Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-217-RC1, 2020

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "OpenIFS@home version 1: a citizen science project for ensemble weather and climate forecasting" *by* Sarah Sparrow et al.

Anonymous Referee #1

Received and published: 1 November 2020

The paper on OpenIFS@home is an interesting proof of concept paper. The paper neither goes into technical nor in scientific detail of the concept, but shows that producing coarse resolution ensembles OpenIFS on personal computers of citizens is possible and seems scientifically reasonable. As a model description paper for GMD I consider it too concise though.

This citizen science approach is limited to the use of external resources. The paper does not discuss how the citizens get involved, what there incentives are and whether there is feedback between the scientific community and the citizen science community.

However, I do applaud the approach to use OpenIFS on personal computers and in this way facilitate the generation of very large initialized ensembles. The application of OpenIFS is particular compared to earlier projects (climateprediction.net and

Printer-friendly version

Discussion paper

weather@home) in that it uses a code that is very close to the world-leading operational medium range prediction system. So, I think it is valuable to publish this and this specific application could be highlighted more.

The presented solution has strong limitations as well. The use of only one core limits the applicability to coarse resolution models due to memory constraints. While this is understandable it should be outlined clearly that this set up is only suitable for a limited set of scientific questions related to weather and climate issues. One also wonders whether this can be extended, given that multicore personal computers with different chips are omnipresent. It would be good have a discussion on the future scope related to this.

The scientific analysis is very limited. It is valuable to see some diagnostics for the 3 systems, coarse resolution with small and high number of ensembles and otherwise high resolution with small number of ensembles. The results show the limited applicability. Clearly resolution matters for the meteorological variables chosen and hence it is very hard to interpret the reliability of the 2000 member ensemble, although it is definitely an improvement compared to the coarse resolution, small ensemble.

The paper needs a bit more flesh in either science or technical implementation, and I think given the scope of GMD and the fact it is submitted as a model description paper it should be the latter. I would like to see more detail on the actual implementation of the software and the dependencies on the hardware and future development. So this needs more discussion on issues as efficiency, scalability, latency, energy use etc etc, so the performance of the code. This paper should include a more detailed comparison with climateprediction.net and weather@home. These could serve as a relevant benchmark, otherwise this paper is merely an introduction to a system. It would be really valuable to the scientific community to see how different, or similar, these applications are. Already in the abstract the innovation should be highlighted.

So my major comments are: 1) The paper needs to include more technical details and

Printer-friendly version

Discussion paper

diagnostics of this specific application of weather code in conjunction with BOINC and the application on personal computers. Currently, only Figure 2 gives some information on the timing, but there is no comparison to other systems (see also comment #2). There is hardly information on the performance of this particular code on the PCs (speed, energy efficiency, etc etc). 2) A benchmark misses, while there are good candidates. It would be good to compare the application with other known projects, such as weather@home and climateprediction.net 3) In the discussion extend the discussion to options and limitations for scientific use, both in terms of the type of questions that can and cannot be answered and the limits of not having the code itself. For instance, to what extend can a model with T159 horizontal resolution simulate cyclones and the extratropical transitions that such a cyclone goes through. The author show 'the power of large ensembles', but for this particular ensemble the resolution may be just as important. That is, a discussion on structural uncertainty misses and should be included as part of the discussion for what types of scientific questions can be addressed. The event chosen is in this regard interesting. 4) Related to the scientific use in comment #3, the statements on the simulated distributions, and in particular whether an event is part of the distribution, are not addressed with sufficient rigor. I suggest to perform a fit to the distributions, and a more in depth assessment of the probability of the event given each fitted distribution and whether one can actually state whether the distributions are different and whether the sampling is sufficient (the authors seem to indicate that the smaller ensemble is too small for that). Also, what is the impact of two different cycles of IFS? This likely affects the interpretation. 5) Extend the citizen science discussion. This a very limited approach to citizen science, only through using resources. What do the citizens get back from it? Is there interaction between the scientist and the citizen that leads to more insight in meteorology or more general in insight in science and social acceptance of science? Already in the introduction the authors state that the public awareness is raised, but is there proof of this?

Smaller remarks: The paper needs some more references. Already in the first paragraph of the introduction several projects are mentioned with no reference.

Citizen science and open science are strongly related. GMD is an open access journal and aims to have its data and code open. In this regards it is disappointing that the code has a license as it has. This is more a remark and the authors can hardly be blamed for this.

It would be good if more consistent use of persistent identifiers were used in the references of manuals and data. For instance for ECMWFs documentation, for the data that is stored at CEDA where the paper currently puts weblinks and email addresses which tend to break down quickly.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-217, 2020.