

ClimGen output file formats, v1-02

ClimGen produces text output files. The format of those stored in `~/out/main/` always follows the same basic structure, regardless of whether it contains observations or scenarios, monthly, seasonal or temporally-averaged data, grid box patterns or spatially-averaged data.

There are always a number of header lines at the beginning of the file, the ‘information block’. This gives information about the data set as a whole, and describes the syntactic and semantic format. Each data block which follows is preceded by a single ‘sub-header line’, with information specific to that block (e.g., the co-ordinates of a particular grid cell).

Information block

The information block contains 13 lines, plus one line for each ‘season’ (or month or annual average, depending on the data frequency within each year) giving specific information about that season.

Information block lines 1 to 3: Description of ClimGen, including version and date/time.

Information block line 4: Brief indication of data sets used that are not already indicated in the file name.

Information block line 5: Brief indication of how the file was constructed (see the ClimGen user guide and scientific description for further details).

Information block line 6: The variable short name, full name and units.

Information block line 7: Usually blank.

Information block line 8: Description of the spatial grid (domain and dimensions) on which the data set was constructed. Note that this is not necessarily the grid of this output file data, since spatial averaging may have been carried out by ClimGen.

Information block line 9: Contains (i) *Nregions*, the number of time series included in the output file (e.g., the number of grid boxes, or the number of regional averages); (ii) *Ntimes*, the number of time periods in the output file (e.g., 30 for a time series of 30 yearly values, or 1 for a single 30-year mean); (iii) the *Multiplier* that should be applied to each output value to obtain the units given on line 6; and (iv) the *Missing Code*. Note that the data values must be compared with the missing code before the multiplier is applied.

Information block line 10: A description of the subheader contents and the Fortran-style format used for the actual data blocks.

Information block line 11: Usually blank.

Information block lines 12 to 12+*Nseasons*: A definition of the months/seasons/annual-means included in the output file. Each line in this header block defines the output for each of the columns in the output data. A T (=true) under a month indicates that this month is included in the output for this column, while an F (=false) indicates that this month is not included. If the output contains individual months only, then each line in this header block will have only one T and eleven Fs, with the T indicating which month corresponds to this column of data. If there are Ts only under, say, Jun, Jul and Aug, then a June–July–August seasonal average value (or total value for precipitation and wetdays) will be contained in the output file. If there are Ts under all months, then an annual mean will be contained in the output file for this column. The final value in each line in this header block (labelled BEG) indicates which month the seasonal/annual-mean was started from. If there are Ts under all months and BEG=1, it will be a January–December annual mean, while if BEG=10, it will be an October–September annual mean. For a boreal winter seasonal mean, there will be Ts under Jan, Feb and Dec, and BEG will be set to 12. Note that in this latter case (and similar cases), the final time period in the output file will be set to the missing code for this column, because this is an average of the December value from within the selected period and the January and February values in the next year (outside the selected period, and thus unavailable). Any seasonal average that spans the December–January boundary will be recorded under the year in which the seasonal average began (i.e., a boreal winter seasonal mean for 2040 will comprise December 2040 and January–February 2041).

Information block line 13+*Nseasons*: Usually blank.

Data blocks

Data values for each regional average or each grid box are given in individual data blocks; the number of data blocks in the file is given by *Nregions* (see ‘Information block line 9’ above). Each data block contains one subheader line and *Ntimes* lines of data.

The subheader line contains (i) an index number for each region/grid box; the row (for north/south) and column (for east/west) of the grid box included in this region that is farthest (ii) north, (iii) east, (iv) south and (v) west, respectively; the (vi) latitude and (vi) longitude of the centre of the box/region; and (vii) a name for the box/region. Latitudes north of the equator are positive, south of the equator are negative. Longitudes west of the Greenwich meridian are negative, east of the meridian are positive.

If the output file contains individual grid boxes (e.g., if it contains all land grid boxes across the globe) rather than regional averages, then the grid box that is farthest north, east, south and west in each “region” will of course all be exactly the same, single grid box. In this case, the (ii) north and (iv) south values will be identical, as will the (iii) east and (v) west values. The latitude and longitude will be the exact centre of the grid box. Because it is not possible to give a proper name to each grid box across the globe, the “name” will usually be the grid box column and row.

If the output file contains regional averages (e.g., country means), then the subheader lines indicate the overall span of data locations and an approximate centre for each region, together with a meaningful name. However, the full definition of which grid boxes contribute to each regional mean is not given in the CLIMGEN output file.

Each data block contains *Ntimes* lines of data. Each data line begins with the start and end year of the time period whose data is contained on this line. If the file contains time series with individual values for every year, the start and end year on a particular data line will be equal, because that line contains data for that single year. If the file contains 30-year averages, the start and end year of the 30-year period will be given on each data line). The remainder of each data line will contain *Nseasons* actual values (one per “season”, where each season is defined in the ‘Information block’ and may be a single month, an average of some months, or an annual average of all months).

Example output file with some annotation

An output file with individual grid box time series from the 0.5° latitude/longitude resolution grid.

Nregions (in this case 67420 is the number of grid boxes containing some land, excluding Antarctica, across the whole world)

```

ClimGen v1.42 data file created on 09.05.08 at 11:47:13
ClimGen developed by Dr Tim Osborn (Climatic Research Unit) and Dr Tim Mitchell (Tyndall Centre), ENV, UEA, UK
http://www.cru.uea.ac.uk/~timo/climgen/
Data used: climate change & observed variability cru_ts_3_00 1961-1990 starting 1961
Pattern scaling: obs mean + obs variability + mean change
.tmp = near-surface temperature (degrees Celsius)

[Long=-180.00, 180.00] [Lati= -90.00, 90.00] [Grid X,Y= 720, 360]
[Regis= 67420] [Periods= 30] [Multi= 1.00000] [Missing=-999]
[Regions=INDEX,N,E,S,W,Lat,Lon,Name] [Format='(2i5,12f6.1)']

COL Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec BEG
1 T F F F F F F F F F F F 1
2 F T F F F F F F F F F F 1
3 F F T F F F F F F F F F 1
4 F F F T F F F F F F F F 1
5 F F F F T F F F F F F F 1
6 F F F F F T F F F F F F 1
7 F F F F F F T F F F F F 1
8 F F F F F F F T F F F F 1
9 F F F F F F F F T F F F 1
10 F F F F F F F F F T F F 1
11 F F F F F F F F F F T F 1
12 F F F F F F F F F F F T 1

1 148 1 148 1 -16.25 -179.75 1, 148
2040 2040 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2041 2041 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2042 2042 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2043 2043 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2044 2044 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2045 2045 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2046 2046 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2047 2047 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2048 2048 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2049 2049 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2050 2050 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9
2051 2051 29.1 29.5 29.3 28.8 28.3 27.7 26.9 26.5 26.9 27.3 28.1 28.9

```

The T under Apr indicates that column 4 contains April values

Latitude and longitude for the first block of data

“Name” for this grid box / region

Row contains data averaged over the period 2047-2047 (i.e., no time averaging in this case, just values for that single year)

Column and row of the grid box for the first block of data, where row runs from 1 to 720, and column runs from 1 to 360, with the western edge of column 1 starting at 180°W (and hence its centre will be at 179.75°W) and the southern edge of row 1 starting at 90°S (and hence its centre will be at 89.75°S)

Time series of data near-surface temperatures for April for 16.25°S, 179.75°W