Author's response

We highly appreciate the constructive comments and literature advice to further improve the paper. We considered all suggestions as prescribed below. The different colors indicate: (grey:) editors comment and (black:) author's response.

I find this a very interesting and robust study, which unfortunately still needs a bit of fine tuning, in particular in the presentation. By reading the abstract I was under the impression that this study was "merely" an application of a data assimilation method result to larger scales, whereas in truth the study expands to uncertainty in met-forcing and initial conditions, which is typically ignored in large-scale applications. I think this achievement deserves a bit more presentation in abstract, results and also discussion.

Thanks. We will revise the abstract, results and discussion accordingly.

In the introduction, the text could be streamlined to focus on the current study and avoid side-lines as the use of atmospheric inversions or the future predicted climate of the Rur valley.

The introduction has been revised as suggested.

I've been missing an introduction section that addresses in some more details the motivation for not only using DA to obtain model parameters, but at the same time use of perturbed boundary conditions (meteorological and soil), and a discussion of this in Section 4. I think this is really a novel contribution, which deserves somewhat more place.

We agree and will extend the introduction and discussion sections to address this.

Methodologically, I am unsure whether the assimilation procedure included the spin-up period, or not? In the latter case, one should caution the interpretation of the effect of the assimilation on the cumulative NEE?

This point is not clear to us. Can the reviewer please clarify this?

The discussion is simply too cursory and needs to better take account of existing literature and explaining the added insights gained from this study. This could be partially achieved by taking considerations from the conclusions and expanding these ideas in the light of the existing literature (or potential applications for regional modelling).

We agree that the discussion was too short and will extend/extended it as suggested.

Minor comments:

The abstract is quite long, consider shortening.

The abstract is revised now and somewhat shorter than before.

P1 L19: give average error reduction in mol CO2 / m2 / s ?

We did not include absolute error reductions here, because we used non gap filled data to calculate the NEE sum. The absolute values here may be misleading.

p1 L21: a) is this in agreement with the observations (the fact that the NEE goes positive); b) add "simulated" before regional carbon balance estimates

a) We do not say that NEE itself but $\sigma_{\sum NEE}$ (the model uncertainty of the NEE sum) did increase, b) "simulated" is added as suggested.

P1 L22: here and elsewhere it would be important to understand whether you have brought the model into equilibrium in terms of the carbon cycle with the parameter set used, or whether you relied on a different way to initialise carbon stocks

As shown e.g. in Post et al. (2016), parameters strongly depend on the set of initial conditions. A main conclusion of this paper was that strictly speaking, estimated CLM parameters cannot necessarily be considered valid when combined with initial states very different from the states that were used during parameter estimation and model evaluation. The strong correlation of initial states like carbon-nitrogen pools and parameters is a central problem in the current usage of complex land surface models. Unfortunately, due to the high computational demand, it is not possible include the spin-up in the parameter estimation procedure for land surface models like CLM.

P1 L26: It would be helpful to add in brackets the difference in annual integrated NEE in mol / m2 / year

We will add this as suggested.

P1 L29: If the uncertainty was indeed reduced, then this would need to be demonstrated by comparing prior and posterior distributions of the regional NEE. I guess you rather would like to state that the accuracy of the projection has increased because the model better fits site-level observations?

Since we did not present a reduction of model uncertainty by comparing the prior and posterior uncertainty, we agree this sentence should be reformulated. Indeed, the uncertainty of modeled NEE that can be attributed to the parameter uncertainty was reduced. Figure 6 highlights how the model uncertainty increases if only one key parameter (here Q_{10}) is perturbed. We are going to rephrase the conclusions here.

P2 L11: define what "conventional interpolation methods" are

We meant simple spatial interpolation methods and can modify or delete this sentence.

P2 L12ff: all correct, but I don't think this paragraph is necessary for the sake of this paper.

We intended to highlight the importance of regional scale and high resolution LSM modeling approaches to examine and predict carbon flux dynamics and uncertainties, as it was done study, but can delete this paragraph.

P3 L2: please define more precisely what "error" is here. There is no principle error in averaging the input data, and then obtaining model output from this. The question is whether the aggregation method is sufficiently representing the average regional flux, is of course relevant.

We will reformulate: "The model-data discrepancy as well as the reliability of model input data"

P3 L32: I disagree that the uncertainty of carbon fluxes has been overlooked so far. I agree that the uncertainty hasn't been sufficiently quantified, and/or reduced.

This was mainly referring to regional or larger scale modeling studies. We will reformulate the sentence.

P4L1: well, if you counted conference papers, it actually has. I don't think the question of "who was first" is really relevant, and would focus more on the value and design of the regional set-up

We agree that the question of "who is first" is not relevant but just intended to highlight the significance of our study in this context. We will delete this sentence.

P4 L6: define what a "validated parameter" is. I assume that this is a parameter derived from data-assimilation? It is unclear to me why new estimates of parameters have been obtained for one PFT. Is this because you extended your PFT set to a new PFT (from 2016), or because the DA method of 2016 did not yield good parameters. In the latter case, would it not have been appropriate to recalibrate the model for all sites?

"Validated parameter estimates" refers to the parameter estimates which have been successfully estimated and evaluated in the previous study, i.e. yielded NEE model outputs closer to the measured values than the default CLM parameter values. Correct, for one PFT (C3-crop) the estimated parameter values from the previous study could not be successfully evaluated, i.e. did not clearly reduce the model-data misfit. Therefore, we estimated new parameter values for C3-crop in this study.

P6 L24: add "average" before percentage PFT cover.

It is added now.

P7 L11: Please briefly explain, why it was necessary to add a second spin-up

This is the default procedure suggested in the CLM user manual and is mainly necessary due to technical reasons. We will clarify in the paper:: “The model states obtained after the 1200-year spin-up were then used as input for a second three years “exit spin-up” also using the meteorological data for the years 2008-2010. The “exit spin-up” in CLM is necessary for technical reasons and switches the CLM settings from the (accelerated) spin-up mode to the “normal” mode in terms of the calculated carbon-nitrogen cycling.”

P7 L22: The statement of robustness of the method either requires a proof or (more likely) a reference

We added “(Vrugt, 2015)”.

P11 L 25: I make this note here, although this probably needs mentioning later in that by not manipulating the “stable” nitrogen pool you basically conserve the amount of N available from net N mineralisation across the ensemble. That’s fine (in particular, because the system are fertilised), but probably implies that you underestimate the true uncertainty in the net N availability at the sites.

We agree and can add a sentence in this respect.

P15 L7 (and elsewhere): please refer to other parts of the manuscript as sections, not chapters.

We will modify this.

P15L11: Please be more precise, this is not really a general finding, it is specific to Q10. Please have the revised manuscript cross-checked by a native speaker, and remove unnecessary words (such as “however” in the middle of a sentence, e.g. P15 L8).

We will correct this and carefully recheck the manuscript for unnecessary words and grammatical errors.

P15: This discussion section is a bluntly speaking a bit too skinny. More could be made for instance of the extremely relevant and very nicely demonstrated point that minor modifications in input meteorology lead to markable differences in modelled NEE.

We agree that the discussion is too short and we will extend it.

However, generally, a critical assessment of the novelty of the study approach and finding compared to the existing literature is missing.

It was difficult to compare our findings with the existing literature given the very few studies which evaluated LSM uncertainty in a similar way as was done here for carbon fluxes. However, we will recheck the literature again and try to extend the comparison with existing literature in the discussion section of the paper.

Table 3: Please arrange the table such that it is immediately clear which CLM4.5 default simulation belongs to which site

We will re-organize the table.

Table 4: please briefly explain the abbreviations used to denote “grid cells”, and how the “n” were calculated (or what they are).

In Table 4 “n” is the amount of non-gap filled half-hourly measurement data that was available to calculate these evaluation indices. We will clarify this.

Table 5: define “n”

Here, n in the number of grid cells in the catchment covered by more than 80% by one specific PFT. We will clarify this in the paper.

Figure 2: define “evaluation period”. It would also be preferable to somewhat make a clearer distinction between different model ensembles and the observations

The dates of the evaluation periods are summarized in Table 1.