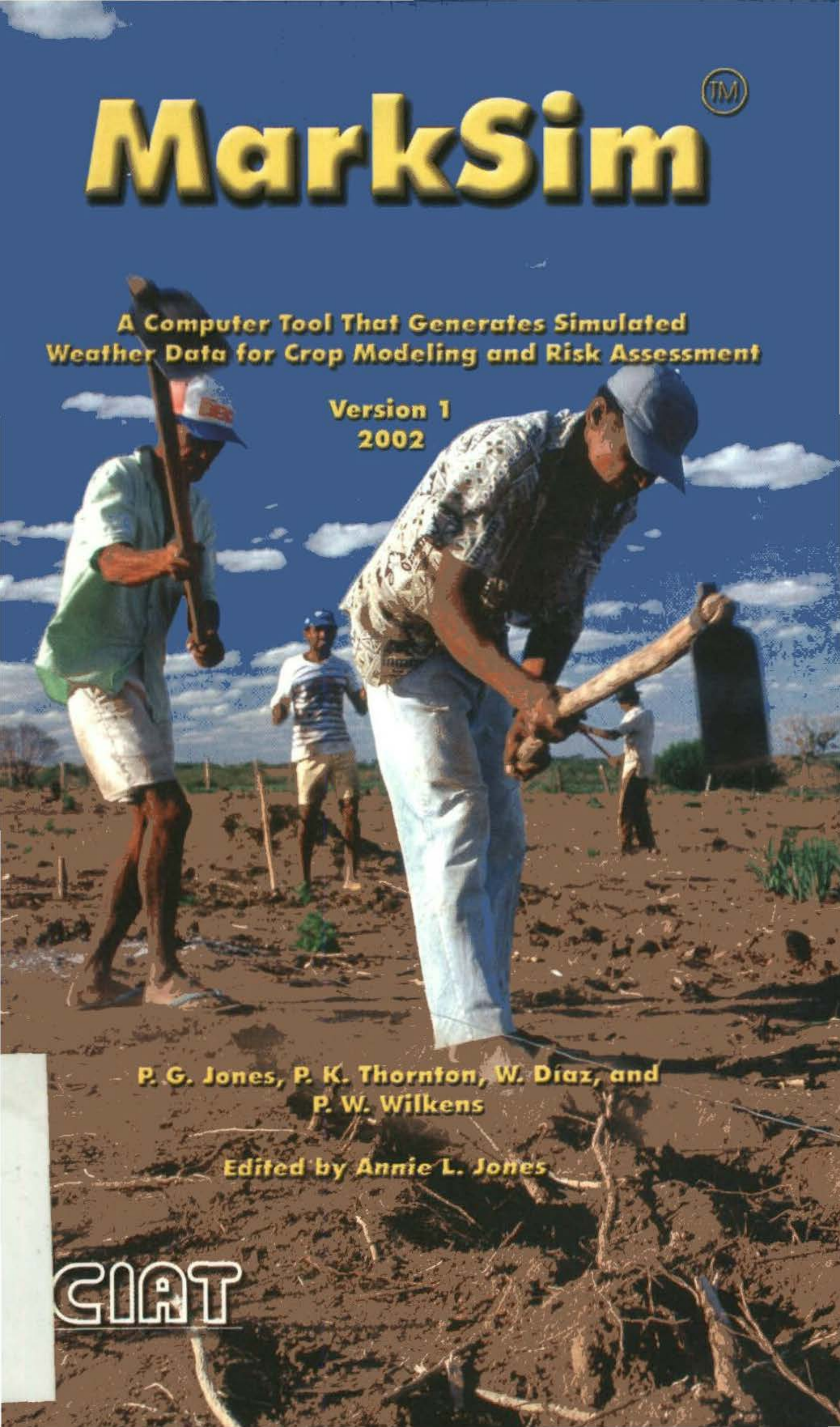


MarkSim™

A Computer Tool That Generates Simulated Weather Data for Crop Modeling and Risk Assessment

Version 1
2002

P. G. Jones, P. K. Thornton, W. Díaz, and P. W. Wilkens

Edited by Annie L. Jones

CIAT

The International Center for Tropical Agriculture (CIAT, its Spanish acronym) is one of 16 food and environmental research organizations known as the Future Harvest centers. The centers, located around the world, conduct research in partnership with farmers, scientists, and policymakers to help alleviate poverty and increase food security while protecting the natural resource base. The Future Harvest centers are principally funded through the 58 countries, private foundations, and regional and international organizations that make up the Consultative Group on International Agricultural Research (CGIAR).

In 2001, CIAT received support from the following donor countries: Australia, Belgium, Brazil, Canada, Colombia, Denmark, France, Germany, Iran, Italy, Japan, Mexico, the Netherlands, New Zealand, Norway, Peru, South Africa, Spain, Sweden, Switzerland, Thailand, the United Kingdom, and the United States of America.

Our work was also financed by the following organizations and foundations: the Asian Development Bank (ADB), the European Union (EU), the United Nations Food and Agriculture Organization (FAO), the Ford Foundation, Fundación Polar, the International Fund for Agricultural Development (IFAD), the Kellogg Foundation, the Nippon Foundation, the Rockefeller Foundation, the Wallace Foundation, and the World Bank.

CIAT also receives funds for research and development services provided under contract to a growing number of institutional clients.

ISBN 958-694-046-2

MarkSim™

A Computer Tool That Generates Simulated Weather Data for Crop Modeling and Risk Assessment

Version 1
2002

P. G. Jones, P. K. Thornton, W. Díaz, and P. W. Wilkens

Edited by Annie L. Jones

CD-ROM Series

Centro Internacional de Agricultura Tropical
International Center for Tropical Agriculture
Apartado Aéreo 6713
Cali, Colombia

Fax: +57 (2) 4450000
E-mail: p.jones@cgiar.org

CIAT Publication No. 330
ISBN 958-694-047-0 (CD-ROM)
ISBN 958-694-046-2 (manual)
Press run: 500
Printed in Colombia
September 2002

Jones, Peter G.
MarkSim : a computer tool that generates simulated weather data for crop modeling and risk assessment : version 1 / P.G. Jones, P.K. Thornton, W. Díaz, and P.W. Wilkens ; edited by Annie L. Jones. -- Cali, CO : Centro Internacional de Agricultura Tropical, 2002.
1 CD-ROM + Manual : MarkSim : a computer tool that generates simulated weather data for crop modeling and risk assessment : version 1. (87 p.) -- (CD-ROM series)
ISBN 958-694-047-0 (CD-ROM)
ISBN 958-694-046-2 (manual)

AGROVOC descriptors in Spanish:
1. Tiempo meteorológico. 2. Modelos de simulación. 3. Bases de datos. 4. Condiciones atmosféricas. 5. Clima.

AGROVOC descriptors in English:
1. Weather. 2. Simulation models. 3. Databases. 4. Meteorological elements. 5. Climate.

I. Title. II. Thornton, Philip K. III. Díaz, William. IV. Wilkens, Paul W. V. Jones, Annie L. VI. Centro Internacional de Agricultura Tropical. VII. Ser.

AGRIS subject categories: P40 Meteorology and climatology
U10 Mathematical and statistical methods

LC classification: S 600 .4 J66

Contents

Introduction

MarkSim has a long history. The rotation algorithm was written on the 6th of March 1978, not long after I had joined CIAT and started construction of the CIAT Climate Database. Markov models of rainfall have been used in many areas. A survey of the literature that I made in the mid 1980s came up with more than a hundred references. However, they have never been particularly successful in the tropics. I wondered why and eventually came to the conclusion that the weather systems prevalent in the tropics do not include the frontal weather with travelling highs and lows that you find at temperate latitudes. This means that the weather generating forces are completely different and need a different order of model to fit them. I eventually showed that this requires at least a third order model, where a first or second order would produce a good fit in temperate climes.

I pursued these investigations as a minor part of my studies in CIAT. One could almost say it was a hobby—until Phil Thornton noticed what I was doing in the early 1990s. He saw its application to crop modeling and pushed me to publishing the first paper, Jones and Thornton (1993). We have been strong collaborators ever since, producing a series of papers and working to craft MarkSim as a part of the CIAT Climate Database tools.

The MarkSim beta release, written for DOS operating systems, went to over 20 scientists in 1998. The response was good; indeed, Jeff White of the International Maize and Wheat Improvement Center (CIMMYT) used it to produce a rainfall reliability map for the whole of Africa. It has taken a disappointing number of years to go from there to this release for Windows. A lot of work has gone on in the meantime. The basic model has been revised. The database to which it is fitted has grown and been substantially cleaned. The station algorithm has been rewritten to incorporate difficult climates where rotation on rainfall pattern is not valid. We have incorporated new batch processing options that will greatly facilitate its use with geographic information systems (GIS).

I am writing this introduction, but MarkSim would not have happened without Phil Thornton. It has been done with remarkably little outside funding. John Lynam of the Rockefeller Foundation has given us a couple of small, but incredibly useful, grants. We would like to thank Paul Wilkens of the International Fertilizer Development Center (IFDC) for his programming in Delphi of the first version of the Windows interfacer. (Paul, you will recognize some parts of it.) William Díaz, as my system analyst and programmer, has born the brunt of my quixotic decisions on the look and feel of the software for over a year now.

This version works. I am sure that the next will be better, but as it is already 2 years late, this is what you get.

Peter G. Jones

Getting Started

MarkSim is a Windows application that will be installed from the CD-ROM and registered automatically. The program files will normally be installed in the directory C:\Program Files\CIAT\, and unless you have a good reason for installing in another directory we strongly recommend that you let the install package go ahead and do so.

✓ Insert the CD-ROM in your CD drive.

✓ Go to the run prompt and type X:\setup where X is the drive letter of your CD drive.

The MarkSim system comes with large data files. The first window of the install procedure shows an analysis of the disk space available on your system and subsequent windows will allow you to tailor the installation to make best use of this space.

✓ Note where best to install MarkSim.

✓ Hit Yes to proceed.

✓ Read the notes on the following screens.

✓ Then choose the relevant installation type.

The largest set of data files is the map coverages. These are Environmental Systems Research Institute (ESRI) shapefiles that are used to create the backgrounds for the maps you will use with MarkSim. The directory is called \coverages\ and is 582 Mb. You can elect to leave it on the CD-ROM if you are short of disk space. In this case choose the option 'Typical' when the Install shield requests it. Leaving it on the CD-ROM will not slow MarkSim operations to any great extent, but it does mean that you have to have the CD-ROM in the drive whenever you work. If you choose to install it on the hard disk, the install shield will attempt to put it in a directory \MarkSimFiles\ on a disk with sufficient space. You may override this and choose another site for it if you wish.

The climate grid files and all the model parameters are stored in the directory \markdat\; this is currently 336 Mb. It has to be installed on a disk and will be flagged read only. We suggest that you install it, if possible, away from the program files on your C disk. The install shield will attempt to put it in the directory \MarkSimFiles\ as above.

The last choice you have is where to put the working directories \dat\ and \output\. These will contain your input and output files. It is also best to keep these away from the program files directory. MarkSim output can be voluminous so make sure that wherever you decide to put them there is sufficient disk space.

✓ Hit Finish to start the installation.

1. Tutorial

This tutorial gives a quick introduction to some of the common operations you may be doing with MarkSim. The software is designed to produce simulated daily weather data for any point in the tropics. It runs off interpolated climate surfaces and operates in two parts. The first creates a file (CLX file) of model parameters. The second runs the MarkSim simulation to produce the daily weather data files. For details of the operations see User Reference Section (p. 20). For how the model works see Theory (p. 44).

MarkSim uses three main subdirectories, two for working files and one for map coverages.

MarkSim offers you two types of input. If you know the monthly average climate data for the point you wish to simulate, you can enter them. This type of input restricts you to points with actual climate data, but it operates fully independently of the interpolated climate grids so it will work for anywhere in the world. The second type of input is where you do not have climate data, but know the whereabouts of the point you wish to model. This works from the climate grids and will simulate any point in the tropics provided that it is on a climate grid. This method is somewhat restricted at present; it works for Latin America, Africa, and South East Asia, including Asia and southern China below 34º N.

Grid Independent Climate Data

This form of input depends on the .DAT file to input data to MarkSim. A .DAT file looks like this:

```
mex07160   17.130 -92.720      70    471    471     471
   42. 43.   23.   47.  115.  264.  188.  236.  275.  172.  86.   56.
 18.0  18.2 20.8  22.4  22.8  22.1   21.7  21.5  21.1  20.8  19.5  17.9
 10.4  11.1 13.0  13.3  12.9  11.7   11.6  11.6  10.5   9.6  10.1   9.9
```

See Appendix A for a full description of the format. You can produce this fixed format ASCII file in a number of ways. If you have data tabulated and wish to write out a series of DAT files, the FORTRAN format for the file is:

(a8,2f8.3,i6,/12f5.0,/12f5.1,/12f5.1)

Running single DAT files

If you have only a few files to prepare, you may type them in directly in an ASCII editor, or use the MarkSim editor. For grid independent data entry, you do not need to load a map, although if you have one displayed by default, there is no harm in leaving it there.

✓ Select the spatial input tool to bring up the spatial input window.

If you wish to enter data as a single DAT file, select the DAT option in the third panel. If the DAT file exists, you can browse for it in the DAT directory. We have placed palmira.dat there for you to try.

✓ Browse for palmira.dat.

✓ Open it.

✓ Create the file palmira.clx by clicking on Run MarkSim in the lower left corner of the window.

If there were no errors, MarkSim will tell you so and ask if you would like to see the log file. This is a file to record the process of the runs. If you want the full information on the run, choose full on the panel clxgen.log. Selecting errors will give you a minimal output with only the error messages. It is best to change to this option once you are processing large quantities of data.

Now to practice entering the data with the MarkSim editor.

✓ Select the DAT panel.

This editor icon will light up on the right of the panel.

✓ Type in the data from the Mexican station given in the example above. Type –500 for the January rainfall and try to save the file.

We have included some rudimentary data checks to trap errors. If you want to check your typing as you progress, use the cloud question mark icon to do a running check on the file.

✓ Correct the January rainfall, save the file, and run the job.

✓ Check with the log and have a look at the CLX file.

You will find a detailed description of the file contents in Appendix A.

☞ *When viewing a file, you are offered the option of editing it or viewing the data as graphics. You may edit a CLX file, but we highly recommend that you do not do so. The parameters are interlinked and editing one without adjusting the set may result in serious errors.*

You can also use the DAT editor to correct files.

✓ Enter the editor.

✓ Select open a file.

✓ Browse to find the Mexican file you have just made.

✓ Open it and change the data.

✓ Change the site name and save it as another file.

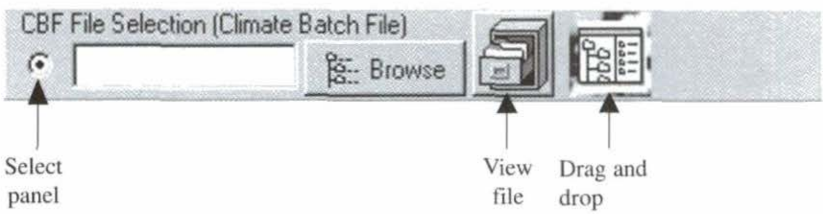
In this way you can use the base data in one file as a template for another. Only make sure that you have changed all the data necessary to completely define the new file.

Running multiple DAT files

It often happens that you will want to simulate a lot of points at a time. Using MarkSim along with a GIS is a good way of testing model results over a study area. It is also a good way to be left handling very large quantities of data. For this reason, we have included a number of batch processing options. The Climate Batch File (CBF) is one example.

The CBF is a sequential ASCII file containing, in each record, the FULL path to a DAT file. It looks like this:

```
C:\Program Files\CIAT\MarkSim\dat\K9238003.dat
C:\Program Files\CIAT\MarkSim\dat\Mex07160.dat
C:\Program Files\CIAT\MarkSim\dat\Hendersn.dat
```



You can construct a CBF in many ways. You can type it in to an ASCII editor, construct it from the DOS DIR instruction, or you can use the handy drag and drop facility provided in MarkSim.

✓ Select the CBF panel on the spatial input window.

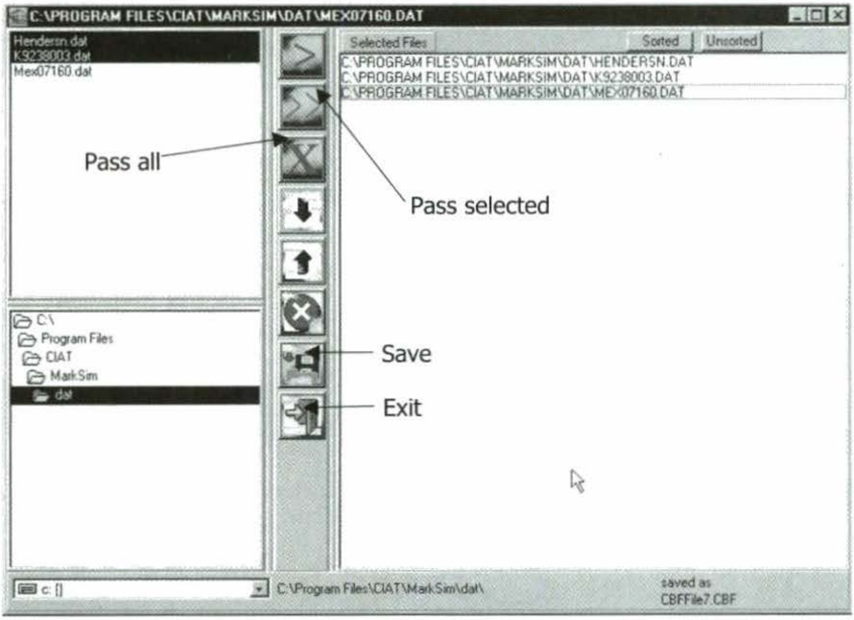
✓ Click on the drag and drop icon at the right of the panel.

You will see the DAT files available in the DAT directory displayed in the top left window. You can search for DAT files in other directories or on other drives by altering the path and drive in the lower windows. To select a single DAT file:

✓ Highlight it by clicking on the filename.

✓ Use the pass selected button to transfer it to the file building list in the right hand window under selected files.

To select all DAT files in the directory use the pass all button. You can change directory to add more DAT files from elsewhere in your system.



✓ Save the CBF and exit.

✓ Browse and select the created CBF with the browse function.

✓ Open it, and check it with the view file button.

When you now run it, MarkSim will create three CLX files in the output directory.

Grid Dependent Data

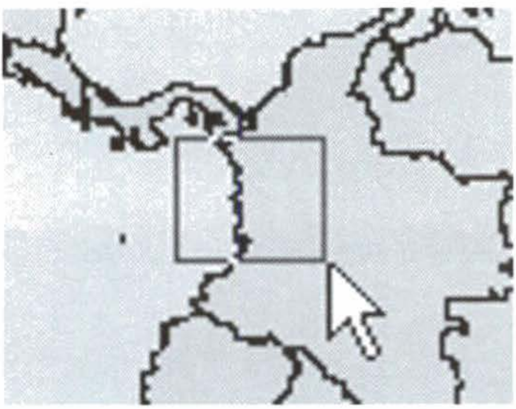
This is the main purpose of MarkSim. From the interpolated grids you can produce a simulated daily output for most points in the tropical world.

Setting up the map

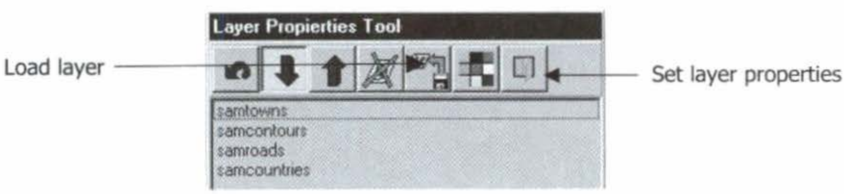
The coverages directory contains ESRI shapefiles of map background information that you can use to display a map to navigate the climate grids. We will start by making a map to use with the Latin America climate grid. The background layer samcountries will be loaded automatically in the newly installed version of MarkSim. You can change this default with the configuration tool, but for the present lets leave it as default.

✓ Select the layer control tool to display the layer control window.

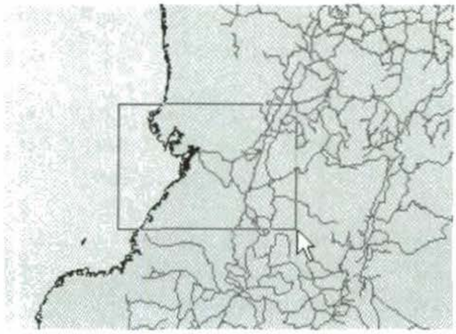
✓ Use the zoom in tool to zoom into a window in western Colombia.



We are going to add layers until you can see a detailed map that you can navigate to find the relevant pixel for CIAT, which is situated 23 km northeast of Cali on the road to Palmira. We will zoom in as we go because the layers we are going to add will cover the continent with a clutter of information.



✓ Zoom in again as at right.



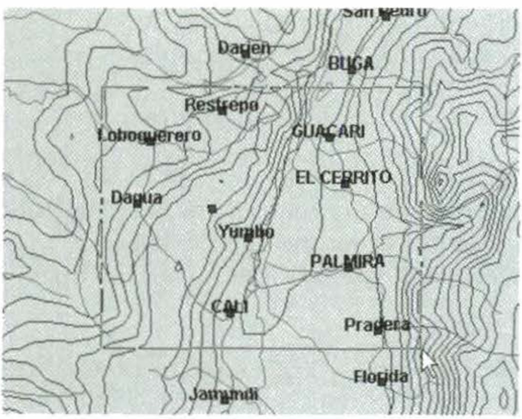
✓ Select the load layer icon in the layer properties tool.

You will be shown the layers available in the coverages directory.

✓ Select samroads.shp and change the color to red.

✓ Zoom to the window shown on the right, then select the coverages samrivers, and change the color to blue.

✓ Select samtowns; use set layer properties to set on the name in the labels fields.



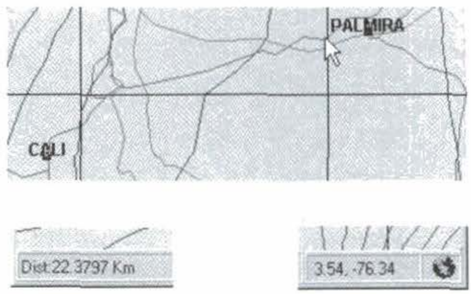
Now we can see where we are. CIAT lies in the Valle del Cauca, or valley of the river Cauca, between two large Andean mountain ranges—the Cordillera Central to the east and the Cordillera Occidental to the west, also called the Farallones or cliffs of Cali.

✓ Make sure you have the zoom in tool selected and place the cursor over the place symbol for Cali, which appears directly below the "A" in the city name.

✓ Left click.

The map will redraw and a small blue dot will appear where the cursor was placed.

✓ Track the cursor along the road to Palmira (the Recta in local parlance) until the distance (at the lower left of the map window) registers 23 km.



You have now arrived at the front gates of CIAT and the coordinates in latitude and longitude appear at the lower right of the window. You are nearly ready to construct the CLX file for the location. However, the climate grid you are working from has pixels of 10 arc minutes on the side (about 18 km at this latitude). The valley at this point is only about 30 km wide (check this with the MarkSim measuring tool just like you measured the distance down the Recta from Cali). There is therefore one last check to make.

✓ Go to the layer control tool and select america_grid from the shapefiles.

This contains the pixel boundaries of the climate grid. You will have to go to the layer properties tool to set the fill to transparent because it is a polygon shapefile and you will need to see the map through it.

The grid pixel boundaries show that CIAT is almost exactly on a pixel boundary. The eastern pixel includes some of the foothills of the Cordillera Central, whereas the western pixel is almost all valley floor.

Climate diagram tool

✓ Check on the climate data to which you will be fitting.

✓ Select the climate diagram tool and click on the western pixel.

✓ Now click on the eastern pixel.

(The first climate diagram will disappear behind the map window. You will have to shift the map window to pick it up. I am sorry about this; it is a glitch that we have not been able to fix as yet.) You will notice that there is very little difference. This is because the National Oceanographic and Atmospheric Administration (NOAA) digital elevation model (DEM) to which the climate grid is fitted holds the modal elevation, not the average, so it is approximating well to the valley floor. The small difference you will notice is that the valley floor (western) pixel is slightly drier.

This is actually masking a larger effect that we would expect in this valley. The Valle del Cauca is a large tropical valley and exhibits the typical large tropical valley circulation where there is a predominance of descending air in the valley center because of differential solar heating at the sides. This results in a rainfall gradient that is wetter at the sides and drier in the middle. MarkSim will shortly be linked to high precision (1 km or 20 arc second) grids, but we have to fix some problems of data storage and access before this can be implemented.

✓ Choose which pixel you want and select the select a latitude, longitude point tool, point at the relevant pixel, and left click.

Select a latitude, longitude point tool

The spatial input window will appear with the coordinates and elevation of the pixel filled in for you.

✓ Type in a name for the CLX file.

✓ Choose full reporting in the clxgen.log panel.

✓ Hit the run clxgen button.

You should see a message saying no errors were encountered and asking you if you would like to see the log.

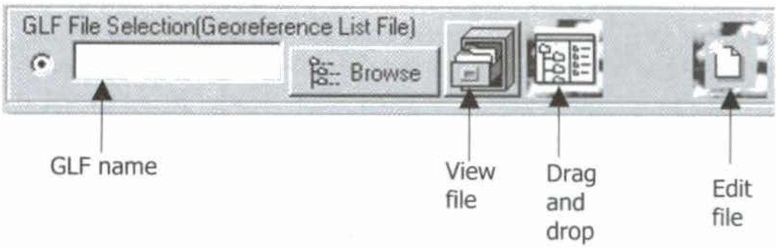
✓ Say yes and check what MarkSim has done for you.

When you are more confident about what is happening, you can change the reporting option to errors only to save creating a large log file.

Multiple georeferenced point data

If you are a power user, perhaps running with a GIS system to simulate points sampled over an area or along a transect, you will want a batch running system where all you do is specify the latitude, longitude, perhaps the elevation, and a name for the point. The Georeference List File (GLF) is designed to do just that.

✓ Select the spatial input tool and go to the GLF panel.



You can prepare the GLF as a comma-delimited sequential ASCII file with any ASCII editor; use the drag and drop facility or the MarkSim GLF editor. A GLF could look like this (spaces are not significant and missing elevation is recorded as -999):

```
23.602,  -46.948,  853,   ITAPEVI
3.460,   -76.525,  1523,  CALI
4.340,   -72.316,  213,   CARIMAGU
-32.918  -68.854,  1219,  CLXFILE3
3.895    -77.073,  -999,  BUENAVEN
```

The drag and drop facility is of relatively limited use here because it searches for CLX files from which to extract the filename. Since the object of the exercise is to create CLX files this seems a roundabout way to do it. It does, however, have some use when you might wish to recreate a set of CLX file or correct location data. This could possibly be of use if you change from one climate grid to an updated one and you wish to recreate a set of CLX files with the new data. Note that the latitude and longitude are in decimal degrees.

✓ Go to the GLF editor and type in all or part of the GLF shown above.

✓ Save it and run it by selecting the GLF panel option.

✓ Open the file, and run with the run clxgen button.

✓ Check that all the CLX files were created and that the missing elevations were filled in from the climate grid.

If you used the GLF in the example above, you will have noticed that it finishes with an error. If you look in the log, you will see that almost all of the CLX files were created correctly. ITAPEVI.CLX, however, was not produced. There is a warning in the log, but MarkSim carried on to process the rest. If you look at the coordinates for ITAPEVI, you will notice that this point actually falls in the middle of the Atlantic Ocean. Itapevi is actually in the state of São Paulo in Brazil. Unfortunately, the validation routine in the GLF editor can only cope with checking if the latitude and longitude are *possible*. It cannot check if they are *correct*. Someone has left off the negative sign. That is no problem now that you have found it.

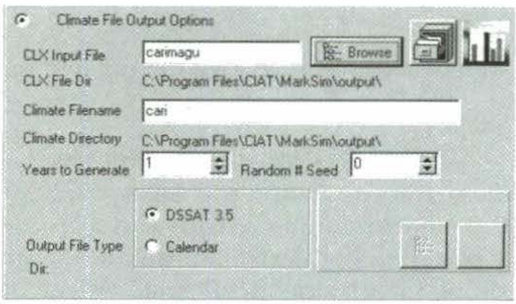
✓ Go into the editor, correct it, and run the job again with just Itapevi because all the others ran correctly.

Running the Simulation

Running a single site

Once you have created your CLX files you need to move to the rungen phase to run the simulation and produce your simulated daily data output files. The generate data tool will take you there, or if you are in the climate input window merely switch to the second page.

The top panel will allow you to run a simulation from a single CLX file.



✓ Use the browse key to find one of the CLX files that you created in the first part of the tutorial.

We will choose to generate DSSAT 3.5 style output for use with a DSSAT crop growth model.

✓ Select the CLX file CARIMAGU.CLX from the output directory.

✓ Type in a climate filename of four characters or less.

This will be the name of the DSSAT CLI file that will be produced. In this example it will be called CARI.CLI and each year of the daily data output will be called CARInn01.WTG, where nn is the number of years.

✓ Set the random number seed to 1243.

✓ Set the number of years you require and hit the run button.

✓ Check the log to ensure that everything worked correctly.

You can now select the output file to check the data. The start of the file CARI0101.WTG should look like this:

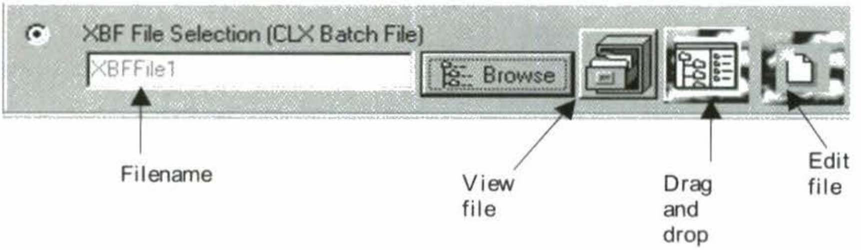
```
*WEATHER : cari From Interpolated Surfaces

@ INSI   LAT     LONG     ELEV    TAV  AMP   REFHT   WNDHT
  cari 4.340 -72.316   213    27.2 11.6  -99.0   -99.0

@DATE     SRAD    TMAX    TMIN   RAIN
01001     22.0    33.2    22.4    0.0
01002     27.2    38.6    23.1    0.0
01003     27.2    38.6    23.0    0.0
01004     24.7    37.8    23.0    0.0
01005     27.2    38.7    24.0    0.0
```

✓ Go back and change the random number seed and rerun the job.

The file CARI0101.WTG will now contain different simulated data. You can, however, exactly duplicate the original run by setting the random number seed back to 1243. If you leave the default seed, the actual seed used will be shown in the log file, so even if you did not specify it you can always repeat a run if you so require.

Running multiple sites

The last exercise is to run the simulation for multiple sites. You have already prepared a number of CLX files. You can now run these from a batch facility. This uses XBF or CLX batch file.




```
C:\CIAT\MARKSIM\OUTPUT\AFRICA.CLX,AFRI,4003,12,c
C:\CIAT\MARKSIM\OUTPUT\ASIA.CLX,ASIA,2919,12,c
C:\CIAT\MARKSIM\OUTPUT\BRASIL.CLX,BRAS,5336,12,c
```

Full path to CLX file

DSSAT site name

Random number seed

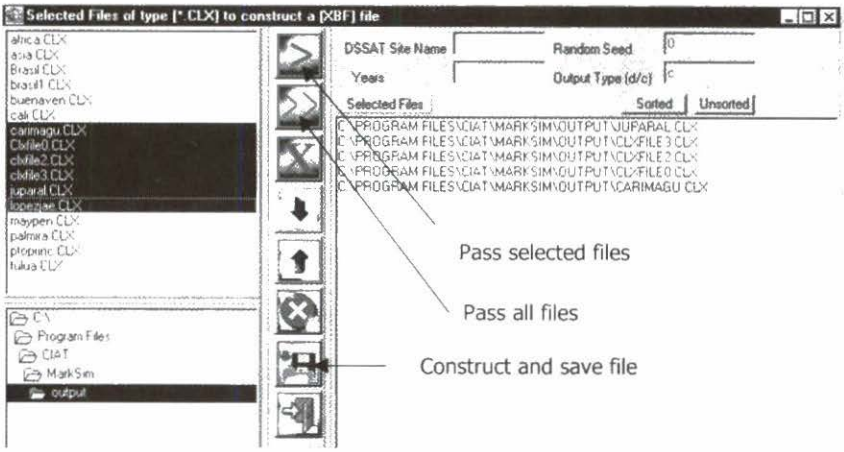
Years

Output type

The XBF is a comma-delimited sequential file with the fields as shown above. You can type the file into any ASCII editor, but because the full path is needed on the filenames it is much more efficient to use the drag and drop facility. The CLX files that are in the output directly will be displayed. You can search for other files by changing the drive and path. You can construct an XBF with CLX files drawn from various sources.

✓ Select files to be incorporated and transfer them to the file building window on the right.

You have various options. You must choose a number of years and output type, but the other fields are optional. If you leave the DSSAT site field blank, the first four characters of the CLX filename will be used. However, the site name must be unique, so if duplicates exist the name is incremented alphanumerically. Thus, if two fields result with the site name CLXF, as in the example, the second is incremented to CLXG, the third to CLXH. If you enter a site name, then that is used as the first site name in the file, and all subsequent ones are derived by incrementing. If you leave the random number seed blank or zero, then the first is derived from the system clock, and subsequent ones from the random number generator.



✓ Select DSSAT output, enter 8 years, leave the DSSAT name blank and the random number seed at zero.

✓ Hit construct and save file, then exit.

You will get a warning that the DSSAT name is blank, but you can ignore it this time. Your XBF will look like this:

```
C:\PROGRAM FILES\CIAT\MARKSIM\OUTPUT\JUPARAL.CLX,JUPA,6859,8,d
C:\PROGRAM FILES\CIAT\MARKSIM\OUTPUT\CLXFILE3.CLX,CLXF,7353,8,d
C:\PROGRAM FILES\CIAT\MARKSIM\OUTPUT\CLXFILE2.CLX,CLXG,7473,8,d
C:\PROGRAM FILES\CIAT\MARKSIM\OUTPUT\CLXFILE0.CLX,CLXH,1849,8,d
C:\PROGRAM FILES\CIAT\MARKSIM\OUTPUT\CARIMAGU.CLX,CARI,2866,8,d
```

Now you can run the file, but also you can still edit it before running if you wish. This has the advantage that you can alter the details of the run for each record in the file. The output type or number of years does not have to be constant throughout the XBF. If you wish to change one or more lines, go to the XBF editor and make whatever changes you need before running the file. Just for fun:

✓ Change the output for Juparal to "c", and the years to 10.

- ✓ Change the years for Carimagu (actually Carimagua, a CIAT station in the Colombian Llanos) to 4.
- ✓ Hit run rungen; check the files that appear in the output directory.

There should be one called JUPARAL.GEN with 10 years of calendar output in it. CLXF0101 to CLXF0801, CLXG0101 to CLXG0801, CLXH0101 to CLX0801, and finally CARI0101 to CARI0401, all containing 1 year each of DSSAT 3.5 output.

2. User Reference Section

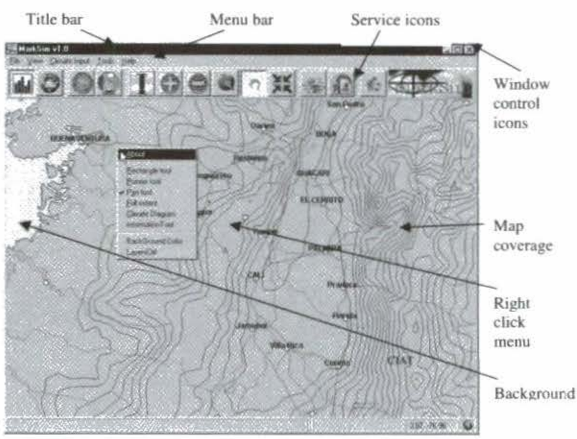
Overview of MarkSim Operation

MarkSim is a daily weather generator based on a third order Markov model for rainfall that is especially adapted to the tropics. It runs off interpolated climate grids to estimate the parameters of the model. It runs two parts that can be operated separately. The model parameter estimation is the first part. This produces an intermediate file known as a .CLX file that contains the model parameters. The .CLX file is then used as input to the second stage where the simulated daily data are produced.

The system offers a variety of input forms including an option to choose a point from the map. The output comes in two standard forms, the MarkSim calendar format and the DSSAT model input format. See Appendix B for descriptions of the file formats.

The Map Window

The basis window of MarkSim is termed the map window. Although you do not need to load a map in order to use MarkSim, this is the first window that appears when you fire up the software. The window contains the menu bar and the service icons that you will use to do the job. You can also access some of the service functions through the right click menu.



✓ Place the cursor anywhere on the map window and right click. This small menu will appear.

A right click on the title bar will give you the standard Windows ® control – (move, size, minimize, maximize, close). You can also control the window with the standard window control icons.

The menu bar consists of pull-down menus that will activate the various services and tools. All of these except help are available directly from the service icons. The about box gives you information about the authors and about various copyright considerations for the software used in MarkSim development.

The right click menu gives you an alternative route to some of the tools that are found in the main menu service icons.

On the CD-ROM you will find a range of shapefiles that you can overlay on the map using the layer control tool. You can find these in the directory \coverages\. They are not placed automatically on your hard disk at installation because you may not want to use them all. They will supply map features such as roads, rivers, and towns to help you navigate about the map. Beware! They are for use with the map zoomed well in, to present sufficient detail. If you apply them to the map at full extent, they will be so dense that they will practically color the map. The map in the illustration is composed of sammunicip.shp, samcountry.shp, samtowns.shp, samrivers.shp, and samroads.shp, and shows the area around CIAT. (You will not find CIAT in the shapefiles; I just put it there to let you know where we are.) The layer control tool allows you to color the map to your liking.

Go to the layer control tool and choose the layers you need to give you enough background to localize the area in which you are working. If you are in doubt as to where the cursor is pointing, the latitude and longitude appear in the lower right corner of the window. In this case, the cursor is on the right click menu header bar and hence is actually a few kilometers east of Buenaventura. To measure distances on the map:

✓ Select the zoom to area tool.

✓ Left click on the map from where you want to measure.

A small blue dot will appear on the map at that point. The distance from this point to the cursor is continuously displayed at the lower left corner of the map window. Do not hold the left button down while moving the cursor or you will draw out the rectangular extent for the zoom tool.

MarkSim selects the data from which to calculate the model parameters from an interpolated climate grid. These vary in pixel size and hence in precision. For Latin America and Africa these are currently 10 arc minutes (about18 km), and for Asia 2.5 arc minutes (about 4 km). In mountainous areas, this pixel size may not allow a full description of the terrain, and in coastal areas, there may be slivers of land that are not covered by the grid. To check exactly where you are on the grid, a set of shapefiles is provided that displays the grid outlines. These are called america_grid.shp, africa_grid.shp, and asia_grid.shp. You will find them on the CD-ROM with the other coverages.

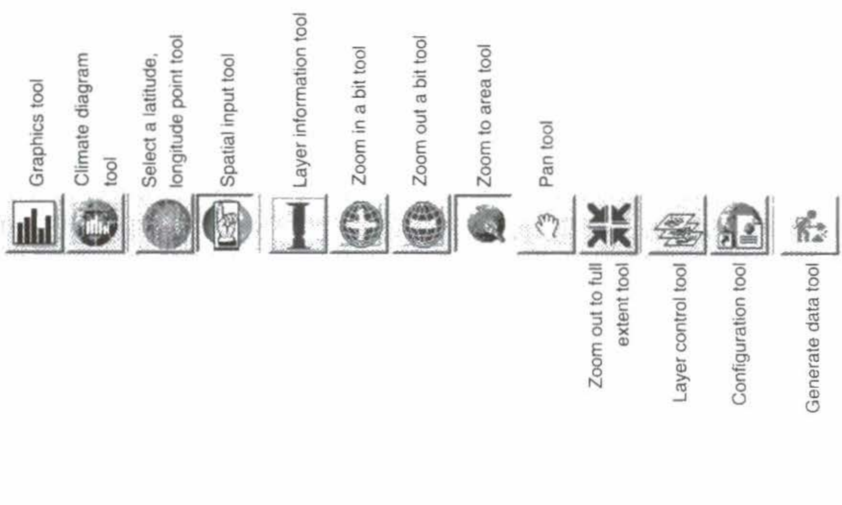
✓ Load them with the layer control tool.

✓ Set the fill to transparent and the outline on in the color of your choice.

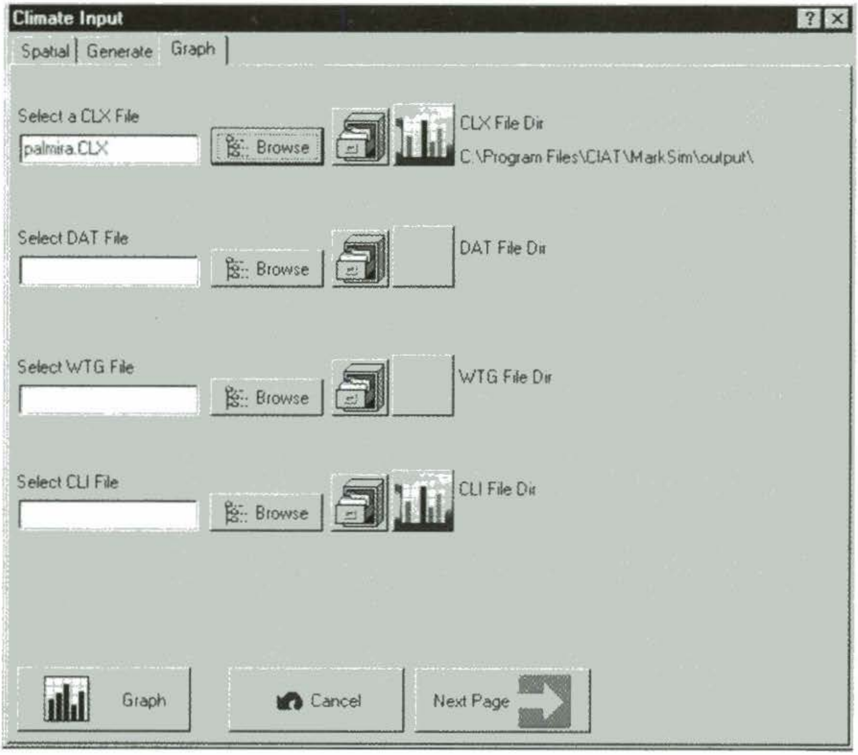
This will show you the exact position of the grid pixels. Note, however, that at small scales the grid will completely cover the continent with outline color. To see the pixels you have to zoom in considerably.

The Main Menu Service Icons

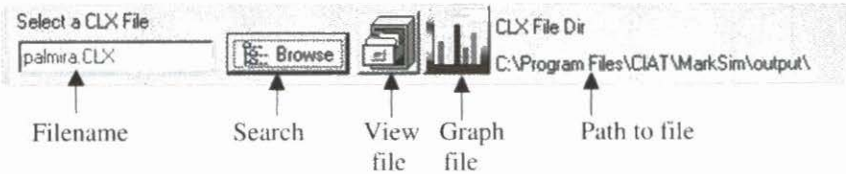
The figure shows the main service icons. We will explain them each in turn, moving from left to right.



The graphics tool



Certain of the MarkSim operating files have climate data associated with them. These data can be shown graphically from various windows. The graphic tool provides direct access to these graphics. The CLX files are used to transfer model parameters from the parameter estimation phase (clxgen) to the stochastic weather generation phase (rungen). The DAT file is a method of presenting climate data to the clxgen phase. The WTG files are DSSAT standard weather files produced in the rungen phase, and the CLI files are DSSAT files associated with the generated data files and are necessary to run a DSSAT model. See Appendix A for the file format definitions.



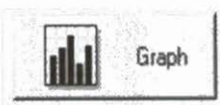
Use the browse facility to identify the file to be displayed. The view file icon will display the file in the MarkSim editor.

☞ *When viewing a file, you are offered the option of editing it or viewing the data as graphics. You may edit a CLX file, but we highly recommend that you do not do so. The parameters are interlinked and editing one without adjusting the set may result in serious errors.*

The graph file icon will display the available climate data from the file. Those available from a CLX file are monthly rainfall, mean monthly temperature, mean diurnal temperature range, and solar radiation. A DAT file contains the same variates less solar radiation. A WTG file contains the simulated daily values of solar radiation, maximum and minimum temperatures, and rainfall for a whole year. The graphs are presented month by month. The CLI file contains monthly values for solar radiation, maximum and minimum temperature, number of raindays, and sunshine hours. In MarkSim, sunlight hours are not estimates so this variate always shows missing values (-99).

These graphic displays are produced by TeeChart. This software gives the user considerable control over the type of display produced.

✓ To invoke the TeeChart graphics control, press ctrl T.

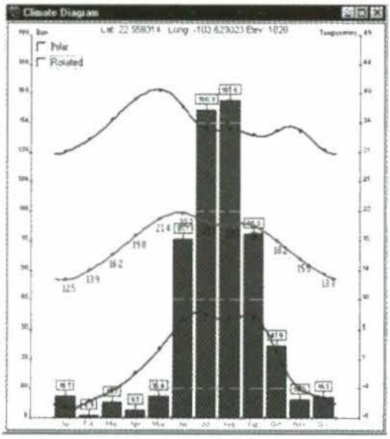


This graph button at the lower left of the window gives access to a different form of display for the CLX file data. This is the climate diagram tool, which is described in the next section.

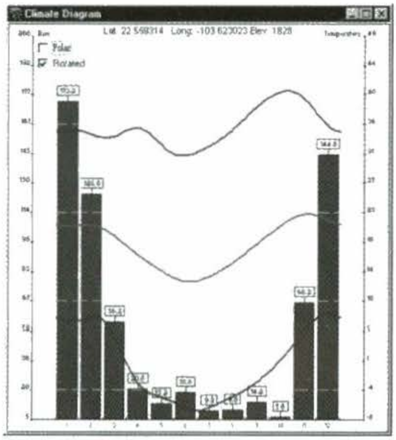
The climate diagram tool

✓ Select the climate diagram tool and click on any point on the map.

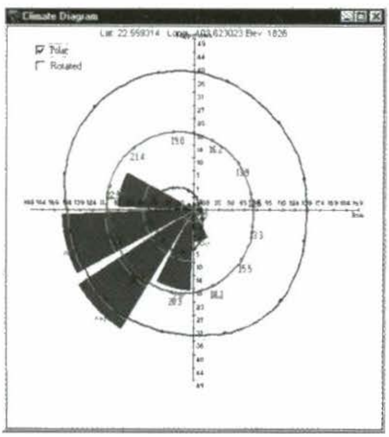
The diagram will be displayed with maximum, minimum, and mean temperatures, and monthly mean total rainfalls. There are options for Cartesian or polar coordinates and for standard or rotated displays. (See Theory section for explanation of the rotation). Under rotation the month names are meaningless so the months are merely numbered.



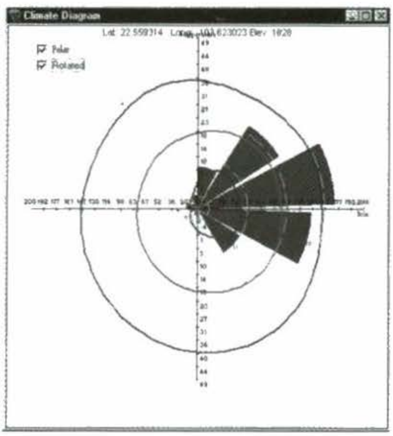
Cartesian normal



Cartesian rotated



Polar normal

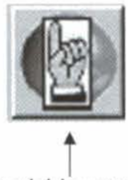


Polar rotated

The tools to input spatial coordinates

Two icons control this function. They both bring up the same window, but by different operations.

Select latitude, longitude point tool

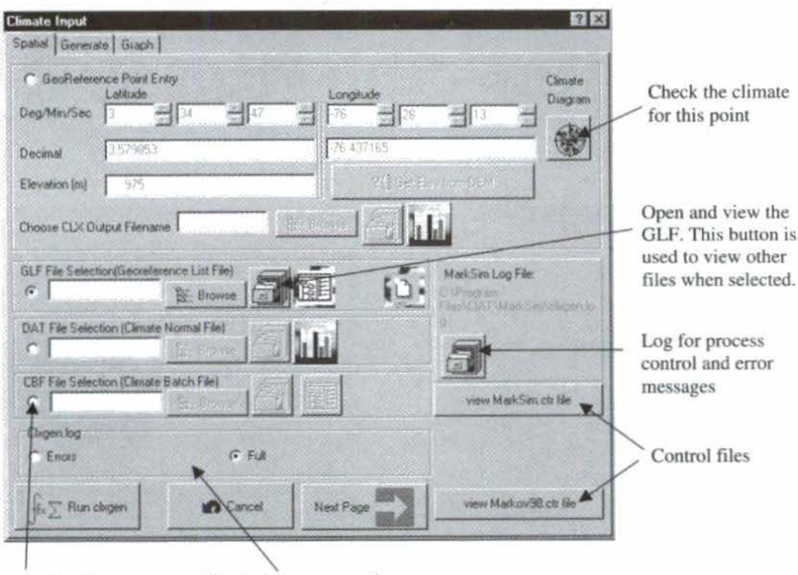


Spatial input tool

The select latitude, longitude point tool does exactly what it says. The spatial input tool brings up the climate input window directly to allow you to choose the form of entry you require.

✓ Point and click on the map.

The climate input window will appear with the selected coordinates and elevation showing in the georeference point entry section of the window.



The climate input window controls the creation of the intermediate model parameter file known as a CLX file. This needs the climate data from a point on the interpolated climate surface as input. The process uses two control files that can be viewed from the window after the CLX file has been created. A record of the run is kept in the log file that can also be viewed after the run. The log file can contain a full informative listing of the various operations in the process, or can contain just error messages. Once you are sure that everything is correct with a run or set of runs, we recommend that you set the error reporting control to errors only, because the log file can become large on long runs that create many CLX files.

☞ *When viewing a file, you are offered the option of editing it or viewing the data as graphics. You may edit a CLX file, but we highly recommend that you do not do so. The parameters are interlinked and editing one without adjusting the set may result in serious errors.*

The CLX file must be given a name of up to eight characters and the point to be simulated, or the climate data for a simulation point must be provided. This can be done in four different ways. Use the panel select button to choose between the options.

1. *Georeference point entry*

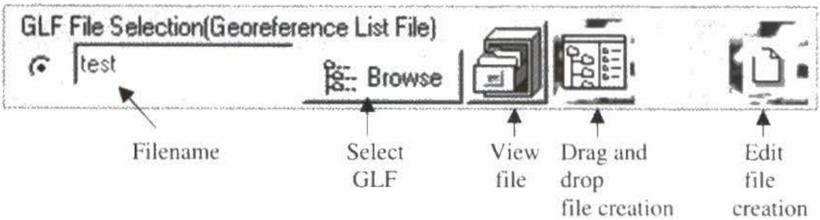
The simplest form of spatial entry is controlled in the upper panel of the climate input screen. Latitude and longitude are shown in degrees, minutes, and seconds, and as decimal degrees. If you have entered via the spatial icon, these fields will be blank. If you entered from choosing a point on the map, they will show the values for that point.

If you decide to enter the latitude and longitude from the keyboard, you can enter them as decimal degrees, or as degrees, minutes, and seconds. The elevation in meters is necessary for the operation of MarkSim. However, if you do not know it, then you can use the key provided to fetch it from the DEM that is an integral part of the climate surfaces.

☞ *In this version of the software, the climate grids for Latin America and for Africa are at a resolution of 10 minutes of arc. This is about 18 km at the equator. In mountainous regions, this resolution can give a poor estimate of the actual elevation of your chosen point so it is better to enter the known elevation if you have it.*

If you enter a location by pointing at the map or typing in the coordinates, you must enter a name for the CLX file. This should be a valid DOS filename (one to eight characters) without a file extension.

2. *Georeference list file selection*



The next option for georeferenced points entry is controlled by the second entry panel and is the GLF selection.

✓ With an ASCII editor, such as Notepad, prepare a file containing a list of latitude, longitude points, with or without elevation data, and put a CLX filename on each line.

The data should be comma separated and could look like this:

```
-12.45, -67.1, -999, PtoVelho
3.5, -76.5, 967, Palmira
-2.33, 37.5, 1800, Nairobi
```

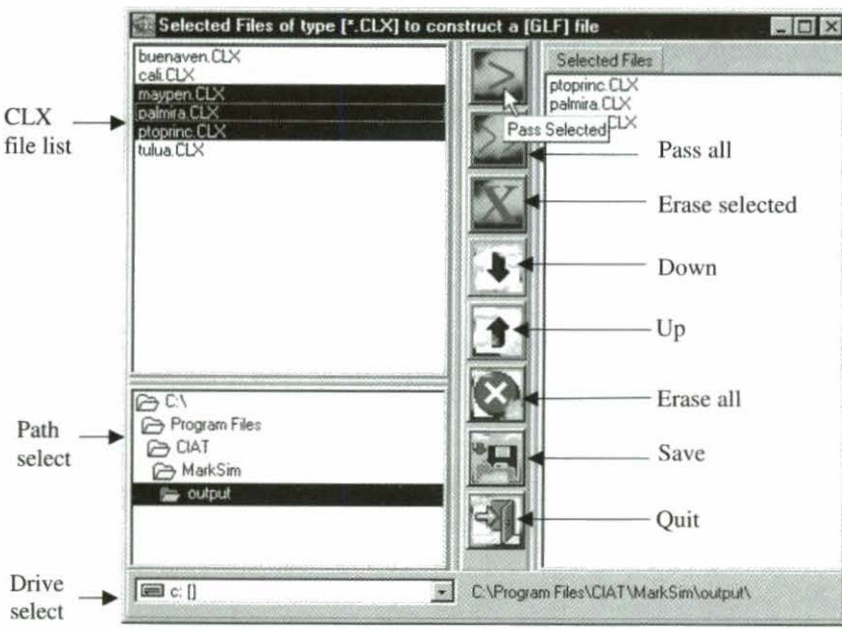
✓ Name the file filename.GLF and put it in the data directory.

✓ Now check the GLF selection option and browse the data directory to pick up the filename.

✓ Alternatively, use the convenient GLF data entry form by clicking on the page symbol at the right of the panel.

This will construct the comma-delimited file as you type in the fields and also check the coordinates and filenames for validity. You can also use this to edit a GLF and to validate one that has been prepared by an ASCII editor. It checks that the CLX filenames are acceptable and that all coordinates and elevation are within realistic bounds.

Another way to enter georeference data into a GLF is to drag and drop the location information from a list of CLX files. This would appear to be a circular argument because what clxgen is going to do is to create the CLX files. However, this is a quick way of updating a long list of CLX files, which you might want to do if they have been damaged in some way, or if the underlying climate grids have been updated. This will happen from time to time as the basic database improves and interpolations are redone.



The list of CLX files that appears automatically will be from the default output directory. You can search for other sets by changing the path or drive.

3. *Climate normal file selection*

The third panel allows you to enter data from another climate data source. The data are entered in a special fixed format file known as a DAT file. These are used internally in the creation of the MarkSim models, hence the fixed format. The file is a fixed format file with the following FORTRAN format (a8,2f8.3,i6,/12f5.0,/12f5.1,/12f5.1).

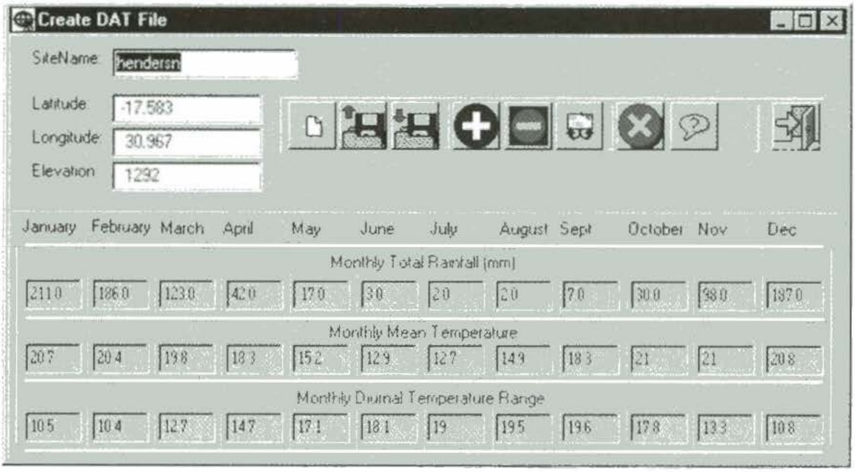
Here is an example:

```
hendersn -17.583  30.967  1292
 211.    186. 123.42.  17.  3.   2.   2.   7.   30.  98.  187.
 20.7    20.4 19.818.3 15.212.912.714.918.321.021.0 20.8
 10.5    10.4 12.714.7 17.118.119.019.519.617.813.3 10.8
```

The values are filename, latitude, longitude, elevation (meters), 12 monthly rainfalls, 12 monthly mean temperatures, and 12 monthly mean diurnal temperature range. In MarkSim, the diurnal temperature range is defined as the difference between mean monthly maximum and mean monthly minimum. Latitude is decimal degrees with Southern latitudes negative. Longitude is decimal degrees with longitudes west of Greenwich negative. A DAT file can be created using an ASCII editor such as Notepad. However, in this case too we have included an input facility to help with the formatting.

✓ Click on the page symbol at the right of the panel and the DAT data entry window will appear.

You may use this to enter a new DAT file, or, if you have selected one with the browse facility, you can use it to edit the file and validate the data. At present, MarkSim merely checks that the site name is valid and that coordinates, elevation, and data are within real world limits. In later releases we will be including more sophisticated checks.



✓ Make sure that the DAT file is in the MarkSim data entry directory defined in the configuration window.

4. *Climate batch file selection*

You may have a large number of DAT files to use to create CLX files. These may be in the MarkSim data entry directory or they may be elsewhere. The climate batch file selection panel allows you to select a CBF that gives the names and paths to these files. It may be created with an ASCII editor or, more easily, by drag and drop from a list of DAT files. The drag and drop window is identical to that described above to form the GLF. The only difference is the format of the file created. In this case, it is the name of each DAT file including the full path.

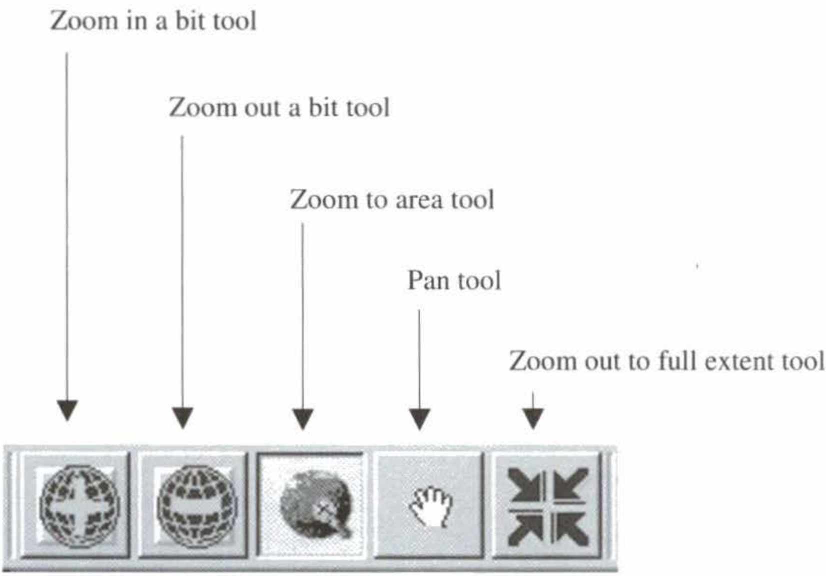
Error reporting

The clxgen process produces a file called clxgen.log as it runs. At the end of a run a window will be displayed asking if you wish to see the log file. You can also open it at any subsequent time by clicking on the log file button. It is overwritten each time you run the process, so if you wish to keep the information it should be renamed. It is produced in the c:\program files\ciat\MarkSim directory. This directory holds sensitive files so observe caution when accessing it with software other than MarkSim. At the foot of the spatial input window two buttons select the type of reporting. Errors will give a short report including only error messages and warnings. Full yields a complete listing including all control file records and the resulting CLX output. When processing large numbers of points, this can result in a large log file. It is best to use full only when you encounter difficulties.

Other mysterious files

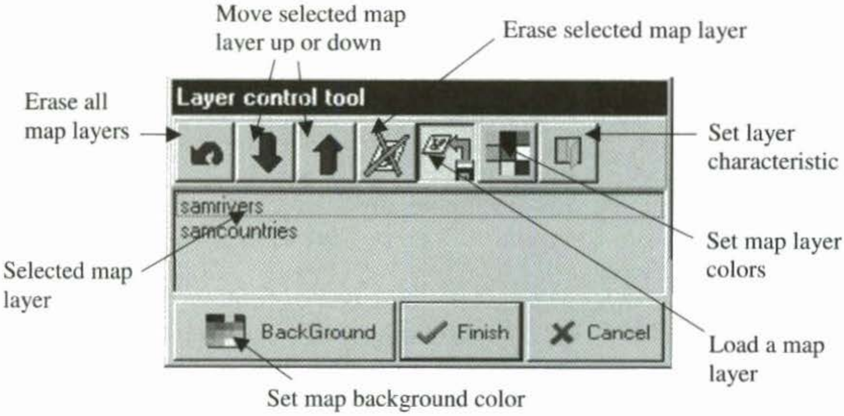
MarkSim.ctr and Markov98.ctr are control files that are used to transfer information to the stochastic weather generation process. You usually only need to view these files for debugging purposes; their structure is given fully in Appendix A. They are also found in the c:\program files\ciat\MarkSim directory.

The zoom tools



The zoom and pan operations are reasonably standard. Zoom in by drawing the desired window on the display map. Pan by pushing the map with the hand cursor. The zoom out feature is a map reset; it will zoom out to the full extent of the map. For a gradual zoom in or out, the zoom in a bit and zoom out a bit tools are available.

The layer control tool



Each map layer is an ESRI shapefile representing geographic features that will help you to identify the points you want to choose on the map. Shapefiles are described in Appendix B. A number of files are included on the CD-ROM. MarkSim will, however, accept shapefiles from any source provided that they are in geographic coordinates (latitude, longitude) and are at a scale appropriate for the window in which you wish to work.

☞ ***NOTE:*** *The files of roads and rivers included with MarkSim v1.00 are not suitable for display at the scales of the full map extent. However, they are useful when displaying at the department or district level, that is to say, zoomed in to large scales.*

The icon erase all map layers clears the map completely for you to start anew. It only removes them from the display map and does not affect the files themselves.

The move map layer up and move map layer down icons shift the selected layer up and down in the layer stack. When the map layers cover different regions, this has no effect on the map. However, when the layers are displayed over the same area, the stack order matters. The upper layers will obscure the lower layers. This has a variety of effects depending on the type of layer you are displaying. Closed polygon layers such as samcountries (the country shapes for Latin America) will obscure everything beneath them. Obviously line and point files sit happily on top of closed polygon files, but would be completely obscured if they were underneath.

☞ ***NOTE:** You can also drag and drop layers up or down with the cursor and mouse. However, this does not result in redrawing the image so you may not see a layer appear or disappear by this method until the map is redrawn.*

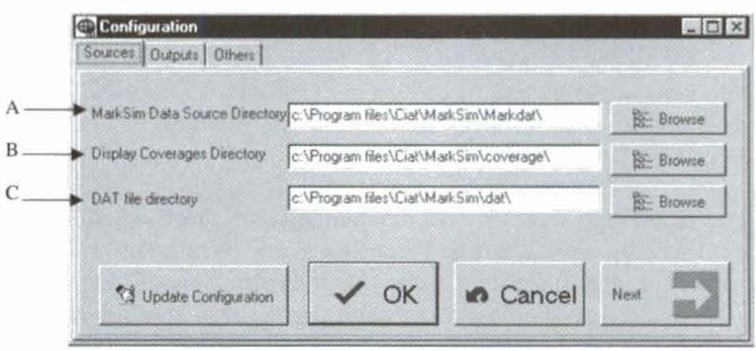
The erase a map layer icon does precisely that, and removes the selected layer from the map. It does not delete the shapefile file.

Load a map layer will load a shapefile. You will be cued to browse for the file to load. It can be any shapefile that is appropriate to the map and anywhere that is accessible to the application. Be careful that it is compatible with the layers you are displaying. For example, you cannot see two closed polygon layers at the same time. If you wish to display a closed polygon layer (i.e.; topography) below the country limits, you should use samboundaries and not samcountries.

Set map layer color will take you to a color selection menu.

Set map background color allows you to change the background color. This is usually the ocean, and hence blue, but it is in fact any area not covered by a loaded layer, so this is not always the case.

The configuration tool



Under the configuration icon are three screens. The first is the most important. This defines where the input files for MarkSim reside.

The file directories you see displayed here are as they will be loaded in the standard installation of MarkSim. If you have decided on another place to load MarkSim, then they will display the new directory site.

The MarkSim data source directory (A) contains all the files that define the MarkSim model, together with the interpolated climate surfaces that allow the model to interpolate to a given point. This directory is 362Mb. The contents are described in more detail in Appendix A. You may move the physical directory and files to another directory or disk unit if you wish, but if you do so, please note to update the configuration to denote its new address.

☞ ***NOTE:*** *The final \ on the directory address is mandatory. Without this, MarkSim will not recognize the path.*

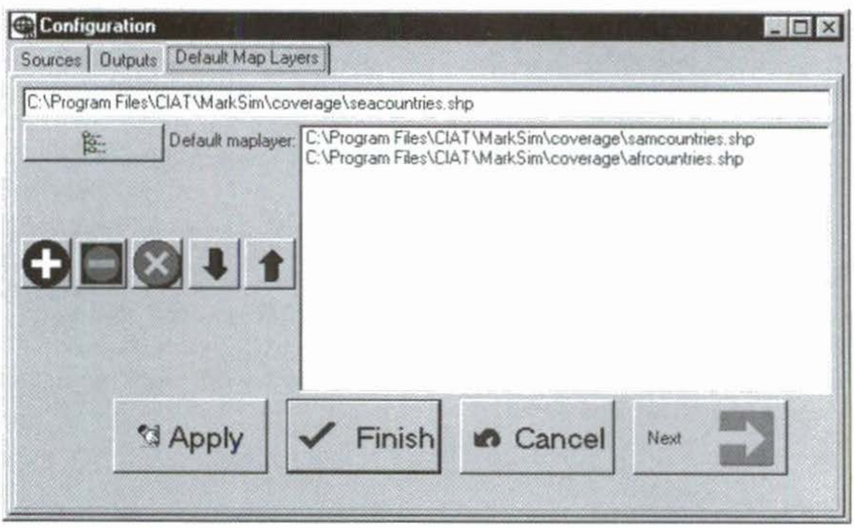
The display coverages directory (B) contains the background coverages that you can load to help you navigate the maps to find your sample points. These are ESRI shapefiles and consist of various subitems. These are explained in Appendix B. The size of this directory will depend on how many coverages were loaded with your version of MarkSim. These will be changing as we develop better backgrounds. For more up to date information, check on the MarkSim Web site and/or wait for notice on the listserver. For details of these, see the front of this manual. Again, you can move these to another directory, but remember to update the configuration. Other standard ESRI shapefiles can be loaded into MarkSim from this directory or others. The only restriction is that the projection and coordinates be geographic (i.e., latitude and longitude), that is to say, MarkSim will not accept shapefiles in other projections (UTM, Lambert, etc.).

User input files are read from the DAT file directory, which is loaded with the program files for your convenience and contains example files to get you started. We strongly recommend that you move the directory out of the C:\program files\...\ path, because it is not good practice to mix registered programs with users' data, even if Microsoft does so!

The definitions made in the configuration window are stored in the file MarkSim.INI that can be found in the c:\Program Files\MarkSim directory. They can be edited there using an ASCII editor (e.g., Notepad), but will not be applied until you leave MarkSim and restart the application. Changing them through the configuration window will make them valid for the current session.

The output directory should definitely be moved to another disk if at all possible. MarkSim can produce voluminous data files, which will often be used onwards in modeling applications. These are best kept separate from the disk partition used for the program files because filling the disk space may mean that your applications no longer run.

The default map layers define those shapefiles that you require, to be loaded automatically when you fire up MarkSim.



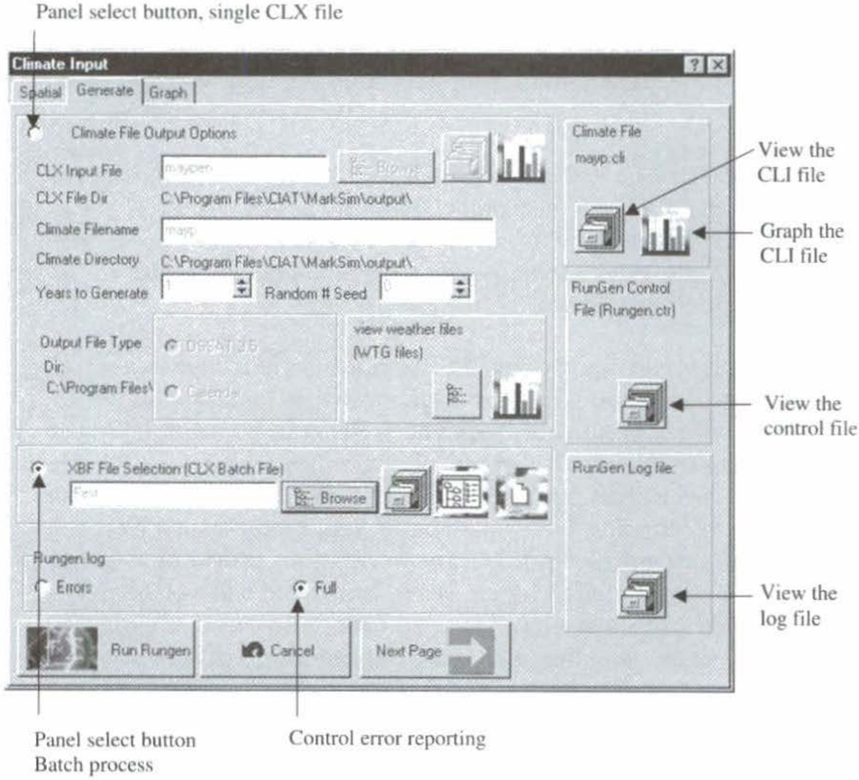
To compile a set for the default map, use the browse facility to open shapefiles into the top window. Use the plus button to add the latter to the default map. To delete a layer, select it with the arrow keys or by clicking on it in the list and using the minus button. You can clear the complete map and start from scratch with the X button.

☞ ***NOTE:*** *The default map is defined with the complete paths to the map layers. If you change the position of these files, then delete them, or do not have the CD-ROM in the reader if that is where they reside, MarkSim will not be able to complete the map.*

MarkSim can handle any ESRI shapefiles that are in geographic coordinates (latitude, longitude). It cannot handle projected layers, because it does not pretend to be a full GIS. The software is provided with a range of shapefiles to use as guides as to where you are in the areas selected. In many cases, these have far too much information for displaying at very small scales (i.e., continental or world levels). We are working on providing a range of products for use as you change scale, but these will not be fully implemented until a later version. You may, however, import any coverage you like as long as it is a shapefile in geographic coordinates.

The climate grids are not a complete match for the continental land coverages supplied with MarkSim. This is partly because the grids have square pixels, but also because some gaps occur where there are large lakes or wide rivers. It is therefore possible that when you ask for a point that appears to be on land you may get an error response saying that there are no data for that point. To make this explicit, the shapefiles of the climate grid pixels are supplied on the CD-ROM. You can use these to check the detail of what is available in the climate grid files. When you load them, the continent will turn black; however, if you select transparent fill in the layer properties window of the layers control tool and zoom in, you will see the pixel outlines appear.

The generate data tool



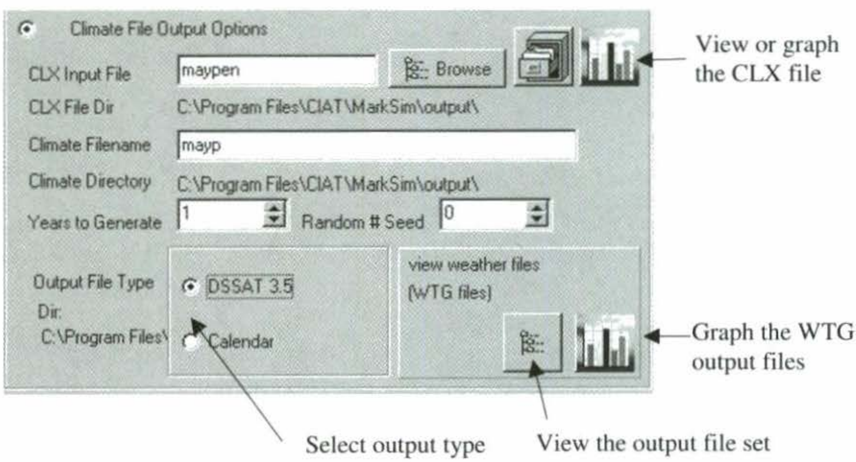
The generate data tool appears as the second page of the climate input window although the icon from the service menu will take you straight to it if you already have a CLX file constructed.

This tool takes the model parameters from the CLX file and uses them to simulate daily rainfall. If you choose the DSSAT 3.5 output option, then maximum and minimum temperatures and global radiation are also simulated. You need to specify how many years of data you would like simulated, a random number seed, and the output file type. If you do not specify a random number seed, then a seed will be calculated from the system clock. This will be reported in the log file so you can exactly duplicate the run at a later date by entering this seed. If you specify DSSAT 3.5 output, then you must also specify a site CLI filename. The CLI file is not used in this simulation, but is required for running DSSAT models so one is constructed for you if it does not exist. See Appendix B for the file structure.

As with the spatial data entry there are options for running a single CLX file or a batch run of many. Use the panel select button to choose between these options.

Single CLX file input

When running a single CLX, you enter the data required directly in the upper panel. Type in or browse for the filename of the CLX file from which you wish to run. If your output option is DSSAT 3.5, then you must specify a four-character DSSAT site name in the window under climate filename. This is used to name the CLI file and also the sequence of WTG output files.

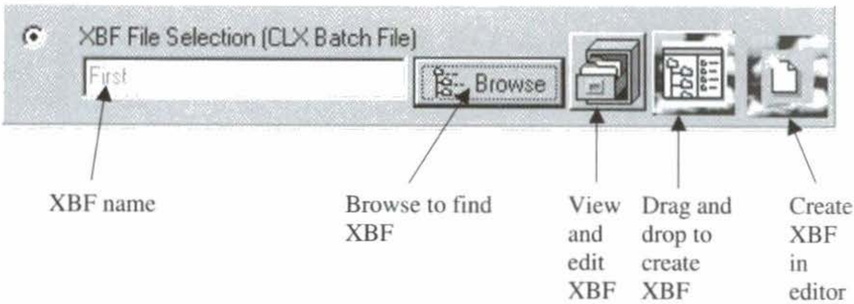


The view weather files panel is not activated until you have run the simulation. Once the run is complete you can look at the output file list. Select a file and you can inspect it with the TeeChart graphics.

Calendar-style output cannot be viewed graphically. It is a shorthand output of rainfall data in the format used to create the MarkSim models. Most users will not have much use for this output style. For the format see Appendix A.

Multiple CLX file input

A batch mode of operation is provided for the user who wishes to run many CLX files at once. This is the mode to use where multiple simulations will be run to cover a geographic region or to simulate many points such as a set of regional trials. The XBF is a list of CLX filenames with all of the data needed for each CLX file run.




```
BUENAVEN,BUEN,423,c
CALI,CALI,271,d
CLXFILE0,CLXF,0,c
CLXFILE2,CLXG,0,c
CLXFILE3,CLXH,0,c
MAYPEN,MAYP,8135,d
PALMIRA,PALM,231,d
PTAUPRINC,,6868,c
TULUA,TULU,8971,c
```

Example of an XBF

Output type

Random number seed

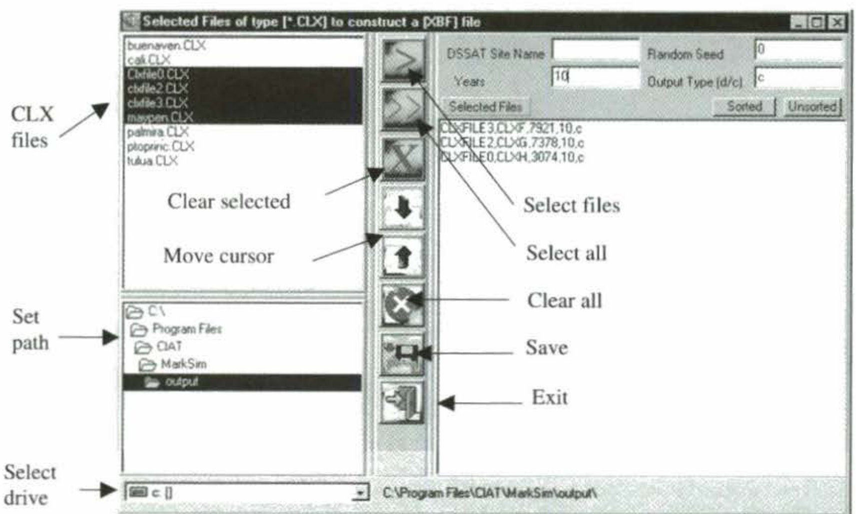
CLI filename

CLX filename

Here is an example of an XBF. It is a sequential, ASCII comma-delimited file that can be prepared in any ASCII editor, in the MarkSim editor, or by drag and drop from a list of CLX files in one or more directories. The first field is the CLX filename. This must start with an alphabetic character, contain no special characters, and be eight or less characters long. The next field is the DSSAT site name. This must start with an alphabetic character, contain no special characters, and be exactly four characters long. This field can be blank if calendar output is requested (see the case of Port au Prince in the eighth record). Next comes the random number seed. This must be an integer with four or less digits. It can be zero (see records 3, 4, and 5), in which case MarkSim will assign a random number seed calculated from the system clock and report it in the rungen.log file. The output type is "c" for MarkSim calendar style output or "d" for DSSAT 3.5 format output. You can mix types of output throughout the XBF.

If you have a large number of CLX files to run, the drag and drop feature will allow you to form the XBF with the minimum of effort.

✓ Click on the drag and drop icon and the following window will appear.



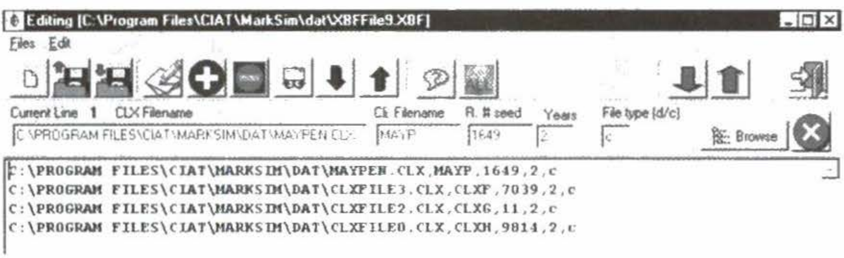
Select files individually, in groups, or all from the list of CLX files in the left hand window. The list that you see initially is of all the CLX files in the MarkSim default:output directory. To select from CLX files in other directories, change the path or drive in the windows provided. Each line of the XBF must contain sufficient information for the simulation run. This can be entered in the panel at the top right. If DSSAT output is required, a DSSAT site name will be needed. If you leave the option blank, the name will be taken from the first four characters of the CLX filename. Site names cannot be duplicated or there would be a confusion of CLI files. Therefore, if the first four characters of the CLX filename would cause duplication, the name is incremented alphanumerically. Thus ABCD becomes ABCE, XXX0 becomes XXX1, and ABZZ becomes AC00. If you enter a DSSAT site name in the space provided, this is used to name the first site in the XBF and each subsequent one is derived by alphanumeric incrementation.

The number of years to simulate is mandatory; there is no default and the creation of the XBF will not proceed until you do. The same number of years is used for every record in the file. The random number seed is optional. If you leave this field blank or zero, a seed is taken from the system clock independently for each record in the file. The output type is set to c or d for every record in the file. If you wish to change the number of years or output type for specific records in the XBF, you can proceed to editing it after you have exited from this window.

✓ Click on save to create and save the file.

You will see it fill out in the right hand window. There is an option above for sorting the records by CLX filename.

There is an editor for creating or customizing your XBF after you have created it with drag and drop.



✓ Use the green arrows to scroll through the file.

As you scroll, the current record appears with the fields selected in the editing windows. Or you can select a record for editing by merely clicking on it with the mouse. The drag and drop facility will have placed a random number seed, which is either constant through the file if you specified it, or one calculated from the system clock if you did not. The number of years will be the same number you specified throughout the file and likewise the output type will be constant. You now have the opportunity to change all those at will to tailor your XBF to exactly what you want. All of these options can vary from line to line as MarkSim interprets each line individually at run time.

You can also add or delete lines, or change their order with the blue arrow keys. These work by dragging the selected line up or down the file. You can search for extra CLX files to include with the browse button. Beware! When you open the selected file, the filename and path will be included in the editing workspace. However, the MarkSim editor does not know which site name, random number seed, years, or type of output you would like so it will leave these fields untouched. If you add the record without modifying these, they will contain the same information as the last selected record. This will be OK for the last three fields because they can be the same throughout the file, but the site name will be duplicated and will cause problems when you run the file with DSSAT output; the previous outputs with that name will be overwritten.

We strongly recommend that you use the validation functions before you save and run the file.

You can do this record by record with the icon 

Or you can validate the complete file with the icon 

3. Theory

In essence, there are two parts to MarkSim. One is a reliable stochastic rainfall generator to drive a weather simulation model. This is all very well when the user has the required parameters to generate synthetic weather records. But what about the situation (normal) when one does not? The second part of MarkSim is a set of surfaces of parameters that can be sampled by the user. More correctly, the parameters of the weather generator are not stored themselves, but rather an "intermediate" set of parameters is stored that can be used to reconstitute a full set of weather generator parameters. The reasons for this intermediate set of parameters are primarily to save space and to enhance efficiency. More details on the methods used in MarkSim can be found in Jones and Thornton (1993; 1997; 1999; and 2000). We summarize these below.

The Rainfall Model

Rainfall is modeled using a two-stage third order Markov chain. First, it is determined whether any particular day is wet; this depends on whether there was any rainfall on the 3 previous days. If so, then the amount of rainfall is determined.

Probability of a wet day

The probability of day i being wet is defined as:

$$P(W / D_1 D_2 D_3) = \Phi^{-1}(b_i + a_{i-1} d_1 + a_{i-2} d_2 + a_{i-3} d_3) \qquad (1)$$

where Φ^{-1} is the inverse of the normal probability (probit) function, b_i is the monthly baseline probit of a wet day following 3 consecutive dry days, a_m are binary coefficients for rain (1) or no rain (0) on day m, and d_m are lag constants. Thus, for example, the probability of a wet day following 3 dry days is $\Phi^{-1}(b_i)$, and the probability of a wet day following 3 wet days is $\Phi^{-1}(b_i + d_1 + d_2 + d_3)$. This part of the model is thus specified by 15 parameters: The baseline probabilities, b_i, derived for each month, and three lag constants, d_1, d_2, and d_3, which are unchanging from month to month.

The model uses a binomial error term and a probit link function. The occurrences of rain on day *i*-1, day *i*-2, and day *i*-3 are treated as the independent variables and the monthly total as another variable. This allows us to test the significance of the lag constants by using a chi-squared statistic. The results showed conclusively that a third-order Markov rainfall model was necessary, because the chi-squared statistic related to the inclusion of the third-order lag in the model was highly significant for 92% of the tropical locations that we have studied. This method of fitting the model also allowed us to test the significance of any interaction between the lag constants and the probabilities for the 12 months. Although certain data sets did show small interaction effects, this was generally not the rule, and it was concluded that under a probit transform the lag effects could be considered additive to the monthly effects (see Equation 1). The residual deviance, tested as a chi-squared statistic, was insignificant in almost all cases.

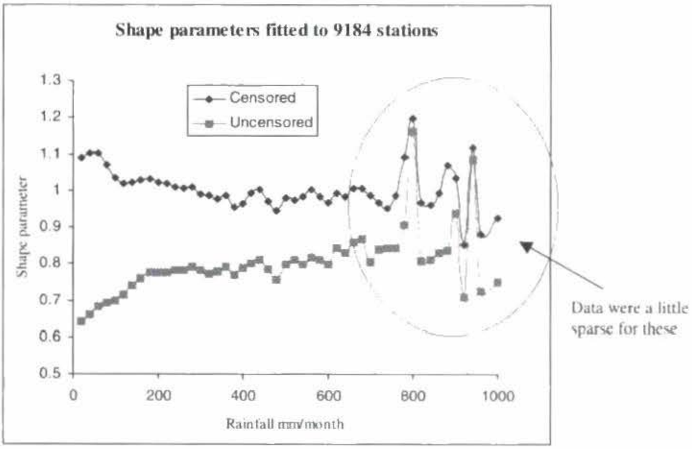
Rainfall on a wet day

Rainfall is modeled by using the censored gamma distribution, restricted below 1 mm, to determine daily rainfall amounts on those days that rainfall is experienced (Sterne and Coe, 1982). The method of maximum likelihood is used to estimate the mean and shape parameters of this distribution for each calendar month, thus giving rise to 24 additional model parameters.

The censoring of the gamma distribution means truncating the lower part of the distribution. This is especially important in the case of the gamma distribution, because if the shape parameter is low, then there is a large proportion of small values (small rainfall events). Differences in the rainfall measurements or reporting mean that these small events are reported differently in different data sets. Sterne and Coe (1982) used a censoring at 0.1 mm; all values including trace records were discarded. They used a series of data where measurements greater than 0.1 mm were all reported more or less the same. Unfortunately, the widely differing sets of data from all parts of the globe that we have used in MarkSim (almost 11,000 station records) means that there is different reporting, with the data not uncommonly being truncated below 1 mm. It is a great shame to lose the well-reported data that go below this level, but in the interests of consistency we had to eliminate them.

This is rather high for a censoring level and we were worried that it might have a large effect on the fitted gamma distribution models. We therefore took data from just over 9000 stations and fitted the gamma distribution to both censored and uncensored data.

The results showed clearly (see Figure 3.1) that, although there was not too much of a shift in mean rainfall size, there was indeed a large effect on the gamma shape parameter.



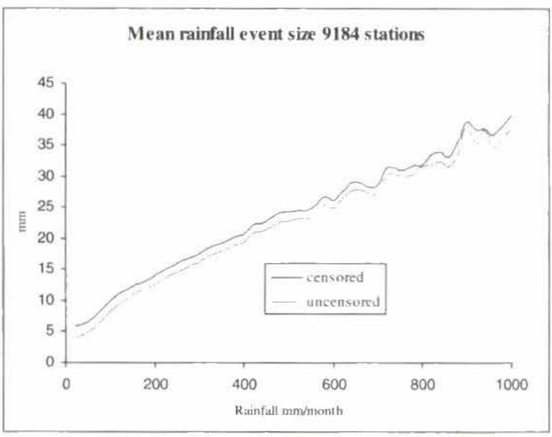


Figure 3.1. The effects of censoring on the gamma distribution parameters actual data from stations throughout the world.

We therefore needed a way to correct for the effect of censoring because we could not disregard it.

We ran 182 (14 x 13) Monte Carlo simulations producing 100,000 samples from each of the gamma populations on the intersections of the rectangular (lower) grid in Figure 3.2. We calculated the mean and shape factor for each simulation to check the sampling. We had to use a censoring to 0.000001 mm to avoid taking logs of very small numbers. For some of the populations, typically one sample in 100,000 was rejected because of this, and the sample parameters matched the population parameters within 0.001 for the shape parameter and about 0.02 for the mean. We then censored the sample data to 1 mm and recalculated the parameters.

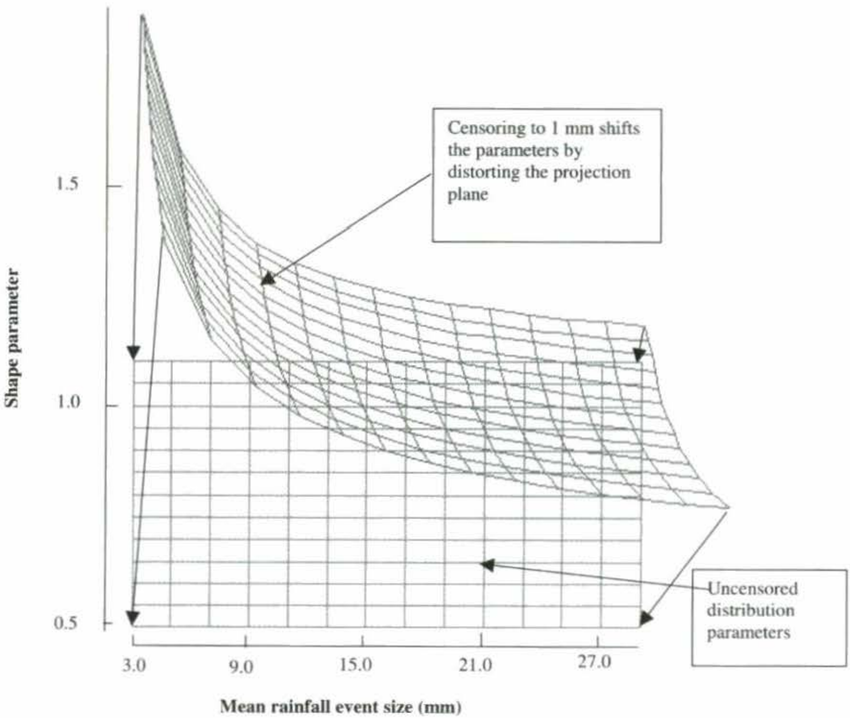


Figure 3.2. Distortion produced in the gamma distribution parameters by censoring to 1 mm.

The distorted (upper) grid shows the distortion introduced by the censoring. As can be seen, it is a monotonic distortion of the plane, like a map projection. The arrows at the corners show the movement needed at those points to undo the distortion. We can therefore correct for it by working out the projection functions.

We used Genstat to fit stepwise regressions to the complete set of 6th order polynomial variables. These are $x \dots x^6$, $y \dots y^6$ plus all cross products. The fitted functions are shown in Appendix B. Now, we know that by censoring the rainfall we have eliminated all the events with less than 1 mm rainfall so we have to adjust the frequency of events also (if we just use the corrections above, the overall amount of rain per month will fall in the model).

The answer is to divide the rainfall probability by the probability of gamma (*p*, *av*) exceeding 1 mm after reconstituting the probabilities from the probits.

Interpolating Back to Daily Data

In generating rainfall records, the monthly baseline probabilities (the probability of rain after no rain for 3 successive days) are interpolated to daily probabilities by using the 12-point Fourier transform described in Jones (1987). The lag effects are then added to each day's probit transform of the baseline probability to produce a matrix of 365 or 366 days by eight states (wet or dry conditions on 3 successive days). The inverse probit transform is then used to transform this matrix to normal probabilities. Similarly, the monthly mean and shape parameters of the gamma distribution of rainfall amounts are interpolated to daily values by using the 12-point Fourier transform.

Annual Variance and the Variability of Parameters

The parameters of the model, being simply estimates obtained from sometimes short data sets, have associated standard errors. To introduce sufficient variability into the model, any random sampling should be based on the uncertainty of the parameter estimates themselves. The 12 monthly baseline probabilities, b_i, are autocorrelated because of the yearly progression of weather, even in the tropics; thus, a resampling scheme must take these correlations into account. This is done by randomly sampling from a 12-variate normal distribution. The resampling scheme can be represented by:

$$b^*_i = s_i RN_i + b_i, i = 1,12 \qquad (2)$$

where b^*_i is the sampled value of b_i, the baseline probability of rain, s_i is the standard deviation of b_i, and RN_i is a random normal number. The resampling algorithm involves the Cholesky square root decomposition of the correlation matrix of monthly rainfall. The correct correlation matrix to use would be that of the baseline probabilities in their probit transform. In practice, however, this is difficult to calculate with short data sets. We thus assumed a surrogate correlation matrix and used the standard errors per year obtained in the original GLIM analysis multiplied by the square root of n-1, where n is the number of years.

The pseudo-random normal number generator of Marsaglia and Bray (1964) is used for rapid resampling of the 12 monthly baseline probabilities in their probit transform. The algorithm then adds in the lag constants and produces a new matrix of 365 or 366 days by eight states for each year for which rainfall records are required.

In the course of testing the model with random resampling, we found that it did not work well when the rainfall probabilities were very low. Subsequent analysis showed that the use of the probit transform produces a systematic bias. When resampling is used, low probabilities are overestimated and high probabilities are underestimated after retransformation. Simulations of completely random numbers were used to evaluate the empirical relationship of the standard error to the overall probability level. Probits produced from runs of up to 200 years were summed to monthly means and retransformed to probabilities. The variances of the retransformed monthly mean probabilities were then compared with the actual variances introduced in the simulations. The bias in the monthly probabilities was found to be related completely (explaining 100% of the variance) and simply, although empirically, to the probability level and the standard deviation. In the algorithm for the rainfall model with sampling, this relationship is used to correct the monthly baseline probabilities by adding to them the correction factor Di, defined as:

$$D_i = b_i (0.55228\, s_i^2 - 0.26154\, s_i^3) \qquad (3)$$

where for month i, b_i is the baseline probability of a wet day following 3 dry days, and s_i is the standard deviation of the baseline probability.

Simulating Temperatures and Solar Radiation

MarkSim uses the DSSAT weather generator (Pickering et al., 1994), based on routines of Richardson (1985) and Geng et al. (1988) to generate daily values of maximum and minimum temperatures based on whether the day is wet or dry. The parameters for generating these variables are the long-term monthly means stored in the CLX site file. The original code was part of the WGEN weather estimator (Richardson and Wright, 1984), and this was modified for DSSAT version 3 (Tsuji et al., 1994). The DSSAT modifications use standard deviations rather than coefficients of variation, which make the estimator more stable than the original version. If monthly climate parameters are used as input, the routines use a combination of the regression equations in SIMMETEO (Geng et al., 1988; Pickering et al., 1988) to compute the standard deviations.

Solar radiation data are generated from monthly mean values for daily solar radiation (or from sunshine hour means, if these exist in the CLI site file). MarkSim uses the routines in the DSSAT generator, which are again based on the equations in Geng et al. (1988) and Pickering et al. (1988). The monthly values of solar radiation are generated from the temperature normals using the model of Donatelli and Campbell (1997), which is a modification and improvement of the earlier model of Bristow and Campbell (1984). Briefly, this model calculates daily solar radiation at the earth's surface as the product of potential radiation and an estimate of the atmospheric solar radiation transmissivity coefficient (the ratio of the value of solar radiation outside the earth's atmosphere and its value at the earth's surface). Potential radiation outside the earth's atmosphere is estimated as a function of the declination, the half-day length, a factor accounting for the distance to the sun, the day of year, and the latitude. Potential solar radiation is then modified by the transmissivity to produce an estimate of radiation at the earth's surface. The transmissivity is estimated as a function of clear sky transmissivity, daily maximum and minimum air temperatures, and two empirical parameters.

The Climate Surfaces

Spatially interpolated climate surfaces are now available for many areas. These usually handle long-term climate normals interpolated over a DEM by various methods (Jones, 1991; Hutchinson, 1997). Pixel size depends on the underlying elevation model. It may be as little as 90 m (Jones, 1996), which results in a massive data set, or 10 minutes of arc (about 18 km), which is as large as is practicable in many instances. In the latter case, the normal elevation model is the NOAA TGPO006 (NOAA, 1984). We have produced interpolated data sets at CIAT using data from about 10,000 stations for Latin America, 7000 for Africa, and 4500 for Asia. Each set of surfaces consists of the monthly rainfall totals, monthly average temperatures, and monthly average diurnal temperature range. This makes 36 climate variates in three groups of 12.

We use a simple interpolation algorithm based on the inverse square of the distance between the station and the interpolated point. For each interpolated pixel we find the five nearest stations. Then the inverse distance weights are calculated and applied to each monthly value of the data type being interpolated. Thus, for five stations with data values *x* and distances from the pixel distance *d*:

$$x_{pixel} = \frac{1}{\sum_{i=1}^{5} d_i^{-2}} \times \sum_{i=1}^{5} \frac{x_i}{d_i^2} \qquad (4)$$

Temperature data are standardized to the elevation of the pixel in the DEM using a lapse rate model (Jones, 1991). Using this simple interpolation has various advantages. First, it is the fastest of all the common methods. Second, it puts the interpolated surface exactly through each station point, because the weight 1/(*d*(*I*)**2) becomes infinite as d approaches zero. Third, the interpolation is highly stable in areas of sparse data. It approaches the mean of the nearest stations while they all become equally distant. Fourth, it is relatively stable against errors in station elevation; only the local region of that station is affected. On the other hand, laplacian spline techniques and co-Kriging both propagate these errors more extensively. This is one advantage of using a proven lapse rate model instead of fitting a local one, as do both of these latter techniques.

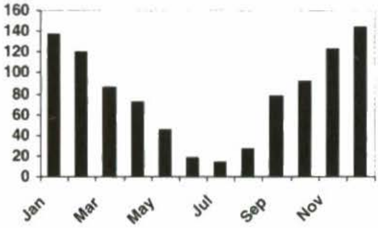
The method has two small disadvantages. First, the derivative of the surface becomes zero as it passes through the station point. In other words, each station is on a small plateau or step in the interpolated surface. This is usually much smaller than the pixel size and hence is not noticeable. Second, a (usually small) step occurs in the fitted surface as stations come into or drop out of the fitting window. Where the station density is high with respect to the pixel size, this is almost impossible to see. Where the stations are not so dense, it can produce unsightly straight lines or smooth arcs in the fitted rainfall data that are not tied to elevation. Inspection of the surface's profile usually shows that these are negligible artefacts, but they are unsightly and can undermine confidence in the surface maps.

Climate date standardization (rotation)

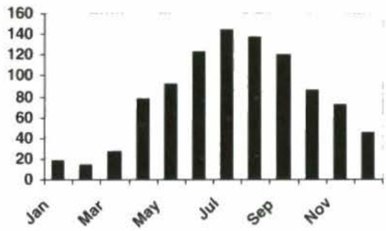
The climatic events that occur through the year, such as summer/ winter and start/finish of the rainy season, are of prime importance when comparing one climate with another. Unfortunately, they occur at different dates in many climate types. The most obvious case is where climates are compared between points in the Northern and Southern Hemispheres, but more subtle differences can be seen in climate event timing throughout the tropics. What we need is a method of eliminating these differences to allow us to make comparisons free of these annual timing effects.

Let us look at two hypothetical climate stations. They are in a typical Mediterranean climate—warm wet winters, hot dry summers. Northville could be somewhere in California, and Southville might be in Chile. The August rainfall in Southville is received in January in Northville (Figure 3.3). If we plot these rainfalls in polar coordinates, we can readily see that to compare them we need to rotate them to a standard time.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Northville	137	120	87	72	46	18	14	27	78	92	123	145
Southville	18	14	27	78	92	123	145	137	120	87	72	46



Northville monthly rainfall



Southville monthly rainfall

Figure 3.3. Monthly rainfalls for Northville and Southville.

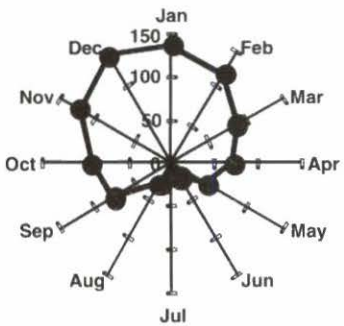
How do we do this automatically? The answer is the 12-point Fourier transform. This is fortunately the simplest of all the possible Fourier transform algorithms. It is highly computationally efficient and fast. In fact, it is the basis of nearly all Fast Fourier transform algorithms that break the problem down sequentially into the simple 12-point case. It takes the 12 monthly values and converts them to a series of sine and cosine functions. The one used in MarkSim has a modification to make it conserve the monthly total values (Jones, 1987). The equation produced is:

$$r = a_0 + \sum_{i=1}^{6} a_i \sin(ix) + b_i (ix) \tag{5}$$

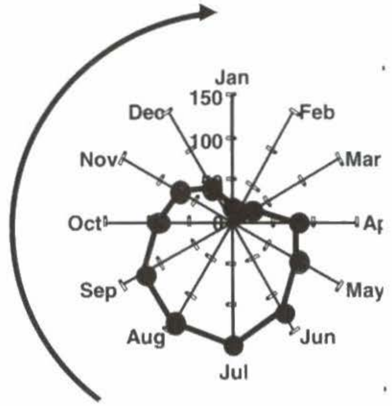
This can be rewritten as a series of frequency vectors, each with an amplitude α_i and a phase angle, θ_i:

$$\alpha_i = \sqrt{(a_i^2 + b_i^2)} \qquad \theta_i = \sin\left(\frac{b_i}{\alpha_i}\right) = \cos\left(\frac{a_i}{\alpha_i}\right) \tag{6}$$

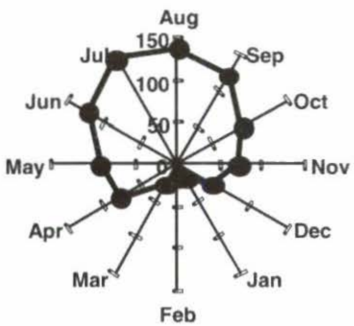
If we subtract the first phase angle from all the other vectors in the set, then we have produced a rigid rotation of the vectors. This is the rotation that we are seeking. It puts the maximum of the first frequency at a phase angle of zero and places the rest in positions equivalent to their angular separation in the original data. We then use the first phase angle for rainfall to rotate the data for temperature and diurnal temperature range, and these variates are rigidly rotated along with the rainfall.



Northville



Southville



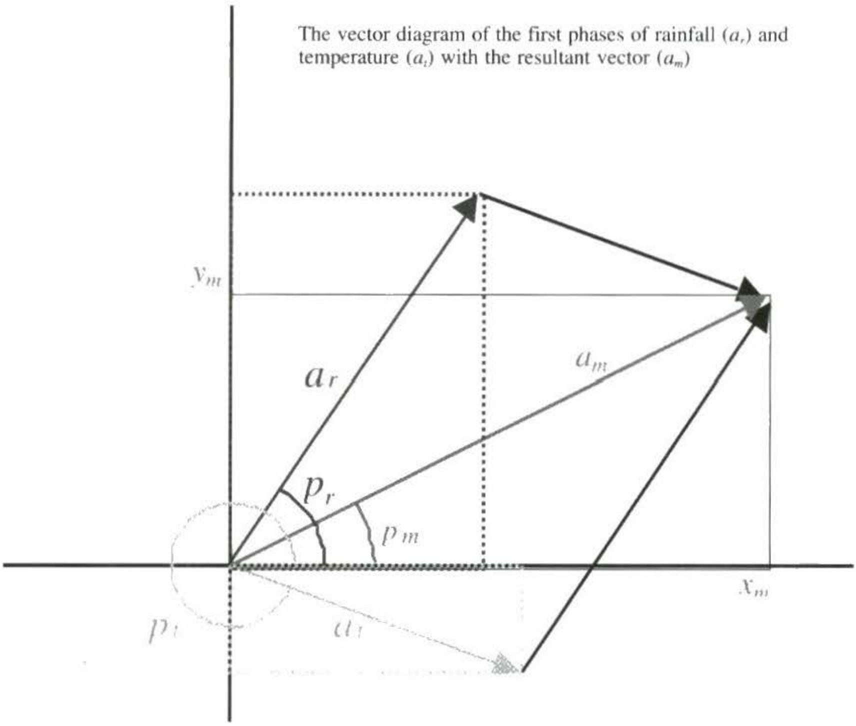
Southville rainfall rotated to coincide with timing of Northville

This explanation works well for the tropics. There was a small chance of the procedure going off the rails if the rainfall record did not have a seasonal peak. This was the case in some records from tropical desert regions, in these cases the rotation was ambiguous and sometimes resulted in pixels allocated to the wrong cluster.

The beta release of MarkSim went out with this type of rotation algorithm, as did the first release of FloraMap. When the climate grids of the latter were extended to Europe, the case arose where annual climate pattern was dominated by temperature and not rainfall.

We therefore have the possibility of rotating on rainfall or temperature, but when to decide which is the dominant? We tried many combinations of rules, but unfortunately came to the conclusion that none were acceptable. They all resulted in a hard line across the map at some point where the rotation basis changed. This led to climates that should have been grading imperceptibly from one type to another suddenly jumping at a discontinuity. This would have given the users serious problems when fitting models in these areas.

The best solution found is to use BOTH the rainfall and the temperature in calculating the rotation phase angle. Thus:



The resultant phase angle and amplitude are then:

$$y_m = a_r \cos p_r + a_t \cos p_t$$

$$x_m = a_r \sin p_r + a_t \sin p_t$$

$$a_m = \sqrt{y_m^2 + x_m^2} \tag{7}$$

$$p_m = angle\left({}^{x_m}\!/_{a_m}, {}^{y_m}\!/_{a_m} \right)$$

Unfortunately, this does not completely solve the problem of fitting a model to climates with different weather determinants. However, the vast majority of climates in the world are either:

(1) Rainfall determined where temperature is not an important seasonal effect (large areas of the tropics and subtropics);

(2) Temperature determined where rainfall is even throughout the year (most of the rest of the tropics and some temperate climates); or

(3) Rainfall and temperature determined when the two variates are highly correlated (summer rains—most of the rest of the world).

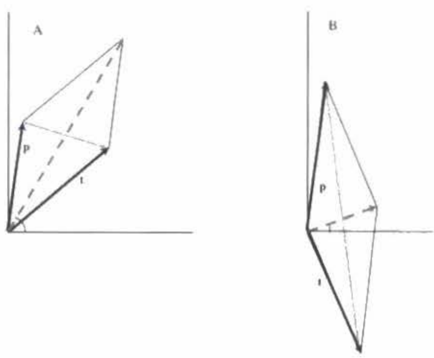
The Odd Man Out is:

(4) Winter rains and hot dry summers (almost only Mediterranean climates).

Luckily, the Mediterranean climates are at moderately high latitudes and we can afford to have the rotation dominated by temperature without losing generality in the rotations and comparisons. We therefore need to increase the weighting for the temperature vector smoothly as we approach the Mediterranean climates (in order to avoid a sudden swing).

The following weightings were found to work well:

$$\mathbf{p} = \text{rainfall mm}$$
$$\mathbf{t} = \text{temperature} \times 2 \times \text{abs(latitude)}$$



There is a potential trap when the two vectors almost cancel each other. This could result in wild swings of the rotation angle for small changes in the rainfall and temperature vectors. This becomes more likely as the situation passes from that in A (above) to B and beyond. The dashed arrows are the rotation vectors as before, but calculated on the weighted rainfall and temperature vectors.

Where the rotation vector is the vector sum **r** + **t**, the counter-diagonal vector is the difference **r** - **t**. It can be readily seen that the dangerous areas will be when **r** - **t** is much greater than **r** + **t**. We can therefore use a handy index of stability, *s*.

$$s = \arctan\left(\frac{|r-t|}{|r+t|}\right) \tag{8}$$

This will be zero for stable states where the rotation angle is dominated by rainfall, by temperature, or by both acting in concert. It will approach $\pi/2$ as the vectors tend towards canceling their effects. Because we can map this index, we can check for areas where this indeterminate rotation might occur. Areas of relatively high *s* (potential instability) occur on the US Pacific Coast, in Chile, northeastern Brazil, Sri Lanka, and through some areas of Central Africa. However, in no area does the index reach 80 degrees. Although this appears high, the phase angles are rotated correctly, and in fact there is little chance of a spurious rotation.

To save computing time, the whole climate surface is rotated according to these rules and all operations in MarkSim are done in the rotated phase space.

☞ ***The only exception to this is when the user requests a climate diagram for a climate surface point.***

Surface interpolation

As noted above, the rainfall model requires an extensive set of parameters: 12-monthly baseline probits (termed β) and monthly mean (**av**) and shape (**ps**) parameters for the rainfall event gamma distribution. Twelve monthly standard deviations and the 66 off-diagonal elements of the 12 x 12-correlation matrix for β are also required. Three lag parameters (**d**) allow us to calculate a 12 x 8 probit transition matrix.

Interpolated climate surfaces commonly hold only climatic normals for monthly rainfall and maximum and minimum temperatures. We therefore need some help to get 117 parameters from 36 monthly values. This help comes from the structure that is inherent in the Markov process and similarities in climate processes within climate types that, although not included explicitly in the model, affect the model parameters in consistent ways.

To produce the surfaces, the first step consisted of clustering the available historical station data. We used the rotated data in a two-pass leader cluster algorithm analysis. The first pass allocated stations as cluster leaders whenever they exceeded a minimum cluster distance. The second pass reallocated the stations to their respective cluster leaders. The distance measure was the Euclidean distance in the 36-dimensioned climate space. We tested various exponential transformations on the rainfall data and chose the exponent 0.5 (square root), based subjectively on the evenness of cluster sizes. Cluster sizes varied from 1 to 307 stations with a mean of 13.9 stations per cluster.

To calculate the expected parameter values of the model for any pixel in the interpolated climate surface, first we need to know to which cluster the pixel belongs and second, how the climate normals of the pixel adjust the parameter values within each cluster relative to the cluster mean values. We use the cluster seed as the type climate for each cluster and calculate the Euclidean distance in climate space for each pixel. The pixel is then associated with the closest cluster seed. This need not be geographically close. For each of the parameter types, we fitted a regression submodel within each cluster to trim the parameters estimated for the pixel to the best estimate we could make from the limited data recorded for each pixel of the climate surface. We dealt separately with two of the parameter types; rainfall event averages (**av**) and correlation matrices (see below).

Derivation of parameter estimates

The parameters for which we need regression submodels fall into two classes: **β**, **ps**, and **se** have 12 monthly values; the lag parameters **d** are single valued for each station or pixel. We therefore created two sets of independent variates for their estimation. The sets were derived from the basic station information and scaled as follows:

β, **ps**, and **se**
rm = monthly rainfall/200
tm = (monthly temperature - 15)/10
dm = (monthly diurnal temp. range -11)/4
srm = **sqrt**(monthly rainfall)/14
tmsq = **tm**^{2}
rmsq = **rm**^{2}
dmsq = **dm**^{2}
lat = station latitude /90
elev = (Ln(station elevation+10)-5)/3
sra = **sqrt**(**ra**)

$\mathbf{d}_1$ $\mathbf{d}_2$ $\mathbf{d}_3$
ra = annual rainfall/200
ta = (annual temperature -15)/10
da = (annual diurnal temp. range - 11)/4
rar = (annual range rainfall)/200
tar = (annual range temp. - 15)/10
dar = (annual range diurnal temp. -11)/4
rasq, **tasq**, **dasq** = **ra**^{2}, **ta**^{2}, **da**^{2}
rarsq, **tarsq**, **darsq** = **rar**^{2}, **tar**^{2}, **dar**^{2}
lat = station latitude /90

The scaling was designed to place regression parameter estimates within a reasonable range for the subsequent selection process.

We ran a five-stage stepwise regression for each cluster for **β**, **ps**, and **se** and a six-stage stepwise regression for the **d** lag parameters. Inspection of the results showed that correlations between the independent variates often resulted in large regression coefficients as a result of differential effects of the variates. Although the effects of fitting both terms were often statistically significant, their inclusion would have led to an undesirable instability of the regression as predictor when we present new data with slightly different values. Because we know the bounds of the clusters, we did not want a model predicting values outside these bounds. Inspection of each cluster for each of the parameters would have been far too time consuming. We therefore compiled a list of the independent variates ordered by the number of times that they occurred in each parameter set of parameter regressions. We then fitted the maximal model for each parameter and progressively eliminated variates until none showed a regression coefficient that would force a prediction out of the cluster bounds. Details of the regression analyses can be found in Jones and Thornton (1999).

Rainfall event averages

If we were to have fitted climate surfaces to rain days per month, the **av** parameters could be easily calculated as the monthly rainfall total divided by the rain days. Unfortunately, the main sources of monthly climate data used in the interpolated climate surfaces rarely contain the number of rain days. We therefore have to estimate these from the model. The probability coefficients used in the model are transition probabilities. They are the probability of the system passing from one triad state to another. The probabilities that we need to calculate the rain days per month are the state or stationary probabilities, which, except for the calibration stations, we do not have.

As a fortunate consequence of some structural redundancies in the model these can be calculated from the monthly average rainfall and the estimates of $\boldsymbol{\beta}$. As noted above, the model works in two parts: One decides whether today will be a rain day; the other decides how much rain should fall. The two parts have a hidden link. A triad is a binary form of three digits denoting rain on each of 3 days. Thus triad $\mathbf{t}$ = 101 means it rained yesterday, it did not rain the day before yesterday, but it did rain 3 days ago. Within the model, there are two classes of probability. One, the transition probability $\mathbf{p}$(t), shows the probability of rain today given that the system is in triad state $\mathbf{t}$. The other, the state probability $\mathbf{s}$(t), shows the probability of the system being in a certain triad state. The model calculates the transition probabilities as probits. Thus the transition probability for a given triad $\mathbf{t}$ in month $\mathbf{m}$ is:

$$P_{t,m} = \Phi^{-1}\left(\beta_m + \sum_{i=1}^{3} t_i d_i\right) \tag{9}$$

where $\boldsymbol{\Phi}^{-1}$ transforms from the probit form to a probability. We can write a transition matrix that governs the relationship between these two types of probabilities. Because we can calculate the $\mathbf{p}(t)$ from the equation above, we can use the transition matrix to calculate $\mathbf{s}(t)$.

$$S_{-1}\begin{pmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{pmatrix} \times \begin{pmatrix} 1-p_{000} & p_{000} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p_{001} & p_{001} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p_{010} & p_{010} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p_{011} & p_{011} \\ 1-p_{100} & p_{100} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p_{101} & p_{101} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p_{110} & p_{110} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p_{111} & p_{111} \end{pmatrix} = S_1^{\mathrm{T}} \begin{pmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{pmatrix}$$

Unfortunately, this matrix is singular. However, the frequency of $s_{110} = s_{011}$, and that of $s_{100} = s_{001}$. The proof of this is simple. Any rainfall sequence longer than 1 day must start with the triad 011 and finish with the triad 110. Thus, in any sequence, the frequencies must be equal if we discount a possible difference of one depending on the starting condition. That is to say, if the sequence starts with a rain period and finishes with a dry period, there will be exactly one more 110 than 011, irrespective of the length of the sequence. The same argument holds for triads 001 and 100 where dry days rather than rain days are counted. The state probabilities sum to unity as do the transition probabilities and the state outcomes. Adding alternative rows of the matrix eliminates four rows. We can therefore apply these restrictions by adding in four rows to the matrix. This then becomes positive definite and has a viable inverse.

$$\begin{pmatrix} 1 & 1 & 1 & 1+p_{011} & 1 & 1 & 1 & p_{111} \\ 1 & 1 & 1+p_{010} & 1 & 1 & 0 & 1+p_{100} & 1 \\ 1 & 1+p_{001} & 1 & 0 & 1 & 1+p_{101} & 1 & 1 \\ 1+p_{000} & 0 & 1 & 1 & 1+p_{100} & 1 & 1 & 1 \\ 2 & 2 & 2 & 3 & 2 & 2 & 1 & 2 \\ 2 & 2 & 3 & 2 & 1 & 1 & 3 & 2 \\ 2 & 3 & 1 & 1 & 2 & 3 & 2 & 2 \\ 2 & 1 & 2 & 2 & 3 & 2 & 2 & 2 \end{pmatrix} \times \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = S$$

We thus have a reliable algorithm to pass from transfer probabilities to state probabilities. Calculating the average rainfall event (**av**) now requires only the baseline probabilities, the lag parameters, and the monthly rainfall normals. The rain-day probabilities are found by summing s_{001}, s_{011}, s_{101} and s_{111} and are divided into the monthly rainfall normals. This eliminates 12 unwanted degrees of freedom and we have constrained the model to simulate actual long-term monthly rainfall normals.

Correlation matrices

As noted in Jones and Thornton (1997), we can see distinct patterns in the correlation matrices of many climate clusters. These patterns can, however, be highly complex. We therefore decided not to try to refine the estimate of the correlation matrices by fitting submodels within climate clusters, but to accept the correlation matrix calculated from the pooled variance/covariance matrices of the cluster members as being representative of all pixels allocated to that cluster.

References

Bristow, K.L.; Campbell, G.S. 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agric Forest Meteorol 31:159-166.

Donatelli, M.; Campbell, G.S. 1997. A simple model to estimate global solar radiation. PANDA Project, Subproject 1, Series 1, Paper 26. International Symposium on Computational Intelligence (ISCI), Bologna, IT. 3 p.

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, Crop Extension, Davis, CA, USA.

Hutchinson, M.F. 1997. ANUSPLIN Version 3.2 Users guide. The Australian National University, Centre for Resource and Environmental Studies, Canberra, AU. 39 p.

Jones, P.G. 1987. Current availability and deficiencies in data relevant to agro-ecological studies in the geographic area covered by the IARCS. *In*: Bunting, A.H. (ed.). Agricultural Environments. CAB International, Wallingford, GB. p. 69-83.

Jones, P.G. 1991. The CIAT Climate Database Version 3.41. Machine readable data set. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO.

Jones, P.G. 1996. Climate Database for Haiti. Machine readable data set. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO.

Jones, P.G.; Thornton, P.K. 1993. A rainfall generator for agricultural applications in the tropics. Agric Forest Meteorol 63:1-19.

Jones, P.G.; Thornton, P.K. 1997. Spatial and temporal variability of rainfall related to a third-order Markov model. Agric Forest Meteorol 86:127-138.

Jones, P.G.; Thornton, P.K. 1999. Fitting a third-order Markov rainfall model to interpolated climate surfaces. Agric Forest Meteorol 97:213-231.

Jones, P G.; Thornton, P.K. 2000. MarkSim: Software to generate daily weather data for Latin America and Africa. Agron J 92:445-453.

Marsaglia, G.; Bray, T.A. 1964. A convenient method for generating normal variables. SIAM Rev 6(3):260-264.

NOAA (National Oceanographic and Atmospheric Administration). 1984. TGP-OO6 D. Computer compatible tape. Boulder, CO, USA.

Pickering, N.B.; Stedinger, J.R.; Haith, D.A. 1988. Weather input for nonpoint-source pollution models. J Drain Eng 114(4):674-690.

Pickering, N.B.; Hansen, J.W.; Wells, C.M.; Chan, V.K.; Godwin, D.C. 1994. WeatherMan: A utility for managing and generating daily weather data. Agron J 86:332-337.

Richardson, C.W. 1985. Weather simulation for crop management models. Trans ASAE 28(5):1602-1606.

Richardson, C.W.; Wright, D.A. 1984. WGEN: A model for generating daily weather variables. United States Department of Agriculture (USDA), Agricultural Research Service, ARS-8, USA. 83 p.

Sterne, R.D.; Coe, R. 1982. The use of rainfall models in agricultural planning. Agric Meteorol 26:35-50.

Tsuji, G.Y.; Uehara, G.; Balas, S. (eds.). 1994. DSSAT Version 3. University of Hawaii, Honolulu, USA.

Appendix A

MarkSim File Structures

The MarkSim parameter file (CLX) is the heart of the MarkSim application. It holds the model parameters calculated in the first phase, clxgen, for transfer to the simulation phase rungen. It is also a critical file used in the construction of the model and hence holds some information that is not actually used in the operation of MarkSim. The file is fixed format and should never be edited by the user because there are complex relationships between the parameters. Do not succumb to the temptation to alter the climate data or model parameters, as the results can be unpredictable. If you wish to adjust the climate information, use the DAT file format described below. You can simply cut and paste the data records from the foot of the CLX file to construct a DAT file.

```
palmira  Interpolated  3.544  -76.306  1005
1.000  0.072 0.081 0.092 0.090 0.066 0.050 0.053 0.065 0.091 0.094 0.083
0.072  1.000 0.091 0.103 0.101 0.074 0.056 0.059 0.072 0.103 0.106 0.093
0.081  0.091 1.000 0.116 0.114 0.083 0.064 0.066 0.082 0.116 0.120 0.106
0.092  0.103 0.116 1.000 0.133 0.097 0.071 0.076 0.096 0.138 0.140 0.121
0.090  0.101 0.114 0.133 1.000 0.095 0.071 0.074 0.093 0.134 0.137 0.119
0.066  0.074 0.083 0.097 0.095 1.000 0.054 0.057 0.068 0.099 0.101 0.088
0.050  0.056 0.064 0.071 0.071 0.054 1.000 0.043 0.050 0.071 0.075 0.066
0.053  0.059 0.066 0.076 0.074 0.057 0.043 1.000 0.054 0.076 0.078 0.069
0.065  0.072 0.082 0.096 0.093 0.068 0.050 0.054 1.000 0.096 0.098 0.085
0.091  0.103 0.116 0.138 0.134 0.099 0.071 0.076 0.096 1.000 0.142 0.122
0.094  0.106 0.120 0.140 0.137 0.101 0.075 0.078 0.098 0.142 1.000 0.125
0.083  0.093 0.106 0.121 0.119 0.088 0.066 0.069 0.085 0.122 0.125 1.000
MONTH    AV    P     BETA    RAINDAYS    S.E.
   1    6.6  0.353  -0.736    0.316    0.26106
   2    6.7  0.341  -0.672    0.349    0.25922
   3    7.9  0.329  -0.500    0.436    0.25657
   4    8.7  0.334  -0.342    0.526    0.25183
   5    8.8  0.327  -0.439    0.468    0.25641
   6    7.3  0.369  -0.802    0.290    0.25440
   7    6.9  0.484  -1.176    0.150    0.23316
   8    6.1  0.391  -0.952    0.220    0.27601
   9    6.6  0.366  -0.771    0.305    0.26120
```

```
  10      9.0  0.329  -0.308     0.540      0.25395
  11      8.4  0.334  -0.518     0.429      0.25382
  12      7.6  0.344  -0.609     0.380      0.25737
D1-3 0.5150   0.1492   0.1073    N=   2 Cluster 132   Phase   0.452
rain   65.    71.  97.   142.  124.  66.   32.   41.   64.  144.  112.  86.
temp  23.7  24.0  24.1  23.8  23.5  23.5  23.7  23.7  23.7  23.2  23.2  23.5
rang  11.4  11.6  11.2  10.5  10.1  10.5  11.7  12.0  11.7  10.4   9.8  10.7
radn  18.5  19.5  19.2  18.2  18.0  18.9  19.3  19.7  18.6  17.3  17.7  18.2
```

The first line consists of an identifier, an indicator that it is an original data file or a model intermediate file (in this case it is the latter and is labeled interpolated), then the latitude, longitude, and elevation of the point. The matrix that follows is the correlation matrix for the baseline probits or Beta variates.

AV is the mean rainfall event amount and P is the gamma distribution shape parameter. Beta is the baseline probit. RAINDAYS is the average rain days per month expressed as a proportion of the days in the month, and S.E. is the standard error of the Beta value. D1, D2, and D3 are the lag parameters and, in an original data CLX, N is the number of years in the raw data. In an interpolated file it is always 2. Cluster is the cluster number associated with the interpolated point and Phase is the angle of rotation for season date standardization. Rain, temp, rang, and radn are the mean rainfall, daily temperature, diurnal temperature range, and solar radiation for the interpolated point. In an original data CLX file there is no record for solar radiation.

Climate definition file (DAT) is another file format used in the original calculation of the MarkSim model, but also used as a data entry format for the end user. The file extension originally stood for "data". Unfortunately, Mr. Gates has usurped it for use as a system file extension in Windows 2000. It allows a user to have complete control over the climate simulated rather than rely on the interpolated climate surface. If, for example, you wish to adjust the climate for a change in elevation within an interpolated pixel, copy the data from the CLX file, adjusting the elevation and temperature data by the standard lapse rate (subtract 6 degrees per 1000 m of elevation). Then resubmit the data as a DAT file. Alternatively, if you wish to enter the exact data from a known climate station, use the MarkSim editor or any ASCII editor to enter the data. The DAT file is a fixed format file conforming to the Fortran format (a8,2f8.3,i6,/12f5.0,/12f5.1,/12f5.1).

```
15353013  -15.700  35.183  1143
 253.  201.  159.   73.   15.   11.    9.    5.    4.   21.  103.  235.
 21.4  21.4  21.0  19.9  18.1  16.0  15.7  17.3  20.1  22.6  22.6  21.9
  7.8   7.8   8.0   8.8   9.8   9.6  10.0  10.9  11.8  11.8  10.3   8.2
```

MARKOV98.CTR is a MarkSim control file.

```
-2.588   -65.585   30     Brasil
-5.423   -64.774   30     brasil1
3.895    -77.073   60     buenaven
3.460    -76.525   1523   cali
```

The user should have little reason to encounter this file. It is used to communicate between the Pascal Delphi shell and the clxgen.dll. It consists of one or more lines in a fixed format with latitude, longitude, elevation and an identifier.

MARKSIM.CTR is a MarkSim control file.

```
C:\Program files\Ciat\MarkSim\Markdat\          Image directory
C:\Program Files\CIAT\MarkSim\output\           Output directory
C:\Program Files\CIAT\MarkSim\dat\              data file directory
GLFFile6.GLF                                    GLF name
Markov98.ctr                                    CTR file

3                                               mode of action
1                                               switch for verbosity
```

Likewise, the user should have little contact with this communication file. It varies a little depending on the action. It can be viewed from the spatial input window, and may be of use in debugging applications.

RUNGEN.CTR is a MarkSim control file.

```
CC:\Program files\Ciat\MarkSim\Markdat\         Image directory
C:\Program Files\CIAT\MarkSim\output\           Output directory
C:\Program Files\CIAT\MarkSim\dat\              Data file directory
XBFFile2                                        XBF name
CLIM                                            CLI filename
1
3
0
1
```

This file mediates between the Pascal Delphi shell and rungen.dll. It varies a little depending on the action. It can be viewed from the generate window and may be of use in debugging applications.

Georeference list file (GLF) contains a list of points with latitude, longitude, elevation, and a CLX filename. It is a sequential, ASCII comma-delimited file that is written by the user with a standard ASCII editor, the MarkSim editor, or may be constructed with the drag and drop facility on the spatial input window. It is used to specify a number of points for which a CLX file is to be produced.

```
-2.588 , -65.585, 30,    Brasil.CLX
-5.423 , -64.774, 30,    brasil1.CLX
 3.895 , -77.073, 60,    buenaven.CLX
 3.460 , -76.525, 1523,  cali.CLX
-4.890 , -64.774, 30,    Clxfile0.CLX
```

Climate batch file (CBF) specifies the full path to a set of DAT files for which CLX files are required. It is a sequential ASCII file that can be written by the user using a standard ASCII editor, or can be produced using the drag and drop facility on the spatial input window.

```
C:\Program Files\CIAT\MarkSim\dat\09333000.dat
C:\Program Files\CIAT\MarkSim\dat\15353013.dat
C:\Program Files\CIAT\MarkSim\dat\806082.dat
C:\Program Files\CIAT\MarkSim\dat\H1308001.dat
```

CBFs must always reside in the data directory.

CLX batch file (XBF) is a free format, comma-delimited sequential file that specifies a set of run orders for rungen. Each record gives the full path name of a CLX file, a DSSAT site name, random number seed, number of years to simulate, and output type. DSSAT site names must be unique. The random number seed is optional, but the field must exist as zero or null. Output types are "c" or "d".

```
C:\PROGRAM FILES\CIAT\MARKSIM\DAT\MAYPEN.CLX,MAYP,1649,2,c
C:\PROGRAM FILES\CIAT\MARKSIM\DAT\CLXFILE3.CLX,CLXF,7039,2,c
C:\PROGRAM FILES\CIAT\MARKSIM\DAT\CLXFILE2.CLX,CLXG,11,2,c
C:\PROGRAMFILES\CIAT\MARKSIM\DAT\CLXFILE0.CLX,CLXH,9814,2,c
```

The XBFs must always reside in the data file directory.

DSSAT climate definition file (CLI) is used by the DSSAT crop model driver. It is not used by MarkSim, but if it does not exist it is created by rungen.

```
*CLIMATE : PTOP
@ INSI   LAT     LONG   ELEV TAV  AMP   SRAY  TMXY   TMNY   RAIY
  PTOP 18.538-72.324  60  25.9 11.2 222.6 31.2   20.7   990
@ START  DURN   ANGA   ANGB  REFHT   WNDHT
    0     0    0.25   0.50  0.00    0.00
@  GSST  GSDU
     1    365

*MONTHLY AVERAGES
@  MONTH  SAMN  XAMN  NAMN  RTOT  RNUM  SHMN
     1    15.7  29.3  18.7  18.0  2.2  -99.0
     2    17.9  29.8  18.6  33.0  3.0  -99.0
     3    19.1  30.4  19.2  53.0   4.8 -99.0
     4    21.0  30.8  20.6 111.0  8.8  -99.0
```

The variate codes are as follows in order of appearance:

INSI	The DSSAT site name. In this case PTOP representing Port au Prince, Haiti
LAT	Latitude, decimal degrees, negative south
LONG	Longitude, decimal degrees, negative west
ELEV	Elevation, meters above sea level
TAV	Mean temperature, °C
AMP	Mean diurnal temperature range °C
SRAY	Solar radiation, yearly average, MJ m^{-2} day^{-1}
TMXY	Temperature maximum, yearly average, °C

TMNY	Temperature minimum, yearly average, °C
RAIY	Rainfall, yearly total, mm
START	Start of summary period for climate (CLI) files, Year *
DURN	Duration of summarization period for climate files, Years *
ANGA	Angstrom 'a' coefficient, yearly, unitless
ANGB	Angstrom 'b' coefficient, yearly, unitless
REFHT	Reference height for weather measurements, m *
WNDHT	Reference height for windspeed measurements, m *
GSST	Growing season start day, Day #
GSDU	Growing season duration, Days #
MONTH	Month number
SAMN	Solar radiation, all days, monthly average, MJ m^{-2} d^{-1}
XAMN	Temperature maximum, all days, monthly average, °C
NAMN	Temperature minimum, all days, monthly average, °C
RTOT	Rainfall total, mm $month^{-1}$
RNUM	Rainy days, # $month^{-1}$
SHMN	Daily sunshine duration, monthly average, percent *

* In this file missing data.

\# In this file dummy data.

Calendar format simulated rainfall file (GEN) consists of a header followed by 31 records for each year and a final trailer record at the end of file. The header is two records, the year number, filename, latitude, longitude, and elevation followed by a record of month labels. The trailer is similar to the first header record, but with the word END in the first three characters.

```
—0001 buenaven Interpolated  3.895  -77.073   60
   JAN FEB  MAR APR  MAY JUN JUL  AUG  SEP OCT  NOV DEC
1    7 118    0 109  244 253  60   42  230  32  207 124
2  191 500  286 193   23 100 171   14  207  67   79  16
3  226   0   95  21   22  30 252   61   78  70   33  59
4    1   0    5 495   47   8  50  145  641 164   53 120
5   94   0   50   0  120 181  60  202  284 196  142 120
6  227 338  158   0  227 407  94   86  131 105  154  32
7    0 116    6 120  340   0 140   37  262 151  443 189
8    0  61   50 170  336  11 132    0    0  47  105 313
9    0   0  329 243   36 153   0   88  198 109    0 259
10   0   0  204 190   83 209   0    0   91 374   28 256
11 252   0    0 313  106  93   0  284  179 118  122 125
12  22   0   19  56  173  37 154   18   75 396  257 110
13  88   0   30 437   54  70  52   74   57 172  370  98
14 300   0  245   0  900 237  50  150  229   0   52  75
15 166  65  100 108   50   0   4   88   41   0  931 131
16  76  22  167   2  278 186 258  133   28   0  247  70
17  40 248   98  12  351  37  89  124   35 136  167 158
18 120  74  121   0   99 130   1   58   37  72  127  36
19  47   0   24   0  181 128  12   80  198 357  107 166
20 264   0  290 201    6   0 135  453  188 255  102  39
21  44 108   95 550  275 150 174  402   47 287  208 358
22 253  81  515  26  156  47   0  124   44 356  128  18
23  36 201  131  86   95  37   0    0  474 172    0 239
24  10   2   34  72  128 364 103  272   22   0   31 365
25  70 136   63 346   53  61   0   21   54 657   73  58
26  49 217   85 176  105 310   0   71  528 130  336 134
27 169   0  198  86  173 103  29  128  299  13  124 168
28  45   0   29  67    0 219   0  167   19 171    0 252
29  28      352  64   33 109 362  135  307 123  167   0
30   3      116 227   20 160  11  155  552 327   13   0
31   0      147         0      67   34      52          0
END       buenaven Interpolated       3.895   -77.073   60
```

Each data record contains the day number and the rainfall values for that day in each month in integer format in tenths of millimeters. Missing days are blank.

DSSAT daily weather output (WTG)

```
*WEATHER : BUEN From Interpolated Surfaces

@ INSI   LAT    LONG   ELEV  TAV   AMP  REFHT   WNDHT
  BUEN 3.895-77.073  60   26.9 11.5 -99.0   -99.0

@DATE  SRAD  TMAX  TMIN  RAIN
01001  23.5  31.8  21.219.1
01002  14.7  32.5  23.821.4
01003  11.0  31.6  21.913.6
01004  10.2  32.2  21.7 2.4
01005  20.4  33.4  22.513.2
01006  15.9  33.6  22.114.8
01007  27.4  33.4  21.2 0.0
01008  27.5  29.3  20.437.6
```

One year's data constitutes one file. The DSSAT naming convention is Site name NN01, where NN is the year number. It is hence not possible to simulate more than 99 years for any site. The Site name is the same as the CLI filename. The header consists of six records, the title with site name, latitude, longitude, and elevation. TAV is the average temperature and AMP the monthly temperature amplitude. The reference height (REFHT) and wind measurement height (WNDHT) are always set to missing values in MarkSim generated files. The header is followed by 365 daily weather records (366 in leap years). The date field is year-day; the data are solar radiation, maximum and minimum temperatures, and rainfall in millimeters.

SHP, SHX, DBF, SBN, SBX coverage files are ESRI shapefiles that are provided with MarkSim to give background detail to the map displays. They are not used in MarkSim operations, but allow the user to identify features when looking for a particular place. There are files of country boundaries, roads, rivers, towns, contours, and municipal boundaries. Many of the coverages have been reworked from the Digital Chart of the World and are not complete. In addition, a set of shapefiles shows the grid cell bounds for the climate grids. Because the climate grids are quite coarse they cannot match coastlines exactly and in mountainous areas are only an approximation to the relief detail. Displaying the grid bounds on the map can help in choosing a site position, or explaining why in some cases the error message appears saying no climate data are available for a point that appears to be on land near a coastline.

Appendix B

Functions for Correcting the Censored Gamma Distribution

The functions pc and avc approximate a stable value for the gamma shape parameter and mean as calculated from daily rainfall data censored below 1 mm from the relationship with mean monthly rainfall (mm). Sdf give an estimate of the standard deviation of the betas. These functions are used as a check for the validity of the censored values.

NOTE: As m tends to 0, pc tends to 35.31; this leaves reality behind by quite a long way. Do not expect a reliable estimate below m = 1, whereas avc is stable right down to m = 0.

```
real function pc(m)
real m,q,x
x = m/1000
q = (m/10)+0.02
pc = 1.07707-(0.3756-0.3175*x)*x+(0.01291+0.013435/q)/q
return
end

real function avc(m)
real m,x
x = m/1000
avc= 5.967 + 39.71*x - 7.12*x**2
return
end

real function sdf(m)
real m
if(m.eq.0) m = 1
sdf = 0.22757+0.02638*m/500+(1.4238-(1.057-0.503/m)/m)/m
sdf = amin0(sdf,1.0977)
return
end
```

The functions pu and avu give the uncensored gamma shape parameter estimated from the monthly rainfall calculated from the censored rain data. The function pu is valid only for m.ge. 1.0. Below this value the function has no meaning because m is log transformed. However, monthly rainfalls below 1 mm may exist (in fact they do not in the censored datasets). It is truncated below 1.125 where it starts to climb to an infinite limit. These functions are used to produce a stable estimate for the uncensored values when the parameters are out of range for the better correction functions gamma_av and gamma_p.

```
real function pu(rain)
real rain
if(rain.lt.1.125) rain = 1.125
r = log(log(rain))
pu = 1.2969 - (.1009 + .009*r)/(1 -(1.2264 -0.4363*r)*r)
return
end

real function avu(rain)
real rain
r = log(rain)
avu = 7.99 + (1.045*r-4.78)/(1 -(0.2389 -0.01541*r)*r)
return
end
```

The functions gamma_av and gamma_p correct the mean and shape parameter for a fitted gamma distribution when the samples are censored below the value of 1. They are for use with rainfall event values in the range 2 to 40 mm and gamma distribution shape parameters 0.3 to 2.5. Because the input parameters are distorted from these uncensored limits, the function inside checks for valid inputs. This functional fit breaks down very fast outside its fitted area. These functions are fitted by stepwise regression in GENSTAT from all powers and cross products of the independent variates to the sixth power for a and p, to the fourth power for ai and pi. They are based on a Monte Carlo simulation of 100,000 samples for each of 14 by 13 points in the range.

For gamma_av, the fit gives Abs Max residual 0.184, Standard Deviation 0.03576.

```
real function gamma_av(average,shape,error)
logical     error,gamma_limit
real      average,shape,a,p,ai,pi

error = .false.
if(.not.gamma_limit(average,shape)) then
    error = .true.
    return
end if
a = average/40
p = shape/2.75
ai = .1/a
pi = .1/p
gamma_av = 0.0119698+1.000303*a+(-0.2358158+(2.973477 &
-11.64334*pi)*pi+0.3975927*ai*ai+(+(-22.32169 &
-23.56441*ai*ai)*pi-1.026232*ai*ai*ai)*pi*ai)*pi*pi
gamma_av = gamma_av *40
end
```

For gamma_p, the function was not fitting at all well. Taking the whole range I have split the fit, parting the data file at average = 9.

```
Abs Max residual. 0.0226 fit to lower part,
Standard Deviation 0.00225

Abs Max residual 0.0110 fit to upper part,
Standard Deviation 0.00125
```

```
real function gamma_p(average,shape,error)
logical     error,gamma_limit
real      average,shape,a,p,ai,pi,p2,p3,p4,a5

error = .false.
if(.not.gamma_limit(average,shape)) then
    error = .true.
    return
end if
a = average/40
p = shape/2.75
ai = .1/a
pi = .1/p
```

```
p2 = pi*pi
p3 = p2*pi
p4 = p3*pi
a5 = a**5
if(average.le.9.0) then
  gamma_p = 0.6707273+(26.57797*p4+(1.319750- 6.289515*pi &
+(-0.5363049-30.09750*p3+(0.0588292                      &
+201.5118*p4)*ai)*ai)*ai)*ai+(-6.980662+20.77201*p2)*p2  &
+(-0.3990897+(0.5076262+(2.686214*a+(-0.9675789*a        &
-236.1536*a5*p)*p*p)*p)*p)*p
 else
  gamma_p = -0.0179229+(-0.6905722+1.141806*pi)*p3+
(-7.837731*p2                                             &
+(158.0031*p4-35.21537*p3+(-1.737783*pi                   &
+0.4967813*ai)*ai)*ai)*ai+(1.179822+(-0.2637289*a         &
+(0.2545756*a*a*a+(-0.1941607*a5+(0.0637321*a*a5          &
+0.0246295*p)*p)*p)*p)*p)*p
   end if
gamma_p = gamma_p * 2.75
return
end
```

MarkSim operational function determines if a censored av,shape parameter pair falls within the competence area of the correction functions gamma_av and gamma_p. The boundaries coincide with the uncensored limits 2 < av > 40 mm, 0.3 < shape > 2.5. These are the bounds of the fit for gamm_av and gamma_p, the fit of which is highly unreliable outside these limits. These boundary functions are calculated from the boundary points of the Monte Carlo function fitting set. They therefore hold for slightly more or less than their mean fitted curve. Hence, the small adjustments after each limit is calculated.

```
logical function gamma_limit(av,shape)
real        a,p,x,av,shape
gamma_limit = .true.

c        First screen - equations may be out of range
if(shape.lt.0.604.or.shape.gt.3.48) then
   gamma_limit = .false.
      return
end if
(av.lt.1.9.or.av.gt.54) then
```

```
    gamma_limit = .false.
    return
end if

x = av/40
p = 3*(0.0633 - (13.0 + 80.0*x)/(1-(522.0 + 96.2*x)*x))   !left
p = p-0.005
if(shape.lt.p) then
    gamma_limit = .false.
    return
end if
x = (av/40)**2
p = 3*(1.068 - (0.675 - 250.0*x)/(1-(858.9+126.0*x)*x))   !right
p = p+.05
if(shape.gt.p) then
    gamma_limit = .false.
    return
end if
x = shape/3
a=40*(1.00390+(0.393-0.999*x)/(1-(14.4 -61.3*x)*x))        !top
a = a+0.05
if(av.gt.a) then
    gamma_limit = .false.
    return
end if
if(shape.lt.1.35) return ! Lower limit of bottom. If not failed
here then fit is good
a=40*(0.026763-(.10448-0.08326*x)/(1-(5.9241-4.091*x)*x))!bottom
a = a-0.05
if(av.lt.a) then
    gamma_limit = .false.
    return
end if
return
end
```

Correcting the transition probabilities for censoring requires that we know the relationship between the transition probabilities and the rainfall distribution. The transition probabilities are assumed to be distributed in probit transform with a binomial error. The rain event size distribution is quite different; it is a gamma distribution. Nonetheless, they are inextricably linked; in fact the probabilities of individual rain event sizes sum to the probability of rain. However, we have censored below a value of 1 mm so we can calculate the number of rain events that we have lost directly from the gamma distribution. Then we can add that back in to the rainfall event probability to correct for the loss.

For any given day i let the probability that it will rain be p_i. Then the probability that it will rain given the system state (001) can be written $p_{i|001}$. If the gamma mean and shape parameters for the day are α_i and γ_i and the gamma probability function of observing a rainfall event greater than x is gamma(α,γ,x), then the probability of observing a rainfall event less than 1 mm in system state 001 is $p_{i|001}$ times 1-gamma(α_i,γ_i,1) and the corrected value is $p_{i|001}$(2-gamma(α_i,γ_i,1)).

Index

L

M

N

O

P

R

S

T

LICENSE AGREEMENT

READ THE TERMS AND CONDITIONS OF THIS LICENSE AGREEMENT CAREFULLY BEFORE OPENING THE PACKAGE CONTAINING THE PROGRAM DISKETTES. BY OPENING THE PACKAGE CONTAINING THE PROGRAM, YOU ARE ACCEPTING AND AGREEING TO THE TERMS OF THIS LICENSE AGREEMENT. IF YOU ARE NOT WILLING TO BE BOUND BY THE TERMS OF THIS LICENSE AGREEMENT, YOU SHOULD PROMPTLY RETURN THE PACKAGE IN UNOPENED FORM, AND YOU WILL RECEIVE A REFUND OF YOUR MONEY.

LICENSE GRANT

The enclosed software is a proprietary product of CIAT, and protected under US copyright law. The software may be used only on computers owned, leased, or otherwise controlled by you. Neither concurrent use on two or more computers nor use in a local area network or other network is permitted without separate authorization and the payment of other license fees. You agree that you will not assign, sublicense, transfer, pledge, lease, rent, or share your rights under this License Agreement. You agree that you may not reverse assemble, reverse compile, or otherwise translate the software.

Upon loading the software into your computer, you may retain the program CD-ROM for backup purposes. In addition, you may make one copy of the software for the purpose of backup in the event the program CD-ROM is damaged or destroyed. You may not copy the user manual or parts of the user manual. Except as authorized under this paragraph, no copies of the software or the user manual may be made by you or any person under your authority or control.

LIMITED WARRANTY

CIAT MAKES NO WARRANTIES, INCLUDING WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. ANY REMEDY UNDER THIS WARRANTY IS LIMITED TO RETURN OF THE SOFTWARE AND USER MANUAL TO THE DEALER OR TO CIAT FOR REPLACEMENT OR REFUND. THIS WARRANTY IS VALID FOR 30 DAYS FROM YOUR DATE OF PURCHASE. CIAT EXCLUDES ANY WARRANTY COVERAGE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES. SOME STATES DO NOT ALLOW LIMITATIONS ON IMPLIED WARRANTIES, SO SOME OF THE ABOVE LIMITATIONS MAY NOT APPLY TO YOU.