Geoscientific
Model Development
Discussions

Interactive comment on "The Marine Virtual Laboratory: enabling efficient ocean model configuration" *by* P. R. Oke et al.

Anonymous Referee #2

Received and published: 18 April 2016

The functionality of the WebMARVL system described in this article is a truly valuable step toward accelerating basic model configuration for novice users. I applaud the MARVL developers for creating such an interface. If it were for only one of the models it would be valuable, but that the same interface works for multiple ocean models is exceptional. The ability to create forcing and boundary conditions inputs from multiple external models will facilitate studies of how downscaled models are sensitive to external model conditions, which is a valuable higher level metric of external model skill.

But I have some reservations about the way that the authors have chosen to present WebMARVL to the readers of GMDD. I have two chief concerns. These are:

(i) The results section focuses on descriptions of test cases that don't rigorously con-

trast differences in model skill with respect to verification data (if there is any) that result from using WebMARVL generated configurations versus configurations customized by expert users, and (ii) the underlying technologies WebMARVL uses to access its data bases, and the algorithms it uses for interpolation and remapping of fields, are not described.

To expand on these issues:

(i) Results section:

Section 3 presents several examples of WebMARVL use. These show that many users have successfully used the system to set up models, which is good, but do not rigorously demonstrate whether the automatically configured files are as good as, or perhaps better than, customized efforts.

Example 3.1 goes some way toward addressing this issue by comparing two configurations of SHOC, and admittedly qualifies the comparison with the remark "Recall that the goal of these comparisons is to establish the credibility of WebMARVL."

In Example 3.2 it is said the model agrees "reasonably well" with observations "with differences of 0.5-0.8C." Viewing Fig. 6a I see large areas in the core of southward flow (roughly 153.4E) in the observations where temperatures are ∼25.5 C but in the model they are 23.5 C. These are errors of 2 C (not 0.5-0.8). This claim of agreeing "reasonably well" is not substantiated by Figs 6b,c.

Example 3.3 makes a useful comparison to ADCP observations, but this is really an evaluation of the ROMS model. It is not an evaluation of the WebMARVL configuration of ROMS compared to a configuration made by other means.

Example 3.4 adds little.

Example 3.5 goes some way toward addressing this issue.

Everything that is said in section 3 could be said with examples 3.1 and 3.5.

But much more directed and illustrative information could have given. For example:

(a) Show a difference plots for the two SHOC bathymetries in Example 3.1 to illustrate what the automatic system produced compared to an expert user, and comment on the potential ramifications of this.

(b) Show some details of an automatically generated coastline land/sea mask and comment on the potential limitations, or not, to alert users of issues they might need to confront before simply accepting WebMARVL output.

(ii) Underlying technologies:

WebMARVL is offered as a convenient user interface for configuring inputs common to most hydrodynamic modeling codes. This begs the question of what might be lost when this all-purpose code is used in place of codes created for specific models by expert users.

Fundamental differences between ocean models are the placement of variables on staggered grids (e.g. Arakawa-B versus Arakawa-C grids) and vertical coordinate (terrain-following versus level coordinates). These differences exist across the target models (ROMS, MOM, SHOC) and external source models (Mercator, BRAN, FOAM). To interpolate from one to another necessitates a choice of interpolation scheme (linear, spline, kriging, smoothing, etc.), choices on respecting conservation issues (e.g. depth integrated transport conservation where source model and target model bathymetries do not exactly match), adjusting for mismatch in land/sea masks (conservation/extrapolation issues), smoothing of bathymetry for stability and accuracy in terrain-following coordinate models, and so on.

What map projection is used for the grid generation?

Weather models use graduated land/sea masks that give the fraction of a near coastal cell that is land or water, whereas ocean models do not. How does WebMARVL handle this aspect of the re-gridding of meteorological model data? Can extremely cold (or, in

Australia, extremely hot) land temperatures spill over into the interpolated ocean fields? These details are important to the readers of this article. More so, I would say, than a long list of example applications.

While many of these issues can be addressed after creating a starter configuration, others (especially horizontal and vertical interpolation between source and target) are fundamental. Expert users are not given any details on what algorithms WebMARVL uses, and therefore cannot evaluate if these might be limiting or a source of "error" that propagates into a downscaled model.

WebMARVL includes a modest database of meteorological forcing, open boundary conditions, and regional in situ data. Are satellite data sets included, and what is their provenance?

The available options are suited to local Australian applications, but GMDD is an international journal. There will be interest from readers in plugging in alternative data sources. If WebMARVL requires a specialized format for those data (e.g. in terms of coordinate or variable names, supporting metadata, file structure or file type, etc.), or requires local storage proximate to the WebMARVL server to achieve acceptable performance, then these limitations could render WebMARVL of limited interest to new users regardless of its features.

Presumably WebMARVL reads data meeting widely used metadata and file format conventions for geophysical data (CF-conventions, the UNIDATA Common Data Model, WMO Grib format, NetCDF/HDF, etc.). The reader deserves a section describing database format assumptions, and possible limitations that stem from them, that are built in to WebMARVL. Can WebMARVL access a web service (e.g. THREDDS, OPeN-DAP) in lieu of local files? Would this significantly impact performance? These are issues that are important to the readership of GMDD and the intended wider user community for WebMARVL. Section 4 claims "we describe the technical steps that are automated by WebMARVL." But there are no technical details.

The Australian emphasis of the WebMARVL portal is understandable. Other groups and communities might similarly wish to implement regional portals based on the Web-MARVL technology. Is the system portable, and documented for others to implement? What software license terms are imposed on usage? Is it open-source? Does it require any specialized hardware or supporting software environment? These are further issues that are important to the readership of GMDD and the intended wider user community for WebMARVL.

Section 4 states "WebMARVL is freely available to researchers associated with Australian research organisations, and there are plans to make it available globally." By "available" does this mean the software that runs WebMARVL, or only global access to an Australian hosted portal? I hope it is the former, just as so many services like ERD-DAP, Live Access Server (LAS), THREDDS, etc. are open source for the geophysical community, and benefit from a distributed developer community culture.

Details:

Figure 1: This diagram doesn't really depict workflow, or algorithmic elements. The arrows do not trace the trajectory of anything meaningful, like data or code objects. The blue balloons at right should lie between the left blue balloons and red balloons because the user choice sets the database, and then it passes to the extractor/re-gridder. The green balloon is not "generic" because it relates directly to the choices of parent model boundary conditions and meteorology forcing (in the blue balloons), and the variables that must be extracted for each specific model. Why is the green balloon linked with a two-way arrow to the red boxes? The red boxes and black boxes should be at the end of the processing flow, so decide: is the flow left to right, or top to bottom, and put the boxes in this order.

Page 9749: "Neither configuration includes freshwater fluxes." Does this refer only to E minus P, or also include river inflows? There is no mention of river flows in the article.

Page 9751 (line 19): Please clarify, does WebMARVL not produce tide boundary con-

ditions, or just not for this example?

Somewhere in the paper it would be nice to see what is in the take away bundle for each target model so an expert user could see clearly what WebMARVAL does and does not deliver. So perhaps create a table that lists the set of files created for each model, and files for that model that are NOT created.

- a grid file (YES)

- a boundary conditions file derived from an external model (YES)

- a harmonic boundary tide constituents file (NO) (??)

- terrestrial river inflow file (NO)

- air-sea heat flux data: - as air-sea fluxes (YES), - as marine boundary conditions for bulk formulae (NO) (??)

- air-sea freshwater flux data (??): - precipitation (NO), - evaporation (NO)

- ice (NO) (??)

─────────────────────────

Interactive comment on Geosci. Model Dev. Discuss., 8, 9741, 2015.