Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-311-RC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-311

Andrew Porter (Referee)

Referee comment on "CP-DSL: Supporting Configuration and Parameter Selection of Ocean Models with UVic (2.9) and MITgcm (67w)" by Reiner Jung et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-311-RC1, 2021

General points

The paper describes the development and evaluation of a Domain-Specific Language (DSL) for use in the configuration and parameterisation of ocean models. The new language ('CP-DSL') is evaluated in the context of the UVic and MITgcm ocean models but is intended to be generally applicable to others. Configuration and parameterisation of ocean models can be complex due to the number of scientific components involved and the historical evolution of the code base. This complexity presents a considerable barrier to adoption of a model and can also be the source of errors meaning that a scientist may not be running a simulation with precisely the options that they intend.

CP-DSL abstracts out the various mechanisms that may be used to configure a model (e.g. CPP directives, include files, Fortran namelists) and simply presents a user with named options and settings which may be grouped appropriately. As such, I think it is an approach that has value although I feel that the paper itself could do with making this case more strongly. Related to this, although there is a discussion of the evaluation of the features and use of CP-DSL, there is no mention of what the model developers themselves think - are they keen to adopt CP-DSL or do they have reservations?

It seems that configuration/control of diagnostic outputs is at an early stage in CP-DSL but this is a critical and complex part of production jobs. I can see value in a common way of specifying diagnostics as this extends beyond ocean modelling: there is a relatively small number of IO systems and these tend to be common between e.g. atmosphere and ocean models.

Specific points

Although the discussion of the various roles played in the development of ocean models in Section 3.1 is interesting, I don't think it adds any value to the paper (which is primarily about the new DSL) and could be removed.

Figure 2 and its accompanying text mention that there are options that are common between ocean models. Given the subtleties that can occur when different scientists implement the same numerical scheme I think that determining such options could be problematic. I think some examples of such options would be helpful here. Is there a need for an agreed set of named quantities?

The UK Met Office uses 'Rose' for job configuration (see https://metomi.github.io/rose/doc/html/tutorial/rose/index.html#rose-tutorial) which has some similarities with the approach described in this paper. It is likely that other meteorological centres have similar configuration systems.

Technical corrections

(References are to page_number:line_number)

2:38 No comma after "as well as" and throughout the text.

3:87 PSyclone is developed by the UK Science and Technology Facilities Council's Hartree Centre in collaboration with the UK Met Office and the Australian Bureau of Meteorology. PSyclone has two 'modes' of operation: as an internal DSL (as used for the UK Met Office's LFRic atmosphere model) and as a code transformation tool (as used with the full NEMO ocean model).

4:89 "model developers what," should be "model developers which,"

4:105 "diagnoses" should be "diagnostics"

5:143 "responsible to control" should be "responsible for controlling"

7:186-189 "densities" should be "resolutions", "wide mesh" => "coarse mesh", "denser mesh" => "finer mesh",

"denser and wider meshes" => "fine and coarse meshes", "as it is the case" => "as is the case"

8:205 Not all models set grid sizes at compile time. This is a run-time option in NEMO for example.

8:214 Deployment of models is, I believe, something that the CYLC workflow engine (https://cylc.github.io/) does. Please compare.

9:231 No comma after e.g.

9:235 "built tools" => "build tools"

9:236 "whereby some are not available" => "which are not always available"

9:239 SVN is used as well as git (e.g. by NEMO)

9:252 "allows to draw" => "allows users to draw"

10:258 "derived form" => "derived from"

11:281 "look like to" => "look to"

11:294 "allows to have" => "allows"

14:325 "for a ocean" => "for an ocean"

15:330 "allows to import other" => "allows the import of other"

18:422 "Xtend" => "XTend"

20:473 "utilized in a joint session" - a joint session of what or whom?

20:474 What does "UVic is the reference simulation by GEOMAR of the University of Victoria model." mean?

A better reference for GOcean may be found at http://nora.nerc.ac.uk/id/eprint/521162/