A Hitchhikers Guide to the Black Arts
(of Earth system modelling)

I+II: cGENIE modelling basics + experiments in climate hysteresis

Stuff to keep in mind:

- *c*GENIE is a model. Models ARE NOT the 'real World'. (Don't get confused!)
- The very low (for a 3-D ocean circulation model) resolution of *c*GENIE limits its applicability to very short time-scales, and because the simplified atmospheric component cannot calculate winds, there is no atmospheric dynamics or inter-annual variability.
- *c*GENIE is best thought of as a 'discovery and exploring' tool for learning how the Earth system works rather than a detailed 'simulation' tool.
- Have fun (or at least try) ☺

Relevant reading:

(1) Primary model description + evaluation/discussion papers:

Edwards and Marsh [2005] (Climate Dynamics 24, 415-433)
→ description and calibration of the climate model component of (c)GENIE

Hargreaves *et al.* [2004] (Climate Dynamics 23, 745-760)
→ description of data assimilation methodology and calibrated climatology of the climate model

Ridgwell *et al.* [2007a] (Biogeosciences 4, 87-104)
→ description of basic ocean carbon cycle and its calibration against observations

Ridgwell *et al.* [2007b] (Biogeosciences 4, 481-492)
→ description of the calibration and application of the pH-responsive pelagic calcification

Ridgwell and Hargreaves [2007] (Global Biogeochemical Cycles 21, doi:10.1029/2006GB002764)
→ description and calibration of the sediment model component and response to fossil CO2 release

(2) Some potted model applications (mostly paleo):

Chikamoto *et al.* [2008] (JGR 113, doi:10.1029/2007JG00066)
→ deep-sea $CaCO_3$ sediment and atmospheric CO_2 response to AMOC shutdown

Panchuk *et al.* [2008] (Geology 36, 315-318)
→ configuration of (c)GENIE for Palaeogene marine carbon cycling; assessing PETM CO2 release

Ridgwell [2007] (Paleoceanography 22, doi:10.1029/2006PA001372)
→ description of sediment core modelling; application to the interpretation of PETM CCD changes

Singaraye *et al.* [2008] (GRL 35, doi:10.1029/2008GL034074)
→ analysis of ocean circulation and atmospheric radiocarbon variability during the Younger Dryas

(3) Snowball Earth and ice-albedo feedback & hysteresis

Hoffman and Schrag [2002] (*Terra Nova* **14**, 129-155)
→ Snowball review.

Hyde et al. [2000] (*Nature* **405**, 425-429)
→ Model analysis of the inception of a snowball Earth and ice-albedo thresholds.

Some of these references can be obtained from: www.seao2.info/pubs.html or with other *c*GENIE resources from: http://mycgenie.seao2.info

Andy (andy@seao2.org)

0.   Before anything else …

0.0   For some of you, the mechanics of running the model will be about as much fun as sticking you tongue in an electrical outlet (a popular hobby in England). (*However, if you are an experienced linux/unix/tounge-in-electrical-socket user, you can skip on to **Section 0.1**.*)

*c*GENIE has traditionally been configured and accessed at the 'command line' of the linux (or Mac equivalent) operating system. The command line is a place where you type text and when you press Return, something (hopefully, good!) happens. Typically the stuff you type started with a 'Command' word, and often followed by one or more options and parameters. The Command word and any options / parameters MUST be separated by SPACEs.

The command line is indicated with something like: `$`. The `$` is call the 'prompt' and is 'prompting' you to type some input (Commands, swear words, etc.). See – the computer is just sat there waiting for you to command it to go do something (stupid?)! Often, you will also be informed (reminded) of the username and computer name and current directory at the prompt, e.g.:

`[username@iwan ~]$`

Which is in this example is user '`username`' (yours will be different) on computer '`iwan`' and the current directory is the 'home' directory (`~`).

If you are not or not very familiar with the linux/unix command line – Appendix 1 contains a quick summary of some of the more important/useful Commands you can use at the command line.

NOTE: In the following workshop instructions, example lines to be typed at the command line <mark>are highlighted in yellow</mark> (or grey when photocopied ...). Be VERY CAREFUL that spaces are not missed out. Also be careful not to confuse the number one (`1`) for the letter el (`l`). Mis-spelling/typing will probably be the primary reason for any wailing and gnashing of teeth ...

NOTE-the-second: Text highlighted in grey represents instructions specific to using Windoz PCs, and hence may or will differ on different computers and/or under operating systems. Other software choices (e.g., for editing files or viewing results) and platforms (e.g. Macs) are possible.

0.1   Some warnings and reminders are repeated over and over and over and … over again.
Some warnings and reminders are repeated over and over and over and … over again.
This is because you will forget immediately each time! ;)

0.2   Additional documentation (of varying degrees of up-to-date-ness) lives in the `cgenie/genie-docs` directory, as well as on the interweb (http://www.seao2.info/mycgenie.html):
(i) The *c*GENIE *User Manual* (the least up-to-date of all!) [`cGENIE.User_manual.pdf`]
(ii) A set of *Examples* [`cGENIE.Examples.pdf`]
(iii) A *HOW-TO* (potted explanations of how to get useful stuff done) [`cGENIE.HOWTO.pdf`]
(iv) A table of model parameters [`cGENIE.Parameter_reference_guide.pdf`]
(v) A Guide to installing *c*GENIE [`cGENIE.QuickStartGuide.pdf`]

03   **OK – now we are ready to start …**

1. Starting (dozing?) off …

1.0 You are going to be installing the model from scratch – why? Why not? Hell, it saves me installing it a dozen times! Actually, having gone through the install and testing procedure once, you should be able to install cGENIE 'for real' another time. (If you dare ever use it again ...)

1.1 Having logged into the PC terminal, log in to the account that has been created for you on the almond computing cluster. To do this – first start the SSH Secure Shell program (you can find this via the All Programs menu tree from the Start icon in Windoz). Run the Secure Shell Client (rather than the Secure File Transfer Client) to start off.

Click on the Quick Connect icon in the main Secure Shell Client window. A 'Connect to Remote Host' dialogue box will appear. In the first box, 'Host Name', enter:

iwan.ggy.bris.ac.uk

and your computing cluster user-name on the line below this ('User Name').

When (and only when) you click on the Connect button will you be asked for your password in a new dialogue window that will appear.

You should now have a dull, blank-look window with the command prompt ($) (see previous page).

For displaying directory contents and transferring model results files we are going to be using the SSH file transfer program. One of the icons towards the middle of the icon bar in the SSH terminal window you have opened is called 'New File Transfer Window' (it should be the 2nd icon to the right of the search (binoculars) icon). Click on this to open a file transfer window.

Or: Again from the All Programs menu tree of Windoz `Start`, select `Secure File Transfer Client`.

You should now have TWO windows open – a 'shell' window (lines of text on an otherwise blank screen) and a file manager (transfer) window. Ensure that you have both these before moving on. It is recommended that you maximize both these windows to full screen. (But no-one will die horribly for not doing so. Probably.).

1.2 The next step is to download a copy of the source code for the *c*GENIE model as follows (all this is done in the shell window of your `iwan` computer account):

(1) To get a copy of *c*GENIE: from your home directory type:

```
$ svn co https://svn.ggy.bris.ac.uk/subversion/genie/branches/cgenie.EC4.2012
  --username=genie-user cgenie
```

All this must be on a <u>continuous</u> line, with a S P A C E before '`--username`', and before `cgenie`.

(If you get asked about a 'certificate' – enter '`p`' to accept 'permanently'.)

Depending on the weather and phase of the moon, you may (or may not!) be asked for a password. If so – it is: `g3n1e-user`. Again – don't mix up the ONE ('`1`') with an 'el' ('`l`'). (The mystery character in question is a 'one' if you must know …)

Some of the files that are installed are quite large, and at times, it may appear like it is 'stuck'. Be patient! Installation should be complete within about 2-3 minutes (longer if everyone else is trying to do this at exactly the same time …)

(2) Change directory (see: Appendix I and also Figure 1.1) to `~/cgenie/genie-main` – if you are not 'linux-friendly' – maybe at first do this in steps – list the contents of the directory (ls) to check where you are, then change to cgenie (cd cgenie), then list again (ls), then change to genie-main (cd genie-main), and then type:

```
$ make testbiogem
```

This compiles a carbon cycle enabled configuration of *c*GENIE and runs a short test, comparing the results against those of a pre-run experiment (also downloaded alongside the model source code). It serves to check that you have the software environment correctly configured.

There may be some 'Warnings' reported (== someone's sloppy programming) but these are not detrimental to the ultimate science results (we hope!). 'Success' of this test is indicated by:

```
**TEST OK**
```

3

You can then be certain that the model you have installed is producing identical (within tolerance) results to everyone else in the World who has ever installed cGENIE. Note that the model will pause for a l o n g time at the line: `./genie.job -t -k -f configs/eb_go_gs_ac_bg_test.xml -o /home/genie00/cgenie_output -c /home/genie00/cgenie -g ../../cgenie -m "" > testbiogem.out`; – this is quite 'normal' – the model is thinking!

(3) At this point, the science modules have been compiled conforming in size to a grid and/or number of tracers configuration that is unlikely to be what you want for running experiments. Clean up (remove) all the compiled cGENIE modules used to run the test so that the model can be recompiled afresh from the source code. This clean-up and complete re-setting of the model is achieved by typing:

`$` `make cleanall`

(4) That is it as far basic installation goes.

You should now have the following the following directories (command: ls to list from your home directory) in your account home directory:

`cgenie`   [a directory containing model source code and experiment configuration]
`cgenie_output`   [a directory containing the results of the model experiments]

Refer to Appendix I for the relevant linux commands if you need to.

1.3   Later on you will be editing some configuration files. So now might be a good time to check that you can use the editor! (You will also be using the same editor to view some of the model output.)

You have two alternative options for editing and viewing text files, depending on whether you are a UNIX nerd with no life, or prefer anything to do with computers to be wrapped in cotton wool and covered with dollops of treacle.

EITHER: Use the linux `vi` application (or similar) if you are familiar with it. I think that this pretty much sucks as a text editor and life is far too short and brutal … so I will not *make* you ;)

OR ... Use a suitable linux-friendly text editor (NOT Micro$oft Notepad) in conjunction with the Secure File Transfer Client. For example: '`SciTE`' http://www.scintilla.org/SciTE.html is suitable.

To set SciTE to automatically open the model configuration files: in the Secure File Transfer Client – go to Edit, then Settings… and from the SECOND File Transfer section of the list in the left-hand panel (there are TWO File Transfer section entries and you are after the one near the bottom of the lest), click on the button for If a file association is missing, use this application to open the file. Then from `Program Files` on the C drive of the Windoz machine, find the SciTE editor directory (called: Scintilla Text Editor) and its program executable inside (`SciTE.exe`). Click on `OK` to close the `Settings` dialogue box. Then under `File` in the main menu: `Save Settings`.

It should now be possible to double-click on a file in the Secure File Transfer Client and it will open like magic (almost)! Try opening `runcgenie.sh` (directory: `cgenie/genie-main`) in this way. If you edit and save the file, you will be asked whether you want to transfer it back to the remote machine and also whether you want to over-write the original file. Simply click `Yes` to both.

If you log out of Windoz then you may have to re-do these settings … ☹

1.4   Next, we are going to be going through how to run and analyze the results of the model, playing around with the 'snowball Earth' hypothesis and assessing the critical ice-albedo feedback in the climate system as an example.
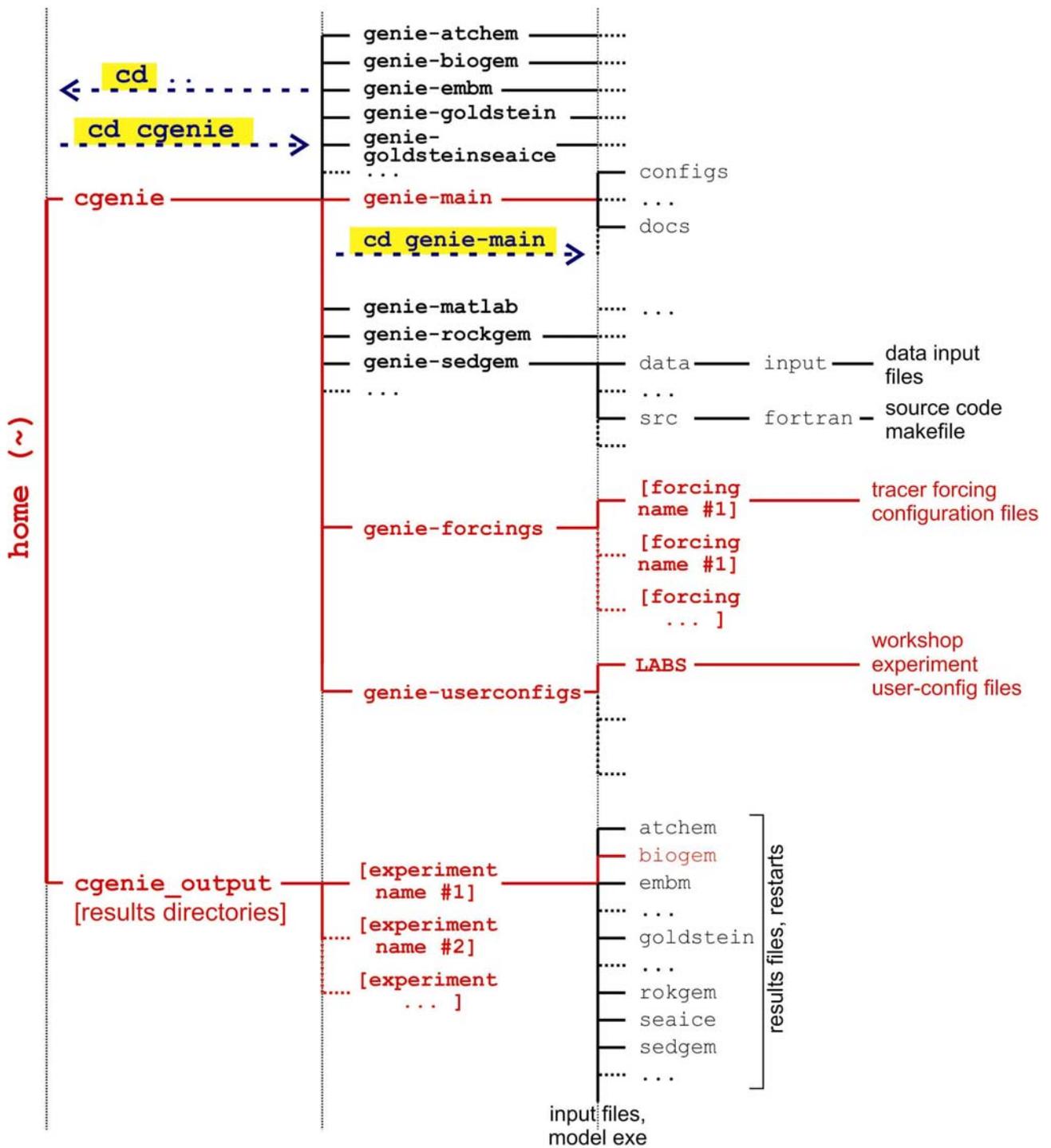
Figure 1.1 *c*GENIE model directory structire

Directory structure of the *c*GENIE model. Highlighted in red are directories and sub-directories that you will need to access at some point. Highlighted in blue are thieving down and up the directory tree.

2.   Running the model ('interactively', in a shell window)

2.1   The strategy for running the model is as follows:

At the command-line (`$`) in the `genie-main` directory (<u>not</u> your home directory), you will be entering in a command (`runcgenie.t100.sh`) together with a list of parameters that will be passed to the model, and as if by magic the model will run (or sometimes not):

`$ ./runcgenie.t100.sh #1 #2 #3 #4 (#5)`

You must list at least 4 parameters after `./runcgenie.t100.sh`, separated by S P A C E S and on a <u>single continuous line</u> (even if it 'wraps' around across 2 lines of the screen). For example:

`$` <mark>`./runcgenie.t100.sh cgenie_eb_go_gs_ac_bg.p0650a.NONE.t100 LABS exp4_snowball 20`</mark>
parameter number =   ↑ #1   ↑ #2 ↑ #3   ↑ #4

(Take care not to confuse an el ('l') with a one ('1') when typing this in ... (it is a 'one' here).)

These parameters are:

#1   … is the name of the required base (or 'basic') configuration ('*base-config*') of the model.

Initially, we will mostly be using the base config: **`cgenie_eb_go_gs_ac_bg.p0650a.NONE.t100`**, which specifies the cGENIE climate model components: GOLDSTEIN ocean (`go`) + (GOLDSTEIN) sea-ice (`gs`) + EMBM atmosphere (`eb`) together with ocean (`bg`) and atmosphere (`ac`) carbon cycle modules (and a FULL-ish selection of ocean geochemical tracers). The ocean circulation component is configured with 8 vertical levels and is forced with annual average insolation.

#2   … is the name of the subdirectory (if any) containing the user configuration ('*user-config*') file (i.e., the file containing the specification of a particular experiment). In the particular file structure adopted here, it will be: **`LABS`**. (If a *user-config* to be used does not live in a subdirectory of `cgenie/genie-userconfigs`, then a '/' is passed as this parameter (a blank or space will be hideously mis-interpreted …).)

#3   … is the name of the experiment itself. There must be a file in the directory specified by parameter #2 (`cgenie/genie_userconfigs/LABS`) with exactly the same name as you enter here for parameter #3. i.e. the actual experiment is defined by the *user-config* file. Here, the experiment is called: **`exp4_snowball`**.

#4   … is the run length of the experiment in years – this must be entered as an integer (even though cGENIE will actually be treating it as a real number). In this example, we are running for only **`20`** years.

There is also one optional (5th) parameter (not used in this particular example and so left out (blank)):

#5   … is the full path (and name) of any model experiment that you wish to continue on from the end of (called a '*restart*' file). If the 5th (optional) parameter is not passed then cGENIE will run from 'cold'. If the 5th parameter is present then cGENIE will attempt to run from a previously generated (*restart*) state. *Restarts* will be discussed later ...

2.2   As an example of running the cGENIE Earth system model, you are going to start a 'modern' (preindustrial) experiment. You will need the following details in order to construct the appropriate run command:

parameter #1: The base config is: `cgenie_eb_go_gs_ac_bg.p0650a.NONE.t100`

parameter #2: The user config directory is: LABS

parameter #3: The user config file (the experiment name) is: `exp4_snowball`

parameter #4: Run the experiment for eleven years: `20`

parameter #5: (There is no restart file, and so no 5th parameter needs to be passed …)

The full command, which you are going to issue from the `~/cgenie/genie-main` directory, looks like:

`$` <mark>`./runcgenie.t100.sh  cgenie_eb_go_gs_ac_bg.p0650a.NONE.t100  LABS  exp4_snowball  20`</mark>

REMEMBER: This must be entered on a single CONTINUOUS LINE. The (single) S P A C E S are vital. Take care not to confuse an el ('l') with a one ('1') when typing this in ... (it is a 'one' here). (<u>The single spaces between the parameters have been somewhat exaggerated here.</u>)

2.3   What should happen is: First, you will end up twiddling your thumbs a while, as all the components of *c*GENIE are compiled from the raw source code (FORTRAN). When it has finished doing this, the model will initialize and carry out some brief self-checking. Only then will it start actually 'running' and doing something, starting with a header describing the columns of numbers that follow:

`model year`   – (Guess!)
`AMO(Sv)`   – Atlantic meridional overturning circulation (Sv) (assuming modern continents)
`ice(%)`   – global sea-ice fraction (%)
`<SST>`   – global sea surface temperature 'SST' (°C)
`<SSS>`   – global sea surface salinity 'SSS' (‰)

The choice of what information to display on screen as the model is running is rather arbitrary, but the chosen metrics do tend to summarize some of the main properties of the climate system and carbon cycle – for my own personal convenience rather than reflecting any fundamental scientific truth ... In later experiments involving the carbon cycle, you will also see columns of information for:

`pCO2(uatm)`   – mean atmospheric CO_2 concentration (in units of μatm)
`d13CO2`   – mean δ^{13}C value of atmospheric CO_2 (‰) (NOTE: only valid if ^{13}C tracer is selected)
`<DIC>`   – global mean ocean dissolved inorganic carbon (DIC) concentration (μmol kg^{-1})
`<ALK>`   – global mean ocean alkalinity (ALK) concentration (μeq kg^{-1})

This information is reported at the same intervals as time-series data (see later & refer to the *User Manual*) is saved and is indicated by:

>>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :

Interleaved with these lines are lines reporting the saving of time-slice data (the 2- and 3-D model states – more of which later and in the *User Manual*). These appear as:

>>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year:

When the model ends (as per the value you specified at the command line), a bunch more lines of diagnostics things are reported. If you scroll up the screen of the terminal window you can get back to where the time-dependent information appears while the model was running.

You can also stop the model at any point (all data up to that time will have been saved) by hitting: `<Ctrl-C>` (CONTROL key + 'C' key).

Just from examining the screen output: how close to steady state does the system appear to have come after this time? i.e., do are SST and/or sea-ice extents appear to be converging towards stable (constant) values? This is an important question to think about later on: 'has the model reached steady-state (and does it matter)?'

The output should look something like the following:

```
      ********************************************************
       *** Initialisation complete: simulation starting ...
      ********************************************************

           model year  *  AMO(Sv)  ice(%)   <SST>   <SSS>

      ***        0.00     0.000   0.000   -0.000  34.900
      >>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year  :       0.500
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       0.500
      ***        1.00     2.182   0.000    1.737  34.902
      >>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year  :       1.500
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       1.500
      ***        2.00     2.343   0.000    2.866  34.904
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       2.500
      ***        3.00     2.473   0.000    3.698  34.905
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       3.500
      ***        4.00     2.582   0.000    4.361  34.908
      >>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year  :       4.500
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       4.500
      ***        5.00     2.675   0.000    4.912  34.910
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       5.500
      ***        6.00     2.755   0.000    5.385  34.912
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       6.500
      ***        7.00     2.827   0.000    5.800  34.914
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       7.500
      ***        8.00     2.889   0.000    6.170  34.915
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       8.500
      ***        9.00     2.881   0.000    6.512  34.917
      >>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year  :       9.500
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :       9.500
      ***       10.00     2.929   0.000    6.826  34.918
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      10.500
      ***       11.00     2.971   0.000    7.116  34.919
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      11.500
      ***       12.00     3.010   0.000    7.389  34.920
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      12.500
      ***       13.00     3.046   3.271    7.960  34.922
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      13.500
      ***       14.00     3.089   6.161    8.506  34.923
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      14.500
      ***       15.00     3.179   8.227    8.967  34.925
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      15.500
      ***       16.00     3.152   8.329    9.188  34.926
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      16.500
      ***       17.00     3.126   8.357    9.393  34.928
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      17.500
      ***       18.00     3.097   8.370    9.586  34.929
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      18.500
      ***       19.00     3.065   8.377    9.770  34.930
      >>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year  :      19.500
      >>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :      19.500
      ***       20.00     3.031   8.382    9.946  34.931
      >>> SAVING BIOGEM RESTART <<<
      >>> SAVING ATCHEM RESTART <<<

      ********************************************************
       *** Simulation complete: shutdown starting ...
      ********************************************************
```

3. Model output

3.1   The first thing to note about output (i.e., saved results files) from *c*GENIE is that every science module saves its own results in its own sub-directory (and sometimes in very different and difficult-to-fathom ways …) – see Figure 1.1. All the sub-directories of results, plus copies of input parameters and the model executable, are gathered together in a directory that is assigned the same name as the experiment (= *user-config* file name). The experiment results directories live in:

`~/cgenie_output`

and will be assigned a directory name something like:

`exp4_snowball`

Within this are each module's results sub-directories.

We will consider primarily only results saved by the ocean biogeochemical module 'BIOGEM' (subdirectory: `biogem`). The results files will thus be found in:

`~/cgenie_output/exp4_snowball/biogem`

BIOGEM has a flexible and powerful facility of saving results by means of spatially explicit '*time-slices*', and as a semi-continuous '*time-series*' of a single global (or otherwise representative mean) variable. In contrast, ATCHEM does not save its own results (BIOGEM can save information about atmospheric composition and air-sea gas exchange) while SEDGEM essentially saves results only at the very end of a model experiment (BIOGEM can also save the spatial distribution of sediment composition as time-slices as well as mean composition as a time-series). Furthermore, in order to attain a common format for both ocean physical properties and biogeochemistry, BIOGEM can save a range of ocean results in addition to temperature and salinity, such as: velocities, sea-ice extent, mixed layer depth, convective frequency, etc

Time-slices

One of the most informative data sets that can be saved is that of the spatial distribution of properties (such as tracers or physical ocean attributes). However, saving full spatial distributions (a 36×36×8 array) for any or all of the tracers each and every time-step is clearly not practical; not only in terms of data storage but also because of the detrimental effect that repeated file access has on model run-time.

Instead, BIOGEM will save the full spatial distribution of tracer properties only at one or more predefined time points (in units of years). These are termed *time-slices*. At the specified time points, a set of spatially-explicit data fields are saved for all the key tracer, flux, and physical characteristics of the system. However, rather than taking an instantaneous snapshot, the *time-slice* is constructed as an average over a specified integration interval (the default is set to 1.0 years, i.e. an annual average).

Time-series

The second data format for model output is much more closely spaced in time. Model characteristics must then be reducible to a single meaningful variable for this to be practical (i.e., saving the time-varying nature of 3-D ocean tracer distributions is not). Suitable reduced indicators would be the total inventories in the ocean and/or atmosphere of various tracers (or equivalently, the mean global concentrations / partial pressures, respectively). Like the *time-slices*, the data values saved in the *time-series* files represent averages over a specified integration interval (the default is set to 1.0 years (annual average)).

3.2   The results directory will contain files with names of the form:

(i)   `biogem` (is the re-start file created form the run you have just complete, and can be ignored).

(ii)   `biogem_series_*.res` – these are the time-series files (in ASCII / plain text format).

(iii)   `biogem_year_*_diag_GLOBAL.res` – these contain (global diagnostics) summary information and are saved at the same frequency as the time-slices (also as ASCII / plain text).

(iv)   `fields_biogem_2d.nc` – 2-D fields of ocean and atmosphere properties, as NetCDF.

(v)   `fields_biogem_3d.nc` – 3-D fields of ocean properties, as NetCDF.

4. Viewing time-series output

4.1 A descriptive summary of all the time-series (`biogem_series_*.res`) data files is given in the cGENIE *User Manual* if you are really that bored. The files of most immediate use/relevance are:

`biogem_series_atm_humidity.res` - mean atmospheric (surface) humidity

`biogem_series_atm_temp.res` - mean atmospheric (surface) air temperature

`biogem_series_misc_opsi.res` - mean ocean overturning stream-function; Atlantic/global min, max

`biogem_series_misc_seaice.res` - mean ocean sea-ice cover and thickness

`biogem_series_ocn_sal.res` - mean ocean surface and whole ocean salinity

`biogem_series_ocn_temp.res` - mean ocean surface and whole ocean temperature

4.2 The contents of the files can be viewed by changing directory into the results directory, i.e.,:

`$ cd ~/cgenie_output/exp4_snowball/biogem`

and opening a file in the `vi` editor. But then I promised that I would not suggest anything quite so unpleasant ...

Instead – change to the experiment results directory and then to the BIOGEM sub-directory in the Secure File Transfer Client, and try double-clicking on one of the .res files (listed above). For `biogem_series_ocn_temp.res`, you should see 4 columns – time, mean (whole) ocean temperature (°C), mean (sea) surface temperature 'SST' (°C), and mean benthic temperature (°C). Other results files may differ in the numbers of columns but all should be identifiable from the header information.

Note: The Secure File Transfer Client **does not** automatically refresh the directory listing. If you cannot see the results sub-directory with the experiment name you have just run 99 times out of 100, it is because the display of the Secure File Transfer Client needs to be refreshed – there is an icon at the top of the program window or hit the 'F5' key.

4.3 <u>For your information</u> and edification: Excel, or MUTLAB if you prefer, can be used to graph the time-series results. Either way you will have to deal with the header line(s) that are present at the top of the file (and preceding the rows of data).

In Excel: Chose File then Open. You will want to select Files of Type 'All Files (*.*)'. In the Text Import Wizard window you can request that Excel skips the first few lines to start the import on the 2nd or 3rd line of the text file. Alternatively, set appropriate column widths manually in Excel to ensure that the columns of data are correctly imported.

MUTLAB will ignore lines starting with a `%`, which the time-series starts with. However, it may be that the header line wraps-around and there is in effect a 2nd header line but without a `%`. In this case, extra care (or a quick edit of the header in the ASCII file) will be required to load the data into MUTLAB.

5. Viewing time-slice output of 2-D and 3-D environmental property fields

5.0 For the time-slice NetCDF (`*.nc`) files you will be using a program called 'Panoply'. If you want your own (FREE!) copy of this utility, you can get it here (and is available for: Windoz, Mac, and linux operating systems): http://www.giss.nasa.gov/tools/panoply/.
Alternatively: there is a Windoz copy sat on the cGENIE website (www.seao2.info/mycgenie.html).

5.1 The first time that you try and view a NetCDF file you will need to 'associate' the file format (the `.nc` file format extension) with the Panoply program. To do this: (1) copy across to a local directory either the 2-D or 3-D NetCDF results file from the BIOGEM results sub-directory, (2) right-mouse-button click over the file, and chose `Open With…`, (3) `Browse …` and find the Panoply program file – this is mapped to your 'T' drive and is located at: `T:\Panoply\Panoply.exe`. Select and tick: `Always use the selected program to open this kind of file`. Hopefully … all you have to do now is to double-click on the `.nc` file icon in the File Transfer window and the time-slice will be displayed like … something that is "*indistinguishable form magic in a sufficiently retarded civilization*" (mis-adapted from A. C. Clarke).

5.2 When you open the NetCDF file, you will be presented with a 'Datasets and Variables' window (on the left hand side of the application window). This contains a list of all the parameters available that you can display. You will find that the 'Long Name' description of the variable will be the most helpful to identify the one you want. Simply double-click on a variable to display. For the 3-D fields you will be asked first whether you want a 'Lon-Lat' or 'Lat-Vert' plot (for the 2-D fields, the plot display will immediately open).

For the 'Lon-Lat' plots – there are multiple levels (depth layers) in the ocean of data that can be plotted, form the surface to the abyssal ocean.

For the 'Lat-Vert' plots – there are multiple possible longitudes at which to plot slices. The default is the global mean meridional distribution.

For both: there may be multiple time-slices (i.e., you can plot data saved from different years).

You can interpolate the data or not (often you may find that it is clearer not to interpolate the data but to leave it as 'blocky' colors corresponding to the resolution of the model), change the scale and colors, overlay continental outline, change the projection, etc etc. Gray cells represent 'dry' grid points, i.e., continental or oceanic crust.

NOTE: The default settings in Panoply can mislead:

(1) the 1st time-slice (often year mid-point 0.5) rather than the experiment end,

(2) auto-scaling of the color scale,

(3) global zonal averaging of lat-depth plots.

Be careful when opening a new plot that you are looking at what you *think* you are looking at …

5.3 Explore different data fields and play with different ways of displaying them. Aim for a set of display properties that show the information you are interested in / want to present in the clearest possible manner. Try different years (time-slice number), depth level (for a lat-long plot), or longitude (for a vertical section). Note that gray cells represent 'dry' grid points, i.e., continental or oceanic crust.

For example -- go to the results directory of the experiment and open up the 2-D NetCDF file (from the `biogem` directory). View the topography (ocean depth) field (`grid_topo`) – land is shown in grey by default. Also by default, Panoply will try and plot on the modern continental outline – turn this off by selecting `<none>` for Overlay in the Map tab. The model has been configured for an idealized super-continent, positioned symmetrically about the Equator, so be shocked (and have expected the modern continental configuration, which you will see in subsequent experiments).

5.4 To save plots in Panoply:
```
File
Save Image As …
```
Then select the location, filename, and graphics format.

6. Submitting experiment 'jobs' to a cluster

6.0 This bit is no particular fun at all, but it is a very handy 'trick' for running the model in the background, and maximizes drinking time in the pub vs. sat bored watching a computer screen ☺

6.1 Running jobs interactively is all very well, but there are three important limitations:

(1) The connection between your terminal and the server computer running the model must remain unbroken. Anything more than a fleeting loss of internet connectively may result in the experiment terminating.

(2) You can only run one experiment at a time … unless you want to have thousands of separate X-Windows open …? I thought not …

(3) Any cluster or computer you are likely to be accessing using a shell will not have many computing cores, either because it is a single machine with only one or two processors, or if a cluster, by using an X-Window you a running on the 'head node', which will have similar computing core limitations to running on a single machine. The more experiments you run simultaneously, the slower they will all run …

6.2 The alternative is to submit your experiment as a 'job' to a queuing system which then manages what compute resources are used to run the model. Once you have submitted the experiment, that is it – you can go straight to the pub :)

Unless you want to read the cGENIE *User Manual*, for now, take the following suggested way of submitting jobs on trust: to run a similar experiment to before (`exp4a_snowball`) for 20 years (or longer if you wish) by submitting the experiment as a job to the cluster queue, type:

```
$ qsub -q cat.q -j y -o cgenie_output -V -S /bin/bash runcgenie.t100.sh
 cgenie_eb_go_gs_ac_bg.p0650a.NONE.t100 LABS exp4a_snowball 20
```

(Again: <u>SINGLE, CONTINUOUS LINE</u>.)

Note that now you should omit the './' bit before `runcgenie.t100.sh`.

(If you are interested (I know that you are not): the options following `qsub` and before `runcgenie.t100.sh` do things like re-directing screen output and error messaging to a file and specify which linux 'shell' to assume. It is even possible to receive an email when the job is done ☺)

The status of the cluster queue and how you experiment job is getting on (e.g., "Is it finished yet?") can be checked by typing:

```
$ qstat -f
```

(`qstat -f -u "*"` will show all jobs on the cluster.)

After submitting an experiment, you receive a job number. This number appears in the first column in the queue status information when you issue a `qstat -f` command. You should see your job appear on one of 2 compute nodes of the 'cat' queue, numbered `0-2` through `0-9` (although there are 10 compute nodes in total on the full queue, numbered from `0-0` to `0-9`), although it might briefly reside as a '`PENDING JOB`'. For each node, there are 8 processing cores, meaning that up to 8 instances of cGENIE can run simultaneously on each node (i.e., a total of 80 across the entire cluster). For an 8-level ocean based configuration of *c*GENIE, being run for 10 years, the job should remain there in the queue for a few good tens of seconds before 'disappearing' (your clue that it has finished, or died …). If you periodically re-issue a `qstat -f` command you can follow your job's progress.

A rough rule of thumb is that 8-level ocean *c*GENIE will simulate about 1000 years per CPU hours. The 16-level version (which you will use later), about 300-400 years per CPU hour.

NOTE: If your experiment appears on the queue but vanishes after a few seconds, it has most likely died ☹

7. 'Restarts'

7.0 Not much fun here either … but again – an important and time-saving (== increased drinking time!) modelling technique to learn to use.

7.1 By default, model experiments start from 'cold', i.e., the ocean is at rest and uniform in temperature and salinity while the atmosphere is uniform in temperature and humidity. All biogeochemical tracers in the ocean have uniform concentrations and/or are zero and there are no biogenic materials in deep-sea sediments. From this state it will take several thousand years (kyr) for the climate system to reach steady-state, and closer to 5 kyr (or more) for ocean biogeochemical cycles and atmosphere CO_2 to reach steady-state, and exceeding 100 kyr for sediment composition to re-balance weathering ... Reaching this the equilibrium state is called the 'spin-up' phase of the model.

There is evidently little point in repeating the spin-up for each and every model experiment that are similar except in a single detail (e.g., testing a variety of different CO_2 emissions scenarios all starting from current year 2012 conditions). A facility is thus provided for requesting that a 'restart' is used – starting a new experiment from the end of a previous one, usually a 'spin-up' that has been run explicitly for the purpose of generating a starting point (*restart*) of the system at steady-state (equilibrium) for subsequent experiments to continue on from.

7.1 You will also need to download a *restart* file that has been created (a pre-run 10,000 year spin-up). To fetch this: Change to the `cgenie_output` directory, and type:

```
$ wget http://www.seao2.info/cgenie/labs/EC4.2012/exp0_snowball_SPINUP.tar.gz
```

This downloads an archived/compressed copy of the restart (`exp0_snowball_SPINUP.tar.gz`) from a location on the interweb. Extract the contents of this archive by typing:

```
$ tar xfzv exp0_snowball_SPINUP.tar.gz
```

You'll need to change directory back to `genie-main` to run the model.

7.3 A *restart* can be requested by setting the 5th (optional) parameter when entering in the `runcgenie.t100.sh` command. A spin-up of a pre-snowball sort of climate state is provided: `exp0_snowball_SPINUP` (which can be found in the `~/cgenie_output` directory).

The experiment `exp4b_snowball` can be told to use this *restart* state by adding an appropriate 5th parameter at the command line, e.g.:

```
$ ./runcgenie.t100.sh cgenie eb_go_gs_ac_bg.p0650a.NONE.t100 LABS
 exp4b_snowball 20 exp0_snowball_SPINUP
```

The run-time output should now look noticeably different. There should be no (very little) drift in any of the various variable values outputted to the screen – this is because you have (re-)started from the end of a run that had already ready an equilibrium, steady-state.

8. Brrrrrrrrrrrr – it's chilly on ... snowball Earth!

8.0 To illustrate how 'easy' it can be to configure an Earth system model such as *c*GENIE and explore the behavior of the Earth system and its response to perturbation – you are going to induce an extreme cooling of climate and see what happens. Solar output was weaker during the late Neoproterozoic, a time when the Earth experienced a series (2 ish) of extreme glaciations. Thus, having a mild climate state to start with must have been dependent on sufficient CO_2 and/or CH_4 in the atmosphere and hence presumably highly elevated compared to the modern World, so sort of the opposite of the problem we have today …

8.1 You are going to be running experiments similar to before and using the *restart*:

```
$ ./runcgenie.t100.sh cgenie_eb_go_gs_ac_bg.p0650a.NONE.t100 LABS
exp5_snowball 100 exp0_snowball_SPINUP
```

Your task will be to determine the radiative forcing (or pCO_2 equivalent) threshold required to drive the climate system into a full ice-covered ocean (snowball Earth) state.
(Read the *Hyde et al.* [2000] paper as a minimum.)

Useful 2-D (netCDF—Panoply) variables to view are surface air temperature and sea-ice extent (and/or thickness). Ocean surface temperature and salinity can be viewed in the 3-D NetCDF results file (apologies for ocean temperature being in units of K …). Time-series (ASCII `.res` files) are useful for providing simple mean indicators of global climate such as global ocean fractional sea-ice covered.

Note that the model configuration of an idealized super-continent, positioned symmetrically about the Equator, is pretty unrealistic. But the further you go back in time, the more uncertain it becomes as to exactly where and in what orientation the continents were. Sometimes modelers have to resort to somewhat idealized experiments if the uncertainties are too great. In addition, one can conduct sensitivity experiments to test whether the continental configuration is important to the results – in this case, we might have had a second alternative configuration with a super continent positioned over a pole. And/or we could have had a set of very fragmented continents. For instance, *Hoffman and Schrag* [2002] discuss the potential importance of continental configuration, while the entire hypothesis of *Donnadieu et al.* [2004] rests on specific details of the continental configuration being realistic.

Other questions to think about with regards to numerical modeling (and this experiment) are:

- *(Is the model configuration and experimental design 'realistic' ... ?)*
- *What is 'missing' in the model and what might the implications for your predictions and conclusions be?*
- *Are the simulations being run for sufficiently long? Why not if not (i.e., justify your choices of parameter values and experimental assumptions)? How might the results and conclusions be biased (if at all)?*
- *How would you test model predictions and your overall conclusions?*
- *How could the experimental design be improved?*

8.2 To search for the atmospheric CO_2 concentration (or rather, radiative forcing equivalent) that would lead to a 'snowball Earth' state in the Neoproterozoic ('**How low does CO_2 have to be to trigger a 'snowba**ll'?'), you are going to edit the file that controls the specific details of the experiment. This is the *user-config* file. From the `genie-userconfigs/LABS` directory, open one of the snowball experiments (e.g. `exp5_snowball`) in the SciTE text editor. At the top you should see:

```
#
#
# --- CLIMATE -------------------------------------------------
#
# solar constant reduced by 5.6% for late Neoproterozoic: 1285.92 W m-2
ma_genie_solar_constant=1285.92
# set no seasonal cycle
ea_dosc=.false.
```

```
go_dosc=.false.
gs_dosc=.false.
# scaling for atmospheric CO2 radiative forcing, relative to 278 ppm
ea_radfor_scl_co2=20.0
```

Each line that is not commented out (i.e., no #) contains a parameter of the format:

PARAMETER=VALUE

The value of each parameter can be edited to form a new experiment. (Additional parameter value specifications can also be added, or existing ones deleted.)

For instance, the line:

`ma_genie_solar_constant= 1285.92`

specifies a reduction in the solar constant of about 6% (from the modern value of 1368 W m^{-2}), appropriate for the late Neoproterozoic, while the line:

`ea_radfor_scl_co2=20.0`

specifies a radiative forcing of climate by CO_2 equivalent to ×20 modern (20×278 = 5560 ppm).

Note that CO_2 is not being explicitly modeled in this experiment, but the long-wave radiative forcing associated with a specified concentration of CO_2 (as a ratio to modern concentrations) is being set instead.

Edit the value of `ea_radfor_scl_co2` (lower or higher) and save the file. Re-run the experiment to see whether sea-ice extent is approaching a new steady state. You may want to try even longer simulations if it becomes clear that the model is still far from steady-state. You can judge how close to equilibrium things have got by following (and/or plotting) the evolution of e.g., global surface air temperature or sea-ice extent (both time-series files). Note that you *might* want to run the experiment longer than 100 years …

HINT: You can make copies (`cp` command) of the *user config* file (e.g. `exp5_snowball`) with different names differently, e.g., `exp5a`, `exp5b`, `exp5c`, …) and in each copy, specify a different radiative forcing assumption. By submitting the experiments to the cluster will allow you to run all these experiments simultaneously.

8.3   For each experiment you want to be assessing how far towards the Equator the sea-ice limit encroaches through some of the time-series and time-slice files or even the on-screen summary lines (assuming running interactively rather than via a job submission to the cluster queue). Informative time-series variables include (but not necessarily be limited to: atmospheric temperature and sea-ice cover. (Sea-ice thickness, on account of the simple physics in the model, low resolution and long time-step, can be a little bonkers at times, but sea-ice volume should be more trustworthy.)

DON'T PANIC! if the model 'crashes' – once it is in a 'snowball' state, the sea-ice steadily builds up and gets thicker and thicker. Because of the low resolution and relatively long time-step used here (chosen to accelerate the exercise), at some point it becomes impossible to adequately solve e.g. heat diffusion through the sea-ice and the model crashes. (It is at least an indication that a snowball state has been achieved.)

For the time-slice data: atmospheric and ocean surface temperature and sea-ice extent (2-D biogem NetCDF file) may be informative.

HINT: Be careful with the default 'auto-scaling' feature in Panoply. At near complete sea-ice cover, you may find Panoply scaling min and max sea-ice between 99.1 and 99.9% or something – it always tries to maximize color contrasts by default and will be at all helpful in this case.

In answering the question think about what an appropriate degree of accuracy might be for your experiments. Just because computer models generally calculate to around 16 significant places of precision, does not mean you have 16 significant figures of accuracy (or realism). For instance – how many significant figures is the solar constant quoted to and what do you think is the uncertainty in this? Harder to judge is how the assumed (incorrect) continental configuration creates additional uncertainty, or the physics assumed in the ocean or sea-ice, or lack of snow on land …

Appendix 1 – linux 101

Viewing directories; moving around the file system

When logging in, you always start from your 'home' directory. This is represented by a '~' before the command prompt (`$`).

At the command line prompt (`$`) in linux, you can view the current directory contents:

`$ ls`

or for a more complete output:

`$ ls -la`

To go down a directory (e.g. `cgenie_output`) relative to where you already are:

`$ cd cgenie_output`

and to go back up one is:

`$ cd ..`

It is often safer/easier at first, if you need to change more than one directory level to do this in stages. e.g. to change to `cgenie_output/exp0_modern_SPINUP`, change to `cgenie_output` (`cd cgenie_output`) but then check that you are in the place you think you are and/or remind yourself of the spelling of the next directory you need to change to by typing `ls`. Then move on, e.g. `cd exp0_modern_SPINUP`. Note that to go straight from your home directory in one jump, type: `cd cgenie_output/exp0_modern_SPINUP`. If you start from somewhere other than your home directory, give the full path: `cd $HOME/cgenie_output/exp0_modern_SPINUP`.

You can always return to your home directory (`~`) by typing:

`$ cd`

or:

`$ cd $HOME`

(or `cd ~`)

Copying and moving files

To copy a file `myconfig` to `myconfig_new`, assuming you are in the same directory where both the old file is and the new file will be:

`$ cp myconfig myconfig_new`

To move myconfig to the cGENIE user-config directory, assuming you are in the directory where the old file is but with the new file in a different directory, give the full path of the new directory:

`$ mv myconfig ~/cgenie/genie-userconfigs/LABS/myconfig`

To rename `myconfig` to `useless_config`:

`$ mv myconfig useless_config`

Creating directories

To create a directory mydirectory:

`$ mkdir mydirectory`

Repeating command lines

You do not have to re-enter lines of commands and options in their entirety each time – by pressing the UP cursor key you get the last command you issued. If you keep pressing the UP cursor key you can recover progressively older commands you have previously entered. When you have recovered a helpful line you can simply just edit it, navigating along with the LEFT and RIGHT cursor keys (press RETURN when you are done).