EVALUATION OF AN ENHANCED WEATHER GENERATION TOOL FOR SAN ANTONIO CLIMATE STATION IN TEXAS

Ju Young Lee

School of Environmental and Water Resource Engineering, Department of Civil Engineering, Texas A&M University, College Station, TX 77840

Abstract: Several computer programs have been developed to make stochastically generated weather data from observed daily data. But they require fully dataset to run WGEN. Mostly, meteorogical data frequently have sporadic missing data as well as totally missing data. The modified WGEN has data filling algorithm for incomplete metrological datasets. Any other WGEN models have not the function of data filling. Modified WGEN with data filling algorithm is processing from the equation of Matalas for first order autoregressive process on a multi dimensional state with known cross and auto correlations among state variables. The parameters of the equation of Matalas are derived from existing dataset and derived parameters are adopted to fill data. In case of WGEN (Richardson and Wright, 1984), it is one of most widely used weather generators. But it has to be modified and added. It uses an exponential distribution to generate precipitation amounts. An exponential distribution is easier to describe the distribution of precipitation amounts. But precipitation data with using exponential distribution has not been expressed well. In this paper, generated precipitation data from WGEN and Modified WGEN were compared with corresponding measured data as statistic parameters. The modified WGEN adopted a formula of CLIGEN for WEPP (Water Erosion Prediction Project) in USDA in 1985. In this paper, the result of other parameters except precipitation is not introduced. It will be introduced through study of verification and review soon

1. INTRODUCTION

Modified WGEN model has invented for several objectives. Especially, the missing data filling is significant in modified WGEN compared classic WGEN. The classic WGEN model required 3 independents parameters as well as two 3 by 3 correlation matrices; the modified WGEN requires expressions for 5 independent parameters as well as the same sized correlation matrices. The second modification to the classic weather generator was to convert to fill in missing weather data values using the modified WGEN. Any classic WGEN does not include data filling algorithm.

The objective of using weather generators is to quantify for uncertainty analysis as a result of climatic variability. For stochastic daily precipitation evaluation of San Antonio Weather Station, classic WGEN and modified WGEN are used. Precipition data generated from classic WGEN and modified WGEN were compared with corresponding measured 7 year data (from 1988 to 1994). In addition, the modified WGEN includes data filling algorithm for incomplete meteorological dataset, and then, historical measured precipitation data and artificial missing dataset in modified WGEN were compared with corresponding 5 year data known as the same season(from 1983 to 1987).

2. METHODOLOGY

In classic WGEN and modified WGEN, daily precipitation occurrence is a two state Markov chain.

Markov chains specify the state of each day as 'wet' or 'dry' and develop a relation between the state of the current day and the state of the preceding days. A first order Markov chain (Bailey, 1964) was used to describe the occurrence of wet or dry days. The order of the Markov chain is the number of preceding days taken into account. A first order Markov chain referred in the literature is first order (Gabriel and Newmann, 1962; Caskey, 1963; Weiss, 1964; Hopkins and Robillard, 1964; Feyerherm and Bark, 1965, 1967; Lowry and Guthrie, 1968; Selvalingam and Miura, 1978; Stern, 1980a, b; Garbutt et al., 1981; Richardson, 1981; Stern and Coe, 1984; Ju Young Lee, 2003).

For this study a first order Markov chain with only two states, wet or dry, was used. A day with total rainfall of 0.2mm (0.008 in) or a wet day on day t given a wet day t-1; let $\Pr(W_t / W_{t-1})$ be the probability that the process at time t will be in "state" i given that at time t-1 the process was in "state" j such as probability of a wet day on day i given a wet day on day i-1; let $\Pr(W_t / D_{t-1})$ be the probability that the process at time t will be in "state" i given that at time t-1 the process was in "state" j such as probability of a wet day on day i given a dry day on day i-1. Then

$$\Pr(D_t / W_{t-1}) = 1 - \Pr(W_t / W_{t-1}) \quad (1)$$

$$\Pr(D_t / D_{t-1}) = 1 - \Pr(W_t / D_{t-1}) \quad (2)$$

where $\Pr(D_t / W_{t-1})$ and $\Pr(D_t / D_{t-1})$ are the probability of a dry day given a wet day on day i-1 and the probability of a dry day given a dry day on day i-1, respectively. Therefore these are commonly called the transition probability.

In case of precipitation amount, two methods have difference to represent the daily precipitation amounts for each day.

In WGEN, the shape of general distribution of precipitation resembles an exponential distribution. Todorovic and Woolhiser (1974) have often used an exponential distribution. The probability density function is

$$p_T(t:\lambda) = \frac{dP_T(t:\lambda)}{dt} = \lambda e^{-\lambda t} \quad (3)$$

where $p_T(t:\lambda)$ is the density function and λ is the distribution parameter.

On the other hand, in modified WGEN, the daily precipitation amounts are generated as eq (4):

$$X = \xi_i \sigma_i + \mu \quad (4)$$

where σ_i the standard deviation and μ the mean for given sample precipitation, ξ_i is modified standard normal variate and Wilson and Hilferty(1931) defined the following relationship;

$$\xi_i = \frac{2}{\gamma_\xi}\left(1+\frac{\gamma_\xi t_i}{6}-\frac{\gamma_\xi^2}{36}\right)^3 - \frac{2}{\gamma_\xi} \quad (5)$$

where γ_ξ is coefficient of skewness and γ_ξ is defined as

$$\gamma_\xi = \frac{1-\rho^3}{(1-\rho^2)^{1.5}}\gamma_x \quad (6)$$

And γ_x is the skewness coefficient, ρ is the lag-one correlation coefficient and t_i is a normally distributed random variable with mean 0 and variance of 1. Especially, feeding the output from a uniform pseudo random number generator into a standard number generator generates it.

Before the determination of filling dataset, the wet day or dry day is determined from following equation from first order Markov chain,

$$P\left(\frac{Wet_i\_or\_Dry_i}{Wet_{i-1}}\right) \text{ Or } P\left(\frac{Wet_i\_or\_Dry_i}{Dry_{i-1}}\right) \quad (7)$$

After this determination of wet or dry day, missing data filling algorithm is processed. For data filling of precipitation, transition matrices and λ are determined and the seasonal variation was described by using Fourier series.

A and B matrices are obtained from measured data for 7 years data measured (from 1988 to 1994), also they get from Matalas equation,

$$\chi_{p,t}(j) = A\chi_{p,t-1}(j) + B\varepsilon_{p,t}(j) \quad (8)$$

where $\chi_{p,t}(j)$ and $\chi_{p,t-1}(j)$ are (5×1) matrices of the maximum temperature(j=1) and minimum temperature(j=2), relative humidity(j=3), wind speed(j=4), and solar radiation residuals(j=5) for days t and t-1 of year p; $\varepsilon_{p,t}(j)$ is a (5×1) matrix of independent random components that are normally distributed with a mean of zero and a variance of unity. A and B are (5×5) matrices whose elements are functions of the lag 0 and lag 1 serial and cross correlation coefficients of the observed residuals, defined so that any series of residuals generated by a series of standard normal errors exhibits the same serial and cross correlation as the observed residuals.

The given dataset gives precipitation information for the years from 1983 to 1994. The focus of this study is to use markov probability theory to determine historical data sets from a known set of data. For example, this study uses the precipitation data sets ranging from the years 1988-1994 to reasonably estimate historical data for the years from 1983 to 1987. The accuracy of the model is tested by comparing the estimated data with the actual data in terms of percentage error and standard deviation.

Table 3 shows the result of historical data mean and standard deviation measured and the result of mean and standard deviation of filled dataset from modified WGEN as the same season from 1983 to 1987. The results of other parameters generated by using modified WGEN with data filling algorithm will be introduced through Journal of KWRA.

3. RESULTS

Measured daily weather data for San Antonio, Texas in US are obtained for testing the comparative study of generated precipitation data by using WGEN technique (Richardson and Wright, 1984) and modified WGEN in National Climate Data Center (NCDC). Daily values of precipitation, maximum temperature, minimum temperature, relative humidity, wind speed, and solar radiation data are obtained. For San Antonio weather station, 7 years data are used to estimate six variables. The year is partitioned into 14day period (26 periods a year). However, the results of others parameters except precipitation will be introduced through Journal of KWRA.

Figure 1 illustrates the measured and Fourier series λ value of each period on actual daily precipitation totals at San Antonio, Texas for 7 years.

To apply measured λ value to WGEN, the distribution of Fourier λ value tended to well express the measured precipitation data on all period at San Antonio, Texas.

On the other hand, to apply measured data into modified WGEN with data filling function algorithm, table 1 illustrates the measured adequate ξ_i , measured mean and standard deviation by using modified WGEN with data filling algorithm at San Antonio, Texas.

Where, adequate ξ_i is used for good expression of measured precipitation distribution on modified WGEN.

Figure 2 shows that precipitation data generated by WGEN and modified WGEN did compare well with the measured precipitation data for San Antonio, Texas.

It is important to note that this generated precipitation data came from measured precipitation data by using WGEN and modified WGEN. Especially, measured data matched well with data generated by the modified WGEN. However, the WGEN model tended to underpredict the daily precipitation on San Antonio site.

In general, WGEN have been criticized for their inability to generate the extreme conditions and special precipitation pattern. For using modified WGEN, the equation (4) and (5) referred by Wilson and Hilferty showed well extreme condition and special precipitation pattern.

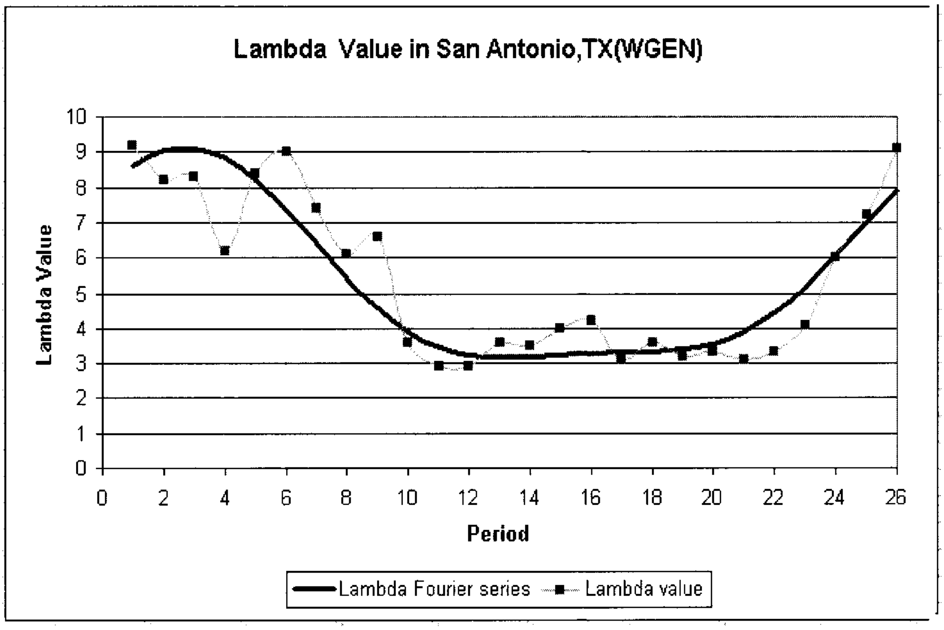
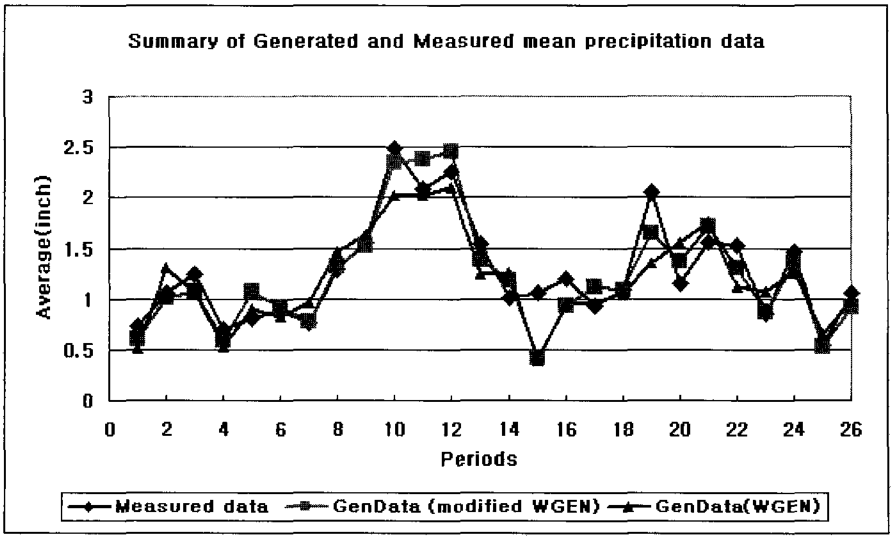


Figure 1. Measured and Fourier series λ distribution of wet day periods extraction from 7 years

Table 1. ξ_i, measured mean and standard deviation for application of modified WGEN with data filling algorithm during 26 periods

Period	ξ_i	Measured Mean	Measured Std
1	3.360	0.733	0.740
2	3.040	1.059	1.432
3	2.925	1.239	1.456
4	2.630	0.694	0.757
5	2.350	0.808	1.278
6	3.240	0.882	0.906
7	2.995	0.762	0.842
8	2.665	1.287	1.386
9	2.280	1.551	1.171
10	3.060	2.484	2.636
11	2.230	2.086	2.216
12	2.315	2.251	2.852
13	2.805	1.533	1.431
14	2.976	1.009	1.394
15	2.800	1.064	1.745
16	2.475	1.203	1.946
17	1.163	0.930	0.933
18	1.646	1.085	1.080
19	2.515	2.048	2.120
20	4.685	1.154	1.780
21	2.808	1.551	1.765
22	2.938	1.523	1.805
23	2.300	0.861	0.894
24	2.085	1.468	1.346
25	2.100	0.548	0.554
26	4.738	1.053	2.576



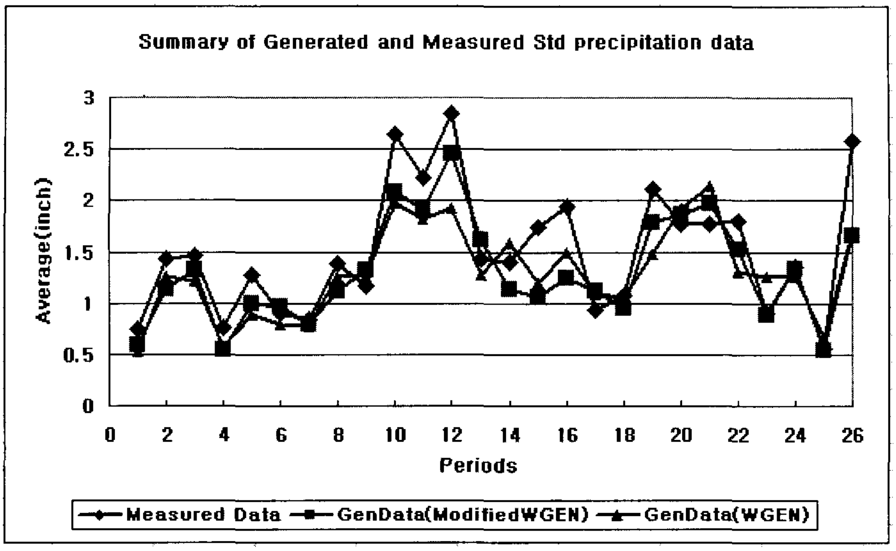


Figure 2. Relationships among precipitation data measured and generated using WGEN and modified WGEN

Table 2. Statistic analysis of precipitation amounts

	Measured Data	WGEN Generated Data	modified WGEN Generated Data
Mean(in)	1.26	1.19	1.23
Std(in)	0.51	0.37	0.55
Skew Coeff.	0.94	0.38	0.92
Kurtosis Coeff.	0.28	0.59	0.38

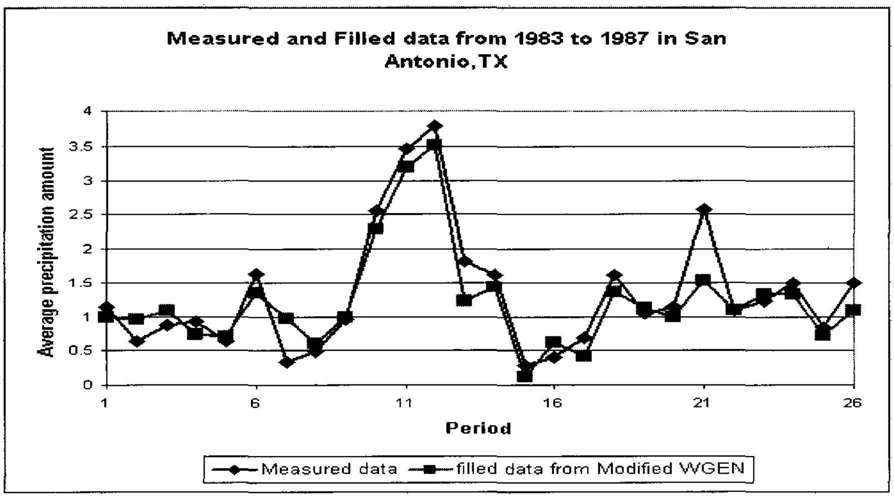


Figure 3. Relationships between measured data and filled data from modified WGENCreated by test

The WGEN model tended to underpredict the precipitation means and standard deviations of all periods on San Antonio, where the mean and standard deviation were underpredicted by 6% and 27.4%. The periodic distribution was normal, and the skewness moderately skewed.

In general, an absolute skew coefficient of <1 is considered moderately skewed and <0.5 fairly symmetric (Evans and Olson, 2002).

The modified WGEN model tends to predict the precipitation means and standard deviation of all periods, where the mean was a little bit underpredicted by 2.4%. However, it is negligible.

On the other hand, the standard deviation was overpredicted by 7.2%. Most of all, skew coefficient and kurtosis coefficients were well simulated at this site by using modified WGEN more than with WGEN.

Figure 3 and Table 3 illustrate that filled data from missing dataset from 1983 to 1987 (missing 5 years) by using modified WGEN with data filling algorithm is similar to historical measure ed data in the same years. This results show that modified WGEN is useful tool in the aspect of data filling for incomplete meteorological datasets.

4. CONCLUSION

The modified WGEN model adequately preserved means and standard deviations of periodic and daily precipitation amounts. The distribution of measured precipitation data was reasonably well matched by the modified WGEN because it has ability to generate the extreme condition and particularly precipitation pattern.

However, in case of WGEN, the use of exponential distribution cannot overcome some of special situations and it has not function for missing data filling.

This study found considerable strengths and weakness in both of weather generator such as classic WGEN and modified WGEN.

In San Antonio, Texas, WGEN consistently underestimated precipitation amounts; on the other hand, generated data were actually replicated as well with modified WGEN.

The data filling version of the weather generator, modified WGEN, is useful tool. It is processing from the equation of Matalas for first order autoregressive process on a multidimensional state with known cross and auto correlations among state variables. The parameters of the equation of Matalas are derived from existing dataset and derived parameters are adopted to fill data.

REFERENCES

Bailey, N.T. J., *The Elements of Stochastic Processes*, p.39, John Wiley, New York, 1964.

Caskey, J. E., Jr., A Markov chain model for the probability of precipitation occurrence in intervals of various lengths, *Mon. Weather Rev.*, 91, 298-301, 1963

Evans, J. R, and D.L. Olson. 2002. *Introduction to Simulation and Risk Analysis*. 2nd ed.

Table 3. Statistics analysis of precipitation amounts

	Measured data	Filled data from modified WGEN
Mean(in)	1.142	1
Std	0.81	0.92
Skewness	1.387	1.42
Kurtosis	1.705	1.758

Upper Saddle River, N.J.: Prentice Hall.

Feyerherm, A.M. and Bark, L.D., 1965. Statistical methods for persistent precipitation pattern. *J. Appl. Meteorol.*, 4, 320–328.

Garbutt, D.J., Stern, R.D., Dennet, M.D. and Elston[1], J., 1981. A comparison of rainfall climate of eleven places in West Africa using a two-part model for daily rainfall. *Arch. Met. Geoph. Biokl., Ser. B.,* 29, 137–155.

Gabriel, R., and J. Neuman, A Markov chain model for daily rainfall occurrence at Tel Aviv, Israel, *Q. J. R Meteorol, Soc.*, 88, 90-95, 1962

Hopkins, J. W., and P. Robillard, Some statistics of daily rainfall occurrence for the Canadian prairie province, *J. appl. Meteorol.*, 3, 600-602, 1964.

Lee, J.Y., and J.K.Brumbelow, *Stochastic simulation of daily weather variables*. Korea Water Engineering Research 2003.07 v.4, n.3, pp.111-126 1229-6503

Lane, L.J. and Nearing, M.A, 1989, Profile Model Documentation: USDA-Water Erosion Prediction Project, Hillslope Profile Version. NSERL Report No.2.National Soil Erosion Research Labortary, Purdue University, West Lafayette, IN.47907

Lowry, W.P. and Guthrie, D., 1968. Markov chains of order greater than one. *Mon. Weath. Rev.*, 96, 798–801.

Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. *Water Resour. Res.*, 3, 937–945

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resour.Res.*, 17, 182–190.

Richardson, C.W. and D.A.Wright. 1984. *WGEN, A model for Generating Daily Weather Variables, USDA ARS Bulletin No. ARS*-8. Washington, DC, U.S.A.: Government Printing Office.

Selvalingam, S. and Miura, M., 1978. Stochastic modelling of monthly and daily rainfall sequences. *Water Resour. Bull.*, 14, 1105–1120.

Stern, R.D. and Coe, R., 1984. A model fitting analysis of daily rainfall data. *J. Roy. Statist. Soc. A*, 147, Part 1, 1–34.

Todorovic, P., and D. A. Woolhiser, Stochastic model of daily rainfaill, Proceedings of the Symposium on Statistical Hydrology, *Misc. Publ.* 1275, pp. 223-246, U.S. Dep. of Agric., Washington, D. C., 1974.

Weiss, L. L., Sequence of wet and dry days described by a Markov chain probability model, *Mon. Weather Rev.*, 92, 169-176, 1964

Wilson, E.B., and M.M. Hilferty, 1931: Distribution of chi-square. *Proc. Nat. Acad. Sci., USA*, **17**, 684-688.

School of Environmental and Water Resource Engineering, Department of Civil Engineering, Texas A&M University, College Station, TX 77840