



Delft University of Technology

An Insight to the Cornucopia of Possibilities in Calibration Data Collection

Vonach, Tanja; Kleidorfer, Manfred; Rauch, Wolfgang; Tscheikner-Gratl, Franz

DOI

[10.1007/s11269-018-2163-6](https://doi.org/10.1007/s11269-018-2163-6)

Publication date

2019

Document Version

Final published version

Published in

Water Resources Management

Citation (APA)

Vonach, T., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2019). An Insight to the Cornucopia of Possibilities in Calibration Data Collection. *Water Resources Management*, 33(5), 1629-1645. <https://doi.org/10.1007/s11269-018-2163-6>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



An Insight to the Cornucopia of Possibilities in Calibration Data Collection

Tanja Vonach¹ · Manfred Kleidorfer¹ · Wolfgang Rauch¹ · Franz Tschekner-Gratl²

Received: 1 December 2017 / Accepted: 4 December 2018 /

Published online: 18 December 2018

© The Author(s) 2018

Abstract

The calibration of models for urban drainage systems has become more and more important as especially the usage of detailed models has increased considerably over the last years as the basis for planning and design. Still the effects originating from the choice of data used for model calibration are little known and advice on planning measurement campaigns for model calibration is limited, especially for small and medium-sized municipalities. The choice of measurement sites (number and location) within a sewer system is affecting the robustness of the calibration and in consequence the assessment of the modelled system behaviour. This paper discusses the calibration of a hydrologic-hydrodynamic model using the representative example of a small municipality. Different calibration scenarios were created using a model-based approach, focusing on varying availability of in-sewer measurement data. To assess the performance of different scenarios and validate the respective models, different model outputs were compared. The different calibration scenarios resulted in high variations in the model performances. The number and location of used calibration points influence model performance significantly. Predicted CSO volumes deviate from a set of given reference values in ranges between 1% and 253% for one, −21% to −5% for two and 1% to 237% for five used calibration points, depending on the rainfall data input. Consequently, the design of measurement campaigns for calibration data is a very sensitive decision in the modelling process. The model performance further influences design and decision-making processes, which are then perceptible in economic and functional aspects.

Keywords Calibration · Urban drainage · Uncertainties · Hydrodynamic models · Measurement campaigns

✉ Tanja Vonach
tanja.vonach@uibk.ac.at

¹ University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria

² Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands

1 Introduction

The usage of hydrodynamic models, not only for flood forecasting but also as a planning tool in urban drainage has increased considerably over the last decades and with it the importance of understanding a model's ability to reproduce the system behaviour. To ensure that the model performance is sufficient to be a reliable foundation for any planning procedure, the calibration process is a crucial and fundamental component of the model development process (Muschalla et al. 2009; Tscheikner-Gratl et al. 2017). Consequently, the process of model calibration has been the topic of many research activities and publications. For example, Di Pierro et al. (2005) investigated the development of calibration algorithms, Kleidorfer et al. (2009a) highlighted the impact of data accuracy and Deletic et al. (2009) focussed on the sources and propagation of uncertainties.

However, uncalibrated or insufficiently calibrated models are still in use in engineering practice, with data availability often being the limiting factor. Calibration usually requires measurement campaigns, which in turn can increase the economic cost of the simulation projects up to an unachievable level, especially for smaller operators (Freni et al. 2009). Calibration uncertainties relate to the data used for calibration and their selection (Notaro et al. 2013) and to the calibration methods (Leonhardt 2015). They stem of measurement errors for both input and calibration data, the selection of appropriate calibration and validation datasets, the applied calibration algorithms and the objective functions used during the calibration process (Deletic et al. 2012).

Another possible deficit of urban water management studies is that the case studies in scientific literature are often the same, usually larger cities which have the financial and human resources to participate in research projects and to provide an appropriate data-background (i.e. Los Angeles in Barco et al. (2008), Melbourne in Bach et al. (2013) or Shenzhen in Gong et al. (2017)). They are selected for providing this good data background, e.g. measurement data over longer periods of time, and/or the required infrastructure for further data collection and management. Such case studies are not always representative for the entire situation of the living environment in a country. At the very least, there is the risk that research outcomes are biased towards large and more affluent municipalities (Tscheikner-Gratl et al. 2016a).

It is apparent that there are a manifold of factors influencing data availability. In this paper, the influence of data availability on calibration performance is investigated for the hydrodynamic drainage model of a small Austrian municipality. For this purpose, different scenarios of data availability for calibration are simulated. Scenarios for varying input data (different number of calibration events, different rainfall input) have already been considered in Tscheikner-Gratl et al. (2016a). There, the model was calibrated with different rainfall events and data sampled according to empirically based measurement campaigns. Additional scenarios also considered uncertainties in the measurement data, by assuming systematic errors in the collection of water level monitoring data. While the influences of the usage of different model input data (i.e. rainfall recordings) for calibration were evaluated in a very detailed way, effects of the varying spatial distribution of the calibration data (in-sewer measurements) were only marginally discussed. All scenarios used one single measurement data set, which is a water level measurement at one point of the system. Consequently, this work cannot answer the question of the effects of using only one measurement site for calibration in contrast to spatially distributed measurements. Although the small size of the case study and the limitation of funds mimicked nicely an engineering approach, a distributed measurement campaign may lead to a more differentiated outcome and consequently a better representation of the case study.

Existing studies, e.g. Kleidorfer et al. (2009b), investigated the influence of an increasing number of measurement stations for the calibration of conceptual sewer models. The effects of locating calibration points for hydrodynamic models were investigated by Vonach et al. (2018), resulting in a proposed heuristics for measurement site placement concluding with the intention of further research to improve the used methodology. With this work, we want to enhance the understanding of these effects by investigating the influence of a different number and different combinations of calibration points for hydrodynamic drainage models. For this, simulation results from a reference scenario are taken as synthetic measurement data, due to limited data availability. This paper describes three scenarios (scenario I, II and III – varying in amount and distribution of measurement sites) to highlight the influences originating from the design of in-sewer measurement campaigns.

2 Methods

The basis for the used methodology is an existing hydrodynamic model of the case study's urban drainage network (Kleidorfer et al. 2014; Muschalla et al. 2015; Tscheikner-Gratl et al. 2016a; Tscheikner-Gratl et al. 2016b; Vonach et al. 2018). The performance of different scenarios was assessed using the Storm Water Management Model (SWMM 5.1.012) software tool (Burger et al. 2014; Gironás et al. 2010), which is widely used (e.g. by Yazdi (2017) and Gong et al. (2017)).

2.1 Case Study

The analysed case study Telfs is a small municipality with 15,000 inhabitants in Tyrol, Austria at an altitude of about 630 m above sea level. It has an average annual rainfall of about 1000 mm.

The here modelled urban drainage network of Telfs consists of 52 km of combined sewers, 28 km of wastewater sewers and 12 km of stormwater sewers. These stormwater sewers have nine outfalls (in the following figures referred to as *RW* for *rainwater*) into the receiving water bodies, while in comparison only three combined sewer overflows (CSO) exist. In total, a catchment area of approx. 73 ha (1251 subcatchments) is connected to the sewer system. For model calibration and validation, precipitation was measured over a period of 1 year with a temporal resolution of 5 min at three sites (rain gauges RG 1-3) within the catchment area and the water level at one site near the inflow to the wastewater treatment plant. This measurement setup also represents limited data availability, inherent to smaller operators due to limited budget.

The ten rain events with the highest occurring intensities, which surpassed an event threshold of 3 mm for all three rain gauges with an inter-event time of 24 h (listed in Table 1), are consolidated to one continuous rain series. Interconnection of the events is avoided by adding dry-weather periods of 4 h (DWA-A 118 2006) between the individual events (see Fig. 1).

A wastewater treatment plant (WWTP) is located southeast of the town. This plant additionally treats the wastewater of four nearby communities and its capacity is designed for 40,000 population equivalents. Accordingly, the case study's drainage network must cope with conveying also the wastewater of the other association members (the four nearby communities) to the WWTP.

Table 1 Characteristics of the rain events (avg. peak: [mm/5 min] (averaged maximum of the three gauges), max. peak: [mm/5 min] (maximum occurring at one of the gauges), max. at (rain gauge where the maximum peak occurs), avg. sum: [mm])

	RE01	RE02	RE03	RE04	RE05	RE06	RE07	RE08	RE09	RE10
Avg. peak	1.7	3.6	6.4	1.9	1.7	1.6	1.5	1.1	0.3	1.6
Max. peak	2.5	5.4	9.2	3.1	2.8	3.1	2.0	2.6	2.7	2.0
Max. at	RG3	RG3	RG3	RG3	RG3	RG2	RG3	RG3	RG3	RG2
Avg. sum	57.0	152.8	141.1	16.2	30.7	141.2	93.3	22.8	53.8	129.4

Regarding the model, it is important to mention that there are two different and independent outlets connected to the WWTP.

2.2 Calibration Procedure

For the initial model calibration, a measurement campaign for water level data at one location and precipitation data at three locations (shown in Fig. 2) was executed in 2014. Due to this restriction in data availability, the calibration datasets used in this paper are derived from the model, which was calibrated to the measured data and is subsequently referred to as ‘reference scenario’. This reference scenario has been calibrated by Tschekner-Gratl et al. (2016a) with the rain series shown in Fig. 1, including the consideration of all three rain gauges.

Calibration scenarios I, II and III representing a different number and location of measurement sites were established by considering weak points in the uncalibrated model (pipes where the agreement of simulated water level courses between the uncalibrated model and the model of the reference scenario is low) and considering the operator’s empirical knowledge. Basic considerations regarding this heuristic scenario development approach can be found in Vonach et al. (2018).

Figure 3 exemplifies abstractions, which were made during the procedure. The basis is the real system with the real water level measurement. A first abstraction is made by calibrating a model to this measurement. For this, input data is varied in different calibration scenarios

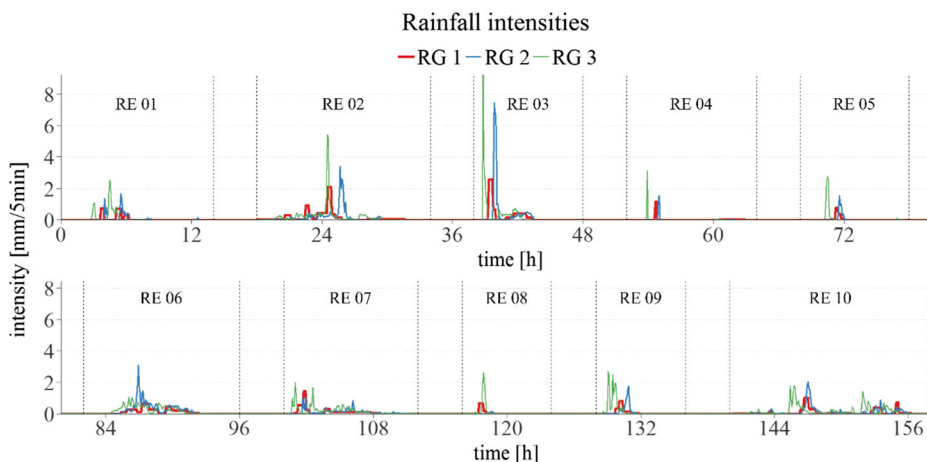


Fig. 1 Consolidated rain series with 10 measured rain events (RE01-RE10) (RG 1: weighing bucket (precipitation) gauge, RG 2 and RG 3: tipping bucket gauges)

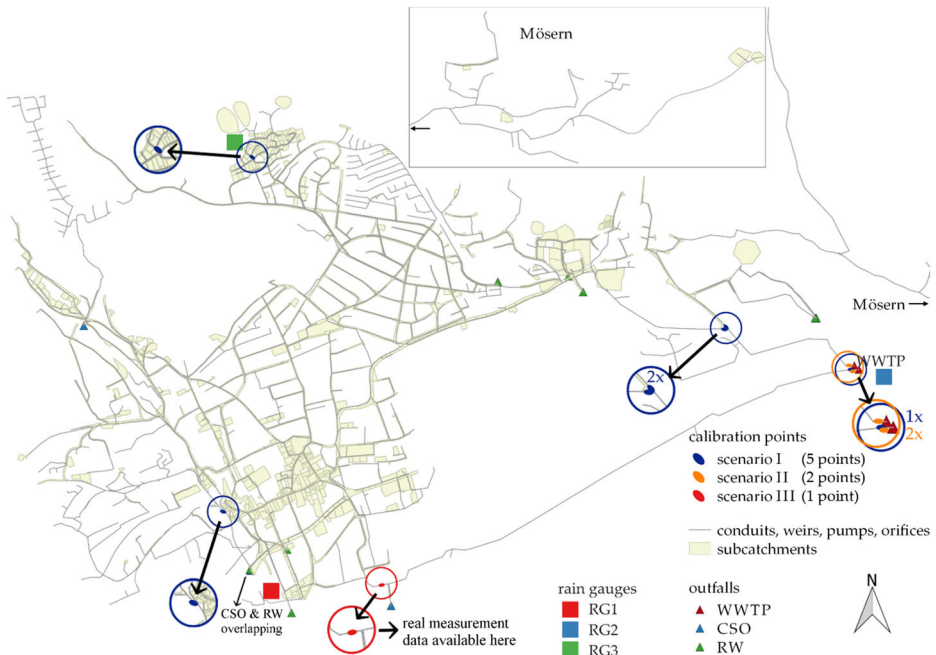


Fig. 2 Model of the case study Telfs with synthetic measurement points for scenario I, II and III (measurement points encircled with arrows leading to enlarged details)

(Tscheikner-Gratl et al. 2016a). The calibration scenario, which resulted in the best overall agreement between the measurement and the model, i.e. the scenario with a calibration to the entire rain series of 10 events and of all three rain gauges with agreement expressed by the Nash-Sutcliffe Efficiency NSE (McCuen et al. 2006; Nash and Sutcliffe 1970), is then used as the reference scenario.

In contrast to Tscheikner-Gratl et al. (2016a), all scenarios of the present work are calibrated with the three rain events RE03, RE06 and RE09 of all three gauges (see Fig. 1). These rain events turned out to be representative, as models calibrated with only one of those delivered a high model performance compared to the reference scenario in terms of their ability to predict the CSO and flooding volumes (Tscheikner-Gratl et al. 2016a). To obtain appropriate synthetic measurement data, i.e. model outputs of the reference scenario, the reference model is simulated with a consolidated rain series of these three rain events.

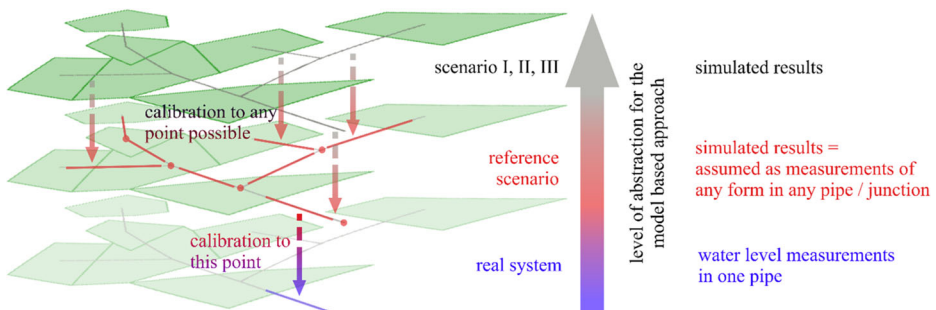


Fig. 3 Necessary abstractions for the model-based approach

By this means, synthetic measurement data and system performance is available in any form (e.g. water levels, CSO and flooding volumes, etc.) and at every point in the system. The model is then calibrated in three separate scenarios (I, II and III), each using different data from different calibration points (shown on Fig. 2) which is extracted from the reference scenario. A similar approach but for a different objective is also used and described in Kleidorfer et al. (2009b). In this work, a temporal resolution of the synthetic measurement data of 5 min is used.

The locations of the measurement sites for the simulated calibration data for the three scenarios I to III are shown in Fig. 2. Scenario I consists of five calibration points scattered throughout the network. Scenario II has two calibration points right at the outlets to the wastewater treatment plant. The (one) investigated calibration point for scenario III is the same point where also the real data was collected. Nevertheless, the synthetic water levels from the reference scenario were used for scenario III to keep the scenarios comparable on the same level of abstraction.

Only subcatchment related parameters concerning the runoff concentration and the total runoff volume are varied. The runoff model implemented in SWMM uses two main parameters, the subcatchment width and the subcatchment imperviousness. The width of a subcatchment determines the shape of the area and consequently significantly affects the concentration time while the imperviousness mainly relates to the total runoff volume. There are more parameters determining the timing (i.e. catchment slope or pipe roughness), but subcatchment width is a non-measurable value and is thus only determinable by calibration, and choosing more parameters to calibrate would reduce parameter identifiability. To avoid a further deficiency in the identifiability of the parameters, subcatchments are clustered, according to their initially estimated values, into three groups distinguishing the imperviousness and four groups differentiated by their width (the ratio between width and area, i.e. their shape), see also Vonach et al. (2018). Each cluster has its own factor being multiplied with the initial width or imperviousness. Consequently, these seven factors represent the calibration parameters (instead of $1251 \times 2 = 2502$ possible unique values).

Parameter adaptation is implemented and automated with R (R Development Core Team 2008). We used an optimization algorithm based on a Nelder-Mead simplex (Duan et al. 1994; Nelder and Mead 1965). The Nash-Sutcliffe efficiency (NSE) was chosen as the objective function to compare measured and predicted water levels. NSE is a measure to compare time series. It ranges from $-\infty$ (no agreement) to 1 (perfect match). For the calibrations in this paper, a threshold of $\text{NSE} = 0.9$ is chosen (Shamseldin 1997). When the optimization algorithm finds a parameter set, which exceeds this threshold at the calibration point, the algorithm terminates and the model is considered as calibrated for the currently regarded calibration point.

As there are several points to calibrate the model to, calibration to each synthetic measurement station is performed in a downstream order. By this means, only subcatchments lying upstream of the current and downstream of the previous calibration point are modified. As an example, the subcatchments varied for each calibration step of scenario I are shown in Fig. 4. This methodology highly increases the number of calibration parameters compared with a calibration to all points simultaneously, as every step establishes a new set of seven parameters.

Each point used for calibration (in total 7 different points of the network, because scenario I and II use the same point at the wastewater treatment plant) is additionally tested as a single calibration point to investigate the effect of a preceding stepwise calibration procedure.

A calibration to systems outlets can lead to a loss of information about high-resolution system behaviour. To highlight the possible extent of this loss, a sensitivity analysis is

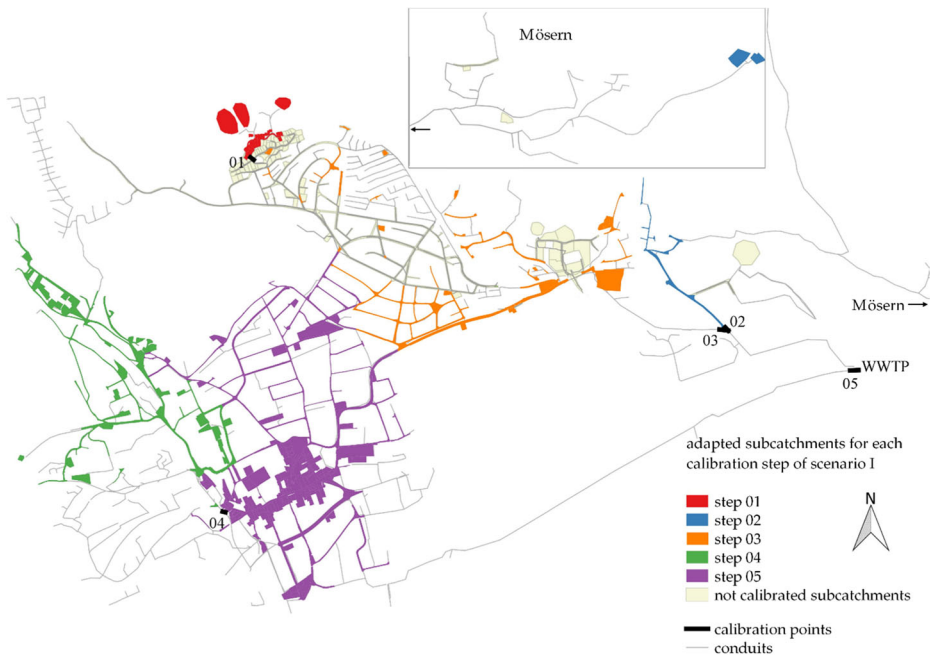


Fig. 4 Systematic order of adapted subcatchments for calibration scenario I

performed for scenario II (a calibration to measurements at the outlets to the WWTP) as an addition to the calibration scenarios. In this sensitivity analysis, 300 sets of calibration parameters were sampled randomly within defined ranges. The resulting models were then simulated three times with three different rainfall inputs. Two of the rainfall inputs are design storms Euler II with a return period of 5 and 10 years, respectively, prepared according to Austrian design guidelines (ÖWAV RB11 2009). More information about the application of these design storm events can be found e.g. in Mikovits et al. (2017) or De Toffol et al. (2006). They are characterized by a high peak (here 12.9 and 15.7 mm/5 min) and a short duration (here 120 min). Required statistical data for these rain events are available from the Austrian rainfall database (eHYD) (Weilguni 2009). The third simulation is done with the measured rain event RE03 from all three rain gauges, where the highest intensities of the measurement series occur (9.2 mm/5 min, measured at RG3).

2.3 Model Validation and Evaluation

Influences caused by the choice of calibration points are evaluated with scenarios I, II and III. The models of the intermediate calibration steps are simulated with six different rainfall inputs (resulting in $(1 + 2 + 5) \times 6 = 48$ simulations with the intermediate scenario models and six simulations with the reference scenario model). These rainfall inputs include our own measurements (the entire measured data of three rain gauges, not the consolidated rain series from Fig. 1), each gauge is used for an individual simulation, a design storm event of type Euler II with a return period (rp) of 10 years (y) and the precipitation data sets ZAMG1 and ZAMG2. ZAMG1 and ZAMG2 are data sets of the Austrian Central Institute for Meteorology and Geodynamics (ZAMG) and are chosen from the nearest measurement sites available (ZAMG1

10 km from the catchment and ZAMG2 30 km). A spatial distribution of the occurring rainfall is not considered for validation. All rain series used for validation last 200 days, except the design storm events, which have a duration of 120 min. Our own measurements and the design storms have a temporal resolution of 5 min, whereas ZAMG1 and ZAMG2 have a time step of 10 min.

Subsequently, model outputs are compared to the reference scenario regarding the occurring flooding and CSO volumes.

To show the effects of generalizing the system behaviour, accompanied by information loss due to a smaller number of measurement stations, simulated CSO and flooding volumes (taken from the randomly created models of the sensitivity analyses for scenario II) were compared with the results from the reference scenario. Only those models were considered which had a good agreement with the reference scenario ($NSE > 0.8$) for the water level courses in the scenario II calibration points (the pipes just before both outlets to the WWTP). So, the occurring effect of information loss about the upstream system's behaviour while maintaining a good agreement at the calibration points can be highlighted. For these evaluations now, the threshold is lowered compared to the calibration scenarios. The authors are aware of the fact, that a NSE of 0.9 is an ambitious value for calibration, as this value indicates a "very satisfactory model performance" according to Shamseldin (1997). By lowering the threshold to a still appropriate value of 0.8 (indicating a "fairly good model" (Shamseldin 1997)), the benefits of having more evaluable data points might outweigh the disadvantage of a slightly lower agreement for the significance of results.

3 Results and Discussion

The following elaboration of results is divided into three parts. First, the final validation performance of the calibration scenarios is compared. Secondly, the performance of the intermediate calibration steps are evaluated to provide an insight to the impact of a varying complexity of a calibration process on the model accuracy. Thirdly, results of the sensitivity analysis performed on scenario II are given.

3.1 Performance of Calibrated Models

Tables 2, 3 and 4 give an overview of the simulated CSO and flooding volumes of the calibration scenarios when simulating the resulting models with the five measured rainfall series (RG1, RG2, RG3, ZAMG1, ZAMG2) and one design storm (Euler II) used for validation.

Out of the measured rainfall series, ZAMG1 caused the largest CSO and flooding volumes in the reference scenario.

Rain series RG3 and RG2 result in very low flooding volumes (4 and 36 m³) for the reference scenario and the relative deviations of the calibrated models (Table 4) are therefore even higher but accordingly less significant. Rain series ZAMG2 and RG1 did not elicit any flooding in the reference scenario, therefore percentage deviations are not possible to evaluate.

Figure 5 shows the results for models simulated with the measured rainfall of ZAMG1. It shows the deviations of predicted flooding and CSO volumes from the reference scenario not only for the scenarios evaluated in this paper but also in comparison with the calibration scenarios with different rainfall input or systematic errors in the water level measurements

Table 2 CSO and flooding volumes for different calibration scenarios (Sc. I, II, III and reference scenario) and rainfall inputs (depicted as: CSO / flooding in [m³])

Sc.	RG1	RG2	RG3	ZAMG1	ZAMG2	Euler II
I	796 / 0	4223 / 183	4182 / 29	9911 / 1698	1049 / 0	10,322 / 5340
II	182 / 0	2604 / 178	2226 / 21	7870 / 2074	351 / 0	9602 / 5279
III	804 / 0	4203 / 265	4142 / 29	9896 / 2172	985 / 0	10,311 / 5977
Ref.	228 / 0	2727 / 36	2417 / 4	8448 / 1075	444 / 0	10,192 / 4143

(Tscheikner-Gratl et al. 2016a). By this, the effects of different calibration points on model performance can be compared directly to the effect of using different rainfall inputs for calibration or a systematic measurement error.

Only one of all models (the scenario that was calibrated with rain event RE05 and including all three rain gauges) deviates less than 25% for both volumes. Scenarios I, II and III all show less than 25% deviation from the reference scenario for the CSO volume. Flooding volume is overestimated throughout for these three scenarios with up to 102%. Scenario III, which has the same amount of calibration points (one) as the other calibration scenarios from Tscheikner-Gratl et al. (2016a), shows a better performance for CSO volumes than most of those scenarios. In return, the agreement of flooding volume is inferior.

No scenario exceeds a deviation of 100% for CSO volume, but three (including scenario III) exceed a deviation of 100% for flooding volume.

The good agreements of CSO volumes for scenarios I, II and III with the reference scenario could elicit from an advantageous sampling of model input data. A spatially distributed rainfall as well as different rain events were used for these calibrations. This contrasts with the majority of the other scenarios, where mostly either only one rain gauge or only one rain event is used for calibration.

Looking at the resulting models themselves, scenario II agrees best in terms of the connected impervious area. The reference scenario has a mean imperviousness of 45.6%. Scenario II approximates this mean imperviousness with 49.3%. Nevertheless, according to the results in Fig. 5, it still underestimates CSO volumes by 7% while overestimating flooding volumes by 93%. Scenarios I and III both overestimate the mean imperviousness (I_m) with $I_m = 54.2\%$ (scenario I) and $I_m = 57.5\%$ (scenario III), flooding volumes (I: 58% and III: 102% deviation from the reference scenario) and CSO volumes (I: 17% and III: 17% deviation from the reference scenario).

3.2 Uncertainties Due to a Spatial Difference in Calibration Data Availability

Concerning the necessary amount of measurement stations to gather data for calibration, Fig. 6 shows the change in the model behaviour (CSO and flooding volumes compared to the

Table 3 Percentage deviations of CSO volumes from the reference scenario for different calibration scenarios (Sc. I, II, III) and rainfall inputs

Sc.	RG1	RG2	RG3	ZAMG1	ZAMG2	Euler II
I	+237.3%	+54.9%	+73.0%	+17.3%	+136.3%	+1.3%
II	-20.2%	-4.5%	-7.9%	-6.8%	-20.9%	-5.8%
III	+252.6%	+54.1%	+71.4%	+17.1%	+121.8%	+1.2%

Table 4 Percentage deviations of flooding volumes from the reference scenario for different calibration scenarios (Sc. I, II, III) and rainfall inputs

Sc.	RG1	RG2	RG3	ZAMG1	ZAMG2	Euler II
I	–	+408.3%	+625.0%	+58.0%	–	+28.9%
II	–	+394.4%	+425.0%	+92.9%	–	+27.4%
III	–	+636.1%	+625.0%	+102.0%	–	+44.3%

reference scenario) with a different number of measurement sites. To investigate a larger number of calibration points as well as the final model, the models resulting from the intermediate steps in the calibration procedure are used. For flooding, only results for the two rainfall inputs ZAMG1 and Euler II are shown on Fig. 6, as they result in the most significant flooding volumes in the reference scenario.

For both scenarios I and II, adding a further calibration point improves model performance for flooding volumes throughout. This is also the case for CSO volumes for scenario II. For scenario I, there are some exceptions, where model performance worsens temporarily during the step-wise calibration procedure (RG1: step 3, RG2: step 2 RG3: step 2, ZAMG1: step 2). Nevertheless, in absolute numbers, these interim degradations of performance for CSO volumes are only marginally, e.g. for a simulation with RG1 the CSO volume increases only by 133 m³ from step 2 (1893 m³) to step 3 (2026 m³). For other rainfall inputs, degradations are only occurring with less than 30 m³ increase in CSO volume.

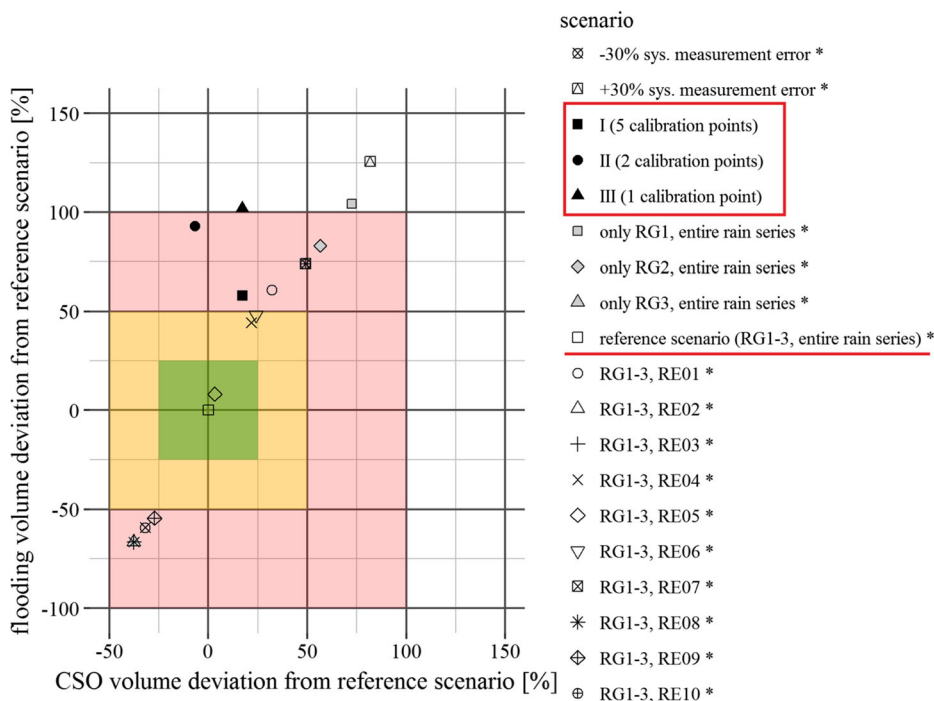


Fig. 5 Flooding volume and CSO volume deviation for measured 200-days rain series ZAMG1 (asterisked scenarios are results simulated with models according to Tscheikner-Grat et al. (2016a); these scenarios are named according to their rainfall input (RainGauge and RainEvent) used for calibration)

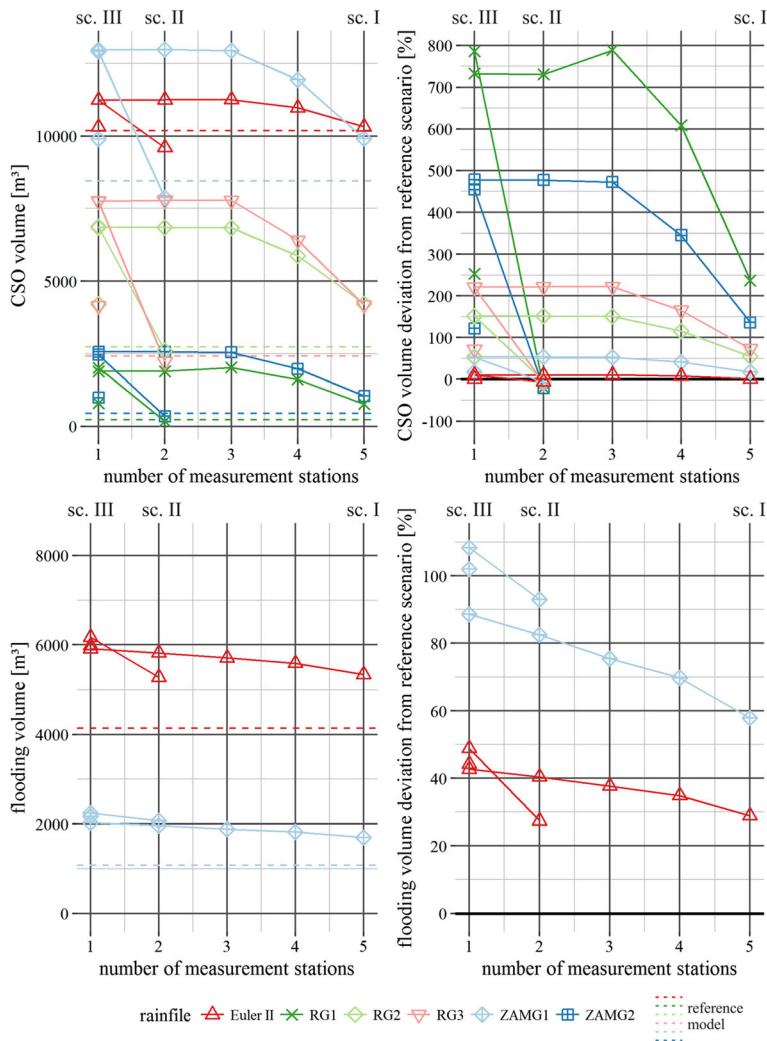


Fig. 6 CSO and flooding volumes and deviations from the reference scenario for six (CSO) and two (flooding) validation rainfall inputs and the intermediate calibration steps of the three scenarios

Regarding the significant improvements of the last points of scenario I and II (which are the same points of the network) and the fact that there is only one CSO but no additional subcatchments between the endpoints of scenario I and II and the calibration point of scenario III, it can be assumed that this branch is favourable for calibration. Thus, Fig. 7 shows the resulting CSO volumes for models calibrated to only one calibration point on this branch. Models calibrated only to the endpoint of scenario I and II show very similar results as calibration scenario III, where the calibration point is approx. 2.5 km more upstream with an interposed CSO construction. In addition, scenario I, where the model is calibrated to 5 different points subsequently, is not superior to a model, which is calibrated only to the last of these 5 points.

Both points used in scenario II would result in a positive deviation when used as a single calibration point. Thus, the switch from an over- to an underestimation of CSO volumes within the calibration procedure of scenario II (Figs. 6 and 7) does not stem from one of these two points but is resulting from the combination of both used calibration points. The inferior model performance from a calibration to the first point of scenario II compared to a calibration to both points or only the second point of scenario II might be related to the high increase of the calibrated connected area. After the subsequent calibration of scenario II, the second measurement point (south eastern outlet) is connected to 18.3 ha of impervious area and the first (north eastern outlet) to only 5.6 ha.

Using the reference scenario model to produce synthetic measurement data also enables us to compare the resulting water level courses in all other pipes. Figure 8a to h exemplify how the model fit changes apart from the considered calibration point(s). For this, the Nash-Sutcliffe efficiency is calculated in selected pipes of the network at the end of a branch with no change in the inflow (includes about one third of all pipes in the system). It shows the stepwise improvement in specific pipes for the intermediate steps of each calibration scenario.

3.3 Sensitivity Analysis for Scenario II

Out of 300 models, 71 models resulted in a $NSE > 0.8$ at both calibration points for an Euler II design storm of $rp = 5y$ and 88 models for $rp = 10y$. The measured rainfall (rain event RE03) input evoked only very little flooding volumes between 0 and 10 m^3 and is neglected for further evaluations. Therefore, Fig. 9 shows results for simulations with the two design storms.

All but one of the random but calibrated models overestimated the flooding volume while underestimating the CSO volume. The deviations in the CSO volume range between -16% and $+9\%$, this is a significantly smaller range than for the flooding volume with a range from -9% to $+109\%$. The higher return period rainfall events are shown in this case to result in lower relative errors than rainfall events of a lower return period, especially in flooding

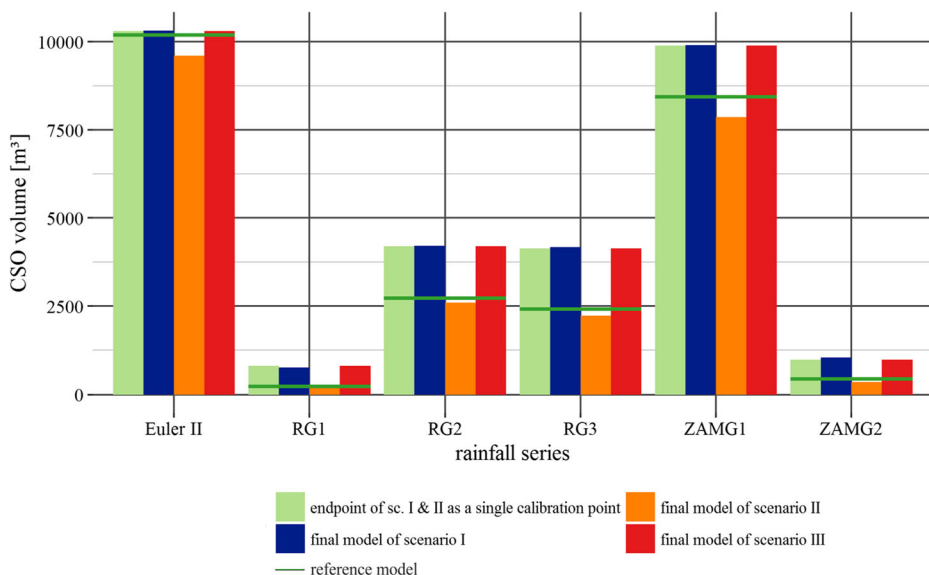


Fig. 7 CSO volumes when using one single calibration point (all on the same sewer branch)

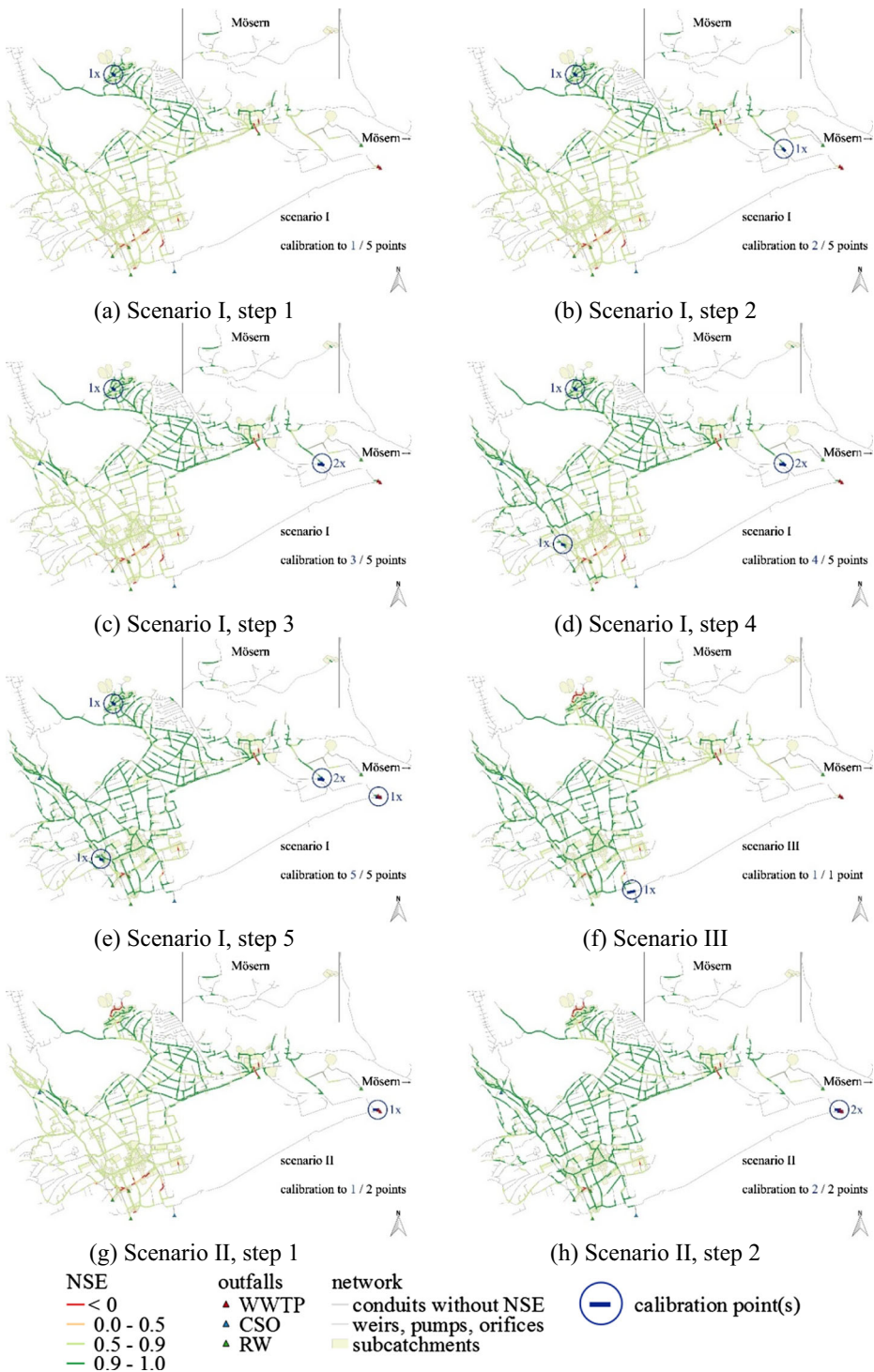


Fig. 8 Overall agreements for different calibration steps in scenarios I, III and II

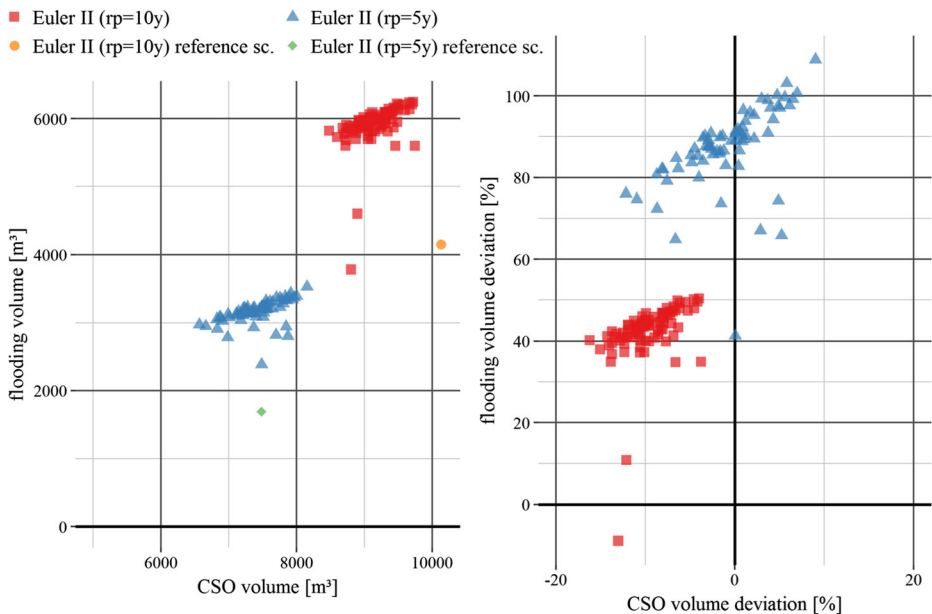


Fig. 9 Absolute values (left) and deviation from the reference scenario (right) of flooding volume and CSO volume for design storm events Euler II (rp=5y and 10y) and random calibration parameter variations for scenario II

volume. These results agree with the findings of Ahmed (2012), who also showed an improved model performance at higher flows in the context of calibrating an integrated river basin model.

The range of CSO volumes is about the same for both rainfall events (1589 m³ for rp=5y and 1262 m³ for rp=10y). Also, if we neglect the flooding volume's outliers, the occurring range for the absolute flooding volume is in the same order of magnitude for both return periods (with 742m³ for rp=5y and 645 m³ for rp=10y).

The impact of coarse-grained sensor distributions (coarse-grained in terms of the flow path within the system, not in a geographic sense) is shown to be relatively minor when regarding the CSO volume but is more significant for the resulting flooding volume. This shows that the flooding volume is more sensitive to changes in the connected area than the CSO volume. The here simulated return period of 5 years is close to the design return period (3 to 5 years depending on land-use) meaning that additional surface runoff causes flooding before it can reach a CSO outlet. The resulting ranges for both volumes are remarkably smaller than their absolute differences to the values from the reference scenario. Consequently, a dense distribution of models with similar deviations can be seen. This could be a result of unsuitable boundary conditions of the calibration parameters. However, the presence of occurring outliers preclude this possibility.

4 Conclusion

A model-based approach was used to evaluate different calibration scenarios with 1, 2 and 5 calibration points. They were validated subsequently with different rainfall sets, including

measured rainfall series from available surrounding measurement stations and design storm events with different return periods. Then they were compared to scenarios processed in Tschekner-Gratl et al. (2016a). Even for different rainfall inputs and neglecting spatial distributions of occurring intensities, their validations resulted in very good agreements and low deviations from the used reference scenario when looking at the CSO volumes. Concerning flooding volumes, the here established scenarios can keep up with those from Tschekner-Gratl et al. (2016a), but are not able to prevail.

The performed sensitivity analysis exemplified a possible error range due to the loss of information. For design storms with return periods of 5 and 10 years, it shows higher possible deviation ranges for flooding than for CSO volumes. Possible ranges of 100% broadness for predicted flooding volumes while having about the same agreement at the calibration points show that there is a quite significant uncertainty inherent to the calibration procedure when using few calibration points as a reference. This result of the study is especially interesting when linking the model calibration with the modelling aims, which are usually either the prediction of CSO volumes (to protect receiving water quality) or to predict flooding volume (to limit flood risk). While prediction of CSO volume is rather stable, high deviations of flooding volume have to be expected even for “calibrated” models. In current modelling practise, this is not a problem as hydrodynamic models are still often used in a way that flooding is just not allowed for certain design return periods. If flooding happens, the pipes have to be re-designed. But with changes in the design approach toward a risk approach in which flooding volumes, flooded areas and water on the surface is accepted to a certain extend for certain design return periods, the instability in predicting flooding volumes might cause problems. Currently the main recommendation only is that distributed measurements improve the understanding of the overall system behaviour. They also improve the calibration performance, but for sure further investigations of calibration of hydrodynamic urban drainage models under high flow conditions are required.

The evaluation of different scenarios for measurement campaigns showed significant differences of model performance with a varying number and location of calibration points. Often, economic or practical reasons restrict the execution of extended measurement campaigns. This study shows that with a careful selection of input data also one well-chosen calibration point can express and predict the system behaviour of a small case study to a satisfiable extent for engineering purposes. A favourable sampling of output data (i.e. a well-planned measurement site selection) can reduce uncertainties regarding an unfavourable choice of input data (i.e. the choice of calibration rainfall events). This emphasizes the importance of conferring resources to the calibration process. These resources are meant in terms of ensuring the availability of different kinds of measurement data (e.g. water levels and CSO volumes) as well as the time used to enable a well-considered planning of data collection.

Acknowledgements A previous shorter version of the paper has been presented in the 10th World Congress of EWRA” Panta Rei” Athens, Greece, 5-9 July 2017.

This work was funded by the Austrian Climate and Energy Fund in the project CONQUAD (9th Call of the Austrian Climate Research Program project number KR16AC0K13143).

Franz Tschekner-Gratl is financed by the Marie Skłodowska Curie Initial Training Network QUICS. The QUICS project has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.

We would also like to acknowledge the input and good cooperation with the Gemeindewerke Telfs, the operating company of the case study.

Funding Information Open access funding provided by University of Innsbruck and Medical University of Innsbruck.

Compliance with Ethical Standards

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Ahmed F (2012) A hydrologic model of Kemptville Basin - calibration and extended validation. *Water Resour Manag* 26:2583–2604. <https://doi.org/10.1007/s11269-012-0034-0>
- Bach PM, Deletic A, Urich C, Sitzenfrie R, Kleidorfer M, Rauch W, McCarthy DT (2013) Modelling interactions between lot-scale decentralised water infrastructure and urban form – a case study on infiltration systems. *Water Resour Manag* 27:4845–4863. <https://doi.org/10.1007/s11269-013-0442-9>
- Barco J, Wong KM, Stenstrom MK (2008) Automatic calibration of the U.S. EPA SWMM model for a large urban catchment. *J Hydraul Eng* 134:466–474. [https://doi.org/10.1061/\(asce\)0733-9429\(2008\)134:4\(466\)](https://doi.org/10.1061/(asce)0733-9429(2008)134:4(466))
- Burger G, Sitzenfrie R, Kleidorfer M, Rauch W (2014) Parallel flow routing in SWMM 5. *Environ Model Softw* 53:27–34. <https://doi.org/10.1016/j.envsoft.2013.11.002>
- De Toffol S, Kleidorfer M, Rauch W (2006) Vergleich hydrodynamischer und hydrologischer Simulationsmodelle bei der Berechnung der Emissionen von Mischwasserbehandlungsanlagen [Comparison of hydrodynamic and hydrologic simulation models for quantifying emissions of combined sewer systems]. *Wiener Mitteilungen* 196:H1–H20
- Deletic A, Dotto CBS, McCarthy DT, Kleidorfer M, Freni G, Mannina G, Uhl M, Fletcher TD, Rauch W, Bertrand-Krajewski JT, Tait S (2009) Defining Uncertainties in Modelling of Urban Drainage Systems. Paper presented at the 8th International Conference on Urban Drainage Modelling and 2nd International Conference on Rainwater Harvesting and Management, 7th - 12th September 2009, Tokyo, Japan
- Deletic A, Dotto CBS, McCarthy DT, Kleidorfer M, Freni G, Mannina G, Uhl M, Henrichs M, Fletcher TD, Rauch W, Bertrand-Krajewski JL, Tait S (2012) Assessing uncertainties in urban drainage models. *Phys Chem Earth A/B/C* 42:44:3–10. <https://doi.org/10.1016/j.pce.2011.04.007>
- Di Piero F, Djordjević S, Kapelan Z, Khu ST, Savic DA, Walters GA (2005) Automatic calibration of urban drainage model using a novel multi-objective genetic algorithm. *Water Sci Technol* 52:43–52
- Duan Q, Sorooshian S, Gupta VK (1994) Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J Hydrol* 158:265–284. [https://doi.org/10.1016/0022-1694\(94\)90057-4](https://doi.org/10.1016/0022-1694(94)90057-4)
- Freni G, Mannina G, Viviani G (2009) Assessment of data availability influence on integrated urban drainage modelling uncertainty. *Environ Model Softw* 24:1171–1181. <https://doi.org/10.1016/j.envsoft.2009.03.007>
- Gironás J, Roesner LA, Rossman LA, Davis J (2010) A new applications manual for the storm water management model (SWMM). *Environ Model Softw* 25:813–814. <https://doi.org/10.1016/j.envsoft.2009.11.009>
- Gong Y, Li X, Zhai D, Yin D, Song R, Li J, Fang X, Yuan D (2017) Influence of rainfall, model parameters and routing methods on stormwater modelling. *Water Resour Manag*. <https://doi.org/10.1007/s11269-017-1836-x>
- Kleidorfer M, Deletic A, Fletcher TD, Rauch W (2009a) Impact of input data uncertainties on urban stormwater model parameters. *Water Sci Technol* 60:1545–1554. <https://doi.org/10.2166/wst.2009.493>
- Kleidorfer M, Moederl M, Fach S, Rauch W (2009b) Optimization of measurement campaigns for calibration of a conceptual sewer model. *Water Sci Technol* 59:1523–1530. <https://doi.org/10.2166/wst.2009.154>

- Kleidorfer M, Tschiesche U, Tscheikner-Gratl F, Sitzenfrei R, Kretschmer F, Muschalla D, Ertl T, Rauch W (2014) Von den Daten zum Modell: Anforderungen an hydraulische Entwässerungsmodelle in kleinen und mittleren Gemeinden [From data to model - requirements for hydraulic urban drainage models applied for small and medium sized case studies]. *Wiener Mitteilungen* 231:J1–J14
- Leonhardt G (2015) Development and application of software sensors and reverse models for urban drainage systems: Model-based approaches for gaining more information from measurement data. *Forum Umwelttechnik und Wasserbau* 19. Innsbruck Univ. Press, Innsbruck, Austria
- McCuen RH, Knight Z, Cutter AG (2006) Evaluation of the Nash Sutcliffe efficiency index. *J Hydrol Eng* 11: 597–602. [https://doi.org/10.1061/\(asce\)1084-0699\(2006\)11:6\(597\)](https://doi.org/10.1061/(asce)1084-0699(2006)11:6(597))
- Mikovits C, Rauch W, Kleidorfer M (2017) Importance of scenario analysis in urban development for urban water infrastructure planning and management. *Comput Environ Urban Syst*. <https://doi.org/10.1016/j.compenvurbsys.2017.09.006>
- Muschalla D et al (2009) The HSG procedure for modelling integrated urban wastewater systems. *Water Sci Technol* 60:2065–2075. <https://doi.org/10.2166/Wst.2009.576>
- Muschalla D et al. (2015) Auf effizientem Wege von den Daten zum Modell (DATMOD) - Sanierungs- und Anpassungsplanung von kleinen und mittleren Kanalnetzen [An efficient way from data to model - renovation and adaptation planning for small and medium size sewer networks (DATMOD)]. Vienna, Austria
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I - a discussion of principles. *J Hydrol* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7:308–313. <https://doi.org/10.1093/comjnl/7.4.308>
- Notaro V, Fontanazza CM, Freni G, Puleo V (2013) Impact of rainfall data resolution in time and space on the urban flooding evaluation. *Water Sci Technol* 68:1984–1993. <https://doi.org/10.2166/wst.2013.435>
- ÖWAV RB11 (2009) Regelblatt 11 - Richtlinien für die abwassertechnische Berechnung und Dimensionierung von Abwasserkanälen [Guidelines for the calculation, dimensioning and design of sewers]. Österreichischer Wasser- und Abfallwirtschaftsverband, Vienna, Austria
- R Development Core Team (2008) R: a language and environment for statistical computing. Vienna, Austria
- Shamseldin A (1997) Application of a neural network technique to rainfall-runoff modelling. *J Hydrol* 199:272–294
- Tscheikner-Gratl F, Zeisl P, Kinzel C, Leimgruber J, Ertl T, Rauch W, Kleidorfer M (2016a) Lost in calibration: why people still don't calibrate their models, and why they still should – a case study from urban drainage modelling. *Water Sci Technol* 74:2337–2348. <https://doi.org/10.2166/wst.2016.395>
- Tscheikner-Gratl F, Zeisl P, Kinzel C, Leimgruber J, Ertl T, Rauch W, Kleidorfer M (2016b) Uncertainties in hydrodynamic modelling regarding rainfall measurements and calibration. In: *Computing and Control for the Water Industry Conference (CCWI 2016)*, Amsterdam, The Netherlands
- Tscheikner-Gratl F, Zeisl P, Kinzel C, Leimgruber J, Ertl T, Rauch W, Kleidorfer M (2017) Effect of varying calibration scenarios on the performance of a hydrodynamic sewer model. *European Water Journal* 57:287–291
- Vonach T, Tscheikner-Gratl F, Rauch W, Kleidorfer M (2018) A heuristic method for measurement site selection in sewer systems. *Water* 10. <https://doi.org/10.3390/w10020122>
- Weilguni V (2009) Bemessungsniederschläge in Österreich [Design storms in Austria]. *Wiener Mitteilungen Wasser-Abwasser-Gewässer* 216:71–84
- Yazdi J (2017) Rehabilitation of urban drainage systems using a resilience-based approach. *Water Resour Manag*. <https://doi.org/10.1007/s11269-017-1835-y>