

# ECCO v4 development notes

Gaël Forget

Department of Earth, Atmospheric and Planetary Sciences  
Massachusetts Institute of Technology

November 21, 2014

## abstract

These notes pertain to the revamped and augmented estimation modules of MITgcm; to the global setups and files used in regression tests; and to the analysis of the state estimate solution, including its re-runs. The intention is to compile and condense practical information for users and developers of ECCO v4 setups. Some of the included material will find its way in Forget and co authors (in preparation), and some should be added to the MITgcm [manual](#).

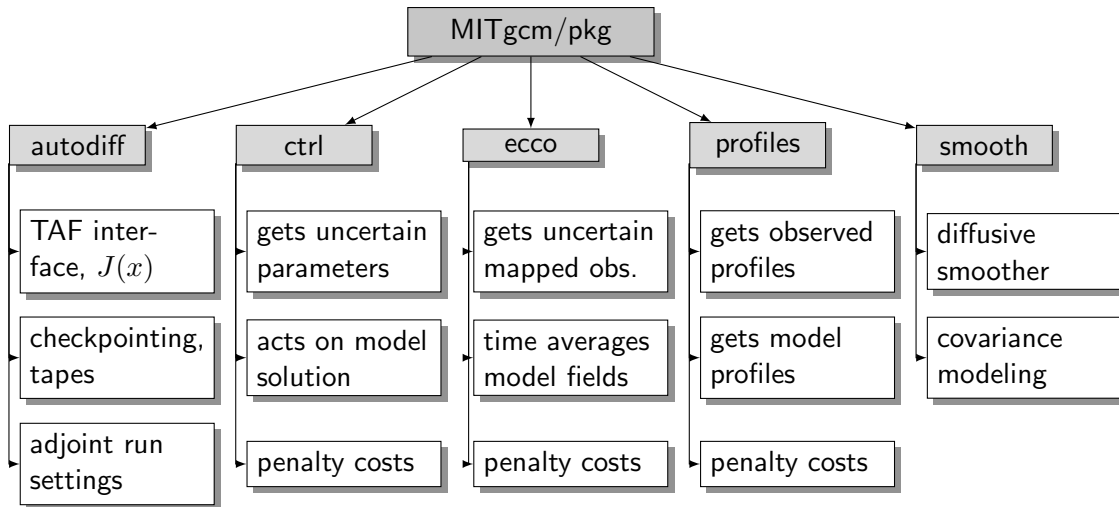


Figure 1: Organization and roles of MITgcm estimation modules.

## Contents

|          |                             |           |
|----------|-----------------------------|-----------|
| <b>1</b> | <b>downloads</b>            | <b>3</b>  |
| 1.1      | MITgcm . . . . .            | 3         |
| 1.2      | setup . . . . .             | 3         |
| 1.3      | solution . . . . .          | 4         |
| 1.4      | analysis . . . . .          | 4         |
| <b>2</b> | <b>MITgcm runs</b>          | <b>6</b>  |
| 2.1      | regression tests . . . . .  | 6         |
| 2.2      | full ECCO v4 runs . . . . . | 6         |
|          | <b>References</b>           | <b>11</b> |

# 1 downloads

## 1.1 MITgcm

Pre-requisites are cvs, gcc, gfortran (or alternatives), and mpi (only for parallel runs). Then :

- The MITgcm web-page is [mitgcm.org](http://mitgcm.org)
- Install MITgcm using cvs as explained @ [cvs](#)
- Run MITgcm using testreport as explained @ [manual](#), [howto](#)

For example, my laptop setup, including mpi and netcdf, involved the following mac ports :

- cvs @1.11.23\_1 (active)
- wget @1.14\_5+ssl (active)
- gcc48 @4.8.2\_0 (active)
- mpich-default @3.0.4\_9+gcc48 (active)
- mpich-gcc48 @3.0.4\_9+fortran (active)
- netcdf @4.3.0\_2+dap+netcdf4 (active)
- netcdf-fortran @4.2\_10+gcc48 (active)

Side note – overriding the default mac gcc and mpich with the above, further requires

- sudo port select –set gcc mp-gcc48
- sudo port select –set mpich mpich-gcc48-fortran

Side note – using mpi and netcdf within MITgcm requires two environment variables :

- export MPI\_INC\_DIR=/opt/local/include
- export NETCDF\_ROOT=/opt/local

## 1.2 setup

Pre-requisites are MITgcm (see above) and mpi (except for small setup). User can then install the ECCO v4 setups, as explained @ [README](#), using the [setup\\_these\\_exps.csh](#) shell script. This script downloads [global\\_oce\\_cs32/](#) (small setup), [global\\_oce\\_llc90/](#) (bigger setup) and [global\\_oce\\_input\\_fields.tar.gz](#) binary model inputs to [global\\_oce\\_tmp\\_download/](#) (local sub-directory). User can then move these directories to MITgcm/verification/ to allow for automated execution by [testreport](#) using [genmake2](#) (Fig.2).

Figure 2: MITgcm directory structure downloaded using [cvs](#). The ECCO v4 directories indicated with "+" were downloaded separately using [setup\\_these\\_exps.csh](#) script and moved to MITgcm/verification/.

```
MITgcm/
├── model/ (core of MITgcm)
├── pkg/ (MITgcm modules)
├── verification/
│   ├── testreport (shell script)
│   ├── aim.5l_cs (mitgcm regression test)
│   ├── + global_oce_cs32/ (for laptops)
│   ├── + global_oce_llc90/ (for computers)
│   ├── + global_oce_input_fields/ (inputs)
│   ├── hs94.128x64x5 (mitgcm regression test)
│   └── ...
└── tools/
    ├── genmake2 (shell script)
    ├── build_options (wrt compilers)
    └── ...
```

### 1.3 solution

The release 1 solution directory linked to [ecco-group.org](#) contains :

- 20 year solution output : [readme](#), [fields](#), [profiles](#), [grid](#)
- additional input files required to run the full 20 year solution (coming soon...).

### 1.4 analysis

Tools (e.g. matlab scripts) available to analyze the release1 solution, and others, include :

- download, set-up [gcmfaces](#) + [MITprof](#) using [shell script](#) or manually (see [getting\\_started.m](#))
- download [MITgcm/utls](#) using [cvs](#) (basic functionalities only).

The so-called [standard analysis.pdf](#) is generated in matlab by means of [diags\\_driver.m](#) and [diags\\_driver\\_tex.m](#) in the following sequence :

```
diags_driver('release1/', 'release1/mat/', 1992:2011);
diags_driver_tex('release1/mat/', {}, 'release1/tex/standardAnalysis');
```

assuming the conventional directory structure shown in Fig.3.

Figure 3: Directory structure as expected by gcmfaces and MITprof toolboxes. The toolboxes themselves can be relocated anywhere as long as their locations are included in the matlab path. Advanced analysis using diags\_driver.m and diags\_driver.tex.m will respectively generate the mat/ directory (for intermediate computational results) and the tex/ directory (for [standard analysis](#)). This diagnostic process relies on the depicted organization of GRID/ and solution/ for automation (user will otherwise be prompted to enter directory names) and depends on downloaded copies of [fields](#) to nctiles/ (local subdirectory).

```

./
├── gcmfaces/ (matlab toolbox)
│   ├── sample_input/ (binary files)
│   ├── @gcmfaces/ (matlab codes)
│   ├── gcmfaces_calc/ (matlab codes)
│   └── ...
├── MITprof/ (matlab toolbox)
│   ├── profiles_samples/ (netcdf files)
│   ├── profiles_process_main_v2/ (matlab codes)
│   ├── profiles_stats/ (matlab codes)
│   └── ...
├── GRID/ (binary output)
├── release 1 solution/
│   ├── diags/ (binary output)
│   ├── nctiles/ (netcdf output)
│   ├── MITprof/ (netcdf output)
│   ├── mat/ (created by gcmfaces)
│   ├── tex/ (created by gcmfaces)
│   └── other solution/
│       ├── diags/ (binary output)
│       └── ...
└── ...

```

## 2 MITgcm runs

Here I document a few procedures, commands and submission scripts that may be relevant to run the ECCO v4 MITgcm setup – either in short regression tests or for multi-decadal simulations such as the full 20 year state estimate.

### 2.1 regression tests

MITgcm and ECCO v4 regression tests are run using testreport utility (see Fig.3; [howto](#)). Serial regression tests can always be executed simply with, e.g.

```
./testreport -t global_oce_cs32
```

```
./testreport -skipdir global_oce_llc90
```

If something goes wrong and/or interrupts the process it is often safer to clean up experiment directories (e.g., by executing `./testreport -clean -t global_oce_*`) and start over. For example, the `global_oce_llc90` experiments require 12 processors in forward (96 in adjoint), and may crash your laptop if you attempted to run them in serial mode.

Often in massive computing environments, however, mpi jobs can only be run within a queuing system. The, machine specific, submission script in Fig.4 provides an example. It contains 3 hard-coded switches :  `fwdORad = 1` (2 for adjoint);  `numExp = 1` (2 for `llc90`);  `excludeMpi = 0` (1 for serial). This script is located and submitted from `MITgcm/verification`. If compute nodes cannot access the remote adjoint compiler (TAF), then proceed in two steps :

1. compile outside of the queuing system using e.g.

```
./testreport -of ../tools/build_options/linux_amd64_ifort+mpi_ice_nas \
-j 4 -MPI 96 -command 'mpiexec -np TR_NPROC ./mitgcmuv_ad' -ad \
-t global_oce_llc90 -norun
```

2. Before submitting the Fig.4 script, add `-q` to the `'opt'` variable to skip compilation.

Adjoint test require access to the TAF compiler. Then the call to testreport only needs to be altered by appending the `'-ad'` option. The ECCO v4 regression tests do not include the common, adjoint specific `'code_ad/'` directory, which is generally un-necessary. Since testreport relies on the existence of `'code_ad/'` for its adjoint option though, it is necessary to soft link `'code/'` to `'code_ad/'` in both `global_oce_cs32/` and `global_oce_llc90/` to run `'testreport -ad'` on those experiments.

### 2.2 full ECCO v4 runs

There are three main differences between regression tests and full model runs (see [howto](#)) :

- compilation and run are done without testreport and with compiler optimization.
- additional forcing, control vectors and/or observational inputs are often necessary.
- additional memory and disk space are often necessary.

75 The typical compilation sequence, taking ECCO v4 and pleiades as an example, is shown  
76 in Fig.5. On a different machine, another opt file (see build\_options) should be used. The  
77 tamc.h\_itXX and profiles.h\_itXX headers allow for additional time steps, and additional in situ  
78 data input, respectively. Also note that compiling the adjoint requires a TAF license. Once  
79 that is done, user creates and enters a run directory, links everything into place (see Fig.6), and  
80 finally submits a job to the queueing system (see Fig.7).

81 A mechanism, analogous to testreport but for long runs, has been introduced recently  
82 (Forget and co authors, in preparation) that is [testreport\\_ecco.m](#). Within matlab, its use re-  
83 quires 'MITgcm/verification/global\_oce\_llc90/results\_itXX/' to be in the path. By itself it com-  
84 pares cost functions and global mean time series to the state estimate values. These can be  
85 extended to meridional transports, which requires [gcmfaces](#)

86 **material for separate paragraph :** memory and disk space requirements (provide specs,  
87 timings); results of testreport; results of testreport\_ecco.m

Figure 4: Example script to run mpi testreport via a queueing system.

```

#PBS -S /bin/csh
#PBS -l select=1:ncpus=16:model=ivy+4:ncpus=20:model=ivy
#PBS -l walltime=02:00:00
#PBS -q devel
#PBS -m n

#environment variables and libraries
#-----
limit stacksize unlimited
module purge
module load modules comp-intel/2013.1.117 mpi-sgi/mpt.2.10r6 netcdf/4.0
#
setenv LD_LIBRARY_PATH ${LD_LIBRARY_PATH}:${HOME}/lib
setenv MPI_IB_TIMEOUT 20
setenv MPI_IB_RAILS 2
setenv MPI_IB_FAILOVER 1
setenv MPI_CONNECTIONS_THRESHOLD 2049

#local variables and commands
#-----
set fwdORad = 1
set numExp = 1
set excludeMpi = 0
#
if ( ${numExp} == 1 ) then
    set nameExp = global_oce_cs32
    set NBproc = 6
else
    set nameExp = global_oce_llc90
    set NBproc = 96
endif
#
if ( ${excludeMpi} == 1 ) then
    set opt = '-of ../tools/build_options/linux_amd64_ifort -j 4'
else
    set opt = '-of ../tools/build_options/linux_amd64_ifort+mpi_ice_nas -j 4'
endif
#
if ( ${fwdORad} == 1 && ${excludeMpi} == 0 ) then
    ./testreport ${opt} -MPI \
    ${NBproc} -command 'mpiexec -np TR_NPROC ./mitgcmuv' -t ${nameExp}
else if ( ${fwdORad} == 2 && ${excludeMpi} == 0 ) then
    ./testreport ${opt} -MPI \
    ${NBproc} -command 'mpiexec -np TR_NPROC ./mitgcmuv_ad' -ad -t ${nameExp}
else if ( ${fwdORad} == 1 && ${excludeMpi} == 1 ) then
    ./testreport ${opt} -t ${nameExp}
else if ( ${fwdORad} == 2 && ${excludeMpi} == 1 ) then
    ./testreport ${opt} -ad -t ${nameExp}
endif

exit

```

Figure 5: Compilation directives, outside testreport, for intensive model runs.

```
cd verification/global_oce_llc90/build
../../../../tools/genmake2 -optfile=\\
../../../../tools/build_options/linux_amd64_ifort+mpi_ice_nas -mpi

make depend

\rm tamc.h profiles.h
cp ../code/tamc.h_itXX tamc.h
cp ../code/profiles.h_itXX profiles.h

make -j 4

(or for the adjoint :
make adtaf
make adall -j 4
)
```

Figure 6: Example script to setup the 20 year ECCO v4 state estimate. It is implied that user has filled directories /bla, /blaa, /blaaa and /blaaa with appropriate forcing, observational, control vector, and pickup files.

```
#!/bin/csh -f

set forcingDir = ~/bla
set obsDir     = ~/blaa
set ctrlDir    = ~/blaaa
set pickDir    = ~/blaaaa

source ../input_itXX/prepare_run
cp ../build/mitgcmuv .
\rm pick*ta EIG*
ln -s ${forcingDir}/EIG* .
ln -s ${obsDir}/* .
ln -s ${ctrlDir}/xx* .
ln -s ${pickDir}/pick* .

exit
```

Figure 7: Example script to run the 20 year ECCO v4 state estimate on 96 processors.

```
PBS -S /bin/csh
#PBS -l select=1:ncpus=16:model=ivy+4:ncpus=20:model=ivy
#PBS -l walltime=12:00:00
#PBS -q long

#environment variables and libraries
#-----
limit stacksize unlimited
module purge
module load modules comp-intel/2013.1.117 mpi-sgi/mpt.2.10r6 netcdf/4.0
#
setenv LD_LIBRARY_PATH ${LD_LIBRARY_PATH}:${HOME}/lib
setenv MPI_IB_TIMEOUT 20
setenv MPI_IB_RAILS 2
setenv MPI_IB_FAILOVER 1
setenv MPI_CONNECTIONS_THRESHOLD 2049

#run MITgcm
#-----
mpiexec -np 96 dplace -s1 ./mitgcmuv
exit
```

## References

- Forget, G. and co authors, in preparation: ECCO version 4: A global ocean modeling and state estimation framework. *Geoscientific Model Development*.
- Forget, G. and R. Ponte, submitted: The partition of regional sea level variability. *Progress in Oceanography*.
- Hansen, P. C., 1992: Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, **34** (4), 561–580.
- Weaver, A. and P. Courtier, 2001: Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, **127** (575), 1815–1846.