Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-158-RC1, 2018

Geoscientific
Model Development
Discussions

Interactive comment on "ScientificWorkflows Applied to the Coupling of a Continuum (Elmer v8.3) and a Discrete Element (HiDEM v1.0) Ice Dynamic Model" *by* Shahbaz Memon et al.

Anonymous Referee #1

Received and published: 16 August 2018

In their manuscript, Memon et al. describe the use of the workflow management system (WMS) UNICORE for modularizing the coupling between the ice flow model Elmer and the Calving model HiDEM. According to the manuscript, this coupling has previously been done with one 400-line long shell script calling various processing scripts and the models. The paper now develops requirements for a rewrite of this coupling that are met by the WMS UNICORE, that Memon propagates in several of his previous publications. Some of these requirements seem constructed to match the specifications of the WMS, and not arising from the reality of modeling (see below). I am not convinced that the 27-page manuscript about the conversion of a 400-line ($\sim$10 pages in this format) shell-script into a high-level system, is advance in Geoscientific Model

Development, and suggest rejection of the manuscript.

Having coupled models - even across different supercomputers - before, I can definitely see a great value in easing this process. UNICORE may be a contribution to solving difficulties arising here, and with the portability of solutions, although, telling from the manuscript, the installation of UNICORE on a supercomputer seems to be far from trivial ("Maintaining a server-side deployment is not trivial because it needs a dedicated server that manages workflows and atomic jobs" (p. 23 ll 22), "The authors are grateful to [. . .] for his support and patience in making the UNICORE services available of CSC's computing resources"). Furthermore, my personal experience tells that getting the models themselves to run on different supercomputers tends to be the agonizing part, with scripts being no match, however poorly they may be written. In the end, the authors have managed to replace 400 lines of shell script (including job-headers) with a dependency on a high-level workflow management system with some services hosted at yet another supercomputing center. I am not convinced that a thorough clean-up/rewrite of the shell script would not have solved the main problems of the old script with way less effort and overhead.

The manuscript itself is somehow entangled between the scientific aspects of the ice dynamical problem, and the technical aspects of the software solution for job control on a set of supercomputers. It lacks a clear focus on either side, as can easily be seen from the mixture of plots that describe the glaciological problem (Figs. 1, 2, 3, 9) and the software solution for abstracting the scripting (Figs. 4-8, with Fig. 7 duplicating the right half of Fig 6). While we are introduced to the basic reasons why one would run the different glaciological models for the different sub-problems, and get some insights on the model grids, and which piece of code solves which sub-problem, there is no interpretation of the results shown in fig 8, or direct glaciological relevance of re-writing the coupling. On the other hand, for the description of how the authors took a shell script and turned it into something more high-level, the details of the glaciological problem are largely irrelevant.

Some of the complaints against the shell script solution are rather surprising to read. On page 10 lines 11f the authors complain "Thirdly, the simu_coupling.sh script was implemented using the Bash shell script language, whose knowledge was necessary to understand and enhance the application." Similarly, on page 23 lines 25-27 "On the other hand, a shell-script-based implementation would require SSH-based remote access, which may cause inconvenience for some users who are not used to interact with systems through a command-line interface." I find it hard to image that somebody would successfully use ELMER on a supercomputer without being able to read and manipulate a shell script. The fact that all sub-steps in their new solution are done in shell or python scripts, is ignored. Similarly I don't see a problem with the user having to take care of the code that copies files from one supercomputer to the other, or to adjust a job header for a new computer/queuing system. These tasks are usually adjusted within a day, with individual tasks being on the order of 10 minutes for anybody used to the supercomputer.

Reading the manuscript, the reader is taken through the same processes that the authors must have gone through when re-working the coupling script(s). The presentation of the initial state as well as the solution duplicates a lot of the content of the manuscript, with the second presentation being better structured than the first one, just as the scripts have gained in structure. The detailed specifications of the I/O and jobs of all sub-components are important for the author of this script, but rather belong into a manual than into a scientific manuscript.

In the end, the authors present a solution with different model parts being run on different supercomputers so each of them can run with maximum efficiency. Sadly, we never learn, how much resources each of these bits requires and whether there is a point in increasing the complexity of the problem to two supercomputers with remote file transfers for gaining computational efficiency, or whether simply letting the smaller model run in a separate (slightly inefficient) job on the same computer would have been faster in the end. While the authors claim a focus of their work was on improving

performance, (p.10 l26 "focused in particular on improving overall runtime [,. . .]"), they never provide any information about the results of this endeavor in terms of reduction of overall runtime, or similar metrics.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-158, 2018.