Reply to the editor:

I remain concerned that the identical twin tests do not recover the parameters. I was disappointed that you did not resolve this in your revision. If identical twin tests fail, it is not usually appropriate to proceed with using real data. Instead the experimental design should be adjusted until the test succeeds. In such a situation I would probably initially try reducing the number of parameters being estimated in order to make the problem more tractable. Did you try this? The reviewer does still think that the methodology and results presented are useful to the community, but there needs to be much more discussion of the limitations. I hope you can see their second review on the manuscript overview, and will be able to provide a revision in due course.

We would like to extend our sincere gratitude for your assistance in guiding us through the publication process of our paper 'Optimising CH4 Simulations from the LPJ- GUESS Model v4.1 Using an Adaptive MCMC Algorithm' in GMD. We appreciate your response and the decision to call for a second revision, which we believe would be instrumental in refining the manuscript.

With regard to the issue you raised in your last letter, we agree that, in principle, the twin experiment should be able to recover all the parameters completely. We acknowledge (now also in the manuscript, section 4.6) that the twin experiment in this study did not fully recover the true values of certain parameters, especially the parameter CH4/CO2. As mentioned in the manuscript and in responses to the reviewers, LPJ-GUESS is a highly complex and highly non-linear model. There is a significant likelihood of different parameter combinations yielding similar results for the same problem (known as equifinality). Based on our understanding, we assume that the equifinality problem might be inherent to the parametrisation in LPJ-GUESS rather than the optimisation algorithm. We provided some indications of this in our response to Review 1, demonstrating the offsetting behaviour of lambda_root and CH4/CO2. In two prior twin experiments, we observed that when one of these parameters converged, the other one consistently exhibited offsetting behaviour and did not converge.

The high dimensionality of the problem and limited variability in the data for its resolution could be another reason for the poor convergence of some parameters. Considering the high degree of non-linearity in the model, we believe that relying on a single type of flux from a single site for the twin experiment might not be sufficient for identifying all the parameters. Therefore, additional sites spanning a wide climatic variability and/or different species of fluxes from the same site may be necessary. For instance, in the case of the parameter CH4/CO2, one reason for the poor recovery could be that the study only constrain the CH4 component, while the part of CO2 is significantly off. Incorporating both CH4 and CO2 fluxes to address both sides of the ratio can lead to a more accurate convergence. These considerations, however, are beyond the scope of this paper, and further studies are currently in progress in these aspects. Another reason for the incomplete convergence could be that the model is not fully constrained by the limited dimensions in the parameter space. Achieving complete convergence may require incorporating additional parameters from different modules of the model that represent various processes.

In response to your suggestion to redo the twin experiment with a reduced number of parameters, technically, it is possible to decrease the number of parameters and rerun the experiment. However, this approach introduces some challenges. A reduced number of parameters would lead to the derivation of an entirely different subsystem of the model, which may not necessarily represent the complete system examined in this study. In the current setup, the parameters co-vary simultaneously after each accepted iteration, resulting in a set of parameters that undergo offsetting and adjustments throughout the exploration process. If we were to reduce the number of parameters, the system would perform the same operations in a more simplistic/low-dimensional manner and potentially recover the original parameters. The problem here is that it won't have much scientific relevance if we use a subsystem or even multiple subsystems separately to have all the parameters included in twin experiments that recover the true values, but then use the full system with all the parameters for the real data optimisation. Alternatively, if we choose a new subsystem for the twin and real data experiment, this would necessitate redoing the entire experiment and rewriting the entire paper with a different set of parameters. We are concerned that this is not a minor revision, but rather turns into substantial revisions that would ultimately change the entire paper. Another significant issue here is the large computational time involved.

Instead of redoing the twin experiment now, we would like to explore the possibility of addressing the reviewer's comment and explicitly tackling the issue of the non-converging behaviour of some parameters, especially of CH4/CO2. This is also along the lines of the reviewer's request. In response, we have now added a detailed description of this issue in Section 4.6. We hope this will sufficiently address your concerns and the reviewer's suggestions.

Apart from this, we kindly request your attention to a minor change we made to the paper at this stage. We have altered the name of the MCMC algorithm we developed from G-RB AM to GRaB-AM for the ease of use and readability.

Once again, we would like to express our gratitude to you and the dedicated reviewers for your and their time and effort in revising our manuscript. We eagerly await your positive decision and the publication of our work.

Reply to the referee #2

I still have some concerns and confusion about the CH4/CO2 parameter. Your sensitivity analysis shows that CH4 flux is most sensitive to this parameter, yet the twin experiment fails to recover it (or even really explore the full space of this parameter). When real observations are used the CH4/CO2 parameter is estimated to be considerably lower than the prior, but this parallels the incorrect underestimation of this parameter by the twin experiment, eroding trust that the optimisation framework is improving simulated fluxes for the right reasons. Given that many readers may have similar concerns and confusion, I would suggest incorporating some of the discussion in your review response (i.e. the impact of model equifinality, future work on simultaneous optimisation of CH4 and CO2) into the paper. Perhaps this could fit in an expanded Section 4.6 that includes discussion of future work.

Many thanks, your observation is true. Though the parameter CH4/CO2 in the experiment shown in the paper has been observed to tend to approach the true value, it has failed to converge to the true value. On the other hand, the parameter has shown convergence to the true value in our other experiments, as shown and described in Revision 1 (Section: Specific Comment). We assume this parameter might exhibit an offsetting behaviour to the parameter lambda_root.

Considering your valuable suggestion and the possible confusion readers might have, we have now extended Section 4.6 with a more detailed explanation of the nonlinearity of the model, the possibility of equifinality in this context, and the need for more species of flux integration in the system. On the other hand, we believe that redoing a twin experiment with different starting points, longer chain lengths, or different sets of parameters would require a tremendous amount of work and would ultimately result in significant changes to the entire paper (please see the reply to the editor above).