Automated urban rainfall-runoff model generation with detailed land cover and flow routing

Tero J. Niemi, D.Sc.\(^1\); Teemu Kokkonen, D.Sc.\(^2\); Nora Sillanpää, D.Sc.\(^3\); Heikki Setälä, Ph.D.\(^4\); Harri Koivusalo, D.Sc.\(^5\)

\(^1\) Postdoctoral Researcher, Department of Built Environment, Aalto University School of Engineering, PO Box 15200, FI-00076 Aalto, Finland (corresponding author). Phone: +358 50 562 5427, E-mail: tero.niemi@aalto.fi

\(^2\) Senior University Lecturer, Department of Built Environment, Aalto University School of Engineering, PO Box 15200, FI-00076 Aalto, Finland. E-mail: teemu.kokkonen@aalto.fi

\(^3\) Postdoctoral Researcher, Department of Built Environment, Aalto University School of Engineering, PO Box 15200, FI-00076 Aalto, Finland. E-mail: nora.sillanpaa@aalto.fi

\(^4\) Professor, Faculty of Biological and Environmental Sciences, University of Helsinki, Niemenkatu 73, FI-15140 Lahti, Finland. E-mail: heikki.setala@helsinki.fi

\(^5\) Professor, Department of Built Environment, Aalto University School of Engineering, PO Box 15200, FI-00076 Aalto, Finland. E-mail: harri.koivusalo@aalto.fi
Abstract

Constructing hydrological models for large urban areas is time consuming and laborious due to the requirements for high-resolution data and fine model detail. An open-source algorithm using adaptive subcatchments is proposed to automate Storm Water Management Model (SWMM) construction. The algorithm merges areas with homogeneous land cover and common outlet into larger subcatchments, while retaining small-scale details where land cover or topography is more heterogeneous. The method was tested on an 85 ha urban catchment in Helsinki, Finland. A model with adaptive subcatchments reproduced the observed discharge at the catchment outlet with high model-performance indices emphasizing the strength of the proposed method. Computation times of the adaptive model were substantially lower than those of a corresponding model with uniformly sized high-resolution subcatchments. Given that high-resolution land cover and topography data are available, the proposed algorithm provides an advanced method for implementing SWMM models automatically even for large urban catchments without substantial manual workload. Simultaneously, the high-resolution land cover details of the catchments can be maintained where they matter the most.

Keywords

SWMM, urban hydrology, stormwater, subcatchment delineation, automation, flow routing
1. Introduction

Urban areas are characterized by fragmented, mosaic land cover leading to altered hydrological cycle when compared to natural areas. The changing landscape due to urbanization has impacts on heat balance and evaporation (Whitford et al., 2001; Zhou et al., 2011), snow cover and snowmelt (Bengtsson and Semádeni-Davies, 2011), infiltration and storm runoff generation (Sillanpää and Koivusalo, 2015), and biodiversity (Pauleit et al., 2005) amongst other effects. To understand the hydrology-related processes in urban areas, accurate description and understanding of the land cover spatial configuration is crucial (Fletcher et al., 2013; Salvadore et al., 2015).

The high-resolution description of catchment details is important in urban hydrological models (Cantone and Schmidt, 2009). While for runoff volumes the impact of spatial resolution may be modest or even negligible (Ghosh and Hellweger, 2012; Goldstein et al., 2016; Krebs et al., 2014; Park et al., 2008), the high detail in land cover description is particularly important for accurate simulation of peak flow rates (Elliott et al., 2009; Ghosh and Hellweger, 2012; Krebs et al., 2014). In addition to the increased accuracy of runoff simulations, describing subcatchments in high-resolution as detailed units with homogeneous land cover simplifies the model calibration procedure and narrows parameter ranges (Krebs et al., 2013; Sun et al., 2014).

In urban areas, impervious surfaces contribute the most to urban runoff and understanding their connection to surrounding areas and to the stormwater network is fundamental (Jacobson, 2011; Mejía and Moglen, 2010; Shuster et al., 2005). To accurately represent the flow routing between contributing surfaces in urban hydrological models, flow paths have to be described in detail requiring high-resolution data (Gironás et al., 2010; Rodriguez et al., 2013). The demand for high-resolution spatial descriptions of urban areas is also driven by the assessment...
of stormwater management systems, such as low-impact development and nature-based solutions. These are often spatially distributed to individual surfaces or outlets of impervious plots and need to be described in great spatial detail in models (Tuomela et al., 2019).

The US EPA Storm Water Management Model (SWMM) (Rossman, 2015) is a widely used open-source urban hydrological simulation model used for both event-based (e.g., Kong et al., 2017; Niemi et al., 2017) and long-term (e.g., Guan et al., 2015; Peleg et al., 2017; Taka et al., 2017) hydrological assessments in urban areas. Manual construction of high-resolution urban hydrological models, where each contributing surface is individually described (Krebs et al., 2014, 2013), is only feasible when the studied area is small. For larger areas, such as entire suburbs, automated methods are necessary to keep the task manageable.

Several tools have been proposed to facilitate the task of urban hydrological model construction. Kertesz et al. (2007) developed a tool to compile and transfer subcatchment information from ArcGIS Geographical Information System (GIS) to SWMM. Pina et al. (2011) introduced an open-source tool inp.PINS for both creating SWMM input files directly from GIS and for visualizing SWMM results in GIS, but the tool has since become deprecated. Dongquan et al. (2009) presented a digital elevation model (DEM)-based automated batch process for subcatchment discretization in ArcGIS without accounting for different land covers within subcatchments and tested the results with SWMM. In addition, several commercial modelling packages built around the EPA SWMM computational engine exist (e.g., InfoSWMM and XPSWMM by Innovyze, PCSWMM by Computational Hydraulics International) to aid the modeler e.g., by incorporating superior GIS capabilities over the standard EPA SWMM user interface or by allowing integrated 1D-2D modelling using 2D surface flow descriptions. Nevertheless, even with commercial packages, subcatchment delineation and routing of water between subcatchments and into the stormwater network are
still largely manual tasks. Clearly, there is still room for improvement as none of the tools facilitate automatic model building while retaining the detailed land cover characteristics of the urban environment.

Easily available, remotely sensed high-resolution data on topography and land cover are abundant (Bates, 2004; French, 2003; Tarolli et al., 2013). This shifts the focus in hydrological modelling from describing the environment in as much detail as possible into such models where simulation times remain feasible (Bates, 2012; Sampson et al., 2012). High-resolution description of land cover and flow paths becomes important in urban environments and requires a modeler to find a balance between necessary level of detail and acceptable computational burden.

Describing a sizeable urban area in high-resolution often results in unfeasibly long simulation times necessitating for a method to aggregate adjacent surfaces into larger computational units. Warsta et al. (2017) proposed an automatic method for building SWMM models in a manner where each DEM/land cover raster cell corresponded to one subcatchment. The method also allowed combining individual grid cells into larger rectangular subcatchments in a rudimentary manner. While this decreased computation times and facilitated application to large urban catchments (Rautiainen, 2016), averaging catchment properties (e.g., elevations) while combining grid cells led to problems with surface runoff routing and to a loss of fine-scale detail in describing land cover and topography.

To tackle the challenge of automatically constructing SWMM models in high-resolution while minimizing the computational burden, the main objective of this paper was to propose a new open-source algorithm to automate SWMM model construction. Following the requirements for accurate flow path description and homogenous subcatchment land cover, the proposed algorithm automatically discretizes the studied area in an adaptive manner based on land cover
and flow routing using high-resolution land cover and DEM data. The result is a SWMM model with a minimum number of subcatchments where each subcatchment is covered by a single land cover type.

The performance of the new discretization method was demonstrated by comparison against a uniform discretization scheme where each raster cell corresponds to one subcatchment. Simulation results were also compared against field measurements to validate model performance.

2. Materials

The studied Länsi-Pakila catchment (Fig. 1) is an 85 ha urban area in Helsinki, Finland. Helsinki has a boreal climate with a mean annual air temperature of 5.9°C and mean annual precipitation of 655 mm, with most of the rainfall falling in late summer and early autumn (Pirinen et al., 2012).

The Länsi-Pakila catchment is a medium-density residential area characterized by detached houses. The area is relatively green, with vegetation covering 53.5%, asphalt 27.5%, and roofs 13.5% of the area (Table 1), resulting in a total imperviousness of 43%. The area is prone to stormwater flooding (Raukola, 2012). Länsi-Pakila is subject to urban development and faces a risk of more severe urban flooding in the future unless due attention is paid to stormwater management.

|TABLE 1|

An openly available 1 × 1 m² DEM from the City of Helsinki was used for catchment delineation. The catchment land cover description was based on the openly available land cover...
classification data from the Helsinki Region Environmental Services Authority HSY. As is
often the case with urban runoff studies, several site visits were required to complement the
scattered stormwater network information available from the network map.

Rainfall was measured at the Länsi-Pakila catchment during summer 2017 using three co-
located fully automatic tipping-bucket rain gauges (ECRN-100 High Resolution Rain Gauge)
with 0.2 mm tip size and 1 min temporal resolution. The gauges were located on top of a low-
rise nursing home building to keep them safe from vandalism and to minimize obstruction from
the urban surroundings (Fig. 1). Daily air temperature and wind speed data were available from
the Finnish Meteorological Institute’s weather station in Kumpula, Helsinki, approximately 5
km south-east from the catchment.

Catchment discharge information was obtained by measuring water level and flow velocity at
the catchment outfall (Fig. 1) in an 800 mm concrete pipe using a Starflow Ultrasonic Doppler
Instrument Model 6526. The time resolution of the discharge measurements was 1 min. The
instantaneous discharge measurements caused velocity fluctuations, which were smoothed
using a 5 min central moving average in further data preparation.

3. Methods

3.1. Adaptive subcatchment discretization

The proposed subcatchment discretization algorithm extends the automatic SWMM model
construction tool introduced by Warsta et al. (2017). They divided the investigated area into
subcatchments using a uniform computation grid with a desired spatial resolution. The grid
cells were then connected to each other and into the stormwater network. Following Krebs et
al. (2014), the generated subcatchments were small enough to be hypothesized to consist of a
single homogenous land cover type, e.g., a green area, a rooftop, or a paved road. This
simplifies the calibration of the resulting SWMM model by allowing parameters to be linked to distinct surface types. In the proposed algorithm, the subcatchments are still assumed to be of a homogeneous land cover type, but their sizes depend on the underlying land cover and flow routing. This reduces the number of subcatchments greatly in areas where land cover and topography are uniform.

The proposed algorithm proceeds as follows:

1. A model with a uniform computation grid where each grid cell corresponds to one subcatchment is created for the studied area.

2. Cells with open storm sewer nodes are initially saved as a set of one-cell-sized subcatchments. SWMM parameters for each subcatchment are adopted from the node cell.

3. Subcatchments are processed one by one.

4. All adjacent upstream cells routed into the currently processed subcatchment are listed:
   a. If an upstream cell has the same land cover as the currently processed subcatchment, the cell is merged to the subcatchment. Subcatchment parameters (area, elevation, slope) are updated.
   b. If an upstream cell has different land cover than the currently processed subcatchment, a new subcatchment is created. Subcatchment properties are copied from the underlying cell for the newly created subcatchment. Downstream subcatchment is set as the outlet of the new subcatchment.

5. Subcatchments and upstream cells are traversed until all cells contributing to any of the open storm sewer nodes are processed.

6. Neighbouring roof cells sharing their outlet are merged together to form a new subcatchment. Subcatchment parameters (area, elevation, slope) are computed for each
merged roof subcatchment. Depending on the given land cover class for the roof, the roof subcatchments are routed either

a. to the nearest adjacent (non-roof) subcatchment (disconnected roof in Fig. 1) or

b. directly to the nearest storm sewer node (connected roof).

Subcatchment cells with an open storm sewer node are connected into the stormwater network. Furthermore, a $3 \times 3$ cell area surrounding each open storm sewer node is used as a collecting area for the node to account for errors in flow routing resulting from inaccuracies in the DEM. Cells not contributing to any downstream subcatchment or storm sewer node, e.g., at the borders of the study area, are disregarded from further analysis. Subcatchment area is the combined area of the contributing cells, whereas subcatchment elevation and slope are assigned the average of the contributing cells. The flow width ($FW$) parameter for adaptive subcatchments is approximated after Krebs et al. (2014) as

$$FW = k \sqrt{A}$$  

(1)

where $A$ is the subcatchment area and $k = 0.7$.

The proposed algorithm requires three raster files with equal dimensions and resolution as inputs: a land cover raster where different land cover classes are identified with integers, a DEM raster depicting the topography of the studied area, and a flow direction raster with integers from 1 (north-east) to 8 (north) indicating the direction of flow from each raster cell. It is assumed that all cells in the flow direction raster are routed, i.e., there are no pits. Other input files to the tool are identical to those reported by Warsta et al. (2017), consisting of geometry files for the stormwater network, a file relating the land cover raster indices to SWMM subcatchment parameters, and various settings files. Given the input files, the tool
produces a SWMM input file (.inp) ready to be used in simulations and a set of GIS-compatible files that facilitate visualization of model set-up in GIS software.

3.2. Model implementations

Two models for the Länsi-Pakila catchment were created; a model where all subcatchments are rectangular and have dimensions of $1 \times 1 \text{ m}^2$ (referred to as $1x1$) and a model with adaptive subcatchments ($adap$) according to the proposed algorithm. The $1x1$ model acts as the reference model for assessing the subcatchment discretization impact on simulation performance.

Six largest rainfall-runoff events from the Länsi-Pakila catchment were selected for analysis (Table 2). Three of the events were used for model calibration and the remaining three were validation events. Earlier, Sillanpää and Koivusalo (2014, 2015) showed a difference in urban runoff response between minor and major storms due to the runoff-contributing area expanding from impervious to pervious areas during major storms. Because the storm size may affect model parameterization, and because the main interest was in potential urban flood-producing events, the selected events were all major storms (rainfall accumulation $>17$ mm). The threshold defining a major storm corresponds with the rainfall threshold set by Sillanpää and Koivusalo (2014) and Guan et al. (2016) for similar climate and catchment conditions.

[TABLE 2]

Land cover in both $1x1$ and $adap$ was represented with 6 classes. All model parameter values except for infiltration parameters corresponded to those used by Warsta et al. (2017) and Krebs et al. (2014) for similar urban catchments in Finland (Appendix A). Initial tests showed SWMM to be sensitive to the Green-Ampt infiltration model parameters, i.e., the suction head ($\psi_s$), the saturated hydraulic conductivity ($K_s$), and the maximum soil moisture deficit ($\theta_{dmax}$), and therefore these parameters were selected for the model calibration. Infiltration parameters
of the underlying soil were uniform for all land cover types and PEST (v.13) software (Doherty, 2016) with Tikhonov regularization was used to calibrate the *adap* model. The parameters for loamy sand from Rawls et al. (1992) were used as initial values. The calibration was conducted by minimizing the sum of squared errors of simulated flow against observed flow for time steps when observed flow exceeded a threshold of 7 – 15 l/s depending on the event. The threshold was selected due to the focus on potential flood-producing events and the desire to match peak flows rather than base flow. The same calibrated parameters were then used for the 1x1 model.

The flow direction raster was created from the DEM. As a pre-processing step, the stormwater network information was integrated into the DEM using “stream burning” (e.g., Saunders, 1999) to ensure maximum collecting area for the catchment. Subsequently, the original DEM without network burning from the corresponding area was used to produce the flow direction raster using r.watershed tool from GRASS GIS (GRASS Development Team, 2017). Cells with land cover classified as buildings were set to block the overland flow and cells with open storm sewer nodes were set to collect the water.

The model performance was evaluated against observations using the Nash-Sutcliffe efficiency *NSE* (-) (Nash and Sutcliffe, 1970), volume error (relative bias) *VE* (%), and peak flow error *PFE* (%).

\[
NSE = 1 - \frac{\sum_t (Q_{o,t} - Q_{s,t})^2}{\sum_t (Q_{o,t} - \bar{Q}_o)^2}
\]  
(2)

\[
VE = 100 \frac{V_s - V_o}{V_o}
\]  
(3)

\[
PFE = 100 \frac{Q_{s,max} - Q_{o,max}}{Q_{o,max}}
\]  
(4)
where $Q_{o,t}$ and $Q_{s,t}$ are the observed and simulated discharge (l/s), respectively, at time $t$, $\overline{Q_o}$ is the average observed discharge (l/s) during an event, $V_o$ and $V_s$ are the observed and simulated flow volumes (m$^3$), respectively, during an event, and $Q_{o,max}$ and $Q_{s,max}$ are the observed and simulated maximum discharges (l/s), respectively. The performance of $\text{adap}$ was evaluated against $1x1$ using the Pearson correlation coefficient $r$ (-):

$$r = \frac{\sum (Q_{1,t} - \overline{Q_1})(Q_{a,t} - \overline{Q_a})}{\sqrt{\sum (Q_{1,t} - \overline{Q_1})^2} \sqrt{\sum (Q_{a,t} - \overline{Q_a})^2}}$$

(5)

where $Q_{1,t}$ and $Q_{a,t}$ are the simulated discharges (l/s) at time $t$ from the $1x1$ and the $\text{adap}$ models, respectively, and $\overline{Q_1}$ and $\overline{Q_a}$ are the average simulated discharges (l/s) during an event using the $1x1$ and the $\text{adap}$ models, respectively. In addition, the performance was evaluated using volume difference $VD$ (%) and the peak flow difference $PFD$ (%) by substituting $1x1$ and $\text{adap}$ for observed and simulated in Eqs. (3) and (4), respectively.

4. Results

4.1. Adaptive subcatchment discretization

Table 3 presents the subcatchment statistics for $\text{adap}$ and $1x1$ subcatchment discretizations. The $\text{adap}$ model resulted in only 10% (82 554) of the number of uniform $1 \times 1$ m$^2$ subcatchments in $1x1$ (848 258). The subcatchment sizes for $\text{adap}$ ranged up to 9 322 m$^2$ with a mean size of 10.3 m$^2$. As both $\text{adap}$ and $1x1$ share the same input DEM data, the mean subcatchment elevation and slope are equal. Differences in the range of subcatchment elevations and slopes are due to some individual raster cell subcatchments of $1x1$ being merged in $\text{adap}$. The maximum subcatchment slope of over 400% is explained by local errors in the DEM. For $1x1$, with all the subcatchments having constant dimensions of $1 \times 1$ m$^2$, the subcatchment flow width is always either 1 or 0.7 m depending on whether the flow is
perpendicular or diagonal from the cell. For *adap* subcatchments, the flow width was computed using Eq. (1) resulting in a larger range of flow widths for individual subcatchments. The average subcatchment flow width was however similar for both models; 0.9 and 1.4 m for *lxl* and *adap*, respectively.

**TABLE 3**

Creating the *adap* models using the proposed algorithm consumed 30% more time (29.3 min) than creating the *lxl* model (22.6 min). The reduction in the number of subcatchments led to a corresponding reduction in SWMM computation time; the average computation time of an *adap* SWMM model was 10% of the corresponding *lxl* model computation time for the calibration and validation events (Fig. 2).

**FIGURE 2**

Fig. 3 shows the reduction in the number of subcatchments and the effect of combining individual cells into larger subcatchments on flow routes when moving from *lxl* to *adap*. The routing from one land cover type to another follows the same paths in *adap* and in *lxl*. However, in *adap* the number of subcatchments is substantially lower as cells with the same land cover type and sharing a common flow path have been merged. Note that the flow routing in Fig. 3 is displayed between subcatchment mass centers, creating an illusion that not all *adap* subcatchments are routed when in fact some subcatchment mass centers are situated outside the subcatchment and/or the figure.

**FIGURE 3**
4.2. Model calibration and validation

Fig. 4 illustrates the *adap* and *1x1* simulation results for the calibration and validation events, while Table 4 presents the corresponding performance statistics (Eqs. 2 – 5). The *adap* model performed well for both calibration and validation events. For all events except v1, *NSE* coefficients exceeded 0.90 indicating "very good" *adap* model performance, and *NSE* of 0.8 for event v1 still indicated "good" model performance according to the recommended model-performance classes by Ritter and Muñoz-Carpena (2013). Regardless of the event, simulations from *adap* tended to slightly underestimate flow volumes with the underestimate somewhat larger for the validation events (average *VE* −13.1%) than for the calibration events (−7.1%). This was partly explained by a more rapid return to pre-storm flow levels in the simulations than in the observed data, although some of the high flows were also underestimated. Peak levels were mainly well captured, with *PFE* less than 10% except for events c3 (*PFE* −11.3%) and v3 (19.3%). Only for event v3 did *adap* overestimate the peak flow.

[FIGURE 4]

[TABLE 4]

Performance of the *1x1* model in terms of *NSE* varied from "acceptable" for events v1 and c1 to "very good" for c2 and c3 (Table 4). Unlike in *adap* simulations where volume error was negative for all events, in *1x1* the flow volume was overestimated for all events with *VE* ranging from 5.1% for v1 to 22.9% for c1. For events c1 and v3, *1x1* overestimated the peak flow by roughly 20%. However, for the other events the peak flow was accurately simulated. The statistics between *adap* and *1x1* (Table 4) highlight the similar reaction of both models to rainfall events, as demonstrated by the correlation coefficients between *adap* and *1x1* approaching unity. Still, *adap* constantly produced 15 – 25% lower flow volumes than *1x1*. 
Although in absolute terms $VE$ of both models was similar, $adap$ on average underestimated observed flow volumes by 10.1% while $lxl$ overestimated by 11.6%. The $adap$ model generally predicted roughly 5% lower peak flows than $lxl$ with the maximum $PFD$ being $-17.4\%$ for event c1. For this event, the difference was almost entirely due to discharge overestimation of $lxl$.

The differences in peak flows and flow volumes are consistent with catchment mass balance differences between $adap$ and $lxl$ (Table 5). In each event, share of surface runoff was less for $adap$ than for $lxl$, whereas infiltration was greater for $adap$ than for $lxl$. In addition, $adap$ produced slightly less evaporation than $lxl$ while final stored water volume was slightly larger in $adap$ than in $lxl$.

5. Discussion

Automated DEM-based methods for SWMM subcatchment generation have been proposed before, but they differ from the current algorithm. Dongquan et al. (2009) aimed for a low number of computational units by using a high-resolution $2 \times 2 \text{ m}^2$ DEM but combining all cells belonging to the same drainage basin into subcatchments regardless of land cover or flow routing details. Their approach resulted in 113 subcatchments for a 13.65 ha study area, yielding a hundred times coarser average subcatchment size of $1 200 \text{ m}^2$ than in the $adap$ model here ($10.3 \text{ m}^2$). Warsta et al. (2017) described the catchments in fine detail but because each cell was considered an individual subcatchment, computation times were long with a large number of redundant cells in areas with a homogenous land cover type. This is analogous to the $lxl$ simulations here. In the proposed adaptive algorithm, more subcatchments are generated in areas where either land cover or flow routes are heterogeneous whereas in more
homogenous areas the subcatchments are allowed to have a larger size. The rudimentary grid cell aggregation procedure of Warsta et al. (2017) yielded shorter computation times, but changed the land cover and flow routing patterns in the catchment. This resulted in a moderate reduction of simulated peak flows and flow volumes. Here, the computation time was greatly reduced as *adap* simulations took on average only 10% of the corresponding 1x1 simulation time, while both *adap* and 1x1 produced good simulation results. The smaller computational burden associated with the proposed algorithm allows model construction for large urban catchments.

Because no manually constructed SWMM models exist for the studied catchment, direct comparison of computation times to a corresponding manual model was not possible. However, a crude estimate of a roughly five-fold increase in computation time between a manual model and a corresponding *adap* model was approximated by comparing models from earlier studies. In the work of Niemi et al. (2019), the proposed algorithm was used to create SWMM models with adaptive subcatchments for three small urban catchments (5.87, 6.63, and 12.59 ha catchment areas) in Lahti, Finland. Earlier, Krebs et al. (2014) manually constructed high-resolution SWMM models for the same catchments. The *adap* models in Lahti had, on average, 14.7 times the number of subcatchments when compared to the corresponding manual models, and required, on average, 4.8 times as long to compute.

The adaptive subcatchment discretization algorithm retains the high spatial resolution of the input DEM and land cover data where necessary, but creates larger subcatchments where such spatial detail is not crucial. This allows for an accurate spatial representation of land cover, deemed important by Cantone and Schmidt (2009) and Petrucci and Bonhomme (2014). However, it also relieves the computational burden that can become excessive with a uniformly high spatial resolution model. Given that input land cover data are in raster format, the
developed algorithm retains the land cover description from the input data. Otherwise, the
accuracy of the land cover description depends on the rasterization of non-raster-format input
data.

When building stormwater models manually, surfaces are usually assumed to drain entirely
into a single inlet node unless there is a compelling reason to resolve the routing in other way.
Therefore, an entire impervious surface, such as a parking lot, may be routed into a single inlet
although actual topography-driven flow paths would drain a part of the area to adjacent yards.
In the method of Warsta et al. (2017), routing of pit cells depended on their location in either
pervious or impervious areas. All water routed into pits residing in pervious areas was
infiltrated whereas water routed into pits in impervious areas was routed directly to the nearest
storm sewer node. As a result, some areas did not contribute to the catchment runoff as the
water had been infiltrated into a pit along its flow path. On the other hand, contribution of other
areas was unduly exaggerated as flow from them was routed directly to the stormwater
network. The proposed new algorithm allows the water to follow topography-driven flow
paths, and the use of a depressionless DEM ensures that water is routed through local pits. This
refined routing, compared to Warsta et al. (2017), should offer better runoff predictions during
major storms when pervious surfaces get saturated and start to convey runoff (Sillanpää and
Koivusalo, 2014; Yao et al., 2016).

Both adap and 1x1 models appropriately reproduced the observed runoff at the studied
catchment. The slight underestimation of flow volumes by adap was expected, as the
underestimation by SWMM in simulating hydrograph tails and low flows is commonly
encountered (e.g., Guan et al., 2015, 2016). This behaviour was accentuated by calibration of
adap focusing on high flows in lieu of low flows to more accurately simulate potential urban
flood-producing events.
Because the model parameters in *adap* and *1x1* were identical, differences in simulated flow volumes and peak flows between the implementations are explained by those model characteristics that were different: flow width, subcatchment slope, and subcatchment area. As these variables appear in the Manning equation SWMM uses to express the surface runoff for each computation time step, the dynamics of runoff production are altered when the parameters change. More importantly, the volume of infiltrated water within a subcatchment depends on its size. As the volume of the infiltrated water is the product of the area and infiltration depth, a larger subcatchment can infiltrate more water than an equally parameterized but a smaller subcatchment. In *adap*, the average subcatchment size was larger than in *1x1* and the likelihood of runon being completely or mostly infiltrated was larger. It is noteworthy, that the dependence of infiltration volume on subcatchment size involves all SWMM models, regardless of their construction procedure. The matter concerns especially models that treat infiltration as a loss from the system without consideration of the storage capacity of the underlying ground.

Adjusting only the infiltration parameters while taking other input parameters from Warsta et al. (2017) and Krebs et al. (2014) was sufficient to yield a well-performing model, with relative uncertainty variance reductions of 0.61, 0.99, and 0.96 for $\psi_s$, $K_s$, and $\theta_{d_{max}}$ respectively. These results are in line with earlier research suggesting that extensive calibration of a hydrological model may be unnecessary if representative parameter sets are available from similar catchments (Bárdossy, 2007; Gao et al., 2015; Kokkonen et al., 2003; Krebs et al., 2016). The results also support the findings of Petrucci and Bonhomme (2014) stating that an uncalibrated SWMM model may perform comparably to a calibrated model as long as land cover is described accurately. The slightly less accurate discharge simulations from *1x1* than *adap* were because infiltration parameters were calibrated using *adap* and applied to *1x1*
SWMM performance is often found to be sensitive to the subcatchment flow width parameter (Niazi et al., 2017). Here, the sensitivity to $FW$ was assessed by evaluating the performance of $adap$ in event $c1$ with calibrated infiltration parameter values while allowing coefficient $k$ in $FW$ (Eq. 1) to vary from 0.3 to 1.1 in steps of 0.2. The results showed $adap$ to be rather insensitive to $FW$ ($NSE$ variation between 0.91 – 0.94, $PFE$ $-1.36\% – 1.43\%$, and $VE$ $-3.99\% – -4.44\%$), justifying the decision to use Eq. (1) with $k = 0.7$ to describe the flow width in this study. However, due to the often encountered importance of proper flow width parameterization in SWMM modelling, and the possibility to calculate it explicitly in the proposed algorithm that traverses through raster cells, this should be considered as one of the first improvements to the presented algorithm.

6. Conclusions

This study presented a new algorithm for automating SWMM model construction with a novel solution to delineate subcatchments based on shared land cover and outlet. The algorithm creates subcatchments adaptively by merging small subcatchments having homogeneous land cover and common outlet into larger areas while retaining small-scale details where land cover is heterogeneous. While pre-processing the input files for the proposed tool is convenient to perform in a GIS software, the proposed tool itself is platform-independent, open-source, and not tied to any specific GIS software. The tool facilitates urban hydrological assessments by substantially reducing the required manual workload.

Based on the results obtained in this study, the following conclusions were drawn:
• The proposed algorithm facilitates rapid model construction even for large urban areas while retaining the high-resolution details where necessary.

• SWMM simulation results obtained using the proposed algorithm matched well with catchment discharge observations.

• Use of adaptive subcatchments resulted in a substantial reduction in the computational burden while yielding similar simulation results to a model having a uniformly high-resolution subcatchment delineation.

• Good model performance obtained with an existing parameter set from similar catchments conditions, adjusting only infiltration parameters, is encouraging regarding stormwater predictions in ungauged urban areas.

• The main limitation of the proposed tool is the requirement for high-resolution and high-quality land cover and DEM data.

7. Appendix A

Table A1 presents the land cover parameter values used in adap and 1x1 adopted from Warsta et al. (2017) and Krebs et al. (2014) for similar urban catchments in Finland. The Green-Ampt infiltration parameters are based on model calibration.

[TABLE A1]

8. Acknowledgements

This research is a part of the EU WaterJPI project ‘Multi-scale urban flood forecasting’ (MUFFIN). The funding was provided by Maa- ja vesiteknikan tuki ry. The tool is available from GitHub (https://github.com/AaltoUrbanWater/GisToSWMM5). Luode Consulting Oy is acknowledged for the discharge measurements. The City of Helsinki and the Helsinki Region Environmental Services Authority HSY are acknowledged for the DEM and the land cover
data (available from https://kartta.hel.fi), and for the stormwater network information. The
temperature data from the Kumpula weather station of the Finnish Meteorological Institute are
available from https://en.ilmatieteenlaitos.fi/open-data. Helsingin Seniorisäätiö and the City of
Helsinki are acknowledged for allowing rainfall measurements on their premises. Thanks are
due to Lassi Warsta for sharing his ideas about automated SWMM construction and to Ambika
Khadka for discussions regarding stormwater modelling.

9. References


have we got?” *Hydrol. Process.* 26, 2515–2521.

2593–2597.

Netherlands, pp. 1211–1217.

Hydrologic/Hydraulic Models.” *J. Hydrol. Eng.* 10.1061/(ASCE)HE.1943-5584.0000023,
596–605.

Watermark Numerical Computing, Brisbane, Australia.


Tables

Table 1. Land cover fractions (%) in the Länsi-Pakila catchment.

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Fraction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation</td>
<td>53.50</td>
</tr>
<tr>
<td>Asphalt</td>
<td>27.51</td>
</tr>
<tr>
<td>Connected roofs</td>
<td>7.87</td>
</tr>
<tr>
<td>Disconnected roofs</td>
<td>5.66</td>
</tr>
<tr>
<td>Sand and gravel</td>
<td>5.17</td>
</tr>
<tr>
<td>Water</td>
<td>0.23</td>
</tr>
<tr>
<td>Rock outcrops</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 2. Summary statistics of the studied rainfall-runoff events. Events c1-c3 were calibration and v1-v3 validation events.

<table>
<thead>
<tr>
<th>Event code</th>
<th>Date</th>
<th>Event duration (h)</th>
<th>Rainfall depth (mm)</th>
<th>Peak rain intensity (mm/min)</th>
<th>Flow volume (m³)</th>
<th>Peak flow (l/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>6 Jun 2017</td>
<td>20</td>
<td>30.0</td>
<td>0.6</td>
<td>4 321</td>
<td>540</td>
</tr>
<tr>
<td>c2</td>
<td>2 Aug 2017</td>
<td>9</td>
<td>17.6</td>
<td>0.4</td>
<td>2 567</td>
<td>368</td>
</tr>
<tr>
<td>c3</td>
<td>9 Sep 2017</td>
<td>13</td>
<td>19.8</td>
<td>0.4</td>
<td>2 882</td>
<td>364</td>
</tr>
<tr>
<td>v1</td>
<td>12 Jun 2017</td>
<td>19</td>
<td>23.2</td>
<td>0.2</td>
<td>3 263</td>
<td>309</td>
</tr>
<tr>
<td>v2</td>
<td>4 Aug 2017</td>
<td>13</td>
<td>31.4</td>
<td>1.0</td>
<td>5 368</td>
<td>509</td>
</tr>
<tr>
<td>v3</td>
<td>12 Sep 2017</td>
<td>7</td>
<td>23.6</td>
<td>1.0</td>
<td>4 035</td>
<td>615</td>
</tr>
</tbody>
</table>

Table 3. Subcatchment statistics in adap (82,554 subcatchments) and 1x1 (848,258 subcatchments) SWMM models.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>adap min</th>
<th>adap mean</th>
<th>adap max</th>
<th>1x1 min</th>
<th>1x1 mean</th>
<th>1x1 max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (m²)</td>
<td>1.0</td>
<td>10.3</td>
<td>9,322.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Elevation (m.a.s.l.)</td>
<td>19.1</td>
<td>27.5</td>
<td>45.8</td>
<td>13.4</td>
<td>27.7</td>
<td>45.8</td>
</tr>
<tr>
<td>Flow width (m)</td>
<td>0.7</td>
<td>1.4</td>
<td>67.6</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>0.2</td>
<td>5.5</td>
<td>417.7</td>
<td>0.1</td>
<td>5.5</td>
<td>464.0</td>
</tr>
</tbody>
</table>
Table 4. Performance statistics of the *adap* and the *1x1* model simulation results against observations (*obs*) and of the *adap* against the *1x1* model simulation results for the calibration (c1-c3) and the validation (v1-v3) events.

<table>
<thead>
<tr>
<th>Event</th>
<th><em>adap</em> vs. <em>obs</em></th>
<th><em>1x1</em> vs. <em>obs</em></th>
<th><em>adap</em> vs. <em>1x1</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSE(-)</td>
<td>VE (%)</td>
<td>PFE (%)</td>
</tr>
<tr>
<td>c1</td>
<td>0.92</td>
<td>−4.1</td>
<td>0.7</td>
</tr>
<tr>
<td>c2</td>
<td>0.97</td>
<td>−9.4</td>
<td>−2.2</td>
</tr>
<tr>
<td>c3</td>
<td>0.96</td>
<td>−7.7</td>
<td>−11.3</td>
</tr>
<tr>
<td>v1</td>
<td>0.80</td>
<td>−9.6</td>
<td>−5.2</td>
</tr>
<tr>
<td>v2</td>
<td>0.92</td>
<td>−16.3</td>
<td>−3.5</td>
</tr>
<tr>
<td>v3</td>
<td>0.91</td>
<td>−13.3</td>
<td>19.3</td>
</tr>
</tbody>
</table>
Table 5. Mass balance statistics of the *adap* and the *1x1* model simulation results for the calibration (c1-c3) and the validation (v1-v3) events.

<table>
<thead>
<tr>
<th>Event</th>
<th><em>P</em> (mm)</th>
<th><em>E</em> (mm)</th>
<th><em>I</em> (mm)</th>
<th><em>R</em> (mm)</th>
<th><em>S</em> (mm)</th>
<th><em>CE</em> (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1-v3</td>
<td>23.600</td>
<td>31.400</td>
<td>19.800</td>
<td>17.600</td>
<td>30.000</td>
<td>0.140</td>
</tr>
<tr>
<td>c2-v2</td>
<td>19.045</td>
<td>25.479</td>
<td>18.714</td>
<td>14.253</td>
<td>23.927</td>
<td>0.512</td>
</tr>
<tr>
<td>c3-v1</td>
<td>10.474</td>
<td>24.082</td>
<td>15.713</td>
<td>3.735</td>
<td>6.298</td>
<td>0.816</td>
</tr>
</tbody>
</table>

| *adap* | 30.000 | 10.39 | 23.927 | 4.924 | 0.133 | -0.077 |
| *1x1*  | 30.000 | 1.047 | 22.972 | 6.298 | 0.119 | -0.452 |

<table>
<thead>
<tr>
<th>Event</th>
<th><em>P</em> (mm)</th>
<th><em>E</em> (mm)</th>
<th><em>I</em> (mm)</th>
<th><em>R</em> (mm)</th>
<th><em>S</em> (mm)</th>
<th><em>CE</em> (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1-v3</td>
<td>0.141</td>
<td>0.512</td>
<td>0.249</td>
<td>0.257</td>
<td>-0.056</td>
<td>-0.086</td>
</tr>
<tr>
<td>c2-v2</td>
<td>0.141</td>
<td>0.512</td>
<td>0.249</td>
<td>0.257</td>
<td>-0.056</td>
<td>-0.086</td>
</tr>
<tr>
<td>c3-v1</td>
<td>0.141</td>
<td>0.512</td>
<td>0.249</td>
<td>0.257</td>
<td>-0.056</td>
<td>-0.086</td>
</tr>
</tbody>
</table>

Note: *P* = precipitation; *E* = evaporation; *I* = infiltration; *R* = surface runoff; *S* = final storage; *CE* = continuity error.
<table>
<thead>
<tr>
<th>Surface type</th>
<th>$I$ (%)</th>
<th>$D$ (mm)</th>
<th>$n$ (-)</th>
<th>$K_s$ (mm/h)$^a$</th>
<th>$\psi_s$ (mm)$^a$</th>
<th>$\theta_{dmax}$ (-)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt</td>
<td>100</td>
<td>0.42</td>
<td>0.011</td>
<td>24.965</td>
<td>55.832</td>
<td>0.350</td>
</tr>
<tr>
<td>Rock outcrop</td>
<td>100</td>
<td>2.49</td>
<td>0.030</td>
<td>24.965</td>
<td>55.832</td>
<td>0.350</td>
</tr>
<tr>
<td>Roof</td>
<td>100</td>
<td>0.87</td>
<td>0.012</td>
<td>24.965</td>
<td>55.832</td>
<td>0.350</td>
</tr>
<tr>
<td>Sand, gravel</td>
<td>33</td>
<td>2.49</td>
<td>0.030</td>
<td>24.965</td>
<td>55.832</td>
<td>0.350</td>
</tr>
<tr>
<td>Vegetation</td>
<td>0</td>
<td>4.22</td>
<td>0.238</td>
<td>24.965</td>
<td>55.832</td>
<td>0.350</td>
</tr>
<tr>
<td>Water</td>
<td>100</td>
<td>0.10</td>
<td>0.011</td>
<td>24.965</td>
<td>55.832</td>
<td>0.350</td>
</tr>
<tr>
<td>Concrete pipe</td>
<td>-</td>
<td>-</td>
<td>0.015</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PVC pipe</td>
<td>-</td>
<td>-</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Open channel</td>
<td>-</td>
<td>-</td>
<td>0.049</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: $I =$ imperviousness; $D =$ depression storage; $n =$ Manning’s roughness; $K_s =$ saturated hydraulic conductivity; $\psi_s =$ suction head; $\theta_{dmax} =$ maximum moisture deficit; $^a$ calibrated parameter.
Figure 2

Event | SWMM run time (min)
--- | ---
cl | 60
cl | 30
cl | 40
cl | 50
v2 | 40
v3 | 20
v1 | 10
adap | 5
adap | 10
adap | 15
adap | 20
Figure 3

- Sand and gravel
- Asphalt
- Vegetation
- Storm sewer node, open
- Storm sewer node, closed
- Rooftop, connected
- Rooftop, disconnected

Layers include:
- Subcatchment routing
- Stormwater network
- N

Scale: 0 5 10 15 20 m
Figure 4

(a) Observed flow
(b) adapt simulation
(c) $|x|I$ simulation
(d) Rainfall

Discharge (l/s)
Rainfall intensity (mm/min)

Time

03:00 05:00 07:00 09:00 11:00 13:00 15:00 17:00 19:00 21:00 23:00
04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00
Fig. 1. Land cover and layout of the stormwater network in the Länsi-Pakila catchment. Surface runoff is routed to open storm sewer nodes, representing storm drain inlets and channel inlets, whereas runoff from connected roofs is routed both into open and closed storm sewer nodes, the latter representing manholes and pipe connections.

Fig. 2. SWMM computation times (min) for the calibration and validation events using 1x1 and adap models (Desktop PC, Intel Xeon 3.20 GHz CPU, Ubuntu Linux 16.04 LTS).

Fig. 3. Comparison of subcatchments and routing between (a) 1x1 and (b) adap models for the Länsi-Pakila catchment. The arrows depicting subcatchment routing are drawn between the subcatchment mass centers.

Fig. 4. Observed (5 min moving average) and adap and 1x1 simulated discharges for the calibration events (a) c1, (b) c2, and (c) c3 and for the validation events (d) v1, (e) v2, and (f) v3.