Worksheet 1: File locations and pre-processing

The following exercises demonstrate some of the tools available for data analysis, and how to prepare PRECIS output for analysis. This can be time consuming for large amounts of data, so in this worksheet a small subset is used to demonstrate the steps involved. In the worksheets that follow, data that has already been processed will be used.

PRECIS output data tables are in PP format, a Met Office binary data format. This worksheet converts data to NetCDF format (a standard format in climate science) in order that it can be used in post processing packages such as CDO.

Contents

Note: *Please ignore the ⅋ sign when typing the commands.*

1.1 Data locations and file names

Identify and list the names of PRECIS output data in PP format using standard Linux commands.

The dataset used here is a three year subset of monthly PRECIS data over south east Asia driven by the HadCMQ0 GCM.

```
% DATA=/media/sf_kl2/
```

1. b) Move to the directory (i.e. folder) called `$DATA/practise_ppfiles/cahpa`. This directory contains **monthly** output data from experiment with RUNID `cahpa`.

`cd` stands for 'change directory' and `pwd` stands for 'print working directory'. If you are not sure where in your directory tree you are, `pwd` will tell you.

```
% cd $DATA/practise_ppfiles/cahpa
% pwd
```

1. c) List the contents of this directory; `ls` stands for 'list' and using the `-l` option gives a 'longer' listing with more information, such as file and size and modification date.

```
% ls
% ls -l
```

2. a) List all the files containing data for September.

```
% ls *sep*
% ls
```

How many files contain data for September?

_____

Note: *The character* `*` *translates as 'any characters'*

2. b) List all the files containing data from 1982 (i.e. all files which begin `cahpaa.pmi2`.)

```
% ls cahpaa.pmi2???.pp
```

Note: *The character* `?` *translates as 'any single character'*

3.) Move up two levels in the directory tree and list the directories.

```
% cd ..
% cd ..
% pwd
% ls
```

The directories **daily** and **monthly** contain data used in the worksheets which follow this one.

1.2 Visualization

`xconv -i` is a utility which allows PP and NetCDF fields to be visualized very quickly. It is very easy to use.

1.) View the contents of the monthly PRECIS file for July 1983 (`i3jul`.) What variables does it contain? What are the latitude and longitude dimensions (in number of grid boxes)?

```
% cd $DATA/practise_ppfiles/cahpa
% xconv -i cahpaa.pmi3jul.pp # Monthly file
```

Variables:

_____

_____

Latitude and longitude dimensions:

_____

_____

Note: *Anything after the character # is just a comment and does not affect the command being run*

1.3 Remove the rim from PP fields

The edges (or rim) of RCM outputs are biased due to the linear relaxation used on certain variables to apply the GCM lateral boundary conditions. This rim of 8 grid points from each edge needs to be excluded from any analysis.

1a.) Remove the 8-point rim from all data in the `practise_ppfiles/cahpa` directory (with the option of automatically deleting the original full-sized files).

> **`pprr` removes a rim from PP fields. A number of grid boxes will be removed from each edge and different edges may have different numbers of grid boxes removed. Type `pprr -helpx` for its full functionality.**

Note: *The `-X` option to `pprr` allows files with particular names to be ignored.*

```
% cd $DATA/practise_ppfiles
% pprr -r 8 -X ".rr8.pp" cahpa
```

1b.) What are the latitude and longitude dimensions (in number of grid boxes) of July 1983 now?

```
% ls cahpa/*
% xconv -i cahpa/cahpaa.pmi3jul.rr8.pp
```

Latitude and longitude dimensions:

_____

Note: *In all further exercises, it is assumed that the 8-point rim has been removed from all of the data being used and that the original, full-sized fields are no longer present.*

1.4 Select variables and convert PP files to NetCDF

The monthly data we are using has multiple variables in each file, we can use a pptool to separate the variables. The rest of the worksheets use CDO, a post processing package which needs files to be in NetCDF format. This means it is necessary for us to convert our PP files into NetCDF format using a pptool.

1a.) Separate the variables in the all of the monthly files into separate directories.

`ppss` splits the PP fields in input PP files into subdirectories based on each field's STASH code and processing code. Type `ppfile -helpx` for its full functionality.

```
% cd $DATA/practise_ppfiles/cahpa
% ppss cahpaa.pm?????.rr8.pp
```

1b.) Change into the directory containing temperature at 1.5m files and view the December 1981 file.

```
% stash 03236
% ls
% cd 03236
% xconv -i cahpaa.pmi1dec.rr8.03236.pp
```

Note: *03236 is the STASH code of temperature at 1.5m.*

2.) Put the monthly temperature files into a single file. This process is necessary for later CDO commands to function properly.

```
% cat cahpaa.pm?????.rr8.03236.pp > cahpaa.pm.1981_1983.rr8.03236.pp
% ls -lrt
```

Note: *the* `cat` *command concatenates multiple files together.*

3.) Convert the 1981-1983 temperature file to NetCDF format so that we may carry out analysis with CDO.

`pp2cf` converts a PP file to NetCDF format (a standard format in climate science,) see http://www.met.reading.ac.uk/~david/pp2cf.html for further details.

```
% pp2cf cahpaa.pm.1981_1983.rr8.03236.pp
% xconv -i cahpaa.pm.1981_1983.rr8.03236.nc
```