COMPARISONS OF HISTORICAL VERSUS SYNTHETIC WEATHER INPUTS TO WATERSHED MODELS AND THEIR EFFECT ON POLLUTANT LOADS

Authors: Fred D. Theurer, Agricultural Engineer, Natural Resources Conservation Service (NRCS), USDA;
Principal Author: phone—(301) 869-7195; or email—Fred.Theurer@verizon.net
Daniel S. Moore, Hydraulic Engineer, Natural Resources Conservation Service (NRCS), USDA,
Presentation & Correspondence Author: phone—(503) 414-3054; or email—dan.moore@por.usda.gov;
William H. Merkel, Hydraulic Engineer, Natural Resources Conservation Service (NRCS), USDA;
Quan D. Quan, Hydraulic Engineer, Natural Resources Conservation Service (NRCS), USDA;
Helen Fox Moody, Hydraulic Engineer, Natural Resources Conservation Service (NRCS), USDA;
Ronald L. Bingner, Agricultural Engineer, Agricultural Research Service (ARS), USDA;
James Bonta, Hydraulic Engineer, Agricultural Research Service (ARS), USDA; and
Dennis Flanagan, Research Agricultural Engineer, Agricultural Research Service (ARS), USDA

ABSTRACT: Synthetic weather generators (SWGs) are used extensively in continuous-simulation, water quality, watershed modeling. Many watersheds lack sufficient daily weather data and must be augmented, transposed, or synthetically generated. If risk analyses are an objective, long periods of weather are necessary to guarantee dependable downstream water quality statistics. Consequently, reliable SWGs must allow any length of simulation period for continuous-simulation watershed modeling. Because water quality models must estimate streamflow, erosion, sediment transport, and pollutant loads, the SWG's precipitation statistics should reflect all three precipitation-probability series of their indigenous parent's historical statistics including the average annual precipitation. Precipitation-probability statistics for the annual, partial duration, and complete duration series were created from historical records and two SWGs at three different locations and compared. The climate data for the historical and two SWGs were used with a continuous-simulation, water quality, watershed model and the resulting runoff, erosion, sediment yield, & sediment load were compared.

INTRODUCTION

How well do synthetic weather generators (SWG) represent historical precipitation for use with water quality evaluations? SWGs are important for continuous-simulation risk analyses of water quality in agricultural watersheds even if the sole objective is only "average annual" downstream water quality statistics. Many watersheds are sparsely or totally ungauged and daily weather data must either be augmented, transposed, or synthetically generated. Because water quality models must estimate streamflow, erosion, sediment transport, and pollutant loads, SWGs' precipitation statistics should match the precipitation-probability statistics of their indigenous parent's historical climate. This paper is concerned with how well two of USDA's SWGs reproduce their parent's historical weather's precipitation-probability statistics that will be used for subsequent simulation of watershed streamflow, erosion, and pollutant loads. An additional advantage of synthetically generated climate is that it can create time series of daily weather data variables that are sufficiently long to conduct risk analyses; i.e., probability analyses of predicted streamflow, erosion, and pollutant loads for more than predicting the "average annual" statistics. Historical records of daily precipitation frequently are serially incomplete; but, more importantly, they are generally too short, or often may not be available.

ARS watershed models are used to evaluate water quality impacts of soil & water conservation programs. They include: (1) Annualized AGricultural Non-Point Source Pollution computer model (AnnAGNPS) (Bingner & Theurer, 2001); (2) Soil and Water Assessment Tool (SWAT) (Arnold et al, 1998), and (3) Water Erosion Prediction Project (WEPP) (Flanagan & Nearing, 1995). Although each model may be used for different sized watersheds and were designed for agricultural-related watersheds with mixed landuse, all are continuous-simulation, water quality, watershed models that require daily inputs of weather data. The ARS watershed model AnnAGNPS was used for the analyses in this study to simulate streamflow, erosion, and pollutant loads.

Two SWGs developed by the Agricultural Research Service (ARS) are frequently used with agricultural-related watershed models: (1) CLImate GENerator (CLIGEN) (Nicks & Gander, 1994: Nicks et al, 1995); and (2) Generation of weather Elements for Multiple applications (GEM) (Johnson et al, 1996; Johnson et al, 2000). In this study, CLIGEN version 5.3 (USDA-ARS, 2010) and GEM version 6 (GEM6, USDA-NRCS, 2010) were used. Two important questions regarding these two SWGs are: (1) Do they reproduce the same precipitation-probability statistics as the historical data from which they were created? (2) What effect does the precipitation from SWGs have on the resulting runoff, erosion, & yield and downstream streamflow & pollutant loads compared to historical

records? Many additional precipitation-related questions ultimately must be answered satisfactorily such as number of rain days, mimicking seasonal precipitation distributions, grouping of storm systems, etc. However, this paper is limited to analyzing only statistics from three precipitation-probability series: (1) annual series; (2) partial duration; and (3) complete duration.

The objective of this paper is to compares precipitation-probability series statistics between the SWGs' output and the those from their corresponding underlying (parent) historical precipitation-probability series; and, using AnnAGNPS, compares simulated runoff, erosion, and pollutant loads using the original historical, CLIGEN, & GEM6 weather data. This was done by: (1) obtaining long term historical daily weather data; (2) conducting probability analyses of their three historical precipitation-probability series; (3) determining the remaining five weather variables that are necessary for continuous-simulation watershed modeling; (4) using CLIGEN & GEM6 to synthetically generate 1000-years of daily weather data; (5) conducting the same probability-statistical analysis for each of the 1000-years of synthetic precipitation; (6) comparing probability statistics between synthetic and historical precipitation; (7) using the historical and two synthetically generated precipitation datasets and their five remaining weather variables with AnnAGNPS for three physiographical different watersheds and comparing the predicted results. Three watersheds were chosen to represent different climates: (1) Stemple Creek on the coast of California (51.5 mi^2); (2) Cheney Lake, Red Rock subwatershed in Kansas (46.9 mi2); and (3) Maumee River Basin, Upper Auglaize watershed in Ohio (38.1 mi^2).

Precipitation is the dominant weather parameter driving runoff, erosion, & yield and their subsequent streamflow & pollution loads. If the objective of water quality modeling is to determine the probability of downstream streamflow (runoff & baseflow) and pollutant loads (sediment, nutrients, and pesticides), then as much as 100-years or more of daily weather data may be necessary to ensure a sufficiently large streamflow & pollutant load datasets for probability analyses that will lead to reliable risk analyses. Only data created from a statistically-accurate SWG may be able to meet this need because historical records often do not approach the length of record required. Also, historical data represent only one probable experience of a weather sequence. A weather generator could be able to generate many statistically desired sequences, for any record length. If climate change is important to water quality evaluation, then only synthetic weather could meet the changed climate statistics.

WEATHER VARIABLES

There are six basic daily weather variables recognized by most water-quality watershed models. They are: (1) precipitation; (2) maximum air temperature; (3) minimum air temperature; (4) dew point temperature; (5) wind speed; and (6) sky cover, cloud cover, or solar radiation. Some water quality models such as WEPP can use time periods less than one day, but all need at least daily information. Sky cover, cloud cover, and solar radiation are surrogates for the same weather parameter since solar radiation is easily predicted (Theurer et al, 1982). Sky cover can be calculated from solar radiation in conjunction with the other weather variables. This is done so that it will reproduce the original daily solar radiation.

While the effect of the magnitude and distribution of precipitation is more obvious, the other weather variables affect the soil moisture which leads to when and how much runoff occurs. The magnitude and timing of the originating soil erosion & pollutant loads are a function of the runoff.

SYNTHETIC WEATHER GENERATOR REQUIREMENTS

All SWGs' output should closely match their parent's historical probability statistics for the three precipitation-probability series. They are an ordered dataset of precipitation depths within the simulation period and were defined by Haan (1977) as the: (1) annual series which is the maximum precipitation for each year; (2) partial duration series which is all precipitation depths greater than zero and results in only the rain days; and (3) complete duration series which is all precipitation depths, including zero depths, for all of the days in the entire simulation period. The annual series may be sufficient for some models that are meant for use with annual rainfalls only, but all three series are important for continuous-simulation, water quality models.

Continuous-simulation, water quality, watershed models have strict requirements beyond just the three precipitation-probability series. These requirements include precipitation patterns that affect the antecedent moisture conditions (AMCs) used with daily runoff curve numbers to estimate runoff such as seasonal precipitation and association of wet days. They include: (1) matching partial durations for seasonal groupings of Julian days; (2) number of rainfall days within the chosen partial durations; and (3) the correlation of rainfall following a rainfall day (wet-wet, dry-wet, etc.) to reflect storm systems lasting for more than one (1) day. However, this paper will limit its analyses to

just the precipitation-probability series and the effects of their respective climate datasets on the runoff, erosion, & yield and downstream streamflow & pollutant loads.

TERMINOLOGY

There are a few terms that require precise definitions because the commonly used terms are ambiguous.

Streamflow is a term used by the U.S. Geological Survey (USGS) with their instream flow measurements. In this paper, it means all water passing a recording gauge regardless of its originating source. There are two sources for this streamflow: (1) baseflow which originates from the groundwater, but not all groundwater appears in the stream as baseflow; and (2) runoff which originates from storm events as surface water and quick return flow. Quick return flow is a term used by the Natural Resources Conservation Service (NRCS) to mean near surface, lateral flow that reappears in the stream during the storm's runoff. Tile drainage can significantly increase quick return flow.

Sediment has three major sources: (1) bed material load originating from the stream bed; (2) stream bank material originating from bank erosion along the stream system; and (3) wash load from erosion of the landscape upslope from the stream system. Wash load is the sediment runoff originating from the fields outside of the stream system. It is called wash load because it is composed mostly of fine-particle sizes; i.e., clay, silt, and occasionally some fine sand. This paper will refer to the first two sources from the stream system as bed & bank load.

Wash load ordinarily originates in fields upslope of the stream system from a combination of sheet & rill erosion and gully erosion and is the sediment runoff. The originating sediment runoff is transported overland where the larger particles deposit and the smaller particles are delivered to its receiving stream reach as wash load. Sediment yield is the wash load. Once in the stream system, the wash load mixes with any bed & bank material load where all identical particle sizes & shapes behave exactly the same regardless of their point of origin. The wash load will be transported in suspension while bed & bank material may be transported either as bed load, suspended load, or any combination depending upon the individual particle sizes & shapes. The amount of suspended sediment in transport varies according to its availability as wash load and amount of fine sediment found in the bed material and the individual particle size's transport capacity.

Yield is a term that is frequently misapplied because unrecognized downstream sources add to the load. Therefore yield is defined to be the amount of originating material on the landscape (e.g., erosion) that is transported to the landscape's receiving stream during the storm event that originates the material (e.g., sheet & rill or gully erosion). Load is the amount of material being transported within the stream. Both units can be either mass or mass per unit time.

PRECIPITATION-PROBABILITY COMPARISONS

All precipitation-probabilities were calculated using the same tool (Moore, 2009) which curve-fitted all ordered precipitation depths to a Log Pearson Type III probability density function according to the procedures in Bulletin 17B of the Interagency Committee on Water Data for "Guidelines for Determining Flood Flow Frequency". The final "station" skew was visually fitted to the ordered precipitation depths at the more extreme probabilities.

The more important precipitation-probability series for analyses are the partial duration & complete duration series because the annual series does not recognize precipitation for each Julian day of each year. Daily historical precipitation statistics reflect seasonal distributions and correlation of storm systems (i.e., wet-wet days, dry-dry-days, wet-dry days, & dry-wet days, etc.), while the annual series does not. The partial duration and complete duration series could, but CLIGEN & GEM6 may not have taken full advantage of these precipitation patterns.

All three synthetic precipitation-probability series should have the same statistics as their parent's historical series allowing for the difference in the size of the datasets. These statistics include the seasonal distribution and grouping of storm systems. If this was true, then the precipitation-probability relationship for the synthetic weather would be similar to its parent historical weather's relationship and would result in the same downstream water quality probabilities—again allowing for the difference in the size of the datasets.

Figure 1, Figure 2, & Figure 3 compare the three statistical series precipitation-probabilities for the historical years of record and 1000-years of synthetically generated precipitation from CLIGEN & GEM6 for: (1) 1913 to 2008 (96-years) at Petaluma, California; (2) 1896 to 2008 (113-years) at Pratt, Kansas; and (3) 1894 to 2008 (115-years) at Findlay, Ohio. The series probability scale has different probability time units for each series, and the time units

are: (1) annual series—years; (2) partial duration—rain days, while the number of rain days should not, they do, vary for each probability dataset; and (3) complete duration—number of days in the year.

The annual series relationship between the precipitation & the probability can be interpreted as the probability of the precipitation being equaled or exceeded in any one year. Alternative ways of referring to the probability is percent chance when the probability is multiplied by 100 or as a return period when using the reciprocal of the probability; e.g., a 0.01 probability is also a 1% chance and has a 100-year return period.

The partial duration precipitation series includes only the rain days for each probability dataset. While the probability time period's average number of rain days per year for the synthetic's precipitation series should be the same as the historical's at each location's, they are not (see Table 4 for the historical average number of rain days).

However the average number of rain days can legitimately vary greatly among different locations (CA, KS, & OH). The probability can be interpreted as the likelihood of the precipitation occurring on any one "rain day" of any year. There are multiple chances of occurrence each year according to the number of rain days each year.

The complete duration series includes all days in the entire simulation period for each probability dataset at each location. The probability can be interpreted as the likelihood of the precipitation occurring on any day of any year. There are 365 chances of occurrence each non-leap year and 366 chances each leap year.

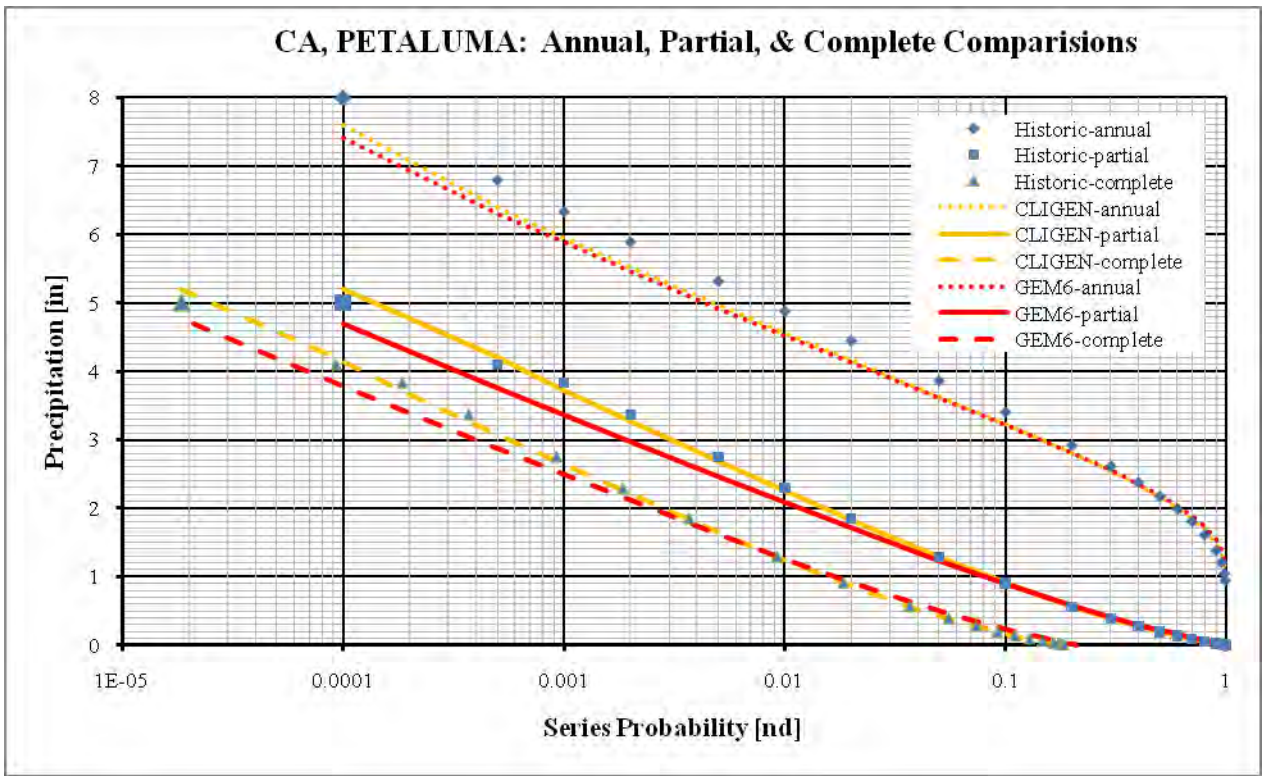

Figure 1: Historic & synthetic (CLIGEN & GEM6) annual (maximum precipitation each year), partial duration (rain days), & complete duration (daily) series precipitation-probability at Petaluma, CA (1913-2008).

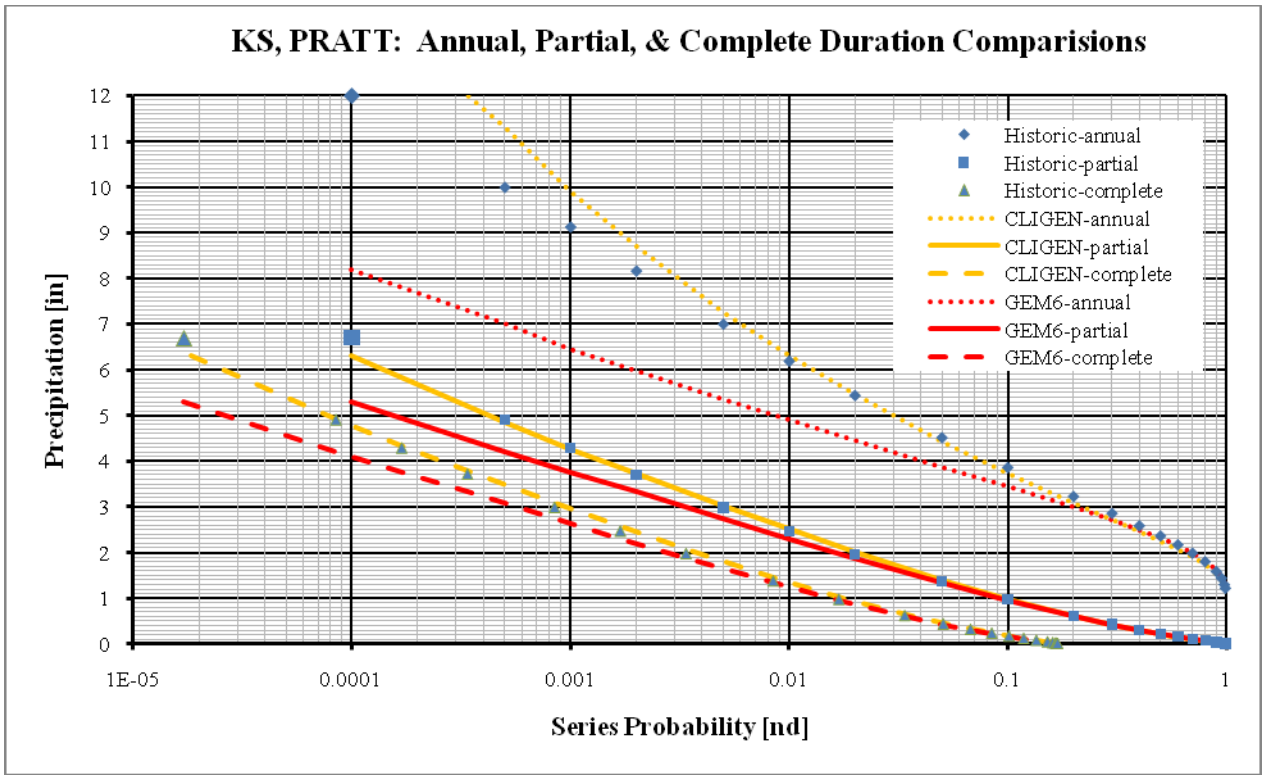

Figure 2: Historic & synthetic (CLIGEN & GEM6) annual (maximum precipitation each year), partial duration (rain days), & complete duration (daily) series precipitation-probability at Pratt, KS (1896-2008).

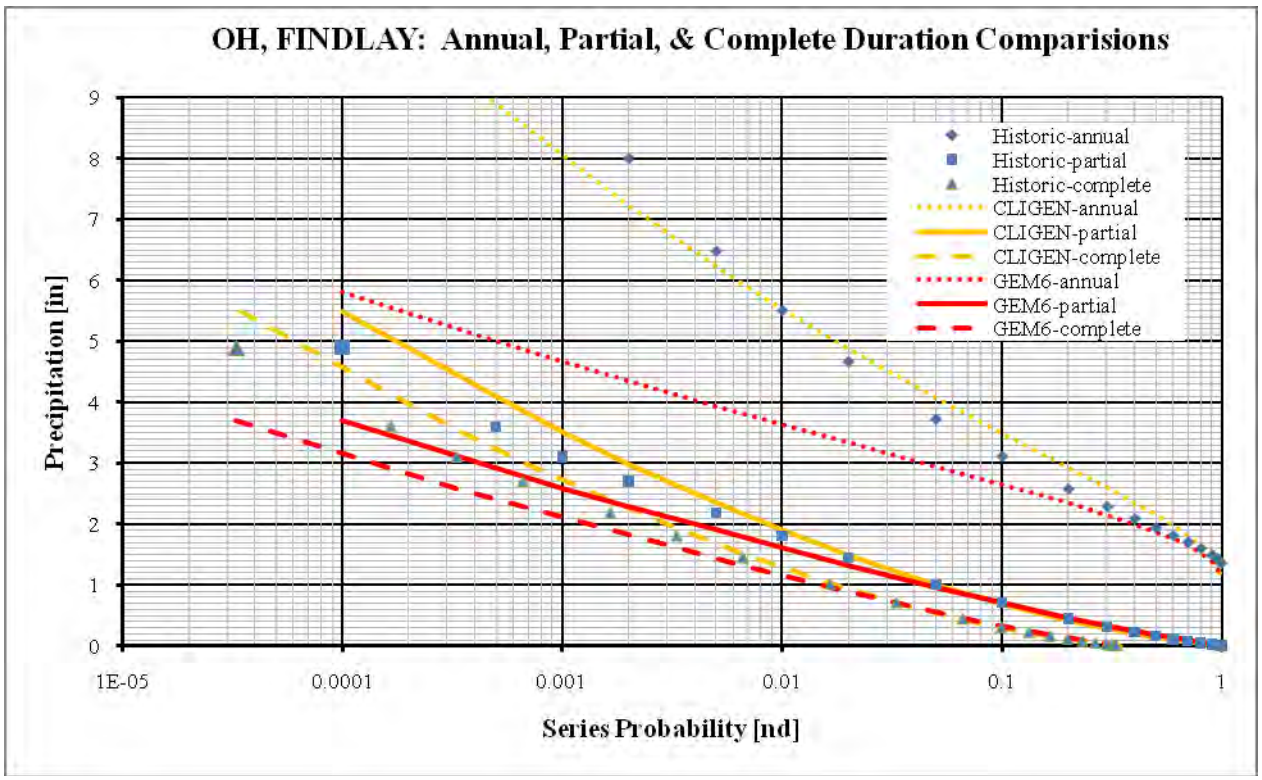

Figure 3: Historic & synthetic (CLIGEN & GEM6) annual (maximum precipitation each year), partial duration (rain days), & complete duration (daily) series precipitation-probability at Findlay, OH (1894-2008).

Table 1 shows the probability time period for each location and the three probability datasets for the partial duration & complete duration precipitation series. The annual series is not shown because it is simple to obtain as it consists of only the maximum precipitation in each year of record and its probability time period's average number of days per year is the number of years in the time period. The more useful portion of Table 1 is the equivalent return period probability

Location	Probability Dataset	Precipitation Series	Probability Time Period			Equivalent Return Period Probabilities			
			[years]	[days]	[day/year]	0.001	0.01	0.1	0.5
CA	Historical	Partial	96	6,496	67.667	1.48E-05	1.49E-04	1.56E-03	1.02E-02
		Complete	96	35,064	365.250	2.74E-06	2.75E-05	2.88E-04	1.90E-03
	CLIGEN	Partial	1,000	67,214	67.214	1.49E-05	1.50E-04	1.57E-03	1.03E-02
		Complete	1,000	365,242	365.242	2.74E-06	2.75E-05	2.88E-04	1.90E-03
	GEM6	Partial	1,000	76,731	76.731	1.30E-05	1.31E-04	1.37E-03	8.99E-03
		Complete	1,000	365,242	365.242	2.74E-06	2.75E-05	2.88E-04	1.90E-03
KS	Historical	Partial	113	7,009	62.027	1.61E-05	1.62E-04	1.70E-03	1.11E-02
		Complete	113	41,273	365.248	2.74E-06	2.75E-05	2.88E-04	1.90E-03
	CLIGEN	Partial	1,000	67,247	67.247	1.49E-05	1.49E-04	1.57E-03	1.03E-02
		Complete	1,000	365,242	365.242	2.74E-06	2.75E-05	2.88E-04	1.90E-03
	GEM6	Partial	1,000	61,976	61.976	1.61E-05	1.62E-04	1.70E-03	1.11E-02
		Complete	1,000	365,242	365.242	2.74E-06	2.75E-05	2.88E-04	1.90E-03
OH	Historical	Partial	115	13,913	120.983	8.27E-06	8.31E-05	8.70E-04	5.71E-03
		Complete	115	42,003	365.243	2.74E-06	2.75E-05	2.88E-04	1.90E-03
	CLIGEN	Partial	1,000	126,557	126.557	7.91E-06	7.94E-05	8.32E-04	5.46E-03
		Complete	1,000	365,242	365.242	2.74E-06	2.75E-05	2.88E-04	1.90E-03
	GEM6	Partial	1,000	120,439	120.439	8.31E-06	8.34E-05	8.74E-04	5.74E-03
		Complete	1,000	365,242	365.242	2.74E-06	2.75E-05	2.88E-04	1.90E-03

Table 1: Time periods and equivalent return period probabilities for the partial & complete duration series.

The equivalent return period probabilities were calculated by the following equation assuming that the probability was equal on any day.

$$p_r = 1 - \left[\left(1 - p_e\right)^{1/n} \right]$$

Equation 1

Where: p_r = equivalent return period probability [nd];
p_e = series event probability [nd}; and
n = probability time period [days/year].

Equation 1 produced exactly the same equivalent return period probabilities for both the partial & complete duration precipitation series; i.e., the curves overlay each other (see Figure 4, Figure 5, & Figure 6). CLIGEN agrees closely with the historical data throughout the entire series probability for all three locations. GEM6 does not.

The equivalent return period probability is approximately equal to the annual series probability for infrequent occurrences; e.g., there is a 1% chance for a complete duration's precipitation at a probability of 2.75E-05 in any one year span to occur. An important difference between the annual series and the equivalent return period probabilities for the partial & complete durations is that only the maximum precipitation is included in the annual series while all precipitation depths are included in the partial & complete duration series. This allows all years, who's second, third, etc. largest precipitation depths, to be included in the ordered statistics.

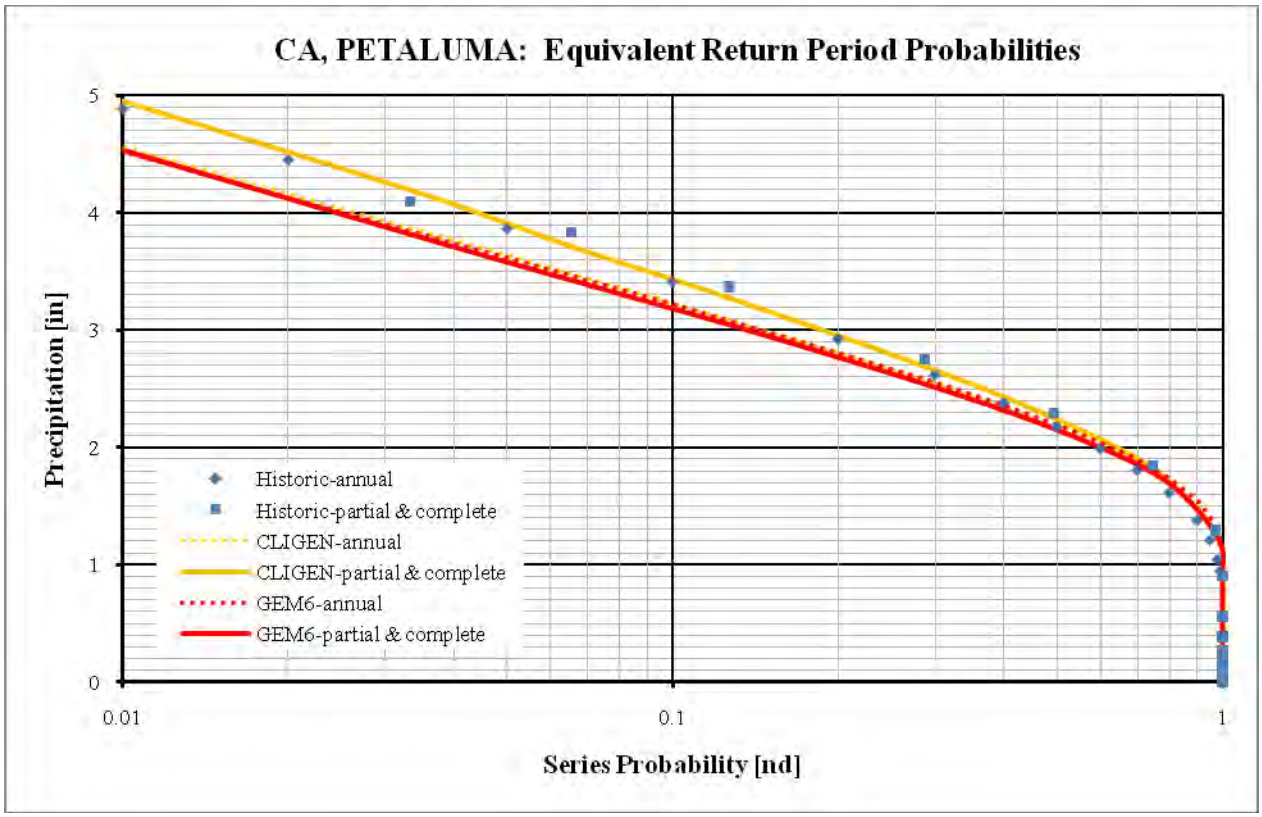

Figure 4: Equivalent return period probabilities for Petaluma, CA

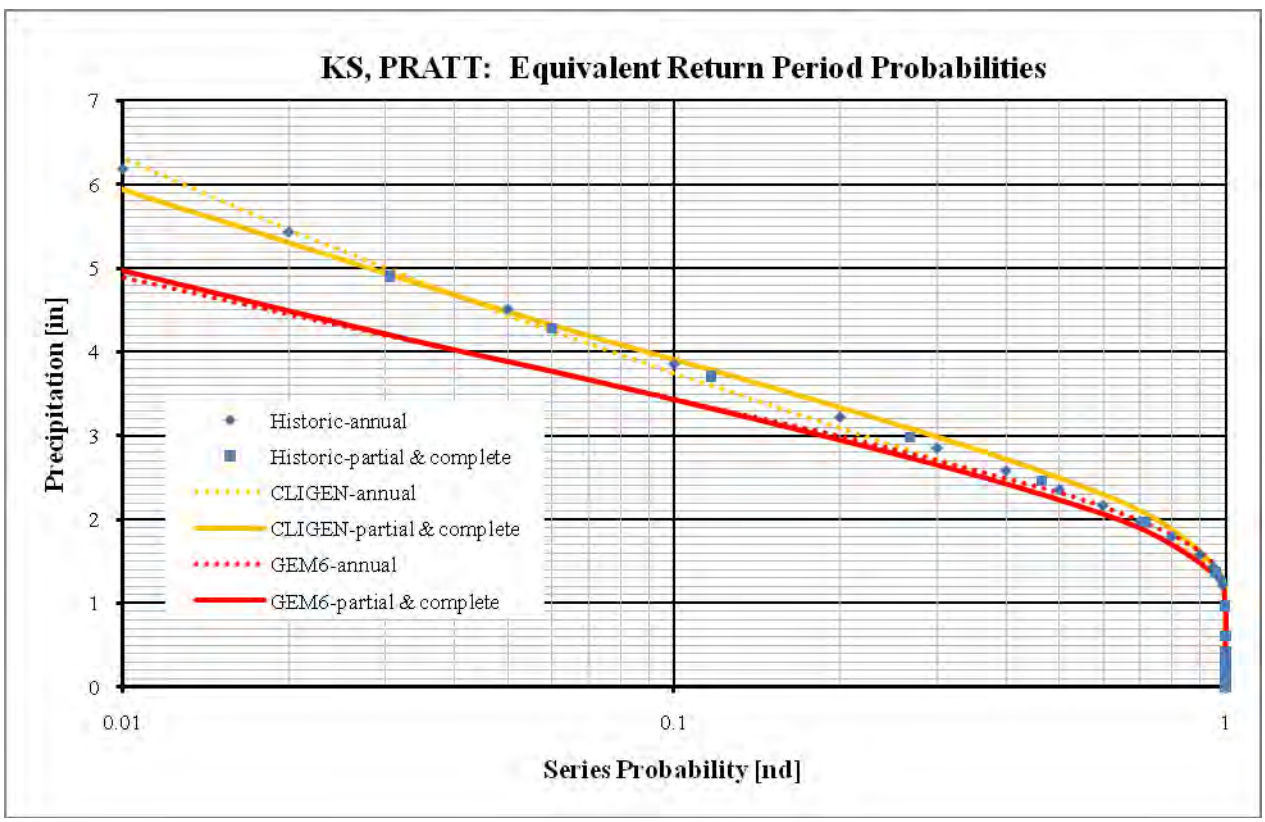

Figure 5: Equivalent return period probabilities for Pratt, KS

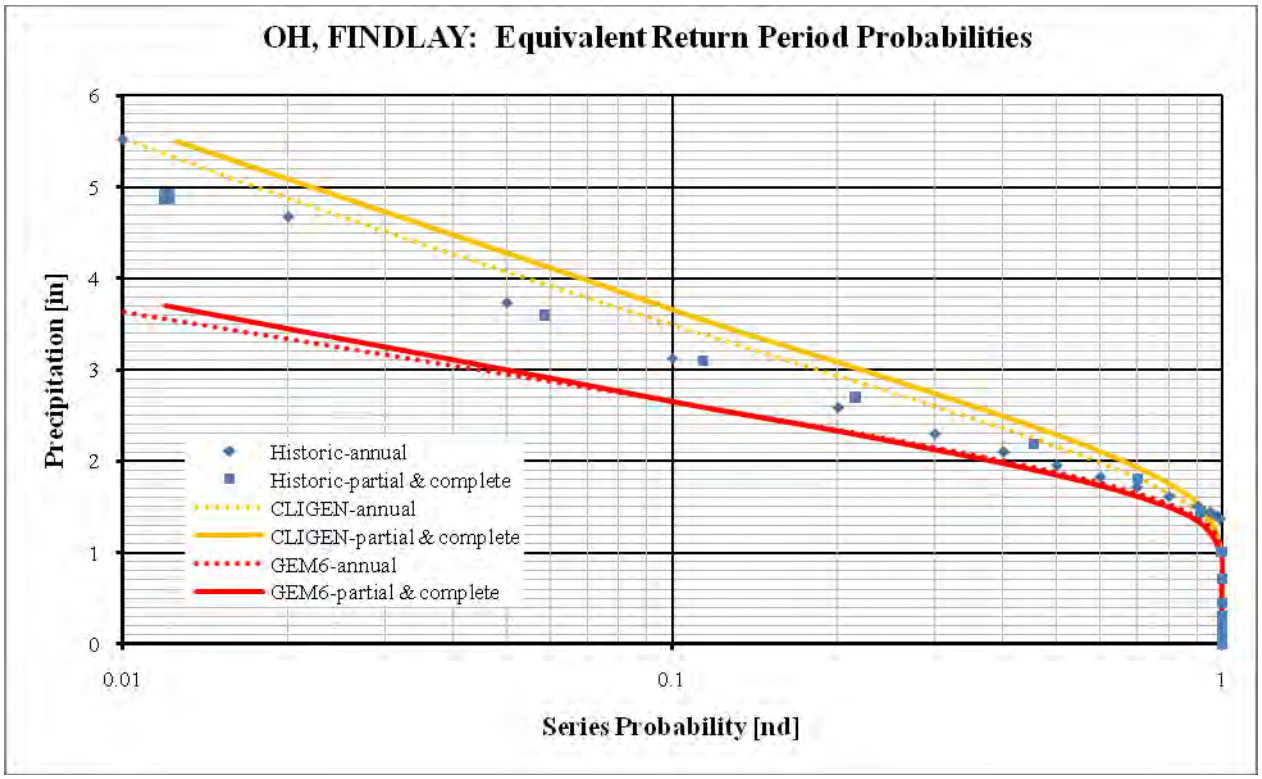

Figure 6: Equivalent return period probabilities for Pratt, KS

Table 2 shows basic annual series, precipitation-probability statistics. In addition to their unique hydro-geomorphic region, each watershed and its weather station were chosen for its length of recorded precipitation. The least length of record was 96-years at Petaluma, CA and the maximum was 115-years at Findlay, OH. All SWGs created 1000-years of daily weather data to capture the probability statistics of the historical climate when executing the watershed model and the likelihood of landuse coinciding with precipitation during watershed modeling execution.

Watershed Location			Climate Data Source	Accumulated Average Precipitation-Probability					
				Drainage Area	Time Period	Annual Series			
						0-0.5 Limits		0-1 Limits	
State	Watershed	Weather Station		[ac]	[yr]	[in/yr]	[%δ]	[in/yr]	[%δ]
CA	Stemple Creek	Petaluma (1913-2008)	historical	33,290	96	1.478	–	2.312	–
			CLIGEN		1000	1.422	-3.8	2.292	-0.9
			GEM6		1000	1.425	-3.6	2.318	0.3
KS	Cheney Lake, Site 5	Pratt (1896-2008)	historical	30,023	113	1.667	–	2.598	–
			CLIGEN		1000	1.616	-3.1	2.511	-3.4
			GEM6		1000	1.519	-8.9	2.457	-5.4
OH	Upper Auglaize Watershed	Findlay (1894-2008)	historical	24,396	115	1.370	–	2.199	–
			CLIGEN		1000	1.513	10.5	2.368	7.7
			GEM6		1000	1.187	-13.3	1.971	-10.4

[1] NA–not available

Table 2: Watersheds' historical and synthetic climate data showing average precipitation for the annual series" for all events and greater than the 2-year return period.

Table 3 shows the percentage of the mix of the sources of wash load erosion, yield, & load; i.e., sheet & rill versus concentrated flow (gully) erosion & sediment from the landscape; i.e., does not include any bed & bank sources. It indicates the importance of the two different erosion processes (Toledo Harbor AGNPS Project Team, 2005; Frees et al, 2009). Sheet & rill erosion is rainfall driven while gully erosion is runoff driven. CLIGEN's statistics more closely follows the historic statistics for Cheney Lake; GEM6's statistics do not. AnnAGNPS for the Upper Auglaize watershed was not executed for CLIGEN because there was a problem with the solar radiation units in its climate file, so no comparison was possible for this study for CLIGEN.

Watershed Location			Climate Data Source	Percentage of Total			
				Erosion		Yield/Load	
State	Watershed	Weather Station		Sheet & Rill	Ephemeral Gullies	Sheet & Rill	Ephemeral Gullies
				[%]	[%]	[%]	[%]
CA	Stemple Creek	Petaluma (1913-2008)	historical	100.00	0.00	100.00	0.00
			CLIGEN	100.00	0.00	100.00	0.00
			GEM6	100.00	0.00	100.00	0.00
KS	Cheney Lake, Site 5	Pratt (1896-2008)	historical	30.33	69.67	22.54	77.46
			CLIGEN	35.26	64.74	27.85	72.15
			GEM6	11.92	88.08	51.31	48.69
OH	Upper Auglaize Watershed	Findlay (1894-2008)	historical	9.43	90.57	26.26	73.74
			CLIGEN	[1]NA	NA	NA	NA
			GEM6	8.71	91.29	27.56	72.44

[1] Not available

Table 3: Watersheds' historical and synthetic climate wash load's erosion, yield, & load results showing the breakdown between "Sheet & Rill" and "Ephemeral Gully" erosion sources.

Watershed Location			Climate Data Source	AnnAGNPS Model Results: Runoff (water)							
				Average Annual Precipitation		[1]Average Number of Rain Days		Average Annual Runoff		[2]Average Number of Runoff Events	
State	Watershed	Weather Station		[in/yr]	[%δ]	[days/yr]	[%δ]	[in/yr]	[%δ]	[days/yr]	[%δ]
CA	Stemple Creek	Petaluma (1913-2008)	historical	24.98	–	67.67	–	5.077	–	37.1	–
			CLIGEN	24.72	-1.0	67.21	-0.7	5.308	4.5	42.9	15.7
			GEM6	28.94	15.8	76.73	13.4	5.535	9.0	45.5	22.8
KS	Cheney Lake, Site 5	Pratt (1896-2008)	historical	25.32	–	62.03	–	1.829	–	22.6	–
			CLIGEN	26.91	6.3	67.25	8.4	1.362	-25.6	17.9	-20.7
			GEM6	24.37	-3.8	61.98	-0.1	1.616	-11.6	22.4	-0.7
OH	Upper Auglaize Watershed	Findlay (1894-2008)	historical	35.76	–	120.98	–	7.659	–	61.2	–
			CLIGEN	36.61	2.4	126.56	4.6	NA	NA	NA	NA
			GEM6	36.23	1.3	120.44	-0.4	7.386	-3.6	64.7	5.7

[1] When daily precipitation exceeds 0.00 [in]; [2] when daily runoff exceeds 0.0001 [in]

Table 4: Historical and synthetic precipitation at three locations showing precipitation & rain days and AnnAGNPS' predictions of runoff, & runoff days.

Table 4 shows the AnnAGNPS' "Average Annual Precipitation" (integration of precipitation from 0 to 1 probability) and runoff statistics. The average annual precipitation depths are close for all but GEM6's CA Petaluma weather station. This anomaly for GEM6 created a variation in the "Average Number of Rain Days per Year" which suggests that there is a problem with GEM6's partial duration series. The "Average Number of Runoff Events" is the number of days of precipitation that resulted in surface runoff. Differences from historical climate

could be related to the other weather variables' influence on the AMCs. The differences between the annual series synthetic precipitation and its parent's historical precipitation are minor for all but GEM6's CA Petaluma weather station. The annual series may not reflect all the problems regarding the failure to recognize daily precipitation-probability and these failures may magnify problems with runoff & erosion and subsequent streamflow & pollutant loads. The Cheney Lake & Upper Auglaize GEM6 statistics do agree closely with the historical statistics.

Table 5 compares AnnAGNPS' model results for the average annual erosion, yield, load, and number of wash load events. CLIGEN's statistics agree more often than GEM6's. In fact, GEM6's statistical agreement was poor for Stemple Creek.

Watershed Location			Climate Data Source	AnnAGNPS Model Results: Wash Load							
State	Watershed	Weather Station		Average Annual Erosion		Average Annual Yield		Average Annual Load		[3]Average Number of Wash Load	
				[tn/ac/yr]	[%δ]	[tn/ac/yr]	[%δ]	[tn/ac/yr]	[%δ]	[days/yr]	[%δ]
CA	Stemple Creek	Petaluma (1913-2008)	historical	3.410	–	2.183	–	1.170	–	26.1	–
			CLIGEN	3.103	-9.0	1.976	-9.5	1.050	-10.3	26.8	2.5
			GEM6	1.471	-56.9	0.943	-56.8	0.480	-59.0	29.1	11.3
KS	Cheney Lake, Site 5	Pratt (1896-2008)	historical	0.921	–	0.397	–	0.114	–	20.7	–
			CLIGEN	0.750	-18.6	0.326	-18.1	0.093	-19.1	19.6	-4.9
			GEM6	1.140	23.8	0.510	28.4	0.094	-18.0	20.0	-3.2
OH	Upper Auglaize Watershed	Findlay (1894-2008)	historical	15.678	–	2.179	–	0.980	–	40.5	–
			CLIGEN	NA	NA	NA	NA	NA	NA	NA	NA
			GEM6	11.764	-25.0	1.468	-32.6	0.676	-31.1	42.4	4.6

[3] When daily wash load exceeds 0.0001 [t/ac]

Table 5: Historical and synthetic precipitation at three locations showing AnnAGNPS' predictions of erosion, sediment yield, & sediment load.

SUMMARY

SWGs are useful for three reasons, to: (1) serially complete a given length of record; (2) extend the length of record when the historical record is too short to adequately represent the impact of conservation measures on water quality loadings ; and (3) adjust climate factor(s) for climate change simulation.

Soil & water conservation professionals and water quality watershed modelers frequently encounter climate datasets that are serially incomplete, insufficiently long, or both. SWGs would be useful to overcome these deficiencies, but an SWG must not change the statistical probability associated with its historical parent's results. This includes seasonal distribution and correlation of storm systems. For example, a soil conservation professional may be interested in soil erosion in a farm field (sheet & rill and ephemeral gully). It would be incorrect if the use of SWG weather resulted in significantly different predictions of soil losses than the use of historical weather. The same is true for all weather probability statistics regarding any pollutant's source of runoff, yield, and loading.

The most important of the six daily weather variables for agricultural-related modeling is the precipitation. Important precipitation statistics include the: (1) precipitation-probability statistics for the annual, partial duration, and complete duration series; (2) rain day probability within each year; (3) seasonal variation of these same two sets of statistics by a seasonal grouping of Julian or solar days; (4) proximity of rain days as a function of storm systems; and (5) yearly cycles of these same two sets of statistics. The first two statistics directly affect runoff. The remaining two statistics, and all of the other five weather variables, affect the AMC which is important because it affects runoff and very few precipitation events result in significant runoff associated with erosion.

DISCUSSION & CONCLUSIONS

Table 1 is included to illustrate that there is a relationship between the three precipitation-probability series but this relationship is more complex than what Equation 1 would indicate. Seasonal precipitation and the correlation of precipitation days would make the probability of precipitation vary from day-to-day each year and not equal

throughout any year and was beyond the scope of this study. But Table 1 does offer limited information. CLIGEN's equivalent return periods probabilities shows that its partial/complete series tracks closely to its annual series probabilities while GEM6's does not. Does this indicate that CLIGEN uses the partial duration series but without any recognizing seasonal variation? Does GEM6 not even consider either the partial or complete duration series? More detailed analyses would be required to answer these important questions.

Table 2 shows the accumulated average precipitation- probability (area under the precipitation-probability curve) for separate two integration limits for the annual series. The 0 to 1 limits (most extreme probability or 0% chance to the 50% chance or 2-year return period) is the same statistics as an average annual precipitation while the 0 to 0.5 limits are more likely to reflect the precipitations that result in runoff. The differences between the 0 to 0.5 probability events are more pronounced than the total from 0 to 1. The lack of incorporating seasonal distributions, grouping of storm systems, and influence of the other five weather variables in the analyses on the AMCs is not known.

The differences in the accumulated average precipitation-probability limits between the historical and both CLIGEN & GEM6 were small for Petaluma & Pratt. But the differences for Findlay are larger than for the other two locations. GEM6 under predicted the precipitation for Findlay throughout the probability limits while CLIGEN over predicted but with less error. The difference between the 0 to 0.5 probability limits is more important because these storms cause the vast majority of runoff. However, the lack of incorporating all daily precipitation events, as well as any seasonal distribution, correlation of storm systems, and influence of the other five weather variables on the AMC are not apparent with the annual series precipitation-probability comparison. Annual series tests alone are not sufficiently comprehensive to say if an SWG is satisfactory.

Table 2 shows the precipitation-probability for the three series for each watershed. Since most pollutant loads are caused by the more extreme events, the precipitation-probability relationships between 0 & 0.5 probabilities are more important. Precipitation more frequent than 0.5 probability has little to no impact on the runoff and therefore very little erosion, wash loads, or pollutant loads happen because runoff occurs only when there is sufficient rainfall and with wet enough AMC to produce runoff.

Figure 2 shows the three precipitation-probability series—annual, partial duration, and complete duration—analyses for the historical years of record from 1896 to 2008 (113-years) at Pratt, Kansas as well as 1000-years of synthetically generated precipitation from CLIGEN & GEM6 which was based upon the same 113-years of historical record. The difference between Pratt's historical and the GEM6 "Average Annual Series Precipitation" for the entire probability range is great. The difference between the probable most extreme event (zero probability or 0% chance) and the 0.5 probability (2-year return period or 50% chance) is even more noticeable because the differences are entirely due to the extreme events. Yet the lack of incorporating all precipitation events, seasonal distribution, grouping of rain days, and influence of the other five weather variables do not appear with the annual series precipitation-probability comparison.

CLIGEN generally performed better than GEM6. CLIGEN closely matches the historical's equivalent return period probability curves at all three locations; GEM6 does not (see Figure 4, Figure 5, & Figure 6). The percent errors for the precipitation data's annual series' probability for the two SWGs in Table 2 show that the CLIGEN results closely match the parents' historical statistics while the GEM6 results are not as close.

Table 3 is meant to show the importance of ephemeral gully erosion with respect to the total erosion from the landscape. The original users for Stemple Creek did not consider the existence of ephemeral gullies—maybe because of the amount of uncontrolled sheet & rill erosion. But field investigations, coupled with careful analysis, may indicate otherwise as it has in the other two locations.

Table 4 shows the relationship between average annual precipitation & runoff and the average number of rain days & runoff events. Here CLIGEN out performed GEM6 for Stemple Creek but not for Cheney Lake. Why CLIGEN did not do well for Cheney Lake would take additional analyses—including the effect of the weather on the AMC. (And it is unfortunate that the climate data for CLIGEN at Findlay had a serious problem with the solar radiation preventing predicting the runoff, erosion, yield, & loads because GEM6 did perform well here.)

Table 5 shows that CLIGEN's runoff, erosion, yields, & loads vary from fair to good while GEM6's vary from poor to fair. GEM6 did not do particularly well predicting the average annual erosion, yield, & load for Stemple Creek in spite of the fact that it predicted more average number wash load events. The precipitation for GEM6 must have been less intense than the historic precipitation because GEM6 averaged more rain days in spite of the fact that GEM6 had a greater average annual precipitation.

REFERENCES

Arnold JG, R. Srinivasan, R.S. Muttiah, J.R. Williams. 1998. Large area hydrologic modelling and assessment part I: model development. Journal of American Water Resources Association 34(1): 73–89.

Bingner, R.L. & F.D. Theurer. 2001. AGNPS 98: a suite of water quality models for watershed use. *In Proceedings of* Sediment: Monitoring, Modeling, and Managing, Pgs. VII-1 to VII-8, 7th Federal Interagency Sedimentation Conference, Reno, NV. 8 pp.

Flanagan DC & M.A. Nearing (Eds). 1995. USDA–Water Erosion Prediction Project: hillslope profile and watershed model documentation. USDA–ARS–NSERL Report No. 10. West Lafayette, IN.

Frees, Lyle, Lisa French, and Fred Theurer. (2009). Conservation Effects Assessment Program: Special Emphasis Watersheds, Cheney Lake Watershed, Kansas. 17 March 2009. USDA, NRCS, Kansas. 50 pp.

Haan, C.T. 1977. Statistical Methods in Hydrology. The Iowa State University Press. Ames, IA. 378 pp.

Hanson, C.L., K.A. Cumming, D.A. Woolhiser and C.W. Richardson. 1994: Microcomputer program for daily weather simulation. U.S. Dept. Agric., Agric. Res. Svc. Pub. No. ARS-114, 38 pp.

Johnson GL, C.L. Hanson, S.P. Hardegree, E.B/ Ballard. 1996. Stochastic weather simulation: overview and analysis of two commonly used models. Journal of Applied Meteorology 35: 1878–1896.

Johnson, G.L., C. Daly, G.H. Taylor and C.L. Hanson. 2000: Spatial variability and interpolation of stochastic weather simulation model parameters. J. Appl. Meteor., 39, 778-796.

Moore, D. 2009. Frequency Curve: Log Pearson Type III Spreadsheet, version 2.13, 2 Oct 2009. http://www.wsi.nrcs.usda.gov/products/W2Q/H&H/Tools_Models/freq.html. Contact: Dan Moore, Hydraulic Engineer, NRCS Portland, OR; dan.moore@por.usda.gov; (503) 414-3054.

Nicks, A.D., and G.A. Gander. 1994. CLIGEN: A weather generator for climate inputs to water resource and other models. Computers in Agriculture 1994: Proc. 5th Int. Conf., pp. 903-909. Eds. D. G. Watson, F. S. Zazueta, and T. V. Harrison, St. Joseph, Mich.: ASAE.

Nicks A.D., L.J. Lane LJ and G.A. Gander. 1995. Chapter 2: Weather generator. In USDA–Water Erosion Prediction Project: hillslope profile and watershed model documentation, Flanagan, D.C., and M.A. Nearing (Eds). NSERL Report #10, USDA–ARS National Soil Erosion Research Laboratory; 2D1–2D22.

Toledo Harbor AGNPS Project Team. 2005. Upper Auglaize Watershed AGNPS Modeling Project Final Report. Prepared for: U.S. Army Corps of Engineers—Buffalo District. 14 February 2005, 68 pg.

Theurer, F.D. and K.A.Voos. 1982. IFG's instream water temperature model validation. *In* Conference on water and energy: technical and policy issues. ASCE Proc. of the Hydraulic Conf., Pittsburgh, PA and Ft. Collins, CO, May and June 1982, pg. 513-518.

USDA-ARS. 2010. National Soil Erosion Research Lab: Cligen. United States Department of Agriculture – Agricultural Research Service. URL: http://www.ars.usda.gov/Research/docs.htm?docid=18094 (Accessed 2/2010)

USDA-NRCS. 2010. Weather Generator Technology (GEM). United States Department of Agriculture – Natural Resources Conservation Service, National Climate and Water Center, Portland, Oregon. URL: http://www.wcc.nrcs.usda.gov/climate/gem.html (Accessed 2/2010)