



# Evaluating the Thiessen polygon approach for efficient parameterization of urban stormwater models

Zhaokai Dong<sup>1</sup> · Daniel J. Bain<sup>2</sup> · Murat Akcakaya<sup>3</sup> · Carla A. Ng<sup>1,4</sup> 

Received: 18 July 2022 / Accepted: 8 November 2022 / Published online: 25 November 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Catchment discretization plays a key role in constructing stormwater models. Traditional methods usually require aerial or topographic data to manually partition the catchment, but this approach is challenging in areas with poor data access. Here, we propose an alternative approach, by drawing Thiessen polygons around sewer nodes to construct a sewershed model. The utility of this approach is evaluated using the EPA's Storm Water Management Model (SWMM) to simulate pipe flow in a sewershed in the City of Pittsburgh. Parameter sensitivities and model uncertainties were explored via Monte Carlo simulations and a simple algorithm applied to calibrate the model. The calibrated model could reliably simulate pipe flow, with a Nash–Sutcliffe efficiency (NSE) of 0.82 when compared to measured flow. The potential influence of sewer data availability on model performance was tested as a function of the number of nodes used to build the model. No statistical differences were observed in model performance when randomly reducing the number of nodes used to build the model (up to 40%). Based on our analyses, the Thiessen polygon approach can be used to construct urban stormwater models and generate good pipe flow simulations even for sewer data limited scenarios.

**Keywords** Stormwater · SWMM · Discretization · Calibration · Monte Carlo simulations · Uncertainty

## Introduction

Urbanization dramatically changes the landscape by increasing impervious surfaces (Arnold and Gibbons 1996). Higher impervious cover reduces infiltration, increases the velocity and volume of stormwater runoff, and accelerates the transport of nutrients and contaminants (Arnold and Gibbons 1996; Lee et al. 2012; Schueler et al. 2009; Li et al. 2009). Urban stormwater is one of the leading sources of water pollution in the USA (National Research Council 2009). Studies

evaluating stormwater impact mitigation have pursued innovative stormwater management methods, in particular green infrastructure (GI), which aims to protect, mimic and maintain natural hydrologic conditions by using decentralized retrofit practices (Coffman et al. 1999; Ahiablame et al. 2012).

To evaluate the effectiveness of urban stormwater mitigation methods a number of modeling tools (Eckart et al. 2017) have been developed. Among these, the US Environmental Protection Agency (EPA) Storm Water Management Model (SWMM) has been widely applied for urban rainfall–runoff simulations (Eckart et al. 2017; Tsai et al. 2017). In recent years, SWMM has been used to study the hydrologic impacts of urbanization (Jang et al. 2007), to model the effects of green and grey infrastructure practices to reduce combined sewer overflows (CSO) (Alves et al. 2016), to investigate the impacts of GI on hydrological responses in urban areas (Palla and Gnecco 2015; Aad et al. 2010) and to simulate chemical transport (McCall et al. 2017).

Despite these practical applications, improved efficiency in model parameterization while maintaining model performance remains a critical need, particularly for large urban systems and areas with poor data availability. In recent years, high-resolution data (e.g., aerial images) have

Responsible Editor: Marcus Schulz

✉ Carla A. Ng  
carla.ng@pitt.edu

<sup>1</sup> Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA

<sup>2</sup> Department of Geology and Environmental Science, University of Pittsburgh, Pittsburgh, PA 15261, USA

<sup>3</sup> Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA

<sup>4</sup> Department of Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

been used to identify and describe detailed surface information in SWMM simulations (Krebs et al. 2013; 2014; Sun et al. 2013), such as buildings and roads, which can help to build a model that closely reflects reality. This is especially important in urban systems where the details of building topology and road edges may influence the routing of surface flows to storm sewers in ways not captured by lower resolution topography. However, delineation of subcatchments based on high-resolution data and sewer maps is time-intensive for large-scale simulations (beyond a few city blocks). In addition, some areas may lack high-quality topographical data. Information about the distributions and geometries of sewer networks may also be unavailable or incomplete. Missing data makes describing the relationship between urban watersheds and sewer systems challenging and data needs make traditional approaches less efficient for building sewershed-scale models and specifying corresponding parameters.

More efficient discretization methods exist, but have specific shortcomings when used in urban areas. For example, hydrologic partitioning based on digital elevation model (DEM) (Tsai et al. 2017) preserves terrain information, but neglects the surface connections to subsurface drainage networks which are crucial for sewer flow hydraulic simulations. On the other hand, pipe network-guided division (Huang and Jin 2019) provides the links between surface and sewer nodes, but the surface flow routing is uncertain because it does not consider the influence of topography. The limits of these easily applied approaches for flow routing between subcatchments and in-pipe conveyance need to be understood and evaluated in urban areas.

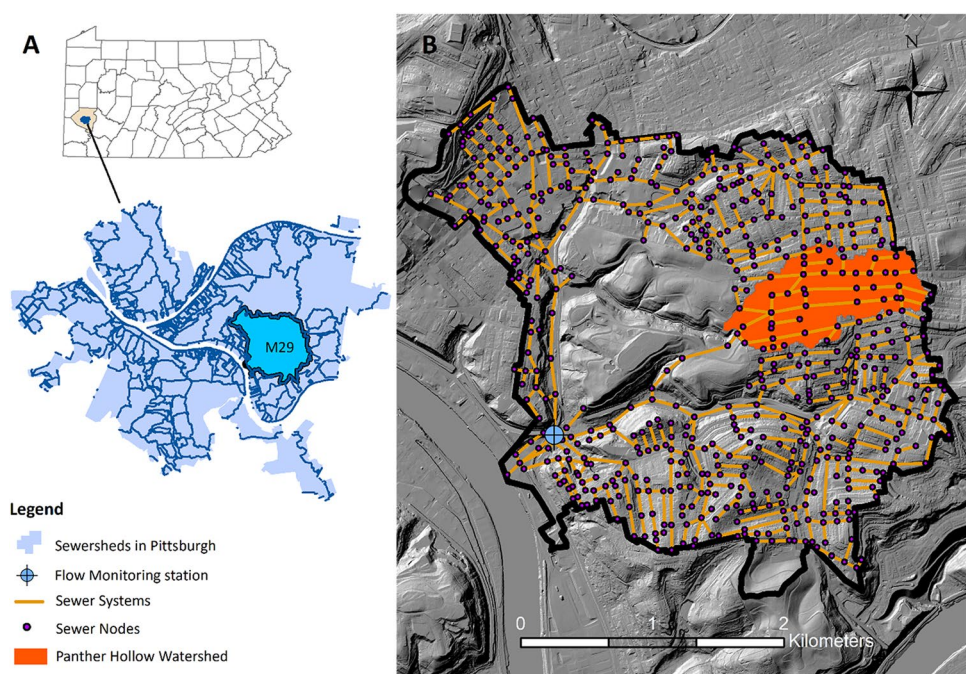
This study tests the utility of using the Thiessen polygon approach (Okabe et al. 2000) to efficiently partition an urban sewershed. We built a sewershed model in the City of Pittsburgh, a city vulnerable to flash floods and combined sewer overflows (Hopkins et al. 2014). The Thiessen polygon approach was used to generate subcatchments. To connect subcatchments to the sewer system a simplifying assumption was made that all surface water within the subcatchment drains to the corresponding node. Parameter distributions were constructed based on the frequencies of initial estimates for each parameter across all the model elements. Further, parameter sensitivities and model uncertainties were analyzed via Monte Carlo simulations and a simple model calibration algorithm combined with generalized likelihood uncertainty estimation was applied to calibrate the model. Based on these analyses, we evaluated the capability of proposed approaches for urban stormwater simulations, which might enable stormwater model development even in areas that lack data.

## Materials and methods

### Study area and data sources

Sewershed M29 in the city of Pittsburgh (Fig. 1) was modeled with a drainage area of approximately 9.7 km<sup>2</sup>. Land-cover in this sewershed varies from highly urbanized to parkland, and the terrain is rugged. In addition, M29 has a large sewer network with over 600 nodes and pipe segments. These characteristics provide the opportunity to investigate

**Fig. 1** Location and pipe network of sewershed M29 in the city of Pittsburgh. **A** Sewersheds in Allegheny County (Allegheny County Division of Computer Services Geographic Information Systems Group 2016) (PA, USA). **B** The sewer network in M29. Hill shade derived from 1.8 m LiDAR DEM. (Allegheny County Division of Computer Services Geographic Information Systems Group 2017)



the relationship among various surface physical properties, the sewer networks, and model responses.

The model was built using multiple spatial data (Table 1). The main input data included: (1) 15-min rainfall intensity data (3 Rivers Wet Weather (3RWW) 2008) obtained from the University of Pittsburgh rain gauge (April 2008); (2) sewer data digitized on the basis of a system-wide sewer map (Allegheny County Sanitary Authority (ALCOSAN) 2022) and then corrected and validated by comparing with GIS data provided by the Allegheny County Sanitary Authority (ALCOSAN); (3) 30 m land cover data (US Geological Survey 2014); (4) 1.8-m Lidar digital elevation model (DEM) (Allegheny County Division of Computer Services Geographic Information Systems Group 2017). We evaluated and calibrated the model by comparing simulation results with monitored pipe flow (April 2008, provided by ALCOSAN), which included both dry and wet weather conditions.

To capture the input sewage base flow during dry weather conditions, we randomly picked two days of pipe flow data from each month with no rainfall records. We then averaged these flows to a daily flow pattern to serve as the sewage base flow (Supplemental Information, SI Fig. S1).

### SWMM governing equations

The general simulation in SWMM is of a system that first receives precipitation and then generates surface runoff within subcatchments. At the outlet of the subcatchment, the pipe node or the channel receives the surface water and routes this runoff as pipe flow through the drainage system. SWMM surface runoff is mainly governed by precipitation, infiltration, and evaporation, modeled as a nonlinear reservoir to derive the governing equation (Rossman and Huber, 2016) (Eq. 1).

$$\frac{\partial d}{\partial t} = i - e - f - \frac{kW}{An} S^{\frac{1}{2}} (d - ds)^{\frac{5}{3}} \quad (1)$$

where  $\frac{\partial d}{\partial t}$  is the change in water depth  $d$  over time  $t$ ;  $i$  is the precipitation rate;  $e$  is the evaporation rate;  $f$  is the infiltration rate;  $k$  is the unit conversion constant (1 for SI units and 1.49 for English units);  $W$  is the width of the subcatchment;  $A$  is the surface area of subcatchment;  $n$  is the Manning's roughness coefficient;  $S$  is the slope;  $ds$  is the depression storage.

The pipe flow in SWMM is derived from the St. Venant conservation of mass and momentum equations in Eq. (2) (Rossman 2017).

$$\frac{\partial Q}{\partial t} = 2v \frac{\partial A}{\partial t} + v^2 \frac{\partial A}{\partial x} - gA \frac{\partial H}{\partial x} - g \frac{Q^2}{\frac{k^2}{n^2} AR^{\frac{4}{3}}} \quad (2)$$

where  $\frac{\partial Q}{\partial t}$  is the change of flow rate per unit of time;  $v$  is the velocity;  $g$  is the gravitational acceleration;  $H$  is the hydraulic head;  $Q$  is the flow rate;  $R$  is the hydraulic radius. Note that in Eq. (2)  $A$  is the flow cross-section area rather than the surface area of the subcatchment.

### Model delineation and parameterization

In sewershed (M29), stormwater drains, described as nodes hereafter, are distributed densely throughout the developed areas. To simplify the model, we merged sets of adjacent nodes (within 5 m) into a single node. Corresponding pipe segments were then merged as well. The simplified sewer network consists of 638 nodes and 636 pipes. Thiessen polygons were used to partition the subcatchments in ArcMap (Esri Inc. 2016), based on the simplified sewer nodes, with each plane enclosing only one node (Fig. S2). This method is widely used to analyze precipitation data from unevenly distributed rain gauges, by finding points within a plane that are closest to the target point (Brassel and Reif 2010). In our case, the target point is a node (storm drain). To route surface flow, we assumed that surface water generated within a subcatchment drains to the corresponding node, selecting the option in SWMM that routes all surface water to outlets (here representing the pipe nodes).

The width of each subcatchment polygon was estimated by dividing the subcatchment area by the longest distance from any point in the subcatchment (polygon) to the node. To estimate imperviousness, land types were extracted from land use data, and each type was assigned imperviousness values based on literature (Rossman and Huber 2016; US EPA 2014). Then, an area-weighted average was used to estimate the imperviousness for each subcatchment. Similar strategies were used to estimate depression storage and Manning's roughness of subcatchments followed by assigning literature values (e.g., Manning's roughness (American Society of Civil Engineers and Water Environment Federation 2018; Bizier 2007) and depression storage (American

**Table 1** Data sources

| Data       | Resolution    | Sources                                       | Year |
|------------|---------------|-----------------------------------------------|------|
| Land cover | 30 m × 30 m   | United States Geological Survey (USGS)        | 2011 |
| DEM        | 1.8 m × 1.8 m | Allegheny County Lidar and Terrain Products   | 2017 |
| Sewer data | N.A           | Allegheny County Sanitary Authority (ALCOSAN) | 2022 |

Society of Civil Engineers and Water Environment Federation 2018)) associated with land uses.

Evapotranspiration was not included in this model, as fast flow storm periods were the primary interest in the study. Slope maps were derived from DEM, and the average slope was assigned to each subcatchment. Infiltration-related parameters were based on Horton's method (Rossman and Huber 2016; Horton 1941) as built into SWMM.

Based on the sewer data from ALCOSAN, pipe diameters in M29 range from 0.24 m up to 4.3 m. Simplifying the sewer network by combining nodes and pipes would result in changes to actual pipe geometric dimensions. We assumed circular cross-sections and an initial pipe diameter of 2.4 m, then created distributions from which diameters could be drawn for the Monte Carlo sampling process described in the next section.

### Sensitivity analysis

Monte Carlo-based sensitivity analysis (Saltelli et al. 2000) was conducted to identify the influence of model inputs on predictions and to create simulations for model calibration discussed in the next sections. We focused on impacts of parameters on two types of model results: (1) single model outputs (within subcatchments or pipes), e.g., surface peak flow, runoff volume and maximum pipe flow; (2) the overall goodness-of-fit between simulated pipe flow and monitored pipe flow near the outlet of the sewershed. To quantify the influence of model parameters on single

model outputs, we calculated the Spearman correlation coefficient (Spearman 1904) between model outputs — surface runoff peak flow, total runoff volume and maximum pipe flow — and parameters for each subcatchment or pipe. Goodness-of-fit is indicated by the Nash–Sutcliffe efficiency (NSE) in Eq. (3) (McCuen et al. 2006).

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^n (Q_{o,i} - \overline{Q_o})^2} \quad (3)$$

here  $Q_{o,i}$  and  $Q_{s,i}$  are the observed and simulated flow discharge values, respectively, and  $\overline{Q_o}$  is the observed mean flow. An NSE value closer to 1 suggests a better prediction was achieved.

The distributions used for the input parameters related to surface runoff and pipe flow are listed in Table 2. To create the distributions for each parameter, two strategies were used. First, lognormal distributions were used for the parameters extracted and estimated from spatial data and those estimated using literature values across the subcatchments. The mean was set to the estimated parameter values based on the area-weighted average. The variance was calculated using the estimated parameter values of all the subcatchments or pipes. For example, for the imperviousness, the mean is the weighted mean imperviousness, and the variance is calculated based on the mean imperviousness values from the 638 subcatchments. Therefore, in each subcatchment, the distribution of a specific parameter will have a different mean but the same variance.

**Table 2** Distributions of parameters included in sensitivity analysis for sewershed M29

| Parameter; units      |                                         | Distribution | (a, b)                                 |
|-----------------------|-----------------------------------------|--------------|----------------------------------------|
| Subcatchment          | Width; m                                | Uniform      | (0.5×Estimation, 1.5×Estimation)       |
|                       | Imperviousness (Imper); %               | Lognormal    | (Individual subcatchment mean, 0.24)   |
|                       | Slope; %                                | Lognormal    | (Individual subcatchment mean, 0.51)   |
|                       | Impervious N (N-imp); [-]               | Lognormal    | (Individual subcatchment mean, 0.559)  |
|                       | Pervious N (N-perv); [-]                | Lognormal    | (Individual subcatchment mean, 0.0867) |
|                       | Impervious D (D-imper); in              | Uniform      | (0.05, 0.1)                            |
|                       | Pervious D (D-perv); in                 | Lognormal    | (Individual subcatchment mean, 0.369)  |
| Infiltration (Horton) | Maximum infiltration rate (MaxIR); in/h | Uniform      | (4, 8)                                 |
|                       | Minimum infiltration rate (MinIR); in/h | Uniform      | (0.1, 0.3)                             |
|                       | Decay rate (Decay); 1/h                 | Uniform      | (2, 7)                                 |
| Sewer system          | Pipe length; m                          | Lognormal    | (Individual pipe mean, 0.04)           |
|                       | Roughness;                              | Uniform      | (0.011, 0.015)                         |
|                       | Diameter; m                             | Lognormal    | (2.4, 0.47)                            |

(a, b) for lognormal distribution: a — mean of lognormal distribution (the subcatchment mean or pipe mean), b — standard deviation of corresponding transformed normal distribution; (a, b) for uniform distribution: a — minimum, b — maximum. As parameter values vary by subcatchment, the mean of the distribution is not reported

The second strategy was to use uniform distributions for the parameters that were constant across subcatchments or for which no information on variability was available. The two bounds of the uniform distributions were assumed by considering typical literature values (Rossman and Huber 2016; Horton 1941; American Society of Civil Engineers and Water Environment Federation 2018; Bizier 2007). Once distributions were established for all relevant parameters, 1000 Monte Carlo simulations were run using Latin hypercube sampling (LHS) (McKay et al. 2000) to evaluate the relationship between the model parameters and predicted outputs.

To understand how model discretization and parameterization influence overall simulation results, rather than assess their influence within each subcatchment or pipe segment, we chose the 5% best and 5% worst simulations (highest NSE and lowest NSE, respectively) to evaluate the change in parameter values that most improved overall model performance (NSE). As the model became more accurate (i.e., higher NSE values) the corresponding parameters selected across the LHS-based Monte Carlo samples were assumed to represent better estimates. In addition, those parameters changing the most when comparing the worst performance to best performance were of particular interest (e.g., they may have had particularly poor initial settings).

We calculated the difference (Eq. 4) in parameter values between the best and worst simulations (5%) for each model element (i.e., each subcatchment or pipe). The differences were then normalized by dividing the mean of the parameters (based on all generated samples) to eliminate the influence of different units. The differences in Eq. (4) were calculated by both types of parameters and each specific subcatchment or pipe.

$$\Delta_{p,c} = \frac{\overline{Best_{5\%}} - \overline{Worst_{5\%}}}{\overline{Mean}} \quad (4)$$

We then calculated the sum of the absolute differences as the contributions of model parameters ( $p$ ) or model elements ( $c$ ) (the specific pipe or subcatchment) to improve NSE values from the worst to best based on the 1000 simulations. The change of parameters ( $\Delta_{p,c}$ ) relates to the specific parameter or model element. Therefore, the contribution of each element or parameter (contribution <sub>$c$</sub> , or contribution <sub>$p$</sub> , Eq. (5)) is represented by the sum of the changes related to specific parameter or element ( $\Delta_p$  or  $\Delta_c$ ) divided by the total changes. We also normalized the contribution of each subcatchment (contribution <sub>$c$</sub> ', Eq. (6)) by area to control for the influence of the subcatchment size.

$$\text{Contribution}_c = \frac{\sum(|\Delta_{p/c}|)}{\sum(|\Delta_{p,c}|)} \quad (5)$$

$$\text{Contribution}'_c = \frac{\text{Contribution}_c}{\text{Area}_c} \quad (6)$$

In these equations,  $p$  indicates the parameters whereas  $c$  indicates model elements (i.e., subcatchment or pipe).

## Model uncertainties on surface delineation

Sewer network simplification and the use of the Thiessen polygon approach inevitably introduced uncertainties to the model outputs. A clear drawback of the Thiessen polygon-based partition approach is that it ignores surface topography to route surface flow to the sewer system. We assumed the water from each subcatchment was routed to the corresponding node, which simplifies the process to determine which nodes the surface inflow drains to. However, we do not consider the influence of topography on subcatchment partitioning; this assumption can be violated if the subcatchment elevation is lower than the node elevation. Therefore, after subcatchment discretization, we checked this assumption by comparing the subcatchment average elevation with the node elevation across the sewershed.

In addition, we evaluated the proposed approach by comparing it to the more common approach of partitioning subcatchments according to topography. We built two small-scale models within the Panther Hollow watershed, which lies within our greater M29 study area, by using delineated subcatchments provided by the Pittsburgh Water & Sewer Authority (PWSA). The Panther Hollow watershed has an area of 185 acres (Fig. 1). To compare the land area covered by our polygon-based model to the PWSA delineated topography-based Panther Hollow model, we merged all the polygons related to that section of the sewer network. We then compared the simulations with observed pipe flow (April–May 2014) at the outfall point.

Finally, the proposed subcatchment partitioning approach was evaluated as a function of the amount of sewer network data available, represented by number of nodes. In some areas, detailed sewer data may be unavailable, leading modelers to use street intersections to estimate information about the sewer network. To explore the potential influence of sewer system structure uncertainty on model performance, we tested two scenarios: (1) randomly removing nodes but using the original sewer segments to routing flow; and (2) randomly removing nodes and building new sewer segments by connecting the remaining nodes based on their elevation. For each scenario, we conducted 4 trials, randomly removing 10%, 20%, 30%, and 40% of the total nodes (except for nodes on major trunk sewers). One hundred simulations were made for each trial, and the NSE values were calculated. Two locations with monitored flow were considered for NSE calculations: one near the outlet of the sewershed; the other on the northwest side of the sewer network (Fig. S3).

For the comparison between the Thiessen polygon and topography-based polygons within the Panther Hollow sub-watershed, we could not use the NSE as a goodness-of-fit measure, because observed fluctuations in sanitary base flow was not captured in the simulation and the NSE calculation heavily penalizes the model for these small flow errors, limiting the value of the assessment for evaluating stormflow predictions. Instead, the comparison of peak flow errors provided a reasonable method to evaluate the behavior of the two delineation methods relative to one another.

## Model calibration

A Monte Carlo-based calibration algorithm was applied based on generalized likelihood uncertainty estimation (GLUE) (Beven 1993), which has been proposed as a Bayesian approach to derive model uncertainties. This approach often uses informal likelihood functions rather than a correct statistical error model, which is not entirely consistent with the Bayesian inference process (Stedinger et al. 2008; Mantovan and Todini 2006), but its ease of implementation and flexibility for combining different model inputs provide capability to deal with a large number of parameters for model calibration. Thus, we combined

the GLUE procedure into the proposed calibration algorithm described as follows (Fig. 2):

(1) Define distributions of model parameters ( $x_i$ ). The initial distributions were defined as shown in Table 2. (2) Draw parameter samples by using Latin hypercube sampling (LHS) and run Monte Carlo simulations. (3) Choose objective functions as response variables ( $Y$ ). Here, the single objective function of NSE (Eq. 3) was used. (4) Divide simulations defined as “behavioral model runs” and “non-behavioral model runs”, in this case based on the median of the NSE values. That is, behavioral model runs were simulations with NSE values larger than the median. (5) Calculate the NSE weights  $W_j$  (Eq. 7) based on the behavioral model runs;  $j$  is the index of behavioral model runs. (6) Calculate the new mean  $\mu_{xi}$  and variance  $\sigma_{xi}^2$  for parameter  $x_i$ . (7) Assume the family of distributions corresponding to parameters does not change and then use the new mean (Eq. 8) and variance (Eq. 9) to draw new samples. (8) Repeat steps (2)–(7) until acceptable NSE values are achieved. The final means of the distributions are treated as optimal values for parameters

$$W_j = \frac{NSE_j}{\sum NSE_j} \quad (7)$$

**Fig. 2** Pseudocode table of calibration algorithm

### Algorithm 1: Calibration

---

**Input:** initial parameter distributions  $f_{x_i, initial}$ , demand samples ( $n = 1000$ ), number of iteration  $N$

**Result:** optimal parameter values  $x_i$ ,

```

1   $n \leftarrow 1$ ;
2   $f_{x_i} \leftarrow f_{x_i, initial}$ ;
3  while  $n \leq N$  do
4      draw samples with LHS based on  $f_{x_i}$ ;
5      run Monte Carlo simulations;
6      calculate  $NSE$  values;
7       $threshold \leftarrow NSE_{median}$ ;
8      if  $NSE \geq threshold$  then
9          | behavioral runs  $\leftarrow NSE_j$ ;
10     else
11         | non-behavioral runs  $\leftarrow NSE_k$ 
12     end
13      $W_j \leftarrow NSE_j / \sum NSE_j$ ;
14     update parameter ( $x_i$ ) distribution  $f_{x_i, new}$ ;
15      $\mu_i \leftarrow \sum W_j \times x_{ij}$ ;
16      $\sigma_i \leftarrow \sum W_j \times (x_{ij} - \mu_i)^2$ ;
17      $f_{x_i} \leftarrow f_{x_i, new}$ ;
18      $n \leftarrow n + 1$ ;
19 end
20 optimal  $x_i \leftarrow \mu_i$ ;

```

---

$$\mu_{xi} = \sum W_j \times x_{ij} \quad (8)$$

$$\sigma_{xi}^2 = \sum W_j \times (x_{ij} - \mu_{xi})^2 \quad (9)$$

## Results and discussion

### Parameter sensitivity analyses for subcatchments and pipes

Based on the Spearman correlation coefficients for total runoff volume, imperviousness had the highest correlation coefficient and dominated total runoff predictions (Fig. S5 (A)). The remaining parameters all have correlation coefficient distributions with a median of zero that cover positive and negative signs (Fig. S5 (A)), indicating that these parameters have no consistent effects on total runoff.

The influence of parameters on the predictions of peak flow is much clearer in both their magnitudes and signs (Fig. S5 (B)). The most influential parameter is also the imperviousness, with a median of 0.63 (range of 0.5 to 1.0). Manning's roughness for pervious area has a negative influence, while width and slope have positive influence that are comparable to one another. Our findings are similar to previous work documenting the fundamental role of imperviousness (Muleta et al. 2013; Barco et al. 2008) and consistent with observations that the Manning's roughness and width have effects on SWMM hydrology simulations (Tsai et al. 2017; C. Li et al. 2014) and flow is less sensitive to infiltration parameters (Li et al. 2014).

All Spearman correlation coefficients for pipe parameters have a zero median with a range between  $-0.1$  and  $0.1$ , indicating that the three pipe parameters influence pipe flow minimally, inconsistent with previous studies that indicated pipe Manning's roughness (Sun et al. 2013; Li et al. 2014) and the conduit geometry (Peterson and Wicks 2006) can impact the flow rates. The disagreements in estimations of

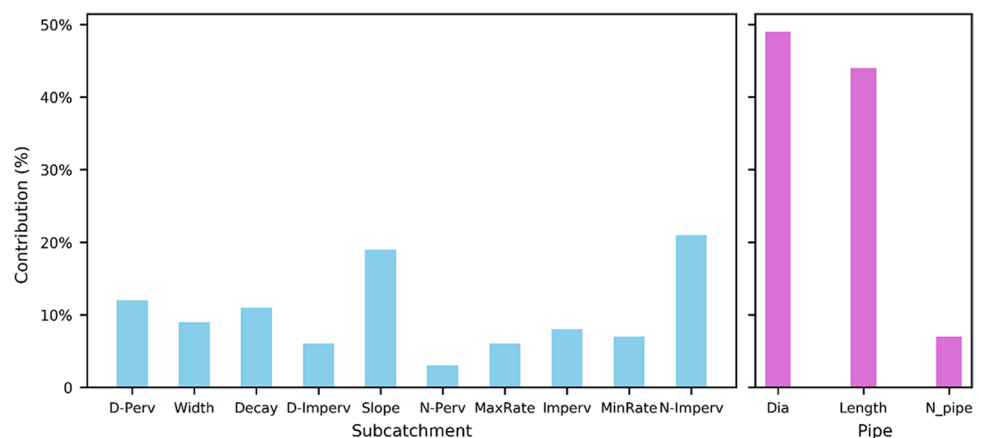
parameter importance across different studies may be caused by assuming different ranges for parameter distributions. For example, this study assumed a uniform distribution for Manning's roughness in pipes with a range of 0.011 and 0.015, a typical range for pipes (American Society of Civil Engineers and Water Environment Federation 2018; Bizier 2007). A much wider distribution, between 0 and 0.1, was assumed by Sun (Sun et al. 2013), which resulted in a higher influence of pipe roughness. However, it is not clear these roughness values are plausible in real pipe environments.

### Parameter contribution to overall model performance

Global parameter contributions show the Manning's roughness for impervious area contributed substantially to improving overall model performance (Fig. 3). However, the imperviousness did not contribute as much as the slope and width. Parameters related to infiltration together contributed nearly 25% to higher model performance when parameter values were varied. The influence of pipe length and diameter contributed as much as 92% to improved NSE values, highlighting the importance of the geometric parameters. However, the influence of pipe geometric parameters is small based on their Spearman correlation coefficients (Fig. S5). This may suggest that the contribution of parameters to improve NSE includes correlated information among subcatchments and pipes that has not been captured by the Spearman correlation coefficient, because they are calculated independently within each specific subcatchment or pipe line, but the parameter contributions to higher NSE values reflect the influence of parameters from all the subcatchments and pipes together.

We evaluated the contribution of each model element (subcatchment or pipe) to model performance, to explore cases where the parameters varied widely when comparing best-performing model runs with worst-performing ones. Subcatchments in the center of our study area did

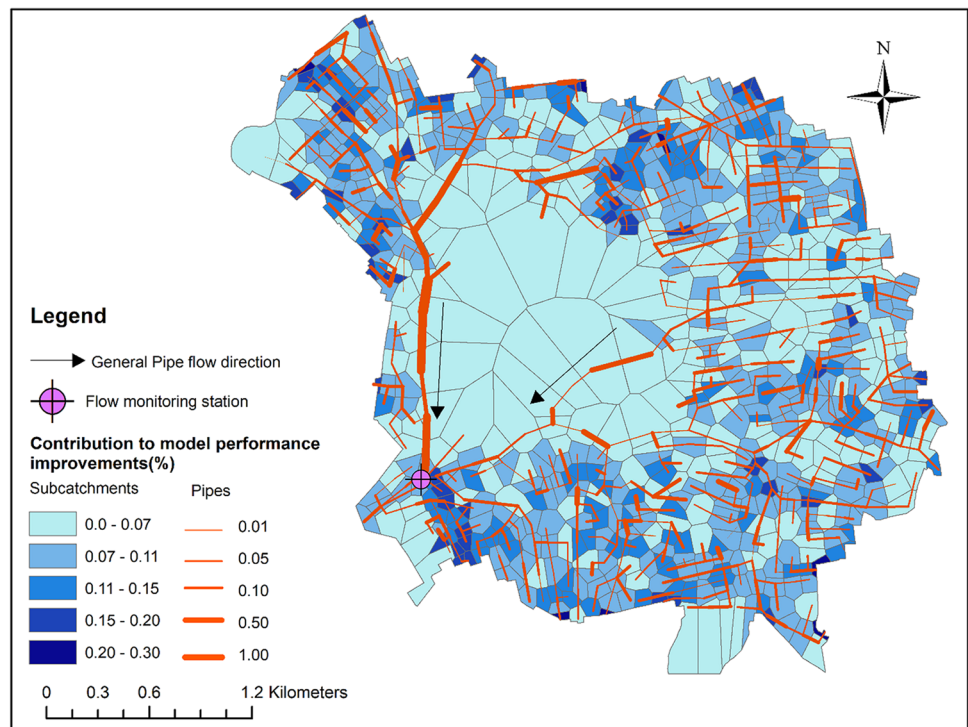
**Fig. 3** Parameter contributions for subcatchments and pipes to achieve better model performance (higher NSE values)



not contribute much to model improvement (Fig. 4). These subcatchments are in Schenley Park, a relatively undeveloped area (parkland) with low imperviousness. The higher impervious areas in subcatchments surrounding the park contribute relatively more, requiring greater changes from worst-performing parameter sets before model performance improvements were observed.

We assumed the mean of pipe diameters was 2.4 m. Under Monte Carlo simulations, the smallest diameter observed in the parameters from the best simulation is 0.65 m. The actual upstream pipe diameters are also less than 0.91 m based on available data from ALCOSAN. After simplification during model construction, some of those pipes effectively had larger diameters when multiple pipes were combined into a simplified network. However, based on this analysis 2.4 m may still be too large a value for upstream pipes and may cause pipe segments at the top of the sewer system to have inordinately large contributions to model performance (darker orange lines, Fig. 4). The pipe segments close to the end-node reach a diameter as large as 4.2 m based on the sewer data from ALCOSAN, but in our system, the diameter could also be much larger ( $> 8.4$  m) due to the effect of merged pipe segments. Therefore, in Monte Carlo runs with good model performance (high NSE values), downstream pipe diameters were sampled from the higher end of the assumed parameter distribution and functioned to allow more water to flow through these lower portions of the sewershed.

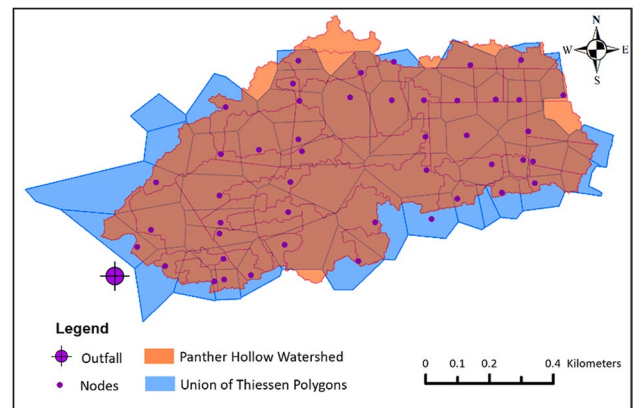
**Fig. 4** Contribution of model elements (subcatchments or pipes) to NSE improvements



## Uncertainties associated with model discretization

The subcatchments delineated by using two approaches are shown in Fig. 5. The area of merged polygons within Panther Hollow watershed is 221 acres, which is larger than the area of Panther Hollow watershed of 185 acres.

A comparison between the two small-scale Panther Hollow models shows similar results for pipe flow simulations (Fig. S6). For simulated peak errors, the terrain-based model has a peak flow error with a median of  $-1.05\%$  (mean  $-0.019\%$ , range  $-39$  to  $40\%$ ), while the



**Fig. 5** Comparison of model subcatchments, overlaying Thiessen polygons and terrain-based subcatchments delineated by PWSA in the Panther Hollow watershed

polygon-based model has a median error of 1.27% (mean 7.67%, range – 31 to 61%). Compared to the delineation that considers topography, the Thiessen polygon-based model can simulate peaks with similar fidelity to observed flows, although it tends to generate higher peaks and leads to more positive peak errors than those simulated by the topography-based model.

When comparing the subcatchment average elevations with the node elevations, there are 230 subcatchments (35% of the total subcatchments) that violate the surface routing assumption (from surface to node), with elevation differences between 0.004 and 15% (Fig. S7). However, some small elevation differences can be caused by the intrinsic errors in DEM. Thus, we only considered routing errors for subcatchments with mean elevations more than 1% lower than the node elevation (Fig. 6). This resulted in 44 subcatchments (6.8%) that violated the routing assumption, many of which are located in rugged portions of the sewershed. This finding suggests even though the naïve assumptions that the nodes receive surface flow can be violated, in the majority of cases the routing assumption is satisfied. Based on these results, the fast and efficient Thiessen polygon approach is appropriate to use for relatively flat areas, but care must be taken for rugged areas in which polygons may need to be further manually adjusted.

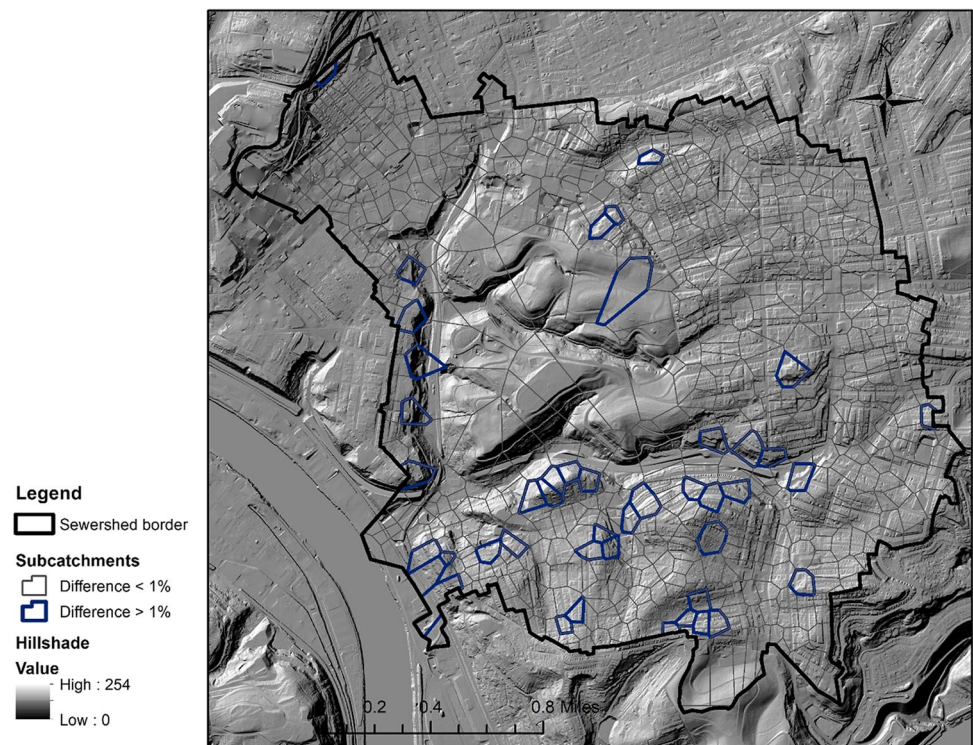
The results of tests on the influence of level of sewer network information on model performance show there are no statistical differences among these simulations. We simulated increasing uncertainty regarding the sewer

network by conducting simulations with 10–40% of nodes randomly removed. In the first case, we kept the information about the pipe network the same as the original model, so that the lack of knowledge simulated was only about the location of the nodes (stormwater inlets). In the second case, we removed the nodes and reconnected the pipe sewers only around the “known” nodes. Even though the variance of the NSE values for predicted flow at the two flow monitoring locations increased as more nodes were removed (Fig. 7), the values remained larger than 0.7 for all 800 simulations (slightly higher at location 2). According to these simulations, no obvious influence of the pipe nodes used to draw polygons on overall model performance (NSE) could be observed, for both fixed sewer segments (Fig. 7A, B) and rebuilt sewer segments (Fig. 7C, D). This suggests that the polygon approach can robustly estimate urban sewershed surface-pipe linkages even when using limited data, provided the major trunk lines are known.

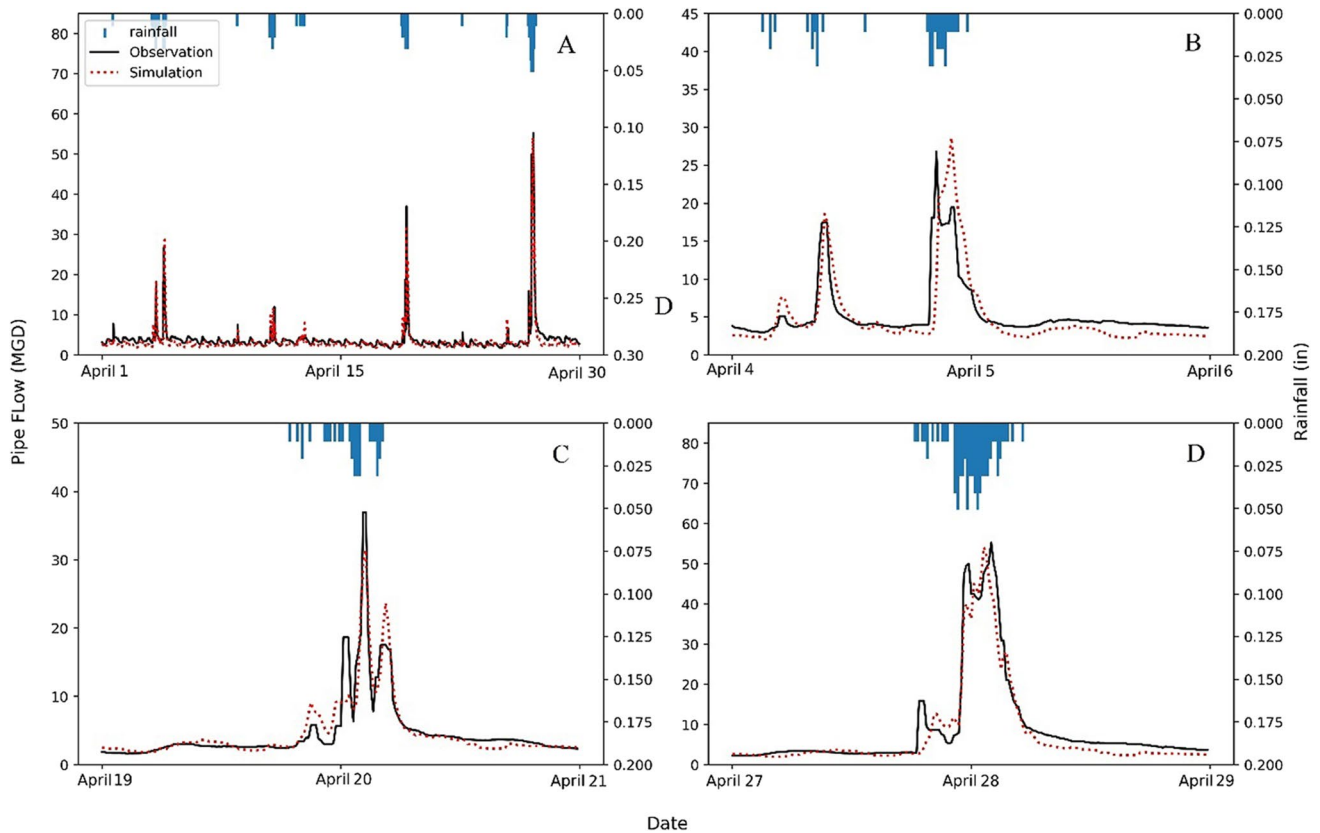
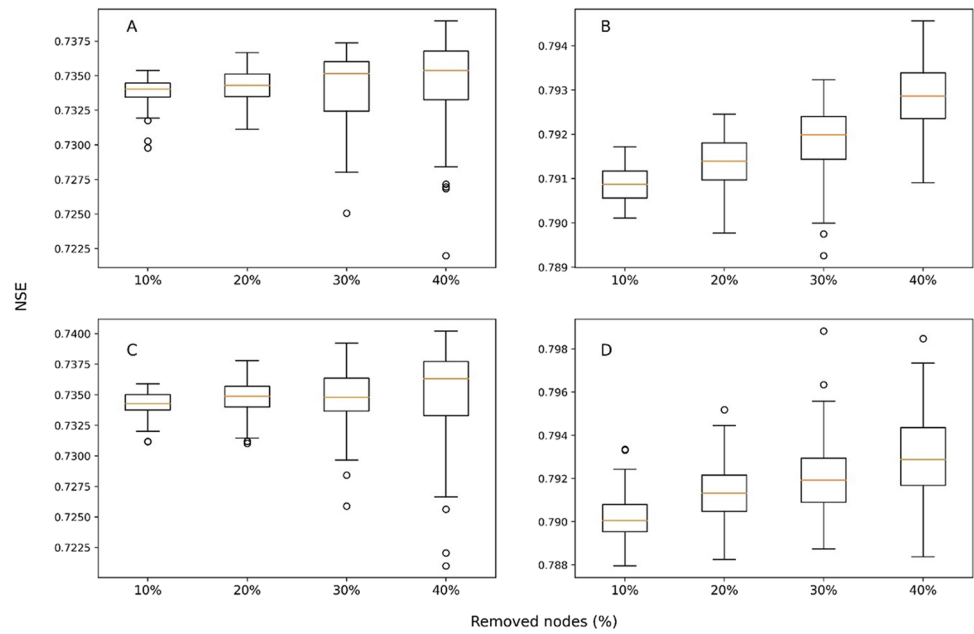
### Model calibration

Based on the results of sensitivity analyses, eight parameters were chosen to calibrate the model, including six subcatchment parameters: slope, width, imperviousness, Manning’s roughness in impervious area, depression storage in pervious area and decay coefficient of infiltration. Two pipe parameters were also chosen: pipe diameter and Manning’s roughness. Three iterations (3000 simulations) of the calibration procedure

**Fig. 6** Comparison between subcatchment average elevation and node elevation with 1% difference threshold



**Fig. 7** Model performance (NSE) as a function of number of nodes used to build the model: **A** location 1 with fixed sewer network; **B** location 2 with fixed sewer network; **C** location 1 with rebuilt sewer network; **D** location 2 with rebuilt sewer network



**Fig. 8** Model performance after calibration compared to observed pipe flow: **A** Simulation results compared with observed pipe flow (black line) for the month of April 2008. **B** Detail of simulation, April

4 to April 6. **C** Detail of simulation, April 19 to April 21. **D** Detail of simulation, April 27 to April 29

were conducted and achieved a median NSE larger than 0.8. Changes of the variance of parameters are shown in Fig. S8, in which the variance of parameters within most of the subcatchments or pipes decreased after calibration, leading to smaller 95% confidence intervals (CI) for simulated pipe flows (shown in Fig. S9). This suggests it is feasible to use high-performance model runs to update parameters. After calibration, the model has an NSE of 0.82 and RMSE of 1.46. The shapes and magnitudes of flows are well predicted and the time to peak also corresponds well with most observed data, particularly for large rain events in late April (Fig. 8C–D).

## Conclusion

In this study, we demonstrated an efficient means to build an urban sewershed stormwater model by drawing Thiessen polygons around sewer nodes. We assumed that the surface flow within subcatchments drained to their corresponding nodes and did not re-route any flow between subcatchments. After calibration, predicted flow patterns near the outlet of the sewershed during the month of April 2008 showed good agreement with monitored pipe flow data (NSE of 0.82). The number of nodes used to build the sewershed model did not dramatically influence model performance (NSE), provided the location of key trunk sewers was known. The proposed calibration algorithm was easy to apply even with a model with a large number of parameters. Therefore, this simple and efficient partitioning approach can be successfully used to build a reasonable sewershed model for predicting pipe flow.

We characterized the influence of SWMM parameters on total surface runoff volume, surface runoff peak flow, maximum pipe flow and NSE values. The four most important parameters influencing peak flow were imperviousness > Manning's roughness for impervious area > width > slope. Maximum pipe flow was not sensitive to the pipe geometry when evaluated based on the Spearman correlation coefficients, but its influence was clear when parameter contributions were analyzed based on comparison of low and high NSE Monte Carlo runs. This suggests, for model calibration, that the focus on improving parameter estimates can be different with different model output objectives.

One limitation of our approach is that it cannot fully analyze the correlated influence among parameters on the model outputs. We used the Spearman correlation coefficients to evaluate the relationship between parameters and their related individual elements (subcatchments or pipes) separately. However, the influences of surface parameters on pipe peak flow were not captured by the Spearman correlation coefficients. In **model calibration**, we only considered a single objective function (NSE). Although it is believed parameter space can be generated by combining behavior model runs with multiple criteria, we did not test

the capability of algorithm to calibrate multiple objectives. Our model assumption that subcatchments route surface water to nodes should be tested in other urban areas to validate its feasibility and limits of such an approach. As evidenced by the violation of routing for particularly rugged areas of our sewershed, when the subcatchment straddles a topographic feature like a sheer wall, the averaged elevation does not reflect the flow dynamics throughout the subcatchment.

In all, the proposed Thiessen polygon approach could generate good pipe flow simulations even when minimal detail is provided on topography and surface cover. Therefore, it is possible to build urban stormwater models using the Thiessen polygon approach in areas that lack sewer data, particularly if the focus is to simulate pipe flow near the outlet of an urban sewershed.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11356-022-24162-7>.

**Acknowledgements** We thank the Allegheny County Sanitary Authority and Pittsburgh Water & Sewer Authority for graciously providing sewer flow data and field data.

**Author contribution** Zhaokai Dong contributed to the investigation, methodology, validation, visualization, software, and wrote the original draft. Dr. Daniel Bain contributed to the methodology, resources, review, and editing. Dr. Murat Akcakaya contributed to the funding acquisition, methodology, review, and editing. Dr. Carla Ng contributed to the funding acquisition, methodology, project administration, resources, supervision, review, and editing.

**Funding** This work was supported by the National Science Foundation, USA (NSF grant number 1854827).

**Data availability** The simplified sewer network and model parameterization is available from the authors on request.

## Declarations

**Ethics approval** This study did not involve human subjects or animal research.

**Consent to participate** Not applicable.

**Consent for publication** All authors agree with the content to submit and publish.

**Competing interests** The authors declare no competing interests.

## References

- 3 Rivers Wet Weather (3RWW) (2008) Rain gauge and calibrated radar rainfall data. <https://www.3riverswetweather.org/municipalities/calibrated-radar-rainfall-data>. Accessed 2 Oct 2021
- Aad A, Maya P, Suidan MT, Shuster WD (2010) Modeling techniques of best management practices: rain barrels and rain gardens using EPA SWMM-5. *J Hydrol Eng* 15(6):434–443. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000136](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000136)

- Ahiablame LM, Engel BA, Chaubey I (2012) Effectiveness of low impact development practices: literature review and suggestions for future research. *Water Air Soil Pollut* 223(7):4253–4273. <https://doi.org/10.1007/s11270-012-1189-2>
- Allegheny County Division of Computer Services Geographic Information Systems Group (2016). Allegheny county boundary. [https://www.pasda.psu.edu/uci/FullMetadataDisplay.aspx?file=AlleghenyCounty\\_Boundary2016.xml](https://www.pasda.psu.edu/uci/FullMetadataDisplay.aspx?file=AlleghenyCounty_Boundary2016.xml). Accessed 2 Oct 2021
- Allegheny County Division of Computer Services Geographic Information Systems Group (2017) Allegheny county lidar and terrain products. <https://www.pasda.psu.edu/uci/FullMetadataDisplay.aspx?file=AlleghenyCountyLidar2017.xml>. Accessed 2 Oct 2021
- Allegheny County Sanitary Authority (ALCOSAN) (2022) Conveyance and treatment system with regional collection system outfalls. [https://www.alcosan.org/docs/default-source/system-mapping/sanitary-combined-sewer-outfalls\\_october2022\\_i.pdf](https://www.alcosan.org/docs/default-source/system-mapping/sanitary-combined-sewer-outfalls_october2022_i.pdf). Accessed 17 Nov 2022
- Alves A, Sanchez A, Vojinovic Z, Seyoum S, Babel M, Brdjanovic D (2016) Evolutionary and holistic assessment of green-grey infrastructure for CSO reduction. *Water* 8(9):402. <https://doi.org/10.3390/w8090402>
- American Society of Civil Engineers and Water Environment Federation (2018) Design and construction of urban stormwater management systems. American Society of Civil Engineers, Reston, VA. <https://doi.org/10.1061/9780872628557>
- Arnold CL, Gibbons CJ (1996) Impervious surface coverage: the emergence of a key environmental indicator. *J Am Plann Assoc* 62(2):243–258. <https://doi.org/10.1080/01944369608975688>
- Barco J, Wong KM, Stenstrom MK (2008) Automatic calibration of the U.S. EPA SWMM model for a large urban catchment. *J Hydraul Eng* 134(4):466–474. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:4\(466\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:4(466))
- Beven K (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv Water Resour* 16(1):41–51. [https://doi.org/10.1016/0309-1708\(93\)90028-E](https://doi.org/10.1016/0309-1708(93)90028-E)
- Bizier P (ed) (2007) Gravity sanitary sewer design and construction. In: ASCE Manuals and Reports on Engineering Practice, 2nd edn, no. 60. American Society of Civil Engineers, Reston, VA; Alexandria, VA
- Brassel KE, Reif D (2010) A procedure to generate Thiessen polygons. *Geogr Anal* 11(3):289–303. <https://doi.org/10.1111/j.1538-4632.1979.tb00695.x>
- Coffman LS, Robert G and Rod F (1999). “Low-impact development: an innovative alternative approach to stormwater management.” In *Low-Impact Development: An Innovative Alternative Approach to Stormwater Management*, 1–10. [https://doi.org/10.1061/40430\(1999\)118](https://doi.org/10.1061/40430(1999)118)
- Eckart K, McPhee Z, Bolisetti T (2017) Performance and implementation of low impact development – a review. *Sci Total Environ* 607–608(December):413–432. <https://doi.org/10.1016/j.scitotenv.2017.06.254>
- Esri Inc (2016) ArcMap (Version 10.5.1) (version 10.5.1). Redlands, CA: Esri Inc
- Hopkins KG, Bain DJ, Copeland EM (2014) Reconstruction of a century of landscape modification and hydrologic change in a small urban watershed in Pittsburgh, PA. *Landscape Ecol* 29(3):413–424. <https://doi.org/10.1007/s10980-013-9972-z>
- Horton Robert E (1941) An approach toward a physical interpretation of infiltration-capacity. *Soil Sci Soc Am J* 5(C):399–417. <https://doi.org/10.2136/sssaj1941.036159950005000C0075x>
- Huang M, Jin S (2019) A methodology for simple 2-D inundation analysis in urban area using SWMM and GIS. *Nat Hazards* 97(1):15–43. <https://doi.org/10.1007/s11069-019-03623-2>
- Jang S, Cho M, Yoon J, Yoon Y, Kim S, Kim G, Kim L, Aksoy H (2007) Using SWMM as a tool for hydrologic impact assessment. *Desalination* 212(1–3):344–356. <https://doi.org/10.1016/j.desal.2007.05.005>
- Krebs G, Kokkonen T, Valtanen M, Koivusalo H, Setälä H (2013) A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. *Urban Water J* 10(6):394–410. <https://doi.org/10.1080/1573062X.2012.739631>
- Krebs G, Kokkonen T, Valtanen M, Setälä H, Koivusalo H (2014) Spatial resolution considerations for urban hydrological modelling. *J Hydrol* 512(May):482–497. <https://doi.org/10.1016/j.jhydrol.2014.03.013>
- Lee JG, Selvakumar A, Alvi K, Riverson J, Zhen JX, Shoemaker L, Lai F-H (2012) A watershed-scale design optimization model for stormwater best management practices. *Environ Model Softw* 37(November):6–18. <https://doi.org/10.1016/j.envsoft.2012.04.011>
- Li C, Wang W, Xiong J, Chen P (2014) Sensitivity analysis for urban drainage modeling using mutual information. *Entropy* 16(11):5738–5752. <https://doi.org/10.3390/e16115738>
- Li H, Sharkey LJ, Hunt WF, Davis AP (2009) Mitigation of impervious surface hydrology using bioretention in North Carolina and Maryland. *J Hydrol Eng* 14(4):407–415. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2009\)14:4\(407\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:4(407))
- Mantovan P, Todini E (2006) Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. *J Hydrol* 330(1–2):368–381. <https://doi.org/10.1016/j.jhydrol.2006.04.046>
- McCall A-K, Palmitessa R, Blumensaat F, Morgenroth E, Ort C (2017) Modeling in-sewer transformations at catchment scale – implications on drug consumption estimates in wastewater-based epidemiology. *Water Res* 122(October):655–668. <https://doi.org/10.1016/j.watres.2017.05.034>
- McCuen RH, Knight Z, Gillian Cutter A (2006) Evaluation of the Nash-Sutcliffe Efficiency Index. *J Hydrol Eng* 11(6):597–602. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:6\(597\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597))
- McKay MD, Beckman RJ, Conover WJ (2000) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42(1):55–61. <https://doi.org/10.1080/00401706.2000.10485979>
- Muleta MK, McMillan J, Amenu GG, Burian SJ (2013) Bayesian approach for uncertainty analysis of an urban storm water model and its application to a heavily urbanized watershed. *J Hydrol Eng* 18(10):1360–1371. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000705](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000705)
- National Research Council (2009) Urban stormwater management in the United States. The National Academies Press: National Research Council, Washington, DC. <https://doi.org/10.17226/12465>
- Okabe A, Boots B, Sugihara K, Chiu SN (2000) Spatial Tessellations: concepts and applications of Voronoi Diagrams, 2nd edn. John Wiley & Sons, Chichester
- Palla A, Gnecco I (2015) Hydrologic modeling of low impact development systems at the urban catchment scale. *J Hydrol* 528(September):361–368. <https://doi.org/10.1016/j.jhydrol.2015.06.050>
- Peterson EW, Wicks CM (2006) Assessing the importance of conduit geometry and physical parameters in Karst systems using the storm water management model (SWMM). *J Hydrol* 329(1–2):294–305. <https://doi.org/10.1016/j.jhydrol.2006.02.017>
- Rossman LA (2017) Storm water management model reference manual volume II – hydraulics. EPA/600/R-17/111: U.S. Environmental Protection Agency, Washington, DC. [https://cfpub.epa.gov/si/si\\_public\\_record\\_report.cfm?Lab=NRMRL&dirEntryId=337162](https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=337162). Accessed 2 Oct 2021
- Rossman LA, Huber WC (2016) Storm water management model reference manual volume I – hydrology. EPA/600/R-15/162A: US EPA Office of Research and Development, Washington, DC. [https://cfpub.epa.gov/si/si\\_public\\_record\\_report.cfm?Lab=NRMRL&dirEntryId=309346](https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=309346). Accessed 2 Oct 2021

- Saltelli A, Tarantola S, Campolongo F (2000) Sensitivity analysis as an ingredient of modeling. *Stat Sci* 15:377–395
- Schueler TR, Fraley-McNeal L, Cappiella K (2009) Is impervious cover still important? Review of recent research. *J Hydrol Eng* 14(4):309–315. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2009\)14:4\(309\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:4(309))
- Spearman C (1904) ‘General intelligence’, objectively determined and measured. *Am J Psychol* 15:201–292
- Stedinger JR, Vogel RM, Lee SU, and Batchelder R (2008) “Appraisal of the generalized likelihood uncertainty estimation (GLUE) method: APPRAISAL OF THE GLUE METHOD.” *Water Resources Res* 44 (12). <https://doi.org/10.1029/2008WR006822>
- Sun N, Hong B, Hall M (2013) Assessment of the SWMM model uncertainties within the generalized likelihood uncertainty estimation (GLUE) framework for a high-resolution urban sewershed. *Hydrol Process* 28(May):3018–3034. <https://doi.org/10.1002/hyp.9869>
- Tsai L-Y, Chen C-F, Fan C-H, Lin J-Y (2017) Using the HSPF and SWMM models in a high pervious watershed and estimating their parameter sensitivity. *Water* 9(10):780. <https://doi.org/10.3390/w9100780>
- U.S. Environmental Protection Agency (2014) Estimating Change in Impervious Area (IA) and Directly Connected Impervious Areas (DCIA) for Massachusetts Small MS4 Permit. Boston, MA. <https://www3.epa.gov/region1/npdes/stormwater/ma/MADCIA.pdf>. Accessed 2 Oct 2021
- U.S. Geological Survey (2014) NLCD 2001 Land Cover (2011 Edition, Amended 2014) - National Geospatial Data Asset (NGDA) Land Use Land Cover. Sioux Falls, SD. <https://www.mrlc.gov/downloads/sciweb1/shared/mrlc/metadata/landcover.html>. Accessed 2 October 2021
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.