Review of Automated model optimisation using the Cylc workflow engine (Cyclops v 1.0).

Simon Tett

I think a potentially interesting paper that should eventually be published. The paper describes a method to use generic optimisation methods to optimise a wave model. In theory the approach could be used for other models though the paper does not really describe the challenges involved in doing this.

I worry the paper is quite close to be a minimum publishable unit and so I am pushing the authors to do it more work. In essence to show their approach does indeed work. To that end I ask that the authors trial two or more additional algorithms. For purely selfish reasons I would be interested in seeing results of the Gauss-Newton approach trailed in Tett et al, 2013 & Tett et al, 2017. However, I understand that the algorithms available to the authors through the NLopt toolkit do not include this. I think the study would also benefit from doing another study in which they started from extreme parameters and see if they end up in the same local optimum or some other one.

The authors do not really deal with the challenge of interfacing the optimisation algorithm to the model. Simply telling us that they generate a simple namelist which gets passed through to the wave model is insufficient detail. I think it would also help the reader if they provided a bit more detail on how the set of previous cases (and cost function values) are passed around. I've done something similar for HadAM3 and much of the effort was in modifying the model namelist variables. HadAM3 has many namelists, each with several variables spread across a few files.

The authors should describe how concurrency happens. I suspect it depends on the optimisation algorithm. If they found a good solution to that that is worth sharing.

One issue that worried us in Tett et al , 2017 was the effect of noise in the optimisation algorithm. If the evaluations needed to fit the 2nd order polynomial in BOBYQA are too close to one another then the difference will largely be chaotic noise. How does the authors approach mitigate against that?

Minor comments

P1, L15 – I don't think the URL belongs in the abstract.

L21 – don't think TM belongs in the abstract (and the text uses (R)).

P2, L10. Note that Roach et al used the system described in Tett et al, 2017.

P2, L12 – I personally don't like 1 sentence paragraphs. Can this sentence be wrapped into the following or preceding paragraph?

P3, L24 read -> reads

P4, l6 A bit more detail on how Cyclops tasks interact would be useful as I don't see a peer reviewed paper describing it. As the optimisation is implemented with special messages being sent some more discussion on messages would be helpful.

P4, l12 interleaveseveral – insert some spaces

P5, line 14 – agree for cases where cost function is some squared difference then –ve values are reasonable. However, I think in the python world returning None to signal need to generate values would be more natural.

P5, L15 – more detail on how the namelist is generated would be helpful. Looking at the code it looks like the text is simply generated. My experience with the Unified Model is that with multiple namelists in multiple files there is a bit of setup to be done to map optimisation variables to namelist variables (in some cases one optimisation variable can modify multiple namelist variables.) Some models may not use namelists so what would be done in this case?

P5, L25 -- Some more detail on how Cycl iterates would be helpful. I think being explicit (and showing how) that Cycl can run several jobs in parallel would be helpful. I think discussing that in the context of the algorithms would also be helpful. I think many algorithms are coded to work serially so won't make use of the ability to run several model simulations in parallel. But clearly authors report doing this so a bit of discussion would help here.

P7, L7 – cite for the model please and don't see the need for the (R)… But I leave it to GMD editors to decide that.

P7, L35 Note this such a cost function (spatial average RMSE) gives high weight to shortest spatial features which are close to model grid scale and thus very likely strongly affected by model grid and chaotic variability. This is one reason Tett et al, 20113 & 2017 focused on RMS error of large spatial averages. It is a mystery to me why people continue to focus on spatial average RMSE for model evaluation given the smallest scales are dominated by chaotic variability and thus not strongly related to parameter choice or model fidelity.

P8, L2 – can this be typeset larger – probably display would help. Does the dot mean d/dt? If so I think better to spell it out.

P8, line 35 – surely not **zero** impact. Imagine it is very small.

P8 – I found the discussion on the two different packages rather confusing. The authors should rewrite to make this clearer.

P9, L9 – why 0.02 rather than 0.05 or 0.01? Would algorithm terminate if any parameter changed by less than 0.02 or would all need to have changed by less than 0.02?

P9, L11 – why introduce two more parameters?

P9, L24 a bit more discussion about parameter sensitivity here would be useful. For which parameters is the cost function most sensitive?

Table 2 would benefit from some description of the parameters – what do they represent? I don't think readers need to know about "n". It is an implementation detail. Table should also explain what the bold labels are – perhaps better to break up into multiple tables with titles given by meaning of bold labels.

Tables 3&4 – only show parameters that were modified. This would reduce the size considerably and make them less confusing.

Figure 1 – text is small and unreadable (and I don't think the colour is necessary). I suggest just showing one iteration of the work flow with some arrows showing the work flow looping back.

References

Simon F. B. Tett, Michael J. Mineter, Coralia Cartis, Daniel J. Rowlands, and Ping Liu. Can top of atmosphere radiation measurements constrain climate predictions? part 1: Tuning. J. Climate, 26:9348–9366, 2013. doi: 10.1175/JCLID-12-00595.1.

Simon F. B. Tett, Kuniko Yamazaki, Michael J. Mineter, Coralia Cartis, and Nathan Eizenberg. Calibrating climate models using inverse methods: Case studies with HadAM3, HadAM3P and HadCM3. Geoscientific Model Development, 10:3567–3589, September 2017. doi: 10.5194/gmd-2016-305.