Atmos. Chem. Phys., 11, 10349–10365, 2011
www.atmos-chem-phys.net/11/10349/2011/
doi:10.5194/acp-11-10349-2011

Atmospheric Chemistry and Physics

A comparison of different inverse carbon flux estimation approaches for application on a regional domain

L. F. Tolk[1], **A. J. Dolman**[1], **A. G. C. A. Meesters**[1], **and W. Peters**[2]

[1]VU University Amsterdam, Amsterdam, The Netherlands
[2]Wageningen University and Reseach Centre, Wageningen, The Netherlands

Received: 10 December 2010 – Published in Atmos. Chem. Phys. Discuss.: 31 January 2011
Revised: 17 June 2011 – Accepted: 24 August 2011 – Published: 18 October 2011

Abstract. We have implemented six different inverse carbon flux estimation methods in a regional carbon dioxide (CO_2) flux modeling system for the Netherlands. The system consists of the Regional Atmospheric Mesoscale Modeling System (RAMS) coupled to a simple carbon flux scheme which is run in a coupled fashion on relatively high resolution (10 km). Using an Ensemble Kalman filter approach we try to estimate spatiotemporal carbon exchange patterns from atmospheric CO_2 mole fractions over the Netherlands for a two week period in spring 2008. The focus of this work is the different strategies that can be employed to turn first-guess fluxes into optimal ones, which is known as a fundamental design choice that can affect the outcome of an inversion significantly.

Different state-of-the-art approaches with respect to the estimation of net ecosystem exchange (NEE) are compared quantitatively: (1) where NEE is scaled by one linear multiplication factor per land-use type, (2) where the same is done for photosynthesis (GPP) and respiration (R) separately with varying assumptions for the correlation structure, (3) where we solve for those same multiplication factors but now for each grid box, and (4) where we optimize physical parameters of the underlying biosphere model for each land-use type. The pattern to be retrieved in this pseudo-data experiment is different in nearly all aspects from the first-guess fluxes, including the structure of the underlying flux model, reflecting the difference between the modeled fluxes and the fluxes in the real world. This makes our study a stringent test of the performance of these methods, which are currently widely used in carbon cycle inverse studies.

Our results show that all methods struggle to retrieve the spatiotemporal NEE distribution, and none of them succeeds in finding accurate domain averaged NEE with correct spatial and temporal behavior. The main cause is the difference between the structures of the first-guess and true CO_2 flux models used. Most methods display overconfidence in their estimate as a result. A commonly used daytime-only sampling scheme in the transport model leads to compensating biases in separate GPP and R scaling factors that are readily visible in the nighttime mixing ratio predictions of these systems.

Overall, we recommend that the estimate of NEE scaling factors should not be used in this regional setup, while estimating bias factors for GPP and R for every grid box works relatively well. The biosphere parameter inversion performs good compared to the other inversions at simultaneously producing space and time patterns of fluxes and CO_2 mixing ratios, but non-linearity may significantly reduce the information content in the inversion if true parameter values are far from the prior estimate. Our results suggest that a carefully designed biosphere model parameter inversion or a pixel inversion of the respiration and GPP multiplication factors are from the tested inversions the most promising tools to optimize spatiotemporal patterns of NEE.

1 Introduction

Carbon cycle studies today rely on a wide range of methods with purely observation-based studies on one side of the spectrum and pure modeling on the other side. In between, there are many studies that use a combination of observations and modeling techniques. A special branch of these combined methods is inverse modeling, in which information is

Correspondence to: L. F. Tolk
(lieselotte.tolk@acaciawater.com)

derived from observations using Bayesian statistical methods to minimize the difference between model predictions and observations. Recent inverse modeling studies include for instance efforts to derive the net carbon exchange across the globe or at smaller scales from mixing ratio observations of CO_2 (e.g. Bousquet et al., 2000; Gurney et al., 2003; Rödenbeck et al., 2003; Mueller et al., 2008; Lauvaux et al., 2009; Ciais et al., 2010; Göckede et al., 2010) or attempts to constrain biophysical parameters from eddy-covariance methods (Papale and Valentini, 2003; Carvalhais et al., 2008).

One recurring issue in inverse studies is the large number of choices to be made concerning amongst others, the selection and weighting of observations, the magnitude and correlations of the uncertainties, the treatment of the time and space domain, and even which unknowns to solve for in the application. As a result, no two inverse studies use the same assumptions and the outcome of inverse studies always needs to be evaluated within the limits of the modeling framework chosen.

Many authors have used atmospheric CO_2 mole fraction observations from around the globe to reconstruct the spatiotemporal patterns of net ecosystem exchange (NEE), each in a different way. Peters et al. (2007) used NEE multipliers across large areas of similar vegetation to scale calculated NEE from a biosphere model over each week over many years. Lokupitiya et al. (2008) estimated NEE for a similar time frame, but estimated separate multipliers for simulated gross photosynthetic production (GPP) and ecosystem respiration (R), and for each model grid box. Rayner et al. (2005) in contrast chose to modify a set of physical parameters in the underlying biosphere model directly, thus adjusting NEE to match with observed CO_2 mole fractions. That study however did not estimate separate parameters for separate parts of the globe. The large resulting NEE differences between these three estimates shows that methodology used is an important part of the final result. Clearly, the question which method is most appropriate to estimate NEE has not been yet been resolved (if it ever will) and remains one of the critical issues in estimating source and sink distributions from observations.

In this paper, we want to further investigate the impact of different methodological choices on estimated reanalysis of NEE. The application we chose for this purpose is a regional inversion of CO_2 mole fractions using a high-resolution transport model, and a realistic spatiotemporal distribution of NEE. We plan to use this framework for actual NEE inversions at a later stage, after determining the optimal approach through a set of pseudo-data inversions where the true answer is known. The regional character of this inversion allows us to disregard some of the issues related to carbon cycling on longer time scales, and from long-range atmospheric transport of CO_2. The four methods we want to test are related to the studies mentioned in the previous paragraph: (1) where we estimate NEE multipliers per vegetation type, (2) where we estimate GPP and R multipliers per vegetation type, (3) where we estimate GPP and R multipliers for each grid box, and (4) where we estimate biophysical model parameters for each vegetation type. The specific questions we want to address are: what is the best strategy to determine the spatiotemporal pattern and magnitude of NEE in our domain? What are the strengths and weaknesses of each method?

Pseudodata studies always carry the danger of oversimplifying the real problem, or to be designed in a way to favor one outcome over another. We have tried to design our study to prevent this issue by using a "true" CO_2 exchange distribution from a different biosphere model (FACEM; Pieterse et al., 2007) than the one we use to retrieve the exchange patterns (5PM; Groenendijk et al., 2011). Differences exist between the models in physical formulations, plant functional types, driving parameter values, and in driving meteorological input data, minimizing the a-priori expected similarity. However, both models are based on similar principles and equations and even though they do not share the same land-use map for the domain, the prescribed land-use maps in each model are realistic and thus similar.

New in our approach is that we test all inversion approaches, each with different underlying assumptions, at a high resolution with the same meteorology, whereas in previous comparisons (e.g. Gurney et al., 2003) both the inversion method and the transport model could differ. This enables us to isolate the impact of the inversion methodology. We expect our results to be applicable to similar setups (short time periods, large flux heterogeneity, large CO_2 variations, small transport errors) but caution against extrapolation to the larger scales.

After describing the details of each of the methods included in our tests in Sect. 2.1, we will describe the general characteristics of all inversions in Sects. 2.2 to 2.5. We present our results next in Sect. 3, using a set of five metrics applied to each solution. Special attention is given to the non-linearity of the biosphere model parameter inversion. The strengths and weaknesses revealed in the result sections are further discussed in Sect. 4. Finally, we revisit our research questions and we present general conclusions and recommendations in Sect. 5.

2 Methods

2.1 Inversion methods

In this study we compare four different optimization methods that are used in current state-of-the-art inverse systems for CO_2. We will briefly describe their main characteristics, which are also summarized in Table 1. For each inversion we applied the same methodology (Ensemble Kalman filter). This approach was necessary as the biosphere model parameter inversion is non-linear and could not be solved

Table 1. Summary of the six inverse methods tested, indicating which variables are optimized in the inversions including their correlation and for which spatial unit they are optimized, the degrees of freedom (D.o.f.) as calculated from the a-priori covariance matrix of each system and the χ^2 of innovations.

Inversion name	Variables in state vector	Correlation	Optimization units	D.o.f.	χ^2 of innovations
β_{NEE}	β_{NEE}	NA	ecoregions	6	0.4
βRG0.0	$\beta_{\mathrm{resp}}, \beta_{\mathrm{GPP}}$	0.0	ecoregions	11	0.7
βRG0.5	$\beta_{\mathrm{resp}}, \beta_{\mathrm{GPP}}$	0.5	ecoregions	9	0.6
βRG1.0	$\beta_{\mathrm{resp}}, \beta_{\mathrm{GPP}}$	1.0	ecoregions	6	0.3
βRGpixel	$\beta_{\mathrm{resp}}, \beta_{\mathrm{GPP}}$	0.0	pixels	62	0.3
Parameter	$\beta_{E_0}, \beta_{R10}, \beta_{V_{\mathrm{max}}}, \beta_{\alpha}$	0.0	ecoregions	22	0.6

with typical linear Bayesian solution methods. A description of the Ensemble Kalman filter can be found in Appendix A. Further we refer to Peters et al. (2005) and references therein for a description of additional details of the inversion procedure.

The first inverse method is one where pre-calculated patterns of NEE from a biosphere model are linearly scaled across larger areas, similar to the CarbonTracker system (Peters et al., 2007, 2010). This can be denoted as:

$$\mathrm{NEE}_{\mathrm{post}}(x,y,t) = \beta_{\mathrm{NEE}}(e)\mathrm{NEE}_{\mathrm{prior}}(x,y,t), \quad (1)$$

where β is a scaling factor for each land-use type (e), with an a-priori value of 1.0. $\mathrm{NEE}_{\mathrm{prior}}(x,y,t)$ is a high resolution NEE field from a biosphere model. This method has as advantage that it is straightforward to implement and needs little extra assumptions to stabilize the solution. Disadvantage is that the β factors offer little possibility to change sources into sinks (the sign of β then needs to change), or to scale small (near zero) fluxes to large fluxes (β needs to change a lot).

To overcome some of these disadvantages, systems were suggested that linearly scale gross fluxes instead (Zupanski et al., 2007; Lokupitiya et al., 2008; Schuh et al., 2009):

$$\mathrm{NEE}_{\mathrm{post}}(x,y,t) = \beta_{\mathrm{resp}}(e)R_{\mathrm{prior}}(x,y,t) - \beta_{\mathrm{GPP}}(e)\mathrm{GPP}_{\mathrm{prior}}(x,y,t) \quad (2)$$

R and GPP, which are large and stem from mostly independent processes at short time scales, then each carry a scaling factor β. An advantage is that this system does more justice to the actual processes in the carbon cycle, but a disadvantage is that the separation of the large and opposing fluxes using atmospheric CO_2 is very difficult, and might need extra regularization of the solution in the form of prescribed covariances. In addition to the uncorrelated version of this inversion, we therefore also test this option with correlations of 0.5 and 1.0 between β_{resp} and β_{GPP} for each land-use type.

A final variant of the approach above is to make the β_{resp} and β_{GPP} spatially explicit:

$$\mathrm{NEE}_{\mathrm{post}}(x,y,t) = \beta_{\mathrm{resp}}(x,y)R_{\mathrm{prior}}(x,y,t) - \beta_{\mathrm{GPP}}(x,y)\mathrm{GPP}_{\mathrm{prior}}(x,y,t) \quad (3)$$

This offers the advantage that the β_{resp} and β_{GPP} spatial patterns are allowed to vary within each ecoregion. However, additional regularization is necessary because the number of unknown bias parameters is too large to estimate from the limited atmospheric observations. In this study, we apply a spatial covariation between all grid points in the same land-use type that decreases exponentially with distance, similar to methods used in Rödenbeck et al. (2003), Peylin et al. (2005), Peters et al. (2005), Schuh et al. (2009) with a smaller length scale ($L = 100$ km) to fit with the more detailed regional setup of the inversion.

Another class of inverse methods uses atmospheric CO_2 not to constrain the surface exchange patterns, but to directly optimize the parameters of the underlying biosphere model (Rayner et al., 2005; Scholze et al., 2007). The optimized parameters then control the new surface CO_2 exchange. An advantage of this approach is the seamless extrapolation of information across the space and time domain and the physical relevance of the optimized result (new biosphere model parameter values instead of scaling factors). Disadvantages come in the form of several pitfalls: limitations in the model structure are difficult to overcome, the model parameters are rarely directly constrained by atmospheric CO_2 and aliasing of information into the wrong parameter is possible. In addition, the biosphere model and often contains non-linearities that conflict with the inversion assumptions, as we will discuss elaborately in Sect. 3.1. The biosphere model optimization can be written as:

$$\mathrm{NEE}_{\mathrm{post}}(x,y,t) = f\,(\,\beta_{E_0}(e)E_{0,\mathrm{prior}}, \beta_{R10}(e)R_{10,\mathrm{prior}}, \beta_{Vc_{\mathrm{max}}}(e)V_{\mathrm{m,prior}}, \beta_{\alpha}(e)\alpha_{\mathrm{prior}}\,) \quad (4)$$

where we have selected to optimize 4 parameters for each land-use class: β_{E_0} is a scaling factor for the respiration activation energy E_0 , β_{R10} is a scaling factor for the respiration rate at reference temperature, $\beta_{Vc_{\mathrm{max}}}$ scales the carboxylation capacity, and β_{α} scales the quantum yield for light limited assimilation. The first two parameters are used to adjust respiration while the latter two control photosynthesis (see next section).

The resulting six inverse methods described above will be referred to as the β_{NEE} inversion, the βRG0.0 inversion (no GPP and R correlations), the βRG0.5 inversion (correlations of 0.5), the βRG1.0 inversion (fully correlated GPP and R), the βRGpixel inversion (estimates for each pixel), and the parameter inversion. Note that the βRG1.0 inversion is not the same as the β_{NEE} inversion because the flux covariance is distributed differently in space and time.

For each inversion method the degrees of freedom are estimated with the simple formula from Patil et al. (2001) that considers the number of significant eigenvalues in the correlation matrix (normalized covariance matrix). For each of these inversions, the d.o.f. is relatively small compared to the number of observations (336). The estimated number of degrees of freedom is given in Table 1.

2.2 Modeling system, simulation period and domain

All inversions are based on simulation with the same model setup in which the non-hydrostatic mesoscale model RAMS (Pielke et al., 1992) is used to simulate the atmospheric transport. The version used in this study is B-RAMS-3.2, including adaptations to ensure mass conservation (Meesters et al., 2008). The prior NEE flux estimates are calculated with the simple biosphere model 5PM (Groenendijk et al., 2011) in which photosynthesis is calculated following Farquhar et al. (1980) and respiration is calculated with the relationship by Lloyd and Taylor (1994). The input for the biosphere model is summarized in Table 2. Further details on the transport and biosphere modeling system can be found in Tolk et al. (2009).

The simulations are performed for an area of 400×400 km at a relative high resolution of 10 km, centered in the Netherlands at 52.25° N and 5.2° E (Fig. 1). The Corine 2000 land use maps are used (http://dataservice.eea.europa.eu). Most of the area consists of cultivated land, in which the most abundant land use type is "Agricultural areas with complex cultivation patterns", which is further referred to as "crops 1". Second most abundant is grassland, referred to as "grass", and the third is "Agricultural land with significant areas of natural vegetation", referred to as "crops2". The simulation period 19 May 2008–2 June 2008 was selected to contain various weather types and thus flux regimes, including cloudy, rainy, and sunny days at the beginning of the growing season.

2.3 Control inversions

The simulations are performed in a pseudo-data environment, so that the "true" fluxes are known and the results of the different inversion methods can be evaluated against them. As a check on the inversion system, and as a reference for the performance of the inversions with a perfect or near perfect structure of the NEE flux pattern, we performed two control simulations. In the first, pseudo data were created as a multiplication of the prior NEE fluxes. This pattern is within the solution capacity of each inversion option and it could be retrieved by all inversions, confirming that the inversion system worked correctly.

In addition, we created a flux field based on a simulation with 5PM with parameters that were a realization of the a-priori covariance of the parameters. The flux field is therefore fully within the statistical properties of each inversion method (see also Sect. 2.5), and the spatial structure is represented perfectly in all methods. This is the control inversion referred to in the rest of the text and serves in our setup as a reference for the performance of the inversions in absence of spatial biosphere model structure differences.

Table 2. Summary of the differences between the FACEM biosphere model ("true") and the 5PM biosphere model (a-priori). In our pseudo-data study these differences mimic the expected structural differences between real CO_2 fluxes and those simulated with a biosphere model, and limit the performance of the inversions substantially.

	Prior fluxes	"True" fluxes
Resolution	10 km	6′
Landuse map	Corine2000	SYNMAP
LAI	MODIS-2006	MODIS-2008
Soil map	UN-FAO*	IGBP-DIS
Meteorology	RAMS	ECMWF

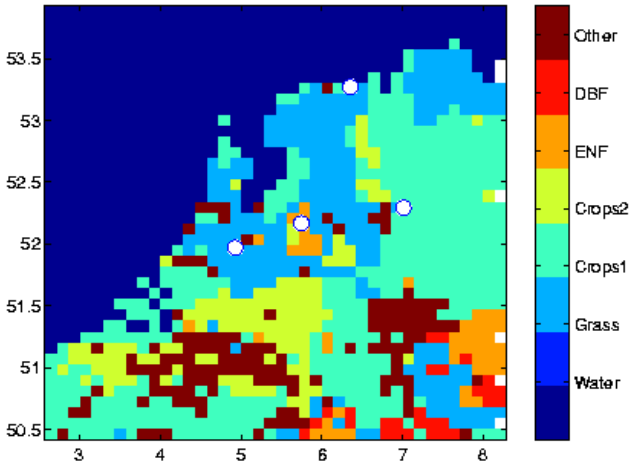

Fig. 1. The domain of the study showing the distribution of six different land-use categories according to the Corinne database, as well as 4 locations where continuous, calibrated measurements of CO_2 are available (Cabauw, west; Lutjewad, north; Loobos, middle; Hengelman; east). The area displayed, shown as lat long on the x and y axis is used in the RAMS meso-scale transport model, and is resolved on 10×10 km spatial resolution.

2.4 True fluxes and pseudo observations

Usually, in pseudo-data studies true fluxes are chosen as some realization of the underlying biosphere model (Zupanski et al., 2007; Schuh et al., 2009), as done in our control inversions. This choice is not very realistic when the structure of the underlying biosphere model itself is part of the inverse problem, and different structures will work better with different optimization strategies. For example, a parameter inversion when the truth was created with perturbed parameters will perform better than a β_{NEE} inversion against the same truth. To prevent this in our study, and also to make the pseudo-data study more realistic, we have used as "true" fluxes those from a different biosphere model.

Hourly biogenic respiration and photosynthesis flux fields from the FACEM model (Pieterse et al., 2007) were used. These were calculated at 6′ resolution and have a different underlying land-use description, different soil type and LAI map, and different meteorological driver data as summarized in Table 2. As a result, all six test inversions have to overcome a difference in model structure that causes the simulated pseudo-CO_2 time series to never perfectly match with the true NEE distribution (Fig. 2a, b). This situation mimics reality in which a biosphere model never grasps the full complexity nor heterogeneity of the true NEE distribution, which can be an important source of error (Kaminsky et al., 2001; Gerbig et al., 2006; Carvalhais et al., 2008).

To create pseudo CO_2 mixing ratio data, the true CO_2 fluxes are coupled to the atmospheric model RAMS. The simulated 3-D atmospheric CO_2 field is sampled at the locations where also in reality CO_2 mixing ratios observations are available (Fig. 1), where the highest observation level at the towers is used (Cabauw, 200 m; Lutjewad, 60 m; Loobos, 24 m; Hengelman; 18 m). Real observations will be used in a companion paper to obtain a real flux estimate. The inversions use hourly CO_2 mixing ratios sampled from the well mixed PBL between 11:00 and 16:00 UTC. In the control inversion we applied a model-data-mismatch of 0.2 ppm and with the FACEM truth we assumed a standard deviation of 1.2 ppm to account for the possible differences in the biosphere structures. The observation selection and uncertainty are the same in all different inversion methods. The total number of observations assimilated is 4 towers times 14 days times 6 h, equaling 336.

2.5 Prior flux covariance

A correct comparison between the different inversion options requires that the overall prior covariance of all options is equal. We require that the NEE, integrated over ecoregion and time, has the same ensemble-variance for all inversion options, and that the ratios between variance of ecoregion-time integrated respiration and GPP are also the same for the inversions. The standard is the parameter inversion, for which mean and variances are prescribed based on our previous work (Tolk et al., 2009; their Table 2), which was in turn based on an eddy covariance study (Groenendijk et al., 2011). For the other inversions, the prior covariance is scaled such that the above-mentioned similarity between the inversions is satisfied (see Appendix A for details).

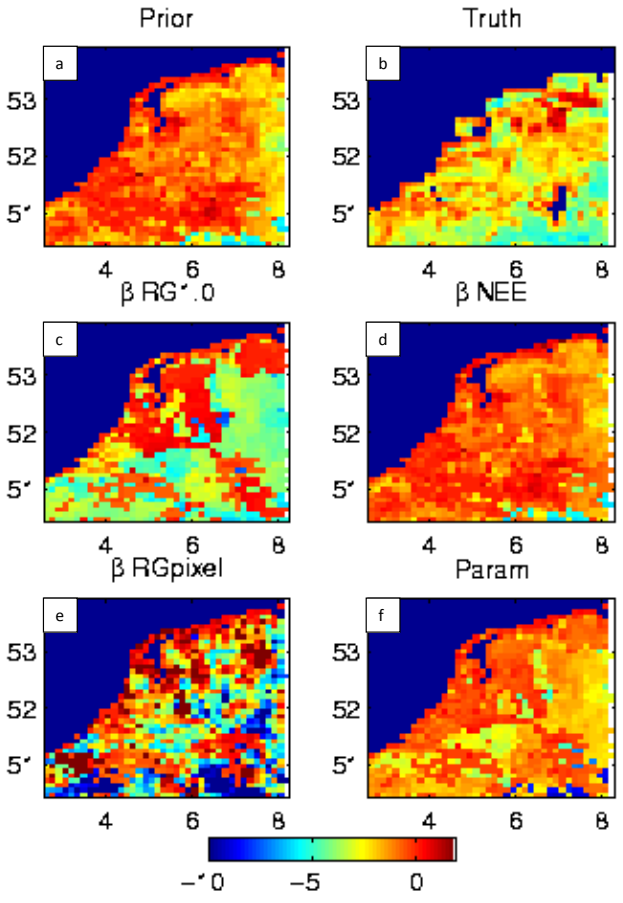

Fig. 2. The spatial distribution of the two-week average fluxes: top left: prior (from 5PM model); top right: "true" flux (from FACEM model); the other pictures show the averaged posterior estimates of the different inversion methods. Both sets of fluxes are used with hourly temporal resolution.

This approach is an important choice in our experiment design. It is interesting to note that equal variance of the time/ecoregion integrated NEE does not ensure the same uncertainty in the inversions at each point in space and time. Our choice of covariance treatment has ensured that (a) the inversions have equal covariance in the quantity that matters most to our CO_2 observations (NEE), (b) the spatial gradients in NEE variance between land-use types is conserved, and (c) the time integral covariance over the inversion are conserved suggesting that all inversions had an equal chance to find the mean NEE of the truth.

The χ^2 metric compares the a-priori model performance to the specified error structure by dividing the squared forecast residuals $(\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^2$ by the total covariance $(\mathbf{HPH}^T+\mathbf{R})$

of fluxes and measurements. It is thus a measure of the balance between expected skill and achieved skill. An innovation chi-squared of close to 1.0 indicates a correct balance, while smaller chi-squared values suggest that the model performed better than specified in the covariance structure and hence the inversion was conservative in its prescription of covariance (Michalak et al., 2004).The innovation χ^2 statistics indicate that flux and observation uncertainties were well balanced. The χ^2 values range from 0.34 to 0.78 (Table 1) indicating that the model skill was high relative to the assumed uncertainty. The inversions were thus conservative in their flux adjustments and not over-constrained.

3 Results

In this section we present our results according to the performance of all the different inversion methods against a set of different metrics that each highlight a particular aspect of the inverse results (Sects. 3.2–3.6). Before the overview of the performance of all methods is presented, the special behavior of one of the methods which is partly non-linear (the parameter inversion) is highlighted in Sect. 3.1. An overall assessment of each individual method is given in the discussion section.

3.1 Optimizations and non-linearity

Five out of six inversion systems used in this study are linear, in the sense that a Gaussian set of parameters will translate to a set of similar Gaussian set of CO_2 flux fields. The exception is the parameter inversion. Out of the four parameters chosen for optimization only the reference rate for respiration (R_{ref}) is linear, while the other ones are nonlinear. In addition, the model follows the Farquhar et al. (1980) photosynthesis limitation principle, in which either light or carboxyl becomes limiting for photosynthesis. The transition from one regime to another presents an important nonlinear step in the simulation of NEE.

Figure 3 shows the distribution of fluxes resulting from a chosen distribution of parameter values. It shows that the activation energy parameter (E_0) in particular affects the fluxes in a nonlinear fashion when the chosen value approaches zero. The carboxylation capacity ($Vc_{\max}$) and quantum yield (α_{jv}) parameters are weakly nonlinear across the chosen range, and the reference respiration rate (R_{10}) is fully linear.

The nonlinearity in the parameter inversion should in principle be dealt with in the ensemble system as it implicitly linearizes over the full model (H). However, we could clearly see the effect of imperfect linearization in our results. When we fed the posterior parameter set back into our flux model, and consequently propagated the solution through the RAMS transport model ($H(x^a)$) we did not obtain the distribution of CO_2 mixing ratios that the linearized inverse solution (Hx^a) had predicted. Generally, the propagated mean was further

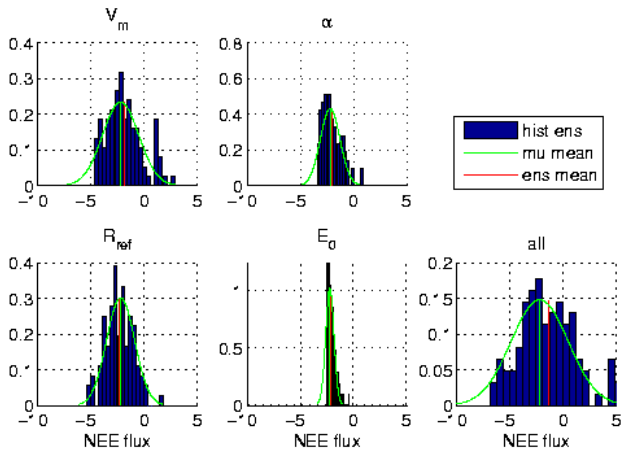

Fig. 3. The distribution of calculated NEE from 5PM given a Gaussian distribution of values for four different parameters in the model. The figure shows the linear translation of for instance $Vc_{\max}$ and R_{10} to a Gaussian shaped NEE distribution, and the non-linear response of α and E_0.

away from the observations than the linearized mean, and the propagated standard deviation of the ensemble was larger than the linearized one.

We explored this further with an offline inverse system that had only three parameters, as is further shown in Appendix C. A simple equation in which we varied the degree of non-linearity related the parameters to observations. The parameters were optimized against a truth obtained with one realization of the parameters, and some additional noise. We observed in our simple non-linear system that the propagated posterior parameter spread always correctly included the true parameter value. Additionally, the simplified model showed that introducing a non-linear parameter does not affect the ability of the inversion to return the correct values for the other, linear parameters in the model. Nonetheless, this simplified model showed the same behaviour with a poorer match to observations, and a larger spread in the concentration when using the propagated posterior parameters instead of the linearized model. This degraded performance of the linearized model is caused by the tails of the a-priori parameter probability density function (PDF), which do not follow the linearized propagation of the mean, or values close to that mean. This results in the observed larger spread in the non-linearly propagated CO_2, and also in the overconfidence of the posterior parameters (the linearized ensemble lacks spread). We found that this effect could be reduced in several ways: (1) by reducing the degree of non-linearity in the model, (2) by starting with a good a-priori parameter value around which the model is linearized, or (3) by limiting the uncertainty of the non-linear parameter to a space where the effect is mostly linear. These three strategies may be generally applicable in future studies attempting non-linear inversions.

3.2 Metric 1: domain integrated fluxes

One of the important metrics in evaluating the performance of the inversions is the ability to retrieve the NEE summed over time and space. This is a final goal of the inversions, but not the only one as more detail may be desired, and the summed values may hide opposing errors. These issues are addressed in the other metrics in the next sections. All except one of the inversions have managed to find an improved posterior time average flux for the whole domain (Fig. 4). The two results closest to the truth are from the βRG inversions without correlations (βRG0.0), or with partial correlations (βRG0.5; not shown), followed by the βRGpixel inversion and the parameter inversion. The β_{NEE} inversion was the only inversion that had a worse posterior time mean flux than prior time mean flux for the whole domain.

If we consider the root-mean-square-difference (RMSD) of the true and estimated domain average flux over time, a similar picture emerges: all inversions show an improvement (Table 3, first column). This agreement was expected under the current design of the study, in which many observations were available to constrain the hourly NEE. Again, the βRGpixel and βRG0.0/βRG0.5 perform best, but also the parameter inversion has captured the temporal structure of the domain total flux better than the prior. The β_{NEE} inversion struggles not only to find an improved time mean flux, but also provides a poor match to the hourly domain averaged time series.

The degree of RMSD improvement is rather small, especially considering the large number of observations available. Further investigation showed that the main limitation to improving the flux estimate is the different structure of the a-priori model (5PM) and the truth (FACEM). In the control inversion, where this model structure difference was not present, all methods (including the β_{NEE} inversion) gave much better time mean fluxes, and RMSD of the time series (Table 3). This suggests that synthetic studies using the same land surface model to generate and retrieve fluxes may overemphasize the ability of the model.

3.3 Metric 2: fluxes per land-use type

We can further separate the results above into each individual land-use class considered. Figure 4 shows these results. The most striking feature is the lack of improvement in mean NEE for most land-use classes in most inversions. This suggests that overall, the inversions have failed to find a correct distribution of time mean NEE *within* the domain (see also Fig. 2). The discrepancy is relatively weak for the parameter inversion and the βRG1.0 inversion, which show improved posterior mean β_{NEE} in 3 out of 6 land-use classes. The two inversions that did best on the total domain average NEE (βRG0.0 and βRG0.5) now reveal a lack of improvement within the dominant land-use classes, suggesting

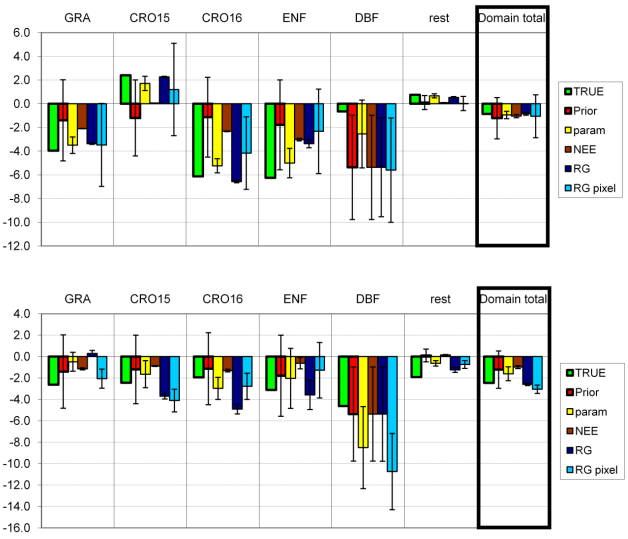

Fig. 4. True, prior and posterior NEE fluxes and their uncertainties for the different inversion methods averaged for the whole domain or per land use class [μmol m^{-2} s^{-1}]. **(a)** Results of the control inversion, where the truth fits the biosphere model, **(b)** results of the inversions with an imperfect biosphere model.

that they might have gotten domain total fluxes right for the wrong reason.

Assessing the RMSD of true and estimated flux time series per land-use type is consistent with the domain total picture of the previous paragraph: the parameter inversion and βRGpixel inversion generally show relatively low posterior RMSD and improve in 4 out of 6 land-use classes (Table 3). The other inversions improve substantially in RMSD in only 2 classes. An example of posterior flux performance for the largest land-use class (Crops1) is shown in Fig. 5. It shows that poor RMSD is caused mostly by an inability to capture the difference in the daytime signal at the different days, and that nighttime NEE is poorly simulated in most of the inversions.

Considering that the instantaneous fluxes per land-use class are most closely connected to the CO_2 mole fraction observations, the performance of the inversions is disappointing, and also alarming. Also in this metric the performance of the same inversion methods against the control fluxes is much better, and agrees with the expectation that domain total and land-use fluxes improve in time (both flux average and RMSD; Fig. 4 and Table 3).

3.4 Metric 3: uncertainty estimate

The contrasting performance of most inversions for individual land-use classes (bad) versus the domain integral (reasonable) suggests significant spatial correlations between classes, with canceling flux errors. Here we investigate the posterior uncertainties.

Table 3. The performance of the six inverse methods expressed as the RMSD of the optimized NEE time series compared to the truth, for the total domain and for each separated land use class [μmol m^{-2} s^{-1}]. The upper part of the table shows the results of the control inversions, the lower part shows the results of the inversions where the truth is generated with a different biosphere model than the priors.

			Temporal RMSD of the NEE flux						
	Prior/Posterior	Inversion name	Domain total	Grass	Crops1	Crops2	ENF	DBF	Rest
	Prior		0.9	3.7	5.5	6.9	6.9	6.3	0.71
Truth fits biosphere model	Posterior	β_{NEE}	0.7	2.4	2.7	4.1	4.3	6.3	0.7
		βRG0.0	0.6	1.5	0.6	1.3	5.6	6.3	0.2
		βRG0.5	0.5	1.5	0.6	1.4	5.8	6.3	0.3
		βRG1.0	0.8	3.4	4.0	4.9	4.6	6.3	0.8
		βRGpixel	0.7	1.6	2.4	3.5	6.1	6.7	0.8
		Parameter	0.1	0.6	1.0	1.3	1.9	2.8	0.1
	Prior		2.7	2.9	2.9	3.1	3.7	4.0	3.4
Imperfect biosphere model	Posterior	β_{NEE}	2.6	2.8	2.6	3.6	4.7	4.0	3.2
		βRG0.0	2.0	4.5	2.0	3.5	2.9	4.0	2.5
		βRG0.5	2.0	4.5	2.0	3.1	5.0	4.0	2.6
		βRG1.0	2.4	5.6	2.3	3.9	4.0	4.0	3.2
		βRGpixel	1.8	2.7	2.5	2.5	3.8	9.9	2.8
		Parameter	2.1	3.2	1.9	2.0	3.5	5.8	2.9

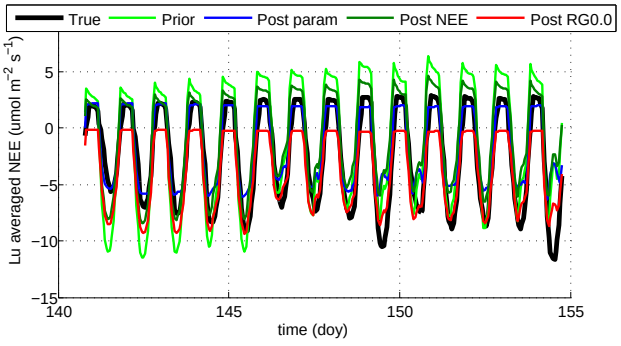

Fig. 5. Time series of aggregated true, prior, and posterior fluxes for the dominant land use type in the domain, Crops1. The figure illustrates the different temporal behavior of each solution on hourly time scales, leading to differences in two-week aggregated RMSD.

The inversions based on multiplication factors per land-use class (β_{NEE} and βRG) all produce posterior uncertainties that are too small. This means they do at most locations not encompass the true flux within one or two standard deviations (Fig. 6). This might have been expected since the degrees of freedom in the inversion are much smaller (∼6–62) than the number of observations assimilated (∼336). However, the χ^2 of innovations suggests a fair balance between CO_2 residuals ($\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x}^b)$) and prescribed uncertainty ($\mathbf{HPH}^T+\mathbf{R}$) and the posterior uncertainty does not scale directly with the degrees of freedom in each inversion. Thus, the relatively large number of observations does not seem the main cause of too low posterior uncertainty.

In fact, when we reduced the number of assimilated observations to 1 per day we saw only minor effects on the posterior variances. Only if we reduced the number of observations and additionally increased the model-data mismatch did we see an increase in posterior variances for the individual land-use classes. But even then the domain total variance remained much too small to accommodate the large difference in flux mean. The small posterior uncertainty is thus not simply an artifact of the inverse method setup.

If we consider the control inversion, the posterior flux estimates are much better for all inversions, and 4 out of 6 inversions report ±1 sigma uncertainties that include the truth (Fig. 4). This again points at an important role of model structure in determining the outcome of an inversion, in this case leading to overconfidence that the true value is retrieved.

The parameter and the βRGpixel inversions generally produce error estimates that are conservative and more realistic compared to the other inversions, in the sense that the posterior estimate (Fig. 6) is in large parts of the domain within 2 sigma of the truth and they encompass the truth within ±2.5 sigma in 5 out of 6 land use classes. Note that in the 6th land use class (rest) also the prior uncertainty is too low, and a realistic posterior uncertainty could therefore not be expected for this land use class.

Posterior covariances are large in all inversions. In the βRG inversions the covariances occur between β parameters for R and GPP. The covariances reveal an inability to separate the effect of photosynthesis and respiration based on CO_2 alone. This was demonstrated in earlier work too (Ahmadov et al., 2009; Schuh et al., 2009). This type of

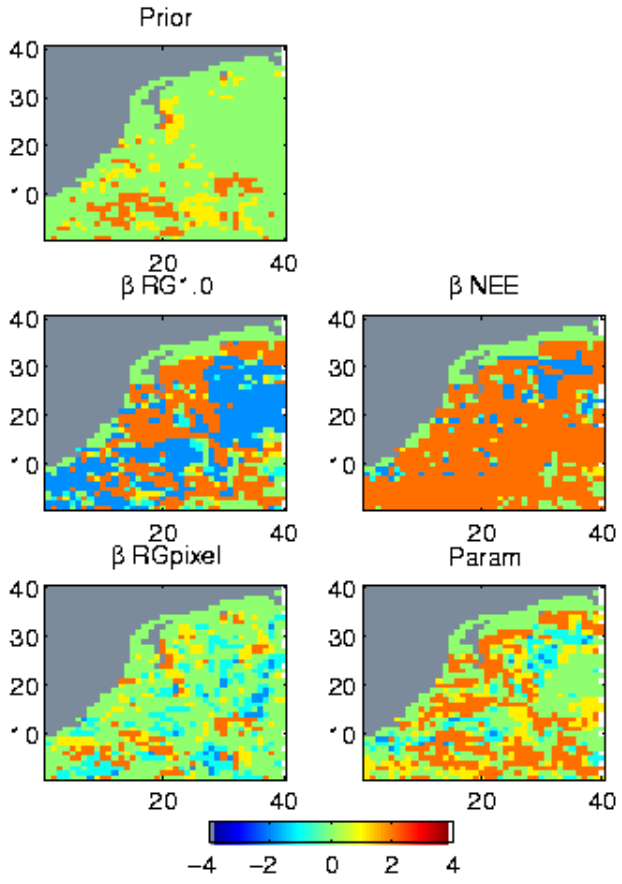

Fig. 6. Spatial maps of the posterior flux residuals (true-minus-optimized fluxes) scaled by the posterior standard deviation of the flux estimate. Green colors indicate that the true flux is within ± 1 standard deviation of the estimated flux, yellow/light blue in dicate within 2 or -2 standard deviation respectively and orange/blue colors are outside this range. The figure shows the general tendency of the inversions to be overconfident, because structural differences between 5PM and FACEM are not taken into account in the treatment of uncertainty.

covariances is more pronounced than those between β's for different land-use classes.

3.5 Metric 4: match to CO_2 mole fractions

All six inverse methods reduce the mismatch between pseudo-observations and simulated CO_2 mole fractions for those observations that were assimilated. This is expected from an inverse calculation. Not all methods obtain equal RMSD though, as each inversion adjusts fluxes differently. Table 4 shows that the posterior RMSD in CO_2 mole fractions is largest for the β_{NEE} optimization and smallest for the βRGpixel optimization.

The χ^2 of innovations (Table 1), which indicates the balance between a-priori mismatch and assumed uncertainties, is smaller for the β_{NEE} and βRG1.0 inversion than for all other optimizations per ecoregion. This is because these two inversions had a much larger uncertainty in hourly NEE than the other inversions (necessary to maintain the same uncertainty over the full period, see Sect. 2.5), translating to more simulated uncertainty in CO_2 mole fractions.

More interesting is the comparison to CO_2 mole fractions that were not assimilated as they show the performance against independent data. The second column of Table 4 shows that only three inversions perform better when assessed against the full CO_2 time series, while for three others it deteriorates. This is an important result that suggests that the estimated NEE field reflects only a limited part of the CO_2 mole fraction time series, as a result of the daytime-only subsampling we used.

Moreover, the β_{NEE} and βRG1.0 inversion – that improve RMSD of non-assimilated CO_2 – show the opposite result in the control inversion, suggesting that the improvement in the real inversion was fortuitous. This leaves only the parameter inversion to improve the RMSD of assimilated and independent CO_2 observations, both in the control inversion and the real inversion.

For the βRG inversions the good performance on the temporal structure of NEE (Tables 3 and 4) contrasts with the poor RMSD of the non-assimilated CO_2 observations . The reason is that the posterior nighttime flux is simulated poorly which does not affect the RMSD of fluxes much (the prior was also rather poor at night, and the nighttime NEE is relatively small compared to the daytime NEE), but is strongly amplified in CO_2 concentration space due to the shallow nighttime boundary layer. Now, the posterior variances in β_{resp} and β_{GPP} mentioned in the previous paragraph really come to expression: they give an acceptable aggregated flux (NEE) but at the expense of incorrect nighttime CO_2 mixing ratios. The propagation of this incorrect nighttime CO_2 signal through the domain will likely result in compensating signals further downwind, or later in time. These effects are not visible yet in our short experiment though.

The results above suggest: (1) that nighttime CO_2 observations are needed to separate respiration from photosynthesis fluxes, and (2) that interpretation of these observations depends critically on the adequate simulation of the nocturnal boundary layer.

3.6 Limitations of the inversions

The NEE inversion is the most simple inversion method tested in this study, the number of degrees of freedom is with 6 the smallest compared to the other inversion methods (Table 1). The results of this inversion did not correctly fit with the true fluxes and were overconfident. This inversion appeared to lack the flexibility to capture the correct fluxes. In the other inversions tested in this study the flexibility is increased in a number of ways indicating the limitations of the inversion.

Table 4. The performance of the six inverse methods expressed as the RMSD of the time series of the atmospheric CO_2 mixing ratio related to the optimized NEE flux fields compared to the pseudo observations [ppm]. For the times the observations are used in the inversions (daytime only) and for the total time series (all hours). The left part of the table shows the results of the control inversions, the right part shows the results of the inversions where the truth is generated with a different biosphere model than the priors.

Prior/Posterior	Inversion name	CO_2 mixing ratio RMSE			
		Truth fits biosphere model		Imperfect biosphere model	
		Daytime only	All hours	Daytime only	All hours
Prior		2.3	2.5	2.6	4.3
Posterior	β_{NEE}	1.1	5.1	2.5	3.6
	βRG0.0	0.4	0.9	1.1	6.0
	βRG0.5	0.4	0.9	1.1	6.5
	βRG1.0	1.3	6.8	1.9	4.2
	βRGpixel	0.5	1.0	0.8	5.6
	Parameter	0.4	0.5	1.6	2.8

In the βRG inversions the flexibility is increased compared to the β_{NEE} inversion as the respiration and GPP can be separately optimized. These βRG inversions provided a better estimate of the domain total fluxes than the β_{NEE} inversion (Fig. 4). The different covariance strengths between β_{resp} and β_{GPP} that were applied resulted in different d.o.f's. The performance of the inversions with 9 and 11 d.o.f. were comparable.

In the control inversions, where the truth fits the biosphere model inversions, the βRG inversion performed correctly improving the fluxes and the RMSE, for both the domain total fluxes and the fluxes per ecoregion, with a realistic uncertainty estimate (e.g. Fig. 4a). However, when the structure of the truth and the prior were non-similar, the flux distribution over space and time was not well captured. The flexibility of the system in time and/or space appeared to be limiting for the inversion. Below, first the temporal limitation of the βRG inversion is discussed based on an analysis of the results of the βRG inversion and the more flexible in time parameter inversion, and secondly the spatial limitations of the inversion are discussed based on the results of the pixel inversion.

The βRG inversion performed badly when looking at the CO_2 mixing ratio RMSE for all hours (Table 4). This is caused by a difficulty in capturing the nighttime fluxes (Fig. 5). The CO_2 mixing ratios during the day, which are used to constrain the fluxes, contained a mixed signal of the respiration and GPP fluxes. The results suggest that the difference between the temporal structure of the truth and the prior caused the respiration fluxes to be altered by the inversion to overcome the difference in the flux structure during daytime, but thereby the nighttime flux estimate was worsened (Fig. 5). In an additional test, in which the inversions were performed for three days only instead of a 15 day period the effect on the nighttime fluxes was smaller. In these three days the weather was steady anti-cyclonic. This suggests that the structure mismatch of the fluxes between the prior and the truth over the 15 days time period due to differences in the response patterns to cyclonic and anti-cyclonic weather influences the estimate of the fluxes, next to the influence of the different diurnal pattern of the fluxes between the prior and the truth

In the parameter inversion, the flexibility to change the fluxes in time is increased. In this inversion four instead of the previous one (in the β_{NEE} inversion) or two (in the βRG inversion) parameters can be altered by the inversion for each ecoregion, which means a d.o.f. of 22. The additional temporal flexibility improved the results and was able to avoid the incorrect change in the nocturnal fluxes that was seen in the βRG inversions (Fig. 5, Table 4).

In the inversions where the structure of the truth did not fit the structure of the prior, the βRG inversions showed a lack of improvement within the dominant land-use classes. This could be caused by the mismatch of variability of the fluxes within one land use type. Therefore, in the βRGpixel inversion, the flexibility of the system to alter the fluxes in space was increased. Instead of one adjustable β_{resp} and β_{GPP} per landuse type, these parameters could be optimized for every pixel, with a correlation length scale of 100 km, which increased the d.o.f. to 62. Despite the increased flexibility, the inversion still could not capture the fluxes per land use type. Nonetheless, the βRGpixel inversion provided acceptable results because it was realistic in its uncertainty estimate, only changing the fluxes in a more limited part of the domain.

The fact that increasing the temporal flexibility in the parameter inversion and the spatial flexibility in the pixel inversion does not fully solve the problems faced in the inversions suggests that both temporal and spatial structure differences between the truth and the prior were limiting to obtain results at a smaller scales than a few hundred kilometer. For the coarser scale, which is in this inversion the full domain,

the inversions performed well despite these limitations. The structure mismatch is thus limiting for the detailed analysis, and not so much for the aggregated results.

4 Discussion

With the steady expansion of (continuous) CO_2 observation sites comes a tendency to estimate carbon exchange patterns at increasingly higher resolutions. Inversion methods are thereto equipped with state-of-the-art meso-scale transport models and detailed ecosystem models. The methodology to optimize carbon exchange at regional scales is often transferred from existing global systems, thereby inheriting their known strengths and weaknesses. At these regional scales however, other considerations come into play that potentially turn weaknesses into strengths and vice versa. An example from this work is the scaling of the diurnal cycle amplitude in the NEE inversion. This was a rather robust way to maintain a balance between respiration and photosynthesis over long time periods in global inversions (Peters et al., 2010), but seems to fail on the smaller scales. We suspect that this is caused by the dominance of the CO_2 diurnal cycle as a source of information for the inversion, and the sensitivity of the inversions to changes in the response to short term weather influences, while in continental scale inversions the average CO_2 mixing ratios were more controlling.

In contrast, we see that at the regional scale the shortcomings of biosphere model structure is expressed more strongly than at continental scales. The potential to alias CO_2 signals onto the wrong parameter because they are simply not reproducible with the prescribed structure was demonstrated and discussed convincingly in Carvalhais et al. (2008). Our study corroborates their conclusions, and shows that spatial mismatches in model structure can lead to incorrect mean flux estimates, with error bars that are overconfident. Since there is currently no metric, nor a place in the inversion for this type of error to be included, we suggest that model structure is assessed critically when optimizing biosphere model parameters or precalculated flux patterns on regional scales. Comparison of resulting fluxes and uncertainties against independent data (i.e. not used on the inversion) is one way to detect model structure errors for inversions using real data (e.g. Lauvaux et al., 2009).

In our assessment of the inversions based on biosphere model parameter optimizations we have seen an important role of non-linearity in the model equations, similar to previous studies (Trudinger et al., 2007; Scholze et al., 2007; Rüdiger et al., 2010). In this regional optimization based on CO_2 mixing ratio observations, we noticed that the non-linear model parameters in particular were difficult to constrain. To use them correctly, they required a good first-guess, a small uncertainty, or a full non-linear model propagation of the solution (rather than a linearized one). These parameters were often also the ones least constrained by daytime atmospheric CO_2, and thus likely suffered from the specific setup of our experiment. Since the estimation of non-linear parameters did not affect the retrieval of the linear ones, we suspect that a different setup (other observations, such as water, energy, and CO_2 fluxes and isotopes, more temporal constraints) might perform better.

In contrast, estimates of bias scaling factors on photosynthesis and respiration remains linear, depend less on model structure, and have more freedom to use the diurnal cycle information on regional scales. In our studies it also proves a good alternative to the NEE scaling and the biosphere model parameter optimization. Also here, the daytime CO_2 sampling scheme used makes it difficult to confidently separate the two processes. Simply including night time CO_2 mixing ratio observations is however not an option, because of the limited skill of transport models to simulate the stable boundary layer (Tolk et al., 2009; Gerbig et al., 2008; Law et al., 2008; Steeneveld et al., 2008). Also, there is a large potential for erroneous photosynthesis and respiration bias scaling factors to propagate in time, and destabilize the inversion after a few weeks. The short time window used in this study does not incorporate this complication.

Perhaps the most important result from our pseudo-data tests is that despite the relatively large number of observations, the high resolution of the (perfect) transport model, and the increased freedom to fit spatiotemporal flux patterns, we still have not achieved a correct estimate of carbon exchange at the local scale. Similar to previous studies (Carouge et al., 2010; Schuh et al., 2009; Gerbig et al., 2006; Ahmadov et al., 2009), we find that significant aggregation of results is needed to come to robust numbers. The aggregation scale is on the order of 100×100 km as in previous work. This suggests that the simple translation of methods from the large scale to the small scale might not be sufficient. A re-evaluation of inversion methods might be needed, with an eye for nonlinear behavior, model structure, and multiple constraints. In that respect, recent work where model mean structure is relaxed in favor of extensive covariance structure based on multiple auxiliary datasets (Michalak et al., 2004; Gourdji et al., 2010; Yadav et al., 2010) is of great interest.

5 Conclusions

We started this paper asking which of six inversion approaches is the most suited for a regional inversion, and what the pitfalls are of each method. From our analysis we have learned that:

- With prior fluxes that have the same structure as the true fluxes, all inversion methods improved the estimate of the NEE, both for the domain total fluxes as for the fluxes per ecoregion.
- When the structure of the priors differed from that of the truth, the full domain estimates improved with all

inversions except for the β_{NEE} option, but all inversion approaches had difficulties in obtaining the fluxes per ecoregion.

– Model structure is therefore an important consideration for inverse estimates that can lead to incorrect spatiotemporal patterns of fluxes, and overconfidence in posterior results. An assessment of model structure error, and its inclusion in the quoted uncertainty would make any regional inversion more plausible.

– Inversions that scale NEE from prescribed spatiotemporal patterns are most susceptible to these errors (which include aggregation errors), and perform worst in the realistic tests presented. We do not recommend using this method for regional NEE estimates.

– Inversions that separately estimate photosynthesis and respiration perform better on NEE, at least on these short time scales, even though they cannot obtain realistic gross flux estimates, which might lead to problems later. We recommend to use them only if the realism of the gross fluxes can be assessed after the inversion, or maintained by other means such as through nighttime observations of fluxes or CO_2 mixing ratios.

– The results with the smallest deviations from the pseudo-truth over all metrics were obtained when the land-use class concept was applied least strictly by allowing spatial variations in bias corrections on gross fluxes (RGpixel), or when the bias parameter approach was abandoned altogether such as in the parameter inversion. Nonetheless, also these inversions had difficulties in estimating the specific fluxes per ecoregion.

– The parameter optimization approach has some appealing features. However, it can only be used if the non-linear behavior of the system is dealt with.

– When optimizing non-linear parameters we recommend to (a) start from a good a-priori mean estimate, (b) keep the uncertainty on the parameter small, and (c) check posterior results carefully using the full non-linear model.

Appendix A

Ensemble Kalman filter method

All the inversions are performed with the Ensemble Kalman filter (EnKF). In this Bayesian approach the optimum value between the prior knowledge and the information in the observations is established by minimizing the cost function:

$$J = (\boldsymbol{y} - H(\boldsymbol{x}))^T \mathbf{R}^{-1} (\boldsymbol{y} - H(\boldsymbol{x})) + \left(\boldsymbol{x} - \boldsymbol{x}_{\text{prior}}\right)^T \mathbf{P}^{-1} \left(\boldsymbol{x} - \boldsymbol{x}_{\text{prior}}\right) \quad \text{(A1)}$$

In which $\boldsymbol{x}$ denotes the state vector (the biospheric parameters or the β's in the inversions), $\boldsymbol{y}$ the observation vector (CO_2 mixing ratios), $\mathbf{P}$ the error covariance matrix of $\boldsymbol{x}_{\text{prior}}$, $\mathbf{R}$ the error covariance matrix of the observations. $\boldsymbol{H}$ is the observation operator, that contains the influence of the variables in the state vector ($\boldsymbol{x}$) on the CO_2 mixing ratio at the observation locations. The optimum posterior state vector that minimizes the cost function, and its error covariance matrix are:

$$\boldsymbol{x}_{\text{post}} = \boldsymbol{x}_{\text{prior}} + \mathbf{K}(\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}_{\text{prior}})) \quad \text{(A2)}$$

$$\mathbf{P}_{\text{post}} = (\mathbf{I} - \mathbf{KH})\mathbf{P}_{\text{prior}} \quad \text{(A3)}$$

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

$$\mathbf{K} = (\mathbf{P}_{\text{prior}}\mathbf{H}^T)(\mathbf{H}\mathbf{P}_{\text{prior}}\mathbf{H}^T + \mathbf{R})^{-1} \quad \text{(A4)}$$

In the Ensemble Kalman Filter method the information in the error covariance matrix $\mathbf{P}$, and its projection in observational space $\mathbf{HP}$ and $\mathbf{HPH}^T$ are not calculated based on independently determined $\mathbf{H}$ and $\mathbf{P}$ matrixes. Instead of the full calculation, an ensemble of state vectors that represent the statistical properties of $\boldsymbol{x}_{\text{prior}}$ and $\mathbf{P}_{\text{prior}}$ is used. Normally this is done to reduce the size of the matrixes, which may become very large if the amount of unknowns is large. Here, the Ensemble Kalman Filter is applied because of another advantage: this method of directly calculating $\mathbf{HP}$ and $\mathbf{HPH}^T$ from $\boldsymbol{x}$ and $\boldsymbol{H}(\boldsymbol{x})$ allows the use of a non-linear relation between the parameters ($\boldsymbol{x}$) and the CO_2 mixing ratios ($\mathbf{H}(\boldsymbol{x})$) as is the case in the parameter inversion.

The ensemble of state vectors, with N ensemble members, was created such that the normalized ensemble of deviations define the columns of matrix $\mathbf{X}$ (Peters et al., 2005):

$$\mathbf{X} = \frac{1}{\sqrt{N-1}}(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}, \boldsymbol{x}_2 - \bar{\boldsymbol{x}}, ..., \boldsymbol{x}_N - \bar{\boldsymbol{x}}) \quad \text{(A5)}$$

which is the square root of the covariance matrix:

$$\mathbf{P} = \mathbf{X}\mathbf{X}^T \quad \text{(A6)}$$

In the ensemble members x contains the parameter values of the biosphere model in case of the parameter inversion, or the multiplication factors β_{RESP} and β_{GPP} in the βRG inversions, or the multiplication factor β_{NEE} in the β_{NEE} inversion. All inversions are performed with ensembles containing 100 ensemble members. For each ensemble member the corresponding CO_2 mixing ratios at the observation locations were calculated. This is done in the coupled biosphere-atmosphere model (5PM+ coupled to B-RAMS3.2). Thus an ensemble of CO_2 mixing ratios was created:

$$\mathbf{H}(\mathbf{X}) = [\boldsymbol{H}(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}), \boldsymbol{H}(\boldsymbol{x}_2 - \bar{\boldsymbol{x}}), ..., \boldsymbol{H}(\boldsymbol{x}_N - \bar{\boldsymbol{x}})] \quad \text{(A7)}$$

From the ensemble of state vectors ($\mathbf{X}$) and the ensemble of corresponding CO_2 mixing ratios ($\mathbf{H}(\mathbf{X})$) the Kalman gain

matrix, and the posterior optimized values including their uncertainty was calculated with Eqs. (A2–A4) combined with (Peters et al., 2005):

$$\mathbf{HPH}^T = \frac{1}{N-1}\mathbf{H}(\mathbf{X})(\mathbf{H}(\mathbf{X}))^T \tag{A8}$$

and

$$\mathbf{PH}^T = \frac{1}{N-1}\mathbf{X}(\mathbf{H}(\mathbf{X}))^T \tag{A9}$$

Due to the use of an ensemble instead of the full KF, small extra covariances may be created. The impact of these small artificial covariances on the inversion result can be diminished by localization. In this study we used the localization method and values established in Zupanski et al. (2007), a threshold value for the ratio between the prior uncertainty and the posterior uncertainty of 1.05 was applied.

In the control inversions in this study the observation vector ($\boldsymbol{y}$) was created based on one realization of the parameter state vector by selecting one of the columns of ensemble $\mathbf{X}$. In the other inversions the observation vector ($\boldsymbol{y}$) was based on the hourly NEE flux fields created by the biosphere model FACEM (Pieterse et al., 2007). In both cases the corresponding CO_2 mixing ratios were calculated with the atmospheric model RAMS in which the sea, the fossil fuel, and the boundary fluxes of CO_2 were kept constant.

The size of the state vector differed in the different inversion methods, in the parameter inversion it had a dimension of 24 (4 parameters times 6 land use classes), in the βRG inversions its dimension was 12 (twice the number of land use classes), in the β_{NEE} inversion its size was 6 (once the number of land use classes) and in the βRGpixel inversion its size was 2218 (twice the number of land pixels). The dimension of the observation vector was the same in all inversions with 336.

For the parameter inversion, βRG0.0 and β_{NEE} inversions all off-diagonals in $\mathbf{P}_{\mathrm{prior}}$ were zero. Additional inversions were performed with a correlation between β_{resp} and β_{GPP} of 0.5 and 1.0. In this case all off-diagonals were zero except the ones denoting the correlations between β_{resp} and β_{GPP} of the same land use type. Also in the pixel inversions only correlations within one land use type are applied with correlations calculated based on distance (D) with a length scale (L) of 100 km:

$$\mathbf{P}_{i,j} = \sigma_i \sigma_j \exp\left(-\frac{D}{L}\right) \tag{A10}$$

By the way, a comparable correlation length (130 km) was found for the prior-truth residuals. The means and variances for the parameter inversion are prescribed based on Tolk et al. (2009). The other inversions use an ensemble of β's with mean one and variances which are scaled to achieve the required similarity between the inversions (Sect. 2.5). First the uncertainty related to the respiration fluxes on the one hand and to the GPP fluxes on the other hand were separated. This was done by running the biosphere model with two different ensembles: (1) containing only variations in the parameters determining respiration and (2) containing only variations in the parameters determining GPP. To convert this to the variance related with the β factors, each ensemble member is scaled with the flux per ecoregion, separately for respiration and GPP in the βRG inversions and for NEE as a whole for the βNEE inversion. This ensures that the ratio between the uncertainty in respiration and GPP per ecoregion is the same in all inversion options. In the inversion where applicable correlations were added to $\mathbf{P}$. The new variances were subsequently scaled with a multiplication factor per ecoregion, with the same multiplication factor for β_{resp} and β_{GPP}. These multiplication factors were chosen such that the uncertainty in NEE integrated over ecoregion and time became the same in all inversion options, taking into account the correlations between respiration and GPP in the βRG0.5 and βRG1.0 options and the reduced correlation within one ecoregion in the pixel inversion.

Appendix B

Biosphere model 5PM

The biosphere model used in this study to calculate the prior NEE fluxes is 5PM (Groenendijk et al., 2011) extended with the use of LAI to upscale the fluxes from leaf to the canopy scale. In this model the photosynthesis is calculated based on the Farquhar model (Farquhar et al., 1980) and heterotrophic respiration is based on the relationship by Lloyd and Taylor (1994).

In the Farquhar approach assimilation of CO_2 by the vegetation is either limited by the amount of radiation or by the availability of the enzyme Rubisco, which is involved in the conversion of CO_2 into glucose and oxygen. Photosynthesis is formulated as the minimum of the light limited (w_{j}) or enzyme limited assimilation rate (w_{c}), corrected for the maintenance respiration of the vegetation (R_{d}):

$$A = \min(w_{\mathrm{c}}, w_{\mathrm{j}}) - R_{\mathrm{d}} \tag{B1}$$

The assimilation rate depends on the CO_2 concentration inside the leaf available for photosynthesis (C_{i}), the internal oxygen concentration (O), the compensation point for CO_2 (Γ^*) and the Michaelis-Menten parameters for CO_2 (K_{c}) and O_2 (K_{o}). The latter are temperature dependent. The first option, Rubisco-limited assimilation is calculated as:

$$w_{\mathrm{c}} = Vc_{\max}\frac{C_{\mathrm{i}} - \Gamma^*}{C_{\mathrm{i}} + K_{\mathrm{c}}\left(1 + \frac{O}{K_{\mathrm{o}}}\right)} \tag{B2}$$

where $Vc_{\max}$ is the maximum carboxylation capacity ($\mu\mathrm{mol\,m^{-2}\,s^{-1}}$). The second option, light limited assimilation is calculated as:

$$w_{\mathrm{j}} = J\frac{C_{\mathrm{i}} - \Gamma^*}{4(C_{\mathrm{i}} + 2\Gamma^*)} \tag{B3}$$

Table B1. Prior values of the parameters in the biosphere model for the different land use classes.

Assimilation		
	$Vc_{\max}$ µmol m^{-2} s^{-1}	α mol mol^{-1}
Grass	70	0.4
Crops	100	0.4
Needle leaf forest	80	0.5
Broadleaf forest	100	0.5
Urban vegetation	80	0.4
Respiration		
	$E_0/\Re$	R_{10}
	K	µmol m^{-2} s^{-1}
All land use classes	200 ± 110	4.0 ± 1.2

where J is the electron yield, specified by:

$$J = \frac{\alpha I_{\mathrm{PAR}} J_{\mathrm{m}}}{\alpha I_{\mathrm{PAR}} + 2.1 J_{\mathrm{m}}} \tag{B4}$$

in which I_{PAR} is the photosynthetic active radiation (µmol photons m^{-2} s^{-1}), J_{m} the maximum potential electron transport rate (µmol m^{-2} s^{-1}) and α the quantum yield (mol mol^{-1}). Assumed is that the plants aim for an optimum between the energy allocated to the potential electron transport rate and to the carboxylation capacity, and J_{m} is linked to V_{m} by (Collatz et al., 1991):

$$J_{\mathrm{m}} = 2.5 V c_{\max} \tag{B5}$$

Leaf internal CO_2 is estimated with the method described in Arneth et al. (2002) in which the value for lambda was kept constant at 700 mol mol^{-1}. The atmospheric CO_2 mixing ratio is assumed to be 380 ppm during photosynthesis. Integration of the photosynthetic flux to the full canopy is based on MODIS leaf area index (LAI) observations (Sellers et al., 1996):

$$A_{\mathrm{c}} = \Pi A_{n0}, \text{ with } \Pi = \frac{1 - e^{-\bar{k}\mathrm{LAI}}}{\bar{k}} \tag{B6}$$

where A is the assimilation, subscript $n0$ refers to leafs at the top of the canopy, subscript "c" refers to total canopy and k is the PAR extinction coefficient.

Respiration is calculated with the temperature dependent relationship by Lloyd and Taylor (1994):

$$R = R_{10}\ e^{\frac{E_0}{\Re}\left(\frac{1}{283.15 - T_0} - \frac{1}{T - T_0}\right)} \tag{B7}$$

where R_{10} is the respiration rate at a reference temperature of 10 °C, $\frac{E_0}{\Re}$ is the activation energy divided by the universal gas constant, T_0 is a constant of 227.13 K and T is soil temperature.

In a previous study (Tolk et al., 2009; Groenendijk et al., 2011) the main parameters of this model ($Vc_{\max}$, α, R_{10} and E_0) were optimized for the full canopy based on a large number of Fluxnet observations (Baldocchi et al., 2001). To determine our prior fluxes, we applied parameter values optimized for the temperate zone, for the period of May–July for all years (Tabel B1).

The relationship between the parameters in the biosphere model and the NEE fluxes is for most parameters non-linear. The strength of this non-linearity was estimated for each parameter, by running the biosphere model with ensemble of Gaussian distributed parameter values and showing the accompanying distribution of the NEE fluxes (Fig. 3). For all parameters except R_{10} it was shown that the resulting NEE distribution was non-Gaussian, affecting the inversions if the prior had a large deviation from the true values.

Appendix C

Simple non-linear model

To test the impact of non-linearity on the inversions we applied simple forward models that calculate a series of observations from a triplet (a, b, c) of arbitrary parameters. Subsequently, several inversion methods were used to estimate (a, b, c) from the observations: a regular minimum least-squares without priors, a full Bayesian solution with a Kalman filter, a serial ensemble KF, and a matrix based ensemble KF. In the Bayesian solutions, the three parameters had prior values that varied between realistic and unrealistic relative to the truth. Also, we varied the degree of nonlinearity in the forecast model (from fully linear to strongly nonlinear).

We applied all methods to a fully linear problem first, and confirmed that each estimation method gave the same (correct) result as long as enough observations were available. Prior values that were reasonably chosen (i.e. with enough uncertainty to accommodate the truth) were thus correctly modified. The statistics of the posterior solution were also as expected: uncertainty on all 3 parameters was reduced in accordance with the specified noise, and the propagated posterior solution satisfied the observations to within the specified uncertainty.

Next, we applied the ensemble KF (the only system to handle nonlinear problems) to the nonlinear function $f(x) = a + \sin(b)x + cx^2$. We noticed here first that the mean of the linear parameters a, c was estimated correctly, but the mean of the nonlinear one (b) was not. Uncertainty on the nonlinear parameter was also miscalculated: the truth was far outside the posterior error. Moreover, we noticed that if we placed the posterior parameter values back in the nonlinear model, the match to the observations was much worse than the statistics of the filter suggested. This was because the linearization that is contained in the ensemble method was not

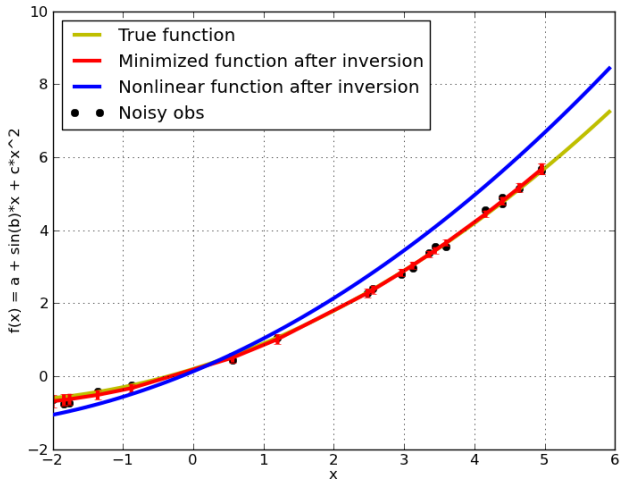

Fig. C1. An illustration of the effect of non-linearity on the results of the inversion.

able to overcome the nonlinearity and therefore mispredicted the mixing ratios and their spread given a set of parameters.

In Fig. C1 we show this feature for the typical nonlinear estimation problem as above. The yellow line is the true function, which. This we "observe" at the 20 black dots, and then we add a little bit of noise. The noisy observations are fed to the ensemble KF to estimate the 3 parameters (a, b, c) that underly the yellow curve. The red line (± 1 standard deviation) is the match to the observations that the system thinks it will achieve given would correspond to the ensemble statistics and its posterior estimate. This coincides with the true curve in yellow. But when the posterior values (a, b, c) are fed into the function $f(x)$, the blue curve is the actual result. This is much less accurate than was predicted by the filter, and actually outside the specified uncertainty range.

The above problem could be remedied through the exact solutions we suggest in the main text: making the more problem more linear, starting from a better prior, or reducing uncertainty on the nonlinear parameter. The figure moreover led us to suggest to always use the full nonlinear model to check your the accuracy of the result, rather than the filter statistics.

Acknowledgements. This work has been executed in the framework of the Dutch project "Climate changes Spatial Planning", BSIK-ME2 and the Carboeurope Regional Component (GOCE CT2003 505572). W. Peters was supported by NWO VIDI grant 864.08.012. We acknowledge A. Vermeulen for providing the FACEM biosphere flux data.

Edited by: M. Heimann

References

Ahmadov, R., Gerbig, C., Kretschmer, R., Körner, S., Rödenbeck, C., Bousquet, P., and Ramonet, M.: Comparing high resolution WRF-VPRM simulations and two global CO_2 transport models with coastal tower measurements of CO_2, Biogeosciences, 6, 807–817, doi:10.5194/bg-6-807-2009, 2009.

Arneth, A., Lloyd, J., Santruckova, H., Bird, M., Grigoryev, S., Kalaschnikov, Y. N., Gleixner, G., and Schulze, E.-D.: Response of central Siberian Scots pine to soil water deficit and long-term trends in atmospheric CO_2 concentration, Global Biogeochem. Cy., 16, 1005, doi:10.1029/2000GB001374, 2002.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, 5 A., Katul, G., Law, B., Lee, X., Mahli, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, B. Am. Meteorol. Soc., 82, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001.

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

Carouge, C., Rayner, P. J., Peylin, P., Bousquet, P., Chevallier, F., and Ciais, P.: What can we learn from European continuous atmospheric CO_2 measurements to quantify regional fluxes – Part 2: Sensitivity of flux accuracy to inverse setup, Atmos. Chem. Phys., 10, 3119–3129, doi:10.5194/acp-10-3119-2010, 2010.

Carvalhais, N., Reichstein, M., Seixas, J., Collatz, G.J., Pereira, J.S., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M. J., and Valentini, R.: Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval, Global Biogeochem. Cy., 22, GB2007, doi:10.1029/2007GB003033, 2008.

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., and Ramonet, M.: Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives, Climatic Change, 103, 69–92, doi:10.1007/s10584-010-9909-3, 2010.

Collatz, G., Ball, J., Grivet, C., and Berry, J.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agr. Forest Meteorol., 54, 107–136, 1991.

Farquhar, G. D., Caemmerer, S. V., and Berry, J. A.: A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species, Planta, 149, 78–90, 1980.

Gerbig, C., Lin, J. C., Munger, J. W., and Wofsy, S. C.: What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, Atmos. Chem. Phys., 6, 539–554, doi:10.5194/acp-6-539-2006, 2006.

Gerbig, C., Körner, S., and Lin, J. C.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, Atmos. Chem. Phys., 8, 591–602, doi:10.5194/acp-8-591-2008, 2008.

Göckede, M., Michalak, A. M., Vickers, D., Turner, D. P., and Law, B. E.: Atmospheric inverse modeling to constrain region-scale CO2-budgets at high spatial and temporal resolution, J. Geophys. Res., 115, D15113, doi:10.1029/2009JD012257, 2010.

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E., and Michalak, A. M.: Regional-scale geostatistical inverse modeling of North American CO_2 fluxes: a synthetic data study, Atmos. Chem. Phys., 10, 6151–6167, doi:10.5194/acp-10-6151-2010, 2010.

Groenendijk, M., Dolman, A. J., van der Molen, M. K., Leuning, R., Arneth, A., Delpierre, N., Gash, J. H. C., Lindroth, A., Richardson, A. D., Verbeeck, H., and Wohlfahrt, G.: Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data, Agr. Forest Meteorol., 151, 22–38, doi:10.1016/j.agrformet.2010.08.013, 2011.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S. M., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Kowalczyk, E., Maki, T., Maksyutov, S., Peylin, P., Prather, M., Pak, B. C., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, Tellus B, 55, 555–579, 2003.

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

Lauvaux, T., Gioli, B., Sarrat, C., Rayner, P. J., Ciais, P., Chevallier, F., Noilhan, J., Miglietta, F., Brunet, Y., Ceschia, E., Dolman, H., Elbers, J. A., Gerbig, C., Hutjes, R., Jarosz, N., Legain, D., and Uliasz, M.: Bridging the gap between atmospheric concentrations and local ecosystem measurements, Geophys. Res. Lett., 36, 1–17, doi:10.1029/2009GL039574, 2009.

Law, R. M., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S. J., Lokupitaya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Patra, P. K., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Takigawa, M., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO_2: Experimental overview and diurnal cycle results for 2002, Global Biogeochem. Cy., 22, GB3009, doi:10.1029/2007GB003050, 2008.

Lloyd, J. and Taylor, J. A.: On the temperature dependence of soil respiration, Funct. Ecol., 8, 315–323, 1994.

Lokupitiya, R. S., Zupanski, D., Denning, A. S., Kawa, S. R., Gurney, K. R., and Zupanski, M.: Estimation of global CO_2 fluxes at regional scale using the maximum likelihood ensemble filter, J. Geophys. Res, 113, D20110, doi:10.1029/2007JD009679, 2008.

Meesters, A. G. C. A., Tolk, L. F., and Dolman, A. J.: Mass conservation above slopes in the Regional Atmospheric Modelling System (RAMS), Environmental Fluid Mechanics, 8, 239–248, 2008.

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

Mueller K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, J. Geophys. Res.-Atmos., 113(D21), D21114, doi:10.1029/2007JD009734, 2008.

Papale, D. and Valentini, R.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Glob. Change Biol., 9, 525–535, 2003.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E.: Local low dimensionality of atmospheric dynamics, Phys. Rev. Lett., 86, 5878–5881, 2001.

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res, 110, D24304, doi:10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., and Hirsch, A. I.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci., 104, 18925–18930, 2007.

Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Scheafer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., Van der Laan, S., Neubert, R. E. M., Jordan, A., Rodó, X., Morguí, J.-A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M., and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, Glob. Change Biol., 16, 1317–1337, 2010.

Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., Ciais, P., and AEROCARB contributors: Daily CO_2 flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology, Atmos. Chem. Phys., 5, 3173–3186, doi:10.5194/acp-5-3173-2005, 2005.

Pielke, R. A., Cotton, W. R., Walko, R. L., Tremback, C. J., Lyons, W. A., Grasso, L. D., Nicholls, M. E., Moran, M. D., Wesley, D. A., Lee, T. J., and Copeland, J. H.: A comprehensive meteorological modeling system RAMS, Meteorol. Atmos. Phys., 49, 69–91, 1992.

Pieterse, G., Bleeker, A., Vermeulen, A. T., Wu, Y., and Erisman, J. W.: High resolution modelling of atmosphere canopy exchange of acidifying and eutrophying components and carbon dioxide for European forests, Tellus B, 59, 412–424, 2007.

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global Biogeochem. Cy., 19, GB2026, doi:10.1029/2004GB002254, 2005.

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmos. Chem. Phys., 3, 1919–1964, doi:10.5194/acp-3-1919-2003, 2003.

Rüdiger, C., Albergel, C., Mahfouf, J. F., Calvet, J. C., and Walker, J. P.: Evaluation of the observation operator Jacobian for leaf area index data assimilation with an extended Kalman filter, J. Geophys. Res., 115, D09111, doi:10.1029/2009JD012912, 2010.

Scholze, M., Kaminski, T., Rayner, P., Knorr, W., and Giering, R.: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, J. Geophys. Res, 112, D17305, doi:10.1029/2007JD008642, 2007.

Schuh, A. E., Denning, A. S., Uliasz, M., and Corbin, K. D.: Seeing

the forest through the trees: Recovering large-scale carbon flux biases in the midst of small-scale variability, J. Geophys. Res., 114, G03007, doi:10.1029/2008JG000842, 2009.

Sellers, P., Randall, D., Collatz, G., Berry, J., Field, C., Dazlich, D., Zhang, C., Collelo, G., and Bounoua, L.: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, J. Climate, 9, 676–705, 1996.

Steeneveld, G. J., Mauritsen, T., de Bruijn, E. I. F., de Arellano, J. V. G., Svensson, G., and Holtslag, A. A. M.: Evaluation of limited-area models for the representation of the diurnal cycle and contrasting nights in CASES-99, J. Appl. Meteorol. Climatol., 47, 869–887, 2008.

Tolk, L. F., Peters, W., Meesters, A. G. C. A., Groenendijk, M., Vermeulen, A. T., Steeneveld, G. J., and Dolman, A. J.: Modelling regional scale surface fluxes, meteorology and CO_2 mixing ratios for the Cabauw tower in the Netherlands, Biogeosciences, 6, 2265–2280, doi:10.5194/bg-6-2265-2009, 2009.

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., and Nikolova, S.: OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, J. Geophys. Res, 112, G02027, doi:10.1029/2006JG000367, 2007.

Yadav, V., Mueller, K. L., Dragoni, D., and Michalak, A. M.: A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, Biogeosciences, 7, 2655–2671, doi:10.5194/bg-7-2655-2010, 2010.

Zupanski, D., Denning, A. S., Uliasz, M., Zupanski, M., Schuh, A. E., Rayner, P. J., Peters, W., and Corbin, K. D.: Carbon flux bias estimation employing maximum likelihood ensemble filter (MLEF), J. Geophys. Res., 112, D17107, doi:10.1029/2006JD008371, 2007.