Optimized Localization and Hybridization to Filter Ensemble-Based Covariances

BENJAMIN MÉNÉTRIER AND THOMAS AULIGNÉ

Mesoscale and Microscale Meteorology Laboratory, National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 10 February 2015, in final form 28 May 2015)

ABSTRACT

Localization and hybridization are two methods used in ensemble data assimilation to improve the accuracy of sample covariances. It is shown in this paper that it is beneficial to consider them jointly in the framework of linear filtering of sample covariances. Following previous work on localization, an objective method is provided to optimize both localization and hybridization coefficients simultaneously. Theoretical and experimental evidence shows that if optimal weights are used, localized-hybridized sample covariances are always more accurate than their localized-only counterparts, whatever the static covariance matrix specified for the hybridization. Experimental results obtained using a 1000-member ensemble as a reference show that the method developed in this paper can efficiently provide localization and hybridization coefficients consistent with the variable, vertical level, and ensemble size. Spatially heterogeneous optimization is shown to improve the accuracy of the filtered covariances, and consideration of both vertical and horizontal covariances is proven to have an impact on the hybridization coefficients.

1. Introduction

The formulation of accurate background error covariances is critical for most data assimilation (DA) schemes used in numerical weather prediction (NWP). Two main methods can be used in practice to define such covariances: parameterized or ensemble based. In the former, a model of covariance matrix is built, whose parameters can be tuned using different datasets. A limited basis of functions is defined to express the matrix, which is not fully flow dependent. On the other hand, the ensemble-based method uses data from an updated ensemble that samples the fully flow-dependent background error distribution. However, because of its computational cost, the ensemble size is limited, which generates sampling noise. Several techniques can be used to filter this noise. The most popular is the covariance localization, which relies on a Schur (i.e., element by element) product of the sample covariance matrix with a correlation matrix. In the hybrid formalism, the two covariance models can be combined linearly to take advantage of both parameterized and ensemble-based methods.

Since the early work of Hamill and Snyder (2000), hybridization is getting more and more popular, especially with the founding work of Lorenc (2003) and Buehner (2005) allowing its implementation in variational schemes [ensemble variational (EnVar)]. Hybridization tries to get the best of both cited methods by combining the robustness of a parameterized covariance matrix with the flow dependence of an ensemble-based one. Recently, the work of Bishop and Satterfield (2013) and Bishop et al. (2013) about hidden variance has provided a theoretical justification about the need for hybridization and also a method to estimate parameters of an assumed distribution of hidden variance. However, this method has been developed and tried on a low-order model only. To the authors’ knowledge, existing studies about hybridization in high-dimensional systems have performed empirical sensitivity tests to determine the hybridization coefficients. Because of the huge cost of such an operation, the number of tested combinations has been generally limited to \(O(10)\). However, several operational configurations are now using hybrid background error covariance matrices (Wang et al. 2008a,b; Clayton et al. 2013; Buehner et al. 2013; Kuhl et al. 2013; Wang et al. 2013; Gustafsson et al. 2014; Kleist and Ide 2015a,b; Lorenc et al. 2015; Buehner et al. 2015).

In this paper, we show that localization and hybridization can be considered together in the more general framework of the linear filtering of sample covariances.
In the two-part paper Ménétretier et al. (2015a,b, hereafter M15a,b), two theories were merged to obtain optimality criteria for the filtering of sample covariances: the theory of optimal linear filtering and the theory of centered moments sampling. Two practical applications were then detailed: variance filtering and covariance localization. We now extend this formalism of sample covariance filtering to include the hybridization with a parameterized covariance matrix, hereafter called “static covariance matrix.” We show that the localization can be optimized simultaneously with the hybridization coefficients.

An important result is the demonstration that this joint optimization provides a localized-hybridized covariance matrix that is systematically more accurate than its localized-only counterpart, whatever the static covariance matrix that is systematically more accurate than its optimality provides a localized-hybridized covariance matrix.

We assume that an ensemble of \( N \) forecasts, denoted \( \{ \mathbf{x}_p^b \}_{p=1}^N \), is provided to sample the forecast error distribution. The ensemble perturbations \( \delta \mathbf{x}_p^b \) are defined by

\[
\delta \mathbf{x}_p^b = \mathbf{x}_p^b - \langle \mathbf{x}^b \rangle ,
\]

where the index \( p \) refers to an ensemble member and the angle brackets \( \langle \cdot \rangle \) denote the sample mean:

\[
\langle \mathbf{x}^b \rangle = \frac{1}{N} \sum_{p=1}^{N} \mathbf{x}_p^b .
\]

Estimations of the centered moments of Eqs. (2a) and (2b), denoted \( \bar{B} \) and \( \bar{\Xi} \), respectively, are computed from these perturbations by

\[
\bar{B}_{ij} = \frac{1}{N-1} \sum_{p=1}^{N} \delta x_{ip}^b \delta x_{jp}^b ,
\]

\[
\bar{\Xi}_{ijkl} = \frac{1}{N-1} \sum_{p=1}^{N} \delta x_{ip}^b \delta x_{jp}^b \delta x_{kp}^b \delta x_{lp}^b .
\]

If the ensemble size goes to infinity, the sample covariance matrix \( \bar{B} \) converges to its asymptotic value \( \mathbf{B}^* \):

\[
\mathbf{B}^* = \lim_{N \to \infty} \bar{B} .
\]

2) OPTIMAL COVARIANCE LOCALIZATION

To remove most of the sampling noise \( \bar{B}^* = \bar{B} - \mathbf{B}^* \) present in the sample covariance matrix \( \bar{B} \), a localization can be performed through a Schur product (i.e., element-by-element):

\[
\bar{B} = \mathbf{L} \bullet \bar{B} \iff \bar{B}_{ij} = L_{ij} \bar{B}_{ij} ,
\]

where \( \mathbf{L} \) denotes the localization and \( \bar{B} \) is the localized covariance matrix. The objective determination of the localization has been studied by various authors (Anderson 2007, 2012; Anderson and Lei 2013; Bishop and Hodyss 2007, 2009a,b, 2011). In M15a,b, these methods have been compared to a new framework proposed under the following assumptions:

- the sampling error \( \bar{B}^* \) is unbiased and orthogonal to the asymptotic covariance matrix \( \mathbf{B}^* \) (which amounts to a zero correlation assumption);
- the random processes generating the asymptotic covariance matrix \( \mathbf{B}^* \) and the sample distribution are independent (see M15a,b for details).

It has been shown that the localization \( \mathbf{L} \) minimizing the expected quadratic error

\[
e = \mathbb{E}[\| \bar{B} - \mathbf{B}^* \|^2] ,
\]
where a Frobenius matrix norm is used, is given by

\[
L_{ij} = \frac{(N - 1)^2}{N(N - 3)} \frac{N}{(N - 2)(N - 3)} \frac{\mathbb{E}[\tilde{v}]}{\mathbb{E}[B_{ij}^2]}
\]

\[
+ \frac{N - 1}{N(N - 2)(N - 3)} \frac{\mathbb{E}[\tilde{v} \dot{v}]}{\mathbb{E}[B_{ij}^2]},
\]

(9)

where \( \tilde{v} \) is the vector containing the sample variances: \( \tilde{v}_i = \tilde{B}_{ii} \). If the ensemble distribution is Gaussian, then Eq. (9) can be simplified to

\[
L^G_{ij} = \frac{N - 1}{(N + 1)(N - 2)} \left( N - 1 - \frac{\mathbb{E}[\tilde{v} \dot{v}]}{\mathbb{E}[B_{ij}^2]} \right),
\]

(10)

Denoting \( \tilde{C} \) the sample correlation matrix, the additional assumption that Cov(\( \tilde{v} \), \( \tilde{C}_v \)) \( \ll \mathbb{E}[\tilde{v} \dot{v}]\mathbb{E}[\tilde{C}_v^2] \) leads to a further simplification of Eq. (10):

\[
L^{GC}_{ij} = \frac{N - 1}{(N + 1)(N - 2)} \left( N - 1 - \frac{1}{\mathbb{E}[C_{ij}^2]} \right),
\]

(11)

which involves sample correlations only.

b. Hybridization with static covariances

1) SCALAR HYBRIDIZATION

With the previous formulation of the localization, a square root of the localized covariance matrix \( \tilde{B} \) is easily available (Lorenc 2003; Buehner 2005):

\[
\delta x = (L \cdot \tilde{B})^{1/2} \nu^a = \frac{1}{\sqrt{N - 1}} \sum_{p=1}^N \delta x_p^h \cdot (L^{1/2} \nu_p^a),
\]

(12)

where \( \nu_p^a \) is the segment of the control variable \( \nu^a \) corresponding to the ensemble perturbation \( \delta x_p^h \). The hybridization of this ensemble-based increment with a term \( \tilde{B}^{1/2} \nu^c \), where \( \tilde{B} \) is the static covariance matrix, lets the increment explore directions that are not spanned by the ensemble members. Thus, in most of recent publications, the hybrid increment \( \delta x^h \) (the superscript \( h \) denoting the hybrid case) is given by

\[
\delta x^h = \beta^e (I^h \cdot \tilde{B})^{1/2} \nu^a + \beta^c \tilde{B}^{1/2} \nu^c,
\]

(13)

where \( I^h \) is a normalized localization, whose diagonal coefficients are equal to one. The main limitation of using scalar weighting coefficients \( \beta^e \) and \( \beta^c \) is that they do not allow for any formal dependence regarding the location, as done in Buehner et al. (2013), where they are varying with the vertical level.

2) VECTOR HYBRIDIZATION

This limitation can be overcome by using vectors \( \beta^e \) and \( \beta^c \) of size \( n \) instead of scalar coefficients:

\[
\delta x^h = \beta^e \cdot (I^h \cdot \tilde{B})^{1/2} \nu^a + \beta^c \cdot \tilde{B}^{1/2} \nu^c.
\]

(14)

Extending the demonstration of Wang et al. (2007) to the case where \( \beta^e \) and \( \beta^c \) are vectors, it is straightforward to show that the localized-hybridized covariance matrix \( \tilde{B}^h \) implied by Eq. (14) is given by

\[
\tilde{B}^h = L^h \cdot \tilde{B} + \tilde{B}^{tot} \Leftrightarrow \tilde{B}_{ij}^h = L_{ij}^h \tilde{B}_{ij} + \tilde{B}_{ij}^{tot},
\]

(15)

where

- \( L^h \) is the effective localization:

\[
L^h = [\beta^e (\beta^e)^T] \cdot \tilde{B} \Leftrightarrow L_{ij}^h = \beta_{ij}^e \beta_{ij}^e \tilde{B}_{ij},
\]

(16)

- \( \tilde{B}^{tot} \) is the total static covariance matrix:

\[
\tilde{B}^{tot} = [\beta^e (\beta^e)^T] \cdot \tilde{B} \Leftrightarrow \tilde{B}_{ij}^{tot} = \beta_{ij}^e \beta_{ij}^e \tilde{B}_{ij},
\]

(17)

c. Joint optimization of localization and hybridization coefficients

The aim of this subsection is to optimize the localization and the hybridization coefficients simultaneously, so that the expected quadratic error on the localized-hybridized matrix \( \tilde{B}^h \),

\[
e^h = \mathbb{E}[||\tilde{B}^h - \tilde{B}^*||^2],
\]

(18)

is minimized.

1) OPTIMIZATION OF THE LOCALIZATION

Similarly to Eq. (28) of M15a,b, the optimality condition on coefficients of the localization \( L^h \) is given by

\[
\frac{\partial e^h}{\partial L_{ij}^h} = 0 \Leftrightarrow \mathbb{E}[(\tilde{B}_{ij}^h - \tilde{B}_{ij}^{tot}) \tilde{B}_{ij}] = 0.
\]

(19)

This condition is exactly the same for localizations \( L \) and \( L^h \), respectively, optimized alone or simultaneously with the static covariance matrix. An interesting consequence is that Eq. (30c) of M15a,b is valid in both cases:

\[
\mathbb{E}[\tilde{B}_{ij} \tilde{B}_{ij}] = \mathbb{E}[\tilde{B}_{ij}^{tot} \tilde{B}_{ij}] = \mathbb{E}[(\tilde{B}_{ij}^*)^2].
\]

(20)
We can develop \( \hat{B}_{ij} \) and \( \hat{B}_{ij}^h \) in the previous equation to get the following:

\[
\mathbb{E}[\hat{B}_{ij}(L_{ij}\hat{B}_{ij})] = \mathbb{E}[\hat{B}_{ij}(L_{ij}^{h}\hat{B}_{ij} + \hat{B}_{ij}^{\text{tot}})]
\]

\[
\Rightarrow L_{ij}^{h} = \frac{\mathbb{E}[B_{ij}]}{\mathbb{E}[B_{ij}]} \hat{B}_{ij}^{\text{tot}} = L_{ij}.
\] (21)

Thus, the localization \( L_{ij}^{h} \) can be expressed via the localization \( L \), and a second contribution involving the total static covariance matrix \( \bar{B} \).

### 3. Properties of optimal localization and hybridization coefficients

#### a. System formulation and approximations

1) **GENERAL NONLINEAR SYSTEM**

The system constituted by Eqs. (21) and (22) is nonlinear, and can be expressed as \( \mathcal{A}(L, \gamma) = 0 \), with

\[
\mathcal{A}_{L_{ij}}(L^{h}, \gamma) = L_{ij}^{h} + \frac{\mathbb{E}[\hat{B}_{ij}]}{\mathbb{E}[\hat{B}_{ij}]} \beta^{c}_{i}(\gamma) \beta^{c}_{j}(\gamma) \hat{B}_{ij} - L_{ij}.
\] (23a)

\[
\mathcal{A}_{\gamma_{m}}(L^{h}, \gamma) = \sum_{ij} \beta^{c}_{i}(\gamma) \beta^{c}_{j}(\gamma) \hat{B}_{ij} - \gamma_{m}(L_{ij} - 1)\mathbb{E}[\hat{B}_{ij}]
\]

\[
+ \beta^{c}_{i}(\gamma) \beta^{c}_{j}(\gamma) \hat{B}_{ij}.
\] (23b)

where \( \mathcal{A}_{L_{ij}} \) and \( \mathcal{A}_{\gamma_{m}} \) are the respective parts of \( \mathcal{A} \) concerning the respective optimizations of \( L_{ij}^{h} \) and \( \gamma_{m} \).

2) **LEVEL-DEPENDENT NONLINEAR SYSTEM**

The system in Eq. (23) can be simplified for usage-specific particular cases. For instance, we are interested, as in Buehner et al. (2013), in a coefficient \( \beta^{c}_{i} \) that is varying with the vertical level only, not horizontally. In this case, the number of parameters \( M \) in \( \gamma \) is, therefore, equal to the number of vertical levels. Defining the function \( \text{lev}(i) \) that gives the vertical level of the point of index \( i \), we have

\[
\beta^{c}_{i}(\gamma) = \gamma_{\text{lev}(i)}.
\] (24)

2) **OPTIMIZATION OF HYBRIDIZATION COEFFICIENTS**

As shown in appendix A, it is not possible to optimize the total static covariance matrix \( \bar{B}^{\text{tot}} \) in practice. However, it is possible to consider the static covariance matrix \( \bar{B} \) as given, and to optimize the \( \beta^{c}_{i} \) vector. To get exact expressions (e.g., \( \beta^{c}_{i} \) being a function of the vertical level only), we assume that this vector is a function of \( M \) parameters gathered in a vector \( \gamma \). The optimality condition for each parameter \( \gamma_{m} \) is given by

\[
\frac{\partial \hat{e}_{ij}}{\partial \gamma_{m}} = 0 \Rightarrow \left[ \sum_{ij} (\hat{B}_{ij} - \hat{B}_{ij}^{*}) \frac{\partial \beta^{c}_{i} \beta^{c}_{j}(\gamma)}{\partial \gamma_{m}} \hat{B}_{ij} \right] = 0
\]

\[
\Rightarrow \sum_{ij} \frac{\partial \beta^{c}_{i} \beta^{c}_{j}(\gamma)}{\partial \gamma_{m}} \hat{B}_{ij} \mathbb{E}[\hat{B}_{ij} - \hat{B}_{ij}^{*}] = 0
\]

\[
\Rightarrow \sum_{ij} \frac{\partial \beta^{c}_{i} \beta^{c}_{j}(\gamma)}{\partial \gamma_{m}} \hat{B}_{ij} (L_{ij}^{h} - 1)\mathbb{E}[\hat{B}_{ij}] + \beta^{c}_{i}(\gamma) \beta^{c}_{j}(\gamma) \hat{B}_{ij} = 0.
\] (22)

hence,

\[
\frac{\partial \beta^{c}_{i}}{\partial \gamma_{m}} = \delta_{\text{lev}(i),m},
\] (25a)

\[
\frac{\partial^{2} \beta^{c}_{i}}{\partial \gamma_{m} \partial \gamma_{k}} = 0,
\] (25b)

where \( \delta_{ij} \) is the Kronecker delta (equal to 1 if \( i = j \) and 0 otherwise). Thus, the system in Eq. (23) can be simplified in

\[
\mathcal{A}_{L_{ij}}(L, \gamma) = L_{ij}^{h} + \frac{\mathbb{E}[\hat{B}_{ij}]}{\mathbb{E}[\hat{B}_{ij}]} \gamma_{\text{lev}(i)} \gamma_{\text{lev}(j)} \hat{B}_{ij} - L_{ij},
\] (26a)

\[
\mathcal{A}_{\gamma_{m}}(L, \gamma) = \sum_{ij} [\delta_{\text{lev}(i),m} \gamma_{\text{lev}(j)} + \gamma_{\text{lev}(i)} \delta_{\text{lev}(j),m}] \hat{B}_{ij}
\]

\[
\times [(L_{ij}^{h} - 1)\mathbb{E}[\hat{B}_{ij}] + \gamma_{\text{lev}(i)} \gamma_{\text{lev}(j)} \hat{B}_{ij}].
\] (26b)

3) **LINEAR SYSTEM**

In the special case where the vector \( \beta^{c}_{i} \) is uniform, it is equivalent to say that it depends on one parameter \( \gamma \) only: \( \beta^{c}_{i} = \gamma \). The system in Eq. (23) becomes then

\[
\mathcal{A}_{L_{ij}}(L, \gamma) = L_{ij}^{h} + \frac{\mathbb{E}[\hat{B}_{ij}]}{\mathbb{E}[\hat{B}_{ij}]} \gamma^{2} \hat{B}_{ij} - L_{ij},
\] (27a)

\[
\mathcal{A}_{\gamma}(L, \gamma) = 2 \gamma \sum_{ij} \hat{B}_{ij} [(L_{ij}^{h} - 1)\mathbb{E}[\hat{B}_{ij}] + \gamma^{2} \hat{B}_{ij}].
\] (27b)

If we exclude the case where \( \beta^{c}_{i} = \gamma = 0 \), Eq. (27b) can be divided by \( \gamma \) and the system in Eq. (27) becomes linear in \( L_{ij} \) and \( \gamma^{2} \).
b. Properties of the linear system

1) HYBRIDIZATION COEFFICIENT BEHAVIOR

Equation (27b) can be rewritten to express the hybridization coefficient:

$$\gamma^2 = \frac{\sum_{ij} \overline{B}_{ij}(1 - L_{ij}) \mathbb{E}[\overline{B}_{ij}]}{\sum_{ij} \overline{B}_{ij}^2} = \frac{\sum_{ij} \overline{B}_{ij} B_{ij}^{opt}}{\sum_{ij} \overline{B}_{ij}^2},$$  \hspace{1cm} (28)

where $\overline{B}_{ij}^{opt} = (1 - L_{ij}^h) \cdot \mathbb{E}[\overline{B}]$ is the optimal total static covariance matrix given by Eq. (A2). Equation (28) shows two interesting properties of the optimal hybridization coefficient $\gamma^2$:

- $\gamma^2$ takes the amplitude of the specified static covariance matrix $\overline{B}$ into account. For instance, a specified static covariance matrix $B = 2\overline{B}$ would lead to a new hybridization coefficient $\gamma'$ such as $\gamma'^2 = \gamma^2 / 2$.
- $\gamma^2$ reflects the match between the specified static covariance matrix $\overline{B}$ and the optimal total static covariance matrix $\overline{B}_{ij}^{opt}$. If $\overline{B} = \overline{B}_{ij}^{opt}$, then $\gamma^2 = 1$.

2) ASYMPTOTIC BEHAVIOR

We can also analyze the asymptotic behavior of the system in Eq. (27) for an infinite ensemble size. In this case, the localization $L_{ij}$, optimized alone, goes to 1, so that Eq. (27a) becomes

$$\lim_{N \to \infty} \mathcal{A}(L^h, \gamma) = L^h + \frac{\mathbb{E}[\overline{B}^*]}{\mathbb{E}[\overline{B}_{ij}^*]^2} \gamma^2 \overline{B}_{ij} - 1.$$  \hspace{1cm} (29)

Plugging this result into Eq. (27b), we get the following:

$$\lim_{N \to \infty} \mathcal{A}_m(L^h, \gamma) = 2 \sum_{ij} \gamma^2 \overline{B}_{ij} \left[1 - \frac{\mathbb{E}[\overline{B}^*]}{\mathbb{E}[\overline{B}_{ij}^*]^2}\right].$$  \hspace{1cm} (30)

Since $\gamma \geq 0$ and $\mathbb{E}[(\overline{B}_{ij}^*)^2] = \mathbb{V}ar(\overline{B}_{ij}^*) + \mathbb{E}[\overline{B}_{ij}^*]^2 > \mathbb{E}[\overline{B}_{ij}^*]^2$, the only solution verifying $\mathcal{A}(L^h, \gamma) = 0$ when $N$ goes to infinity is $L^h = 1$ and $\gamma_m = 0$. This is the expected result since there is no need for localization nor hybridization for an infinite ensemble size.

3) BENEFITS OF HYBRIDIZATION

In appendix B, we show that the difference between the expected quadratic errors $e^h$ and $e$, respectively, measuring the accuracy of the optimal localized-hybridized covariance matrix $\overline{B}^h$ and the optimal-localized only covariance matrix $\overline{B}$ is given by

$$e^h - e = -\gamma^4 \sum_{ij} \frac{\overline{B}_{ij}^2 \mathbb{V}ar(\overline{B}_{ij})}{\mathbb{E}[\overline{B}_{ij}^2]} \leq 0.$$  \hspace{1cm} (31)

This result is very important since it shows the superiority of the hybrid formalism in the linear case. Indeed, the optimal localized-hybridized covariance matrix $\overline{B}^h$ is always more accurate than its optimal localized-only counterpart $\overline{B}$, whatever the static covariance matrix $\overline{B}$ specified for the hybridization.

c. Practical solution of the nonlinear system

Several issues arise when we try to solve the nonlinear system $\mathcal{A}(\mathbf{L}, \gamma) = 0$ given by Eq. (23):

- the system is nonlinear, and there is no guarantee that an iterative method such as Newton’s method will converge since there may be several solutions;
- there is no guarantee that these solutions will provide nonnegative values for $L_{ij}^h = (\beta_{ij}^*)^2$ and $\beta_{ij}^*$, which is a required property.

These problems can be mitigated by switching the root finding problem into a bound-constrained nonlinear optimization problem. From this perspective, we are now looking for the minimum of the function:

$$J(\mathbf{L}, \gamma) = \| \mathcal{A}(\mathbf{L}, \gamma) \|^2,$$  \hspace{1cm} (32)

while keeping $L_{ij}^h$ and $\beta_{ij}^*$ positive. Public algorithms such as M2QN1 (Gilbert and Lemaréchal 1989) or L-BFGS-B (Morales and Nocedal 2011) are well suited for this task. They only require the gradient of $J$:

$$\nabla J(\mathbf{L}, \gamma) = 2(\nabla \mathcal{A})^T \mathcal{A}(\mathbf{L}, \gamma).$$  \hspace{1cm} (33)

where the Jacobian matrix $\nabla \mathcal{A}$ is given in appendix C.

The same method can be applied to solve the nonlinear-level-dependent equation in Eq. (26), whose Jacobian matrix is also given in appendix C.

4. Experimental design

a. Data generation

1) ENSEMBLE SIZE

To get an experimental validation of the theory, an accurate estimation of the asymptotic covariance matrix $\overline{B}^*$ is required (i.e., sampled from a large enough ensemble). As shown in M15a,b, the signal-to-noise ratio (SNR) of a sample covariance can be easily computed, assuming that the ensemble distribution is Gaussian:

$$\text{SNR}_{\overline{B}_y} = \sqrt{\frac{N - 1}{1 + 1/(\mathbf{C}_{y}^*)^2}},$$  \hspace{1cm} (34)
where $\tilde{C}_{ij}^*$ is the asymptotic correlation. This formula gives an objective criterion to define what should be the size of a “large enough ensemble.” Thus, Fig. 1 shows that for a 1000-member ensemble, the SNR is higher than 10 for correlations over 0.11, which encompasses most of the signal of interest.

2) Ensemble of Perturbed Forecasts

Building such a 1000-member ensemble, with a full state-of-the-art ensemble data assimilation system, would require a huge computational effort. However, our goal is only to validate our theory about sampling error filtering, so we do not need to use an ensemble whose distribution would be precisely representative of a real data assimilation system. We only need an ensemble whose covariances have realistic features in term of heterogeneity and anisotropy. Thus, a long-range ensemble forecast starting from perturbed initial conditions should exhibit the required properties.

In this experiment, we are using the ARW Model with a 25-km resolution and 40 vertical levels over a large domain covering North America and large parts of the Pacific and Atlantic Oceans (see Fig. 2). Our target date is 0000 UTC 30 May 2013 when a deep trough occurred on North America, associated with strong thunderstorms over the central part of the United States, as shown in Fig. 2. A 48-h-range ensemble forecast has been started from perturbed initial conditions at 0000 UTC 28 May 2013. The ensemble mean of initial conditions is an interpolation onto the WRF grid of the NCEP-FNL analysis produced by the Global Data Assimilation System (NOAA/National Centers for Environmental Prediction 2000). Ensemble perturbations of initial conditions are randomized using the global NCEP background error covariances, which are described in Wu et al. (2002).
3) BOUNDARY CONDITIONS

To get a proper ensemble forecast on a limited-area domain, each member should use perturbed boundary conditions. However, computing 1000 sets of perturbed boundary conditions would have been very costly, so another strategy has been followed:

1) the same boundary conditions are used for all members;
2) the domain is cropped before computing covariances, to exclude areas where variances are decreased due to the similar boundary conditions.

To estimate the width of the area that should be cropped, sample variances $v^*$ have been computed for each vertical level, as well as their horizontal average $v$. Then, the normalized departure to horizontal average $(v^* - v)/v$ has been computed and averaged over all vertical levels and non-hydrometeor control variables ($u, v, T, q_s$). Figure 3 displays this quantity, which clearly shows the effect of similar boundary conditions. It indicates that excluding a band of 500 km on each side of the domain should be enough to get rid of most of the border effects. The same method has been used vertically, leading to the exclusion of the seven higher stratospheric levels before the covariance computation.

4) FLOW DEPENDENCY

The flow dependency of the forecast error covariances can be checked by looking at standard deviations and correlation functions. Thus, Fig. 4 shows that standard deviations are increasingly heterogeneous with the forecast range. This heterogeneity is confirmed for

Fig. 3. Normalized departure of variances to their horizontal average, vertically averaged, in percentage (see text for details). The solid black rectangle indicates the cropped domain.

Fig. 4. Standard deviation of temperature (°C) at level 7 (~1 km above ground) for (from top left to bottom right) time range = 12, 24, 36, and 48 h.
correlation functions in Fig. 5, where large discrepancies can be observed between the largest and smallest correlation length scales, and where strong anisotropies are associated with local flow regimes.

b. Objective evaluation of localization and hybridization

1) Ergodicity assumption

The theory presented in sections 2 and 3 is fully general since it uses statistical expectations $E$. However, as in M15a,b, an ergodicity assumption is required to use all the formulas in practice. Various ergodicity assumptions could be made based on the location, the scale, or the coordinate of any suitable base. Following M15a,b method, we start with the simplest one: the spatial and angular ergodicity assumption, which leads to homogeneous and isotropic estimations of localizations. Practically, these spatial and angular averages can be computed through a random sampling for large systems.

2) Localization fit

Even if the raw estimated localizations can be applied to sample covariance in our study, since we are directly computing Schur products, it is not the case in variational systems where Eq. (14) is used instead. As a consequence, a square root of the localization and, therefore, a positive definite estimate of this function is required. In this study, we have chosen to perform a Gaussian fit of diagnosed localizations via an exhaustive search method: a set of close discrete values within a reasonable range are all tested, and the best fit is kept. This fit might degrade the localization accuracy.

3) Static covariance definition

In the hybrid case, the static covariance matrix $\mathbf{B}$ has to be specified. In our case, we do not have a static background error covariance matrix that would be representative of our forecast error. Thus, a stationary background error covariance matrix “of the day” is used instead, defined as a homogeneous and isotropic average of the sample covariance matrix. Indeed, since spatial and angular averages of various quantities have to be computed to estimate localizations, an additional estimation of the spatial and angular average of the sample covariance matrix is an almost free by-product of this process.
4) EVALUATION METRIC

To validate the theory developed in sections 2 and 3, we ran 25-, 50-, and 100-member ensembles, independently from the reference 1000-member ensemble. For each ensemble, univariate horizontal covariances are computed on nine tiles, shown in Fig. 5. The size of tiles has been chosen so that it is large enough to represent long-distance effects of localization, yet small enough to see heterogeneity. These covariances will be possibly localized or hybridized, and their difference with the reference covariances is estimated through the L2 norm, giving a sampled estimation of the expected quadratic errors $e$ and $e^h$ used in theoretical sections. The sensitivity of the results to the arbitrary choice of the metric will be studied in the next section.

All scores were then normalized by the norm of the sampling error. For instance, the improvement between the raw sample covariance $\tilde{B}$ and its localized counterpart $B$ can be measured either with an $L_1$ or $L_2$ norm:

$$L_1 = 1 - \frac{\sum_{ij} |\tilde{B}_{ij} - B_{ij}^*|}{\sum_{ij} |B_{ij} - B_{ij}^*|},$$  \hspace{1cm} (35a)$$

$$L_2 = 1 - \frac{\sqrt{\sum_{ij} (\tilde{B}_{ij} - B_{ij}^*)^2}}{\sqrt{\sum_{ij} (B_{ij} - B_{ij}^*)^2}}.$$  \hspace{1cm} (35b)$$

These normalized quantities can be averaged for the four available forecast ranges (12, 24, 36, and 48 h), and over the 32 vertical levels taken into account, in order to increase their statistical robustness. For convexity reasons, the norm $L_1$ gives relatively more weight to small errors, and the norm $L_2$ gives more weight to large errors.
5. Experimental results

a. Evaluation of diagnosed localizations

1) QUALITY OF LOCALIZED-ONLY COVARIANCES

To assess the quality of the diagnosed localizations, the following procedure is applied:

1) the reference covariance matrix $\mathbf{B}_\text{w}$ is computed with the 1000-member ensemble;
2) the sample covariance matrix $\mathbf{B}$ is computed with independent 25-, 50-, and 100-member ensembles;
3) nonnormalized Gaussian localizations are fitted from the diagnosed localizations $L$, $L^G$, and $L^{GC}$, respectively, given by Eqs. (9), (10), and (11), and a normalized Gaussian localization that optimizes the $L_2I$ score is computed with an exhaustive search;
4) the localized covariance matrix $\mathbf{B}_l$ are evaluated with the $L_2I$ score.

Figure 6 shows the results for zonal wind ($u$), meridional wind ($v$), temperature ($T$), and specific humidity ($q_s$). The $L_2I$ scores obtained with raw and fitted localizations are both shown. It appears that

- Raw localization obtained with the general equation in Eq. (9) is quasi optimal. In some cases (e.g., zonal wind with 25 members), it is even better than the optimized normalized Gaussian localization. This is due to the normalization of the localization, which is set to one for the normalized Gaussian localization, but is optimized at a lower value for the diagnosed localization.
- The fit of diagnosed localizations is decreasing their accuracy, by a small amount though ($\sim 1\%$ of the $L_2$ score).
- Both $L$ and $L^G$ are leading to similar results, whereas $L^{GC}$ seems less accurate because of its additional approximation.
- As expected, the impact of localization decreases if the ensemble size increases since there is less sampling noise to filter. However, diagnosed localizations seem relevant whatever the ensemble size.

Figure 7 shows that there is a quasi-linear relationship between $L_1I$ and $L_2I$ scores. Thus, previous results would have been similar if the $L_1I$ score had been used instead of the $L_2I$ score.

2) ACCURACY OF LOCALIZATION LENGTH SCALES

The fit of diagnosed localizations with Gaussian functions provides a length scale. The vertical profiles of localization length scales obtained in the four previous
cases are shown for several ensemble sizes (Fig. 8) and for several variables (Fig. 9). In Fig. 8, it appears that zonal wind localization length scales deduced from formulas (9), (10), and (11) are very close to those found by an exhaustive search optimization of \( L_{2I} \), actually within a 2% margin error on \( L_{2I} \). As expected, the optimal localization length scale increases with the ensemble size, but another interesting feature is the widening of the error margin with the ensemble size: a lower accuracy on the localization length scale is required for a larger ensemble.

Figure 9 shows that localization length scales diagnosed for other variables are not always as accurate as for zonal wind, but still remain inside the 5% margin error on \( L_{2I} \). A significant variability can be observed for localization length scales between different ensemble sizes, different vertical levels, or different variables, which highlights the need for an objective determination of localizations.

**b. Superiority of the hybrid case**

We have proven in section 3b that an optimally localized-hybridized covariance matrix should always be more accurate than its localized-only counterpart. To validate this theory, we compute the localization \( L \) from Eq. (9), and then optimize the localization and hybridization coefficients by solving the linear system in Eq. (27) independently for each vertical level (thus \( \beta^c \) is a scalar).

The profiles of optimal hybrid coefficients are given for several ensemble sizes (Fig. 10) and for several variables (Fig. 11). As expected, the weight \( \beta^c \) given to the static covariance matrix decreases as the ensemble size increases, and is compensated for by an increase of the weight \( \beta^e \) given to the ensemble term. However, we can see that the common practice of letting \( (\beta^c)^2 + (\beta^e)^2 = 1 \) [e.g., Eq. (19) of Lorenc et al. (2015)] is not exactly verified here. Even if this constraint is not always used in recent studies, the usual justification for it relies on the conservation of variance, assuming that the ensemble-based and the static covariance matrices give independent estimations of the true background error covariance matrix. In the author’s opinion, the static covariance matrix is only considered as a regularization term in the linear filtering of sample covariances, whose purpose is to offset some deficiencies of the localized sample covariances. Thus, there is no objective reason to keep the sum \( (\beta^c)^2 + (\beta^e)^2 \) equal to one. In our experiments, the static covariance matrix is taken as a homogeneous and isotropic average of the sample covariances, which
explains why $(\beta^o)^2 + (\beta^e)^2$ remains close to one. However, in a case where variances of the static covariance matrix would be very different from those of the sample, $(\beta^o)^2 + (\beta^e)^2$ could take any value.

In Fig. 11, we can notice that hybridization coefficients vary depending on the considered variable, but that they remain in the same range. Thus, a common set of coefficients for all variables could be an acceptable approximation in our case, and might help maintain internal balances.

We can verify in Fig. 12 that adding a static covariance matrix, with optimal hybridization coefficients, allows for an improvement of $L_2I$ between $1\%$ and $3\%$ for all variables.

c. Heterogeneous optimization

The formalism developed to optimize the localization and the hybridization coefficients is not linked to a particular ergodicity assumption. Therefore, as in M15a,b, we can compute these quantities separately on distinct subdomains, while keeping a homogeneous and isotropic ergodicity assumption inside each subdomain. For simplicity, we choose to use the tiles over which the scores $L_1$ are computed as subdomains (illustration in Fig. 5). Thus, we obtain a first degree of heterogeneity, yet very basic. It should be noted that the static covariance matrix $\mathbf{B}$ is still estimated as a homogeneous and isotropic average of the sample covariance matrix, but is now computed independently for each subdomain.

Figure 12 shows that this heterogeneous optimization leads to almost no improvement for the localization-only case, but brings a significant improvement in the hybrid case. Figure 13 provides some clues to explain this contrast. In the localization-only case, the average of heterogeneously optimized localization length scales is very close to the value obtained for a homogeneous optimization, and their standard deviation is rather small. Conversely, the heterogeneous average differs significantly from the homogeneous length scale in the hybrid case, with an important standard deviation. This difference is linked to the behavior of hybridization coefficients in Fig. 14: $\beta^e$ is always higher with a heterogeneous optimization than with a homogeneous one, whereas $\beta^o$ consistently behaves in the opposite way. This indicates that the optimization of hybridization coefficients puts more weight on the static covariance matrix in the heterogeneous case, which is
considered more accurate than in the homogeneous case. A probable cause is the better representation of local conditions with a smaller averaging domain, which should remain large enough to ensure that

- the static covariance correlation length scale is comparable or smaller than the averaging domain size;
- the averaging domain size is large enough to get a reliable sampling.

This indicates that the method is sensitive to the quality of the specified static covariance matrix, and can benefit from a heterogeneous, and thus more accurate, one.

We can notice that a smaller $b_c$ (and accordingly a larger $b_e$) generally goes with a smaller localization length scale. Thus, when the optimization gives a larger weight to the sample covariance matrix, it also imposes a tighter localization length scale. It should be remembered that both have an impact on the effective localization [Eq. (16)]. This clear link between $b_c$ and $b_e$, and the localization length scale emphasizes the relevance of a joint estimation of these quantities.

d. 3D optimization

At the end of section 3c, we showed that it was possible to optimize vertically varying hybridization coefficients by solving the nonlinear-level-dependent system in Eq. (26). Staying in a horizontally homogeneous framework, we want to assess the differences between the following:

- the joint optimization of a horizontal localization and scalar hybridization coefficients, independently for each level (2D), as done in section 5b; and
- the joint optimization of a 3D localization and level-dependent hybridization coefficients, via the solving of the system in Eq. (26).

In the 3D case, the static covariance, as well as the localization, is homogeneous and isotropic in the horizontal plane, but dependent on the two vertical levels of the considered pair of grid points. Some differences can be expected since the hybridization coefficients are taking both horizontal and vertical covariances into account in the 3D case, whereas they are only considering horizontal covariances in the 2D case.

In Fig. 15, we can notice that the weight $b_c$ given to the static covariance matrix is larger in 2D than in 3D. The weight $b_e$ is consistently varying in the opposite direction. Figure 16 shows that the localization length scale is varying in the same direction as $b_c$, as expected.

A possible interpretation of this difference can be supported by the horizontally homogeneous static covariance matrix $B$. If this horizontal ergodicity assumption is better verified for horizontal covariances than for vertical covariances, then $B$ can be considered more accurate in the 2D case than in the 3D case. Thus,
the optimization is giving more weight to $\mathbb{B}$ in the 2D case than in the 3D case.

6. Conclusions

Generally considered to be two separate techniques available to enhance an ensemble-based covariance matrix, localization and hybridization can actually be optimized simultaneously. Following the work of M15a,b, we have shown that the theoretical framework of linear filtering of sample covariances could be extended to compute consistent localization functions and hybridization coefficients. We have also shown that hybridization could provide a systematically more accurate covariance matrix if optimal weights were found, whatever the static covariance matrix specified in the system. A variational formulation to solve the system in practice has been provided, which ensures that the localization and hybridization coefficients remain inside a valid range.

Using a realistic 1000-member ensemble of forecasts to compute a reference covariance matrix, we have shown experimentally that the localization formulas given by M15a,b were quasi optimal in filtering sampling error, and also that the optimal hybridized-localized sample covariance matrix was more accurate than its localized-only counterpart. Several results have highlighted the need for a common and objective optimization of localization and hybridization coefficients: variability with the considered variable, vertical level and ensemble size, and correlation between optimal localization and hybridization coefficients. A heterogeneous experiment has obtained improved results, with a larger weight on the static covariance matrix. This has been interpreted as a sign that the method was sensitive to the quality of the static covariance matrix, and that enhanced one could bring further improvements. Finally, it has been shown that the optimized hybridization coefficients were different according to whether they were optimized to take horizontal and vertical covariances into account, or only horizontal covariances.

It should be outlined that this work has been focusing on the filtering of sampling noise present in any ensemble-based covariance matrix. In the present theory, the ability of hybridization to alleviate other issues such as model error has not been taken into account. However, this limitation has made possible the determination of objective formulas for localization and hybridization coefficients. To our knowledge, it is the first method allowing such a joint optimization. In future work, the impact of this method in a cycled data assimilation scheme should be evaluated, within a realistic operational system framework.

Acknowledgments. This work has been partially funded by the Air Force Weather Agency. The authors thank Michelle Menard for her careful proofreading.

APPENDIX A

Optimization of the Total Static Covariance Matrix

At first sight, it seems possible to optimize the total static covariance matrix $\mathbb{B}^{tot}$ involved in Eq. (15). This optimal total static covariance matrix is denoted $\mathbb{B}^{opt}$. Its optimality condition would be given by

$$\frac{\partial e^h}{\partial \mathbb{B}_{ij}^{opt}} = 0 \Leftrightarrow \mathbb{E}[\mathbb{B}_{ij}^h - \mathbb{B}_{ij}^*] = 0,$$

where $e^h$ is the quadratic error given by Eq. (18). The absence of bias in the sampling error would imply that $\mathbb{E}[\mathbb{B}_{ij}^*] = \mathbb{E}[\mathbb{B}_{ij}]$, so that

$$\frac{\partial e^h}{\partial \mathbb{B}_{ij}^{opt}} = 0 \Leftrightarrow \mathbb{B}_{ij}^{opt} = (1 - L^h_{ij})\mathbb{E}[\mathbb{B}_{ij}] \leftrightarrow \mathbb{B}^{opt} = (1 - L^h)\mathbb{E}[\mathbb{B}],$$

where $L^h$ is the localization length scale (km).
where $\mathbf{I}$ is a matrix whose coefficients are all 1. Unfortunately, it appears that $(\mathbf{1} - \mathbf{L}^h)$ can have negative eigenvalues, and so does $\mathbf{B}^{opt}$. Since a positive definite (or at least semidefinite) matrix is required in the variational framework, it is not practical to optimize the total static covariance matrix with this technique.

\[ e^h = \mathbb{E} \left[ \sum_{ij} (L^h_{ij} \bar{B}_{ij} + \gamma^2 \bar{B}_{ij} - \bar{B}_{ij}^*)^2 \right] \]
\[ = \sum_{ij} \mathbb{E} \left[ \left( L^h_{ij} \bar{B}_{ij} - \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \gamma^2 \bar{B}_{ij} + \gamma^2 \bar{B}_{ij} - \bar{B}_{ij}^* \right)^2 \right] \]
\[ = \sum_{ij} \mathbb{E} \left\{ \left( L^h_{ij} \bar{B}_{ij} - \bar{B}_{ij}^* + \gamma^2 \bar{B}_{ij} \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \bar{B}_{ij} \right) \right)^2 \right\} \]
\[ = \sum_{ij} \mathbb{E}[(\bar{B}_{ij} - \bar{B}_{ij}^*)^2] + 2\gamma^2 \sum_{ij} \mathbb{E} \left[ \left( L^h_{ij} \bar{B}_{ij} - \bar{B}_{ij}^* \right) \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \bar{B}_{ij} \right) \right] + \gamma^4 \sum_{ij} \mathbb{E}[B^2_{ij}] \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \bar{B}_{ij} \right)^2 \].

In the previous equation, the first term is the expected quadratic error $e$ measuring the accuracy of localized-only covariance matrix $\mathbf{B}$, given in Eq. (8). Keeping in mind the following:

- the sampling error is unbiased: $\mathbb{E}[\bar{B}_{ij}] = \mathbb{E}[\bar{B}_{ij}^*]$.
- the sampling error is not correlated with the asymptotic covariance matrix: $\mathbb{E}[\bar{B}_{ij} \bar{B}_{ij}^*] = \mathbb{E}[(\bar{B}_{ij}^*)^2]$.
- the localization optimized alone is such as: $\mathbb{E}[(\bar{B}_{ij}^*)^2] = \mathbb{E}[\bar{B}_{ij} \bar{B}_{ij}] = L_{ij} \mathbb{E}[B^2_{ij}]$.

then the second term can be simplified in

\[ 2\gamma^2 \sum_{ij} \mathbb{E} \left[ (L^h_{ij} \bar{B}_{ij} - \bar{B}_{ij}^*) \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \bar{B}_{ij} \right) \right] \]
\[ = 2\gamma^2 \sum_{ij} \mathbb{E} \left[ (L^h_{ij} \bar{B}_{ij} - \mathbb{E}[\bar{B}_{ij}]) - L_{ij} \mathbb{E}[\bar{B}_{ij}] - L_{ij} \mathbb{E}[\bar{B}_{ij}] \mathbb{E}[B^2_{ij}] \right] \]
\[ + \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \mathbb{E}[(\bar{B}_{ij}^*)^2] \]
\[ = 2\gamma^2 \sum_{ij} \mathbb{E} \left[ \mathbb{E}[\bar{B}_{ij}] \right] \left[ \frac{\mathbb{E}[(\bar{B}_{ij}^*)^2]}{\mathbb{E}[B^2_{ij}]} - 1 \right] \]
\[ = 2\gamma^2 \sum_{ij} \mathbb{E} \left[ \bar{B}_{ij} \right] (L_{ij} - 1). \]

APPENDIX B

**Benefits of the Hybridization**

In the linear case, the $\beta$ vector is uniform: $\beta_i = \gamma$. Using Eq. (21) to develop $L^h_{ij}$, the expected quadratic error $e^h$ measuring the accuracy of the optimal localized-hybridized covariance matrix $\mathbf{B}^h$ can be expressed as

The third term is equal to

\[ \gamma^4 \sum_{ij} \mathbb{E}[B^2_{ij}] \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]}{\mathbb{E}[B^2_{ij}]} \bar{B}_{ij} \right)^2 \]
\[ = \gamma^4 \sum_{ij} \mathbb{E}[B^2_{ij}] \left( 1 - 2 \frac{\mathbb{E}[\bar{B}_{ij}]^2}{\mathbb{E}[B^2_{ij}]} + \frac{\mathbb{E}[\bar{B}_{ij}]^2}{\mathbb{E}[B^2_{ij}]} \right) \]
\[ = \gamma^4 \sum_{ij} \mathbb{E}[B^2_{ij}] \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]^2}{\mathbb{E}[B^2_{ij}]} \right). \] (B3)

The difference between the quadratic errors of localized-hybridized covariance matrix $\mathbf{B}^h$ and the localized-only covariance matrix $\mathbf{B}$ is thus given by

\[ e^h - e = \gamma^2 \left[ 2 \sum_{ij} \mathbb{E}[\bar{B}_{ij}] (L_{ij} - 1) \right] \]
\[ + \gamma^2 \sum_{ij} \mathbb{E}[\bar{B}_{ij}] \left( 1 - \frac{\mathbb{E}[\bar{B}_{ij}]^2}{\mathbb{E}[B^2_{ij}]} \right). \] (B4)

Then, using Eq. (28) to replace $\sum_{ij} \mathbb{E}[\bar{B}_{ij}] (L_{ij} - 1)$ in Eq. (B4), we get the following:
\[ e^h - e = -\gamma^4 \sum_{ij} B_{ij}^2 \left( 1 - \frac{\mathbb{E}[\hat{B}_{ij}]^2}{\mathbb{E}[\hat{B}_{ij}^2]} \right) = -\gamma^4 \sum_{ij} \mathbb{E}[\hat{B}_{ij}^2] \] 

Appendix C

Jacobian Matrices

a. Nonlinear system

The Jacobian matrix of the nonlinear system in Eq. (23) is given by

\[
\frac{\partial \mathbf{A}_{ij}}{\partial \gamma_k} = \sum_{ij} \left[ \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_m \partial \gamma_k} \frac{\partial \beta^c_{ij}(\gamma)}{\partial \gamma_m} + \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_k \partial \gamma_m} \frac{\partial \beta^c_{ij}(\gamma)}{\partial \gamma_m} + \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_m \partial \gamma_k} \right] \times \mathbb{B}_{ij} \left( (L^h_{ij} - 1)\mathbb{E}[\hat{B}_{ij}] + \beta^c_{ij}(\gamma) \frac{\partial \beta^c_{ij}(\gamma)}{\partial \gamma_m} \right) \mathbb{B}_{ij}^2 \left[ \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_k \partial \gamma_m} \right] + \beta^c_{ij}(\gamma) \frac{\partial \beta^c_{ij}(\gamma)}{\partial \gamma_k} \mathbb{B}_{ij}^2 \left[ \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_k \partial \gamma_m} \right] \mathbb{B}_{ij}^2 + \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_m \partial \gamma_k} \mathbb{B}_{ij} \left( (L^h_{ij} - 1)\mathbb{E}[\hat{B}_{ij}] + \beta^c_{ij}(\gamma) \frac{\partial \beta^c_{ij}(\gamma)}{\partial \gamma_m} \right) \mathbb{B}_{ij}^2 \left[ \frac{\partial^2 \beta^c_{ij}(\gamma)}{\partial \gamma_k \partial \gamma_m} \right] \mathbb{B}_{ij}^2,
\]

where \( \delta \) is the Kronecker delta:

\[
\delta_{ij} = \begin{cases} 
0 & \text{if } i \neq j \\
1 & \text{if } i = j.
\end{cases}
\]

b. Nonlinear-level-dependent system

The Jacobian matrix of the nonlinear-level-dependent system in Eq. (26) is given by

\[
\frac{\partial \mathbf{A}_{ij}}{\partial \gamma_k} = \sum_{ij} \left[ \delta_{\text{lev}(i),m} \delta_{\text{lev}(j),k} + \delta_{\text{lev}(i),k} \delta_{\text{lev}(j),m} \right] \times \mathbb{B}_{ij} \left( (L^h_{ij} - 1)\mathbb{E}[\hat{B}_{ij}] + \gamma_{\text{lev}(i)} \gamma_{\text{lev}(j)} \mathbb{B}_{ij} \right) \mathbb{B}_{ij}^2 \left[ \delta_{\text{lev}(i),m} \delta_{\text{lev}(j),k} + \delta_{\text{lev}(i),k} \delta_{\text{lev}(j),m} \right] \mathbb{B}_{ij}^2.
\]

References


