



HAL
open science

Impact of correlated observation errors on inverted CO₂ surface fluxes from OCO measurements

Frederic Chevallier

► **To cite this version:**

Frederic Chevallier. Impact of correlated observation errors on inverted CO₂ surface fluxes from OCO measurements. *Geophysical Research Letters*, 2007, 34 (24), 10.1029/2007GL030463 . hal-02948201

HAL Id: hal-02948201

<https://hal.science/hal-02948201>

Submitted on 7 Oct 2020

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

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

Impact of correlated observation errors on inverted CO₂ surface fluxes from OCO measurements

Frédéric Chevallier¹

Received 23 April 2007; revised 5 October 2007; accepted 14 November 2007; published 20 December 2007.

[1] Proper assignment of error statistics is essential in the field of Bayesian inference. This paper studies the impact of correlated observation errors in the case of the estimation of CO₂ surface fluxes from NASA's forthcoming Orbiting Carbon Observatory (OCO). Using a series of observation simulation system experiments, it is shown that hypothetical observation error correlations of 0.5 in neighbouring observations have a rather limited impact on the accuracy of the inverted fluxes when they are correctly taken into account. The information loss induced by commonly-used approximate treatments of the observation error correlations (neglecting, observation thinning and error inflating), that are computationally more efficient, is quantified. Error inflation has the least detrimental impact among the suboptimal set-ups and limits the loss in uncertainty reduction to a few per cent, in spite of its very low reduced chi-squared. **Citation:** Chevallier, F. (2007), Impact of correlated observation errors on inverted CO₂ surface fluxes from OCO measurements, *Geophys. Res. Lett.*, *34*, L24804, doi:10.1029/2007GL030463.

1. Introduction

[2] Statistical inference systems are vulnerable to any structure in their random variables that is not well accounted for. Systematic errors, or biases, are a prominent example of such structures and receive much attention. Error correlations are another expression of organized patterns in the inference systems. A proper inference system should link correlated information pieces and weight them properly. For instance, correlated errors in the prior information reduce the effective dimension of the inversion problem and therefore theoretically should induce more accurate solutions. Correlated observation errors have a similar beneficial effect when each observation corresponds to a different variable to estimate. However, for a series of observations of a single variable x , the final uncertainty on x is larger for positively-correlated than for uncorrelated errors (the effect is opposite for negative correlations). In practice, correlations are often ignored, both because there are difficult to detect and quantify, and because properly taking them into account slows down the inversion systems to a large extent. A prominent illustration is given by the numerical weather prediction systems, since most of them assume uncorrelated observation errors (but correlated prior errors). To attenuate the effects of such a rough simplification, these systems include two empirical adjustments [e.g., Liu and Rabier, 2003, and references therein]: the observation

density is thinned (i.e., only a subset of all possible remotely-sensed weather data is assimilated) and the errors assigned to the assimilated ones are usually inflated. For the estimation of CO₂ surface fluxes from measurements of atmospheric concentrations, diagonal error matrices have been empirically used until recently [e.g., Gurney *et al.*, 2002]. Prior error correlations have been introduced in some studies [e.g., Rödenbeck *et al.*, 2003], but observation errors are still assumed to be uncorrelated, even though all components of the observation errors can be affected by correlations. For instance, error correlations in remote sensing measurements are generated by misinterpreted features in the electromagnetic spectrum and by the ancillary data used in retrieval [see, e.g., Chevallier *et al.*, 2005, Figure 2]. Since the observations are interpreted in the inversion system with the help of an atmospheric transport model (that links the fluxes to the measurements), such models also contribute to the observation error budget and induce space-time correlations between observation errors [e.g., Kaminski *et al.*, 2001].

[3] Before the end of the decade, two groundbreaking CO₂-dedicated instruments will be launched to monitor CO₂ concentrations from space: Orbiting Carbon Observatory (OCO) [Crisp *et al.*, 2004] and Greenhouse Gases Observing Satellite (GOSAT) [Inoue *et al.*, 2006]. Several articles [e.g., Rayner and O'Brien, 2001; Pak and Prather, 2001; Houweling *et al.*, 2004; Chevallier *et al.*, 2007] have highlighted the potential of such data to significantly reduce the uncertainties related to flux variations. These successive studies describe observing system simulation experiments (OSSEs) with increasing realism, but all of them made the assumption of null observation error correlations. At best, the observation density was preliminarily thinned.

[4] This paper aims at investigating the impact of correlated errors in inverse modelling. OCO serves as a case study, with hypothetical error correlations of +0.5 introduced for observations 280 km apart. Such correlation errors could be present in the remote sensing product itself or could be caused by the transport model used in the inversion system. Despite the large number of observations, an analytical form of the covariance matrix inverse is found, given the instrument measurement configuration. This form serves us as a rigorous reference to assess the impact of commonly-used simplified treatments of the correlations: neglecting, observation thinning and error inflating. The set-up of our OSSEs is described in the next section. Results are shown in section 3, followed by a conclusion in section 4.

2. Method

[5] The simulation of the observing system impact on flux estimation closely follows the method of Chevallier *et al.* [2007]. We recall here its successive steps: (1) use a

¹Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, CEA, CNRS, UVSQ, Gif-sur-Yvette, France.

climatology of CO₂ surface fluxes as boundary conditions to a transport model and generate a set of pseudo observations accounting for the satellite orbit and cloud cover, (2) perturb the pseudo-observations consistently with assumed observation error statistics, (3) perturb the surface flux climatology consistently with assumed error statistics, (4) perform a Bayesian inversion of the surface fluxes using the perturbed pseudo-observations as data and the perturbed climatology as the prior field, (5) compare the estimate of the inversion to the flux climatology to get the errors in the estimate.

[6] The method is actually applied several times with different perturbations each time, in order to compute the inversion error statistics. The present study relies on an ensemble of four one-year inversions of surface fluxes in eight-day segments. Doing that, a series of 180 fluxes is available at each location of the world that provides stable statistics. In the case of large-dimensional systems, like the present one, the errors in the estimate can also be computed with the Lanczos algorithm [Chevallier *et al.*, 2005], but only in the case of optimal systems. The Monte Carlo approach used here does not have such a limitation.

[7] The atmospheric transport is simulated by the general circulation model of the Laboratoire de Météorologie Dynamique (LMDZ) [Hourdin *et al.*, 2007] at 3.75° × 2.5° (longitude-latitude) resolution, nudged to winds from weather analyses. Atmospheric conditions correspond to year 2003. The CO₂ flux climatology includes 3-hourly biospheric fluxes, monthly oceanic fluxes and yearly fossil fuel emissions, but, by construction of our OSSEs, the results presented here are not sensitive to their definition (but rather to the definition of their errors, which is given later in this section).

[8] The observations are individual column-averaged dry air mole fractions of CO₂, denoted X_{CO₂}, in clear spots and in the sunlit hemisphere, like those of the forthcoming OCO instrument [Crisp *et al.*, 2004]. Their time-space sampling emulates the OCO planned orbitography and accounts for cloud cover statistics. It is assumed that the instrument is in the glint observing mode. Before being processed, the observations are binned per orbit at the 3.75° × 2.5° model resolution. This preliminary thinning is motivated by the large observation error correlations that the transport model may induce at short distances. In this study, we investigate the impact of hypothetical mid-range correlations (+0.5) between neighboring model grid boxes. The simulated OCO orbit and cloud cover give about 243,500 observations at the horizontal resolution of the LMDZ transport model for the whole year 2003.

[9] The Bayesian inversion is achieved by the variational scheme of Chevallier *et al.* [2005]. This system finds the optimal fluxes \mathbf{x}_a that fit both the observations \mathbf{y} with their specified error covariance matrix \mathbf{R} and the prior fluxes \mathbf{x}_b with their specified error covariance matrix \mathbf{B} , by iteratively minimizing the cost function J defined by:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{H}(\mathbf{x}) - \mathbf{y}) \quad (1)$$

\mathbf{H} is the transport model convolved with a uniform vertical weighting function.

[10] The control variables \mathbf{x} are the CO₂ surface fluxes both at daytime and night-time, at each point of the 3.75° × 2.5° model grid every eight days. The inversion also retrieves the CO₂ concentrations at the initial time step of the LMDZ simulation, but this has a negligible impact on the surface flux results presented here. Spatial correlations of the individual flux errors are specified as a function of distance, with correlation e-folding lengths of 500 and 1000 km over land and ocean respectively. No temporal correlations are considered for these fluxes at eight-day resolution. The square root of the sum of the error covariances in \mathbf{B} is set to 0.8 and 2.0 Gt C per year for ocean and land respectively. The errors are spread in space proportionally to grid size over ocean and to an annual-mean heterotrophic respiration flux pattern over land. Error standard deviations at the grid point level correspond to about 0.4 gC.m⁻² per day over ocean and 4 gC.m⁻² per day over vegetated areas. This configuration, detailed and discussed by Chevallier *et al.* [2007], expresses the current uncertainty of the carbon budget at the Earth surface in a simple, but rather realistic manner. However, it is subjective and the absolute values of the figures presented in the result section should be interpreted with caution.

[11] The observation error statistics of the 243,500 independent X_{CO₂} observations are the focus of the present study and five configurations (C1 to C5) are considered. They differ in the definition of the covariance matrix \mathbf{R} for the generation of the simulated observations and for the inversion formulation (equation (1)).

[12] The first two cases are statistically optimal, i.e. the inversion finds the most likely value of the fluxes given the information provided by the prior fluxes and by the observations. Consistent with the theory, the covariance matrix \mathbf{R} used to generate the observations is also used in the minimisation (equation (1)).

[13] C1 corresponds to the set-up of Chevallier *et al.* [2007], with uniform observation-error standard deviations of 2 ppm (accounting for the combination of expected measurement and transport errors) and no observation error correlations. Observations are simply generated from the Gaussian distribution centered on the ‘true’ X_{CO₂} values and a 2 ppm standard deviation.

[14] In C2, the standard deviations have the same values (2 ppm), but the observation error correlations are arbitrarily set to +0.5 from one observation to the next along the satellite track, which corresponds to a distance of about 280 km. Distant correlations are set accordingly (e.g., 0.25 between two observations separated by one). Correlated observation errors are generated with a Markov chain. Note that, at the model resolution, the OCO subtracks follow a mostly South-North line from and to the high-latitudes of the two hemispheres, with about 14 orbits per day separated by about 25 degrees of longitude (measurements are made in the sunlit part of the orbits): the linear correlation pattern defined here may not be appropriate for instruments with large swaths. Technically, this configuration involves inverting the non-diagonal \mathbf{R} matrix for the computation of the cost function (equation (1)). This is made possible by the simple form of the exact inverse in our case: without any approximation \mathbf{R}^{-1} is simply a tridiagonal matrix with about 1.67 ppm⁻² in the main diagonal, and about -0.67 ppm⁻² in the two sub-diagonals.

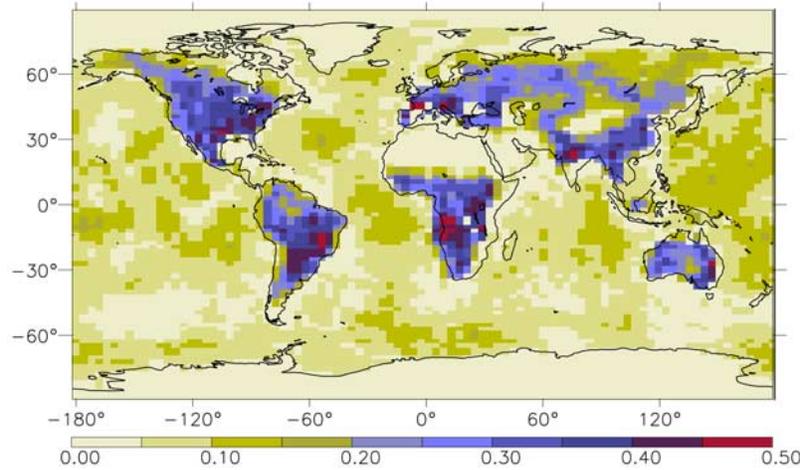


Figure 1. Fractional error reduction of the eight-day mean grid point CO₂ surface fluxes for configuration C1 (no observation error correlations). The error reduction is defined as $(1 - \sigma_a/\sigma_b)$, with σ_a the posterior error standard deviation and σ_b the prior error standard deviation.

[15] The third, fourth and fifth inversion configurations are approximate (and suboptimal) treatments of the error-correlated observations (configuration C2), using an *ad hoc* diagonal \mathbf{R} matrix in the inversion, rather than the non-diagonal one used to generate the observations. Those three configurations are borrowed from the usual practice in variational data assimilation.

[16] Configuration C3 simply ignores the existence of correlations: the off-diagonal terms of \mathbf{R} are suppressed for the inversion (equation (1)) without any further adaptation.

[17] In C4, the observation density is thinned in order to remove the largest correlations from the system. Keeping one observation out of every two, the remaining correlations are less than 0.25 and are neglected. Note that a correlation threshold of 0.2 gave an optimal thinning interval in the study by *Liu and Rabier* [2003], in the context of numerical weather prediction.

[18] The last configuration (C5) inflates the observation errors of the 243,500 X_{CO_2} observations in the inversion, so that the system trusts them less: variances in \mathbf{R} are arbitrarily multiplied by 2 for the inversion (i.e., the assigned standard deviations are about 2.8 ppm) and the off-diagonal terms are neglected. The twofold factor was chosen by trial-and-error so as to make the standard deviation of the flux increments ($\mathbf{x}_a - \mathbf{x}_b$) in this configuration about that of the optimal configuration C1.

3. Results

[19] Figure 1 presents the maps of the uncertainty reduction from the prior to the analysis for the eight-day fluxes, in the case of configuration C1. This optimal set-up without observation error correlations corresponds to the results of *Chevallier et al.* [2007, Figure 2], the main difference being that four years instead of six are used in the statistics (which are, therefore, slightly less representative). The uncertainty reduction is defined as one minus the ratio of the posterior error standard deviation to the prior error standard deviations. A value of zero indicates that the observations have not provided any information to the prior. A value of one would be reached if the observations gave a perfect knowl-

edge about the fluxes. Negative values occur when the analysis is worse than the prior and happen only in suboptimal configurations (see configuration C3 hereafter). Figure 1 shows that the OCO-type observing system should significantly reduce the flux uncertainty over the vegetated areas (by up to about 45% with our set-up). The reduction is smaller over the oceans, but still exceeds 10% in some regions, like the Western Pacific.

[20] The impact of observation error correlations is displayed in Figure 2 for the optimal scenario C2. The error correlations slightly degrade the analysis and correspondingly diminish the fractional error reduction, mainly over ocean, by up to 0.06. The largest impact over ocean may be explained by the length of the prior error structures which is smaller over land (500 km vs. 1000 km, see section 2): there are fewer observations for a given flux error structure over land than over ocean and therefore the impact is less. The different a priori uncertainties assigned to the land and to the ocean may also play a role. An indication of our proper treatment of the off-diagonal terms in \mathbf{R} is given by the value of the cost function J (equation 1) at its minimum: as expected for an optimal system [e.g., *Chevallier et al.*, 2007], it is very close to the number of observations (243,500) and the reduced chi-squared (J divided by the number of observations) is one.

[21] Ignoring the observation error correlations (configuration C3) further degrades the results by up to about 0.1 over both land and ocean (Figure 3a). Since the error reduction in the reference set-up C2 is usually less than 0.10 over sea (Figure 2), the results indicate that the suboptimal flux analysis provides little information in those parts of the globe. It is even slightly worse than the prior fluxes in some oceanic regions (e.g., in the Pacific about French Polynesia and about latitude 50S, or off the coasts of Somalia and of the Arabic Peninsula). The standard deviation of the inversion flux increments is about 30% larger than for the optimal configuration C2 over ocean and about 10% larger over land (not shown). The cost function minimum is not close to the number of observations, as expected from a suboptimal system. It reduces to about

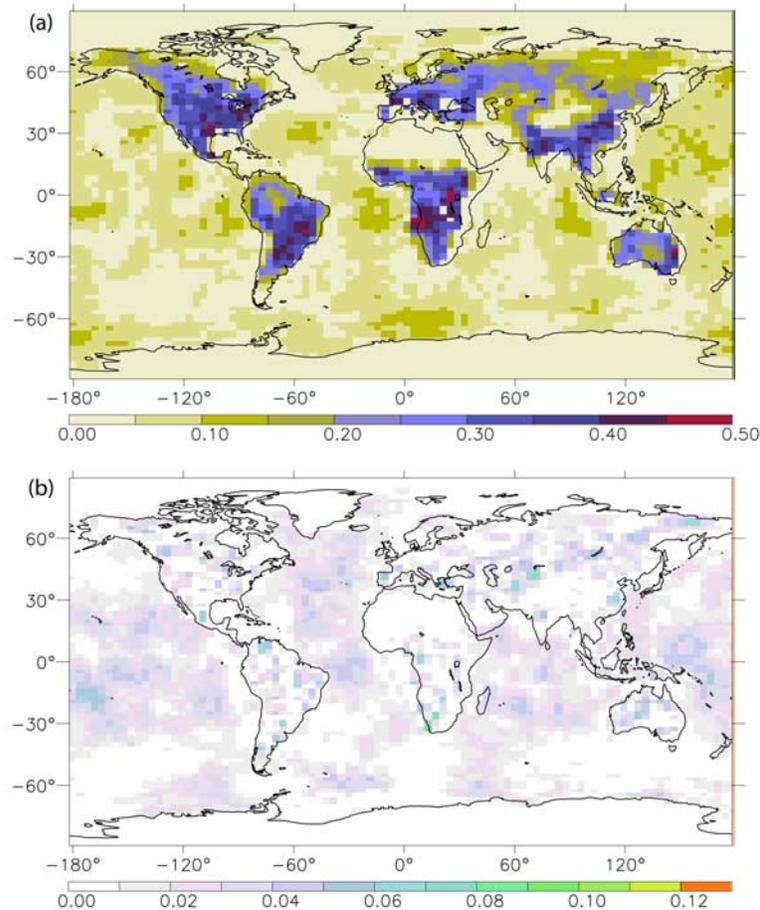


Figure 2. (a) The same as Figure 1, but for the optimal configuration C2, where observation error correlations exist and are properly taken into account. (b) The difference between Figure 1 and Figure 2a (Figure 1 minus Figure 2a).

238,500 vs. 243,500 because the system trusts the observations too much. The reduced chi-squared is about 0.98.

[22] Configuration C4 relies on data thinning to perform the inversion without dealing with the off-diagonal terms of \mathbf{R} . In this case (Figure 3b), while the observations never drive the estimate further from the truth than the prior (as happens in some places in C3), they generally only correct it by two-thirds of the way as in C2, resulting in larger errors than in C3. The twofold-reduced number of observations makes the cost function about 121,000 and therefore a reduced chi-squared about 0.99.

[23] The alternative of inflating the observation error variances (configuration C5) appears to be the least detrimental among the three approximate (and computationally-simple) solutions, with degradations usually less than 0.03 over ocean and 0.06 over land. Note that this confirms the strategy chosen for the Atmospheric Tracer Transport Model Intercomparison Project [Gurney *et al.*, 2002]. In contrast with C3 and C4, the relatively small spatial scale of the difference patterns implies that the quality of the inverted carbon fluxes is hardly affected by the approximate C5 set-up when integrated over continental-size regions. The standard deviation of the inversion flux increments are within 10% of those of the optimal configuration C2 over both land and ocean (not shown). The twofold-inflated

observation error variances make J about half the number of observations at its minimum and induce a chi-squared about 0.5. This very low value despite the rather good performance of the suboptimal configuration emphasises the ambiguity of the chi-squared diagnostic. In particular, it cannot be used alone to tune prior or observation errors safely.

4. Discussion and Conclusions

[24] Performing Bayesian inversion involves modelling the statistical characteristics of the errors. The chosen models potentially include a huge number of univariate and multivariate moments. Multivariate moments, like covariances, are even more difficult to assign than univariate ones, like biases and variances. Further, technical considerations may cause one to neglect some of them. Based on the Monte Carlo approach of Chevallier *et al.* [2007], this paper quantifies the detrimental impact of unaccounted observation error correlations in the inversion of CO_2 surface fluxes for a particular set-up of the inference problem: in some (limited) cases, the inverted fluxes are of lesser quality than the prior ones. Observation thinning and error inflation are safer alternatives, but leave out some of the observation information content, as illustrated with our

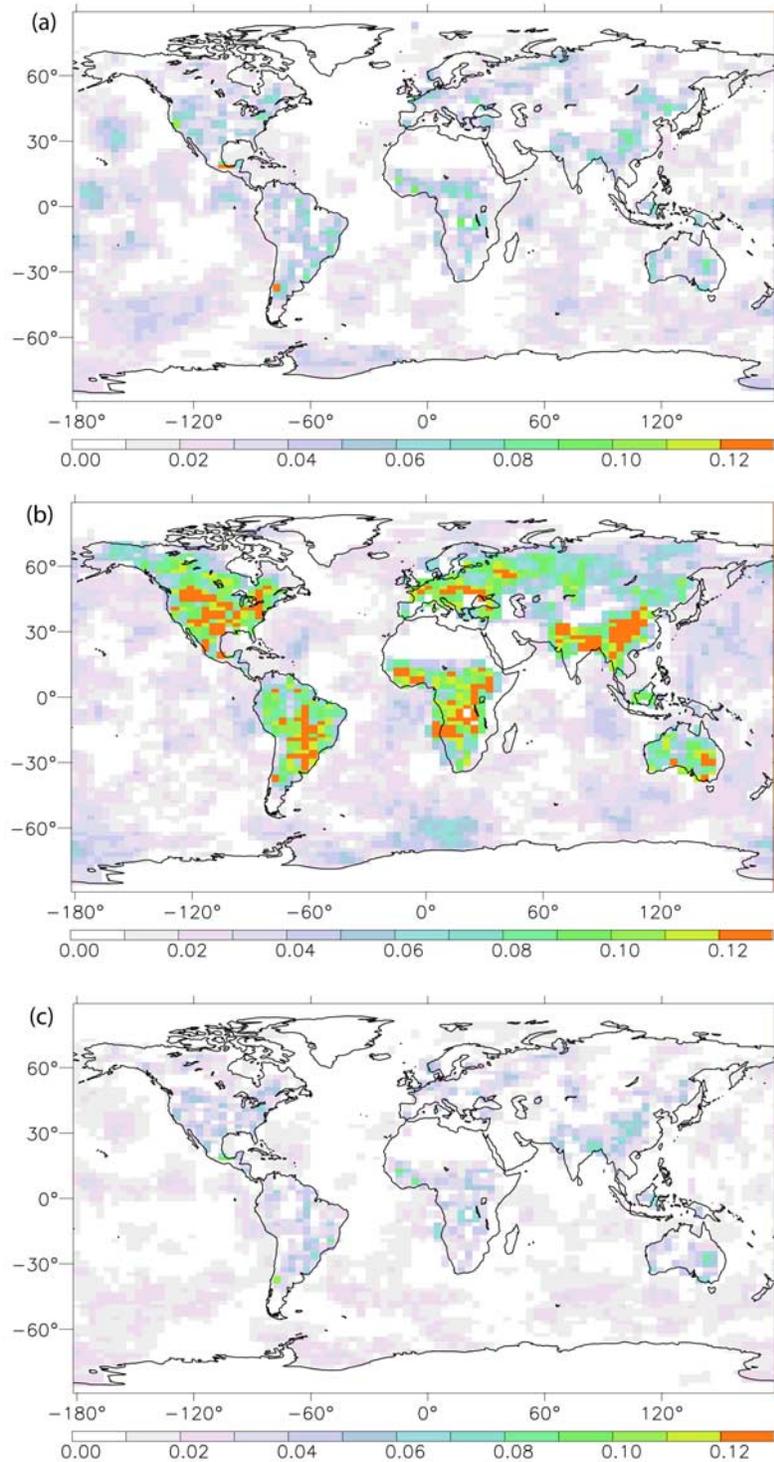


Figure 3. Degradation of the fractional error reduction (FER) shown in Figure 2a (configuration C2), induced by the suboptimal configurations (a) C3, (b) C4, and (c) C5. The degradation is expressed as the FER of configuration C2 minus the suboptimal FER.

inversion configurations C4 and, to a lesser extent, C5. The practical consequence is the importance of rigorous knowledge of spatial error correlations not only for the prior [Chevallier *et al.*, 2006], but also for the observations. Such task is particularly challenging for observation errors since they combine measurement, model and representativeness

errors. Multi-sensor observing systems (with, e.g., OCO, GOSAT and *in situ* measurements together) would be less sensitive to inadequate specification of the measurement errors of the individual instruments because independent biases would partially cancel out. Representativeness and transport error correlations should be studied, e.g., along the

lines of the methods described by Gerbig *et al.* [2003] and Lin and Gerbig [2005] for standard deviations. The present paper focussed on spatial error correlations in the context of a satellite instrument with a sampling repeat cycle of several days (16 for OCO). Temporal correlations of the transport model errors are an important issue for the processing of *in situ* high-frequency measurements and should be carefully studied as well.

[25] **Acknowledgments.** The author would like to thank F.-M. Bréon, P. Peylin and three anonymous reviewers for their constructive comments on an earlier version of this paper, and F. Marabelle for computer support. This study was co-funded by the European Commission under the project GEMS, contract SIP4-CT-2004-516099.

References

- Chevallier, F., M. Fisher, P. Peylin, S. Serrar, P. Bousquet, F.-M. Bréon, A. Chédin, and P. Ciais (2005), Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS data, *J. Geophys. Res.*, *110*, D24309, doi:10.1029/2005JD006390.
- Chevallier, F., N. Viovy, M. Reichstein, and P. Ciais (2006), On the assignment of prior errors in Bayesian inversions of CO₂ 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 CO₂ sources and sinks: Theoretical study in a variational data assimilation framework, *J. Geophys. Res.*, *112*, D09307, doi:10.1029/2006JD007375.
- Crisp, D., et al. (2004), The Orbiting Carbon Observatory (OCO) mission, *Adv. Space Res.*, *34*(4), 700–709.
- 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 CO₂ with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res.*, *108*(D24), 4756, doi:10.1029/2002JD003018.
- Gurney, K. R., et al. (2002), Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models, *Nature*, *415*(6872), 626–630.
- Hourdin, F., et al. (2007), The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dyn.*, *27*, 787–813, doi:10.1007/s00382-006-0158-0.
- Houweling, S., F.-M. Bréon, I. Aben, C. Rödenbeck, M. Gloor, M. Heimann, and P. Ciais (2004), Inverse modeling of CO₂ sources and sinks using satellite data: A synthetic inter-comparison of measurement techniques and their performance as a function of space and time, *Atmos. Chem. Phys.*, *4*, 523–538.
- Inoue, G., et al. (2006), Global carbon dioxide and methane column observation by GOSAT (Greenhouse gases observing satellite), *Geophys. Res. Abstr.*, *8*, 10624.
- Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, *106*(D5), 4703–4716.
- Lin, J. C., and C. Gerbig (2005), Accounting for the effect of transport errors on tracer inversions, *Geophys. Res. Lett.*, *32*, L01802, doi:10.1029/2004GL021127.
- Liu, Z., and F. Rabier (2003), The potential of high-density observations on Numeric Weather prediction: A study with simulated observations, *Q. J. R. Meteorol. Soc.*, *129*, 3013–3035.
- Pak, B. C., and M. J. Prather (2001), CO₂ source inversions using satellite observations of the upper troposphere, *Geophys. Res. Lett.*, *28*, 4571–4574.
- Rayner, P. J., and D. M. O'Brien (2001), The utility of remotely sensed CO₂ concentration data in surface source inversions, *Geophys. Res. Lett.*, *28*, 175–178.
- Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, *3*, 1919–1964.

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