



HAL
open science

On the assignment of prior errors in Bayesian inversions of CO₂ surface fluxes

Frédéric Chevallier, Nicolas Viovy, Markus Reichstein, Philippe Ciais

► **To cite this version:**

Frédéric Chevallier, Nicolas Viovy, Markus Reichstein, Philippe Ciais. On the assignment of prior errors in Bayesian inversions of CO₂ surface fluxes. *Geophysical Research Letters*, 2006, 33 (13), 10.1029/2006GL026496 . hal-02924878

HAL Id: hal-02924878

<https://hal.science/hal-02924878>

Submitted on 17 Sep 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.

On the assignment of prior errors in Bayesian inversions of CO₂ surface fluxes

Frédéric Chevallier,¹ Nicolas Viovy,¹ Markus Reichstein,² and Philippe Ciais¹

Received 4 April 2006; revised 16 May 2006; accepted 23 May 2006; published 1 July 2006.

[1] For the estimation of surface CO₂ fluxes from atmospheric concentration measurements, most often Bayesian approaches have been adopted. As with all Bayesian techniques the definition of prior probability distributions is a critical step in the analysis. However, practical considerations usually guide the definition of prior information rather than objective criteria. In this paper, in situ CO₂ flux pointwise measurements made by the eddy-covariance technique are used to estimate the errors of prior fluxes provided by the prognostic carbon-water-energy model ORCHIDEE. The results contradict the usual convenient assumption of a multivariate Gaussian distribution. The errors of ORCHIDEE have a heavier-tail distribution with a linear temporal dependency after the second lag day and no particular spatial structure. Such error distribution significantly complicates the inversion of CO₂ surface fluxes. **Citation:** 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.

1. Introduction

[2] Quantifying the spatio-temporal variations of CO₂ surface fluxes over continents has been a scientific target of primary importance since the end of the 1980s [Keeling *et al.*, 1989; Tans *et al.*, 1990]. Several bottom-up methods have been developed, based on local observations [e.g., Baldocchi *et al.*, 2001], on ecosystem modelling [e.g., Krinner *et al.*, 2005], or on a combination of both. Despite dramatic improvements, the uncertainty of each estimate is still too large for these estimates to be reliably used for detailed regional estimates of carbon fluxes. One alternative and complementary approach consists in inferring the fluxes from the atmospheric concentration measurements, knowing how the movement of air parcels link the former to the latter (the top-down approach implemented by, e.g., Enting *et al.* [1995], Gurney *et al.* [2002], and Rödenbeck *et al.* [2003]). The diffusive nature of atmospheric transport makes such an inversion problem mathematically ill-posed: the measurements have to be combined with some other information to regularize it, usually through the Bayes' formula. This essential extra information consists of what one knows about the CO₂ surface fluxes prior to the examination of the concentration measurements. If one knew nothing, one should choose uniform prior distribution

probabilities for the fluxes (i.e., any flux state is all equally likely). There actually exists some (limited) knowledge of the biogeochemical processes that govern the fluxes and which are gathered in numerical models of the terrestrial carbon cycle. In situ pointwise measurements of the ecosystem fluxes at flux towers, large scale inventories of fossil fuel emissions, of carbon stocks changes, and satellite-based observations of vegetation activity and disturbances also provide some prior information about the fluxes.

[3] Empiricism has dominated the assignment of prior flux errors in Bayesian inversions so far, which bears consequences on the reliability of the inferred fluxes. For convenience, errors are usually modelled by tuneable multivariate (space-time) normal (Gaussian) distributions. Two different strategies could improve on the current situation. The first one, called "marginalization", consists in treating the unknown characteristics of the prior errors, like the standard deviations, as unknown variables in the Bayes' rule. Michalak *et al.* [2005] developed a simplified approach along this path. This method is still difficult to implement for large dimension problems. Another strategy, which is favoured here, consists in estimating the prior error characteristics based on actual flux observations.

[4] In this paper, surface flux measurements made by the eddy-covariance technique [Aubinet *et al.*, 2000; Baldocchi *et al.*, 2001] on a continuous basis are used to investigate the errors of prior fluxes of the terrestrial biosphere. Prior CO₂ fluxes (i.e., fluxes prior to the analysis of any concentration observation) are provided here by a numerical carbon cycle model: the Organizing Carbon and Hydrology In Dynamic Ecosystems model (ORCHIDEE) described by Krinner *et al.* [2005]. The model and the observations are presented in the next section. Section 3 shows the results, which are discussed in the last section.

2. Model and Data

[5] Developed at Institut Pierre-Simon Laplace (IPSL), the carbon-water-energy model ORCHIDEE explicitly simulates the principal processes of the continental biosphere influencing the global carbon cycle and the fluxes of CO₂ exchanged with the atmosphere, like photosynthesis and respiration of plants and soils. The model handles short-term (half-hourly) to long-term (yearly and beyond) flux and pool variations. It is fully described by Krinner *et al.* [2005]. We focus here on the configuration of the model which is being used as prior information for flux inversions at Laboratoire des Sciences du Climat et de l'Environnement (LSCE). It relies on prescribed atmospheric conditions, a static land-cover distribution, a prognostic observation-independent phenology and a simple two-layer hydrology module.

¹Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CEA-CNRS-UVSQ, Gif-sur-Yvette, France.

²Max-Planck Institute for Biogeochemistry, Jena, Germany.

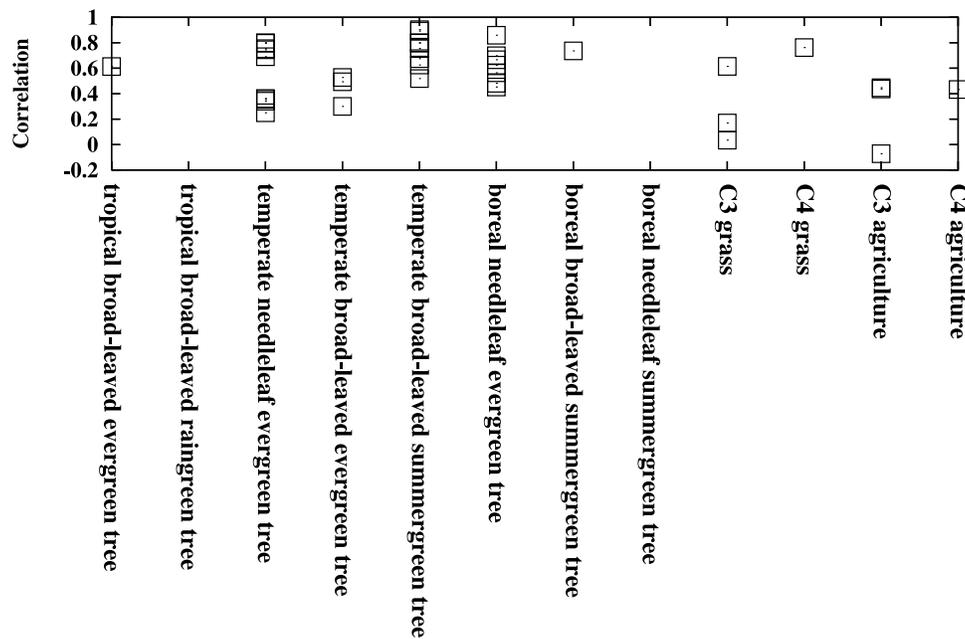


Figure 1. Correlations between the modelled and the observed daily fluxes as a function of the dominant plant functional type on each measurement site. The 12 plant functional types of ORCHIDEE are used. Note that a plant functional type may relatively dominate a site even though it only covers 40% of the site.

[6] The direct measurement of CO₂ surface fluxes is provided by the eddy covariance method. This method deduces fluxes from the covariance between fluctuations in anomalous vertical wind velocity and CO₂ mixing ratio [e.g., Aubinet *et al.*, 2000]. Some limitations of the method in unsteady atmospheric conditions and over complex landscapes induce substantial uncertainty in the fluxes. Random errors are about 0.4 gC.m⁻² for daily totals, based on Hollinger and Richardson [2005], i.e., of smaller amplitude than the departures between the ORCHIDEE simulations and the measurements presented below. Biases occur in some atmospheric conditions and are difficult to quantify. Despite their uncertainties, the flux towers are considered as the reference standard for CO₂ flux measurement and a network of them has been developed across representative ecosystems [Baldocchi *et al.*, 2001]. We use quality-controlled records obtained at 34 flux tower stations located in the Northern hemisphere, for which we assume negligible errors for daily (24-hour) averages in comparison to those of ORCHIDEE. These sites data were from the FluxNet archive at Oak Ridge [Baldocchi *et al.*, 2001] and from a separate collection of European forest sites recently used by Ciais *et al.* [2005]. Each record spans several years between 1994 and 2004, and consists of observations of 2-meter temperature, 10-meter wind, precipitation and radiation fluxes in addition to the CO₂ fluxes, with a time step of 30 minutes. Incomplete records of the CO₂ fluxes have not been gap-filled whereas meteorological variables have been interpolated when needed. These meteorological variables have been used as a boundary condition for ORCHIDEE simulations at each site. Vegetation is distributed in the simulations according to the site characteristics. Note that most sites include more than one vegetation type. The initial plant and soil carbon reservoirs are not known and, following the common practice, have been set at the initial time

step of each simulation so that the simulated ecosystems are carbon-neutral on a yearly basis.

[7] The eddy-covariance flux observations make it possible to investigate the errors of the simulated fluxes at various temporal resolutions. Most inversion studies up to now have inferred monthly fluxes [Gurney *et al.*, 2002, and references therein]. However, the specification of fixed temporal flux patterns within a month necessarily induces spatial and temporal correlations of the model errors at the observation locations that are difficult to take into account, and are usually not. The technical limitations (i.e., computer memory and power) that prevented to infer fluxes at higher temporal resolutions are being circumvented thanks to the introduction of new formulations of the Bayesian inference problem [e.g., Chevallier *et al.*, 2005; Peters *et al.*, 2005]. Consequently, we focus here on 24-hour flux averages.

3. Results

[8] Altogether, the database of daily-mean eddy-covariance fluxes consists of 31,500 quality controlled samples. Figure 1 displays the correlation between the modelled and the observed daily fluxes as a function of the dominant plant functional type (PFT) on each site. The 12-PFT classification of ORCHIDEE is used. The scatter of the points illustrates the diversity of processes that are involved and that are reproduced in ORCHIDEE with various skills, with correlations ranging from about 90% for some forest sites to about zero for some crop sites.

[9] Figure 2 shows the distribution of the model-minus-observation differences after combining the data from all sites. The negative bias of the distribution (0.6 gC.m⁻² per day) was expected because vegetation at most eddy covariance sites is in growing phase, and therefore acts as sink for carbon whereas the model has been initialized to be carbon-

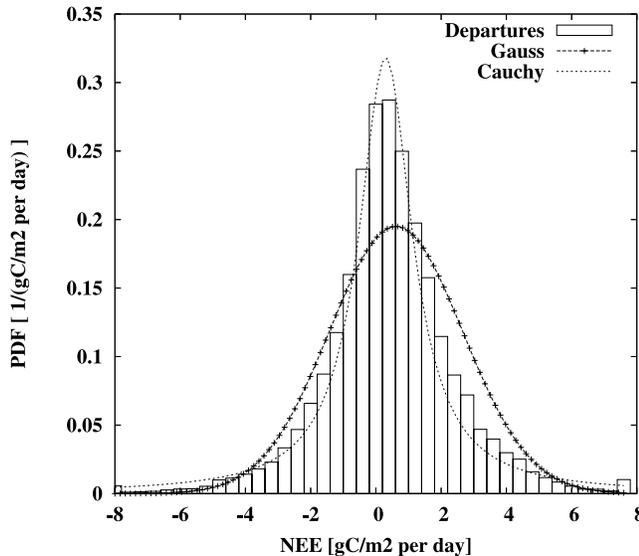


Figure 2. Probability Density Function (PDF) of the ORCHIDEE-minus-observation departures for daily CO₂ fluxes. The Gaussian distribution with the same mean and standard deviation, as well as the Cauchy distribution with location parameter 0.3 and scale factor 1.0, are also reported on the graph.

neutral on the long-term mean. However, a mean $0.6 \text{ gC}\cdot\text{m}^{-2}$ per day sink over the whole vegetated land surface of the Earth, which is about 10^{14} m^2 , would translate into a global sink of 22 GtC per year. This excessive figure [Intergovernmental Panel on Climate Change (IPCC), 2001] indicates that the database of flux measurements over-represents productive vegetation stands and that different stages of ecosystem disturbance regimes are not covered by the FluxNet network well enough. This is consistent with the fact that most sites are affected by human activities, like cutting, planting and nitrogen fertilisation.

[10] The differences spread around the mean with a $2 \text{ gC}\cdot\text{m}^{-2}$ per day standard deviation. Note that such random errors translate to much less than 22 Gt C per year because uncorrelated random errors evolve as the square root of the time and space scales of the aggregation whereas the bias have a linear behaviour. The actual value of the impact of random errors depends on the time and space correlations. As a corollary, the much-looked-for land biospheric sink [IPCC, 2001] may be negligible compared to the daily 24-hours flux at the scale of a model grid point, even though it dominates the uncertainty of the global annual budget.

[11] In Figure 2, two theoretical distributions have been superimposed to the one of the model-data flux differences. The first one is the Gaussian distribution with the same mean and standard deviation. The second one is the Cauchy distribution (also called Lorentz distribution) with location parameter 0.3 and scale parameter 1. Obviously, approximating the ORCHIDEE errors by the Gaussian distribution is a poor approximation whereas the Cauchy distribution would be more appropriate. This is because the daily fluxes are well simulated at some sites during some periods but poorly in other cases (as seen in Figure 1). Errors at individual sites are more normally distributed (not shown). In other words, the simulation error is not purely random

but depends on nuisance variables (in the statistical sense), like the start of the growing season or the plant hydric stress. Therefore a Cauchy distribution, with a relatively narrow peak compared to the tails, better fits the error distribution than a normal distribution. In order to keep Gaussian errors for individual fluxes, one would have to use different widths of the Gaussian distribution, depending on sites or periods.

[12] Correlations of the differences between the simulations and the measurements of the in situ daily fluxes are shown in Figure 3 as a function of time and space. Time correlations drop down to about 70% at lag-day 2 and behave rather linearly afterward, which is far from a Gaussian decay. The correlations at lag-day 30 are about 30%. Spatially, the correlation between the differences at distinct sites is below 50% in absolute value even when considering nearby sites. Given the large spread of the space correlations at any distance, no obvious spatial coherence can be identified.

4. Discussion and Conclusions

[13] Assuming normally-distributed prior errors is common practice for Bayesian inversions for several reasons. First, this approximation makes the problem analytically solvable, either by matrix operations or by the minimization of a cost function [e.g., Lorenc, 1986]. Second, it is the least committal choice when one only knows the mean and the standard deviation of the actual distribution [Jaynes, 1957]. Third, under certain conditions, the distribution of the sum of a large number of independent variables is indeed

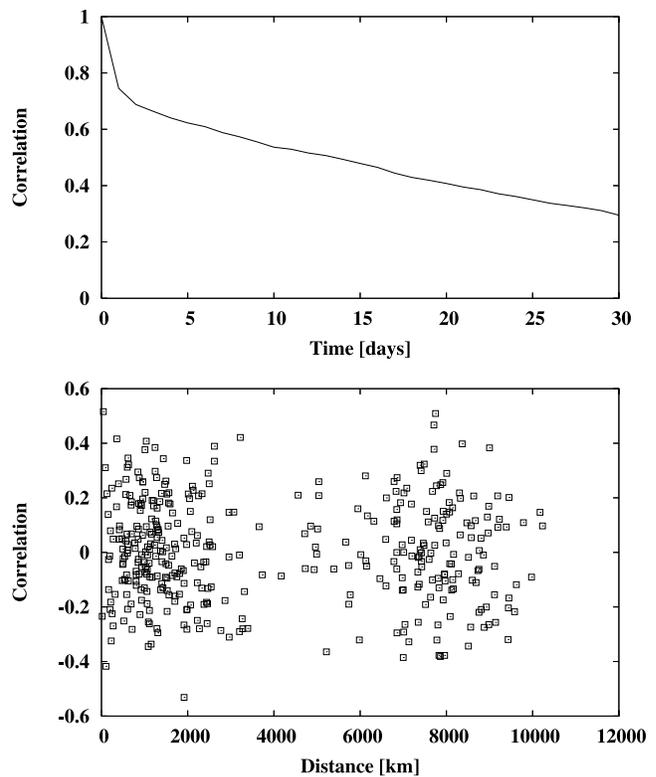


Figure 3. (top) Time and (bottom) space correlations of the differences between the ORCHIDEE simulations and the observations.

approximately Gaussian, as stated by the central limit theorem. None of these reasons justifies a systematic use of Gaussian error distributions and there is a need to investigate the properties of the prior errors, at least to estimate the mean and the variance of their distribution. In this study, pointwise continuous measurements made by the eddy covariance technique have been used to estimate the characteristics of the prior errors for CO₂ flux inversions. The prior is provided by the model of the terrestrial biosphere ORCHIDEE. This strategy is not fully exhaustive. First, the eddy covariance measurements are affected by some errors. Second, the observation towers consider areas of size typically about 1 km², whereas ORCHIDEE is used at a lower spatial resolution to provide flux maps, typically about 100 × 100 km² (see for instance http://www.lsce-orchidee.cea.fr/orchidee_gb.html). Furthermore, for large areas, the atmospheric forcing needed to run the model cannot be obtained from local observations but is provided by short-range weather forecasts of lower accuracy. Last, our flux database is biased toward temperate middle-aged forest ecosystems. However, in spite of these limitations, local observations by the eddy-covariance technique may be the only reliable benchmark and only they can currently provide some evidence about the structure of the simulation errors. The present study focuses on the daily CO₂ fluxes simulated by ORCHIDEE and its results may not be valid for other time scales or other models, that require specific attention based on a similar methodology.

[14] This study indicates an error standard deviation of 2 gC.m⁻² per day when considering all ecosystems together. Combined with the temporal correlations of Figure 3, this number corresponds to a monthly error budget (i.e., the square root of the sum of the covariances within a month) of about 60 gC.m⁻² and a yearly one of about 200 gC.m⁻² at one site (without any respect to the nature of its vegetation). As a comparison, Rödenbeck *et al.* [2003] guessed errors to be of similar or smaller amplitude (depending on latitude) for the prior information in their flux inversion (their Figures 9d and 9e) whereas Houweling *et al.* [2004] supposed twice as large values. These studies included biomass burning in addition to the biosphere photosynthesis and respiration CO₂ fluxes, and hypothesised perfect knowledge of fossil fuel emissions. Both types of processes deserve a specific investigation. Our study also shows that the temporal correlations of the ORCHIDEE errors slowly decrease in a linear way after lag-day 2. No particular spatial structure could be identified. The absence of spatial correlation emphasizes the importance to perform flux inversions at horizontal resolutions as high as possible, since subgrid scale errors are correlated by construction. Obviously, the sparse eddy-covariance network may not reveal some correlation structure that may actually exist. In particular, the fact that each site contains a mixture of vegetation types prevented us to analyze the arguably-larger correlations within a given vegetation type. Nevertheless, there seems to be less justification in introducing spatial correlations than to ignore them. If the results hold for other models of the biosphere and other sites, one would wonder whether surface measurements of CO₂ concentrations actually contain much information about the spatial distribution of biospheric fluxes. Indeed, in the absence of spatial correlations, inversion increments generated by the surface

observations are mainly confined to the vicinity of the measurements [e.g., Bocquet, 2005]. Such issue highlights the importance of the forthcoming spatial instruments dedicated to the observation of atmospheric CO₂, because they will provide measurements well above the surface and a much denser coverage of the globe.

[15] Finally, our results seem to indicate that there is no ground for choosing Gaussian prior error distributions in atmospheric inversion, at least when using daily fluxes from ORCHIDEE. Gaussian distributions can still be justified by their analytical properties from a pragmatic point of view. Indeed a distribution with heavier tails makes the Bayesian cost function non-quadratic and therefore increases the computational burden of the inversion. Further, there is no closed-form solution to the inversion problem any more. However, properly assigning the prior errors in flux inversions would make the flux inversions closer to what they are supposed to be: the best solutions given the evidence provided.

[16] **Acknowledgments.** Authors wish to thank F.-M. Bréon, P. Naveau, P. Peylin and P. Rayner (LSCE) for fruitful discussions about the topic. C. Rödenbeck (MPI-Jena) and two anonymous reviewers made useful comments on an earlier version of the paper. This study was made possible by the work of numerous scientists, students and technicians involved in data collection and analysis at the various Fluxnet sites. Most data have been downloaded from <http://www.daac.ornl.gov/FLUXNET/fluxnet.html>. The other ones have been kindly provided by M. Aubinet, J. Banza, C. Bernhofer, A. Carrara, A. Granier, W. Kutsch, D. Loustau, D. Papale, J. S. Pereira, K. Pilegaard, M. J. Sanz, G. Seufert, J.-F. Soussana and T. Vesala. This study was co-funded by the European Union under projects GEMS and GEOLAND.

References

- Aubinet, M., et al. (2000), Estimates of the annual net carbon and water exchange of forests: The EUROFLUX methodology, *Adv. Ecol. Res.*, **30**, 113–175.
- Baldocchi, D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, **82**, 2415–2434.
- Bocquet, M. (2005), Grid resolution dependence in the reconstruction of an atmospheric tracer source, *Nonlinear Processes Geophys.*, **2**, 219–233.
- Chevallier, F., M. Fisher, P. Peylin, S. Serran, 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.
- Ciais, P., et al. (2005), Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, **437**, 529–533.
- Enting, I. G., C. M. Trudinger, and J. R. Francey (1995), A synthesis inversion of the concentration and δ¹³C atmospheric CO₂, *Tellus, Ser. B*, **47**, 35–52.
- Gurney, K. R., et al. (2002), Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models, *Nature*, **415**(6872), 626–630.
- Hollinger, D. Y., and A. D. Richardson (2005), Uncertainty in eddy covariance measurements and its application to physiological models, *Tree Physiol.*, **25**, 873–885.
- 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.
- Intergovernmental Panel on Climate Change (2001), *Climate Change 2001: The Scientific Basis—Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, 881 pp., Cambridge Univ. Press, New York.
- Jaynes, E. T. (1957), Information theory and statistical mechanics, *Phys. Rev.*, **106**, 620–630.
- Keeling, C. D., R. B. Bacastow, A. F. Carter, S. C. Piper, T. P. Whorf, M. Heimann, W. G. Mook, and H. Roeloffzen (1989), A three-dimensional model of atmospheric CO₂ transport based on observed winds: 1. Analysis of observational data, in *Aspects of Climate Variability in the Pacific and the Western Americas*, *Geophys. Monogr. Ser.*, vol. 55, edited by D. H. Peterson, pp. 165–236, AGU, Washington, D. C.

- Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.
- Lorenç, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*, 1177–1194.
- Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, *J. Geophys. Res.*, *110*, D24107, doi:10.1029/2005JD005970.
- Peters, W., J. B. Miller, J. Whitaker, A. S. Denning, A. Hirsch, M. C. Krol, D. Zupanski, L. Bruhwiler, and P. P. Tans (2005), An ensemble data assimilation system to estimate CO₂ surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, *110*, D24304, doi:10.1029/2005JD006157.
- 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.
- Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the global atmospheric CO₂ budget, *Science*, *247*, 1431–1438.
-
- F. Chevallier, P. Ciais, and N. Viovy, Laboratoire des Sciences du Climat et de l'Environnement/IPSL, CEA-CNRS-UVSQ, F-91191 Gif-sur-Yvette, France. (frederic.chevallier@cea.fr)
- M. Reichstein, Max-Planck Institute for Biogeochemistry, P.O. Box 10 01 64, D-07701 Jena, Germany.