Geosci. Model Dev. Discuss., 8, C2470–C2472, 2015
www.geosci-model-dev-discuss.net/8/C2470/2015/

Geoscientific
Model Development
Discussions
Open Access

Interactive comment on “Towards convection-resolving, global atmospheric simulations with the Model for Prediction Across Scales (MPAS): an extreme scaling experiment” *by* D. Heinzeller et al.

D. Heinzeller et al.

heinzeller@kit.edu

Received and published: 13 October 2015

First we would like to thank Anonymous Referee #1 for the evaluation of our manuscript and for the valuable comments, suggestions and corrections. While we will be waiting for the comments of referee #3 before providing an updated version of our manuscript, we would like to answer to the questions and concerns raised by referee #1:

(2) Specific Comments

Referee #1 comments "The ’number of cells owned by each task’ as a robust metric

for the scalability limit does not appear similarly convincing to me". We admit that the explanation of the transition zone can and will be improved in the manuscript. We suggest to alter the definition of the transition zone such that it describes the point at which the absolute runtimes reach a minimum and increase again for larger numbers of tasks. For this definition, there is indeed a fairly robust number of 150 cells owned per task for all three test cases in Section 2 - and for the Intel-based systems only. For the Bluegene system Juqueen, this limit is lower and lies around 40-60 owned cells per task, but the picture is less clear since for one or even two of the test cases the total number of cells is too small for a propoer application on this system. Also, tests on NCAR’s Yellowstone show that this number can be around 80 cells/task even on Intel architectures when switching off the I/O. Hence, we will weaken our conclusion here and state, according to the new definition of the transition zone, that for the three test cases considered here, 150 cells per task is a good estimate below which the absolute runtimes increase and thus a parallelisation on larger numbers of tasks becomes pointless on the Intel systems. This also fits better to Table G2, in which we compare "cheapest" and "fastest" parallelisations for the different meshes. We will also rephrase the parts concerning the good scaling (here defined as 70% parallel efficiency or better) in a way that it fits to the above statements. Please note also that the numbers mentioned here will become clearly visible in additional plots requested by Anonymous Referee #2, which display the scaling of the 120km and the 100-25km test case in the same way as it is already done for the 60-12km test case (Fig. 12).

We will also add the missing cases to the Tables D1-F1 and incorporate all the corrections mentioned as "more specific comments".

Regarding the question on pg 7007, lines 18-22: we agree that choosing two cases within the transition zone is not particularly useful. We repeated another profiling excercise for the 60-12km mesh with 4096 tasks on Juqueen (130 owned cells per task). This will be used together with the 2048-task profiling on Juqueen for the same mesh and consequently the 100-25km profiling run will be dropped. The differences are clearly visible between the two 60-12km test cases, since the two communication patterns (all-to-all and point-to-point) have significantly larger percentages for the 4096-task run than for the 2048-task run.

(3) Technical corrections

The fact that we are dealing with very different problem sizes in section 2 and 4 make it difficult to follow the referee's suggestion. Since performance metrics have been measured in "Realtime [s] per 24h model integration" for the three test cases in Sect. 2, we would prefer to keep this metric. In particular, we do not have split-up measurements (how much time for initialisation, how much for I/O, how much for time integration) for all test cases. This makes it difficult to scale the results to "Simulated years per day".

For the extreme scaling experiment in Sect. 4, simulated years per day is not a particularly useful measure either, even though we do have all the required information to calculate this number. For instance, the fastest execution on 24 racks (393216 tasks) runs at 6.3 x real time, which means the number of simulated years per day is 6.3/365.25 = 0.017. To avoid confusion, we suggest to drop the speedup column in Table G2 and instead add a column "CPUh fastest run" for a 24 h model integration with disk I/O enabled (similar to column 3, CPUh cheapest run).

Interactive comment on Geosci. Model Dev. Discuss., 8, 6987, 2015.