



HAL
open science

Selected topics in multiscale data assimilation

Marc Bocquet, Lin Wu, Frédéric Chevallier, Mohammad Reza Koohkan

► **To cite this version:**

Marc Bocquet, Lin Wu, Frédéric Chevallier, Mohammad Reza Koohkan. Selected topics in multiscale data assimilation. International Summer School - Advanced Data Assimilation for Geosciences, Ecole de Physique des Houches, May 2012, Les Houches, France. pp.415-431, 10.1093/acprof:oso/9780198723844.003.0018 . hal-01094384

HAL Id: hal-01094384

<https://inria.hal.science/hal-01094384>

Submitted on 26 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Selected topics in multiscale data assimilation

M. BOCQUET¹, L. WU¹, F. CHEVALLIER²,
and M. R. KOOKHAN¹

¹Université Paris-Est, CEREIA joint laboratory École des Ponts ParisTech
and EdF R&D, INRIA, Paris Rocquencourt Research Center
École des Ponts ParisTech
Champs-sur-Marne, France

²Laboratoire des Sciences du Climat et de l'Environnement
CEA-CNRS-UVSQ
Gif-sur-Yvette, France

18.1 Introduction

In geophysical sciences, physics unfolds and should be modelled at many spatial and temporal scales. This stems from the very large scope of the systems often considered in ambitious applications, from the global synoptic motions of the atmosphere accounting for heat transfer and dominant winds, down to the microphysics accounting for clouds, hydrometeors, and aerosols. This is also intrinsic to the physics of turbulence that shapes geophysical flows and that correlates scales so that it becomes difficult to model them separately. In addition to the complexity of multiscale models that need to be dealt with in geophysical data assimilation systems, one needs to incorporate observations pertaining to many scales, such as in situ observations, satellite observations, radar, and lidar. Even though all of this is to a large extent ignored in current data assimilation systems, it is easy to envision future data assimilation systems accounting for these multiscale aspects, not only at a technical level by interfacing models and observations at different scales, but also at a more fundamental mathematical level in that respect. Potentially useful analysis tools exist, such as multigrid methods and wavelets, but developing a fully consistent data assimilation system based on them is still a challenge.

This chapter provides an exposition of a selection of problems considered in this field. Most of the results were obtained in the context of the ANR (French National Science Foundation) MSDAG project.

There are a few insightful papers in the geophysical literature addressing mathematical approaches to multiscale data assimilation, for instance Liu and Rabier (2002), Janjić and Cohn (2006), Oke and Sakov (2008), and Koohkan and Bocquet (2012). They offer a different, often complementary, view on the topic to what is presented here, although several concepts are common.

This chapter deals with a multiscale extension of the best linear unbiased estimate (BLUE) inference, and its application to several problems. It also deals with a more pragmatic and perhaps more rewarding approach of the same problem, focusing on the multiple-scales treatment of innovation statistics. These methods are applied to atmospheric transport and chemistry problems.

18.2 Bayesian multiscale analysis

In the following, we will discuss how to account for multiple scales in the BLUE estimation principle at the heart of most current data assimilation methods. Our starting point is the observation equation

$$\boldsymbol{\mu} = \mathbf{H}\boldsymbol{\sigma} + \boldsymbol{\epsilon}, \quad (18.1)$$

with the simplest assumptions, where $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ are the observation and error vectors defined in \mathbb{R}^p and $\boldsymbol{\sigma}$ is the control space state vector in \mathbb{R}^N . In the following discussion, $\boldsymbol{\sigma}$ also coincides with the state vector. The observation errors are assumed to follow a normal distribution, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, whereas the background errors on the state vector $\boldsymbol{\sigma}$ are assumed to be distributed according to a Gaussian, $\boldsymbol{\sigma} \sim \mathcal{N}(\boldsymbol{\sigma}_b, \mathbf{B})$.

We present some partial results drawn from Bocquet (2009), Bocquet et al. (2011), Wu et al. (2011), and Koohkan et al. (2012).

18.2.1 Scaling operators

To climb up or down the spatial or temporal scales, one needs to define a *restriction operator* that tells how a state vector σ is coarse-grained and a *prolongation operator* that tells how the state vector is refined through the scales. See Rodgers (2000) for an in-depth discussion of the topic.

First, let us consider the restriction operator. Assume σ is a state vector that is known in the finest regular grid Ω isomorphic to $\mathbb{R}^{N_{\text{fg}}}$. Let ω be a discretization of the control space, which we also call a *representation*. It is isomorphic to \mathbb{R}^N , meaning that it has N grid cells. These N grid cells are possibly of different sizes and shapes, but they form a partition of control space. It could be a regular grid, or an adaptive grid that partitions the space and possibly the time domain. The coarse-graining of σ in ω is defined by $\sigma_\omega = \Gamma_\omega \sigma$, where $\Gamma_\omega : \mathbb{R}^{N_{\text{fg}}} \rightarrow \mathbb{R}^N$ stands for the linear coarse-graining operator. This operator is supposed to be unambiguously defined. It could stand for a simple averaging. But it could also be a more complex coarse-graining, with an associated prolongation operator given by a spline interpolation, or model-specific coarser Jacobian matrices, such as \mathbf{H} .

A state vector can also be refined thanks to a prolongation operator $\Gamma_\omega^* : \mathbb{R}^N \rightarrow \mathbb{R}^{N_{\text{fg}}}$ that refines σ_ω into $\sigma = \Gamma_\omega^* \sigma_\omega$. This linear operator is ambiguous, since additional information is needed to reconstruct a source at higher resolution. A schematic of the use of the restriction and prolongation operators is displayed in Fig. 18.1

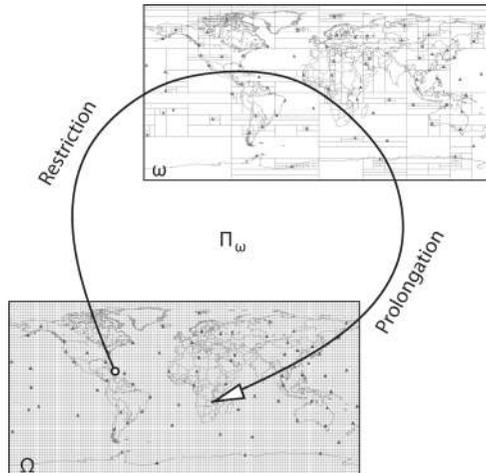


Fig. 18.1 Schematic of the restriction and prolongation operators from the finest regular grid to a representation (adaptive grid) ω , and vice versa.

One simple choice, which we shall call the deterministic one, is to set $\mathbf{\Gamma}_\omega^* = \mathbf{\Gamma}_\omega^T$. However, in this data assimilation framework, one has prior information on the control state vectors that may be exploited, in the form of a probability density function $q(\boldsymbol{\sigma})$. Following our initial assumption, q is Gaussian: $q(\boldsymbol{\sigma}) \sim \mathcal{N}(\boldsymbol{\sigma}_b, \mathbf{B})$. From this prior defined in the finest regular grid, one can infer, thanks to $\mathbf{\Gamma}_\omega$, the prior pdf of $\boldsymbol{\sigma}$ in the representation ω :

$$q_\omega(\boldsymbol{\sigma}_\omega) \sim \mathcal{N}(\boldsymbol{\sigma}_\omega^b, \mathbf{B}_\omega), \quad (18.2)$$

with

$$\boldsymbol{\sigma}_\omega^b = \mathbf{\Gamma}_\omega \boldsymbol{\sigma}_b, \quad \mathbf{B}_\omega = \mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^T. \quad (18.3)$$

Conversely, assume that one knows $\boldsymbol{\sigma}_\omega$ in the representation ω . Since the problem is underdetermined, one could opt for the most likely refinement. This is given by the mode of $q(\boldsymbol{\sigma}|\boldsymbol{\sigma}_\omega)$. From Bayes' rule, it is clear that

$$q(\boldsymbol{\sigma}|\boldsymbol{\sigma}_\omega) = \frac{q(\boldsymbol{\sigma})}{q_\omega(\boldsymbol{\sigma}_\omega)} \delta(\boldsymbol{\sigma}_\omega - \mathbf{\Gamma}_\omega \boldsymbol{\sigma}), \quad (18.4)$$

where δ is the Dirac distribution in \mathbb{R}^N . Then the mode of this posterior Gaussian distribution is given by

$$\boldsymbol{\sigma}^* = \boldsymbol{\sigma}_b + \mathbf{B} \mathbf{\Gamma}_\omega^T (\mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^T)^{-1} (\boldsymbol{\sigma}_\omega - \mathbf{\Gamma}_\omega \boldsymbol{\sigma}_b). \quad (18.5)$$

Thus, $\mathbf{\Gamma}_\omega^*$ would be an affine operator. We denote by $\mathbf{\Lambda}_\omega^*$ its tangent linear component:

$$\mathbf{\Lambda}_\omega^* \equiv \mathbf{B} \mathbf{\Gamma}_\omega^T (\mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^T)^{-1}. \quad (18.6)$$

Moreover, we define

$$\mathbf{\Pi}_\omega \equiv \mathbf{\Lambda}_\omega^* \mathbf{\Gamma}_\omega = \mathbf{B} \mathbf{\Gamma}_\omega^T (\mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^T)^{-1} \mathbf{\Gamma}_\omega, \quad (18.7)$$

so that we can choose as a prolongation operator

$$\mathbf{\Gamma}_\omega^* \equiv (\mathbf{I}_{N_{\text{fg}}} - \mathbf{\Pi}_\omega) \boldsymbol{\sigma}_b + \mathbf{\Lambda}_\omega^*, \quad (18.8)$$

where $\mathbf{I}_{N_{\text{fg}}}$ is the identity operator in the finest grid control space. Since the refinement is now a probabilistic process, errors are attached to it and can be estimated. The corresponding error covariance matrix is

$$\mathbf{P}_\omega^* = \mathbf{B} - \mathbf{B} \mathbf{\Gamma}_\omega^T (\mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^T)^{-1} \mathbf{\Gamma}_\omega \mathbf{B} = (\mathbf{I}_{N_{\text{fg}}} - \mathbf{\Pi}_\omega) \mathbf{B}.$$

As expected, if the representation ω is close to the finest grid

$$\frac{N_{\text{fg}} - \text{Rank}(\mathbf{\Pi}_\omega)}{N_{\text{fg}}} \ll 1, \quad (18.9)$$

then the refinement error is negligible. If the representation is coarse, $\text{Rank}(\mathbf{\Pi}_\omega)/N_{\text{fg}} \ll 1$, the refinement error is limited by that of the background.

These scaling operators first satisfy

$$\mathbf{\Gamma}_\omega \mathbf{\Gamma}_\omega^* = \mathbf{I}_N, \quad (18.10)$$

which is a consistency identity. Any reasonable prolongation operator should satisfy it. Then, one verifies that

$$\mathbf{\Gamma}_\omega^* \mathbf{\Gamma}_\omega = (\mathbf{I}_{N_{\text{fg}}} - \mathbf{\Pi}_\omega) \boldsymbol{\sigma}_b + \mathbf{\Pi}_\omega. \quad (18.11)$$

The linear operator $\mathbf{\Pi}_\omega$ is a projector, since it can be checked that $\mathbf{\Pi}_\omega^2 = \mathbf{\Pi}_\omega$. Besides, it is \mathbf{B}^{-1} -symmetric, since

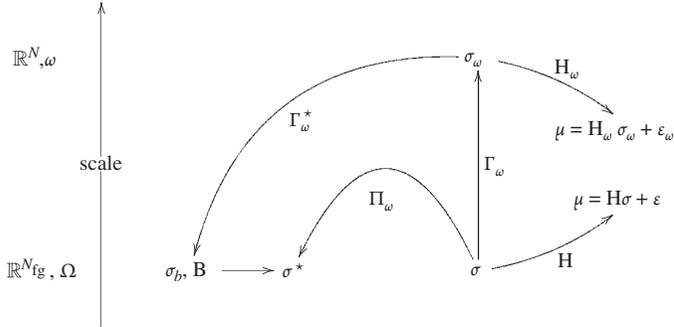
$$\begin{aligned} \langle \mathbf{x}, \mathbf{\Pi}_\omega \mathbf{y} \rangle_{\mathbf{B}^{-1}} &= \mathbf{x}^\top \mathbf{B}^{-1} \mathbf{B} \mathbf{\Gamma}_\omega^\top (\mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^\top)^{-1} \mathbf{\Gamma}_\omega \mathbf{y} \\ &= \mathbf{x}^\top \mathbf{\Pi}_\omega^\top \mathbf{B}^{-1} \mathbf{y} = \langle \mathbf{\Pi}_\omega^\top \mathbf{x}, \mathbf{y} \rangle_{\mathbf{B}^{-1}}, \end{aligned} \quad (18.12)$$

where $\langle \cdot, \cdot \rangle_{\mathbf{B}^{-1}}$ is the scalar product built on \mathbf{B}^{-1} . In matrix form, this is equivalent to

$$\mathbf{\Pi}_\omega \mathbf{B} = \mathbf{B} \mathbf{\Pi}_\omega^\top. \quad (18.13)$$

$\mathbf{\Pi}_\omega$ cannot be the identity, because the coarse-graining implies a loss of information that, in general, cannot be fully recovered.

Schematically, we have the following:



18.2.2 Observation equation in any representation

The mathematical formalism having been laid, the observation equation (18.1) can be written in any representation ω . The Jacobian \mathbf{H} becomes $\mathbf{H}_\omega = \mathbf{H} \mathbf{\Gamma}_\omega^*$. By inheritance from $\mathbf{\Gamma}_\omega^*$, \mathbf{H}_ω is an affine operator. The observation equation reads

$$\boldsymbol{\mu} = \mathbf{H}_\omega \boldsymbol{\sigma}_\omega + \boldsymbol{\epsilon}_\omega = \mathbf{H} \mathbf{\Gamma}_\omega^* \mathbf{\Gamma}_\omega \boldsymbol{\sigma} + \boldsymbol{\epsilon}_\omega = \mathbf{H} \boldsymbol{\sigma}_b + \mathbf{H} \mathbf{\Pi}_\omega (\boldsymbol{\sigma} - \boldsymbol{\sigma}_b) + \boldsymbol{\epsilon}_\omega. \quad (18.14)$$

The error $\boldsymbol{\epsilon}_\omega$ has been made scale-dependent, because several sources of errors depend on the scale, such as the aggregation errors (to be defined soon), or the errors in model subgrid physical parameterizations.

18.2.3 General scale-dependent errors

Indeed, a general decomposition of the error that account for scales could involve (i) the scale-independent observation error $\boldsymbol{\epsilon}$, which would also include model error that could be scale-independent, (ii) an error due to discretization $\boldsymbol{\epsilon}_\omega^c$ (coarse-graining), and (iii) a model error that would be scale-dependent $\boldsymbol{\epsilon}_\omega^m$:

$$\boldsymbol{\epsilon}_\omega = \boldsymbol{\epsilon} + \boldsymbol{\epsilon}_\omega^c + \boldsymbol{\epsilon}_\omega^m. \quad (18.15)$$

18.2.4 Representativeness/aggregation errors

Let us assume that errors are specified in the finest grid level, $\boldsymbol{\epsilon} = \boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}$, and that they may originate from many sources, including scale-independent model error. One assumes that there is no scale-dependent model error ($\boldsymbol{\epsilon}_\omega^m = 0$), so that errors at larger scales $\boldsymbol{\epsilon}_\omega = \boldsymbol{\mu} - \mathbf{H}_\omega\boldsymbol{\sigma}_\omega$ are supposed to be due solely to this original error, plus an error due entirely to coarsening, or aggregation error, which is a form of representativeness error, albeit from the control-space viewpoint. In that case, the model scaling is entirely described by the coarsening $\mathbf{H}_\omega = \mathbf{H}\boldsymbol{\Pi}_\omega^*$. Since $\boldsymbol{\mu} = \mathbf{H}\boldsymbol{\sigma} + \boldsymbol{\epsilon} = \mathbf{H}\boldsymbol{\sigma}_b + \mathbf{H}\boldsymbol{\Pi}_\omega(\boldsymbol{\sigma} - \boldsymbol{\sigma}_b) + \boldsymbol{\epsilon}_\omega$, the scale-dependent error reads

$$\boldsymbol{\epsilon}_\omega = \boldsymbol{\epsilon} + \boldsymbol{\epsilon}_\omega^c = \boldsymbol{\epsilon} + \mathbf{H}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)(\boldsymbol{\sigma} - \boldsymbol{\sigma}_b), \quad (18.16)$$

with $\boldsymbol{\epsilon}_\omega^c = \mathbf{H}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)(\boldsymbol{\sigma} - \boldsymbol{\sigma}_b)$ identified as the *scale-covariant* representativeness/aggregation error.

Assuming independence of the error and source error priors, the computation of the covariance matrix of these errors yields

$$\mathbf{R}_\omega = \mathbf{R} + \mathbf{R}^c = \mathbf{R} + \mathbf{H}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)\mathbf{B}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)\mathbf{H}^\text{T} = \mathbf{R} + \mathbf{H}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)\mathbf{B}\mathbf{H}^\text{T}. \quad (18.17)$$

Since $\mathbf{H}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)\mathbf{B}\mathbf{H}^\text{T}$ is a positive matrix, the mean variance of the errors always increases because of the aggregation.

Intuitively, the statistics of the innovation vector $\boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}_b$ should not depend on the scale. However, when written in terms of errors, the innovation depends formally on the representation ω :

$$\begin{aligned} \boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}_b &= \boldsymbol{\mu} - \mathbf{H}_\omega\boldsymbol{\sigma}_\omega + \mathbf{H}_\omega\boldsymbol{\sigma}_\omega - \mathbf{H}\boldsymbol{\sigma}_b \\ &= \boldsymbol{\epsilon}_\omega + \mathbf{H}_\omega\boldsymbol{\sigma}_\omega - \mathbf{H}_\omega\boldsymbol{\sigma}_\omega^b = \boldsymbol{\epsilon}_\omega + \mathbf{H}_\omega(\boldsymbol{\sigma}_\omega - \boldsymbol{\sigma}_\omega^b). \end{aligned} \quad (18.18)$$

We have used the fact that

$$\boldsymbol{\mu}_b = \mathbf{H}_\omega\boldsymbol{\sigma}_\omega^b = \mathbf{H}\boldsymbol{\Pi}_\omega^*\boldsymbol{\Gamma}_\omega\boldsymbol{\sigma}_b = \mathbf{H}\boldsymbol{\sigma}_b + \mathbf{H}\boldsymbol{\Pi}_\omega(\boldsymbol{\sigma}_b - \boldsymbol{\sigma}_b) = \mathbf{H}\boldsymbol{\sigma}_b. \quad (18.19)$$

Fortunately, this paradox is only superficial, since one can check that the statistics of the innovation are truly scale-independent:

$$\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^\text{T} = \mathbf{R} + \mathbf{H}(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega)\mathbf{B}\mathbf{H}^\text{T} + \mathbf{H}\boldsymbol{\Pi}_\omega\mathbf{B}\boldsymbol{\Pi}_\omega^\text{T}\mathbf{H}^\text{T} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\text{T}. \quad (18.20)$$

18.2.5 Estimation of the performance in the finest grid Ω

More generally, an analysis performed in the representation ω is obtained by coarsening the analysis at the finest scale. Hence, in this case, the multiscale formalism has no theoretical benefit compared with performing data assimilation in the finest grid. (There are major practical advantages, though, such as dividing the computational load.) This can be understood by directly applying Bayes' rule, using Gaussian statistics:

$$q(\boldsymbol{\sigma}_\omega | \boldsymbol{\mu}) = \frac{q(\boldsymbol{\mu} | \boldsymbol{\sigma}_\omega)q(\boldsymbol{\sigma}_\omega)}{q(\boldsymbol{\mu})} \propto \exp \left[-\frac{1}{2}(\boldsymbol{\sigma}_\omega - \boldsymbol{\sigma}_\omega^b)^T \mathbf{B}_\omega^{-1}(\boldsymbol{\sigma}_\omega - \boldsymbol{\sigma}_\omega^b) - \frac{1}{2}(\boldsymbol{\mu} - \mathbf{H}_\omega \boldsymbol{\sigma}_\omega)^T \mathbf{R}_\omega^{-1}(\boldsymbol{\mu} - \mathbf{H}_\omega \boldsymbol{\sigma}_\omega) \right]. \quad (18.21)$$

This leads to the estimate

$$\begin{aligned} \boldsymbol{\sigma}_\omega^a &= \boldsymbol{\sigma}_\omega^b + \mathbf{B}_\omega \mathbf{H}_\omega^T \left(\mathbf{R}_\omega + \mathbf{H}_\omega \mathbf{B}_\omega \mathbf{H}_\omega^T \right)^{-1} (\boldsymbol{\mu} - \mathbf{H} \boldsymbol{\sigma}_b) \\ &= \boldsymbol{\Gamma}_\omega \left[\boldsymbol{\sigma}_b + \mathbf{B} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T \right)^{-1} (\boldsymbol{\mu} - \mathbf{H} \boldsymbol{\sigma}_b) \right] = \boldsymbol{\Gamma}_\omega \boldsymbol{\sigma}_a, \end{aligned} \quad (18.22)$$

with $\boldsymbol{\sigma}^a$ the state estimation in the finest grid. The analysis error covariance matrix transforms similarly according to

$$\mathbf{P}_\omega^a = \boldsymbol{\Gamma}_\omega \left[\mathbf{B} - \mathbf{B} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T \right)^{-1} \mathbf{H} \mathbf{B} \right] \boldsymbol{\Gamma}_\omega^T = \boldsymbol{\Gamma}_\omega \mathbf{P}^a \boldsymbol{\Gamma}_\omega^T, \quad (18.23)$$

where \mathbf{P}^a is the analysis error covariance matrix in the finest grid. This can also be consistently obtained, through the finest scale:

$$q(\boldsymbol{\sigma}_\omega | \boldsymbol{\mu}) = \int d\boldsymbol{\sigma} q(\boldsymbol{\sigma}_\omega | \boldsymbol{\sigma}, \boldsymbol{\mu}) q(\boldsymbol{\sigma} | \boldsymbol{\mu}) = \int d\boldsymbol{\sigma} \delta(\boldsymbol{\sigma}_\omega - \boldsymbol{\Gamma}_\omega \boldsymbol{\sigma}) q(\boldsymbol{\sigma} | \boldsymbol{\mu}), \quad (18.24)$$

which yields (18.22) and (18.23) by a simple convolution of Gaussian probability density functions.

18.2.6 Estimation of the performance in the coarser grid ω

We have just shown that, theoretically, it is better to estimate the truth $\boldsymbol{\sigma}_t$ (living in the finest grid) within the finest grid. First, that was assuming that (i) it is numerically affordable, and (ii) there is no scale-dependent model error. Therefore, in practice and with real data, the finest grid will not necessary be the best space in which to perform the analysis, and we believe that there should be an optimal degree of refinement of control space. The concept is illustrated Fig. 18.2.

Secondly, we could be interested in the more modest but safer objective to estimate the truth projected in a coarser grid. Therefore, we are interested in the analysis of the error $\|\boldsymbol{\sigma}_\omega^a - \boldsymbol{\sigma}_\omega^t\|$, within the targeted representation ω coarser than Ω .

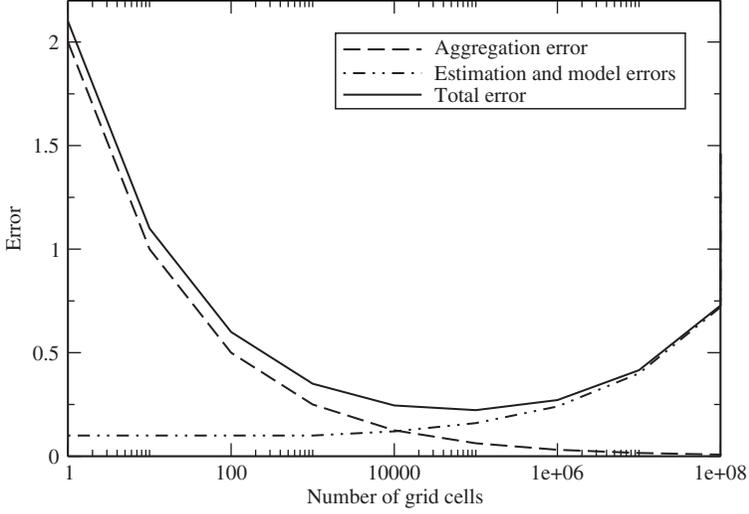


Fig. 18.2 Schematic of the posterior error (arbitrary units) or of a reasonable criterion as a function of the resolution resulting from the aggregation, model, and estimation errors.

We evaluate the inversion errors at different scales by decomposing the errors at the finest scale in \mathbf{B}^{-1} -norm into two parts: one for the variability at the finest scale and the other for coarser variations. Substituting the equation

$$\boldsymbol{\sigma}^a = \boldsymbol{\Gamma}_\omega^* \boldsymbol{\sigma}_\omega^a = \boldsymbol{\sigma}^b + \boldsymbol{\Lambda}_\omega^* (\boldsymbol{\sigma}_\omega^a - \boldsymbol{\sigma}_\omega^b), \quad (18.25)$$

into $\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|_{\mathbf{B}^{-1}}^2$, we obtain

$$\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|_{\mathbf{B}^{-1}}^2 = \|(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega) (\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t) + \boldsymbol{\Lambda}_\omega^* \boldsymbol{\sigma}_\omega^a - \boldsymbol{\Pi}_\omega \boldsymbol{\sigma}^t\|_{\mathbf{B}^{-1}}^2 \quad (18.26)$$

$$= \|(\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega) (\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t) + \boldsymbol{\Lambda}_\omega^* (\boldsymbol{\sigma}_\omega^a - \boldsymbol{\sigma}_\omega^t)\|_{\mathbf{B}^{-1}}^2. \quad (18.27)$$

Letting $\boldsymbol{\gamma}^f = (\mathbf{I}_{N_{\text{fig}}} - \boldsymbol{\Pi}_\omega) (\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t)$ and $\boldsymbol{\gamma}^c = \boldsymbol{\Lambda}_\omega^* (\boldsymbol{\sigma}_\omega^a - \boldsymbol{\sigma}_\omega^t)$, one can verify that $\boldsymbol{\gamma}^f$ and $\boldsymbol{\gamma}^c$ are \mathbf{B}^{-1} -orthogonal, that is, $\boldsymbol{\gamma}^{f,T} \mathbf{B}^{-1} \boldsymbol{\gamma}^c = 0$. Therefore, we have

$$\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|_{\mathbf{B}^{-1}}^2 = \|\boldsymbol{\gamma}^f\|_{\mathbf{B}^{-1}}^2 + \|\boldsymbol{\gamma}^c\|_{\mathbf{B}^{-1}}^2. \quad (18.28)$$

Here $\boldsymbol{\gamma}^f$ characterizes the variability at the finest scale and $\boldsymbol{\gamma}^c$ is related to the departure $\boldsymbol{\sigma}_\omega^a - \boldsymbol{\sigma}_\omega^t$ at coarse scales. As the representation ω is refined, the error $\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|_{\mathbf{B}^{-1}}^2$ is bound to decrease as fine details of the background state are better and better exploited. Meanwhile, the error $\|\boldsymbol{\gamma}^c\|_{\mathbf{B}^{-1}}^2$ is bound to increase, since retrieval at finer and finer scale is becoming erroneous.

18.3 Application to Bayesian control space design

One application of the above theoretical results is the Bayesian design of adaptive representations of control space. A typical problem is to know whether there is an optimal representation ω of control space using N grid cells. The above mathematical formalism can help us solve this problem. However, we first need to define ‘optimal’, by choosing an optimality criterion.

18.3.1 Design criteria

The Fisher criterion

Given our original incentive, to construct an adaptive grid of control space that is optimal for data assimilation, the optimality criterion must be a measure of the quality of the analysis. In Bocquet (2009), the following criterion was chosen:

$$\mathcal{J} = \text{Tr} \left(\mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right). \quad (18.29)$$

It is inspired by the Fisher information matrix,¹ normalized by the background error covariance matrix, so that the criterion is invariant under a change of coordinate in control space (for a given grid). Specifically, it measures the reduction of uncertainty granted by the observations.

In a representation ω , the criterion reads

$$\mathcal{J}_\omega = \text{Tr} \left(\mathbf{B}_\omega \mathbf{H}_\omega^T \mathbf{R}_\omega^{-1} \mathbf{H}_\omega \right). \quad (18.30)$$

The operator $\mathbf{H}_\omega = \mathbf{H} \mathbf{\Lambda}_\omega^*$ is the tangent linear operator of the affine operator \mathbf{H}_ω (which explains the difference of notation). Because only the linear part of \mathbf{H}_ω survives when averaging over the errors to obtain second-order moments, \mathbf{H}_ω appears in the criterion rather than \mathbf{H}_ω .

If one assumes that the errors are essentially scale-independent, then $\mathbf{R}_\omega \simeq \mathbf{R}$. In that case, \mathcal{J}_ω can be written in terms of $\mathbf{\Pi}_\omega$ using the machinery developed earlier:

$$\mathcal{J}_\omega = \text{Tr} \left[\mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}_\omega^T (\mathbf{\Lambda}_\omega^*)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{\Lambda}_\omega^* \right] = \text{Tr} \left(\mathbf{\Pi}_\omega \mathbf{B} \mathbf{\Pi}_\omega^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right). \quad (18.31)$$

Using the Bayesian prolongation operator $\mathbf{\Gamma}_\omega^*$, which makes use of the prior, one further obtains

$$\mathcal{J}_\omega = \text{Tr} \left(\mathbf{\Pi}_\omega \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right), \quad (18.32)$$

owing to the \mathbf{B}^{-1} -symmetry of $\mathbf{\Pi}_\omega$.

But, if the errors are scale-covariant following (18.16), the Fisher criterion (18.29) reads

$$\mathcal{J}_\omega = \text{Tr} \left(\mathbf{B}_\omega \mathbf{H}_\omega^T \mathbf{R}_\omega^{-1} \mathbf{H}_\omega \right) = \text{Tr} \left\{ \mathbf{\Pi}_\omega \mathbf{B} \mathbf{H}^T \left[\mathbf{R} + \mathbf{H} (\mathbf{I}_N - \mathbf{\Pi}_\omega) \mathbf{B} \mathbf{H}^T \right]^{-1} \mathbf{H} \right\}, \quad (18.33)$$

which is more difficult to optimize because of the nonlinear dependence of \mathcal{J}_ω on $\mathbf{\Pi}_\omega$.

¹ Ronald Aylmer Fisher was a famous English statistician (1890–1962) who did pioneering work in fundamental statistics, mathematical genetics, and evolutionary biology.

Fisher criterion and representativeness error

In the absence of scale-dependent model error, one can identify the normalized scale-covariant aggregation error, which is a measure of the representativeness error induced by the aggregation of grid cells:

$$\underbrace{\text{Tr}[\mathbf{R}^{-1}(\mathbf{R}_\omega - \mathbf{R})]}_{\text{normalized aggregation error}} = \underbrace{\text{Tr}(\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}_{\text{Fisher criterion in } \Omega} - \underbrace{\text{Tr}(\mathbf{\Pi}_\omega\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}_{\text{Fisher criterion in } \omega}. \quad (18.34)$$

It is clear that the larger the Fisher criterion, the smaller the aggregation error. Hence, optimal representations for the Fisher criterion are representations that minimize the representativeness errors.

Degrees of freedom for the signal

This criterion $\mathcal{J}_\omega = \text{Tr}(\mathbf{I}_N - \mathbf{B}_\omega^{-1}\mathbf{P}_\omega^a)$ is known to measure the number of degrees of freedom for the signal (DFS), i.e. the information load that helps resolve the parameter space. It is actually more common in data assimilation literature than the cost function (18.32). In the absence of any source of errors, the DFS are equal to the number of scalar observations that are assimilated (p here). In the presence of errors, the DFS range between 0 and the number of observations p , because the information of the observations is also used to resolve the noise (Rodgers, 2000). So the maximization of \mathcal{J}_ω entails maximizing these degrees of freedom, which seems very natural.

Assuming scale-covariant errors of the aggregation form described earlier, the dependence on $\mathbf{\Pi}_\omega$ of the DFS criterion is actually simple:

$$\begin{aligned} \mathcal{J}_\omega &= -\text{Tr}(\mathbf{B}_\omega^{-1}\mathbf{P}_\omega^a - \mathbf{I}_N) = \text{Tr}\left[\mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^T\left(\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^T\right)^{-1}\right] \\ &= \text{Tr}\left[\mathbf{\Pi}_\omega\mathbf{B}\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T\right)^{-1}\mathbf{H}\right], \end{aligned} \quad (18.35)$$

using the innovation statistics scaling (18.20).

Note that the criterion (18.32) is the limiting case of this DFS criterion when \mathbf{R} is inflated or when \mathbf{B} vanishes.

18.3.2 Numerical solution

The explicit dependence of the criteria on the representation ω , through $\mathbf{\Pi}_\omega$ is the key to an optimization of the criteria on the ω in a *library* of potential discretizations of control space. This permits an algebraic formulation of any adaptive discretization that follows a hierarchy of discretization. This formulation enables the definition of a conventional objective function to be minimized. All the mathematical details can be found in Bocquet (2009), Bocquet et al. (2011), and Bocquet and Wu (2011).

18.3.3 Example: the United Nations International Monitoring Network

Optimal representations of the control space can be constructed by the Bayesian multiscale framework and the design methodology introduced above. These optimal representations, under the information criteria such as the Fisher criterion or DFS, are in fact information maps indicating how information from observation sites could be propagated towards control space. One direct application of such information maps is to evaluate the performance of a monitoring network in the context of inverse modelling. In this section, we summarize one such example applied to the International Monitoring System (IMS) radionuclide network (Koochkan et al., 2012).

The IMS radionuclide network is operated by the United Nations CTBT Organisation (CTBTO) to enforce the Comprehensive Nuclear-Test-Ban Treaty (CTBT), which bans nuclear explosions. There is a total of 80 stations, of which 79 have assigned locations. We shall consider these 79 stations. As of June 2011, 60 stations were certified and operational to collect seismic, infrasound, and hydroacoustic data as well as radionuclide (particulate matter and noble gases) activity concentrations (triangles in Fig. 18.4 mark their locations). This IMS network is usually assessed by its detection capability using global atmospheric transport models (ATM), whereas, in our approach, we evaluate the potential of the IMS network for inverse modelling of the source.

The year 2009 is the focus of the study. Activity concentrations measurements are integrated over 24 hours. Therefore, $79 \times 365 = 28\,835$ observations are considered. The relationship between the emission source parameters and the observations is derived from the influence functions (also known as adjoint solutions or footprints or retroplumes) obtained using the FLEXPART Lagrangian transport model. The temporal extent of each influence function is 14 days, with a time step of $\Delta t = 3$ hours. In the absence of significant correlations between observation errors, and between background errors, we assume diagonal observation and background error covariance matrices.

We compute optimal adaptive grids of the source parameter space by maximizing the Fisher criterion or, alternatively, the DFS. This optimization takes into account the monitoring network, the uncertainty prescription, and the relationship between the source parameters and the observations derived from atmospheric transport over 2009.

The comparison of the performance of the regular and optimal representations shown in Fig. 18.3 demonstrates the efficiency of the optimal grids in terms of uncertainty reduction. We present one resulting optimal grid under the Fisher criterion in Fig. 18.4. Areas of the domain where the grid cells of the optimal adaptive grid are large emphasize zones where the retrieval is more uncertain, whereas areas where the grid cells are smaller and denser stress regions where more source variables can be resolved.

The observability of the globe through inverse modelling has been studied in cases with strong, realistic, and small model error. The strong-error and realistic-error cases yield heterogeneous adaptive grids, indicating that information does not propagate far from the monitoring stations, whereas in the small-error case, the grid is much more

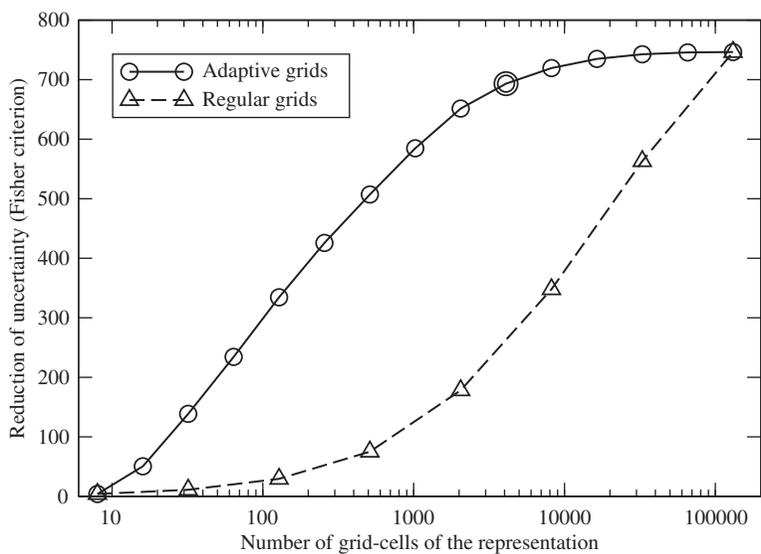


Fig. 18.3 Fisher criterion of optimal adaptive grids and regular grids against the number of grid cells in the representation. The representation in Fig. 18.4 corresponds to the point indicated by a double circle.

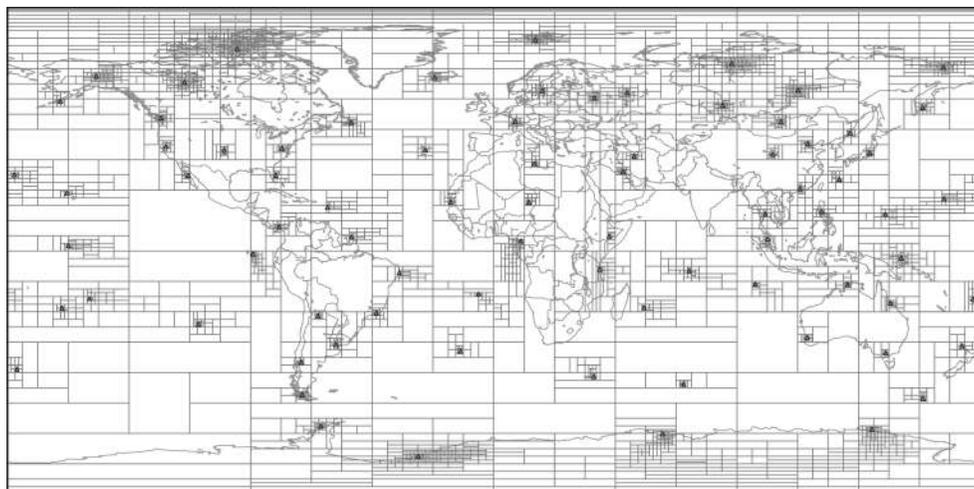


Fig. 18.4 An optimal adaptive grid for $N = 4096$ grid cells derived from the optimization of the Fisher criterion.

homogeneous. In all cases, several specific continental regions remain poorly observed, such as Africa and the tropics, because of the trade winds. The Northern Hemisphere is better observed through inverse modelling, mostly because it contains more IMS stations. This unbalance leads to a better performance of inverse modelling in the Northern Hemisphere winter.

18.3.4 Example: Ring 2

The inversion of CO₂ surface fluxes from atmospheric concentration measurements involves discretizing the flux domain in time and space. Because of the sparsity of the available concentration observations, the spatial extent of fluxes, and the dispersive nature of the atmospheric transport, the inversion of carbon fluxes is an ill-posed inverse problem. One popular remedy addressing this observation–flux imbalance is to reduce the effective degrees of freedom of fluxes by aggregation of flux variables within large regions (so-called ecoregions). However, the failure to take into account the flux variations at fine scales leads to the aggregation error that, in some cases, can be of the same order as the flux magnitude.

Thanks to the Bayesian multiscale framework and the control space design methodology, optimal multiscale representations can be obtained by maximizing the DFS. Moreover scale-dependent aggregation errors can be identified and explicitly formulated for more reliable inversions (Wu et al., 2011). We demonstrate this by performing inversions using synthetic continuous hourly CO₂ concentration data in the context of the Ring 2 experiment in support of the North American Carbon Program Mid Continent Intensive (MCI, <<http://www.ring2.psu.edu>>).

A ring of eight towers around the state of Iowa collects hourly averaged CO₂ concentration observations (in ppm) in and out of the corn belt area. The locations of these towers are shown in Fig. 18.5. The time period of the experiment is from 1 June 2007 at 0000 UTC to 16 June 2007 at 0000 UTC. The time length is 15 days. The total number of hourly observations p is thus 2880 ($8 \times 24 \times 15$). The two-dimensional spatial domain (an area of 980 km \times 980 km) is discretized into a finest regular grid of 128×128 points. Simulations of a vegetation model SiBcrop within this spatio-temporal domain are used as the reference true fluxes. Backward particle trajectories over 15 days are generated using the Lagrangian transport model LPDM driven by the atmospheric transport simulated by the meteorological model WRF. These particles within the surface grid cells are recorded to compute the influences of the fluxes on concentration observations to form the Jacobian matrix \mathbf{H} . The background error covariance matrix \mathbf{B} either is diagonal or follows a Balgovind model that parameterizes the isotropic exponential decaying correlation. We assume \mathbf{R} to be diagonal; that is, the observation errors are spatio-temporally independent.

We plot optimal representations obtained with the DFS criterion in Fig. 18.5, which demonstrate the heterogeneity of optimal propagation of information from observation sites to the whole domain. The Balgovind correlation in background errors implicitly imposes aggregation on fluxes. This results in a more uniformly distributed optimal

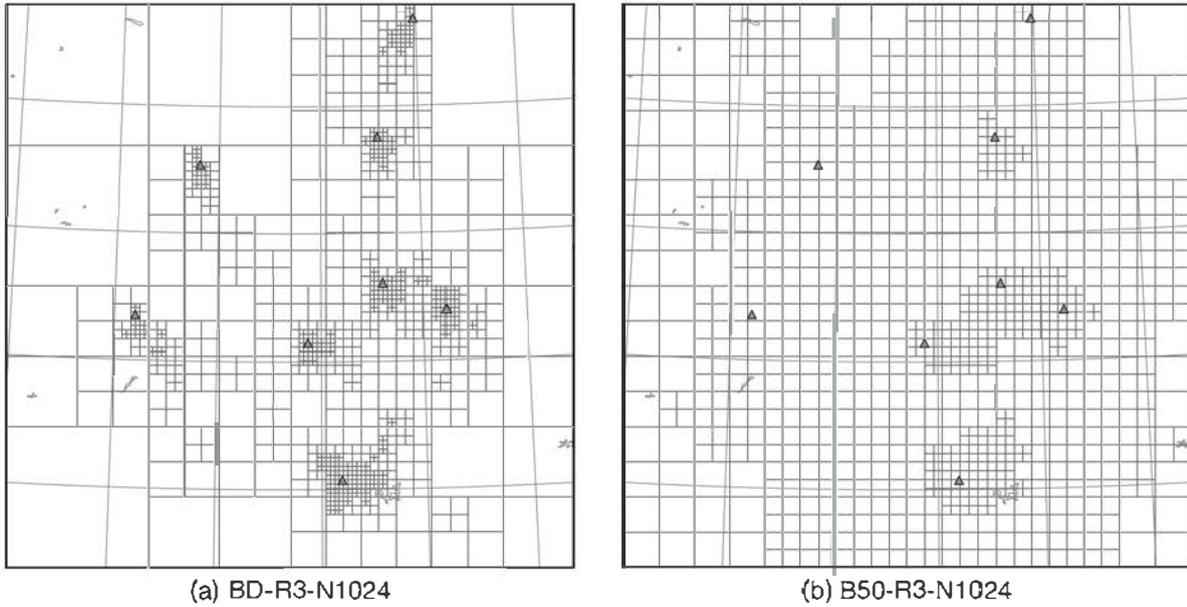


Fig. 18.5 ● Optimal multiscale representations under the DFS criterion. There is a total of 1024 grid cells. The observation error standard deviation is 3 ppm. (a) Diagonal background error covariance matrix. (b) Balgovind background error covariance matrix with correlation length equal to 50 km.

grid than that in the case of diagonal \mathbf{B} . We compare the performance of the regular and ● optimal representations with or without taking into account the aggregation error in Fig. 18.6. For diagonal \mathbf{B} , the optimal grids are more efficient than regular grids in terms of information gain (similar to Fig. 18.3). For Balgovind \mathbf{B} , this gain of optimal grids is less obvious (Fig. 18.6(a)), since those optimal grids are more uniformly distributed. When the correlations in the errors of a priori fluxes are physically unrealistic (Fig. 18.6(b)), it is recommended that the aggregation errors should be taken into account explicitly. In general, only a small part of observations from the sparse Ring 2 network are effectively assimilated ($\text{DFS}/p < 20\%$).

18.4 Empirical multiscale statistics

In previous sections, we have introduced restriction and prolongation operators to transfer fluxes between scales. The aggregation error related to these operators has been identified and formulated for viable inversions. In CO_2 inversions, we have shown the important role of the background error covariance matrix.

In this section, we sketch a quite different approach (Chevallier et al., 2012) addressing the multiscale issue, in which objective statistics are first explored to estimate the background error correlations at the finest site-level scale. Then the background error covariance matrices at coarser scales are constructed by averaging using established statistical models at the finest scale.

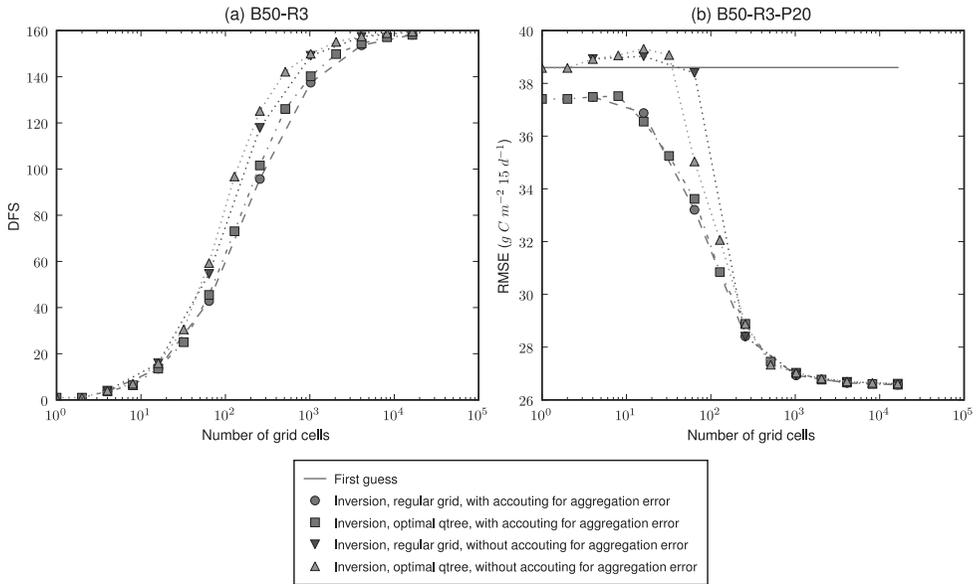


Fig. 18.6 Performance of the regular and optimal multiscale grids. The background error covariance matrix \mathbf{B} takes the form of a Balgovind correlation with correlation length equal to 50 km. The observation error standard derivation is 3 ppm. (a) DFS value. (b) Root-mean-square Error (RMSE) with a first guess generated by perturbing the true flux with a smaller correlation length of 20 km.

18.4.1 Objective statistics at the finest scale

The objective statistics are based on the comparison between the surface fluxes calculated by a process-based terrestrial ecosystem model ORCHIDEE and daily averages of CO_2 flux measurements at 156 sites across the world in the global FLUXNET network. We use flux observations collected between 1991 and 2007. ORCHIDEE simulations are performed for these sites and are driven by meteorological data (Interim Reanalysis of the European Centre for Medium-Range Weather Forecasts). The background flux errors have been estimated by the statistics of the model–observation differences (the innovation).

At the daily scale, the standard deviation of the model–data fit is $2.5 \text{ gC m}^{-2} \text{ d}^{-1}$. Temporal autocorrelations (Fig. 18.7(b)) are significant at the weekly scale (>0.3 for lags less than four weeks). The spatial structure of the error appears to be a function of the lag distance between pairs of sites based on the Pearson correlation coefficient of the model–observation differences (Fig. 18.7(a)). The median reveals spatial structure at short distances (<100 km). The correlation median is 0.33 for distances less than 100 km, 0.26 for distances between 100 and 200 km, and negligible beyond 400 km.

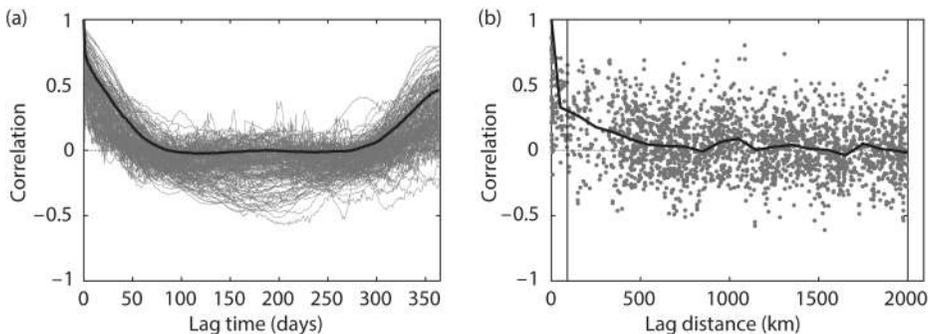


Fig. 18.7 (a) Time correlogram, i.e. autocorrelation between the errors of the ORCHIDEE model at distant times for the same site. Each thin line corresponds to a different site. The thick line represents the all-site autocorrelation; (b) Distance correlogram, i.e. correlation between the daily net ecosystem exchange (NEE) errors of the ORCHIDEE model at pairs of distant sites for the same time. Each point includes all the common years of one of the site pairs. The thick line represents the median of the points per 100 km bin.

The short spatial correlation length estimated from FLUXNET data does not seem to justify the aggregations by imposing biome-dependent correlations.

A statistical model of all-site autocorrelation can be established by fitting the thick lines in Fig. 18.7,

$$r_{\tau}(\tau) = 0.767 - 0.018\tau + (0.01222\tau)^2 - (0.00809\tau)^3 + (0.00513\tau)^4 - (0.0012\tau)^5 + 0.216/(1 + 10\tau), \quad (18.36)$$

$$r_{\delta}(\delta) = 1/(0.032\delta + 1), \quad (18.37)$$

with r_{τ} , the dependence of the correlation as a function of lag time τ in days, and r_{δ} , the correlation as a function of the lag distance δ in km.

18.4.2 Background error at coarser scales

If each flux variable at a coarse scale is assumed to be the arithmetic mean of fluxes at finest pixels of 1 day and 1 km², then the variance of the flux error on a coarser grid cell can be computed by multiplying the error variance of the finest pixel with the arithmetic mean of correlations of flux errors between two arbitrary finest pixels in that coarser grid cell. The correlation of errors of fluxes between two coarser-scale grid cells can be computed in a similar way.

In Fig. 18.8, we show the effect of temporal and spatial aggregation on the error statistics at various aggregation scales. In general, aggregation reduces error variances, while increasing correlations. As an example, for a typical inversion at a global scale of grid-point (300 km × 300 km) monthly fluxes, the prior flux error follows an approximate e -folding correlation length of 500 km only, with correlations from one month to the next as large as 0.6.

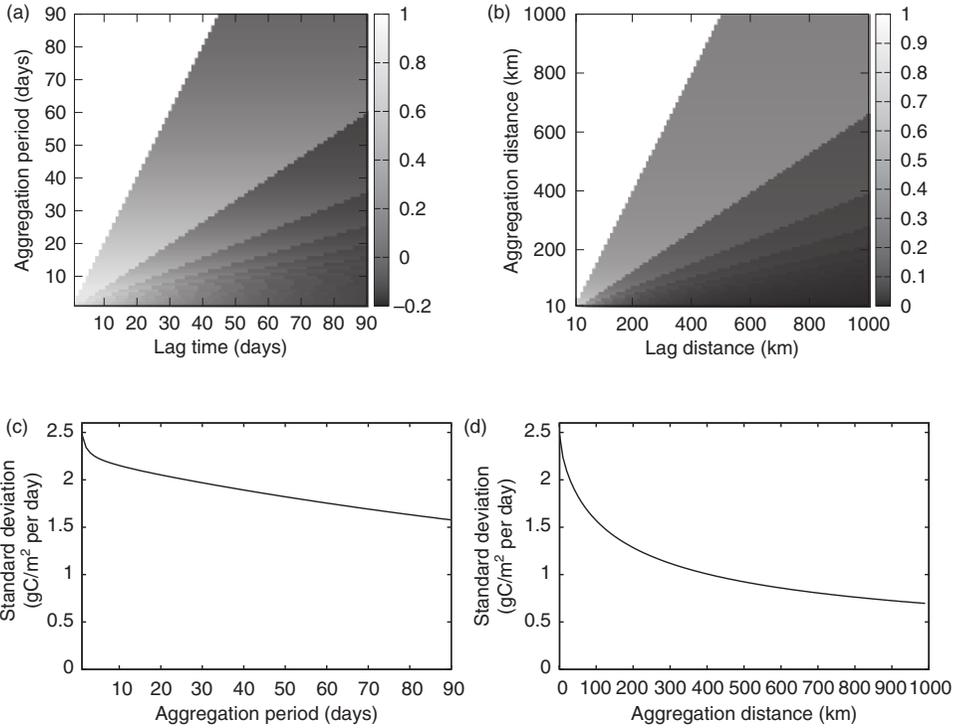


Fig. 18.8 Effect of (a, c) temporal and (b, d) spatial aggregation of the fluxes on error correlation (a, b) and error standard deviation (c, d). The aggregation distance (b, d) is defined as the length of the side of a square on which the aggregation is performed.

18.5 Conclusion

These selected topics have offered only a restricted view on a few facets of the multiscale aspects of geophysical data assimilation. In spite of the apparent heterogeneity of the formalism and concepts, a few pivotal ideas and requirements have emerged. The first is the need for properly formalized operators for changing scales and converting/transferring information through the scales. These operators need to be able to consistently transfer not only the fields but also their statistics. Another key idea is the importance of accounting and formalizing representativeness errors or alternatively aggregation errors. Finally, the paradigm of information budget that underlies modern data assimilation is even more pregnant when dealing with the multiscale aspects of data assimilation, since changing scales may involve losing or gaining information, such as the forecast and analysis steps of standard data assimilation.

References

- Bocquet, M. (2009). Towards optimal choices of control space representation for geophysical data assimilation. *Mon. Weather. Rev.*, **137**, 2331–2348.
- Bocquet, M. and Wu, L. (2011). Bayesian design of control space for optimal assimilation of observations. II: Asymptotics solution. *Q. J. R. Meteorol. Soc.*, **137**, 1357–1368.
- Bocquet, M., Wu, L., and Chevallier, F. (2011). Bayesian design of control space for optimal assimilation of observations. I: Consistent multiscale formalism. *Q. J. R. Meteorol. Soc.*, **137**, 1340–1356.
- Chevallier, F., Wang, T., Ciais, P., Maignan, F., Bocquet, M., Arain, A., Cescatti, A., Chen, J.-Q., Dolman, H., Law, B. E., Margolis, H. A., Montagni, L., and Moors, E. J. (2012). What eddy-covariance flux measurements tell us about prior errors in CO_2 -flux inversion schemes. *Global Biogeochem. Cycles*, **26**, GB1021.
- Janjić, T. and Cohn, S. E. (2006). Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Weather. Rev.*, **134**, 2900–2915.
- Koohkan, M. R. and Bocquet, M. (2012). Accounting for representativeness errors in the inversion of atmospheric constituent emissions: application to the retrieval of regional carbon monoxide fluxes. *Tellus B*, **64**, 19047.
- Koohkan, M. R., Bocquet, M., Wu, L., and Krysta, M. (2012). Potential of the international monitoring system radionuclide network for inverse modelling. *Atmos. Environ.*, **54**, 557–567.
- Liu, Z.-Q. and Rabier, F. (2002). The interaction between model resolution, observation resolution and observation density in data assimilation: a one-dimensional study. *Q. J. R. Meteorol. Soc.*, **128**, 1367–1386.
- Oke, P. R. and Sakov, P. (2008). Representation error of oceanic observations for data assimilation. *J. Atmos. Ocean. Tech.*, **25**, 1004–1017.
- Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*, Vol. 2. World Scientific, Singapore.
- Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., Rayner, P., and Davis, K. (2011). Optimal representation of source–sink fluxes for mesoscale carbon dioxide inversion with synthetic data. *J. Geophys. Res.*, **116**, D21304.