Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data

Lin Wu,[1,2] Marc Bocquet,[1,2] Thomas Lauvaux,[3] Frédéric Chevallier,[4] Peter Rayner,[5] and Kenneth Davis[3]

Received 3 May 2011; revised 15 August 2011; accepted 16 August 2011; published 5 November 2011.

[1] The inversion of CO_2 surface fluxes from atmospheric concentration measurements involves discretizing the flux domain in time and space. The resolution choice is usually guided by technical considerations despite its impact on the solution to the inversion problem. In our previous studies, a Bayesian formalism has recently been introduced to describe the discretization of the parameter space over a large dictionary of adaptive multiscale grids. In this paper, we exploit this new framework to construct optimal space-time representations of carbon fluxes for mesoscale inversions. Inversions are performed using synthetic continuous hourly CO_2 concentration data in the context of the Ring 2 experiment in support of the North American Carbon Program Mid Continent Intensive (MCI). Compared with the regular grid at finest scale, optimal representations can have similar inversion performance with far fewer grid cells. These optimal representations are obtained by maximizing the number of degrees of freedom for the signal (DFS) that measures the information gain from observations to resolve the unknown fluxes. Consequently information from observations can be better propagated within the domain through these optimal representations. For the Ring 2 network of eight towers, in most cases, the DFS value is relatively small compared to the number of observations d (DFS/d < 20%). In this multiscale setting, scale-dependent aggregation errors are identified and explicitly formulated for more reliable inversions. It is recommended that the aggregation errors should be taken into account, especially when the correlations in the errors of a priori fluxes are physically unrealistic. The optimal multiscale grids allow to adaptively mitigate the aggregation errors.

Citation: Wu, L., M. Bocquet, T. Lauvaux, F. Chevallier, P. Rayner, and K. Davis (2011), Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, *J. Geophys. Res.*, *116*, D21304, doi:10.1029/2011JD016198.

1. Introduction

[2] Top-down approaches allow inferring the spatiotemporal distribution of carbon dioxide fluxes at the Earth surface by combining diverse sources of information in a statistically optimal way, namely prior estimates of surface fluxes, CO_2 concentration observations, and atmospheric transport models that link concentrations with surface fluxes [*Tans et al.*, 1990]. Due to 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 [*Enting*, 2002].

[3] The imbalance between the fluxes and observations can be alleviated either by making more observations or by reducing the effective degrees of freedom of fluxes. New observations have been increasingly collected from extended networks or satellites [*Lauvaux et al.*, 2011; *Chevallier et al.*, 2007], and continuous observations from towers may provide additional gains [*Law et al.*, 2003; *Peylin et al.*, 2005].

[4] Assigning correlations in errors of a priori (or background) fluxes, either implicitly or explicitly, reduces the number of degrees of freedom of the flux variables. For instance, the usual prescription of the flux variations within large regions (so-called ecoregions) [*Fan et al.*, 1998; *Bousquet et al.*, 2000] implements such correlations. However, imposing prior error correlations can generate aggregation errors that, in some cases, can be of the same order as the flux magnitude [*Kaminski et al.*, 2001]. There are too few independent estimates of flux variables to allow reliable modeling of the spatial statistics for flux variations at fine scales.

[5] Mesoscale or regional inversions, which enable simulation-observation comparisons [*Lauvaux et al.*, 2009b]

[1]CEREA, Joint Laboratory École des Ponts ParisTech - EDF R&D, Université Paris-Est, Marne la Vallée, France.
[2]INRIA, Paris-Rocquencourt Research Center, Paris, France.
[3]Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.
[4]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, IPSL, Gif-sur-Yvette, France.
[5]School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia.

0148-0227/11/2011JD016198

and probably capture local meteorological or orographic scenarios, have been recently developed aiming at regional constraints on anthropogenic and biogenic carbon emissions and the coupling between regional and global scales [*Gerbig et al.*, 2003; *Lauvaux et al.*, 2008].

[6] The number of flux variables increases with finer spatiotemporal scales, which degrades the conditioning of the carbon inverse problem. As mentioned above, the dimension of the flux vector can be reduced through the aggregation of flux variables. However, it is often expected that the aggregation does not cause great loss of information. Sensitivity analyzes have been conducted with several different settings of regular resolutions for either temporal [*Gourdji et al.*, 2010] or spatial aggregations [*Tolk et al.*, 2008]. The aggregation errors, although qualified or even quantified, are not formulated explicitly for most carbon inversions.

[7] *Gerbig et al.* [2003, 2006] were the first to use heterogeneous spatial grids to lessen aggregation errors. Unrealistic correlations make the estimation of the a posteriori uncertainties of the fluxes less accurate.

[8] The adaptive spatial grid from *Gerbig et al.* [2006] is fixed and obtained with a polar projection, which is centered around one tower to adapt to the heterogeneous influence of observations. We will revise this heterogeneous inverse problem using general adaptive spatial grids with more towers that cover the domain. In this paper, the following questions are addressed: Can the adaptive grids be optimized so that the information from observations can be better propagated within the domain? What if the aggregation errors are explicitly formulated for carbon inversions? How about the role of correlations in background errors for inversions using optimal adaptive grids?

[9] Such questions are seldom investigated due to the lack of a multiscale framework for analysis. Based on a recent consistent Bayesian multiscale formalism to optimally design control space (in which control variables are to be estimated) [*Bocquet*, 2009; *Bocquet et al.*, 2011; *Bocquet and Wu*, 2011], we will construct the optimal adaptive representations of the fluxes for inversions using synthetic concentration data. Such representations are taken from a large dictionary of adaptive multiscale grids. The criterion for representation optimization is chosen to be the number of degrees of freedom for the signal (DFS) that measures the information gain from observations to resolve the unknown fluxes. Consequently the information from observations is expected to be better propagated within the domain through these optimal representations. We will then conduct carbon inversions on the optimal representations. Several issues, e.g., the information propagation, the correlations in background errors, the explicit formulation of aggregation errors, will be examined in the optimal multiscale settings. Hopefully, such optimal adaptive representations would be helpful to set up fixed multiscale grids for practical carbon flux inversions.

[10] The paper is organized as follows. In Section 2, we present the methodology for the representation optimization and the inversion in the multiscale setting. Inversions are performed in the context of the Ring 2 experiment in support of the North American Carbon Program Mid Continent Intensive (http://www.ring2.psu.edu). The experimental setup is detailed in Section 3. We report the resulting optimal representation and the corresponding inversion results in Section 4. Finally conclusions are given in Section 5.

2. Methodology

2.1. Inversion at Finest Resolution

[11] When properly discretized into N_{fg} regular grid cells at finest resolution, the surface flux vector $\boldsymbol{\sigma} \in \mathbb{R}^{N_{\text{fg}}}$ can be related to the observation (or receptor) vector $\boldsymbol{\mu} \in \mathbb{R}^{d}$ as:

$$\boldsymbol{\mu} = H[\boldsymbol{\sigma}] + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon}$ is the vector of errors originating from the imprecision of observations, the representativeness errors and the deficiency of transport models, and H is an operator that includes the atmospheric transport. In CO_2 flux inversion, the transport models are usually assumed to be linear. Hence H take the form of a source-receptor matrix $\mathbf{H}$ (the so-called Jacobian matrix) of size $d \times N_{\text{fg}}$.

[12] If one assumes Gaussian and independent errors for the fluxes and observations, a BLUE (Best Linear Unbiased Estimator) analysis reads

$$\boldsymbol{\sigma}^{a} = \boldsymbol{\sigma}^{b} + \mathbf{K}\left(\boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}^{b}\right), \tag{2}$$

where $\mathbf{K}$ is the gain matrix

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{3}$$

$\boldsymbol{\sigma}^{b}$ is the a priori flux vector, and $\boldsymbol{\sigma}^{a}$ is the a posteriori flux vector. Here $\mathbf{R}$ is the error covariance matrix for observations $\boldsymbol{\mu}$, and $\mathbf{B}$ is the error covariance matrix for the background fluxes $\boldsymbol{\sigma}^{b}$. The corresponding a posteriori error covariance matrix for fluxes $\boldsymbol{\sigma}^{a}$ after inversion is

$$\mathbf{P}^{a} = \left(\mathbf{I}_{N_{\text{fg}}} - \mathbf{K}\mathbf{H}\right)\mathbf{B}, \tag{4}$$

where $\mathbf{I}_{N_{\text{fg}}}$ is the identity matrix in $\mathbb{R}^{N_{\text{fg}}}$.

2.2. Source-Receptor Matrix

[13] If the number of observations d is significantly smaller than the number of flux components N_{fg}, it is more efficient to compute the source-receptor matrix $\mathbf{H}$ row-wise. For transport models, the rows of $\mathbf{H}$ identifies with the set of adjoint solutions indexed by the observations.

[14] The influence of an upstream flux $\sigma(\mathbf{s}, t)$ on one tracer concentration observation at receptor location $\mathbf{s}_r$ and a later time t_r can be evaluated using stochastic Lagrangian transport models, in which an ensemble of particles are released from the receptor $(\mathbf{s}_r, t_r)$ and transported *backward* to source location $(\mathbf{s}, t)$ [*Uliasz*, 1994; *Lin et al.*, 2003; *Seibert and Frank*, 2004]. The particle distribution is supposed to be well mixed (which implies time reversibility), and the turbulence is accounted for using Markov chain process. Local meteorological fields are used to drive the displacements of each particle. Each particle has its position indexed exactly rather than only up to the grid resolution of the model and consequently Lagrangian models are less diffusive than their Eulerian counterparts. Therefore, Lagrangian transport models are quite popular for the computation of $\mathbf{H}$ in mesoscale inversion [*Gerbig et al.*, 2003; *Lauvaux et al.*, 2008].

[15] For each observation at ($\mathbf{s}_r$, t_r), the influence of a flux defined at a discrete spatiotemporal grid cell ($\mathbf{s}_i$, t_n) with a finite volume and a finite time interval can be characterized by the density of the particles released from receptor location $\mathbf{s}_r$ during the time period of the measurement around t_r, e.g. one hour. Integrating up all possible flux influences multiplied by the corresponding flux $\sigma(\mathbf{s}_i, t_n)$, one obtains the variation in the concentration value at ($\mathbf{s}_r$, t_r) resulting from all the flux influences.

[16] A practical flux variable for inversion is the mass surface source-sink flux in units of (g C m^{-2} $time^{-1}$), rather than the volume flux in units of (ppm m^{-3} $time^{-1}$). If the concentration observations are in units of (ppm) (volume mixing ratio), and if the surface fluxes are diluted within a surface layer of height h in units of meters, then the footprint which relates a flux at a surface spatiotemporal grid cell to its resulting concentration variation at ($\mathbf{s}_r$, t_r) can be calculated by:

$$f(\mathbf{s}_r, t_r | x_i, y_i, t_n) = \frac{g \cdot m_{\text{air}}}{\Delta P \cdot m_{\text{CO}_2}} \frac{1}{N_{\text{tot}}} \sum_{p=1}^{N_{\text{tot}}} \Delta\tau_{p,i,n}, \tag{5}$$

where x_i, y_i, t_n are the spatiotemporal coordinates of that surface grid cell, m_{CO_2} and m_{air} are the molecular mass for CO_2 and dry air respectively, g is the gravitational acceleration, ΔP is the air pressure difference between ground level and height h above ground level, $\Delta\tau_{p,i,n}$ is the residence time for particle p staying in the spatiotemporal grid cell at (x_i, y_i, t_n), and N_{tot} is the total number of particles released for the observation at ($\mathbf{s}_r$, t_r). Here the hydrostatic approximation is assumed.

[17] Similarly to the surface fluxes, the footprint can be computed for the concentrations at boundary grid cells or at the initial time. In this study, synthetic data are used. In this synthetic case, the boundary and initial conditions can be assumed to be perfectly known, therefore they do not contribute to the variations in observation values. For each concentration observation, one can thus compute the footprint elements for the surface fluxes at all upstream spatiotemporal grid cells. These footprint elements form one row of the source-receptor matrix $\mathbf{H}$.

2.3. Error Parameterization

[18] The background error covariance matrix $\mathbf{B}$ is a key ingredient of carbon inversions, because the correction $\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^b$ lies in the column space of $\mathbf{B}$. For sparse observations, the correlations in $\mathbf{B}$ spread the information from the observations sites to their neighborhood.

[19] Unfortunately in most cases $\mathbf{B}$ is not well established. The only objective study to date [*Chevallier et al.*, 2006], which compares the simulation of vegetation models and the in situ flux observations from eddy-covariance sites, shows that, for terrestrial models run at the resolution of global transport models, the length of the spatial error correlation is a few hundred kilometers at most, and that the temporal correlation is strong (up to several weeks, even months). This motivates a spatially diagonal $\mathbf{B}$ with perfect temporal correlations for a few days in some studies [*Lauvaux et al.*, 2008]. One alternative popular choice is the exponential decaying correlation model [*Rödenbeck et al.*, 2003; *Michalak et al.*, 2004; *Peylin et al.*, 2005]. Note that *Lauvaux et al.* [2011] have combined the spatial correlations with ecosystem considerations [*Peters et al.*, 2007].

[20] In this study, the two above mentioned assumptions on $\mathbf{B}$ are tested: a spatial diagonal $\mathbf{B}$ and isotropic correlations in background errors. In the latter case, the error covariance between two spatial points $\mathbf{s}_1$ and $\mathbf{s}_2$ is computed according to the Balgovind parameterization [*Balgovind et al.*, 1983]:

$$C(\mathbf{s}_1, \mathbf{s}_2) = \kappa(\mathbf{s}_1)\kappa(\mathbf{s}_2)\left(1 + \frac{h_s}{L_s}\right)\exp\left(-\frac{h_s}{L_s}\right), \tag{6}$$

where L_s is the characteristic correlation length, $h_s = \|\mathbf{s}_1 - \mathbf{s}_2\|$ is the spatial distance between two points, and κ is the background error standard deviation which can be heterogeneous, for instance, when the information from local ecosystems is considered. In this study, we assume homogeneous κ (thus a constant). Note that similar formulations can also be introduced for temporal correlations [*Gourdji et al.*, 2010].

[21] The estimation of the observational error covariance matrix $\mathbf{R}$ is a difficult issue, because the error of the transport model is involved [*Lauvaux et al.*, 2009a]. As in most carbon inversions, we assume $\mathbf{R}$ diagonal, that is, the observation errors are spatiotemporally independent.

2.4. Multiscale Formalism

[22] We summarize the multiscale formalism in this subsection. For a detailed presentation, please refer to *Bocquet et al.* [2011].

2.4.1. Multiscale Representation Structure

[23] Let Ω be a spatiotemporal 3D (2D + T) domain of surface fluxes discretized into a regular grid at a finest scale of N_{fg} grid cells. A hierarchical scale of resolution can be obtained by successive dyadic coarse-grainings of the grid cells at the finest scale. The source-receptor matrix $\mathbf{H}$ is computed at the finest scale. Its dyadic coarse-grainings can be obtained by simple averaging or summation.

[24] Let us define a representation ω as a set of N multiscale grid cells that cover Ω. For *admissible* representations, each point in Ω corresponds to one and only one grid cell of ω. Trees are appropriate tools to describe such multiscale representations. For instance in a 1D domain, a coarse grid cell (mother tree node) can be divided into two refined grid cells (daughter tree nodes). This forms a binary tree (Figure 1a). Multiscale representations in a 2D domain can be constructed using grid cells that are Kronecker products of grid cells of two binary trees. This leads to the so-called *tiling* representation (Figure 1b), which is anisotropic, since each grid cell in this representation can have arbitrary scale levels in each direction of the 2D domain. A more numerically effective representation is to use a quaternary tree, hereafter called *qtree* (Figure 1c), instead of products of grid cells of binary trees. That is, each mother grid cell is divided into four daughter grid cells.

2.4.2. Restriction and Prolongation

[25] In order to switch scale in this multiscale setting, we define a *restriction operator* that describes how a source (a spatiotemporal flux vector) is coarse-grained, and a *prolongation operator* that describes how a source is refined to the finer scales.

[26] Let $\boldsymbol{\sigma}$ be the source at the finest scale. The coarse-graining of $\boldsymbol{\sigma}$ on ω is described by $\boldsymbol{\sigma}_\omega = \boldsymbol{\Gamma}_\omega \boldsymbol{\sigma}$, where $\boldsymbol{\Gamma}\omega$:

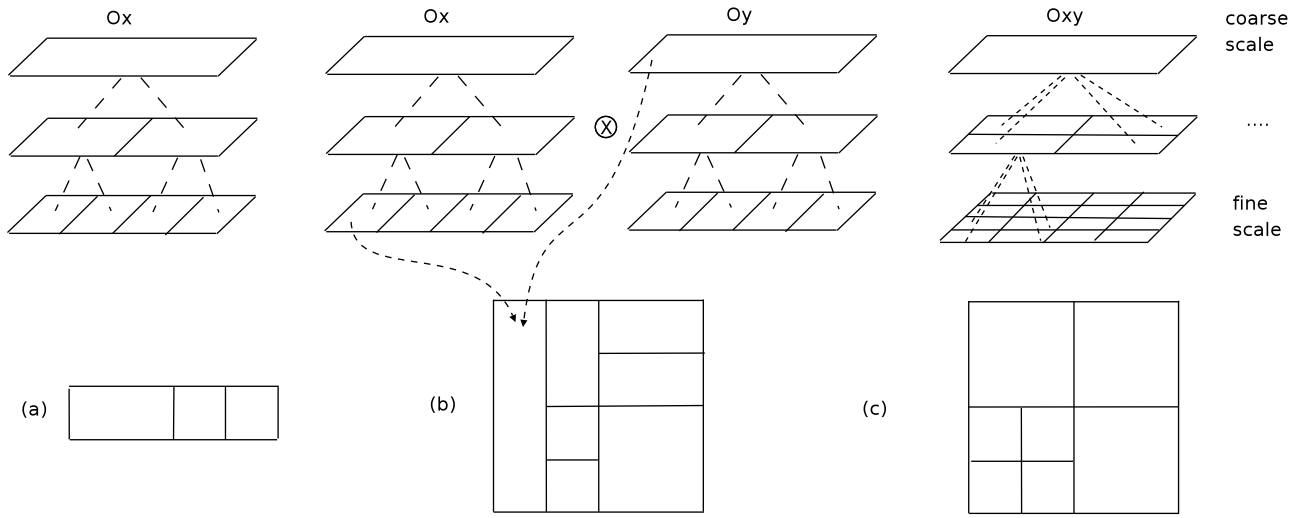

Figure 1. Schematic of the multiscale structures described by trees. (a) A binary tree for Ox. (b) The Kronecker product of grid cells of two binary trees for Ox and Oy. This corresponds to the tiling structure. (c) A quaternary tree, which corresponds to the qtree structure. For each case, an example of a generated grid is drawn. Each grid cell is related to one and only one node of each of its corresponding trees, as is illustrated with the leftmost grid cell in Figure 1b.

$\mathbb{R}^{N_{\text{fg}}} \rightarrow \mathbb{R}^N$ is the restriction operator that can be unambiguously defined as simple dyadic averaging. Refining a source $\boldsymbol{\sigma}_\omega$ on ω back to the finest scale is described by $\boldsymbol{\sigma} = \boldsymbol{\Gamma}^*_\omega \boldsymbol{\sigma}_\omega$, where $\boldsymbol{\Gamma}^*_\omega : \mathbb{R}^N \rightarrow \mathbb{R}^{N_{\text{fg}}}$ is the prolongation operator. We sketch the restriction and prolongation operators in Figure 2.

[27] The prolongation operator $\boldsymbol{\Gamma}^*_\omega$ is ambiguous, since, given only a source $\boldsymbol{\sigma}_\omega$, no information is available at finer scales to refine that source. Therefore additional information or assumptions have to be exploited to reconstruct a source $\boldsymbol{\sigma}^*$ at finest scale that corresponds to $\boldsymbol{\sigma}_\omega$. Such additional information can be the prior probability density function (pdf) $q(\boldsymbol{\sigma})$ on $\boldsymbol{\sigma}$. Thus the source can be reconstructed using the posterior pdf $q(\boldsymbol{\sigma}|\boldsymbol{\sigma}_\omega)$. A Bayesian analysis gives:

$$q(\boldsymbol{\sigma}|\boldsymbol{\sigma}_\omega) = \frac{q(\boldsymbol{\sigma})}{q_\omega(\boldsymbol{\sigma}_\omega)} q(\boldsymbol{\sigma}_\omega|\boldsymbol{\sigma}), \tag{7}$$

where $q_\omega(\boldsymbol{\sigma}_\omega)$ is the prior pdf of $\boldsymbol{\sigma}$ on ω, and the conditional pdf $q(\boldsymbol{\sigma}_\omega \mid \boldsymbol{\sigma})$ can be defined by $\delta(\boldsymbol{\sigma}_\omega - \boldsymbol{\Gamma}_\omega \boldsymbol{\sigma})$. Here δ is the Dirac distribution.

[28] If we assume a Gaussian prior at finest scale: $q(\boldsymbol{\sigma}) \sim \mathcal{N}(\boldsymbol{\sigma}^b, \mathbf{B})$, the prior on ω is also Gaussian: $q_\omega(\boldsymbol{\sigma}_\omega) \sim \mathcal{N}(\boldsymbol{\sigma}^b_\omega, \mathbf{B}_\omega)$, where $\boldsymbol{\sigma}^b_\omega = \boldsymbol{\Gamma}_\omega \boldsymbol{\sigma}^b$ and $\mathbf{B}_\omega = \boldsymbol{\Gamma}_\omega \mathbf{B} \boldsymbol{\Gamma}^{\mathrm{T}}_\omega$. In this case, the corresponding maximum likelihood estimation based on $q(\boldsymbol{\sigma}|\boldsymbol{\sigma}_\omega)$ is:

$$\boldsymbol{\sigma}^* = \boldsymbol{\sigma}^b + \mathbf{B}\boldsymbol{\Gamma}^{\mathrm{T}}_\omega \left(\boldsymbol{\Gamma}_\omega \mathbf{B} \boldsymbol{\Gamma}^{\mathrm{T}}_\omega\right)^{-1} \left(\boldsymbol{\sigma}_\omega - \boldsymbol{\Gamma}_\omega \boldsymbol{\sigma}^b\right). \tag{8}$$

This defines an unambiguous *affine* prolongation operator:

$$\boldsymbol{\Gamma}^* = \left(\mathbf{I}_{N_{\text{fg}}} - \boldsymbol{\Pi}_\omega\right)\boldsymbol{\sigma}^b + \boldsymbol{\Lambda}^*_\omega, \tag{9}$$

where $\boldsymbol{\Lambda}^*_\omega = \mathbf{B}\boldsymbol{\Gamma}^{\mathrm{T}}_\omega \left(\boldsymbol{\Gamma}_\omega \mathbf{B} \boldsymbol{\Gamma}^{\mathrm{T}}_\omega\right)^{-1}$ and $\boldsymbol{\Pi}_\omega = \boldsymbol{\Lambda}^*_\omega \boldsymbol{\Gamma}_\omega$. An application of $\boldsymbol{\Gamma}^*_\omega$ on $\boldsymbol{\sigma}_\omega$ is given by a linear transform $\boldsymbol{\Lambda}^*_\omega \boldsymbol{\sigma}_\omega$ shifted by $(\mathbf{I}_{N_{\text{fg}}} - \boldsymbol{\Pi}_\omega)\boldsymbol{\sigma}^b$, which reproduces equation (8). We use the name of affine operator to emphasize the translation related to $\boldsymbol{\sigma}^b$. It can be verified that the linear operator $\boldsymbol{\Pi}_\omega$ is a projector and has the property of $\mathbf{B}^{-1}$-symmetry which means that $\boldsymbol{\Pi}_\omega \mathbf{B} = \mathbf{B}\boldsymbol{\Pi}^{\mathrm{T}}_\omega$. The projector $\boldsymbol{\Pi}_\omega$ is a composition of a coarse-graining $\boldsymbol{\Gamma}_\omega$ and a projection $\boldsymbol{\Lambda}^*_\omega$ back from ω to the finest grid. Therefore it characterizes the variations at a coarse-scale representation ω. Consequently $\mathbf{I}_{N_{\text{fg}}} - \boldsymbol{\Pi}_\omega$ is also a projector which conserves the small-scale variations smoothed out by $\boldsymbol{\Pi}_\omega$.

[29] For efficient multiscale representation optimizations, $\mathbf{B}$ is preferred to be diagonal. When cross-correlations in background errors are present between different grid cells, $\mathbf{B}$ is non-diagonal. However, the multiscale optimization machinery can be kept by introducing a new coarse-graining operator $\widetilde{\boldsymbol{\Gamma}}_\omega = \boldsymbol{\Gamma}_\omega \mathbf{B}^{-1/2}$ where $\boldsymbol{\Gamma}_\omega$ is the original restriction operator. For a source $\boldsymbol{\sigma} \in \mathbb{R}^{N_{\text{fg}}}$, the new coarse graining is described by $\boldsymbol{\sigma}_\omega = \widetilde{\boldsymbol{\Gamma}}_\omega \boldsymbol{\sigma} = \boldsymbol{\Gamma}_\omega \mathbf{B}^{-1/2} \boldsymbol{\sigma}$. Here a linear transform

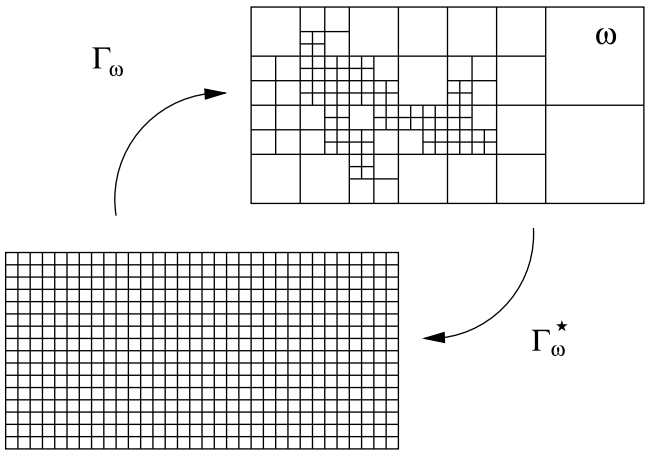

Figure 2. Schematic of the restriction operator $\boldsymbol{\Gamma}_\omega$ and of the prolongation operator $\boldsymbol{\Gamma}^*_\omega$, which relate sources at finest scale with those on a multiscale representation ω.

$\widetilde{\boldsymbol{\sigma}} = \mathbf{B}^{-1/2}\,\boldsymbol{\sigma}$ is implicitly introduced to remove the cross-correlations in the sense that the error covariance of $\widetilde{\boldsymbol{\sigma}}^b = \mathbf{B}^{-1/2}\boldsymbol{\sigma}^b$ is an identity matrix. The maximum likelihood estimation still applies with the new coarse graining operator $\widetilde{\boldsymbol{\Gamma}}_\omega$:

$$\begin{aligned}\boldsymbol{\sigma}^* &= \boldsymbol{\sigma}^b + \mathbf{B}\widetilde{\boldsymbol{\Gamma}}_\omega^{\mathrm{T}}\left(\widetilde{\boldsymbol{\Gamma}}_\omega\mathbf{B}\widetilde{\boldsymbol{\Gamma}}_\omega^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{\sigma}_\omega - \widetilde{\boldsymbol{\Gamma}}_\omega\boldsymbol{\sigma}^b\right)\\ &= \boldsymbol{\sigma}^b + \mathbf{B}^{1/2}\boldsymbol{\Gamma}_\omega^{\mathrm{T}}\left(\boldsymbol{\Gamma}_\omega\boldsymbol{\Gamma}_\omega^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{\sigma}_\omega - \boldsymbol{\Gamma}_\omega\mathbf{B}^{-1/2}\boldsymbol{\sigma}^b\right).\end{aligned} \tag{10}$$

The new prolongation operator can thus be derived as:

$$\widetilde{\boldsymbol{\Gamma}}_\omega^* = \left(\mathbf{I}_{N_{\mathrm{fg}}} - \widetilde{\boldsymbol{\Pi}}_\omega\right)\boldsymbol{\sigma}^b + \widetilde{\boldsymbol{\Lambda}}_\omega^*, \tag{11}$$

where $\widetilde{\boldsymbol{\Lambda}}_\omega^* = \mathbf{B}^{1/2}\,\boldsymbol{\Gamma}_\omega^{\mathrm{T}}\,(\boldsymbol{\Gamma}_\omega\boldsymbol{\Gamma}_\omega^{\mathrm{T}})^{-1}$ and $\widetilde{\boldsymbol{\Pi}}_\omega = \widetilde{\boldsymbol{\Lambda}}_\omega^*\widetilde{\boldsymbol{\Gamma}}_\omega$. Note that in this case the dyadic averaging of $\boldsymbol{\Gamma}_\omega$ is applied to the sources $\widetilde{\boldsymbol{\sigma}} \in \mathbb{R}^{N_{\mathrm{fg}}}$ with decorrelated errors, therefore the adaptive grid of the representation ω cannot be related directly to the original spatiotemporal lat-lon domain.

2.4.3. Aggregation Error

[30] For a given representation ω, the source-receptor matrix $\mathbf{H}$ becomes $\boldsymbol{H}_\omega = \mathbf{H}\boldsymbol{\Gamma}_\omega^*$, which is also an *affine* operator. Its linear part is $\mathbf{H}_\omega = \mathbf{H}\boldsymbol{\Lambda}_\omega^*$. The multiscale fluxes on ω are related to the observations by: $\boldsymbol{\mu} = \boldsymbol{H}_\omega\boldsymbol{\sigma}_\omega + \boldsymbol{\epsilon}_\omega = \mathbf{H}\boldsymbol{\sigma}^b + \mathbf{H}\boldsymbol{\Pi}_\omega\,(\boldsymbol{\sigma} - \boldsymbol{\sigma}^b) + \boldsymbol{\epsilon}_\omega$. Note that at finest scale $\boldsymbol{\mu} = \mathbf{H}\boldsymbol{\sigma} + \boldsymbol{\epsilon}$. Therefore we can identify the total scale-covariant error [*Bocquet et al.*, 2011]:

$$\boldsymbol{\epsilon}_\omega = \boldsymbol{\epsilon} + \mathbf{H}\left(\mathbf{I}_{N_{\mathrm{fg}}} - \boldsymbol{\Pi}_\omega\right)\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}^b\right) = \boldsymbol{\epsilon} + \boldsymbol{\epsilon}_\omega^{\mathrm{agg}}, \tag{12}$$

where $\boldsymbol{\epsilon}_\omega^{\mathrm{agg}} = \mathbf{H}(\mathbf{I}_{N_{\mathrm{fg}}} - \boldsymbol{\Pi}_\omega)(\boldsymbol{\sigma} - \boldsymbol{\sigma}^b)$ is the part of the total error resulting from the aggregation. Assuming independence between the observational error $\boldsymbol{\epsilon}$ at finest scale and the background error $\boldsymbol{\epsilon}^b = \boldsymbol{\sigma} - \boldsymbol{\sigma}^b$, one obtains

$$\mathbf{R}_\omega = \mathbf{R} + \mathbf{H}\left(\mathbf{I}_{N_{\mathrm{fg}}} - \boldsymbol{\Pi}_\omega\right)\mathbf{B}\mathbf{H}^{\mathrm{T}}. \tag{13}$$

It can then be verified that the statistics of the innovation vector $\boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}^b$ is scale-invariant when the aggregation error is identified and formulated according to equation (12): $\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ (for details see Section 3.2 of *Bocquet et al.* [2011]). Failing to taking into account $\boldsymbol{\epsilon}_\omega^{\mathrm{agg}}$ leads to an inconsistent innovation statistics $\mathbf{R} + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}$.

2.5. Inversion With Multiscale Representations

[31] The innovation vector $\boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}^b$ is scale-invariant due to the fact that $\boldsymbol{H}_\omega\boldsymbol{\sigma}_\omega^b = \mathbf{H}\boldsymbol{\sigma}^b$. The BLUE analysis on ω gives the following update:

$$\boldsymbol{\sigma}_\omega^a = \boldsymbol{\sigma}_\omega^b + \mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}\left(\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}^b\right). \tag{14}$$

The posterior error covariance matrix is

$$\mathbf{P}_\omega^a = \left(\mathbf{I}_N - \mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}\left(\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}\right)^{-1}\mathbf{H}_\omega\right)\mathbf{B}_\omega. \tag{15}$$

When the aggregation error $\boldsymbol{\epsilon}_\omega^{\mathrm{agg}}$ is considered, the term $\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}$ can be computed by $\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ in the above two formulae. By contract, when the aggregation error fails to be formulated, the inconsistent innovation statistics $\mathbf{R} + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}$ is used.

[32] The inverted flux vector $\boldsymbol{\sigma}_\omega^a$ on ω can be transformed into a flux vector $\boldsymbol{\sigma}^a$ at the finest scale by the prolongation operator:

$$\begin{aligned}\boldsymbol{\sigma}^a &= \boldsymbol{\Gamma}_\omega^*\boldsymbol{\sigma}_\omega^a = \left(\mathbf{I}_{N_{\mathrm{fg}}} - \boldsymbol{\Pi}_\omega\right)\boldsymbol{\sigma}^b + \boldsymbol{\Lambda}_\omega^*\boldsymbol{\sigma}_\omega^a\\ &= \boldsymbol{\sigma}^b + \boldsymbol{\Lambda}_\omega^*\left(\boldsymbol{\sigma}_\omega^a - \boldsymbol{\sigma}_\omega^b\right).\end{aligned} \tag{16}$$

For non-diagonal $\mathbf{B}$, $\boldsymbol{\Gamma}_\omega$ is redefined as $\widetilde{\boldsymbol{\Gamma}}_\omega = \boldsymbol{\Gamma}_\omega\mathbf{B}^{-1/2}$ to obtain the inverted fluxes at the finest scale:

$$\boldsymbol{\sigma}^a = \boldsymbol{\sigma}^b + \widetilde{\boldsymbol{\Lambda}}_\omega^*\left(\widetilde{\boldsymbol{\sigma}}_\omega^a - \widetilde{\boldsymbol{\sigma}}_\omega^b\right). \tag{17}$$

In practice, all the expressions in decorrelated space, such as $\widetilde{\boldsymbol{\sigma}}_\omega^a$, can be calculated using the same formulae obtained in the diagonal case, except that $\mathbf{B}$ and $\mathbf{H}$ should be replaced by $\mathbf{I}_{N_{\mathrm{fg}}}$ and $\mathbf{H}\mathbf{B}^{1/2}$ respectively.

2.6. Criteria for Optimal Representation of Sources

[33] Diverse criteria have been proposed to evaluate the performance of a multiscale representation ω [*Bocquet et al.*, 2011]. In this study, we choose the number of degrees of freedom for the signal (DFS) as the criterion, which is defined by [*Rodgers*, 2000]

$$\mathcal{J} = \mathrm{E}\left[\left(\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^b\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^b\right)\right]. \tag{18}$$

where E denotes the expectation operator on the background and observational errors, and $\boldsymbol{\sigma}^a$ is the BLUE analysis in equation (2). Note that $\boldsymbol{\sigma}^a$ can be obtained by the following variational problem:

$$\begin{aligned}\boldsymbol{\sigma}^a &= \underset{\boldsymbol{\sigma}}{\mathrm{argmin}}\,\chi^2(\boldsymbol{\sigma})\\ &= \underset{\boldsymbol{\sigma}}{\mathrm{argmin}}\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}^b\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}^b\right) + \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\epsilon},\end{aligned} \tag{19}$$

where the observational error is defined by the observation equation (1): $\boldsymbol{\epsilon} = \boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}$. The DFS is thus the part of χ^2 that measures the relative correction of $\boldsymbol{\sigma}^a$ to $\boldsymbol{\sigma}^b$. In the simplest case, for which a measurement is made of a scalar: $\mu = \sigma + \epsilon$, the DFS is $v^b/(v^b + v^\epsilon)$, where v^b and v^ϵ are the prior and measurement error variances respectively. If the measurement is exact ($v^\epsilon = 0$), or if there is no prior information ($v^b = \infty$), we have one degree of freedom for the signal which is provided by the measurement. By contrast, if $v^\epsilon = \infty$, we have zero DFS, or one degree of freedom for the noise.

[34] In the general vector case, the DFS can be computed by $\mathrm{Tr}(\mathbf{A})$ where $\mathbf{A} = \mathbf{K}\mathbf{H}$ is the averaging kernel matrix. (The averaging kernel is defined as the sensitivity of the inversion to the true state from *Rodgers* [2000]. In this paper, it is referred to simply as a mathematical term, and we do not seek its interpretation like smoothing functions as in the retrieval community.) This equals to $\mathrm{Tr}[(\mathbf{B} - \mathbf{P}^a)\mathbf{B}^{-1}]$, which measures the relative reduction of uncertainty for the BLUE analysis. In the presence of noise $\boldsymbol{\epsilon}$, it can be demonstrated that the DFS value ranges between 0 and the number of observations d.

[35] In summary, the DFS measures the information gain from observations to resolve the unknown parameters.

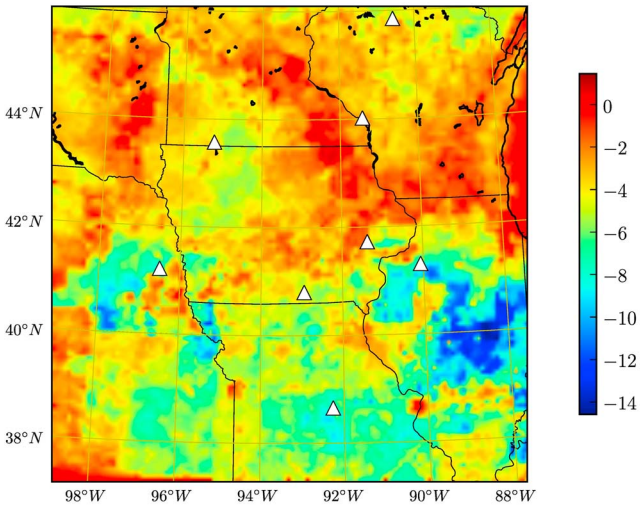

Figure 3. Reference "true" SiBcrop fluxes over the domain of the Ring 2 campaign averaged for 15 days between 1 and 16 June 2007 in units of g C m^{-2} 15 d^{-1}. The triangles indicate the location of the eight towers.

Using the property of scale-invariance of the innovation statistics, the DFS on ω can be computed by:

$$\mathcal{J}_\omega = \mathrm{Tr}\left(\mathbf{I}_N - \mathbf{P}_\omega^a \mathbf{B}_\omega^{-1}\right) = \mathrm{Tr}\left[\mathbf{\Pi}_\omega \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{H}\right], \quad (20)$$

where $\mathbf{\Pi}_\omega = \mathbf{B}\mathbf{\Gamma}_\omega^{\mathrm{T}}\left(\mathbf{\Gamma}_\omega \mathbf{B}\mathbf{\Gamma}_\omega^{\mathrm{T}}\right)^{-1}\mathbf{\Gamma}_\omega$ following equation (9). For non-diagonal $\mathbf{B}$, the DFS become

$$\mathcal{J}_\omega = \mathrm{Tr}\left[\widetilde{\mathbf{\Pi}}_\omega \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{H}\right] \quad (21)$$

$$= \mathrm{Tr}\left[\mathbf{\Pi}_\omega \mathbf{B}^{1/2}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{H}\mathbf{B}^{1/2}\right], \quad (22)$$

where $\mathbf{\Pi}_\omega$ is reduced to $\mathbf{\Gamma}_\omega^{\mathrm{T}}\left(\mathbf{\Gamma}_\omega \mathbf{\Gamma}_\omega^{\mathrm{T}}\right)^{-1}\mathbf{\Gamma}_\omega$.

[36] The DFS is to be maximized for optimal multiscale representation ω^*. In practice, $\mathbf{\Pi}_\omega$ has explicit algebraic formula for an efficient evaluation of $\mathcal{J}_\omega$.

2.7. Representation Optimization

[37] The optimization of $\mathcal{J}_\omega$ over admissible representations is a constrained optimization problem. The numerical procedure for solutions is mainly composed of three parts: the calculation of the DFS $\mathcal{J}_\omega$, a statistical regularization scheme and a gradient-based optimization routine (see *Bocquet et al.* [2011] and references therein for details). Thanks to the algebraic form of Π_ω, the efficient evaluation of $\mathcal{J}_\omega$ enables efficient optimizations less costly than the inversion at the finest scale. For representations with a larger number of grid cells, fast asymptotic solutions are possible, which for most cases last only for seconds [*Bocquet and Wu*, 2011].

[38] It is possible that the numerical solutions lead to suboptimal representations. We do not require the strict optimality. The resulting optimal representations can be validated by their improvements in the DFS value, and one can also check whether they are physically coherent.

3. Experimental Setup

[39] The Ring 2 campaign in support of the North American Carbon Program Mid Continent Intensive (MCI) will be used to test and discuss the concepts introduced earlier. The spatial domain covers an area of size 980 km × 980 km centered at [37.1906°N, 98.5925°W] (Figure 3). A ring of eight towers (one from PSU (Pennsylvania State University), five from MCI, and two from NOAA (National Oceanic and Atmospheric Administration)) around the state of Iowa collect hourly averaged CO_2 concentration observations (in ppm) in and out of the corn belt area. The locations of these towers are shown in Figure 3. The time period of the experiment is from 1st June 2007 at 0000 UTC to 16th June 2007 at 0000 UTC. The time length is 15 days. Simulations of a vegetation model SiBcrop [*Lokupitiya et al.*, 2009] within this spatiotemporal domain are used as the reference true fluxes (e.g. the fluxes during 15 days in Figure 3). The total number of observations d is thus 2880 (8 × 24 × 15).

[40] The atmospheric transport is simulated using the meteorological WRF model [*Skamarock et al.*, 2005] with a horizontal resolution of 10 km. There are 60 vertical levels. Forty of them are in the lower 2 km, and the top of the first level is 20 m. Backward particle trajectories over 15 days are generated using the Lagrangian transport model LPDM [*Uliasz*, 1994] with an integration time step of 20 s. At each time step, 10 particles are released from the tower locations. Therefore, for each hourly averaged observation, the total number of particles N_{tot} is 1800. The surface layer height h is taken to be 50 m. This height represents the atmospheric surface layer depth where the well-mixed criteria allows us to consider that particles are influenced by the surface. In theory, one should count the touchdowns at the surface, but the misrepresentation of the lowest level dynamics in the model (from 0 to 55 m in our case or the first 3 levels) may cause unrealistic footprints during stable atmospheric conditions in particular, mainly due to under-estimated vertical mixing velocities. Larger scale models use even deeper layer to compensate for this issue, which is problematic when using nighttime mixing ratios but reasonable for observations in the daytime well-mixed convective PBL.

[41] The 2D spatial domain is discretized into a finest regular grid of 128 × 128 points. In this finest reference grid, each grid cell is of size about 8 km × 8 km. The particles within these surface grid cells are recorded to compute the influences of the fluxes on concentration observations.

[42] The temporal correlations between fluxes are usually considered to be significant over days for mesoscale inversions [*Carouge et al.*, 2010; *Gourdji et al.*, 2010]. This leads to long time aggregations for inversions, e.g. over one week [*Schuh et al.*, 2010; *Lauvaux et al.*, 2011]. In this study, the mean 15-day fluxes are to be inverted. The flux dimension is the total number of grid cells: $N_{\mathrm{fg}} = 16384$.

[43] Two settings are tested for the background error covariance matrix $\mathbf{B}$: a diagonal one and a Balgovind form. We set the standard deviation κ of background flux errors to 10 g C m^{-2} 15 d^{-1}. In our case, the corn crop had not yet fully developed in June, therefore κ is set to a value smaller

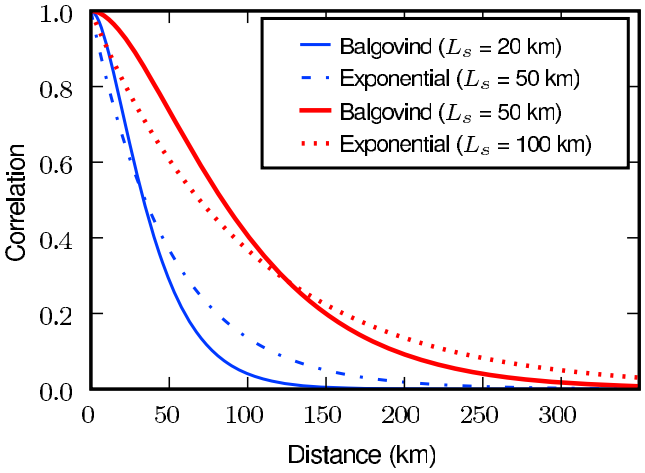

Figure 4. Correlations computed using the Balgovind and exponential correlation functions. The Balgovind correlation function is: $(1 + h_s/L_s)\exp(-h_s/L_s)$, whereas the exponential correlation function is: $\exp(-h_s/L_s)$. Here h_s is the distance and L_s is the correlation length.

than those of *Lokupitiya et al.* [2009] and *Lauvaux et al.* [2011]. Realistic correlations are assessed by testing two Balgovind correlation length (L_s) values: 20 and 50 km. This is approximately equivalent to an exponential model with correlation length set to 50 and 100 km respectively (Figure 4). *Lauvaux et al.* [2011] found little impact on inversion results with correlation length larger than 100 km for exponential models. Note that the effective degrees of freedom of the background fluxes are reduced significantly with increasing correlation lengths. Quantitative results are omitted since the degrees of freedom (information content) of background fluxes are not under the same metric as that of DFS for a fair comparison.

[44] The observational error covariance matrix **R** is assumed to be diagonal. Two values of the standard deviation of the observational error are tested: 3 and 0.5 ppm. The larger value 3 ppm is supposed to include both the atmospheric transport error and the aggregation error (that leads to the representativity error); the smaller value only accounts for the instrumental error. These observational error values are consistent with those of *Carouge et al.* [2010] and *Schuh et al.* [2010].

[45] The synthetic observations are generated by the right multiplication of the source-receptor matrix **H** with the true reference fluxes. These synthetic observations are perturbed according to the observational error. We employed 10 different seeds for the random number generation to obtain different realizations of observation perturbations. Our preliminary tests showed that there are about 2% relative variations in the root mean square errors (RMSE) for inverted fluxes resulting from this randomness. For simplicity, we present inversion results based on only one realization of observation perturbation. The background fluxes $\boldsymbol{\sigma}^b$ are generated by perturbing the true reference fluxes according to the given error structure: $\boldsymbol{\sigma}^b = \boldsymbol{\sigma}^t + \mathbf{B}^{1/2}\mathbf{n}$, with $\boldsymbol{\sigma}^t$ the true reference flux and **n** a draw from a random vector of independent normal components of standard deviation 1.

[46] In our preliminary tests, we found no significant difference when adopting a tiling structure or a qtree structure. The results obtained with the DFS criterion are also quite similar to those obtained with the criterion described by *Bocquet* [2009]. That is why we shall only present the results obtained with qtree, based on the DFS criterion.

4. Results

[47] In this section, we present the results on representation optimization and multiscale inversion. Different experimental setups are listed in Table 1, where we specify the background and observational error covariance matrices, as well as the generation of the first guess for inversions. These experiments help to demonstrate the impact of aggregation errors (Section 4.1 for regular grids and Section 4.3 for optimal grids), the information propagation from observation sites to the whole domain (Section 4.2 on optimal representations and Section 4.3 on reduction of uncertainties), and the importance of a realistic correlation length for inversions (Section 4.1 for regular grids and Section 4.3 for optimal grids). We discuss specific issues, e.g., the non-diagonal observational error covariance matrix and the inversion errors at different scales in Sections 4.4 and 4.5.

4.1. Regular Representation at Different Scales

[48] We first consider regular grids at different coarse scales, for instance: a regular grid of 64 × 64 points in which four adjacent grid cells at finest scale are aggregated into one large grid cell. It is straightforward to vary scales and perform inversions on these resulting regular grids. We list their inversion performance in Figure 5.

[49] The expectation of the root mean square error of inverted fluxes is $(\mathrm{E}[(\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t)^{\mathrm{T}}(\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t)]/N_{\mathrm{fg}})^{1/2} = [\mathrm{Tr}(\mathbf{P}^a)/N_{\mathrm{fg}}]^{1/2}$. Similarly, the expected RMSE of the first guess is $[\mathrm{Tr}(\mathbf{B})/N_{\mathrm{fg}}]^{1/2}$. In general, the inverted fluxes have smaller RMSEs than those of the first guesses, since the information from concentration observations are assimilated. Note that the DFS that measures the information gain from observations is evaluated by $\mathrm{Tr}[(\mathbf{B} - \mathbf{P}^a)\mathbf{B}^{-1}]$, whereas the RMSE of inverted fluxes measures the residual uncertainty after analysis.

[50] For a diagonal **B**, the improvement in RMSE is not significant (Figures 5a and 5c). However, the corresponding

Table 1. Different Experimental Setups[a]

Case	B	R	Perturbation
BD-R3-PD	diagonal	std = 3 ppm	diagonal **B**
BD-R0.5-PD	diagonal	std = 0.5 ppm	diagonal **B**
B20-R3-P20	Balgovind, L_s = 20 km	std = 3 ppm	Balgovind **B**
B20-R0.5-P20	Balgovind, L_s = 20 km	std = 0.5 ppm	Balgovind **B**
B50-R3-P50	Balgovind, L_s = 50 km	std = 3 ppm	Balgovind **B**
B50-R0.5-P50	Balgovind, L_s = 50 km	std = 0.5 ppm	Balgovind **B**
B20-R3-PD	Balgovind, L_s = 20 km	std = 3 ppm	diagonal **B**
B20-R0.5-PD	Balgovind, L_s = 20 km	std = 0.5 ppm	diagonal **B**
B20-R3-P50	Balgovind, L_s = 20 km	std = 3 ppm	Balgovind **B**
B50-R3-P20	Balgovind, L_s = 50 km	std = 3 ppm	Balgovind **B**

[a]We detail the configurations for the background and observational error covariance matrices (**B** and **R**), as well as the covariance matrices of the zero-mean Gaussian perturbation vectors added to the true reference fluxes to generate the first guesses. For all the settings, the standard deviations (std) of background errors are set to 10 g C m^{-2} 15 d^{-1}, and **R** is assumed to be diagonal.

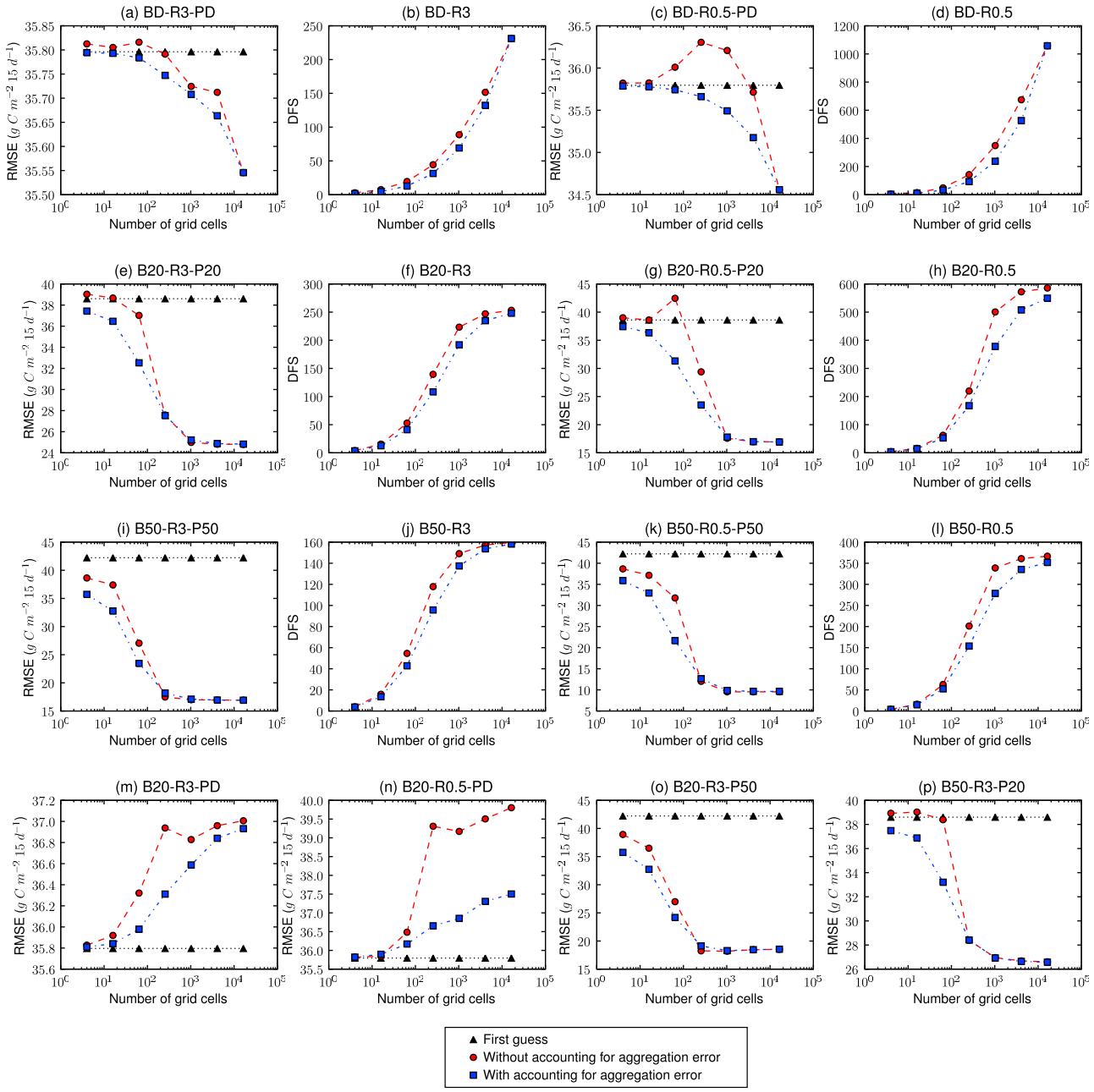

Figure 5. Performance of inversions with regular grids at different scales. The x-axis shows the number of grid points for each regular grid. The y-axes are either the DFS or the RMSE of inverted fluxes. The observational error is set to 3 and 0.5 ppm respectively. The background error covariance matrix and the generation of first guess are defined in Table 1. The triangles indicate the RMSE of the background fluxes (first guess) to serve as a baseline; the circles present the inversion performance for the cases in which the aggregation error is not considered (inconsistent innovation statistics $\mathbf{R} + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}$ is used); whereas the squares show the performance when the aggregation error is formulated explicitly for inversions using scale-independent innovation statistics $\mathbf{R} + \mathbf{HBH}^{\mathrm{T}}$.

DFSs are considerable (e.g. DFS/d ~ 8% in Figure 5b and DFS/d ~ 37% in Figure 5d for the finest grid). Since there are no correlations in the background errors, the number of degrees of freedom of the system is quite large. In addition, the information is not spread to the domain away from the observation sites in the absence of correlations in background errors. The DFS values indicate that the observations are effectively assimilated locally. Nevertheless, there are still many unresolved degrees of freedom of system. This explains the weak improvement in RMSE.

[51] For **B** in Balgovind form, the correlations of the background errors play a role similar to the aggregation of variables, but decaying with distance. The number of degrees of freedom of the system is therefore decreased, and the information from observations is more propagated within the domain. Consequently, one observes better improve-

ments in RMSE (e.g., Figure 5e). The best improvements in RMSE are obtained with the longest length in error correlations (e.g. Figure 5i). In this case, we have the smallest number of unresolved degrees of freedom of system. When the standard deviation of the observational error is set to the realistic value 3 ppm, the DFSs for Balgovind **B** (DFS/d ~ 9% in Figure 5f and DFS/d ~ 5% in Figure 5j) are comparable with that for the diagonal case (Figure 5b). When the ratio between the background and observational errors increases, the DFSs for Balgovind **B** (DFS/d ~ 19% for Figure 5h and DFS/d ~ 12% for Figure 5l) are inferior to that for the diagonal case (Figure 5d).

[52] In practice, the error structure of **B** is unknown and difficult to parameterize. We simulate this fact by deliberately using misspecified error structures to perturb the reference fluxes in order to obtain the first guess $\boldsymbol{\sigma}^b$. The RMSE is affected by this mis-specification, but the DFS do not depend on the perturbation error structure. The theoretical improvement in RMSE of the BLUE analysis is not guaranteed with perturbations that break the assumption on background error structure. For instance, if in reality **B** bears little correlation structures, and if the inversion is performed with a **B** with a correlation length of 20 km, we have catastrophic increasing RMSE for inverted fluxes (Figures 5m and 5n). The imposed aggregations on independent regions lead to dominant aggregation errors. The corresponding inversions may diverge and produce spurious and large a posteriori error covariance matrix $\mathbf{P}^a$. The mis-specification of a 20 km correlation length by 50 km or vice versa leads to poorer inversions (Figures 5o and 5p compared with Figures 5i and 5e respectively).

[53] The impact resulting from explicitly formulating the aggregation error for inversion is also shown in Figure 5. The inversions at the finest scale are the best, because the aggregation error $\boldsymbol{\epsilon}_\omega^{\text{agg}}$ is null at that scale. Inversions taking into account aggregation errors perform systematically better than those without consideration of aggregation errors, especially when **B** is not physically realistic (Figures 5m–5p) or when the ratio between the background and observational errors increases (e.g. Figures 5c, 5g and 5k). This latter case may play important roles in practice, since our background error has been underestimated. It is also noticeable that the aggregation of variables with diagonal **B** also generates considerable aggregation errors. Based on these results, it is recommended that, if possible, the aggregation error should be considered explicitly for inversions using the scale-independent innovation statistics $\mathbf{R} + \mathbf{HBH}^{\mathrm{T}}$ rather than the inconsistent term $\mathbf{R} + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^{\mathrm{T}}$.

4.2. Optimal Representations

[54] We show optimal multiscale representations under the DFS criterion in Figure 6. The number of grid cells in these representations are far fewer than that of the finest scale N_{fg}. Equally important, the multiscale representations are optimal in the sense that the gain of information from observations is maximized. In other words, the optimal representation characterizes how information from observations are propagated spatiotemporally within the domain.

[55] For diagonal **B**, there are no correlations in background errors introducing the aggregation effect. However, the information from μ can still be spread to the regions around observation sites by advection and diffusion through the source-receptor matrix **H** in the innovation vector $\boldsymbol{\mu} - \mathbf{H}\boldsymbol{\sigma}^b$. As a result, the adaptive optimal grids are dense around the observation sites, with shapes following the wind conditions (leftmost column in Figure 6). Since little information from observations can be used to resolve the fluxes in regions distant from the observation sites, the optimal grids are sparse in these distant regions.

[56] By contrast, for **B** in Balgovind form, the information from observations can be propagated to distant regions due to the aggregation effect introduced by the correlation in background errors. The adaptive optimal grids are more uniformly distributed than those for diagonal **B** (right two columns in Figure 6). The longer the correlation length is, the more uniform the optimal representation becomes. The influence of the variations in meteorological conditions seems to be smoothed by the correlations in background errors. Nevertheless, the optimal representations are still dense around the observation sites (e.g. Figure 6h), which results from the balance between the meteorological conditions and the introduced aggregation effect [*Saide et al.*, 2011].

[57] It has been found by preliminary tests by *Lauvaux et al.* [2011] that the spatial correlation structure for crops is rather short. Inversion results with correlation lengths longer than 100 km are very similar. Further studies on the exact optimal correlation length are needed, e.g., an objective analysis using in-site flux observations and/or cross validations using observation sets from different towers.

4.3. Inversion With Multiscale Representations

[58] Once the multiscale representation is fixed, inversions can be performed as described in Section 2.5. For both regular and optimal multiscale grids, we perform diverse inversions with **B** chosen to be either diagonal or in Balgovind form. The inversion results are shown in Figures 7 and 8. It is verified that, for diagonal **B**, the gain through inversions peaks around the observation sites (Figure 8a), and the information from observations cannot be conveyed to distant regions. The corrections of inverted fluxes to background fluxes are small, especially in regions distant from observation sites (Figure 8b).

[59] By contrast, for Balgovind **B** even with spatial correlations (L_s = 20 and 50 km respectively), the gain through inversions can be spread to regions far from the observation sites (Figures 8e and 8k). Considerable corrections are obtained, and there are significant improvements in RMSE for the inverted fluxes (Figures 8g and 8l). In these cases, the first guesses are generated by perturbing the true reference fluxes with a zero-mean Gaussian vector whose covariance matrix is taken to be the same Balgovind **B**. In other words, **B** is assumed to be well tuned.

[60] The relative gain of variance at i-th grid cell is computed by: $([\mathbf{B}]_{ii} - [\mathbf{P}^a]_{ii})/[\mathbf{B}]_{ii}$, where $[\mathbf{B}]_{ii}$ denotes the i-th diagonal element of **B**. For the cases with non-diagonal **B**, the sum of relative gains over all the grid cells is bigger than the DFS, because the gains are expected to be correlated. By contrast, the sum of relative gains of variance for diagonal **B** equals to the DFS.

[61] For diagonal **B**, since little gain is found in regions far from the observation sites, there is no need to allocate computational resources in these regions. Therefore efficient optimal multiscale representations can be obtained against the regular grids. Accordingly, one observes large improve-

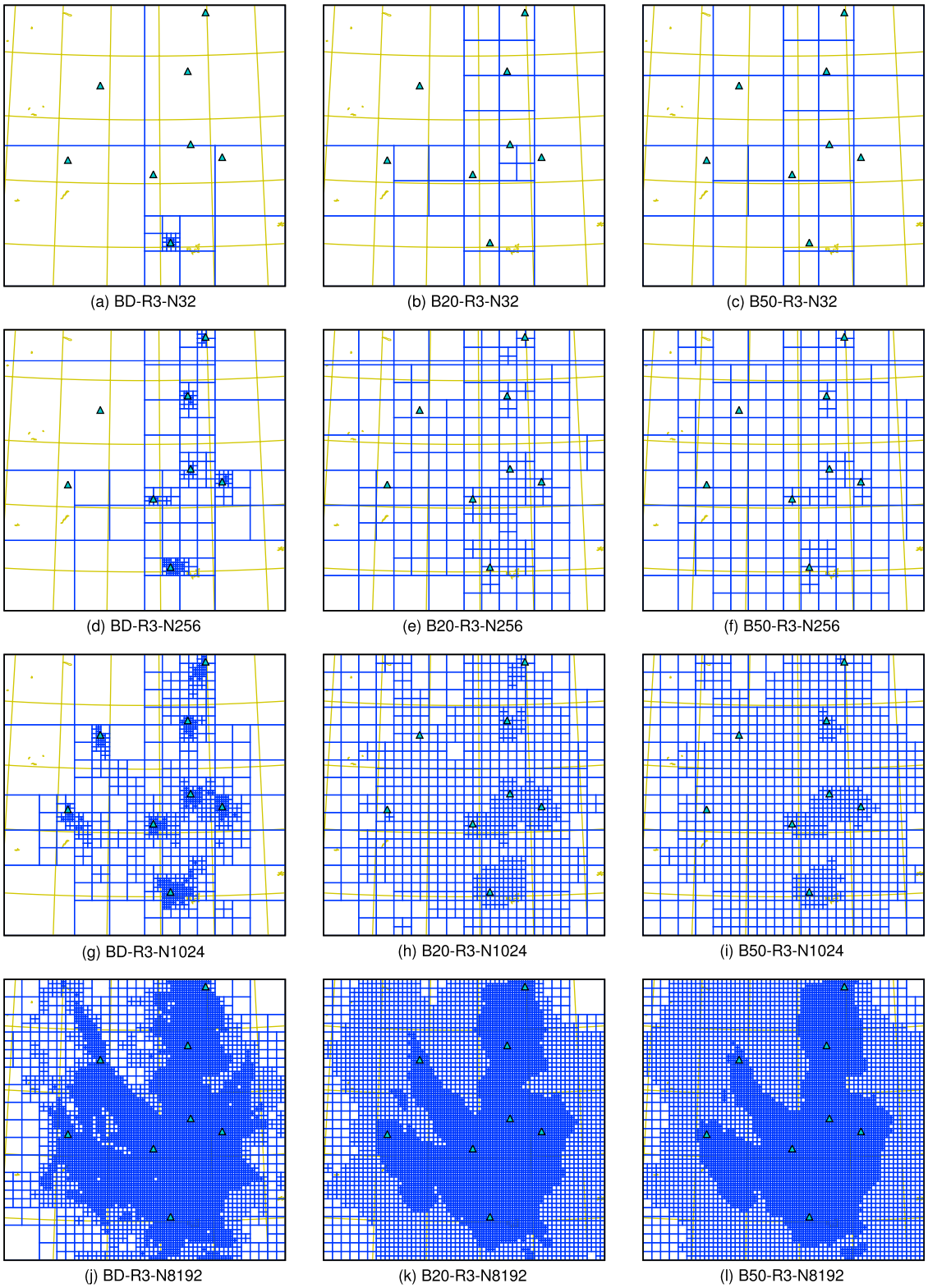

Figure 6. Optimal multiscale representations under the DFS criterion. The background and observational error covariance matrices are defined in Table 1. The total targeted number of grid cells is identical for the three optimal representations at each row: $N = 32$, 256, 1024, 8194 from top to bottom.

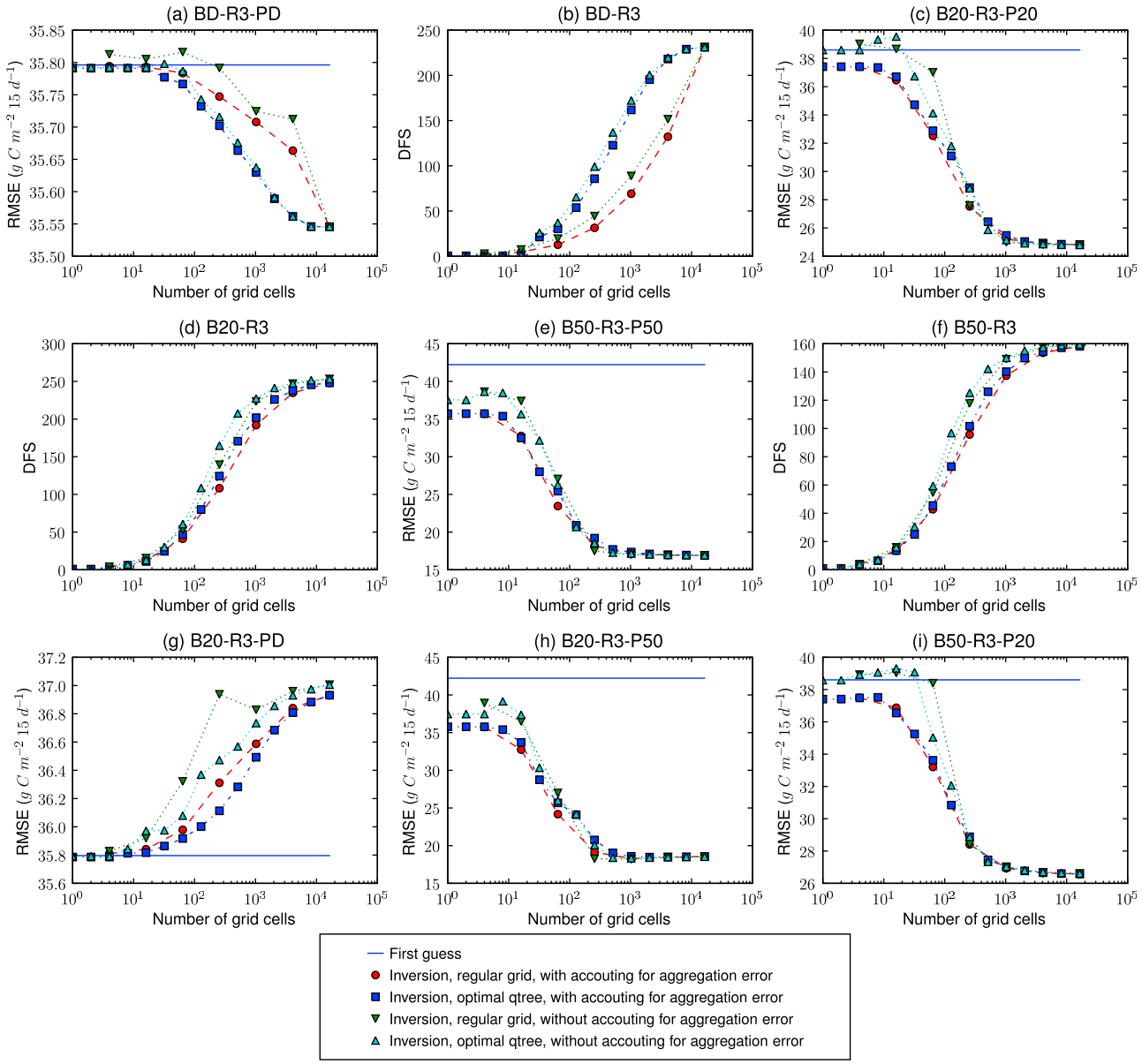

Figure 7. Performance of inversions for optimal multiscale grids with different total number of grid cells. The squares present these inversion results. Performance of inversions for regular grids are plotted as circles for comparison. The x-axis shows the total number of grid cells for multiscale grids. The y-axes are either the DFS or the RMSE of inverted fluxes. The generation of first guess and the background and observational error covariance matrices are defined in Table 1.

ment in DFS for optimal representations against regular representations in Figure 7b. Compared with the grid at finest scale, an optimal representation may have only 25% of the total number of grid cells, but can keep 94% of the DFS value (Figure 7b) when the aggregation error is taken into account explicitly. The correction of the inverted fluxes to the background fluxes for optimal representations (Figure 8d) is similar to that for the finest grid (Figure 8b) and much better than that for regular representations of the same size (Figure 8c).

[62] For Balgovind **B**, the corrections of inverted fluxes to background fluxes for the regular and optimal representations are very similar (Figures 7d and 8g). This is not a surprise, since the optimal representations are more uniformly distributed in this case. The influence of observations transported by the meteorological conditions is smoothed by the imposed correlations in background errors. The optimal representation may have only 25% of the total number of grid cells, but can keep at least 96% of the DFS value (Figures 7d and 7f) when the aggregation error is taken into account explicitly.

[63] Without accounting for the aggregation error, inversions with optimal adaptive grids perform systematically better than those with regular grids (Figure 7). The optimal

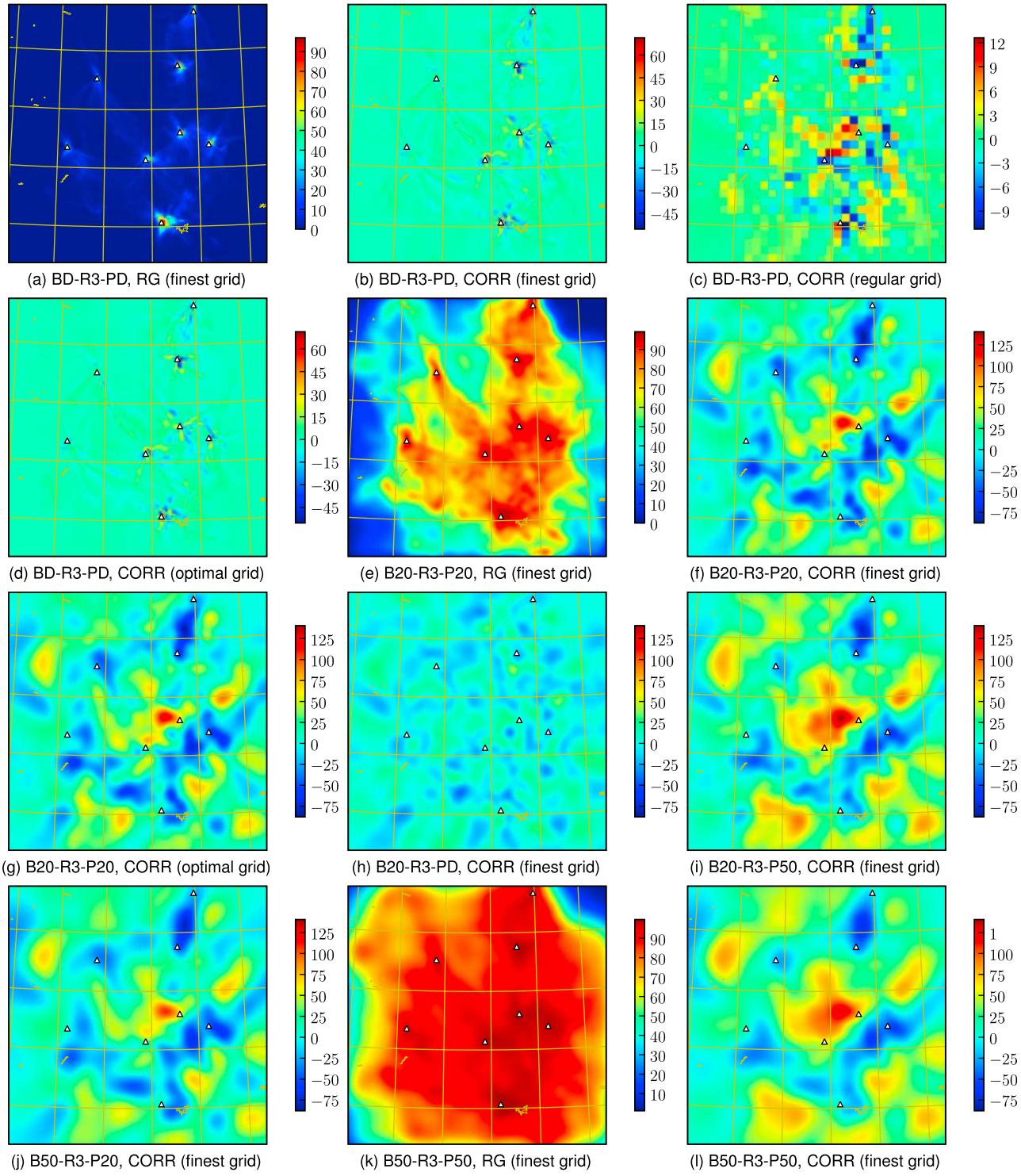

Figure 8. Inversion results for the regular and optimal multiscale grids. The generation of first guess and the background and observational error covariance matrices are defined in Table 1. The total number of grid cells is identical for the regular and optimal grids. For each grid cell, the relative gain (RG) of variance is computed as $(v^b - v^a)/v^b \times 100\%$, where v^b, v^a are the variances of the errors of the background and inverted fluxes respectively. The corrections (CORR) are the differences between background and inverted fluxes (in units of g C m^{-2} 15 d^{-1}).

grids seem to be less affected by the aggregation effect. In fact, the aggregation effect can be quantified by:

$$\widehat{\mathcal{J}}_\omega = \mathrm{Tr}\left[\mathbf{R}^{-1}(\mathbf{R}_\omega - \mathbf{R})\right] = \mathrm{Tr}\left(\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right) - \mathrm{Tr}\left(\mathbf{\Pi}_\omega\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right). \quad (23)$$

Here $\mathbf{R}_\omega$ is given by equation (13). To minimize the aggregation effect is hence equivalent to the maximization of the Fisher criterion $\mathrm{Tr}(\mathbf{\Pi}_\omega\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})$. *Bocquet et al.* [2011] gave more details on the Fisher criterion, which is the limiting case of the DFS criterion when **R** is inflated or when **B** vanishes. Our preliminary tests showed that the optimal grids are quite similar under the Fisher and DFS criteria (results omitted). In summary, the optimal grids seem to have the desirable property that mitigates the aggregation effect.

[64] Most of the CO_2 inversions [*Rödenbeck et al.*, 2003; *Michalak et al.*, 2004; *Peylin et al.*, 2005] rely on lengthy correlations in background errors for effective corrections of fluxes. However, such lengthy correlations may not be realistic as indicated by *Chevallier et al.* [2006]. Indeed, when the first guess is perturbed differently with a diagonal **B**, the correction (Figure 8h) becomes dramatically different from that with the perturbation using correct Balgovind **B** (Figure 8f). The RMSE of inverted fluxes can even be worse than the background RMSE (Figure 5m). This result is consistent with the investigations of *Gerbig et al.* [2006] and *Carouge et al.* [2010] using regular grids. It is clear from our study that the choice of such an unphysical correlation length is likely to yield a significant over-estimation of the inversion gain.

[65] Suppose that the realistic **B** has a short correlation length, and that the inversion is performed using a **B** with longer correlation length, the resulting correction (Figure 8j) tends to omit the small scale variations (compared with Figure 8f). In reverse, underestimation of the correlation length of **B** (Figure 8i) will improperly impose small scale variations in the corrections (compared with Figure 8l). Therefore, the specification of a realistic and robust **B** is a crucial problem in top-down CO_2 inversions that needs further in-depth investigations.

4.4. On Correlated R

[66] In this study using synthetic data, the model transport errors are not formulated explicitly, and the observational error covariance matrix **R** is assumed to be diagonal. Lacking of in-depth investigations on model errors, most recent regional inversions adopt a diagonal **R** [*Carouge et al.*, 2010; *Gourdji et al.*, 2010; *Schuh et al.*, 2010].

[67] In practice, **R** is probably non-diagonal if the transport model errors are considered. The transport errors due to the imprecision in horizontal transport, vertical mixing, and sub-grid physics may persist across hours, which probably leads to spatiotemporal correlations in **R**. In this regard, *Gerbig et al.* [2003] used exponential decaying models for a parameterization of spatial correlations. Based on an ensemble of atmospheric transports [*Lauvaux et al.*, 2009a], *Lauvaux et al.* [2011] introduced temporal correlations within 12 hours in the observational error to account for its impact on inversions. In general, such observational spatiotemporal correlations would make the inversions less constrained by the observations [*Lauvaux et al.*, 2011].

[68] One may roughly assess the effect of observational spatiotemporal correlations by artificially inflating the error variances of a diagonal **R**. Recall that the Fisher criterion is the limiting case of the DFS criterion with larger error variances in **R**, and the optimal grids under the two criteria were found to be stable. We conjecture that the optimal grids are less influenced by the introduction of observational spatiotemporal correlations than the inversion on regular grids.

4.5. Inversion Errors at Different Scales

[69] All the errors (RMSEs related to the Euclidean norm $\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|_2$) in Figures 5 and 7 are compared at the finest scale. This provides an equal basis to compare all coarse-scale inversions, which helps to assess the aggregation errors. However, when inversions are performed on coarser representation ω, it would be interesting to compare the inversion errors at the same coarse scale, that is, some quantities related to $\|\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega\|$.

[70] 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. In fact, substituting equations (16) or (17) into $\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|^2_{\mathbf{B}^{-1}}$, we have

$$\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|^2_{\mathbf{B}^{-1}} = \left\|\left(\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega\right)\left(\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t\right) + \mathbf{\Lambda}^*_\omega\boldsymbol{\sigma}^a_\omega - \mathbf{\Pi}_\omega\boldsymbol{\sigma}^t\right\|^2_{\mathbf{B}^{-1}} \quad (24)$$

$$= \left\|\left(\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega\right)\left(\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t\right) + \mathbf{\Lambda}^*_\omega\left(\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega\right)\right\|^2_{\mathbf{B}^{-1}}. \quad (25)$$

Let $\gamma^f = (\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega)(\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t)$ and let $\gamma^c = \mathbf{\Lambda}^*_\omega\,(\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega)$. One can verify that γ^f and γ^c are $\mathbf{B}^{-1}$-orthogonal, that is, $\gamma^{f,\mathrm{T}}\mathbf{B}^{-1}\gamma^c = 0$. Therefore we have

$$\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|^2_{\mathbf{B}^{-1}} = \|\gamma^f\|^2_{\mathbf{B}^{-1}} + \|\gamma^c\|^2_{\mathbf{B}^{-1}}. \quad (26)$$

Here γ^f characterizes the variability at the finest scale, and γ^c is related to the variation $\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega$ at coarse scales.

[71] We compare the $\mathbf{B}^{-1}$-norm of $\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t$, $\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t$, γ^f and γ^c in Figure 9. The inversion errors are normalized by **B**, consequently the improvements in $\mathbf{B}^{-1}$-norm are less significant (especially for the case of diagonal **B**) than those in RMSE in Figures 5 and 7. When the numbers of grid cell is small, the inversion error is mainly due to the failure of coarser representations in accounting for the variability γ^f at the finest scale. When increasing the number of grid cells, γ^f decreases since finer grid cells are used, and γ^c increases because there are more grid cells to be compared for the term $\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega$.

[72] In general, the inversion error for optimal grids has a greater γ^f but smaller γ^c and $\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t$ compared with those for regular grids with the same total number of grid cells. The sum of $\|\gamma^f\|^2_{\mathbf{B}^{-1}}$ and $\|\gamma^c\|^2_{\mathbf{B}^{-1}}$ is slightly greater than $\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|^2_{\mathbf{B}^{-1}}$, because the approximation of the square root of **B** in inversions slightly breaks the $\mathbf{B}^{-1}$-orthogonality between γ^f and γ^c. In summary, the RMSEs in Figures 5 and 7 are related to $\|\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t\|_{\mathbf{B}^{-1}}$, which is the composition of the variations in errors at the finest and coarser scales. Therefore they are reasonable criteria to evaluate the inversion per-

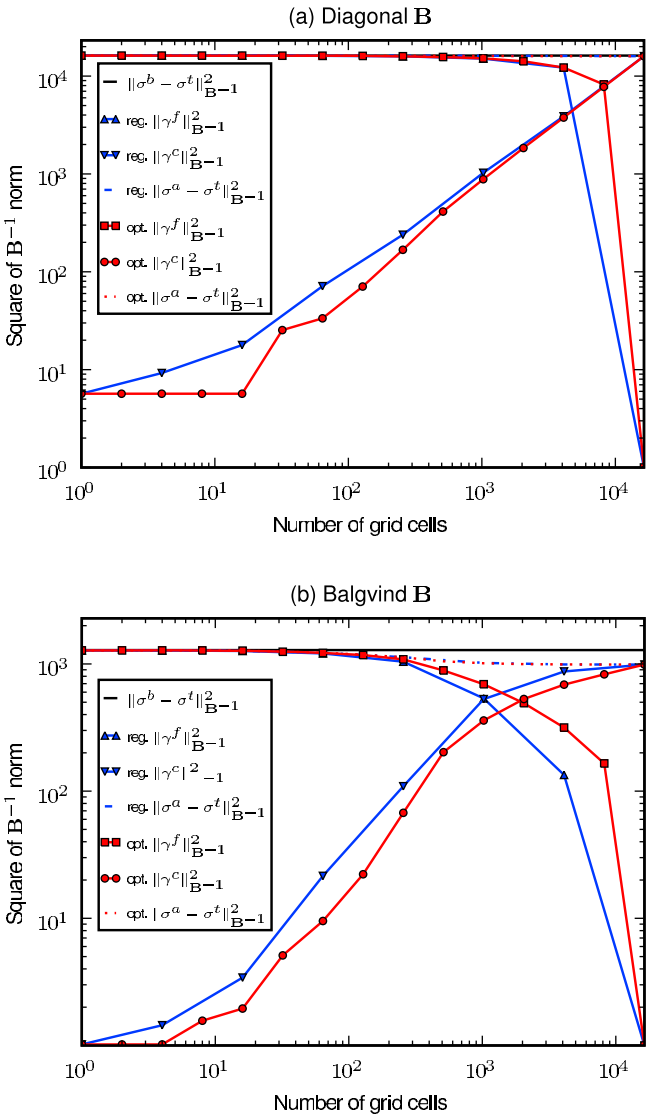

Figure 9. B^{-1}-norm of $\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t$, $\boldsymbol{\sigma}^a - \boldsymbol{\sigma}^t$, γ^f and γ^c. Here $\gamma^f = (\mathbf{I}_{N_{fg}} - \boldsymbol{\Pi}_\omega)\ (\boldsymbol{\sigma}^b - \boldsymbol{\sigma}^t)$ characterizes the variability at the finest scale, and $\gamma^c = \boldsymbol{\Lambda}^*_\omega\ (\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega)$ is related to coarser variations. The correlation length for the Balgovind case is set to 20 km. Both regular (reg.) and optimal (opt.) grids are tested.

formances. Although it is not possible to compute $\boldsymbol{\sigma}^a_\omega - \boldsymbol{\sigma}^t_\omega$ directly through those RMSEs, one can assess the variations at coarse scales γ^c which typically follow the curves plotted in Figure 9. For the case of diagonal **B**, the optimization algorithm is less efficient and produces zigzag for very small number of grid cells in Figure 9a.

5. Conclusion

[73] We have implemented a consistent Bayesian formalism to construct optimal multiscale space-time representation of carbon fluxes for mesoscale inversions under an information criterion: the number of degrees of freedom for the signal (DFS). This methodology has been tested using synthetic CO_2 concentration data in the context of the Ring 2 experiment in support of the North American Carbon Program Mid Continent Intensive.

[74] The DFS, ranging from 0 to the number of observations d, measures the information gain from observations to resolve the unknown fluxes. It has been found that, for continuous hourly concentration observations from the Ring 2 network of eight towers with realistic observational errors, in general only a small part of observations are effectively assimilated (DFS/d < 20%). By contrast, the root mean square errors (RMSE) of inverted fluxes measure the residual uncertainty after analysis.

[75] In the absence of correlations in the errors, there are many more degrees of freedom of the system to be resolved compared to the relatively small DFS value. Therefore, one observes large residual uncertainties (small improvements of RMSE). In general, aggregation of flux variables or introducing correlations in background errors are needed to reduce the effective degrees of freedom of system. In this multiscale setting, we have formulated the scale-dependent aggregation errors explicitly. The inversion performances have been systematically improved by explicitly taking into account the aggregation error for inversions. The multiscale representation of carbon fluxes allow to adaptively mitigate the aggregation errors. This is justified by the better performance of the optimal grids against the regular grids in the inversions without accounting for the aggregation error.

[76] The optimal multiscale representations have been found to have far fewer grid cells (e.g. 25% of the total grid cells), but can keep most of the DFS value (e.g. 94%). This enables more efficient inversions, since far fewer flux variables are to be inverted. Equally important, the optimal multiscale representations characterize how information from observations can be optimally spread to the whole domain. For instance, in the absence of correlations in background errors, the information from observations has little impact on regions distant from observation sites. Consequently the optimal representations are dense around observation sites but very sparse in distant regions. By contrast, when correlations in background errors are introduced, the optimal multiscale representations are more uniformly distributed, which results from the balance between atmospheric transport and the imposed aggregation effect.

[77] The correlations in background errors have been shown to be crucial for carbon inversions. Failure in the specification of realistic correlations leads to significant aggregation errors. In this case, scale-dependent aggregation errors should be formulated explicitly for more reliable carbon inversions.

[78] In this study, transport errors were not taken into account in the formulation of the representativity error. With finer spatial scales, the aggregation errors are supposed to decrease since fewer regions are aggregated. Nevertheless the smoothing of carbon fluxes by the complex mesoscale atmospheric transport makes it more difficult to retrieve flux variations from concentration observations. Therefore the estimation errors are expected to increase [*Peylin et al.*, 2001; *Enting*, 2002]. When considering scale-dependent transport errors, the balance between the aggregation and estimation errors may result in an optimal multiscale representation with an inversion performance better than that of the regular grid at finest scale [*Bocquet et al.*, 2011]. Moreover, the transport model error probably arouses spatiotemporal

correlations in the observational error. We will examine its impact on the representation optimization in details in the subsequent studies.

[79] Further in-depth investigations on realistic correlations in background errors are needed for reliable inversions. More sources of information may be exploited, e.g. the carbon flux observations, simulations of diverse vegetation models, and different correlation parameterizations. Non-stationary correlations may be helpful to incorporate additional constraints (e.g. ecosystem information). Clues on correlations may also be inferred through leave-one-out cross validations, in which we compare the concentration observations from one excluded tower and the simulated concentrations for this tower using observations from other towers (see *Pickett-Heaps et al.* [2011] for an example of validation studies).

[80] This study will be extended to real concentration observations. This means that the boundary and initial conditions will also be included in the inversion. The methodology for multiscale analysis in this paper is not limited to spatial aggregations. We can deal with multiscale temporal aggregations within the same framework. Inversions with finer time scales will be a future subject, which needs temporal correlations in background errors [*Gourdji et al.*, 2010]. The optimal multiscale representations depend on the meteorological scenarios. Experiments during a longer time period, e.g. several months, could be performed. Statistics on the resulting optimal representations will help to set up fixed multiscale grids for practical carbon flux inversions.

[81] **Acknowledgments.** This paper is a contribution to the MSDAG project supported by the Agence Nationale de la Recherche, grant ANR-08-SYSC-014. The authors would like to thank the reviewers for their substantial suggestions that helped to improve the manuscript.

References

Balgovind, R., A. Dalcher, M. Ghil, and E. Kalnay (1983), A stochastic-dynamic model for the spatial structure of forecast error statistics, *Mon. Weather Rev.*, *111*, 701–722.

Bocquet, M. (2009), Towards optimal choices of control space representation for geophysical data assimilation, *Mon. Weather Rev.*, *137*, 2331–2348.

Bocquet, M., and L. Wu (2011), Bayesian design of control space for optimal assimilation of observations. II: Asymptotic solutions, *Q. J. R. Meteorol. Soc.*, *137*, 1357–1368.

Bocquet, M., L. Wu, and F. Chevallier (2011), Bayesian design of control space for optimal assimilation of observations. I: Consistent multiscale formalism, *Q. J. R. Meteorol. Soc.*, *137*, 1340–1356.

Bousquet, P., P. Peylin, P. Ciais, C. L. Quéré, P. Friedlingstein, and P. P. Tans (2000), Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, *290*(5495), 1342–1346.

Carouge, C., P. J. Rayner, P. Peylin, P. Bousquet, F. Chevallier, and P. Ciais (2010) What can we learn from European continuous atmospheric CO_2 measurements to quantify regional fluxes – Part 2: Sensitivity of flux accuracy to inverse setup, *Atmos. Chem. Phys.*, *10*(6), 3119–3129.

Chevallier, F., N. Viovy, M. Reichstein, and P. Ciais (2006), On the assignment of prior errors in Bayesian inversions of CO_2 surface fluxes, *Geophys. Res. Lett.*, *33*, L13802, doi:10.1029/2006GL026496.

Chevallier, F., F.-M. Bréon, and P. J. Rayner (2007), Contribution of the orbiting carbon observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, *J. Geophys. Res.*, *112*, D09307, doi:10.1029/2006JD007375.

Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, Cambridge Univ. Press, Cambridge, U. K.

Fan, S., M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento, T. Takahashi, and P. Tans (1998), A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models, *Science*, *282*(5388), 442–446.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003), Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, *108*(D24), 4757, doi:10.1029/2003JD003770.

Gerbig, C., J. C. Lin, J. W. Munger, and S. C. Wofsy (2006), What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, *Atmos. Chem. Phys.*, *6*(2), 539–554.

Gourdji, S. M., A. I. Hirsch, K. L. Mueller, A. E. Andrews, and A. M. Michalak (2010), Regional-scale geostatistical inverse modeling of North American CO_2 fluxes: A synthetic data study, *Atmos. Chem. Phys.*, *10*, 6151–6167.

Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, *106*(D5), 4703–4715.

Lauvaux, T., et al. (2008), Mesoscale inversion: first results from the ceres campaign with synthetic data, *Atmos. Chem. Phys.*, *8*(13), 3459–3471.

Lauvaux, T., O. Pannekoucke, C. Sarrat, F. Chevallier, P. Ciais, J. Noilhan, and P. J. Rayner (2009a), Structure of the transport uncertainty in mesoscale inversions of CO_2 sources and sinks using ensemble model simulations, *Biogeosciences*, *6*(6), 1089–1102.

Lauvaux, T., et al. (2009b), Bridging the gap between atmospheric concentrations and local ecosystem measurements, *Geophys. Res. Lett.*, *36*, L19809, doi:10.1029/2009GL039574.

Lauvaux, T., A. E. Schuh, M. Uliasz, S. Richardson, N. Miles, L. I. Diaz, D. Martins, P. Shepson, and K. J. Davis (2011), Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, *Atmos. Chem. Phys. Discuss.*, *11*, 20,855–20,898.

Law, R. M., P. J. Rayner, L. P. Steele, and I. G. Enting (2003), Data and modelling requirements for CO_2 inversions using high-frequency data, *Tellus, Ser. B*, *55*(2), 512–521.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger (2003), A near-field tool for simulating the upstream influence of atmospheric observations: The stochastic time-inverted Lagrangian transport (STILT) model, *J. Geophys. Res.*, *108*(D16), 4493, doi:10.1029/2002JD003161.

Lokupitiya, E., et al. (2009), Incorporation of crop phenology in simple biosphere model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands, *Biogeosciences*, *6*, 969–986.

Michalak, A., L. Bruhwiler, and P. Tans (2004), A geostatistical approach to surface flux estimation of atmospheric trace gases, *J. Geophys. Res.*, *109*, D14109, doi:10.1029/2003JD004422.

Peters, W., et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: Carbontracker, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(48), 18,925–18,930.

Peylin, P., P. Bousquet, and P. Ciais (2001), Inverse modeling of atmospheric carbon dioxide fluxes - Response, *Science*, *294*(5550), 2292–2292.

Peylin, P., P. J. Rayner, P. Bousquet, C. Carouge, F. Hourdin, P. Heinrich, P. Ciais, and AEROCARB contributors (2005), Daily CO_2 flux estimates over Europe from continuous atmospheric measurements: 1. Inverse methodology, *Atmos. Chem. Phys.*, *5*(12), 3173–3186.

Pickett-Heaps, C. A., et al. (2011), Atmospheric CO_2 inversion validation using vertical profile measurements: Analysis of four independent inversion models, *J. Geophys. Res.*, *116*, D12305, doi:10.1029/2010JD014887.

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, *3*, 1919–1964.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding*, World Sci., Hoboken, N. J.

Saide, P., M. Bocquet, A. Osses, and L. Gallardo (2011), Constraining surface emissions of air pollutants using inverse modeling: method intercomparison and a new two-step two-scale regularization approach, *Tellus, Ser. B*, *63*, 360–370.

Schuh, A. E., A. S. Denning, K. D. Corbin, I. T. Baker, M. Uliasz, N. Parazoo, A. E. Andrews, and D. E. J. Worthy (2010), A regional high-resolution carbon flux inversion of North America for 2004, *Biogeosciences*, *7*(5), 1625–1644.

Seibert, P., and A. Frank (2004), Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, *Atmos. Chem. Phys.*, *4*(1), 51–63.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers (2005), A description of the advanced research WRF version 2, *Tech. Rep. 468+STR*, NCAR, Boulder, Colo.

Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the global atmospheric carbon dioxide budget, *Science*, *247*(4949), 1431–1438.

Tolk, L. F., A. G. C. A. Meesters, A. J. Dolman, and W. Peters (2008), Modelling representation errors of atmospheric CO_2 mixing ratios at a regional scale, *Atmos. Chem. Phys.*, *8*(22), 6587–6596.

Uliasz, M. (1994), Lagrangian particle dispersion modeling in mesoscale applications, in *Environmental Modeling II*, edited by P. Zannetti, pp. 71–102, Comput. Mech., Southampton, U. K.

M. Bocquet and L. Wu, CEREA, Joint Laboratory École des Ponts ParisTech - EDF R&D, 6-8 Avenue Blaise Pascal, Cité Descartes Champs-sur-Marne, F-77455 Marne la Vallé CEDEX 2, France. (marc.bocquet@cerea.enpc.fr; lin.wu@cerea.enpc.fr)

F. Chevallier, Laboratoire des Sciences du Climat et de l'Environnement, CEA-Orme des Merisiers, F-91191 Gif-Sur-Yvette CEDEX, France. (frederic.chevallier@lsce.ipsl.fr)

K. Davis and T. Lauvaux, Department of Meteorology, Pennsylvania State University, 412A Walker Bldg., University Park, PA 16802, USA. (davis@meteo.psu.edu; tul5@meteo.psu.edu)

P. Rayner, School of Earth Sciences, University of Melbourne, Melbourne, VIC 3010, Australia. (prayner@unimelb.edu.au)