Author response to comment on “An inversion approach for determining production depth and temperature sensitivity of soil respiration” by R. N. C. Latimer and D. A. Risk
By anonymous (Referee 1)

(referee comments in black, author responses in blue)

The manuscript presents a synthetic data study on how model inversion could help determine soil CO2 production parameters (in particular, temperature sensitivity and production depth) from field measurements avoiding error sources of existing approaches, and how field measurement configurations could be optimized to serve best such an inversion effort. The following general comments are numbered according to the BGD review criteria list, their explanations are given in the detailed comments below.

This is (1) a highly relevant question within the scope of BG, and (2) fairly original (meaning not completely novel but an important step towards a model that is simple enough to be used operationally across groups). The conclusions are (3) worth publishing, but with respect to the simplicity of the model at least one of them (the comparatively low usefulness the authors found for surface flux measurements) might be somewhat premature (5). The scientific methods are (4) sound, but not everywhere outlined clearly enough, which should be changed to ensure reproducibility (6), this applies especially to three subsections of section 2 (13). The authors make clear their own contribution (7) but related previous work could in places (e.g. inversion, temperature sensitivity determination problem) be referenced more comprehensively (14). The title very well reflects the contents (8), although the authors might want to add some scope towards what appears to be their future study plans (validation/sensitivity study with synthetic data vs. application to field data). The abstract is complete and correct (9), except for a clarity issue with respect to the possible combination of surface flux and in-soil concentration measurements. The language is very fluent (11) and most of the presentation is clear (10), except for a part of the model set-up description (13). There is one missing symbol explanation in an Equation (12) and no supplementary material (15).

Thank you. As these are all addressed in the detailed comments, we respond to them one at a time.

Detailed comments:

P10138L06: maybe insert something like "and/or in-soil" concentrations" (to keep readers from mistaking the "surface" for relating to both the fluxes and the concentrations). Concerning the "and/or", see also comment on P10146 / Table 2.

Done, as suggested.

P10139L05: maybe insert 'mainly' (think of macrobiota) ? P10139LL27: maybe replace "normally" by "frequently" or "often"

Done, as suggested.

P10140L12: ‘Though not done to date…’: This has been done at least once (Bauer et al. 2012, Biogeochemistry 108:119–134).

We have added that reference and changed the text. Interesting paper – thank you! That study is different from ours in that the authors aimed to undertake a thorough exploration of the model sensitivities. Our manuscript of course aims primarily at methodological issues, which will allow us to
seek our the best suited field datasets. We have added not only the reference, but the following text which clarifies the different role of these studies: “To our awareness, cite(bauer) are the only others to attempt a soil respiration inversion, though the authors used only field data for their study, and did not start with synthetic tests to explore the accuracy of their approach, or its sensitivities to the available types of CO$_2$ respiration input data. ’

P10140L24: "Working exclusively with synthetic...": While this is a strength (in giving the MS a very clear scope) it is also a weakness - so much can go wrong when applying an approach only tested on synthetic data to the real word. While the first approach has the advantage that we know the truth, the second (in which this is often not the case) can add important information on the robustness of the approach that even adding noise to the synthetic data often cannot mimic. Think e.g. of the pressure pumping discussion (Takle et al. 2004, Agric. Forest Meteorol. 124:193-206), which might underlie also the fact that to my knowledge no group succeeded so far in deriving efflux from soil concentration profiles only without using surface flux (chamber) measurements for calibration. Is there any chance to add a brief case study with field data?

We certainly understand the interest, and shared that perspective at the outset of the study, and even actually designed and conducted a field data collection study in hopes that we could plug it in directly to this modeling work. However, we recognized over time that this would eventually exist as another study. To do this well needs careful translational work to take this theoretical approach and to make it work for the real world. In particular, we need to understand how closely we must characterize certain phenomena, and we have been conducting experiments to investigate the consequences of doing things in x or y manner. Our approach has been to take datasets collected over different years in the same region, using different techniques (surface flux and/or subsurface concentration). Presumably we should get a similar answer for most of these studies, or at least differences should be explainable by the current study. We feel that this approach will do better justice to the topic of real-world applicability. We hope to conclude that study in early 2016.

P10141L06: Are there any plans to make any of these versions available to the public?

We have discussed an R Shiny portal that might launch model runs, but there are no specific plans for that now. Once we are comfortable with its performance in field settings, we would be happy to collaborate with others, by sharing the code, or processing data on their behalf (via web portal, or personally).

P10141L11: 1 m and 100 layers: The way it is written now, it sounds as if those settings were hardcoded very deep inside the model. Is this true? If not, I would recommend a rephrasing that indicates: Here we have the way the model describes physical processes, and there we have the settings used for this particular study.

Thanks for pointing this out. These values are certainly not hardcoded inside the model, and can easily be changed. We have clarified the text to explain that the model uses $n$ number of layers, and that we used 100 for this study.

P10141L19: Same as above: The formulation leaves unclear whether the model a) is constrained to parameterizing the temperature time series as a few sine waves, b) similar but could be run on real-world temperature time series after Fourier transformation (i.e. can use a very large number of sine waves efficiently), or c) is prepared to "eat" real-world temperature time series already (as would be desirable for an inversion model) and transform (if necessary) them itself, but this feature wasn’t used for the current study.
Similarly to above, the choice of temperature profile can easily be changed to another function. Thank you for pointing out this confusion. To clarify this confusion, we have partially re-written the section called “Incorporating External Data” to explain how the model will be used with real world data in the future. Basically, as described in the section in studies with real data we have been “eating” the comparatively coarse real world data in pre-processing, and this rewritten section explains how we populate such data into the layers and timesteps. While real world data use is NOT the topic of this manuscript of course, we agree it is useful to discuss this so that readers can look forward.

The re-written section reads as follows:

**Incorporating external data**

In order to model soil conditions at field sites, real soil measurements will eventually be used to drive the simulation. Measurements of temperature through depth, soil volumetric water content, CO$_2$ surface flux and CO$_2$ concentrations take place at 1800 s intervals at our particular field sites. Soil temperature is an explicit model driver, while CO$_2$ surface flux and concentration would be used as model constraints. As volumetric water content is assumed in the model to sit as a physical variable related to porosity, it does not currently sit as a formal driver of biological respiration. When using real world data, our would (at least initially) be to perform inversions for temperature sensitivity only during periods during which soil volumetric water content (and thus porosity) was more or less constant. For simplification, we would also normally assume constant soil volumetric water content through depth, though the model could readily accommodate different depth parameterizations for either moisture or porosity.

We expect that soil temperature will be a crucial variable for accurate real-world simulations, but measured data are coarse relative to the model in terms of depth and temporal resolution. Our practice in test simulations involving real-world data has been to perform a regression of measured temperature through depth (to populate layer temperatures), followed by a linear interpolation through time to populate layer temperatures for every modelled time step. The resultant temperature values replace our originally sinusoidally varying temperature function in the model. The value of thermal diffusivity is implicitly built into these measurements and is no longer required as a direct model input.

Near Eq. 2, though it is trivial, for completeness the symbol t should be described (must be time, and if important for the model implementation, also an added phase lag for the respective sinewave).

Done, as suggested.

Eq. 2-3: Do I get it right that this is meant for soil thermal properties that do not change with depth?

The reviewer is correct that in this study, all modeled soil layers were assigned the same value of thermal diffusivity. But, it could be done otherwise. To clarify, we have added the following text:

“$D_\{T\}$ is coded into the model as a depth dependent variable, though these simulations assumed a constant value through depth.”

Also, the depth diffusivity only affects the temperature profile of our synthetic data sets, as is now inferred in the new section 2.3 on Incorporating External data:

“When using real data the “value of thermal diffusivity is implicitly built into these measurements and is no longer required as a direct model input”.
Eq. 4: Again, this is a reasonable choice for a single synthetic study but for future applications of the model (think e.g. of agricultural sites which are more likely to have a depth-independent source across the tillage horizon) it would be good to learn that the model is modular, and that this profile shape is the particular choice done for this study.

Yes, it would be straightforward to change the production profile. We have added the following text:

“Any function that relates production to depth could be used in place of this parameterization, but we chose this one as it has been used in most of our past soil model efforts.”

Eq. 5: Are the porosities assumed constant over depth and time?

Again, for this study they were. We have also clarified how we might approach this issue when using real-world data, in the new section “Incorporating External Data” (added in response to an earlier comment):

“As volumetric water content is assumed in the model to sit as a physical variable related to porosity, it does not currently sit as a formal driver of biological respiration. When using real world data, our would (at least initially) be to perform inversions for temperature sensitivity only during periods during which soil volumetric water content (and thus porosity) was more or less constant. For simplification, we would also normally assume constant soil volumetric water content through depth, though the model could readily accommodate different depth parameterizations for either moisture or porosity.”

To summarize most of the above comments on section 2.1, maybe it would be good to give section 2 a structure (or maybe it is already sufficient to move some information from here to the next subsection 2.2) that distinguishes more clearly between what’s at the heart of the model (including any limitations on the vertical and temporal variability of parameters that might result from the chosen equations) on the one hand, and parameter choices for this particular study on the other hand. For example, the time step (P10144L12) is mentioned only in 2.2 although it is probably not more or less fix than the number of layers which is already mentioned in 2.1. By the way, maybe you could comment on the interdependence between layer thickness and time step to obtain stability? Since it is mentioned later that the study used up a lot of computation time, it would be helpful to check that both values (layer thickness and time step) are not much smaller than needed for robust results.

We agree, and in response to this comment we changed the section titles, and some text, to more clearly define the “Base Model” from the “Inversion Process”. Hopefully this delineation helps to alleviate any confusion.

Regarding instability and efficiency, we have been using variants of the base model for nearly a decade, and generally understand its limits of stability. In every study, we do conduct some limited sensitivity tests to increase efficiency, but we don’t push the envelope too far. Our interest is mainly in error avoidance, and not efficiency. For each study study, we also always conduct an error analysis specific to the study model and parameterization, using synthetic datasets – in the same way as an instrument operator runs blank samples. Our error tests for this study were mentioned in the text, but we have clarified the statement to express that we did so across the ranges of parameters used in the study:

“Deviations of modelled from analytic concentrations after spin-up were found to be far less than 1\%, across the ranges of parameters used in this study.”
We acknowledge computational efficiency than could be enhanced, and the path towards that is a straightforward one, involving full sensitivity tests. In the future when we might look toward releasing the model publicly, this would have to be done. But in our view this process would still be less focused on maximizing efficiency, than guarding against error. Unfortunately other users will not take the time to error test model output for their particular parameter set.

Section 2.3: Unclear how this agrees with section 2.1. When/where/why were the sinusoidal temperature waves replaced? Capitalize "soil" at beginning of new sentence (P10145L01).

In response to a comment above, we have reworked the text in section 2.3, particularly around the use of temperature data.

Section 2.4: Maybe one or some references on inverse modelling in soil physics and/or biogeosciences would be helpful. Have you considered the possibility to weigh observations according to their reliability in the objective function (e.g. make all concentration measurements together as important as the surface flux measurements - if such combinations were tested, see comment on Table 2. In this case it would also be somewhat arbitrary to weigh the problem arising from the different units of flux and concentration, e.g. by normalizing differences in Eq. 9 by the magnitude or variance of the measurements). The word "pair" near the end is somewhat confusing (since you suggest in the Equation to compare more than one time series of modelled and measured values), maybe "parameter set" would be more consistent here?

Regarding the objective function, we agree that alternative forms could work, that that the form depends on aim of the modeling study, and available data. To address this, we have added the following text:

“This form of objective function is well suited to synthetic data testing, where we have abundant constraining data. But admittedly in future studies involving applications to real data, we could examine alternative forms of the objective function that could be weighted towards particular parameters of interest.”

We have replaced the word “pair” with the word “simulation” which is more specific in this instance and which should help clarify here.

Section 2.5: Again, this section, section 2.3 and section 2.1 somehow together seem to try to give a complete picture of what was actually done - but it is hard to understand, at least to me.

This comment is an aggregate of the more detailed comments above. In response to the earlier comments we have tried to address the reviewer’s uncertainties – and we thank the reviewer for these requests for clarity, because this definitely helps improve the manuscript.

In addition to those detailed changes, we have also slightly restructured the information in these sections. For example in the section title “Base Model Description”, we removed the profile depth and layer thickness, and instead included these in section 2.2 Model Execution and Validation. In section 2.1 we attempted to clarify that almost all of the governing model equations can be changed depending on the soil type one wishes to represent:

“For every modelled time step, each soil layer has a defined temperature $T[i]$, biological $CO_2$ production $P[i]$, $CO_2$ flux $F[i]$, thermal diffusivity $D_{(T)}[i]$ and gas diffusivity $D[i]$. The functions
used to generate these profiles through depth and time can be changed depending on the soil type one is trying to represent. The functions chosen for the purposes of this study will now be discussed.

P10146 / Table 2: The table suggests that surface flux measurements were never tested in combination with any in-soil concentration measurements, while the wording of the abstract (especially "and/or") suggests that combinations were tested.

Thank you for pointing this out. The wording has been fixed.

P10147L06: To be reproducible the noise-adding needs to be described in somewhat more detail. Did you add normally distributed random values, and were they correlated in time or not?

The added noise was not normally distributed. It was added randomly to all measurements in the synthetic data, with no correlation in time, using a random number generator in C. This was meant to simulate the random uncertainties in field sensors. As such we have changed the text to read:

“Field-deployable CO$_2$ sensors typically have 1-5 \% error. To see how the model and inversion would perform under these conditions, errors of 1, 5 and 10 \% were added randomly into all components of the synthetic data. This error, which had no correlation in time or across variables, was generated using a random number generator in C. So with a 1\% error in synthetic data, each value would be anywhere in the range of 1\% of it's value in the original data file. The effect of these errors on the inverse method were observed.”

P10149L07: Clarify if you really want to say that deep soil or surface flux measurements worsen the results, or if they just add no or little value.

It was found that deep soil and surface flux measurements worsen the results, the wording has been changed to:

“Deep soil measurements and surface flux constraints resulted in the largest errors, and should therefore be avoided if the aim is the minimize overall error.”

Figure 4: When using small letters in the caption, do so in the figure subpanels as well. I'm not sure, but when looking at the dependence of the Q10 estimate deviation (especially on Q10 itself), it might be more logical to look at the relative rather than the absolute error. Increases of the error with Q10 (or a part of such an increase) might otherwise be a quite trivial result of the lognormal nature of Q10 (with a lower bound at zero and theoretically infinite maximum values).

We appreciate this comment, but have left this figure as-is, because we generally find that most readers can more readily mentally digest the absolute values.

P10152L05: delete "a" from "which is a fairly high". Maybe, to avoid confusion, you should add that x $\mu$ mol m-3 s-1 CO2 production are equivalent to x $\mu$ mol m-2 s-1 of efflux / column-averaged production, due to the particular choice of a 1 m deep modeling domain. I guess more readers will be more familiar with surface flux magnitudes than those of volume-related production.

Done, as suggested. Thank you.

Section 3.3: The title might be a bit misleading. At first I was expecting typical errorscapes from the inversion to be discussed here (i.e. how the objective function depends on the tested parameters). In
fact showing those would be an interesting add-on because it would help to decide whether your inversion really needs to scan the whole parameter space, or can make use of a simple optimizer that quickly runs into the next local minimum of the objective function. What is shown instead is just a 3-D-version of what was already shown in the previous section, right? In this case maybe it should better be included into that section and shortened a bit.

We interpret two comments here – one related to clarity, and the other related to the choice of optimization process.

Just in case it is misleading, we have shortened the title to read simply “Error Landscape”, so as not to suggest more than we presented. We do show some 3D plots for error in the parameters of specific interest here, but we restrict ourselves to showing the parameters which were defined as being of specific interest at the start of the study (depth of production, Q10, and different sensor placements).

We certainly appreciate the reviewer’s excitement regarding real-world implementation of the approach, because this suggests to us that the reviewer is already looking past this study. Many of our previous studies using soil models have used simulated annealing and other techniques (see in particular publications from our group involving Chance Creelman), but Brute Force tends to be our typical beginning point for first-order work where we might be more concerned initially with the applicability of the base model within an inversion approach. Particularly when moving to long timeseries of real-world data, it would definitely be logical to investigate alternative optimization practices. But, we see this as an issue for future studies: The reviewer is certainly right in that there is a suite of issues that can/will be addressed in subsequent studies involving real-world implementation.

P10153L08: "which soil conditions should be avoided": Better write something like "which soil conditions will be most challenging / pose the biggest problems etc.". If the ultimate motivation behind the inversion study is to characterize global variability in soil respiration parameters (see introduction), the final application cannot afford to systematically skip sites with certain properties that might make them exceptional also in terms of their respiration parameters. Rather, it should be aware that results from such sites will be associated with the largest errors.

This has been changed as suggested, thank you. It now reads as:

"It remains evident which soil conditions are at risk of producing higher inaccuracies, and should therefore be approached most cautiously."