Atmos. Chem. Phys. Discuss., 11, C5084–C5099, 2011
www.atmos-chem-phys-discuss.net/11/C5084/2011/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "A comparison of different inverse carbon flux estimation approaches for application on a regional domain" *by* L. F. Tolk et al.

L. F. Tolk et al.

lieselotte.tolk@falw.vu.nl

Received and published: 17 June 2011

Reply to referee #2

We thank the reviewer for his comments and the constructive review. The questions and remarks raised by the referee are answered below, and the suggestions are included in a new version of the paper.

"- The approaches you use include very different assumptions which are key elements of the system (error correlations, number of unknowns, ...). You explore here all these aspects at the same time, and without really identifying the true causes. For example, when you assume one correction factor per vegetation type (case 1), you assume

that one factor can be optimized to improve the fluxes in each region. It includes large error correlations within the regions and few unknowns over the domain, leading to a reduced flexibility and a prescribed diurnal cycle. Then, you split this factor for respiration and GPP (case 2), increasing the degree of freedom, and changing the diurnal cycle definition."

In the current paper four often used inversion approaches are integrally compared. The aim of this setup of this study is to explore the importance of the inversion setup for the regional scale. Since all the inversions employed in this study are regular methods used in previous studies, they all seem justified to use at the regional scale. The scope of this study is to compare a range of often used inversion methods, rather than separately investigating the details within a single inversion study. This indeed means that different aspects vary at the same time between two inversions setups. We actually see this as a strength of the set-up of the study, since these variations are realistic in the choice between the different approaches. Instead of separately testing the different aspects of the inversion setup we compared over the broad range of integral methods, which aims to provide insight in the realistically used methods at the regional scale. Of course many more details can be explored, some of which would require separated studies. In this response we will extract the details which are included in the current study but have been underexposed so far. We thank the reviewer for highlighting these.

"You have to clarify and organize your results."

In the results we show the performance at all important metrics to provide a full overview of the strong and weak points of the different methods. We decided to maintain this structure, since it gives the opportunity to the reader to look into the behavior of the different inversions. For the additional questions raised in this review, we added a new part to the results section (the new paragraph 3.6).

"What comes from the diurnal cycle? What is the impact of the increase of unknowns? Could you define shorter periods of time in NEE to fit better the diurnal cycle? Is

ACPD

11, C5084–C5099, 2011

the choice of 15 days relevant? What is the aggregation error in time due to this assumption?"

The impact of the different assumptions in the different cases, including the number of unknowns and the number of degrees of freedom, is described in the new paragraph 3.6. To better fit the diurnal cycle by using shorter periods of time, these periods of time should become smaller than one day. For NEE the nocturnal and daily values should be split apart to better fit the diurnal cycle, which resembles the RG inversion. The choice of 15 days is relevant in the sense that it includes both anti-cyclonic and cyclonic weather conditions. The difference in the response of the fluxes to different meteorological circumstances is, next to the diurnal cycle, an important source of the structure mismatch between the fluxes of the truth and the prior. The aggregation error in time during the 15 day period is thus a combination of the structure mismatch due to differences in the diurnal cycle and due to a difference in the response to different weather circumstances. This is addressed in section 3.6.

"- In your conclusions, you recommend the use of case 3 and case 4 compared to the other solutions. But results in figure 4 show large differences between the different methods and no real improvement in either case. One can notice large biases for crops using two factors in case 3 (pixel based inversion with factors for respiration and GPP), and very little changes everywhere in case 4 and case 1."

In this study we addressed several different metrics, next to the posterior fluxes, to evaluate the performance of the different options based on which we conclude that we recommend case 3 (pixel inversion) and 4 (parameter inversion). This is stated more explicitly in the updated Conclusions ("smallest deviations ... over all metrics"). We agree, however, that the (dis)agreement for the regional fluxes is an important criterion. Hence, the phrase "best results" is removed, and in particular, it has been added that "also these inversions had difficulties in estimating the specific fluxes per ecoregion".

"One can wonder if your results are limited by the vegetation type homogeneity (ecore-

ACPD

11, C5084–C5099, 2011

gions) assumption, or by the lack of data in critical areas (over the crops), or by a lack of flexibility in case 1 and more corrections but no better results in the other cases. Does it mean that splitting in two factors is better? It is a change of assumptions but it is hard to conclude clearly looking at your results."

The results are limited both by the temporal and the spatial structure, and thus the heterogeneity, of the truth compared to the priors, which can for example clearly be seen by an comparison with the control inversions, which returned much better posterior results. This is stated in the results (end of each section), the discussion (first and second paragraph) and the conclusions (previously first, now third point). The amount of data is not the most limiting factor. In the relatively small domain of 400x400km 4 observation towers are available, which is a relatively dense network. One of the towers (the eastern tower) is located in an area with crops2 and the middle and eastern tower are located close to and downwind of crop1 areas. The posterior results of both crop species encompass the localization, and both the posterior mean and the uncertainty are altered by the inversions. The only ecoregion where the results are limited by the lack of observations is deciduous broad leaf forest, which in βNEE and βRG inversions does not encompass the localization. We showed that the standard methods of increasing the flexibility, by splitting in βrespiration and βGPP instead of only βNEE, and by using a pixel inversion instead of ecoregion approach, does improve the results but still have to be handled with caution.

"Then you perform one case on a pixel based system (case 3), increasing your degree of freedom and changing the underlying assumptions (less flux error correlations or more distributed in space). One more time, this is a change in the assumptions more than a change of "approach". You have to clarify all these elements."

These elements are clarified in the method section (2.1), where we introduce the pixel based inversion in equation 3. The flux error correlation, with a length scale of 100km within one ecoregion is also described in this section. At the end of this section is referred to Table 1 in which degrees of freedom of the different options used in this

ACPD

11, C5084–C5099, 2011

study are shown. The use of the word “assumptions” or “approach” is a semantic question, were we feel that approach is an applicable word to use, but we added in de introduction (second last paragraph) that the approaches all have different underlying assumptions.

“Figure 2 shows that lots of noise is introduced in the results without any structure with the pixel based system. Is it due to your flux error correlations? Do you have enough members to capture relevant signals in the EnKF? “

The strong noise on a relatively small scale is indeed a consequence of the choice of a small correlation length. This choice is considered realistic as it suits the very fine structure of the land use for this area. The correlation between neighboring pixels is easily visible in the pictures, but on a somewhat longer scale the results have indeed little structure. Concerning the number of members, we followed the conclusions by Zupanski et al., 2007, who performed inversions with a comparable number of unknowns as in the current study. Further it follows from basic principles that the small-scale noise is a consequence of the constraining of many unknowns with few effectively-independent constraints. This problem cannot be remedied by enhancing the ensemble size (though this will cause a slight smoothing), but only by aggregating the fluxes to larger units. This is confirmed by experiments of authors such as Carouge et al. (2010), who use a Lagrangian approach with a large ensemble. Unfortunately, the possibilities for systematic study of the influence of the number of members are limited, as we use an Eulerian approach, in which each new member requires a new 3D field. We were not able to perform runs for an ensemble size much larger than 100 (150 did not work). But, again, an ensemble of say the double size cannot be expected to yield results that are really better on the small scale.

“You should perform a third inversionas case 1 but at the pixel scale to really explore what is the impact of the correlations, and what is the impact of the ecoregion assumption. Case 2 is not really better in general.“

ACPD

11, C5084–C5099, 2011

We agree with the reviewer that such an inversion could potentially tease apart some of the differences between the methods as caused by differences in correlations and differences in stratification of unknowns. However, we also feel that a pixel scale NEE inversion is not a very likely approach for anyone to adopt in an actual regional application. This is because NEE is heterogeneous and may change sign over the landscape. Scaling factors for these fluxes can change sign leading to erroneous (flipped) diurnal cycles, as is shown for example in the behaviour of the βNEE results for crops in the control inversion, in some or many gridboxes which is difficult to deal with physically and numerically. Because this case is thus unlikely to be used, we decided not to pursue this suggestion.

"You could also test different correlation lengths or structures to identify their importance."

We note that the impact of correlation lengths and structures has been studied in depth by various authors (Michalak et al., 2005; Peters et al., 2005; Carouge et al., 2008; Chevallier et al., 2006, to name just a few).

"- Similarly, Figure 2 leads to several critical issues and questions. Basically, your case 1 shows no significant change, suggesting too few unknowns and no flexibility. Then case 2 generates very positive fluxes over grass, and very negative fluxes over crops, both far from the truth (indicated in fig 4 as well). Case 3 presents very noisy results, with almost no structure in space. It seems clear that flux error correlations are critical for the different cases. Are you sure this has anything to do with the approach? I would suggest that this is a direct consequence of the flux error correlation assumption, the temporal behaviour you assume (diurnal cycle), and the number of unknowns."

The reviewer is absolutely right in his last sentence: the behavior of each method is a function of number of unknowns, their assumed relation to the fluxes, and the correlations between the unknowns. All these aspects vary when one switches from one approach to another. Of course, if one selects one approach (say grid based optimiza-

ACPD

11, C5084–C5099, 2011

tion of respiration and photosynthetic scaling factors) there is again the possibility to vary all these aspects and one could conceivable pull apart some of their individual influences with a set of well designed simulations. However, those results would only be applicable for that one method, and under the assumed setup. As stated before this is not what we intend to do in this paper. Rather, we describe the results of a number of popular methods under similar assumptions as made in current real world applications.

“Finally,the case 4 shows little changes compared to the prior and no real structure in the posterior compared to the truth. Figure 2 also shows clearly that this method is not far from the prior.“

We do not agree with this remark. The posteriors in case 4 are clearly altered compared to the priors, for example the posterior grass fluxes are only a third of the prior ones, the carbon uptake by crops2 is more than doubled and in all ecoregions the posterior uncertainty is strongly reduced (see figure 4). The effect of the non-linearity in case 4 on the changes in the posteriors and the reduction of the uncertainty is discussed in section 3.1.

“Defining which "approach" is best is overstated, mainly based on a statistical criteria, the RMS from table 3, even if all cases give biased estimates for the main regions (crops being a large component in terms of annual NEE).“

We agree that all inversions have difficulties in optimizing the fluxes for the individual ecoregions in the inversions where the structure of the prior deviates from the structure of the truth. This is an important message of the paper, as stressed for example in the abstract (third paragraph). Based on this defining the best approach may indeed be overstated, therefore we rephrased this more modestly in the conclusions and the abstract.

“Finally, in figure 5, it seems clear that for the RG 1.0 case, the nighttime fluxes are unrealistic, with a negative value almost constant. Why the other cases give more reasonable estimates? Is it only due to the perfect correlation between resp and GPP?”

ACPD

11, C5084–C5099, 2011

First, we found that the legend contained a type-error, instead of RG1.0 it should note RG 0.0. This is corrected in the new version and we apology for this flaw. The reason for the bad performance of the RG inversion to capture the nocturnal fluxes is related to the difference between the temporal structure of the fluxes in the truth and the priors. This is added to the new paragraph 3.6.

"- You have to identify more clearly which assumptions are critical (flux error correlations, number of unknowns, flexibility of the system,...)."

The difference in assumptions between the different inversions, and the impact on the results is further analyzed and described in the new section 3.6. The fact that increasing the temporal flexibility in the parameter inversion and the spatial flexibility in the pixel inversion does not fully solve the problems faced in the inversions suggests that both temporal and spatial structure differences between the truth and the prior were limiting to obtain results at a smaller scales than a few hundred kilometer.

"What signals can be retrieved with your current observation network?"

This is of course an interesting question and as the reviewer likely knows, difficult to answer. Based on our results we can say that in the absence of any errors (the control case), we can retrieve the two-week average integrated flux well, the individual ecoregion fluxes too, and even get decent temporal behavior with all methods. Pixel based fluxes only make sense after post-aggregation as in other studies. Unique to our results though is that as soon as we include a realistic difference in the structure of the underlying flux model, all methods struggle to perform comparably well. None of them excels at all metrics, and only two seem to be a promising way forward. So what does that mean for the current observation network and its retrievable signals? We believe the answer is that the current observation network is only one of the limiting factors in this study, and likely not the dominant one.

"Are the fluxes strong enough that your inversion can detect them? May-June is the very beginning of the growing season, with 5-10 umol.m-2.s-1 for crops."

ACPD

11, C5084–C5099, 2011

Interactive Comment

During 2008, net daytime CO2 fluxes were regularly measured by aircraft over the Netherlands (6 trajectories; to be published). For the last 10 days of May an average net flux of -6.3 μmol m-2 s-1 was found (average over the daytime hours for which measurements were done). This is already close to the mean for June-July-August which was -7.7 μmol m-2 s-1, according to the same set of measurements. So the fluxes for the interval of the present paper, though small, are already approaching the summer average. That the fluxes are (in principle) detectable by the inversion system, in spite of their weakness, is shown by the results of the control run.

"What is the impact of a 15-day time period? Is it too long and might lead to biased estimates? Many aspects of the system should be explored before you conclude about methods, results, and potentials of such a system."

We'd like to refer to the answer on the third question of the reviewer; also see the new section 3.6.

"And finally, your discussion section is out of the scope of the paper. How can you argue about longer term issues, correct carbon flux balance at regional scale or not, and aggregation errors, with your pseudo data test over 15 days?"

The last-but-one paragraph of the discussion section alluded to data aquisition to improve the modeling of the change of biotic parameters on the long term. This paragraph has been deleted in the new version. Otherwise, the content of the discussion section is restricted to what can be inferred from the present results. We do not make statements about the accuracy of long term budgets etc. but only on means for the considered time scale. The use of only 15 days of concentration measurements is not limiting as this is many times the Lagrangian time scale of the considered area.

"You also discuss non linearity without showing consistent results with this statement."

We agree that we discussed the effect of nonlinearity on our results without showing figures that we have used to come to our conclusions. This is because we came to our

ACPD

11, C5084–C5099, 2011

insights through a long period of trial-and-error with our nonlinear model to finally arrive at the strategy we describe in the paper. One of the tools that we used to investigate failed inversions was a simple inversion code where we could optimize three unknowns as part of either a linear, or a nonlinear forecast model. In the new version, we add a result from this toy model as appendix C, to convince the reviewer that we discuss nonlinearity based on actual investigations and not just on general principle. We have included a figure and some discussion in an appendix, so as not to distract too much from the main text. "This study is based on different potential approaches that one may use to perform a regional scale inversion. You have to seperate clearly what is the impact of each component on the final results. A more rigorous analysis is required to identify the importance of the diurnal cycle, the spatial distribution, the number of unknowns without splitting in different factors simultaneously, the importance of the flux strength (very weak during your period), the correlation structures in the flux errors (length scale value), and the prescription of the diurnal cycle."

In addition to the analysis already done in the previous version of the paper, we added in section 3.6 an extra description of the impact of each component and the impact of the different components on the final results, the impact of the diurnal cycle, the spatial distribution and the d.o.f. are discussed here. However, much of these questions are simply beyond the scope of the paper. The fluxes are not particularly weak for the domain under scope (see the answer to the reviewers previous question "Are the fluxes strong enough that your inversion can detect them?").

"Technical comments: -introduction: your bibliography doesn't include any of the regional scale inverse studies already published. Some examples are Gerbig et al., 2003, Lauvaux et al., 2009, Schuh et al., 2009, Goekede et al, 2010, Gourdji et al., 2010)"

We apologize for not including some of these papers in the introduction. We note that we have drawn significantly (and referenced) the papers from Gourdji et al., Schuh et al. Lauvaux et al. 2009 was indeed forgotten in the list of references (though cited on

ACPD

11, C5084–C5099, 2011

page 3373 line 21). This has been fixed in the new version. A reference to Göckede has been added in the new version. “3358-1: "recently" is inappropriate, or explain why the 3 studies you cite are different” compared to past inversions. We removed “recently” from the text. We intended to indicate that in the last decade many inversion studies are performed, not specifically just these three examples.

“13: missing "has"”

“has” is included in the text.

“3360-9: refer to appendix A”

We included a reference to appendix A in the text.

“3361-12: your length scale seems much smaller compared to the different studies you cite. Can you explain why?”

The cited studies employ a coarser inversion set-up, with coarse gridboxes and a larger domain over which the inversion is performed. In the regional scale inversion in the current study a smaller length scale is used that fits with the smaller inversion setup, and the higher level of detail in this study. When expressed in gridbox size, instead of in kilometer, the length scales are in the same order of magnitude. This difference is added to the text. “Why don’t you diagnose the flux error correlations from your prior-truth residuals?”

Based on prior-truth residuals (synthetic truth of course; all day and night hours together), the error correlation length would be about 130 km, only slightly more than the 100 km assumed for the present modeling work (see figure 2 in this response). This is added to the text after equation A10.

“3363-1: what crops are included in "crops1"?”

As explained in section 2.2 crops1 refers to ‘Agricultural areas with complex cultivation patterns’ and crops2 refers to ‘Agricultural land with significant areas of natural vege-

ACPD

11, C5084–C5099, 2011

tation'. These are the two most important crops classes in the Netherlands according to the Corina database.

"3365-1-2: the prior error covariance can't be equal for all cases, only the overall error."

We already stressed this point in 3365-12-13 were we stated that "equal variance of the time/ecoregion integrated NEE does not ensure the same uncertainty in the inversions at each point in space and time". To emphasize this even more we added 'overall' to 3365-1.

"3.6: What is the purpose of this paragraph? You Want to show that early June is not very different in terms of fluxes compared to late May? This metric has no clear value as written here. You could run a second period (early June) of inversion and propagate the information from the first one to see if it helps. Otherwise, you just detect if the two periods are similar or not."

We agree that this paragraph might seem useless because the metric presented does not discriminate between the different methods. We previously decided to include it because the predictive power of methods that are based on underlying biosphere characteristics rather than just bias correction factors is often used as an argument to develop and improve such methods. But indeed due to the similarity of the two periods of the month May we did not identify this possible advantage. Extending towards a period further in the growing season might however reduce the validity of the parameters. Therefore, we decided to remove this paragraph and replace it with the new paragraph in the result section.

"-Discussion: 3373-21: "non-synthetic" = real"

We replaced 'non-synthetic' with 'real'.

"3374-12-14: only true if you include error correlation between periods, or propagate the optimized factors. But is it a good solution? Does it really impact your system? You can't answer this point with the actual analysis. This is an assumption."

ACPD

11, C5084–C5099, 2011

Interactive Comment

The reviewer is right that the propagation of errors can only occur through propagation of the mean or the covariance of the solution. This is generally done though in carbon cycle studies because atmospheric CO2 inherently constrains fluxes over larger spatiotemporal domains. Therefore, the problem of propagating erroneous parameter estimates is real, and has been experienced by many inverse modelers that focus on periods beyond the two weeks in this study. This includes co-authors Wouter Peters and Antoon Meesters, and we know from personal communication that others (L. Bruhwiler, D. Baker, A. Schuh, S. Houweling, R Lokupitiya to name a few) have also wrestled with this. The statement thus does not feel out of place to us, and is not merely an assumption.

"3374-15-22: This paragraph is out of the scope of your paper. You suggest that a highly sophisticated system able to optimize model parameter by ingesting several types of data and information in addition of co2 concentrations will perform better. This is a very general comment, not related to this study."

In this paragraph of the discussion section we provided a broader reflection on one of the approaches tested in this study. However, we can agree with reviewers view that it is a too general comment. Therefore, we removed this paragraph from the text.

"3374-23-26: other papers have compared their results to other methods and found reasonable results using real data compared to bottom-up approaches (e.g Goekede et al., 2010). Could you discuss or explain the differences in your case? Several pseudo data experiment have found that one can retrieve reasonable patterns and signals (Gourdji et al., 2010). They tested different time windows for the fluxes, and found large differences between one value, 4 time windows per day, and every 3 hours. You conclude with only four cases that the method is not suitable as it is now. Why your conclusions are different from other studies?"

Gourdji et al. 2010 performed inversions for whole North America, and the retrieved details are of a size which is comparable about to the size of our whole domain. It is

ACPD

11, C5084–C5099, 2011

well known that inversions have more “ambiguity” problems to the extent that the size of the investigated details becomes smaller. Further, may the work for North America also have profited from the coarser land use scale for this continent and the larger number of observation sites, but probably this is of less importance.

“3375-1-3: The two sentences are beyond your analysis. Your study is limited to 4 cases and 15 days of pseudo data. Do you think this is enough to conclude about the needs for future inversions?”

We have changed “will not be” to “might not be” as to stress the speculative nature of this statement. We do however think that our analysis supports questioning the applicability of the methods presented on the smaller scales of the future. The end of the discussion seems to us also the appropriate place to do so.

“-Conclusions: See above the general comments.”

The conclusions are adapted to the general comments above.

Figure captions:

Figure 1, The effect of non-linearity on the results of the inversion.

Figure 2, Prior-truth-residuals : Correlation between values at two points at the same time, as a function of the distance. Stars belong to distances 120, 130, 140 km, the line indicates a value of 1/e.

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 3355, 2011.

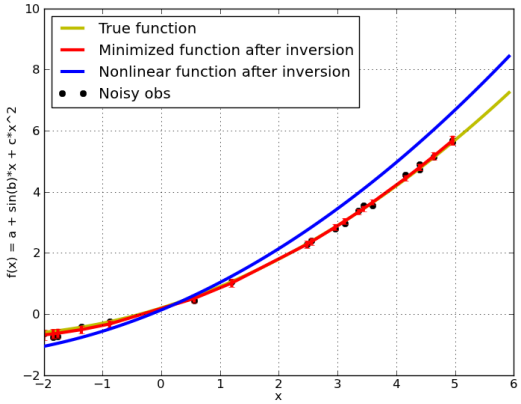

Fig. 1. The effect of non-linearity on the results of the inversion.

ACPD
11, C5084–C5099, 2011

Interactive Comment

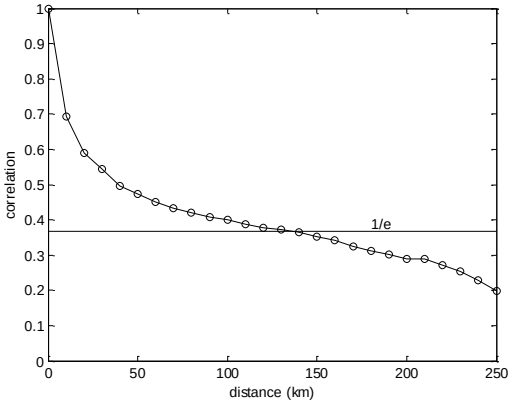

Fig. 2. Prior-truth-residuals : Correlation between values at two points at the same time, as a function of the distance. Stars belong to distances 120, 130, 140 km, the line indicates a value of 1/e.

ACPD

11, C5084–C5099, 2011