqrt{2}$ (and thus $L_+ = L/\sqrt{2}$ and $L_- = L$). These modified C_{ll} and C_{tt} retain the duality associated with the

inversion of β, so they can be plotted compactly by the five curves in Fig. 5 for five different settings of β as those in Fig. 2. The unmodified C_{tt} ($=C_{ll}$) for $\beta = 1$ plotted by the thick blue curve in Fig. 2 is duplicated by the thin dashed blue curve in Fig. 5. This unmodified C_{tt} ($=C_{ll}$) has the same decorrelation length and thus the same curvature at $r = 0$ as the modified C_{tt} ($=C_{ll}$) for $\beta = 1$ plotted by the thick blue curve in Fig. 5, but it decreases rapidly and becomes negative as r increases to $1.4L$ and beyond, while the modified C_{tt} ($=C_{ll}$), given by $\sigma^2 G(r/L)$ for $\beta = 1$, stays above zero as r increases unboundedly. As shown in Fig. 5, the modified C_{tt} for $\beta > 1$ still contains a negative sidelobe but its negative sidelobe is much shallower than that of the unmodified C_{tt} for the same value of $\beta > 1$ in Fig. 2. Similarly, according to the duality between C_{ll} and C_{tt}, the modified C_{ll} for $\beta < 1$ contains a negative sidelobe and its negative sidelobe is much shallower than that of the unmodified C_{ll}. Thus, the modifications made to C_+ and C_- in (15) can reduce but not eliminate the negative sidelobe in C_{tt} for $\beta > 1$ or in C_{ll} for $\beta < 1$.

b. Type-2 modified covariance functions with no negative sidelobe

As shown in (8a), the divergent parts of C_{ll} and C_{tt} are defined by $C_{ll}^d(r) \equiv -\partial_r^2 C_{\chi\chi}(r)$ and $C_{tt}^d(r) \equiv -r^{-1}\partial_r C_{\chi\chi}(r)$, respectively. Based on these definitions, $C_{tt}^d(r)$ can be related to $C_{ll}^d(r)$ by $C_{tt}^d(r) = r^{-1}\int_0^r C_{ll}^d(r') \, dr'$ directly with $C_{\chi\chi}(r)$ dropped from this relationship. Using this relationship, we can set $C_{ll}^d(r) = \sigma_d^2 G(r/L_d)$ and obtain $C_{tt}^d(r) = \sigma_d^2 E(r/L_d)$, where

$$E(r/L_d) \equiv (L_d/r)\int_0^{r/L_d} G(r') \, dr'.$$

Similarly, we can set $C_{tt}^r(r) = \sigma_r^2 G(r/L_r)$ with $C_{ll}^r(r) = r^{-1}\int_0^r C_{tt}^r(r') \, dr'$ and obtain $C_{ll}^r = \sigma_r^2 E(r/L_r)$ for the rotational parts of $C_{ll}(r)$ and $C_{tt}(r)$ defined in (8b).

Combining the above results gives the type-2 modified covariance functions defined below:

$$C_d(r) = \sigma_d^2 [G(r/L_d) + E(r/L_d)] \tag{20a}$$

and

$$C_r(r) = \sigma_r^2 [G(r/L_r) + E(r/L_r)], \tag{20b}$$

or

$$C_{ll}(r) = \sigma^2 [G(r/L_d) + \beta E(r/L_r)]/(\beta + 1) \tag{21a}$$

and

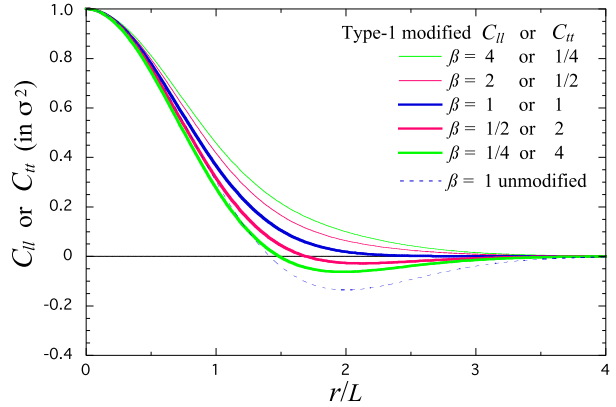$$C_{tt}(r) = \sigma^2 [\beta G(r/L_r) + E(r/L_d)]/(\beta + 1), \tag{21b}$$

FIG. 5. As in Fig. 2, but for type-1 modified $C_{ll} = (C_+ + C_-)/2$ computed from (15) and (18) for $\beta = 4, 2, 1, \frac{1}{2},$ and $\frac{1}{4}$ [or type-1 modified $C_{tt} = (C_+ - C_-)/2$ computed from (15) and (18) for $\beta = \frac{1}{4}, \frac{1}{2}, 1, 2,$ and 4] with $L_d = L_r = L/\sqrt{2}$. The thin blue dashed curve duplicates the thick blue curve in Fig. 2 for the unmodified C_{tt} ($=C_{ll}$) with $\beta = 1$.

where $\sigma^2 \equiv \sigma_r^2 + \sigma_d^2 = (1+\beta)\sigma_d^2$ as defined in (13). For C_d (or C_r) in (20) to have the same decorrelation length as the unmodified C_d (or C_r) in (11), we need to set $L_d = L_\chi/\sqrt{3}$ (or $L_r = L_\psi/\sqrt{3}$) as shown in appendix C, where L_χ (or L_ψ) is the decorrelation length of $C_{\chi\chi}$ (or $C_{\psi\psi}$) in (10a) [or (10b)], so $L_\chi/\sqrt{2}$ (or $L_\psi/\sqrt{2}$) is the decorrelation length of the unmodified C_d (or C_r).

The type-2 modified covariance functions C_{ll} and C_{tt} in (21) retain the duality associated with the inversion of β, so they can be plotted compactly by the five curves in Fig. 6 for five different settings of β with $L_r = L_d = L/\sqrt{3}$. As shown in Fig. 6, the type-2 modified C_{ll} and C_{tt} contain no negative sidelobe for any given β (≥ 0). Their derived C_{uu} (or C_{vv}) also contains no negative sidelobe as shown by the black contours (vs the colored contours) in Fig. 3a for $\beta = 2$. As shown by the thick blue curve in Fig. 6, the type-2 modified C_{tt} ($=C_{ll}$) for $\beta = 1$ has the same decorrelation length and thus the same curvature at $r = 0$ as the type-1 modified C_{tt} ($=C_{ll}$) for $\beta = 1$ (plotted by the thin dashed blue curve in Fig. 6), but it decreases more slowly than the type-1 modified C_{tt} ($=C_{ll}$) as r increases. All the type-2 modified covariance functions in Fig. 6 have fat tails, like the fat-tailed hyper-Gaussians (or multi-Gaussians) that allow a broader dynamical range of scales in the analysis increments to be assimilated (Purser et al. 2003b). This can be an advantage of the type-2 modified covariance functions over the type 1 for variational data assimilation over a wide mesoscale range.

c. Generalized covariance functions for (χ, ψ)

For the type-1 modified covariance functions in (15), $C_{\chi\chi}$ and $C_{\psi\psi}$ can be related to C_+ and C_-, according to (8), by the following equations:
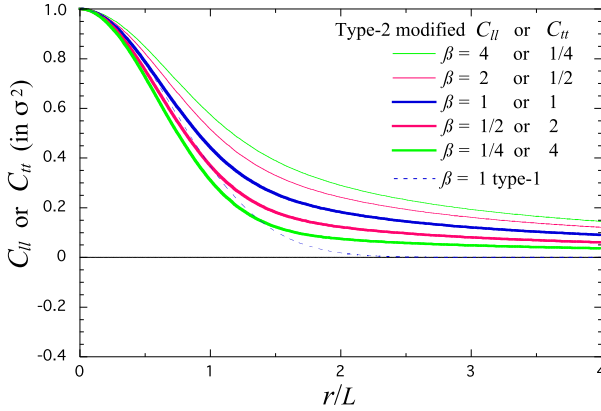
FIG. 6. As in Fig. 5, but for type-2 modified C_{ll} computed from (21a) for $\beta = 4, 2, 1, \frac{1}{2}$, and $\frac{1}{4}$ [or type-2 modified C_{tt} computed from (21b) for $\beta = \frac{1}{4}, \frac{1}{2}, 1, 2$, and 4] with $L_d = L_r = L/\sqrt{3}$. The thin blue dashed curve duplicates the thick blue curve in Fig. 5 for type-1 modified C_{tt} ($=C_{ll}$) with $\beta = 1$.

$$(\partial_r^2 + r^{-1}\partial_r)(C_{\chi\chi} + C_{\psi\psi}) = -C_+ = -2\sigma^2 G(r/L_+), \quad (22a)$$

$$(\partial_r^2 - r^{-1}\partial_r)(C_{\chi\chi} - C_{\psi\psi}) = -C_-$$
$$= -2a\sigma^2(r/L_-)^2 G(r/L_-). \quad (22b)$$

For differentiable $C_{\chi\chi}$ and $C_{\psi\psi}$, $\partial_r C_{\chi\chi}$ and $\partial_r C_{\psi\psi}$ must approach zero as $r \to 0$. Using these conditions with $C_{\chi\chi}(r)|_{r=0} \equiv \sigma_\chi^2$ and $C_{\psi\psi}(r)|_{r=0} \equiv \sigma_\psi^2$, $C_{\chi\chi}$ and $C_{\psi\psi}$ can be solved from (22). The results are

$$C_{\chi\chi} = \sigma_\chi^2 - \sigma^2 L_+^2 \int_0^{r/L_+} dr'[1 - G(r')]/r'$$
$$+ a\sigma^2 L_-^2 [1 - G(r/L_-)], \quad (23a)$$

$$C_{\psi\psi} = \sigma_\psi^2 - \sigma^2 L_+^2 \int_0^{r/L_+} dr'[1 - G(r')]/r'$$
$$- a\sigma^2 L_-^2 [1 - G(r/L_-)]. \quad (23b)$$

Note that the integrand $[1 - G(r)]/r$ is positive for $r > 0$ and approaches zero in proportion to $1/r$ as $r \to \infty$, so the integral approaches ∞ (in proportion to $\ln r$) as $r \to \infty$ [see (D4) in appendix D for a rigorous proof]. Therefore, $C_{\chi\chi}$ and $C_{\psi\psi}$ derived in (23) approach $-\infty$ as $r \to \infty$, so they are not classically valid covariance functions. In this case, as shown in appendix E, the random (χ, ψ) background error fields are no longer homogeneous and isotropic although their gradient fields (associated with the velocity error fields) are homogeneous and isotropic, and this explains why they have no classically valid homogeneous and isotropic covariance functions. However, the above derived $C_{\chi\chi}$ and $C_{\psi\psi}$ can still be used as generalized covariance functions for (χ, ψ)

chosen as momentum control variables, because the relationship between $\mathbf{C_{vv}}$ and $\mathbf{C_{cc}}$ derived in (2) remains valid as shown in appendix F.

For the type-2 modified covariance functions in (21), $C_{\chi\chi}$ can be derived by integrating $\partial_r^2 C_{\chi\chi}(r) = -C_{ll}^d(r) = -\sigma_d^2 G(r/L_d)$ directly from $r = 0$ with $\partial_r C_{\chi\chi}(r)|_{r=0} = 0$ and $C_{\chi\chi}(r)|_{r=0} \equiv \sigma_\chi^2$. The result is

$$C_{\chi\chi}(r) = \sigma_\chi^2 - \sigma_d^2 \int_0^r \int_0^{r'} G(r''/L_d)\, dr''\, dr'. \quad (24a)$$

Note that $\int_0^r G(r'/L_d)\, dr' > 0$ for $r > 0$ and $\int_0^r G(r'/L_d)\, dr'$ approaches the constant value of $(\pi/2)^{1/2} L_d$ rapidly as r increases, so the integral on the right-hand side of (24a) approaches infinity (in proportion to r) as $r \to \infty$. Therefore, $C_{\chi\chi}$ derived in (24a) approaches $-\infty$ as $r \to \infty$ and thus is not a classically valid covariance function. Similarly,

$$C_{\psi\psi}(r) = \sigma_\psi^2 - \sigma_r^2 \int_0^r \int_0^{r'} G(r''/L_r)\, dr''\, dr' \quad (24b)$$

derived from $\partial_r^2 C_{\psi\psi}(r) \equiv -C_{tt}^r(r) = -\sigma_r^2 G(r/L_r)$ is also not a classically valid covariance function. Again, the above derived $C_{\chi\chi}$ and $C_{\psi\psi}$ can be used as generalized covariance functions for (χ, ψ) chosen as momentum control variables, because the relationship between $\mathbf{C_{vv}}$ and $\mathbf{C_{cc}}$ derived in (2) remains valid as shown in appendix F.

The above results indicate that when the velocity components are chosen as momentum control variables with the modified velocity covariance functions of either type, the derived $C_{\chi\chi}(r)$ and $C_{\psi\psi}(r)$ become invalid in the classical sense but still can be used as generalized covariance functions for (χ, ψ) chosen as momentum control variables. However, when δ and ζ are chosen as momentum control variables with $C_{\delta\delta}$ and $C_{\zeta\zeta}$ modeled by Gaussian functions, not only $C_{\chi\chi}(r)$ and $C_{\psi\psi}(r)$ but also $C_d(r)$ and $C_r(r)$ derived from $C_{\delta\delta}(r)$ and $C_{\zeta\zeta}(r)$ become unbounded (see appendix D), so the analyzed incremental wind field becomes unbounded [as indicated by the solution form in (A9)]. This implies that $C_{\delta\delta}(r)$ and $C_{\zeta\zeta}(r)$ cannot be modeled merely as classically valid covariance functions and they must be also derivable at least from classically valid velocity covariance functions.

5. Conclusions

In this paper the issue from recent studies (Xie and MacDonald 2012; Sun et al. 2016) on the choice of momentum control variables and related covariance modeling for mesoscale high-resolution variational data assimilation is addressed from theoretical perspectives.

In particular, the relationships between background error covariance functions are derived for three different choices of momentum control variables; that is, (i) the horizontal velocity (u, v), (ii) the velocity potential χ and streamfunction ψ, and (iii) the divergence δ and vorticity ζ. These different choices of momentum control variables with their error covariance functions satisfying the derived relationships are proven to be theoretically equivalent in the sense that they lead to the same optimally analyzed incremental wind field in the continuous limit. As the relationships between covariance functions are derived in general nonhomogeneous and nonisotropic forms, the proved theoretical equivalence is applicable to not only variational data assimilation (with homogeneous and isotropic background error covariance formulations) but also ensemble-based data assimilation (with flow-dependent nonhomogeneous and nonisotropic background error covariance formulations).

Although the aforementioned different choices of momentum control variables are equivalent in theory, they are not equivalent in practical applications, as their required numerical schemes and computational costs can be very different for obtaining numerical solutions in discrete spaces, but this topic is beyond this paper. Furthermore, their theoretical equivalence requires the velocity covariance tensor function to be valid in the classic sense (i.e., to not only satisfy the required symmetry and positive semidefiniteness but also approach zero as the separation distance increases boundlessly). Note that the classical validity of a (χ, ψ) covariance tensor function ensures the classical validity of (u, v) and (δ, ζ) covariance tensor functions, as the latter functions are derived from the former via differentiations [see (2) and (3)]. The inverse, however, does not hold true, and this can limit the (δ, ζ) covariance modeling but not the choices of control variables (see section 4c). In particular, when (δ, ζ) are chosen as momentum control variables with their covariance functions modeled by Gaussian functions, the derived velocity covariance functions become unbounded (see appendix D) and so does the analyzed incremental wind field. This implies that if (δ, ζ) are used as momentum control variables, their error covariance functions must be derivable from classically valid velocity covariance functions.

It is also shown in this paper that the (u, v) autocovariance functions derived from single-Gaussian (χ, ψ) autocovariance functions contain significant negative sidelobes. These negative sidelobes can represent background wind error structures associated with the recirculations arising from the near geostrophy and prevailing nondivergence as well as with the dominant baroclinic waves on the synoptic scale, but they become unrepresentative on the mesoscale (as

shown in sections 3b and 3c). Using (u, v) directly as independent momentum control variables with their autocovariance functions modeled by Gaussian functions can remove negative sidelobes, but this velocity covariance modeling is not consistent with the assumed isotropy in variational data assimilation in general [unless $C_d = C_r$ or, equivalently, $C_{ll} = C_{tt}$ is further assumed as shown by case 1 in section 3a].

To solve the above problem, Gaussian covariance functions are modified in two ways to model the isotropic canonical-form velocity autocovariance functions with negative sidelobes reduced or removed in consistency with the assumed isotropy for mesoscale variational data assimilation. In this case, (χ, ψ) covariance functions can be derived by integrating the modified velocity covariance functions and used as generalized covariance functions for (χ, ψ) chosen as momentum control variables, although the derived (χ, ψ) covariance functions are not classically valid as they approach negative infinity in the limit of infinitely large separation distance (see section 4c).

When negative sidelobes are reduced or removed in the modified velocity covariance functions, the velocity power spectra become nonzero and their derived (χ, ψ) power spectra become singular at zero wavenumber [see (E1) in appendix E]. These properties are further explored mathematically in appendixes E and F, and the results can be summarized in the following implications that can be important and/or useful for mesoscale data assimilation:

1) The zero-wavenumber and near-zero-wavenumber components in the power spectra of the modified velocity covariance can represent, to a certain extent, the synoptic-scale components not represented by the velocity covariance functions derived from single-Gaussian mesoscale (χ, ψ) covariance functions.

2) Generalized (χ, ψ) covariance functions derived from the modified velocity covariance functions contain unbounded structures, but these unbounded structures are bounded in a mesoscale domain and thus can represent, to a certain extent, the synoptic-scale components [as shown explicitly by the one-dimensional example in (G8)–(G15) of appendix G].

3) For mesoscale data assimilation, the covariance structures should be modeled by function forms more general than single Gaussians. These include not only the two types proposed in this paper but also those comprising multiple Gaussians with their decorrelation lengths ranging up to the synoptic scale (at which point the presence of negative sidelobes becomes acceptable). Multi-Gaussians (Purser et al. 2003b) are more flexible and thus can be more

accurate than the two types proposed in this paper for mesoscale data assimilation, especially if the true covariance can be well estimated and represented by a multi-Gaussian on the mesoscale and synoptic scale. Since the true covariances are often largely unknown or not well estimated on the mesoscale, the modifications proposed in this paper provide simple theoretically justifiable choices for mesoscale data assimilation.

4) Background velocity covariance functions C_+ and C_- represented by spectral expansions can be estimated directly from radar radial wind innovations on the mesoscale [see (12)–(14) of X07a]. These spectral representations, if well estimated, can be used directly with their zero-wavenumber components retained to represent the synoptic-scale components that are not resolved because of the limited innovation data coverage. Furthermore, the synoptic-scale components represented by the zero-wavenumber components can be estimated separately from radiosonde innovations and then used to replace the zero-wavenumber components and thus expand the mesoscale spectral representations to the synoptic scale [as shown by the one-dimensional example in (G16) with its two-dimensional extension suggested for real-data assimilation at the end of appendix G]. In this case, the geostrophy can be conveniently built into multivariate covariance on the synoptic scale.

5) Covariance localizations have been commonly used for ensemble-based data assimilation. When such an ensemble-based method is used for mesoscale data assimilation, the localization radiuses often cover only small fractions of the mesoscale domain and thus inevitably eliminate all the large-scale structures beyond the localization radiuses. This may partially explain in theory why a localized flow-dependent covariance estimated by ensemble perturbations needs to be combined with prescribed (or preestimated) covariance with smooth structures well beyond the localization radiuses via a hybrid approach.

The two types of modified velocity covariance functions formulated in this paper (see sections 4a and 4b) can be used directly to construct the background covariance matrix in 3DVar for mesoscale wind analyses and data assimilation. In this way the utility in 3DVar and advantages of the type-2 modified velocity covariance functions over their unmodified counterparts have been demonstrated and evaluated by comparative experiments with radar radial-velocity observations. The detailed results were presented in section 5 of Xu et al. (2017). The modified velocity covariance functions will be incorporated, as enhanced options, into the recently developed radar wind analysis system (Xu et al. 2015).

APPENDIX A

Equivalence of Different Choices of Momentum Control Variables in the Continuous Limit

a. Cost functional for choice 1 and its minimizer in the continuous limit

For choice 1, we have $\Delta \mathbf{a} \rightarrow \mathbf{v}(\mathbf{z})$ and $\mathbf{B} \rightarrow \mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}')$ [see (1)] in the continuous limit. The space for $\mathbf{v}(\mathbf{z})$ can be completed as a Hilbert space (Courant and Hilbert 1962) [or, specifically, a Sobolev space (Adams 1975) as shown by the examples in sections 4–5 of Xu (2005)], denoted by $H_{\mathbf{v}}$, with the following inner product:

$$\langle \mathbf{f}(\mathbf{z}), \mathbf{g}(\mathbf{z}) \rangle_{\mathbf{v}} \equiv \int d\mathbf{z} \int d\mathbf{z}' \mathbf{f}(\mathbf{z})^{\mathrm{T}} \mathbf{Q_{vv}}(\mathbf{z}, \mathbf{z}') \mathbf{g}(\mathbf{z}'), \quad \text{(A1a)}$$

where $\mathbf{f}(\mathbf{z})$ and $\mathbf{g}(\mathbf{z})$ denote two arbitrary vector functions in $H_{\mathbf{v}}$, $\mathbf{Q_{vv}}(\mathbf{z}, \mathbf{z}')$ is the inverse of $\mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}')$ defined [see section 5.3 of Bennett (1992)] by

$$\int d\mathbf{z}' \mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}') \mathbf{Q_{vv}}(\mathbf{z}', \mathbf{z}'') = \delta(\mathbf{z} - \mathbf{z}'') \mathbf{I}, \quad \text{(A1b)}$$

the integral $\int d\mathbf{z}$ is over the entire space of $\mathbf{z}$, and $\mathbf{I} \equiv (1, 1)^{\mathrm{diag}}$. In this case we have

$$\Delta \mathbf{a}^{\mathrm{T}} \mathbf{B}^{-1} \Delta \mathbf{a} \rightarrow \int d\mathbf{z} \int d\mathbf{z}' \mathbf{v}(\mathbf{z})^{\mathrm{T}} \mathbf{Q_{vv}}(\mathbf{z}, \mathbf{z}') \mathbf{v}(\mathbf{z}') = \|\mathbf{v}(\mathbf{z})\|_{\mathbf{v}}^2$$
$$\text{(A2)}$$

in the continuous limit, where $\|\mathbf{v}(\mathbf{z})\|_{\mathbf{v}}^2 \equiv \langle \mathbf{v}(\mathbf{z}), \mathbf{v}(\mathbf{z}) \rangle_{\mathbf{v}}$ is the squared norm of $\mathbf{v}(\mathbf{z})$ in $H_{\mathbf{v}}$.

Note that $\mathbf{H}\Delta\mathbf{a}$ is a vector in the observation space $\mathbf{R}^M$ of finite dimension M, where M is the total number of horizontal wind component observations. The mth element of $\mathbf{H}\Delta\mathbf{a}$, denoted by $\{\mathbf{H}\Delta\mathbf{a}\}_m$, corresponds to the mth observed velocity component at the observation point $\mathbf{z}_m$. In the continuous limit of $\Delta\mathbf{a}$, $\mathbf{H}\Delta\mathbf{a}$ is still a vector in $\mathbf{R}^M$ with its mth element given by

$$\{\mathbf{H}\Delta\mathbf{a}\}_m \rightarrow \int d\mathbf{z}\delta(\mathbf{z}_m - \mathbf{z})\mathbf{r}_m^{\mathrm{T}}\mathbf{v}(\mathbf{z}) = \langle \mathbf{r}_m^{\mathrm{T}}\mathbf{C_{vv}}(\mathbf{z}_m, \mathbf{z}), \mathbf{v}(\mathbf{z}) \rangle_{\mathbf{v}},$$
$$\text{(A3)}$$

where $\mathbf{r}_m$ denotes the unit vector along the mth observed horizontal wind component at the observation point $\mathbf{z}_m$. The last step of (A3) is derived by using $\delta(\mathbf{z}_m - \mathbf{z})\mathbf{r}_m^{\mathrm{T}} = \mathbf{r}_m^{\mathrm{T}}\int d\mathbf{z}'\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q}_{\mathbf{vv}}(\mathbf{z}', \mathbf{z})$, according to (A1b), with the definition of $\langle \cdot, \cdot \rangle_{\mathbf{v}}$ in (A1a). From (A3), we have

$$\mathbf{H}\Delta\mathbf{a} \rightarrow \langle \mathbf{L}_{\mathbf{v}}(\mathbf{z}), \mathbf{v}(\mathbf{z}) \rangle_{\mathbf{v}}, \qquad (\text{A4})$$

where $\mathbf{L}_{\mathbf{v}}(\mathbf{z})$ is a $M \times 2$ matrix with its mth row given by $\mathbf{r}_m^{\mathrm{T}}\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_m, \mathbf{z})$.

Substituting (A2) and (A4) into (4a) in the continuous limit gives

$$J(\mathbf{v}) = \|\mathbf{v}(\mathbf{z})\|_{\mathbf{v}}^2 + (\langle \mathbf{L}_{\mathbf{v}}(\mathbf{z}), \mathbf{v}(\mathbf{z}) \rangle_{\mathbf{v}} - \mathbf{d})^{\mathrm{T}}$$
$$\times \mathbf{R}^{-1}(\langle \mathbf{L}_{\mathbf{v}}(\mathbf{z}), \mathbf{v}(\mathbf{z}) \rangle_{\mathbf{v}} - \mathbf{d}). \qquad (\text{A5})$$

As a function in $H_{\mathbf{v}}$, $\mathbf{v}(\mathbf{z})$ in (A5) can be expressed in general by

$$\mathbf{v}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L}_{\mathbf{v}}(\mathbf{z}) + \mathbf{g}_{\mathbf{v}}(\mathbf{z}), \qquad (\text{A6})$$

where $\mathbf{p} \in \mathbf{R}^M$ is an intermediate vector to be determined in the observation space and $\mathbf{g}_{\mathbf{v}}(\mathbf{z})$ is an element in $H_{\mathbf{v}}$ orthogonal to $\mathbf{L}_{\mathbf{v}}(\mathbf{z})$, that is, $\langle \mathbf{g}_{\mathbf{v}}(\mathbf{z}), \mathbf{L}_{\mathbf{v}}(\mathbf{z}) \rangle_{\mathbf{v}} = 0$. Substituting (A6) into (A5) gives

$$J(\mathbf{v}) = \mathbf{p}^{\mathrm{T}}\mathbf{\Pi}\mathbf{p} + \|\mathbf{g}_{\mathbf{v}}(\mathbf{z})\|_{\mathbf{v}}^2$$
$$+ (\mathbf{\Pi}\mathbf{p} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{\Pi}\mathbf{p} - \mathbf{d}), \qquad (\text{A7})$$

where $\mathbf{\Pi} = \langle \mathbf{L}_{\mathbf{v}}(\mathbf{z}), \mathbf{L}_{\mathbf{v}}^{\mathrm{T}}(\mathbf{z}) \rangle_{\mathbf{v}}$ is an $M \times M$ symmetric matrix in the observation space with its diagonal (or off diagonal) element given by $\mathbf{r}_m^{\mathrm{T}}\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_m, \mathbf{z}_m)\mathbf{r}_m$ [or $\mathbf{r}_m^{\mathrm{T}}\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_m, \mathbf{z}_n)\mathbf{r}_n$] corresponding to the mth observation (or the paired mth and nth observations).

Clearly, $\|\mathbf{g}_{\mathbf{v}}(\mathbf{z})\|_{\mathbf{v}}^2$ is nonnegative and thus must be zero if $\mathbf{v}(\mathbf{z})$ minimizes J in (A7). This implies that $\mathbf{g}_{\mathbf{v}}(\mathbf{z})$ is unrepresented by observations and must be discarded, so the infinite-dimensional cost functional $J(\mathbf{v})$ in (A7) reduces to the following finite-dimensional form:

$$J = \mathbf{p}^{\mathrm{T}}\mathbf{\Pi}\mathbf{p} + (\mathbf{\Pi}\mathbf{p} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{\Pi}\mathbf{p} - \mathbf{d}). \qquad (\text{A8})$$

The minimizer of $J(\mathbf{v})$ in (A5) is thus given by

$$\mathbf{v}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L}_{\mathbf{v}}(\mathbf{z}) = \mathbf{L}_{\mathbf{v}}(\mathbf{z})^{\mathrm{T}}\mathbf{p} = \sum_m \mathbf{C}_{\mathbf{vv}}(\mathbf{z}, \mathbf{z}_m)\mathbf{r}_m p_m, \quad (\text{A9})$$

where p_m is the mth component of $\mathbf{p}$, and $\mathbf{p}$ is the solution of the following equation [derived from $\partial J/\partial\mathbf{p} = 0$ for J in (A8)]:

$$(\mathbf{\Pi} + \mathbf{R})\mathbf{p} = \mathbf{d}. \qquad (\text{A10})$$

Here, the advanced method of representer solution for imperfect-model 4DVar (Bennett 1992, chapter 5) is simplified for 3DVar (or perfect-model 4DVar) to derive the solution in (A9).

b. Choice 2 and its equivalence to choice 1 in the continuous limit

For choice 2, we have $\mathbf{a} \rightarrow \mathbf{c}(\mathbf{z})$ and $\mathbf{B} \rightarrow \mathbf{C}_{\mathbf{cc}}(\mathbf{z}, \mathbf{z}')$ in the continuous limit. The space for $\mathbf{c}(\mathbf{z})$ can be completed as a Hilbert space, denoted by $H_{\mathbf{c}}$, with the inner product, denoted by $\langle \cdot, \cdot \rangle_{\mathbf{c}}$, defined similarly to that in (A1) except that $\mathbf{C}_{\mathbf{vv}}(\mathbf{z}, \mathbf{z}')$ and $\mathbf{Q}_{\mathbf{vv}}(\mathbf{z}, \mathbf{z}')$ are replaced by $\mathbf{C}_{\mathbf{cc}}(\mathbf{z}, \mathbf{z}')$ and $\mathbf{Q}_{\mathbf{cc}}(\mathbf{z}, \mathbf{z}')$, respectively. In this case, we have

$$\Delta\mathbf{a}^{\mathrm{T}}\mathbf{B}^{-1}\Delta\mathbf{a} \rightarrow \|\mathbf{c}(\mathbf{z})\|_{\mathbf{c}}^2, \qquad (\text{A11})$$

$$\{\mathbf{H}\Delta\mathbf{a}\}_m \rightarrow \int d\mathbf{z}\,\delta(\mathbf{z}_m - \mathbf{z})\mathbf{r}_m^{\mathrm{T}}[(\nabla, \mathbf{k} \times \nabla)\mathbf{c}(\mathbf{z})] = \int d\mathbf{z}\,\mathbf{r}_m^{\mathrm{T}}[-(\nabla, \mathbf{k} \times \nabla)\delta(\mathbf{z}_m - \mathbf{z})\mathbf{I}]\mathbf{c}(\mathbf{z})$$

$$= \int d\mathbf{z}\,\mathbf{r}_m^{\mathrm{T}}[(\nabla_m, \mathbf{k} \times \nabla_m)\delta(\mathbf{z}_m - \mathbf{z})\mathbf{I}]\mathbf{c}(\mathbf{z}) = \int d\mathbf{z}\,\mathbf{r}_m^{\mathrm{T}}\left[(\nabla_m, \mathbf{k} \times \nabla_m)\int d\mathbf{z}'\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q}_{\mathbf{cc}}(\mathbf{z}', \mathbf{z})\right]\mathbf{c}(\mathbf{z})$$

$$= \int d\mathbf{z}\int d\mathbf{z}'\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q}_{\mathbf{cc}}(\mathbf{z}', \mathbf{z})\mathbf{c}(\mathbf{z}) = \langle \mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_m, \mathbf{z}), \mathbf{c}(\mathbf{z}) \rangle_{\mathbf{c}}. \qquad (\text{A12})$$

In the derivation of (A12), $\mathbf{v}(\mathbf{z}) = (\nabla, \mathbf{k} \times \nabla)\mathbf{c}(\mathbf{z})$ is used in the first step, integration by parts is used in the second step, $-(\nabla, \mathbf{k} \times \nabla)\delta(\mathbf{z}_m - \mathbf{z}) = (\nabla_m, \mathbf{k} \times \nabla_m)\delta(\mathbf{z}_m - \mathbf{z})$ is used (see section 2.2 of Lighthill 1958) in the third step, $\int d\mathbf{z}'\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q}_{\mathbf{cc}}(\mathbf{z}', \mathbf{z}) = \delta(\mathbf{z}_m - \mathbf{z})\mathbf{I}$ is used in the fourth step, and the definition of $\langle \cdot, \cdot \rangle_{\mathbf{c}}$ is used in the last step. From (A12), we have

$$\mathbf{H}\Delta\mathbf{a} \rightarrow \langle \mathbf{L}_{\mathbf{c}}(\mathbf{z}), \mathbf{c}(\mathbf{z}) \rangle_{\mathbf{c}}, \qquad (\text{A13})$$

where $\mathbf{L}_{\mathbf{c}}(\mathbf{z})$ is an $M \times 2$ matrix with its mth row given by $\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_m, \mathbf{z})$.

Substituting (A11) and (A13) into (4a) in the continuous limit gives

$$J(\mathbf{c}) = \|\mathbf{c}(\mathbf{z})\|_{\mathbf{c}}^2 + [\langle \mathbf{L_c}(\mathbf{z}), \mathbf{c}(\mathbf{z}) \rangle_{\mathbf{c}} - \mathbf{d}]^{\mathrm{T}}$$
$$\times \mathbf{R}^{-1}[\langle \mathbf{L_c}(\mathbf{z}), \mathbf{c}(\mathbf{z}) \rangle_{\mathbf{c}} - \mathbf{d}]. \qquad \text{(A14)}$$

Similar to $\mathbf{v}(\mathbf{z})$ in (A6), $\mathbf{c}(\mathbf{z})$ can be expressed by $\mathbf{c}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L_c}(\mathbf{z}) + \mathbf{g_c}(\mathbf{z})$, where $\mathbf{g_c}(\mathbf{z})$ is an element in $H_{\mathbf{c}}$ orthogonal to $\mathbf{L_c}(\mathbf{z})$. Using this expression together with the relationship of $\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)[(\nabla_n, \mathbf{k} \times \nabla_n)\mathbf{C_{cc}}(\mathbf{z}_n, \mathbf{z}_m)]^{\mathrm{T}}\mathbf{r}_n = \mathbf{r}_m^{\mathrm{T}}\mathbf{C_{vv}}(\mathbf{z}_m, \mathbf{z}_n)\mathbf{r}_n$ derived from (2) or, equivalently, $\langle \mathbf{L_c}(\mathbf{z}), \mathbf{L_c}(\mathbf{z})^{\mathrm{T}} \rangle_{\mathbf{c}} = \langle \mathbf{L_v}(\mathbf{z}), \mathbf{L_v}(\mathbf{z})^{\mathrm{T}} \rangle_{\mathbf{v}} = \boldsymbol{\Pi}$, the cost functional in (A14) can be transformed into the same form as that in (A7) but with $\|\mathbf{g_v}(\mathbf{z})\|_{\mathbf{v}}^2$ replaced by $\|\mathbf{g_c}(\mathbf{z})\|_{\mathbf{c}}^2$. Again, $\|\mathbf{g_c}(\mathbf{z})\|_{\mathbf{c}}^2$ is nonnegative and thus must be zero at the minimum of $J(\mathbf{c})$, so $\mathbf{g_c}(\mathbf{z})$ must be discarded, reducing $J(\mathbf{c})$ to the same finite-dimensional form as in (A8). Thus, similar to the solution in (A9), the minimizer of $J(\mathbf{c})$ in (A14) is given by

$$\mathbf{c}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L_c}(\mathbf{z}), \qquad \text{(A15)}$$

with $\mathbf{p}$ solved from (A10). According to (2), we have $(\nabla, \mathbf{k} \times \nabla)[\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\mathbf{C_{cc}}(\mathbf{z}_m, \mathbf{z})]^{\mathrm{T}} = (\nabla, \mathbf{k} \times \nabla)$

$$[(\nabla_m, \mathbf{k} \times \nabla_m)\mathbf{C_{cc}}(\mathbf{z}_m, \mathbf{z})]^{\mathrm{T}}\mathbf{r}_m = \mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}_m)\mathbf{r}_m = \mathbf{r}_m^{\mathrm{T}}\mathbf{C_{vv}}(\mathbf{z}_m, \mathbf{z}),$$
which gives $(\nabla, \mathbf{k} \times \nabla)\mathbf{L_c}(\mathbf{z})^{\mathrm{T}} = \mathbf{L_v}(\mathbf{z})^{\mathrm{T}}$, so

$$(\nabla, \mathbf{k} \times \nabla)\mathbf{c}(\mathbf{z}) = (\nabla, \mathbf{k} \times \nabla)[\mathbf{p}^{\mathrm{T}}\mathbf{L_c}(\mathbf{z})]$$
$$= (\nabla, \mathbf{k} \times \nabla)\mathbf{L_c}(\mathbf{z})^{\mathrm{T}}\mathbf{p} = \mathbf{L_v}(\mathbf{z})^{\mathrm{T}}\mathbf{p} = \mathbf{p}^{\mathrm{T}}\mathbf{L_v}(\mathbf{z}).$$
$$\text{(A16)}$$

This proves that $\mathbf{c}(\mathbf{z})$ in (A15) is equivalent to $\mathbf{v}(\mathbf{z})$ in (A9).

c. Choice 3 and its equivalence to choice 2 in the continuous limit

For choice 3, we have $\mathbf{a} \to \mathbf{q}(\mathbf{z})$ and $\mathbf{B} \to \mathbf{C_{qq}}(\mathbf{z}, \mathbf{z}')$ in the continuous limit. The space for $\mathbf{q}(\mathbf{z})$ can be completed as a Hilbert space, denoted by $H_{\mathbf{q}}$, with the inner product, denoted by $\langle \cdot, \cdot \rangle_{\mathbf{q}}$, defined similarly to that in (A1) except that $\mathbf{C_{vv}}(\mathbf{z}, \mathbf{z}')$ and $\mathbf{Q_{vv}}(\mathbf{z}, \mathbf{z}')$ are replaced by $\mathbf{C_{qq}}(\mathbf{z}, \mathbf{z}')$ and $\mathbf{Q_{qq}}(\mathbf{z}, \mathbf{z}')$, respectively. In this case, we have

$$\Delta \mathbf{a}^{\mathrm{T}}\mathbf{B}^{-1}\Delta \mathbf{a} \to \|\mathbf{q}(\mathbf{z})\|_{\mathbf{q}}^2, \qquad \text{(A17)}$$

$$\{\mathbf{H}\Delta\mathbf{a}\}_m \to \int d\mathbf{z}\,\delta(\mathbf{z}_m - \mathbf{z})\mathbf{r}_m^{\mathrm{T}}[(\nabla, \mathbf{k} \times \nabla)\nabla^{-2}\mathbf{q}(\mathbf{z})] = \int d\mathbf{z}\,\mathbf{r}_m^{\mathrm{T}}[\nabla^{-2}(-\nabla, -\mathbf{k} \times \nabla)\delta(\mathbf{z}_m - \mathbf{z})\mathbf{I}]\nabla^{-2}\mathbf{q}(\mathbf{z})$$

$$= \int d\mathbf{z}\,\mathbf{r}_m^{\mathrm{T}}[(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\delta(\mathbf{z}_m - \mathbf{z})\mathbf{I}]\mathbf{q}(\mathbf{z}) = \int d\mathbf{z}\,\mathbf{r}_m^{\mathrm{T}}\left[(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\int d\mathbf{z}'\mathbf{C_{qq}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q_{qq}}(\mathbf{z}' - \mathbf{z})\right]\mathbf{q}(\mathbf{z})$$

$$= \int d\mathbf{z}\int d\mathbf{z}'\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\mathbf{C_{qq}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q_{qq}}(\mathbf{z}' - \mathbf{z})\mathbf{q}(\mathbf{z}) = \langle \mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\mathbf{C_{qq}}(\mathbf{z}_m, \mathbf{z}), \mathbf{c}(\mathbf{z}) \rangle_{\mathbf{q}}, \qquad \text{(A18)}$$

where $\nabla^{-2} = \int d\mathbf{x}' \ln|\mathbf{x}' - \mathbf{x}|/(2\pi)$ is the inverse of ∇^2 [see (3.1) of Xu et al. 2011], the integral $\int d\mathbf{x}'$ is over the entire horizontal space of $\mathbf{x}'$, and similarly ∇_m^{-2} is the inverse of ∇_m^2. In the derivation of (A18), $\mathbf{v}(\mathbf{z}) = (\nabla, \mathbf{k} \times \nabla)\nabla^{-2}\mathbf{q}(\mathbf{z})$ is used in the first step, integration by parts is used in the second step, $\nabla^{-2}(-\nabla, \mathbf{k} \times \nabla)\delta(\mathbf{z}_m - \mathbf{z}) = (\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\delta(\mathbf{z}_m - \mathbf{z})$ is used in the third step, $\int d\mathbf{z}'\mathbf{C_{qq}}(\mathbf{z}_m, \mathbf{z}')\mathbf{Q_{qq}}(\mathbf{z}', \mathbf{z}) = \delta(\mathbf{z}_m - \mathbf{z})\mathbf{I}$ is used in the fourth step, and the definition of $\langle \cdot, \cdot \rangle_{\mathbf{q}}$ is used in the last step. From (A18) we have

$$\mathbf{H}\Delta\mathbf{a} \to \langle \mathbf{L_q}(\mathbf{z}), \mathbf{q}(\mathbf{z}) \rangle_{\mathbf{q}}, \qquad \text{(A19)}$$

where $\mathbf{L_q}(\mathbf{z})$ is an $M \times 2$ matrix with its mth row given by $\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\mathbf{C_{qq}}(\mathbf{z}_m, \mathbf{z})$.

Substituting (A17) and (A19) into (4a) in the continuous limit gives

$$J(\mathbf{q}) = \|\mathbf{q}(\mathbf{z})\|_{\mathbf{q}}^2 + [\langle \mathbf{L_q}(\mathbf{z}), \mathbf{q}(\mathbf{z}) \rangle_{\mathbf{q}} - \mathbf{d}]^{\mathrm{T}}$$
$$\times \mathbf{R}^{-1}[\langle \mathbf{L_q}(\mathbf{z}), \mathbf{q}(\mathbf{z}) \rangle_{\mathbf{q}} - \mathbf{d}]. \qquad \text{(A20)}$$

Again, as $\mathbf{v}(\mathbf{z})$ in (A6), $\mathbf{q}(\mathbf{z})$ can be expressed by $\mathbf{q}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L_q}(\mathbf{z}) + \mathbf{g_q}(\mathbf{z})$, where $\mathbf{g_q}(\mathbf{z})$ is an element in $H_{\mathbf{q}}$ orthogonal to $\mathbf{L_q}(\mathbf{z})$. Using this expression together with the relationship of $\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}[(\nabla_n, \mathbf{k} \times \nabla_n)\nabla_n^{-2}\mathbf{C_{qq}}(\mathbf{z}_n, \mathbf{z}_m)]^{\mathrm{T}}\mathbf{r}_n = \mathbf{r}_m^{\mathrm{T}}\mathbf{C_{vv}}(\mathbf{z}_m, \mathbf{z})\mathbf{r}_n$ derived from (2) and (3) or, equivalently, $\langle \mathbf{L_q}(\mathbf{z}), \mathbf{L_q}(\mathbf{z})^{\mathrm{T}} \rangle_{\mathbf{q}} = \langle \mathbf{L_v}(\mathbf{z}), \mathbf{L_v}(\mathbf{z})^{\mathrm{T}} \rangle_{\mathbf{v}} = \boldsymbol{\Pi}$, the cost functional in (A20) can be transformed into the same form as that in (A7) but with $\|\mathbf{g_v}(\mathbf{z})\|_{\mathbf{v}}^2$ replaced by $\|\mathbf{g_q}(\mathbf{z})\|_{\mathbf{q}}^2$. Again, $\|\mathbf{g_q}(\mathbf{z})\|_{\mathbf{q}}^2$ is nonnegative, so $\mathbf{g_q}(\mathbf{z})$ must be discarded, reducing $J(\mathbf{q})$ to the same finite-dimensional form as in (A8). The minimizer of $J(\mathbf{q})$ in (A20) is thus given by

$$\mathbf{q}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L_q}(\mathbf{z}) \qquad \text{(A21)}$$

with $\mathbf{p}$ solved from (A10). Note that $\nabla^{-2}[\mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\nabla_m^{-2}\mathbf{C_{qq}}(\mathbf{z}_m, \mathbf{z})] = \mathbf{r}_m^{\mathrm{T}}(\nabla_m, \mathbf{k} \times \nabla_m)\mathbf{C_{cc}}(\mathbf{z}_m, \mathbf{z})$ according to (3) or, equivalently, $\nabla^{-2}\mathbf{L_q}(\mathbf{z}) = \mathbf{L_c}(\mathbf{z})$, so

$$\nabla^{-2}\mathbf{q}(\mathbf{z}) = \nabla^{-2}[\mathbf{p}^{\mathrm{T}}\mathbf{L_q}(\mathbf{z})] = \mathbf{p}^{\mathrm{T}}\nabla^{-2}\mathbf{L_q}(\mathbf{z}) = \mathbf{p}^{\mathrm{T}}\mathbf{L_c}(\mathbf{z}).$$
$$\text{(A22)}$$

This proves that $\mathbf{q}(\mathbf{z})$ in (A21) is equivalent to $\mathbf{c}(\mathbf{z})$ in (A15).

d. Extension to three-dimensional velocity observations

For simplicity, we have so far considered only horizontal wind observations. This simplification is conventional in operation data assimilation, since operationally assimilated wind observations are either essentially horizontal or treated approximately as horizontal velocities. In particular, operationally assimilated radar radial-component velocity observations are commonly limited to those scanned at low-elevation angles (nearly horizontal) and thus treated as horizontal velocities. However, when radar radial-component velocity observations scanned at high-elevation angles ($>5°$) are also assimilated, a bias correction should be made first to each observed radial-component velocity (often as a part of data quality control) by subtracting the projection of precipitate hydrometeors' downward terminal velocity (estimated from radar observed reflectivity) on the radar beam direction (because it is the precipitate hydrometeors' velocity that is measured by a radar and the precipitate hydrometeors largely follow the air motion in the horizontal but fall downward relative to the air motion in the vertical). The bias-corrected radial-component velocity observation can be then related, still linearly, to the projection of the model's three-dimensional velocity on the radar beam direction. In this case the theoretical equivalence between the three different choices of momentum control variables can be proven in the same way as shown in the previous subsections except that the observation operators in (A4), (A13), and (A19) need to be extended for three-dimensional velocity observations. The detailed extensions are described below.

For three-dimensional velocity observations, the two-dimensional unit vector $\mathbf{r}_m$ introduced in the previous subsections should be extended to denote the three-dimensional unit vector along the mth observed component of three-dimensional velocity at the observation point $\mathbf{z}_m$. In this case (A3) is extended to the following form:

$$\{\mathbf{H}\Delta\mathbf{a}\}_m \rightarrow \int d\mathbf{z}\,\delta(\mathbf{z}_m - \mathbf{z})\mathbf{r}_m^T[\mathbf{v}(\mathbf{z})^T, w(\mathbf{z})]^T$$
$$= \langle \mathbf{r}_m^T \mathbf{P}_m \mathbf{C}_{\mathbf{vv}}^T(\mathbf{z}_m, \mathbf{z}), \mathbf{v}(\mathbf{z}) \rangle_{\mathbf{v}},$$

where $w(\mathbf{z}) = -\rho^{-1}\int_0^\eta d\eta' \nabla^T[\rho\mathbf{v}(\mathbf{z})]$ is the vertical velocity computed from the horizontal velocity $\mathbf{v}(\mathbf{z})$ by vertically integrating the mass continuity equation of $\nabla^T(\rho\mathbf{v}) + \partial_\eta(\rho w) = 0$ with the boundary condition of $w = 0$ at $\eta = 0$, ρ is the pseudodensity associated with the

vertical coordinate η (e.g., ρ is the air density if height z is used for η, but $\rho = 1$ if pressure p is used for η), $\mathbf{P}_m = (\mathbf{I}, -\rho^{-1}\int_0^\eta d\eta' \nabla_m \rho)^T$ is a 3×2 matrix operator consisting of the 2×2 identity matrix on the top and the 1×2 row vector operator $-\rho^{-1}\int_0^\eta d\eta' \nabla_m^T \rho$ on the bottom. Consequently, the observation operators $\mathbf{L}_{\mathbf{v}}(\mathbf{z})$ in (A4), $\mathbf{L}_{\mathbf{c}}(\mathbf{z})$ in (A13), and $\mathbf{L}_{\mathbf{q}}(\mathbf{z})$ in (A19) are extended (from $M \times 2$ matrices) to $M \times 3$ matrices with their mth row given by $\mathbf{r}_m^T\mathbf{P}_m\mathbf{C}_{\mathbf{vv}}(\mathbf{z}_m, \mathbf{z})$, $\mathbf{r}_m^T\mathbf{P}_m(\nabla_m, \mathbf{k}\times\nabla_m)\mathbf{C}_{\mathbf{cc}}(\mathbf{z}_m, \mathbf{z})$, and $\mathbf{r}_m^T\mathbf{P}_m(\nabla_m, \mathbf{k}\times\nabla_m)\nabla_m^{-2}\mathbf{C}_{\mathbf{qq}}(\mathbf{z}_m, \mathbf{z})$, respectively.

APPENDIX B

Derivations of (17a) and (17b)

Applying the spectral formulations in (3.2) and (3.3) of XG to $C_+(r)$ and $C_-(r)$ in (15) gives

$$S_{ll}(k) + S_{tt}(k) \equiv \int_0^\infty C_+(r)J_0(kr)r\,dr = 2\sigma^2 L_+^2 G(kL_+),$$
$$\text{(B1a)}$$

$$S_{ll}(k) - S_{tt}(k) \equiv -\int_0^\infty C_-(r)J_2(kr)r\,dr$$
$$= -2a\sigma^2 k^2 L_-^4 G(kL_-), \quad \text{(B1b)}$$

where $k = |\mathbf{k}|$, $\mathbf{k} \equiv (k_x, k_y)^T$ is the two-dimensional vector wavenumber associated with $\mathbf{x} \equiv (x, y)^T$, and $J_0(\cdot)$ [or $J_2(\cdot)$] denotes the zero-order (or second-order) Bessel function of the first kind. Recombining (B1a) and (B1b) gives

$$S_{ll}(k) = 2\sigma^2 L_+^2 G(kL_+)[1 - ak^2 L_-^4 L_+^{-2} G(kL_-)/G(kL_+)],$$
$$\text{(B2a)}$$

$$S_{tt}(k) = 2\sigma^2 L_+^2 G(kL_+)[1 + ak^2 L_-^4 L_+^{-2} G(kL_-)/G(kL_+)],$$
$$\text{(B2b)}$$

Here, $S_{ll}(k)$ [or $S_{tt}(k)$] is the power spectrum associated with the autocorrelation function $C_{ll} = (C_+ + C_-)/2$ [or $C_{tt} = (C_+ - C_-)/2$] derived from (15).

For C_{ll} and C_{tt} to be positive semidefinite, $S_{ll}(k)$ and $S_{tt}(k)$ must be nonnegative. This requires $1 \geq f(k^2) \equiv |a|k^2 L_-^4 L_+^{-2} G(kL_-)/G(kL_+)$ in (B2). A necessary condition for this is that $L_- > L_+$, so $k^2 G(kL_-)/G(kL_+) = k^2 \exp[-k^2(L_-^2 - L_+^2)/2] \rightarrow 0$ as $k^2 \rightarrow \infty$. For $L_- > L_+$, one can verify that $f(k^2)$ reaches its maximum of $|a|(2/e)L_-^4/(L_+^2 L_-^2 - L_+^4)$ at the point of $k^2 = 2/(L_-^2 - L_+^2)$ where $df(k^2)/dk^2 = 0$. Thus, $1 \geq f(k^2)$ can be satisfied if and only if $L_- > L_+$ and $|a|(2/e)L_-^4/(L_+^2 L_-^2 - L_+^4) \leq 1$. The latter two conditions yield $2|a|(L_-/L_+)^4 - e(L_-/L_+)^2 + e \leq 0$ or, equivalently,

$$[(L_-/L_+)^2 - R_-][(L_-/L_+)^2 - R_+] \leq 0, \quad \text{(B3)}$$

where $R_{\pm} = 2|a_m/a|[1 \pm (1 - |a/a_m|)^{1/2}]$ are the two roots of $2|a|(L_-/L_+)^4 - e(L_-/L_+)^2 + e = 0$, and $a_m = e/8$ ($\approx 0.333\,98$). Clearly, (B3) can be satisfied if and only if $1 \geq |a|/a_m$ (so $R_{\pm}$ can be real) and $R_- \leq (L_-/L_+)^2 \leq R_+$. This gives the two conditions in (17).

APPENDIX C

Decorrelation Lengths of $C_d(r)$ and $C_r(r)$ in (20)

In the limit of $r \to 0$, we have

$$E(r/L_d) \equiv r^{-1} \int_0^r G(r'/L_d)\, dr' \to G(r/L_d) \to 1, \quad \text{(C1)}$$

$$
\begin{aligned}
r^{-1} \partial_r E(r/L_d) &= [G(r/L_d) - E(r/L_d)]/r^2 \\
&\to [-(r/L_d^2)G(r/L_d) - \partial_r E(r/L_d)]/(2r) \\
&\to -L_d^{-2}/2 - r^{-1}\partial_r E(r/L_d)/2, \quad \text{(C2)}
\end{aligned}
$$

where the L'Hôpital's rule is used. Note that (C2) gives $r^{-1}\partial_r E(r/L_d) = -L_d^{-2}/2 - r^{-1}\partial_r E(r/L_d)/2$ as $r \to 0$ or, equivalently,

$$r^{-1}\partial_r E(r/L_d) \to -L_d^{-2}/3 \quad \text{as} \quad r \to 0. \quad \text{(C3)}$$

Note also that $r^{-1}\partial_r E(r/L_d) \to \partial_r^2 E(r/L_d)$ as $r \to 0$, so

$$
\begin{aligned}
-\nabla^2 E(r/L_d) &= -(\partial_r^2 + r^{-1}\partial_r)E(r/L_d) \\
&\to 2L_d^{-2}/3 \quad \text{as} \quad r \to 0, \quad \text{(C4)}
\end{aligned}
$$

where (C3) is used. It is easy to see that

$$
\begin{aligned}
-\nabla^2 G(r/L_d) &= -(\partial_r^2 + r^{-1}\partial_r)G(r/L_d) \\
&= (2/L_d^2 - r^2/L_d^4)G(r/L_d) \\
&\to 2/L_d^{-2} \quad \text{as} \quad r \to 0. \quad \text{(C5)}
\end{aligned}
$$

Substituting (C1) into (20a) gives

$$C_d(r) = \sigma_d^2[G(r/L_d) + E(r/L_d)] \to 2\sigma_d^2 \quad \text{as} \quad r \to 0. \quad \text{(C6)}$$

Substituting (C4) and (C5) into $-\nabla^2(20a) = -(\partial_r^2 + r^{-1}\partial_r)(20a)$ gives

$$
\begin{aligned}
-\nabla^2 C_d(r) &= -\sigma_d^2(\partial_r^2 + r^{-1}\partial_r)[G(r/L_d) \\
&\quad + E(r/L_d)] \to 8\sigma_d^2 L_d^{-2}/3 \quad \text{as} \quad r \to 0. \quad \text{(C7)}
\end{aligned}
$$

The decorrelation length of the modified C_d in (20) is thus given by

$$[2C_d(r)|_{r=0}]^{1/2}/[-\nabla^2 C_d(r)|_{r=0}]^{1/2} = L_d(3/2)^{1/2} \quad \text{(C8)}$$

according to (C6) and (C7). The decorrelation length of the unmodified C_d in (11a) is $L_\chi/\sqrt{2}$, where L_χ is the decorrelation length of $C_{\chi\chi}$ in (10a). According to (C8), for the modified C_d in (20a) to have the same decorrelation length as the unmodified C_d in (11a), we need to set $L_d = L_\chi/\sqrt{3}$ for the modified C_d derived in (20a). Similarly, $L_r = L_\psi/\sqrt{3}$ for the modified C_r derived in (20b), where L_ψ is the decorrelation length of $C_{\psi\psi}$ in (10b).

APPENDIX D

$C_d(r)$ and $C_r(r)$ Derived from $C_{\delta\delta}(r)$ and $C_{\zeta\zeta}(r)$ Modeled by Gaussian Functions

From (8) we obtain

$$C_d = C_{ll}^d + C_{tt}^d = -(r^{-1}\partial_r + \partial_r^2)C_{\chi\chi},$$

$$C_r = C_{ll}^r + C_{tt}^r = -(r^{-1}\partial_r + \partial_r^2)C_{\psi\psi}.$$

Substituting these into (9) gives

$$(\partial_r^2 + r^{-1}\partial_r)C_d = -C_{\delta\delta}, \quad \text{(D1)}$$

$$(\partial_r^2 + r^{-1}\partial_r)C_r = -C_{\zeta\zeta}. \quad \text{(D2)}$$

For $C_{\delta\delta} = \sigma_\delta^2 G(r/L_\delta)$, C_d can be solved from (D1) with $\partial_r C_d|_{r=0} = 0$ and $C_d|_{r=0} = \sigma_d^2$. The solution is

$$C_d = \sigma_d^2 - \sigma_\delta^2 L_\delta \int_0^{r/L\delta} dr'[1 - G(r')]/r'. \quad \text{(D3)}$$

Note that $1 \geq 1 - G(r) \geq 0$ and $[1 - G(r)]/r \to O(1/r)$ as $r \to \infty$, so the integral in (D3) approaches ∞ (in proportion to $\ln r$) as $r \to \infty$. A rigorous prof of this is given by

$$
\begin{aligned}
&\int_0^\infty dr[1 - G(r)]/r > \int_R^\infty dr[1 - G(r)]/r > \int_R^\infty dr(1 - 1/r^b)/r \\
&\quad = [\ln r + (br^b)^{-1}]\big|_R^\infty = \infty, \quad \text{(D4)}
\end{aligned}
$$

where $G(R) = 1/R^b$, so $G(r) \leq 1/r^b$ for $r \geq R$. Therefore, C_d derived from (D1) approaches $-\infty$ as $r \to \infty$. Similarly, C_r derived from (D2) is also not a classically valid covariance function.

APPENDIX E

Pseudospectral Representations and Generalized Covariance Functions for χ and ψ

By applying the Hankel transformation to (9) with the identity $(\partial_r^2 + r^{-1}\partial_r)J_0(kr) = -k^2 J_0(kr)$ and applying the spectral formulations in (2.11) of XW to $S_{ll}(k)$ and $S_{tt}(k)$ in (B2), we obtain

$$S_{\delta\delta}(k)/k^4 = S_{ll}(k)/k^2 = S_{\chi\chi}(k) = S_{sg}(k) - S_{rg}(k), \quad \text{(E1a)}$$

$$S_{\zeta\zeta}(k)/k^4 = S_{tt}(k)/k^2 = S_{\psi\psi}(k) = S_{sg}(k) + S_{rg}(k). \quad \text{(E1b)}$$

Here, $S_{\delta\delta}(k) \equiv \int_0^\infty C_{\delta\delta}(r) J_0(kr) r\, dr$ and $S_{\zeta\zeta}(k) \equiv \int_0^\infty C_{\zeta\zeta}(r) J_0(kr) r\, dr$ are the power spectra of δ and ζ, respectively, $S_{\chi\chi}(k) \equiv \int_0^\infty C_{\chi\chi}(r) J_0(kr) r\, dr$ and $S_{\psi\psi}(k) \equiv \int_0^\infty C_{\psi\psi}(r) J_0(kr) r\, dr$ are the power spectra of χ and ψ, respectively, formally defined as in (2.7) of XW. Also, $S_{sg}(k) \equiv \sigma^2(L_+/k)^2 G(kL_+)$ is the singular part of $S_{\psi\psi}(k)$ or $S_{\chi\chi}(k)$, and $S_{rg}(k) \equiv \sigma^2 a L_-^4 G(kL_-)$ is the regular part of $S_{\psi\psi}(k)$ or the negative of the regular part of $S_{\chi\chi}(k)$.

Note that $S_{sg}(k) \rightarrow \infty$ in proportion to $1/k^2$ as $k \rightarrow 0$, so $S_{\chi\chi}(k)$ and $S_{\psi\psi}(k)$ are not integrable in the wavenumber space and their inverse Hankel transformations do not converge. This implies that χ and ψ are no longer homogeneous and isotropic and thus cannot have classically valid $C_{\chi\chi}(r)$ and $C_{\psi\psi}(r)$. However, $k^2 S_{\chi\chi}(k) = S_{ll}(k)$ and $k^2 S_{\psi\psi}(k) = S_{tt}(k)$ are integrable, so $S_{\chi\chi}(k)$ and $S_{\psi\psi}(k)$ satisfy the convergence condition for locally homogeneous and isotropic random fields that have homogeneous and isotropic increments or gradients [see (13.34) in Panchev 1971, chapter 3] and thus have valid structure functions [see (13.4) in Panchev 1971, chapter 3]. The structure functions can be derived for χ and ψ and then used to formulate $C_{\chi\chi}(r)$ and $C_{\psi\psi}(r)$ by generalizing the relationship between the covariance and structure functions established for homogeneous and isotropic random fields to locally homogeneous and isotropic random fields. The detailed derivations are given below for ψ.

The random background error field ζ is homogeneous and isotropic, so it can be represented by the following stochastic integral of Fourier–Stieltjes:

$$\zeta(\mathbf{x}) = (2\pi)^{-1/2} \int \exp(i\mathbf{k}^T\mathbf{x})\, dZ_\zeta(\mathbf{k}), \quad \text{(E2)}$$

where $i = \sqrt{-1}$ is the imaginary unit, and $dZ_\zeta(\mathbf{k}) \equiv Z_\zeta(\mathbf{k} + d\mathbf{k}) - Z_\zeta(\mathbf{k})$, where $Z_\zeta(\mathbf{k})$ is a complex random function of $\mathbf{k}$ satisfying the following conditions [see sections 9 and 10 of Yaglom (1962) or section 10 in chapter 2 of Panchev (1971) for one-dimensional cases]:

$$\langle Z_\zeta(\mathbf{k}) \rangle = 0,$$
$$\langle [Z_\zeta(\mathbf{k}+\Delta\mathbf{k}) - Z_\zeta(\mathbf{k})][Z_\zeta(\mathbf{k}' + \Delta\mathbf{k}') - Z_\zeta(\mathbf{k}')]^* \rangle = 0 \quad \text{for any nonoverlapped } \Delta\mathbf{k} \text{ and } \Delta\mathbf{k}',$$
$$\langle |dZ_\zeta(\mathbf{k})|^2 \rangle = S_{\zeta\zeta}(|\mathbf{k}|) d\mathbf{k}, \quad \text{(E3)}$$

where $[\cdot]^*$ denotes the complex conjugate of $[\cdot]$. With $\zeta(\mathbf{x})$ given by (E2) in the infinite domain of $\mathbf{x}$, a pseudospectral representation of $\psi(\mathbf{x})$ can be obtained as a solution of $\nabla^2\psi(\mathbf{x}) = \zeta(\mathbf{x})$ in the following general form:

$$\psi(\mathbf{x}) = (2\pi)^{-1/2} \int |\mathbf{k}|^{-2}[c(\mathbf{k}) - \exp(i\mathbf{k}^T\mathbf{x})]\, dZ_\zeta(\mathbf{k}) + c_0\xi, \quad \text{(E4)}$$

where ξ is a random variable independent of $dZ_\zeta(\mathbf{k})$ with $\langle \xi \rangle = 0$ and $\langle |\xi|^2 \rangle = 1$, and $c(\mathbf{k})$ and c_0 are the constants of integration. The solution of $\nabla^2\psi(\mathbf{x}) = \zeta(\mathbf{x})$ in (E4) can be further generalized by changing $c(\mathbf{k})$ and c_0 to $h(\mathbf{x}, \mathbf{k})$ and $h_0(\mathbf{x})$, respectively, where $h(\mathbf{x}, \mathbf{k})$ and $h_0(\mathbf{x})$ are arbitrary harmonic functions of $\mathbf{x}$ satisfying $\nabla^2 h(\mathbf{x}, \mathbf{k}) = 0$ and $\nabla^2 h_0(\mathbf{x}) = 0$. This generalization, however, is excluded, as it yields no valid structure function.

Unlike the spectral representation of $\zeta(\mathbf{x})$ in (E2), the pseudospectral representation of $\psi(\mathbf{x})$ in (E4) contains nonsinusoidal terms and thus is not purely spectral unless $c(\mathbf{k}) = c_0 = 0$ in (E4). In the latter case, $\psi(\mathbf{x})$ in (E4) reduces to

$$\psi_s(\mathbf{x}) = -(2\pi)^{-1/2} \int \exp(i\mathbf{k}^T\mathbf{x})\, dZ_\zeta(\mathbf{k})/|\mathbf{k}|^2. \quad \text{(E5)}$$

This pure spectral form is not convergent in the mean-square sense because of the singularity of $S_{sg}(k)$ at $k = 0$. This problem can be seen from the unbounded variance of $\psi_s(\mathbf{x})$ derived below,

$$\langle |\psi_s(\mathbf{x})|^2 \rangle = (2\pi)^{-1} \int \langle |dZ_\zeta(\mathbf{k})|^2 \rangle/|\mathbf{k}|^4 = (2\pi)^{-1} \int d\mathbf{k} S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4 = \int_0^\infty k\, dk S_{\zeta\zeta}(k)/k^4 = \int_0^\infty k\, dk [S_{sg}(k) + S_{rg}(k)], \quad \text{(E6)}$$

where $|\exp(i\mathbf{k}^T\mathbf{x})| = 1$ is used in the first step, (E3) is used in the second step, $(2\pi)^{-1} \int d\mathbf{k} = \int_0^\infty k\, dk$ is used for the isotropic integrand $S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4$ in the third step, (E1b) is used in the last step, and $\int d\mathbf{k}$ denotes the integration over the entire two-dimensional wavenumber space of $\mathbf{k}$. It is easy to see that $k S_{sg}(k) \rightarrow \infty$ in proportion to $1/k$ as $k \rightarrow 0$, so the integral in (E6) does not converge. This implies that

the covariance of $\psi_s(\mathbf{x})$ defined by $C_{\psi_s\psi_s}(\mathbf{x}, \mathbf{x}') \equiv \langle\psi_s(\mathbf{x})\psi_s(\mathbf{x}')^*\rangle$ is not convergent as $|\mathbf{x}' - \mathbf{x}| \to 0$. The above problem is also seen for the general pseudospectral representation of $\psi(\mathbf{x})$ in (E4) unless $c(\mathbf{k}) = 1$ as shown later.

On the other hand, the structure function of $\psi(\mathbf{x})$ in (E4) defined by $D_{\psi_s\psi}(\mathbf{r}) \equiv \langle|\psi(\mathbf{x} + \mathbf{r}) - \psi(\mathbf{x})|^2\rangle$ is convergent (clearly to zero) as $|\mathbf{r}| \to 0$, regardless of the values of $c(\mathbf{k})$ and c_0. Substituting $\psi(\mathbf{x})$ in (E4) into the definition of $D_{\psi_s\psi}(\mathbf{r})$ gives

$$D_{\psi_s\psi}(\mathbf{r}) = (2\pi)^{-1}\int|\exp(i\mathbf{k}^T\mathbf{r}) - 1|^2\langle|dZ_\zeta(\mathbf{k})|^2\rangle/|\mathbf{k}|^4 = 2(2\pi)^{-1}\int d\mathbf{k}[1 - \cos(\mathbf{k}^T\mathbf{r})]S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4$$
$$= 2\int_0^\infty k\,dk[1 - J_0(kr)]S_{\zeta\zeta}(k)/k^4 = D_{sg}(r) + D_{rg}(r), \tag{E7}$$

where $D_{sg}(r) \equiv 2\int_0^\infty k\,dk[1 - J_0(kr)]S_{sg}(k)$ and $D_{rg}(r) \equiv 2\int_0^\infty k\,dk[1 - J_0(kr)]S_{rg}(k)$. The derivation of (E7) is similar to that of (E6) but $(2\pi)^{-1}\int d\mathbf{k}\cos(\mathbf{k}^T\mathbf{r}) = \int_0^\infty k\,dkJ_0(kr)$ is used in addition to $(2\pi)^{-1}\int d\mathbf{k} = \int_0^\infty k\,dk$ in the third step. The integral in $D_{rg}(r)$ can be further derived to give

$$D_{rg}(r) = 2\int_0^\infty k\,dk[1 - J_0(kr)]\sigma^2 aL_-^4\,G(kL_-)$$
$$= 2\sigma^2 aL_-^2[1 - G(r/L_-)]. \tag{E8}$$

Note that $1 - J_0(kr) \to 0$ in proportion to k^2 as $k \to 0$, so the singularity of $S_{sg}(k)$ at $k = 0$ is offset completely by $1 - J_0(kr)$ in $D_{sg}(r)$ and the integral in $D_{sg}(r)$ is convergent (to a finite value for a given $r < \infty$). Because of this, $D_{sg}(r)$ can be further derived by applying the Laplace operator $(\partial_r^2 + r^{-1}\partial_r)$ first and then its inverse, denoted by $(\partial_r^2 + r^{-1}\partial_r)^{-1}$, to the integral in $D_{sg}(r)$. This gives

$$D_{sg}(r) = (\partial_r^2 + r^{-1}\partial_r)^{-1}(\partial_r^2 + r^{-1}\partial_r)2\int_0^\infty k\,dk[1 - J_0(kr)]S_{sg}(k) = (\partial_r^2 + r^{-1}\partial_r)^{-1}2\int_0^\infty k\,dkJ_0(kr)\sigma^2 L_+^2\,G(kL_+)$$
$$= (\partial_r^2 + r^{-1}\partial_r)^{-1}2\sigma^2 G(r/L_+) = 2\sigma^2 L_+^2\int_0^{r/L_+} dr'[1 - G(r')]/r', \tag{E9}$$

where the identity $(\partial_r^2 + r^{-1}\partial_r)[1 - J_0(kr)] = k^2J_0(kr)$ is used in the second step, and the boundary conditions $D_{sg}(r) = \partial_r D_{sg}(r) = 0$ at $r = 0$ are used for the inversion of $(\partial_r^2 + r^{-1}\partial_r)$ in the last step.

By generalizing the relationship between the covariance and structure functions established for homogeneous and isotropic random fields [see (13.5) in chapter 3 of Panchev (1971)] to locally homogeneous and isotropic random fields, we obtain the following generalized covariance function for $\psi(\mathbf{x})$:

$$C_{\psi_s\psi}(r) = C_{\psi_s\psi}(0) - D_{\psi_s\psi}(\mathbf{r})/2 = \sigma_\psi^2 - [D_{sg}(r)/2 + D_{rg}(r)/2], \tag{E10}$$

which recovers the result in (23b). Similarly, from (E1a) we can derive the following generalized covariance function for χ:

$$C_{\chi\chi}(r) = \sigma_\chi^2 - D_{\chi\chi}/2 = \sigma_\chi^2 - [D_{sg}(r) - D_{rg}(r)]/2, \tag{E11}$$

which recovers the result in (23a).

Setting $c(\mathbf{k}) = 1$ can make $\psi(\mathbf{x})$ in (E4) convergent. In this case, we have $\psi(0) \equiv \psi(\mathbf{x})|_{|\mathbf{x}|=0} = c_0\xi$ and $\sigma_\psi^2 \equiv \langle|\psi(0)|^2\rangle = c_0^2\langle|\xi|^2\rangle = c_0^2$, and $\psi(\mathbf{x})$ is regularized to

$$\psi(\mathbf{x}) = (2\pi)^{-1/2}\int[1 - \exp(i\mathbf{k}^T\mathbf{x})]\,dZ_\zeta(\mathbf{k})/|\mathbf{k}|^2 + \sigma_\psi\xi. \tag{E12}$$

The variance of $\psi(\mathbf{x})$ is then given by

$$\sigma_\psi(\mathbf{x})^2 \equiv \langle|\psi(\mathbf{x})|^2\rangle$$
$$= \sigma_\psi^2 + (2\pi)^{-1}\int|1 - \exp(i\mathbf{k}^T\mathbf{x})|^2\langle|dZ_\zeta(\mathbf{k})|^2\rangle/|\mathbf{k}|^4$$
$$= \sigma_\psi^2 + 2(2\pi)^{-1}\int d\mathbf{k}[1 - \cos(\mathbf{k}^T\mathbf{x})]S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4$$
$$= \sigma_\psi^2 + 2\int_0^\infty k\,dk[1 - J_0(k|\mathbf{x}|)]S_{\zeta\zeta}(k)/k^4$$
$$= \sigma_\psi^2 + D_{sg}(|\mathbf{x}|) + D_{rg}(|\mathbf{x}|), \tag{E13}$$

where $D_{sg}(|\mathbf{x}|)$ and $D_{rg}(|\mathbf{x}|)$ have the same function forms as those in (E9) and (E8), respectively, except that r is replaced

by $|\mathbf{x}|$. The derivation of (E13) is similar to that in (E7) except that r is replaced by $|\mathbf{x}|$. The covariance of $\psi(\mathbf{x})$ is given by

$$
\begin{aligned}
C_{\psi\psi}(\mathbf{x}, \mathbf{x}') &\equiv \langle \psi(\mathbf{x})\psi(\mathbf{x}')^* \rangle = \sigma_\psi^2 + (2\pi)^{-1}\int d\mathbf{k}[1 - \exp(i\mathbf{k}^T\mathbf{x}) - \exp(-i\mathbf{k}^T\mathbf{x}') + \exp(-i\mathbf{k}^T\mathbf{r})]S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4 \\
&= \sigma_\psi^2 + (2\pi)^{-1}\int d\mathbf{k}\{[1 - \cos(\mathbf{k}^T\mathbf{x})] + [1 - \cos(\mathbf{k}^T\mathbf{x}')] - [1 - \cos(\mathbf{k}^T\mathbf{r})]\}S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4 \\
&= \sigma_\psi^2 + [D_{sg}(|\mathbf{x}|) + D_{rg}(|\mathbf{x}|)]/2 + [D_{sg}(|\mathbf{x}'|) + D_{rg}(|\mathbf{x}'|)]/2 - [D_{sg}(r) + D_{rg}(r)]/2 \\
&= C_{\psi\psi}(r) + [D_{sg}(|\mathbf{x}|) + D_{rg}(|\mathbf{x}|)]/2 + [D_{sg}(|\mathbf{x}'|) + D_{rg}(|\mathbf{x}'|)]/2, \quad\quad\quad\quad\quad\text{(E14)}
\end{aligned}
$$

where $\mathbf{r} = \mathbf{x}' - \mathbf{x}$ and $C_{\psi\psi}(r)$ is given in (E10). The derivation of (E14) is similar to that in (E13) except that $\int d\mathbf{k}\sin(\mathbf{k}^T\mathbf{x})S_{\zeta\zeta}(|\mathbf{k}|)/|\mathbf{k}|^4 = 0$ is used in the second step due to the antisymmetry of $\sin(\mathbf{k}^T\mathbf{x})$ with respect to $|\mathbf{k}| = 0$ for any given $\mathbf{x}$. When $\mathbf{x}' = \mathbf{x}$, $C_{\psi\psi}(\mathbf{x}, \mathbf{x}')$ in (E14) recovers $\sigma_\psi(\mathbf{x})^2$ in (E13). Similarly, we can derive the following covariance function for χ:

$$
\begin{aligned}
C_{\chi\chi}(\mathbf{x}, \mathbf{x}') = C_{\chi\chi}(r) &+ [D_{sg}(|\mathbf{x}|) - D_{rg}(|\mathbf{x}|)]/2 \\
&+ [D_{sg}(|\mathbf{x}'|) - D_{rg}(|\mathbf{x}'|)]/2, \quad\text{(E15)}
\end{aligned}
$$

where $C_{\chi\chi}(r)$ is given in (E11).

Clearly, the covariance function in (E14) [or (E15)] is neither homogeneous nor isotropic, and it consists of three parts: a correlated part given by $C_{\psi\psi}(r)$ [or $C_{\chi\chi}(r)$], and two uncorrelated parts that are functions of $|\mathbf{x}|$ and $|\mathbf{x}'|$, respectively. When these covariance functions are substituted for $\mathbf{C_{cc}} = [C_{\chi\chi}(\mathbf{x}_i, \mathbf{x}_j), C_{\psi\psi}(\mathbf{x}_i, \mathbf{x}_j)]^{\text{diag}}$ in (2), their uncorrelated parts that are independent of $\mathbf{x}_i$ (or $\mathbf{x}_j$) are eliminated by $(\nabla_j, \mathbf{k} \times \nabla_j)$ [or $(\nabla_i, \mathbf{k} \times \nabla_i)$] applied to $\mathbf{C_{cc}}$. Thus, the uncorrelated parts do not affect the analyzed velocity increments from velocity observations. Furthermore, the unbounded structure in $C_{\psi\psi}(\mathbf{x}_i, \mathbf{x}_j)$ [or $C_{\chi\chi}(\mathbf{x}_i, \mathbf{x}_j)$] becomes bounded over the mesoscale range and it should represent the synoptic-scale covariance structures that are localized and thus not fully resolvable on the mesoscale. As the synoptic-scale covariance structures contain no uncorrelated part, the uncorrelated parts in $C_{\psi\psi}(\mathbf{x}_i, \mathbf{x}_j)$ [or $C_{\chi\chi}(\mathbf{x}_i, \mathbf{x}_j)$] must be nonrepresentative and nonphysical. This point is shown explicitly by the one-dimensional example in (G14) and (G15) of appendix G.

APPENDIX F

Validity of (2) for Generalized (χ, ψ) Covariance Functions

When $\mathbf{C_{vv}}$ and $\mathbf{C_{cc}}$ are homogeneous and isotropic as assumed in section 2c, (2) reduces to

$$
\begin{aligned}
\mathbf{R}_\alpha^T[C_{ll}(r), C_{tt}(r)]^{\text{diag}}\mathbf{R}_\alpha \\
= (\nabla_i, \mathbf{k} \times \nabla_i)\{(\nabla_j, \mathbf{k} \times \nabla_j)[C_{\chi\chi}(r), C_{\psi\psi}(r)]^{\text{diag}}\}^T, \quad\text{(F1)}
\end{aligned}
$$

where $\mathbf{C_{cc}} = [C_{\chi\chi}(r), C_{\psi\psi}(r)]^{\text{diag}}$ and $\mathbf{C_{vv}} = \mathbf{R}_\alpha^T[C_{ll}(r), C_{tt}(r)]^{\text{diag}}\mathbf{R}_\alpha$ are used [see (7)]. In this case, as shown in (12), $C_{ll}(r)$ and $C_{tt}(r)$ are related to $C_{\chi\chi}(r)$ and $C_{\psi\psi}(r)$ by

$$
C_{ll} = -\partial_r^2 C_{\chi\chi} - r^{-1}\partial_r C_{\psi\psi}, \quad\text{(F2a)}
$$

$$
C_{tt} = -\partial_r^2 C_{\psi\psi} - r^{-1}\partial_r C_{\chi\chi}. \quad\text{(F2b)}
$$

It is easy to see from the derivations in section 4 that the modified velocity covariance functions [see (15) and (21)] and their derived covariance functions for (χ, ψ) [see (23) and (24)] satisfy (F2). Thus, to show that (2) remains valid for these covariance functions, we need only to prove that (F2) can be derived from (F1).

Note that $r \equiv |\mathbf{r}|$ and $\mathbf{r} \equiv \mathbf{x}_j - \mathbf{x}_i = (x_j - x_i, y_j - y_i)^T$, so

$$
\begin{aligned}
(\nabla_j, \mathbf{k} \times \nabla_j)[C_{\chi\chi}(r), C_{\psi\psi}(r)]^{\text{diag}} \\
= [(\mathbf{x}_j - \mathbf{x}_i)/r, \mathbf{k} \times (\mathbf{x}_j - \mathbf{x}_i)/r]\partial_r[C_{\chi\chi}(r), C_{\psi\psi}(r)]^{\text{diag}}, \\
\text{(F3)}
\end{aligned}
$$

where $\nabla_j a(r) = [(\mathbf{x}_j - \mathbf{x}_i)/r]\partial_r a(r)$ is used and $a(r)$ denotes an arbitrary differentiable function of r.

Substituting (F3) into the right-hand side of (F1) gives

$$
\begin{aligned}
(\nabla_i, &\mathbf{k} \times \nabla_i)\{[\mathbf{x}_j - \mathbf{x}_i, \mathbf{k} \times (\mathbf{x}_j - \mathbf{x}_i)][\partial_r C_{\chi\chi}(r)/r, \partial_r C_{\psi\psi}(r)/r]^{\mathrm{diag}}]\}^{\mathrm{T}} \\
&= \{(\nabla', \mathbf{k} \times \nabla')[(\mathbf{x}_j - \mathbf{x}')\partial_r C_{\chi\chi}(r)/r, \mathbf{k} \times (\mathbf{x}_j - \mathbf{x}')\partial_r C_{\psi\psi}(r)/r]^{\mathrm{T}}\}|_{\mathbf{x}'=\mathbf{x}_i} \\
&\quad + \{(\nabla_i, \mathbf{k} \times \nabla_i)[\partial_r C_{\chi\chi}(r)/r, \partial_r C_{\psi\psi}(r)/r]^{\mathrm{diag}}\}[(\mathbf{x}_j - \mathbf{x}_i), \mathbf{k} \times (\mathbf{x}_j - \mathbf{x}_i)]^{\mathrm{T}} \\
&= -[\partial_r C_{\chi\chi}(r)/r + \partial_r C_{\psi\psi}(r)/r]\mathbf{I} - \mathbf{R}_\alpha^{\mathrm{T}}\{\partial_r[\partial_r C_{\chi\chi}(r)/r, \partial_r C_{\psi\psi}(r)/r]^{\mathrm{diag}}\}(r\mathbf{R}_\alpha) \\
&= \mathbf{R}_\alpha^{\mathrm{T}}[-\partial_r^2 C_{\chi\chi}(r) - \partial_r C_{\psi\psi}(r)/r, -\partial_r^2 C_{\psi\psi}(r) - \partial_r C_{\chi\chi}(r)/r]^{\mathrm{diag}}\mathbf{R}_\alpha. \quad (\mathrm{F4})
\end{aligned}
$$

In the above derivation, the gradient operator ∇' is applied to $\mathbf{x}'$ and then $\mathbf{x}'$ is set to $\mathbf{x}_i$ for the first term in the first step; $[(\mathbf{x}_j - \mathbf{x}_i)/r, \mathbf{k} \times (\mathbf{x}_j - \mathbf{x}_i)/r] = \mathbf{R}_\alpha^{\mathrm{T}}$, $\nabla_i a(r) = -[(\mathbf{x}_j - \mathbf{x}_i)/r]\partial_r a(r)$ and $(\nabla', \mathbf{k} \times \nabla')[(\mathbf{x}_j - \mathbf{x}')a(r), \mathbf{k} \times (\mathbf{x}_j - \mathbf{x}')b(r)]^{\mathrm{T}} = -[a(r) + b(r)]\mathbf{I}$ are used in the second step, where $a(r)$ and $b(r)$ denote arbitrary differentiable functions of r and $\mathbf{I} \equiv (1, 1)^{\mathrm{diag}}$; and $\mathbf{R}_\alpha^{\mathrm{T}}\mathbf{R}_\alpha = \mathbf{I}$ is used in the last step. Substituting (F4) back into the right-hand side of (F1) gives (F2).

APPENDIX G

One-Dimensional Examples of Generalized Covariance Functions

Consider a one-dimensional random background error field $v(x)$ of meridional velocity over the domain of $0 \le x \le L$ that represents a global latitude circle. Assume that this random field is homogeneous and constrained by $\int_0^L v\, dx = 0$ (say, as a result of the global mass conservation), so its discrete power spectrum, denoted by $S_n = S_{vv}(k_n)$, has no zero-wavenumber component and its covariance has the following general form:

$$C_{vv}(x) = \sum_1^N S_n \cos(k_n x), \quad (\mathrm{G1})$$

where $\sum_1^N$ denotes the summation over n from 1 to N, $k_n = n\delta k$, $\delta k = 2\pi/L$ is the global minimum wavenumber, and $k_N = N\delta k = 2\pi/\delta x$ is the maximum wavenumber resolved by the background grid (spaced every $\delta x \equiv L/N$). The spectrum representation of $v(x)$ is then given by

$$v(x) = \sum_1^N [a_n \cos(k_n x) + b_n \cos(k_n x)], \quad (\mathrm{G2})$$

where a_n and b_n are random variables satisfying $\langle a_n \rangle = \langle b_n \rangle = \langle a_n b_n \rangle = 0$, $\langle a_n b_{n'} \rangle = 0$ for $n \ne n'$, and $\langle a_n^2 \rangle = \langle b_n^2 \rangle = S_n/2$ [see (2.29) of Yaglom (1962) or (9.1) in chapter 2 of Panchev (1971)]. It is easy to see that $C_{vv}(x) \equiv \langle v(x' - x)v(x') \rangle$ derived from $v(x)$ in (G2) recovers (G1).

When $C_{vv}(x)$ is estimated from coarse-resolution observation innovations (such as radiosonde innovations sampled over one season), its spectral representation in (G1) is coarsened and truncated to the following form:

$$C_{vv}(x) = \sum_1^M S_m \cos(k_m x), \quad (\mathrm{G3})$$

where $\sum_1^M$ denotes the summation over m from 1 to M ($\ll N$), $k_m = m\Delta k$, $\Delta k = 2\pi/l$ is the minimum wavenumber, l ($<$ or $\ll L$) is the synoptic-scale range covered by the coarse-resolution innovations, and $k_M = M\Delta k = 2\pi/\Delta x$ is the maximum wavenumber resolved by the coarse-resolution innovations (spaced about every $\Delta x \equiv l/M$). Note that S_m in (G3) should estimate and represent the sum of S_n for $(m - \frac{1}{2})\Delta k < k_n \le (m + \frac{1}{2})\Delta k$ or, equivalently, $(m - \frac{1}{2})L/l < n \le (m + \frac{1}{2})L/l$ in (G1). However, because k_M ($=2\pi/\Delta x$) in (G3) is much smaller than k_N ($=2\pi/\delta x$) in (G1) and the high-wavenumber components with $k_n > 2\pi/\Delta x$ in (G1) are aliased into the low-wavenumber components sampled and estimated by S_m in (G3), the estimated sum (by S_m) can become larger than the true sum and this leads to

$$S_m > \sum_{n(m)} S_n, \quad (\mathrm{G4})$$

where $\sum_{n(m)}$ denotes the summation over n for $(m - \frac{1}{2})L/l < n \le (m + \frac{1}{2})L/l$. Strictly speaking, a zero-wavenumber component S_0 should be added to the spectral representation in (G3) to estimate and represent the sum of S_n for $1 \le n \le \Delta k/(2\delta k) = L/(2l)$ in (G1). However, since the power spectrum S_n for $C_{vv}(x)$ in (G1) is peaked around wavenumbers on the synoptic scale and so is the estimated S_m in (G3), the sum of S_n

estimated by S_0, if added in (G3), is very small and thus can be neglected [see (4.4), Figs. 4 and 6 of XW].

In consistence with $C_{vv}(x)$ in (G3), the spectrum representation of $v(x)$ in (G2) is also coarsened and truncated to

$$v(x) = \sum_1^M [a_m \cos(k_m x) + b_m \sin(k_m x)], \quad (G5)$$

where a_m and b_m are random variables similar to those in (G2) but associated with S_m in (G3). The spectrum representation for the streamfunction ψ, defined by $\partial_x \psi = v$, can be derived by integrating $v(x)$ in (G5), which gives

$$\psi(x) \sum_1^M [a_m \sin(k_m x) - b_m \cos(k_m x)]/k_m. \quad (G6)$$

Substituting (G6) into $C_{\psi\psi}(x) \equiv \langle \psi(x' - x)\psi(x')\rangle$ gives

$$C_{\psi\psi}(x) = \sum_1^M S_m \cos(k_m x)/k_m^2. \quad (G7)$$

Note that $\partial_x^2 C_{\psi\psi}(x) = -C_{vv}(x)$ can be derived by substituting $\partial_x \psi = v$ into $C_{vv}(x) \equiv \langle v(x' - x)v(x')\rangle$ as those in section 2a, and this relationship is satisfied by $C_{\psi\psi}(x)$ in (G7) with $C_{vv}(x)$ in (G3).

When $C_{vv}(x)$ is estimated from high-resolution observation innovations (such as radar radial velocity innovations sampled over a mesoscale time window up to 1/2 day), the spectral representation in (G3) is further coarsened to

$$C_{vv}(x) = \sum_0^{M'} S'_m \cos(k'_m x), \quad (G8)$$

where $\sum_0^{M'}$ denotes the summation over m' from 0 to M' $(\ll N)$, $k'_m = m'\Delta k'$, $\Delta k' = 2\pi/l'$ is the minimum wavenumber, l' $(\ll l)$ is the mesoscale range covered by the high-resolution innovations, and $k'_M = M'\Delta k' = 2\pi/\delta x'$ is the maximum wavenumber resolved by the high-resolution innovations (spaced about every $\delta x' = l'/M'$). Note that S'_0 retained by $C_{vv}(x)$ in (G8) estimates and represents the sum of S_n for $1 \le n \le \Delta k'/(2\delta k) = L/(2l')$ in (G1) and this sum, according to (G4), is overestimated by the sum of S_m for $1 \le m \le \Delta k'/(2\Delta k) = l/(2l')$ in (G3), so

$$S'_0 \approx \sum_1^{n_0} S_n < S_{m_0} \equiv \sum_1^{m_0} S_m, \quad (G9)$$

where $\sum_1^{n_0}$ denotes the summation over n for $1 \le n \le n_0$, $\sum_1^{m_0}$ denotes the summation over m for $1 \le m \le m_0$,

$n_0 \equiv \text{int}[L/(2l')] \gg m_0 \equiv \text{int}[l/(2l')] \gg 1$, and $\text{int}[(\cdot)]$ denotes the integer part of $(\cdot)$.

In consistence with $C_{vv}(x)$ in (G8), the spectrum representation of $v(x)$ in (G2) is also further coarsened to

$$v(x) = \sum_0^{M'} [a'_m \cos(k'_m x) + b'_m \sin(k' x)], \quad (G10)$$

where a'_m and b'_m are random variables similar to those in (G2) but associated with S'_m in (G8). Integrating (G10) gives

$$\psi(x) = c_0\xi_0 + a'_0 x + \sum_1^{M'} [a'_m \sin(k'_m x) - b'_m \cos(k'_m x)]/k'_m, \quad (G11)$$

where c_0 is a constant of integration to be specified [see (G14)], ξ_0 is a random variable independent of a'_m and b'_m and satisfying $\langle \xi_0 \rangle = 0$ and $\langle \xi_0^2 \rangle = 1$. Clearly, $\psi(x)$ in (G11) is no longer homogeneous and thus has no classically valid homogeneous covariance, but it is locally homogeneous (with homogeneous increments or derivative) and thus has a valid structure function defined by $D_{\psi\psi}(x) \equiv \langle [\psi(x' + x) - \psi(x')]^2 \rangle$ [see (3.40) of Yaglom (1962) or (11.6) in chapter 2 of Panchev (1971)]. Substituting $\psi(x)$ in (G11) into this definition gives

$$D_{\psi\psi}(x) = S'_0 x^2 + 2\sum_1^{M'} S'_m [1 - \cos(k'_m x)]/k'^2_m$$

$$= S'_0 x^2 + 2\sigma'^2_\psi - 2\sum_1^{M'} S'_m \cos(k'_m x)/k'^2_m, \quad (G12)$$

where $\sigma'^2_\psi \equiv \sum_1^{M'} S'_m/k'^2_m$ is the mesoscale variance resolved by the sinusoidal terms of $\psi(x)$ in (G11). By generalizing the relationship between the covariance and structure functions established for homogeneous random fields [see (3.42) of Yaglom (1962) or (11.9) in chapter 2 of Panchev (1971)] to locally homogeneous random fields, we obtain the following generalized covariance function for $\psi(\mathbf{x})$ in (G11):

$$C_{\psi\psi}(x) = C_{\psi\psi}(0) - D_{\psi\psi}(x)/2 = \Delta\sigma^2_\psi - S'_0 x^2/2$$

$$+ \sum_1^{M'} S'_m \cos(k'_m x)/k'^2_m, \quad (G13)$$

where $\Delta\sigma^2_\psi = \sigma^2_\psi - \sigma'^2_\psi$ is the residual variance not resolved by the sinusoidal terms of $\psi(x)$ in (G11), and $\sigma^2_\psi \equiv C_{\psi\psi}(0)$ is the total variance of $\psi(x)$.

Clearly, $C_{\psi\psi}(x)$ in (G13) satisfies $\partial_x^2 C_{\psi\psi}(x) = -C_{vv}(x)$ for $C_{vv}(x)$ in (G8). This relationship or, equivalently, $\partial_{x''}\partial_{x'}C_{\psi\psi}(x'' - x') = C_{vv}(x'' - x')$ is also satisfied by the following covariance function derived directly from the pseudospectrum representation of $\psi(x)$ in (G11):

$$C_{\psi\eta\psi}(x', x'') \equiv \langle \psi(x')\psi(x'') \rangle = c_0^2 + S_0'(x'^2 + x''^2 - x^2)/2$$

$$+ \sum_1^{M'} S_m' \cos(k_m'\chi)]/k_m'^2$$

$$= C_{\psi\eta\psi}(x) + S_0'x'^2/2 + S_0'x''^2/2, \quad \text{(G14)}$$

where $x = x'' - x'$, the identity of $x'x'' = (x'^2 + x''^2 - x^2)/2$ is used in the first step, and $C_{\psi\eta\psi}(x)$ in (G13) is used by

specifying $c_0^2 \equiv \Delta\sigma_\psi^2$ in the last step. The covariance function in (G14) consists of three parts: a homogeneously correlated part given by $C_{\psi\eta\psi}(x)$, and two uncorrelated parts that are functions of x' and x'', respectively. The two uncorrelated parts are contained in the nonsinusoidal part of $C_{\psi\eta\psi}(x', x'')$, that is, $\Delta\sigma_\psi^2 + S_0'(x'^2 + x''^2 - x^2)/2$. According to (G9), this nonsinusoidal part should represent $\sum_1^{m_0} S_m \cos(k_m x)/k_m^2$ and the latter can be estimated by

$$C_{0\psi\eta\psi}(x) \equiv (S_0'/S_{m0}) \sum_1^{m_0} S_m \cos(k_m x)/k_m^2 = (S_0'/S_{m0}) \sum_1^{m_0} S_m (1/k_m^2 - x^2/2! + x^4 k_m^2/4! - x^6 k_m^4/6! + \cdots) \approx S_0'(\alpha - x^2/2),$$

$$\text{(G15)}$$

where the Taylor expansion of $\cos(k_m x)$ is used in the first step, and $S_{m_0} \equiv \sum_1^{m_0} S_m$ and $a \equiv (\sum_1^{m_0} S_m/k_m^2)/S_{m_0}$ are used in the second step after the expansion is truncated. The truncation is valid for $|x| \le l'/2$ because $1 \le m \le m_0 \equiv \text{int}[l/(2l')]$ and thus $|k_m x| \le \pi/2$ for all $\cos(k_m x)$ terms in the summation denoted by $\sum_1^{m_0}$. By setting $\Delta\sigma_\psi^2 = \alpha S_0'$, the truncated expansion in the last step of (G15) gives the nonsinusoidal part of $C_{\psi\eta\psi}(x)$ in (G13) rather than $C_{\psi\eta\psi}(x', x'')$ in (G14). This implies the followings: (i) $\Delta\sigma_\psi^2$ can be set to aS_0' with a and S_0' estimated from the coarse-resolution innovations although $\Delta\sigma_\psi^2$ cannot be estimated by (G8) from the high-resolution innovations; (ii) $S_0'(x'^2 + x''^2)$ in $C_{\psi\eta\psi}(x', x'')$ is nonphysical, as it can neither represent nor approximate any part of $C_{0\psi\eta\psi}(x)$ in (G15); (iii) $\Delta\sigma_\psi^2 - S_0'x^2/2$ in $C_{\psi\eta\psi}(x)$ is a valid approximation of $C_{0\psi\eta\psi}(x)$ for $|x| \le l'/2$ in the mesoscale range; and (iv) $C_{0\psi\eta\psi}(x)$ can replace $\Delta\sigma_\psi^2 - S_0'x^2/2$ in $C_{\psi\eta\psi}(x)$ with S_m estimated from the coarse-resolution innovations, so $C_{\psi\eta\psi}(x)$ in (G13) is modified and expanded to the synoptic scale into the following form:

$$C_{\psi\eta\psi}(x) = C_{0\psi\eta\psi}(x) + \sum_1^{M'} S_m' \cos(k_m'x)]/k_m'^2. \quad \text{(G16)}$$

As the high-resolution innovations are sampled over a mesoscale time window (≤ 0.5 days), the true synoptic-scale innovation variance is partially attributed to the innovation mean and thus not fully detectable. This implies that the true synoptic-scale background error variance that should be represented by S_0' is only partially extracted by S_0' from the partially detected innovation variance, and this leads to $S_0' < \sum_1^{n_0} S_n$ instead of $S_0' \approx \sum_1^{n_0} S_n$ in (G9). Thus, when $C_{vv}(x)$ in (G8) or $C_{\psi\eta\psi}(x)$ in (G13) or (G16) is used for mesoscale data assimilation, S_0' (estimated from the high-resolution innovations) may need to be adjusted properly toward (but not to) its upper bound S_{m_0} (estimated from the coarse-resolution innovations) in (G9). Such an

adjustment may become unnecessary if the mesoscale data assimilation is performed in the second step after the background errors are reduced on the synoptic scale by analyzing coarse-resolution observations in the first step via a multistep and multiscale approach.

The above one-dimensional results can be extended to the two-dimensional space for real-data assimilation, in which case the truncated one-dimensional spectral expansions formulated above are replaced by their respective two-dimensional versions (similar to those in XW, X07a, and X07b). The detailed formulations are omitted here.

REFERENCES

Adams, R. A., 1975: *Sobolev Spaces.* Academic Press, 268 pp.

Bennett, A. F., 1992: *Inverse Method in Physical Oceanography.* Cambridge University Press, 346 pp.

Courant, R., and D. Hilbert, 1962: *Partial Differential Equations.* Vol. II, *Methods of Mathematical Physics*, Interscience, 830 pp.

Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123**, 2449–2461, https://doi.org/10.1002/qj.49712354414.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data assimilation method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469, https://doi.org/10.1175/1520-0426(2004)021<0457: ATVDAM>2.0.CO;2.

——, and Coauthors, 2013: A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Wea. Forecasting*, **28**, 727–745, https://doi.org/10.1175/WAF-D-12-00093.1.

Hollingsworth, A., and P. Lonnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136, https://doi.org/10.1111/j.1600-0870.1986.tb00460.x.

Li, Z., J. C. McWilliams, K. Ide, and J. D. Farrara, 2015: A multiscale variational data assimilation scheme: Formulation and illustration. *Mon. Wea. Rev.*, **143**, 3804–3822, https://doi.org/10.1175/MWR-D-14-00384.1.

Lighthill, M. J., 1958: *Introduction to Fourier Analysis and Generalized Functions.* Cambridge University Press, 79 pp.

Lorenc, A. C., 2003: Modeling of error covarinces by 4D-Var data assimilation. *Quart. J. Roy. Meteor. Soc.*, **129**, 3167–3182, https://doi.org/10.1256/qj.02.131.

——, and Coauthors, 2000: The Met Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012, https://doi.org/10.1002/qj.49712657002.

Panchev, S., 1971: *Random Functions and Turbulence.* Pergamon Press, 444 pp.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524–1535, https://doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.

——, ——, ——, and ——, 2003b: 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, https://doi.org/10.1175//2543.1.

Sun, J., H. Wang, W. Tong, Y. Zhang, C.-Y. Lin, and D. Xu, 2016: Comparison of the impacts of momentum control variables on high-resolution variational data assimilation and precipitation forecasting. *Mon. Wea. Rev.*, **144**, 149–169, https://doi.org/10.1175/MWR-D-14-00205.1.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Xie, Y., and A. E. MacDonald, 2012: Selection of momentum variables for a three-dimensional variational analysis. *Pure Appl. Geophys.*, **169**, 335–351, https://doi.org/10.1007/s00024-011-0374-3.

——, S. Koch, J. McGinley, S. Albers, P. E. Bieringer, M. Wolfson, and M. Chan, 2011: A space–time multiscale analysis system: A sequential variational analysis approach. *Mon. Wea. Rev.*, **139**, 1224–1240, https://doi.org/10.1175/2010MWR3338.1.

Xu, Q., 2005: Representations of inverse covariances by differential operators. *Adv. Atmos. Sci.*, **22**, 181–198, https://doi.org/10.1007/BF02918508.

——, and L. Wei, 2001: Estimation of three-dimensional error covariances. Part II: Analysis of wind innovation vectors. *Mon. Wea. Rev.*, **129**, 2939–2954, https://doi.org/10.1175/1520-0493(2001)129<2939:EOTDEC>2.0.CO;2.

——, and J. Gong, 2003: Background error covariance functions for Doppler radial-wind analysis. *Quart. J. Roy. Meteor. Soc.*, **129**, 1703–1720, https://doi.org/10.1256/qj.02.129.

——, K. Nai, and L. Wei, 2007a: An innovation method for estimating radar radial-velocity observation error and background wind error covariances. *Quart. J. Roy. Meteor. Soc.*, **133**, 407–415, https://doi.org/10.1002/qj.21.

——, ——, ——, H. Lu, P. Zhang, S. Liu, and D. Parrish, 2007b: Estimating radar wind observation error and NCEP WRF background wind error covariances from radar radial-velocity innovations. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction.* Park City, UT, Amer. Meteor. Soc., 1B.3, https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_123419.htm.

——, J. Cao, and S. Gao, 2011: Computing streamfunction and velocity potential in a limited domain. Part I: Theory and integral formulae. *Adv. Atmos. Sci.*, **28**, 1433–1444, https://doi.org/10.1007/s00376-011-0185-6.

——, L. Wei, K. Nai, S. Liu, R. M. Rabin, and Q. Zhao, 2015: A radar wind analysis system for nowcast applications. *Adv. Meteor.*, **2015**, 264515, https://doi.org/10.1155/2015/264515.

——, ——, J. Gao, Q. Zhao, K. Nai, and S. Liu, 2016: Multistep variational data assimilation: Important issues and a spectral approach. *Tellus*, **68A**, 31110, https://doi.org/10.3402/tellusa.v68.31110.

——, ——, and K. Nai, 2017: Generalized covariance functions for mesoscale wind analysis and data assimilation. *17th Conf. on Mesoscale Processes*, San Diego, CA, Amer. Meteor. Soc., 8.7, https://ams.confex.com/ams/17MESO/webprogram/Paper320236.html.

Yaglom, A. M., 1962: *An Introduction to the Theory of Stationary Random Functions.* Dover Publications, 235 pp.