Geosci. Model Dev. Discuss., 8, C3100–C3105, 2015
www.geosci-model-dev-discuss.net/8/C3100/2015/

Geoscientific Model Development Discussions

Interactive comment on "The infrastructure MESSy submodels GRID (v1.0) and IMPORT (v1.0)" *by* A. Kerkweg and P. Jöckel

A. Kerkweg and P. Jöckel

kerkweg@uni-mainz.de

Received and published: 27 November 2015

We thank the referee #2 for his/her helpful comments. Here is our reply.

GRID and IMPORT should be important submodels in the infrastructure MESSY. This manuscript presents the details of GRID and IMPORT. It may be interested to MESSY developers, users, as well as the ones who develop couplers or other model infrastructures. I have already implemented similar common modules in the latest version of our software recently while it is still very difficult for me to follow the details in the manuscript because the presentation is not good or even poor. I recommend that revisions are required before this work can be published

in GMD. Authors should reorganize the manuscript so as to significantly improve the presentation.

This manuscript focuses on the GRID and IMPORT submodels in the infrastructure MESSY. However, as a paper, it should not be limited to MESSY. For revision, authors should well address the following questions:

1. The key idea of GRID and IMPORT is to make part of preprocessing online in model integration.

Yes, indeed. We will add a paragraph to the introduction, stating the advantages of performing (parts of) the "classical pre-processing" on-line, thus better motivating our developments.

What are the requirements of preprocessing according to the current or even future status of Earth system modelling in the world? Which requirements are considered for the design and implementation of GRID and IMPORT, and why? There should be some discussions about the requirements that are not included in current GRID and IMPORT. Examples are welcome for the discussion of the requirements.

This will also be addressed in the introduction of the revised manuscript. We will strengthen this discussion (including examples) and show more clearly the current status of GRID and IMPORT.

2. How about the related works? It may be difficult to go through all related works because engineers always do not write papers. Many models already have modules for online "preprocessing". Authors try to achieve common

modules for various models. I believe that authors can quickly know whether a model have common modules according to the code or configuration system. Similar models in some well known models and infrastructures (if have) such as CESM, WRF, FMS and ESMF should be discussed and compared.

The referee is right. We hardly found any literature on data import. From what we found, mostly by asking people working with the models, and not by published model descriptions, it seems that mostly off-line pre-processing and direct import on the models grid is applied. Model couplers are discussed more often, but this is off-topic for this article. Nevertheless, we will address this issue in the introduction of the revised article.

3. The common modules in this manuscript generally focuses on online interpolation. What are the requirements of online interpolation according to the current or even future status of Earth system modelling in the world? To answer this question, various types of grids (including 3-D grids) and various remapping algorithms (or requirement for interpolation) should be discussed.

Note that "interpolation" describes only one class of grid transformation (or grid remapping) algorithms. In most applications (e.g., emission flux remapping for CCMs) conservative remapping is required. As discussed by Jöckel (2006)[1], the corresponding algorithms might differ, depending on if the quantities to be remapped are intensive or extensive. Ideally, a common module should be able to handle both and provide a variety of algorithms, which can be selected depending on the application. Our impression is that most models / couplers utilise the remapping / interpolation tools based on the SCRIP toolkit.

[1] Jöckel, P.: Technical note: Recursive rediscretisation of geo-scientific data in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 3557-3562, 2006.

We will discuss this in the revised version.

4. About the implementation of GRID and IMPORT. Here authors should answer how to make GRID and IMPORT support various types of grids, various remapping algorithms, 3-D interpolation in parallel, and various expressions of time information.

Most of the information answering those questions are part of the supplement. Nevertheless, we will provide more on that in the revised manuscript. GRID does support various types of grids and performs grid transformations in distributed memory parallelisation, depending on the domain decomposition of the used model. No interpolation in time is implemented so far, mostly for the reason that conservation constraints (e.g., time integrated flux (= mass) conservation), are difficult to achieve within time-stepping procedures, if concurrent access to the entire time series of data to be imported is to be avoided.

There may be some limitations in GRID and IMPORT, while authors should clearly discuss these limitations. For example, are these limitations because of the whole MESSY or other reasons, and how to solve these limitations in the future? Why and how about the design of the API and configuration format should be presented, corresponding to what are supported in GRID and IMPORT.

Indeed, GRID in IMPORT also have limitations. We will extend the manuscript by discussing those (e.g., the current lack of interpolation in time) wherever appropriate.

The last sentence above is unfortunately unclear. The API is documented in detail in the supplement.

5. How to use GRID and IMPORT? Some comprehensive examples are required.

The detailed supplement comprises information about (1) the usage of the stand-alone tools, and (2) how GRID can be used from within other MESSy submodels. This supplement is referred to in the manuscript. In the manuscript itself, we provide example namelists showing how to add data to be imported with IMPORT_GRID and IMPORT_TS. So, it is unclear what "comprehensive examples" should be?

6. How about the performance of GRID and IMPORT, especially the scalability of parallel interpolation? How about the comparison to the offline solution? I/O should be a bottleneck for both online and offline solution. How about the performance comparison when parallel I/O is used? It is possible that the online solution and offline solution outperforms in some cases and then the hybrid solution (for example, horizontal interpolation is processed offline and vertical and time interpolation is processed in parallel online) should be much better. Authors should discuss about that.

These are indeed important issues. And yes, the I/O is a bottleneck for both, on-line and off-line, solutions. The advantages of the on-line solution are a high flexibility w.r.t. to resolution. Initial and boundary condition data needs to be stored on disk only once, and not at every model resolution. Most important, some parts of the re-mapping can hardly be pre-processed off-line. For example, it is desirable to distribute emissions in global CCMs in the vertical. This vertical distribution might depend on the actual meteorological situation, or the vertical grid (e.g., a hybrid-pressure grid) might be time dependent. An adequate off-line pre-processing is not straightforward to achieve in those cases. Separating the horizontal and vertical remapping into off-line and on-line pre-processing, respectively, might be an option, though it will complicate the approach. Nevertheless, such split is naturally possible with IMPORT, because

the grid resolution of the imported data is arbitrary, implying the possibility that the horizontal grid matches the model grid.

The performance of the different approaches (off-line vs. on-line vs. hybrid) depends very much on the application, e.g., the amount of data to be imported, the applied model resolution, the parallel domain decomposition of the model and last, but not least on the computer architecture. Thus, a universal statement on performance cannot be provided. Nevertheless, we will add a short discussion about these issues in the revised manuscript.

According to the scalability: Again it depends on the overlying model, because the distributed memory parallelisation of IMPORT utilises the domain decomposition of the overlying model. Two possibilities are implemented: either one task performs the grid-transformation for the entire grid and distributes (scatters) the results, or each task performs the grid-transformation only for its own part of the entire grid. It is obvious that the first option does hardly scale, but the second option does. Again, a universal statement cannot be provided, but we will add this discussion to the revised manuscript.

As IMPORT is currently implemented, we use serial netCDF for data input. Neither parallel I/O, nor asynchronous I/O are implemented. These are promising options for the future and we will state this in the outlook section.

Interactive comment on Geosci. Model Dev. Discuss., 8, 8607, 2015.