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

1 General things:

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

We acknowledge that standard practice in data assimilation is to split the data into

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

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

This is now discussed more thoroughly the conclusion.

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

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

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

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

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

Lambda added to equation.

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

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

"Preliminary experiments showed very narrow uncertainties whilst running the optimisation scheme over multiple sites i.e. the background term was found to dominate the cost function. In previous multisite studies (Kuppel et al., 2012, 2014), the prior range was also used to defined the background covariance matrix B. The range was variously further multiplied by a factor of 40% (Kuppel et al., 2012) and 1/6 (Kuppel et al., 2014). Experiments were run to find a similar factor to use in this study (the constant of proportionality in Eq. 5). In each of the multisite experiments, the lowest value of

such that the Hessian is positive definite at the optimal parameter value was used. This allows uncertainties to be generated around each parameter and prevents the gradient descent algorithm from reaching the boundaries of the prescribed prior range."

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

The parameters are assumed to start off uncorrelated. Text added: *"...The matrix $\mathbf{B}$ 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."*

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

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

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

(1)

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

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

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

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

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

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

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

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

is a single dimensionless measure of model misfit. The fractional error ϵ can than be interpreted as the typical (root-mean-square) error expressed as a fraction of the (root-mean-square) magnitude of the observed seasonal cycle. Thus, $\epsilon = 0$ represents a

perfect match to the observations, while $\epsilon = 1$ corresponds to the error in a null model whose prediction $\vec{m}_{i,t}$ always equals the observational mean $\vec{\bar{o}}_i$. "

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

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

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

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

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

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

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

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

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

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

Figure 6 - I'd make the lower y-axis bound 0.0 (rather than what looks like 0.001)?

Gives a better feel then of the improvement in absolute terms.

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

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

The conclusion has been expanded to emphasise this point.

2 Small things:

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

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

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

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

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

Text added.

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

Text added.

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

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

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

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

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

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

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

Added.

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

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

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

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

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

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

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

Text added to picture.

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

Of the 500 FLUXNET sites, we were only able to obtain 160. Of those, 50% were rejected. As well as the criteria listed in this section, sites were also excluded based on number of years available (2 years minimum in order to perform a spin-up) and whether

the sites were dominated by particular PFT (e.g. more than 50% coverage with the exception of C4 grasses). The crop sites were excluded since in the newer versions of JULES these are considered separately. These extra exclusions are described in section 2.4. This is made explicit in the paper using the following text: *"Data from 160 sites were made available for this study by M.Groenendijk."*

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

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

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

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

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

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

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

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

C11

No longer relevant with the removal of section 3.2.

References:

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

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