

## Response to comments

We kindly thank the editor and the two reviewers for their helpful and constructive comments. The response is given in red, italics below each comment, the changes in the manuscript are indicated with bold letters.

### Comments from Executive editor of GMD, Astrid Kerkweg

"The main paper must give the model name and version number (or other unique identifier) in the title."  
"If the model development relates to a single model then the model name and the version number must be included in the title of the paper. If the main intention of an article is to make a general (i.e. model independent) statement about the usefulness of a new development, but the usefulness is shown with the help of one specific model, the model name and version number must be stated in the title. The title could have a form such as, "Title outlining amazing generic advance: a case study with Model XXX (version Y)"."

*OK, done in the revised version*

### Comments from Anonymous Referee #1

Dear Authors, this is an interesting study, but there are still a lot of issues in the presentation of the study as well as with analysis and discussion. The application of one model on an single test site is quite specific, which makes it more important to distribute between site and model specific result and general findings. Especially the later ones I would like to see worked out and highlighted more. Please find my more detailed comments in the supplement.

Please also note the supplement to this comment:

<http://www.geosci-model-dev-discuss.net/gmd-2016-116/gmd-2016-116-RC1-supplement.pdf>

### General comments:

The manuscript "The importance of process interactions and parameter sensitivity for modelling the carbon dynamics in a natural peatland" describes a calibration and sensitivity analysis of the CoupModel, using several variables measured on an eddy covariance test site in Sweden. Several variables, which describe carbon, energy and water fluxes, are used for calibration and a sensitivity analysis.

Overall the manuscript is not well written and difficult to follow. In wide parts corrections by a native speaker is required.

*English copy editing will be provided by the journal in a later state of the manuscript processing*

However, the study is interesting and contains relevant aspects. Unfortunately, the actual presentation of the study is not convincing. Several parts are too fuzzy and too general, while other parts are too detailed. The objective is not clear and the conclusion does not provide any new information despite general knowledge about this field. I am really puzzled to rate this manuscript, as there are many concerns in almost all parts of the manuscript.

However, as I see also the potential of the study, I rate it acceptable with major revisions, but I have to be clear, that only answering the comments below won't be enough to get the publication to an acceptable form. As there were too many issues, it was not possible to comment all in detail, but I tried to explain my concerns on some parts in more detail.

### Overall comments:

The objective is not really clear. Reading the paper it seems like that all variables are needed to improve the quality of calibration, several parameters are interacting and more measurements are needed. I do not need a study to come to this conclusion.

*The main message of the study is that parameters are interlinked, not only within, but also between different modules. It implies that parameter ranges might not be transferrable between studies that use*

different models or even same models with a different set of calibrated parameters or included processes. This hasn't been shown before, as previous C-cycle studies on peatlands usually calibrate only parameters of the Carbon module, or from few additional modules. Further, while multi-criteria constrain is widely used in e.g. hydrological modelling, this technique is still hardly found in C-cycle modelling studies, especially on peatlands. Instead, it is common practice that parameter values are transferred between studies and between models without questioning the covariance between parameters and dependence on the variable and criteria used to reject not acceptable performance of the model.

**In the revised version, we reformulated the objectives to emphasize the understanding of the dependencies between parameter distributions and between parameters and model performance. In the results we added a sentence, telling that using several measurement variables helps to identify if a parameter range is not robust. Also in the discussion and conclusions we reformulated some parts to emphasize the importance of using many variables and criteria to constrain the model.**

I miss out more numbers, that rate the quality, and real values, like how much quality do I miss out, if I calibrate only on one variable rather than on all variables (for R2, ME and NSE).

Calibrating using only one observation variable and criteria will normally create the highest performance for that particular variable and the particular index. However, the result can easily be unique for only that particular variable and time period used and lead to worse performance in other variables or if other indices are used. The advantage of using several variables and indices is to be able to identify which of the resulting parameter ranges vary depending on the chosen criteria and which are robust in this respect (still, this doesn't include the robustness in respect to transferability between models and sites) but is difficult to quantify. Figure 4 and 5 show how much the performance in a certain variable is reduced, if criteria for another variable or performance index is set. It would be possible to create such figures for all combinations of multiple criteria, but this would be several pages of figures.

If only some few parameters are calibrated, the same or similar goodness of fit might be achieved, depending on which parameters are chosen, but parameters will be constraint to a range, which may be misleading because of the tendency of equifinality. To identify the correlation structure between parameters we have to define a list of parameters that have the change to be both correlated to other parameters and sensitive to the criteria and data available.. Fig. 3 tells how many parameters can be constrained depending on which variables are used to constrain the model in the calibration procedure.

Out of 27 sensitive parameters, 15 could not be constrained to an unambiguous range. This means, that in more than 50% of the cases, a parameter range constrained by only one variable or index is not robust because it depends on the chosen criteria. The more variables and performance indices are used and the more parameters are calibrated, the more a statement is possible if the resulted parameter range is robust or not and to which factors it is connected to. **This number was added to the results and the meaning added to the discussion.**

There is no discussion about transferability of the results and the robustness of the results. Which results can be used in general and which are related to the CoupModel. I do not see the list of other studies as a discussion of transferability. If the authors want to include these studies, there need to be an analysis of the differences for the different approaches used for the different processes.

Interactions, also between different modules, certainly exist on other sites/ecosystems and with other models as well. The same applies to the problem of different resulting ranges depending on performance index, measurement variable and it's sub period used for calibration.

But the specific results, i.e. which parameter interact in which way, the constrained parameter ranges, the rank of parameter uncertainty and therefore importance of additional needed measurement variables, and the parameters identified as most sensitive are probably to a large extend model, ecosystem and maybe site specific. As we tested only one model on one site, the only way to make a statement about transferability is to compare with other studies. We mention the model name and the ecosystem of these studies, but analyzing all differences between the studies would include differences between models (used equations, processes that are implemented or not, ...), between applied methods

(calibration procedure, selection of other parameters that are calibrated simultaneously, performance indices, calibration variable, tested value range of the parameter, ...) and between sites (ecosystem, climate zone, soil conditions, vegetation, ...) - all of them might play an important role why this parameter was found to be most sensitive. A full list of the differences would just be too long, especially as this is not the main message of the study. All studies differ from ours in at least one point (e.g. ecosystem type, which is already mentioned in the manuscript), indicating that some results might be transferrable to some extent

**We added at several positions in the discussion the information about whether a certain result relates to CoupModel and site conditions or can be used more general.**

I am also not sure if the picked indicators describing the goodness of fit are well picked. The mean error will compensate strong negative and positive disagreements in the overall value, which do not reflect the quality of the model performance. I would like to see the root mean square error used instead. Also R2 is not a good value for model performance as it might be sensitive to extreme values, if not all parts of the data range are represented equally.

*It is true that strong negative and positive disagreements are compensated in mean error, but they are reflected in the R2. The root mean square error has the disadvantage that it doesn't tell if there is an over- or underestimation. There are many other performance indicators, some of them calculated on base of R2 or ME or the combination of both. We chose R2 and ME because they are simple and we think they are sufficient to show the main message. We agree that a single performance indicator should be easier. The reason for selection both R2 and ME was that we would like to distinguish errors related to the mean bias and the ability to reflect the variability in itself. The ability to reflect the full range between high and low values and being sensitive to the magnitude of the range was part of conceptual thinking behind the criteria chosen.*

The state of the art method for calibration is Bayesian calibration, which is not mentioned in this study. At least in the introduction and maybe in the discussion this needs to be mentioned and explained why the here used method is as good, better or worse than the Bayesian calibration and what are the advantages and disadvantages. Using several variables for calibration of models or a sensitivity analysis the pareto optimization would be an appropriate multi-criteria approach to address the subjective judgement of the model performance. However, at least this technique and/or other approaches for multi-criteria optimizations should be mentioned and discussed.

*The Bayesian approach has a lot of advantages providing that we have a well defined error model and that the multiple variables can be combined into one single log-likelihood value. The high number of different variables and especially the risk for converting into posterior distribution without covering the full range of combinations for all parameters was the main reason for not selecting the Bayesian approach. The Bayesian approach does not show any substantial advantages when we have many different measurement variables and we would like to have an unbiased investigation of all parameter combination rather than searching for a singly highest probability of the entire model.*

**We mention the Bayesian approach in the revised version in the introduction and the discussion.**

#### **Specific comments:**

##### **Comments on the title:**

First, the model used in this study is not mentioned in the title.

**OK, done in the revised version**

Second, the title contains only the carbon dynamics, while most of the variables that are considered in the analysis are energy and water fluxes. The title needs to be reformulated and more precise.

**We reformulated the title to include also heat and water fluxes.**

##### **Comments on the objective:**

- For point 1. the authors do not identify processes, but parameters and variables, which are most sensitive in the model to simulate the target variables.

Processes are described by equations, containing parameters. Parameters determine if an equation results in a high or low value. If the result of an equation doesn't matter for the fit of the model output to a variable, it means that the underlying process doesn't play an important role for the variable in the tested scenario. Therefore, identifying the most sensitive parameters means identifying the sensitive processes. An exception would be if several parameters of the same equation would be calibrated, but this was avoided in this study.

***We reformulated this objective to make it more clear and added the explanation to the discussion.***

-Point 2. is not well formulated and it is difficult to understand the objective.

10 ***Reformulated in the revised version.***

- I am not sure about point 4. Why do you test the usability of measurements? The measured data are usable.

***Translation error. Should be usefulness or potential. Replaced in the revised version.***

15

Additional, it is not true that you can detect the missing measurement variables. You can detect model sensitivity and required data for the used version of CoupModel. Other models might need other parameter sets and might show different sensitivities. Also, the authors work out improved model performance by adding more variables in the calibration process, it still rise the question, if additional measurements are need, the model approach that simulates the process needs to be improved or the calibration approach needs to be improved.

20

***We identify parameter that are highly sensitive and at the same time not constrainable with the available data. As we calibrate only one parameter per equation, it means that the process described by this equation plays an important role. If it is possible to measure a variable that describe this process, we found a "missing variable". E.g. the high concern of a parameter describing the soil water retention curve. This is used for calculation of the soil water content. So either having measured soil water retention, or soil water content, would improve the modeling. However the improvement is not in a better model performance, but in the possibility to constrain this and connected parameters to a more narrow range (and therefore improve predictions which might be performed with this model). Of course we tested it only for CoupModel, but it is probably also an important variable for other models that have some dependence of decomposition or plant growth from soil water content. Only for models that do not have this dependency it indicates that including such a dependency/adding corresponding processes might improve the model performance.***

25

***A much larger limitation might be the dependence on site conditions. E.g. we know from measurements and correlation analyses that water level does not play an important role at every peatland, which might indicate that also water content might not play an equally important role on all peatland sites - e.g. because the water content is not much fluctuating. However, for natural peatlands with hydrological regime related to climate there are strong reasons to believe that our results are general.***

30

***We added "by identifying sensitive or interacting parameters that cannot be constrained by the available data" to this objective and incorporated the response to this comment in the discussion.***

35

- In the objectives it is not mentioned what the authors are actually doing. The model is not mentioned and the four points are not linked to any land use, model or analysis approach.

***We reformulated the objectives to be more precise and added a sentence about what we are doing.***

40

- The sentences after point 4 do not contain any useful information about the actual study, but only general information what you can do with an outcome from a sensitivity analysis.

***It is right, that this information applies to sensitivity analyses in general, but this information might be still be valuable for readers, that are not very familiar with sensitivity analyses.***

45

***We restricted the sentence after point 4 to Carbon models and peatlands.***

#### **Comments on the method section :**

Section 2.2: The gap - filling of the climate data is explained, but not the gap filling of the EC data .



*The EC data was not gap-filled, as mentioned in the last sentence of Section 2.2. Only measured data were used for calibrating the model.*

5 The model description is far too long, but leaves crucial aspects out at the same time. There is also lack the scientific terminology.

*See response to specific comments below*

Page 4 line 20: EC is not defined. Please add this on line 15 the same page.

**OK**

10 Page 4 line 28-29: C uptake by the ecosystem from the atmosphere

**OK**

15 Section 2.3.3 There is no need to give a general introduction into soil hydrology.

*This section describes how soil hydrology is realized in the CoupModel in the used setup. CoupModel provides many possibilities for the user to select between different sub models, different equations and different complexities of the used equations. E.g. ground water flow as well as evaporation can be included or discarded and there is no need for using the Richards equation or simulating soil water vapor in CoupModel. The number of hydrological soil horizons is flexible; instead of Brooks & Corey, the van Genuchten equation can be used for description of the water retention curve, etc. This is all configured by switches through the user - the text describes how these switches were set, which is relevant information that cannot be found in the manual.*

20 *We added a sentence in Section 2.3. to make it clearer, that the following sections describe the applied, study specific configuration.*

25 Section 2.3.4 There is no need for a general introduction into phenological models, but provide the key information: used phenological model, the model is based on temperature sum and day length, parameters and settings. Also the description of allocation of carbon in the plant is too long and not well formulated. Especially, the labelling of parameters in the model do not contribute to a better understanding of the study.

30 *Also for vegetation, CoupModel provides a wide range of opportunities. Mosses as additional plant layer had never been simulated before with the CoupModel, which makes it necessary to describe how the existing C pool scheme in the model was applied to mosses, that do not have roots and a seasonality comparable to vascular plants. Also for vascular plants, the carbon pools were used in an unconventional way that allows considering stems as photosynthetically active and that allows senescence to be dependent on both, temperature sum and growth stage. These are also the reasons for the labeling, explaining how the model was configured and how to understand parameters and equations, that still use a labeling that was originally intended for vascular plants, in particular trees. This information is relevant for reproducing the study and of interest for other CoupModel users that would like to apply the model on a moss/sedge dominated site. **But we agree, that this section is very long and therefore moved large parts to the supplementary material.***

45 Section 2.3.5 The section is too long, in some parts the scientific terminology is not used, the description of the processes is too casual and essential information is missing (e.g. turnover rates of the pools, decomposition follows first order kinetics).

*Turnover rates of pools were calibrated parameters. For a better readability we did not mention any values of fixed parameter as well as value ranges of calibrated ones in the text. Instead they can be found in tables S2 and S3 in the supplement as mentioned in section 2.3, last sentence. A quite large part of this section is occupied by the description of how peat growth was simulated. This functionality was newly developed for the site in this study and therefore not described anywhere else.*

50 *We added the information about first order kinetics.*

Based on the description I am not sure, if the model considers really different temperature sensitivities for fungi and bacteria (which would surprise me) and where the data about community size are coming from. I assume the authors mean that the SOC module contains a parameter that controls the impact of temperature on the decomposition rates and this factor was calibrated and tested for fungi and bacteria dominated soils. For the here presented study this doesn't matter.

*There is no difference between bacteria and fungi in the used setup of the model. The bacteria and fungi in the text refers to the applicability of the Ratkowsky function that was used to describe the temperature dependence. This function was originally developed for bacteria, we mentioned the previous application to fungi by others, to justify why we applied it on a peatland, where fungal decomposition plays an important role. Other peatland models and previous CoupModel applications often use a more simple Q10 approach to describe temperature dependence.*

*In the revised manuscript we reformulated this sentence and moved the description of the temperature dependence more to the top of the section.*

Section 2.4.3 A couple of problems with the NEE values could be sorted by using the correction approach by Papale et al., 2006 (Biogeosciences, 3, 571–583). This would enable to solve the problems with extreme day values and the peaks for the night periods.

*We did not apply a filter for friction velocity or any spike removal as suggested in Papale et al 2006 for following reasons: We did not see any effect of friction velocity on the fluxes which is likely due the EC measurements being conducted at 2m above an open mire surface. Furthermore, friction velocity filtering is only valid if the turbulent transport and biological sources of measured fluxes are coupled. During nighttime, however, biological activity might continue while the turbulent transport is absent leading to accumulation of e.g. CO<sub>2</sub>. The accumulated concentrations might be released and detected as 'spikes' in the morning when turbulent movement sets in. Removing these spikes would therefore introduce an error in the C budget (i.e. in the emission component). In addition, inherent noise in EC data leads to occasional spikes which are presumably randomly and evenly distributed around the mean. Selectively removing spikes might introduce artificial and subjective bias into the flux balance. We have clarified and rephrased the relevant text in section 2.4.2.*

I also wondering, if the gap filling tool, develop by Reichstein and Falge, is used to fill gaps for NEE? *Although gapfilled data was available based on the Reichstein et al 2005 approach, this study only used the measured NEE, H and LE fluxes and omitted all gapfilled periods, as stated in section 2.2. The model should be calibrated with measured data, not with another model.*

### Comments on the result section

Page12 lines 11-14 I understand that the soil water content is an important variable, which is difficult to measure and to simulate. This is not new and as this is known, this should be a central part of a sensitivity analysis. I think it is not enough to ask for more measurements, which is always a good answer to all problems with simulations. First, I miss a discussion of the measurements of the soil water content, which is often done on a single spot rather than spatial distributed or in different depths. Second, there is no discussion of the footprint area of the EC measurements. If the footprint changes and the soil type or hydraulic properties differ on the test site, this might explain differences.

*The referee is correct in that continuous measurements of the water table depth is conducted at just one spot and spatially distributed measurements would give a measure of the variation. However, the mire surface within the entire footprint is totally (100%) covered by Sphagnum mosses. The most important functional trait separating Sphagnum mosses into different functional groups is the architecture of the plant determining both the capillary forces as well as the water holding capacity and thus at what distance to the water table the different Sphagnum species grow. The plant community distribution within the foot print areas (see below) is very homogenous and totally dominated by Sphagnum species (see below) reflecting a growing season average water table of ~5-15 cm below the moss surface. Thus, even if we have conducted continuous measurements of both water table at one spot and soil water*

content at a few spots the dominating *Sphagnum* species composition within the footprint clearly reflects a spatially average water table equal to the measurement spot.

The position of the EC tower is in the center of a mire unit totally dominated by lawns, i.e. the growing season average water table varies between ~8-15 cm below the mire surface (see e.g. Sagerfors et al 2008). The lawn plant communities have a close to 100% cover of *Sphagnum* mosses (*Sphagnum balticum*, *Sphagnum majus* and *Sphagnum lindbergii*) and a limited contribution of vascular plants, totally dominated by the sedges *Eriophorum vaginatum* and *Trichophorum cespitosum* and the dwarf shrubs *Vaccinium oxycoccus* and *Andromeda polifolia*.

Both the day time and night time foot prints are well within this very homogenous lawn dominated unit of the mire (see Sagerfors et al 2008). The footprint areas are most narrow with daily average 90%-tile boundaries <<50m radius to the tower with most limited seasonal variation (seasonal footprint modelled by Kljun, unpublished).

The need for measuring water content on several spots and in several depths is already mentioned in the manuscript: "Thereby, the horizontal and vertical variability in peat hydraulic properties needs to be accounted for (Baird et al., 2012, Waddington et al., 2015)."

**We added the footprint area problematic to the discussion: "measured NEE is not the CO2 exchange between biosphere and atmosphere at a certain point, but is a calculation based on turbulent vertical fluxes measured several meters above the ground and resulting from a diurnal and seasonally changing area that includes different soil conditions and vegetation."**

Third, as the authors make a sensitivity analysis, it is possible to detect the most sensitive soil property and give at least the advice, which soil property should be measured to get better results with the CoupModel.

We advise to measure the water retention curve. This is mentioned in the text. As we calibrated only one parameter of this curve to avoid equifinalities within the same equation, the sensitivity to this parameter represents the sensitivity to the result of the equation. **This was added in the revised version to the discussion section 4.5.**

### 3.1 Parameter sensitivity:

I do not understand why the authors highlight the module dependency so strong. This analysis makes the study extremely model dependent. I think the authors should relate the sensitivity to processes. I assume that the modules represent separate processes, but this is not necessarily the case.

**Processes is an ambiguous term, as it can refer to a single equation, or a set of several equations. We used module when we were talking about a process described by a set of several equations. But this seems to be ambiguous as well. Therefore, we replaced it in the revised version by "category of processes" which we define in the beginning of the manuscript.**

Page13 lines 27-29 R2 and ME are contradicting in their goodness of fit: Is this an indication that these are not the best indicators to detect the quality of performance?

**They measure the performance in different ways: R2 measures the performance in the dynamics, whereas ME shows how well the magnitude was simulated. When they constrain a parameter to different value ranges, it means that there is no value that can produce a perfect fit in both, dynamics and magnitude. That's what we wanted to show: parameter ranges that are constrained by calibration might depend on the performance index that was chosen for calibration. R2 and ME are simple, but sufficient to show this. Of course we could compare the resulting parameter ranges with further other indices - and would get other resulting ranges. Taking only one index for calibration will give one resulting range, but does not tell the user, if there were shortcomings in either magnitude or dynamics, or something else. Reasons for the mismatching ranges is not a bad performance index but the limitation of the model to produce a perfect fit of the model output to the measured values simultaneously in both magnitude and dynamics in the certain variable. Models are always a simplification, not perfect and include parameters for which a perfect value is not existing.**

*The response to this comment is added in the revised version of the manuscript.*

Section 3.4 Usefulness might be not a good word to describe the measured variables.

*Translation error. Replaced by potential.*

5

**Comments on the discussion:**

Wide parts of the discussion are not really a discussion, but do only compare qualitative findings of the study with other studies.

10 4.1 Parameter sensitivity

It is correct that the detection of sensitivity of parameters enable to concentrate the calibration on the main drivers, but how robust are the findings on this test site and how transferable are the results to other ecosystems or to other climate zones? Peatland in Northern Europe is a quite specific test site, so, is it possible to transfer the results to mineral soils? How transferable are the results to Central Europe or to the Mediterranean area? It is no problem, if the results are not transferable, but at least there need to be a discussion.

15

*We tested only one model on one site, therefore we cannot name which of the most sensitive parameters, parameter ranges, interactions, etc. might be transferrable and also not to what extent/to which other ecosystems or models. The only indication we have, is when comparing with other studies: as mentioned in the manuscript, some parameters that we identified as most sensitive that were also among the most sensitive in studies on other ecosystem, using other models.*

20

Page 19 lines 3-5: “While the existence of interactions between the processes and their parameters is supposed to be less dependent on site conditions and model structure, the exact shape of the connections as well as constraint parameter ranges might strongly depend on these factors. “ This might be correct as the sensitivity analysis only represents effects of the model structure. However, by applying the analysis on a specific test site, the relevance of processes depends on the climate zone, ecosystem, land use, soil type, etc. This also effects the limitations for the data range of the considered parameters and variables. The relation and interaction might be different outside this range. Therefore, I wouldn’t exclude the site conditions as relevant factors.

25

30

*We fully agree, but that's more or less what we are saying. We didn't mention the relevance, but added it in the revised version as site and model dependent finding.*

35

Page19 lines 14-16: It depends: Several models using the same approaches to describe processes. Therefore, the formulated hypothesis needs to be tested by compare the approaches used in the different models to be sure, that this correlations are really independent of the model structure.

40

*Models often use same or similar equations, but the combination of equations, which processes are simulated and which replaced by a constant value, the number and type of parameters calibrated together and used variables for calibration differ between the studies. A detail presentation of all differences is outside the scope of this study.*

45

Page 19 line 27 to page 20 line 2: I do not really understand how the implementation of open water bodies should explain the differences in the correlations. In the measurements H is more related to temperature and LE more to the water flows. Photosynthesis is the main driver for growth and photosynthesis is calculated by a light use efficiency function and, as written in this manuscript “...total plant growth is proportional to the net global radiation absorbed.....”. Is it possible that the correlation of H and NEE can be explained by the calculation of photosynthesis by radiation, which is also the main driver for H, while LE is calculated in more complex equations with less direct correlation to radiation and temperature?

50

*It is not H and NEE that correlate, but the parameter values that lead to a good fit in both. As we mention in ln 31 the same page, we tested only the effect of parameters, not the effect of input variables (like the sensitivity to radiation), which would be an interesting study as well.*



Open water bodies is just an example for missing processes. The fit for LE is not good in spring, whereas this pattern cannot be seen in NEE. In the real world, there might be a lot of evaporation from open water bodies, so the model underestimates LE in spring - this could be compensated with parameters that lead to a higher plant transpiration (=> better fit in LE), but these parameters would also lead to an overestimation of NEE in spring (=> worse fit in NEE). **We reformulated the sentence in the revised manuscript to make it more clear.**

Page 20 lines 3-5: No, not necessarily. If you try to understand the pattern of data in advance, the used indicator for the goodness of fit can be picked sensible. E.g. there are variables with several values (e.g. night values) at zero or around zero. These values will have a strong impact on the ME as the models, usually, simulate the zero values during night quite well. The R2 can cope with the clouds around zero, but it is sensitive to single extreme values.

To reduce the effect of extreme values, we had additionally the R2 of accumulated values. As stated before, there are many more complex indices, and they would probably result in different parameter ranges - this only supports our statement: the choice of the index has an effect on the resulting range. Values around zero do not have a strong impact on ME, as the modeled values during this periods are also low. That's why we decided to add ME of winter values - values are low and if you only look to the whole year, parameters that influence winter fluxes have no/low sensitivity. In case of NEE (where we differentiate between day and night values), the night values are dominated by respiration, whereas during daytime photosynthesis plays an important role - therefore it is not surprising, that different parameter and parameter ranges lead to the best fit for either day or night. This cannot be solved by a more sophisticated performance index - the underlying problem is, that the model is not able to give simultaneously a very good fit in daytime, nighttime, as well as in magnitude and dynamics - it remains a decision of the user to calibrate to NEE only, or separately to night and daytime values and to decide if a good model result in magnitude or in dynamics is more important. Same for the seasonality - if none of the runs shows the best fit in both spring and summer, it is not a question of another performance index - instead it hints to limitations in the model, e.g. a process that is not implemented or at least not included in the calibration. But models are never perfect, therefore a best value or value range is not existing for many parameters. **We added some discussion on this in section 4.2 in the revised version**

Bottom line the used indicator for goodness of fit influences the outcome of the analysis and if the indicator is well picked, there are subjective judgements. Controversial results of different indicators need to be analysed to understand the reasons for the contradiction. Unfortunately, this analysis is missing in this manuscript.

The most pronounced controversial results are analysed in the subsections of 4.5., but a detailed analysis for each parameter and each variable would be extremely extensive and outside the scope of the manuscript.

#### **Possible reasons for controversial results, which we added now to section 4.2.:**

- Most important: The model does not reflect the real world (e.g. decomposition rate coefficient is not a constant value, but depends on the activity of soil microbes which is influenced by many factors that vary in time, e.g. community structure, community size, stress factors, food availability and quality, etc). A parameter with very high discrepancy between performance indices is the aerodynamic resistance dependency on LAI  $r_{alal}$ . For a good magnitude of temperature, this value has to be extraordinary high - much higher than a value that was actually measured at this site, see discussion to this in 4.5.3.

- Measurements do not reflect the real world. Measurements have limitations, e.g. NEE is not the real exchange between Eddy fetch - not a point like the model, but instead an area, and the area changes during the day and during the year. Also, not the CO2 exchange between biosphere and atmosphere is measured, but turbulent vertical fluxes at the sensor (several meters above the ground), which further include a lot of calculations to receive NEE.

- Indicator not good (this is the case for snow - timing of snow melt is most important, but not well enough reflected in R2 and not at all in ME, see section 4.5.4)

Page 20 Lines 6 -15: Of course there are lot's of correlation between LAI and other variables, because these parameters use LAI. However, an analysis and discussion of the cited publications is missing. This would be a chance to bring the here presented study in the context of other studies. Instead of only mention the correlation, the authors could explain the different dependencies. E.g. I assume that LAI

5

*Also on peatlands, LAI correlates with water content due to transpiration. Such dependencies are nicely described in Schulze 2006. How they are realised by the different equations that have LAI as parameter is described in the supplementary material and the CoupModel description. To all three references, we refer in the manuscript, page 20, ln 13-15: "These relationships can be explained by the many*

10

*dependencies between LAI and e.g. photosynthesis, transpiration, heat insulation and water uptake (Schulze 2006), of which several are also implemented in the model (see model description and equations, Sect. 2.3, Table S2 in the supplement and Jansson and Karlberg, 2010)."*

Page 20 line 17 temporal or spatial resolution? What means high resolution mm, cm or m; seconds, hours, days?

15

*It refers to temporal, which we added in the manuscript. "High" depends on what one is interested in. We worked with hourly values, so that it is sufficient to measure a time series of hourly measurements in one layer (for simulating the dynamics) plus -for the magnitude - theoretically one single measurement in the upper and on in the lower layer.*

20

Page 23 Lines 10 – 15: I see the strong sensitivity of the soil hydraulic properties as relevant factor, but first, it is not that easy to measure these parameters and, second, I think the authors should provide an alternative method to derive better fits and quantify the reduction of quality by missing out soil hydraulic properties. An alternative method would be to calculate the soil hydraulic properties by pedo-transfer-functions (as mentioned in the model description). If do so, the sensitivity of single parameters (soil type, bulk density, field capacity (by itself) etc.) can be tested and it might be possible to get better calibration using this information or detect the most sensitive of these parameters.

25

*This doesn't work on peatlands, only on mineral soil.*

30

### **Comments on the references**

The publication of He et al. needs to be updated

**Done**

### **Comments on figures:**

35

- I would like to see a figure like Fig.5 also for actual values and not only for a prior and posterior comparison.

*Plotting the dependencies between different output variables would require many dimensions, as they are all connected between each other and also depend on the different parameter sets. There is an enormous amount of combinations, which makes it not visualisable.*

40

- The quality of the figures is not good

**Figures will be uploaded with higher resolution in the revised version.**

### **Comments on the supplement:**

45

Table S1 I think there is no need to present parameter name in the model.

*This information would be very helpful for other CoupModel users, as these are the names given in the user interface.*

I am even not sure if the module name provides any useful or needed information, but it might be better to group the parameters instead (e.g. soil, hydrology, snow, vegetation/growth).

50

*The parameters are sorted for the module which gives shortly in which calculation the parameter is used. Of course this could be also read from the equation number, but the text is easier to read.*

Table S2 is really needed, if you develop a model and publish it, but I do not see the use for the actual study. Most of the equations are standard approaches that are already described in the text.

5 *As mentioned before, there are many possibilities to configure CoupModel. Therefore the used equations vary between studies, and in some cases also the terms within an equation are modified, deepening on switches and parameters that might set a term to zero. It further shows where the specific parameters are used in an equation*

## Comments from Anonymous Referee #2

10 Received and published: 5 August 2016

Metzger et al present an interesting study addressing process interactions and parameter sensitivity for model carbon dynamics in a natural peatland. This is a “heavy” topic and the authors did a good job. Their findings are important and meaningful for both model users and model developer, the latter of whom they overlooked. There are some aspects needs substantial revision.

15 a) There are too many small paragraphs with only one or two sentences. I would suggest the authors to combine them.

**OK**

20 b) The authors claimed “interactions between parameters” “limited transferability of parameter values between models and even between studies”. I am not quite understand the connections between the two topics. It could be great if the authors can elaborate more on this.

25 *If parameters interact, the value range resulting from a calibration depends on the values of other parameters. This demonstrate that parameters are not independent. Therefore, one cannot transfer the information obtained from a single parameter without also considering the value of the correlated parameters. The correlation obtained may be a phenomena that is related to a possible coexistence for this particular ecosystem, But it can also be because of the problem to constrain the model by not enough of data. Note that all parameters for the posterior distribution are uncertain and we do not expect to find a narrow range for single parameters since also the real world system is expected to have a range of parameters that represent the certain temporal and spatial variability of the system considered*

30 *We reformulated this (in the first part of the discussion) to be more clear.*

c) The authors mentioned many times of “CO2 model(s)”, which seems improper because the CoupModel is more like a C cycling model, rather than CO2 model.

35 *We agree that the study is not with emphasize on CO2 instead it will try to understand the full carbon turnover at the specific site. However, the use of NEE from flux measurements is of course the major response to all the ongoing processes and fluxes of the ecosystem.*

*We changed the title to include heat and water fluxes.*

40 d) This work is not only meaningful for model users, but also for model developers. Nowadays, for example, many researchers develop and use models to predict impacts of climate change on carbon cycling or hydrology, and others. However, many of these models are not integrated or balanced enough representing all aspects (processes/modules). Such model predictions lack of credit for me. I could suggest the authors also discuss this aspect in the discussion section. Overall, I think the paper is publishable after major revision. Some specific comments are:

45 *With "modellers" we mean not only model users but also model developers. We included them more explicitly at several points in the revised manuscript.*

50 1) Line 9-10: From my understanding, most previous models focused only one or few modules because their model emphasized only on these module(s) and simplified (overlook) others. Interestingly, this could highlights the importance of the present study. The authors may want to elaborate this point more.

*Models are always a simplification, and even that we show the interactions between the different modules, we would not like to devalue simpler models - it always depends on how accurate the model prediction need to be.*

5 *We included the importance of considering the different processes together in the last sentence of the revised abstract.*

2) Line 13: Please specify the modules to make the reader to easy understand.

**OK**

10 3) Line 20: This sentence is hard to understand. Please revise.

**OK**

4) The introduction contains too many paragraphs and they are not very well logically connected. Please consider to reduce them into 4-5 paragraphs.

15 *We reordered the paragraphs in the introduction in a more logical order and combined them, including some reformulations.*

5) Line 28: I think these findings will be of critical importance for model development as well.

*We added the model development at several places in the manuscript.*

20

6) Line 1 in Page 9: What do you mean of “uniform random distribution”?

*The values are randomly taken, whereas all values in the range have the same probability of being used - this is added in the revised version*

25 7) Line 9 in page 9: Has this definition of sensitivity been used by others?

*There are several possibilities to quantify sensitivity. Most common are measures of the difference between prior and posterior parameter distribution. As we use a simple uniform distribution, it is not necessary to use sophisticated methods like Kolmogorov D statistic or stepwise regression analysis. The simplest way is to just compare the range of posterior and prior distribution. This has certainly be done in one or another way by other studies as well. In contrast to the R2 value between parameter values and performance, this accounts also for parameters that have an optimum range around in the centre of the prior distribution.*

35 8) Line 21 in page 9: Please explain clearer how the equifinalities was quantified.

*Reformulated to: "Equifinalities were quantified by the R2 value of a simple linear regression through the values of the interacting parameter pair in the accepted runs."*

Figures quality/resolution are low. It is hard to read these figures

40 *Higher resolution will be provided in revised version.*

### **List of relevant changes**

45 In all parts of the manuscript we connected paragraphs to larger ones, replaced some translation errors and reformulated sentences that were misunderstood to make them more clear. The term "modules" was replaced by "process categories", which we defined with its first occurrence. Some important sentences that were overlooked by a referee were moved further to the top of a section. In several parts of the manuscript, we included model developers more explicitly.

50 **Title:** The name and version of the model are added to the title and heat and water fluxes mentioned additionally to the carbon fluxes.



**Abstract:** We listed the different modules (now process categories), highlight the interactions across them and reformulated a misunderstood sentence.

5 **Introduction:** We reordered the paragraphs more logically and connected them. We mention the Bayesian approach

10 **Objectives:** We reformulated the objectives to make them more precise and to emphasize the understanding of the dependencies between parameter distributions and between parameters and model performance. We added the method that was applied to fulfill the objective and reformulated objective 2 to make it clearer.

15 **Methods:** The information was added that CoupModel offers many different configuration possibilities was added to the methods, to clarify that the model description is necessary to make the study reproducible. The model description for vegetation and decomposition was reformulated and shortened by moving parts to the supplementary material.

Page 4 line 20: "EC" was added and Page 4 line 28-29: C uptake by the ecosystem from the atmosphere was added.

20 We clarified that input data was gap filled, whereas calibration data was not gap filled and spikes were not removed to avoid introducing a bias.

**Results:**

25 We highlighted the found dependencies across process categories. A sentence was added, telling that using several measurement variables helps to identify if a parameter range is not robust. The results about controversial parameter ranges were moved from the sensitivity section to the section about confounding and supporting effects.

**Discussion:**

Also in the discussion we reformulated some parts to emphasize the importance of the identified dependencies across modules.

30 We added a sentence to clarify that sensitivity to a parameter means sensitivity to a process, as only one parameter per equation was calibrated.

35 We added at several positions in the discussion the information about whether a certain result relates to CoupModel and site conditions or can be used more general. We further mention that we tested only one model on one site, so that our statements about robustness refers only to robustness in respect to criteria selection, but not to site or model dependency.

We added a more detailed discussion about the possible reasons for the contradicting parameter ranges to section 4.2 .

We mentioned the Bayesian approach and added the EC footprint area problematic to the discussion.

40 **Conclusions:**

Also in the discussion and conclusions we reformulated some parts to emphasize the importance of using many variables and criteria to constrain the model.

**Figures:**

45 The figures are now provided in a higher resolution.

**References:**

Publication of He et al. was updated and references that belong to the sections of the model description that were moved to the supplement were removed.

50

## Marked up manuscript

# ~~The importance of process~~ Parameter interactions and ~~parameter~~ sensitivity analysis for modelling ~~the~~ carbon heat and water fluxes dynamics in a natural peatland, using CoupModel v5

Christine Metzger<sup>1\*</sup>, Mats B. Nilsson<sup>2</sup>, Matthias Peichl<sup>2</sup> and Per-Erik Jansson<sup>1</sup>

<sup>1</sup>Department of Land and Water Resources Engineering, Royal Institute of Technology, Stockholm, 100 44, Sweden

<sup>2</sup>Department of Forest Ecology & Management, Swedish University of Agricultural Sciences, Umeå, 901 83, Sweden

\*now at: Institute for Meteorology and Climate Research/Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute for Technology, Garmisch-Partenkirchen, 82467, Germany

Correspondence to: Christine Metzger (cmetzger@kth.se)

**Abstract.** In contrast to previous peatland carbon dioxide (CO<sub>2</sub>) model sensitivity analyses, usually focusing on only one or few ~~modules~~ processes, this study investigates interactions between various biotic and abiotic processes and their parameters by comparing CoupModel v5 results with multiple observation variables.

Many interactions were found not only within, but also between ~~the~~ various model modules process categories simulating plant growth, decomposition, radiation interception, soil temperature, aerodynamic resistance, transpiration, soil hydrology, and snow. Each measurement variable was sensitive to up to ten (out of 54) parameters, from up to seven different process es-categories. The constrained parameter ranges varied, depending on the variable and performance index chosen as criteria, and on other calibrated parameters (equifinalities).

Therefore, transferring parameter ranges between models needs to be done with caution, especially if such ranges were achieved by considering only few processes. The identified interactions and constrained parameters will be of high interest to use for comparisons with model results and data from similar ecosystems. All of the available measurement variables (net ecosystem exchange, leaf area index, sensible and latent heat fluxes, net radiation, soil temperatures, water table depth and snow depth) improved model constraint. ~~Additional measurements of soil hydraulic properties or water content would reduce equifinalities and constrain additional parameters that showed high range of uncertainty. If hydraulic properties or water content were measured, further parameters could be constrained, resolving several equifinalities and reducing model uncertainty.~~

The presented results highlight the importance of considering biotic and abiotic processes together and can help modellers and experimentalists to design and calibrate models ~~model calibrations and~~ as well as to direct experimental setups in peatland ecosystems towards modelling needson peatlands.

**Keywords.** Parameter uncertainty, equifinalities, net ecosystem exchange (NEE), carbon dioxide (CO<sub>2</sub>), boreal mire

## 1 Introduction

Understanding and quantification of interactions between different processes and between different parameters is required for reducing uncertainty in prognostic modelling in carbon (C) cycle research. Undisturbed peatlands act as carbon sink and

have accumulated at least 550 Gt of C, which is equivalent to twice the C stock in the forest biomass of the world (Gorham, 1991; Parish, 2008). A more recent estimate for exclusively northern peatlands amounts to 436 Gt of C (Loisel et al., 2014). Management or climate change can cause this carbon to be released as CO<sub>2</sub> emissions as has been shown from measurements (Maljanen et al., 2010; Drösler et al., 2013; Petrescu et al., 2015). Process oriented models are necessary to transfer the knowledge gained from measurements to different locations, management or future climate scenarios. Further, such models can help to understand the processes underlying the observations. But oOnly few of the parameters used in process models are known as site independent, unambiguous constants from laboratory experiments. All others need to be either assumed, or gained from calibration procedures (e.g. Kennedy and O'Hagan, 2001, Wang and Chen 2012). But not all parameters have a strong impact on model output and performance (i.e. fit between modelled and measured variables, whereas in this manuscript, variable always refers to a time series that is either the output of the model or the measurement to which the model output is compared to). Monte Carlo based sensitivity analyses are used to identify key parameters for both, the performance and the impact on various major model outputs (e.g. Verbeeck et al., 2006; Van Oijen et al., 2011; Santaren et al., 2014).

Many studies investigated single processes and their parameters, while only few consider different biotic and abiotic processes and multiple calibration variables. Several modelling studies have explored peatland hydrology (e.g. Dimitrov et al., 2010; Dettmann et al., 2014) and heat fluxes in peatlands (e.g. Granberg et al., 1999; Keller et al., 2004), while others concentrate on carbon fluxes and pools (e.g. Frolking et al., 2002; Verbeeck et al., 2006; Wu et al., 2013) where the focus is sometimes on heterotrophic respiration only (e.g. Abdalla et al., 2014). However, many processes are involved in the C-cycle of peatlands: Net ecosystem exchange (NEE) is the balance of photosynthesis, and autotrophic respiration from plants as well as heterotrophic respiration from nd-microbes. All three-NEE component fluxes are strongly interconnected in several ways with the amount of plant biomass, temperature, radiation, nutrients and moisture availability (e.g. Clymo, 1984; Lindroth et al., 2007). Photosynthesis, soil temperature (Ts) and moisture depend among others on incoming radiation, transpiration and plant coverage. Heterotrophic respiration further depends on quality and quantity of plant litter (e.g. Yeloff and Mauquoy, 2006). In addition, phenological events such as the timing of snow melt are important for soil temperature dynamics, biologic activity and peatland CO<sub>2</sub> fluxes (Aurela, 2004; Peichl et al., 2015). Different biotic and abiotic processes are realised in some modelling studies on peatlands, though, only the sensitivity to carbon fluxes or pools was tested (e.g. Yurova et al., 2007; St. Hilarire et al., 2010; Quillet et al., 2013; Webster et al., 2013; Wu and Blodau, 2013; Kim et al., 2014). Also, models are continuously extended or coupled with other models (e.g. Wang et al., 2005; Prentice et al., 2007; Giltrap et al., 2009; Hidy et al., 2012; Jansson 2012; Tang et al., 2015), developing to more and more holistic models, accounting for plant and soil carbon processes, water and energy flows and biochemistry. However, often only parameters of the new module are tested (e.g. Belassen et al., 2010; Wania et al., 2010, Zhu et al., 2014; Tang et al., 2015), ignoring possible interactions between processes.

~~Models are continuously extended or coupled with other models (e.g. Wang et al., 2005; Prentice et al., 2007; Giltrap et al., 2009; Hidy et al., 2012; Jansson 2012; Tang et al., 2015), developing to more and more holistic models, accounting for plant and soil carbon processes, water and energy flows and biochemistry. However, often only parameters of the new module are tested (e.g. Belassen et al., 2010; Wania et al., 2010, Zhu et al., 2014; Tang et al., 2015), ignoring possible interactions between modules.~~

~~Many peatland studies investigate only the sensitivity to carbon fluxes or pools, despite their models include different biotic and abiotic processes (e.g. Yurova et al., 2007; St. Hilarire et al., 2010; Quillet et al., 2013; Webster et al., 2013; Wu and Blodau, 2013; Kim et al., 2014). Though, the profit of using multiple constraints for model calibration and the importance of~~

~~interactions between parameters and across modules has been shown by e.g. sensitivity analyses on forest ecosystems (Carvalhais et al., 2010; Santaren et al., 2014; Tian et al., 2014).~~

~~Further~~Another limitation of previous peatland modelling studies is the use of, often only local sensitivity analyses, ~~are performed,~~ changing only one parameter or one input ~~variable-driver~~ at a time (e.g. Hilbert et al., 2000; Yu et al., 2001; Zhang et al., 2002; Wania et al., 2009; Frolking et al., 2010; Tang et al., 2010; St-Hilaire et al., 2010). This approach does not account for possible interactions and non-linearity in equations (e.g. Saltelli et al., 2008; Quillet et al., 2013), but peatland processes are often non-linear and interact in many ways (Belyea, 2009). Therefore, we performed a global sensitivity analysis, calibrating parameter simultaneously and accounting for interactions.

~~Net ecosystem exchange is the balance of photosynthesis, and respiration from plants and microbes. All three NEE components are strongly interconnected in several ways with the amount of plant biomass, temperature, radiation, nutrients and moisture availability (e.g. Clymo, 1984; Lindroth et al., 2007). Photosynthesis, soil temperature (Ts) and moisture depend among others on incoming radiation, transpiration and plant coverage. Heterotrophic respiration further depends on quality and quantity of plant litter (e.g. Yeloff and Mauquoy, 2006). In addition, phenological events such as the timing of snow melt are important for soil temperature dynamics, biologic activity and peatland CO<sub>2</sub> fluxes (Aurela, 2004; Peichl et al., 2015).~~

~~Such processes interactions are realised in complex ecosystem models, but lead~~This allows to inter-correlation between the different parameters ~~and-which~~ complicates the parameter constraint to an unambiguous solution: several combinations of different parameter values can lead to a similar good fit of model output to measured variables, which is defined as equifinality (Beven and Freer, 2001). The model sensitivity to such parameters might be hidden if equifinalities are not considered. Constraining a model based on multiple observation variables can help to resolve or reduce equifinalities (Carvalhais et al., 2010). The profit of using multiple constraints for model calibration and the importance of interactions between parameters and across different processes has been shown by sensitivity analyses on e.g. forest ecosystems (Carvalhais et al., 2010; Santaren et al., 2014; Tian et al., 2014). Unlike previous ~~peatland-modelling studies on peatlands,~~ we therefore investigate the sensitivity to parameters from several different modules simultaneously, in their effect on not only on NEE, but also on LE, sensible heat (H), net radiation (Rn), leaf area index (LAI), Ts, WT and snow, and identify parameter interactions.

However, criteria based on multiple variables imply a subjective weighting of variables and performance indices. Fitting the model to a certain variable might improve or worsen the performance in another variable (Carvalhais et al., 2010) and might therefore have implications for the parameter range judged as valid (e.g. Schulz and Beven, 2003). In this study, the effects of selecting a certain criteria on the resulting parameter range will be investigated. We avoided to use a Bayesian approach, which was tested by Van Oijen et al. (2011) with several models including the CoupModel, using a data set of more than one variable. The single probability of the model as the summation of many different variables requires a detailed understanding of an error model that is typically not available in field measurements representing many different errors for each set of variables.

~~We use the~~The detailed ecosystem model CoupModel (Jansson and Karlberg, 2010) was used in this study for the following reasons: It is a well-established and widely used model (Jansson, 2012). Its ~~model~~ structure is flexible and allows simulation of different abiotic and biotic processes based on well-established physical equations, which can be selected by the user. The CoupModel includes all main components expected to have an impact on the carbon cycle: i) A detailed module for simulation of heat and water fluxes in the soil and at the interface to the atmosphere, ii) plant growth from photosynthesis, limited by water availability and temperature, iii) plant respiration and litter fall and iv) a module for soil organic carbon (SOC) decomposition. A user defined time step allows using the full information contained in measurements with high temporal resolution (i.e. hourly) on site scale.



## 1.1 Objectives

The aim was to identify and explore the connections within and between biotic and abiotic processes and parameters which are relevant for modelling NEE in a natural open peatland. Therefore, 54 parameters of the CoupModel v5 from different plant, decomposition, energy and water flux processes were calibrated to, by investigating several different output variables and several different sets of criteria for selecting acceptable runs were tested. The specific objectives were:

1. To identify which processes impact which measured variable, by testing the sensitivity of model performance to the parameters of the different processes
2. To evaluate ~~implications of different criteria selection choices on the dependence of~~ model performance and resulting parameter ranges on the performance index, the measured variable and the time period of the variable, that are chosen as criteria
3. To identify and describe equifinalities between parameters from different processes simulating carbon, energy and water fluxes
4. To test the ~~usability-potential~~ of all available observation data for model constrain and identify missing measurement variables by identifying sensitive or interacting parameters that cannot be constrained by the available data

The answers to these questions will be crucial for ~~future-model~~ development and future calibrations of carbon models on peatlands similar ecosystems: They will represent most valuable information for selecting processes that need to be taken into account, for selecting parameters and their value ranges and considering parameter connections, as well as selecting sites and observed variables. They further help experimentalists to decide on the measurement of variables to make their site suitable for modelling.

## 2 Materials and methods

### 2.1 Site description

Degerö Stormyr (64.182016 N, 19.55663 E) is an oligotrophic, minerogenic, mire, located on a highland, 270 m.a.s.l, in the county of Västerbotten, Sweden. A detailed description of the site and the measurements can be found in Peichl et al. (2014) and references therein. “The mire catchment is predominantly drained by the small creek Vargstugbäcken towards north-west. The depth of the peat is generally between 3–4 m, but depths up to 8 m have been measured. ... The micro-topography is dominated by mainly carpets and lawns, with only sparse occurrences of hummocks” (Peichl et al., 2014). The plant community of the mire is dominated by cottongrass (*Eriophorum vaginatum* L), tufted bulrush (*Trichophorum cespitosum* L. Hartm.) and *Sphagnum* mosses (Nilsson et al., 2008; Laine et al., 2012). Total aboveground biomass (moss capitula and vascular plants) is  $141 \pm 45 \text{ g m}^{-2}$  (Laine et al., 2012). Seasonal maximum leaf area index of vascular plants was estimated at  $0.8 \text{ m}^2 \text{ m}^{-2}$  in 2012 (Peichl et al., 2015).

The 30-year (1961–1990) mean annual precipitation and air temperature are 523 mm and  $+1.2^\circ\text{C}$ , respectively, while the mean air temperatures in July and January are  $+14.7^\circ\text{C}$  and  $-12.4^\circ\text{C}$ , respectively (Alexandersson et al., 1991). The snow cover normally reaches a depth of up to 0.6 m and lasts for approximately 6 months (Peichl et al., 2014). The peatland was continuously a sink for atmospheric  $\text{CO}_2$  during twelve years of eEddy covariance (EC) measurements, with a 12-year average ( $\pm$  standard deviation) NEE of  $-58 \pm 21 \text{ g C m}^{-2} \text{ yr}^{-1}$  (Peichl et al., 2014).

### 2.2 Data used in this study

Hourly values of global radiation, air temperature, relative humidity, precipitation and wind speed were used as meteorological input data (Table 1). They were measured at the same tower as-to which the EC sensors were mounted. ~~An~~

~~overview of the data used for calibration can be found in Table 2; a more detailed description in Peichl et al. (2014) and references therein.~~ For gap filling (due to instrument failure) of the input data, as well as for the pre-evaluation period 1991-2000, daily data from the nearby (13 km away) standard climate station at the Svartberget field station were obtained. In case of air temperature and relative humidity, seasonal regression relationships were applied to account for temperature and humidity differences between the site and the standard climate station.

An overview of the data used for calibration can be found in Table 2, a more detailed description is provided by Peichl et al. (2014) and references therein. Measured carbon concentrations per soil layer were used for estimation of pool sizes as described in Sect. 2.3.5. The model was calibrated based on measured NEE, LE, H, WT, Rn, soil temperatures in -2 cm ( $T_{s1}$ ) and -42 cm ( $T_{s2}$ ) depth, snow depth and LAI of vascular plants ~~as listed in (Table 2).~~ NEE, LE and H were measured using the eddy covariance technique and details for data processing were previously described in Peichl et al. (2014). In this study, only the measured values of NEE, LE and H were used for calibration (i.e. gapfilled values were omitted). ~~and described in Peichl et al. (2014), and references therein.~~ Negative NEE values indicate net CO<sub>2</sub> uptake by the ecosystem from the atmosphere while positive NEE values indicate emission from the ecosystem to the atmosphere of CO<sub>2</sub>. All calibration data were averaged to hourly values, except snow depth and LAI values ~~and snow depth which had a daily and biweekly to monthly resolution. In this study, only measured values were used for calibration: gap filled values during measurement gaps were omitted.~~

### 2.3 Model description and application to Degerö Stormyr

CoupModel v5 from 12<sup>th</sup> December 2014 was used for simulations. The current version can be downloaded from the CoupModel homepage (CoupModel, 2015). A detailed description can be found in Jansson and Karlberg (2010). The CoupModel allows the user to select between different sub models, different equations and different complexities of the used equations. The following sections describe the configuration as applied in this study. The model represents the ecosystem by a description of C and N fluxes in the soil and in the plants. It includes the main abiotic fluxes, such as soil heat and water fluxes that represent the major drivers for regulation of the biological components of the ecosystem. For application to Degerö Stormyr, the vegetation canopy was defined as two layers: vascular plants and mosses. The soil profile was divided into 16 layers with an increasing layer depth from 4 cm for the upper nine layers to 60 cm in the lowest layer, resulting in a total depth of 3.4 m. The model internal time step was half-hourly for abiotic processes and hourly for nitrogen and carbon related processes. The simulations were started ten years prior to the evaluation period, so the system could adapt to the site conditions and become more independent of initial values.

The most important equations and the corresponding calibrated parameters can be found in Table S1 and S2 in the supplement. The major model assumptions relating to the model application to the peatland are described below. Detailed assumptions in respect to fixed parameter values can be found in Table S3 in the supplement.

#### 2.3.1 Radiation interception, evapotranspiration and snow

An interception model for both, radiation and precipitation, a snow model and a surface pool model was used to provide boundary conditions at the soil surface. Cloud fraction was calculated from global radiation input and latitude. Incoming radiation was partitioned between one part, which was absorbed by the plant canopy and another part, which reached the soil according Beer's law (cf. Impens and Lemeur, 1969). Radiation absorbed by the canopy was partitioned between the two plant layers (Fig. 1), depending on their height and surface cover, whereas it was assumed that leaves are uniformly

distributed within the total height of the canopy. Interception and plant evaporation depended on the simulated leaf area index of the vegetation as well as the degree of area coverage. Transpiration depended additionally on the simulated plant water uptake. Soil evaporation was derived from an iterative solution of the soil surface energy balance of the soil surface, using an empirical parameter for estimating the vapour pressure and temperature at the soil surface. Vapour pressure deficit was calculated from the relative humidity input. Snow fall was simulated from precipitation and air temperature, while snow melt was estimated from global radiation, air temperature and simulated soil heat flux.

### 2.3.2 Soil temperatures and heat fluxes

Surface temperature was simulated based on an energy balance approach, where the radiation reaching the soil equals the sum of sensible and latent heat flux to the air and heat flux to the soil. The same approach was used for the snow surface temperature. Heat flow between adjacent soil layers were calculated based on thermal conductivity functions accounting for the content of ice and water. The heat flow equation is based on a coupled equation also accounting for the freezing and thawing in the soil (Jansson and Halldin, 1979). Convection heat flows were not accounted for. The lower boundary temperature was calculated based on a sine variation including parameters for the annual mean temperature and amplitude at the site.

### 2.3.3 Soil hydrology

Soil water flows and water contents were calculated for each of the 16 soil layers. Soil water depended on infiltration to the soil, soil evaporation, water uptake by plants, and ground water flow. Soil moisture represented as liquid water content, wa<sub>l</sub> calculated based on the water storage and temperature in the corresponding soil layer. Water flows between adjacent soil layers were calculated based on Richards' equation (Richards, 1931), considering hydraulic conductivity, water potential gradient and vapour diffusion. Saturation conductivity was assigned depending on the mean measured dry bulk density values of the corresponding layers (cf. Päivänen, 1973).

In respect to hydrologic characteristics, the soil profile was divided in two horizons representing the acrotelm and the catotelm (cf. Ivanov, 1981), whereas the boundary between these horizons was positioned at -30 cm as suggested for Degerö Stormyr, based on visual differences in the soil profile and water table depth measurements (Granberg et al., 1999). The soil water characteristics were described by the Brooks & Corey equation (Brooks and Corey, 1964) and unsaturated conductivity by the Mualem function (Mualem, 1976). When the current simulated ground water table is above the assumed drainage level, outflow of saturated layers above that level was simulated, based on a linear model.

Surface runoff was controlled by a surface pool of water that covers various fractions of the soil surface. During periods of a fully saturated soil profile the flow of water in the upper soil compartment could be directed up-wards, towards the surface pool. Surface runoff was calculated as a function of the amount of water in the surface pool.

### 2.3.4 Vegetation

Two plant layers were simulated, representing vascular plants and mosses. They differed in their parameters for size, shape, carbon allocation, litter fall and temperature response for assimilation and respiration. A detailed description of the carbon pools of the two plant types and the partitioning of assimilates to the pools can be found in the supplement.

~~Vascular plants consisted of three functional parts: roots, photosynthetically active biomass (i.e. green leaves and green stems that are labelled as leaves in equations and parameter names), and photosynthetically passive biomass (i.e. brown, senescent leaves and woody stems that are labelled as stems). Mosses were considered to consist of two parts: an upper, photosynthetically active part (labelled as leaves) and a lower, photosynthetically passive part (labelled as roots) representing~~

pale or brown, belowground leaves and stems that are still living. Each plant constitutes a biomass pool for each of its parts. Vascular plants had additionally a pool for mobile reserves, that was filled during litter fall. LAI was proportional to leaf biomass by using a constant specific leaf area as conversion factor. Vascular plants They were assumed to have a maximal height of 50 cm compared to 2 cm for mosses.

Plant development was temperature sum and day length dependent. Senescence and litter fall for vascular plants depended on growth stage, temperature sum and day length. In case of mosses, litter fall was proportional to assimilation. Litter from above ground carbon pools went through a surface litter pool and then to the upper soil litter pool, litter from below ground to the corresponding soil layer. In case of mosses, litter fall occurred only in belowground parts.

Plant development started every spring when the accumulated sum of air temperatures above a threshold value reached a certain value. The accumulation of temperatures started when the day length (geometric estimated time of sun above horizon) exceeded 10 hours. Snow cover hindered leafing out by reducing the radiation supply to the plant, while low soil temperatures reduced plant water uptake.

Senescence and litter fall differed between the two plant types. For vascular plants, beside a small amount of litter fall occurring during the whole plant growth period (cf. Fulkerson and Donaghy, 2001), senescence was assumed to start after the plant reached maturity and therefore depended on growth stage (cf. Thomas and Stoddart, 1980) and dormancy temperatures (cf. Davidson and Campbell, 1983). New assimilates were constantly allocated to the roots and to the photosynthetically active part. After maturity, existing green biomass was reallocated to the photosynthetically passive part. A third stage of litter fall was configured depending on a temperature threshold: Five consecutive days in the autumn with day lengths shorter than 10 hours and with temperatures below a threshold temperature parameter terminated the growing season; Increased litter fall took place and vascular plants went to dormancy. During vascular plant litter fall, part of the carbon was stored in the mobile pool, which could be then reused for leafing out in the next year (cf. White, 1973; Wiegler, 2005). The litter from above ground biomass was inserted to a surface litter pool, while root litter was inserted to the corresponding litter pools of the soil layers in which the roots were located. The litter in the surface pool was inactive and transferred with a constant rate to the litter pool of the uppermost layer.

A different approach for senescence and litter fall was applied for mosses, as they largely differ in these processes from vascular plants: Sphagnum mosses produce new leaves in the top (capitula), while litter fall occurs on the lower leaves, when they become shaded and die (cf. Clymo and Hayward, 1982). This leads to a permanent moss cover and a litter fall that is proportional to assimilation. In the model, this was realised by keeping the photosynthetically active part of mosses to a fixed static value. Any losses (i.e. respiration and litter fall) or gains (incorporation of assimilates) were restricted to the belowground moss parts. Moss litter was produced with a constant rate coefficient throughout the year and was directly inserted to the corresponding soil litter pools. The dormancy period for mosses was initiated in the same way as for vascular plants, but affected only assimilation.

For both plant types, assimilation was simulated using the light use efficiency approach (cf. Monteith, 1972), at which total plant growth is proportional to the net of global radiation absorbed by the canopy but limited by unfavourable temperature and limited soil water. The response to soil water was defined from the ratio of actual to potential transpiration. Potential transpiration depended on vapour pressure, temperature, wind speed and aerodynamic resistance of the plant. Actual transpiration was assumed to equal water uptake from soil layers, depending on relative amount of roots, the specific response to soil water potential, and soil temperature of each layer. Both plant layers were assumed to be well adapted to wet conditions (cf. Keddy, 1992; Steed et al., 2002) and therefore experiencing water stress only due to too dry conditions, which was supported by pre-study modelling results.

Plant respiration was assumed to be proportional to assimilation (growth respiration) and to amount of biomass (maintenance respiration), whereas maintenance respiration depended also on temperature trough a -in active leaves and roots. In case of mosses, maintenance respiration took place only in belowground parts, therefore a higher range for the



~~parameter sealing growth respiration was calibrated (cf. Table S1 in the supplement). A simple Q10 approach was used to simulate the response of plant maintenance respiration on temperature.~~

### 2.3.5 SOC decomposition

5 The organic substrate was represented by three C and N pools for each of the 16 soil layers: one representing more stable, partly decomposed material ( $SOM_s$ ), one representing fresh or little decomposed moss litter ( $SOM_m$ ) and one representing fresh or little decomposed litter from vascular plants ( $SOM_v$ ). Initial conditions were selected to fulfil the measured total carbon per layer and partitioned into the pools in the way that they were approximately in equilibrium for a certain parameter combination that produces a reasonable fit to NEE (prior calibration). ~~Decomposition followed first order kinetics with pool specific rates which were reduced under unfavourable soil temperature and moisture conditions. Temperature dependence was described with a function which was developed by the Ratkowsky function, that was originally developed for bacteria (Ratkowsky et al. (1982) for bacteria, but has also been~~ has also been applied to fungal growth by (Bazin and Prosser, (1988). The response to Soil moisture response was assumed to be zero at moisture contents below the wilting point, rising to 100% between two threshold moisture contents and falling to a certain level under saturated conditions.

15 Decomposition products from the  $SOM_m$  and  $SOM_v$  pools were partitioned into  $CO_2$  ~~which-that~~ was released to the atmosphere and C ~~which-that~~ is partly moved to the  $SOM_s$  pools and partly returned to the  $SOM_m$  and  $SOM_v$  pools. Decomposition products from the  $SOM_s$  pools were partly released as  $CO_2$  and partly returned to the  $SOM_s$  pools. Under saturated conditions, carbon could leave the pools as methane ( $CH_4$ ), which was later oxidised to  $CO_2$  or transported to the atmosphere via plants or through ebullition. Nitrogen and methane related processes were considered by a model including the most important pathways and fluxes, but ~~However no emphasize on the calibration of these processes were made in this study.~~

~~The rate at which carbon was transferred between pools and towards the atmosphere was pool specific and reduced under unfavourable soil temperature and moisture conditions. Temperature dependence was described with a function which was developed by Ratkowsky et al. (1982) for bacteria, but has also been applied to fungal growth (Bazin and Prosser, 1988). The response to moisture was assumed to be zero at moisture contents below the wilting point, rising to 100% between two threshold moisture contents and falling to a certain level under saturated conditions.~~

Peat depth growth during the simulation period was considered by the following: The initial organic concentration was preserved for each layer but the lowest in the profile. Instead, the difference in the total amount of C in all pools in one layer between start and end of each year was moved to or from the layer below, to simulate growth or decrease of the peat depth. Thereby, carbon was taken from the different pools according to the relative abundance of each pool in the source layer and inserted to the corresponding pool in the target layer to allow dynamic changes in litter quality. The lowest layer (-2.8 to -3.4 m below the surface) represented the entire depth change of the whole profile, but was excluded from a constant concentration to avoid adjustments of the number of layers.

~~Nitrogen and methane related processes were considered by a model including the most important pathways and fluxes. However no emphasize on the calibration of these processes were made in this study.~~

### 2.4 Calibration procedure

A Monte Carlo calibration including acceptance criteria was performed to identify process and parameter interactions. T and ~~the~~ resulting parameterisations were analysed for correlations between different parameters, between parameters and model performance and between performances in different variables, ~~to identify process and parameter interactions.~~ 50'000 runs were performed to calibrate 54 parameters from different processes. Parameter values were randomly assigned from, using a

uniform ~~random~~-distribution within assumed prior ranges (i.e. all values had the same probability of being used). ~~for 54 selected parameters from different modules.~~ The parameters were selected as candidates to demonstrate the role of various regulating processes, which we group into eight different process categories: processes that describe 1) plant growth, 2) decomposition, 3) radiation interception, 4) soil temperature, 5) aerodynamic resistance, 6) transpiration, 7) soil hydrology, and 8) snow. ~~Many parameters were still considered with fixed single values (Table S3 in the supplement).~~ Prior ranges for calibrated parameters were selected according to literature values or experiences from previous model runs, in most cases a certain range around the default values (Table S1 in the supplement). Many parameters were still considered with fixed single values (Table S3 in the supplement). Model outputs were compared with measured field data including many variables in high temporal resolution, spanning up to 12 years of observations (Table 2). Several combined criteria were defined to select runs (behavioural models) with an acceptable performance (see Sect. 2.4.2) in different variables. Resulting parameter value ranges of the accepted runs were then compared with the prior ranges and between the different criteria selections to examine the effect of criteria selection. Correlations between parameter values and model performance in the different measurement variables were analysed, as well as between accepted values of different parameters. Parameters were ranked in their effect on model performance, their correlation with other parameters and their constrain ability from the available data.

#### 2.4.1 Splitting of calibration variables into sub-periods

Additional to the calibration data for the whole period we introduced further sub-variables for certain sub periods and times of the day. NEE was separated into night time values (22:30 – 02:30), representing ecosystem respiration, and day time values (09:30 – 15:30), representing the sum of the respiration component and the assimilation component. Additionally, spring time values were considered separately for NEE and snow depth, and spring and winter time values for Rn, Ts, H, and LE. This is justified as low values with little dynamic during winter and the critical transition of plant emerge and snow melt in spring might not be properly accounted for, if only the whole period was considered. WT was calibrated and analysed in the whole profile and additional in lower soil layers (one sub-variable for WT depths > -0.15 m and one for > -0.2 m). This was motivated, as WT in the upper soil layers showed high fluctuations in the modelled, and also partly the measured WT, while our interest was to achieve a good overall water table with good representations of dry summer periods.

#### 2.4.2 Performance indices

Selection of runs and evaluation of model performance were based on three indices: coefficient of determination ( $R^2$ ) assesses how well the dynamics in the measurement derived values are represented by the model. Mean error (ME) is the difference between the average of the simulated compared to the average in the measured, i.e. it shows the error in the magnitude. Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) accounts for both, deviation of dynamics and magnitude. It ranges from  $-\infty$  to 1, whereas 1 means the best fit of modelled to measured data. Values < 0 indicate that the mean measured value is a better predictor than the simulated value (Moriassi et al., 2007). As NSE may be understood as a combination of  $R^2$  and ME, it was only evaluated, if  $R^2$  and ME alone did not narrow the parameter range.

The decoupling of turbulent transport and biological activity during night time may introduce spike-type fluxes in NEE, LE and H if accumulated concentrations during calm night-time conditions are released during the onset of turbulence in the morning hours ~~NEE showed a spiky record, especially during night time, probably caused by transport processes in the atmosphere, which were not represented in the model.~~ To attenuate the effect of these spikes, the simulated and measured

values were transformed to cumulated total amounts, starting from the beginning of the observation period. An additional  $R^2$  value was calculated for the cumulated values ( $AR^2$ ).

### 2.4.3 Criteria for posterior selection

Criteria were applied in two steps. In the first step, a basic set of 1285 behavioural models was selected. Out of these, several sets of 50 runs each were selected in the second step in two different ways: one for sensitivity analyses and parameter ranges which was based on single criteria and the other for identification of equifinalities, based on multiple criteria.

#### Basic selection

The basic selection was applied, as the lowest summer water levels and a reasonable representation of the plant was assumed to be crucial for most of the processes of interest. Criteria were on performance in WT and vascular plant LAI (Table 3). The criteria on water level below 0.2 m was chosen, as a correct representation of summer drought conditions was of higher interest in this study than a correct water level during e.g. frozen conditions in winter, causing water table drop downs to 0.15 m. The criteria on LAI ME of  $\pm 0.2 \text{ m}^2 \text{ m}^{-2}$  was a relatively wide range, as the mean of measured values was  $0.4 \text{ m}^2 \text{ m}^{-2}$ , i.e. a underestimation of LAI by  $-0.2 \text{ m}^2 \text{ m}^{-2}$  would result in a maximum LAI of 0.2–0.4, which was close to the minimum for being able to re-establish new biomass after a low productive year. A wide range of day-time NEE ME was additionally applied to exclude outliers due to numerical problems, which reached an ME in NEE up to  $8 \cdot 10^{27} \text{ gCO}_2\text{-C day}^{-1} \text{ m}^{-2}$  in the prior.

#### Single criteria to identify parameter range

For sensitivity analyses and to test if, and how, parameter ranges depend on the selected criteria, the best 50 behavioural models for each performance index of each variable were selected out of the basic selection. Thereby, best means highest in case of  $R^2$  and NSE, but closest to zero in case of ME. We defined posterior parameter range as the interval between the 5% and the 95% percentile of the distribution of parameter values of the runs selected. Posterior parameter ranges were compared with the ranges resulting from the basic selection. If the upper or lower limit of a posterior parameter range of the final selections differed by  $\geq 10\%$  from the upper or lower limit of the posterior range of the basic selection, the parameter was assumed to be sensitive to the selected criteria and further analysed.

The same was done for each best 200 behavioural models, but as the results were similar, they were only plotted in respect to parameter ranges. Further, all parameters were plotted against all performance indices of each variable and checked visually for discrepancies with the resulting ranges (results are not shown).

#### Multiple criteria to identify parameter correlations

For identification of equifinalities, a set of multiple criteria for each variable (Table 3) was applied to select sets of 50 behavioural models each. Again, these selections were based on the basic selection. Parameter ensembles of these accepted behavioural models were then analysed to identify covariance between parameters. A pair of parameters was considered to interact, if their values correlated with an  $R^2$  of at least 0.1 in the basic selection, respectively 0.2 in the final selection. If a pair showed correlations in several criteria sets, the highest  $R^2$  value was reported in the results.

### 2.4.4 Evaluation and measures

To rank the parameters in their concern, several measures were used to quantify parameter sensitivities and constrainabilities, as well as equifinalities. The sensitivity ( $S$ ) of a parameter to each performance index of each variable was quantified by the sum of the differences between posterior range and prior range (range reduction). If a parameter was sensitive to more than one period of each variable, the highest value for each variable was chosen for further analysis. To

identify trade-offs and supporting effects between different criteria, correlations of the performances between different variables and indices were plotted and visually analysed. Due to limited computer capacity, this was based on a random set of 3200 runs. Further, the parameter value ranges resulting from the different criteria were compared with each other and determined how well they were overlapping, i.e. how unambiguously they could be constraint. Overlap ( $O$ ) for each parameter was defined as the difference between the minimum of the upper limits of the posterior ranges of the different criteria, minus the maximum of the lower limits of posterior ranges and therefore become negative, if ranges were not overlapping. Further it was compared how well overlapping ranges differed between performance indices within the same variable and between different variables. The overlapping range of each parameter was normalized by dividing it by the average of the posterior ranges of this parameter, so that a value of 1 would be reached if all posterior ranges of that parameter would be identical for all performance indices and variables. Equifinalities were quantified by the  $R^2$  value of a simple linear regression through the values of the interacting parameter pair in the accepted runs. of the correlation between each parameter pair. Parameter concern ( $P$ ) was defined based on three components: the sensitivity of the parameter, how unambiguously it could be constraint and the sum of correlation coefficients of equifinalities with other parameters:

$$P = (S_{R^2} + S_{ME}) \times (1 - O) + \sum 2 \times \frac{2^{10 \times R^2_{equi}}}{10} \quad (1)$$

Thereby, sensitivity was the sum of the range reduction for  $R^2$  and for ME, respectively NSE in case no sensitivity was detected for  $R^2$  and ME but NSE. The sensitivity was multiplied by the factor one minus the normalized overlapping range, so that the sensitivity of parameters which could be unambiguously constrained are down weighted, and such with high uncertainty due to different results for different performance indices or variables are up weighted. Equifinalities were considered by the sum of  $R^2$  values for each correlation of that parameter with another parameter, displayed in exponential form and weighted, so that strong correlations were emphasised and the contribution of equifinalities were in a comparable scale to the sensitivity measures.

### 3 Results

Processes as well as parameters were strongly interacting, which was reflected in sensitivities of each variable to several different modulesprocess categories, correlations between the performance in different variables, and in equifinalities between parameters of different modulesprocess categories.

About half of the parameters were sensitive to model performance in one or more variables, but only ~~very~~ few had a distinct range (Sect. 3.1). Instead they affected several processes, causing trade-offs in model performance between the different measurement variables and between the different performance indices, but also several supporting effects could be identified (Sect. 3.2). A lot of equifinalities were identified between parameters. Parameters were correlated with up to seven other parameters, often from different modulesprocess categories. Therefore, a good performance often requires certain combinations of parameter values, rather than specific parameter values (Sect. 3.3).

Each of the available measurement variables (NEE, LAI, sensible and latent heat fluxes, net radiation, soil temperatures, water table depth and snow depth) constrained ~~several~~ parameters from several different process categories, without any variable being redundant (Sect. 3.4). Nevertheless, large uncertainty remained in especially the unsaturated water distribution ( $\psi_a$ ) in the soil (Fig. 2), which affected all considered processes and hindered further parameter constrain. This might be solved by additional measurements of i.e. soil hydraulic properties. Other important parameters that could not be constrained, define aerodynamic resistance, radiation interception (in particular moss albedo), timing of snow melt, and in case of NEE mostly the leaf litter fall rate of vascular plants during the growing season (Fig. 2).

A detailed description of the key parameters for each process and the detected interactions can be found in Sect. 3.5. Results for model fits to the different variables can be found in Fig. S1 in the supplement.

### 3.1 Parameter sensitivity

Model performance was sensitive to parameter across the different process categories: Most of the Out of 27 sensitive parameters were 21 that affected model performance in more than one variable, but For 15 of the sensitive parameters, resulting value ranges differed strongly (less than 50% overlapping range), depending on both, the variable and the performance index (Fig. S2 in the supplement). Performance in Ts and WT was determined by 12 key parameters belonging to seven and six different modprocess categories, respectively (Fig. 3). In contrast, snow depth and LAI depended mainly on parameters from their own modprocess categories. Large differences in resulting accepted ranges depended on the selected performance index and the considered sub-period: On average, accepted value ranges overlapped with 35% between different performance indices and between different sub periods of the same variable and with 6% if additionally the differences between different variables were considered (Fig. 4). Radiation and LAI refer to the simplest processes in respect to number of connected parameters (Fig. 3). However, radiation was, together with snow depth, the variable with the strongest average disagreement in parameter value ranges between the different selection criteria (Fig. 4). In case of eleven parameters, the accepted ranges did not overlap at all (Fig. S2 in the supplement). Four parameters were sensitive to at least half of the considered variables (Fig. 2): The parameter defining the water retention curve and unsaturated soil hydraulic conductivity ( $\psi_a$ ) affected model performance in variables of all eight considered variables. Moss transpiration coefficient ( $g_{max,moss}$ ), vascular plant respiration coefficient ( $k_{gresp,vasc}$ ) and litter fall rate ( $I_{Lcl}$ ) were important parameters for not only LAI and NEE, but also H, LE and WT,  $g_{max,moss}$  and  $k_{gresp,vasc}$ , additional for Ts. The sensitivities of the single parameters are described in more detail in Sect. 3.5. The full table of the correlation coefficients between parameters and performance can be found in the supplement (Table S4).

### 3.2 Confounding and supporting effects of interacting processes

The performances of several variables were connected in supporting and cofounding ways (Fig 5 and 6). Especially ME of LE and WT were strongly connected, but also ME of LAI had an impact on the performance in many other variables. Trade-offs existed not only between the performances of different variables, but also within a variable, depending on chosen performance index or seasonality sub-period. This was also reflected in the Large differences in resulting accepted ranges depended on the selected performance index and the considered sub period: On average, accepted value ranges overlapped with 35% between different performance indices and between different sub periods of the same variable and with 6% if additionally the differences between different variables were considered (Fig. 4). In case of eleven parameters, the accepted ranges did not overlap at all (Fig. S2 in the supplement).

Strong connections existed especially between ME of LE and WT were strongly connected, but also ME of LAI had an impact on the performance in many other variables:

The magnitude of vascular plant LAI was strongly correlated with magnitude of LE, WT, H and NEE, especially if daytime and night time values were considered (Fig. 5). Thereby the lowest ME in day and night time NEE, as well as ME and dynamics of H, went along with a slight underestimation, and for LE and WT with a slight overestimation of vascular plant LAI. Best performance for WT dynamics was reached if the magnitude of vascular plant LAI was correct (Fig. 6). A noticeable existence of the vascular plants (LAI ME > -0.4) increased the fit in NEE  $R^2$  to at least 0.2, but this was not a necessary precondition for good NEE performance (Fig. 6). Highest performance in dynamics of WT, H and Ts in the upper layer coincided with a good fit in NEE magnitude (Fig. 6). This relationship was even stronger if these variables were compared to ME in NEE night time and NEE daytime.



A correct representation of WT dynamics and depth coincided with high performance in H dynamics and a correct or slightly underestimated H (Fig. 5 and 6). A small ME in H correlated with high performance in WT dynamics. Performances in soil temperatures of different layers were strongly correlated with each other in both, dynamics and magnitude.

Underestimation of LE was connected to an overestimation of H, but also to better dynamics in H (Fig. 5). ME in Net radiation was positively correlated with ME in H. A good fit between modelled and observed snow depth did not correlate with the performance in any other variable. The only exception was a negative correlation between the dynamics in snow depth and H, if exclusively performance during spring time was considered (Fig. S3 in the supplement).

~~Trade-offs existed not only between different variables but also between different performance indices of the same variable.~~

Especially for snow, Rn, and in case of some parameters also for Ts, accepted ranges were contradictory depending on whether  $R^2$  or ME was chosen. In case of moss albedo ( $a_{pve,moss}$ ) and aerodynamic resistance dependency on LAI ( $r_{alai}$ ), the ranges also strongly depended on the season during which the variable was considered. For two aerodynamic resistance and one soil parameter ( $z_{0M,snow}$ ,  $c_{H0,canopy}$ ,  $S_k$ ) ranges differed between  $R^2$  of actual values and  $R^2$  of accumulated values.

Additional to the uncertainty from unambiguous parameter ranges, further uncertainty results from equifinalities between parameters.

### 3.3 Equifinalities

Parameters were strongly inter-correlated, often with several parameters, and often ~~from across~~ different ~~model process~~ categories. Equifinalities can hinder the identification of sensitivities, which was especially true for the basic selection: Despite reducing the number of runs by 97.5%, posterior and prior ranges differed hardly (Table S5 in the supplement).

Instead certain value triples for photosynthetic efficiency ( $\epsilon_{L,vasc}$ ) with the respiration coefficient ( $k_{gresp,vasc}$ ) and with the storage fraction for plant regrowth in spring ( $m_{retain}$ ) were crucial for the survival of the vascular plant layer. Certain value pairs for the moss transpiration coefficient ( $g_{max,moss}$ ) with the shape parameter of soil water retention ( $\psi_a$ ) were crucial for a reasonable water table depth.

Equifinalities existed not only between parameters from the same ~~model process~~ categories, but even more often between parameters from different ~~model process~~ categories (Fig. 7). Parameters defining radiation interception, soil temperature, aerodynamic resistance, transpiration, and soil hydrology correlated with exclusively parameter from different ~~model process~~ categories. Parameters defining radiation interception were mostly correlated with parameters defining aerodynamic resistance. Only in case of plant and SOC decomposition parameters, equifinalities existed mainly between parameters of the same ~~model process~~ categories.

Except  $\rho_{smin}$ , all sensitive parameters and further other parameters were detected to correlate with up to five other parameters in the final selections,  $\psi_a$  correlated with even seven others (Fig. 2). Two parameters had very strong correlations ( $R^2 \geq 0.3$ ) with two other parameters each, which belong to different ~~model process~~ categories ( $\psi_a$  with  $c_{H0,canopy}$  and  $g_{max,moss}$  and  $a_{pve,moss}$  with  $z_{0M,snow}$  and  $r_{alai}$ ) (Table S6 in the supplement).

### 3.4 Usefulness of measurement variables

All available measured variables (NEE, LAI, LE, H, Rn, Ts, WT and snow depth) were helpful in constraining parameter ranges (Fig. 2). None of the supporting effects was strong enough, to make one variable fully replaceable by another. Even for the strongest correlation between soil temperatures of the different layers, the remaining uncertainty in one temperature

when knowing the other would be in the magnitude of 0.5°C, which corresponds to more than 25% of the total uncertainty resulting from the tested parameter ranges (Fig. 5).

In case of 15 variables, the usage of several variables revealed that constrained ranges were not robust. Twelve parameters could be unambiguously constrained to a more narrow range, as their resulting ranges were had at least 50% well overlapping, or affected only one variable (Fig. S2 in the supplement). The performance on each variable constrained was correlated with many parameters from several different processes-process categories (Fig. 3). The highest number of correlations was detected for the performance in WT and Ts, which constrained 12 parameters from different modulprocess categories. Also the available data for LE, H, and NEE constrained many parameters.

Still, large uncertainty remained due to equifinalities and differences in accepted ranges: The largest uncertainty was caused by a parameter defining the shape of the water retention curve (air entry,  $\psi_a$ ). As this was the only calibrated parameter of the water retention curve, it determined the unsaturated hydraulic conductivity of the soil.  $\psi_a$  was sensitive to all considered variables and had many strong interactions with other parameters, while it was not possible to constrain it to an unambiguous value range (Fig. S2 in the supplement). Therefore it would be of great value to be able to deduce such parameters from additional measurements. This applies also to following parameters, which could not be constrained unambiguously: Leaf litter fall rate of vascular plants during the growing season ( $l_{Lci}$ ) was the second most sensitive parameter, affecting the performance in NEE, H, LE and WT. Moss albedo ( $a_{pve,moss}$ ), aerodynamic resistance dependency on LAI ( $r_{alai}$ ) and transpiration coefficients ( $g_{max,vasc}$ ,  $g_{max,moss}$ ,  $g_{maxwin}$ ) had similar importanceconcern, due to their equifinalities to other parameters. Plant respiration ( $k_{gresp,vasc}$ ) had strong sensitivity, but could be constrained unambiguously by the available data.

### 3.5 Detailed description of sensitivities and interactions per process

Detected sensitivities, connections between performances, and equifinalities showed all strong interactions between the different processes and parameters of different modulprocess categories. Connections existed between all variables and modulprocess categories, but most strongly interlinked were LE with WT, Rn with H and Ts (Fig. 2). H, LE, WT were also linked to each other and to NEE. The impact of the plant is further reflected in the correlations between performances in LAI with performances in many other variables (Fig. 5). The implications on the performance for each considered variable will be described in the following sections.

#### 3.5.1 Water level depth and soil moisture conditions

Performance in water level depth was determined by 12 key parameters (Table S4 in the supplement). It was most strongly connected to the shape of the soil water retention curve ( $\psi_a$ ) as well as to the transpiration coefficients for mosses and winter transpiration ( $g_{max,moss}$ ,  $g_{maxwin}$ ). The transpiration coefficient from vascular plants played a smaller role due to the high sensitivities of parameters defining the growth and therefore magnitude of the vascular plant (i.e.  $k_{gresp,vasc}$ ,  $m_{retain}$ ,  $l_{Lci}$ ). Equifinalities existed between several of these parameters.

$\psi_a$  had strong effect on the performance of all variables and several strong equifinalities, in particular with parameters defining aerodynamic resistance and transpiration; On the other hand  $\psi_a$  could not be constrained to an unambiguous range and was therefore the parameter causing the largest overall uncertainty (Fig. 2).

Performance in WT was further sensitive to parameters defining aerodynamic resistance, i.e.  $-r_{alai}$  and  $c_{H0,canopy}$ . Both parameters had equifinalities with  $\psi_a$  and moss albedo ( $a_{pve,moss}$ ) as well as with timing of snow melt ( $m_T$ ) and thermal conductivity of snow ( $s_k$ ). Also the distance between drainage ( $d_p$ ), showed some sensitivity.

### 3.5.2 Transpiration and evaporation

The nine most important parameters for WT performance were also key parameters for LE ( $\psi_a$ ,  $g_{max,vasc}$ ,  $g_{max,moss}$ ,  $g_{maxwin}$ ,  $k_{gresp,vasc}$ ,  $m_{retain}$ ,  $l_{Lc1}$ ,  $r_{alai}$ ,  $c_{H0,canopy}$ ). This explains the strong correlation between the performance in WT and LE ME (Fig. 5) and shows the connections with plant, WT and H. Another parameter, sensitive to LE was the roughness length of snow ( $z_{0M,snow}$ ), belonging to the aerodynamic resistance ~~model~~process category and correlating with moss albedo, hinting to the connections between LE and R associated processes.

Dynamics in WT and LE, but also magnitude of H was improved if the transpiration coefficient was on its lower range in case of mosses and on its upper range in case of vascular plants (Fig. S2 in the supplement). Despite the lower values for mosses, transpiration prior criteria selection was dominated by mosses, due to their higher LAI and coverage (Fig. S4 in the supplement).

Crucial for LE performance was also a parameter defining the aerodynamic resistance of the canopy under stable conditions ( $c_{H0,canopy}$ ): a very small value improved the  $R^2$  of LE and spring LE, but downgraded  $R^2$  of accumulated LE and of winter radiation.

Spring LE was overestimated in most of the runs (see Fig. S1 in the supplement). The strongest sensitivity on spring LE was by the coefficient for winter transpiration ( $g_{maxwin}$ ): the higher the better  $R^2$  and ME. Together with ( $z_{0M,snow}$ ) this was also the most important parameter for winter LE.

### 3.5.3 NEE & LAI

Seven of the nine parameters which were common for LE and WT were also among the most effective parameters for NEE ( $\psi_a$ ,  $g_{max,moss}$ ,  $g_{max,vasc}$ ,  $k_{gresp,vasc}$ ,  $m_{retain}$ ,  $l_{Lc1}$ ,  $r_{alai}$ ) and belong to four different ~~model~~process categories: plant, transpiration, soil hydrology and aerodynamic resistance (Table S4 in the supplement). However the most sensitive parameter for NEE was the rate coefficient for heterotrophic respiration ( $k_H$ ), which was especially important for night time NEE. Further sensitive parameters for night time NEE were the growth respiration coefficient for mosses ( $k_{gresp,moss}$ ) and the temperature dependency coefficient for heterotrophic respiration ( $t_{min}$ ).

The rates of photosynthesis and its temperature dependence ( $\epsilon_{L,vasc}$ ,  $\epsilon_{L,moss}$ ,  $p_{mn,vasc}$ ) were key parameters for LAI, NEE magnitude or temporal NEE dynamics, respectively. Many strong interactions existed between plant parameters, which were especially visible in the basic selection (see Sect. 3.3).

The rate of leaf litter fall during the growing season  $l_{Lc1}$  was one of the parameters with the highest concern, due to its sensitivity on many different processes, its equifinalities and as it could not be constrained to an unambiguous solution (Fig. 2). Resulting ranges for  $l_{Lc1}$  differed especially between the different performance indices within NEE and within LAI, but also between NEE and LAI (Fig. S2 in the supplement).

### 3.5.4 Sensible heat fluxes, soil temperatures and net radiation

Many inter-connections existed between H, Ts and Rn, but all three were also linked with LE, WT, snow and NEE. A snow parameter, determining the timing of snow melt ( $m_T$ ) was the most crucial parameter for heat fluxes, not only in spring time, but also for the whole year period. Further,  $m_T$  was important for Ts in spring time (cf. Sect. 3.5.5). The shape of the soil water retention curve ( $\psi_a$ ) was the second most sensitive parameter for both variables.

The aerodynamic resistance dependency factor on LAI ( $r_{alai}$ ) was the most sensitive parameter for Ts, and affected also LE, WT and night time NEE, while it strongly correlated with moss albedo ( $a_{pve,moss}$ ), the third most sensitive parameter for H and most sensitive parameter for Rn. Accepted ranged for  $r_{alai}$  contradicted within the soil temperature variables, depending on the chosen performance index and considered season: high values were important for Ts ME and  $R^2$  during winter, but

low ones improved Ts  $R^2$  during spring and during the whole period. Therefore,  $r_{lai}$  was the parameter causing the largest overall uncertainty after  $\psi_a$ . This was followed by  $a_{pve,moss}$ , which had low values for accepted ranges in case of H, Rn and Ts during the whole period, but high values in case of winter H and Rn. It further showed strong ~~equifinalities~~ equifinalities with the roughness length of snow ( $z_{0M,snow}$ ), which was the second most sensitive parameter for Rn, but also affected H and LE.

The coefficient for thermal conductivity of snow ( $s_k$ ) affected Rn and Ts, but not H.

The thermal conductance coefficient of soil organic material ( $h_2$ ), the lower boundary mean temperature ( $T_{amean}$ ), the snow melt dependency to radiation coefficient ( $m_{Rmin}$ ) and the density of new and old snow ( $\rho_{smin}$ ,  $S_{dw}$ ) affected only soil temperatures, the latter two also snow depth.

Parameters defining moss and winter transpiration ( $g_{max,moss}$ ,  $g_{maxwin}$ ) and the growth respiration coefficient of vascular plants with its effect on vascular plant biomass and LAI ( $k_{gresp,vasc}$ ) were sensitive to Ts,  $g_{max,moss}$  and  $k_{gresp,vasc}$  also to H. The most important parameter for LE,  $c_{H0,canopy}$  was another key parameter for Rn and H.

### 3.5.5 Snow

The temperature coefficient in the snow melt function ( $m_T$ ) was the most important parameter for ME in snow and determined timing of snow melt. However, resulting parameter ranges did not overlap between the different performance indices within the snow depth variable and between different other variables. A longer lasting snow cover (low  $m_T < 3$ ) was crucial for spring H and reduced mean error in snow depth, but lowered  $R^2$  values in spring Ts and snow depth.  $m_T$  interacted with another snow parameter ( $T_{RainL}$ ) as well as with parameters from the temperature and transpiration ~~modulprocess~~ categorye ( $T_{amean}$ ,  $g_{maxwin}$ ). The density coefficients for old ( $S_{dw}$ ) and new snow ( $\rho_{smin}$ ) had medium effect on snow depth performance, and affected also spring and winter soil temperatures in all layers, but the latter could be unambiguously constrained by the available data.

## 4 Discussion

Unlike many previous sensitivity studies for ~~carbon~~ CO<sub>2</sub> modelling that often focus on only one or few calibration variables and parameters of the associated ~~modulprocess~~ categorye, we considered many different abiotic and biotic measurements (NEE, LAI, Rn, Ts, H, LE, WT and snow depth) to investigate the interactions between various process ~~categorieses~~ categorieses (SOC decomposition, plant growth related processes, radiation interception, soil temperature, aerodynamic resistance, transpiration, soil hydrology and snow) in a peatland ecosystem.

Similarly to results from a forest modelling study using the DRAINMOD-Forest model (Tian et al., 2014) and a N<sub>2</sub>O study using CoupModel on a drained peatland forest (He et al., 2016), we found that processes were sensitive to parameters from several different ~~modulprocess~~ categories. Together with the discovered supporting effects between model performances in different variables, this confirms the connections and dependencies between different processes as implemented in the model (cf. Model description and equations, Sect. 2.3, Table 2 in the supplement and Janson and Karlberg, 2010). The many ~~interactions between parameters of both, between the same and also~~ equifinalities within and between different ~~modulprocess~~ categories, reveal the dependency of constrained parameter ranges as well as parameter sensitivities to model structure, calibration setup and parameters with fixed values: a deviation in one of these factors leads to different optimal value ranges, whereas a non-sensitive parameter might become sensitive if an interacting parameter is set constant. This implies a limited transferability of parameter values between models in general and even between studies using the same model in a different configuration. Resulting parameter ranges were moreover affected by the applied criteria for selecting runs. Yet, it is quite common practice to adopt at least some parameter values from other modelling studies (e.g. Frohling et al., 2002, Yurova et

al., 2007; St. Hilaire et al., 2010; Wania et al., 2010; Gong et al., 2013; Kim et al., 2014; Kurnianto et al., 2014; Zhu et al., 2014), which includes the usage of model default values that were estimated under a different model configuration.

~~Further, T~~he strong interactions across different ~~modul~~process categories also emphasize the importance of measurements of ancillary data additionally to the variable of interest and model input data (meteorological and SOC data). Measurements of NEE, LAI, LE, H, Rn, Ts, WT and snow were all found to be valuable for constraining parameters from several different ~~modul~~process categories and can therefore reduce uncertainty in model predictions. Further constraint of the parameters in this study would be possible, if especially additional water content or soil hydraulic properties were measured.

Beside parameter uncertainty, also uncertainty in model structure and in measured input and calibration data contribute to model uncertainty (Thorsen et al., 2001; Beven and Freer, 2001). This was tested for other peatland models (e.g. model structure: Tang et al., 2015; input drivers: Wania et al., 2009; St-Hilaire et al., 2010; Grant et al., 2011, Kim et al., 2014), but goes beyond the scope of this study. Here, only one model and one site was investigated. A previous study using CoupModel investigated the differences of parameter ranges between several different peatland sites (Metzger et al., 2015).

#### 4.1 Parameter sensitivity

~~The sensitivity of variables to parameters from many different processes revealed the importance of process interactions. The gained knowledge on parameter sensitivities can help to simplify future calibrations (Saltelli et al., 2000), by focussing on the most striking parameters and narrowing the ranges for parameter which could be successfully constrained. Further it helped to identify process interactions.~~ Especially abiotic processes were strongly inter-linked, but also biotic variables showed sensitivities to parameters from up to seven different ~~modul~~process categories, suggesting that parameter sensitivities and model performance of a certain process depend on which other ~~modul~~process categories are considered in ~~the a~~ model and in ~~the a~~ calibration. This is an important finding, as many studies investigate the sensitivity of often only few parameters from mainly the same ~~modul~~process category as the output variable (e.g. Yu et al., 2001; Frolking et al., 2002; Belassen et al., 2010; Wania et al., 2010; Morris et al., 2012; Wu and Blodau, 2013; Zaho et al., 2013; Zhu et al., 2014), which might lead to sensitivities and resulting ranges that are not robust. The ~~knowledge on these dependencies identified interactions~~ can help modellers to develop or select an appropriate model including the parameters, processes and ~~modul~~process categories which need to be considered together, depending on the variable of interest.

~~The gained knowledge on p~~Parameter sensitivity analyses can also help to simplify future calibrations (Saltelli et al., 2000), by focussing on the most striking parameters and narrowing the ranges for parameter which could be successfully constrained.

~~Though, w~~While the existence of interactions between the processes and their parameters is supposed to be less dependent on site conditions and model structure, the exact shape of the connections, as well as constraint parameter ranges, as well as the relevance of the specific processes and the specific interactions might strongly depend on these factors.

~~Still, o~~One or more of the following parameters that we identified as ~~-most influential~~, correspond to key parameters in other studies using other models and partly different ecosystems: The respiration rate coefficients, radiation use efficiency, transpiration coefficients or the soil water retention capacity were among the most sensitive parameters for NEE, its components, or yield, respectively, in e.g. the PCARS (Frolking et al., 2002) and the GUESS-ROMUL (Yurova et al., 2007) model on peatland, the SiB v2.5 model on a forest area including some wetlands (Prihodko et al., 2008), the LPJ-GUESS model on forest and herbaceous vegetation (Pappas et al., 2013), the EPIC model on cropland (Wang et al., 2005), the BIOME-BGC model for different tree species (Tatarinov and Cienciala, 2006), or the ACASA (Staudt et al., 2010), the 3-PG (Esprey et al., 2004; Xenakis et al., 2008), the FORUG (Verbeeck et al., 2006) or the DRAINMOD-FOREST (Tian et al., 2014) model on forest. These sensitivities seem to be therefore quite independent of model structure, included processes and parameters used for calibration and apply to different types of ecosystems. The resulting value ranges of these parameters



should be compared between ecosystems and models to find out to what extent they can be related to site conditions and therefore used for predictions and upscaling.

~~However, one has to bear in mind that resulting constrained ranges~~ They might be connected to the environmental scenario (Hidy et al., 2012; Ben Thouhami et al., 2013; Sulman et al., 2013) and the chosen prior distributions of the parameters (e.g. Tatarinov & Cienciala, 2006). Further, our results have shown that the parameter ranges depend on model structure, on the selection of parameters for calibration and on the selected acceptance criteria. Thereby, not only the selected variable, but also the selected sub-period was relevant, as has been shown by other studies as well (e.g. Prihodko et al., 2008; Van Huisteden et al., 2009; Safta et al., 2014).

#### 4.2 Confounding and supporting effects of interacting processes

Criteria selection is a subjective choice of the modeller if multiple output variables are available. The identified supporting effects and trade-offs between the performances in different variables allow modellers to assess the implications of a certain criteria on model performance and parameter ranges and to choose criteria according to the processes of interest—, however some of them might be ecosystem or model specific. Trade-offs existed not only between different variables but also within the same variable, depending on whether ME,  $R^2$  of actual or  $R^2$  of accumulated values was chosen and which season was considered. This implies that the problems of a subjective criteria selection also exist if only one time series variable is considered. Even if a standardised multi-criteria optimization algorithm like Bayesian calibration or a more sophisticated performance index combining several performance measures is used, the choices and the corresponding weightings are moved to the developer of the algorithm or index, but still remain subjective.

More than half of the sensitive parameters in this study could not be constrained to an unambiguous range. Constraining such a parameter by only one variable and one index would result in a range that is not robust. Using several measurement variables and several indices can therefore help to test the robustness of calibrated parameters. A parameter that is robust might better represent a physical constant, whereas controversial resulting ranges might hint to a not well represented system: There is no value for this parameter that leads to simultaneously the best performance for dynamics and magnitude in all variables and during all periods. Instead of a physical constant this parameter might correspond to a dynamic process. Beside model inadequacy, mismatching ranges could be caused in some cases by an inappropriate performance index (cf. discussion in Sect. 4.5.4) or measurements that do not truly represent the modelled variable. E.g. with the EC technique, NEE is not directly measured as the  $\text{CO}_2$  exchange between biosphere and atmosphere at a certain point, but rather results from calculations of the turbulent exchange of vertical fluxes measured several meters above the ground. Moreover, fluxes may originate from a footprint area that changes diurnally and seasonally and thus may include different soil conditions and vegetation.

Usually, LE is assumed to be closely connected to NEE due to the coupling of transpiration and carbon assimilation in vascular plants (e.g. Schulze et al., 2006), but has also been shown to correlate for mosses (e.g. Robroek et al., 2009). Our study reveals much stronger relations between parameters defining H and NEE, than between LE and NEE. Trade-offs between performance in LE and NEE were also found by Staudt et al. (2010) and Prihodko et al. (2008) in a forest and a forest complex including wetlands. However, only the effect of parameters, not the effect of model input (i.e. meteorological input data) variables on these processes were tested in both studies, as well as in ours. Such a confounding effect might also be the effect of result from a parameter value compensating for a process not implemented in the model. For example, parameter values that lead to an overestimation of NEE in spring result in higher transpiration and therefore better LE, whereas the reason for the underestimated LE during mid-April to mid-June (Fig. S1 in the supplement) might in fact be caused by

~~evaporation from, like e.g. evaporation from~~ open water bodies ~~that form on the peatland~~ during spring and early summer, ~~a process not implemented in the applied version of CoupModel.~~

~~, which might also explain the model data mismatch in LE during mid-April to mid-June (Fig. S1 in the supplement).~~

~~Trade-offs existed not only between different variables but also within the same variable, depending on whether ME,  $R^2$  of actual or  $R^2$  of accumulated values was chosen and which season was considered. This implies that the problems of a subjective criteria selection also exist if only one time series variable is considered.~~

~~Also several~~The detected supporting effects ~~were detected~~, indicating that some measurement variables can partly compensate absence or low resolution of a connected variable, even though they were not strong enough to make one variable fully redundant. For example, LAI measurements could reduce uncertainty in model predictions of the magnitudes of NEE, LE, H and WT on locations where these variables are not available. Tight relationships between plant and LAI, soil hydrology, C-fluxes and soil temperatures have been found by other model sensitivity studies as well (e.g. Ben Thouhami et al., 2013; Quillet 2013; Tian et al., 2014; Sándor et al., 2016) and strong correlations between LAI and NEE (Lund et al., 2010), and NEE and water availability (Reichstein et al., 2007) have also been found by data syntheses of eddy covariance sites. These relationships can be explained by the many dependencies between LAI and e.g. photosynthesis, transpiration, heat insulation and water uptake (Schulze 2006), of which several are also implemented in the model (see model description and equations, Sect. 2.3, Table S2 in the supplement and Jansson and Karlberg, 2010).

Other examples for detected supporting effects indicate that if H fluxes are available, the model is constrainable to produce improved WT dynamics, even if WT measurements were missing. High temporal resolution of soil temperature measurements in one layer are sufficient to model good temperatures if just the magnitude of soil temperature in an upper and a lower layer is known, e.g. due to short time or low resolution measurements.

The knowledge on supporting effects helps modellers in their site selection and in uncertainty estimation of model predictions depending on available ancillary data. It further can help experimentalists in their decisions which variables should and which need to be measured if the site should be usable for model constraint.

### 4.3 Equifinalities

The fit of model output to measured data in complex models is often not driven by a particular parameter but instead by interactions among parameters (e.g. Beven and Freer, 2001), which was also the case for several parameters in our study, hindering the constraint of parameters to a more narrow range. Also other carbon modelling studies found, that parameter values and sensitivities depend on the values of other parameters (e.g. Tatarinov & Cienciala 2006; Verbeeck et al., 2006; Quillet et al., 2013). This implies that especially if only few parameters and processes are calibrated (as in e.g. Yu et al., 2001; Wania et al., 2010, Zhu et al., 2014; Kim et al., 2014, Tang et al., 2015), resulting constrained ranges might not be comparable and transferable between models differing in their constant parameter values. Many equifinalities were identified, not only between parameters from the same module process category, but also ~~from across~~ different module process categories. This means that the problem of limited transferability also applies, if parameters from only one module process category are calibrated (as e.g. in Wang et al., 2005, Belassen et al., 2010, Wania et al., 2010, Sándor et al., 2016), or if models differ in the structures and implementations of their modules.

The knowledge on equifinalities is needed for a better parameter constraint in future calibrations as it allows calibration of the connected parameters in dependency of each other. Another way to respond to identified equifinalities is to calibrate only one of the connected parameters. However the resulting range will then not be transferable to other models using different values for connected, constant parameters.

Some equifinalities included several parameters, making their visualisation impossible and simple regression an insufficient tool for fully detecting and describing them (cf. Saltelli et al., 2008). These equifinalities need to be further investigated in additional calibrations which incorporate those parameter interactions and constrained ranges which were unambiguous, to

achieve a higher number of acceptable runs. This is needed, because the numbers of accepted runs in the final selections (50) did not allow a much more detailed analysis in such a complex model, as was apparent in comparison with the basic selection: An  $R^2$  threshold value of 0.1 was sufficient to identify equifinalities in the basic selection of 1286 accepted runs, but with just 50 accepted runs in the final selections, this threshold value could easily be exceeded by a random distribution, even that a higher threshold value of 0.15 was used. A threshold of 0.15 was on the other hand already too high, to detect for example the strong relationships between the plant parameters which were only clearly visible in the basic selection. Nevertheless the six equifinalities with  $R^2$  of higher than 0.30 are unambiguous [in this application of the CoupModel](#) and those with lower values are still very useful to design future calibrations to further investigate and describe these equifinalities.

#### 10 4.4 Usefulness of measurement variables

Models can be improved and their uncertainty reduced by calibrating their parameters to measurement data (e.g. Friend et al., 2007; Wang et al., 2009; Williams et al., 2009). We tested the [usefulness-potential](#) of several measurement variables (NEE, LAI, LE, H, Rn, Ts, WT and snow depth) and found all contributing to a better parameter constraint.

15 Thereby none of the variables could be fully replaced by another. Due to the strong interactions and as parameters of each [modelprocess categorie](#) were constrained by several different variables, ancillary variables are valuable even if only one certain process is of interest. In case of snow, our results suggest that data on snow cover might be sufficient, if snow depth is not available.

20 In a forest site simulation with the ORCHIDEE model, H and Rn were found to be redundant for constraining energy balance parameters if NEE and LE were available (Santaren et al., 2007). In contrast, some energy balance related parameters in our study were constrained by exclusively Rn and H, or additionally by LE but with different resulting ranges. This reveals the usefulness of Rn and H measurements for model constraints and shows that variables which might have been identified as redundant in one study could be of high importance on another ecosystem or for another model calibrating a different parameter selection.

25 Several influential parameters could not be unambiguously constrained or showed equifinalities and need additional measurements to be further investigated. This includes soil water content or soil water retention properties, as well as canopy albedo and leaf litter fall during the growing season. Except for water retention properties these variables are needed as time series throughout the year. A more detailed discussion of the benefit of such measurements can be found in the following sections.

#### 4.5 Detailed discussion of sensitivities and interactions per process

30 The parameters that were identified as most influential or that showed the strongest equifinalities were related to soil hydrology and water content, to a stable representation of the plant, to radiation, temperature and heat fluxes or to snow. [As only one parameter per equation was calibrated, a high sensitivity to this parameter means a high sensitivity to the corresponding process. Some of such process sensitivities might be also be interesting for other models and similar ecosystems.](#)

35 The introduced index to measure parameter concern includes subjective choices like weighting factors, the choice of considered calibration variables and their sub periods as well as the chosen performance indices. However several tested variations in especially the weighting did not noticeable change the results:  $\psi_a$  was always the most important parameter, followed by the group of parameters with medium importance which differed slightly in their ranking among each other.

#### 4.5.1 Unsaturated water distribution & soil moisture conditions

Our results suggest that model uncertainty could be greatly reduced if data for either soil hydraulic properties, water content or plant transpiration characteristics were available: Despite available data of detailed WT and LE in our study, large uncertainty remained in simulated water content due to the combined uncertainty in estimates of soil hydraulic properties ( $\psi_a$ ) and plant water uptake ( $g_{max,vasc}$ ,  $g_{max,moss}$ ,  $g_{maxwin}$ ). Their sensitivity to many variables and the high number of equifinalities hindered the constraint of other parameters and therefore the uncertainty reduction in all involved processes. For example this might explain why the water response functions for neither plant assimilation nor soil respiration could be constrained.

The shape parameter of the water retention curve ( $\psi_a$ ) was among the top two most sensitive parameters for NEE, WT, LE, H, Ts, and the third and fifth most sensitive parameter in case of Rn and snow. That confirms the importance of the implemented interactions of soil moisture with water and heat fluxes, soil temperature, assimilation and respiration processes, as reported from empirical studies (Kim and Verma, 1996; Bridgham et al., 1999; Tezara et al., 1999; Kellner, 2001; Flangan and Johnson 2005; Lafleur et al., 2005; Schulze, 2006; Belyea 2009).

Also, the transpiration coefficients ( $g_{max,vasc}$ ,  $g_{max,moss}$ ,  $g_{maxwin}$ ) were among the top ~~10~~-ten most important and influential parameters. In case of vascular plants, they correspond to the stomatal conductance parameter in other models, which was shown to be crucial for modelling NEE, biomass, LE or H in other studies (Esprey et al., 2004 for forest stand volume, Tatarinov and Cienciala, 2006 for NEE and carbon pools; Staudt et al., 2010 for NEE, LE and H; Hidy et al., 2012 for carbon fluxes and LE; Bonan et al., 2011 and Tian et al., 2014 for LH and H). The control of stomatal conductance on transpiration and photosynthesis has also been emphasised by several empiric studies (e.g. Jarvis & Morison 1981, Quick et al., 1992, Tezara et al., 1999, Yordanov et al., 2000).

The strong sensitivity of  $\psi_a$ ,  $g_{max,vasc}$ ,  $g_{max,moss}$ ,  $g_{maxwin}$  for many processes is especially remarkable as parameters and parameter combinations could only vary to such an extent that the water level fitted the measurements as restricted by the basic selection.

The importance of water table on NEE fluxes has widely been mentioned (e.g. Silvola et al., 1996; Yurova et al., 2007, Kurbatova et al., 2009; Dušek et al., 2012) but our results point out that the knowledge on WT alone is not sufficient for model calibration and reliable predictions. In addition also measurements of soil hydraulic properties are crucial for model calibration. The usefulness of water retention properties for modelling carbon dynamics was also found by other sensitivity analyses on peatlands as well as on mineral soils (e.g. Wang et al., 2005; Pappas et al., 2013, Quillet et al., 2013). Nevertheless, many of the available peatland sites in current databases (e.g. European Fluxnet Database Cluster, <http://gaia.agraria.unitus.it>) still do not contain information on water retention properties or water content.

We ~~therefore~~ strongly recommend experimentalists to include water retention measurements in their experimental set up. Thereby, the horizontal and vertical ~~variability~~variability in peat hydraulic properties needs to be accounted for (Baird et al., 2012, Waddington et al., 2015). Such measurements might also help to resolve the strong equifinalities of  $\psi_a$  with transpiration coefficients and a parameter in the calculation of aerodynamic resistance of the plant canopy, defining the minimum exchange under stable conditions ( $C_{H0,canopy}$ ).

#### 4.5.2 C balance of vascular plants

A stable vascular plant that establishes a reasonable amount of biomass every year throughout the simulation period, could only be achieved by certain value combinations for the photosynthetic efficiency ( $\epsilon_{L,vasc}$ ), the respiration coefficient ( $k_{gresp,vasc}$ ) and the storage fraction for plant regrowth in spring ( $m_{retain}$ ). Despite their high impact in the basic selection, neither equifinalities, nor sensitivities of these parameters reached high measures in final selections, probably because several parameters were interacting simultaneously. This indicates the need for either calibrating these parameters in

dependency of each other or setting at least one of them to a constant value, as the available data was not sufficient to resolve these equifinalities. Many studies on other ecosystems have found NEE or biomass to be strongly sensitive to a parameter corresponding to photosynthetic efficiency ( $\epsilon_{L,vasc}$ ) (Esprey et al., 2004, Verbeeck et al., 2006; Prihodko et al., 2008; Staudt et al., 2010; Bonan et al., 2011; Pappas et al., 2013, Tian et al., 2014, Xenakis et al., 2008), but were performed without a simultaneous calibration of parameters related to plant respiration and storage for regrowth. Pappas et al. (2013) discussed a possible overestimation of model sensitivity to photosynthetic efficiency due to processes that are not implemented like the active simulation of plant growth including growth limitations. A strong negative correlation between two of the parameters (plant respiration and photosynthetic efficiency) was also found in a sensitivity analysis using the LPJ model (Zaehle et al., 2005).

10 Despite their effect on model performance,  $\epsilon_{L,vasc}$ ,  $k_{gresp,vasc}$  and  $m_{retain}$  had a low rank in parameter concern, as ranges for these parameters could be narrowed unambiguously due to well overlapping ranges between the different variables. Nevertheless, these parameters would be of high importance for predictions, if none of the constraining variables are available.

15 Compared to a previous application of the CoupModel on five different open peatlands including different management intensities (Metzger et al., 2015), vascular plants had to have a much more effective C household to produce the measured leaf area given a limited amount of assimilates. This can be realised by low respiration and litter fall losses and a large storage pool for regrowth in spring. Even if respiration losses from vascular plants were 1/10 of the ones used at the sites in Metzger et al. (2015), the model tended to either underestimate vascular plant LAI, or overestimate CO<sub>2</sub> uptake (Fig 2). A possible explanation for the differences in parameter value combination of vascular plants might lie in the vegetation communities. Despite Metzger et al. (2015) included several different types of treeless peatland vegetation communities, none of these sites had a similar vegetation community typical for nutrient poor habitats, consisting of mainly mosses and *Eriophorum vaginatum*, as at Degerö Stormyr. *Eriophorum vaginatum* is known to be much more effective in maintaining C compared to other sedges and having a highly efficient remobilization from senescing leaves (Shaver and Laundre, 1997; Jonasson and Chapin III, 1985). Uncertainties in measurements and the distribution of modelled respiration over the hours of the day might accelerate or diminish this effect. Explanations by differences in model structure can be excluded, as the same effect was observed when using exactly the same structure (unpublished data). To identify the difference between the sites, which causes the deviations in the combined parameter value ranges, the model need to be applied to further open peatland sites differing in vegetation community, nutrient status and plant productivity. This might allow finding trends in parameter ranges, which is a necessary precondition for estimation and reducing model uncertainty in predictions on other peatland sites.

30 Another plant parameter which was important for a stable vascular plant layer and was ranked as one of the overall most important parameters was the rate coefficient for the leaf litter fall during the growing season ( $l_{Lcl}$ ). Probably due to the high number of correlations with other parameters, these correlations did not exceed the threshold value.  $l_{Lcl}$  is directly connected to the filling of the storage pool, but also for maintaining C in the leaves. The strong sensitivity of LAI to  $l_{Lcl}$  affects transpiration and thereby water uptake which explains the strong sensitivity to WT depths below -0.2 m and the equifinalities with a transpiration parameter and a parameter describing the response of heterotrophic respiration to water. In Metzger et al. (2015), a value of  $l_{Lcl} = 0.01 \text{ day}^{-1}$  could be used site independent. This contradicts the much lower ranges of  $l_{Lcl}$  in our study, necessary for acceptable performance in several variables, in particular R<sup>2</sup> of LAI, WT depths below -0.2 m and ME of spring time NEE. However, species in nutrient-poor habitats are associated with longer-lived leaves than those of nutrient-rich habitats (Ryser 1996) and fast growing species (Reich et al., 1992), whereas *Eriophorum vaginatum* in particular is known for long-lived leaves and therefore have a very low litter fall rate (Jonasson & Chapin 1985). Less complex models as the GUESS-ROMUL model, which was also applied to this site, use annual accumulated NEE as estimate for litter fall (Yurova et al., 2007) which is therefore directly dependent on site productivity. Only one site in



Metzger et al. (2015) had lower annual NEE compared to Degerö Stormyr, but this is probably a result of the shorter vegetation period at that site, whereas a site with similar annual NEE was formerly drained, so that the soil respiration contribution to NEE is much larger, compensating the larger productivity. A high sensitivity of litter fall rate to plant biomass and soil carbon pools was also found by Xenakis et al. (2008) using the 3-PG model on forest.

- 5 Further investigations including model applications to additional sites are needed to resolve the differences in resulting ranges and equifinalities with other parameters. Thereby, measurements of leaf litter fall throughout the year would be of high value.

#### 4.5.3 Sensible heat fluxes, soil temperatures and net radiation

- 10 The large number of strong connections between H, Ts and Rn and the equifinalities between their determining parameters indicate the importance to consider, model and calibrate the related processes together. However the constraint of two of the most important parameters (aerodynamic resistance dependency on LAI,  $r_{alai}$  and moss albedo,  $a_{pve,moss}$ ), failed not due to different ranges between variables but due to the differences depending on which performance index and season was considered. This emphasises the importance of the subjective criteria choice, even if only one variable is considered.

- 15 Accepted values for  $r_{alai}$  were exceptionally high (200 s m<sup>-1</sup> for Ts R<sup>2</sup> and 550 to 800 s m<sup>-1</sup> for Ts<sub>1</sub> ME, whereas a  $r_{alai}$  of 200 multiplied with the moss LAI of 1.8 leads to an aerodynamic resistance of 360 s m<sup>-1</sup>). Mosses might form a well insulating layer, but still the values are much higher than the aerodynamic resistance estimates for this site (approximately 50 s m<sup>-1</sup>, Peichl et al., 2013) or of a bog in South-Sweden (60 s m<sup>-1</sup>, Kellner, 2001). Price (1991) reported very high resistance, when moss surface moisture is low, e.g. during dry periods, but these values were still lower than ours. A possible explanation  
20 might be an interaction with a non-calibrated, fixed parameter. A high aerodynamic resistance causes better temperature insulation leading to higher summer soil temperatures with lower diurnal oscillations. Further, it leads to strongly reduced soil evaporation and therefore reduced LE, even though this is partly compensated by slightly higher transpiration from mainly mosses, which profit from the higher water contents in upper soil layers. This explains the sensitivities to WT and LE which also supported a higher  $r_{alai}$  value. The main cause for the much lower optimum range for dynamics in Ts compared to  
25 magnitude in Ts is probably an overestimation of the diurnal amplitude. A lower moss LAI can reduce this overestimation, but the corresponding parameter was not calibrated to avoid further equifinalities:  $r_{alai}$  showed already strong interactions with  $a_{pve,moss}$  and  $z_{0M,snow}$ . The correlations of the conductivity of organic material ( $h_2$ ) with plant, LE and WT parameters might be explained by the dependency of thermal conductivity from peat wetness (Kellner, 2001).

- Seasonal differences in moss albedo ( $a_{pve,moss}$ ) could be expected as their radiation reflection properties vary with moss water  
30 content (Graham et al., 2006). However higher values would be expected in summer, when the moss surface is dry and lighter, but our calibration resulted in higher values during spring and winter. These values were much higher (>22%) compared to literature values (11–16.5%, Berglund and Mace, 1972; 16.4%, Zhao et al., 1997; 11%, Kellner, 2001) and therefore rather compensate for values of interacting parameters (in particular  $z_{0M,snow}$  and  $r_{alai}$ ) or not implemented processes. Especially the effect on winter H and Rn might result from the strong interaction with  $z_{0M,snow}$ , as the mosses in winter are  
35 covered with a thick snow cover, so that their albedo shouldn't show any sensitivity in winter. Further, H in spring tended to be overestimated, which would be compensated by a high albedo during this time, but might be caused in the real world by open water over frozen soil, which was not realized in the model. Interestingly, albedo of vascular plants did not show any sensitivities, neither during vegetative stage ( $a_{pve,vase}$ ), nor after start of senescence ( $a_{pgrain}$ ) when a higher value would have been expected due to leaf yellowing. Direct measurements of plant albedo were not available in this study. A time series  
40 observation of those would be very helpful for clarification, as this parameter is known to vary substantially within and between peatlands (Belyea et al., 2009).

#### 4.5.4 Snow

The model performance in simulating snow depth was not connected to performance in any other variable, except to performance in H if exclusively spring time values were considered. This was surprising, as the uncertainty for timing of snow melt ranged for about two weeks but determined the start of temperature rise, water table dropping and biotic activity.

5 A possible explanation might be the poor ability of snow depth  $R^2$  and ME to assert a good fit in duration of snow cover. This is supported by the fact that the most important parameter for timing of snow melt ( $m_T$ ) strongly affected performance in dynamics of H, NEE and Ts during spring time. Parameters defining timing of snow depth might be better constrained if future calibrations include an additional variable with a stronger conclusiveness to the timing of snow melt, e.g. by a boolean time series indicating if snow cover is present or not. It needs to be tested if this could also help to solve the disagreements in  
10 value ranges between the performance indices in case of the density coefficient of old snow ( $S_{dw}$ ) which caused in combination with  $m_T$  the low average overlap within snow depth sensitive parameters.

According to Jansson and Karlberg (2010), a high value for  $m_T$  ( $4\text{--}6 \text{ kg } ^\circ\text{C}^{-1} \text{ m}^{-2} \text{ day}^{-1}$ ) could be expected for open fields. A possible explanation for the low accepted values ( $<3 \text{ kg } ^\circ\text{C}^{-1} \text{ m}^{-2} \text{ day}^{-1}$ ) of  $m_T$  in case of criteria on H in contrast to the high values if criteria were on Ts, could be that high values compensate for overestimated spring time H (cf. Fig. S1 in the  
15 supplement). However, the overestimation of spring H might be connected to different reflection properties of mosses during spring time or to missing consideration of radiation reflection and evaporation from open water which might be formed during snow melt on still frozen soils. The latter is further supported by the underestimation of LE during April and May (Fig. S1 in the supplement), which cannot be connected to underestimated plant transpiration, as the model even tended to overestimate  $\text{CO}_2$  uptake during this period.

## 20 5 Conclusions

~~CO<sub>2</sub> models are often commonly calibrated on NEE as only measurement variable. Here, w~~We investigated the interactions between different abiotic and biotic processes and their parameters, as well as the implications and usefulness of data on not only NEE, but also LAI, sensible and latent heat fluxes, radiation, water table depth, soil temperatures and snow depth for model calibration on a boreal peatland. ~~Different P~~processes and ~~their parameters as well as~~ model performance ~~between in~~  
25 ~~the~~ different observation variables were strongly interlinked across process categories. This means parameter ranges that result from calibration depend on model structure, included processes, other parameter values and calibration setup, and might therefore not be transferable between studies. It further implies that a study aiming to understand and interpret parameter values need to calibrate processes and parameters of many different process categories, using a wide range and multiple criteria on various observation variables. ~~This needs to be taken into account in model calibrations and when transferring calibrations results between models differing in their structure or in their constant parameters.~~

The key parameters identified will help to simplify future model calibrations by selecting only the most influential parameters for the variable of interest and using a narrower range for the constrained parameters. This means a simpler calibration and faster computation and in turn, allows the inclusion of a more detailed investigation of a process of certain interest. Further, it helps model developers to include the most sensitive processes for simulating a certain variable. ~~On the other hand, our results revealed the strong dependence of constrained parameter ranges to other parameters and to the chosen criteria. This means, that a study aiming to understand and interpret parameter values need to calibrate processes and parameters of many different modules, using a wide range and multiple criteria on various observation variables.~~  
35

Parameter interactions were found to be more important than parameter value ranges, revealing the need for accounting for equifinalities, also across different biotic and abiotic processes: Either by calibrating correlated parameters in dependence of  
40 each other or by calibrating only one of the correlated parameters. The latter will lead to a narrower constrained range, but this range might not be transferable to other sites and other models.

The gained knowledge on trade-offs will be useful to avoid modelling studies with too many purposes and helps model users assessing the implications of their criteria choice. The validity of calibrated models is always restricted and robustness of obtained parameter ranges should be questioned.

The identified supporting effects between some variables indicated that some measurement variables can partly compensate absence or low resolution of the connected variable. This information tells experimentalists which measurement variables are helpful and which are obligatory if a certain process should be understood from the underlying regulating principles. It further helps modellers to decide if a site has enough available data for model calibration and to estimate uncertainties in model predictions depending on available ancillary data.

All observed calibration variables (NEE, LAI, sensible and latent heat fluxes, net radiation, soil temperatures, water table depth and snow depth) helped for model constraint and interpretation. Ancillary variables are in particular important for evaluating the robustness of calibrated parameter ranges. They should therefore be measured on sites used for calibration of complex process oriented models. Additional measurements of, in particular, soil hydraulic properties or water content would largely reduce uncertainty and help for a better parameter constraint.

### Code and data availability

The model and extensive documentation can be downloaded from the CoupModel homepage <http://www.coupmodel.com/>. The source code can be requested for non-commercial purposes from Per-Erik Janson ([pej@kth.se](mailto:pej@kth.se)). The simulation files including the model and calibration setup, the used parameterisation and corresponding input and validation files can be requested from Christine Metzger ([cmetzger@kth.se](mailto:cmetzger@kth.se)). They cannot be made freely public available, as they include climate and site data that require authorisation from the data owners.

The flux data and ancillary data are available from the European Flux Database Cluster (<http://www.europe-fluxdata.eu/>), site name: Degerö, Site code: SE-Deg, with open data access for the years 2001–2006, and restricted data access (the Principal Investigator of the site has to authorize the data request) for the years 2007–2015.

### Acknowledgements

This study was financed by FORMAS (grant no. 2009-872), project LAGGE – Landscape Greenhouse Gas Exchange – Integration of Terrestrial and Freshwater sources and sinks, coordinated by Leif Klemmedtsson.

The authors thank Mikael Ottosson Löfvenius and Jörgen Sagerfors from the Swedish University of Agricultural Sciences ([SLU](http://www.slu.se)) for their great effort in collecting and providing the model input and calibration data. Climate data from 1990 to 2000 was provided by “reference climate monitoring program at Vindeln experimental forests, SLU, Umeå”.

The measurements were funded by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (grant no. 21.4/2003-0876 and no. 2007-666) and the Swedish Research Council (grant no. 621-2003-2730). We also acknowledge the Kempe Foundation for the grants supporting the micrometeorological instrumentation. Support from the ICOS Sweden (Integrated Carbon Observation System) and SITES (Swedish Infrastructure for Ecosystem Research) research infrastructure (Swedish Research Council) is also acknowledged.

## References

- Abdalla, M., Hastings, A., Bell, M. J., Smith, J. U., Richards, M., Nilsson, M. B., Peichl, M., Löfvenius, M. O., Lund, M., Helfter, C., Nemitz, E., Sutton, M. A., Aurela, M., Lohila, A., Laurila, T., Dolman, A. J., Belelli-Marchesini, L., Pogson, M., Jones, E., Drewer, J., Drosler, M., and Smith, P.: Simulation of CO<sub>2</sub> and Attribution Analysis at Six European Peatland Sites Using the ECOSSE Model, *Water Air Soil Pollut*, 225, 1–14, doi:10.1007/s11270-014-2182-8, 2014.
- Alexandersson, H., Karlström, C., and Larsson-McCann S.: Temperaturen och nedercörden i sverige 1961–1990 (Swedish). Temperature and Precipitation in Sweden 1961–1990. Reference Normals Meteorologi 81, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden, 1991
- Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the Annual Net Carbon and Water Exchange of Forests: The EUROFLUX Methodology, in: *Advances in Ecological Research Volume 30, Advances in Ecological Research*, Elsevier, 113–175, 1999.
- Aurela, M.: The timing of snow melt controls the annual CO<sub>2</sub> balance in a subarctic fen, *Geophys. Res. Lett.*, 31, doi:10.1029/2004GL020315, 2004.
- Baird, A. J., Morris, P. J., and Belyea, L. R.: The DigiBog peatland development model 1: Rationale, conceptual model, and hydrological basis, *Ecohydrol.*, 5, 242–255, doi:10.1002/eco.230, 2012.
- Bazin, M. J. and Prosser, J. I.: *Physiological models in microbiology*, CRC series in mathematical models in microbiology, CRC Press, Boca Raton, Fla., 2 volumes, 1988.
- Bellassen, V., Le Maire, G., Dhôte, J. F., Ciais, P., and Viovy, N.: Modelling forest management within a global vegetation model—Part 1: Model structure and general behaviour, *Ecological Modelling*, 221, 2458–2474, doi:10.1016/j.ecolmodel.2010.07.008, 2010.
- Belyea, L. R.: Nonlinear dynamics of peatlands and potential feedbacks on the climate system, in: *Carbon Cycling in Northern Peatlands*, Baird, A. J., Belyea, L. R., Comas, X., Reeve, A. S., Slater, L. D. (Eds.), *Geophysical Monograph Series*, American Geophysical Union, Washington, D. C., 5–18, 2009.
- Ben Touhami, H., Lardy, R., Barra, V., and Bellocchi, G.: Screening parameters in the Pasture Simulation model using the Morris method, *Ecological Modelling*, 266, 42–57, doi:10.1016/j.ecolmodel.2013.07.005, 2013.
- Berglund, E. R. and Mace, A. C.: Seasonal Albedo Variation of Black Spruce and Sphagnum-Sedge Bog Cover Types, *Journal of applied meteorology*, 806–812, 1972.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249, 11–29, doi:10.1016/s0022-1694(01)00421-8, 2001.
- Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, *J. Geophys. Res.*, 116, doi:10.1029/2010JG001593, 2011.
- Bridgman, S. D., Pastor, J., Updegraff, K., Malterer, T. J., Johnson, K., Harth, C., and CHEN, J.: ECOSYSTEM CONTROL OVER TEMPERATURE AND ENERGY FLUX IN NORTHERN PEATLANDS, *Ecological Applications*, 9, 1345–1358, doi:10.1890/1051-0761(1999)009[1345:ECOTAE]2.0.CO;2, 1999.
- Brooks, R. H. and Corey, A. T.: *Hydraulic Properties of Porous Media*, Hydrology Papers, Colorado State University, 1964.
- Carvalho, N., Reichstein, M., Ciais, P., Collatz, G. James, Mahecha, M. D., Montagnani, L., Papale, D., Rambal, S., and Seixas, J.: Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints, *Glob Change Biol*, 16, 2813–2829, doi:10.1111/j.1365-2486.2010.02173.x, 2010.
- ~~Clymo, R. S. and Hayward, P. M.: The Ecology of Sphagnum, in: *Bryophyte Ecology*, Smith, A. J. E. (Ed.), Springer Netherlands, 229–289, 1982.~~
- Clymo, R. S.: The Limits to Peat Bog Growth, *Royal Society Publishing*, 303, 605–654, 1984.
- CoupModel, Current version of COUP model for download, available at model homepage: “<http://www.coupmodel.com>”, last accessed: 20 October 2015
- ~~Davidson, H. R. and Campbell, C. A.: The effect of temperature, moisture and nitrogen on the rate of development of spring wheat as measured by degree days, *Canadian journal of plant science*, 63, 833–846, 1983.~~
- Dettmann, U., Bechtold, M., Frahm, E., and Tiemeyer, B.: On the applicability of unimodal and bimodal van Genuchten–Mualem based models to peat and other organic soils under evaporation conditions, *Journal of Hydrology*, 515, 103–115, doi:10.1016/j.jhydrol.2014.04.047, 2014.
- Dimitrov, D. D., Grant, R. F., Lafleur, P. M., and Humphreys, E. R.: Modeling the effects of hydrology on ecosystem respiration at Mer Bleue bog, *Journal of Geophysical Research: Planets*, 115, 2010.
- Drösler, M., Verchot, L. V., Freibauer, A., Pan, G., Evans, C. D., Bourbonniere, R. A., Alm, J. P., Page, S., Agus, F., Hergoualch, K., Couwenberg, J., Jauhiainen, J., Sabiham, S., Wang, C., Srivastava, N., Borgeau-Chavez, L., Hooijer, A., Minkinen, K., French, N., Strand, T., Sirin, A., Mickler, R., Tansey, K., and Larkin, N.: Drained Inland Organic Soils (chapter 2), in: *2013 Supplement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Wetlands, Methodological Guidance on Lands with Wet and Drained Soils, and Constructed Wetlands for Wastewater Treatment*, Hiraiishi, T., Krug, T., Tanabe, K., Srivastava, N., Jamsranjav, B., Fukuda, M., Troxler, T. (Eds.), IPCC, Switzerland, 2.1-2.76, 2013.

- Dušek, J., Čížková, H., Stellner, S., Czerný, R., and Květ, J.: Fluctuating water table affects gross ecosystem production and gross radiation use efficiency in a sedge-grass marsh, *Hydrobiologia*, 692, 57–66, doi:10.1007/s10750-012-0998-z, 2012.
- 5 Esprey, L., Sands, P., and Smith, C.: Understanding 3-PG using a sensitivity analysis, *Forest Ecology and Management*, 193, 235–250, doi:10.1016/j.foreco.2004.01.032, 2004.
- Flanagan, L. B. and Johnson, B. G.: Interacting effects of temperature, soil moisture and plant biomass production on ecosystem respiration in a northern temperate grassland, *Agricultural and Forest Meteorology*, 130, 237–253, doi:10.1016/j.agrformet.2005.04.002, 2005.
- 10 Friend, A. D., Arneeth, A., Kiang, N. Y., Lomas, M., OGÉE, J., Rödenbeck, C., Running, S. W., Santaren, J.-D., Sitch, S., Viovy, N., Woodward, I. F., and Zaehle, S.: FLUXNET and modelling the global carbon cycle, *Glob Change Biol*, 13, 610–633, 2007.
- Frolking, S. E., Roulet, N. T., Moore, T. R., Lafleur, P. M., Bubier, J. L., and Crill, P. M.: Modeling seasonal to annual carbon balance of Mer Bleue Bog, Ontario, Canada, *Global Biogeochem. Cycles*, 16, 4-1–4-21, doi:10.1029/2001gb001457, 2002.
- 15 Frolking, S. E., Roulet, N. T., Tuittila, E., Bubier, J. L., Quillet, A., Talbot, J., and Richard, P. J. H.: A new model of Holocene peatland net primary production, decomposition, water balance, and peat accumulation, *Earth System Dynamics*, 1, 1–21, doi:10.5194/esd-1-1-2010, 2010.
- ~~Fulkerson, W. J. and Donaghy, D. J.: Plant soluble carbohydrate reserves and senescence key criteria for developing an effective grazing management system for ryegrass-based pastures: A review, *Animal Production Science*, 41, 261–275, 2001.~~
- 20 Giltrap, D. L., Li, C., and Sagar, S.: DNDC: A process-based model of greenhouse gas fluxes from agricultural soils, *Agriculture, Ecosystems & Environment*, 136, 292–300, doi:10.1016/j.agee.2009.06.014, 2010.
- Gong, J., Shurpali, N. J., Kellomäki, S., Wang, K., Zhang, C., Salam, M. M. A., and Martikainen, P. J.: High sensitivity of peat moisture content to seasonal climate in a cutaway peatland cultivated with a perennial crop (*Phalaris arundinacea*, L.): A modeling study, *Agricultural and Forest Meteorology*, 180, 225–235, doi:10.1016/j.agrformet.2013.06.012, 2013.
- 25 Gorham, E.: Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming, *Ecological Applications*, 1, 182–195, 1991.
- Graham, E. A., Hamilton, M. P., Mishler, B. D., Rundel, P. W., and Hansen, M. H.: Use of a Networked Digital Camera to Estimate Net CO<sub>2</sub> Uptake of a Desiccation-Tolerant Moss, *Int. J Plant Sci.*, 167, 751–758, doi:10.1086/503786, 2006.
- 30 Granberg, G., Grip, H., Löfvenius, M. O., Sundh, I., Svensson, B. H., and Nilsson, M.: A simple model for simulation of water content, soil frost, and soil temperatures in boreal mixed mires, *Water Resour. Res.*, 35, 3771–3782, doi:10.1029/1999WR900216, 1999.
- Grant, R. F., Humphreys, E. R., Lafleur, P. M., and Dimitrov, D. D.: Ecological controls on net ecosystem productivity of a mesic arctic tundra under current and future climates, *Journal of Geophysical Research: Planets*, 116, 2011.
- 35 Hansen, G. K. and Jensen, C. R.: Growth and Maintenance Respiration in Whole Plants, Tops, and Roots of *Lolium multiflorum*, *Physiol. Plant.*, 39, 155–164, 1977.
- ~~He, H., Jansson, P.-E., Svensson, M., Björklund, J., Tarvainen, L., Klemedtsson, L., and Kasimir, Å.: Forests on drained agricultural peatland are potentially large sources of greenhouse gases—insights from a full rotation period simulation, *Biogeosciences Discuss.*, 12, 19673–19710, doi:10.5194/bgd-12-19673-2015, 2015.~~
- 40 ~~He, H., Jansson, P.-E., Svensson, M., Björklund, J., Tarvainen, L., Klemedtsson, L., and Kasimir, Å.: Forests on drained agricultural peatland are potentially large sources of greenhouse gases – insights from a full rotation period simulation, *Biogeosciences*, 13, 2305–2318, doi:10.5194/bg-13-2305-2016, 2016.~~
- Hidy, D., Barcza, Z., Haszpra, L., Churkina, G., Pintér, K., and Nagy, Z.: Development of the Biome-BGC model for simulation of managed herbaceous ecosystems, *Ecological Modelling*, 226, 99–119, doi:10.1016/j.ecolmodel.2011.11.008, 2012.
- 45 Hilbert, D. W., Roulet, N., and Moore, T.: Modelling and analysis of peatlands as dynamical systems, *J Ecology*, 88, 230–242, doi:10.1046/j.1365-2745.2000.00438.x, 2000.
- Impens, I. and Lemeur, R.: Extinction of net radiation in different crop canopies, *Arch. Met. Geoph. Biokl. B.*, 17, 403–412, doi:10.1007/BF02243377, 1969.
- 50 Ivanov, K. E., Thomson, A., and Ingram, H.: Water movement in mirelands, Academic Press, London, 276 pp., 1981.
- Jansson, P.-E.: CoupModel: Model use, Calibration, and Validation, *Transactions of the ASABE*, 55, 1335–1344, 2012.
- Jansson, P.-E.: CoupModel: Model use, Calibration, and Validation, *Transactions of the ASABE*, 55, 1335–1344, 2012.
- Jansson, P.-E. and Halldin, S.: Model for the annual water and energy flow in a layered soil, Comparison of forest and energy exchange models. Society for Ecological Modelling, Copenhagen, 145–163, 1979.
- 55 Jansson, P.-E. and Karlberg, L.: Coupled heat and mass transfer model for soil–plant–atmosphere systems. Royal Institute of Technology, Stockholm, 484 pp., accessed: 17 November 2015, <https://drive.google.com/file/d/0B0-WJKp0fmYCYZ0JVeVgzRWFbUk/view?pli=1>, 2010.
- Jarvis, P. G. and Morison, J. I.: The control of transpiration and photosynthesis by the stomata, *Stomatal physiology*, 247–279, 1981.
- 60 Jonasson, S. and Stuart Chapin, F. III: Significance of sequential leaf development for nutrient balance of the cotton sedge, *Eriophorum vaginatum* L, *Oecologia*, 67, 511–518, doi:10.1007/BF00790022, 1985.
- Keddy, P. A.: Assembly and response rules: two goals for predictive community ecology, *Journal of Vegetation Science*, 3, 157–164, doi:10.2307/3235676, 1992.



- Keller, J. K., White, J. R., Bridgman, S. D., and Pastor, J.: Climate change effects on carbon and nitrogen mineralization in peatlands through changes in soil quality, *Global Change Biology*, 10, 1053–1064, doi:10.1111/j.1529-8817.2003.00785.x, 2004.
- 5 Kellner, E.: Surface energy fluxes and control of evapotranspiration from a Swedish Sphagnum mire, *Agricultural and Forest Meteorology*, 110, 101–123, doi:10.1016/S0168-1923(01)00283-0, 2001.
- Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, *J Royal Statistical Soc B*, 63, 425–464, doi:10.1111/1467-9868.00294, 2001.
- Kim, J. and Verma, S. B.: Surface exchange of water vapour between an open sphagnum fen and the atmosphere, *Boundary-Layer Meteorol*, 79, 243–264, doi:10.1007/BF00119440, 1996.
- 10 Kim, Y., Roulet, N., Peng, C., Li, C., Frolking, S., Strachan, L. B., and Tremblay, A.: Multi-Year carbon dioxide flux simulations for mature Canadian black spruce forests and ombrotrophic bogs using Forest-DnDc, *Boreal Environ. Res*, 19, 417–440, 2014.
- Kurbatova, J., Li, C., Tatarinov, F. A., Varlagin, A., Shalukhina, N., and Olchev, A.: Modeling of the carbon dioxide fluxes in European Russia peat bogs, *Environmental Research Letters*, 4, 45022, 2009.
- 15 Kurnianto, S., Warren, M., Talbot, J., Kauffman, B., Murdiyarso, D., and Frolking, S.: Carbon accumulation of tropical peatlands over millennia: a modeling approach: A modeling approach, *Global change biology*, 2014.
- Lafleur, P. M., Moore, T. R., Roulet, N. T., and Frolking, S. E.: Ecosystem Respiration in a Cool Temperate Bog Depends on Peat Temperature But Not Water Table, *Ecosystems*, 8, 619–629, doi:10.1007/s10021-003-0131-2, 2005.
- Laine, A. M., Bubier, J., Riutta, T., Nilsson, M. B., Moore, T. R., Vasander, H., and Tuittila, E.-S.: Abundance and composition of plant biomass as potential controls for mire net ecosystem CO<sub>2</sub> exchange, *Botany*, 90, 63–74, doi:10.1139/b11-068, 2012.
- 20 Lindroth, A., Lund, M., Nilsson, M., Aurela, M., Christensen, T. R., Laurila, T., Rinne, J., Riutta, T., Sagerfors, J., Ström, L., Tuovinen, J.-P., and Vesala, T.: Environmental controls on the CO<sub>2</sub> exchange in north European mires, *Tellus B*, 59, 812–825, doi:10.1111/j.1600-0889.2007.00310.x, 2007.
- 25 Loisel, J., Yu, Z., Beilman, D. W., Camill, P., Alm, J., Amesbury, M. J., Anderson, D., Andersson, S., Bochicchio, C., Barber, K., Belyea, L. R., Bunbury, J., Chambers, F. M., Charman, D. J., Vleeschouwer, F. de, Fialkiewicz-Koziel, B., Finkelstein, S. A., Galka, M., Garneau, M., Hammarlund, D., Hinchcliffe, W., Holmquist, J., Hughes, P., Jones, M. C., Klein, E. S., Kokfelt, U., Korhola, A., Kuhry, P., Lamarre, A., Lamentowicz, M., Large, D., Lavoie, M., MacDonald, G., Magnan, G., Makila, M., Mallon, G., Mathijssen, P., Mauquoy, D., McCarroll, J., Moore, T. R., Nichols, J., O'Reilly, B., Oksanen, P., Packalen, M., Peteet, D., Richard, P. J., Robinson, S., Ronkainen, T., Rundgren, M., Sannel, A. B. K., Tarnocai, C., Thom, T., Tuittila, E.-S., Turetsky, M., Valiranta, M., van der Linden, M., van Geel, B., van Bellen, S., Vitt, D., Zhao, Y., and Zhou, W.: A database and synthesis of northern peatland soil properties and Holocene carbon and nitrogen accumulation, *The Holocene*, 24, 1028–1042, doi:10.1177/0959683614538073, 2014.
- 30 Lund, M., Lafleur, P. M., Roulet, N. T., Anders, L., Christensen, T. R., Aurela, M., Chojnicki, B. H., Lawrence, F. B., Humphreys, E. R., Laurila, T., Oechel, W. C., Olejnik, J., Rinne, J., Schubert, P. E. R., and Nilsson, M. B.: Variability in exchange of CO<sub>2</sub> across 12 northern peatland and tundra sites, *Glob Change Biol*, 2436–2448, doi:10.1111/j.1365-2486.2009.02104.x, 2010.
- 35 Maljanen, M., Sigurdsson, B. D., Guðmundsson, J., Óskarsson, H., Huttunen, J. T., and Martikainen, P. J.: Greenhouse gas balances of managed peatlands in the Nordic countries – present knowledge and gaps, *Biogeosciences*, 7, 2711–2738, doi:10.5194/bg-7-2711-2010, 2010.
- 40 Metzger, C., Jansson, P.-E., Lohila, A., Aurela, M., Eickenscheidt, T., Belelli-Marchesini, L., Dinsmore, K. J., Drewer, J., van Huissteden, J., and Dröslér, M.: CO<sub>2</sub> fluxes and ecosystem dynamics at five European treeless peatlands – merging data and process oriented modeling, *Biogeosciences*, 12, 125–146, doi:10.5194/bg-12-125-2015, 2015.
- Monteith, J. L.: Solar radiation and productivity in tropical ecosystems, *Journal of Applied Ecology*, 9, 747–766, 1972.
- 45 Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *American Society of Agricultural and Biological Engineers*, 50, 885–900, 2007.
- Morris, P. J., Baird, A. J., and Belyea, L. R.: The DigiBog peatland development model 2: Ecohydrological simulations in 2D, *Ecohydrol.*, 5, 256–268, doi:10.1002/eco.229, 2012.
- 50 Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resources Research*, 12, 513–522, doi:10.1029/WR012i003p00513, 1976.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, *Journal of Hydrology*, 10, 282–290, 1970.
- 55 Nilsson, M., Sagerfors, J., Buffam, I., Laudon, H., Eriksson, T., Grelle, A., Klemedtsson, L., Weslien, P. E. R., and Lindroth, A.: Contemporary carbon accumulation in a boreal oligotrophic minerogenic mire – a significant sink after accounting for all C-fluxes, *Glob Change Biol*, 14, 2317–2332, doi:10.1111/j.1365-2486.2008.01654.x, 2008.
- Päivänen, J.: Hydraulic conductivity and water retention in peat soils, *Acta Forestalia Fennica*, 129, 1973.
- Pappas, C., Fatichi, S., Leuzinger, S., Wolf, A., and Burlando, P.: Sensitivity analysis of a process-based ecosystem model: Pinpointing parameterization and structural issues, *J. Geophys. Res. Biogeosci.*, 118, 505–528, doi:10.1002/jgrg.20035, 2013.
- 60 Parish, F. (Ed.): Assessment on peatlands, biodiversity and climate change, 203, Wetlands International, Wageningen, 179 S, 2008.

- Peichl, M., Sagerfors, J., Lindroth, A., Buffam, I., Grelle, A., Klemedtsson, L., Laudon, H., and Nilsson, M. B.: Energy exchange and water budget partitioning in a boreal minerogenic mire, *J. Geophys. Res. Biogeosci.*, 118, 1–13, doi:10.1029/2012JG002073, 2013.
- 5 Peichl, M., Öquist, M., Löfvenius, M. O., Ilstedt, U., Sagerfors, J., Grelle, A., Lindroth, A., and Nilsson, M. B.: A 12-year record reveals pre-growing season temperature and water table level threshold effects on the net carbon dioxide exchange in a boreal fen, *Environmental Research Letters*, 9, 55006, 2014.
- Peichl, M., Sonnentag, O., and Nilsson, M.: Bringing Color into the Picture: Using Digital Repeat Photography to Investigate Phenology Controls of the Carbon Dioxide Exchange in a Boreal Mire, *Ecosystems*, 18, 115–131, doi:10.1007/s10021-014-9815-z, 2015.
- 10 Petrescu, A. M. R., Lohila, A., Tuovinen, J.-P., Baldocchi, D. D., Desai, A. R., Roulet, N. T., Vesala, T., Dolman, A. J., Oechel, W. C., Marcolla, B., Friborg, T., Rinne, J., Matthes, J. H., Merbold, L., Meijide, A., Kiely, G., Sottocornola, M., Sachs, T., Zona, D., Varlagin, A., Lai, D. Y. F., Veenendaal, E., Parmentier, F.-J. W., Skiba, U., Lund, M., Hensen, A., van Huissteden, J., Flanagan, L. B., Shurpali, N. J., Grünwald, T., Humphreys, E. R., Jackowicz-Korczyński, M., Aurela, M. A., Laurila, T., Grüning, C., Corradi, C. A. R., Schrier-Uijl, A. P., Christensen, T. R., Tamstorf, M. P., Mastepanov, M., Martikainen, P. J., Verma, S. B., Bernhofer, C., and Cescatti, A.: The uncertain climate footprint of wetlands under human pressure, *Proceedings of the National Academy of Sciences of the United States of America*, 112, 4594–4599, doi:10.1073/pnas.1416267112, 2015.
- Prentice, I. C., Bondeau, A., Cramer, W., Harrison, S. P., Hickler, T., LUCHT, W., Sitch, S., Smith, B., and Sykes, M. T.: Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change, in: *Terrestrial ecosystems in a changing world*, Canadell, J. G., Pataki, D. E., Pitelka, L. (Eds.), Global change--the IGBP series, Springer, Berlin, New York, 175–192, 2007.
- 20 Price, J.: Evaporation from a blanket bog in a foggy coastal environment, *Boundary-Layer Meteorol*, 57, 391–406, doi:10.1007/BF00120056, 1991.
- Prihodko, L., Denning, A. S., Hanan, N. P., Baker, I., and Davis, K.: Sensitivity, uncertainty and time dependence of parameters in a complex land surface model, *Agricultural and Forest Meteorology*, 148, 268–287, doi:10.1016/j.agrformet.2007.08.006, 2008.
- 25 Quick, W. P., Chaves, M. M., Wendler, R., David, M., Rodrigues, M. L., Passaharinho, J. A., Pereira, J. S., Adcock, M. D., Leegood, R. C., and Stitt, M.: The effect of water stress on photosynthetic carbon metabolism in four species grown under field conditions, *Plant Cell Environ*, 15, 25–35, doi:10.1111/j.1365-3040.1992.tb01455.x, 1992.
- 30 Quillet, A., Frohking, S., Garneau, M., Talbot, J., and Peng, C.: Assessing the role of parameter interactions in the sensitivity analysis of a model of peatland dynamics, *Ecological Modelling*, 248, 30–40, doi:10.1016/j.ecolmodel.2012.08.023, 2013.
- Ratkowsky, D. A., Olley, J., McMeekin, T. A., and Ball, A.: Relationship between temperature and growth rate of bacterial cultures, *J. Bacteriol*, 149, 1–5, 1982.
- 35 Reich, P. B., Walters, M. B., and Ellsworth, D. S.: Leaf Life-Span in Relation to Leaf, Plant, and Stand Characteristics among Diverse Ecosystems, *Ecological Monographs*, 62, 365, doi:10.2307/2937116, 1992.
- Reichstein, M., Papale, D., Valentini, R., Aubinet, M., Bernhofer, C., Knohl, A., Laurila, T., Lindroth, A., Moors, E., Pilegaard, K., and Seufert, G.: Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites, *Geophys. Res. Lett.*, 34, doi:10.1029/2006GL027880, 2007.
- 40 Rice, S. K., Aclander, L., and Hanson, D. T.: Do bryophyte shoot systems function like vascular plant leaves or canopies?: Functional trait relationships in Sphagnum mosses (Sphagnaceae), *American journal of botany*, 95, 1366–1374, doi:10.3732/ajb.0800019, 2008.
- Richards, L. A.: Capillary conduction of liquids through porous mediums, *Journal of Applied Physics*, 1, 318–333, doi:10.1063/1.1745010, 1931.
- 45 Robroek, Bjorn J. M., Schouten, Matthijs G. C., Limpens, J., Berendse, F., and Poorter, H.: Interactive effects of water table and precipitation on net CO<sub>2</sub> assimilation of three co-occurring Sphagnum mosses differing in distribution above the water table, *Glob Change Biol*, 15, 680–691, doi:10.1111/j.1365-2486.2008.01724.x, 2009.
- Roulet, N., Hardill, S., and Comer, N.: Continuous measurement of the depth of water table (inundation) in wetlands with fluctuating surfaces, *Hydrol. Process.*, 5, 399–403, doi:10.1002/hyp.3360050407, 1991.
- 50 Ryser, P.: The Importance of Tissue Density for Growth and Life Span of Leaves and Roots: A Comparison of Five Ecologically Contrasting Grasses, *Functional Ecology*, 10, 717, doi:10.2307/2390506, 1996.
- Safta, C., Ricciuto, D., Sargsyan, K., Debusschere, B., Najm, H. N., Williams, M., and Thornton, P.: Global sensitivity analysis, probabilistic calibration, and predictive assessment for the Data Assimilation Linked Ecosystem Carbon model, *Geosci. Model Dev. Discuss.*, 7, 6893–6948, doi:10.5194/gmdd-7-6893-2014, 2014.
- 55 Sagerfors, J., Lindroth, A., Grelle, A., Klemedtsson, L., Weslien, P., and Nilsson, M.: Annual CO<sub>2</sub> exchange between a nutrient-poor, minerotrophic, boreal mire and the atmosphere, *J. Geophys. Res. Biogeosci.*, 113, G01001, doi:10.1029/2006JG000306, 2008.
- Saltelli, A., Tarantola, S., and Campolongo, F.: Sensitivity Analysis as an Ingredient of Modeling, *Statistical Science*, 15, 377–395, 2000.
- 60 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.
- Sándor, R., Barcza, Z., Hidy, D., Lellei-Kovács, E., Ma, S., and Bellocchi, G.: Modelling of grassland fluxes in Europe: Evaluation of two biogeochemical models, *Agriculture, Ecosystems & Environment*, 215, 1–19, doi:10.1016/j.agee.2015.09.001, 2016.

- Santaren, D., Peylin, P., Bacour, C., Ciais, P., and Longdoz, B.: Ecosystem model optimization using in situ flux observations: Benefit of Monte Carlo versus variational schemes and analyses of the year-to-year model performances, *Biogeosciences*, 11, 7137–7158, doi:10.5194/bg-11-7137-2014, 2014.
- Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, *Global Biogeochem. Cycles*, 21, n/a, doi:10.1029/2006GB002834, 2007.
- Schulz, K. and Beven, K.: Data-supported robust parameterisations in land surface-atmosphere flux predictions: towards a top-down approach, *Hydrol. Process.*, 17, 2259–2277, doi:10.1002/hyp.1331, 2003.
- Schulze, E.-D.: Biological control of the terrestrial carbon sink, *Biogeosciences*, 3, 147–166, doi:10.5194/bg-3-147-2006, 2006.
- Shaver, G. R. and Laundre, J.: Exsertion, elongation, and senescence of leaves of *Eriophorum vaginatum* and *Carex bigelowii* in Northern Alaska, *Glob Change Biol*, 3, 146–157, doi:10.1111/j.1365-2486.1997.gcb141.x, 1997.
- Silvola, J., Alm, J., Ahlholm, U., Nykanen, H., and Martikainen, P. J.: CO<sub>2</sub> Fluxes from Peat in Boreal Mires under Varying Temperature and Moisture Conditions, *Journal of Ecology*, 84, 219–228, 1996.
- Staudt, K., Falge, E., Pyles, R. D., Paw, U. K. T., and Foken, T.: Sensitivity and predictive uncertainty of the ACASA model at a spruce forest site, *Biogeosciences*, 7, 3685–3705, doi:10.5194/bg-7-3685-2010, 2010.
- Steed, J. E., DeWald, L. E., and Kolb, T. E.: Physiological and Growth Responses of Riparian Sedge Transplants to Groundwater Depth, *International Journal of Plant Sciences*, 163, 925–936, doi:10.1086/342634, 2002.
- St-Hilaire, F., Wu, J., Roulet, N. T., Frolking, S. E., Lafleur, P. M., Humphreys, E. R., and Arora, V.: McGill wetland model: evaluation of a peatland carbon simulator developed for global assessments, *Biogeosciences*, 7, 3517–3530, doi:10.5194/bg-7-3517-2010, 2010.
- Sulman, B., Desai, A., and Mladenoff, D.: Modeling Soil and Biomass Carbon Responses to Declining Water Table in a Wetland-Rich Landscape, *Ecosystems*, 16, 491–507, doi:10.1007/s10021-012-9624-1, 2013.
- Tang, J., Miller, P. A., Crill, P. M., Olin, S., and Pilešjö, P.: Investigating the influence of two different flow routing algorithms on soil-water-vegetation interactions using the dynamic ecosystem model LPJ-GUESS, *Ecohydrol.*, 8, 570–583, doi:10.1002/eco.1526, 2015.
- Tang, J., Zhuang, Q., Shannon, R. D., and White, J. R.: Quantifying wetland methane emissions with process-based models of different complexities, *Biogeosciences*, 7, 3817–3837, doi:10.5194/bg-7-3817-2010, 2010.
- Tatarinov, F. A. and Cienciala, E.: Application of BIOME-BGC model to managed forests, *Forest Ecology and Management*, 237, 267–279, doi:10.1016/j.foreco.2006.09.085, 2006.
- Tezara, W., Mitchell, V. J., Driscoll, S. D., and Lawlor, D. W.: Water stress inhibits plant photosynthesis by decreasing coupling factor and ATP, *Nature*, 401, 914–917, 1999.
- ~~Thomas, H. and Stoddart, J. L.: Leaf Senescence, *Ann. Rev. Plant Physiol.*, 31, 83–111, 1980.~~
- Thorsen, M., Refsgaard, J. C., Hansen, S., Pebesma, E., Jensen, J. B., and Kleeschulte, S.: Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale, *Journal of Hydrology*, 242, 210–227, doi:10.1016/S0022-1694(00)00396-6, 2001.
- Tian, S., Youssef, M. A., Amatya, D. M., and Vance, E. D.: Global sensitivity analysis of DRAINMOD-FOREST, an integrated forest ecosystem model, *Hydrol. Process.*, 28, 4389–4410, doi:10.1002/hyp.9948, 2014.
- van Huissteden, J., Petrescu, A. M. R., Hendriks, D., and Rebel, K. T.: Sensitivity analysis of a wetland methane emission model based on temperate and arctic wetland sites, *Biogeosciences*, 6, 3035–3051, 2009.
- van Oijen, M., Cameron, D. R., Butterbach-Bahl, K., Farahbakhshzad, N., Jansson, P.-E., Kiese, R., Rahn, K. H., Werner, C., and Yeluripati, J. B.: A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest, *Agricultural and Forest Meteorology*, 151, 1609–1621, doi:10.1016/j.agrformet.2011.06.017, 2011.
- Verbeeck, H., Samson, R., Verdonck, F., and Lemeur, R.: Parameter sensitivity and uncertainty of the forest carbon flux model FORUG: A Monte Carlo analysis, *Tree Physiology*, 26, 807–817, doi:10.1093/treephys/26.6.807, 2006.
- Waddington, J. M., Morris, P. J., Kettridge, N., Granath, G., Thompson, D. K., and Moore, P. A.: Hydrological feedbacks in northern peatlands, *Ecohydrol.*, 8, 113–127, doi:10.1002/eco.1493, 2015.
- Wang, G. and Chen, S.: A review on parameterization and uncertainty in modeling greenhouse gas emissions from soil, *Geoderma*, 170, 206–216, doi:10.1016/j.geoderma.2011.11.009, 2012.
- Wang, X., He, X., Williams, J. R., Izaurralde, R. C., and Atwood, J. D.: Sensitivity and uncertainty analyses of crop yields and soil organic carbon simulated with EPIC, *American Society of Agricultural and Biological Engineers*, 48, 1041–1054, 2005.
- Wang, Y.-P., Trudinger, C. M., and Enting, I. G.: A review of applications of model–data fusion to studies of terrestrial carbon fluxes at different scales, *Agricultural and Forest Meteorology*, 149, 1829–1842, doi:10.1016/j.agrformet.2009.07.009, 2009.
- Wania, R., Ross, I., and Prentice, I. C.: Integrating peatlands and permafrost into a dynamic global vegetation model: 1. Evaluation and sensitivity of physical land surface processes, *Global Biogeochem. Cycles*, 23, n/a, doi:10.1029/2008GB003412, 2009.
- Wania, R., Ross, I., and Prentice, I. C.: Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1, *Geosci. Model Dev.*, 3, 565–584, doi:10.5194/gmd-3-565-2010, 2010.
- Webster, K. L., McLaughlin, J. W., Kim, Y., Packalen, M. S., and Li, C. S.: Modelling carbon dynamics and response to environmental change along a boreal fen nutrient gradient, *Ecological Modelling*, 248, 148–164, doi:10.1016/j.ecolmodel.2012.10.004, 2013.

- White, L. M.: *Carbohydrate Reserves of Grasses: A Review*, *Journal of Range Management*, 26, 13–18, 1973.
- Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., and Wang, Y.-P.: Improving land surface models with FLUXNET data, *Biogeosciences*, 6, 1341–1359, doi:10.5194/bg-6-1341-2009, 2009.
- 5 | ~~Wingler, A.: The role of sugars in integrating environmental signals during the regulation of leaf senescence, *Journal of Experimental Botany*, 57, 391–399, doi:10.1093/jxb/eri279, 2005.~~
- Wu, J., Roulet, N. T., Sagerfors, J., and Nilsson, M. B.: Simulation of six years of carbon fluxes for a sedge-dominated oligotrophic minerogenic peatland in Northern Sweden using the McGill Wetland Model (MWM), *J. Geophys. Res. Biogeosci.*, 118, 795–807, doi:10.1002/jgrg.20045, 2013.
- 10 | Wu, Y. and Blodau, C.: PEATBOG: a biogeochemical model for analyzing coupled carbon and nitrogen dynamics in northern peatlands, *Geoscientific Model Development*, 6, 1173–1207, doi:10.5194/gmd-6-1173-2013, 2013.
- Xenakis, G., Ray, D., and Mencuccini, M.: Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model, *Ecological Modelling*, 219, 1–16, doi:10.1016/j.ecolmodel.2008.07.020, 2008.
- 15 | Yeloff, D. and Mauquoy, D.: The influence of vegetation composition on peat humification: Implications for palaeoclimatic studies, *Boreas*, 35, 662–673, doi:10.1080/03009480600690860, 2006.
- Yordanov, I., Velikova, V., and Tsonev, T.: Plant Responses to Drought, Acclimation, and Stress Tolerance, *Photosynthetica*, 38, 171–186, doi:10.1023/A:1007201411474, 2000.
- Yu, Z., Turetsky, M. R., Campbell, I. D., and Vitt, D. H.: Modelling long-term peatland dynamics. II. Processes and rates as inferred from litter and peat-core data, *Ecological Modelling*, 145, 159–173, doi:10.1016/s0304-3800(01)00387-8, 2001.
- 20 | Yurova, A., Wolf, A., Sagerfors, J., and Nilsson, M.: Variations in net ecosystem exchange of carbon dioxide in a boreal mire: Modeling mechanisms linked to water table position, *J. Geophys. Res. Biogeosci.*, 112, G02025, doi:10.1029/2006JG000342, 2007.
- Zaehle, S., Sitch, S., Smith, B., and Hatterman, F.: Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics, *Global Biogeochem. Cycles*, 19, doi:10.1029/2004GB002395, 2005.
- 25 | Zhang, Y., Li, C., Trettin, C. C., Li, H., and Sun, G.: An integrated model of soil, hydrology, and vegetation for carbon dynamics in wetland ecosystems, *Global Biogeochem. Cycles*, 16, 9-1–9-17, doi:10.1029/2001gb001838, 2002.
- Zhao, G., Bryan, B. A., and Song, X.: Sensitivity and uncertainty analysis of the APSIM-wheat model: Interactions between cultivar, environmental, and management parameters, *Ecological Modelling*, 279, 1–11, doi:10.1016/j.ecolmodel.2014.02.003, 2014.
- 30 | Zhao, W., Hidenori, T., and Zhao, H.: Estimation of vegetative surface albedo in the Kushiro Mire with Landsat TM data, *Chin. Geograph.Sc.*, 7, 278-288, doi:10.1007/s11769-997-0055-4, 1997.
- Zhu, Q., Liu, J., Peng, C., Chen, H., Fang, X., Jiang, H., Yang, G., Zhu, D., Wang, W., and Zhou, X.: Modelling methane emissions from natural wetlands by development and application of the TRIPLEX-GHG model, *Geoscientific Model Development*, 7, 981–999, 2014.
- 35

## Tables

**Table 1. Measurement data used as model input**

Variable	Period	Resolution as used for model input <sup>a</sup>	Method <sup>b</sup>	Measurement height
Global radiation	1991–2013	Hourly; 1991-2000: hourly values calculated from daily values by assuming a sinusoidal distribution between 07:30 and 19:30.	2001-2013: Li200sz sensor (LI-COR, Lincoln, NE, USA)	3 m
Air temperature	1991–2013	Hourly	MP100 temperature and moisture sensor (Rotronic AG, Bassersdorf, Switzerland) equipped with a ventilated radiation shield	3 m
Relative humidity		Hourly; 1991-2000: hourly values calculated from daily values by assuming equally distribution during each day	MP100 temperature and moisture sensor (Rotronic AG, Bassersdorf, Switzerland) equipped with a ventilated radiation shield	3 m
Precipitation	1991–2013	Hourly; 1991-2000 and November to April: the total daily precipitation was assumed to fell at 12:00 each day	Rainfall tipping-bucket (ARG 100, Campbell Scientific, Logan, UT, USA).	1 m
Wind speed	1991–2013	Hourly; 1991-2000: hourly values calculated from daily values by assuming equally distribution during each day	2001-2013:3-d wind anemometer (Gill Instruments Ltd., Hampshire, UK)	1.8 m
C content per soil layer	1994	One time in 1994	Every 4 cm between 0 and -32 cm, and every 12 cm between -60 and -338 cm	0 to -338 cm

<sup>a</sup>: Measurement resolution was the same or higher, except where mentioned differently.

- 5 <sup>b</sup>: The method description of meteorological input data applies to the climate station at the site. For gap-filling and for the pre-evaluation period, the data was obtained from the nearby standard climate station (Svartberget field station).

**Table 2: Measurement data used for model calibration**

<b>Variable</b>	<b>Period</b>	<b>Resolution as used for calibration</b>	<b>Method</b>	<b>Measurement height</b>
NEE	2001–2012	hourly	EC system, consisting of a three-dimensional sonic anemometer (1012R3 Solent, Gill Instruments, UK; heated during winter months) and a closed path infrared gas analyzer (IRGA 6262, LI-COR, Lincoln, Nebraska USA). Fluxes were calculated by the EcoFlux software (In Situ Flux AB, Ockelbo, Sweden) according to the EUROFLUX methodology (Aubinet et al., 1999, Sagerfors et al., 2008, Nilsson et al., 2008)	1.8 m
LE & H	2001–2009	hourly	Same EC system as above (Peichl et al., 2014)	1.8 m
Water table	2001–2009	daily	Float and counterweight system attached to a potentiometer (Roulet et al., 1991)	
Soil temperature	2001–2012	hourly	TO3R thermistors mounted in sealed, waterproof, stainless steel tubes (TOJO Skogsteknik, Djäkneboda, Sweden) in a lawn community 100 m northeast of the flux tower	–2 cm, –42 cm
Net radiation	2001–2011		NR-Lite sensor (Kipp&Zonen, Delft, the Netherlands)	4 m
Snow depth	2001–2012	daily	Sr-50 ultrasonic sensor (Campbell Scientific, Logan, UT, USA) nearby the flux-tower	
LAI of vascular plants	May–Sept. 2012	biweekly	Destructive sampling (Peichl et al., 2015)	

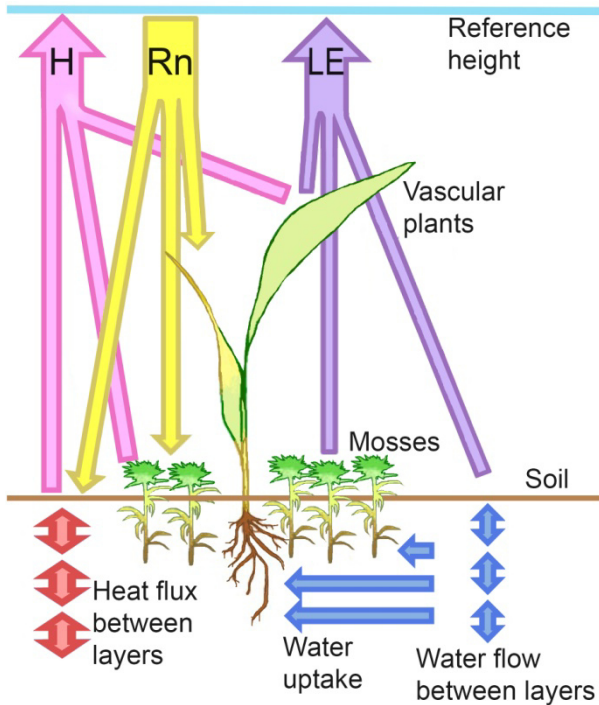


**Table 3: Different criteria sets for the selections of accepted runs**

Main component	Variable	R <sup>2</sup>	Mean error (ME)
Basic selection (these criteria are applied additionally in all following criteria sets)	WT < -0.2 m	≥0.40	±0.02 m
	LAI vascular plants	≥0.40	±0.02 m <sup>2</sup> m <sup>-2</sup>
	Daytime NEE		±2 gCO <sub>2</sub> -C m <sup>-2</sup> d <sup>-1</sup>
NEE	Accumulated NEE	≥0.98	
	Daytime NEE		±0.02 gCO <sub>2</sub> -C m <sup>-2</sup> d <sup>-1</sup>
	Night time NEE		±0.07 gCO <sub>2</sub> -C m <sup>-2</sup> d <sup>-1</sup>
Sensible heat	H		±3·10 <sup>5</sup> J m <sup>-2</sup> d <sup>-1</sup>
	Accumulated H	≥0.97	
Latent heat	LE		±1·10 <sup>5</sup> J m <sup>-2</sup> d <sup>-1</sup>
	Accumulated LE	≥0.98	
Net radiation	Net radiation	≥0.82	±4·10 <sup>4</sup> J m <sup>-2</sup> d <sup>-1</sup>
Soil temperature	Temperature -2 cm	≥0.95	±0.22 °C
	Temperature -42 cm		±0.22 °C
Snow	Snow depth	≥0.76	
Water table	WT < -0.15 m	≥0.51	

5

**Figures**



**Figure 1. Energy flux partitioning and related soil water flows in the CoupModel as applied to a peatland using two plant canopies and root systems. Rn: Incoming radiation, LE: latent heat fluxes (sum of actual transpiration, interception evaporation and soil evaporation), H: sensible heat fluxes**

10

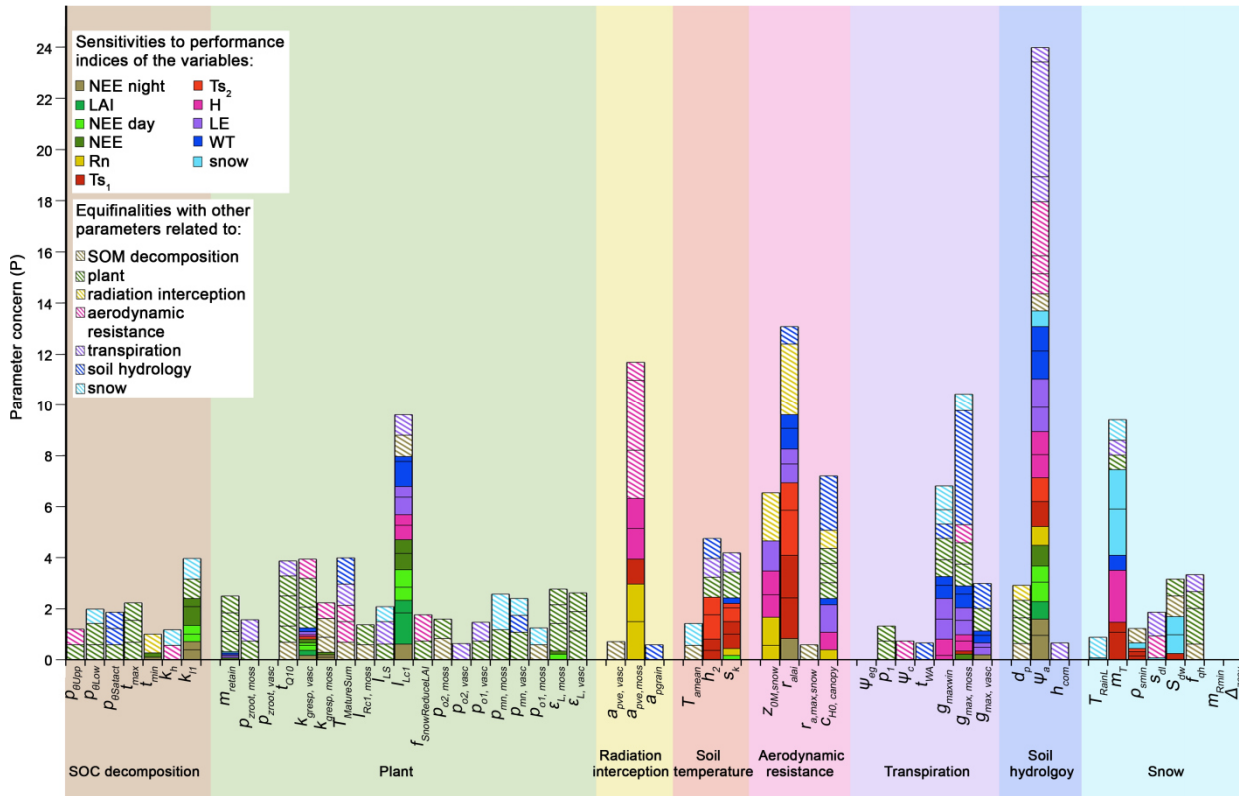


Figure 2. Parameter concern is shown on the y axis as sum of equifinalities (hatched) and sensitivities that could not be constrained unambiguously (solid). The x-axis shows the parameters, which belong to the **model process category** of the background colour.

5

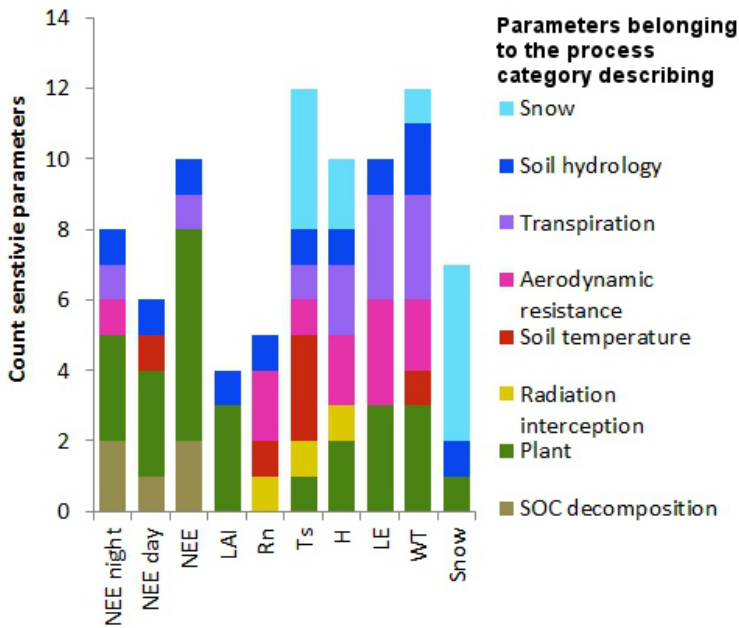


Figure 3. Connections between processes and parameters of different **model process categories**. The y-axis shows the count of parameters from the different **model process categories** (colours) that are sensitive to model performance in the various variables (x-axis).

10

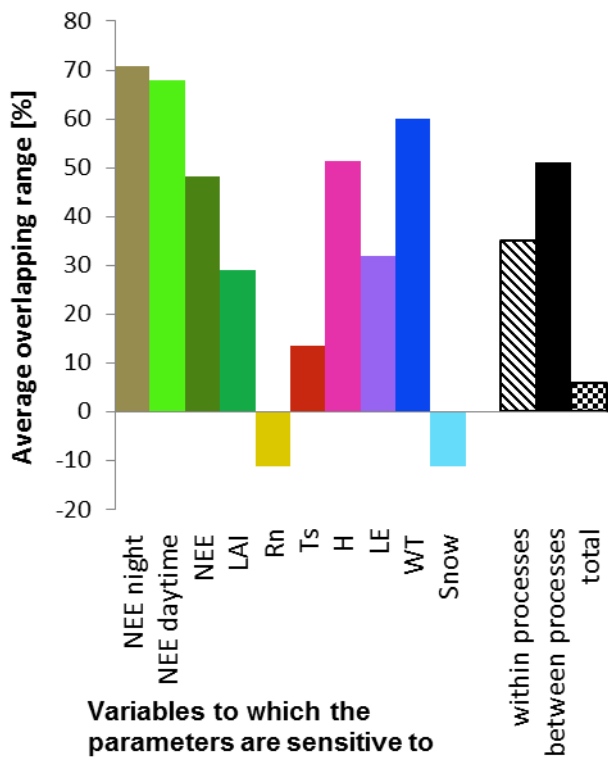


Figure 4. Average overlap of accepted ranges per parameter within each process and between processes, i.e. how unambiguously the parameters could be constrained. Negative values indicate the distance between accepted ranges when ranges did not overlap at all.

5

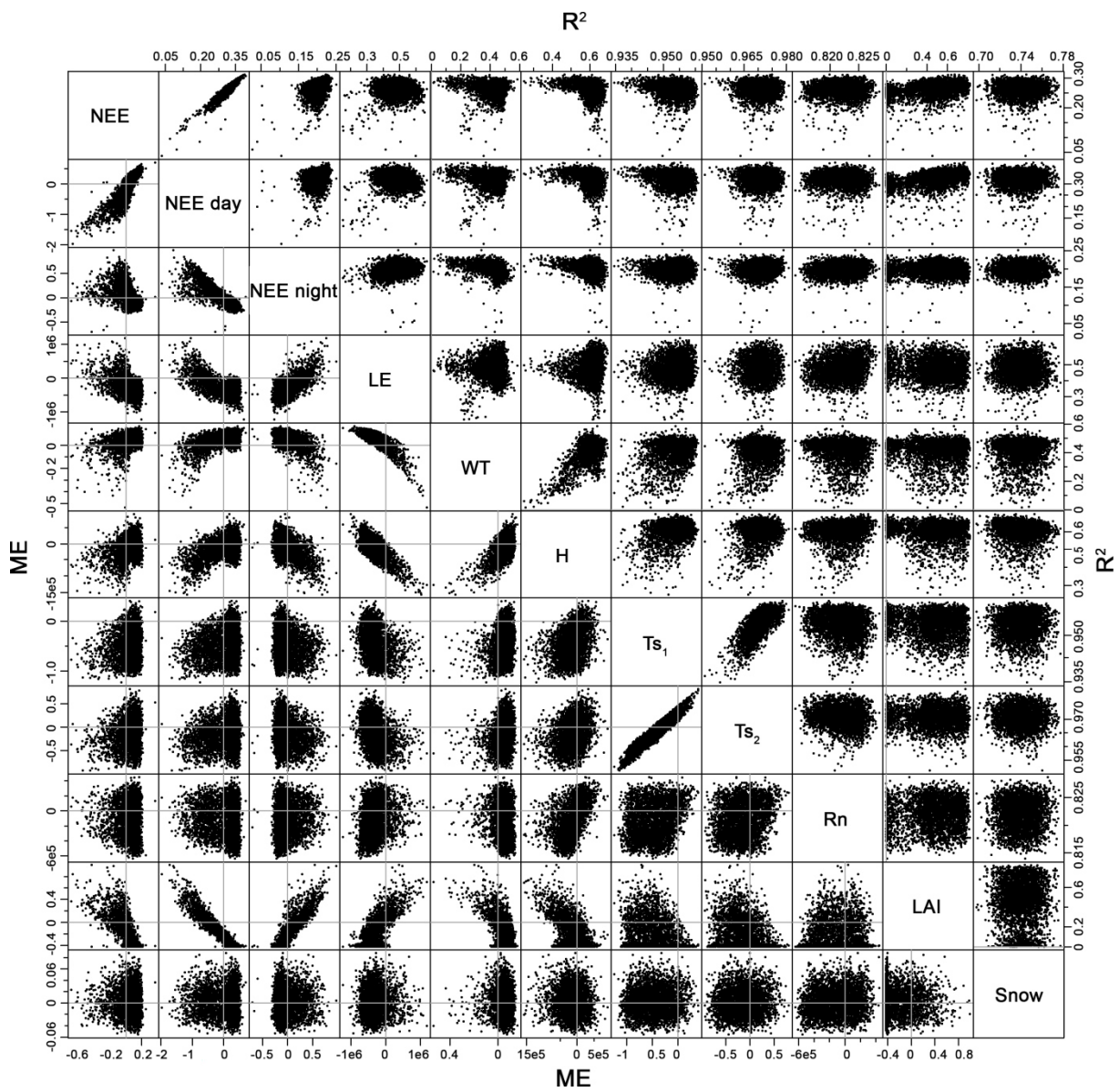


Figure 5. Correlations between performance indices in the prior distribution (3200 random runs):  $R^2$  versus  $R^2$  (upper panel); mean error (ME) versus ME (lower panel). Each of the dots represents a parameter set. Grey lines indicate the axes through zero.

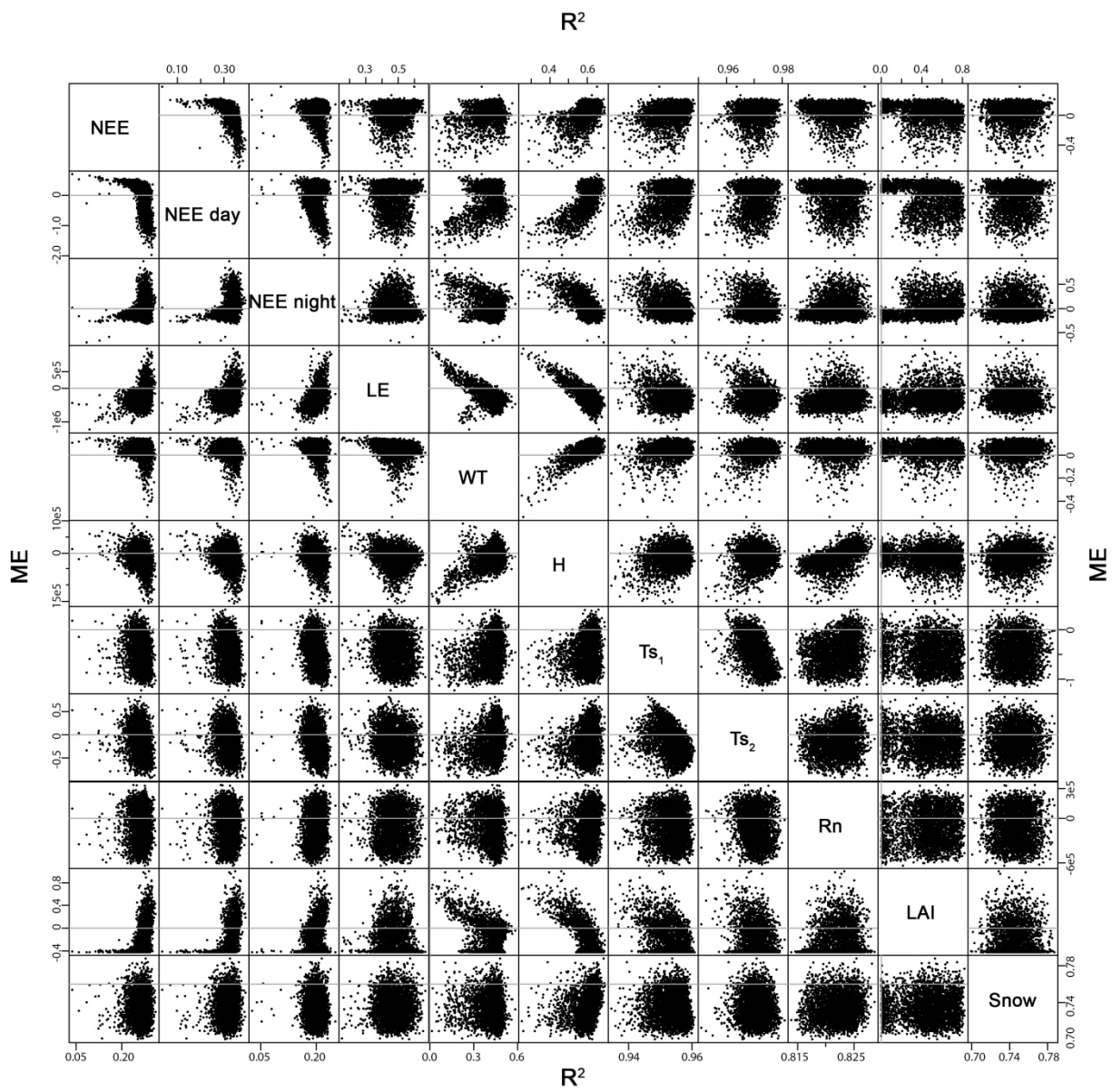


Figure 6. Correlations between performance indices in the prior distribution (3200 random runs):  $R^2$  (columns) versus mean error (ME) (rows). Each of the dots represents a parameter set. Grey lines indicate the axes through zero.

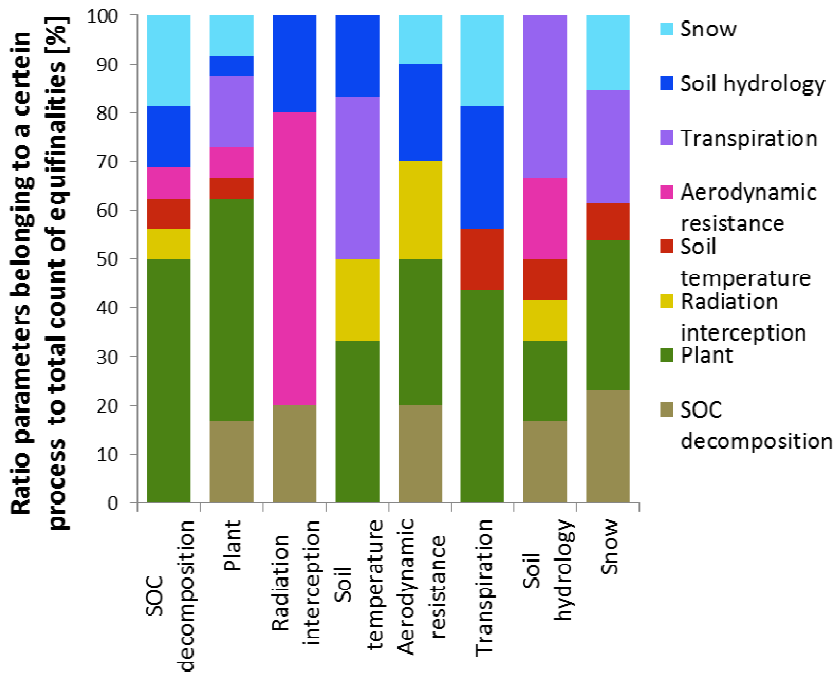


Figure 7. Module Process category belongings of parameters that correlated with parameters of a certain module process category.