

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2022-141-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-141

Anonymous Referee #1

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-RC1>, 2022

This manuscript presents results from a modified WAVEWATCH III code, which off-loads the spectral source terms to GPUs. The authors investigate the scaling and two different platforms, and validates the results against a CPU-only run.

This is a very timely, important and well written manuscript. I can recommend that it is accepted for publication after some minor changes. I also want to mention, that the way the problem is broken down and presented step by step made the manuscript easy to follow. Please find my comments and questions below:

lines 24-25: "despite the growing literature on their in the simulation of weather and climate."

Seems to be missing a word

line 71: I'm not sure what this is a reference to, but usually Komen et al. 1994 is used as a WAM reference?

line 86-88: These sentences are slightly confusing, since first we are talking about modules that calculate source terms (right hand of Eq. 1), and then we talk about discretizing (left hand side).

line 94: Can the relative computational intensiveness change if we are using different propagation schemes. Can extremely small time steps alter/tip this balance? (Probably not, I guess.)

line 188-189: I'm a bit surprised that S_{in} takes half of the source term computational time, since I would have expected the non-linear interactions to be the heaviest. The cumulative breaking term in ST4 is supposed to be quite resource consuming, while possible not having a large effect on the end results. Is that turned on or off here? If it's turned on, then that would perhaps be an obvious candidate to try to speed up the model (not directly related to GPU porting).

line 196-198 Is this a hard requirement, or does the lack of communication just mean that the source terms are trivially parallelizable? The word "suitable" suggests that any communication here would make GPU porting a non-option, but I'm not sure that is the case (although it probably becomes a lot more complex).

line 241-266: very long paragraph. Even though the paper is generally very well written, this aspect could be checked.

line 275: It seems like the order the figures are presented might be wrong, since Fig. 8 has already been referenced?

line 295: Would it be possible to increase occupancy by reorganizing the loops? Now we loop over all grid points, and then loop over one spectrum, but perhaps it would be more efficient to define an array that has both spatial and spectral dimensions (and perhaps slice that up into some blocks, if needed)? Can you comment on this?

The paper is missing a discussion section. Although it might not strictly be needed in this kind of more technical paper, it would perhaps be interesting for the reader to know what kind of impact these speed-ups might have in practical terms. Several days of wall time was mentioned, but is this a "game changer" to allow for including wave models in ESMs, or do we still need to optimize? I'm also wondering how well the exact non-linear solution might scale (if the authors can comment), since this might have consequences to very basic research into e.g. wave growth that might be affected by the crude approximations of DIA. Finally, would it every be viable to port any other parts of the wave model, such as the propagation, to GPUs, or is the communication needed between the grid points a complete deal breaker?