On the use of multi-objective optimization for multi-site calibration of extensive green roofs

Abdalla Elhadi Mohsen Hassan[1], Alfredsen Knut[1], and Muthanna Tone Merete[1]

[1]Norwegian University of Science and Technology

November 16, 2022

Abstract

Conceptual hydrological models are practical tools for estimating the performance of green roofs, with respect to stormwater management. Such models require calibration to obtain parameter values, which limits their use in cases when measured data are not available. One approach that has been thought to be useful is to transfer parameters from a gauged roof calibrated locally (i.e., single-site calibration) to a similar ungauged roof located in a different location. This study tested this approach by transferring calibrated parameters of a conceptual hydrological model between sixteen extensive green roofs located in four Norwegian cities with different climatic conditions. The approach was compared with a multi-site calibration scheme that explores trade-offs of model performances between the different sites. The results showed that single site calibration could yield optimal parameters for one site and perform poorly in other sites. In contrast, obtaining a common parameter set that yields satisfactory results (Kling Gupta Efficiency >0.5) for different sites, and roof properties could be achieved by multi-site calibration. The practical implications of multi-site calibration have been discussed in the context of stormwater management. The multi-site calibration scheme is recommended not only for transferability amongst roofs in different sites but also when applying conceptual models for evaluating climate change scenarios in which the climatic variables are significantly different from the ones used for calibration.

On the use of multi-objective optimization for multi-site calibration of extensive green roofs

Elhadi Mohsen Hassan Abdalla [1], Knut Alfredsen [1], Tone Merete Muthanna [1]

[1] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Andersens vei 5, Trondelag 7031, Norway.

Corresponding author: **Elhadi Mohsen Hassan Abdalla** (elhadi.m.h.abdalla@ntnu.no)

Key Points:

- This study compared single-site and multi-site calibration of conceptual hydrological models of green roofs
- Single-site calibration yielded optimal parameters for one site that produced poor simulation results in different locations
- Multi-site calibration yielded satisfactory modelling results in different locations and roof properties.

Abstract

Conceptual hydrological models are practical tools for estimating the performance of green roofs, with respect to stormwater management. Such models require calibration to obtain parameter values, which limits their use in cases when measured data are not available. One approach that has been thought to be useful is to transfer parameters from a gauged roof calibrated locally (i.e., single-site calibration) to a similar ungauged roof located in a different location. This study tested this approach by transferring calibrated parameters of a conceptual hydrological model between sixteen extensive green roofs located in four Norwegian cities with different climatic conditions. The approach was compared with a multi-site calibration scheme that explores trade-offs of model performances between the different sites. The results showed that single site calibration could yield optimal parameters for one site and perform poorly in other sites. In contrast, obtaining a common parameter set that yields satisfactory results (Kling Gupta Efficiency >0.5) for different sites, and roof properties could be achieved by multi-site calibration. The practical implications of multi-site calibration have been discussed in the context of stormwater management. The multi-site calibration scheme is recommended not only for transferability amongst roofs in different sites but also when applying conceptual models for evaluating climate change scenarios in which the climatic variables are significantly different from the ones used for calibration.

1 Introduction

In the last few decades, green roofs have emerged as a sustainable stormwater infrastructure option. Green roofs reduce the volume and intensity of stormwater runoff entering the sewer network, through retention and detention processes (Hamouz et al., 2018; Johannessen et al., 2018; Stovin et al., 2013). Furthermore, green roofs reduce the urban heat island effect (Susca et al., 2011); enhance urban biodiversity (Wooster et al., 2022); improve the visual amenity of urban catchments (Jungels et al., 2013); reduce the energy consumption of buildings (Jim, 2014; Refahi & Talkhabi, 2015).

Hydrological models are practical tools for evaluating the efficiency of various design configurations of green roofs under different climatic conditions. Thus, they can assist practitioners aiming at quantifying the hydrological impact, i.e., retention and detention processes, of green roof implementation in urban catchments. Numerous hydrological models of green roofs have been developed and tested in the literature. The models can be classified into physically-based (Bouzouidja et al., 2018; Yanling Li & Babcock, 2015; Palla et al., 2009), conceptual (Abdalla et al., 2022; Palla et al., 2012; Vesuviano et al., 2014) and data-driven (Abdalla et al., 2021). The use of conceptual hydrological models has been favored by many studies due to their simplicity, accuracy, and computational efficiency (Abdalla et al., 2022; Palla et al., 2012).

Conceptual hydrological models apply simplified equations to simulate the hydrological processes of green roofs. Due to the simplification of these equations, they depend on empirical parameters that are not physically measurable. Therefore, calibration is required to obtain optimal values for these parameters. The high dependency on calibration limits the application of conceptual models in cases when measured data are not available for calibration. Several studies have attempted to obtain explicit relationships between conceptual model parameters and physically measurable characteristics of green roofs. For instance, a handful of studies concluded that conceptual model parameters representing internal green roof storages could be estimated from the field capacity of green roof substrates (Abdalla et al., 2022; Stovin et al., 2013), which can be physically measured

(Fassman & Simcock, 2012). Moreover, parameters controlling flow movements and dynamics within the green roof layers were found to be correlated with roof properties such as the depth of the substrate layer (Soulis et al., 2017; Yio et al., 2013),the drainage layer type, and the slope of the roof (Vesuviano & Stovin, 2013). However, no explicit formulas were obtained that could estimate flow parameters solely from the physical roof characteristics.

Few studies have attempted to transfer calibrated models amongst similar roofs located in different cities, with the premise of physical similarity, a common approach in predicting flows for ungauged basins (Oudin et al., 2008; Tsegaw et al., 2019). For example, Johannessen et al. (2019) tested the transferability of calibrated parameters of the SWMM model (Rossman, 2015) between similar green roofs located in four Norwegian cities with different climatic conditions. However, only calibrated models from wetter cities (higher amount of precipitation) showed to yield satisfactory modelling results for the green roofs of the drier cities, but not vice versa, indicating an influence of climatic inputs on model parameters. Abdalla et al. (2021) attempted to transfer trained machine learning models between the same set of similar green roofs located in four Norwegian cities. They found the transferred models to yield satisfactory results only between cities with similar rainfall events characteristics.

The effect of climatic variables on conceptual model parameters has not been thoroughly discussed in the context of green roof modelling. It is particularly important not only for transferring calibrated parameters of conceptual models amongst similar green roofs located in different locations but also for utilizing calibrated conceptual models of green roofs for evaluating climate change scenarios in which the climatic variables are significantly different from the ones used for model calibration. Abdalla et al. (2022) tested and evaluated the performance of a conceptual green roof model for 16 green roofs located in the Norwegian cities. They discussed the effect of climatic data on the calibrated model parameters, in particular the flow parameters. They found high values of flow parameters for cities that receive rainfall events with higher amount and intensity and have shorter anticipant dry weather periods (ADWP), in comparison to cities with low precipitation amounts and longer ADWP. They acknowledged the difficulties of estimating flow parameters from climatic conditions.

Many studies that conducted hydrological modelling of large basins found the performance of conceptual models to reduce significantly when evaluated using different climatic conditions compared to the calibration period (Coron et al., 2012; Hartmann & Bárdossy, 2005). Fowler et al. (2016) discussed the effect of the calibration method on producing robust parameter sets that are applicable for contrasting climatic conditions. They recommended a calibration strategy based on multi-objective optimization to explore trade-offs between model performance in different climatic conditions. Similarly, Saavedra et al. (2022) found the hydrological models in their study to produce poor flow simulations in contrasting climatic conditions from calibration periods and proposed a model calibration strategy based on multi-objective optimizations for reducing the dependency of model parameters on climatic inputs.

A multi-objective optimization aims at approximating a Pareto front that contains a set of optimal solutions. In early hydrological modelling studies using Pareto front, the Pareto front was estimated by aggregating objective functions into one scalar value and running a series of independent optimization runs of the scalar value with varying weights of the objective functions (H. Madsen, 2000; Henrik Madsen, 2003). The development of algorithms that are customized for multi-objective problems, such as the nondominated sorting genetic algorithm II (NSGA-II) (Deb et al., 2002) and the multi-objective Shuffled Complex Evolution Metropolis (Vrugt et al., 2003) allows for efficient estimation of Pareto front. In recent years, several multi-objective algorithms were developed and evaluated in hydrological modelling studies. Examples include the multi-objective Artificial Bee Colony optimization algorithm (Huo & Liu, 2019), the differential evolution with adaptive Cauchy mutation and Chaos searching (MODE-CMCS) (Liu et al., 2016), and the multi-objective Bayesian optimization (M T M Emmerich et al., 2006). Some studies attempted to compare the performance of algorithms in the context of hydrological modelling (Guo et al., 2014; Wang et al., 2010).

This research sought to investigate a multi-objective optimization scheme for multi-site calibrations of sixteen extensive green roofs located in four Norwegian cities with different climatic conditions. The primary aim of this study is to demonstrate the possible advantages of multi-site calibration over single-site calibration for conceptual hydrological models of green roofs. Moreover, the study provides insights on the practical implication of multi-site optimization for urban stormwater management.

2 Green roof data

Sixteen extensive green roofs located in four Norwegian cities were used in this study. The cities are Bergen, Sandnes, Trondheim and Oslo. Bergen city receives the highest amount of annual precipitation of 3110 mm, followed by Sandnes city which receives annual precipitation of around 1700 mm. Both Sandnes and Bergen are classified as temperate oceanic climate (Cfb), according to Köppen–Geiger climate classification (Kottek et al., 2006). Trondheim is the northmost city with annual precipitation of around 1100 mm and has a subpolar oceanic climate (Dfc). The driest city in the study is Oslo, receiving annual precipitation of 970 mm, with a temperate oceanic climate (Cfb). A comparison between the rainfall characteristics of the four cities can be found in Abdalla et al. (2021).

The green roofs vary in geometries (i.e., width, length, and slope) among the four cities. According to similarities in configurations, they were categorized into five types, as shown in **Table 1**. Precipitation, outflow, and temperature were collected between 2015-2017 in one-minute resolution. The roofs in Oslo have a long record of data (from 2011-2017). The reader is directed to Johannessen et al. (2018) for more details about field measurements and data pre-processing.

3 Materials and Methods

3.1 The rationale for multi-site calibration

The performance of the calibration is typically assessed via objective functions such as the Nash Sutcliffe efficiency (Nash & Sutcliffe, 1970) and the Kling Gupta efficiency (Gupta et al., 2009). A single site calibration yields solutions that are near-optimal for the specific site. Many optimization algorithms used in hydrological modelling are stochastic, such as the shuffled complex evolution (SCE-UA) (Duan et al., 1992), resulting in different solutions for the same site and calibration setup. When these solutions are applied at another site with contrasting climatic conditions, they might result in poor solutions reflected by low values of objective functions. This was reported by Johannessen et al. (2019), attempting to transfer unchanged model parameters between similar green roofs located in different locations. On the other hand, the multi-site calibration explores trade-offs between model performance in different climatic conditions.

Table 1: Green roofs geometries and configurations

Roof type	Roof ID	Geometry			Configuration	
		Width (m)	Length (m)	Slope (%)	Substrate	Drainage mat
Type A	**BERG1**	1.6	4.9	16	VM (10 mm)	TR (10 mm)
	OSL3	2	4	5.5	VM (10 mm)	TR (10 mm)
	SAN1	1.6	5.3	27	VM (10 mm)	TR (10 mm)
	TRD1	2	7.5	16	VM (10 mm)	TR (10 mm)
Type B	**BERG3**	1.6	4.9	16	VM (10 mm) + MW (50 mm)	EPS (75 mm) + TR (5 mm)
	OSL2	2	4	5.5	VM (10 mm) + MW (50 mm)	HDPE (40 mm) + TR (5 mm)
	SAN2	1.6	5.3	27	VM (10 mm) + MW (50 mm)	EPS (75 mm) + TR (5 mm)
	TRD3	2	7.5	16	VM (10 mm) + MW (50 mm)	HDPE (25 mm) + TR (5 mm)
Type C	**BERG2**	1.6	4.9	16	VM (10 mm)	L+B (50 mm)
	SAN4	1.6	5.3	27	VM (10 mm)	L+B (50 mm)
	TRD2	2	7.5	16	VM (10 mm)	L+B (50 mm)
Type D	**BERG4**	1.6	4.9	16	VM (10 mm)	TR (3 mm)
	SAN3	1.6	5.3	27	VM (10 mm)	TR (3 mm)
Type E	**BERG5**	1.6	4.9	16	VM (10 mm) + Pumice (50 mm)	TR (3 mm)
	OSL1	2	4	5.5	VM (10 mm)	HDPE (25 mm)
	TRD4	2	7.5	16	VM (10 mm) + MW (50 mm)	PE (30 mm)

VM: vegetation mats (sedum)
MW: a mineral wool plate
TR: Textile retention fabric
L+B: a mixture of Leca and bricks
PE: plastic drainage layers of polyethylene
EPS: plastic drainage layers of expanded polystyrene
HDPE: plastic drainage layers of high-density polyethylene

Multi-site calibration aims to approximate the Pareto front containing non-dominated solutions for the objective functions used for calibration. Figure 1 presents a hypothetical Pareto front for two green roofs with contrasting climatic conditions. According to Figure 1, calibration solutions can be classified into one of five classes: i) theoretically possible that are neither acceptable for both green roofs, ii) theoretically impossible solutions, iii) solutions that are only acceptable for green roof 1, iv) solutions that are only acceptable for green roof 2 and v) solutions that are acceptable for both green roofs. The latter class is desirable for yielding parameters that are applicable to different climatic conditions.

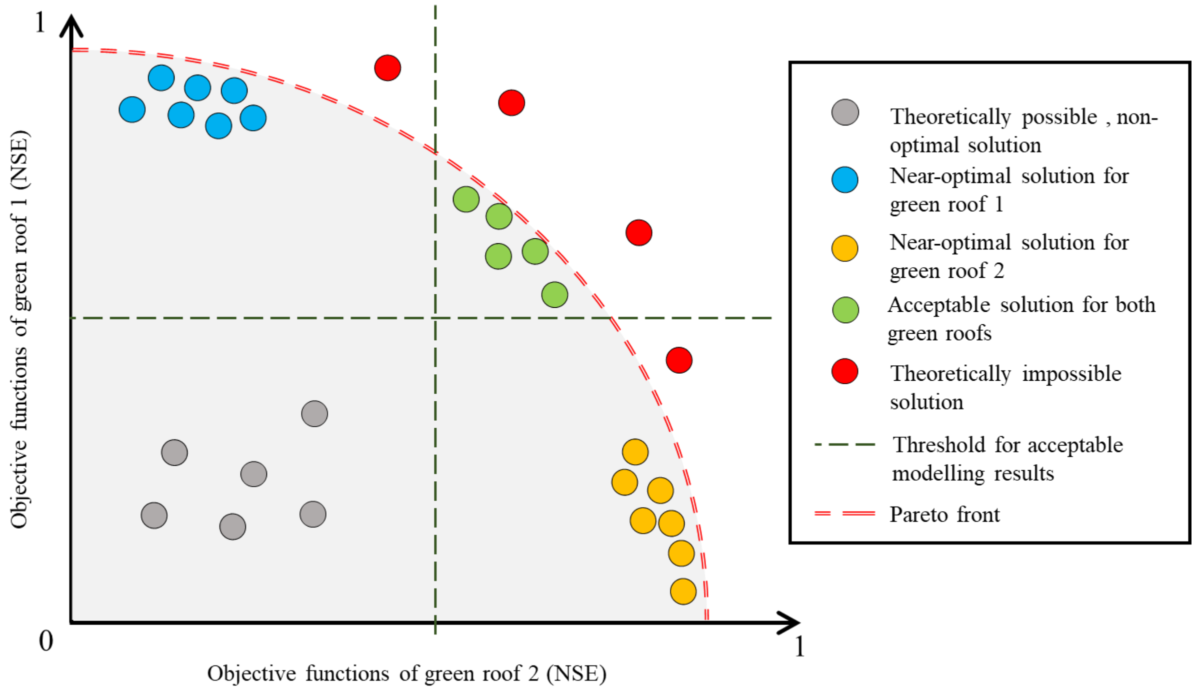


Figure 1: Results of two hypothetical calibrations of two green roofs plotted in two-dimensional objective space.

3.2 Calibration methods

3.2.1 Single-site calibration

In this study, single-site calibration refers to the process of obtaining optimal values of model parameters for a single green roof, using a single objective optimization algorithm (SOO). The differential evolution algorithm (DE) was used for single-site calibration (Storn & Price, 1997). DE is a stochastic algorithm that belongs to a family of optimization methods, referred to as evolutionary algorithms. These methods are suitable for global optimization and do not require the optimized function to be differentiable or continuous. DE generates populations of candidate solutions iteratively until a certain stoppage criterion is met. Each solution contains a vector of model parameters, and each population evolves from the previous one in such a way that each solution is either improved or remained the same. The initial generation is formed through random sampling of parameters from the user-defined ranges. To generate the next population, the DE applies a differential mutation process for each member of the current generation. In this process, three solutions (x0, x1, and x2) are randomly selected from the current population to produce a population of mutant solutions (v) for each member of the population as follows:

$$v = x_0 + F \,.\, (x1 - x2) \qquad \textit{Equation 1}$$

F, called the mutation factor, is a positive scale value typically less than 1. After the mutation process is done for each member of the population, the DE applies the crossover process which controls the fraction of parameters that are copied from the mutant or the original solution. A trial solution (u) is formed for each member of the population as follows:

$$u_{j,i,G+1} = \begin{cases} v_{j,i,G+1}, & if\ rand_{j,i} \leq CR\ or\ j = I_{rand} \\ x_{j,i,G}, & if\ rand_{j,i} > CR\ or\ j \neq I_{rand} \end{cases} \qquad \text{Equation 2}$$

Where $rand_{j,i}$ is a random real value between (0,1), j is the index of the parameter in the solution vector, i is the index of the solution in the population, G is the index of the population, CR is the cross-over probability, and I_{rand} is a random integer number between (1, D) where D is the number of solutions for each population. I_{rand} ensure that $u_{i,G+1} \neq x_{i,G}$. After the cross-over process, the DE applies the selection process, in which each solution from the current population is compared with its associated trial solutions from the cross-over process. the solution with the best objective value is selected for the next population. If the two objective functions are equal, the trial solution is selected for the next population.

In this study, the DEopim library in R was used (Mullen et al., 2011), and the KGE of the simulated outflow was selected as an objective function. The hyperparameter of the optimizer were selected as follow: CR = 0.5, F = 0.8, D = 10 * number of model parameters. The stoppage criterion was running the DE until the maximum number of populations (N = 200) was reached. Typically, the best solution in the last population is considered optimal in single-site calibration. In this study, however, the best solution for each population was considered a near-optimal solution. Hence, for each single site calibration, a group of 200 parameter sets was selected. Note that some solutions were duplicated since the best solution could remain the same in several populations.

3.2.2 Multi-site calibration

In this study, multi-site calibration refers to the process of estimating the Pareto front for two green roofs using a multi-objective optimization algorithm (MOO). For multi-site optimization, multi-objective Bayesian optimization (MBO) was selected. This algorithm requires a fewer number of model evaluations to approximate the Pareto front, in comparison to other multi-objective optimization methods (Binois & Picheny, 2019). The steps of the MBO are as follows:

i. Select an initial population of candidate solutions based on random sampling from the pre-defined parameter limits and determined the value of the objective functions of each solution.
ii. Apply the Pareto dominance test to extract non-dominance solutions, forming an initial Pareto front. A solution x1 is said to dominate solution x2 if and only if i) solution x1 is not worse than x2 in all objective functions and ii) x1 is better than x2 in at least one objective function. Non-dominated solutions are solutions that are not dominated by any member of the solution set.
iii. Build a surrogate model for each objective function from the candidate solutions. The Gaussian process was selected for building the surrogate model in this study (Binois & Picheny, 2019; Snoek et al., 2012; Worland et al., 2018).
iv. Select a new solution based on the surrogate models. The new solution is selected following a specific criterion that improves the Pareto front of the current iteration.
v. The selected solution is evaluated in the hydrological model, and its objective functions are determined and used to update the surrogate models of the objective functions.
vi. Repeat steps iii to v for N iteration (1000 in this study).

This study applied a common criterion for selecting potential solutions from surrogate models, termed the expected hypervolume improvement (Emmerich et al., 2011) which is presented in Figure 2.

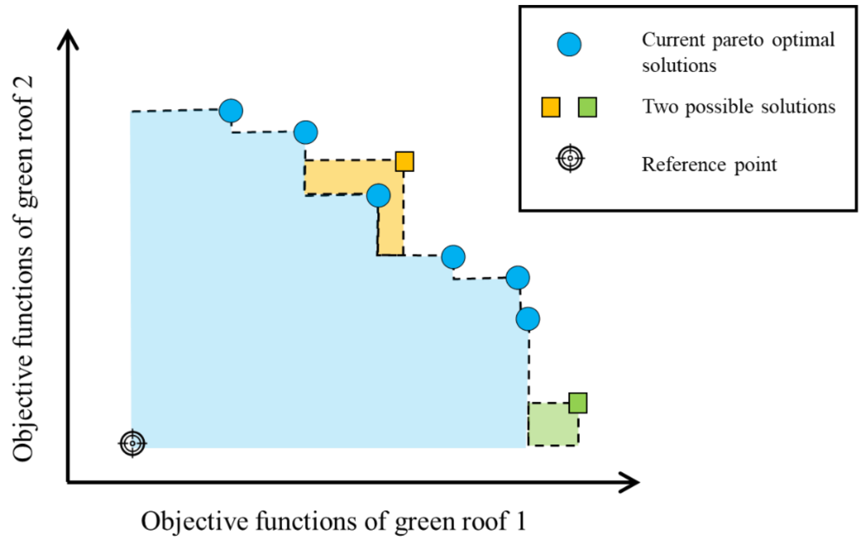


Figure 2: Hypervolume criterion for selecting potential solutions. The orange solution is better than the green solution based on the method. The orange solution maximizes the hypervolume which is measured from the reference point. modified from (Binois & Picheny, 2019)

The GPareto library in R (Binois & Picheny, 2019) was used for the multi-objective optimization in this study. The objective functions used were the KGE of simulated outflows for each green roof.

3.3 The hydrological model (CRRM linear)

The green roofs were modelled with a linear reservoir model (Figure 3). The model was developed and tested by Abdalla et al. (2022). It applies several equations (Equation 3 - Equation 10) to calculate infiltration (INF), drainage flow (Q), actual evapotranspiration (AET), soil moisture (SW), and drainage storage (DW). The potential evapotranspiration (PET) is determined using the Oudin formula (Oudin et al., 2005), which is suitable for cold climates and was found to be suitable by Johannessen et al. (2017) for cities in this study. The model contains five calibrated parameters; S1 (available storage of the soil layer), S2 (available storage of the drain layer), S11 (the threshold of soil water after which AET is equal to PET), k1 (flow parameter of the soil layer) and k2 (flow parameter of the drainage layer).

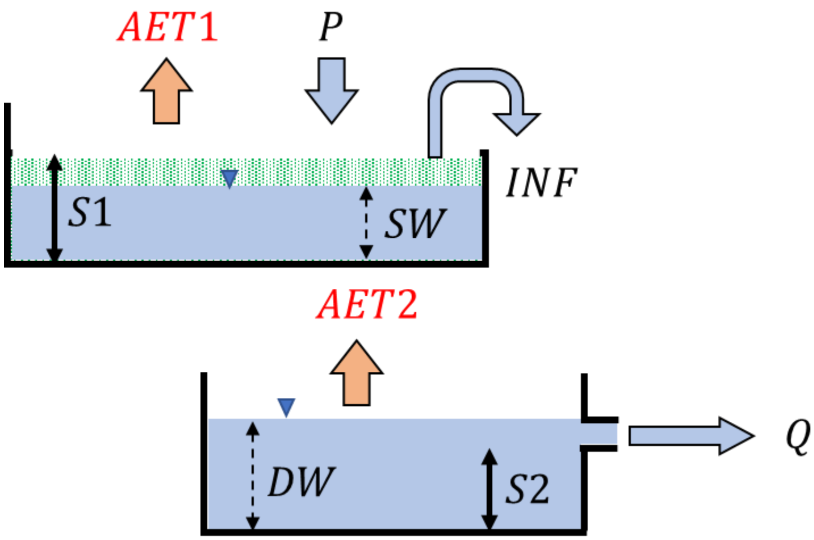


Figure 3: The linear reservoir model

$$INF_t = k1 \times \max(SW_t - S1, 0)$$ *Equation 3*

$$Q_t = k2 \times \max(DW_t - S2, 0)$$ *Equation 4*

$$SW_t = \max(SW_{t-1} + P_t - AET1_t - S1, 0)$$ *Equation 5*

$$DW_t = \max(DW_{t-1} + INF_t - AET2_t - S2, 0)$$ *Equation 6*

$$PET\left[\frac{mm}{day}\right] = \begin{cases} 0 \ if\ T_{mean} \leq 5℃ \\ \frac{Ra}{\lambda\rho} \times 0.01 \times (T_{mean} + 5)\ if\ T_{mean} > 5℃ \end{cases}$$ *Equation 7*

$$f_t = \min(1, \frac{SW_{t-1}}{S11})$$ *Equation 8*

$$AET1_t = f_t \times PET_t$$ *Equation 9*

$$AET2_t = \min(DW_{t-1}, PET_t - AET1_t)$$ *Equation 10*

3.4 Study experiments

To investigate the performance of the multi-objective optimization, three experiments were conducted as follows:

- Experiment one: calibration of two similar green roofs configurations in different sites
- Experiment two: calibration of two different green roofs configurations in the same site
- Experiment three: calibration of four similar green roofs configurations in different sites

In all experiments, the Pareto optimal solutions were compared with the results of single-site calibrations. Based on the value of KGE, the model results were classified as:

- Poor (KGE<0.5)
- Satisfactory (0.5<KGE<0.75)
- Good (KGE>0.75)

The classification followed the recommendation of Thiemig et al. (2013). It should be emphasized that such classification is based on the consensus of what is considered “good” or “poor” modelling results in the literature.

Measurements from 2017 were selected for model calibration while 2016 data were used for model validation. Snow periods (i.e., October to March) were excluded since the model does not simulate snow accumulation and melting. A 5 min time-step was use for the modelling. Hence, data were aggregated accordingly.

4 Results and discussion

4.1 Two-site calibration (similar roofs configurations on different sites)

The optimal solutions for the two-site and the single-site calibration schemes were plotted and compared. Figure 4 presents the comparisons for type A and type B roofs. Some solutions were found by the single-site calibration to yield poor model results when transferred to different sites. For instance, all parameter sets of OSL3 yielded KGE values below 0.1 for the TRD1 roof. In contrast, multi-site calibration yielded solutions that were satisfactory for both OSL3 and TRD1 roofs. In some cases, single-site calibration yielded satisfactory to good results for other roofs than the one used for calibration. For instance, all solutions of OSL3 roof yielded good to satisfactory results for BERG1, and vice versa. However, solutions found by the multi-site calibration for the two roofs were closer to the 1:1 line (i.e., best compromised solutions).

For some roofs, different parameter sets gave the same results for the same site, indicating equifinality (Beven, 1993). These solutions, however, yielded different results when transferred to different sites. For instance, optimal solutions that produced the same model performance at BERG3 yielded poor to satisfactory results for the OSL2 roof. This shows that single-site calibration could potentially miss promising solutions which produce satisfactory results in different locations. A similar conclusion was drawn in a study by Fowler et al. (2016), where they assessed the transferability of model parameters between dry and wet conditions.

4.2 Two-site calibration (different roofs configurations on the same site)

The comparison between the two calibration schemes (single vs multi-sites) is presented in Figure 5 for Bergen roofs. Almost all parameter sets found by the single-site calibration could yield satisfactory to good results in the other roofs with different configurations. Only a few parameter sets of BERG2 roofs yielded poor results for BERG3. On the other hand, results from the two-site calibration yielded better compromised results (closer to the 1:1 line).

It can be noted that climatic variables (i.e., location) could have a greater influence on model parameters than the roof's physical characteristics, as shown in Figure 5 as opposed to Figure 4. Similarly, Abdalla et al. (2021) found that ML trained in one location could yield satisfactory model performance for the different roof properties that are located in the same location.

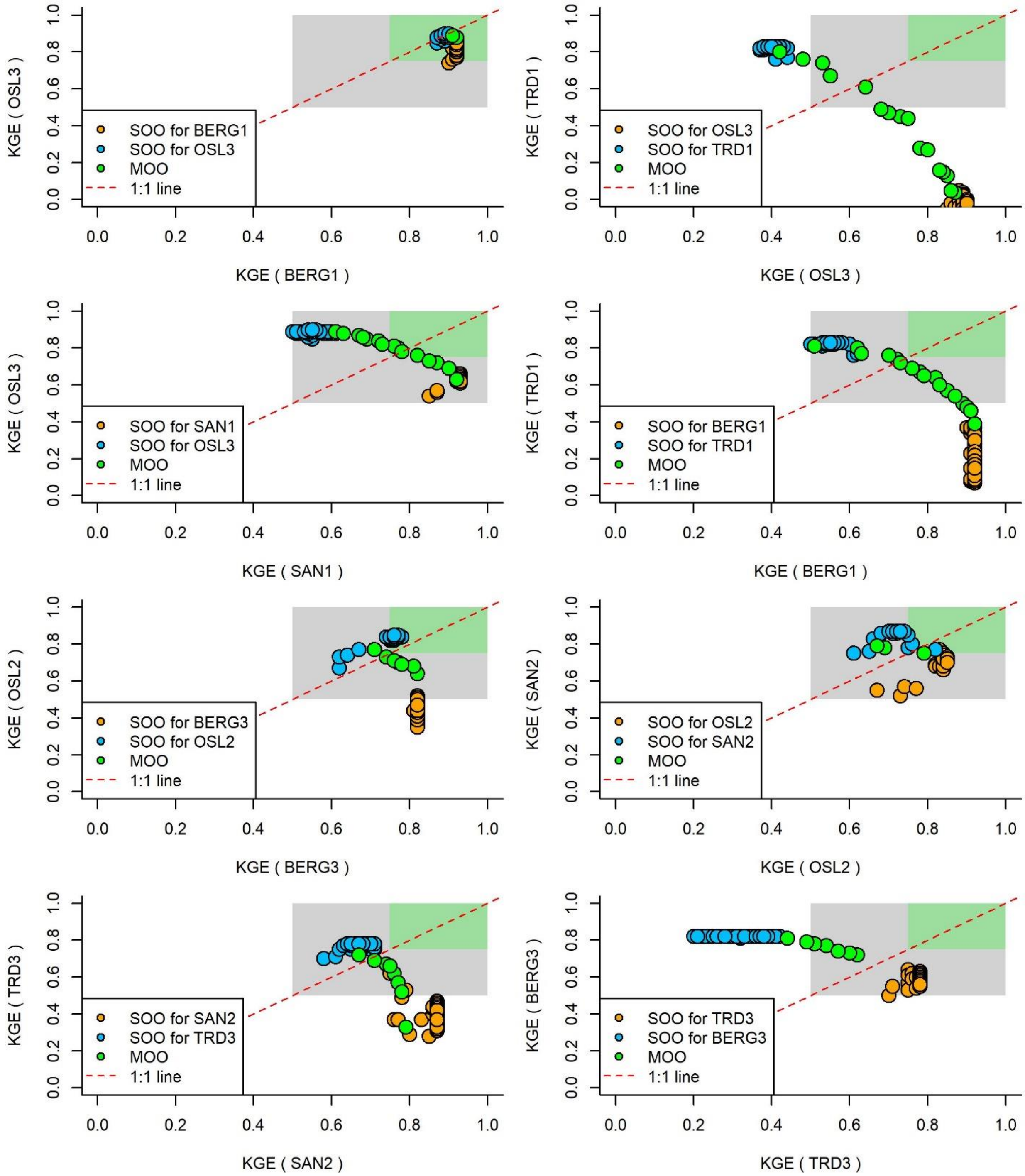


Figure 4: The comparison between two-site (MOO) and single-site (SOO) calibrations for similar green roofs located in different cities. Solutions that are close to the 1:1 line are considered the best compromised solutions. The grey-shaded area represents solutions that are considered satisfactory for both sites (0.5<KGE<0.75). The green-shaded area represents solutions that are considered good for both sites (KGE>0.75)

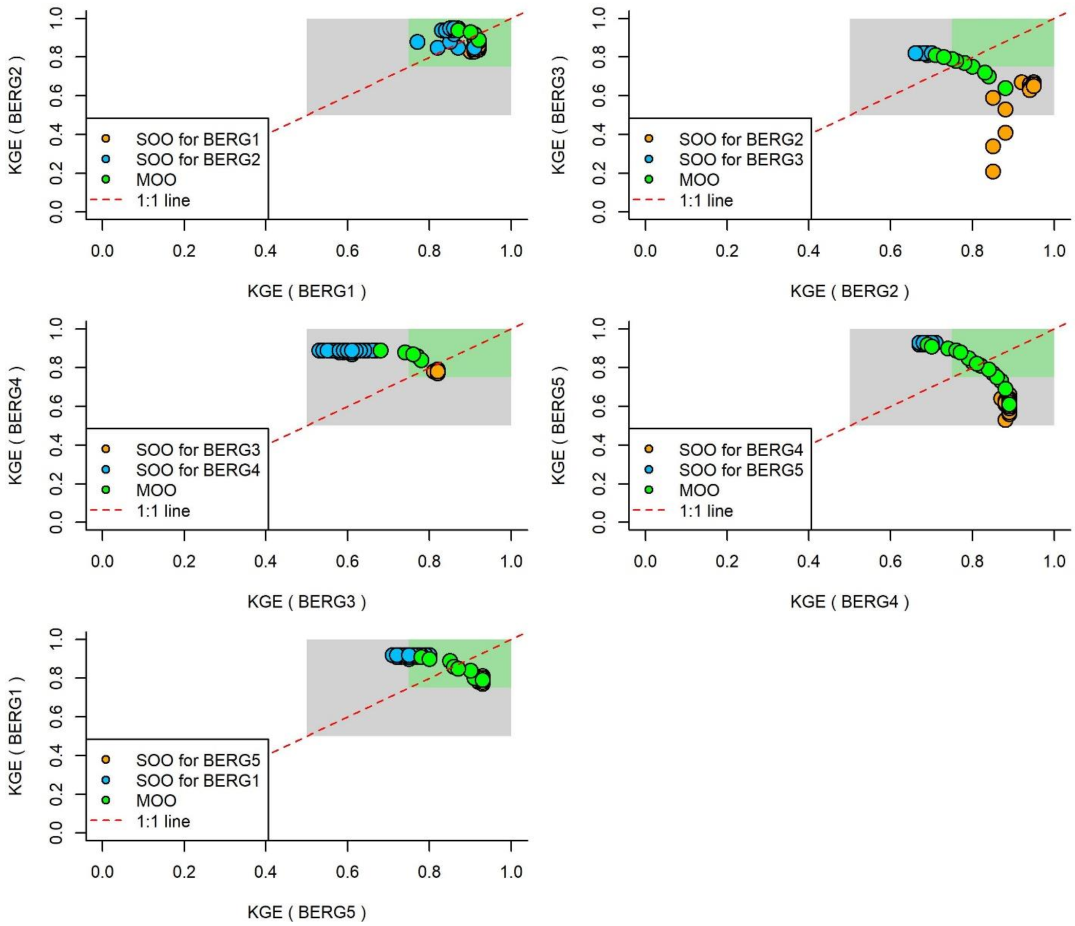


Figure 5: The comparison between two-site (MOO) and single-site (SOO) calibrations for different green roof types located at the same site (Bergen). The grey-shaded area represents solutions that are considered satisfactory for both sites (0.5<KGE<0.75). The green-shaded area represents solutions that are considered good for both sites (KGE>0.75)

4.3 Four site calibration (similar roofs in different cities)

The solutions of the single site calibration were used to simulate outflows for the green roofs in the other cities in the study. Figure 6 presents the performance of these simulations for type A and type B roofs. The result showed that transferring single site calibration results into different locations could yield poor modelling results. A similar finding was reported in the study of Johannessen et al. (2019) in which calibrated SWMM models were found to yield poor results when validated in multiple locations. As shown in Figure 6, transferability could yield satisfactory results between some cities (for instance, Bergen and Oslo). However, obtaining one parameter set from single-site calibration that produces satisfactory results in the four cities is very difficult, if not impossible. On the other hand, multi-site calibration resulted in a set of non-dominated solutions that allowed for exploring trade-offs of model performance amongst cities.

One parameter set that yielded the highest minimum KGE between the four locations was selected (S1 = 6.794, S11 = 8.378, k1 = 0.435, k2 = 0.031, S2 = 3.989). The selected set yielded KGE values ranging between 0.62 to 0.89 for the calibration periods and 0.6-0.82 for the validation periods, as shown in Table 2, for the four roofs which are considered satisfactory to good results. Figure 7 illustrates the simulated and observed outflows of type A roofs for the validation periods.

The simulated outflows matched well with observation, although some of the peak values were underestimated.

The selected parameter set was used to simulate outflows from the sixteen roofs in the study. Table 2 presents the performance of these simulations, as measured by KGE. All simulations yielded KGE values that were higher than 0.5 and some scored KGE above 0.75, indicating satisfactory to good results. Therefore, in contrast to single-site calibration, it is possible to obtain a common parameter set that yields satisfactory model results for different locations, by evaluating Pareto optimal solutions from multi-site calibration.

It could be noted that the variation of KGE values between locations and modelling period was slightly higher than between the different roof properties. For instance, the common set scored KGE values that ranged only between 0.58-0.63 for Oslo roofs (calibration periods), and only between 0.67-0.89 for Bergen roofs (validation periods). This further strengthens the conclusion that the influence of climatic variables on conceptual model parameters is higher than the influence of the roof properties.

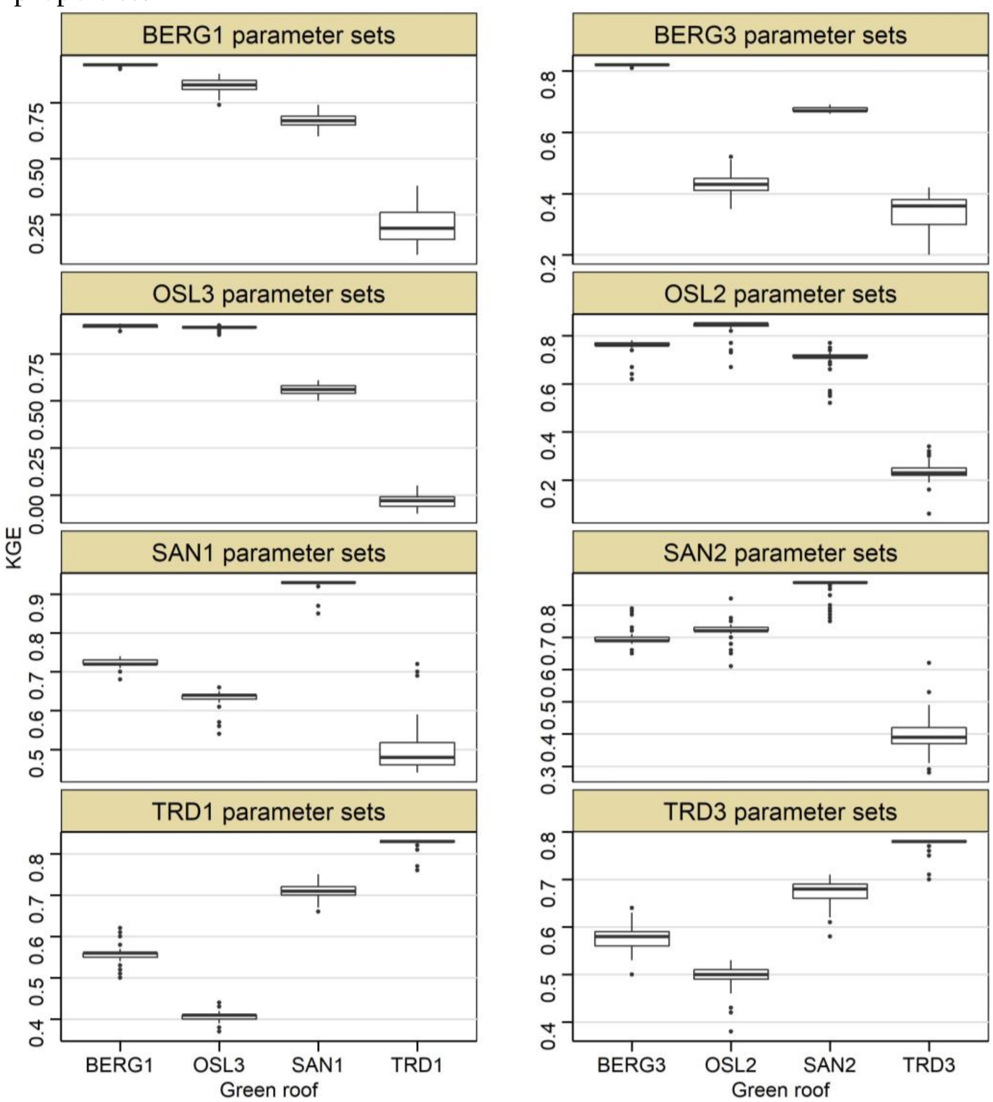
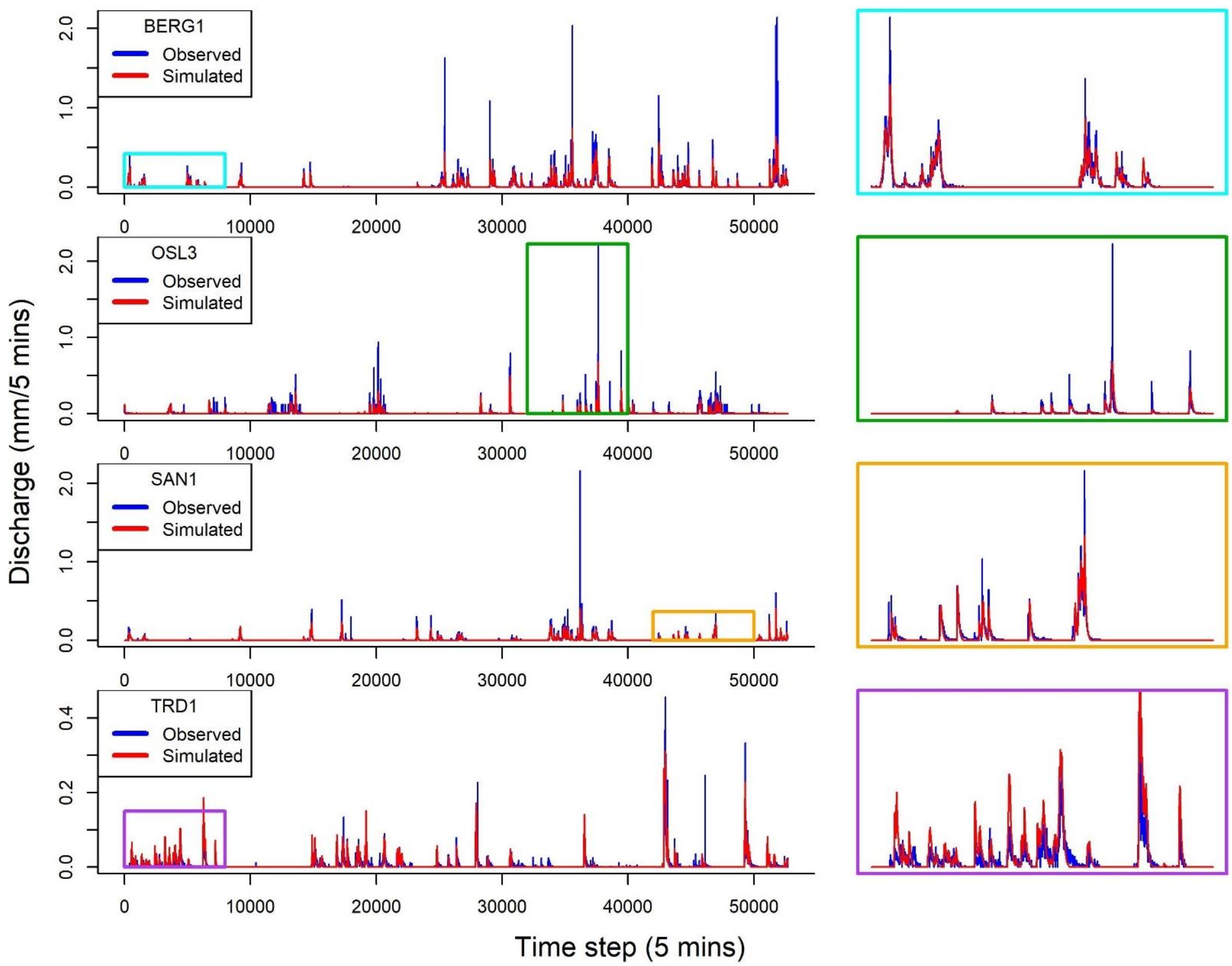


Figure 6: The performance of parameter sets obtained from single-site calibration in similar roofs located in other cities



331
332 Figure 7: The performance of the best compromised parameter set on the validation periods of the four roofs.

333 Table 2: The performance of the best compromised parameter set from the four site-calibration on the 16 green roofs

GR	KGE (Calibration periods)	KGE (validation periods)
BERG1	0.77	0.82
BERG2	0.86	0.86
BERG3	0.6	0.67
BERG4	0.66	0.68
BERG5	0.89	0.89
OSL1	0.59	0.65
OSL2	0.58	0.63
OSL3	0.63	0.67
SAN1	0.89	0.78
SAN2	0.76	0.61
SAN3	0.68	0.86
SAN4	0.82	0.84
TRD1	0.62	0.6
TRD2	0.68	0.79
TRD3	0.51	0.77
TRD4	0.68	0.86

334

4.4 Implications for stormwater management

Single-site calibration was found to yield optimal parameters for one location which performed poorly in the other sites, due to the different climatic conditions. In the future, climatic variables are expected to change significantly due to climate change (Sun et al., 2006). Therefore, a conceptual model calibrated with the current climate variables using a single-site scheme is likely to yield poorer simulations for the future. Nevertheless, this argument has rarely been discussed in the context of modelling sustainable stormwater measures, such as green roofs. It is a common practice in sustainable stormwater modelling studies to investigate climate change scenarios using a model calibrated with the current conditions. Therefore, caution should be exercised when interpreting the results of a model that is calibrated in contrasting climatic conditions from those used in model scenarios.

The results of this study are in-line with the common consensus in catchment modelling studies, in which hydrological models were found to score poor simulation results when evaluated on contracting climatic compared to those used for model calibration (Coron et al., 2012; Hartmann & Bárdossy, 2005. A solution which has been suggested by some scholars, is to calibrate models on climatic conditions similar to those used in model scenarios (C. Z. Li et al., 2012). For instance, if the model is intended to simulate wet conditions it must be calibrated on a wet condition period from the historical data. However, as argued by Fowler et al. (2016), this limits the applicability of calibrated model beyond the climatic conditions available in the historic periods. In green roof studies, observations are even more scarce than in large catchments which further limits the applicability of such an approach.

The results of this study show that obtaining a common parameter set that fit "reasonably well" for different locations and roof properties could be achieved by multi-site calibration. This is valuable for stormwater management, as it provides a fast and reliable tool for quantifying the hydrological impact of green roofs in different locations and climate change scenarios. It should be noted, however, that such a common parameter set typically will yield lower performance for one roof than the best parameter set from the single site calibration of that roof. A question is whether this decrease in performance affects the usefulness of the model for stormwater management.

Before answering this question, it is useful to discuss the common metrics used to quantify the hydrological benefits of green roofs. Typically, green roof performance is measured by assessing retention and detention. The former is the measure of how much water is retained (i.e., removed) via roof evapotranspiration. In the literature on green roof modelling, simple water balance models with hourly or daily time steps and suitable evapotranspiration equations were found to be sufficient for estimating retention (Abdalla et al., 2021; Bengtsson et al., 2005; Stovin et al., 2013). On the other hand, green roof detention refers to the reduction and delay of outflows due to the temporal storage of water in the green roof. Estimating detention requires calibrated models and short time steps (sub-hourly). Typically, detention is measured by event-based metrics, such as peak reduction, peak delay, etc. However, recent studies discussed issues of event-based metrics and suggested alternative approaches based on long term-simulations (Stovin et al., 2017). Among these alternatives, flow duration curves (FDCs) were found to provide an unambiguous estimation of green roof detention (Hernes et al., 2020). Hence, it was adopted in the study.

We investigated the accuracy of simulated FDCs from the common parameter set in Table 2 (regional set) and whether these FDCs are comparable with those derived from the best parameter set from the single-site calibration setup (best fit). Figure 8 presents the observed outflow and FDC of the BERG1 roof compared with the simulated results from the best fit, and the regional parameter sets. Both parameters sets underestimated the high flows. However, the best fit set produced better estimates of the high flows than the regional set. This represents the part of the FDC with low durations (e.g., less than 5 hours). For medium and low parts of the FDC (duration > 5 hours), the regional set produced slightly better estimates for medium and low values. Figure 9 presents the simulated and observed FDCs for the sixteen roofs in the study. For visualization purposes, the log-log scale was used. The regional parameter set produced FDCs that were comparable to those derived by the best fit sets for each roof. for cities with high and intense precipitation, such as Bergen, the best-fit parameters produced better estimates of high values while the regional set slightly produce better simulations for medium and low values. On the other hand, for Trondheim city, which receives lower precipitation amount and intensity, the regional set overestimated low values and provided a better estimate for high values.

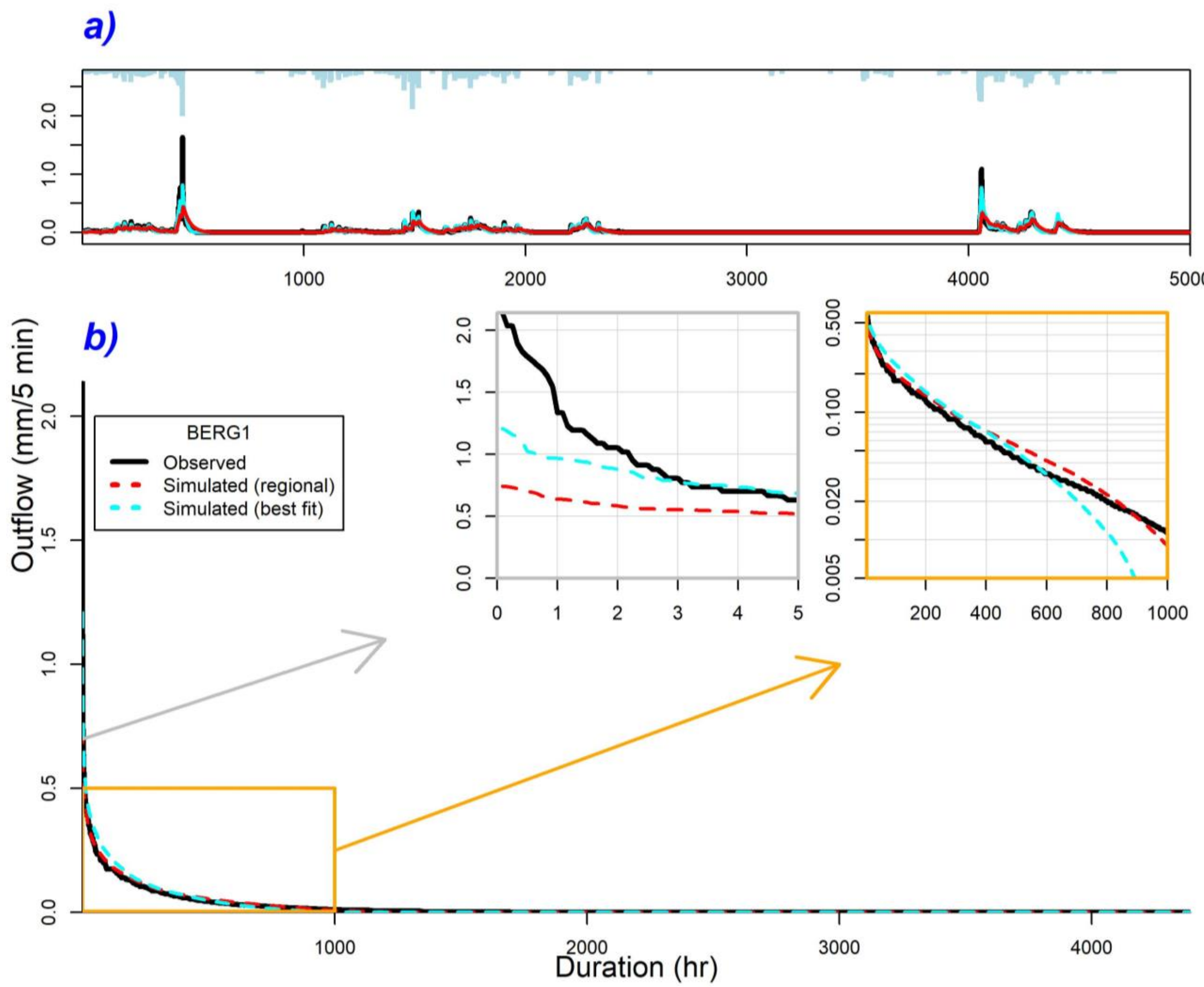


Figure 8: a) observed outflows of BERG1 roof compared by simulated outflows from the best parameter set (best fit) and the four-site calibration (Regional) for the selected period. b) Observed flow duration curve (FDC) of BERG1 compared by the simulated FDC obtained from the parameter set that produces the best fit at BERG1 (single site) and from the best compromised parameter set from the four-site calibration (Regional).

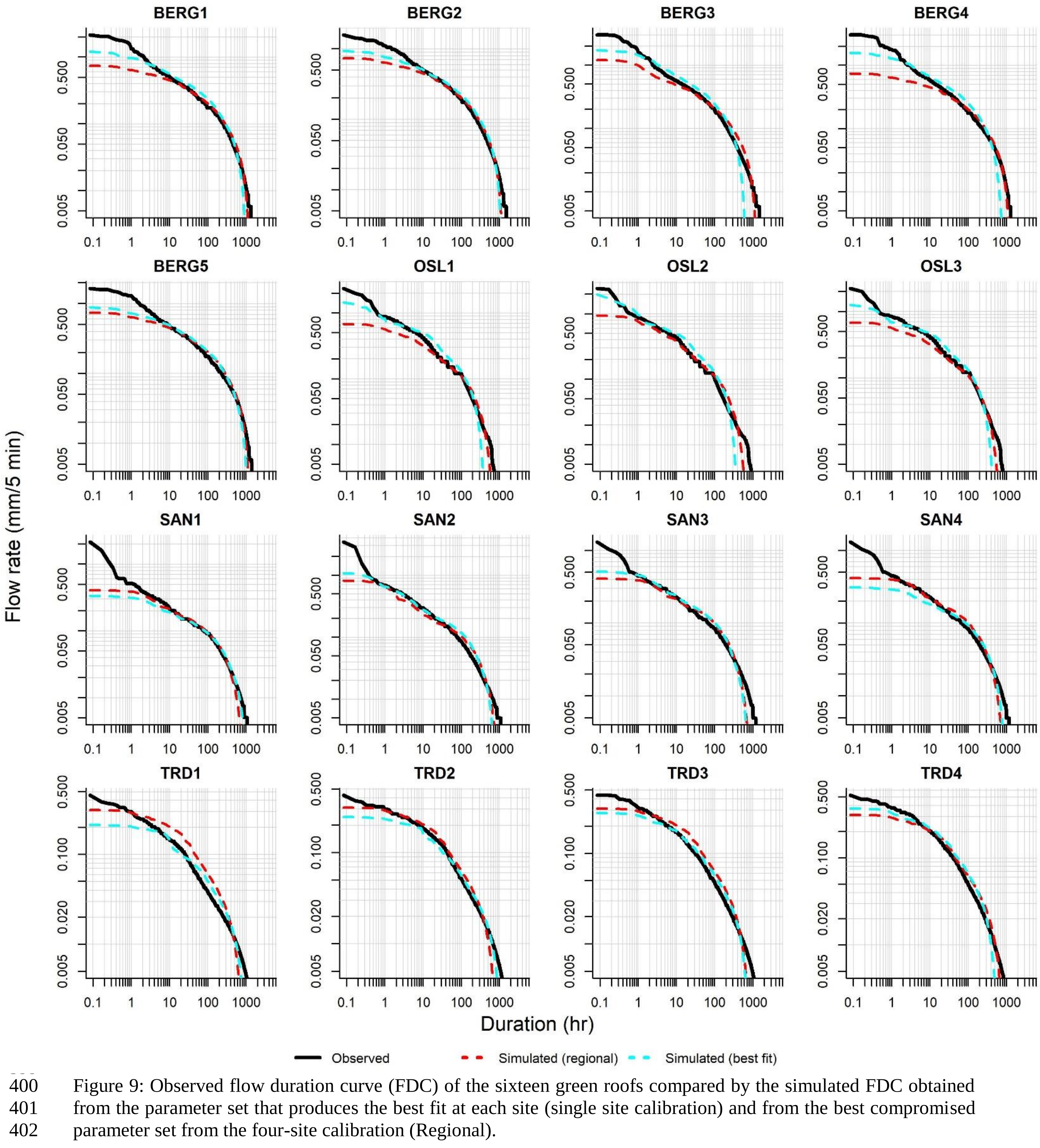


Figure 9: Observed flow duration curve (FDC) of the sixteen green roofs compared by the simulated FDC obtained from the parameter set that produces the best fit at each site (single site calibration) and from the best compromised parameter set from the four-site calibration (Regional).

5 Summary and conclusions

The current study aimed to evaluate the potential of multi-site calibration for conceptual hydrological models of green roofs. Additionally, the study provided insights on the practical implication of multi-site calibration, concerning stormwater management. Based on the results of the study, the following conclusions can be drawn:

- Single site calibration obtains optimal parameters for one site that perform poorly for other locations and climate conditions.
- The variation of model performance due to climatic variables is greater than due to roof properties.
- Obtaining a common parameter set that yields satisfactory (Kling Gupta Efficiency >0.5) for different locations and roof properties can be achieved by multi-site calibration. Such a parameter set provides flow durations curves that are comparable in accuracy to those derived from the best parameter sets from single-site calibration
- The multi-site calibration scheme is recommended not only for transferability among roofs in different cities but also when applying conceptual models for evaluating climate change scenarios for which the climatic variables are significantly different from the ones used for calibration.

Acknowledgments

The authors would like to acknowledge the financial support from the Research Council of Norway through the Centre for Research-based Innovation “Klima 2050” (www.klima2050.no)

References

Abdalla, Alfredsen, & Merete Muthanna. (2022). Towards improving the calibration practice of conceptual hydrological models of extensive green roofs. *Journal of Hydrology*, *607*(January), 127548. https://doi.org/10.1016/j.jhydrol.2022.127548

Abdalla, Pons, Stovin, De-Ville, Fassman-Beck, Alfredsen, & Muthanna. (2021). Evaluating different machine learning methods to simulate runoff from extensive green roofs. *Hydrology and Earth System Sciences*, *25*(11), 5917--5935. https://doi.org/10.5194/hess-25-5917-2021

Bengtsson, Grahn, & Olsson. (2005). Hydrological function of a thin extensive green roof in southern Sweden. *Nordic Hydrology*, *36*(3), 259–268. https://doi.org/10.2166/nh.2005.0019

Beven. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*(1), 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Binois, & Picheny. (2019). Gpareto: An r package for gaussian-process-based multi-objective optimization and analysis. *Journal of Statistical Software*, *89*(1). https://doi.org/10.18637/jss.v089.i08

Bouzouidja, Séré, Claverie, Ouvrard, Nuttens, & Lacroix. (2018). Green roof aging: Quantifying the impact of substrate evolution on hydraulic performances at the lab-scale. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2018.07.032

Coron, Andréassian, Perrin, Lerat, Vaze, Bourqui, & Hendrickx. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resources Research*, *48*(5). https://doi.org/https://doi.org/10.1029/2011WR011721

Deb, Pratap, Agarwal, & Meyarivan. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. https://doi.org/10.1109/4235.996017

Duan, Sorooshian, & Gupta. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/https://doi.org/10.1029/91WR02985

Emmerich, M T M, Giannakoglou, & Naujoks. (2006). Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, *10*(4), 421–439. https://doi.org/10.1109/TEVC.2005.859463

Emmerich, Michael T.M., Deutz, & Klinkenberg. (2011). Hypervolume-based expected improvement: Monotonicity properties and exact computation. *2011 IEEE Congress of Evolutionary Computation*, *CEC 2011*, 2147–2154. https://doi.org/10.1109/CEC.2011.5949880

Fassman, & Simcock. (2012). Moisture Measurements as Performance Criteria for Extensive Living Roof Substrates. *Journal of Environmental Engineering*. https://doi.org/10.1061/(asce)ee.1943-7870.0000532

Fowler, Peel, Western, Zhang, & Peterson. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, *52*(3), 1820–1846. https://doi.org/https://doi.org/10.1002/2015WR018068

Guo, Zhou, Lu, Zou, Zhang, & Bi. (2014). Multi-objective optimization of empirical hydrological model for streamflow prediction. *Journal of Hydrology*, *511*, 242–253. https://doi.org/10.1016/j.jhydrol.2014.01.047

Gupta, Kling, Yilmaz, & Martinez. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hamouz, Lohne, Wood, & Muthanna. (2018). Hydrological performance of LECA-based roofs in cold climates. *Water (Switzerland)*, *10*(3), 1–16. https://doi.org/10.3390/w10030263

Hartmann, & Bárdossy. (2005). Investigation of the transferability of hydrological models and a method to improve model calibration. *Advances in Geosciences*, *5*, 83–87. https://doi.org/10.5194/adgeo-5-83-2005

Hernes, Gragne, Abdalla, Braskerud, Alfredsen, & Muthanna. (2020). Assessing the effects of four SUDS scenarios on combined sewer overflows in Oslo, Norway: Evaluating the low-impact development module of the Mike Urban model. *Hydrology Research*, *51*(6), 1437–1454. https://doi.org/10.2166/nh.2020.070

Huo, & Liu. (2019). Application research of multi-objective Artificial Bee Colony optimization algorithm for parameters calibration of hydrological model. *NEURAL COMPUTING \& APPLICATIONS*, *31*(9, SI), 4715–4732. https://doi.org/10.1007/s00521-018-3483-4

Jim. (2014). Air-conditioning energy consumption due to green roofs with different building thermal insulation. *APPLIED ENERGY*, *128*, 49–59. https://doi.org/10.1016/j.apenergy.2014.04.055

Johannessen, Hamouz, Gragne, & Muthanna. (2019). The transferability of SWMM model parameters between green roofs with similar build-up. *Journal of Hydrology*, *569*(October 2018), 816–828. https://doi.org/10.1016/j.jhydrol.2019.01.004

Johannessen, Hanslin, & Muthanna. (2017). Green roof performance potential in cold and wet regions. *Ecological Engineering*, *106*, 436–447. https://doi.org/10.1016/j.ecoleng.2017.06.011

Johannessen, Muthanna, & Braskerud. (2018). Detention and Retention Behavior of Four Extensive Green Roofs in Three Nordic Climate Zones. *Water*, *10*(6), 671. https://doi.org/10.3390/w10060671

Jungels, Rakow, Allred, & Skelly. (2013). Attitudes and aesthetic reactions toward green roofs in the Northeastern United States. *LANDSCAPE AND URBAN PLANNING*, *117*, 13–21. https://doi.org/10.1016/j.landurbplan.2013.04.013

Kottek, Grieser, Beck, Rudolf, & Rubel. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Li, C. Z., Zhang, Wang, Zhang, Yu, & Yan. (2012). The transferability of hydrological models under nonstationary climatic conditions. *Hydrology and Earth System Sciences*, *16*(4), 1239–1254. https://doi.org/10.5194/hess-16-1239-2012

Li, Yanling, & Babcock. (2015). Modeling Hydrologic Performance of a Green Roof System with HYDRUS-2D. *Journal of Environmental Engineering (United States)*, *141*(11), 1–9. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000976

Liu, Guo, Sun, Zhang, Wang, & Zhou. (2016). Multiobjective Optimal Algorithm for Automatic Calibration of Daily Streamflow Forecasting Model. *Mathematical Problems in Engineering*, *2016*, 8215308. https://doi.org/10.1155/2016/8215308

Madsen, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, *235*(3–4), 276–288. https://doi.org/10.1016/S0022-1694(00)00279-1

Madsen, Henrik. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, *26*(2), 205–216. https://doi.org/10.1016/S0309-1708(02)00092-1

Mullen, Ardia, Gil, Windover, & Cline. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v040.i06

Nash, & Sutcliffe. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Oudin, Andréassian, Perrin, Michel, & Le Moine. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*(3), 1–15. https://doi.org/10.1029/2007WR006240

Oudin, Hervieu, Michel, Perrin, Andréassian, Anctil, & Loumagne. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, *303*(1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

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

Palla, Gnecco, & Lanza. (2012). Compared performance of a conceptual and a mechanistic hydrologic models of a green roof. *Hydrological Processes*, *26*(1), 73–84. https://doi.org/10.1002/hyp.8112

Refahi, & Talkhabi. (2015). Investigating the effective factors on the reduction of energy consumption in residential buildings with green roofs. *RENEWABLE ENERGY*, *80*, 595–603. https://doi.org/10.1016/j.renene.2015.02.030

Rossman. (2015). STORM WATER MANAGEMENT MODEL USER'S MANUAL Version 5.1. *National Risk Management Laboratory Office of Research and Development. United States Environmental Protection Agency, Cincinnati, Ohio.*, *September*, 352. http://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100N3J6.TXT

Saavedra, Mendoza, Addor, Llauca, & Vargas. (2022). A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrological Processes*, *36*(1), e14446. https://doi.org/https://doi.org/10.1002/hyp.14446

Snoek, Larochelle, & Adams. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*.

Soulis, Valiantzas, Ntoulas, Kargas, & Nektarios. (2017). Simulation of green roof runoff under different substrate depths and vegetation covers by coupling a simple conceptual and a physically based hydrological model. *Journal of Environmental Management*. https://doi.org/10.1016/j.jenvman.2017.06.012

Storn, & Price. (1997). Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

Stovin, Poë, & Berretta. (2013). A modelling study of long term green roof retention performance. *Journal of Environmental Management*, *131*, 206–215. https://doi.org/10.1016/j.jenvman.2013.09.026

Stovin, Vesuviano, & De-Ville. (2017). Defining green roof detention performance. *Urban Water Journal*, *14*(6), 574–588. https://doi.org/10.1080/1573062X.2015.1049279

Sun, Solomon, Dai, & Portmann. (2006). How often does it rain? *Journal of Climate*. https://doi.org/10.1175/JCLI3672.1

Susca, Gaffin, & Dell'Osso. (2011). Positive effects of vegetation: Urban heat island and green roofs. *ENVIRONMENTAL POLLUTION*, *159*(8–9, SI), 2119–2126. https://doi.org/10.1016/j.envpol.2011.03.007

Thiemig, Rojas, Zambrano-Bigiarini, & De Roo. (2013). Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2013.07.012

Tsegaw, Alfredsen, Skaugen, & Muthanna. (2019). Predicting hourly flows at ungauged small rural catchments using a parsimonious hydrological model. *Journal of Hydrology*, *573*(March), 855–871. https://doi.org/10.1016/j.jhydrol.2019.03.090

Vesuviano, Sonnenwald, & Stovin. (2014). A two-stage storage routing model for green roof runoff detention. *Water Science and Technology*, *69*(6), 1191–1197. https://doi.org/10.2166/wst.2013.808

Vesuviano, & Stovin. (2013). A generic hydrological model for a green roof drainage layer. *Water Science and Technology*. https://doi.org/10.2166/wst.2013.294

Vrugt, Gupta, Bastidas, Bouten, & Sorooshian. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, *39*(8), 1–19. https://doi.org/10.1029/2002WR001746

Wang, Lei, Jiang, & Wang. (2010). Performance comparison of three multi-objective optimization algorithms on calibration of hydrological model. *Proceedings - 2010 6th International Conference on Natural Computation, ICNC 2010*, *6*(Icnc), 2798–2803. https://doi.org/10.1109/ICNC.2010.5583573

Wooster, Fleck, Torpy, Ramp, & Irga. (2022). Urban green roofs promote metropolitan biodiversity: A comparative case study. *BUILDING AND ENVIRONMENT*, *207*(A). https://doi.org/10.1016/j.buildenv.2021.108458

Worland, Farmer, & Kiang. (2018). Improving predictions of hydrological low-flow indices in ungaged basins using machine learning. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2017.12.021

Yio, Stovin, Werdin, & Vesuviano. (2013). Experimental analysis of green roof substrate detention characteristics. *Water Science and Technology*, *68*(7), 1477–1486. https://doi.org/10.2166/wst.2013.381