Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-301-RC3, 2018

Geoscientific Model Development Discussions
Open Access

Interactive comment on "Simulation of the Performance and Scalability of MPI Communications of Atmospheric Models running on Exascale Supercomputers" *by* Yongjun Zheng and Philippe Marguinaud

Anonymous Referee #2

Received and published: 4 May 2018

The article presents an important aspect often ignored in NWP model development. It studies the impact of network topology, not only for one particular algorithm, but for multiple representative algorithms found in NWP models. It illustrates that the choice of equivalent but different numerical algorithms may well depend on the available network layout. In this case a semi-Lagrangian approach using nearest-neighbour communication for wide halo-exchange is studied. Further, a spectral transform method is studied consisting of large distributed matrix transpositions, and finally a Krylov solver consisting of multiple AllReduce operations. The results are presented in a detailed yet clear

manner.

I have attached a edited PDF of the original article containing comments and suggestions. If these are addressed, I am happy to see the article published.

For clarity, I will report the major comments and questions below (besides being present already in the attached PDF).

1) Throughout the article the term "radix" is used. It would be good to formulate a definition of it in this paper's context.

2) Line 135: I recommend following more representative citation instead of "Kuhnlein et al., 2017": Smolarkiewicz et al., 2016: A finite-volume module for simulating global all-scale atmospheric flows, J. Comput. Phys., 314, pp. 287-304, doi:10.1016/j.jcp.2016.03.015

3) Line 363: It's worth noting that this regularity is only possible for structured grids. Even then there are differences between regular and reduced grids. Unstructured grids would not have a preferred x or y sweeping, and communication must be done in a single sweep. Does the following analysis still hold in this case?

4) Line 603, whole paragraph: Can MPI tasks be carefully pinned to cores using knowledge of the domain decomposition to reduce congestion?

5) Line 639: "However, the bandwidth of memory limits the performance and scalability of computation for multi-core or many-core systems". This statement seems taken without reasoning. Surely this cannot apply to any algorithm. Could the authors elaborate?

6) Acknowledgements: The Horizon 202 program ESCAPE acknowledgement has more strict rules on how to acknowledge (e.g. mention of EU and program number). I recommend asking the project manager for details.

Please also note the supplement to this comment:

https://www.geosci-model-dev-discuss.net/gmd-2017-301/gmd-2017-301-RC3-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-301, 2018.