Atmospheric CO_2 inversions at the mesoscale using data driven prior uncertainties. Part1: Methodology and system evaluation

Panagiotis Kountouris[1], Christoph Gerbig[1], Christian Rödenbeck[1], Ute Karstens[1,*], Thomas F. Koch[2], Martin Heimann[1]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany

[2]Meteorological Observatory Hohenpeissenberg, Deutscher Wetterdienst, Germany

[*]Now at ICOS Carbon Portal, Lund University, Lund, Sweden

Correspondence to: P. Kountouris (pkount@bgc-jena.mpg.de)

Abstract

Atmospheric inversions are widely used in the optimization of surface carbon fluxes at regional scale using information from atmospheric CO_2 dry mole fractions. In many studies the prior flux uncertainty applied to the inversion schemes does not reflect directly the true flux uncertainties but ~~it~~ is used ~~in such a way~~ to regularize the inverse problem. Here, we aim to implement an inversion scheme using the Jena inversion system and applying a prior flux error structure derived from a model – data residual analysis using high spatial and temporal resolution over a full year period in the European domain. We analyzed the performance of the inversion system with a synthetic experiment, where the flux constraint is derived following the same residual analysis but applied to the model-model mismatch. The synthetic study showed a quite good agreement between posterior and "true" fluxes at European/Country and annual/monthly scales. Posterior monthly and country aggregated fluxes improved their correlation coefficient with the "known truth" by 7% compared to the prior estimates when compared to the reference, with a mean correlation of 0.92. Respectively, the ratio of the standard deviation between posterior/reference and prior/reference was also reduced by 33% with a mean value of 1.15. We identified temporal and spatial scales where the inversion system maximizes the derived information; monthly temporal scales at around 200 km spatial resolution seem to maximize the information gain.

1 Introduction

The continuous rise of the abundance of greenhouse gases in the atmosphere, especially due to fossil fuel combustion, alerted the scientific community to systematically monitor these emissions. The challenge is not limited only to revealing the spatial distribution of CO_2 sources and sinks on continental scales, but also to accurately quantifying CO_2 emissions and their uncertainties at country scales. In situ atmospheric measurements of the atmospheric CO_2 variability combined with inverse atmospheric models are used as an independent method to provide "top down" flux estimates for comparison with estimates from "bottom up" methods. The latter use local observations (e.g. eddy covariance), and combine these with ancillary data, e.g. soil maps, satellite data, and terrestrial ecosystem models in order to spatially scale up local flux estimates to larger regions (Jung et al., 2009). Both approaches act complementary, for optimal comprehension of carbon sources and sinks in a "multiple constraint" (Schulze et al., 2010) approach and emission inventories assessment. As these inventories are used to deduce national emission estimates, in compliance with the Kyoto protocol requirements, accuracy is essential.

An atmospheric inverse modeling system provides the link from atmospheric concentrations to surface fluxes. However, the limited number of observations available for solving the system for quite a number of unknowns (spatially and temporally resolved fluxes) makes the inverse problem strongly under-determined. To solve the inverse problem the system incorporates Bayes' theorem and uses a-priori knowledge, provided by e.g. biosphere models and emission inventories accompanied by corresponding uncertainty estimates. Then, the system optimizes the a-priori fluxes by minimizing the difference between model predictions and observed concentrations. For the current study only the biospheric fluxes were optimized, and emissions from fossil fuel combustion are assumed to be known much better, as it is the case in almost all published regional inversion studies. Inversion systems have been extensively used to derive spatiotemporal flux patterns at global (e.g. Enting et al., 1995; Kaminski et al., 1999a; Gurney et al., 2003; Mueller et al., 2008), and regional scale (e.g. Gerbig et al., 2003a; Peylin et al., 2005; Lauvaux et al., 2012; Broquet et al., 2013).

The challenge in regional inversions is to reconstruct at high resolution the spatiotemporal flux patterns, usually of the net ecosystem exchange (NEE). For that purpose currently deployed global or regional inverse modeling schemes use different state spaces (i.e. the set of variables to be optimized through the inversion process). Peters et al. (2007) split the domain of interest into regions according to ecosystem type. Subsequently fluxes are optimized by using linear multiplication factors to scale NEE for each week and each region. The pitfall of this system is that a zero prior flux has no chance to be optimized and remains zero. Zupanski et al. (2007) divided the NEE into two components, i.e. the gross photosynthetic production (GPP) and ecosystem respiration (R). Then multiplicative factors for the gross fluxes were derived on the grid scale, under the assumption of being constant in time. A step further made by Lokupitiya et al. (2008) used the same approach but with an 8-week time window allowing for temporal variations for the multiplicative factors. A different approach introducing the carbon cycle data assimilation system (CCDAS) was implemented by Rayner et al. (2005) and Kaminski et al. (2012) by constraining global parameters within a biosphere model able to control surface-atmosphere exchange fluxes, against observed atmospheric CO_2 mole fractions, instead of the fluxes themselves; this CCDAS approach alwo allows for nonlinear dependencies of the fluxes on the parameters. Lauvaux et al. (2012) used a Bayesian approach based on matrix inversion, separately optimizing day and night time fluxes at a weekly time scale for a limited simulation period and domain. An attempt to assess which of these approaches better reproduces NEE was made by Tolk et al. (2011). This study investigated the impact of different inversion approaches via a synthetic experiment utilizing an ensemble Kalman filter technique and the same transport model for all cases. They found that inversions which separately optimize gross fluxes within a pixel inversion concept perform better on reconstructing the NEE, although they fail to obtain the gross fluxes. Taking into consideration these findings we also choose the pixel based inversions but optimizing the net biogenic fluxes as we are mainly interested in the total carbon flux budget.

Introducing proper prior flux uncertainties is crucial for meaningful posterior estimates, as these uncertainties weight the prior knowledge between different locations and times, as well as with respect to the data constraint. The uncertainties have the form of a covariance matrix and can be categorized in uncertainties of the prior fluxes, and uncertainties of the observational constraint, which includes measurement and transport model uncertainties. While the measurement uncertainty in the observational constraint ~~may be more easily~~is usually defined with the main diagonal of the covariance matrix representing the uncertainty of the observations and the model at a specific time and location, our knowledge for the prior uncertainty is limited, especially regarding temporal and spatial correlations that effectively control the state space. Early inversions assumed fully uncorrelated flux uncertainties (Kaminski et al., 1999b), while spatial and temporal correlations were used later by Rödenbeck et al. (2003), who investigated the autocorrelation of monthly CO_2 fluxes calculated by a set of terrestrial and ocean models. In Rödenbeck (2005), spatial correlations for land fluxes were assigned to a state space of 4° latitude x 5° longitude resolution. Slightly different correlation length scales were considered for the meridional and zonal direction, assuming that the climate ~~zone~~ of the latter varies less than of the former. Flux correlations on land were determined by assuming an exponential pulse response function with a length of 1275 km. This leads to correlation lengths ~~with~~ approximately two~~ice~~ times larger compared to the pulse ~~the correlation~~ length. Typically the spatial correlations are considered more as a tool to regularize the inverse problem, rather than an uncertainty feature. Schuh et al. (2010) obtained correlation lengths from Rödenbeck et al. (2003) but with a much higher state space resolution of 200 km. Lauvaux et al. (2008) neglected the spatial correlations to enlarge the impact of the data. Carouge et al. (2010a) inferred spatial and temporal correlation lengths based on the agreement between posterior and "true" fluxes in the framework of a synthetic experiment, where the "truth" is known. A different approach was used in Peters et al. (2007) study where they interpret the length scale from a climatological and ecological perspective, and use it to spread information within regions, which the network is incapable to constrain. In particular correlations are applied such that the same ecosystem types in different TransCom regions (basis function regions, see also http://transcom.project.asu.edu/transcom03_protocol_basisMap.php) decrease exponentially with distance (L=2000km), and thus assumes a coupling between the behavior of the same ecosystem. Ad-hoc solutions have also been used, assuming that daily fluxes have smaller correlation lengths than monthly fluxes which are used by other studies (Peylin et al. 2005). More specifically Peylin et al. (2005) assumed 500 km for daily temporal resolution compared to the much larger correlation lengths used by Rödenbeck for monthly flux resolution. Michalak et al. (2004) implemented a geostatistical approach to describe the prior error structure. Specifically the prior error covariance describes at which degree deviations of the surface fluxes from their mean behavior at two different locations or times are expected to be correlated as a function of the distance in space or in time. They simultaneously estimate posterior fluxes as well as parameters controlling the model-data mismatch uncertainty and the prior flux uncertainty, including variance as well as spatial and temporal correlation lengths. Although this approach may be considered as an objective way to infer spatial and temporal correlation lengths, it forces the structural parameters of the error covariance to be statistically consistent with the atmospheric data. In other words, flux parameters are optimized from atmospheric concentration data, and they are forced to have values which can reproduce the atmospheric data. ~~from the few regions where station-to-station distances are small enough to be comparable to the correlation length scales.~~In a similar approach Ganesan et al. (2014) and Lunt et al. (2016), applied a hierarchical Bayesian model using atmospheric concentrations, to estimate both fluxes, and a set of hyper-parameters (e.g. mean and standard deviation of a priori emissions PDF as well as model – measurement standard deviation and autocorrelation scales). In those studies direct information for the tracers of interest (sulfur hexafluoride (SF_6) and methane (CH_4)) is not available. Eddy Covariance stations (EC) can provide a more direct method to infer spatial and temporal flux correlations. Chevallier et al. (2006) and Chevallier et al. (2012) introduced autocorrelation analysis of the residual between fluxes simulated by biosphere models ~~or~~ and fluxes measured by EC to infer spatial and temporal error correlations. The derived error statistics were implemented in a regional CO_2 inversion by Broquet et al. (2013).

Daily NEE flux residuals from model - data comparisons showed temporal correlations up to 30 days but very short spatial correlations up to 40 km (Kountouris et al. 2015). In such a case the a-priori integrated uncertainty over time and space, e.g. annually and EU wide domain integrated, according to the error propagation will be exceptionally small. For example a variance of 1.82 $\mu mole.m^{-2}.s^{-1}$ (from model – data differences) combined with the abovementioned correlation scales yields an uncertainty of 0.12 GtC y^{-1} for the total flux over Europe. This value is significantly smaller than the assumed uncertainty which is typically used by the inversion systems. For comparison we refer to studies from Rivier et al. (2010) and Peylin et al. (2005) (for a slightly larger domain than ours) where an a priori uncertainty of approximately 1.4 GtC y^{-1} and 1 GtC y^{-1} respectively was used. Further, Peylin et al. (2013) found that the variance of the posterior NEE fluxes ~~for~~ integrated over the European domain among 11 global inversions is also 3 to 4 times larger (0.45 GtC y^{-1}). Although it is not yet entirely clear what would be the

Formatted: Subscript

Formatted: Subscript

“correct” value for the prior uncertainty, it seems that in our study it should be increased not only to give enough flexibility to the system to adjust but also to be at least comparable with other posterior uncertainty estimates. A typical method is to inflate the spatiotemporal component by scaling accordingly the prior error covariance. In a study by Lauvaux et al. (2012) two correlation lengths were used at 300 and 50 km, and for the shorter scale the uncertainty was inflated by increasing the RMS of the prior error covariance. The model - data analysis (Kountouris et al. 2015) does neither justify the use of large correlation scales nor largely inflated variances which exceed the model-data flux mismatches, however it is consistent with an additional overall bias error which can not be captured from the estimated spatiotemporal error structure. Hence an appropriate approach would be to introduce two adjustable terms into the inversion system. One term to reflect the data-derived error structure without error inflation (prior error covariance matrix which describes the spatiotemporal component) and one term to represent a bias component. To the best of our knowledge such an approach has not yet been used in inversion systems.

This study primarily aims to use the information extracted from the model-EC data residuals (spatiotemporal error structure) to define a data-driven error covariance rather than simply assuming one, adopting a conservative one or an expert knowledge solution. For that, we implement our previous methodology and findings regarding the prior uncertainty to atmospheric inversions following Kountouris et al. (2015). As explained above, we implement two uncertainty terms; the first one to reflect the true spatiotemporal error structure and the second term ~~referred~~ to reflect~~to~~ a bias term. We use the Jena inversion system (Rödenbeck, 2005; Rödenbeck et al., 2009) for the regional scale consisting of a fully coupled system as described in Trusilova et al. (2010), ~~between~~ which couples the global three-dimensional atmospheric tracer transport model TM3 (Heimann and Körner, 2003) and the regional stochastic Lagrangian transport model STILT (Lin et al., 2003). This scheme allows retrieving surface fluxes at much finer resolution (0.25^o) compared to global models. The first part of this study details the methodology of the prior error implementation, and evaluates the system’s performance through a synthetic data experiment. The system evaluation is an extension of Trusilova et al. (2010) where the evaluation was limited to the observation space only. We extend that to the flux space by comparing flux retrievals at various spatial and temporal scales against synthetic “true” fluxes. Station locations and observation times (including gaps) were created as in the real observation time series presented in the second part of this study (Kountouris et al., 2016). That way we can use the synthetic experiment to evaluate to what extent we can trust the results, if a real-data inversion is performed. In the second part of this study (Kountouris et al., 2016) the regional inversion system is applied to real observations of atmospheric CO_2 mole fractions from a network of 16 stations.

This paper is structured as follows. In Section 2 we present the inversion scheme and introduce the settings of the atmospheric inversions. In Section 3 we present the results from a synthetic inversion experiment aimed to assess the prior error setup, considering it as a step towards atmospheric inversions using real atmospheric data with an objective, state of the art prior error formulation. Discussion and conclusions are following in Section 4 and 5 respectively.

2 Methods

2.1 Inversion scheme

The Jena Inversion System (Rödenbeck 2005; Rödenbeck et al., 2009) was used for the current study. The scheme is based on the Bayesian inference and uses two transport models, the TM3 model (Heimann and Körner, 2003) for global, and the STILT model (Lin et al., 2003) for regional simulations. The advantage of the system is that it combines a global transport model with a regional one without the need of a direct coupling along the boundaries. The global is used to calculate fluxes from the far field (outside of the regional domain of interest), and subsequently this information can be used to provide lateral boundary information for the regional model. Primary input of the system is the observed mixing ratios c_{meas}. This vector contains all measured mixing ratios at different times and locations. The modeled mixing ratios c_{mod} given from a temporally and spatially varying discretized flux field f are computed from an atmospheric transport model and can be formally expressed as

$$c_{\mathrm{mod}} = Af + c_{ini} \tag{1}$$

where c_{ini} is the initial concentration and A the transport matrix which maps the flux space to the observation space. For the regional domain the transport matrix A has been pre-computed by the STILT transport model. The system calculates the modeled concentrations when and where a measurement exists in the c_{meas} vector. The initial concentration assumed to be well mixed and remains constant throughout the simulation. The assumption of the well mixed initial concentration is considered to be valid, since any spatial structure would be lost during the spin-up period.

In the following, we briefly describe the inverse modeling approach. For more details the reader is referred to Rödenbeck (2005).

In grid-based atmospheric inversions the number of unknowns (spatially and temporally resolved fluxes) is larger than the number of measurements (hourly dry mole fractions at different sites), making the inverse problem ill-posed. In the Bayesian concept this can be remedied by adding a-priori information. This information can be written as

$$f = f_{fix} + F \cdot p \tag{2}$$

where f_{fix} is the a-priori expectation value of the flux, matrix F contains all the a-priori information about flux uncertainties and correlations (implicitly defining the covariance matrix) and p is a vector representing the adjustable parameters. The parameters p are uncorrelated with zero mean and unit variance. This flux model represents just a different way to define the a-priori probability distribution of the fluxes, than the traditional way where the a-priori error covariance matrix is explicitly specified. The cost function describing the observational constraint is expressed as

$$J_c = \frac{1}{2}(c_{meas} - c_{mod})^T \cdot Q_c^{-1} \cdot (c_{meas} - c_{mod}) \tag{3}$$

where Q_c is the observation error covariance matrix. This diagonal matrix weights the mixing ratio values considering measurement uncertainty, location-dependent model uncertainty and a data density weighting. The latter ensures that the higher amount of data from continuous measurements compared to the data from flask measurements would not lead to a considerably stronger impact of these corresponding sites (Rödenbeck, 2005). This can also be formally interpreted as a temporal correlation scale which ensures that the model-data-mismatch error is not independent within a week, corresponding roughly to time scales of synoptic weather patterns.

The inversion system seeks to minimize the following cost function that combines the observational (Eq. 3) and the prior flux constrain

$$J = J_c + \frac{1}{2} \cdot p^T \cdot p \tag{4}$$

The minimization of the cost function is done iteratively with respect to the parameters p by using a Conjugate Gradient algorithm with re-orthogonalization (Rödenbeck 2005).

2.2 Characteristics of the inversion set up

2.2.1 A-priori information and uncertainties

The a-priori CO_2 flux fields were derived from the Vegetation Photosynthesis and Respiration Model, VPRM (Mahadevan et al., 2008). VPRM uses ECMWF (European Centre for Medium Range Weather Forecasting) operational meteorological data for radiation (downward shortwave radiative flux) and temperatures (T2m), the SYNMAP landcover classification (Jung et al., 2006), and EVI (enhanced vegetation index) and LSWI (land surface water index) derived from MODIS (Moderate Resolution Imaging Spectroradiometer). Model parameters were re-optimized for Europe using eddy covariance measurements made during 2007 from 47 sites (a full site list is given in Kountouris et al. (2015); we excluded some sites due to insufficient temporal data coverage or lack of representativeness). To mediate the impact of data gaps, a data density weighting was introduced that takes into account the coverage of different times of the day (using 3-hour bins) in the different seasons. Optimized parameters are shown in Table 1. The net ecosystem exchange at hourly scale and at 0.25°x 0.25° spatial resolution for 2007 was simulated with the optimized parameters for the European domain shown in Fig. 1. The domain-wide aggregated biospheric carbon budget for 2007 derived that way from VPRM was found to be -0.96 GtC y^{-1} (i.e. uptake by the biosphere). Note that without the density weighting an even stronger flux of -1.35 GtC y^{-1} was derived, indicating the importance of proper treatment of data gaps by either gap-filling or by the inclusion of weights.

Additionally, biogenic CO_2 fluxes were simulated with the BIOME-BGC model, specifically its global implementation as GBIOME-BGCv1 (Trusilova and Churkina 2008) at the same 0.25°x 0.25° spatial and hourly temporal resolution. The purpose of the second flux field is to provide a perfectly known flux distribution as “true” fluxes that can be used to generate synthetic observations. The BIOME-BGC model is a terrestrial ecosystem process model which requires only standard meteorological data like, daily maximum-minimum temperature, precipitation, incoming shortwave solar radiation, vapor pressure deficit (VPD), and the day length (DLn). How accurate the modeled fluxes are is difficult to say, since this would require a validation against observed fluxes from eddy covariance stations. Nevertheless, biospheric models still suffer from large uncertainties. The remarkably diverge results between models confirm how uncertain models are (see Friedlingstein et al. (2014)). However in the current experiment the accuracy of the “true” fluxes is not of a concern, since we aim only to create a synthetic flux field that we perfectly know.

The a-priori flux in a real-data inversion would have three components including fossil fuel and ocean fluxes

$$f_{pr} = f_{pr,nee} + f_{pr,ff} + f_{pr,oc} \tag{5}$$

We note that for the synthetic case the last two a priori terms are set to zero. Similarly the deviation term (the data-derived correction to the a-priori fluxes) of the flux model (Eq. 6) consists of the terms referring to NEE, fossil fuel, and ocean fluxes. ~~but equivalently~~Here in the synthetic case the last two terms are set to zero (i.e. they are not optimized) ~~for the synthetic inversion~~.

$$F\,\delta s = (F_{nee}, F_{oc}, F_{ff})\begin{pmatrix}\delta s_{nee}\\ \delta s_{oc}\\ \delta s_{ff}\end{pmatrix} \quad (6)$$

~~Note that the a-priori error covariance matrix does not explicitly appear in the inversion, but is included though the second term in Eq. 8 (see section 2.2.2).~~

~~According to this formulation the columns of G_{tcor} and G_{xycor} contain the spatiotemporal extents of the individual NEE pulses (range of values between 0 and 1) and the diagonal matrix $f_{sh}(x,y,t)$ contains the pixel-wise a priori uncertainties. These uncertainties were chosen to be flat (constant) in space and time. For more detailed information the reader is referred to Rödenbeck et al. (2005).~~

The total prior uncertainty was chosen according to the mismatch between VPRM and BIOME-BGCv1, calculated as the annual and domain wide integrated flux mismatch. Prior fluxes and the fluxes representing the synthetic truth are strongly different (-0.96 GtC y^{-1} and -0.31 GtC y^{-1} for VPRM and GBIOME-BGCv1, respectively). The error structure used for the synthetic study is estimated according to the method applied in Kountouris et al. (2015). Time-series of daily fluxes were extracted for both biosphere models at grid cell locations where an EC station exists. Fluxes from GBIOME-BGCv1 can also be regarded as synthetic EC fluxes. Then spatial and temporal autocorrelation analysis was performed on the daily model-model flux residuals, yielding a spatial correlation length scale of 566 km and a temporal correlation scale of 30 days. We note that the current study does not directly make use of the error structure derived in Kountouris et al. (2015), since this is applicable for real data inversions. Instead we use the same methodology to derive the actual model-model error structure since here we perform a synthetic data inversion, exploring amongst others the accuracy of this method.

The eddy covariance station locations used for this analysis were exactly the same as in Kountouris et al. (2015) ensuring similarity in the derivation of the error structure for the synthetic data inversions. Following this approach apart from the similarity, we also ensure that results from the synthetic experiment, would be informative for a real data inversion, by using exactly the same information to characterize the prior uncertainties. ~~However o~~Of note is that for the synthetic data inversions, prior fluxes from VPRM model were not optimized against GBIOME-BGCv1 “true” fluxes.

The implicitly defined prior error covariance matrix contains diagonal elements of (1.45 µmol m^{-2} s^{-1})^2, which reflect the variance from model-model flux mismatches at the 50 km spatial resolution of the state space. Exponentially decaying spatial correlations were implemented with a correlation scale of 766 km at the zonal and 411 km at the meridional direction, roughly corresponding to the 566 km correlation scale yielded from the model-model residual autocorrelation analysis and preserving the same zonal/meridional ratio as in the global inversion. Temporal autocorrelation was set to 31 days, which is consistent with the Kountouris et al. (2015) analysis. These scales result in an uncertainty for the spatiotemporal component (E_{st}) domain-wide and annually integrated of 0.44 GtC y^{-1}. We chose two different approaches to increase the prior uncertainty at domain-wide and annually integrated scale such that it matches the mismatch of 0.65 GtC y^{-1} between the two biosphere models. First we inflate the error by scaling the error covariance matrix, this case is referred to as base case B1 hereafter. The second approach, referred to as scenario S1, could be considered as a more formal way: we introduce an additional degree of freedom to the inversion system by allowing for a bias term. This term is spatially distributed according to the annually averaged VPRM respiration component, and is kept constant in time. The idea behind the implementation of this term is that at large scales a bias might exists, which can not be captured in the model-data residual autocorrelation analysis (EC measurements are representative at scales ~ 1 km). This assumption avoids the artificial inflation of the uncertainty at pixel scale, and restricts the pixel to pixel corrections to be statistically consistent with the actual error structure. The bias shape selection (respiration shape) was preferred over the NEE fluxes, as otherwise a priori neutral pixels (with zero NEE) could not be bias corrected. Further, allowing bias to have a spatial shape might be sound, since regions with stronger fluxes might be also more biased. The error E_{BT} of the bias component was adjusted such that the total prior error E_{tot} for annually and domain-wide integrated fluxes matches the targeted total uncertainty:

$$E_{tot}^2 = E_{ST}^2 + E_{BT}^2 \qquad (7)$$

This resulted in an overall uncertainty E_{tot} of 0.65 GtC y^{-1}, which is identical to the mismatch between the two biosphere models.

2.2.2 State space

The inversion system optimizes additive corrections to three-hourly fluxes in a sense that the posterior flux estimate can be given by the sum of a fixed a priori term (first term of the right hand side in Eq. 8) and an adjustable term (second term in Eq. 8). The latter has a-priori a zero mean ~~and unit variance~~. The biogenic fluxes can be defined as follows:

$$f(x,y,t) = f_{fix}(x,y,t) + f_{sh}(x,y,t) \cdot \sum_{m_t}^{N_t} \sum_{m_s}^{N_s} G_{tcor,m_t}(t) \cdot G_{xycor,m_s}(x,y) \cdot p_{inv,m_t,m_s} \qquad (8)$$

where f_{sh} is a shape function which defines the adjustable term. The spatial and temporal correlation structures of the uncertainty are described by the pulse response functions G_{xycor} and G_{tcor} respectively. The term p_{inv} contains the adjustable parameters which ~~they~~ a~~-~~ priori have~~,~~ a Gaussian distribution with zero mean and unit variance.

Note that the a-priori error covariance matrix ($Q_{f,pr}$) does not explicitly appear in the inversion, but is included though the second term in Eq. 8 ~~(see section 2.2.2)~~.

According to this formulation the columns of G_{tcor} and G_{xycor} contain the spatiotemporal extents of the individual NEE pulses (range of values between 0 and 1) and the diagonal matrix $f_{sh}(x,y,t)$ contains the pixel-wise a priori uncertainties. These uncertainties were chosen to be flat (constant) in space and time. For more detailed information the reader is referred to Rödenbeck ~~et al.~~ (2005).

For the S1 case the posterior flux estimates can be ~~derived~~ expressed by adding the optimized bias flux field to Eq. 8

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

$$f(x,y,t) = f_{fix}(x,y,t) + f_{sh}(x,y,t) \cdot \sum_{m_t}^{N_t} \sum_{m_s}^{N_s} G_{tcor,m_t}(t) \cdot G_{xycor,m_s}(x,y) \cdot p_{inv,m_t,m_s} + f_{sh}^{BT}(x,y) \cdot \sum_{m_t}^{N_t} G_{tcor,m_t}(t) \cdot p_{BT} \quad (9)$$

The bias term f^{BT} follows a flux shape (here we used annually averaged respiration, with no temporal variation).

Following Rodgers 2000, the posterior flux uncertainties are contained in the covariance matrix of the posterior probability distribution which can be estimated from eq. (10)

$$Q_{f,post} = ((A \cdot F)^T \cdot Q_c^{-1} \cdot (A \cdot F) + Q_{f,pr}^{-1})^{-1} \quad (10)$$

where Q_c is the measurement error covariance matrix.

2.2.3 Observation vector and uncertainties

The observation vector c_{meas} contains mixing ratio observations at all site locations and sampling times. A common procedure to derive synthetic observations is to create a "true" flux field by adding some error realizations to the a-priori fluxes (Schuh et al., 2009; Broquet et al., 2011) or to perturb the resulting synthetic observations (Wu et al., 2011). For the current study instead we use a different biosphere model, the GBIOME-BGCv1 model, to derive biogenic CO_2 fluxes at hourly scale. Such an approach is also used by Tolk et al. (2011). Then a forward transport model run was performed to create synthetic mixing ratios at hourly resolution for each station location. We note that the synthetic data were derived without adding error realizations. This choice of using two different biosphere models for deriving the a-priori and the "true" fluxes is expected to increase the realism of the synthetic data study, given the fact that the real spatiotemporal flux distribution is highly unknown (though the model-to-model difference may not accurately reflect the model errors either). For the synthetic study, observations were created for the same station locations and observation times as in the real observation time series which are used in the second part of this study (Kountouris et al., (2016)). An overview of the atmospheric stations is given in table 2 and Fig. 1. The data coverage per station is shown in

Field Code Changed

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

Figure 2~~Figure 2~~. Only daytime observations were considered (11:00 – 16:00 local time) since the transport model is expected to perform worse during night when a stable boundary layer forms. An exception is made for mountain stations that measure the free troposphere, where only nighttime observations (23:00 – 04:00 local time) were considered, as this time can be better represented by the transport model (Geels et al., (2007)). In total 20273 hourly observations from the year 2007 were used.

The model-data mismatch uncertainty associated with each measurement is expressed as a diagonal covariance matrix, and contains measurement errors and errors from different components describing the modeling framework (i.e. model errors due to imperfect transport, aggregation errors, etc.) (Gerbig et al., 2003b). For the current study, all sites are classified according to their characteristics (e.g. tall tower, mountain sites etc.), and uncertainties were defined depending on the site class (Figure 2~~Figure 2~~, legend on the right). The uncertainties are considered as representative for current inverse modeling systems. Although the measurement error covariance is a diagonal matrix, transport error correlations might be present. Although we do not explicitly introduce off-diagonal terms in the measurement error covariance matrix, we do consider for temporal correlations via a data density weighting function which inflates the uncertainty. (see Section 2.1 and more information in Rödenbeck, C., 2005).

Formatted: Font: Not Bold, Check spelling and grammar

Formatted: Font: Not Bold, Check spelling and grammar

2.2.4 Atmospheric transport

For the synthetic data study only the regional atmospheric model STILT was used to create the observations with a forward run, and to perform the inversion. This was feasible since the synthetic CO_2 observations are only influenced by fluxes occurring within the Domain of Interest (DoI), hence global runs to retrieve boundary conditions at the edge of DoI are not necessary. The transport matrix for the regional inversions was generated in form of pre-calculated footprints (sensitivities of atmospheric observations to upstream fluxes) at 0.25 degrees spatial and hourly temporal resolution for the full year 2007. STILT trajectory ensembles were driven by ECMWF meteorological fields (Trusilova et al., 2010), and computed for 10 days backwards in time, ensuring that nearly all trajectories have left the domain of interest.

With respect to the assumed model height, STILT uses surface elevation maps from ECMWF (European Centre for Medium Range Weather Forecasting) with a resolution of 0.25 x 0.25 degrees. As the model orography represents an average over the whole grid cell, it is, in particular at steep mountain sites, significantly smaller compared to the real orographic height at the station location. In order to better represent the location of the station in the large scale flow, usually a model height is assumed that more closely represents the real height (above sea level) of the measurements. However, using exactly the measurement height (a.s.l.) in the model would decouple the CO_2 signal too strongly from the surface fluxes and hence lead to a systematic underestimation of the surface influence on the concentrations (Geels et al., 2007). A compromise was reached by adjusting the model height (above ground) by half the distance between the model orographic height and the real station height.

~~2.3 Metrics for performance evaluation~~

~~Following Rödenbeck et al. (2003) we evaluate the goodness of fit for each station (station specific χ^2). The modeled dry mole fractions should be with 68% probability within the ± 1σ range from the observed mole fractions. This is equivalent to the requirement that the dry mole fraction part of the cost function defined as the sum of hourly squared differences, divided by the uncertainty interval and the number of observations n (Eq. 10), should be close to unity.~~

$$\cancel{\chi_c^2 = \frac{\sum_i \frac{(\Delta c_i)^2}{\sigma_i^2}}{n} \qquad (10)}$$

~~Another important aspect is the reduced χ_r^2 metric that compares the a-priori model performance with the specified error structure by dividing the squared residuals of optimized minus observed dry mole fractions by the squared specified uncertainties. This is also equivalent to two times the cost function at its minimum divided by the number of degrees of freedom (effective number of observations) (Thompson et al., 2011):~~

$$\cancel{\chi_r^2 = 2\frac{J_{\min}}{n}}$$

~~(11)~~

~~Again, a correct balance should be close to unity. Smaller values suggest that the model performance was better than specified in the covariance structure and hence the assumed uncertainties (denominator) were conservative.~~

~~In flux space, we evaluate the inversion performance, by comparing the retrieved flux estimates against the synthetic fluxes ("true") at different temporal and spatial scales: annually and monthly integrated fluxes, domain-wide and at country scale. In particular we are interested in capturing the "true" fluxes down to country scale. For that we assess monthly posterior retrievals which we compare to reference data ("true" fluxes), country aggregated, using a Taylor diagram. This diagram provides a concise statistical summary of how well patterns match each other in terms of their correlation and the ratio of their variances.~~

3 Results

The purpose of the synthetic study is to evaluate the system set-up with a realistic approach. To evaluate the ability of the system to retrieve the synthetic true fluxes we visualize spatially distributed fluxes and we study spatially integrated (domain and national scale) as well as temporally (annual and monthly scale) integrated fluxes.

3.1 CO_2 mole fractions

A comparison of true and modeled CO_2 dry mole fractions from forward runs of the optimized fluxes can reveal the goodness of fit, realized through the optimization process. Such a comparison is presented in Figure 3~~Figure 3~~ for the Schauinsland (SCH) continuous station. Both B1 and S1 inversions significantly reduce the misfit between the synthetic (truth) and the a-priori mole fractions. As expected from the optimization (i.e. minimization of the cost function),~~T~~ the RMSD between the prior/posterior from the "true" timeseries for all stations (Table 3) shows an average reduction of around 74% and 76% for the S1 and B1 inversions respectively. Prior correlations (prior vs. true dry mole fractions), have an averaged value of 0.46 which is increased

Formatted: Font: Not Bold, Check spelling and grammar to 0.93 for both inversions. Significant differences between the two inversions were not found apart from a slightly larger decrease of the RMSD for the B1 case. Figure 4~~Figure 4~~ summarizes the capability of the inversions to capture the true signal at each station location in form of a Taylor diagram, indicating that the inversions showed a significant increase of the correlation for all sites. Further the variance of the modeled time-series is significantly closer to the variance of the true signal.

Formatted: Font: Not Bold

To estimate the goodness of fit we consider the station specific χ_c^2 values (Eq. 1~~0~~1) following Rödenbeck et al. (2003).~~,~~ We us~~inge~~ here 7-day aggregated residuals instead of hourly to match the temporal scale of one week of the observation error.~~-~~ The modeled dry mole fractions should be with 68% probability within the ± 1σ range from the observed mole fractions. This is equivalent to the requirement that the dry mole fraction part of the cost function defined as the sum of hourly squared differences, divided by the uncertainty interval and the number of observations n (Eq. 11), should be close to unity.

$$\chi_c^2 = \frac{\sum_t \frac{(\Delta c_t)^2}{\sigma_t^2}}{n} \qquad (11)$$

Field Code Changed

Values smaller than 1 are found for most of the stations with a mean value of 0.28 and 0.32 for the B1 and S1 cases respectively, suggesting a good fitting performance for all stations and for both inversions. The results are comparable with those found in the Rödenbeck et al. (2003) study.

Another important aspect is the reduced χ_r^2 metric, which we use to assess the model performance. By definition the reduced χ_r^2 can be obtained by dividing the squared residuals of optimized minus observed dry mole fractions by the squared specified uncertainties. This is also equivalent to two times the cost function at its minimum divided by the number of degrees of freedom (effective number of observations) (Tarantola 2005):

$$\chi_r^2 = 2\frac{J_{\min}}{n}$$

(12)

Field Code Changed

Again, a correct balance should be close to unity. The reduced chi-squared (Eq. 12~~1~~) was found to be 0.21 for both cases, indicating that the error variance is overestimated, making the error assumption rather conservative.

3.2 Flux estimates and uncertainties

In flux space, we evaluate the inversion performance, by comparing the retrieved flux estimates against the synthetic fluxes ("true") at different temporal and spatial scales: annually and monthly integrated fluxes, domain-wide and at country scale. In particular we are interested in capturing the "true" fluxes down to country scale. For that we assess monthly posterior retrievals which we compare to reference data ("true" fluxes), country aggregated, using a Taylor diagram. This diagram provides a concise statistical summary of how well patterns match each other in terms of their correlation and the ratio of their variances.

The spatial distributions of the annual biosphere-atmosphere exchange fluxes for the prior, the known truth, and the posterior cases are presented in Figure 5~~Figure 5~~. Note that annual fluxes between the two biosphere models used for prior fluxes and true fluxes are substantially different. The inversion significantly adjusts the spatial flux distribution mainly in central Europe and in southern Scandinavia, where a denser atmospheric network exists. The absolute annual mean difference in fluxes (|mean(true – prior)| and |mean(true – posterior)|) is greatly reduced from 70.8 $gCm^{-2}y^{-1}$ to 14.7 $gCm^{-2}y^{-1}$ and 24.6 $gCm^{-2}y^{-1}$ for the B1 and S1 inversions respectively. Detailed patterns, however, are not well reproduced: the fraction of explained spatial variance in the true fluxes (measured~~s~~ as squared Pearson correlation coefficient) decreases from the prior (0.17) to the posterior (0.07 and 0.06 for the cases B1 and S1, respectively). When evaluating this at monthly scales, the fraction of explained spatial variance increases in the posterior estimates compared to the prior for winter months from around 0-15% to about 15-50%, while during the growing season typically a decrease from around 10-35% to about 0-34% is found. The accumulated footprint of the atmospheric network is shown in Figure 6~~Figure 6~~, clearly indicating the strongest constraint on fluxes in central Europe. Interestingly both error structures from S1 and B1 inversions produce posterior fluxes that have approximately the same spatial distribution. When separating the spatiotemporal component from the bias

Formatted: Font: Not Bold, Check spelling and grammar

Formatted: Font: Not Bold, Check spelling and grammar component (in S1 case) we can identify differences between the two inversions. Significant deviations of the spatial flux distribution between the spatiotemporal components were found: The spatiotemporal component in the S1 case has a domain wide annual flux correction of 0.39 GtC y^{-1} (prior – posterior) while the corresponding term in the B1 case has a correction of 0.78 GtC y^{-1}. Nevertheless standard deviations of the corrections with respect to the true spatial flux distribution (true – posterior) was found to have no significant difference ($6.88*10^{-5}$ and $7.38*10^{-5}$ GtC $y^{-1}cell^{-1}$ for S1 and B1 respectively). We do not observe any strong correction in the south-eastern part of Europe as it cannot be “seen” from the atmospheric network due to the distance to the observing sites and the prevailing westerly winds. This could also be inferred from the flux innovation plots (see Figure 5~~Figure 5~~) defined as the difference between prior and posterior fluxes. Only very small or even no corrections occurred in this area.

Formatted: Font: Not Bold, Check spelling and grammar

We are specifically interested in the ability of the inversion system to capture integrated fluxes over time and space. Figure 7~~Figure 7~~ shows an overview of the domain-integrated fluxes at a monthly and annual scale. Despite the remarkably larger a-priori (VPRM) sink compared to the synthetic truth (GBIOME-BGCv1) during the growing season, both inversions, with and without the bias term, produce posterior flux estimates that fully capture the "true" monthly and annually integrated fluxes. While the monthly posterior estimates give no clear evidence on which inversion performs better, retrievals at annual scale slightly favor the inversion without the bias term (B1 case). A difference was observed in the prior uncertainties between the two inversions. While both were scaled to have the same prior annual uncertainty, the B1 inversion has systematically larger prior monthly uncertainties than the S1 as a result of the inflated spatiotemporal component of the prior error covariance. Posterior uncertainties were found to be similar, and include or are close to including (S1 case) the true flux estimates. The uncertainty reduction for annually and domain-wide integrated fluxes, defined as the difference between prior and posterior uncertainties normalized by the prior uncertainty, was found to be 73% and 69% for the S1 and B1 respectively. Note that whilst the prior uncertainty refers only to the flux space, the posterior uncertainty depends on the uncertainty of prior fluxes, measurements, and transport.

Formatted: Font: Not Bold

In order to assess how well the posterior estimates agree with the true fluxes, root mean square difference (RMSD) between true and posterior monthly integrated gridded fluxes were computed

(Table 4). Both inversions B1 and S1 show a similar reduction in the RMSD values compared to the prior. The same picture emerges for the annually integrated fluxes.

Of particular interest is the performance of the system at regional scale, specifically at national level. Figure 8~~Figure 8~~ shows monthly fluxes for selected European countries, including the prior, true and posterior estimates with the corresponding uncertainties. Both error structures show a similar performance. Despite the large prior misfit, the system succeeded in retrieving monthly fluxes at country level. Better constrained regions mainly located in central Europe show the ability to broadly capture the temporal flux variation at monthly scale. Figure 9 summarizes in a Taylor diagram the inversion performance for the S1 case and for each EU-27 country, showing the improvement of monthly and country aggregated fluxes (perfect match would be if the head of the arrow coincides with the reference point marked as green bullet). It is worth mentioning that also for regions that are less constrained by the network, such as Great Britain, Spain, Poland and Romania, the inversions still improved the posterior estimates compared to the prior estimates (see also Fig. 9).

Formatted: Font: Not Bold

3.3 Evaluation with synthetic eddy covariance data

In order to investigate the potential of using eddy covariance measurements for evaluating the retrieved CO_2 fluxes, monthly fluxes from the prior (VPRM), the truth (GBIOME-BGCv1), and the posterior for cases B1 and S1 were extracted at the grid cell locations where eddy covariance stations exist, using the same 53 sites as in Kountouris et al. (2015). The corresponding fluxes were then aggregated over all sites, using a weight that compensates for the asymmetry between number of flux towers for specific vegetation types and the fraction of land area covered by the specific vegetation type. Prior fluxes show a systematically larger uptake compared to the truth, predominantly during the growing season with maximum differences of 0.8 gCm^{-2}day^{-1} (Figure ~~Figure~~ 10). Posterior estimates for both cases captured the magnitude of the true fluxes, with maximum differences of around 0.3 gCm^{-2}day^{-1} during June/July. A significantly larger correction is apparent during spring and summer compared to winter and fall. The very close correspondence of these results with those shown in Figure 7~~Figure 7~~ for the domain-wide

Formatted: Superscript

Formatted: Superscript

monthly flux budget ~~clearly~~ potentially shows that eddy covariance measurements can principally be used for validation of the inverse estimates at monthly timescales.

4 Discussion

4.1 Performance in flux space

Results from the synthetic experiment showed the strengths but also the weaknesses of the system to retrieve the "true" spatial flux distribution. Although the error structure applied to this experiment was statistically coherent with the mismatch between prior and true fluxes, we note a limited ability of the current atmospheric network to retrieve fluxes at local scales. For coarser spatial scales (country level) the carbon budget estimates in the synthetic inversion showed a quite good performance at monthly and annual temporal scales. Further we observed an average reduction of the monthly uncertainties of 65% for the B1 case, and 64% for the S1 case. In combination with the fact that the flux estimates reproduce the "truth" within the posterior uncertainties, this gives us confidence in the accuracy of our estimates.

In the current study we do not excessively assess the transport error but it is rather included as diagonal elements in the measurement error covariance, which is typical in atmospheric inversions. The chi square values confirm that there is no underestimation of the uncertainties. We note though that erroneous flux estimates are likely to be estimated, especially at finer spatial scales where the transport model is not able to resolve the real transport (e.g. individual eddys, complicate terrain etc). However, for coarser spatial scales transport morels perform better, and as long as the fitting performance shows good results, flux estimates should be more reliable.

Prior error correlation in time and space limits the scale, at which information can be retrieved from the inversion. The spatial correlation of several hundred kilometers implies that fluxes at scales smaller than this cannot be significantly improved by the inversion, as the results clearly showed. To assess this more quantitatively, the spatial correlation between a priori or retrieved and true monthly fluxes is calculated for different spatial aggregation scales (starting at 0.25

degree, fluxes were aggregated to 0.5, and then in 1-degree steps up to 8 degree). Results shown in Fig. 11 a) indicate a nearly ~~continuous~~ monotonous increase of the spatial correlation of prior and posterior fluxes with increasing aggregation scale. The additional explained variance brought about by the inversion, i.e. the difference between posterior (red/blue line) and prior (grey line) flux correlation (r-square) with the truth, starts at low values around 0.1, and reaches values around 0.2 for scales larger or equal 2 degrees. Similarly, the spatial correlation between a priori and true fluxes for a given spatial aggregation of 2 degrees, but for different temporal aggregation scales ranging from 1 day to 128 days (Fig. 11 b) shows a continuous increase from about 0.23 to 0.42 (r-square), while the spatial correlation between retrieved and true fluxes only varies slightly between 0.4 and 0.53 (Fig. 11 b), red and blue lines). Here, the additional spatial variance explained by the retrieved fluxes is largest at around monthly time scales (differences between prior and posterior r-square around 0.2), while at seasonal scales this additional explained variance is only around 0.1. Overall, this analysis confirms that there are preferred spatial and temporal scales at which the inversion retrieves the flux distribution best and where thus most information is gained. This is not dependent on whether or not a bias term is included in the state vector, as results for case B1 and S1 do not differ in this regard. It is important to realize that all other scales, at which the inversion does not provide much information, need to be properly represented by the a priori flux distribution. Thus the a priori fluxes need to be realistic at short spatial scales below about 200 km, at seasonal temporal scales, and of course at hourly time scales which are not retrieved by the inversion.

The annual spatial flux distribution of the B1 and S1 cases was found to be quite similar, indicating that inflating the uncertainty by a factor of 1.5 (B1 case, see also 2.2.1 section) or adding a bias component to compensate the inflation (S1 case) lead to a similar flux constraint. This could be explained due to the long correlation length (566 km) which drastically reduces the effective number of degrees of freedom, forcing the fluxes to be smoothly corrected, regardless the use of the bias component.

4.2 Performance in observation space

The high RMSD reduction in combination with the high correlation values and the captured variability between posterior and true dry mole fractions in the synthetic experiment suggest a good performance of the inversion system to retrieve the “true” mixing ratios. Nevertheless this is not surprising, as the atmospheric data are “fitted” by the inversion, and furthermore the forward and the inverse runs used identical transport, without any impact from imperfections in transport simulations.

The uncertainties in the flux space are statistically consistent with the model-model flux mismatch. However the reduced χ_r^2 values obtained from the inversions were rather small (around 0.21). This indicates that overall conservative uncertainties were assumed, and the small χ_r^2 values are a result from the assumed uncertainties in the observation space. Indeed uncertainties in the observation space include also transport uncertainties; however, given that the same transport is used to create synthetic observations and to perform the inversion, there is no actual model-data mismatch related to transport uncertainties, and so the assumed uncertainties are overestimated. In the current study we assumed a diagonal measurement error covariance matrix. Concerns might rise that the observational uncertainties are underestimated due to the absence of the error correlations. However the χ_r^2 values prove the opposite.

5 Conclusions

This paper describes the setup and the implementation of prior uncertainties as derived from model-eddy covariance data comparisons into an atmospheric CO_2 inversion. The inversion system assimilates hourly dry air mole fractions from 16 ground stations to optimize 3-hourly NEE fluxes for the study year 2007. Two different error structures were introduced to describe the prior uncertainty by either inflating the error or by adding an additional degree of freedom allowing for a long term bias. The need of this error inflation comes from the fact that the spatiotemporal model - data error structure alone underestimates prior uncertainties typically assumed for inversion systems at continental/annual scale. In this study we evaluate the Jena inversion system by performing a synthetic experiment and expanding the evaluation also to the retrieved fluxes, whilst only the observation space was evaluated in Trusilova et al. (2010).

Further we assess the impact when adding a bias term in the flux error structure. This study is a preparatory step to retrieving European biogenic fluxes using a data driven error structure consistent with model-flux data mismatches, which is described in the companion paper (Kountouris et al. 2016).

Significant flux corrections and error reductions were found for larger aggregated regions (i.e. domain-wide and countries), giving us confidence on the reliability of the results for a real data inversion at least for aggregated scales up to the country level. We found a similar performance for both error structures. A more detailed analysis of the spatial and temporal scales, at which the inversion provides a significant gain in information on the distribution of fluxes, clearly confirms that a) fluxes at spatial scales much smaller than the spatial correlation length used for the a prior uncertainty cannot be retrieved; b) the inversion performs best at temporal scales around monthly, and c) especially the small spatial scales need to be realistically represented in the a priori fluxes.

Acknowledgments

This work contributed to the European Community's Seventh Framework Program (FP7) project ICOS-INWIRE, funded under grant agreement no. 313169. The authors would also like to thank the Deutsches Klimarechenzentrum (DKRZ) for using the high performance computing facilities. This publication is an outcome of the International Space Science Institute (ISSI) Working Group on "Carbon Cycle Data Assimilation: How to consistently assimilate multiple data streams.

Appendix

The exponentially decaying temporal autocorrelations is a feature newly implemented into the Jena Inversion System. Temporal correlations are not directly defined as off-diagonal elements in the a-priori error covariance, as the latter does not appear explicitly in the inversion. Rather, the inversion system involves time series filtering in terms of weighted Fourier expansions. More specifically the columns of matrix G_{tcor} contain Fourier modes, weighted according to the frequency spectrum that corresponds to the desired autocorrelation function. The reader is referred to Rödenbeck (2005) for more information. Following Rödenbeck (2005) we define the following spectral weight w:

$$w = \frac{v_{low}}{\sqrt{v_{low}^2 + (2\pi v)^2}} \qquad \text{A1}$$

where v_{low} is the characteristic frequency. The characteristic frequency v_{low} can be calculated from the desired temporal autocorrelation time (30 days) of the exponential decay and is expressed in years:

v_{low} = 1/(1/12) where 1/12 is the autocorrelation time in years. Hence the characteristic frequency corresponding to a monthly autocorrelation is 12.

To test numerically whether the implemented autocorrelation decay shape approximates an exponential decay, an error realization of the characteristic frequency was added to the prior fluxes, and the autocorrelation function as described in Kountouris et al. (2015) was calculated numerically simultaneously for the flux time series of all grid cells. Then an exponentially decaying function was fitted (Fig. A1) to derive the autocorrelation scale for the corresponding frequency. The resulting autocorrelation shape indeed approximates very well an exponential decay, with an e-folding time of precisely 30 days. The tight confidence bounds of the fitted parameter (29.3 and 30.6 days within 95 % confidence interval), in combination with the small residual sum-of-squares (0.14) suggests a very good approximation of the exponential decay.

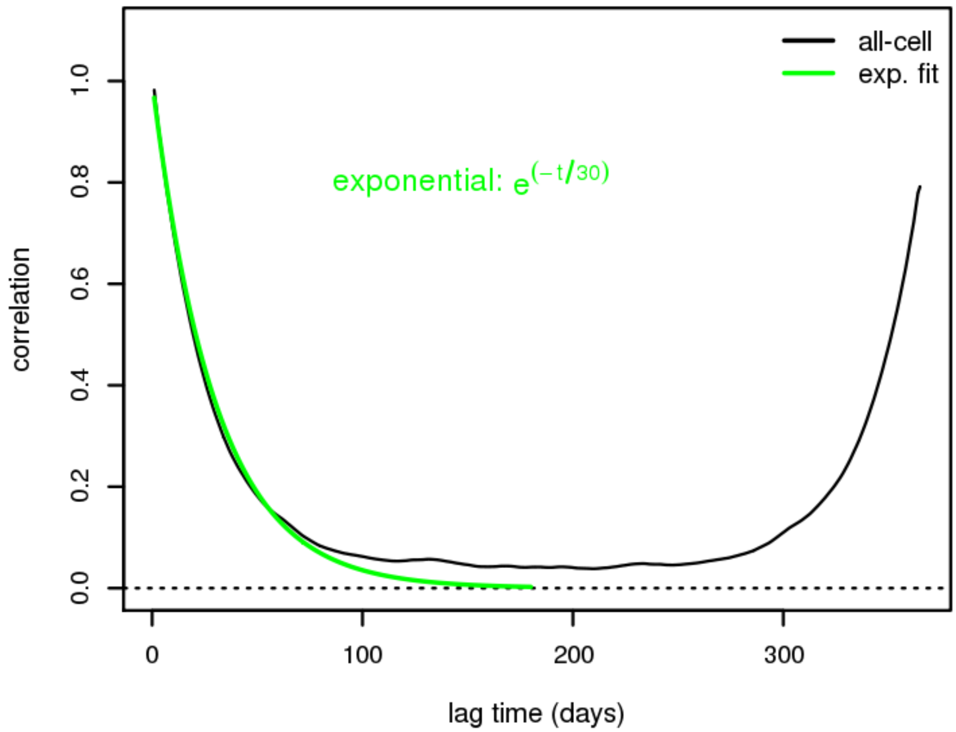

Figure A1: Autocorrelation function for a characteristic frequency of the exponential filter. The autocorrelation is calculated simultaneously for all the domain grid cells. The numerical realization of the autocorrelation does not decay to zero because of the flux seasonality.

References

Broquet, G., Chevallier, F., Bréon, F. M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO_2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, Atmos. Chem. Phys., 13, 9039-9056, doi:10.5194/acp-13-9039-2013, 2013.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T.and Ciais, P.: A European summertime CO_2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, J Geophys. Res., 116, D23303, doi:10.1029/2011JD016202, 2011.

Carouge, C., Bousquet, P., Peylin, P., Rayner, P. J., and Ciais, P.: What can we learn from European continuous atmospheric CO_2 measurements to quantify regional fluxes – Part 1: Potential of the 2001 network, Atmos. Chem. Phys., 10, 3107-3117, doi:10.5194/acp-10-3107-2010, 2010a.

Carouge, C., Peylin, P., Rayner, P. J., 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, 2010b.

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

Chevallier, F., Wang T., Ciais P., Maignan F., Bocquet M., Altaf A. M., Cescatti A., Chen J., Dolman A., J., Law B. E., Margolis, H. A., Montagnani, L., Moors, E. J.: What eddy-covariance measurements tell us about prior land flux errors in CO_2 flux inversion schemes, Glob. Biogeochem. Cy., 26, GB1021, doi:10.1029/2010GB003974, 2012.

Ciais, P., Peylin, P. and Bousquet, P.:, Regional biospheric carbon fluxes as inferred from atmospheric CO_2 measurements, Ecol Appl, 10, 1574-1589, doi: 10.2307/2641225, 2000.

Ciais, P., Borges, A. V., Abril, G., Meybeck, M., Folberth, G., Hausglustaine, D., and Janssens, I. A.: The impact of lateral carbon fluxes on the European carbon balance, Biogeosciences, 5, 1259-1271, doi: 10.5194/bg-5-1259-2008, 2008.

Ciais, P., Paris, J. D., Marland, G., Peylin, P., Piao, S. L., Levins, I., Pregger, T., Scholz, Y., Friedrich, R., Rivier, L., Houwelling, S., Schuldze, E. D., and members of the CARBOEUROPE SYNTHESIS TEAM: The European carbon balance. Part1: fossil fuel emissions, Glob Change Biol, 16, 1395-1408, doi: 10.1111/j.1365-2486.2009.02098.x, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597. doi: 10.1002/qj.828, 2011.

Enting, I. G., Trudinger, C. M., and Francey, R. J.: A synthesis inversion of the concentration and $\delta^{13}C$ of atmospheric CO_2, Tellus, ser.B, 47, 35-52, 1995.Ferrarese S., Apadula F., Bertiglia F., Cassardo C., Ferrero A., Fialdini L., Francone C., Heltai D., Lanza A., Longhetto A., Manfrin M., Richiardone R., Vannini C.: Inspection of high–concentration CO2 events at the Plateau Rosa Alpine station, Atmospheric Pollution Research 6, 415-427, doi:10.5094/APR.2015.046, 2015.

Fiedler, V., Dal Maso, M., Boy, M., Aufmhoff, H., Hoffmann, J., Schuck, T., Birmili, W., Hanke, M., Uecker, J., Arnold, F., and Kulmala, M.: The contribution of sulphuric acid to atmospheric particle formation and growth: a comparison between boundary layers in Northern and Central Europe, Atmos. Chem. Phys., 5, 1773-1785, doi:10.5194/acp-5-1773-2005, 2005.

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U.,

Formatted: Line spacing: 1.5 lines

Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R.: Comparing atmospheric transport models for future regional inversions over Europe - Part 1: mapping the atmospheric CO2 signals, Atmos. Chem. Phys., 7, 3461–3479, doi:10.5194/acp-7-3461-2007, 2007.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S. and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms , J Geophys Res – Atmos, 108, 4756, doi: 10.1029/2002JD003018, 2003a.

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

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.

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., 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 and CO_2 inversion intercomparison 1. Annual and mean control results and sensitivity to transport and prior flux information, Tellus 55B , 555-579, doi: 10.1034/j.1600-0889.2003.00049.x, 2003.

Gurney, K. R., Rachel M. L., Denning, A. S., Rayner, P. J., Bernard C. P., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyotov, S., Peylin, P., Prather, M. and Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, Global Biogeochem. Cy., 18, GB1010, doi:10.1029/2003GB002111, 2004.

Formatted: Space Before: 12 pt

Heimann, M. and Körner, S.: The global atmospheric tracer model TM3, Tech. Rep. 5, MPI BGC, Jena (Germany), online available at: http://www.bgc-jena.mpg.de/mpg/websiteBiogeochemie/Publikationen/Technical Reports/tech report5.pdf, 2003.

Houweling, S., Aben, I., Breon, F.-M., Chevallier, F., Deutscher, N., Engelen, R., Gerbig, C., Griffith, D., Hungershoefer, K., Macatangay, R., Marshall, J., Notholt, J., Peters, W., and Serrar, S.: The importance of transport model uncertainties for the estimation of CO2 sources and sinks using satellite measurements, Atmos. Chem. Phys., 10, 9981-9992, doi:10.5194/acp-10-9981-2010, 2010.

Jung, M., Reichstein, M., Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, 2009.

Jung, M., Henkel, K., Herold, M. and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sensing of Environment, 101(4), 534–553, doi:10.1016/j.rse.2006.01.020, 2006.

Kaminski, T. and Heimann, M.: A coarse grid three-dimensional global inverse model of the atmospheric transport 1. Adjoint model and Jacobian matrix, J. Geophys. Res. 104, D15, 18,535-18,553, doi: 10.1029/1999JD900147, 1999a.

Kaminski, T., Heimann, M. and Giering, R.: A coarse grid three-dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO_2 in the 1980s, J. Geophys. Res.: Atmospheres (1984--2012) 104, D15, 18,555-18,581, doi: 10.1029/1999JD900146, 1999b.

Kaminski, T., Rayner, P. J., Voßbeck, M., Scholze, M. and Koffi, E.: Observing the continental-scale carbon balance assessment of and sampling complementarity and redundancy in a terrestrial and assimilation system by means of quantitative network design, Atmos. Chem. and Phys. 12 , 7867-7879 , doi: 10.5194/acp-12-7867-2012, 2012.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, F. Th., Heimann, M.: Atmospheric CO2 inversions at the mesoscale using data driven prior uncertainties. Part2: the European terrestrial CO_2 fluxes, submitted to Atmos. Chem. Phys.

Kountouris, P., Gerbig, C., Totsche, K. U., Dolman, A. J., Meesters, A. G. C. A., Broquet, G., Maignan, F., Gioli, B., Montagnani, L., Helfter, C.: An objective prior error quantification for regional atmospheric inverse applications, Biogeosciences, 12, 7403-7421, doi: 10.5194/bg-12-7403-2015, 2015.

Lauvaux, T., Schuh, A. E., Bocquet, M., Wu, L., Richardson, S., Miles, N. and Davis, K. J.: Network design for mesoscale inversions of CO_2 sources and sinks, Tellus B 64, 17980, doi: 10.3402/tellusb.v64i0.17980, 2012.

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K., Ciais, P., Denning, A. and Rayner, P.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. and Phys., 8 , 3459-3471, doi:10.5194/acp-8-3459-2008, 2008.

Lin, J. C., and C. Gerbig: Accounting for the effect of transport errors on tracer inversions, Geophys. Res. Lett., 32, L01802, doi:10.1029/2004GL021127, 2005.

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

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.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y. and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Glob. Biogeochem. Cy. 22, GB2005, doi: 10.1029/2006GB002735, 2008.

Meesters, A. G. C. A., Tolk, L. F., Peters, W., Hutjes, R. W. A., Vellinga, O. S., Elbers, J. A., Vermeulen, A. T., van der Laan, S., Neubert, R. E. M., Meijer, H. A. J., Dolman, A. J.: Inverse carbon dioxide flux estimates for the Netherlands, J. Geophys. Res.-Atmos. 117, D20306, 1984-2012 , doi: 10.1029/2012jd017797, 2012.

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.

Michalak, A., Hirsch, A., Bruhwiller, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, J. Geophys. Res., 100, D24107, doi:10.1029/2005JD005970, 2005.

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. 113, D21114, doi: 10.1029/2007JD009734, 2008.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., B.Miller, J., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Wennberg, J. T. R. P. O., Krol, M. C. and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proceedings of the National Academy of Sciences, 104, 48, 18,925-18,930, doi: 10.1073pnas.0708986104, 2007.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences 10 , 6699-6720 , doi: 10.5194/bg-10-6699-2013, 2013.

Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P., and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling: model intercomparison, Atmos. Chem. Phys., 11, 6607-6622, doi:10.5194/acp-11-6607-2011, 2011.

Peylin, P., Rayner, P., 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.

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), Glob. Biogeochem. Cy. 19 , GB2026 , doi: 10.1029/2004GB002254, 2005.

Rivier, L., Peylin, P., Ciais, P., Gloor, M., Roedenbeck, C., Geels, C., Karstens, U., Bousquet, P., Brandt, J. and Heimann, M.: European CO_2 fluxes from atmospheric inversions using regional and global transport models, Climatic Change, 103, 93-115, doi: 10.1007/s10584-010-9908-4, 2010.

Rodgers, C., D. Inverse methods for Atmosphere Sounding: Theory and Practice, World Sci., River Edge, N. J., 2000.

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

Rödenbeck, C.: Estimating CO_2 sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport, Technical Report 6, Max Planck Institute for Biogeochemistry, Jena, http://www.bgc-jena.mpg.de/mpg/websiteBiogeochemie/ Publikationen/Technical Reports/tech report6.pdf, 2005.

Rödenbeck, C., Gerbig, C., Trusilova, K. and Heimann, M.: A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models, Atmos. Chem. and Phys. 9, 5331-5342, doi:10.5194/acp-9-5331-2009, 2009.

Rödenbeck, C., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., Cassar, N., Reum, F., Keeling, R. F. and Heimann, M.: Interannual sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme, Biogeosciences, 11(17), 4599–4613, doi:10.5194/bg-11-4599-2014-supplement, 2014.

Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., Andrews, A. E. and Worthy, D. E. J.: A regional high-resolution carbon flux inversion of North America for 2004, Biogeosciences, 7, 1625-1644 , doi: 10.5194/bg-7-1625-2010, 2010.

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, doi: 10.1029/2008JG000842, 2009.

Schulze, E. D., Ciais, P., Luyssaert, S., Schrumpf, M., Janssens, I. A., Thiruchittampalam, B., Theloke, J., Saurat, M., Bringezu, S., Lelieveld, J., Lohila, A., Rebmann, C., Jung, M., Bastviken, D., Abril, G., Grassi, G., Leip, A., Freibauer, A., Kutsch, W., Don, A., Nieschulze, J., Börner, A., Gash, J. H., and Dolman, A. J.: The European carbon balance. Part 4: integration of carbon and other trace-gas fluxes, Glob. Change Biology, 16, 1451-1469, doi: 10.1111/j.1365-2486.2010.02215.x, 2010.

~~Thompson, R. L., Gerbig, C. and Rödenbeck, C.: A Bayesian inversion estimate of N_2O emissions for western Europe and the assessment of aggregation errors, Atmos. Chem. Phys., 11, 3443-3458, doi: 10.5194/acp-11-3443-2011, 2011.~~

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, ISBN: 0-89871-572-5, siam, 2005.

Tolk, L. F., Dolman, A. J., Meesters, A. G. C. A. and Peters, W.: A comparison of different inverse carbon flux estimation approaches for application on a regional domain, Atmos. Chem. Phys., 11, 10349-10365, doi: 10.5194/acp-11-10349-2011, 2011.

Trusilova, K., Rödenbeck, C., Gerbig, C., and Heinmann, M.: Technical Note: A new coupled system for global to regional downscaling of CO_2 concentration estimation, Atmos. Chem. Phys. 10, 3205-3213, doi:10.5194/acp-10-3205-2010, 2010.

Trusilova, K., and Churkina, G.: The Terrestrial Ecosystem Model GBIOME-BGCv1, Max-Planck Institute for Biogeochemistry, Technical Report 14, http://www.db-thueringen.de/servlets/DerivateServlet/Derivate-20689/tech_report14.pdf, 2008.

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.

Table 1. Optimized VPRM parameters SW_0, λ_{SW}, α, β for different vegetation classes[a]

	SW_0	λ_{SW}	α	β
Evergreen forest	275	0.226	0.288	-1.10
Deciduous forest	254	0.215	0.181	0.84
Mixed forest	446	0.163	0.244	-0.49
Open shrub	70	0.293	0.055	-0.12
Crop	1132	0.086	0.092	0.29
Grass	528	0.119	0.125	0.017

[a]Units are as follows: SW_0 : W m^{-2}; λ_{SW}: μmole CO_2 $m^{-2}s^{-1}$ / (W m^{-2}); α: μmole CO_2 $m^{-2}s^{-1}$ / ^{0}C; β: (μmole CO_2 $m^{-2}s^{-1}$).

Table 2. Information on the stations used for the regional inversions. Same network applied for the synthetic, and the real data inversions in Kountouris et al. (2016). In first column the term "type" stands for continuous (C) or flask (F) data.

Site Code / type	Name	Latitude (°)	Longitude (°)	Height (m.a.s.l.) (m)	Measurement height (above ground) (m)	Model height
BAL/F	Baltic Sea, Poland	55.50	16.67	8	57	28
BIK/C	Bialystok, Poland	53.23	23.03	183	90	90
CBW/C	Cabauw, Netherlands	51.58	4.55	-2	200	200
CMN/C	Monte Cimone, Italy	44.18	10.7	2165	12	670
HEI/C	Heidelberg, Germany	49.42	8.67	116	30	30
HPB/F	Hohenpeissenberg, Germany	47.80	11.01	934	50	10
HUN/C	Hegyhatsal, Hungary	46.95	16.65	248	115	96
JFJ/C	Jungfraujoch, Switzerland	46.55	7.98	3572	10	720
KAS/C	Kasprowy Wierch	49.23	19.93	1987	5	480
LMU/C	La Muela, Spain	41.36	-1.6	570	79	80
MHD/C	Mace Head, Ireland	53.33	-9.90	25	10	15
OXK/C	Ochsenkopf,	50.03	11.81	1022	163	163

	Germany					
PRS/C	Plateau Rosa, Italy	45.93	7.71	3480	-	500
PUY/C	Puy De Dome, France	45.77	2.97	1465	10	400
SCH/C	Schauinsland, Germany	47.92	7.92	1205	~~-~~8	230
WES/C	Westerland, Germany	54.93	8.32	12	-	15

Table 3. RMSD (first column in ppm) and correlation coefficients (second column) between known truth and prior/posterior CO_2 dry mole fractions for daily "daytime" or "nighttime" averaged values and for each station. The third column shows χ^2, the normalized dry mole fraction mismatch per degree of freedom for 7-day averaged residuals, as a measure of how well the data were fitted. The format for each station is as follows: RMSD | r^2 | χ^2.

	Prior	B1	S1
BAL	4.78 | 0.07 | 18.44	0.89 | 0.97 | 0.48	1.02 | 0.96 | 0.37
BIK	5.28 | 0.43 | 15.50	1.20 | 0.97 | 0.18	1.29 | 0.97 | 0.25
CBW	8.60 | 0.04 | 74.29	0.99 | 0.99 | 1.31	1.06 | 0.99 | 1.34
CMN	2.68 | 0.33 | 6.31	0.74 | 0.93 | 0.08	0.78 | 0.92 | 0.10
HEI	11.39 | 0.37 | 12.97	1.83 | 0.98 | 0.36	1.84 | 0.98 | 0.37
HPB	7.73 | 0.35 | 26.58	1.01 | 0.99 | 0.21	1.19 | 0.99 | 0.31
HUN	6.50 | 0.63 | 31.89	1.36 | 0.98 | 0.21	1.46 | 0.98 | 0.25
JFJ	3.12 | 0.21 | 3.93	1.24 | 0.86 | 0.24	1.31 | 0.84 | 0.27
KAS	4.00 | 0.32 | 10.67	0.73 | 0.98 | 0.11	0.80 | 0.97 | 0.15
LMU	3.42 | 0.19 | 6.5	0.79 | 0.95 | 0.12	0.86 | 0.94 | 0.16
MHD	1.53 | 0.0002 | 0.83	0.65 | 0.09 | 0.16	0.68 | 0.06 | 0.17
OXK	6.10 | 0.21 | 38.50	3.35 | 0.76 | 0.76	3.40 | 0.75 | 0.80
PRS	2.32 | 0.15 | 2.46	0.70 | 0.92 | 0.30	0.74 | 0.91 | 0.33
PUY	4.27 | 0.15 | 12.06	0.68 | 0.97 | 0.06	0.73 | 0.~~15~~96 | 0.09
SCH	4.76 | 0.26 | 21.17	0.90 | 0.97 | 0.07	0.95 | 0.97 | 0.09

Table 4. Performance of the two error structures expressed as the spatial RMSD of the optimized monthly and annual NEE fluxes compared to the truth for the whole domain in μmole m^{-2} s^{-1}.

	Annual	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
prior	0.38	0.61	0.53	0.55	1.06	1.26	1.56	1.17	0.94	0.65	0.57	0.63	0.63
B1	0.33	0.46	0.40	0.45	0.84	0.99	1.21	1.00	0.86	0.63	0.43	0.46	0.44
S1	0.34	0.48	0.41	0.45	0.86	1.01	1.24	1.03	0.86	0.63	0.45	0.47	0.45

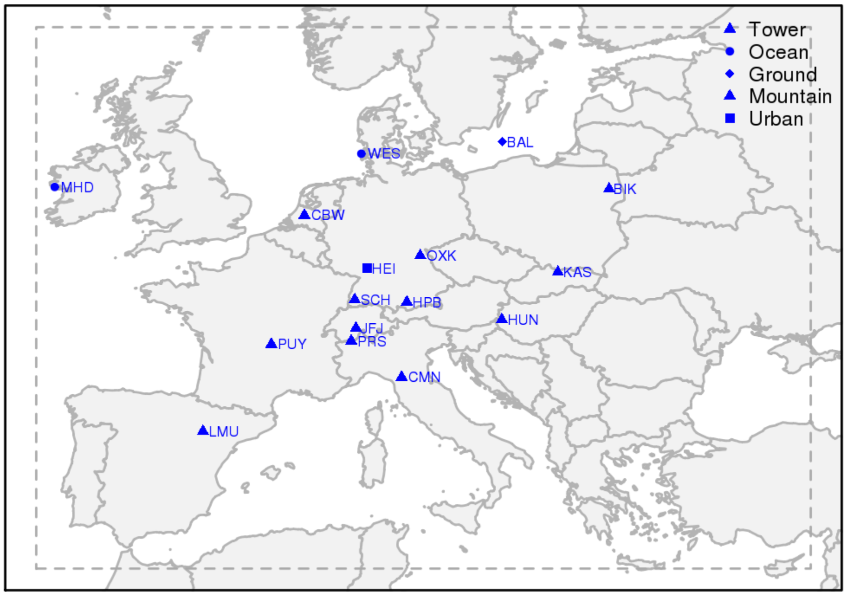

Figure 1. Domain of the inversions (dashed rectangle). Locations of the atmospheric measurement stations are shown with blue marks. Red stars denote the eddy covariance locations used for flux comparisons at grid scale.

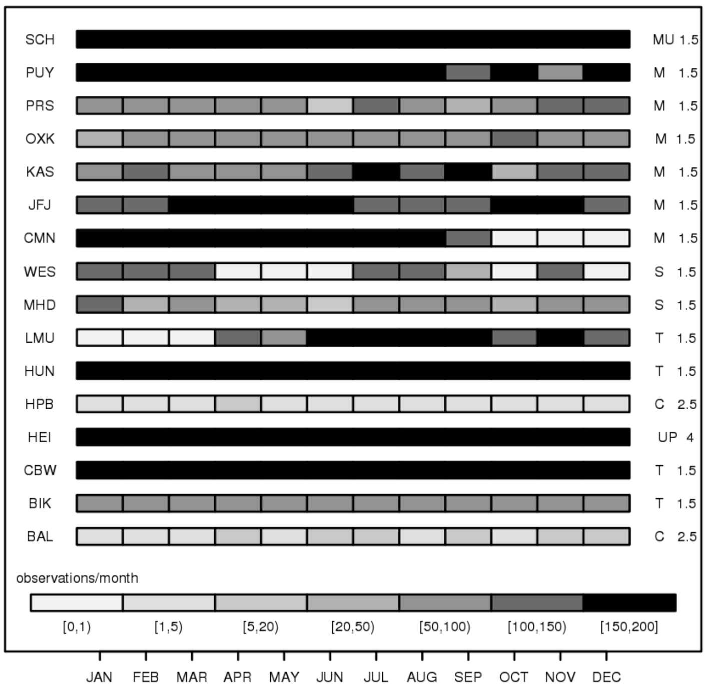

Figure 2. Monthly data coverage plot for the atmospheric stations used in the regional inversions. Left column shows the code name and the right columns show the station class and the assigned uncertainty in units of ppm. “C” stands for continental sites near the surface, “T” for continental tall towers, “S” for stations near shore, “M” for mountain sites, “MU” for mountain sites with diurnal upslope winds and “UP” for urban pollutant.

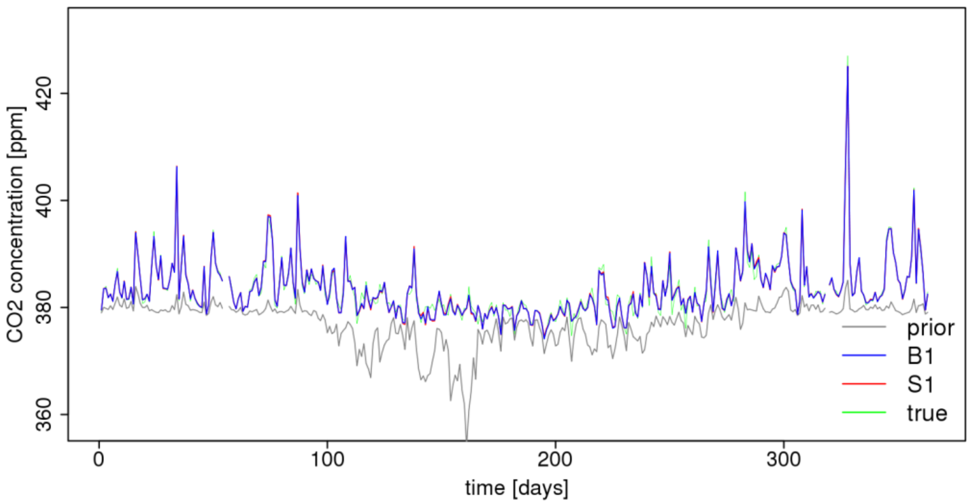

Figure 3. Daily nighttime (23:00-4:00 UTC) averages for prior, true, and posterior CO_2 dry mole fraction time series for the mountain site Schauinsland. Time starts at 1st January 2007. Note due to the almost perfect fit posterior and true time-series overlap to each other.

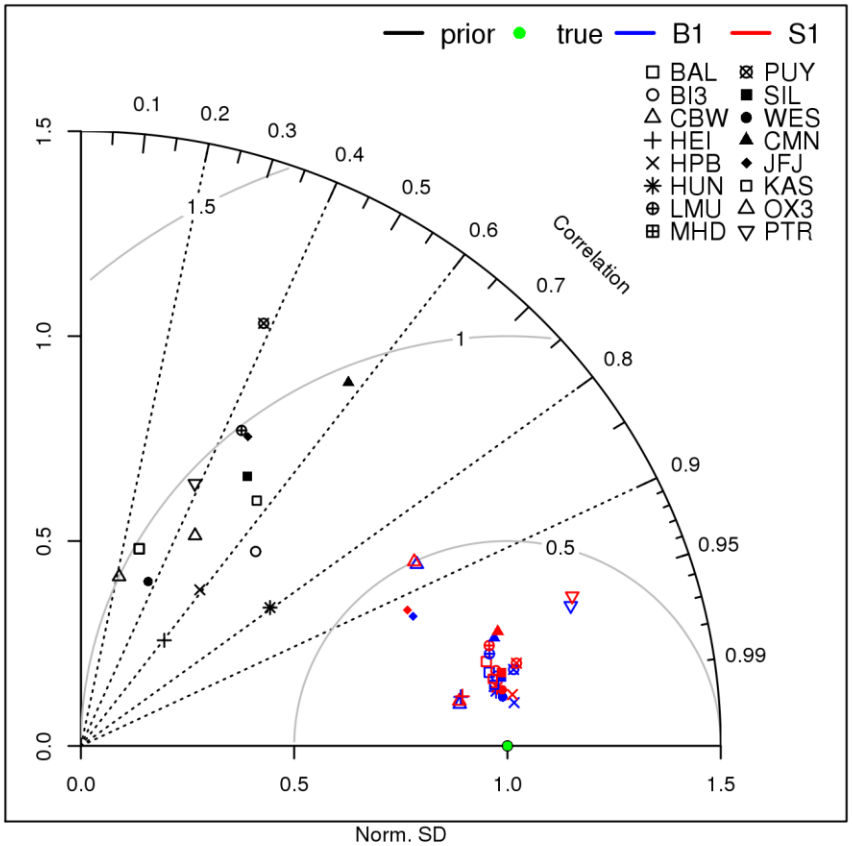

Figure 4. Taylor diagram for daily averaged modeled and measured time-series (annual basis) of CO_2 dry mole fractions. Prior (black), true (green, the perfect match of modeled and true time-series) and the different inversion cases (B~~R0~~1 blue; S~~R~~1 red) are displayed. Different symbols denote different atmospheric stations. The normalized SD was calculated as the ratio~~n~~ of the SD of the modeled time-series to the SD of observations.

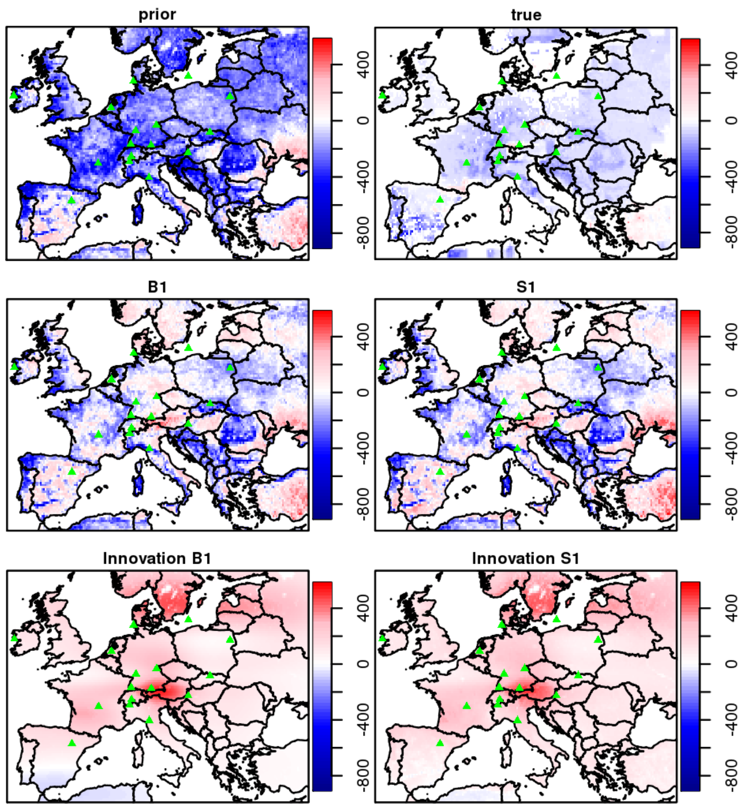

Figure 5. Annual spatial distribution for the prior, true, and posterior biogenic flux estimates for the two synthetic inversions S1 and B1 (top two rows), and flux innovation defined as the difference posterior - prior (bottom row). Fluxes are given in units of gC~~y^{-1}~~m^{-2} y^{-1}.

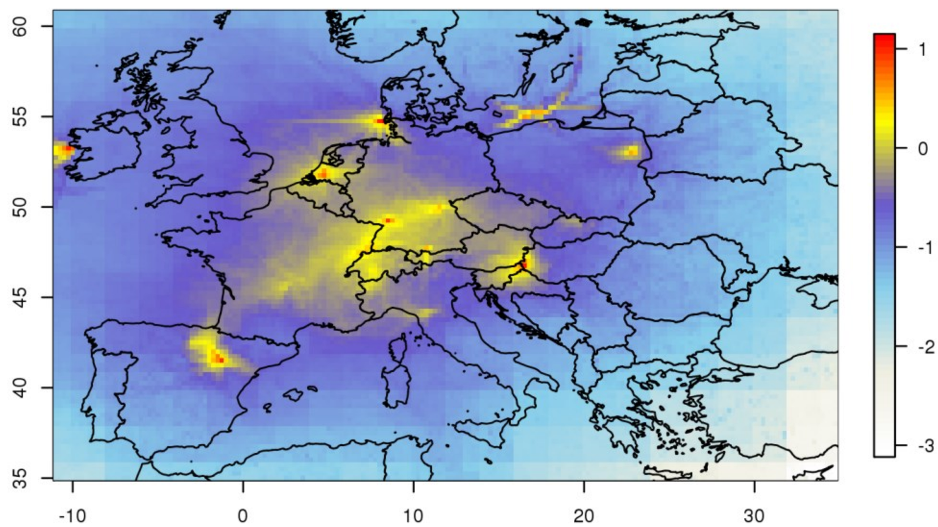

Figure 6. Annual integrated influence for 2007 of the current atmospheric network. Footprint influence is presented in a logarithmic scale and units are in $\log_{10}[\mathrm{ppm}/(\mu\mathrm{mol}/\mathrm{m}^2/\mathrm{s})]$

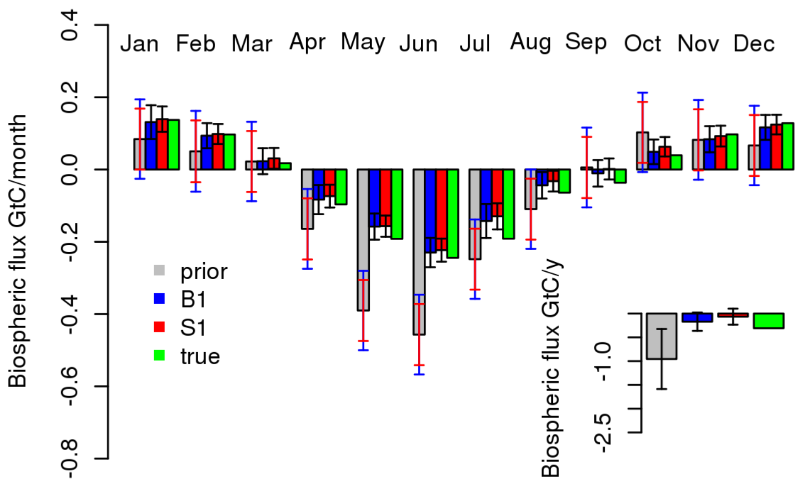

Figure 7. Monthly and annual carbon flux budget, integrated over the European domain. Note that both inversions share the same annual prior uncertainty but monthly uncertainties differ. Blue and red error bars denote the prior uncertainty for the B1 and S1 scenarios respectively.

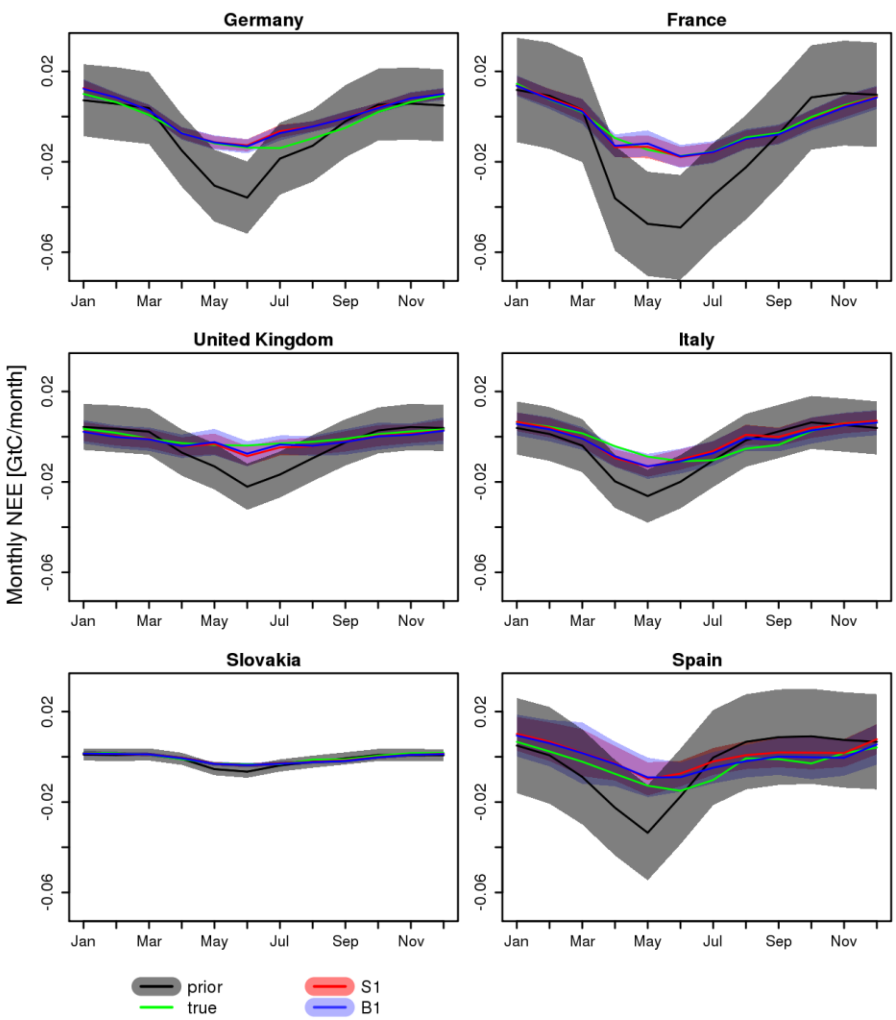

Figure 8. Temporal evolution of monthly NEE for selected European countries for the synthetic data inversion.

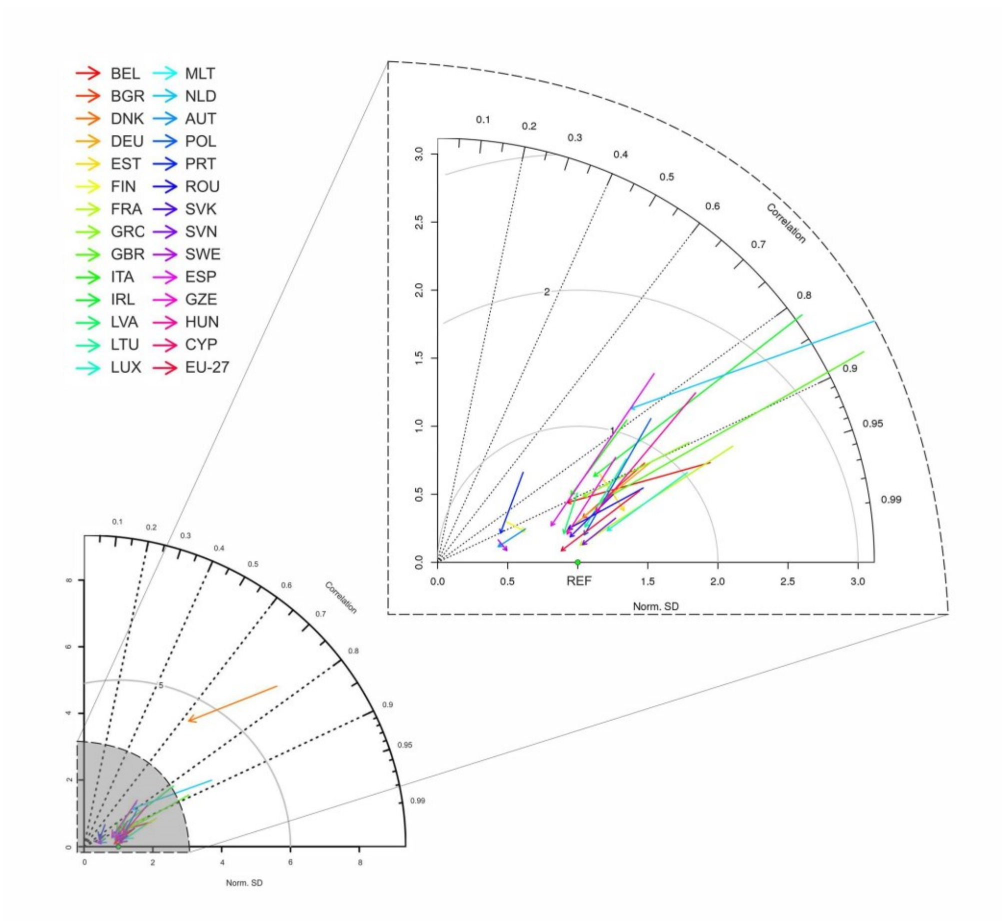

Figure 9. Overview of the model performance (S1 case) summarized in a Taylor diagram. Posterior and prior monthly and country scale aggregated biospheric fluxes are compared against the reference fluxes ("true"). Each line corresponds to a different country. The starting point of each arrow shows prior/reference comparison and the ending point the posterior/reference comparison. Ideally the ending point should coincide with the green point which represents the reference model.

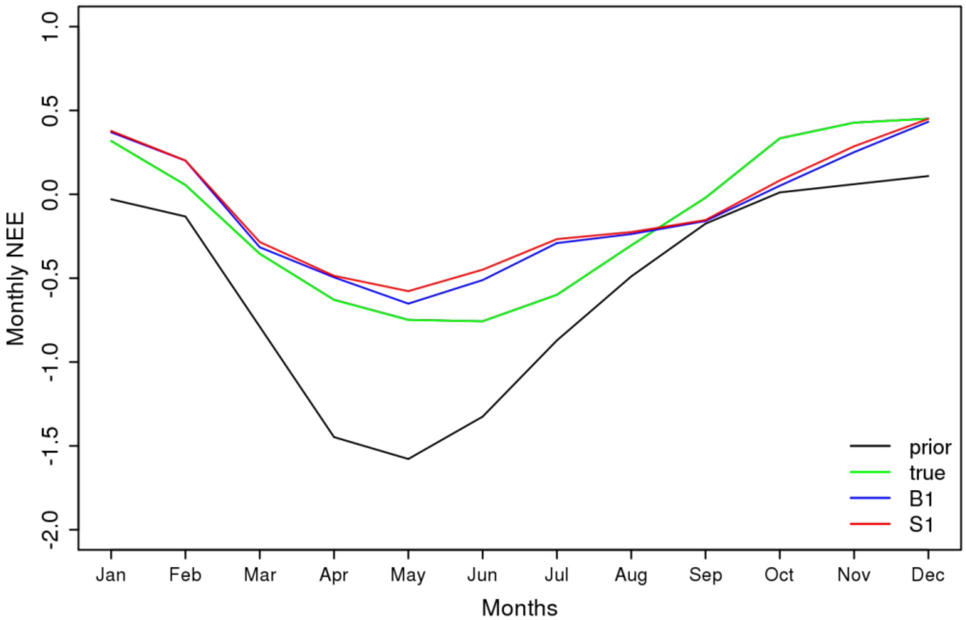

Figure 10. Mean monthly NEE averaged over the 53 different eddy covariance site locations as reported in Kountouris et al. (2015). A priori (black), true (green), and posterior fluxes for scenarios B1 (blue) and S1 (red) are shown. Units are in $gCm^{-2}day^{-1}$.

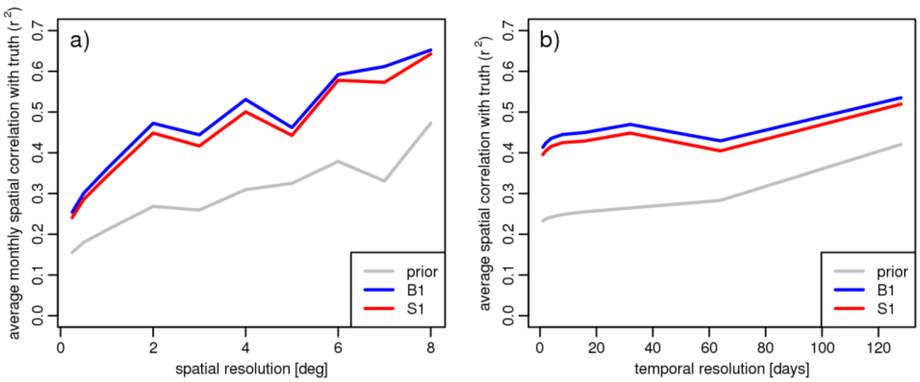

Figure 11. a): Mean spatial correlation of monthly fluxes with true fluxes as function of spatial flux aggregation scale for prior fluxes (grey), and for posterior fluxes from scenarios B1 (blue) and S1 (red). b): Mean spatial correlation of fluxes with true fluxes at 2 deg. spatial resolution as function of temporal flux aggregation scale for prior fluxes (grey), and for posterior fluxes from scenarios B1 (blue) and S1 (red).

Clarification on how the author response is structured:

With bold letters we present the comments and questions of the referees. The page and line numbering is linked to the published paper on the public discussion. Response from the authors follows a non bold typesetting. Note that page and line numbers are linked to the corrected and change tracked document.

Anonymous Referee #1

We thank the referee for the comments on our manuscript, which helped improving our study. We hope that our answers and the modifications are satisfactory.

General comments: The paper is quite well written. The inversion system seemed to do its job pulling the posterior fluxes toward the true flux. However, how accurate is the "true" flux? Perhaps, the author can add more information on how reliable the "truth" was.

"True" fluxes were created from the GBIOME_BGC biosphere model. This is a terrestrial ecosystem process model which requires only standard meteorological data like, daily maximum-minimum temperature, precipitation, incoming shortwave solar radiation, vapor pressure deficit (VPD), and the day length (DLn). How accurate the modeled fluxes are, is difficult to say, since this would require a validation against observed fluxes from eddy covariance stations. The accuracy though of the modeled fluxes is still debatable. We would like to refer to a study from Friedlingstein et al. (2014). Especially in figure 4 panel d, we see the large range that terrestrial carbon flux estimates between the different models exhibit. However, in the current experiment the accuracy of "true" fluxes is not of a concern. We are interested just to recover the known spatial and temporal flux field using only atmospheric observations. Accuracy is essential in the part 2 of this study, where real measurements are used, and the optimized flux field is being validated against real eddy covariance measurements.

P11, L10 we added:

"The BIOME-BGC model is a terrestrial ecosystem process model which requires only standard meteorological data like, daily maximum-minimum temperature, precipitation, incoming shortwave solar radiation, vapor pressure deficit (VPD), and the day length (DLn). How accurate the modeled fluxes are, is difficult to say, since this would require a validation against observed fluxes from eddy covariance stations. Nevertheless, biospheric models still suffer from large uncertainties. The remarkably diverge results between models confirm how uncertain models are (see Friedlingstein et al. (2014)). However, in the current experiment the accuracy of the "true" fluxes is not of a concern, since we aim only to create a synthetic flux field that we perfectly know."

Specific comment:

P12, L14-15: what was the purpose of setting the bias term according to the annually averaged VPRM respiration only?

The bias shape selection (respiration shape) was preferred over the NEE fluxes, as otherwise a priori neutral pixels would not receive any flux correction. Further, allowing the bias to have a spatial shape instead of a flat one, might be sound, since regions with stronger fluxes might be also more biased." However in part 2 of this study we use different shapes of the bias term to investigate how the bias spatial structure impacts the posterior flux estimates.

P13, L18 we added: "The idea behind the implementation of this term is that at large scales a bias might exists, which can not be captured in the model-data residual autocorrelation analysis (EC measurements are representative at scales ~ 1 km). This assumption avoids the artificial inflation of the uncertainty at pixel scale, and restricts the pixel to pixes corrections to be statistically consistent with the actual error structure. The bias shape selection (respiration shape) was preferred over the NEE fluxes, as otherwise a priori neutral pixels would not receive any flux correction. Further, allowing bias to have a spatial shape might be sound, since regions with stronger fluxes might be also more biased."

P19, L10: Specify the Figure number.

It is defined but the number is separated and it is in the next line. We correct though and now reads "Fig. 10" instead of Figure 10.

P19, L10 and 12: gCm2y-1 → gCm-2y-1

P22, L25 and 27 we corrected : "$gCm^{-2}day^{-1}$"

P34, Table 2: At some sites, the model and the measurement heights are significantly different. What was the reason for that? And at some sites, the measurement heights were not specified ("-").

Mountain stations are more difficult to represent in models because the orography is generally not adequately resolved. In STILT we use surface elevation maps from ECMWF (European Centre for Medium Range Weather Forecasting) with a resolution of 0.25 x 0.25 degrees. As the model orography represents an average over the whole grid cell, it is, in particular at steep mountain sites, significantly smaller compared to the real orographic height at the station location. In order to better represent the location of the station in the large scale flow, usually a model height is assumed that more closely represents the real height (above sea level) of the measurements. However, using exactly the measurement height (a.s.l.) in the model would decouple the CO2 signal too strongly from the surface fluxes and hence lead to a systematic underestimation of the surface influence on the concentrations (Geels et al., 2007). We try to find a compromise and adjust the model height (above ground) by half the distance between the model orographic height and the real station height. This assumption is supported by comparisons of modeled and observed diurnal cycles of CO2 concentration at mountain sites. For sites that a dash (-) appears, means that we have no information about the height

P17 L1 we added: "With respect to the assumed model height, STILT uses surface elevation maps from ECMWF (European Centre for Medium Range Weather Forecasting) with a resolution of 0.25 x 0.25 degrees. As the model orography represents an average over the whole grid cell, it is, in particular at steep mountain sites, significantly smaller compared to the real orographic height at the station location. In order to better represent the location of the station in the large scale flow, usually a model height is assumed that more closely represents the real height (above sea level) of the measurements. However, using exactly the measurement height (a.s.l.) in the model would decouple the CO_2 signal too strongly from the surface fluxes and hence lead to a systematic underestimation of the surface influence on the concentrations (Geels et al., 2007). A compromise was reached by adjusting the model height (above ground) by half the distance between the model orographic height and the real station height."

P36, Table 3: I noticed that the statistical values of B1 and S1 are quite close except at PUY site r2 (0.97 and 0.15 for B1 and S1, respectively). I am wondering what happened there.

We thank the reviewer for spotting this mistake. We corrected it, the correct value is 0.96 also for the S1 case at this station.

P40, Figure 3: I cannot clearly see the "true" flux line (light green). Perhaps, make it more visible.

The reason is that the modeled dry mole fractions (running the model forward using the optimized fluxes) fit almost perfect to the true. Only by plotting the true last would be visible but then, we could not see the posterior estimates. Unfortunately (or maybe not) changing the color would not work.

Figure 3 we added a note in the caption: "Note due to the almost perfect fit posterior and true time-series overlap to each other."

P42, L4: Usually see gCm-2y-1 instead of gCy-1m-2 P44,

P46, L4 we corrected : "$gCm^{-2}\ y^{-1}$"

Figure 7: Why in many months posterior flux values are not in between prior and true fluxes?

Please see explanation in response to Referee 3, general comment 3.

Anonymous Referee #2

We thank the referee for the comments on our manuscript, which helped improving our study. We hope that our answers and the modifications are satisfactory.

General comments:

This paper describes calculations of CO2 fluxes for Europe based on inversion from synthetic concentrations. It serves as preparation of a second part where observed concentrations are used. The title announces that "data driven prior uncertainties" will be used. But there is a substantial issue with this. It is important to note that the paper has a precursor in Kountouris et al. (2015) where prior flux errors are estimated based on comparison of model results and real (eddy correlation) flux observations. There, remarkably small flux error correlation lengths of up to 40 km are found (see page 6 L 7 in the present paper). When this is imposed on the prior flux error matrix, this leads to "exceptionally small" (L 9) estimates of the error in the continental integrated prior flux. Apparently, this constitutes a problem: in the end, the authors decide to use a much larger correlation length (of 566 km on average, see page 12 Ls 3-7), which is based on an investigation of model-model residuals (page 11 Ls 19-21, and Abstract). Unfortunately, this means that the "data driven prior uncertainties" claim in the title no longer holds. This also undermines the innovative pretention expressed in the title. An interesting innovation is the use of an extra "bias" term in the flux, consisting of a "known" spatial flux field multiplied with an unknown time series to be determined by optimal fitting. This avoids the artificial inflation of errors to obtain an acceptable result. Maybe, more could be said about its proposed physical interpretation (which is now indicated very briefly on page 13 in Ls 11-12).

The authors would like to clarify the scope of this paper. In part 1 (current paper) of this study, we evaluate the inversion system in a synthetic experiment by using the same methodology as in the real data inversion. In the synthetic experiment, the known truth, as well as the prior are derived from biospheric models. The real error structure for this case clearly can not be derived from the analysis of model-data mismatch as in Kountouris et al. (2015), but instead from the analysis of model-model mismatch, i.e. the mismatch between true fluxes and a priori fluxes in the synthetic experiment. This paper should be considered as an evaluation of the method we follow, to quantitatively characterize the prior error structure, rather than making assumptions or using expert knowledge. Then in part 2 of this study, which uses real data, we do make use of data driven uncertainties, as those investigated in Kountouris et al. (2015).

Regarding the bias term we added more information in the paper.

P13, L18-25: now reads:

"…The idea behind the implementation of this term is that at large scales a bias might exists, which can not be captured in the model-data residual autocorrelation analysis (EC measurements are representative at scales ~ 1 km). This assumption avoids the artificial inflation of the uncertainty at pixel scale, and restricts the pixel to pixel corrections to be statistically consistent with the actual error structure. The bias shape selection (respiration shape) was preferred over the NEE fluxes, as otherwise a priori neutral pixels (with zero NEE) could not be bias corrected. Further, allowing bias to have a spatial shape might be sound, since regions with stronger fluxes might be also more biased."

Minor comments

P2, L5: "it is used in such a way", "is used"

P2, L5 we corrected: "…but is used to.."

P4, Ls 12-16: does this involve nonLinearity ? comment on this.

Yes, this involves nonlinearity. We have added this in the introduction:

P4, L16 we added: "…instead of the fluxes themselves; this CCDAS approach also allows for nonlinear dependencies of the fluxes on the parameters."

P5, L8: delete "zone"; "later", "latter"

P5, L10 we corrected: "…the climate of the latter varies…"

P5, L10: "with", "for distances up to" ?

P5, L12 we clarify: "This leads to correlation lengths approximately two times larger compared to the pulse length."

P5, Ls 16-19: be more specific

P5, L22- L25 we clarify: “. In particular correlations are applied such that the same ecosystem types in different TransCom regions (basis function regions, see also http://transcom.project.asu.edu/transcom03_protocol_basisMap.php) decrease exponentially with distance (L=2000km), and thus assumes a coupling between the behavior of the same ecosystem.”

P6, Ls 26-end: this is somewhat difficult to follow.

P6 L2-L7 we clarify: “They simultaneously estimate posterior fluxes as well as parameters controlling the model-data mismatch uncertainty and the prior flux uncertainty, including variance as well as spatial and temporal correlation lengths. Although this approach may be considered as an objective way to infer spatial and temporal correlation lengths, it forces the structural parameters of the error covariance to be statistically consistent with the atmospheric data. In other words, flux parameters are optimized from atmospheric concentration data, and they are forced to have values which can reproduce the atmospheric data.”

P6, L3: “or”, “respectively” ?

P6, L17 we clarify: “…models and fluxes measured…”

P6, L16: “for”, “integrated over”

P6, L30: we corrected: “…NEE fluxes integrated over the””

P6, L17: “Although is”, “Although it is”.

P6, L31: we corrected: “…Although it is…”

P7, L7: “term referred to a bias term”, “term to reflect the bias” ?

P7, L21: we corrected: ”…term to reflect a bias term.”

P7, L9: “between” : another word is needed here.

P7, L23 : we clarify: “…which couples…”

P7, L27: “conclusions are following in Section 4”: these are presently in Section 5.

P8, L10 we clarify: “…4 and 5, respectively.”

P8, L17: “cini is the initial concentration”: is this correct? With f =0, cmod would still evolve in time.

We assume a well-mixed (uniform) initial concentration at the beginning of the model run, thus it just remains a constant concentration offset throughout the simulation. A well-mixed initial concentration can be assumed because any spatial structure would be lost during the spin-up period anyway. Deviations of the true initial conditions from c_{ini} are taken into account through flux adjustment during the spin-up period.

P9 L4 we added: “The initial concentration assumed to be well mixed and remains constant throughout the simulation. The assumption of the well mixed initial concentration is considered to be valid, since any spatial structure would be lost during the spin-up period.”

P9, L6: “constrain”, “constraint”

P9, L20: we corrected: “…constraint…”

P11, Ls 1-4: the wording is a bit confused.

P11, L22 we clarify: “We note that for the synthetic case the last two a priori terms are set to zero. Similarly the deviation term (the data-derived correction to the a-priori fluxes) of the flux model (Eq. 6) consists of the terms referring to NEE, fossil fuel, and ocean fluxes. Here in the synthetic case the last two terms are set to zero (i.e. they are not optimized).”

P11, equation 6: apparently not referred to and of unknown use.

P11, L23: we added the equation in the text

P11, Ls 8-12: this is an errant block, it should come later.

P12, L4-8: the text has been moved.

P11, L12: delete “et al.”

P14, L1 we corrected: “…to Rödenbeck (2005)”

P11, Ls 16-21: there is a difference in method here: Kountouris et al. (2015) used model-data instead of model-model comparison. And the resulting correlation lengths are also very different, which should be indicated.

P12, L18-21 we clarify: "We note that the current study does not directly make use of the error structure derived in Kountouris et al. (2015), since this is applicable for real data inversions. Instead we use the same methodology to derive the actual model-model error structure since here we perform a synthetic data inversion, exploring amongst others the accuracy of this method."

However, we do not claim that we use model-data error structure in this study. We have explicitly written in the same paragraph, that the error structure is estimated according to the method followed by Kountouris et al. (2015) and that the residual autocorrelation analysis is performed for the model-model residuals.

P11, L23: "ensuring similarity": same remark.

Similarity does not mean the same error structure but rather the same methodology used to characterize the error structure (e.g. which analysis is used, which stations are assumed and at which locations, at which spatial temporal scales etc).

P12, Ls 12-13: Not sure if the acronyms "B1" and "S1" would be the best choice, one might think of more telling names.

Although the reviewer is right about the names those acronyms serve a distinct role. First we make sure we are consistent with the acronyms of the second part of this study. Secondly in the second part, quite a number of inversions are performed. We consider that this frugal acronym would be the best choice for the reader to distinguish later the different inversions. A table explaining all the different scenarios is given in the second part of this work (table 2 Kountouris et al., 2016)

P12, L28: "and unit variance": this pertains not to the adjustable term but to the p-coefficients.

P14, L9 we corrected by deleting the sentence "and unit variance"

P13, L4: "which they a-priori have, a", "which a priori have a"

P14, L13: we corrected: "parameters which a priori have a"

P13, L6: "derived", "expressed" (nothing is said yet about how values are derived)

P14, L23 we corrected and now it reads "expressed"

General about section 2.2.2: It remains unclear in the paper how posterior errors and covariances are derived.

P15, L6-9 we clarified: "Following Rodgers 2000, the posterior flux uncertainties are contained in the covariance matrix of the posterior probability distribution which can be estimated from eq. (10)

$$Q_{f,post} = ((A \cdot F)^T \cdot Q_c^{-1} \cdot (A \cdot F) + Q_{f,pr}^{-1})^{-1} \tag{10}$$

where Q_c is the measurement error covariance matrix."

P13, L21: "use a different biosphere model": add eventually references to literature where the same is done, like in the previous sentences.

P15, L18 we clarify: "Such an approach is also used by Tolk et al. (2011)."

P14, L3: "table 2": and figure 1.

P15, L27 we corrected: "table 2 and Fig. 1. "

P14, L24: "DoI": explain that this means domain of interest.

P16, L23 we clarify: "…within the Domain of Interest (DoI)"

Section 2.3: a separate subsection may be superfluous, instead the content could be built in within the results section.

We modified by deleting the section 2.3. The content in now incorporated into sections 3.1 and 3.2

P15, L9 and 10: Unclear sentence. "a-priori" in L9 and "optimized" in L10 seem to contradict each other.

In statistics the reduced chi square is a weighted sum of squared deviations. The nominator consists of the squared differences between the observations (i.e "true" fluxes) and the calculated data (i.e. the optimized fluxes). The denominator represents the input variance (i.e the a priori uncertainties). Hence by definition, this metric evaluates the a priori uncertainties by comparing them to the squared residuals between observed and optimized data.

P19, L19 we clarified: "Another important aspect is the reduced χ_r^2 metric, which we use to assess the model performance. By definition the reduced χ_r^2 can be obtained by dividing the squared residuals of optimized minus observed dry mole fractions by the squared specified uncertainties."

P17, L6: "central Europe": also south Scandinavia

P20, L17 we corrected: "in central Europe and in southern Scandinavia,"

P17, L10: "measures", "measured".

P20, L22 we corrected: "measured"

P17, Ls 11-12: is this shown anywhere in the paper?

We computed and present within the text the squared Pearson correlation coefficients. The authors feel that it would be superfluous to make a plot or a table for those 3 coefficients.

P17, L24: "found", "was found".

P21, L5 we corrected: "…(true – posterior) was found to have…"

P18, L24: inversion performance: for which of the two inversions? see also question at figure 9.

P22, L9 we clarify: "…performance for the S1 case and for each…"

P19, L10: "Figure": Figure 10.

We corrected: “Fig. 10”

P19, L28: “65 %”, “64 %”: where is this stated ?

This is a trivial calculation which we state directly here.

P20, L9: “nearly continuous”, “nearly monotonous” ?

P24, L2 we corrected: “nearly monotonous”

P23, L14: “years”: reciprocal years.

The autocorrelation time is given in years, and amounts to 1/12 years. So it is not “reciprocal years”.

Figure 4: “R0”, “R1”: wrong acronyms. “ration”, “ratio”. With which time base were the results obtained ?

Figure 4 : We corrected and clarified: “Taylor diagram for daily averaged modeled and measured time-series (annual basis) of CO_2 dry mole fractions. Prior (black), true (green, the perfect match of modeled and true time-series) and the different inversion cases (B1 blue; S1 red) are displayed. Different symbols denote different atmospheric stations. The normalized SD was calculated as the ratio of the SD...”

Figure 5: “gCy-1m-2 “: usually this is written as “gCm-2yr-1 “.

We corrected and now reads “gCm^{-2} y^{-1}”

Figure 9: colors will be often indiscernible in practice (maybe no problem !); why is one arrow seen when there are two ways to calculate a posterior?

This figure is meant to give an overview of the model performance at country scale. For a more detailed and country specific assessment we used figure 8, but extending to all countries might be superfluous. Hence we make use of the tailor diagram (fig. 9) to summarize the results. We use only the S1 case for the sake of simplicity, since another 27 arrows in the plot would make it not readable. Further, as the results suggest, the two methods are very close to each other, hence it would not add any significant information.

Anonymous Referee #3

We thank the referee for the comments on our manuscript, which helped improving our study. We hope that our answers and the modifications are satisfactory.

Introductory remark:

Whilst the paper is generally well written, I was left wondering what we've really learnt from a study such as this. At present, the abstract and conclusions largely focus on the outcome of the synthetic data inversion, which I don't believe represent a major innovation, or provide a framework that could readily be used in other work (see below). Perhaps the paper can be re-focused on elements that the authors feel represent a true advance, that could be applied beyond the inversion system described. Alternatively, it appears that the authors have attempted to split this work into two publications: whilst I haven't read the companion paper, I wonder whether the work in this paper is too incremental to stand on its own, and could instead be folded into the other work (provided the below comments can also be addressed)?

We disagree with the reviewer in that the paper is too incremental to stand on its own. It is correct that the work has been split into two publications, as this would increase the readability of the paper. The current split helps also a reader that is only interested in the methodological part of the study, including the prior error characterization and the inverse system description. A reader that is interested more in the real data results and the regional European carbon budget including series of sensitivity and case runs, can directly refer to the second part, avoiding all the theory and methodology used behind the inversions.

General comments:

1.

I'm not convinced that, with a synthetic data experiment such as this, it is possible to show whether a particular prior flux uncertainty covariance is closer to the "truth" than another (aside from demonstrating that one or another was obviously very under- or over-constraining), or, put another way, that one inversion set up would perform better using real world data. The paper describes various metrics of the posterior solution. However, most of these (e.g. RMSE and correlation compared to the known fluxes), simply show that the gradient descent is probably working (i.e. these factors must improve unless there is something obviously wrong with the algorithm). The only metric that might have some ability to demonstrate that the prior uncertainty covariance is appropriate to the real world are the chi-squared tests. However, as the authors note, since this is a synthetic data study, the model is

"perfect", so the model-data mismatch will be much smaller than would be achieved in the real world, making this test uninformative for real-world applications.

The reviewer argues that the metrics of the posterior solution just support the obvious, that the algorithm works properly. We disagree that the solution in the flux space (RMSE and correlation were explicitly stated in the comment) will obviously converge to the synthetic one. The conjugate gradient algorithm optimizes flux scaling parameters by minimizing the model-data mismatch in the concentration space and not in the flux space. Hence, metrics assessing the posterior solution in concentration space, they should indeed improve, and confirm that the algorithm works. However in the flux space quite different flux patterns would lead to almost the same value of the cost function (Kaminski and Heimann 2001). Bayesian inversions set a limit to the flux field, by accounting the a-priori information. Nevertheless, priors do have a Gaussian a-priori probability distribution and they can deviate from their mean value (best guess value). Hence, convergence of the fluxes to the right direction can not be considered as granted. Thorough analysis is needed to explore first, if the fluxes have indeed converged to the "known truth" and secondly, at which spatial/temporal scales can we retrieve the "known truth".

Regarding the second part of the comment, indeed in the current synthetic experiment, transport uncertainties are not included since the same transport model was used for both, the synthetic data creation and the inversion. However we try to keep the experiment as realistic as possible by assuming two totally different biosphere models, to simulate the synthetic observations (BIOME-BGC, process based model), and to provide the prior flux field (VPRM, diagnostic model). Further, we include data gaps in the synthetic mixing ratios in accordance with the gaps appeared in the real observations, for the same time period (see also 2.2.3). The model-data mismatch is calculated for both, the synthetic case and the real data inversion in the companion paper. Whilst the reviewer expects a large difference between those two inversions, this appears not to be the case. Comparing the mixing ratio time-series plots for the Schauinsland station and also the summarized statistics in the Taylor diagrams for both inversions (synthetic and real data), the model-data mismatch is not dramatically different. Posterior mixing ratios from both inversions share same correlations (above 0.9) with the (true respectively pseudo-) observations; furthermore, the normalized standard deviations show a lot of similarity between the pseudo-data inversion and the real-data inversion. We therefore disagree that the current test is uninformative for a real-world application.

2.

Several relevant papers have not been referenced here. Ganesan et al. (2014) tackle essentially the same problem in a hierarchical Bayesian framework. They show that

inclusion of a set of hyper-parameters describing the prior uncertainty covariance necessarily moves the posterior uncertainty closer to the "truth", compared to an inversion without these factors. They were also able to include transport model-data mismatch uncertainties in the inversion. Whilst I don't believe they included a spatial or temporal component in the prior uncertainty covariance, they did explore this in the model-data mismatch, and I don't see why the framework couldn't be extended to do so with the prior (similarly the inclusion of a "bias" hyper-prior would also be possible). In a related approach, Lunt et al. (2016) included the spatial disaggregation of the flux field (and hence, presumably, the level of spatial correlation in the posterior solution) as an unknown in the inversion. Finally, Zammit-Mangion et al. (2015; 2016) present a solution to the flux inverse problem in which only the spatial correlation lengths are used a priori, and the inversion is not constrained to a mean flux field. In summary, I think that these papers demonstrate some significant advances in this area in recent years. Ideally, this article would build on these developments, or demonstrate why the advocated approach is preferable. At the very least, these papers should be cited.

Ganesan et al. (2014) perform a hierarchical Bayesian inversion for a totally different tracer (SF_6). SF_6 flux information from direct observations is not available (no eddy covariance (EC) measurements), hence describing the prior error from comparisons to flux observations is rather impossible. For that they use atmospheric mixing ratio measurements to derive optimized fluxes and hyper-parameters. The latter is introduced as an uncertainty term optimized by the atmospheric data (as well as the fluxes are). With this term Ganesan et al. claim to obtain better results than with a traditional Bayesian inversions that use expert knowledge to determine prior error structure. We note two things: 1. we do not use expert knowledge for the prior error covariance but instead a fully characterized error structure, using an autocorrelation analysis in flux residuals. 2. We are able to perform this analysis simply because flux data is available through the EC measurements, something that Ganesan et al. (2014) can not use since there is no SF_6 flux information. Since we do have spatial and temporal information for CO_2 fluxes, and we can directly quantify the prior error structure we do not see the reason to use mixing ratio measurements to indirectly correct posterior flux estimates by a hyper-parameter which is again optimized from the mixing ratio measurements.

Zammit-Mangion et al. (2015; 2016) and Lunt et al. (2016) studies describe CH_4 inversions. Again there is no EC flux information available. Further, CH_4 fluxes taken from inventories are quite uncertain. For that, they use a spatially invariant prior flux field claiming that the optimization will be predominately data driven. However the covariance function which describes the spatial dependence in the flux field was obtained by a variogram analysis with fluxes derived from different inventories. In our study instead, we look into spatial and temporal autocorrelation patterns of residuals between flux observations (or pseudo observations) and prior fluxes. This is a direct way to fully characterize the prior error structure, as long as the available information exists (i.e EC flux measurements).

We added a reference also to those papers:

P6, L10-15 : "In a similar approach Ganesan et al. (2014) and Lunt et al. (2016), applied a hierarchical Bayesian model using atmospheric concentrations, to estimate both fluxes, and a set of hyper-parameters (e.g. mean and standard deviation of a priori emissions PDF as well as model – measurement standard deviation and autocorrelation scales). In those studies direct flux information for the tracers of interest (sulfur hexafluoride (SF_6) and methane (CH_4)) is not available."

3.

In Figure 7, it appears that, for several months, the derived fluxes are not between the prior and the "truth". I'm not sure how this could be the case, since the pseudo-data should always pull the solution towards the truth, and the prior should pull towards itself. Therefore, shouldn't our expectation value of the posterior fluxes be somewhere in between? Has some random error been added to the pseudo-data (this should be clarified in Section 2.2.3)? If so, is this feature a product of this particular random realisation of the pseudo-dataset? Therefore, do you need to run an ensemble of inversions to "average out" sampling errors?

Contrary to intuition, the posterior expectation value of a synthetic experiment does not necessarily need to be in between the prior and the truth. A possible reason to cause the a-posteriori fluxes to fall outside this bracket are a-priori flux correlations. For illustration, let us consider 2 example pixels with a mutual distance within the spatial correlation radius. Assume that pixel 1 is well constrained by the atmospheric data, and that its true flux is smaller than the prior.

Consequently, the flux correction at such a constrained pixel will have a negative sign. In contrast, assume pixel 2 to have a true flux larger than the prior, and to be poorly constrained. Due to the weak data influence (the network sensitivity is uneven), the flux correction at pixel 2 will mainly follow the constrained pixel 1 via the spatial correlation and be negative as well, even if that brings the a-posteriori flux further away from the true flux.

This scenario is consistent with the behavior of the S1 case relative to the B1 case. In S1, a bias term has been assumed, which simultaneously shifts the flux field at all pixels, whether they are well or poorly constrained. This introduces an additional spatial correlation, possibly causing the S1 fluxes to be outside prior and truth more frequently than in the B1 case.

We note that the pseudo-data does not contain any kind of random error realization, therefore an ensemble of inversions is not required. We clarify that also in the paper:

P15, L20: "…We note that the synthetic data were derived without adding error realizations."

Specific comments:

P4, L31: I don't see why model errors will be more easy to define that prior uncertainties? I don't think we have a very good handle on transport model error. Furthermore, this term does not need to be diagonal, as this sentence implies (see references above).

We clarify that uncertainties in the measurements may be easier to quantify. We added:

P4, L31: "While the measurement uncertainty in the observational constraint is usually defined with the main diagonal of the covariance matrix representing the uncertainty of the observations and the model at a specific time and location, our knowledge for the prior uncertainty is limited, especially regarding temporal and spatial correlations that effectively control the state space."

We agree with the reviewer that the measurement error covariance matrix (includes measurement and transport uncertainties) does not have to be strictly diagonal as correlations are probably present. Whilst we do not explicitly introduce off-diagonal terms in the measurement error covariance matrix, the Jena Carbonscope system implicitly assumes that correlations exist. In fact the system contains the so called "density weighting function" (Rödenbeck, C., 2005). The role of this weighting is to combine flask (~weekly) and continuous (hourly) data with a consistent way, as otherwise the high frequency data would lead to a stronger impact at those particular sites. To avoid that, the density weighting inflates the uncertainty by the square root of the number of the observations at weekly basis. This density weighting plays one more role. It implicitly takes into account correlations in transport uncertainties which might be present. More information can be found in Rödenbeck, C., 2005.

P16, L14 we added: "…transport error correlations might be present. Although we do not explicitly introduce off-diagonal terms in the measurement error covariance matrix, we do consider for temporal correlations via a data density weighting function which inflates the uncertainty. (see Section 2.1 and more information in Rödenbeck, C., 2005)."

P6, L30: See references above.

This is not described in the references mentioned by the referee, as those publications do not make use of flux observations to constrain the a priori flux error structure.

P9, L9: Why limit this matrix to being diagonal? As noted on L 13, the transport model will certainly exhibit temporal and spatial uncertainty correlations.

As with the comment above, we agree with the reviewer that the transport model will exhibit temporal and spatial error correlations. We have also explained above, the role of the density weighting. The reviewer might have some concerns regarding the concentration mismatch uncertainty, that without considering the correlations we might have underestimate it. At this point, we would like to refer to the reduced chi square values at site level (eq. 11). This metric by definition (posterior mismatch over assumed uncertainties) assures us that the assumed uncertainties were rather conservative (values smaller than 1). Hence we do not believe that the measurement error covariance is mishandled.

P25, L14 we added: "In the current study we assumed a diagonal measurement error covariance matrix. Concerns might rise that the observational uncertainties are underestimated due to the absence of the error correlations. However the χ_r^2 values prove the opposite."

P11, L5: This equation is not referenced explicitly in the text. What does it show?

The equation refers to the deviation term of the flux model. We clarify that by adding:

P11, L23: "the flux model (Eq.6)…."

P11, L6 – L12: These terms are discussed before being introduced (they refer to an equation in the following subsection). I think the order needs to be changed here.

We corrected by deleting the text. The text is moved to P14, L15.

P11, L19: If I understand this correctly, synthetic eddy covariance (EC) data were extracted at several locations in both models, and these pseudo-fluxes were used to calculate the spatial and temporal correlation lengths for use in the inversion (please clarify that this is synthetic EC). So essentially, we are using the difference between two models as a proxy for the uncertainty correlation in the real world? I think this is fine. However, two things come to mind: 1) if we were to use "real" eddy covariance data, we would sample very much smaller length scales than the model (i.e. typically <1km, rather than 50km), so I would not expect that the derived correlations would be comparable to the same experiment using real data (as the text seems to indicate on P6); 2) since we're in model world, and in light of point (1), why not use every grid cell to calibrate the correlations? Would this come out as being very different?

We clarify that the eddy covariance data is synthetic. We added:

P12, L15 :"Fluxes from GBIOME-BGCv1 can also be regarded as synthetic EC fluxes."

We note that the current paper describes the synthetic experiment. The aim is to perform the real data inversion (see companion paper Kountouris et al., 2016) using data driven uncertainties and for that we need a methodology. The methodology is the same for both papers based on Kountouris et al. (2015). By no means, we do want to imply that we use model-model differences as a proxy for the uncertainty in the real inversions. To evaluate correctly the system we need to apply though the same methodology but to different data sets. For the synthetic case the correct error structure will be derived by estimating the error correlations between the models which took part in the inversion. In the real world though, we perform a model-data analysis since this would be the appropriate for the real error structure. Indeed as the reviewer expected, we calculated spatial correlations significantly smaller than those derived from the model-model residual analysis. The text in P6 explicitly refers to that finding (L7). We disagree with the reviewer that P6 indicates the opposite.

Regarding the second concern the reviewer makes a good point. We have tested also the spatial correlations by calculating all the potential pixel pairs. No significant difference was found. But even if a difference was present, we selected to extract modeled fluxes only at the same locations where an EC station exist, for the sake of comparability to real data inversions.

In the companion paper we perform the same flux residual analysis for real EC data. The spatially resolved flux distribution is known only at the EC measurement sites. By selecting the same grid-cells in the synthetic case we make sure that we do not add additional information in the error structure, information that we do not have in the real world. With this approach we make sure that the synthetic case is not over constrained and hence, it is a fair experiment comparable with a real data inversion.

P12 L24 we added: "Following this approach apart from the similarity, we also ensure that results from the synthetic experiment, would be informative for a real data inversion, by using exactly the same information to characterize the prior uncertainties."

P12, L12: The two experiments that are carried out focus on "tuning" the covariance matrix in two ways, so as to match the overall difference between the two models: B1, scale the covariance matrix uniformly; S1 add a bias. What is the reasoning for choosing only these two methods? Couldn't this mismatch be closed in several other ways, e.g. by increasing the correlation lengths or adding a "nugget" term to the diagonal elements, etc.?

Methods like increasing the correlation lengths or adding a nugget term to the diagonal elements are already used by Lauvaux et al. (2012) and cited P6 L21. In Kountouris et al. (2015) the analysis leaves no room to assume much larger correlation scales. Indeed we could have chosen larger correlations to increase the aggregated uncertainty but at expense of the validity of the error structure.

P14, L4-L8: Please provide a reference for these choices of data filtering.

P16, L5 We added : "…Geels et al. (2007)". The citation is also added in P29 L27.

P15, L13: I don't think Thompson et al., 2011 is the most appropriate reference here.

P19, L23 we corrected: "Tarantola 2005". The Thompson citation was deleted and instead we used the following citation: "Tarantola, A.: Inverse problem theory and methods for model parameter estimation, ISBN: 0-89871-572-5, siam, 2005."

P16, L17 – L22: The improved correlation and "variance" is simply a product of the cost function descent. This should be clarified.

Certainly the improved statistics are directly related to the minimization of the cost function, as the Bayesian inversion balances between data constraint and prior constraint.

We have added the following in P18, L25 to remind the reader of this: "As expected from the optimization (i.e. minimization of the cost function), the…"

P16, L23: Does "chi-squared" show us anything here that we can extend to the real world, given that the model is perfect (see general point 1 above)?

The chi-squared values did not intend here to show anything regarding a real data inversion. This is just a metric to evaluate how well the algorithm fits the observational (dry mole fractions) data, and also to evaluate our prior error assumptions. This is a rather important step before proceeding with a real data inversion. The fact that there are no transport uncertainties does not make the whole model perfect. We refer here to the answers to the first comment.

P17, L7: Again, isn't this a trivial result showing that the gradient descent is working?

As explained in the very first comment, the gradient algorithm minimizes the model-data mismatch (concentration space), and not the fluxes themselves. We refer here to the answers to the first comment.

P18, L11: Probably should be noted that this will largely be determined by the model-measurement mismatch uncertainty covariance, rather than the prior uncertainty.

The uncertainty reduction in flux space is defined as [1 – (posterior flux uncertainty)/(prior flux uncertainty)]. The posterior uncertainty depends on the prior, the measurement and the transport uncertainty. However it is not obvious that the uncertainty reduction is largely determined by the measurement error covariance. Hence we would like to avoid this statement. Instead we clarify in the paper the dependence of the posterior uncertainties. We added:

P21, L26: "…and B1 respectively. Note that whilst the prior uncertainty refers only to the flux space, the posterior uncertainty depends on the uncertainty of prior fluxes, measurements, and atmospheric transport".

P19, L15: I think this is a very strong conclusion to draw here. I'd contend that the suitability of EC data for "validation" of inverse model fluxes is dominated by scaling issues. In this paper, it is assumed that the EC data is representative of 50km^2. In reality, EC data will sample scales that are orders of magnitude smaller.

We agree with the reviewer that EC data is representative for much smaller scales on the order of ~1 km^2. We note that the retrieved fluxes are at ~25 km resolution not 50 km. Nevertheless, still the scales are not directly comparable. However this method is also used in Broquet et al. (2013) where posterior flux estimates were compared against EC data. Despite the scale mismatch they found that posterior mismatches are in good agreement with the theoretical uncertainties.

We corrected in text the word clearly and now read:

P23, L1: "…potentially shows…"

P20, L1: I think this shows that your inversion algorithm is working, not that you would get any closer to the truth in the real world.

We believe this is answered in the very first comment made by the reviewer.

P21, L13: See general point 1.

We refer to our response for the general point 1 and 2.

P22, L11: I don't think we can comment on the reliability of the results of a real world inversion here. A real world inversion will likely be dominated by chemical transport model errors, which are not quantified here.

With respect to the error assumptions and whether we underestimated the uncertainties in the measurement error covariance, we refer to the specific comment P4, L31 from referee 3. We note that we follow a standard approach for the characterization of the transport model uncertainties and more information can be found in Rödenbeck (2005). Inversions are typically adding the transport error in the measurement error covariance matrix as a diagonal element. Off diagonal elements usually are not considered, since there is no direct method to fully characterize (spatial and temporal autocorrelation lengths) the transport error. Potentially, one could have a rough estimation of the transport uncertainty, by running a number of different transport models, and comparing the simulated atmospheric mole fractions with observations or by calculating the range of the posterior flux estimates. Since all metrics show a very good fit to the atmospheric but also to the flux data, we have no reason to reject those results. Regarding the real data inversions, if transport uncertainties are dominating, we would expect also a bad data fitting. In that case we would agree that probably the posterior flux information would not be informative at finer scales, where complex atmospheric transport patterns can not be fully captured by the atmospheric transport models (e.g. mesoscale circulations etc). However, at coarser aggregated scales fluxes would still be informative as long as the data show a good fitting performance.

P23 L17 we added: “In the current study we do not excessively assess the transport error but it is rather included as diagonal elements in the measurement error covariance, which is typical in atmospheric inversions. The chi square values confirm that there is no underestimation of the uncertainties. We note though that erroneous flux estimates are likely to be estimated, especially at finer spatial scales where the transport model is not able to resolve the real transport (e.g. individual eddys, complicate terrain etc). However, for coarser spatial scales transport models perform better, and as long as the fitting performance shows good results, flux estimates should be more reliable.”

P26 L7 we added: “…at least for aggregated scales up to the country level”

References

Broquet, G., Chevallier, F., Bréon, F. M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, Atmos. Chem. Phys., 13, 9039-9056, doi:10.5194/acp-13-9039-2013, 2013.

Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., Knutt, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, American Meteorological Society, 27, doi:10.1175/JCLI-D-12-00579.1, 2014.

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U., Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R.: Comparing atmospheric transport models for future regional inversions over Europe - Part 1: mapping the atmospheric CO2 signals, Atmos. Chem. Phys., 7, 3461–3479, doi:10.5194/acp-7-3461-2007, 2007.

Jacobson, M.Z.: Correction to "Control of fossil-fuel particulate black carbon and organic matter, possibly the most effective method of slowing global warming.". J. Geophys. Res. 110, D14105, doi:10.1029/2005JD005888, 2005.

Johnston, N.A.C., Blake, D.R., Rowland, F.S., Elliott, S., Lackner, K.S., Ziock, H.J., Dubey, M.K., Hanson, H.P., Barr, S.: Chemical transport modeling of potential atmospheric CO_2 sinks, Energy Conversion and Management, 44, 681–689, doi: 10.1016/S0196-8904(02)00078-X, 2003.

Kaminski, T. and Heimann, M.: Inverse modeling of atmospheric carbon dioxide fluxes, Science, 294, 5541, doi: 10.1126/science.294.5541.259a, 2001.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, F. Th., Heimann, M.: Atmospheric CO2 inversions at the mesoscale using data driven prior uncertainties. Part2: the European terrestrial CO2 fluxes, Atmos. Chem. Phys., acp-2016-578, submitted on 04 Jul 2016, 2016b.

Rodgers, C., D. Inverse methods for Atmosphere Sounding: Theory and Practice, World Sci.,

River Edge, N. J., 2000.

Rödenbeck, C.: Estimating CO2 sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport, Technical Report 6, Max Planck Institute for Biogeochemistry, Jena, http://www.bgc-jena.mpg.de/mpg/websiteBiogeochemie/ Publikationen/Technical Reports/tech report6.pdf, 2005.

Tolk, L. F., Dolman, A. J., Meesters, A. G. C. A. and Peters, W.: A comparison of different inverse carbon flux estimation approaches for application on a regional domain, Atmos. Chem. Phys., 11, 10349-10365, doi: 10.5194/acp-11-10349-2011, 2011.