Geoscientific

Model Development

Discussions

[GMDD]

Interactive

comment

Interactive comment on "Towards in-situ visualization integrated earth system models: RegESM 1.1 regional modelling system" *by* Ufuk Utku Turuncoglu

C. Lemmen (Referee)

carsten.lemmen@hzg.de

General comments

This single-author manuscript by Turuncoglu describes the scientific motivation, technical extension, and the application of the Regional Earth System Model (RegESM 1.0), originally also by the same author, with online visualization techniques. RegESM is originally a four-component modularly coupled system of ocean, wave, atmosphere and river routing science models. The convincingly argues that such coupled systems's growing data storage demands do not align with increases in compute efficiency, such

that storage of data for later analysis will necessarily have to compromise in how much or how often data can be stored. If data was stored at less time intervals, information on processes that are scientifically interesting and occur on very short time scales, could not be analysed with conventional tools, for example, storm events. By adding a modularly coupled visualization pipeline as part of the model system, events on short time scales could be recorded and condensed efficiently without the need to store the full model outputs at such short time scales. Consequently, Turuncoglu adds a fifth component to RegESM (now in version 1.1) that uses the widely used and parallelized visualization toolkit ParaView for recording fast time-scale data.

The contribution is within the scope of the Special Issue "Model infrastructure integration and interoperability (MI3)" in Geoscientific Model Development. The manuscript contributes to "combining multiple models into a coupled system", relates to "parallel data transfer and grid interpolation", and provides "supporting capabilities", here in the form of an already established visualization tool. The innovativity of the manuscript is the combination of the visualization with a modularly coupled system; this work has the potential to be of use to many other geoscientific model applications.

The paper is overall well written, but suffers from severe language issues pertaining to the use and misuse of articles, in particular, which inhibits the flow of reading. Wording, grammar, style and other aspects of language are mostly good, but this manuscript needs to be fixed by a native English speaker. Figures are appropriate with some potential for improvement to be more intelligible. The supplementary material is well presented and useful. The paper is well organized and has a clear and logical structure

I recommend that this paper is published after some scientific and substantial language revisions. Below I comment on specific aspects of the manuscript.

Title, Abstract

title Prepositions/articles missing. Also, I don't agree with "integrated" but suggest
to use the word "modular". As the visualization has been proven, it is not a
perspective "towards", but already production ready, so don't use "towards". I
would rephrase the title (this is a mere suggestion) as "Co-processing online
visualization: the Regional Earth System RegESM version 1.1".

abstract Fix language issues, such as: articles, avoid "being" before participles, check
use of due to/because of, avoid overly use of "used", avoid "have to", do pro-
cesses literally "move"? Avoid use of "basically", and many more. It is a major
revision task to carefully check each sentence

abstract Avoid jargon especially in the abstract, it may not be clear to all interested
readers what "multi-scale", "processing pipeline", "run process". Find alterna-
tive/easier wording or explain (as these terms are necessary and useful later in
the manuscript). Define I/O (and consistently all abbreviations), define ESMF
correctly.

Introduction

p1 l23 Use consistent spelling for Earth System Model (capitalized), introduce abbre-
viation already here. Also use consistent spelling of Earth (name of our planet)
versus earth.

p2 l4 Unsubstantiated claims "more demanding compared to conventional standalone
models, and in many cases, to global climate models". Also standalone mod-
els can be computationally demanding and though less complex may employ

a higher resolution in uncoupled contexts and lower resolution in coupled context, arriving at the same computational demand. And if global ESM were less demanding, also their resolution would increase aligned to available compute resources and user requirements.

p2 l4 You have not yet defined "obvious needs". What are "truthful predictions" in a model context? Aren't they rather "accurate hindcasts" and "realistic projections".

p2 l10f avoid "in order", just use "to". Avoid "moreover" (throughout manuscript). Try to condense sentences and leave out all unnecessary words throughout the manuscript.

p2 l12 "are also simplified" is not necessarily true. This "can" be simplified or is sometimes even "enforced", such as the grid specification in ESMF, or the standard names by CF conventions.

p2 l16ff ESMF suggests to use Theurich 2016 et al. for referencing ESMF. "in addition" does not align with "can be given as examples". MCT only provides part of the functionality that is present in EMSF (also FMS), as does OASIS (in its different); OASIS may or may not contain MCT, depending on version. There are also more recent developments, such as Jcup3, OMUSE, MOSSCO, MESSy, FABM, CSDMS which you may mention. Please clean up the list, discriminate couplers from coupling frameworks, sort by relationship and highlight key differences or commonalities. A review of older technologies is provided by Bert Jagers 2010.

p2 l19 "The ESMs make use of state-of-art modeling frameworks, model components and libraries to have better representation of the earth system processes". This sentence is irrelevant. "The key component of coupled modeling systems is often the driver". Please refer to Alexander/Easterbrook (2015) for typical coupling constellations in ESMs. There, it is quite obvious that many models do not employ a driver. You could also use the term "hub-and-spoke" architecture (consistent

with your Figure 1) where the driver is the hub. The word "driver" itself is jargon and should be introduced beforehand.

p2 l19 "such as six hourly or daily averages". This is true only for coupled physical systems. For geology ESMs, typical timescales may be longer (years, centuries), see e.g., CSDMS.

p2 l23 "Coupling between individual earth system components adds extra overhead... network bandwidth usage, computation, disk I/O and data storage". Again, not necessarily. Please substantiate or give an example or reference where this applies. In my opinion the real overhead is in the loss of flexibility/modularity: it is preferable to have 1 hub speak a common language to 4 components than have each of the 4 components speak to each of the other three. Though the latter may be less complex, it hinders extensibility of a coupled system.

p3 l3f Fix articles. As everywhere.

p3 l7 Don't overly use "in short", also throughout manuscript.

p3 l7 Don't use judging language such as "brilliant". This entire sentence could be deleted (and the keyword "exascale" mentioned before.

p4 l4ff This paragraph is not needed.

Design

p4 l9 "couple four different model components". I don't understand. Version 1.0 coupled four components, version 1.1 adds a co-processing component, totalling five. This is also confusing in Figure 1. Modular frameworks such as

C5

NUOPC/ESMF do not discriminate scientific models, data components, output/visualization per se. All are just "components". I would suggest to use the term "component" for all five (four of them science models).

p4 l11 delete "state-of-art"

p4 l11 You have not yet introduced "the coupling interfaces" (this may be an article problem). So far, a driver and four models have been introduced. The necessity to add also coupling interfaces, and possibly NUOPC caps (CMI in the framework of Peckham 2013), has not yet been mentioned.

p4 l16ff delete this paragraph

p4 l section heads add ATM (later OCN, WAV, COP) to section heads

p4 l22 For in-text enumerations, please use (1) with double parentheses, and do not write the number in bold face.

p4 l26 "it is well tested with verified in". This does not make sense grammatically, please fix. Also, you have a verified model system for South and Eastern European/West Asian regions, but not for the Gulf of Mexico, which you provide as an example for the visualization.

p4 l27f Avoid "Additionally, ". What relevance has "does not support online nesting yet" for the application discussed here. Is online nesting needed for coupling to the co-processing model. Or is it just not "operational" in a sense that RegESM/WRF has not been scientifically proved?

p5 sections 2.1.1/2.1.2 Align in model descriptions of RegCM and WRF the content. Use the same structure: dynamical core, spatial representation, explicitly modeled states and subgrid-scale parameterizations, sub-modules. This parallel constructions helps to compare the ATM science models.

Interactive comment

p5 l9 "the driver and coupling interface are developed by ITU". The driver should be independent of RegCM, why is this mentioned? What is the coupling interface? Is it the CMI/NUOPC cap? You may choose to introduce those two terms to make clear what you mean by coupling interface. Please describe the cap (e.g. what state variables does it export/import, which transformations does it encompass). You may also elaborate on the changes internal to RegCM (and WRF/MITgcm, WAM/HD)

p5 l21 "model is modified to exchange data". Discriminate work done on cap (CMI) and within model (BMI, Peckham 2013).

p5 l23ff All that was said in the ATM section also applies here (parallel structure, elaboration ...)

p6 l2ff Avoid judging language like "fortunately". Avoid "in general" if possible. "In some studies", please reference those studies or give a motivation for both cases.

p6 l7f "In the current design of the coupled modeling system, the driver allows to select the desired exchange fields from the predefined list". Please explain how this works. Is the list like a database with all known fields from all possible coupled models and there is an automated selection of suitable fields?

p6 l15 Leave out example technical detail "REGCM_COUPLING". I would appreciate a list of changes in the supplementary material, if that is practical to do.

p6 l23 define PROTEUS abbreviation

p6 l26 PBL has already been defined

p6 l26f third-generation *wave* model. What is "pure" physics? Use an en-dash to indicate a relationship in wave-wave / wave–wave (the second one is correct typography)

p6/7 l32 WAM is not by ENEA but by GKSS (now HZG) in Germany, I believe. Please attribute correctly. Describe changes to original model (here in text or as table in SOM).

p7 l 5 Unit of z0 (m).

p7 l20f Explain your motivation for the two different ways. Do they pertain to ROMS and MITgcm coupling, respectively? How are they chosen?

p7 l22f The change to output format is irrelevant in this context.

p7 l30f "The ESMF is chosen because ... of the NUOPC". Please explain the relationship between ESMF and NUOPC. Why do you chose NUOPC in addition to ESMF, this needs elaboration for readers who are not familiar with NUOPC.

p8 l2f "It also provides the capability of transferring computational grids"; here, "it" should refer to ESMF, not NUOPC.

p8 l2f "NUOPC layer to support various configuration of component interactions such as defining multiple coupling time steps (fast and slow time steps; Fig. 2)". There are other frameworks that provide add this capability to ESMF, namely MAPL and MOSSCO, please reference and contrast.

p8 l9f "connector components provided by NUOPC layer. Connector components are mainly used to create link between individual model components and driver". As far as I know, RegESM includes its own instance of a NUOPC connector, which is a single one for all connections between the components. Please elaborate on the design and functionality of the connector.

p8 l18f You name stability as an advantage of semi-implicit coupling. Also add drawbacks of semi-implicit and advantages of implicit time stepping.

p8 l20f delete "mainly" and "In this case"

p8 l28 You may want to add the term "load balancing" for describing the challenges in concurrent coupling. There should be plenty of literature that describes load-balancing and possible ways to minimise this. In the geosciences, usually all OASIS-coupled models face this challenge.

p8 l31f The description of regridding could be a subsection, section 2.6 "Connector". How do other coupled model systems handle this (any references available?). Recently, I implemented into MOSSCO a very similar scheme using bilinear for interpolation and nearest neighbour for extrapolation, so I believe this is a very good way to handle this (but it is not published yet). Possibly, describe how ESMF 8 handles this rather than pointing vaguely to ongoing development.

p9 l14 Align your description with the terms used in Figure 4 (see comments below on this figure).

Integration of co-processing

p9 l25 Use an em-dash "—" instead of a hyphen. One sentence is not a paragraph, so join with next paragraph.

p9 l28 Define "conventional" or give a reference to such approaches. Define pipeline in this context.

p9 l30 How is "wrapper layer" different from Component Model Interface (CMI) or NUOPC cap or ESMF component interface?

p10 l3 On what "other hand"? Describe the relationship between the adaptor and its NUOPC cap

Benchmark

p14 l26 The "driver component introduces additional 5-10%". In ESMF documentations, usually below 5% overhead is assumed. this makes sense as you to the SMM twice for each interpolation of mask extrapolation. It would be interesting to see whether the generic implementation of extrapolation is more efficient in ESMF 8 (out of scope of this MS).

p14 l35 " includes vertical interpolation to map data from sigma coordinates to height coordinates". Why is this interpolation step necessary? In principle, the visualization should equally work on sigma layers. And is the interpolation performed within the ESMF connector, or within a Paraview pipeline?

p15 l2 the computational demand for visualization "require 10-40% extra" is rather high and its evaluation would probably change depending on whether it is more 10% (acceptable) or 40% (not acceptable). The benchmark using software rendering with Mesa is not ideal to demonstrate this, especially as many new HPC systems come with dedicated GPU nodes that could much reduce the visualization overhead.

Results and Conclusions

p15 l31 This section is not well named. I suggest "Demonstration application" but make sure to choose something more elaborate than "results"

p15 l35 oblivious => obvious

p16 l3 Explain how you choose between live mode and co-processing modes. Are these two different models? Or is this done by way of a configuration file. Or,

is it possible to attach live viewing to any running system that has co-processing enabled?

Figures

Figure 1 Very clear and readable, consistent color and design with NUOPC web page. Please see my suggestions to treat everything as component, regardless of science "MODEL" or visualization "COMP". It is not clear from the text what "redist" is; also, this feature is not used when running sequentially. I also do not understand the many "regrid" arrows, as regridding occurs when the driver transports a field from one to another component, not from the component to the driver. This is confusing.

Figure 4 Rather unreadable. Please redesign. (1) Avoid bridges (this is possible by routing the arrow "Create ESMF FIELD_B" north to "Create ESMF FIELD_C". (2) don't use same line thickness/color for outlining the boxes around src/dst grid, rather make them visually less intrusive to highlight the flowchart. (3) Abstract away 1.0, LARGEST, "over sea" or other value-based masking, if possible.

Figure 7 It is confusing that the visualization 2x2 tiles don't geographically overlap for data from OCN and ATM in CPL. This is not needed, but confusing at first. Maybe give a hint to the reader that this is not a problem.

Figures 9/10 It is very difficult to compare ATM versus OCN performance using different y scales. I recommend to use the same ylog-scale for all panels in figs 9 and 10. Also try loglog. Tile sizes are hard to read, please move text from lines. Calculate all speedup relative to 140 cores, so that you don't have different reference points. Explain model versions like RegCM_r6274, delete UHeM. Avoid overlapping graphics and text.

Figure 10 8 lines in a graph is too messy. Even if on ylog, this may be difficult. Possibly choose different or more subpanels (e.g. CPL_LR + CPL_HR in one panel, then P1, P2, P3 each in one panel). Or use more colors for lines or different line styles ...

Figure 11 Zoom in to Paraview Window only (delete lower and upper OS information bands)

Figure 13 Use kg kg^{-1} and the like, not the slash for division in units.

1 Code

I have been able to download and compile all components, except for getting the catalyst to run. It would be helpful to provide a fully functional reference system as a docker container, for example.

2 Supplementary videos

Useful and nice! Please make sure the video dois are referenced in the text, and add to the videos a link to your publication.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-179, 2018.