# Performance comparison of green roof hydrological models for full-scale field sites

Ico Broekhuizen, Santiago Sandoval, Hanxue Gao, F. Mendez-Rios, Günther Leonhardt, Jean-Luc Bertrand-Krajewski, Maria Viklander

Journal of Hydrology X 12 (2021) 100093

Contents lists available at ScienceDirect

Journal of Hydrology X

journal homepage: www.sciencedirect.com/journal/journal-of-hydrology-x

Research papers

# Performance comparison of green roof hydrological models for full-scale field sites

Ico Broekhuizen [a,*], Santiago Sandoval [b], Hanxue Gao [b], Felipe Mendez-Rios [b], Günther Leonhardt [a], Jean-Luc Bertrand-Krajewski [b], Maria Viklander [a]

[a] Luleå University of Technology, Luleå, Sweden
[b] University of Lyon, INSA Lyon, France

ARTICLE INFO

*Keywords:*
Green roof
Modelling
Predictive uncertainty
Parameter identifiability
Model structure

ABSTRACT

Green roofs can be valuable components in sustainable urban drainage systems, and hydrological models may provide useful information about the runoff from green roofs for planning purposes. Various models have been proposed in the literature, but so far no papers have compared the performance of multiple models across multiple full-size green roofs. This paper compared 4 models: the conceptual models Urbis and SWMM and the physically-based models Hydrus-1D and Mike SHE, across two field sites (Lyon, France and Umeå, Sweden) and two calibration periods for each site. The uncertainty and accuracy of model predictions were dependent on the selected calibration site and period. Overall model predictions from the simple conceptual model Urbis were least accurate and most uncertain; predictions from SWMM and Mike SHE were jointly the best in terms of raw percentage observations covered by their flow prediction intervals, but the uncertainty in the predictions in SWMM was smaller. However, predictions from Hydrus were more accurate in terms of how well the observations conformed to probabilistic flow predictions. Mike SHE performed best in terms of total runoff volume. In Urbis, SWMM and Hydrus uncertainty in model predictions was almost completely driven by random uncertainty, while parametric uncertainty played a significant role in Mike SHE. Parameter identifiability and most likely parameter values determined with the DREAM Bayesian algorithm were found to be inconsistent across calibration periods in all models, raising questions about the generalizability of model applications. Calibration periods where rainfall retention was highly variable between events were more informative for parameter values in all models.

## 1. Introduction

Green roofs are an increasingly popular design feature in urban areas that can limit the increase in runoff peaks and volumes caused by urbanization (e.g. Carson et al., 2013) while providing additional benefits for e.g. air quality (e.g. Tomson et al., 2021) and building insulation (e.g. Susca, 2019). The hydrology of green roofs has been studied both in full-scale roofs (for overviews see e.g. Carson et al., 2013; Cipolla et al., 2016; Mentens et al., 2006) and in purpose-built test beds (e.g. Stovin et al., 2012). Factors that influence the hydrologic performance include substrate properties (Berretta et al., 2014; Fassman-Beck et al., 2013; Szota et al., 2017), evapotranspiration (Cascone et al., 2019; Ebrahimian et al., 2019), roof slope (Getter et al., 2007), vegetation properties (Berretta et al., 2014) and roof dimensions (Hakimdavar et al., 2014). For example, Liu et al. (2019) found, based on irrigation lab experiments, that substrate material was the most important factor influencing the runoff retention in green roofs, followed by substrate depth, roof slope, and vegetation.

Planning of new green roofs and drainage systems that include green roofs requires predictions of roof runoff (e.g. Palla and Gnecco, 2020; Versini et al., 2015). This interest has led several authors to propose models of varying complexity and levels of detail. For specific green roofs, regression equations (e.g. Carson et al., 2013; Fassman-Beck et al., 2013) or unit hydrographs (Villarreal and Bengtsson, 2005) can be developed to quantify their long-term performance on either an event basis or for longer periods (Mentens et al., 2006). The downside of such approaches is that they remain site specific and thus might lack generality. Such simplified rainfall-runoff relationships may not be directly applicable in forecasting contexts.

A different approach is represented by conceptual models, i.e. simplified descriptions retaining a physical basis, which have been

* Corresponding author at: Laboratorievägen 14, 971 87 Luleå, Sweden.
*E-mail address:* ico.broekhuizen@ltu.se (I. Broekhuizen).

https://doi.org/10.1016/j.hydroa.2021.100093
Received 24 February 2021; Received in revised form 12 July 2021; Accepted 14 July 2021
Available online 21 July 2021

**Nomenclature**

*List of symbols*

| Symbol | Description |
|---|---|
| $A$ | Roof area [$L^2$] |
| $A_{cell}$ | Area of a single cell in the SHE model domain ($L^2$) |
| $A_{reg}$ | Coefficient controlling drainage flow (Urbis) [$L^2$] |
| $B_{dr}$ | Fraction of potential evapotranspiration made available to the drainage layer (Urbis) [–] |
| $B_{season}$ | Seasonal adjustment factor for evapotranspiration (Urbis) [–] |
| $C_{sub}$ | Maximum water storage capacity of the substrate layer (Urbis) [L] |
| $d_{sur}$ | Current depth of water on surface [L] |
| $d_{dr}$ | Current depth of water in the drainage layer [L] |
| $D_{dr}$ | Thickness of the drainage layer [L] |
| $D_{sur}$ | Surface depression storage [L] |
| $D_{sub}$ | Thickness of the substrate layer [L] |
| $E_{dr}$ | Actual evapotranspiration from the drainage layer [$L\ T^{-1}$] |
| $E_{pot}$ | Potential evapotranspiration rate [$L\ T^{-1}$] |
| $E_{sub}$ | Actual evapotranspiration from the substrate layer [$L\ T^{-1}$] |
| $E_{sur}$ | Actual evapotranspiration from the surface [$L\ T^{-1}$] |
| $E_{tot}$ | Total actual evapotranspiration [$L\ T^{-1}$] |
| $g$ | Gravitational acceleration, $g = 9.81\ m\ s^{-2}$ |
| $H$ | Sink term in Richards' equation [$T^{-1}$] |
| $I$ | Infiltration rate into the substrate [$L\ T^{-1}$] |
| $J$ | Accumulated infiltration volume during a rainfall event (SWMM) [L] |
| $K$ | Hydraulic conductivity [$L\ T^{-1}$] |
| $K_s$ | Saturated hydraulic conductivity [$L\ T^{-1}$] |
| $L$ | Flow of water from substrate to drainage layer [$L\ T^{-1}$] |
| $n$ | Manning's coefficient for surface flow (SWMM) [$T\ L^{-1/3}$] |
| $n_{dr}$ | Manning's coefficient for drainage layer flow (SWMM) [$T\ L^{-1/3}$] |
| $n_{she}$ | Manning's coefficient for surface flow (SHE) [$T^{-1}\ L^{1/3}$] |
| $N$ | Number of observations |
| $M$ | Roof slope [$L\ L^{-1}$] |
| $P$ | Precipitation rate [$L\ T^{-1}$] |
| $Q_{sur}$ | Surface runoff [$L^3\ T^{-1}$] |
| $Q_{dr}$ | Drainage layer outflow [$L^3\ T^{-1}$] |
| $Q_{ov}$ | Overflow from full drainage layer [$L^3\ T^{-1}$] |
| $Q_{obs}$ | Observed roof runoff [$L^3\ T^{-1}$] |
| $Q_{sim}$ | Simulated roof runoff [$L^3\ T^{-1}$] |
| $Q_{stoch}$ | Simulated roof runoff with added random error [$L^3\ T^{-1}$] |
| $R_{dr}$ | Coefficient controlling drainage flow (SHE) [$T^{-1}$] |
| $R_U$ | Coefficient controlling drainage flow (Urbis) [-] |
| $S_{eff}$ | Effective saturation [$L^3\ L^{-3}$] |
| $t$ | Time [T] |
| $V$ | Averjanov coefficient (SHE) [–] |
| $W$ | Roof width [L] |
| $z$ | Vertical coordinate [L] |
| $\alpha$ | Van Genuchten parameter [$L^{-1}$] |
| $\beta$ | Van Genuchten parameter [–] |
| $\gamma$ | Coefficient controlling change in K as function of $\theta$ (SWMM) [–] |
| $\eta$ | Coefficient controlling relationship between water depth in the drainage layer and actual evapotranspiration (Urbis). |
| $\theta$ | Current substrate volumetric water content [$L^3\ L^{-3}$] |
| $\theta_{fc}$ | Volumetric water content of substrate at field capacity [$L^3\ L^{-3}$] |
| $\theta_r$ | Residual volumetric water content [$L^3\ L^{-3}$] |
| $\theta_s$ | Volumetric water content of substrate at saturation [$L^3\ L^{-3}$] |
| $\theta_{top}$ | Volumetric water content of top layer of the substrate [$L^3\ L^{-3}$] |
| $\theta_{wp}$ | Volumetric water content of substrate at wilting point [$L^3\ L^{-3}$] |
| $\lambda$ | Van Genuchten parameter [–] |
| $\mu$ | Van Genuchten parameter [–] |
| $\sigma$ | Standard deviation of the observation-simulation error distribution [$L^3\ T^{-1}$] |
| $\sigma_0$ | Intercept of the heteroscedastic error model [$L^3\ T^{-1}$] |
| $\sigma_1$ | Slope of the heteroscedastic error model [$L^{1/2}\ T^{-1/2}$] |
| $\sigma_Q$ | Standard deviation of ensemble of flow values [$L^3\ T^{-1}$] |
| $\phi_{dr}$ | Void fraction of the drainage layer [$L^3\ L^{-3}$] |
| $\phi_{sur}$ | Void fraction of the surface layer [$L^3\ L^{-3}$] |
| $\psi$ | Soil pressure head [L] |
| $\psi_E$ | Threshold pressure head for evapotranspiration in Hydrus [L] |
| $\psi_{init\_bot}$ | Initial pressure head at bottom of substrate (Hydrus) [L] |
| $\psi_{GA}$ | Green-Ampt suction head (SWMM) [L] |
| $\Omega_{spread}$ | Relative spread of model prediction ensemble [–] |
| $\Omega_{rel}$ | Reliability of model prediction ensemble [–] |
| $\mathscr{L}$ | Log-likelihood [–] |

proposed by several authors in the green roof context (e.g. Carbone et al., 2014; Herrera et al., 2018; Kasmin et al., 2010; Li et al., 2019; Locatelli et al., 2014; Rossman and Huber, 2016; She and Pang, 2010; Sherrard and Jacobs, 2012; Skala et al., 2019; Vesuviano et al., 2014). Some of these modelling approaches mainly focused on sub-models for specific parts or processes of green roofs, such as the drainage layer (Vesuviano and Stovin, 2013) and evapotranspiration (Feng and Burian, 2016; Jahanfar et al., 2018).

One commonly used model for green roofs is the U.S. E.P.A. Storm Water Management Model (SWMM), which has had official support for green roofs since version 5.1 (Rossman and Huber, 2016), although green roofs could be simulated in older versions using the curve-number approach or representing the green roof as a storage node in the drainage network (Alfredo et al., 2010). Attempts to model green roofs using the SWMM bioretention cell module were successful in some cases (Burszta-Adamiak and Mrowiec, 2013) but not in others (Versini et al., 2015). Different studies have confirmed that the SWMM green roof module added in version 5.1 is applicable to individual rainfall events (e.g. Hamouz and Muthanna, 2019; Palla et al., 2012) and over longer periods (e.g. Cipolla et al., 2016; Hamouz and Muthanna, 2019; Krebs et al., 2016). However, further studies have also raised concerns about this software, regarding e.g. the description of evapotranspiration (Feng and Burian, 2016; Hamouz and Muthanna, 2019; Peng and Stovin, 2017), the inability of calibrated parameters to be transferred to similar roofs in different locations (Johannessen et al., 2019), and the use of a large number of parameters to which the modelling outputs are insensitive (Leimgruber et al., 2018). Similar considerations are likely to apply to some other conceptual models as well, but have rarely been investigated in detail or reported.

Another type of descriptions used for green roof modelling are physically-based models, mainly based on the Richards equation, operating most commonly in one (i.e. the vertical) dimension (e.g. Avellaneda et al., 2014; Hilten et al., 2008; Peng et al., 2019; Sims et al., 2019; Skala et al., 2020), although two dimensional (Li and Babcock, 2015; Palla et al., 2009) or three dimensional (Brunetti et al., 2016) descriptions have also been reported. One frequently-named advantage of physically-based models is that parameter values represent physical characteristics which can be measured in situ (e.g. De-Ville et al., 2018b;

Sims et al., 2019) or in the lab (e.g. Brunetti et al., 2016; Peng et al., 2020). However, the representativeness of parameter values measured in the laboratory under controlled conditions has been questioned (Szota et al., 2017). This finding raises concerns about whether laboratory measurements of model parameters may also be applicable for roofs that have already been built. Another concern is that different models may interpret the same physical parameter in different ways in the model equations.

An ideal model would be capable of reproducing the runoff from green roofs with different designs, under different climatic conditions and for all times of the year. In addition, in an ideal situation models would be able to reliably predict runoff from green roofs without calibration, e.g. when considering the implementation of new green roofs. Comparing multiple models using the same dataset may be one way of generating insights into shortcomings of and potential improvements to model structures that might help achieve this goal. Some studies have tried to identify the best-performing models by applying multiple models to the same dataset. Carson et al. (2017) found that the (empirical) curve number method and characteristic runoff equation performed slightly better (0.03 difference in Nash-Sutcliffe Efficiency – NSE) than SWMM for two full-scale roofs. However, the site-specific nature of these first two models is less appealing in practice, and more complex models were not considered. Palla et al. (2012) performed event-based calibrations of a 16 × 22 m roof, comparing a simple conceptual model (consisting of three linear reservoirs) with the mechanistic model Hydrus-1D (Šimůnek et al., 2008). Both models closely reproduced the observed runoff hydrographs, but the mechanistic model's predictions were somewhat more accurate. By contrast, Soulis et al. (2017) reported better performance (for 2 × 1 m test modules) for the conceptual SWMM model than for Hydrus-1D. Xie et al. (2020) performed irrigation experiments on laboratory modules and found that Hydrus-1D obtained a higher NSE than SWMM in calibration, but worse scores for testing periods (with higher rainfall intensities), regarding runoff volumes and peak flows. Brunetti et al. (2020) performed a detailed calibration study of a month-long laboratory irrigation experiment with a 57 × 36 cm green roof module. They found that a conceptual model and a relatively simple mechanistic model were equally suited to describe the hydrological behaviour of the test module, while the use of more complex mechanistic models was not justified given the limited improvements in performance.

Although these studies contributed to elucidate some aspects of modelling green roofs, they might still present certain limitations. First, only one of them (Palla et al., 2012) is based on an actual green roof and not purpose-built lab or pilot modules, dismissing aspects of real green roofs such as scale effects, heterogeneity and non-vertical water movement. Second, the simulations were most commonly event-based descriptions (Palla et al., 2012; Xie et al., 2020), ignoring the processes that take place in the roof during dry periods and being sensitive to initial conditions. Brunetti et al. (2020) did use a continuous simulation, but their irrigation experiment only used block rainfalls of identical intensity which is not representative of the variability in natural rainfall and runoff from green roofs. Third, the mentioned comparative studies (Brunetti et al., 2020; Palla et al., 2012; Soulis et al., 2017) were all carried out under a Mediterranean climate (except Xie et al. (2020) who only simulated different rain intensities and ignored evapotranspiration), so their conclusions might not be applicable to other climates. Fourth, a number of the previous studies present a limited assessment of uncertainties in model parameters and predictions, which might also have a significant influence in the conclusions thereby reported.

Hydrological simulations of green roofs provided by reliable models seem to be an essential tool for the operation and design of these structures, which ultimately encourages more sustainable urban drainage systems. Therefore, the goal of this study is to propose a comparative evaluation of models of varying complexity (from simple conceptual to mechanistic ones), applied to two full-scale green roofs with different physical layouts and in two different climates. The models are compared under a context of continuous simulation, assessing modelling uncertainties and accuracy by means of Bayesian methods and monitored data of rainfall, potential evapotranspiration and output flow rate from different time periods. A formal Bayesian calibration approach is used to estimate values of model parameters that fit best with the observed flow data (including parameters that are not directly linked to observable characteristics of the green roof system or may be difficult to determine in-situ) while evaluating the inherent uncertainties in such estimates and the resulting uncertainties in the model predictions.

## 2. Methods

### 2.1. Green roofs & data sets

For this study two green roofs were selected, one in Lyon, France, and one in Umeå, Sweden, representing different designs and different climate conditions.

The green roof in Lyon was built in the Congress Centre of Lyon, France in 1995 and covers 280 $m^2$. The roof is flat and consists of a 50 mm egg-box shaped drainage layer underneath the substrate layer, which has a depth varying between 40 and 140 mm and is planted with sedum. Outflows from both the substrate and the drainage layer join in a single drain in which discharge was measured using an electromagnetic flow meter (Krohne Optiflux 2000), allowing to measure flow rates up to 330 L $min^{-1}$. The specified accuracy of the flow meter depends on the flow value and varies from 2.9% at 10 L $min^{-1}$ to 1.0% at 30 L $min^{-1}$ (i.e. for typical low and high flow rates for this roof). Rainfall was measured using a 0.2 mm tipping bucket rain gauge installed on the roof. Flow rate and rainfall measurements were reported with a constant one-minute time step from September 2012 to May 2013 (Bertrand-Krajewski and Vacherie, 2014). Other meteorological data (temperature, wind speed, atmospheric pressure, solar radiation and relative humidity) were provided by the French national meteorological office (Météo France), from the Bron weather monitoring station located 9 km from the green roof. The 2012–2013 green roof monitoring campaign was carried out for other objectives than model calibration and no local measurements of potential evapotranspiration were available. The Bron station is the closest station where daily measurements of potential evapotranspiration ($E_{pot}$) were also available. Therefore, $E_{pot}$ values were initially estimated on an hourly basis from temperature, wind speed, atmospheric pressure, solar radiation and relative humidity records, employing the Penman-Monteith equation (e.g. Zotarelli et al., 2010). These hourly $E_{pot}$ values were further corrected to sum to the Meteo France available total daily $E_{pot}$ values considered as reference values. Afterwards, hourly $E_{pot}$ corrected values were linearly interpolated to a one-minute time step (Principato, 2015).

The second roof was built in Umeå, Sweden in 2014 and covers 2469 $m^2$. A 25 mm drainage layer underlies a substrate layer whose depth varies between 42 and 80 mm with an average of 60 mm, planted with a mixture of grass and sedum. The original design depth of the substrate was 70 mm (BG Byggros AB, 2018a). The roof consists (see Fig. 1) of 7 V-shaped sections (i.e. 14 sloped sections in total, with half facing due north, half facing due south) that are angled between 15° and 38°. In the narrow section of the roof, a drain (receiving flow from both the surface and the drainage layer) is installed on either side of the roof at the bottom of the V; in the wider section drains are installed at ¼ and ¾ of the total width, each receiving flow from half of the V-shaped section. A network of pipes collects the water from these roof sections and from two small impervious roofs (169 $m^2$ in total). Flow was measured using a 1 L tipping bucket for flow rates up to 25 L $min^{-1}$ (HyQuest, 2016) and an area-velocity flow meter for higher flow rates (Teledyne ISCO, 2010). The accuracy of the tipping bucket gauge was not quantified precisely, but based on a flow-rate dependent correction from the manufacturer (HyQuest, 2016) and in-situ checks of the volume per tip it is believed to be reasonably small. Previous laboratory tests of this model of area-

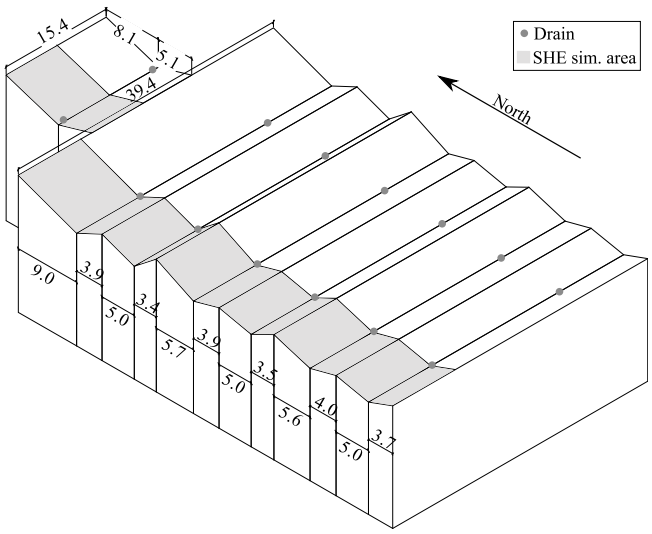


**Fig. 1.** Dimensions of the Umeå green roof, all measures in m. The part of the roof represented in Mike SHE simulations is indicated, the resulting flows were multiplied by 2 (narrow part of the roof) and 4 (wide part of the roof) to obtain flow rates for the whole roof. In SWMM the entire roof was simulated, while in Urbis and Hydrus (1-D models) the resulting flows given per $m^2$ were multiplied by the area of the whole roof.

velocity sensor showed that the accuracy of the flow measurement is typically within ±5% in the flow range of 120–540 L $min^{-1}$. (Broekhuizen, 2021). Rainfall was measured using a 0.2 mm tipping bucket. Measurements were carried out from late 2017 to early November 2020, although some gaps were present in the data. In addition, periods influenced by snow fall or snow melt (occurring as early as October, and as late as May) were excluded since they pose additional challenges outside the scope of this paper (e.g. Hamouz and Muthanna, 2019). Evapotranspiration was calculated using the Penman-Monteith equation and data from two stations from the Swedish Meteorological and Hydrological Institute (station 140,615 for solar radiation (approx. 2.7 km from the green roof) and 140,480 for other data (approx. 1.1 km from the green roof)).

For both sites, two independent one-month long periods were selected to calibrate and test different green-roof models. Independent calibrations were undertaken for each period and then cross-tested against the data for the other period. This was done to test whether a model calibrated for a given period is usable for other periods with potentially different meteorological conditions. For each site, one period was selected that showed high variability among the retention rates for individual events, while the other had similar total rainfall depth but with less varied retention. For Lyon, the selected periods were November 2012 (reporting low retention rates with a low inter-event retention variability) and April 2013 (with high variability of the inter-event retention rates). For Umeå, the selected periods were from 9 September 2020 to 11 October 2020 (reporting a high inter-event retention variability; hereafter referred to as just September) and from 29 June 2020 to 28 July 2020 (with low inter-event retention variability; hereafter July).

When analysing monthly time series including data from dry periods, zero-registered rainfall and flow-rate values were found to represent the majority of the values. However, the inclusion of these zero-registered values in the model calibrations might tend to favour the representation of low flow-rate values, penalizing the peaks of the hydrographs (e.g. Oliveira et al., 2018). In addition, the predictive capabilities of the models could also be overestimated during these dry periods, since almost any model will predict no flow if it has not rained for some time. On the other hand, close-to-zero flow rate values during rainy periods can also represent valuable information about the model performance, as models or parameter setups that predict flows when none occurred are not desirable. Therefore, flow rate values lower than the detection limit of the measurement devices (0.79 and 2 L $min^{-1}$ for Lyon and Umea, resp.) were not considered in the calculations, as long as no rainfall was registered within the preceding 15 min.

### 2.2. Hydrological models

Four models of varying complexity were selected for this study. The model structures are shown visually in Fig. 2 and described in more detail in the sections below.

All models used in the study assume that flow through the substrate is vertical, with lateral flow occurring on the surface and in the drainage layer. The slopes of the roof in Umeå may give rise to some lateral flow in the substrate. However, this is believed to play at most a minor role for a number of reasons. First, water input (precipitation) and output (drainage) from the substrate take places uniformly across the entire roof. Second, the permeability of the geotextile separating the substrate from the drainage layer (90 mm $hr^{-1}$ (BG Byggros AB, 2018a, BG Byggros AB, 2018b)) is higher than the observed rainfall intensities, so the formation of a saturated layer on top of the geotextile (i.e. in the bottom of the substrate) is unlikely. Third, the substrate is very thin compared to the length (in the direction of the slope) of the roof segments so water moving down through the substrate has little time to travel laterally before reaching the bottom of the substrate and the drainage layer. Fourth, the thin substrate means that should the drainage layer capacity be exceeded, the substrate will quickly become saturated at which point drainage outflow and surface runoff will be larger than lateral flow through the substrate. For these reasons, the assumption of only vertical flow in the substrate is considered appropriate. In addition, this is also the assumption most commonly used in green roof modelling studies, and it is therefore appropriate here given that the study goal is to test existing models.

#### 2.2.1. Urbis

Urbis is a scenario simulation and decision support tool aimed at simulating long-term hydrological dynamics of different stand-alone and interconnected Stormwater Control Measures (SCMs), such as green roofs, swales, retention basins or permeable surfaces, at the block and building scales. This free software (https://deep.insa-lyon.fr/fr/content/urbis) was developed in the Python programming language, by INSA Lyon, France in collaboration with the companies Nidaplast, Siplast, and AS2C (Sandoval et al., 2019). The model is based on conceptual simplifications of these systems into a set of connected boxes or reservoirs that represent e.g. the storage zone or the substrate of each structure. Focusing on the case of green roofs, the option "alveolar flat green roof" was chosen in Urbis to model both the Lyon and Umeå green roofs. Urbis uses the term alveolar to refer to the egg box like layer of plastic compartments forming the bottom of the roof that allow for some storage volume and for drainage; SWMM uses the term drainage layer. Urbis was selected as an example of a simple conceptual model, because computer code for it was already available and the authors were already familiar with it, helping to avoid errors in the model setup. Other simple conceptual models have been proposed (see Section 1) and could be used as well, but given the large number of available models it was not feasible to compare all models here.

In the following lines a brief description is given of this "alveolar flat green roof" modelling setup in Urbis. The potential evapotranspiration is calculated according to:

$$E_{tot} = B_{season} E_{pot} \quad (1)$$

where $E_{tot}$ is the total actual evapotranspiration, $E_{pot}$ is a user-supplied potential evapotranspiration time series (e.g. calculated using the Penman-Monteith equation) and default values of the coefficient $B_{season}$ are 0.5 for winter and 1 for summer. Evapotranspiration from the

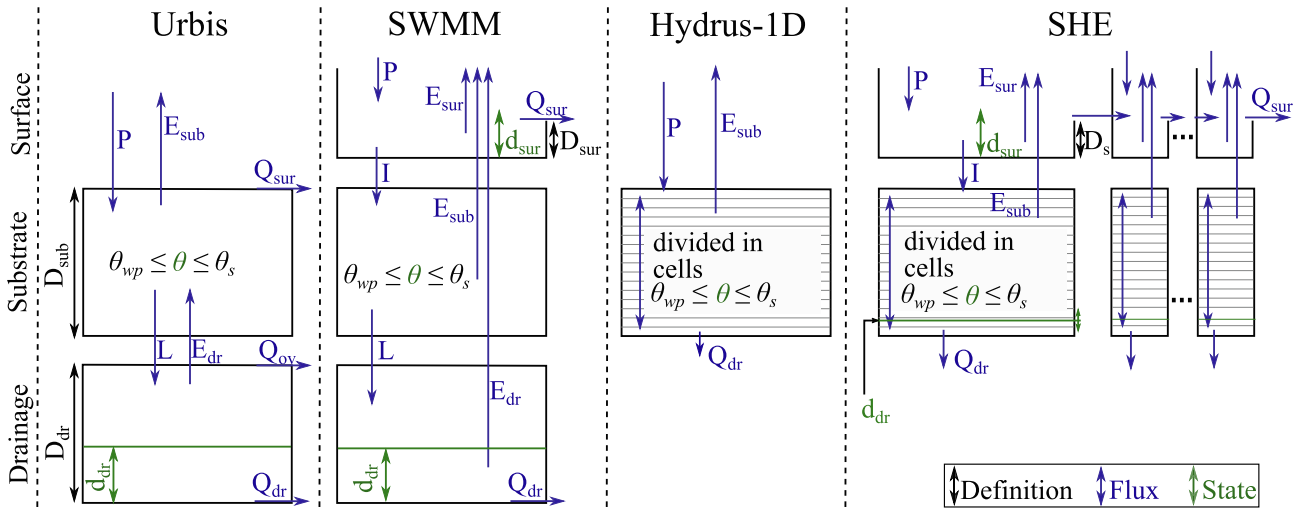


**Fig. 2.** Schematics of the models used in this study. The symbols are defined in the list of symbols at the beginning of the article. Note that for SHE two dimensions (the vertical and one horizontal dimension) are shown, but the model actually operates in three dimensions (vertical + two horizontal), and with more cells than shown here.

alveolus layer to the substrate is calculated from the following equations:

$$E_{dr} = \min\left(B_{dr}E_{tot}, d_{dr}\frac{3600}{\Delta t}\right) \tag{2}$$

$$B_{dr} = \left(\frac{d_{dr}}{\phi_{dr}D_{dr}}\right)^{\eta} \tag{3}$$

where $B_{dr}$ is a fraction of the potential evapotranspiration that is available for the drainage (alveolus) layer, $d_{dr}$ is the current water stored in the alveolus layer (since alveolus layers in Urbis are similar to drainage layers in SWMM, the subscript $dr$ is used for both), $D_{dr}$ is the drainage layer thickness (50 mm for Lyon and 25 mm for Umeå) and $\phi_{dr}$ is the void index of the layer structure, set as a calibration parameter for the case of Lyon and as a fixed value of 0.54 for Umeå based on available information from the manufacturer (BG Byggros AB, 2019). The coefficient η describes the relationship between water level in the alveolus layer and movement of water from the alveolus layer to the substrate, see next paragraph.

Furthermore, evapotranspiration from the substrate ($E_{sub}$) to the atmosphere is calculated according to:

$$E_{sub} = \begin{cases} E_{dr} & d_{dr} > 0 \\ \min\left(E_{tot}, \theta D_{sub}\frac{3600}{\Delta t}\right) & d_{dr} = 0 \end{cases} \tag{4}$$

where $D_{sub}$ is the substrate depth and $\theta$ the current water content of the substrate. $D_{sub}$ was set to 60 mm for Umeå and it was established as a calibration parameter for the green roof in Lyon, given the high variability of this value reported from in situ measurements. For the studied green roofs, no evapotranspiration or re-humidification occurs from the alveolus layer to the substrate and no storage exists in the drainage layer during dry periods. This was explicitly considered in Urbis by setting $\eta = 10000$ (a very high value) to make $B_{dr}$ equal to zero (Eq. (3)), leading to set$E_{dr} = 0$ (Eq. (2)). Indeed, this conduces to cancel the water transfer from the alveolus (drainage) layer to the substrate, leading to estimations of $E_{sub}$ (Eq. (4)) only dependent on the potential evapotranspiration and the substrate water content $\theta$. Eq. (4) allows evapotranspiration to occur up to the potential rate, but reducing when the substrate moisture contents becomes low.

The maximum storage capacity of the substrate $C_{sub}$ is calculated as follows:

$$C_{sub} = (\theta_{fc} - \theta_{wp})D_{sub} \tag{5}$$

where $\theta_{fc}$ and $\theta_{wp}$ are the field capacity and wilting point of the substrate, respectively. The wilting point was set to Urbis default value $\theta_{wp} = 0.05$ and $\theta_{fc}$ was established as a calibration parameter for both green roofs. If inflow into the substrate exceeds $C_{sub}$ (saturation of the substrate), the excess water $L$ from the saturated substrate, transferred to the alveolus (drainage) layer, is calculated as follows:

$$L = \max\left(\theta D_{sub}\frac{3600}{\Delta t} + P - E_{sub} + E_{dr} - C_{sub}\frac{3600}{\Delta t}, 0\right) \tag{6}$$

If the maximum storage capacity of the alveolus layer is reached, the excess becomes overflow ($Q_{ov}$), and if the maximum storage of the substrate is reached, any excess water on the surface of the substrate becomes runoff ($Q_{sur}$). This alveolar structure can also include a controlled outflow, aimed to simulate the output flows from the roof delivered to the pipe system. This flow rate $Q_{dr}$ is then calculated by an orifice equation as follows:

$$Q_{dr} = R_U \cdot A_{reg}\sqrt{2gd_{dr}} \tag{7}$$

where $R_U$ is a coefficient (set to 0.6 by default in Urbis). The area of the regulation orifice $A_{reg}$ was set as a calibration parameter because $A_{reg}$ is normally calculated in the software based on the design flow rate to be delivered from the roof to the drainage system, which was unknown in this case. It was not possible to calculate $A_{reg}$ directly since conceptually the value of the parameter refers to an orifice in the wall of a tank while in the actual system the opening was a drainage outlet in the bottom of the roof. The regulation orifice is set to be located at the bottom of the alveolar layer in order to avoid stormwater storage in the drainage layer during dry periods, as neither roof was designed for this purpose. The roof runoff that was studied in this paper is the sum of $Q_{dr}$, $Q_{ov}$ and $Q_{sur}$.

### 2.2.2. SWMM

SWMM (v 5.1.013; Rossman and Huber, 2016) is a conceptual model that still relies on some physically-based parameters and process descriptions. SWMM is commonly used in both research and practice and it is therefore of particular interest to evaluate its capabilities. A short description of the main process equations is given below, for details see Rossman and Huber (2016). Infiltration is represented using the Green-Ampt model (Green and Ampt, 1911):

$$I = K_s\left(1+\frac{(\theta_s-\theta_{top})(d_{sur}+\psi_{GA})}{J}\right) \quad (8)$$

where $K_s$ is the saturated hydraulic conductivity, $\theta_s$ the saturated water content, $\theta_{top}$ the water content of the top layer of the soil, $d_{sur}$ the depth of water on the surface,$\psi_{GA}$ the soil suction head at the wetting front, and $J$ the accumulated infiltration volume during the rainfall event. The percolation of water from the substrate layer to the drainage layer is described by Darcy's law:

$$L = K_s e^{-\gamma(\theta_s-\theta)} \quad (9)$$

where $\gamma$ is a calibration parameter describing the change of conductivity as a function of soil moisture content and $\theta$ is the current substrate moisture content. Percolation is disabled if $\theta \leq \theta_{fc}$. Surface runoff is described using Manning's equation:

$$Q_{sur} = \frac{1}{n}\sqrt{M}W\phi_{sur}(d_{sur}-D_{sur})^{\frac{5}{3}} \quad (10)$$

where $n$ is Manning's roughness coefficient, $M$ and $W$ the slope and width of the roof, $\phi_{sur}$ the void fraction of the surface (i.e. fraction of volume not filled by plants), and $D_{sur}$ the depression storage. Flow in the drainage layer is also described using Manning's equation:

$$Q_{dr} = \frac{1}{n_{dr}}\sqrt{M}\mathrm{W}\phi_{dr}d_{\mathrm{dr}}^{\frac{5}{3}} \quad (11)$$

with $n_{dr}$ a separate Manning's coefficient, $\phi_{dr}$ the void fraction of the drainage layer, and $d_{dr}$ the water depth in the same layer. Evapotranspiration takes place at the user-supplied potential rate $E_{pot}$, using first any water available on the surface (the evapotranspiration is labelled $E_{sur}$), then from the substrate (labelled $E_{sub}$), and finally from the drainage layer (labelled $E_{dr}$). The roof runoff that was studied in this paper is the sum of $Q_{dr}$ and $Q_{sur}$.

The roof in Lyon was represented as a single green roof. For Umeå the individual roof sections were included as individual sub-catchments with individual green roofs, so that the dimensions and slope of each section could be input directly. The pipe network was included explicitly in the SWMM model, although it had a negligible influence on the output of the model. The small impervious roof areas in Umeå were simulated only in SWMM. The calculated flow was added to the total outflow from the roof in the other models. In this way, the impermeable roof sections were treated the same way in all models and not subject to any calibration, following the scope of the paper.

#### 2.2.3. Hydrus-1D

Hydrological modelling of the studied green roofs is also simulated by using the Hydrus-1D model (Šimůnek et al., 2018). Two- and three-dimensional versions of Hydrus also exist, but in this paper we refer to Hydrus-1D as simply "Hydrus". This model describes one-dimensional vertical water movement using Richards equation:

$$\frac{\partial\theta(\psi)}{\partial t} = \frac{\partial}{\partial z}\left[K(\psi)\left(\frac{\partial\psi}{\partial z}+1\right)\right]-H \quad (12)$$

where $K(\psi)$ is the unsaturated hydraulic conductivity as a function of the tensiometer pressure potential $\psi$; $z$ is the vertical coordinate (positive upward); and $H$ is a sink term representing evaporation and root water uptake.

The unsaturated soil hydraulic properties can be simulated in Hydrus with six different models. In this study, the Van Genuchten and Mualem model is adopted (van Genuchten, 1980), as in similar green roof modelling settings (e.g. Brunetti et al., 2020). The Van Genuchten relationships can be written as:

$$\theta(\psi) = \begin{cases} \theta_r + \dfrac{\theta_s-\theta_r}{\left[1+|\alpha\psi|^{\beta}\right]^{\mu}} & \psi < 0 \\ \theta_s & \psi \geq 0 \end{cases} \quad (13)$$

$$K(\psi) = K_s S_{eff}^{\lambda}\left[1-\left(1-S_{eff}^{1/\mu}\right)^{\mu}\right]^2 \quad (14)$$

$$S_{eff} = \frac{\theta-\theta_r}{\theta_s-\theta_r} \quad (15)$$

where $\theta_r$ and $\theta_s$ are the residual and the saturated water content of the soil, $K_s$ is the saturated hydraulic conductivity and $S_{eff}$ is the effective saturation,$\alpha$ (related to the inverse of the air entry suction), β (pore size distribution), $\lambda$ (tortuosity) and $\mu = 1–1/\beta$ (dimensionless) are fitting parameters of the soil–water retention curve. Coarse-grained soils have a high value of $\alpha$ and $\beta$, whereas fine-textured soils have a lower α and $\beta$ (Rassam et al., 2018). $\beta$ is used instead of the more common symbol $n$ to avoid confusion with Manning's coefficient.

At the surface of the substrate the boundary condition (BC) "Atmospheric BC with surface run off" was applied where observed time series of precipitation and evapotranspiration are used by the model. $\psi_E$ denotes the minimum surface pressure head at which actual evapotranspiration can still take place at the potential rate $E_{pot}$; below this value soil moisture availability limits the actual evapotranspiration. $\psi_E$ was set to a fixed value of −300,000 mm (Brunetti et al., 2016), since evapotranspiration parameters were not calibrated in any of the other models. A "constant pressure head" BC was applied at the bottom of the substrate. For this case, the pressure head at the bottom node $\psi_{init\_bot}$ is set constant and equal to the value given by the initial conditions at the bottom of the substrate. This $\psi_{init\_bot}$ value was included as a calibration parameter.

It is worth clarifying that alternative BCs at the bottom of the substrate have been applied for modelling green roofs with Hydrus, namely the "free drainage" (Palla et al., 2012; Hakimdavar et al., 2014) or "seepage face" (Brunetti et al., 2016) BCs. The constant pressure head BC was selected because it was found to have better performance compared to the two mentioned alternatives by Peng et al. (2019) and in initial calibration runs using the data sets from this paper.

The initial conditions for every model run were linearly interpolated from a typical pressure head of −3000 mm (Brunetti et al., 2016) at the top of the substrate, to the provided $\psi_{init\_bot}$ parameter value at the bottom. This approach aimed to maintain similar initial conditions for all model runs, and to avoid numerical problems related to zero-valued or abruptly changing vertical gradients. In addition, a 14-day warming-up period was applied to reduce furthermore the effect of these initial conditions on the results. The Python library phydrus (phydrus.readthedocs.io) was used as interface to Hydrus in the calibrations (Collenteur et al., 2020). The egg box-like drainage layers in the roofs do not have a suitable physical representation in Hydrus, so it was not included for this model in any of the studied roofs.

### 2.3. Mike SHE

The physically-based model Mike SHE (version 2019, not updated; DHI, 2019a, 2019b) was selected as a more extensive mechanistic model (newer versions of Mike SHE were not used since they produced identical results but with approximately double model runtimes). The model uses a regular rectangular grid (1 m grid spacing was used here) and vertically discretizes the unsaturated zone into a user-specified number of cells (here 1 cm cells were used). Overland flow across the grid cells is described using the Saint-Venant equations (using the diffusive wave approximation) where Mannings coefficient describes the friction from the surface (note that Mike SHE uses the inverse of Manning's $n$ used in SWMM). Richards equation (see Eq. (12)) is used to describe vertical flow in the unsaturated zone. Different options are available for the

hydraulic conductivity function. For reasons of numerical stability, the Averjanov equation was used here:

$$K(\theta) = K_s\left(\frac{\theta-\theta_r}{\theta_s-\theta_r}\right)^V \quad (16)$$

where $\theta_r$ is the residual soil water content and $V$ a calibration coefficient ($V$ is used instead of the more common symbol $n$ to avoid confusion with Manning's coefficient). The relation between $\psi$ and $\theta$ is expressed by calibrating saturated moisture content $\theta_s$ (corresponding to a pressure head $\psi = -10$ mm) and $\theta_{fc}$ ($\psi = -10^3$ mm) while the wilting point moisture content $\theta_{wp}$ ($\psi = -10^{4.2}$ mm) was fixed at 0.1 and the residual moisture content $\theta_r$ ($\psi = -10^6$ mm) at 0.02. Intermediate values were linearly interpolated between adjacent points. The use of classical hydraulic conductivity curves (e.g. van Genuchten, 1980) was avoided since first calibration runs showed that this sometimes led to failed model runs or very long model run times which made it infeasible to perform calibrations in a reasonable amount of time.

The egg box-like drainage layers in the roofs do not have a suitable physical representation in Mike SHE (initial attempts to model the drainage layers using SHE's saturated zone module were unsuccessful). A simple conceptual drainage function is thus used:

$$Q_{dr} = d_{dr}R_{dr}A_{cell} \quad (17)$$

where $d_{dr}$ is the hydraulic head in the saturated zone and $R_{dr}$ a (calibrated) constant and $A_{cell}$ the horizontal area of a single cell in the model domain. Drainage is calculated for each cell individually (since $d_{dr}$ varies spatially) and then summed to obtain the drainage outflow for the entire model domain. The roof runoff that was studied in this paper is the sum of $Q_{dr}$ and the overland flow ($Q_{sur}$) across the model boundaries.

To decrease model runtime, symmetry in the roofs was used to reduce the size of the simulation domain. In Lyon, half the roof was simulated (and the resulting flow multiplied by 2). In Umeå, one half of the narrow roof section was simulated (i.e. the contributing area for one of the two drains installed at the edge) and one quarter of the wider part of the roof (i.e. the area contributing to a roof drain from one side, since drains were installed at ¼ and ¾ of the total width), see Fig. 1.

### 2.4. Calibration approach

All tested models were calibrated for the selected sites and calibration periods using the Bayesian algorithm DREAM$_{(ZS)}$ (Laloy and Vrugt, 2012) as implemented in the Python package PyDREAM (Shockley et al., 2018) and with the following log-likelihood function:

$$\mathscr{L} = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{i=N}\ln(\sigma_i) - \sum_{i=1}^{i=N}\frac{\left(Q_{obs,i}-Q_{sim,i}\right)^2}{2\sigma_i^2} \quad (18)$$

where $N$ is the number of observations, $Q_{obs}$ is the observed flow, $Q_{sim}$ is the simulated flow from the hydrological model, and $\sigma$ is the standard deviation of the distribution which needs to be established during calibration. Since it is expected that this standard deviation will be higher for larger flows (i.e. the errors show heteroscedasticity), it was parameterized following Ammann et al. (2019):

$$\sigma_i = \sigma_0 + \sigma_1 Q_{sim,i}^c \quad (19)$$

where $\sigma_0$ and $\sigma_1$ are calibration parameters. For this study $c$ was set to 0.5 to limit the expansion of $\sigma$ for high flows, as preliminary runs showed it becoming so large that the simulated runoff underestimated peak flows to an undesirable extent (while still obtaining good likelihood values due to the large $\sigma$). This approach was also used in earlier research for natural catchments to improve the performance of model predictions (McInerney et al., 2017). More complex likelihood functions that account for bias, skew and autocorrelation of the residuals are an on-going topic of research (e.g. Ammann et al., 2019; Evin et al., 2013; McInerney et al., 2017; Oliveira et al., 2018; Schoups and Vrugt, 2010; Wani et al., 2017). However, there are (as of yet) no universal recommendations on what more complex likelihood functions are applicable in what conditions. Problems have been reported with accounting for autocorrelation at the short time steps common in urban hydrology (Métadier and Bertrand-Krajewski, 2012), and for natural catchments it has been reported that accounting for autocorrelation can produce worse predictions, especially for dry catchments (Evin et al., 2014; Schoups and Vrugt, 2010) with which urban catchments share some similarities. In this light it was preferred to use a relatively simple likelihood function in this study. For comparison, the assumptions of normality and independence of the residuals were also employed in earlier green roof modelling work (Brunetti et al., 2020). Calibration parameters and prior ranges (i.e. the minimum and maximum allowed values) are given in Table 1. For the green roof in Lyon, the maximum substrate depth in the calibration range of $D_{sub}$ was increased from 140 mm to 175 mm, in order to include the remaining substrate at the inner part of the egg-box drainage, which could represent a higher storage capacity than *in-situ* measurements.

### 2.5. Probabilistic predictions of roof outflow

A successful calibration with DREAM results in a sample of parameter sets that form a representative set of samples of the (joint) parameter posterior distribution. The hydrological model output ($Q_{sim}$) for each of those parameter sets is available. For each observation time step, this ensemble of hydrological model outputs can be summarized by a confidence bound (here, the 90% central bound was used) to visualize the predicted outflow over time and the uncertainty that arises from uncertainty in the estimates of the hydrological model parameters. In addition, the total uncertainty (parametric + model structure uncertainty, see e.g. Kuczera et al., 2006) can be visualized. Stochastic model outputs ($Q_{stoch}$) are sampled from a normal distribution, where the mean is the hydrological model output for a parameter sample ($Q_{sim}$), and the flow-dependent standard deviation is given by Eq. (19); any negative flow values are set to zero. The resulting ensemble of stochastic model outputs is again summarized using the 90% confidence bound at each observational time step, thus visualizing total uncertainty.

### 2.6. Model evaluation criteria

#### 2.6.1. Nash-Sutcliffe efficiency

The Nash-Sutcliffe efficiency (NSE) is calculated using:

$$NSE = 1 - \frac{\sum_{i=1}^{i=N}\left(Q_{obs,i}-\widehat{Q_{sim,i}}\right)^2}{\sum_{i=1}^{i=N}\left(Q_{obs,i}-\overline{Q_{obs}}\right)^2} \quad (20)$$

where $\widehat{Q_{sim}}$ is the flow for the parameter sample with the highest posterior probability (MAP) and $\overline{Q_{obs}}$ the mean of all flow observations.

#### 2.6.2. Volume error

The volume error (VE) is calculated as:

$$VE = \frac{\int\widehat{Q_{sim}} - \int Q_{obs}}{\int Q_{obs}}100\% \quad (21)$$

where $\int\widehat{Q_{sim}}$ and $\int Q_{obs}$ are the simulated (MAP) and observed flow volumes, calculated taking into account the non-constant time steps between observations in the data from Umeå, where a tipping bucket was used for measuring low flow rates (Section 2.1).

#### 2.6.3. Reliability

For perfectly reliable probabilistic model predictions (i.e., probability distributions), the actual observations should be proportionately distributed across the corresponding predictive distributions, i.e. with

**Table 1**
Calibration parameters and prior ranges.

| Parameter | Symbol | Unit | Min | Max | SWMM | SHE | Urbis | Hydrus |
|---|---|---|---|---|---|---|---|---|
| Manning's coefficient | $n$ | s $m^{-1/3}$ | 0.01 | 0.60 | X | | | |
| Manning's coefficient | $n_{she}$ | $m^{1/3}$ $s^{-1}$ | 0.1 | 100 | | X | | |
| Detention storage | $D_{sur}$ | mm | 0 | 5 | X | X | | |
| Porosity | $\theta_s$ | – | 0.4 | 0.6 | X | X | | X |
| Field capacity | $\theta_{fc}$ | – | 0.11 | 0.35 | X | X | | |
| | | | 0.1 | 0.5 | | | X | |
| Saturated hydraulic conductivity* | $K_s$ | mm $h^{-1}$ | 5 | 1200 | X | | | |
| Saturated hydraulic conductivity* | $K_s$ | m $s^{-1}$ | 1.4e-6 | 3.3e-4 | | X | | |
| Saturated hydraulic conductivity | $K_s$ | mm $min^{-1}$ | 2 | 100 | | | | x |
| Conductivity slope | $\gamma$ | – | 20 | 60 | X | | | |
| Averjanov coefficient | $V$ | – | 1 | 30 | | X | | |
| Green-Ampt suction head | $\psi$ | mm | 25 | 200 | X | | | |
| Initial potential pressure head | $\psi_{init}$ | mm | −5000 | −1 | | | | X |
| Manning drainage layer | $n_{dr}$ | s $m^{-1/3}$ | 0 | 4 | X | | | |
| Saturated zone drain constant | $R_{dr}$ | $s^{-1}$ | 1e-6 | 1e-2 | | X | | |
| Flowrate regulation coefficients | $A_{reg}$ | $mm^2$ | 0 | 10,000 | | | x | |
| Van Genuchten parameter | α | $mm^{-1}$ | 0.0005 | 0.005 | | | | X |
| Van Genuchten parameter | β | – | 1 | 2.5 | | | | X |
| Van Genuchten parameter | λ | – | 0 | 1 | | | | x |
| Substrate depth | $D_{sub}$ | mm | 40 | 175 | Lyon | Lyon | Lyon | Lyon |
| Void index of the drainage layer | $\phi_{dr}$ | – | 0.4 | 0.6 | | | Lyon | |
| Heteroscedasticity slope | $\sigma_1$ | $l^{1/2}m^{-1/2}$ | 0 | 5 | X | X | X | X |
| Heteroscedasticity intercept | $\sigma_0$ | l/min | 0 | 5 | X | X | X | X |

* Calibrated in log-scale.

more observations in the peak(s) of the predictive distributions and fewer observations falling in the tails of the predictive distributions. This can be checked by first calculating for each observation where it falls along the cumulative distribution function (CDF) of the predictions (for the same time step), i.e. resulting in a value in the interval [0, 1]. Then, if the observations are appropriately distributed across the predictive distributions, all the CDF values together should follow a Uniform(0,1) distribution. This can be measured by the distance between the CDFs of a U(0,1) distribution and of the observed distribution of CDF values. If this is evaluated at each observation's CDF value, it results in the following metric for reliability (Ammann et al., 2019; Evin et al., 2014):

$$\Omega_{rel} = 1 - \frac{2}{n+1}\sum_{i=1}^{i=N}\left|F_{Qi}\left(Q_{obs,i}\right) - F_{\zeta}\left(F_{Qi}\left(Q_{obs,i}\right)\right)\right| \quad (22)$$

where $F_{\zeta}$ is the empirical cumulative distribution function of $\{F_{Qi}(Q_{obs,i})|i = 1, 2, \cdots, N\}$, i.e. the set of the CDF values of the observations for all time steps $i$. This value can vary from 0 to 1, where 1 indicates perfect reliability.

### 2.7. Relative spread of model predictions

The precision of model predictions (based either on parametric or total uncertainty) can be measured by the spread of the predictions relative to the observed flows (Ammann et al., 2019; McInerney et al., 2017):

$$\Omega_{spread} = \frac{\sum_{i=1}^{i=N}\sigma_{Q,i}}{\sum_{i=1}^{i=N}Q_{obs,i}} \quad (23)$$

where $\sigma_{Q,i}$ is the standard deviation of the ensemble of runoff predictions at time step $i$. Generally, a smaller spread is desirable as it means model predictions are more precise (i.e. less uncertain), although increased precision may come at the cost of reduced reliability.

## 3. Results

### 3.1. Posterior parameter distributions

The marginal posterior distributions (PD hereafter) of the model parameters are shown in Fig. 3 for Lyon and Fig. 4 for Umeå. Since uniform prior distributions were used, the prior distribution is simply a horizontal line. If no information is gained about a certain parameter in the calibration period the posterior distribution will be the same as the prior and the parameter is considered non-identifiable from the data. The more information about suitable values for a parameter is gained in calibration (i.e. the more identifiable it is), the more its PD will diverge from a horizontal line, moving instead towards a single peak. Parameters that showed to be well identifiable in all calibrations were the field capacity $\theta_{fc}$ and porosity $\theta_s$, γ (SWMM only), Van Genuchten's $\alpha$, $\lambda$, $\beta$, (Hydrus), drainage parameters ($n_{dr}$ in SWMM, $\psi_{init\_bot}$ in Hydrus, $R_s$ in SHE,) and the error model parameters $\sigma_0$ and $\sigma_1$. The only parameter which remained practically unidentifiable in all calibrations was the void index of the drainage layer ($\phi_{dr}$) in Lyon for Urbis (not calibrated in Umeå since it was known from manufacturer's information). For the other parameters, the identifiability was found to be site- and calibration period-dependent.

The information content of the different calibration periods for each site can be compared using the Kullback-Leibler divergence from the prior to the posterior. This value is indicated for each individual parameter in Figs. 3 and 4. The sum of the values for all parameters in each calibration is given in Table 2. For Umeå the September calibration period was the most informative for all four models and for Lyon the April period was the most informative. These periods were selected (Section 2.1) because they include high variability in the retention of the individual rainfall events, indicating that this may be a useful method for identifying suitable calibration periods, although this assumption would need to be verified for additional sites.

In some cases, the parameter posterior distributions collapse towards the boundaries imposed in calibration e.g. for $\theta_{fc}$ and $\theta_s$, the drainage constant $t_{dr}$ in SHE, and $K_s$ in SWMM (Umeå only). In these cases, it could be considered to extend the prior range, but this was not implemented

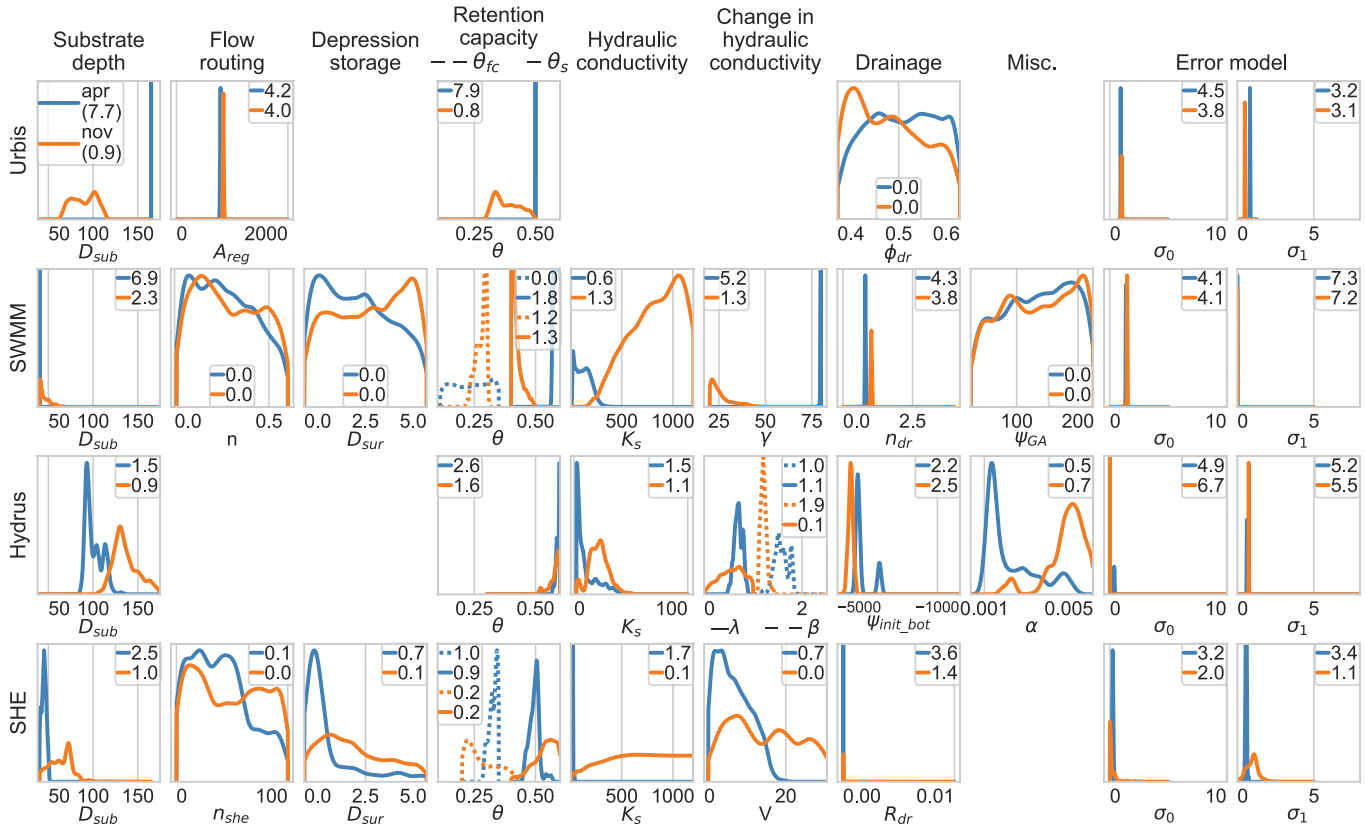


**Fig. 3.** Marginal posterior distributions (kernel density estimates) of hydrological and error model parameters for Lyon (in orange: Nov. 2012, in blue: April 2013) with Kullback-Leibler divergence from the prior indicated. The y-axis in each subplot shows probability density (dimensionless); since these values are not comparable between different subplots they are not shown here to conserve space. Parameters with similar roles are grouped in the same column, even if their interpretation is not completely the same in the different models; miscellaneous parameters that are not related to each other are shown in the column Misc.. Note that lines for the two different calibration periods overlap in some cases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

because the bounds of these parameters aimed to retain a physically realistic representation.

### 3.2. Predictive uncertainty

The model predictions (i.e. central 90% confidence bounds showing both parametric and total uncertainty) for the calibration phase are shown in Fig. 5 for Lyon and Fig. 6 for Umeå. Uncertainties in Urbis, SWMM and Hydrus seemed to be driven by the uncertainties related to the model structure, as parametric uncertainty (narrow dark grey band) presented a minor contribution to the total predictive uncertainty (average of 2–7% of the total uncertainty). For SHE the contribution of parametric uncertainty was larger (24% of the total predictive uncertainty, on average). In rainfall events with multiple peaks and troughs (e.g. the event on May 2nd-3rd in Lyon) the total uncertainty bounds in Urbis did not follow the same pattern of peaks and troughs, while the other models managed to capture these patterns with their narrower uncertainty bounds. For several events (especially in Lyon), runoff from Hydrus started too early and initial peaks in the hydrograph were also sometimes overestimated (see e.g. the events in Lyon on April 27th and November 26th). For the first events in each period this could be possibly related to the initial conditions of the model run (despite using a 14 day warmup period), but this seems unlikely as the issue still occurs later in the simulation period, e.g. on November 26th in Lyon.

Considering the testing phase (Fig. 7 for Lyon, Fig. 8 for Umeå) the wider hydrological parameter uncertainty bounds in SHE stood out, especially for the April period in Lyon and the September period in Umeå, where they formed most of the total uncertainty bound. On two occasions (around October 7th 00:00 and around October 9th 12:00) the uncertainty bounds are much wider than for other similar flow rates and driven completely by the hydrological parameter uncertainty. This behaviour may be partially related to some numerical problems encountered in SHE (see Section 3.3). Urbis completely missed one small event in Umeå (July 23rd) and two modest to large events in Lyon (November 4th and 8th), but otherwise the uncertainty bounds behaved similarly to the calibration phase. SWMM had some problems where it over- or underestimated flow at the start of events in Lyon. The problem with Hydrus predicting too high runoff early in some rainfall events that was present in the calibration phase persisted in the testing phase.

The predictive performance of the models is summarized in Table 3. SHE had the largest relative spread (i.e. the widest uncertainty bounds) of flow predictions resulting from uncertainty in the hydrological model parameters alone, but Urbis had the largest total uncertainty, with SWMM uncertainties being the smallest. Parametric uncertainty relative to the observations was small for Urbis, SWMM and Hydrus ($\leq$0.03) but somewhat larger for SHE (0.14 on average). Total uncertainty was smallest in SWMM, followed by Hydrus, SHE, and finally Urbis. Predictive uncertainty in SHE was driven by uncertainty in hydrological model parameters to a greater extent (24% in calibration, 32% in testing) than in Urbis (2% in both), SWMM (3% and 8% respectively), and Hydrus (7% in calibration, 3% in testing).

The agreement between the probabilistic model predictions (total uncertainty) and the observations was measured using the reliability metric (see Eq. (20)) and the percentage of observations falling inside the 90% prediction interval. Reliability varied between the different sites and calibration periods. Urbis was the most reliable model in Lyon

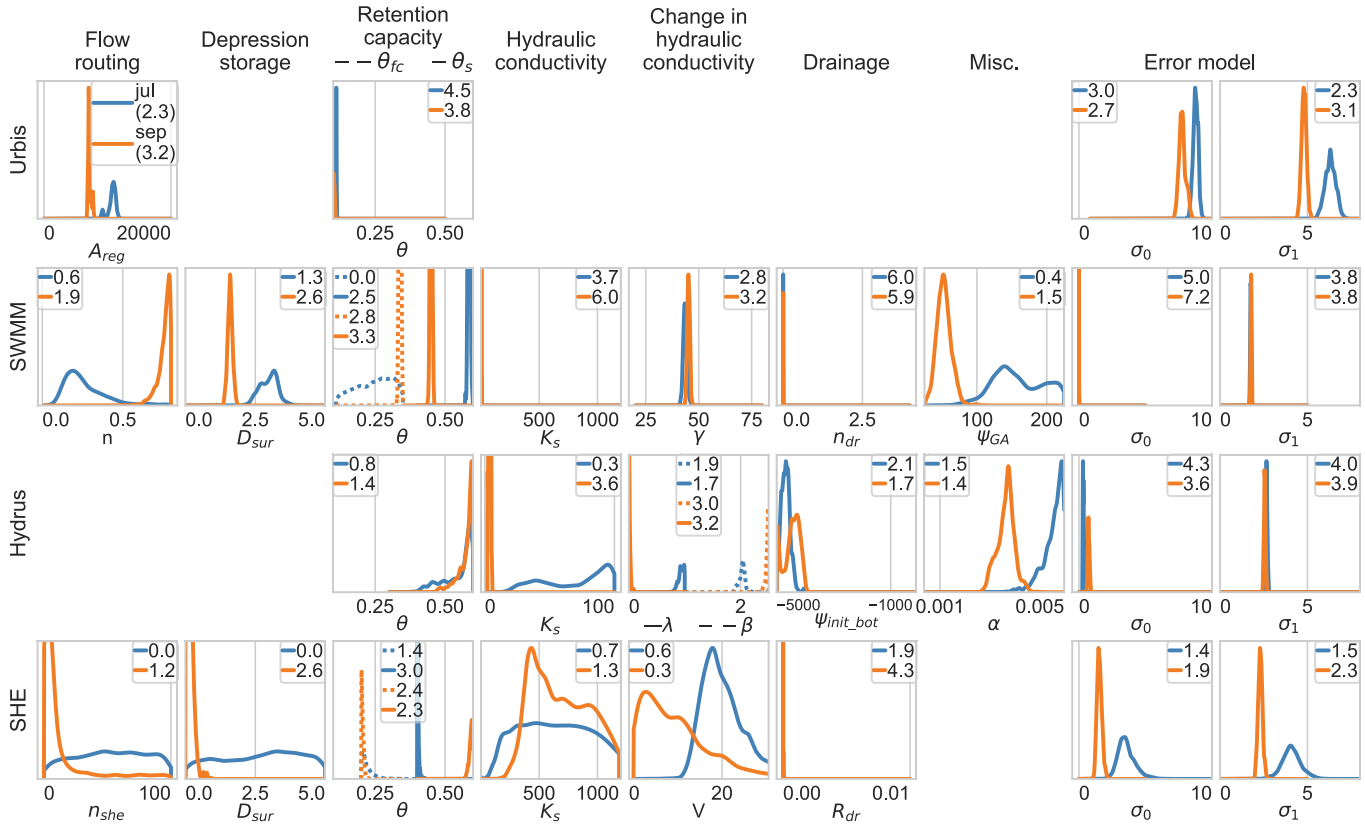


**Fig. 4.** Marginal posterior distributions (kernel density estimates) of hydrological and error model parameters for Umeå (in orange: Sept. 2020, in blue: July 2020) with Kullback-Leibler divergence from the prior indicated. The y-axis in each subplot shows probability density (dimensionless); since these values are not comparable between different subplots they are not shown here to conserve space. Parameters with similar roles are grouped in the same column, even if their interpretation is not completely the same in the different models; miscellaneous parameters that are not related to each other are shown in the column Misc. Note that lines for the two different calibration periods overlap in some cases; for $K_s$ in SWMM both lines fall at the left border of the plot (the minimum value imposed in calibration). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
Sum of Kullback-Leibner divergence between prior and posterior distributions for all calibration parameters. The period where this value is highest is marked in bold font.

| Site | Umeå | | Lyon | |
|---|---|---|---|---|
| Period | Jul | Sep | Apr | Nov |
| Urbis | 12.3 | **13.1** | **27.9** | 12.8 |
| SWMM | 26.1 | **38.4** | **30.7** | 23.3 |
| Hydrus | 16.8 | **22.1** | **22.4** | 21.2 |
| SHE | 10.4 | **16.7** | **17.3** | 6.7 |

in April and a close second to Hydrus in November, but the least reliable in Umeå. Reliability of SWMM and SHE was similar for the different calibration scenarios. Hydrus had the highest reliability overall (0.80 on average), although it did not score well (0.61) in Umeå – September. In testing, reliability was lower than in calibration for Urbis (from 0.73 to 0.60), SWMM (0.74 to 0.72) and SHE (0.77 to 0.73), but Hydrus had the same average reliability as in the calibration phase (0.80).

In terms of the percentage of observations captured by the total uncertainty bounds the models showed generally good performance, with 13 out of 16 calibrations capturing ≥90% of the observations in the 90% confidence bounds of the predictions. In testing, the percentage cover was somewhat lower, reaching >90% for only 6 out of 16 calibrated models. SWMM, Hydrus and SHE still consistently exceeded 80% observation cover, while for Urbis this varied from 72% to 83%.

The Nash-Sutcliffe efficiency (calculated for the MAP parameter set) gives an overall indication of the goodness of fit of the model. The July period in Umeå shows relatively low scores in both calibration and testing for all models, suggesting that some processes were not properly accounted for by any of the studied models or that there may be some issue with the data in this period. The NSE for SWMM and SHE was generally high (≥0.9 in calibration, ≥0.8 in testing) except for SHE in the July period in Umeå (0.6). Urbis had worse performances in this metric (≤0.8 in calibration, ≤0.7 in testing). Performance for Hydrus was rather variable, ranging in calibration from – 0.51 (Umeå July) to 0.91 (Umeå September) and in testing from −0.03 (Umeå September) to 0.84 (Umeå July). NSE for Hydrus in Lyon was always low compared to SWMM and SHE. These relatively poor NSE scores for Hydrus can be linked to occasional issues with reproducing flow rates in the beginning of the events discussed above and visible in Figs. 5–8.

Finally, the volume errors (calculated for the MAP parameter set) indicate how well the models approximate the water balance in the roof during the simulations. For Lyon, SWMM and SHE showed good performances (≤8% error in calibration, ≤17% in testing). Urbis had similar performances in calibration (+8% and +10% for April and November respectively) but considerably larger errors in testing (−33% in April and +42% in testing). Volume error for Umeå was rather high across all models: in calibration, only SHE had acceptable errors (≤7%) with larger errors for SWMM (up to +24% in July), Hydrus (+28% in July) and Urbis (−52% in July). In testing, volume errors were generally large (≥20%). Note that volume errors in one part of the record may be compensated by errors in the other direction in another part of the record, so the volume error is primarily an indication of model bias and should not be interpreted without considering other performance indicators.

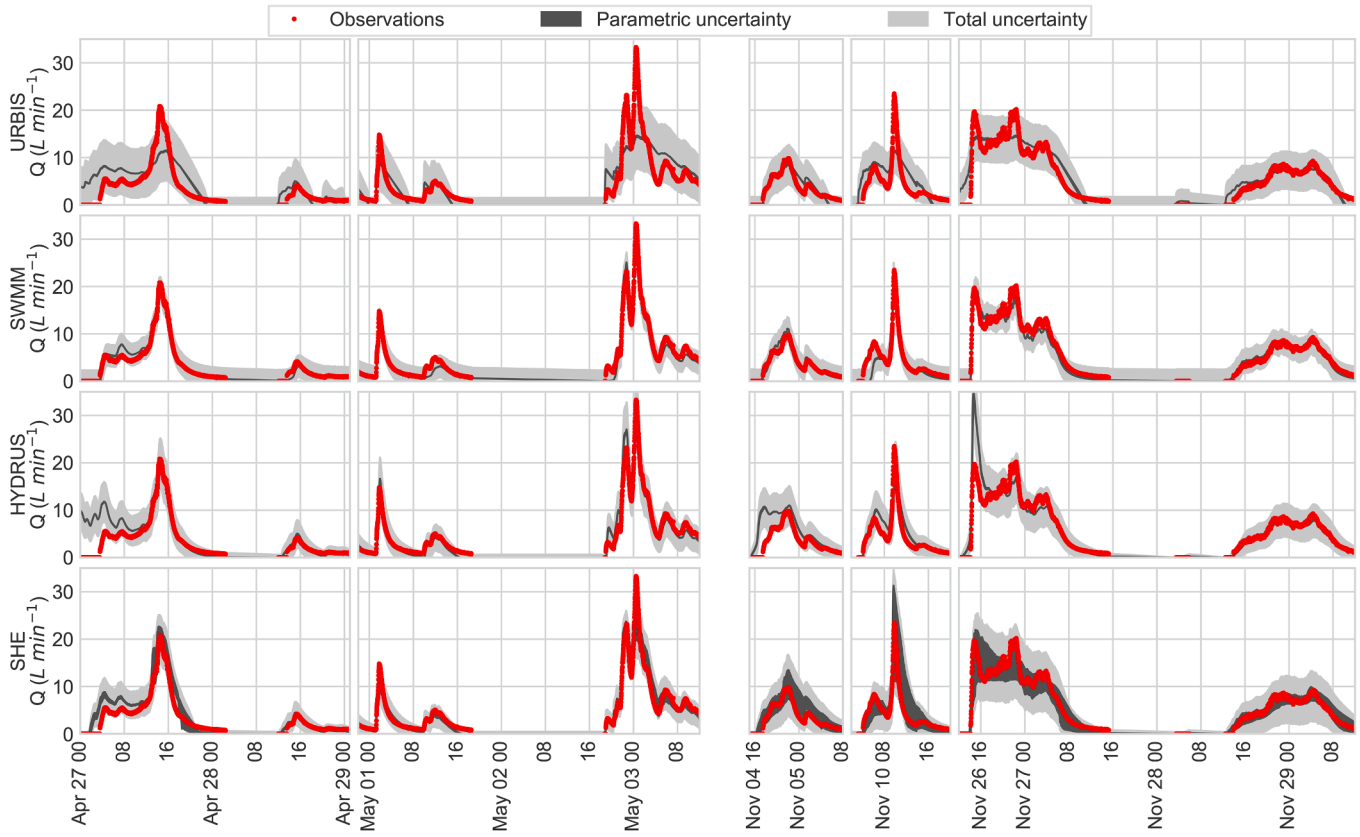


**Fig. 5.** Model predictions for the Lyon calibration phase, zoomed in on two or three time periods within each month.

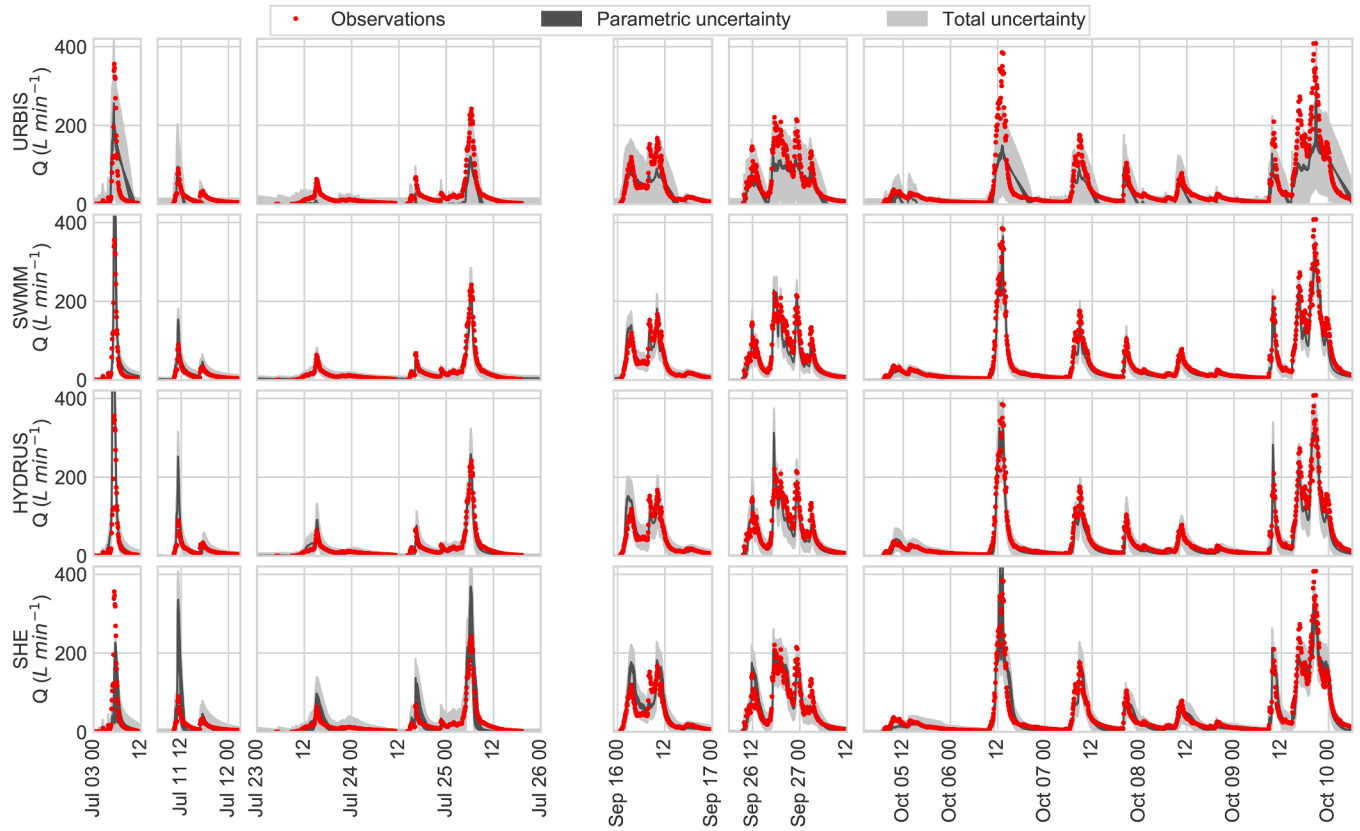


**Fig. 6.** Model predictions for the Umeå calibration phase, zoomed in on three time periods within each month.

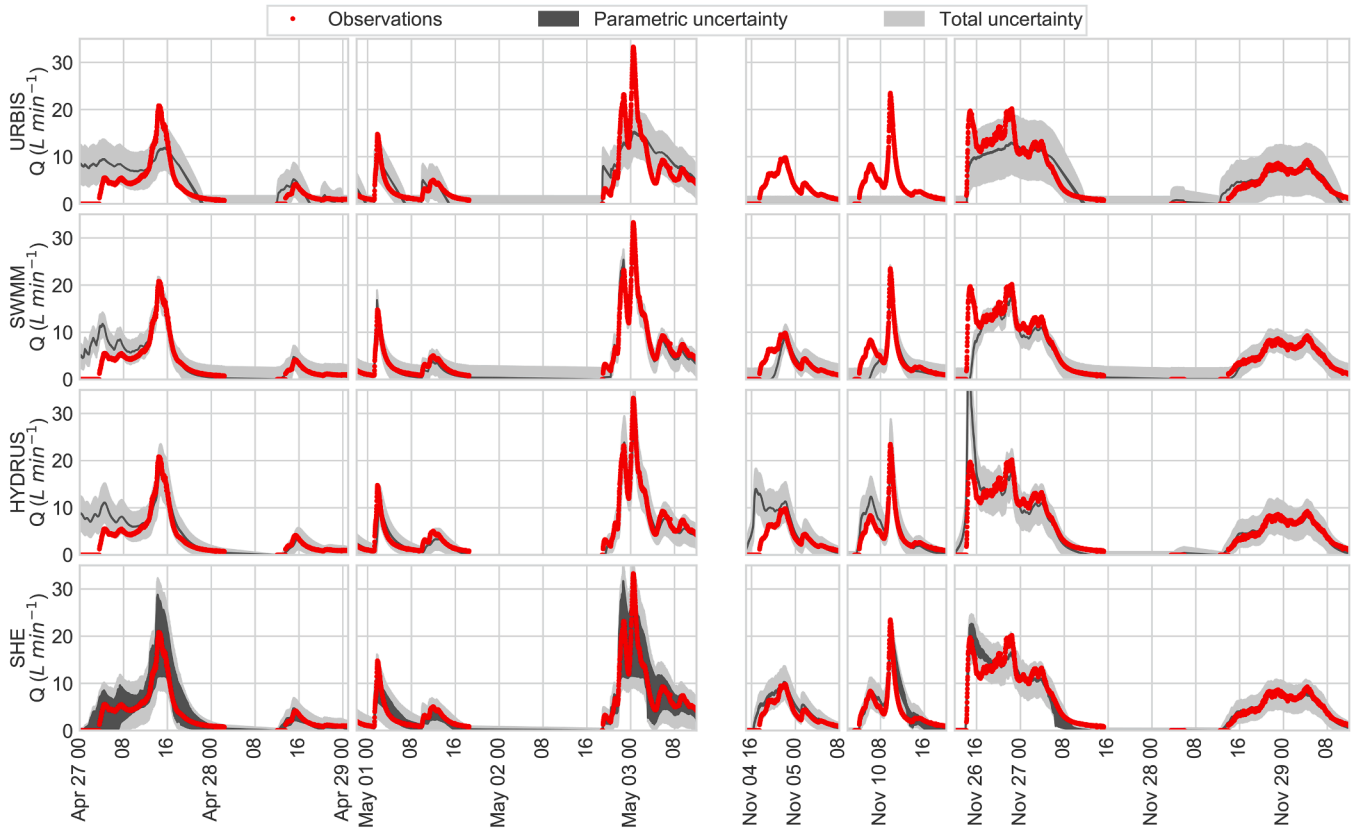


**Fig. 7.** Model predictions for the Lyon testing phase, zoomed in on two or three time periods within each month.

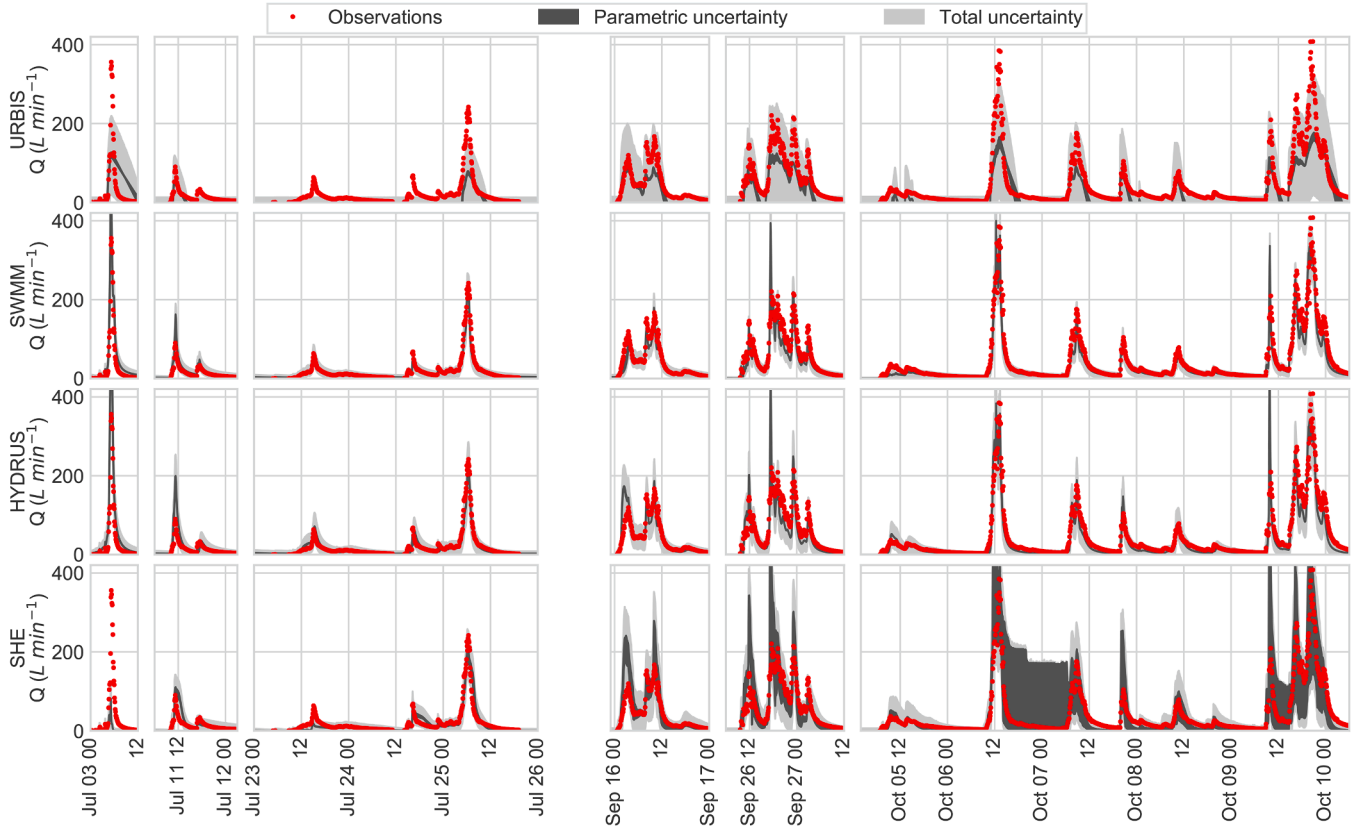


**Fig. 8.** Model predictions for the Umeå testing phase, zoomed in on three time periods within each month.

**Table 3**
Summary statistics for predictive performance in calibration and testing. Explanation of columns: (1) relative spread of the hydrological model outputs, i.e. reflecting only parametric uncertainty; (2) relative spread of hydrological model outputs + randomly sampled uncertainty, i.e. total uncertainty; (3) is (1) as percentage of (2). *: the MAP from the Umeå July SHE calibration resulted in a failed model run in testing, so the parameter set with the second-highest likelihood in calibration was used to calculate these values.

| | | | Calibration | | | | | | | Testing | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Site | Cal. month | Model | $\Omega_{spread,par}$(1) | $\Omega_{spread,total}$(2) | Percentage parame-tric uncertainty (3) | $\Omega_{rel}$ | % observations covered | NSE (MAP) | Volume error (MAP) | $\Omega_{spread,par}$(1) | $\Omega_{spread,total}$(2) | Percentage parame-tric uncertainty (3) | $\Omega_{rel}$ | % observations covered | NSE (MAP) | Volume error (MAP) |
| Lyon | Apr | Urbis | 0.01 | 0.71 | 1% | 0.88 | 0.93 | 0.69 | 8% | 0.00 | 0.39 | 1% | 0.65 | 0.72 | 0.45 | −33% |
| | | SWMM | 0.00 | 0.52 | 1% | 0.73 | 0.96 | 0.96 | −8% | 0.00 | 0.28 | 1% | 0.76 | 0.90 | 0.76 | −17% |
| | | Hydrus | 0.01 | 0.50 | 2% | 0.90 | 0.83 | 0.66 | 30% | 0.01 | 0.33 | 2% | 0.95 | 0.87 | 0.60 | 14% |
| | | SHE | 0.06 | 0.36 | 16% | 0.83 | 0.88 | 0.91 | 4% | 0.04 | 0.28 | 15% | 0.91 | 0.86 | 0.92 | 3'% |
| | Nov | Urbis | 0.01 | 0.41 | 2% | 0.90 | 0.93 | 0.81 | 10% | 0.01 | 0.69 | 1% | 0.77 | 0.79 | 0.46 | 42% |
| | | SWMM | 0.00 | 0.31 | 1% | 0.80 | 0.96 | 0.94 | −7% | 0.01 | 0.57 | 1% | 0.77 | 0.93 | 0.87 | 5% |
| | | Hydrus | 0.03 | 0.31 | 11% | 0.86 | 0.93 | 0.79 | 12% | 0.01 | 0.41 | 3% | 0.85 | 0.83 | 0.75 | 22% |
| | | SHE | 0.20 | 0.49 | 41% | 0.77 | 1.00 | 0.88 | −4% | 0.29 | 0.64 | 45% | 0.73 | 0.97 | 0.78 | −7% |
| Umeå | Jul | Urbis | 0.03 | 1.08 | 2% | 0.57 | 0.89 | 0.47 | −52% | 0.03 | 0.66 | 4% | 0.48 | 0.83 | 0.71 | −45% |
| | | SWMM | 0.02 | 0.45 | 5% | 0.71 | 0.93 | 0.78 | 24% | 0.03 | 0.21 | 13% | 0.64 | 0.82 | 0.89 | −20% |
| | | Hydrus | 0.06 | 0.63 | 10% | 0.85 | 0.94 | −0.47 | 29% | 0.01 | 0.31 | 3% | 0.53 | 0.88 | 0.84 | −16% |
| | | SHE | 0.24 | 1.04 | 23% | 0.77 | 0.95 | 0.51 | 3% | 0.41 | 0.77 | 54% | 0.64 | 0.97 | 0.76* | −3%* |
| | Sep | Urbis | 0.02 | 0.62 | 4% | 0.58 | 0.90 | 0.67 | −36% | 0.02 | 0.84 | 3% | 0.52 | 0.79 | 0.29 | −54% |
| | | SWMM | 0.01 | 0.22 | 4% | 0.73 | 0.90 | 0.93 | −16% | 0.08 | 0.47 | 16% | 0.70 | 0.93 | −0.73 | 35% |
| | | Hydrus | 0.01 | 0.32 | 4% | 0.61 | 0.93 | 0.91 | −16% | 0.04 | 0.69 | 6% | 0.86 | 0.94 | −0.04 | 40% |
| | | SHE | 0.05 | 0.33 | 15% | 0.69 | 0.93 | 0.90 | −7% | 0.05 | 0.48 | 11% | 0.64 | 0.80 | 0.43 | −28% |
| Mean | | Urbis | 0.02 | 0.71 | 2% | 0.73 | 0.91 | 0.66 | −18% | 0.02 | 0.64 | 2% | 0.60 | 0.78 | 0.48 | −22% |
| | | SWMM | 0.01 | 0.37 | 3% | 0.74 | 0.94 | 0.90 | −2% | 0.03 | 0.38 | 8% | 0.72 | 0.90 | 0.45 | 1% |
| | | Hydrus | 0.03 | 0.44 | 7% | 0.80 | 0.91 | 0.47 | 14% | 0.02 | 0.44 | 3% | 0.80 | 0.88 | 0.54 | 15% |
| | | SHE | 0.14 | 0.55 | 24% | 0.77 | 0.94 | 0.80 | −1% | 0.20 | 0.54 | 32% | 0.73 | 0.90 | 0.72 | −9% |

### 3.3. Computational requirements and errors

The practical value of a model depends in part on its computational demands. The runtime for the various calibrations is therefore given in Table 4. Runtimes were up to 10 times longer for the more complex SHE model than for the simpler Urbis and SWMM. For SWMM and Urbis run times were similar, although it should be pointed out that Urbis has not been optimized to the same degree as the other models and it may therefore be possible to lower its runtime further. Although both SHE and Hydrus use the computationally intensive Richards equation for unsaturated zone flow, the fact that Hydrus operated in only one dimension (i.e. simulating just a single vertical profile, compared to 600 profiles in SHE) meant that this model was still 2–3 times faster than SHE.

In addition to the pure computational time, the numerical stability of the models can also affect their applicability. Urbis directly showed good convergence of the Markov chains in calibrations, while some further experiments were required in SWMM and Hydrus to ensure good performance of the calibration algorithm. On the other hand, two types of major numerical problems were encountered with SHE. First, the Markov chains of the DREAM algorithm sometimes got stuck in a state with a relatively high likelihood value where even small proposed jumps would have a much lower likelihood value, which led to a systematic rejection of all the proposals. For example, in one such state changing only $\theta s$ by $10^{-5}$ reduced the probability of accepting the new point to effectively zero. This unexpected behaviour can be attributed to numerical instability of the SHE model that may be associated with the model's origin in natural catchment modelling rather than green roofs, i.e. with larger time steps and spatial scales. It was therefore necessary to include an additional step in the calibration where the likelihood would be re-evaluated with a very small perturbation to the parameter values, and the new likelihood value used in evaluating the acceptance probability. Second, the dependence between model numerical stability and parameter values meant that some of the parameter samples from the calibration period resulted in an abnormal termination of the model when applied to the testing period. Specifically, after calibrating the SHE model for Umeå for the July period, 45% of the obtained parameter samples resulted in failed model runs (i.e. the simulation either stopped prematurely or got completely stuck) when applied in testing to the September period. The same happened with 0.5% of the samples from the Lyon November calibration; no such errors occurred in the other models.

## 4. Discussion

### 4.1. Parameter consistency and identifiability

In an ideal model where the structure adequately describes the green roof processes and model parameters describe physical aspects of the system, model parameters are expected to be stationary, i.e. independent of the selected calibration period. This property can be defined as parametric consistency (Thyer et al., 2009). However, some changes over time have been reported for physical properties of green roofs (De-Ville et al., 2018a, 2017) which could be reflected in some of the model parameters. Further, parametric identifiability can be defined as the amount of information obtained while updating the parameters' prior distribution to the posterior distribution (PD) during the calibration process. For the case study these properties of consistency and identifiability were unsatisfactory for various parameters in all of the models and for both sites (see Figs. 3 and 4). This is an indication that the four models still suffer from inadequacies in their representations of all physical processes in green roofs.

**Table 4**
Model runtimes (in hours) per 1000 iterations using an Intel Xeon 8168 processor. Note that multiple chains are run in parallel (here 5 chains were used), so 1000 iterations would give 5000 total samples.

| | Lyon | | Umeå | |
|---|---|---|---|---|
| | April | November | July | September |
| SHE | 4.9 | 4.4 | 5.9 | 6.3 |
| SWMM | 0.4 | 0.4 | 1.1 | 1.5 |
| Urbis | 0.5 | 1.0 | 1.1 | 1.0 |
| Hydrus | 2.3 | 2.1 | 2.4 | 2.0 |

The dependence of parameter PDs on the selected calibration period (lack of consistency) can be interpreted as evidence that parameter values are dependent not just on physical characteristics of the green roof, but also on the meteorological conditions. The dependence of parameter values on meteorological conditions may have also played a role in an earlier study (Johannessen et al., 2019), where it was found that calibrated parameter values for one site gave poor model performance when applied to a roof with the same design in a different location. However, other factors (e.g. vegetation, which will always have some dependence on local the meteorological conditions) may have also played a role in that study. The variability of estimated parameter values from different calibration periods implies that models might deliver different results when applied to new periods or for forecasting. This is already visible to some extent in the testing phase where model predictions are less accurate compared to calibration results. The fact that parameter values obtained for the different calibration periods differ as much as they do raises the question of what the impact of this will be when the meteorological differences between the calibration period and the prediction period become larger, e.g. when models are used to forecast the behaviour of green roofs under climate change scenarios (e.g. Karlsson et al., 2016).

The identifiability of parameter values was previously investigated by Brunetti et al. (2020) in an experiment with a green roof module in a climate chamber. As in the current paper, they found that soil porosity $\theta_s$ was well-identifiable using outflow measurements for both conceptual and mechanistic models. They also found that both the identifiability and the "optimal" values (i.e. the parameter values with the maximum posterior likelihood (MAP)) depended on the type of calibration data used (substrate water content, outflow, or tracer transport); the results from this paper provide evidence that these values are further dependent on properties of the specific calibration period chosen, e.g. different meteorological and vegetation conditions. Identifiability of model parameters is related to the model's sensitivity to changes in the values of parameters which was previously investigated for SWMM (in a sensitivity analysis, rather than a calibration context) by Leimgruber et al. (2018). For example, they found that substrate depth and porosity were the most sensitive parameters, which is confirmed by their high identifiability (see Figs. 3 and 4). By contrast, Leimgruber et al. reported low sensitivity of field capacity and $\gamma$ (called "conductivity slope" in their paper), but in our results, the identifiability of field capacity was site-dependent while $\gamma$ was well-identifiable in all cases. Furthermore, they reported that hydraulic conductivity was almost completely insensitive, but this parameter was nonetheless well-identifiable in all calibrations except November in Lyon. The Green-Ampt suction head was identifiable in Umeå (more so for September than July) despite being reported as an insensitive parameter by Leimgruber et al. Most notably they also reported that the Manning's coefficient for the drainage layer was completely insensitive (unidentifiable) while this parameter was identifiable in all calibrations for both field sites in this study.

Compared to the other models, Urbis uses a relatively simple structure with fewer calibration parameters, which can explain the good identifiability of the model parameters in all calibrations (except for the drainage layer void index in Lyon). However, as stated in similar calibration settings within the hydrological context (Thyer et al., 2009), parametric consistency for highly identifiable parameters is more likely to be unsatisfactory, as the posteriors delivered from different calibration periods are less likely to overlap. Higher parametric identifiability at the expense of the lack of consistency, besides from Urbis, is also

appreciable for almost all parameters in SWMM and Hydrus in the Umeå site. And, to a lower extent, for the parameters $n_{dr}$ (in SWMM), $D_{sub}$ (in SHE) and $\psi_{init\_bot}$ (in Hydrus) in Lyon. Complementarily, the observed stability and rapid convergence of the Markov chains in Bayesian calibrations for Urbis may also be related to the high parameter identifiability and lower complexity of the inference, given the lower number of parameters used to by this model.

### 4.2. Physical interpretation of parameter values

The differences in identifiability and low consistency that was found for parameters included in multiple of the studied models (e.g. substrate depth (Lyon) and porosity in all models, Manning's coefficient, depression storage, field capacity, for SWMM and SHE, saturated hydraulic conductivity in all models except Urbis) can be considered a further indication that model parameters are not exclusively representing the variability of the physical characteristics that they are aimed to represent. The obtained parameter MAP values also raise some questions. For SWMM (both calibration periods) and Hydrus (September only) in Umeå the PD of the hydraulic conductivity $K_s$ was concentrated towards the minimum value allowed in calibration. This minimum was set based on literature values for natural soils (Rawls et al., 1983), but it is rather low compared to values reported for green roof substrate (Peng et al., 2020). This could be an unexpected result of the calibration process (e.g. compensating for some other parameter or some process not well-accounted for in the model), or it may indicate that laboratory measurements of $K_s$ are not necessarily indicative of effective field values. The latter could for example depend on the effects of vegetation, compaction and aging of the substrate (e.g. De-Ville et al., 2017), or the lower permeability of the geotextile separating the substrate from the drainage layer (reported as 90 mm $hr^{-1}$ by the manufacturer (BG Byggros AB, 2018b)). Unrealistic calibrated values of green roof model parameters have been found in earlier studies as well (Avellaneda et al., 2014), even in carefully controlled laboratory experiments (Brunetti et al., 2020). The current study provides further evidence that calibrated parameters for even rather detailed green roof models may not be exclusively linked to physical characteristics of the system.

It is not completely clear why different models results in different estimates of the same parameter, but some potential explanations can be found by examining the model structures. Saturated hydraulic conductivity ($K_s$) is also estimated differently for the three models that include it. One aspect that may influence this is that the models used $K_s$ for different calculations: Hydrus used Van Genuchten's model and SHE used the Averjanov equation to describe unsaturated hydraulic conductivity, while SWMM used $K_s$ to calculate both infiltration (Green-Ampt equation) and percolation to the drainage layer (Darcy equation). These equations all have different forms, which may be part of the reason why the estimated values of Ks differ between the models. The maximum storage capacity of the green roofs depends on the substrate thickness $D_{sub}$ and the saturated moisture content $\theta_s$, but these parameters also calibrated to different values in the different models. This may be explained by the fact that the models were calibrated against only outflow data, while the relationship between outflow and current water storage is different across the models. In addition, for SWMM and URBIS part of the difference in maximum water storage capacity may instead be represented by the storage in the drainage layer; this separate compartment does not exist in the other two models. In SWMM and SHE the estimates for Manning's coefficient ($n$ and $n_{she}$ respectively) are remarkably similar, keeping in mind that $n_{she}$ is the inverse of $n$ so the regions of highest likelihood (peaks in the distribution) for the September calibration in Umeå are actually around the same value. The depression storage is not well identifiable for either model in Lyon, while in Umeå the values are somewhat different, which may be related to the differences in maximum substrate storage capacity between the two roofs.

### 4.3. Predictive uncertainty

In calibration all models included close to or more than 90% of observations in the 90% confidence intervals of the model predictions. The relative spread of the model predictions was however lower in SWMM, making its predictions more precise. In testing, the performance was only slightly lower for SWMM, Hydrus and SHE while the relative spread of the predictions was the same; for Urbis the prediction bounds were slightly narrower than in calibration (but still larger than the other models), but at the cost of lower coverage of the observations (78%). The larger uncertainty bounds and lower cover in Urbis suggest that the model contains the highest structural deficiencies in attempting to capture the main dynamics of the hydrological process for the studied green-roofs. This is also confirmed by the lower Nash-Sutcliffe values. Overall, SWMM and SHE had similar performances and which one performed better depended on the calibration site and period and on the considered performance metric. In most cases (Umeå September being the exception), Hydrus had slightly higher reliability values than SWMM and SHE, but similar or lower cover of the observations. Hydrus also frequently had lower NSE values than SWMM and SHE. Palla et al. (2012) found that Hydrus performed better than a still acceptable conceptual model; in the present study Hydrus usually performs better than the simple conceptual model Urbis, except for the calibration phase in Lyon. Compared to SWMM, Hydrus usually has better reliability but lower observation cover and NSE. Finally, in terms of volume error, SHE clearly performed better than the other models.

One reason for the differences in performance between the models may be the varying ways in which lateral flow is represented. Urbis lacks a routing component for overland flow, but since this also applies to Hydrus (which performed better) it is likely that other factors (e.g. the simple model structure with few calibration parameters) are influencing the performance for Urbis. Comparing SWMM, Hydrus and SHE, which model performs best depends on the evaluation criterion used. Because of this, and because the models differ in their descriptions of other aspects (e.g. vertical flow) as well, it is not possible to say in general which conceptualization of lateral flow is more appropriate: Manning's equation for surface and drainage layer flow (SWMM), no explicit routing (Hydrus, although the calibrated lower boundary condition may act to provide some flow routing) or a calibrated drainage constant and 2D overland flow (SHE).

Differences between the studied models were most noticeable around high flow peaks, which are also practically relevant in sizing of drainage systems. For example, in Umeå (calibration), the upper bound of the predicted peak flow on July 11th varies strongly: approx. 210 L $min^{-1}$ in Urbis, 300 L $min^{-1}$ in Hydrus, 400 L $min^{-1}$ in SHE, and 180 L $min^{-1}$ in SWMM compared to the observed value of approx. 100 L $min^{-1}$ (see Fig. 6). This shows how large the uncertainty arising from the choice of a specific model structure can be and that this source of uncertainty should perhaps receive more attention in future urban drainage studies (see also Broekhuizen et al., 2019). Xie et al. (2020) reported that SWMM always had higher peak flow rates than Hydrus; by contrast we found that peak flows in Hydrus are similar to or larger than in SWMM. Furthermore, the simplest conceptual model Urbis usually showed the lowest peak flows while SHE would usually predict larger peak flows than SWMM, with some exceptions. For example, for the event on July 3rd in Umeå (see Figs. 4 and 6) the SWMM peak flow was more than twice that of SHE.

Deviations between observed and simulated flows are usually the largest at the beginning of events, where models may not accurately predict when runoff starts or how large the outflow will be initially. This further leads to single-peak events being generally less well represented than events with multiple peaks. This suggests that the studied models are unable to properly represent the processes in the roofs during dry periods (e.g. Berretta et al., 2014; Cascone et al., 2019), leading to under- or overestimations of the initial conditions of the roofs at the beginning of rainfall events. Further investigations into the

representation of evapotranspiration in models may be valuable to remedy this issue, as may the collection of relevant additional data such as substrate moisture content. For Hydrus the choice of boundary conditions may also be affecting the model behaviour early in rainfall events.

### 4.4. Influence of calibration site and period

For most performance metrics, the models reported more satisfactory results in Lyon than in Umeå, especially regarding runoff volume. This may be caused by a number of different factors, e.g. larger flow measurement uncertainties, different climate, and greater complexity and less typical construction of the roof in Umeå. The latter could be especially relevant for Urbis and Hydrus, since it was not possible to represent the inclination of all the individual sections in Umeå. These will make the large green roof behave more like a collection of smaller roofs. The variability of the model performances between the two roofs should also serve as a reminder that study findings about models developed/tested for a single roof should be interpreted carefully, given their low generality for different climates and experimental settings.

It is remarkable that for all four models and both calibration sites, the calibration period which was selected for its high variability in event retention, was the period that was most informative in terms of maximizing the distance between the prior and posterior distributions of model parameters (Kullback-Leibler divergence, see Table 2). Although the generality of this finding should be checked for additional sites, two potential implications arise. First, it underlines the importance of long-term measurement campaigns in actual field sites, with the aim to provide datasets containing a richer mapping of the inter-event variability. Second, in laboratory irrigation experiments (e.g. Brunetti et al., 2020), more information about the behaviour of the system may be gained by using varied rainfall events (e.g. longer and shorter, more and less intense, single and multiple peaks), dry periods and initial conditions.

### 4.5. Impact of computational challenges on model predictions

Perhaps the most noticeable effect of the occasional numerical problems with SHE is the problematic behaviour of the error bands in the testing for Umeå on the 6th of October. Because 45% of the posterior parameter samples caused an abnormal termination of the model run, these samples had to be excluded from the construction of the uncertainty bounds. Apparently, numerical stability of models depends on the parameter values to some extent, so the removed parameter sets were not evenly distributed in the sample which led to the abnormally large error bounds for some time periods in October. It can even be questioned if these uncertainty bounds for the testing period should be used, since they do not accurately reflect the inferences made about the parameter values in the calibration period. The inability to apply the calibrated model to other time periods rather limits its utility. Although model numerical errors have not received much attention in urban drainage studies (see Alfredo et al. (2010) for an example), this indicates that numerical problems may occur and need to be considered.

Partially as a consequence of the numerical issues with SHE the time required for the calibration of this model was up to a factor 10 larger than for SWMM and Urbis (Table 4). This requirement may prove problematic in some studies or practical applications. There was also no significant difference in performance between SHE and the much faster SWMM that might have justified the difference in computational time. However, there are also potential benefits associated with SHE. First, it can explicitly handle green roofs with almost any geometrical design in a relatively straightforward way. Second, it can easily handle spatially variable model parameters (e.g. for roof slope, substrate depth or varying vegetation) that may be more difficult to represent in other models. Finally, the explicit representation of overland flow may be a benefit where such flow occurs, as was the case for some combinations of parameter values in the Umeå roof. It may be worth considering the use of a model emulator to speed up the calibration process (e.g. Machac et al., 2016).

## 5. Conclusions

Planning and control of green roofs in order to realize their potential as a stormwater control measure can benefit from accurate predictions of roof runoff, but so far it has not been investigated extensively what hydrological models are most suitable for this task. Therefore, this study compared the performance of four different models (the conceptual models Urbis and SWMM and the mechanistic models Hydrus-1D and Mike SHE) for two full-scale roofs in Lyon, France and Umeå, Sweden across two one-month calibration periods.

The model predictions of roof runoff were least accurate for the simple conceptual model Urbis, although the predictions from this model still captured >90% of observations during calibration and >70% during testing. In terms of reliability, i.e. how well the observations conform to probabilistic flow predictions, the Hydrus model had best performance, while SWMM and SHE had the best performance in terms of what percentage of observations fell anywhere in the model prediction interval and in terms of Nash-Sutcliffe efficiency. SHE had the best performance in terms of total runoff volume. The differences between the models' predicted flow rates are large enough that they could lead to different designs of green roofs to meet e.g. runoff reduction targets.

For all models, but especially for SWMM, Hydrus and SHE, there were large differences between the posterior distributions and most likely values of most model parameters obtained for different calibration periods. In some cases, the posterior distributions also tended towards values that are less likely for the physical characteristics that the parameters are meant to represent. These two results indicate that for all four models tested here the parameter values do not fully reflect physical properties of the studied roofs, but are also dependent on the meteorological conditions of the different calibration periods. Even if model performance was still good when switching to the testing phase in this study, the question remains what the impact of these discrepancies in parameter values may be when models are used e.g. in long-term forecasts of green roof performance under climate change scenarios. The discrepancies in parameter estimates also imply that the model structures do not adequately describe the relevant processes in the roofs and more work is needed to improve these descriptions. Such future studies might also benefit from the use of more measured variables, e.g. substrate water content or evapotranspiration.

For both Umeå and Lyon the information about parameters gained during the calibration (i.e. the distance between the prior and posterior distributions of parameters, as measured by the Kullback-Leibler divergence) was the largest when calibrating with the period containing the highest variability in event retention. This could provide a useful method for selecting calibration periods in future studies.

The differences in complexity between the models also manifested in differing computational requirements and problems. The simpler models Urbis and SWMM took considerably less time for the calibration than the physically-based models Hydrus and SHE. In the case of SHE, there were some challenges in regards to the model numerical schemes which resulted in e.g. failed model runs and unrealistically wide error bounds in the testing phase.

## CRediT authorship contribution statement

**Ico Broekhuizen:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Santiago Sandoval:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - review & editing. **Hanxue Gao:** Investigation. **Felipe Mendez-Rios:** Investigation. **Günther Leonhardt:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Jean-Luc**

**Bertrand-Krajewski:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Maria Viklander:** Conceptualization, Methodology, Writing - review & editing, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was financed by the Swedish Research Council Formas (grants no. 2015-121 (co-funded in-kind by VAKIN) and 2015-778) and by VINNOVA as part of DRIZZLE – Centre for Stormwater Management (grant no. 2016-05176). We would like to thank Anna Magnusson, Joel Lönnqvist and Peter Rosander for their work on the flow measurements in Umeå and VAKIN for providing access to the site and for providing rainfall data.

## References

Alfredo, K., Montalto, F., Goldstein, A., 2010. Observed and modeled performances of prototype green roof test plots subjected to simulated low- and high-intensity precipitations in a laboratory experiment. J. Hydrol. Eng. 15 (6), 444–457. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000135.

Ammann, L., Fenicia, F., Reichert, P., 2019. A likelihood framework for deterministic hydrological models and the importance of non-stationary autocorrelation. Hydrol. Earth Syst. Sciences 23 (4), 2147–2172. https://doi.org/10.5194/hess-23-2147-2019.

Avellaneda, P., Leön, E., Donado, L.D., Rodríguez, E., Ballestero, T., 2014. Evaluation of an unsaturated flow model for flow attenuation in green roofs. In: World Environmental and Water Resources Congress 2014, Proceedings, pp. 2156–2164. https://doi.org/10.1061/9780784413548.215.

Berretta, C., Poë, S., Stovin, V., 2014. Moisture content behaviour in extensive green roofs during dry periods: the influence of vegetation and substrate characteristics. J. Hydrol. 511, 374–386. https://doi.org/10.1016/j.jhydrol.2014.01.036.

Bertrand-Krajewski, J.-L., Vacherie, S., 2014. Projet ECCLAIRA rapport de fin de contrat - Partie INSA Lyon LGCIE-DEEP: suivi expérimental de la toiture végétalisée du Centre des Congrès de Lyon [ECCLAIRA project final report - INSA Lyon LGCIE-Deep: experimental study of vegetated roof of the Lyon Congress Centre].

BG Byggros AB, 2019. Produktblad - gröna tak: dräneringsplattor (Product data sheet green roofs: drainage layers). <www.byggros.com/se/draneringsplattor>. (last accessed 2021-06-07).

BG Byggros AB, 2018a. Systemkonstruktion gröna tak - BGreen-it DiaCell (System configuration green roofs BGreen-it DiaCell). <https://www.byggros.com/se/nedladdningar-pdf-systemkonstruktioner-sedumtak>. (last accessed 2021-06-07).

BG Byggros AB, 2018b. Produktblad - gröna tak: filter- och skyddsdukar (Product data sheet green roofs: filter and protection textiles). <www.byggros.com/se/filterduk>. (last accessed 2021-06-07).

Broekhuizen, I., 2021. Hydrological modelling of green urban drainage systems : Advancing the understanding and management of uncertainties in data, model structure and objective functions. Doctoral Thesis. Luleå University of Technology, Luleå.

Broekhuizen, I., Muthanna, T.M., Leonhardt, G., Viklander, M., 2019. Urban drainage models for green areas: structural differences and their effects on simulated runoff, 100044 J. Hydrol. X 5. https://doi.org/10.1016/j.hydroa.2019.100044.

Brunetti, G., Papagrigoriou, I.-A., Stumpp, C., 2020. Disentangling model complexity in green roof hydrological analysis: a Bayesian perspective. Water Res. 182, 115973. https://doi.org/10.1016/j.watres.2020.115973.

Brunetti, G., Šimůnek, J., Piro, P., 2016. A comprehensive analysis of the variably saturated hydraulic behavior of a green roof in a Mediterranean climate. Vadose Zone J. 15 https://doi.org/10.2136/vzj2016.04.0032.

Bursztа-Adamiak, E., Mrowiec, M., 2013. Modelling of green roofs' hydrologic performance using EPA's SWMM. Water Sci. Technol. 68, 36–42. https://doi.org/10.2166/wst.2013.219.

Carbone, M., Garofalo, G., Nigro, G., Piro, P., 2014. A Conceptual model for predicting hydraulic behaviour of a green roof. In: Procedia Engineering, 12th International Conference on Computing and Control for the Water Industry, CCWI2013 70, pp. 266–274. https://doi.org/10.1016/j.proeng.2014.02.030.

Carson, T., Keeley, M., Marasco, D.E., McGillis, W., Culligan, P., 2017. Assessing methods for predicting green roof rainfall capture: a comparison between full-scale observations and four hydrologic models. Urban Water J. 14 (6), 589–603. https://doi.org/10.1080/1573062X.2015.1056742.

Carson, T.B., Marasco, D.E., Culligan, P.J., McGillis, W.R., 2013. Hydrological performance of extensive green roofs in New York City: observations and multi-year modeling of three full-scale systems. Environ. Res. Lett. 8 (2), 024036. https://doi.org/10.1088/1748-9326/8/2/024036.

Cascone, S., Coma, J., Gagliano, A., Pérez, G., 2019. The evapotranspiration process in green roofs: a review. Build. Environ. 147, 337–355. https://doi.org/10.1016/j.buildenv.2018.10.024.

Cipolla, S.S., Maglionico, M., Stojkov, I., 2016. A long-term hydrological modelling of an extensive green roof by means of SWMM. Ecol. Eng. 95, 876–887. https://doi.org/10.1016/j.ecoleng.2016.07.009.

Collenteur, R., Vremec, M., Brunetti, G., 2020. Interfacing FORTAN Code with Python: an example for the Hydrus-1D model. In: Presented at the EGU General Assembly 2020, Copernicus Meetings. https://doi.org/10.5194/egusphere-egu2020-15377.

De-Ville, S., Menon, M., Jia, X., Reed, G., Stovin, V., 2017. The impact of green roof ageing on substrate characteristics and hydrological performance. J. Hydrol. 547, 332–344. https://doi.org/10.1016/j.jhydrol.2017.02.006.

De-Ville, S., Menon, M., Jia, X., Stovin, V., 2018a. A longitudinal microcosm study on the effects of ageing on potential green roof hydrological performance. Water (Switzerland) 10 (6), 784. https://doi.org/10.3390/w10060784.

De-Ville, S., Menon, M., Stovin, V., 2018b. Temporal variations in the potential hydrological performance of extensive green roof systems. J. Hydrol. 558, 564–578. https://doi.org/10.1016/j.jhydrol.2018.01.055.

DHI, 2019a. MIKE SHE Volume 1: User Guide. Danish Hydrological Institute, Hørsholm, Denmark.

DHI, 2019b. MIKE SHE Volume 2: Reference Guide. Danish Hydrological Institute, Hørsholm, Denmark.

Ebrahimian, A., Wadzuk, B., Traver, R., 2019. Evapotranspiration in green stormwater infrastructure systems. Sci. Total Environ. 688, 797–810. https://doi.org/10.1016/j.scitotenv.2019.06.256.

Evin, G., Kavetski, D., Thyer, M., Kuczera, G., 2013. Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration: technical note. Water Resour. Res. 49 (7), 4518–4524. https://doi.org/10.1002/wrcr.20284.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resour. Res. 50 (3), 2350–2375. https://doi.org/10.1002/2013WR014185.

Fassman-Beck, E., Voyde, E., Simcock, R., Hong, Y.S., 2013. 4 Living roofs in 3 locations: does configuration affect runoff mitigation? J. Hydrol. 490, 11–20. https://doi.org/10.1016/j.jhydrol.2013.03.004.

Feng, Y., Burian, S., 2016. Improving evapotranspiration mechanisms in the U.S. environmental protection agency's storm water management model, 06016007 J. Hydrol. Eng. 21 (10). https://doi.org/10.1061/(ASCE)HE.1943-5584.0001419.

Getter, K.L., Rowe, D.B., Andresen, J.A., 2007. Quantifying the effect of slope on extensive green roof stormwater retention. Ecol. Eng. 31 (4), 225–231. https://doi.org/10.1016/j.ecoleng.2007.06.004.

Green, WH., Ampt, G.A., 1911. Studies on soil phyics. J. Agric. Sci. 4 (1), 1–24. https://doi.org/10.1017/S0021859600001441.

Hakimdavar, R., Culligan, P.J., Finazzi, M., Barontini, S., Ranzi, R., 2014. Scale dynamics of extensive green roofs: quantifying the effect of drainage area and rainfall characteristics on observed and modeled green roof hydrologic performance. Ecol. Eng. 73, 494–508. https://doi.org/10.1016/j.ecoleng.2014.09.080.

Hamouz, V., Muthanna, T.M., 2019. Hydrological modelling of green and grey roofs in cold climate with the SWMM model. J. Environ. Manage. 249, 109350. https://doi.org/10.1016/j.jenvman.2019.109350.

Herrera, J., Flamant, G., Gironás, J., Vera, S., Bonilla, C.A., Bustamante, W., Suárez, F., 2018. Using a hydrological model to simulate the performance and estimate the runoff coefficient of green roofs in semiarid climates. Water 10, 198. https://doi.org/10.3390/w10020198.

Hilten, R.N., Lawrence, T.M., Tollner, E.W., 2008. Modeling stormwater runoff from green roofs with HYDRUS-1D. J. Hydrol. 358 (3-4), 288–293. https://doi.org/10.1016/j.jhydrol.2008.06.010.

HyQuest, 2016. TB1L & TB/0.5L Tipping Bucket Flow Gauge [WWW Document]. HyQuest Solutions Europe. URL <https://www.hyquestsolutions.eu/products/hardware/water-flow/tb1l-tb05l-tipping-bucket-flow-gauge> (accessed 11.24.20).

Jahanfar, A., Drake, J., Sleep, B., Gharabaghi, B., 2018. A modified FAO evapotranspiration model for refined water budget analysis for Green Roof systems. Ecol. Eng. 119, 45–53. https://doi.org/10.1016/j.ecoleng.2018.04.021.

Johannessen, B.G., Hamouz, V., Gragne, A.S., Muthanna, T.M., 2019. The transferability of SWMM model parameters between green roofs with similar build-up. J. Hydrol. 569, 816–828. https://doi.org/10.1016/j.jhydrol.2019.01.004.

Karlsson, I.B., Sonnenborg, T.O., Refsgaard, J.C., Trolle, D., Børgesen, C.D., Olesen, J.E., Jeppesen, E., Jensen, K.H., 2016. Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change. J. Hydrol. 535, 301–317. https://doi.org/10.1016/j.jhydrol.2016.01.069.

Kasmin, H., Stovin, V.R., Hathway, E.A., 2010. Towards a generic rainfall-runoff model for green roofs. Water Sci. Technol. 62, 898–905. https://doi.org/10.2166/wst.2010.352.

Krebs, G., Kuoppamäki, K., Kokkonen, T., Koivusalo, H., 2016. Simulation of green roof test bed runoff. Hydrol. Process. 30 (2), 250–262. https://doi.org/10.1002/hyp.v30.210.1002/hyp.10605.

Kuczera, G., Kavetski, D., Franks, S., Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. J. Hydrol. 331 (1-2), 161–177. https://doi.org/10.1016/j.jhydrol.2006.05.010.

Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. Water Resour. Res. 48 (1) https://doi.org/10.1029/2011WR010608.

Leimgruber, J., Krebs, G., Camhy, D., Muschalla, D., 2018. Sensitivity of model-based water balance to low impact development parameters. Water 10, 1838. https://doi.org/10.3390/w10121838.

Li, S., Qin, H., Peng, Y., Khu, S.T., 2019. Modelling the combined effects of runoff reduction and increase in evapotranspiration for green roofs with a storage layer. Ecol. Eng. 127, 302–311. https://doi.org/10.1016/j.ecoleng.2018.12.003.

Li, Y., Babcock, R.W., 2015. Modeling hydrologic performance of a green roof system with HYDRUS-2D, 04015036 J. Environ. Eng. 141 (11). https://doi.org/10.1061/(ASCE)EE.1943-7870.0000976.

Liu, W., Feng, Q., Chen, W., Wei, W., Deo, R.C., 2019. The influence of structural factors on stormwater runoff retention of extensive green roofs: new evidence from scale-based models and real experiments. J. Hydrol. 569, 230–238. https://doi.org/10.1016/j.jhydrol.2018.11.066.

Locatelli, L., Mark, O., Mikkelsen, P.S., Arnbjerg-Nielsen, K., Bergen Jensen, M., Binning, P.J., 2014. Modelling of green roof hydrological performance for urban drainage applications. J. Hydrol. 519, 3237–3248. https://doi.org/10.1016/j.jhydrol.2014.10.030.

Machac, D., Reichert, P., Rieckermann, J., Albert, C., 2016. Fast mechanism-based emulator of a slow urban hydrodynamic drainage simulator. Environ. Modell. Software 78, 54–67. https://doi.org/10.1016/j.envsoft.2015.12.007.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., Kuczera, G., 2017. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. Water Resour. Res. 53, 2199–2239. https://doi.org/10.1002/2016WR019168.

Mentens, J., Raes, D., Hermy, M., 2006. Green roofs as a tool for solving the rainwater runoff problem in the urbanized 21st century? Landscape Urban Plann. 77 (3), 217–226. https://doi.org/10.1016/j.landurbplan.2005.02.010.

Métadier, M., Bertrand-Krajewski, J.-L., 2012. Interest of Bayesian learning principle for stormwater quality modelling based on turbidity time series. In: Presented at the 9th International Conference on Urban Drainage Modelling, p. 11.

Oliveira, D.Y., Chaffe, P.L.B., Sá, J.H.M., 2018. Extending the applicability of the generalized likelihood function for zero-inflated data series. Water Resour. Res. 54 (3), 2494–2506. https://doi.org/10.1002/wrcr.v54.310.1002/2017WR021560.

Palla, A., Gnecco, I., 2020. A continuous simulation approach to quantify the climate condition effect on the hydrologic performance of green roofs. Urban Water J. 17 (7), 609–618. https://doi.org/10.1080/1573062X.2019.1700287.

Palla, A., Gnecco, I., Lanza, L.G., 2012. Compared performance of a conceptual and a mechanistic hydrologic models of a green roof. Hydrol. Process. 26 (1), 73–84. https://doi.org/10.1002/hyp.v26.110.1002/hyp.8112.

Palla, A., Gnecco, I., Lanza, L.G., 2009. Unsaturated 2D modelling of subsurface water flow in the coarse-grained porous matrix of a green roof. J. Hydrol. 379 (1-2), 193–204. https://doi.org/10.1016/j.jhydrol.2009.10.008.

Peng, Z., Smith, C., Stovin, V., 2020. The importance of unsaturated hydraulic conductivity measurements for green roof detention modelling. J. Hydrol. 590, 125273. https://doi.org/10.1016/j.jhydrol.2020.125273.

Peng, Z., Smith, C., Stovin, V., 2019. Internal fluctuations in green roof substrate moisture content during storm events: monitored data and model simulations. J. Hydrol. 573, 872–884. https://doi.org/10.1016/j.jhydrol.2019.04.008.

Peng, Z., Stovin, V., 2017. Independent validation of the SWMM green roof module. J. Hydrol. Eng. 22 (9), 04017037. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001558.

Principato, F., 2015. Vegetated Roofs as a Low Impact Development (LID) Approach: Hydrologic and Hydraulic Modeling for Stormwater Runoff Mitigation in Urban Environment. University of Calabria.

Rassam, D., Šimůnek, J., Mallants, D., van Genuchten, M.T., 2018. The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media: Tutorial, version 1.00. CSIRO Land and Water, Australia.

Rawls, W.J., Brakensiek, D.L., Miller, N., 1983. Green-ampt infiltration parameters from soils data. J. Hydraul. Eng. 109 (1), 62–70. https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(62).

Rossman, L.A., Huber, W.C., 2016. Storm Water Management Model Reference Manual Volume III – Water Quality. National Risk Management Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, USA.

Sandoval, S., Filippi, R., Houssin, E., Beauvisage, L., Bournique, R., Bertrand-Krajewski, J.-L., 2019. A simulation tool for comparing the hydrological performance of various associated stormwater source-control techniques at the scale of buildings and blocks. Proceedings of Novatech, Presented at the Novatech 2019. Graie, Lyon, France.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resour. Res. 46 (10) https://doi.org/10.1029/2009WR008933.

She, N., Pang, J., 2010. Physically based green roof model. J. Hydrol. Eng. 15 (6), 458–464. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000138.

Sherrard, J.A., Jacobs, J.M., 2012. Vegetated roof water-balance model: experimental and model results. J. Hydrol. Eng. 17 (8), 858–868. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000531.

Shockley, E.M., Vrugt, J.A., Lopez, C.F., 2018. PyDREAM: high-dimensional parameter inference for biological models in python. Bioinformatics 34, 695–697. https://doi.org/10.1093/bioinformatics/btx626.

Sims, A.W., Robinson, C.E., Smart, C.C., O'Carroll, D.M., 2019. Mechanisms controlling green roof peak flow rate attenuation. J. Hydrol. 577, 123972. https://doi.org/10.1016/j.jhydrol.2019.123972.

Šimůnek, J., Šejna, M., Saito, H., Sakai, M., van Genuchten, M.Th., 2018. The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media.

Šimůnek, J., Genuchten, M.T., Šejna, M., 2008. Development and applications of the HYDRUS and STANMOD software packages and related codes. Vadose Zone J. 7 (2), 587–600. https://doi.org/10.2136/vzj2007.0077.

Skala, V., Dohnal, M., Votrubová, J., Jelínková, V., 2019. The use of simple hydrological models to assess outflow of two green roofs systems. Soil Water Res. 14 (2), 94–103.

Skala, V., Dohnal, M., Votrubova, J., Vogel, T., Dusek, J., Sacha, J., Jelinkova, V., 2020. Hydrological and thermal regime of a thin green roof system evaluated by physically-based model, 126582 Urban For. Urban Green. 48. https://doi.org/10.1016/j.ufug.2020.126582.

Soulis, K.X., Valiantzas, J.D., Ntoulas, N., Kargas, G., Nektarios, P.A., 2017. Simulation of green roof runoff under different substrate depths and vegetation covers by coupling a simple conceptual and a physically based hydrological model. J. Environ. Manage. 200, 434–445. https://doi.org/10.1016/j.jenvman.2017.06.012.

Stovin, V., Vesuviano, G., Kasmin, H., 2012. The hydrological performance of a green roof test bed under UK climatic conditions. J. Hydrol. 414–415, 148–161. https://doi.org/10.1016/j.jhydrol.2011.10.022.

Susca, T., 2019. Green roofs to reduce building energy use? A review on key structural factors of green roofs and their effects on urban climate. Build. Environ. 162, 106273. https://doi.org/10.1016/j.buildenv.2019.106273.

Szota, C., Fletcher, T.D., Desbois, C., Rayner, J.P., Williams, N.S.G., Farrell, C., 2017. Laboratory tests of substrate physical properties may not represent the retention capacity of green roof substrates in situ. Water 9, 920. https://doi.org/10.3390/w9120920.

Teledyne ISCO, 2010. 2150 Area Velocity Flow Module and Sensor: Installation and Operation Guide.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: a case study using Bayesian total error analysis. Water Resour. Res. 45 (12) https://doi.org/10.1029/2008WR006825.

Tomson, M., Kumar, P., Barwise, Y., Perez, P., Forehead, H., French, K., Morawska, L., Watts, J.F., 2021. Green infrastructure for air quality improvement in street canyons. Environ. Int. 146, 106288. https://doi.org/10.1016/j.envint.2020.106288.

van Genuchten, M.T., 1980. A Closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. Soil Sci.Soc. Am. J. 44 (5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x.

Versini, P.-A., Ramier, D., Berthier, E., de Gouvello, B., 2015. Assessment of the hydrological impacts of green roof: from building scale to basin scale. J. Hydrol. 524, 562–575. https://doi.org/10.1016/j.jhydrol.2015.03.020.

Vesuviano, G., Sonnenwald, F., Stovin, V., 2014. A two-stage storage routing model for green roof runoff detention. Water Sci. Technol. 69, 1191–1197. https://doi.org/10.2166/wst.2013.808.

Vesuviano, G., Stovin, V., 2013. A generic hydrological model for a green roof drainage layer. Water Sci. Technol. 68, 769–775. https://doi.org/10.2166/wst.2013.294.

Villarreal, E.L., Bengtsson, L., 2005. Response of a Sedum green-roof to individual rain events. Ecol. Eng. 25 (1), 1–7. https://doi.org/10.1016/j.ecoleng.2004.11.008.

Wani, O., Scheidegger, A., Carbajal, J.P., Rieckermann, J., Blumensaat, F., 2017. Parameter estimation of hydrologic models using a likelihood function for censored and binary observations. Water Res. 121, 290–301. https://doi.org/10.1016/j.watres.2017.05.038.

Xie, H., Wu, Y., Wang, L., Luo, W., Zhou, W., Zhou, H., Yan, Y., Liu, J., 2020. Comparing simulations of green roof hydrological processes by SWMM and HYDRUS-1D. Water Supply 20, 130–139. https://doi.org/10.2166/ws.2019.140.

Zotarelli, L., Dukes, M.D., Romero, C.C., Migliaccio, K.W., 2010. Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method) 10.