Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-56-RC2, 2020

Geoscientific Model Development Discussions

Interactive comment on "Necessary conditions for algorithmic tuning of weather prediction models using OpenIFS as an example" *by* Lauri Tuppi et al.

Joakim Kjellsson (Referee)

jkjellsson@geomar.de

Received and published: 13 September 2020

Overall:

The manuscript describes and evaluates the usability of algorithmic tuning for weather prediction models, in this case the convection scheme in OpenIFS. The manuscript uses two methods although only one is discussed and tested in detail. I'm curious about the method, in particular when applied to low-resolution versions of atmosphere models, since it could be a good way forward to improve climate models. Given that OpenIFS is used in a few climate models and that similar convection schemes are used in other models, this paper can be of interest also outside the weather forecasting

community.

I recommend this paper for publication after some comments below have been addressed.

Major comments:

The paper is well structured and the results interesting. The writing needs some work, in particular the descriptions of experiments and the latter half of the results section.

I also feel that the paper should clarify right from the start that any closure scheme, in this case convection, is only a parametrisation of the real world, i.e. there is no "true" parameter value.

The choice of 850 hPa geopotential as a cost function seems odd and the authors also state that they expected this to be a bad choice. While its good to compare a bad choice to a good choice, I think it would make more sense to have used something else, e.g. precipitation or total-column water etc. I don't think all the experiments need to be re-done to accommodate this, but I'd like to know exactly why 850 hPa geopotential was chosen and if the authors think some other pressure level or single variable would be more appropriate.

Minor comments:

Overall: The authors use "Firstly" and "Secondly" a few times in the paper. This should be replaced by "First" and "Second".

Line 15: A closure approximates a complex process with an equation and a parameter. As an example, we approximate diffusion/mixing with some high-order derivative and a coefficient. It don't think it's correct to say "the parameter values are not known exactly" because they are only approximations anyway. There is no magic value of e.g. autoconversion rate that will result in a perfect cloud scheme. What you are looking for is the optimal parameter, i.e. the one that results in the smallest error.

Line 45: How about time of year, i.e. is there a seasonal dependence of the optimal parameters. Also, how are the optimal parameters dependent on model resolution, if at all?

Line 54: I'm a bit confused as to the number of forecasts here. Each test is a 51 member ensemble, done for 52 weeks in a year. And Fig 1 looks like you are using 52 iterations. Does this mean 51 members x 52 weeks x 52 iterations > 100 000 forecasts? Or is it 51 members and each weeks is an iteration, i.e. 51 x 52 2600 forecasts? If it is the latter, then your cost function for early iterations are based on how well OpenIFS simulates winter conditions, while iterations 30 are based on how well OpenIFS simulates summer. I could imagine that the optimal parameter and the convergence are quite different in different seasons, so you would probably want to estimate the optimal parameter set for each season separately.

Line 48: Could the authors please add a reference to a paper or documentation for OpenIFS.

Line 54 (contd): I would also really like to see an estimate of the computational cost of these optimisation experiments, e.g. number of core-hours and total number of simulated days.

Line 98: Is "dD = dx * dy"? Then why not use that? Or "dA" for area. Would make more sense.

Line 105: I'm confused about this equation. The individual terms seem to have different units. u^2 is m2/s2, while $L^2q^2/(cpT)$ would be in Joule, and the last term I think is in [kg/m3]?

Line 106: In eq (1), primes denote an individual ensemble member. In (2), primes denote the difference between ensemble member and pseudo observations. First, it is confusing that primes denote two different things on the same page. Second, please explain what pseudo-obs are.

Line 120: What is "n"? Is it an index over all tunable parameters?

Line 149: How do the authors measure convergence? Is it done using parameter mean or uncertainty? A "parametric" way could be to smooth curves of mean parameter values and then see where the derivative $\partial\theta/\partial N$ approaches zero, where N is iteration number. Or do you simply judge this by eye here? I agree that L0 seems to converge the fastest (mean stabilises the quickest and uncertainty decreases quickest). On the other hand L1, L2, L3 converge equally fast in terms of the mean parameter value but L1 has the largest uncertainty in Fig 1b.

Line 151 and 160: What is meant by "fully-realistic"? L3 should be the most "realistic" in the sense that it better represents the full complexity of the entire OpenIFS model. "Fully realistic" would be optimising all parameters (probably > 100) in OpenIFS with SPPT?

Line 166: again, "fast convergence" based on what metric?

Line 166: How about changing the sentence to "is explained by two factors. First ... " followed by "Second ...".

Line 171: I would remove this sentence and then start the next sentence with "... recommend using a more comprehensive cost function which accounts for more than one... ".

Line 174: Since you are tuning convection parameters it would be a more fair comparison to use a ΔZ taken at 200 or 300 hPa. I'm not saying it needs to be done, but it is worth noting that it would be a better comparison.

Line 176: I would change the first sentence to "The process of finding a forecast range ... " and then merge with the next sentence so that it ends "of computational resources is done in two steps." The next sentence would be "First ... forecast ranges. " followed by "Second we take the forecast range ... ensemble size. ".

Line 179: Merge "Using L2 realism" with either the previous or following sentence.

Line 182: I'm struggling to understand what is done here. Do you take the final converged state from Fig 1 c,d, then set parameters to default, and compare how the estimated optimal parameters compare to those you got from Fig 1 c,d? Or do you run a convergence test and compare to the default parameter value as in eq 3? If you are doing the former, please explain this more clearly. If you are doing the latter, then why is low bias good? A large bias could mean that the tuning process worked really well and you got a more optimal parameter value.

Line 186: Could it also be due to the fact that you are focusing on convection parameters and that convection is short lived? If you run 72 hours then you get an influence of synoptic variability which adds extra complexity. If you were tuning orographic wave drag you might want longer forecasts?

Line 194: Change to "We now focus on the forecast range of 24 hours and perform convergence tests with ... ".

Line 195: Change to "Fig 4 indicates that convergence tests with ensemble size > 20 are stable since convergence tests with smaller ensemble sizes do not show... ".

Line 197: What is meant by the two sentences strarting with "sampling variance". What is "sampling variance" here?

Line 198: Change to "Fig 4 also enables comparison ... "

Line 201: Change to "not seem to be necessary to achieve good convergence. "

Line 202: Change to "Results here are for θ_2 but the same conclusions can be drawn from θ_1 ... "

Line 211: I would remove this sentence.

Line 213: The black box In Fig 3 is clearly visible, but the gray one is not. Please use some other colours, e.g. green and yellow, to make sure the boxes stand out. Also explain in the caption for Fig 3 what the boxes are.

Line 217: Rewrite to "shows that both L1 set-up s yield fairly reproducible convergence.".

Line 224: L1 does not use SPPT, i.e not stochastics in OpenIFS. So why are A1-A4 not identical? Are the initial conditons perturbed differently each time, or is there some other stochastic component activated?

Line 226: The language switches from present to past tense here. Please commit to one throughout the paper.

Line 229: I would rewrite sentence to "The opposite is true for θ_1." Also, I would like to see the plots of this (like Fig 1) in a supplementary materials or appendix.

Line 231: I would replace "(to the left)" by Fig 6a and similarly for the other panel. Then I would add letters a,b to the two subplots.

Line 231 232: Words "especially" and "actually" are superfluous and should be removed.

Line 232: It is interesting that θ_2 depends so strongly on forecast length, but it is also interesting how it is not so sensitive to ensemble size. Could this be an effect of not using SPPT so that the system lacks spread? Also, why is θ_1 results all over the place? Could the authors comment on this?

Line 232: I think the rest of this paragraph needs a good rewrite. Here I would rephrase to something like "We now examine the cost functions for two sets of ensemble forecasts (please specify length), one using the default parameters and one using parameters obtained from the optimisation (Fig 6)."

Line 236: Rewrite to "Results show that globally optimal parameter values are different from the their respective default values ... "

Line 241: The sentence includes a statement "we are unsure whether .. " but ends with a question "does the dependency... ". Split the sentences.

Line 241: This is a complete speculation from me: Could it be that you have found parameters that optimise the inter-ensemble spread, i.e. the atmospheric states in the ensemble with optimised parameters might be different from the control forecast, but have small spread which could give a low cost function. For instance if the optimisation produces less convection...

Line 249: If you use "First" in a previous sentence, you should use "Second" here.

Line 250: "with respect to"

Line 250: Merge sentences.

Lines 253-256: This part is very vague. For instance, do you mean that EPPES and DE are both robust? Or is one robust while the other is not? What is excessively large? How can a user know this in advance?

Line 259: This makes sense if I'm only looking to optimise 1-2 variables. But if a forecasting centre wants to optimise their entire convection scheme (10 parameters) they would need to go for something larger.

Line 265: How are the five-parameter tests different from the L3 tests before?

Line 266: Which additional three parameters are used?

Line 267: Remove last sentence and make first sentence on line 263 be " ...demanding convergence tests with EPPES."

Lines 269-270: When you say 20 of 30 or 25 of 30, which ones do you mean? Do some variables converge to off-default variables more often than others? A plot or some more text would help here.

Line 270: "uncertainty" = "model spread" ?

Line 270: Do you compute cost function of difference to control run? So we would expect variables to converge to default values?

Line 266: How is the uncertainty chosen? Is this taken from previous optimisation runs, or do you use a guess?

Line 273: This sentence adds no information. Remove it.

Line 279: I disagree that all parameters converge. Parameter 8 does not seem to converge at all. Parameter 1 converges to some extent, but very slowly. How is convergence gauged? By eye? And by what metric, convergence of value or model spread?

Line 282: If it was expected to be bad, why not try another level? 300 hPa would have made more sense maybe, or using surface field like precipitation or total-column precipitable water etc.

Line 284: Could the authors please expand on this? Why are they not equally sensitive?

Line 295: Could it be that some parameters (e.g. convection which is fast) should be tuned with short forecasts while other parameters (e.g. gravity wave drag which is slower) should use longer forecast times?

Line 345: How do you guess the covariance matrix? If you start with a bad guess of the covariance, does that affect the results?

Table 1: It would also be good with a short description of how each parameter should impact the two different cost functions, which would help the argument on line 284.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-56, 2020.