Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2015-281-AC1, 2016

Geoscientific Model Development Discussions

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

Nina M. Raoult et al.

nr278@exeter.ac.uk

Received and published: 2 May 2016

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

1 General comments:

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

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

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

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

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

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

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

but are more computationally costly."

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

Given the small number of sites available to us, we decided to use all available sites in finding the multisite parameters sets. Sites used in this study required at least two consecutive years, one to spin-up the model and one to calibrate against. Re-examining the data, we found that the majority of the sites had more than two years and so a different year could be used to validate the optimised set of parameters. The year used for validation was chosen to be the second most complete, the first most complete having been used in the calibration. The results of the validation, which are very positive, are now shown in the results section alongside the results for the calibration.

2 Specific comments:

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

Corrected: now defined on P2 L31.

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

Factor added.

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

Removed "observed" from the sentence.

P4 L19: How does lambda enter Eq 1?

An explanation has been added.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The purpose of section 3.2 was to shows that given a set of 5 randomly selected sites, the optimised parameter vector found by optimising over these sites also improves the

rest of the sites not used in the calibration. This experiment is now obsolete since we have the ability to validate the PFT-specific parameters properly in our improved result section. As a result, this section has been removed.

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

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

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

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

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

The site identification code has been added to the plots and the units moved from the top of the figure to the side of each individual panel for clarity. The sites picked were the ones that captured best the general trends for each of the PFTs. This was done manually.

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

As our focus is on the carbon cycle, we choose to only optimised parameters directly

relating to the photosynthesis equations in the JULES model. Given more time and more computing power, more parameters could be used in the optimisation.

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

Corrected.

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

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

P10 L4: The sentence need to rephrased.

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

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

No longer relevant with the removal of section 3.2.

P10 L8: What are these sets?

No longer relevant with the removal of section 3.2.

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

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

"For some sites, US-Blo and BW-Ma1 for example, the PFT-generic parameter vector over-performs the parameter vector found locally. This phenomenon was also noted in Kuppel et al. (2014). The study further suggested that the added simultaneous constraints placed on the parameters by increasing the number of sites used in the

cost function caused the cost function to become 'smoother' and so the optimisation scheme is less likely to get stuck in local minima."

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

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

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

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

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

Corrected.

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

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

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

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

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

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

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

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

Issue addressed in response to general comments.

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

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

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

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

Fig 6: What is the difference between top and bottom panel and what to the

vertical lines denote? What are the outliers that have been removed and why did you remove them?

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

References:

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

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

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

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

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2015-281, 2016.