USGS
science for a changing world

Prepared in Cooperation with
Earth Surface Dynamics Program—Effects of Climate Variability on American Drylands Project

USER'S GUIDE

Software for Reduction and Analysis of Daily Weather and Surface-Water Data:

Tools for Time Series Analysis of Precipitation, Temperature, and Streamflow Data

By Richard Hereford

2006

Open-File Report 2006–1101

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Dirk Kempthorne, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia 2006

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Hereford, Richard, 2006, User's guide, software for reduction and analysis of daily weather and surface-water data: Tools for time series analysis of precipitation, temperature, and streamflow data: U.S. Geological Survey Open-File Report 2006-1101 [available on the World Wide Web at URL http://pubs.usgs.gov/of/2006/1101/].

Contents

Figures

Tables

USER'S GUIDE

Software for Reduction and Analysis of Daily Weather and Surface-Water Data

By Richard Hereford[1]

Introduction

The software described here is used to process and analyze daily weather and surface-water data. The programs are refinements of earlier versions that include minor corrections and routines to calculate frequencies above a threshold on an annual or seasonal basis. Earlier versions of this software were used successfully to analyze historical precipitation patterns of the Mojave Desert and the southern Colorado Plateau regions, ecosystem response to climate variation, and variation of sediment-runoff frequency related to climate (Hereford and others, 2003; 2004; in press; Griffiths and others, 2006).

The main program described here (Day_Cli_Ann_v5.3) uses daily data to develop a time series of various statistics for a user specified accounting period such as a year or season. The statistics include averages and totals, but the emphasis is on the frequency of occurrence in days of relatively rare weather or runoff events. These statistics are indices of climate variation; for a discussion of climate indices, see the Climate Research Unit website of the University of East Anglia (*http://www.cru.uea.ac.uk/projects/stardex/)* and the Climate Change Indices web site (*http://cccma.seos.uvic.ca/ETCCDMI/indices.html*). Specifically, the indices computed with this software are the frequency of high intensity 24-hour rainfall, unusually warm temperature, and unusually high runoff. These rare, or extreme events, are those greater than the 90th percentile of precipitation, streamflow, or temperature computed for the period of record of weather or gaging stations. If they cluster in time over several decades, extreme events may produce detectable change in the physical landscape and ecosystem of a given region.

Although the software has been tested on a variety of data, as with any software, the user should carefully evaluate the results with their data. The programs were designed for the range of precipitation, temperature, and streamflow measurements expected in the semiarid Southwest United States. The user is encouraged to review the examples provided with the software.

The software is written in Fortran 90 with Fortran 95 extensions and was compiled with the *Digital Visual Fortran* compiler version 6.6. The executables run on *Windows 2000* and *XP*, and they operate in a MS-DOS console window that has only very simple graphical options such as font size and color, background color, and size of the window. Error trapping was not written into the programs. Typically, when an error occurs, the console window closes without a message.

[1] Flagstaff Science Center, Arizona

Input data format

Sources of data and measurement units

Daily weather data, referred to as the NCDC Cooperative Station Data, are commercially available from *EarthInfo*, *Inc*. (*http://www.earthinfo.com*) Boulder, CO or the *National Climate Data Center* (NCDC, *http://www.ncdc.gov/oa/ncdc.html*) Asheville, NC. These data consist of precipitation, snowfall, and maximum and minimum temperature. The weather data is explained in Dataset 3200 at the NCDC website (*http://www.ncdc.noaa.gov/oa/documentlibrary/surface-doc.html*). The NCDC data have been processed with an automated quality control system described by Reek and others (1992). Surface-water data expressed as average daily discharge can be obtained without cost from the U.S. Geological Survey Surface-Water Website (*http://water.usgs.gov/nwis/sw/*).

Data from these sources are in English units of measurement. Precipitation and snowfall are in hundredths (HI) and tenths of an inch (TI), respectively. Temperature is in degrees Fahrenheit and streamflow is in cubic feet per second ($ft^3 s^{-1}$). The software processes the data in English units. After processing, the output can be converted to inches or metric units.

Data fields

The daily data must have a two-column format written as an ASCII text file. The first column contains the date field and the second column, following a space, contains the numeric data field. The date format is a 10-character string MM/DD/YYYY that must include leading zeroes. The numeric field can be up to 15 numeric characters in length. Non-numeric characters are not allowed in the numeric data field with the exception of "T", which stands for trace (< 0.005 in) of precipitation. The programs automatically replace "T" with zero. A text editor can be used to remove other unwanted text characters that occasionally occur as flags in weather and streamflow data. Precipitation, snowfall, and temperature are integers, whereas, the streamflow data are real (floating point) numbers. The missing value code is -99. Numerous examples of this data format are provided in the **Example** subdirectory of the SetUpTable_v1.1 program.

The daily time series file should begin on the first day (01/01) of the first year and end on the last day (12/31) of the last year. If the program does not execute correctly, it may be necessary to complete the daily time series by padding the beginning and end of the file with the dates and the missing value code (-99). The utility program Cal_Date_Str can be used to generate the dates and a missing value code for placement in the input file.

File names

File names must contain "qw", "ppt", "snw", "tmax", or "tmin" (do not use quotation marks) to identify the data type (e.g., bakerppt.txt). These abbreviations are used by the software to determine the appropriate computational procedure. File names cannot exceed 24 characters including the ".txt" extension. The path and file name together cannot exceed 120 characters. Duplicate file names are not allowed and all the necessary files must be present in the data directory. The programs will not overwrite existing files and they do not prompt the user for missing files. If an error occurs while reading or writing a file, the console window will close without an error message or return to the first prompt.

Batch-file format

Most of the programs have the option to process one file at a time or to process multiple files, which is called batch processing. Batch processing is done with a batch file, a text file that contains all

the parameters necessary to operate the program. Batch processing provides a means to rapidly process numerous input files. The batch file can be troublesome to create, but it can be reused with different accounting periods and for other programs. Each line of the batch file contains the input file name, output file name, first and last date of the *beginning* sample interval, percent missing, and plotting position of the first interval. The user determines the acceptable number of missing values for the accounting period, which is expressed as a percent of the total days in the period. The file format is described below and examples are given in the **Example** subdirectories.

Columnar format of the batch processing file

Column No(s)., Descriptor

1–24, Input file name
25, Space
26–49, Output file name
50, Space
51–60, Beginning date of the *first* season or year
61, Space
62–71, Ending date of the *first* season or year
72, Space
73–74, Number of missing values per interval, in percent
75, Space
76–79, Plotting position of first interval (year of the first accounting period)

Example format

```
independenceall.ppt      independppt.txt          01/01/1898 01/31/1898 10 1898
greenlandrchtmin.txt     greenlandtmin.txt        01/01/1912 12/31/1912  5 1912
indioppt.txt             indioppt_WY.txt          10/01/1927 09/30/1928 10 1928
```

A simple and accurate way to make the list of input files utilizes the DOS DIR command. Open the appropriate directory and from the DOS Command Prompt type

DIR *.txt /b > drive:\ …directory\FILES.txt.

The file names will be listed in FILES.txt. Check to be sure that the file names begin on the first line and first column of the batch file.

Another way to make the input file name column that avoids the DOS DIR command (a real advantage in some cases) is to use MS Word (*PC Magazine*, May 24, 2005, p. 70):

In *Windows Explorer*, select the files (include only text files output with NCDC_v4 or those in the two-column format described above) and drag and drop the file names onto a word document using the right mouse button. Select the option *Create Hyperlink Here*. After this, select all the resulting hyperlinks, press Ctrl-x to cut them to the clipboard, and select *Edit|Paste Special* from the menu. In the list of formats, select *Unformatted Text*, then click OK. At this point each file name is preceded by the full path, which will not be accepted by the program. To remove the path, highlight the path name of any file including the final back slash, and press Ctrl-c to copy it to the clipboard. Highlight all of the files and press Ctrl-h to bring up the *Find and Replace* dialog. Paste the path into the *Find what* box, and enter ^p in the *Replace with* box. Click on *Replace all* and the file names will be listed without the path information. Check the listing for any extraneous formatting and make sure the names begin on the first line, then save the file as *Text Only*. To complete the batch file, add to each line the output file name, beginning and ending date of the first accounting period, number of missing values in percent, and the plotting position of the first accounting period.

The software

After the data are assembled, the work flow proceeds as follows: 1) The raw NCDC files are processed with NCDC_v4, or with surface-water data, the file header information is removed with a word processor. 2) The processed files in two-column format can be analyzed with Annual_Cycle_v2 to search for seasonal patterns in precipitation or streamflow. 3) The daily time series output in step 1 are then processed with Day_Cli_Ann_v5.3, which produces seasonal or annual statistical summaries of the daily precipitation or streamflow data. 4) If multiple stations were processed, the resulting files can be assembled into a single multi-column table using SetUpTable_v1.1, which simplifies further analysis with statistical software.

NCDC_v4

This program, as mentioned above, is the first step in analyzing daily weather data. The program is used to reformat raw NCDC CR/LF (carriage return/line feed) data files that are the input to Day_Cli_Ann_v5.3, Month_from_Day_v2.2, Annual_Cycle_v2, and SetUpDailyTable. The NCDC CR/LF format is one of several format options available from data vendors. The other formats are not supported by this software. The NCDC CR/LF format is almost visually incomprehensible. Each line has 403 characters that contain the identifying number of the weather station, the type of data, the units of measurement, the year, the month, the day of the month, and one month of daily data. The output from NCDC_v4 is the two-column time series described above. The NCDC file protocol excludes long runs of missing data that occur, for example, when a weather station was inoperative for several years. NCDC_v4 checks for absent years and fills in the series with the missing dates and the missing value code.

Operation—The program processes one file at a time. First, the user is prompted to enter the name and complete path of the input file. This is a NCDC CR/LF text file with the extension ".crd" when the file was output with the *EarthInfo, Inc.* software. *It is not necessary to type the path and file name in this program or any of the others.* Typing is easily avoided by dragging the input file name from *Windows Explorer* and dropping it on the console prompt. At the second prompt, requesting the output path and file name, repeat the drag and drop procedure, but modify the file name on the console to include the data type identifiers "qw", "ppt", "snw", "tmax", or "tmin" (do not use quotation marks) and change the extension to ".txt". After the program has processed the file, the console displays the input and output file names, the number of months (or lines) and days in the file, and prompts the user for more input. If the input file has missing lines, the user is notified on the console, and the unadjusted and adjusted number of lines is displayed.

Day_Cli_Ann_v5.3

The input to this program is the daily time series of weather data created by NCDC_V4 or surface-water data in the appropriate format downloaded from the U.S. Geological Survey Surface-Water web site. The output of the program is a time series (see figures 1–5) of various climate indices that are calculated for any accounting period that repeats every 365(6) days. The accounting period can be addressed to the day; for example the calendar year (01/01 to 12/31), water year (10/01 to 09/30), winter (12/21 to 03/20), and summer (06/21 to 09/20) are obvious possibilities. For computational purposes, the dates are assigned a Julian day using the **julday** function of Press and others (1992, p. 13). The **julday** function used in this version was corrected as described by Ahlquist (1999).

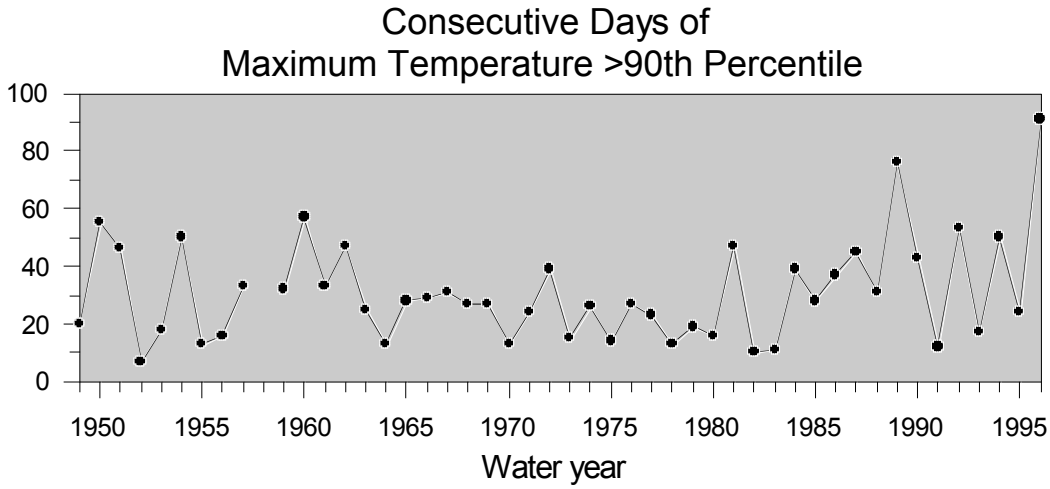

Figure 1. Consecutive days with detrended maximum temperature greater than the 90th percentile, 1949–1996, Furnace Creek, Death Valley National Park.

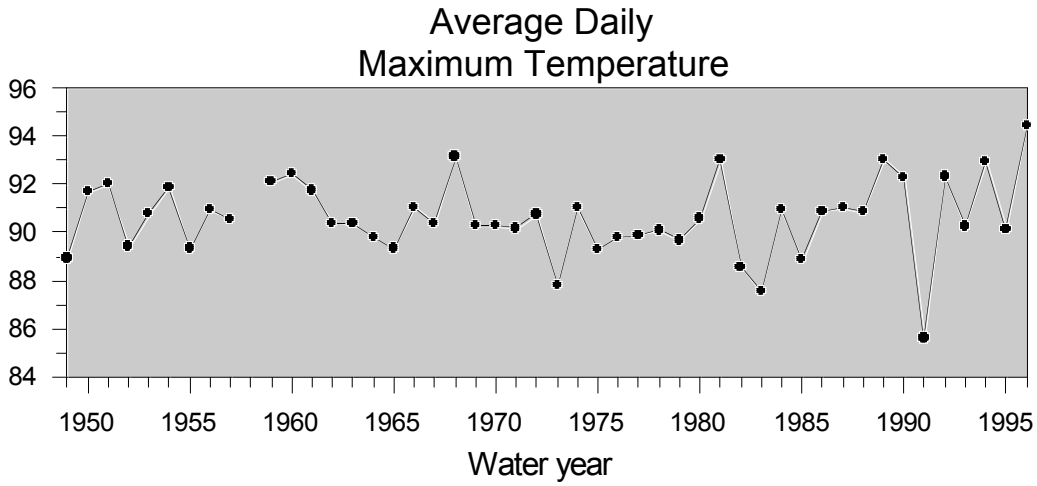

Figure 2. Average daily maximum temperature, 1949–1996, Furnace Creek, Death Valley National Park.

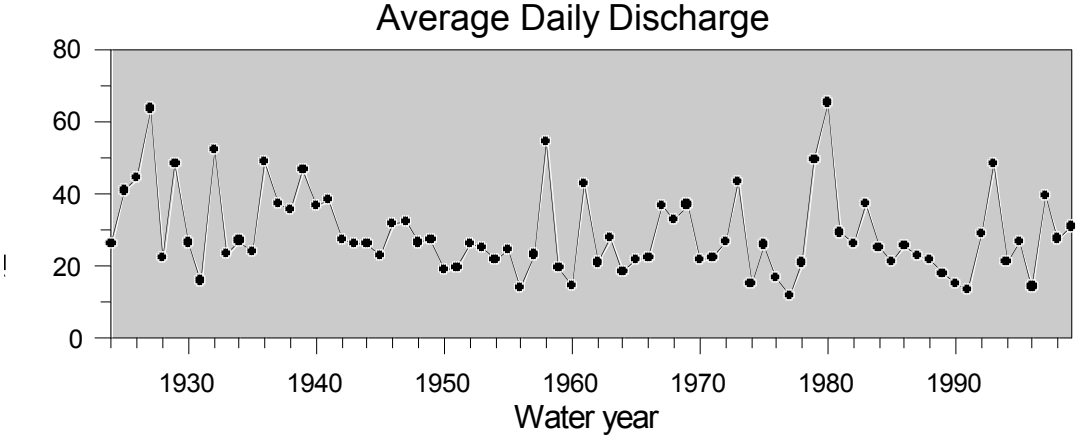

Figure 3. Average daily discharge, 1924–1999, Paria River near Lees Ferry, Arizona.

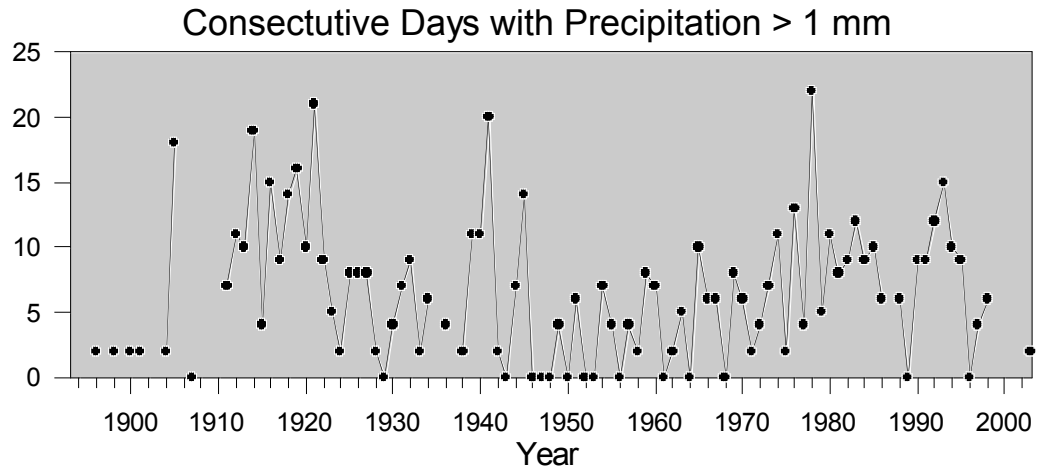

Figure 4. Consecutive days of the calendar year with precipitation greater than 1 mm (4HI), 1893–2003, Parker, Arizona.

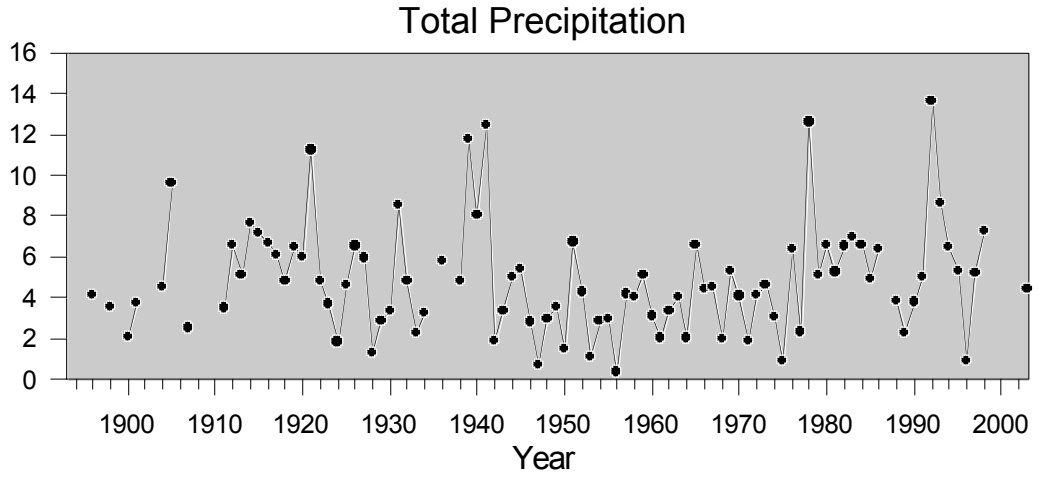

Figure 5. Total precipitation of the calendar year, 1893–2003, Parker, Arizona.

The time series output by the program depends on the type of data. Table 1 lists the output by column and data type. The program computes the usual statistics of total rainfall (TotPpt in Table 1), total snowfall (TotSnw), average daily temperature (Avg), and average daily streamflow (AvgQ) for the specified accounting period. The minimum (Min or Min#0, which is the minimum rainfall, streamflow, and snowfall greater than zero) and maximum (Max) values of rainfall, snowfall, streamflow, and temperature are also output for each period. In addition, the order statistics are calculated for the specified accounting period based on the period of record. For rainfall, snowfall, and streamflow, zeroes are not included in the ordering. With temperature data, the annual or solar heating cycle (fig. 6) is removed from the accounting period. This detrending is done by subtracting the average temperature of the calendar day (*i.e.,* 01/01, modulo 366) from the recorded temperature of the corresponding day (*i.e.*, 01/01/YYYY), which yields the deviation from the average daily temperature. These detrended daily temperatures are then sorted, the percentiles are calculated, and the deviations are averaged (AvgDev) for each accounting period.

Table 1. Output of Day_Cli_Ann_v5.3 by column for each accounting period.

Column No.	Precipitation (HI)[1]	Snowfall (TI)[2]	Streamflow (ft^3s^{-1})	Temperature (Fahrenheit)
1	Year	Year	Year	Year
2	TotPpt	TotSnw	AvgQ	Avg
3	Min#0	Min#0	Min#0	Min
4	Max	Max	Max	Max
5	Days>90[3]	Days>90	Days>90	Days>Q3[4]
6	Ppt>90	Snw>90	Days≤90	Days>90
7	Days≤90	Days≤90	Days>95	Days>95
8	Ppt≤90	Snw≤90	Days>99	AvgDev
9	Days>95	Days>95	ZeroDays	Missing
10	Ppt>95	Snw>95	Missing	Single>90[9]
11	Days≥1in[5]	ZeroDays	Single>95	Consec>90
12	Ppt≥1in	Missing	Consec>95	
13	ZeroDays[6]	Single>Q1		
14	Missing[7]	Consec>Q1		
15	Single>4[8]			
16	Consec>4			

(1) Hundredths of an inch, includes rain and snow
(2) Tenths of an inch
(3) Days with precipitation, snowfall, or streamflow exceeding the 90th, 95th, or 99th (streamflow only) percentile (zeroes excluded) and accumulated precipitation and snowfall (Ppt>90, Snw>90, Ppt>95, and Snw>95, respectively) where the percentiles are computed from the period of record
(4) Days with temperature exceeding the 75th, 90th, or 95th percentile of the deviation from the average daily temperature
(5) Days with precipitation greater than or equal to 1 inch and accumulated precipitation (Ppt 1in)
(6) Days without precipitation, snowfall, or streamflow
(7) Days with missing data
(8) Single day or consecutive day (Consec>4) occurrences of precipitation greater than 4 HI (ca. 1 mm)
(9) Single or consecutive occurrences of temperature deviation from the average daily temperature that exceed the 90th percentile

Various frequencies or occurrences in days of weather or streamflow relative to a threshold are also calculated. In most cases, the threshold is the value of a specified percentile; these are, depending on the type of data, the 25th (Q1), 75th (Q3), 90th, 95th, and 99th percentiles. For the precipitation and snowfall data, the accumulated precipitation or snowfall above or less than or equal to the specified percentile are also calculated. In addition, the accumulated number of single and consecutive days above a threshold is calculated for each accounting period. The precipitation threshold is presently set at 4 HI, or about 1mm, which is appropriate for the semiarid climate of the study area. The temperature thresholds are various percentiles of the detrended daily temperature.

Finally, missing entries are numerous in many of the daily time series. For this reason, the user must determine an acceptable number of missing daily entries in the accounting period. The number is expressed as a percent of the days in the accounting period. When the percent is exceeded, the missing value code (-99) is assigned to the accounting period.

Operation—The program proceeds on the console as follows:

Process a single or multiple files?

s = single file

m = multiple files [use batch file]

If s or S then–

Enter Name of Data File to Process–Drag and drop the file name at the prompt as described above in NCDC_v4

Enter Output File Name–Drag and drop the file name and modify it with a descriptive term such as (without quotation marks) "cal" (calendar), "wint" (winter), or "wy" (water year). Append the ".txt" extension to the file name, if necessary. If the file is opened successfully, the file name, the number of days (or lines), the number of missing days, and the first date and last date in the file are displayed.

Enter beginning date of first season or year (i.e. 01/01/1953)–Enter the beginning date of the first season.

Enter ending date of first season or year (i.e. 12/31/1953)–Enter the ending date of the first season.

Enter acceptable number of missing values per interval, in percent–Determine and enter an acceptable number of missing days expressed as a percent of the number of days in the accounting period.

Enter plotting position of first interval–This is the first year of the first accounting period.

Process more data Y/N

If y or Y then loop

If m or M then–

Enter path and name of batch processing file–Drag and drop the name of the batch processing file onto the prompt.

Enter number of files to process–Enter the number of input files to process.

Enter drive letter and directory of input files–With *Windows Explorer*, open the directory containing the input files. These are the files listed in column one of the batch file. From the *Address line*, drag and drop the folder icon onto the prompt.

Enter drive letter and directory of output files–With *Windows Explorer*, open the directory that will contain the output files. These are the files listed in column two of the batch file. From the *Address line*, drag and drop the folder icon onto the prompt. After processing the batch file, the program prompts for further processing.

Month_from_Day_v2.1

Use this program to create a monthly table of totals (precipitation and snowfall) or averages (streamflow and temperature) from the daily time series generated with NCDC_v4 or the surface-water data time series. At the prompts, drag and drop the input and output file names, respectively, as described above. It is useful to modify the output file name with "mth" for identification purposes. When operating in single-file mode, a third prompt will request the number of acceptable missing days per month in percent. Many files can be processed quickly using drag and drop. But, the program will process multiple files with the project batch file. In batch-processing mode, the four prompts are identical to those of Day_Cli_Ann_v5.3. The first column of the batch file is used to select the input files and the fifth column sets the acceptable percent of missing days per month.

Annual_Cycle_v2

This program computes average daily statistics of discharge, precipitation, or temperature for each of the 365 (366) days in a year (figures 6 and 7). Plots such as figure 7 are useful to determine the runoff or precipitation cycle of a region. The input data are surface-water data or the daily time series generated with NCDC_v4. Missing data are ignored. Drag and drop the input and output data as explained above. Modify the output file name with "ann" or another descriptive modifier, or place the results in an appropriately named directory. Many files can be processed quickly using drag and drop. The program will also process multiple files with a project batch file. The first column of the batch file used with Day_Cli_Ann_v5.3 specifies the input files.

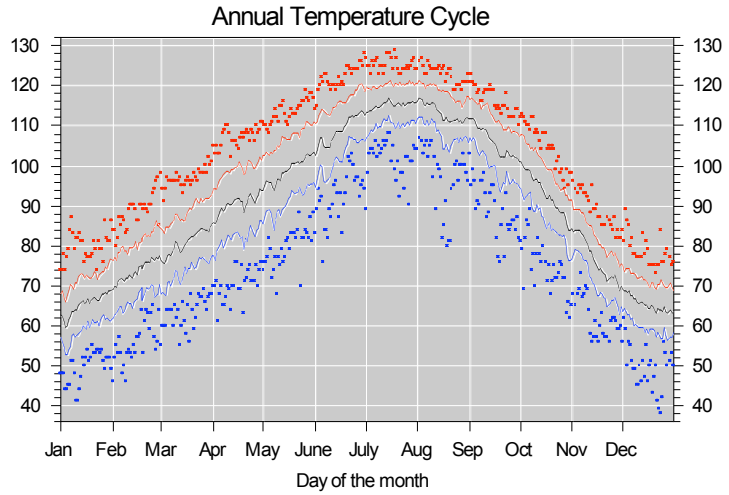

Figure 6. Annual cycle of maximum temperature, 1949–1996, Furnace Creek, Death Valley National Park. Solid black is average daily temperature; solid blue and solid red lines are the ± 1 standard deviation; and blue and red circles are the record minimum and maximum temperature for the day.

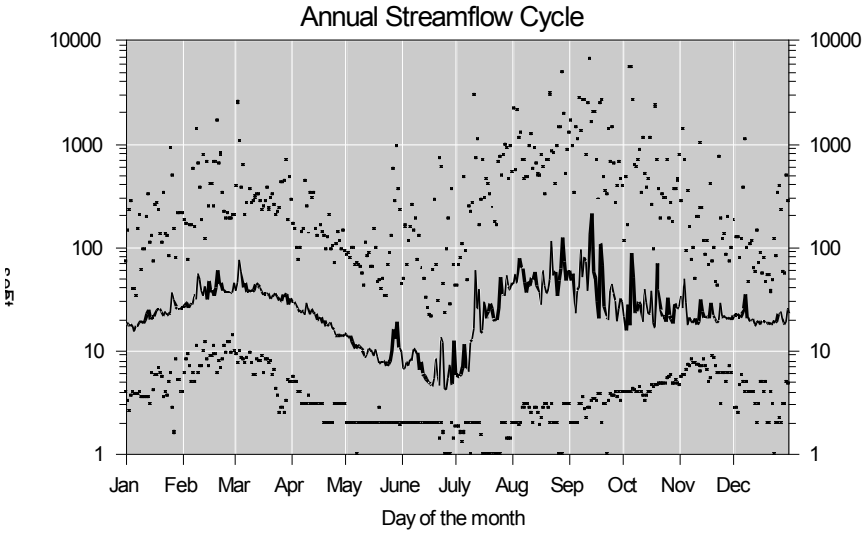

Figure 7. Annual streamflow cycle, Paria River near Lees Ferry, Arizona, 1924–1999. Solid line is the average daily discharge, solid circles are the minimum and maximum discharge of the day, respectively.

SetUpDailyTable

For certain types of analysis, it is handy to have a table containing all the daily time series in a project. This program generates a table with calendar dates in the first column followed by data columns. The data columns, which have the name of the weather or gaging station, are placed (or synchronized) in the table at their beginning dates. The user specifies the beginning and ending date of the table that must include the earliest and latest dates among all of the input files. Failure to correctly specify these dates will produce an array bounds error. The first column of the batch file used by Day_Cli_Ann_v5.3 is used to collect the files in the table.

Operation—With minimal typing, the program proceeds as follows:

Enter path and name of batch processing file–From *Windows Explorer*, drag and drop the file name of the batch file onto the prompt.

Enter drive letter and directory of input files–From the *Address line* of *Windows Explorer*, drag and drop the file icon onto the prompt. This is the path to the input files.

Enter drive letter and directory of output file–From the *Address line* of *Windows Explorer*, drag and drop the file icon onto the prompt. This is the path to the output files.

Enter name of output file–Type the name of the output table at the prompt. Add the ".txt" extension, which is useful for opening the file.

Enter first date of output table (i.e., 01/01/1890)–Type the earliest calendar date of the output table at the prompt.

Enter last date of output table (i.e., 12/31/2004)–Type the last calendar date of the output table at the prompt.

Process Another Batch File Y/N

SetUpTable_v1.1

This program is used to generate a multi-column table using the batch-processed files output from Day_Cli_Ann_v5.3. This table is readily imported into plotting and statistical software for further analysis. The first column in the output table contains the year of the accounting period. This is followed by data columns corresponding to one of the selected statistics (Table 1). The data columns are given the name of the weather or streamflow station. The program uses column two of the batch file to select the files used in the table.

Operation—The user prompts are essentially the same as for SetUpDailyTable except that the program prompts for the type of data (Table 1) to place in the output table.

Utility software

Cal_Date_Str_v1

This utility program generates a two-column table of calendar dates and missing values given a beginning and ending year. Enter the first and last year at the appropriate prompts. The output is a file with multiple lines of MM/DD/YYYY -99 that is always placed in the TEMP folder of your C: drive. The output table is useful for populating a spreadsheet with calendar dates or for padding a time series. Be sure to have a TEMP folder on the C: drive. If necessary, delete the previous CAL_DATES.txt file from the C: TEMP directory before running the program.

Date_Calculator_v2

Use this program to find the Julian day of a calendar date or the calendar date of a Julian day (selection 1), or to find the number of days (inclusive) between two calendar dates (selection 2). This latter selection is useful for checking that the number of lines in the two-column data file equals the number of days between the first and last date.

DecimalDates

Use this program to generate a table of daily dates in decimal format given a beginning and ending year. At the first prompt, enter the first year and last year separated by a space. The table is useful for populating a spreadsheet with decimal dates for applications that do not accept calendar dates. Be sure to have a TEMP folder on the C: drive. If necessary, delete the DEC_DATES.txt file from the C: TEMP directory before running the program.

References Cited

Ahlquist, J.E., 1999, Calendars and software: Bulletin of the American Meteorological Society, v. 81, p. 69-75.

Griffiths, P.G., Hereford, R., and Webb, R.H., 2006, Sediment yield and runoff frequency of small drainage basins in the Mojave Desert, U.S.A.: Geomorphology, v. 74, p. 232-244.

Hereford, R., Webb, R.H., and Graham, S., 2002, Precipitation history of the Colorado Plateau region, 1900–2000: U.S. Geological Survey Fact Sheet 119-02, 4p.

Hereford, R., Webb, R.H., and Longpré, C.I., 2004, Precipitation history of the Mojave Desert region, 1893–2001: U.S. Geological Survey Fact Sheet 03-117, 4p.

Hereford, R., Webb, R.H., and Longpré, C.I., in press, Precipitation history and ecosystem response to multidecadal precipitation variability in the Mojave Desert region, 1893–2001: Journal of Arid Environments, Special Issue, 46 manuscript pages, 15 figs., 3 tables.

Press, W.H., Teulosky, S.S., Vetterling, W.T., and Flannery, B.B., 1992, Numerical recipes in FORTRAN: The art of scientific computing: Cambridge, Cambridge University Press, 961 p.

Reek, T., Doty, S.R., and Owen, T.W., 1992, A deterministic approach to the validation of daily temperature and precipitation data from the Cooperative Network: Bulletin of the American Meteorological Society, v. 73, p. 753-762.