

Geosci. Model Dev. Discuss., referee comment RC3
<https://doi.org/10.5194/gmd-2022-141-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-141

Anonymous Referee #3

Referee comment on "Porting the WAVEWATCH III (v6.07) wave action source terms to GPU" by Olawale James Ikuyajolu et al., Geosci. Model Dev. Discuss.,
<https://doi.org/10.5194/gmd-2022-141-RC3>, 2022

General Comments:

The authors present a GPU acceleration of the source term part within the parallel context of the WW3 Framework. The work is innovative and important. However, there are some flaws within the theoretical approach and the tests that have been done when evaluating the new parallelization option.

In eq. (1) the wave action equation we have certain terms that are local parts (e.g. source terms and spectral advection), and global parts, which is the geographical advection. The geo. advection needs some parallel exchange either for the CD or the DD approach.

Now, when expanding this asymptotically using Amdahl's law for the given problem it can easily be seen that ultimately for a infinite number of computational cores the only cost that remains would be the parallel exchange since all other workload tends to zero.

Introducing now the communicators to the GPU, this would remain as well an overhead and add up to the global exchange of the advection part itself.

Now the most important question is how does the scheme scale for various constellation. Since the GPU layer was introduced the scalability analysis becomes twodimensional in terms of number of GPU and CPU. This question remain open in this paper even if the authors have sufficient acces to the needed computational resources. In the sense of the above also the quantification of the computation cost of the source terms is rather linked to give testcase constellation investigated in this paper.

In conclusion I think that the work is interesting and the implementation is an important topic for GMD but the work lacks in depth scalability analysis and therefore no final statement can be made in terms of efficiency. Especially, in the context that only 8 cores have been used from possible NSPEC cores within the CD approach.

In conclusion I think that much more work must be done to evaluate the performance of this approach before the given conclusion can be made and the general contribution of the work can be evaluated, which is now not the case.

Specific Comments:

- 94: "is the most computationally intensive part of WW3" Can this be quantified? Moreover, I can not agree on this, since this would be rather linked to the given configuration and the used schemes. For implicit schemes and various other constellations with high resolution geographical space this must not be true. I think that too much general statements have been derived by the given configuration and testcase. I could imagine that with high spatial resolution and a lot of computational cores, which have unfortunately not been used here, the communication itself will take more time than the computation of the source terms itself, as explained in the general comments above.

The authors are using CD for their parallelization strategy, but I do not see why one could not use domain decomposition in combination with GPU. How would the scheme perform with DD approach? Why so few computation cores. Why so much development and such little evaluation of the performance?

The authors have limited their simulation to just 8 processors, however, I do not see why it could not be used on say 1000 processors if each of those processors has access to a GPU. This cannot be repeated often enough. It remains the major concern of this work. The Kodiak and Summit supercomputer have several thousand processors. Why is the limit to 8?

The interaction between explicit and implicit computations is not considered. I would think that the code can be used for explicit and implicit and the scalability should be evaluated for both.

Further code moving to the GPU could be the frequency shifting and refraction in explicit

mode, has this been considered as well?

Technical issues:

- There is a typo in Equation (1), the "+" is missing
- The literature reference in the paper is not done properly. Here are two examples:
- 19: Either refer to the paper with „e.g.“ or put the original reference, which is not Hasselmann 1991. It was Gelci et al. 1957 if I remember right.
- 146: Again, Chawla et al. Was referenced but those authors did not derive the scaling of 1.1. The original publication should be cited Hasselmann & Hasselmann, 1981 or it can be add „e.g.“. However, the latter I do not find very useful since we want to honor and cite the original work. This should be cleaned throughout the paper.

Missing reference:

- 104: Brus et al., 2021 reference is missing.