From the Department of Landscape Ecology and Resources Management
Professorship of Landscape, Water and Biogeochemical Cycles of the
Justus Liebig University Giessen

Application of machine learning for the prediction of stable isotopes of water concentrations in streams and groundwater

Dissertation
for the degree Doctor of Natural Science (Dr. rer. nat.)
submitted by Amirhossein Sahraei, M. Sc.

Referees from the Justus Liebig University Giessen:
Prof. Dr. Lutz Breuer
Prof. Dr. Jan Siemens

Submitted: 03.09.2021

Abstract

This dissertation introduces the application of machine learning to isotope hydrology. The recent development of laser spectroscopy has made it feasible to measure stable isotopes of water in high temporal resolutions up to sub-hourly scales. High-resolution data provides the opportunity to identify the fine-scale, short-term transport and mixing processes that are not detectable at coarser resolutions. Despite such advantages, routine and long-term sampling of streams and groundwater sources at high temporal resolution is still far from being widespread. Novel approaches that are able to predict and interpolate infrequently measured data at multiple sources would be a major breakthrough. This dissertation focuses on the application of machine learning and hyperparameter optimization to efficiently predict high-resolution isotope concentrations of multiple stream and groundwater sources in the Schwingbach Environmental Observatory (SEO), Germany. In a first step, an automated mobile laboratory was utilized to automatically sample and analyse stable isotopes and water quality for multiple water sources at 20 min intervals in situ. Prompt responses of isotope concentrations to precipitation revealed that shallow subsurface flow pathways rapidly delivered water to the stream. A Spearman rank analysis indicated that precipitation is a main driver of event and pre-event water contribution. In a second step, Artificial Neural Network (ANN) and Support Vector Machine (SVM) were optimized to predict maximum event water fractions in streamflow on independent precipitation events using only precipitation, soil moisture and air temperature as input features. The optimized SVM outperformed that of ANN with an RMSE of 9.43%, MAE of 7.89%, R^2 of 0.83, and NSE of 0.78. A systematic hyperparameter optimization approach showed that an adequate number of hidden nodes and a suitable activation function enhanced the ANN performance, whereas the performance of SVM was directly related to the selection of the kernel function. Finally, a Long Short-Term Memory (LSTM) deep learning model was optimized using a Bayesian optimization algorithm to predict high-resolution time series of isotope concentrations in multiple stream and groundwater sources using a set of explanatory data that are more straightforward and less expensive to measure compared to the stable isotopes. The LSTM could successfully predict isotope concentrations of stream and groundwater sources using only short-term sequence (6 h) of measured water temperature, pH and electrical conductivity with an RMSE of 0.7‰, MAE of 0.4‰, R^2 of 0.90, and NSE of 0.70. In conclusion, machine learning methods are promising tools for the prediction of variables that are difficult, expensive or cumbersome to measure.

Table of Contents

1 Extended Summary

1.1 Introduction

It is challenging to comprehend how a catchment responds to precipitation events owing to its complex underlying hydrological processes (Uhlenbrook et al., 2002; Zhou et al., 2021). A comprehensive understanding of runoff generation mechanisms and hydrological response characteristics of a catchment is crucial for the sustainability, protection and management of its water resources (Hrachowitz et al., 2011). Stable isotopes of water (δ^2H and $\delta^{18}O$) are commonly used as conservative natural tracers in catchment hydrology to provide insight into hydrological processes. They are ideal tracers of water sources because they are naturally occurring constituents of water, they maintain hydrological information and they are able to characterize catchment-scale hydrological responses (Stadnyk et al., 2013). Such tracers have proved to be useful tools to study the source and formation of recharged water, the interactions between surface and groundwater, or the partitioning of evaporation, transpiration and mixing processes between different water sources (Kendall and McDonnell, 2012; Orlowski et al., 2016). Specifically at the catchment scale, stable isotopes of water have been widely employed to separate runoff components via hydrograph separation techniques (Klaus and McDonnell, 2013), to estimate mean transit times (McGuire and McDonnell, 2006), to explore flow pathways (Tetzlaff et al., 2015), to calculate groundwater recharge rates (Koeniger et al., 2016), to comprehend soil water mixing processes (Sprenger et al., 2016) and to enhance hydrological model simulations (Windhorst et al., 2014).

Isotope studies typically involve manual or automated sample collection followed by transport, storage and subsequent laboratory analysis. The time and effort involved in sample handling are often major constraints limiting the frequency and duration of sampling and thus scope of studies. Most of previous isotope studies dealing with the investigation of hydrological processes and catchment responses have collected daily to monthly data, at which the higher resolution behaviour is unobservable. As pointed out by Kirchner et al. (2004), sampling at intervals much longer than the hydrological response times of a catchment might lead to a significant loss of information. Particularly at the event scale, weekly or even daily data are often not able to accurately partition total runoff into event water (i.e., the new water from incoming precipitation) and pre-event water (i.e., the old water stored in the catchment) components through isotopic hydrograph separation

technique owing to the fact that the stable isotopes of water are highly variable during precipitation events. A few studies manually collected sub-daily water isotopes for subsequent laboratory analysis (McGlynn et al., 2004; Wissmeier and Uhlenbrook, 2007; James and Roulet, 2009; Fischer et al., 2017b). However, these studies tended to focus on a few precipitation events since manually sampling of sub-daily water isotopes is laborious and cost intensive. Although such studies are useful, they are still limited in giving insight into short-term response variability and mixing processes over a longer-term catchment behaviour and might fail in capturing of relevant events due to their erratic and difficult-to-plan nature.

The recent development of compact and robust isotope analysers has significantly reduced the analytical costs and made it conceivable to automatically measure stable isotopes of water in high temporal resolution (i.e., sub-hourly) directly in the field. To date, only a few studies have implemented field-based automated systems for continuous high-resolution monitoring of water isotopes (Heinz et al., 2014; von Freyberg et al., 2017; Mahindawansha et al., 2018; Quade et al., 2019; Sahraei et al., 2020). The result of these studies emphasized that the high-resolution data provide an opportunity to identify fine-scale, short-term transport and mixing processes that might be masked at coarser resolutions. However, despite its advantages, routine monitoring of water isotopes with high-resolution is still far from being common. Long-term, high-resolution sampling of multiple sources such as precipitation, stream and groundwater is even less common, although such measurements are likely to shed new light on hydrological processes and may help to delineate particular flow pathways that contribute to runoff generation. Thus, a proper predictive model that is able to interpolate infrequently measured data at multiple sampling sites with a set of explanatory data that are more straightforward and less expensive to measure compared to the stable isotopes of water would be a major breakthrough in the field of isotope hydrology. Here, machine learning could be an effective tool to fulfil such a purpose.

Machine learning is a data-driven Artificial Intelligence (AI) approach that aims to give computers the ability to automatically learn and extract patterns from data (Samuel, 1959; Goodfellow et al., 2016). In the last twenty years, machine learning models have been extensively used to efficiently simulate non-linear systems and capture noise complexity in corresponding data. A key advantage that has raised the application of machine learning models is their ability to represent highly non-linear and complex systems through simple-to-measure explanatory features and without any a priori assumption about the data structure and knowledge of the underlying physical processes

(Liu and Lu, 2014). In recent years, machine learning has been increasingly applied to predict streamflow, groundwater and lake water level, water quality parameters such as phosphorus, nitrogen and dissolved oxygen, soil moisture, soil hydraulic conductivity, water temperature and many other hydrological variables (Lange and Sippel, 2020; Zounemat-Kermani et al., 2020).

Artificial Neural Network (ANN) is an immensely popular AI algorithm for the prediction of hydrological variables (Maier and Dandy, 2000; Maier et al., 2010). ANN is a parallel-distributed information processing system, which has been essentially inspired by the biological neural network (Haykin, 1994). One of the key advantages of ANNs is that they are universal function approximators (Khashei and Bijari, 2010). This means that they are able to automatically approximate a large class of functions with a high degree of accuracy. Moreover, ANN is developed through learning and not through programming meaning that no preliminary specification of fitting function is required in the model development process. Instead, the network function is derived from the patterns and characteristics of the data through the learning process (Khashei and Bijari, 2010). Another key advantage of ANNs is that they have an intrinsic generalization capability. This means that they can recognize and respond to patterns that are analogous, but not identical to the patterns, with which they have been trained (Benardos and Vosniakos, 2007). With increasing research on ANN models, the limitations of ANN have been highlighted, such as local optimal solutions and gradient disappearance, which constrain the application of the model (Yang et al., 2017; Zhang et al., 2018a). As an alternative, the Support Vector Machine (SVM) algorithm was introduced by Cortes and Vapnik (1995) offering a new structure in the field of machine learning. The major advantage of SVM is that it is based on the structural risk minimization principle, which aims to minimize an upper bound to the generalization error instead of traditional local training error, from which SVM is able to achieve good generalization ability (Liu and Lu, 2014). Furthermore, for the SVM implementation, a convex quadratic constrained optimization problem must be solved so that its solution is always unique and globally optimal (Schölkopf et al., 2002). Another major advantage of SVM is that it automatically identifies and incorporates support vectors during the training process and prevents the impact of the non-support vectors on the model. This allows the model to cope well with noisy conditions (Han et al., 2007). Reviews of numerous studies have shown strong promise of success for ANN and SVM models in various areas of hydrology and earth system science (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000; Maier and Dandy, 2000; Dawson and Wilby, 2001; Maier et al.,

2010; Raghavendra and Deka, 2014). However a drawback of ANN and SVM models, which have mainly been used for time series analysis in the past, is that any information on the sequential order of the inputs is lost (Kratzert et al., 2018). Therefore, for the problems that the sequential order of the inputs matters, more advanced machine learning models are required to efficiently handle the temporal dependencies.

Deep learning is an advanced subfield of machine learning that has experienced an explosive growth in recent years due to improvements in computer technology, especially through Graphic Processing Units (GPU), and the availability of large data sets (Schmidhuber, 2015). Deep learning generally refers to large, multilayer neural networks that can extract high-level representations from complex and high-dimensional data through a hierarchical learning process using multiple non-linear transformations (Shen, 2018; Zuo et al., 2019). Deep learning is capable of digesting large quantities of data and is able to successfully generalize to new instances (Shen, 2018). Furthermore, it can extract substantial features without being explicitly instructed, and hence, is immune to raw and noisy data (Sahiner et al., 2019). Long Short-Term Memory (LSTM) is the state-of-the-art deep learning architecture that is widely employed to simulate sequential data like time series (Gers et al., 2002). LSTM is a type of Recurrent Neural Network (RNN) that can learn when to forget and how long to retain the state information (Hochreiter and Schmidhuber, 1997). Unlike the traditional RNNs, LSTM does not suffer from exploding and vanishing gradients, which allows the network to learn long-term dependencies (Hochreiter and Schmidhuber, 1997). This is desirable for modelling catchment processes like storage effects that have relatively long timescales compared to input-driven processes like direct surface runoff, and may play an important role in hydrological processes (Kratzert et al., 2018). So far, only a few attempts have been made to apply LSTM to predict time series of hydrological variables such as soil moisture (Fang et al., 2017), streamflow (Kratzert et al., 2018), water table depth (Zhang et al., 2018b), dissolved oxygen, turbidity and chemical oxygen demand (Liu et al., 2019a). The potential of LSTM has not yet been explored in the field of isotope hydrology; and in general, the application of machine learning in this field has rarely been investigated (Cerar et al., 2018).

1.2 General Objectives of the Dissertation

The main aim of this dissertation is *to predict high temporal resolution of the stable isotope concentrations in multiple stream and groundwater sources using machine learning.* For this investigation, the dissertation is subdivided in three objectives, which will be addressed in separate chapters:

1) Investigation of the response of runoff components to precipitation and the role of underlying controlling factors on the runoff generation process.

To accomplish the first objective, an automated mobile laboratory, the Water Analysis Trailer for the Environmental Research (WATER), is utilized to automatically sample multiple water sources and analyse the stable isotopes of precipitation, stream and groundwater (δ^2H and $\delta^{18}O$) and their water quality (water temperature, pH and electrical conductivity) in high temporal resolution (20-min intervals) directly in the field (Chapter 2).

2) Prediction of the response of runoff components to precipitation using machine learning methods.

To fulfil the second objective, two widely used machine learning algorithms, namely ANN and SVM, are optimized to predict maximum event water fractions in streamflow for independent precipitation events at times when no direct measurements of stable isotopes of water are available. Furthermore, the influence of hyperparameter configurations on the model performance is investigated (Chapter 3).

3) Simulation of high temporal resolution of stable isotope concentrations in multiple water sources using deep leaning.

To achieve the third objective, a Long Short-Term Memory (LSTM) deep learning model is optimized through a Bayesian optimization approach to predict high-resolution stable isotope concentrations of multiple stream and groundwater sources with a set of explanatory data that are more straightforward and less expensive to measure compared to the stable isotopes of water. A sensitivity analysis is further conducted to examine how different input features and sequence lengths affect the performance of the model (Chapter 4).

The investigation of the objectives for this dissertation requires data of high-resolution, multiple sources and a large number of variables. Particularly, the availability of high-resolution data of stable isotopes of various water sources is a limiting factor. Worldwide, such information is only available for a very few sites. One such catchment is the Schwingbach headwater, for which the analyses of dissertation were carried out.

1.3 Materials and Methods

1.3.1 Study area and monitoring network

The study area is located in the headwater catchment of the Schwingbach Environmental Observatory (SEO) (50°28′40″N, 8°32′40″E) in Hesse, Germany (Figure 1.1a-e). The research area comprises a total area of 1.03 km^2 with elevations ranging from 310 m in the north to 415 m a.s.l. in the south (Figure 1.1b). The climate is classified as temperate oceanic (Köppen climate classification), with a mean annual air temperature of 9.6°C and mean annual precipitation of 623 mm (1969-2019) recorded at the near-by weather station Giessen-Wettenberg operated by the German Weather Service. The catchment area is covered by 76% of forest that is mainly located in the east and south, 15% of arable land in the north and west and 7% of meadows along the perennial Schwingbach stream (Figure 1.1c). The soils are classified as Cambisols mainly covered by forests and Stagnosols under arable land. Under the forest, soil's texture is dominated by silt, fine sand and gravel from 0 to 3 m depths, with a low clay content between 0.6 and 3 m depths. Soils lay on slightly weathered clay shale bedrock. Under meadows and arable land, the soil texture is rich in silt and fine sand from 0 to 5 m depths with small amounts of clay and gravel between 0.8 and 5 m depths. Here, soils developed on greywacke and weathered clay shale.

An automatic climate station (AQ5, Campbell Scientific Inc., Shepshed, UK) operated with a CR1000 data logger measured precipitation depth, air temperature at 2 m height, relative humidity, air pressure, solar radiation and wind speed at 5 min intervals (Figure 1.2a). Six remote telemetry loggers (A753, Adcon, Klosterneuburg, Austria), three at the toeslope (SM1, SM2 and SM3) and three at the footslope (SM4, SM5 and SM6), were equipped with capacitance sensors (ECH2O 5TE, METER Environment, Pullman, USA) to automatically measure soil moisture at 5, 15, 30, 50 and 70 cm depths at 5 min intervals (Figure 1.2b-c). A stream gauge (RBC flume, Eijkelkamp Agrisearch Equipment, Giesbeek, Netherlands) equipped with a pressure transducer (Mini-Diver,

Eigenbrodt Inc., Königsmoor, Germany) automatically recorded water levels at 10 min intervals at the outlet (SW2) of the catchment (Figure 1.2d). The transducer readings were calibrated against manual measurements and continuous stream discharge was derived through the calibrated stage-discharge relationship provided by the manufacturer

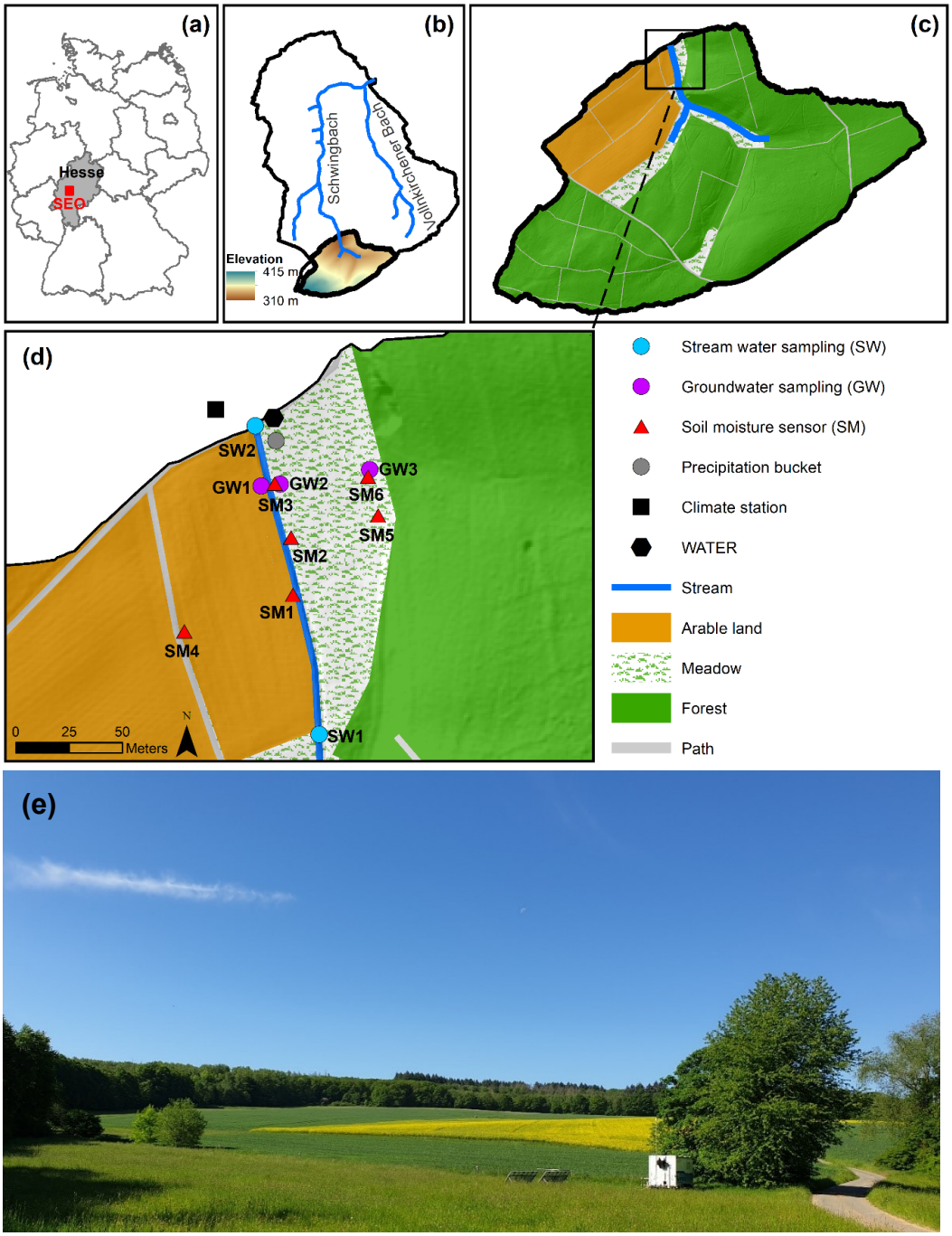

Figure 1.1 (a) Location of the Schwingbach Environmental Observatory (SEO) in Hesse, Germany (red square), (b) Schwingbach Environmental Observatory (SEO), (c) Study area in the Schwingbach headwater, (d) Measuring network along the stream reach of the Schwingbach headwater and (e) Landscape of the study area, with the mobile laboratory WATER and two solar power panels in the foreground right.

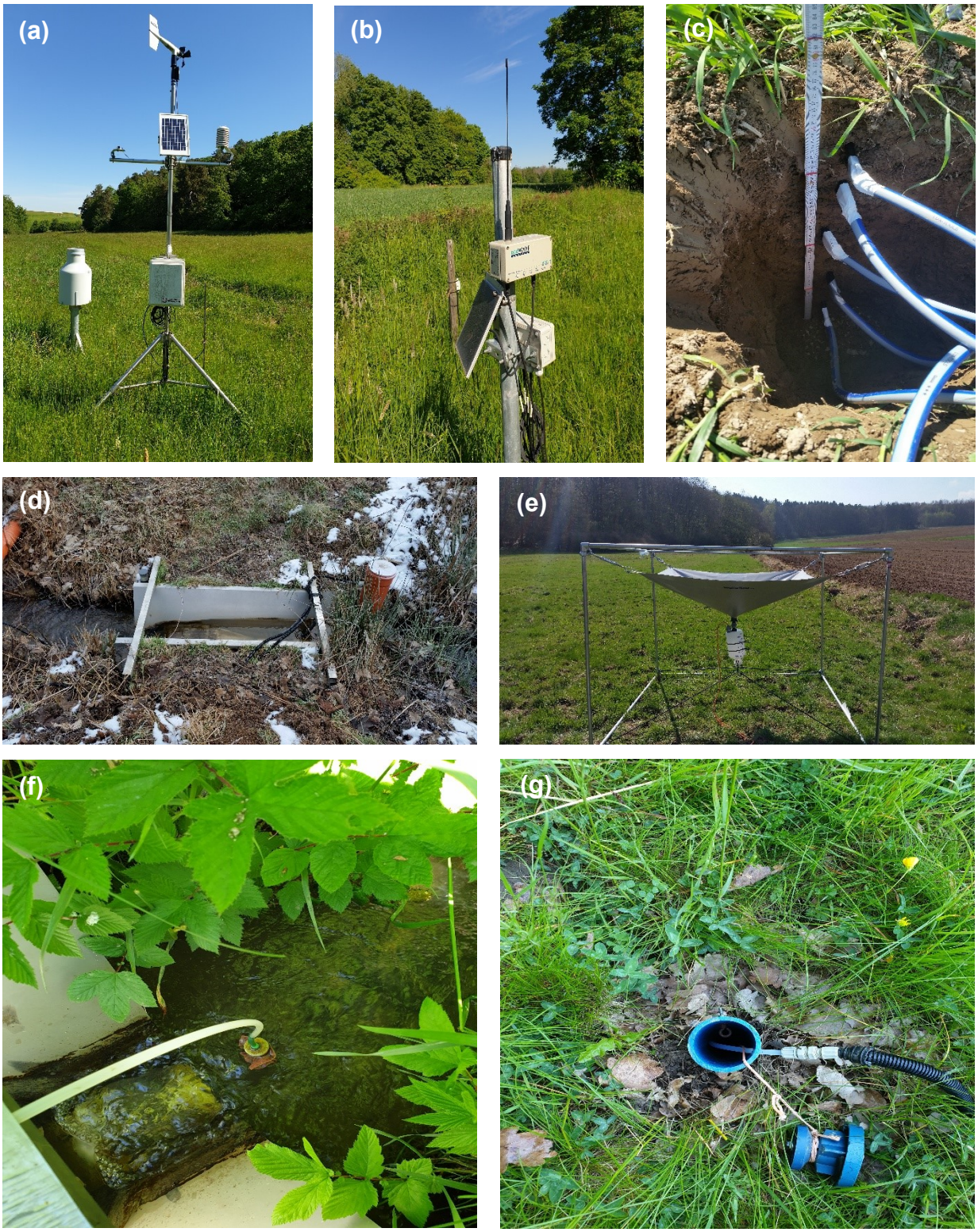

Figure 1.2 Monitoring network: (a) Climate station, (b) Remote telemetry logger, (c) Soil moisture sensors at 5, 15, 30, 50 and 70 cm depths, (d) Stream gauge at SW2 (outlet), (e) Precipitation bucket, (f) Stream water sampling station at SW2 and (g) Groundwater sampling station at GW3**.**

1.3.2 Water Analysis Trailer for Environmental Research (WATER)

An automated mobile laboratory, the Water Analysis Trailer for the Environmental Research (WATER), was utilized to automatically sample and analyse the stable isotopes of water (δ^2H and $\delta^{18}O$) and water quality for multiple water sources at 20 min intervals in situ (Figure 1.3a-c). The automatic sampling system allows to sample up to 12 water sources. A peristaltic pump (M1500, Verder GmbH, Haan, Germany) draws water from the sources and delivers it through a 5 µm polyester membrane (polyester filter, Pieper Filter GmbH, Germany) to a 1,000 mL sample reservoir, which is connected to a continuous water sampler (CWS) (A0217, Picarro Inc., Santa Clara, USA), coupled to a wavelength-scanned cavity ring-down spectrometer (WS-CRDS) (L2130-i, Picarro Inc., Santa Clara, USA) to analyse stable isotopes of water. A multi-parameter water quality probe (YSI600R, YSI Inc., Yellow Springs, USA) is installed on the sampling board of the WATER to measure temperature, pH and EC of sampled water.

In the setup of this dissertation, the WATER automatically sampled two stream water reaches (SW1 and SW2), three groundwater sources (GW1, GW2 and GW3) and one precipitation collector (Figure 1.1d and Figure 1.2e-g). SW1 was sampled approximately 145 m upstream of the WATER close the arable land and water from SW2 was taken at the outlet of the catchment next to the location of the WATER. Piezometers of GW1 and GW2 were located at the toeslope on arable land and meadow, respectively. The piezometer of GW3 was installed at the footslope at the edge of the forest. Precipitation was sampled from a 5 L bucket falling on a funnel-type 3.3 m^2 funnel-shaped tarp. The bucket was equipped with a sensor (tank electrode, Votronic GmbH, Lauterbach, Germany) to monitor the amount of the collected precipitation inside the bucket. Given the sampling resolution of 20 min, the system setup allowed the measurement of 72 water isotope samples per day. The sampling cycle was scheduled to sample $n = 16$ (with n the number of samples) for each stream water reach, $n = 8$ for each groundwater source and $n = 8$ for each standard water sample in case of no precipitation event. In the case of a precipitation event, the sampler automatically took a precipitation sample whenever the sample volume exceeded 1 L (0.3 mm) inside the precipitation bucket. After sampling, the precipitation bucket was emptied, and the sampler was blocked for 60 min, allowing to take samples from other sources.

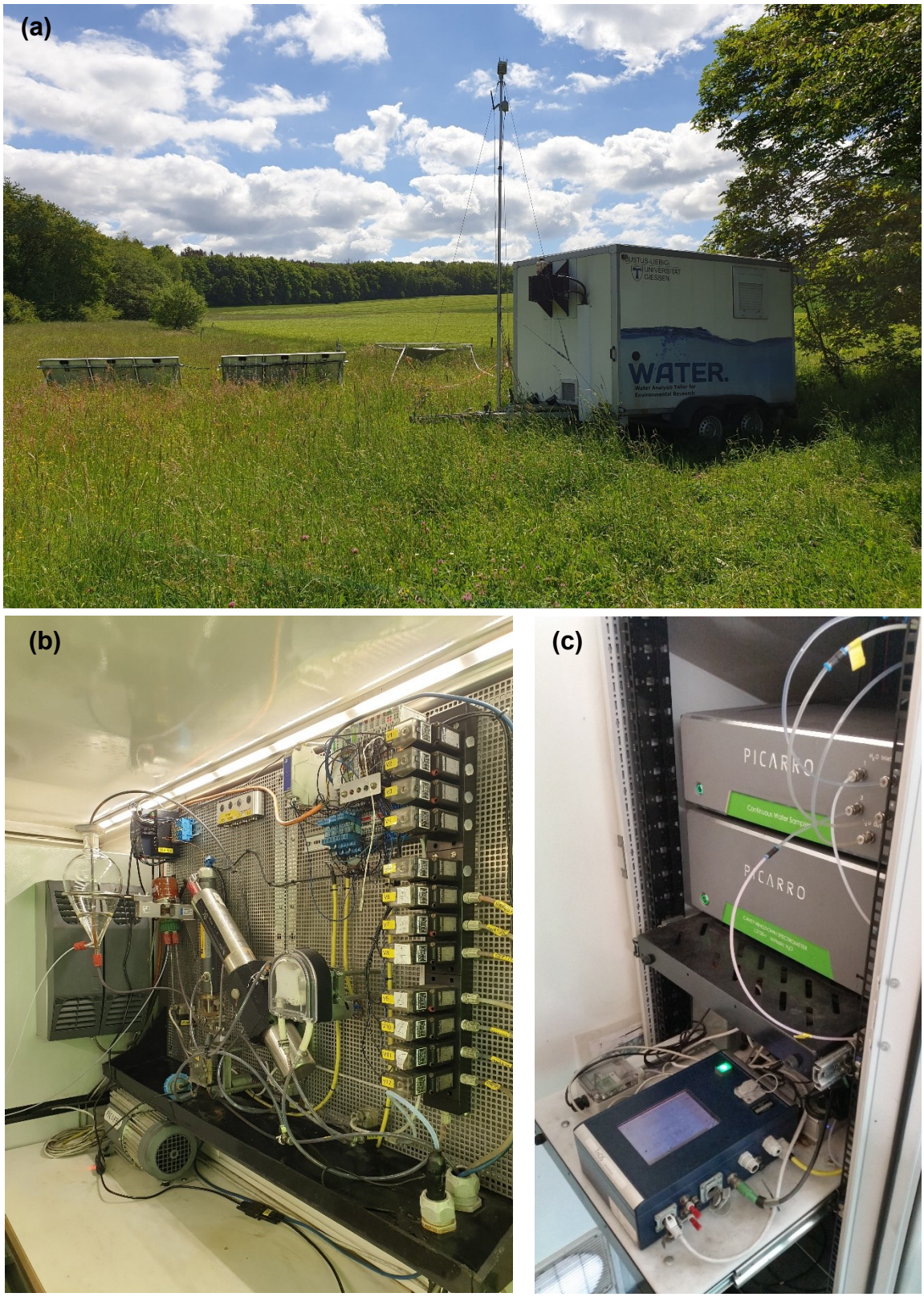

Figure 1.3 (a) Water Analysis Trailer for Environmental Research (WATER), (b) Sampling board and analytical instrumentation inside the WATER and (c) Continuous water sampler (CWS) coupled to a wavelength-scanned cavity ring-down spectrometer (WS-CRDS) for stable water isotopes analysis.

1.3.3 Artificial Neural Network (ANN)

Multilayer Perceptron (MLP) feedforward neural network is one of the most popular ANN architectures in hydrology (Tanty and Desmukh, 2015; Oyebode and Stretch, 2018). The MLP network consists of a set of nodes (i.e., neurons) organized into three different types of layers, i.e., input, hidden, and output layers (Figure 1.4). The input layer contains nodes, through which the input data is incorporated into the network. In the hidden layers, the nodes receive signals from the nodes in the previous layer via weighted connections, at which the weights determine the strength of the signals. The weighted values from the previous layer are summed together with a bias assigned with the nodes. The result is then passed through an activation function to produce a signal (i.e., activation level) for each node. The signals are then transmitted to the subsequent layer and the process is continued until the information reaches the output layer. The signals, which are produced in the output layer, are the model outputs. To train the MLP network, the backpropagation algorithm (Lecun et al., 2015) was used, in which the outputs generated in the output layer for the given inputs, are compared to the target values (i.e., observations) and the error is calculated through a loss function. If the output layer does not generate the desired values, the output errors are propagated backwards through the network, while the parameters of the nodes (i.e., weights and bias) in each layer are updated along the way with an optimizer. The learning process repeats in this manner for several rounds (i.e., epochs) until the loss function reaches an optimal value.

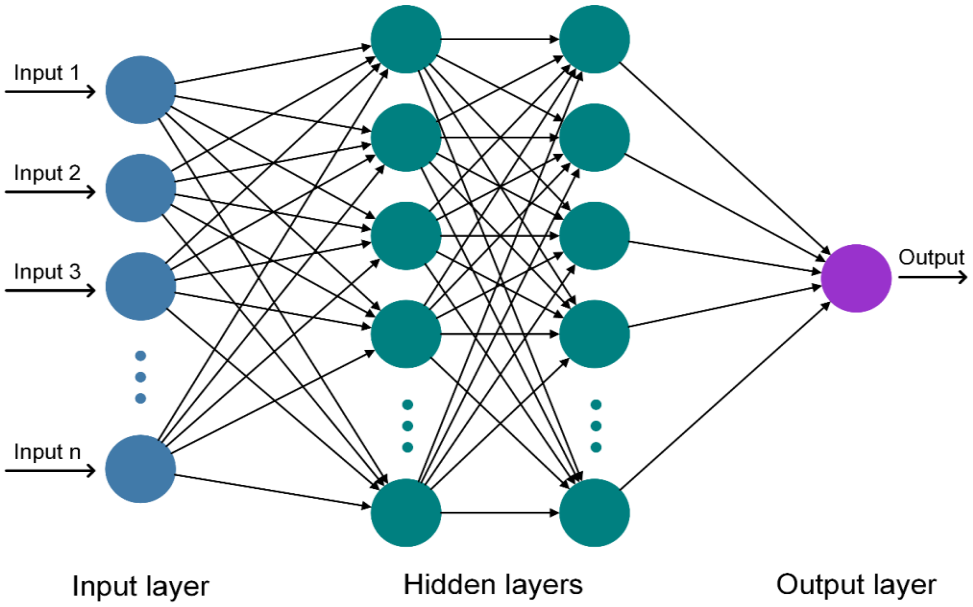

Figure 1.4 Schematic diagram of an Artificial Neural Network (ANN), where the network consists of a set of nodes organized into input, hidden and output layers.

1.3.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) was developed by Cortes and Vapnik (1995) based on the statistical learning theory. The SVM is a novel classification algorithm, which employs the principle of structural risk minimization and transforms it into quadratic programming. It uses suitable kernel function to non-linearly map the original data in a higher dimensional feature space and solve a linear regression problem in the feature space. In the feature space, the SVM establishes decision boundaries to separate data points of different classes. Its intent is to create an optimal separating hyperplane between two classes to minimize the generalization error and thereby maximize the margin (Figure 1.5). If any two classes are separable from among the infinite number of linear classifiers, the SVM determines that hyperplane, which minimizes the generalization error. Conversely, if the two classes are non-separable, the SVM tries to search that hyperplane, which simultaneously maximizes the margin and minimizes a quantity proportional to the number of misclassification errors. Thus, the selected hyperplane will have the maximum margin between the two classes. The margin is defined as a summation of the distances between the separating hyperplane and the nearest points on either side of the two classes (i.e., support vectors). The SVM algorithm was extended to Support Vector Regression (SVR) to solve the regression problems with the introduction of the ε-insensitive loss function (Vapnik et al., 1997; Smola and Schölkopf, 2004). In this case, the SVR attempts to determine the optimal hyperplane, which minimizes the distance of all data points.

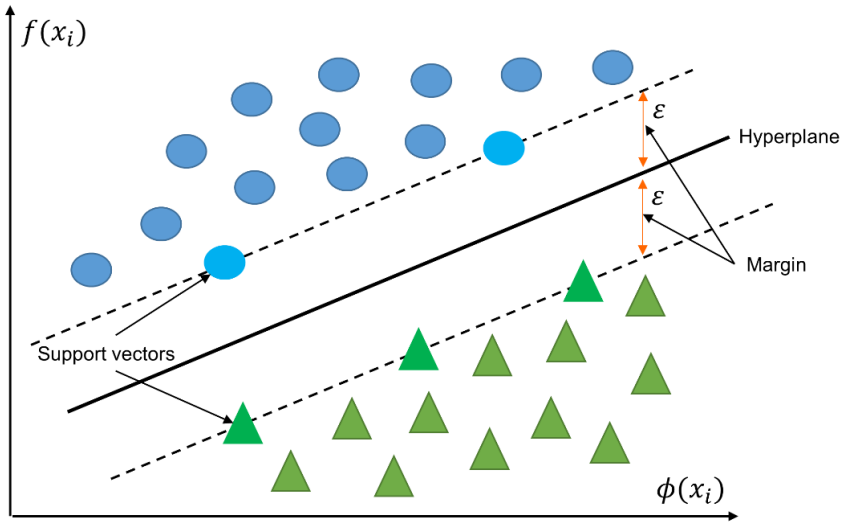

Figure 1.5 Schematic diagram of a Support Vector Regression (SVM), with the hyperplane splitting up two classes of support vectors by maximizing the margin.

1.3.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is considered an evolutionary form of the Recurrent Neural Network (RNN) that is capable of learning long-term dependencies by overcoming the exploding and vanishing gradient problems of traditional RNN networks (Hochreiter and Schmidhuber, 1997). The main characteristics of the LSTM are the specially designed units so called *memory cell* and *gates*. The memory cell consists of forget, input and output gates that together control the flow of information within the LSTM network. Each memory cell is interconnected through two vital components, at which the hidden state h_{t-1} is known as short-term memory and the cell state c_t is referred to as long-term memory. By using the cell state, the LSTM is able to store essential information for a prolonged period. Figure 1.6a-d illustrates the LSTM learning process. In the first step, the forget gate parameter f_t decides how much of the information is removed from the cell state. This decision is made by a sigmoid function (σ) with a linear calculation of the current input x_t and the previous output h_{t-1}. In the second step, the input gate parameter i_t controls which new information is updated to the cell state by the sigmoid function with a linear relation on x_t and h_{t-1} as well. The new cell state candidate c_t' is calculated by a hyperbolic tangent function (*tanh*) with a linear relation on x_t and h_{t-1}. In the third step, the old cell state c_{t-1} is updated into the new cell state c_t through an element-wise multiplication operator. Finally, the output parameter o_t is calculated by the sigmoid function with a linear relation on x_t and h_{t-1}. The final output at the current time step h_t is the production of o_t and the $tanh$ function value of the cell state c_t.

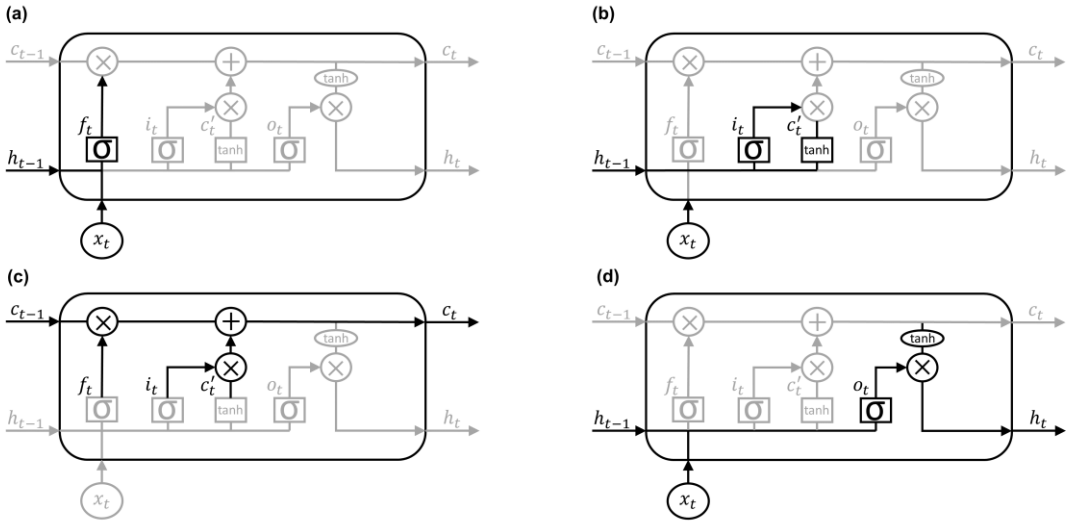

Figure 1.6 Schematic of Long Short-Term Memory (LSTM) learning process in four steps, where the memory cell consists of forget, input and output gates that together control the flow of information.

1.3.6 Hyperparameter optimization of machine learning algorithms

Most of the machine learning algorithms have several settings that govern the entire learning process (Goodfellow et al., 2016). These settings are called *hyperparameters*. The hyperparameters are external to the model, they are not learned from the data and must be set prior to the learning process (Géron, 2019). The performance and computation complexity of machine learning models are crucially dependent on the hyperparameter setting. Hence, it is necessary to optimize the hyperparameters in order to enhance the model performance. However, investigations of different hyperparameters and comparison of their influence on the model performance are rarely reported in hydrology. This dissertation provides detailed discussion with respect to the optimization process of the machine learning algorithms.

1.3.6.1 ANN setup

As the base architecture, the ANN model applied in this work was defined by seven nodes in the input layer, and two hidden layers followed by a linear output layer with one node. The Adam optimizer was used for the optimization of the learning process (Kingma and Ba, 2014). Among the hyperparameters, the number of nodes in hidden layers, learning rate, and activation functions in hidden layers have the greatest impact on the performance of the model and its robustness (Smithson et al., 2016; Klein and Hutter, 2019). Therefore, a set of the aforementioned hyperparameters was defined to evaluate the performance of the ANN model configurations. Too few hidden nodes implies that the network does not have enough capacity to solve the problem. Inversely, too many hidden nodes implies that the network memorizes noise within training data, leading to poor generalization capability (Maier and Dandy, 2000; Thomas et al., 2016b). Given that, five different configurations were tested for the number of hidden nodes. Further, too small learning rates slow down the learning process and may cause the learning to become stuck with a high learning error. On the other hand, too large learning rates can lead the network to go through large oscillations during the learning process or that the network may never converge (Maier and Dandy, 1998; Goodfellow et al., 2016). The default learning rate of the Adam optimizer is set to 0.001. Smaller (0.0001) and larger (0.01) learning rates were also tested. Finally, several activation functions where assessed. Sigmoid and tanh are the most common activation functions used in feedforward networks (Maier and Dandy, 2000). However, Rectified Linear Unit (ReLU) is usually a more suitable choice (Maier and Dandy, 2000). Hence, sigmoid, tanh and ReLU were checked as

varieties for activation functions of hidden layers. Each network was trained for a maximum of 10,000 epochs to minimize the mean absolute error (MAE).

1.3.6.2 SVM setup

The performances of four commonly employed kernels in SVM studies, including the Linear, Sigmoid, Radial Basis Function (RBF), and Polynomial kernels combined with their optimal hyperparameters were examined in detail. The Linear kernel has two hyperparameters (C, ε), the Sigmoid and RBF kernels have three hyperparameters (C, ε, γ) and the Polynomial kernel has four hyperparameters (C, ε, γ, d) to be optimized. The penalty coefficient C determines the tolerance of deviations larger than ε from the real value; i.e., smaller deviations are tolerable for larger values of C. The permitted error threshold ε affects the number of support vectors used in the regression function, i.e., the smaller the value of ε, the greater the number of support vectors that will be selected. The kernel coefficient γ, controls the influence radius of the support vectors, i.e., for high values of gamma, the radius of the influence area of the support vectors only includes the support vector itself. The degree of the polynomial function d, is the largest of the degrees of polynomial's monomials with non-zero coefficient. 20 values for each of C, ε and γ based on geometric progression and 3 values for d were tested. Therefore, 400 configurations for the Linear kernel, 8,000 configurations for the Sigmoid and RBF kernels and 24,000 configurations for the Polynomial kernel were studied. Finally, the performances between the kernel functions in combination with their optimal hyperparameter configurations were compared.

1.3.6.3 LSTM setup

The LSTM architecture consisted of an input layer with as many neurons as input features, one LSTM layer, a dropout layer and a fully connected dense layer with a single unit for the output feature. The Adam optimizer was used for the optimization of the learning process (Kingma and Ba, 2014) and tanh and sigmoid functions as state and gate activation functions, respectively. The LSTM model was run in a sequence-to-one mode so that an input sequence of a fixed length was used to predict output value at the next time step. The sequence length, i.e., the look-back window, is the length of past input observations that the LSTM looks back to predict an output for the next time step. The Sequential Model-Based Optimization (SMBO) search with the Tree-structured Parzen Estimator (TPE) algorithm were used in the frame of a Bayesian optimization approach (Bergstra et al., 2011) to optimize the number of hidden, dropout rate, learning rate, number of epochs and batch size. Bayesian optimization is a very efficient optimization algorithm that has

been demonstrated to outperform well-established approaches i.e., grid and random search (Bergstra et al., 2013; Eggensperger et al., 2013). It develops a statistical model between the hyperparameters and the objective function and makes the assumption that there is a smooth but noisy function that maps between the hyperparameters and the objective function (Reimers and Gurevych, 2017). Given the search history of hyperparameters and the objective function, SMBO-TPE suggests hyperparameters for the next trial that are expected to improve the objective function. As the number of trials grows, the search history grows and eventually the hyperparameters become optimized. The optimization evaluation was run for 1,000 trials through the search space of the hyperparameters to minimize the objective function, in this case the mean squared error (MSE).

1.3.7 Model evaluation

To evaluate the generalization ability of the machine learning models, the data was divided into train, validation and test sets. The train set was used to learn the internal parameters (i.e., weights and bias) of the model. The validation set was used for the hyperparameter optimization and the test set was used to estimate the generalization capability of the model. For ANN and SVM models, ten times 10-fold cross-validation was performed such that eight folds were considered as a train set, one fold as the validation set and one fold as a test set. For the LSTM model, the data was split into 70%, 15% and 15% of train, validation and test sets, respectively, while maintaining the temporal order of the observations. The prediction performances of machine learning models were evaluated in terms of the root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2) and Nash-Sutcliffe efficiency (NSE).

1.4 Main Results

1.4.1 Objective 1: Investigation of the response of runoff components to precipitation and the role of underlying controlling factors on the runoff generation process

The isotopic response of stream and groundwater during the precipitation events revealed a rapid mixing of event water (i.e., precipitation) with pre-event water (i.e., water stored in the sampling sources), indicating that the Schwingbach is a highly responsive catchment (Figure 1.7). It took,

on average, 6 h for stream water and 7 h for groundwater at the toeslope that the isotope signatures reach peak values implying that shallow subsurface flow pathways rapidly delivered water to the stream. Macropore and soil pipe networks along relatively flat areas in stream banks were likely relevant pathways for the rapid transmission of water.

A two-component isotopic hydrograph separation technique was applied to quantify event and pre-event water fractions in stream and groundwater sources. The hydrograph of 20 precipitation events showed that more than 79% of runoff consisted of pre-event water. Due to the highly permeable soils, the contribution of overland flow was negligible and hence precipitation infiltrated to the subsurface and mobilized pre-event water instead of running off to the stream. The runoff coefficient, i.e., runoff divided by the corresponding precipitation amount, was very low across the events with a mean of 2.2%. The high porosity of the soils in the upper layers of meadows and arable land as well as the extremely dry conditions in summer 2018 were likely the major reasons for the low runoff coefficients.

A Spearman rank correlation analysis was applied to investigate the control of precipitation and antecedent wetness conditions on the response characteristics of the catchment. The correlation values were considered statistically significant if $p < 0.05$. Total event water volume and event water fraction in total runoff significantly increased with the precipitation amount pointing out that the event water contribution increased with the precipitation amount. The positive correlation of the maximum event water fraction in groundwater sources with precipitation amount and initial soil moisture at the toeslope pointed out that not only the higher precipitation amount led to an increase of event water contribution but also that wetter conditions at the toeslope facilitated groundwater recharge mechanisms to switch on. Total pre-event volume moderately correlated with total precipitation amount suggesting that increasing precipitation led to the mobilization of pre-event water. Nevertheless, no significant correlation was observed between pre-event water fraction in total runoff and antecedent wetness of the catchment. This indicates that the effect of antecedent wetness was too low on pre-event water contribution, which was likely caused by the extreme dry conditions in 2018.

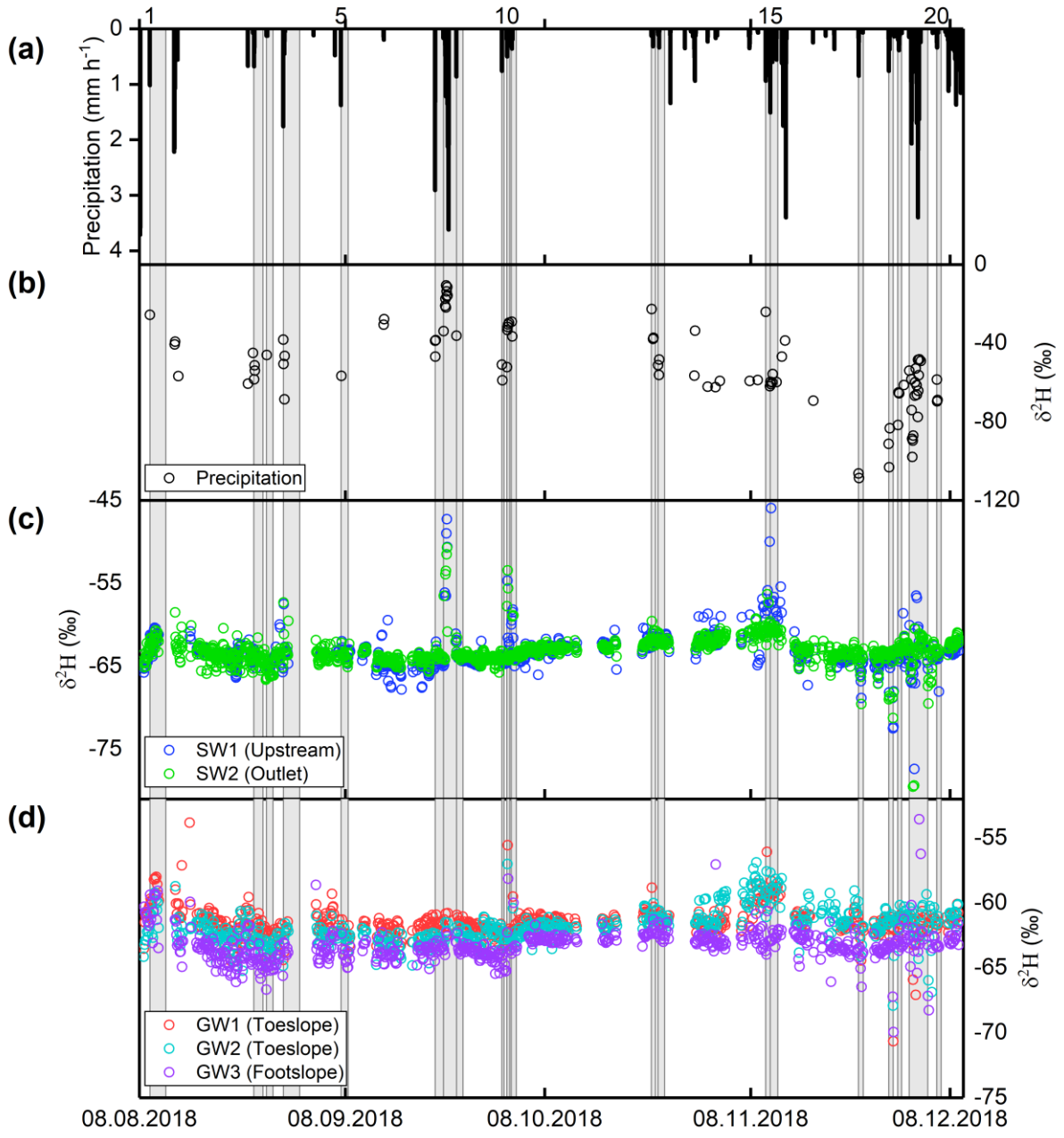

Figure 1.7 Time series of (a) Precipitation, (b) δ^2H in precipitation, (c) δ^2H in stream water sources and (d) δ^2H in groundwater sources. The vertical grey bars indicate the 20 precipitation events investigated in detail. The numbers on top of panel (a) represent the event ID.

1.4.2 Objective 2: Prediction of the response of runoff components to precipitation using machine learning methods

ANN and SVM algorithms were optimized to predict maximum event water fractions in streamflow on independent precipitation events in the SEO using only precipitation, soil moisture, and air temperature. The optimized ANN and SVM models were selected based on the average performance over 100 iterations derived from ten times 10-fold cross validation. The optimized ANN was chosen based on a combination of five nodes in the first hidden layer and three nodes in the second hidden layer [5, 3], a learning rate of 0.0001 and the tanh activation function in hidden layers. This configuration resulted in an RMSE of 12.91%, MAE of 10.27%, R^2 of 0.70, and NSE of 0.63. The optimized SVM was achieved by a Polynomial kernel function with kernel hyperparameters of ($C = 8.8587$, $\varepsilon = 0.0162$, $\gamma = 0.1833$, $d = 2$). The optimized SVM outperformed the ANN, with an RMSE of 9.43%, MAE of 7.89%, R^2 of 0.83, and NSE of 0.78. SVM could better capture the dynamics of maximum event water fractions across the events (Figure 1.8). ANN tended to underestimate the events with high maximum event water fractions and to overestimate the events with low maximum event water fractions.

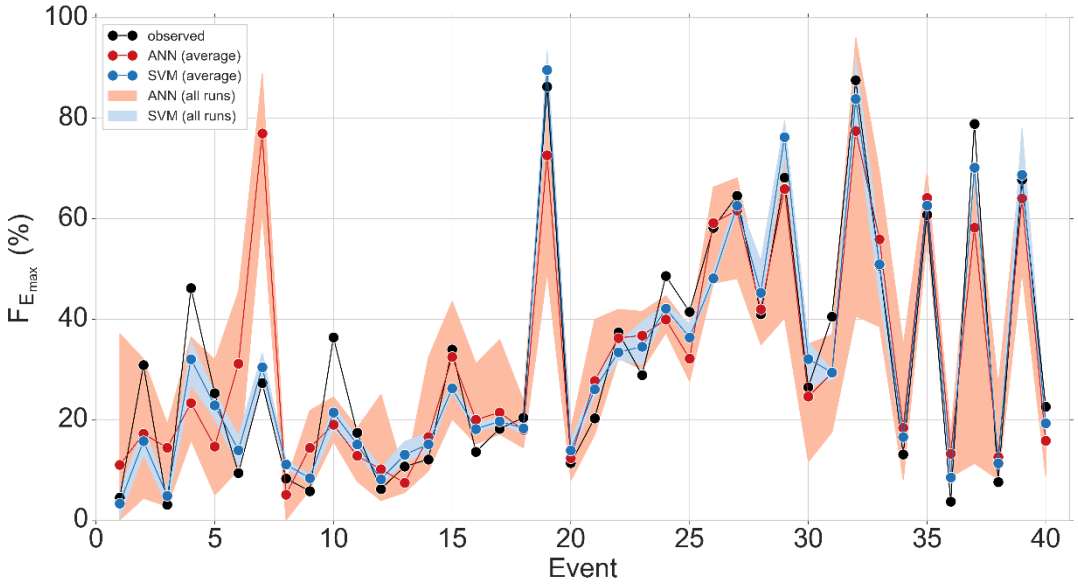

Figure 1.8 Comparison of observed and predicted maximum event water fractions $\boldsymbol{F_{E_{max}}}$ at 40 precipitation events simulated by the optimized ANN and SVM models. The red and blue circles indicate the average of ten predicted values at each event and red and blue bands indicate the range of all ten predicted values (i.e., all runs) at each event derived from ten times 10-fold cross-validation by the optimized ANN and SVM models, respectively.

The influence of the hyperparameter configurations on the performance of the ANN and SVM models was further investigated. Choosing an appropriate number of hidden nodes and a proper activation function enhanced the ANN model performance, whereas changes of the learning rate did not have a major impact on the model performance when using the Adam optimizer. The result highlighted that increasing the complexity of the network not only declined the model performance but also decreased its stability. For the SVM, the model performance was directly related to the selection of the kernel function. The Polynomial and RBF kernels were found to be the most appropriate kernel functions for modelling of the non-linear behaviour, whereas the Linear and Sigmoid kernels were weaker and less stable over differently shuffled train and test sets.

1.4.3 Objective 3: Simulation of high temporal resolution of stable isotope concentrations in multiple water sources using deep learning

A LSTM deep learning model was developed to predict the complex and non-linear high-resolution (3 h) isotope concentrations of multiple stream and groundwater sources under different landuse and hillslope positions in the SEO. The core concept was to explore the prediction performance of an LSTM that was trained on multiple sites, with a set of explanatory data that were more straightforward and less expensive to measure compared to the stable isotopes of water. The explanatory data consisted of meteorological data, catchment wetness conditions and natural tracers, i.e., water temperature, potential of hydrogen (pH) and electrical conductivity. A sensitivity analysis was run to examine how different input feature scenarios and sequence lengths influence the prediction performance of the LSTM. The result showed that using the meteorological data and catchment wetness conditions alone as the input features did not lead to a satisfying prediction performance. The relations of these variables with the water isotope concentrations were not strong enough to transfer enough information that was required for an adequate learning of the LSTM to predict isotope concentrations. In contrast, using short sequence lengths (6-12 h) of the natural tracers, i.e., water temperature, pH and electrical conductivity of the sampling sites as the input features achieved good prediction performance. It indicates that these simple-to-measure tracers were able to indirectly model complex interactions between controlling drivers and water isotope concentrations in the catchment. This could be associated with the rapid response characteristics

of the Schwingbach catchment as well as the synchronized behaviour of isotopic compositions with electrical conductivity.

A single LSTM model was trained and optimized with the input feature scenario and sequence length, at which the model performed on average the best for all of the sampling sites, rather than training and optimizing a separate LSTM model for each of the sampling sites. Based on the sensitivity analysis, the LSTM achieved the best performance when only measurements of the last 6 h of water temperature, pH and EC were used as input features. On average across all sampling sites, an RMSE of 0.7‰, MAE of 0.4‰, R^2 of 0.90 and NSE of 0.70 was achieved. The optimized LSTM was identified through the Bayesian optimization algorithm by a combination of 60 hidden units, a batch size of 40, a dropout rate of 0.4, and a learning rate of 0.0001 over 40 training epochs. The LSTM captured the timing of hydrologic events and base flow conditions of δ^2H quite well, but struggled to predict the peak of the hydrograph (Figure 1.9). The overall performance of the LSTM indicated that it efficiently performed in stream and groundwater sources. A single LSTM that was simultaneously trained on multiple sampling site with different landuse and hillslope positions with a range of different isotope concentrations and magnitudes of response to precipitation events tended to balance the error so that it minimized the error between predicted and observed isotope concentrations for all of the sampling sites at the same time.

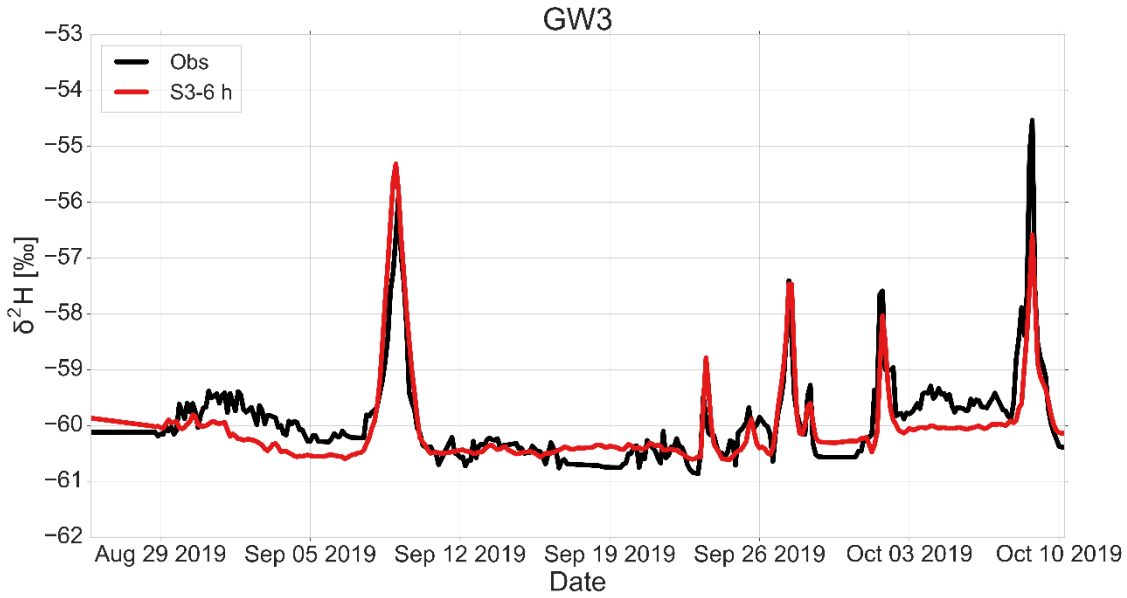

Figure 1.9 Comparison of observed and predicted δ^2H values by the best LSTM using a sequence length of 6 h for water temperature, pH and electrical conductivity measurements (scenario S3-6 h) at GW3 site.

1.5 Implications and Outlook

Powered by machine learning, AI is rapidly reshaping research and our lives. On the scientific front, machine learning is transforming many scientific disciplines, and is widely adopted and regarded as an indispensable tool for the future. In hydrology, the number of machine learning-related publications is skyrocketing in recent years. Several initiatives are pressing for paradigm shifts in the way hydrological modelling is done. The two key frontiers are the use of machine learning and the push for higher-resolution simulations (Klaus and McDonnell, 2013; Shen, 2018). Penna et al. (2018) call for a better incorporation of high-resolution modelling and observations in isotope hydrology. This dissertation aims to propel the isotopic modelling research by demonstrating the application of machine learning for the simulation of high-resolution stable isotopes in stream and groundwater sources.

"If we want to understand the full symphony of catchment hydrochemical behaviour, then we need to be able to hear every note" (Kirchner et al., 2004). Long-term, high-resolution isotope measurements are essential to capture the full spectrum of complexity within hydrological processes so that notes from the seasonal down to the event scale can be detected. The study presented in Chapter 2 demonstrated that the automated high-resolution sampling revealed fine-scale hydrological responses that were not possible to be detected at coarser resolutions in previous studies. This provides new insights on catchment functioning and flow pathways. The hydrograph separation technique clearly showed that long-term, high-resolution isotopic analyses are crucial for capturing the unusual but informative events to more precisely constrain the event and pre-event water contribution. It furthered stressed the importance of sampling multiple sources that have been overlooked so far. The sampling of groundwater along the stream water underlined the vital role of shallow subsurface flow in runoff generation mechanisms. High-resolution isotope observations have a great potential for developing and testing of hydrological models and their process representation. Typical weekly or monthly monitoring programmes cannot capture the short-term dynamics that most closely reflect these hydrological processes. Therefore, high-resolution isotope observations coupled with hydroclimatic data could be beneficial in calibrating the next generation of hydrological models. Advancing the future research to carry out high-resolution sampling over longer periods under different hydroclimatic conditions would help to uncover the diversity of hydrological symphonies.

The estimation of maximum event water fractions in streamflow based on isotopic hydrograph separation provides significant insight into the mechanisms of runoff generation and hydrologic response characteristics of a catchment. However, such an estimate is not always possible due to the spatiotemporal difficulties in sampling and measuring of stable isotopes of water. Machine learning showed a strong promise of success for predicting the maximum event water fractions at times when no stable isotopes of water are available (Chapter 3). It used only precipitation, soil moisture and air temperature as explanatory data. This is convenient because such variables are available for far more catchments than measurements of stable isotopes of water.

Hand tuning of the machine learning models can be a cumbersome process especially when the search space of hyperparameters is large (Wu et al., 2019; Sahu et al., 2020). A key innovation of this dissertation is the use of a systematic hyperparameter optimization approaches through cross-validation to test different model architectures for developing prediction models. The results showed that the optimization of hyperparameters had a great influence on the model performance that has been widely neglected in hydrological studies so far. Therefore, it is essential to examine the model performance by a large set of hyperparameters for future studies to determine the optimal hyperparameters and ensure the prediction ability of the model. Future work could apply the proposed ANN and SVM models to other catchments to allow assessing how well the models are transferable to catchments with different topography, landuse and climatic conditions. Furthermore, the proposed models could be compared to other machine learning models using different input scenarios of hydroclimatic data.

Recently, deep learning has emerged as a revolutionary and versatile tool that is generating new and improved capabilities for scientific discovery and modelling. The adoption of deep learning in hydrology has so far been gradual, but the field is now ripe for breakthroughs. For the first time in the field of isotope hydrology, this dissertation adopted deep learning in isotope hydrology (Chapter 4). The proposed LSTM model could successfully predict stable isotopes of stream and groundwater sources when using only short-term sequence of measured water temperature, pH and electrical conductivity. The LSTM prediction can be utilized to predict and interpolate the continuous isotope concentration time series either for data gap filling or in case where no continuous data acquisition is feasible. This is very valuable in practice because measurements of these tracers as proxies are still much cheaper than stable isotopes of water and can be carried out continuously with relatively little maintenance. The strength of LSTM as a deep learning model is

that it can synthesize information from multiple sites and situations into a single model. An LSTM trained on multiple sites under different landuse and hillslope conditions can learn different patterns of isotopic behaviour. The trained LSTM is able to spatially estimate the isotope concentrations for multiple sites across the catchment with only simple-to-measure tracers. This leads not only to a significant reduction in measurement costs, but also to an increase in spatiotemporal knowledge of hydrological processes in the catchment.

With the increased deployment of LSTM, there is a need to understand the sensitivity of its prediction performance to different input features and sequence lengths. The sensitivity analysis applied in this dissertation stressed the importance of sensitivity analysis for the model performance. Such an analysis could be used to inform about which data should be collected, and where data acquisition should be made to effectively optimize the investment in monitoring infrastructure. For future research, more data will be collected to improve the prediction performance of the LSTM. New input features like groundwater table could be also used to enhance the model performance.

The data-hungry nature of deep learning models is a potential barrier to their application in data-scarce catchments. As many catchments of potential applications may lack the data length that is available in the SEO where this work has been carried out, the sensitivity of prediction performance to the length of the training data warrants further investigation. The use of "pre-trained LSTM" is a promising way to reduce the large data-demand for an individual catchment. The pre-trained model could be exploited to transfer learning from data-rich to data-scarce catchments since hidden layers that have been trained to distill shape information are also effective even when ported to data-scarce environments.

Finally, future research direction also includes the application of LSTM to predict further water quality parameters for efficient managing water resources and protecting ecosystems by providing an early warning against water quality deterioration. This particularly addresses the measurement of variables that are cumbersome to measure or expensive to analyse, such as antibiotics, pesticides, or microplastics, among many others.

2 High-resolution, in situ monitoring of stable isotopes of water revealed insight into hydrological response behavior

This chapter is published in the journal *Water* as:

Sahraei, A.[1], Kraft, P.[1], Windhorst, D.[1], & Breuer, L.[1,2] (2020). High-resolution, in situ monitoring of stable isotopes of water revealed insight into hydrological response behavior. Water, 12(2), 565. https://doi.org/10.3390/w12020565

[1]Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, 35392 Giessen, Germany

[2]Centre for International Development and Environmental Research (ZEU), Justus Liebig University Giessen, Senckenbergstrasse 3, 35390 Giessen, Germany

Abstract

High temporal resolution (20-min intervals) measurements of stable isotopes from groundwater, stream water and precipitation were investigated to understand the hydrological response behavior and control of precipitation and antecedent wetness conditions on runoff generation. Data of 20 precipitation events were collected by a self-sufficient mobile system for in situ measurements over four months in the Schwingbach Environmental Observatory (SEO, temperate climate), Germany. Isotopic hydrograph separation indicated that more than 79% of the runoff consisted of pre-event water. Short response times of maximum event water fractions in stream water and groundwater revealed that shallow subsurface flow pathways rapidly delivered water to the stream. Macropore and soil pipe networks along relatively flat areas in stream banks were likely relevant pathways for the rapid transmission of water. Event water contribution increased with increasing precipitation amount. Pre-event water contribution was moderately affected by precipitation, whereas, the antecedent wetness conditions were not strong enough to influence pre-event water contribution. The response time was controlled by mean precipitation intensity. A two-phase system was identified, at which the response times of stream water and groundwater decreased after reaching a threshold of mean precipitation intensity of 0.5 mm h^{-1}. Our results suggest that high temporal resolution measurements of stable isotopes of multiple water sources combined with hydrometrics

improve the understanding of the hydrological response behavior and runoff generation mechanisms.

2.1 Introduction

Understanding the response of runoff components to precipitation and its controlling factors gives insight into runoff generation mechanisms. Total runoff responds to incoming precipitation via event water or pre-event water, the latter being water stored in the catchment prior to the onset of precipitation (Klaus and McDonnell, 2013). The fraction of event and pre-event water in total runoff and the timing of its responses may vary depending on the controlling factors such as topography, land use, precipitation, and antecedent wetness characteristics.

The use of stable isotopes of water ($\delta^{18}O$ and $\delta^{2}H$) and other conservative tracers in tracer-based hydrograph separation techniques enables the differentiation of source components of event and pre-event water and provides the opportunity to evaluate its controlling factors (Klaus and McDonnell, 2013). Several studies investigated the correlation of event and pre-event water contribution with topographical, land use, precipitation, and antecedent wetness characteristics. Shanley et al. (2002) noted that event water contribution correlated positively with catchment size and open land cover. Šanda et al. (2017) addressed the influence of land use on pre-event water contribution. They found that the pre-event water contribution decreased with increasing forest cover, particularly due to the retention of water via interception and transpiration losses of soil water. Several studies reported a positive correlation between precipitation amount and event water contribution (Pellerin et al., 2008; James and Roulet, 2009; Fischer et al., 2017a; von Freyberg et al., 2018). In contrast, Renshaw et al. (2003) showed a positive correlation between pre-event water contribution and precipitation amount, but only during peak discharge. Generally, wetter antecedent conditions result in an increase of pre-event water contribution, as they increase the connectivity of contributing areas (Pellerin et al., 2008; James and Roulet, 2009; Muñoz-Villers and McDonnell, 2012; Penna et al., 2015; von Freyberg et al., 2018). In contrast to those findings, Shanley et al. (2002) described positive correlations between antecedent wetness conditions and event water contribution. They hypothesized that infiltration must first fill storage deficits before event water can be rapidly transported via lateral flow pathways to stream. Response times (i.e., the time lag between the first detection of precipitation and maximum event water fraction) and their controlling factors have been only scarcely investigated, without finding a consensus

(McGlynn et al., 2004; James and Roulet, 2009). McGlynn et al. (2004) observed that response times increased with catchment size, whereas James and Roulet (2009) reported no strong correlation of response times with catchment characteristics. The latter work focused on small, forested headwater catchments. In such catchments, the canopy could possibly buffer the incoming precipitation signal. Another likely explanation is that different response times of groundwater blur the land use signal.

Most of the previous isotope-based studies dealing with the investigation of catchment responses used daily to monthly data, at which the fine-scale response behavior of runoff components is overlooked. Kirchner et al. (2004) noted that a sampling period longer than the hydrological response time of a catchment might lead to a significant loss of information. Only a few studies employed high temporal resolution sampling (i.e., sub-daily) of stable isotopes of water (McGlynn et al., 2004; James and Roulet, 2009; Fischer et al., 2017a; von Freyberg et al., 2017, 2018). However, these studies are limited to sampling a low number of water sources, i.e., mainly stream water and precipitation.

Manual high-resolution, multiple-source sampling is restricted to short durations and might fail in the sampling of relevant events due to their erratic and difficult-to-plan nature. Therefore, a few groups developed and utilized automatic sampling and analytical systems for in situ monitoring recently to overcome these problems (Heinz et al., 2014; von Freyberg et al., 2017; Mahindawansha et al., 2018). Such systems allow gaining insight into previously unknown hydrological dynamics. Von Freyberg et al. (2017) sampled stable water isotopes of stream water and precipitation every 30 min to estimate event water fractions of eight precipitation events using hydrograph separation. However, their study is limited to sampling two water sources, i.e., stream water and precipitation. Heinz et al. (2014) presented the technical setup of an automated sampling system and provided an initial proof-of-concept to monitor multiple water sources (surface water and groundwater) in up to ten rice paddies of a field trial in the Philippines. Building on this system, Mahindawansha et al. (2018) investigated seasonal and crop effects on isotopic compositions of surface water and groundwater in these rice paddies. An important, yet poorly known mechanism is the response of several water sources in real catchments, including that of shallow groundwater to precipitation.

Here we report about a newly developed system for high temporal resolution sampling of multiple sources, including groundwater, stream water, and precipitation. Previous studies in the Schwingbach catchment reported a fast response of the system to precipitation inputs (Orlowski et

al., 2014, 2016). They found that shallow groundwater head levels and streamflow responded rapidly to precipitation inputs. However, no obvious response was observed by Orlowski et al. (2016) in isotopic compositions of stream water and groundwater, potentially due to a rather coarse weekly sampling resolution. We, therefore, measured stable isotope composition of water in high temporal resolution (20-minutes) and of various sources. Sources included two stream sections, three different groundwater sources and precipitation. In order to investigate the hydrological response behavior and the role of controlling factors on the runoff generation process, we used a new, trailer-based mobile automatic sampling system for high-resolution water quality analysis. In particular, we studied the following objectives:

I. Investigation if the short-term response behavior observed in the previous studies in the Schwingbach Environmental Observatory (SEO) is reflected in isotopic signatures of stream water and groundwater.
II. Estimation of event-based contribution of event and pre-event water to total runoff.
III. Quantification of the event-based response time of maximum event water fraction in stream water and groundwater.
IV. Evaluation of the role of precipitation and antecedent wetness conditions as drivers of event and pre-event water contribution and response time.

2.2 Materials and Methods

2.2.1 Study area

The study was conducted in the headwater catchment of the Schwingbach Environmental Observatory (SEO) in Hesse, Germany (Figure 2.1a–c). The 1.03 km^2 catchment area is covered by 76% forests, 15% arable land, and 7% meadows, mainly found along the small perennial Schwingbach stream. The elevation ranges from 310 m in the north to 415 m a.s.l. in the south. The climate is temperate oceanic (Köppen climate classification), with a mean annual air temperature of 10.1 °C and total annual precipitation of 452 mm in the year 2018. Our study took place from August 8th to December 9th in 2018, which was an unusually dry and warm year (Vogel et al., 2019). Soil types and geology are similar to the neighbouring valley of the Vollnkirchner Bach (Orlowski et al., 2014). Cambisols dominate on forests stands, whereas Stagnosols are mainly found on arable land. Under the forests, the soil texture is dominated by silt, fine sand and gravel

at 0 to 3 m depths, with low clay content between 0.6 and 3 m depths. Soils lay on slightly weathered clay shale bedrock. Under meadows and arable land, the soil texture is rich in silt and fine sand (0–5 m depths) with small amounts of clay and gravel between 0.8 and 5 m depths. Here, soils developed on greywacke and weathered clay shale.

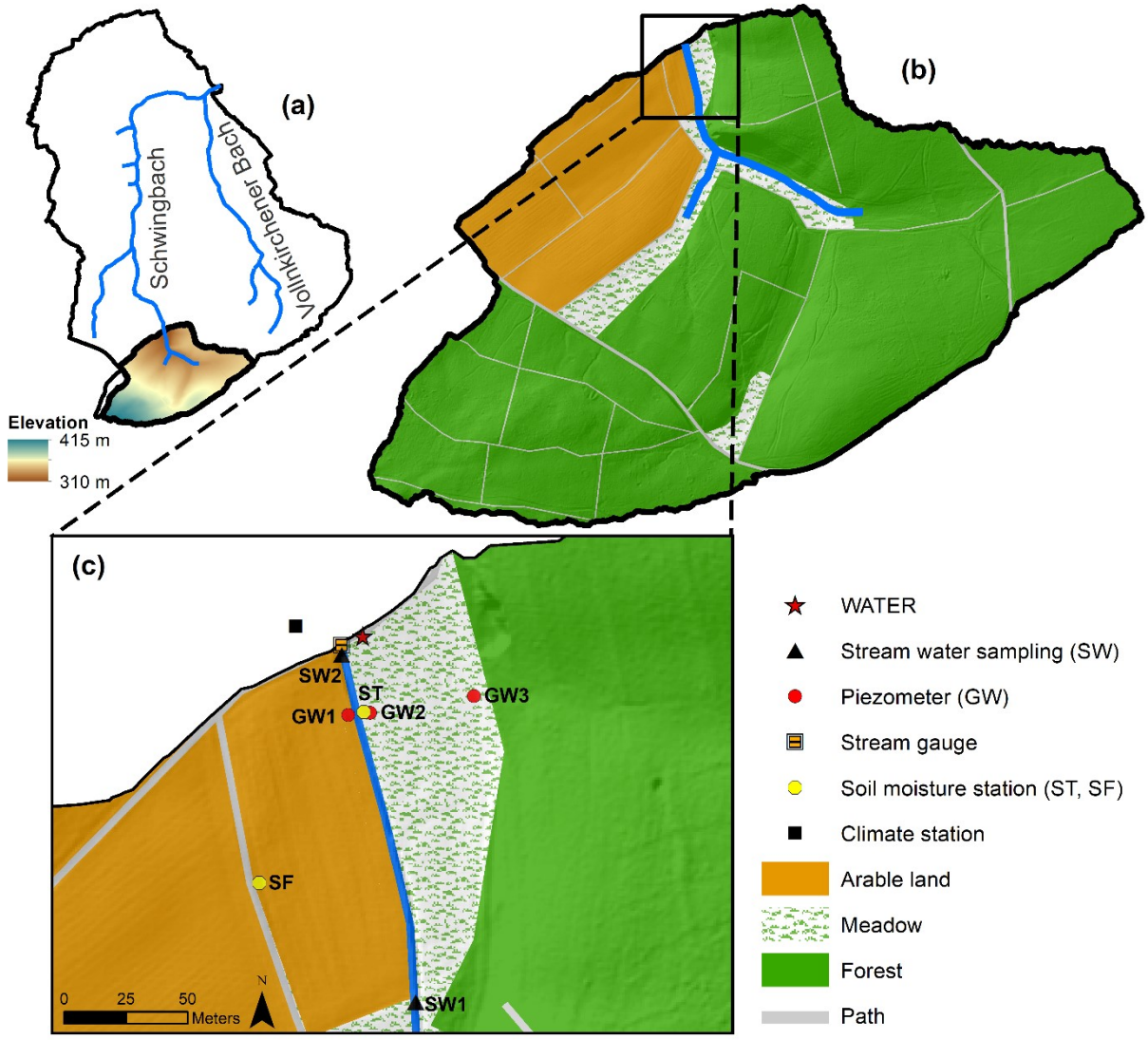

Figure 2.1 (a) Schwingbach Environmental Observatory (SEO), (b) Study area in the Schwingbach headwater and (c) Measuring network along the stream reach of the Schwingbach headwater. ST and SF are the soil moisture stations at the toeslope and footslope, respectively.

An automatic climate station (AQ5, Campbell Scientific Inc., Shepshed, UK) operated with a CR1000 data logger measured precipitation depth and air temperature at 5 min intervals. Groundwater (GW) table depth was manually measured biweekly to obtain an estimate of the groundwater depth in the unconfined aquifer over the year. The mean groundwater table depth (i.e., mean of biweekly measurements) from November 2017 to September 2018 was 0.44 ± 0.33 m (mean ± standard deviation), 0.48 ± 0.33 m and 1.41 ± 0.48 m for GW1, GW2, and GW3, respectively. A stream gauge at SW2 (RBC flume, Eijkelkamp Agrisearch Equipment, Giesbeek, Netherlands) equipped with a pressure transducer (Micro-Diver, Eigenbrodt Inc., Königsmoor, Germany) automatically recorded water levels at 10 min intervals. The transducer readings were calibrated against manual measurements and continuous discharge was derived through the calibrated stage-discharge relationship of the RBC flume provided by the manufacturer (Eijkelkamp RBC flumes, 2013). The discharge was converted to mm h^{-1} (i.e., divided by the area of the catchment) to allow for a direct comparison with precipitation in the same unit. Two remote telemetry loggers (A753, Adcon, Klosterneuburg, Austria), one installed at the toeslope (ST) and one at the footslope (SF), were equipped with sensors (ECH2O 5TE, METER Environment, Pullman, USA) to measure soil moisture at 5, 30, and 70 cm soil depths, representing the densely rooted organic horizon, the topsoil, and subsoil, at 5 min intervals.

2.2.2 Water Analysis Trailer for Environmental Research (WATER)

A custom-made, automated mobile laboratory, the Water Analysis Trailer for Environmental Research (WATER), was utilized to automatically sample multiple water sources (e.g., stream water, groundwater, and precipitation) and analyse the stable isotopes of water ($\delta^{18}O$ and $\delta^{2}H$) and water chemistry in situ (Figure 2.2). Constructing the trailer was based on previous experiences made by Heinz et al. (2014) and Mahindawansha et al. (2018). Figure 2.3 illustrates the technical set up of the WATER. The automatic sampling system allows us to sample up to 12 water sources. A peristaltic pump (M1500, Verder GmbH, Haan, Germany) draws water from the sources and delivers it through a 5 µm polyester membrane (polyester filter, Pieper Filter GmbH, Germany) to a 1000 mL sample reservoir, which is connected to a continuous water sampler (CWS) (A0217, Picarro Inc., Santa Clara, USA), coupled to a wavelength-scanned cavity ring-down spectrometer (WS-CRDS) (L2130-i, Picarro Inc., Santa Clara, USA) to analyze the water's isotopic composition. The deployment of the Picarro A0217 and L2130-i requires the supply of at least two standard water samples with a known isotopic signature. We used heavy (+3.13‰ for $\delta^{18}O$ and −1.52‰ for

δ^2H) and light (−22.43‰ for $\delta^{18}O$ and −164.60‰ for δ^2H) standard water samples, which were calibrated via a WS-CRDS (precision of 0.02‰ for $\delta^{18}O$ and 0.18‰ for δ^2H) at the laboratory of the Institute for Landscape Ecology and Resources Management at the Justus Liebig University Giessen, Germany. To prevent the isotope analyser from falling dry between switching the samples, a backup reservoir is installed. A 10 L bucket filled with deionized water flushes the system between the sampling cycles. A 5 L bucket filled by precipitation falling on a funnel-type 3.3 m^2 tarp is connected to the WATER for automatic sampling (Figure 2.2). The bucket is equipped with a sensor (tank electrode, Votronic GmbH, Lauterbach, Germany) to monitor the amount of the collected precipitation inside the bucket.

The WS-CRDS measured the isotopic composition of water vapour every 1.7 s. Due to the carry-over effects within the CWS, it took approximately 10 min until a steady isotope signal was reached. We averaged the last 3 min of the following 10 min sampling period to report the final isotope values. The overall sampling interval per sample summed up to 20 min. To minimize the carry-over effect, the sample reservoir was automatically rinsed with 60 mL sample water, emptied and again filled with 60 mL water to be analysed. Results are reported in (‰) for $\delta^{18}O$ and δ^2H relative to the Vienna Standard Mean Ocean Water (VSMOW).

Further analytical instruments of the WATER (Figure 2.3) measure water quality parameters such as electrical conductivity, pH, and water temperature (via multi-parameter water quality probe, YSI600R, YSI Inc., Yellow Springs, USA) and NO_3, DOC, TOC, and total suspended sediments via UV spectrometry (ProPS, Trios GmbH, Rastede, Germany). Results from water quality analysis are not reported in this study.

An in-house developed software written in Python runs on an industry PC (NISE 101, SEPCTRA, Reutlingen, Germany) with the Ubuntu Mate operation system to control the sampling scheme and the delivery system, as well as to store results and communicate with the WATER. The measurement processes are controlled by a programmable logic controller (WAGO-I/O-SYSTEM 750, WAGO Kontakttechnik GmbH, Minden, Germany). Raw data is stored every 5 min and automatically uploaded to a web-database for visualization, post-processing, and storage. Alarm messages (e.g., a low fuel gauge or system failure) are automatically triggered by SMS. A diesel backup-generator (with 5 kW peak), a Lithium-ion buffer battery (7.2 kWh), and two solar panels (with 1.7 kW peak) (Figure 2.2) supply the AC and DC power of the system and allow the system to support a continuous load of approximately 400 W.

Figure 2.2 Water Analysis Trailer for Environmental Research (WATER) (right), two solar panels supporting power supply (left) and the funnel-type tarp for collection of precipitation (further back in the middle) installed in the Schwingbach Environmental Observatory (SEO), Hesse (Germany).

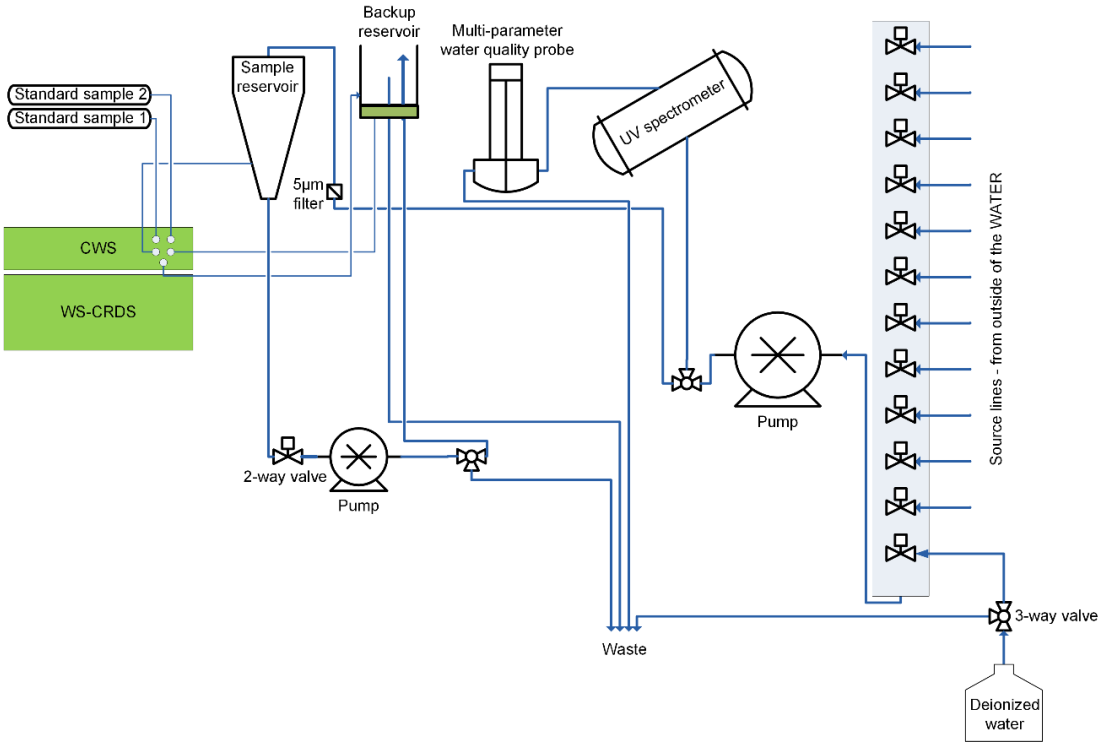

Figure 2.3 Schematic of the sampling board and analytical instrumentation inside the Water Analysis Trailer for Environmental Research (WATER). CWS = continuous water sampler, WS-CRDS = wavelength-scanned cavity ring-down spectrometer.

2.2.3 Sampling schedule

In the setup of this project, the WATER sampled automatically from two stream water sources (SW1 and SW2), three groundwater sources (GW1, GW2 and GW3), and one precipitation bucket (Figures 2.1c and 2.2). SW1 was sampled approximately 145 m upstream of the WATER at the edge of arable land and water from SW2 was taken next to the location of the WATER. Shallow groundwater was sampled from piezometers made from perforated PVC tubes sealed with bentonite clay at the upper part of the tube to prevent contamination by surface water. Piezometers of GW1 and GW2 were located at the toeslope on arable land and the meadow, respectively. The piezometer of GW3 was installed at the footslope at the edge of the forest.

Given the sampling resolution of 20 min, the setup of the system allowed the measurement of 72 samples per day. The sampling cycle was scheduled to sample n = 16 (with n the number of samples) for each stream water source, n = 8 for each groundwater source and n = 8 for each standard water sample in case of no precipitation event. In the case of a precipitation event, the sampler automatically took a precipitation sample whenever the sample volume exceeded 1 L (0.3 mm) inside the precipitation bucket. After sampling, the precipitation bucket was emptied to avoid carry-over effects, and the sampler was blocked for 60 min, allowing us to take samples from other sources.

2.2.4 Event definition and characteristics

We analysed autocorrelation for time series of precipitation depth measured over the sampling period to define independent precipitation events (Wenzel and Voorhees, 1981; Gaál et al., 2014). A plot of the correlation coefficient with respect to the lag time indicated the minimum lag time, for which the autocorrelation coefficient was close to zero. This lag time defined the minimum inter-event time, at which the precipitation events were considered independent. The autocorrelation plot for the precipitation depth at 5 min intervals (Figure 2.4) shows that correlation coefficients begin to level off close to zero at 5.7 h. We, therefore, considered precipitation events as independent if the inter-event time exceeded 6 h. In total, 20 events were selected, at which the application of hydrograph separation was possible due to the availability of isotope concentration data of precipitation, stream water, and groundwater. The beginning of an event was defined as the onset of precipitation and the end of an event as the time when $\frac{Q_{e_i}}{Q_{e_{max}}} \leq 0.05$ (i.e., when the fraction

of event water relative to its maximum value reached 5% or less) or when a new event began, whichever happened first.

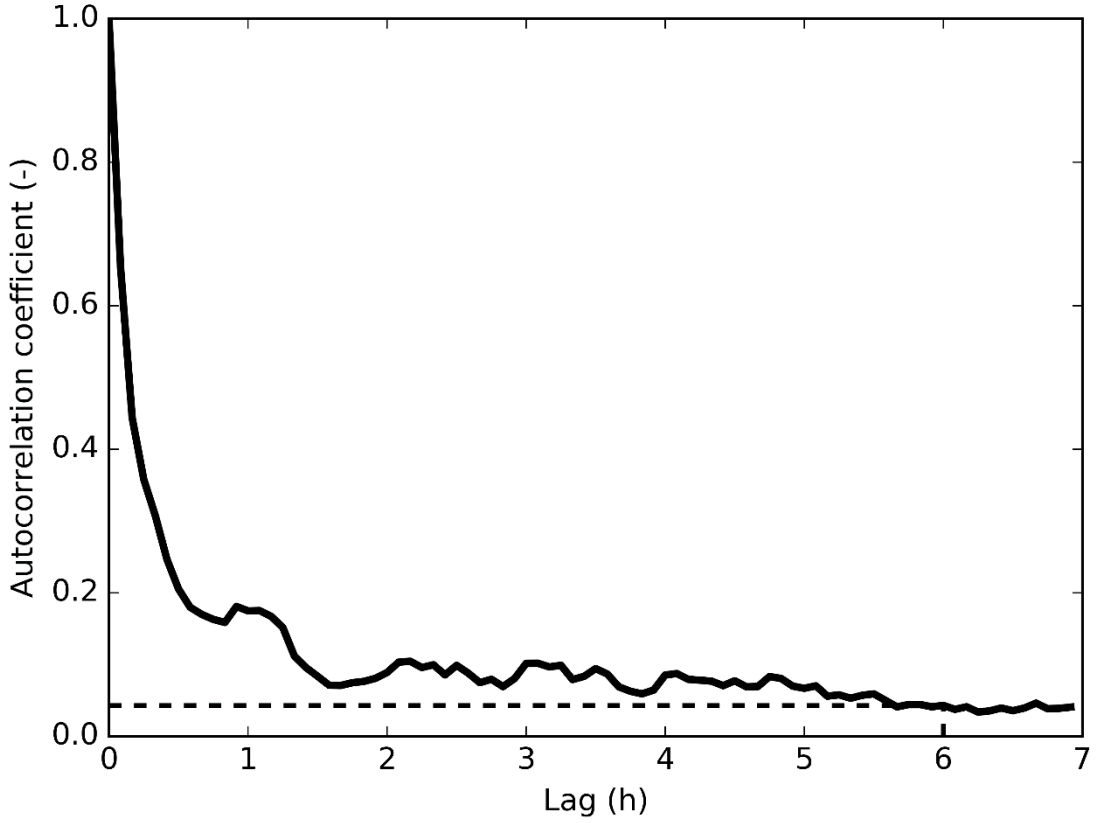

Figure 2.4 Autocorrelation plot for the precipitation depth at 5 min intervals. The precipitation events were considered independent if the inter-event time exceeded 6 h.

We calculated a number of hydrometrics to describe precipitation, antecedent wetness, and response characteristics of each event (Table 2.1). A Spearman rank correlation (ρ) analysis was used to assess the control of precipitation and antecedent wetness hydrometrics on response characteristics. The correlation values were considered statistically significant if $p < 0.05$. We described strength of the correlation strong if $\rho \geq 0.7$, moderate if $0.4 < \rho < 0.7$, and weak if $\rho \leq 0.4$.

Table 2.1 Hydrometrics to describe precipitation, antecedent wetness, discharge characteristics, maximum event water fraction, and response time for each event.

Metric	Description	Unit
Precipitation		
P	Total precipitation per event	mm
T	Precipitation duration	h
P_{int}	Mean precipitation intensity ($\frac{P}{T}$)	mm h^{-1}
Antecedent Wetness		
Q_{ini}	Initial discharge: 1 h averaged discharge before onset of precipitation	mm h^{-1}
$ST5_{ini}, ST30_{ini}, ST70_{ini}$	Initial soil moisture: 1 h averaged soil moisture before onset of precipitation at the toeslope (ST) at 5, 30 and 70 cm soil depths	%
$SF5_{ini}, SF30_{ini}, SF70_{ini}$	Initial soil moisture: 1 h averaged soil moisture before onset of precipitation at the footslope (SF) at 5, 30 and 70 cm soil depths	%
Discharge		
Q_{total}	Total discharge volume per event	mm
Q_e	Total event water volume	mm
Q_{pe}	Total pre-event water volume	mm
$\frac{Q_{total}}{P}$	Runoff coefficient	%
$\frac{Q_e}{Q_{total}}$	Event water fraction in total runoff	%
$\frac{Q_{pe}}{Q_{total}}$	Pre-event water fraction in total runoff	%
$F_{E_{max}}$	Maximum event water fraction	%
$T_{F_{E_{max}}}$	Response time: time lag between the first detection of precipitation and the maximum event water fraction	h

2.2.5 Response characteristics

A two-component hydrograph separation (Equation (2.1)) was applied to quantify event-water fractions F_E in stream water and groundwater sources using isotopic composition of stream water, groundwater, and precipitation:

$$F_E = \frac{C_{SW/GW} - C_{pe}}{C_e - C_{pe}} \tag{2.1}$$

$C_{SW/GW}$, C_e and C_{pe} are the isotopic concentrations in the sampling sources (i.e., stream water *SW* and groundwater *GW*), event water (*e*) (i.e., precipitation) and pre-event (*pe*) water of sampling sources, respectively. C_{pe} was calculated as the average of the last five samples before the onset of precipitation. C_e is the incremental weighted mean of precipitation samples (Mcdonnell et al., 1990). The maximum event water fractions $F_{E_{max}}$ in stream water and groundwater were derived for each event at the time steps that resulted from the sampling schedule. The uncertainty of the event water fractions W_{F_E} was quantified according to the Gaussian error propagation technique (Genereux, 1998) (Equation (2.2)):

$$W_{F_E} = \left\{ \left[\frac{C_e - C_{SW/GW}}{\left(C_e - C_{pe}\right)^2} W_{C_{pe}} \right]^2 + \left[\frac{C_{SW/GW} - C_{pe}}{\left(C_e - C_{pe}\right)^2} W_{C_e} \right]^2 + \left[\frac{-1}{\left(C_e - C_{pe}\right)} W_{C_{SW/GW}} \right]^2 \right\}^{\frac{1}{2}} \tag{2.2}$$

$W_{C_{pe}}$, the uncertainty in pre-event water was estimated using the standard deviation of the last five measurements before the onset of precipitation. W_{C_e}, the uncertainty of the event water is the standard deviation of the incremental weighted mean of precipitation. $W_{C_{SW/GW}}$, the uncertainty in the sampling sources is the measurement precision of the CWS coupled to the WS-CRDS (0.23‰ for $\delta^{18}O$ and 0.57‰ for $\delta^{2}H$), which we derived from the standard deviation of measurements during the last 3 min of the sampling period of stream water and groundwater sources. The uncertainty of event water fractions was higher using $\delta^{18}O$ measurements than those of $\delta^{2}H$. Therefore, we performed an analysis based on the $\delta^{2}H$ values only.

We estimated the response time of the maximum event water fraction $T_{F_{E_{max}}}$ in stream water and groundwater sources for each event. The response time was defined as the time lag between the first detection of precipitation and maximum event water fraction.

Total event water volume Q_e (Equation (2.3)) and total pre-event water volume Q_{pe} (Equation (2.4)) of each event were calculated from the event water fraction and discharge at each time step *i* (hourly) as:

$$Q_e = \sum F_{E_i} \times Q_i \tag{2.3}$$

$$Q_{pe} = \sum (Q_i - Q_{e_i}) \tag{2.4}$$

where F_{E_i}, Q_i and Q_{e_i} are the event water fraction in the stream water, the discharge recorded at the stream gauging site and the event water volume at time step i, respectively. In this study, we estimated Q_{e_i} only for SW2 as stream discharge was only recorded at this site. Event water fractions F_E were interpolated linearly with respect to time step i. F_E values were also interpolated linearly between the measurement gaps that happened due to system maintenance. Following, the event water fraction in total runoff $\frac{Q_e}{Q_{total}}$, pre-event water fraction in total runoff $\frac{Q_{pe}}{Q_{total}}$ and the runoff coefficient $\frac{Q_{total}}{P}$ were calculated for each event.

2.3 Results

2.3.1 Time series of isotopic and hydrometric observations

The dynamics of precipitation, discharge, soil moisture as well as the δ^2H of precipitation, stream water, and groundwater sources from August 8th to December 9th in 2018 are shown in Figure 2.5a–e. Over the sampling period, precipitation ranged between 0.01 and 3.71 mm h^{-1}, and discharge between 0.0002 and 0.004 mm h^{-1} with the highest peak at event #19. The discharge time series shows some distinct peaks, whereas diurnal variations in discharge resulted in a rather flickering behavior of the hydrograph. The discharge response pattern to precipitation during the sampling period partly matches the shallow soil moisture development at the toeslope and footslope. Measured soil moisture indicated generally wetter conditions at all three soil depths at

the toeslope compared to the one at the footslope. Further, the temporal dynamics at the toeslope were also more pronounced, particularly at 5 cm soil depth. At both sites, the lowest soil moistures were recorded in late summer at 5 cm soil depth.

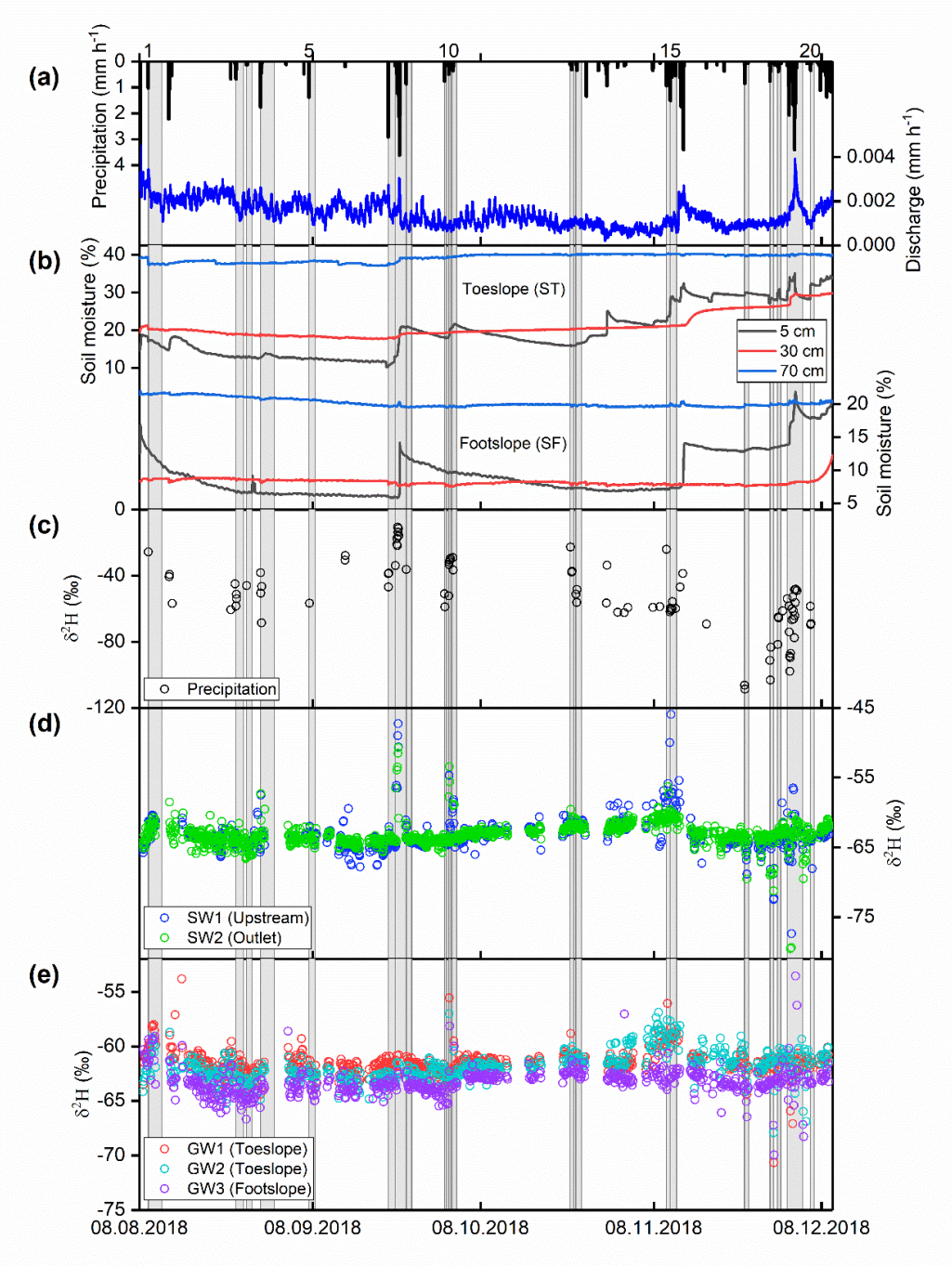

Figure 2.5 Time series of (a) precipitation and discharge, (b) soil moisture measured at 5, 30 and 70 cm depths at the toeslope and footslope, (c) δ^2H in precipitation, (d) δ^2H in stream water sources and (e) δ^2H in groundwater water sources. The vertical grey bars indicate the 20 events. The numbers on top of panel (a) represent the event ID.

The values of δ^2H in precipitation ranged from −10.7‰ to −108.6‰ with high within-storm event variation. From the beginning of the sampling period until early November, the values of δ^2H in precipitation were heavier than those in stream water and groundwater, whereas they became significantly lighter over the remaining sampling period. Isotopic concentrations in stream water and groundwater (i.e., $C_{SW/GW}$) as well as in precipitation (i.e., C_e) were heavier than those in pre-event water (i.e., C_{pe}) until event 15, whereas they became lighter thereafter. Over the entire sampling period, stream water and groundwater isotopes were almost stable. The mean values of δ^2H for the stream sampling positions SW1 and SW2 were −63.3 ± 2.1‰ (mean ± standard deviation) and −63.2 ± 1.7‰, respectively. On average, GW1 (−61.5 ± 1.2‰) and GW2 (−61.8 ± 1.3‰) at the toeslope depicted slightly heavier mean δ^2H values compared to the GW3 at the footslope (−63.3 ± 1.2‰). Distinct responses to precipitation were observed in stream water as well as groundwater during events #10, #17, and #19. However, different response behaviours were found during events #4 and #7, at which stream water isotopes showed clear responses to precipitation, whereas a slight reaction could be observed in the groundwater isotopes.

2.3.2 Hydrometrics

Hydrometrics of precipitation, antecedent wetness, and discharge characteristics, as well as maximum event water fraction, its uncertainty and response time of stream water and groundwater sources, were calculated for each of the 20 events (Tables S 2.1-2.3 in the Supplementary Materials). Figure 2.6a–f shows the variation of the total precipitation, mean precipitation intensity, initial soil moisture in shallower layers (i.e., 5 and 30 cm) at the toeslope and footslope, runoff coefficient, and event and pre-event water fraction in the total runoff for the 20 events. Total precipitation P ranged between 0.3 and 19.6 mm (3.2 ± 5.2 mm, mean ± standard deviation). The largest total precipitation was observed at event #19 followed by event #7 (Figure 2.6a). Mean precipitation intensity P_{int} ranged between 0.1 and 6.1 mm h^{-1} (1.4 ± 1.9 mm h^{-1}). Larger P_{int} was observed among the first nine events, with the largest value at event #1, followed by event #5 and #9 (Figure 2.6b).

Initial soil moisture at the toeslope position showed an increasing trend over the sampling events at both, 5 and 30 cm soil depths (Figure 2.6c). $ST5_{ini}$ values were lower than those in $ST30_{ini}$ until event #7, and subsequently reached slightly higher values over the last events. By contrast, the initial soil moisture at the footslope was quite stable with a slight decrease over the sampling events

at 30 cm depth, whereas an increasing trend with variation in between was observed at 5 cm depth (Figure 2.6d).

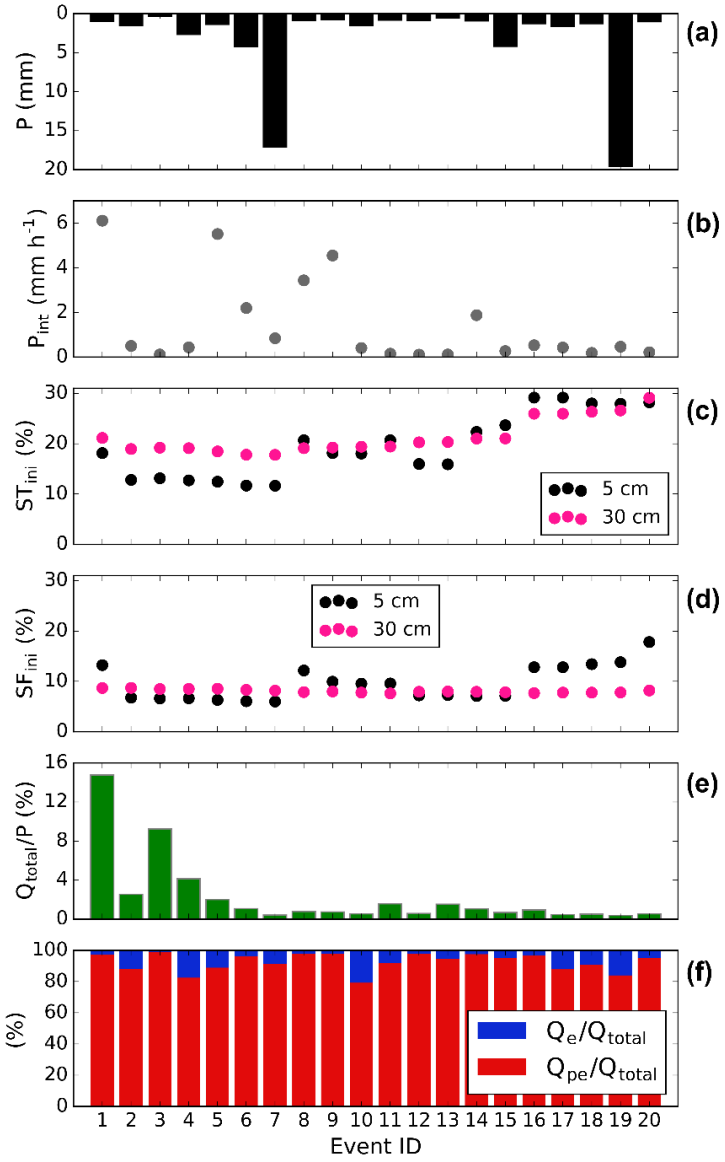

Figure 2.6 (a) Total precipitation per event $\boldsymbol{P}$, (b) mean rainfall intensity $\boldsymbol{P_{int}}$, (c) initial soil moisture at the toeslope $\boldsymbol{ST}$ in 5 and 30 cm depths, (d) initial soil moisture at the footslope $\boldsymbol{SF}$ in 5 and 30 cm depths, (e) runoff coefficient $\frac{\boldsymbol{Q_{total}}}{\boldsymbol{P}}$ and (f) event water fraction $\frac{\boldsymbol{Q_e}}{\boldsymbol{Q_{total}}}$ and pre-event water fraction $\frac{\boldsymbol{Q_{pe}}}{\boldsymbol{Q_{total}}}$ of the 20 precipitation events.

The runoff coefficient was relatively low, less than 2% for most of the events and with a mean of 2.2 ± 3.5% (Figure 2.6e). The highest runoff coefficient amounted to 14.8% at event #1, with a total precipitation input of only 1 mm but with a relatively high $SF5_{ini}$ of around 13%. The overall trend of runoff coefficient across the events showed that the events with relatively higher total precipitation did not necessarily result in higher rainfall-runoff ratios. Very low runoff coefficients (<1%) were observed even at events with the highest total precipitation (i.e., event #7 and #19), and distinctly different values for $SF5_{ini}$, $ST5_{ini}$, and $ST30_{ini}$.

The event water fraction in total runoff was also relatively low for the 20 events and varied from 1.4 to a maximum of 20.8% (7.6 ± 5.6%). Only three events exceeded a value of 15%, and on average, 92.4 ± 5.6% of the discharge contained pre-event water (Figure 2.6f). Lowest event water fractions (i.e., less than 3%) were observed for events with low total precipitation inputs (less than 1 mm), whereas highest event water fractions (i.e., more than 15%) were observed for events with the total precipitation higher than 1.5 mm.

Figure 2.7a,b shows the maximum event water fraction $F_{E_{max}}$ in stream water and groundwater across the 20 events. $F_{E_{max}}$ ranged from 3.3 to 74.5% (20.6 ± 17.2%) and from 3.1 to 86.3% (21.4 ± 18.95%) in SW1 and SW2, respectively. Despite a slightly higher $F_{E_{max}}$ at SW2 within most of the events, the difference between $F_{E_{max}}$ of SW1 and SW2 was insignificant for most of the events (3.4%, mean absolute difference). The highest $F_{E_{max}}$ in the stream water was recorded at event #19 with the highest total precipitation P among all events (19.6 mm). For the remaining events, $F_{E_{max}}$ was less than 50%. Those $F_{E_{max}}$ of stream water with higher values than the mean were observed at the events with P higher than 1.4 mm. At those events, in which $F_{E_{max}}$ in groundwater was higher than the mean values, $F_{E_{max}}$ in the stream water was also higher than the mean values (i.e., event #2, #10, #15, #18, and #19; only exception was event #17). The highest $F_{E_{max}}$ in GW1 and GW2 was observed at event #19, whereas the highest $F_{E_{max}}$ in GW3 was quantified for event #15. Differences between $F_{E_{max}}$ of groundwater sources were marginal within most of the events. However, groundwater at the toeslope (GW1 and GW2) reached on average slightly higher $F_{E_{max}}$ than the at the footslope (GW3).

Relatively large uncertainties of maximum event water fraction of 20% and more existed at events #15, #18, #19, and #20 (Figure 2.7a,b) due to the large temporal variation of isotope values in

precipitation and concurrent small differences between the isotopic composition of event and pre-event water. For the remaining events, the uncertainty was on average roughly 6% and 5% for stream water and groundwater, respectively.

The response time of maximum event water fraction $T_{F_{E_{max}}}$ ranged from 0.3 h to 24 h and 0.6 h to 33 h in stream water and groundwater, respectively (Figure 2.7c,d). On average, the event water fraction reached its maximum after 6 h in the stream water (i.e., SW1 and SW2). Differences between $T_{F_{E_{max}}}$ in the two stream water sources were small (1.3 h mean absolute difference), except at event #12, #17, and #20, at which the differences were larger than 5 h, 2 h, and 3 h, respectively. In comparison, the event water fractions in groundwater at the toeslope (GW1 and GW2) and the footslope (GW3) took a little longer to reach their maxima with 7–8 h. The longer response times in the stream water (i.e., at event #2, #7, #15, and #19) as well as in the groundwater (i.e., at event #2, #17, and #19) were observed at times when $F_{E_{max}}$ was above its mean value.

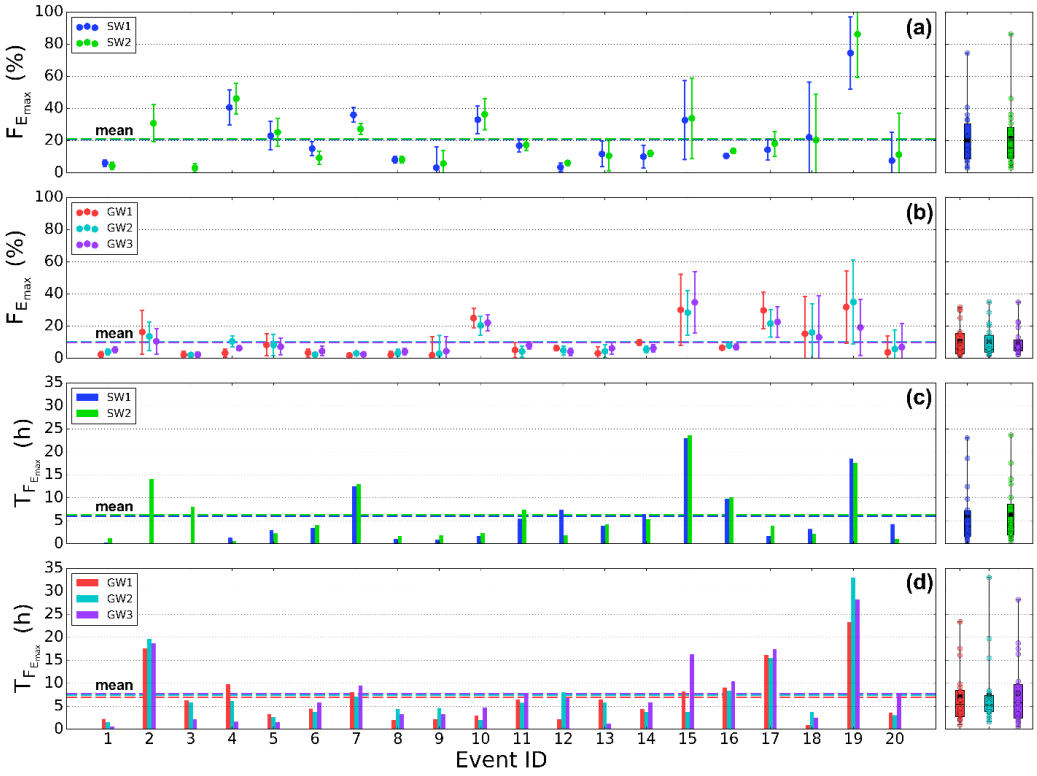

Figure 2.7 (a,b) Maximum event water fraction $\boldsymbol{F_{E_{max}}}$ for stream water (SW) and groundwater (GW) sources, including the uncertainty of estimations (error bars) and (c,d) response time of maximum event water fraction $\boldsymbol{T_{F_{E_{max}}}}$ for the same water sources. SW1 was not sampled at event #2 and #3. Boxplots (right panels) represent the interquartile range, black bars the median, black squares the mean and the whiskers the range of all 20 events.

2.3.3 Correlations of hydrometrics

In order to investigate the role of precipitation and antecedent wetness conditions on discharge characteristics, maximum event water fraction and response time, we tested their correlations (Table 2.2). Event water contribution increased significantly with the precipitation amount. In turn, total event water volume Q_e and event water fraction in total runoff $\frac{Q_e}{Q_{total}}$ correlated positively with precipitation amount P, whilst total pre-event water volume Q_{pe} was strongest related to the initial discharge Q_{ini}. No significant correlation was found for any of the soil moisture characteristics at the toeslope and discharge characteristics. By contrast, footslope soil moisture in deeper soil layers indicated a significant impact on total pre-event water contribution and the runoff coefficient.

Maximum event water fraction $F_{E_{max}}$ in the stream water increased with the amount and duration of precipitation. $F_{E_{max}}$ in groundwater was also related to the precipitation amount but with a less strong correlation. Particularly noteworthy was the importance of initial soil moisture conditions at the toeslope.

The response times $T_{F_{E_{max}}}$ of GW1 and GW3 correlated moderately with total precipitation similar to $T_{F_{E_{max}}}$ of SW1 and GW3 with precipitation duration and initial soil moisture of 70 cm depth at the toeslope (Table 2.2). No other significant correlations were established for any of the hydrometrics listed in Table 2.1. Nevertheless, we observed a clear two-state system with respect to the mean precipitation intensity P_{int} (Figure 2.8). In the first state, no relationship could be depicted between $T_{F_{E_{max}}}$ and P_{int}. After a certain threshold of P_{int}, the system swapped into a second state, at which $T_{F_{E_{max}}}$ decreased with increasing P_{int}. In order to identify the threshold, we calculated the Spearman rank correlation between $T_{F_{E_{max}}}$ and P_{int} by leaving out one event after another starting from the event with the lowest P_{int}. The P_{int} where we found the highest Spearman rank correlation coefficient was considered as the threshold. The Spearman rank correlation coefficients ranged between −0.99 and −0.88 ($p < 0.001$) for all water sources. The threshold values were 0.47 mm h^{-1} for SW1, SW2, GW2, and GW3 and 0.44 mm h^{-1} for GW1 (Figure 2.8).

Table 2.2 Spearman rank correlation coefficients (ρ) for the relationships of precipitation and antecedent wetness characteristics with discharge characteristic, maximum event water fraction and its response time (see Table 2.1 for acronyms). Green bolded cells represent statistically significant correlations with $p < 0.05$.

	P	T	P_{int}	Q_{ini}	$ST5_{ini}$	$ST30_{ini}$	$ST70_{ini}$	$SF5_{ini}$	$SF30_{ini}$	$SF70_{ini}$
Q_{total}	**0.54**	0.10	0.33	**0.69**	−0.39	−0.30	−0.37	−0.38	0.43	0.44
Q_e	**0.73**	0.29	0.23	**0.53**	−0.36	−0.25	−0.34	−0.25	0.28	0.33
Q_{pe}	**0.48**	0.05	0.37	**0.74**	−0.43	−0.36	−0.44	−0.41	**0.46**	**0.45**
$\frac{Q_{total}}{P}$	−0.39	**−0.53**	0.17	**0.49**	−0.42	−0.39	−0.32	−0.38	**0.59**	**0.48**
$\frac{Q_e}{Q_{total}}$	**0.64**	0.41	−0.06	0.05	−0.04	0.02	−0.04	0.04	−0.10	0.00
$\frac{Q_{pe}}{Q_{total}}$	**−0.64**	−0.41	0.06	−0.05	0.04	−0.02	0.04	−0.04	0.10	0.00
$F_{E_{max}}$ SW1	**0.75**	**0.58**	−0.12	0.25	−0.18	−0.17	−0.10	−0.29	−0.14	0.05
$F_{E_{max}}$ SW2	**0.75**	**0.56**	−0.06	−0.01	0.03	0.01	0.09	−0.07	−0.21	−0.09
$F_{E_{max}}$ GW1	**0.46**	0.33	−0.20	−0.27	**0.47**	**0.46**	**0.50**	0.26	−0.42	−0.18
$F_{E_{max}}$ GW2	**0.52**	**0.44**	−0.20	−0.21	**0.46**	**0.48**	**0.49**	0.34	−0.34	−0.07
$F_{E_{max}}$ GW3	**0.45**	0.30	−0.17	−0.28	**0.50**	**0.46**	**0.45**	0.36	−0.43	−0.23
$T_{F_{E_{max}}}$ SW1	0.32	**0.69**	−0.42	−0.27	0.18	0.24	**0.51**	−0.15	−0.33	−0.11
$T_{F_{E_{max}}}$ SW2	0.31	0.34	−0.13	−0.08	0.03	−0.03	0.13	−0.26	−0.23	−0.16
$T_{F_{E_{max}}}$ GW1	**0.53**	0.39	−0.11	0.11	0.03	0.02	0.01	−0.16	0.03	0.08
$T_{F_{E_{max}}}$ GW2	0.19	0.43	−0.26	−0.01	0.01	−0.03	−0.15	−0.04	−0.13	0.00
$T_{F_{E_{max}}}$ GW3	**0.53**	**0.45**	−0.13	−0.37	0.38	0.23	0.15	0.15	−0.38	−0.33

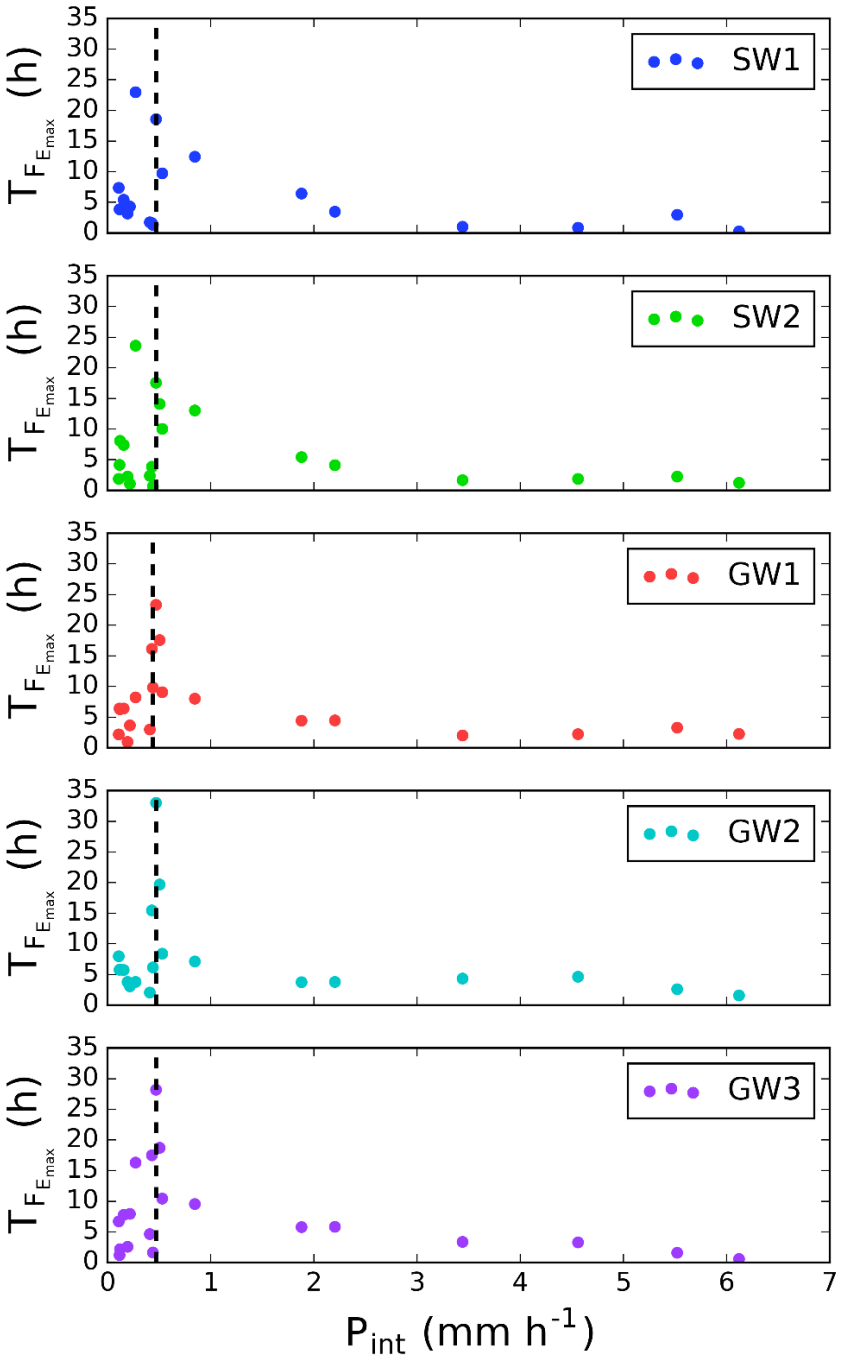

Figure 2.8 Response time $\boldsymbol{T_{F_{E_{max}}}}$ of stream water and groundwater sources versus mean precipitation intensity $\boldsymbol{P_{int}}$. The dashed line is the threshold, at which the response time decreased with increasing mean precipitation intensity. The threshold value was 0.47 mm h^{-1} for SW1, SW2, GW2, and GW3 and 0.44 mm h^{-1} for GW1.

2.4 Discussion

2.4.1 The Schwingbach is highly responsive to precipitation

Our findings show that the short-term response behavior observed in the previous studies in the SEO (Orlowski et al., 2014, 2016), is reflected in the isotopic signatures of stream water and groundwater. The isotopic responses in stream water and groundwater during the precipitation events often disclosed a rapid mixing of water in the sampling sources with event water indicating that the Schwingbach is a highly responsive catchment to incoming precipitation. The results reported that the average response time was 6 h in stream water and 7 h and 8 h for groundwater at the toeslope and footslope, respectively. Capturing event-based isotopic variation in stream water and groundwater emphasized the advantage of high-resolution sampling of multiple sources, giving insight into catchment response and runoff generation mechanisms. Firstly, the automated high-resolution sampling revealed the fine-scale hydrological responses of the stream water and groundwater, which could not be captured with manual isotope sampling in the previous study (Orlowski et al., 2016). Secondly, the high-resolution data constrained the interpretation of event water contribution derived from hydrograph separation by the continuous sampling of stream water. Thirdly, sampling of groundwater along the stream water, showed that the groundwater response times, particularly at the toeslope, were close to the stream water response times. This underlines the linkage between stream water and shallow subsurface flow and consequently, the vital role of shallow subsurface flow in runoff generation in the Schwingbach.

2.4.2 Pre-event water dominates runoff generation

Tracer-based hydrograph separations indicated that more than 79% of runoff consisted of pre-event water. Due to the highly permeable soils, the contribution of overland flow was negligible and hence precipitation infiltrated to the subsurface and mobilized pre-event water. Similar findings were reported in previous studies in other catchments, where pre-event water was also the major contributor to streamflow (Buttle, 1994; McGlynn and McDonnell, 2003; Jones et al., 2006; von Freyberg et al., 2018).

The runoff coefficients we found were very low (mean of 2.2%). Low runoff coefficients are mainly a result of high soil permeability and dry catchment wetness conditions (Merz and Blöschl, 2009). The high porosity of the soils in the upper layers of meadows and arable land (i.e., silty

sand), as well as the extremely dry conditions in summer 2018, were likely reasons for the low runoff coefficients we measured. Other studies also reported low event runoff coefficients of less than 2% (Blume et al., 2007; Scherrer et al., 2007). They found that the high porosity of the soil, lack of anthropogenic effects such as soil compaction and dry antecedent wetness conditions were the main reasons for low runoff coefficients.

2.4.3 Shallow subsurface flow pathways rapidly deliver water to the stream

The short response times of groundwater and stream water suggests the rapid movement of water vertically in the soil profile and in lateral downslope direction through shallow subsurface flow pathways. The short response times in the stream water (on average 6 h) as well as in the groundwater, particularly at the toeslope with on average 7 h, confirmed the linkage between stream water and shallow subsurface flow pathways. This linkage became likely stronger during events, at which maximum event water fractions in groundwater were above mean values such as event #10 and #19 (Figure 2.7a,b), and in turn, increased the event water fraction in total runoff (Figure 2.6f). Other studies also identified shallow subsurface as the dominant contributor to runoff generation (Brown et al., 1999; McGlynn and McDonnell, 2003; James and Roulet, 2009; von Freyberg et al., 2018). Possible flow pathways leading to fast response times in stream water and groundwater can be vertical and lateral macropore flow pathways through unsaturated or partially saturated soil matrix (Beven and Germann, 1982; McDonnell, 1990; Weiler and Naef, 2003; Jones, 2010). During the dry summer of 2018, an extended crack network developed that had the potential to act as a drainage system. Water transmission can be very high through this vertically-oriented continuous network of macropores, even higher than rainfall intensities (Weiler and Naef, 2003). We conclude this to have happened in our system in view of the short response times we found in the groundwater. Furthermore, the relatively flat areas at the toeslope and the stream banks likely contributed to the rapid transmission of water. These areas have the potential to quickly saturate, store water, and rapidly release water to the stream network even during small precipitation events (Sidle et al., 2000; Penna et al., 2011). This assumption was supported by the distinct response of shallow soil moisture to precipitation events at the toeslope, which was not clearly visible at the footslope.

2.4.4 Variable controls of runoff generation

Total event water volume and event water fraction in total runoff correlated positively with total event precipitation, pointing out that event water contribution increased with the precipitation amount (Table 2.2). Our results are consistent with previous studies, which also reported growing event water contribution with rising precipitation amounts (Pellerin et al., 2008; James and Roulet, 2009; Penna et al., 2015; von Freyberg et al., 2018). Higher precipitation amounts led to an increase of maximum event water fraction, total event water volume, and event water fraction in total runoff. This suggests that the maximum event water fraction can be considered as a qualitative proxy to indicate the overall contribution of event water to runoff. The positive correlation of the maximum event water fraction in groundwater sources with precipitation amount and initial soil moisture at the toeslope underlined that not only the higher precipitation amount led to an increase of event water contribution but also that wetter conditions at the toeslope facilitated groundwater recharge mechanisms to switch on.

Total pre-event volume correlated moderately with total precipitation amount (Table 2.2) suggesting that increasing precipitation led to the mobilization of pre-event water. Albeit the rising total pre-event water volume, the pre-event water fraction in total runoff decreased, indicating the gaining importance of event water contribution at the same time. Wetter antecedent conditions often lead to an increase of pre-event water contribution (Pellerin et al., 2008; James and Roulet, 2009; Muñoz-Villers and McDonnell, 2012; Penna et al., 2015; von Freyberg et al., 2018). Despite significant correlations of initial discharge and initial soil moisture of deeper layers at the footslope with total pre-event volume, we did not observe significant correlations between pre-event water fraction in total runoff and antecedent wetness hydrometrics in our system. Moreover, for catchments where pre-event water prevails in runoff generation, it is usually expected that antecedent wetness conditions show a major influence on the runoff coefficient. Although we detected significant correlations of the runoff coefficient with storm duration and antecedent wetness conditions in the Schwingbach, correlations were not strong and in the case of soil moisture conditions at the toeslope, they were even insignificant. This could be due to the dry and warm weather conditions during the study period, which resulted in only a small range of antecedent wetness conditions impacting the correlation analysis. Therefore, we conclude that pre-event water contribution was only moderately affected by precipitation amount, whereas under extremely dry

conditions such as in the year 2018, the antecedent wetness was too low to impact the pre-event water contribution.

With a closer look at the response time variation with respect to the mean precipitation intensity (Figure 2.8), we detected a threshold of mean precipitation intensity, at which the response time behavior changed. Whereas there was no clear pattern before the threshold, the response times in stream water and groundwater decreased with increasing mean precipitation intensity after the threshold. This suggests that an increase in the mean precipitation intensity after a certain precipitation sum enhanced infiltration and that water rapidly drained causing event water fraction to reach its peak faster. Surprisingly, this threshold was very low with only 0.5 mm h^{-1}, from where rainfall intensity likely controlled the initiation of macropore flow. It underlines that macropore networks governed flow control beyond the threshold. For mean precipitation intensities lower than the threshold, the trigger was not strong enough to control the response behavior.

2.5 Conclusions

Here we reported on high temporal resolution (20-min intervals) measurements of stable isotopes from groundwater, stream water, and precipitation to investigate the response of runoff components and the underlying controlling factors of precipitation and antecedent wetness characteristics. The analysis focused on 20 precipitation events in the Schwingbach Environmental Observatory (SEO) in Germany.

We conclude that high temporal resolution sampling of multiple sources, especially in systems with rapid hydrological responses, has a large potential to gain further insight into the runoff generation process. High temporal resolution sampling uncovered the event-based isotopic variation in stream water and groundwater and helped to constrain event and pre-event water contribution. The results of this investigation stress the importance of shallow subsurface flow contribution to the runoff in headwater systems. The sampling of groundwater in different hillslope positions and under different land uses increased the spatial resolution and led to better identify the main flow pathways during rapid delivery of water to the stream. We suggest that further studies be carried out over longer periods and that different land uses and slopes be taken into account in varying weather conditions.

Supplementary Materials

Table S 2. 1 Precipitation, antecedent wetness and discharge characteristics for the 20 events.

ID	Event date	P (mm)	T (h)	P_{int} (mm h^{-1})	Q_{ini} (mm h^{-1})	$ST5_{ini}$ (%)	$ST30_{ini}$ (%)	$ST70_{ini}$ (%)	$SF5_{ini}$ (%)	$SF30_{ini}$ (%)	$SF70_{ini}$ (%)	Q_{total} (mm)	Q_e (mm)	Q_{pe} (mm)	$\frac{Q_{total}}{P}$ (%)	$\frac{Q_e}{Q_{total}}$ (%)	$\frac{Q_{pe}}{Q_{total}}$ (%)
1	9 Aug	1.0	0.2	6.1	0.003	18.2	21.2	39.3	13.3	8.7	21.6	0.151	0.0044	0.146	14.8	3.0	97.0
2	25 Aug	1.5	3.0	0.5	0.002	12.8	19.0	37.6	6.8	8.7	21.2	0.038	0.0045	0.034	2.5	11.9	88.1
3	27 Aug	0.3	2.6	0.1	0.002	13.1	19.2	37.8	6.6	8.5	21.1	0.029	0.0004	0.028	9.3	1.4	98.6
4	29 Aug	2.7	6.1	0.4	0.002	12.7	19.1	37.6	6.6	8.5	20.8	0.110	0.0194	0.091	4.1	17.6	82.4
5	7 Sep	1.4	0.2	5.5	0.002	12.5	18.5	39.2	6.3	8.5	20.7	0.028	0.0031	0.024	2.0	11.1	88.9
6	21 Sep	4.2	1.9	2.2	0.002	11.7	17.8	37.2	6.1	8.3	19.7	0.045	0.0016	0.043	1.1	3.7	96.3
7	22 Sep	17.1	20.3	0.8	0.001	11.7	17.8	37.6	6.0	8.2	19.7	0.067	0.0059	0.061	0.4	8.8	91.2
8	24 Sep	0.9	0.3	3.4	0.001	20.7	19.1	39.0	12.2	7.9	19.6	0.007	0.0002	0.007	0.8	2.4	97.6
9	1 Oct	0.8	0.2	4.6	0.001	18.1	19.2	38.9	10.0	8.0	19.7	0.006	0.0001	0.005	0.7	2.4	97.6
10	2 Oct	1.5	3.7	0.4	0.001	18.0	19.4	39.5	9.5	7.8	19.5	0.008	0.0017	0.006	0.5	20.8	79.2
11	3 Oct	0.8	5.4	0.2	0.001	20.7	19.5	39.4	9.6	7.6	19.5	0.013	0.0011	0.012	1.6	8.1	91.9
12	24 Oct	0.9	7.8	0.1	0.001	16.0	20.3	39.8	7.2	7.9	19.9	0.005	0.0001	0.005	0.6	2.2	97.8
13	25 Oct	0.6	4.9	0.1	0.001	15.9	20.3	40.1	7.3	8.0	19.9	0.009	0.0005	0.008	1.5	5.5	94.5
14	10 Nov	0.9	0.5	1.9	0.001	22.3	21.0	40.0	7.1	7.9	19.7	0.010	0.0002	0.010	1.1	2.5	97.5
15	10 Nov	4.2	15.4	0.3	0.001	23.7	21.1	40.3	7.1	7.8	19.7	0.029	0.0014	0.028	0.7	4.9	95.1
16	24 Nov	1.3	2.4	0.5	0.001	29.2	25.9	39.9	12.8	7.7	19.5	0.012	0.0004	0.012	0.9	3.4	96.6
17	28 Nov	1.6	3.8	0.4	0.001	29.1	25.9	39.7	12.8	7.8	19.5	0.007	0.0009	0.006	0.5	12.2	87.8
18	30 Nov	1.3	6.7	0.2	0.001	28.0	26.3	39.9	13.4	7.8	19.9	0.006	0.0006	0.006	0.5	9.2	90.8
19	1 Dec	19.6	41.8	0.5	0.001	27.9	26.6	39.9	13.8	7.8	19.9	0.070	0.0113	0.059	0.4	16.0	84.0
20	6 Dec	1.0	4.8	0.2	0.001	28.2	29.1	40.0	17.8	8.2	20.0	0.006	0.0003	0.005	0.5	5.1	94.9

Table S 2. 2 Maximum event water fraction $F_{E_{max}}$, uncertainty of maximum event water fraction W_{F_E} and response time $T_{F_{E_{max}}}$ of stream water and groundwater sources derived from δ^2H values for the 20 events.

	$F_{E_{max}} \pm W_{F_E}$ (%)															$T_{F_{E_{max}}}$ (h)				
ID	SW1			SW2			GW1			GW2			GW3			SW1	SW2	GW1	GW2	GW3
1	6.2	±	2.0	4.5	±	2.2	2.4	±	1.8	4.1	±	1.9	5.4	±	1.8	0.3	1.2	2.3	1.6	0.6
2	-	±	-	30.9	±	11.6	16.3	±	13.7	13.7	±	9.0	10.7	±	8.0	-	14.1	17.6	19.7	18.7
3	-	±	-	3.1	±	2.5	2.3	±	2.2	2.1	±	1.3	2.4	±	1.5	-	8.1	6.3	5.8	2.2
4	40.7	±	10.9	46.2	±	9.6	3.3	±	2.6	10.5	±	3.5	6.3	±	1.5	1.3	0.7	9.8	6.2	1.6
5	23.1	±	8.9	25.3	±	8.5	8.4	±	6.8	8.3	±	6.7	7.3	±	5.3	3.0	2.3	3.3	2.6	1.6
6	15.1	±	4.4	9.4	±	4.0	3.5	±	2.5	2.4	±	1.2	4.6	±	3.0	3.5	4.1	4.5	3.8	5.8
7	36.0	±	4.5	27.3	±	3.4	1.8	±	1.4	3.1	±	1.5	2.6	±	1.3	12.4	13.0	8.0	7.1	9.6
8	8.1	±	2.0	8.3	±	2.1	2.3	±	2.1	3.4	±	2.2	4.3	±	2.1	1.0	1.7	2.0	4.4	3.4
9	3.3	±	12.8	5.8	±	8.2	2.0	±	11.5	2.8	±	11.4	4.5	±	8.9	0.9	1.9	2.2	4.6	3.3
10	33.1	±	8.6	36.4	±	9.6	25.1	±	5.9	20.5	±	5.9	22.2	±	4.9	1.7	2.4	3.0	2.1	4.7
11	16.9	±	4.1	17.4	±	3.5	5.3	±	4.8	4.3	±	3.3	8.1	±	2.4	5.4	7.4	6.4	5.8	7.8
12	3.5	±	2.5	6.2	±	1.5	6.3	±	1.7	5.0	±	2.7	4.1	±	2.3	7.4	1.9	2.2	8.0	6.7
13	11.8	±	7.9	10.7	±	9.4	3.1	±	4.0	4.4	±	4.0	6.3	±	3.5	3.9	4.2	6.4	5.8	1.2
14	10.2	±	7.0	12.1	±	1.9	9.8	±	1.8	5.6	±	2.1	6.3	±	2.2	6.5	5.4	4.4	3.8	5.8
15	32.8	±	24.5	34.0	±	25.0	30.3	±	22.0	28.4	±	13.9	34.9	±	19.2	23.0	23.6	8.3	3.8	16.3
16	10.6	±	1.4	13.6	±	1.5	6.6	±	1.4	7.9	±	1.4	7.1	±	1.7	9.8	10.1	9.1	8.4	10.4
17	14.5	±	6.4	18.2	±	7.7	29.8	±	11.5	21.8	±	8.5	22.6	±	9.4	1.6	3.9	16.1	15.5	17.5
18	22.1	±	34.3	20.4	±	28.3	15.3	±	23.1	16.1	±	17.7	13.2	±	25.9	3.2	2.3	1.0	3.8	2.6
19	74.5	±	22.6	86.3	±	27.0	31.9	±	22.4	35.0	±	25.9	19.2	±	17.3	18.6	17.6	23.3	33.0	28.2
20	7.6	±	17.5	11.4	±	25.6	3.8	±	10.2	5.8	±	11.9	7.0	±	14.8	4.3	1.1	3.7	3.1	8.0

Table S 2. 3 Maximum event water fraction $F_{E_{max}}$, uncertainty of maximum event water fraction W_{F_E} and response time $T_{F_{E_{max}}}$ of stream water and groundwater sources derived from $\delta^{18}O$ values for the 20 events.

	$F_{E_{max}} \pm W_{F_E}$ (%)														$T_{F_{E_{max}}}$ (h)				
ID	SW1			SW2			GW1			GW2			GW3			SW1	SW2	GW1	GW2	GW3
1	6.4	±	3.9	4.1	±	4.1	2.6	±	4.0	4.9	±	3.8	3.6	±	3.9	0.3	1.2	2.3	1.6	0.6
2	-	±	-	33.2	±	18.6	13.1	±	22.9	10.3	±	17.8	11.9	±	16.7	-	14.1	17.6	19.7	18.7
3	-	±	-	5.7	±	7.8	4.2	±	8.1	4.8	±	7.8	3.8	±	8.0	-	8.1	6.3	5.8	2.2
4	40.2	±	22.7	46.9	±	15.9	3.0	±	11.7	7.8	±	11.3	6.6	±	11.0	1.3	0.7	9.8	6.2	1.6
5	24.1	±	19.8	26.7	±	21.3	8.0	±	18.0	8.9	±	17.1	7.8	±	16.6	3.0	2.3	3.3	2.6	1.6
6	15.4	±	10.1	10.3	±	9.7	5.3	±	9.0	3.5	±	9.0	6.1	±	7.8	3.5	4.1	4.5	3.8	5.8
7	38.9	±	6.5	28.3	±	6.1	1.5	±	5.2	1.6	±	4.8	1.6	±	4.7	12.4	13.0	8.0	7.1	9.6
8	8.6	±	6.9	7.7	±	7.0	3.6	±	7.8	2.6	±	7.3	4.1	±	7.2	1.0	1.7	2.0	4.4	3.4
9	4.0	±	26.6	6.9	±	26.5	3.0	±	23.7	2.9	±	20.4	1.6	±	22.5	0.9	1.9	2.2	4.6	3.3
10	27.1	±	10.4	25.5	±	10.0	26.7	±	9.2	24.4	±	10.0	24.1	±	8.3	1.7	2.4	3.0	2.1	4.7
11	14.1	±	7.6	19.8	±	7.2	8.0	±	10.2	4.4	±	11.5	10.9	±	10.9	5.4	7.4	6.4	5.8	7.8
12	6.0	±	6.1	4.5	±	4.7	4.7	±	4.8	4.3	±	6.1	3.8	±	5.7	7.4	1.9	2.2	8.0	6.7
13	12.3	±	12.9	10.0	±	13.8	6.1	±	11.9	4.1	±	12.8	7.9	±	9.6	3.9	4.2	6.4	5.8	1.2
14	9.6	±	9.5	10.4	±	4.3	10.0	±	4.3	5.2	±	4.2	7.0	±	4.0	6.5	5.4	4.4	3.8	5.8
15	33.5	±	36.4	34.3	±	38.6	33.0	±	35.1	31.0	±	15.9	35.5	±	22.2	23.0	23.6	8.3	3.8	16.3
16	9.3	±	3.9	12.0	±	4.0	5.2	±	3.9	5.3	±	4.1	4.0	±	4.4	9.8	10.1	9.1	8.4	10.4
17	16.2	±	7.2	19.9	±	8.5	31.8	±	14.7	22.7	±	9.2	21.5	±	11.4	1.6	3.9	16.1	15.5	17.5
18	19.5	±	45.6	18.4	±	38.7	13.3	±	26.4	16.2	±	25.9	13.1	±	28.3	3.2	2.3	1.0	3.8	2.6
19	70.5	±	25.0	83.6	±	31.6	31.5	±	24.6	32.2	±	27.1	17.5	±	20.3	18.6	17.6	23.3	33.0	28.2
20	9.4	±	20.2	11.3	±	28.0	4.0	±	15.0	6.3	±	15.4	8.0	±	17.8	4.3	1.1	3.7	3.1	8.0

3 Application of machine learning models to predict maximum event water fraction in streamflow

This chapter is published in the journal *Frontiers in Water* as:

Sahraei, A.[1], Chamorro, A.[2], Kraft, P.[1], & Breuer, L[1,3]. (2021). Application of machine learning models to predict maximum event water fractions in streamflow. Frontiers in Water, 3, 52. https://doi.org/10.3389/frwa.2021.652100

[1]Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, 35392 Giessen, Germany

[2]Department of Ecosystems and Environment, Pontifical Catholic University of Chile, Santiago, Chile

[3]Centre for International Development and Environmental Research (ZEU), Justus Liebig University Giessen, Senckenbergstrasse 3, 35390 Giessen, Germany

Abstract

Estimating the maximum event water fraction, at which the event water contribution to streamflow reaches its peak value during a precipitation event, gives insight into runoff generation mechanisms and hydrological response characteristics of a catchment. Stable isotopes of water are ideal tracers for accurate estimation of maximum event water fractions using isotopic hydrograph separation techniques. However, sampling and measuring of stable isotopes of water is laborious, cost intensive, and often not conceivable under difficult spatiotemporal conditions. Therefore, there is a need for a proper predictive model to predict maximum event water fractions even at times when no direct sampling and measurements of stable isotopes of water are available. The behavior of maximum event water fraction at the event scale is highly dynamic and its relationships with the catchment drivers are complex and nonlinear. In last two decades, machine learning algorithms have become increasingly popular in the various branches of hydrology due to their ability to represent complex and nonlinear systems without any a priori assumption about the structure of the data and knowledge about the underlying physical processes. Despite advantages of machine learning, its potential in the field of isotope hydrology has rarely been investigated. Present study investigates the applicability of Artificial Neural Network (ANN) and Support Vector Machine (SVM) algorithms to predict maximum event water fractions in streamflow using precipitation, soil moisture and air temperature as a set of explanatory input features that are more straightforward and less expensive to measure compared to stable isotopes of water, in the Schwingbach

Environmental Observatory (SEO), Germany. The influence of hyperparameter configurations on the model performance and the comparison of prediction performance between optimized ANN and optimized SVM are further investigated in this study. The performances of the models are evaluated using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2) and Nash-Sutcliffe Efficiency (NSE). For the ANN, the results showed that an appropriate number of hidden nodes and a proper activation function enhanced the model performance, whereas changes of the learning rate did not have a major impact on the model performance. For the SVM, Polynomial kernel achieved the best performance, whereas Linear yielded the weakest performance among the kernel functions. The result showed that maximum event water fraction could be successfully predicted using only precipitation, soil moisture and air temperature. The optimized ANN showed a satisfactory prediction performance with MAE of 10.27%, RMSE of 12.91%, R^2 of 0.70, and NSE of 0.63. The optimized SVM was superior to that of ANN with MAE of 7.89%, RMSE of 9.43%, R^2 of 0.83, and NSE of 0.78. SVM could better capture the dynamics of maximum event water fractions across the events and the predictions were generally closer to the corresponding observed values. ANN tended to underestimate the events with high maximum event water fractions and to overestimate the events with low maximum event water fractions. Machine learning can prove to be a promising approach to predict variables that are not always possible to be estimated due to the lack of routine measurements.

3.1 Introduction

Estimating how runoff reacts to precipitation events is fundamental to understand the runoff generation mechanisms and hydrological response characteristics of a catchment (Klaus and McDonnell, 2013). For this, the total runoff can be partitioned into event and pre-event water components during a precipitation event. The event water is referred to the "new" water from incoming precipitation and pre-event water is referred to the "old" water stored in the catchment prior to the onset of precipitation. Stable isotopes of water ($\delta^{18}O$ and $\delta^{2}H$) are ideal tracers of these water components, since they are naturally occurring constituents of water, they preserve hydrological information and they characterize basin-scale hydrological responses (Stadnyk et al., 2013). Many studies have utilized stable isotopes of water in hydrograph separation techniques to differentiate the event and pre-event water during precipitation events (Kong and Pang, 2010; Klaus and McDonnell, 2013). Particularly in responsive catchments, the "maximum event water

fraction", at which the event water contribution to streamflow reaches its peak value during a precipitation event, can provide valuable information on the runoff generation mechanisms and response behavior for better selection and implementation of water management strategies. The maximum event water fraction (or vice versa, the minimum pre-event water fraction) mainly occurs at the peak streamflow and hence provides information to what extend the event and pre-event water components dominate the runoff during the peak streamflow (Buttle, 1994). This fraction can be used as a proxy to indicate overall contribution of event and pre-event water during a precipitation event and can be related to hydroclimatic variables such as precipitation, antecedent wetness conditions, land use or catchment size (Klaus and McDonnell, 2013; Fischer et al., 2017a). Many studies have used the maximum event water fraction to better comprehend runoff generation mechanisms and catchment hydrological responses to precipitation events (Sklash et al., 1976; Eshleman et al., 1993; Buttle, 1994; Genereux and Hooper, 1998; Brown et al., 1999; Hoeg et al., 2000; Renshaw et al., 2003; Carey and Quinton, 2005; Brielmann, 2008; Pellerin et al., 2008; Penna et al., 2015, 2016; von Freyberg et al., 2017; Fischer et al., 2017a,b; Sahraei et al., 2020).

The most accurate way to separate event water from pre-event water is to use isotopic hydrograph separation techniques (Klaus and McDonnell, 2013). If the two end-members (i.e., streamflow and precipitation) have a distinct difference in their isotopic signature, the stormflow hydrograph can be separated in their contributions based on a mass balance approach. Unlike the graphical techniques, isotopic hydrograph separation is measureable, objective and based on components of the water itself, rather than the pressure response in the channel (Klaus and McDonnell, 2013). Isotopic hydrograph separation techniques have been widely used in previous studies to accurately differentiate event and pre-event water components of flow in a wide range of climate, geology and land use conditions (Klaus and McDonnell, 2013). However, estimation of maximum event water fractions derived from isotopic hydrograph separation techniques is not always conceivable owing to the fact that sampling and measuring of stable isotopes of water is laborious, cost intensive, and often not feasible under difficult spatiotemporal conditions. Although the advent of laser spectroscopy technology has significantly reduced analytical costs of these isotopes (Lis et al., 2008), routine measurements are still far from being common. A proper predictive model based on a set of hydroclimatic variables would allow the estimation of maximum event water fractions in streamflow even at times when no direct sampling and measurements of stable isotopes of water are available.

In the last two decades, machine learning algorithms have been widely applied for efficient simulations of nonlinear systems and capturing noise complexity in respective datasets. The behavior of maximum event water fraction at the event scale is highly dynamic, which is derived from interaction of various catchment drivers such as precipitation, antecedent wetness conditions, topography, land use and catchment size (Klaus and McDonnell, 2013). The relationships between maximum event water fraction and the catchment drivers are complex and nonlinear. It is difficult to model these relationships accurately using process-based models. Process-based models are often limited by strict assumptions of normality, linearity, boundary conditions and variable independence (Tapoglou et al., 2014). One of the main advantages that has increased the application of machine learning algorithms is their ability to represent complex and nonlinear systems without any a priori assumption about the structure of the data and knowledge about the underlying physical processes (Liu and Lu, 2014). Nevertheless, machine learning is not a substitute for process-based models that provide valuable insights into the underlying physical processes (Windhorst et al., 2014; Kuppel et al., 2018; Zhang et al., 2019). In recent years, machine learning algorithms have been widely employed for efficient simulations of high dimensional and nonlinear relationships of various hydrological variables in surface and subsurface hydrology. They have been employed to predict streamflow (Wu and Chau, 2010; Rasouli et al., 2012; Senthil Kumar et al., 2013; He et al., 2014; Shortridge et al., 2016; Abdollahi et al., 2017; Singh et al., 2018; Yuan et al., 2018; Adnan et al., 2019b; Duan et al., 2020; Adnan et al., 2021b), groundwater and lake water level (Yoon et al., 2011; Tapoglou et al., 2014; Li et al., 2016; Sahoo et al., 2017; Sattari et al., 2018; Malekzadeh et al., 2019; Sahu et al., 2020; Yaseen et al., 2020; Kardan Moghaddam et al., 2021), water quality parameters such as nitrogen, phosphorus and dissolved oxygen (Chen et al., 2010; Singh et al., 2011; Liu and Lu, 2014; Kisi and Parmar, 2016; Granata et al., 2017; Sajedi-Hosseini et al., 2018; Najah Ahmed et al., 2019; Knoll et al., 2020), soil hydraulic conductivity (Agyare et al., 2007; Das et al., 2012; Elbisy, 2015; Sihag, 2018; Araya and Ghezzehei, 2019; Adnan et al., 2021a), soil moisture (Gill et al., 2006; Ahmad et al., 2010; Coopersmith et al., 2014; Matei et al., 2017; Adeyemi et al., 2018), water temperature in rivers (Piotrowski et al., 2015; Zhu et al., 2018, 2019; Qiu et al., 2020; Quan et al., 2020) and many other hydrological variables. Comprehensive reviews have been published for the application of machine learning algorithms in hydrology and earth system science (Lange and Sippel, 2020; Zounemat-Kermani et al., 2020).

Artificial Neural Network (ANN) is an extremely popular algorithm for the prediction of water resources variables (Maier and Dandy, 2000; Maier et al., 2010). ANN is a parallel-distributed information processing system, which has been basically inspired by the biological neural network, consisting of a myriad of interconnected neurons in the human brain (Haykin, 1994). One of the main advantages of ANN models is that they are universal function approximators (Khashei and Bijari, 2010). This means that they can automatically approximate a large class of functions with a high degree of accuracy. Their power comes from the parallel processing of the information from the data. Furthermore, ANN is developed through learning rather than programming. This means that no prior specification of fitting function is required in the model building process. Instead, the network function is determined by the patterns and characteristics of the data through the learning process (Khashei and Bijari, 2010). Another main advantage of ANNs is that they possess an inherent generalization ability. This means that they can identify and respond to patterns that are similar, but not identical to the ones with which they have been trained (Benardos and Vosniakos, 2007). Minns and Hall (1996) successfully applied ANNs in earlier years to simulate streamflow responses to precipitation inputs. Chen et al. (2010) used an ANN with backpropagation training algorithm to forecast concentrations of total nitrogen, total phosphorus and dissolved oxygen in response to agricultural non-point source pollution for any month and location in Changle River, southeast China. They concluded that ANN is an easy-to-use modelling tool for managers to obtain rapid preliminary identification of spatiotemporal water quality variations in response to natural and artificial modifications of an agricultural drainage river. Wu and Chau (2010) compared the performance of ANN with that of Auto-Regressive Moving Average (ARMA) and K-Nearest-Neighbors (KNN) for predication of monthly streamflow. The result revealed that ANN outperformed the two other models. Coopersmith et al. (2014) investigated the application of ANN, classification trees and KNN in order to simulate the soil drying and wetting process at a site located in Urbana, Illinois, the USA. They reported that reasonably accurate predictions of soil conditions are possible with only precipitation and potential evaporation data using ANN model. Tapoglou et al. (2014) tested the performance of an ANN model using cross-validation technique to predict dynamic of groundwater level in Bavaria, Germany. They concluded that ANN can be successfully employed in aquifers where geological characteristics are obscure, but variety of other, easily accessible data, such as meteorological data can be easily found. Piotrowski et al. (2015) examined different types of ANNs such as Multilayer Perceptron (MLP), product-units, Adaptive Neuro Fuzzy Inference System (ANFIS) and wavelet neural networks to predict water temperature in

rivers based on various meteorological and hydrological variables. The result showed that simple MLP neural networks were in most cases not outperformed by more complex and advanced models. Sihag (2018) successfully applied ANN to simulate unsaturated hydraulic conductivity. Yuan et al. (2018) utilized the long short-term memory (LSTM) neural network for monthly streamflow prediction in comparison with backpropagation and Radial Basis Neural Network (RBNN). They found that the LSTM neural network performed better than the other ANN models. Sahu et al. (2020) analysed the accuracy and variability of groundwater level predictions obtained from a MLP model with optimized hyperparameters for different amounts and types of available training data. Yaseen et al. (2020) developed a MLP with Whale optimization algorithm for Lake water level forecasting. The result indicated that the proposed model was superior to other comparable models such as Cascade-Correlation Neural Network Model (CCNNM), Self-Organizing Map (SOM), Decision Tree Regression (DTR) and Random Forest Regression (RFR). Comprehensive reviews have been published for the application of ANN (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000; Maier and Dandy, 2000; Dawson and Wilby, 2001; Maier et al., 2010) in the field of hydrology.

With increased ANN model research, the limitations of ANN have been highlighted, such as local optimal solutions and gradient disappearance, which limit the application of the model (Yang et al., 2017; Zhang et al., 2018a). Therefore, SVMs were introduced (Cortes and Vapnik, 1995) as a relatively new structure in the field of machine learning. The main advantage of SVM is that it is based on the structural risk minimization principle that aims to minimize an upper bound to the generalization error instead of the training error, from which SVM is able to achieve good generalization capability (Liu and Lu, 2014). Furthermore, for the implementation of SVM, a convex quadratic constrained optimization problem must be solved so that its solution is always unique and globally optimal (Schölkopf et al., 2002). Another main advantage of SVM is that it automatically identifies and incorporates support vectors during the training process and prevents the influence of the non-support vectors over the model. This causes the model to cope well with noisy conditions (Han et al., 2007). Lin et al. (2006) investigated the potential of SVM in comparison with ANN and ARMA models to predict streamflow. They found that prediction performance of SVM was superior to that of ANN and ARMA models. Ahmad et al. (2010) explored application of SVM to simulate soil moisture in the Lower Colorado River Basin, the USA. Results from the SVM were compared with the estimates obtained from ANN and

Multivariate Linear Regression model (MLR); and showed that SVM model performed better for soil moisture estimation than ANN and MLR models. Behzad et al. (2010) optimized SVM using cross-validation technique and compared its performance to that of ANN in order to predict transient groundwater levels in a complex groundwater system under variable pumping and weather conditions. It was found that even though modelling performance for both approaches was generally comparable, SVM outperformed ANN particularly for longer prediction horizons when fewer data events were available for model development. They concluded that SVM has the potential to be a useful and practical tool for cases where less measured data are available for future prediction. Yoon et al. (2011) successfully applied SVM to forecast groundwater level fluctuations at a coastal aquifer in Korea. Das et al. (2012) examined the application of SVM and ANN for prediction of field hydraulic conductivity of clay liners. The reported that the developed SVM model was more efficient compared with the ANN model. Jain (2012) found that SVM is also suitable to model the relationship between the river stage, discharge, and sediment concentration. Liu and Lu (2014) compared the optimized SVM with optimized ANN for forecasting total nitrogen and total phosphorus concentrations for any location of the river polluted by agricultural pollution in eastern China. The results revealed that even for the small sample size, SVM model still achieves good prediction performance. He et al. (2014) investigated the potential of ANN, ANFIS and SVM models to forecast river flow in the semiarid mountain region, northwestern China. A detailed comparison of the overall performance indicated that the SVM performed better than ANN and ANFIS in river flow forecasting. The results also suggest that ANN, ANFIS and SVM method can be successfully applied to establish river flow with complicated topography forecasting models in the semiarid mountain regions. Mohammadpour et al. (2015) employed SVM, ANN and RBNN to predict the water quality index. The research highlights that the SVM and ANN can be successfully applied for the prediction of water quality in a free surface constructed wetland environment. Adnan et al. (2019a) compared the prediction accuracy of Optimally Pruned Extreme Learning Machine (OP-ELM), Least Square Support Vector Machine (LSSVM), Multivariate Adaptive Regression Splines (MARS) and M5 model Tree (M5Tree) in modelling monthly streamflow using precipitation and temperature inputs. The test results showed that the LSSVM and MARS-based models provide more accurate prediction results compared to OP-ELM and M5Tree models. Malik et al. (2020) optimized SVM by six meta-heuristic algorithms to predict daily streamflow in Naula watershed, India. They reported that SVM optimized by Harris Hawks optimization algorithm showed superior performance to the other optimized SVM models.

Quan et al. (2020) proposed SVM for predicting water temperature in a reservoir in western China. They found that the prediction performance of optimized SVM was superior to that of ANN model. Comprehensive reviews have been published for the application of SVM (Raghavendra and Deka, 2014) in the field of hydrology.

So far, the potential of machine learning in the field of isotope hydrology has rarely been explored. Cerar et al. (2018) compared the performance of MLP with that of ordinary kriging, simple and multiple linear regression to predict isotope composition of oxygen ($\delta^{18}O$) in groundwater in Slovenia. Based on validation data sets, the MLP model proved to be the most suitable method for predicting $\delta^{18}O$ in the groundwater. However, they did not optimize the employed MLP model to enhance the prediction performance. Most of the machine learning algorithms have several settings that govern the entire training process (Goodfellow et al., 2016). These settings are called "hyperparameters". The hyperparameters are external to the model that are not learned from the data and must be set prior to the training process (Shahinfar and Kahn, 2018; Géron, 2019). The performance and computational complexity of ANN and SVM models are heavily dependent on configuration of their hyperparameters (Smithson et al., 2016; Zhang et al., 2018a). Hence, it is necessary to optimize the hyperparameters in order to enhance the model performance. However, investigations of different hyperparameters and comparison of their influence on the model performance are rarely reported in the field of hydrology.

To the best of our knowledge, there is no study that applies machine learning algorithms for prediction of maximum event water fractions at the event scale. The novelty of this research is: (1) to investigate for the first time the applicability of two widely used machine learning algorithms, namely ANN and SVM, to predict maximum event water fractions in streamflow on independent precipitation events using precipitation, soil moisture and air temperature as a set of explanatory input features that are more straightforward and less expensive to measure compared to stable isotopes of water; (2) to summarize the influence of hyperparameter configurations on model performance; and (3) to compare the prediction performance of optimized ANN with optimized SVM.

3.2 Materials and Methods

3.2.1 Study area and data set

The study area is located in the headwater catchment (Area = 1.03 km^2) of the Schwingbach Environmental Observatory (SEO) (50°28'40"N, 8°32'40"E) in Hesse, Germany (Figure 3.1). The elevation ranges from 310 m in the north to 415 m a.s.l. in the south. The climate is classified as temperate oceanic, with a mean annual air temperature of 9.6°C and mean annual precipitation of 618 mm obtained from the German Weather Service (Deutscher Wetterdienst, Giessen-Wettenberg station, period 1969-2019). The catchment area is dominated by 76% of forest that is mainly located in east and south, 15% of arable land in the north and west and 7% of meadows along the stream. The soil is classified as Cambisol mainly covered by forest and Stagnosols under arable land. The soil texture is dominated by silt and fine sand with low clay content. For more details refer to Orlowski et al. (2016).

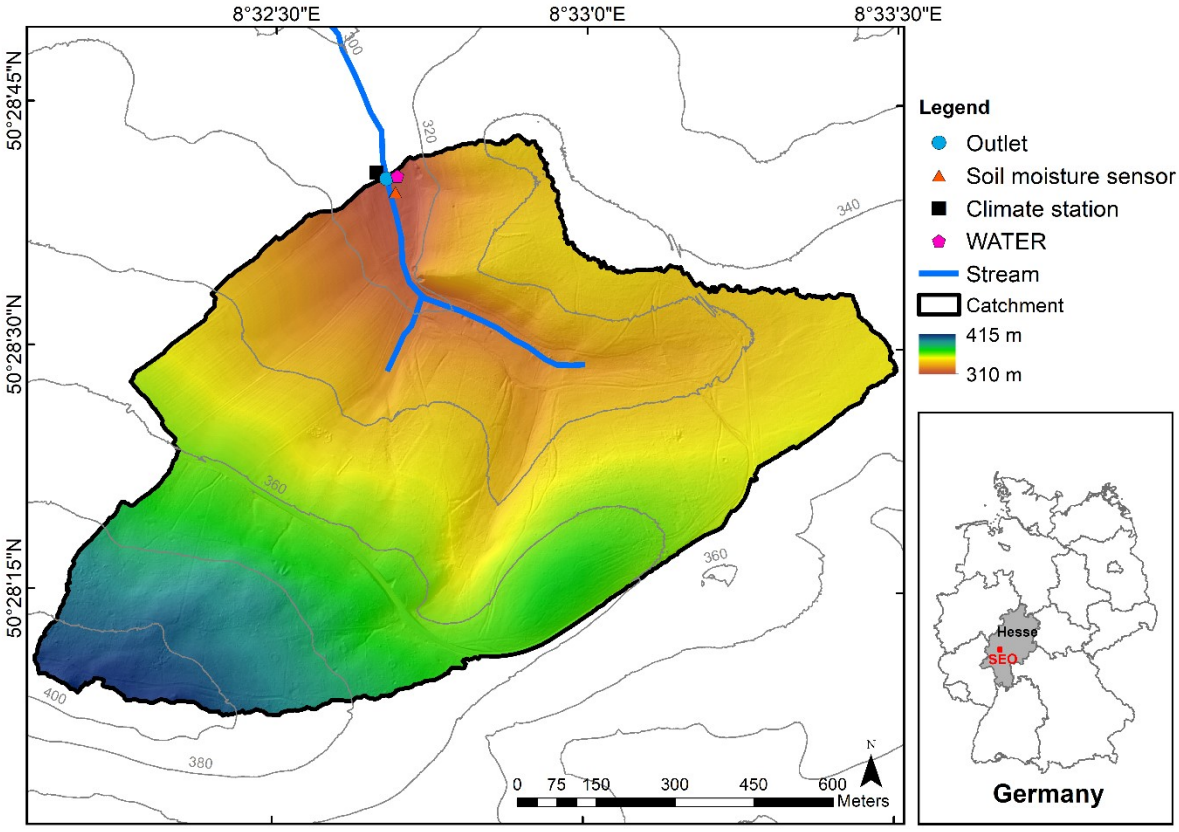

Figure 3.1 Location of the Schwingbach Environmental Observatory (SEO) in Hesse, Germany (red square, right panel) and a map of the study area in the headwater catchment of the SEO including the measuring network.

Precipitation depth and air temperature were recorded at 5 min intervals with an automated climate station (AQ5, Campbell Scientific Inc., Shepshed, UK) operated with a CR1000 data logger. Soil moisture was measured at 5, 30 and 70 cm depths at the toeslope at 5 min intervals with a remote telemetry logger (A753, Adcon, Klosterneuburg, Austria) equipped with capacitance sensors (ECH2O 5TE, METER Environment, Pullman, USA). Stream water at the outlet of the catchment and precipitation were automatically sampled for stable isotopes of water in situ using an automated mobile laboratory, the Water Analysis Trailer for Environmental Research (WATER), from August 8th until December 9th in 2018 and from April 12th until October 10th in 2019. If no precipitation occurred, stream water was sampled approximately every 90 min. Precipitation was automatically sampled if its depth exceeded 0.3 mm. It should be noticed that the SEO is not a snow-driven catchment. The precipitation events were all in form of rainfall with no snow included over the sampling period. After sampling, the precipitation sampler was blocked for 60 min. Isotopic composition of stream water and precipitation were measured using a continuous water sampler (CWS) (A0217, Picarro Inc., Santa Clara, USA), coupled to a wavelength-scanned cavity ring-down spectrometer (WS-CRDS) (L2130-i, Picarro Inc., Santa Clara, USA) inside the mobile laboratory. For detailed description of the WATER and the sampling procedures see Sahraei et al. (2020). The SEO was chosen in this study because of its responsive characteristic to precipitation inputs as well as the available extensive hydrological knowledge and data on the catchment (Lauer et al., 2013; Orlowski et al., 2014, 2016; Aubert et al., 2016; Sahraei et al., 2020). The runoff generation processes of the Schwingbach are characterized by a rapid mixing of streamflow with event water through shallow subsurface flow pathways. The high-resolution sampling of isotope concentrations and hydroclimatic data employed in this study allowed the detection of fine-scale, short-term responses and mixing processes, which in turn provided the opportunity to capture a wide range of maximum event water fractions under different flow regimes and hydroclimatic conditions. The mean, standard deviation, Kurtosis and Skewness of the maximum event water fraction data are 32.4%, 24.2%, -0.4 and 0.8, respectively. These statistical values imply that the maximum event water fractions are highly dynamic in the SEO, which in turn provides the opportunity to test the capability and effectiveness of the ANN and SVM models for the prediction of maximum event water fractions at the event scale.

In this work, the precipitation events were defined as independent if the inter-event time exceeded 6 h, similar to the separation made by Sahraei et al. (2020). The dry period between two wet periods

is known as the inter-event time, and if the dry period is equal to or longer than the desired inter-event time, the two wet periods are considered as two independent events (Balistrocchi and Bacchi, 2011; Chin et al., 2016). In total, 40 events were selected over the sampling period (Figure 3.2), for which the application of isotopic hydrograph separation was possible. The isotopic hydrograph separation was not possible for other events because either the isotopic compositions of stream water and precipitation were not available or because the difference between the isotopic composition of stream water and precipitation was too small. For the selected events, the beginning of the event was defined as the onset of precipitation and the end of the event as the time when either the event water fraction declined to 5% of its peak value or another precipitation event began, whichever first occurred.

The event water fractions F_E was quantified for each event at approximately 90 min intervals using a two-component isotopic hydrograph separation according to Equation (3.1):

$$F_E = \frac{C_S - C_P}{C_E - C_P} \tag{3.1}$$

where C_S, C_E and C_P are the isotopic concentrations (δ^2H) in the stream water, event water (i.e., precipitation) and pre-event water, respectively. C_S was measured at approximately 90 min intervals. C_E is the incremental weighted mean of precipitation samples (Mcdonnell et al., 1990), which were measured if the precipitation depth exceeded 0.3 mm. After sampling, the precipitation sampler was blocked for 60 min. C_P was calculated as the average of the last five samples, which were measured at approximately 90 min intervals before the onset of precipitation. The maximum event water fraction $F_{E_{max}}$, at which the event water contribution to streamflow reaches its peak value during the precipitation event, was obtained from maximum of F_E values over each of the 40 events. The $F_{E_{max}}$ was the target variable predicted by the ANN and SVM models. Sahraei et al. (2020) reported that higher precipitation amounts generally lead to an increase of the maximum event water fraction in the streamflow as can be observed in Figure 3.3. This behavior is consistent with previous studies, which also reported growing maximum event water fractions with rising precipitation amounts (Pellerin et al., 2008; Penna et al., 2015; von Freyberg et al., 2017). This suggests that the Schwingbach is highly responsive to precipitation inputs and precipitation is a main driver of hydrological response behavior in the stream.

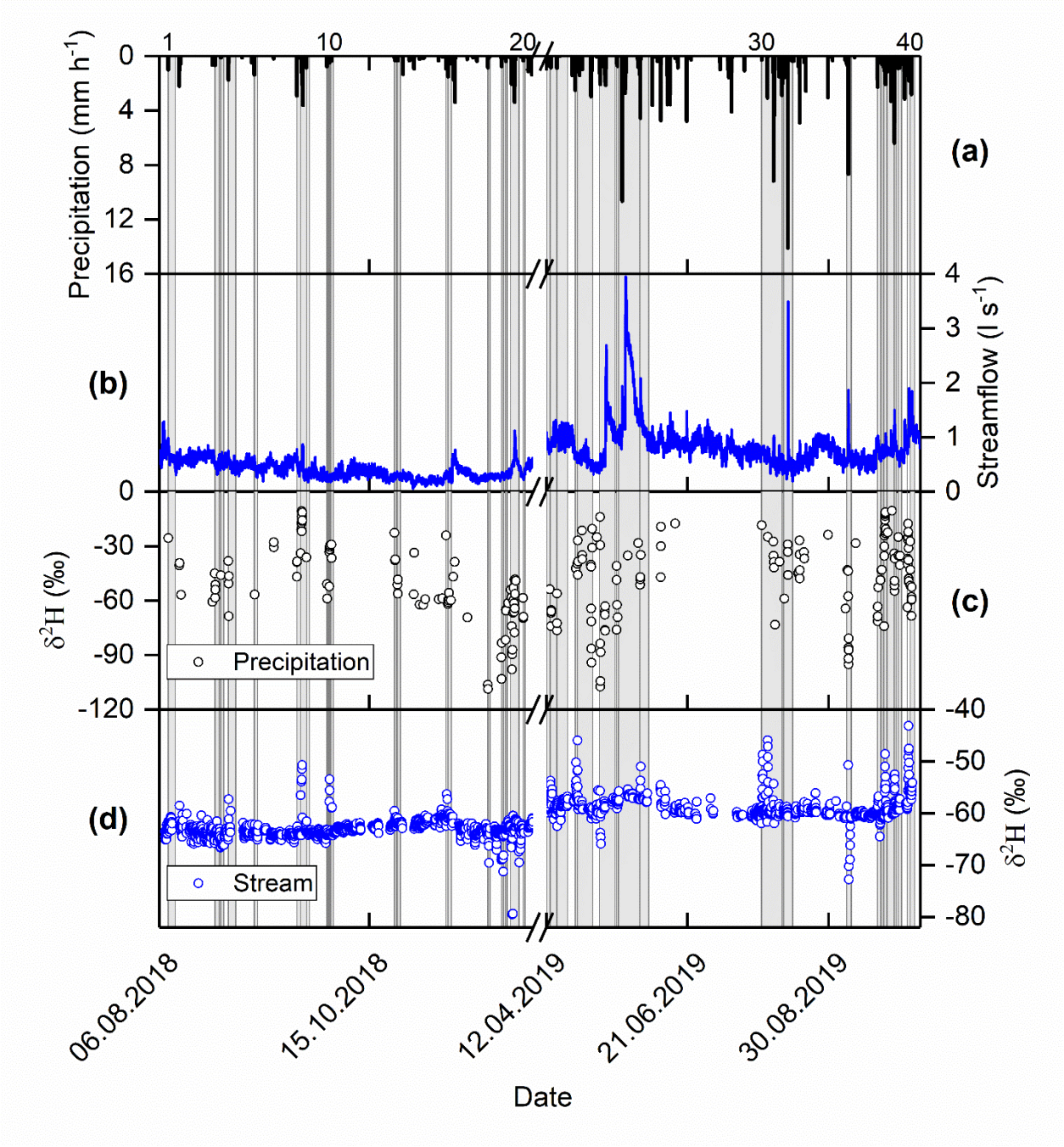

Figure 3.2 Time series of (a) precipitation, (b) streamflow, (c) δ^2H in precipitation and (d) δ^2H in streamflow. The vertical grey bars indicate the 40 events. The numbers on top of panel (a) represent the event ID.

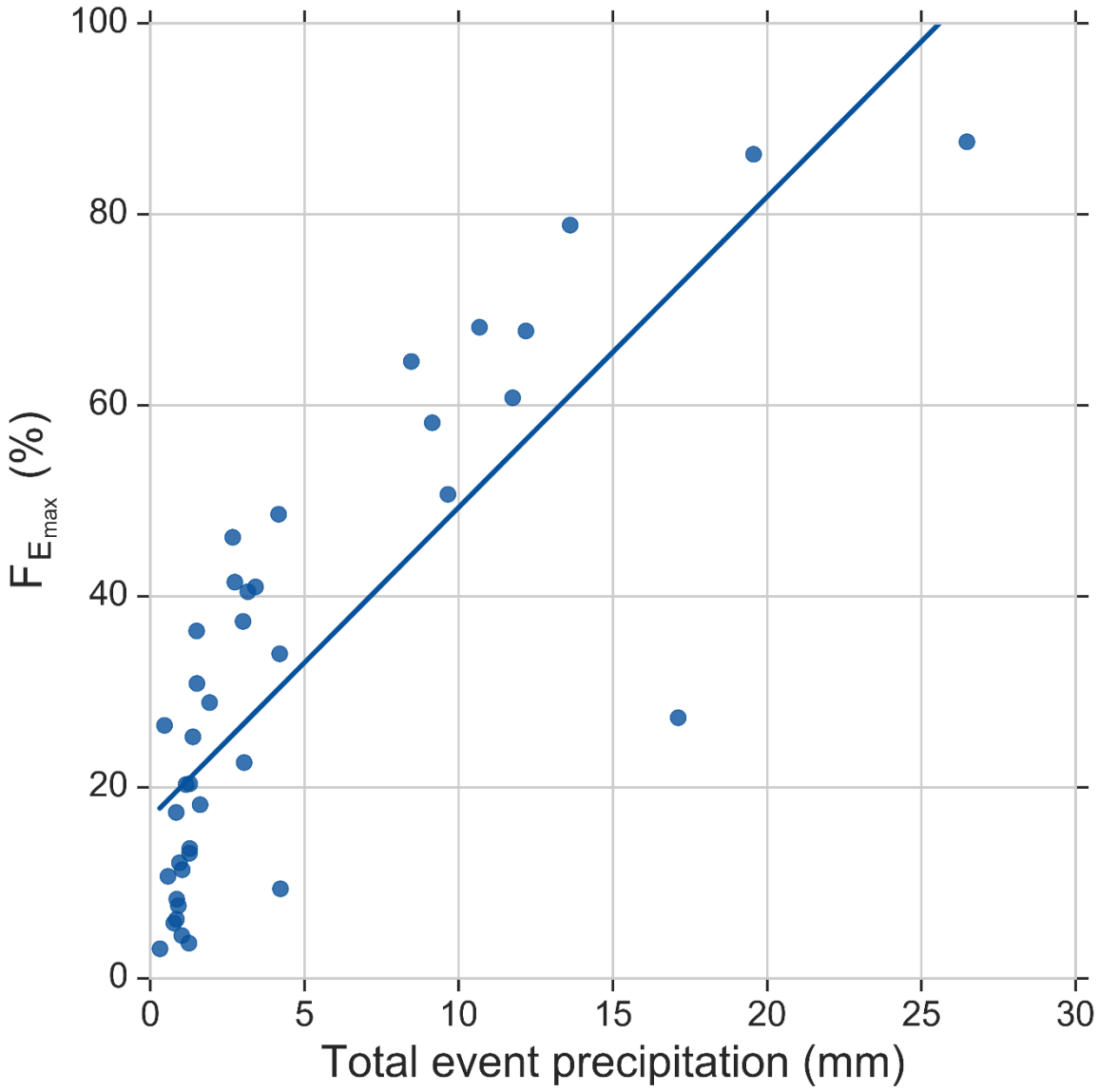

Figure 3.3 Maximum event water fraction $\boldsymbol{F_{E_{max}}}$ vs. total event precipitation.

The uncertainty of maximum event water fraction $W_{F_{E_{max}}}$ was quantified according to the Gaussian error propagation technique (Genereux, 1998):

$$W_{F_{E_{max}}} = \left\{ \left[\frac{C_E - C_S}{(C_E - C_P)^2} W_{C_P} \right]^2 + \left[\frac{C_S - C_P}{(C_E - C_P)^2} W_{C_E} \right]^2 + \left[\frac{-1}{(C_E - C_P)} W_{C_S} \right]^2 \right\}^{\frac{1}{2}} \quad (3.2)$$

W_{C_P}, the uncertainty in pre-event water, was estimated using the standard deviation of the last five measurements before the onset of precipitation. W_{C_E}, the uncertainty of the event water, is the standard deviation of the incremental weighted mean of precipitation. W_{C_S}, the uncertainty of the stream water measurements, is the measurement precision of the CWS coupled to the WS-CRDS (0.57‰ for δ^2H), which we derived from the standard deviation of the measurements during the last 3 min of the sampling period of stream water. The uncertainty of maximum event water fractions ranged from 1.2% to 28.3% with a mean of 6.7 ± 7.2% (mean ± standard deviation).The quantified uncertainty for maximum event water fractions of all 40 events are reported in Supplementary Table S 3.1.

As the inputs for the models, a set of hydroclimatic variables representing the precipitation and antecedent wetness characteristics of the Schwingbach catchment was selected. Total event precipitation (P, mm), precipitation duration (D, h), 3-day antecedent precipitation defined as the accumulated precipitation over a 3-day window before the onset of a precipitation event ($AP3$, mm), 1 hour averaged soil moisture before onset of precipitation at 5, 30 and 70 cm depths (SM, %) and mean daily air temperature (T, °C) were selected for each of the 40 events. The selection of input variables was based on the combination of a priori knowledge obtained from Sahraei et al. (2020) and expert knowledge. The values of input and output variables are presented in the Supplementary Table S 3.1.

3.2.2 Machine learning algorithms

3.2.2.1 Artificial Neural Network (ANN) theoretical background

In this study, the Multilayer Perceptron (MLP) feedforward neural network was applied, which is one of the most widely used ANNs in hydrology (Tanty and Desmukh, 2015; Oyebode and Stretch, 2018). The MLP network consists of a set of nodes organized into three different types of layers, i.e. input, hidden and output layers (Figure 3.4). The input layer includes nodes, through which the input data is incorporated into the network. In the hidden layers, the nodes receive signals from the

nodes in the previous layer via weighted connections, at which the weights determine the strength of the signals. The weighted values from the previous layer are summed together with a bias associated with the nodes. The result is then passed through an activation function to generate a signal (i.e., activation level) for each node. Next, the signals are transmitted to the subsequent layer and the process is continued until the information reaches the output layer. The signals, which are generated in the output layer are the model outputs. In general, the mathematical operation of a MLP feedforward neural network is given by Equation (3.3):

$$y_j = f\left(\sum_{i=1}^{n}(w_{ij}x_i) + b_j\right) \tag{3.3}$$

where x_i is the i^{th} nodal value in the previous layer, y_j is the j^{th} nodal value in the present layer, b_j is the bias of the j^{th} node in the present layer, w_{ij} is a weight connecting x_i and y_j, n is the number of nodes in the previous layer, and f denotes the activation function in the present layer.

The backpropagation algorithm (Lecun et al., 2015) was used to train the network, in which the outputs produced in the output layer for the given inputs, are compared to the target values (i.e., observations) and the error is calculated through a loss function. If the output layer does not produce the desired values, the output errors are propagated backwards through the network, while the parameters of nodes (i.e., weights and bias) in each layer are updated along the way with an optimizer. The learning process repeats in this way for several rounds (i.e., epochs) until the loss function reaches an optimal value.

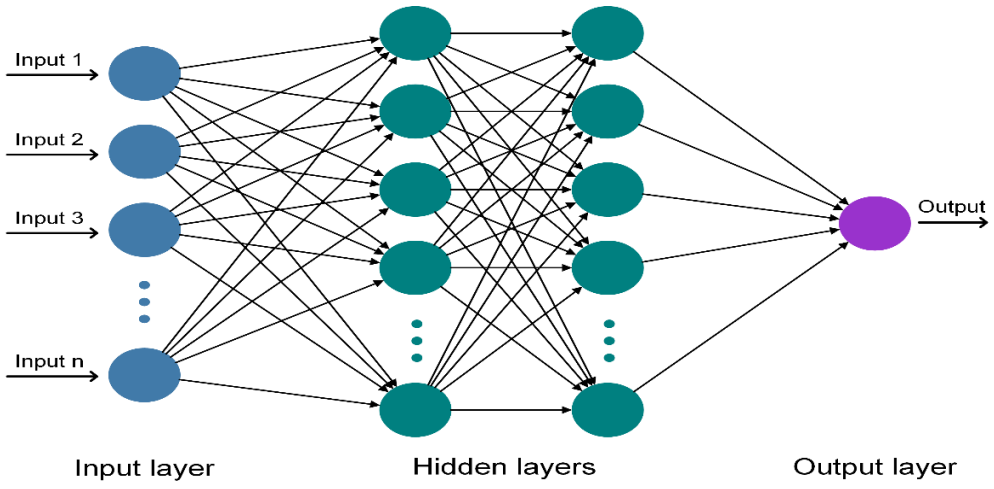

Figure 3.4 Schematic diagram of an Artificial Neural Network (ANN).

3.2.2.2 ANN setup

The performance of the ANN model is strongly dependent on the configuration of its hyperparameters. Among the hyperparameters, the number of nodes in hidden layers, learning rate and activation functions in hidden layers have the greatest influence on the model performance and its robustness (Smithson et al., 2016; Klein and Hutter, 2019). Therefore, a set of the aforementioned hyperparameters was defined to evaluate performance of 45 ANN model configurations resulting from discretizing the hyperparameters shown in Table 3.1. As the base architecture, an ANN model with seven nodes in the input layer, two hidden layers followed by a linear output layer with one node was defined. Previous studies have reported that networks with two hidden layers generally outperform those with a single hidden layer, since they are more stable, flexible and also sufficient to approximate complex nonlinear functions (Sontag, 1992; Flood and Kartam, 1994; Heaton, 2015; Thomas et al., 2017). The most challenging and time consuming aspect of neural network design is choosing the optimal number of hidden nodes (Thomas et al., 2016b). Too few hidden nodes implies that the network does not have enough capacity to solve the problem. Conversely, too many hidden nodes implies that the network memorizes noise within training data, leading to poor generalization capability (Maier and Dandy, 2000; Thomas et al., 2016b). The number of hidden nodes in the first and second hidden layers should be kept nearly equal to yield better generalization capability (Karsoliya, 2012). Given that, five different configurations shown in Table 3.1 were used for the number of hidden nodes according to Equation (3.4) and Equation (3.5) expressed by Thomas et al. (2016a):

$$n_h = n_1 + n_2 \tag{3.4}$$

$$n_1 = int(0.5n_h + 1) \tag{3.5}$$

where n_h represents total number of hidden nodes, n_1 and n_2 represent number of nodes in the first and second hidden layer, respectively. The upper bound on the total number of hidden nodes was considered as half of the total number of samples (i.e., $n_h = 20$) suggested by previous studies (Tamura and Tateishi, 1997; Huang, 2003). The "Adam" optimizer was used for the optimization of the learning process, which is effective for dealing with nonlinear problems, including outliers (Kingma and Ba, 2014). Learning rate is a positive number below 1 that scales the magnitude of the steps taken in the weight space in order to minimize the loss function of the network (Maier and Dandy, 2000; Heaton, 2015; Goodfellow et al., 2016). Too small learning rates slow down the

learning process and may cause the learning to become stuck with a high learning error. On the other hand, too large learning rates can lead the network to go through large oscillations during the learning process or that the network may never converge (Maier and Dandy, 1998; Goodfellow et al., 2016). The default learning rate of the Adam optimizer is set to 0.001. The smaller (0.0001) and larger (0.01) learning rates were also tested. Sigmoid and hyperbolic tangent (Tanh) are the most common activation functions used in the feedforward networks (Maier and Dandy, 2000). However, Rectified Linear Unit (ReLU) is usually a more suitable choice (Heaton, 2015). Sigmoid, Tanh and ReLU as the activation functions of hidden layers were therefore tested. Each network was trained for a maximum of 10,000 epochs to minimize the mean absolute error (MAE).

Table 3.1 Hyperparameter set for the ANN model.

Hyperparameters	Choices
Number of nodes in two hidden layers [m, n]	[3, 1], [5, 3], [7, 5], [9, 7], [11, 9]
Learning rate	0.0001, 0.001, 0.01
Activation functions in hidden layers	ReLU, Sigmoid, Tanh

Note: [m, n] refers to m nodes in the first hidden layer and n nodes in the second hidden layer.

3.2.2.3 Support Vector Machine (SVM) theoretical background

The Support Vector Machine (SVM) was first proposed by Cortes and Vapnik (1995). It is based on statistical learning theory and is derived from the structural risk minimization hypothesis to minimize both empirical risk and the confidence interval of the learning machine in order to achieve good generalization capability. In this study, the Support Vector Regression (SVR) was used, which is an adaptation of the SVM algorithm for regression problems (Vapnik et al., 1997; Smola and Schölkopf, 2004). The basic concept of the SVR is to nonlinearly map the original data in a higher dimensional feature space and to solve a linear regression problem in the feature space (Figure 3.5). Suppose a series of data points $\{x_i, y_i\}_i^n$, where n is the data size, x_i is the input vector and y_i represents the target value, the SVR function is as follows:

$$f(x_i) = w \cdot \phi(x_i) + b \tag{3.6}$$

where w is a weight vector and b is a bias. $\phi(x_i)$ denotes a nonlinear mapping function that maps the inputs vectors to a high-dimensional feature space. w and b are estimated by minimizing the regularized risk function, as shown in Equation (3.7):

$$R(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} L_\varepsilon\big(y_i, f(x_i)\big) \tag{3.7}$$

where $\frac{1}{2}\|w\|^2$ is a regularization term, C is the penalty coefficient, which is considered to specify the trade-off between the empirical risk and the model flatness and $L_\varepsilon\big(y_i, f(x_i)\big)$ is called the ε-insensitive loss function, which is defined according to Equation (3.8):

$$L_\varepsilon\big(y_i, f(x_i)\big) = \max\{0, |y_i - f(x_i)| - \varepsilon\} \tag{3.8}$$

where ε denotes the permitted error threshold. ε will be ignored if the predicted value is within the threshold; otherwise, the loss equals a value greater than ε. In order to represent the distance from actual values to the corresponding boundary values of the ε-tube, two positive slack variables ξ and ξ^* are introduced. Then Equation (3.7) is transformed into the following constrained form:

$$\min f(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi + \xi^*) \tag{3.9}$$

Subject to:

$$\begin{cases} y_i - [w \cdot \phi(x_i)] - b \le \varepsilon + \xi, \xi \ge 0 \\ [w \cdot \phi(x_i)] + b - y_i \le \varepsilon + \xi^*, \xi^* \ge 0 \end{cases} \tag{3.10}$$

This constrained optimization problem can be solved by the following Lagrangian function:

$$\begin{aligned} \max H(\partial_i, \partial_i^*) = &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\partial_i - \partial_i^*)\big(\partial_j - \partial_j^*\big)K\big(x_i, x_j\big) \\ &+ \sum_{i=1}^{n} y_i(\partial_i - \partial_i^*) - \varepsilon\sum_{i=1}^{n} y_i(\partial_i + \partial_i^*) \end{aligned} \tag{3.11}$$

Subject to:

$$\sum_{i=1}^{n}(\partial_i - \partial_i^*) = 0, \partial_i, \partial_i^* \in [0, C] \tag{3.12}$$

Therefore, the regression function is as follows:

$$f(x) = \sum_{i=1}^{n} (\partial_i - \partial_i^*) K\left(x_i, x_j\right) + b \tag{3.13}$$

where $K\left(x_i, x_j\right)$ is the kernel function and ∂_i and ∂_i^* are the nonnegative Lagrangian multipliers, respectively. The kernel function can nonlinearly map the original data into a higher dimensional feature space in an implicit manner and solve a linear regression problem in the feature space.

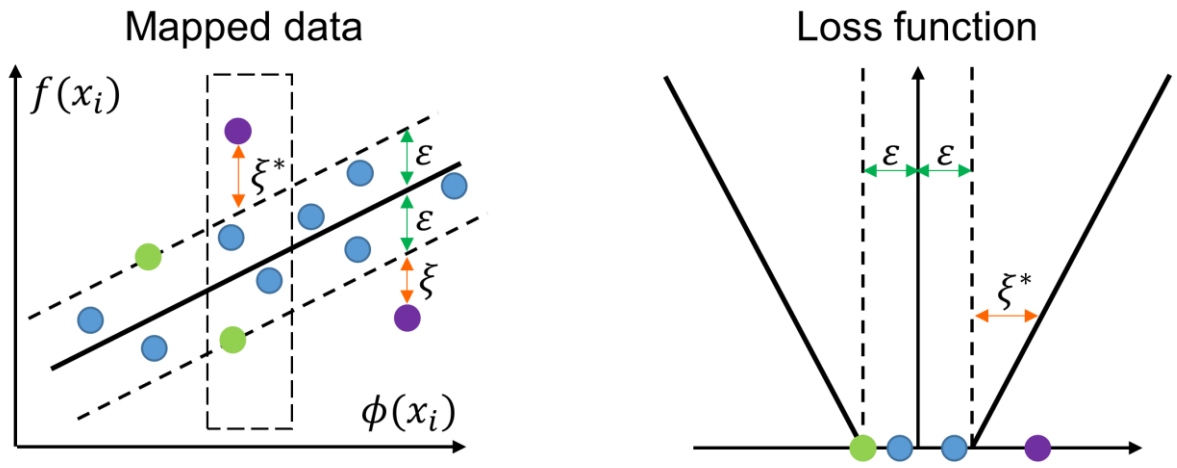

Figure 3.5 Schematic diagram of a Support Vector Regression (SVR).

3.2.2.4 SVM setup

The performance of the SVM model is directly related to the selection of the kernel function (Raghavendra and Deka, 2014; Zhang et al., 2018a). In this study, the performances of four commonly employed kernels in SVM studies including the Linear, Sigmoid, Radial Basis Function (RBF) and Polynomial kernels combined with their optimal hyperparameters were tested. The Linear kernel has two hyperparameters (C, ε), the Sigmoid and RBF kernels have three hyperparameters (C, ε, γ) and the Polynomial kernel has four hyperparameters $(C, \varepsilon, \gamma, d)$ to be optimized. The penalty coefficient C, determines the tolerance of deviations larger than ε from the real value; i.e., smaller deviations are tolerable for larger values of C. The permitted error threshold ε, affects the number of support vectors used in the regression function, i.e., the smaller the value of ε, the greater the number of support vectors that will be selected. The kernel coefficient γ, controls the influence radius of the support vectors, i.e., for high values of gamma, the radius of

the influence area of the support vectors only includes the support vector itself. The degree of the polynomial function d, is the largest of the degrees of polynomial's monomials with non-zero coefficient. 20 values for each of C, ε and γ based on geometric progression and 3 values for d were tested according to the set of hyperparameters shown in Table 3.2. Therefore, 400 configurations for the Linear kernel, 8,000 configurations for the Sigmoid and RBF kernels and 24,000 configurations for the Polynomial kernel were tested. Finally, the performances between the kernel functions in combination with their optimal hyperparameter configurations were compared.

Table 3.2 Hyperparameter set for the kernel functions of the SVM model.

Hyperparameters	Choices
C	1 – 100
ε	0.01 – 1
γ	0.01 – 1
d	2, 3, 4

3.2.3 Model evaluation

Repeated 10-fold cross-validation was performed to estimate the performance of the ANN and SVM models with different hyperparameter configurations. The data was first randomly shuffled and split into ten folds such that each fold contained four samples. For the ANN model, eight folds were considered for train set, one fold for validation set and one fold for test set. For the SVM model, nine folds were considered for train set and one fold for test set. The procedure was repeated 10 iterations so that in the end, every instance was used exactly once in the test set. Within each iteration of the cross-validation, the train set was normalized to be in the range [0-1]. To avoid data leakage, the test and validation sets were normalized using the parameters derived from train set normalization (Hastie et al., 2009). The train set was used to learn the internal parameters (i.e., weights and bias) of the network. The validation set was used for the hyperparameter optimization and the test set was used to estimate the generalization capability of the model (Goodfellow et al., 2016). The ANN model was first trained on the train set for 10,000 epochs to minimize the MAE on the validation set. The number of epochs, at which MAE of the validation set reached its minimum value, was saved. The validation set was then added to the train set and the whole data

set was used to train the model for the fixed number of epochs derived from previous step. The test set was used to estimate the prediction performance of the ANN and SVM models. The estimation of model performance via one run of k-fold cross-validation might be noisy. This means that for each iteration of cross-validation, a different split of data set into k-folds can be implemented and in turn, the distribution of performance scores can be different, resulting in a different mean estimate of model performance (Brownlee, 2020). In order to reduce the noise and allow a reliable estimate of prediction performance, it is recommended to repeat the cross-validation procedure such that the data is re-shuffled before each round (Refaeilzadeh et al., 2009). The common numbers for repetition of cross-validation procedure are three, five and ten (Kuhn and Johnson, 2013; Brownlee, 2020). To obtain a robust and reliable estimate for the model performance, the 10-fold cross-validation was repeated ten times with a different seed for the random shuffle generator, resulting in 100 iterations for each configuration. The performance of the ANN and SVM models were evaluated in terms of the MAE, the root mean squared error (RMSE), the coefficient of determination (R^2) and Nash-Sutcliffe Efficiency (NSE) according to Equations (3.14), (3.15), (3.16) and (3.17), respectively. The optimized (i.e., best) configurations of the ANN and SVM were selected based on the average performance over 100 iterations. The performance of the optimized configuration of the ANN was then compared with that of the optimized configuration of the SVM to choose the final predictive model.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|O_i - P_i| \tag{3.14}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \tag{3.15}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}\sqrt{\sum_{i=1}^{N}(P_i - \overline{P})^2}}\right]^2 \tag{3.16}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2} \tag{3.17}$$

where N is the number of observations (with N=4 for the test set in each iteration of the cross-validation), O_i is the observed value, P_i is the predicted value, $\overline{O}$ is the mean of observed values and $\overline{P}$ is the mean of predicted values. MAE and RMSE describe average deviation of predicted from observed values with smaller values indicating better performance. R^2 indicates correspondence between predicted and observed values with higher values [0 to 1] indicating stronger correlations. NSE evaluates the model ability to predict values different from the mean and gives the proportion of the initial variance accounted for by the model (Nash and Sutcliffe, 1970). NSE ranges between [-∞ to 1]. The closer the NSE value is to 1, the better is the agreement between predicted and observed values. A negative value of the NSE indicates that the mean of observed values is a better predictor than the proposed model. According to Moriasi et al. (2007), model performance can be evaluated as: very good (0.75 < NSE ≤ 1.00), good (0.65 < NSE ≤ 0.75), satisfactory (0.50 < NSE ≤ 0.65) and unsatisfactory (NSE ≤ 0.50).

3.2.4 Model implementation

In this research, the numerical experiments were conducted with Python 3.7 programming environment (van Rossum, 1995) on Ubuntu 20.04 with AMD EPYC 7452 32-core processor, 125 GB of random access memory (RAM) and NVIDIA GTX 1050 Ti graphical processing unit (GPU). The ANN model was built with Keras 2.3.1 library (Chollet, 2015) on top of Theano backend 1.0.4 (Al-Rfou et al., 2016). The SVM model was built with Scikit-Learn 0.22.2 library (Pedregosa et al., 2011). Numpy 1.18.1 (Van Der Walt et al., 2011), Pandas 1.0.1 (McKinney, 2010) and Scipy 1.4.1 libraries (Virtanen et al., 2020) were used for preprocessing and data management. The Matplotlib 3.1.3 (Hunter, 2007) and Seaborn 0.9.1 (Waskom et al., 2020) libraries were used for the data visualization.

3.3 Results and Discussion

3.3.1 ANN prediction performance

The performance of 45 different ANN configurations were evaluated based on the number of nodes in hidden layers, learning rate and activation functions in hidden layers. Table 3.3 shows the average prediction performance over 100 iterations of the cross-validation for each configuration in terms of MAE, RMSE, R^2, and NSE. The best performance was obtained by a combination of

five nodes in the first hidden layer and three nodes in the second hidden layer [5, 3], a learning rate of 0.0001 and the Tanh activation function in hidden layers (configuration #12). This configuration resulted in a MAE of 10.27%, RMSE of 12.91%, R^2 of 0.70, and NSE of 0.63.

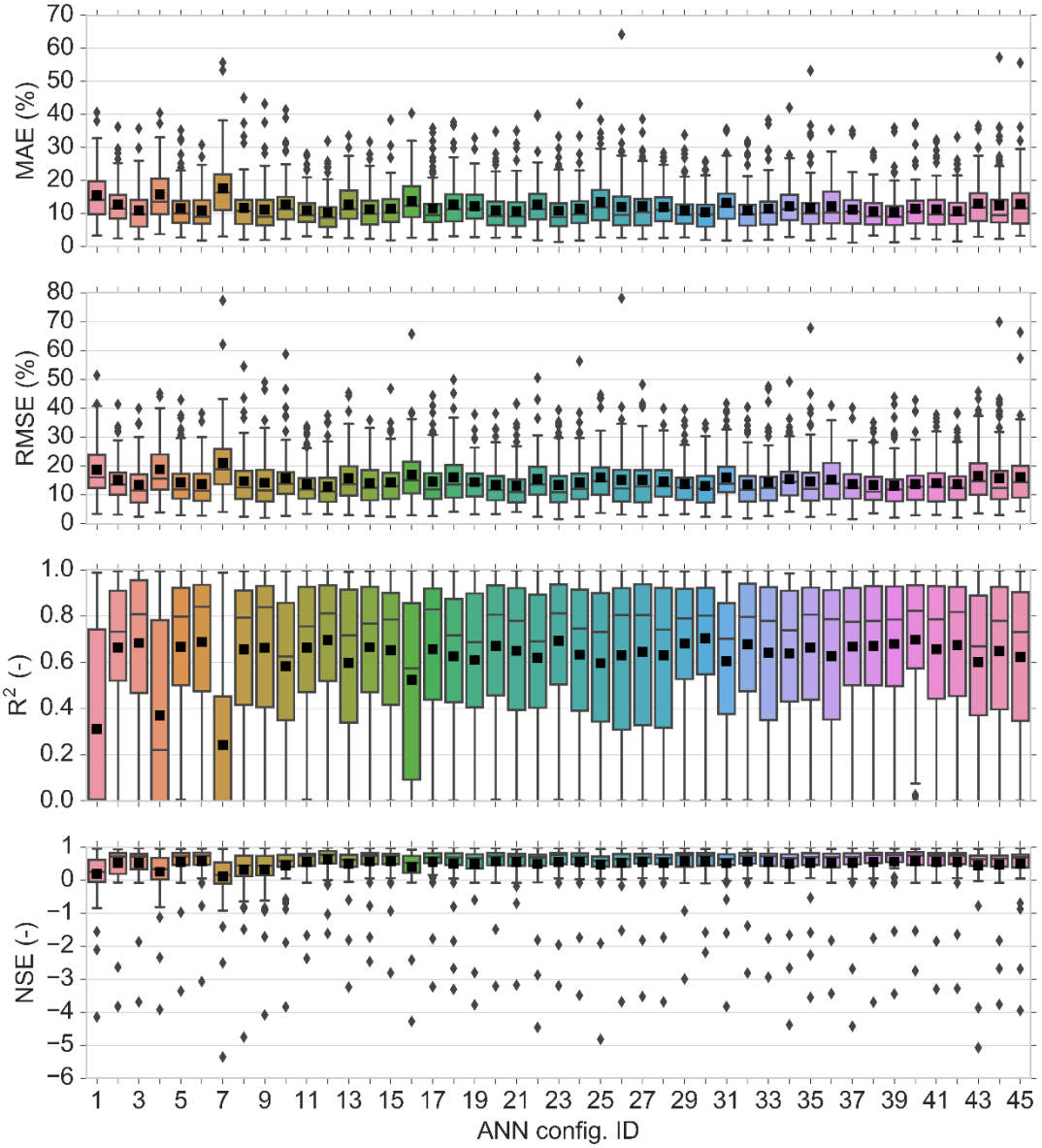

Figure 3.6 Boxplots of the performance criteria for prediction performance of 45 ANN configurations over 100 iterations derived from the ten times 10-fold cross-validation. The average is indicated by a black square and the median by the bar separating a box.

Table 3.3 Average prediction performance of the ANN model with different hyperparameter configurations over 100 iterations of the cross-validation. The best configuration is bolded.

Hidden nodes	Learning rate	Activation function	Config. ID	MAE (%)	RMSE (%)	R^2	NSE
[3-1]	0.0001	ReLU	1	15.57	18.78	0.31	0.19
		Sigmoid	2	12.72	15.18	0.66	0.53
		Tanh	3	10.91	13.55	0.68	0.53
	0.001	ReLU	4	15.73	18.85	0.37	0.26
		Sigmoid	5	11.57	14.34	0.67	0.55
		Tanh	6	10.91	13.74	0.69	0.57
	0.01	ReLU	7	17.66	21.00	0.24	0.11
		Sigmoid	8	11.77	14.75	0.66	0.32
		Tanh	9	11.14	14.18	0.66	0.33
[5-3]	**0.0001**	ReLU	10	12.82	15.82	0.58	0.45
		Sigmoid	11	11.01	13.65	0.66	0.56
		Tanh	**12**	**10.27**	**12.91**	**0.70**	**0.63**
	0.001	ReLU	13	12.71	15.68	0.60	0.50
		Sigmoid	14	11.28	14.14	0.67	0.56
		Tanh	15	11.47	14.32	0.65	0.57
	0.01	ReLU	16	13.80	17.07	0.52	0.41
		Sigmoid	17	11.54	14.57	0.66	0.55
		Tanh	18	12.69	16.08	0.63	0.51
[7-5]	0.0001	ReLU	19	12.12	14.43	0.61	0.50
		Sigmoid	20	10.85	13.39	0.67	0.56
		Tanh	21	10.67	13.22	0.65	0.55
	0.001	ReLU	22	12.68	15.47	0.62	0.49
		Sigmoid	23	10.78	13.45	0.69	0.55
		Tanh	24	11.39	14.30	0.63	0.54
	0.01	ReLU	25	13.42	15.98	0.60	0.48
		Sigmoid	26	12.06	15.10	0.63	0.52
		Tanh	27	12.15	15.15	0.65	0.54
[9-7]	0.0001	ReLU	28	11.87	14.60	0.63	0.53
		Sigmoid	29	10.90	13.63	0.68	0.57
		Tanh	30	10.40	13.00	0.70	0.57
	0.001	ReLU	31	13.23	15.96	0.61	0.51
		Sigmoid	32	10.87	13.62	0.68	0.56
		Tanh	33	11.51	14.34	0.64	0.56
	0.01	ReLU	34	12.21	15.38	0.64	0.49
		Sigmoid	35	11.71	14.69	0.66	0.52
		Tanh	36	12.15	15.34	0.63	0.53
[11-9]	0.0001	ReLU	37	11.12	13.79	0.67	0.52
		Sigmoid	38	10.71	13.37	0.67	0.56
		Tanh	39	10.39	13.09	0.68	0.56
	0.001	ReLU	40	11.38	13.64	0.70	0.57
		Sigmoid	41	11.32	14.13	0.66	0.55
		Tanh	42	10.69	13.68	0.67	0.55
	0.01	ReLU	43	13.06	16.52	0.60	0.45
		Sigmoid	44	12.48	15.83	0.65	0.48
		Tanh	45	12.87	16.24	0.62	0.49

Note: number of nodes in input and output layers are seven and one, respectively and are kept fixed for all of the 45 ANN configurations.

Figure 3.6 shows the boxplots of statistical measures for prediction performance of 45 ANN configurations over 100 iterations of the cross-validation. The average is indicated by a black square and the median by the bar separating a box. It is clear from Figure 3.6 that configuration #1, #4 and #7 with three nodes in the first hidden layer and one node in the second hidden layer and ReLU activation function are the least stable networks, whereas the configuration #12 is the most stable network. Furthermore, it indicates that increasing the complexity of the network not only declines average model performance but also decreases the model stability as can be clearly seen on configurations #43 - #45. Table 3.3 shows that the performance of the 45 configurations based on the number of hidden nodes showed that increasing the number of nodes in hidden layers from eight [5, 3] to twenty [11, 9] did not have a major impact on the prediction performance and even the network with the highest number of hidden nodes [11, 9] did not yield the best prediction performance, whereas decreasing the number of hidden nodes to less than eight declined the model performance, particularly when the ReLU was used as the activation function. This implies that increasing the number of hidden nodes does not necessarily improve the ability of network to generalize, whereas decreasing the number of hidden nodes below a certain threshold may not be sufficient for the network to learn well. Similar behavior has been reported in the previous studies (Wilby et al., 2003; Yoon et al., 2011; Liu and Lu, 2014; Mohammadpour et al., 2015), emphasizing the importance of choosing an appropriate number of hidden nodes to ensure the model performance. Although the optimal number of hidden nodes depends on the essence and complexity of the problem (Maier and Dandy, 2000), the common challenge is to find an appropriate number of hidden nodes, which is sufficient to attain desired generalization capability in a reasonable training time (Maier and Dandy, 2000; Hagan et al., 2014; Goodfellow et al., 2016; Thomas et al., 2016b). The Adam optimizer is generally robust to the choice of its hyperparameters. However, the learning rate needs to be changed sometimes from the suggested default to improve the model performance (Kingma and Ba, 2014; Goodfellow et al., 2016). The smaller (0.0001) and larger (0.01) learning rates were therefore tested. In general, a change of the learning rate, while the number of hidden nodes and activation function were kept fixed, did not have a remarkable effect on the model performance. Nevertheless, the default learning rate still did not yield the best performance. An increase of the learning rate to 0.01 with fixed hidden nodes and activation function declined the performance, whereas the smaller learning rates (i.e., 0.0001 and 0.001) achieved slightly better performance. This implies that increasing the learning rate may cause the network to converge too quickly to a suboptimal solution, whereas smaller learning rate causes the

network to take smaller steps towards minimum of the loss function leading to a better performance (Maier and Dandy, 1998). Smaller learning rates usually guarantee the convergence of the neural network; however, it is a trade-off between the model performance and the time required for the learning process (Heaton, 2015; Liu et al., 2019b). It may be possible that our network achieves slightly better performance with the learning rates lower than 0.0001, but too long training time does not justify the use of smaller learning rates. It is almost impossible to make sure that the selected learning rate is the optimal learning rate, since any value falling between 0 and 1 can be considered for the learning rate (Najah Ahmed et al., 2019). However, it is suggested that testing the learning rate between 0.00001 and 0.1 by an order of magnitude usually ensures near-global optimal solution (Larochelle et al., 2007; Goodfellow et al., 2016; Géron, 2019). This range is usually shorter between 0.0001 and 0.01 when using the Adam optimizer (Poornima and Pushpalatha, 2019; Prasetya and Djamal, 2019; Feng et al., 2020). The Adam optimizer uses an adaptive learning rate algorithm leading the network to converge much faster compared to classical stochastic gradient descent (Kingma and Ba, 2014). Stochastic gradient descent maintains a single learning rate for all of its weight updates and the learning rate does not change during the learning process, whereas the Adam optimizer computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients (Kingma and Ba, 2014). Among the activation functions, ReLU achieved the weakest performance with fixed hidden nodes and learning rate, whereas Sigmoid and Tanh activation functions yielded better, but nearly similar performances. The significant limitation of ReLU is that it easily overfits compared to the Sigmoid and Tanh activation functions (Nwankpa et al., 2018). ReLU is sometimes fragile during learning process causing some of the gradients to die. This leads to several nodes being dead as well, thereby causing the weight updates not to activate in future data points and hence, hindering a further learning as dead nodes give zero activation (Goodfellow et al., 2016; Nwankpa et al., 2018). Sigmoid and Tanh, both are monotonically increasing functions that asymptote a finite value as $\pm\infty$ is approached (LeCun et al., 2012). The only difference is that Sigmoid lies between 0 and 1, whereas Tanh lies between 1 and -1. One of the main advantages of Sigmoid and Tanh activation functions is that the activations (i.e., the values in the nodes of the network, not the gradients) may not explode during the learning process, since their output range is bounded (Feng and Lu, 2019; Szandała, 2020). However, it should be pointed out that each activation function has its own strengths and limitations and its performance may be different based on the network complexity and data structure (Nwankpa et al., 2018; Feng and Lu, 2019). Therefore, choosing a proper

activation function should be prioritized to enhance the model performance. ANN model was trained for a maximum of 10,000 epochs to automatically optimize the number of epoch, at which the loss function (i.e., MAE) reaches its minimum on the validation set. Figure 3.7 shows MAE values over 10,000 training epochs on the normalized train and validation data over 100 iterations of the cross-validation for the optimized ANN model. The model converged within 6,000 epochs. The convergence behavior of the model indicates that the optimized model was robust and efficient in the learning process and that MAE was a proper loss function for the model to learn the problem.

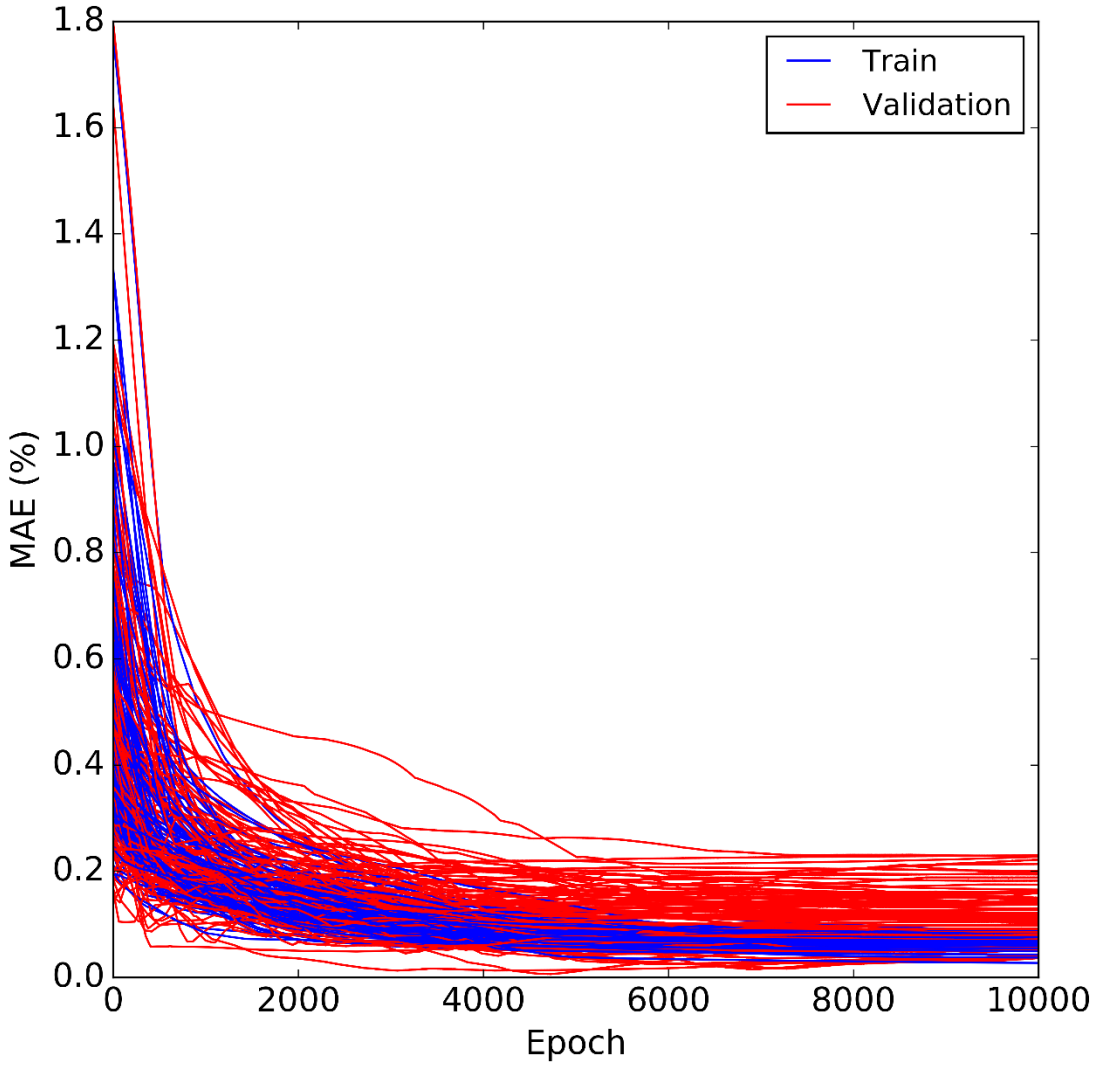

Figure 3.7 Loss function (MAE) versus epoch on the normalized train and validation sets over 100 iterations derived from the ten times 10-fold cross-validation for the optimized ANN.

3.3.2 SVM prediction performance

The prediction performance of SVM was evaluated using various types of kernel functions in combination with the optimal hyperparameters of the kernel functions. Table 3.4 shows the average prediction performance over 100 iterations of the cross-validation for each kernel function with its optimal hyperparameters. The best prediction performance was obtained by a Polynomial kernel function with kernel hyperparameters of (C = 8.8587, ε = 0.0162, γ = 0.1833, d = 2). This configuration achieved a MAE of 7.89%, RMSE of 9.43%, R^2 of 0.83, and NSE of 0.78. The RBF kernel ranked second among the kernel functions. The Linear and Sigmoid kernels performed very similar to each other but slightly weaker than the RBF kernel. Figure 3.8 shows the boxplots of statistical measures for prediction performance of SVM kernels over 100 iterations of the cross-validation. It is observed that Linear and Sigmoid are the least stable kernels, whereas the Polynomial is the most stable kernel. The RBF kernel is more stable than Linear and Sigmoid kernels but less stable than Polynomial kernel. Our findings are in line with previous studies where Polynomial and RBF were found to be the most appropriate kernel functions that tend to give good performance for the modelling of the nonlinear systems (Raghavendra and Deka, 2014). However, the RBF kernel is the mostly used kernel in previous studies (Raghavendra and Deka, 2014), since it has less hyperparameters than the Polynomial kernel that affect the complexity of model selection. Unlike the linear kernel, it can capture the nonlinear relation between class labels and attributes and it tends to exhibit satisfactory performance under general smoothness assumptions (Raghavendra and Deka, 2014; Yaseen et al., 2015). The degree of polynomial function d defines the smoothness of the function, which gives the polynomial kernel the advantage of increasing the dimensionality (Raghavendra and Deka, 2014). For d = 2, the separation surfaces correspond to conic surfaces (i.e., an ellipse or hyperbola) in the feature space. Higher degrees can yield more complex decision boundaries, however, it may increase the risk of overfitting and hence, a decline in model performance (Yaman et al., 2012). Polynomial kernel takes the feature combinations implicitly into account instead of combining the features explicitly. By setting d, different numbers of feature conjunctions can be implicitly computed. In this way, Polynomial kernel performs often better than linear kernels, which do not utilize feature conjunctions (Wu et al., 2007).

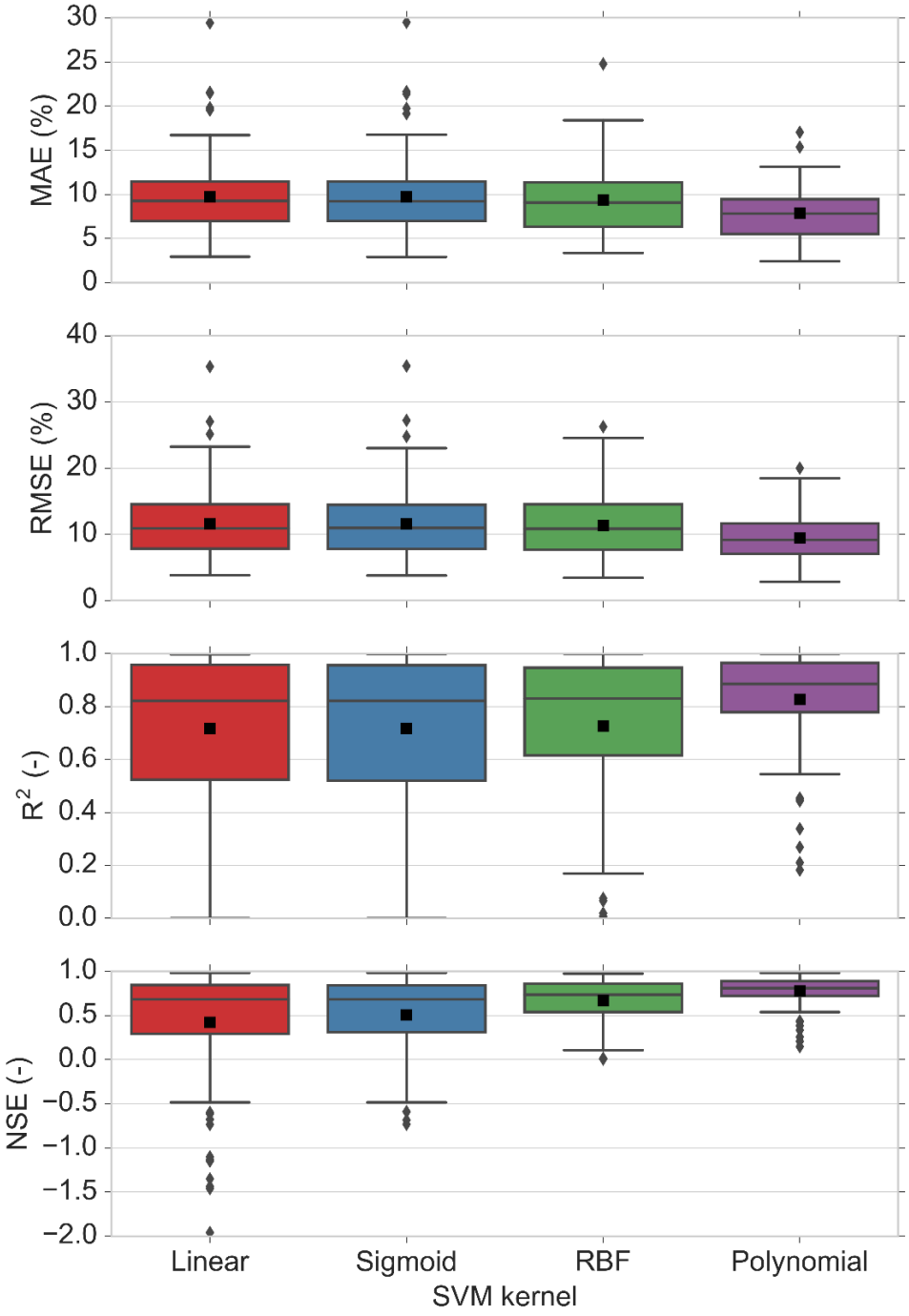

Figure 3.8 Boxplots of the performance criteria for prediction performance of SVM kernels over 100 iterations derived from the ten times 10-fold cross-validation. The average is indicated by a black square and the median by the bar separating a box.

Table 3.4 Average prediction performance of the SVM model using different kernel functions with their optimal hyperparameters (C, $\boldsymbol{\varepsilon}$, $\boldsymbol{\gamma}$, d) over 100 iterations of the cross-validation. The best configuration is bolded.

Kernel	C	ε	γ	d	MAE (%)	RMSE (%)	R^2	NSE
Linear	2.0691	0.0886	-	-	9.72	11.61	0.72	0.42
Sigmoid	61.5848	0.0886	0.0336	-	9.71	11.60	0.72	0.51
RBF	1.6238	0.0428	0.2976	-	9.38	11.30	0.73	0.67
Polynomial	**8.8587**	**0.0162**	**0.1833**	**2**	**7.89**	**9.43**	**0.83**	**0.78**

3.3.3 Comparative performance of machine learning models

The optimized ANN and SVM models were selected based on the average performance over 100 iterations derived from ten times 10-fold cross-validation for an in-depth comparison of the models' capability in prediction of maximum event water fractions in the Schwingbach. Figure 3.9 shows the distribution of statistical measures for learning and prediction performance of the optimized ANN and SVM models over 100 iterations of the cross-validation. The optimized SVM outperformed the optimized ANN model in terms of performance and its stability. In terms of performance capability, the SVM achieved better performance in the train as well as in the test sets. The average and median values over the train and test sets were lower in terms of the MAE and RMSE, and higher in terms of the R^2 and NSE for the SVM compared to those of the ANN. In terms of performance stability, fewer variations were observed for the performance of the SVM model over the 100 iterations of cross-validation indicating that the SVM is a more stable model with respect to different train and test sets.

The 10-fold cross-validation strategy provides the opportunity for each data point to be predicted exactly once. Repeating this procedure ten times, results in ten predicted values for each data point, though over differently shuffled train and test sets. Figure 3.10 shows the comparison of observed and predicted maximum event water fractions of the 40 events by the optimized ANN and SVM models. The red and blue circles indicate the average of ten predicted values at each event by the optimized ANN and optimized SVM, respectively. The red and blue bands depict the range of all ten predicted values at each event by the optimized ANN and optimized SVM, respectively. Both

models were able to reproduce the overall dynamic of maximum event water fractions across the events quite well. However, the optimized SVM model showed superiority over the optimized ANN model for predicting of the maximum event water fractions. Good match between observed and predicted maximum event water fractions was found using SVM model and generally, the predictions were closer to the corresponding observed values than those of the ANN model. The ANN tended to underestimate the events with high maximum event water fractions (e.g., events 4, 19, 32 and 37) and to overestimate the events with low maximum event water fractions (e.g., events 3, 6, 7 and 36). The main reason that the ANN may not be able to capture extreme values is the unavailability of the high number of extreme values in the training data, and hence the ANN cannot adequately learn the process in respect of extremes (Adnan et al., 2019a). The other main reason may be the fact that the ranges of extreme values in the training data are lower than those of the validation and/or test data (Adnan et al., 2019a; Malik et al., 2020). This leads to extrapolation difficulties in machine learning models (Kisi and Aytek, 2013; Kisi and Parmar, 2016). Several researchers have also reported this constraint in the application of the ANN in previous studies (Karunanithi et al., 1994; Minns and Hall, 1996; Dawson and Wilby, 1998; Campolo et al., 1999; Jimeno-Sáez et al., 2018; Singh et al., 2018; Najah Ahmed et al., 2019). It is evident that the range of all ten predicted values at each event was relatively high for the ANN model, particularly for the events with high maximum event water fractions. In contrast, this range was relatively low for the SVM model. It indicates that the performance of the SVM model for the prediction of maximum event water fraction is more reliable under randomly shuffled data compared to that of the ANN model.

The overall prediction ability of the SVM model was superior to that of the ANN model in this study. SVM implements the structural risk minimization principle, which gives it superior generalization ability in the situation of small sample size, whereas the ANN implements empirical risk minimization principle, which makes it more vulnerable to overfitting and susceptible to consider noise in the data as a pattern (Wu et al., 2008; Behzad et al., 2010; Yoon et al., 2011; Liu and Lu, 2014).

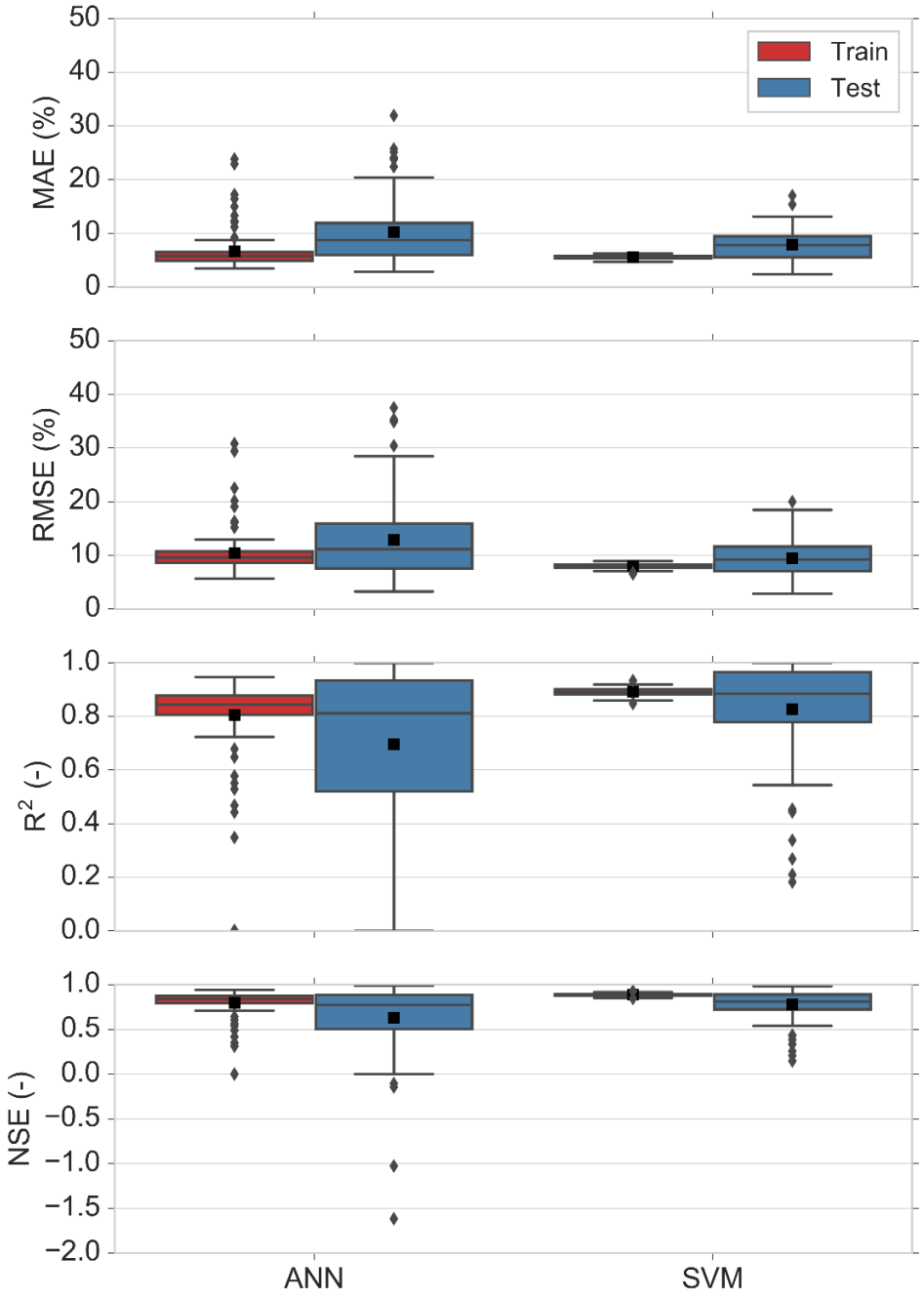

Figure 3.9 Boxplots of the performance criteria of the optimized ANN and optimized SVM for the train and test sets of 100 iterations derived from the ten times 10-fold cross-validation. The average is indicated by a black square and the median by the bar separating a box.

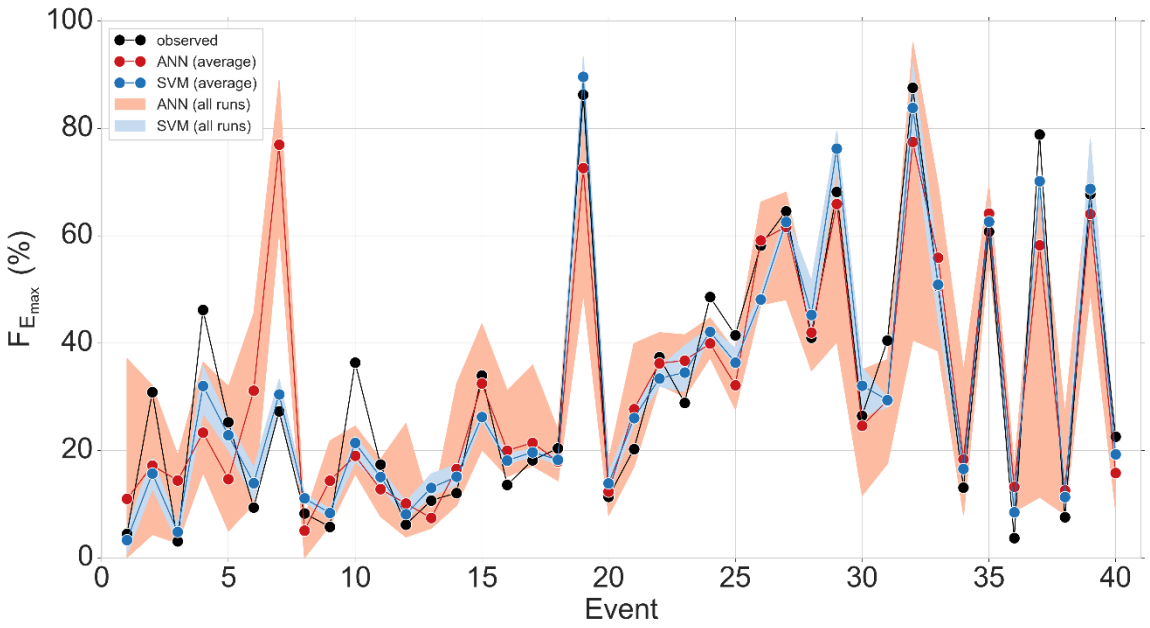

Figure 3.10 Comparison of observed and predicted maximum event water fractions $\boldsymbol{F_{E_{max}}}$ of the 40 events predicted by the optimized ANN and optimized SVM models. The red and blue circles indicate the average of ten predicted values at each event derived from ten times 10-fold cross-validation by the optimized ANN and SVM models, respectively. The red and blue bands indicate the range of all ten predicted values (i.e., all runs) at each event derived from ten times 10-fold cross-validation by the optimized ANN and SVM models, respectively.

3.3.4 Application of machine learning on small data set: challenges and strategies

Generally, the performance of machine learning models excels when having more training data available (Schmidhuber, 2015). Machine learning has exhibited promising performance in many fields such as natural language processing and image recognition where big data set is employed in training procedure (Feng et al., 2019). However, collecting training data set that is large enough is a challenge for some research fields such as medical, material and natural sciences due to the complexity and high costs of data collection (Feng et al., 2019; Brigato and Iocchi, 2020). Compared to other areas of hydrological research, the available data is still limited in the field of isotope hydrology likely due to the laborious and cost intensive measurements of water isotopes. This lack of data is even more pronounced in case of high-resolution isotopic hydrographs at the event scale.

The key challenge on the application of machine learning on small data sets is that the model generalizes patterns in training data such that it correctly predicts unseen data on the test data set. The other main challenge is to ensure a reliable estimation for the generalization ability of the machine learning model. With the advance of machine learning during recent years, several strategies have been established to tackle data limitations that may affect the performance of the machine learning models. Resampling strategies like repeated k-fold cross-validation are the most appropriate for a small sample size (Kim, 2009; Refaeilzadeh et al., 2009; Beleites et al., 2013; Goodfellow et al., 2016). This strategy estimates the model performance by repeating the k-fold cross-validation procedure for number of times, where each time the data is shuffled randomly and subsequently the average performance over all iterations is reported. This is not only an unbiased estimate of model performance but also a reliable estimate of model performance and its stability with respect to different train and test sets (Kohavi, 1995; Beleites et al., 2013; Géron, 2019). Previous studies have successfully utilized cross-validation technique on small sample size ranging from 35 to 60 instances using SVM and ANN models (Behzad et al., 2010; Das et al., 2012; Khovanova et al., 2015; Shaikhina et al., 2015).

Model complexity is a critical factor when only limited data set is available. Using too complex models increases the likelihood of overfitting (Maier et al., 2010). Overfitting occurs when the model learns the details and noises in the training data to the extent that it fails to generalize on new data. On the other hand, using too simple models may lead to underfitting. Underfitting occurs when the model is too simple and does not have enough capacity to learn the patterns in the training data set. Ideally, it is desired to select a model at the sweet spot between underfitting and overfitting. In order to reduce the risk of underfitting and overfitting, a validation set was used to optimize the complexity of the models. Figure 3.6 shows that increasing the complexity of the ANN architecture does not enhance the performance of the model, whereas using too simple architectures does not lead to a promising performance too. This implies the importance and necessity of model complexity optimization in order to enhance the model performance especially in situation of small data set, for which the risk of underfitting and overfitting is even higher compared to larger data sets.

The choice of machine learning algorithm is another important factor that needs additional attention when dealing with limited data sets. SVM has shown to be a more suitable algorithm compared to ANN in the case of small sample sizes (Wu et al., 2008; Behzad et al., 2010; Das et al., 2012; Liu

and Lu, 2014). Structural risk minimization and statistical machine learning process are the bases of SVM. The structural risk minimization aims to minimize the upper bound to the generalization error instead of traditional local training error. Moreover, SVM offers a unique and globally optimal solution towing to the convex characteristic of the ideal problem and employs high-dimensional spaced set of kernel functions that subtly includes nonlinear transformation. Therefore, it has no hypothesis in functional transformation, which makes it essential to have linearly divisible data. At the same time, overfitting is unlikely to occur with the SVM, if the parameters are properly selected. Contrasting, ANN implements empirical risk minimization principle, which makes it more vulnerable to overfitting and falling into a local solution. The results showed that ANN was less stable compared to SVM (Figure 3.9). This might be due to the sensitivity of ANN to the initialization parameter values and training order (Bowden et al., 2002). ANN initialization and backpropagation training algorithms commonly contain deliberate degrees of randomness in order to improve convergence to the global optimal of the associated loss function (Wasserman, 1989; LeBaron and Weigend, 1998). Moreover, the order with which the training data is fed to the ANN may affect the level of convergence and produce erratic outcomes (LeBaron and Weigend, 1998). Such ANN volatilities limit the reproducibility of the results. The success of k-fold cross-validation to reduce the stability problems in ANNs has been discussed in previous studies (Maier and Dandy, 1998, 2000; Bowden et al., 2002). Nevertheless, additional care for optimization of hyperparameters and complexity of network needs to be taken when applying ANN to small sample size.

3.4 Conclusion

Estimating maximum event water fractions in streamflow based on the isotopic hydrograph separation gives valuable insight into runoff generation mechanisms and hydrological response characteristics of a catchment. However, such estimation is not always possible due to the spatiotemporal difficulties in sampling and measuring of stable isotopes of water. Therefore, there is a need for a proper predictive model to predict maximum event water fractions in streamflow even at times when no direct sampling and measurements of stable isotopes of water are available. As the relationships between the maximum event water fraction and its drivers are complex and nonlinear, predictions using process-based models are difficult. Recently, machine learning algorithms have become increasingly popular in the field of hydrology due to their ability in

representing complex and nonlinear systems without any a priori assumption about the structure of the data and knowledge about the underlying physical processes. However, the potential of machine learning in the field of isotope hydrology has rarely been investigated. This study investigates the applicability of ANN and SVM algorithms to predict maximum event water fractions on independent events in the Schwingbach Environmental Observatory (SEO), Germany. The result showed that maximum event water fraction could be successfully predicted using only precipitation, soil moisture and air temperature. This is very important in practice because sampling and measuring of stable isotopes of water are cost intensive, laborious and often not feasible under difficult spatiotemporal conditions. It is concluded that SVM is superior to the ANN in the prediction of maximum event water fractions on independent precipitation events. SVM could better capture the dynamics of maximum event water fractions across the events and the predictions were generally closer to the corresponding observed values. ANN tended to underestimate the events with high maximum event water fractions and to overestimate the events with low maximum event water fractions. Detailed discussion is provided with respect to the influence of hyperparameters on the model performance.

Directions of future research include benchmarking the performance of the proposed machine learning models against process-based models in the field of isotope hydrology. Applying the proposed machine learning models with more training data and to other catchments in other parts of the world would allow assessing how well our method is transferable to catchments with different characteristics (e.g., weather patterns, topography, land use and catchment size). Further, the proposed algorithms could be compared to other machine learning algorithms using different input scenarios of hydroclimatic data.

Using machine learning algorithms for the prediction of target variables that are difficult, expensive or cumbersome to measure provides a valuable tool for future applications. Such algorithms can also be of great value for gap filling procedures. Our study focuses on machine learning in the field of isotope hydrology, but further applications in hydro-chemistry, e.g. for the prediction of pesticides or antibiotics have great potential for future research.

Machine learning does not substitute monitoring efforts as the model development is based on the robust data that covers diverse flow situations. However, machine learning is seen as a promising supplement for the establishment of long-term monitoring strategies of hydrological systems.

Supplementary Materials

Table S 3. 1 Values of input and output variables.

Event ID	Date	P (mm)	D (h)	$AP3$ (mm)	SM_5 (%)	SM_30 (%)	SM_70 (%)	T (°C)	$F_{E_{max}} \pm W_{F_{E_{max}}}$ (%)
1	09/08/2018	1.0	0.2	19.0	18.2	21.2	39.3	20.9	4.5 ± 2.2
2	25/08/2018	1.5	3.0	0.8	12.8	19.0	37.6	12.4	30.9 ± 11.6
3	27/08/2018	0.3	2.6	2.3	13.1	19.2	37.8	16.1	3.1 ± 2.5
4	29/08/2018	2.7	6.1	0.3	12.7	19.1	37.6	17.6	46.2 ± 9.6
5	07/09/2018	1.4	0.3	0.5	12.5	18.5	39.2	14.4	25.3 ± 8.5
6	21/09/2018	4.2	1.9	0.0	11.7	17.8	37.2	15.3	9.4 ± 4.0
7	22/09/2018	17.1	20.3	4.2	11.7	17.8	37.6	11.5	27.3 ± 3.4
8	24/09/2018	0.9	0.3	17.1	20.7	19.1	39.0	8.3	8.3 ± 2.1
9	01/10/2018	0.8	0.2	0.0	18.1	19.2	38.9	7.8	5.8 ± 8.2
10	02/10/2018	1.5	3.7	0.8	18.0	19.4	39.5	6.9	36.4 ± 9.6
11	03/10/2018	0.8	5.4	2.3	20.7	19.5	39.4	6.9	17.4 ± 3.5
12	24/10/2018	0.9	7.8	0.0	16.0	20.3	39.8	11.6	6.2 ± 1.5
13	25/10/2018	0.6	4.9	0.9	15.9	20.3	40.1	10.4	10.7 ± 9.4
14	10/11/2018	0.9	0.5	1.4	22.3	21.0	40.0	9.5	12.1 ± 1.9
15	10/11/2018	4.2	15.4	2.3	23.7	21.1	40.3	9.5	34.0 ± 25.0
16	24/11/2018	1.3	2.4	0.0	29.2	25.9	39.9	3.1	13.6 ± 1.5
17	28/11/2018	1.6	3.8	0.0	29.1	25.9	39.7	1.8	18.2 ± 7.7
18	30/11/2018	1.3	6.7	1.8	28.0	26.3	39.9	2.6	20.4 ± 28.3
19	01/12/2018	19.6	41.8	1.9	27.9	26.6	39.9	5.9	86.3 ± 27.0
20	06/12/2018	1.0	4.8	9.4	28.2	29.1	40.0	7.8	11.4 ± 25.6
21	13/04/2019	1.2	0.5	0.0	42.7	30.9	39.2	1.0	20.3 ± 1.2
22	14/04/2019	3.0	9.0	1.2	42.7	30.9	39.3	1.0	37.4 ± 7.9
23	17/04/2019	1.9	4.3	1.0	42.3	30.7	39.3	10.6	28.9 ± 1.9
24	26/04/2019	4.2	9.3	2.7	38.7	30.4	39.8	9.0	48.6 ± 3.7
25	27/04/2019	2.7	8.3	6.9	40.6	30.5	39.7	7.4	41.5 ± 4.9
26	04/05/2019	9.1	4.8	0.0	40.8	30.3	39.7	3.1	58.2 ± 3.7
27	08/05/2019	8.5	13.9	0.1	41.8	30.4	39.6	7.5	64.6 ± 2.1
28	16/05/2019	3.4	13.2	0.0	44.5	34.5	39.7	7.4	41.0 ± 3.4
29	28/05/2019	10.7	12.8	3.0	44.1	32.2	39.5	11.5	68.2 ± 3.8
30	27/07/2019	0.5	2.8	0.0	24.9	30.3	37.8	21.2	26.5 ± 1.9
31	30/07/2019	3.2	0.5	0.5	23.5	30.2	37.8	20.9	40.5 ± 2.7
32	08/09/2019	26.5	25.3	0.4	18.1	30.1	38.2	11.7	87.6 ± 8.0
33	23/09/2019	9.7	38.3	0.0	18.6	30.3	38.9	14.5	50.7 ± 4.3
34	25/09/2019	1.3	6.4	9.7	25.9	30.3	39.4	14.1	13.1 ± 3.3
35	26/09/2019	11.8	22.8	2.9	26.5	30.3	39.4	14.8	60.8 ± 6.5
36	28/09/2019	1.3	3.9	12.8	30.8	30.4	39.5	14.0	3.7 ± 1.6
37	01/10/2019	13.6	7.6	6.0	30.9	30.4	39.5	14.2	78.9 ± 3.8
38	03/10/2019	0.9	0.7	15.3	32.0	30.5	39.6	8.5	7.6 ± 2.8
39	08/10/2019	12.2	19.8	11.1	34.2	30.6	39.6	11.2	67.8 ± 4.6
40	09/10/2019	3.0	1.8	20.9	37.1	31.1	39.6	10.1	22.6 ± 1.5

4 Deep learning for isotope hydrology: The application of Long Short-Term Memory (LSTM) to estimate high temporal resolution of the stable isotope concentrations in stream and groundwater

This chapter is accepted for the publication in the journal *Frontiers in Water* as:

Sahraei, A.[1], Houska, T.[1] & Breuer, L.[1,2]. (2021). Deep learning for isotope hydrology: The application of Long Short-Term Memory (LSTM) to estimate high temporal resolution of the stable isotope concentrations in stream and groundwater. *Frontiers in Water* (under publication). https://doi.org/10.3389/frwa.2021.740044

[1]Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, 35392 Giessen, Germany

[2]Centre for International Development and Environmental Research (ZEU), Justus Liebig University Giessen, Senckenbergstrasse 3, 35390 Giessen, Germany

Abstract

Recent advances in laser spectroscopy has made it feasible to measure stable isotopes of water in high temporal resolution (i.e., sub-daily). High-resolution data allow the identification of fine-scale, short-term transport and mixing processes that are not detectable at coarser resolutions. Despite such advantages, operational routine and long-term sampling of stream and groundwater sources in high temporal resolution is still far from being common. Methods that can be used to interpolate infrequently measured data at multiple sampling sites would be an important step forward. This study investigates the application of a Long Short-Term Memory (LSTM) deep learning model to predict complex and non-linear high-resolution (3 h) isotope concentrations of multiple stream and groundwater sources under different landuse and hillslope positions in the Schwingbach Environmental Observatory (SEO), Germany. The main objective of this study is to explore the prediction performance of an LSTM that is trained on multiple sites, with a set of explanatory data that are more straightforward and less expensive to measure compared to the stable isotopes of water. The explanatory data consist of meteorological data, catchment wetness conditions, and natural tracers (i.e., water temperature, pH and electrical conductivity). We analyse the model's sensitivity to different input data and sequence lengths. To ensure an efficient model

performance, a Bayesian optimization approach is employed to optimize the hyperparameters of the LSTM. Our main finding is that the LSTM allows for predicting stable isotopes of stream and groundwater by using only short-term sequence (6 h) of measured water temperature, pH and electrical conductivity. The best performing LSTM achieved, on average of all sampling sites, an RMSE of 0.7‰, MAE of 0.4‰, R^2 of 0.90 and NSE of 0.70. The LSTM can be utilized to predict and interpolate the continuous isotope concentration time series either for data gap filling or in case where no continuous data acquisition is feasible. This is very valuable in practice because measurements of these tracers are still much cheaper than stable isotopes of water and can be continuously conducted with relatively minor maintenance.

4.1 Introduction

Catchment hydrological processes are complex and it is challenging to comprehend how the catchment responds to precipitation (Uhlenbrook et al., 2002; Zhou et al., 2021). Stable isotopes of water (δ^2H and $\delta^{18}O$) have been widely employed as conservative natural tracers in catchment hydrology to shed light on the hydrological processes. Such tracers have proved to be valuable tools to investigate the origin and formation of recharged water, surface-groundwater interactions, mixing processes between various water sources, and differentiation of evaporation and evapotranspiration (Kendall and McDonnell, 2012; Orlowski et al., 2016). Particularly at the catchment scale, the stable isotopes of water have been used to differentiate runoff components via hydrograph separation techniques (Klaus and McDonnell, 2013), to estimate mean transit times (McGuire and McDonnell, 2006), to identify flow pathways (Tetzlaff et al., 2015), to explore groundwater recharge rates (Koeniger et al., 2016), to understand soil water mixing processes (Sprenger et al., 2016) and to improve hydrological model simulations (Windhorst et al., 2014).

Recent advances in laser spectroscopy has made it feasible to measure stable isotopes in high temporal resolution (i.e., sub-daily) and in situ. High-resolution data allow the identification of fine-scale, short-term transport and mixing processes that are not detectable at coarser resolutions (Birkel et al., 2012). Previously, studies using sub-daily isotope data focused on single precipitation events (McGlynn et al., 2004; Wissmeier and Uhlenbrook, 2007; Berman et al., 2009). Whilst useful, they are still limited in providing insight into short-term response variability and mixing processes over a longer-term catchment behaviour. To overcome these limitations, a few research

groups recently developed automated systems for continuous monitoring of water isotopes directly in the field. Von Freyberg et al (2017) analysed isotopes of precipitation and stream water every 30 min over 28 days to derive fractions of event water from hydrograph separation at eight precipitation events in the Erlenbach catchment, Switzerland. The result indicated that the high-resolution measurements allowed an in-depth comparison of event water fractions through endmember mixing analysis. Heinz et al. (2014) introduced the technical design of an automated system and reported a primary proof-of-concept to monitor isotopes of stream and groundwater in rice paddies in the Philippines. They concluded that the high-resolution measurements provided the foundation for insights into hydrological interactions that could not be studied previously, particularly with respect to spatially distributed sampling. Based on this setup, Mahindawansha et al. (2018) reported impact of seasons and crops on surface and groundwater isotope concentrations in these rice paddies. The result showed that groundwater isotopes reacted rapidly to irrigation under maize dry season, suggesting the process of preferential flow through deep roots and cracks. Quade et al. (2019) monitored soil water isotopic composition every 30 min during one growing season of sugar beet to partition evapotranspiration flux at the Selhausen agricultural research site, Germany. The comparison between non-destructive high-resolution and destructive coarser resolution sampling of soil water showed significant discrepancies between the isotopic compositions of evaporation led in turn to significant differences in evapotranspiration flux estimation. Sahraei et al. (2020) analysed isotopes of stream, groundwater and precipitation every 20 min over approximately five months to investigate hydrological response behaviour and role of precipitation and antecedent wetness conditions in runoff generation in the Schwingbach Environmental Observatory (SEO), Germany. The result revealed that maximum event water fractions of stream and groundwater responded rapidly to precipitation events, indicating the fast delivery of water to the stream through shallow subsurface flow pathways.

Despite the advantages of high-resolution water isotope data, the routine measurements are still far from being common. Long-term, high-resolution sampling of multiple sources are even less common despite the fact that such measurements are likely to provide new insight into hydrological processes and can help to constrain individual endmembers and flow pathways that contribute to runoff generation. Methods that can be used to interpolate infrequently measured data at multiple sampling sites would be an important step forward. Statistical time series methods such as simple exponential smoothing (SES), autoregressive (AR), moving average (MA) or autoregressive

integrated moving average (ARIMA) have been traditionally employed to predict time series problems. The main drawback of the statistical methods is that they assume that the series are derived from linear processes and hence they might be inadequate for water isotope time series that are non-linear (Zhang et al., 1998; Khashei et al., 2009). Another major limitation is that the statistical methods are local models, in which the free parameters are individually estimated for each time series. It means that it is not possible to share the learning over multiple time series to extract patterns that cannot be distinguished at an individual level (Calkoen et al., 2021). In contrast, machine learning provides a useful tool for the joint extraction of non-linear patterns form a collection of time series. The core idea is to predict isotope concentrations with a set of explanatory data that are more straightforward and less expensive to measure. If the machine learning algorithm is able to predict isotope concentrations from the explanatory data, data-driven interpolations of continuous isotope concentration time series can be acquired.

Machine learning is a data-driven approach that aims to give computers the ability to automatically learn and extract patterns from data (Samuel, 1959; Goodfellow et al., 2016). Machine learning has been increasingly applied in hydrology and earth system science in recent years, owing to its capability to efficiently simulate highly non-linear and complex systems without any a priori knowledge of underlying physical processes. Extensive reviews have been released for the application of machine learning in water resources (Lange and Sippel, 2020; Zounemat-Kermani et al., 2020). Artificial Neural Network (ANN) is the most commonly used machine learning algorithm for the prediction of hydrological variables (Maier and Dandy, 2000; Maier et al., 2010). The main advantage of ANN models is that they are universal function approximators, meaning that they can automatically fit a wide range of functions with a high accuracy level (Khashei and Bijari, 2010). The other major advantage of ANNs is that they have an inherent generalization capability, meaning that they are able to recognize and respond to the patterns that are analogous, but not identical to those on which they have been trained (Benardos and Vosniakos, 2007). Nevertheless, a drawback of ANNs, which have primarily been employed for the analysis of time series in the past, is that any information about the sequence of input features is lost (Kratzert et al., 2018). Therefore, more advance machine learning models are required to efficiently handle these temporal dependencies.

Deep learning is an advance sub-field of machine learning that has drawn significant attention recently. Deep learning generally refers to deep neural networks with multilayer structures that can

extract high-level representations from complex and high-dimensional data via a hierarchical learning process applying multiple non-linear transformations (Shen, 2018; Zuo et al., 2019). Long Short-Term Memory (LSTM) is the current state-of-the-art deep learning architecture that is widely adopted to simulate sequential data like time series (Gers et al., 2002). LSTM is a type of Recurrent Neural Network (RNN) that was originally developed by Hochreiter and Schmidhuber (1997). Unlike the traditional RNN networks, LSTM does not suffer from exploding and vanishing gradients, which allows the network to learn long-term dependencies (Hochreiter and Schmidhuber, 1997). This is beneficial to capture dynamics of catchment processes like storage effects, which may play an important role in hydrological processes (Kratzert et al., 2018). Fang et al. (2017) successfully applied LSTM for the first time in hydrological research to predict soil moisture using meteorological forcing data, static physiographic attributes and model-simulated moisture as inputs. They concluded that the LSTM generalizes well across regions with different climates and environmental conditions. Kratzert et al. (2018) used LSTM to model daily runoff using the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) data set over hundreds of catchments in the USA. The result demonstrated that the LSTM showed better prediction performance than traditional RNN due to its ability of learning and storing long-term dependencies. Zhang et al. (2018b) compared the performance of LSTM with that of a feed-forward ANN for the prediction of water table depths in agricultural areas in northwestern China. The result revealed that the LSTM outperformed the traditional feed-forward ANN. Liu et al. (2019) investigated the application of LSTM to predict water quality parameters such as turbidity, dissolved oxygen and chemical oxygen demand in China. They reported that the LSTM is a feasible and effective approach for water quality prediction.

To the best of our knowledge, the potential of LSTM has not been investigated in the field of isotope hydrology yet; and in general, only few studies have explored the application of machine learning in this field. Cerar et al. (2018) compared the performance of multilayer feed-forward ANN network with that of ordinary kriging, simple and multiple linear regression for predicting the isotope composition ($\delta^{18}O$) of groundwater over several locations across Slovenia. They collected 83 groundwater samples from two campaigns, first in spring and second in autumn under base flow conditions. The result showed that feed-forward ANN achieved better performance than the other three models. Sahraei et al. (2021) investigated the potential of Support Vector Machine (SVM) and multilayer feed-forward ANN to predict maximum event water fractions of streamflow

in the Schwingbach Environmental Observatory (SEO), Germany. They found that the SVM outperformed the ANN model as it could better captured the dynamics of maximum event water fractions under distinct hydroclimatic conditions and flow regimes.

For the first time in the field of isotope hydrology, we investigate the application of deep leaning to predict stable isotopes of water. We apply an LSTM to estimate complex and non-linear high-resolution isotope concentrations of multiple stream and groundwater sources under different land use and hillslope positions in the Schwingbach Environmental Observatory (SEO), Germany. We use an automated in situ mobile laboratory, the Water Analysis Trailer for Environmental Research (WATER), to sample and measure high-resolution (3 h) isotope concentrations in two stream reaches and three groundwater sources. Explanatory data comprise meteorological data, catchment wetness conditions, and natural tracers, i.e., water temperature, potential of hydrogen (pH) and electrical conductivity (EC) that are more straightforward and less expensive to measure compared to the stable isotopes of water. In particular, we report on: (1) how different combinations of input data affect the prediction accuracy of the LSTM; and (2) how short-, medium- and long-term dependencies relate to the sequence length of input data in the LSTM.

4.2 Materials and Methods

4.2.1 Study area and data collection

The study was carried out in the headwater catchment of the Schwingbach Environmental Observatory (SEO) in Hesse, Germany (Figure 4.1). The catchment area is 1.03 km^2 with the elevation ranging from 310 m in the north to 415 m a.s.l. in the south (Figure 4.1a). The climate is categorized as temperate oceanic, with a mean annual precipitation of 623 mm and a mean annual air temperature of 9.6°C (Deutscher Wetterdienst, Giessen-Wettenberg station, period 1969-2019). 76% of catchment area is covered by forest that is mostly located in the east and south, 15% by farmland in the north and west and 7% by meadows alongside the stream (Figure 4.1b). The soil is categorized as Cambisol, covered mainly by forests and Stagnosols under farmland. The soil texture is predominantly consists of silt and fine sand with a low clay content. Further details can be found in Orlowski et al. (2016).

An automated climate station (AQ5, Campbell Scientific Inc., Shepshed, UK) equipped with a CR1000 data logger recorded precipitation depth, air temperature, relative humidity, air pressure, solar radiation and wind speed at 5 min intervals (Figure 4.1c). Six remote-controlled data loggers (A753, Adcon, Klosterneuburg, Austria), three at the toeslope (SM1, SM2 and SM3) and three at the footslope (SM4, SM5 and SM6), were connected to sensors (ECH2O 5TE, METER Environment, Pullman, USA) to automatically monitor soil moisture at 5 and 15 cm depths at 5 min intervals. A stream gauge (RBC flume, Eijkelkamp Agrisearch Equipment, Giesbeek, Netherlands) equipped with a pressure transducer (Mini-Diver, Eigenbrodt Inc., Königsmoor, Germany) automatically measured water levels at the outlet (SW2) of the catchment at 10 min intervals. The transducer measurements were calibrated against manual readings and continuous stream discharge was obtained via the calibrated stage-discharge relationship provided by the manufacturer.

An automated mobile laboratory, the Water Analysis Trailer for the Environmental Research (WATER), was used to automatically sample and analyse the stable isotopes of water (δ^2H and $\delta^{18}O$), water temperature, potential of hydrogen (pH) and electrical conductivity (EC) for multiple water sources in situ from August 8th until December 9th in 2018 and from April 12th until October 10th in 2019. The WATER was equipped with a continuous water sampler (CWS) (A0217, Picarro Inc., Santa Clara, USA), coupled to a wavelength-scanned cavity ring-down spectrometer (WS-CRDS) (L2130-i, Picarro Inc., Santa Clara, USA) to analyse the isotopic composition of sampled water. Isotopic ratios are reported in per mill (‰) deviations from the Vienna Standard Mean Ocean Water (VSMOW). We only used δ^2H time series in the LSTM modelling, as $\delta^{18}O$ had a similar variation but lower precision (precision 0.23‰ for $\delta^{18}O$ and 0.57‰ for δ^2H). The high-resolution δ^2H data were verified with water samples that were manually collected from the sampling sites on a weekly basis. A multi-parameter water quality probe (YSI600R, YSI Inc., Yellow Springs, USA) was installed on the sampling board of the WATER to measure temperature, pH and EC of sampled water. For a detailed description of the WATER and the sampling setup refer to Sahraei et al. (2020).

We measured the stable isotope composition (δ^2H), water temperature, pH and EC of two stream water reaches (SW1 and SW2) and three groundwater sources (GW1, GW2 and GW3) (Figure 4.1c). SW1 was sampled approximately 145 m upstream of the WATER at the edge of farmland and SW2 was sampled at the outlet of the catchment next to the WATER. Groundwater was

sampled from piezometers made from perforated PVC tubes sealed in the upper part with bentonite clay to prevent contamination by surface water. The piezometers of GW1 and GW2 were located at the toeslope on the farmland and meadow, respectively. The piezometer of GW3 was located at the footslope at the edge of the forest. The sampling schedule of the WATER allowed measuring isotopic composition, water temperature, pH and EC for each of the stream water reaches at 1.5-hour and for each of the groundwater sources at 3-hour intervals.

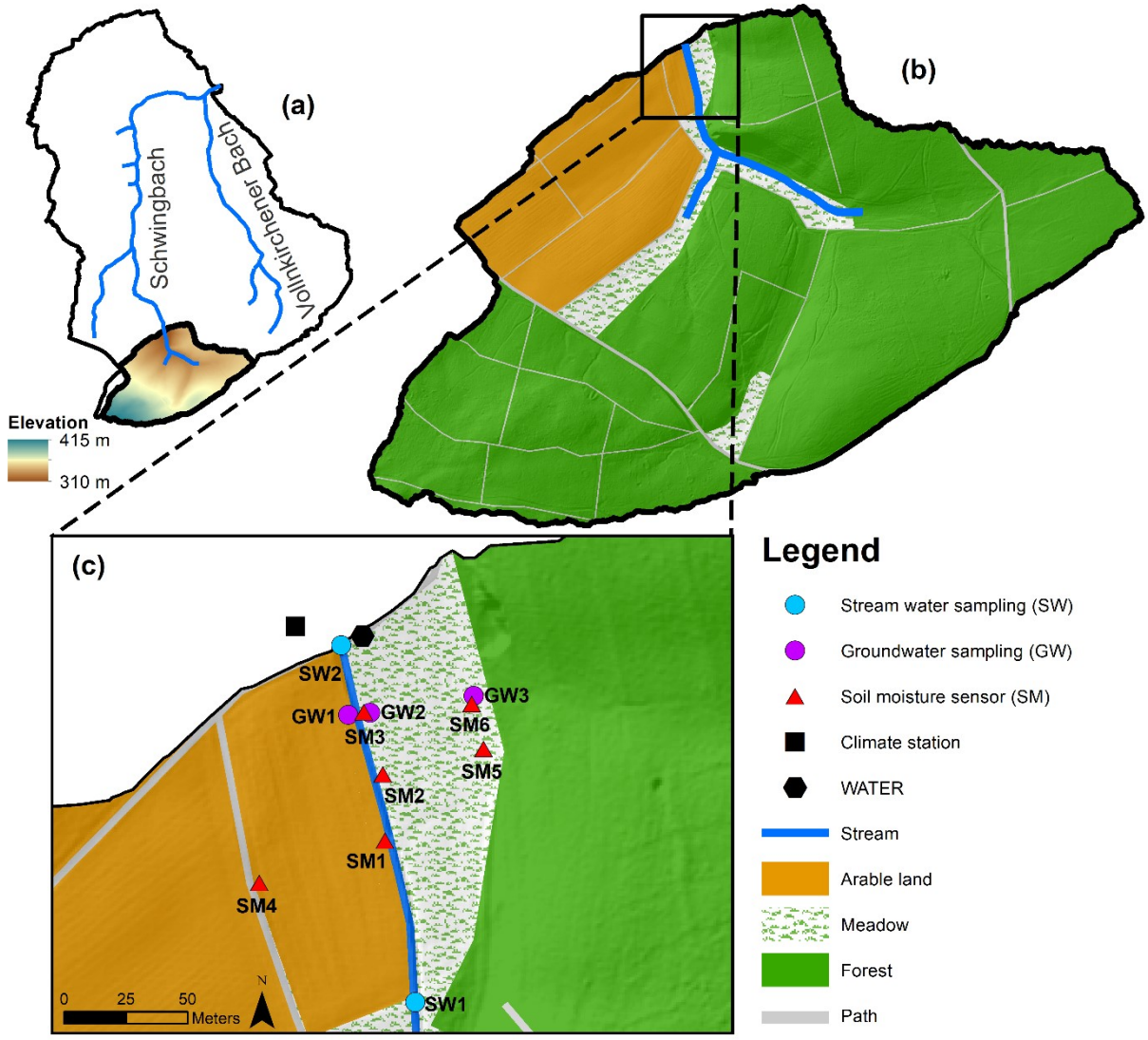

Figure 4.1 (a) Schwingbach Environmental Observatory (SEO), (b) Study area in the Schwingbach headwater and (c) Measuring network along the stream reach of the Schwingbach headwater.

4.2.2 Data pre-processing

As the input features for our LSTM model, we selected a set of explanatory data that are more straightforward and less expensive to measure compared to the stable isotopes of water. The input features were categorized into meteorological, catchment wetness and natural tracer variables (Table 4.1). Our collected data set contained different sampling frequencies. To obtain a uniform frequency, which matches the frequency of the output variable, we aggregated all of the observed data to 3-hour intervals by averaging, except the precipitation, which was aggregated to 3-hour intervals by summing up. The winter period (10th December 2018-11th April 2019) was excluded from the data set because the WATER did not sample during this period due to the freezing weather conditions. When gaps occurred over small timescales due to faulty sensors or equipment maintenance, we used linear interpolation to estimate the missing values. The proportions of observations, which were gap filled through linear interpolation, accounted for at most 3% (73 observations) of total length of the time series for each of the features. We split the observations of all the features into train (70%, 1700 observations), validation (15%, 365 observations) and test (15%, 365 observations) sets, while maintaining the temporal order of the observations. The train set was used to learn the internal parameters (i.e., weights and bias) of the model. The validation set was used to optimize the hyperparameter, and the test set was used to evaluate the generalization capability of the model. To stabilize the learning process and to speed up the convergence, the data needs to be normalized before feeding to the model. For this, the train set was normalized to be in the range [0-1]. The validation and test sets were normalized using the parameters obtained from the train set normalization to avoid data leakage (Hastie et al., 2009).

Table 4.1 Input and output features. The input features are categorized into meteorological, catchment wetness and natural tracer variables. The output feature is isotope composition (δ^2H) at SW1, SW2, GW1, GW2 and GW3 sampling sites.

Input features	Symbol	Unit
Meteorological data		
Precipitation	P	mm
Air temperature	T_a	°C
Relative humidity	RH	%
Air pressure	PR	hPa
Solar radiation	SR	W m^{-2}
Wind speed	WS	m s^{-1}
Catchment wetness		
Stream discharge	Q	l s^{-1}
Soil moisture at 5 cm depth	SM_5	%
Soil moisture at 15 cm depth	SM_{15}	%
Natural tracers		
Temperature	T_w	°C
Potential of hydrogen	pH	-
Electrical conductivity	EC	µS cm^{-1}
Output feature		
Isotopic composition of water	δ^2H	‰

4.2.3 Long Short-Term Memory (LSTM) model

The LSTM is a type of a Recurrent Neural Network (RNN) that is capable of learning long-term dependencies by overcoming the exploding and vanishing gradient problems of traditional RNN networks (Hochreiter and Schmidhuber, 1997). The main characteristics of the LSTM are the specially designed units so called *memory cell* and *gates*. The memory cell consists of forget, input and output gates that together control the flow of information within the LSTM network. The structure of the LSTM memory cell and the algorithms in the cell are shown in Figure 4.2. The memory cell contains a specific status for each time step, the so-called cell state c, which contains the information for long-term memory. The forget gate controls which information is removed from the cell state. The input gate defines which information is updated to the cell state and the

output gate specifies which information is used from the cell state. Suppose a sequence of inputs $x = [x_1, \ldots, x_T]$ with T time steps, where each element x_t is a vector that contains input at time step $(1 \leq t \leq T)$, the process within the LSTM cell is represented with the Equations (4.1-4.6). The LSTM cell updates six parameters at each time step. The first parameter is the forget gate parameter f_t that decides how much of the information from the previous cell state c_{t-1} needs to be forgotten by a sigmoid function (σ) with a linear calculation of the current input x_t and the previous output h_{t-1}. W's and b's with different subscripts represent the gate-specific network weights and bias parameters for the linear calculations. The second parameter is the input gate parameter i_t that controls which new information is updated to the cell state by the sigmoid function with a linear relation on x_t and h_{t-1} as well. The new cell state candidate c_t' is calculated by a hyperbolic tangent function ($tanh$) with a linear relation on x_t and h_{t-1}. The cell state c_t is then updated through an element-wise multiplication $\odot$ operator. In the end, the output parameter o_t is calculated by the sigmoid function with a linear relation on x_t and h_{t-1}. The final output at the current time step h_t is the production of o_t and the $tanh$ function value of the cell state c_t.

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \tag{4.1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4.2}$$

$$c_t' = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4.3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t' \tag{4.4}$$

$$o_t = \sigma(W_o x_t \cdot [h_{t-1}, x_t] + b_o) \tag{4.5}$$

$$h_t = o_t \odot tanh(c_t) \tag{4.6}$$

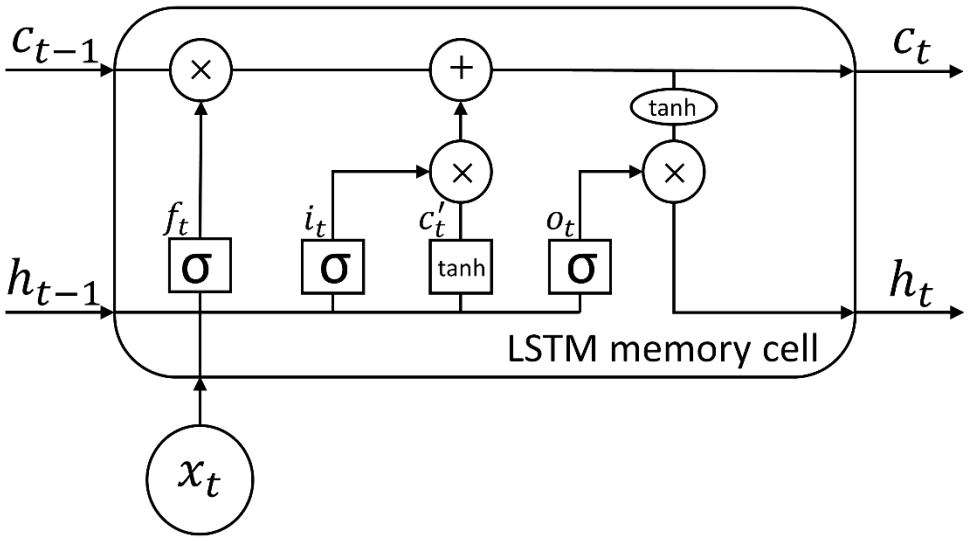

Figure 4.2 The architecture of LSTM memory cell.

4.2.4 Hyperparameter optimization

The majority of machine learning algorithms possess several settings that control the entire learning process (Goodfellow et al., 2016). These settings are referred to as *hyperparameters*. The hyperparameters are exterior to the model and need to be set before the learning process (Géron, 2019). The performance and computational complexity of LSTM models strongly depend on the set of hyperparameters that determine many aspects of the algorithm's behaviour (Nakisa et al., 2018). Therefore, it is essential to optimize hyperparameters to boost the LSTM performance. In this study, we used a Sequential Model-Based Optimization (SMBO) search with the Tree-structured Parzen Estimator (TPE) algorithm, a Bayesian optimization approach (Bergstra et al., 2011). Bayesian optimization is a very effective optimization algorithm that has been shown to outperform well-established methods i.e., grid and random search (Bergstra et al., 2013; Eggensperger et al., 2013). It develops a statistical model between the hyperparameters and the objective function and makes the assumption that there is a smooth but noisy function that maps between the hyperparameters and the objective function (Reimers and Gurevych, 2017). Given the search history of hyperparameters and objective function, SMBO-TPE suggests hyperparameters for the next trial that are expected to improve the objective function. As the number of trials grows, the search history expands and eventually the hyperparameters become optimized. In this study,

we optimized the number of hidden units (i.e., neurons), dropout rate, learning rate, number of epochs and batch size (Table 4.2). We ran the optimization evaluation for 1,000 trials through the search space of the hyperparameters to minimize mean squared error (MSE) on the validation set. The complete set of optimized hyperparameters can be found in Table S 4.1 in the supplementary material.

Table 4.2 Hyperparameter set for the LSTM model.

Hyperparameter	Choices
Number of hidden units	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5
Learning rate	0.00001, 0.0001, 0.001, 0.01, 0.1
Number of epochs	2, 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100
Batch size	32, 64, 128, 256

4.2.5 LSTM model setup

The architecture that we used in this study consists of an input layer with as many neurons as input features, one LSTM layer, a dropout layer and a fully connected dense layer with a single unit for the output feature. We used the Adam optimizer for the optimization of the learning process (Kingma and Ba, 2014) and *tanh* and sigmoid functions as state and gate activation functions, respectively. We ran the LSTM model in a sequence-to-one mode so that an input sequence of a fixed length was used to predict a δ^2H value at the next time step. The sequence length, i.e., the look-back window, is the length of past input observations that the LSTM looks back to predict an output for the next time step.

The performance of deep learning models improves by making more training data available (Schmidhuber, 2015). Building a single LSTM model that is trained and optimized on all of the sampling sites instead of training and optimizing a separate LSTM model for each of the sampling sites, allows the network to learn more general and abstract patterns of the input-output relationship (Kratzert et al., 2018). We therefore built a single LSTM model to predict the δ^2H values at SW1, SW2, GW1, GW2 and GW3 sampling sites. The LSTM was simultaneously trained on train sets of all sampling sites to learn the internal parameters of the model. It simultaneously used validation

sets of all the sampling sites to optimize the hyperparameters and predict the δ^2H values on each test set of the sampling sites to estimate the generalization capability of the model.

4.2.6 LSTM model sensitivity to input features and sequence length

We examined how different combinations of input features affect the prediction performance of the LSTM model. It allows us to identify which combinations give us the most accurate predictions of isotope concentrations in the stream and groundwater sampling sites. We tested seven different input feature scenarios shown in Table 4.3. The first three scenarios (S1, S2 and S3) investigate the effect of meteorological data, catchment wetness conditions and natural tracers on the prediction performance separately. S4, S5 and S6 test the prediction performance when we combine the first three scenarios together and S7 examines the predication performance when we use all of the input features for training of the LSTM. We also used the labels of the sampling sites as input features in all of the seven scenarios. By using the labels of sampling sites as input features, we allow the model to know on which site it trains and hence it produces a unique output for each individual site. This is especially beneficial when using the meteorological data as input features. Since the meteorological data is the same for all sites, the model would otherwise produce the same outputs for them if we do not consider their labels as input features. We used one-hot encoding technique to encode the site labels to the numeric features (Heaton, 2015). This technique creates unique binary columns for each of the labels so that SW1, SW2, GW1, GW2 and GW3 labels are encoded to [1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0], [0, 0, 0, 1] and [0, 0, 0, 0], respectively. Moreover, we used the month of the year of the measurements' time stamp as an input feature in all seven scenarios to represent the inherent seasonality of the data.

We also examined the effect of the sequence length of input data on the prediction performance of our LSTM model. The sequence length is intrinsically connected to the underlying physical processes driving the dynamics of output variables (Duan et al., 2020). A proper sequence length should be large enough to incorporate all historical information relevant to the prediction of the isotopic compositions. However, too large sequence lengths increase the model complexity and training time that can in turn reduce the performance (Duan et al., 2020). Some previous studies have set the sequence length to an arbitrary number (Zhang et al., 2018a; Le et al., 2019; Tennant et al., 2020), whereas some other studies have reported that the sequence length affects the model performance (Fan et al., 2020; Meyal et al., 2020; Xiang et al., 2020). We therefore investigated

the LSTM sensitivity to the sequence length. We defined nine different sequence lengths into three categories to represent short-, medium- and long-term dependencies. We tested sequence lengths of 6, 12 and 24 hours for short-term, 72 (3), 168 (7) and 336 (14) hours (days) for medium-term and 720 (30), 1,440 (60) and 2,160 (90) hours (days) for long-term dependencies. We examined all of these nine sequence lengths for each of the seven input feature scenarios resulting in 63 different scenarios for the sensitivity analysis. We trained the LSTM model and optimized its hyperparameters for each of these scenarios separately to ensure a fair comparison between them.

Table 4.3 Input feature scenarios for the LSTM model.

Scenario label	Input features
S1	Meteorological data
S2	Catchment wetness
S3	Natural tracers
S4	Meteorological data, Catchment wetness
S5	Meteorological data, Natural tracers
S6	Catchment wetness, Natural tracers
S7	Meteorological data, Catchment wetness, Natural tracers

Note: Month of the year and label of sampling sites are used as input features in all of seven scenarios.

4.2.7 Evaluation metrics

We evaluated the prediction performance of the LSTM model using four statistical metrics (goodness-of-fit criteria). We used root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2) and Nash-Sutcliffe efficiency (NSE) according to Equations (4.7-4.10).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \tag{4.7}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|O_i - P_i| \tag{4.8}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}\sqrt{\sum_{i=1}^{N}(P_i - \overline{P})^2}} \right]^2 \quad (4.9)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2} \quad (4.10)$$

where N is the number of observations, O_i is the observed value, P_i is the predicted value, $\overline{O}$ is the mean of observed values and $\overline{P}$ is the mean of predicted values. RMSE and MAE describe the average deviation of predicted from observed values with smaller values indicating a better performance. R^2 is a good indicator for the temporal correspondence between predicted and observed values ranging between [0 to 1], with higher values indicating stronger correlations. NSE evaluates the model ability to predict values different from the mean and gives the proportion of the initial variance accounted for by the model (Nash and Sutcliffe, 1970). The NSE is a preferred criterion in many hydrological studies as it particularly describes a model's capability of matching higher values, extremes and outliers. NSE ranges between [-∞ to 1]. The closer the NSE value is to 1, the better is the agreement between predicted and observed values. A negative value of the NSE indicates that the mean of observed values is a better predictor than the proposed model. According to Moriasi et al. (2007), model performance can be evaluated as: very good ($0.75 < NSE \leq 1.00$), good ($0.65 < NSE \leq 0.75$), satisfactory ($0.50 < NSE \leq 0.65$) and unsatisfactory ($NSE \leq 0.50$).

4.2.8 Setup of numerical experiments

We conducted the numerical experiments with Python 3.7 programming environment (van Rossum, 1995) on Ubuntu 20.04 with AMD EPYC 745-core processor, 125 GB of random access memory (RAM) and NVIDIA GTX 1050 Ti graphical processing unit (GPU). The LSTM model was built with Keras 2.3.1 deep learning framework (Chollet, 2015) on top of TensorFlow backend 2.0 (Abadi et al., 2016). Hyperopt 0.2.5 (Bergstra et al., 2013) and Hyperas 0.4.1 (Pumperla, 2019) libraries were used to implement hyperparameter optimization. Scikit-Learn 0.22.2 (Pedregosa et al., 2011), Numpy 1.18.1 (Van Der Walt et al., 2011), Pandas 1.0.1 (McKinney, 2010) and Scipy 1.4.1 (Virtanen et al., 2020) libraries were used for pre-processing and data management. The

Matplotlib 3.1.3 (Hunter, 2007) and Seaborn 0.9.1 (Waskom et al., 2020) libraries were used for the data visualization.

4.3 Results and Discussion

4.3.1 Temporal dynamics

In the following, we briefly describe the dynamics of the observed data over the sampling period (Figure 4.3-4.6). During the sampling period of 2018, precipitation is very low with respect to the long-term historical record (Deutscher Wetterdienst, Giessen-Wettenberg station, period 1969-2019, mean annual precipitation of 623 mm), with a total annual precipitation of 452 mm (Figure 4.3), stating the unusual dry conditions in 2018 (Vogel et al., 2019). More precipitation is observed during the sampling period of 2019 with an annual sum of 624 mm. Air temperature, relative humidity, air pressure, solar radiation and wind speed show strong diurnal fluctuations during the sampling period (Figure 4.3). However, during the precipitation periods, these fluctuations tend to decrease particularly in case of air temperature and relative humidity, and tend to increase in case of wind speed. Air temperature, relative humidity and solar radiation represent seasonal trends, whereas air pressure and wind speed do not depict a clear trend during the sampling period.

Stream discharge and soil moisture reflect the drought of 2018 with only few small peaks during the precipitation periods (Figure 4.4). The catchment wetness increases in 2019 and stream discharge as well as soil moisture react similarly to precipitation inputs. Soil moisture of shallower depth (i.e., 5 cm) generally shows more clear responses to precipitation than at deeper depths (i.e., 15 cm) at all soil moisture stations. This response is more pronounced in the meadow (SM2, SM3 and SM6) than in the farmland (SM1 and SM4). However, soil moisture of deeper depths in the farmland generally reveal a higher responsiveness to rainfall than equivalent depths in the meadow. The shallower depths of SM2 and SM3 at the toeslope depict a similar reaction pattern to SM6 in footslope suggesting the extension of linkage between toeslope and footslope during precipitation events. On average, SM1 and SM3 show the lowest (8 ± 4.3%, mean ± standard deviation, for 5 cm depth, 14 ± 4.6% for 15 cm depth) and highest (29.4 ± 11%, 24.3 ± 6.8%) soil moisture among the SM stations, respectively.

Figure 4.5 displays that stream water temperature strongly matches the diurnal fluctuations of air temperature, whereas the groundwater temperature follows the overall long-term pattern of air temperature without any high-temporal fluctuations. SW1 and SW2 exhibit similar temperature patterns with means of 11.8 ± 2.7 °C and 11.7 ± 2.7 °C over the sampling period, respectively. However, groundwater temperatures at the sampling sites slightly differs from each other with the mean of 11.3 ± 1.8 °C, 13.0 ± 2.1 °C and 12.3 ± 1.8 °C for GW1, GW2 and GW3, respectively. The patterns of pH dynamics in stream and groundwater sources are almost identical over the sampling period (Figure 4.5). The pH shows an increasing trend while the air temperature decreases in 2018 and then remains almost stable in 2019 with some fluctuations. The EC of stream and groundwater is highly responsive to precipitation inputs (Figure 4.5). Rapid reductions of EC are observed upon in the course of precipitation events that are followed by a fast recovery. The EC of stream as well as groundwater is relatively high in 2018, followed by a smooth decline towards the end of the sampling period. The EC of SW1 (336.9 ± 59.3 µS cm^{-1}) is slightly higher than that of SW2 (325.5 ± 58.3 µS cm^{-1}), whereas the difference between the EC of groundwater sampling sites is higher with means of 228.7 ± 61.2 µS cm^{-1}, 321.9 ± 59.9 µS cm^{-1} and 239.3 ± 50 µS cm^{-1} for GW1, GW2 and GW3, respectively.

Figure 4.6 presents the high temporal resolution of δ^2H at the stream and groundwater sampling sites. The δ^2H of stream and groundwater responds promptly to precipitation, and is strongly synchronized with the response patterns of EC. These rapid isotopic responses in stream as well as groundwater indicate the strong linkage between stream and subsurface flow pathways during the precipitation events (Sahraei et al., 2020). The δ^2H values are lighter with stronger diurnal variations in 2018 compared to those during sampling period of 2019. Seasonal variations are observed during the sampling period with heavier δ^2H values in May-June and lighter in November-December. Figure 4.7 shows the distribution of the isotopic compositions in the stream and groundwater sources over the sampling period. Both stream sampling sites indicate similar variations for δ^2H. The mean values of δ^2H for SW1 and SW2 are -60.7 ± 2.7‰ and -60.6 ± 2.5‰, respectively. However, the variations of δ^2H tend to increase from GW1 to GW3. On average, GW1 (-60.1 ± 2.1‰) and GW2 (-60 ± 1.9‰) at the toeslope depict slightly heavier mean δ^2H values compared to GW3 at the footslope (60.5 ± 2.6‰).

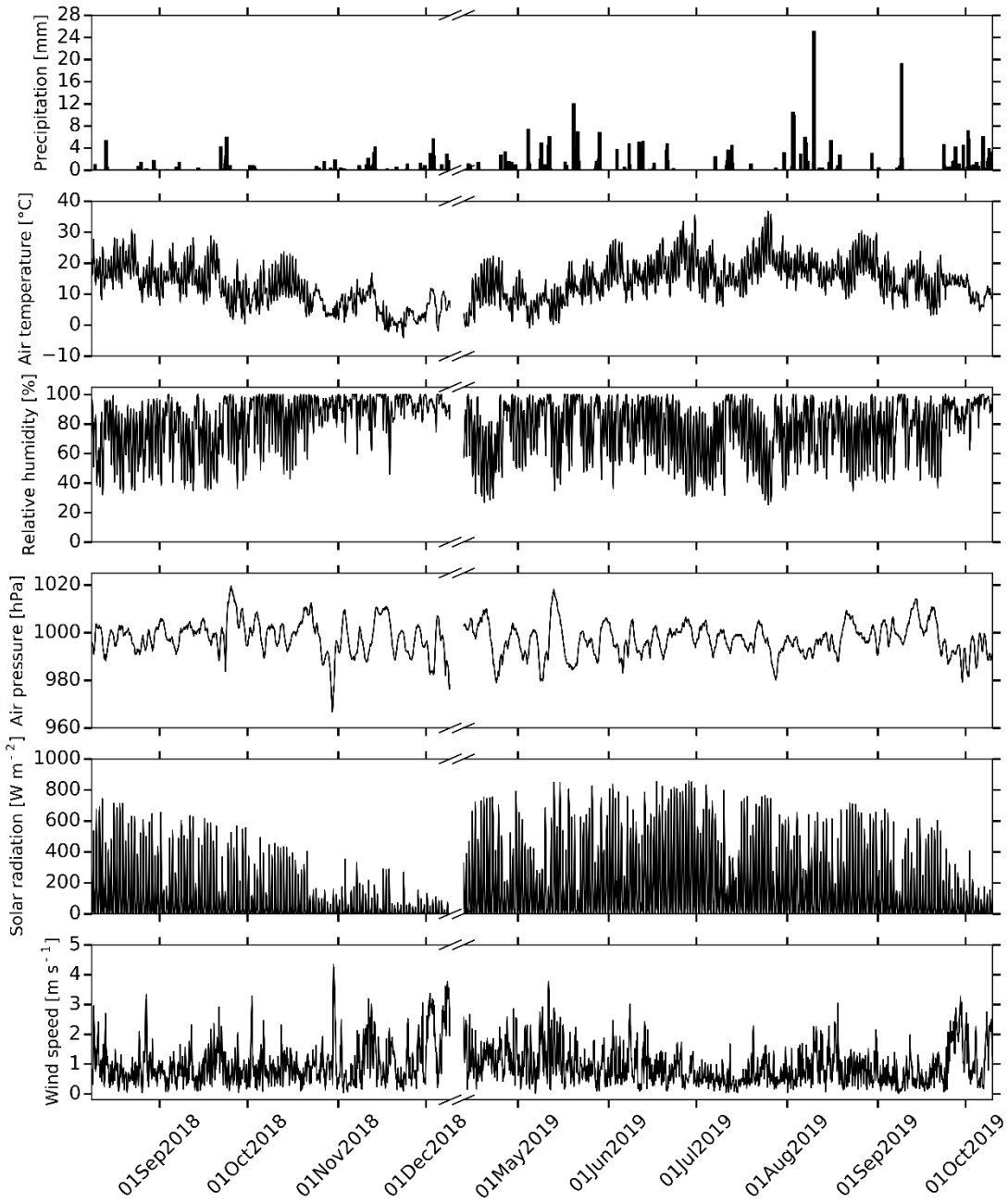

Figure 4.3 Time series of meteorological data. From the top, we report precipitation, air temperature, relative humidity, air pressure, solar radiation and wind speed.

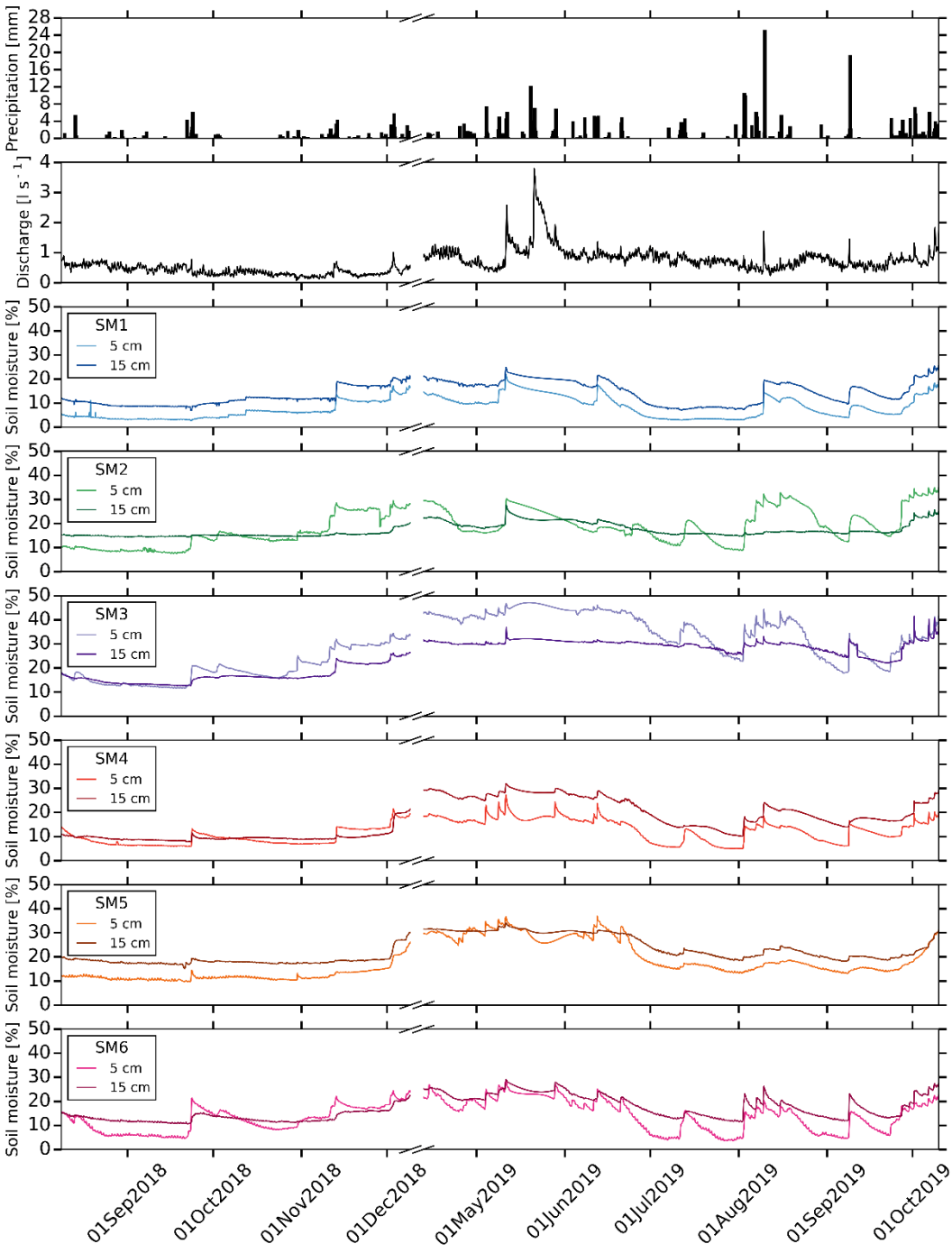

Figure 4.4 Time series of catchment wetness conditions. From the top, we report precipitation, discharge and soil moisture at SM1, SM2, SM3, SM4, SM5 and SM6 stations.

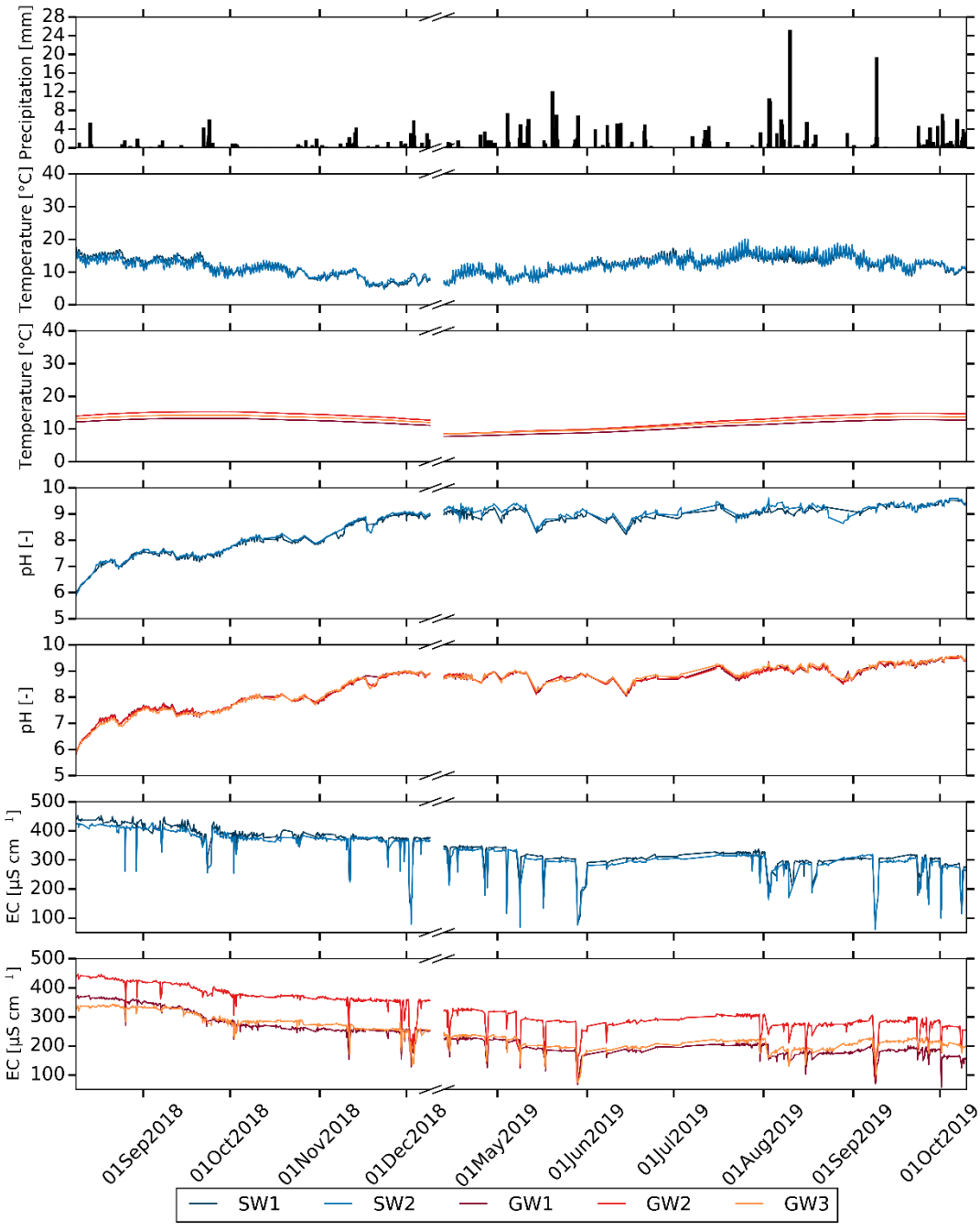

Figure 4.5 Time series of natural tracers at stream and groundwater sampling sites. From the top, we report precipitation, water temperature, pH and electrical conductivity (EC) at stream and groundwater sampling sites.

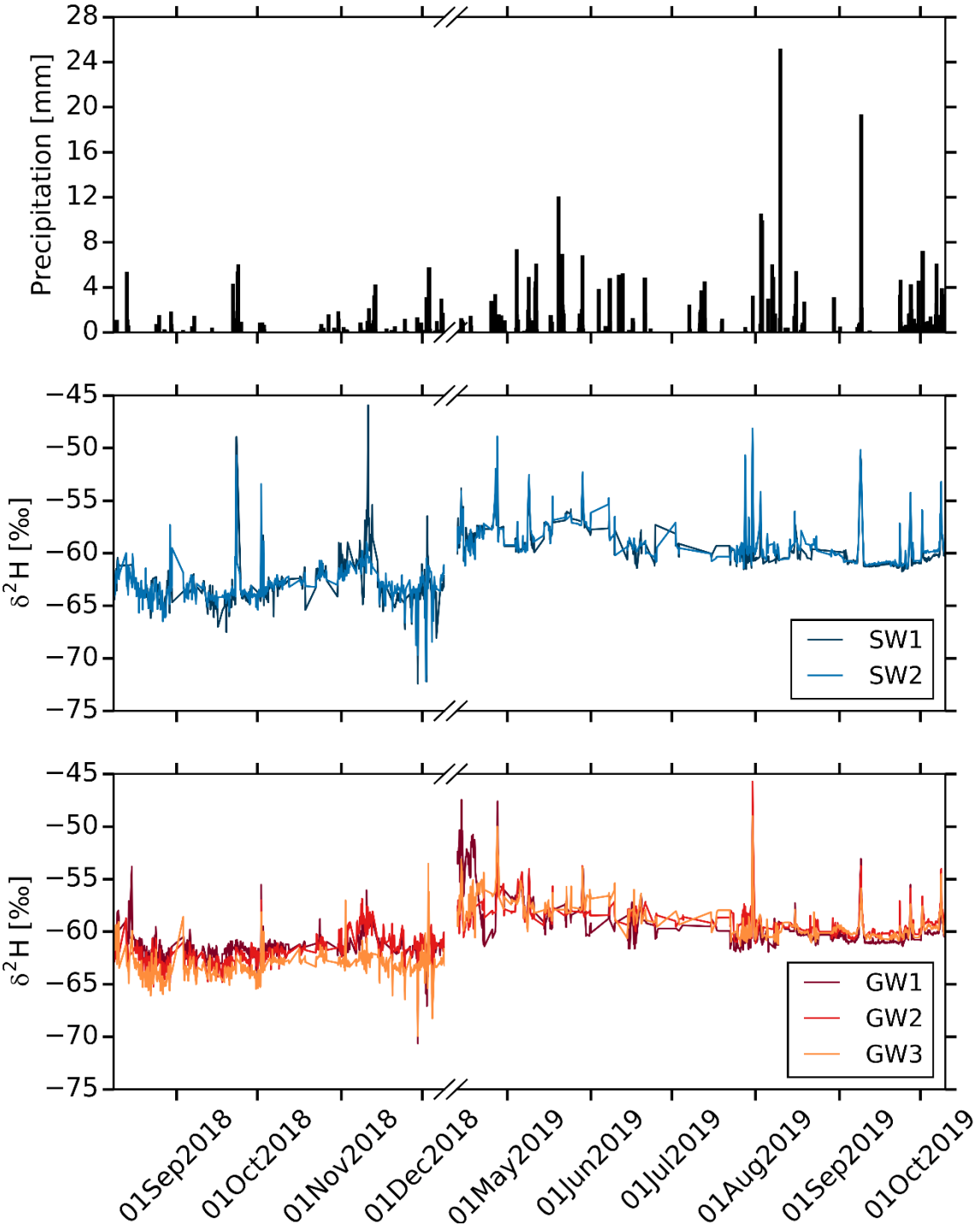

Figure 4.6 Time series of precipitation and stable isotope concentrations at stream and groundwater sampling sites. From the top, we report, precipitation and δ^2H values at stream and groundwater sampling sites.

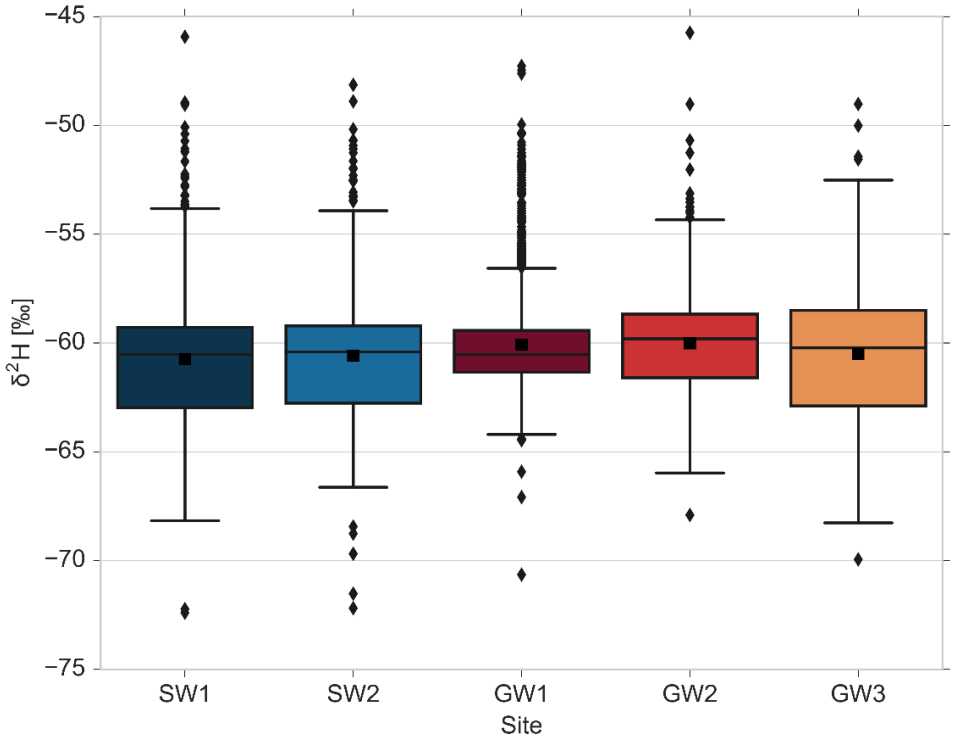

Figure 4.7 Boxplots of the isotopic compositions at stream and groundwater sampling sites. The average is indicated by a black square and the median by the bar separating a box.

4.3.2 LSTM model sensitivity to input features and sequence length

We investigated the impact of input features and sequence length on the prediction performance of the optimized LSTM model. Figure 4.8 shows the heatmaps of the prediction performance using four statistical metrics for the combination of seven input feature scenarios and nine sequence lengths at the stream and groundwater sampling sites (SW1, SW2, GW1, GW2 and GW3). It is apparent that using the meteorological data (scenario S1) and catchment wetness conditions (scenario S2) alone as the input features does not lead to a good prediction performance. Although using meteorological and catchment wetness data together (scenario S4) slightly decreases prediction errors for some cases, yet the performance is not satisfying. For these scenarios, the NSE values are mainly near to zero and even negative. This emphasizes that the model is not able to predict the peak values properly. It suggests that the relations of meteorological data and catchment wetness variables with the water isotope concentrations are not strong enough to transfer enough information that is required for an adequate learning of the LSTM to predict peak values of isotope concentrations.

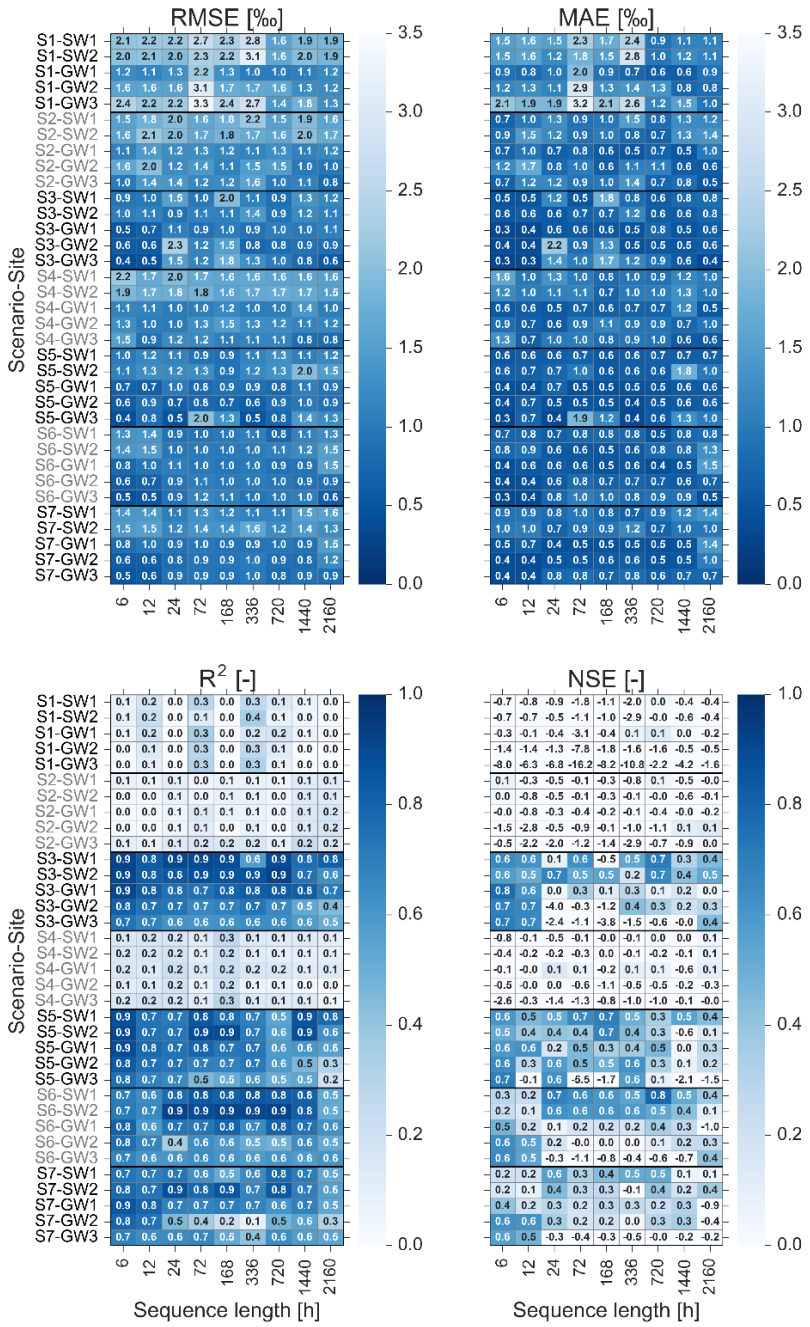

Figure 4.8 Heatmaps of the prediction performance using four statistical metrics for the combination of seven input feature scenarios and nine sequence lengths at stream and groundwater sampling sites. The dark blue indicates good prediction performance (low values of RMSE and MAE and high values of R^2 and NSE), whereas the light blue indicates poor prediction performance (high values of RMSE and MAE and low values of R^2 and NSE).

In contrast, using the natural tracers (scenario S3), i.e., water temperature, pH and EC of the sampling sites, as the input features achieves good prediction performance. It indicates that these easy-to-measure tracers are able to indirectly simulate complex interactions between controlling drivers and water isotope concentrations in the Schwingbach catchment. With a closer look to the prediction errors across the sequence lengths of this scenario, we can see that the model performs the best at all of the sampling sites when we use short sequence lengths (6 and 12 h). This implies that the most recent input data contain the most relevant information for the prediction of isotope concentrations and that there is no significant benefit using older input data. This behaviour can also be noticed for the prediction of peak values, which is measured by NSE metric. The NSE values are higher for shorter sequence lengths indicating that the model can better predict the reaction of isotope signatures during precipitation events by using the recent input data. This can be associated to the rapid response characteristics of the Schwingbach catchment (Orlowski et al., 2014, 2016; Sahraei et al., 2020). The responses of isotopes in stream and groundwater during precipitation events often reveal a rapid mixing of water at the sampling sites with the event water. It takes, on average, 6 h for stream water and 7 h to 8 h for groundwater at the toeslope (GW1 and GW2) and footslope (GW3), respectively, that the isotope signatures reach peak values. After that, they immediately return to pre-event water conditions (Sahraei et al., 2020). This short-term response behaviour together with the synchronized behaviour of isotopic compositions with electrical conductivity makes it sufficient for the model to extract the dynamics of isotopic compositions. In contrast, the model performance declines when we use medium and long sequence lengths. This is possibly due to the noise generated by irrelevant inputs (Xiang et al., 2020).

We combined the natural tracers with meteorological and catchment wetness variables (scenarios S5 and S6) to test if this boosts the prediction performance of the LSTM model. Comparing scenario S3 with S5, we can see that the LSTM still performs better at scenario S3 when using short sequence lengths. However, the model tends to perform better at scenario S5 for medium and long sequence lengths. Similarly, the inclusion of catchment wetness (scenario S6) provides some improvements in the model performance for medium and long sequence lengths. The physical explanation underlying the improved performance of the model when we provide longer historical meteorological and catchment wetness information, may be related to the climatological and storage characteristic of the catchment. It could well be that long-term dependencies exist between the dynamics of isotope concentrations and climatic conditions as well as soil water content of the

catchment. The ability of the LSTM to learn and remember the long-term dependencies provides the opportunity to identify such storage effects within the catchments (Kratzert et al., 2018, 2019). A comparison between prediction errors of scenario S5 and S6 indicates that the model generally performs better at scenario S5. This suggests that the meteorological condition has a higher impact on the dynamic of isotope concentrations compared to the soil moisture state in the catchment, which is in line with the previous findings in the Schwingbach catchment (Orlowski et al., 2014; Sahraei et al., 2020).

We also tested the prediction performance for all meteorological, catchment wetness and natural tracer variables as input features (scenario S7). In general, the model performance does not improve remarkably and it even deteriorates in some cases compared to performance of model at scenario S3, S5 and S6. The inclusion of too many input features increases the model complexity. Using too complex models augments the chance of overfitting (Maier et al., 2010). Overfitting arises when the model learns the details and noises in the training data to an extent that it is unable to generalize to new data.

4.3.3 Visualization of the LSTM prediction performance

We built a single LSTM model to predict the high-resolution δ^2H values at SW1, SW2, GW1, GW2 and GW3 sampling sites in the Schwingbach catchment. Ideally, we do not want to train and optimize the model specifically for each sampling site to achieve the best performance, but rather, we intend to train and optimize the model once for all of the sampling sites using only one specific input feature scenario and sequence length. Therefore, we trained and optimized the model with the input feature scenario and sequence length, at which the model performs on average the best for all of the sampling sites. According to the results presented in Figure 4.8, LSTM achieves the best performance when measurements of the last 6 h of water temperature, pH and EC are used as input features (scenario S3). It achieves, on average of all sampling sites, an RMSE of 0.7‰, MAE of 0.4‰, R^2 of 0.90, and NSE of 0.70.

In the following, we visualize the prediction performance of the best LSTM model on the test sets to better illustrate its generalization ability on the unseen data. Figures 4.9-4.10 show the performance of the LSTM using a sequence length of 6 h for scenario S3 at SW1 and GW3 sampling sites, respectively. The LSTM captures the timing of hydrologic events and base flow conditions of δ^2H at stream and groundwater sites quite well, but it still underestimates the peak of

the hydrograph. This is a commonly known issue with LSTMs and in general, ANN models that they cannot learn adequately the phenomenon in respect of extremes. The major reason that the LSTM may not be able to capture extreme values is the lack of a large number of extremes in the training data (Adnan et al., 2019). The other major reason could be the fact that the range of extreme values in the training data is smaller than those of the validation and test data (Adnan et al., 2019; Malik et al., 2020). This results in difficulties for extrapolation in ANN models (Kisi and Aytek, 2013; Kisi and Parmar, 2016). Several scholars have also reported this limitation in the implementation of the LSTM in previous studies (Kratzert et al., 2018; Chen et al., 2020; Fan et al., 2020; Müller et al., 2020; Xiang et al., 2020). The prediction performances of the best LSTM at SW2, GW1 and GW2 sampling sites are provided in the supplementary material (Figures S 4.1-4.3). The development of the LSTM model assumed stable boundary conditions with regard to land use and cover. Changes of these during the study period have not been taking place in the SEO. However, such changes could be part of long-term simulations in the LSTM model if information were available on this.

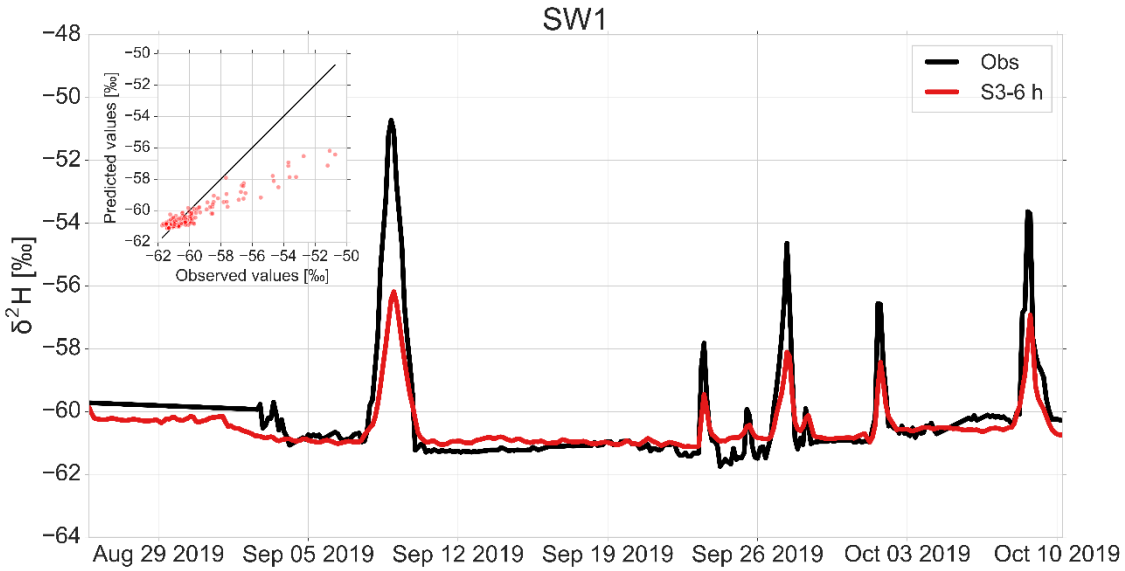

Figure 4.9 Comparison of observed and predicted δ^2H values by the best LSTM using a sequence length of 6 h for water temperature, pH and electrical conductivity measurements (scenario S3-6 h) with an RMSE of 0.9‰, MAE of 0.5‰, R^2 of 0.9 and NSE of 0.6 at SW1 site.

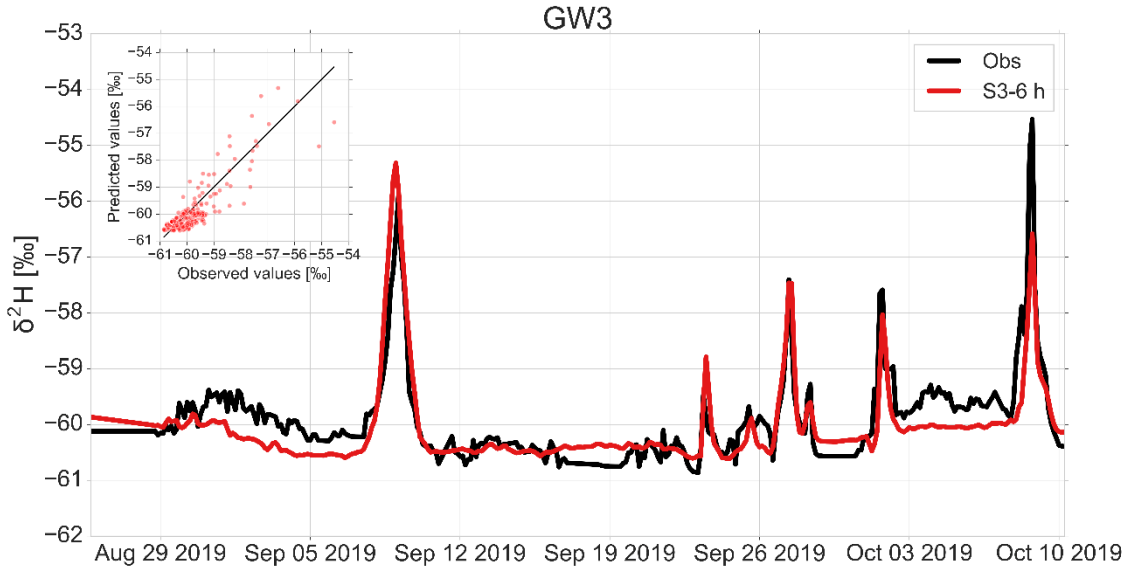

Figure 4.10 Comparison of observed and predicted δ^2H values by the best LSTM using a sequence length of 6 h for water temperature, pH and electrical conductivity measurements (scenario S3-6 h) with an RMSE of 0.4‰, MAE of 0.3‰, R^2 of 0.7 and NSE of 0.7 at GW3 site.

The power of LSTM as a deep learning model is that it can synthesize information from multiple sampling sites and situations into a single model. An LSTM trained on multiple sites under different landuse and hillslope conditions can learn different patterns of isotopic behaviour. It is evident from Figures 4.9-4.10 and Figures S 4.1-4.3 that the LSTM performs efficiently at the stream as well as groundwater sites. Since we trained a single LSTM on stream and groundwater sites with different landuse and hillslope positions with a range of different isotope concentrations and magnitudes of response to precipitation events, the LSTM tends to balance the error so that it minimizes the error between predicted and observed isotope concentrations for all of the sampling sites at the same time. The trained LSTM can be used to spatially estimate the isotope concentrations for multiple sites across the catchment with only simple-to-measure tracers. This leads not only to a substantial reduction of the cost of measurements, but also to an increase in the spatiotemporal knowledge of hydrological processes in the catchment.

One of the challenging tasks in hydrology is the spatial transferability of the models, particularly to data-scarce catchments. Most of hydrological models need to be rebuilt from scratch using newly collected data when the distribution of data is changed in the feature space. However, collecting adequate data is still challenging in many catchments due to the time consuming and expensive

measurement procedures. “Transfer learning” is a powerful technique of deep learning that reduces the need and effort to collect the data in data-scarce catchments. Transfer learning is a method that transfers the knowledge obtained in the source domain to the target domain when the latter has few data (Pan and Yang, 2010). Generally, if the deep learning model exhibits satisfying performance in the source catchment, it can be then generalized to target catchments and achieve good prediction performances without much additional data. The transfer learning is efficient in the LSTM because the hidden layers that have been trained to digest shape information, are still effective even when applied to data-scarce catchments (Shen, 2018). In order to apply the proposed LSTM to another catchment, the last fully connected layers are trained on the new data with initial random weights (Yosinski et al., 2014). Although the data is different from the one on which the model is already trained, the low-level features are similar. Therefore, transferring parameters from the “pre-trained LSTM” can provide the new target model with powerful feature extraction ability and reduce the data demand as well as computation or monitoring costs.

The prediction performance of LSTM and in general deep learning models could be enhanced when having more training data available (Schmidhuber, 2015). Providing longer historical data from multiple years for the training process in future research may allow the model to capture long-term seasonal trends under various flow regimes and hydroclimatic conditions, and hence extract more related information that can effectively reproduce the dynamic of the isotope concentrations. It will be particularly interesting to investigate how extreme weather events impact on the LSTM outcome. For the future, we might be able to test this given the current rather wet weather patterns of 2021, which are substantially different to the study period of our current work, which considers rather dry weather conditions from 2018 and 2019.

4.4 Conclusion

This study investigates the application of Long Short-Term Memory (LSTM) deep learning model to predict high-resolution (3 h) isotope concentrations of multiple stream and groundwater sources under different landuse and hillslope positions in the Schwingbach Environmental Observatory (SEO), Germany. The core objective of this study is to explore the prediction performance of an LSTM that is trained on multiple sites, with a set of explanatory data that are more straightforward and less expensive to measure compared to the stable isotopes of water. The explanatory data

consist of meteorological data, catchment wetness conditions and natural tracers (i.e., water temperature, pH and electrical conductivity). This study further conducts a sensitivity analysis to examine how different input features and their sequence lengths affect the performance of the LSTM. The ability of the LSTM that inherently considers the impact of environmental factors such as evaporation and evapotranspiration on fractionation of water isotopes without an explicit representation of the underlying processes provides the opportunity to efficiently apply the proposed model to isotope hydrology. The result showed that the LSTM could successfully predict stable isotopes of stream and groundwater sites when using only short-term sequence (6 h) of measured water temperature, pH and electrical conductivity. The LSTM prediction can be utilized to predict and interpolate the continuous isotope concentration time series either for data gap filling or in case where no continuous data acquisition is feasible. This is very valuable in practice because measurements of these tracers are still much cheaper than stable isotopes of water and can be continuously carried out with relatively minor maintenance.

For the future research, we will be collecting more data to enhance the prediction performance of the LSTM. There is still room to capture the peak isotope concentrations better. New input features like groundwater table should be used to potentially improve the model performance in the future. The data-hungry nature of deep learning models is a potential barrier for applying them in data-scarce catchments. Since many catchments of potential application may lack the length of the data available in this work, the sensitivity of prediction performance to the length of the training data warrants further investigation. The use of pre-trained LSTM is a promising approach to mitigate the large demand for data in a single catchment. The future research direction also includes applying the LSTM to predict high-resolution water quality parameters such as nitrate, pH and water temperature.

Supplementary Materials

Table S 4. 1 Optimized hyperparameters for the sensitivity analysis of the LSTM model.

	Sequence length (h)								
Scenario	**6**	**12**	**24**	**72**	**168**	**336**	**720**	**1440**	**2160**
S1	NH=40 DR=0.3 LR=1e-5 NE=80 BS=32	NH=40 DR=0.3 LR=1e-5 NE=80 BS=32	NH=50 DR=0.3 LR=1e-4 NE=10 BS=128	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=50 DR=0.3 LR=1e-4 NE=10 BS=128	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=80 DR=0.2 LR=1e-3 NE=70 BS=256	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128
S2	NH=80 DR=0.2 LR=1e-3 NE=70 BS=256	NH=80 DR=0.2 LR=1e-3 NE=70 BS=256	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=70 DR=0.3 LR=1e-2 NE=70 BS=64	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=70 DR=0.1 LR=1e-2 NE=90 BS=64	NH=70 DR=0.3 LR=1e-2 NE=70 BS=64
S3	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=70 DR=0.1 LR=1e-2 NE=90 BS=64	NH=70 DR=0.3 LR=1e-2 NE=70 BS=64	NH=70 DR=0.3 LR=1e-2 NE=70 BS=64
S4	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128	NH=60 DR=0.3 LR=1e-2 NE=35 BS=128	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128	NH=80 DR=0.2 LR=1e-3 NE=70 BS=256	NH=70 DR=0.3 LR=1e-2 NE=70 BS=64
S5	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128	NH=70 DR=0.1 LR=1e-2 NE=90 BS=64	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32
S6	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=80 DR=0.2 LR=1e-3 NE=70 BS=256
S7	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-4 NE=40 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=60 DR=0.4 LR=1e-3 NE=70 BS=32	NH=70 DR=0.4 LR=1e-4 NE=90 BS=128

Note: NH, DR, LR, NE and BS stand for number of hidden units, dropout rate, learning rate, number of epochs and batch size, respectively.

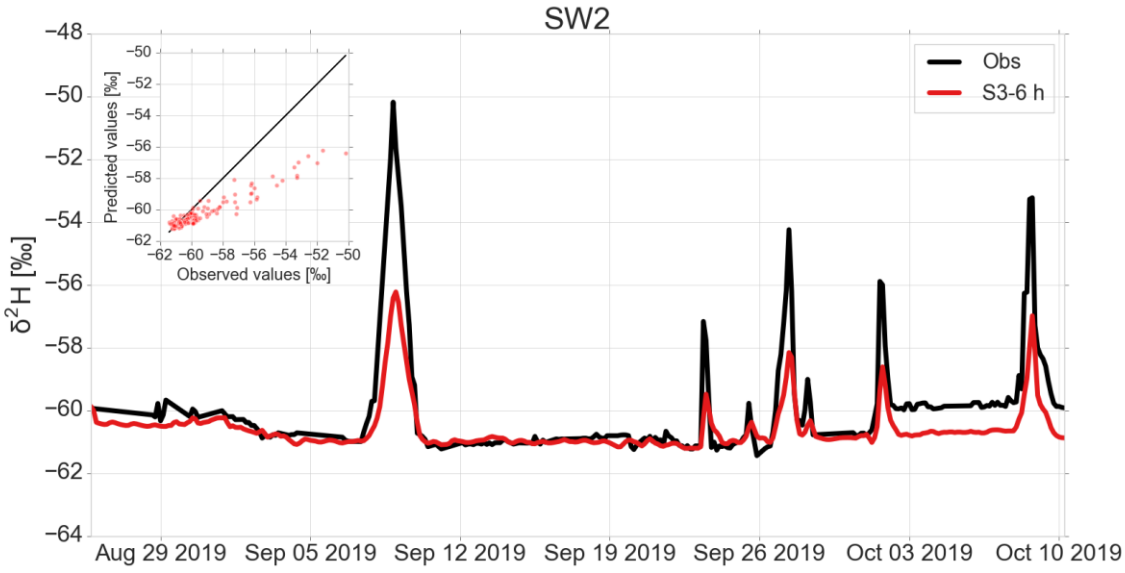

Figure S 4. 1 Comparison of observed and predicted δ^2H values by the best LSTM using a sequence length of 6 h for water temperature, pH and electrical conductivity measurements (scenario S3-6 h) with an RMSE of 1.0‰, MAE of 0.6‰, R^2 of 0.9 and NSE of 0.6 at SW2 site.

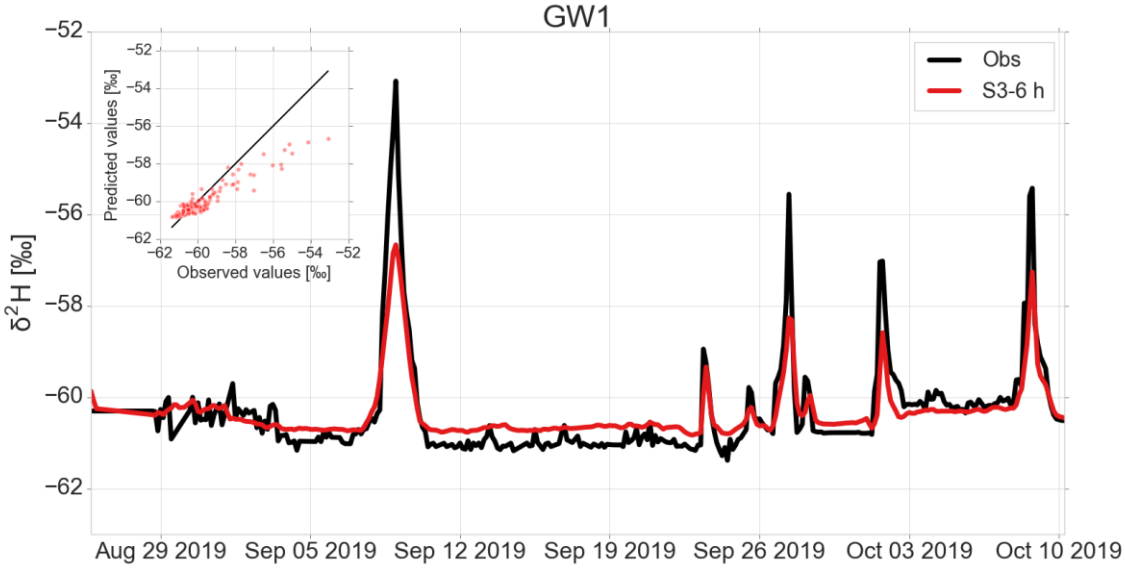

Figure S 4. 2 Comparison of observed and predicted δ^2H values by the best LSTM using a sequence length of 6 h for water temperature, pH and electrical conductivity measurements (scenario S3-6 h) with an RMSE of 0.5‰, MAE of 0.3‰, R^2 of 0.9 and NSE of 0.8 at GW1 site.

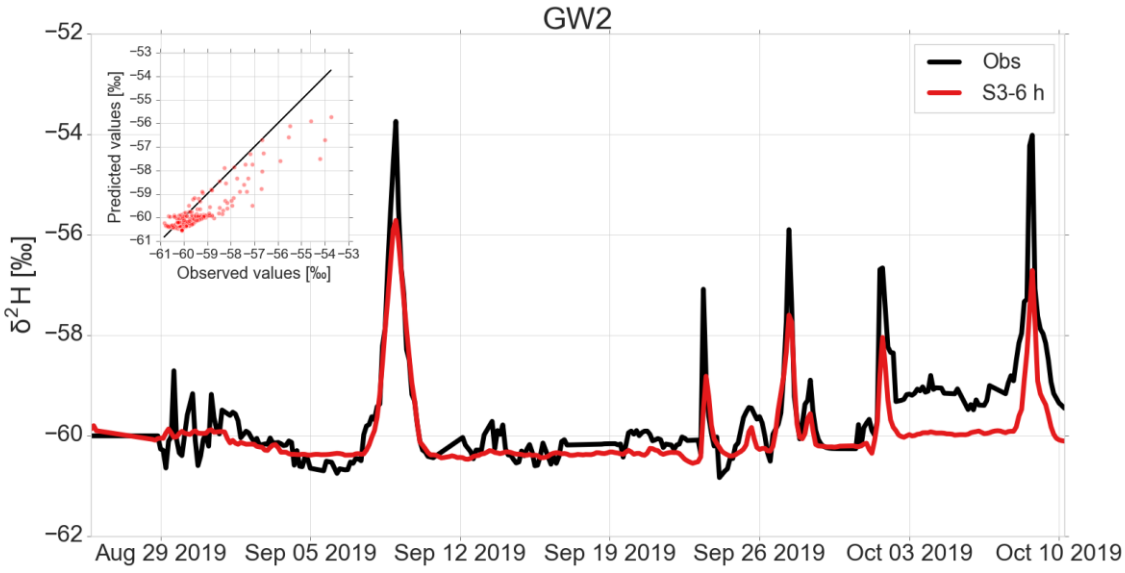

Figure S 4. 3 Comparison of observed and predicted δ^2H values by the best LSTM using a sequence length of 6 h for water temperature, pH and electrical conductivity measurements (scenario S3-6 h) with an RMSE of 0.6‰, MAE of 0.4‰, R^2 of 0.8 and NSE of 0.7 at GW2 site.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Preprint arXiv1603.04467*.

Abdollahi, S., Raeisi, J., Khalilianpour, M., Ahmadi, F., and Kisi, O. (2017). Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. *Water Resour. Manag.* 31, 4855–4874. doi:10.1007/s11269-017-1782-7.

Adeyemi, O., Grove, I., Peets, S., Domun, Y., and Norton, T. (2018). Dynamic neural network modelling of soil moisture content for predictive irrigation scheduling. *Sensors* 18:3408 doi:10.3390/s18103408.

Adnan, R. M., Khosravinia, P., Karimi, B., and Kisi, O. (2021a). Prediction of hydraulics performance in drain envelopes using Kmeans based multivariate adaptive regression spline. *Appl. Soft Comput.* 100:107008. doi:10.1016/j.asoc.2020.107008.

Adnan, R. M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., and Li, B. (2019a). Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrol.* 586: 124371. doi:10.1016/j.jhydrol.2019.124371.

Adnan, R. M., Liang, Z., Parmar, K. S., Soni, K., and Kisi, O. (2021b). Modeling monthly streamflow in mountainous basin by MARS, GMDH-NN and DENFIS using hydroclimatic data. *Neural Comput. Appl.* 33, 2853–2871. doi:10.1007/s00521-020-05164-3.

Adnan, R. M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., and Kisi, O. (2019b). Daily streamflow prediction using optimally pruned extreme learning machine. *J. Hydrol.* 577: 123981. doi:10.1016/j.jhydrol.2019.123981.

Agyare, W. A., Park, S. J., and Vlek, P. L. G. (2007). Artificial neural network estimation of saturated hydraulic conductivity. *Vadose Zo. J.* 6, 423–431. doi:10.2136/vzj2006.0131.

Ahmad, S., Kalra, A., and Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour.* 33, 69–80.

doi:10.1016/j.advwatres.2009.10.008.

Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., et al. (2016). Theano: A Python framework for fast computation of mathematical expressions. *arXiv Preprint arXiv1605.02688v1*.

Araya, S. N., and Ghezzehei, T. A. (2019). Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. *Water Resour. Res.* 55, 5715–5737. doi:10.1029/2018WR024357.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000). Artificial neural networks in hydrology. I: preliminary concepts. *J. Hydrol. Eng.* 5, 115–123. doi:10.1061/(ASCE)1084-0699(2000)5:2(115).

Aubert, A. H., Thrun, M. C., Breuer, L., and Ultsch, A. (2016). Knowledge discovery from high-frequency stream nitrate concentrations: Hydrology and biology contributions. *Sci. Rep.* 6:31536. doi:10.1038/srep31536.

Balistrocchi, M., and Bacchi, B. (2011). Modelling the statistical dependence of rainfall event variables through copula functions. *Hydrol. Earth Syst. Sci.* 15, 1959–1977. doi:10.5194/hess-15-1959-2011.

Behzad, M., Asghari, K., and Coppola, E. A. (2010). Comparative study of SVMs and ANNs in aquifer water level prediction. *J. Comput. Civ. Eng.* 24, 408–413. doi:10.1061/(ASCE)CP.1943-5487.0000043.

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., and Popp, J. (2013). Sample size planning for classification models. *Anal. Chim. Acta* 760, 25–33. doi:10.1016/j.aca.2012.11.007.

Benardos, P. G., and Vosniakos, G. C. (2007). Optimizing feedforward artificial neural network architecture. *Eng. Appl. Artif. Intell.* 20, 365–382. doi:10.1016/j.engappai.2006.06.005.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems,* 24.

Bergstra, J., Yamins, D., and Cox, D. (2013). "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *International Conference on Machine Learning*, 115–123.

Berman, E. S., Gupta, M., Gabrielli, C., Garland, T., and McDonnell, J. J. (2009). High-frequency field-deployable isotope analyzer for hydrological applications. *Water Resour. Res.* 45. doi:10.1029/2009WR008265.

Beven, K., and Germann, P. (1982). Macropores and water flow in soils. *Water Resour. Res.* 18, 1311–1325.

Birkel, C., Soulsby, C., Tetzlaff, D., Dunn, S., and Spezia, L. (2012). High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles. *Hydrol. Process.* 26, 308–316. doi:10.1002/hyp.8210.

Blume, T., Zehe, E., and Bronstert, A. (2007). Rainfall—runoff response, event-based runoff coefficients and hydrograph separation. *Hydrol. Sci. J.* 52, 843–862. doi:10.1623/hysj.52.5.843.

Bowden, G. J., Maier, H. R., and Dandy, G. C. (2002). Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* 38. doi:10.1029/2001wr000266.

Brielmann, H. (2008). *Recharge and Discharge Mechanism and Dynamics in the Mountainous Northern Upper Jordan River Catchment, Israel*. Available at: https://edoc.ub.uni-muenchen.de/9972/1/Brielmann_Heike.pdf.

Brigato, L., and Iocchi, L. (2020). A Close Look at deep learning with small data. *arXiv PreprintarXiv2003.12843.*

Brown, V. A., McDonnell, J. J., Burns, D. A., and Kendall, C. (1999). The role of event water, a rapid shallow flow component, and catchment size in summer stormflow. *J. Hydrol.* 217, 171–190. doi:https://doi.org/10.1016/S0022-1694(98)00247-9.

Brownlee, J. (2020). *Repeated k-Fold Cross-Validation for Model Evaluation in Python*. Available online at: https://machinelearningmastery.com/repeated-k-fold-cross-validation-with-python/ [Accessed March 19, 2021].

Buttle, J. M. (1994). Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins. *Prog. Phys. Geogr.* 18, 16–41. doi:https://doi.org/10.1177/030913339401800102.

Calkoen, F., Luijendijk, A., Rivero, C. R., Kras, E., and Baart, F. (2021). Traditional vs. Machine-

learning methods for forecasting sandy shoreline evolution using historic satellite-derived shorelines. *Remote Sens*. 13. doi:10.3390/rs13050934.

Campolo, M., Andreussi, P., and Soldati, A. (1999). River flood forecasting with a neural network model. *Water Resour. Res.* 35, 1191–1197. doi:10.1029/1998WR900086.

Carey, S. K., and Quinton, W. L. (2005). Evaluating runoff generation during summer using hydrometric, stable isotope and hydrochemical methods in a discontinuous permafrost alpine catchment. *Hydrol. Process.* 19, 95–114. doi:10.1002/hyp.5764.

Cerar, S., Mezga, K., Žibret, G., Urbanc, J., and Komac, M. (2018). Comparison of prediction methods for oxygen-18 isotope composition in shallow groundwater. *Sci. Total Environ.* 631–632, 358–368. doi:10.1016/j.scitotenv.2018.03.033.

Chen, D., Lu, J., and Shen, Y. (2010). Artificial neural network modelling of concentrations of nitrogen, phosphorus and dissolved oxygen in a non-point source polluted river in Zhejiang Province, southeast China. *Hydrol. Process.* 24, 290–299. doi:10.1002/hyp.7482.

Chen, X., Huang, J., Han, Z., Gao, H., Liu, M., Li, Z., et al. (2020). The importance of short lag-time in the runoff forecasting model based on long short-term memory. *J. Hydrol.* 589, 125359. doi:10.1016/j.jhydrol.2020.125359.

Chin, R. J., Lai, S. H., Chang, K. B., Jaafar, W. Z. W., and Othman, F. (2016). Relationship between minimum inter-event time and the number of rainfall events in Peninsular Malaysia. *Weather* 71, 213–218. doi:10.1002/wea.2766.

Chollet, F. (2015). Keras. Available online at: https://keras.io.

Coopersmith, E. J., Minsker, B. S., Wenzel, C. E., and Gilmore, B. J. (2014). Machine learning assessments of soil drying for agricultural planning. *Comput. Electron. Agric.* 104, 93–104. doi:10.1016/j.compag.2014.04.004.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi:10.1007/BF00994018.

Das, S. K., Samui, P., and Sabat, A. K. (2012). Prediction of field hydraulic conductivity of clay liners using an artificial neural network and support vector machine. *Int. J. Geomech.* 12, 606–611. doi:10.1061/(asce)gm.1943-5622.0000129.

Dawson, C. W., and Wilby, R. L (1998). An artificial neural network approach to rainfall-runoff modelling. *Hydrol. Sci. J.* 43, 47–66. doi:10.1080/02626669809492102.

Dawson, C. W., and Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Prog. Phys. Geogr.* 25, 80–108. doi:10.1177/030913330102500104.

Duan, S., Ullrich, P., and Shu, L. (2020). Using convolutional neural networks for streamflow projection in California. *Front. Water* 2:28. doi:10.3389/frwa.2020.00028.

Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Holger Hoos, et al. (2013). Towards an empirical foundation for assessing bayesian optimization of hyperparameters. In *NIPS Workshop on Bayesian Optimazation in Theory and Practice* 10, 3.

Eijkelkamp RBC flumes (2013). Available online at: https://en.eijkelkamp.com/products/sensors-monitoring_uk/rbc-flumes.html [Accessed February 6, 2020].

Elbisy, M. S. (2015). Support Vector Machine and regression analysis to predict the field hydraulic conductivity of sandy soil. *KSCE J. Civ. Eng.* 19, 2307–2316. doi:10.1007/s12205-015-0210-x.

Eshleman, K. N., Pollard, J. S., and O'Brien, A. K. (1993). Determination of contributing areas for saturation overland flow from chemical hydrograph separations. *Water Resour. Res.* 29, 3577–3587. doi:10.1029/93WR01811.

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., and Jiang, J. (2020). Comparison of long short term memory networks and the hydrological model in runoff simulation. *Water* 12, 175. doi:10.3390/w12010175.

Fang, K., Shen, C., Kifer, D., and Yang, X. (2017). Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network. *Geophys. Res. Lett.* 44, 11,030–11,039. doi:10.1002/2017GL075619.

Feng, J., and Lu, S. (2019). Performance analysis of various activation functions in artificial neural networks. *J. Phys. Conf. Ser.* 1237:022030. doi:10.1088/1742-6596/1237/2/022030.

Feng, S., Zhou, H., and Dong, H. (2019). Using deep neural network with small dataset to predict material defects. *Mater. Des.* 162, 300–310. doi:10.1016/j.matdes.2018.11.060.

Feng, X., Ma, G., Su, S. F., Huang, C., Boswell, M. K., and Xue, P. (2020). A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. *Ocean Eng.* 211:107526. doi:10.1016/j.oceaneng.2020.107526.

Fischer, B. M. C., Stähli, M., and Seibert, J. (2017a). Pre-event water contributions to runoff events of different magnitude in pre-alpine headwaters. *Hydrol. Res.* 48, 28–47. doi:10.2166/nh.2016.176.

Fischer, B. M. C., van Meerveld, H. J., and Seibert, J. (2017b). Spatial variability in the isotopic composition of rainfall in a small headwater catchment and its effect on hydrograph separation. *J. Hydrol*. 547, 755–769. doi:10.1016/j.jhydrol.2017.01.045

Flood, I., and Kartam, N. (1994). Neural networks in civil engineering. I: principles and understanding. *J. Comput. Civ. Eng.* 8, 131–148. doi:10.1061/(ASCE)0887-3801(1994)8:2(131).

Gaál, L., Molnar, P., and Szolgay, J. (2014). Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland. *Hydrol. Earth Syst. Sci.* 18, 1561–1573. doi:10.5194/hess-18-1561-2014.

Genereux, D. (1998). Quantifying uncertainty in tracer-based hydrograph separations. *Water Resour. Res.* 34, 915–919. doi:https://doi.org/10.1029/98WR00010.

Genereux, D. P., and Hooper, R. P. (1998). "Oxygen and hydrogen isotopes in rainfall-runoff studies," in *Isotope Tracers in Catchment Hydrology,* eds C. Kendall and J. J McDonnell (Amsterdam: Elsevier), 319–346.

Géron, A. (2019). *Hands-On Machine Learning With Scikit-Learn, Keras, And TensorFlow: Concepts, Tools, And Techniques To Build Intelligent Systems*. Sebastopol, CA: O'Reilly Media.

Gers, F. A., Eck, D., and Schmidhuber, J. (2002). "Applying LSTM to time series predictable through time-window approaches", in *Neural Nets WIRN Vietri-01* (London: Springer), 193–200.

Gill, M. K., Asefa, T., Kemblowski, M. W., and McKee, M. (2006). Soil moisture prediction using support vector machines. *J. Am. Water Resour. Assoc.* 42, 1033–1046. doi:10.1111/j.1752-

1688.2006.tb04512.x.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge: MIT Press.

Granata, F., Papirio, S., Esposito, G., Gargano, R., and de Marinis, G. (2017). Machine learning algorithms for the forecasting of wastewater quality indicators. *Water* 9:105. doi:10.3390/w9020105.

Hagan, M. T., Demuth, H. B., and Beale, M. H. (2014). *Neural Network Design*. Stillwater, OK: Martin Hagan.

Han, D., Chan, L., and Zhu, N. (2007). Flood forecasting using support vector machines. *J. Hydroinformatics* 9, 267–276. doi:10.2166/hydro.2007.027.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements Of Statistical Learning: Data Mining, Inference, And Prediction*. New York, NY: Springer Science & Business Media.

Haykin, S. (1994). *Neural Networks: A Aomprehensive Foundation*. Hoboken, NY: Prentice Hall PTR.

He, Z., Wen, X., Liu, H., and Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.* 509, 379–386. doi:10.1016/j.jhydrol.2013.11.054.

Heaton, J. (2015). *Artificial Intelligence For Humans, Volume 3: Deep Learning And Neural Networks*. Chesterfield, MO: Heaton Research, Inc.

Heinz, E., Kraft, P., Buchen, C., Frede, H.-G., Aquino, E., and Breuer, L. (2014). Set up of an automatic water quality sampling system in irrigation agriculture. *Sensors* 14, 212–228. doi:10.3390/s140100212.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hoeg, S., Uhlenbrook, S., and Leibundgut, C. (2000). Hydrograph separation in a mountainous catchment — combining hydrochemical and isotopic tracers. *Hydrol. Process.* 14, 1199–1216. doi:10.1002/(SICI)1099-1085(200005)14:7<1199::AID-HYP35>3.0.CO;2-K.

Hrachowitz, M., Bohte, R., Mul, M. L., Bogaard, T. A., Savenije, H. H. G., and Uhlenbrook, S.

(2011). On the value of combined event runoff and tracer analysis to improve understanding of catchment functioning in a data-scarce semi-arid area. *Hydrol. Earth Syst. Sci.* 15, 2007–2024. doi:10.5194/hess-15-2007-2011.

Huang, G. Bin (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Trans. Neural Networks* 14, 274–281. doi:10.1109/TNN.2003.809401.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *IEEE Ann. Hist. Comput.* 9, 90–95. doi:10.1109/MCSE.2007.55.

Jain, S. K. (2012). Modeling river stage-discharge-sediment rating relation using support vector regression. *Hydrol. Res.* 43, 851–861. doi:10.2166/nh.2011.101.

James, A. L., and Roulet, N. T. (2009). Antecedent moisture conditions and catchment morphology as controls on spatial patterns of runoff generation in small forest catchments. *J. Hydrol.* 377, 351–366. doi:10.1016/j.jhydrol.2009.08.039.

Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J., and Pulido-Velazquez, D. (2018). A comparison of SWAT and ANN models for daily runoff simulation in different climatic zones of Peninsular Spain. *Water* 10:192. doi:10.3390/w10020192.

Jones, J. A. A. (2010). Soil piping and catchment response. *Hydrol. Process.* 24, 1548–1566. doi:10.1002/hyp.7634.

Jones, J. P., Sudicky, E. A., Brookfield, A. E., and Park, Y.-J. (2006). An assessment of the tracer-based approach to quantifying groundwater contributions to streamflow. *Water Resour. Res.* 42. doi:10.1029/2005WR004130.

Kardan Moghaddam, H., Ghordoyee Milan, S., Kayhomayoon, Z., Rahimzadeh kivi, Z., and Arya Azar, N. (2021). The prediction of aquifer groundwater level based on spatial clustering approach using machine learning. *Environ. Monit. Assess.* 193:173. doi:10.1007/s10661-021-08961-y.

Karsoliya, S. (2012). Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *Int. J. Eng. Trends Technol.* 3, 714–717.

Karunanithi, N., Grenney, W. J., Whitley, D., and Bovee, K. (1994). Neural networks for river flow prediction. *J. Comput. Civ. Eng.* 8, 201–220. doi:10.1061/(ASCE)0887-3801(1994)8:2(201).

Kendall, C., and McDonnell, J. J. (2012). *Isotope Tracers in Catchment Hydrology*. (Amsterdam: Elsevier).

Khashei, M., and Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Syst. Appl.* 37, 479–489. doi:10.1016/j.eswa.2009.05.044.

Khashei, M., Bijari, M., and Raissi Ardali, G. A. (2009). Improvement of auto-regressive integrated moving average models using fuzzy logic and Artificial Neural Networks (ANNs). *Neurocomputing* 72, 956–967. doi:10.1016/j.neucom.2008.04.017.

Khovanova, N. A., Shaikhina, T., and Mallick, K. K. (2015). Neural networks for analysis of trabecular bone in osteoarthritis. *Bioinspired, Biomim. Nanobiomaterials* 4, 90–100. doi:10.1680/bbn.14.00006.

Kim, J. H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput. Stat. Data Anal.* 53, 3735–3745. doi:10.1016/j.csda.2009.04.009.

Kingma, D. P., and Ba, J. L. (2014). Adam: a method for stochastic optimization. *arXiv Preprint arXiv1412.6980*.

Kirchner, J. W., Feng, X., Neal, C., and Robson, A. J. (2004). The fine structure of water-quality dynamics: the (high-frequency) wave of the future. *Hydrol. Process.* 18, 1353–1359. doi:10.1002/hyp.5537.

Kisi, O., and Aytek, A. (2013). Explicit neural network in suspended sediment load estimation. *Neural Netw. World* 23, 587–607. doi:10.14311/NNW.2013.23.035.

Kisi, O., and Parmar, K. S. (2016). Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J. Hydrol.* 534, 104–112. doi:10.1016/j.jhydrol.2015.12.014.

Klaus, J., and McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: review and evaluation. *J. Hydrol.* 505, 47–64. doi:10.1016/j.jhydrol.2013.09.006.

Klein, A., and Hutter, F. (2019). Tabular benchmarks for joint architecture and hyperparameter optimization. *arXiv Preprint arXiv1905.04970*.

Knoll, L., Breuer, L., and Bach, M. (2020). Nation-wide estimation of groundwater redox

conditions and nitrate concentrations through machine learning. *Environ. Res. Lett.* 15:064004. doi:10.1088/1748-9326/ab7d5c.

Koeniger, P., Gaj, M., Beyer, M., and Himmelsbach, T. (2016). Review on soil water isotope-based groundwater recharge estimations. *Hydrol. Process.* 30, 2817–2834. doi:10.1002/hyp.10775.

Kohavi, R. (1995). A Study of cross-validation and bootstrap for accuracy estimation and model selection. *Int. Jt. Conf. Artif. Intell.* 14, 1137–1145.

Kong, Y., and Pang, Z. (2010). Isotope hydrograph separation in alpine catchments: A review. *J. Glaciol. Geocryol.* 32, 619–625. doi:10.3724/SP.J.1226.2011.00086.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* 22, 6005–6022. doi: 10.5194/hess-22-6005-2018.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G. (2019). Towards learning universal , regional , and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* 23, 5089–5110. doi: 10.5194/hess-23-5089-2019.

Kuhn, M., and Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer.

Kuppel, S., Tetzlaff, D., Maneta, M. P., and Soulsby, C. (2018). EcH2O-iso 1.0: Water isotopes and age tracking in a process-based, distributed ecohydrological model. *Geosci. Model Dev.* 11, 3045–3069. doi:10.5194/gmd-11-3045-2018.

Lange, H., and Sippel, S. (2020). “Machine learning applications in hydrology,” in *Forest-Water Interactions*, eds D. F. Levia, D. E. Carlyle-Moses, S. Iida, and B. Michalzik (Cham: Springer), 233–257.

Larochelle, H., Erhan, D., Courville, A., Bergstra, J., and Bengio, Y. (2007). "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proceedings of the 24th international conference on Machine learning* (Corvallis, Or), 473–480.

Lauer, F., Frede, H. G., and Breuer, L. (2013). Uncertainty assessment of quantifying spatially concentrated groundwater discharge to small streams by distributed temperature sensing. *Water Resour. Res.* 49, 400–407. doi:10.1029/2012WR012537.

Le, X., Ho, H. V., Lee, G., and Jung, S. (2019). Application of Long Short-Term Memory (LSTM) neural network for flood forecasting. *Water* 11, 1387. doi:10.3390/w11071387.

LeBaron, B., and Weigend, A. S. (1998). A bootstrap evaluation of the effect of data splitting on financial time series. *IEEE Trans. Neural Networks* 9, 213–220. doi:10.1109/72.655043.

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539.

LeCun, Y., Bottou, L., Orr, G. B., and Müller, K. R. (2012). "Efficient Backpro," in *Neural Networks: Tricks of the Trade*, eds G. Montavon, G. B. Orr, and K. -R. Müller (Berlin, Heidelberg: Springer), 9–48.

Li, B., Yang, G., Wan, R., Dai, X., and Zhang, Y. (2016). Comparison of random forests and other statistical methods for the prediction of lake water level: A case study of the Poyang Lake in China. *Hydrol. Res.* 47, 69–83. doi:10.2166/nh.2016.264.

Lin, J. Y., Cheng, C. T., and Chau, K. W. (2006). Using support vector machines for long-term discharge prediction. *Hydrol. Sci. J.* 51, 599–612. doi:10.1623/hysj.51.4.599.

Lis, G., Wassenaar, L. I., and Hendry, M. J. (2008). High-precision laser spectroscopy D/H and 18O/16O measurements of microliter natural water samples. *Anal. Chem.* 80, 287–293. doi:10.1021/ac701716q.

Liu, M., and Lu, J. (2014). Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? *Environ. Sci. Pollut. Res.* 21, 11036–11053. doi:10.1007/s11356-014-3046-x.

Liu, P., Wang, J., Sangaiah, A., Xie, Y., and Yin, X. (2019a). Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability* 11, 2058. doi:10.3390/su11072058.

Liu, Y., Zhao, Q., Yao, W., Ma, X., Yao, Y., and Liu, L. (2019b). Short-term rainfall forecast model based on the improved BP–NN algorithm. *Sci. Rep.* 9, 1–12. doi:10.1038/s41598-019-56452-5.

Mahindawansha, A., Breuer, L., Chamorro, A., and Kraft, P. (2018). High-frequency water isotopic analysis using an automatic water sampling system in rice-based cropping systems. *Water* 10,

1327. doi:10.3390/w10101327.

Maier, H. R., and Dandy, G. C. (1998). The effect of internal parameters and geometry on the performance of back-propagation neural networks: An empirical study. *Environ. Model. Softw.* 13, 193–209. doi:10.1016/S1364-8152(98)00020-6.

Maier, H. R., and Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Model. Softw. 15*, 101–124. doi:10.1016/S1364-8152(99)00007-9.

Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environ. Model. Softw.* 25, 891–909. doi:10.1016/j.envsoft.2010.02.003.

Malekzadeh, M., Kardar, S., and Shabanlou, S. (2019). Simulation of groundwater level using MODFLOW, extreme learning machine and wavelet-extreme learning machine models. *Groundw. Sustain. Dev.* 9:100279. doi:10.1016/j.gsd.2019.100279.

Malik, A., Tikhamarine, Y., Souag-Gamane, D., Kisi, O., and Pham, Q. B. (2020). Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. *Stoch. Environ. Res. Risk Assess.* 34, 1755–1773. doi:10.1007/s00477-020-01874-1.

Matei, O., Rusu, T., Petrovan, A., and Mihuţ, G. (2017). A data mining system for real time soil moisture prediction. *Procedia Eng.* 181, 837–844. doi:10.1016/j.proeng.2017.02.475.

McDonnell, J. J. (1990). A rationale for old water discharge through macropores in a steep, humid catchment. *Water Resour. Res.* 26, 2821–2832. doi:https://doi.org/10.1029/WR026i011p02821.

Mcdonnell, J. J., Bonell, M., Stewart, M. K., and Pearce, A. J. (1990). Deuterium variations in storm rainfall: implications for stream hydrograph separation. *Water Resour. Res.* 26, 455–458. doi:10.1029/WR026i003p00455.

McGlynn, B. L., and McDonnell, J. J. (2003). Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resour. Res.* 39. doi:10.1029/2003WR002091.

McGlynn, B. L., McDonnell, J. J., Seibert, J., and Kendall, C. (2004). Scale effects on headwater

catchment runoff timing, flow sources, and groundwater-streamflow relations. *Water Resour. Res.* 40. doi:10.1029/2003WR002494.

McGuire, K. J., and McDonnell, J. J. (2006). A review and evaluation of catchment transit time modeling. *J. Hydrol.* 330, 543–563. doi:10.1016/j.jhydrol.2006.04.020.

McKinney, W. (2010). "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference (*Austin, TX), 56–61.

Merz, R., and Blöschl, G. (2009). A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resour. Res.* 45. doi:10.1029/2008WR007163.

Meyal, A. Y., Versteeg, R., Alper, E., and Johnson, D. (2020). Automated cloud based long short-term memory neural network based SWE prediction. *Front. Water* 2:574917. doi:10.3389/frwa.2020.574917.

Minns, A. W., and Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrol. Sci. J.* 41, 399–417. doi:10.1080/02626669609491511.

Mohammadpour, R., Shaharuddin, S., Chang, C. K., Zakaria, N. A., Ghani, A. A., and Chan, N. W. (2015). Prediction of water quality index in constructed wetlands using support vector machine. *Environ. Sci. Pollut. Res.* 22, 6208–6219. doi:10.1007/s11356-014-3806-7.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. Am. Soc. Agric. Biol. Eng.* 50, 885–900. doi:10.13031/2013.23153.

Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., et al. (2020). Surrogate optimization of deep neural networks for groundwater predictions. *J. Glob. Optim.* doi:10.1007/s10898-020-00912-0.

Muñoz-Villers, L. E., and McDonnell, J. J. (2012). Runoff generation in a steep, tropical montane cloud forest catchment on permeable volcanic substrate. *Water Resour. Res.* 48. doi:10.1029/2011WR011316.

Najah Ahmed, A., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., et al. (2019). Machine learning methods for better water quality

prediction. *J. Hydrol.* 578:124084. doi:10.1016/j.jhydrol.2019.124084.

Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F., and Chandran, V. (2018). Long short term memory hyperparameter optimization for a neural network based emotion recognition framework. *IEEE Access* 6, 49325–49338. doi:10.1109/ACCESS.2018.2868361.

Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* 10, 282–290. doi:10.1016/0022-1694(70)90255-6.

Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv Preprint arXiv1811.03378*.

Orlowski, N., Kraft, P., Pferdmenges, J., and Breuer, L. (2016). Exploring water cycle dynamics by sampling multiple stable water isotope pools in a developed landscape in Germany. *Hydrol. Earth Syst. Sci.* 20, 3873–3894. doi:10.5194/hess-20-3873-2016.

Orlowski, N., Lauer, F., Kraft, P., Frede, H.-G., and Breuer, L. (2014). Linking spatial patterns of groundwater table dynamics and streamflow generation processes in a small developed catchment. *Water* 6, 3085–3117. doi:10.3390/w6103085.

Oyebode, O., and Stretch, D. S. (2018). Neural network modeling of hydrological systems : A review of implementation techniques. *Nat. Resour. Model.* 32:e12189. doi:10.1002/nrm.12189.

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng*. 22, 1345–1359. doi:10.1109/TKDE.2009.191.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Pellerin, B. A., Wollheim, W. M., Feng, X., and Vörösmarty, C. J. (2008). The application of electrical conductivity as a tracer for hydrograph separation in urban catchments. *Hydrol. Process.* 22, 1810–1818. doi:10.1002/hyp.6786.

Penna, D., Hopp, L., Scandellari, F., Allen, S. T., Benettin, P., Beyer, M., et al. (2018). Tracing ecosystem water fluxes using hydrogen and oxygen stable isotopes: challenges and

opportunities from an interdisciplinary perspective. *Biogeosciences.* 15, 6399–6415. doi:10.5194/bg-15-6399-2018.

Penna, D., Meerveld, H. J. T., Gobbi, A., Borga, M., and Dalla Fontana, G. (2011). The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment. *Hydrol. Earth Syst. Sci.* 15, 689–702. doi:10.5194/hess-15-689-2011.

Penna, D., van Meerveld, H. J., Oliviero, O., Zuecco, G., Assendelft, R. S., Dalla Fontana, G., et al. (2015). Seasonal changes in runoff generation in a small forested mountain catchment. *Hydrol. Process.* 29, 2027–2042. doi:10.1002/hyp.10347.

Penna, D., van Meerveld, H. J., Zuecco, G., Dalla Fontana, G., and Borga, M. (2016). Hydrological response of an Alpine catchment to rainfall and snowmelt events. *J. Hydrol.* 537, 382–397. doi:10.1016/j.jhydrol.2016.03.040.

Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., and Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *J. Hydrol.* 529, 302–315. doi:10.1016/j.jhydrol.2015.07.044.

Poornima, S., and Pushpalatha, M. (2019). Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units. *Atmosphere* 10:668. doi:10.3390/atmos10110668.

Prasetya, E. P., and Djamal, E. C. (2019). "Rainfall forecasting for the natural disasters preparation using recurrent neural networks," in *International Conference on Electrical Engineering and Informatics (ICEEI)* (Bandung), 52–57.

Pumperla, M. (2019). Hyperas. Available online at: https://pypi.org/project/hyperas/ [Accessed May 10, 2021].

Qiu, R., Wang, Y., Wang, D., Qiu, W., Wu, J., and Tao, Y. (2020). Water temperature forecasting based on modified artificial neural network methods: Two cases of the Yangtze River. *Sci. Total Environ.* 737:139729. doi:10.1016/j.scitotenv.2020.139729.

Quade, M., Klosterhalfen, A., Graf, A., Brüggemann, N., Hermes, N., Vereecken, H., et al. (2019). In-situ monitoring of soil water isotopic composition for partitioning of evapotranspiration during one growing season of sugar beet (Beta vulgaris). *Agric. For. Meteorol.* 266-267, 53–

64. doi:10.1016/j.agrformet.2018.12.002.

Quan, Q., Hao, Z., Xifeng, H., and Jingchun, L. (2020). Research on water temperature prediction based on improved support vector regression. *Neural Comput. Appl* 1-10. doi:10.1007/s00521-020-04836-4.

Raghavendra, S., and Deka, P. C. (2014). Support vector machine applications in the field of hydrology: a review. *Appl. Soft Comput.* 19, 372–386. doi:10.1016/j.asoc.2014.02.002.

Rasouli, K., Hsieh, W. W., and Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol.* 414–415, 284–293. doi:10.1016/j.jhydrol.2011.10.039.

Refaeilzadeh, P., Lei, T., and Liu, H. (2009). "Cross-validation," in *Encyclopedia of Database Systems.* eds L. Liu and M. T. Özsu (Boston, MA: Springer). doi:10.1007/978-0-387-39940-9_565.

Reimers, N., and Gurevych, I. (2017). Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. *arXiv Preprint arXiv1707.06799.*

Renshaw, C. E., Feng, X., Sinclair, K. J., and Dums, R. H. (2003). The use of stream flow routing for direct channel precipitation with isotopically-based hydrograph separations : the role of new water in stormflow generation. *J. Hydrol.* 273, 205–216. doi:10.1016/S0022-1694(02)00392-X.

Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., et al. (2019). Deep learning in medical imaging and radiation therapy. *Med. Phys.* 46, e1–e36. doi:10.1002/mp.13264.

Sahoo, S., Russo, T. A., Elliott, J., and Foster, I. (2017). Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. *Water Resour. Res.* 53, 3878–3895. doi:10.1002/2016WR019933.

Sahraei, A., Chamorro, A., Kraft, P., and Breuer, L. (2021). Application of machine learning models to predict maximum event water fractions in streamflow. *Front. Water* 3:652100. doi:10.3389/frwa.2021.652100.

Sahraei, A., Kraft, P., Windhorst, D., and Breuer, L. (2020). High-resolution , in situ monitoring

of stable isotopes of water revealed insight into hydrological behavior. *Water* 12:565. doi:10.3390/w12020565.

Sahu, R. K., Müller, J., Park, J., Varadharajan, C., Arora, B., Faybishenko, B., et al. (2020). Impact of input feature selection on groundwater level prediction from a multi-layer perceptron neural network. *Front. Water* 2:573034. doi:10.3389/frwa.2020.573034.

Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., et al. (2018). A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci. Total Environ.* 644, 954–962. doi:10.1016/j.scitotenv.2018.07.054.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* 3, 210–299. doi:10.1147/rd.33.0210.

Šanda, M., Sedlmaierová, P., Vitvar, T., Seidler, C., Kändler, M., Jankovec, J., et al. (2017). Pre-event water contributions and streamwater residence times in different land use settings of the transboundary mesoscale Lužická Nisa catchment. *J. Hydrol. Hydromechanics* 65, 154–164. doi:10.1515/johh-2017-0003.

Sattari, M. T., Mirabbasi, R., Sushab, R. S., and Abraham, J. (2018). Prediction of groundwater level in Ardebil plain using support vector regression and M5 tree model. *Groundwater* 56, 636–646. doi:10.1111/gwat.12620.

Scherrer, S., Naef, F., Faeh, A. O., and Cordery, I. (2007). Formation of runoff at the hillslope scale during intense precipitation. *Hydrol. Earth Syst. Sci.* 11, 907–922. doi: 10.5194/hess-11-907-2007.

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks* 61, 85–117. doi:10.1016/j.neunet.2014.09.003.

Schölkopf, B., Smola, A. J., and Bach, F. (2002). *Learning With Kernels: Support Vector Machines, Regularization, Optimization, And Beyond*. Cambridge, MA: MIT Press.

Senthil Kumar, A. R., Goyal, M. K., Ojha, C. S. P., Singh, R. D., and Swamee, P. K. (2013). Application of artificial neural network, fuzzylogic and decision tree algorithms for modelling of streamflow at Kasol in India. *Water Sci. Technol.* 68, 2521–2526. doi:10.2166/wst.2013.491.

Shahinfar, S., and Kahn, L. (2018). Machine learning approaches for early prediction of adult wool growth and quality in Australian Merino sheep. *Comput. Electron. Agric.* 148, 72–81. doi:10.1016/j.compag.2018.03.001.

Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., and Khovanova, N. (2015). Machine learning for predictive modelling based on small data in biomedical engineering. *IFAC-PapersOnLine* 48, 469–474. doi:10.1016/j.ifacol.2015.10.185.

Shanley, J. B., Kendall, C., Smith, T. E., Wolock, D. M., and McDonnell, J. J. (2002). Controls on old and new water contributions to stream flow at some nested catchments in Vermont , USA. *Hydrol. Process.* 16, 589–609. doi:10.1002/hyp.312.

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* 54, 8558–8593. doi:10.1029/2018WR022643.

Shortridge, J. E., Guikema, S. D., and Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol. Earth Syst. Sci.* 20, 2611–2628. doi:10.5194/hess-20-2611-2016.

Sidle, R. C., Tsuboyama, Y., Noguchi, S., Hosoda, I., Fujieda, M., and Shimizu, T. (2000). Stormflow generation in steep forested headwaters: a linked hydrogeomorphic paradigm. *Hydrol. Process.* 14, 369–385. doi:10.1002/(SICI)1099-1085(20000228)14:3<369::AID-HYP943>3.0.CO;2-P

Sihag, P. (2018). Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Model. Earth Syst. Environ.* 4, 189–198. doi:10.1007/s40808-018-0434-0.

Singh, A., Malik, A., Kumar, A., and Kisi, O. (2018). Rainfall-runoff modeling in hilly watershed using heuristic approaches with gamma test. *Arab. J. Geosci.* 11:261. doi:10.1007/s12517-018-3614-3.

Singh, K. P., Basant, N., and Gupta, S. (2011). Support vector machines in water quality management. *Anal. Chim. Acta* 703, 152–162. doi:10.1016/j.aca.2011.07.027.

Sklash, M. G., Farvolden, R. N., and Fritz, P. (1976). A conceptual model of watershed response to rainfall, developed through the use of oxygen-18 as a natural tracer. *Can. J. Earth Sci.* 13,

271–283. doi:10.1139/e76-029.

Smithson, S. C., Yang, G., Gross, W. J., and Meyer, B. H. (2016). "Neural networks designing neural networks: multi-objective hyper-parameter optimization," in *Proceedings of the 35th International Conference on Computer-Aided Design* (Austin, TX), 1–8.

Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Stat. Comput.* 14, 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

Sontag, E. D. (1992). Feedback stabilization using two-hidden-layer nets. *IEEE Trans. Neural Networks* 3, 981–990. doi:10.1109/72.165599.

Sprenger, M., Leistert, H., Gimbel, K., and Weiler, M. (2016). Illuminating hydrological processes at the soil-vegetation- atmosphere interface with water stable isotopes. *Rev. Geophys.* 54, 674–704. doi:10.1002/ 2015RG000515.

Stadnyk, T. A., Delavau, C., Kouwen, N., and Edwards, T. W. D. (2013). Towards hydrological model calibration and validation: simulation of stable water isotopes using the isoWATFLOOD model. *Hydrol. Process.* 27, 3791–3810. doi:10.1002/hyp.9695.

Szandała, T. (2020). “Review and comparison of commonly used activation functions for deep neural networks,” in *Bio-inspired Neurocomputing* (Singapore: Springer), 203–224.

Tamura, S., and Tateishi, M. (1997). Capabilities of a four-layered feedforward neural network: four layers versus three. *IEEE Trans. Neural Networks* 8, 251–255. doi:10.1109/72.557662.

Tanty, R., and Desmukh, T. S. (2015). Application of artificial neural network in hydrology- a review. *Int. J. Eng. Res. Technol.* 4. 184-188. doi:10.17577/ijertv4is060247.

Tapoglou, E., Karatzas, G. P., Trichakis, I. C., and Varouchakis, E. A. (2014). A spatio-temporal hybrid neural network-kriging model for groundwater level simulation. *J. Hydrol.* 519, 3193–3203. doi:10.1016/j.jhydrol.2014.10.040.

Tennant, C., Larsen, L., Bellugi, D., Moges, E., Zhang, L., and Ma, H. (2020). The utility of information flow in formulating discharge forecast models: A case study from an arid snow-dominated catchment. *Water Resour. Res.* 56. doi:10.1029/2019WR024908.

Tetzlaff, D., Buttle, J., Carey, S. K., Mcguire, K., Laudon, H., and Soulsby, C. (2015). Tracer-

based assessment of flow paths, storage and runoff generation in northern catchments: A review. *Hydrol. Process.* 29, 3475–3490. doi:10.1002/hyp.10412.

Thomas, A. J., Petridis, M., Walters, S. D., Gheytassi, S. M., and Morgan, R. E. (2017). "Two hidden layers are usually better than one," in *International Conference on Engineering Applications of Neural Networks* (Cham: Springer). doi:10.1007/978-3-319-65172-9_24.

Thomas, A. J., Walters, S. D., Gheytassi, S. M., Morgan, R. E., and Petridis, M. (2016a). On the Optimal node ratio between hidden layers: a probabilistic study. *Int. J. Mach. Learn. Comput.* 6, 241–247. doi:10.18178/ijmlc.2016.6.5.605.

Thomas, A. J., Walters, S. D., Petridis, M., Gheytassi, S. M., and Morgan, R. E. (2016b). "Accelerated optimal topology search for two-hidden- layer feedforward neural networks," in *International Conference on Engineering Applications of Neural Networks* (Cham: Springer), 253–266. doi:10.1007/978-3-319-44188-7_19.

Uhlenbrook, S., Frey, M., Leibundgut, C., and Maloszewski, P. (2002). Hydrograph separations in a mesoscale mountainous basin at event and seasonal timescales. *Water Resour. Res.* 38, 31-1-31–14. doi:10.1029/2001wr000938.

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. doi:10.1109/MCSE.2011.37.

van Rossum, G. (1995). *Python Tutorial*. Amsterdam. Centrum voor Wiskunde en Informatica (CWI).

Vapnik, V., Golowich, S. E., and Smola, A. (1997). "Support vector method for function approximation, regression estimation, and signal processing," in *Adv. Neural Inf. Process. Syst.* (Cambridge, MA), 281–287.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2.

Vogel, M. M., Zscheischler, J., Wartenburger, R., Dee, D., and Seneviratne, S. I. (2019). Concurrent 2018 hot extremes across northern hemisphere due to human-induced climate change. *Earth's Futur.* 7, 692–703. doi:10.1029/2019EF001189.

von Freyberg, J., Studer, B., and Kirchner, J. W. (2017). A lab in the field: high-frequency analysis of water quality and stable isotopes in stream water and precipitation. *Hydrol. Earth Syst. Sci.* 21, 1721–1739. doi:10.5194/hess-21-1721-2017.

von Freyberg, J., Studer, B., Rinderer, M., and Kirchner, J. W. (2018). Studying catchment storm response using event- and pre-event-water volumes as fractions of precipitation rather than discharge. *Hydrol. Earth Syst. Sci.* 22, 5847–5865. doi:https://doi.org/10.5194/hess-22-5847-2018.

Waskom, M., Botvinnik, O., Gelbart, M., Ostblom, J., Hobson, P., Lukauskas, S., et al. (2020). Seaborn: Statistical data visualization. *Astrophys. Source Code Libr. ascl-2012*.

Wasserman, P. D. (1989). *Neural Computing: Theory and Practice*. New York, NY: Van Nostrand Reinhold Co.

Weiler, M., and Naef, F. (2003). An experimental tracer study of the role of macropores in infiltration in grassland soils. *Hydrol. Process.* 17, 477–493. doi:10.1002/hyp.1136.

Wenzel, H. G. J., and Voorhees, M. L. (1981). *Evaluation of the Urban Design Storm Concept*. University of Illinois at Urbana-Champaign: Champaign, IL, USA

Wilby, R. L., Abrahart, R. J., and Dawson, C. W. (2003). Detection of conceptual model rainfall-runoff processes inside an artificial neural network. *Hydrol. Sci. J.* 48, 163–181. doi:10.1623/hysj.48.2.163.44699.

Windhorst, D., Kraft, P., Timbe, E., Frede, H. G., and Breuer, L. (2014). Stable water isotope tracing through hydrological models for disentangling runoff generation processes at the hillslope scale. *Hydrol. Earth Syst. Sci.* 18, 4113–4127. doi:10.5194/hess-18-4113-2014.

Wissmeier, L., and Uhlenbrook, S. (2007). Distributed, high-resolution modelling of 18O signals in a meso-scale catchment. *J. Hydrol.* 332, 497–510. doi:10.1016/j.jhydrol.2006.08.003.

Wu, C. L., and Chau, K. W. (2010). Data-driven models for monthly streamflow time series prediction. *Eng. Appl. Artif. Intell.* 23, 1350–1367. doi:10.1016/j.engappai.2010.04.003.

Wu, C. L., Chau, K. W., and Li, Y. S. (2008). River stage prediction based on a distributed support vector regression. *J. Hydrol.* 358, 96–111. doi:10.1016/j.jhydrol.2008.05.028.

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., and Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* 17(1), 26-40. doi:10.11989/JEST.1674-862X.80904120.

Wu, Y.-C., Yang, J.-C., and Lee, Y.-S. (2007). "An approximate approach for training polynomial kernel SVMs in linear time," in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions* (Prague) 65–68.

Xiang, Z., Yan, J., and Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Environ. Res.* 56. doi:10.1029/2019WR025326.

Yaman, S., Pelecanos, J., and Omar, M. K. (2012). "On the use of nonlinear polynomial kernel SVMs in language recognition," in *13th Annual Conference of the International Speech Communication Association* (Portland, OR).

Yang, T., Asanjan, A. A., Faridzad, M., Hayatbini, N., Gao, X., and Sorooshian, S. (2017). An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis. *Inf. Sci.* 418, 302–316. doi:10.1016/j.ins.2017.08.003.

Yaseen, Z. M., El-shafie, A., Jaafar, O., Afan, H. A., and Sayl, K. N. (2015). Artificial intelligence based models for stream-flow forecasting: 2000-2015. *J. Hydrol.* 530, 829–844. doi:10.1016/j.jhydrol.2015.10.038.

Yaseen, Z. M., Naghshara, S., Salih, S. Q., Kim, S., Malik, A., and Ghorbani, M. A. (2020). Lake water level modeling using newly developed hybrid data intelligence model. *Theor. Appl. Climatol.* 141, 1285–1300. doi:10.1007/s00704-020-03263-8.

Yoon, H., Jun, S., Hyun, Y., Bae, G., and Lee, K. (2011). A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* 396, 128–138. doi:10.1016/j.jhydrol.2010.11.002.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *arXiv Preprint arXiv1411.1792*.

Yuan, X., Chen, C., Lei, X., Yuan, Y., and Adnan, R. M. (2018). Monthly runoff forecasting based on LSTM–ALO model. *Stoch. Environ. Res. Risk Assess.* 32, 2199–2212.

doi:10.1007/s00477-018-1560-y.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., et al. (2018a). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* 565, 720–736. doi:10.1016/j.jhydrol.2018.08.050.

Zhang, G., Eddy Patuwo, B., and Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast*. 14, 35–62. doi:10.1016/S0169-2070(97)00044-7.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., and Yang, J. (2018b). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* 561, 918–929. doi:10.1016/j.jhydrol.2018.04.065.

Zhang, Z., Chen, X., Cheng, Q., and Soulsby, C. (2019). Storage dynamics, hydrological connectivity and flux ages in a karst catchment: Conceptual modelling using stable isotopes. *Hydrol. Earth Syst. Sci.* 23, 51–71. doi:10.5194/hess-23-51-2019.

Zhou, J., Liu, G., Meng, Y., Xia, C., Chen, K., and Chen, Y. (2021). Using stable isotopes as tracer to investigate hydrological condition and estimate water residence time in a plain region, Chengdu, China. *Sci. Rep.* 11. doi:10.1038/s41598-021-82349-3.

Zhu, S., Heddam, S., Wu, S., Dai, J., and Jia, B. (2019). Extreme learning machine-based prediction of daily water temperature for rivers. *Environ. Earth Sci.* 78, 1-17. doi:10.1007/s12665-019-8202-7.

Zhu, S., Nyarko, E. K., and Hadzima-Nyarko, M. (2018). Modelling daily water temperature from air temperature for the Missouri River. *PeerJ* 6:e4894. doi:10.7717/peerj.4894.

Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., et al. (2020). Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *J. Hydrol.* doi:10.1016/j.jhydrol.2020.125085.

Zuo, R., Xiong, Y., Wang, J., and Carranza, E. J. M. (2019). Deep learning and its application in geochemical mapping. *Earth-Science Rev.* 192, 1–14. doi:10.1016/j.earscirev.2019.02.023.

Acknowledgements

First and foremost, I would like to express my special thanks and appreciation to my supervisor Prof. Dr. Lutz Breuer for his continuous support, constant motivation, invaluable advice and immerse knowledge. Thank you for always being there for me to listen to my ideas and answer my questions. I truly appreciate the free rein you gave me when I took off on my own endeavours. Special thanks to Prof. Dr. Jan Siemens for his agreement to be the second supervisor of this dissertation.

I would like to extend my sincere thanks to Dr. Philipp Kraft and Dr. David Windhorst for their insightful comments and substantial support throughout this dissertation. I am also grateful to Dr. Alejandro Chamorro for his profound help to extend my knowledge of machine learning. I would also like to thank Johannes Laufer, Beate Lindenstruth and Heike Weller for their technical support in the lab as well as field work.

Next, I would like to thank all my friends and colleges in the Department of Landscape Ecology and Resources Management for the fruitful discussion and all the fun activities we had together in the last years.

Last but not least, my deepest thanks go to my parents, my sister and my beloved Katha, for their spiritual support, continuous motivation and unconditional love. You were always there for me, believed in me and supported me all the way throughout my journey.

Declaration

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an anti-plagiarism software to check my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ in carrying out the investigations described in the dissertation.”

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

_____

Amirhossein Sahraei

Giessen, 03.09.2021