This is not a peer-reviewed paper.

Pp. 107-110 in Soil Erosion Research for the 21st Century, Proc. Int. Symp. (3-5 January 2001, Honolulu, HI, USA). Eds. J.C. Ascough II and D.C. Flanagan. St. Joseph, MI: ASAE.701P0007

Evaluation and Enhancement of the CLIGEN Weather Generator

D.C. Flanagan, C.R. Meyer, B. Yu, D.L. Scheele[1]

Abstract

Stochastic weather generators are often employed with process-based soil erosion prediction models. Long-term observed weather station statistics (precipitation, temperature, etc.) can be used to provide a daily series of temperature, precipitation, solar radiation, and wind parameters needed by an erosion model in its calculations for evapotranspiration, plant growth, soil frost and thaw development, snow melt, infiltration, runoff, and soil erosion. The Water Erosion Prediction Project (WEPP) erosion model typically uses the CLIGEN weather generator to provide the daily climatic inputs that drive its hydrologic and erosion processes. This presentation describes recently identified problems with CLIGEN and its weather station database, as well as enhancements to improve the generator program and supporting database.
Keywords. Climate generation, Stochastic weather generator, Erosion models, CLIGEN, WEPP.

Introduction

The Water Erosion Prediction Project (WEPP) is an effort by the United States Departments of Agriculture (USDA) and Interior (USDI) to improve technology for prediction of soil erosion by water. WEPP is constructed to simulate the important physical processes related to infiltration, percolation, runoff, evapotranspiration, residue decomposition, soil detachment by raindrops and flowing water, sediment transport and sediment deposition (Flanagan and Nearing, 1995). The model operates on a continuous basis, with daily climate inputs of precipitation, temperature, solar radiation, and wind, which drive the hydrologic, plant growth, residue decomposition, and soil erosion processes. The CLIGEN (CLImate GENeration) model was developed by Nicks et al. (1995), and uses historical observed weather data statistics for a station to generate a time series of simulated daily weather. Table 1 lists the input parameters required by CLIGEN. For determination of the precipitation (P) and temperature (T) parameters, observed long-term daily P and T values (>25 years) for each individual station are used, while interpolation methods are employed for the other factors (solar radiation, dew point, etc.) because of the lack of detailed data at all desired locations. CLIGEN was released in 1995 with a set of 1078 weather station parameter files for the 50 United States as well as other U.S. territories (Pacific Islands, etc.). Investigations over the past 2 years have indicated some problems in the CLIGEN software and database. This presentation will briefly describe the problem areas and efforts to correct and enhance CLIGEN.

Problem Identification and Solution

CLIGEN Database Errors and Enhancements

The USDA Forest Service (FS) Rocky Mountain Research Station in Moscow, ID, began examining the CLIGEN databases after an extremely questionable precipitation depth was observed by one of their users. CLIGEN v4.2 predicted a depth over 999 mm (39 in.) for Warren, ID, which is practically impossible for that location. Staff at the FS worked with WEPP personnel at the USDA-ARS National Soil Erosion Research Laboratory (NSERL) and with Dr. Jane Thurman at the USDA-ARS Hydrology Laboratory Water Data Center, and determined that the error was not in the CLIGEN FORTRAN model, but was instead in the input parameters (precipitation depth, standard deviation, skew coefficient values) for Warren, ID. This led to examination of the entire CLIGEN input parameter database, as well as the raw daily precipitation and temperature values (station data) and other data used to derive the CLIGEN parameters. These efforts were complicated by the death of Dr. Arlin Nicks (main developer of CLIGEN), and sparse documentation of procedures for developing parameters.

[1] Dennis C. Flanagan, Agricultural Engineer, USDA-ARS-MWA, National Soil Erosion Research Laboratory, West Lafayette, IN; Charles R. Meyer, Computer Specialist, USDA-ARS-MWA, National Soil Erosion Research Laboratory, West Lafayette, IN; Bofu Yu, Senior Lecturer, Griffith University, Faculty of Environmental Sciences, Nathan, Qld., Australia; Dayna L. Scheele, Civil Engineer, USDA-FS, Rocky Mountain Research Station, Moscow, ID. **Corresponding author**: Dr. Dennis C. Flanagan, USDA-ARS-MWA, NSERL, 1196 Building SOIL, West Lafayette, IN 47907-1196; tel.: (765) 494-7748; fax: (765) 494-5948; e-mail: <flanagan@purdue.edu>.

Table 1. Abbreviated list of weather station input parameters for the CLIGEN program

Station name, state number, station number, latitude, longitude, elevation
Maximum 30 minute and 6 hour precipitation depths
Average depth, standard deviation, and skew coefficient of precipitation on a wet day (by month)
Probabilities of a wet day following wet and dry days (by month)
Maximum, minimum and standard deviation of average daily air temperatures (by month)
Mean and standard deviation of daily solar radiation (by month)
Maximum 30 minute precipitation depth (by month)
Mean daily dew point temperature (by month)
Time-to-peak intensity cumulative distribution curve values for the station
Wind velocity, standard deviation, skew coefficients for 16 cardinal directions (by month)

A major problem in the database was caused by the use of fixed format field widths and use of "9999" to denote missing daily values. In some cases values of "99999" had been used. This shifted the precipitation depth values and caused errors when the utility program that converted the daily values to CLIGEN station parameter files was executed. Techniques and test programs were developed by the FS to test all of the raw data files, so that format errors could be corrected, and updated CLIGEN station parameter files could be created.

There are five separate databases used to create a CLIGEN station parameter file: station, solar radiation, dewpoint, time-to-peak, and wind data. Deficiencies were found in each of the databases. The FS expanded the databases of solar radiation, dewpoint, time-to-peak, and wind data available for interpolation, and used an additional 1570 station data files to create more than 2600 CLIGEN station parameter files. 868 of the station data files used had more years of observed data than those used in to generate the CLIGEN files distributed with WEPP in 1995. Changes to latitude and longitude values in the different databases altered which stations were used by the interpolation routines. The most significant change was the correction in 155 of the station data files, where the precipitation skew coefficients were off by at least 0.5, with some in error by as much as 22.86. The corrections and changes resulted in at least one change in each of the generated parameter files. There were significant changes to 347 values of monthly maximum 30 minute precipitation depth, 96 dewpoint temperatures (difference > 1.0 ^oF), 29 values of average mean precipitation on a wet day (difference > 0.1 inch), and changes to many wind values. In addition, 33 elevations, 16 station names, 14 station longitudes, and 5 station latitudes were corrected.

Storm Rainfall Intensity Predictions

Rainfall intensity can affect many variables in an erosion model simulation, including runoff volume, peak runoff rate, interrill detachment rate, rill detachment rate and sediment transport capacity. Previous sensitivity analyses of the WEPP model using single storm inputs found the erosion model output to be relatively sensitive to climate parameters related to either average or peak rainfall intensity (Flanagan and Nearing, 1991). However, during evaluations of WEPP model performance in a continuous simulation mode using CLIGEN-generated weather inputs during the summer of 1999, problems in the climate input files to WEPP were identified (Yu, 2000). Yu found that daily values for storm peak intensity predicted by CLIGEN v4.2 were insensitive to the values for monthly maximum 30 minute precipitation depths in the input parameter file.

Further investigation revealed that CLIGEN was computing a ratio of $w = R_{0.5}/R$, where both $R_{0.5}$ and R were rainfall depths. However, $R_{0.5}$ was in mm, while R was in inches; thus the ratio w was 25.4 times larger than it should have been. The overly large values for the ratio w usually exceeded an internal range check in CLIGEN, so that almost all values of w were set to a constant of 0.95 for all months and all locations. This resulted in the average predicted storm duration being 3.0 h and the peak rainfall intensity ($mm \cdot h^{-1}$) in a storm being about 3.1 times larger than daily rainfall depth (mm), regardless of location. However, simply correcting the initial w ratio coding error caused predicted storm durations to become extremely large, and predicted runoff and soil loss to become too low (Yu, 2000).

Fourteen sets of WEPP hillslope model input files for which previous WEPP model validation studies had been conducted were then used by Yu (2000) to develop a full correction to the intensity problem in CLIGEN. All of these sites had break-point rainfall data available for use as climate input files to the erosion model, as an alternative for synthetic weather generated by CLIGEN. Three soils (Caribou, Providence, Tifton) were used to provide a wide range of infiltration characteristics. Slope input was a uniform 22 m long profile at 9% gradient, and management was a continuous tilled fallow treatment. The locations and soils chosen resulted in 42 combinations for climate generator testing and verification. CLIGEN was modified in three ways to improve predicted storm duration, runoff and soil erosion: (1) the previously mentioned *w* ratio error was corrected, (2) new code was added to compute the monthly mean of the maximum 30 minute rainfall depth, and (3) new parameter values for storm duration and coefficient of variation for the ratio of maximum 30 minute rainfall depth to daily rainfall depth were incorporated, based upon the break-point rainfall data. Complete details of the changes made to CLIGEN v4.2 are provided in Yu (2000).

Figure 1 shows the runoff and soil loss results found when conducting 100 year WEPP (v99.5) model simulations with the modified CLIGEN (Simulated values on X-axes) climate input files, as compared to those obtained when using the break-point precipitation input files (Reference values on Y-axes). There was very good agreement overall (r^2 =0.97 for both runoff and soil loss), indicating that the modified CLIGEN model with the storm intensity corrections resulted in runoff and erosion predictions as good as those from break-point rainfall data model climate inputs.

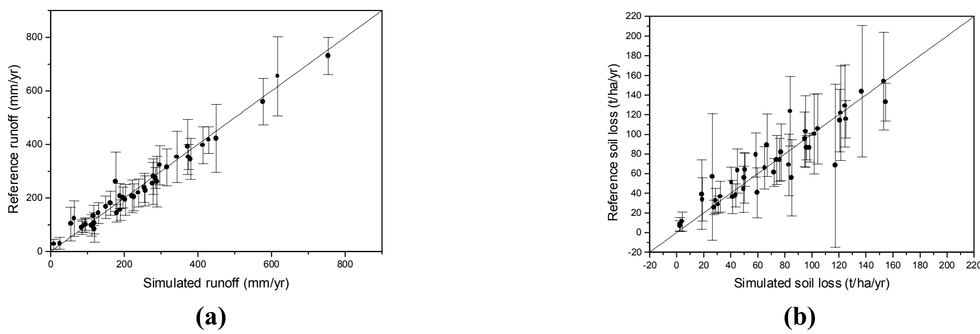

Figure 1. A comparison of the mean annual runoff (a) and soil loss (b) using observed break-point precipitation data versus climate generated with CLIGEN modified with Yu (2000) intensity corrections.

Results of additional WEPP model simulations are provided in Table 2. These 30 year model simulations were conducted for 4 locations, for a Miami silt loam soil, an S-shaped profile with average 10% slope, and continuous corn under a fall plow tillage system. The results illustrate that with the storm intensity changes, in some locations runoff and soil loss predictions increase (College Station, TX; Indianapolis, IN), while in other locations runoff and soil loss predictions decrease (Moscow, ID; Tucson, AZ).

Random Number Generation

In late 1999, the CLIGEN program was extensively recoded to facilitate maintenance, testing, and incorporation of enhancements. Following the recoding, the CLIGEN generator was evaluated to test how well the output numbers it generated for daily precipitation, minimum daily temperature, maximum daily temperature, and daily solar radiation matched the inputs. The Central Limit Theorem from statistics states that the means of samples drawn from a population approximate a standard normal distribution, regardless of the underlying distribution. This justifies application of “confidence interval tests on the mean” to a set of numbers allegedly from a particular population, to determine at what level of probability one can say those numbers are from that population (Ross, 1993, p. 69). This approach was applied to CLIGEN v4.2 to determine how well the means of the stochastically generated output values matched the monthly means from which they were generated. The results indicated that the generated distributions of precipitation depths, temperatures, and solar radiation that CLIGEN produced were not statistically equivalent to the input distributions at the 90% level of probability. Further testing revealed a problem with the two routines which produce the random numbers that all of CLIGEN’s outputs depend upon.

CLIGEN generates all its values from monthly means and standard deviations of observed data at the site. For each day for each parameter, CLIGEN uses a ***uniform random number generator*** to generate a random number between zero and one. That is fed into a ***standard normal deviate generator***, to produce a set of numbers closely approximating a standard normal distribution (mean=0, standard deviation=1). The generated ***standard normal deviate*** is scaled by the historically observed standard deviation, and added to the historically observed mean, to produce a daily value. Because their distributions are known, confidence interval tests can be performed on both the mean and standard deviation of the standard normal deviates. By definition, two normal populations are equivalent if they have the same mean and standard deviation.

CLIGEN was modified to generate a month’s worth of random numbers for a parameter at one time. The new code tests the mean and standard deviation of the numbers it produces. If they are not acceptable at the specified level of confidence, it generates a new set. This is analogous to quality control in a factory, where the standard normal deviate generator produces random numbers that CLIGEN consumes. Confidence interval testing on the means of output values generated with the random number screening showed that the predicted output distributions are now equivalent to the input ones. Table 2 shows the results of the random number modifications to CLIGEN in 30 year simulations, as well as the average annual precipitation derived from the base CLIGEN parameters. One can see that the random number screening technique brought the predicted values for precipitation closer to the expected mean values. The random number change also resulted in slightly decreased runoff and soil loss at all 4 locations. For Moscow, ID, predicted average annual sediment yield decreased by 26%.

Table 2. Summary of 30 year WEPP erosion model (v99.5) predictions using climate input files generated from original CLIGEN v4.2 and modifications

Location (Observed Prcp.)	CLIGEN Version*	Ave. Annual Precipitation (mm)	Ave. Annual Runoff (mm)	Ave. Annual Detachment ($kg\cdot m^{-2}$)	Ave. Annual Deposition ($kg\cdot m^{-2}$)	Ave. Annual Sed. Yield ($kg\cdot m^{-1}$)
College Station,	Original	1040	308	33.6	67.4	215
Texas	Orig+Int	1040	323	36.2	73.1	220
(P_{ave} = 960 mm)	Rand	988	265	28.1	61.0	174
	Rand+Int	988	271	30.2	72.0	169
Indianapolis,	Original	1035	145	10.1	29.4	65.6
Indiana	Orig+Int	1035	149	10.8	35.5	66.3
(P_{ave} = 1013 mm)	Rand	1012	131	9.7	30.5	60.8
	Rand+Int	1012	138	10.7	36.5	64.8
Moscow,	Original	643	21.7	1.9	8.0	11.3
Idaho	Orig+Int	643	8.1	0.8	5.1	3.1
(P_{ave} = 622 mm)	Rand	644	21.4	1.5	6.5	8.4
	Rand+Int	644	4.8	0.3	1.7	1.8
Tucson,	Original	310	25.5	4.2	11.3	21.9
Arizona	Orig+Int	310	21.6	3.7	11.1	16.2
(P_{ave} = 293 mm)	Rand	291	21.2	3.9	10.0	19.4
	Rand+Int	291	13.4	2.5	7.6	9.8

* Original - CLIGEN version 4.2; Orig+Int - includes storm intensity corrections; Rand - includes random number checks; Rand+Int - includes both random number checks and storm intensity corrections.

Summary

The CLIGEN weather generator and database commonly used to produce climate inputs for the WEPP erosion model has undergone significant recoding, evaluation and enhancement in the past two years. The new CLIGEN program is significantly improved in its ability to reproduce distributions of input parameters, as well as correctly simulate storm rainfall intensity. Also, the CLIGEN weather station parameter database has been corrected and expanded to over 2600 stations.

References

Flanagan, D.C. and M.A. Nearing. 1991. Sensitivity analysis of the WEPP hillslope profile model. ASAE Paper No. 91-2074. American Society of Agricultural Engineers, St. Joseph, MI. 26 pp.

Flanagan, D.C. and M.A. Nearing (eds.). 1995. *USDA-Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*. NSERL Report No. 10. West Lafayette, IN: USDA-ARS National Soil Erosion Research Laboratory.

Nicks, A.D., L.J. Lane and G.A. Gander. 1995. Chapter 2. Weather generator. In (D.C. Flanagan and M.A. Nearing, eds.): *USDA-Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*. NSERL Report # 10. West Lafayette, IN: USDA-ARS National Soil Erosion Research Laboratory.

Ross, S.M. 1993. *Introduction to probability models, 5th ed*. Academic Press, San Diego, CA.

Yu, B. 2000. Improvement and evaluation of CLIGEN for storm generation. *Transactions of the ASAE* 43(2): 301-307.