Comment on gmd-2021-311
Stephan Kramer (Referee)

The manuscript "CP-DSL: Supporting Configuration and Parametrization of Ocean Models with UVic (2.9) and MITgcm (67w)" describes the implementation of a Domain Specific Language approach for the configuration of ocean models in the form of a new language CP-DSL. The subject of this paper addresses an important topic regarding the often complex configuration process of ocean models with a number of strongly conflicting issues like user friendliness for users with a wide range of expertises, the wide variety of competences involved in the construction of accurate and efficient ocean models which affect both the way the software is developed but also its optimal configuration, and the fact that different models have quite different configuration processes, where the expertise of an advanced user of a specific model does not easily translate to that of another model. Although I believe a number of interesting ideas and techniques to address these challenges are presented in this manuscript, a number of issues in the presentation make it quite hard to evaluate what has actually been achieved.

The paper spends a good amount of time discussing the requirements of an ocean model configuration system (important), but there is considerable less attention to explaining what the actual objectives of this project are and how the chosen approach helps to achieve these. As the authors quite rightly state the added complexity and risks (in terms of new dependencies) need to provide benefits, but there is only a limited discussion of what these are, and in particular how the specific choices in this project deliver these. This, and the fact that the actual implementation is only described in a rather abstract way, with only a few restricted excerpts and no concrete examples, make it hard to judge to what extent such benefits are delivered by this project.

Specific issues throughout the manuscript:

- the most problematic section in my view is section 5 which provides a very abstract
overview of the syntax of the proposed language but only through a few excerpts that do not give a very clear picture of the language as a whole. I also do not find the UML diagrams to be particularly enlightening (figures 3-6). I would really like to see a more complete overview of the features that have been implemented, and a lot more concrete examples of actually what goes into the "configuration model" vs. the "declaration model" so we can get a better idea how universal the language is and able to specify things in a model independent way.

- it's only in section 6 that we finally get told what CP-DSL is actually made of, but the description of the key components in section 6.1 is very terse. Please explain for instance what EMF is. As a side note, it also seems that some of these components bring in a dependency on a specific version of Java which seems to be in contrast with one the requirements (line 251-253) and isn't particularly well supported on some of the HPC systems that ocean models run on.

- as mentioned before the authors do not really evaluate or discuss what has actually been achieved in this work. The evaluation by other users in section 7 is described in a rather superficial way. The first evaluation in section 7.1 seems to be about a quite different version of the language, so the only information we get, about a second evaluation is "The division in general parameters and modules was considered useful. Also the reworked YAML syntax was rated easy to understand." This is actually the very first place (right at the end of the manuscript) where it is actually mentioned that the CP-DL syntax is closely (?) related to YAML - the other places YAML is only mentioned to contrast with XML and JSON. From the conclusions:

"As this is an ongoing research project, we aim to further extend and improve CP-DSL in close contact with users and active scientists from the domain. We initially developed the DSL for a representative subset of MITgcm ocean modeling scenarios and are currently evaluating it to be able to support all modules of MITgcm. However, the current syntax for diagnostics caused a larger comment by our domain experts, as diagnostics are an important topic in climate modeling for various purposes."

Having read the paper I still have little idea what representative subset has actually been implemented.

- I tried to evaluate the software, also to get a better idea about the structure and functionality of the language, but I'm afraid I didn't get very far. There is no documentation at all that is directly accessible, just some installation instructions that were unfortunately insufficient to guide a user, like me, with no experience of using Java, maven, etc. For me it is very unclear what the different parts (cp-dsl, cp-dsl-replication, cp-dsl-jupyter-kernel) consists of and how they are supposed to work together, and how they should be installed such that the different components can access each other.

- As discussed ocean modelling already brings together a variety of expertises (e.g. oceanography, numerical analysis and HPC) and the CS flavoured approach followed in this paper with DSLs, context-free grammars and metamodels adds a whole layer on top of that. This makes it important to have a clear view of the intended audience and adjust the language to it, briefly explaining key concepts. In particular if the intended audience is ocean model developers, who may be familiar with many advanced
computational and numerical techniques, but not with tools and terminology common in DSL approaches, some more guidance would be helpful. Here it is also important to be aware of how terminology various between different communities and be specific about what definition is being used. As an example, the authors already point out that the concept of language models adds a new meaning to word model in the ocean modelling context, but within the ocean modelling community the word already has a range of meanings: a conceptual model of the global ocean, a description of its physics, a translation of that into a mathematical model, which in turn are translated into numerical equations whose specific software implementation is also referred to as an ocean model, and finally a specific configuration of such a model for a specific scenario is again referred to as an ocean model. In this paper the authors choose the (appropriate) definition of a specific software implementation (line 16), but then in line 140 we have "The Model Developer is a software developer and responsible for transferring the ocean models into code" which contradicts that definition and makes it hard to understand what the difference between a "Scientific Modeler" and a "Model Developer" is. As another example, the authors make a distinction between configuration and parameterisation. Here it is important to note that "parameterisation" already has a very specific meaning in the ocean (and atmosphere) modelling community, it refers to processes that are not modelled through PDEs on a numerical grid, but rather through empirical parameterisation of these processes (typically on the sub-grid scale). "parameter selection" might be a better description of what is meant in the paper.

Configuration is defined in line 37 as "the selection of features and code to be used for the model, as well as, the build configuration." but then on line 268:

"For each simulation experiment, we need to define a configuration and a parametrization. Independent of a specific experiment, we declare settings that are specific to an ocean model. Thus, the parameters and configurable features are declared with CP-DSL in a Declaration Model specific to each supported ocean model, as depicted in Figure 2. The Configuration Model is independent of a specific an ocean model, it defines the settings of a concrete experiment, whereby the declarations in the Declaration Model for the specific ocean model are referenced. This way, we separate the ocean-model-independent and the ocean-model-dependent settings"

which seems to bring an entire new definition of configuration which is used along with the old in the same paragraph.

As a final example, the word "deploy" and "deployment" is used in a number of places: "Model developers also deploy the software" (line 141), "the deployment is merely configuration and parametrization" (line 163) - and I don't really understand what is meant there - I'm more familiar with its usage as in line 207-214.

Some suggestions for further references
Note: these are suggestions from personal experience only, I don't think the current
Section 2 provides a fair overview of previous DSL approaches in the ocean modelling context. Although section 4 gives a good overview of the specific requirements for a configuration system for an ocean model, there isn't a clear separation between those requirements that are specific to ocean models, and those which are in common with the configuration of other types of scientific modelling software, e.g. atmosphere models. As one of the key decisions in the design of DSLs is based around finding the right level of abstraction, it would be worth to extend this discussion to scientific models in general and explain why an ocean-modelling specific language is required, or whether it could be built on top of a more generic approach for the configuration of scientific models in general. As an example of the latter, I have personally been involved in SPUD [5] an XML + RELAX NG based configuration system for scientific computer models (I do share the author's preference for more plain text formats btw).

Already mentioned are Psyclone, Dusk/Dawn and Sprat which target other layers of the software stack, in particular PDE discretisation in combination with automated code generation. In this context it might be worth mentioning the popular FEnics [1] and Firedrake [2], and DUNE [3] projects which make extensive use of such approaches (for context I'm one of the authors of Thetis [4] a coastal ocean model based on Firedrake). You do mention ICON in the context of diagnostic configuration, but I believe there is more DSL-based ICON development described in [6]. Finally the Atmospheric Modelling Language (ATMOL) [7, 8] developed with the Royal Netherlands Meteorological Institute may be worth a mention.

[1] M. S. Alnaes, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes and G. N. Wells. The FEniCS Project Version 1.5, Archive of Numerical Software 3 (2015). [doi.org/10.11588/ans.2015.100.20553]
[2] Florian Rathgeber, David A. Ham, Lawrence Mitchell, Michael Lange, Fabio Luporini, Andrew T. T. Mcrae, Gheorghe-Teodor Bercea, Graham R. Markall, and Paul H. J. Kelly. Firedrake: automating the finite element method by composing abstractions. ACM Trans. Math. Softw., 43(3):24:1–24:27, 2016. doi:10.1145/2998441
[3] Peter Bastian et al, "The Dune framework: Basic concepts and recent developments", https://doi.org/10.1016/j.camwa.2020.06.007
[4] Tuomas Kärnä et al, "Thetis coastal ocean model: discontinuous Galerkin discretization for the three-dimensional hydrostatic equations", GMD, 11, 4359-4382, 2018, https://doi.org/10.5194/gmd-11-4359-2018
[5] Ham, D. A., et al. "Spud 1.0: generalising and automating the user interfaces of scientific computer models." Geoscientific Model Development 2.1 (2009): 33-42
[6] R. Torres et al. "ICON DSL: A Domain-Specific Language for climate modeling" http://sc13.supercomputing.org/sites/default/files/WorkshopsArchive/pdfs/wp127s1.pdf
[7] Robert A. van Engelen, "ATMOL: A Domain-Specific Language for Atmospheric Modelling", Journal of computing and information technology 9.4 (2001): 289-303.
[8] Paul van der Mark, "A Case Study for Automatic Code Generation on a Coupled Ocean—Atmosphere Model." International Conference on Computational Science. Springer, Berlin, Heidelberg, 2002.

Some smaller specific comments:
line 89: "by model developers what, in principle, allows" - I think you mean "by model developers which, in principle, allows"
lines 186-190: I would suggest replacing grid density, dense and wide with grid resolution, fine (grid resolution), and coarse which are more common
section 3.5: I don't really understand the difference being made between versions and variants, and how it is relevant to the rest of the paper
line 203: why mention openmp but not MPI (the Message Passing Interface not Max Planck!)

line 239-240: "Code management and sharing via tar-balls, ssh and email is mostly deprecated and replaced by Git, often with the Git Large File Storage (Git LFS) extension." I think that's a little too strong; the use of ssh and shared file systems is still very common practice.

line 242: what is meant by two layers?