



A catchment-scale model of river water quality by Machine Learning

Maria Grazia Zanoni, Bruno Majone, Alberto Bellin *

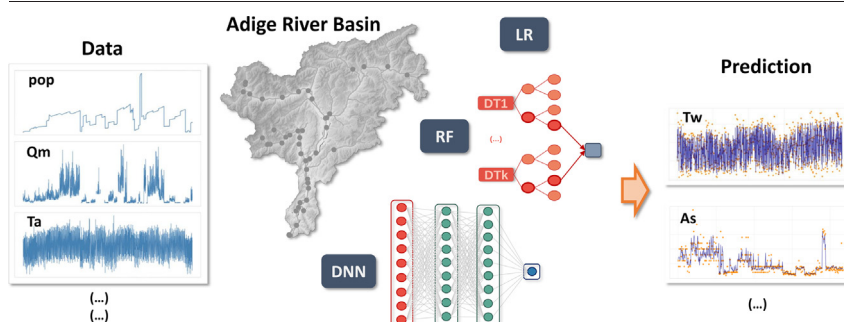
Department of Civil, Environmental and Mechanical Engineering, University of Trento, via Mesiano 77, I-38123 Trento, Italy



HIGHLIGHTS

- Machine Learning techniques are tested for predicting water quality parameters in an Alpine watershed.
- Deep feed-forward Neural Network showed the best performances compared to Random Forest and Linear Regression techniques.
- The role played by the drivers has been investigated by means of a Feature Importance Assessment procedure.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Fernando A.L. Pacheco

Keywords:

Regression Features Importance Assessment
Random Forest
Deep feed-forward
Neural Network
Linear Regression

ABSTRACT

Water quality is a concern in most river basins worldwide due to the widespread release of pollutants which impacts the freshwater ecosystems. Exploring the relationships between drivers and water quality parameters at the regional scale is key in the identification of appropriate actions for the reduction of water pollution. Regional models are the appropriate tool to achieve this, though their development poses relevant challenges because of the complexity and non-linearity of such relationships. Among the available approaches, Machine Learning (ML) is promising because of its capability to detect complex nonlinear relationships and flexibility in the parameterization, which is learned from data. In this work, we developed regional models of water temperature, dissolved oxygen, arsenic, sulfate and chloride concentrations, as well as electrical conductivity, by using two ML algorithms, Random Forest and Deep feed-forward Neural Network, and compared their performances against the standard Linear Regression model. Our results indicate that the two ML algorithms are much more accurate models for such variables than the classical Linear Regression model, with Deep feed-forward Neural Network being the most effective in identifying the reciprocal importance of the drivers and capturing nonlinear relationships between drivers and water quality variables. Our analysis also revealed that the Julian day and year at which the sample was taken surrogate the air temperature in modeling water temperature and dissolved oxygen, with only a slight performance reduction. Arsenic, sulfate, and chloride show more complex behaviors in which geogenic and anthropogenic sources are intertwined. Dilution exerts a role chiefly for arsenic concentration, which suggests a non-uniform, in space, geogenic origin for this variable.

1. Introduction

The deterioration of river water quality is one of the most relevant challenges that humanity is facing since the dawn of the industrial revolution.

Among the most relevant pollutants, chemicals create concern to regulatory authorities, protection agencies, and the public as well for their adverse impact on ecosystems and human health (Schwarzenbach et al., 2010; Landrigan et al., 2018). Likewise, metals, metalloids, and solutes of

Abbreviations: AI, Artificial Intelligence; ML, Machine Learning; DT, Decision Tree; RF, Random Forest; LR, Linear Regression; ANNs, Artificial Neural Networks; DNN, Deep feed-forward Neural Network; AIC, Akaike's Information Criterion; PCA, Principal Component Analysis.

* Corresponding author: Department of Civil, Environmental and Mechanical Engineering, University of Trento, via Mesiano 77, 38123 Trento, Italy.

E-mail address: alberto.bellin@unitn.it (A. Bellin).

<http://dx.doi.org/10.1016/j.scitotenv.2022.156377>

Received 28 February 2022; Received in revised form 27 May 2022; Accepted 27 May 2022

Available online 03 June 2022

0048-9697/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

geogenic origin, arsenic and sulfates for example, create concern as well, given the similar risks they pose to human and freshwater ecosystems' health (Neumann et al., 2010; Harvey et al., 2002; Smedley and Kinniburgh, 2002). Other important physical parameters are water temperature and pH because they are indicators of the river ecological status (see, e.g., Caissie, 2006).

Modeling the spatial and temporal evolution of these and other water quality parameters is challenging but at the same time crucial in assessing river functionality and protecting the health of riverine ecosystems (see, e.g., Rinaldo et al., 2011; Rügner et al., 2013; Navarro-Ortega et al., 2015; Ma et al., 2020) and ultimately human health (Landrigan et al., 2018). Detecting the relationships between water quality parameters and their drivers is also key in developing strategies for water resources management in human-impacted watersheds (Guillet et al., 2019). In the recognition of the importance of achieving a good ecological status in European rivers, monitoring networks have been established by river authorities under the impulse of the European Union and this effort has been intensified after the introduction of the European Water Framework Directive (WFD, 2000/60/EC) in 2000 (Council of European Union, 2000) and research initiatives have been developed (see, e.g., Grathwohl et al., 2013). This common effort produced a wealth of data, though with spatial and temporal resolutions varying among countries and in some cases also among regions of the same country. Despite this large investment in data collection, which has been accompanied by a similar one in data interpretation, a full understanding of the relationships between drivers and water quality parameters has not yet been achieved, thereby impairing remediation actions.

Predictive tools are useful in assessing the impact of the hydrological and anthropogenic stressors on water quality variables (Navarro-Ortega et al., 2015; Kormann et al., 2015; Chiogna et al., 2016; Lutz et al., 2016; Diamantini et al., 2018). Dilution and mixing-controlled biochemical reactions smooth the concentration signal with respect to that of the external forcing, and quantifying these effects is one of the main objectives of physically-based models (see, e.g., Kirchner et al., 2000, 2001; Kirchner and Neal, 2013; Botter et al., 2006; McDonnell et al., 2010). However, the lack of a shared blueprint concerning data collection hampers the application of these models, which are better constrained when data are available at a relatively high time resolution. In fact, model's parameters controlling concentration attenuation are loosely constrained when data are available at a frequency too low to resolve variability at scales comparable with those of hydrological fluxes. Unfortunately, this is the case with most concentration databases (see, e.g., Chappell et al., 2017).

Traditionally river water quality modeling studies resort to the solution of the Advection Dispersion Equation along the river reaches, with the inclusion of geochemical and hyporheic exchange processes at different levels of complexity (see, e.g., Wörman, 1998; Marion et al., 2008; Zarnetske et al., 2012; Diamantini et al., 2019). Other approaches adopt a catchment point of view considering the river corridor as part of a more complex natural system working as a reactor that elaborates the material entering from the Earth's surface (Basu et al., 2010; Rinaldo et al., 2011; Grathwohl et al., 2013). Besides uncertainty stemming from the parameterization of the many processes controlling river water quality, the success of process-based modeling approaches is limited by the large errors and uncertainties characterizing sources of geogenic and anthropogenic origin.

An alternative approach, less sensitive to the unavoidable errors in source quantification, is to use data-driven models such as Machine Learning (ML), which learn directly from the available data (Mitchell et al., 1997; Nearing et al., 2021). These techniques are less affected by source errors and irregular sampling than process-based models, though they require a relatively large amount of data to infer the cause-effect relationship. Furthermore, ML techniques pose fewer restrictions to the parameterization of the often poorly known physical and biogeochemical processes driving transport and transformation of contaminants in the riverine environment, and they are designed to detect non-linear behaviors (Shen, 2018). Under this perspective, ML techniques are expected to be effective in the prediction of water quality parameters for their capability to "learn" from data

the intertwined and nonlinear relationships between river water quality parameters and their drivers (Piggott et al., 2012). The epistemic error affecting process-based models as a consequence of fixing a-priori the relationship between input and output is here limited to the selection of the hyper-parameters with the functional relationship between the input information that is learned from data.

Early applications of ML date back to the 90s of the last century with the introduction of Artificial Neural Networks (ANNs) in various scientific disciplines, among which hydrology (Buch et al., 1993; Hsu et al., 1995). More recently, deep learning algorithms such as Recurrent Neural Network and Convolutional Neural Network, constituted by hierarchical architectures of ANNs, have been used in hydrology (see, e.g., Shen, 2018; Xu and Liang, 2021, for a review). Since their appearance, ANNs and their various extensions have been used in water quality studies concerning both surface and subsurface water bodies (Basant et al., 2010; Chen and Liu, 2014; Nemati et al., 2015; Csábrági et al., 2017; Zhu et al., 2018; Al-Mukhtar and Al-Yaseen, 2019; Shah et al., 2021; Kamrava et al., 2021). These modeling studies focus on the prediction of a single or few water quality variables, and few attempts have been done so far to develop regional models for the class of water quality variables identified as relevant for assessing the chemical and ecological status of a river in a multi-stressors situation (Tiyasha et al., 2020; Navarro-Ortega et al., 2015).

Contributions combining ANNs with process-based models or wavelets transformation have been proposed in several hydrological applications (see, e.g., Hasan et al., 2011; Tiyasha et al., 2020; Xu and Liang, 2021). For example, Shukla et al. (2022) compared a feed-forward ANNs model with algorithms having a more complex architecture, obtained by coupling the ANN algorithm with wavelets or with an adaptive-neuro fuzzy inference system, obtaining with the latter an accurate reproduction of stream flow in an Indian river. Furthermore, Shamshirband et al. (2019) presented an ensemble model composed of multiple wavelets-ANNs, which improved individual models' performances of chlorophyll and salinity in coastal water, especially for multi-step ahead modeling. Nourani et al. (2018) combined streamflow simulations by the Hydrologic Modeling System (HEC-HMS) with a ML model to infer Storage Coefficient variations from the simulated runoff. In addition, Dalla Libera et al. (2020) applied Self Organizing Map to explore the correlations between arsenic, groundwater levels, and other chemical compounds, and afterward simulated arsenic mobility in an aquifer under different redox conditions. The combination with wavelets, and to some extent also with process-based models, requires time series of observational variables with time steps small enough to capture all the relevant scales of variability. This characteristic is seldom respected in available datasets of water quality data, whose frequency of acquisition is in most cases seasonal (i.e. 4 values per year) given the cost of sampling and successive chemical analyses.

The capabilities of ML techniques have been exploited also in studies dealing with water quality classification. Nafi et al. (2020) applied Random Forest and Random Tree techniques to classify rivers into quality classes according to the values assumed by electrical conductivity, temperature, concentration of fecal and total coliforms, Biological Oxygen Demand, nitrate, supplemented by the year and the location of the measurements. Moreover, Fan et al. (2020) applied Support Vector Machine to assess the response of the phytoplankton community to multiple environmental stressors both at the catchment- and reach-scale, trying also to identify those mostly affecting the ecological status of the river. Massei et al. (2018) investigated, through k-means clustering, toxic risk in major European rivers in terms of the effects induced by the presence of pesticides and biocides.

In existing ML applications, limited attention has been given to the importance of identifying the most relevant drivers to be used as model input. We consider this important because the number of drivers used as input is a structural parameter, a hyper-parameter in the ML jargon, influencing the number of fitting parameters. Given that more drivers as input lead to more parameters to be fitted, the inclusion of a driver with a weak influence on the modeled variable may lead to overfitting, a situation not conducive to reliable modeling output, though the fitting error may be small. The capability of ML to learn from the data, thereby giving small weights to less

influential drivers, compensates only partially for this conundrum. There is therefore a paramount interest in including only the drivers with a significant influence on the modeled variable, and we are not aware of systematic procedures aiding the modeler in this identification. In addition, ML performed with the most influential drivers allows for evaluating their synergistic effects on water quality variables, which is key to identifying actions for reducing pollution. Again, to the best of our knowledge, only a few studies analyzed the synergistic effect of multiple drivers on river water quality (Heddad and Kisi, 2017; Hu et al., 2020; Krishnaraj and Deka, 2020; Kouadri et al., 2021).

As mentioned above, time series of water quality data are typically available at a few positions along the river network and at a frequency that is seldom higher than monthly or seasonal, thereby methodologies of spatialization and possibly temporal filling are useful in river quality management and in identifying possible remediation actions. In the present work, we explore the capability of data-driven ML techniques to construct effective regionalization models of water quality data. For this purpose, we applied two widely used ML techniques, Random Forest (RF) and Dense feed-forward Neural Network (DNN) to a water quality dataset of the Adige river, a large Alpine River Basin in northeastern Italy (Fig. 1), with drivers time series collected in a previous work by Diamantini et al. (2018), which focused on the statistical analysis of water quality data. Water quality monitoring in this catchment is performed according to the guidelines of the European Water Framework Directive and thereby it is representative of the data available in most of the European rivers. In the present work, we seek for a data-driven modeling approach, general enough to be applied in other similar contexts in terms of data availability, able to make the best possible use of data collected according to monitoring protocols adopted by Environmental Protection Agencies, and not necessarily developed to answer specific research questions.

The objective of the present study is therefore twofold. Firstly we will explore the benefits of using Random Forest and Dense feed-forward Neural Network instead of standard regression to model spatial and temporal variability of water quality parameters selected among those monitored across Europe according to the Water Framework Directive. A second, more important, objective is to develop an approach for ranking the drivers from the most influential to the less influential and excluding those bringing a marginal improvement at the cost of overfitting because more drivers call

for more parameters. We showed that the most relevant drivers are variable specific and the analysis we proposed in the present work was able to identify them.

2. Data

2.1. Study basin

With a contributing area of 12100 km² the Adige river basin, located in the southern oriental Alpine region northeast of Italy (Fig. 1), is the third-largest Italian river basin after Po and Tiber. The Adige river rises from a spring in the proximity of the Resia lake in the northwestern part of the watershed at the elevation of 1586 m a.s.l.. It first flows eastward to Bolzano municipality and then south, through the Adige valley to Trento and Verona. Downstream of Verona it turns east to reach the Adriatic sea, after a total path of 409 km. The most relevant tributaries are in its upper (Isarco) and middle (Noce and Avisio) course (see Fig. 1). Sampling sites considered in this study are located in the upper and middle course of the river in the territories of the Autonomous Provinces of Trento and Bolzano, which cover about 91 % of the total contributing area.

We selected this river basin because sampled by the two Environmental Protection Agencies insisting on the territory according to the indications of the Water Framework Directive and with a sufficient spatial resolution to obtain samples representative of the different pressures an Alpine river is subject to. In addition, the river basin is located in an area with significant human pressures from agricultural and industrial activities along the main stem and by a quite intensive tourism in the headwaters. This is a common situation in a number of mountain areas of the world.

Streamflow presents a typical alpine regime, with high flows in early summer, due to snow melting, or in autumn, triggered by cyclonic storms (Lutz et al., 2016; Laiti et al., 2018; Mallucci et al., 2019). The geographical location of the study area, in the southern oriental Alps, and its high altitudinal variation (from about 130 to approximately 3900 m a.s.l.) define the climate of the watershed, with dry winters, snow and glacier melt in spring and high humidity in summer, and autumn. The annual average precipitation ranges from about 500 mm in the northwestern area to 1600 mm in the southern portion of the basin (Chiogna et al., 2016; Mallucci et al., 2019). A long-term annual mean air temperature of 9.3 °C was obtained from the available time series of the investigated sites in the period 1956–2014 (see Section 2.2).

2.2. Water quality variables and drivers

The water quality dataset comprises six parameters monitored seasonally between 1994 and 2014 at the 45 sites shown in Fig. 1. The monitored parameters are arsenic (As), sulfates (SO₄), and chloride (Cl), supplemented by water temperature (Tw), electrical conductivity (Ec), and dissolved oxygen (DO). The data were provided by the Environmental Protection Agencies of the Autonomous Provinces of Trento (<http://www.appa.provincia.tn.it>) and Bolzano (<http://www.provincia.bz.it/agenzia-ambiente>). Other available variables were excluded from the analysis for the following reasons. Biological Oxygen Demand (BOD) is strongly related to the Dissolved Oxygen, which dataset was more reliable and was included in the analysis, while Chemical Oxygen Demand (COD) and suspended solids were excluded because very few values were available. pH, along with COD and BOD, are strongly related to DO and Ec, and therefore we decided to focus on the latter. The total phosphorus and orthophosphates (PO₄), such as total nitrogen (N_{tot}), provided unsatisfactory performances, and nitrate is not a real concern in the Adige catchment and was already studied in the paper by Lutz et al. (2016). This study concluded that nitrate concentration is reducing, especially during the last decades, likely due to local agricultural practices improvements. Tw is an important variable for its influence on ecosystem functioning and was the object of several modeling efforts considering air temperature as the primary driver of change and simplified process-based models (Toffolon and Piccolroaz, 2015; Piccolroaz et al., 2016). The skills of these models have been proved to be comparable

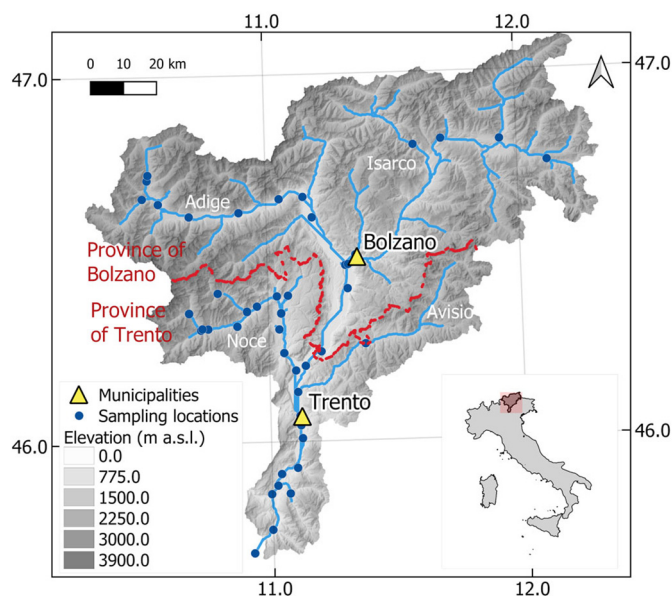


Fig. 1. Map of the upper and middle Adige river basin in WGS84 coordinate system with the 45 sampling locations indicated by blue dots and the Municipalities of Trento and Bolzano. The lower right inset shows the location of the Adige river basin in the Italian territory.

with those obtainable by ML (see, e.g., Qiu et al., 2021; Zhu et al., 2019; Lencioni et al., 2022), and in the present work, we will show that the information carried by the date (Year and Julian day) of the sampling is a good proxy of air temperature when modeling is performed by ML.

The basic statistics of the six variables are provided in Table 1, and the seasonal maps of their distribution are presented in the Supplementary Material (from Figs. S1 to S4). The statistics have been computed considering the data collected at all sampling sites within the river basin. Average values of chemical contaminants are 0.0028 mg/L, 34.2 mg/L, and 4.4 mg/L, for As, SO₄, and Cl, respectively. With these values, the chemical status of the Adige river basin can be classified as good according to the national limits for surface waters (Decreto Legislativo 13 ottobre 2015, n.172, 2015). As and Cl are positively skewed, with skewness between 0.83 and 1.46 (see Table 1). On the contrary, DO, Ec, and Tw are almost symmetrically distributed with a smaller skewness and average values of 11.6 mg/L, 220.4 μ S/cm, and 8.5°C, respectively. The total number of collected measurements for each variable ranges from 3392 to 3924, except As, for which only 1685 samples are available.

Since one of the main objectives of this work is to identify which stressors are more influential in modeling water quality parameters routinely monitored in the Adige watershed, drivers of hydrological and anthropogenic nature have been identified and collected. Changes in water quality variables in the Adige are driven by changes in the releases due to human activities, climate change, and changes in stream dilution, which is indeed controlled by streamflow (see e.g., Navarro-Ortega et al., 2015; Chiogna et al., 2016). In particular, changes in streamflow occur as the combined effect of climate change and water uses due to human activities. All these effects are present in the Adige river basin and can be epitomized by the air temperature, streamflow, land use, and the distribution of the population. In addition to these drivers, lithology may have an impact on solutes with a geogenic, though not exclusive, origin. A driver describing the relevant lithological characteristics, which may change depending on the considered water quality variable, cannot be identified, but the introduction of the spatial coordinates of the sampling points acts as a proxy for this driver. In addition, we introduced the Julian day and the year of the measure as additional drivers to include into the analysis the effects of seasonality.

Monthly aggregated streamflow data (Qm) were retrieved from the Hydrological Offices of the Provinces of Trento (<http://www.floods.it/public/index.php>) and Bolzano (<http://www.provincia.bz.it/hydro/index.i.asp>). Daily air temperatures Ta, for the period 1994–2014, were extracted from the database of gridded air temperatures (1956–2013) provided by Mallucci et al. (2019) via interpolation of daily air temperature data at 350 meteorological stations within the Adige watershed. The other drivers were: the human population density (pop), obtained from the National Aeronautics and Space Administration-Socioeconomic-Data-and-Applications-Center (CIESIN, 2017) at an interval of 5 years and the percentage of agricultural and artificial land use, indicated with agr and art, respectively, retrieved from the Corine Land Cover dataset, with the reference inventory of 1990 and the relative updates produced in 2000, 2006 and 2012 (EEA, 2013). The driver artificial land use includes the following classes from the Corine land Cover product: “urban fabric”, “industrial, commercial and transport units”, “mine, dump and construction sites”, and “artificial, non-agricultural vegetated areas”. In addition, the

Julian Day and the year at which the sample was taken (JD and Y, respectively), the altitude (Z) and the coordinates (coord, expressed in the WGS84 coordinate system) of the sampling sites have been considered, such as including spatio-temporal dependencies into the ML models.

Additional details on the water quality dataset can be found in Diamantini et al. (2018) and Mallucci et al. (2019) for what concerns the spatialization of temperature data.

3. Methods

In the present work, we modeled separately the six selected water quality variables described in Section 2.2 by adopting the same model parameterization for all sampling locations. Compared to the widely used site-dependent parameterization, the advantage of our approach is that the resulting regional models can be employed to simulate the variables of interest at unsampled locations.

Both RF and DNN offer great flexibility to accommodate for (mostly unknown) non-linear relationships between drivers and modeled parameters, but this comes with a cost: the risk of overfitting that grows with the number of drivers. Consequently, we analyzed the performances of RF and DNN by including the drivers progressively to evidence at each step the gain in performance granted by the increase of model complexity due to the introduction of the new driver. These ML models suffer from the limitation of not considering the temporal correlation of the modeled variable. However, this limitation is not of concern in the present work, given the seasonal scale at which the data are collected, as it often happens to make affordable the cost of long-term water quality monitoring programs. Although more advanced ML techniques, such as Long Short-Term Memory (LSTM, see, e.g., Shen, 2018), may have been applied, we kept the analysis intentionally simple by employing the above widely used algorithms, owing to our main objective of developing regional models of water quality, properly balancing overfitting and regularization, rather than identifying the best possible ML algorithm for the data at hand. Furthermore, the additional ingredient that LSTM introduces is the capability to learn how long the state information should be retained (i.e., the correlation time scale), a characteristic not crucial for our analysis due to the seasonal time scale of the water quality data. In the present work, the capability of RF and DNN to exploit nonlinear relationships between drivers and water quality data was evidenced by comparing their performance against that of LR for all the investigated water quality variables. A detailed description of the models RF, DNN and LR is provided in the Appendix A.

3.1. Data organization, efficiency metrics and cross-validation

As stated above, our main objective in this work was to build a regional water quality model that uses all the available information. Taking advantage of the fact that RF, DNN, and LR do not consider the temporal correlation between the observational variables, the time series of the six water quality variables were merged separately following the topological order of the river network from the headwaters to the main stem. The resulting models are regional thanks to the inclusion of the spatial coordinates and elevation of the sampling points as drivers and the spatial variability of the other drivers, such as the air temperature.

Each time series obtained with this merging was split into two parts: the training and the test sets, as customary when the number of data is not enough to extract also a validation set to be used for an independent validation after training (Raschka, 2015). To accurately select the hyperparameters (see Section Appendix A for their definition) a k -fold cross-validation procedure was applied by randomly splitting the training set into $k = 30$ folds without replacement. The number of folds was chosen after a preliminary sensitivity analysis that indicated that $k = 30$ folds yield comparable efficiency metrics in the folds and the training sets. A similar result with $k = 30$ folds was also obtained in Hadjisolomou et al. (2021), where the authors investigated a range of k values and also the leave-one-out cross-validation, noticing the risk of overfitting with this extreme case. The cross-validation requires that the time series elements

Table 1

Main statistics of the investigated water quality time series: mean, median, minimum, maximum, standard deviation and skewness. The statistics were computed considering the time series available at all the Adige river basin sampling sites.

	Mean	Median	Min	Max	St. Dev.	Skewness
As (mg/L)	0.0028	0.0020	0.0005	0.0620	0.0028	1.463
SO ₄ (mg/L)	34.2	30	0.8	288	23.6	0.829
Cl (mg/L)	4.4	3.4	0.1	118	4.6	1.308
Ec (μ S/cm)	220.4	218	5.4	636	78.6	0.41
Tw (°C)	8.5	8.5	0	26.4	4.4	0.087
DO (mg/L)	11.6	11.5	6.8	16.1	1.4	0.289

are statistically independent and identically distributed, a tenet that is satisfied here owing to the considerable time lag between successive measurements. Iteratively, $k - 1$ folds were used as a training set and the remaining one as the test set.

As customary in ML applications, training was performed by using the Mean Squared Error (MSE) as the efficiency metric:

$$MSE = \frac{1}{N} \sum_{i=1}^N [y_{pred,i} - y_{meas,i}]^2 \quad (1)$$

where N is the number of data in the training set.

The hyper-parameters providing the largest mean Nash-Sutcliffe (NSE) index in the k -fold cross-validation are then selected for the successive optimization conducted with the entire training set (i.e., 70 % of the data) in order to obtain the final model's parameters. The NSE index is defined as follows (Nash and Sutcliffe, 1970):

$$NSE = 1 - \frac{\sum_{i=1}^{N_T} [y_{pred,i} - y_{meas,i}]^2}{\sum_{i=1}^{N_T} [y_{meas,i} - \mu_{meas}]^2} \quad (2)$$

where N_T is the total number of samples in the fold used as test set, $y_{meas,i}$ and $y_{pred,i}$ are the measured (observed) and predicted (by the model) values of the water quality parameter, with the first two sampling statistics (mean and standard deviation) indicated by μ_{meas} , σ_{meas} and μ_{pred} , σ_{pred} , respectively. NSE (Eq. (2)) can assume values in the range $(-\infty, 1]$ with negative values identifying models to be rejected because not performing better than the mean of the observed values. Models with $NSE > 0.5$ are considered behavioral and, therefore, potentially acceptable (see, e.g., Moriasi et al., 2007). Notice that we used this metric to select the hyper-parameters instead of MSE because it is sensitive to the variance of the signal and is commonly used in hydrological applications.

Performance analysis of the final models (RF, DNN, and LR), obtained by training with the entire training set on both training and test sets, was then evaluated a-posteriori by considering four different efficiency metrics widely used in applications. The first selected metric is the NSE index, introduced in Eq. (2). The second metric is the classic Root Mean Square Error $RMSE$:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [y_{pred,i} - y_{meas,i}]^2} \quad (3)$$

Since $RMSE$, as NSE , is insensitive to the bias between measured and modeled values, we also considered the following two additional metrics. The percent bias ($PBIAS$):

$$PBIAS = 100 \frac{\sum_{i=1}^N (y_{pred,i} - y_{meas,i})}{\sum_{i=1}^N y_{meas,i}}; \quad (4)$$

and the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009):

$$KGE = 1 - ED, \quad (5)$$

$$ED = \sqrt{(r - 1)^2 + \left(\frac{\sigma_{pred}}{\sigma_{meas}} - 1\right)^2 + \left(\frac{\mu_{pred}}{\mu_{meas}} - 1\right)^2};$$

where r is the linear correlation coefficient between predictions and observations. According to Eq. (5), KGE performs a trade-off between the following optimal conditions: modeled and observed quantities perfectly correlated ($r = 1$), with the same variance ($\sigma_{pred}/\sigma_{meas} = 1$) and model output not biased ($\mu_{pred} = \mu_{meas}$).

3.2. Data preprocessing and correlation analysis

A salient characteristic of ML is the combined use of different data types, each one with its range of variability. To avoid bias and other undesired effects, outliers, here identified as values exceeding the 99-th percentile of the empirical probability distribution, are first removed, and then the

resulting data set is transformed in the attempt to make it as close as possible to the Gaussian distribution, while normalization serves to remove dimensional effects and make all variables comparable.

In the present work, the variable of interest x is mapped into the transformed target variable $y = f(x)$, with the function f being one of the following three simple functions: the logarithm $f(x) = \ln(x)$, the square root $f(x) = \sqrt{x}$ and the cubic root $f(x) = \sqrt[3]{x}$. The logarithmic transformation was applied to As and DO, the squared root transformation to SO_4 , Ec, pop and agr, and finally, the cubic root transformation was applied to Cl, Qm and art. The remaining variables, Ta, Tw, JD, Y, coord, and Z were not transformed. These preliminary steps are key to a good performance of the models, in particular for LR and RF, which accuracy deteriorates when trained with skewed data (Raschka, 2015; Chollet, 2018). The transformed data are then standardized by subtracting the mean μ_y and dividing by the standard deviation σ_y :

$$y_{stand} = \frac{y - \mu_y}{\sigma_y} \quad (6)$$

with y_{stand} being the standardized variable. After transformation and standardization, features with different ranges of variability in their original form are equally influencing the efficiency criteria presented above.

Prior to applying the ML algorithms, we performed a preliminary analysis of the correlations between identified features and variables. The analysis was conducted by computing the Spearman's non-parametric correlation coefficient r_s (see Section 4.1 for the results), which is defined as follows (Hipel and McLeod, 1994):

$$r_s = 1 - \frac{6 \sum_{i=1}^n [R(X_i) - R(Y_i)]^2}{n(n^2 - 1)} \quad (7)$$

where n is the sample size and $R(X_i)$, and $R(Y_i)$ indicate the positions in the ranking of the i -th value of the analyzed X and Y variables, respectively, assumed as untied. This coefficient provides a primary indication of the influence of selected features on the observed water quality parameters. The value of r_s varies between -1 and 1 , with the former indicating variables perfectly anticorrelated and the latter variables perfectly correlated.

4. Results

4.1. Exploratory correlation analysis

As a preliminary step, the rank correlation analysis between all the variables was performed, which results are shown in Fig. 2. The grey rectangle evidences the portion of the matrix where drivers (x-axis) and water quality variables (y-axis) are compared.

All the correlation coefficients were statistically significant with p values lower than 0.05, except for agr and JD with As, coord with SO_4 , Y with Tw, and art and Z with DO. The rank correlation matrix shown in Fig. 2 evidences that arsenic (As) is not correlated to most of the factors, except for artificial land use (art, $r_s = -0.31$), population density (pop, $r_s = -0.13$) and measurement year (Y, $r_s = -0.19$). All these correlations are weak and the latter suggests a tendency of As concentration to reduce with time. No seasonality is observed in the signal, given the very low correlation ($r_s = -0.04$) with the Julian day (JD). A weak correlation is shown with the elevation (Z, $r_s = 0.10$) and a slightly stronger one with the spatial coordinates (coord, $r_s = 0.29$), suggesting a spatially non uniform release of arsenic. SO_4 is weakly correlated with the air temperature (Ta, $r_s = -0.24$) and the fraction of agricultural area (agr, $r_s = 0.14$), while the correlation with As is moderate ($r_s = 0.48$). Similarly, a moderate correlation is observed also with the electrical conductivity (Ec, $r_s = 0.43$).

Chloride (Cl) is positively correlated with land use features: agr ($r_s = 0.45$) and art ($r_s = 0.59$) and with the anthropogenic pressure (pop, $r_s = 0.51$), while it is negatively correlated with the spatial coordinates (coord, $r_s = -0.36$), elevation (Z, $r_s = -0.45$) and Julian day (JD, $r_s =$

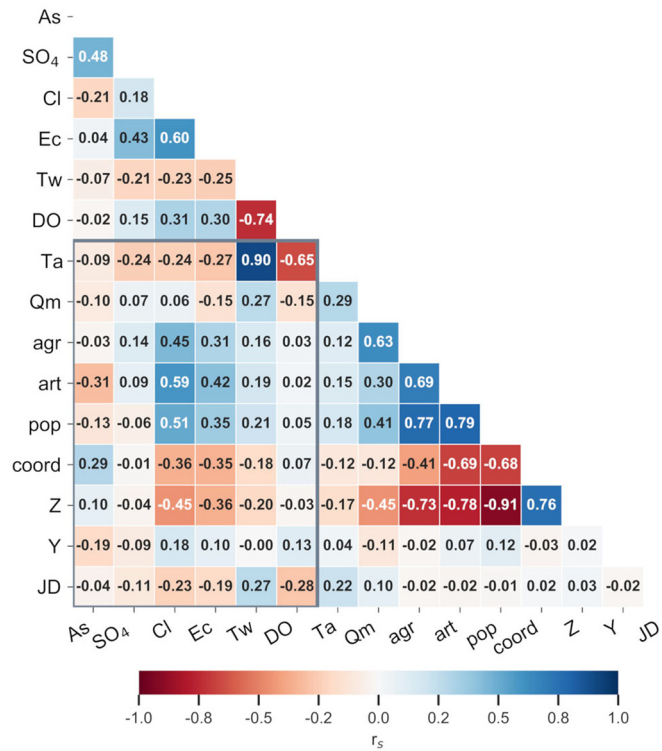


Fig. 2. Correlation Matrix between variables and drivers as introduced in Section 3.2 (As = arsenic concentration, SO₄ = sulfate concentration, Cl = chloride concentration, Ec = electrical conductivity, Tw = water temperature, DO = dissolved oxygen concentration, Ta = air temperature, Qm = monthly averaged water discharge, agr = percentage of agricultural land use, art = percentage of artificial land, pop = population density, coord = coordinates of the station, Z = elevation of the station, Y = year of the measurement, JD = Julian day of the measurement). Warm and cold colors indicate negative and positive correlations, respectively, and darker colors indicate stronger correlations.

– 0.23), though less strongly. Furthermore, Cl shows the highest correlation with Ec ($r_s = 0.60$) and is positively correlated with DO ($r_s = 0.31$) as well.

Importantly, Tw showed to be highly correlated with Ta, which is promordial to the satisfying performance of the Linear Regression model for this variable, as discussed later. The high correlation of DO with Tw is expected from physical principles as it is the slightly smaller correlation of DO with Ta. This is the consequence of the high correlation between Ta and Tw. It is indeed well known that DO concentration reduces as Tw increases (see, e.g., Stumm and Morgan, 1996). In addition, Tw is weakly correlated with the Julian day. The high negative correlation between population density and elevation ($r_s = -0.91$) complicates the attribution of changes in quality parameters to geology rather than anthropogenic effects, similar to the correlation between agricultural land use and population ($r_s = 0.77$). On the other hand, these correlations help in increasing the identifiability of the relevant drivers, as will be clarified in Section 4.3.

The negative correlations of SO₄, Cl, Ec, and DO with Ta are in line with a similar analysis conducted by Diamantini et al. (2018). Notice that in Diamantini et al. (2018) the rank correlation was computed separately at each sampling location, while in the present work, it was computed for the entire merged time series.

These observations suggest that Tw and DO should be adequately predicted by classical LR models, with the features showing the highest correlations. Instead, chemical water quality variables are only weakly correlated or not, with the drivers, thereby anticipating that the LR model may not be suitable for these variables. However, the lack of evident

correlations does not exclude the possibility of significant non-monotonic and mutual correlations that Spearman's correlation analysis is unable to detect, while ML may actually succeed, as will be highlighted in the following Section 4.2.

4.2. ML techniques performance

A first application of the ML models described in the Appendix A was performed by considering the entire set of drivers as input features. This was done to evaluate in a comparative way the prediction performances of the three models for each investigated water quality variable.

Following Raschka (2015), the k -fold cross-validation was used for hyper-parameters tuning. As described in Section 3.1 the hyper-parameters were chosen such as to maximize the mean NSE of the $k = 30$ folds while keeping low the standard deviation. The mean values of NSE and standard deviations are displayed in Table 2 for all the three models. The small standard deviations are indicative of a small variability between the NSE values of the folds, thereby resulting in a limited parametric uncertainty, i.e., in a small variability of the modeling output among the different parameterizations obtained with the folds. Similarly, the choice of the hyper-parameters that produced the highest mean NSE value minimized the epistemic uncertainty through the identification of the structure of the algorithm that better exploits the information contained in the data. The hyper-parameters that produced the statistics shown in Table 2 were then used in a new training extended to the entire training set (without folding) with the model parameters optimized by minimizing MSE (Eq. (1)).

The values of the efficiency metrics (i.e. NSE, RMSE, PBIAS and KGE) introduced in Section 3.1 are reported in Table 3 for both the training and test sets. A visual comparison of NSE values is offered as an example in Fig. 3. For all the variables, the NSE of the test set is comparable with the corresponding optimized value of the training set (in some cases, also indicating better performance). A similar behavior is also observed for all the other efficiency metrics, shown in Table 3 and Figs. S5, S6, and S7 of the Supplementary Material. Consequently, to facilitate the presentation of the results, in the following, performances will be discussed by referring only to the metrics obtained with the test set.

Inspection of Fig. 3 reveals that DNN provided the best performances with NSE values ranging from 0.6 for DO to 0.89 for Tw. The gain in using DNN is remarkable for most variables, except DO and Tw, for which the improvement with respect to LR and RF is minor. Focusing on the DNN model, we notice that As, SO₄, Cl, and Ec show similar performances, with NSE ranging from 0.68 for Cl to 0.78 for As. Tw shows the best performance with all the models, though DNN is characterized by the highest NSE of 0.89 with LR and RF reaching the values of 0.84 and 0.83, respectively.

KGE and PBIAS coefficients prove that simulations performed with DNN are characterized by the smallest biases for all the variables, except for Tw, which shows the smallest bias with RF, but that obtained with DNN is however small (PBIAS = 0.08% for RF and 0.47% for DNN). In more detail, simulations performed with DNN show negligible biases for Ec and DO, which increase for Tw, As, Cl, and SO₄, remaining small in absolute terms. The largest bias is observed for Cl (PBIAS = 5.64%), which is however smaller than those obtained with RF (PBIAS = 8.35%) and LR (PBIAS

Table 2

Mean values \pm standard deviation of NSE coefficients obtained in the cross-validation phase, computed over $k = 30$ folds of the training set, the latter composed of 70 % of the available data.

	LR	RF	DNN
As	0.40 \pm 0.022	0.69 \pm 0.014	0.84 \pm 0.008
SO ₄	0.23 \pm 0.020	0.54 \pm 0.030	0.76 \pm 0.013
Cl	0.53 \pm 0.012	0.61 \pm 0.013	0.78 \pm 0.012
Ec	0.46 \pm 0.013	0.56 \pm 0.022	0.80 \pm 0.008
Tw	0.84 \pm 0.003	0.83 \pm 0.003	0.89 \pm 0.003
DO	0.54 \pm 0.004	0.53 \pm 0.005	0.63 \pm 0.009

Table 3

PBIAS, RMSE, NSE and KGE values for training and test sets for each water quality variable and the three models: LR, RF and DNN. The RMSE values are reported in the unit of measurement of the corresponding variable (mg/L for As, SO₄, Cl and DO, $\mu\text{S/cm}$ for Ec, and °C for Tw).

	NSE						KGE					
	Training set			Test set			Training set			Test set		
	LR	RF	DNN	LR	RF	DNN	LR	RF	DNN	LR	RF	DNN
As	0.29	0.51	0.76	0.33	0.56	0.78	0.35	0.52	0.77	0.38	0.56	0.80
SO ₄	0.20	0.54	0.74	0.20	0.58	0.75	0.23	0.57	0.77	0.24	0.61	0.77
Cl	0.44	0.57	0.76	0.48	0.51	0.68	0.57	0.61	0.79	0.53	0.57	0.73
Ec	0.44	0.54	0.80	0.42	0.47	0.76	0.55	0.53	0.83	0.56	0.48	0.83
Tw	0.84	0.83	0.90	0.84	0.83	0.89	0.88	0.85	0.92	0.87	0.85	0.91
DO	0.54	0.53	0.62	0.54	0.51	0.60	0.62	0.57	0.69	0.61	0.55	0.66

	PBIAS [%]						RMSE					
	Training set			Test set			Training set			Test set		
	LR	RF	DNN	LR	RF	DNN	LR	RF	DNN	LR	RF	DNN
As	15.30	10.69	5.67	12.34	5.13	2.93	0.0017	0.0014	0.0010	0.0016	0.0013	0.0009
SO ₄	5.76	3.94	2.39	5.60	2.29	1.97	15.83	12.02	8.99	15.20	11.00	8.49
Cl	6.16	6.61	3.29	7.38	8.35	5.64	2.00	1.75	1.30	1.89	1.84	1.49
Ec	1.23	1.55	0.56	0.67	1.72	0.01	48.25	43.50	29.14	51.06	48.49	33.08
Tw	0.00	0.10	−0.01	−0.23	0.08	0.47	1.67	1.73	1.35	1.67	1.73	1.35
DO	0.26	0.31	0.23	0.07	0.13	−0.01	0.84	0.85	0.76	0.84	0.87	0.78

= 7.38%). The good performance of DNN in terms of both RMSE (Eq. (3)) and bias (Eq. (4)) is summarized by the high values of the KGE (Eq. (5)) in both the training and the test sets for all the investigated variables.

Finally, it can be observed, by comparing the rank correlation matrix of Fig. 2 with the performance indices of Table 3, that variables showing the highest r_s coefficients, namely Tw ($r_s = 0.90$ with Ta) and DO ($r_s = -0.65$ with Ta), are also those with the smallest reduction in performance when LR is used instead of DNN. The other variables are characterized by much lower absolute values of r_s and consistently their LR performances are much poorer than those obtained by using DNN for all the metrics: the largest loss of performance of LR with respect to DNN is observed for As and SO₄, which are the variables with the smallest absolute values of r_s (Fig. 2). Nonlinearity in ReLU transformation and the flexibility in exploring connections between the variables are the ingredients that allowed DNN to learn the complex relationships between drivers and water quality variables that the LR model cannot resolve.

The metrics shown in Table 3 lead to the conclusion that the model providing the best performances for all the variables and efficiency indexes is

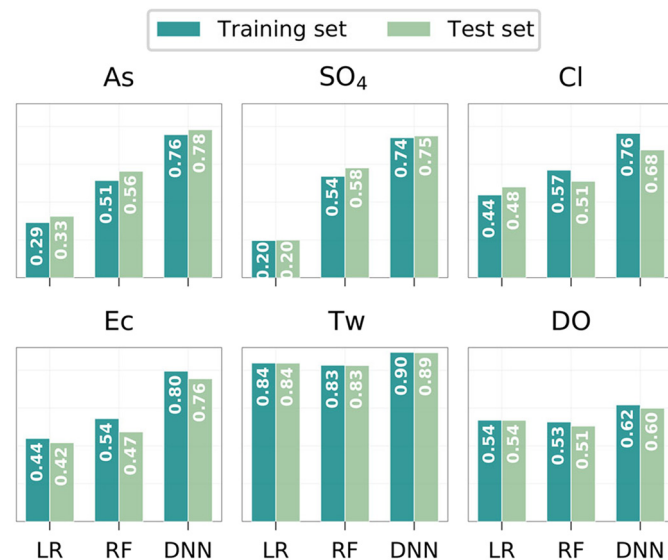


Fig. 3. NSE coefficients of the six water quality variables for the Linear Regression (LR), Random Forest (RF) and Deep feed-forward Neural Network (DNN) models. Dark and light bars refer to training and test sets, respectively.

DNN. Simulated variables of the test set with DNN are then compared with the measurements in Figs. 4 and 5.

In the left panels of Fig. 4 the simulations by DNN (continuous blue line) are compared with the observations (orange bullets). The time series was composed by sequentially merging the test set time series of all available sampling sites, as detailed in Section 3.1. The overall behavior is well captured by DNN for all the water quality variables, even though a tendency to smooth the variability emerges from visual inspection of the left panels. However, the bias is slight to negligible, as demonstrated by the low values of the PBIAS metric reported in Table 3, and the small to insignificant asymmetry around the 1:1 line in the scatterplots of measured versus predicted values as shown on the right panels of Fig. 4. The smoothing was somewhat expected, given that DNN is regression-based like the other models. Still, the improvement in terms of both bias and prediction errors compared to the classic LR method is remarkable, as clearly shown in Table 3.

The tendency of DNN of smoothing the time series is better evidenced in Fig. 5 where the measured data are presented in ascending order (ranked) with the simulated values reordered such as to maintain the temporal correspondence. This arrangement of the data allows appreciating the model's accuracy, and it better evidences the tendency to underestimate high concentrations and overestimate low ones, with DO being the variables more affected by smoothing.

Notice in Fig. 5 (second panel from the top in the left column) the staircase behavior of Cl measurements for small to medium concentrations. This somewhat non-physical behavior occurs because Cl measurements in the northern portion of the watershed are available at a resolution of 1 mg/L, which partially justifies the lower NSE of Cl with respect to the other variables. A similar behavior is observed for As at intermediate concentrations. Specifically, the measurement resolution for As varies between the two Provinces. The Autonomous Provinces of Trento and Bolzano collected the measurements with a resolution of 0.0001 mg/L and 0.0005 mg/L, respectively. This leads to a staircase behavior of As above 0.002 mg/L, which is indeed the minimum value detected in Bolzano. The other variables, except DO, are reproduced by DNN with a less pronounced smoothing and smaller oscillations of the predicted values around the measured ones.

4.3. Importance features assessment

ML is particularly flexible in dealing with a multiplicity of drivers, but on the other hand, more drivers also mean a larger risk of overfitting. Under the likely hypothesis that not all the drivers have a comparable influence on the modeled variables, the existence of an optimal selection

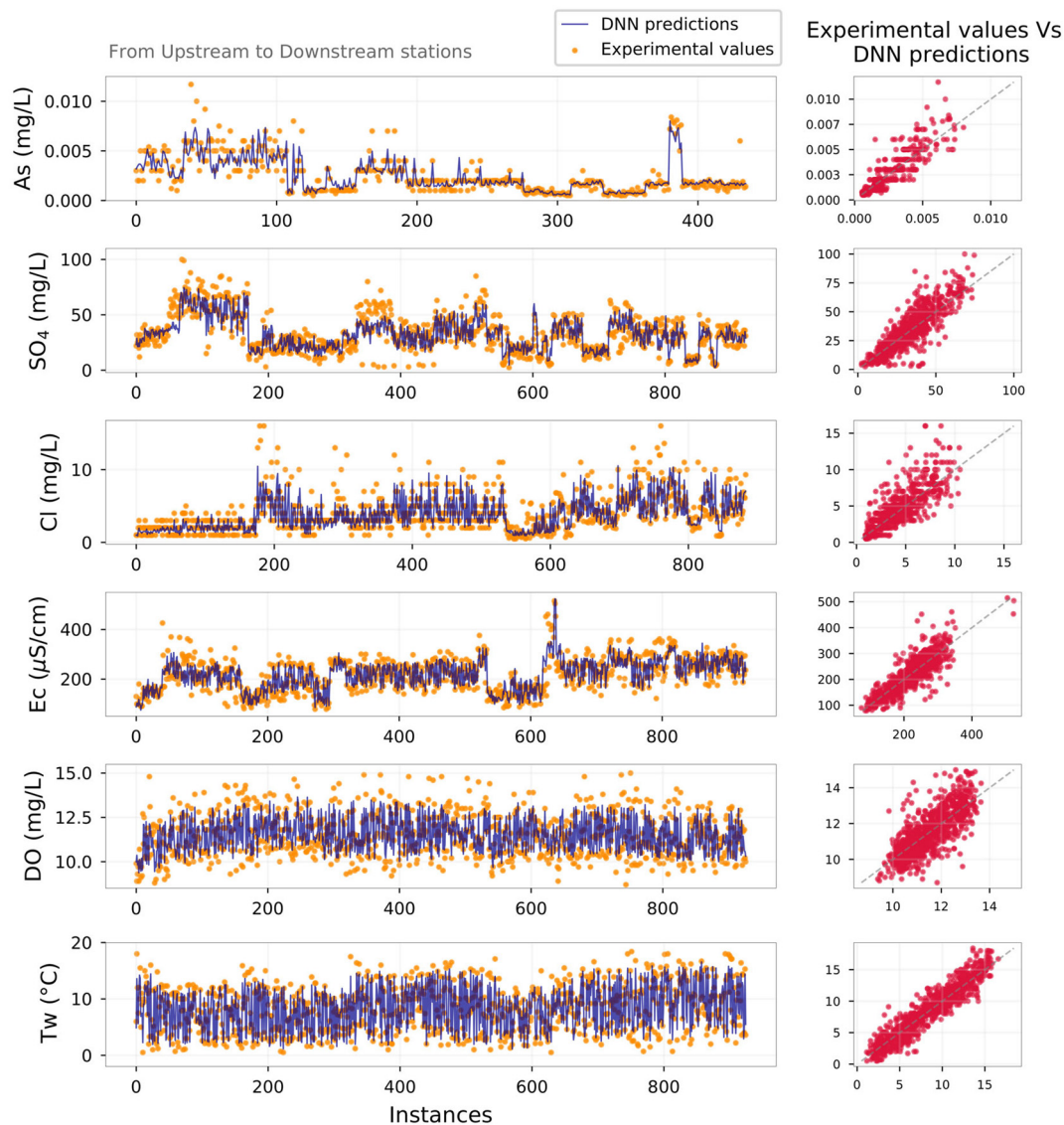


Fig. 4. Comparison between experimental data and DNN predictions for the six water quality variables and with reference to the test set. The left panels show the sequence of instances obtained by merging the time series of all the sampling sites, and the right panels show scatterplots of observed versus predicted values, with the dashed lines indicating the perfect match.

of drivers is arguable. Under this perspective, in this section, we analyze the relative importance of the features identified as potential drivers of change in water quality variables to identify the most influential ones. We started with a first model using the spatial information (“Sp”) solely, i.e. coordinates (coord) and elevation (Z) as drivers. A spatio-temporal (“SpT”) model was then developed by adding as features the Julian day (JD) and the year of the samplings (Y). Successively, the monthly water discharge Q_m , which epitomizes dilution, was added (“SpT + Q_m ”), and subsequently also the air temperature T_a (“SpT + Q_m + T_a ”). In the last stage, population density (pop) and the percentages of agricultural and artificial land use, agr and art, were added, leading to a configuration employing the complete set of drivers (“Full”), which resembles the structure discussed in Section 4.2.

The three models, LR, RF, and DNN, were then applied with reference to the aforementioned five configurations by using the same hyper-parameters set-up used for the Full configuration. Given the different number of input features, all the models were retrained on the training set for each investigated configuration and the models parameters optimized. As before, reliability was then evaluated with reference to the test set.

The NSE obtained in both training and test sets are shown for all configurations and models in Fig. 6, with the box at the bottom indicating the combination of drivers adopted in each case. NSE indexes for each model and investigated configuration are similar between training and test sets, thus confirming the reliability of previously presented hyper-parameters tuning. As expected, the highest values of NSE are obtained for each water quality variable by considering all the drivers (“Full” configuration), which coincide with those reported in Table 3 and displayed in Fig. 4.

In general, RF and DNN performed better than LR for all driver configurations. The difference between RF and DNN is small for Tw and DO but appreciable for all the other water quality variables. As expected, for LR, a significant performance improvement occurred with the inclusion of the drivers showing the highest rank correlation with the variable of interest. For As, Cl, and Ec, these are pop, agr, and art. Differently, for Tw and DO , a significant improvement was obtained when T_a was added to the drivers.

The highest difference in the performance of RF and DNN with respect to LR, occurred when only the spatio-temporal information (“SpT” configuration) was included. For both models, NSE increases with the number of drivers, but the way this improvement occurs depend on the variable.

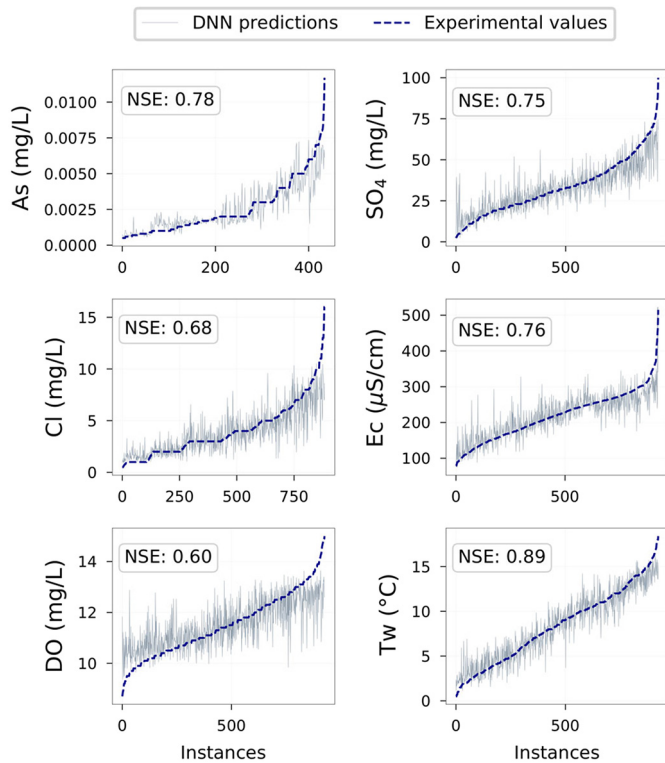


Fig. 5. Comparison between measured variables and DNN simulations, for each investigated water quality variable and with reference to the test set. The measured values are depicted in ascending order (dashed cyan line), while the grey line shows the DNN simulations of the corresponding instance. Notice that the abscissa of all the panels represents the position assigned to the observed values from the smallest to the largest one in the ordered series of the test set. The *NSE* coefficients between measured and modeled values are also displayed inside the panels.

Both *Tw* and *DO* showed a remarkable improvement when temporal information was added to the spatial one (“SpT” configuration), with only a small further gain obtained when the other drivers were added. Indeed, *NSE* increased from 0.81 to 0.89 and from 0.60 to 0.61 for *Tw* and *DO*, respectively, when *Ta* was added to the drivers used in the DNN model.

Figs. 7a and b show the Quantile-Quantile (Q-Q) plots of *DO* and *Ta*, respectively, with the quantile of the fitted normal distribution shown in the abscissa and that of the sample in the ordinate. Three samples were considered: the measured values (grey triangles) and the modeled values obtained by using DNN with the spatial information “SpT” (red dots) and the spatial information supplemented by the air temperature “Sp + Ta” (blue squares), respectively. Transformed and normalized values are shown in the figures. These plots allow checking separately the correspondence of the models with the measurements and the departure from normality of the transformed data. The objective here is to compare the performance of the “SpT” model with a new model obtained by replacing the time information with air temperature. We indicate this new alternative model with “Sp + Ta”. The inclusion of air temperature in place of time information (i.e., *JD* and *Y*), causes only a slight modification of the model's accuracy, as shown by the small difference between the red and blue quantile curves of Figs. 7a and b. Inspection of these figures reveals that temporal information (the Julian day and the year of the measurements) is capable to predict *DO* and *Tw*, at comparable accuracy to air temperature. This is also confirmed by the small differences between the *NSE* values of the test set for *Tw* and *DO* under these two configurations. Indeed, “Sp + Ta” configuration leads to *NSE* coefficients of 0.84 and 0.50 for *Tw* and *DO* respectively, with the corresponding *NSE* indexes for the “SpT” configuration being 0.81 and 0.56. These small differences allow concluding that the time of

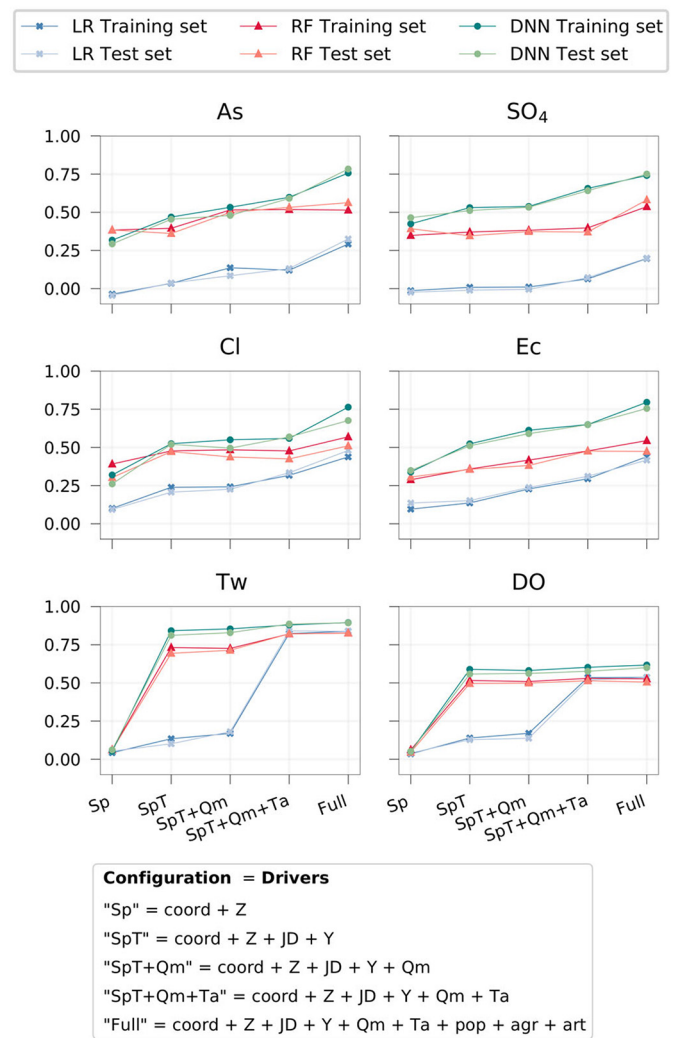


Fig. 6. *NSE* values for the three models and for all the six water quality parameters, evaluated with reference to different configurations of drivers. The various combinations of drivers employed for each configuration are detailed in the box at the bottom.

sampling (*JD* + *Y*) may be used in this case as a surrogate of air temperature, which, however, given the higher value of *NSE*, is a slightly better descriptor of the observed variability.

5. Discussion

The analysis presented in Section 4.3 identified the drivers exerting the major influence on the selected water quality variables and the drivers' combination that explains most of their variability. This assessment is important because reducing the number of parameters renders the model less prone to overfitting (see, e.g., Liu et al., 2022). An alternative approach, often used in Artificial Intelligence applications, is to transform the drivers through Principal Component Analysis (PCA), a powerful complexity reduction technique. However, PCA transforms the input signals by combining them such that the identification of the most influential drivers is no longer possible. This is a drawback, since the identification of the most influential drivers is key for the hints it provides in developing an effective strategy for improving river water quality.

For all the six water quality variables considered in the present work, DNN detected the nonlinear relationships between drivers and variables of interest better than RF, though both performed better than LR when a limited number of drivers were used. A peculiarity that differentiates the present work from previous contributions is the use of the same model for

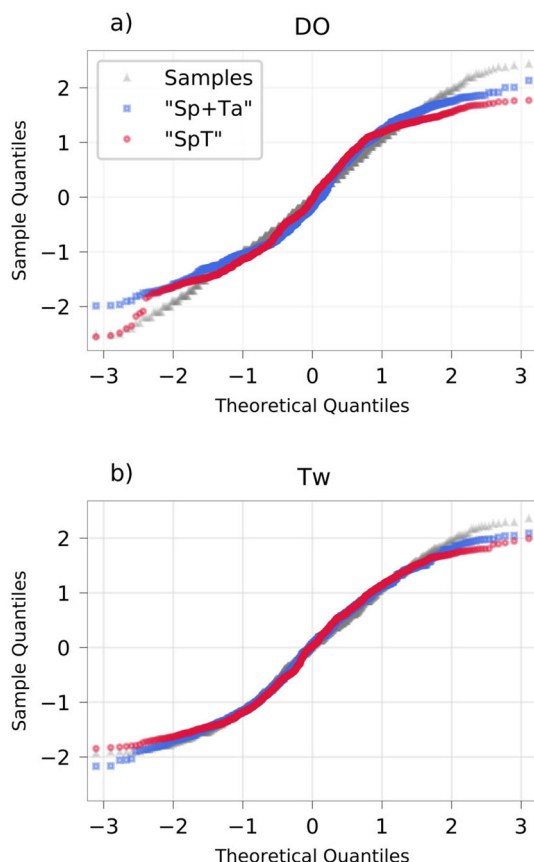


Fig. 7. Quantile-Quantile plots of the samples and the fitted normal distribution for a) the dissolved Oxygen (DO) and b) the water temperature (Tw) in their transformed and normalized form. The three curves refer to the sample of measured test set values (grey triangles) and the corresponding samples obtained by the “SpT” (red dots) and the “Sp + Ta” (blue squares) DNN models. The normal distribution, used as a reference, was fitted to the three data sets by the Maximum Likelihood method.

all the sampling sites. In doing that, we obtained a regional model, which can be applied at unsampled locations within the same region, owing to the inclusion of spatial coordinates and space-dependent variables as drivers. The *NSE* values we obtained with the six regional models are in line with those of existing studies that utilized site specific models (see, e.g., Anmala and Venkateshwarlu, 2019; Chen and Liu, 2014; Shah et al., 2021; Green et al., 2021; Shamsirband et al., 2019), though some of them resulted in a slightly poorer fit, i.e. smaller *NSE* values (see, e.g., Nemati et al., 2015; Al-Mukhtar and Al-Yaseen, 2019; Zhu et al., 2019). In general, our models were successful in capturing the spatial and temporal dynamics of the six variables and these payoffs, together with the possibility to perform simulations at unsampled locations, the slight reduction of *NSE* coefficients resulting from using the same regional models for all the sampling sites.

Figs. 6 and 7 show that when either RF or DNN was chosen, the most influential driver of water temperature is the air temperature, but that the date of sampling (Y plus JD), surrogates for this information. This is not the case for LR, given that the model “SpT” (spatial plus temporal information) results in a much poorer reproduction of the water quality variables than with RF and DNN. Similar considerations can be drawn for Dissolved Oxygen. This evidences that seasonality and air temperature carry similar information, but that only RF and DNN are able to distill it from the date at which the sample is taken. The dominance of seasonality on water temperature is the consequence of the buffering effect of the water against changes in air temperature due to a specific thermal capacity four times

higher than that of the air. However, the slight improvement of the *NSE* values when air temperature is added in both the RF and DNN models (Fig. 6) indicates that variations of air temperature at the daily time scale have an influence on water temperature, but much smaller than the effect of changes at the seasonal scale. The nonlinear dependence between water temperature and seasonality is in line with results recently proposed by Toffolon and Piccolroaz (2015) and Piccolroaz et al. (2016).

As shown in Fig. 6 and discussed in Section 4.3 both RF and DNN perform better than LR in modeling the other four water quality variables (As, SO₄, Cl and Ec) for all the driver combinations. When only the spatial information (“Sp”) is employed as driver, moderate values of *NSE* are obtained, and this highlights a fundamental difference with Tw and DO, which are characterized by *NSE* values close to zero, and thereby are insensitive to the spatial information. The influence of spatial coordinates hinges on the lithological characteristics of the drainage area and possibly on the distribution of the human activities, both potential sources of the above dissolved species, which jointly affect Ec. Indeed, As, SO₄, and Cl are all present in igneous and sedimentary rocks, to a different extent depending on the lithology (see, e.g., Freeze and Cherry, 1979, ch. 3), and are also dissolved in both surface runoff and treated and untreated waste waters from urbanized areas (Bhattacharya et al., 2007). In this respect, our results contribute in closing the research gap evidenced in the recent review by Tiyyasha et al. (2020), which highlighted the limited use of ML algorithms to investigate sulfate or chloride dynamics in riverine environments (only 4 out of the 209 investigated papers adopted a ML algorithm). Most importantly, none of the reviewed studies employed ML techniques to simulate arsenic concentration in rivers.

Typical concentrations of SO₄ in freshwaters, due to the cumulative contribution of rocks and soils weathering as well as precipitations and atmospheric depositions, are in the range of 2 ÷ 80 mg/l (see, e.g., Chapman, 1992, chs. 3 and 6). Marine aerosols also deposit Cl, causing concentrations usually lower than 10 mg/L. In the Adige river, the concentrations of SO₄ were always in the aforementioned range, except for a site in the northwestern portion of the basin, where a concentration of 105 mg/l was observed in summer (see Fig. S2 in Supplementary Material). On the other hand, the concentrations of As were not exceeding 0.008 mg/L, and the sites reporting the highest values are located in the northwest portion of the basin and along the main stem of the Adige river, close to the city of Bolzano (see Figs. S1, S2, S3, and S4 in the Supplementary Material). These concentrations are compatible with the hypothesis of a dominating geogenic origin of As, SO₄ and Cl, provided that lithology is also compatible. To verify this we performed a careful inspection of local grey literature and geological maps of the area (https://www.isprambiente.gov.it/Media/carg/note_illustrative/60_Trento.pdf). The analysis revealed that lithologies containing As and SO₄ are widespread in the Adige catchment but limited in extension. For example, isolated deposits of gypsum have been identified in the Avisio valley, one of the main tributary of the Adige river (Antolini, 1984), and in the northern part of the river basin (Perna, 2003), and pyrite was identified in several locations within the catchment (see, e.g., Maiello, 2006; Cozzi, 2019). The observed concentrations are thereby consistent with the hypothesis of a natural or geogenic origin, though additional contributions by anthropogenic sources cannot be excluded, owing to the relatively high population density of the region. Differentiating geogenic and anthropogenic sources of SO₄ is possible through isotopic analysis (Cortecci et al., 2002), but unfortunately, this information was not available in the Adige river basin.

Finally, the concentration of Cl was always below 12 mg/l, with most of the sites reporting concentrations smaller than 10 mg/l and showing a positive gradient from North to South in all seasons (see Figs. S1, S2, S3, and S4 in the Supplementary Material), thereby suggesting atmospheric deposition as the main source of Cl in the Adige waters.

The combined analysis of river basin lithology and concentration levels identified a likely geogenic origin of As, SO₄ and Cl, but at the same time, it cannot exclude an anthropogenic influence. In fact, when the information on the population and land use was added, moving from the fourth to the fifth configuration level (see Fig. 6), an

appreciable increase of the *NSE* value was observed for all the variables, except for *Tw* and *DO*, as discussed above. This reveals an anthropogenic signature in the spatial and temporal concentration distributions of *As*, *SO₄*, and *Cl*. In other words, the analysis performed by separating the drivers revealed that the monitored dissolved species are of geogenic origin, but that they are also influenced by the human activities, of which the population density and land use are a proxy. In this respect, both *RF* and *DNN* behave similarly, except for *As*, for which the improvement is observed only when *DNN* is used. On the other hand, *LR* is characterized by much smaller *NSE* values, with a reduction of efficiency larger for *As* and *SO₄*, than for the other variables.

In the following we focus on the role of the single drivers in the observed dynamics of *SO₄* and *As*. Both *RF* and *DNN* models of *SO₄*, obtained by including only the spatial information (“*Sp*” model), produced *NSE* values close to 0.5. The progressive inclusion of the temporal information, streamflow, and air temperature did not provide any noticeable benefit for *RF*, while for *DNN* the values *NSE* increased from 0.46 to 0.51, when temporal information was added, remained constant after the inclusion of stream flow (no benefit), and increased to *NSE* = 0.64 after the addition of air temperature. A dependence of *SO₄* concentration on air temperature was observed by Todd et al. (2012) and Crawford et al. (2019) for the Alpine environment. They attributed this dependence to the more rapid weathering rate of rocks and soils at the higher temperatures caused by global warming. This is confirmed by the insets of Fig. 8 where the aggregated temporal trends of *SO₄* and *Ta*, characteristics of the main sub-catchments, show an almost parallel

behavior, particularly in the northwest and south of the basin and in the Isarco valley. The negative Spearman correlation between *SO₄* and *Ta*, shown in Fig. 2, does not invalidate this interpretation because it was performed by considering the daily air temperature time series, which oscillate wildly, while weathering processes are expected to be sensitive to cumulate annual or seasonal temperatures. As already commented above, the inclusion of anthropogenic drivers (“*Full*” model) leads to a noticeable increase of *NSE* for both *DNN* and *RF*.

The role of the drivers in shaping *As* concentrations is more balanced. The predominant role of the spatial information remains, though with a *NSE* smaller than for *SO₄* (*NSE* ≈ 0.35). However, differently from *SO₄* the impact of stream flow (*Qm*) is evident (when *DNN* is used, less for *RF*), suggesting that dilution plays a role in controlling *As* concentration. The weakly negative Spearman correlation between *Qm* and *As* shown in Fig. 2 is compatible with this interpretation. The noticeable increase of *NSE* when population and land use information were added to the *DNN* model suggests the presence of anthropogenic sources. However, no improvements were observed when the *RF* model was used, and local grey literature identified cases in which *As* is of geogenic origin, but without excluding a possible anthropogenic contribution (see, e.g., Fuganti et al., 2005). These somewhat divergent conclusions with respect to our analysis can be due to the inherent local scale at which this study was conducted.

Cl shows a behavior similar to that of *As* with land use, population density, and time of sampling exerting the larger effects on its concentration. The noticeable increase of *NSE* observed in Fig. 6 when the temporal information is added (“*SpT*” model) is in line with the hypothesis that a large

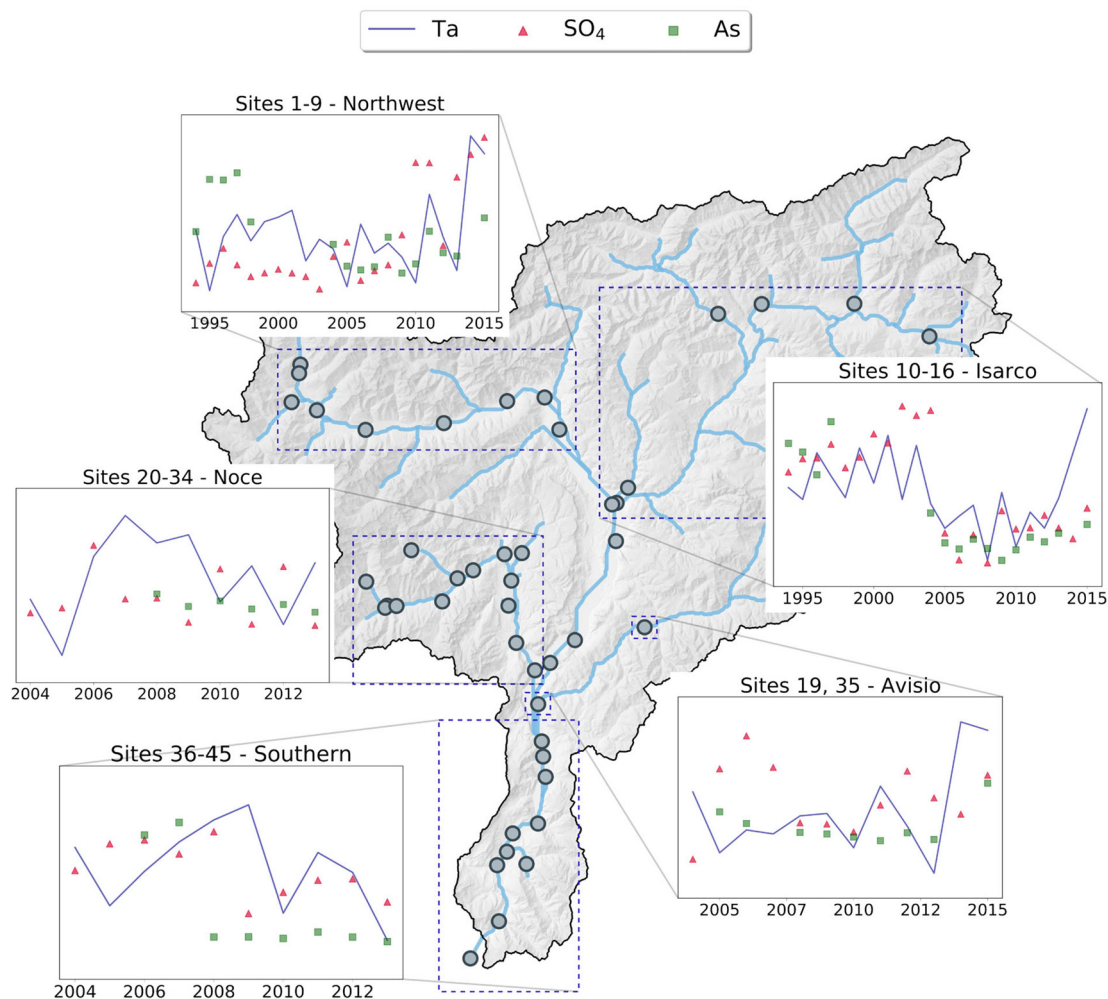


Fig. 8. Temporal trends of *SO₄*, *As*, and *Ta* averaged across specific zones of the Adige River Basin.

fraction of Cl has marine origin through the deposition with the precipitations (Gabrielli et al., 2006). The tendency of Cl concentrations to increase along the main stem of the Adige river (see Figs. S1, S2, S3, and S4 in the Supplementary Material) is consistent with the North-South trend for the precipitations in the region as reported by Mallucci et al. (2019) and Borsato et al. (2007). Furthermore, the similar increase of *NSE* occurring when the information on population and land use was added suggests a significant anthropogenic influence, particularly along the main stem of the Adige river, where the larger concentrations of Cl have been detected and population density is the highest of the river basin.

6. Conclusions

We compared two ML algorithms, namely Random Forest and Deep feed-forward Neural Network with Linear Regression, assessing their performance in modeling As, SO₄, and Cl concentrations and a few physical parameters: Ec, Tw and DO in a number of sampling locations within the Adige River catchment. The six water quality variables were modeled independently by using for each of them the same model structure, thereby creating six regional models and obtaining a common ground for the analysis of the relative importance of the drivers.

Random Forest and Deep feed-forward Neural Network performed better than the Linear Regression model for all the water quality parameters. The comparison between DNN and RF showed that the former is more flexible and effective in detecting non-linear relationships and therefore is preferable in modeling the investigated water quality variables. LR performed similarly to RF for both water temperature and dissolved oxygen when all the drivers are included but with the largest influence exerted by the air temperature. This is in line with the preliminary data analysis, which showed a relatively high rank correlation between air temperature and both water temperature and dissolved oxygen.

The learning process was repeated by progressively adding the drivers from the mere inclusion of the spatial coordinates and elevation, then supplemented by the time at which samples have been taken, which brings information on seasonality, the streamflow, the air temperature and finally population density and soil use. This analysis highlighted that for both Tw and DO, the most influential driver is air temperature, but that timing of the measurements (Year and Julian day) is a very effective surrogate when air temperature is not available. However, in this case, DNN should be used and the inclusion of air temperature produced a further, though slight, improvement. This leads to the conclusion that water temperature and variables strictly depending on it, such as DO, can be predicted with good accuracy by using only the information provided by the day of sampling (which reflects the seasonality of the signal, possibly because of the buffering effect of water), but that only DNN and, to a lesser extent, RF are able to distill this information, owing to their capability to unveil non-linear behaviors.

For the other water quality variables (As, SO₄, Cl, and Ec), RF and DNN performed remarkably better than LR at all the levels of information introduced by the drivers. For these variables, the influence of the drivers is more balanced than for Tw and DO. A clear geogenic origin is observed with anthropogenic disturbance that influences the spatial and temporal variability, given that the addition of population density and land use information prompted an increase of *NSE* for all these variables. The inclusion of isotopic data is expected to help in this, though they are rarely available to an amount sufficient to be effective in ML applications. As epitomized by streamflow, dilution has a low impact on these variables, except for As, probably due to its uneven release within the catchment.

The proposed analysis with different combinations of drivers can be applied to identify the external forcing exerting the most relevant impacts on the dynamics of water quality variables. However, disentangling their effect is complicated by the interdependence of the drivers. Nonetheless, ranking the drivers by their importance and deducing their synergistic effects is key to identifying policy actions aiming to reduce pollution according to the EU Water Framework Directive.

CRedit authorship contribution statement

Maria Grazia Zanoni: Software, Data curation, Visualization, Writing - Original Draft, Writing - Review & Editing.

Bruno Majone: Methodology, Conceptualization, Writing - Review & Editing, Supervision.

Alberto Bellin: Methodology, Conceptualization, Writing - Review & Editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially supported by the Italian Ministry of Education, Universities and Research (MIUR) under the Departments of Excellence, grant L.232/2016. Streamflow and meteorological data were provided by the Hydrological and Meteorological Offices of the Autonomous Provinces of Trento (<http://www.floods.it/public/index.php>) and Bolzano (http://www.provincia.bz.it/hydro/index_i.asp). Water quality data were provided by the Environmental Protection Agencies of the Autonomous Provinces of Trento (<http://www.appa.provincia.tn.it>) and Bolzano (<http://www.provincia.bz.it/agenzia-ambiente/>).

Appendix A. Machine Learning and Linear Regression modeling

A.1. Random Forest regression model

A Decision Tree (DT) algorithm automatically subdivides a dataset by employing decision rules on the target variable. The tree is grown through optimal splits, obtained by minimizing the *MSE* metric (Eq. (1)) between prediction and measured values (Breiman, 2001). The splits of the decision nodes are repeated until the tree reaches a given depth, or the leaves (i.e., the outcomes of the model) are pure (end node with a threshold value that includes just a single observation). The tree may also be pruned, removing sections or nodes with low explanatory power to reduce the model's complexity level (Breiman, 2001).

Random Forest (RF) regression is an ensemble of multiple DTs, applied to a bootstrap resampling (random sampling with replacement), usually with a size set to 2/3 of the training dataset (Breiman, 2001; Raschka, 2015). In each DT, the nodes are split with a randomized subset of the available features to maximize the accuracy. The prediction value of a given measurement obtained by each DT is saved, and then the RF prediction is computed as the unweighted average of the modeled values over all the DTs. In doing that, a stronger learner is built starting from the combination of weak learners (Breiman, 2001; Raschka, 2015). Indeed RF is less sensitive to outliers than the standard DT process and typically provides a better generalization performance given the randomness of the selection process of both features and bootstrap resamples (Raschka, 2015). Furthermore, RF is appealing because of its capability to exploit non-linear relationships among data, which LR cannot identify.

The most critical hyper-parameters for this algorithm are the number of decision trees, the maximum depth, and the minimum leaf sample, in the present work fixed as 18, 3, and 20, respectively. These values were set by repeating the *k*-fold analysis presented in Section 3.1 with the number of decision trees, maximum depth, and minimum samples for leaves in the ranges 1–50, 1–20, and 1–50, respectively. The selected parameters indicated above provided the best combination of *NSE* values in the training and validation sets, as described in Section 3.1. Given the number of maximum depth of layers fixed to 3, each DT can only use for classification a subset of 3 randomly selected features. The RF model was implemented by using the python package *scikit-learn* (Pedregosa et al., 2011).

A.2. Deep feed-forward Neural Network model

A typical ANN is composed of layers and nodes, the latter called neurons. The transfer of the information among the layers occurs at the level of the neurons, which transform a combination of input signals arriving from the upstream layer through a non-linear activation function, thereby operating sequentially (moving forward) from the input layer to the output layer (Chollet, 2018; Goodfellow et al., 2016). Hence the ANN acts as a dynamic filter, which distills the information progressively through the various layers and finally provides the prediction in the output layer. The structure of the network and the connections (e.g., activation functions, number of hidden layers, and neurons) need to be decided in advance by the modeler, representing the hyper-parameters. Instead, the neurons' weights are obtained by minimizing the MSE metric given by Eq. (1).

We implemented a Deep feed-forward Neural Network, which is an ANN with dense (fully) and unidirectional interconnection between the neurons, with more than one hidden layer. For the input layer, we considered nine neurons, each one receiving one of the nine drivers introduced in Section 2.2 (JD, Y, coord, Z, Ta, Qm, agr, art, pop). While the output layer is composed of a single node, representing the prediction of the target water quality variable. When less than nine drivers are used (see Section 4), the number of neurons reduces, and so the number of parameters, thereby reducing overfitting and increasing regularization.

Hyper-parameters were selected evaluating the loss score function during the training step and the NSE index in the k -fold cross-validation procedure. Specifically, for the activation function, we chose the Rectified Linear Unit function (ReLU) (Goodfellow et al., 2016; Glorot et al., 2011):

$$g(z) = \max(0, z) \quad (\text{A.1})$$

where z is the input signal processed by the activation function in each neuron, in other words, the sum of each unit of the precedent layer with the corresponding weight and the bias unit. Afterward, the choice of the number of hidden layers and neurons was performed according to Akaike's Information Criterion (AIC), as described in Appendix A.2.1, which penalizes network complexity through a regularization term. The application of the Akaike information criterion in ML learning was discussed in Zhao et al. (2008), among others. The optimal structure of the network, according to the Akaike information criterion, was composed of two hidden layers, each one formed by nine neurons, in addition to the input and output layers (see Fig. A.9). Finally, the weighting coefficients were obtained by minimizing the MSE (Eq. (1)) between predicted and observed water quality parameters by means of the Adam optimizer; an algorithm designed explicitly for high-dimensional parameter spaces problems (Kingma and Ba, 2014). Kingma and Ba (2014) showed that the Adam optimizer performed better than other stochastic regularization techniques. The DNN model was implemented with the deep-learning library Keras, developed in Python (Chollet, 2018).

A.2.1. AIC for DNN model selection

A sufficient number of neurons are required to learn from the data properly, but, on the other hand, exceeding their number leads to overfitting. AIC or other criteria (such as Bayesian Information Criterion, for example) allow to rationally choose among competing models with a different number of parameters (Zhao et al., 2008; Panchal et al., 2010; Hou et al., 2018). Since one of the main objectives of this work is to investigate the role of the drivers predicting the Adige water quality dataset, we opted for using a fixed DNN architecture for all the variables.

For the conventional least-squares regression with normally distributed errors and for a small sample size (where the ratio N_T/K is lower than 40), AIC is given by the following expression (Sugiura, 1978):

$$AIC = N_T \cdot \log \left(\frac{\sum_{i=1}^{N_T} [y_{pred,i} - y_{meas,i}]^2}{N_T} \right) + \frac{2KN_T}{N_T - (K + 1)} \quad (\text{A.2})$$

where N_T is the training sample size and K is the number of parameters used in the regression. The first right-hand side term of Eq. (A.2) is a measure of the goodness-of-fit of the model, while the second term penalizes the complexity of the models. For $N_T \gg K$, the second right hand term simplifies to $2K$.

As observed by Zhao et al. (2008) the strong dependence of AIC on the sample size may lead to a suboptimal choice of the model (see also, Burnham and Anderson, 1998):

$$\lambda_i = \frac{1/(\Delta_m)_i}{\sum_{j=1}^m 1/(\Delta_m)_j} \quad (\text{A.3})$$

where i indicates the i -th model of the ensemble S of models (with size equal to m) obtained by the combination of different numbers of hidden layers and neurons, and

$$(\Delta_m)_i = 1 + \frac{\Delta_i}{\Delta_{max}} \beta \quad (\text{A.4})$$

with $\Delta_i = AIC_i - AIC_{min}$ being the difference between the AIC coefficient of the i -th model and the minimum one, and Δ_{max} the Δ_i value for the worst DNN configuration. Finally $\beta = \sigma_{max} - \sigma_{min}$, where

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^N (y_{pred,j}^i - \mu_{pred}^i)^2}{N}} \quad (\text{A.5})$$

and $\sigma_{min} = \min_{i \in S} \sigma_i$ and $\sigma_{max} = \max_{i \in S} \sigma_i$. In Eq. (A.5) $y_{pred,j}^i$ is the predicted j -th value of the target variable y performed with the i -th model and μ_{pred}^i is the average of all the predictions performed with the same model.

Fig. A.9 shows λ for the various combinations of 1, 2, and 3 hidden layers (hl) and 9, 18, and 27 neurons for each of the six investigated water quality parameters: As, SO₄, Cl, Ec, Tw and DO. We remark that a model with a larger λ is statistically preferable to a model with a smaller λ . Given this, we selected a network with two hidden layers and 9 neurons for each layer since it produced the largest λ for almost all the water quality parameters (see red symbols in the Fig. A.9), except for As. This architecture of the DNN model was the only considered in the following operations, from the regression with the complete set of drivers to the Importance Features Assessment (in this second step, the input neurons will change according to the features employed in the regression, but the neurons in the hidden layers remain 9).

A.3. Linear Regression model

A possible simple and broadly employed alternative to the above ML algorithms is the classical Linear Regression (LR) model. In the present work, we also considered this model to better evidence the advantage offered by ML techniques with respect to an algorithm of comparable complexity, widely used in interpreting environmental data. LR seeks for an optimal linear relationship between a dependent variable, y , and a number of independent variables x_i , $i = 1, \dots, m$, called regressors, assumed reciprocally independent:

$$y = w_0 x_0 + w_1 x_1 + \dots + w_m x_m = \sum_{i=0}^m w_i x_i \quad (\text{A.6})$$

where w_1, \dots, w_m are the regression coefficients (i.e., the weights that should be optimized). Eq. (A.6) envisions a multiple linear relationship between the dependent variable and the regressors. In this work, the independent variables x_i are the regressors (or drivers) described in Section 2.2. The LR model was implemented by using the python package *scikit-learn* (Pedregosa et al., 2011).

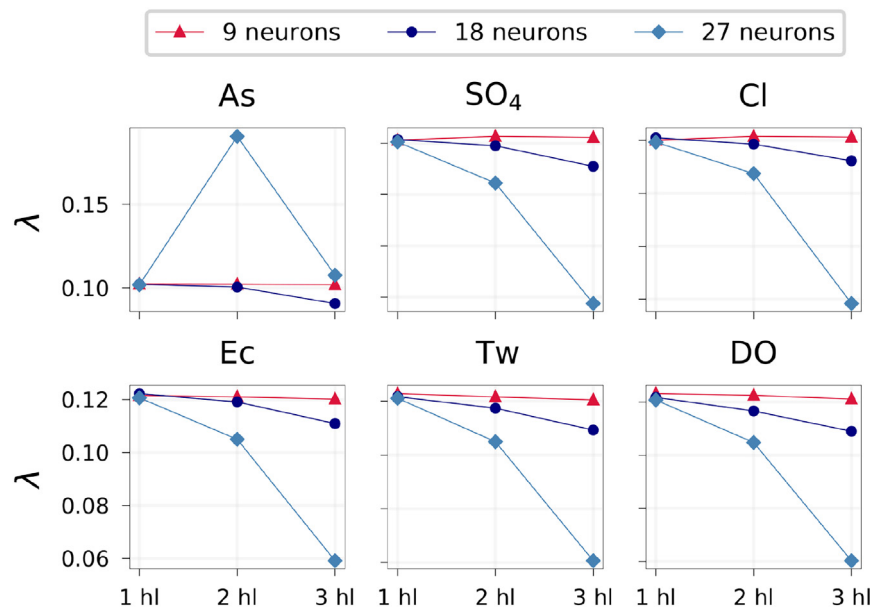


Fig. A.9. Lambda weights for the DNN model related to the AIC for each investigated water quality variable, considering different numbers of hidden layers (hl as reported in the x-axis) and neurons (displayed with different colors and lines). (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.156377>.

References

- Al-Mukhtar, M., Al-Yaseen, F., 2019. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq marsh in south of Iraq. *Hydrology* 6, 24. <https://doi.org/10.3390/hydrology6010024>.
- Anmala, J., Venkateshwarlu, T., 2019. Statistical assessment and neural network modeling of stream water quality observations of Green River watershed, KY, USA. *Water Sci. Technol. Water Supply* 19, 1831–1840. <https://doi.org/10.2166/ws.2019.058>.
- Antolini, P., 1984. Rassegna dei principali affioramenti di gesso in Italia. *Att. Acc. Agiati*. XXIV (B). Accademia Roveretana degli Agiati, pp. 83–117. <https://media.agiati.org/page/attachments/agiati-atti-b-1984-p.antolini-p.83.pdf>.
- Basant, N., Gupta, S., Malik, A., Singh, K.P., 2010. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water - a case study. *Chemom. Intell. Lab. Syst.* 104, 172–180. <https://doi.org/10.1016/j.chemolab.2010.08.005>.
- Basu, N.B., Destouni, G., Jawitz, J.W., Thompson, S.E., Loukinova, N.V., Darracq, A., Zanardo, S., Yaeger, M., Sivapalan, M., Rinaldo, A., Rao, P.S.C., 2010. Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. *Geophys. Res. Lett.* 37. <https://doi.org/10.1029/2010GL045168>.
- Bhattacharya, P., Welch, A.H., Stollenwerk, K.G., McLaughlin, M.J., Bundschuh, J., Panaullah, G., 2007. Arsenic in the environment: biology and chemistry. *Sci. Total Environ.* 379, 109–120. <https://doi.org/10.1016/j.scitotenv.2007.02.037>.
- Borsato, A., Miorandi, R., Corradini, F., Frisia, S., 2007. Idrochimica delle acque ipogee in Trentino: specie, variabilità stagionale, gradient altitudinale e implicazioni per gli studi climatico-ambientali da speleotemi. *Studi Trentini Sci. Nat. Acta Geol.* 82, 123–150. https://www2.muse.it/publicazioni/6/actaG82/MUSCC_Vol_Acta_Geo_12.pdf.
- Botter, G., Settin, T., Marani, M., Rinaldo, A., 2006. A stochastic model of nitrate transport and cycling at basin scale. *Water Resour. Res.* 42. <https://doi.org/10.1029/2005WR004599>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Buch, A., Mazumdar, H., Pandey, P., 1993. Application of artificial neural networks in hydrological modeling: a case study of runoff simulation of a Himalayan Glacier basin. *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*. IEEE, pp. 971–974. <https://doi.org/10.1109/IJCNN.1993.714073>.
- Burnham, K.P., Anderson, D.R., 1998. *Model Selection And Inference. A Practical Information-theoretic Approach*. Springer, New York. <https://doi.org/10.1007/978-1-4757-2917-7>.
- Caissie, D., 2006. The thermal regime of rivers: a review. *Freshw. Biol.* 51, 1389–1406. <https://doi.org/10.1111/j.1365-2427.2006.01597.x>.
- Chapman, D., 1992. In: Chapman, D. (Ed.), *Water Quality Assessments: A Guide to the Use of Biota, Sediments And Water in Environmental Monitoring*. on behalf of UNESCO, WHO and UNEP, Chapman & Hall, London. 585 pp. https://apps.who.int/iris/bitstream/handle/10665/41850/0419216006_eng.pdf?sequence=1&i.
- Chappell, N.A., Jones, T.D., Tych, W., 2017. Sampling frequency for water quality variables in streams: systems analysis to quantify minimum monitoring rates. *Water Res.* 123, 49–57. <https://doi.org/10.1016/j.watres.2017.06.047>.
- Chen, W.B., Liu, W.C., 2014. Artificial neural network modeling of dissolved oxygen in reservoir. *Environ. Monit. Assess.* 186, 1203–1217. <https://doi.org/10.1007/s10661-013-3450-6>.
- Chiogna, G., Majone, B., Paoli, K.C., Diamantini, E., Stella, E., Mallucci, S., Lencioni, V., Zandonai, F., Bellin, A., 2016. A review of hydrological and chemical stressors in the Adige catchment and its ecological status. *Sci. Total Environ.* 540, 429–443. <https://doi.org/10.1016/j.scitotenv.2015.06.149>.
- Chollet, F., 2018. *Deep Learning With Python*. Manning Shelter Island. <https://doi.org/10.1007/978-1-4842-2766-4>.
- CIESIN, 2017. Center for International Earth Science Information Network - Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10. NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY. <https://doi.org/10.7927/H4DZ068D>.
- Decreto Legislativo 13 ottobre 2015, n.172, 2015. Attuazione della direttiva 2013/39/UE, che modifica le direttive 2000/60/CE per quanto riguarda le sostanze prioritarie nel settore della politica delle acque. (15G00186). URLGazzetta ufficiale. www.gazzettaufficiale.it/eli/id/2015/10/27/15G00186/sg.
- Cortecchi, G., Dinelli, E., Bencini, A., Adorni-Braccesi, A., La Ruffa, G., 2002. Natural and anthropogenic so₄ sources in the Arno river catchment, northern Tuscany, Italy: a chemical and isotopic reconnaissance. *Appl. Geochem.* 17, 79–92. [https://doi.org/10.1016/S0883-2927\(01\)00100-7](https://doi.org/10.1016/S0883-2927(01)00100-7) URL: <https://www.sciencedirect.com/science/article/pii/S0883292701001007>.
- Council of European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32000L0060>.
- Cozzi, M., 2019. Cartografia geochimica del fiume Adige e dei suoi affluenti a monte di Bolzano. Università di Bologna - Analisi e Gestione dell'Ambiente Master's thesis. https://amslaurea.unibo.it/18081/1/cozzi_marco_tesi.pdf.
- Crawford, J.T., Hinkley, E.L.S., Litaor, M.I., Brahney, J., Neff, J.C., 2019. Evidence for accelerated weathering and sulfate export in high alpine environments. *Environ. Res. Lett.* 14. <https://doi.org/10.1088/1748-9326/ab5d9c>.
- Csábrági, A., Molnár, S., Tanos, P., Kovács, J., 2017. Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube. *Ecol. Eng.* 100, 63–72. <https://doi.org/10.1016/j.ecoleng.2016.12.027>.
- Dalla Libera, N., Pedretti, D., Tateo, F., Mason, L., Piccinini, L., Fabbri, P., 2020. Conceptual model of arsenic mobility in the shallow alluvial aquifers near Venice (Italy) elucidated through machine learning and geochemical modeling. *Water Resour. Res.* 56. <https://doi.org/10.1029/2019WR026234>.
- Diamantini, E., Lutz, S.R., Mallucci, S., Majone, B., Merz, R., Bellin, A., 2018. Driver detection of water quality trends in three large European river basins. *Sci. Total Environ.* 612, 49–62. <https://doi.org/10.1016/j.scitotenv.2017.08.172>.
- Diamantini, E., Mallucci, S., Bellin, A., 2019. A parsimonious transport model of emerging contaminants at the river network scale. *Hydrol. Earth Syst. Sci.* 23, 573–593. <https://doi.org/10.5194/hess-23-573-2019>.

- EEA, 2013. European Environment Agency. Corine Land Cover (CLC) 2006, Version 2020. Available online <http://land.copernicus.eu/pan-european/corine-land-cover/clc-2006/view> Copenhagen, Denmark.
- Fan, J., Wang, S., Li, H., Yan, Z., Zhang, Y., Zheng, X., Wang, P., 2020. Modeling the ecological status response of rivers to multiple stressors using machine learning: a comparison of environmental DNA metabarcoding and morphological data. *Water Res.* 183, 116004. <https://doi.org/10.1016/j.watres.2020.116004>.
- Freeze, A., Cherry, J.A., 1979. *Groundwater*. Prentice Hall, Englewood Cliffs, NJ07632.
- Fuganti, A., Morteani, G., Bazzoli, G., Cocco, S., Santuliana, E., Visintainer, M., 2005. L'arsenico nelle rocce, nelle acque superficiali e nelle acque sotterranee della valle dell'adige fra mezzolombardo e mattarello e presso roveré della luna (trento). *Atti Acc. Rov. Agiati* 5 (B), 59–94. <https://media.agiati.org/page/attachments/agiati-atti-b-2005-art-04-fuganti.pdf>.
- Gabrielli, P., Cozzi, G., Torcini, S., Cescon, P., Barbante, C., 2006. Atmospheric trace elements in Alpine snow source and origin of atmospheric trace elements entrapped in winter snow of the Italian Eastern Alps. *Atmospheric trace elements in Alpine snow*. *Atmos. Chem. Phys. Discuss* 6, 8781–8815. <http://www.atmos-chem-phys-discuss.net/6/8781/2006/>.
- Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. In: Gordon, G., Dunson, D., Dudík, M. (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. PMLR, Fort Lauderdale, FL, USA, pp. 315–323. <https://proceedings.mlr.press/v15/glorot11a.html>.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT press. <http://www.deeplearningbook.org>.
- Grathwohl, P., Rügner, H., Wöhling, T., Osenbrück, K., Schwientek, M., Gayler, S., Wollschläger, U., Selle, B., Pause, M., Delfs, J.O., Grzeschik, M., Weller, U., Ivanov, M., Cirpka, O.A., Maier, U., Kuch, B., Nowak, W., Wulfmeyer, V., Warrach-Sagi, K., Streck, T., Attinger, S., Bilke, L., Dietrich, P., Fleckenstein, J.H., Kalbacher, T., Kolditz, O., Rink, K., Samaniego, L., Vogel, H.J., Werban, U., Teutsch, G., 2013. Catchments as reactors: a comprehensive approach for water fluxes and solute turnover. *Environ. Earth Sci.* 69, 317–333. <https://doi.org/10.1007/s12665-013-2281-7>.
- Green, M.B., Pardo, L.H., Bailey, S.W., Campbell, J.L., McDowell, W.H., Bernhard, E.S., Rosi, E.J., 2021. Predicting high-frequency variation in stream solute concentrations with water quality sensors and machine learning. *Hydrol. Process.* 35. <https://doi.org/10.1002/hyp.14000>.
- Guillet, G., Knapp, J.L., Merel, S., Cirpka, O.A., Grathwohl, P., Zwiener, C., Schwientek, M., 2019. Fate of wastewater contaminants in rivers: using conservative-tracer based transfer functions to assess reactive transport. *Sci. Total Environ.* 656, 1250–1260. <https://doi.org/10.1016/j.scitotenv.2018.11.379>.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hadjisolomou, E., Stefanidis, K., Herodotou, H., Michaelides, M., Papatheodorou, G., Papastergiadou, E., 2021. Modelling freshwater eutrophication with limited limnological data using artificial neural networks. *Water (Switzerland)* 13, 1–15. <https://doi.org/10.3390/w13111590>.
- Harvey, C.F., Swartz, C.H., Badruzzaman, A., Keon-Blute, N., Yu, W., Ali, M.A., Jay, J., Beckie, R., Niefan, V., Brabander, D., et al., 2002. Arsenic mobility and groundwater extraction in Bangladesh. *Science* 298, 1602–1606. <https://doi.org/10.1126/science.1076978>.
- Hasan, R., Raghav, A., Mahmood, S., Hasan, M.A., 2011. Artificial intelligence in river quality assessment. 2011 International Conference on Information Management, Innovation Management And Industrial Engineering. IEEE, pp. 491–495. <https://doi.org/10.1109/ICIMI.2011.125>.
- Heddam, S., Kisi, O., 2017. Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. *Environ. Sci. Pollut. Res.* 24, 16702–16724. <https://doi.org/10.1007/s11356-017-9283-z>.
- Hipel, K.W., McLeod, A.I., 1994. *Time Series Modelling of Water Resources And Environmental Systems*. Elsevier.
- Hou, M., Yang, Y., Liu, T., Peng, W., 2018. Forecasting time series with optimal neural networks using multi-objective optimization algorithm based on AICc. *Front. Comput. Sci.* 12, 1261–1263. <https://doi.org/10.1007/s11704-018-8095-8>.
- Hsu, K.L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.* 31, 2517–2530. <https://doi.org/10.1029/95WR01955>.
- Hu, J.H., Tsai, W.P., Cheng, S.T., Chang, F.J., 2020. Explore the relationship between fish community and environmental factors by machine learning techniques. *Environ. Res.* 184, 109262. <https://doi.org/10.1016/j.envres.2020.109262>.
- Kamrava, S., Im, J., de Barros, F.P.J., Sahimi, M., 2021. Estimating dispersion coefficient in flow through heterogeneous porous media by a deep convolutional neural network. *Geophys. Res. Lett.* 48, e2021GL094443. <https://doi.org/10.1029/2021GL094443>.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. 3rd International Conference for Learning Representations (ICLR). <https://arxiv.org/abs/1412.6980v9>.
- Kirchner, J.W., Neal, C., 2013. Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *Proc. Natl. Acad. Sci.* 110, 12213–12218. <https://doi.org/10.1073/pnas.1304328110>.
- Kirchner, J.W., Feng, X., Neal, C., 2000. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature* 403, 524–527. <https://doi.org/10.1038/35000537>.
- Kirchner, J.W., Feng, X., Neal, C., 2001. Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations. *J. Hydrol.* 254, 82–101. [https://doi.org/10.1016/S0022-1694\(01\)00487-5](https://doi.org/10.1016/S0022-1694(01)00487-5).
- Kormann, C., Francke, T., Renner, M., Bronstert, A., 2015. Attribution of high resolution streamflow trends in Western Austria - an approach based on climate and discharge station data. *Hydrol. Earth Syst. Sci.* 19, 1225–1245. <https://doi.org/10.5194/hess-19-1225-2015>.
- Kouadri, S., Elbeltagi, A., Islam, A.R.M.T., Kateb, S., 2021. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). *Appl. Water Sci.* 11. <https://doi.org/10.1007/s13201-021-01528-9>.
- Krishnaraj, A., Deka, P.C., 2020. Spatial and temporal variations in river water quality of the Middle Ganga Basin using unsupervised machine learning techniques. *Environ. Monit. Assess.* 192, 1–18. <https://doi.org/10.1007/s10661-020-08624-4>.
- Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., Nikulin, G., Majone, B., 2018. Testing the hydrological coherence of high-resolution gridded precipitation and temperature data sets. *Water Resour. Res.* 54, 1999–2016. <https://doi.org/10.1002/2017WR021633>.
- Landrigan, P., Fuller, R., Acosta, N., Adeyi, O., Arnold, R., Basu, N., Baldé, A., Bertollini, R., Bose-O'Reilly, S., Boufford, J., Breyse, P., Chiles, T., Mahidol, C., Coll-Seck, A., Cropper, M., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K., McTeer, M., Murray, C., Ndahimananjara, J., Perera, F., Potočnik, J., Preker, A., Ramesh, J., Rockström, J., Salinas, C., Samson, L., Sandilya, K., Sly, P., Smith, K., Steiner, A., Stewart, R., Suk, W., van Schayck, O., Yadama, G., Yumkella, K., Zhong, M., 2018. The lancet commission on pollution and health. *Lancet* 391, 462–512. [https://doi.org/10.1016/S0140-6736\(17\)32345-0](https://doi.org/10.1016/S0140-6736(17)32345-0) Erratum in: *Lancet*. 2018 Feb 3;391(10119):430. PMID: 29056410.
- Lencioni, V., Stella, E., Zanolini, M.G., Bellin, A., 2022. On the delay between water temperature and invertebrate community response to warming climate. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2022.155759>.
- Liu, X., Lu, D., Zhang, A., Liu, Q., Jiang, G., 2022. Data-driven machine learning in environmental pollution: gains and problems. *Environ. Sci. Technol.* 56, 2124–2133. <https://doi.org/10.1021/acs.est.1c06157>.
- Lutz, S.R., Mallucci, S., Diamantini, E., Majone, B., Bellin, A., Merz, R., 2016. Hydroclimatic and water quality trends across three Mediterranean river basins. *Sci. Total Environ.* 571, 1392–1406. <https://doi.org/10.1016/j.scitotenv.2016.07.102>.
- Ma, T., Sun, S., Fu, G., Hall, J.W., Ni, Y., He, L., Yi, J., Zhao, N., Du, Y., Pei, T., Cheng, W., Song, C., Fang, C., Zhou, C., 2020. Pollution exacerbates China's water scarcity and its regional inequality. *Nat. Commun.* 11. <https://doi.org/10.1038/s41467-020-14532-5>.
- Maiello, F., 2006. Molti, diversi per origine, sorprendenti... i minerali del trentino. *Natura Alpina* 1, 23–52. https://www2.muse.it/publicazioni/8/57a/natalp.57_23-53.pdf.
- Mallucci, S., Majone, B., Bellin, A., 2019. Detection and attribution of hydrological changes in a large alpine river basin. *J. Hydrol.* 575, 1214–1229. <https://doi.org/10.1016/j.jhydrol.2019.06.020>.
- Marion, A., Zaramella, M., Bottacin-Busolin, A., 2008. Solute transport in rivers with multiple storage zones: the stir model. *Water Resour. Res.* 44. <https://doi.org/10.1029/2008WR007037>.
- Massei, R., Busch, W., Wolschke, H., Schinkel, L., Bitsch, M., Schulze, T., Krauss, M., Brack, W., 2018. Screening of pesticide and biocide patterns as risk drivers in sediments of major European river mouths: ubiquitous or river basin-specific contamination? *Environ. Sci. Technol.* 52, 2251–2260. <https://doi.org/10.1021/acs.est.7b04355>.
- McDonnell, J., McGuire, K., Aggarwal, P., Beven, K., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., et al., 2010. How old is streamwater? open questions in catchment transit time conceptualization, modeling and analysis. *Hydrol. Process.* 24, 1745–1754. <https://doi.org/10.1002/hyp.7796>.
- Mitchell, T.M., et al., 1997. *Machine Learning*. McGraw-Hill, New York.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50, 885–900. <https://pubag.nal.usda.gov/download/9298/pdf>.
- Nafi, S.N.M.M., Mustapha, A., Mostafa, S.A., Khaleefah, S.H., Razali, M.N., 2020. Experimenting two machine learning methods in classifying river water quality. In: Khalaf, M.I., Al-Jumeily, D., Lisitsa, A. (Eds.), *Applied Computing to Support Industry: Innovation And Technology*. Springer International Publishing, Cham, pp. 213–222. https://doi.org/10.1007/978-3-030-38752-5_17.
- Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part i - a discussion of principles. *J. Hydrol.* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Navarro-Ortega, A., Acuña, V., Bellin, A., Burek, P., Cassiani, G., Choukr-Allah, R., Dolédec, S., Eloegi, A., Ferrari, F., Ginebreda, A., Grathwohl, P., Jones, C., Rault, P.K., Kok, K., Koundouri, P., Ludwig, R.P., Merz, R., Milacic, R., Muñoz, I., Nikulin, G., Paniconi, C., Paunović, M., Petrovic, M., Sabater, L., Sabater, S., Skoulidakis, N.T., Slob, A., Teutsch, G., Voulvoulis, N., Barceló, D., 2015. Managing the effects of multiple stressors on aquatic ecosystems under water scarcity. The globaqua project. *Sci. Total Environ.* 503–504, 3–9. <https://doi.org/10.1016/j.scitotenv.2014.06.081>.
- Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V., 2021. What role does hydrological science play in the age of machine learning? *Water Resour. Res.* 57, e2020WR028091. <https://doi.org/10.1029/2020WR028091>.
- Nemati, S., Fazelifard, M.H., Terzi, Ö., Ghorbani, M.A., 2015. Estimation of dissolved oxygen using data-driven techniques in the Tai Po River, Hong Kong. *Environ. Earth Sci.* 74, 4065–4073. <https://doi.org/10.1007/s12665-015-4450-3>.
- Neumann, R.B., Ashfaq, K.N., Badruzzaman, A., Ali, M.A., Shoemaker, J.K., Harvey, C.F., 2010. Anthropogenic influences on groundwater arsenic concentrations in Bangladesh. *Nat. Geosci.* 3, 46–52. <https://doi.org/10.1038/ngeo685>.
- Nourani, V., Roushangar, K., Andalib, G., 2018. An inverse method for watershed change detection using hybrid conceptual and artificial intelligence approaches. *J. Hydrol.* 562, 371–384. <https://doi.org/10.1016/j.jhydrol.2018.05.018>.
- Panchal, G., Ganatra, A., Kosta, Y., Panchal, D., 2010. Searching most efficient neural network architecture using Akaike's Information Criterion (AIC). *Int. J. Comput. Applic.* 1, 54–57. <https://doi.org/10.5120/126-242>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python.

- J. Mach. Learn. Res. 12, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://githubhelp.com>.
- Perna, G., 2003. Trentino ed Alto Adige. In: Madonia, G., Forti, P. (Eds.), Aree Carsiche Gessose d'Italia. Istituto Italiano di Speleologia, Bologna (Italy), pp. 143–148. <http://www.venadelgesso.it/2019/ac/4t.pdf>.
- Piccolroaz, S., Calamita, E., Majone, B., Gallice, A., Siviglia, A., Toffolon, M., 2016. Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. *Hydrol. Process.* 30, 3901–3917. <https://doi.org/10.1002/hyp.10913>.
- Piggott, J.J., Lange, K., Townsend, C.R., Matthaei, C.D., 2012. Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. *PloS one* 7, e49873. <https://doi.org/10.1371/journal.pone.0049873>.
- Qiu, R., Wang, Y., Rhoads, B., Wang, D., Qiu, W., Tao, Y., Wu, J., 2021. River water temperature forecasting using a deep learning method. *J. Hydrol.* 595, 126016. <https://doi.org/10.1016/j.jhydrol.2021.126016>.
- Raschka, S., 2015. *Python Machine Learning*. Packt Publishing Ltd.
- Rinaldo, A., Beven, K.J., Bertuzzo, E., Nicotina, L., Davies, J., Fiori, A., Russo, D., Botter, G., 2011. Catchment travel time distributions and water flow in soils. *Water Resour. Res.* 47. <https://doi.org/10.1029/2011WR010478>.
- Rügner, H., Schwientek, M., Beckingham, B., Kuch, B., Grathwohl, P., 2013. Turbidity as a proxy for total suspended solids (tss) and particle facilitated pollutant transport in catchments. *Environ. Earth Sci.* 69, 373–380. <https://doi.org/10.1007/s12665-013-2307-1>.
- Schwarzenbach, R., Egli, T., Hofstetter, T.B., von Gunten, U., Wehrli, B., 2010. Global water pollution and human health. *Annu. Rev. Environ. Resour.* 35, 109–136. <https://doi.org/10.1146/annurev-environ-100809-125342>.
- Shah, M.I., Javed, M.F., Abunama, T., 2021. Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environ. Sci. Pollut. Res.* 28, 13202–13220. <https://doi.org/10.1007/s11356-020-11490-9>.
- Shamshirband, S., Jafari Nodoushan, E., Adolf, J.E., Abdul Manaf, A., Mosavi, A., Wing Chau, K., 2019. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Eng. Appl. Comput. Fluid Mech.* 13, 91–101. <https://doi.org/10.1080/19942060.2018.1553742>.
- Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* 54, 8558–8593. <https://doi.org/10.1029/2018WR022643>.
- Shukla, R., Kumar, P., Vishwakarma, D.K., Ali, R., Kumar, R., Kuriqi, A., 2022. Modeling of stage-discharge using back propagation ANN-, ANFIS-, and WANN-based computing techniques. *Theor. Appl. Climatol.* 147, 867–889. <https://doi.org/10.1007/s00704-021-03863-y> URL: 10.1007/s00704-021-03863-y.
- Smedley, P., Kinniburgh, D., 2002. A review of the source, behaviour and distribution of arsenic in natural waters. *Appl. Geochem.* 17, 517–568. [https://doi.org/10.1016/S0883-2927\(02\)00018-5](https://doi.org/10.1016/S0883-2927(02)00018-5).
- Stumm, W., Morgan, J., 1996. *Aquatic Chemistry, Chemical Equilibria And Rates in Natural Waters*. John Wiley & Sons Inc, New York.
- Sugiura, N., 1978. Further analysts of the data by Akaike's Information Criterion and the finite corrections. *Commun. Stat. Theory Methods* 7, 13–26. <https://doi.org/10.1080/03610927808827599>.
- Tiyasha, Tung, T.M., Yaseen, Z.M., 2020. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J. Hydrol.* 585, 124670. <https://doi.org/10.1016/j.jhydrol.2020.124670>.
- Todd, A.S., Manning, A.H., Verplanck, P.L., Crouch, C., McKnight, D.M., Dunham, R., 2012. Climate-change-driven deterioration of water quality in a mineralized watershed. *Environ. Sci. Technol.* 46, 9324–9332. <https://doi.org/10.1021/es3020056>.
- Toffolon, M., Piccolroaz, S., 2015. A hybrid model for river water temperature as a function of air temperature and discharge. *Environ. Res. Lett.* 10. <https://doi.org/10.1088/1748-9326/10/11/114011>.
- Wörman, A., 1998. Analytical solution and timescale for transport of reacting solutes in rivers and streams. *Water Resour. Res.* 34, 2703–2716. <https://doi.org/10.1029/98WR01338>.
- Xu, T., Liang, F., 2021. Machine learning for hydrologic sciences: an introductory overview. *Wiley Interdiscip. Rev. Water* 8, 1–29. <https://doi.org/10.1002/wat2.1533>.
- Zarnetske, J.P., Haggerty, R., Wondzell, S.M., Bokil, V.A., Gonzzález-Pinzón, R., 2012. Coupled transport and reaction kinetics control the nitrate source-sink function of hyporheic zones. *Water Resour. Res.* 48. <https://doi.org/10.1029/2012WR011894>.
- Zhao, Z., Zhang, Y., Liao, H., 2008. Design of ensemble neural network using the Akaike Information Criterion. *Eng. Appl. Artif. Intell.* 21, 1182–1188. <https://doi.org/10.1016/j.engappai.2008.02.007>.
- Zhu, S., Nyarko, E.K., Hadzima-Nyarko, M., 2018. Modelling daily water temperature from air temperature for the Missouri River. *PeerJ* 2018, 1–19. <https://doi.org/10.7717/peerj.4894>.
- Zhu, S., Nyarko, E.K., Hadzima-Nyarko, M., Heddam, S., Wu, S., 2019. Assessing the performance of a suite of machine learning models for daily river water temperature prediction. *PeerJ* 7. <https://doi.org/10.7717/peerj.7065>.