

# Fast multi-output relevance vector regression for joint groundwater and lake water depth modeling

Mir Jafar Sadegh Safari<sup>a,\*</sup>, Shervin Rahimzadeh Arashloo<sup>b</sup>, Babak Vaheddoost<sup>c</sup>

<sup>a</sup> Department of Civil Engineering, Yaşar University, İzmir, Turkey

<sup>b</sup> Department of Computer Engineering, Bilkent University, Ankara, Turkey

<sup>c</sup> Department of Civil Engineering, Bursa Technical University, Bursa, Turkey

## ARTICLE INFO

### Keywords:

Fast multi-output relevance vector regression  
Groundwater  
Lake urmia  
Lake water depth  
Multi-output regression  
Support vector regression

## ABSTRACT

Fast multi-output relevance vector regression (FMRVR) algorithm is developed for simultaneous estimation of groundwater and lake water depth for the first time in this study. The FMRVR is a multi-output regression analysis technique which can simultaneously predict multiple outputs for a multi-dimensional input. The data used in this study is collected from 34 stations located in the lake Urmia basin over a 40-year time period. The performance of the FMRVR model is examined in contrast to the support vector regression (SVR) and multi-linear regression (MLR) benchmarks. Results reveal that FMRVR is able to generate more accurate estimation for groundwater and lake water depth with coefficient of determination ( $R^2$ ) of 0.856 and 0.992 and root mean square error (RMSE) of 0.857 and 0.083, respectively. The outperformance of FMRVR can be linked to its capability for a joint estimation of multiple relevant outputs by taking into account possible correlations among the outputs.

## 1. Introduction

It is known that groundwater and lake water depth are two most important parameters that are effective in the hydrological and ecological behavior of lake systems (Vaheddoost and Aksoy, 2018; Javadzadeh et al., 2020; Kozekalani Sales et al., 2021). Accordingly, lakes play an important role as indicators for climate-born events such as climate change or climate variability (Tong et al., 2016; Zhang et al., 2019). In a closed-basin hydrologic environment, each water drop eventually ends up in the water body at the lowest parts of the terrain. This water body is usually a permanent or seasonal lake that is under the influence of micro-climate, surrounding aquifers (e.g. groundwater storage), streamflow, precipitation, and evaporation. In this regard, almost all of the hydro-meteorological parameters except groundwater have a one-way relationship with lake water depth, as groundwater represents a dynamic act of water-loss and -gain between nearby aquifers and the lake. Owing to the complexity of the process between the groundwater and the lake water depth, most of the studies in the relevant literature only investigated the lake water level behavior. This is mostly due to the complexity of the groundwater systems, high precision and bias in estimation of the groundwater state, and the unavailability of

the data representing the spatiotemporal behavior of the system.

Lake Urmia is a Ramsar site and a UNESCO Biosphere Reserve declared respectively in 1971 and 1976 (Pengra, 2012). It has a pivotal role in the socio-economy, and hydrology of the regions and a home to species such as *Artemia Urmiana* and Brine shrimp. In recent years however, lake faced with major environmental issues, most of which are related to the continuous encroachment in the hydrology of the basin, anthropogenic changes and gradual climate change. Dealing with a gradual decrease in the water depth of lake Urmia, much effort has been made for modeling and investigation of the lake behavior due to its critical role in the ecosystem of the region (Shadkam et al., 2016; Jekhouni et al., 2017; Chaudhari et al., 2018; Khazaei et al., 2019). For instance, by considering the climate change and via applying a hydro-dynamic model through utilizing hydro-meteorological parameters, the drying condition of lake Urmia is investigated by Abbaspour et al. (2012). Hassanzadeh et al. (2012) applied a system dynamic method to scrutinize the reasons for declining water levels of lake Urmia. It was confirmed that the variation in inflow to the lake as a consequence of the climate change has a great impact, while dam construction and precipitation decrease have also significant roles in the decrease of lake Urmia water level. Sima and Tajrishy (2013) utilized remote sensing data to study the volume-area-elevation relationship in lake Urmia. A

\* Corresponding author.

E-mail addresses: [jafar.safari@yasar.edu.tr](mailto:jafar.safari@yasar.edu.tr) (M.J.S. Safari), [s.rahimzadeh@cs.bilkent.edu.tr](mailto:s.rahimzadeh@cs.bilkent.edu.tr) (S. Rahimzadeh Arashloo), [babak.vaheddoost@btu.edu.tr](mailto:babak.vaheddoost@btu.edu.tr) (B. Vaheddoost).

<https://doi.org/10.1016/j.envsoft.2022.105425>

Received 7 April 2021; Received in revised form 26 April 2022; Accepted 18 May 2022

Available online 21 May 2022

1364-8152/© 2022 Elsevier Ltd. All rights reserved.

### Abbreviations

ANFIS	Adaptive network-based fuzzy inference system
ANN	Artificial neural network
EM	Expectation maximisation algorithm
ELM	Extreme learning machine
FMRVR	Fast multi-output relevance vector regression
FFA	Firefly algorithm
GPR	Gaussian process regression
GP	Genetic programming
MPMR	Minimax probability machine regression
MP	Most probable hyper-parameters
MLP	Multi-layer perceptron artificial neural networks
MLR	Multi-linear regression
PSO	Particle swarm optimization
RBNN	Radial basis neural networks
RVM	Relevance vector machine
SLR	Simple linear regression
SMLR	Stepwise multiple linear regression
SVM	Support vector machines
SVR	Support vector regression
WA	Wavelet

decline in the area and the volume of lake Urmia was emphasized followed by recommendations on the required water level to satisfy ecological requirements. [Marjani and Jamali \(2014\)](#) examined the impact of flow exchange in the Urmia lake dynamics. To this end, the impact of constructed causeway on the salt water balance of Urmia lake was studied. Comparing the results with the collected field data revealed that the variation in the lake water levels at both sides of causeway has significant impact on the flow exchange within the lake. [Arkian et al. \(2016\)](#) performed trend analysis based on climate parameters to investigate a possible relationship between different meteorological parameters and the reduction in lake Urmia water level. [Dehghanipour et al. \(2020\)](#) studied the impact of Urmia lake on the region's climate through examining the time series data of different meteorological parameters. The effect of an increase in the temperature was identified as an important parameter in the regional climate.

Groundwater in the lake Urmia basin has been relatively less studied in the literature compared to the corresponding water levels. Among the few studies, [Zarghami \(2011\)](#) highlighted the importance of water usage for irrigation in the reduction of groundwater level in the basin. [Tourian et al. \(2015\)](#) utilized satellite data and analyzed water balance in the lake Urmia. It was found that the runoff coefficient declined significantly in proportion with the reduction in groundwater level in the basin. [Amirataee and Zeinalzadeh \(2016\)](#) considered autocorrelation coefficients and applied modified Mann–Kendall test to analyze the quantitative and qualitative parameters of the Urmia lake basin groundwater. [Amiri et al. \(2016a, 2016b\)](#) studied the potential saltwater intrusion to the groundwater aquifers and highlighted its critical role in wet seasons. [Ashraf et al. \(2017\)](#) investigated the depletion behavior of several aquifers in Urmia lake basin while [Vaheddoost and Aksoy \(2018\)](#) studied the interaction between the Urmia lake water level and groundwater through a base-flow separation approach and concluded that a significant part of the inflow to the lake is in the form of base-flow, and accordingly, groundwater level reduction may accelerate the reduction in the lake water level. [Javadzadeh et al. \(2020\)](#) applied cross-correlation analysis and found strong correlation between the Urmia lake water level and groundwater. [Sheibani et al. \(2020\)](#) investigated the effect of bed sedimentation on the interaction between groundwater and Lake Urmia. Several scenarios including steady-state conditions are investigated and concluded that the lake bathymetry, sediment thickness, and seasonal state of water level in the lake cause

either salinity intrusion from the lake to the coastal aquifer or the lake water body being fed by the aquifer.

Finding an analytical solution to develop a hydrological model based on water budget parameters that affect the groundwater and lake water depth is quite a challenging task and may be impossible based on the existing knowledge. Among a variety of different approaches, machine learning methods may provide a reliable method to find a possible relationship between the effective parameters involved in the problem ([Yaseen et al., 2020](#); [Pandey et al., 2020](#); [Safari, 2020](#); [Kumar et al., 2020](#); [Safari et al., 2020](#); [Rehannia et al., 2021](#)). Such a modeling approach builds on the data to approximate a function based on system information. Machine learning techniques have been widely applied for lake water depth modeling. For instance, [Çimen and Kisi \(2009\)](#) applied support vector machines (SVM) and artificial neural network (ANN); [Buyukyildiz et al. \(2014\)](#) implemented support vector regression (SVR), particle swarm optimization-artificial neural networks (PSO-ANN), radial basis neural networks (RBNN), multi-layer artificial neural networks (MLP) and adaptive network-based fuzzy inference system (ANFIS); [Yadav and Eliza \(2017\)](#) utilized wavelet-support vector machine (WA-SVM); [Ghorbani et al. \(2018\)](#) applied firefly algorithm (FFA) with the multilayer perceptron (MLP) algorithm; [Bonakdari et al. \(2019\)](#) used relevance vector machine (RVM), minimax probability machine regression (MPMR), extreme learning machine (ELM) and Gaussian process regression (GPR) and, [Long et al. \(2019\)](#) used simple linear regression (SLR) and stepwise multiple linear regression (SMLR) for lake water level modeling. For groundwater modeling, [Gorgij et al. \(2016\)](#) utilized wavelet-artificial neural network-genetic programming (WA-ANN-GP) while [Nourani and Mousavi \(2016\)](#) made use of ANN and ANFIS techniques. More recently, [Rabet et al. \(2020\)](#) used ANN, random forest, SVR, and linear regression to investigate the groundwater potential around the Lake Urmia. Several parameters including slope, slope direction, altitude, curvature, stream and fault density, distance to stream and fault, average rainfall, lithology, land use, relative slope position, topographic wetness index, and topographic position index were used. It was concluded that the ANN is the superior model, while the obtained results indicate that 31.4% of the basin has relatively high groundwater potential. Also, [Radman et al. \(2021\)](#) incorporated the rainfall, groundwater, and lake area variations from satellite measurement together with deep learning and InSAR methods in evaluation of the lands in the Lake Urmia vicinity. The deformations in the ground were also estimated using MLP, long short-term memory networks, and convolutional neural network. It was concluded that an ensemble model is capable of improving the land deformation anticipation by the help of networks in various conditions.

In contrast to the exiting methods, in this study, we advocate the use of a well-established method for regression analysis, i.e., relevance vector regression. The method provides a number of appealing characteristics as follows. First, it is built on probabilistic assumptions, and thus, better reflects the uncertainties associated with the problem in hand. Second, drawing on its multi-output formulation, it is better suited to handling problems where multiple relevant outputs should be simultaneously predicted for a given input. A simultaneous prediction of multiple relevant outputs provides the necessary ground for modeling any existing correlations between several output variables and the applied method in this study models such correlations via potentially non-diagonal covariance matrices incorporated into the model. And last but not the least, in contrast to some other multi-output regression analysis techniques, the multi-output relevance vector regression applied in this work enjoys better computational complexity properties. Most of the studies in the literature neglect the role of groundwater in the lake water depth modeling and there are a few studies that only investigated the interaction and correlation between the groundwater and the lake water depth parameters. It is mostly due to the complex nature of the water-loss and –gain in the coastal zone between coastal aquifers and lake water body which prevents strict forward extrapolation or an explicit modelling. More importantly, the reported machine

learning models are confined to a single output considered as the lake water depth while in a few other cases the single output variable is considered as the groundwater level. In addition to incorporating different hydro-meteorological parameters into the model, this study recommends a multi-output learning-based hydrological model for simultaneous estimation of groundwater and lake water levels by applying the FMRVR algorithm for the first time in the literature.

## 2. Materials and methods

### 2.1. Study area and data

Lake Urmia, with a semi-arid climate, is located in the northwest of Iran (44°50'~46°10' E; 36° 45'~38°20' N) and known as the second largest hypersaline lake in the World (Vaheddoost and Aksoy 2019; Dehghanipour et al., 2020). The Lake Urmia has an approximate area of 52,000 km<sup>2</sup> to 6100 km<sup>2</sup> at its wet periods. The temperature varies between 0 °C and -20 °C in winter, while reaches up to 40 °C in the summer. The annual precipitation, is mostly observed as rainfall in the lowlands and snowfall in highlands, reaching up to 326 mm annually (Vaheddoost and Aksoy 2017). It is a UNESCO Biosphere Reserve, a Ramsar sight, and a protected national park that deserve protection. It has a closed basin (Fig. 1), with a relatively rich data recorded at the meteorological stations (more than 200), hydrometric stations (more than 100), or groundwater wells (more than 500).

In this study the hydrometeorological data are acquired from the Iranian Water Resources Management Company (IWRM Co.) in a monthly average format for the period between November 1974 to September 2014 (479 month) in terms of evaporation (*E*), precipitation (*P*), lake water depth (*L*), streamflow (*S*), and groundwater level (*G*) time series. The missing data were reconstructed by means of frequency domain analysis of the additive parameters, while the stationarity, randomness,

trend and consistency in the time series respectively were tested using Augmented Dickey–Fuller test, run test, Spearman rank order correlation and double mass curve.

Lake water depth is preferred to the lake water level to reduce the precision error associated with the obtained results (Vaheddoost and Aksoy, 2019). For this purpose, the average bottom level of the lake (i.e. 1267.08 m above sea level) is subtracted from the monthly lake water levels to obtain the monthly lake water depth time series. Afterwards, the summation of streamflow records related to seventeen rivers streams ending to the Lake Urmia (Table 1) is used as single time series representing the streamflow (*S*). On the other hand, 7 out of 156 meteorological stations in the vicinity of the lake are used (Table 1) to estimate

**Table 1**  
Summary for the data used and the allocated stations in the analysis.

Variable	Stations used	Average	Standard deviation
Lake water depth	Golmankhaneh	7.99 m	1.89 m
Streamflow	Aji, Azarshahr, Baranduz, CheshmeDul, Dariyan, Gadar, Ghala, Khorkhore, Leylan, Mahabad, Maraghe, Nazloo, Ruzeh, Shahar, Simineh, Sufi, Zarrineh, Zola	123.95 m <sup>3</sup> /s	127.88 m <sup>3</sup> /s
Precipitation	Abajalu-Sofla, Azarshahr, Gharalar, Moghanjugh, Peyghala, PoleSorkh, Sharafkhane	26.73 mm	22.92 mm
Evaporation	Abajalu-Sofla, Azarshahr, Gharalar, Moghanjugh, Peyghala, PoleSorkh, Sharafkhane	4.01 mm	3.22 mm
Groundwater	Adeh, Ali-Molk, Begi-Osbi, Hesar, Majd-Abad, Sheikh-Vali, Shishvan, Soltan-Ahmad, Vorjuy	40.87 m	2.22 m

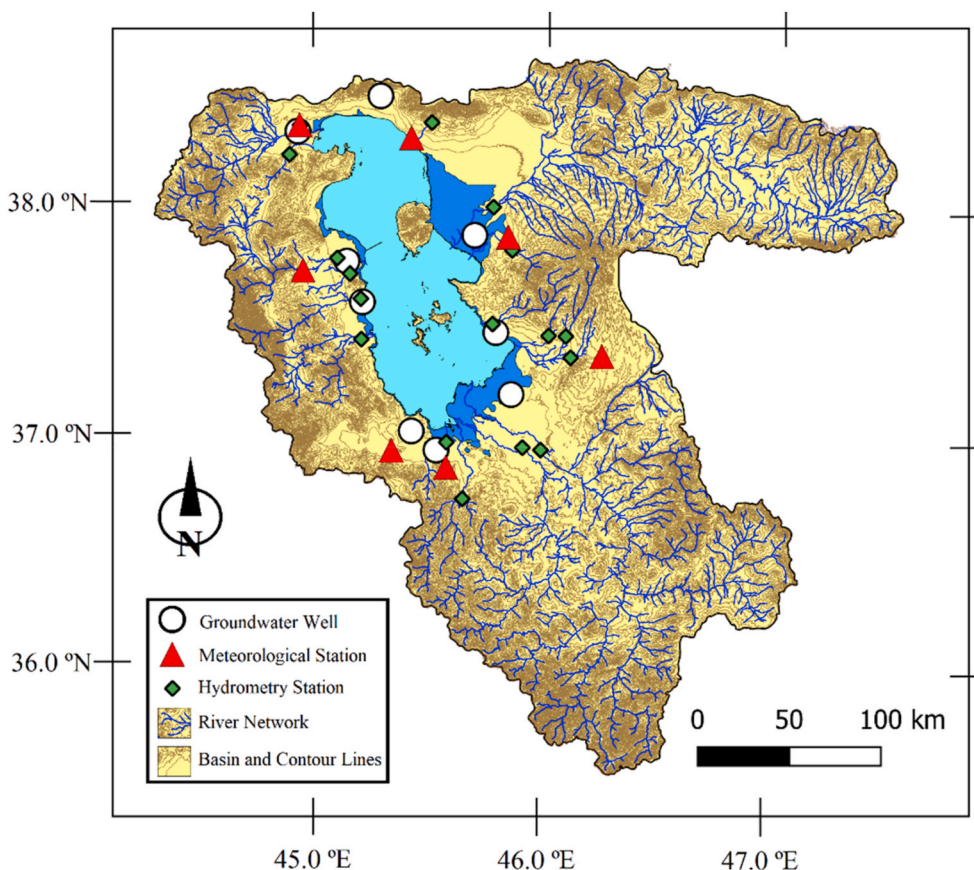


Fig. 1. Study area and selected stations in this study.

both the precipitation ( $P$ ) on and evaporation ( $E$ ) from the surface of the lake while the share of each station on the representative time series is estimated using Thiessen polygon method (Fig. 2). It is also important to note that a 0.76 pan coefficient is used on evaporation values to transfer Class-A pan values to saturated lake surface condition. Likewise, the difference between average groundwater level at 9 observation wells (Table 1) with lake water level in each month is used to obtain the potential groundwater head that can be discharged into the lake.

### 2.2. Modeling scenarios and assumptions

Models are developed based on a number of preliminary assumptions as follows:

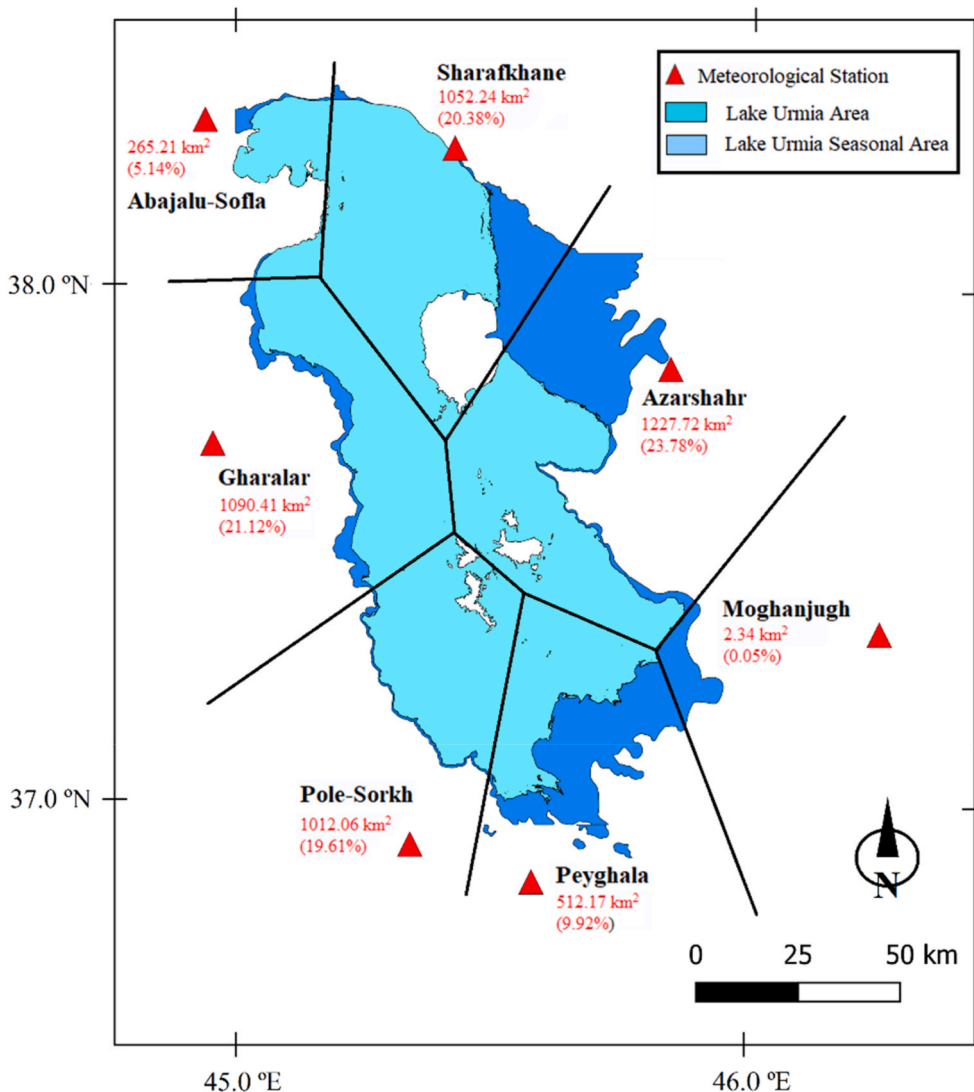
- I. Lake is the control volume which can be studied as a dependent variable, while other phenomena that take place in the basin can be neglected since the allocated variables represents the behavior of the control volume.
- II. Precipitation and evaporation are the independent variables in the model, while groundwater can act either as a dependent or independent variable if the conducted analysis is the outcome of a single- or multi-output model.

- III. The spatial contribution of the predictor(s) on the outcome is lumped into the contribution of variables in different stations, while the temporal contribution is handled by using lagged time in the predictors.
- IV. The relationships between variables do not vary in time and the dynamics of the system are tangible through conceptual model development.

The scenarios given in Table 2 are used to select the combination of

**Table 2**  
Input and output combinations for FMRVR, SVR and MLR models.

Algorithm	Inputs	Outputs
FMRVR	S1: $S_t, P_t, E_t, G_{t-1}, L_{t-1}$	$G_t, L_t$
	S2: $S_t, S_{t-1}P_t, P_{t-1}, E_t, E_{t-1}, G_{t-1}, L_{t-1}$	$G_t, L_t$
SVR	S1: $S_t, P_t, E_t, G_{t-1}, L_{t-1}$	$G_t$
	S1: $S_t, P_t, E_t, G_{t-1}, L_{t-1}$	$L_t$
	S2: $S_t, S_{t-1}P_t, P_{t-1}, E_t, E_{t-1}, G_{t-1}, L_{t-1}$	$G_t$
	S2: $S_t, S_{t-1}P_t, P_{t-1}, E_t, E_{t-1}, G_{t-1}, L_{t-1}$	$L_t$
MLR	S1: $S_t, P_t, E_t, G_{t-1}, L_{t-1}$	$G_t$
	S1: $S_t, P_t, E_t, G_{t-1}, L_{t-1}$	$L_t$
	S2: $S_t, S_{t-1}P_t, P_{t-1}, E_t, E_{t-1}, G_{t-1}, L_{t-1}$	$G_t$
	S2: $S_t, S_{t-1}P_t, P_{t-1}, E_t, E_{t-1}, G_{t-1}, L_{t-1}$	$L_t$



**Fig. 2.** Selected meteorological station for estimating precipitation, and together with the allocated Thiessen polygons to acquire the share of each station. The red values given for each station represent the coverage area of the Thiessen polygons and the share of each station in the total precipitation/evaporation time series.

spatial and/or temporal predictors needed to develop the best model. Two scenarios are used to represent the FMRVR models with multiple outputs, while the same scenarios are examined by 4 scenarios with a single-output approach to define a comprehensive investigation in terms of the role of each predictor on the lake water depth in time and space. The first scenarios (S1) are selected based on the primarily study of [Vaheddoost and Aksoy \(2021\)](#) indicating that the effect of monthly  $S$ ,  $P$ , and  $E$  on  $L$  decays very fast, but the effect of  $G$  persist for at least one month. Alternatively, the second scenarios (S2) are used to define a one-month lag for each variable, to let the method decide if it is necessary or not. Then 80% of data (383 months out of 479 months) are used for model training, while the remaining 20% of the data (96 months out of 479 months) are used to test the performance of the models. To this end, [Fig. 3](#) represents the core of the relationship between inputs and output in which the groundwater can play either as an input or output based on the followed approach.

### 2.3. Fast multi-output relevance vector regression

Let us consider a set of  $N$  input-output data pairs  $\{\mathbf{x}_i \in \mathbb{R}^{d \times 1}, \mathbf{y}_i \in \mathbb{R}^{1 \times V}\}_{i=1}^N$ . In the multi-output regression, the responses  $\mathbf{y}_i$ 's, are assumed to be observations from the model  $\mathbf{z}(\mathbf{x}_i; \mathbf{G})$  plus some noise, i.e.

$$\mathbf{y}_i = \mathbf{z}(\mathbf{x}_i; \mathbf{G}) + \varepsilon_i \quad (1)$$

where  $\mathbf{G} \in \mathbb{R}^{(N+1) \times V}$  denotes the coefficient matrix and  $\varepsilon_i \in \mathbb{R}^{1 \times V}$  stands for independent observations from a zero-mean Gaussian process with

$$p(\mathbf{G} | \mathbf{Y}, \alpha, \mathbf{C}) = 2\pi \left( 2\pi \right)^{-\frac{V(N+1)}{2}} \left| \mathbf{C} \right|^{-\frac{N+1}{2}} \left| \boldsymbol{\Sigma} \right|^{-\frac{V}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{C}^{-1} (\mathbf{G} - \mathbf{M})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{G} - \mathbf{M}) \right\} \quad (6)$$

the covariance matrix  $\mathbf{C} \in \mathbb{R}^{V \times V}$ .

Using matrix algebra, the equation above can be expressed as

$$\mathbf{Y} = \mathbf{\Phi} \mathbf{G} + \mathbf{E} \quad (2)$$

where  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^\top \in \mathbb{R}^{N \times V}$  denotes the response and,  $\mathbf{E} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N]^\top \in \mathbb{R}^{N \times V}$  stands for the noise,  $\mathbf{\Phi} = [\mathbf{\Phi}(\mathbf{x}_1), \mathbf{\Phi}(\mathbf{x}_2), \dots, \mathbf{\Phi}(\mathbf{x}_N)]^\top \in \mathbb{R}^{N \times (N+1)}$  represents a design matrix,  $\mathbf{\Phi}(\mathbf{x}) = [1, k(\mathbf{x}, \mathbf{x}_1), k(\mathbf{x}, \mathbf{x}_2), \dots, k(\mathbf{x}, \mathbf{x}_N)]^\top \in \mathbb{R}^{(N+1) \times 1}$ , and  $k(\mathbf{x}, \mathbf{x}')$  denotes

a kernel function associated with the non-linear basis function  $\mathbf{\Phi}(\cdot)$ . The goal in the multi-output regression is to determine the coefficient matrix  $\mathbf{G}$  for a given basis function.

Assuming a matrix Gaussian distribution, the likelihood of the observations may be expressed as

$$p(\mathbf{Y} | \mathbf{G}, \mathbf{C}) = 2\pi \left( 2\pi \right)^{-\frac{VN}{2}} \left| \mathbf{C} \right|^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{\Phi} \mathbf{G})^\top (\mathbf{Y} - \mathbf{\Phi} \mathbf{G}) \right\} \quad (3)$$

where  $\mathbf{C} = \frac{\mathbb{E}[\mathbf{E}^\top \mathbf{E}]}{N}$ , and  $\text{tr}(\cdot)$  denotes matrix trace.

In the FMRVR approach ([Ha and Zhang, 2019](#)), the assumption made to prevent overfitting in the estimation process of the coefficient matrix  $\mathbf{G}$  is

$$p(\mathbf{G} | \alpha, \mathbf{C}) = (2\pi)^{-\frac{V(N+1)}{2}} \left| \mathbf{C} \right|^{-\frac{N+1}{2}} \left| \mathbf{A} \right|^{\frac{V}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (\mathbf{C}^{-1} \mathbf{G}^\top \mathbf{A} \mathbf{G}) \right\} \quad (4)$$

where  $\mathbf{A}^{-1} = \text{diag}(\alpha_0^{-1}, \alpha_1^{-1}, \dots, \alpha_N^{-1}) = \frac{\mathbb{E}(\mathbf{G} \mathbf{G}^\top)}{\text{tr}(\mathbf{C})}$  which corresponds to the assumption of a zero-mean Gaussian prior distribution for  $\mathbf{G}$  with variances  $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_n]^\top \in \mathbb{R}^{(N+1) \times 1}$ , that give rise to  $N + 1$  hyper-parameters in the FMRVR model.

Using the Bayes' theorem ([Joyce, 2003](#)) and the fact that  $p(\mathbf{Y} | \mathbf{G}, \alpha, \mathbf{C}) = p(\mathbf{Y} | \mathbf{G}, \mathbf{C})$ , the posterior probability of  $\mathbf{G}$  may be written as

$$p(\mathbf{G} | \mathbf{Y}, \alpha, \mathbf{C}) = \frac{p(\mathbf{Y} | \mathbf{G}, \mathbf{C}) p(\mathbf{G} | \alpha, \mathbf{C})}{p(\mathbf{Y} | \alpha, \mathbf{C})} \quad (5)$$

which based on the assumption of Gaussian distribution is given as

where the covariance  $\mathbf{C}$  matrix and the mean  $\mathbf{M}$  vector are:

$$\boldsymbol{\Sigma} = (\mathbf{\Phi}^\top \mathbf{\Phi} + \mathbf{A})^{-1} \quad (7)$$

$$\mathbf{M} = \boldsymbol{\Sigma} \mathbf{\Phi}^\top \mathbf{Y} \quad (8)$$

A uniformity assumption for the hyper-parameters  $\alpha$  and  $\mathbf{C}$  makes maximising a posteriori probability  $p(\alpha, \mathbf{C} | \mathbf{Y}) \propto p(\mathbf{Y} | \alpha, \mathbf{C}) p(\alpha) p(\mathbf{C})$  equivalent to the maximisation of the likelihood  $p(\mathbf{Y} | \alpha, \mathbf{C})$  defined as

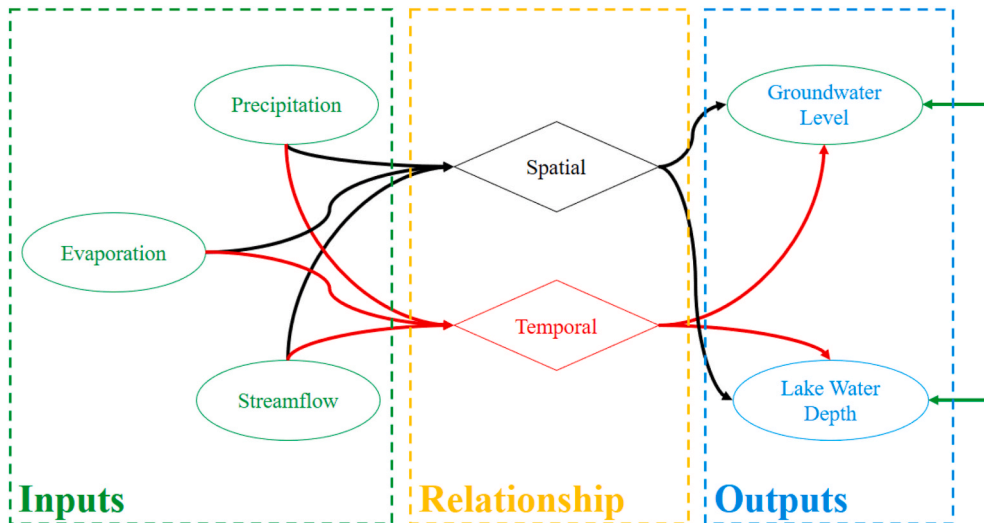


Fig. 3. The core of the relationship between input and output variables.

$$p(\mathbf{Y}|\boldsymbol{\alpha}, \mathbf{C}) = 2\pi \left(2\pi\right)^{-\frac{M}{2}} \left|\mathbf{C}\right|^{-\frac{M}{2}} \left|\mathbf{I} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^\top\right|^{-\frac{M}{2}} \exp\left[-\frac{1}{2}\text{tr}\mathbf{C}^{-1}\mathbf{Y}^\top\left(\mathbf{I} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^\top\right)^{-1}\mathbf{Y}\right] \quad (9)$$

In order to learn the FMRVR model, an expectation maximisation (EM) algorithm (Dempster et al., 1977) is applied to maximise the likelihood.

### 2.3.1. Making predictions

Using the FMRVR, for a new input vector  $\mathbf{x}_* \in \mathbf{R}^{d \times 1}$  one may estimate both a mean vector  $\mathbf{z}_* \in \mathbf{R}^{1 \times V}$  and a covariance matrix  $\mathbf{C}_* \in \mathbf{R}^{V \times V}$  using the most probable (MP) hyper-parameters  $\boldsymbol{\alpha}_{MP} \in \mathbf{R}_{>0}^{M \times 1}$  and  $\mathbf{C}_{MP} \in \mathbf{R}^{V \times V}$ . The distribution of  $\mathbf{y}_*$  is then a Gaussian distribution as

$$p(\mathbf{y}_*|\mathbf{Y}, \boldsymbol{\alpha}_{MP}, \mathbf{C}_{MP}) = \mathcal{N}(\mathbf{y}_*|\mathbf{z}_*, \mathbf{C}_*), \quad (10)$$

with

$$\mathbf{z}_* = \boldsymbol{\Phi}(\mathbf{x}_*)^\top \mathbf{M} \quad (11)$$

and

$$\mathbf{C}_* = \mathbf{C}_{MP}(\mathbf{I} + \boldsymbol{\Phi}(\mathbf{x}_*)^\top \boldsymbol{\Sigma} \boldsymbol{\Phi}(\mathbf{x}_*)) \quad (12)$$

where  $\boldsymbol{\Phi}(\mathbf{x}) \in \mathbf{R}^{M \times 1}$  relies only on  $M$  basis functions which are incorporated into the model during the training process.

## 2.4. Benchmarks

In order to effectively gauge any advantages offered by a multi-output modelling approach (i.e. the FMRVR method), we examine two widely used methods for regression analysis, namely, the Support Vector Regression (SVR) (Drucker et al., 1997) and the Multi-Linear Regression analysis (Alpaydin, 2010), discussed next.

### 2.4.1. Support vector regression

Support vector regression (SVR) (Drucker et al., 1997) is a variant of the Support Vector Machine classifier (Cortes and Vapnik, 1995) that is designed and widely used for regression analysis (Pesantez et al., 2020; Safari and Rahimzadeh Arashloo, 2021). The prediction for a sample  $\mathbf{x}$  in the SVR model is obtained as

$$y = \mathbf{G}^\top \boldsymbol{\varphi}(\mathbf{x}) + b \quad (13)$$

where  $\mathbf{G}$  denotes the model's coefficient vector;  $\boldsymbol{\varphi}(\mathbf{x})$  is the mapped input into the feature space and  $b$  is the bias term.

The SVR model is learned via a constrained minimisation problem as

$$\min_{\mathbf{G}} \frac{1}{2} \mathbf{G}^\top \mathbf{G} + k \sum_j (\zeta_j^* + \zeta_j) \quad (14)$$

$$\text{s.t. } y_j - \mathbf{G}^\top \boldsymbol{\varphi}(\mathbf{x}) - b \leq \varepsilon + \zeta_j \quad (15)$$

$$\mathbf{G}^\top \boldsymbol{\varphi}(\mathbf{x}) + b - y_j \leq \varepsilon + \zeta_j^* \quad (16)$$

$$\zeta_j, \zeta_j^* \geq 0 \quad (17)$$

where  $k$  stands for the penalty coefficient,  $\zeta_j, \zeta_j^*$  denote the empirical risk and  $\varepsilon$  is the band area width. In order to solve the constrained problem above, the corresponding Lagrangian is considered.

In the experiments, we used a Gaussian (RBF) kernel function, i.e.  $k(\mathbf{x}, \mathbf{x}') = e^{-\gamma \|\mathbf{x} - \mathbf{x}'\|^2}$  where  $\gamma$  is set to the reciprocal of the average pair-

wise Euclidean distance between all training samples. For the FMRVR algorithm, the maximum number of iterations, and the tolerance value to check convergence of the algorithm are set to default values of the method, i.e. 1000 iterations and 0.1 for the tolerance value. Regarding the SVR, we used the widely used LIBSVM implementation of epsilon-SVR where epsilon set to its default value 0.1.

### 2.4.2. Multi-linear regression

One of the baseline techniques in regression analysis is that of linear regression (Alpaydin, 2010). As the name implies, in this method, a linear relationship between the input  $\mathbf{x}$  and the output  $y$  is considered:

$$y = \mathbf{G}^\top \mathbf{x} \quad (18)$$

where  $\mathbf{G}$  denotes model coefficients. The linear regression model is learned by minimising a sum of squared errors over the training samples:

$$\min_{\mathbf{G}} \sum_j (y_j - \mathbf{G}^\top \mathbf{x}_j)^2 \quad (19)$$

The optimal solution to the problem above is then given as

$$\mathbf{G} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad (20)$$

where  $\mathbf{X}$  is a matrix collection of training samples and  $\mathbf{y}$  is the corresponding vector responses.

## 2.5. Performance metrics

Both quantitative and visual performance metrics are considered for assessing the proposed model's performance. The determination coefficient ( $R^2$ ), root mean square error (RMSE), mean absolute error (MAE) and Kling-Gupta efficiency (KGE) are used as quantitative performance metrics while time series and scatter plots of observed and estimated values are used for visual examination of the model's performance. The developed model provides near perfect performance having RMSE and MAE of zero and,  $R^2$  and KGE of unity. The performance metrics of  $R^2$ , RMSE, MAE and KGE and their ranges are given as follows (Legates and McCabe, 1999; Gupta et al., 2009; Moriasi et al., 2015; Safari and Rahimzadeh Arashloo, 2021; Malik et al., 2021a,b).

$$R^2 = \left[ \frac{\sum_{i=1}^n (x_i^o - \bar{x}^o)(x_i^e - \bar{x}^e)}{\sqrt{\sum_{i=1}^n (x_i^o - \bar{x}^o)^2} \sqrt{\sum_{i=1}^n (x_i^e - \bar{x}^e)^2}} \right]^2 \quad 0.0 \text{ to } 1.0 \quad (21)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i^o - x_i^e)^2}{n}} \quad 0.0 \text{ to } \infty \quad (22)$$

$$MAE = \frac{\sum_{i=1}^n |x_i^o - x_i^e|}{n} \quad 0.0 \text{ to } \infty \quad (23)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad 0.0 \text{ to } 1.0 \quad (24)$$

where  $x_i^o$ ,  $x_i^e$ ,  $\bar{x}^o$  and  $\bar{x}^e$  are the observed, estimated and mean of the observed and estimated values while  $n$  is the number of data items;  $r$  denotes the correlation coefficient and  $\alpha$  represents the ratio of standard deviation of the estimated to the standard deviation of the observed values.  $\beta$  corresponds to the ratio of mean of the estimated to the mean of the observed values.

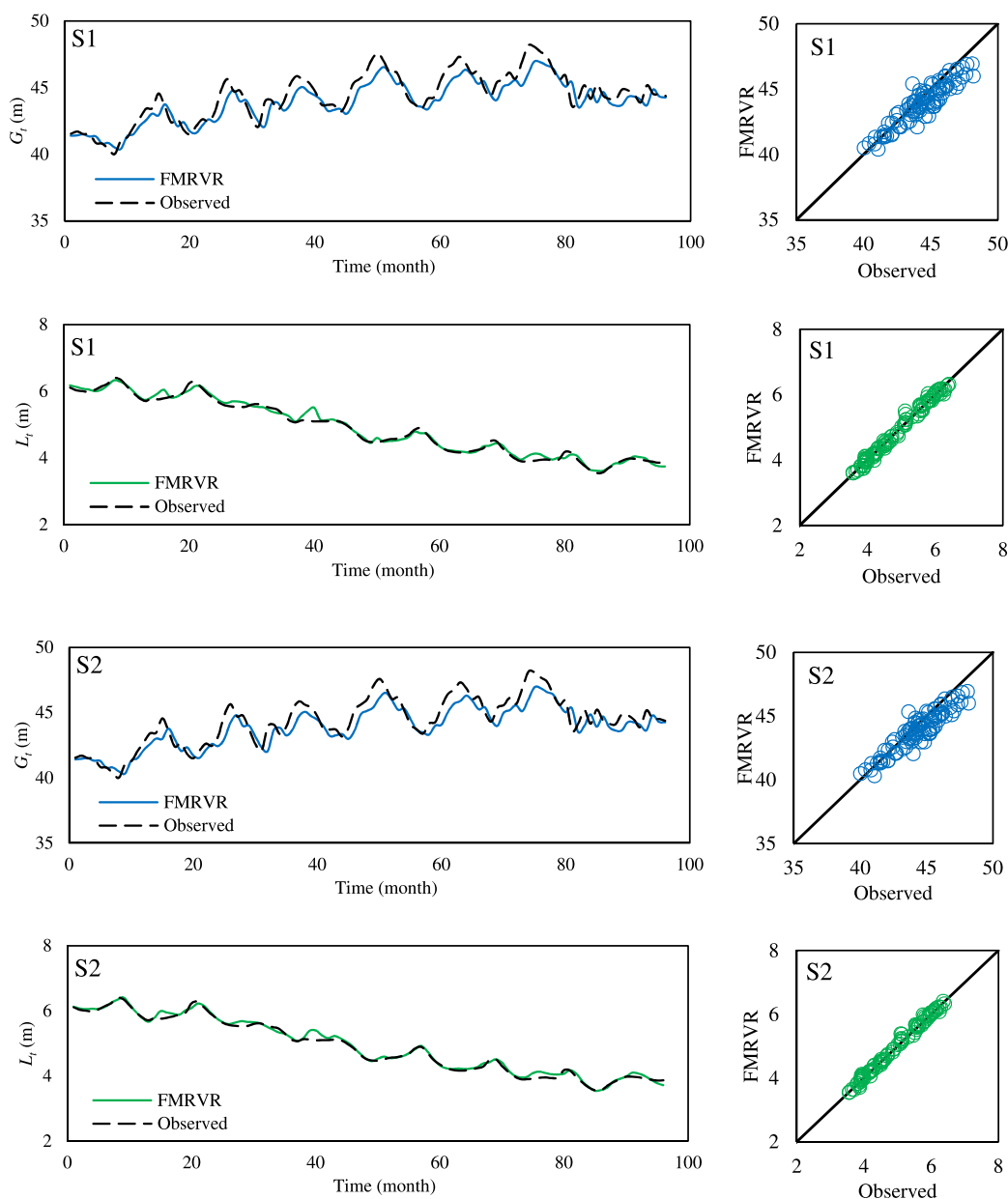
**Table 3**  
Performance of FMRVR, SVR and MLR models for groundwater and lake water depth modeling in terms of  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$ .

Algorithm	Scenario	Groundwater				Lake water depth			
		$R^2$	$RMSE$ (m)	$MAE$ (m)	$KGE$	$R^2$	$RMSE$ (m)	$MAE$ (m)	$KGE$
FMRVR	S1	0.850	0.851	0.673	0.833	0.987	0.098	0.074	0.992
	S2	0.856	0.857	0.675	0.832	0.992	0.083	0.059	0.992
SVR	S1	0.841	1.753	1.513	0.588	0.982	0.192	0.161	0.945
	S2	0.858	1.829	1.590	0.570	0.982	0.236	0.205	0.925
MLR	S1	0.824	1.714	1.534	0.788	0.981	0.468	0.444	0.851
	S2	0.839	1.723	1.558	0.793	0.988	0.477	0.457	0.849

**3. Results**

As explained in the previous section, two alternative input scenarios are considered for modeling purposes where for the S1 case  $S_t, P_t, E_t, G_{t-1}$ ; and  $L_{t-1}$ , and for S2,  $S_t, S_{t-1}, P_t, P_{t-1}, E_t, E_{t-1}, G_{t-1}$  and  $L_{t-1}$  are fed into the model as independent variables. For FMRVR as a multi-output

model, the groundwater ( $G_t$ ) and lake water depth ( $L_t$ ) are simultaneously estimated as model outputs while for the SVR and MLR approaches, the groundwater ( $G_t$ ) and lake water depth ( $L_t$ ) are modeled separately. In order to compare the performance of the developed models, the statistical performance criteria of  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$  are utilized. Table 3 reports the calculated  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$



**Fig. 4.** Performance of FMRVR model for groundwater and lake water depth computation.

values for the proposed multi-output approach based on FMRVR as well as the SVR and MLR techniques in groundwater and lake water depth modeling.

The results presented in Table 3 show that a joint modeling of multiple relevant outputs based on the FMRVR gives  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$  of 0.850, 0.851, 0.673 and 0.833, respectively, demonstrating its superior performance compared to the single output SVR and MLR techniques in groundwater modeling for the S1 scenario. The developed SVR and MLR models provide almost similar results in terms of different performance metrics. Similar outcomes can be seen for the S2 scenario for groundwater modeling where the multi-output model provides a better performance in comparison to the SVR and MLR techniques having  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$  of 0.856, 0.857, 0.675 and 0.832, respectively. Although the performances of all models are slightly better for the S2 scenarios, there are no significant differences between the two scenarios in groundwater modeling.

An examination of different models' performances for lake water

depth modeling illustrates that all the developed models perform well for both the S1 and the S2 scenarios in terms of  $R^2$  and  $KGE$  as may be observed from Table 3. However, a detailed analysis reveals that the multi-output model is more accurate providing  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$  of 0.987, 0.098, 0.074 and 0.992, respectively for the S1 scenario and  $R^2$ ,  $RMSE$ ,  $MAE$  and  $KGE$  of 0.992, 0.083, 0.059 and 0.992, respectively for the S2 scenario for lake water depth modeling. In terms of  $R^2$ , the SVR and MLR approaches provide almost similar results. By examining Table 3, it can be seen that the SVR model has lower  $RMSE$  and  $MAE$  values compared to the MLR approach, confirming the superiority of SVR to MLR model in lake water depth modeling.

For a visual evaluation of the developed models' performances, a comparison of the observed and estimated groundwater and lake water depth values of the S1 and S2 scenarios for the FMRVR, SVR and MLR models in the form of time series and scatter plots are depicted in Figs. 4-6. Such an analysis is helpful to understand the models under- and over-estimation performances which are considered to be an

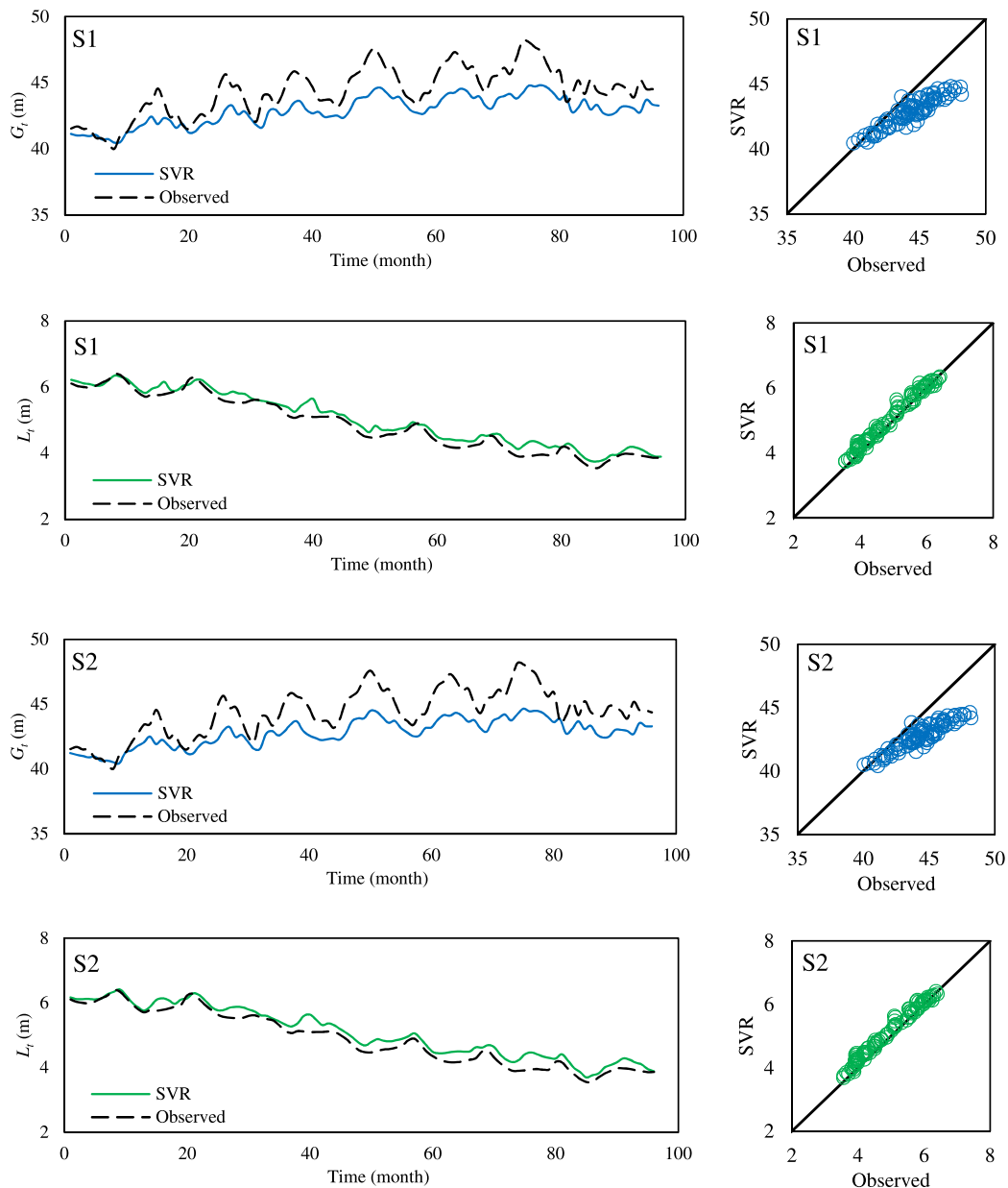


Fig. 5. Performance of SVR model for groundwater and lake water depth computation.



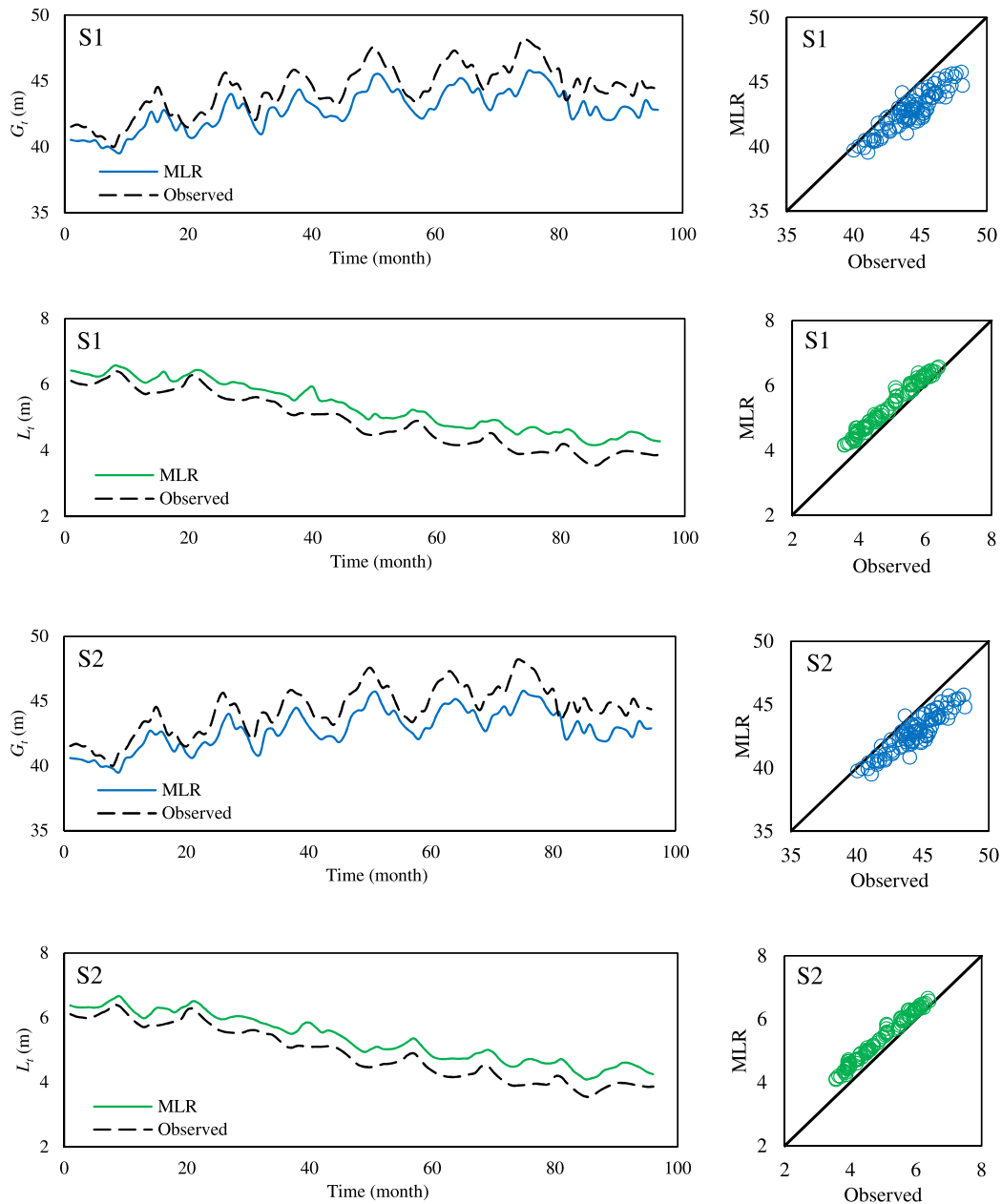


Fig. 6. Performance of MLR model for groundwater and lake water depth computation.

essential criterion for model examination in hydrological modeling. On the other hand, through visual inspection of the results, a model's behavior for capturing the peak values may be also investigated.

By inspecting Fig. 4, regarding the performance of the proposed multi-output model, in the corresponding time series plots it can be seen that for groundwater modeling for both the S1 and S2 scenarios, FMRVR values tend to fit the observed curves relatively well which demonstrates its high capability for groundwater modeling. It should be pointed out that although FMRVR outcomes are quite close to the observed values, some slight underestimations are observed for groundwater modeling. Scatter plots support the findings obtained in time series plots where predictions made by the multi-output algorithm are quite close to the observed true values with a slight underestimation. The performance of the FMRVR technique in lake water depth modeling is near to perfect where for both scenarios the estimated values fit the observed curves in the time series plots very well which is much better for the S2 scenario.

The cloud of the data almost falls on the best fit lines in scatter plots showing the superior performance of FMRVR for lake water depth modeling purposes.

Fig. 5 shows that the SVR model provides acceptable performance for lake water depth estimation. However, its estimations for groundwater modeling are not favorable where a significant under-estimation is observed in groundwater modeling for both the S1 and S2 scenarios. Although SVR can detect the general trend of the data, but it fails to provide accurate predictions especially for peak groundwater values. For lake water depth modeling, SVR provides acceptable performance. Nevertheless, a slight over-estimation is noticeable which is more tangible for the S1 scenario. It is seen in Fig. 6 that the MLR model provides poor performances for both groundwater and lake water depth modeling. Noticeable under-estimation and over-estimation are seen for groundwater and lake water depth modeling for both the S1 and S2 scenarios for the MLR technique.

Summarizing the findings discussed above, it can be said that the multi-output modeling results based on FMRVR have higher  $R^2$  and  $KGE$  and, lower  $RMSE$  and  $MAE$  values in contrast to the single-output SVR and MLR models. While SVR and MLR fail to generate accurate results having noticeable under-estimation in groundwater and over-estimation in lake water depth modeling, the multi-output FMRVR model predictions are in agreement with their corresponding observed values. Although the SVR and MLR models demonstrate poor performances in detecting peak values, FMRVR can appropriately model the local groundwater and lake water depth peak values. Consequently, it can be concluded that the multi-output FMRVR approach is superior to both the SVR and MLR models in terms of quantitative as well as visual performance criteria for groundwater and lake water depth modeling.

#### 4. Discussion

The importance of studying Lake Urmia behavior with respect to different hydro-meteorological variables comes from the fact that alteration in ecological and hydrological parameters in the basin can alter the biota and the socio-economy of the region. In this respect, the explicit nature of water-loss and -gain between groundwater and lake water body limits the performance of the single-output models developed either for estimating the behavior of groundwater or for lake water depth estimation. It cannot be overemphasized that most of the previous studies have only modeled lake water depth neglecting the groundwater on the lake system behavior while only a few considered a possible interaction between the groundwater and lake water depth parameters such as [Vaheddoost and Aksoy \(2018\)](#), [Javadzadeh et al. \(2020\)](#), [Kozekalani Sales et al. \(2021\)](#). In contrast, this study addressed the aforementioned point by considering groundwater and lake water depth as outputs of a joint model where the correlations are explicitly captured via a non-diagonal covariance matrix.

Typically, the hydrological models developed for lake systems in the literature mainly focus on the estimation of a single parameter, namely the lake water depth. This study proposes FMRVR as a multi-output model for simultaneous estimation of groundwater and lake water depth considering two different scenarios that incorporate the role of precipitation, streamflow and evaporation in time and space. Based on different quantitative and visual statistical performance criteria, the multi-output FMRVR model is found superior in contrast to the SVR and MLR for groundwater and lake water depth modeling purposes in the applied scenarios. A brief literature review confirms the superior performance of SVR in comparison to the particle swarm optimization-artificial neural networks (PSO-ANN), radial basis neural networks (RBNN), multi-layer artificial neural networks (MLP) and adaptive network-based fuzzy inference system (ANFIS) as reported by [Buyukildiz et al. \(2014\)](#) for lake water depth modeling. Based on the results obtained in this study, as FMRVR outperforms the well-known SVR benchmark in groundwater and lake water depth modeling, it may be concluded that FMRVR may serve as a more reliable choice for modeling lake system behavior.

Considering the conceptualization of the hydrological water budget in the form of multiple-output model, it can be concluded that a joint estimation of the groundwater and lake water depth is the outcome of the implicit nature of the water loss and gain in the Lake Urmia. Therefore, in accordance with the results obtained by [Ashraf et al. \(2017\)](#), [Vaheddoost and Aksoy \(2018\)](#), [Javadzadeh et al. \(2020\)](#) and [Kozekalani Sales et al. \(2021\)](#) it is concluded that neglecting groundwater effect in the analysis of lake water level/depth in Lake Urmia would lead to a propagation of bias in the suggested models.

The majority of the existing machine learning approaches for hydrological modeling apply single-output regression analysis techniques whereas the FMRVR is formulated as a multi-output method. The advantages offered by such a formulation lies in the capability of the approach to deal with problems where multiple outputs are influenced by a joint input. In this context, a joint model is beneficial from a

computational complexity perspective. More importantly, a joint modelling of multiple outputs in the FMRVR enables capturing any dependencies which might exist between the outputs. In the FMRVR approach, this is realized through a probabilistic modeling formalism and by considering Gaussian distributions for the output variables where the covariance matrix may be non-diagonal and thus captures the possible correlations that might exist between output parameters.

#### 5. Conclusions

In this study, groundwater and lake water depth are modeled through considering different hydro-meteorological parameters of precipitation, streamflow and evaporation. Fast multi-output relevance vector regression (FMRVR) algorithm is applied for simultaneous modeling of groundwater and lake water depth for the first time in the relevant literature. FMRVR results are compared against the well-known support vector regression (SVR) and multi-linear regression (MLR) benchmarks. An examination of the models' accuracies based on quantitative and visual performance criteria showed that a multi-output regression model where the correlations between multiple outputs is modeled (i.e. the FMRVR approach) is able to provide better predictions without significant under- and overestimation. The single-output SVR and MLR methods, on the other hand, fail to make accurate predictions and tend to significantly under-estimate and/or over-estimate groundwater and lake water depth in different scenarios. Since the FMRVR approach as a multi-output model, it performs better than single-output models including the SVR and MLR methods. Thus, it may be concluded that the groundwater and lake water depth parameters tend to be correlated variables. This dependency is well captured by the FMRVR approach through a probabilistic modeling mechanism while the single output techniques fail to benefit from any existing correlations among multiple outputs. The promising findings obtained in this study seem to be useful and applicable in evolutionary hydrological modeling to construct models with higher number of output parameters. One limitation of this study from a hydrological perspective is the possible change in the seasonal behavior of the system where droughts, climate-variability, climate-change, and groundwater depletion may take place. Therefore, more sophisticated models may lump the effect of climate-born events as well as anthropogenic changes in the models. The limitation of the FMRVR approach, however, is that of making a simple Gaussianity assumption in its formulation. Although such an assumption tends to provide reasonable performances in practice, a better probabilistic modeling approach, such as using a mixture of Gaussian distributions, may lead to an even better performance. A development of more sophisticated models considering droughts, climate-variability and climate-change are considered as future research directions. The successful application of the joint modeling approach for modeling multiple related hydrological phenomena in this study, may draw a projection for future studies in evolutionary environmental modeling.

#### 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.

#### References

- [Abbaspour, M., Javid, A.H., Mirbagheri, S.A., 2012. Investigation of lake drying attributed to climate change. Int. J. Environ. Sci. Technol. 9, 257–266.](#)
- [Alpaydin, E., 2010. Introduction to Machine Learning, 2nd. ed. The MIT Press.](#)
- [Arkan, F., Nicholson, S.E., Ziaie, B., 2016. Meteorological factors affecting the sudden decline in Lake Urmia's water level. Theor. Appl. Climatol. 131, 641–651.](#)
- [Amirataee, B., Zeinalzadeh, K., 2016. Trends analysis of quantitative and qualitative changes in groundwater with considering the autocorrelation coefficients in west of Lake Urmia, Iran. Environ. Earth Sci. 75 \(5\), 371.](#)
- [Amiri, V., Nakhaei, M., Lak, R., Kholghi, M., 2016a. Geophysical, isotopic, and hydrogeochemical tools to identify potential impacts on coastal groundwater](#)

- resources from Urmia hypersaline Lake, NW Iran. *Environ. Sci. Pollut. Res.* 23 (16), 16738–16760.
- Amiri, V., Nakhaei, M., Lak, R., Kholghi, M., 2016b. Assessment of seasonal groundwater quality and potential saltwater intrusion: a study case in Urmia coastal aquifer (NW Iran) using the groundwater quality index (GQI) and hydrochemical facies evolution diagram (HFE-D). *Stoch. Environ. Res. Risk Assess.* 30 (5), 1473–1484.
- Ashraf, B., Agha Kouchak, A., Alizadeh, A., Baygi, M.M., Moftakhari, H.R., Mirchi, A., Anjileli, H., Madani, K., 2017. Quantifying anthropogenic stress on groundwater resources. *Sci. Rep.* 7 (1), 12910.
- Bonakdari, H., Ebtehaj, I., Samui, P., 2019. Lake water-level fluctuations forecasting using minimax probability machine regression, relevance vector machine, Gaussian process regression, and extreme learning machine. *Water Resour. Manag.* 33, 3965–3984.
- Buyukyildiz, M., Tezel, G., Yilmaz, V., 2014. Estimation of the change in lake water level by artificial intelligence methods. *Water Resour. Manag.* 28, 4747–4763.
- Chaudhari, S., Felfelani, F., Shin, S., Pokhrel, Y., 2018. Climate and anthropogenic contributions to the desiccation of the second largest saline lake in the twentieth century. *J. Hydrol.* 560, 342–353.
- Çimen, M., Kisi, O., 2009. Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *J. Hydrol.* 378 (3–4), 253–262.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297.
- Dehghanipour, A.H., Moshir Panahi, D., Mousavi, H., Kalantari, Z., Tajrishy, M., 2020. Effects of water level cline in lake Urmia, Iran, on local climate conditions. *Water* 12 (8), 2153.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Series B Stat. Methodol.* 39 (1), 1–22.
- Drucker, H., Burges, C.J., Kaufman, L., Smola, A., Vapnik, V., 1997. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* 9, 155–161.
- Ghorbani, M.A., Deo, R.C., Karimi, V., Yaseen, Z.M., Terzi, O., 2018. Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stochastic Stoch Environ. Res. Risk Assess.* 32 (6), 1683–1697.
- Gorgij, A.D., Kisi, O., Moghaddam, A.A., 2016. Groundwater budget forecasting, using hybrid wavelet-ANN-GP modelling: a case study of Azarshahr Plain, East Azerbaijan, Iran. *Hydrol Res* 48 (1), 202–224.
- 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 (1–2), 80–91.
- Ha, Y., Zhang, H., 2019. Fast multi-output relevance vector regression. *Econ. Modell.* 81, 217–230.
- Hassanzadeh, E., Zarghami, M., Hassanzadeh, Y., 2012. Determining the main factors in declining the Lake Urmia by using system dynamics modeling. *Water Resour. Manag.* 26, 129–145.
- Javadzadeh, H., Ataie-Ashtiani, B., Hosseini, S.M., Simmons, C.T., 2020. Interaction of lake-groundwater levels using cross-correlation analysis: a case study of Lake Urmia Basin, Iran. *Sci. Total Environ.* 729, 138822.
- Jeihouni, M., Toomanian, A., Alavipanah, S.K., Hamzeh, S., 2017. Quantitative assessment of Urmia Lake water using spaceborne multisensor data and 3D modeling. *Environ. Monit. Assess.* 189 (11), 572.
- Joyce, J., 2003. Bayes' theorem. In: Zalta, Edward N. (Ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.
- Khazaei, B., Khatami, S., Alemohammad, S.H., Rashidi, L., Wu, C., Madani, K., Kalantari, Z., Destouni, G., AghaKouchak, A., 2019. Climatic or regionally induced by humans? Tracing hydro-climatic and land-use changes to better understand the Lake Urmia tragedy. *J. Hydrol.* 569, 203–217.
- Kozekalani Sales, A., Gul, E., Safari, M. J.S., Ghodrati Gharehbagh, & H., Vaheddoost, B., 2021. Urmia lake water depth modeling using extreme learning machine-improved grey wolf optimizer hybrid algorithm. *Theoretical and Applied Climatology* 146, 833–849. <https://doi.org/10.1007/s00704-021-03771-1>. <https://link.springer.com/article/10.1007/s00704-021-03771-1#citeas>.
- Kumar, M., Kumari, A., Kushwaha, D.P., Kumar, P., Malik, A., Ali, R., Kuriqi, A., 2020. Estimation of daily stage–discharge relationship by using data-driven techniques of a perennial river, India. *Sustainability* 12 (19), 7877.
- Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* 35 (1), 233–241.
- Long, Y., Tang, R., Wu, C., Jiang, C., Hu, S., 2019. Estimating real-time water area of Dongting lake using water level information. *Water* 11, 1240.
- Marjani, A., Jamali, M., 2014. Role of exchange flow in salt water balance of Urmia Lake. *Dynam. Atmos. Oceans* 65, 1–16.
- Malik, A., Tikhomarine, Y., Sammen, S.S., Abba, S.I., Shahid, S., 2021a. Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. *Environ. Sci. Pollut. Res.* 28 (29), 39139–39158.
- Malik, A., Tikhomarine, Y., Souag-Gamane, D., Rai, P., Sammen, S.S., Kisi, O., 2021b. Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. *Meteorol. Atmos. Phys.* 133 (3), 891–909.
- Moriassi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: performance measures and evaluation criteria. *Trans. ASABE (Am. Soc. Agric. Biol. Eng.)* 58 (6), 1763–1785.
- Nourani, V., Mousavi, S., 2016. Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method. *J. Hydrol.* 536, 10–25.
- Pandey, K., Kumar, S., Malik, A., Kuriqi, A., 2020. Artificial neural network optimized with a genetic algorithm for seasonal groundwater table depth prediction in Uttar Pradesh, India. *Sustainability* 12 (21), 8932.
- Pengra, B., 2012. The drying of Iran's lake Urmia and its environmental consequences. UNEP-GRID, Sioux falls, UNEP global environmental alert service (GEAS). *Environ. Dev.* 2 (1), 128–137.
- Pesantez, J.E., Berglund, E.Z., Kaza, N., 2020. Smart meters data for modeling and forecasting water demand at the user-level. *Environ. Model. Software* 125, 104633.
- Rabet, A., Dastranj, A., Asadi, S., Asadi Nalivan, O., 2020. Determination of groundwater potential using artificial neural network, Random forest, support vector machine and linear regression models (Case study: lake Urmia watershed). *Iranian J. Ecohydrol.* 7 (4), 1047–1060.
- Radman, A., Akhoondzadeh, M., Hosseini, B., 2021. Integrating InSAR and deep-learning for modeling and predicting subsidence over the adjacent area of Lake Urmia, Iran. *GIScience Remote Sens.* 58 (8), 1413–1433.
- Rehannia, I., Benlaoukli, B., Jamei, M., Karbasi, M., Malik, A., 2021. Simulation of seepage flow through embankment dam by using a novel extended Kalman filter based neural network paradigm: case study of Fontaine Gazelles Dam, Algeria. *Measurement* 176, 109219.
- Safari, M.J.S., 2020. Hybridization of multivariate adaptive regression splines and random forest models with an empirical equation for sediment deposition prediction in open channel flow. *J. Hydrol.* 590, 25392 <https://doi.org/10.1016/j.jhydrol.2020.125392>. <https://doi.org/10.1016/j.jhydrol.2020.125392>.
- Safari, M.J.S., Rahimzadeh Arashloo, S., 2021. Kernel ridge regression model for sediment transport in open channel flow. *Neural Comput. Appl.* 33, 11255–11271. <https://doi.org/10.1007/s00521-020-05571-6>. <https://doi.org/10.1007/s00521-020-05571-6>.
- Safari, M.J.S., Rahimzadeh Arashloo, S., 2021. Sparse kernel regression technique for self-cleansing channel design. *Adv. Eng. Inf.* 47, 101230 <https://doi.org/10.1016/j.aei.2020.101230>. <https://doi.org/10.1016/j.aei.2020.101230>.
- Safari, M.J.S., Rahimzadeh Arashloo, S., Danandeh Mehr, A., 2020. Rainfall-runoff modeling through regression in the reproducing kernel Hilbert space algorithm. *J. Hydrol.* 587, 125014 <https://doi.org/10.1016/j.jhydrol.2020.125014>. <https://doi.org/10.1016/j.jhydrol.2020.125014>.
- Shadkam, S., Ludwig, F., Van Oel, P., Kirmitt, C., Kabat, P., 2016. Impacts of climate change and water resources development on the declining inflow into Iran's Urmia Lake. *J. Great Lake Res.* 42 (5), 942–952.
- Sheibani, S., Ataie-Ashtiani, B., Safaie, A., Simmons, C.T., 2020. Influence of lakebed sediment deposit on the interaction of hypersaline lake and groundwater: a simplified case of lake Urmia, Iran. *J. Hydrol.* 588, 125110.
- Sima, S., Tajrishy, S., 2013. Using satellite data to extract volume–area–elevation relationships for Lake Urmia, Iran. *J. Great Lake Res.* 39, 90–99.
- Tong, X., Pan, H., Xie, H., Xu, X., Li, F., Chen, L., Luo, X., Liu, S., Chen, P., Jin, Y., 2016. Estimating water volume variations in Lake Victoria over the past 22 years using multi-mission altimetry and remotely sensed images. *Remote Sens. Environ.* 187, 400–413.
- Tourian, M.J., Elmi, O., Chen, Q., Devaraju, B., Roohi, S., Sneeuw, N., 2015. A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran. *Remote Sens. Environ.* 156, 349–360.
- Vaheddoost, B., Aksoy, H., 2021. Regressive-stochastic models for predicting water level in Lake Urmia. *Hydrol. Sci. J.* 66 (13), 1892–1906.
- Vaheddoost, B., Aksoy, H., 2019. Reconstruction of hydrometeorological data in lake Urmia basin by frequency domain analysis using additive decomposition. *Water Resour. Manag.* 33 (11), 3899–3911.
- Vaheddoost, B., Aksoy, H., 2018. Interaction of groundwater with lake Urmia in Iran. *Hydrol. Processes* 32 (21), 3283–3295.
- Vaheddoost, B., Aksoy, H., 2017. Structural characteristics of annual precipitation in lake Urmia basin. *Theor. Appl. Climatol.* 128 (3), 919–932.
- Yadav, B., Eliza, K., 2017. A hybrid wavelet-support vector machine model for prediction of lake water level fluctuations using hydro-meteorological data. *Measurement* 103, 294–301.
- Yaseen, Z.M., Naghshara, S., Salih, S.Q., Kim, S., Malik, A., Ghorbani, M.A., 2020. Lake water level modeling using newly developed hybrid data intelligence model. *Heor. Appl. Climatol.* 141 (3), 1285–1300.
- Zarghami, M., 2011. Effective watershed management; case study of Urmia Lake, Iran. *Lake Reservoir Manag.* 27 (1), 87–94.
- Zhang, G., Yao, T., Chen, W., Zheng, G., Shum, C.K., Yang, K., Piao, S., Sheng, Y., Yi, S., Li, J., O'Reilly, C.M., Qi, S., Shen, S.S.P., Zhang, H., Jia, Y., 2019. Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes. *Remote Sens. Environ.* 221, 386–404.