Response to Reviewer #1:

The authors would like to thank Reviewer #1 for taking the time to write such helpful, thorough and constructive comments. The comments have been taken into consideration in the revised manuscript. We answer them individually as follows:

1 General comments:

Sometimes they refer with adJULES to the adjoint of JULES and sometimes to the whole optimisation system. The two are certainly very different and as such should also be clearly distinguished in the manuscript.

This has now been clarified, adJULES is used to refer to the whole optimisation scheme. The bracket containing 'called adJULES' was removed from page 4, line 6, which we believe was the source of this confusion.

There is an established terminology in the data assimilation community and it would improve the readability if the authors would use this terminology, e.g. 'posterior' instead of 'new' parameter.

This has been addressed, the text has been changed in places where established terminology would improve readability, for example in section 3.2.1 'Assessment of PFT-specific optimal parameters'.

The authors claim that any residual differences between the observations and model output using the optimised parameter vector are due to structural errors in the model and not to the parameter values. This may be true if they have really identified the best possible fit, i.e. if they have found the global cost function minimum. Since with such complex models the cost function usually has a multimodal structure it is not clear that a gradient-based optimisation approach finds the global minimum. The authors need to comment on that in the manuscript. In fact, the manuscript would benefit from including some posterior diagnostics, such as the final cost function and gradient values. It is not clear if they've always found a minimum, and if so if that is the global minimum.

It is true that the limitations of a gradient-based optimisation approach is missing from the manuscript. The following text has been added to the conclusion to address this omission and to reduce the emphasis placed on model structure errors

" A limitation of gradient descent methods, such as the optimisation scheme used in this study, is the fact that sometimes a local minimum in found instead of the global minimum. However, as discussed in section [..] Kuppel et al (2014)'s hypothesis that the cost function becomes smoother with additional sites may be a solution in avoiding local minima. Alternative methods, including ensemble methods, could avoid this issue, but are more computationally costly."

The study also lacks some independent validation. The authors only calculate the improvement in RMSE for the same data streams they also assimilate. A careful validation against independent data is especially important because by calibrating the model parameters against a specific data set the model's performance may be deteriorated compared to other independent data.

Given the small number of sites available to us, we decided to use all available sites in finding the multisite parameters sets. Sites used in this study required at least two consecutive years, one to spin-up the model and one to calibrate against. Re-examining the data, we found that the majority of

the sites had more than two years and so a different year could be used to validate the optimised set of parameters. The year used for validation was chosen to be the second most complete, the first most complete having been used in the calibration. The results of the validation, which are very positive, are now shown in the results section alongside the results for the calibration.

2 Specific comments:

P3 L3: The term 'adJULES' should be defined before using it.

Corrected: now defined on P2 L31.

P4 Eq 1: The cost function is missing the factor 1/2. The omission of this factor in the calculations leads to a wrong estimation of the posterior uncertainties.

Factor added.

P4 L17: What do you mean by 'observed covariance in the error (m-o)'? How can you observe this?

Removed "observed" from the sentence.

P4 L19: How does lambda enter Eq 1?

An explanation has been added.

P4 L28/29: This sentence needs to be reformulated. It is not clear how reverse and forward mode relate to the adjoint. The adjoint calculates the derivative in reverse mode.

Removed part of the sentence: "in 'reverse mode' (rather than 'forward mode') for computational efficiency", and the following text was added: *"Automatic differentiation relies on using the chain rule, the choice of forward or reverse mode refers to the order in which the derivatives are computed."*

P5 Fig 1: Essentially the figure is incomprehensive and does not show an interative loop.

There are two iterative loops in our system, one found within the minimisation scheme itself (BFGS) and one created by re-feeding zin the system. This second loop is need since the covariance matrix $\mathbf{R}$ is dependent on $\vec{z}$. This fact has now been explained more explicitly in the text. Eq(1) now reads:

$$f\left(\vec{z};\hat{\vec{z}},\vec{z}_0\right) = \frac{1}{2}\left[\sum_t \left(\mathbf{m}_t(\vec{z}) - \mathbf{o}_t\right)^T \mathbf{R}\left(\hat{\vec{z}}\right)^{-1}\left(\mathbf{m}_t(\vec{z}) - \mathbf{o}_t\right) + \lambda(\vec{z}-\vec{z}_0)^T\mathbf{B}^{-1}(\vec{z}-\vec{z}_0\right)\right]. \qquad (1)$$

Here, $\mathbf{R}(\hat{\vec{z}}) = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{m}(\hat{\vec{z}})_t - \mathbf{o}_t)(\mathbf{m}(\hat{\vec{z}})_t - \mathbf{o}_t)^T$ *denotes the error cross product matrix produced by a JULES run with parameter value* $\hat{\vec{z}}$*. In an optimisation,* $\vec{z}$ *and* $\hat{\vec{z}}$ *are updated separately in nested loops, having both been initialised to the default JULES parameter value* $\vec{z}_0$*. In the inner loop,* $\vec{z}$ *is varied to minimise the cost function (termination criterion:* $\nabla f \approx 0$*) for the current value of* $\hat{\vec{z}}$*. In the outer loop,* $\hat{\vec{z}}$ *is reset to the new value of* $\vec{z}$ *from the inner loop (termination criterion: change in* $\hat{\vec{z}}$ *negligible). At the end of an optimisation, therefore, the matrix* $\mathbf{R}$ *conveys information about the error correlation structure in a JULES run with optimal parameter values.*

The figure has also been amended, removing the criterion $\nabla f \approx 0$ from the question box, since it was incorrectly referring to the BFGS terminating condition and not the z terminating condition.

P6 L1: The data selection criteria should be specified exactly. What does 'significant

gaps' mean. There is also the danger of introducing biases by certain data selection criteria. This should be taken into account.

Sites with data gaps of more than 50% during the growing season or missing input variables were excluded from the analysis. This has been clarified in the text replacing "significant gaps" with "data gaps of more than 50%".

P6 L3: Why does one require NEE and LE fluxes to model photosynthesis? Please clarify.

Sentence rephrased: "To constrain photosynthetic parameters, Net Ecosystem Exchange (NEE) and Latent Heat flux (LE), among other fluxes, are helpful."

P6 L5: The eddy covariance technique measures the net exchange flux and not GPP. The net flux is partitioned into GPP and respiration by a model. So essentially, in this study the authors calibrate the JULES model against another model, which is used to obtain GPP from eddy covariance measurements. This needs to be discussed.

Text added: *"GPP data are model-derived estimates, which could introduce an additional uncertainty into the results. This is kept in mind during the analysis."*

P6 L6/7: This procedure may lead to inconsistencies between the actual vegetation at a given site and the vegetation structure and soil type used in the model. This should be discussed in the manuscript.

Sentence added: *"This could lead to inconsistencies between the actual vegetation at a given site and the vegetation structure and soil type used in the model. This is kept in mind during the analysis."*

P6 L8: Please provide a reference for the LAI product. Here again, this may lead to another inconsistency, see point above.

Reference for MODIS data added: *"Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R. and Lotsch, A., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote sensing of environment, 83(1), pp.214-231."*

L31/32: Please rephrase. The adjoint does not find the second derivative.

Rephrased: *"The second derivative of the cost function found by differentiation of the adjoint code..."*

P6 L33: How did you determine the weights? What do you mean by 'low enough'?

Text added to the Experiment setup section, explaining the tuning of lambda for the multisite cases:

"Preliminary experiments showed very narrow uncertainties whilst running the optimisation scheme over multiple sites i.e. the background term was found to dominate the cost function. In previous multisite studies (Kuppel et al., 2012, 2014), the prior range was also used to defined the background covariance matrix B. The range was variously further multiplied by a factor of 40% (Kuppel et al., 2012) and 1/6 (Kuppel et al., 2014). Experiments were run to find a similar factor to use in this study (the constant of proportionality in Eq. 5). In each of the multisite experiments, the lowest value of such that the Hessian is positive definite at the optimal parameter value was used. This allows uncertainties to be generated around each parameter and prevents the gradient descent algorithm from reaching the boundaries of the prescribed prior range."

P7 Ll11-17: This is an interesting way to calculate the posterior parameter uncertainties, but it is not clear why and what exactly you do there. What is the advantage of using

this method over calculating the posterior uncertainties from the inverse of the Hessian directly? When you calculate the full Hessian you also get the full error covariance matrix! Do you get a semi-definite Hessian (see also general comment on obtaining a minimum)?

The adJULES system is run using box constraints on the prior, giving it a (multivariate) top-hat distribution. The methodology was picked due to the fact the posterior PDFs will be truncated multivariate normal distributions due to the prescribed prior bounds given to each of the parameters. Text added on L11 to clarify this:

"... Hessian is used to generate samples from the posterior distribution.This is a truncated multivariate normal distribution because of the box constraints placed on the prior. "

Sect 2.5.2: What is the advantage of the metric you define here over calculating the relative uncertainty reduction with respect to the prior? This also provides and assessment of the quality of the fit and is a common diagnostic in data assimilation. It is also not clear how a complete mismatch looks like.

This metric was chosen because not only does it show the improvement made by the optimised parameter vectors but could also be used to see how different sites performed compared to each other. The metric has been amended slightly to define the fraction of variance unexplained, which is more intuitive. Paragraph added to this effect on line 19:

"This metric was chosen to show not only the improvement made by the optimal parameter vectors at each site but also to show how each site performed relative to others."

P8 L5: This is not a validation, but rather an assessment of the how good the fit against the data is. A real validation would be against independent data and not the data used for assimilation.

The purpose of section 3.2 was to shows that given a set of 5 randomly selected sites, the optimised parameter vector found by optimising over these sites also improves the rest of the sites not used in the calibration. This experiment is now obsolete since we have the ability to validate the PFT-specific parameters properly in our improved result section. As a result, this section has been removed.

P8 L25: Why does JULES not perform very well for C4 grasses. You should elaborate this.

Text added P8 L26 to the effect: *"The original stomatal conductance-photosynthesis model within JULES was developed based on fluxes measured over C4 grass as part of the FIFE field experiment (Cox and Huntingford, 1998). However, there are relatively few Fluxnet sites over C4-dominated landscapes, and only two even in the extended dataset that we use. As a result ,the sensitivity of stomatal conductance and photosynthesis to environmental factors has been less well tested for C4 grasses. The results presented in this paper therefore highlight the need to reassess JULES and other land-surface models for predominantly C4 landscapes."*

P8 L 31: What do you mean by the 'adjoint performs well'? Does it perform well in terms of efficiency? And if so, how efficient is the adjoint?

Sentence changed to: *"the adJULES system works well in finding optimal parameter vectors which improve the performance of JULES at individual sites, regardless of PFT".*

P9 Fig 2: Which sites are you showing and what are the units? On what basis did you select the shown sites?

The site identification code has been added to the plots and the units moved from the top of the figure

to the side of each individual panel for clarity. The sites picked were the ones that captured best the general trends for each of the PFTs. This was done manually.

P9 last sentence: Why didn't you include these parameters in the optimisation?

As our focus is on the carbon cycle, we choose to only optimised parameters directly relating to the photosynthesis equations in the JULES model. Given more time and more computing power, more parameters could be used in the optimisation.

P10 Fig 2 caption: Please remove the extra 'vector'.

Corrected.

P10 L2: Again, 'validate' is the wrong word here. And why only for broadleaf sites?

With the removal of section 3.2, this is no longer relevant.

P10 L4: The sentence need to rephrased.

Similarly, this sentence was removed when section 3.2 was suppressed.

P10 L6: What do you mean by 'training sets'? This sounds a bit like as if you were using a neural network approach, which has to be trained.

No longer relevant with the removal of section 3.2.

P10 L8: What are these sets?

No longer relevant with the removal of section 3.2.

P11 L1-3: Why should adding more sites render the cost function more smoothly? It could also be the opposite, please explain in the manuscript.

With the removal of section 3.2, this is no longer relevant. However, this is a phenomenon hypothesised in Kuppel et al. (2014) and there are a couple of examples of this happening in figure 6 (old). As a result, the text on P11 L1-3 has been suppressed and the following text added to page 21:

"For some sites, US-Blo and BW-Ma1 for example, the PFT-generic parameter vector over-performs the parameter vector found locally. This phenomenon was also noted in Kuppel et al. (2014). The study further suggested that the added simultaneous constraints placed on the parameters by increasing the number of sites used in the cost function caused the cost function to become 'smoother' and so the optimisation scheme is less likely to get stuck in local minima."

P11 Sec 3.3: This section is really only a description of the posterior parameters but they need to be discussed as well and put in context of a) their prior values, b) their physical meaning and c) the covariances with respect to the resulting fluxes and a successful optimisation.

This section has been expanded to include a more thorough analysis of the correlations in the context of physical interpretation. In order to achieve this, a description of the relevant JULES equations has been added to section 2.1. This puts the parameters in terms of the equations they govern. These equations are then used in section 3.2 to explain why some of the parameters vary the way they do.

P11 L9: What do you mean by 'new uncertainties'?

Sentence changed to: "the prior parameter value is found outside the posterior uncertainty bounds".

P11 L12/13: The uncertainties cannot be skewed, it's the PDF that can be skewed.

Corrected.

P11 L15: What is the 80% confidence interval, how did you calculate this?

The 80% fraction interval was calculated by taking the difference between the 90th and 10th quantile and dividing by the prescribed range. To make this clearer, this interval has been renamed the 80% "quantile" interval. The following description is added to section 2.5.1:

"In order to illustrate the parameter uncertainties, error bars are used to represent the 80% quantile range (10th to 90th percentile) for each optimal parameter."

P11 L30/31: Why are the correlations related to the number of sites used in the optimisation? Please explain in the manuscript.

This was a trend that was observed - the more sites used in the calibration, the more pronounced the correlations seemed to be. However, there was not enough time to run this experiment fully for the paper. The higher correlations found between the parameters for the BT and NT compares to the grass PFTs has now been address in the response to comment P11 Sec 3.3. As a results, this hypothesis has been removed from the text.

P12 L10/11: What makes the UK-PL3 site different? Please explain in the manuscript.

Text added: *"This UK site is in the Pang/Lambourn catchment, which has chalk soil with macropores that permit significant lateral subsurface flows of soil moisture. These horizontal flows cannot be captured in a model like JULES which is essentially one-dimensional in the vertical below the soil surface."*

P12 Sec 3.3.3: As mentioned in the general comments, the calibrated parameter set should be evaluated against independent data.

Issue addressed in response to general comments.

P12 L25/26: What do you mean here? Please rephrase the sentence.

The conclusion was been reworded due to the addition of validation to our study.

Fig 4: Please label the rows. Maybe increase the bar size to improve readability.

We have been experiencing issues with some printed version of the PDF, sometimes some of the information is missing. We will address this. On the online version, the rows are labelled with each parameter symbol. In order to declutter this figure, the prior and posterior values have been removed since this values are made explicit in Table B1. This has allowed us to increase the error bar plots. The lines have been increased to improve readability.

Fig 6: What is the difference between top and bottom panel and what to the vertical lines denote? What are the outliers that have been removed and why did you remove them?

As mentioned above, printed versions of this paper may show incomplete plots, the online version should still contain all the information. The two panels are the same, with Broadleafs and C3 grasses shown in the top panel, and Needleleafs, Shrubs and C4 grasses shown in the bottom panel. The vertical lines are there to break up the different PFTs. The outliers were removed from the plot because they made plot unreadable with much higher errors than the rest in the plot (x10). The sites removed are listed in the caption to the figure. With the slightly adjusted metric, as discussed previously, there are now only 2 such outliers and the data has now been split into 4 panels. An "i.e."

was added into the brackets to clarify that this list contains the outliers.

References:

Cox PM, Huntingford C, Harding RJ. (1998) A canopy conductance and photosynthesis model for use in a GCM land surface scheme, Journal of Hydrology, volume 212-213, pages 79-94.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F. and Richardson, A.D., 2012. Constraining a global ecosystem model with multi-site eddy-covariance data. Biogeosciences, 9(10), pp.3757-3776.

Kuppel, S., Peylin, P., Maignan, F., Chevallier, F., Kiely, G., Montagnani, L. and Cescatti, A., 2014. Model?data fusion across ecosystems: from multisite optimizations to global simulations. Geoscientific Model Development, 7(6), pp.2581-2597.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R. and Lotsch, A., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote sensing of environment, 83(1), pp.214-231.

Response to Reviewer #2:

The authors would like to thank Reviewer #2 for taking the time to write such helpful, thorough and constructive comments. The comments have been taken into consideration in the revised manuscript. We answer them individually as follows:

1 General things:

The Abstract is clear, although just reading the Abstract, a question might be asked as to why the data is not split into training and test data?

We acknowledge that standard practice in data assimilation is to split the data into training and validation data however, originally, we had thought this not possible. Due to the small number of sites available to us, especially with regards to Shrubs and C4 grasses, we chose to use all sites in the calibration to insure the best possible parameters. On closer inspection, the majority of the sites were found to have extra year of data which we now use to validate the model. The following text has been added to abstract now that validation in a key part of the study *" ...both at the calibration and validation stages"*

Maybe expand just slightly on "a third of which give similar reduction in errors as site specific optimisations". The point being made here is that this suggests parameters are similar and robust between sites. This is always good news for climate modelling, suggesting it is possible to reduce to relatively small numbers of PFTs. Maybe stress this point a bit more? (However, if this is stressed more, then need to explain Groenendijk et al 2010?).

This is now discussed more thoroughly the conclusion.

Lines 34, page 2 - Lines 3, page 3. This feels as if it undersells the adjoint approach! I would make a key bullet point that this is a more sophisticated approach (via matrix inversion) to finding rapidly minima across multiple parameters. It would be almost impossible to replicate these findings using some sort of brute-force optimisation, with nested loops over different parameters.

Text has been added on line 33 to this effect: *"The adJULES system uses the adjoint method which finds minima rapidly across multiple parameters via matrix inversion and has the advantage of reproducibility. Replicating these findings using brute-force optimisation would be prohibitively expensive computationally."*

Around Eqn (1), line 11. Sentence "A cost function f(z). . ." looks like it has remained in by accident, and then the correct sentence is the next one. "The cost consists. . .." (The second sentence correctly identifies that the z-z0 differences also contribute to cost function in Eqn(1)).

Sentence starting with *"A cost function $f(z)$..."* has been removed and the second sentence rephrased as follows: *"The cost function, $f(\vec{z})$, consists of a weighted sum of squares of the difference between $\vec{m}_t$ (the vector of model outputs at time t), and $\vec{o}_t$ (the vector of observations at time t), combined with a term quadratic in the difference between parameter values $\vec{z}$ and initial parameter values $\vec{z}_0$"*

Eqn(1) - Has Lambda been accidentally dropped from Eqn (1) . It should multiple the second term? (I realise line 21 states it is taken as unity, but I?d still put it in Eqn(1), and state line 21 "All parameters and observations are equally weighted in this cost

function - i.e. lambda=1"

Lambda added to equation.

Is there a good reason for selecting lambda=1 (or its implications)? Does it imply we put equal trust in the FLXUNET measurements (left term) as the local measurements that give the local parameters (right term). A couple of words on this might help the reader?

The cost function was set up such that both terms are equally weighted. This is used in the single site experiment. Due to the number of sites, it was not possible to tune the value of λ for each individual site. For the multisite experiments, more time was spent tuning this value. What value λ should have is still something we need to look at, hopefully we will get a chance to understand it properly in further study. Text added to the Experiment setup section, explaining the tuning of λ for the multisite cases:

"Preliminary experiments showed very narrow uncertainties whilst running the optimisation scheme over multiple sites i.e. the background term was found to dominate the cost function. In previous multisite studies (Kuppel et al., 2012, 2014), the prior range was also used to defined the background covariance matrix B. The range was variously further multiplied by a factor of 40% (Kuppel et al., 2012) and 1/6 (Kuppel et al., 2014). Experiments were run to find a similar factor to use in this study (the constant of proportionality in Eq. 5). In each of the multisite experiments, the lowest value of such that the Hessian is positive definite at the optimal parameter value was used. This allows uncertainties to be generated around each parameter and prevents the gradient descent algorithm from reaching the boundaries of the prescribed prior range."

Possibly me being confused, but if B is a diagonal matrix, then this isn't about covariances? which imply off-diagonal terms?

The parameters are assumed to start off uncorrelated. Text added: *"...The matrix* **B** *describes the prior covariances assigned to the parameters, and is here chosen to be a diagonal matrix proportional to the inverse square of the ranges allowed for each parameter. The prior uncertainties are therefore assumed to be uncorrelated between the parameters."*

Section "Multisite Implementation". Could tighten slightly to say something like "and this would introduce a double summation in Eqn(1), over n locations. Hence R and B become matrices of size [n*s x n*s]?" Is that correct? This would fit with, as stated, to find "values for a common set of parameters". This gives single values for each z parameter. The wording of the last sentence is slightly ambiguous? "Similarly, the first and second derivative. . ...using the sum of the derivatives at the individual sites". This reads as if the derivatives are calculated locally, and then a mean taken. Would in fact a single sweep across all n*s data points be used, and the derivatives calculated once, if common parameters are investigated. [Maybe eqn term cancellation implies they are the same, but. . ...?].

The following equation has been added to the section to clarify the methodology with s denoting different sites:

$$f\left(\vec{z};\hat{\vec{z}},\vec{z}_0\right) = \frac{1}{2}\left[\sum_s\sum_t \left(\mathbf{m}_{t,s}(\vec{z}) - \mathbf{o}_{t,s}\right)^T \mathbf{R_s}\left(\hat{\vec{z}}\right)^{-1}\left(\mathbf{m}_{t,s}(\vec{z}) - \mathbf{o}_{t,s}\right) + S\lambda(\vec{z}-\vec{z}_0)^T\mathbf{B}^{-1}(\vec{z}-\vec{z}_0)\right]. \quad (1)$$

Somewhere in Section 2.3 or 2.4 ? possibly remind readers that FLUXNET also comes with the meteorological data. (In other words, it's not just the fluxes and then something like NCEP or ECMWF data was used additionally to give met drivers).

Text added to P6 L16: *"... using the meteorological forcing data.."*

Section 2.5.2. Would need to be confident that outlier points didn't do something odd in Eqn (4)? I guess the initial sweep of data ensures this is OK? (The alternative would be to normalise with SD of (mi,t)), and then get the percentage of variance explained).

Thank you very much for this suggestion. By normalising with the standard deviation, we realised how similar our metric was to the fraction of variance unexplained. The fraction of variance unexplained is a useful metric used in statistical analysis and suits our problem well, therefore was picked has an alternative. The values are more intuitive, with 0 still representing a perfect match to the observations. The following text now replaces lines 22-27 on page 7:

"For each data stream i, the fraction of variance unexplained by the model is

$$\epsilon_i^2 = \frac{\sum_{t=1}^{k}(\vec{o}_{i,t} - \vec{m}_{i,t})^2}{\sum_{t=1}^{k}(\vec{o}_{i,t} - \vec{\bar{o}}_i)^2}, \quad \text{where} \quad \vec{\bar{o}}_i = \frac{1}{k}\sum_{t=1}^{k}\vec{o}_{i,t} \tag{2}$$

It follows that the mean fraction of variance unexplained across data streams,

$$\epsilon^2 = \frac{\epsilon_1^2 + \epsilon_2^2}{2}, \tag{3}$$

is a single dimensionless measure of model misfit. The fractional error ϵ can than be interpreted as the typical (root-mean-square) error expressed as a fraction of the (root-mean-square) magnitude of the observed seasonal cycle. Thus, $\epsilon = 0$ represents a perfect match to the observations, while $\epsilon = 1$ corresponds to the error in a null model whose prediction $\vec{m}_{i,t}$ always equals the observational mean $\vec{\bar{o}}_i$. "

Figure 3. This is a great figure. However, pictures and captions often get pulled out of papers and shown in isolation. To ensure information is safely contained, mention in caption (or across top of plot), these are broadleaf trees only? At first glance, I thought the y-axis was some sort of physical unit (for LE or NPP). However caption says this is from the 2.5.2 Section metric. The normalised values are very small, and don't ever get near unity. Could this be the outliers mentioned above? Not a problem, but bottom of page 7 says "1 ? a complete mismatch" Wouldn't we expect some of the parameters to perform quite badly, and get a bit nearer to unity? Or - does this mean that in general, even without parameter fitting, then JULES is an exceptionally good model? Fitting reduces that last small error down further? Figure 3 mentions training versus validation. This appears different to the impression of the Abstract that all data is used to train?

Section 3.2 was put in to asses the multisite methodology. It showed that given a set of 5 randomly selected sites, the optimised parameter vector found by optimising over these sites also improves the rest of the sites not used in the calibration. This experiment is now obsolete since we have the ability to validate the PFT-specific parameters properly in the improved result section. As a result, this figure has been removed. The altered metric, as described in the previous comment, is now more intuitive. A value 1 no longer represents a complete mismatch, which was unclear, but corresponds to the misfit of a null model whose prediction is equal to the mean observation at every time point.

Figure 3. Usual practice is to put the legend inside the plot ? there is space for the 5 symbols, top left hand maybe?

Though Figure 3 no longer exists (see response above), Figure 6 is very similar and the legend has been moved inside the plots for this figure.

Figure 4 is great. But on my print out, the vertical lines cannot be seen in many instances. Thicken them maybe? As always, a matter of style, so just a suggestion. To make Figure 4 less crowded, would it make sense to not put the value of original & optimised as text annotations as this repeats information in the plots. Then the plots can be made bigger and bolder? Maybe put units in left column?

These are nice ideas. In order to declutter this figure, the prior and posterior values have been removed since this values are made explicit in Table B1. This has allowed us to increase the error bar plots. The lines have been thickened to improve readability. Units have been left out in order not to re-clutter the figure.

Section 3.3 and Figure 5. What is so remarkable about Figure 5 is that the strong correlations between parameters are not consistent across the PFTs. Maybe not for this paper, but some sort of physical interpretation of that would be really interesting. Returning to the governing equations and their scaled amounts might help. Is "correlation" the best word ? "collinearity" might be more appropriate?

This section has been expanded to include a more thorough analysis of the correlations in the context of physical meaning. In order to achieve this, a description of the relevant JULES equations has been added to section 2.1. This puts the parameters in terms of the equations they govern. These equations are then used in section 3.2 to explain why some of the parameters vary the way they do.

Figure 6 is like Figure 3, but a lot less cluttered. The data on Figure 6, BTs is same as that on Fig 3, except the "multi-site". Looking back at Fig 3, need to understand better the "five sites" algorithm (again, page 10, line 4 - some text accidently deleted?).

As mentioned in response to the comment above Fig 3, the section containing Fig 3 has been suppressed and only Figure 6 remains.

Figure 6 - I'd make the lower y-axis bound 0.0 (rather than what looks like 0.001)? Gives a better feel then of the improvement in absolute terms.

With the amended metric, a log scale is now used in this figure. It follows that a fixed vertical improvement in this figure represents a fixed (multiplicative) reduction in fractional error.

Conclusions - To my eye, Figure 6 says it all, and I would stress far more the real headline findings that: (!) There is a general reduction in error of around 50% (2) Possibly of more importance, using cross-PFT parameters, often get very similar improvements than local fits. This implies robust parameterisations independent of geography - which GCM modellers always like to see.

The conclusion has been expanded to emphasise this point.

2 Small things:

Maybe get the words "Data Assimilation" used a few times on the paper on page 1 / Abstract, so it gets picked up for anyone using that expression in an Internet search. (It?s an older terminology used for this sort of approach, but is still valid).

Added the term *"data assimilation"* in the abstract.

Abstract: Line 2, maybe mention that JULES is also used comprehensively as an impacts tool, sometimes forced with known climatologies and/or alternative GCMs in to the future. So it is not used just coupled to UK Met Office models.

Text added to abstract: *"JULES is also extensively used offline as a land-surface scheme impacts tool, sometimes forced with known climatologies into the future."*

Abstract - could "automatically differentiated" be expanded slightly to "automatically differentiated with respect to JULES parameters. . .. . ."

Text added.

Maybe line 25, page 1. To make topical post COP21, could add something like: "Any future decreased ability of the land surface to draw-down atmospheric CO2 could imply fewer "permissible emissions" in order to stay below key warming thresholds such as two degrees"

Text added.

Top page 2. Is there a process other than nitrogen cycling that can be mentioned ? preferably one that has been introduced in to the JULES model version used here?

Canopy light interception has been added to the text as a process modelled in the version 2.2 of JULES: *"..or canopy light interception (Mercado et al., 2009)."*

Sentence "Given the small spatial footprint. . .". Maybe clarify why this gives overtuning? Presumably because it might be see some anomalous plants in the small footprint, and that are not representative of PFTs over a broader area?

Text added to clarify this *"This over tuning may be due to the fact that a single site may not represent the full range of a PFT, given different tree types, ages and aboveground biomass found at each site. There may be some anomalous plants in the small footprint that are not representative of the PFTs over a broader area."*

Bottom of page 3. Line 30. Could mention that "available observations" are about independent large-scale measurements (such as FLUXNET)? These are different to the specific process measurements used to calibrate the individual components tha are mentioned in line 27.

Text added: *"...available observations, such as eddy covariance flux data."*

Page 4, line 5. Possibly: "As used widely in weather forecasting, along with other disciplines".

Added.

Page 4, line 14. Maybe: " . . ..outputs at time 1 <= t <= s" (so defines s).

Since s is not used anywhere in the paper, it has been removed from the sum.

Page 4, line 22. To anyone new to data assimilation, could say: "optimal vector. . ..minimizes the cost function (Eq. 1) via JULES model itself though m=m(z) (left terms) and directly via z in the right terms"

The sentence has been left unchanged, instead the term '$\mathbf{m}(z)$' has been added to Eq. 1. Hopefully this addition clarifies that the model time-series part of the cost function changes with different iterations of z.

Page 4, line 25. I can understand box constraints = upper or lower bounds. But what does "limited memory" refer to?

Compared to the full BFGS algorithm which stores a full approximation to the inverse Hessian, the limited memory version will only store a few vectors to represent the approximation (Bryd et al.,1995). For optimisation problems with large numbers of variables, this linear memory requirement makes the L-BFGS the variant of choice. Texted added to line 25: *"... to use limited memory, for computational affordability, and box constraints..."*

On the diagram, Figure 1, right-hand side, maybe word as "Hessian to give uncertainty bounds"

Text added to picture.

Page 5, line 14. Table A1 has a lot fewer than 500 entries, so quite a lot of data is rejected?

Of the 500 FLUXNET sites, we were only able to obtain 160. Of those, 50% were rejected. As well as the criteria listed in this section, sites were also excluded based on number of years available (2 years minimum in order to perform a spin-up) and whether the sites were dominated by particular PFT (e.g. more than 50% coverage with the exception of C4 grasses). The crop sites were excluded since in the newer versions of JULES these are considered separately. These extra exclusions are described in section 2.4. This is made explicit in the paper using the following text: *"Data from 160 sites were made available for this study by M.Groenendijk."*

Page 7, line 15. "One or two" refers to whether a plot is a standard plot, or a contour plot in two parameters?

This was referring to both and has hopefully been clarify with the addition of the text below:

"Consequently the optimal parameter values (which are modes of the full high dimensional distribution) may not coincide with modes of the one- and two-dimensional marginal distributions."

Page 8, line 5. The analysis here is more testing the concept of common parameters between sites, rather than testing the methodology?

This sentence has been suppressed with the removal of section 3.2.

Figure 2 is really nice. Just a few small things. Is there are reason black is also dashed? Style thing, but I've have maybe put as the individual panel titles "Broadleaf LE, Broadleaf GPP", etc. So across the top of the panels. And then the y-axes, put the units ? so W/m2 left panels etc. Then possibly not bothered with the labels (a)-(e)? Maybe make the lines with slightly larger line width? Inside each box, give the site ID as annotated text, as these are time-series for just single sites.

The black line denotes the observations; each point represents a FLUXNET observation and the dashed part of connects them. The figure has been changed to include the rest of the suggestions and the text changed according.

Page 10, line 4. Some text missing from sentence?

No longer relevant with the removal of section 3.2.

References:

Byrd, R., Lu, P., Nocedal, J., and Zhu, C.: A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific Computing, 16, 1190?1208, 1995.

Mercado, L.M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M. and Cox, P.M., 2009. Impact of changes in diffuse radiation on the global land carbon sink. Nature, 458(7241), pp.1014-1017.

Land surface parameter optimisation through data assimilation: the adJULES system

Nina M. Raoult[1], Tim E. Jupp[1], Peter M. Cox[1], and Catherine M. Luke[1]

[1]National Centre for Earth Observation, University of Exeter, Exeter EX4 4QF, UK

Correspondence to: Nina Raoult (nr278@exeter.ac.uk)

Abstract. Land-surface models (LSMs) are crucial components of the Earth System Models (ESMs) which are used to make coupled climate-carbon cycle projections for the 21st century. The Joint UK Land Environment Simulator (JULES) is the land-surface model used in the climate and weather forecast models of the UK Met Office. JULES is also extensively used offline as a land-surface impacts tool, forced with climatologies into the future. In this study, JULES is automatically differentiated with respect to JULES parameters using commercial software from FastOpt, resulting in an analytical gradient, or adjoint, of the model. Using this adjoint, the adJULES parameter estimation system has been developed, to search for locally optimum ~~parameter sets~~ parameters by calibrating against observations. This paper describes adJULES in a data assimilation framework and demonstrates its ability to improve the model-data fit using eddy covariance measurements of gross primary production (GPP) and latent heat (LE) fluxes. adJULES also has the ability to calibrate over multiple sites simultaneously. This feature is used to define new optimised parameter values for the 5 Plant Functional Types (~~PFTS~~PFTs) in JULES. The optimised PFT-specific parameters improve the performance of JULES over ~~90~~85% of the sites used in the study ~~, a third of which give similar reduction in errors as site specific optimisations~~both at the calibration and validation stages. The new improved ~~parameter set~~ parameters for JULES is presented along with the associated uncertainties for each parameter.

1 Introduction

Land-surface models (LSMs) have formed an important component of climate models for many decades now (Pitman, 2003). First generation land-surface schemes focussed on providing the lower boundary condition for atmospheric models by calculating the land-atmosphere fluxes of heat, moisture and momentum, and updating the surface state variables that these fluxes depend on (e.g. soil temperature, soil moisture, snow-cover). In the mid to late 1990s some land-surface modelling groups began to introduce additional aspects of biology into their schemes, most notably the dynamic control of transpiration by leaf stomata and the connected rates of leaf photosynthesis (Sellers et al. (1997); Cox et al. (1999)).

In the early 2000s, climate modelling groups began to use the carbon fluxes simulated by LSMs within first generation climate-carbon cycle models (Cox et al. (2000), Friedlingstein et al. (2001)). These early results, and a subsequent model intercomparison (Friedlingstein et al., 2006), highlighted the uncertainties associated with land carbon-climate feedbacks. The 5th Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5 (Stocker et al., 2013)) for the first time routinely included models with an interactive carbon cycle (now called Earth System Models or ESMs), confirming that land

responses to climate and CO_2 are amongst the largest of the uncertainties in future climate change projections (Arora and Boer (2005); Brovkin et al. (2013); Jones et al. (2013); Friedlingstein et al. (2013)). Any future decreased ability of the land surface to draw-down atmospheric CO_2 could imply smaller "compatible emissions" in order to stay below key warming thresholds such as two degrees.

Uncertainties in LSMs arise from two major sources: ~~(a) process uncertainty , and (b)~~ process uncertainty and parameter uncertainty. Process uncertainty includes the misrepresentation of land-surface processes and also the neglect of important processes (such as nitrogen-limitations on ~~plants~~ plant growth, see for example ~~.~~ Thornton et al. (2007); Zaehle et al. (2010)), or canopy light interception (Mercado et al., 2009) . The drive to reduce process uncertainty almost invariably leads to increases in LSM complexity, which typically leads to the introduction of additional internal model parameters. Parameter uncertainty arises from uncertainty in these internal model parameters. The evolution of LSMs has therefore involved an attempt to reduce process uncertainty by increasing model realism/complexity, but at the cost of increasing parameter uncertainty. This paper concerns the development and application of a technique to reduce parameter uncertainty in the widely used Joint UK Land Environment Simulator (JULES) LSM (Best et al. (2011); Clark et al. (2011)).

Optimisation techniques come under the umbrella of model-data fusion and range from simple ad-hoc parameter tuning to rigorous data assimilation frameworks. These approaches have been used in a number of studies, covering various ~~LSM~~LSMs, to derive vectors of parameters that improve model-data fit significantly (e.g. Wang et al. (2001, 2007); Reichstein et al. (2003); Knorr and Kattge (2005); Raupach et al. (2005); Santaren et al. (2007); Thum et al. (2008); Williams et al. (2009); Peng et al. (2011)). Many of these studies calibrate the model at individual measurement sites. Given the small spatial footprint of each flux tower, this can often result in ~~over tuning~~over-tuning. This over-tuning may occur when a single site does not represent the full range of a PFT, given different tree types, tree ages and aboveground biomass found at each site. There may be some anomalous plants in the small footprint that are not representative of the PFTs over a broader area. The optimised model parameters are site-specific and often struggle to perform as well when generalised over other sites (Xiao et al., 2011).

The majority of LSMs group vegetation into a small number of plant functional types (PFTs). Model parameters are assumed to be generic over each PFT. Through different optimisation techniques, some studies have tried to assess the robustness of PFT-specific parameters (e.g. (Kuppel et al., 2014)). Medvigy et al. (2009) and Verbeeck et al. (2011) both show that parameters derived at one site can perform well on a similar site and ~~, in a later study (Medvigy and Moorcroft (2011)),~~ over the surrounding region (Medvigy and Moorcroft (2011)). However, a contradictory study by Groenendijk et al. (2010) found that there was cross-site parameter variability after optimisation within the PFT groupings.

In the last few years, there has been a move towards deriving PFT-specific parameters using data from multiple sites, the results of which have been generally positive ~~,~~ (e.g. Xiao et al. (2011) and Kuppel et al. (2012)). Both of these studies used data from multiple sites in their optimisation (calling it multisite optimisation) and have commented on the robustness of this technique, showing that the choice of the initial parameter vector had little effect on the optimised values.

Kuppel et al. (2012) compared different approaches for finding generic PFT-specific parameters, such as averaging optimised parameter vectors over PFTs and directly optimising over multiple sites. They found that the latter method was best for finding

PFT-specific parameters. The multisite optimisation procedure was refined in Kuppel et al. (2014), extended to other PFTs, and evaluated at a global scale.

For global modelling, there is a clear need to find generic parameters and associated uncertainties ~~by~~ for each PFT, by optimising against observations in a reproducible way. This paper presents a model-data fusion framework, called adJULES, that allows data from multiple sites to be used simultaneously in order to improve the JULES land surface model. The adJULES system uses the adjoint method which finds minima rapidly across multiple parameters via matrix inversion and has the advantage of reproducibility. Replicating these findings using brute-force optimisation would be prohibitively expensive computationally.

This paper aims to answer the following questions:

- Can an optimum vector of generic parameters for each of the JULES PFT classes be found?
- How does the ~~new~~ optimal PFT parameter vector compare to parameter vectors found by optimising each site individually?
- How robust is the adJULES system when optimising over multiple sites?
- What uncertainty is associated with each parameter?

In section 2, methods and data used in the study are described. The JULES land surface model and our new data assimilation system (adJULES), are ~~described. The data used , and~~ introduced, along with the data used and the parameters chosen to be optimised in the study~~, are also discussed~~. In section 3, the results are presented. The methodology for optimising over multiple sites simultaneously is validated, and optimum parameter values are provided for each JULES PFT. The performance of the new parameter sets is assessed and shown to ~~significantly improve~~ improve significantly the fit of the JULES model to the observations. The conclusions are laid out in section 4.

2 Methods and Data

2.1 The JULES land-surface model

The JULES land-surface model (Best et al., 2011; Clark et al., 2011) simulates the interactions between the land and atmosphere. Originally developed from the Met Office Surface Exchange Scheme (MOSES) (Cox et al., 1999), JULES can be used 'offline' with observed atmospheric forcing data, or can be coupled into a global circulation model (GCM). JULES is the land surface model used in the UK Met Office Unified Model.

JULES is a mechanistic land surface model including physical, biophysical, and biochemical processes that control the radiation, heat, water, and carbon fluxes in response to time-series of the state of the overlying atmosphere (Best et al., 2011; Clark et al., 2011). Processes such as photosynthesis, evaporation, plant growth and soil microbial activity are all linked through mathematical equations that quantify how ~~soil moisture and temperature govern~~ environmental conditions affect evapotranspiration, heat balance, respiration, photosynthesis and carbon assimilation (Best et al., 2011; Clark et al., 2011). JULES runs at a

given sub-daily step (typically 30 minutes), using meteorological drivers of rainfall, incoming radiation, temperature, humidity, and windspeed as inputs.

Vegetation in the JULES model is categorised into five ~~plant functional types (PFTs)~~PFTs; broadleaf trees (BT), needleleaf trees (NT), C3 grasses (C3G), C4 grasses (C4G), and shrubs (Sh). Default parameters for these PFT classes are taken from a previous ~~ad hoc calibration~~ study (Blyth et al., 2010) .

The eight parameters that are calibrated within this study (see Table 1) relate predominantly to leaf-level stomatal conductance (g) and photosynthesis (A). Four of these parameters control the responses of g and A to environmental conditions such as surface temperature (T_{upp}, T_{low}), solar radiation (α), and atmospheric humidity deficit (dq_c). Two other calibration parameters (f_0, n_{l0}) essentially control the maximum values of g and A. The remaining two calibration parameters influence the hydrological partitioning at the land-surface and relate to the amount of rainfall intercepted by the plant canopy ($\delta c/\delta L$), and the "rootdepth" (d_r) from which each PFT can access soil water for transpiration. The simulated latent heat flux and gross primary productivity have been found to be especially senstive to these parameters in previous studies (Blyth et al., 2010).

The full set of equations within the JULES model is documented in the papers by Best et al. (2011) and Clark et al. (2011) , but the key equations are highlighted below. In JULES , leaf-level photosynthesis and stomatal conductance are treated with a coupled model (Cox et al., 1998) . Based on the models of **??** , leaf-level photosynthesis A is controlled by the carboxylation rate (which depends on n_0, T_{low}, T_{upp}) and light-limited photosynthesis (which depends on α). It follows that:

$$A = A(n_0, \alpha, T_{low}, T_{upp}, c_i, \beta) \tag{1}$$

where c_i is the internal CO_2 concentration inside the leaf, and β is a soil moisture stress factor which depends on the vertical soil moisture profile θ, and the plant rootdepth d_r:

$$\beta = \beta(\theta, d_r) \tag{2}$$

The internal CO_2 concentration c_i is assumed to be dependent on the external CO_2 concentration c_a and the atmospheric humidity deficit δq (Cox et al., 1998) via the equation:

$$\frac{c_i - c_*}{c_a - c_*} = f_0 \left(1 - \frac{\delta q}{\delta q_c}\right) \tag{3}$$

where c_* is the CO_2 compensation point, and f_0 and δq_c are parameters that are calibrated in this study. The stomatal conductance for water vapour g is diagnosed in JULES from the leaf-level photosynthesis A and the internal and external CO_2 concentrations:

$$g = 1.6 \frac{A}{c_a - c_i} \tag{4}$$

The factor of 1.6 converts the stomatal conductance for CO_2 into a stomatal conductance for water vapour. The scaling-up from leaf to canoy level in this version of JULES uses a "big-leaf" approach (Cox et al., 1999) .

2.2 Data assimilation system

Even a relatively simplistic land-surface representation such as JULES has over a hundred internal parameters representing the environmental sensitivities of the various land-surface types and PFTs within the model. In general these parameters are chosen to represent measurable ~~quantities within the real world~~'real world' quantities (e.g. aerodynamic roughness length, surface albedo, plant root-depth)~~, which~~ . This allows observationally-based estimates of these parameters to be made in the early stages of the model development process. However, the detailed performance of a land-surface model can be very sensitive to such internal parameters. It is therefore common for land-surface modellers to calibrate their models against available observations~~(e. g. Blyth et al. (2010)).~~ , such as eddy covariance flux data. This is typically carried-out in a rather ad hoc manner with the modeller varying the parameters that he/she believes are most relevant to the model performance. Such model tuning is by its very nature subjective, lacks reproducibility, and is often sub-optimal because the modeller is unable to explore the full feasible parameter space through such a manual technique.

This paper describes a more objective approach to land-surface model calibration, adopting ideas from the applied mathematics of data assimilation as used widely in weather forecasting and other disciplines, and motivated by pioneering attempts at carbon cycle data assimilation (Rayner et al. (2005); Kaminski et al. (2013)). It utilises the adjoint of the JULES model~~(called adJULES)~~, derived by automatic differentiation, which enables efficient and objective calibration against observations. Importantly, adJULES also allows the uncertainties in the best-fit parameters to be estimated. Such uncertainties are important information for model users, and can also form the basis for observation-constrained estimates of ~~prior~~ posterior probability density functions for the land-surface parameter perturbations used in climate model ensembles (e.g. Booth et al. (2012)).

2.2.1 The theory of adJULES

JULES generates a modelled time-series for a given vector of internal parameters, $\boldsymbol{z}$. ~~A~~ The cost function, $f(\boldsymbol{z})$~~is defined as~~ , consists of a weighted sum of squares of ~~differences between the modelled and the observed time-series. The cost consists of the difference between~~ the difference between $\boldsymbol{m}_t$ (the vector of model outputs at time t~~, $\boldsymbol{m}_t$ and~~), and $\boldsymbol{o}_t$ (the vector of observations at time t~~, $\boldsymbol{o}_t$,~~), combined with a term quadratic in the difference between parameter values $\boldsymbol{z}$ and initial parameter values $\boldsymbol{z}_0$~~and values $\boldsymbol{z}$~~:

$$f(\boldsymbol{z}) = \sum_{t=1}^{s} (\mathbf{m}_t - \mathbf{o}_t)^T \mathbf{R}^{-1} (\mathbf{m}_t - \mathbf{o}_t) + (\boldsymbol{z} - \boldsymbol{z}_0)^T \mathbf{B}^{-1} (\boldsymbol{z} - \boldsymbol{z}_0).$$

~~Here $\mathbf{R}$ is the observed covariance in the errors $(\mathbf{m}_t - \mathbf{o}_t)$and~~

$$f(\boldsymbol{z}; \hat{\boldsymbol{z}}, \boldsymbol{z}_0) = \frac{1}{2} \left[\sum_t (\mathbf{m}_t(\boldsymbol{z}) - \mathbf{o}_t)^T \mathbf{R}(\hat{\boldsymbol{z}})^{-1} (\mathbf{m}_t(\boldsymbol{z}) - \mathbf{o}_t) + \lambda (\boldsymbol{z} - \boldsymbol{z}_0)^T \mathbf{B}^{-1} (\boldsymbol{z} - \boldsymbol{z}_0) \right]. \quad (5)$$

Here, $\mathbf{R}(\hat{z}) = \frac{1}{n} \sum_{t=1}^{n} (\mathbf{m}(\hat{z})_t - \mathbf{o}_t)(\mathbf{m}(\hat{z})_t - \mathbf{o}_t)^T$ denotes the error cross product matrix produced by a JULES run with parameter value $\hat{z}$. In an optimisation, $\boldsymbol{z}$ and $\hat{\boldsymbol{z}}$ are updated separately in nested loops, having both been initialised to the default JULES parameter value $\boldsymbol{z}_0$. In the inner loop, $\boldsymbol{z}$ is varied to minimise the cost function (termination criterion: $\nabla f \approx 0$)

for the current value of $\hat{z}$. In the outer loop, $\hat{z}$ is reset to the new value of z from the inner loop (termination criterion: change in $\hat{z}$ negligible). At the end of an optimisation, therefore, the matrix $\mathbf{R}$ conveys information about the error correlation structure in a JULES run with optimal parameter values.

The matrix $\mathbf{B}$ describes the prior covariances ~~in the parameterswith~~ assigned to the parameters, and is here chosen to be a diagonal matrix proportional to the inverse square of the ranges allowed for each parameter. The prior uncertainties are therefore assumed to be uncorrelated between the parameters. The constant of proportionality λ ~~, with default value 1, controls the width of the prior distribution and ultimately the~~ controls the relative importance of the background term ~~.~~ (i.e. the right-hand term in Eq. 5) and the error term (i.e. the left-hand term in Eq. 5). Larger values of λ help condition the problem ~~(Bouttier and Courtier, 1999)~~ and force parameter values to be close to the initial value z_0 (Bouttier and Courtier, 1999) . All parameters and observations are equally weighted in this cost function.

The optimal vector of parameters is the vector z that minimises the cost function (Eq. 5). The aim of adJULES is to find this vector. adJULES minimises the cost function iteratively using the gradient descent algorithm L-BFGS-B (Byrd et al. (1995), optim: R Development Core Team (2015)). This algorithm is based on the BFGS quasi-Newton method but is modified to use limited memory, for computational affordability, and box constraints, so each parameter is given an upper and lower bound based on expert opinion or on physical reasoning (Byrd et al., 1995).

At each iteration, the gradient $\nabla f(z)$ of the cost function $f(z)$ is computed with respect to all parameters, using the adjoint model of JULES. The adjoint is generated with the automatic differentiator tool TAF (Transformation of Algorithms in Fortran; see Giering et al. (2005))~~in 'reverse mode' (rather than 'forward mode') for computational efficiency~~. Automatic differentiation relies on using the chain rule, the choice of forward or reverse mode refers to the order in which the derivatives are computed. Calculating $\nabla f(z)$ is most efficient in reverse mode as only one sweep is needed to generate the derivative with respect to all parameters (Bartholomew-Biggs et al., 2000).

Once the cost function reaches the minimum, a locally optimal parameter vector z_1 is returned and the second derivative of the cost function with respect to the parameters can be used to calculate posterior uncertainties. This process is then repeated, the locally optimised parameters are fed back through JULES, generating a new modelled time-series and hence a new cost function. The loop ~~is~~ terminates when the modelled time series no longer improves (Fig. 1).

2.2.2 Multisite Implementation

In its simplest form, adJULES runs at a single grid-point location and so the derived optimal parameter vector is site specific. On the other hand, multisite optimisation aims to find values for a common set of parameters, using data from multiple locations. The definition of the cost function (Eq. 5) can be extended to include the observations from all S sites, and its derivative found in order to use the L-BFGS-B algorithm again. The extended cost function is the sum of the individual cost functions for each

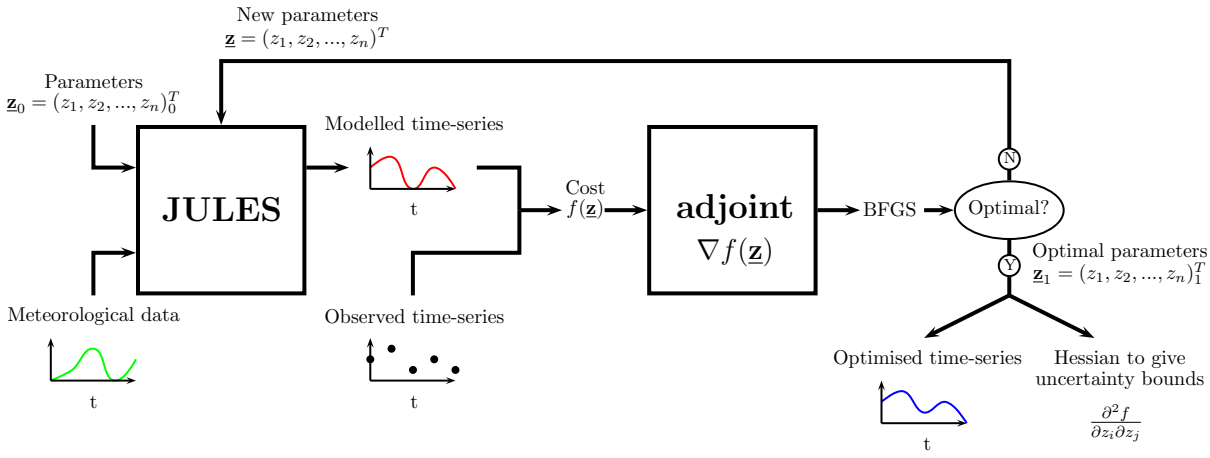


Figure 1. Schematic of the adJULES parameter estimation system starting with the initial parameter vector $\boldsymbol{z}_0$. This is usually based on default JULES parameter values (Blyth et al., 2010). The optimised parameter vector is denoted $\boldsymbol{z}_1$.

site s. Similarly, the first and second derivatives of this new cost function can be defined using the sum of the derivatives at the individual sites.

$$f\left(\boldsymbol{z};\hat{\boldsymbol{z}},\boldsymbol{z}_0\right) = \frac{1}{2}\left[\sum_s\sum_t \left(\mathbf{m}_{t,s}(\boldsymbol{z}) - \mathbf{o}_{t,s}\right)^T \mathbf{R_s}\left(\hat{\boldsymbol{z}}\right)^{-1}\left(\mathbf{m}_{t,s}(\boldsymbol{z}) - \mathbf{o}_{t,s}\right) + S\lambda(\boldsymbol{z}-\boldsymbol{z}_0)^T\mathbf{B}^{-1}(\boldsymbol{z}-\boldsymbol{z}_0)\right]. \quad (6)$$

2.3 Eddy covariance flux data

The eddy-covariance flux data used in this study are part of ~~the FLUXNET network~~ FLUXNET (Baldocchi et al., 2001). The FLUXNET database contains more than 500 locations worldwide, and all the data ~~is~~ are processed in a harmonised manner using the standard methodologies including correction, gap-filling and partitioning (Papale et al., 2006). Data from 160 sites were made available for this study by M. Groenendijk. The sites used in this study were selected based on data availability: sites with missing input variables or ~~significant data gaps~~ data gaps of more than 50% during the growing ~~the~~ season were omitted.

To ~~model photosynthesis~~ constrain photosynthetic parameters, Net Ecosystem Exchange (NEE) and Latent Heat Flux (LE), among other fluxes, are ~~required~~helpful. The NEE flux, defined as the net flux of CO_2, is partitioned into gross primary production (GPP) and ecosystem respiration (~~Re~~Resp) (Reichstein et al., 2005). In this study this GPP flux is used, along with the LE flux to constrain the model. GPP data are model-derived estimates, which could introduce an additional uncertainty into the results.

In an attempt to run the experiments as closely to a standard JULES run as possible, input fields of vegetation structure and soil type were drawn from the UK Met Office ancillary files used in the HadGEM2 configurations. The LAI seasonal cycle used is derived from a MODIS product (?) from Boston University. The values taken for each of the experiment sites correspond to the closest grid point ~~with values.~~ at which data are available. This could lead to inconsistencies between the actual vegetation at a given site and the vegetation structure and soil type used in the model.

2.4 Experimental setup

Version 2.2 of JULES is implemented in the current version of adJULES. This version is set up to calibrate a subset of JULES soil and vegetation parameters against up to six observables in the vectors $\boldsymbol{m}_t$ and $\boldsymbol{o}_t$ (Eq. 5): net ecosystem exchange (NEE), sensible heat (H), latent heat (LE), surface temperature (T^{*}_{s}), gross primary productivity (GPP) and ecosystem respiration (Resp).

This study aims to improve the parameters used to define PFTs and therefore it concentrates on vegetation parameters. Table 1 ~~outline~~ outlines the parameters chosen.

One year of FLUXNET data is used for each site considered in this study at the calibration stage. Where multiple years are available, the most complete year was chosen. For each site ~~,~~ the model is spun up to a steady soil moisture and temperature state. Where possible, the two years of data preceding the year of comparison were repeatedly applied in the spin-up. Where this was not possible, the first year of data was repeatedly applied. Only sites with at least two years of data are used in this study, so that the spin-up year is different from the experiment year. In each case, the model was spun up for at least 50 years. For deciduous sites and crop sites, leaf area index values are taken from MODIS data for the appropriate year. Where possible, a second year of FLUXNET data was spun up to be used at the validation stage of this study. This second year was chosen to be the second most complete year when more than one year was available.

The sites used in each of the PFT classes are described in Appendix A. The FLUXNET database used in this study did not distinguish between the different types of grasslands. Using Met Office ancillary files, the grasslands were partitioned into C3 grasses and C4 grasses according to fractional cover. In the case of C3 grasses, sites were picked only when the fractional cover was over 60%. Since the C4 grasses are ~~under represented~~ under-represented in the FLUXNET database, this boundary was lowered to include all sites where C4 grass was the dominant PFT. Crops were not included in either grass class. The photosynthesis model used in JULES is based on scaling up observed processes at the leaf scale to represent the canopy. The scaling to canopy level can be done in several ways~~, in~~ . In this study the simple big leaf approach was adopted (Clark et al., 2011), although optimisations can also be ~~carried-out~~ carried out for more complex canopy radiation options (Mercado et al., 2009) .

All of the sites in each PFT class are used to find the optimal values for the PFT. The second derivative of the cost function found by differentiation of the adjoint code is then used to quantify the uncertainties associated with these ~~new~~ optimal parameter vectors.

~~For the multisite experiments, the background term was weighted~~ Preliminary experiments showed very narrow uncertainties whilst running the optimisation scheme over multiple sites (i.e. the background term was found to dominate the cost function).

In previous multisite studies (Kuppel et al., 2012, 2014) , the prior range was also used to defined the background covariance matrix $\mathbf{B}$. The range was variously further multiplied by a factor of 40% (Kuppel et al., 2012) and $\frac{1}{6}$ (Kuppel et al., 2014) . Experiments were run to find a similar factor to use in this study (the constant of proportionality λ in Eq. 5). In each of the multisite experiments, the lowest value of λ such that the ~~problem remained conditioned but low enough for useful~~ Hessian is positive definite at the optimal parameter value was used. This allows uncertainties to be generated ~~.~~ around each parameter and prevents the gradient descent algorithm from reaching the boundaries of the prescribed prior range.

2.5 ~~Introducing~~ Analysis tools~~for analysis~~

2.5.1 ~~Different ways to represent parameter~~ Parameter uncertainty

As well as generating optimal parameter values, adJULES estimates the uncertainty associated with each parameter. The second derivative (Hessian) of the cost function,

$$H_{ij} = \frac{\partial^2 f}{\partial z_i \partial z_j} \tag{7}$$

where $f(\boldsymbol{z})$ is given by equation (5), evaluated at the optimal parameter value, yields information about the curvature of the cost function at the local minimum. A 'sharp' cost function, where the cost function is steep either side of the optimal parameter value, indicates lower parameter uncertainty. This can also be interpreted as meaning that a small deviation from the optimal parameter value yields a large increase in cost. Conversely, a 'flat' cost function indicates higher parameter uncertainty, or little change in cost caused by deviation from the optimal parameter value.

In order to generate statistics associated with the curvature of the cost function, the Hessian is used to generate samples from the posterior distribution. This is a truncated multivariate normal distribution (Genz et al., 2015) because of the box constraints placed on the prior. Using Gibbs sampling (Geman and Geman, 1984), an ensemble of plausible parameter vectors is generated from this distribution, for a statistically satisfactory match between observations and modelled time series. The multivariate normal parameter distribution allows for marginal density plots to be generated for each parameter. When considering these marginal density plots, it is important to remember that they represent only one or two ~~dimension of an n-dimensional multivariate normal distributions. The~~ dimensions of a high dimensional multivariate normal distribution which is truncated. Consequently the optimal parameter values ~~may not correspond to the peak of the one-dimensional distributionsas a result~~(which are modes of the full high dimensional distribution) may not coincide with modes of the one- and two-dimensional marginal distributions.

~~Metric~~ In order to illustrate the parameter uncertainties, error bars are used to represent the 80% quantile range (10th to ~~quantify model~~ 90th percentile) for each optimal parameter.

2.5.2 Fractional error - a metric of model-data fit~~to data~~

To measure the improvement exhibited by different parameter vectors, ~~a normalised root-mean-square deviation (ϵ) is used .~~ the fraction of variance unexplained ϵ^2 is used to define the fractional error ϵ. This metric was chosen to show not only the improvement made by the optimal parameter vectors at each site but also to show how each site performed relative to others.

Given a parameter vector, $\boldsymbol{z}$, a modelled time series $\boldsymbol{m}_{i,t}$ with k data points is generated using JULES, where i denotes one of the observable data streams (in this case LE and GPP~~data streams~~). For each data stream i, the ~~ϵ normalised error is calculated as follows~~

$$\epsilon_i = \sqrt{\frac{\sum_{t=1}^{k}(\boldsymbol{m}_{i,t} - \boldsymbol{o}_{i,t})^2}{k}},$$

~~and then normalised:~~

$$\hat{\epsilon}_i = \frac{\epsilon_i}{\max(\boldsymbol{m}_{i,t}, \boldsymbol{o}_{i,t}) - \min(\boldsymbol{m}_{i,t}, \boldsymbol{o}_{i,t})}.$$

~~After non-dimensionalising both~~ fraction of variance unexplained by the model is

$$\epsilon_i^2 = \frac{\sum_{t=1}^{k}(\boldsymbol{o}_{i,t} - \boldsymbol{m}_{i,t})^2}{\sum_{t=1}^{k}(\boldsymbol{o}_{i,t} - \bar{\boldsymbol{o}}_i)^2}, \quad \text{where} \quad \bar{\boldsymbol{o}}_i = \frac{1}{k}\sum_{t=1}^{k}\boldsymbol{o}_{i,t} \tag{8}$$

It follows that the mean fraction of variance unexplained across data streams, ~~the final error is given by~~

$$\hat{\epsilon} = \frac{\hat{\epsilon}_1 + \hat{\epsilon}_2}{\sqrt{2}}.$$

~~This ensures values are between 0 and 1; 0 representing a~~

$$\epsilon^2 = \frac{\epsilon_1^2 + \epsilon_2^2}{2}, \tag{9}$$

is a single dimensionless measure of model misfit. The fractional error ϵ can than be interpreted as the typical (root-mean-square) error expressed as a fraction of the (root-mean-square) magnitude of the observed seasonal cycle. Thus, $\epsilon = 0$ represents a perfect match to the observations, ~~1 a complete mismatch. The closer the value is to 0, the better the set of parameters used is at creating modelled time-series resembling the observed time-series~~while $\epsilon = 1$ corresponds to the error in a null model whose prediction $\boldsymbol{m}_{i,t}$ always equals the observational mean $\bar{\boldsymbol{o}}_i$.

3 Results and discussion

In this section, the site-specific optimisations are first considered. By considering each PFT separately, the misfits between the model and the observations are discussed and the effect of optimising over each site individually to improve model-observation agreement is considered.

Next, the multisite methodology is ~~validated. This is then~~ used to perform optimisations over each of the PFTs. All of the sites in a given PFT are optimised simultaneously to find a generic parameter vector appropriate to the PFT. The new optimised parameter vectors are presented, along with associated uncertainties. Some of the uncertainties and correlations found between parameters are discussed, especially in the context of the equations described in section 2.1. The rest of the section considers the improvement found using these optimised parameter vectors ~~, and discusses some of the uncertainties and correlations found~~both on the calibration year and the validation year for each of the sites.

3.1 Single-site optimisations

First, each of the sites was optimised individually in order to find site-specific parameter vectors. As described in section 2.4, one year runs at the different sites were optimised against monthly averaged LE and GPP. The constant of proportionality λ is set to 1 for all sites, in order to give equal weighting to both terms in Eq. 5. A site dominated by each PFT was picked to represent the general improvements made. The main seasonal cycles of ~~latent heat~~ LE and GPP for the different sites are shown in Fig. 2.

Most broadleaf sites follow the pattern ~~of~~ illustrated (Fig. 2~~(a~~, top row). Normally, for broadleaf sites, a standard JULES run will underestimate GPP. The optimisation does a good job in fixing this, bring the modelled time-series closer to the observations. In contrast, LE does not improve as much.

Similarly for the needleleaf sites (Fig. 2~~(b))~~, second row), the JULES model output tends to overestimate LE and underestimate GPP. The parameter vector found in the optimisation improves the fit of both data streams, most notably GPP. At sites for which a double peak seasonality is apparent, the optimised model captures this better than the original model.

GPP is also underestimated for the C3 grass sites (Fig. 2~~(c))~~, middle row) and, for the majority of the sites, the optimisation does a good job of correcting this. The LE flux tends to ~~be at~~ have the right magnitude before optimisation, unlike the GPP flux, but adJULES does not manage to improve this output significantly. In the example shown, the JULES model using the default parameter vector already performs very well, so little improvement is needed, but this is not always the case. The new set of parameters is also good at simulating multiple peaks in the LE and GPP fluxes, when they are observed.

There are only two C4 grass sites in the set and JULES does not perform very well on these before or after optimisation (Fig. 2~~(e)).~~

~~The shrub sites , Fig. 2(f),~~ , fourth row). The original stomatal conductance-photosynthesis model within JULES was developed based on fluxes measured over C4 grass as part of the FIFE field experiment (Cox et al., 1998) . However, there are relatively few FLUXNET sites over C4-dominated landscapes, and only two even in the extended dataset used here. As a result, the sensitivity of stomatal conductance and photosynthesis to environmental factors has been less well tested for C4 grasses. The results presented in this paper therefore highlight the need to reassess JULES and other land-surface models for predominantly C4 landscapes.

The shrub sites show no general pattern ~~.~~ (Fig. 2, fourth row). Some sites overestimate LE, whilst others ~~underestimated~~ underestimate it, and similarly for GPP. The ~~levels~~ level of improvement varies over sites. For some of the sites in this PFT,

the magnitude of GPP fails to get close to the magnitude of the observations, both before and after optimisation. However, it is hard to pick out a general pattern for this PFT, since there are only 5 sites in this set.

Overall, the ~~adjoint performs well in improving~~ adJULES system works well in finding optimal parameter vectors which improve the performance of JULES at individual sites, regardless of PFT. The systematic underestimation of GPP in default JULES ~~,~~ improves the most. This larger improvement in GPP fit reflects the larger set of optimised parameters that are exclusively related to the carbon cycle. Different parameters may need to be incorporated, for example some soil ones, for the LE flux to improve further.

3.2 ~~Multisite Validation~~PFT-specific optimal parameter values

~~Broadleaf sites were used to validate the multisite methodology. This PFT is the best represented in the FLUXNET network, though since the broadleaf set is large and spans a wide range of climatologies, only deciduous sites were considered.~~

~~Optimisation was performed four randomly selected sets of five sites were. The optimal parameter vectors were then tested at the remaining sites. The results are shown in Fig. **??**.~~

~~The effect of parameter vectors z vectors on the overall model-data fit at each of the sites tested, using the metric described in section 2.5.2. Original default JULES parameters ($*$), site-specific optimal parameters ($*$), and the multisite parameters found by optimising over each set of five sites (,$\bullet * 0 \circ$,,$\bullet * 0 \circ$), denoted set 1, set 2, set 3, set 4respectively. Sites in the training set (filled circles), sites in validation set (open circles).~~

~~The optimised parameter vectors generally perform well, both on the sites used in the training sets and the sites used in the validations sets. Indeed 15/18 of the sites improve no matter which of the optimised parameter vectors are used. The parameter vector optimised over set 3 performs even better than the individual optimisations for some of the sites. JULES performed worse on just two sites (UK-PL3, US-Ha1) using these parameter values compared to the default JULES parameters. These two sites also start off with relatively small errors, so even with the slight increase in errors they are still among the best performing sites in the set. UK-PL3 does not improve with any of the 5-site parameter sets, but observations from this site appears to be somewhat unusual (e.g. it has a very different seasonality to the rest of the sites for this PFT).~~

~~It seems likely that the adding more sites to a multisite optimisation smooths the cost function and makes it less likely for the optimisation to get stuck in local minima. This may be one of the reasons that some of the 5-site optimisation works better than the single-site optimisation for certain sites.~~

~~Overall the results are promising, showing that the optimised parameters, even when calibrated from a small subset of sites, can be generalised over the rest of the set.~~

3.3 ~~New PFT parameter values~~

Optimisations were performed over all available sites for each of the PFTs simultaneously. The optimised model parameters for each of the PFTs are presented in Fig. 3.

For half of the parameters, the ~~original parameter value is found outside the new~~ prior parameter value lies outside the posterior uncertainty bounds. The $\frac{\delta c}{\delta l}$ parameter, which determines the efficiency of rainfall interception by the plant canopy,

does not change much from its original value for any of the PFTs. The uncertainty bounds are relatively tight and symmetrical. The rest of the parameters show more variation. As described in section 2.5.1, the optimal values need not be in the centre of the uncertainty range, the ~~uncertainties~~ PDF can be skewed. Most of the PFTs display high uncertainty in at least one of the ~~the~~ parameters optimised; for the optimised broadleaf set for example, dq_c is highly unconstrained. For C4 grasses, d_r is so unconstrained ~~, even~~ that the optimal value found ~~is outside of~~ lies outside the 80% confidence interval. C3 grasses ~~shows~~ show large uncertainty in n_0 and for shrubs, the parameter with the largest uncertainty is α.

The uncertainties shown in Fig. 3 are one-dimensional marginal distributions. To understand further how the parameters are correlated, consider the two-dimensional representation in Fig. 4. For all of the PFTs, the ~~new~~ posterior parameter uncertainties exclude a large part of the prior ranges. The cloud of plausible points tends to be restrictive and tight for most parameters.

~~The majority of the broadleaf parameters, shown in Fig. 4(a), are highly correlated with each other. The d_r and $\frac{\delta c}{\delta l}$ parameters are the only ones to be uncorrelated with other parameters. Similarly, the needleleaf parameters (Fig. 4(b)) are all highly correlated, either positively or negatively, except for T_{low}, which is completely uncorrelated with any of the other parameters.~~

~~For C3 grasses, (Fig.4(c)), the parameters which show no correlation between themselves and any other parameters are T_{low}~~

Fig. 4 shows clear correlation of some parameters, especially for the tree PFTs. Many of these correlations can be understood in terms of the underlying structure of the JULES model (Section 2.1). For example, the positive correlation of n_0 with f_0, and the negative correlation of n_0 with dq_c, are consistent with adJULES attempting to fit the stomatal conductance g, which controls the transpiration flux from taller vegetation. The stomatal conductance has the approximate form

$$g \approx 1.6 \frac{A}{c_a} \left(\frac{1}{(1-f_0) + f_0 \frac{dq}{dq_c}} \right) \tag{10}$$

if it is assumed that $c_* \ll c_i$ and $\frac{\delta c}{\delta l}$. ~~The d_r parameters shows varying levels of correlations with the other parameters. The rest of the parameters are highly correlated.~~ $c_* \ll c_a$.

~~All the C4 grass parameters (Fig. 4(d)) are completely uncorrelated, with the exception of of the parameter pair d_r and dq_c, which covary. For the shrub optimised parameters (~~The maximum rate of leaf photosynthesis is controlled largely by the leaf nitrogen content n_0, especially in this big-leaf version of JULES (Cox et al., 1999). The best fit parameters for tree PFTs also seem to imply that the second term in the denominator dominates over the first. As a result, maintaining a realistic g value, and therefore a realistic latent heat flux, will require that n_0 and f_0 vary proportionally, and that n_0 and dq_c values are negatively correlated. This is consistent with Fig. 4(~~e)) , n_0 is negatively correlated with f_0 and dq_c. The rest of the parameters are not correlated.~~ a) and (b).

~~The parameter vectors showing the highest correlations belong to the broadleaf and the needleleaf optimisations, for which there are more measurement sites. Such a high correlation between parameters may therefore be related to the number of sites used in the optimisation~~Such correlation of parameters is less obvious for the grass PFTs, because evapotranspiration is controlled less by stomatal conductance and more by the smaller aerodynamic conductances associated with shorter vegetation.

3.2.1 ~~The performance~~ Assessment of ~~the new PFT~~ PFT-specific optimal parameters

The performance of the PFT-specific parameters ~~are~~ is now compared to the default JULES values and ~~the results of the model optimised~~ to the parameters found by optimising independently at each measurement site. ~~This is shown in~~ For each site, the fractional error in both the calibration year and the validation year is displayed Fig. 5. ~~The lower the error, the better the model fits the observations, and so the better performing the parameter vector is.~~

~~All sites are improved using the locally optimised parameter vectors.~~

By definition, the fractional error in calibration years decreases when moving from default to site-specific optimal parameters in the calibration years. Remarkably, the site-specific optimal parameters also improve the model-data fit in validation years for 59/64 of sites (over 92%). Similarly, the PFT-specific optimal parameter vector improves the fit (in both calibration and validation years) for 85% of the sites; 75/79 sites for the calibration years and 55/64 sites for the validation years.

Consider first the broadleaf sites (Fig. 5, top two rows). For the majority of sites ~~, this decrease in error is substantial. Only the outliers, which start with large initial errors, and the C4 grass site show little improvement. For the C4 grass sites, the initial error is low due to the fact these sites have incomplete data.~~

~~The new~~ displayed in the top broadleaf panel, the reduction in fractional error in moving from default to site-specific optimal parameters is substantial and sometimes as much as a factor of 2. In the calibration year, the PFT-specific ~~parameter vectors improve JULES performance over 92% of the sites used in this study. The new broadleaf parameter vector significantly improves 25 of the 28 broadleaf sites , and a further two~~ optimal parameter vector improves 26 of the 27 broadleaf sites shown although one of the sites~~give errors similar to when the default parameters are used. Only~~ , IT-Lec, the fit shows no change. The improvement is typically about half as good (on a log scale) as the improvement using the site-specific optimal parameters. In other words, the reduction in fractional error moving from default to PFT-specific optimal parameters is sometimes as much as a factor of $\sqrt{2}$. Amongst broadleaf sites, only UK-PL3 gets notably worse. ~~Considering this site more closely, it can be seen to behave~~ Investigation shows that this site behaves differently from the rest of the sites in the set, both in the magnitude of the fluxes and seasonality. This UK site is in the Pang/Lambourn catchment, which has chalk soil with macropores that permit significant lateral subsurface flows of soil moisture. These horizontal flows cannot be captured in a model like JULES which is essentially one-dimensional in the vertical below the soil surface.

~~The needleleaf sitesimprove greatly when using the new needleleaf parameter values, with~~ Similar levels of fit and error reduction can be seen in the validation years in the Broadleaf set. Only IT-Col and US-MMS show no improvement, the PFT-specific optimal parameter vector does not worsen the fit at these locations. For AU-Tum, the PFT-specific parameter vector outperforms the site-specific vector. This illustrates that the PFT-specific vector can be robust, whereas the locally-optimised vectors might over-tune to the specific behaviour of the calibration year.

Results are similar for the Needleleaf sites, the majority of the sites show noticeable improvements in both the calibration and validation years when using site-specific optimal parameter vectors. For over a third of the sites ~~nearly performing as well as the single-site optimisations. The only site this new parameter vector~~ in this PFT, the improvement when using the PFT-specific parameter vector is similar to that obtained with the site-specific parameter vector. This illustrates that these sites

fit well together as a single PFT. For these sites, the PFT-specific vector sometimes outperforms the site-specific vector on the validation years. Some sites in the Needleleaf PFT remain unchanged regardless of the parameter vector used. Anomalous sites that should be noted are CA-Qcu, CA-SF3 and US-Blo. The CA-Qcu site is the only one in this PFT that does not improve ~~is CA-Qcu, which was one of the sites with the lowest initial error. As with averaging, sites with the best fitmay have to be sacrificed to achieve a generic parameter set across the PFT.~~ when using the PFT-specific vector, for either the calibration or validation years. This site has a lower annual cycle of GPP than the rest in this set. The CA-SF3 site improves when using the site-specific parameter vector in the validation year, but not using the PFT-specific vector. The ~~new error still remains relatively low.~~ US-Blo site improves in the calibration year, but when confronted with the validation year, both the site-specific vector and PFT-specific vector worsen the fit. This validation year has unusually high LE which might be causing this discrepancy.

~~For~~ For some sites, (e.g. US-Blo and BW-Ma1), the PFT-specific optimum outperforms the site-specific optimum in the calibration year. This phenomenon was also noted by Kuppel et al. (2014) , who suggest that the added constraints placed on the parameters by increasing the number of sites causes the cost function to become 'smoother'. This may render the optimisation scheme less likely to become trapped in local minima.

The last panel of Fig. 5 shows the C3 grass ~~site, there is a reduction in error for 9 of the 11 sites when using the new parameters. The last two sitesin the set act similarly to when the default parameters were used.~~ sites, the C4 grass sites and the Shrub sites. For the C3 grass sites, the majority of the validation years have a better fit with the PFT-specific parameter vector than with site-specific parameter vector. This suggests that the seasonal cycle differs over the different years at these sites. For the C4 grass sites, which started with relatively ~~low~~ high errors, the new parameter ~~vector improves~~ vectors improve the sites slightly ~~. However, the~~ for the calibration year but hardly at all for the validation year. This set of two sites is too small to draw any proper conclusion about the C4 grass parameters. There is a clear need for more data from C4 grass sites. Finally, the Shrubs can be seen to improve for all the sites.

~~In the case of the outliers, the new~~ For the Shrub sites, both the site-specific and the PFT-specific ~~parameter vectors improves JULES performance even relative to the single-site optimisations. A further 9 sitesof the whole set of sites improve to a greater extend than the local optimisations.~~ provide a better fit of the model to the observations of the calibration year. The improvement is minor for these sites, except for CA-Mer which halves its fractional error. When confronted with observations from the validation years, the model also improves the fit of these sites for both site-specific and PFT-specific parameters (with the exception of US-Los, where the site-specific optimal vector increases error but the PFT-specific vector reduces it). This is another example of the PFT-specific parameter vector being more robust.

4 Conclusions

~~adJULES~~

This study introduces the adJULES system, which has been developed to tune the internal parameters of the JULES land surface model. adJULES enables objective calibration of JULES against observational data, providing best fit internal parameters and ~~the~~ associated uncertainty ranges. ~~The adJULES fits of JULES against~~

For individual FLUXNET sites~~show significant improvements in the performance of JULES compared to default parameters, typically in both the simulation of~~ , adJULES has the ability to find local (site-specific) optimal parameter vectors that significantly improve the performance of the JULES model compared to runs generated using the default parameters. The data streams used in the calibration, LE and GPP~~. All~~ , are both modelled more accurately with the optimal parameter vectors, with the GPP flux improving the most. The greater improvement in the GPP flux is due to the fact that the parameters considered in this study are mainly related to photosynthesis. For the LE flux to improve more significantly, more water and energy related parameters would need to be considered in the optimisation.

When optimised locally to find site-specific parameters, all of the sites in this study improve ~~when optimised locally , with the GPP flux improving most significantly.~~ the model-data fit for the calibration year. In addition, when confronted with independent data from a validation year, the locally optimised parameter vectors decreased the error in model-data fit for 92% of the sites. This validation of the site-specific parameter vectors is promising, and suggests that the adJULES system is robust. It also gives confidence that the parameter vectors found can be generalised over different locations.

~~The study is partially motivated~~ This study is motivated partly by the desire to improve the performance of JULES within the Hadley Centre's Earth System Models, which means needing to find best fit parameters for a relatively small number of PFTs. ~~This is achieved by~~ The adJULES system has the ability to calibrate multiple locations simultaneously in order to find best-fit parameters. This 'multisite' optimisation is a relatively new feature in terrestrial data assimilation. By classifying the FLUXNET sites into groups dominated by each JULES PFT (BT, NT, C3G, C4G, Sh)~~and using adJULES~~ , adJULES was used to find the ~~best fit parametersfor each of these PFT groupings.~~ optimal PFT-specific parameters.

Although the PFT-specific ~~parameters inevitably do not~~ optimal parameters do not always fit the data as well as site-specific optimal parameters, they still offer significant improvements ~~of~~ over the default JULES parameters. For over ~~90~~85% of the sites, ~~the new~~ PFT-specific ~~parameters are~~ optimal parameters perform better than default parameters ~~giving closer model-data fit.~~

~~For some PFTs (notably C4G and Shrubs)there are insufficient FLUXNET sites to determine optimal parameters satisfactorily. Additional data and sites for these PFTs are therefore urgently required~~when confronted with independent validation data. For 50% of the sites, the PFT-specific optimal parameters perform at least as well as site-specific optimal parameters. This implies that the multisite methodology is less susceptible to over-tuning, both in terms of variablity across sites (e.g. different overground biomass and tree ranges), and in terms of variability through time (e.g. unusually high rainfall in the calibration year).

The PFT-specific parameters found in this study represent a significant improvement on the default ones. The fact that such parameters could be found implies robust parameterisations independent of geography. This supports the idea that it is possible to represent global vegetation with a relatively small number of PFTs.

A successful and robust multisite optimisation assumes that sites can be grouped and parameter values can apply to several sites at once. Whilst the PFT-specific parameters show great improvement, agreeing with the use of five PFTs in JULES, it would be possible to rethink the PFT definitions and group sites differently. This could be done either by looking more closely

at the site specifics detailed in the FLUXNET database, or by considering single-site optimisations and performing a cluster analysis in parameter space to identify PFTs empirically.

It is ~~, however ,~~ however clear that there are some limitations to the success of the optimisation results. ~~Certain~~ Some sites still show significant differences between model output and observations. ~~These issues indicate~~ This suggests that improvement to model physics may be necessary in order to produce better model output. This is because adJULES produces the (locally) best possible fit to observations, given the existing model physics and the prescribed driving data. If the fit is still inadequate, ~~it is down~~ this may be due to the model and data themselves, rather than parameter values. adJULES ~~therefore also enables~~ can therefore be used in the identification of model structural errors~~to be identified.~~

~~A successful and robust multisite optimisation assumes that sites can be grouped and parameter values can apply to several sites at once. Whilst the PFT generic parameters show great improvement, agreeing with the general 5-PFTs definition found in JULES, there is a possibility to rethink the PFT definitions and group sites differently. This could be done either by looking more closely at the site specifics detailed by the FLUXNET database, or by considering the single-site optimisations and performing a cluster analysis to empirically identify PFTs~~ . Another reason for inadequate fit may be due to the method used. A limitation of gradient descent methods, such as the optimisation scheme used in this study, is that sometimes a local minimum is found instead of the global minimum. However, as discussed in section **??**, the fact that the cost function becomes smoother with additional sites may be a solution to becoming trapped in local minima (Kuppel et al., 2014) . Alternative methods, including ensemble methods, could avoid this issue, but are more computationally costly. For some PFTs (notably C4G and Shrubs) there are insufficient FLUXNET sites to determine optimal parameters satisfactorily. Additional data and sites for these PFTs are therefore urgently required.

Code availability

The source code of the adJULES data assimilation system is available at http://adjules.ex.ac.uk/. The JULES land surface model is freely available to any researcher for non-commercial use. Version 2.2 used in this study can be requested at jules.jchmr.org. The main documentation for the JULES system can also be found at this site. The adjoint of the JULES model has been generated using commercial software TAF (sect. 2.2.1). For licensing reasons, the recalculation of the adjoint following code changes can ~~only be done by us here at~~ be done only by the authors at the University of Exeter.

Appendix A

Appendix B

Acknowledgements. This work was supported by the UK Natural Environment Research Council (NERC) through the National Centre for Earth Observation (NCEO).

This study used eddy-covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC.

~~Further acknowledgements goes to the financial support to the~~ Support for eddy covariance data harmonisation was provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California - Berkeley, University of Virginia.

~~Thanks~~ The authors are grateful to T. Kaminski and R. Giering from FastOpt for their contribution to the development of the adjoint model~~. Finally, thanks~~, and to M. Groenendijk, A. Harper, and the UK Met Office for processing and sharing their data.

The authors are particularly grateful to two anonymous referees for their thoughtful and constructive reviews which greatly improved this manuscript.

References

Arora, V. and Boer, G.: A parameterization of leaf phenology for the terrestrial ecosystem component of climate models, Global Change Biology, 11, 39–59, 2005.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., et al.: FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, Bulletin of the American Meteorological Society, 82, 2415–2434, 2001.

Bartholomew-Biggs, M., Brown, S., Christianson, B., and Dixon, L.: Automatic differentiation of algorithms, Journal of Computational and Applied Mathematics, 124, 171 – 190, doi:http://dx.doi.org/10.1016/S0377-0427(00)00422-2, http://www.sciencedirect.com/science/article/pii/S0377042700004222, numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations, 2000.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), Model description - Part 1: Energy and water fluxes, Geoscientific Model Development Discussions, 4, 595–640, doi:10.5194/gmdd-4-595-2011, 2011.

Blyth, E., Lloyd, A., Gash, J., Pryor, M., Weedon, G., and Shuttleworth, J.: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data, Journal of Hydrometeorology, 2010.

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N.: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, Nature, 484, 228–232, http://dx.doi.org/10.1038/nature10946, 2012.

Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, Training, pp. 1–59, 1999.

Brovkin, V., Boysen, L., Raddatz, T., Gayler, V., Loew, A., and Claussen, M.: Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, Journal of Advances in Modeling Earth Systems, 5, 48–57, doi:10.1029/2012MS000169, http://dx.doi.org/10.1029/2012MS000169, 2013.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C.: A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific Computing, 16, 1190–1208, 1995.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), Model description - Part 2: Carbon fluxes and vegetation, Geoscientific Model Development Discussions, 4, 641–688, doi:10.5194/gmdd-4-641-2011, http://www.geosci-model-dev-discuss.net/4/641/2011/, 2011.

Cox, P., Huntingford, C., and Harding, R.: A canopy conductance and photosynthesis model for use in a GCM land surface scheme, Journal of Hydrology, 212, 79–94, 1998.

Cox, P., Betts, R., Jones, C., Spall, S., and Totterdell, I.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, 408, 184–187, 2000.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Climate Dynamics, 15, 183–203, 1999.

Friedlingstein, P., Bopp, L., Ciais, P., Dufresne, J.-L., Fairhead, L., LeTreut, H., Monfray, P., and Orr, J.: Positive feedback between future climate change and the carbon cycle, Geophysical Research Letters, 28, 1543–1546, doi:10.1029/2000GL012015, http://dx.doi.org/10.1029/2000GL012015, 2001.

Friedlingstein, P., Bopp, L., Rayner, P., Cox, P. M., Betts, R., Jones, C., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., et al.: Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison, Journal of Climate, 19, 3337–3353, 2006.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, Journal of Climate, 27, 511–526, doi:10.1175/JCLI-D-12-00579.1, http://dx.doi.org/10.1175/JCLI-D-12-00579.1, 2013.

Geman, S. and Geman, D.: Pattern Analysis and Machine Intelligence, IEEE Transactions on, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, 1984.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T.: mvtnorm: Multivariate Normal and t Distributions, 2015.

Giering, R., Kaminski, T., and Slawig, T.: Generating efficient derivative code with TAF: Adjoint and tangent linear Euler flow around an airfoil, Future Generation Computer Systems, 21, 1345 – 1355, doi:http://dx.doi.org/10.1016/j.future.2004.11.003, http://www.sciencedirect.com/science/article/pii/S0167739X04001785, 2005.

Groenendijk, M., Dolman, A., van der Molen, M., R. Luning, A. A., Delpierre, N., Gash, J., Lindorth, A., Richardson, A., Verbeeck, H., and Wohlfahrt, G.: Assessing parameter variability in a photosynthesis model within and between plany functional types using global FLUXNET eddy covariance data, Agricultural and Forest Meteorology, 2010.

Jones, G. S., Stott, P. A., and Christidis, N.: Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations, Journal of Geophysical Research: Atmospheres, 118, 4001–4024, doi:10.1002/jgrd.50239, http://dx.doi.org/10.1002/jgrd.50239, 2013.

Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P. J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E. N., Köstler, C., Mathieu, P.-P., Pinty, B., Reick, C. H., Rödenbeck, C., Schnur, R., Scipal, K., Sebald, C., Stacke, T., van Scheltinga, A. T., Vossbeck, M., Widmann, H., and Ziehn, T.: The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges, Journal of Geophysical Research: Biogeosciences, 118, 1414–1426, doi:10.1002/jgrg.20118, http://dx.doi.org/10.1002/jgrg.20118, 2013.

Knorr, W. and Kattge, J.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, Global Change Biology, 11, 1333–1351, 2005.

Kuppel, S., Peylin, P., F.Chevallier, Bacour, C., Maignan, F., and Richardson, A.: Constraining a global ecosystem model with multi-site eddy covariance data, Biogeosciences, 2012.

Kuppel, S., Peylin, P., Maignan, F., F.Chevallier, Kiely, G., Montagnani, L., and Cescatti, A.: Mode-data fusion across ecosystems: from multi-site optimizations to global simulations, Geosci. Model Dev. Discuss., 7, 2961–3011, 2014.

Medvigy, D. and Moorcroft, P. R.: Predicting ecosystem dynamics at regional scales: an evaluation of a terrestrial biosphere model for the forests of northeastern North America, Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367, 222–235, doi:10.1098/rstb.2011.0253, 2011.

Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y., and Moorcroft, P. R.: Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2, Journal of Geophysical Research: Biogeosciences, 114, n/a–n/a, doi:10.1029/2008JG000812, http://dx.doi.org/10.1029/2008JG000812, g01002, 2009.

Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., and Cox, P. M.: Impact of changes in diffuse radiation on the global land carbon sink, Nature, 458, 1014–1017, 2009.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., et al.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571–583, 2006.

Peng, C., Guilot, J., Wu, H., Jiang, H., and Luo, Y.: Integrating models with data in ecology and palaeoecology advances towards a model-data fusion approach, Ecology Letters, 14, 522–536, 2011.

Pitman, A.: The evolution of, and revolution in, land surface schemes designed for climate models, International Journal of Climatology, 23, 479–510, 2003.

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2015.

Raupach, M., Rayner, P., Barrett, D., Defries, R., Heimann, M., Ojima, D., Quegan, S., and Schmullius, C.: Model-data synthesis in terrestrial carbon observation methods, data requirements and data uncertainty specifications, Global Change Biology, 11, 378–297, 2005.

Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global Biogeochem. Cycles, 19, 2005.

Reichstein, M., Tenhunen, J., Roupsard, O., Ourcival, J.-M., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G., and Valentini, R.: Inverse modeling of seasonal drought effects on canopy CO2/H2O exchange in three Mediterranean ecosystems, Journal of Geophysical Research: Atmospheres (1984–2012), 108, 2003.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biology, 11, 1424–1439, 2005.

Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, Global biochemical cycles, 21, 2007.

Sellers, P., Dickinson, R., Randall, D., Betts, A., Hall, F., Berry, J., Collatz, G., Denning, A., Mooney, H., Nobre, C., et al.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, Science, 275, 502, 1997.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., eds.: Summary for Policymakers, book section SPM, p. 1–30, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.004, www.climatechange2013.org, 2013.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., and Mahowald, N. M.: Influence of carbon-nitrogen cycle coupling on land model response to CO2 fertilization and climate variability, Global Biogeochemical Cycles, 21, n/a–n/a, doi:10.1029/2006GB002868, http://dx.doi.org/10.1029/2006GB002868, gB4018, 2007.

Thum, T., Aalto, T., Laurila, T., Aurela, M., Lindroth, A., and Vesala, T.: Assessing seasonality of biochemical CO 2 exchange model parameters from micrometeorological flux observations at boreal coniferous forest, Biogeosciences, 5, 1625–1639, 2008.

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., and Ciais, P.: Seasonal patterns of CO2 fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model, Journal of Geophysical Research, 116, 2011.

Wang, Y.-P., Leuning, R., Cleugh, H. A., and Coppin, P. A.: Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful?, Global Change Biology, 7, 495–510, 2001.

Wang, Y. P., Baldocchi, D., Leuning, R., Falge, E., and Vesala, T.: Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight Fluxnet sites, Global Change Biology, 13, 652–670, 2007.

Williams, M., Richardson, A., Reichstein, M., Stoy, P., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C., and Wang, Y.-P.: Improving land surface models with FLUXNET data, Biogeosciences, 2009.

Xiao, J., Davis, K. J., Urban, N. M., Keller, K., and Saliendra, N. Z.: Upscaling Carbon Fluxes from Towers to the Regional Scale: Influence of Parameter Variability and Land Cover Representation on Regional Flux Estomates, Journal of Geophysical Research, 2011.

Zaehle, S., Friedlingstein, P., and Friend, A. D.: Terrestrial nitrogen feedbacks may accelerate future climate change, Geophysical Research Letters, 37, n/a–n/a, doi:10.1029/2009GL041345, http://dx.doi.org/10.1029/2009GL041345, l01401, 2010.

Table 1. Parameters in optimisation vector, with descriptions.

Symbol	Name in code	Description	Units
n_0	`nl0`	Top leaf nitrogen concentration	kg N (kg C)$^{-1}$
f_0	`f0`	Maximum ratio of internal to external CO_2	-
d_r	`rootd_ft`	Root depth	m
α	`alpha`	Quantum efficiency	mol CO_2 per mol PAR photons
$\frac{\delta c}{\delta l}$	`dcatch_dlai`	Rate of change of canopy interception capacity with LAI	kg m^{-2}
T_{low}	`tlow`	Lower temperature for photosynthesis	°C
T_{upp}	`tupp`	Upper temperature for photosynthesis	°C
dq_c	`dqcrit`	Humidity deficit at which stomata close	~~kg kg^{-1}~~ kg kg^{-1}

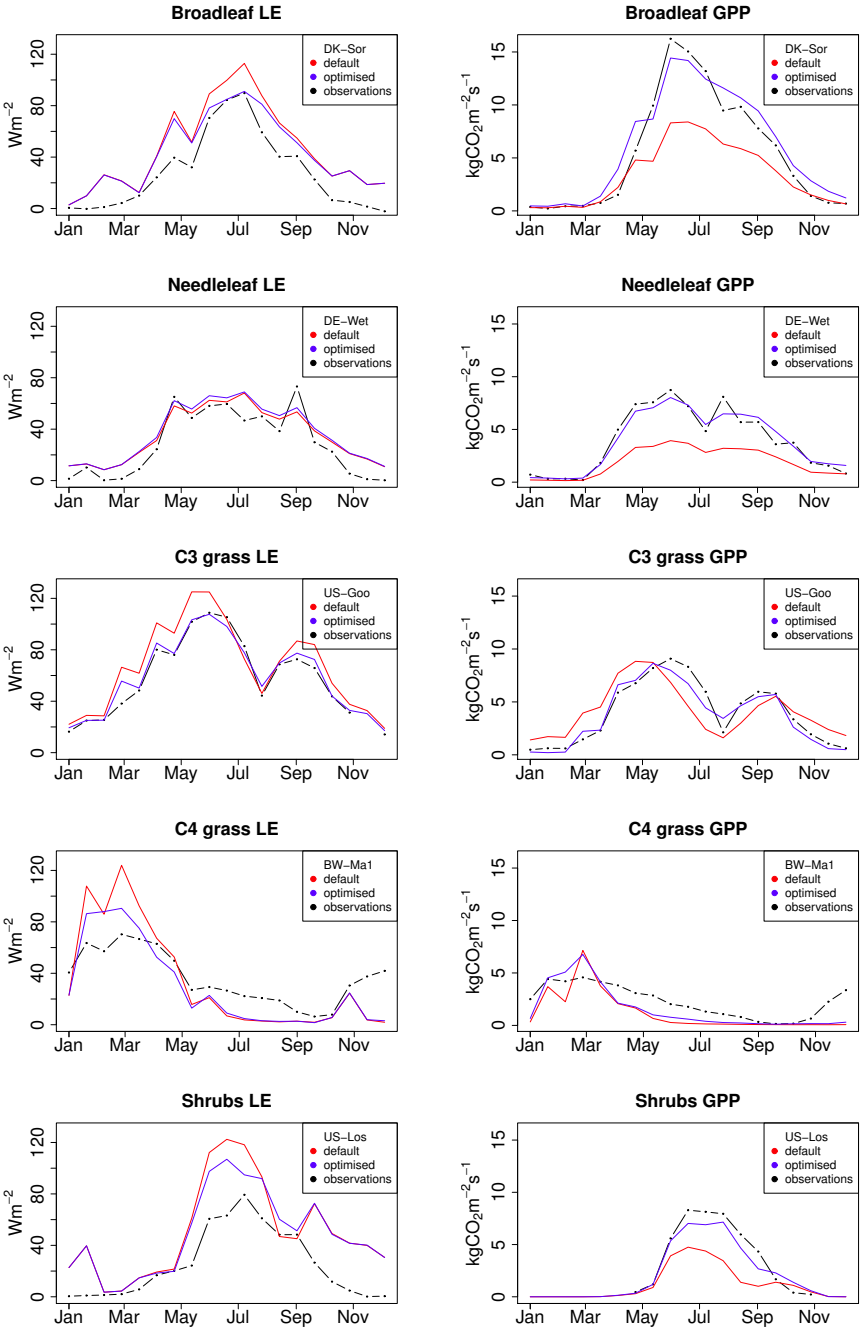


Figure 2. Time-series plots for illustrative site-specific validations showing LE (left) and GPP (right) for ~~a single site in~~ each of the different PFTs. Observations (black) are compared to ~~the original~~ JULES runs using default parameters (red) and ~~the runs using the~~ site-specific optimal ~~parameters found at each individual site locally~~ paremeters (blue).

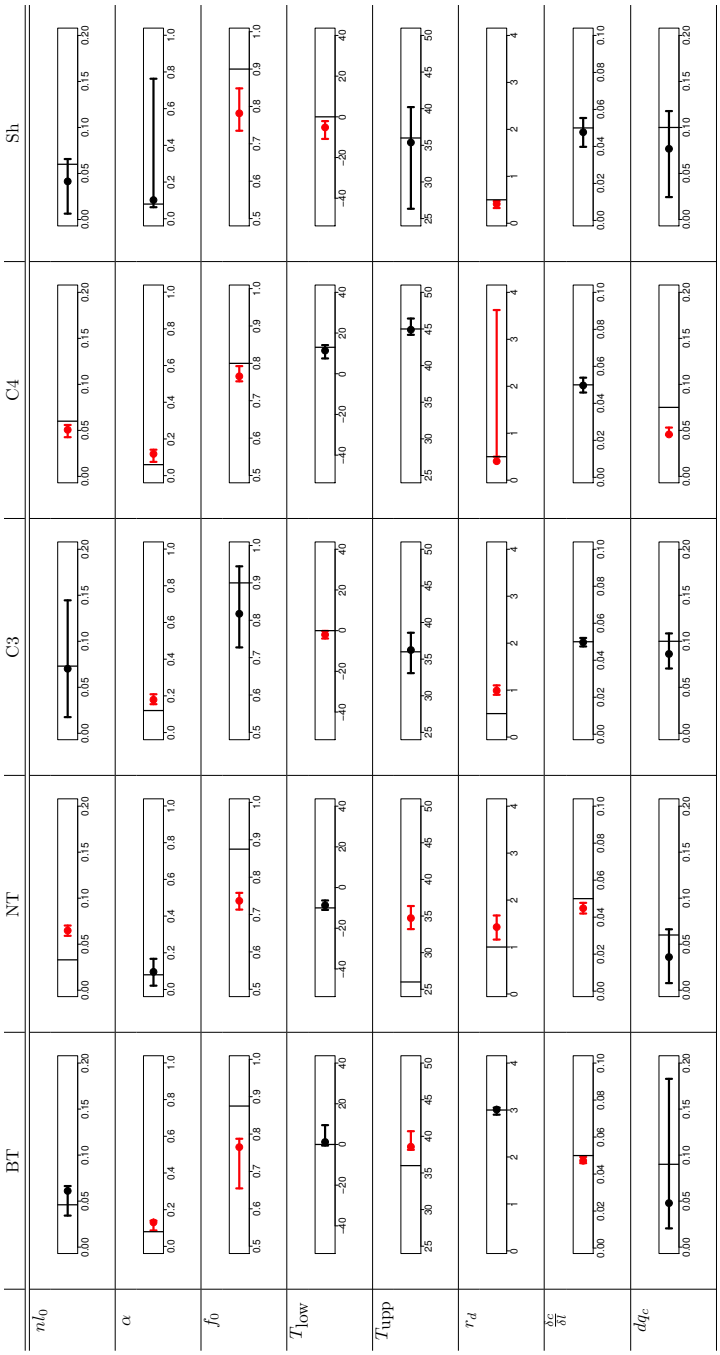
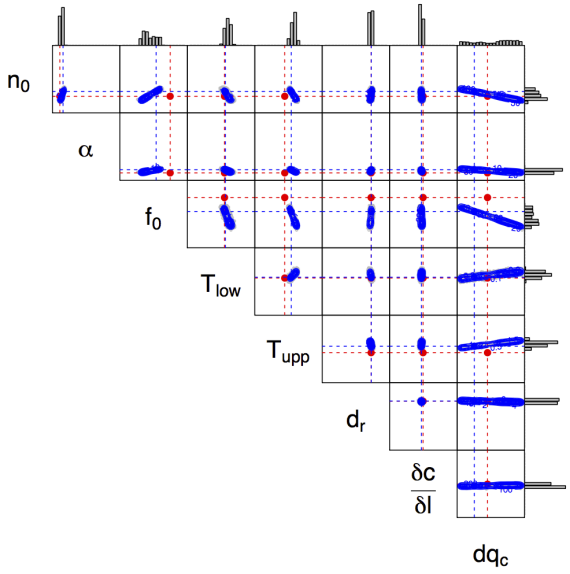
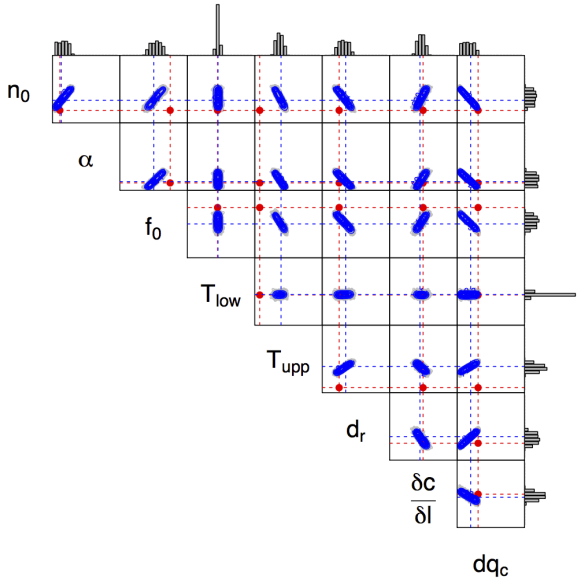


Figure 3. Summary of PFT-specific optimal JULES parameters ~~optimised~~ found in this study (Table 1). ~~Initial values for each PFT are given, and below in bold are optimised values.~~ The error bars show the uncertainty ranges given as ~~a~~ an 80% ~~confidence~~ quantile interval. The range of ~~the~~ each box is the ~~allowed~~ prior range of the parameters~~were allowed to vary over~~. Highlighted in red are the error bars for which the prior values (~~dotted~~ vertical line) are found outside the ~~new~~ posterior uncertainty bounds. A numerical version of this ~~table exists~~ figure is given in ~~Appendix B~~Table B1.

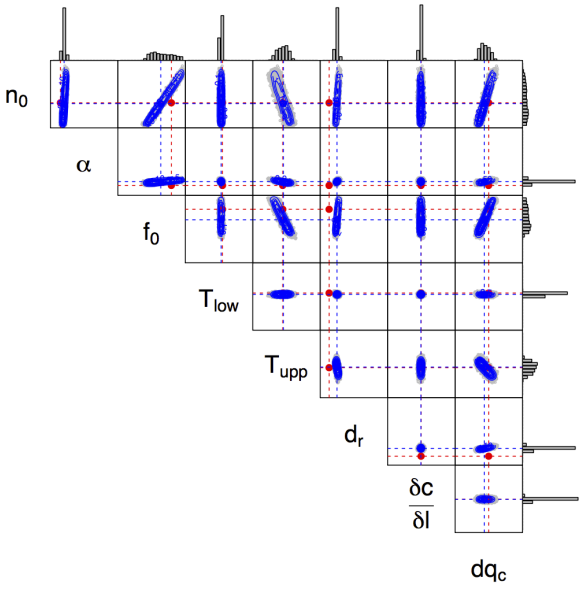
Figure 4. The correlations between ~~the optimised~~ parameters ~~found at each of the PFTs~~for PFT-specific parameter optimisations. Each subfigure shows a two-dimensional correlation map~~and with each subfigure~~, within which each box is a 2-D marginal plot. Bar graphs show 1-D marginal distributions ~~of the~~ for individual parameters. The dimensions of the boxes represent the ~~allowed~~ prior range of each parameter. Red points/dashed lines represent initial parameter values. Blue points/dashed lines represent ~~the new~~ optimised parameter values~~and the blue~~ . Blue contours ~~define~~ illustrate the ~~cloud of possible parameter values~~posterior distribution.



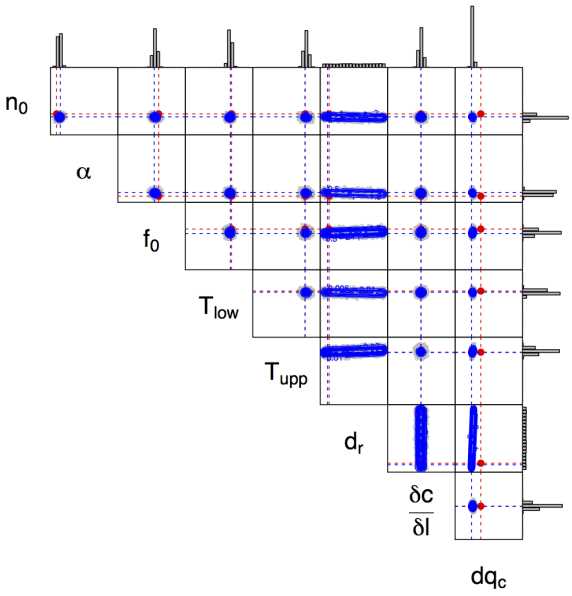
(a) Broadleaf



(b) Needleleaf

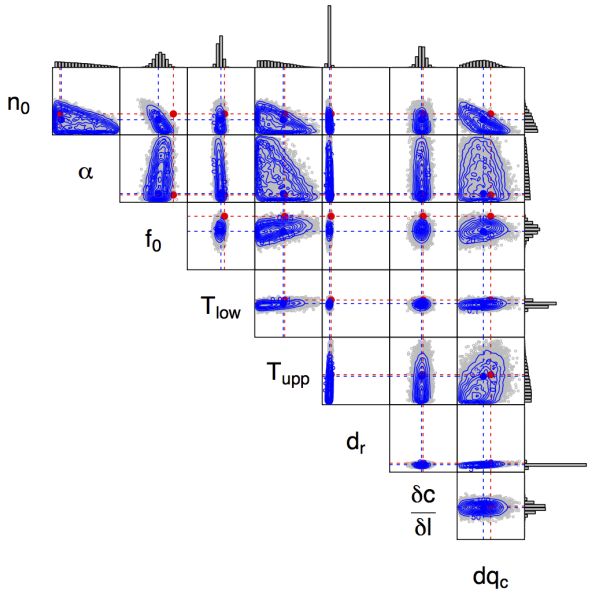


(c) C3 grasses



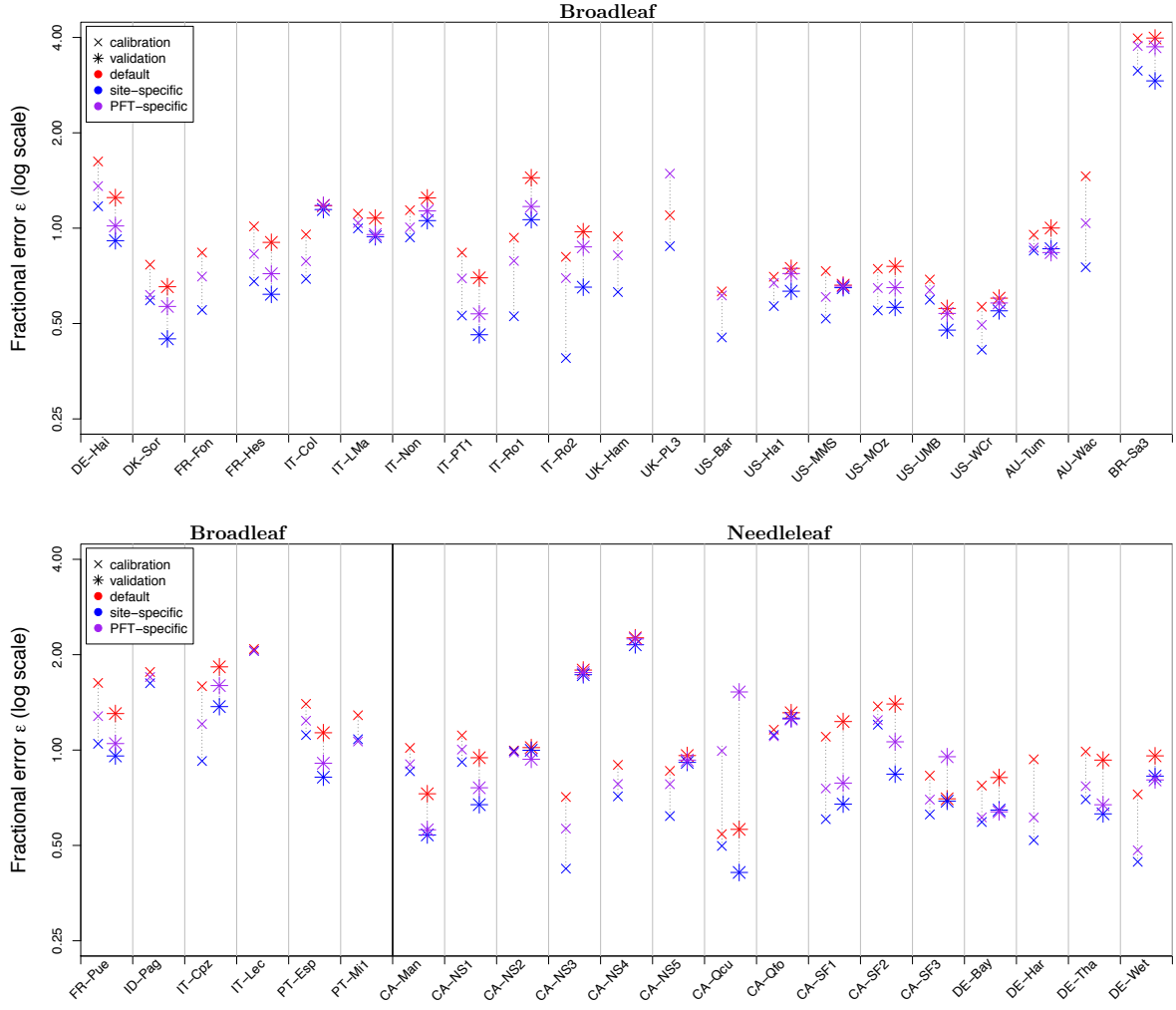
(d) C4 grasses

Figure 4. continued



(e) Shrubs

Figure 4. continued



~~The effect of different parameter vectors on the overall model-data fit at each of the site tested, using the metric described in section 2.5.2. The three z vectors tested: the original default JULES parameters ($*$), the parameter vector found by optimising at the individual sites ($*$), the new PFT-specific parameter vector found by optimising over the given PFT ($\bullet$). Outliers with very large initial errors have been removed from the plot shown (Broadleaf: BR-Sa3, IT-Lec, Needleleaf: CA-NS2, SE-Sk2, IT-Yat, IT-Lav).~~

Figure 5. Calibration and validation of site-specific and PFT-specific parameter optimisation at FLUXNET sites, using the metric described in section 2.5.2. Fractional error shown for: default JULES parameters ($*$), site-specific optimal parameters ($*$), PFT-specific optimal parameters ($\bullet$). Results are shown both for the calibration year ($\times$, on left) and for the validation year ($*$, on right). No validation year was available for some sites (Broadleaf: FR-Fon, UK-Ham, UK-PL3, US-Bar, ID-Pag, IT-Lec, PT-Mi1, Needleleaf: SE-Sk2, UK-Gri, US-Me4, US-SP1, Shrubs: DE-Gri, DK-Lva, PL-wet). Sites with very large initial errors have been removed from the plot (Broadleaf: BR-Sa1, Shrubs: IT-Pia).

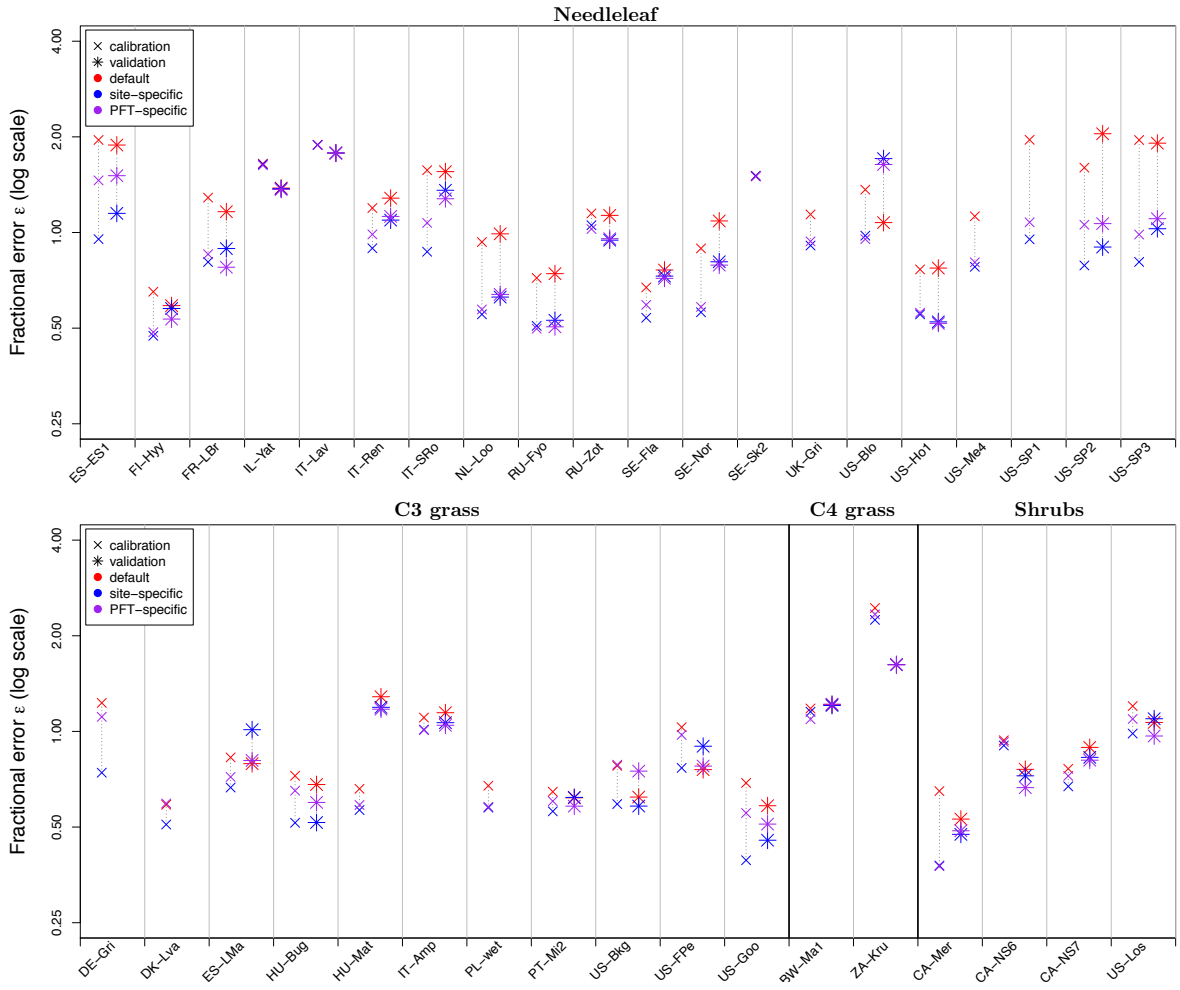


Figure 5. continued.

Table A1. ~~Sites~~ FLUXNET sites used in this study, ~~the name code is made from the~~ labelled by a country code (first two letters) and site name (last three letters). The period corresponds to the available years of data for each of the sites.

Site	Period	~~Experiment~~ Calibration Year	Validation Year	Latitude	Longitude
Broadleaf sites (BT)					
DE-Hai	(2000,2006)	2005	~~51.0793~~ 2004	51.079	10.452
DK-Sor	(1996,2006)	2006	~~55.4869~~ 2004	~~11.6458~~55.487	11.646
FR-Fon	(2005,2006)	2006	~~48.4763~~ -	~~2.78015~~48.476	2.780
FR-Hes	(1997,2006)	2003	~~48.6742~~ 1998	~~7.06462~~ 48.674	7.065
IT-Col	(1996,2006)	2005	~~41.8494~~ 2001	~~13.5881~~ 41.849	13.588
IT-LMa	(2003,2006)	2006	~~45.5813~~ 2004	~~7.15463~~ 45.581	7.155
IT-Non	(2001,2006)	2002	~~44.6898~~ 2003	~~11.0887~~ 44.690	11.089
IT-PT1	(2002,2004)	2003	~~45.2009~~ 2004	~~9.06104~~ 45.201	9.061
IT-Ro1	(2000,2006)	2006	~~42.4081~~ 2005	~~11.93~~ 42.408	11.930
IT-Ro2	(2002,2006)	2004	~~42.3903~~ 2006	~~11.9209~~ 42.390	11.921
UK-Ham	(2004,2005)	2005	~~51.1208~~ -	~~-0.86083~~ 51.121	-0.861
UK-PL3	(2005,2006)	2006	~~51.45~~ -	~~-1.26667~~ 51.450	-1.267
US-Bar	(2004,2005)	2005	~~44.0646~~ -	~~-71.28808~~ 44.065	-71.288
US-Ha1	(1991,2006)	1996	~~42.5378~~ 1998	~~-72.1715~~ 42.538	-72.171
US-MMS	(1999,2005)	2002	~~39.3231~~ 2003	~~-86.4131~~ 39.323	-86.413
US-MOz	(2004,2006)	2006	~~38.7441~~ 2005	~~-92.2~~ 38.744	-92.200
US-UMB	(1999,2003)	2003	~~45.5598~~ 2002	~~-84.7138~~ 45.560	-84.714
US-WCr	(1999,2006)	2005	~~45.8059~~ 2000	~~-90.0799~~ 45.806	-90.080
AU-Tum	(2001,2006)	2003	~~-35.6557~~ 2005	-35.656	148.152
AU-Wac	(2005,2007)	2006	-	-37.429	145.187
BR-Sa1	(2002,2004)	2003	~~-2.85667~~ 2004	~~-54.9589~~ -2.857	-54.959
BR-Sa3	(2000,2003)	2002	~~-3.01803~~ 2003	~~-54.9714~~ -3.018	-54.971
FR-Pue	(2000,2006)	2006	~~43.7414~~ 2005	~~3.59583~~ 43.741	3.596
ID-Pag	(2002,2003)	2003	-	2.345	114.036
IT-Cpz	(1997,2006)	2004	~~41.7052~~ 2006	~~12.3761~~ 41.705	12.376
IT-Lec	(2005,2006)	2006	~~43.3046~~ -	~~11.2706~~ 43.305	11.271
PT-Esp	(2002,2004)	2004	~~38.6394~~ 2003	~~-8.6018~~ 38.639	-8.602
PT-Mi1	(2003,2005)	2005	~~38.5407~~ -	~~-8.0004~~ 38.541	-8.000
C3 grasses sites (C3G)					
DE-Gri	(2005,2006)	2006	~~50.9495~~ -	~~13.5125~~ 50.950	13.512
DK-Lva	(2005,2006)	2006	~~55.6833~~ -	~~12.0833~~ 55.683	12.083
ES-LMa	(2004,2006)	2006	~~39.9415~~ 2005	~~-5.77336~~ 39.941	-5.773
HU-Bug	(2002,2006)	2006	~~46.6911~~ 2005	~~19.6013~~ 46.691	19.601
HU-Mat	(2004,2006)	2006	~~47.8469~~ 2005	47.847	19.726
IT-Amp	(2002,2006)	2006	~~41.9041~~ 2005	~~13.6052~~ 41.904	13.605
PL-wet	(2004,2005)	2005	~~52.7622~~ -	~~16.3094~~ 52.762	16.309
PT-Mi2	(2004,2006)	2006	~~38.4765~~ 2005	~~-8.02455~~ 38.477	-8.025
US-Bkg	(2004,2006)	2006	~~44.3453~~ 2005	~~-96.8362~~ 44.345	-96.836
US-FPe	(2000,2006)	2002	~~48.3079~~ 2004	48.308	-105.101
US-Goo	(2002,2006)	2006	~~34.25~~ 2004	~~-89.97~~ 34.250	-89.970

Table A1. continued

Site	Period	~~Experiment~~ Calibration Year	Validation Year	Latitude	Longitude
Needleleaf sites (NT)					
CA-Man	(1997,2003)	2001	~~55.8796~~ 2002	~~-98.4808~~ 55.880	-98.481
CA-NS1	(2002,2005)	2004	~~55.8792~~ 2003	~~-98.4839~~ 55.879	-98.484
CA-NS2	(2001,2005)	2002	~~55.9058~~ 2004	~~-98.5247~~ 55.906	-98.525
CA-NS3	(2001,2005)	2004	~~55.9117~~ 2002	~~-98.3822~~ 55.912	-98.382
CA-NS4	(2002,2004)	2004	~~55.9117~~ 2003	~~-98.3822~~ 55.912	-98.382
CA-NS5	(2001,2005)	2004	~~55.8631~~ 2002	55.863	-98.485
CA-Qcu	(2001,2006)	2005	~~49.2671~~ 2006	~~-74.0365~~ 49.267	-74.037
CA-Qfo	(2003,2006)	2006	~~49.6925~~ 2005	~~-74.3421~~ 49.693	-74.342
CA-SF1	(2003,2005)	2004	2005	54.485	-105.818
CA-SF2	(2003,2005)	2004	~~54.2539~~ 2005	54.254	-105.878
CA-SF3	(2003,2005)	2005	~~54.0916~~ 2004	54.092	-106.005
DE-Bay	(1996,1999)	1999	~~50.1419~~ 1998	~~11.8669~~ 50.142	11.867
DE-Har	(2005,2006)	2006	~~47.9344~~	47.934	7.601
DE-Tha	(1996,2006)	2005	~~50.9636~~ 2004	~~13.5669~~ 50.964	13.567
DE-Wet	(2002,2006)	2006	~~50.4535~~ 2004	~~11.4575~~ 50.453	11.457
ES-ES1	(1999,2006)	2005	2000	39.346	~~-0.31881~~ -0.319
FI-Hyy	(1996,2006)	2006	~~61.8474~~ 2004	~~24.2948~~61.847	24.295
FR-LBr	(2003,2006)	2006	~~44.7171~~ 2005	~~-0.7693~~ 44.717	-0.769
IL-Yat	(2001,2006)	2005	2006	31.345	~~35.0515~~ 35.051
IT-Lav	(2000,2002)	2001	~~45.9553~~ 2002	~~11.2812~~ 45.955	11.281
IT-Ren	(1999,2006)	2005	~~46.5878~~ 2006	~~11.4347~~ 46.588	11.435
IT-SRo	(1999,2006)	2006	~~43.72786~~ 2005	~~10.28444~~ 43.728	10.284
NL-Loo	(1996,2006)	2006	~~52.1679~~ 2003	~~5.74396~~ 52.168	5.744
RU-Fyo	(1998,2006)	2005	~~56.46167~~ 2006	~~32.92389~~ 56.462	32.924
RU-Zot	(2002,2004)	2003	~~60.8008~~ 2004	~~89.3508~~ 60.801	89.351
SE-Fla	(1996,1998)	1998	~~64.1128~~ 1997	~~19.4569~~ 64.113	19.457
SE-Nor	(1996,1999)	1997	1999	60.086	17.480
SE-Sk2	(2004,2005)	2005	~~60.12967~~	~~17.84006~~ 60.130	17.840
UK-Gri	(1997,1998)	1998	~~56.60722~~	~~-3.79806~~ 56.607	-3.798
US-Blo	(1997,2006)	2006	~~38.8952~~ 2000	38.895	-120.633
US-Ho1	(1996,2004)	2004	~~45.2041~~ 2003	~~-68.7403~~ 45.204	-68.740
US-Me4	(1996,2000)	2000	~~44.4992~~	44.499	-121.622
US-SP1	(2000,2001)	2001	~~29.7381~~	~~-82.2188~~ 29.738	-82.219
US-SP2	(1998,2004)	2001	~~29.7648~~ 2004	~~-82.2448~~ 29.765	-82.245
US-SP3	(1999,2004)	2001	~~29.7548~~ 2002	~~-82.1633~~ 29.755	-82.163
Shrubs sites (Sh)					
CA-Mer	(1998,2005)	2004	~~45.4094~~ 2005	~~-75.5186~~ 45.409	-75.519
CA-NS6	(2001,2005)	2003	~~55.9167~~ 2004	~~-98.9644~~ 55.917	-98.964
CA-NS7	(2002,2005)	2003	~~56.6358~~ 2004	~~-99.9483~~56.636	-99.948
IT-Pia	(2002,2005)	2003	~~42.5839~~ 2004	~~10.0784~~ 42.584	10.078
US-Los	(2001,2005)	2005	~~46.0827~~ 2003	~~-89.9792~~ 46.083	-89.979
C4 grasses sites (C4G)					
BW-Ma1	(1999,2001)	2000	~~-19.9155~~ 2001	~~23.5605~~ 19.916	23.561
ZA-Kru	(2001,2003)	2002	~~-25.0197~~ 2003	~~31.4969~~ 25.020	31.497

Table B1. ~~Parameters of~~ PFT-specific JULES parameters optimised in this study ~~as described in table~~ (Table 1). The prior values and ranges for each PFT are given~~along with the initial ranges allowed~~. Below in bold are the optimised values and posterior uncertainty ranges given as ~~a~~ an 80% confidence interval (in parentheses). Optimised values for which the prior values ~~are found~~ lie outside the ~~new uncertainty~~ posterior range are highlighted by (*). A graphical version of this table is shown in Figure 3.

	BT	NT	C3	C4	Sh
n_0	0.046	0.033	0.073	0.06	0.06
	(0.001,0.2)	(0.001,0.2)	(0.001,0.2)	(0.001,0.2)	(0.001,0.2)
	0.061	**0.065***	**0.07**	**0.051***	**0.041**
	(0.034,0.066)	**(0.059,0.07)**	**(0.018,0.145)**	**(0.043,0.056)**	**(0.006,0.066)**
α	0.08	0.08	0.12	0.06	0.08
	(0.001,0.999)	(0.001,0.999)	(0.001,0.999)	(0.001,0.999)	(0.001,0.999)
	0.131*	**0.096**	**0.179***	**0.118***	**0.102**
	(0.087,0.14)	**(0.021,0.167)**	**(0.155,0.209)**	**(0.075,0.141)**	**(0.063,0.763)**
f_0	0.875	0.875	0.9	0.8	0.9
	(0.5,0.99)	(0.5,0.99)	(0.5,0.99)	(0.5,0.99)	(0.5,0.99)
	0.765*	**0.737***	**0.817**	**0.765***	**0.782***
	(0.655,0.787)	**(0.713,0.758)**	**(0.727,0.944)**	**(0.752,0.793)**	**(0.735,0.848)**
T_{low}	0	-10	0	13	0
	(-50,40)	(-50,40)	(-50,40)	(-50,40)	(-50,40)
	1.203	**-8.698**	**-1.985***	**11.37**	**-5.208***
	(-0.555,9.492)	**(-10.98,-6.342)**	**(-3.877,-0.13)**	**(7.522,14.072)**	**(-10.855,-2.106)**
T_{upp}	36	26	36	45	36
	(25,50)	(25,50)	(25,50)	(25,50)	(25,50)
	38.578*	**34.721***	**36.242**	**44.897**	**35.385**
	(38.157,40.698)	**(33.214,36.365)**	**(33.087,38.599)**	**(44.201,46.426)**	**(26.339,40.216)**
d_r	3	1	0.5	0.5	0.5
	(0.1,4)	(0.1,4)	(0.1,4)	(0.1,4)	(0.1,4)
	3.009	**1.425***	**0.991***	**0.404***	**0.411***
	(2.901,3.052)	**(1.159,1.672)**	**(0.901,1.101)**	**(0.5,3.623)**	**(0.324,0.473)**
$\frac{\delta c}{\delta l}$	0.05	0.05	0.05	0.05	0.05
	(0.001,0.1)	(0.001,0.1)	(0.001,0.1)	(0.001,0.1)	(0.001,0.1)
	0.047*	**0.045***	**0.05**	**0.05**	**0.048**
	(0.046,0.049)	**(0.042,0.048)**	**(0.047,0.052)**	**(0.046,0.054)**	**(0.04,0.055)**
dq_c	0.09	0.06	0.1	0.075	0.1
	(0.001,0.2)	(0.001,0.2)	(0.001,0.2)	(0.001,0.2)	(0.001,0.2)
	0.048	**0.036**	**0.086**	**0.046***	**0.077**
	(0.02,0.183)	**(0.008,0.066)**	**(0.07,0.109)**	**(0.045,0.053)**	**(0.024,0.118)**