Corrections and Improvements to the CLIGEN Climate Database

Dayna L. Scheele and David E. Hall
USDA Forest Service Rocky Mountain Research Station, Moscow Lab
May 18, 2000

Introduction

Errors were found in the CLIGEN input files, called parameter files, distributed with the WEPP 95 CD. These parameter files are used by the stochastic weather generator CLIGEN to create the climate input file for WEPP. The files and the file generation program were analyzed and errors found were corrected. In addition, many new parameter files were generated.

Background

The CLIGEN computer program is a stochastic weather generator used to simulate weather data for use by many programs, primarily, the Water Erosion Prediction Project (WEPP) model. CLIGEN uses parameter files, compiled from weather station data, to generate simulated weather data for any number of years. The

Figure 1: The CLIGEN data path

parameter files list monthly statistical values for temperature, precipitation, dewpoint, solar radiation, and wind information. This analysis began when a WEPP user found the unrealistic value of 272 mm (10 ¾ in.) of precipitation in one day while using the Warren, ID, climate.

The goal of this analysis was to find and correct identifiable errors in the CLIGEN parameter files. Later, we discovered the ability to generate new parameter files from files provided by Jane Thurman at the Water Data Center in Beltsville, MD. The Water Data Center originally obtained the files from Arlin Nicks' computer--the developer of CLIGEN. Many of the corrections made are based on comparison to the 1991 station data inventory (*sod)* files obtained from the Western Regional Climate Center website (WRCC 1991).
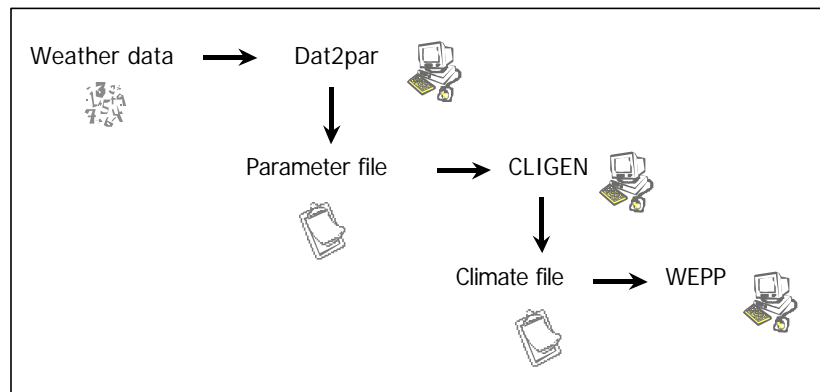
Analysis and Correction of Files

Dat2par Source Code and Files

Dat2par is our modified version of the FORTRAN program written to generate the parameter files. This program was originally called *calparms.f,* and was coded using older practices of programming FORTRAN, so it gave errors with our compiler. *Calparms.f* was recoded to run without the errors and named *dat2par* to distinguish it from *calparms.f.*

While examining and modifying the *calparms.f* code, we learned how a parameter file is generated. A *data* file (.dat) of daily values for precipitation and temperature from the weather station is read, and statistics are generated from the values. These statistics are written to the parameter file, and include:

- Mean precipitation on a wet day (MEAN P)
- Standard deviation of the precipitation on a wet day (S DEV P)
- Skew of precipitation on a wet day (SKEW P)

- Probability of a wet day following a wet day (P(W/W))
- Probability of a wet day following a dry day (P(W/D))
- Mean maximum temperature (TMAX AV)
- Mean minimum temperature (TMIN AV)
- Standard deviation of the maximum temperature (SD TMAX)
- Standard deviation of the minimum temperature (SD TMIN)

Descriptive information about the station comes from a station data file, *stations.dat*, and includes the station name, identification number, latitude, longitude, years of record, 24-hour rainfall distribution type, and elevation. The additional climate information in the parameter files was obtained by triangulating nearby weather stations and interpolating a new value using proximity to the parameter file station for weighting. The files and values that underwent this process are as follows:

- *statparm (142 stations)*
    - Maximum 30 minute and 6 hour precipitation depth (TP5 and TP6)
    - Solar radiation and standard deviation of solar radiation (SOL. RAD and SD SOL)
    - Maximum 30 minute precipitation depth (MAX .5 P)
- *dewpoint (273 stations)*
    - Dewpoint temperature (DEW PT)
- *timepeak.tot (1,548 stations)*
    - The time to peak cumulative distribution (Time Pk)
- *idxall and wind files in state directories (852 stations)*
    - Wind data for each of the 16 compass directions (%N through %NNW, MEAN, STD DEV, SKEW, and CALM)

The wind stations used for interpolation and their weighting are listed in the parameter files on the WEPP95 CD. Since other values are also interpolated, *dat2par* was modified to list all the stations and weightings used for interpolation in the parameter file.

Initial attempts to create the parameter files using dat2par were first run with Hawaii. We noticed some of the Hawaii stations were using a station in California or Alaska to interpolate values. The routine used for interpolation worked well in the contiguous 48 states, but had problems when there was not a large distribution of stations available for interpolation. This routine attempts to form an enclosing triangle of stations around the parameter file station location using the ten closest stations. In this case, all the stations within reasonable distance of the location were on one side of it and numbered less than ten. The next closest stations to Hawaii are in the Aleutian Islands of Alaska, and then Southern California. The routine exhausted all the Hawaii stations and then went on to using the Alaska or California stations. We applied a distance limit of 2000 km for finding nearby stations to fix the problem.

Precipitation and Temperature Data File

Apparent errors were found in several of the CLIGEN parameter files, mainly the skew values for precipitation. The files used to create the parameter files are not distributed with WEPP, however, a partial set of the weather station *data* (.dat) files were among the files from the Water Data Center. Some investigation found that the unusual skew values were generated because the precipitation and temperature *data* file used to create the flawed parameter file was improperly formatted. The *data* files were in fixed width format allowing four digits for each entry and a "missing value" notation of "9999." However, some of the files had a mix of "9999" and "99999" values, which lead to shifted values that were misread by *dat2par* and incorrect statistics were generated. A PERL program called *clicheck.pl* was written to search the *data* files

for this problem. *Clicheck.pl* was provided to the Water Data Center to assist them in identifying which *data* files needed correction. They corrected the identified files and provided them to us along with a list of the bad *data* files (*bad_sta.lst*).

Solar Radiation and Precipitation Depth

The values for TP5, TP6, SOL. RAD, SD SOL, and MAX .5 P were interpolated from the same file, *statparm*. The latitude and longitude values used for interpolation were examined and compared with the values listed in the *sod* files. The majority of the values were found to be in units of degrees and minutes when they should have been in decimal degrees (42° 24' listed as 42.24° and should have been 42.40°). In some cases, the latitude and longitude values were in different units in the same file. The required units for input into *dat2par* were decimal degrees, so the latitude and longitude values were assessed for accuracy and converted to decimal degrees when required. This had a big effect on the resulting parameter files in many cases; for example, Dubuque, IA was listed as 42.24 degrees when it should have been 42.40 degrees. This is a difference of 9.6 minutes (0.16°) and approximately 38.4 km (23.9 mi.) on the ground (2.5 minutes is approximately 4 km). This led to differences in weighting factors and stations used for interpolation of the solar radiation and precipitation depth values.

MAX .5 P was corrected, however, no adjustments were made to TP5, TP6, SOL. RAD, and SD SOL because we saw no obvious errors and lacked data for comparison. The values for MAX .5 P had two major cases of error. First, two of the values listed for Boise, ID, 5.0 for November and 9.0 inches for December, were many times larger than the other values (0.07-0.55 inches) listed for Idaho. These two months were changed to 0.08 inches based on information from Burns, OR, (a climate station located in the same ecoregion as Boise), and Pocatello, ID (another station physically close to Boise). This adjustment affected the November and December MAX .5 P values in climate files that used Boise, ID for interpolation. La Crosse, WI, contained the second case of error. All twelve values for MAX .5 P were listed as 0.00. This led to interpolation for locations near La Crosse—for example, Viroqua, WI—resulting in numbers much lower than the correct values as shown in Table 1. To correct this error, the data entry for La Crosse was removed from the file *statparm*.

Table 1.

Maximum 30 minute precipitation depth interpolated **with** La Crosse, WI, for Viroqua, WI											
MX .5 P	0.02	0.03	0.08	0.14	0.27	0.32	0.30	0.31	0.12	0.07	0.04	0.03

Maximum 30 minute precipitation depth interpolated **without** La Crosse, WI, for Viroqua, WI											
MX .5 P	0.10	0.20	0.37	0.82	1.25	1.30	1.39	1.48	0.64	0.35	0.34	0.19

Time-to-peak

The time-to-peak values were evaluated by visually comparing graphs of the values for stations in the same region. All the values had reasonable twelve-month trends and correlated with other values in the same area. No incorrect latitude and longitude values were discovered in the time to peak file, and only a few minor spelling corrections were needed.

Dewpoint

The dewpoint values were evaluated in the same graphical manner as the time-to-peak values. The values correlated with other values in the same regional area and followed reasonable yearly trends. Nine dewpoint station's latitude and longitude values required correction (Table 2). Most of the stations in the dewpoint file are located at major cities and the names were consistent with

Table 2.

Dewpoint Stations with Corrected Latitude and Longitude
Sandberg, CA
Washington, D.C.
Portland, ME
Blue Hill, MA
Albuquerque, NM
Binghamton, NY
Canton, NY
Winston Salem, NC
Pittsburgh, PA

the data for that station. However, the latitude and longitude specified for Albuquerque, NM, was for a location approximately 140 miles east of Albuquerque near Sumner Lake State Park. This was an unusual location but near enough to Albuquerque to cause indecision whether the location name or the location was incorrect, or both. The state climatologist for New Mexico was consulted and he provided dewpoint values for Albuquerque, which were used in place of the old values, and the latitude and longitude were set for Albuquerque, NM.

Figure 2. Graphical comparison of dewpoint stations in Indiana

Wind

Only some aspects of the wind data could be verified. The wind direction percentage and calm values should sum to 1.00, because either wind blows in a direction or the air is calm. A PERL computer program was written to read each file and sum the percent direction and calm values for each month, and when the value did not equal 1.00, the file was examined and corrected or removed. This method helped find several files where the values for the entire month were 0.00. These five files were removed from the database (Table 3). The latitude and longitude for the wind files were listed in an index file called *idxall* along with the pathname for the wind data file. The latitude and longitude values were checked for accuracy, and corrected (Table 4).

Station Data

The station data file lists the station name, number, latitude, longitude, elevation, and type for each station. This file was originally called *sdata*. This file was reformatted with corrections and named *stations.dat*. The elevation was fixed by comparing the stations file to *sod* files. Elevations of zero were evaluated and changed to the correct elevation. Then, elevation values were compared for differences of 20 feet or more and corrected. Station names were also checked by comparison with the *sod* files. Some stations listed in *stations.dat* had a name or station identification number that did not match the *sod* file. When this happened, the *sod* file was searched for another station with the same name or number and the two listings were compared by latitude, longitude, and elevation. The closest matching station listing in the *sod* file was used in stations.dat. Spelling errors in the station names were also corrected.

Several station temperature and precipitation *data* files did not have corresponding station information listed in *sdata*. New station entries were created in *stations.dat* using the *sod* files. All of the information needed for the new entries was available from *sdata* file and the *sod* files except TYPE. The TYPE values correspond to the four synthetic rainfall distributions described by the Soil Conservation Service in Technical Release 55. Types 1 and 2 (I and IA) are
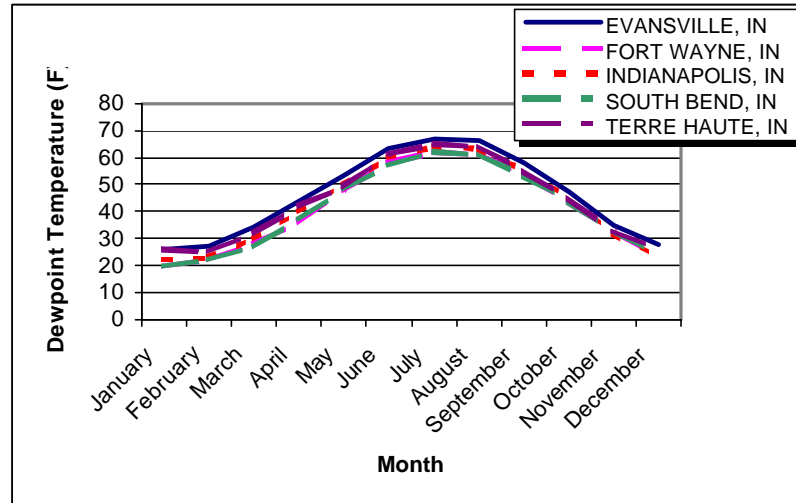
Table 3.

Wind Stations with Zero Values
Newhall, CA
Burney, CA
Fort Benning, GA
El Morro, NM
Killeen, TX

Table 4.

Wind Stations with Corrected Latitude and Longitude
Burbank, CA
Barber's Point, HI
Jolon, CA
Fort Knox, KY
Tucson/Davis, AZ
Guam/Agana

for the pacific maritime region with wet winters and dry summers. Type 4 (III) represents areas where large 24-hour rainfall amounts are brought on by tropical storms in the Gulf of Mexico and the Atlantic coast. Type 3 (II) covers the rest of the country (SCS 1986). The source of the value TYPE was unknown at the time, so a method of estimating it was developed. The TYPE values were plotted
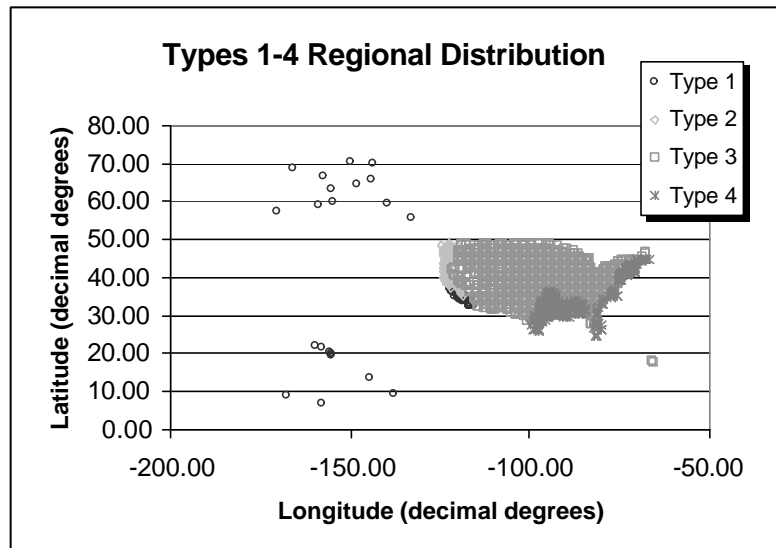

Figure 2.

on a graph by latitude and longitude. From this graph, we could see there were four regions. From this plot and the listed data, a value of 1,2,3, or 4 for TYPE could be determined using the latitude and longitude of the station. When station was on a border, the closest stations were used to get a better idea of what the value should be and typically, the closest station's TYPE value was used.

Summary

Once corrections were made to the program and data files, the parameter files were regenerated using *da2par*. A PERL program named *difference.pl* was written to determine how many parameter files had actually been changed from the originals and how many new files were produced. The results showed that all 1078 original files had at least one value different and 1570 new stations had been generated. 868 files had different years of record than before, producing different values for precipitation and temperature. The most significant change was the correction of the 155 improperly formatted precipitation and temperature *data* files, which changed skew values in the parameter files by 0.5 to 22.86. The three stations with the largest differences in skew were: Everglades, FL (21.24); Kainaliu Airport, HI (22.86); and Rowlesburg, WV (20.4). Due to the numerous changes in latitude and longitude values in *statparm*, the values interpolated from this file had the highest number of differences including 347 MAX .5 P values (difference greater than 0.1). Many of the wind values had large differences due to the increase in available stations for interpolation, and variability of wind measurements. Further differences include 16 station names, 5 latitudes, 14 longitudes, 33 elevations, 29 MEAN P values (difference greater than 0.1), and 96 dewpoint values (difference greater that 1.0).

Overall, the errors found in the CLIGEN database were not easily detectable without the precipitation and temperature data files and the files used for interpolation. Correcting the interpolation files has generated a more solid database for future use.

Future Work

Many of the stations used for interpolation are located at airports where the terrain is flat, open, and low in elevation. Many climate factors change with elevation and the terrain. If more interpolation stations with measurements at higher elevations could be added, the parameter files would be improved. In addition, many of the precipitation and temperature data files end in the year 1992 or 1993; so, more years of record could be added to the files. More stations with precipitation and temperature measurements, such as SNOTEL stations, could also be added as the years of data recorded at the stations increase.

References

USDA Soil Conservation Service. 1986. <u>Urban Hydrology for Small Watersheds</u>. Technical
Release 55.

Western Regional Climate Center. 1991. "Old NCDC Cooperative Observer Network
<u>WRCC – Active Coop Station Data Inventory Listings</u>.
http://www.wrcc.dri.edu/inventory/inventact.html