Carbon flux bias estimation employing Maximum Likelihood Ensemble Filter (MLEF)

Dusanka Zupanski

Colorado State University, Cooperative Institute for Research in the Atmosphere,
Fort Collins, Colorado, USA

A. Scott Denning and **Marek Uliasz**

Colorado State University, Department of Atmospheric Science,
Fort Collins, Colorado, USA

Milija Zupanski

Colorado State University, Cooperative Institute for Research in the Atmosphere,
Fort Collins, Colorado, USA

Andrew E. Schuh

Colorado State University, Department of Atmospheric Science,
Fort Collins, Colorado, USA

Peter J. Rayner

Laboratoire des sciences du Climat et de l'Environnement,
Gif-sur-Yvette, France

Wouter Peters

NOAA Earth System Research Laboratory,
Boulder, Colorado, USA

Katherine D. Corbin

Colorado State University, Department of Atmospheric Science,
Fort Collins, Colorado, USA

Abstract

[1]   We evaluate the capability of an ensemble based data assimilation approach, referred to as Maximum Likelihood Ensemble Filter (MLEF), to estimate biases in the CO_2 photosynthesis and respiration fluxes. We employ an off-line Lagrangian Particle Dispersion Model (LPDM), which is driven by the carbon fluxes, obtained from the Simple Biosphere - Regional Atmospheric Modeling System (SiB-RAMS). The SiB-RAMS carbon fluxes are assumed to have errors in the form of multiplicative biases. Our goal is to estimate and reduce these biases and also to assign reliable posterior uncertainties to the estimated biases. Experiments of this study are performed using simulated CO_2 observations, which resemble real CO_2 concentrations from the Ring of Towers in northern Wisconsin. We evaluate the MLEF results with respect to the "truth" and the Kalman Filter (KF) solution. The KF solution is considered theoretically optimal for the problem of this study, which is a linear data assimilation problem involving Gaussian errors. We also evaluate the impact of forecast error covariance localization based on a new "distance" defined in the space of information measures. Experimental results are encouraging, indicating that the MLEF can successfully estimate carbon flux biases and their uncertainties. As expected, the estimated biases are closer to the "true" biases in the experiments with more ensemble members and more observations. The data assimilation algorithm has a stable performance and converges smoothly to the KF solution when the ensemble size approaches the size of the model state vector (i.e., the control variable of the data assimilation problem).

1. Introduction

[2]   The pioneering work of *Evensen* [1994] defined the theoretical background for Ensemble Kalman filter (EnKF) methods and opened new avenues for research and applications in the geosciences. Since then, EnKF methods have been continuously advancing and had many successful applications to the problems of atmospheric, oceanic, hydrological, chemical and carbon transport sciences [e.g., *Houtekamer and Mitchell*, 1998; *Lermusiaux and Robinson*, 1999; *Hamill and Snyder*, 2000; *Keppenne*, 2000; *Mitchell and Houtekamer*, 2000; *Anderson*, 2001; *Bishop et al.*, 2001; *van Leeuwen*, 2001; *Reichle et al.*, 2002; *Whitaker and Hamill*, 2002; *Tippett et al.*, 2003; *Zhang et al.*, 2004; *Ott et al.*, 2004; *Szunyogh et al.*, 2005; *Zupanski*, 2005; *Peters et al.*, 2005; *Dunne and Entekhabi*, 2005].

[3]   One of the newer applications of the EnKF methods is in the area of carbon flux inversion problems [e.g., *Peters et al.*, 2005]. *Peters et al.* [2005] developed a fixed-lag ensemble Kalman smoother algorithm for estimation of surface fluxes of atmospheric trace gases. They applied this algorithm to estimate global surface fluxes of CO_2 using pseudo-observations of CO_2 concentrations, located at the real observing sites. In earlier studies, similar problems have been addressed employing more traditional Bayesian [also referred to as "matrix inversion", e.g., *Rayner et al.*, 1999; *Gurney et al.*, 2002], geostatistical [e.g., *Michalak et al.*, 2004] and Kalman Filter [KF, e.g., *Bruhwiler et al.*, 2005] inversions. As indicated by *Peters et al.* [2005], ensemble-based data assimilation methods are an especially promising alternative to the traditional methods due to their applicability to large-scale non-linear problems, while they still maintain the benefits of updating forecast error covariance as in the traditional methods (e.g., KF). Additionally, ensemble-based algorithms do not require adjoint model development, which is an important advantage over variational methods, which are also novel methods capable of addressing large-scale non-linear carbon inversion problems [e.g., *Chevallier et al.*, 2005; *Engelen and McNally*, 2005
and references therein]. However, carbon inversion studies that compare different data assimilation methods under the same conditions are still lacking, so advantages/disadvantages of different approaches still remain to be evaluated.

[4]   There are also outstanding problems in the carbon inversion that need further attention, independently of the data assimilation approach taken. One of these problems is that the current data assimilation approaches often assume that the geophysical forecast models (including carbon transport models) are perfect. This assumption is not justified, since geophysical forecast models have errors, which could often be in the form of large systematic errors (biases). Estimation and correction of model biases by employing information from observations has become an active area of data assimilation research, ever since the pioneering works of *Sasaki* [1970] and *Derber* [1989] appeared. Since then, the problem of model bias estimation was further investigated under various data assimilation methods, including Kalman filter, variational, geostatistical and ensemble-based methods [e.g., *Bennett et al.*, 1993; *Dee*, 1995; *Reichle et al.*, 2002; *Nichols*, 2003; *Michalak et al.*, 2004; *Tsyrulnikov*, 2005; *Zupanski and Zupanski*, 2006]. Nevertheless, more research in this area is still needed, especially on assigning appropriate uncertainties to the estimated biases.

[5]   In this paper, we estimate the model bias and its uncertainty for a carbon flux inversion problem employing an ensemble-based data assimilation approach. The paper is organized as follows. In section 2 we explain the motivation for this study. In section 3 we define the carbon flux bias estimation problem. In section 4, the approach to the problem taken in this study is shortly described.

Experimental results are presented and discussed in section 5. Finally, in section 6, the conclusions are drawn and future research directions and applications are discussed.

2. Motivation

[6]   This study is motivated by the need to address the model error estimation problem in general, and also by the need to compare different data assimilation approaches for carbon inversion problems. We have chosen to address a data assimilation problem involving a linear carbon transport model and Gaussian probability density functions (PDFs) for which the theoretically optimal solution is known. The optimal solution could be obtained by using the classical KF approach [e.g., *Jazwinski*, 1970], or any other approach that becomes identical to the KF approach for this particular problem.

[7]   We will evaluate an ensemble-based data assimilation approach [Maximum Likelihood Ensemble Filter, MLEF - *Zupanski*, 2005; *Zupanski and Zupanski*, 2006; *Zupanski et al.*, 2006] by comparing it to the KF approach. As explained by *Zupanski* [2005, Appendix A] the MLEF solution is identical to the KF solution when the ensemble size is equal to the size of the control variable. The control variable is the variable we alter in order to find the "optimal" solution (defined in this case as the maximum likelihood solution) to the data assimilation problem of interest. We will refer to the MLEF solution using ensemble size equal to the size of the control variable as the full-rank MLEF solution (since the full-rank forecast and analysis error covariances are used), or simply the KF solution. The equivalence between the KF and the full-rank MLEF was practically confirmed when independent computational algorithms are used to compare the KF and the MLEF results (Uliasz 2006, personal communication). Conversely, we will refer to the MLEF solution using ensemble size smaller than the size of the control variable as the reduced-rank MLEF solution. The reduced-rank MLEF solution is related to the reduced-rank KF solution, but the two solutions are not necessarily identical, since the reduced-rank KF is typically defined in the subspace of orthogonal vectors, which is not necessarily the case for the reduced-rank MLEF.

[8]   The focus of this study is two-fold: (i) to examine if a reduced-rank MLEF, with an ensemble size considerably smaller than the size of the control variable, could still produce useful bias estimation results, and (ii) to examine if the reduced-rank MLEF smoothly converges to the optimal KF solution when the ensemble size approaches the size of the control variable. The former issue is relevant for applications to problems involving large size control variables, since it is only feasible to employ a relatively small ensemble size (i.e., reduced rank approaches). The latter issue is also important to address, since, in order to be able to choose any ensemble size that is practical for the available computing resources, a smooth convergence (as the ensemble size increases) to the theoretically optimal full-rank solution is desirable. These two issues were not addressed in the previous studies, at least not for the carbon flux bias estimation problem.

3. The Problem

[9]   We define the bias estimation problem in terms of estimation of multiplicative bias correction terms applied to the carbon fluxes, which are used as forcing to the carbon transport model. The carbon transport model is an off-line Lagrangian Particle Dispersion Model [LPDM, *Uliasz and Pielke*, 1991 and *Uliasz et al.*, 1996], which is driven by the carbon fluxes obtained from a coupled model, the Simple Biosphere - Regional Atmospheric Modeling System [SiB-RAMS, e.g., *Denning et al.*, 2003; and references there in]. Even though we assimilate simulated observations, real observations (e.g., meteorology, soil and vegetation characteristics) are used to initialize and force SiB-RAMS. This ensures that the SiB-RAMS fluxes are realistic, but not error free.

[10]   To estimate fluxes from atmospheric mixing ratios, we assume that the SiB-RAMS fluxes have errors defined in the form of spatially varying multiplicative correction terms (biases) to the component fluxes. We account for high-frequency time variations of respiration and photosynthesis (or Gross Primary Production, GPP) by assuming that they are obtained by well-understood and relatively easily modeled processes (radiation, temperature, and soil moisture) from SiB-RAMS, then

solve for unknown multiplicative biases β_{RESP} and β_{GPP} in each component flux. Being multiplicative factors to the component fluxes, β_{RESP} and β_{GPP}
are defined in non-dimensional units. The net ecosystem exchange (NEE) is composed of these two component fluxes:

$$N(x, y, t) = \beta_{RESP}(x, y) S(x, y, t) - \beta_{GPP}(x, y) G(x, y, t), \tag{1}$$

where N, S, and G denote NEE, respiration and GPP, respectively; x and y represent grid coordinates and t represents time. Equation (1)
allows for reducing the complex problem of estimation of space and time varying fluxes (S and G) to the estimation of only space varying bias components (β_{RESP} and β_{GPP}).

[11]   The rationale for definition (1)
is as follows. A persistent bias in photosynthesis might result from underestimation of leaf area, available nitrogen, or soil moisture, whereas a persistent bias in respiration might result from overestimation of soil carbon or coarse woody debris. Also, the total soil respiration and the fraction of autotrophic respiration, which are not fully known, can contribute to the persistent (or slow varying) bias. Thus in order to avoid possible cancellation of the two types of errors, it seems reasonable to account for biases in the two flux components, rather than trying to correct the NEE flux via a single bias component. Sub-hourly variations in the simulated component fluxes S and G are primarily controlled by the weather (especially changes in radiation due to clouds and the diurnal cycle of solar forcing), whereas seasonal changes are derived from phenological calculations parameterized from satellite imagery. Fine-scale variations in space are driven by variations in vegetation cover, soil texture, soil moisture, and soil temperature. In any case, it is reasonable that biases β_{RESP} and β_{GPP} vary much more slowly than the fluxes themselves.

[12]   Next, we define the discrete form of the partial derivative of the observed mixing ratio C^{obs} with respect to the NEE, referred to as the surface influence function $C_{k,i,n}* = \frac{\Delta C_k^{obs}}{\Delta N_{k,i,n}}$ (where index k
defines observation location in time and space, index i defines a grid cell, and index n defines a time step). By convolving $C_{k,i,n}*$ with respiration and GPP fluxes $S_{i,n}$ and $G_{i,n}$ obtained from SiB-RAMS and by integrating the backward-in-time particle trajectories from LPDM we obtain the surface influence functions for component fluxes $C_{RESP,k,i}*$ and $C_{GPP,k,i}*$ as

$$C_{RESP,k,i}^* = \Delta t \Delta x \Delta y \sum_n S_{i,n} C_{k,i,n}^*; \tag{2a}$$

$$C_{GPP,k,i}^* = \Delta t \Delta x \Delta y \sum_n G_{i,n} C_{k,i,n}^*. \tag{2b}$$

where the length scales Δx and Δy are the sizes of the grid cells in the zonal and meridional direction, and Δt is the time step over which the fluxes are applied (i.e., Δx, Δy and Δt define spatial and temporal resolution of LPDM). The summation is applied over a time interval, which will be used as a data assimilation interval in the data assimilation experiments, explained later.

[13]   Finally, we can use the influence functions defined in (2a), (2b) and represent the mixing ratio at the observation locations (C_k) as

$$C_k = \sum_{i=1}^{N_{cell}} \beta_{RESP,i} C_{RESP,k,i}^* + \beta_{GPP,i} C_{GPP,k,i}^* + C_{BKGD,k}, \tag{3}$$

where the summation is performed over all grid cells in the LPDM domain (N_{cell}).

[14]   The term $C_{BKGD,k}$ in (3) represents the "background" CO_2 mixing ratio, which includes the

contribution from the initial carbon flux (i.e., the initial conditions), and contribution from the carbon flux through the boundaries of the LPDM domain (i.e., boundary conditions). For simplicity, the "background" carbon flux contribution is assumed perfectly known in the experiments of this paper. Note, however, that in the experiments with real observations the uncertainty of the background flux (especially the boundary conditions part of it, if using a regional domain) could easily become larger then the uncertainties of the flux components S and G. In such case, neglecting the background flux uncertainty may not be justified.

[15]   We seek the maximum likelihood solution to the bias estimation problem, which is in this case equivalent to the KF solution. The method used to find the solution is described in the next section.

4. The Method: MLEF

4.1. Maximum Likelihood Solution and its Uncertainty

[16]   To find the maximum likelihood solution for the bias parameters β_{RESP} and β_{GPP}, we employ an ensemble-based data assimilation approach, MLEF. The basic theoretical background of the MLEF is defined in *Zupanski* [2005]
and generalization of it to include model bias and parameter estimation is given in *Zupanski and Zupanski* [2006]. Here we will only explain the specifics of the MLEF as they apply to the particular bias estimation problem. The problem of finding the optimal bias parameters β_{RESP} and β_{GPP} reduces to the minimization of the following cost function J:

$$J(\boldsymbol{\beta}) = \frac{1}{2}[\boldsymbol{\beta} - \boldsymbol{\beta_b}]^T P_f^{-1}[\boldsymbol{\beta} - \boldsymbol{\beta_b}] + \frac{1}{2}[y - H(\boldsymbol{\beta})]^T R^{-1}[\boldsymbol{y} - H(\boldsymbol{\beta})], \tag{4}$$

where we represented the bias parameters as elements of a vector $\beta = (\beta_{RESP}, \beta_{GPP})$, which size is $N_{state} = 2 \times N_{cell}$, and we used the vector equation $\mathbf{C} = \boldsymbol{H}(\boldsymbol{\beta})$ instead of (3). Note that the $N_{obs} \times N_{state}$ observation operator (denoted $\boldsymbol{H}$) is in general a non-linear operator, but in the experiments of this study we use a linear observation operator. Vector y, of dimension equal to the number of observations N_{obs}, defines simulated observations of the CO_2 mixing ratio, collected over a predefined time interval (referred to as data assimilation interval). Subscript b denotes a background (i.e., prior) estimate of β, and superscript T denotes a transpose. The $N_{obs} \times N_{obs}$ matrix $\boldsymbol{R}$ is a prescribed observation error covariance, and it includes instrumental and representativeness errors [e.g., *Cohn,* 1997]. In the experiments of this study, we use a diagonal, constant in time observation error covariance matrix. The matrix $\boldsymbol{P}_f$ of dimension ($N_{state} \times N_{state}$) is the forecast error covariance matrix, which defines the prior uncertainty of β. Note that in ensemble data assimilation in general, and also in this study, the use of the large matrix $\boldsymbol{P}_f$

is usually avoided by employing a rank-reduced square-root formulation $\boldsymbol{P}_f = P_f^{\frac{1}{2}} (P_f^{\frac{1}{2}})^T$, where $P_f^{\frac{1}{2}}$ is an $N_{state} \times N_{ens}$ square-root matrix defined in ensemble subspace (N_{ens} being the ensemble size).

[17]   To close the equations for the bias estimation problem, we also have to define a dynamical model M
for the bias to transport the bias and its uncertainty from the current data assimilation cycle to the next data assimilation cycle. Assuming that the bias is slowly varying with time, it is appropriate to use the identity operator as a dynamical model for the bias [e.g., *Dee,* 1995; *Zupanski,* 1997; *Dee and da Silva,* 1998; *Nichols,* 2003; *Zupanski and Zupanski,* 2006]. Thus assuming $M = I$, we have

$$\boldsymbol{\beta}_{m+1} = M(\boldsymbol{\beta}_m) = \boldsymbol{\beta}_m = \boldsymbol{\beta}, \tag{5}$$

where m is a time index and denotes a data assimilation cycle.

[18]   We minimize (4)
via an iterative conjugate-gradient algorithm, which in this case converges in a single iteration to the

KF solution [*Zupanski*, 2005, Appendix A]:

$$\boldsymbol{\beta} = \boldsymbol{\beta}_b + P_f H^T \left(H P_f^T H^T + R \right)^{-1} [\boldsymbol{y} - H(\boldsymbol{\beta}_b)]. \quad (6)$$

The prior and the posterior uncertainties of the solution (6) are defined in ensemble subspace as square roots of the forecast error covariance $P_f^{\frac{1}{2}}$ and the analysis error covariance $P_a^{\frac{1}{2}}$:

$$P_a^{\frac{1}{2}} = \left[\begin{array}{cccc} \mathbf{p}_a^1 & \mathbf{p}_a^2 & \cdots & \mathbf{p}_a^{N_{ens}} \end{array} \right] = P_f^{\frac{1}{2}} (I + A)^{-\frac{1}{2}} , \quad (7)$$

$$A = Z^T Z \qquad ; \quad z^i = R^{-\frac{1}{2}} H \left(\boldsymbol{\beta} + \mathbf{p}_f^i \right) - R^{-\frac{1}{2}} H(\boldsymbol{\beta}) = R^{-\frac{1}{2}} H \mathbf{p}_f^i, \quad (8)$$

$$P_f^{\frac{1}{2}} = \left[\begin{array}{cccc} \mathbf{p}_f^1 & \mathbf{p}_f^2 & \cdots & \mathbf{p}_f^{N_{ens}} \end{array} \right] \quad ; \quad \mathbf{p}_f^i = M \left(\boldsymbol{\beta} + \mathbf{p}_a^i \right) - M(\boldsymbol{\beta}) = \mathbf{p}_a^i. \quad (9)$$

where matrix A of dimension $N_{ens} \times N_{ens}$ is the so-called information matrix in ensemble subspace [*Zupanski et al.*, 2007], and it is defined using ensemble vectors in observational space (z^i). Matrix A will be used in the experiments of this study to evaluate Degrees of Freedom (DOF) for signal of assimilated observations [e.g., *Rodgers*, 2000; *Engelen and Stephens*, 2004; *Zupanski et al.*, 2007]. The square root in (7) is calculated via eigenvalue decomposition of A. It is defined as a symmetric positive semi-definite square root [*Zupanski*, 2005]. Vectors $\mathbf{p}_f^i$ and $\mathbf{p}_a^i$ are forecast and analysis perturbations of $\boldsymbol{\beta}$ in ensemble subspace. Note that, according to (9), for $M = I$, the forecast error covariance in data assimilation cycle m
+ 1 is equal to the analysis error covariance in the data assimilation cycle m:

$$P_{f,m+1}^{\frac{1}{2}} = P_{a,m}^{\frac{1}{2}}. \quad (10)$$

Equation (10)
does not implicate that the forecast error covariance remains constant at all times, because the analysis error covariance changes in time due to the impact of assimilated observations involved in the information matrix A (according to equation (7)).

[19]   In addition to the minimization problem described above, a special care has to be taken to define adequate covariance localization to avoid negative impact of spurious long distance correlations in the forecast error covariance matrix, when ensemble size is small. Also, covariance inflation might be helpful to increase insufficient variance in $P_f^{\frac{1}{2}}$ due to insufficient ensemble size [e.g., *Houtekamer and Mitchell*, 1998; *Hamill et al.*, 2001; *Whitaker and Hamill*, 2002]. These problems are addressed in the next sub-section.

4.2. Dynamic Covariance Localization

[20]   Covariance localization is often used in ensemble-data assimilation applications to better constrain the data assimilation problems with either insufficient observations or insufficient ensemble size [e.g., *Houtekamer and Mitchell*, 1998; *Hamill et al.*, 2001; *Whitaker and Hamill*, 2002; *Zupanski et al.*, 2007]. We are seeking a solution for covariance localization that is sensitive to dynamical changes in the analysis and forecast uncertainties. To define a "distance" for covariance localization, we employ the ratio r
between the forecast and the analysis uncertainty [or in other words, the ratio between the prior (σ_0) and the posterior (σ) uncertainty] defined as $r = \sigma_0/\sigma$, where the prior uncertainty is defined as $\sigma_0 = [\mathrm{diag}(P_f)]^{\frac{1}{2}}$ and the posterior uncertainty as $\sigma = [\mathrm{diag}(P_a)]^{\frac{1}{2}}$. According to the information theory [e.g., *Rodgers*, 2000], and also equation (7)

of this paper, the ratio between the prior and the posterior error covariance matrices measures the information content of the assimilated observations. Thus one can interpret the ratio r as a distance defined in the space of the information measures. Note that this distance is different from the conventionally used geodesic distance in most covariance localization approaches in the current literature.
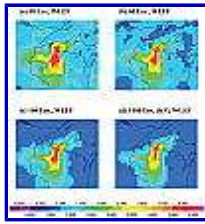

Figure 1. "Distance" r, defined as the ratio between the prior (σ_0) and the posterior (σ) standard deviation ($r = \sigma_0/\sigma$). Values are shown for β_{RESP}, obtained in the first data assimilation cycle in the experiments with (a) 40, (b) 60, (c) 100, and (d) 1800 ensemble members. Small circles with X's indicate sampling sites (indicating all tall towers from the Ring of Towers). Observations from the WLEF tower only, located in the center of the domain, are used in the experiments shown in this figure.

[21] The ratio r calculated for the bias component β_{RESP} is shown in Figure 1. The entire LPDM domain, covering 30×30 grid points, is shown. Simulated observations form WLEF tall tower of northern Wisconsin are used in the experiments shown in Figure 1. Small circles with X's indicate locations of the towers from the Ring of Towers (used in the experiments presented later). A circle located in the center of the model domain indicates the WLEF tower. Results in Figure 1, are shown for different ensemble sizes (40, 60, 100, and 1800). Figure 1 indicates that the ratio r is positive (and also ≥ 1) in all grid points. It has largest values in the areas close to the observations, and it decreases in the areas further from the observations. This suggests that ratio r may be appropriate for measuring a distance from the observations.

[22] Results using ensemble sizes of 100 and 1800 (Figures 1c and 1d) indicate an area of higher values of r
around the tower, where the influence of observations is strongest and the uncertainty reduction is largest, which is clearly separated from the area of smaller values of r, where the influence of observations is weak, or non-existent. Similar strong influence of the observations around the tower can also be seen in the results using smaller ensemble sizes (40 and 60), where the isolines of r are quite similar to the ones obtained with larger ensemble sizes (100 and 1800), however, the distinction between the areas of strong and weak data influence is less clear, indicating problems due to insufficient ensemble size. Therefore covariance localization, which would limit the data influence only within the areas of strongest information content, might be beneficial. Assuming that the KF experiment produced an optimal solution, we can also notice that the ratio r is too large in the experiments with smaller ensemble sizes (red/pink area around the tower is larger), meaning that the posterior error covariance is smaller, thus indicating underestimation of the posterior error covariance due to small ensemble size. This is an indication that covariance inflation would be beneficial in order to account, at least in an approximate way, for the deficient variance when the ensemble size is small.
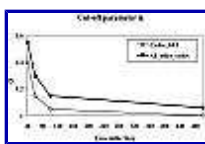

Figure 2. Empirically determined cut-off parameter α used in the experiments of this study. It is defined as a monotonically decaying function of the ensemble size.

[23] Based on the properties of the ratio r we conclude that it is appropriate to localize the influence of the observations within the domain where r is larger than a cut-off ratio:

$$r \geq \alpha \ r_{min} = r_{cut-off}, \qquad\qquad\qquad (10')$$

where scalars r_{min} and $r_{cut\text{-}off}$ represent the minimum and the cut-off values of the ratio r, and α is a cut-off parameter, which is ≥ 1 and it is empirically determined. We also use the cut-off ratio $r_{cut\text{-}off}$ to define covariance inflation, which ensures that the magnitude of the posterior error covariance is equal to the magnitude of the prior error covariance in the areas far from the observations ($r \leq r_{cut\text{-}off}$). The cut-off parameter α

used in the experiments of this study is plotted as a function of the ensemble size in Figure 2. Note that α

decreases with the increasing ensemble size, and approaches the value equal to 1 (no localization) for larger ensemble sizes. Note also that we use different (smaller) cut-off parameters in the first two data assimilation cycles, compared to the cut-off parameters in all other cycles. This is because of the negative influence of the initially prescribed inadequate forecast error covariance matrix.

[24]   Since the covariance localization described above is based on the distance r defined by the dynamically changing information content of data we use the term "dynamic covariance localization" for this particular type of covariance localization. It has to be noted that this localization does not strictly localize the forecast error covariance itself; it localizes the data influence, which is an effect equivalent to localizing the forecast error covariance. The above-described approach for dynamic localization and covariance inflation is evaluated in this study.

5. Experimental Design

[25]   The experiments of this study are performed for a single tall tower (WLEF) and for the Ring of Towers, over a 70-day period starting on 1 June 2004, with a data assimilation interval of 10 days (7 data assimilation cycles). We use the LPDM influence functions on a 20-km grid over a 600×600 km area centered on the WLEF tall tower. The influence functions were generated by running the LPDM backward in time for 2-h mean "samples" from six surface layer towers in the Ring of Towers, plus five levels on the WLEF tower (all but the 11 m level).

[26]   To generate synthetic observations we employ "true" $\boldsymbol{\beta}_{RESP}$ and $\boldsymbol{\beta}_{GPP}$ in (3) and also add Gaussian noise to the "true" observations, with a mean of zero and a variance that depends on the tower height and time of day. The error assigned to the observations is a diagonal matrix $\boldsymbol{R}^{\frac{1}{2}}$ with magnitudes ranging from 1 ppm above 200 m during daytime to 45 ppm below 50 m at night. We use N_{obs}

= 1200 (i.e., 1200 observations per data assimilation cycle) in the experiments with WLEF tower, and N_{obs} = 2640 in the experiments with the Ring of Towers. The size of the control variable is N_{state} = 1800. As will be shown later, the number of independent pieces of observed information is much smaller than N_{state}, which makes the bias estimation problem severely under-constrained by the available observations.
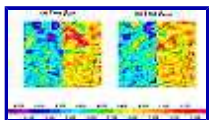

Figure 3.   True biases: (a) $\boldsymbol{\beta}_{GPP}$ and (b) $\boldsymbol{\beta}_{RESP}$.

[27]   The true $\boldsymbol{\beta}$'s are defined to have the following characteristics. On the eastern half of the domain, the mean values of both $\boldsymbol{\beta}$'s are equal to 1.1 non-dimensional units, and on the western half they are equal to 0.5. To make the problem more difficult, we also include random deviations in each $\boldsymbol{\beta}$ chosen from a Gaussian distribution with a mean of zero and a standard deviation of 0.1. Additionally, we apply a smoothing to both $\boldsymbol{\beta}$'s using a compactly supported second-order correlation function of *Gaspari and Cohn* [1999], with decorrelation length scales of 80 km in the southern and 160 km in the northern halves of the domain. The true $\boldsymbol{\beta}_{RESP}$ and $\boldsymbol{\beta}_{GPP}$ are shown in Figure 3.

[28]   As a background value for the model bias $\boldsymbol{\beta}_b$, we use a uniform field of $\boldsymbol{\beta}_b = 0.75$ in every grid

cell. We use this value in the first data assimilation cycle only, since in each subsequent data assimilation cycle the optimal value of β from the previous cycle is used as β_b (equation (5)). To define ensemble perturbations (i.e., $P_f^{\frac{1}{2}}$

) in the first data assimilation cycle we impose a random noise to β_b with a mean of zero and standard deviation of σ_0

= 0.2, in all grid cells. Note that the assumed standard deviation is on average 45% smaller the true standard deviation of β_b. We also smooth the initial random perturbations using *Gaspari and Cohn*'s [1999]

correlation function with the spatial decorrelation length-scale of 120 km (recall that this is different from the "true" decorrelation length scales of 80 km in the southern part and 160 km in the northern part). The smoothing is appropriately normalized to preserve σ_0. In subsequent data assimilation cycles $P_f^{\frac{1}{2}}$ is updated using (10), with application of an inflation factor in some experiments, and no additional smoothing is applied.

[29]   The experimental design described above is chosen for the following reasons. The reason for splitting the domain into two parts, with distinctly different magnitudes in the eastern and western parts, is to be able to examine the method's capability to describe discontinuous transition between the fluxes of different magnitudes. Sharp changes in the fluxes often occur in nature, either due to sharp differences between different ecoregions over land, or due to sharp differences between the ocean and the land. Also, sharp changes occur in the meteorological fields (e.g., associated with the passages of the frontal systems), which are difficult to handle even by the most advanced data assimilation systems available in geosciences today.

[30]   The rationale for choosing *Gaspari and Cohn*'s [1999] correlation function for smoothing was its generality: it can be easily manipulated (by different choices of parameters) in order to define different covariance structures and decorrelation lengths for the fields of interest. The same correlation function (with different parameters) is often used in geosciences. We define larger decorrelation length scales in northern half of the domain, than in the southern half, for the purposes of evaluating the method's capability to recover different scales, which is an additional difficulty of the data assimilation problem.

[31]   In the following subsections we present and evaluate the reduced-rank MLEF results employing varying ensemble size with the condition that $N_{ens} \ll N_{state}$. We also present the full-rank MLEF results ($N_{ens} = N_{state}$

= 1800) and use them to evaluate the reduced-rank results. The experiments of this study are summarized in Table 1.
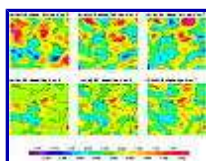
6. Results

6.1. Impact of Dynamic Localization

Figure 4. Estimated β_{RESP}, obtained in the first, third, and seventh data assimilation cycle in the experiments with 40 ensemble members. Results without localization are shown in panels (a), (b), and (c), and with localization in panels (d), (e) and (f).
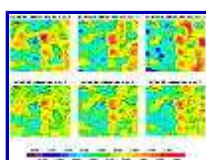

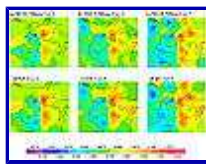**Figure 5.** As in Figure 4, but for ensemble size of 100.

Figure 6. Estimated β_{RESP}, obtained in the first, third, and seventh data assimilation cycle in the experiments with 500 and 1800 ensemble members. Results from the experiment with 500 ensemble members are shown in panels (a), (b), and (c), and from the experiment with 1800 ensemble members (KF) are shown in panels (d), (e), and (f).

[32] The impact of covariance localization on the solution for one of the two biases (β_{RESP}) is shown in Figures 4, 5, and 6
as a function of ensemble size and data assimilation cycles. Experimental results using 40 ensemble members are given in Figure 4, using 100 ensemble members in Figure 5, and using 500 ensemble members in Figure 6. Reference KF solution is also given in Figure 6. By comparing the results with and without localization in Figure 4, where the smallest ensemble size is used (40 ensembles), we can notice large differences between the two solutions in the areas far from the observations (recall that simulated observation from the WLEF tower, located in the center of the domain are used in the experiments of this subsection). Comparing the results with the truth, given in Figure 3b, and also with the KF solution given in Figures 6d–6f, indicates that the experiment with localization (panels d, e, and f) is in better agreement with the truth and with the KF solution, thus indicating a positive impact of covariance localization. The improvements are mostly confined to the areas far from the observations, which is an expected impact of covariance localization, since it does not change the solution in the vicinity of the observations. One can also notice localized impact of data in the first data assimilation cycle in (Figure 4d), where the estimated β_{RESP} is equal to the background value (of 0.75) in the large portion of the model domain. In later cycles (Figures 4e–4f), the data influence gradually spreads out, which results in further improvements of the solution. Note that the solution of the experiment without localization also improves with time, however, remains worse than the solution with localization.

[33] One should be aware, however, that even the optimal KF solution departs considerably from the truth in the areas far from the observations, since little information about the truth is given to the system in these areas. These results could be further improved by employing additional data (additional towers, as shown later). The solution is also weakly constrained over the Great Lakes, but it is not clear if additional observations would improve the results over the Great Lakes because both GPP and respiration are zero there.

[34] Experimental results with 100 ensemble members (Figure 5) still indicate larger differences between the experiments with and without localization in the less observed areas, however these differences are smaller than in the experiments with 40 ensemble members. This is because covariance localization is less critical when the ensemble size is larger.

[35] Finally, comparing the experimental results using 500 ensemble members with the KF results in Figure 6, one can notice a striking similarity between the two solutions, thus indicating that, when larger ensemble sizes are used, covariance localization might not be needed, or it might be helpful only in the initial data assimilation cycles. Note that dynamic localization is applied only in the first two data assimilation cycles in the experiment with 500 ensemble members (see also Table 1).

[36] Figures 4, 5, and 6
also indicate that the MLEF is capable in describing discontinuous transition in the fluxes from east to west portion of the model domain. This discontinuity is better described in the well-observed area, where the data information content is larger (ratio r is larger, as shown in Figure 1). The increasing ensemble size also improves the MLEF capability to recover this discontinuity, however, it is not well recovered in the non-observed or weakly observed areas (where ratio r is small). The MLEF capability to realistically describe different scales is not obvious in this experimental set up, though there are hints of smaller scales in the southern portion and larger scales in the northern portion in the experiments with larger ensemble size.
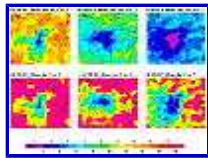
Figure 7. Analysis (posterior) uncertainty σ_{RESP} of the estimated β_{RESP}, obtained in the first, third, and seventh data assimilation cycle in the experiments with 40 ensemble members. Results without localization are shown in panels (a), (b), and (c), and with localization in panels (d), (e), and (f). The values of σ_{RESP} are multiplied by 100.
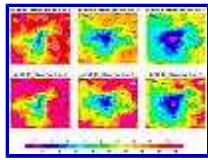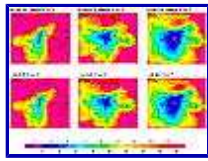

Figure 8. As in Figure 7, but for ensemble size of 100.

Figure 9. Analysis uncertainty σ_{RESP} of the estimated β_{RESP}, obtained in the first, third, and seventh data assimilation cycle in the experiments with 500 and 1800 ensemble members. Results from the experiment with 500 ensemble members are shown in panels (a), (b), and (c), and from the experiment with 1800 ensemble members (KF) are shown in panels (d), (e), and (f). The values of σ_{RESP} are multiplied by 100.

[37] Let us now examine the impact of dynamic localization on the posterior uncertainty of the estimated solution. In Figures 7, 8, and 9 posterior uncertainty σ_{RESP} is shown as a function of ensemble size and data assimilation cycles. Results using 40 ensembles are given in Figure 7, 100 members in Figure 8, and 500 and 1800 members in Figure 9. Assuming, as before, that the results using KF are optimal, one can immediately notice that σ_{RESP} is severely underestimated in the experiment with 40 ensemble members (Figures 7a–7c). This is significantly improved when applying dynamic localization (Figures 7d–7f), since these results, even though not perfect, are much closer to the optimal results, given in Figures 9d–9f. The positive impact of dynamic localization on σ_{RESP} is still present in the experiment with 100 ensemble members (Figure 8), but it is less pronounced. Finally, when the ensemble size is very large (500 ensembles), the estimated σ_{RESP} is almost indistinguishable from the optimal σ_{RESP}, obtained using the KF (Figure 9).
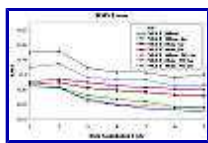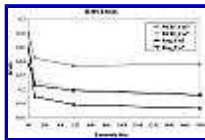

Figure 10. Total RMS errors of the estimated biases (β_{GPP} and β_{RESP}) calculated with respect to the truth, shown as functions of data assimilation cycles. Results from the experiments with varying ensemble size (from 40 to 1800), with and without dynamic localization, are shown (experiments listed in Table 1).

[38] To end this subsection, we summarize the results by plotting the total Root Mean Square (RMS) errors of the estimated β_{GPP} and β_{RESP} as functions of data assimilation cycles in Figure 10. The total RMS errors are calculated with respect to the truth, in all model points, and for all experiments listed in Table 1. Figure 10 indicates clear positive impact of dynamic localization in reducing RMS errors in all data assimilation cycles when the ensemble size is small (40 and 60 ensembles). For ensemble size of 100, the impact of covariance localization is slightly negative, since the RMS errors are slightly smaller in the experiment without localization, indicating that for ensemble sizes of around 100 or larger covariance localization might not be needed (at least not in all cycles). Finally, the RMS errors of the experiment with 500 ensemble members, which is performed using localization only in the first two cycles, are comparable to the RMS errors of the KF. Additionally, posterior error covariance is in

good agreement with the KF results (Figure 9), thus indicating superior performance of the experiment with 500 ensembles in both aspects. The results of this subsection are quite encouraging indicating not only positive impact of dynamic localization when the ensemble size is small, but also smooth convergence of the reduced-rank MLEF solution to the KF solution. We will further examine the convergence of the MLEF solution toward the KF solution in the experiments with the Ring of Towers, presented in the next subsection.

6.2. Impact of More Data

Figure 11. Total RMS errors of the estimated biases (β_{GPP} and β_{RESP}) calculated with respect to the truth, shown as functions of the ensemble size. Results from the experiments with WLEF tower only (denoted by prefix WLEF) and the experiments with the Ring of Towers (with prefix Ring) are shown for third and seventh data assimilation cycle.

[39]   To further examine the convergence of the MLEF solution to the KF solution, let us examine total RMS errors of the estimated biases (β_{GPP} and β_{RESP}), plotted in Figure 11 as functions of ensemble size. Results from third and seventh data assimilation cycle, using observations from a single tower (WLEF) and the Ring of Towers, are shown. As the figure indicates, all experiments converge to the KF solution. The convergence is fast up to the ensemble size of 100 (the lines are steep), and after that, it dramatically slows down. Thus in the range of ensemble sizes 100–1800 only small improvements could be expected when increasing the ensemble size. Thus from the cost-benefit point of view, one can decide that ensemble size of 100 is the best choice for the data assimilation problem of this study.

[40]   Examining the impact of more data in Figure 11, we can notice that more data from the Ring helps reducing the RMS errors for all ensemble sizes in both 3rd and 7th data assimilation cycle. The improvements are larger in the 3rd cycle (after 30 days) than in the 7th cycle (after 70 days), which leads to the results of the Ring after 30 days being of similar quality as the results of the WLEF after 70 days. These results indicate that more observations would be helpful when estimating biases that have shorter timescales.

[41]   Let us now examine if the more observations from the Ring brings more independent pieces of information. This can be accomplished by evaluating DOF for signal (d_s), defined as [e.g., *Rodgers*, 2000; *Zupanski et al.*, 2007]:

$$d_s = \sum_i \frac{\lambda_i^2}{(1 + \lambda_i^2)}, \qquad (11)$$

where λ_i^2 are eigenvalues of the information matrix A, given in (6) and (7). DOF for signal in (11) measures the number of independent pieces of information of assimilated observations that are above the noise ($\lambda_i^2 > 1$). Since dimensions of the matrix A are small ($N_{ens} \times N_{ens}$) it is easy to calculate its eigenvalues, even when the model state vector is large (i.e., N_{state} is large). Note that d_s cannot exceed the ensemble size, thus, it can be underestimated in the experiments using insufficient ensemble sizes. Nevertheless, as demonstrated in *Zupanski et al.* [2007], it is still valid to compare d_s values within the same ensemble size.
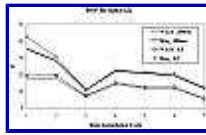
Figure 12. DOF for signal obtained in the experiments with the WLEF tower and with the Ring of Towers, shown as a function of data assimilation cycles. Results using 100 ensemble members (WLEF_100ens and Ring_100ens) and 1800 members (WLEF_KF and Ring_KF) are shown.

[42]   In Figure 12 we show d_s, obtained in the experiments with the WLEF tower and with the Ring of Towers, as a function of data assimilation cycles. Results using ensemble size of 100 and 1800 members are shown as examples. As the figure indicates, assimilation of more observations brings more independent information to the system. The information content of data is the highest in the first couple of data assimilation cycles, since the system's knowledge about the truth is poor (similar results are also obtained using a different dynamical model in *Zupanski et al.*, 2007). The information content also changes from one data assimilation cycle to another due to dynamical changes in the flow patterns and other variables that drive the LPDM influence functions. We can also notice a slight underestimation of the true information content in the experiments with 100 ensembles (lines are slightly below the lines obtained using the KF); however, the essential characteristics of the information content are well captured with 100 ensembles.
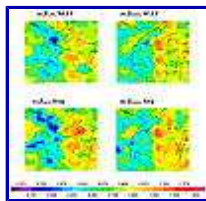

Figure 13. Estimated β_{GPP} and β_{RESP}, obtained in the seventh data assimilation cycle in the experiments with 100 ensemble members. Results with observations from the WLEF tower are shown in panels (a) and (b) and with the observations from the Ring of towers in panels (c) and (d).
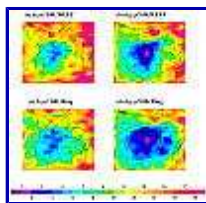

Figure 14. Analysis uncertainties σ_{GPP} and σ_{RESP}, obtained in the seventh data assimilation cycle in the experiments with 100 ensemble members. Results with observations from the WLEF tower are shown in panels (a) and (b) and with the observations from the Ring of towers in panels (c) and (d). The values of σ_{GPP} and σ_{RESP} are multiplied by 100.
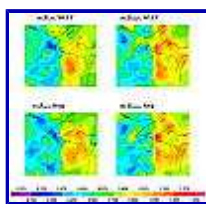

Figure 15. As in Figure 13, but using KF.
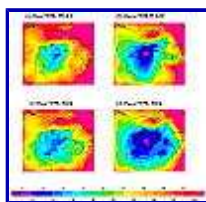

Figure 16. As in Figure 14, but using KF.

[43]   We illustrate the impact of more data on the solutions for β_{GPP} and β_{RESP} and their posterior uncertainties in Figures 13, 14, 15, and 16. Results obtained after 7th data assimilation cycle are shown. Figures 13 and 14 show estimated β_{GPP} and β_{RESP} and their uncertainties, obtained using 100 ensemble members (experiment MLEF_100ens_loc of Table 1, which includes dynamic localization).

Figures 15 and 16

show the same results, but using the KF approach. The figures indicate that more observations bring further improvements to the estimated biases and their uncertainties, for both the MLEF with 100 ensemble members and the KF. The east-west discontinuity line is slightly better defined, there is more distinctions between small scales in the south and large scales in the north, and the minimum values of the posterior uncertainties match the locations of the towers. These are all theoretically expected, and thus quite encouraging results, indicating good performance of the data assimilation system, and benefits of having more towers. Furthermore, the locations of the towers in the Ring seem to be appropriately chosen, since this choice results in a rather uniform wide area around the towers of strong error reduction. This, however, has to be confirmed in the experiments with real observations, which will be done in the next stage.

[44]   Finally, by comparing the uncertainties obtained using 100 ensemble members and 1800 members (Figures 14 and 16), one can notice a slight underestimation of the errors in the experiment with 100 ensembles. This underestimation did not have an obvious detrimental impact on the solution, since the MLEF remained stable when repeating data assimilation cycles (it did not show any signs of divergence). Considering the computational cost, which is directly proportional to the ensemble size, slight deficiencies in the reduced-rank solution are still a small price to pay for the wealth of valuable information offered by this solution, as demonstrated in the experimental results of this study.

7. Conclusions

[45]   In this study we have evaluated the potential of ensemble based data assimilation approaches, using the MLEF approach as an example, to estimate multiplicative bias correction terms, applied to photosynthesis and respiration CO_2 fluxes. The CO_2 fluxes, obtained by running a complex atmosphere-biosphere model (SiB-RAMS), are used as forcing to a relatively simpler (and computationally less expensive) carbon transport model (LPDM), which produced CO_2 concentrations at available observation locations.

[46]   The experimental results, using simulated CO_2 concentrations from a single tall tower (WLEF) and the Ring of Towers of northern Wisconsin, indicated that the MLEF could successfully estimate carbon flux biases and their uncertainties. The data assimilation algorithm had a stable performance over a wide range of ensemble sizes, and converged smoothly to the KF solution as the ensemble size approached the size of the control variable. These favorable results indicated that assimilation of CO_2 concentrations using an ensemble-based approach, such as the MLEF, could reduce model bias errors in CO_2

fluxes, at least when using similar process models as in our study and when observations are available.

[47]   One can also ask a question if the estimated biases could be applied to improve CO_2 fluxes in the future (i.e., the forecasted CO_2

fluxes), when observations are not available. We expect that the biases estimated at the current time should be applicable for correcting the CO_2

fluxes in the future, as long as the biases have slow time variability. However, the question regarding how long in the future the estimated bias correction terms could remain constant is still open, since we would need to perform data assimilation experiments with real observations to learn about time variability of the "real" biases. In this study we assumed that the "true" biases remained constant over a long period of time (70 days), but in reality the biases could change on a shorter timescale, which would make the bias estimation problem more difficult.

[48]   We also examined the impact of covariance localization, formulated using a new "distance" function defined in the information space. This distance, being sensitive to the flow patterns, is different from the classical geographical definition of distance. The impact of covariance localization was positive for small ensemble sizes (40–60). For larger ensemble sizes we have found that the

localization was not essential, at least not after a couple of the initial data assimilation cycles. These are expected results of a covariance localization approach.

[49]   Verifications with respect to the "truth" indicated that, even when using the best data assimilation approach (KF), there is room for further improvements of the results, which can be achieved by including more observations (e.g., more towers). Adding more observations from the Ring of Towers further improves the data assimilation results in terms of both the estimated biases and their uncertainties. The experiments with more towers also indicated that the "true" bias could be recovered after a shorter time period (e.g., Figure 11), which is important for estimating biases with shorter timescales (then 70 days).

[50]   The results of this study, even though encouraging, will have to be confirmed in the experiments using real observations. This is planned for the future research.

Acknowledgments