

*Fine Scale Rainfall Disaggregation in Greece using RBLRPM*

Sofia SKROUFOUTA1, Evangelos BALTAS2

National Technical University of Athens, Greece

email: sofiaskroufouta@chi.civil.ntua.gr

email: baltas@chi.civil.ntua.gr

Abstract

The fine-scale disaggregation of rainfall is an essential tool for hydrological applications because of the strong stochasticity that underlies natural hydrological phenomena and the scarcity of available data. In two study areas in Greece, the Random Bartlett-Lewis Rectangular Pulse Model (RBLRPM) is applied for two distinct distributions of the impulse intensity (Gamma and Exponential). The corresponding daily time series is divided into hourly ones, and the HYETOS-R package is used to assess the behavior of the alternative distributions based on the required statistical quantities.

1. Introduction

Data at fine temporal scales are essential to the accuracy and dependability of many hydrological applications, including the design of drainage systems, reservoirs, and flood control projects (Hanaish et al., 2013). These applications are based on understanding and modeling hydrological processes, and therefore they often require precise data as input that in many cases do not exist. Thus, the scarcity of precipitation data on fine time scales, particularly in Greece, is a serious issue. Hydrological applications therefore necessitate the use of stochastic rainfall models at various time scales (Olsson, Burlando, 2002; Hanaish, Ibrahim, 2011). Even though the stochastic simulation of precipitation is very interesting and intriguing, the finer the scale the more complex the problem becomes. This is because the structure of the observed time series appears to show intermittent zero and non-zero precipitation values. Modeling, therefore, using classical linear stochastic models becomes impossible, since the aforementioned discontinuous structure cannot be satisfactorily simulated and reproduced.

The models that best adapt the behavior of this hydrological process and are able to reproduce synthetic data with the same statistical behavior, appear to implicate the Poisson distribution (as Todorovic and Vevjevich (1969) first observed), with the point process models being the pioneers. To fully understand these models, it is essential to explore and comprehend the daily and intraday precipitation structure. Essentially, the variables that must be simulated for the application of the models under consideration are directly connected to the comprehension of individual rain events and their relationships; variables that are ignored in second-order linear models. The maximum quantity of rain per storm, the duration and intensity of a rain event, and the sequence of dry-rainy days are examples of such factors. In addition to the variables themselves, the degree of stochastic dependency of the variables' consecutive values, as well as the variables among themselves, is a significant aspect of the modeling that also significantly augments the computational load of the process.

In addition, point process models are stochastic procedures, which describe a sequence of random events in continuous space or time, as in the case of hydrological processes (Waymire, Gupta, 1981b). Accordingly, in the case of point process-based models for rainfall, this pattern is used for the mathematical analysis of the random variable, where it is associated with the onset time of individual rain events. These models consider each episode as an instantaneous event, with duration and intensity of rain, as well as time of occurrence. Furthermore, it is noted that most point process models ensure stochastic independence between the random variables of the time of occurrence, the duration, and the amount of rain of an episode, but also between the consecutive values of each variable (Koutsoyiannis, Xanthopoulos, 1990).

The evolution of point process models leads to the creation of rectangular pulse models. The primary observable characteristics of precipitation processes in time and space are essentially represented conceptually by these models, which are essentially simplified (Waymire, Gupta, 1981a). These models have the advantage of being able to accurately describe the precipitation process using a limited number of parameters, which allows for the deduction of other physical process properties (Onof, Wheater, 1993; Onof, Wheater, 1994). In the case of the Bartlett-Lewis Rectangular Pulse Model (BLRPM), the model derives from the Markovian model of rectangular pulses, treating pulses as clusters rather than individual ones. Like the point evolution model on which it is based, it also calculates the required parameters, relating them to the key statistical characteristics of the historical time series of each time scale. In addition, it accepts overlap, both between successive pulses and between successive rain episodes, defining the starting point of each pulse as the starting point of the previous pulse. An important, also, advantage of the BLRPM is that it manages to adequately capture both the structure of the selected historical statistical parameters, as well as the probability of dry for various time scales, for an exclusive set of parameter values for each month, ensuring the temporal independence of the model (Rodriguez-Iturbe et al., 1984).

Since in many cases in Greece (and worldwide), historical data is non-existent or available at the required fine scale, this paper implements the disaggregation of rainfall data, using the above-mentioned BLRPM for the production of synthetic time series. Essentially, an important part of the process is choosing the right precipitation simulation model. The present paper focuses on fine time scales, where the problem is more complex, as rainfall time series exhibit a strongly intermittent character and their variables exhibit strong dependence and variation. This makes linear models unsuitable for their simulation as it is already mentioned, unlike the BLRPM, as the results of this paper prove.

1. Material and Methods

The disaggregation algorithm implemented by the HYETOS-R package is based on the Koutsoyiannis and Onof (2001) algorithm. In order to accurately replicate the historical time series with a finer time scale, the model consists of a Bartlett-Lewis model, in this case the RBLRPM, and an adjusting procedure that modifies the shorter synthetic time series (Grygier, Stedinger, 1988). The process is repeated until the number of iterations crosses the chosen threshold or a smaller-scale synthetic time series deviating by a given threshold da (or less) is produced (Koutsoyiannis, Manetas, 1996). Applying the correction procedure to optimally adjust the statistical parameters once the synthetic time series is derived requires a substantial computational effort. This problem can be resolved by simulating the time series with shorter periods for rainy events as opposed to the full duration. The assumption that storms follow a Poisson distribution and that clustered rain days are independent are two of the highly realistic assumptions that the Bartlett-Lewis model is found to accept.

In more detail, the algorithm starts by dividing the input daily time series (either historical or synthetic) into clusters of non-zero rainy days, until their length equals a selected value. Then, the intensities of the generated rectangular pulses of the episodes are derived and the synthetic heights of rainy days are calculated, through discretization and summation. Between the first two steps, it is checked how much the sectional synthetic time series deviates from the cumulative one, to ensure that too high or too low synthetic rain heights do not dominate. If the deviation is greater than the allowed one, then the intensities for the existing pulses are generated again as in the first step, without affecting their time arrangement, while in the opposite case the synthetic daily rain heights resemble the real ones, and the sequence is assembled. Then, it is checked whether the number of repetitions of the check for the sequence exceeds the desired limit. If not, the generated sequence is discarded entirely and a new one is generated. If so, the group splits into two groups of length less than the aforementioned selected value and the procedure continues by recalculating the intensities of the generated rectangular pulses of the episodes and so on. Finally, since the iterative process is completed, a final check takes place, where in case the generated group has been divided into 2 or more subgroups, it joins them into one rainy day group.

The RBLRPM is an improved version of the BLRPM (Rodriguez-Iturbe et al., 1987a; Rodriguez-Iturbe et al., 1987b; Rodriguez-Iturbe et al., 1988), which differs in the way it models the random variable of pulse duration. The statistical parameters of the improved model are five. The parameter λ of the initial Poisson evolution describes the overall stochastic modeling process and is related to the starting time points ti of the events i. Since the arrival time of the first pulse is the same as the arrival time of the corresponding storm, the parameter βi of the second Poisson evolution refers to the starting time points of the pulses tij of each event. The Exponential distribution's parameter γi is linked to the random variable of the time interval νi, encompassing the beginning times of the pulses in every event i. The Exponential distribution's parameter ηi describes the random variable of the time interval wij of event i's pulses. The Exponential distribution parameter λX is associated with the random variable Xij, which represents the rain height during the event i pulses (Fig. 1).

** Fig. 1.** Temporal representation of the elements that characterize the rectangular intensity pulses of rain events (Kossieris et al., 2015)

In addition, an enhanced version of the evolutionary annealing-simplex algorithm (Efstratiadis, 2001; Efstratiadis, Koutsoyiannis, 2002) is used to estimate the unique set of parameters of the Barlett-Lewis model for the two study areas (Thessaloniki and Heraklion). This algorithm is a probabilistic heuristic optimization method that combines the effectiveness of hill-climbing techniques in convex regions with the robustness of rough response surface annealing simulation. The primary characteristic of the evolutionary annealing-simplex algorithm, which is based on the Simplex method, is that none of the moves are completely deterministic, which increases the possible rough search field of the iterative process. Concisely, the algorithm follows a process where the “temperature” is gradually reduced, according to an appropriate annealing cooling schedule, which is automatically adjusted during evolution. Consequently, the probability of replacing poor performers increases as the process gradually transitions from a random path to a local search. The recombination operator is based on the well-known simple descent transitions (Nelder, Mead, 1965). According to the relative objective function values at the vertices, the simplex is reflected, expanded, contracted, or contracted, where quasi-stochastic scale factors are used instead of constant ones.

To make the algorithm more flexible, additional transformations are introduced, i.e. multiple expansions in the direction of reflection, when a downward path is found (i.e., the slope of the function) and similar expansions, but in the opposite (upward) direction, to escape from the nearest local minimum. If any of the above transitions improve the value of the function, the new individual is created through deletion. The associated operator uses a random perturbation scheme outside the usual population range, as determined by the mean and standard deviation values of its coordinates.

1. Results and Discussion

The simulation occurs during the driest (non-zero) and wettest months in each study area. In these months, the behavior of the rainfall presents the most extreme statistical characteristics, hence it can be evaluated whether the model can satisfactorily understand the process’ statistical behavior, but also produce the corresponding stochastic and disaggregated time series. For each region, the statistical characteristics of the synthetic daily time series are first presented compared to the characteristics of the historical time series. While following are the corresponding results for the disaggregated hourly time series.

For Thessaloniki, the Gamma distribution appears to fit the best for the driest month (June), but the Exponential distribution appears to preserve better the historical statistical features of the time series for the wettest month (January). The structure of the autocorrelation coefficient is not satisfactorily preserved, even though the Gamma distributions' deviations are not great in the case of the dry probability. Maintaining the structure of the autocorrelation coefficient is a crucial requirement, but the deviations of the coefficients of variation and skewness are comparable to those of an Exponential distribution. For example, the structure of the skewness coefficient and the probability of dry conditions are fully preserved, when creating a daily synthetic time series of January assuming an Exponential distribution. The variation coefficient, in particular, appears to be overestimated by the model by 18%, with a 45% deviation from the historical coefficient in the case of skewness (Fig. 2a). In contrast, the coefficients of variation and skewness (37% and 38%) are underestimated and deviate more from the historical ones, the deviations of the synthetic statistical characteristics for June seem to be larger than those of the corresponding rainy month and already of the autocorrelation coefficient (Fig. 2b).



**Fig. 2.** Statistical Characteristics of Thessaloniki for: a) January, considering that the intensity of the pulses follows the exponential distribution, and b) June, considering that the intensity of the pulses follows the Gamma distribution (the diagrams for the volatility coefficient (left) and the asymmetry coefficient (right)).

In the case of Heraklion, there are two months (June and July) with zero precipitation in all years of the available historical data; therefore, the simulation takes place for the driest non-zero month, August, and the wettest January. Conversely, the Gamma distribution is proved to be more appropriate for both months in the Heraklion region, since it maintains more satisfactorily the coefficients of variability and asymmetry, in addition to the structure of the autocorrelation coefficient. Both distributions simulate dry probabilities equally well, fully preserving the structure of the skewness coefficient. For both the wettest and the driest month, the variation's deviation is less than 5% and the skewness is less than 15% (Fig 3).



**Fig. 3.** Statistical Characteristics of Heraklion, considering that the intensity of the pulses follows the Gamma distribution for: a) January and b) August (the diagrams for the volatility coefficient (left) and the asymmetry coefficient (right)),

The model satisfactorily maintains the conditions under study statistical characteristics at the various time scales. Notably, deviations show a slight escalation during the drier months, according to the comparison of the probability of rain over the entire time series, the coefficient of variability, and the asymmetry coefficient. These factors are evaluated at different time scales, ranging from 1 to 24 hours, as illustrated in Fig. 4. This figure highlights the model's ability to maintain the statistical integrity of the parameters under control, further emphasizing the robustness observed over these different time scales.

**Fig. 4.** Comparison of statistical characteristics of the synthetic and historical time series at different aggregation scales, for Thessaloniki, for January, where: a) the probability of dry diagram, b) the variability coefficient diagram, and c) the diagram of the asymmetry factor. The elements of the synthetic disaggregated time series are in green, while in red are the historical ones.

Even though the model achieves to maintain the essential statistical characteristics of the historical data in different time scales, it is crucial, for every model of natural processes, to examine whether the produced time series appears to be realistic in a continuous time. In both study areas, the synthetic hourly time series have a satisfactory and realistic behavior. As shown in Fig.5a, the model seems to produce more clustered rain events, when the probability of dry is high (for the driest months), while for the wettest months, the rain events seem to be better distributed.



**Fig. 5.** Diagram of the produced synthetic hourly time series for Thessaloniki: a) for June, for the gamma distribution, b) for January, for the exponential distribution.

Although the results in Fig. 5 refer only to the region of Thessaloniki, similar results exist in the case of Heraklion, with the average monthly rainfall (both for the driest and the wettest) being higher than in Thessaloniki. This is optative, however, in the case of a very dry month in other study areas, where the average precipitation is very low, the model might face difficulty in producing realistic time series.

So far, the model seems to perform adequately, concerning solely the production of synthetic time series in the same time scale as the historical data. It is evident to examine whether the same ability occurs in the case of disaggregation on a finer scale. Both in Thessaloniki and Heraklion, the model produces disaggregated synthetic time series close to the available historical ones (used only for comparison), with acceptable deviations in all the substantial statistical characteristics. The results for the region of Thessaloniki for June are displayed indicatively in Fig. 6, as this is when the most significant deviations occur. In the case of Heraklion, with the largest deviation occurring in the skewness and being only 7%, the model's performance in disaggregating the daily synthetic time series into hourly is also very impressive, given that the structure of the statistical characteristics under study is fully preserved.



**Fig. 6.** Statistical Characteristics for January for Thessaloniki, considering that the intensity of the pulses follows the exponential distribution, where: a) the diagram of the autocorrelation coefficient, b) the diagram of the probability of no rain and c) the diagrams for the coefficient of variability (left) and the coefficient of asymmetry (right) for all elements of the respective time series. The elements of the synthetic disaggregated time series are in green, while in red are the historical ones.

1. Conclusions

Based on the abovementioned, the RBLRPM can preserve the essential statistical properties that control the daily and intraday structure of the precipitation. The model's ability to simulate and allocate precipitation to multiple aggregation scales using a single set of parameters, regardless of the historical time series' time scale, is a significant advantage. Specifically, the outcomes of this study confirm that the RBLRPM is appropriate for the distribution of daily rainfall time series in the Greek region at finer time scales. In particular, at different aggregation scales, it can preserve the coefficients of autocorrelation, variability, and asymmetry for the entire time series, as well as, for the rainy days individually. Moreover, it maintains the probability of dry, which is another crucial feature of the precipitation time series. Given that the independent intensities of the rectangular pulses follow a Gamma distribution, the results also demonstrate that most of the months of the study areas in Greece are better simulated. Nonetheless, the time series can also be disaggregated using the Exponential distribution.

References

*Hanaish I. S., Ibrahim K. and Jemain A. A., On the potential of Bartlett Lewis rectangular pulse models for simulating rainfall in Peninsular Malaysia, Hydrological Sciences Journal, 58:8, 1690-1703, 2013, DOI: 10.1080/02626667.2013.838322*

*Olsson J., Burlando P., Reproduction of temporal scaling by a rectangular pulses rainfall model, Hydrological Processes, 16, 3, 0885-6087, 2002,* [*https://doi.org/10.1002/hyp.307*](https://doi.org/10.1002/hyp.307)

*Hanaish I. S. and Ibrahim K., Daily rainfall disaggregation using HYETOS model for Peninsular Malaysia, International Conference on Applied Mathematics, Simulation, Modelling, 2011 - Proceedings. 146-150*

*Todorovic P., Yevjevich V., Stochastic process of precipitation, Hydrol. Pap. Ser., No. 35, Colorado State Univ, Fort Collins, Colo, 1969*

*Waymire E. and Gupta V. K., The Mathematical Structure of Rainfall Representations: 2. A review of the Theory of Point Processes, Water Resources Research, 17, 1273-1285, 1981b*

*Koutsoyiannis D. and Xanthopoulos Th., A dynamic model for short-scale rainfall disaggregation, Hydrological Sciences Journal, 35 (3), 303–322, doi:10.1080/02626669009492431, 1990*

*Waymire E. and Gupta V. K., The Mathematical Structure of Rainfall Representations: 1. A review of the Stochastic Rainfall Models, Water Resources Research, 17, 1261-1272, 1981a*

*Onof C., Wheater H. S., Modeling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model, Journal of Hydrology, 149, 67-95, 1993*

*Onof, C. and Wheater H. S., Improvements to the modelling of British rainfall using a Modified Random Parameter Bartlett-Lewis Rectangular Pulse Model, Journal of Hydrology, 157, 177-195, 1994*

*Rodríguez - Iturbe I., Gupta V. K. and Waymire E., Scale Considerations in the Modeling of Temporal Rainfall, Water Resources Research, 20, 1611-1619, 1984*

*Koutsoyiannis D. and Onof, C., Rainfall disaggregation using adjusting procedures on Poisson cluster model, Journal of Hydrology, 246(1):109-122, 2001, DOI: 10.1016/S0022-1694(01)00363-8*

*Grygier J. C. and Stedinger J. R., Condensed disaggregation procedures and conservation corrections for stochastic hydrology, Water Resour. Res., 24( 10), 1574– 1584, 1988, doi:10.1029/WR024i010p01574*

*Koutsoyiannis D., Manetas A., Simple disaggregation by accurate adjusting procedures, Water Resources Research, 32 (7), 2105 – 2117, 1996*

*Rodríguez - Iturbe I., D. R. Cox, F.R.S. and V. Isham, Some models for rainfall based on stochastic point processes, Proc. R. Soc. Lond. A 417, 269 – 288, 1987a*

*Rodríguez - Iturbe I., B. Febres De Power and J. B. Valdes, Rectangular pulses point process models for rainfall: analysis of empirical data, Journal of Geophysical Research, 92, 9645 – 9656, 1987b*

*Rodríguez - Iturbe I., D. R. Cox, F.R.S. and V. Isham, A point process model for rainfall: further developments, Proc. R. Soc. Lond. A 417, 283 – 298, 1988*

*Kossieris P., Efstratiadis A,. Tsoukalas I. and Koutsoyiannis D., Assessing the performance of Bartlett-Lewis model on the simulation of Athens rainfall, Conference: European Geosciences Union General Assembly 2015At: Vienna, AustriaVolume: Geophysical Research Abstracts, Vol. 17, EGU2015-8983, DOI: 10.13140/RG.2.2.14371.25120*

*Efstratiadis A., Investigation of global optimum seeking methods in water resources problems, MSc thesis, 139 pages, Department of Water Resources, Hydraulic and Maritime Engineering – National Technical University of Athens, Athens, 2001*

*Efstratiadis, A. and Koutsoyiannis D., An evolutionary annealing-simplex algorithm for global optimisation of water resource systems, Proceedings of the Fifth International Conference on Hydroinformatics, Cardiff, UK, 1423-1428, International Water Association, 2002. (*[*http://itia.ntua.gr/en/docinfo/524/*](http://itia.ntua.gr/en/docinfo/524/)*)*

*Nelder, J. A. and Mead R., A simplex method for function minimization. Computer Journal, 7(4), 308-313, 1965*