Review of “ROMSOC: A regional atmosphere-ocean coupled model for CPU-GPU hybrid system architectures” by Eirund, Leclair, Muennich, and Gruber.

General comments

This work presents the regional coupled atmosphere ocean land model ROMSOC, which is based on the existing components COSMO for the regional atmosphere and land, ROMS for the regional ocean, and OASIS3 for the coupling. While similar such models have been constructed earlier, this article points out a novel practical implementation, that is the hybrid computational setup with the ocean component running on CPUs and the atmosphere component running on GPUs of the same nodes on the computing system they have available. In their setup, where the ocean model is slower, than the atmosphere, and the billed computing costs depend only on the number of nodes and length of computation, this means that the computations for the atmosphere and land are essentially for free. Thus the article shows how the exploitation of the GPU computing power and the readiness of the COSMO code to run on GPUs can drastically reduce the costs of such simulations.

The manuscript is well structured. The introduction is followed by a first result section focused on the computational aspects, for which the hybrid CPU/GPU setup is essential. While the main results are obvious, the discussion of the strong scaling properties of the model system would be more interesting if additional details were provided. More detailed comments follow later.

Thereafter the model simulations of the coupled ROMSOC are analyzed and compared to observations and uncoupled simulations using ROMS only. This section covers a number of aspects, but is kept a bit superficial. Where differences to ROMS are discussed, the discussion is often too general, and difference figures are missing, which makes the assessment of the differences difficult for the reader. This section needs to be improved. More detailed comments follow later.

Overall this article is worth a publication, but needs some improvement.

Detailed comments and questions

Abstract

L8 … ROMSOC, a newly developed regional coupled atmosphere-ocean model. …

“newly developed” is misleading, as both model components, ROMS and COSMO, exist already, as well as the coupling software OASIS3. Further the authors cite earlier work by Byrne et al. (2016) that is also based on a coupled COSMO ROMS model. Please change the phrasing.

1 Introduction

L73 delete “… for the California Current System (CalCS)” and add “(CalCS)” earlier in the same sentence.

L130 … The shallow convection was switched off because this reduced our too-low cloud cover bias in the south of our model domain (section 4.5). …

As COSMO operates on a 7 km grid, one may expect that COSMO can simulate explicit deep convection to some degree realistically. But it is unlikely that shallow convection can be simulated faithfully. This makes it likely that better results – the reduced cloud cover bias in the south of the model domain – found without parameterized shallow convection are obtained for wrong reasons. Or one would have to conclude that the shallow convection parameterization of COSMO has no skill at 7km resolution, although this parameterization was operationally used at similar and higher resolution for NWP. Do you know why you get better results, although this is unexpected?

2 Model configuration

L153 … the surface net heat flux, …
= sensible and latent turbulent heat flux + net LW flux at the surface?

L157 … to COMSO … → … to COSMO …

L173 … ocean and atmosphere models use ERA5 … COSMO uses at its lateral boundaries ERA-derived … at 6-hour frequency. ROMS is forced daily at the surface outside the COSMO region …

What is the reason for using a 6-hour frequency for COSMO and a 24-hour frequency for ROMS? Wouldn't it be more consistent to use the ERA boundary conditions at the same frequency for both model components? Further, ERA5 data are available at hourly frequency. Would it have been advantageous to drive the simulation at this higher frequency, for atmosphere and ocean?

L174 … and surface temperature, … and the snow temperature and thickness …

Why does COSMO need surface related data from ERA? The surface of the COSMO domain is entirely over the ocean, where SST is coupled from ROMS, or land, for which COSMO can compute the surface temperature from its land model (TERRA) and the energy exchange between atmosphere and land?

Why is ERA5 snow temperature and thickness needed for COSMO?

L178 … In addition, ROMS SST and sea surface salinity (SSS) fields are being restored to monthly Reynolds SST fields (NOAA OI v2) and climatological SSS (ICOADS, Worley et al., 2005). …

Restoring SST seems necessary because the effect of turbulent heat fluxes at the surface on SST would otherwise be missing in the described setup of ROMS. For SSS it is less clear because surface freshwater fluxes are prescribed. Can you please provide brief justifications for the need to restore SST and SSS?

3 Compute setup and Performance

L185 … comprising 12 CPU (Intel Xeon E5-2690 v3 @ 2.60GHz) …

The XC50 nodes of Piz Daint have only 1 CPU per node, each having 12 cores. Please clarify this in the text.

L191 … Figure 2b shows the strong scaling results for ROMSOC on setups ranging from 4 to 84 nodes, with 4 nodes being the lowest node number required for ROMSOC to run. The coupled model scales very well up to 64 nodes, where it reaches a parallel efficiency of 80.8%. …

The speed-up and the implied strong scaling factor are hard to read or guess from Figure 2. For instance the "Factor of 12" indicated in panel (b) would suggest a strong scaling factor of 12/16 = 0.75 or 75%, which is less than the 80.8% stated in the text. It would be more appropriate to present the speed-up in a log-log plot, and the strong scaling factor in a separate plot with linear axis for the scaling (and still with log axis for nodes). This would be strongly appreciated.

It would also be of interest to know the scaling of the components separately. Based on the information that ROMS is dominating the integration time, the strong scaling up to 64 nodes is probably also dictated by ROMS. But this is only a guess. This could be shown in the same or in additional panels for speed-up and strong scaling.

The authors show in their table A1 that there exist a considerable unbalance between the compute times for COSMO and ROMS in their "ideal" 64 nodes setup. Lines [5] and [6] of Table A1 indicate about a factor 2. How is this ratio for the time integration part of both models alone, thus after the initialization has been completed?

This unbalance could be exploited, without further total costs by making the atmospheric integration more expensive, for instance by making the domain larger (or by increasing the horizontal grid resolution or similar actions, which could improve the expected quality of the simulation. Has this been considered?

Obviously such heterogeneous setups are very sensitive to the ratio of the computational power of the CPU and GPU units on the same node. CSCS has recently started its new Alps system, and Piz Daint will be decommissioned soon. Therefore I wonder if the authors can add information on the ideal setup and scaling for a ROMSOC setup on the Alps (on an architecture of their choice). Would this unbalance change, if yes in which direction?

Such unbalances could be avoided if both models ran on GPUs. Then the resources could be distributed as needed to get a balance. But ROMS has not been porte to GPUs. Do you plan to port ROMS to GPUs to overcome this limitation? If not, is this related to missing resources or is it based on the expected difficulties for the GPU port of some components (solver for the barotropic mode)?

L213 … This efficient scaling of energy usage is also shown in Figure 2c, where the increase in energy consumption for the 64 node setup increases by only a factor of 1.3 with respect to the simulation on 4 nodes (as a result of more nodes being used), while the speed-up in runtime increases by a factor of 12 (Figure 2b). …

The energy consumption is hard to understand. More explanations are needed. The total number of operations for the model integration is the same, whatsoever the parallelization. Only the communication related to the data exchange for the parallelization increases with the number of processes. But this is usually a minor cost. Still panel (c) suggest that doubling the number of nodes from 4 to 8 and to 16 adds most strongly to the energy consumption, although the strong scaling is best at this end, meaning that both components should scale very well, meaning that the time spent per operation is nearly stable. But when increasing from 16 to 32 and to 48 nodes the energy consumption flattens, though the communication should become increasingly more expensive. Which seems to happen when increasing the node number further from 48 to 64 and to 84(?). How can this be explained?

4 Model evaluation

Figure 3. Please add a panel showing the differences ROMSOC – Reynolds and ROMS – Reynolds, and ROMSOC – ROMS. This would be helpful to see more directly the effect of adding the atmosphere model and the coupling to ROMS.

L242 … abormously … → … anomalously … (?)

L246 … An important reason for this increasing bias is our turning off of the SST restoring in the coupled simulation. In the uncoupled simulation, this Newtonian restoring keeps the model’s SST closer to the observation, while in the coupled simulation, the model’s SST is permitted to evolve freely. …

Using a Newtonian relaxation of the SST in the coupled domain would be entirely against the spirit of coupling atmosphere and ocean, as the energetic consistency would be destroyed. Therefore this explanation is not helpful. (An important reason to use the SST restoring in ROMS is that the effect of the turbulent heat fluxes on the surface temperature needs to be accounted for, somehow.) A more interesting explanation would be that the energy fluxes determining the SST are biased. And this is what is addressed to some degree in the later parts of this article. Please rephrase your explanation for the SST bias in ROMSOC.

The SST field of ROMSOC shows a cold SST bias near the northern border, for which some explanations are given. But there seems also to exist an additional bias in a narrow strip along the north western edge (Figure 3a), which looks like an artifact from the assimilation of the model fields to the ERA5 boundary conditions. Such boundary artifacts are also evident in Figure 12. Does the framed area in Figure 3 (and in all other similar figures) include the lateral boundary regions of COSMO, where fields are modified by external boundary condition data? If so, it would be helpful to either exclude these areas (and explain this clearly) or to add a second frame that shows the internal edge of the boundary region. Then the discussion can be focused on the performance in the interior of COSMO. (And boundary problems can be discussed separately, if the authors wish to do so.)

L255 … The 10-m wind speed is also increased in the north of the domain, possibly favoring advection of cold water masses, which could lead to overall colder SST in the north of the coupling domain in ROMSOC (Figure 3). …

Following the description ERA5 wind is used in the lateral boundary conditions. Still the 10m wind maximum at the north western edge of the ROMSOC domain is substantially stronger than the ERA5 wind, and quite similar to the independent ccmp winds. How can this be understood? Are the ERA5 winds inefficiently used for the lateral boundary conditions of COSMO? Or is this a matter of the parameterization of the 10m wind based on model level winds and surface friction, so that the 10m wind can be substantially different despite of the assimilation of ERA5 winds in the boundary region of ROMSOC atmosphere?

It is also a kind of astonishing that the wind vectors in Fig.5b, for ERA5, show some strange small scale changes in directions along the north western edge, at ca. 47N, of the ROMSOC domain, although this panel shows a 12 year average.

Currently the discussion is limited to the strength of the 10m wind. The discussion should also cover directional differences. For example the ROMSOC winds close to the coast are mostly tangential from the North or slightly towards off-shore. In contrast, the ERA5 10m winds are mostly westerly, thus towards the shore, where the wind is strongest. And CCMP winds are again tangential to the coast. These directional differences should matter for the coastal upwelling in ROMSOC compared to a ERA5 forced ROMS simulation.

L258 Please include the wind directions in the discussion of Figure 6a+b.

L263 … Figure 6c shows the coastal wind drop-off, …

The data shown in this panel is unrelated to the strong-wind event on March 7th, 2011. It is a 12 year time average, as in Figure 5. Please present Figure 6c as a separate Figure, or include it in Figure 5. Then there would be less of a risk of confusion about a specific day versus a multi year average.

Figure6c and Figure5 show time averages for the same period, and therefore Figure 6c should be a projection onto the northerly direction of the vector fields in Figure 5a+b. Figure 5a suggests a maximum northerly wind component next to the coast and a drop-off towards offshore. Figure 6c however shows a zero northerly wind next to the coast and an increase towards offshore to about -6 m/s. Is the scale of the “Alongshore wind speed” in Figure 6c correct? If yes, why is this Figure inconsistent with the vector field of Figure5a?

A similar problem seems to exist for ERA5 vector field in Figure 5b and the alongshore wind speed in Figure 6c. Figure 5b shows essentially no northerly wind component near 40N.

L302 … The EKE averaged in a band off California (as outlined in Figure 9a) is lower in ROMSOC as compared to ROMS, although not statistically significant. …

Beside the general low bias of ROMSOC vs. ROMS, the other difference is the bump in EKE in ROMSOC at scales of about 500 to 700 km. Do you have any speculation on the nature of this bump? Why should a 12 year average of daily EKE develop such a bump at ~500 km scale? (ROMS does not.)

L323 … However, this bias in cloud cover does not lead to enhanced incoming shortwave (SW) radiation at the surface in that region as compared to ERA5 (Figure 12a,e) and reflects the very common too-few-too-bright cloud bias known from global models (e.g., Nam et al., 2012). …

This is a too simple description of the ROMSOC – ERA5 difference. Some areas in ROMSOC, near the Southern California coast, seem to receive a stronger downward SW flux than in ERA5, while others obviously receive less. Please add a difference plot, see also the following comment.

Figure 12: Please add “ROMSOC” and “ERA” to the panel titles. Concerning the organization, please put the panels for the same variable on the same row, and add a difference plot. This would make the differences more visible.

Figure 12: The ROMSOC panels for SW, LW and the latent turbulent heat flux show boundaries along the domain edges (see earlier comment on SST). What is the reason for theses boundaries?

L328 … Turbulent surface fluxes exhibit similar patterns in ROMSOC and ERA5, … and hence increase the latent heat flux (Stevens, 2007). …

This paragraph is a bit difficult to follow because difference plots are missing. Some features described in the current text are relevant for a relatively small fraction of the area (pos. sensible heat flux). Differences in patterns are not mentioned (strongest sea-to-air sensible heat flux near south west corner of the ROMSOC domain, absent in ERA5). Finally the LW flux downward is shown but not discussed at all.

Overall the description is not accurate. The “sensible flux downward” is positive, i.e. “from the atmosphere into the ocean”, in a minor area near the north western edge and at a few places on the coast only. In most of the areas the “sensible flux downward” is negative, out of the ocean to the atmosphere. Further the pattern is different. ROMSOC shows a maximum ocean-to-atmosphere sensible heat flux along in the south west along the edge. ERA5 has no such feature. Also here a difference plot would be very helpful.

5 Discussion

L341 … We related this cold bias to stronger wind forcing in the coupled model as compared to its uncoupled counterpart. At the same time though, this strong wind forcing leads to increased ocean mixing and a deeper and more realistically simulated MLD throughout the domain in ROMSOC as in ROMS compared to Argo observations. …

This reads as if the wind forcing allows a more realistic MLD simulation, and at the same time the SST gets to cold although the MLD is “correct”. Wouldn’t this simply suggest that the radiative forcing of the ocean surface temperature are a primary concern?

L362 … we want to highlight that both the spatial and temporal resolutions of COSMO are higher compared to ERA5 and that the coupling time step of ROMSOC is 144 times higher than the forcing time step applied to ROMS. Hence, we cannot identify with certainty whether the spatial or temporal resolution of the wind forcing is responsible for oceanic differences between ROMS and ROMSOC. …

This question could be addressed. For instance a sensitivity simulation where the COSMO resolution is degraded substantially could give insight into the role of horizontal resolution. Or ROMS could be forced by hourly ERA5 data instead of 24 hourly data, which would decrease the factor of the coupling interval from 144 to 6. Have such sensitivity experiments been considered?

L366 … However, we performed a sensitivity simulation for one year where we applied the coupling only daily in ROMSOC. The difference in MLD between our default ROMSOC simulation and the sensitivity run is mainly positive, indicating a 8-10 m deeper MLD in ROMSOC with a higher temporal coupling resolution …

What is the effect on SST? If the assumption is that the deeper mixing results from a higher coupling frequency, one would also expect an effect on the SST, that depends also on mixing.

L369 … Hence, we suggest that the temporal resolution of the coupling is mainly responsible for forcing stronger ocean mixing and coastal upwelling in the coupled model. …

Would you expect that a coarser resolution in COSMO, still using the same coupling frequency of the presented ROMSOC, would allow to simulate at the same quality level, despite of the increasing difficulty to represent orographic details of the coast and islands?

Further, how would a change in the ERA5 data frequency used for ROMS boundary conditions influence the quality of the ROMS simulation (MLD, upwelling)?