Efficient job list creation for long-term statistical modelling of combined sewer overflows

Allen, Kiara Marie; Mollerup, Ane Loft; Rasmussen, Søren Feilberg; Sørup, Hjalte Jomo Danielsen

Published in:
Water Science and Technology

Link to article, DOI:
10.2166/wst.2022.065

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Allen, K. M., Mollerup, A. L., Rasmussen, S. F., & Sørup, H. J. D. (2022). Efficient job list creation for long-term statistical modelling of combined sewer overflows. *Water Science and Technology*, *85*(5), 1424–1433. Article wst2022065. https://doi.org/10.2166/wst.2022.065

Water Science & Technology

Water Science & Technology Vol 85 No 5, 1424 doi: 10.2166/wst.2022.065

Efficient job list creation for long-term statistical modelling of combined sewer overflows

Kiara Marie Allen [a,b], Ane Loft Mollerup [a], Søren Feilberg Rasmussen [a] and Hjalte Jomo Danielsen Sørup [b,*]

[a] Novafos A/S, Plan Rens & Afløb, Birkerød, Denmark
[b] Department of Environmental Engineering, Climate & Monitoring, Technical University of Denmark, Lyngby, Denmark
*Corresponding author. E-mail: hjds@env.dtu.dk

KMA, 0000-0003-4687-5331; HJDS, 0000-0002-7110-6975

ABSTRACT

The modelling of urban drainage systems is an important aspect of their design process and long-term statistical modelling using historical rain series is commonly used. The objective of this study is to determine whether logistic regression models that use rainfall event statistics can be a viable alternative to create job lists with fewer extraneous events. Two methods are used to develop a regression model; both use iterative stepwise algorithms to select the rain variables to include and both perform similarly. The resulting model is able to capture ~90% of the relevant events with ~50% fewer jobs compared to the reference job list. The results suggest that there is no right threshold to use, but instead this methodology facilitates balancing the number of jobs with the desired level of precision of the results. In all cases, it is possible to greatly decrease the number of jobs that need to be run. The methodology works relatively well on different nodes in the system, though node characteristics appear to impact the amount of CSOs captured.

Key words: CSO, logistic regression models, LTS simulations, sewer system, urban water management

HIGHLIGHTS

- Job lists for LTS simulations can be targeted using logistic regression models.
- The logistic regression models draws on statistical data from the rainfall input data.
- The number of jobs can be drastically reduced while mostly maintaining the relevant jobs.
- The proposed methodology can be used to run more efficient LTS simulations.

INTRODUCTION

The modelling of urban drainage systems is an important aspect of their design process, as this helps to ensure that the final system design will meet its service criteria and function adequately (Botturi *et al.* 2021). Due to the complexity of these systems, there are different data requirements depending on which aspect is to be tested. For instance, the pipe sizes in the system can be adequately tested using an artificial rain series, such as a Chicago design storm (Keifer & Chu 1957; Jean *et al.* 2018). Others, such as the frequency of combined sewer overflows (CSOs) or implementation of nature-based solutions are best evaluated using long-term historical data sets, due to non-linear responses (Locatelli *et al.* 2015; Andersen *et al.* 2017; Jean *et al.* 2018; Sørup *et al.* 2018; Sørup & Lerer 2021). Further, climate change impacts on urban drainage systems, and complex blue-green-grey mitigation strategies responding to these impacts (Sørup *et al.* 2019), require intensive modelling to ensure proper hydraulic function of the complete system (Cristiano *et al.* 2018; Jean *et al.* 2018; van Uytven *et al.* 2020).

Some studies have focused on how to reduce the computational burden when analyzing urban drainage systems; either by exploring the link between rainfall properties and overflow phenomena in the sewer systems (Jean *et al.* 2018; McGrath *et al.* 2019), by utilizing machine learning techniques to avoid physics based models for flooding (Löwe *et al.* 2021; Vorobevskii *et al.* 2020) or by introducing model predictive or real time control into the system (Rathnayake and Faisal Anwar 2019; Bachmann-Machnik *et al.* 2021; El Ghazouli *et al.* 2021). Common for all studies is a recognition of the complexity involved

in determining the performance of urban drainage systems, and a goal of minimizing the efforts involved in designing, analyzing and operating these systems.

Continuous modelling of urban water systems is most often not feasible, so to minimize the amount of simulation time required for these data sets long-term statistical (LTS) modelling is commonly used (Thorndahl 2009; Davidsen *et al.* 2017; Jean *et al.* 2018) if it is not desirable to reduce the complexity of the model through lumping (Ochoa-Rodriguez *et al.* 2015; Thrysøe *et al.* 2019). In LTS modelling only the rain events that are likely to result in a relevant response are simulated while the others are skipped. These events are specified before the start of the simulation in a job list. Typically, events are selected for inclusion in the job list based on a minimum rain depth or rainfall intensity criteria based on the type of output that is of interest (Jean *et al.* 2018; McGrath *et al.* 2019). This selection process often works well, but can result in relatively few jobs that result in a desired response in relation to the total number of jobs included on the job list if the response sought for is very non-linear. Looking at the urban systems and their statistical properties has been done historically in order to best determine which input data to use to observe a given response (Schaarup-Jensen *et al.* 2009; Fontanazza *et al.* 2011; van Bijnen *et al.* 2012; Ochoa-Rodriguez *et al.* 2015; Cristiano *et al.* 2018). As such, techniques that filter out some of these superfluous jobs could result in much more efficient LTS simulations with a larger percentage of jobs resulting in relevant responses.

Jean *et al.* (2018) pointed out that continuous simulations will result in the least erroneous result while Vorobevskii *et al.* (2020) showed how rainfall characteristics could be used to select relevant events in flood modelling. The objective of this study is to determine the rain events to be included on an LTS simulation job list to simulate CSOs, given statistical information from the rainfall time series only. In order to do so objectively, logistic regression models are formulated to link the statistical properties of the rainfall to the CSO output of the LTS simulations. This type of model is chosen as they are more selective than the typical job list creation method as it allows for more flexibility in what variables can be considered and more sophistication in variable values than whether they exceed a defined criterion. The full job list is used as a reference for comparison to the model results to evaluate how large a reduction in jobs that need to be run can be without missing too many relevant CSO causing events.

METHODS

Data and model

This study uses a MIKE URBAN model (DHI 2020) of the eastern part of Rudersdal municipality to obtain CSO information. This area was selected due to the 41 years of rainfall data available from The Water Pollution Committee of The Society of Danish Engineers' (SVK) rain gauge system for this area from a single tipping bucket gauge at the Vedbæk wastewater treatment plant. This was the rain data used in this analysis and is assumed to be applicable for and uniformly distributed over the whole model region.

The model area is shown in Figure 1 along with the location of the rain gauge and the nodes of interest. The characteristics of these nodes are summarized in Table 1. The six nodes described in Table 1 are identified as being representative for frequent overflow behavior in the system, as they each have several CSOs during each of the calibration periods and varying storage volumes and emptying times. Due to this their overflow data is used throughout this article.

This rain data is split into equivalent calibration and test subsets, where the calibration set is used to build the logistic models and the test set is used to test them afterwards. To avoid the changes in precipitation patterns due to climate change, the data is split into calibration and test periods in alternating chronological order (Gregersen *et al.* 2015). As the focus of this analysis is on frequent return periods, the data is split into periods of approximately five years. The allocations of these periods are shown in Table 2. A separate simulation is run for each of these periods. As shown, all periods have comparable numbers of rain events and CSOs but with expected differences between different periods. Both the calibration and test data sets include periods with below and above average numbers of each.

A dry weather separation time of 6 hours is used to separate rain events, and any rain events that occur within 6 hours of each other are grouped together and considered as a single event to account for the impacts of coupled rain events. Rain events are included on the reference job lists for these simulations if they have a minimum rain depth of 3 mm after this grouping is considered. The simulation of each event stops when the total system volume reaches 14,000 m^3, as all the basins can be considered empty at this stage. Using these settings, a simulation for one year of data takes approximately 3.5 full days to complete using MIKE URBAN on a dedicated calculation server. For 41 years of data, this means an effective time of

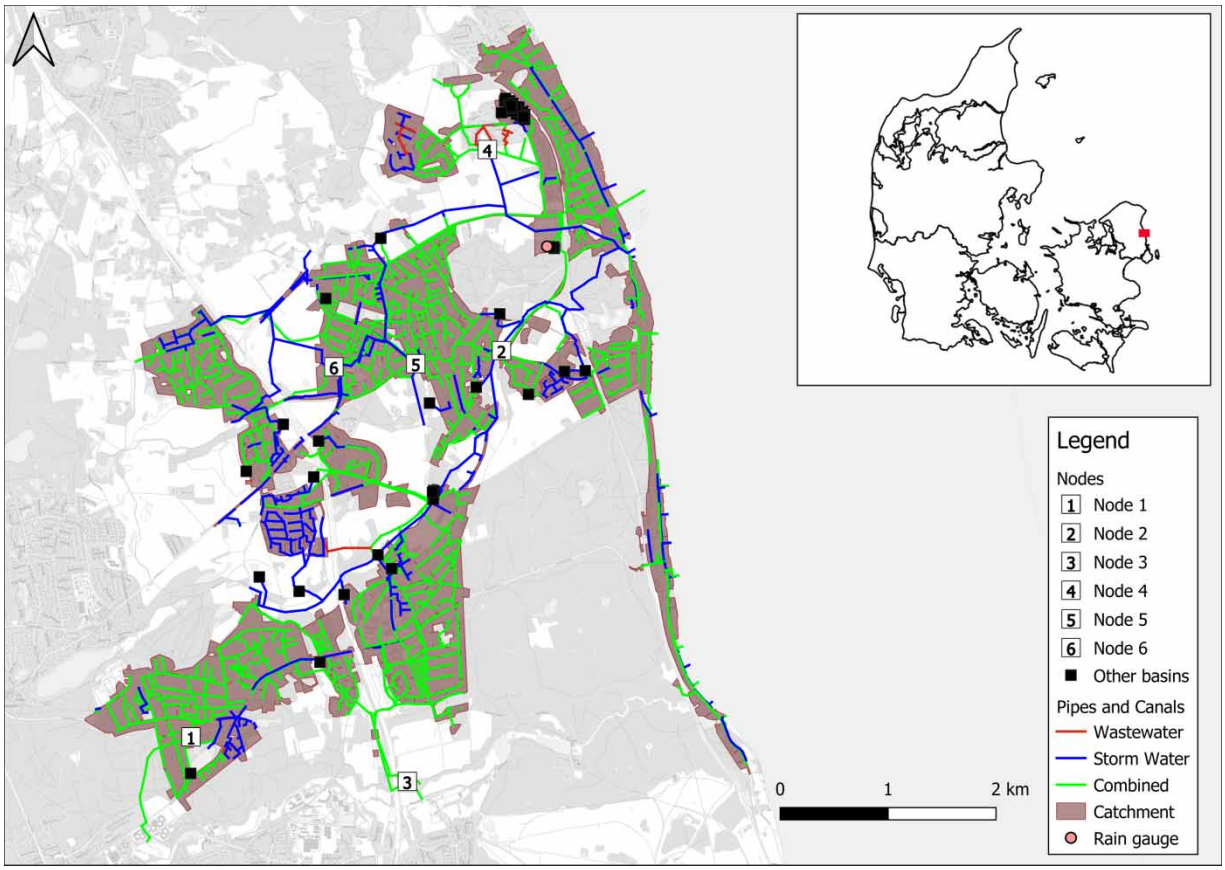

Figure 1 | Location of the used model area, rain gauge, and nodes of interest.

Table 1 | Summary of node characteristics

Node	Type	Volume [m^3]	Emptying time [hr.]
1	Basin	569	15
2	Basin	700	26
3	Basin	1185	16
4	Basin	60	7
5	Manhole	3	2
6	Basin	6640	84

calculation of approximately 140 days. In our case we were able to run these simulations in batches in parallel due to rich access to both computational resources and DHI licenses.

Predictive model development

Logistic regression models predict the probability of a data point to fall within a certain class or category based on the variables included in the model (Kassambara 2017; Dalpiaz 2021). In this way they are well-suited to qualitative binary outputs, such as whether a rain event is likely to cause CSO or not. Additionally, they are simple to implement, and the number of rain events one wants to include can be easily changed by changing the minimum probability that is used as the threshold. The generic logistic regression equation is shown in Equation (1):

$$p(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \ \ldots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ \ldots + \beta_n x_n)} \tag{1}$$

Table 2 | Rain and CSO data summary for the six nodes of interest in the system

Calibration	C1	C2	C3	C4	Average	Total
Period	1979–1983	1989–1993	1999–2003	2009–2013	–	–
Rain events per year	57.4	53.2	66.2	61.4	59.6	1191
CSO events per year	20.0	20.6	27.4	23.0	22.8	455
Test	**T1**	**T2**	**T3**	**T4**[a]	**Average**	**Total**
Period	1984–1988	1994–1998	2004–2008	2014–2019	–	–
Rain events per year	62.0	58.6	68.4	66.5	63.9	1,344
CSO events per year	19.6	24.8	25.8	24.4	23.7	473

[a]Please note that the V4 period is six years compared to all the other periods that are five years.

where $p(x)$ is a number between 0 and 1 where a higher value is associated with a higher probability of an event resulting in a CSO, β_0 is the intercept, β_1 is the parameter for the first model variable, x_1 is the data values for the first model variable, β_n is the parameter for the nth model variable, and x_n is the data values for the nth model variable (Kassambara 2017; Dalpiaz 2021).

As the units and value ranges for the different variables vary considerably, all the quantitative variables are log-transformed to improve the normality of the data before it is used in the model. Categorical variables are transformed into binary values, where the most influential category is equivalent to a value of 1. For instance, as summer rain events appeared to be more correlated to CSOs, a rain event in summer has a season variable value of 1, where one happening in spring, fall, or winter has a season value of 0.

Two logistic regression models are developed in this study where forwards-step and backwards-step algorithms are used in conjunction with the glm package in R (R Core Team 2020) to build and fit the models. Similar work has been done where rain variables are selected and used to fit a model that predicts CSO characteristics such as volume (Sandoval *et al.* 2013) and frequency (Thorndahl & Willems 2008; Mailhot *et al.* 2015; Zhao *et al.* 2017, among others). This work differs from those referenced in that two simple objective algorithms are used to select which variables are included in the predictive model. Selection via algorithm has the advantages of being faster and reducing variable selection bias as they are inherently objective with respect to the possible variable to include.

The forwards-step algorithm works by defining a model that only contains an intercept (also known as a null model). It then iteratively adds one variable at a time to the model, based on its significance and how it impacts the Akaike information criterion (AIC) (Akaike 1998) value of the model. The AIC value considers the fit and the number of parameters in the model, where a lower value suggests a better fit and less extraneous parameters. In the algorithm, parameters are iteratively added to the model if they reduce the AIC value. In each iteration, a parameter is added if it results in the largest reduction in AIC value from the null model or the model from the previous iteration. If the addition of a parameter would not change or would increase the AIC value, then none is added and this model is then determined to be the best iteration according to the algorithm.

The backwards-step algorithm works by starting with a model that contains all possible variables (also known as a full model). It then iteratively removes one variable at a time, based on the significance of its parameter value and how its removal would impact the AIC value of the model. This is similar to the process described for the forwards-step algorithm, but parameters are removed from the model based on how they can reduce the AIC, instead of being added. This process of removing parameters continues until removing another parameter would have a negligible impact on the AIC value. This model is determined to be the best iteration of the model according to the algorithm.

The variables considered for inclusion in the models are confined to those that are easily obtained directly from the measured rainfall data. These variables are summarized in Table 3.

Comparison metrics

The main success criterion for the models is how few rain events need to be included in the job list to ensure that all relevant events are included. The perfect model would only include rain events that cause CSOs, without excluding any CSO-causing rain events, where CSO-causing rain events are defined using the full LTS results. However, this is an unlikely outcome, and

Table 3 | Variables considered for inclusion in the logistic predictive models

Variable	Unit	Note
Total rain event depth	mm	To account for the volume of the rain event
1-minute rain intensity	μm/s	To account for peak flow levels in the system
5-minute rain intensity	μm/s	
10-minute rain intensity	μm/s	
30-minute rain intensity	μm/s	
1-hour rain intensity	μm/s	
3-hour rain intensity	μm/s	
6-hour rain intensity	μm/s	
12-hour rain intensity	μm/s	
24-hour rain intensity	μm/s	
48-hour rain intensity	μm/s	
Rain event rank	–	Ranked according to volume, with largest volume having rank 1 and so forth
Season of rain event	–	Some seasons statistically have more CSO events than others
Number of rain events in the season	–	A wet season with many events could have more CSO events as well
Accumulated rain depth 48-hours before the start of the rain event.	mm	A proxy of the amount of water already present in the sewer system at the beginning of the event

the goal is to therefore come as close as possible to this ideal, where there is a significant reduction in the number of required rain events for the model without excluding CSO-causing rain events.

The success of this goal is measured by comparing the length of the reference job list and occurrence of CSOs from the LTS simulations to the required number of jobs, included CSOs, and excluded CSOs for each model. If the number of required jobs for a model is close to the number in the reference jobs list, then the success criterion is not met, and logistic regression is not a meaningful alternative. The same is true if there are only few required jobs, but many excluded CSO events.

The sensitivity of the models is also evaluated by comparing the results each model fit using the whole calibration dataset, with four of the same models fit with only a portion of the dataset (C1–C4 referenced in Table 2). This is done to see how radically choosing a shorter dataset to fit the model would impact the results.

RESULTS AND DISCUSSION

Predictive model development

A list of the selected variables for both models is shown in Table 4, along with parameter estimates, standard error, *p*-values, and AIC. Both models follow the generic equation shown in Equation (1) and could have included all variables shown in Table 3. However, both algorithms result in the same final model with the same six parameters in addition to the intercept. Both algorithms result in a model with exclusively statistically significant parameters with *p*-values below 0.05, and are comprised of several different intensity variables, a duration variable, and two depth variables. The 6-hour intensity, the 24-hour intensity, and the event depth are the most influential parameters as they have the highest estimated values.

Table 4 | Estimated parameter values, standard error, *p*-values, and Akaike information criterion (AIC) scores for both models

Model	Parameter	Estimate	Standard Error	*p*-value	AIC
Both forwards and backwards model	β_0 (Intercept)	−8.5224	2.9038	3.34×10^{-3}	765.75
	β_1 (30-min. int.)	0.8434	0.2419	4.90×10^{-4}	
	β_2 (6-hour int.)	3.2624	0.5885	2.97×10^{-8}	
	β_3 (24-hour int.)	−3.1984	0.8015	6.60×10^{-5}	
	β_4 (Duration)	−1.0102	0.2001	4.44×10^{-7}	
	β_5 (Depth)	5.6746	0.5289	$<2\times10^{-16}$	
	β_6 (Acc. depth)	−0.6855	0.138	6.80×10^{-7}	

It was unexpected that both algorithms produced the same resulting models, but it is evidence that exactly these parameters result in the strongest model and none of the other parameters could act as viable proxies. As both algorithms resulted in the same final model, only one is further investigated and we simply call it the model from hereon.

Performance on calibration data

In order to determine the optimal threshold value for this model, several different values were investigated along with their respective percentage of total jobs and CSOs captured, as shown in Figure 2 for the calibration data. A lower threshold results in a higher CSO capture rate, but also results in more jobs. Conversely, a high threshold has a lower CSO capture rate, but fewer extraneous jobs. The optimal threshold value should therefore minimize the included number of jobs without excluding any CSOs. A potential optimum threshold value can be the threshold value when the percentage of job reductions and the percentage of CSOs captured intersect. From Figure 2, this optimum is shown to be approximately 0.65. However, one could also focus on the performance of CSO prediction, as Figure 2 shows that a 90% performance can be achieved with a reduction in jobs of more than 50% with a threshold value of approximately 0.4. In our case this would save approximately 70 days of efficient model computations.

Figure 2 also shows the sensitivity of the model when fit with the different five-year sub-periods of the calibration dataset instead of the complete set originally used. This is done to see how using a shorter period would impact the results. There is some visible variation between the main model and the sub-models, but the overall tendency is clearly the same; the models fitted to any of the short periods result in virtually the same model, even though dependence of singular CSO causing events do become more pronounced, when shorter periods are used for calibration of the model.

Figure 3 shows the capture of CSOs and the overall job reductions for the six nodes of interest. The job reduction is the same for all six nodes, because the jobs are collectively selected and applied to all nodes as opposed to being applied differently to different nodes. When looking at the results for the individual nodes, it is clear that there is a lack of agreement on an optimal threshold value. Instead of the CSO-capture rates converging at a similar threshold, they form three distinct groups. One group, comprising nodes 1, 2 and 4, have a high threshold close to 0.9. Another group, comprising of nodes 3 and 5, have a threshold of

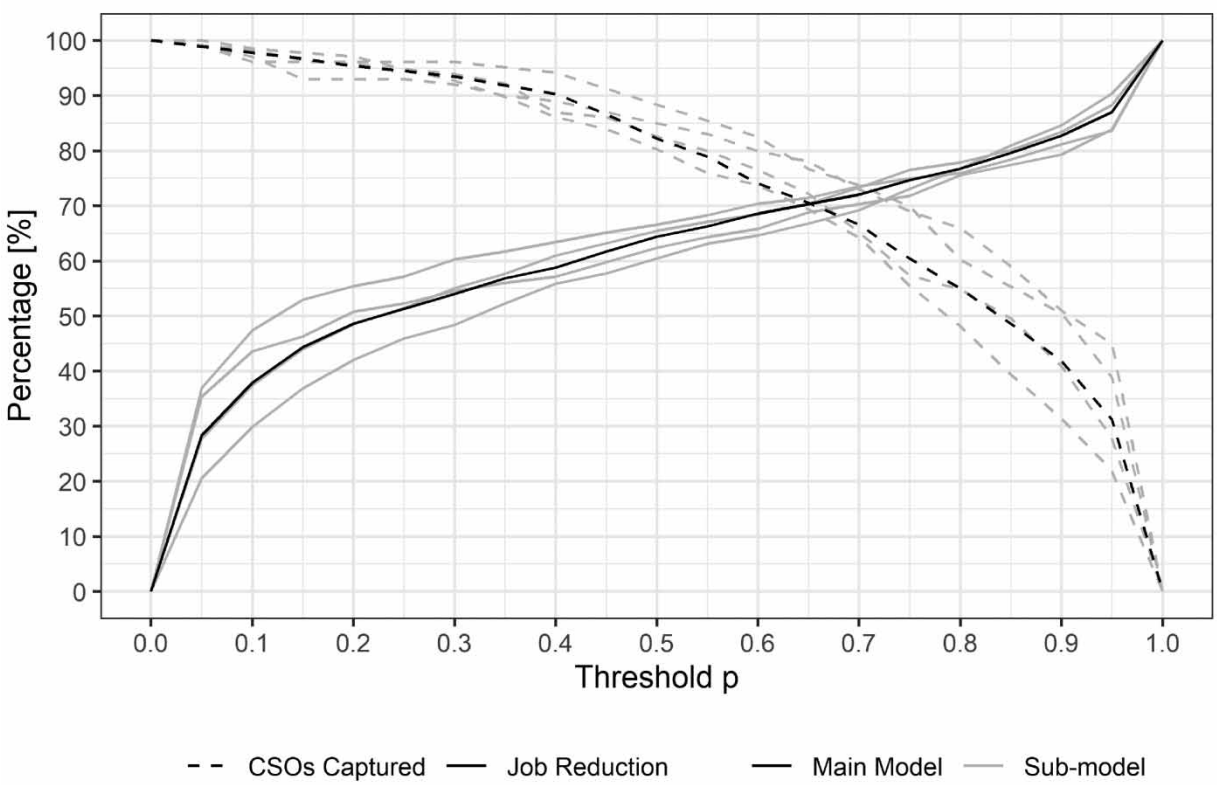

Figure 2 | Percentage of CSOs captured and job reduction for various threshold values for the main forward-model fit to the full calibration period, and forward-models fit to sub-periods C1, C2, C3, and C4.

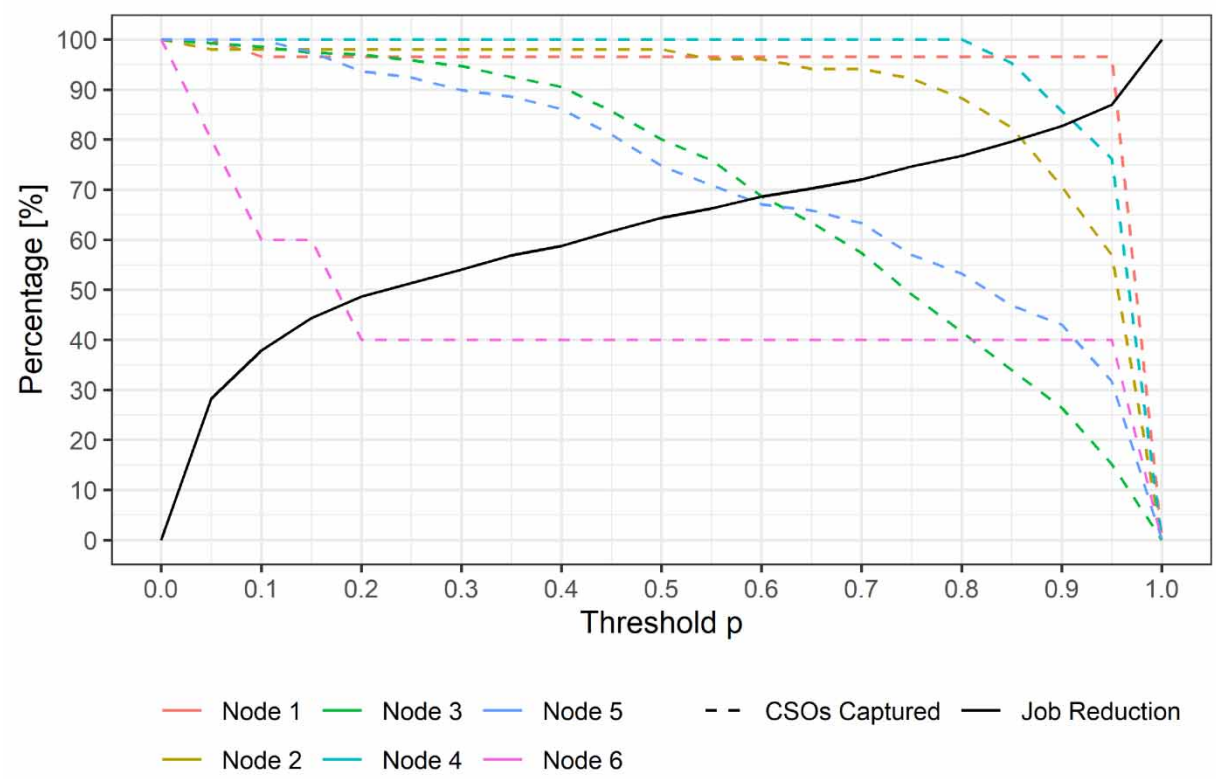

Figure 3 | Percentage of CSOs captured and job reduction for various threshold values for the nodes of interest using the full model and the calibration period.

approximately 0.6. Lastly, node 6 has a low threshold close to 0.2. It appears that the number of CSO events at each node influences the model results more than the storage volume and emptying time, as there is a more apparent correlation between the model results and the number of CSO events for each node. For instance, nodes 1, 2 and 4 have a moderate number of CSO events per year, which range from approximately one to 2.5 at each node in total. Conversely, nodes 3 and 5 have higher numbers of CSO events, which are respectively 13 and four per year. Meanwhile, node 6 has a very low number of CSO events, with only one every second year. This indicates that the model will be more effective for nodes with a moderate level of CSO-events, as a higher threshold is more effective, which means there can be a significant job reduction, while excluding few CSO events. For instance, with a threshold of approximately 0.85, there can be an 80% reduction in the number of jobs at node 2, while only excluding 20% of the CSOs. Conversely, the model is not as effective for nodes with a high number of CSO events, as a 65% reduction in jobs using a threshold of approximately 0.6 would result in excluding approximately 35% of CSO events. Finally, the results are similar for nodes with very few CSO events, such as node 6, as using a threshold of approximately 0.2 would reduce the number of jobs by 50%, but it would also exclude 50% of the CSOs.

Performance on test data

The different threshold values are also tested for the test data as shown in Figure 4. Similar to the results for the calibration data, reducing the threshold value results in a higher capture rate of the CSO events. A threshold value of approximately 0.65 is again found to be the optimum value for balancing the reducing of jobs with the capture of CSO events. The results are generally similar to those for the calibration period, with a CSO-capture rate of 70%.

The amount of captured CSO events for each node is shown in Figure 5 and are mostly similar in the test period as in the calibration period. Nodes 1, 2 and 4 again have high optimal thresholds and moderate numbers of CSO events, with a value range of 1.5–2.5 per year. Nodes 3 and 5 also have test results similar to their calibration results, with optimal thresholds close to 0.6 and high numbers of CSO events, approximately 22.5 and five per year. However, in the test data node 6 is also grouped close to nodes 1, 2 and 4 with an optimal threshold of approximately 0.8 and approximately 0.7 CSO events per year. This number is slightly higher than in the calibration period, but the events are apparently much better captured here. It is

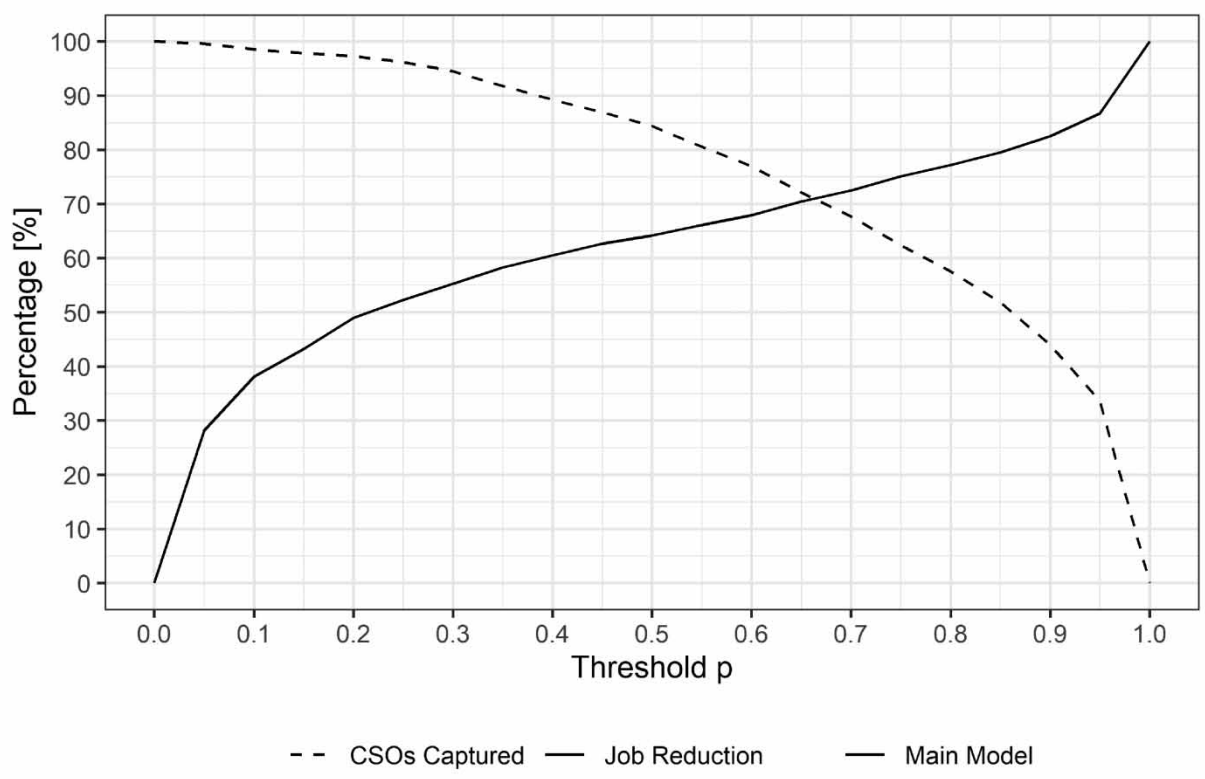

Figure 4 | Percentage of CSOs captured and job reduction for various threshold values for the full model run with the full test data.

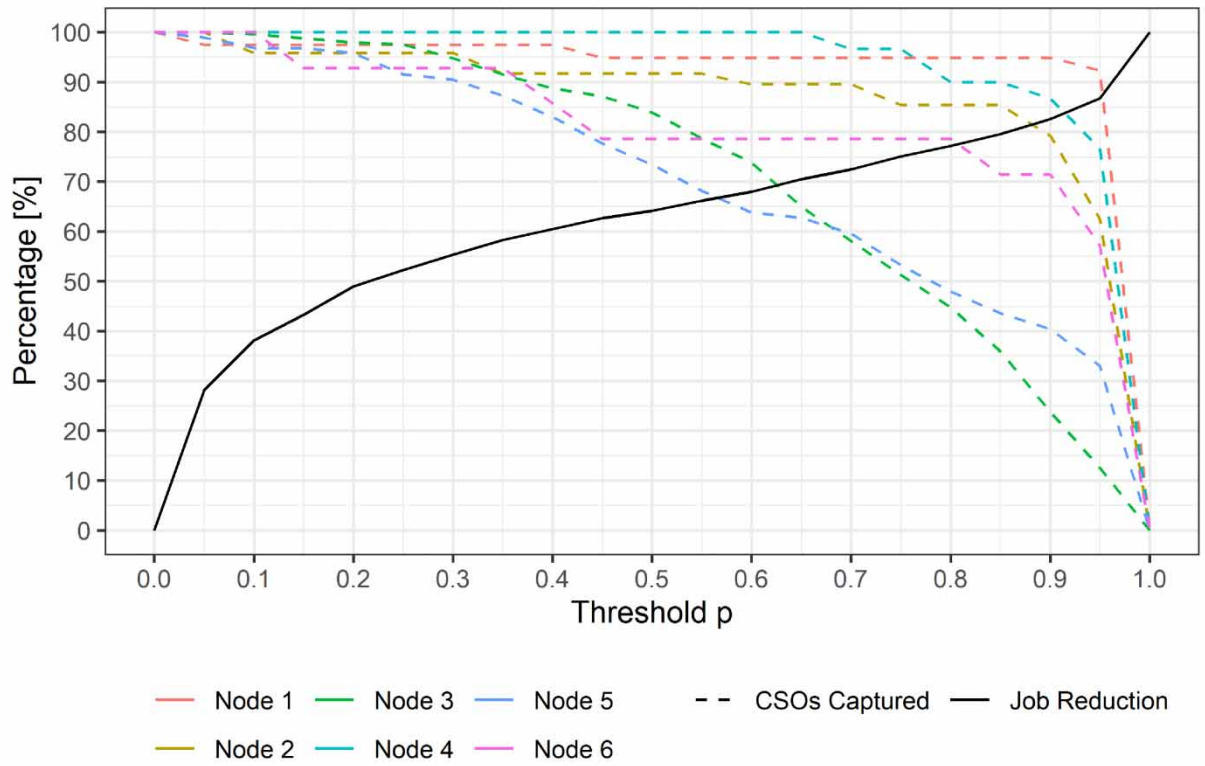

Figure 5 | Percentage of CSOs captured and job reduction for various threshold values for the nodes of interest using the full model and the full test period.

surprising that the performance at the node can change so drastically, although it confirms the strong impact that individual events have on the model results.

In terms of the overall percent of CSO events captured, the model performs relatively well for both the calibration and test datasets, as both have similar optimal thresholds of approximately 0.65, where approximately 70% of the CSOs for all of the nodes are captured. The similar performance for the test and calibration data sets, and the fit to shorter periods of the calibration data set, indicate that it should be possible to fit these models to a smaller amount of data and then successfully use them to make a job list for a much longer period of data. The events not captured by the models tend to be smaller, which suggests that the resulting job lists that exclude them would still be adequate for uses such as dimensioning basins. Additionally, this number of missed CSOs can be estimated from the capture percentage of the calibration data set.

It was expected that a clear trend in model performance would be seen based on the characteristics of the different nodes. Though the number of CSO events at a node seem to impact the model performance, such patterns related to basin volume or emptying time are not obvious. The results clearly show that it is different rainfall events with different characteristics that lead to CSOs at the different nodes in the system, and that the events that appear very often or very seldom are the ones less likely to be captured by the model. This suggests that stronger models could be fitted to either individual nodes or, possibly, groups of nodes with similar behavior. Future research could consider this as well as the impacts of site-specific characteristics. One way of doing that is the inclusion of static node specific data in the models, such as emptying time, volume, and catchment area in addition to the rainfall related data. This could result in much more tailored job lists, which specifies relevant events to analyze for specific nodes or nodes with specific characteristics, but also much shorter job lists.

CONCLUSIONS

The use of a logistic regression model with this methodology can be used to greatly decrease the number of jobs needed for an LTS simulation without losing much information about the system behavior. As these simulations are very computational demanding, this is a major benefit when analyzing the CSO behavior. A minimal loss of information is expected, as capturing all of the CSO-causing events for all nodes of interest does not seem feasible, while also reducing the number of jobs. It is possible to capture a high percentage of CSO events while reducing the number of jobs for nodes with a moderate amount (approximately 1–2.5 per year) of CSOs. This job reduction can also be increased if only one node is of interest. The ability to easily define and re-define the threshold value allows for control over the number of jobs to be included. Therefore, this methodology makes it possible to balance the number of jobs with the desired level of precision of the results. In all cases, it is possible to greatly decrease the number of jobs that need to be run.

ACKNOWLEDGEMENTS

This research was supported by the Foundation for Development of Technology in the Danish Water Sector, Project: VÆRDI, Grant/Award Number:57.2019. The dataset used is a product of The Water Pollution Committee of The Society of Danish Engineers made freely available for research purposes. Access to data is governed by the Danish Meteorological Institute, and they should be contacted for enquiries regarding data access. The MIKE URBAN model used for the LTS simulation was made available for this project by the utility company Novafos.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Akaike, H. 1998 Information theory and an extension of the maximum likelihood principle. In: (Akaike, H., Parzen, E., Tanabe, K. & Kitagawa, G., eds) *Selected Papers of Hirotugu Akaike*. Springer, New York, pp. 199–213.

Andersen, J. S., Lerer, S. M., Backhaus, A., Jensen, M. B. & Sørup, H. 2017 Characteristic rain events: a methodologly for improving the amenity value of stormwater control measures. *Sustainability* **9** (10), 1793.

Bachmann-Machnik, A., Brüning, Y., Bakhshipour, A. E., Krauss, M. & Dittmer, U. 2021 Evaluation of combined sewer system operation strategies based on highly resolved online data. *Water* **13** (6), 751.

Botturi, A., Ozbayram, E. G., Tondera, K., Gilbert, N. I., Rouault, P., Caradot, N., Gutierrez, O., Daneshgar, S., Frison, N., Akyol, Ç., Foglia, A., Eusebi, A. L. & Fatone, F. 2021 Combined sewer overflows: a critical review on best practice and innovative solutions to mitigate impacts on environment and human health. *Critical Reviews in Environmental Science and Technology* **51** (15), 1585–1618.

Cristiano, E., ten Veldhuis, M.-C., Gaitan, S., Ochoa Rodriguez, S. & van de Giesen, N. 2018 Critical scales to explain urban hydrological response: an application in Cranbrook, London. *Hydrology and Earth System Sciences* **22** (4), 2425–2447.

Dalpiaz, D. 2021 *Applied Statistics with R*. GitHub repository. Available from: https://daviddalpiaz.github.io/appliedstats/

Davidsen, S., Lowe, R., Ravn, N. H., Jensen, L. N. & Arnbjerg-Nielsen, K. 2017 Initial conditions of urband permeable surfaces in rainfall-runoff models using Horton's infiltration. *Water Science & Technology* **77** (3), 662–669.

DHI 2020 *MIKE URBAN*. Hørsholm, Denmark. Available from: https://www.mikepoweredbydhi

El Ghazouli, K., El Khatabi, J., Soulhi, A. & Shahrour, I. 2002 Model predictive control based on artificial intelligence and EPA-SWMM model to reduce CSOs impacts in sewer systems. *Water Science & Technology* **85** (1), 398–408.

Fontanazza, C. M., Freni, G., La Loggia, G. & Notaro, V. 2011 Uncertainty evaluation of design rainfall for urbaan flood risk analysis. *Water Science & Technology* **63** (11), 2641–2650.

Gregersen, I. B., Madsen, H., Rosbjerg, D. & Arnbjerg-Nielsen, K. 2015 Long term variations of extreme rainfall in Denmark and southern Sweden. *Climate Dynamics* **44**, 3155–3169.

Jean, M.-È., Duchesne, S., Pelletier, G. & Pleau, M. 2018 Selection of rainfall information as input data for the design of combined sewer overflow solutions. *Journal of Hydrology* **565**, 559–569.

Kassambara, A. 2017 *Machine Learning Essentials: Practical Guide in R*. STHDA (available online).

Keifer, C. J. & Chu, H. H. 1957 Synthetic storm pattern for drainage design. *Journal of the Hydraulics Division* **83** (4), 1332.

Locatelli, L., Gabriel, S., Mark, O., Mikkelsen, P., Arnbjerg-Nielsen, K., Taylor, H., Bockhorn, B., Larsen, H., Kjølby, M. J., Blicher, A. S. & Binning, P. J. 2015 Modelling the impact of retention–detention units on sewer surcharge and peak and annual runoff reduction. *Water Science & Technology* **71** (6), 898–903.

Löwe, R., Böhm, J., Jensen, D. G., Leandro, J. & Rasmussen, S. H. 2021 U-FLOOD – topographic deep learning for predicting urban pluvial flood water depth. *Journal of Hydrology* **603**, 126898.

Mailhot, A., Talbot, G. & Lavallee, B. 2015 Relationships between rainfall and Combined Sewer Overflow (CSO) occurrences. *Journal of Hydrology* **523**, 602–609.

McGrath, G., Kaeseberg, T., Reyes-Silva, J. D., Jawitz, J. W., Blumensaat, F., Borchardt, D., Mellander, P.-E., Paik, K., Krebs, P., Suresh, P. & Rao, C. 2019 Network topology and rainfall controls on the variability of combined sewer overflows and loads. *Water Resources Research* **55** (11), 9578–9591.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P. & ten Veldhuis, M.-C. 2015 Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: a multi-catchment investigation. *Journal of Hydrology* **531** (2), 389–407.

Rathnayake, U. & Faisal Anwar, A. 2019 Dynamic control of urban sewer systems to reduce combined sewer overflows and their adverse impacts. *Journal of Hydrology* **579**, 124150.

R Core Team 2020 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL Available at: www.R-project.org/.

Sandoval, S., Torres, A., Rawlowsky-Reusing, E., Riechel, M. & Caradot, N. 2013 The evaluation of rainfall influence on combined sewer overflows characteristics: the Berlin case study. *Water Science & Technology* **68** (12), 2683–2690.

Schaarup-Jensen, K., Rasmussen, M. R. & Thorndahl, S. 2009 To what extent does variability of historical rainfall series influence extreme event statistics of sewer system surcharge and overflows? *Water Science & Technology* **60** (1), 87–95.

Sørup, H. & Lerer, S. 2021 Principles for distributing infiltration-based stormwater control measures in series. (M. Scholz, Ed.). *Water* **13** (8), 1029.

Sørup, H., Davidsen, S., Lowe, R., Thorndahl, S., Borup, M. & Arnbjerg-Nielsen, K. 2018 Evaluating catchment response to artificial rainfall from four weather generators for present and future climate. *Water Science & Technology* **77** (11), 2578–2588.

Sørup, H., Fryd, O., Liu, L., Arnbjerg-Nielsen, K. & Jensen, M. 2019 An SDG-based framework for assessing urban stormwater management systems. *Blue-Green Systems* **1** (1), 102–118.

Thorndahl, S. 2009 Stochastic long term modelling of a drainage system with estimation of return period uncertainty. *Water Science & Technology* **59** (12), 2331–2339.

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.

Thrysøe, C., Arnbjerg-Nielsen, K. & Borup, M. 2019 Identifying fit-for-purpose lumped surrogate models for large urban drainage systems using GLUE. *Journal of Hydrology* **568**, 517–533.

van Bijnen, M., Korving, H. & Clemens, F. 2012 Impact of sewer condition on urban flooding: an uncertainty analysis based on field observations and Monte Carlo simulations on full hydrodynamic models. *Water Science & Technology* **65** (12), 2219–2227.

van Uytven, E., Wampers, E., Wolfs, V. & Willems, P. 2020 Evaluation of change factor-based statistical downscaling methods for impact analysis in urban hydrology. *Urban Water Journal* **17** (9), 785–794.

Vorobevskii, I., Janabi, F. A., Schneebeck, F., Bellera, J. & Krebs, P. 2020 Urban floods: linking the overloading of a storm water sewer system to precipitation parameters. *Hydrology* **7** (2), 35.

Zhao, W., Beach, T. H. & Rezgui, Y. 2017 Automated model construction for combined sewer overflow prediction based on efficient LASSO algorithm. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **49** (6), 1254–1269.

First received 23 September 2021; accepted in revised form 11 February 2022. Available online 24 February 2022