

# Hydrodynamic and probabilistic modelling of storm overflow discharges

Szeląg Bartosz, Adam Kiczko, Jan Studziński and Lidia Dąbek

## ABSTRACT

The study compares an annual number of weir overflows calculated using a hydrodynamic model by continuous simulations and a probabilistic model. The weir outflow for a single precipitation event was successfully modelled using logistic regression. Performed numerical experiments showed that the calculated number of weir outflows with the hydrodynamic model falls within confidence intervals of the probabilistic model. This suggests that the model of the logistic regression can be used in practice. The probabilistic simulations revealed that a model with a probabilistic description of a number of annual precipitations and a model with an assumed average number of such events are not consistent. The proposed methodology can be applied for the design of outflow weirs and other storm devices.

**Key words** | hydrodynamic model, logistic regression, overflow discharges, probabilistic model, stormwater, SWMM

**Szeląg Bartosz** (corresponding author)

**Lidia Dąbek**

Faculty of Environmental, Geomatic and Energy Engineering,  
Kielce University of Technology,  
Tysiaclecia Państwa Polskiego Av. 7, 25-314 Kielce,  
Poland  
E-mail: [b.szelag@tu.kielce.pl](mailto:b.szelag@tu.kielce.pl)

**Adam Kiczko**

Faculty of Civil and Environmental Engineering,  
Warsaw University of Life Sciences,  
Nowoursynowska St. 159, 02-787 Warszawa,  
Poland

**Jan Studziński**

Centre for Computer Science Applications in  
Environmental Engineering,  
Systems Research Institute Polish Academy of  
Sciences,  
Newelska 6 St. 7, 01-447 Warszawa,  
Poland

## INTRODUCTION

Growing urbanization, climate change and extending terms of running of stormwater systems cause an essential increase in the amount of stormwater pouring on the surface area and the number of discharges by storm overflows (Doglioni *et al.* 2009). That leads to an increase in the level of the water table in receivers, to deterioration of the quality of the receivers' waters and to failure of their biological and chemical balance. Therefore, when designing storm overflows, the criteria for protection of waters against pollution should be taken into account; those criteria are expressed among other means by the permitted number of annual stormwater discharges (Rauch *et al.* 1998; Mantegazza *et al.* 2010).

Multiannual results of measurements which are usually used to develop regression models may constitute the basis for assessment of the number of stormwater discharges. However, since these models are of a local nature, their scope of application is limited and they can be used only for estimation operations (Veldkamp & Wiggers 1997;

Fidala-Szope *et al.* 1999). A number of studies have therefore been undertaken to develop methods of a universal nature, but due to the simplifications they have adopted, the result received could be significantly flawed (Benoist & Lambertus 1989; Urcikán & Rusnák 2006). At present, the basis for the multiplicity of the operation of a storm overflow is usually continuous hydrodynamic simulations (Price 2000; Thorndahl 2009; Andrés-Doménech *et al.* 2010) made by means of a calibrated model on the basis of multiannual rainfall time series (DWA-A 118E 2006).

Due to strong interactions between parameters describing the characteristics of catchment areas considered in the models, there are problems with determining the values of empirical parameters influencing a satisfactory mapping of processes that take place during the stormwater outflow. The high cost of measuring rainfall and stormwater flows as well as problems with model calibration make the approaches described above not always economically

justified. An additional problem is the access to multiannual continuous rainfall measurements of high resolution (DWA-A 118E 2006).

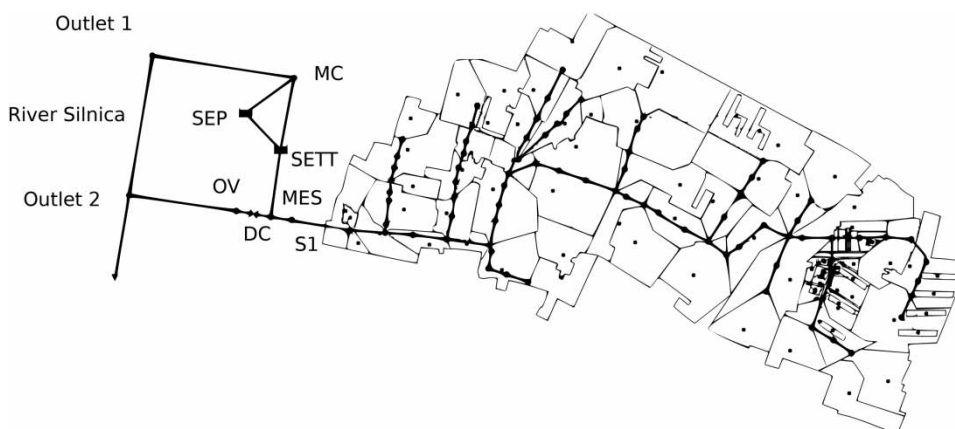
Therefore, probabilistic models were developed for the analysis of operation of storm overflows; in the models the necessity for continuous simulations was eliminated and statistical methods, like for example FORM (first order reliability method), were used for forecasting the occurrence of a stormwater discharge. The estimation of parameters in this method requires the implementation of complex numerical algorithms, so its application to model stormwater systems operation is limited (Thorndahl & Willems 2008). At the same time, in the methods, the stochastic nature of the number of rainfall events in the annual cycle has not been taken into account which may have a significant impact on the simulation results.

In view of the above considerations, it is appropriate to continue the search for models for forecasting the annual number of discharges by storm overflows. Those models should take into account the complex nature of surface runoff and the stochastic nature of rainfall and they should reduce the costs of rainfall and stormwater flow measuring; the estimation of their parameters should be relatively simple. Moreover, these models should not be highly complex for only then would they be successfully applied in engineering practice. In the paper, an attempt was made to develop a model in which a logistic regression model was used to assess the occurrence of a discharge in a storm overflow. This way hydrodynamic modelling was

unnecessary. The Iman-Conover (IC) method was used to simulate the rainfall event, assuming that the relationship between the analysed variables describing the operation of the overflow could be determined by the Spearman correlation factor. At the same time, two solutions were considered in the paper: the first of them adopted an average annual number of rainfall events which is common practice in probabilistic models (Osorio *et al.* 2009; Fu & Butler 2014; Szeląg & Bąk 2017). In the latter case, it was assumed that the number of rainfall events per year is random and described by a uniform distribution which in this type of deliberation constitutes an authorial approach.

## OBJECT OF INVESTIGATION

The subject of the analysis is an urban catchment area located in the south-eastern part of Kielce in Poland (Figure 1). The total surface area of the catchment concerned is 62 ha. The analysed catchment area is covered by housing estates, public utility buildings and main and side streets. Sealed areas extend over an area of 29.26 ha (with a retention height of 2.50 mm) and unsealed ones over an area of 32.74 ha (with a retention height of 5 mm) and the weighted average field retention value of the catchment is equal to  $d_r = 3.81$  mm (Szeląg *et al.* 2016). The separate sewer system designed ensures complete drainage of the existing infrastructure. A detailed description of the urban catchment area can be found in the paper of Szeląg *et al.* (2013).



**Figure 1** | Diagram of analysed urban catchment in SWMM program (Storm Water Management Model).

The sewer system is built of 200 sewage wells, 100 pipe sections with a total length of 22 km, one storm overflow and two outlets. The roughness coefficient of Manning regarding the stormwater drainage pipes changes between 0.013–0.018  $\text{m}^{-1/3}\cdot\text{s}$  (Szeląg *et al.* 2016). Rain waters from the catchment area flow out by the collectors  $\Phi 1.25$  m to the receiver – Silnica River, and they are previously treated in the stormwater treatment plant (Figure 1). If the filling of the diversion chamber (DC) does not exceed 0.42 m, then the entire stormwater via four pipes  $\Phi 0.400$  m is discharged to the treatment plant. In the first place it is sent to the horizontal settling tank (SETT) and then a part of it is transported through a  $\Phi 0.20$  m pipe to the coalescing separator (SEP) and to the measuring chamber (MC), from which the treated stormwater is introduced into the receiver. When the filling of the DC exceeds 0.42 m, the stormwater is discharged via the storm overflow (OV) to the  $\Phi 1.25$  m discharge channel (S2), from which it is directly discharged to Silnica River.

In the framework of continuous measurements carried out in 2009–2011, a filling probe was installed in the DC and an ultrasonic flow meter (MES) was placed on the  $\Phi 1.25$  m pipe to the DC at a distance of 3.0 m from the collector outlet (S1).

The investigated urban catchment area has already been the subject of numerous studies (Szeląg *et al.* 2013; Szeląg & Bąk 2017), which dealt with forecasting the quantity and quality of stormwater effluent on the outflow from the catchment area and the operation of the stormwater treatment plant and storm overflow by means of a hydrodynamic model developed in the SWMM programme (Storm Water Management Model). The hydrodynamic model of the catchment area and of the DC used in the paper has been previously calibrated and an analysis of its sensitivity and uncertainty was performed using the GLUE + GSA method (Szeląg *et al.* 2016).

## RAINFALL DATA

The evaluation of storm overflow operation and determination of independent rainfall events were based on rainfall data gained from a rainfall station (collector S1 in Figure 1) located 2 km away from the catchment basin and

established in 2008. A minimum antecedent period of 4 h (DWA-A 118E 2006) was used for the calculation. In the work, a rainfall event was assumed as a fall episode when the total minimum precipitation depth is not less than 3.0 mm (Fu *et al.* 2011; Fu & Kapelan 2013). The duration of rainfall ( $t_d$ ) of observed events was between 20 min and 2.366 min, while the antecedent period was in the range of 0.16–60 days. The total depth of rainfall in precipitation events varied between 3.0 mm and 45.2 mm, while the maximum rainfall depth in the precipitation events lasting 15, 30 or 45 minutes was respectively  $P_{t=15} = 0.3$ –20.4 mm,  $P_{t=30} = 0.6$ –25.4 mm and  $P_{t=45} = 1.1$ –28.5 mm.

On the basis of the observed rainfall in the period 2008–2017, independent rainfall events were determined. The number of rainfall events ( $N$ ) in individual years varied from 36 to 58, while their average value was equal to 45 episodes. Using the methodology to identify precipitation events described above, 463 precipitation episodes were selected in the available data series. The selected rainfall events have been parameterized by: total precipitation depth ( $P_c$ ); event duration ( $t_d$ ); maximum rainfall depth in the rainfall events lasting respectively 15, 30 or 45 minutes ( $P_{t=15}$ ,  $P_{t=30}$ ,  $P_{t=45}$ ); the variability of these characteristics was described by empirical distributions. Then, in order to obtain the best possible theoretical and empirical data matching, the following statistical distributions (Adams & Papa 1999; Bacchi *et al.* 2008; Andrés-Doménech *et al.* 2010) were considered: Weibull, chi-square statistic, exponential, Generalized Extreme Value (GEV), Gumbel, Fisher-Tippet, gamma, log-normal, Pareto and beta. The Kolmogorov-Smirnov test was used to assess the conformity of the empirical and theoretical distributions.

## METHODOLOGY

The article provides a probabilistic model for forecasting the annual number of stormwater discharges in a storm overflow facility under investigation. It should be clearly stated here that the model developed and examined in the paper does not concern forecasting the flow of stormwater in the drainage system, but concerns forecasting the events consisting of stormwater overflows occurring in the storm overflow installed on a pipe of that system. The models used to

forecast the stormwater flow itself were presented in detail in the works by Szeląg *et al.* (2018). The model developed in the paper assumes that the operation of the overflow is random and depends on the variability of rainfall in the annual cycle. The logistic regression model was used for forecasting storm overflow operation, while the Monte Carlo (MC) method was used for simulation of rainfall series. If the storm overflow operation was explained by only one variable or the explained variables were independent, then MC sampling was performed by means of theoretical distributions describing rainfall characteristics, as presented in works by Adams & Papa (1999) and Szeląg & Bąk (2017). Where the analysed explanatory variables  $x_i$  were correlated, and the relationship between the variables characterizing rainfall was determined by calculating the correlation coefficient of Spearman  $R$ . Assuming that this is a sufficient description of the relationship between the variables, rainfall events were generated using the Iman & Conover (1982) method, described below. The calculation diagram of the model proposed in the paper is presented in Figure 2.

From the diagram presented, the calculation of the annual number of discharges by storm overflow is carried out in the following stages:

1. identification of parameters ( $\alpha_i$ ) and explanatory variables ( $x_i$  – precipitation characteristics) in the logit model,
2. evaluation of interdependencies between explanatory variables ( $x_i$ ) based on the Spearman correlation coefficient ( $R$ ),
3.  $T$  – th simulation of  $N$  rainfall events in the synthetic rainfall data series by means of Monte Carlo method,
4.  $M$  – th simulation of precipitation characteristics ( $x_i$ ) of rain in the synthetic rainfall data series by means of Monte Carlo method, taking into account the correlation; the analyses conducted assume  $N = M = 5000$ ,
5. determination of the number of discharges by storm overflow in modelled synthetic rainfall series based on Equation (1) (see below),
6. determination of the empirical curve describing the probability of exceeding the annual number of discharges in storm overflow (CDF – cumulative distribution function).

## LOGISTIC REGRESSION

The logistic regression model is a model often used for the analysis of binary data (single-zero-single-one), which means that it is often used in economic sciences, social sciences, medicine, etc. (Bagley *et al.* 2001). Moreover, this model is increasingly being used to assess the performance of biological reactors in wastewater treatment plants (Bayo *et al.* 2006), in assessing the technical condition of water supply and stormwater pipes (Tran *et al.* 2007) and in the sensitivity analysis of statistical models (Bartkiewicz *et al.* 2016). The logistic model discussed in the paper is a special case of a generalized linear model of the following form:

$$p = P(Y = 1 | X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{\exp(\alpha_0 + \sum_{i=1}^k \alpha_i \cdot x_i)}{1 + \exp(\alpha_0 + \sum_{i=1}^k \alpha_i \cdot x_i)} \quad (1)$$

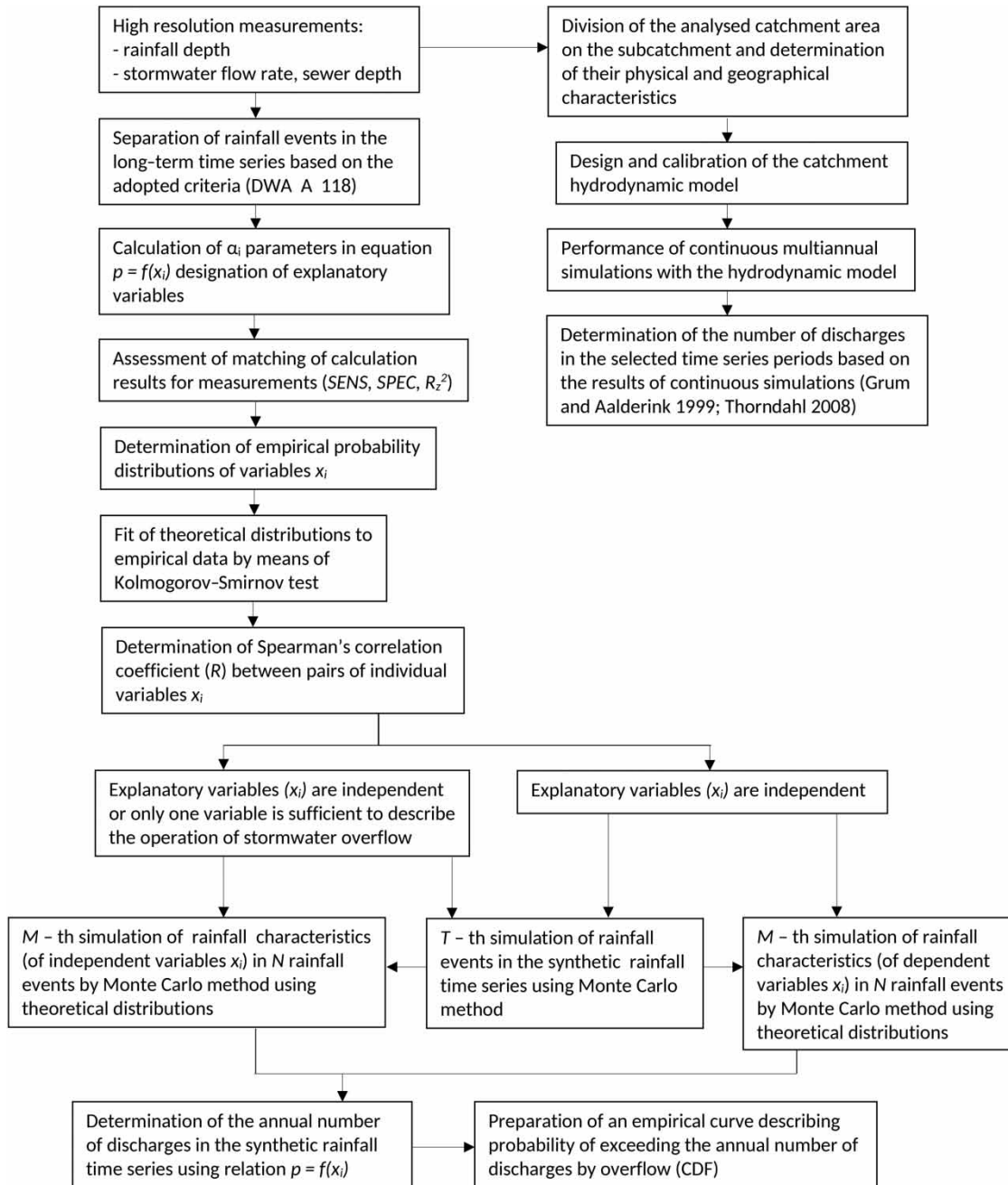
in which the probability ( $P$ ) of occurrence of  $s$  – stormwater discharges in the annual rainfall data series ( $N$ ) is described by the binomial distribution:

$$P(Y = s) = \binom{N}{s} \cdot p_i^s \cdot (1 - p_i)^{N-s} \quad (2)$$

where  $\alpha_0$  – free term,  $\alpha_i$  – empirical coefficients determined by the maximum likelihood method,  $x_i$  – variables, i.e.: total precipitation depth ( $P_c$ ), rainfall duration ( $t_d$ ), antecedent period ( $t_{an}$ ), maximum rainfall depth in rainfall events lasting 15, 30 and 45 minutes ( $P_{t=15}$ ,  $P_{t=30}$ ,  $P_{t=45}$ ),  $s$  – annual number of overflow discharges,  $N$  – number of rainfall events per year,  $p_i$  – probability of the occurrence of a discharge in the storm overflow in  $i$ -th rainfall event.

It was assumed in the analyses performed that stormwater overflowing in single rainfall events will take place when the value  $p$  calculated by Equation (1) will not be less than 0.50, which corresponds to the condition of the form:

$$\alpha_0 + \sum_{i=1}^k \alpha_i \cdot x_i = 0 \quad (3)$$



**Figure 2** | Calculation diagram of a probabilistic model for forecasting the number of discharges by storm overflow. (SENS – Sensitivity; SPEC – Specificity; CDF – Cumulative distribution functions.)

with the multidimensional space of variables ( $x_i$ ) that account for stormwater discharge.

The variables explaining the rainfall characteristics ( $P_c$ ,  $t_d$ ,  $t_{ans}$ ,  $P_{t=15}$ ,  $P_{t=30}$ ,  $P_{t=45}$ ) during the considered rainfall episodes were used for the analysis of the overflow

operation. Only statistically significant variables at the adopted confidence level ( $\alpha = 0.95$ ) are included in the statistical model. Quality of the model, i.e. adjustment of calculation results to the measurements data in the logistic regression model, has been determined on the basis of the following

coefficients: sensitivity (*SENS*, %) , specificity (*SPEC*, %) and reckoning error ( $R_z^2$ ) (Bartkiewicz et al. 2016).

In order to estimate  $\alpha_i$  parameters in the logit model, the results of measurements data concerning storm overflow operation carried out in the years 2009–2011 were used, where overflow discharge occurred in 69 cases in 188 precipitation events, and 42 overflow discharges occurred (2012–2014) in 93 cases of the maximum filling of the DC. The model was validated using 10 independent rainfall events.

## RAINFALL EVENT GENERATOR

In this paper, if the values characterizing rainfall ( $P_c$ ,  $t_d$ ,  $t_{an}$ ,  $P_{t=15}$ ,  $P_{t=30}$ ,  $P_{t=45}$ ) and also the explanatory variables in the logit model show significant dependence, then for its mathematical description the value of Spearman correlation coefficient was used (Iman & Conover 1982). The variability of the rainfall characteristics defined in the form of boundary distributions is described by the functions of theoretical distributions. Based on the established correlation coefficients and boundary distributions, individual events were randomly sampled using the IC method.

The IC method is an algorithm commonly used in the Monte Carlo method. This method is implemented in several statistical packages (Risk, Crystal Ball, STATISTICA), which makes it available to a wide range of users. The method can be properly used when the following conditions are met (Wu & Tsang 2004):

- Mean values ( $\mu_1, \mu_2, \dots, \mu_i$ )<sub>s</sub> and standard deviations ( $\sigma_1, \sigma_2, \dots, \sigma_i$ )<sub>s</sub> of individual variables in the predicted rainfall data series do not differ more than  $\varepsilon < 5\%$  from relevant values of the theoretical distributions, where  $\varepsilon$  is the relative difference between the modelled and measured value.
- The empirical distributions of the modelled variable values ( $x_i$ ) are in line with theoretical distributions; for the verification of this condition it is recommended to use the Kolmogorov-Smirnov test.
- The correlation coefficient ( $R$ ) between individual dependent variables ( $x_i$ ) obtained for data received from the MC simulation do not differ more than about  $\varepsilon_c < 5\%$  from the  $R$  value calculated for empirical data.

A detailed description of the assumptions and theoretical considerations concerning the IC method can be found in the works of Iman & Conover (1982), Wu & Tsang (2004) and Tarpanelli et al. (2012).

## RESULTS

Using the hydrodynamic model of the water catchment with the DC, continuous simulation of storm overflow operation based on continuous measurements series of rainfall from 2008 to 2017 was performed. On the basis of the calculations performed (Table 1), the annual number of discharges by storm overflow located in the DC was determined and compared with the results of measurements collected from 2008 to 2011. The annual number of discharges calculated by means of the developed logit model described by Equation (2) is also presented in Table 1. Based on the tabular data, one can show that in the analysed time period the number of overflow discharges in particular years ranged from 13 to 29. The number of discharges measured and projected with the hydrodynamic model shows a slight variation, which is indicated by the numerical values in Table 1. Moreover, the analyses carried out shows that the number of rainfall events in particular years ranged

**Table 1** | Summary of the annual number of stormwater discharges and rainfall events in the period 2008–2017

Year	Number of discharges by overflow calculated (SWMM)/measured	Number of discharges forecasted by logit model	Number of rainfall events
2008	13/15	14	43
2009	15/16	16	47
2010	17/18	19	47
2011	19/20	19	51
2012	21/13*	20/14*	36
2013	22/13*	20/13*	41
2014	29/16*	28/16*	44
2015	26	26	58
2016	22	21	44
2017	17	17	38

where: numbers of stormwater discharges for years 2008–2011 were obtained using continuous simulations, \*outflow events for 2012–2014 were acquired on the basis of observed DC maximal filling.

from 36 to 58, while the average number of rainfall events was 45.

Based on the results of measurements of overflow operation and on the rainfall characteristics specified for particular stormwater discharges ( $P_c$ ,  $t_d$ ), a logistic regression model was developed, described as follows:

$$p = \frac{\exp(\alpha_1 \cdot P_c + \alpha_2 \cdot t_d + \alpha_0)}{1 + \exp(\alpha_1 \cdot P_c + \alpha_2 \cdot t_d + \alpha_0)} \quad (4)$$

The values of  $\alpha_i$  coefficients and the measures for matching calculation results to measurements are given in Table 2. Determined parameters (*SENS*, *SPEC*,  $R_z^2$ ) indicating the matching of calculation results to the measurements show that the obtained logistic regression model  $p = f(P_c, t_d)$  has satisfactory predictive capability. Taking into account individual matching parameters expressing the accuracy of model classification (*SPEC* = 96.9%) it was found that the model correctly identified 103 events for 106 observed stormwater discharge events. Based on the value of *SENS* = 98.2% it was found that out of 165 events without stormwater discharged, the model appropriately classified 162 events.

Analysing the  $R_z^2$  value it was found that out of 271 observed events the model correctly classified 264 cases. Validation of the logit model showed that out of 10 events simulation results were obtained in 8 cases according to the measurements. A high compliance of the annual discharges number in the period 2008–2011 with the measurement results confirms a satisfactory predictive capability of the obtained logistic regression model.

The obtained dependency described by Equation (4) is presented in Figure 3(c). Analysis of the course of curves in

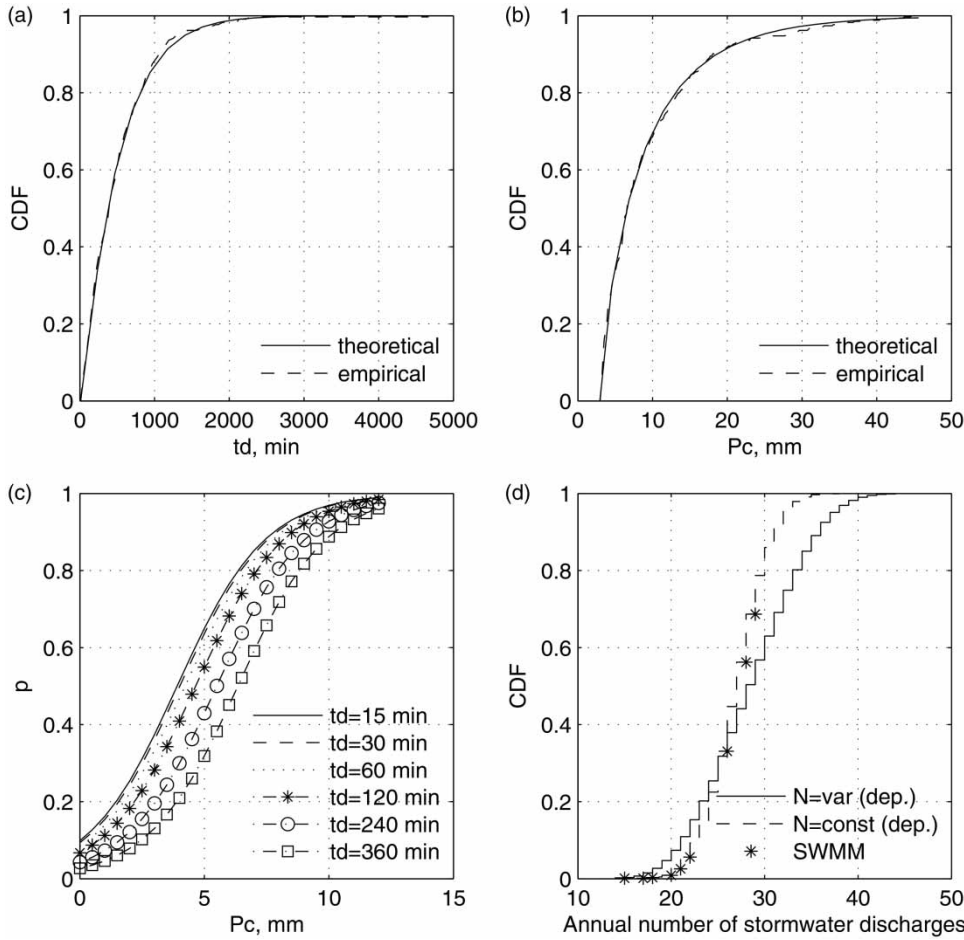
Figure 3(c) indicates that the increase of rain duration by the assumption  $P_c = \text{const.}$  (reduction of the average rainfall intensity  $q = P_c - t_d^{-1}$ ) leads to a decrease of the probability of a discharge in storm overflow. It should be noted that the obtained relation  $p = f(P_c, t_d)$  has been confirmed by Grum & Aalderink (1999), and Thorndahl & Willems (2008), who analysed the operation of storm overflows in urban catchment areas. In view of the above observation and attempting to provide a universal relation to the prediction of the occurrence of a stormwater overflow in a rainwater drainage system, Equation (4) was transformed in the form:

$$P_c - \frac{\alpha_2}{\alpha_1} \cdot t_d - \frac{\alpha_0}{\alpha_1} = P_c - \frac{\alpha_2}{\alpha_1} \cdot t_d - d_{min} = 0 \quad (5)$$

The obtained Equation (5) is equal to the condition for the minimum rainfall ( $d_{min}$ ), the exceeding of which causes the discharge by an overflow and which was obtained by Thorndahl & Willems (2008) analysing the discharge volumes. The calculated value of parameter  $\alpha_0/\alpha_1 = 2.152/0.566 = 3.80$  in Equation (5) has a physical interpretation and is equal to the weighted average value of retention height ( $d_r = d_{min}$ ) of the analysed catchment area. The result obtained is of significant practical importance as it enables the use of the developed logit model for other urban catchments, and in particular for uncontrolled ones. Equation (5) obtained in the paper is very important from the point of view of the forecast of storm overflow operation in uncontrolled basins, where at present, in order to analyse its functioning, a hydrodynamic model is created. It should be noted that despite the development of numerous empirical relationships for the determination of parameters in SWMM models, the results of calculations do not always accurately reflect the actual state of affairs (Guo & Urbonas 2009). At the same time, it should be remembered that the logit or hydrodynamic model is only a tool for forecasting the operation of the storm overflow. In order to determine the annual number of discharges, multiannual rainfall series (30 years according to DWA A-118E) are necessary, in the case of the hydrodynamic model for continuous simulation, and in the case of the probabilistic model for the development of the precipitation generator. Literature (Fu & Butler 2014) shows that a reliable precipitation generator can be built on the basis of a 10 year series.

**Table 2** | Summary of determined empirical  $\alpha_i$  coefficients in Equation (4) and of parameters for matching calculation results to measurements

Variable	$\alpha_i$	Standard deviation	$p$ from test
$P_c$	0.566	0.0620	0.002
$t_d$	-0.004	0.0012	0.009
$\alpha_0$	-2.152	0.1860	0.004
$R_{McFadden}^2 = 0.868$		<i>SENS</i> = 0.982	
$R_{Cox-Snell}^2 = 0.798$		<i>SPEC</i> = 0.969	
$R_{Negelerke}^2 = 0.868$		$R_z^2 = 0.974$	



**Figure 3** | (a) Comparison of empirical and theoretical distributions for rainfall duration. (b) Comparison of empirical and theoretical distributions for total rainfall depth ( $P_c$ ). (c) Logit curves showing relationships  $p = f(P_c, t_d)$ . (d) Theoretical cumulated probability density distribution (CDF) describing probability of exceeding the annual number of discharges by storm overflow.

Using the obtained relation shown in Equation (5) empirical distributions (distribution functions) for the total rainfall depth ( $P_c$ ) and duration ( $t_d$ ) were determined. The theoretical distributions based on Kolmogorov-Smirnov test results were adjusted to the obtained empirical distributions; the results are presented in Table 3 and in Figures 3(a) and 3(b).

Analysing the obtained values  $D$  and  $p$  (see Table 3), by means of a Kolmogorov-Smirnov test, it was found that for the empirical data (concerning the total rainfall depth and duration) the Weibull distribution described by the following formula would be best suited:

$$F(x_i) = 1 - \exp\left\{-\left(\frac{x - \mu}{\gamma}\right)^\beta\right\} \quad (6)$$

**Table 3** | Results of calculations ( $p, D$ ) of Kolmogorov-Smirnov tests for distributions of  $P_c$  and  $t_d$

Probability distribution	$P_c$		$t_d$	
	D-critical	p-test	D-critical	p-test
chi-square	0.231	<0.0001	0.546	<0.0001
beta	0.108	<0.0001	0.068	0.0830
Weibull	0.050	0.3570	0.030	0.8520
exponential	0.273	<0.0001	0.057	0.2080
GEV	0.399	<0.0001	0.128	<0.0001
Pareto	0.375	<0.0001	0.403	<0.0001
Fisher-Tippet	0.485	<0.0001	0.382	<0.0001
log-normal	0.087	<0.0001	0.091	0.0060
gamma	0.951	<0.0001	1.000	<0.0001
Gumbel	0.951	<0.0001	1.000	<0.0001



where  $x$ -variable;  $\beta$ ,  $\gamma$ ,  $\mu$ -empirical parameters of the distributions, in the case of  $P_c$  are  $\mu = 3.0$  mm;  $\beta = 0.714$  and  $\gamma = 5.336$  mm and for  $t_d$  are  $\mu = 12.0$  min;  $\beta = 0.994$  and  $\gamma = 491.94$  min.

The calculations performed showed that the difference between the modelled and measured mean values ( $\mu$ ) and standard deviation values ( $\sigma$ ) for  $P_c$  and  $t_d$  does not exceed 1%. The difference in the value of the correlation coefficient ( $R$ ) derived from MC simulation results and calculated from the measurement data does not exceed 2%. In addition, the assessment of the conformity of the theoretical distributions and those obtained by Monte Carlo simulation using the IC method showed that the probability values  $p$  in the Kolmogorov-Smirnov test are  $p = 0.54$  for the total rainfall depth and  $p = 0.59$ , respectively. Taking into account the above calculation results it can be concluded that the number of  $N = 5000$  samples used is sufficient and the theoretical distributions of  $P_c$  and  $t_d$  and those ones received from MC simulations are compatible.

Based on the logit regression model described by Equation (4) the theoretical distributions ( $P_c$  and  $t_d$ ) were determined and the annual number of stormwater discharges was calculated with a probabilistic model in accordance with the scheme in Figure 1. In the analyses performed, two cases have been considered: the first option takes into account the variable number of rainfall events per year ( $N = \text{var}$ ) and the second option omits this fact and takes into account the average value of rainfall events occurring in singular years ( $N = \text{const.}$ ). The results of the calculations performed with the probabilistic model (Figure 3(d)) were compared with the results of continuous simulations carried out with the SWMM programme for the period 2008–2017.

From data in Figure 3(d), the expected value of the number of discharges in storm overflow ( $n = 28$ ) obtained under the assumption of a uniform distribution of the number of rainfall events in a year ( $N = \text{var.}$ ) is higher than when the average number of rainfall episodes ( $n = 27$ ) is assumed. Moreover, the obtained curves show that for  $p < 0.28$  the number of discharges in storm overflow obtained with the assumption  $N = \text{const.}$  is lower than when the stochastic character of the number of rainfall events per year is taken into account. If the probability value  $p$  (Figure 3(d)) is greater than 0.28 then the number

of discharges in storm overflow obtained under the assumption of the average annual number of rainfall episodes is greater than the discharges number determined while taking into account the random character of the number of rainfall events in the year. Percentile values  $p = 0.05$  and  $p = 0.95$  corresponding to the annual number of discharges in storm overflow are 22 and 32, respectively, when  $N = \text{const.}$ ; when the number of annual rainfall episodes is of a probabilistic nature then the appropriate percentile values are 19 and 37. Summing up the results of the calculations obtained, it can be concluded that at the stage of storm overflow designing by means of the presented probabilistic model, omitting the stochastic nature of the number of rainfall events ( $N = \text{const.}$ ) can lead to under-positioning of the overflow edge and thus to exceeding the permitted number of wastewater discharges. Moreover it is worth noting that the results of simulation of the annual number of discharges by means of the calibrated hydrodynamic model are within the range of solutions obtained by the probabilistic model which indicates that the method presented in the paper can be applied in practice.

## CONCLUSIONS

The paper presents the results of calculations of the annual number of stormwater overflows using the developed probabilistic model and based on the results of continuous simulations with a hydrodynamic model.

The paper presents an innovative application of the logistic regression model to the forecast of the operation of an underground infrastructure facility (storm overflow) located in the rainwater drainage system. The calculations have confirmed that the logit model is a useful tool and can be used to simulate complex flow phenomena during precipitation episodes. In practical terms, the logit model can be used as an alternative tool for forecasting the storm overflow operation by means of a hydrodynamic model, and in particular in the event of problems with its calibration.

It is worth noting that the empirical parameters used in the logit model have a physical interpretation, which makes it possible to apply that model in unmeasured urban basins. However, in order to confirm this, further studies are

needed in catchments with different physical and geographical characteristics.

The probabilistic model presented in the paper for the forecast of the annual number of stormwater overflows is an innovative combination of logit model and rainfall generator (IC method). This model takes into account the stochastic nature of the number of rainfall events in the year, which has so far been ignored in the calculations. The developed probabilistic model makes it possible to estimate the annual number of discharges in the case of limited access to long-term rainfall measurements (30 years); however, data are necessary to develop a model (characteristics of the catchment area) to simulate the operation of the overflow and precipitation generator.

The results obtained in the paper indicate that ignoring the stochastic nature of the number of rainfall events in the annual cycle may lead to lowering the designed storm overflow edge and, consequently, to exceeding the permissible number of discharges.

The analyses of the annual number of discharges by the storm overflow showed that the results of calculations obtained with the use of a hydrodynamic model are in the range of probabilistic solutions presented in the paper with the use of the mathematical model. This confirms the practical character of the method discussed and applied in the paper for the forecast of the annual number of stormwater discharges by means of a storm overflow.

Taking into account the above considerations, it seems justified to conduct further research aimed at the possibility of using the logistic regression model to assess the functioning of other facilities located in the rainwater drainage networks (e.g. retention reservoirs, sewage treatment plants). Moreover, due to the fact that in many cases we have limited data resources (short period of rainfall and flow measurements), further studies are recommended in order to assess the influence of selected boundary distributions and uncertainty of estimated parameters in the logit models on the results of simulations of storm overflow operation.

## REFERENCES

- Adams, B. J. & Papa, F. 1999 *Urban Stormwater Management Planning with Analytical Probabilistic Models*. John Wiley & Sons, New York.
- Andrés-Doménech, I., Múnera, J. C., Francés, F. & Marco, J. B. 2010 [Coupling urban event-based and catchment continuous modelling for combined sewer overflow river impact assessment](#). *Hydrology and Earth System Sciences* **14**, 2057–2072. doi:10.5194/hess-14-2057-2010.
- Bacchi, B., Balistrocchi, M. & Grossi, G. 2008 [Proposal of a semi-probabilistic approach for storage facility design](#). *Urban Water Journal* **5** (3), 195–208. <http://dx.doi.org/10.1080/15730620801980723>.
- Bagley, S. C., White, H. & Golomb, B. A. 2001 [Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain](#). *Journal of Clinical Epidemiology* **54** (10), 979–985. [https://doi.org/10.1016/S0895-4356\(01\)00372-9](https://doi.org/10.1016/S0895-4356(01)00372-9).
- Bartkiewicz, L., Szeląg, B. & Studziński, J. 2016 [Impact assessment of input variables and ANN model structure on forecasting wastewater inflow into sewage treatment plants](#). *Ochrona Środowiska* **38** (2), 29–36 (in Polish).
- Bayo, J., Angosto, J. M. & Serrano-Aniorte, J. 2006 [Evaluation of physicochemical parameters influencing bulking episodes in a municipal wastewater treatment plant](#). *Water Pollution VIII: Modelling, Monitoring and Management*, Bologna, September 4-6, 531–541. doi:10.2495/WP060521.
- Benoist, A. P. & Lambertus, L. 1989 [A methodology for assessment of frequency distributions of combined sewer overflow volumes](#). *Water Resources* **23** (4), 487–493. [https://doi.org/10.1016/0043-1354\(89\)90140-1](https://doi.org/10.1016/0043-1354(89)90140-1).
- Dogliani, A., Promativo, F., Laucelli, D., Monno, V., Khu, S. T. & Giustolisi, O. 2009 [An integrated modelling approach for the assessment of land use change effects on wastewater infrastructure](#). *Environmental Modelling and Software* **24** (12), 1522–1528.
- DWA-A 118E 2006 *Hydraulic Dimensioning and Verification of Drain and Sewer Systems*. DWA German Association for Water, Wastewater and Waste.
- Fidala-Szope, M., Sawicka-Siarkiewicz, H. & Koczyk, A. 1999 *Protection of Surface Waters From the Discharge of Wastewater From Combined Sewer Systems: Guide*. Publishing House of the Institute of Environmental Protection, Warszawa (in Polish).
- Fu, G. & Butler, D. 2014 [Copula-based frequency analysis of overflow and flooding in urban drainage systems](#). *Journal of Hydrology* **510**, 49–58. <https://doi.org/10.1016/j.jhydrol.2013.12.006>.
- Fu, G. & Kapelan, Z. 2013 [Flood analysis of urban drainage systems: probabilistic dependence structure of rainfall characteristics and fuzzy model parameters](#). *Journal of Hydroinformatics* **15** (3), 687–699. doi: 10.2166/hydro.2012.160.
- Fu, G., Butler, D., Khu, S. T. & Sun, S. 2011 [Imprecise probabilistic evaluation of sewer flooding in urban drainage systems using random set theory](#). *Water Resources Research* **47** (2), 1–13. doi: 10.1029/2009WR008944.
- Grum, M. & Aalderink, R. H. 1999 [Uncertainty in return period analysis of combined sewer overflow effects using embedded Monte Carlo simulations](#). *Water Science and Technology* **39** (4), 233–240. [https://doi.org/10.1016/S0273-1223\(99\)00063-3](https://doi.org/10.1016/S0273-1223(99)00063-3).

- Guo, J. & Urbonas, B. 2009 Conversion of natural watershed to kinematic wave cascading plane. *Journal of Hydrologic Engineering* **14** (8), 839–846. doi:10.1061/(ASCE)HE.1943-5584.0000045.
- Iman, R. L., Conover, W. J., Sandia National Laboratories, Albuquerque, NM, 87185 1982 A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics – Simulation and Computation* **11** (3), 311–334. <http://dx.doi.org/10.1080/03610918208812265>.
- Mantegazza, S. A., Alessandro, G., Mambretti, S. & Camylyn, L. 2010 Designing CSO storage tanks in Italy: a comparison between normative criteria and dynamic modelling methods. *Urban Water Journal* **7** (3), 211–216. <http://dx.doi.org/10.1080/1573062X.2010.484499>.
- Osorio, F., Muhaisen, O. & García, P. A. 2009 Copula-based simulation for the estimation of optimal volume for a detention basin. *Journal of Hydrologic Engineering* **14** (12), 1378–1382. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000124](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000124).
- Price, R. 2000 Hydroinformatics and urban drainage: an agenda for the beginning of the 21st century. *Journal of Hydroinformatics* **2** (2), 133–147.
- Rauch, W., Henze, M., Koncsos, L., Reichert, P., Shanahan, P., Somlyódy, L. & Vanrolleghem, P. 1998 River water quality modelling: I. state of the art. *Water Science and Technology* **38** (11), 237–244. [https://doi.org/10.1016/S0273-1223\(98\)00660-X](https://doi.org/10.1016/S0273-1223(98)00660-X).
- Szeląg, B. & Bąk, Ł. 2017 Probabilistic model for the annual number of storm overflow discharges in a stormwater drainage system. *Urban Water Journal* **14** (6), 604–611. <http://dx.doi.org/10.1080/1573062X.2016.1223860>.
- Szeląg, B., Górski, J., Bąk, Ł. & Górka, K. 2013 Modelling of stormwater quantity and quality on the example of urbanised catchment in Kielce. *Ecological Chemistry and Engineering A* **20** (11), 1305–1316. doi: 10.2428/ecea.2013.20(11)118.
- Szeląg, B., Kiczko, A. & Dąbek, L. 2016 Sensitivity and uncertainty analysis of hydrodynamic model (SWMM) for storm water runoff forecasting in an urban basin – a case study. *Ochrona Środowiska* **38** (3), 15–22 (in Polish).
- Szeląg, B., Chmielowski, K. & Wąsik, E. 2018 Simulation of a storm overflow and stormwater treatment plant with probabilistic and hydrodynamic models. *Urban Water Journal* (under review).
- Tarpanelli, A., Franchini, M., Brocca, L., Camici, S., Melone, F. & Moramarco, T. 2012 A simple approach for stochastic generation of spatial rainfall patterns. *Journal of Hydrology* **472**, 63–76. <https://doi.org/10.1016/j.jhydrol.2012.09.010>.
- Thorndahl, S. 2009 Stochastic long term modelling of a drainage system with estimation of return period uncertainty. *Water Science and Technology* **59** (12), 2331–2339. doi:10.2166/wst.2009.305.
- Thorndahl, S. & Willems, P. 2008 Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first-order reliability method and parameterization of local rain series. *Water Research* **42** (1–2), 455–466. doi: 10.1016/j.watres.2007.07.038.
- Tran, D. H., Ng, A. W. M., Perera, B. J. C., Burn, S. & Davis, P. 2007 Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal* **3** (3), 175–184. <http://dx.doi.org/10.1080/15730620600961684>.
- Urcikán, P. & Rusnák, D. 2006 A complex method of calculating critical rain intensity for the needs of storm overflow design. *Ochrona Środowiska* **28** (1), 33–38 (in Polish).
- Veldkamp, R. G. & Wiggers, J. B. M. 1997 A statistical approach to pollutant emissions from combined sewer systems. *Water Science and Technology* **36** (8–9), 95–100. [https://doi.org/10.1016/S0273-1223\(97\)00613-6](https://doi.org/10.1016/S0273-1223(97)00613-6).
- Wu, F. C. & Tsang, Y. P. 2004 Second-order Monte Carlo uncertainty/variability analysis using correlated model parameters: application to salmonid embryo survival risk assessment. *Ecological Modelling* **177** (3–4), 393–414. <https://doi.org/10.1016/j.ecolmodel.2004.02.016>.

First received 8 January 2018; accepted in revised form 31 May 2018. Available online 2 July 2018