

TM5-MP - Feature #8931

removing dependence on HDF4

02/07/2018 03:46 PM - Philippe Le Sager

Status:	In Progress	Start date:	02/07/2018
Priority:	Low	Due date:	
Assignee:		% Done:	30%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour
Description			
<p>I propose to get rid of the dependence on HDF4, at least for EC-Earth. This is very old software (HDF group recommends to switch to HDF5), and it clutters EC-Earth compilation configurations.</p> <p>We can already drive the model with netCDF met fields, and still have the option of running with hdf met fields. The only reason we must compile with HDF4 is for few input files. There are only 12 of them distributed with EC-Earth:</p>			
<pre>cca-login4 ~/perm/ECE3-DATA [1020] >>> find -L . -name "*.hdf" ./tm5/TM5_INPUT/boundary/ODIN/ODIN_CO_O3_ratio_4levels.hdf ./tm5/TM5_INPUT/boundary/ODIN/ODIN_Climatology_HNO3_O3_4levels.hdf ./tm5/TM5_INPUT/boundary/CH4_top/haloe_ch4vmr_91.hdf ./tm5/TM5_INPUT/natural_emissions/tracer/RN222_WMO2004.hdf ./tm5/TM5_INPUT/natural_emissions/reactive_gases/CH4/HYMN/CH4-natural-nonLPJ/CH4-N71-Sanderson-0000-sfc-glb100x100.hdf ./tm5/TM5_INPUT/natural_emissions/reactive_gases/CH4/HYMN/CH4-natural-nonLPJ/CH4-N40-Lambert-0000-sfc-glb100x100.hdf ./tm5/TM5_INPUT/natural_emissions/reactive_gases/CH4/HYMN/CH4-natural-nonLPJ/CH4-N70-Olson-0000-sfc-glb100x100.hdf ./tm5/TM5_INPUT/natural_emissions/reactive_gases/DMS/DMSland.hdf ./tm5/TM5_INPUT/natural_emissions/reactive_gases/DMS/DMSconc.hdf ./tm5/TM5_INPUT/land/lsm lai.hdf ./tm5/TM5_INPUT/land/soilph.hdf ./tm5/TM5_INPUT/land/landfraction.hdf</pre>			
<p>For the standalone version of the model, there are a lot more files, because of the stratospheric ozone data. Then there is also the case of the old "save file", which we should drop altogether. But for that we just need to get rid of the writing in parallel of the netcdf restart files first, because the only reason to keep old save file is for cases users cannot compile netcdf4 with parallel IO enabled. This can be done in the same time we fix the writing of the timeseries with Tommy.</p>			
<p>Ok this is low priority, but to keep in mind. I think we could easily convert to netcdf the few HDF files still needed for EC-Earth.</p>			

History

#1 - 02/12/2018 08:16 AM - Philippe Le Sager

Turns out that when you use *with_budget*, you also write a hdf files like *regionsg_glb300x200.hdf*, which we can get rid of I think.

#2 - 02/20/2018 01:13 PM - Philippe Le Sager

As of [r725](#), the netcdf restart files are not read/written in parallel anymore. Tommi, the same method can be applied to write your output more efficiently.

New hidden feature: if one uses *istart=32* and a tracer is missing in the restart file, the program continues and assigns a very small non-zero value to the missing tracer. Useful when adding tracers, which was one of the two reasons to switch back to "save file", and is irrelevant now.

#3 - 07/02/2018 09:33 AM - Philippe Le Sager

- Status changed from New to In Progress

- % Done changed from 0 to 20

A new task (8.4) to tackle this issue (among others in an effort to simplify the code) has been flagged by the Steering Committee on Friday, June 29th, 2018.

#4 - 07/13/2018 03:49 PM - Philippe Le Sager

- % Done changed from 20 to 10

Looking at [r835](#) and [r834](#). First, this is not working out-of-the-box with the CB05 project and the output it uses: at [r835](#) I got a compilation error, at [r829](#) a runtime error.

My proposal is to keep the use of HDF4 for hdf meteo only, since it is still available. On the long term we may get rid of it, but it should not be a problem to keep it as an option for the time being. For everything else, we should not remove HDF (except few exceptions), but replace it by the equivalent netCDF, since most of the model functionality should remain. We can prioritize and go step-by-step:

- leave hdf meteo available [nothing to do]
- remove the old restart (a/k/a save file) functionality [remove]
- dumpfield (in toolbox.F90) [remove, but somebody will ask for it one day]
- convert MMIX into netCDF (although we do not use it in EC-Earth, so it could be optional and available only if compiled with hdf)
- convert budget into netCDF (this can be optional, i.e. requires HDF4)
- convert all hdf input (except hdf meteo) into netcdf [code and data]
 - first limited set for EC-Earth
 - extend to all other inputs

I have probably missed some other hdf files (particularly some output), but I think that's the general idea. It requires careful testing and will not happen in one day.

We probably need to rollback [r835](#) and [r834](#). I can do it if you want.

#5 - 07/17/2018 02:20 PM - Philippe Le Sager

Rolled back [r835](#) and [r834](#), see [r856](#). Branch is in synced with trunk ([r857](#)). I got the same results (CB05+M7 project) in the [trunk](#) and in the [cleanup-udunits-hdf-ncep](#) branch when using Udunits and when I am not using Udunits. I have reintegrated the branch into the [trunk](#) ([r858](#)), since it addresses the Udunits issue ([#616](#)).

#6 - 11/18/2019 02:07 PM - Philippe Le Sager

- % Done changed from 10 to 30

Some progress have been made towards a HDF4-free model. The subset of hdf data **needed in EC-Earth** has been converted into netCDF, and the code to read them adapted. It is now possible to compile (and run) without a hdf4 library. This was developed in EC-Earth and has been ported to the TM5-MP repository in a new branch (<https://svn.knmi.nl/svn/TM5-MP/branches/no-hdf>). In [r1073](#), the change set from EC-Earth to run without having compiled against HDF4 has been included.

But at this stage, some functionality is missing: some output is not available. This is fine for EC-Earth simulations, but may be limiting for stand-alone TM5. The following output is muted (or has to be set to F) to run without HDF4:

- muted output (base) : budget, mmix (must be set to F explicitly in rc file, else code stop)
- meteo output (base) : possible only in netCDF

- muted output (proj/cb05) : photolysis
- muted output (proj/output): user_output_station.F90 (must be set to F explicitly in rc file, else code stop), user_output_noaa.F90
- user_output_cf.F90, user_output_flask.F90 : not available at all, since need to be ported to TM5-MP first...

Before removing all HDF-based features, we need to get the most needed ones to work with netcdf: mmix, photolysis, and budget.

Finally, this obviously requires new input files. They will be provided as soon as they are all ready.