3rd International Conference - Water resources and wetlands. 8-10 September, 2014 Tulcea (Romania);
Available online at http://www.limnology.ro/wrw2016/proceedings.html
Editors: Petre Gâştescu, Petre Bretcan; ISSN: 2285-7923; Pages:171-176; Open access under CC BY-NC-ND license

# PREDICTING DAILY RUNOFF USING STOCHASTIC RAINFALL DATA GENERATION AND RAINFALL-RUNOFF MODELS

**Aminuddin Baki[1], Ismail Atan[2], Yasmin Ashaari[3], Jurina Jaafar[3], Mohd Bakri Samsudin[4]**
[1]Al Madinah International University, P.O. Box 7866, Shah Alam 40730, Malaysia, Tel: +60169958048
[2]Infrastructure University Kuala Lumpur, Unipark Suria, Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia
[3]Faculty of Civil Engineering, Universiti Teknologi MARA, Shah Alam 40450, Selangor, Malaysia Tel: +60355442000
[4]Envirab Services, 2A-2-1 Jalan Pusat Bandar 1A, Bandar Baru Bangi 43650 Selangor, Malaysia. Tel: +60123060086
*E-mail: aminbaki@msn.com, ismail585@yahoo.com, yasminashaari@yahoo.com, jucyber@yahoo.com, envirab@gmail.com*

**ABSTRACT**

Runoff data generally has short record lengths; however rainfall data with long records is available. Therefore, rainfall data will include wider range of conditions compared to runoff data. This paper will use stochastic rainfall data with conceptual rainfall-runoff model to generate synthetic runoff data for long term water resources design. In this paper, the ability of generating runoff data in the future using statistics of historical rainfall data will be analysed. For this purpose, rainfall data were divided into two samples: referred to as the Earlier Period and the Later Period. The rainfall data will be split according to the length of runoff data. The Later Period will be of the same length with the runoff data, whilst the Earlier Period will be all data before the runoff data. For the stochastic rainfall data generation, Transition Probability Matrices model will be used. For the rainfall-runoff modelling, Soil Dryness Index model will be used. Daily data will be used as the period of available data is long. Catchment used in this study was Kangaroo Valley, Australia. Comparison will be made between synthetic and recorded runoff data (daily, monthly and annual statistics; daily, monthly and annual maxima; monthly and annual minima; lag-one serial correlation coefficients). Daily, monthly and annual means and standard deviations were reproduced satisfactorily. However, lower values of skews tended to be generated indicating more normally distributed data were generated. Daily, monthly and annual maxima generated tend to be higher than the recorded values. Monthly and annual minima generated tend to be lower than the recorded values. This behaviour was desirable since it gave values of runoff outside the range of recorded runoff, which may not cover a complete range of possible values. The lag-one serial correlation coefficient was satisfactory as it is of the same range (slightly lower) as the recorded values. Generally, the use of historical rainfall data (Earlier Period) statistical properties to generate data for the future (Later Period) and then applying rainfall-runoff model to generate runoff data was found to be satisfactory.

**Keywords:** daily rainfall data generation, daily rainfall-runoff modelling, soil dryness index model, split sample, transition probability matrices model

## 1 INTRODUCTION

Runoff data generally has short record lengths; however rainfall data with long records is more widely available (Subramanya 1994). Therefore, rainfall data will include wider range of conditions compared to runoff data. If the rainfall data can be used to generate long synthetic records of rainfall and rainfall-runoff model is used to produce synthetically long runoff data, the issue of short runoff record can be overcome (Boughton 1965). The objective of this paper is to test the viability of this hypothesis by using a split-sample test (Baki 2009), where rainfall records were divided into two portions: the Earlier Period, which will be used to stochastically generate long term synthetic rainfall records; the Later Period, which is of the same period as the runoff records, will be used in the calibration of the rainfall-runoff models to find the optimum set of rainfall-runoff model parameters. These parameters will be used for the rainfall-runoff modelling using synthetic rainfall data.

### 1.1 Catchment

The catchment selected for this study is Kangaroo Valley, which is located about 150 km south of Sydney, and about 50 km west of the east coast of New South Wales, Australia. The map is shown in Figure 1 and catchment characteristics are as listed below (Baki 1996):

- The National Index reference is 215220
- The catchment area is 330 km$^2$.
- The length of the stream (Kangaroo River) is 34.5 km.
- The average slope of the Kangaroo River is 1.35% or 135 in 10,000.
- The annual rainfall for Kangaroo Valley is 1,629.0 mm.
- The annual runoff from the catchment is 934.2 mm.
- The annual pan evaporation is 1,773.4 mm.

• The climatic condition for this catchment is temperate.
• The vegetation in the area is a mixture of rainforest, hedgeland, sedgeland and grassland.

A total of 80 years of daily rainfall data were used. Both regionalised and single-site approaches have been satisfactorily used in rainfall data generation. Benson and Matalas (1967) used regionalised parameters in stochastic runoff data generation. Solomon (1976) used regionalised parameters as he found that regionalised parameters were more suitable than single site parameters because regionalisation reduced operational bias. Baki (1997) found that by using the average rainfall for the catchment, continuity of data could be obtained. Hernáez and Martin-Vide (2011), Mehrotra et al. (2012) and Camberlin et al. (2014) had used regionalised approach to satisfactorily modelled rainfall data. However, Mhanna and Bauwers (2011) had satisfactorily generated rainfall data using single-site approach. In this study, the regionalised approach had been adopted using catchment daily average rainfall. Therefore, the use of catchment average rainfall instead of individual stations allows for better approximations of rainfall stochastic properties and processes.

The location of the catchment is shown in Figure 1. The locations of the rainfall stations are shown in the enlarged inset of Figure 1. Catchment average rainfall was computed using Thiessen (1911) polygons of available data for the day. For the day with available data from all rainfall stations, the Thiessen (1911) polygons will be computed using 6 rainfall stations (as shown in the inset of Figure 1). For days that have missing data (for example if station 1 data is missing), the Thiessen (1911) polygons will be computed using the available data only, namely stations 2, 3, 4, 5 and 6. Similarly if data from stations 1 and 2 are missing, then the Thiessen (1911) polygons will be computed using the available data from stations 3, 4, 5 and 6. There are different polygons for different sets of missing data.


**Figure 1.** Study area (after Baki 1996)

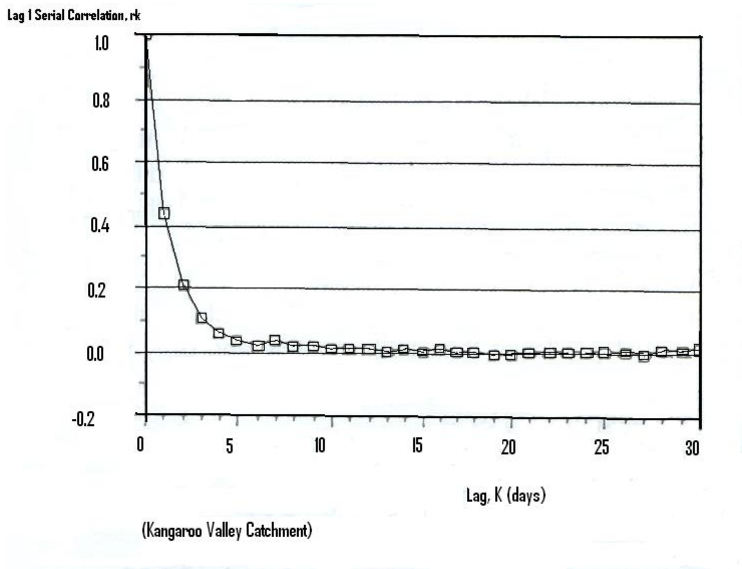
Statistics of daily rainfall for this catchment are shown in Table 1. Table 1 shows that the overall means, standard deviations, skew and coefficient of variations of daily rainfall for this catchment are 4.4mm, 15.6mm, 8.1mm and 3.5mm, respectively. The ratio of skew to coefficient of variation is 2.3, which is close to 2 indicating that gamma distribution can be used to approximate the rainfall distribution (Baki 2002).

**Table 1.** Recorded Daily Rainfall Statistics

| Month | Mean (mm) | Std. Dev. (mm) | Skew (□) | Coeff. Var. ($C_v$) | $\gamma/C_v$ |
|---|---|---|---|---|---|
| Jan | 4.8 | 16.4 | 11.3 | 3.4 | 3.3 |
| Feb | 5.5 | 17.5 | 7.4 | 3.2 | 2.3 |
| Mar | 5.8 | 18.2 | 6.3 | 3.1 | 2.0 |
| Apr | 4.9 | 16.5 | 7.0 | 3.4 | 2.1 |
| May | 4.8 | 17.7 | 7.9 | 3.7 | 2.1 |
| Jun | 6.1 | 19.3 | 5.5 | 3.1 | 1.8 |
| Jul | 4.6 | 18.0 | 8.3 | 3.9 | 2.1 |
| Aug | 3.1 | 11.4 | 8.2 | 3.7 | 2.2 |
| Sep | 3.1 | 9.9 | 6.5 | 3.2 | 2.0 |
| Oct | 3.7 | 15.0 | 9.2 | 4.1 | 2.3 |
| Nov | 2.9 | 8.9 | 6.8 | 3.0 | 2.2 |
| Dec | 4.1 | 12.2 | 7.3 | 3.2 | 2.3 |
| **Overall** | **4.4** | **15.6** | **8.1** | **3.5** | **2.3** |

(after Baki 1997)

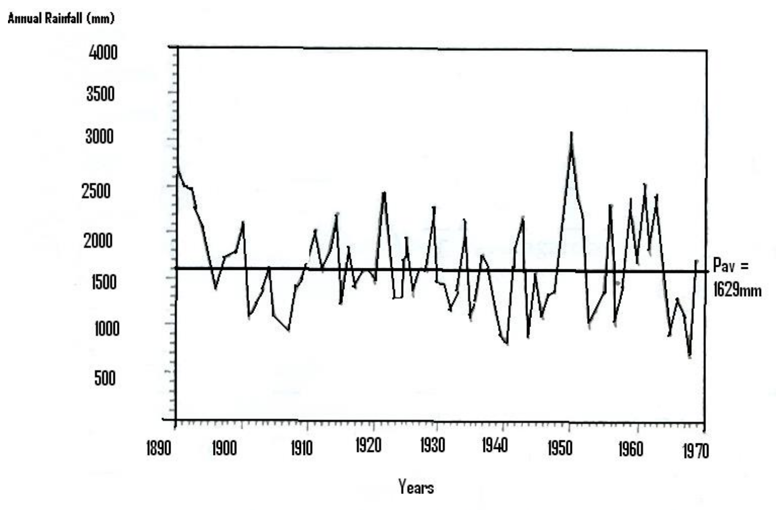
Figure 2 shows a plot of serial correlation coefficient ($r_k$) plotted against the corresponding lag (k). The lag-one value is: $r_1 = 0.436$, while the other $r_k$ values are less than half $r_1$ (Baki, 1997). Fisher (1958) suggested a value of $r_k$ of 0.349 as the conventional minimum value for stochastic analysis of time series. The lag-one serial correlation coefficient (r1) was shown to be satisfactory for this catchment, while the other rk values are much lower than the suggested conventional minimum value. Lag-one correlation was adopted for this paper (Baki, 1997).




(Baki 1997)

**Figure 2.** Plot of Serial Correlation against Lag

Figure 3 shows the plot of annual rainfall values (Baki 1997). No apparent trend can be observed in the values of annual rainfall for this catchment. Therefore, the random variations are assumed to continue into the future. The stochastic rainfall data generation is therefore assumed to be able to reproduce these random variations (Baki 1997). Jaafar et al. (2016) found that the Transition Probability Matrices were the most satisfactory stochastic model for this catchment. The finding was consistent with the findings of Srikanthan and MacMahon (2005) and Srikanthan et al. (2005). Thus, Transition Probability Matrices Model was adopted for the Stochastic Daily Rainfall Data Generation part of the study.




(Baki 1997)

**Figure 3.** Plot of Annual Catchment Rainfall

Figure 4 illustrated the concept of split sample approach. The split sample approach will be used to test the viability of using historical data to generate data in the future.

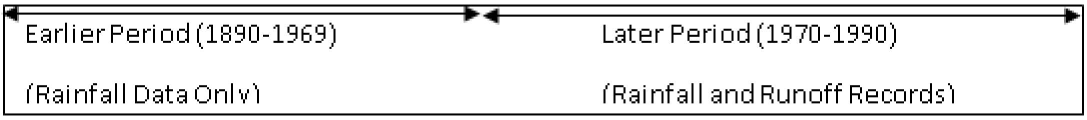



**Figure 4.** Period of Analysis

### 1.1 Rainfall-Runoff Modelling

Rainfall-runoff modelling will be conducted over the Later Period as shown in Figure 4. Baki (2006) compared various rainfall-runoff models on the Kangaroo Valley catchment (Figure 1). Baki (2006) found that Soil Dryness Index (SDI) Streamflow Yield Model was the most satisfactory rainfall-runoff model for this particular catchment (Kuczera 1988). Thus, SDI Model will be adopted for the rainfall-runoff modelling part of this study.

## 2 RESULTS AND DISCUSSIONS

Comparisons were made between the synthetic runoff data produced by the combined stochastic-deterministic approach against the recorded runoff data. Tables 2 and 3 show comparisons of daily, monthly and annual statistics; daily, monthly and annual maximum; monthly and annual minimum; and lag-one serial correlation coefficients ($r_1$).

**Table 2.** Daily Runoff Statistics

| MON | Means (mm) | | | Std Dev (mm) | | | Skews | | | Maxima (mm) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rec | Pred | Syn | Rec | Pred | Syn | Rec | Pred | Syn | Rec | Pred | Syn |
| Jan | 1.7 | 1.6 | 1.8 | 4.8 | 4.7 | 5.1 | 6.3 | 4.5 | 5.5 | 53.3 | 41.0 | 58.4 |
| Feb | 3.0 | 3.1 | 3.0 | 9.7 | 10.7 | 10.9 | 7.9 | 7.8 | 9.7 | 126.1 | 133.4 | 302.3 |
| Mar | 4.5 | 4.4 | 4.6 | 18.0 | 17.7 | 16.7 | 9.7 | 8.2 | 8.3 | 251.5 | 216.3 | 424.0 |
| Apr | 3.6 | 3.5 | 4.3 | 10.1 | 10.1 | 11.4 | 7.2 | 7.3 | 5.2 | 113.2 | 160.7 | 176.0 |
| May | 2.9 | 2.8 | 3.1 | 8.6 | 9.5 | 8.9 | 9.5 | 7.3 | 6.9 | 140.0 | 118.0 | 219.1 |
| Jun | 3.9 | 3.3 | 3.3 | 14.4 | 12.9 | 11.3 | 12.9 | 7.3 | 8.0 | 276.5 | 207.7 | 288.8 |
| July | 1.7 | 1.3 | 1.3 | 6.1 | 5.6 | 5.4 | 12.5 | 11.0 | 15.0 | 106.4 | 80.0 | 252.1 |
| Aug | 2.7 | 2.4 | 2.2 | 16.2 | 21.0 | 14.6 | 11.1 | 13.4 | 11.4 | 269.4 | 355.6 | 497.6 |
| Sept | 1.0 | 1.0 | 1.2 | 3.4 | 3.5 | 4.5 | 12.4 | 8.4 | 9.6 | 58.4 | 53.8 | 84.4 |
| Oct | 2.0 | 1.9 | 2.2 | 8.2 | 9.0 | 9.2 | 9.2 | 9.3 | 9.0 | 115.3 | 118.9 | 236.8 |
| Nov | 1.8 | 1.7 | 1.9 | 6.3 | 6.6 | 7.3 | 12.8 | 8.5 | 9.4 | 120.6 | 103.7 | 241.8 |
| Dec | 1.0 | 1.1 | 1.3 | 4.0 | 3.8 | 5.3 | 14.8 | 8.3 | 10.5 | 76.7 | 56.6 | 155.5 |
| Overall | 2.5 | 2.3 | 2.5 | 10.3 | 11.0 | 10.6 | 13.7 | 12.5 | 13.6 | 269.4 | 355.6 | 497.6 |
| Lag-one Serial Correlation Coefficient ($r_1$): | | | | | | | | | | 0.538 | 0.373 | 0.508 |

**Table 3.** Monthly and Annual Runoff Statistics

| | Means (mm) | | | Std Dev (mm) | | | Skews | | | Maxima (mm) | | | Minima (mm) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mon | Rec | Pred | Syn | Rec | Pred | Syn | Re | Pre | Sy | Rec | Pred | Syn | Re | Pre | Sy |
| Jan | 52.3 | 56.9 | 78.9 | 78.9 | 65.3 | 64.6 | 2.8 | 1.9 | 1.5 | 350.9 | 268.8 | 421.0 | 1.0 | 0.6 | 0.2 |
| Feb | 84.3 | 86.1 | 84.2 | 117.4 | 101.9 | 99.5 | 2.4 | 1.2 | 2.0 | 501.2 | 359.8 | 696.7 | 4.3 | 2.5 | 0.2 |
| Mar | 138.7 | 135.4 | 141.3 | 164.1 | 143.0 | 176.4 | 1.6 | 1.4 | 2.2 | 623.4 | 547.6 | 1659.4 | 10.0 | 5.6 | 0.1 |
| Apr | 108.7 | 105.1 | 129.6 | 117.4 | 118.1 | 132.4 | 1.1 | 1.5 | 1.4 | 382.6 | 447.8 | 733.7 | 3.9 | 3.3 | 0.1 |
| May | 88.6 | 87.2 | 96.7 | 86.0 | 98.4 | 108.9 | 1.1 | 1.6 | 1.9 | 283.3 | 397.1 | 640.3 | 8.4 | 2.0 | 0.6 |
| Jun | 116.6 | 100.0 | 98.2 | 158.0 | 138.0 | 108.9 | 2.0 | 1.8 | 1.4 | 595.0 | 452.3 | 529.9 | 8.1 | 1.2 | 1.3 |
| July | 51.5 | 41.5 | 40.5 | 57.2 | 51.0 | 48.2 | 1.4 | 1.8 | 2.2 | 185.5 | 158.5 | 348.7 | 4.8 | 0.7 | 0.8 |
| Aug | 84.4 | 75.4 | 67.1 | 156.1 | 137.4 | 156.6 | 2.2 | 2.1 | 2.8 | 582.6 | 452.3 | 529.9 | 4.3 | 1.0 | 0.5 |
| Sep | 30.6 | 29.8 | 36.0 | 34.4 | 36.2 | 51.7 | 2.0 | 1.6 | 2.0 | 127.1 | 118.0 | 329.6 | 2.2 | 0.6 | 0.3 |
| Oct | 62.1 | 60.3 | 68.4 | 85.0 | 91.1 | 110.3 | 1.8 | 1.7 | 2.2 | 292.1 | 289.2 | 812.5 | 1.2 | 0.3 | 0.5 |
| Nov | 52.6 | 52.2 | 57.2 | 57.6 | 56.5 | 76.3 | 1.5 | 1.6 | 1.8 | 227.8 | 217.3 | 559.2 | 1.5 | 0.3 | 0.3 |
| Dec | 30.7 | 34.0 | 41.7 | 39.3 | 39.5 | 56.2 | 2.8 | 1.8 | 2.2 | 181.6 | 140.1 | 315.6 | 0.4 | 1.1 | 0.2 |
| All | 75.1 | 71.5 | 76.5 | 108.4 | 100.1 | 117.3 | 2.6 | 2.2 | 3.6 | 623.4 | 547.6 | 1659.4 | 0.4 | 0.3 | 0.1 |
| Ann | 901.2 | 858.2 | 907.9 | 509.9 | 496.3 | 442.6 | 0.8 | 0.7 | 0.5 | 2142.5 | 2167.0 | 2408.0 | 178.3 | 210.0 | 99.1 |

Three-way comparisons were made between Recorded Runoff, Predicted Runoff (results of rainfall-runoff modelling using recorded rainfall in the Later Period - Baki 2006) and Synthetic Runoff (results of rainfall-runoff modelling using stochastically generated rainfall data, Jaafar et al., 2016).

Tables 2 and 3 show that daily, monthly and annual means were satisfactorily reproduced by the combined stochastic-deterministic approach. The values of standard deviations were also satisfactory but lower values of skews tend to be generated indicating that the generated data are more normally distributed compared to the recorded data.

Maximum values of runoff (daily, monthly and annual) generated tend to be higher than the recorded values, while minimum values tend to be lower. This behaviour is desirable since it gave values of runoff outside the range of recorded runoff, which may not cover a complete range of possible values. The value of lag-one serial correlation coefficient $r_1$ was satisfactory.

## 3 CONCLUSIONS

The combined stochastic-deterministic approach was shown to be able to reproduce the statistical properties of the recorded runoff data. The stochastic-deterministic approach can be considered satisfactory and have the potential to be used in catchments lacking in runoff data.

## REFERENCES


Baki, A.B.M. (1996), Objective Functions in the Optimisation of Daily Rainfall-Runoff Modelling, *JURUTERA: The Monthly Bulletin of the Institution of Engineers Malaysia*, No.**9**, Sept.1996, pp.11-15.

Baki, A.B.M. (1997), Stochastic Rainfall Data Generation Using Lag-One Markov Chain Model, *Journal of the Institution of Engineers Malaysia*, **58 (3)**,55-61.

Baki, A.B.M. (2002), Stochastic Rainfall Data Generation Using Two-Step Markov Chain Model: A Case Study, *Proc. of the 20th. Conference of ASEAN Federation of Engineering Organisations (CAFEO20)*, Phnom Penh, Cambodia, 2-4 September 2002, **1**,85-92.

Baki, A.B. M. (2005), Stochastic Rainfall Data Generation Using Transition Probability Model, *Proceedings of the Seventh Annual IEM Water Resources Colloquium 2005*, 18 June 2005, The Institution of Engineers Malaysia, Petaling Jaya, Malaysia, pp.9-1 – 9-9.

Baki, A.B.M. (2006), Rainfall-Runoff Modelling: Comparison of Various Models Performances, *Proceedings of the IWA World Water Congress & Exhibition 2006*, 10-14Sept. 2006, Beijing, China, pp.101-102.

Baki, A.B.M. (2009). Use of Split Sample Approach in Evaluation of Stochastic Daily Rainfall Generation Models. *Journal of the Institution of Engineers Malaysia,* **69**(1), 61-65.

Benson, M.A. and Matalas, N.C. (1967), Synthetic Hydrology Based on Regional Statistical Parameters, *Water Resources Research*, **3**(4), 931-935

Boughton, W.C. (1965), *A new simulation technique for estimating catchment yield*, Report/Water research laboratory. University of New South Wales, Australia No.**78**, NN09488.

Camberlin, P., Gitau, W., Oettli, P., Ogallo, L., Bois, B. (2014), Spatial interpolation of daily rainfall stochastic generation parameters over East Africa, Climate Research **59**(1), 39-60.

Fisher, R.A. (1958), *Statistical Methods for Research Workers*, 13th Edition, Oliver & Boyd, London.

Hernáez, P.F-A. and Martin-Vide, J. (2011), Regionalization of the probability of wet spells and rainfall persistence in the Basque Country (Northern Spain), *International Journal of Climatology*, October 2012, **32**(12), 1909–1920.

Jaafar, J., Baki, A., Abu Bakar, I.A., Tahir,W., Awang, H., Ismail, F. (2016), Evaluation of Stochastic Daily Rainfall Data Generation Models, In: Wardah Tahir, Prof Ir Dr Sahol Hamid Abu Bakar, Marfiah Ab. Wahid, Siti Rashidah Mohd Nasir and Wei Koon Lee (Editors), *Proceedings of the International Symposium on Flood Research and Management* 2015, Springer Singapore, pp. 203-220, DOI 10.1007/978-981-10-0500-8_17.

Kuczera, G. (1988), On the Revision of the Soil Dryness Index Streamflow Yield Model, *Civil Engineering Transactions, I.E.Aust.,* CE, **30**(2), pp.79-86.

Mhanna, M. and Bauwens, W. (2011), Stochastic single-site generation of daily and monthly rainfall in the Middle East, *Meteorological Applications*, March 2012, **19**(1), 111–117.

Mehrotra, R., Westra, S.P., Sharma, A., Srikanthan, R. (2012), Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach, Water Resources Research, **48**:W01536.
Solomon, S. (1976), Parameter Regionalisation and Network Design, In: Shen, H.W. (ed.), *Stochastic Approaches to Water Resources*, Colorado State University Press, Fort Collins, CO, USA, pp.12.1-12.37.
Srikanthan, R. and McMahon, T.A. (2005), Automatic Evaluation of Stochastically Generated Rainfall Data, *Australian Journal of Water Resources*, **8**(2),195-201.
Srikanthan, R., Siriwardena, L., McMahon, T.A. (2005), Comparison of Two Daily Rainfall Data Generation Models, *Australian Journal of Water Resources*, **8**(2),203-212.
Subramanya, K. (1994). *Engineering Hydrology*, Tata McGraw-Hill Publishing Company Ltd. New Delhi, 355p.
Thiessen, A.H. (1911), Precipitation Averages for Large Areas, *Monthly Weather Review*, 1082 p.