https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

Evaluating different machine learning methods to simulate runoff from extensive green roofs

Elhadi Mohsen Hassan Abdalla[1], Vincent Pons[1], Virginia Stovin[2], Simon De-Ville[2], Elizabeth Fassman-Beck[3], Knut Alfredsen[1], and Tone Merete Muthanna[1]

[1]Department of Civil and Environmental Engineering, The Norwegian University of Science and Technology, Trondheim, 7031, Norway
[2]Department of Civil and Structural Engineering, The University of Sheffield, Sheffield, S1 3JD, United Kingdom
[3]Southern California Coastal Water Research Project, Costa Mesa, California, CA 92626, United States

Correspondence: Elhadi Mohsen Hassan Abdalla (Elhadi.m.h.abdalla@ntnu.no)

Abstract. Green roofs are increasingly popular measures to permanently reduce or delay stormwater runoff. Conceptual and physically-based hydrological models are powerful tools to estimate their performance. However, physically-based models are associated with a high level of complexity and computation costs while parameters of conceptual models are more difficult to obtain when measurements are not available for calibration. The main objective of the study was to examine the potential of using machine learning (ML) to simulate runoff from green roofs to estimate their hydrological performance. Four machine learning methods, Artificial Neural Network (ANN), M5 Model tree, Long Short-Term Memory (LSTM) and k-Nearest Neighbour (kNN) were applied to simulate stormwater runoff from sixteen extensive green roofs located in four Norwegian cities across different climatic zones. The potential of these ML methods for estimating green roof retention was assessed by comparing their simulations with a proven conceptual retention model. Furthermore, the transferability of ML models between the different green roofs in the study was tested to investigate the potential of using ML models as a tool for planning and design purposes. The ML models yielded low volumetric errors that were comparable with the conceptual retention models, which indicates good performance in estimating annual retention. The ML models yielded satisfactory modelling results ($NSE > 0.5$) in both training and validation data which indicates an ability to estimate green roof detention. The variations in ML models' performance between the cities was larger than between the different configurations, which was attributed to the different climatic characteristics between the four cities. Transferred ML models between cities with similar rainfall events characteristics (Bergen-Sandnes, Trondheim-Oslo) could yield satisfactory modelling performance ($NSE > 0.5, |PBIAS| < 25\%$) in most cases. However, we recommend the use of the conceptual retention model over the transferred ML models, to estimate the retention of new green roofs, as it gives more accurate volume estimates. Follow-up studies are needed to explore the potential of ML models in estimating detention from higher temporal resolution datasets.

1 Introduction

Green roofs are a type of green infrastructure (GI) that have received significant attention in recent years. In contrast to conventional stormwater infrastructure, green roofs attempt to decrease stormwater runoff while improving its quality, preserve

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

ecosystems and improve urban visual amenity among other benefits (Berndtsson, 2010) . Roof areas represent around 40-50% of impermeable areas in urban catchments (Dunnett and Kingsbury, 2004); therefore, retrofitting current roofs with substrate/growing media and vegetation offers an efficient and area-free GI option. Many studies have confirmed the potential of green roofs to mitigate rainfall events from field measurements (Fassman-Beck et al., 2013; Johannessen et al., 2018; Liu and Chui, 2019; Stovin, 2010).

Quantifying the hydrological performance of a green roof is usually done by estimating **retention**, a permanent reduction of stormwater by evapotranspiration, and **detention**, flow peak reduction and delay. Both retention and detention metrics are needed to justify the widespread implementation of green roofs by the stormwater community, and for planning and design by practicing engineers. Hence, numerous studies have investigated different approaches and tools to simulate runoff from green roofs to estimate retention and detention metrics.

For estimating green roof detention, models that simulate rainfall-runoff events in short time steps (sub-hourly) are required. Several models have been tested successfully in the literature, which can be categorized into physically-based and conceptual models. Physically-based models simulate the water flow in porous media by solving physical equations numerically, such as the Richards equations, either in 1D (Bouzouidja et al., 2018; Hilten et al., 2008; Liu and Fassman-Beck, 2017; Peng et al., 2019), 2D (Li and Babcock Jr, 2015; Palla et al., 2009) or 3D (Brunetti et al., 2016). These models were proven to be accurate and to rely only on measurable parameters (Sims et al., 2019). However, they require a significant amount of work to estimate their parameters through extensive lab measurements (Li and Babcock Jr, 2015) or expensive calibration, which makes these models too complex and computationally costly for practical applications. Moreover, many studies reported issues in estimating unsaturated hydraulic conductivity (Liu and Fassman-Beck, 2017, 2018; Peng et al., 2019), which is crucial in solving the Richards equation.

A less complex physically-based model is the EPA-SWMM (Rossman et al., 2010) which is applied in many studies to estimate green roof detention from event-based simulations (Baek et al., 2020; Hamouz and Muthanna, 2019; Krebs et al., 2016; Peng and Stovin, 2017; Russwurm et al., 2018). SWMM is widely used in green roof modelling, and it is considered as a suitable tool for assessing the hydrological performance of green roofs (Cipolla et al., 2016). However, the SWMM model has major limitations that affect its results. For instance, the model lacks a proper representation of the storage in the drainage layer which is compensated by the model during calibration by assigning unrealistic soil and drainage parameters resulting in equifinality issue (Hernes et al., 2020), in which different parameter sets give the same model outputs (Beven, 1993). These limitations force the SWMM model to rely on calibrated rather than measured parameters. Johannessen et al. (2019) attempted to transfer calibrated SWMM model parameters between similar green roofs located in different locations. However, only parameter sets from wet locations yielded good results in drier locations but not vice versa.

Conceptual models simplify the physical processes using linear or nonlinear equations to simulate green roof runoff. One common type of these models is the reservoir routing model which was applied to estimate runoff detention from event-based simulations in previous literature (Palla et al., 2012; Soulis et al., 2017; Vesuviano et al., 2014). These models were found to produce results that are comparable to physically-based models with much lower computational costs and level of complexity (Peng et al., 2019). Palla et al. (2012) recommended the use of a reservoir routing model instead of physically-based models

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

for design purposes when little information is available about the green roof properties. However, the parameters of conceptual models are not measurable. Hence, calibration is needed to find their optimal values, unlike physically-based models (Peng et al., 2019). A few studies have identified relations between the flow parameters of reservoir models and some physical properties of green roofs, such as slope and substrate depth (Vesuviano and Stovin, 2013; Yio et al., 2013). However, these studies focused on lab-scale green roofs in which detention due to the horizontal flow is not significant (Sims et al., 2019).

For estimating green roof retention, models with water balance equations (in hourly or daily time step) and suitable representation of the actual evapotranspiration process (AET) were found by many studies to be sufficient (Bengtsson et al., 2005; Jahanfar et al., 2018; Johannessen et al., 2017; Stovin et al., 2013). The most common way to model AET is by multiplying the potential evapotranspiration (PET), the maximum evaporation rate assuming unlimited water supply, with reduction functions that account for soil moisture deficit and crop type. The reduction functions require careful parameterization of the maximum storage of the roof and crop factors. The maximum storage of the roof was found by many studies to be related to the measurable field capacity of the substrate (Liu and Fassman-Beck, 2017; Stovin et al., 2013). Crop factors for agricultural crops are well documented and studied (Allen et al., 1998). However, crop factor values for Sedum plants, commonly applied for green roofs, are less known. Previous studies reported different crop factor values for Sedum plants (Berretta et al., 2014; Rezaei et al., 2005; Sherrard Jr and Jacobs, 2012).

Data-driven models, which are derived entirely from observed data, may offer alternative modelling tools that can estimate both retention and detention of green roofs without explicitly accounting for complex hydrological processes. However, the use of data-driven models in green roof studies has been limited to simple regression models (Carson et al., 2013) which are site-specific and not transferable. Arguably, the data conserve all information needed to simulate green roof hydrology, and the only lack is our understanding of how these variables are connected. We hypothesize in this study that a powerful nonlinear data-driven model, such as Machine Learning (ML), could be well suited for simulating runoff from green roofs while avoiding the uncertainties and difficulties in modelling the complex hydrological processes in green roofs, such as evapotranspiration and horizontal runoff routing.

Machine learning methods have been successfully applied in hydrological modelling in recent decades. Previous studies reported better performances of ML models compared to conventional hydrological models in runoff prediction (Solomatine and Dulal, 2003; Yilmaz and Muttil, 2014; Young et al., 2017), runoff simulation (Javan et al., 2015; Kratzert et al., 2018) , and for building relationships between water level and discharge (Bhattacharya and Solomatine, 2005). Some of the popular Machine Learning methods applied in hydrology include Artificial Neural Networks (ANN), M5 model tree, Long Short-Term Memory (LSTM), and k Nearest Neighbours (kNN).

Artificial Neural Network is the most common and among the earliest ML used in hydrological modelling (Daniell, 1991). Early examples of research into ANN includes the study conducted by Hsu et al. (1995), in which ANN outperformed the linear ARMAX and the conceptual Sacramento SAC-SMA model in simulating runoff from a medium-sized catchment. Likewise, Tokar and Johnson (1999) compared an ANN to a simple conceptual model and found the former to outperform the latter. M5 model tree has been applied in different studies. Solomatine and Dulal (2003), reported a satisfactory performance of both M5 model tree and ANN in runoff forecasting. They, however, emphasized the advantages of M5 model tree over ANN

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

due to the better interpretation of M5 model outputs. Goyal et al. (2013b) applied the M5 model tree for flow forecasting in India, among other ML methods and found it to perform satisfactorily. Away from flow simulation, Gharaei-Manesh et al. (2016) used M5 tree and other methods to simulate the spatial distribution of snow depths in Iran, while Goyal et al. (2013a) evaluated M5 model tree on formulating operation rules for a reservoir. Kisi (2016) used M5 model tree to model reference evapotranspiration.

LSTM has been applied in different scientific fields and could provide good results (Shen, 2018). Regarding runoff modelling, Kratzert et al. (2018) investigated the potential of LSTM to predict runoff from ungauged basins. They could achieve good prediction performance that was comparable to the well-known Sacramento model. Similarly,Ayzel (2019) obtained comparable results with LSTM to a conceptual model. Hu et al. (2018) compared between an ANN and an LSTM in runoff simulation and found the latter to outperform the former. Nevertheless, LSTM is computationally expensive, and the training process takes a long time (Ayzel, 2019). k-Nearest Neighbour was applied first by Karlsson and Yakowitz (1987) in runoff forecasting in which it outperformed unit hydrograph forecasters. Modaresi et al. (2018) found the k-Nearest Neighbour to be comparable with ANN in monthly runoff forecasting. Furthermore, Wu et al. (2009) applied the k-Nearest Neighbour in predicting monthly runoff, and they discussed the effect of k value on the performance of kNN.

Few studies have modelled green infrastructure with ML techniques. For instance, Tsang and Jim (2016) applied a Fuzzy-neural network to optimize irrigation of a green roof by estimating soil moisture deficit. The neural network could reproduce the soil moisture well, which indicates the capability of ML models to simulate the nonlinear AET process. Li et al. (2019) developed an artificial neural network model to predict the flow reduction from a catchment with different GI structures. Similarly, Radfar and Rockaway (2016) applied a neural network model to predict flow reduction from a permeable pavement. The satisfactory performances of ML models in two studies demonstrate the potential of ML models in GI hydrological modelling.

This study examines the ability of four machine learning methods, M5 model Tree, Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and k-Nearest Neighbour (kNN), to estimate green roof hydrological performance, specifically by:

1. Evaluating the performance of ML models in simulating the temporal dynamics of green roof runoff and estimating the retention from long term simulations across different climatic locations.

2. Investigating the potential of using ML models as a useful tool for planning that predicts the performance of new green roofs where runoff data are not available for model calibration nor for assessing the suitability of AET equations.

2 Data

Sixteen extensive green roofs located in four Norwegian cities with different climates: Bergen (BERG), Sandnes (SAN), Oslo (OSL) and Trondheim (TRD) were used in the study. Bergen city is located on the western coast of Norway. Bergen is the wettest city among the four with annual precipitation of 3110 mm followed by Sandnes city, which is located on the south-west

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

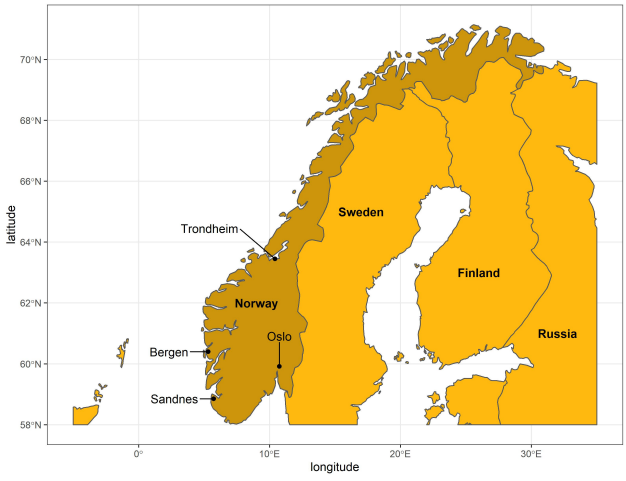


Figure 1. Locations of the four Norwegian cities with green roof field data

coast, with annual precipitation of 1690 mm. Oslo is the driest city with only 970 mm of annual precipitation while Trondheim is the northmost city with annual precipitation of 1070 mm. According to the Köppen–Geiger Climate Classification (Kottek et al., 2006) , both Bergen and Sandnes are classified as temperate oceanic climate (Cfb), while Oslo has the warm-summer humid continental climate (Dfb) and Trondheim has a subpolar oceanic climate (Dfc). The locations of the four cities are shown in Figure 1. Table 1 shows the geometries and configurations of roofs. Roof geometries (areas and slopes) vary between the cities, while the different configurations represent the variety of options in the Norwegian green roof market. Some green roofs in the study have the same configuration, for instance, BERG1, SAN1 OSL3 and TRD1. Continuous times series data were collected from TRD, BERG and SAN roofs between 2015 to 2017, while the green roofs at OSL have a seven-year record of data from 2011 to 2017. Data includes precipitation, runoff, relative humidity and wind speed at a 1 min resolution. In Oslo, the wind speed was not measured at the roofs but collected from a nearby station. For details about roof setup, data collection and processing, please refer to Johannessen et al. (2018).

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Table 1. Roof Geometries and Configurations

Roof	Geometry			Configuration			
	Width	Length	Slope	Vegetation mat thick-ness1	Extra Substrate type and thickness	Drain mat type and thickness	Total roof thickness
	(m)	(m)	(%)	(mm)			(mm)
BERG1	1.6	4.9	16	30	-	Textile retention fabric (10mm)	40
BERG2	1.6	4.9	16	30	-	Substrate mat2 (50mm)	80
BERG3	1.6	4.9	16	30	Separate Substrate3 (50mm)	Drainage layer5 (EPS) (75mm) + Textile retention fabric (5mm)	160
BERG4	1.6	4.9	16	30	-	Textile retention fabric (3mm)	33
BERG2	1.6	4.9	16	30	Pumice (50mm)	Textile retention fabric (3mm)	83
OSL1	2	4	5.5	30	-	Drainage layer (HDPE)6 (25mm)	55
OSL2	2	4	5.5	30	Separate Substrate3 (50mm)	Drainage layer (HDPE)6 (40mm) + Textile retention fabric (5mm)	125
OSL3	2	4	5.5	30	-	Textile retention fabric (10mm)	40
SAN1	1.6	5.3	27	30	-	Textile retention fabric (10mm)	40
SAN2	1.6	5.3	27	30	Separate Substrate3 (50mm)	Drainage layer (EPS)5 (75mm) + Textile retention fabric (5mm)	160
SAN3	1.6	5.3	27	30	-	Textile retention fabric (3mm)	33
SAN4	1.6	5.3	27	30	-	Substrate mat2 (50mm)	80
TRD1	2	7.5	16	30	-	Textile retention fabric (10mm)	40
TRD2	2	7.5	16	30	-	Substrate mat2 (50mm)	80
TRD3	2	7.5	16	30	Separate Substrate3 (50mm)	Drainage layer (HDPE)6 (25mm) + Textile retention fabric (5mm)	110
TRD4	2	7.5	16	30	Separate Substrate3 (50mm)	Drainage layer (PE)4 (30mm)	110

Pre-grown reinforced vegetation mats (sedum)

Substrate mat: a mineral wool plate

Separate Substrate: a mixture of Leca and bricks

Drainage layer (PE): plastic drainage layers of polyethylene

Drainage layer (EPS): plastic drainage layers of expanded polystyrene

Drainage layer (HDPE): plastic drainage layers of high-density polyethylene

2.1 Machine learning models

2.1.1 M5 model Tree

In this approach, the training data are divided into many subsets. For each subset, a piece-wise linear regression equation is built between the output and the input variables (Solomatine and Dulal, 2003). The algorithm used by the model tree is called

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

M5, which was developed in 1992 (Quinlan et al., 1992). It divides the data into subsets based on rules that reduce the intra-variation (variance) within each subset (variables within each subset are as similar as possible). The M5 model tree has an upside-down tree structure. Input variables enter the tree from the top (the tree root) to arrive at the models located at the tree leaves. For a detailed explanation of the M5 model tree, see Solomatine and Dulal (2003).

2.1.2 Artificial Neural Network (ANN)

The ANN applied in this study is the standard feed-forward neural network. It comprises an input layer, a hidden layer(s) and an output layer. The building block of the network is called a neuron, and each neuron is fully connected with all other neurons in the backward and forward layers. Hidden layers are where relations between input variables are revealed. Each neuron in the ANN applies simple mathematical operations for the variable vectors, as represented in equation 1:

$$O = f(X1 \times W1 + X2 \times W2 + B) \quad (1)$$

O is the output from a neuron, $W1$ and $W2$ are the weights of the variables $X1$ and $X2$, respectively, and B is the neuron's bias. $f(.)$ is the neuron's activation function that adds non-linearity to the neuron's output. During the training process, the weights and biases are updated for the whole network to obtain the best fit between simulated and observed outputs. A standard algorithm used for the training is backpropagation, which uses the approach of the steepest gradient descent (Rumelhart et al., 1986). Training of the neural network is done by dividing the training data set into several batches. The weights and biases are updated for each batch until all training data have been visited, and then the same cycle is repeated. This cycle is called an epoch, and the learning performance improves with the increasing number of epochs. However, there is a risk of overfitting for models with high numbers of epochs. To avoid that, a separate data set (validation data set) is often used to optimize the number of epochs by determining the objective function for the validation data while training the model. Overfitting starts when the error increases in the validation data set while decreasing in the training data. Parameters related to the structure of the network (i.e. the number of hidden layers and number of neurons at each hidden layer), the optimizer (learning rate, momentum and objective function) and the type of activation functions at each layer are called the hyperparameters which usually optimized by trial and error.

2.1.3 Long Short-Term Memory (LSTM)

In hydrology, sequential runoff data are often autocorrelated, especially data with a short time step. Autocorrelation is triggered by system memory in hydrology, usually due to the storage effects. A Recurrent Neural Network (RNN) is a special type of neural networks that can tackle sequential data modelling because it includes output from the previous time step as input to the following time step. Nevertheless, it doesn't account for the long-term dependency in the system. Hochreiter and Schmidhuber (1997) discussed the issue of RNN with long term dependency and proposed a unique RNN model called Long Short-Term Memory (LSTM). In this model, a value representing the system memory (S) is calculated and updated each time step to account for the long-term dependency of the system. LSTM cell comprises of three gates (Figure 2): forget gate (f), input gate (i) and output gate (o). The three gates control cell output and update its state for each time step by applying weights (W) and

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

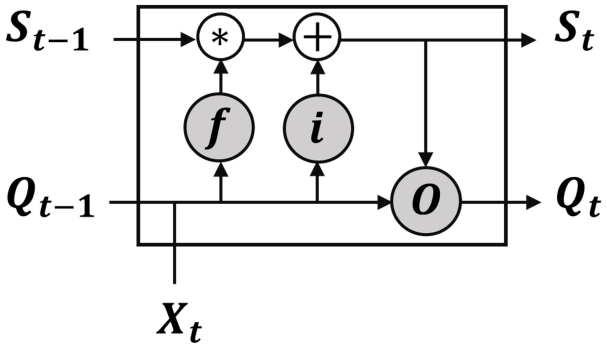


Figure 2. : Structure of the Long Short-Term Memory (LSTM) cell, modified from Kratzert et al. (2018)

biases (B). The first step is to control which information to be forgotten from the previous time step (f_t), which is done by the forget gate using equation 2. Secondly, the updated value for the cell state (ΔS_t) is determined from equation 3. Subsequently, the input gate output (i_t) is derived from equation 4, which controls how much information will be used from ΔS_t to update the cell state S_t. In the following step, the cell state S_t is determined by applying equation 5. Finally, the output from the output gate (O_t) is calculated from equation 6 which used to determine the cell output (Q_t) by using equation 7.

$$f_t = f_f(W_f \times X_t + U_f \times Q_{t-1} + B_f) \tag{2}$$

$$\Delta S_t = f_{\Delta S_t}(W_{\Delta S_t} \times X_t + U_{\Delta S_t} \times Q_{t-1} + B_{\Delta S_t}) \tag{3}$$

$$i_t = f_i(W_i \times X_t + U_i \times Q_{t-1} + B_i) \tag{4}$$

$$S_t = f_t \cdot S_{t-1} + i_t \cdot \Delta S_t \tag{5}$$

$$o_t = f_o(W_o \times X_t + U_o \times Q_{t-1} + B_o) \tag{6}$$

$$Q_t = \tanh S_t \cdot o_t \tag{7}$$

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

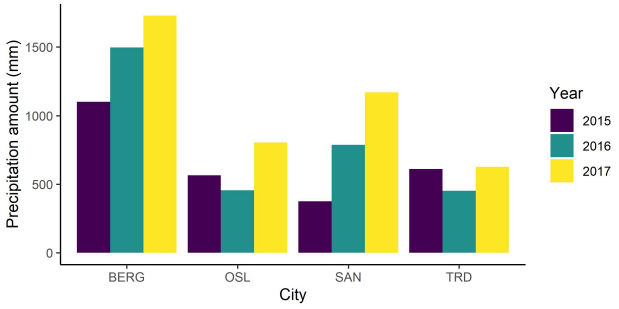


Figure 3. Sum of precipitation between April -September (mm)

2.1.4 k Nearest Neighbours (kNN)

k Nearest Neighbours is a nonparametric method that estimates the output of each time step based on its similarity to the historical time steps. Basically, the algorithm determines similarity distances (such as euclidian distances) between each of the input variables of the new time step to the variables from the training data set. Then it calculates the mean outputs of the k most similar time steps. k is an integer number that must be optimized by trial and error.

2.2 ML Modelling steps

A general equation was developed relating runoff to climatic variables as follow (equation 8):

$$R_t = f(P_t, P_{t-1}, P_{t-2}, \ldots, P_{t-lag}, Ta_t, Ta_{t-1}, Ta_{t-2}, \ldots, Ta_{t-lag}, W_t, W_{t-1}, W_{t-2}, \ldots, W_{t-lag}, Rh_t, Rh_{t-1}, Rh_{t-2}, \ldots, Rh_{t-lag}) \quad (8)$$

R is green roof runoff, P is precipitation, Ta is air temperature, W is wind speed, and Rh is relative humidity. This is a simplification as the physical properties of the green roof also affect its runoff. However, using data from the same green roofs in this study, Johannessen et al. (2018) found only a small variation in the hydrological performances between the different roof configurations and found the climatic variables to have high impacts on their performance. In the ML models in this study, climatic variables were lagged to represent the initial saturation of the green roofs at each time t. A trial and error method was applied to find the optimal value of the lag value for each city and each ML model. Data were aggregated into one-hour resolution, and snow accumulation periods were excluded (1 Oct. – 31 Mar.). One year was used for training and one year for validation. The selection of the training year was based on the sum of precipitation as the wettest year between 2015 to 2017 for each roof, and the second wettest year for validation. The rationale for the selection is that the wettest year covers a broader span of precipitation events which improves the generalization performance of the models. Figure 3 shows the cumulative precipitation between April-September for the four cities.

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

For fair comparisons between the methods, models were tuned to achieve good modelling performance and to avoid overfitting. Model tuning is the process of finding the optimal hyperparameters for the problem (e.g. number of hidden layers in ANN, number of rules in M5, k value in kNN, etc.) which are usually optimized by trial and error. BERG1, OSL1, SAN1 and TRD1 roofs were selected to test different hyperparameters to find the optimal parameters for each city. In this study, the tuning was done in two steps. In the first step, 120 sets of hyperparameters were selected by using Latin Hypercube sampling (McKaya et al., 1979) to identify the structural parameters that give good results, i.e. highest Nash Sutcliffe Efficiency (NSE) (equation 9) with low volumetric error measured by the percentage bias (PBIAS) (equation 10). Secondly, the structural parameters were fixed, and different lag values ranging from 1 hour to 200 hours were tested to identify the optimal lag value. To avoid overfitting, the performance of changing hyperparameters was observed in the validation periods.

$$NSE = 1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2} \tag{9}$$

$$PBIAS = 100 \times \frac{\sum Q_{obs} - \sum Q_{sim}}{\sum Q_{obs}} \tag{10}$$

2.3 The conceptual retention model

The sixteen roofs were modelled using a conceptual model, which was developed and validated by Stovin et al. (2013). The conceptual model is intended to provide a robust tool that estimates green roof retention using simple water balance equations (equations 11,12 and 13) and without the need for prior calibration.

$$R_t = \max(0, P_t - (S_{max} - S_t) - AET_t) \tag{11}$$

$$S_t = min(S_{t-1} + P_t - AET_t, 0)) \tag{12}$$

$$AET_t = PET \times \frac{S_{t-1}}{S_{max}} \tag{13}$$

R_t is the runoff from a green roof at time t, P_t is the precipitation at time t, S_{max} is the maximum storage available in a green roof and S_t is the water stored in a green roof at time t. In our study region, Johannessen et al. (2018) found the Oudin's model for ET to be the most accurate for their water balance model and Almorox et al. (2015) recommended the use of Oudin for cold climates. Hence, the potential evapotranspiration was computed using Oudin's model as follows (equations 14):

$$PET(\frac{mm}{day}) = \begin{cases} 0, & \text{if } Ta_{mean} \leq 5 \\ \frac{Ra}{\lambda\rho} \times 0.01 \times (Ta_{mean} + 5), & \text{if } Ta_{mean} > 5 \end{cases} \tag{14}$$

Ta_{mean} is the daily mean temperature, Ra is extra-terrestrial radiation derived from Julian day and latitude $(MJ.m^{-2})$, $\frac{1}{\lambda\rho} \approx 0.408$,$\lambda$ is the latent heat of water $(MJ.kg^{-1})$, ρ is the volumetric mass of water $(kg.m^{-3})$. S_{max} values were estimated

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

by assuming the field capacities of the roof layers from reported literature values as follow: vegetation mats were assumed to have 20% of the total substrate depth as a field capacity (Johannessen et al., 2018) , brick-based substrates were assumed to have 25 % of the total substrate depth as a field capacity (Stovin et al., 2013) while the drainage mats were assumed to have no permanent storage.

2.4 ML Model evaluation

Methods were evaluated based on the performance on the validation data sets. With respect to retention estimation, flow accumulation curves were plotted for the simulated runoff from ML models against the observed runoff and compared with the results from the conceptual retention model. In addition, the percentage bias ($PBIAS$) values were calculated for each simulation for comparison. To evaluate the performance of ML models in estimating the temporal variation in runoff, the simulated runoff from ML models were plotted against the observed values and the NSE values were determined. Values of $NSE > 0.5$ were considered satisfactory (Moriasi et al., 2007; Rosa et al., 2015). To evaluate the potential of using ML as a useful tool for planning and design purposes, ML models were transferred between the roofs unchanged. The transferred models simulated the validation periods of each roof, and NSE was used to evaluate the transferability performance. Moreover, a volumetric factor (vol) based on the PBIAS was determined by using equation 15 to assess transferability in terms of volume estimation. A vol value of 1 indicates a perfect runoff volume estimation and hence a perfect retention estimation, while A vol value of zero indicates 100% error in volume estimation.

$$vol = 1 - \frac{|PBIAS|}{100} \tag{15}$$

3 Results and Discussion

3.1 Hyperparameter optimization (model tuning)

Different ANN models were tested with a number of hidden layers ranging from 1 to 4 and a number of neurons from 1 to 100. The number of epochs was optimized by running an initial ANN model with 1000 epochs. After 20 epochs, the performance in the validation data didn't improve further and started to decrease after 100 epochs while improving in the training data set, which indicates overfitting. Therefore, 20 epochs were selected as an optimal value. Additional testing resulted in the selection of the ANN with 20 epochs and two hidden layers. The number of neurons varied from one city to another ranging from 40 at Bergen to 90 at Sandnes. LSTM with one hidden layer was selected for this study following the recommendation of the study of Ayzel (2019) in which, a grid search was performed for LSTM hyperparameters which compared thousands of LSTM structures. One hidden layer was determined to perform reasonably well with lower computational cost compared with multiple hidden layers. LSTM structure with 80 units was found to yield good results in the different cities in the current study. For the M5 model tree and the kNN, 20 rules and a k value of 3, respectively were found to fit well with the different locations.

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

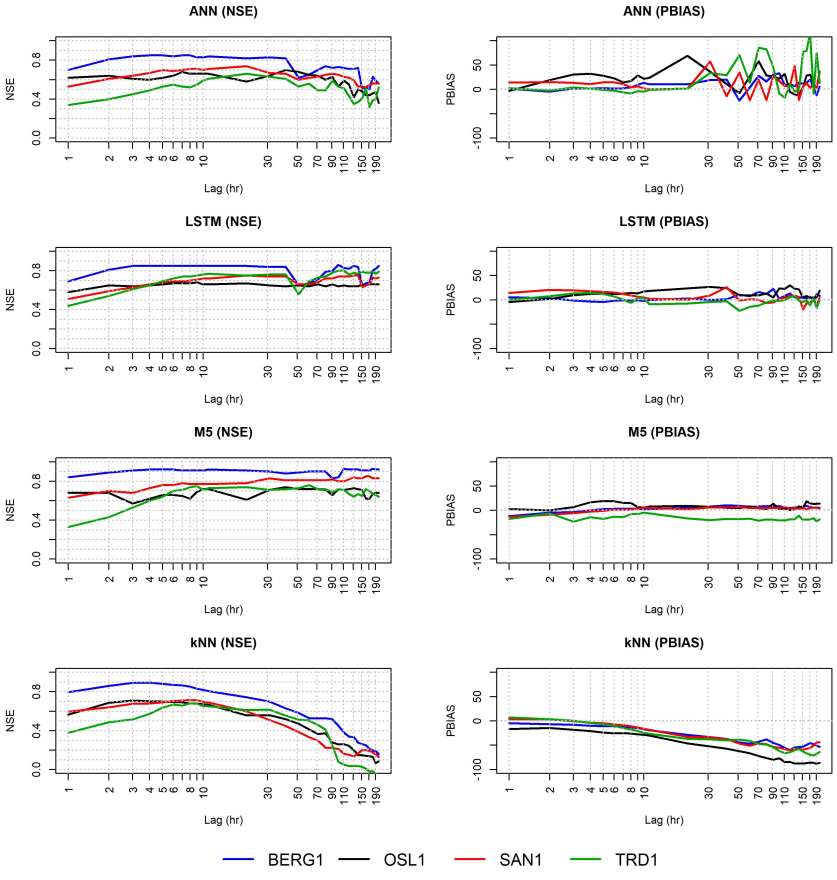


Figure 4. Effect of lagging inputs for ANN, LSTM and M5 on the simulation result (Validation periods)

The optimal structural parameters were fixed, and different lag values were tested. The importance of lagging climatic variables is demonstrated in Figure 4. The lag values were selected by considering both NSE and PBIAS metrics. Table 2 presents the final selection of the ML hyperparameters.

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

Table 2. Results of Hyperparameters tuning

Models	Hyperparameters	BERG	OSL	SAN	TRD
ANN	Number of hidden layers	2	2	2	2
	Number of Neurons	40	60	90	87
	Number of Epochs	20	20	20	20
	lag	4	51	11	21
LSTM	Number of hidden layers	1	1	1	1
	Number of units	80	80	80	80
	Number of Epochs	20	20	20	20
	lag	4	51	11	41
M5	Number of rules	20	20	20	20
	lag	4	41	20	11
kNN	k	3	3	3	3
	lag	4	3	6	6

The lag values varied between the cities and the methods. For LSTM, the lag value was smaller in BERG and SAN compared to TRD and OSL. To interpret this finding, rainfall events, with 6-hour intra-event periods, were extracted from the validation years at the four cities and compared, as shown in Figure 5. BERG roofs received events with higher maximum intensity and duration compared to OSL roofs, whereas the antecedent dry weather periods (ADWP) at OSL are longer than BERG. Hence, due to the longer ADWP in OSL beside the low-intensity rainfall, a longer memory of the system is needed to account for the wider range of possible initial saturations compared to BERG roofs.

3.2 Model evaluations

3.2.1 Retention estimation

Machine learning models were built for all roofs based on the optimized structural parameters presented in Table 2. Figure 6 illustrates the simulated and observed runoff cumulative curves together with the cumulative precipitation for each roof, and Table 3 shows the percentage bias of the models. The results presented in Figure 6 and Table 3 confirm that the ML models and the conceptual models can reproduce the observed runoff volume in most of the green roofs. By comparing the median $|PBIAS|$ values on the validation periods, M5 yielded only 5.65% with a maximum value of 14.2% and a minimum value of 0.2%. Following M5, median values of 6.45%, 6.85% and 6.9% were obtained by ANN, LSTM and the conceptual model, respectively. The kNN yielded the highest volumetric errors with a median $|PBIAS|$ of 11.1% and a maximum value of 22.1% which still classified as acceptable modelling results regarding volumetric error, as per Moriasi et al. (2007).

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

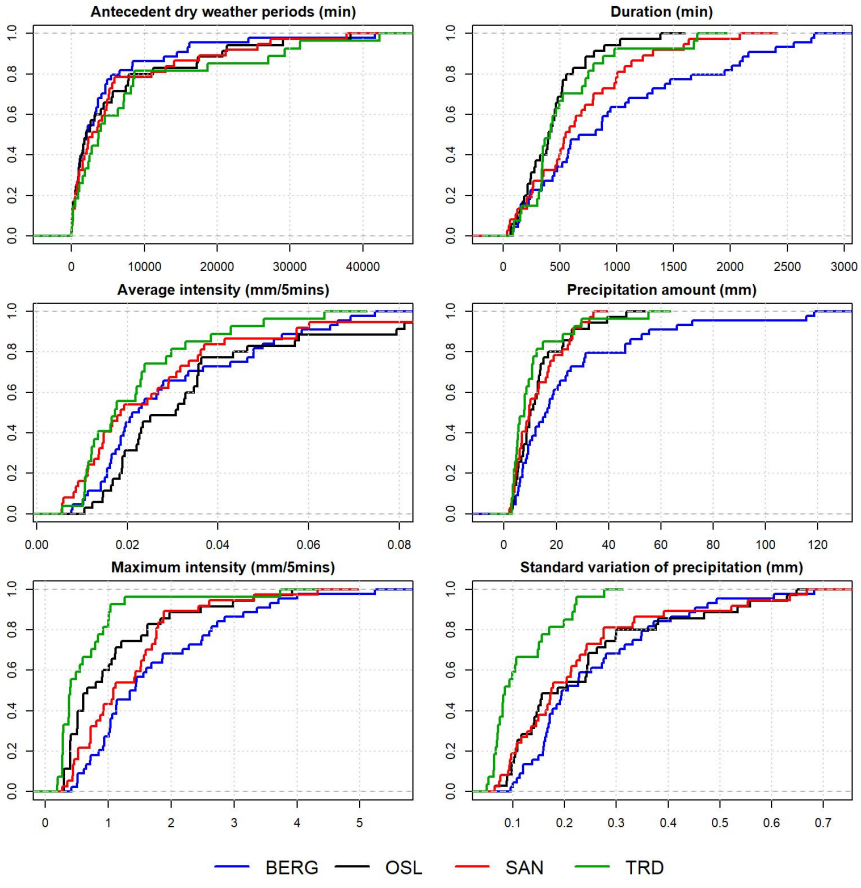


Figure 5. Comparison between the rainfall events during validation years at the four Norwegian cities

3.2.2 Temporal variations in runoff

Table 4 presents the NSE values for training and validation periods for the ML models. All ML models yielded satisfactory results in both training and validation periods ($NSE > 0.5$). Overfitting can be observed more in M5 and ANN simulations compared to LSTM, especially in TRD and OSL roofs, which explains the difference between the performance in the training and validation periods. Figure 7 shows the observed and simulated hydrographs for BERG1 roof, which confirms the ability of the ML models to reproduce the observed runoff. In contrast, the conceptual model produced satisfactory results in only three roofs. We found the green roofs in our study to detain small and medium sized events for up to two hours. The conceptual model failed to simulate these dynamics due to lack of routing.

The performance of ML models varied between the different cities more than between the different configurations. Johannessen et al. (2018), using the same data in this study, observed similar hydrological performance for the different configuration within the same city. It should be noted that, however, the geometries of the roofs are identical at each city (Table 1). The per-

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021


Hydrology and Earth System Sciences
Discussions

formance of the ML methods can be explained based on this comparison between the cities' rainfall characteristics (Figure 5). For instance, the NSE values of the ML models are higher in BERG roofs in comparisons to the other roofs in the study. As mentioned earlier, OSL roofs have a wider range of possible initial saturations. Therefore, one year of training data might not be enough to cover this wide range of runoff possibilities. On the other hand, BERG roofs received more frequent and intense precipitation events resulting in a small range of possibilities of initial saturation that could be covered using one year only. The kNN method produced lower NSE values compared to the other ML models. This was attributed to the relatively small training data used in this study as kNN estimates the performance depending on the similarity to the previous time steps.

Unlike other ML models, LSTM maintains a state value between consecutive time steps which makes it more suitable for modelling green roofs where initial saturation plays an important role in green roof runoff generation process. A comparison was made between ANN and LSTM at TRD1 (Figure 8) to demonstrate the potential of LSTM. ANN was found to produce runoff when no precipitation occurred (Figure 8-A), unlike LSTM. The advantage of maintaining cell state value is clear in Figure 8-B; although the magnitude of the first and third runoff peaks were almost equal, they resulted from different rainfall values due to the different initial saturations. ANN simulations failed to simulate the third peak because it has no memory about the past. Hence ANN did not account accurately for the initial saturation. In contrast, LSTM could simulate the third peak successfully. In Figure 8-C, one can see that the simulation of ANN was noisier (dropping suddenly) while LSTM simulation was smoother. Similarly, Kratzert et al. (2018) found LSTM simulations to be smoother than a normal recurrent neural network and to be better in accounting for the storage capacity (including snow accumulation) of a natural catchment

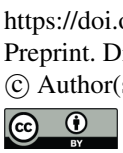
3.2.3 Transferability

Models were transferred between the roofs unchanged to simulate the validation periods. Figure 9 presents the transferability performance measured by NSE. Some models could yield satisfactory results in different locations ($NSE > 0.5$). For instance, M5 and ANN models that were trained using data from TRD4, SAN1, SAN3 and SAN4 roofs could yield satisfactory performance in almost all cities. Figure 10 presents the transferability performance with respect to retention estimation measured by vol. Transferred ML model could simulate result with acceptable accuracy ($vol > 0.75$) (Moriasi et al., 2007) between TRD and OSL cities and between SAN and BRRG cities with few exceptions. This can be attributed to the similarity in climatic conditions between the cities (Figure 5). However, the conceptual models in this study could produce better volume estimates than the transferred ML models in most cases. This implies that using the conceptual model is preferable over the transferred ML to estimate the annual retention for new roofs.

3.3 Machine learning potentials for green roof hydrological modelling

The present paper has demonstrated that well-trained ML models can be applied to estimate retention process (rainfall losses) in a range of different green roof systems. The predictions are comparable in accuracy to a conceptual water balance model based on losses due to evapotranspiration. Additionally, well-trained ML models showed more accurate predictions of runoff hydrographs than the conceptual water balance model which is encouraging for detention modelling. Detention modelling is required to estimate the lag and attenuation of runoff associated with any rainfall that is not retained. Whilst physically-

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021



Hydrology and Earth System Sciences
Discussions

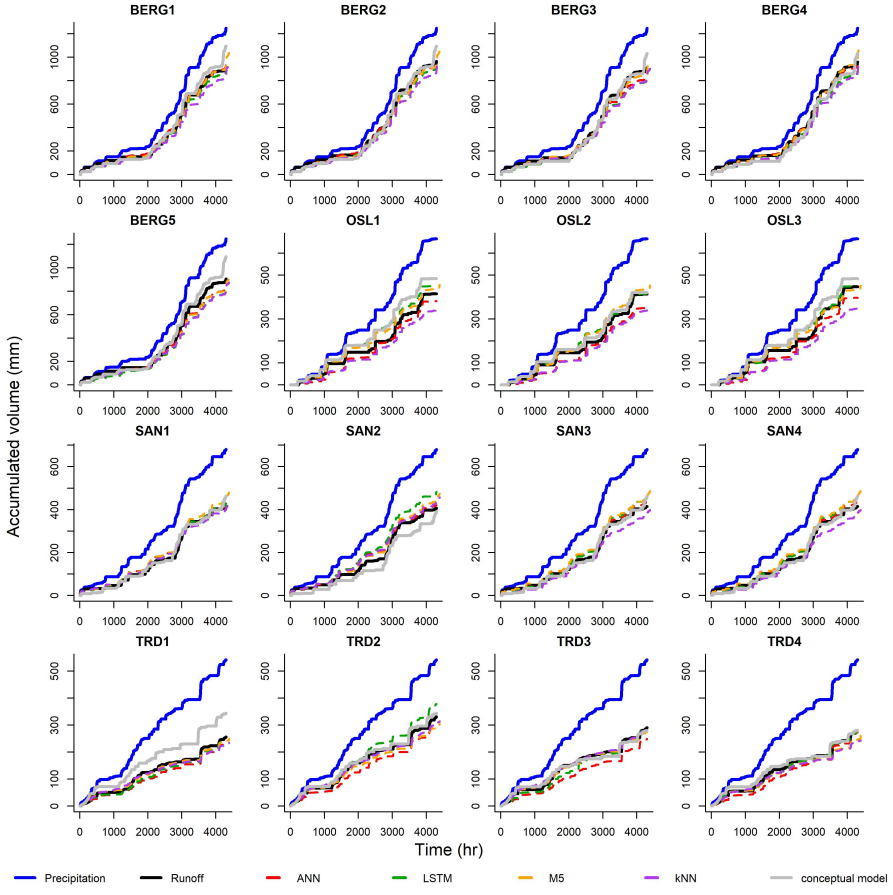


Figure 6. Cumulative precipitation, observed and simulated runoff of the roofs (validation years)

based detention models have been proposed (for example the use of 1D Richards equation-based models to represent vertical flow through the unsaturated growing media), these are computationally expensive, reliant on detailed characterisation of the physical media, and are generally not configured to accommodate complex, multi-layered systems. In practice, many modelling frameworks rely on calibrated reservoir routing models to estimate the cumulative detention effects of multiple interacting component layers, and few (if any) convincing validation cases for a complete detention modelling framework have been presented. It would therefore be very valuable to explore whether the ML models, when trained on higher temporal resolution datasets, have the capability to capture these complex detention effects better than the alternative black-box approaches.

4 Conclusions

Four machine learning models, commonly used in runoff modelling studies, were applied to simulate runoff from sixteen green roofs located in four Norwegian cities with different climatic conditions. We further investigated the potential of using

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

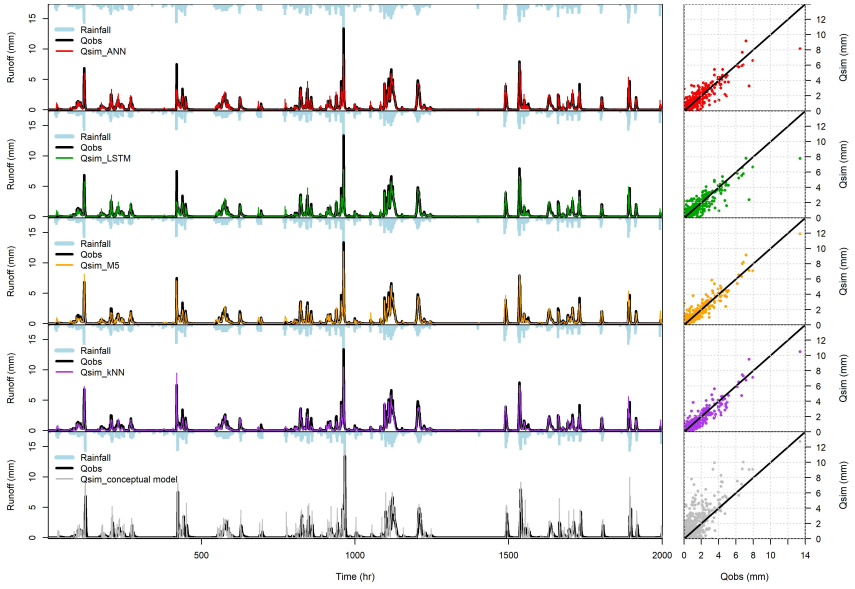


Figure 7. Performance of ML models on the validation period (BERG 1). The hydrographs were plotted for around three months period (2000 hours), while the Q-Q plots were plotted for the entire validation period

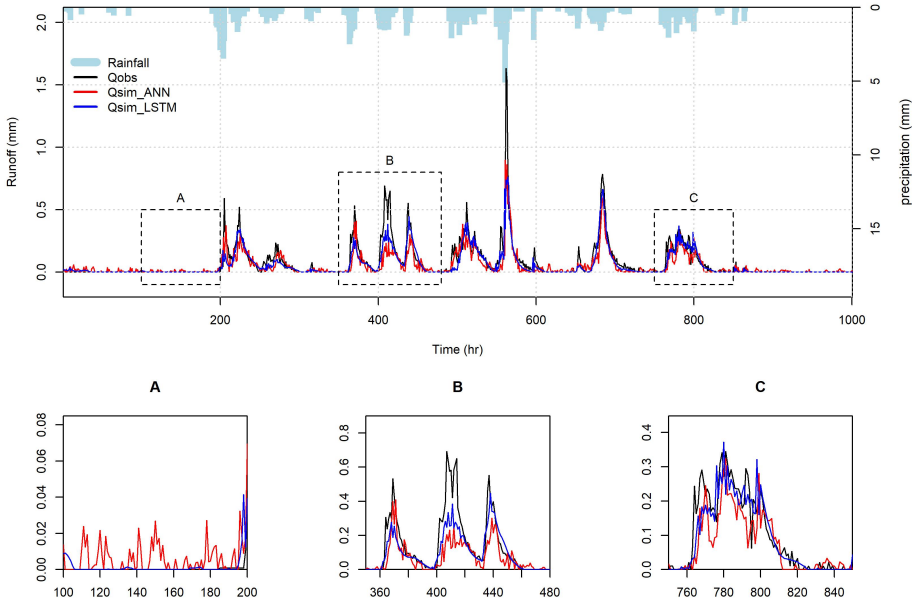


Figure 8. Comparison between LSTM and ANN at TRD1 roof

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

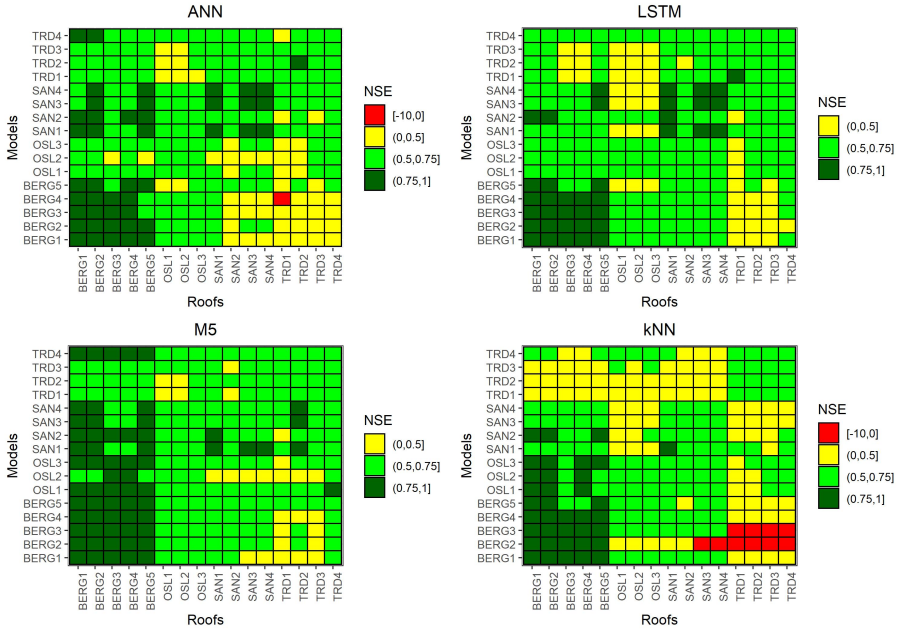


Figure 9. Transferability between the different roofs (NSE). Models in the y-axis are used to simulate the measured green roofs in the x-axis

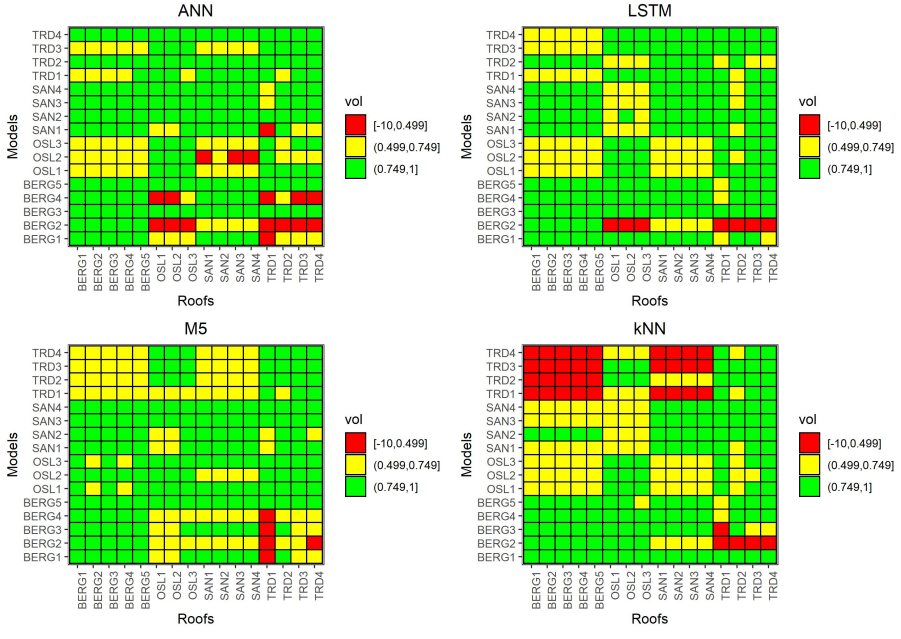


Figure 10. Transferability between the different roofs (vol). Models in the y-axis are used to simulate the measured green roofs in the x-axis

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

ML models to estimate performance of new roofs where runoff data are not available for model training. This was done by means of transferring ML models between the roofs in the study. Our results confirms the ability of well-trained ML models to estimate green roof retention and the temporal runoff dynamics.The estimates of the annual retention were comparable to a proven conceptual model. Despite the 1-hr time step, the ML models provided accurate simulations of runoff dunamics i.e discharge hydrographs (NSE values higher than 0.5 in both training and validation periods) which is encouraging for detention modelling. The LSTM demonstrated better modelling performance by maintaining a state value between consecutive time step, which makes it more appropriate for simulating runoff of green roofs. In future studies, shorter time-steps will be applied to estimate detention metrics.

Some transferred ML models could give acceptable model performance ($NSE > 0.5, |PBIAS| < 25\%$) in different locations. However, we recommend using the conceptual model with literature values of the S_{max} parameter to estimate the annual retention of new roofs over the transferred model as it give accurate volume estimations.

Table 3. Overall modelling performance (PBIAS)

GR	ANN		M5 model tree		kNN		LSTM		Conceptual model
	T	V	T	V	T	V	T	V	
BERG1	-3.2	1.7	-0.9	0.2	-	-9.8	-6.1	-3.3	9
BERG2	-6.7	-1.9	-3.1	-3.4	-	-10.8	-6.8	-4.3	3.8
BERG3	-10.3	-7.6	-2.7	-5.6	-	-12.4	-8.8	-10.3	3.1
BERG4	-3	-1.5	0.2	1.1	-	-11.9	-8.5	-8.2	-1.5
BERG5	-8.2	-7.9	-6.3	-11.4	-	-14.8	-7.9	-11.1	10.1
OSL1	-3.4	-7.9	-8.8	5.5	-	-18.3	1.9	9.4	15.5
OSL2	-5.4	-15.4	-3.9	5.7	-	-18.2	-6.5	0.6	0.5
OSL3	-1.7	-11.2	-6.1	-2.3	-	-22.1	-1.1	1.1	7.4
SAN1	-9.9	0.5	-0.4	5	-	-5.6	-8.7	0.2	6.7
SAN2	-9.9	3.5	-2.8	5.6	-	3.4	-0.9	14.8	-8.8
SAN3	2	5.3	-2	6.9	-	-11.1	5.7	7.7	7.1
SAN4	2	5.3	-2	6.9	-	-11.1	5.7	7.7	7.1
TRD1	-4.2	-10.1	-5.7	-6.1	-	-9.7	2	-6	38
TRD2	-9.3	-11	-7.1	-11	-	-6.4	12	14.3	6.6
TRD3	-21.7	-14.6	-11.2	-5.5	-	-1.9	-4.1	-3	-1.4
TRD4	-6.8	-2.3	-6.8	-11.6	-	-14.3	1.4	-1	2.5

T= Training, V= Validation

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

Table 4. Overall modelling performance (NSE)

GR	ANN		M5 model tree		kNN		LSTM		Conceptual model
	T	V	T	V	T	V	T	V	
BERG1	0.85	0.85	0.95	0.92	-	0.89	0.84	0.85	0.42
BERG2	0.77	0.84	0.86	0.92	-	0.89	0.76	0.84	0.39
BERG3	0.76	0.81	0.84	0.86	-	0.84	0.75	0.81	0.66
BERG4	0.87	0.86	0.95	0.92	-	0.89	0.86	0.85	0.68
BERG5	0.82	0.81	0.92	0.87	-	0.84	0.80	0.80	0.14
OSL1	0.96	0.68	0.94	0.74	-	0.71	0.83	0.65	0.36
OSL2	0.96	0.65	0.97	0.65	-	0.69	0.8	0.67	0.52
OSL3	0.95	0.67	0.96	0.70	-	0.71	0.82	0.67	0.32
SAN1	0.93	0.80	0.95	0.86	-	0.77	0.90	0.83	-0.38
SAN2	0.91	0.59	0.93	0.75	-	0.68	0.86	0.70	0.18
SAN3	0.88	0.76	0.94	0.62	-	0.67	0.86	0.79	-0.66
SAN4	0.88	0.76	0.94	0.62	-	0.67	0.86	0.79	-0.66
TRD1	0.91	0.74	0.9	0.73	-	0.67	0.84	0.77	-1.71
TRD2	0.91	0.76	0.86	0.69	-	0.62	0.85	0.75	-0.76
TRD3	0.87	0.66	0.84	0.69	-	0.67	0.75	0.70	-0.08
TRD4	0.91	0.65	0.81	0.74	-	0.68	0.77	0.72	0.38

T= Training, V= Validation

Author contributions. Elhadi was responsible of the machine learning models, including the selection of the four models, hyperparameter tuning and and models training. Virginia Stovin proposed the comparison with the conceptual retention model. Vincent Pons assisted Elhadi during ML modelling and the conceptual model application. Knut and Tone supervised each step of the research from the start. Elhadi wrote the first manuscript which was revised many times by all co-authors

Competing interests. The authors declare that they have no competing interests

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

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hydrology and Earth System Sciences
Discussions

References

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, Fao, Rome, 300, D05 109, 1998.

Almorox, J., Quej, V. H., and Martí, P.: Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes, Journal of Hydrology, 528, 514–522, 2015.

Ayzel, G.: Does deep learning advance hourly runoff predictions, in: Proceedings of the V International Conference Information Technologies and High-Performance Computing (ITHPC-2019), Khabarovsk, Russia, pp. 16–19, 2019.

Baek, S., Ligaray, M., Pachepsky, Y., Chun, J. A., Yoon, K.-S., Park, Y., and Cho, K. H.: Assessment of a green roof practice using the coupled SWMM and HYDRUS models, Journal of environmental management, 261, 109 920, 2020.

Bengtsson, L., Grahn, L., and Olsson, J.: Hydrological function of a thin extensive green roof in southern Sweden, Hydrology Research, 36, 259–268, 2005.

Berndtsson, J. C.: Green roof performance towards management of runoff water quantity and quality: A review, Ecological engineering, 36, 351–360, 2010.

Berretta, C., Poë, S., and Stovin, V.: Reprint of "Moisture content behaviour in extensive green roofs during dry periods: The influence of vegetation and substrate characteristics", Journal of Hydrology, 516, 37–49, 2014.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in water resources, 16, 41–51, 1993.

Bhattacharya, B. and Solomatine, D. P.: Neural networks and M5 model trees in modelling water level–discharge relationship, Neurocomputing, 63, 381–396, 2005.

Bouzouidja, R., Séré, G., Claverie, R., Ouvrard, S., Nuttens, L., and Lacroix, D.: Green roof aging: Quantifying the impact of substrate evolution on hydraulic performances at the lab-scale, Journal of Hydrology, 564, 416–423, 2018.

Brunetti, G., Šimnek, J., and Piro, P.: A comprehensive analysis of the variably saturated hydraulic behavior of a green roof in a mediterranean climate, Vadose Zone Journal, 15, 1–17, 2016.

Carson, T., Marasco, D., Culligan, P., and McGillis, W.: Hydrological performance of extensive green roofs in New York City: observations and multi-year modeling of three full-scale systems, Environmental Research Letters, 8, 024 036, 2013.

Cipolla, S. S., Maglionico, M., and Stojkov, I.: A long-term hydrological modelling of an extensive green roof by means of SWMM, Ecological Engineering, 95, 876–887, 2016.

Daniell, T.: Neural networks. Applications in hydrology and water resources engineering, in: National Conference Publication- Institute of Engineers. Australia, 1991.

Dunnett, N. and Kingsbury, N.: Planting Green Roofs and Living Walls, 2004.

Fassman-Beck, E., Voyde, E., Simcock, R., and Hong, Y. S.: 4 Living roofs in 3 locations: Does configuration affect runoff mitigation?, Journal of Hydrology, 490, 11–20, 2013.

Gharaei-Manesh, S., Fathzadeh, A., and Taghizadeh-Mehrjardi, R.: Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran, Cold Regions Science and Technology, 122, 26–35, 2016.

Goyal, M. K., Ojha, C., Singh, R., Swamee, P., Nema, R., et al.: Application of ANN, fuzzy logic and decision tree algorithms for the development of reservoir operating rules, Water resources management, 27, 911–925, 2013a.

Goyal, M. K., Ojha, C., Singh, R., Swamee, P., et al.: Application of artificial neural network, fuzzy logic and decision tree algorithms for modelling of streamflow at Kasol in India, Water science and technology, 68, 2521–2526, 2013b.

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Hamouz, V. and Muthanna, T. M.: Hydrological modelling of green and grey roofs in cold climate with the SWMM model, Journal of environmental management, 249, 109 350, 2019.

Hernes, R. R., Gragne, A. S., Abdalla, E. M., Braskerud, B. C., Alfredsen, K., and Muthanna, T. M.: 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, 1437–1454, 2020.

Hilten, R. N., Lawrence, T. M., and Tollner, E. W.: Modeling stormwater runoff from green roofs with HYDRUS-1D, Journal of hydrology, 358, 288–293, 2008.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735–1780, 1997.

Hsu, K.-l., Gupta, H. V., and Sorooshian, S.: Artificial neural network modeling of the rainfall-runoff process, Water resources research, 31, 2517–2530, 1995.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z.: Deep learning with a long short-term memory networks approach for rainfall-runoff simulation, Water, 10, 1543, 2018.

Jahanfar, A., Drake, J., Sleep, B., and Gharabaghi, B.: A modified FAO evapotranspiration model for refined water budget analysis for Green Roof systems, Ecological engineering, 119, 45–53, 2018.

Javan, K., Lialestani, M. R. F. H., and Nejadhossein, M.: A comparison of ANN and HSPF models for runoff simulation in Gharehsoo River watershed, Iran, Modeling Earth Systems and Environment, 1, 1–13, 2015.

Johannessen, B. G., Hanslin, H. M., and Muthanna, T. M.: Green roof performance potential in cold and wet regions, Ecological Engineering, 106, 436–447, 2017.

Johannessen, B. G., Muthanna, T. M., and Braskerud, B. C.: Detention and retention behavior of four extensive green roofs in three nordic climate zones, Water, 10, 671, 2018.

Johannessen, B. G., Hamouz, V., Gragne, A. S., and Muthanna, T. M.: The transferability of SWMM model parameters between green roofs with similar build-up, Journal of Hydrology, 569, 816–828, 2019.

Karlsson, M. and Yakowitz, S.: Nearest-neighbor methods for nonparametric rainfall-runoff forecasting, Water Resources Research, 23, 1300–1308, 1987.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen-Geiger climate classification updated, 2006.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using long short-term memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, 2018.

Krebs, G., Kuoppamäki, K., Kokkonen, T., and Koivusalo, H.: Simulation of green roof test bed runoff, Hydrological processes, 30, 250–262, 2016.

Li, S., Kazemi, H., and Rockaway, T. D.: Performance assessment of stormwater GI practices using artificial neural networks, Science of the total environment, 651, 2811–2819, 2019.

Li, Y. and Babcock Jr, R. W.: Modeling hydrologic performance of a green roof system with HYDRUS-2D, Journal of Environmental Engineering, 141, 04015 036, 2015.

Liu, R. and Fassman-Beck, E.: Hydrologic response of engineered media in living roofs and bioretention to large rainfalls: experiments and modeling, Hydrological Processes, 31, 556–572, 2017.

Liu, R. and Fassman-Beck, E.: Pore structure and unsaturated hydraulic conductivity of engineered media for living roofs and bioretention based on water retention data, Journal of Hydrologic Engineering, 23, 04017 065, 2018.

Liu, X. and Chui, T. F. M.: Evaluation of green roof performance in mitigating the impact of extreme storms, Water, 11, 815, 2019.

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

McKaya, M., Beckmana, R., and Conoverb, W.: Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979.

Modaresi, F., Araghinejad, S., and Ebrahimi, K.: A comparative assessment of artificial neural network, generalized regression neural network, least-square support vector regression, and K-nearest neighbor regression for monthly streamflow forecasting in linear and nonlinear conditions, Water Resources Management, 32, 243–258, 2018.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885–900, 2007.

Palla, A., Gnecco, I., and Lanza, L. G.: Unsaturated 2D modelling of subsurface water flow in the coarse-grained porous matrix of a green roof, Journal of Hydrology, 379, 193–204, 2009.

Palla, A., Gnecco, I., and Lanza, L.: Compared performance of a conceptual and a mechanistic hydrologic models of a green roof, Hydrological Processes, 26, 73–84, 2012.

Peng, Z. and Stovin, V.: Independent validation of the SWMM green roof module, Journal of Hydrologic Engineering, 22, 04017 037, 2017.

Peng, Z., Smith, C., and Stovin, V.: Internal fluctuations in green roof substrate moisture content during storm events: Monitored data and model simulations, Journal of Hydrology, 573, 872–884, 2019.

Quinlan, J. R. et al.: Learning with continuous classes, in: 5th Australian joint conference on artificial intelligence, vol. 92, pp. 343–348, World Scientific, 1992.

Radfar, A. and Rockaway, T. D.: Captured runoff prediction model by permeable pavements using artificial neural networks, Journal of Infrastructure Systems, 22, 04016 007, 2016.

Rezaei, F., Jarrett, A., Berghage, R., and Beattie, D.: Evapotranspiration rates from extensive green roof plant species, in: 2005 ASAE Annual Meeting, p. 1, American Society of Agricultural and Biological Engineers, 2005.

Rosa, D. J., Clausen, J. C., and Dietz, M. E.: Calibration and verification of SWMM for low impact development, JAWRA Journal of the American Water Resources Association, 51, 746–757, 2015.

Rossman, L. A. et al.: Storm water management model user's manual, version 5.0, National Risk Management Research Laboratory, Office of Research and . . . , 2010.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, nature, 323, 533–536, 1986.

Russwurm, I. L., Johannessen, B. G., Gragne, A. S., Lohne, J., Muthanna, T. M., and Fridthjofsdottir, E. G.: Modelling green roof detention performance in cold climates, 2018.

Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources scientists, Water Resources Research, 54, 8558–8593, 2018.

Sherrard Jr, J. A. and Jacobs, J. M.: Vegetated roof water-balance model: experimental and model results, Journal of Hydrologic Engineering, 17, 858–868, 2012.

Sims, A. W., Robinson, C. E., Smart, C. C., and O'Carroll, D. M.: Mechanisms controlling green roof peak flow rate attenuation, Journal of Hydrology, 577, 123 972, 2019.

Solomatine, D. P. and Dulal, K. N.: Model trees as an alternative to neural networks in rainfall—runoff modelling, Hydrological Sciences Journal, 48, 399–411, 2003.

Soulis, K. X., Valiantzas, J. D., Ntoulas, N., Kargas, G., and Nektarios, P. A.: 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, 200, 434–445, 2017.

https://doi.org/10.5194/hess-2021-124
Preprint. Discussion started: 10 March 2021

Stovin, V.: The potential of green roofs to manage urban stormwater, Water and Environment Journal, 24, 192–199, 2010.

Stovin, V., Poë, S., and Berretta, C.: A modelling study of long term green roof retention performance, Journal of environmental management, 131, 206–215, 2013.

Tokar, A. S. and Johnson, P. A.: Rainfall-runoff modeling using artificial neural networks, Journal of Hydrologic Engineering, 4, 232–239, 1999.

Tsang, S. and Jim, C. Y.: Applying artificial intelligence modeling to optimize green roof irrigation, Energy and Buildings, 127, 360–369, 2016.

Vesuviano, G. and Stovin, V.: A generic hydrological model for a green roof drainage layer, Water Science and Technology, 68, 769–775, 2013.

Vesuviano, G., Sonnenwald, F., and Stovin, V.: A two-stage storage routing model for green roof runoff detention, Water Science and Technology, 69, 1191–1197, 2014.

Wu, C., Chau, K. W., and Li, Y. S.: Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques, Water Resources Research, 45, 2009.

Yilmaz, A. G. and Muttil, N.: Runoff estimation by machine learning methods and application to the Euphrates Basin in Turkey, Journal of Hydrologic Engineering, 19, 1015–1025, 2014.

Yio, M. H., Stovin, V., Werdin, J., and Vesuviano, G.: Experimental analysis of green roof substrate detention characteristics, Water Science and Technology, 68, 1477–1486, 2013.

Young, C.-C., Liu, W.-C., and Wu, M.-C.: A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events, Applied Soft Computing, 53, 205–216, 2017.