Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-119-AC3, 2016

Geoscientific Model Development Discussions

Interactive comment on "Towards European-Scale Convection-Resolving Climate Simulations" *by* David Leutwyler et al.

David Leutwyler et al.

david.leutwyler@env.ethz.ch

Received and published: 30 August 2016

Thank you for your review. Since the report mostly consists of a general critique, we chose to individually address the points raised, where appropriate.

The manuscript entitled "Towards European-Scale Convection-Resolving Climate Simulations" describes a new implementation of the COSMO code capable of using GPU cores. In addition, this new implementation is applied by performing two simulations at convection permitting scales over a large domain. These simulations are compared to 12-km ones. Finally, the computational advantages are discussed.

General comments: *The manuscript is easy to read and well written. Most figures are clear and the scientific content seems correct.*

Printer-friendly version

Discussion paper

We are glad that the scientific correctness of the presented results is acknowledged.

However, the goals of this study are very unclear to me. In the introduction, the authors wrote: "we assess the applicability of the convection-resolving COSMO model on continental scales". I do not often read Geoscientific Model Development but the publication of such an assessment in COSMO technical report seems more relevant to me. If every time somebody is increasing the domain size, he/she publishes a paper, then there would be a lot of useless literature out there.

In our view the paper fits very well into the scope of the journal. The website states that GMD is dedicated to the publication of "the description, development, and evaluation of numerical models of the Earth system and its components". Further description is available regarding the specific manuscript type applicable to our paper (model evaluation paper, see http://www.geoscientific-model-development.net/about/manuscript_types.html): "Model evaluation is an important component of most GMD papers. Model development papers in particular often include a large proportion of evaluation".

The description also clarifies the level of evaluation: "It is, however, common for pure evaluation papers to contain substantial conclusions about geoscience rather than about models, and such papers are not suitable for submission to GMD. [...]."

According to these descriptions, our manuscript fits the appropriate manuscript type very well.

The Reviewer also states: *"If every time somebody is increasing the domain size, he/she publishes a paper, then there would be a lot of useless literature out there".*

We consider this statement a bit unfriendly and would like to reply as follows: (1) The current paper is beyond increasing the domain size. Actually, it is the first paper to demonstrate the feasibility of a fully GPU-based model to conduct regional climate experiments. (2) The review makes it sound as if we would increase the domain merely by a little bit. However, we are actually expanding the previous domain of Ban et al. (2015) by about an order of magnitude (in area). The current version enables studying the climate of a continent, the previous that of an intermediate mountain range.

About half of this manuscript describes the methodology and the results of two experiments. The result of these simulations are well-established findings: CPSs can model finer structures and more realistic sub-daily statistics of convective precipitation. This was already found in many studies and is not worse being re-communicated, at least not with so many details. A small part of the paper is, in my opinion, relevant for publication, namely Section 5. [...]

It is correct that some of the results qualitatively agree with recent existing literature (Kendon et al. 2012, Ban et al. 2014, Prein et al. 2015). However, we are using an entirely different code version (with a completely rewritten dynamical core), and a computational domain that is at least 10 times as large as previous high-resolution studies over Europe. We think it is useful to corroborate the previous results and demonstrate the suitability of the GPU-based large-domain approach. Furthermore, our paper also demonstrates the ability of non-hydrostatic models to represent (1) wrap-up of small vorticies, (2) narrow convective cold-frontal rainbands (section 3) as well as (3) propagating cold-air pools and gust fronts associated with thunderstorm outflows (section 4). In fact, these differences not only represent an increase in model detail due to the increased resolution, but also changes in physical behavior. We are not aware of any detailed descriptions of such features in high-resolution simulations spanning the European continent.

That said, it is correct that Section 5 is particularly essential to the paper. Following suggestions of Referees 1 and 3, we have expanded section 5: (1) We clarified the " socket metric" and the implications for the presented experiment (including a new figure, Fig. 11 in the revised version). (2) We elaborate differences in weak-scaling experiments w. r. t. global simulations and (3) w. r. t. to the employed numerical schemes.

Furthermore we made some shortening in sections 3. We have also deleted a figure (former Fig. 4), and reduced the number of panels in another one (current Fig. 10). We think that with these shortenings the level of detail in the presentation is fine for a GMD article.

I am not sure what recommendation to give for this paper. I think it needs strong revision on the motivation. What do you want to communicate? Where should you publish this communication? I do not think that stating that the COSMO can be used on big domain is a communication relevant for publication in a peer review journal. [...]

Again, the review makes it sound as if we would merely increase the domain by just a small little bit. The details are as follows: The computational domain in our experiments (1536x1536x60 grid points) is an order of magnitude larger than what has been established and used in long-term RCM simulations (500x500x60 grid points by Ban et al., 2014). Increasing the problem size by an order of magnitude requires thorough and continued (re-)evaluation of the model capabilities. We think that using and reproducing established results in the process is a useful approach.

Besides confirming the applicability of the model on continental-scale domains, we present a number of rather novel results and ideas. For instance we are currently unaware of (peer-reviewed) publications of:

1. A weather and climate model which is able to execute the entire time stepping algorithm on GPU accelerators.
2. An assessment of the computational performance of such a model with full model physics.
3. A week-long convection-resolving real-case simulation of an entire extratropical synoptic systems (winter storm Kyrill).
4. The cloud visualization technique presented in section 2.4.1.

5. The demonstrationof our model's ability to scale weakly over very large domains.

I know that in the CLM community they also use quite large domains, also at CPS. For example, they are doing big brother experiments with domain of something like 1000 x 1000 grid points.

The Reviewer's comment is very unspecific. We are unaware of a publication about these efforts.

Because, I may not have understood the real motivation of this manuscript, I recommend a major revision. I ask the authors to express the motivation of the paper clearly and to make sure that the communication they want to publish is relevant. In addition, I ask the authors to restructure this paper according to this motivation. Stating that CPSs can model fine structure is most probably not necessary.

In revising the paper, we have partly followed these recommendations.

Line 436: the use of a large domain is motivated by the fact that "large domains provides a tool to study cold pools in heterogeneous ...". I am quite sure that the domain size of Ban et al. (2014) or Kendon et al. (2012) are large enough to reproduce these cold pools. In general, in the manuscript, there is no motivation for the use of large domains. Please motivate the need for such large domain.

We removed "large domains" from the particular sentence. It now reads: "The use of high-resolution models provides a tool to study cold pools in heterogeneous areas. Here we focus on the subdomain indicated . . ."

Furthermore added a concise motivation for using large domains in Section 2:

"The analysis domain excludes grid columns close to or within the relaxation zone (50 km distance to the CTRL2 boundary) and contains 1536x1536 grid points (2900x2900 km^2). It should therefore be large enough for small-scale processes to fully develop (Leduc and Laprise, 2009; Brisson et al., 2016)."

And we extended the paragraph in the end of Section 6 with the following sentences:

“Once established, such simulation capabilities will enable investigations of continental-scale climate feedbacks, sensitive to the treatment of deep convection, or assembling model-climatologies of interactions between convective meso/small-scale and synoptic-scale systems.”

Minor comments:

L320: You indicate “not shown”. Why not using the supplementary material to display this information?

We added the domain-mean of the precipitation rate at 18 UTC: CTRL12: 2.7 mm/day, CTRL2: 3.55 mm/day. This is the same snapshot as in Figure 6 (top row).

L415: Typo: logn should be long? L423: Typo: bahavior should be behavior?

We fixed these spelling mistakes.

L489: I agree with using socket for the comparison. Still, I think you should provide more information on what is on each sockets. Please describe the types of CPU/GPU that are used. You could also write the energy efficiency of these hardware. This would allow you to provide a rough estimation of the energy saved for a similar simulation in a latter part of the manuscript.

The paragraph has been criticized by the other reviewers as well. Therefore we added further clarifications, including a new figure (currently Figure 11). The sections now read:

"The full strong-scaling experiment corresponds to a 24 h simulation on a domain of 1536×1536×60 grid points. Input for this simulation consists of the lateral boundary conditions at hourly resolution, amounting to about 120 GB for the whole simulation. Additionally an output workload consisting of about 6 GB is written to the file system. All performance results have been obtained on a heterogeneous Cray XC30 system, located at the Swiss National Supercomputing Centre (CSCS) in Lugano, Switzerland (Piz Daint). The Piz Daint supercomputer consists of a heterogeneous node architecture with an eight-core Intel Xeon E5-2670 CPU and an NVIDIA Tesla K20X GPU per node (Figure 11), and Cray's Aries interconnect using a three-level dragonfly topology to connect the compute nodes. To normalize the performance metrics, they are defined as per socket. In the case of our configuration (Piz Daint, Figure 11), a socket corresponds to either an eight-core Xeon CPU or an NVIDIA K20x GPU.

A socket is the electrical component that provides the connection between the circuit board and the chip sitting on top of it. The advantage of the per-socket metric is its flexibility across architectures, which also allows comparing with individual sockets on a multi-GPU node (fat node). On a fat node, a socket still hosts only a single GPU chip, even if multiple GPU sockets are installed on a PCI express card or on a node. However, for the node configuration found in Piz Daint, this metric is a bit unfair towards the multi-core systems, since GPUs (today) still need an accompanying CPU hosting the operating system and instructing the GPU. With the socket-based metric, we do not account for that additional CPU. Another metric would be node-to-node comparison, assuming that a node can either consist of one CPU and a GPU, or two CPUs. For such a configuration, the second option would be fairer for the multi-core architecture. In general, node-to-node comparison is useful to compare the various possible node configurations one may find in a supercomputer. However, we believe that for the current study the per-socket performance metric is more useful than node-to-node comparisons, also because nowadays fat-nodes are commercially available."

Energy to solution experiments would require its own dedicated study. An experiment, using a similar version of COSMO, was performed by T. Schulthess et al. on the Piz Daint supercomputer. See here for an online presentation at SC13, including an analysis of energy to solution: https://www.youtube.com/watch?v=X5PqyfXc9pAt=21m20s

L509: "5 times more sockets". Why using 5 times in the text and 4.9 on the figures. Please be consistent.

Rounding the number 4.9 to 5 is indeed inconsistent with Figure 12. We changed that number in the text.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-119, 2016.