Environ Monit Assess (2020) 192: 374
https://doi.org/10.1007/s10661-020-08338-7

Automated calibration of the EPA-SWMM model for a small suburban catchment using PEST: a case study

Roberto Perin · Matteo Trigatti · Matteo Nicolini · Marina Campolo · Daniele Goi

Received: 28 June 2019 / Accepted: 4 May 2020 / Published online: 16 May 2020

Abstract Rainfall-runoff models must be calibrated and validated before they can be used for urban stormwater management. Manual calibration is very difficult and time-consuming due to the large number of model parameters that must be estimated concurrently. Automatic calibration offers as a promising alternative, ideally supporting a user-independent and time-efficient approach to model parameters estimation. In this article, we test the use of a state-of-the-art standard package (PEST, Parameter ESTimation, http://www.pesthomepage.org/) for the automatic calibration of a rainfall-runoff EPA-SWMM (Storm Water Management Model) model developed for a small suburban catchment. Results reported in the paper demonstrate that the performance of automatically calibrated models still depends on a number of user-dependent choices (the level of catchment discretization, the selection of significant parameters, the optimization techniques adopted). Through a systematic analysis of the results, we try to identify the guidelines for the effective use of automatic calibration procedures based on modeling assumptions and target of the analysis.

Keywords Automated calibration · EPA-SWMM · Global and local optimization methods · PEST · Urban stormwater modeling

R. Perin (✉) · M. Trigatti · M. Nicolini · M. Campolo · D. Goi
Dipartimento Politecnico di Ingegneria e Architettura, Università di Udine, Via del Cotonificio 108, Udine, Italy
e-mail: roberto.perin@uniud.it

Introduction

Nowadays accurate and realistic physically based models of many hydrological processes (e.g., the response of urban catchments, the buildup/wash off process, the propagation of pollutants through sewage systems) are available off the shelf. Once properly calibrated, they can be used effectively to solve practical problems in urban stormwater management. This is provided that a number (typically large) of free model parameters are properly estimated, of which only a few are the most relevant for simulating the physical process accurately (Duan et al. 1994; Xu et al. 2019). Some model parameters may be directly measured, and their value can be fixed through onsite investigation. Others, which are only conceptual representations of catchment's features, must necessarily be determined through a trial and error process, until a satisfactory matching is obtained between the model output and measured data (Pauwels 2008).

Uncertainties in hydrologic modeling calibration may arise from model structure, type of parameters, initial conditions, and observational data available to calibrate and validate the model (Liu and Gupta 2007). Model complexity, availability, and accuracy of data and modeler's expertise are critical factors for successful calibration (Dent et al. 2004). This is particularly true when traditional (i.e., manual) trial and error approaches are used. Manual calibration requires both extensive knowledge of the model's numerical formulation and of the system to be modeled: it may lead to good results when performed by an experienced user; yet, it is user-

driven and can be very time-consuming (Duan et al. 1994; Yapo et al. 1998). Moreover, due to the high number of self-interacting calibration parameters, it is difficult to perform a full sensitivity analysis which may prevent from a suboptimal parameter estimation.

The development and use of automatic calibration procedures for estimating model parameters might represent a viable alternative, potentially leading to time savings and to a more user-independent approach. Automatic calibration procedures perform a self-driven search in the model parameter space aimed at minimizing the deviation between model results and experimental data (see Bahrami et al. 2019). Depending on their ability to explore confined/extended portions of the parameter space, they can be classified as either local or global search strategies. The local search can be direct/gradient-based depending on the error surface climbing strategy adopted. Global search methods generally work with a random set (population) of possible values for model parameters. They use deterministic and stochastic rules to converge the population into the subregion(s) of the parameter space where the model error is minimized (Blasone et al. 2007; Bahrami et al. 2019). Due to the number of model parameters, the error surface may exhibit many local minima; in this case, local search methods may be less efficient than global ones because the optimized value of parameters depend on the search starting point (Madsen and Jacobsen 2001), and there is a risk to be trapped into a local minimum (Skahill and Doherty 2006).

In order to evaluate the performance achievable by automated calibration tools, in this paper, we coupled EPA-SWMM (release, 5.1.012) with the model-independent parameter estimation PEST (Parameter ESTimation). PEST implements a particularly robust variant of the Gauss-Marquardt-Levenberg method of nonlinear parameter estimation adjusting the parameters until the discrepancies between selected model outputs and a complementary set of field measurements is reduced to a minimum in the weighted least squares sense (http://www.pesthomepage.org/); it is well documented, free software, and no programming is required since PEST exploits the model's own input and output files as interface for model optimization (Doherty et al. 2010).

A case study is used to discuss the different steps involved in the setup and calibration of a hydrological-hydraulic numerical model. The paper is organized as follows: first, it presents a description of the study area and data sets used; then it describes both rainfall-runoff numerical models and calibration methods. Results of the application of the calibration algorithms are presented together with a sensitivity analysis of PEST control data. Conclusions that can be drawn from the analysis are discussed at the end of the paper.

Study area

The study was performed in Galleriano di Lestizza (Fig. 1), a small suburban catchment (36 ha) characterized by a predominantly flat topography (elevations in the range of 39 to 44 m above mean sea level) located in the middle of the Friuli Venezia Giulia plain in the North East of Italy. The area is a rural/residential: a modest portion of the basin is impervious, and the drainage network conveys both sanitary wastewater and stormwater through the same pipes.

During periods of moderate to heavy rainfall, the capacity of the drainage system to convey the combined flow can be exceeded, resulting in combined sewer overflows (CSOs). In such cases, stormwater is discharged into infiltration ponds. To model the urban rainfall-runoff process for this basin, sewer network geometry and topology was identified via field surveys and GPS data collection. Pipe invert levels, slope, diameter, length, and the coordinates of each manhole were stored in Excel file format. Results of the survey are shown in Fig. 2 and summarized in Table 1. In particular, the suburban catchment can be divided into two macro-catchments A and B (shown in Fig. 2 in red and black, respectively), each with its own sewer network and main conduits (7 and 12, respectively).

Main conduits converge into a concrete circular pipe (length 95 m, diameter 1.00 m) conveying both sanitary wastewater and stormwater to the last hydraulic node of the drainage network, which includes a combined sewer overflow and the pumping station to the treatment plant.

Meteorological and hydrological data

Meteorological and hydrological data gathered from March 2017 to May 2017 were used in this study. As shown in Fig. 1, a rain gauge (Tecnopenta Rain Gauge, M1_PLUV model) installed inside the study area was used to collect rainfall data. A submerged ultrasonic sensor (SIGMA 950 A/V) was installed in the closure section of the basin to measure both velocity and water level. The sampling time was set to 1 min to fully resolve the short concentration time during wet periods,

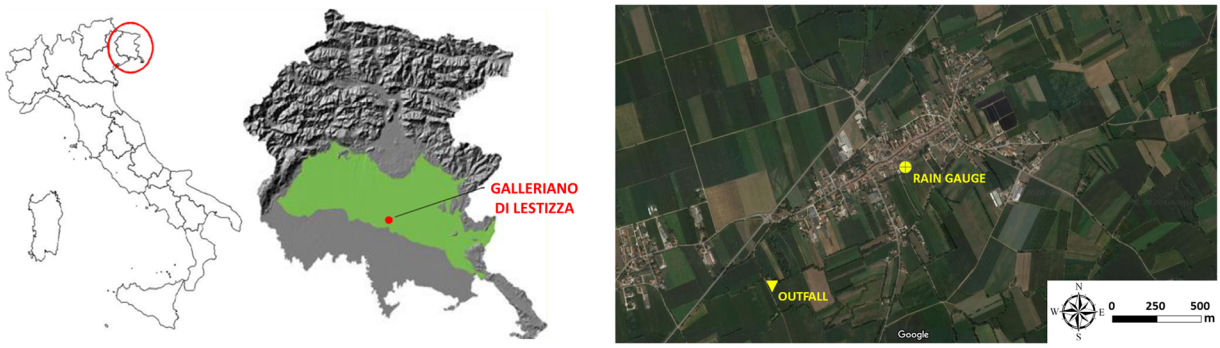

Fig. 1 Basin of Galleriano di Lestizza (North East of Italy)

typical of small suburban catchments. Rain gauge and flowmeter were synchronized to the same reference time. Figure 3 shows the two events recorded.

Methodology

EPA-SWMM model

The EPA Storm Water Management Model (SWMM) is a dynamic hydrologic-hydraulic model used to simulate the quantity and quality of runoff from urban areas. It is a widely accepted tool for planning, analysis and design of stormwater runoff, combined and sanitary sewers, and other drainage systems (Rossman and Huber 2016; Alamdari and Sample 2019; Ngamalieu-Nengoue et al. 2019; Zhou et al. 2019). The runoff module calculates the response of subcatchment areas to precipitation, generating runoff and pollutant loads using a nonlinear reservoir routine. Subcatchments are units of land in which topography and drainage system drive surface runoff to a single discharge point. The routing module conveys runoff through the drainage network (a system of pipes, storage/treatment devices, pumps, and regulators) tracking the flow rate, flow depth, and quality of water along each element of the system. The drainage network is modeled as a series of nodes connected by links. Nodes define the elevation of the drainage system and the time-varying hydraulic head applied at the end of each link. Link elements (pipes, channels, pumping stations, orifices, and weirs) control the flow transferred from one node to the next. The flow transported through links and nodes is ultimately discharged from a final outfall node. Different boundary conditions can be applied at the outfall, including free discharge, fixed water surface, and time-varying water surface.

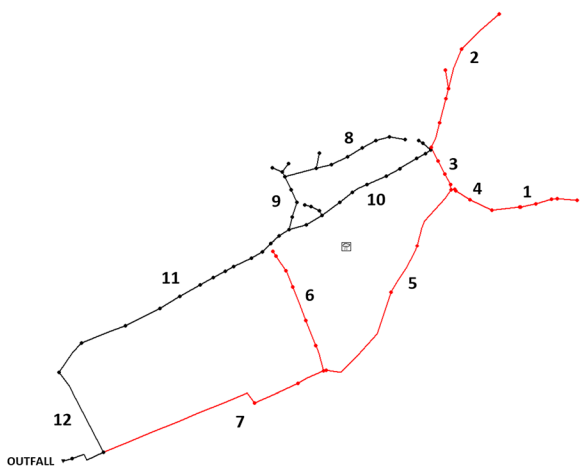

Fig. 2 SWMM representation of Galleriano di Lestizza drainage system network

SWMM schematizes each subcatchment as a rectangular surface that has a uniform slope (S) and width (W): equations used to generate surface runoff are developed on the basis of the idealized rectangular subcatchment. This schematization is very crude compared with the real catchment geometry. One of the most difficult task when modeling an urban basin is therefore how to divide a study area into an appropriate number of subcatchments, identifying the outlet point of each one, since there is some subjectivity in this process. On the other hand, modeling the hydraulic network is driven by survey data and does not require any particular skill. In this paper, PEST is used to drive the automated calibration of SWMM to check model ability in simulating stormwater runoff at the outfall when different levels of catchment discretization and different parameterization are used.

Table 1 Geometric data of drainage network

Macro-catchment A				Macro-catchment B			
Id.	Conduit	Length (m)	Diameter (m)	Id.	Conduit	Length (m)	Diameter (m)
1	Via Castelliere	236	0.40	8	Via del Lavie	355	0.40
2	Via Gorizia (north)	419	0.40–0.50	9	Via Trento	165	0.60
3	Via Trieste (north)	127	0.50	10	Via Gorizia (south)	447	0.40–0.60
4	Via Trieste (south)	129	0.50	11	Via San Giovanni	713	0.60–0.80
5	Via della Viuzza	633	0.60	12	Main conduit B	350	0.80
6	Via Asmara	367	0.30–0.50	Tot.	2030	6	
7	Main conduit A	665	0.60–0.80				
	Tot.	2576					

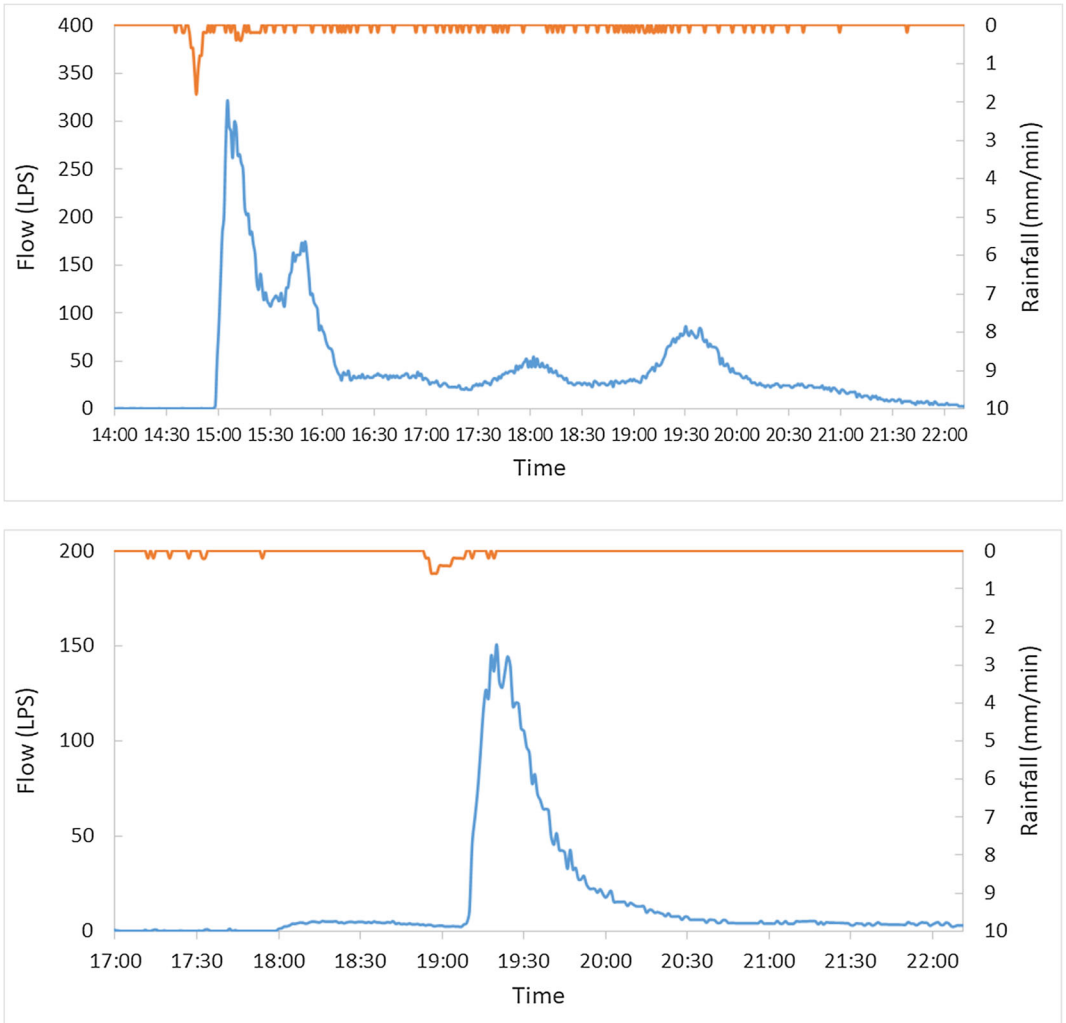

Fig. 3 Rainfall-runoff events recorded during 14 May and 30 May in Galleriano di Lestizza basin

Model setup

Figure 2 shows model layout of Galleriano di Lestizza drainage system. It consists of 78 nodes, 72 links, and 1 outfall node (see Table 1). In this work, we used two different levels of discretization to identify basin subcatchments: at the first level of discretization (FLoD), drainage boundaries were defined based on the use of topographic data; at the second level of discretization (SLoD), subcatchments were defined with no attention to land use or spatial variability in land features. Pervious areas (lawns, fields) were not considered because they are not connected to the sewer system. Figure 4 shows the catchment discretization resulting from FLoD approach. Digital cartography and orthophotos were used to map all the impervious areas. We defined two types of land use: public or private paved roads (gray) and buildings (red and green). Based on this methodology, 142 subcatchments were identified, 93 of which (about 3.35 ha overall) are buildings while 49 (2.31 ha overall) are public/private paved roads. The 93 subcatchments representing buildings were further divided into two groups: 24 of them, whose connection to the urban drainage system is certain since buildings are adjacent to the public street, were included in the model; the remaining 69, whose connection to the urban drainage system is not certain since they are in low density residential zones, were not included. The basin is characterized by a very short concentration time; hence, potential infiltration into the conduits at longer times (due to rising groundwater levels and defective pipe joints or broken pipes) was considered not significant at the model time scale.

The catchment discretization resulting from SLoD approach is based on 54 subcatchments (36 ha overall, corresponding to the entire suburban catchment area), one for each of the nodes of the drainage network (excluding those used to model conduits without inlets from overland flow or main pipes, like main conduit 7 and 12, as indicated in Table 1). In this case, each subcatchment includes both impervious and pervious areas, yet no infiltration module was used in modeling the rainfall-runoff process to be consistent with FloD assumptions. Excess rainfall was calculated by using a runoff coefficient (C) that is related to the subcatchment imperviousness parameter.

Parameter estimates

Rainfall-runoff models typically involve a number of parameters which need to be fixed by the user during model setup. The simpler is the model, the smaller is the number of parameters and the easier is to perform a well-posed calibration procedure. When a large number of parameters is used, e.g., to account for the spatial variability of the process, calibration procedures become more complex (Madsen 2000; Madsen and Jacobsen 2001); according to the principle of parsimony (Hill 1998), the well posedness of the calibration process can still be ensured if the number of calibration parameters is reduced, i.e., only a subset of model parameters is selected for calibration. Parameters should be selected based on a sensitivity analysis to identify the ones contributing much to model output determination.

In this paper, model parameters' needing calibration are subcatchment data such as imperviousness (Imp), width (W), slope (Slope), and roughness (Roughness) that significantly affect hydrograph shape and volume. Their number obviously depends on the subcatchment discretization adopted (FLoD or SLoD model). In this specific implementation, subcatchment imperviousness (Imp) affects hydrograph volume, whereas width, slope,

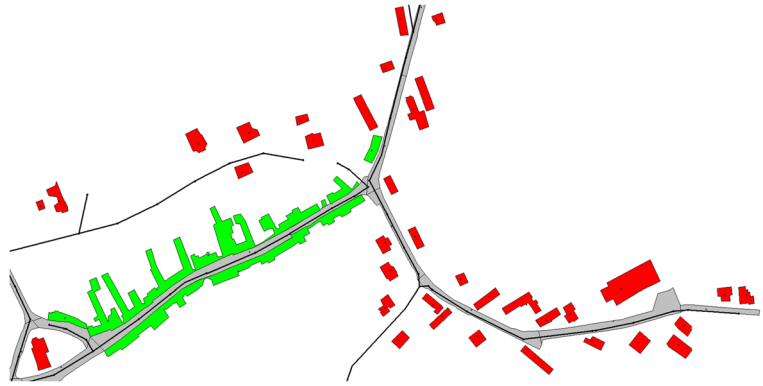

Fig. 4 Example of FLoD subcatchment discretization (gray, public or private paved roads; red, buildings whose connection with urban drainage system is not certain; green, buildings whose connection with the urban drainage system is certain; white, pervious areas)

and roughness affect hydrograph shape. Table 2 summarizes the parameters selected for calibration. Some were not included in the set of calibration parameters, their value being fixed either from calculation or extrapolated from field surveys or scientific literature (Rossman and Huber 2016; Rossman 2017). Table 2 indicates that in the FLoD model, one single calibration parameter was used to represent imperviousness for all the buildings without a proven connection to the drainage system whereas a distinct value of width, W (set in the range 50–200% of the square root of each subcatchment area), was assigned in order to reflect the spatial heterogeneity of this parameter; due to the flat topography of the basin, one single slope was used as calibration parameter for both public and private road slope, and one calibration parameter was adopted to represent building slope conceptually. The overall number of calibration parameters is 96.

In the SLoD model (including 54 subcatchments), the overall number of calibration parameters is 110 (see Table 2, last row). These included one single value for imperviousness and roughness and 54 different calibration parameters for subcatchment width (W) and slope (Slope) to account for the (potential) spatial variability of data in the simpler model. A reduced range of variation was set for the imperviousness value due to extension of permeable areas. Manning's roughness coefficient for all concrete conduits was set equal to 0.014 in both models.

PEST (Parameter ESTimation)

The EPA-SWMM rainfall-runoff model of Galleriano di Lestizza was automatically calibrated using PEST (Doherty 1994). PEST is a nonlinear parameter estimation and optimization package offering model-independent optimization routines (Liu et al. 2005). A detailed description of PEST can be found in Doherty et al. (2010). As shown in Fig. 5, PEST is able to take control of a model, running it as many times as needed while adjusting its parameters until the deviation between model output and field measurements is minimized in the weighted least squares sense. SWMM/PEST data exchange is done using native input/output files provided by SWMM. For each value of the calibration parameters set, SWMM generates a runoff (flow rate) which is compared against available field measurements. The sum of squares deviation between model output and observed flow rates is calculated as:

$$\phi = \sum_{i=1}^{m} (w_i r_i)^2 \quad (1)$$

where ϕ is the objective function; m is the total number of time steps in calibration period; w_i is the *i-th* observation weight; and r_i is the difference between the simulated flow and actual flow rate for *i-th* observation.

Table 2 Number and value constraints of calibration parameters involved in the first and second level of discretization (FLoD and SloD models)

	FLoD			SLoD
	Subcatchment typology			Subcatchment typology
Parameter typology	Public/private paved roads	Buildings connected to drainage system	Buildings w/o proven connection to drainage system	(No land use criteria adopted for subcatchment ident.)
Imperviousness (%)	Fixed (100)	Fixed (100)	1 (range, 50–100)	1 (range, 1–10)
Width (m)	Fixed (RW*)	24 (range, 50–200% of SRSA**)	69 (range, 50–200% of SRSA**)	54 (range, 30–300% of SRSA**)
Slope (%)	1 (range, 0.1–1.0)	1 (range, 1–10)		54 (range, 0.10–1.0)
Roughness ($m^{-1/3}$ s)	Fixed (0.015)	Fixed (0.030)		1 (range, 0.015–0.10)
Initial abstraction (mm)		Fixed (1.5)		
Number of calibration parameters	96			110

*RW, road width

**SRSA, square root of subcatchment area

The value of the objective function ϕ depends on the value of the weights (w_i) used to sum up the deviation between model output and field observation. If the objective of the hydrologic/hydraulic model is to fit the hydrograph as a whole, a constant value for the weights should be adopted; if the objective is to fit the peaks, a larger value of the weights should be given to these events. PEST standard optimization scheme is based on the iterative application of the Gauss-Marquardt-Levenberg (GML) gradient search algorithm. Multiple runs of the model are required to determine the optimal set of model parameters. The process terminates when a stopping criterion (e.g., maximum number of iterations, convergence of the total objective function, convergence of the parameter set) is satisfied (Liu et al. 2005).

Although PEST is quite efficient in finding the minimum of the objective function, being a gradient based method, it may converge into a local rather than a global minimum (Gupta et al. 2003; Skahill and 2006; Blasone et al. 2007) depending on the initial values of the calibration set. In such cases, PEST optimization algorithm can be coupled with a multi-start method (PD_MS2, PEST Driver Multiple Starting Point2): multiple consecutive calibration runs are conducted to ensure extensive exploration of the parameter space (Poeter and Hill 1997; Skahill and Doherty 2006; Blasone et al. 2007). The PD_MS2 driver combines the strengths of the global optimization method and the speed of the GML method adding a certain degree of randomness into the parameter estimation process (Mancipe et al. 2012).

In this paper, both a conventional and a multi-starting PEST calibration method have been used to compare the performance of local/global optimization methods for the calibration of the FLoD and SLoD models developed for Galleriano di Lestizza catchment.

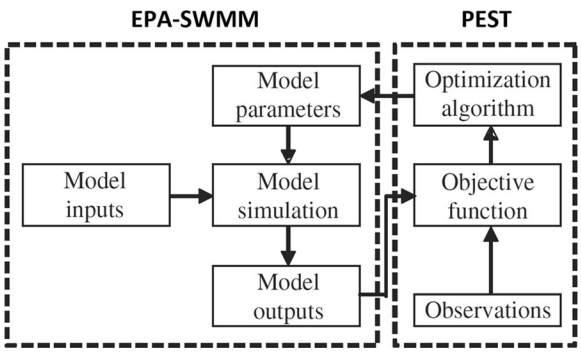

Fig. 5 Schematic of the EPA-SWMM automatic calibration (modified from Liu et al. 2005)

Efficiency criteria used to evaluate calibration performance

Table 3 defines the four efficiency criteria selected to evaluate model calibration performance. They include Nash-Sutcliffe efficiency (NSE) and modified Nash-Sutcliffe efficiency (MNSE) coefficients and relative error in volume (REV) and in peak (REP) which are considered standards indicators to evaluate hydrological models (Chau 2004; da Silva et al. 2015). Each indicator has specific pros and cons (Krause et al. 2005) which should be considered during model calibration and evaluation. The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970), defined as shown in Table 3, is one of the most used in hydrological applications (Biondi et al. 2012). The value of NSE, included in the range $(-\infty,1)$, indicates a perfect fit when NSE = 1; values lower than zero, calculated when the mean value of the observed time series would be a better prediction than the model itself (Krause et al. 2005), indicate poor model performances.

Since the deviation between observed/calculated data is squared before summation, large deviations have a larger impact on NSE value than small deviations (Legates and McCabe 1999), leading to models able to fit peak flows better than low flow conditions (Krause et al. 2005). A logarithmic form of NSE criteria or a modified form of Nash and Sutcliffe (MNSE) in which the deviation between observed/calculated data is elevated to a different power, j, like the one adopted here, can be used to reduce model mismatch at low flow rates. For $j = 1$, MNSE values become lower than NSE, since

Table 3 Efficiency criteria used for model evaluation

Nash-Sutcliffe (NSE)	Modified Nash-Sutcliffe (MNSE)
$NSE = 1-\frac{\sum_{i=1}^{n}(O_i-P_i)^2}{\sum_{i=1}^{n}(O_i-O_m)^2}$	$MNSE = 1-\frac{\sum_{i=1}^{n}\lvert O_i-P_i\rvert^j}{\sum_{i=1}^{n}\lvert O_i-O_m\rvert^j} with\, j\in N$
Relative error in volume (REV)	**Relative error in peak (REP)**
$REV = \frac{V_o-V_s}{V_o}\cdot 100$	$REP = \frac{Q_{po}-Q_{ps}}{Q_{po}}\cdot 100$

O_i, observed flow rate; O_m, average observed flowrate; P_i, predicted simulated flow; V_0, observed runoff volume; V_s, predicted simulated runoff volume; Q_{po}, observed peak flow rate; Q_{ps}, predicted simulated peak flow rate; i, number of time steps in the calibration period

small and large deviations compensate each other. Both NSE and MNSE values are related to the shape of the hydrographs. NSE values higher than 0.70 indicate a good fitting.

The other two indicators defined in Table 3 focus on more specific characteristics of the hydrograph: REV is the relative percent error between observed and simulated runoff volume, normalized by the observed volume, whereas REP is the relative percent error between observed and simulated peak flow, normalized by the observed peak flow. Positive/negative values indicate model under/overestimation. Values of REV in between [± 0.25] and values of REP in between [± 0.20] indicate good fitting (Kean et al. 2008).

EPA-SWMM model calibration and PEST sensitivity analysis

To evaluate PEST performance in the automatic calibration of our suburban rainfall-runoff model, we focused on three factors which may affect performances: (1) catchment discretization level, as represented by FLoD and SLoD models (i.e., number of calibration parameters); (2) use of local/global optimization methods; and (3) use of different objective functions (equal weights, w_e, to fit the hydrograph as a whole versus different weights, w_p, to fit peak flow values).

Whichever the model discretization and number of calibration parameters, PEST performance depends on the values (start values and upper and lower constraints) assumed by calibration parameters and on the control data used to tune PEST internal optimization algorithms (a brief description of PEST control variables considered in this work is available in the Appendix at the end of the paper; for reference values of PEST control data, see Doherty et al. 2010). For these reasons, we performed also a sensitivity analysis to assess the influence of these values on simulation results. Starting from a reference control data set, we changed one of the control data at a time within the range of values suggested by PEST developers, leaving all the others unchanged. Control data whose appropriate values were specifically suggested by PEST developers were not considered (i.e., PHIRATSUF, 0.3; PHIREDLAM, 0.01; FACORIG, 0.001; PHIREDSWH, 0.1; PHIREDSTP, 0.01; NPHISTP, 4; NPHINORED, 4; RELPARSTP, 0.01; NRELPAR, 3; ICOV, 0; ICOR, 0; IEIG, 0; DERINC, 0.01; DERINCLB, 0; DERINCMUL, 1.5). The reference control data set adopted and possible value/type alternatives are shown in Table 4.

Since the logarithmic transformation of some parameters may affect the success of the parameter estimation process, sensitivity analysis of control data always begins with the evaluation of calibration performances when PARTRANS control variable is set to Log or None.

Selection of calibration and validation period

The rainfall-runoff model developed in this work was calibrated over two single events. Monitoring periods (March 2017–May 2017) include runoff events produced by both moderately high and low precipitation events. We decided to use the rainfall event of 14 May 2017 (moderately high flow) for calibration and the event of 30 May 2017 (low flow) for validation.

FLoD model calibration and validation

Table 5 summarizes the variation in efficiency criteria resulting from changes in (a) PARTRANS values (None or Log), (b) use of uniform/non uniform weights (w_p or w_e), and (c) the use of global (GSM) or local (LSM) search methods.

This case study's result shows that the use of the logarithmic transformation (Log) of calibration parameters has no effect on the output. This conclusion should be confirmed by the application of the model on an

Table 4 Reference control data set and possible value constraints/alternatives

Control data	Value/type	Values range/alternatives
RLAMBDA1	5	1–10
RLAMFAC	1	> 1
NUMLAM	7	5–10
RELPARMAX	5	1–5
FACPARMAX	5	1–5
NOPTMAX	25	20–30
INCTYP	Relative	Absolute–rel to Max
FORCEN	Switch	Always2–always3
DERMTHD	Parabolic	Best Fit
PARCHGLIM	Factor	Relative
PARTRANS	None	Log

Table 5 Efficiency criteria values obtained using a reference PEST control data set

FLoD model	Global search method (GSM)				Local search method (LSM)			
	PARTRANS				PARTRANS			
	None		Log		None		Log	
Eff. criteria	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e
NSE	0.90	0.90	0.90	0.90	0.93	0.88	0.93	0.90
MNSE	0.60	0.70	0.60	0.70	0.69	0.72	0.70	0.71
REV (%)	24.60	3.69	24.60	3.96	11.70	− 1.74	11.80	0.70
REP (%)	− 0.57	− 22.43	− 0.57	− 22.80	− 11.80	− 28.30	− 11.90	− 25.10

extended set of additional data. GSM led to very satisfactory REV values when using equal weights (w_e) and to excellent REP values where only peak flow value (w_p) is used in the parameter estimation process. LSM led to very satisfactory REV values using equal weights (w_e) but unsatisfactory REP values when only peak flow data (w_p) are used in the parameter estimation process. LSM with high NSE values (e.g., 0.93) can represent a good compromise in terms of volume and peak flow predictions. GSM can achieve excellent REP (or REV) values but simultaneously unsatisfactory REV (or REP) values.

Table 6 reports the sensitivity analysis to PEST control data when PARTRANS is set to Log and the local search method (LSM) is used. Values of efficiency criteria reported in the last two columns of Table 5 (obtained with PEST control data reference set) are the base case of those reported in Table 6.

Table 6 PEST control data sensitivity analysis using the local search method (LSM)

FLoD model (LSM)	PARTRANS (Log)													
	RLAMBDA				RLAMFAC		NUMLAM				R./FACPARMAX		PARCHGLIM	
	1		10		4		5		10		1		Relative	
Eff. criteria	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e
NSE	0.93	0.90	0.93	0.90	0.93	0.90	0.93	0.90	0.93	0.90	0.93	0.93	0.93	0.90
MNSE	0.69	0.69	0.69	0.70	0.69	0.69	0.69	0.71	0.69	0.69	0.69	0.69	0.69	0.70
REV (%)	11.7	4.90	11.7	4.14	11.7	4.58	11.7	1.89	11.7	4.59	11.7	11.7	11.7	3.74
REP (%)	− 11.9	− 22.4	− 11.9	− 22.6	− 11.9	− 22.4	− 11.9	− 24.3	− 11.9	− 22.4	− 11.9	− 11.9	− 11.9	− 23.1
FLoD model (LSM)	PARTRANS (Log)													
	NOPTMAX				INCTYP				FORCEN				DERMTHD	
	20		30		Absolute		Rel to max		Always2		Always3		Best fit	
Eff. criteria	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e
NSE	0.93	0.90	0.93	0.90	0.93	0.88	0.93	0.92	0.93	0.88	0.93	0.90	0.93	0.90
MNSE	0.69	0.71	0.69	0.69	0.69	0.70	0.69	0.69	0.69	0.72	0.69	0.70	0.69	0.71
REV (%)	11.7	1.87	11.7	5.11	11.7	1.60	11.7	11.1	11.7	−2.02	11.7	2.87	11.7	1.35
REP (%)	− 11.9	− 24.3	− 11.9	− 22.2	− 11.9	− 27.6	− 11.9	− 12.9	− 11.9	− 28.4	− 11.9	− 23.7	− 11.9	− 24.2

Table 6 shows that values of efficiency criteria are insensitive to changes in control parameters when peak flow data (w_p) are used for calibration. LSM with peak flow weighting (w_p) may represent an acceptable compromise in terms of volume and peak flow predictions. Using equal weights (w_e), it is possible to obtain lower REV values but at the expense of high REP values.

Table 7 reports the sensitivity analysis to PEST control data when PARTRANS is set to None, the global search method (GSM) is used. Values of efficiency criteria reported in the second and third columns of Table 5 (obtained with PEST control data reference set) are the base case of those reported in Table 7.

Table 7 shows that values of efficiency criteria do not change (with the exception of three cases: REL/FACPARMAX, 1; INCTYP, Rel to max; DERMTHD, Best Fit) when peak data (w_p) are used for calibration. GSM with peak flow weighting (w_p) can have very low REP values (e.g., − 0.57%), but these are associated with high REV values. Using equal weights (w_e), it is possible to obtain very low REV values (e.g., 1.94%), but these are associated with high REP values (e.g., − 28.6%).

The global optimization process performed by a multi-start method begins by the random selection of a starting point in the parameter space. Uniform random sampling between upper/lower bound of parameters value was used. Multiple model runs are performed until local parameter optimization. A new starting point is then selected, and the search is repeated, converging to a different local optimum. The number of times the process is performed (i.e., the number of points used for pre-inversion random sampling) is usually fixed between 4 and 10 times the number of adjustable parameters. Results discussed so far using the GSM method involved a number of points equal to 4 times the number of adjustable parameters. To test the effect of this parameter, GSM calibration was performed by changing the number of starting points to 2 and 6 times the number of adjustable parameters. In this case, only optimization based on equal weights (w_e) was considered. Results of the sensitivity analysis are shown in Table 8. Base case data are those in Table 7 (w_e columns only).

The sensitivity analysis indicates that increasing the number of random sample of calibration parameters tested leads to a reduction of REV values. This is not the case when INCTYP is set to Rel to max. The base case (× 4) does not give always better results than the minimal case (× 2).

Efficiency criteria reported in Tables 5, 6, 7, and 8 show that global search methods (GSM) allow to achieve excellent calibration performance in terms of REV or REP values whichever the weights distribution (w_e or w_p) adopted. In many cases, it was possible to obtain REP values near zero (e.g., − 0.57%) but also low REV values. Local search methods (LSM) with equal weights (w_e) provided good performance in terms of REV values, but REP values did not achieve satisfactory results in most cases.

Figure 6 illustrates the hydrographs calculated at the outfall of Galleriano di Lestizza by the FLoD model for those calibration parameters which minimize the REP value (− 0.57%), (a) and (b), and those which minimize the REV value (− 0.06%), (c) and (d). Graphs (a) and (c) and (b) and (d) show model results obtained for calibration and validation set, respectively.

If the model objective is to predict high flow conditions (e.g., flood prediction), the set of model calibration parameters corresponding to low REP values is the best choice (Fig. 6, graphs on the top). Conversely, if the main purpose of the model is to predict low flow or the exact overall runoff volume, the set of model calibration parameters corresponding to low REV values (Fig. 6, graphs on the bottom) is the best choice.

Efficiency criteria for the validation set were NSE, 0.88; MNSE, 0.71; REV, 27.29%; and REP, 4.01% for the first set of calibration parameters and NSE, 0.88; MNSE, 0.67; REV, 23.46%; and REP, −17.87% for the second one.

SLoD model calibration and validation

Theoretically, a numerical model with a perfect spatial resolution can be used in a deterministic mode without any calibration (excluding uncertainty related to forcing variables, such as the rainfall spatial distribution) (Duong et al. 2016). As already mentioned, the SLoD model is more simplified than the FLoD model. The spatial discretization of the domain does not consider land use or spatial variability in land features. In other words, while the FLoD model can be considered a distributed model (thanks to its high spatial resolution), the SLoD model is most similar to a lumped model where all of the parameters which impact the hydrologic response of each subcatchment are spatially averaged together to create uniformity inside each subcatchment. Clearly, the more spatially varied are the land

Table 7 PEST control data sensitivity analysis using the global search method (GSM)

FLoD model (GSM)	PARTRANS (None)													
	RLAMBDA				RLAMFAC		NUMLAM				R./FACPARMAX		PARCHGLIM	
	1		10		4		5		10		1		Relative	
Eff. criteria	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e
NSE	0.90	0.91	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.91	0.91	0.87	0.90	0.90
MNSE	0.60	0.69	0.60	0.69	0.60	0.69	0.60	0.70	0.60	0.69	0.70	0.72	0.60	0.69
REV (%)	24.6	6.69	24.6	5.18	24.6	5.29	24.6	3.69	24.6	7.15	4.69	−1.94	24.6	5.41
REP (%)	− 0.57	− 19.5	− 0.57	− 22.0	− 0.57	− 21.9	− 0.57	− 22.4	− 0.57	− 19.1	− 20.7	− 28.6	− 0.57	− 22.0
FLoD model (GSM)	PARTRANS (None)													
	NOPTMAX				INCTYP				FORCEN				DERMTHD	
	20		30		Absolute		Rel to max		Always2		Always3		Best fit	
Eff. criteria	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e	w_p	w_e
NSE	0.90	0.90	0.90	0.90	0.90	0.88	0.92	0.90	0.91	0.90	0.90	0.92	0.91	0.89
MNSE	0.60	0.70	0.60	0.70	0.60	0.71	0.68	0.69	0.70	0.67	0.60	0.68	0.70	0.69
REV (%)	24.6	3.70	24.6	3.68	24.6	−0.43	14.7	4.77	4.69	7.29	24.6	12.1	4.69	3.10
REP (%)	− 0.57	− 22.4	− 0.57	− 22.4	− 0.57	− 27.7	− 8.1	− 22.2	− 20.7	− 22.3	− 0.57	− 12.7	− 20.7	− 26.0

Table 8 Efficiency criteria values obtained for a number of points for pre-inversion random sampling equal to 2 (× 2) and 6 (× 6) times the number of adjustable parameters, assuming equal weights (w_e)

FLoD model (GSM) e. w. f. (w_e)	PARTRANS (None)													
	RLAMBDA				RLAMFAC		NUMLAM				R./FACPARMAX		PARCHGLIM	
	1		10		4		5		10		1		Relative	
Eff. criteria	× 2	× 6	× 2	× 6	× 2	× 6	× 2	× 6	× 2	× 6	× 2	× 6	× 2	× 6
NSE	0.90	0.89	0.89	0.89	0.90	0.88	0.90	0.89	0.90	0.89	0.88	0.88	0.90	0.90
MNSE	0.70	0.69	0.71	0.70	0.70	0.71	0.68	0.69	0.68	0.69	0.72	0.71	0.70	0.69
REV (%)	3.27	3.49	0.26	2.50	3.70	0.24	5.77	3.10	5.75	3.10	− 1.71	− 1.25	4.13	3.83
REP (%)	− 22.1	− 25.7	− 26.7	− 25.9	− 22.3	− 27.9	− 23.1	− 25.6	− 23.1	− 25.6	− 28.7	− 28.4	− 22.4	− 23.8
FLoD model (GSM) e. w. f. (w_e)	PARTRANS (None)													
	NOPTMAX				INCTYP				FORCEN				DERMTHD	
	20		30		Absolute		Rel to max		Always2		Always3		Best fit	
Eff. criteria	× 2	× 6	× 2	×6	× 2	× 6	× 2	× 6	× 2	× 6	× 2	× 6	× 2	× 6
NSE	0.90	0.89	0.90	0.89	0.89	0.88	0.90	0.90	0.90	0.89	0.91	0.89	0.90	0.89
MNSE	0.68	0.69	0.68	0.69	0.70	0.71	0.68	0.67	0.70	0.69	0.68	0.69	0.68	0.69
REV (%)	5.75	3.14	5.75	3.10	1.57	−0.06	6.71	7.70	3.73	3.28	8.31	2.54	5.65	3.03
REP (%)	−23.1	−25.5	−23.1	−25.6	−26.5	−27.9	−20.9	−21.5	−23.2	−26.1	−19.1	−26.2	−23.2	−25.6

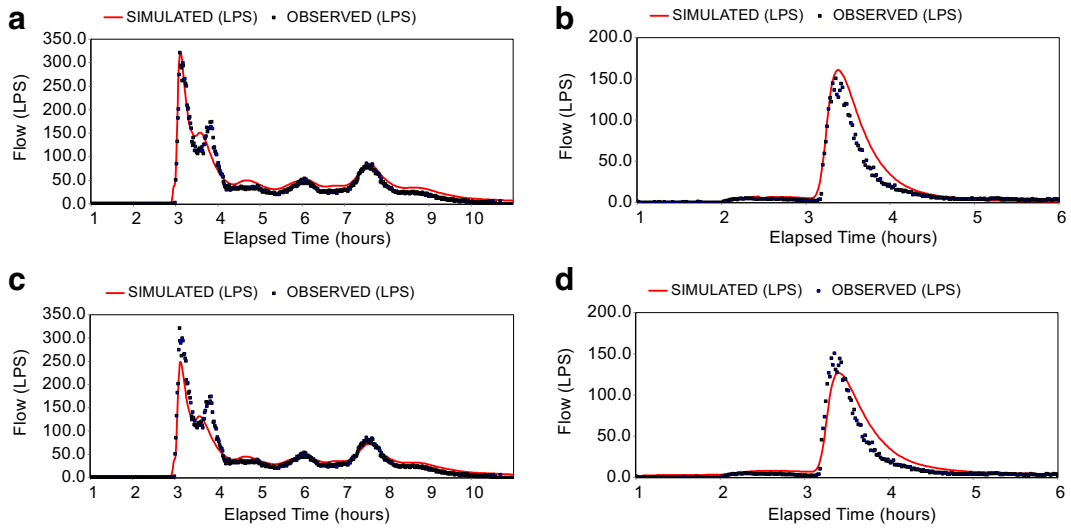

Fig. 6 Comparison between observed data (points) and hydrograph (line) calculated by FLoD model at the outlet of Galleriano basin: model calibration parameters are those corresponding to lower REP values (− 0.57%) as shown in graphs (**a**) and (**b**) and lower REV values (− 0.06%) as shown in graphs (**c**) and (**d**); model results are shown for calibration data in graphs (**a**) and (**c**) and validation data in graphs (**b**) and (**d**)

characteristics, the smaller is calibration effort to achieve good simulation results.

SLoD model sensitivity analysis to calibration parameters started using the global search method (GSM). Only the equal weighting (w_e) approach was used since peak weighting (w_p) provided low values of NSE and MNSE but also high values of REV and REP. Parameters transformation was not further considered, this

Table 9 PEST control data sensitivity analysis using the global search method (GSM)

SLoD model (GSM)	PARTRANS (None)						
	RLAMBDA		RLAMFAC	NUMLAM		R./FACPARMAX	PARCHGLIM
	1	10	4	5	10	1	Relative
Eff. criteria	w_e	w_e	w_e	w_e	w_e	w_e	w_e
NSE	0.86	0.86	0.62	0.85	0.85	0.54	0.83
MNSE	0.67	0.67	0.27	0.67	0.67	0.60	0.68
REV (%)	10.47	10.40	56.5	9.90	9.90	− 24.0	2.14
REP (%)	− 19.5	− 19.5	20.4	− 21.2	− 21.2	− 58.1	− 31.1
SLoD model (GSM)	PARTRANS (None)						
	NOPTMAX		INCTYP		FORCEN		DERMTHD
	20	30	Absolute	Rel to max	Always2	Always3	Best fit
Eff. criteria	w_e	w_e	w_e	w_e	w_e	w_e	w_e
NSE	0.85	0.85	0.55	0.87	0.84	0.87	0.85
MNSE	0.67	0.67	0.62	0.67	0.67	0.67	0.67
REV (%)	9.90	9.90	− 23.6	10.8	7.93	10.9	10.0
REP (%)	− 21.2	− 21.2	− 55.2	− 18.8	− 23.3	− 18.5	− 21.0

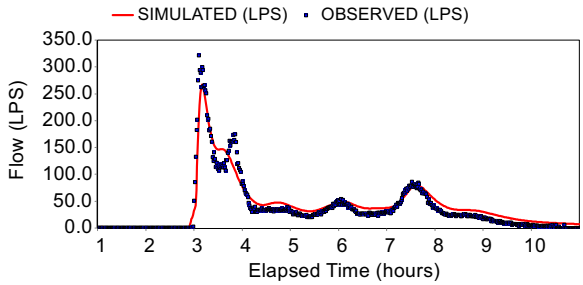

Fig. 7 Observed and SLoD model simulated hydrograph at the outlet of the study catchment using calibration parameters that provide satisfactory compromise in terms of REV (10.8%) and REP (− 18.8%) values

choice resulting in better values for efficiency criteria (NSE, 0.85; MNSE, 0.67; REV, 10.05%; REP, − 21.15% for PARTRANS = None versus NSE, 0.67; MNSE, 0.59; REV, 5.25%; REP, − 34.20% for PARTRANS = Log). Table 9 reports PEST control data sensitivity analysis. Contrary to FLoD model results, in this case the specific setting of PEST control data significantly affects calibration performance. Nevertheless, it was not possible to obtain low REP values in all cases. REV values were generally satisfactory except when the Marquardt lambda adjustment parameter (RLAMFAC) was set equal to 4 (REV, 56.5%); an excellent performance was obtained when the type of parameter change limit (PARCHGLIM) was set to relative (REV, 2.14%). A satisfactory compromise in terms of REV and REP values was obtained when central derivatives calculation (FORCEN, Always3) was used or the method by which parameter increments are calculated (INCTYP) was set to Rel to max (Fig. 7). For all the simulations, the number of points for pre-inversion random sampling was set to 4 times (4×) the number of adjustable parameters.

In particular, Fig. 7 shows the capacity of global search methods (GSM) to adequately calibrate even those numerical models with a very simplified (and inaccurate) spatial discretization, such as SLoD models.

SLoD model sensitivity analysis to calibration parameters was also performed using the local search method (LSM). Also in this case, only an equal weighting (w_e) approach was used because the peak weighting approach (w_p) provided very low values of NSE and MNSE and very high values of REV and REP. Efficiency criteria values were compared on the basis of parameter transformation (PARTRANS) setting, resulting in NSE, 0.61; MNSE, 0.62; REV, − 13.51%; REP, − 55.17% for PARTRANS = Log and NSE, 0.02;

Table 10 PEST control data sensitivity analysis using the local search method (LSM)

SLoD model (LSM)	PARTRANS (Log)						
	RLAMBDA		RLAMFAC	NUMLAM		R./FACPARMAX	PARCHGLIM
	1	10	4	5	10	1	Relative
Eff. criteria	w_e	w_e	w_e	w_e	w_e	w_e	w_e
NSE	0.63	0.61	0.53	0.61	0.61	− 0.18	0.53
MNSE	0.62	0.62	0.60	0.62	0.62	− 0.24	0.59
REV (%)	− 11.1	− 13.6	− 25.9	− 13.6	− 13.6	98.8	− 23.7
REP (%)	− 54.1	− 55.2	− 59.7	− 55.2	− 55.2	39.6	− 61.6
SLoD model (LSM)	PARTRANS (Log)						
	NOPTMAX		INCTYP		FORCEN		DERMTHD
	20	30	Absolute	Rel to max	Always2	Always3	Best fit
Eff. criteria	w_e	w_e	w_e	w_e	w_e	w_e	w_e
NSE	0.61	0.61	0.32	0.50	0.35	0.61	0.33
MNSE	0.62	0.62	0.51	0.59	0.52	0.62	0.51
REV (%)	− 13.5	− 13.5	− 32.8	− 25.7	− 32.7	− 14.7	− 32.8
REP (%)	− 55.2	− 55.2	− 73.1	− 62.5	− 73.2	− 56.3	− 73.1

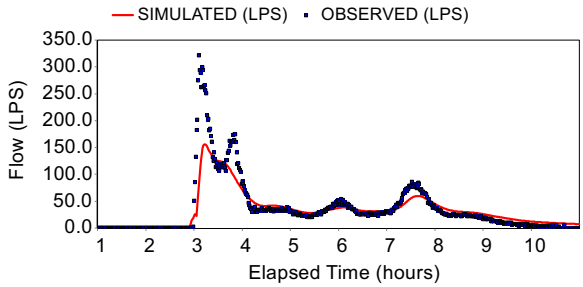

Fig. 8 Observed and simulated hydrograph at the outlet of the study catchment using SLoD model calibration parameters that provide REV, − 11.1% and REP, − 54.1%

MNSE, − 0.18; REV, 85.41%; REP, − 2.01% for PARTRANS = None. PARTRANS = Log was chosen only for further analysis, even if values of efficiency criteria were not satisfactory. For example, it is possible to obtain good REV (or REP) values at the expense of bad REP (or REV). Table 10 reports sensitivity analysis of PEST control data when the local search method (LSM) was used. Comparing Table 9 and Table 10, we observe that global search methods (GSM) outperform local search methods (LSM) when numerical models with a simplified spatial discretization are considered.

Specifically, the calibration set providing the best REV value (REV, − 11.1%) was unsatisfactory in terms of REP values (REP, − 54.1%), as shown in Fig. 8.

Additionally, the calibration set providing the best REP value (REP, − 2.01%) was unsatisfactory in terms of REV values (REV, − 85.41%), as shown in Fig. 9.

SloD model validation (Fig. 10) was performed using a calibration set providing a satisfactory compromise in terms of REV (10.8%) and REP (− 18.8%) values. SLoD model performance on validation data were NSE, 0.75; MNSE, 0.57; REV, 29.9%; and REP, − 4.04%. Despite the simplicity of the SLoD model formulation, the results obtained were quite satisfactory

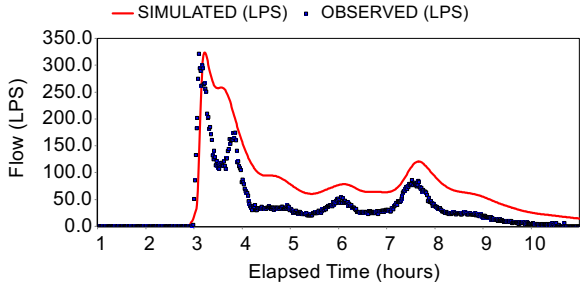

Fig. 9 Observed and simulated hydrograph at the outlet of the study catchment using SLoD model calibration parameters that provide REV, 85.41% and REP, − 2.01%

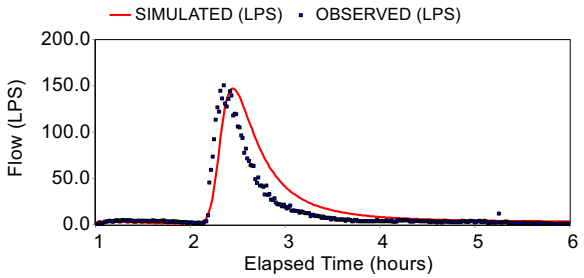

Fig. 10 SLoD model validation results at the outlet of the study catchment using calibration parameters that provided a satisfactory compromise in terms of REV (10.8%) and REP (− 18.8%) values

except for an overestimation of the runoff volume (+ 29.9%) and a lag time (5 min) in the simulated peak hydrograph.

Conclusions

In this work, we tested the use of a state-of-the-art standard package (PEST) for the automatic calibration of a rainfall-runoff EPA-SWMM model developed for the small suburban catchment of Galleriano di Lestizza.

Three factors which may affect calibrated model performances have been considered: (1) catchment discretization level (represented by the two different formulations of the model, FLoD and SLoD, which differ in the number of calibration parameters); (2) use of local/global optimization methods; and (3) use of different objective functions (equal weights, w_e, to fit the hydrograph as a whole, versus different weights, w_p, to fit precisely peak values). Model performances were assessed using four efficiency criteria widely used in hydrology (NSE, MNSE, REV, REP). A sensitivity analysis to PEST control data was also performed.

The following specific conclusions were made:

1. Four efficiency criteria (Table 3) were used to evaluate calibration performance on the basis of the two different rainfall-runoff model formulations (FLoD and SLoD models). The Nash-Sutcliffe (NSE) and modified Nash-Sutcliffe (MNSE) efficiency criteria were noted as being less sensitive than relative volume error (REV) and relative peak error (REP) efficiency criteria in performing PEST control data sensitivity analysis and in comparing model calibration performance.
2. Theoretically, global search methods (GSM) provide more accurate parameter estimates than local

search methods (LMS). However, this study shows this is not always true and that the accuracy depends on the rainfall-runoff model parametrization and spatial discretization adopted. Specifically, by adopting a rigorous spatial discretization of the sub-urban catchment (FLoD model), it was not possible to assess which of the two optimization techniques work better: each had its pros and cons. Table 6 shows that by using local search methods (LSM) with a peak weighting approach (w_p), it is possible to obtain a satisfactory overall rainfall-runoff model behavior in terms of peak flow and runoff volume predictions. At the same time, on the basis of the objective function weighting approach adopted, global search methods (GSM) can provide better performance in terms of peak flow or total runoff volume prediction. For example, global search methods (GSM) with a peak weighting approach (w_p) can provide very good performance in matching observed and simulated peak flow (low REP values) even while providing lower performance in terms of observed and simulated runoff volume comparison (high REV values). Likewise, an equal weighting approach (w_e) provides good performance in matching observed and simulated runoff volume (low REV values) but provides lower performance in terms of observed and simulated peak flow comparison (high REP values). By adopting a rainfall-runoff model with a simplistic spatial discretization (SLoD model), global search methods (GSM) manifest higher efficiency in model calibration than local methods (LSM) as is shown in Table 9 and Table 10. Global optimization models provide better performance when the catchment to be modeled is complex and heterogeneous.

3. Objective functions with different weights were used: equal weighting (w_e) and peak weighting approaches (w_p). FLoD models calibrated with local search methods (LSM) showed that even by using a peak weighting approach (w_p), it was not possible to match observed with simulated peak flow (high REP values), while it was possible using global search methods (low REP values). In relation to runoff volume, both local and global search methods with an equal weighting approach (w_e) allowed an accurate simulation of runoff volumes (low REV values).
4. The global optimization process using PEST begins with a random selection of points in the parameter space and then carrying out model runs on the basis of these points. Normally a number of points between 4 and 10 times the number of adjustable parameters is warranted in this random selection process. In order to assess the influence of the so-called points for pre-inversion random sampling, a number of points equal to 2, 4, and 6 times the number of adjustable parameters was used. By comparing Table 8 with Table 7, it is possible to say a number of points equal to 6 times the number of the adjustable parameter provided better results (lower values) in terms of REV values. It is important to emphasize that increasing the number of points for pre-inversion random sampling can improve calibration performance but will increment considerably the calibration computation time.
5. Clearly, it is not possible to establish a priori what the most appropriate set of PEST control data values is before starting any calibration procedure. PEST control data sensitivity analysis shows how variations in these parameters can influence calibration and then model performance. However, by varying PEST control data (in accordance with the values/alternatives reported in Table 4), a significant and substantial variation of the efficiency criteria values was generally not noticed except in a few specific cases. Nevertheless, considering the advantages automatic calibration offers in terms of time savings, going to the effort to perform PEST parameter sensitivity analysis is recommended in order to best exploit its potential in rainfall-runoff model calibration.

Appendix: PEST control parameter

PHIRATSUF: fractional objective function sufficient for end of current iterations

PHIREDLAM: termination criterion for Marquardt lambda search

FACORIG: minimum fraction of original parameter value in evaluating relative change

PHIREDSWH: sets objective function change for introduction of central derivatives

PHIREDSTP: relative objective function reduction triggering termination

NPHISTP: number of successive iterations over which PHIREDSTP applies

NPHINORED: number of iterations since last drop in objective function to trigger termination

RELPARSTP: maximum relative parameter change triggering termination

NRELPAR: number of successive iteration over which RELPARSTP applies

ICOV: instruct PEST to record covariance matrix in matrix file

ICOR: instruct PEST to record correlation-coefficient matrix in matrix file

IEIG: instruct PEST to record eigenvectors in matrix file

DERINC: absolute or relative parameter increment

DERINCLB: absolute lower bound of relative parameter increment

DERINCMUL: derivative increment multiplier when undertaking central derivatives calculation

RLAMBDA1: initial Marquardt lambda

RLAMFAC: factor by which the Marquardt lambda is adjusted

NUMLAM: upper limit on the number of lambdas that PEST can test during any one optimization iteration

RELPARMAX: maximum relative change that a parameter is allowed to undergo between optimization iterations

FACPARMAX: maximum factor change that a parameter is allowed to undergo

NOPTMAX: maximum number of iterations that PEST is permitted to undertake on a particular parameter estimation run

INCTYP: method by which parameter increments are calculated

FORCEN: determines whether central derivatives calculation is done

DERMTHD: method of central derivatives calculation

PARCHGLIM: type of parameter change limit

PARTRANS: parameter transformation

References

Alamdari, N., & Sample, D. J. (2019). A multiobjective simulation-optimization tool for assisting in urban watershed restoration planning. *Journal of Cleaner Production, 213*, 251–261.

Bahrami, M., Bozorg-Haddad, O., & Loáiciga, H. A. (2019). Optimizing stormwater low-impact development strategies in an urban watershed considering sensitivity and uncertainty. *Environmental Monitoring and Assessment, 191*, 340. https://doi.org/10.1007/s10661-019-7488-y.

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., & Montanari, A. (2012). Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Physics and Chemistry of the Earth, 42*(44), 70–76.

Blasone, R. S., Madsen, H., & Rosbjerg, D. (2007). Parameter estimation in distributed hydrological modelling: Comparison of global and local optimisation techniques. *Nordic Hydrology, 38*(4–5), 451–476.

Chau, K. W. (2004). Selection and calibration of numerical modeling in flow and water quality. *Environmental Modeling and Assesment, 9*, 169–178.

da Silva, M. G., de Oliveira de Aguiar Netto, A., de Jesus Neves, R. J., do Vasco, A. N., Almeida, C., & Faccioli, G. G. (2015). Sensitivity analysis and calibration of hydrological modeling of the watershed Northeast Brazil. *Journal of Environmental Protection, 6*, 837–850.

Dent, S., Blair Hanna, R., & Wright, L. (2004). Automated calibration using optimization techniques with SWMM RUNOFF. *Journal of Water Management Modeling*. https://doi.org/10.14796/JWMM.R220-18.

Doherty, J., Brebber, L., & Whyte, P. (2010). *PEST: Model independent parameter estimation*. Brisbane: Watermark Computing Trademarks.

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology, 158*, 265–284.

Duong, V. N., Binh, N. Q., Cuong, L. X., Ma, Q., & Gourbesville, P. (2016). Applying deterministic distributed hydrological model for stream flow data reproduction. A case study of Cu De catchment, Vietnam. *Procedia Engineering, 154*, 1010–1017.

Gupta, H. V., Sorooshian, S., Hogue, T. S., & Boyle, D. P. (2003). Advances in automatic calibration of watershed models. In Q. Duan, H. V. Gupta, S. Sorooshian, A. Rousseau, & R. Turcotte (Eds.), *Calibration of watershed models*. Washington, D.C.: American Geophysical Union.

Hill, M. C. (1998). *Methods and guidelines for effective model calibration*. Denver, Colorado: U.S. Geological Survey.

Kean, S., Chye, L., Shuy, E., Yat-Man, E., & Wan, L. (2008). Performances of rainfall-runoff models calibrated over single and continuous storm flow events. *Journal of Hydrologic Engineering, 13*(7), 597–607.

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advanced in Geosciences (ADGEO), 5*, 89–97.

Legates, D. R., & McCabe Jr., G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35*, 233–241. https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/1998WR900018

Liu, Y. B., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research, 43*. https://doi.org/10.1029/2006WR005756.

Liu, Y. B., Batelaan, O., De Smedt, F., Poórová, J., & Velciká, L. (2005). Automated calibration applied to a GIS-based flood simulation model using PEST. In *Proceedings of the 3rd international symposium on flood Defence*. Nijmegen: CRC Press.

Madsen, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology, 235*(3–4), 276–288.

Madsen, H., & Jacobsen, T. (2001). *Automatic calibration of the MIKE SHE integrated hydrological modelling system*, *Proceedings of 4th DHI software conference*. Helsinger.

Mancipe, N. A., Buchberger, S., & Suidan, M. (2012). Calibration of distributed rainfall-runoff model in Hamilton County, Ohio. *Journal of Water Management Modeling*. https://doi.org/10.14796/JWMM.R245-11.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models, Part I - A discussion of principles. *Journal of Hydrology, 10*, 282–290.

Ngamalieu-Nengoue, U. A., Iglesias-Rey, P. L., Martinez-Solano, F. J., Mora-Melia, D., & Valderrama, J. G. S. (2019). Urban drainage network rehabilitation considering storm tank installation and pipe substitution. *Water, 11*, 515.

Pauwels, V. R. N. (2008). A multistart weight-adaptive recursive parameter estimation method. *Water Resources Research, 44*. https://doi.org/10.1029/2007WR005866.

Poeter, E. P., & Hill, M. C. (1997). Inverse models: A necessary next step in ground-water modeling. *Ground Water, 35*(2), 250–260.

Rossman, L. A. (2017). *Storm Water Management Model reference manual volume II - hydraulics*. Cincinnati: National Risk Management Laboratory, Office of Research and Development, US Environmental Protection Agency.

Rossman, L. A., & Huber, W. C. (2016). *Storm Water Management Model reference manual volume I - hydrology (revised)*. Cincinnati: National Risk Management Laboratory, Office of Research and Development, US Environmental Protection Agency.

Skahill, B. E., & Doherty, J. (2006). Efficient accommodation of local minima in watershed model calibration. *Journal of Hydrology, 329*(1–2), 122–139.

Xu, Z., Xiong, L., Li, H., Xu, J., Cai, X., Chen, K., & Wu, J. (2019). Runoff simulation of two typical urban green land types with the Stormwater management model (SWMM): Sensitivity analysis and calibration of runoff parameters. *Environmental Monitoring and Assessment, 191*, 343. https://doi.org/10.1007/s10661-019-7445-9.

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology, 204*, 83–97.

Zhou, Q., Lai, Z., & Blohm, A. (2019). Optimising the combination strategies for pipe and infiltration-based low impact development measures using a multiobjective evolution approach. *Journal of Flood Risk Management, 12*. https://doi.org/10.1111/jfr3.12457.