

MarkSim_standalone for DSSAT users

P. G. Jones March 2012, Waen Associates, Gwynedd.

Introduction

Marksim_standalone.exe combines the basic routines from the original MarkSim version, produced at CIAT (International Center for Tropical Agriculture) and distributed on CD (Jones et.al., 2002), and the new MarkSimGCM web application (Jones et al. 2011a, b) that runs off Google Earth. The standalone version is designed for computer users that need to process a large amount of data. It eliminates their picking at a keyboard and abusing their eyesight by searching an on-screen map for the required data point.

It uses the same six GCM model results as MarkSimGCM and the same three scenarios, along with an ensemble of the six.

It is designed to be used with DSSAT <http://dssat.net/> and so it uses CLI files (see description below), existing, or constructed by the user, to create new CLI files and WTG files for the simulated weather data under a range of GCMs and scenarios.

The EXE runs from the DOS prompt (or a system program call). It will read a directory, work out a list of the CLI files in that directory, and create subdirectories for each CLI file. In each of these it will place up to 99 replicates of yearly weather data and a new CLI file describing the climate under the particular situation of GCM/scenario/year for which the user called.

MarkSim

MarkSim was developed in the 1980s and 1990s to simulate weather from known sources of monthly climate data from around the world (Jones & Thornton 1993, 1997, 1999, 2000). It divides the world into 720 clusters of climate that are distinct from one another and fits a third order Markov model to the precipitation data. The temperature data simulation is derived from SIMMETEO (Geng et al., 1988). The radiation data are based on the model of Donatelli and Campbell (1997).

For each of the about 9,200 stations with usable daily data, the third order Markov chain model of rainfall was developed. The results were grouped according to climate cluster, and regression equations for each of the Markov parameters were calculated using the monthly average rainfall and temperature figures for each station within the cluster. The model can now be fitted to any monthly climate data record by determining to which cluster it belongs, and using the relevant regression equations.

Markov models are excellent simulations of the rainfall process, but they do have limitations. It was found quite early in the analysis (Jones & Thornton, 1993) that a simple first or second order process, while adequate for temperate regions, would not fit well enough for the tropics. Hence we went with the third order model. Another deficiency we found was that Markov processes ordinarily underestimate the variance of rainfall. This is solved in MarkSim by resampling the Markov probability coefficients for each year of simulation. This is because the coefficients themselves are only estimates and have an error term. Once we reincorporate this error term by resampling, the rainfall variance agrees with the observed.

GCM interpolation and regression

Table 1 shows the models used in the application. We obtained data from the GCMs for either nine or five time slices: for average monthly precipitation and daily maximum (tmax) and minimum (tmin) air temperatures. These time slice data had been previously interpolated to 0.5 degrees from the original model resolutions. We used the deviations from

the climate standard 1960-1990 period and defined the origin for the regressions as 1975, the mid point of this period. Fourth order polynomial regressions were fitted for deviations of precipitation, tmax and tmin for every pixel in the 720*180 image. Standard deviations and R^2 of the regression were mapped and the images inspected visually for unusual deviations. The regression coefficients were combined into a binary data structure for each GCM, scenario, and variate.

The interpolation procedure is an inverse square distance convolute algorithm working from the nine points closest to the desired coordinates. This is not perfect, but is highly stable and much faster than a thin-plate smoothing algorithm. In practice, faint jumps in the contours can be seen as the convolute moves from one calibration set to another at the 0.25 and 0.75 degree boundaries. In fact, the systematic errors in the original interpolation from the GCM native resolutions far outweigh these effects.

Table 1. Atmosphere-Ocean General Circulation Models used in the work (details from Randall et al., 2007).

Model Name	Institution	Reference	Resolution (degrees)	Time slices in regression (years)
BCCR-BCM2.0	Bjerknes Centre for Climate Research (University of Bergen, Norway)	Furevik et al. (2003)	2.8 x 2.8	9 x 10
CNRM-CM3	Météo-France/Centre National de Recherches Météorologiques, France	Déqué et al. (1994)	1.9 x 1.9	9 x 10
CSIRO-Mk3.5	Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atmospheric Research, Australia	Gordon et al. (2002)	1.9 x 1.9	9 x 10
ECHam5	Max Planck Institute for Meteorology, Germany	Roeckner et al. (2003)	1.9 x 1.9	5 x 20
INM-CM3.0	Institute for Numerical Mathematics, Moscow, Russia.	Diansky & Volodin (2002)	5.0 x 4.0	9 x 10
MIROC3.2 (medres)	Center for Climate System Research (CCSR), National Institute for Environmental Studies (NIES), and Frontier Research Center for Global Change (FRCGC), University of Tokyo, Japan.	K-1 Model Developers (2004)	2.8 x 2.8	5 x 20

Data

21 binary files (.bin) contain 2.91 Gb of data for the GCM regression interpolation. These should be downloaded, unzipped, and placed into a directory accessible from where you intend to run the EXE. Also included in the zip are the MarkSim cluster files. Place them in the same directory as the binary files. The cluster data files are avmats.cor, betas.out, clusstat.out, dlag1.out, dlag2.out, dlag3.out, ps.out, and sds.out.

Running the EXE

MarkSim_standalone.exe can be run from any directory under the DOS prompt or by system call. It requires six arguments with an optional seventh; the call from DOS is:

MarkSim_standalone path1 path2 GCM scenario year nreps (seed)

The arguments are:

Path1: The path to the directory in which you have placed the MarkSimGCM data file (see Data above).

Path2: The path to the directory in which the CLI files reside. Note that this directory will be expanded by a subdirectory for every CLI file and for every use of MarkSim_standalone with different GCM, scenario, and year arguments.

GCM: The GCM that you wish to use. You can choose from six or the composite mean (ensemble) of them all. The codes are:

Code	Model
BCC	BCCR-BCM2.0
CNR	CNRM-CM3
CSI	CSIRO-Mk3.5
ECH	ECHam5
INM	INM-CM3.0
MIR	MIROC3.2 (medres)
AVR	Ensemble of six above

Scenario: Three scenarios are available: the low, mid, and high range of emissions from the IPCC 4th assessment. The available codes are **A2**, 'high' emissions; **A1**, 'medium' emissions – normally written A1b, but these codes are only two characters; and **B1**, 'low' emissions.

Year: Is the year you wish to simulate. This can take the values 2012 to 2095, but if you wish to simulate 'present' data as represented by your CLI file without GCM changes then use the year -1. No others will be accepted. Note that the 'new' CLI file for 'present' values will contain the original rainfall tmax and tmin, but radiation and rain days will be as calculated by MarkSim.

Nreps: Is the number of years to be simulated. NOTE these are replicates of the year that you specify and are NOT a series of years from that date!

Seed: You do not need to specify a seed; this is an optional argument. If you are trying to replicate trials at a number of places or times it might be a good idea to do so. MarkSim

does not yet guarantee complete sequencing of random numbers (that must wait for MarkSim_v2), but the sequencing is quite good. If you do specify a seed then use a four to nine digit odd integer.

CLI files and naming conventions

DSSAT convention is for four character CLI names, thus the program cannot add the codes for the model, scenario, or simulation date. Therefore it creates a new subdirectory with the name of the 'CLI file//GCM code//scenario code//year' and writes the modified CLI file therein under its own name along with the required number of years of simulated data. Since users may be using a CLI name greater than four characters outside of DSSAT, a file name length of 20 characters is permitted.

The simulated data files are named cccnn01.WTG where cccc are the first four characters of the CLI name and nn the year number 01-99 to conform to DDSAT naming. *This means that all files in a run with multiple CLI files carry the same names and are differentiated only by the directory in which they reside.* Note that in MarkSim output these do not represent a series of years, but are replicates of the target year.

MarkSim_standalone will search a directory for all the CLI files it contains. Other files in the directory are ignored. A maximum of 32,000 CLI files is allowed in each directory. If more than these are needed, then we suggest that you place them in multiple directories and call the EXE for each directory.

On a reasonable PC, the EXE takes about 4 seconds to load the GCM regression data and exceeds the simulation run time by more than 10^6 . This happens once for each directory call, so it is advantageous to place as many CLI files together as possible to minimize loading time. *Placing a single CLI file in each directory and calling the EXE with a BAT that produces one simulation run per directory is therefore a slow option.*

Error reporting

The EXE is designed to run in the DOS prompt and will hard stop and write error messages there. Many of these should never be seen by the user, they are listed for completeness. Text of the message listed here is **red**; variables within the message are **blue**, inserted text is **green**.

1. Error. '*path*' is not a valid path

This may be the path to the data, to the CLI files, or it may be the path to the newly constructed directory to hold the results. The error can result if you do not put the drive in it.

2. Error. Unable to list directory, error code *ret*

ret is the return code from a SYSTEM call in Visual Fortran and has the following values:

1. The argument list is too long.
2. The command interpreter cannot be found.
3. The command interpreter file has an invalid format and is not executable.
4. Not enough system resources are available to execute the command.

The user should never see the first of these, but any of the others might occur if your machine configuration is not standard.

3. Error. Unable to open directory listing.

Probably due to errors 1 or 2 and so the user should not see this one.

4. Error. Unable to delete list file, error code, *ret*

as error 2

5. Error. Incorrect arguments

obvious

6. Error. Path to GCM data longer than 80 chars

"

7. Error. Path to CLI directory longer than 80 chars

"

8. Error. GCM code longer than 3 chars

"

9. Error. Scenario code longer than 2 chars

"

10. Error. Year for simulation out of range

"

11. Error. Number of replicates out of range (1 to 99 allowed)

"

'field'

Field length exceeded

11a. Error. Number of replicates out of range (1 to 99 allowed)

n

number out of range

12. Error. Too many digits for seed

Limit is 9 digits

13. Error. Seed out of range

If it is specified, seed must be an odd integer greater than 0, but works best with four or more digits.

14. Error. Unable to open *'filename'*

User should not see this unless a major corruption of the CLI file directory has occurred.

15. Error. Unable to read *'filename'*

This is a general message denoting that something (unknown) is wrong with the indicated file. It may be empty, or corrupted in a way that the program cannot discern.

16. Error. Cannot construct *clx* with an unrotated climate record - *construct_clx*

The user should never see this.

17. Error. Radiation out of range in *camdon*, cannot be negative – *camdon*

This can occasionally happen, but is highly unlikely. In fact, negative net radiation is possible and does happen in polar regions, but this is more likely to be a pathological effect in the Campbell Donatelli model.

18. Error. Non-convergence in *root* – *avs*

This is theoretically possible, but very rare. This indicates a failure in the recalculation of the *avs* parameters. In fact, the routine to calculate the *root* is very robust.

19. Error. *Av* unusually high

Never yet seen, but could happen.

20. Error. Data file not available. Make sure path is correct

The path to the GCM and MarkSim data is not correct and the relevant files could not be opened.

21. Error. Reading GCM data, file integrity

An error occurred reading the GCM data files. Try downloading again.

22. Error. No GCM points within range

This can happen in certain cases where a small island is missing from the file. It is rare.

23. Error. Maths error in dcpmat or probit.

This should be very rare. If it occurs please report it to p.jones@cgiar.org with the input file that produced it.

24. Error. Writing WTG file.

It is difficult to see how this could happen, but it might if creation of the subdirectory had failed.

25. Error. Failure in reading 'filename' 'record'

Could not read the designated CLI file. The offending record is shown after the error message.

26. Error. Unable to open new CLI file 'filename'

This is probably an error in creating the directory, and should have been caught earlier

27. Error. Unable to write new CLI file.

As above

28. Error, Too many CLI files in directory 'path' (limit 32000)

DSSAT CLI file description for MarkSim_standalone

The CLI file required for MarkSim_standalone is a reduced version of the standard DSSAT CLI file. It does not contain either the WGEN parameters, the RANGE CHECK VALUES, or the FLAGGED DATA COUNT. It therefore stops after the listing of MONTHLY AVERAGES.

MarkSim ignores many fields, but some are transferred to the new CLI file constructed with the future climate.

```
*CLIMATE : PALMIRA-CIAT
@ INSI      LAT      LONG    ELEV    TAV    AMP    SRAY    TMXY    TMNY    RAIY
   PALM      3.500   -76.300  965    24.1   11.6   199.4   29.4   18.8   1147
```

In this line only the INSI, LAT, LONG and ELEV fields are actually used. All the others are recalculated from the new climate. If necessary they can take the value -99.0 or -99 and they will be replaced in the new file.

```
@START  DURN  ANGA  ANGB  REFHT  WNDHT  SOURCE
        0    0  0.25  0.50  0.00  0.00  Calculated_from_daily_data
@ GSST  GSDU
        1   365
```

None of the values in these two lines are used, however they are transferred to the new file. The values shown here are the standard default values and should always be set to these if actual values are not available. The exception is the source field. This will be replaced with the relevant GCM model, scenario and simulation year information.

```
*MONTHLY AVERAGES
@MONTH  SAMN  XAMN  NAMN  RTOT  RNUM  SHMN
        1  18.0  29.7  18.7  81.0  9.1  -99.0
        2  18.0  29.8  18.8  75.0  7.8  -99.0
        .  .  .
        12 17.4  29.1  18.6 102.0 10.6 -99.0
```

SAMN and RNUM are recalculated for the new file and can take the missing value

-99.0 if desired. SHMN is not used and will be transferred to the new file. It is, however, read and needs to be present, even if only missing value.

Selected weather data codes relevant to CLI as above

AMP	Temperature amplitude, monthly averages, C
AMTH	Angstrom 'a' coefficient, monthly, unitless
ANGA	Angstrom 'a' coefficient, yearly, unitless
ANGB	Angstrom 'b' coefficient, yearly, unitless
BMTH	Angstrom 'b' coefficient, monthly, unitless
DURN	Duration of summarization period for climate files, Yr
ELEV	Elevation, m
GSDU	Growing season duration, Day
GSST	Growing season start day, Doy
INSI	Institute and site code
LAT	Latitude, degrees (decimals)
LONG	Longitude, degrees (decimals)
MTH	Month, #
NAMN	Temperature minimum, all days, monthly average, C
RAIY	Rainfall, yearly total, mm
REFHT	Reference height for weather measurements, m
RNUM	Rainy days, # month-1
RTOT	Rainfall total, mm month-1
SAMN	Solar radiation, all days, monthly average, MJ m ⁻² d ⁻¹
SHMN	Daily sunshine duration, monthly average, percent
SOURCE	Source of daily weather data, text
SRAY	Solar radiation, yearly average, MJ m ⁻² day ⁻¹
START	Start of summary period for climate (CLI) files, Year
TAV	Temperature average for whole year, C
TMNY	Temperature minimum, yearly average, C
TMXY	Temperature maximum, yearly average,
WIND	Daily wind speed (km d ⁻¹)
WNDHT	Reference height for windspeed measurements, m
XAMN	Temperature maximum, all days, monthly average, C

References

- Déqué, M., Dreveton, C., Braun, A., Cariolle, D. 1994. The ARPEGE/IFS atmosphere model: A contribution to the French community climate modeling. *Clim. Dyn.* 10: 249-266.
- Diansky, N.A. and Volodin, Ye. M. 2002. The reproduction of modern climate by model of atmosphere-ocean general circulation model (in Russian). *Izv. Atmos. Oceanic Phys.* 38: 824-840.
- Donatelli, M. and Campbell, G.S. 1997. A simple model to estimate global solar radiation. PANDA Project, Subproject 1, Series 1, Paper 26. Research Institute for Industrial Crops (ISCI), Bologna, Italy. 3 pp.
- Furevik, T., Bentsen, M., Drange, H., Kindem, I.K.T., Kvamstø, N.G., Sorteberg, A. 2003. Description and evaluation of the Bergen climate model: ARPEGE coupled with MICOM. *Clim. Dyn.* 21: 27-51.
- Geng, S., Auburn, J., Brandstetter, E., Li, B. 1988. A program to simulate meteorological variables. Documentation for SIMMETEO. (Agronomy Report No. 204). University of California, Davis Crop Extension, Davis, California.
- Gordon, H.B., Rotstayn, L.D., McGregor, J.L., Dix, M.R., Kowalczyk, E.A., O'Farrell, S.P., Waterman, L.J., Hirst, A.C., Wilson, S.G., Collier, M.A., Watterson, I.G., Elliot, T.I. 2002. The CSIRO Mk3 Climate System Model. Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atmospheric Research Technical Paper No. 60. CSIRO, Aspendale, Victoria, Australia. 130 pp.
http://www.cmar.csiro.au/e-print/open/gordon_2002a.pdf (accessed December 2009).
- Jones, P.G. and Thornton, P.K. 1993. A rainfall generator for agricultural applications in the tropics. *Agric. For. Meteorol.* 63: 1-19.
- Jones, P.G. and Thornton, P.K. 1997. Spatial and temporal variability of rainfall related to a third order Markov model. *Agric. For. Meteorol.* 86: 127-138.
- Jones, P.G. and Thornton, P.K. 1999. Fitting a third order Markov rainfall model to interpolated climatic surfaces. *Agric. For. Meteorol.* 97: 213-231.

- Jones, P.G. and Thornton, P.K. 2000. MarkSim: Software to generate daily weather data for Latin America and Africa. *Agron. J.* 93: 445-453.
- Jones P.G., Thornton, P.K., Diaz, W., Wilkins, P.W. 2002. MarkSim. A computer tool that generates simulated weather data for crop modeling and risk assessment. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. Theory section available at: <http://gisweb.ciat.cgiar.org/marksim/download/Theory.pdf>
- Jones, P.G., Thornton P.K., Heinke, J. 2011a. Generating characteristic daily weather data using downscaled climate model data from the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment. Available at: <https://hc.box.com/shared/f2gk053td8>
- Jones, P.G., Thornton, P.K., Giron, E. 2011b. Web application. MarkSimGCM, A weather simulator. Available at: <http://gismap.ciat.cgiar.org/MarkSimGCM/>
- K-1 Model Developers. 2004. K-1 Coupled Model (MIROC) Description. K-1 Technical Report No. 1, Hasumi, H., Emori, S. (Eds.) Center for Climate System Research(CCSR), National Institute for Environmental Studies (NIES), Frontier Research Center for Global Change (FRCGC), University of Tokyo, Tokyo, Japan. <http://www.ccsr.u-tokyo.ac.jp/kyosei/hasumi/MIROC/tech-repo.pdf> (accessed December 2009).
- Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichfet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E. 2007. Climate models and their evaluation. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, UK and New York, NY, USA.
- Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, S., Hagemann, S., Kirchner, I., Kornbluch, L., Mauzini, E., Rhodin, A., Schlese, U., Schulzweida, U., Tompkins, A. 2003. The Atmospheric General Circulation Model ECHAM5. Part I: Model Description. MPI Report 349, Max Planck Institute for Meteorology, Hamburg, Germany. 127 pp.