

### Response to Reviewer #3 (our responses in italics)

Ref #3: This paper used the adjoint technology, L-BFGS-B optimization algorithm and Gibbs sampler to optimize the parameters of JULES. Generally speaking, this paper is well written since it has been reviewed and revised, but it also has three problems that I have to point out. (1) Data assimilation. This paper used 'adjoint' to provide the gradient information. Although 'adjoint' has been used in data assimilation, it is actually not 'data assimilation'. So the 'data assimilation' in the title, as well as in the main body, are suggested to be replaced by 'adjoint'. (2) Local optimum. If I understand correctly, the gradient based BFGS algorithm is used to find the local optimum, but if there are multiple local optimums, how to jump out of the local optimums to find the global one? (3) Posterior distribution. The posterior distribution seems really weird and I guess the reason is a too strong assumption: truncated multivariate normal distribution. If the true posterior distribution is not Gaussian, the results given by Gibbs sampler with such a strong assumption might be very misleading.

#### (1) Title: data assimilation

What's the definition of 'data assimilation' in your paper? For my all due respect, this title is really misleading. Usually, data assimilation means adjusting the state variables, such as soil moisture, temperature, and pressure, with observed values; while parameter optimization means tuning the parameters according to the deviation between observed and simulated output. Although they have some common techniques, these two concepts are fundamentally different. Kalman filter, 3D-Var and 4D-Var are very popular in the community of data assimilation, while in the community of parameter optimization, Genetic Algorithms, Metropolis-Hastings, and Gibbs samplers et.al., are more frequently used. Consequently, it is really weird to use 'optimization through data assimilation' in the title of this paper, because the adjoint (Hessian) of JULES, BFGS gradient based optimization and Gibbs sampler, were used to tune the parameters, while the state variables remain untouched. Although the 'adjoint' technique has been used in data assimilation, it is not equals to data assimilation. I think it might be better to replace 'data assimilation' with 'adjoint'.

***Our Response:*** *We acknowledge that we use a broader definition for "Data Assimilation" than is often the case in weather forecasting. Our application of data assimilation involves using data assimilation techniques to estimate internal model parameters, while DA in weather forecasting is typically used to estimate the initial state of the atmosphere. However, we are certainly not the first to use "Data Assimilation" in this sense (see for example Rayner et al., 2005). In addition Reviewer #1 has previously requested that we use the term "Data Assimilation" more rather than less prominently (in contradiction to the suggestion of Reviewer #3). In order to avoid a potential impasse, we have therefore retained the term but explained our use of it more clearly by the addition of the following text at the beginning of the data assimilation section.*

***"The term 'Data Assimilation' is commonly used to describe the process of using observations to refine the initial state within a numerical representation of a system (Bouttier and Courtier, 1999). This is most obviously the case for weather forecasting, in which the temperature, humidity and wind fields define the initial state. However, data assimilation techniques have been also used for parameter estimation, for example in hydrological models (Madsen, 2003, Liu and Gupta, 2007), and carbon cycle data assimilation systems (CCDAS, (Rayner et al. (2005); Kaminski et al. (2013)). In parameter optimisation by data assimilation, the internal parameters of a model take on the role of the dynamical state variables in initial state estimation by data assimilation. Nevertheless, the underlying techniques (e.g. of defining a model adjoint and minimising the error in the fit to data), are very similar in these two applications of data assimilation. This paper is certainly not the first to define parameter estimation of this form as data assimilation (Braswell et al. (2005), Stockli et al (2008), Verbeeck et al. (2011), Kuppel et al (2012), Hararuk et al. (2014)), but the reader should note the subtle difference between our definition of data assimilation and that commonly used in weather forecasting."***

**Ref #3: (2) Page 2, line 5:** two major sources: (a) process uncertainty, and (b) parameter uncertainty...

You are missing another important source, the uncertainty due to initial and boundary conditions (forcing data, for land surface models). [Kavetski et al., 2006a, 2006b][Ajami et al., 2007]

***Our Response:*** *Thanks, this is a good point. We have included the following relevant text: "Uncertainties in LSMs arise from three major sources: parameter uncertainty, process uncertainty and uncertainty due to initial and boundary conditions. Taking these in reverse order, uncertainty due to initial and boundary conditions in the case of LSM refers to uncertainty in the forcing data (Kavetski et al. (2006a, 2006b), Ajami et al. (2007))."*

**(3) Page 3, line 15, Page 5, line 1 and 12, etc.**

Replace 'data assimilation' with 'Adjoint based optimization' or something that can accurately describe the methodology used in this paper. See my previous comment (1).

***Our Response:*** *See response to comment (1).*

**Ref #3: (4) Page 6, line 12:** gradient descent algorithm L-BFGS-B, **Page 6, line 22:** locally optimal parameter vector, **Page 6, line 24:** locally optimized parameters

This paper used the gradient descent algorithm L-BFGS-B as optimization method, and the gradient information was provided by 'adjoint'. My concern is, if the optimization algorithm is gradient base, how to deal with the local optimums? If the response surface has multiple local optimums, the gradient-based methods will be trapped in a local optimum so that they cannot find the global optimum. There have been a lot of approaches to find the global optimum, such as Genetic Algorithm [Wang, 1991], SCE-UA [Duan et al., 1992], or multi-start methods [Krityakierne and Shoemaker, 2015]. Although the performance of JULES has been significantly improved, it is still not perfect for me because the model performance can be further improved if the algorithm can find the global optimum. Please provide convincing information that the algorithm can successfully jump out of local optimums, or use another algorithm to find the global optimum.

***Our Response:*** *Getting stuck in local optima is indeed a potential shortcoming of any gradient-descent methodology. However, the consistency between our single site and multiple site optimisations gives us confidence in the robustness of the convergence of our algorithm for this application. The gradient-descent methodology, albeit with the slower tangent linear method, has also been used in other LSM parameter estimation papers e.g. Kuppel et al. 2012. However, we have added text to acknowledge this issue:*

***"There is always a risk of becoming stuck in local minima when optimising within a high- dimensional parameter space by gradient descent. When an optimisation finds a local minimum, the final optimised state depends on the initial conditions. The consistency between our single site and multiple site optimisations therefore gives us some confidence in the robustness of the convergence of our algorithm for this application ."***

**Ref #3 : (5) Page 9, line 17:** a truncated multivariate normal distribution ... using Gibbs sampling...

This is actually a very strong assumption about the posterior distribution. If you want to obtain the posterior distribution of parameters but do not know too much about the shape of it, more general approaches, such as Metropolis-Hastings[Hastings, 1970], Adaptive metropolis[Haario et al., 2001], or DRAM[Haario et al., 2006], might be better.

I think the reason of strongly correlated joint posterior distribution is your strong assumption of truncated multivariate normal distribution. If you assumed the posterior distribution to be multivariate normal, it is impossible to get any other shape of distribution. Due to my experience, the response surfaces of land surface models are usually not Gaussian, i.e. have multiple local optimums, valleys and peaks, so that such a strong assumption will give totally misleading results. So I suggest to use more general MCMC approaches, such as MH, AM or DRAM to replace Gibbs.

***Our Response:*** *We assume uniform top-hat priors on the initial parameters given their upper and lower bounds. This leads to truncated Gaussian posteriors. Since the posterior distribution is truncated, we cannot simply extract a simple probability density function. At*

*the optimum found using the gradient descent method, the Hessian is used to approximate the local quadratic around the optimum. Gibbs sampling is used to sample this space. This can be done without requiring further model runs.*

**Ref #3 : (6) Page 14. Section 3.2.1** should be section 4.

***Our Response:*** *Section 3 is titled 'Results and discussion' and section 3.2.1 'Assessment of PFT-specific optimal parameters'. Whilst we agree that putting this part as a subsubsection detracts from its importance, it does fit well in the section 3 description. As a compromise, section 3.2.1 has been moved up a level to be a subsection of section 3.*

**Ref #3 : (7) Page 17, line 10:** Alternative methods ... could avoid this issue, but are more computationally costly.

How many model evaluations did it cost in this paper? Some classical global optimization methods, such as SCE-UA [Duan et al., 1994], are able to converge within several thousands of model evaluations, while the most recent surrogate based optimization algorithms, such as [Wang et al., 2014; Gong et al., 2015, 2016], are able to obtain global optimum with only hundreds of model evaluations. The global optimization methods are not as costly as you think.

***Our Response:*** *For this paper, optimisations typically took 150 function evaluations. Whilst we acknowledge that there are other methods, this does remain one of the most effective, and the purpose of this study was to explore what could be found using the model adjoint.*

#### Response to Reviewer #4: Professor Hoshin Gupta (our responses in italics)

##### General Comments:

Ref #4: For example the review of related literature [citing 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), Xiao et al., (2011), Kuppel et al., (2012, 2014), Medvigy et al. (2009), Verbeeck et al. (2011), Medvigy and Moorcroft (2011), Groenendijk et al. (2010)] fails to cite much of the related literature dating back to at least 1999 [e.g., Gupta et al (JGR 1999), Houser et al (JGR 2001), Laplatrier et al (JGR 2002), Xia et al (JHM 2002), Demarty et al (JoH 2004), Liu et al (JGR 2004), Liu et al (JHM 2005), Demarty et al (WRR 2005), Hogue et al (WRR 2006), Abramowitz et al (JHM (2007), Rosolem et al (HP 2012), to name just a few].

While I certainly do not expect the authors to cite all of the above mentioned examples (which would be embarrassingly self-serving), I am generally concerned that the discussion does not show much awareness of the related developments in other closely related fields (such as Hydrologic Science) from which much of the impetus for optimization of the parameters of LSM's derives ... beyond referring to the motivation being "*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))*". Certainly, to my knowledge, much early work in LSM parameter optimization was promoted by the group led by Professor Pitman in Sydney. And as such, the general attention to parameter optimization considerably pre-dates the current interest in the broader concept of "*data assimilation*".

***Our Response:*** We thank you for this extensive insight into the Hydrological literature that will be invaluable. A reference to Pitman already exists at the beginning of the paper but we have now included a few of the other references listed above.

Ref #4: Further, I would generally expect a manuscript of this kind to provide a comparative evaluation of the *model performance improvements* with those of related parameter optimization studies performed by other members of the LSM community (albeit not using derivative based optimization). And while I am not arguing that a comparison with other kinds of *optimization methods* necessarily needs to be performed (although it would be useful and informative) I do think it would be prudent to provide some comments about a) computational cost and b) relative advantages and disadvantages vis-a-vis other optimization approaches that have been used for LSM parameter optimization by the community (beyond simply remarking that the method is prone to premature convergence at local optima).

***Our Response:*** As far as we are aware there are no other such studies performed on the JULES land-surface model. The Blyth et al. benchmarking study outlines the performance of JULES using the default parameters and this study improves the fit at the sites used in the benchmarking amongst many more sites. Computational cost information has now been added to the main text.

##### Section on Methods and Data:

**Ref #4: (1)** Only eight parameters relating predominantly to leaf-level stomatal conductance and photosynthesis (including the hydrological partitioning at the land-surface) are calibrated. Please comment on the fact that given there are "*over a hundred internal parameters*" in JULES that need to be specified, and that "*the detailed performance of a land-surface model can be very sensitive to such internal parameters*", it is quite possible that fixing most of the parameters during the optimization might affect the calibration results (due to parameter interdependence effects). Further, the model likely contains additional coefficients that are fixed (hard-coded) to values that may be generally suspect (see Mendoza et al WRR 2015); please comment on potential the implications of that to the results obtained.

**Our Response:** The following sentence has been added to the end of line 25 in section 2.4:

**"Note that this only represents a modest subset of the parameters available and as such the results could be different when considering different subsets in the calibration."**

**Ref #4: (2)** While you point out the inherent "subjectivity" and lack of reproducibility of LSM parameter calibration by manual adjustment, please comment on the history (since at least 1999) of the application of "objective" automated (multiple-criteria) methods to LSM calibration, albeit not with gradient based algorithms, and also please comment on the relative strengths and weakness of manual versus automated methods. I should also point out that the mathematical basis for such automated calibration significantly pre-dates the "data assimilation" literature that the authors cite, and goes back at least as early as Bard (1974, Nonlinear Parameter Estimation, Academic Press) as a well established reference.

**Our Response:** We are of the view that our manuscript gives sufficient reference to other methods, within the confines of the problem that we address.

**Ref #4: (3)** Please comment on the reliability of the "parameter uncertainty estimates" provided by adjoint methods (linear-Gaussian approximation), given the significant parameter-output nonlinearity associated with LSMs (I note, in particular, your comment that "optimal values need not be in the centre of the uncertainty range, the PDF can be skewed". For example, does empirical Monte-Carlo sampling of the region of the optimum (thereby approximating the true non-Gaussian shape of the posterior parameter pdf) provide similar uncertainty ranges for the parameters?

**Our Response:** We assume uniform top-hat priors on the initial parameters given their upper and lower bounds. This leads to truncated Gaussian posteriors. Since the posterior distribution is truncated, we cannot simply extract a simple probability density function. At the optimum found using the gradient descent method, the Hessian is used to approximate the local quadratic around the optimum. Gibbs sampling is used to sample this space. This is done without running any more JULES runs. The optimal values not lying at the centre of the PDF is due to the truncated and multidimensional nature of the posterior distributions.

**Ref #4: (4)** Continuing on from the above, it appears that for the results you did not actually use the Hessian generated by adjULES to report the uncertainty estimates for the parameters, but instead used sampling of the posterior distribution (Section 2.5.1), which makes perfect sense. Since the earlier part of the paper gives one the impression that the uncertainty estimates are computed directly by adjULES it would be good to modify the presentation to make the actual fact clear (remove possibility for confusion).

**Our Response:** The Hessian is used to generate the uncertainty estimates, see response to comment (3) for more details.

**Ref #4: (5)** Please comment on the fact that use of an additive cost function (where only the total summed cost is minimized, as opposed to an approach where all individual cost functions are required to be improved) means that it is possible for the optimization method to achieve the "best" solution by improving the match at one site while possibly making the match worse (than it could otherwise be) at one or more other sites simply to achieve a better value of the cost function.

**Our Response:** The following sentences have been added to the end of the Multisite Implementation section:

**"An additive cost function, where the optimisation criterion is to minimise the total cost, was chosen over a cost function where all individual cost functions are required to be improved. This was due to the fact that all of the sites were used finding the optimal parameter vector for each PFT. Sites which do not improve with the rest of the PFT will suggest wrong classification of the site or issues with the PFT definitions."**

**Ref #4: (6)** Please comment on the fact that, due to model structural errors, calibration to specific observables could actually cause model simulations of other fluxes (that were not used in tuning) to become worse (this is a potentially very serious problem in multi-flux calibration by weighted single criteria optimization to only some of the model fluxes); please see *Gupta et al (JGR 1999)*.

**Our Response:** The following sentence has been added to page 7 line 13:

**“Also note that due to model structural errors, calibration to these two specific observables could actually cause model simulations of other fluxes (that were not used in tuning) to become worse (Gupta et al. 1999).”**

**Ref #4: (7)** I am concerned that the  $\epsilon^2$  “fractional variance explained” measure is not really a properly informative measure of model performance, given that the benchmark for comparison is the observed seasonal cycle (I note also the related comment by a previous reviewer). In Hydrology this is related to a metric known as the “Nash” efficiency (equivalent to  $1 - \epsilon^2$ ) that has been repeatedly demonstrated to be a poor index of model performance unless  $1 - \epsilon^2 > 0.85$  or  $0.9$  (such values are rarely achievable), and can hide the existence of significant bias in the performance (see, e.g., *Schaefli and Gupta (HP 2007)*). Note that the “Nash” efficiency is also typically justified as enabling cross-site and cross-model comparison, but arguable this is a poor reason for using a poorly informative metric. Instead the component decomposition [e.g., see *Murphy (Monthly Weather Review 1988)*, *Gupta et al (Journal of Hydrology 2009)*] can provide a more meaningful indicator of performance in terms of bias, variability and cross-correlation (see application in *Rosolem et al, Hydrological Processes 2012*).

**Our Response:** Since our cost function is described using RMSE, we felt the natural choice was to use a normalised version to perform the main diagnostic, especially in terms of cross-site comparison. However, we do agree that further breakdown of the metric would be beneficial. As such, we have decided to use Taylor diagrams since it provides a visual methodology to decompose the misfit into standard deviation and correlation.

A sentence has been added to end of section 2.5.2 in order to acknowledge this metric’s shortcomings and section 3.4 has been included displaying a few Taylor diagrams and their assessment.

**“In Hydrology this is related to a metric known as the “Nash-Sutcliffe” efficiency (Nash and Sutcliffe, 1970), equivalent to  $1 - \epsilon^2$ , and has been used by many studies to perform cross-site comparisons.”**

Section on Results and discussion:

**Ref #4: (8)** Please comment about your results in the context of the findings by *Abramowitz et al (JHM 2007)* who show that “... as much as 45% of per-time-step model root-mean-square error in ... flux outputs is due to systematic problems in ... model processes insensitive to changes in vegetation parameters ... These results suggest that efforts to improve the representation of fundamental processes in land surface models, rather than parameter optimization, are the key to the development of land surface model ability”.

**Our Response:** We believe that this stated already in our conclusion on lines 10-14.

**Ref #4: (9)** Please comment on the sensitivity of the optimized results to choice of the constant of proportionality  $\lambda$  (how do the results change if  $\lambda$  is made smaller). Given the importance of using a “prior” on the parameters as constraint, this seems to me to be a rather important issue. limitation?

**Our Response:** The following sentence has been added to the end of section 3.2:

**"Choice of  $\lambda$  had less effect on the values of the optimal parameters than on uncertainties and correlations found. The uncertainty ranges become larger with smaller values  $\lambda$  and correlations less pronounced."**

**Ref #4: (10)** Please comment on the quality of the site-specific performance improvement in comparison with findings obtained by others. For example, *Rosolem et al (Hydrological Processes 2012)* reported that "All sites showed improvements in simulation of the surface energy and carbon fluxes" and "In contrast, the default parameter sets (commonly used in GCM simulations) were found to be unable to reproduce the diurnal variation of energy fluxes at the tropical rainforest sites and showed a tendency to overestimate (underestimate) sensible (latent) heat fluxes. The calibration improved the simulations of these two fluxes by removing bias and variability errors (errors in signal mean and standard deviation)."

**Our Response:** *It is not clear how we can compare diurnal cycles in Rosolem et al to the annual cycle calibrations we perform in our study.*

**Ref #4: (11)** In general it is well accepted that it is not ever possible to "validate" a model, and so the term validation in regards to model performance evaluation would seem to be misleading. Might I politely suggest the use of the more accurate term "evaluation" in its stead?

**Our Response:** *The use of the word validation is due to its established pairing with the word calibration. Many studies use calibration and validation, and whilst substitution would not change the science, we prefer to stick with the word 'validation'.*

#### Section on Conclusions:

**Ref #4: (12)** Perhaps it would be appropriate to comment on the computational cost involved with optimization using the adJULES system and on how many model runs (cost function evaluations) are necessary to achieve convergence (starting from the default parameter values)?

**Our Response:** *Text to this effect has been added to end of the first sentence in section 3.1: '...typically requiring 150 function evaluations to find a local optimum.'*

#### REFERENCES:

Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., 106, 7183-7192, 2001 (also see PCMDI Report 55, <http://www.pcmdi.llnl.gov/publications/ab55.html>)

# Land surface parameter optimisation ~~through~~ using data assimilation ~~techniques~~: the adJULES system

Nina M. Raoult<sup>1</sup>, Tim E. Jupp<sup>1</sup>, Peter M. Cox<sup>1</sup>, and Catherine M. Luke<sup>1</sup>

<sup>1</sup>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 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) in JULES. The optimised PFT-specific parameters improve the performance of JULES ~~at~~ over 85% of the sites used in the study ~~both at~~, ~~at both~~ the calibration and validation stages. The new improved parameters for JULES are 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 on which~~ 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 inter-comparison (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<sub>2</sub> 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<sub>2</sub> could imply smaller “compatible emissions” in order to stay below key warming thresholds such as two degrees.

Uncertainties in LSMs arise from ~~two~~three major sources: parameter uncertainty, process uncertainty and ~~parameter uncertainty~~  
5 ~~–uncertainty due to initial and boundary conditions. Taking these in reverse order, uncertainty due to initial and boundary~~  
~~conditions refers to uncertainty in the forcing data (??, ?).~~ Process uncertainty includes the misrepresentation of land-surface  
processes and also the neglect of important processes (such as nitrogen-limitations on 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 uncer-  
tainty almost invariably leads to increases in LSM complexity, which typically leads to the introduction of additional internal  
10 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  
15 rigorous data assimilation frameworks. These approaches have been used in a number of studies, covering various 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. This over-tuning may occur when a single site does not represent the full range  
20 of a ~~PFT~~Plant Functional Type (PFT), given different tree types, tree ages and ~~aboveground~~above ground 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)~~PFTs. Model parameters  
25 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 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.

30 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  
35 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 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~~ a (locally) optimum vector of generic parameters for each of the JULES PFT classes be found?
- How does the 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 introduced, along with the data used and the parameters chosen to be optimised in the study. 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 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 the 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 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.

30 Vegetation in the JULES model is categorised into five 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 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  
5 such as surface temperature ( $T_{upp}$ ,  $T_{low}$ ), solar radiation ( $\alpha$ ), 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~~ sensitive to these parameters in previous studies (Blyth et al.,  
10 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  $\alpha$ ). It follows that:

$$15 \quad A = A(n_0, \alpha, T_{low}, T_{upp}, c_i, \beta) \quad (1)$$

where  $c_i$  is the internal CO<sub>2</sub> concentration inside the leaf, and  $\beta$  is a soil moisture stress factor which depends on the vertical soil moisture profile  $\theta$ , and the plant rootdepth  $d_r$ :

$$\beta = \beta(\theta, d_r) \quad (2)$$

The internal CO<sub>2</sub> concentration  $c_i$  is assumed to be dependent on the external CO<sub>2</sub> concentration  $c_a$  and the atmospheric  
20 humidity deficit  $\delta q$  (Cox et al., 1998) via the equation:

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

where  $c_*$  is the CO<sub>2</sub> compensation point, and  $f_0$  and  $\delta q_c$  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<sub>2</sub> concentrations:

$$25 \quad g = 1.6 \frac{A}{c_a - c_i} \quad (4)$$

The factor of 1.6 converts the stomatal conductance for CO<sub>2</sub> into a stomatal conductance for water vapour. The scaling-up from leaf to ~~canoy~~ canopy level in this version of JULES uses a “big-leaf” approach (Cox et al., 1999).

## 2.2 Data assimilation system

The term ‘Data Assimilation’ is commonly used to describe the process of using observations to refine the initial state within a numerical representation of a system (Bouttier and Courtier, 1999). This is most obviously the case for weather forecasting, in which the temperature, humidity and wind fields define the initial state. However, data assimilation techniques have also been used for parameter estimation, for example in hydrological models (, ), and carbon cycle data assimilation systems (CCDAS; Rayner et al. (2005), Kaminski et al. (2013)). In parameter optimisation by data assimilation, the internal parameters of a model take on the role of the dynamical state variables in initial state estimation by data assimilation. Nevertheless, the underlying techniques (e.g. of defining a model adjoint and minimising the error in the fit to data), are very similar in these two applications of data assimilation. This paper is certainly not the first to define parameter estimation of this form as data assimilation (, , Verbeeck et al. (2011), Kuppel et al. (2012), ), but the reader should note the subtle difference between our definition of data assimilation and that commonly used in weather forecasting.

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 ‘real world’ quantities (e.g. aerodynamic roughness length, surface albedo, plant root-depth). 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, such as eddy covariance flux data. This is typically ~~carried-out~~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, derived by automatic differentiation, which enables efficient and objective calibration against observations. Importantly, ad-JULES 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 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,  $\mathbf{z}$ . The cost function,  $f(\mathbf{z})$ , consists of a weighted sum of squares of the difference between  $\mathbf{m}_t$  (the vector of model outputs at time  $t$ ), and  $\mathbf{o}_t$  (the vector of observations at time  $t$ ), combined with a term quadratic in the difference between parameter values  $\mathbf{z}$  and initial parameter values  $\mathbf{z}_0$ :

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

Here,  $\mathbf{R}(\hat{\mathbf{z}}) = \frac{1}{n} \sum_{t=1}^n (\mathbf{m}(\hat{\mathbf{z}})_t - \mathbf{o}_t)(\mathbf{m}(\hat{\mathbf{z}})_t - \mathbf{o}_t)^T$  denotes the error cross product matrix produced by a JULES run with parameter value  $\hat{\mathbf{z}}$ . In an optimisation,  $\mathbf{z}$  and  $\hat{\mathbf{z}}$  are updated separately in nested loops, having both been initialised to the default JULES parameter value  $\mathbf{z}_0$ . In the inner loop,  $\mathbf{z}$  is varied to minimise the cost function (termination criterion:  $\nabla f \approx 0$ ) for the current value of  $\hat{\mathbf{z}}$ . In the outer loop,  $\hat{\mathbf{z}}$  is reset to the new value of  $\mathbf{z}$  from the inner loop (termination criterion: change in  $\hat{\mathbf{z}}$  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 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  $\lambda$  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  $\lambda$  help condition the problem and force parameter values to be close to the initial value  $\mathbf{z}_0$  (Bouttier and Courtier, 1999). All parameters and observations are equally weighted in this cost function.

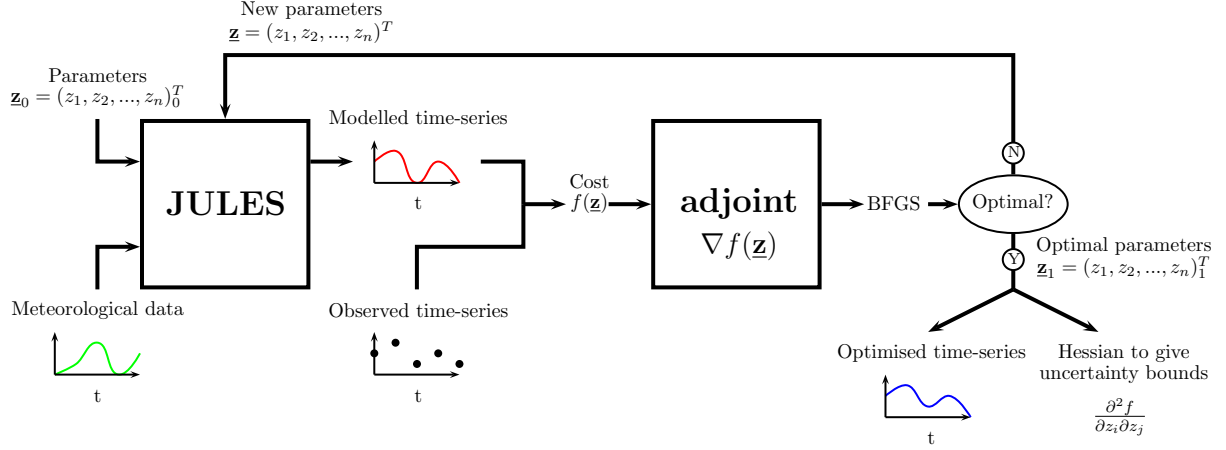
The optimal vector of parameters is the vector  $\mathbf{z}$  that minimises the cost function (Eq. 5). The aim of adJULES is to find this vector. adJULES minimises the cost function iteratively using the gradient descent algorithm L-BFGS-B (Byrd et al. (1995), optim: R Development Core Team (2015)). 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  $\nabla f(\mathbf{z})$  of the cost function  $f(\mathbf{z})$  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)). 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  $\nabla f(\mathbf{z})$  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  $\mathbf{z}_1$  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 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



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

each 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.

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

An additive cost function, where the optimisation criterion is to minimise the total cost, was chosen over a cost function where all individual cost functions are required to be improve. ~~This was due to the fact that all~~ All of the sites were used in finding the optimal parameter vector for each PFT. ~~Sites~~, so that sites which do not improve with the rest of the PFT will suggest wrong suggest incorrect classification of the site or issues with the PFT definitions.

### 2.3 Eddy covariance flux data

The eddy-covariance flux data used in this study are part of FLUXNET (Baldocchi et al., 2001). The FLUXNET database contains more than 500 locations worldwide, and all of the data 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 data gaps of more than 50% during the growing season were omitted.

To constrain photosynthetic parameters, Net Ecosystem Exchange (NEE) and Latent Heat Flux (LE), among other fluxes, are helpful. The NEE flux, defined as the net flux of  $\text{CO}_2$ , is partitioned into gross primary production (GPP) and ecosystem respiration (Resp) (Reichstein et al., 2005). In this study this the GPP flux is used, along with the LE flux to constrain the

15 model. GPP data are model-derived estimates, which could introduce an additional uncertainty into the results. Also note that due to model structural errors, calibration ~~to these two specific observables could actually~~ against these two particular observables could cause model simulations of other fluxes (~~that were~~ not used in the tuning) to become worse ?.

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  
5 the closest grid point 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  $\mathbf{m}_t$  and  $\mathbf{o}_t$  (Eq. 5): net ecosystem exchange (NEE),  
10 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 outlines the parameters chosen. Note that this only represents a modest subset of the parameters available and as such the results could be different when considering different subsets in the calibration.

15 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~~ applied repeatedly 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  
20 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  
25 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 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 this study the simple big leaf approach was adopted (Clark et al., 2011), although optimisations can  
30 also be 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 optimal parameter vectors.

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  $\lambda$  in Eq. 5). In each of the multisite experiments, the lowest value of  $\lambda$  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.

## 2.5 Analysis tools

### 2.5.1 Parameter uncertainty

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

$$H_{ij} = \frac{\partial^2 f}{\partial z_i \partial z_j} \quad (7)$$

- 15 where  $f(z)$  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.

- 20 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 dimensions of a high dimensional multivariate normal distribution which is truncated. 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.

25 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.



### 2.5.2 Fractional error - a metric of model-data fit

To measure the improvement exhibited by different parameter vectors, the fraction of variance unexplained  $\epsilon^2$  is used to define the fractional error  $\epsilon$ . 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,  $\mathbf{z}$ , a modelled time series  $\mathbf{m}_{i,t}$  with  $k$  data points is generated using JULES, where  $i$  denotes one of the observable data streams (in this case LE and GPP). For each data stream  $i$ , the fraction of variance unexplained by the model is

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

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

$$\epsilon^2 = \frac{\epsilon_1^2 + \epsilon_2^2}{2}, \quad (9)$$

is a single dimensionless measure of model misfit. The fractional error  $\epsilon$  can then 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  $\mathbf{m}_{i,t}$  always equals the observational mean  $\bar{\mathbf{o}}_i$ .

In Hydrology this is related to a metric known as the “Nash-Sutcliffe efficiency” (?), equivalent to  $1 - \epsilon^2$ , and has been used by many studies to perform cross-site comparisons. ?

## 3 Results and discussion

In this section, the site-specific optimisations are ~~first-considered~~ considered first. 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 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 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 ~~typically-requiring~~ Typically, this required about 150 function evaluations to find a local optimum. 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  $\lambda$  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 LE and GPP for the different sites are shown in Fig. 2.

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

Similarly for the needleleaf sites (Fig. 2, 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.

5 GPP is also underestimated for the C3 grass sites (Fig. 2, middle row) and, for the majority of the sites, the optimisation does a good job of correcting this. The LE flux tends to 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.

10 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, 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  
15 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 underestimate it, and similarly for GPP. The 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.

20 Overall, the 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 PFT-specific optimal parameter values

25 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 prior parameter value lies outside the posterior uncertainty bounds. The  $\frac{\delta c}{\delta t}$  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  
30 described in section 2.5.1, the optimal values need not be in the centre of the uncertainty range, the PDF can be skewed. Most

of the PFTs display high uncertainty in at least one of the parameters optimised; for the optimised broadleaf set for example,  $dq_c$  is highly unconstrained. For C4 grasses,  $d_r$  is so unconstrained that the optimal value found lies outside the 80% confidence interval. C3 grasses show large uncertainty in  $n_0$  and for shrubs, the parameter with the largest uncertainty is  $\alpha$ .

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 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.

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

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

if it is assumed that  $c_* \ll c_i$  and  $c_* \ll c_a$ .

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(a) and (b).

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.

Choice of  $\lambda$  had less effect on the values of the optimal parameters than on uncertainties and correlations found. The uncertainty ranges become larger with smaller values  $\lambda$  and correlations less pronounced.

### 3.2.1 Assessment of PFT-specific optimal parameters

### 3.3 Assessment of PFT-specific optimal parameters

The performance of the PFT-specific parameters is now compared to the default JULES values and to the parameters found by optimising independently at each measurement site. For each site, the fractional error in both the calibration year and the validation year is displayed Fig. 5.

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 sites (over 92% of sites). 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 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 optimal parameter vector improves 26 of the 27 broadleaf sites shown  
30 although one of the sites, 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. 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.

Similar levels of fit and error reduction can be seen in the validation years in the Broadleaf set. Only IT-Col and US-MMS  
5 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 in this PFT, the improvement  
10 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 when using the PFT-specific vector, for either the calibration or validation years. This site has a lower annual  
15 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 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 some sites, (e.g. US-Blo and BW-Ma1), the PFT-specific optimum outperforms the site-specific optimum in the cali-  
20 bration 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 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  
25 suggests that the seasonal cycle differs over the different years at these sites. For the C4 grass sites, which started with relatively  
high errors, the new parameter vectors improve the sites slightly 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. For the Shrub sites, both the site-specific  
and the PFT-specific 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.

### 5 3.4 Analysis of improvement in fit

The fractional error is a good tool for cross-site comparison but it does not give much information about the way in which the optimised parameter vectors improve the fit at each site. Taylor diagrams ( ? ) provide more insight into how the fit has been improved by considering the relationship between observed variance  $var(\mathbf{o}_t)$ , modelled variance  $var(\mathbf{m}_t)$ , error variance  $var(\mathbf{o}_t - \mathbf{m}_t)$  and model-observation correlation  $cor(\mathbf{o}_t, \mathbf{m}_t)$ .

10 The Taylor diagrams in Fig. 6 illustrate the improvement in performance of the optimised model for both the site-specific and PFT-generic parameters during calibration years (plots for validation years are very similar).

For latent heat at broadleaf sites (left), the improvement is most noticeable in cases where the seasonal cycle was overestimated. The correlation between the modelled time-series and observation time-series does not improve much but for the majority of the sites this starts off relatively high (over 0.6). Other PFTs show less drastic improvements for latent heat.

15 For GPP at needleleaf sites (right), the seasonal cycle is typically underestimated and improves noticeably for both the single-site parameter vectors and the PFT-generic parameter vectors. The correlation between model and observed time-series does not change greatly. The Taylor diagram for GPP at broadleaf sites is very similar. For grasses and shrubs, the change is less drastic, though some of the sites have a more notable increase in correlation.

20 Since Taylor diagrams are based on a decomposition of the variance of the errors they are insensitive to any systematic offset in the model. It therefore makes sense to consider in addition the normalised bias  $|\mu_m - \mu_o|/\sigma_o$ . Calculating this statistic separately shows a reduction in bias for nearly all sites. Taken together, these measures show that the observed improvements in model fit are due mainly to adjustment of the magnitude of the annual cycle and a reduction in bias.

## 4 Conclusions

25 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 associated uncertainty ranges.

30 For individual FLUXNET sites, 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, 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 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.

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. 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), adJULES was used to find the optimal PFT-specific parameters.

Although the PFT-specific optimal parameters do not always fit the data as well as site-specific optimal parameters, they still offer significant improvements over the default JULES parameters. For over 85% of the sites, PFT-specific optimal parameters perform better than default parameters 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 variability 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.

There is always a risk of becoming stuck in local minima when optimising within a high-dimensional parameter space by gradient descent. When an optimisation finds a local minimum, the final optimised state depends on the initial conditions. The consistency between our single site and multiple site optimisations therefore gives us some confidence in the robustness of the convergence of our algorithm for this application.

It is however clear that there are some limitations to the success of the optimisation results. Some sites still show significant differences between model output and observations. 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, this may be due to the model and data themselves, rather than parameter values. adJULES can therefore be used in the identification of model structural errors. 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~~ the local

minimum found depends on the initial parameter vector. However, as discussed in section 3.3, the fact that the cost function becomes smoother with additional sites may ~~be a solution to~~ help with 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  
5 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](http://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  
10 generated using commercial software TAF (sect. 2.2.1). For licensing reasons, the recalculation of the adjoint following code changes can 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  
15 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. Support for  
20 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.

The authors are grateful to T. Kaminski and R. Giering from FastOpt for their contribution to the development of the adjoint model, and  
25 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://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.

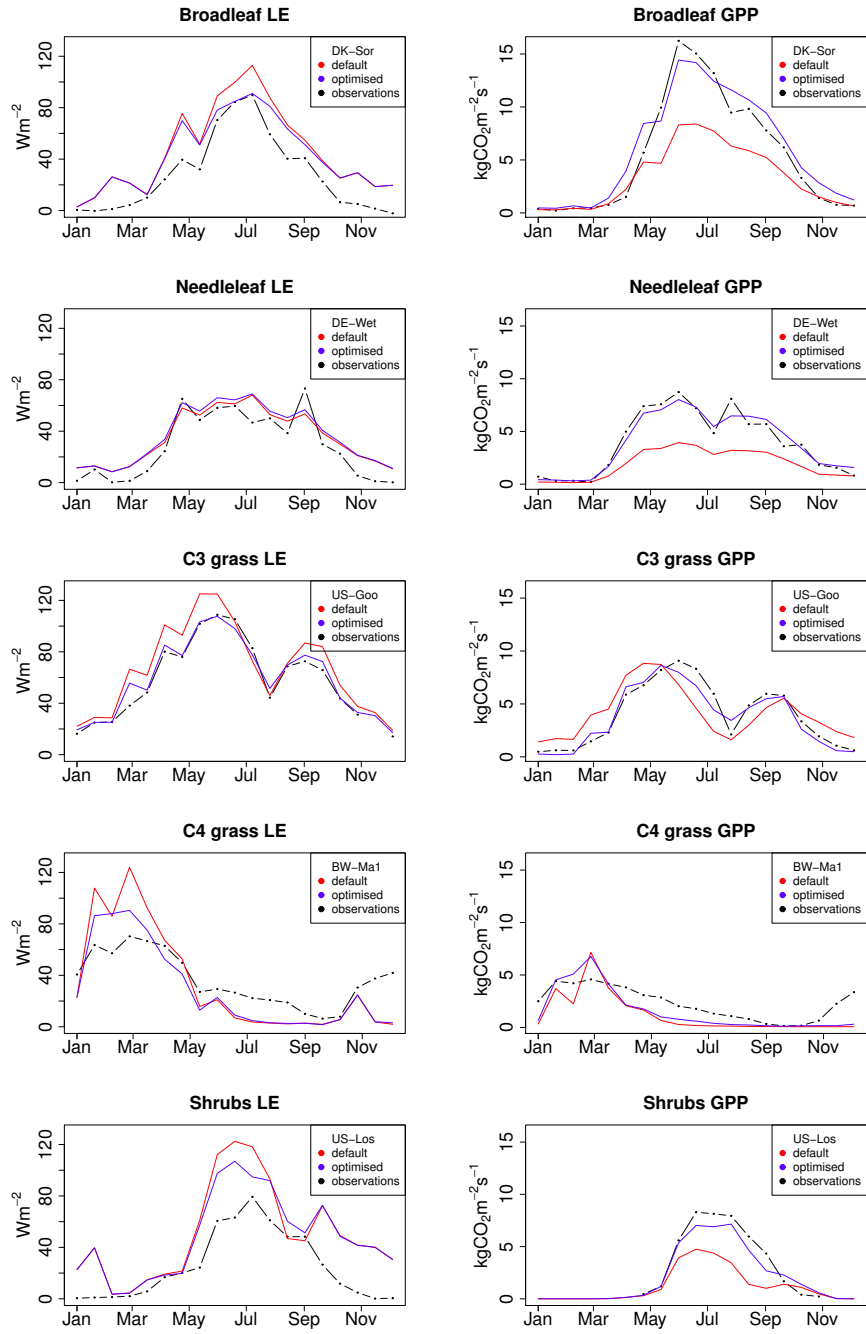
- Friedlingstein, P., Bopp, L., Rayner, P., Cox, P. M., Betts, R., Jones, C., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., et al.: Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison, *Journal of Climate*, 19, 3337–3353, 2006.
- Friedlingstein, P., Meinshausen, M., 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 plant 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.
- 5 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.
- 10 Raupach, M., Rayner, P., Barrett, D., Defries, R., Heimann, M., Ojima, D., Quegan, S., and Schimmlus, 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.
- 15 Reichstein, M., Tenhunen, J., Rouspard, 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 CO<sub>2</sub>/H<sub>2</sub>O 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.
- 20 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.
- 25 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 CO<sub>2</sub> 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.
- 30 Thum, T., Aalto, T., Laurila, T., Aurela, M., Lindroth, A., and Vesala, T.: Assessing seasonality of biochemical CO<sub>2</sub> 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 CO<sub>2</sub> fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model, *Journal of Geophysical Research*, 116, 2011.
- 35 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.

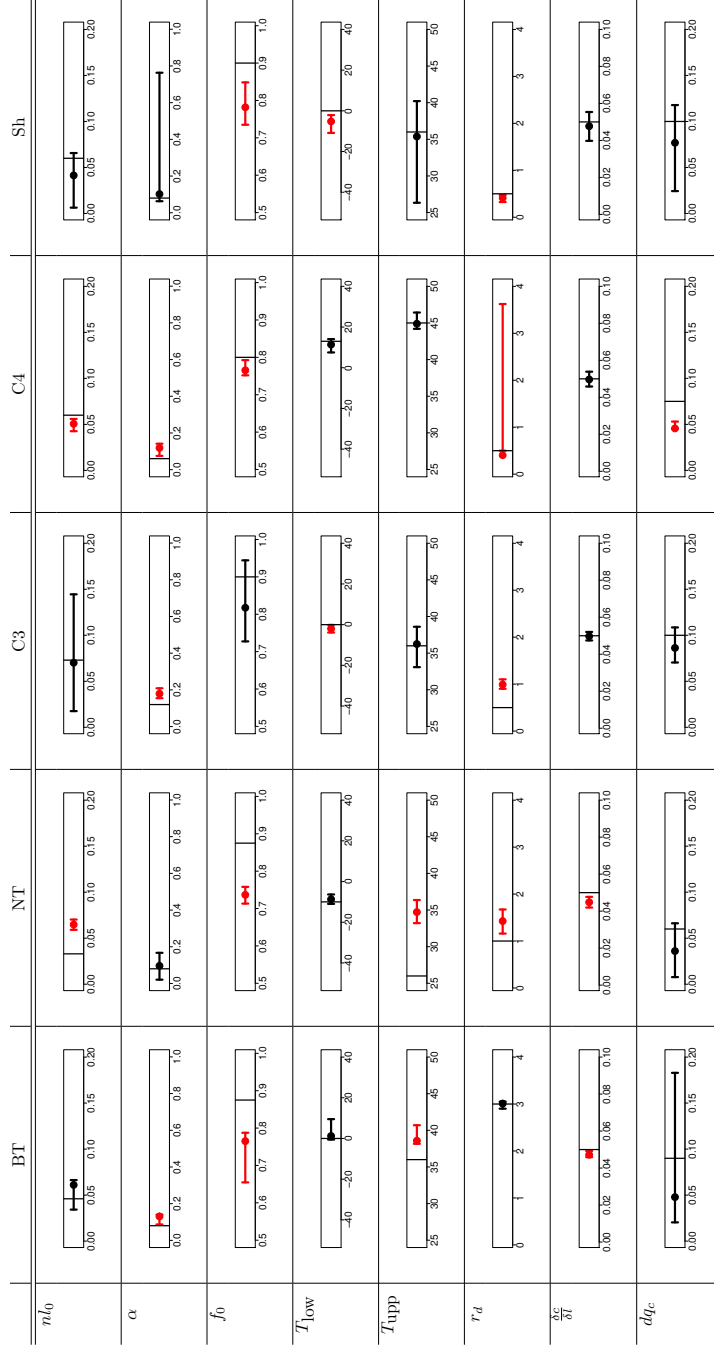
- Williams, M., Richardson, A., Reichstein, M., Stoy, P., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C., and Wang, Y.-P.: Improving land surface models with FLUXNET data, *Biogeosciences*, 2009.
- 5 Xiao, J., Davis, K. 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 Estimates, *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>, 101401, 2010.

**Table 1.** Parameters in optimisation vector, with descriptions.

Symbol	Name in code	Description	Units
$n_0$	n10	Top leaf nitrogen concentration	$\text{kg N (kg C)}^{-1}$
$f_0$	f0	Maximum ratio of internal to external $\text{CO}_2$	-
$d_r$	rootd_ft	Root depth	m
$\alpha$	alpha	Quantum efficiency	mol $\text{CO}_2$ per mol PAR photons
$\frac{\delta c}{\delta l}$	dcatch_dlai	Rate of change of canopy interception capacity with LAI	$\text{kg m}^{-2}$
$T_{\text{low}}$	tlow	Lower temperature for photosynthesis	$^{\circ}\text{C}$
$T_{\text{upp}}$	tupp	Upper temperature for photosynthesis	$^{\circ}\text{C}$
$dq_c$	dqcrit	Humidity deficit at which stomata close	$\text{kg kg}^{-1}$

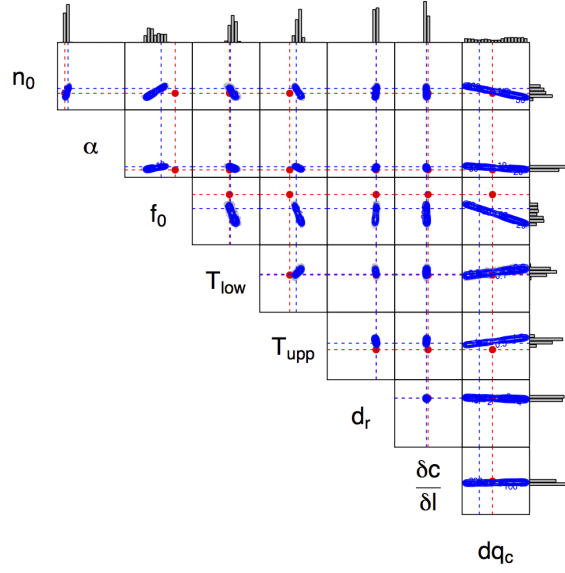


**Figure 2.** Time-series plots for illustrative site-specific validations showing LE (left) and GPP (right) for each of the different PFTs. Observations (black) are compared to JULES runs using default parameters (red) and site-specific optimal parameters (blue).

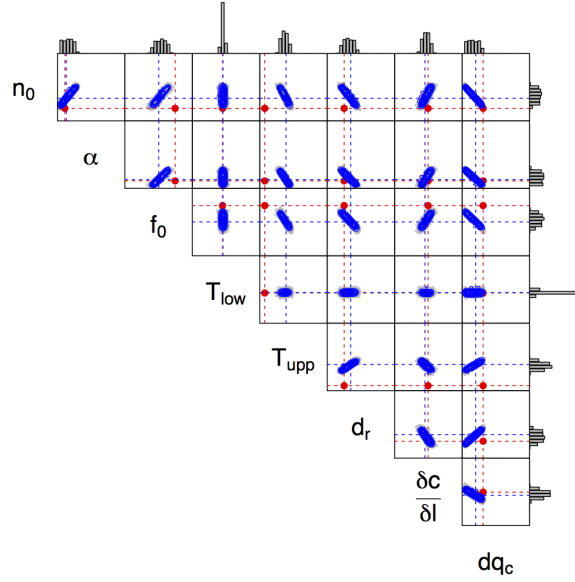


**Figure 3.** Summary of PFT-specific optimal JULES parameters found in this study (Table 1). The error bars show the uncertainty ranges given as an 80% quantile interval. The range of each box is the prior range of the parameters. Highlighted in red are the error bars for which the prior values (vertical line) are found outside the posterior uncertainty bounds. A numerical version of this figure is given in Table B1.

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

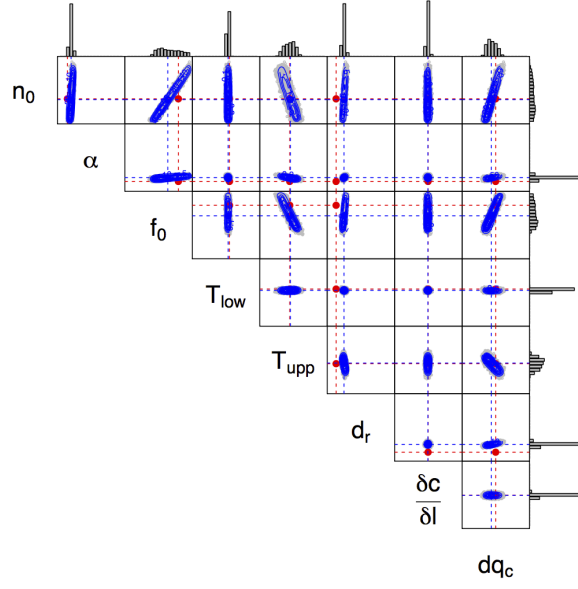


(a) Broadleaf

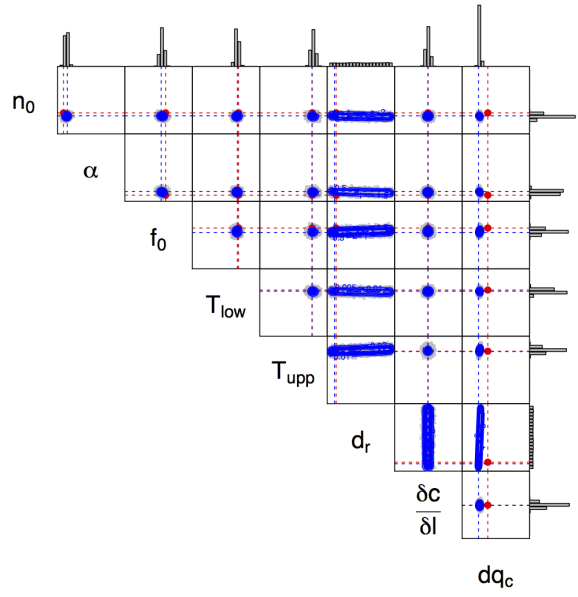


(b) Needleleaf



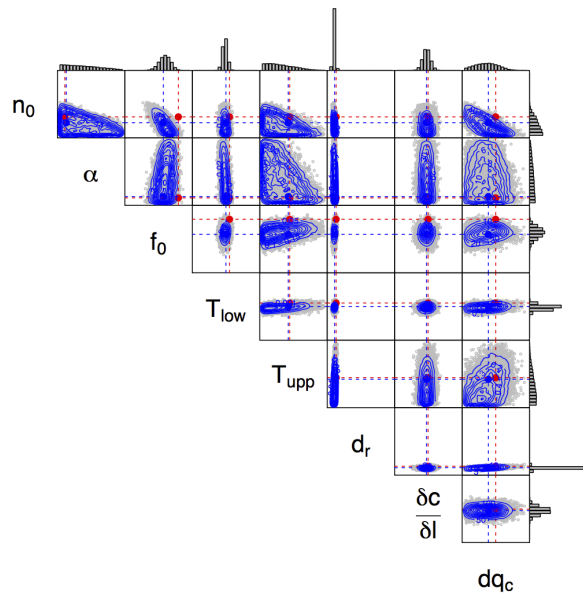


(c) C3 grasses



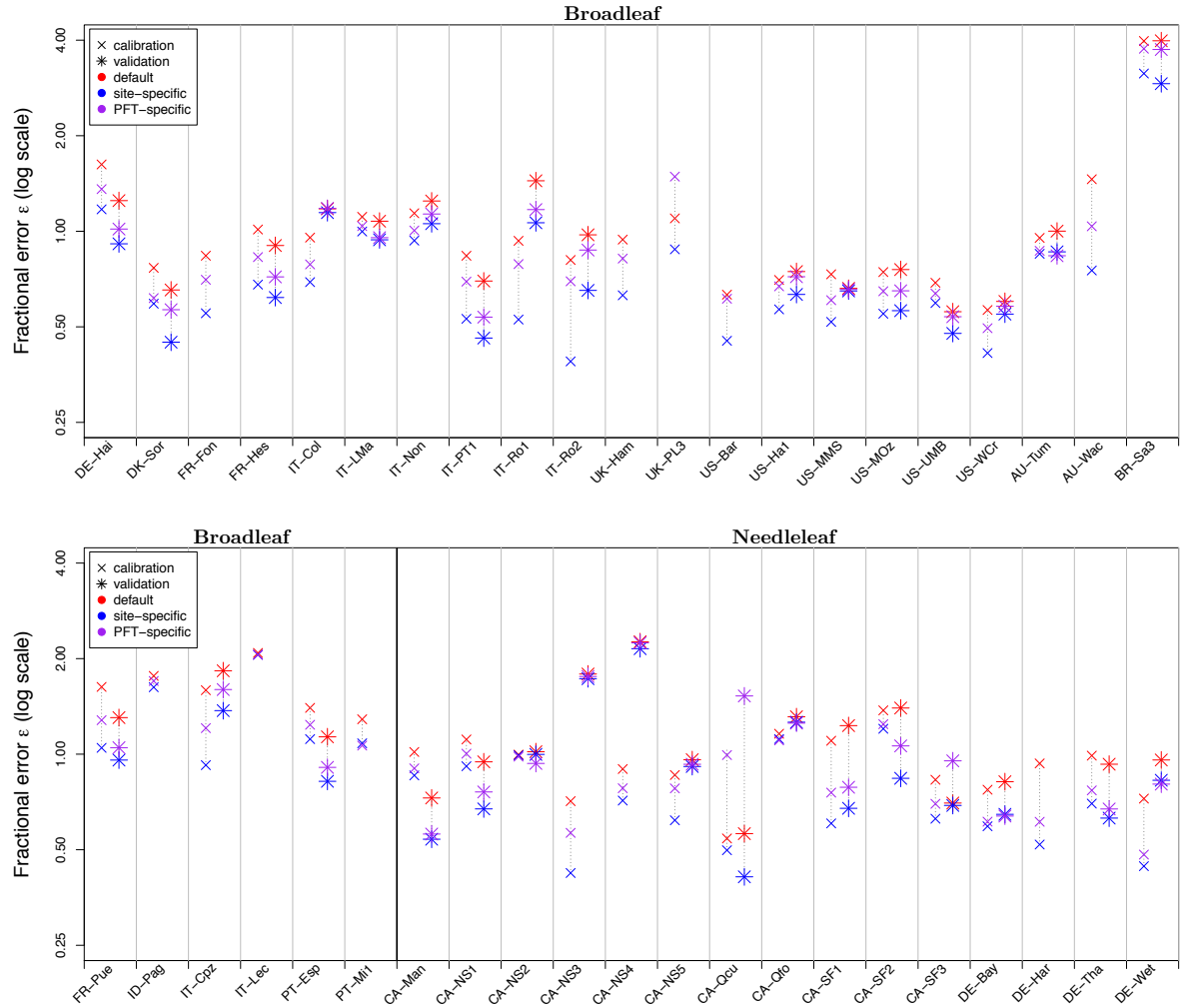
(d) C4 grasses

Figure 4. continued



(e) Shrubs

**Figure 4.** continued



**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 (\*). 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).

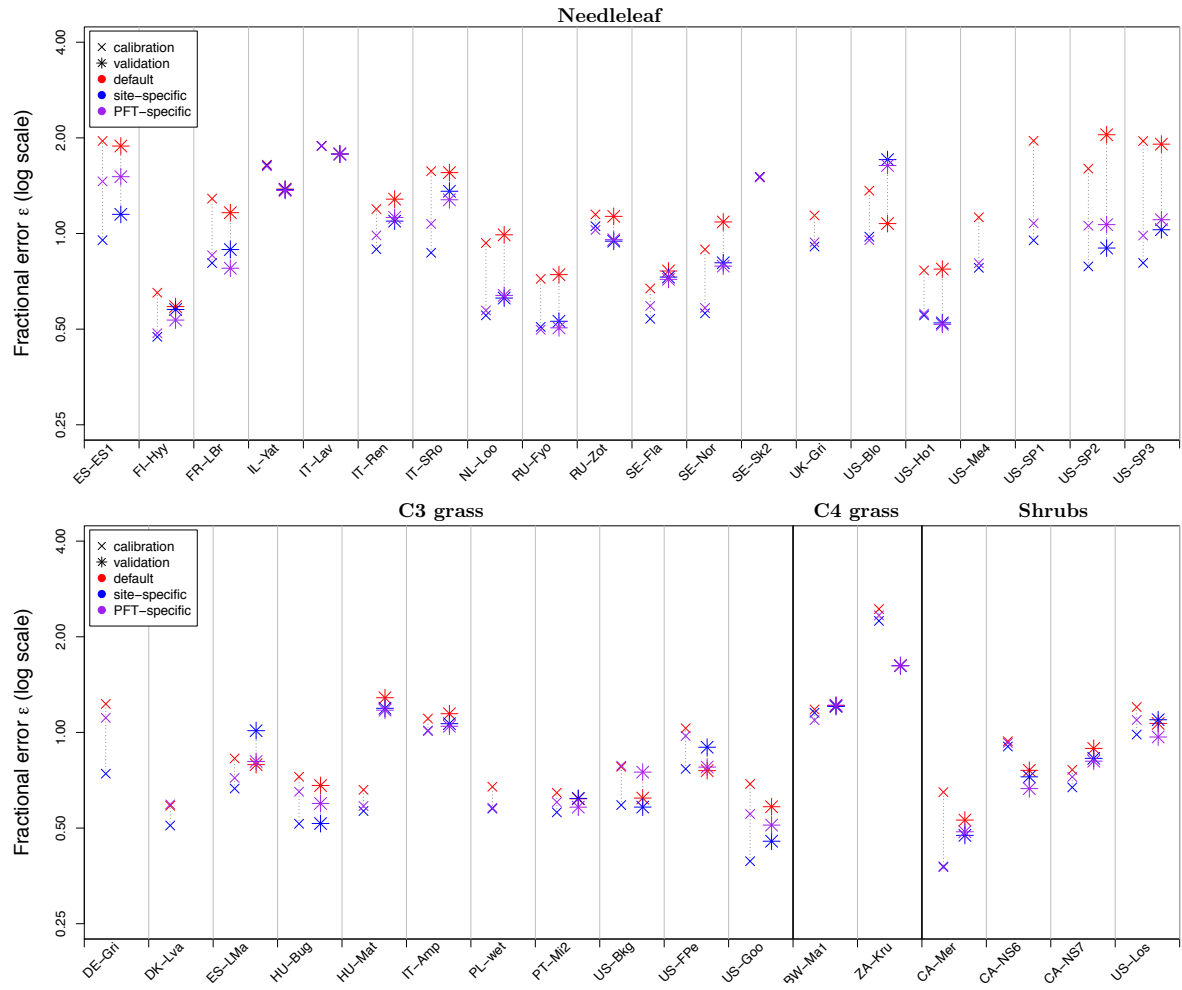
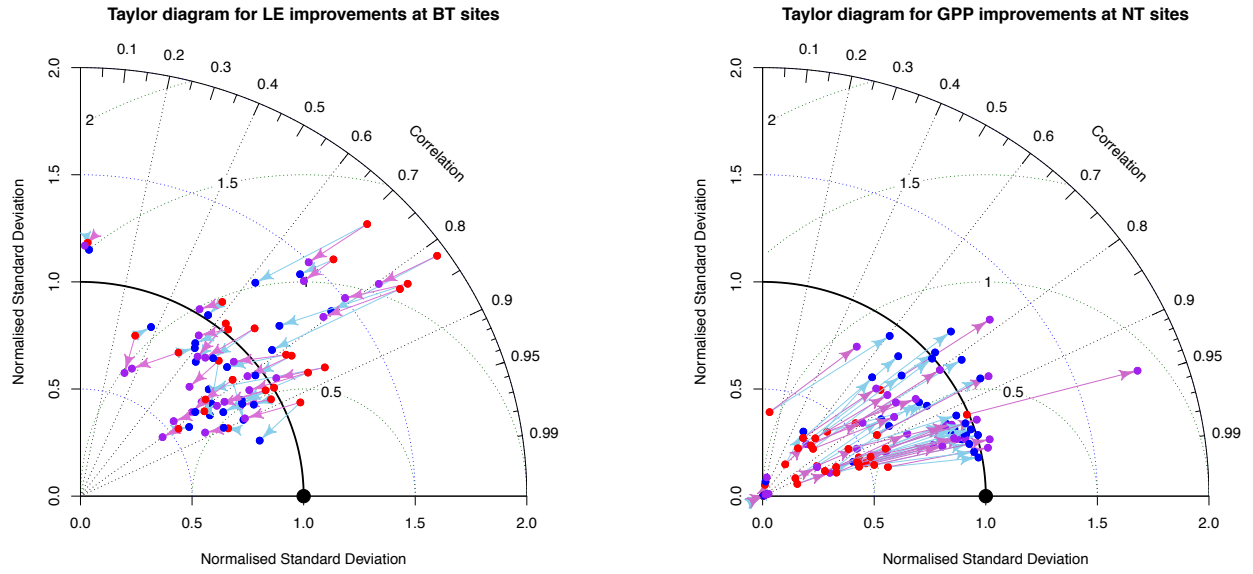


Figure 5. continued.



**Figure 6.** Improvements in fit represented by ‘Taylor diagrams’. Observed timeseries (black dot) can be compared with modelled timeseries for default parameters (red dots), site-specific optimal parameters (blue dots) and PFT-generic optimal parameters (purple dots). Radial distance from the origin (dotted lines) represents normalised standard deviation  $\sqrt{\text{var}(\mathbf{m}_t)/\text{var}(\mathbf{o}_t)}$ , and so a modelled time series with the correct variance lies on the thick black line. Angular position represents the correlation between modelled and observed timeseries. The distance from the black dot (dotted green lines) represents the normalised standard deviation in the errors  $\sqrt{\text{var}(\mathbf{o}_t - \mathbf{m}_t)/\text{var}(\mathbf{o}_t)}$ .

**Table A1.** FLUXNET sites used in this study, 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	Calibration Year	Validation Year	Latitude	Longitude
<b>Broadleaf sites (BT)</b>					
DE-Hai	(2000,2006)	2005	2004	51.079	10.452
DK-Sor	(1996,2006)	2006	2004	55.487	11.646
FR-Fon	(2005,2006)	2006	-	48.476	2.780
FR-Hes	(1997,2006)	2003	1998	48.674	7.065
IT-Col	(1996,2006)	2005	2001	41.849	13.588
IT-LMa	(2003,2006)	2006	2004	45.581	7.155
IT-Non	(2001,2006)	2002	2003	44.690	11.089
IT-PT1	(2002,2004)	2003	2004	45.201	9.061
IT-Ro1	(2000,2006)	2006	2005	42.408	11.930
IT-Ro2	(2002,2006)	2004	2006	42.390	11.921
UK-Ham	(2004,2005)	2005	-	51.121	-0.861
UK-PL3	(2005,2006)	2006	-	51.450	-1.267
US-Bar	(2004,2005)	2005	-	44.065	-71.288
US-Ha1	(1991,2006)	1996	1998	42.538	-72.171
US-MMS	(1999,2005)	2002	2003	39.323	-86.413
US-MOz	(2004,2006)	2006	2005	38.744	-92.200
US-UMB	(1999,2003)	2003	2002	45.560	-84.714
US-WCr	(1999,2006)	2005	2000	45.806	-90.080
AU-Tum	(2001,2006)	2003	2005	-35.656	148.152
AU-Wac	(2005,2007)	2006	-	-37.429	145.187
BR-Sa1	(2002,2004)	2003	2004	-2.857	-54.959
BR-Sa3	(2000,2003)	2002	2003	-3.018	-54.971
FR-Pue	(2000,2006)	2006	2005	43.741	3.596
ID-Pag	(2002,2003)	2003	-	2.345	114.036
IT-Cpz	(1997,2006)	2004	2006	41.705	12.376
IT-Lec	(2005,2006)	2006	-	43.305	11.271
PT-Esp	(2002,2004)	2004	2003	38.639	-8.602
PT-Mi1	(2003,2005)	2005	-	38.541	-8.000
<b>C3 grasses sites (C3G)</b>					
DE-Gri	(2005,2006)	2006	-	50.950	13.512
DK-Lva	(2005,2006)	2006	-	55.683	12.083
ES-LMa	(2004,2006)	2006	2005	39.941	-5.773
HU-Bug	(2002,2006)	2006	2005	46.691	19.601
HU-Mat	(2004,2006)	2006	2005	47.847	19.726
IT-Amp	(2002,2006)	2006	2005	41.904	13.605
PL-wet	(2004,2005)	2005	-	52.762	16.309
PT-Mi2	(2004,2006)	2006	2005	38.477	-8.025
US-Bkg	(2004,2006)	2006	2005	44.345	-96.836
US-FPe	(2000,2006)	2002	2004	48.308	-105.101
US-Goo	(2002,2006)	2006	2004	34.250	-89.970

**Table A1.** continued

Site	Period	Calibration Year	Validation Year	Latitude	Longitude
<b>Needleleaf sites (NT)</b>					
CA-Man	(1997,2003)	2001	2002	55.880	-98.481
CA-NS1	(2002,2005)	2004	2003	55.879	-98.484
CA-NS2	(2001,2005)	2002	2004	55.906	-98.525
CA-NS3	(2001,2005)	2004	2002	55.912	-98.382
CA-NS4	(2002,2004)	2004	2003	55.912	-98.382
CA-NS5	(2001,2005)	2004	2002	55.863	-98.485
CA-Qcu	(2001,2006)	2005	2006	49.267	-74.037
CA-Qfo	(2003,2006)	2006	2005	49.693	-74.342
CA-SF1	(2003,2005)	2004	2005	54.485	-105.818
CA-SF2	(2003,2005)	2004	2005	54.254	-105.878
CA-SF3	(2003,2005)	2005	2004	54.092	-106.005
DE-Bay	(1996,1999)	1999	1998	50.142	11.867
DE-Har	(2005,2006)	2006	-	47.934	7.601
DE-Tha	(1996,2006)	2005	2004	50.964	13.567
DE-Wet	(2002,2006)	2006	2004	50.453	11.457
ES-ES1	(1999,2006)	2005	2000	39.346	-0.319
FI-Hyy	(1996,2006)	2006	2004	61.847	24.295
FR-LBr	(2003,2006)	2006	2005	44.717	-0.769
IL-Yat	(2001,2006)	2005	2006	31.345	35.051
IT-Lav	(2000,2002)	2001	2002	45.955	11.281
IT-Ren	(1999,2006)	2005	2006	46.588	11.435
IT-SRo	(1999,2006)	2006	2005	43.728	10.284
NL-Loo	(1996,2006)	2006	2003	52.168	5.744
RU-Fyo	(1998,2006)	2005	2006	56.462	32.924
RU-Zot	(2002,2004)	2003	2004	60.801	89.351
SE-Fla	(1996,1998)	1998	1997	64.113	19.457
SE-Nor	(1996,1999)	1997	1999	60.086	17.480
SE-Sk2	(2004,2005)	2005	-	60.130	17.840
UK-Gri	(1997,1998)	1998	-	56.607	-3.798
US-Blo	(1997,2006)	2006	2000	38.895	-120.633
US-Ho1	(1996,2004)	2004	2003	45.204	-68.740
US-Me4	(1996,2000)	2000	-	44.499	-121.622
US-SP1	(2000,2001)	2001	-	29.738	-82.219
US-SP2	(1998,2004)	2001	2004	29.765	-82.245
US-SP3	(1999,2004)	2001	2002	29.755	-82.163
<b>Shrubs sites (Sh)</b>					
CA-Mer	(1998,2005)	2004	2005	45.409	-75.519
CA-NS6	(2001,2005)	2003	2004	55.917	-98.964
CA-NS7	(2002,2005)	2003	2004	56.636	-99.948
IT-Pia	(2002,2005)	2003	2004	42.584	10.078
US-Los	(2001,2005)	2005	2003	46.083	-89.979
<b>C4 grasses sites (C4G)</b>					
BW-Ma1	(1999,2001)	2000 <sup>32</sup>	2001	-19.916	23.561
ZA-Kru	(2001,2003)	2002	2003	-25.020	31.497

**Table B1.** PFT-specific JULES parameters optimised in this study (Table 1). The prior values and ranges for each PFT are given. Below in bold are the optimised values and posterior uncertainty ranges given as an 80% confidence interval (in parentheses). Optimised values for which the prior values lie outside the 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)
	<b>0.061</b>	<b>0.065*</b>	<b>0.07</b>	<b>0.051*</b>	<b>0.041</b>
	<b>(0.034,0.066)</b>	<b>(0.059,0.07)</b>	<b>(0.018,0.145)</b>	<b>(0.043,0.056)</b>	<b>(0.006,0.066)</b>
$\alpha$	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)
	<b>0.131*</b>	<b>0.096</b>	<b>0.179*</b>	<b>0.118*</b>	<b>0.102</b>
	<b>(0.087,0.14)</b>	<b>(0.021,0.167)</b>	<b>(0.155,0.209)</b>	<b>(0.075,0.141)</b>	<b>(0.063,0.763)</b>
$f_0$	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)
	<b>0.765*</b>	<b>0.737*</b>	<b>0.817</b>	<b>0.765*</b>	<b>0.782*</b>
	<b>(0.655,0.787)</b>	<b>(0.713,0.758)</b>	<b>(0.727,0.944)</b>	<b>(0.752,0.793)</b>	<b>(0.735,0.848)</b>
$T_{\text{low}}$	0	-10	0	13	0
	(-50,40)	(-50,40)	(-50,40)	(-50,40)	(-50,40)
	<b>1.203</b>	<b>-8.698</b>	<b>-1.985*</b>	<b>11.37</b>	<b>-5.208*</b>
	<b>(-0.555,9.492)</b>	<b>(-10.98,-6.342)</b>	<b>(-3.877,-0.13)</b>	<b>(7.522,14.072)</b>	<b>(-10.855,-2.106)</b>
$T_{\text{upp}}$	36	26	36	45	36
	(25,50)	(25,50)	(25,50)	(25,50)	(25,50)
	<b>38.578*</b>	<b>34.721*</b>	<b>36.242</b>	<b>44.897</b>	<b>35.385</b>
	<b>(38.157,40.698)</b>	<b>(33.214,36.365)</b>	<b>(33.087,38.599)</b>	<b>(44.201,46.426)</b>	<b>(26.339,40.216)</b>
$d_r$	3	1	0.5	0.5	0.5
	(0.1,4)	(0.1,4)	(0.1,4)	(0.1,4)	(0.1,4)
	<b>3.009</b>	<b>1.425*</b>	<b>0.991*</b>	<b>0.404*</b>	<b>0.411*</b>
	<b>(2.901,3.052)</b>	<b>(1.159,1.672)</b>	<b>(0.901,1.101)</b>	<b>(0.5,3.623)</b>	<b>(0.324,0.473)</b>
$\frac{\delta c}{\delta t}$	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)
	<b>0.047*</b>	<b>0.045*</b>	<b>0.05</b>	<b>0.05</b>	<b>0.048</b>
	<b>(0.046,0.049)</b>	<b>(0.042,0.048)</b>	<b>(0.047,0.052)</b>	<b>(0.046,0.054)</b>	<b>(0.04,0.055)</b>
$dq_c$	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)
	<b>0.048</b>	<b>0.036</b>	<b>0.086</b>	<b>0.046*</b>	<b>0.077</b>
	<b>(0.02,0.183)</b>	<b>(0.008,0.066)</b>	<b>(0.07,0.109)</b>	<b>(0.045,0.053)</b>	<b>(0.024,0.118)</b>