

Water Resources Research



RESEARCH ARTICLE

10.1029/2018WR024558

Special Section:

Big Data & Machine Learning in Water Sciences: Recent Progress and Their Use in Advancing Science

Key Points:

- Application of entropy and mutual information reveals the information content gap between discharge derived from joint tracer and discharge analyses
- Understanding the information content of hydrological data enhances the interpretation of machine learning prediction results
- Similarities in information could be used for regionalization of catchment characteristics of karst-affected catchments

Supporting Information:

- Supporting Information S1

Correspondence to:

B. Mewes,
benjamin.mewes@rub.de

Citation:

Mewes, B., Opiel, H., Marx, V., & Hartmann, A. (2020). Information-based machine learning for tracer signature prediction in karstic environments. *Water Resources Research*, 56, e2018WR024558. <https://doi.org/10.1029/2018WR024558>

Received 4 DEC 2018

Accepted 9 JAN 2020

Accepted article online 11 JAN 2020

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Information-Based Machine Learning for Tracer Signature Prediction in Karstic Environments

B. Mewes¹, H. Opiel^{1,2}, V. Marx², and A. Hartmann²

¹Institute of Hydrology, Water Resources and Environmental Engineering, Ruhr-University Bochum, Bochum, Germany,

²Chair of Hydrological Modeling and Water Resources, Albert-Ludwigs-University of Freiburg, Freiburg, Germany

Abstract Karstic groundwater systems are often investigated by a combination of environmental or artificial tracers. One of the major downsides of tracer-based methods is the limited availability of tracer measurements, especially in data sparse regions. This study presents an approach to systematically evaluate the information content of the available data, to interpret predictions of tracer concentration from machine learning algorithms, and to compare different machine learning algorithms to obtain an objective assessment of their applicability for predicting environmental tracers. There is a large variety of machine learning approaches, but no clear rules exist on which of them to use for this specific problem. In this study, we formulated a framework to choose the appropriate algorithm for this purpose. We compared four different well-established machine learning algorithms (Support Vector Machines, Extreme Learning Machines, Decision Trees, and Artificial Neural Networks) in seven different karst springs in France for their capability to predict tracer concentrations, in this case SO_4^{2-} and NO_3^- , from discharge. Our study reveals that the machine learning algorithms are able to predict some characteristics of the tracer concentration, but not the whole variance, which is caused by the limited information content in the discharge data. Nevertheless, discharge is often the only information available for a catchment, so the ability to predict at least some characteristics of the tracer concentrations from discharge time series to fill, for example, gaps or increase the database for consecutive analyses is a helpful application of machine learning in data sparse regions or for historic databases.

1. Introduction

Tracer-based methods are often the only way to separate stream flow components and to determine the origin of water (Kirchner, 2003; Klaus & McDonnell, 2013; Mei & Anagnostou, 2015; Mewes & Opiel, 2019; Rimmer & Hartmann, 2014; Weiler et al., 2017). Especially in karstic environments, tracer investigations allow a deeper understanding of the underlying karstic system and the interdependencies of discharge and the current state of the subterranean processes or storages (Aquilina et al., 2005; Gur et al., 2003; Lee & Krothe, 2001).

The joint analysis of tracer data and discharge measurements is a common tool to derive information about hydrological systems, for example, the identification of the origin of water within a catchment. Despite their advantages, these approaches demand long time series of tracer measurements covering a wide range of hydrological system dynamics (Garvelmann et al., 2017; Lee & Krothe, 2001). To describe catchments with hydrological models, the link between tracer signatures and the system's hydrological state is of interest to set up suitable calibration strategies. Although the dependency on tracer data in model studies is high, the information content of tracer measurements has rarely been analyzed. Furthermore, the information-to-noise ratio in the data has to be high to derive the desired information about the system (Kelleher et al., 2015). Another problem is the lack of available tracer databases that hinders many applications, especially in data sparse regions. Here, machine learning could be useful because of the core concept to predict values that are difficult to measure with input data that are straightforward to measure. If the algorithms are able to predict tracer concentrations from discharge time series, data-driven interpolations of continuous tracer concentrations time series can be obtained.

With the rise of machine learning technologies and further improvements in information technology, the application of new approaches for data analysis and the interplay of data, information content, and results have increased (Goodfellow et al., 2016; Kelleher et al., 2015). Machine learning is the umbrella term for

processes that extract patterns from data automatically (Goodfellow et al., 2016). Machine learning-based algorithms are used in many hydrological applications (Raghavendra & Deka, 2014), like rainfall-runoff modeling with artificial neural networks (Hu et al., 2011; Nourani et al., 2009), precipitation forecasting (Yu et al., 2017), evapotranspiration prediction (Tabari et al., 2012), baseflow separation (Corzo & Solomantine, 2007), measurement setup design (Chacon-Hurtado et al., 2017), streamflow forecasting (Shortridge et al., 2016; Shrestha & Solomatine, 2006; Taormina et al., 2015; Yaseen et al., 2016), the separation of flood events from time series of discharge (Mewes & Oppel, 2019), water resource management (Fotovatikhah et al., 2018), and many more. In these studies, machine learning algorithms were mostly used to replicate a system and transform a certain variable into the future. Machine learning was found a useful tool to manipulate data in complex systems, like catchments, where the rules leading from input to output are not completely describable. For example, using a Multi-Layer-Perceptron neural network, dispersion of a tracer was evaluated for a small river in 1-D profile (Piotrowski et al., 2007).

For machine learning algorithms the information content of training data is important (Han & Kamber, 2010; Kelleher et al., 2015; Vapnik, 2013). The Shannon entropy is a common concept in information theory to analyze the information content of given data (Shannon, 1948; see also Fernando et al., 2009). Until now, no study tried to predict natural tracer concentrations in karstic environments from discharge dynamics by the application of machine learning algorithms to fill gaps between point measurements of tracer concentrations. This strategy was chosen, because discharge is often the only available data source with an appropriate temporal resolution for hydrological modeling at an event scale. In the database we used, some infrequent tracer concentration measurements were available as point measurements. A machine learning tool capable of filling these gaps would allow the application of databases of frequent discharge measurements and non-frequent measured tracer concentrations. Additionally, an already trained algorithm could predict tracer concentrations for catchments in which only a limited number of discharge measurements is available used. Furthermore, it would qualify historic data for application in modeling approaches that require a higher temporal resolution of tracer measurements. In karstic environments, the joint analysis of tracer data is often the key for a deeper understanding of system states and behavior (Mudarra et al., 2019). Therefore, we assume a high information content in the measured tracer data because they describe the complex interaction of subterranean processes. Machine learning algorithms depend on information provided in the data. Consequently, the available data sets of discharge and tracer measurements have to be analyzed on explanatory power, what has not been done before for a database of karst springs. Furthermore, an information content-based analysis of the interpolated tracer measurements can be conducted by comparing the prediction results with the information content of the input data.

In this study, we analyzed observed discharge and natural tracer data (sulfate, SO_4^{2-} , and nitrate, NO_3^-) from seven different karst springs across Europe regarding their information content. We took natural tracers because they exist in varying concentrations and are measurable without any induced injection. We chose nitrate and sulfate because they represent different residence times in the system. While nitrate represents shallow fast flowing water, sulfate represents the opposite origin: slow phreatic processes. We applied different machine learning algorithms such as Support Vector Machines (SVM), Classification and Regression Trees (CART), Extreme Learning Machines (ELM), and Artificial Neural Networks (ANN), to estimate tracer concentrations from discharge dynamics. We selected those four machine learning approaches that (a) are well established in hydrology, (b) are used for pattern recognition in structured data sets, and (c) deliver to a certain degree interpretable structures for the researcher. Furthermore, we compared different concepts of prediction, including the univariate prediction that separately estimates each tracer with a specialized machine and the multivariate estimation that tries to predict a set of tracers with a combined machine. We tested each of the chosen approaches on the prediction capability in seven different catchments and created a strategy to build a data-driven interpolation tool set for the interpolation of continuous time series of tracer measurements. Finally, we linked the prediction results with the observed information content in the data as well as with the mutual information between the chosen tracers.

2. Methods and Data

Sound results from machine learning approaches require data with a high information-to-noise ratio. Moreover, the choice of the appropriate machine learning algorithm for this task is difficult to justify

without understanding the internal structure of the problem. Following the No-Free-Lunch-Theorem, all available approaches should be equally suitable to solve this problem but with a different performance and different demands to the data in terms of amount and quality (Wolpert & Macready, 1997). Accordingly, without information for an a priori selection of the best machine learning approach to use, we chose four structurally different approaches to estimate tracer concentrations in seven catchments. To quantify the information content within the data set, we introduce concepts like continuous entropy and mutual information. After defining these basic concepts, we explain the choice of machine learning algorithms in this study and explain the further scheme of this application.

2.1. Entropy and Mutual Information

Shannon's model of entropy allows to quantify the amount of information gain by adding new data to the analysis (Shannon, 1948). The entropy H is defined by the chance of a sample X_d to be of one of the given classes $\{x_1, \dots, x_{N_d}\}$ with $P(x_n)$ as the probability that $X_d = x_n$ with a sample length N :

$$H(X_d) = - \sum_{n=1}^N P(x_n) \log_2 P(x_n) \quad (1)$$

Because Shannon's entropy is only valid for discrete data, the concept was extended to the continuous entropy for a continuous variable X_c , which is in our case discharge:

$$h(X_c) = - \int_{\Omega} f(x) \log_2 f(x) dx \quad (2)$$

where $f(x)$ is the probability density function (PDF) of X_c and Ω is the defined domain of X_c (Gong et al., 2014). To determine the explanatory power of data concerning a variable, for example, how much of the information of NO_3^- is explained by discharge only, we further extend the concept of continuous entropy to conditional entropy (Thomas & Cover, 2006), where y is the tracer concentration and x is the discharge sequence:

$$H(Y|X) = \sum_{x \in X, y \in Y} P(x, y) \log_2 \frac{P(x)}{P(x, y)} \quad (3)$$

Conditional entropy describes how much of variable y can be explained by variable x . To describe the shared information between two data points given as x and y , we apply the mutual information (Shannon, 1948; Sharma, 2000). In our case, we investigate the shared information between the two chosen tracers NO_3^- and SO_4^{2-} . The mutual information between two measurements is defined as

$$MI = \iint f_{x,y}(x, y) \log_2 \left[\frac{f_{x,y}(x, y)}{f_x(x) f_y(y)} \right] dx dy \quad (4)$$

where $f_x(x)$ and $f_y(y)$ are marginal PDFs of x and y , $f_{x,y}(x, y)$ is the joint PDF of x and y (Sharma, 2000). After Sharma, 2000, the mutual information score from equation (4) can be approximated by

$$MI = \frac{1}{N} \sum_{i=1}^N \log_2 \left[\frac{f_{x,y}(x_i, y_i)}{f_x(x_i) f_y(y_i)} \right] \quad (5)$$

In this approximation $f_x(x_i)$, $f_y(y_i)$, and $f_{x,y}(x_i, y_i)$ are marginal functions and joint densities at the same point of the same sample (Fernando et al., 2009; Sharma, 2000). To estimate the density, we apply a kernel estimator (Fernando et al., 2009). Without the kernel estimator a theoretical distribution of the MI has to be assumed, which adds more bias to the approach. As nearly all models rely on the interplay of input and output data the shared information through mutual information has to be weighted stronger than the internal information represented through the continuous entropy.

2.2. Machine Learning Algorithms

The main aim of the paper is to use discharge data as a predictor for tracer concentrations because discharge in streams and rivers is more commonly measured than tracer concentrations, especially in regions where access to the site is limited and research relies on public databases. Therefore, we train machine learning algorithms using time series of runoff to predict time series of tracer concentrations.

The discharge dynamics are captured by a window of discharge data from the original time series with tracer measurement t' as input for the machine learning algorithms. The machine learning algorithms predict the tracer concentrations based on information from the discharge pattern (Figure 1). For training and validation, the predicted tracer concentrations are compared to the measured data (which is considered to represent the reality). To reduce overfitting due to complex input data, an optimal length for the window of discharge data has to be identified, which is discussed in detail in section 2.3. Without defining a window, a Long-Short-Memory network can be applied, which requires a continuous time series of input and training data. Due to a lack of continuous time series of tracer measurements, this approach was discarded.

Four structurally different machine learning algorithms are used in this study: SVM, CART, ELM, and Multi-Layer-Perceptron ANN. These algorithms were chosen because of their suitability for regression problems and their origin in two of the four main machine learning families: error-based learning and information-based learning (Kelleher et al., 2015). Moreover, they are commonly applied in hydrology and deliver, to a certain degree, structures that can be interpreted by the researcher. SVM and CART are not known to capture temporal patterns in time series data. By the reduction from a complete time series to a window with a variable length, temporal dependencies are reduced to dependencies of the relative position within the window. Thus, the problem is diminished to a pattern recognition problem (Nasrabadi, 2007).

A SVM is an error-based machine learning algorithm that tries to set up a regression to estimate the unknown tracer concentration from the input discharge sequence (solid line in Figure 2(a)). This regression is depicted through a hyperplane, for which the distance to the margin (dashed line in Figure 2(a)) and the most distant feature, the so-called support vector, is maximized (Cortes & Vapnik, 1995; Raghavendra & Deka, 2014). For a linear problem, this fitting of a regression can easily be done, but most of the machine learning problems, as the one presented here, are highly nonlinear. Therefore, we have to transfer the existing problem to a higher dimension where the problem becomes linear with a kernel function (Chang et al., 2010; Kelleher et al., 2015). As the choice of the mapping kernel is highly problem specific, a selection of several kernel functions (radial basis function, linear, polynomial, and sigmoid) was tested and the best kernel was chosen (in terms of numerical stability and computational demands), in our case the radial basis function kernel. For more information on the choice of the kernel, see Vapnik (2013). The created boundary layer is used to predict the unknown tracer concentration C in the feature space through the input discharge dynamic, represented as a green dot (Figure 2). Accordingly, the SVM tries to solve the regression problem by transferring the discharge data into either a single tracer concentration or a set of tracer concentrations in the multivariate output. Hence, the hyperplane represents the regression function to estimate the respective tracer concentration from the discharge sequence.

CART builds decision trees that are guidebooks to estimate the tracer concentration from the discharge values. The tree shows the ramifications of decisions leading to the final regression result (Breiman et al., 1984; Kelleher et al., 2015; Quinlan, 1986). To build the tree, all discharge values are analyzed in their ability to maximize the decrease of the residuum of the regression between observed and estimated tracer concentration at each branch. The branching occurs on the descending order of error reduction. As a result, the structure of the decision tree can be obtained as guidebook for unknown values, in order to get the desired tracer concentration C (Figure 2(b)). In the given example, the discharge value at position 0 has the highest influence on error reduction and results in the decision between the major branches, which are themselves as diversified as certain discharge values, resulting in the final leaves with the target value C represented as a green dot. The error reduction within the tree for each node is calculated with the root-mean-square error (RMSE) of the regression (see error metrics section). The regression tree analyzes the discharge values to find the values that have the highest influence on the regression problem to determine the predicted tracer concentration. The depth of the CART tree was limited to the number of input values from the time series of runoff in order to capture all details of the variability of discharge.

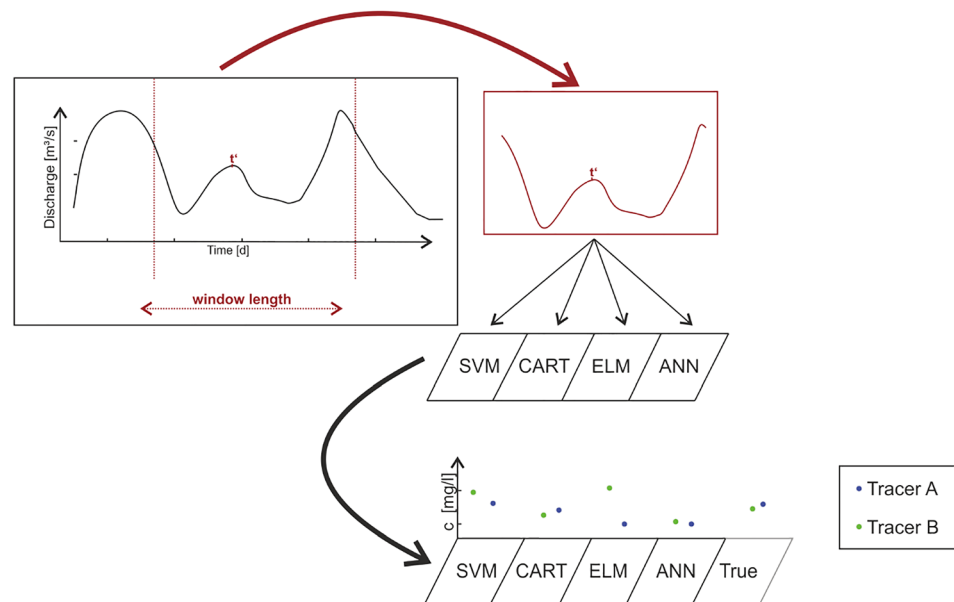


Figure 1. Workflow of the analysis including the clipping of the window for the *discharge* data, the prediction of tracer signatures by the machine learning algorithms, and the following comparison with measured tracer measurements.

ANN and a ELM (Figures 2(c) and 2(d)) are both variations of neural networks that try to solve the regression or classification problem by imitating the structure of the human brain and by guiding the training data through a network of hidden layers equipped with neurons (Haykin, 1999). Here, the input nodes are the discharge values from the window of discharge values for estimation of the desired tracer concentration. The hidden layers and nodes represent the underlying system, in this case the karst subsurface system. The connection between nodes and layers is trained by the optimization of weights in order to minimize the regression error. An ELM is a special case of an ANN: The nodes on the hidden layer receive their weights only once. In the following, they remain constant over the process of network adaption. Only the weights from the hidden layer to the output node are updated, which is called a feedforward network due to the update direction of weight (Huang et al., 2004). Here, the discharge values are sent through the network of nodes and hidden layers to identify the pattern and estimate the tracer amount. The network can either be trained to estimate a single tracer or a set of tracers. Generally, the number of hidden layers is restricted to a single hidden layer with half of the input window length as hidden nodes (and a minimum of three hidden nodes for stability reasons).

To avoid overfitting of the data, the number of input data was reduced to a maximum of half of the available runoff data in the window with a minimum of three remaining runoff values as input data. Furthermore, the random selection of input values was shuffled 10 times and the mean prediction was taken to be representative for the specified window length.

Machine learning algorithms depend on the information content of the data (Goodfellow et al., 2016; Kelleher et al., 2015). Consequently, we assume a link between the performance of the algorithm and the information content of the data (defined in section 2.1). We train the algorithms by two different ways: (1) by a univariate strategy estimating each tracer individually and (2) by a multivariate strategy that trains one algorithm to estimate both tracers simultaneously. We expect that the multivariate strategy performs better than the univariate as the combination (i.e., interaction) of data should lead to more incorporated information than just the information content of a single data set. A globally trained algorithm to predict a set of natural tracers would lower the interpretability of the results. Thus, we discarded the idea of a universal machine for tracer concentration prediction but focused on the two mentioned natural tracers.

2.3. Training

The discharge data have to be reduced to a window with an unknown length. This optimal length might be highly subjective whether all information on the system's behavior is covered in the respective time span. The window to be selected contains the tracer measurements and the number of discharge values depicting the

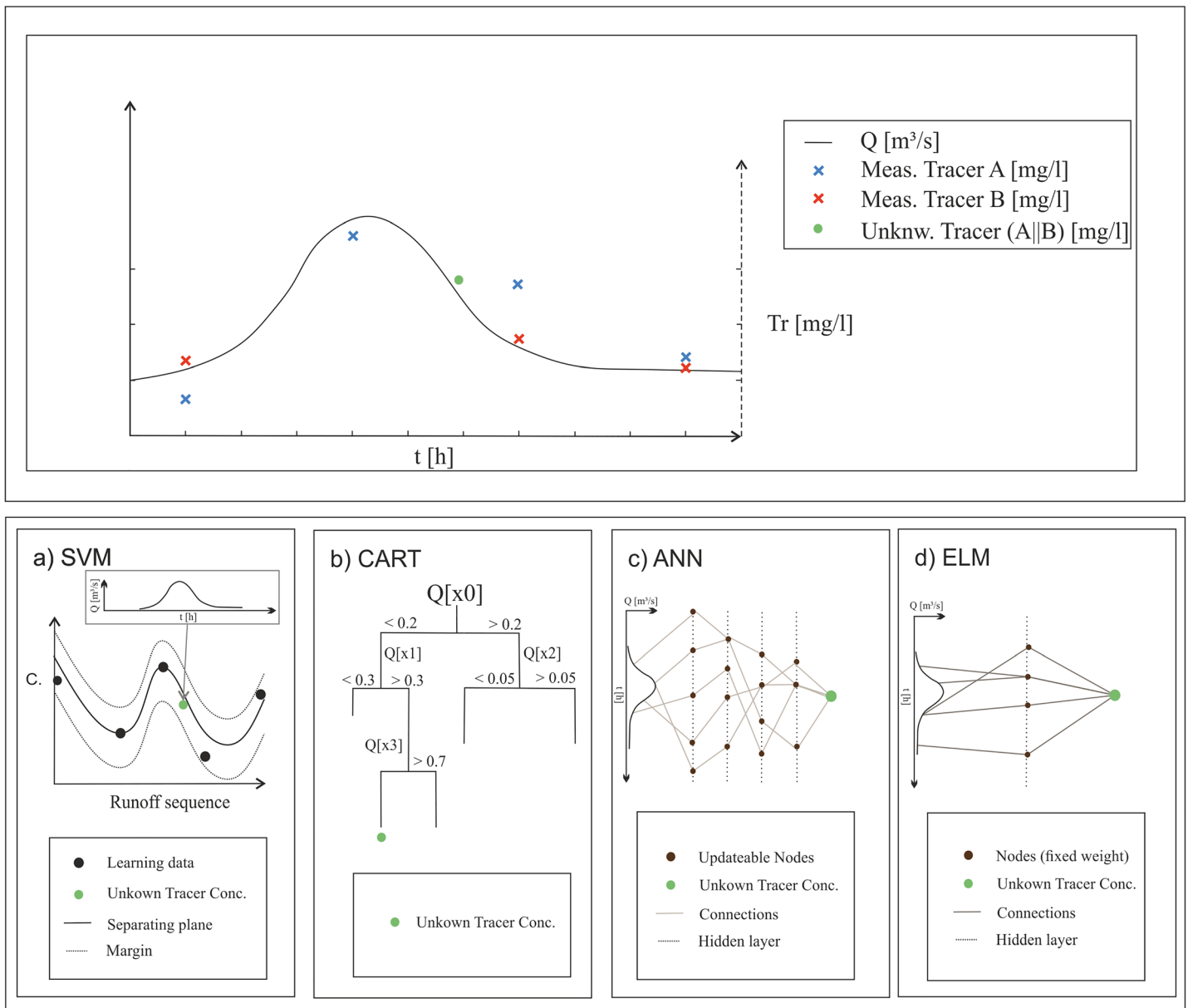


Figure 2. The major task for the machine learning algorithms in this study is presented in the upper part of the figure: To estimate the unknown tracer concentration, C , by training a machine learning algorithm to the pattern formed by a subset of discharge and a measured pair of tracers. The structure of the chosen algorithm for this study are shown in subplots a, b, c and d.

discharge dynamics. As we do not know whether the window length depends on the chosen approach or region we varied it from 1 to 180 days in steps of [1, 3, 6, 30, 60, 90, 180] with equally sized borders to face the unknown optimal length. The window lengths chosen here represent natural breaks within the classification of time to describe a system. We chose these different lengths of the window to include short-, medium-, and long-term processes in the discharge data and to minimize the number of data sets analyzed. Therefore, we focused on time spans like a month, two months, and half a year. The discharge in the sequence is normalized by the catchment specific average discharge to reduce the influence of the peak. The measured tracer concentrations are also normalized by the specific mean of this tracer for the catchment. The share of the training data is increased gradually to understand how simulation performance is influenced by the size of the training data. Therefore, we varied the amount of data used

for training from 10–90% of the available time series for the catchment. Using the length of the covered time series instead would be insufficient because the input data includes runoff sequences that might overlap. Hence, the number of tracer measurements is important.

We train the algorithms with both a univariate and a multivariate strategy. We compare the results from the two learning strategies to quantify the potential improvement of shared information and joint learning. Furthermore, we discuss the influence of the window length on prediction quality. This is relevant as the length of the input sequences can create a bias in the learning process. If we choose the length too short, we might not cover all relevant processes, whereas sequences that are too long might confuse the algorithms in finding a suitable system. In the last step, we elaborate on the transferability of the algorithms to be used as predictors at catchments for which they were not trained. That way, we can test whether machine learning tools and their results might reveal hidden similarities in catchment responses or even more interesting the application of machine learning is suitable for the prediction of missing tracer measurement data.

2.4. Evaluation and Error Measures

To compare the different machine learning approaches, training strategies and window lengths, quantitative performance measures were used.

In order to show the general prediction performance, the RMSE was applied for observed and estimated tracer measurements, which becomes 0 for a perfect prediction. To calculate RMSE for the tracer content, we differentiated between measured and predicted c_T , with N being the number of samples in the validation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (c_{T_{meas}} - c_{T_{pred}})^2}{N}} \quad (6)$$

We apply RMSE for both tracers individually and calculate the mean of both as an indication of the combined error. Because of the variable window length, individual RMSEs are calculated for each approach and each region. As the normalization in RMSE does not show the direction of error in contrast to the mean error which is less robust against outliers, we also analyze the average concentration ratio \bar{c}_T that provides information about the general strength and direction of the error of prediction:

$$\bar{c}_T = \frac{1}{N} \sum_{i=1}^N \frac{c_{T_{pred}}}{c_{T_{meas}}} \quad (7)$$

\bar{c}_T is able to show the direction and the strength of the error by its sign and its difference from one, respectively. Again, because of the multitude of different window lengths, a range of \bar{c}_T values is calculated for each region and approach.

As all the presented measures are merely a measure of quantitative performance, the qualitative performance is measured with the accuracy of the internal ranking of the two tracer signatures. Therefore, we calculated the accuracy by an error matrix of true and false combinations of ranking. The deducted measure of accuracy acc is able to describe the qualitative information between the two tracers as an accuracy with a ranking (Han & Kamber, 2010):

$$acc = \frac{pos_{True}}{pos} \frac{pos}{(pos + neg)} + \frac{neg_{True}}{neg} \frac{neg}{(pos + neg)} = \frac{pos_{True}}{(pos + neg)} + \frac{neg_{True}}{(pos + neg)} \quad (8)$$

With pos_{True} and neg_{True} as the ranking of the pair of tracers in concentration, for example, $c_{T_{Aobs}} > c_{T_{Bobs}}$ but $c_{T_{Aest}} < c_{T_{Best}}$ results in a *neg* prediction, whereas $c_{T_{Aobs}} > c_{T_{Bobs}}$ and $c_{T_{Aest}} > c_{T_{Best}}$ counts as *pos_{True}*. Accuracy shows the ability of the machine learning method to replicate the ranking of the tracer concentrations in order to replicate changing tracer dynamics.

The three measures considered here to judge the performance represent the major key characteristics of the prediction results. The overall goodness represented as the RMSE, deviation from the mean and the ranking between both tracers. So, by a correct ranking the qualitative information that tracer concentration dominates is still captured, even though the variance of the prediction is not high enough.

2.5. Data

The target variables of the machine learning prediction are the concentrations of SO_4^{2-} and NO_3^- that act as a combined tracer signature. While NO_3^- is known as an indicator for fast water fluxes from the soil or epikarst, that is, the shallow subsurface (Hartmann et al., 2016; Mahler & Garner, 2009), SO_4^{2-} in karst systems is usually derived from geogenic processes that dissolve evaporates in the phreatic subsurface that sustains base flow (Hartmann et al., 2017; Mudarra & Andreo, 2011). We chose these two tracers as an example for any tracer combination. Due to their different origins, the shallow subsurface (NO_3^-) and the phreatic zone (SO_4^{2-}), we expect that their observations of dissolved evaporates include different information.

The data for our analyses originate from seven different karst springs in France (Table 1 and Figure 3). Tracer measurements were normalized by individual mean values, leading to seven different means (Eaufrance, 2018a). The tracers analyzed in this study are natural tracers; no human-induced injections were made. The tracer concentrations were measured repeatedly, but not at fixed intervals. There was a strong linear correlation between both tracers SO_4^{2-} and NO_3^- with $r = 0.67$. Measured discharge values were obtained from Banque Hydrologique and have a daily resolution (Eaufrance, 2018b). Banque Hydrologique publically provides daily discharge data of continuously measured rivers and springs collected by French state agencies.

The two springs Baget and Fontestorbes are located in the Pyrénées Mountains (Ariège department) at a median altitude of 1,000 m. The recharge areas are 13 and 80 km² for the Baget and Fontestorbes spring, respectively. Mean daily discharge of the Fontestorbes spring, which is one of the largest intermittent karst springs in the world, is 2.1 and 0.5 m³/s at the Baget spring. Due to the similarity of the two midaltitude basins (Labat et al., 2002), mean annual precipitation of 1,178 mm (Bailly-Comte et al., 2018) can be assumed for both locations. The Durzon spring is located on the Larzac Plateau in the Grands Causses area in the Massif Central (Aveyron department). It is a perennial, vauclusian-type spring with a mean daily discharge of 1.5 m³/s. The recharge area has been determined to be >100 km² (Jacob et al., 2008). The Fontaine de Vaucluse spring is a well described and famous karst spring being the largest karstic outlet in France (Vaucluse department). The mean daily discharge is over 20 m³/s and the low flow discharge is always higher than 4 m³/s. The recharge area is about 1,115 km² (Fleury et al., 2009). The Fontbelle spring is part of the Ouyse karst system (Lot department) (Kavouri et al., 2011). The Source de la Touvre is the second largest karst spring in France and the sole outlet of Rochefoucault karst system (Charente department). The spring, fed by the losses of three large rivers, has a mean daily discharge of 13 m³/s and a recharge area of about 126 km². The water resources are used for the water supply of Angouleme city. The Source du Lez is the main perennial outlet of the Lez karst system (Montpellier department) with a mean daily discharge of 2 m³/s. Pumping for the water supply of Montpellier city puts the aquifer under high anthropogenic pressure (Bicalho et al., 2017).

More details about the springs are provided in Table 1 and Figure S1 (see supporting information) or at data base webpage (hydro.eaufrance.fr).

3. Results

3.1. Entropy and Mutual Information of Available Data Sets

Following the principle of continuous entropy, the information content of discharge and the mutual information of the joint data sets (tracer signatures) was calculated. We resampled the complete set of sequences ten times and looked at the mean entropy of each individual data set and the mutual information of two different tracer signatures, SO_4^{2-} and NO_3^- . Missing or erroneous results are labeled NA, which leads to gaps shown in the information contents of springs like Fontaine de Vaucluse (see supporting information).

The Baget example shows that the entropy of discharge decreases when more data are used for training (Figure 4). The mutual information between the two tracers exceeds the continuous entropy of discharge by far. The information content shared between those two tracers is 35 times higher than the continuous entropy of the discharge. That means that we need a lot of information to fully describe the variability of the interplay between those two tracers and we might not successfully describe this variability with the discharge data alone. Using more than 60% of the available tracer data sets, the mutual information reaches a plateau where no further information is needed to describe the dynamics. The behavior of MI is similar for all other catchments: The information content is by far higher than the continuous entropy of discharge and

Table 1
Overview of Used Data

| Source | Lat | Lon | Department | Mean daily discharge (m ³ /s) | Recharge area (km ²) | Köppen Geiger climate | Mean annual rainfall (mm/a) | Geology | Length of daily discharge measurements | Tracer measurements SO ₄ ²⁻ and NO ₃ ⁻ |
|---------------------------------------|---------|--------|-------------|--|----------------------------------|-----------------------|-----------------------------|--|--|--|
| Baget ^{a-d} | 42.9554 | 1.0304 | Ariège | 0.5 | 13 | Dfb/Dfc | 1,187 | Lower Cretaceous limestone | 1968–2015 | 24 |
| Source de Fontestorbes ^{b-d} | 42.8925 | 1.9271 | Ariège | 2.1 | 80 | Cfb | 1,187 | Cretaceous limestones and marls | 1965–2015 | 43 |
| Durzon ^e | 43.9909 | 3.2617 | Aveyron | 1.5 | 124 | Dfc/Cfb | 400 ^(k) | Middle to upper Jurassic limestones and dolomites | 1996–2016 | 154 |
| Fontaine de Vaucluse ^{a,f} | 43.9177 | 5.1327 | Vaucluse | 20 | 1,115 | Csb/Csa | 960 | Great, lower Cretaceous limestone series | 1966–2016 | 51 |
| Fontbelle ^g | 44.7956 | 1.5640 | Lot | 0.1 | | Cfb | 730 ^h | Middle-to-Late Jurassic tabular carbonate sequence | 2004–2015 | 194 |
| Source de la Touvre ⁱ | 45.6630 | 0.2546 | Charente | 13 | 126 | Cfb | 945 | Upper Jurassic limestones | 1980–2016 | 125 |
| Source du Lez ^{a,j} | 43.7182 | 3.8842 | Montpellier | 2 | | Csb | 942 | Upper Jurassic and early Cretaceous limestones | 1987–2016 | 300 |

^aJourde et al. (2018) ^bLabat et al. (2002) ^cBailly-Comte et al. (2018) ^dBDLISA (2019) ^eJacob et al. (2008) ^fFleury et al. (2009) ^gKavouri et al. (2011) ^hObtained from Hartmann et al. (2015) ⁱLe Moine et al. (2008) ^jBicalho et al. (2017)

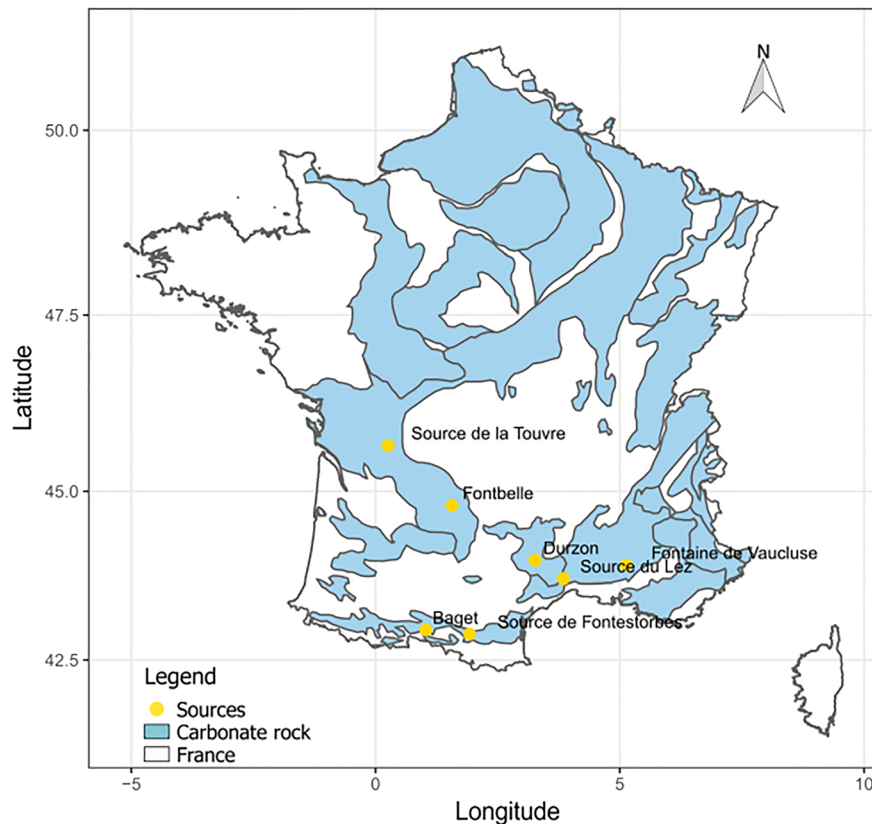


Figure 3. Location of analyzed carbonate rock dominated sources in France.

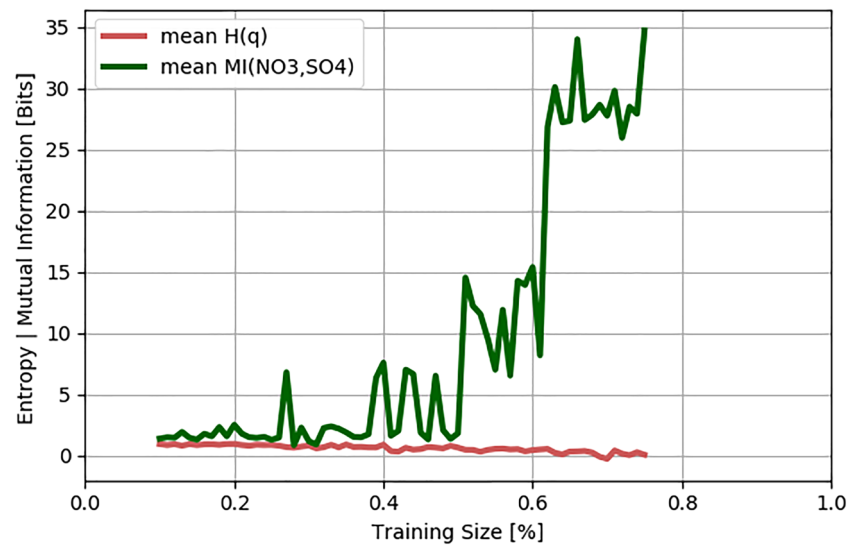


Figure 4. Mean continuous entropy and mutual information between NO_3^- and SO_4^{2-} at Baget spring, showing the shared information between both tracers and discharge and the singular information through the continuous entropy of the isolated data sets.

a plateau is reached using at least 60% of data. Therefore, we assume that we need at least 60% of the available tracer measurements to cover the variability of the system's dynamics in the training. For more details, we refer to the supplement (Figure S2) where the entropy and the mutual information for all catchments is shown in detail.

3.2. Validation of Prediction Accuracy

For the validation of the prediction accuracy, we compared two different learning strategies: the univariate strategy, focusing on only one tracer at a time, and the multivariate strategy, considering both tracers at the same learning phase. The results shown here represent all considered sizes of the discharge window. The prediction results are presented as a boxplot to show the variability and the influence of the different window lengths without going into detail on the specific influence of the window (Figure 5). The average tracer concentration ratio \bar{c}_T indicates that the tracer signatures can be predicted better at some springs than at others. Furthermore, they show a preference toward certain prediction techniques with a \bar{c}_T value close to the optimum value. For the Fontaine de Vaucluse, Fontbelle, Sources de Fontestorbes and Source du Lez, \bar{c}_T converged to the optimal value 1.0. The differences between the machines were marginal, although ELM and ANN results were less variable and thus less influenced the amount of training data. For the Baget catchment, we could not predict the concentrations with any machine as the variability is high for all applied approaches and amounts of training data. For the catchments Durzon and Source de la Touvre either NO_3^- or SO_4^{2-} was overestimated or underestimated, although CART delivered acceptable results for the Source de la Touvre.

The RMSE of the prediction from all investigated window lengths is presented as a boxplot in Figure 6. The RMSE of the tracer concentrations shows similar results like \bar{c}_T . While for some catchments RMSEs were low regardless of the chosen machine, for catchments like Baget the results are worse than for catchments like Fontbelle and Source de la Touvre. If the \bar{c}_T of the catchment does not converge to 1.0 (like the SVM in Source du Lez), the RMSE is higher than in regions like Fontaine de Vaucluse and Fontbelle where \bar{c}_T is also close to the optimum. The choice of the machine has only small influence on the RMSE, apart from Source du Lez where the SVM delivers worse results than any other method. Generally, a RMSE lower than 1.0 is an acceptable value for the prediction of the normalized concentration. This limit is reached for all machines in the catchments Fontaine de Vaucluse, Fontbelle, Source de Fontestorbes, and Source de la Touvre while at Baget, Durzon, and Source du Lez the RMSE remains highly variable. Whether a univariate or a multivariate approach results in a lower mean RMSE cannot be stated with certainty from these results, but in most cases the mean RMSE of the multivariate approach was lower than the mean RMSE of the respective univariate approach.

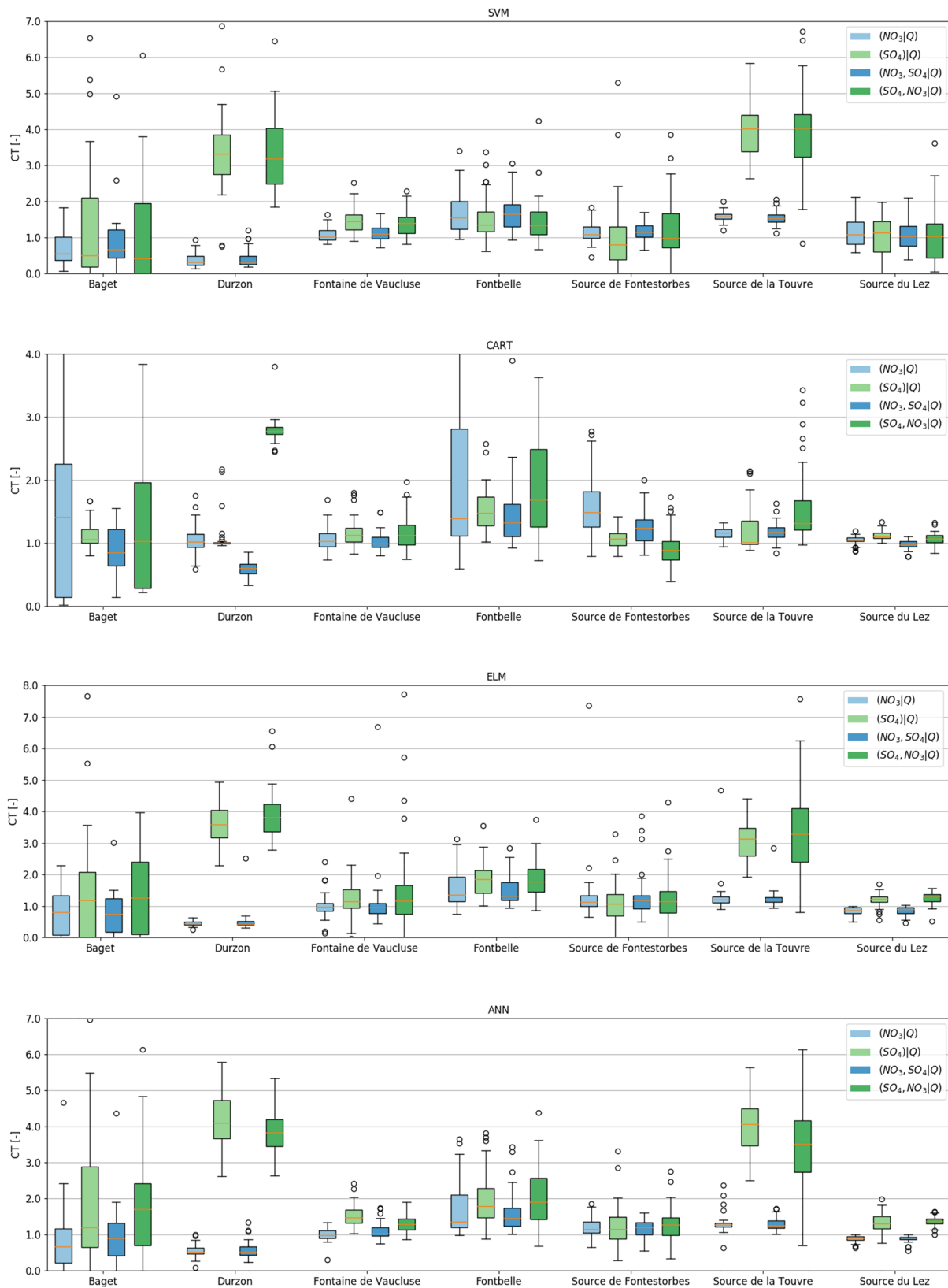


Figure 5. \overline{CT} of SVM, CART, ELM, and ANN. The variability within each boxplot expresses the performance according to the applied type of training data. While the results are good for most catchments, some concentrations like SO_4 in certain catchments, like Source de la Touvre are overestimated while using an ELM or an ANN algorithm.

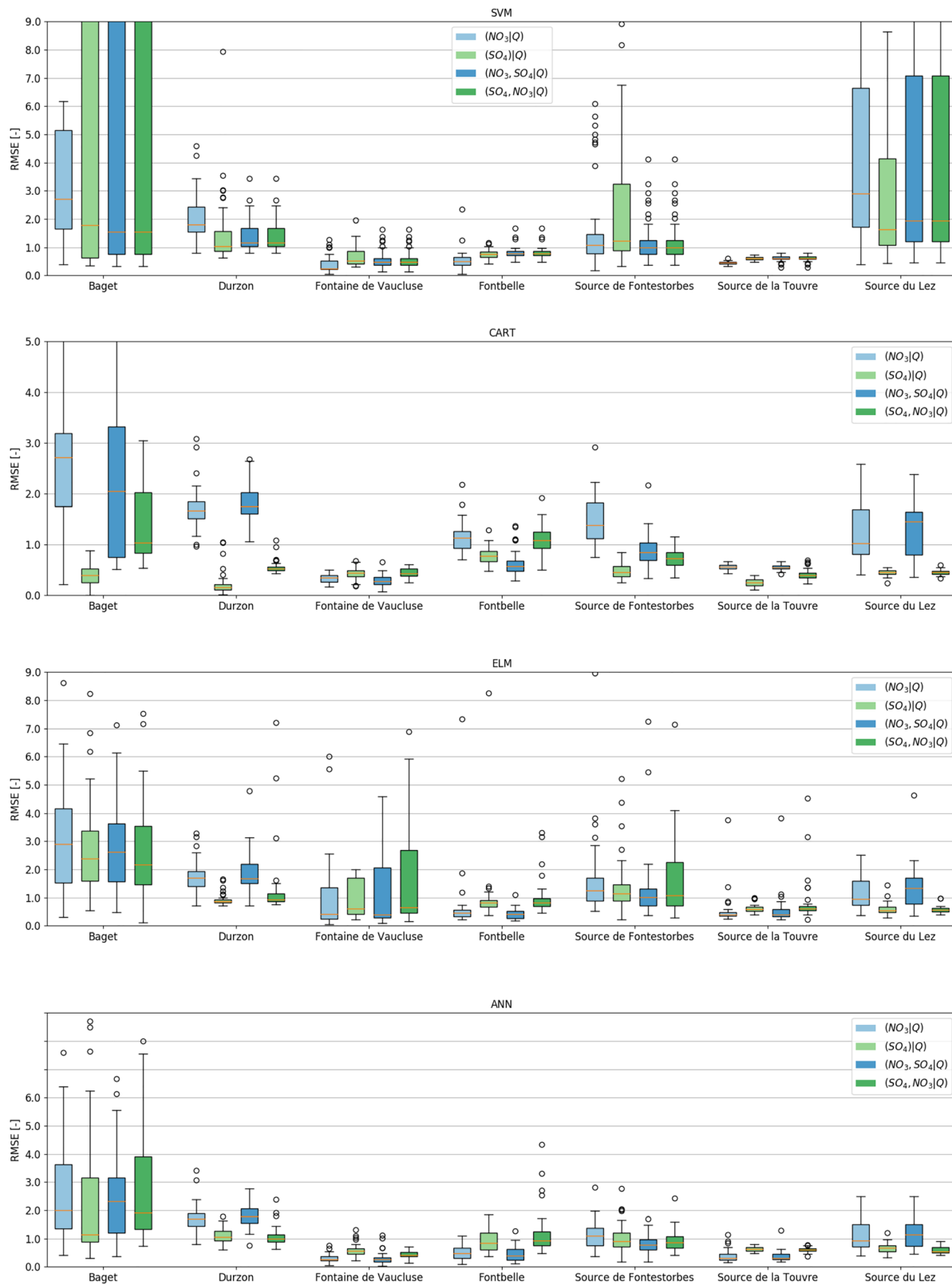


Figure 6. RMSE of the normalized tracer concentrations of SVM, CART, ELM, and ANN for univariate and multivariate algorithms. The variability shows the influence of the learning threshold on the development of RMSE in the catchment. The RMSE results are similar to the results from \bar{c}_T and show that the error relates to the average tracer concentration and that for some catchments problems in the prediction occur, like catchment Baget. The choice of the machine has only a small influence on the error and depends on the region.

The *Acc* value describing the correct ranking of tracer concentrations shows for all catchments that at least 40% of the rankings are estimated correctly (Figure 7). None of the machines reached mean *Acc* values >70%. Here, the choice of machines has an influence on the dynamics of the tracer concentrations. The *Acc* values were highest for catchment Baget compared to all other catchments, while showing the highest variability of \bar{c}_T . The multivariate prediction does not automatically improve the results in terms of *Acc* at all catchments, and the improvement or deterioration varies among the applied approaches (e.g., SVM and ELM in Durzon). The reason behind this might be found in the interplay of information content, regional aspects of the catchment, and the quality of the input data. Therefore, it is out of scope of this paper to check the causality of the preferred choice. Nevertheless, in most catchments, the multivariate machines improve *Acc*. Again, the choice of the machine has less impact on results and it is merely a catchment specific question.

The influence of the chosen window length on the prediction capability of NO_3^- and SO_4^{2-} is exemplified by the \bar{c}_T values of all four (univariate) machines in catchment Source de Fontestorbes (Figure 8). Generally, either very short windows (1–4 days) or long windows (>60 days) lead to good results, while window lengths in between worsen the results for SVM, CART, and ELM. For further information on the window dependency of the other catchments, which are very similar to the information we derived from our example, we refer to the supporting information.

As a good example for choosing an approach with the required number of training data for a catchment, we elaborate the case of Fontbelle (Figure 9). Here, ANN and SVM obtain \bar{c}_T values close to the optimum of 1.0, but the ANN results in lower RMSE values than the SVM. Therefore, we chose the ANN to predict tracer concentrations in this catchment. The resulting time series (predicted by an ANN trained with 70% of the available measurements) reveals that the measured tracer concentrations and the predicted time series show an acceptable agreement with the mean value of concentration captured as well as the general ability to predict concentrations at all levels measured.

Taking a closer look at the prediction capability for SO_4^{2-} , we can see that the multivariate approach interpolates concentration in the same range, even close to a concentration of 0.0 mg/L (red marked area in Figure 9). The multivariate approach is able to cover the peaks, while the univariate approach predicts values close to the mean concentration. Interestingly, the mean tracer concentration rises over time using the univariate approach. However, the behavior NO_3^- is different: The univariate prediction shows a variability that reflects the measured tracer concentrations better, while the multivariate machine predictions show too low variability around the observed mean concentration. As shown by the red marked area of Figure 8, the univariate approach allows interpolating NO_3^- concentrations from Day 2,000 to Day 3,200. The following decreasing trend cannot be interpolated, and thus, the approach lacks a significant performance here from Day 3,200 until the end.

4. Discussion

Missing tracer measurements in terms of gaps or irregular measurement campaigns are the major downside in using these data to develop models for system characterization. In many cases, it is not possible to repeat the measurements for the desired tracers, for instance, when data are obtained from online databases like the U.S. Geological Survey. Furthermore, only limited knowledge is available on the information content of the data used in tracer-aided modeling (Hartmann et al., 2017; Kelleher et al., 2019). Our results indicate that machine learning algorithms represent a valuable technique to predict some characteristics of tracer concentrations in the karstic environments. Even though none of the machine learning methods were able to describe the complete dynamics between the two tracers with high precision, our comparative approach of using different machine learning methods allows us to choose the most appropriate method describing a specific characteristic at a specific site. Hence, we are able to predict key characteristics like the mean concentration and the relative ranking of tracers in a joint tracer analysis. The reason that tracer concentration dynamics could not entirely be predicted by discharge alone is the low information content thereof compared to the shared information of the tracers. The use of ancillary data or more sophisticated approaches to improve our prediction is hampered by data limitations or unsecure quality (in terms of measurement quality). Consequently, the prediction capability of the algorithms is lowered by the limitation to discharge data and the low temporal resolution of concentration measurements. Thus, results have to be interpreted carefully and with special regard to the information content of the underlying data.

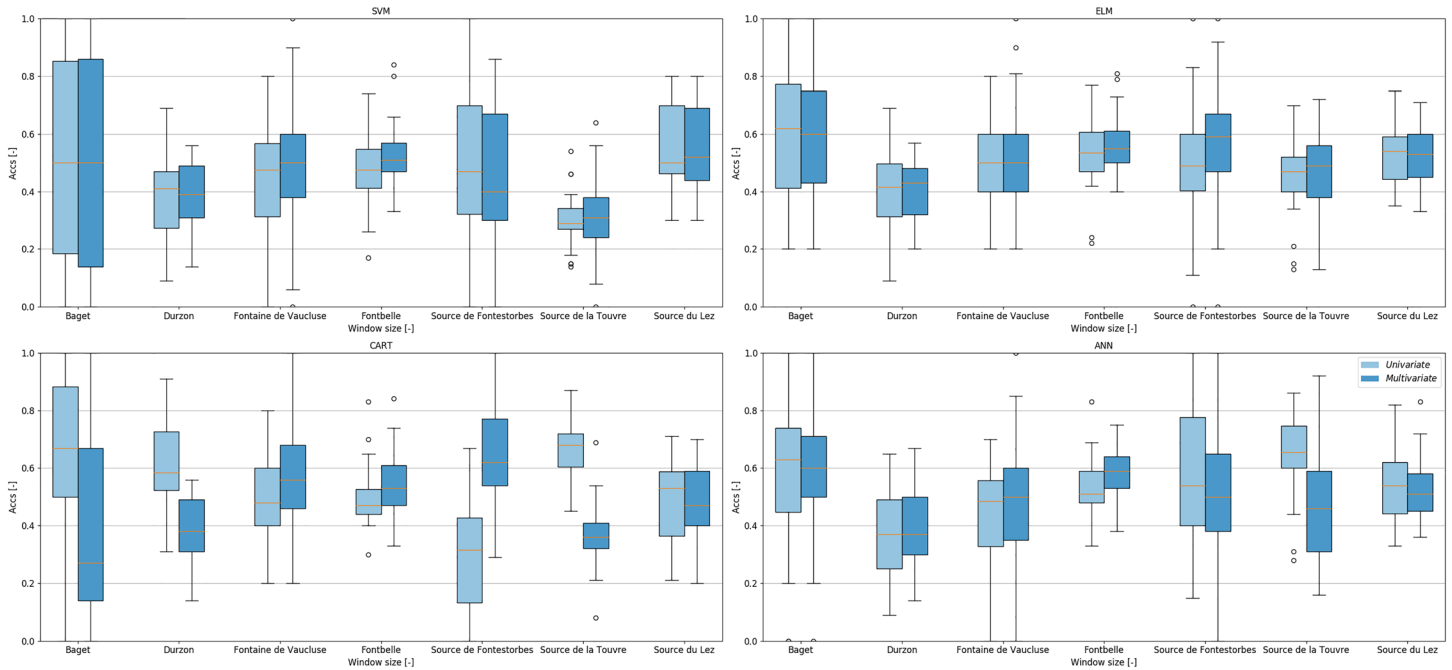


Figure 7. Accuracy Acc of the applied machines with both the univariate and the multivariate approach. The variability of the plots shows the influence of amount of training data used for prediction.

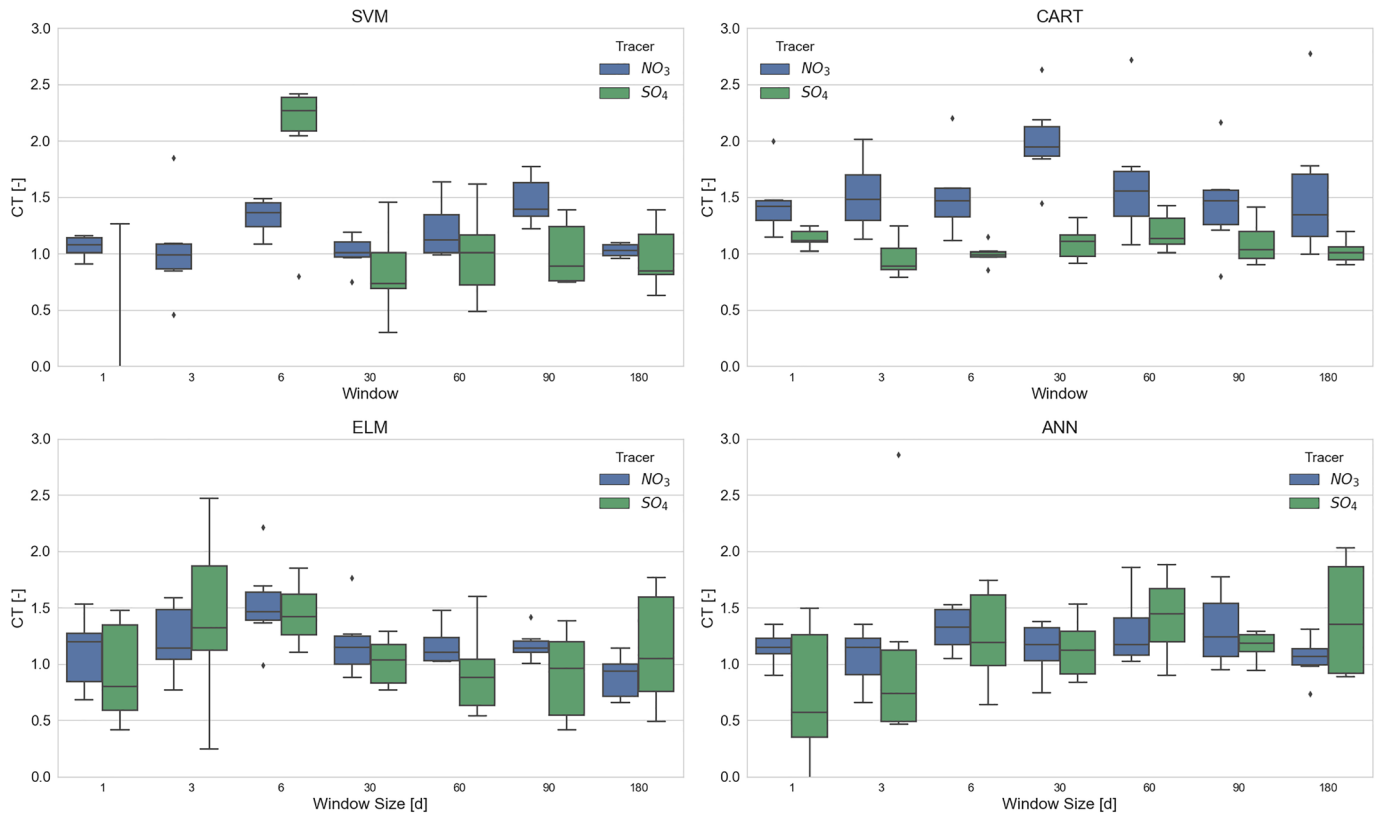


Figure 8. Dependency of chosen window length on \bar{c}_T at source de Fontestorbes of all applied machines. Good results are either achieved by short sequences (1–4 days) or long sequences (>60 days).

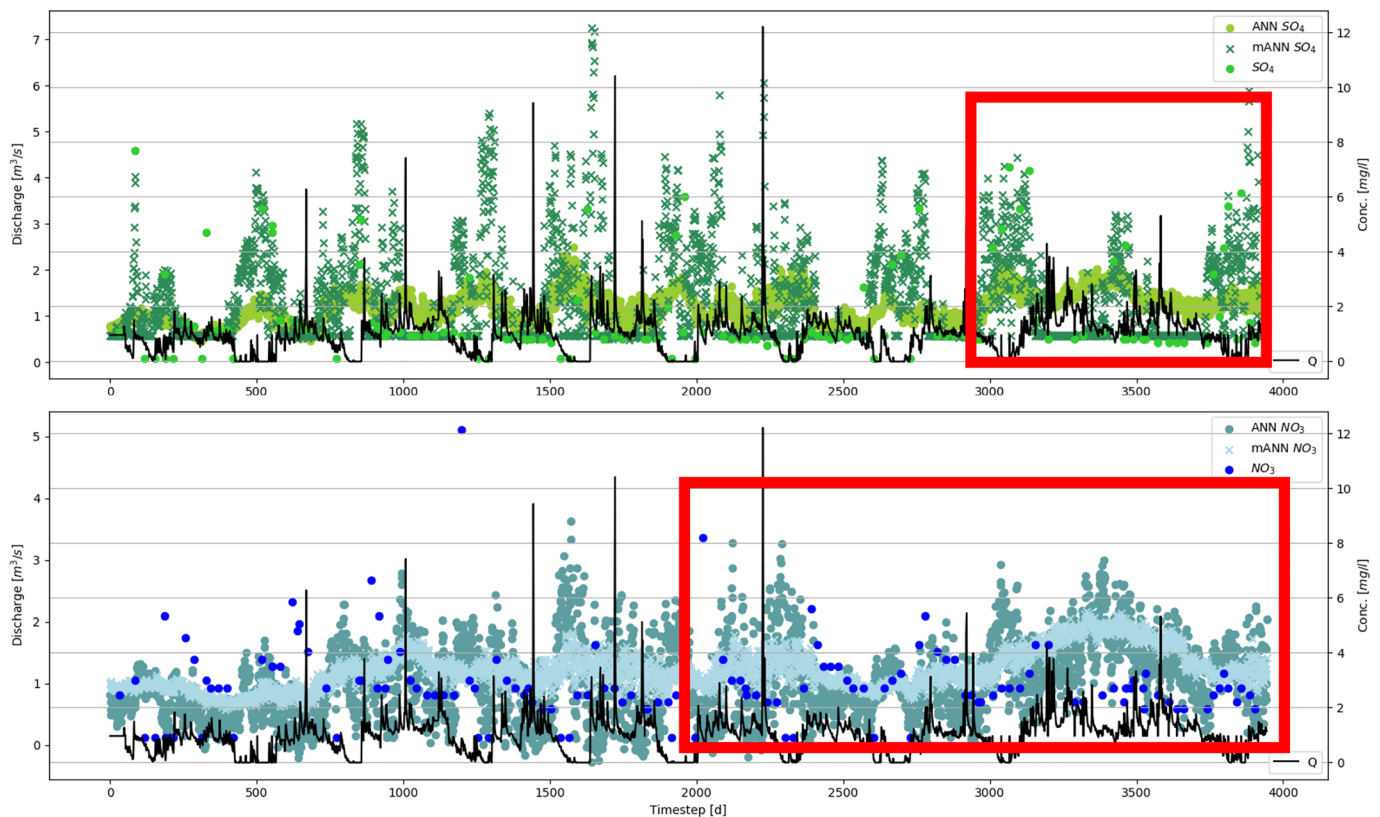


Figure 9. Interpolated time series of SO_4^{2-} and NO_3^- predicted for catchment Fontbelle with a univariate and a multivariate ANN and 80% of data used for training. The black line indicates the observed discharge dynamics.

Like for other machine learning applications in hydrology, the choice of the most promising algorithm has to be found through trial and error (He et al., 2014; Raghavendra & Deka, 2014). Hence, we adapted the research design to the No-Free-Lunch-Theorem (Wolpert & Macready, 1997) and compared four different algorithms from two of the main machine learning families (Kelleher et al., 2015). We assumed that discharge data are able to provide enough information to describe the interplay between tracer measurements and to predict the concentrations. However, the continuous entropy of discharge and mutual information between NO_3^- and SO_4^{2-} emphasized that the information needed to describe the interplay between this pair of tracers is far higher than the continuous entropy of the discharge data alone. Although the algorithms were able to predict certain aspects like the mean concentration and peaks quite well, the complete variability could not be predicted. In contrast to concentration-discharge relations that require distinct knowledge on the measured data and the catchment, our study shows that machine learning algorithms can be trained from databases with few discontinuous measurements to provide continuous reconstructions of tracer concentrations.

With knowledge on the required information content and the delivered information content, we were not able to distinguish properly among the different approaches and a further choice would depend strongly on the focus of the task: Would we like to predict the tracer concentration, or is the ranking of tracer methods for the dynamic description more important? This lack of a clear preference of the chosen machine learning methods can also be observed in other comparative machine learning studies in hydrology, for example, in flood event separation (Mewes & Oppel, 2019) and the simulation of streamflow (Shortridge et al., 2016). Similarly to their results, there might not be a single machine for all purposes that works with our data set, but a set of machines that work together to deliver the desired results, which was shown to be useful for hydrological modeling in general (Clark et al., 2008). We assume that the interplay of the information content of the tracers and discharge determines the choice of the best working algorithm. This assumed link between information content of data, prediction performance, and method preference might be a way to regionalize karst catchments by a data-driven approach (Abdollahi et al., 2017).

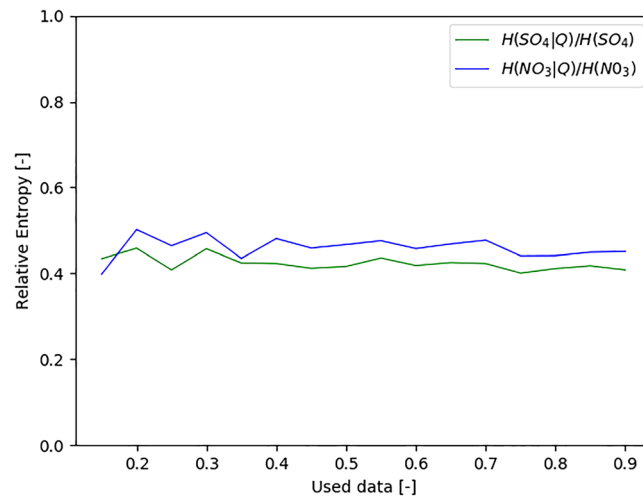


Figure 10. Conditional or relative entropy of NO_3^- and SO_4^{2-} of catchment Fontbelle. Relative entropy shows the explanatory power how much of the entropy of discharge can be used to predict a single tracer.

Consequently, our comparative analysis of algorithms and learning approaches allowed setting up a strategy to use the aforementioned algorithms to predict tracer signatures. Interestingly, the length of the input sequence of discharge consists of two groups: a group that prefers short windows and a group that prefers long windows. This might be related to different processes that relate to the transition time of the karst spring, which means that we use the information of the time spent by the water in the karst system (Hartmann et al., 2016). While SO_4^{2-} requires long times to dissolve from the karstic rock to the water, NO_3^- dissolves faster. This is the reason that the two tracers are investigated: to separate slow from fast water. Here, SO_4^{2-} could be predicted better by long windows of input data, while NO_3^- had higher performances with short input windows.

Apparently, the information that we use right now is sufficient for peak concentrations and the mean values, but concentrations of SO_4^{2-} close to nearly 0.0 mg/L lead to errors (Figure 9). Hence, processes that lead to low SO_4^{2-} concentrations in the discharge are not yet covered by the discharge data and should be included with ancillary data. Such multi-input machine learning applications are widely used in remote sensing and other applications but underrepresented in hydrology because knowledge on the information content of the input data is crucial for their application and that remains unknown in many hydrological studies (Mountrakis et al., 2011; Piotrowski et al., 2007; Zheng et al., 2015).

Overall, our investigations show that we cannot state a clear preference toward a single approach. However, the introduction of a comparative framework helps to identify the most appropriate solution to predict tracer concentrations for a specific catchment. In the following parts of the discussion, we adapt our concept of entropy and present a preliminary framework that could be used to predict tracer concentrations.

4.1. Improvements for Concept of Entropy

Due to the mixed results of the multivariate approach, we analyzed the results of both approaches, univariate and multivariate, as an example and learned that we need one tracer to predict the other. As we can see from the interpolated time series of catchment Fontbelle, the multivariate approach performed better for SO_4^{2-} than for NO_3^- . Therefore, the additional information from NO_3^- helped the algorithm to find the pattern in SO_4^{2-} . Hence, a framework should consist of a univariate ANN to predict NO_3^- , which acts as additional information to predict SO_4^{2-} .

To reveal the relationship of explanatory power between predictors and variables, we transfer the concept of mutual information to conditional, or relative, entropy (Chacon-Hurtado et al., 2017; Corzo & Solomantina, 2007; Keum & Coulibaly, 2017). The conditional entropy shows that NO_3^- has a higher conditional explanatory power than SO_4^{2-} to be predicted by discharge (Figure 10). This means that a univariate approach is

more beneficial to predict NO_3^- than it is for predicting SO_4^{2-} . Consequently, we can use the concept of conditional entropy to decide whether a univariate or a multivariate approach should be preferred and which tracer measurement can be used as ancillary data for the prediction of other tracer concentrations.

4.2. Application of Machine Learning in Interpolation of Tracer Time Series

Discharge separation by tracers relies on tracer observations which are often limited in availability (Birkel & Soulsby, 2015; Klaus & McDonnell, 2013). We assumed that machine learning is a tool to interpolate time series of tracers by discharge observations. Keeping the aforementioned downsides of machine learning in mind, the shown interpolation capability of the algorithms is a valuable addition to discharge separation applications (Garvelmann et al., 2017; Klaus & McDonnell, 2013).

As the explanatory power of discharge alone is too low to describe the interplay between the tracers in all its variations, the question toward the filling of the gaps by machine learning tools has to be precise. In our framework, an extensive preanalysis was conducted to show the general applicability in terms of RMSE and $\overline{c_T}$ for all considered algorithms and amounts of available training data. The length of the input sequence again is a source for uncertainty in our approach, but we were able to link good prediction results with the geochemical residence time of the tracer in the system. So, for hypothesis testing on transit times, the machine learning approach can be utilized. To describe the uncertainty of the prediction, both lengths of input sequences should be used: a short window length of discharge to catch short residence time processes and a long window of discharge to catch slow processes. Nevertheless, the definition of short and long windows is catchment specific and has to be determined either by a data-driven preanalysis or detailed knowledge of the respective catchment, which would be identical to the calibration of a hydrological model (Hartmann et al., 2014; Wu & Chau, 2011).

5. Conclusions

Our initial study focus explored the use of machine learning algorithms for the prediction of tracer measurements. Since time series of tracer measurements are often too sparse for modeling, machine learning tools can potentially be useful for researchers with limited access to environmental tracer data or limited resources to obtain additional measurements. We could show that our selected machine learning tools were able to identify some characteristics of the observed tracer concentrations like average concentrations or the appropriate constellations of tracer concentrations at the selected test sites. Our analysis also revealed that the information content of discharge alone is not sufficient to predict tracer concentrations with all its entire variability, as the mutual information between the pairs of tracers is higher than the continuous entropy of the discharge data. For that reason, the prediction capability of the machine learning algorithms is lowered substantially. The interpretation of the predicted time series has to be done with care, because the predicted time series lack extreme concentrations that are abundant in the observations.

Moreover, we were able to build a preliminary framework that creates an ensemble of predictions addressing the uncertainty of a machine learning-based approach by eliminating the bias of the chosen input sequence length and the learning approach of the algorithms. All methods considered in this paper deliver acceptable results in comparison, but the choice of the most suitable algorithm remains catchment specific and should be based on site-specific knowledge (e.g., residence time estimations) or extensive data-driven preanalysis. We found that the amount of required training data is high, as the mutual information between the pair of tracers requires at least 60% of the available data to reach a plateau. Hence, the training of the machines is not likely to be successful in data-poor regions.

We conclude from our investigations that the setup of a framework to predict tracer concentrations with machine learning tools remains challenging. Nevertheless, we show that the process of setting up the machine learning-based ensemble framework can be facilitated by information-based analyses like the concept of entropy, conditional entropy, and mutual information. Knowledge on the information content of the data helps to justify the nonobvious choice of methods facing “black-box” machine learning approaches. Moreover, they could be the basis for future regionalization of catchments and the transfer of trained machines to data-poor regions, in case the machine learning approaches were trained in information-rich environments. By the training with information-rich training data, linkages between processes that are hidden in data, like discharge data, become transferable and quantifiable. Hydrological models, on the other

hand, require the same amount of data regardless of their information content. So measurements too few for traditional hydrological models may still contain sufficient information to improve machine learning models. Overall, we are just at the doorstep to use data-driven approaches in hydrology, especially in complex environments like karst. Disregarding the problems that we still have to face in the future, advanced data-driven machine learning approaches may allow further improvements of data analysis, model calibration, and model development.

Although there is no silver bullet in predicting tracer concentrations, we could show by the input window analysis that the characteristics of the assumed transit time of tracers becomes visible in the most suitable input window lengths for the prediction. However, through analyzing on how a machine learns data patterns and investigating the results of the prediction, our study highlights the importance of an information content analysis. This opens the field of further entropy-based approaches of data mining in hydrological contexts, especially in often data-sparse applications like karst hydrology.

Author Contributions

The analysis and writing was conducted by B. M. and supported and advised by H. O. and A. H. V. M. conducted the literature review on the investigated karst regions, the site selection, and data preparation.

Acknowledgments

Support to Andreas Hartmann was provided by the Emmy Noether-Programme of the German Research Foundation (DFG; Grant HA 8113/1-1; project "Global Assessment of Water Stress in Karst Regions in a Changing World").

References

- Abdollahi, S., Raeisi, J., Khalilianpour, M., Ahmadi, F., & Kisi, O. (2017). Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. *Water Resources Management*, 31(15), 4855–4874.
- Aquilina, L., Ladouche, B., & Dörfliker, N. (2005). Recharge processes in karstic systems investigated through the correlation of chemical and isotopic composition of rain and spring-waters. *Applied Geochemistry*, 20(12), 2189–2206.
- Bailly-Comte, V., Ladouche, B., Allanic, C., Bitri, A., Moiroux, F., Monod, B., et al. (2018). Evaluation des ressources en eaux souterraines du Plateau de Sault - Amélioration des connaissances sur les potentialités de la ressource et cartographie de la vulnérabilité. Rapport final. BRGM/PR-67528-FR.
- BDLlisa. (2019). Retrieved from <https://bdlisa.eaufrance.fr>
- Bicalho, C. C., Batiot-Guilhe, C., Taupin, J. D., Patris, N., Van Exter, S., & Jourde, H. (2017). *A Conceptual Model for Groundwater Circulation Using Isotopes and Geochemical Tracers Coupled with Hydrodynamics: A Case Study of the Lez Karst System*. Chemical Geology, (February 2016), 0–1: France. <https://doi.org/10.1016/j.chemgeo.2017.08.014>
- Birkel, C., & Soulsby, C. (2015). Advancing tracer-aided rainfall–runoff modelling: A review of progress, problems and unrealised potential. *Hydrological Processes*, 29(25), 5227–5240. <https://doi.org/10.1002/hyp.10594>
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor & Francis.
- Chacon-Hurtado, J. C., Alfonso, L., & Solomatina, D. P. (2017). Rainfall and streamflow sensor network design: A review of applications, classification, and a proposed framework. *Hydrology and Earth System Sciences*, 21(6), 3071–3091. <https://doi.org/10.5194/hess-21-3071-2017>
- Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., & Lin, C.-J. (2010). Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research*, 11, 1471–1490.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, 44(12).
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Corzo, G., & Solomatina, D. (2007). Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. *Hydrological Sciences Journal*, 52(3), 491–507. <https://doi.org/10.1623/hysj.52.3.491>
- Eaufrance (2018a). ADES: Portail nationale d'Accès aux Données sur les Eaux Souterraines, <http://www.ades.eaufrance.fr/>
- Eaufrance (2018b). Banque Hydro, <http://hydro.eaufrance.fr/>
- Fernando, T. M. K. G., Maier, H., & Dandy, G. (2009). Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. *Journal of Hydrology*, 367. <https://doi.org/10.1016/j.jhydrol.2008.10.019>
- Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., & Dörfliker, N. (2009). Modelling the hydrologic functions of a karst aquifer under active water management—The Lez spring. *Journal of Hydrology*, 365(3–4), 235–243. <https://doi.org/10.1016/j.jhydrol.2008.11.037>
- Fotovatikhah, F., Herrera, M., Shamshirband, S., Chau, K.-W., Ardebili, S. F., & Piran, M. J. (2018). Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics*, 12(1), 411–437. <https://doi.org/10.1080/19942060.2018.1448896>
- Garvelmann, J., Warscher, M., Leonhardt, G., Franz, H., Lotz, A., & Kunstmann, H. (2017). Quantification and characterization of the dynamics of spring- and stream water systems in the Berchtesgaden Alps with a long-term stable isotope dataset. *Environmental Earth Sciences*, 76(22), 1–17. <https://doi.org/10.1007/s12665-017-7107-6>
- Gong, W., Yang, D., Gupta, H. V., & Nearing, G. (2014). Estimating information entropy for hydrological data: One-dimensional case. *Water Resources Research*, 50(6), 5003–5018. <https://doi.org/10.1002/2014WR015874>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge: MIT Press.
- Gur, D., Bar-Matthews, M., & Sass, E. (2003). Hydrochemistry of the main Jordan River sources: Dan, Baniyas, and Kezinim springs, north Hula Valley, Israel. *Israel Journal of Earth Sciences*, 52. <https://doi.org/10.1560/RRMW-9WXD-31VU-MWHN>
- Han, J., & Kamber, M. (2010). *Data Mining: Concepts and Techniques, the Morgan Kaufmann Series in Data Management Systems*. Amsterdam: Elsevier.
- Hartmann, A., Barberá, J. A., & Andreo, B. (2017). On the value of water quality data and informative flow states in karst modelling. *Hydrology and Earth System Sciences*, 21(12), 5971–5985. <https://doi.org/10.5194/hess-21-5971-2017>

- Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y., & Wagener, T. (2015). A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. *Geoscientific Model Development*, 8(6), 1729–1746. <https://doi.org/10.5194/gmd-8-1729-2015>
- Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics*, 52(3), 218–242. <https://doi.org/10.1002/2013RG000443>
- Hartmann, A., Kobler, J., Kralik, M., Dirnböck, T., Humer, F., & Weiler, M. (2016). Model-aided quantification of dissolved carbon and nitrogen release after windthrow disturbance in an Austrian karst system. *Biogeosciences*, 13(1), 159–174. <https://doi.org/10.5194/bg-13-159-2016>
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Upper Saddle River: Prentice Hall.
- He, Z., Wen, X., Liu, H., & 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. *Journal of Hydrology*, 509, 379–386. <https://doi.org/10.1016/j.jhydrol.2013.11.054>
- Hu, C., J.-j. Wang, Z.-n. Wu, and Lina-Liu (Eds.) (2011). Application of the support vector machine on precipitation-runoff modelling in Fenhe River, 2011 International Symposium on Water Resource and Environmental Protection, 1099–1103, vol. 2.
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2004). Extreme learning machine: A new learning scheme of feedforward neural networks. *Neural Networks*, 2, 985–990.
- Jacob, T., Bayer, R., Chery, J., Jourde, H., Le Moigne, N., Boy, J.-P., et al. (2008). Absolute gravity monitoring of water storage variation in a karst aquifer on the Larzac plateau (southern France). *Journal of Hydrology*, 359(1–2), 105–117. <https://doi.org/10.1016/j.jhydrol.2008.06.020>
- Jourde, H., Massei, N., Mazzili, N., & Binet, S. (2018). SNO KARST: A French network of observatories for the multidisciplinary study of critical zone processes in karst watersheds and aquifers. *Vadose Zone Journal*, 17(1).
- Kavouri, K., Plagnes, V., Tremoulet, J., Dörfli, N., Rejiba, F., & Marchet, P. (2011). PaPRIKa: A method for estimating karst resource and source vulnerability—Application to the Ouysses karst system (Southwest France). *Hydrogeology Journal*, 19(2), 339–353. <https://doi.org/10.1007/s10040-010-0688-8>
- Kelleher, C., Ward, A., Knapp, J. L. A., Blaen, P. J., Kurz, M. J., Drummond, J. D., et al. (2019). Exploring tracer information and model framework trade-offs to improve estimation of stream transient storage processes. *Water Resources Research*, 55, 3481–3501. <https://doi.org/10.1029/2018WR023585>
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge: MIT Press.
- Keum, J., & Coulibaly, P. (2017). Information theory-based decision support system for integrated design of multivariable hydrometric networks. *Water Resources Research*, 53, 6239–6259. <https://doi.org/10.1002/2016WR019981>
- Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological Processes*, 17(4), 871–874.
- Klaus, J., & McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, 505, 47–64.
- Labat, D., Mangin, A., & Ababou, R. (2002). Rainfall-runoff relations for karstic springs: Multifractional analyses. *Journal of Hydrology*, 256, 176–195. [https://doi.org/10.1016/S0022-1694\(01\)00535-2](https://doi.org/10.1016/S0022-1694(01)00535-2)
- Le Moine, N., Andréassian, V., & Mathevet, T. (2008). Confronting surface- and groundwater balances on the La Rochefoucauld-Touvre karstic system (Charente, France). *Water Resources Research*, 44, W03403. <https://doi.org/10.1029/2007WR005984>
- Lee, E. S., & Krothe, N. C. (2001). A four-component mixing model for water in a karst terrain in south-Central Indiana, USA. Using solute concentration and stable isotopes as tracers. *Chemical Geology*, 179(1), 129–143.
- Mahler, B. J., & Garner, B. D. (2009). Using nitrate to quantify quick flow in a karst aquifer. *Ground Water*, 47(3), 350–360. <https://doi.org/10.1111/j.1745-6584.2008.00499.x>
- Mei, Y., & Anagnostou, E. N. (2015). A hydrograph separation method based on information from rainfall and runoff records. *Journal of Hydrology*, 523, 636–649. <https://doi.org/10.1016/j.jhydrol.2015.01.083>
- Mewes, B., & Oppel, H. (2019). A comparative analysis of machine learning tools for hydrograph separation. *Frontiers in Water Complexity*. submitted
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Mudarra, M., & Andreo, B. (2011). Relative importance of the saturated and the unsaturated zones in the hydrogeological functioning of karst aquifers: The case of Alta Cadena (southern Spain). *Journal of Hydrology*, 397(3–4). <https://doi.org/10.1016/j.jhydrol.2010.12.005>
- Mudarra, M., Hartmann, A., & Andreo, B. (2019). Combining experimental methods and modeling to quantify the complex recharge behavior of karst aquifers. *Water Resources Research*, 55, 1384–1404. <https://doi.org/10.1029/2017WR021819>
- Nasrabadi, N. M. (2007). Pattern recognition and machine learning. *Journal of Electronic Imaging*, 16(4), 49,901.
- Nourani, V., Komasi, M., & Mano, A. (2009). A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water Resources Management*, 23(14), 2877. <https://doi.org/10.1007/s11269-009-9414-5>
- Piotrowski, A., Wallis, S. G., Napiórkowski, J. J., & Rowiński, P. M. (2007). Evaluation of 1-D tracer concentration profile in a small river by means of multi-layer perceptron neural networks. *Hydrology and Earth System Sciences*, 11(6), 1883–1896. <https://doi.org/10.5194/hess-11-1883-2007>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1007/BF00116251>
- Raghavendra, N. S., & Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19, 372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rimmer, A., & Hartmann, A. (2014). Optimal hydrograph separation filter to evaluate transport routines of hydrological models. *Journal of Hydrology*, 514, 249–257. <https://doi.org/10.1016/j.jhydrol.2014.04.033>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sharma, A. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1—A strategy for system predictor identification. *Journal of Hydrology*, 239(1), 232–239.
- Shorridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, 20(7), 2611–2628. <https://doi.org/10.5194/hess-20-2611-2016>
- Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, 19(2), 225–235. <https://doi.org/10.1016/j.neunet.2006.01.012>

- Tabari, H., Kisi, O., Ezani, A., & Hosseinzadeh Talaei, P. (2012). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology*, *444*, 78–89. <https://doi.org/10.1016/j.jhydrol.2012.04.007>
- Taormina, R., Chau, K.-W., & Sivakumar, B. (2015). Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *Journal of Hydrology*, *529*, 1788–1797.
- Thomas, J. A., & Cover, T. M. (2006). *Elements of Information Theory*. NY, USA: Wiley New York.
- Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. New York: Springer science & business media.
- Weiler, M., Seibert, J., & Stahl, K. (2017). Magic components—Why quantifying rain, snow- and icemelt in river discharge isn't easy. *Hydrological Processes*, *32*(1), 160–166. doi:<https://doi.org/10.1002/hyp.11361>
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.
- Wu, C. L., & Chau, K. W. (2011). Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology*, *399*(3–4), 394–409.
- Yaseen, Z. M., Jaafar, O., Deo, R. C., Kisi, O., Adamowski, J., Quilty, J., & El-Shafie, A. (2016). Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. *Journal of Hydrology*, *542*, 603–614. <https://doi.org/10.1016/j.jhydrol.2016.09.035>
- Yu, P.-S., Yang, T.-C., Chen, S.-Y., Kuo, C.-M., & Tseng, H.-W. (2017). Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology*, *552*, 92–104. <https://doi.org/10.1016/j.jhydrol.2017.06.020>
- Zheng, B., Myint, S. W., Thenkabail, P. S., & Aggarwal, R. M. (2015). A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, *34*, 103–112. <https://doi.org/10.1016/j.jag.2014.07.002>