International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:13, No:1, 2019

Disaggregation the Daily Rainfall Dataset into Sub-Daily Resolution in the Temperate Oceanic Climate Region

Mohammad Bakhshi, Firas Al Janabi

Abstract**—**High resolution rain data are very important to fulfill the input of hydrological models. Among models of high-resolution rainfall data generation, the temporal disaggregation was chosen for this study. The paper attempts to generate three different rainfall resolutions (4-hourly, hourly and 10-minutes) from daily for around 20-year record period. The process was done by DiMoN tool which is based on random cascade model and method of fragment. Differences between observed and simulated rain dataset are evaluated with variety of statistical and empirical methods: Kolmogorov-Smirnov test (K-S), usual statistics, and Exceedance probability. The tool worked well at preserving the daily rainfall values in wet days, however, the generated data are cumulated in a shorter time period and made stronger storms. It is demonstrated that the difference between generated and observed cumulative distribution function curve of 4-hourly datasets is passed the K-S test criteria while in hourly and 10-minutes datasets the P-value should be employed to prove that their differences were reasonable. The results are encouraging considering the overestimation of generated high-resolution rainfall data.

Keywords**—**DiMoN tool, disaggregation, exceedance probability, Kolmogorov-Smirnov Test, rainfall.

I. Introduction

THE precipitation plays a significant role in hydrological models because it is used as an indicator of past hydrological variability and takes part in hydrological models to predict the future [2], [3]. High-resolution rainfall accessibility is one of a major problem for hydrologists due to many reasons such as expenses and governmental policies. Therefore, since last decades, scientists have attempted to generate the real high-resolution rainfall data by many computer programs which they are based on statistical probabilities of frequency occurrence and volume quantification [4].

The concept of disaggregation rainfall data from low resolution to high resolution is applied to estimate model parameters and reproduce measured daily rainfall data by either sampling or conditioning the simulated rainfall data [8], [6]. Since 1973, when Valencia and Schaake presented a multivariate model as a linear combination of coarser rainfall data resolution and generated data for the first time, many models were invented to propose the method of preserving the statistical properties of a rainfall event [15].

The latest theory is scale invariance theory which presents two models, multiplicative random cascade (MRC) model and resampling model (method of the fragment). In MRC models, the weight of each cascade shows how the amount of each parent cascade is divided into subdivisions and type of weight distribution divides the MRC into canonical cascade models and microcanonical cascade models. In the canonical cascade models, the mass is preserved in an average of whole cascades during disaggregation process while in microcanonical models the exact mass is preserved in each cascade during the process [7], [11].

Olsson tested a microcanonical model on 17-hour time series and produced hourly rainfall volume [13] and Güntner et al. examined an MRC model in semi-arid tropical and temperate climates to disaggregate the daily resolution into hourly which illustrated that the rainfall properties is conserved better in semi-arid tropical climate [3]. Licznar et al. indicated that not only the canonical and microcanonical model underestimated the 5-minute maximum rainfall intensity but also canonical model cannot generate the time series with high intermittency across a range of timescale [8].

Lisniak et al. proposed a new method named parameterization which based on circulation patterns (CPs) in MRC approach to adjust the parameters crossing to observed and simulated rainfall time series [9]. The generated hourly dataset from daily resolution and indicated the brilliant statistical properties conservation of rainfall time series with CP method compare to simple MRC. The method of Fragment (MoF) is another model of scale invariance theory that is the most popular method for disaggregation of precipitation data. The term "Fragment" represents the ratio of sub-daily (e.g. hourly) to daily rainfall for a particular time-step to measure the disaggregated data; the fragment for each sub-daily time scale should multiple by the daily data being disaggregated to measure the sub-daily time scale values. This method needs both original and disaggregated database at desired timescale. In contrast to other methods, MoF needs much less amount of data of desired time scale compared to data that is disaggregated [14].

Many tools are applied to disaggregate the precipitation data such as Hyetos, MuDRain, and Castalia. In this research, the DiMoN, which is prepared by Lisniak et al. is employed [9]. This tool can extract the required parameters of the short time rainfall time series of the desired time-scale. On the other hand, it preserves the exact generated rainfall amount at

M. Bakhshi (M.Sc.) was with Technische Universität Dresden, Germany (corresponding author, phone: +989154792575; e-mail: bakhshi2250@gmail.com).

F. AL Janabi (PhD) is with Technische Universität Dresden, Germany (e-mail: firas.aljanabi@tu-dresden.de).

desired time scale to the observed rainfall. It means that it works with microcanonical MRC and MoF.

The aim of this study is disaggregating the rainfall data with daily resolution into 4-hourly, hourly and 10-minutes, and then verifying the result of generated data. Formerly, Lisniak et al. did the disaggregation from daily to hourly, but no one prepares the 10-minutes generated data set from daily by applying the DiMoN [9].

II. Case Study and Materials

This study is a single site study and the area which is chosen for the present study is called Lockwitz watershed which is located in the south-east of Dresden, the capital city of Sachsen state, one of east state of Germany.

The total area which is taken for the study extends to 28 ha with around 60% imperviousness. The study area is located at latitude of 50°59'33" North, longitude of 13°48'34" East and at an average altitude of 138 meters above mean sea level. The area has a slope from south to north with level varying from 150 to 130 meters and generally slopes towards Elbe River with average 2%.

The population of Lockwitz area was about 6000 capita, and the population density was more than 200 people per hectare, approximately.

The average rainfall amount of Lockwitz watershed is about 750 millimeter per year and average temperature and humidity are 9.15°C and 73%, respectively. The location and climate condition reveal that the case study is situated in a Temperate Oceanic Climate.

Fig. 1 indicates the map of Dresden and places of its rain gauges. The applied instrument for rain gauge is an OTT Pluvio rain gauge with 400 mm bucket orifice opening and height of 750 mm. The Stadtentwässerung Dresden is a governmental company which is responsible for hydrological data collection and maintains the instruments. The available rain dataset is 5-minutes resolution which begins from 01.01.1996 and extends to 01.02.2015 which means more than 19 years.

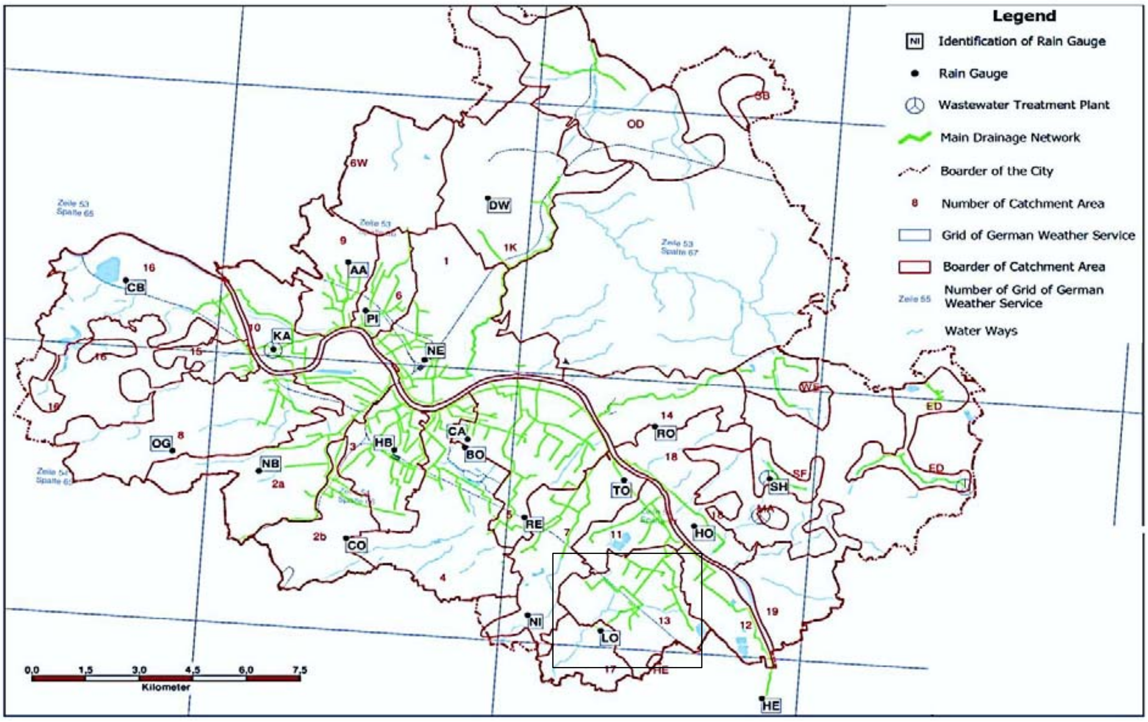


Fig. 1 Location of Rain Gauges in the City of Dresden, Germany (Stadtentwässerung) and the location of case study (black rectangle)

III. Methodology

The methodology consists of three parts:

1) Theoretical disaggregation that the DiMoN tool is based on,
2) Statistical tests to verify the result of synthesized rainfall data,
3) Statistical properties preservation of generated data compares to observed data.

In the scale invariance theory for disaggregation, some data of desired resolution are required for parameterization and calibration of the model by actual data. These high-resolution data indicate that the rainfall distribution algorithm along the dataset denotes the probability distribution of generators in each level of disaggregation process. For this reason, all required data (daily, 4-hourly, hourly and 10-minutes) are prepared by summing the 5-minutes rainfall data (the available data).

According to disaggregation process by DiMoN, the tool spends four levels (with branching number 2 and 3) to generate the finer data resolution. The discrepancy between

DiMoN and other tools is that in DiMoN the first branching number is 3 and the next cascades are divided by 2 while in the others all of the branching numbers are equal to 2. Hence, the DiMoN can convert daily dataset into hourly directly but in the other tools; they produce the 90-minutes or 45-minutes. [5]

According to cascade models, the rainfall volume of each child cell (R_{k+1}) is assigned by multiplying the rain volume of the parent cell (R_k) into weighting factor (W_k) of parent cell's level (1).

$$R_{k+1} = R_k \times W_k \tag{1}$$

The most challengeable part of microcanonical MRC approach is estimating the weighting generator (W) in parent wet cells. It is based on random process (MoF) and statistical probabilities which are achieved by extracting parameters of observed time series in desired resolution.

Corresponding to DiMoN disaggregation process, with two procedures the 4-hourly dataset is generated. First, the tool needs 96 hourly (4-days) resolution which is prepared by summing the available data (5-minutes), then set it as initial input in DiMoN and do the disaggregation process. Second, once the hourly dataset is generated from daily resolution, the disaggregated 4-hourly is produced by summing the generated hourly values. In the current study, the second method is chosen, although another scientific reason (for this selection) will explain in the next section.

The procedure for disaggregation the 10-minutes resolution dataset required an innovative method. According to cascade process, the tool can extrapolate the disaggregation process to each value. It means that the DiMoN releases the result of each input value after 4 steps are passed (cascade process). Hence, the generated 4-hourly dataset is inputted in DiMoN and repeat the process (exactly the same as an hourly dataset) to synthesize the 10-minutes dataset.

According to Ahamed et al., *2013*, it is very important to investigate the goodness of fit the model by summarizing the discrepancy between observed and generated rainfall dataset in all desired resolution [1]. Therefore, in this research, the Kolmogorov-Smirnov (K-S) test, which is a normality test for residual, is employed to examine the similarity between cumulative distribution function (CDF) diagrams of observed and disaggregated datasets with compare the maximum vertical distance between absolute CDF values (2) and the critical value of maximum vertical distance (3) under zero hypothesis (H_0) which is the two datasets has the same distribution.

$$D_{m,n} = \mathrm{Max}\left|F_{(x)} - G_{(x)}\right| \tag{2}$$

where "m" and "n" are the size of observed and generated datasets, "$D_{m,n}$" is the absolute maximum difference between CDF of two observed and generated datasets, "F(x)" and "G(x)"is the observed and generated dataset respectively.

$$D_{m,n,\alpha} = C_\alpha \sqrt{\frac{m+n}{m \times n}} \tag{3}$$

where C(α) is the Kolmogorov distribution value and extracted from Table I.

TABLE I
KOLMOGOROV DISTRIBUTION VALUE FOR EACH SIGNIFICANT LEVEL (ALPHA)

α	0.1	0.05	0.025	0.01	0.005	0.001
$C_{(\alpha)}$	1.22	1.36	1.48	1.63	1.73	1.95

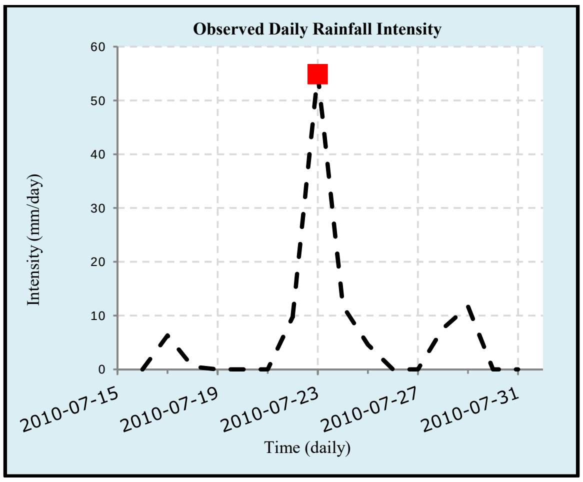
According to the bibliography, the α=0.05 is an acceptable value. In addition, in this research, the P-value is applied to test the mentioned hypothesis. If the P-value is close to 1, then the H_0 is acceptable and if P-value is smaller than α-value and close to zero, then it claims that the H_0 can reject.

The last part of this section includes statistical properties preservation between observed and generated data by some usual statistical tests such as mean value, variance and standard deviation and an empirical test which is Exceedance Probabilities (E-P). The frequency of rainfall event can be measured either by a return period (R-P) or an Exceedance probability, although the E-P is simply the mutual of R-P [10], [12].

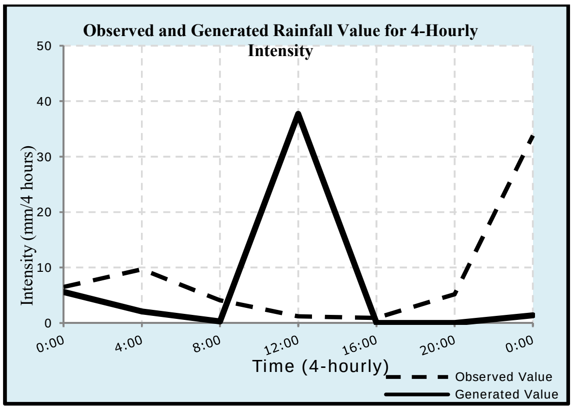
IV. RESULT AND DISCUSSION

The result consists of three parts: 1) A sample of hyetographs comparison between observed and generated rain data in each resolution, 2) Result of K-S test and statistical properties in each time steps, 3) E-P diagrams and table of statistical preservation for each desired time resolution.

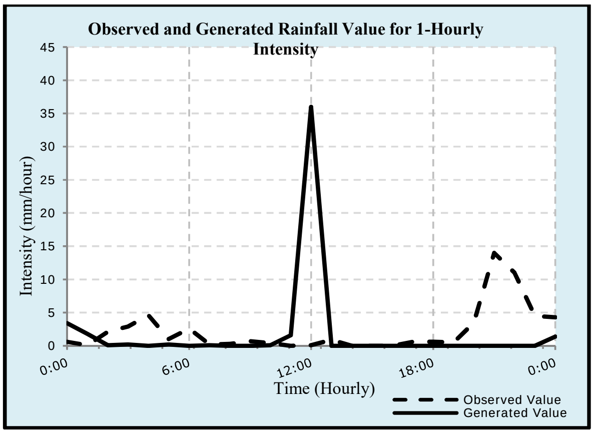
Fig. 2 (a)-(d) illustrate a sample of difference between hyetograph of generated and the observed value of rainfall in the desired date. The date was 23[th] of August, 2010, and it was chosen because it had the highest value within whole time series of available rain data. According to the point process theory and MRC method, generated values are not split uniformly on a wet day; in addition, the area under generated diagram is equal to the observed diagram.



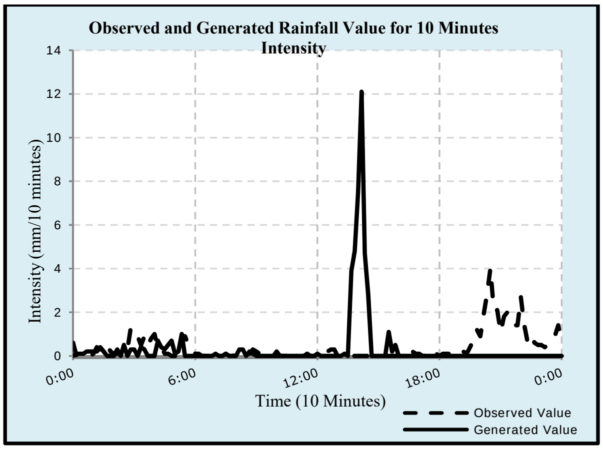
(a)



(b)



(c)



(d)

Fig. 2 An Example of Distribution Properties of Observed and Simulated Rain Data with Daily (a), 4-Hourly (b), 1-Hourly (c) and 10 Minutes (d) Intensity on August 23rd, 2010

It means that after disaggregation process, the rain value of observed and generated are equal during the day. The time period of "day" is important because the daily resolution is the first input dataset in the process. Obviously, in real storm in all intensities, although a rain event has a peak and in most of the cases the rain has been fallen during a longer period of time, while in disaggregated rain data the values cumulate in a short time period, hence the peak of generated rain data is higher than observed data.

Table II reveals the result of K-S test. In the 4-hourly resolution, both K-S test and P-value prove that both generated and observed dataset have the same distribution, while in hourly and 10-minutes resolution datasets just P-values lead to accepting the H_0. One reason for this outcome is related to the tool (DiMoN) which was noted in the methodology section. The DiMoN increases the uncertainties and errors of produced data during disaggregation process while the summing operation can decrease them. Consequently, the result of the generated 4-hourly dataset is accepted directly from K-S test while in hourly and 10-minutes datasets the result should be proved by P-value confirmation. The other reason belongs to the essence of disaggregated data which is cumulated in a short time period. On this account, the vertical difference between observed and generated data is getting larger in finer resolution compare to the coarser resolution dataset.

TABLE II
KOLMOGOROV-SMIRNOV TEST RESULTS FOR LOCKWITZ (LO) STATION IN DIFFERENT TIME RESOLUTION FROM 01.01.1996 UNTIL 01.02.2015

	4-Hourly	**Hourly**	**10-Minutes**
Test Statistics (absolute difference of CDFs)	0.0070	0.0087	0.0025
Critical Value at α=5%	0.0094	0.0047	0.0019
The result of the Test	**Accept the H_0**	Reject the H_0	Reject the H_0
Mean Value of Observed CDF	0.9967	0.9987	0.9997
Mean Value of Generated CDF	0.9968	0.9987	0.9997
Standard Deviation of Generated CDF	0.0135	0.0056	0.0014
Adjusted standard Deviation of Generated CDF	0.00042	0.00026	0.00009
P-Value	0.5323	**0.5006**	**0.4902**

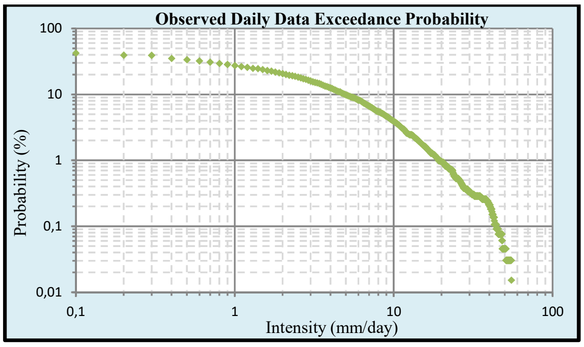
Table III expresses the result of general statistical properties. The sum row of this table claims that the basic of MRC is followed. It means that, during the disaggregation process, the tool preserves the daily value of each wet day. The variance, standard deviation, and skewness are approximately constant in each resolution and getting decreased by increasing the resolution. The correlation coefficient values claim the explanations of hyetographs because they are very small values (near to zero) which mean that the DiMoN cannot follow the rainfall pattern during the day.

Figs. 3 (a)-(d) represent the Exceedance Probabilities (E-P) for historical and disaggregated rainfall data in each desired time resolution. They are logarithmic downward diagrams in which the horizontal and vertical axis are expressing the intensity (mm/time step) and probabilities (%), respectively. All diagrams consist of three parts. In the first part, the spots separate each other which means that in low intensities, although the probability of occurrence is greater than higher intensities, the variation of probabilities is as low as higher intensities. In the second part, the probabilities of intensities are very dense whereas the diagram gets a curve shape. The last part belongs to high intensities which contain separated groups of spots. According to these diagrams, the time step resolution and probabilities of occurrence have the same trend in a constant intensity.

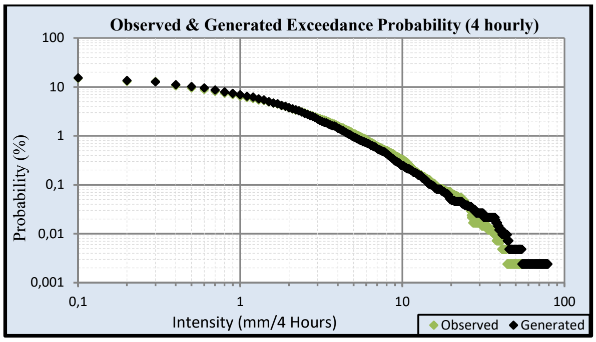
TABLE III
DESCRIPTIVE COMPARISON STATISTICS BETWEEN OBSERVED AND GENERATED RAINFALL IN DIFFERENT TIME RESOLUTION IN LOCKWITZ (LO) STATION FROM 01.01.1996 UNTIL 01.02.2015

	Daily	4-Hourly		Hourly		10-Minutes	
	Observed	**Observed**	**Generated**	**Observed**	**Generated**	**Observed**	**Generated**
Sum	11427.2	11427.3	11427.2	11427.3	11427.3	11427.3	11427.2
Mean Value (mm)	1.64	0.27	0.27	0.068	0.068	0.01138	0.01138
Variance	20.17	1.68	1.74	0.22	0.26	0.01427	0.02234
Standard Deviation	4.50	1.30	1.32	0.47	0.51	0.1194	0.1494
Coefficient of Variance (%)	2.74	4.75	4.83	6.92	7.45	10.49	13.13
Skewness	10.73	12.85	17.56	23.01	29.09	49.70	61.11
Correlation Coefficient (R)	-	0.275		0.137		0.037	

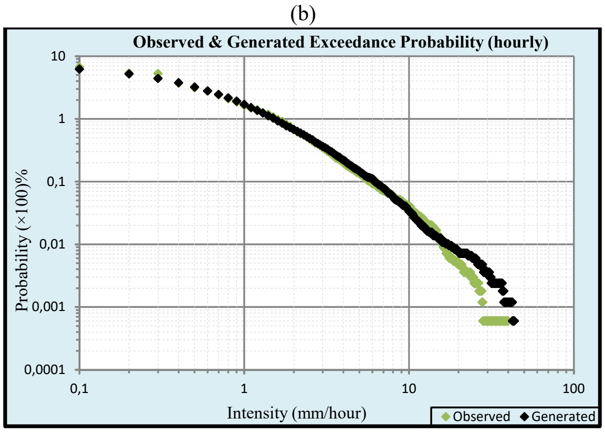
Formerly, it was expressed that the DiMoN increases the errors and uncertainties during disaggregation process. One confirmation for this claim is that the difference between probabilities of observed and generated rainfall data, in the same intensity get increased by increasing the resolution. On the other hand, in high-intensity part of E-P diagrams, the difference of generated and observed rainfall data is getting large where at hourly and 10-minutes resolution, this difference is more obvious compared to 4-hourly resolution. The reason for this differentiation is because of the employed methodology which means the hourly and 10-minutes resolution datasets are generated directly with disaggregation process with one and two times of repeat the process respectively, while the 4-hourly dataset is achieved indirectly from the mentioned process.



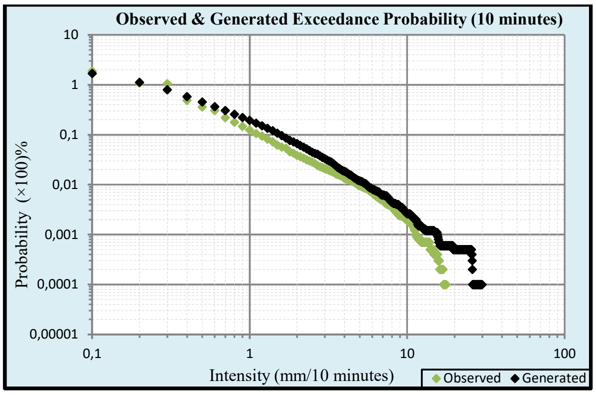
(a)



(b)



(c)



(d)

Fig. 3 Comparison of the result of Exceedance Probabilities of observed and generated rainfall data in different intensities for different time steps daily (a), 4-hourly (b), hourly (c) and 10-minutes (d)

V. Conclusion

In this paper, a framework was described where continues different resolutions (4-hourly, hourly and 10-minutes) rainfall from daily values can be generated at Mediterranean climate. The basis of this approach is microcanonical MRC method in which the wet daily values are constant after disaggregation procedure. The obtained result was compared to the observed values with Exceedance probabilities diagrams and validated with some statistical tests. Presumably, the rainfall produced by this structure would be adequate for input to hydrological models such as rainfall-runoff and infiltration-runoff models.

The approach sought to address several important limitations associated with actual rainfall pattern. First, the generated rainfall values created a time series with higher intensity events, especially in high resolutions, thus the generated rainfall values are overestimated, which should be noticed during the process of other hydrological applications. Second, in many dissertations, two different time series are created, one of them for disaggregation process and the other for result validation, but in this study, the whole period of the available dataset used for disaggregation process. Therefore, in future studies, this framework can be applied to two different datasets. Finally, the proposed structure worked-well with rain data at Temperate Oceanic regions; while maybe in other climates do not work-well which should be inquired.

Acknowledgment

We would like to express our warm thanks to Prof. Peter Krebs, the holder of the chair for urban water management institute of Environmental Sciences Department of Technische Universität Dresden, who support us a lot to write this scientific paper. Moreover, we should thank to Dipl. –Ing. D. Lisniak, Dipl. –Ing. T. Käseberg, Dipl. –Ing. B. Helm and M.Sc. M. Käseberg who help us in data analysis part of this dissertation.

References

[1] Ahamed, S., Piantadosi, J., Agrawal, M. and Boland, J., 2013. Generating synthetic rainfall using a disaggregation model (Doctoral dissertation, Modelling and Simulation Society of Australia and New Zealand).
[2] Gaume, E., Mouhous, N. and Andrieu, H., 2007. Rainfall stochastic disaggregation models: Calibration and validation of a multiplicative cascade model. Advances in Water Resources, 30(5), pp.1301-1319.
[3] Güntner, A., Olsson, J., Calver, A. and Gannon, B., 2001. Cascade-based disaggregation of continuous rainfall time series: the influence of climate. Hydrology and Earth System Sciences Discussions, 5(2), pp.145-164.
[4] Gyasi-Agyei, Y. and Mahbub, S. P. B., 2007. A stochastic model for daily rainfall disaggregation into fine time scale for a large region. Journal of Hydrology, 347(3-4), pp.358-370.
[5] Hanaish, I. S., Ibrahim, K. and Jemain, A. A., 2011. Daily rainfall disaggregation using HYETOS model for Peninsular Malaysia. matrix, 2, p.1.
[6] Kilsby, C. G., Jones, P. D., Burton, A., Ford, A. C., Fowler, H. J., Harpham, C., James, P., Smith, A. and Wilby, R.L., 2007. A daily weather generator for use in climate change studies. Environmental Modelling & Software, 22(12), pp.1705-1719.
[7] Koutsoyiannis, D., 2003, May. Rainfall disaggregation methods: Theory and applications. In Workshop on Statistical and Mathematical Methods for Hydrological Analysis, Rome (Vol. 5270, pp. 1-23).
[8] Licznar, P., Łomotowski, J. and Rupp, D. E., 2011. Random cascade driven rainfall disaggregation for urban hydrology: An evaluation of six models and a new generator. Atmospheric Research, 99(3-4), pp.563-578.
[9] Lisniak, D., Franke, J. and Bernhofer, C., 2013. Circulation pattern based parameterization of a multiplicative random cascade for disaggregation of observed and projected daily rainfall time series. Hydrology and Earth System Sciences, 17(7), pp.2487-2500.
[10] Mansell, M. G., 2003. Rural and urban hydrology. Thomas Telford.
[11] Molnar, P. and Burlando, P., 2005. Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. Atmospheric Research, 77(1-4), pp.137-151.
[12] Nathan, R., Jordan, P., Scorah, M., Lang, S., Kuczera, G., Schaefer, M. and Weinmann, E., 2016. Estimating the exceedance probability of extreme rainfalls up to the probable maximum precipitation. Journal of hydrology, 543, pp.706-720.
[13] Olsson, J., 1998. Evaluation of a scaling cascade model for temporal rain-fall disaggregation. Hydrology and Earth System Sciences Discussions, 2(1), pp.19-30.
[14] Sharma, A. and Srikanthan, S., 2006. Continuous rainfall simulation: A nonparametric alternative. In 30th Hydrology & Water Resources Symposium: Past, Present & Future (p. 86). Conference Design.
[15] Valencia R. D., Schaake Jr. J. C., 1973. Disaggregation Processes in Stochastic Hydrology. Water Resources Research, 9(3):580-585.