

Interactive comment on “Soil respiration compartments on an aging managed heathland: can model selection procedures contribute to our understanding of ecosystem processes?” by G. R. Kopittke et al.

G. R. Kopittke et al.

vanloon@uva.nl

Received and published: 19 February 2013

We very much appreciate Mr. Wutzler’s extensive and constructive comments, and provide answers to the issues & questions raised by him below.

General comments

A) Kopittke et al. study statistical models of soil respiration and soil heterotrophic respiration at a temperate heathland site. In addition to numerous already existing soil respiration modeling studies they make explicit the usually implicit process of model

C8315

building and model validation. This is a significant contribution to the biogeochemistry and soil carbon community. Especially the usage of two levels of validation plots and periods is exemplary for future designs of experimental setups. The analysis is sound and I can follow all the procedure, because they give all the details. However this high level of detail needs to be balanced with still missing clarity of the main points and conclusion. What did we really learn from this model selection?

In order to enhance clarity and better emphasize the overall outcomes of the study, we made the following changes: rewording the abstract, rearrangement of some parts of the discussion and the addition of a conclusion section to more closely link the findings of the model selection procedure and ecosystem processes.

B) Since the model selection process seems to be the part most contributing to community understanding and knowledge, the study would greatly benefit by a comparison to the often encountered process that uses only the calibration data or uses all the data as calibration data. Do the models and conclusions change? What are the reasons in favor or against cross validation to make usage of the full amount of data? Would conclusions change if only fewer measurements were available?

We think this is an excellent suggestion by the reviewer. In addition to the calibration-validation cycle that was already adopted in the original manuscript, we have added a model selection based on a 10-fold cross-validation routine. In addition, we included a model calibration phase, where all data was used to calibrate the model and no separated dataset for validation purposes. The results from this are listed in Appendix C and mentioned in the results and discussion sections of the revised manuscript.

It turns out that calibration, cross-validation and validation errors are highly correlated (Pearson correlations > 0.99) in our study. Apriori we did expect a correlation, but not such high correlations. It implies that the result of this study would not be influenced by using only a calibration cycle for choosing a model. Our study is perhaps special in the fact that the calibration and validation sets used are very similar and located at

C8316

the same site. A more rigorous approach would be a case where different sites (i.e. geographically separated) are used for calibration and validation – this is recommended in the discussion section.

C) Together with focusing the paper, the paper would also benefit from shortening. Can parts of the methods be moved to appendix without compromising the general understandability? Results can be shortened by e.g. stating that for total soil respiration LMM, LMM2 and Selsted models performed much worse than the GLMM models and hence their detailed model variant results do not need to be reported and discussed.

The authors agree that shortening the paper would be beneficial and appreciate the Reviewer's suggestions. The following changes have been made: - A second Appendix document has been created and the full description of the PG method has been moved there, leaving only a shortened version in the method - The results section 3.4 and 3.5 have been shortened - The discussion paragraph about Nitrogen deposition has been removed (page 16266, line 20-29) as, although interesting, does not add to the overall outcomes of the study. - The diurnal variation data has also been removed from the method, results and discussion (based on reviewer #2 comments).

The request to compare different calibration-validation cycles (previous request) did lead to a slight increase in some parts of the results and discussion sections.

Specific comments

1) With some background on forest soil carbon dynamics I at once miss the incorporation of seasonal variation of available substrate for heterotrophic respiration and hence seasonal changes in parameter R_0 in the model variants. Is litter production in line with average rate of decomposition throughout the year in the heathland?

The comment regarding seasonal variation of available substrate for heterotrophic respiration is an interesting question. It partially addressed in the discussion of the peak heterotrophic respiration observed in spring 2012 (page 16268 line 2). It is also again

C8317

referred to as a potential explanatory variable that other studies have used in respiration models (page 16273 line 5). Litterfall data was only available for the Old vegetation between March 2011 and February 2012; that is, not for the Young and Middle and it only overlaps five months of the respiration assessment. Therefore, this was not suitable to be used as an explanatory variable in the models. However, in a comparison of the general seasonal data trends, the litterfall pattern did not correspond to the observed heterotrophic soil respiration pattern. The maximum litterfall rate occurred in January (8 g m⁻² month⁻¹) and the minimum in February (2 g m⁻² month⁻¹) with gradually declining rates recorded from March to November (7 to 5 g m⁻² month⁻¹). This pattern did not correspond to the observed soil respiration rates and suggests RH is more closely associated with temperature than with litterfall patterns. However, if sufficient litterfall data had been available for inclusion as a variable with soil temperature, it may have improved RH model fit by explaining additional data variation. This information has been included into the discussion.

2) The study discusses sequestration without presenting other fluxes such as net ecosystem production or losses other than respiration. Please, present those other fluxes or omit the discussion or present it only shortly in an outlook (but not in abstract).

The C sequestration information included in the abstract has now been changed to remove the comment about C balances and losses other than soil respiration. The final section of the paper has been amended to emphasize that these statements are hypotheses only, based on the RS and PG data used in the current study, and could be tested with the NEE and ER data using a similar modeling approach as described in this study.

3) How general is the result of no further information of plant variables in addition to temperature for soil respiration?

The soil respiration model that used soil temperature alone could certainly be improved, if the important co-varying factor(s) (that that were absent for this study) could be iden-

C8318

tified and included in the modeling process. However, after assessing 7 variables in addition to soil temperature (soil moisture, Calluna biomass, PG, air temperature, and in a preliminary phase also root biomass, microbial biomass and PAR), we did not find any that were statistically significant. To make this finding clearer, it has been included in the revised version of the abstract and in a new section "Conclusions".

4) The age classes are not randomized. I miss a short discussion on whether and which other covariates can cause differences between the young and the other age classes.

Dr Wutzler is correct in his statement that the age classes are not randomised. This is discussed on pages 16244 (line 26)-16245 (line 3), where the experimental layout is explained. A short explanation has been added to the text, stating that: "Due to the inherent nature of the site, randomization of the factor 'community age' is not possible in our experiment. However, collinearity of weather data with the distribution of the three age classes is highly unlikely since the area is small compared to variations in weather-variables. Furthermore, soil data (including soil temperature and soil moisture) appear not to vary much between the age classes."

5) Please, guide the reader a bit more on the usage of log-link function and the GLMM approach. E.g: in the model-method section "By using the log-link function we study the log of the respiration and hence the model with exponential terms becomes a linear model. "

The text now includes a brief explanation at the point where the GLMM is mentioned for the first time: "Generalized linear models (GLMs) extend linear models that involve non-normal error distributions or heteroscedasticity and may also require a transformation to become linear. Linear functions of the predictor variables are obtained by transforming the right side of the equation by a so-called link function. In this case the shape of the relation is exponential, so by taking its logarithm it becomes linear. The data are then fit in this transformed scale (using an iterative routine based on least

C8319

squares), but the expected variance is calculated on the original scale of the predictor variables. The Gamma distribution describes that the error is right-skewed at low values of the predictor variable and becomes symmetric at higher values. In our case, the mean and variance of the model error are equal (McCullagh & Nelder, 1989)." We decided to stop here with the explanation of GLMs and refer to appropriate literature rather than to explain more details. Please note that the Gamma distribution in our study is a 'continuous-variable equivalent' to the Poisson distribution (in the previous version of the manuscript we attempted to keep things simple by stating that it followed a Poisson-distribution, which is strictly speaking incorrect because the Poisson distribution only applies to discrete data).

6) To appreciate the model adequacy a plot of the residuals vs temperature and/or time and a quantile-quantile plot of the residuals and the Poisson distribution would be helpful.

We think this is a good suggestion. In the revised manuscript the suggested diagnostic plots for the models that are listed in Table 5 are shown and discussed briefly in Appendix D. Reference to Appendix D is made to from the caption of Table 5 as well as the result section where Table 5 is discussed.

Technical comments

P16241 L2 second largest flux <after XX?> in the <global?> carbon cycle P16242 L2 use <mostly> recently produced P16254 L15 generalized <linear> mixed effects model The introduction has been amended in a number of locations to clarify sentences. These changes have occurred on Page 16241 Line 2 (of the original document) to read "second largest flux after gross primary productivity in the global carbon cycle."; Page 16242 Line 2 to say "use mostly the recently produced..."; and on Page 16254 Line 15 "generalized linear mixed effects model".

Table 4 lacks information on residual variance and variance of the random effects. I suggest to also report the coefficients of almost as good models (with P or M as a

C8320

predictors respectively). The coefficients of the models with P and M as predictors are reported in the revised manuscript (see Table 5 - Table 4 in the original manuscript). Furthermore, the variance of the random effects and residual variance are reported (as standard deviations, rather than variance, so that these values are at the same scale as the RMSE values) in Table 4.

P16255 Please report the values of Max Biomass and Maximum P_G. This will help meta-analysis that compare dimensional model coefficients (if model coefficients are reported in Table 4) The absolute values for Maximum Biomass and Maximum Pg have now been included in the methodology section.

P16259 To my opinion the p values can be omitted, if they are not close to the significance level. The p values in Section 3.3 (Page 16255) have been removed where they are not close to significance level.

P16263 Can the RMSE values be presented in a Table? The text is really hard to read. A selection of the RMSE (calibration and validation) values have been included in a table and deleted from the text, to improve readability.

How about discussing the observation of overfitting (L 11ff) more prominently? The term over-parameterized (instead of over-fitting) was used in the text, where over-parameterization refers to a model that contains more parameters than is necessary to explain the variation in an independent data set and often performs worse than a simpler (more parsimonious) model. This could also be stated as: a model is over-parameterized if it explains more variation in a calibration data set than is appropriate, based on the model performance on an independent data set.

Over-parameterization is unavoidable if a model fit is only evaluated on a calibration data set (and without the use of goodness of fit criterion with a penalty for additional model parameters like R2adjusted or AIC). This is because, in such a case, more parameters will always lead to a better fit. The difference between misfit for a calibration and a validation data set is generally used as a measure for the degree at which a

C8321

model is over-fitted (i.e. tries to explain more variation in the calibration data set than what is appropriate, judging from validation fit). This aspect is well explained in Hastie et al. (2009).

In the original manuscript, we represented the errors for RMSEC for LMM and GLMM models with locations as fixed factors (rather than random factors), and hence the values for RMSEC were relatively low and not directly comparable to RMSEV (viz. comment D by Reviewer 3 and our answer to that). At this point in the discussion we incorrectly made a direct comparison between RMSEC and RMSEV. In the revised manuscript, we corrected this and the result is that the relationship between RMSEC and RMSEV is quite strong, so there is now no indication of over-parameterization. This is now mentioned in the discussion.

Hastie, T., Robert Tibshirani, R, and Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. 2009

Fig9: Are these predictions of the population or the groups/plots? Get the figures confusing if you indicate different groups by different colour or symbols? The symbols show all observations from all plots (ie in the Young graph, the symbols are from all Young plots) and the red line is the prediction for the community. The graph would become too confusing if separate plots were shown in different symbols.

Please also note the supplement to this comment:

<http://www.biogeosciences-discuss.net/9/C8315/2013/bgd-9-C8315-2013-supplement.zip>

Interactive comment on Biogeosciences Discuss., 9, 16239, 2012.

C8322