WATER SECURITY AND CLIMATE CHANGE CONFERENCE (WSCC2022)
Adaptation for Sustainable and Resilient Development
01–03 December 2022, Novotel Bangkok Future Park Rangsit, Thailand

Precipitation Disaggregated Datasets Evaluation and Its Performance in the Storm Water Management Model SWMM

F.Aljanabi [1*],S, Guemar[2*], P, Krebs[2*] and J, Stamm [1*]

[1] *Institute of Hydraulic Engineering and Technical Hydromechanics, Technische Universität Dresden, Dresden 01062, Germany*
[2] *Institute of Urban and Industrial Water Management, Technische Universität Dresden, Dresden 01062, Germany*
** Corresponding author; E-mail: firas.aljanabi@tu-dresden.de*

ABSTRACT. The understanding of climate change and its impacts gained the attention of many researchers, especially the impacts related to urban drainage and stormwater systems. Flood studies are therefore becoming increasingly important. However, these studies require high temporal resolution of rainfall data mainly on an hourly or sub-hourly scale. Many meteorological stations provide longer daily rainfall data as compared to hourly data, also there are less stations providing hourly data. Furthermore, global circulation model's outputs have lower temporal resolution than required in impact assessment studies. Therefore, using disaggregation tools became necessary to deal with this issue, as they provide higher resolution of rainfall data which aggregates up to the coarser scale data. In this study, the disaggregation method based on the random Bartle-Lewis model is tested on rainfall data of Dresden (Germany). The daily historical rainfall records are disaggregated into hourly data and then compared with the historical ones. In this study, both hourly rainfall data (historical and disaggregated) are input into a rainfall-runoff model (e.g. SWMM model) to evaluate the flooding of an urban drainage subnetwork in both cases. The study shows the good performance of the model in preserving statistical characteristics of the rainfall data. However, the disaggregated hourly data doesn't coincide with the real historical hourly data. Some rainfall events were selected for the simulation of the flooding of the urban drainage subnetwork using both disaggregated and real historical hourly data. Results have showed that the flooding volumes were close to a great extent in both cases, number of flooding nodes was also approached, but in some cases underestimated. Furthermore, the time of flooding occurrence was different due to the shift in the time of rainfall events occurrence simulated by the disaggregation tool. However, using the disaggregation tool's results would still provide an overview about the urban flooding situation, and could be especially used for the disaggregation of GCMs output which will serve in climate change impacts studies, namely on urban drainage networks.

Keywords: *Disaggregation; Rainfall Data; SWMM; Weather Generator;*

Introduction

In many hydrological studies and applications, the data availability at an appropriate temporal or spatial scale is often a major problem. The need of hourly timescale data is for instance very important especially in detailed hydrological models, which is often a big challenge due to the limited availability of historical records in comparison with the daily records that are more widely available. In many countries around the world, a large number of meteorological stations providing daily rainfall data have been operational for decades, the number of rain gauges providing hourly data is on the other hand smaller. Moreover, these stations often provide shorter rainfall timeseries.

In the context of climate change, urban flood studies are becoming increasingly important, as urban flooding risk is getting bigger in many countries around the world. In such studies, high-resolution rainfall data an on hourly or even sub-hourly scale is extremely important. The need arises from the fact that the usage of low temporal resolution rainfall data would lead to the underestimation of flood peaks

(Lyu, H., et al., 2018). However, future climate projections (GCMs outputs) provide rainfall data on a daily scale which can't be directly used in some climate change impact studies. In order to cope with this problem, several disaggregation methods have been developed. The disaggregation method became a major technique used mainly in hydrological studies and provides the user with possible realizations of the required high-resolution data (e.g. hourly rainfall data), which aggregate up to the given coarser scale data (e.g. daily data) (Koutsoyiannis & Onof, 2001). Several approaches are used for the disaggregation, such as Bartlett-Lewis rectangular pulse model presented by RODRIGUEZ-ITURBE et al. (1987), in which rainfall events are assumed to occur according to Poisson process, each event consists of a group of cells of a specific duration and intensity.

The evaluation of the disaggregation tools has already been performed in many studies. However, the disaggregation tools are evaluated mainly based on their ability to preserve the overall statistical characteristics of rainfall data such as the mean and the variance ...etc. Furthermore, there is an increasing trend of using disaggregation tools in flood studies to deal with the lack or absence of high resolution of rainfall data. However, the disaggregated rainfall data obtained by disaggregation tools doesn't necessarily coincide with the actual real ones. The results are considered to be only a likely realization. Using disaggregated data in further models such as rainfall-runoff models should be therefore evaluated as well, to test whether using such disaggregated data as input would lead to significant differences as compared to real historical data.

In this study, the Bartlett-Lewis Rectangular Pulse model is going to be tested on a rainfall dataset from Dresden city area. The disaggregation model is going to be calibrated based on historical daily and hourly data available in the area. The disaggregation is going to be first performed on daily historical data and the results are going to be compared with historical hourly rainfall. In addition to that, a further comparison of the disaggregation tool is being performed in which both historical hourly data and disaggregated hourly data are implemented into a rainfall-runoff model (e.g. SWMM model), focus will be in the comparison of the flooding results in an urban sub-catchment in Dresden.

Material and Methods

Data and study area:

The study area is a relatively small urban sub-catchment located in the city of Dresden in the Saxonian state, in eastern Germany. Dresden climate is classidied as Cfb, according to Köppen classification. The city is characterized by moderately warm summers and cold winters. According to the climatological information from DWD, the precipitation is highly variable, and mainly falling between May and August with an average annual precipitation of around 662.5 mm for the period 1971-2000. A significant amount of precipitation occurs even during the driest month

Three rain gauge stations are installed in the city in different locations to monitor precipitation. The station chosen for this study is Hosterwitz station which provides both hourly and daily rainfall data for a relatively long period of time. Information about this station are presented in the table 1 below. The observed rainfall time series are availble for the period between 1946 and 2020 with a daily temporal resolution, and for the period between 2006 and 2020 with an hourly temporal resolution. The data are obtained from the database of Deutscher Wetterdienst (DWD, German Meterological Service).

Station ID	Location	Longitude [Degrees]	Latitude [Degrees]	Period of measurement of hourly rainfall data	Period of measurement of daily rainfall data	Station height (m)
01050	Hosterwitz	13.85	51.02	2006-2020	1946-2020	112

Table 1 Details of selected rainfall station

The hourly data provided by the station are particularly important and will serve as an input to determine the parameters of the disaggregation tool. Therefore, at first, a comparison of the aggregated hourly data with the daily data is necessary to check the quality of the data. The results are plotted in the following figure, which shows that there is a very good correlation between the aggregated and observed data with a correlation coefficient of R^2=0.9769.

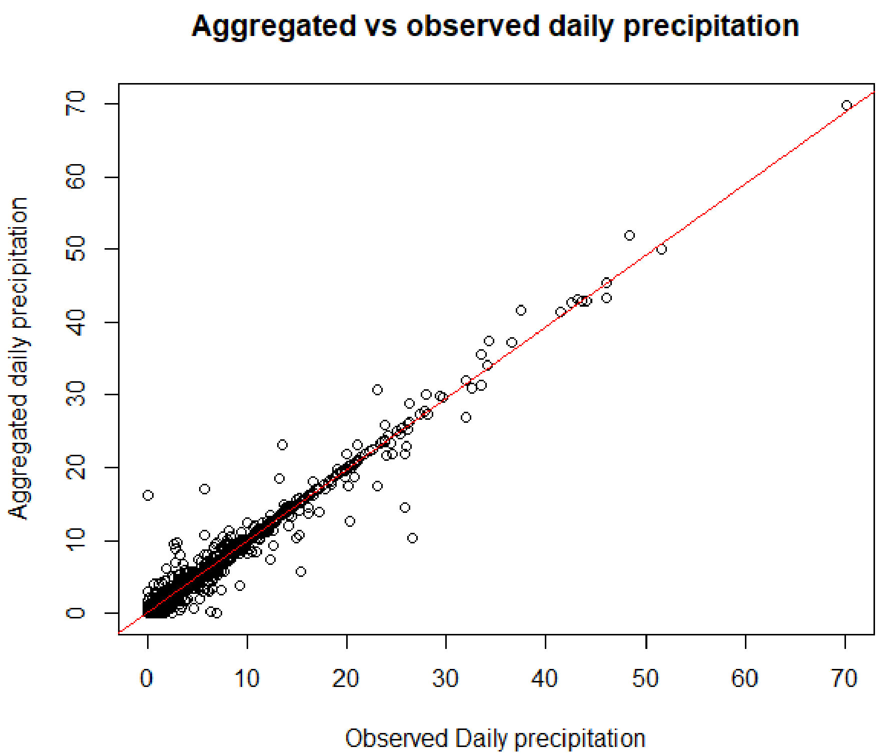

figure 1 Comparison of aggregated and observed daily precipitation provided by Hosterwitz station (R^2=0.9769)

For further use of these rainfall records, an analysis of the extreme value distribution for hourly and daily rainfall records is of a high importance, especially for further use in the disaggregation tool. The analysis is important to determine the rare events and detect the outliers which can affect the results of the study. The analysis is conducted for the period between 2006 to 2020 for the hourly data, and for the period 1980-2020 for the daily data. The results are presented in the figures below:

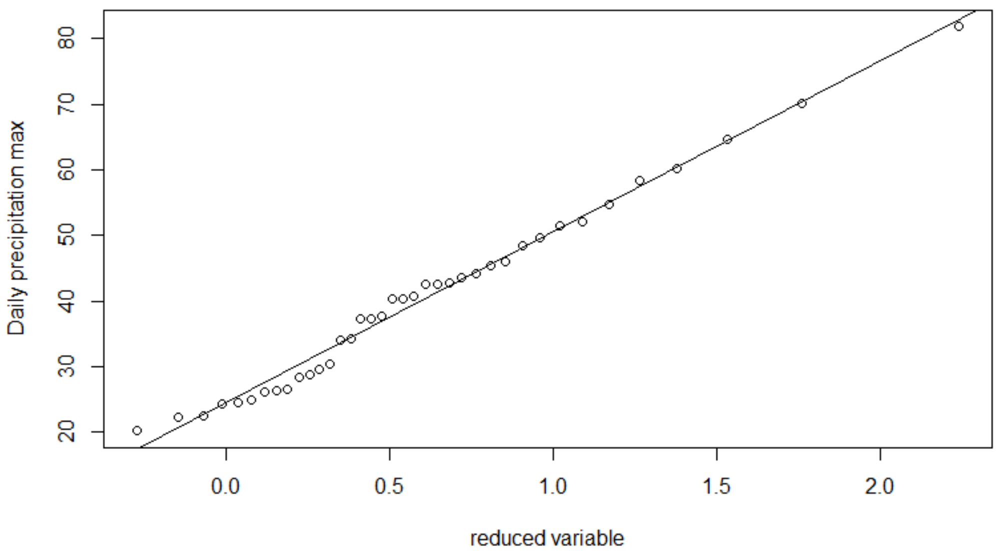

Figure 2 Gumbel plot for annual maximum daily data (1980-2020): R^2=0.9895

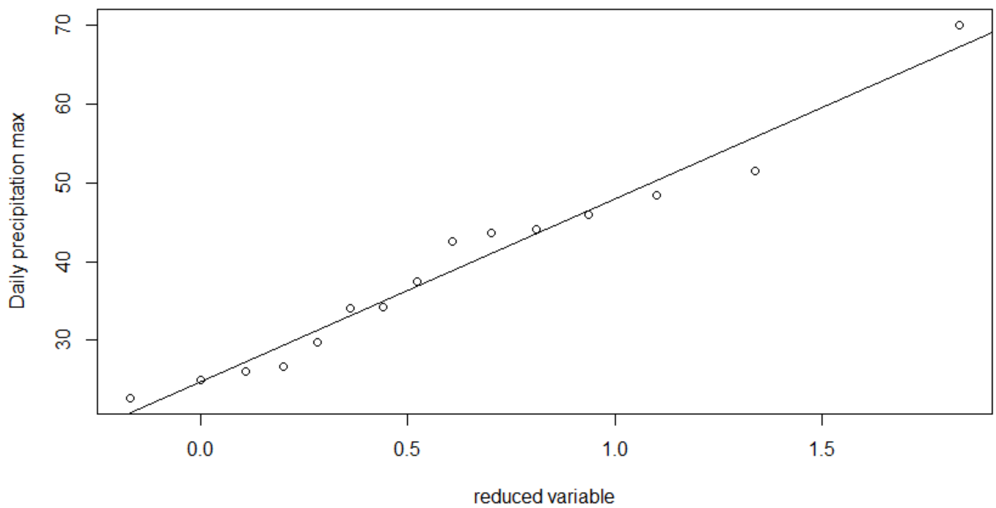

Figure 3 Gumbel plot for annual maximum daily data (2006-2020): R^2=0.9672

The plots present the observed maxima in the period 2006-2020 for the hourly precipitation maxima, and for the period 1980-2020 for the daily precipitation maxima after removing the years with no precipitation records. The maxima are arranged in an increasing sequence. The maxima are plotted in the vertical axis, while the horizontal axis presents the reduced variable which is calculated using the following equation:

$$r_{i:15} = -log(-log(\frac{i-0.5}{15}))$$

where i is the plotting position.

The plot indicates the best fitted Gumbel distribution in case the points are plotted following a straight line, any presence of points significantly far from the straight line would indicate the presence of outliers. In the graphs above, all points are aligned with the straight line which indicated no presence of outliers.

Disaggregation of rainfall data

a) *Bartlett-Lewis Rectangular Pulse model:*

Several approaches and models were proposed recently to disaggregate rainfall data such as Bartlett-Lewis rectangular pulse model (BLRPM) which is used in this study. The model chosen has been widely applied in several climates and has shown good results in its ability to reproduce some important features of the rainfall data from hourly to daily scale. (Rodriguez-Iturbe et al., 1987, 1988) & (Onof and Wheater, 1993, 1994).

The Bartlett-Lewis Rectangular Pulse model belongs to the Poisson-cluster model's category and represents rainfall in continuous time using clusters of rectangular pulses which simulate rainfall events. The clustering mechanism is based on the following assumptions (Rodriguez-Iturbe et al., 1987):

- Each storm origin follows a Poisson process with a rate λ.
- Within each storm, the process generating the associated cells is also a Poisson process. The first cell's origin is identical to the origin of the storm, and the origin of each cell follows the Poisson process with rate β.
- The cells arrivals of each storm are exponentially distributed with parameter γ, and terminate after a time span vi.
- The duration of each cell follows an exponential distribution with the rate η.
- Each cell has uniform intensity and follows specific distribution, with mean μx. The distribution is assumed to be either exponential with parameter $1/\mu x$, or a two-parameter gamma distribution with mean μx and μx^2 representing the mean square of cell intensity.
- In each storm, a number of cells C follows a geometric distribution of mean $\mu_c = 1 + \kappa / \varphi$, with κ and φ being dimensionless parameters calculated as follow: $\varphi = \gamma/\eta$ and $\kappa = \beta/\eta$ (Rodriguez-Iturbe et al. 1987).

The Bartlett-Lewis Rectangular Pulse model is therefore defined with five parameters: λ, κ, μ_x, φ and η which are assumed to be constant (Rodriguez-Iturbe et al. 1987)

The diagram depicted in the figure 4 is an explanatory illustration of the model.

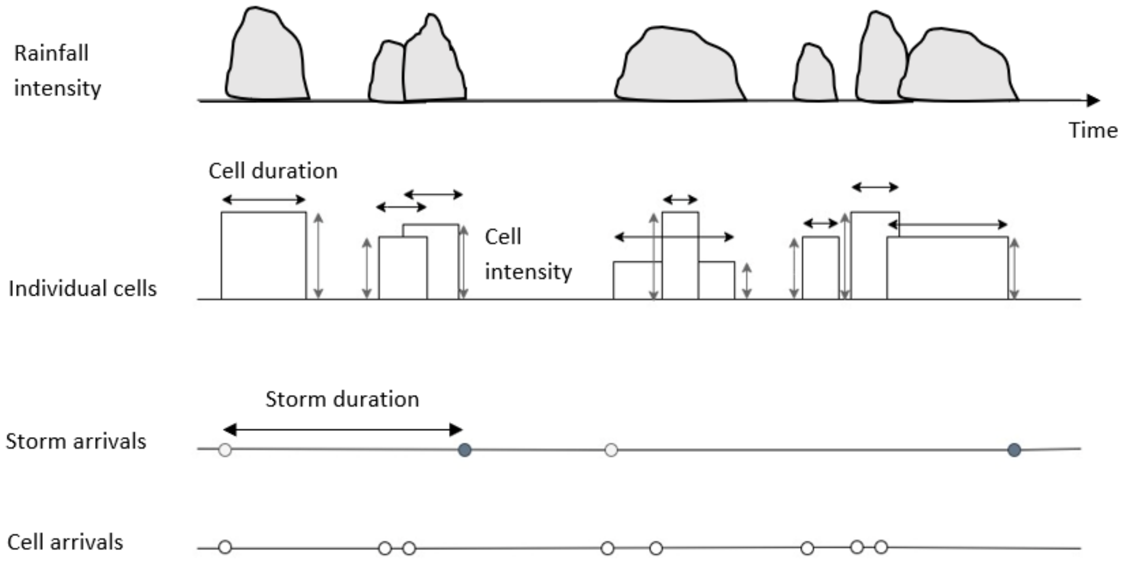

Figure 4 Explanatory illustration of the Bartlett-Lewis Rectangular Pulse model (Koutsoyiannis and Onof, 2001).

A weakness of this BLPR model is its inability to preserve the dry and wet periods properties, which might underestimate the proportion of the dry periods, or generate fewer wet periods than necessary (Rodriguez-Iturbe et al. 1987).

Therefore, an extra parameter has been introduced into the model by Rodriguez-Iturbe et al. 1988 in order to improve the fitness of the statistical parameters. The new model allows the random variation of the mean cell duration from storm to storm, and the parameter η follows two-parameter gamma distribution with shape parameter α and scale parameter v. The parameters β and γ both vary in a way that the following ratios stay constant: $\varphi = \gamma/\eta$ and $\kappa = \beta/\eta$. The depth of each cell is a random constant exponentially distributed with a mean equal to E[x]. The new model is therefore a six-parameter model with parameters λ, κ, μ_x, φ, α and v. In the BLRP model, the distribution of the cell intensity can be assumed to follow the exponential, gamma or Weibull distribution.

b) Model implementation in HyetosMinute software:

Within this work, the model is implemented in a software program named HyetosMinute, which is a package coded in the programing language R allowing a temporal stochastic simulation of fine-resolution rainfall series. The program was developed by Koutsoyiannis and Onof (2001), based on the BLRP model as a background model for the disaggregation of rainfall data. Using iteration, the model produces synthetic rainfall data statistically similar to the given daily rainfall. The input parameters necessary for the model are mainly the statistical properties calculated from the historical rainfall data, namely the mean, variance lag-1 autocovariance and the probability dry at 1, 6, 12 and 24 hours. The program requires entering those statistical properties in addition to the daily rainfall timeseries in order to generate the hourly rainfall data.

Based on the entered statistical properties, the model at first separates the daily data into multiple clusters of wet days, for each cluster the model repeats the Bartlett-Lewis model several times to obtain an initial synthetic data of which daily depths of wet days are similar enough to the original one. The two series of daily depths match each other if the departure d defined by the following formula is smaller than an acceptable limit defined by the user da.

$$d = [\sum_{i=1}^{L} ln^2(\frac{Zi + 0.1}{\widetilde{Zi} + 0.1})]^{\frac{1}{2}}$$

L is the length of the cluster.

Zi is the observed daily depth of day i

$\widetilde{Zi}$ is the synthetic daily depth of day i (Kossieris, P., et al., 2018).

After generating an initial short timescale rainfall series based on these parameters, Hyetos program corrects the time series using the proportional adjusting procedure of generated hourly data in order that the short timeseries terms add up to the high-level longer time series and are therefore consistent with the original daily series. The procedure is defined as:

$$Xs = \widetilde{Xs}\left(\frac{\sum_{i=1}^{L} Zi}{\sum_{i=1}^{L} \widetilde{Zi}}\right) s = 1, \ldots, 24$$

, where $\widetilde{Xs}$ is an initial synthetic depth of an hourly time s and Xs is a final adjusted hourly depth (Koutsoyiannis and Onof.2001).

This step is necessary in order to minimize the error between the sum of the generated time series and the high-level entered time series. The process is repeated several times until obtaining the best values, and generating results close enough to the historical values.

c) Parameter estimation:

The required statistical properties to estimate the BLRPM parameters are: the mean, variance, lag-1 auto-covariance coefficient and the proportion of dry periods. The parameters are then calculated based on the following equations (Bo, Z., Islam, S. et al, 1994):

$$Mean = \lambda\mu \times \mu c \frac{v}{\alpha-1} T \quad (1)$$

$$Variance = \frac{2v^{2-\alpha}T}{\alpha-1}(k1 - \frac{k2}{\varphi}) - \frac{2v^{3-\alpha}}{(\alpha-2)(\alpha-3)}.(k1 - \frac{k2}{\varphi^2}) + \frac{2}{(\alpha-2)(\alpha-3)}.[k1(T+v)^{3-\alpha} - \frac{k2}{\varphi}(\varphi T + v)^{3-\alpha}] \quad (2)$$

Where

$$k1 = (2\lambda\mu c E^2[x] + \frac{\lambda\mu ck\varphi E^2[x]}{\varphi^2 - 1})(\frac{v^\alpha}{\alpha - 1})$$

$$k2 = \frac{\lambda\mu ckE^2[x]}{\varphi^2 - 1}(\frac{v^\alpha}{\alpha - 1})$$

$$\mathrm{Cov}[Y_t^{(T)}, Y_{t+s}^{(T)}] = \frac{k1}{(\alpha-2)(\alpha-3)}\{[T(s-1)+v]^{3-\alpha} + [T(s+1)+v]^{3-\alpha} - 2(Ts+v)^{3-\alpha}\} + \frac{k2}{\varphi^2(\alpha-2)(\alpha-3)}\{2(\varphi Ts+v)^{3-\alpha} - [\varphi T(s-1)+v]^{3-\alpha} - [\varphi T(s+1)+v]^{3-\alpha}\} \quad s \geq 1 \tag{3}$$

$$\mathrm{prob}[\text{zero rainfall}] = \exp\{-\lambda T - [\frac{\lambda v}{\varphi(\alpha-1)}(1+\varphi(\kappa+\varphi) - \frac{1}{4}(\kappa+\varphi).(\kappa+4\varphi) + \frac{\varphi(\kappa+\varphi)(4\kappa^2+27\kappa\varphi+72\varphi^2)}{72}] + \frac{\lambda v}{(\alpha-1)(k+\varphi)}(1-\kappa-\varphi+\frac{3}{2}k\varphi+\varphi^2+\frac{\kappa^2}{2}) + \frac{\lambda v}{(\alpha-1)(\kappa+\varphi)}(\frac{v}{v+(\kappa+\varphi)T})^{\alpha-1}.\frac{\kappa}{\varphi}(1-\kappa-\varphi+\frac{3}{2}\kappa\varphi+\varphi^2+\frac{\kappa^2}{2})\} \tag{4}$$

Where : $Y_t^{(T)}$ represents the time series of rainfall at an accumulation interval T, and s represents the lag time in number of accumulation intervals.

All these equations consist the basis of the fitting procedure of the model by relating the statistical properties of the rainfall time series to the model parameters.

Based on the statistical properties of 1, 6, 12 and 24 hours, the estimation of the Bartlett-Lewis model parameters is possible using Hyetos program, via the enhanced version of the evolutionary annealing-simplex (eas) optimization method (Kossieris et al., 2015). This function is developed originally by Efstratiadis and Koutsoyiannis (2002) (Tsoukalas et al., 2016), and provides all the model parameters required for further disaggregation of the rainfall data, based on an objective function which is computed to select the best set of statistics for each moment combination.

In this study, the statistical parameters were calculated for each month separately. Similarly, the BLRPM parameters were also estimated by minimizing the objective function and using the eas function for each month. Hence, twelve parameter sets were determined for each month, each set contains the six BLRPM parameters λ, κ, μx, φ, α and v.

In this study, the random Bartlett-Lewis model is being used. The model is being first examined to check its ability to reproduce the statistical properties of the observed hourly data as well as its performance in reproducing the extremes.

EPA-SWMM Model Description:

The EPA Storm Water Management Model (SWMM) is a public domain software tool for dynamic rainfall-runoff modelling used for the simulation of single event or long-term (continuous) simulation of the quantity and quality of runoff in urban areas (Rossman, L.A., Huber, W.C., 2016).

The model is designed to produce an integrated approach to urban drainage, and has several applications mainly in the design, planning, and analysis of stormwater systems as well as sanitary and combined sewers and other drainage systems. Several parameters can be tracked through the SWMM model such as runoff quantity and quality in each sub catchment that receives precipitation, flow rate and flow depth in addition to the water quality in each pipe (Rossman, L.A., Huber, W.C., 2016). Several versions of the SWMM model have been developed, the newest version of the model SWMM 5 features an integrated Windows environment and offers several options for analysis such as the use of graphs, tables and maps to display the simulation results. This study uses the release 5.1 of the model to perform the analysis.

Case study:

The stormwater system selected for this study is a subnetwork from the urban drainage network of Dresden in Germany, located in between Geberbach, Lockwitzbach and the Elbe rivers. The system has 679 junction nodes, and 747 conduit links. The model provided has been already calibrated and validated. The calibration of the model was conducted using the multi-objective optimization approach. Further details about the calibration and validation of the model are found in the literature (Reyes-Silva, J.D, et al., 2020).

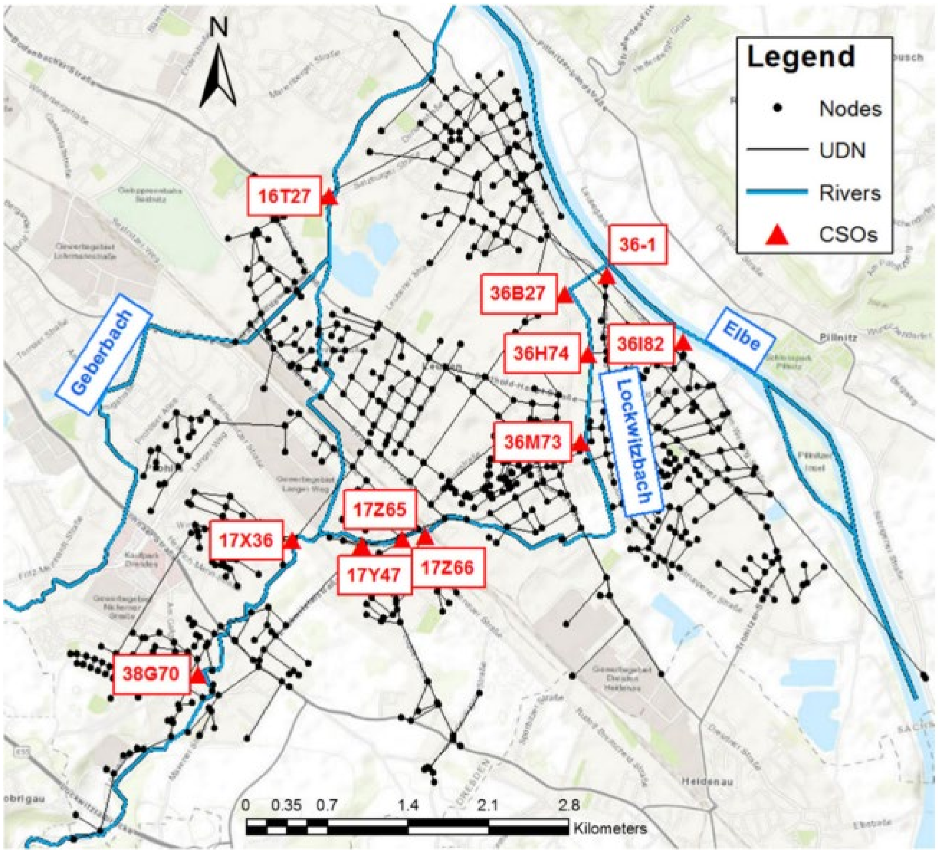

Figure 5 Urban drainage network in the study area (Reyes-Silva, J.D, et al., 2020).

This model is selected as a case study to evaluate the performance of the disaggregation tool and its performance in SWMM model through the evaluation of the flooding in the network. The evaluation of will focus on two main criteria: Node flooding and surcharged conduits. The former occurs when water surface elevation exceeds the rim elevation of the manhole, which results into flooding in the downstream. The later occurs when both the entrance and the exit of the conduit are full, therefore the pipe is flowing full. Therefore, both criteria can be used as index for the assessment of the stormwater system flooding. SWMM model provides information about the flooding nodes in a table presenting information about the total flooding quantity and the duration of flooding in each node. The model also provides information about the conduits surcharged through the “Hours capacity limited” found in the summary results under the “conduit surcharge” section. In this study, the parameters chosen for the comparisons are mainly the number of flooded nodes, the volume of flooding in each node in addition to the number of surcharged conduits. Some thresholds are chosen to facilitate the comparison.

Results

Disaggregation of rainfall data:

The disaggregation method using Bartlett-Lewis Random model was tested on a rainfall time series of daily resolution for a period of 15 years (2006 to 2020) collected from Hosterwitz station in Dresden (Germany). The station provides also a series of hourly rainfall data from the same time period, which is going to be further compared with the disaggregated data in order to evaluate the performance of the disaggregation model. Using the Bartlett-Lewis random model implemented in HyetosMinute package software, the daily data is disaggregated to generate hourly synthetic data that sums up exactly to the historical daily values. The rainfall patterns in Dresden (Germany) vary significantly among seasons, therefore the analysis conducted in the study is performed on monthly basis.

The following table presents the statistical characteristics of rainfall data calculated for different resolutions, namely 1h, 6hours 12 hours and 24 hours.

	Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
	Mean	0.060	0.047	0.049	0.048	0.082	0.103	0.113	0.121	0.075	0.073	0.054	0.059
	Var	0.062	0.044	0.065	0.113	0.272	0.696	0.541	0.656	0.195	0.120	0.067	0.056
	Auto	0.035	0.027	0.028	0.049	0.097	0.116	0.209	0.224	0.089	0.074	0.042	0.036
1h	Pd	0.860	0.894	0.902	0.932	0.914	0.921	0.916	0.912	0.922	0.892	0.901	0.867
	Mean	0.360	0.283	0.291	0.288	0.492	0.615	0.681	0.724	0.450	0.438	0.324	0.357
	Var	1.091	0.844	0.946	1.531	3.221	6.149	6.781	8.069	3.087	2.311	1.262	1.042
	Auto	0.354	0.337	0.237	0.484	0.620	0.900	2.156	1.695	1.209	0.997	0.636	0.304
6h	Pd	0.685	0.767	0.764	0.839	0.794	0.799	0.792	0.792	0.821	0.766	0.780	0.713
	Mean	0.721	0.565	0.583	0.577	0.984	1.230	1.362	1.449	0.899	0.873	0.648	0.720
	Var	2.821	2.378	2.348	4.011	7.929	14.298	18.016	19.647	8.872	6.959	3.747	2.822
	Auto	0.894	0.786	0.425	1.023	0.983	2.028	4.788	3.150	3.212	2.332	1.628	0.582
12h	Pd	0.570	0.668	0.659	0.763	0.689	0.709	0.690	0.690	0.742	0.668	0.695	0.587
	Mean	1.442	1.130	1.165	1.155	1.968	2.460	2.728	2.897	1.799	1.752	1.296	1.442
	Var	7.627	6.423	5.275	9.952	16.818	31.987	46.102	43.487	25.426	17.371	11.821	6.622
24h	Auto	1.531	2.095	1.247	1.520	3.185	4.927	9.368	6.242	7.771	5.011	2.722	0.900

	Pd	0.419	0.523	0.519	0.658	0.546	0.564	0.527	0.534	0.624	0.537	0.567	0.436

Table 2 Historical characteristics of precipitation of Hosterwitz station obtained from the series (2006-2020)

var: Variance, Auto: Autocorrelation lag-1, Pd: Probability dry

Random model Bartlett Lewis Model parameters

Based on the statistical properties calculated for each month, the BLRM parameters are hence estimated assuming local stationarity within the month. The following table shows the produced statistics of the random Bartlett-Lewis model for each month.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$\lambda(d^{-1})$	0.658	0.347	0.546	0.291	0.552	0.517	0.333	0.505	0.747	0.306	1.007	0.676
ϕ	0.159	0.051	0.108	0.058	0.073	0.041	0.050	0.093	0.967	0.124	0.335	0.132
K	3.184	2.503	1.363	2.230	2.190	1.343	2.229	1.735	1.742	1.126	0.149	2.851
α	6.978	10.150	8.241	12.451	13.199	14.117	13.838	14.273	1.946	9.412	2.198	12.493
ν(d)	0.290	0.156	0.180	0.196	0.296	0.221	0.314	0.451	0.038	0.833	0.117	0.315
μ_x (mm d-1)	2.192	3.944	6.110	5.937	4.227	8.338	7.531	7.925	19.825	5.472	7.526	3.388
μ_x/σ_x	2.890	2.070	1.906	3.440	4.696	4.614	4.011	4.010	1.108	2.255	1.774	2.457

Table 3 Bartlett Lewis model's parameters for each month

After determination of the model's statistics, the model then disaggregates the daily rainfall data of each month using the corresponding set of parameters to generate synthetic timeseries of hourly data. It is to mention that the model calibration using the random BLRM is difficult and time consuming, for each month the disaggregation was performed several times until obtaining satisfactory results.

The model performance:

After disaggregation of the data using Bartlett-Lewis random model implemented in HyetosMinute, it is possible to evaluate the performance of the model by performing a comparison of the different statistical properties (mean, variance, autocorrelation and proportion dry) of the observed data against those of the disaggregated data generated by the model. The comparison is conducted on different aggregation levels namely 1 hour, 6 hours and12 hours. These statistical properties calculated for each month are plotted for both observed and synthetic data and are represented in the figures below.

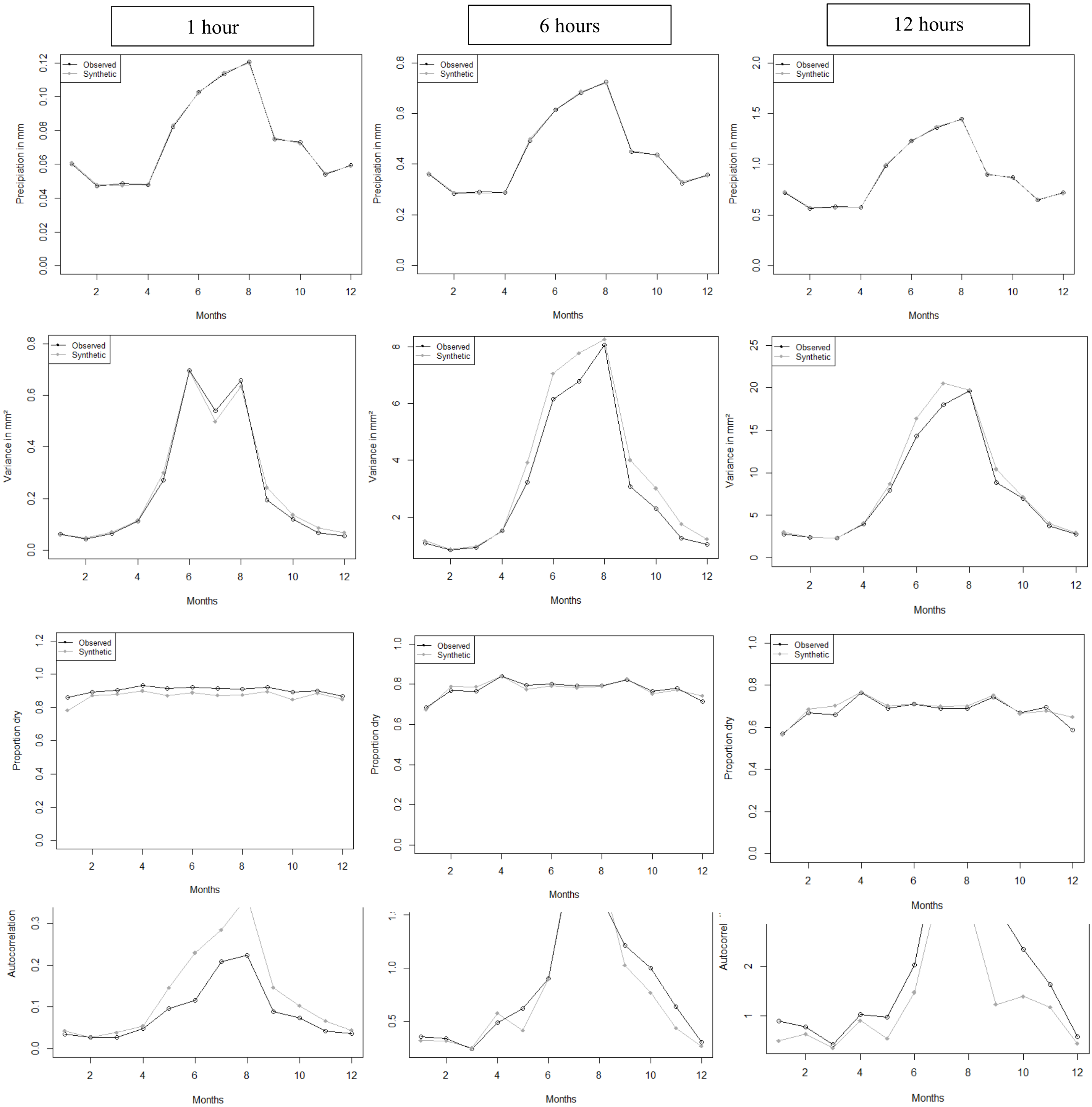

Figure 6 The statistical characteristics of observed against synthetic disaggregated data calculated for each month and for different aggregation level (1h, 6h and 12 hours): Mean value in mm, variance in mm^2, proportion dry and autocovariance.

The figures above show to what extent is the model able to preserve the different statistical properties. The model has proved its very good performance in preserving the mean among different aggregation levels. While the mean value has been preserved, other parameters show a slight bias in some months and at different aggregation levels. For the 1-hour level, results presented in the figure above show that the model performs in a satisfactory level for the variance estimation, the probability dry is a slightly underestimated for all months, the variance is on the other hand estimated to a satisfactory level. While the autocorrelation lag-1 is not preserved with the difference getting to the highest level towards the summer season. For the 6 hours aggregation level, results show that the variance is still overestimated mainly in the summer months while the proportion dry has improved on the other hand, and the autocorrelation has improved. For the 12 hours aggregation level, results show slightly similar results than those of the 6 hours. The variance is overestimated in the months of June and July, the proportion dry is well simulated, and the autocorrelation present some bias, with highest difference in September.

Therefore, the results found concluded that the model showed good adjustment of the statistical properties between historical and simulated data, mainly for the mean, the variance and proportion dry. As for the autocorrelation of lag-1, results present less adequacy for most aggregation levels. However, further evaluation of the model performance is required, especially to evaluate the fitting of the extreme values.

Extreme value analysis:

Another important criterion in the evaluation of the disaggregation model's performance is its ability to reproduce the distribution of maximum rainfall depths. Therefore, an assessment of the model's performance has been made by comparing the observed and simulated hourly annual maxima. For each month, the annual maximum rainfall depth corresponding to that month along with the one from the disaggregated data were visually compared and represented in Gumbel plots. The Gumbel plot represents the annual maximum rainfall in the vertical axis and the reduced variate in the horizontal axis. The reduced variate is given by: $r = -\log(-\log(\frac{i-0.5}{n}))$

Where i is the plotting position, and n is the number of years of the observation.

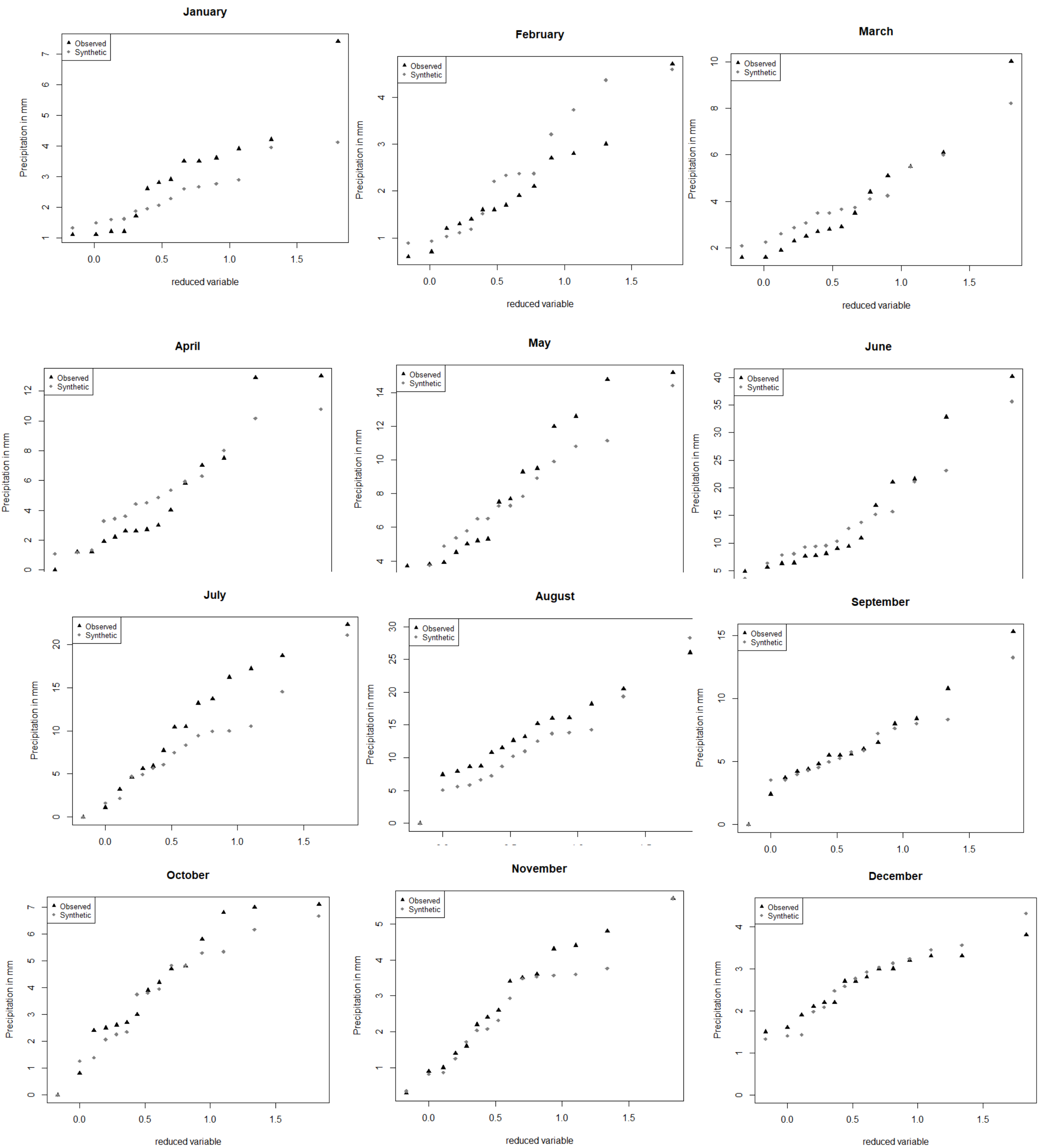
January
February
March
April
May
June
July
August
September
October
November
December
Observed
Synthetic
Precipitation in mm
reduced variable

Figure 7 Gumbel plots of the rainfall extreme values comparing the synthetic disaggregated data and the observed data from the period (2006-2020)

As it is shown, the disaggregation model performance varies from a month to another. The model shows very good performance in the reproduction of the extremes and following the same pattern as the one from the observed data for the months of September, December and November. With the exception that some values have been underestimated in the month of November. The difference is however relatively small and can be accepted. As for August, the model has followed the same pattern of the observed extremes recorded in that month, but almost all values have been a little bit underestimated. In July also, the model has underestimated some extremes. On the contrary, for the month of February, the highest values of rainfall extremes have been overestimated. As for the other months, the model performed to an acceptable level.

Finally, it can be concluded that the synthetic data generated by HyetosMinute using the random Bartlett-Lewis model has overall good performance in the reproduction of the extreme values all with preserving the statistical properties of the mean, variance and proportion dry for different aggregation levels. There is always a room for improvement of the results by trying other distributions of the cell intensities (Weibull or Exponential), which might improve the results of the reproduction of the extremes. However, the calibration is sometimes challenging and time consuming. For this study, the results found are very satisfactory, and the model is considered to have an overall good performance.

Time series comparison

The disaggregated data is going to be further used as an input in the SWMM model. Therefore, a last step in the evaluation of the model is to compare the rainfall time series for short periods of time between the synthetic and observed data, to check whether the rainfall events occur at the same time and with the same intensity. For this reason, some rain events are being selected from the observed data and have been compared to the rain events generated by the disaggregated model for the same period of time. The period of time chosen is relatively big (a month) to better understand how the model approach the real observed patterns of the rainfall events.

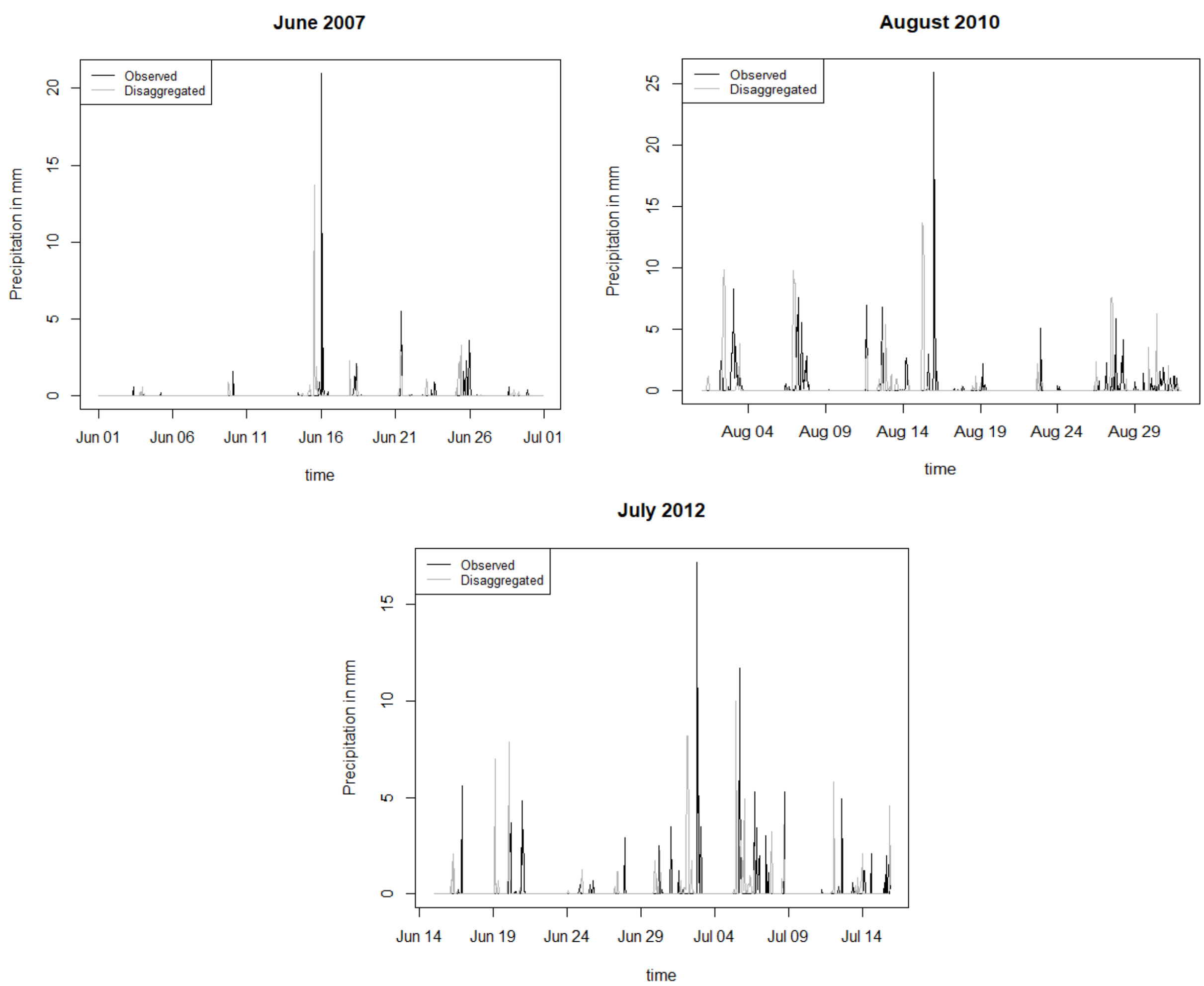

Figure 8 Rain events in June 2007, August 2010 and July 2012, comparison between disaggregated and observed data.

Although the overall statistical properties of the observed rainfall data have been generally preserved for different aggregation levels, the rainfall time series of both disaggregated hourly data and hourly observed data present some differences. From the graphs above, it can be noticed that the rainfall events occur on different times, sometimes occurring before the actual observed rainfall event. Even in the cases where the rainfall events coincide at the same time, there is however difference in the peak value of the rain event. In most cases, the model has underestimated the peak values especially for rain events with high peak value. In some few cases, the peaks were overestimated, as it is the case in the rainfall event of July 2012, furthermore the rain event occurring on June 19th which was a little bit overestimated by the model. In addition to that, the duration of rainfall events can differ between observed and disaggregated data. To evaluate further the disaggregation model in relationship with the rainfall-runoff modelling, a further comparison is conducted following section concerning the stormwater modelling using SWMM.

Evaluation of the model in SWMM

After comparing the different statistical parameters as well as rainfall time series between observed hourly data and the disaggregated hourly data, a last step in the evaluation of the performance of the disaggregation model is to assess the flooding results using both rainfall data series.

For this purpose, several rain events with different return periods are selected. The events are selected from the observed rainfall time series, and for the same period of time, the corresponding disaggregated rainfall are selected for the comparison. Both are implemented in SWMM model, and results of flooding situation are presented in the following:

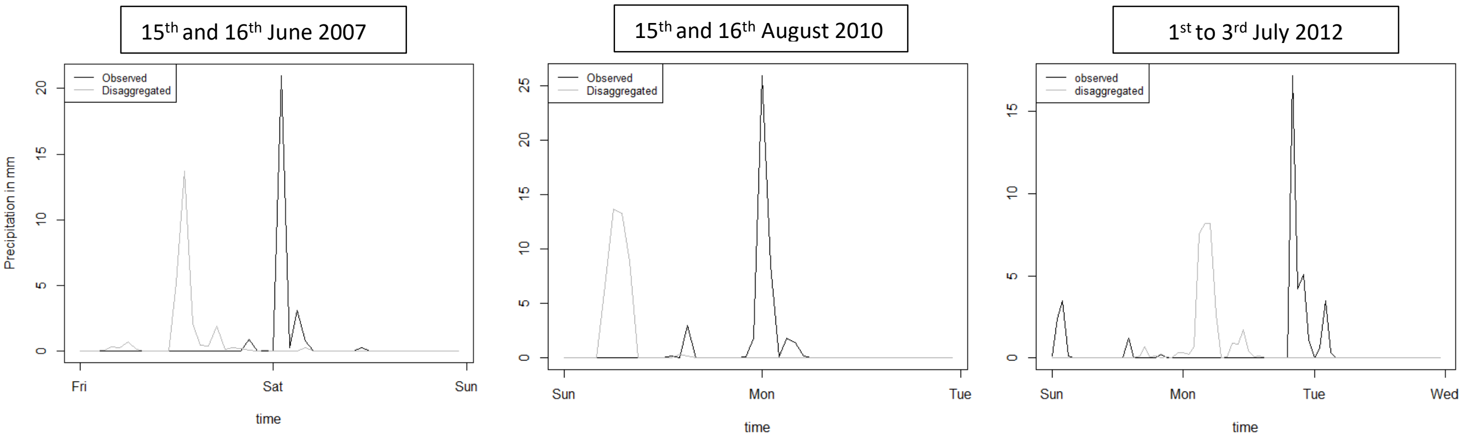

Figure 9 Selected rainfall events from observed and disaggregated data for the simulation in SWMM.

Rainfall event of June 15th/16th: The historical rainfall event has occurred in the first hours of the day June 16th but has occurred the day before according to the disaggregated rainfall data. Here, the rainfall event peak was also underestimated, and the event has longer duration of as compared to the historical one.

Rainfall event of August 15th/16th: Here also, the rainfall event occurred in the last hours of the day August 15th and the first hours of the next day. However, the model counted the whole rainfall event on the day 15th which had smaller peak value of around 14 mm but has occurred for longer period of time as compared to the historical event.

Rainfall event of July 2nd: Here both rainfall events have occurred on the same day (historical and disaggregated), however, the duration and the peak volume of the rainfall events were different.

The rainfall events are input into SWMM model to compare and analyze the flooding situation in both cases. The comparison of flooding is assessed by a comparison of flooding nodes and surcharged conduits results in both simulations.

The results are presented in the following table:

	Observed	Disaggregated	Number of conduits surcharged	Observed	Disaggregated
For more than one hour	63	73	More than 1h	73	76

For more than 10^6 ltr	8	6	More than 5h	8	10

Table 4 Flooding nodes and surcharged conduits results during rain event of June 15th /16th

	Observed	Disaggregated	Number of conduits surcharged	Observed	Disaggregated
For more than one hour	110	114	More than 1h	129	142
For more than 10^6 ltr	23	21	More than 5h	24	15

Table 5 Flooding nodes and surcharged conduits results during rain event of August 15th /16th

	Observed	Disaggregated	Number of conduits surcharged	Observed	Disaggregated
For more than one hour	91	84	More than 1h	106	95
For more than 10^6 ltr	17	12	More than 5h	43	36

Table 6 Flooding nodes and surcharged conduits results during rain event of July 2nd

To better visualize the results of flooding during these rain events, the results of total amount of flooding is plotted in the following figure:

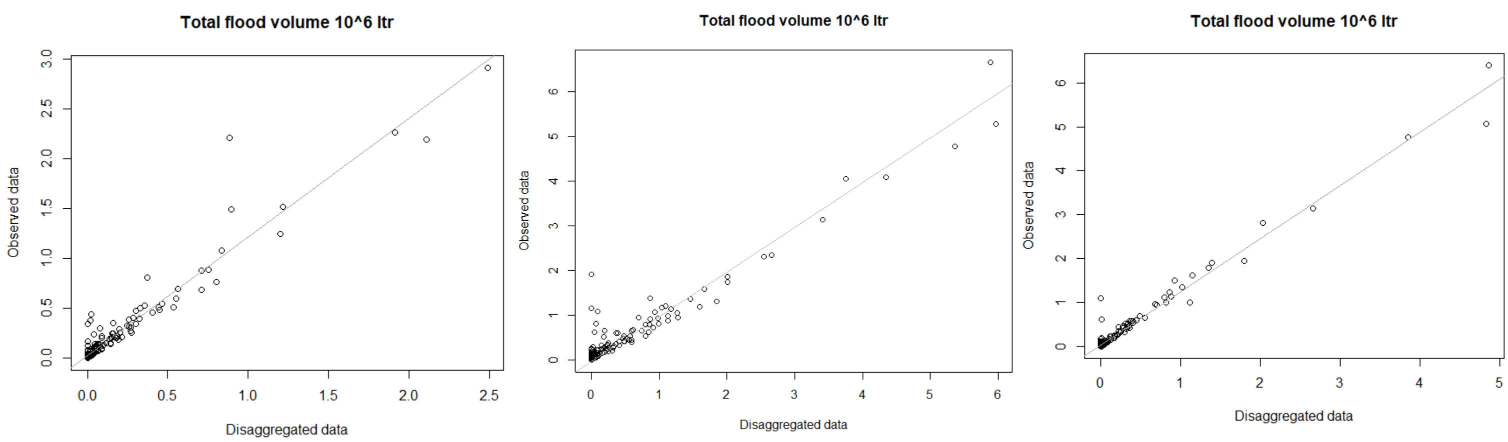

Figure 10 Comparison of total flood volume (10^6 ltr) between disaggregated and observed data. The grey line is the trendline with R^2=0.9267 (June 2007: left figure), R^2=0.9373 (August 2010: middle figure) and R^2=0.9735 (July 2012: right figure).

From the figures and the tables above, it can be noticed that although the rainfall events differ in time and in volume, but the flooding situation resulting from the rainfall events are very similar. The total number of flooding nodes can differ, however in so many cases the volume of the flooding was very small. The number of flooded nodes for more than the specified threshold doesn't differ much, the results were very close in the case of the rainfall event of August 15th/16th, results were on the other hand slightly overestimated in the case of the rainfall event of July 2nd. Furthermore, there were 10 nodes or more which were flooded for more than an hour in the case of disaggregated data, most of them were however flooded with a small volume. Finally, it is to mention that the day and time of the maximum flooding is different, due to the difference in the occurrence of the rainfall events.

The comparison of the surcharged conduits is assessed through comparing the hours of capacity limited in both cases. The tables above show that a quite similar number of conduits were surcharged in both cases in 15th/16th June and an underestimation of the results was generally observed in the results of the July 2nd event.

From the figure 10 which presents the comparison of the total flooding volume in both simulations, the results were very satisfactory, as the correlation value calculated by R^2 ranges between 0.92 to 0.97. It is to mention that flooded nodes in only one simulation (either using historical data, or disaggregated data) are not taken into consideration in this comparison.

Conclusion

The disaggregation of the daily precipitation data was performed based on Bartlett-Lewis model using HyetosMinute software package. First, the performance of the model in reproducing the rainfall characteristics was tested. The rainfall characteristics of observed data on different aggregation levels were input into the model along with the daily rainfall data which was therefore disaggregated into hourly data. To evaluate the model performance, the disaggregated rainfall data provided by the model was compared to the hourly data. The model has generally showed good performance in preservation of the overall statistical characteristics of the rainfall. However, as the results will be implemented in SWMM model for further comparison, the comparison of time series of some rainfall events was first performed. The results showed in many cases an underestimation of the peak values of the rainfall events, the durations of rainfall events were also in general different, and the peak value has occurred in many cases before the actual peak value.

The model generates a possible realization of the hourly rainfall data, it is therefore very likely that the generated synthetic hourly data doesn't coincide with the real hourly data. However, the two series are input into SWMM model to evaluate whether the flooding situation of the urban drainage subnetwork are comparable in both cases. The comparison was based on two criteria, namely the flooding nodes and surcharged conduits**.** The results showed that the flooding volumes occurring during the rainfall events using disaggregated hourly data were close to a great extent to those using actual historical hourly data. However, the time of the flooding occurrence were different, due to differences in the time of rainfall events occurrence. The number of flooded nodes or surcharged conduits were also close enough, with sometimes underestimation or overestimation of the results. The evaluation of the results was restricted on the comparison of flooding volumes, number of flooded nodes and the limited capacity of conduits, therefore, other parameters or criteria should also be taken into consideration in future work to get better understanding of the model performance. Finally, it can be concluded that the usage of disaggregated rainfall data can give an overview of the flooding situation occurring in the urban drainage subnetwork when hourly data are absent. The model can mainly be used when dealing with climate change impact assessment studies, to disaggregate the future projections of rainfall (from GCMs output), to get the higher-resolution of rainfall data required in such studies.

References

DWD. Retrieved 27.12.2021, from http://worldweather.wmo.int/de/city.html?cityId=1343.

Bo, Z., Islam, S., & Eltahir, E. A. B. (1994). Aggregation-disaggregation properties of a stochastic rainfall model. Water Resources Research, 30(12), 3423-3435.

Kossiers, P., Efstratiadis, A., Tsoukalas, l., Koutsoyiannis, D. (2015). Assessing the performance of Bartlett-Lewis model on the simulation of Athens rainfall, in: European Geosciences Union General Assembly 2015, Geophysical Research Abstracts, Vol. 17, Vienna, EGU2015-8983. European Geosciences Union, 2015.

Koutsoyiannis D, Onof C. (2001). Rainfall disaggregation using adjusting procedures on a Poisson cluster model. J Hydrol 246(1-4):109-122.

Lyu, H., Ni, G., Cao, X., Ma, Y., & Tian, F. (2018). effect of temporal resolution of rainfall on simulation of urban flood processes. Water, 10(7), 880.

Onof, C. and H. S. Wheater. (1993). Modelling of British rainfall using a Random Parameter Bartlett-Lewis Rectangular Pulse Model. J. Hydrol., 149, 67-95.

Onof, C. and H. S. Wheater. (1994). Improvements to the modeling of British rainfall using a Modified Random Parameter Bartlett-Lewis Rectangular Pulses Model, . J. Hydrol., 157, 177-195.

Reyes-Silva, J. D., Bangura, E., Helm, B., Benisch, J., & Krebs, P. . (2020). The Role of Sewer Network Structure on the Occurrence and Magnitude of Combined Sewer Overflows (CSOs). Water, 12(10), 2675.

Rodriguez-Iturbe, D. R. Cox, and V. Isham. (1987). Some models for rainfall based on stochastic point processes. Proc. R. Soc. Lond., A 410, 269-298.

Rodriguez-Iturbe, D. R. Cox, and V. Isham. (1988). A point process model for rainfall: Further developments. Proc. R. Soc. Lond., A 417, 283-298.

Rossman, L.A., Huber, W.C. (2016). Storm Water Management Model Reference Manual Volume I—Hydrology (Revised); U.S. Environmental Protection Agency: Cincinnati, OH, USA, 2016.

Tsoukalas, I., Kossieris, P., Efstratiadis, A., Makropoulos, C. (2016). Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. Environ.Model.Softw. 77,122-142.doi:10.1016/j.envsoft.2015.12.008