This document contains the reviewer comments (and our responses) to the second iteration of the manuscript: https://www.geosci-model-dev-discuss.net/gmd-2018-313/

- attached is also a new supplement S3 (see response to general comment 2)
- and the latexdiff file containing the modifications to the previous version of the manuscript (the images have been cut from the file to keep the file size moderate).

The comments are gathered here as indented text and we have added numbering to them, so they can be individually referred to, if need be. Our responses to these comments are shown as unindented text (such as this) after each comment. We have made few additional grammar corrections, and also modified Fig. 4 by changing the axis units (so both are in mass units) and added gridlines to the image.

Comments by Reviewer 1

> I thank the authors for correcting the previous manuscript according to my comments. The procedure of the experiment becomes much clearer, and generally, I agree with the interpretation of the results. However, now I found some contradictions in the results and discussion. Authors need to carefully explain the results for ET and the estimated parameters related to the water stress. As the other reviewer suggested, the conclusions should be supported by the numerical results.
>
> I also suggest some improvements for the manuscript so that the readers can easily understand (notified with “Draft amendment”). They are just my recommendations for the improvements of the manuscript, and there may be misunderstandings. Therefore, the authors needed to reconsider the way of writings. I read these sentences, again and again, therefore, the difficulties should be the same for other readers.

General comments:

The authors need to describe the below results carefully.

1. Improvement of springtime increase

   GPP is greatly improved indeed, but improvement for ET is not so clear.

On this we agree but it should also be noted that there is not such an obvious discrepancy between the modelled and observed springtime ET as there is for GPP. We can see improvements e.g in FI-Hyy and FI-Sod, where the default model early bias has been removed.

2. b and r^2
   - Fig. 2 and S2 show that both b and r^2 are improved for GPP. However, b for ET is not so much improved or sometimes gets worse (especially the biases at FI-Ken and FL-Sod are apparent). I think it is better to show the results which get worse compared to the default using italic letters in Table 6. Then the bias increment for ET at FI-Ken, FL-Sod, and both for ET and GPP at RU-Zot and US-Prr will also be clear.

The focus of Table 6 is to identify the “best” conductance formulation from the results point of view whereas the suggested metric would show differences between default and optimised model versions and thus evaluate the performance of the calibration. We believe that combining the suggested metric to this table would not be beneficial as it would blur the focus that is to compare the conductance model behaviour. However, we have produced a bit more detailed analysis along the suggested lines and added it as a supporting supplement S3. The supplement contains details on both model specific and site specific improvements and deterioration's in the calibration process, separated for GPP and ET and the metrics (b and r^2). The bias increments and deteriorations in the model are clearly apparent in the supplement.

We are treating these results only as supporting material and not including them in the main text because the interpretation of the analysis is not straightforward. In the supplement we are not comparing like-to-like – the JSBACH default parameters reflect the Baseline and Bethy behaviour (as they are originally part of the model) but the BB (and variants) are newly introduced to the JSBACH. The default parameters for the BB models can be seen as arbitrary and we have not done/presented any initial calibration/validation. Therefore, we like to keep the focus of the manuscript in the comparison of the "end results".

- r^2 for ET is improved compared to the default, but much lower than r^2 for GPP. These results are also needed to be explained in the result section. The bold letter may be better for, e.g. from 0.9 to 1.1 so that the readers easily recognize the higher values.

The main reasons for this is that in JSBACH, e.g. the conductance is resolved for carbon assimilation, and the same conductance is then used for transpiration. In the model, GPP always takes a priority and is the determining factor (we could almost say that ET/transpiration is an afterthought). Additionally, GPP is derived from EC measurements by flux partitioning – this tends remove some of the flux instabilities (that are still present in ET). This also plays a role as the more stable GPP flux is easier to model than the more chaotic ET. This explanation was added to discussion in section "Validity of the simulations".

We appreciate the suggestion about the metric and explored the idea but ultimately feel that our original "metric" is more suited for the situation. The purpose of the bolded values is not to draw attention to where the model reproduces the observations well. The focus is on the comparison of the conductance formulations and systematic differences in the model performance – so we are also interested in the best performance on poorly replicated sites. The suggested metric would only focus on the "good" sites.

In addition to the suggestion, we also tested highlighting the best value (for each site and b, r^2 separately) and all values that are within 5% of it (similarly to S3). This approach yielded similar, but slightly obscured results to the original metric (so the interpretation was not as straightforward).

- For model validation, b and r^2 may better to be evaluated separately to prevent the overestimate, because sometimes r^2 is high but b gets worth. From this aspect, it may better to cumulate the number of the "good" result e.g. from 0.9 to 1.1 for b and 0.9 to 1.0 for r^2, respectively. Then b and r^2 is better to be explained and discussed for ET and GPP respectively.

We do not follow, what overestimate the reviewer means. The question about using a different metric to examine the model performance was addressed above. We have separated the ET and the GPP discussion along these lines, in the places where the reviewer has suggested it (see the specific comments).

The "good" results yielded with the suggested metric (with the suggested limits) are presented below and they do not provide clear added value to the results of Table 6:

	Base	Bethy	BB	Leu	F&K	USO
ET: b	5	5	5	4	3	1
ET: r^2	0	0	0	0	0	0
GPP: b	3	4	5	4	4	4
GPP: r^2	3	3	2	4	5	5

3. q, θpwp, θtsp

   The effects of these parameters are complicated. Therefore, a clear explanation is needed so that the readers can understand it easily. The optimised θpwp and θtsp for the general

condition are smaller than the default; that makes β larger (which means, β became ineffective). The oprimised θ tsp for the dry condition is also smaller than the default but larger than that of general condition and θpwp varies between the sites; that makes β different between the sites. Larger q makes β smaller, whereas smaller q makes β larger (because β is between 0 and 1). In my understanding, q changes An, which control GPP but not affect ET. Is that correct?

We have added description about the combined effect of these parameters to the end of the section 2.3 “Modifications to the JSBACH model”. The changes to parameter values do not necessarily make β ineffective, they can also be viewed to change the response intensity to drought (instead of slow gradual restriction we get a more abrupt and strong response). The parameter θpwp does not vary between the sites in our simulations – it varies between the conductance formulations (i.e. Fig. 3 rightmost panels) as the reviewer probably meant.

The β function is first used to calculate the stomatal conductance for carbon assimilation. The same conductance is later used for transpiration calculations (A8). However, the factor β^q only affects the net assimilation, not ET. So q affects only GPP.

4. I understand that the authors used the mean of the site level cost function values for APIS. Then, the parameters of the 50 draw for each IS sampler are the same for all study site? Does Figure 1 show the “global” estimate by using the “mean” of the site level cost function? I think it is better to add these descriptions to P9 Lines 25-27 and Fig. 1 so that the readers can easily understand the procedure.

Essentially, each site generates the ET and GPP fluxes with the same parameter values. Then the actual cost function (which is returned to APIS/optimiser) is calculated as the mean of the site level “cost functions”. So when we refer to the cost function, it is always this mean (except for the dry period). We modified the cost function definition in 2.9 (which we believe is the correct location) to better reflect this.

All of the estimates in Fig. 1 are calculated with the “average cost function”. The global estimate (yellow) is produced with the deterministic mixture approach (item 4 in the MCMC-APIS comparison list) whereas the red is just the mean of the IS sampler locations. We added the mention of the deterministic mixture to the image caption.

Specific Comments:

1. P1. The title is needed to be reconsidered. For example, "Parameter calibration for stomatal conductance and photosynthesis.."

We calibrate also parameters that are not directly linked to stomatal conductance or photosynthesis, so the suggested revision is imprecise. AS a compromise, we could change the title to: “Parameter calibration and stomatal conductance formulation comparison for boreal forests with adaptive population importance sampler in the land surface model JSBACH”

2. P1. Lines 6-8: Draft amendment, also please reconsider the description with regarding general comments 1. “This modification enabled the model to correctly reproduce the springtime increase in GPP for conifers throughout the measurements sites used in this study. However, the improvement for ET was limited. The key parameters identified along with this modification were the parameters which control the soil moisture stress function and the overall rate of carbon fixation.”
3. P1. Lines 8-10: Please reconsider the description, "Overall, … models", concerning general comments 2.

We have amended the manuscript to better reflect the two suggestions above. However, the delayed effect of temperature is mostly correcting the erroneous behaviour of the springtime GPP – the springtime ET is not the “original” problem, so it should not be the focus here. We now state the “improvements” for ET and GPP separately in the manuscript.

4. P5. Line21: What is “new”? Is that mean new to the JSBACH model? The explanation for q (one of the key parameters) is also not found in the main text.

Yes, the text has been amended to reflect this better. We have also added the explanations of q to the end of the section 2.3.

5. P10. Lines 4-11: I do not understand this paragraph. Please reconsider the descriptions. In my understanding, each IS sampler use just one parameter set for spin-up. Then the remaining 49 members use the same spin-up but the parameters are perturbed around the first parameter set. If so, such description may help the readers to understand the spin-up process easily. The initial parameter combination for each IS sampler is also selected randomly from the ranges in Table 2. This information is also needed. I also do not understand the procedure, “We also slightly scale (reduce) the importance weights based on the distance of ..”. Is it the same as procedure 3 or different procedure just for spin-up?

We rearranged the sentenced in the paragraph and added the proposed description. The initial parameter combination is now more explicitly stated in the MCMC-APIS comparison list. The scaling procedure is the same as depicted in the list before.

6. P11. Lines 6-8: I still do not understand “Since we also did not run the model spin-up for .. parameter values”. The authors described that the post procedure for APIS is used for the “correct spin-up” at P10. Lines 4-11. Is not enough? If so, why?

We do believe that the “spin-up correction” should be sufficient but it is also a good practice to verify the results of the more complicated sampling algorithms with simpler optimisers (as we have done here). The optimiser is also the only algorithm used for the drought period (we noticed this detail was missing from the manuscript and added it to this section).

7. P11. Line 8: Does “the same datasets as APIS” mean, same calibration period (the first five years), same climate forcing data, and the same observation data? Is the initial state also generated by APIS, or same as spin-up for APIS? Please make these settings clear.

The dataset is exactly the same as for APIS. The initial state is the mean of the APIS final location parameters. The spin-up is run separately for all the samples the optimiser draws. These details have been added to the manuscript.

8. P12. Line 5: What are the "acceptable values"? Is that mean the final state of the parameter distribution range of the parameter optimisation? Please make it clear.

No, this refers to the absolute range of the parameters. We added a reference to the Table 2, which should make this clear.

9. P14. Lines 6-8: The former sentence indicates that bud burst is not as critical, whereas the latter sentence indicates that the acclimation parameter dominates the phenology parameters. At first, I thought these sentences contradict, but later I understand the difference of these procedures. I think it is better to describe the different functions of these parameters clearly so that the readers easily understand.

We added a description for the function of tau and the LoGro parameters.

10. P15. Lines 10-11: "The annual cycles of the Bethy model are more in line with the Ball-Berry variants than those of the Baseline model (see supplements S2 for the yearly cycles of the

other models)." I think this is not always true. This result is not used in discussion, so I think it is better to remove this sentence. The authors can rewrite that "The results of other stomatal conductance models are shown in S2".

We modified the sentences according to the suggestions. We also added here a mention of the supporting material in supplement S3.

11. P15. Lines 15-18: Please reconsider this paragraph according to my general comment 1 and 2.

We modified this paragraph and stated separately the ET and GPP behaviour.

12. P16. Table 6: “best values” -> “N over threshold” (please refer general comment 2).

Not changed as we did not change the metric (see answer to general comment 2).

13. P17. Line 10-16: The authors do not show the results. Therefore, I do not understand the experimental setting and what is validated.

We produce similar results as in Fig. 2 and supplement S2 but using the dry period optimised parameters. We compare these to Fig. 2 and S2. The paragraph in question merely states that the dry period optimisation does not generally produce better results. We amended the paragraph so the setup is easier to understand.

14. P18. Lines 24-26: “This is mostly a direct result of the normalisation of the cost function that inflates the target distribution and gives too much weight to the initial locations and draws.” I do not understand this sentence. I think the convergence is rather related to the parameter sensitivity to the observations.

Yes, the reviewer is absolutely correct. The cost function depicts the parameter sensitivity to observations. So when we normalise the cost function, we inflate the target distribution – so the parameter (relative) sensitivity to the observations is reduced.

15. P19. Lines 1-5: Draft amendment, also please refer general comment 1.

    “This delay is also reflected in transpiration, and consequently in ET at FI-Hyy and FI-Sod to some extent. However, the effect at the other sites is not clear." (see general comment 1). The description for FI-Sod seems too much detailed, and discussion is rather needed to be done for general comment 2. Also, I wonder why ET is not improved greatly compared to GPP, although both the observations are used for the calibration. There may be some possible reasons for the mismatch. 1) interaction in the optimisaion process. For me, it seems that GPP is optimised by soil water parameters, and that affect ET estimate. 2) Parameter estimation bias (e.g. θhum strongly decreased and get to its lower limit). 3) Bias correction by q. In my understanding, GPP bias can be corrected using q, but q does not affect the bias of ET. Considering these issues are important for the study with multi observations and to improve ET. I recommend the authors to run some additional experiment to clarify the issues mentioned above for the discussion (only one site is enough). If it is beyond the scope, please discuss the possible reasons for future study.

We modified the sentences following the suggestions. We clarified the FI-Sod description and feel that it is an important point to make – improving model behaviour can lead to an increase in the ET bias if the improvement negates a previously erronous behaviour. The differences in ET/GPP improvements have been addressed in our response to general comment 2.

16. P19. Lines 13-16: Please reconsider this paragraph according to my general comment 2.

We modified this paragraph slightly, but as we did not change the metric in Table 6, the changes are minor.

17. P20. Lines 1-3: I do not find "the result of the site level estimates of g1", so I do not understand this sentence. "not only" is needed at Line 1. I also do not understand what does "control" for Wang mean.

This is a general comment regarding e.g. literature values, not specific simulations/results in the manuscript – the sentence was modified to reflect this. The control refers to the setting in the Wang (1996) paper, this has been modified to "(in Table 1, Control)", which should clarify the reference.

18. P20. Lines 18-28: Please reconsider this paragraph with general comments 3.

We added the clarifications in the "Model modifications" section as explained in the answer to general comment 3. Therefore, we mainly added some clarifications to this paragraph.

19. P20, Line 20: Not only θtsp but also θpwp is lower.

Added.

20. P21. Lines 3-4: "The parameters affecting the optimisation process the most were consistent for all stomatal conductance formulations." I do not understand this sentence.

Yes this sentence was missing commas, but we modified the sentence to be more understandable: "The parameters that were most effective in the optimisation processes, were consistent for all stomatal conductance formulations."

21. P21. Lines 8-11: Please reconsider this paragraph with general comments 3.
    - How did the authors evaluate the "importance of q for the Ball-Berry type model"? The same validation of importance in Table 5 may need for Table 7.
    - "Overall, both optimisations strongly indicate that boreal forest transpiration is not limited by soil moisture stress under normal conditions." In the discussion, the authors indicated two reasons (the other is the water retention capabilities of the soil). Also, ET is sometimes underestimated. Therefore, the authors can just indicate the possibility.

The importance can be verified from the sampling states directly, but it can also be seen from Fig. 3 rightmost panels, where we have plotted the β-function values during the drought. Since we are focusing on the dry event, the relative amount of values affected by q (i.e. low soil moisture) is considerable higher than under general conditions (so q becomes more important). The "importance metric" was not used here because it was meant to reflect the APIS simulation identifiability (so it would be detrimental to add it to Table 7, where APIS was not used).

We added a mention about the nonlinearity of the additional reduction (β^q), which also indicates that soil moisture stress is not the limiting factor during normal conditions. We removed the word "strongly" but feel that this formulation should be sufficient as it is indicative. If ET is underestimated, then likely transpiration is also – so this is not a contradiction as such but points to other possible problems in the model.

22. P21. Line 13: Please reconsider this paragraph with general comments 2.

We examined the different metric proposed in GC2, but decided against this. Therefore, only minimal changes were made to this paragraph.

Technical corrections

1. P1. Lines 2-3: Draft amendment

   "The parameter posterior distributions were generated by the adaptive population importance sampler (APIS), then the optimal values were estimated by a simple stochastic optimisation algorithm"

Accepted.

2. P1. Lines 3-5: Draft amendment

   “Using the in-situ measurements of evapotranspiration (ET) and gross primary production (GPP), we calibrated three model parameter groups (**, **, and **), and identified the key parameters. “

This is slightly inaccurate, because we don’t calibrate the groups – APIS does not use the grouping and the optimiser also draws samples from the full parameter space. We modified the previous formulation to include the “in-situ observations”.

3. P1. Lines 11-13: Draft amendment

   “This optimisation improved the model behaviour, but the changes to the parameter values were significant except for the unified stomatal optimization model (USO). Interestingly, the USO model demonstrated the best performance during this event with only small changes to the parameter values."

Accepted with modifications: “This optimisation improved the model behaviour, but resulted in significant changes to the parameter values except for the unified stomatal optimisation model (USO). Interestingly, the USO model demonstrated the best performance during this event.”

4. P2. Lines 20-22: Draft amendment

   "It can be hypothesised that the choice of the stomatal conductance model affects the ecosystem model parameters broadly. Because the stomatal conductance formulations vary in their responses to the different conditions. However, a holistic assessment of the performance of the stomatal conductance models together with other parameters (e.g. photosynthesis parameters) has been missing."

Accepted with slight modifications.

5. P3. Lines 6-7: Draft amendment

   “The APIS algorithm samples the full parameter space (as do MCMC methods) and it can treat a mixture of parameter prior distributions. Therefore, APIS can estimate complicated multidimensional probability distributions.”

Accepted with slight modifications.

6. P3. Lines 11-14: Draft amendment

   “First, we utilise APIS to sample the full parameter space with the different stomatal conductance formulations and to locate different modes of the target parameter distributions (peaks of high probability). Second, using the distributions generated by APIS as the prior distributions, the parameters are optimized using a simple stochastic optimisation method. Finally, we assess the inter-site variability and the robustness of the calibrated parameters together with different stomatal conductance formulations. Optimised parameters for a specific drought is also investigated and compared with the parameters for the general optimisation.”

Accepted with slight modifications.

7. P3. Line 20: Draft amendment

   “The site level half-hourly measurements of eddy covariance (EC)”

Accepted with modified word order.

8. P3. Lines 21-24: Draft amendment

“The gap-filled and low quality (based on FLUXNET data quality flags) measurements were masked, and the daily aggregates (usually means) were accepted as part of the calibration process if at least 60% of values between 4:00 and 20:00 (i.e. daytime measurements) for that day were unmasked. The daily aggregated data (ET and GPP) were used for the calibration and the validation, whereas all of the half-hourly data were used as the climate forcing data (as explained in section 2.4).”

Accepted with slight modifications.

9. P5. Line 32: Draft amendment

   “However, coniferous evergreen trees do not shed all of their leaves for winter, and the original phenology model is not suitable for a boreal forest.”

This is not entirely correct. The purpose of the phenology model is to determine when new leaves start to grow and the consequent growth rate. The state of acclimation that in our simulations corrects the early spring GPP, restricts the net assimilation rate but does not (directly) influence the phenology model. We believe that the phenology model itself is performing adequately.

10. P6. Table 2: The additional parameters for Friend and Kiang model is also needed to be included.

Added.

11. P7. Line 1: There are many “b” in this manuscript: photosynthetic acclimation, additional parameter for Friend and Kiang model, and the slope of the regression line. The authors should change them so that the readers can recognize these parameters are different.

Modifed, the F&K exponent is now d and the curvature in photosynthetic acclimation is k.

12. P7. Line 17: Draft amendment

    “In the original JSBACH formulation (i.e. the Baseline version),”

Accepted.

13. P9. Lines 5-8: Draft amendment

    “Above i is the elements with each IS sampler (described later). Generally, Eq. (4) cannot be analytically solved, hence it is usually estimated numerically. Commonly this is achieved by one of the many Markov chain Monte Carlo (MCMC) methods, but in this study, we apply the adaptive population importance sampler (APIS) defined by Martino et al. (2015). APIS is a Monte Carlo (MC) method that utilises a population of importance samplers (IS) to jointly estimate the target pdf ($p(\theta|x)$) and the normalising constant ($Z(x)$) by a deterministic mixture approach (Veach and Guibas, 1995; Owen and Yi, 2000), whereas the MCMC methods do not care about the value of Z. Importance sampling density $q(\theta)$ is also introduced in APIS algorism.”

    Then P10. Lines 13-15 needed to be removed to here.

Accepted with slight modifications. The “i” refers to the i-th element in the (parameter) vector as is the standard notation (we added this information also). It is just calculated as the marginal integral over the whole parameter space.

14. P11. Line 6: Draft amendment

    “overshadows the calculations” -> I do not understand "overshadow". Appropriate word is needed.

This was reworded to “dominates the calculations”.

15. P13. Line 17: I could not find Ball-Berry results in S1.

The supplement included the image “APIS-S1-posteriors.png” that includes both Bethy and BB posteriors at 20 iterations. Apparently I forgot to add the chains for the BB results when I reproduced the other images. Will be added to the supplement.

16. P13. Table 4: Draft amendment

    “Parameter scale reduction ^R (at APIS iteration) and stability δ(threshold number of the iteration) estimates from the Bethy simulations.”

Accepted.

17. P14. Lines 2-3: Draft amendment

    “There is an overall agreement on the values of the most prevalent parameters (see the bold and the italic letters in Table 5 between the models.”

Accepted, with “letters” changed to characters.

18. P14. Table 5: The order of the parameters is different from Table 2. I think that the same order is better to be understood. “b” also should be reconsidered.

Yes, the g0, g1, a, d (previously b) were moved to the end of the Table. We will change this to the same order as before.

19. P15. Lines 26-27: Draft amendment

    “The values of the relative humidity parameter θhum, the residual stomatal conductance g0, and fC3 have remained nearly unchanged,”

The relative change in the values of f_{C3} is quite large and the parameter is now at the lower limit. We added remarks on this (and the low values of g1) to the text.

20. P15. Lines 28-29: Draft amendment

    “Noticeably the USO optimisation only changes the value of θtsp and q, and leaves the rest of the parameters almost untouched.”

Accepted with slight modifications.

21. P15. Lines 30-33: Draft amendment

    “For ET, the Baseline, Ball-Berry, and USO are greatly improved especially at the drought in summer 2006 when compared to more general optimisation, however too much drawdown was found for Bethy. The Baseline, Ball-Berry, Leuning, and to a lesser degree the Friend and Kiang formulations, now suffer from the too low ET values before the actual drought. GPP was greatly improved both for general and dry period optimisatons except for the drawdown for the Baseline and Bethy at the drought in summer 2006. Drawdown for USO is also clear but successfully reproduce the observed drawdown. The GPP of other formulations has remained roughly the same as with the more generally optimised parameter values. Overall, The Bethy model has a too strong drawdown for both ET and GPP during the drought.”

We did not directly accept this amendment (as there is some repetition), but modified the paragraph in question to clarify it.

22. P17. Line 4: “Fig. 4, right”

We added “rightmost panels”.

23. P18. Line 9: Draft amendment

    “reproducing the fluxes for the validation sites with low LAI (i.e. RU-Zot and US-Prr)”

Accepted.

24. L18. Line 17: Draft amendment

    “We optimised the model for individual (calibration) sites as well (not shown).”

Accepted.

25. P20. Line 1: “The site level estimates of (g0 and) g1 are sensitive not only to”

Accepted.

26. P3. Figure 4: “Bethy (general)” seems better.

We are not entirely certain, what the reviewer indicates here. In Fig. 4, we can understand why Bethy (opt/general) seems better than Bethy (dry), but the b and r^2 values in Fig. 3 show that it is not.

Comments by Reviewer 3

Dear authors,

I checked the authors appropriately addressed all items raised by the referee #2 except following 2 items.

(1) For the following referee's comment.

> 20. 5.5. The sentence is unintelligible. Moreover, the explanation of how parameter ranges (i.e. priors - why don’t you call them priors) are derived is not sufficient. Provide a clear rationale for prior elicitation.

Authors replied as follows.

> Parameter ranges are not priors. This is explained at the beginning of this document.

But I could not find corresponding explanation on the manuscript.

We have modified the section 2.5: “Sampling process”, where the difference between the parameter ranges and priors is now explicitly stated. We also added a focus to the Table 2 (parameter descriptions) caption: “model parameters with default values, range __of acceptable values__”, which should also clarify the role of the parameter ranges.

(2) On the last item from the referee#2, authors replied as follows.

> The code is under MPI-M License agreement and we cannot distribute it. The driving data (approximately 500Mb) and chains can be uploaded e.g. as supplements.

For ensuring computational reproducibility, driving data, at least, should be available on an appropriate repository, which is reasonably accessible for readers.

The driving data (calibration and validation) was uploaded to the Zenodo data portal, as stated in the code availability section of the manuscript. We amended the description in the data availability section to include this information more precisely (instead of just “dataset” we state that it contains the forcing data and observations).

Supplement S3

This supplement is a supporting analysis of the calibration process improvements, when compared to the model with default parametrisations. The analysis is based on the slope of the regression line (b) and the coefficient of determination (r^2) from Fig. 2 and the corresponding supplementary images (S2). We calculated how many times the calibrated parameter values resulted in improvements for these variables (in boldface), how many times these values are roughly the same (the value from the default simulation is within 5 % of the corresponding value from the calibration process) and how many times the calibration has worsened the results (italic).

We urge caution in making detailed or definate conclusions based on these supporting results. This is because the Ball-Berry model and the variants are here coupled to the JSBACH model without any initial calibration. The default parameter values for these models are taken from literature – it is possible that the combination of these values and the JSBACH default parametrisation (for the other parameters) results in inferior behaviour. Therefore, comparing the calibrated results to these simulations may not be meaningful.

Table S3a. Model spesific analysis of the calibration process for the validation period. Improvements are given in boldface, similar behaviour without any accent and deteriorations in italic.

mode	b(ET)	r^2(ET)	b(GPP)	r^2(GPP)	Σ
Base	**2**,3,*5*	**9**,1,*0*	**8**,0,*2*	**8**,2,*0*	**27**,6,*7*
Bethy	**3**,1,*6*	**9**,1,*0*	**8**,0,*2*	**10**,0,*0*	**30**,2,*8*
BB	**7**,1,*2*	**10**,0,*0*	**9**,0,*1*	**9**,1,*0*	**35**,2,*3*
Leu	**5**,0,*5*	**10**,0,*0*	**8**,0,*2*	**8**,2,*0*	**31**,2,*7*
F&K	**2**,2,*6*	**10**,0,*0*	**8**,0,*2*	**10**,0,*0*	**30**,2,*8*
USO	**2**,2,*6*	**9**,1,*0*	**7**,0,*3*	**8**,2,*0*	**26**,5,*9*
Σ	**21**,9,*30*	**57**,3,*0*	**48**,0,*12*	**53**,7,*0*	

Table S3b. Site spesific analysis of the calibration process for the validation period. Improvements are given in boldface, similar behaviour without any accent and deteriorations in italic. Validation site identifiers have also been italised.

site	b(ET)	r^2(ET)	b(GPP)	r^2(GPP)	Σ
CA-Obs	**3**,3,*0*	**6**,0,*0*	**6**,0,*0*	**6**,0,*0*	**21**,3,*0*
CA-Qfo	**2**,1,*3*	**6**,0,*0*	**6**,0,*0*	**6**,0,*0*	**20**,1,*3*
FI-Hyy	**3**,2,*1*	**6**,0,*0*	**6**,0,*0*	**3**,3,*0*	**18**,5,*1*
FI-Ken	**1**,0,*5*	**3**,3,*0*	**5**,0,*1*	**6**,0,*0*	**15**,3,*6*
FI-Sod	**0**,0,*6*	**6**,0,*0*	**6**,0,*0*	**6**,0,*0*	**18**,0,*6*
RU-Fyo	**3**,3,*0*	**6**,0,*0*	**6**,0,*0*	**2**,4,*0*	**17**,7,*0*
CA-Ojp	**2**,0,*4*	**6**,0,*0*	**6**,0,*0*	**6**,0,*0*	**20**,0,*4*
FI-Let	**6**,0,*0*	**6**,0,*0*	**6**,0,*0*	**6**,0,*0*	**24**,0,*0*
RU-Zot	**1**,0,*5*	**6**,0,*0*	**1**,0,*5*	**6**,0,*0*	**14**,0,*10*
US-Prr	**0**,0,*6*	**6**,0,*0*	**0**,0,*6*	**6**,0,*0*	**12**,0,*12*
Σ	**21**,9,*30*	**57**,3,*0*	**48**,0,*12*	**53**,7,*0*	

~~Stomatal conductance, photosynthesis and parameter~~ Parameter calibration and stomatal conductance formulation comparison for boreal forests with adaptive population importance sampler in the land surface model JSBACH

Jarmo Mäkelä[1], Jürgen Knauer[2], Mika Aurela[1], Andrew Black[3], Martin Heimann[2], Hideki Kobayashi[4], Annalea Lohila[1,5], Ivan Mammarella[5], Hank Margolis[6], Tiina Markkanen[1], Jouni Susiluoto[1,7], Tea Thum[2], Toni Viskari[1], Sönke Zaehle[2], and Tuula Aalto[1]

[1]Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland
[2]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany
[3]University of British Columbia, Canada
[4]Institute of Arctic Climate and Environment Change Research, Japan Agency for Marine-Earth Science and Technology, Japan
[5]Institute for Atmospheric and Earth System Research / Physics, P.O. Box 48, Faculty of Science, FI-00014 University of Helsinki, Finland
[6]Department of Forest Sciences, University of Quebec, Canada
[7]School of Engineering Science, Lappeenranta–Lahti University of Technology, P.O. Box 20, FI-53851 Lappeenranta, Finland

Correspondence: Jarmo Mäkelä (jarmo.makela@fmi.fi)

Abstract. We calibrated the JSBACH model with six different stomatal conductance formulations using measurements from 10 FLUXNET coniferous evergreen sites in the Boreal zone. The parameter posterior distributions were generated by the adaptive population importance sampler (APIS)~~and~~ , then the optimal values were estimated by a simple stochastic optimisation algorithm. The ~~observations used to constrain the model are~~ model was constrained with in-situ observations of evapotranspiration (ET) and gross primary production (GPP). We identified the key parameters in the calibration process. These parameters control the soil moisture stress function and the overall rate of carbon fixation.

The JSBACH model was also modified to use a delayed effect of temperature for photosynthetic activity in spring. This modification enabled the model to correctly reproduce the springtime increase in GPP ~~(which was also reflected in ET) for conifers throughout the measurements~~ for all conifer sites used in this study. Overall, ~~we were able to improve~~ the calibration and model modifications improved the coefficient of determination and the model bias for GPP with all stomatal conductance formulations. However, only the coefficient of determination was clearly improved for ET. The optimisation resulted in best performance by the Bethy, Ball-Berry and the Friend and Kiang stomatal conductance models.

We also optimised the model during a drought event in a Finnish Scots pine forest site. This optimisation improved the model behaviour~~but the~~ , but resulted in significant changes to the parameter values ~~were significant. Interestingly,~~ except for the unified stomatal optimisation model ~~demonstrated~~ (USO). Interestingly, the USO model demonstrated the best performance during this event~~with only small changes to the parameter values~~.

1 Introduction

Plants exchange carbon dioxide (CO_2) and water vapour (H_2O) with the atmosphere. Sufficient soil water, irradiance and adequate temperature are required to maintain the exchange ~~rate~~ rates during the growing season. Disturbances in these conditions such as drought, cold temperature or low radiation cause the plants to respond to the environmental stress via stomatal closure and the ~~drawdown of~~ decrease in photosynthesis and transpiration (**???**). The capability of plants to recover from such events depends on species and their adaptation to site conditions (**?**). Stress is part of the normal annual cycle of the plants, but occasionally it may exceed the limits of recovery.

Soil water deficit and high water vapour pressure deficit can result in suppressed plant transpiration (**??**). Globally, soil drought has been recognised as one of the main limiting factors for plant photosynthesis (**?**) and boreal forests are known to occasionally suffer from soil drought (**??**). The recovery of photosynthetic capacity in spring has been connected to temperature history, and to frequency of severe night frosts (**??**) ~~,~~ that can reverse the recovery. Understanding, and correctly modelling, these phenomena are especially important for boreal forests (**?**) under changing environmental conditions.

Ecosystem and land surface models, describing the plant photosynthesis, transpiration and soil hydrology related processes, usually include descriptions and ~~parametrisations~~ parameterisations for various stress effects. These parameters often lack a theoretical foundation (**??**) and descriptions of vegetation drought response and phenology have been recognized to need better ~~representations~~ formulations and design (**????**). These deficiencies restrict ~~the~~ a model's predictive capability under changing environmental conditions, and call for specific ~~parametrisations~~ parameterisations for different plant types and vegetation zones.

~~The stomatal~~ Stomatal conductance models describe the pathway of CO_2 and water through the leaf stomata by an electric circuit analogy (**?**). The variations in stomatal opening and mesophyll structure are interpreted as resistances to ~~the~~ water flow and the process is idealised via generalised parameterisation. ~~The stomatal~~ Stomatal conductance models mainly differ in their choice of ~~variable~~ variables driving the stomatal closure, and their performance has been recently assessed in modelling studies by e.g. **???**. However, it can be hypothesised that the choice of the stomatal conductance model affects the ecosystem model parameters more broadly ~~than just those directly related to the actual~~ as the stomatal conductance formulations ~~as these formulations~~ vary in their responses to the different conditions. A holistic assessment of the performance of the stomatal conductance models together with ecosystem model parameter optimisation has been missing.

In many other studies, where the aim has been to optimise land surface model parameters, the optimisation is based on estimating the gradient of the cost function: **?** for JSBACH, **??** for ORCHIDEE and **?** for JULES. ~~The gradient-based~~ Gradient-based methods are faster than Markov chain Monte Carlo (MCMC) methods as they strongly steer the sampling process to reach a ~~minima of~~ minimum in the cost function (see e.g. **?**). This approach also enables a more indefinite setting of parameter ranges (limits for acceptable parameter values) when compared to methods that sample the full parameter space. However, they are prone to get stuck in ~~a~~ local minima, especially when the dimensionality of the parameter space increases. ~~Lately~~ In the last few years, similar parameter estimations have also been done for CLM by **?** using the DREAM$_{(zs)}$ (MCMC) algorithm with multiple chains, and for JULES by **?** with the BORG algorithm that employs multiple optimisation algorithms

simultaneously. The DREAM algorithm is fully iterative, which limits the number of parallel processes to the number of parallel chains in use (when we do not account for the possibility of the model parallelisation that can be substantial). The applicability of the BORG algorithm is dependent on the algorithms in use and the expertise of the user (to choose the right algorithms etc.).

APIS is a Monte Carlo (MC) method that can be run iteratively as presented by **?** but it is also straightforward to parallelise, since all samples prior to each adaptation (in our simulations 2000 draws) can be drawn and estimated simultaneously. This latter feature is useful to decrease the amount of real time required to run the algorithm when computer resources are not the limiting factor – APIS requires considerably fewer sequential estimates than typical Markov chain methods. In the iterative mode, automatic stopping rules can be easily implemented to indicate when additional samples are not required to improve the estimates. The APIS algorithm samples the full parameter space (as do MCMC methods) ~~, is able to~~ and can utilise a mixture of parameter prior distributions~~and~~ . Therefore, APIS can estimate complicated multidimensional probability distributions with relative ease. These aspects make APIS an attractive alternative to the other sampling and optimisation methods mentioned above.

In this study we apply the land surface model JSBACH for 10 boreal coniferous evergreen forest eddy covariance sites to examine the performance of different stomatal conductance models, and their effect on calibrated parameters related to photosynthesis, phenology and hydrology. ~~We will assess the inter-site variability and focus on a specific drought period at one site. We will provide an assessment of the robustness of the calibration of parameters together with different stomatal conductance descriptions. We~~ First, we utilise APIS to sample the full parameter space with the different stomatal conductance formulations and to locate different modes of the target distributions (peaks of high probability). Second, using the distributions generated by APIS as the prior distributions, we optimise the parameters using a simple stochastic optimisation method. Finally, we assess the inter-site variability and the robustness of the calibrated parameters together with different stomatal conductance formulations. Optimised parameters for a specific drought are also investigated and compared with the parameters for the general optimisation.

2 Materials and methods

We will next introduce the measurement sites, followed by the model and modifications made to it. Afterwards we will give a general overview of the simulations as well as the sampling process, the algorithms and methods used to analyse the results.

2.1 Sites and measurements

We use data from 10 FLUXNET (doi:10.17616/R36K9X) sites characterised as coniferous evergreen forests. Site descriptions with appropriate references are ~~gathered~~ provided in Table 1. The ~~site level~~ site-level half-hourly eddy-covariance (EC) measurements were quality checked and gap-filled when needed to produce continuous half-hourly and daily time series. The gap-filled and ~~low quality~~ low-quality (based on FLUXNET data quality flags) measurements were masked, and the daily aggregates (usually means) were accepted as part of the calibration process if at least 60% of the values between 4:00 and 20:00

(i.e. daytime measurements) for that day were unmasked~~– the acceptance of the daily values for the calibration is based on the quality of the "daytime" measurements but all of the values are used to drive the model~~. The daily aggregates of ET and GPP were used to calibrate and validate the model, whereas the half-hourly data were used as climate forcing (as explained later in Section 2.4.

Based on the quality and quantity of their respective measurements, the sites were divided into calibration and validation sites. Essentially, if we have enough data from a site, it is used for both calibration and validation purposes. We ~~require~~ required the site to have at least eight years of measurements, where the first five ~~are~~ were used for calibration, and the consecutive three for validation. Otherwise we ~~use~~ used the site only for a three year validation. The FLUXNET datasets were missing both the long- and shortwave radiation for the two Russian sites, Fyodorkovskoye (RU-Fyo) and Zotino (RU-Zot). These were generated from ERA Interim data. The soil types of all of these sites can mostly be identified as mineral soils with varying sand, clay and peat contents. Fyodorovskoye and Poker Flat (US-Prr) are natural peatlands and Lettosuo (FI-Let) is a drained peatland site.

The measurement error in the EC flux data ~~can be~~ were separated into systematic and random errors. The main systematic errors (density fluctuations, high-frequency losses, calibration issues) ~~are~~ were taken into account as part of the post-processing of the data, and the random errors tend to dominate the uncertainty of the instantaneous fluxes. The random error is often assumed Gaussian but can be more accurately approximated by a symmetric exponential distribution (**?**). It increases linearly with the magnitude of the flux, with a standard deviation typically less than 20% of the flux (**??**). Our treatment of the measurement (and model) errors is explained in ~~section~~ Section 2.9.

Table 1. Descriptions for the sites used in this study sorted by their FLUXNET identifier. The first six sites are used for both calibration and validation purposes, with the first five years of each site used for calibration. The last three years as well as the last four sites are used for validation only. The reported elevation is in meters above sea level, LAI is the one-sided leaf area index and the average stand age is in years, along with average annual precipitation (P) in mm and temperature (T) in degrees Celsius.

Site id	lat	lon	elev.	dom. species	LAI	age	P	T	years	reference
CA-Obs	53.99	-105.12	629	*Picea mariana*	3.8	135	406	0.8	1999–2006	**?**
CA-Qfo	49.69	-74.34	382	*Picea mariana*	3.7	112	962	-0.4	2003–2010	**?**
FI-Hyy	61.85	24.29	180	*Pinus sylvestris*	3.5	45	709	2.9	1999–2006	**?**
FI-Ken	67.99	24.24	337	*Picea abies*	2.1	100	484	0.4	2003–2010	**?**
FI-Sod	67.36	26.64	179	*Pinus sylvestris*	1.7	150	527	-0.4	2001–2008	**?**
RU-Fyo	56.45	32.90	265	*Picea abies*	4.5	200	711	3.9	2002–2009	**?**
CA-Ojp	53.92	-104.69	579	*Pinus banksiana*	2.6	100	431	0.1	2004–2006	**?**
FI-Let	60.64	23.96	119	*Pinus sylvestris*	6.0	40	627	4.6	2010–2012	**?**
RU-Zot	60.80	89.35	121	*Pinus sylvestris*	1.5	215	493	-3.3	2002–2004	**?**
US-Prr	65.12	-147.49	210	*Picea mariana*	0.7	72	275	-2.0	2011–2013	**?**

2.2 The JSBACH model

JSBACH (**?**) is a process-based ecosystem model and the land surface component of the Earth System model of the Max Planck Institute for Meteorology (MPI-ESM). We ran JSBACH offline using meteorological measurements from the flux towers to force the model. Implications of this one-way coupling with the atmosphere include lack of feedback from the surface energy balance to the atmosphere, i.e. latent and sensible heat fluxes and surface thermal radiation do not directly affect prescribed air temperature or humidity. Similarly, the feedback of the surface to the vertical transfer coefficients within the atmospheric surface layer is missing as the wind speed that drives mixing is prescribed. Furthermore, since we use site level data (each site is represented as a single grid point), the grid resolution does not affect the results.

We focus only on the most essential parts of JSBACH relating to our work. A more complete model description with details on e.g. soil heat transfer, water balance and coupling to the atmosphere can be found in **?**, whereas **?** provides a more descriptive synopsis on land-surface interactions, **?** ~~supplements both with~~ complements both with an addition of land cover change processes, and **?** introduces soil hydrological mechanisms within a multilayer scheme applying five layers.

In JSBACH, the land surface is divided into grid-cells, which are split into bare soil and vegetative areas. The vegetative area is further divided into tiles representing the most prevalent vegetation classes, called plant functional types (PFTs) (**?**). In our site-level simulations, the model was set to use only one PFT, coniferous evergreen trees. The seasonal development of leaf area index (LAI) for the trees is regulated by air temperature and soil moisture with a single limiting value (for all sites) for the maximum of LAI. This maximum value was fixed and the site-specific fractions of vegetative area were adjusted to reproduce the measured site level LAI.

The predictions of phenology are produced by the Logistic Growth Phenology (LoGro-P) ~~model of~~ sub-model in JSBACH (**?**). Photosynthesis is described by the biochemical photosynthesis model (**?**). Following **?**, we set the maximum electron transport rate (J_{max}) at 25 degrees Celsius to 1.9 times the maximum carboxylation rate ($V_{C,max}$), which is in line with e.g. **??**. The photosynthetic rate is dependent on the used stomatal conductance formulation, introduced in ~~chapter~~ Section 2.3. Radiation absorption is estimated by a two stream approximation within a three-layer canopy (**?**). Especially in sparse canopies, ~~the~~ radiation absorption is affected by clumping of the leaves which is here taken into account according to the formulation by **?**.

Parameters detailing site-specific soil properties, such as soil porosity and field capacity, were derived from FLUXNET datasets and the references in Table 1. We approximated the soil ~~compositions~~ composition and generated these properties following **?**.

2.3 Modifications to the JSBACH model

All parameters of interest, presented in Table 2, were extracted from the JSBACH model code to an external file to facilitate the simulations. The default values of ~~new parameters (~~newly added parameters (not originally in JSBACH: τ, q, g_0, g_1) ~~are~~ were derived from a synthesis of literature values. Most of the parameter ranges (limiting values for the parameters) were adapted from our previous work on a similar topic (**?**). The parameter grouping was done to enhance optimisation and the mechanism

is explained in ~~chapter~~ Section 2.7. Group I consists of parameters most directly affecting photosynthesis, group II parameters are intimately involved with soil moisture, and group III are the logistic growth phenology (LoGro-P) model parameters. The equations governed by these parameters are presented in Appendix A.

Table 2. Descriptions of model parameters with default values, range of acceptable values and references to equations in the manuscript or in the ~~appendixes~~appendices. Parameters in the same group were calibrated simultaneously.

Parameter	def	range	Units	Group	Description	Eq.
$V_{C,max}$	62.5	[40,65]	$\diamond$	I	Farquhar model maximum carboxylation rate at 25°C of the enzyme Rubisco (coupled with maximum electron transport rate at 25°C with a factor of 1.9) [$\diamond = \mu$ mol(CO_2) m^{-2} s^{-1}].	A2
α	0.28	[0.26,0.32]	-	I	Farquhar model efficiency for photon capture at 25°C.	A4
τ	10.0	[5,15]	days	I	Adjustment period length in acclimation of photosynthesis.	1
c_b	5.0	[4,7]	-	I	Multiplier in momentum and heat stability functions (**?**).	-
f_{C3}	0.87	[0.7,0.95]	-	I	Ratio of unstressed C3-plant internal/external CO_2 concentration.	A3
q	0.0	[0,1]	-	I	Exponential scaling of water stress in reducing photosynthesis.	A1
g_0	0.001	[1E-5,5E-3]	$\triangledown$	I	Residual stomatal conductance [$\triangledown$ = mol m^{-2} s^{-1}].	B3
g_1	Values in Table 3		-	I	Slope of the stomatal conductance function.	B3
a	2.8	[1.5,3.5]	-	I	Base rate of stomatal conductance response to atmospheric humidity for the Friend and Kiang model.	B3
d	80	[50,120]	-	I	Exponential rate of stomatal conductance response to atmospheric humidity for the Friend and Kiang model.	B3
θ_{dr}	0.9	[0.5,0.95]	-	II	Volumetric soil water content above which fast drainage occurs.	A6
θ_{hum}	0.5	[0.2,0.8]	-	II	Fraction depicting relative surface humidity based on soil dryness.	A9
θ_{pwp}	0.35	[0.15,0.4]	-	II	Volumetric soil moisture content at permanent wilting point.	2
θ_{tsp}	0.75	[0.25,0.8]	-	II	Value of volumetric soil moisture content above which transpiration is unaffected by soil moisture stress (β); and $0.9\theta_{tsp} \geq \theta_{pwp}$.	2
p_{int}	0.25	[0.15,0.35]	-	II	Fraction of precipitation intercepted by the canopy.	A5
s_{sm}	5.9E-3	[1E-4,0.1]	m	II	Depth for correction of surface temperature for snow melt.	-
w_{skin}	2.0E-4	[1E-5,5E-3]	m	II	Maximum water content of the skin reservoir of bare soil.	-
C_{decay}	13.0	[5,25]	days	III	LoGro-P: memory loss parameter for chill days.	A12
S_{min}	10.0	[5,30]	°C days	III	LoGro-P: minimum value of critical heat sum.	A12
S_{range}	150.0	[100,300]	°C days	III	LoGro-P: maximal range of critical heat sum.	A12
T_{alt}	4.0	[2,10]	°C	III	LoGro-P: cutoff in alternating temperature.	A10
T_{ps}	10.0	[3,25]	°C	III	LoGro-P: memory loss parameter for pseudo soil temperature.	A14

The start of the growing season in the JSBACH model is defined by a "spring event" in the LoGro phenology model (appendix A3) that induces leaf growth. The phenology model calculates a sum of ambient temperature (heatsum) since last autumn that is above the cutoff value T_{alt}, presented in Eq. (A10). It also calculates a variable threshold, defined in (A12), for the heatsum to reach. The threshold decreases based on the number of days the ambient temperature is below T_{alt}, whereas the heatsum increases. When the heatsum reaches the threshold, the plant leaves are free to grow.

However, coniferous evergreen trees do not shed all of their leaves for winter ~~. The~~ and the existing foliage enables them to quickly ~~instigate~~ initiate photosynthesis in the following spring. The start of the photosynthetically active season in the model has been observed to occur too early in the Boreal region by e.g **?**. ~~To~~ In order to correct this behaviour i.e. to restrain the respiration and photosynthesis of conifers in the early spring, we utilise a delayed effect of temperature for photosynthetic activity, introduced by **?**. To calculate the reduction, we must first define the state of photosynthetic acclimation that **?**, p.371 present as: "an aggregated measure of the state of those physiological processes of the leaves that determine the current photosynthetic capacity at any moment".

The state of acclimation (S) is calculated from air temperature (T) with a delay prescribed by parameter τ (this is similar to the calculation of T_S in appendix A14). S is then inserted into sigmoidal relation Eq. (1) to calculate a factor γ, a formulation that is adapted here from **?**. Finally, γ is used to reduce the photosynthetic efficiency in Eq. (A1). $T_{1/2}$ denotes the inflection point where γ reaches half of γ_{max}, ~~b~~ k is the curvature of the function and $\gamma = 1$ when $S \geq 10$.

$$\frac{dS}{dt} = \frac{T - S}{\tau}, \qquad \gamma = \frac{\gamma_{max}}{1 + e^{b(S - T_{1/2})}} \frac{\gamma_{max}}{1 + e^{k(S - T_{1/2})}} \tag{1}$$

The JSBACH model was also modified to include altogether six different stomatal conductance formulations following **?**. These formulations include the pre-existing Baseline and Bethy versions as well as the Ball-Berry model and three of its variants. Model information is gathered in Table 3 for easy referencing and the detailed formulations are given in appendix B. The limits of the slope of the stomatal conductance formulation parameter (g_1) were set to reflect commonly observed values from physiological measurements (**?**). The limits of g_1^{USO} reflect the results presented by **?**.

Table 3. Stomatal conductance models with default values and range for g_1 and references to equations in Appendix B as well as related articles. The $\star$ symbol indicates the Ball-Berry model and its variants.

Stomatal conductance model	short	g_1	range	references
Baseline	Base	-	-	B1 **?**
Biosphere-Energy-Transport-Hydrology	Bethy	-	-	B2 **?**
$\star$ Ball-Berry	BB	9.0	[4,10]	B3 **?**
$\star$ Leuning	Leu	8.0	[6,10]	B3 **?**
$\star$ Friend and Kiang	F&K	9.5	[7,11]	B3 **?**
$\star$ Unified stomatal optimisation	USO	2.0	[1.5,3.5]	B3 **?**

We have also included two additional parameters (a and d in Table 2) for the Friend and Kiang (**?**) stomatal conductance formulation in B3. These parameters were not originally included in the optimisation, but the resulting cost function (9) values were poor when compared to the other formulations. At that point, these parameters were included in the optimisation process. This increases the degrees of freedom for the Friend and Kiang model by two and therefore may give it an advantage when compared to the other Ball-Berry type formulations, which has to be considered in the interpretation of the results.

All of the stomatal conductance models contain an empirical water stress factor β, which reduces stomatal conductance as a function of volumetric soil water content (θ).

$$\beta = \begin{cases} 1, & \theta \geq \theta_{tsp} \\ \frac{\theta - \theta_{pwp}}{\theta_{tsp} - \theta_{pwp}}, & \theta_{pwp} < \theta < \theta_{tsp} \\ 0, & \theta \leq \theta_{pwp} \end{cases} \tag{2}$$

In JSBACH, the stomatal conductance (g_s) is primarily resolved to estimate carbon fixation. The same g_s is then later used to calculate transpiration (A8). In the original JSBACH formulation ~~,~~ (i.e. the Baseline version~~, the stomatal conductance (~~), the g_s ~~)~~ is first resolved for unstressed canopy and then scaled by the water stress factor β. The Bethy approach ~~assumes that transpiration is either limited by atmospheric demand or water supply~~ is similar, but the conductance can also be limited by water supply (B2). In cases when the water supply is not the limiting factor, the calculations are similar to the Baseline version. In all of the empirical Ball-Berry variants, the stomatal conductance can be written as $g_s = g_0 + c\beta g_1$. The residual conductance (g_0) and the slope of the function (g_1) are both formulation specific parameters as well as the factor c, that incorporates net photosynthesis and effects of atmospheric humidity and CO_2 concentration. The parameters g_0 and g_1 are part of our sampling and optimisation processes (group I in Table 2 when applicable).

The water stress factor (β) limits the carbon fixation and transpiration via the stomatal conductance formulation. Following **?**, it is also used to directly limit the net assimilation rate (A_n), as seen in (A1). The additional scaling (or limiting) factor for A_n takes the form β^q, so it is a function of both soil water content θ and the parameter q. Maximal reduction is achieved when $q = 1$ and the reduction factor reverts to β. The minimal reduction occurs when $q = 0$ and the reduction factor resembles a step function (at $\theta = \theta_{pwp}$). For any other value of q, it is a continuous convex function between the two extremes $\beta^q : [\theta_{pwp}, \theta_{tsp}] \to [0, 1]$.

2.4 Model simulations

The site level measurements, used as model inputs, are air temperature, air pressure, precipitation, humidity, wind speed and CO_2 concentration as well as short- and longwave and potential shortwave radiation. Additionally, evapotranspiration (ET) and gross primary production (GPP), derived from the eddy covariance (EC) measurements, are used to constrain and evaluate the model (as explained later in ~~sections~~ Sections 2.8 and 2.9). We drive the model with half-hourly data but output daily values.

The initial state of the JSBACH model can be generated from predefined values of state variables (usually empty initial storage pools) or the model can be restarted from a file describing the state of some previous run. Depending on the area of interest, a model spin-up may be required to bring the model into a steady state. In our simulations, some of the more slowly

changing variables (e.g. soil water content and LAI) need to be equilibrated, so a spin-up is required. This can be achieved by running the model over a set of measurements multiple times, each time restarting from the final state of the previous run.

The calibration period consists of the first five years given for the calibration sites in Table 1. The spin-up is achieved by looping over these five years, altogether four times (~~20 year~~ 20-year spin-up) and then saving the state of the model at the end of the run. The actual calibration is started from the beginning of the calibration period, using the previously saved state variables. To reduce any bias this induces, the first year in the calibration run is removed from the cost function calculations. The spin-ups for the validation sites in Table 1 are similarly generated.

During the summer 2006, the Hyytiälä (FI-Hyy) measurement site suffered from a severe drought (**?**), leading ~~e.g.~~ to visible discolouration of needles. These events are difficult for models to capture and hence are of interest to modellers. We have previously and unsuccesfully attempted to optimise the JSBACH model (**?**) for this event. Here we focus directly on the extended dry period (190–260th day of the year in 2006), during which the actual drought is mostly in effect between 210–235th DOY. We ~~fix~~ adjusted some of the parameter values as those uncovered by the more general calibration, presented above. The spin-up was the same as for the calibration period, but at the end of the spin-up, the model was run forward to the start of the year 2006. Only values between the 190–260th day of the year (DOY) in 2006 were used in constraining the model.

2.5 Sampling process

We describe the modelling setup with the equation $\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}, \mathbf{x}) + \mathbf{e}$, where the aim is to reproduce the observations ($\mathbf{y}$) with our model ($\mathcal{M}$), the driving data ($\mathbf{x}$) and the current parameter values ($\boldsymbol{\theta}$). The residuals ($\mathbf{e}$) depict how well the model reproduces the observations and they form the basis of the likelihood function (formulated in ~~section~~ Section 2.9), that is used to derive the parameter posterior distributions.

Using Bayes' rule on conditional probability we can write the parameter posterior density ($p(\boldsymbol{\theta}, \mathcal{M}|\mathbf{x})$) as a function of the likelihood ($\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathcal{M})$), parameter prior distributions ($\pi(\boldsymbol{\theta})$) and the model evidence ($Z(\mathbf{x}|\mathcal{M})$). As usual and from here on, we do not write $\mathcal{M}$ in the Bayes' formula:

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{Z(\mathbf{x})} \tag{3}$$

We can now utilise the posterior density as a probability density for the parameters and infer the expectation values:

$$E[\boldsymbol{\theta_i}] = \frac{1}{Z}\int \boldsymbol{\theta_i} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}, \qquad Z = \int p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \tag{4}$$

Above $\boldsymbol{\theta}_i$ is the i-th element of the parameter vector. Generally, Eq. (4) cannot be analytically solved, hence it is usually estimated numerically. Commonly this is achieved by one of the many Markov chain Monte Carlo (MCMC) methods, but in this study we apply the adaptive population importance sampler (APIS) ~~,~~ defined by **?**. ~~The~~ APIS (**?**) is a Monte Carlo (MC) method that utilises a population of importance samplers (IS) to jointly estimate the target pdf ($p(\boldsymbol{\theta}|\mathbf{x})$) and the normalising constant ($Z(\mathbf{x})$) by a deterministic mixture approach (**??**), whereas the MCMC methods do not care about the value of Z ~~and for APIS we introduce~~ . We denote the importance sampling density as $q(\boldsymbol{\theta})$.

$$E[\boldsymbol{\theta_i}] = \frac{1}{Z}\int \boldsymbol{\theta_i} r(\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta}, \qquad \text{where} \qquad r(\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|\mathbf{x})}{q(\boldsymbol{\theta})} \tag{5}$$

Above r is the reweighing factor that is the driving force in importance sampling. We will next give a summary description of the sampling process with comparison to a general multichain MCMC approach (since MCMC methods are more commonly used in these types of situations).

1. The initialisation of a multichain MCMC sampler and APIS are very similar. In our simulations, APIS is set up as 40 simultaneous and independent importance samplers. This is similar to an independent 40-chain MCMC sampler. Each sampler or chain has a random starting location ~~and sampling distribution (we use truncated Gaussian distributions) that will evolve throughout the process. The starting locations were sampled~~ drawn from a uniform distribution defined by the parameter ~~limits (~~ranges, given in Table 2~~)~~. The initial sampling (or prior) distribution for each sampler is also randomly generated – we use truncated Gaussian distributions with diagonal covariance matrices, where the standard deviations are randomised. The sampling distributions will evolve throughout the process.

2. In an MCMC setup, the model would be run once (for each chain), evaluated and then the draw (parameter values) accepted or rejected accordingly. In APIS, instead of a single element (one run) we use a sample size of 50. This means that we draw 50 elements with each IS sampler (or "chain") independently. These draws are then evaluated and reweighted as presented in Eq. (5).

3. The 50 reweighted draws (for each IS sampler separtely) are used to calculate a new location for the sampling distribution. This location is automatically accepted (no rejection criteria) and we also adapt the shape of the distribution using the self-normalising AMIS estimator by **?**.

4. Additionally, all of the draws in APIS are used to calculate "global" estimates of the parameter expected values. This process utilises the deterministic mixture approach (**??**) and ~~it~~ is fully iterative ~~–~~ with no need for any recalculations as the previous estimates are directly adjusted (no information is lost either).

MCMC chains track the evolution of single elements, and occasionally adjust the sampling distribution. The sample size in APIS is larger (it is not a Markov chain method) and the focus is on the evolution of the locations of the sampling distributions, not on the individually drawn elements. These location parameters are expected to be around all the modes of the target and the deterministic mixture ensures the stability of the estimation of the (global) parameter expected values. As an importance sampler, APIS is also a variance reducing method.

Before taking a more detailed look at APIS, we make some further notes about the sampling process. The first element of the 50 draws (item 2 in the list above) is always fixed as the current mean. ~~This requirement stems from a need to reduce computational time. Running the model to a steady state (chapter 2.4) for each parameter set is costly. Hence we~~ We run the spin-up (Section 2.4) and generate the model starting state only for the proposal means, and use the same state for the other 49 draws ~~. This induces~~ (perturbed around the proposal mean). This requirement stems from a need to reduce computational time as running the model to a steady state is costly. This approach might induce some discrepancies, but they are mitigated by removing the first year of the calibration simulations (as explained in ~~section~~ Section 2.4). We also slightly ~~scale (reduce) reduce~~ the importance weights ~~based on the distance of the corresponding sample to the mean of the proposal. This scaling~~

~~is only used in the~~ of the 49 samples (more reduction for samples further from the proposal mean), when calculating the new location parameters (item 3 in the list above) – the reduction only (slightly) slows the adaptation of the ~~new location. Additionally~~IS sampler locations. Finally, we note that this approach ensures that we run the proposal means, that are the focus in APIS, with the correct spin-up.

2.6 Adaptive population importance sampler

~~APIS (?) is a Monte Carlo (MC) method that utilises a population of importance samplers (IS) to jointly estimate the target pdf ($p(\boldsymbol{\theta}|\mathbf{x})$) and the normalising constant ($Z(\mathbf{x})$) by a deterministic mixture approach (??).~~

Normally, only the location parameters of the IS proposals are adapted, but we also adapt the shape parameters using the self-normalising AMIS estimators by **?**. ~~The~~ APIS is able to utilise different or a mixture of normalised proposals densities, but we use truncated Gaussian proposals with diagonal covariance matrices.

In our simulations, APIS is formed of 40 independent IS estimators. Each estimator draws a sample $\boldsymbol{\theta}_i, i \in \{1, ..., N\}$, of size $N = 50$ at a time from their own proposal distribution $q_j(\boldsymbol{\theta}), j \in \{1, ..., M\}, M = 40$. The estimator then calculates the importance weights ($w_{ij} = \frac{p(\boldsymbol{\theta}_i|\mathbf{x})}{q_j(\boldsymbol{\theta}_i)}$) for each sample. The location ($\boldsymbol{\mu}_j$) and shape ($\mathbf{C}_j$) parameters (**?**) of each proposal are updated using only samples (and weights) drawn from q_j. The new shape parameters are formed as a mean of the previous estimate and $\mathbf{C}_j$, as calculated below.

$$\boldsymbol{\mu}_j = \frac{\sum_i w_{ij}\boldsymbol{\theta}_i}{\sum_i w_{ij}}, \qquad \mathbf{C}_j = \frac{\sum_i w_{ij}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_j)(\boldsymbol{\theta}_i - \boldsymbol{\mu}_j)^T}{\sum_i w_{ij}} \tag{6}$$

The simple IS estimators alone are rarely sufficient if the target is even slightly complicated. One classical way of tackling this problem is to join multiple IS estimators together. The simplest approach is to calculate the weights for each of these estimators separately and to normalise the result by the combined sum of all weights. However, this leaves the estimators susceptible to "bad" proposals. ~~The~~ APIS suppresses the bad proposals by utilising the deterministic mixture approach (**??**) presented in Eq. (7), where each proposal q_j is evaluated at all the drawn samples and weighed by the amount of samples drawn ($N_j = 50$) from that proposal. This is equivalent to joining the normalised proposal densities together and evaluating the joint pdf.

$$w_{ij} = \frac{p(\boldsymbol{\theta}_{ij}|\mathbf{x})}{\sum_j \left(\frac{N_j}{\sum_k N_k}\right) q_j(\boldsymbol{\theta}_{ij})} \tag{7}$$

The parameter expectation values and the normalising constant in Eq. (5) can now be estimated by Monte Carlo integration using weights calculated in Eq. (7).

2.7 Parameter optimisation

The APIS algorithm is a rather robust method meant for examining the full target probability distribution and ~~e.g.~~ locating the modes of the target distribution. Adaptation in APIS utilises multiple draws simultaneously, which can easily lead to few parameters ~~dominating~~ controlling this process (the marginal density of one or few parameters ~~overshadows~~ dominates

the calculations). Since we also did not run the model spin-up for all drawn samples (although the discrepancies should be minimal), we utilise a simple custom stochastic optimiser to locate the optimal set of parameter values. This optimiser is run after the APIS calibration simulations and ~~it utilises the same datasets as APIS~~separately for the drought period. The optimiser utilises the exact same datasets (calibration, validation, observations etc.) as APIS, the spin-up is generated for all drawn samples separately and the initial state of the algorith is the mean value of the APIS final configuration (location parameters).

Our optimiser is a simple random sampler amplified by the "velocity" of the last jump (the idea is similar to Hamiltonian or Hybrid Monte Carlo by **?**). We draw a set of samples from a small Gaussian proposal distribution in the vicinity of the current best estimate and calculate the cost function for the samples. Whenever a better point is found (smaller cost function), we jump to that (update the mean of the proposal distribution). The "velocity" of the jump (for us merely distance of change in each parameter) is then added to the new mean (with a maximal limit of one standard deviation in the proposal distribution), but it is reduced and eventually removed if a better sample is not found.

The covariance matrix of the proposal distribution is recalculated at predefined intervals (for all parameters). Additionally, we utilise a subset sampling procedure, where the samples are first drawn from the full parameter space, in the next step they are drawn only from group I in Table 2 (the rest are kept at their current optimal values), followed by groups II and III and then back to the full parameter space. When the number of parameters is reduced, we are more likely to find a better set of parameter values. We have kept the parameters mostly affecting the same processes in the same group, but some dependencies may not be apparent and hence it is also important to draw samples from the full parameter space.

2.8 Simulation analysis

Even though APIS is not a Markov chain method, we can (naively) interpret the evolution of the location paramaters of each IS sampler as chains. The resulting 40 chains have random starting positions but they are relatively short (we present results from the Bethy calibration, where the chains were adjusted 100 times), hence we did not discard any of the samples. We test the convergence of these chains with the Gelman-Rubin diagnostic tests (**?**), comparing the variance between the chains to the variance within each chain, and calculating the potential scale reduction factors ($\hat{R}$). We also test the stability of the (parameter) global expected value estimate (using the deterministic mixture approach) by calculating the difference of the final global expected value and the mean of the location parameters (at each iteration). We denote this test as δ and report the number of the ~~iteration~~ iterations when this difference is below 5% of the parameters range, given in Table 2.

In order to visualise the results, we have utilised a Gaussian kernel density estimation (KDE) to produce distributions from the APIS simulation location parameters. In practice, KDE places a Gaussian distribution centred at each sample and the constructed composite distribution is an estimate of the underlying actual distribution. The bandwidth for the distributions is calculated using the Scott's rule (**?**): the data covariance matrix is multiplied by a factor $n^{\frac{-1}{d+4}}$, where n is the number of data points and d is the number of dimensions.

The effectiveness of each parameter was calculated from the final state of each optimisation process. This was done by first setting all parameters to their optimised values. Then we (evenly) sampled each parameter separately from their range of acceptable values, given in Table 2, and calculated the corresponding cost functions. For each parameter the maximum

difference in these cost function values (and the optimised value) was recorded. The parameters (within each optimisation) were then ordered by these numbers (with highest difference meaning highest effectiveness) and separated into three groups with highest (most effective) and lowest (least effective) effectiveness values, and the rest. This effectiveness relates to how the APIS "sees" the sampling process – the 50 draws are evaluated simultaneously and a very effective parameter can easily mask the influence of a less effective (the marginal density of one or few parameters dominates the calculations).

We report the slope of the regression line (b) and the coefficient of determination (r^2), between the observations (y_i) and the model output (x_i). The slope of the regression line is highly indicative of the model bias (difference of the expected values of the observations and the model). Hence we interpret the bias directly from b (in our results the regression lines pass near origin so the differences this induces are negligible).

$$b = \frac{\sum_i (x_i - \overline{x_i})(y_i - \overline{y_i})}{\sum_i (y_i - \overline{y_i})^2}, \qquad r^2 = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (y_i - \overline{y_i})^2} \tag{8}$$

2.9 Cost function

The Bayesian framework requires a likelihood function that optimally combines pointwise model and observational errors. The JSBACH model error is unknown as is the (pointwise) observation error. We could use a general type of error estimate (such as that of 20% of the flux value) for the observations, but would have to include a minimal site and instrumentation dependent precision. In this study, the full error is treated as Gaussian white noise. Because of these limitations, we are calling and defining our likelihood as a cost function. It is calculated with the same parameter values for each site~~separately using daily values,~~ , using site spesific daily measurements with the gap-filled, low-quality and winter (between the 315th and the 75th day of the year) values removed (resulting in N_{ET} and N_{GPP} points). ~~The cost functionis then averaged over the sites and~~ These site level estimates are averaged to produce the actual cost function, which is then returned for the algorithm to produce an estimate that is independent of the characteristics of any single site.

The cost function (9) in our simulations is based on the normalised mean squared error (NMSE) estimates of the daily gross primary production (GPP) and the daily evapotranspiration (ET). The residual of each variable is divided by the mean of observations, as has been previously done by e.g. **????**. We make use of this approach since we needed to balance two series of different magnitudes (ET and GPP). The residuals are additionally divided by the (site specific) number of observations so that the cost function is not biased towards any specific site. The cost function (without the normalisation) can be interpreted as a negative log-likelihood function with a (~~gaussian~~Gaussian) error term equal to the observational mean.

$$cf_1 = \overbrace{\frac{1}{N_{ET}} \sum \left(\frac{ET_{mod} - ET_{obs}}{\overline{ET_{obs}}} \right)^2}^{\mathrm{NMSE}_{ET}} + \overbrace{\frac{1}{N_{GPP}} \sum \left(\frac{GPP_{mod} - GPP_{obs}}{\overline{GPP_{obs}}} \right)^2}^{\mathrm{NMSE}_{GPP}} \tag{9}$$

We also use a modified version of this cost function, where the NMSE's are weighted by factors based on coefficients of determination (r^2) defined in Eq. (8). This latter cost function is only used during the separate drought period optimisation for Hyytiälä. During the drought we are more interested in the correct timing of the change in GPP and ET fluxes, rather than the size of the actual change. The aim is to correctly reproduce the changes in the water use efficiency (WUE) of plants, which we

interpret here as the pointwise ratio of (ecosystem level) GPP to ET. The NMSE values ensure that the overall amplitude of the fluxes will remain satisfactory.

$$cf_2 = (1 - r^2_{ET})\text{NMSE}_{ET} + (1 - r^2_{GPP})\text{NMSE}_{GPP} \tag{10}$$

3 Results

First we present the performance of the APIS algorithm and the parameters themselves, followed by site and stomatal conductance model specific results~~and lastly~~, and finally an examination of the Hyytiälä drought event in 2006. For simplicity, we ~~equate~~ use the name of the stomatal conductance model to refer to the JSBACH model ~~version utilising that~~ utilising that stomatal conductance formulation.

The evolution of the APIS algorithmic process is presented in Fig. 1 for three parameters from the calibration of the Bethy model. The chosen parameters highlight different levels of identifiability for the algorithm (with the given cost function). The first parameter (f_{C3}) shows a well identifiable situation, where the algorithm quickly locates the area of high probability. The second parameter (θ_{dr}) is also identifiable but the speed of convergence is diminished. The last example (C_{decay}) represents situations where the parameter is not constrained. We have included images of the APIS chains for the other parameters as supplement S1 along with parameter posterior estimates at 20 iterations with the Bethy and Ball-Berry ~~formulaations~~formulations.

We also report the results of the Gelman-Rubin (**?**) and δ tests in Table 4. Both of these tests indicate that the algorithm is performing well at 20 iterations – the values of $\hat{R} \approx 1$, which means that further simulations are unlikely to improve the variance estimates. However, for some parameters, the convergence of the global estimate is slow (as also seen in the supplementary image S1 for e.g. τ, c_b and q). The APIS sampling process did not reveal any multimodal distributions and thus provided suitable initial conditions for the optimisation.

Table 4. Parameter scale reduction $\hat{R}$ (at APIS iteration) and stability δ (~~with a~~ threshold number of iterations) estimates from the Bethy simulations.

	$V_{C,max}$	α	τ	c_b	f_{C3}	q	θ_{dr}	θ_{hum}	θ_{pwp}
$\hat{R}$ at 20	1.12	0.99	1.02	0.99	1.0	0.99	1.0	1.3	1.08
~~R~~ $\hat{R}$ at 100	1.3	1.03	1.25	1.16	1.03	1.08	1.03	1.52	1.16
δ ($\pm$0.05)	20	21	27	40	0	36	18	14	17
	θ_{tsp}	p_{int}	s_m	w_{skin}	C_{decay}	S_{min}	S_{range}	T_{alt}	T_{ps}
~~R~~ $\hat{R}$ at 20	0.99	1.01	0.99	1.0	0.99	0.99	0.99	0.99	0.99
$\hat{R}$ at 100	1.06	1.13	1.0	0.99	0.99	1.0	0.99	0.99	0.99
δ ($\pm$0.05)	26	35	8	0	12	22	0	1	0

3.1 Optimised parameters

The results of the optimisation process are gathered in Table 5. There is an overall agreement on the values of the most prevalent parameters (~~bolded~~ see the bold and italic characters in Table 5 between the models). Most notably, the permanent wilting point (θ_{pwp}) and the point above which transpiration is unaffected by soil moisture stress (θ_{tsp}) have been significantly lowered. The LoGro phenology parameters, ~~that~~ which affect the timing of the spring and autumn events, are expected to contribute only little to the cost function. The coniferous evergeen trees do not shed all their leaves for winter and therefore the timing of the bud burst is not as critical as for e.g. deciduous trees. Additionally, because of the existing foliage, the state of acclimation parameter τ ~~also affects the vegetation active period and~~ that depicts the reduction of carbon assimimilation in the early spring likely dominates the phenology parameters that determine when new leaves start to grow.

Some of the parameters have converged to their limiting values, which can reflect deficiencies in the model structure or the preset parameter ranges. Convergence to the boundary can also be a problem in model calibration, but in this experiment, the algorithms were able to cope with the situation as APIS located the area of high probability and the optimiser located the maxima. The different parameter effectiveness levels reported in Table 5 can be roughly equated to the identifiability situations in Fig. 1. The effectiveness levels are highly situational (e.g. they depend on the sampling limits in Table 2 given for each parameter) and merely reflect the parameter identifiability in the APIS process. Low effectiveness complements the test results in Table 4, as the tests may indicate good performance for a parameter (e.g. for S_{range}) that is ineffective in the simulations.

3.2 Annual cycles

We present the average annual cycles for the validation period and for all sites in Fig. 2 using the Bethy formulation that is part of the standard model. The annual cycles ~~of the Bethy model are more in line with the Ball-Berry variants than those of the Baseline model (see supplements~~ generated with the other stomatal conductance models are added as supplement S2~~for the yearly cycles of the other models)~~. The parameters of the regression lines (b and r^2) between the measured and modelled ET and GPP fluxes of all the models are gathered in Table 6. These indicators have been calculated using all corresponding values regardless of the quality of the data. The sites are in the same order as in Table 1 with the six calibration sites first, followed by the four sites used only for validation. We have also included a supporting synthesis of the b and r^2 values between the model simulations with the default and optimised parameter values as supplement S3.

The optimisation has improved the model ~~results~~ bias and the correlation coefficients for the GPP in Fig. 2 for ~~all of the calibration sites and at least for half of the validation sites. The~~ nearly every site, with the exception of deteriorating bias for Poker Flat (US-Prr) and Zotino (RU-Zot). Additionally, the improvement in the timing of the springtime increase in the GPP is apparent. ~~The~~ All of the correlation coefficients for ~~ET and GPP have improved for every site and the GPP bias has diminished for all calibration sites – the two validation sites where GPP bias has increased are Poker Flat (US-Prr) and Zotino (RU-Zot)~~the ET in Fig. 2 have also been improved but the model bias has mostly increased.

Table 5. Parameter default and optimised values for the calibration period with corresponding cost function value. The values written in boldface were the most effective and the italic values the least effective for the given experiment. Also presented are the fixed parameter values for the drought period optimisation, with opt referring to the use of the corresponding optimised value from this table.

Parameter	def
$V_{C,max}$	62.5
α	0.28
τ	10.0
c_b	5.0
f_{C3}	0.87
q	0.0
g_0	1.0E-3
g_1	Table 3
a	2.8
d	80
θ_{dr}	0.9
θ_{hum}	0.5
θ_{pwp}	0.35
θ_{tsp}	0.75
p_{int}	0.25
s_m	5.9E-3
w_{skin}	2.0E-4
C_{decay}	13.0
S_{min}	10.0
S_{range}	150
T_{alt}	4.0
T_{ps}	10.0
~~g_0 1.0E-3 - - 4.7E-3 4.7E-3 4.4E-3 4.2E-3 Table 7 g_1 Table 3 - - **9.9** **8.8** 10.9 1.6 Table 7 a 2.8 - - - - 3.2 - opt b 80 - - - - **71** - opt~~ height	cf_1

3.3 Drought event

The resulting parameter values, from the optimisation during the drought conditions in Hyytiälä (FI-Hyy) in the summer of 2006, are presented in Table 7. Setting the maximum carboxylation rate to a constant value ($V_{C,max} = 52.0$) enabled the full use of the dynamical range of q – the idea was to ensure that $V_{C,max}$ does not dominate the optimisation, any value for q is possible and it is able to influence the outcome. The LoGro phenology parameters and τ were fixed to their optimised values, presented in Table 5, as they should not be affected by the drought. Likewise, the values of other parameters (not presented in Table 7) were set as compromises between the stomatal conductance formulations.

Table 6. Slope of the regression line (b) and the coefficient of determination (r^2) for the different stomatal conductance formulations during the validation period with the optimised parameters. We have written the best values of b and r^2 in boldface for each site, and italicised the abbreviations of the separate validation sites.

	Evapotranspiration (ET)											
	b						r^2					
Site	Base	Bethy	BB	Leu	F&K	USO	Base	Bethy	BB	Leu	F&K	USO
CA-Obs	0.91	0.9	**0.91**	0.86	0.81	0.76	0.75	**0.77**	0.76	0.76	0.75	0.74
CA-Qfo	0.96	0.98	**0.99**	0.92	0.89	0.83	0.71	**0.72**	0.7	0.71	0.7	0.69
FI-Hyy	0.97	1.05	1.07	0.95	**0.98**	0.79	0.73	**0.77**	0.77	0.75	0.77	0.69
FI-Ken	0.54	**0.64**	0.62	0.56	0.58	0.48	0.48	0.51	**0.52**	0.49	0.51	0.45
FI-Sod	0.64	0.73	**0.74**	0.63	0.64	0.56	0.58	**0.64**	0.61	0.6	0.62	0.55
RU-Fyo	0.98	1.02	1.01	0.98	**0.99**	0.85	0.7	0.71	0.71	0.71	**0.71**	0.7
CA-Ojp	0.8	**0.84**	0.84	0.75	0.72	0.67	0.64	**0.65**	0.64	0.65	0.64	0.63
FI-Let	1.09	0.98	1.08	1.04	**1.01**	0.94	0.49	0.47	0.49	0.5	**0.51**	0.48
RU-Zot	0.49	0.56	**0.56**	0.47	0.46	0.41	0.45	**0.52**	0.5	0.47	0.48	0.41
US-Prr	0.38	0.37	**0.42**	0.35	0.33	0.35	0.48	0.53	**0.53**	0.46	0.44	0.43
best values	0	2	5	0	3	0	0	6	2	0	2	0
	Gross primary production (GPP)											
	b						r^2					
Site	Base	Bethy	BB	Leu	F&K	USO	Base	Bethy	BB	Leu	F&K	USO
CA-Obs	**0.83**	0.77	0.82	0.81	0.81	0.77	0.87	0.9	0.89	0.89	**0.91**	0.9
CA-Qfo	0.97	0.95	**0.98**	0.96	0.96	0.9	0.84	0.87	0.85	0.86	**0.88**	0.87
FI-Hyy	1.02	**1.01**	1.05	1.03	1.06	0.98	0.94	0.94	0.94	0.95	**0.95**	0.95
FI-Ken	0.9	0.97	**0.97**	0.93	0.95	0.9	0.93	0.9	0.9	0.93	0.93	**0.94**
FI-Sod	0.66	0.71	**0.71**	0.67	0.69	0.65	0.88	0.87	0.86	0.89	**0.9**	0.9
RU-Fyo	0.95	0.88	0.91	0.96	**0.98**	0.91	0.89	0.88	0.88	0.91	0.91	**0.91**
CA-Ojp	0.72	0.74	**0.75**	0.7	0.69	0.66	0.83	0.85	0.84	0.85	0.86	**0.86**
FI-Let	1.27	**0.99**	1.09	1.25	1.26	1.21	0.93	0.88	0.89	0.94	0.94	**0.94**
RU-Zot	0.42	**0.44**	0.44	0.42	0.42	0.4	0.86	0.85	0.84	0.88	**0.88**	0.88
US-Prr	0.2	**0.21**	0.21	0.2	0.19	0.19	0.62	0.6	0.6	0.62	**0.63**	0.62
best values	1	4	4	0	1	0	0	0	0	0	6	4

We can now compare the parameter values in Table 7 to those in Table 5. The values of the relative humidity parameter (θ_{hum}) and the residual stomatal conductance (g_0) have remained nearly unchanged, but for the rest of the ~~parameter we see wildly different results. Noticeably~~parameters have quite varied values. The leaf internal-to-external CO_2 concentration

Table 7. Optimised parameter and corresponding cost function values with different stomatal conductance formulations for the extended dry period.

Parameter	def	Base	Bethy	BB	Leu	F&K	USO
f_{C3}	0.87	0.7	0.7	-	-	-	-
q	0.0	0.09	0.0	0.15	0.57	0.16	0.30
θ_{tsp}	0.75	0.57	0.46	0.48	0.44	0.45	0.41
θ_{pwp}	0.35	0.40	0.38	0.27	0.23	0.28	0.16
θ_{hum}	0.5	0.2	0.2	0.2	0.2	0.2	0.2
g_0	Table 3	-	-	4.9E-3	5.0E-3	3.8E-3	4.6E-3
g_1	Table 3	-	-	7.5	6.0	7.0	1.5
cf_2		0.42	0.44	0.39	0.41	0.41	0.41

(f_{C3}) as well as the slope of the stomatal conductance (g_1) are at the lower bound (expect g_1 for BB). Noticeably, the USO optimisation only changes the ~~value~~ values of θ_{tsp} and q, and leaves the rest of the parameters almost untouched.

The changes these different ~~parametrisations~~ parameterisations have on the model output are visualised in Fig. 3. ~~The Baseline, Bethy and USO formulations demonstrate a considerable increase in the agreement in GPP between the model and observations when compared to the default setting or the previous more general optimisation. The GPP of other formulations has remained roughly the same as with the more generally optimised parameter values. The~~ All of the stomatal conductance models, with default parameterisation, suffer from too low ET values before (and during) the actual drought. This behaviour was corrected during the general optimisation, but has partially re-emerged with the dry period parameters for the Baseline, Ball-Berry, Leuning, and to a lesser degree the Friend and Kiang formulations~~, now suffer from the too low ET values before~~. Most of the models also exhibit too high ET values during the actual drought ~~. The Bethy model has~~ with the generally optimised parameter values. This behaviour was also corrected with the dry period optimisation, but the Baseline and especially the Bethy model now suffer from a too strong drawdown of ~~both ETand GPP during the drought~~ET. These models also demonstrate the too strong drawdown for the GPP. The GPP itself was greatly improved with both optimisations and for all models. The dry period optimisation of the USO model also managed to correct the erroneous GPP of the general optimisation during the actual drought, where as the GPP of other formulations has remained roughly the same as with the general optimisation. The USO formulation results in the best fits for r^2 and b with the dry period optimisation~~and it makes full use of the dynamical range of~~.

The Bethy and the USO models demonstrate the most variability in the β-function ~~. Overall,~~ values in Fig. 3 (rightmost panels), for the dry period optimisation~~was succesful for the USO model and to a lesser for the Bethy formulation as well – the results for the other variants are mixed and inconclusive.~~

~~We selected two of the~~. We selected these two stomatal conductance formulations ~~, Bethy and USO,~~ to examine the changes to the water use efficiency (WUE) of plants during the extended dry period. The highlighted observations in Fig. 4 (rightmost panels) show a clear path of development for the drought where the observations imitate the letter δ. The colourings follow

the β-function values in Fig. 3 between the red vertical lines. Both observational colourings (same as the model colouring) are similar and depict~~first~~ , initially, a linear decrease in both ET and GPP, followed by a rapid decline in ET and a delayed decline in GPP. The recovery of plants from the drought can also be seen as the colouring starts to turn lighter. The models depict a more linear response of GPP to ET as the drought develops, although with USO we can see a bit more similarities in the pattern of the values.

~~Lastly, we inspected~~ Finally, we used both optimised parameter sets (Table 5 and 7) to produce the ET and GPP cycles for all sites and stomatal conductance models. This analysis (not shown) ~~for the whole validation period with both optimised parameter sets, all stomatal conductance formulations and all calibration sites~~verified that in general conditions the Table 5 parameter values produced better estimates in general. The b and r^2 values for ~~ET were~~ the ET were systematically better for all stomatal conductance formulations (except one)~~using the more generally optimised parameter set. There is~~ . There was some variation in the indicators for the GPP, where approximately a third of the values (of mostly r^2) are better with the dry period parameter set. These differences are mostly attributed to increased model bias (decreased b) that is explained by the lower values of g_1. Overall, the more general optimisation provided systematically better or comparable results to the dry period optimisation. The exception is the USO formulation, which had an approximately 1:1 distribution of best values for both variables in-between the parameter sets.

4 Discussion

We will first discuss the validity of our approach and the simulation setup, followed by ~~examinations on~~ examination of the success of the modifications made to the model, and close with some further remarks on the parameter values.

4.1 Validity of the simulations

Before we calibrated the model, we fixed the limiting value for LAI and adjusted the site-specific vegetative area fractions to reproduce the measured site level maximum of LAI. In the simulations, we focused on boreal coniferous forests, where light penetration is deep and the light conditions are homogenous – consequently we could assume a homogenous leaf distribution. Furthermore, the JSBACH model takes into account leaf clumping and we can assume the leaf orientation and shape to be similar throughout the study sites. Therefore, we argue that reproducing the site level maximum of LAI is appropriate approach in this study. Together with parameter calibration it has resulted in improved ET and GPP fluxes as can be verified from the b and r^2 values in Fig. 2. The improvements in b and r^2 are mostly seen in the GPP flux, which can be explained by the fact that the stomatal conductance in JSBACH is primarily resolved for carbon assimilation, and the same conductance is then used for transpiration (A8). Additionally, GPP is derived from the EC measurements by flux partitioning – this tends to remove some of the flux instabilities (that are still present in the ET).

We encountered difficulties in ~~replicating~~ reproducing the fluxes for the validation sites with low LAI (i.e RU-Zot and US-Prr). This can be a consequence of the area scaling as the adjustment linearly changes the proportions between vegetative area and bare soil. Another reason is the lack of the site understory in these simulations. For example, approximately half of

the CO_2 fluxes (and consequently roughly half of the GPP) for Poker Flat are produced by the site understory (**?**). Additionally, there are also many parameters describing site-specific soil properties (such as porosity) that were not part of the optimisation and may be inaccurate. These effects may also be pronounced due to the changes in parameters affecting soil moisture as well as the area scaling.

There were no clear differences between sites dominated by pine or spruce. Neither did we notice any particular effect on the bias, NMSE or correlation coefficient that could be explained by geographical location, stand age or annual precipitation or temperature. We optimised the model for individual (calibration) sites as well (not shown). Mostly this changed the values of parameters (such as $V_{C,max}$ and g_1) affecting the amplitude of the modelled fluxes. These parameters can be viewed to be more site-specific, a characteristics that is reduced in a multi-site calibration – the possibility of highly site-specific properties (and parameter values) can also explain the difficulties in reproducing the validation site observations. We are omitting these results as single-site optimisation can be viewed as overfitting the model and the results do not provide any additional insights.

The APIS performance tests (Gelman-Rubin and δ) indicate that the algorithm is performing well at 20 iterations but the convergence of the global estimate for some parameters is slow. This is mostly a direct result of the normalisation of the cost function that inflates the target distribution, which reduces the parameter sensitivity to observations and gives too much weight to the initial locations and draws. Without the normalisation, the algorithm would also converge faster. Additionally, APIS is meant to examine the full target distribution with only some sequantiality – 20 iterations (or less) should be sufficient for APIS to locate the modes of the target. In longer APIS simulations, the global estimate would likely benefit from ~~e.g.~~ discarding the first half of the samples but this would require the estimate to be recalculated at each iteration (from the drawn samples) as it could not be calculated iteratively.

4.2 Delayed effect of temperature

We modified the JSBACH model by introducing the delayed effect of temperature for photosynthesis to restrain the respiration and photosynthesis of conifers in spring. The effect of this (delayed increase in GPP) is apparent in the annual GPP cycles of CA-Qfo, FI-Hyy, FI-Ken, FI-Sod and RU-Zot in Fig. 2. The delay is in place for the other sites as well, but the effect is less apparent in the figure. This delay is to a lesser extent also reflected in transpiration, and consequently in ET~~. This~~, as can be seen e.g. at FI-Hyy and FI-Sod – for other sites this effect is not clear. The correction in the ET values can lead to an increase in model bias as is the case with Sodankylä (FI-Sod), where the too low autumn values in the default model were previously compensated by too high springtime values ~~with the default parametrisation. This~~ (in the sense of annual ET). Correcting the springtime behaviour leads to an increase in bias, but this should not be viewed as a fault in the optimisation as the model was previously mitigating an erroneous behaviour (too low autumn ET) with another (too high springtime ET).

? used a linear dependency of photosynthetic efficiency to the state of acclimation, and reported 13.75 days to be the best fit for the adjustment period length (τ). **?** utilised a sigmoidal relation and reported the value of 8 days, but noted that the range of values resulting in a good fit was large (5–10.4 days). **?** came to a similar conclusion when they encounter a near-flat distribution for τ in the range of 1–12 days. In our simulations τ exhibits larger optimal values (nearly 15 days), which is most

likely due to the model adapting to the multi-site calibration (as sites have different characteristics, a longer acclimation period accounts better for these variations).

4.3 Stomatal conductance models

We examined the model behaviour with six stomatal conductance formulations ~~. The~~ and the resulting b and r^2 values are presented in Table 6~~indicate that the~~ . The best performance (bolded values) in simulated ET is achieved by the BB model for bias and the Bethy formulation for r^2. These two models also share the best performance in the GPP bias, whereas the best r^2 values for the GPP are demonstrated by the $F\&K$ model, followed by the USO formulation. Calculating the number of best values demonstrated by each model, we obtain that the best performance is shared by the Bethy (12) and F&K (12) formulations, followed by the BB (11) model. ~~The Bethy model dominates the r^2 values of ET, where as the BB model has the highest "score" for the b values of GPP and the F&K formulation leads in the r^2 of GPP. However,~~ However, we note that some of the ~~differences in the indicator valuesare small and we calibrated~~ "best values" are only marginally better than comparable values. Additionally, we used two more parameters (a and d) for the F&K formulation ~~(when compared to~~ than for the other Ball-Berry ~~models)~~formulations. Likewise, we could have~~for example~~ , for example, included the factor D_0 ~~(that depicts stomatal sensitivity to changes in vapour pressure deficit D_s)~~ (B3) in the optimisation, which would have likely improved the performance of the Leuning model. Similarly to the results by **?**, based on this (general) calibration, there is no clear single candidate for the best stomatal conductance formulation.

The model behaviour was also examined during the Hyytiälä drought of 2006. Some of the parameter values were kept fixed during these simulations, most of the fixed parameters should not affect the drought period calibration but there are exceptions, such as the maximum carboxylation rate $V_{C,max}$. It can be argued that e.g. both the parameters $V_{C,max}$ and g_1 should decrease (**?**) during the drought but we decided to fix $V_{C,max}$ to get a better response for q. The best fit to the observations was achieved by the USO formulation, as seen in Fig. **??**, with remarkably similar parameter values to the general optimisation. The USO model was also able to (somewhat) replicate the "δ" shape of the drought in Fig. 4.

The stomatal conductance function ($g_s = g_0 + c\beta g_1$) incorporates also the soil water parameters θ_{tsp} and θ_{pwp} in the form of the β-function as portrayed in Eq. (2). The changes in the values of these parameters (mostly g_1, θ_{tsp} and θ_{pwp}) are intertwined. During the drought, the decrease in the optimised values of g_1 is expected as the plants close their stomata to minimise the loss of water by transpiration (**??**). The same effect is also achieved by increasing the values of θ_{tsp} and θ_{pwp} as this decreases the values of the β-function. The higher values of g_1 during the more general optimisation are better reflected by **?**, whereas the lower values during the drought are more in accordance with physiological observations by **?**. Likewise, **?** found higher values for g_1 (both boreal area and gymnosperm trees) using the USO model.

~~The~~ In general, the site level estimates of (g_0 and) g_1 are sensitive not only to the stomatal conductance formulation but also e.g. to the ~~general~~ structure of the underlying model and the value of other parameters, such as maximum carboxylation rate ($V_{C,max}$). **?** reported $g_1 = 3.78$ (~~control~~in Table 1, Control), using a Leuning model similar to ours, where $(1 + D_S/D_0)$ is replaced by D_S. **?** approximated ~~$g_1^{BB} \approx 5$~~ g_1^{BB} to be 5 for Sodankylä while estimating the variation in the values of $V_{C,max}$ and maximum rate of electron transport J_{max}. We would suggest that the limiting values θ_{pwp} and θ_{tsp} should be optimised

or fixed before introducing additional tuning factors such as mesophyll conductance or scaling the β in multiple ways in the stomatal conductance formulations (**?**). Our simulation setup for q corresponds to the configuration 5 (C5) by **?**, with variables $q = q_B$ and fixed value $q_S = 1$.

4.4 Parameter values

Some of the parameters in this study have been calibrated before by e.g. **??**. Our approach differs from these as we required the model to reproduce the site level maximum of LAI. In contrast e.g. **?** found the structural limit for (all-sided) LAI to be 4.2, which is considerably lower than the measured LAI for many of the sites in Table 1. Our approach directly scales the vegetative area, so it also scales GPP and ~~e.g.~~ also the amount of rain available for plants (as rain is directed to bare soil and vegetative area). This means that the parameter values should not be directly compared without taking the different paradigms into account. However, our optimised $V_{C,max}$ values are in-between 62.5 reported by **?** and 29.3 by **?**, and are in line with the yearly cycle presented by **?**.

The exponential scaling factor q in Eq. (A1) of the β-function (2), was revealed to be ineffective in our optimisation as indicated in Table 5. In our simulations, this situation arises as the effective range of the β-function has been ~~reduced by lowering~~ lowered by reducing θ_{pwp} and θ_{tsp}. The actual soil moisture is rarely below ~~this fraction~~ the fraction θ_{tsp}, so q is constrained with a very limited number of datapoints~~. Therefore~~, and thus has only minimal effect on the fluxes and the cost function. Therefore, the values presented for q in Table 5 can be unreliable and even unrealistic. This situation is remedied in the drought period optimisation ~~but the~~ when the soil moisture is low. The resulting values for q in Table 7 have a wide ~~range~~dispersion, although they are mostly on the lower end. This signifies that the additional GPP reduction is mostly gradual, with a steep decrese near the permanent wilting point θ_{pwp}.

The values of soil water parameters are closely grouped in the optimisations except for the values of θ_{pwp} during the drought. This can occur due to a larger impact, of the different stomatal conductance formulations ~~to~~ on the accumulating soil water content, than assumed – this can also be seen from the differences in the β-function values in Fig. 3. Furthermore, the values of θ_{tsp} and θ_{pwp} have been considerably lowered from their default values in both optimisations. This change can be perceived in at least two different ways. Either the boreal forests are not generally limited by soil moisture stress (except in the case of extreme drought) or the water retention capabilities of the soil (in the model) have been systematically overestimated. The latter seems unlikely, in the light of results by e.g. **?**.

5 Conclusions

The adaptive population importance sampler (APIS) is a recent method, capable of estimating complicated multidimensional probability distributions using a population of different proposal densities. The algorithm was able to produce reasonably stable estimates for most parameters quickly. Prior to calibrating the model, we adjusted the site-specific vegetative area fractions to reproduce the measured site level maximum of LAI. This practical approach resulted in improved ET and GPP fluxes, although we encountered difficulties in replicating these for sites with low LAI. The model parameters were optimised simultaneously

for all sites without any additional site level tuning. The parameters ~~affecting~~ that were most effective in the optimisation processes~~the most~~ , were consistent for all stomatal conductance formulations.

The introduction of the S-function, to delay the start of the vegetation active season, has corrected the springtime increase in GPP for conifers throughout the sites used in this study. The parameters θ_{tsp} and θ_{pwp}, that set the range for the soil moisture stress function β, were both systematically lowered and optimised to nearly identical values for all stomatal conductance models. ~~This~~ The low effective range for the β function rendered the experimental parameter q nearly ineffective in the more general optimisation. The dry period optimisation increased the effective range of the β-function and the importance of q~~for the Ball-Berry type model. The Baseline and Bethy versions optimised q to be practically ineffective~~, which resulted in highly nonlinear (additional) reduction in the net assimilation rate. Overall, ~~both optimisations strongly~~ this fact and both optimisations indicate that boreal forest transpiration is not limited by soil moisture stress under normal conditions.

The optimisation improved the predictive skill of the model with all stomatal conductance formulations as was seen during the validation period. The Bethy, Ball-Berry and Friend and Kiang versions were the most ~~compliant~~ in agreement with the observations, although the differences between these and the other formulations were small. Most of the model versions had some problems during the extended dry period ~~. The unified stomatal optimisation model had~~ and the best b and r^2 values ~~during the drought~~were achieved by the unified stomatal optimisation model. Additionally, the optimised parameter values of the USO model for the dry period were the most ~~coherent~~ alike (of all stomatal conductance formulations) with those of the more general optimisation.

Code and data availability. The data required to calibrate and validate the model is originally part of the FLUXNET2015 dataset that can be accessed through the FLUXNET database (doi:10.17616/R36K9X). Our modified dataset, containing the forcing data and the observations used in this article, is available through Zenodo portal (doi:10.5281/zenodo.3240954). The data depicting the simulations (parameter draws, cost function values etc.) has been added as a supplement. The JSBACH model (branch: cosmos-landveg-tk-topmodel-peat, revision: 7384) can be obtained from the Max Planck Institute for Meteorology, where it is available for scientific community under the MPI-M Sofware License Agreement (http://www.mpimet.mpg.de/en/science/models/license/). The modifications to the model, described in this paper, have been uploaded to Github and they can be accessed by contacting the authors at jarmo.makela@fmi.fi (after access to the actual model has been approved). For any questions, we encourage you to contact the authors at jarmo.makela@fmi.fi.

Appendix A: Parametric equations within JSBACH

In this appendix we present the most relevant equations that are governed by the parameters in Table 2. The appendix is divided into sections that coincide with the parameter groups.

A1 Photosynthesis

The Farquhar model (**?**) is based on the observation that the assimilation rate in the chloroplast is limited either by the carboxylation rate (V_C), induced by the Rubisco enzyme, or the light-limited assimilation rate (J_E). The total rate of carbon fixation

is reduced by the amount of dark respiration (R_d), resulting in net assimilation rate (A_n). The experimental scaling factor β^q (**?**) is based on soil moisture stress in Eq. (2), that takes effect ($\beta < 1$) when soil moisture is significantly reduced. This scaling is used by all stomatal conductance formulations. We have also introduced here in equation form the actual reduction to photosynthesis by γ from the delay in the start of the vegetation active season in Eq. (1).

$$A_n = \beta^q(\min(\gamma V_C, J_E) - \gamma R_d) \tag{A1}$$

Oxygenation of the Rubisco molecule reduces the carboxylation rate, which is given as:

$$V_C = V_{C,max}\frac{C_i - \Gamma_\star}{C_i + K_C(1 + O_i/K_O)} \tag{A2}$$

Here C_i and O_i are the leaf internal CO_2 and O_2 concentrations, $\Gamma_\star$ is the photorespiratory CO_2 compensation point, K_C and K_O are Michaelis-Menten constants ~~parametrizing~~ parameterising the dependence on CO_2 and O_2 concentrations. Furthermore, leaf internal CO_2 concentration depends on the external (ambient) concentration C_a (in the Baseline and Bethy formulations and unstressed conditions) by:

$$C_i = f_{C3}C_a \tag{A3}$$

Likewise, the light-limited assimilation rate can be expressed as a function on electron transport rate (J), which is a function of radiation intensity (I) in the photosynthetically active band, the maximum electron transport rate (J_{max}) and the quantum efficiency for photon capture (α):

$$J_E = J(I)\frac{C_i - \Gamma_\star}{4(C_i + 2\Gamma_\star)}, \qquad J(I) = J_{max}\frac{\alpha I}{\sqrt{J_{max}^2 + \alpha^2 I^2}} \tag{A4}$$

A2 Soil water

In JSBACH the soil water budget is based on several reservoirs (e.g. skin, soil, bare soil, rain intercepted by canopy etc.) and the different formulations are plentiful. We present here only the most crucial of these. Changes in volumetric soil water (θ_s, not to be confused with ~~volumetric~~ relative soil water content $\theta = \frac{\theta_s}{\theta_{fc}}$) due to rainfall ($R$), evapotranspiration ($ET$), snow melt ($M$), surface runoff ($R_s$) and drainage ($D$) are calculated with a geographically varying maximum field capacity (θ_{fc}) ~~,~~ and soil water density (ρ_w).

$$\rho_w\frac{\partial\theta_s}{\partial t} = (1 - p_{int})R + ET + M - R_s - D \tag{A5}$$

The interception parameter (p_{int}) also affects the amount of water intercepted by vegetation and bare soil which further affects evaporation and transpiration. The skin reservoir is limited by w_{skin} and excess water is transferred to soil water. Likewise when the soil water content (θ) is greater than parameter θ_{dr}, the excess water is rapidly drained (in addition to the limited drainage below this threshold), where d, d_{min} and d_{max} are constant parameters:

$$D = d_{min}\theta + (d_{max} - d_{min})\left(\frac{\theta - \theta_{dr}}{1 - \theta_{dr}}\right)^d, \qquad \theta \geq \theta_{dr} \tag{A6}$$

Evaporation from wet surfaces (E_{ws}) depends on air density (ρ), specific humidity (q_a), saturation specific humidity (q_s) at surface temperature (T_s) and pressure (p_s) and aerodynamic resistance (R_a). The aerodynamic resistance depends on heat transfer coefficient (C_h) and horizontal velocity (v_h).

$$E_{ws} = \rho \frac{q_a - q_s(T_s, p_s)}{R_a}, \qquad R_a = C_h |v_h|^{-1} \tag{A7}$$

Transpiration from vegetation (T_v) is likewise formulated but additionally depends on the stomatal resistance of the canopy (R_c), which is an inverse of the stomatal conductance and as such, depends on which conductance model is used.

$$T_v = \rho \frac{q_a - q_s(T_s, p_s)}{R_a + R_c} \tag{A8}$$

Evaporation from dry bare soil (E_s) also has an added dependence on surface relative humidity (h_s) calculated from soil dryness:

$$E_s = \rho \frac{q_a - h_s q_s(T_s, p_s)}{R_a}, \qquad h_s = \max\left[\theta_{hum}(1 - \cos(\pi\theta)), \min\left(1, \frac{q_a}{q_s(T_s, p_s)}\right)\right] \tag{A9}$$

The total evapotranspiration is a weighted average of E_{ws}, T_v and E_s, where the weights are based on fill levels of reservoirs and the vegetative fraction of the grid cell.

A3 Logistic Growth Phenology (LoGro-P) model

The parameters from the LoGro-P are mainly used to determine the spring and autumn events for JSBACH. To determine the date of the spring event we first introduce a few additional variables, namely the heatsum $S_T(d)$, the number of chill days $C(d)$ and the critical heatsum $S_{crit}(d)$. $T(d)$ denotes the mean temperature at day d.

$$S_T(d) = \sum_{d'=d_0}^{d} \max(T(d') - T_{alt}, 0) \tag{A10}$$

Heatsum $S_T(d)$ cumulates the amount of "heat" above the parameter T_{alt} after the previous growing season. The actual starting date d_0 of the summation need not be known since it is enough to start the summation "reasonably late" after the last growth season.

$$C(d) = \sum_{d'=d_a}^{d} H\left(T_{alt} - T(d)\right) \tag{A11}$$

The number of chill days is calculated as the number of days when the mean temperature is below T_{alt}. Here $H()$ denotes the Heaviside step function and the summation starts at the day (d_a) of the last autumn event.

$$S_{crit}(d) = S_{min} + S_{range} e^{-C(d)/C_{decay}} \tag{A12}$$

The critical heatsum (S_{crit}) decreases as the number of chill days $C(d)$ increases, with an exponential memory loss parameter C_{decay}. The spring event happens when:

$$S_T(d) \geq S_{crit}(d) \tag{A13}$$

The autumn event requires the definition of one more variable, the (pseudo) soil temperature ($T_s(t)$), which at time t is calculated as an average air temperature (T) with an exponential memory loss (T_{ps}). The autumn event occurs when T_s falls below a certain threshold. In the equation N is the normalization constant and τ is the length of a time step.

$$T_s(t) = \frac{1}{N} \sum_{n=-\infty}^{t} T(n) e^{-(t-n)\frac{\tau}{T_{ps}}} \tag{A14}$$

Appendix B: Stomatal conductance formulations

In this appendix we present the stomatal conductance model formulations used in this study. In the original JSBACH formulation, the Baseline model (**?**), the photosynthetic rate is resolved in two steps. First the stomatal conductance under conditions with no water stress is assumed to be controlled by photosynthetic activity (**?**). Here the leaf internal CO_2 concentration is assumed to be a constant fraction ($C_{i,pot} = f_{C3} C_a$) of ambient CO_2 concentration (C_a). This allows for an explicit resolution of the photosynthesis (**?**). Then the impact of soil water availability is accounted for by a soil moisture-dependent multiplier (β) that is identical for each canopy layer (**?**).

$$g_{s,pot} = \frac{1.6 A_{n,pot}}{C_a - C_{i,pot}} \qquad \Rightarrow \qquad g_s = \beta g_{s,pot} \tag{B1}$$

After accounting for soil water stress, the net assimilation rate (A_n) and intercellular CO_2 concentration are (C_i) are recalculated using g_s, and integrated over the leaf area index to produce canopy level estimates.

In the Bethy approach (**?**), the unstressed canopy conductance ($G_{c,pot}$) is calculated similarly to the Baseline model, but potentially further limited by the water supply function of the maximum transpiration rate ($T_{supply} = \beta T_{max}$). T_{max} is a fixed and predefined upper limit for transpiration as in **?**.

$$G_c = \begin{cases} G_{c,pot} \frac{T_{supply}}{T_{pot}}, & T_{pot} \geq T_{supply} \geq 0 \\ G_{c,pot}, & T_{pot} < T_{supply} \end{cases}, \qquad T_{pot} = \rho \frac{q_s - q_a}{1/G_a + 1/G_{c,pot}} \tag{B2}$$

The potential (unstressed) transpiration rate (T_{pot}) is a function of air density (ρ), saturation specific humidity (q_s) at given temperature and pressure, specific humidity (q_a), aerodynamic conductance (G_a) and unstressed canopy conductance ($G_{c,pot}$). After this scaling, the net assimilation rate and intercellular CO_2 concentration are recalculated as in the Baseline model.

The Ball-Berry variants relate the stomatal conductance (g_s) to empirically fitted parameters g_0 (mol $m^{-2}s^{-1}$) and g_1 (unitless, except for g_1^{USO} which has units of $\sqrt{\text{kPa}}$) that respectively represent the residual stomatal conductance and the slope of the function. The stomatal conductance is a function of the net assimilation rate (A_n), the water stress factor (β) and the atmospheric CO_2 concentration (C_a). The original Ball-Berry formulation (**?**) also depends on relative humidity at leaf surface (h_s). In the Leuning model (**?**), the CO_2 concentration is reduced by the CO_2 compensation point (Γ) as well as scaled by the vapour pressure deficit (D_s) and a constant (D_0) depicting the stomatal sensitivity to changes in D_s. The Friend and Kiang model (**?**) adds an exponential dependency on the difference of specific (q~~q~~$_a$) and saturation specific humidity (q_{sat}) with empirically fitted constants $a = 2.8$ and $b = 80$. The unified stomatal optimisation model (**?**) also adds a dependency to

the vapour pressure deficit (D_s).

$$g_s^{BB} = g_0^{BB} + g_1^{BB}\beta\frac{A_n h_s}{C_a} \qquad g_s^{Leu} = g_0^{Leu} + g_1^{Leu}\beta\frac{A_n}{(C_a-\Gamma)(1+D_s/D_0)} \tag{B3}$$

$$g_s^{F\&K} = g_0^{F\&K} + g_1^{F\&K}\beta\frac{A_n a^{b(q_s-q_a)}}{C_a}\frac{A_n a^{-d(q_{sat}-q_a)}}{C_a} \qquad g_s^{USO} = g_0^{USO} + 1.6\left(1+\frac{g_1^{USO}\beta}{\sqrt{D_s}}\right)\frac{A_n}{C_a}$$

Author contributions. The experiments were planned by J. Mäkelä, T. Aalto, T. Markkanen and T. Thum. J. Mäkelä ran the simulations and prepared the manuscript with contributions from co-authors. J. Knauer originally implemented the Ball-Berry type stomatal conductance formulations into JSBACH under S. Zaehles supervision. J. Susiluoto maintained the framework for testing the algorithm. M. Aurela, A. Black, M. Heimann, A. Lohila, I. Mammarella, H. Margolis and H. Kobayashi provided the site level observations required in this study. T. Aalto, T. Markkanen and T. Viskari extensively commented and revised the document.

Competing interests. The authors declare that they have no conflicts of interest.

Acknowledgements. This work has been supported by Jenny and Antti Wihuri Foundation, the NordForsk Nordic Centre of Excellence under Grant no. 57001 (eSTICC) and the Academy of Finland under Grant no. 295874 (OPTICA), as well as Academy of Finland Centre of Excellence under Grant no. 307331 and ICOS-Finland (project No. 281255) and EU-Life+ project MONIMET (LIFE12 ENV/FI000409). This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices.