

Interactive comment on “Modeling the vertical soil organic matter profile using Bayesian parameter estimation” by M. C. Braakhekke et al.

Anonymous Referee #2

Received and published: 19 October 2012

General comments

The paper describes a parameter estimation for the model SOMPROF using Bayesian calibration. The model is used to describe the vertical soil organic matter formation and is calibrated by an inverse technique using soil carbon measurements and $^{210}\text{Pb}_{ex}$ -concentration measurements. Two sites with different SOM profiles were independently calibrated and the posterior parameter distributions were analysed.

The study is overall well structured, well written and includes an ambitious statistical modelling. The work does not only include the results of the parameter estimation itself, but also an interesting interpretation with regard to the modelled SOM profile. Additionally, it provides some advices on how to deal with problems arising when using

C4938

Bayesian calibration.

The study has, however, some shortcomings especially with regard to mathematical formulations.

Specific comments

0.0.1 p.11253 line 25:

Equation (6) is wrong as the Likelihood is not a function depending on the observations O , but instead it is depending on the parameters θ . One advantage of using the Bayes theorem is in particular that the probability of the model parameters **given the observations** can be transposed to the product of a constant, the prior probability of the model parameters and the probability of the observations **given the model parameters** (which can often be calculated more easily). Therefore the authors should write:

$$P(\theta|O) = cP(\theta)L(O|\theta). \quad (1)$$

$L(\theta|O)$ has to be changed throughout the article. Furthermore, as the density is used in equation (6), the authors may consider to write P in lower case as it is done in Gelman et al. (2003).

0.0.2 p.11254 line 5:

The authors claim that no analytical expression for the posterior exists, but do not give reasons for that, at this point. It might be better to move the paragraphs from line 5 to line 22 on page 11254, after the next two sections (2.4.1 and 2.4.2), where the likelihood as well as the prior function is explained in more detail. This would make it more easy to explain that no analytical solution for the posterior exists.

C4939

0.0.3 p.11254 line 25:

The authors chose a log-normal distribution for their statistical model, because the observations are bound to zero (I assume they mean the observations are positive or non-negative). But as there are more distributions, which are non-negative e.g. the gamma distribution, why do they chose the log-normal distribution? Furthermore a normalising constant for the Likelihood in equation (7) is missing. In this form it is not the Likelihood of the underlying statistical model.

0.0.4 p.11257 line 12:

The authors say that most of the parameters were constrained by the observations except for k_{RL} and I_m . In Figure 4, however, I do not see a difference between k_{RL} and e.g. v or $\alpha_{FL \rightarrow LS}$ for model 1 and 2 in terms of being constrained.

0.0.5 p.11258 line 12:

To compare the modes they use the minimum misfit in the MCMC sample. This should be an indicator for the maximum posterior. In Table 3, however, this conflicts with the explanation of misfit as the minimum of the (log) posterior. An exact definition of misfit might be helpful.

0.0.6 p.11282:

Table 2 needs more explanation. Especially that the priors are only used for model 3. Is the lower bound always 0?

C4940

0.0.7 p.11286 Fig.3:

The prior of v is illustrated in plot E as uniform between 0 and 3. In table 2, however, the upper bound is 0.1. Which one is correct?

0.0.8 In Appendix A2 p.11270 line 7:

The authors explain that they used the correction factor for the statistical model 1 and 2 but not for model 3. But it is not clear, whether they transformed the model parameters for Model 3 or not. If they did an explanation or a reference why this is still correct would be needed.

Technical corrections

0.0.9 p.11261 line 9:

I would suggest not to use significant in a paper dealing with a lot of statistics, without proving that it is statistically significant.

0.0.10 p.11262 line 5:

There is a missing "to" between compared and material.

0.0.11 p.11263 line 11:

There is a missing "is" between This and in.

C4941

0.0.12 p.11264 line 22:

There is a missing “the” between by and fact.

0.0.13 p.11273 line 18:

The authors probably mean M. van Oijen (instead of M. van Ooien).

References

Gelman, A., Carlin, J., Stern, H., and Rubin, D.: Bayesian Data Analysis, Chapman and Hall, 2 Edn., London, UK, 2003.

Interactive comment on Biogeosciences Discuss., 9, 11239, 2012.