

Interactive comment on “Land surface parameter optimisation through data assimilation: the adJULES system” by Nina M. Raoult et al.

Anonymous Referee #1

Received and published: 24 February 2016

The manuscript “Land surface parameter optimisation through data assimilation: the adJULES system” by N. Raoult et al. presents a parameter calibration system for the JULES land surface model based on a 4D-Var approach employing the adjoint of JULES. The data to be assimilated are eddy covariance measurements of GPP and LE as provided by FLUXNET. The authors also compared the performance of single-site optimisation against multi-site optimisation for the 5 JULES PFTs. The multi-site optimisations (they call it ‘PFT-specific’) improves the fit against the assimilated data for over 90% of the sites used in the optimisation, and for about a third of the sites performs as well as the single-site optimisation.

The manuscript addresses an important topic in Earth System modelling, i.e. the objective calibration of model parameters against observations and reduction of uncertainty for the land surface component of the UK Met Office climate and weather forecast

C1

models. Such objective calibration systems for Earth System Model land surface components are still in their infancy, and thus not routinely performed, which is one of the reasons for the wide spread in the latest climate model intercomparison results especially for the terrestrial carbon cycle components. Hence, this contribution is very timely. Unfortunately, the manuscript suffers from many inconsistencies and lack of precision which makes it sometimes a rather awkward read. The manuscript would have certainly benefited from careful proof reading by all (co-)authors before submission.

One example of this is the authors’ use of the term ‘adJULES’. Sometimes they refer with adJULES to the adjoint of JULES and sometimes to the whole optimisation system. The two are certainly very different and as such should also be clearly distinguished in the manuscript. There is an established terminology in the data assimilation community and it would improve the readability if the authors would use this terminology, e.g. ‘posterior’ instead of ‘new’ parameter.

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

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

C2

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

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

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

P4 L19: How does lambda enter Eq 1?

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

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

P6 L1: The data selection criteria should be specified exactly. What does 'significsnt gaps' mean. There is also the danger of introducing biasas by certain data selection criteria. This should be taken into account.

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

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

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

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

C3

P6 L31/32: Please rephrase. The adjoint does not find the second derivative.

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

P7 L11-17: This is an interesting way to calculate the posterior parameter uncertainties, but it is not clear why and what exactly you do there. What is the advantage of using this method over calculating the posterior uncertainties from the inverse of the Hessian directly? When you calculate the full Hessian you also get the full error covariance matrix! Do you get a semi-definite Hessian (see also general comment on obtaining a minimum)?

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

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

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

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

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

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

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

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

C4

P10 L4: The sentence need to rephrased.

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

P10 L8: What are these sets?

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

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

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

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

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

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

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

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

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

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

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

C5