

## Research Article

# Simulation Using Machine Learning and Multiple Linear Regression in Hydraulic Engineering

M Poursaeid<sup>1\*</sup>, AH Pouraseid<sup>2</sup>, S Shabanlou<sup>3</sup>

Received: September 20, 2021

Accepted: September 18, 2022

Revised: September 17, 2022

Published online: December 22, 2023

1-Ph.D. in Civil Engineering, Deputy of Technical and Engineering, Plan and Budget Organization, Khorramabad, Lorestan, Iran.

2-Ph.D. Student, Faculty of Power Electrical Engineering, Lorestan University, Khorramabad, Lorestan, Iran.

3-Assoc. Prof., Faculty of Water Engineering, Islamic Azad University, Branch of Kermanshah, Iran.

\*Corresponding Author, Email: [mojtaba.poursaeid.ac@gmail.com](mailto:mojtaba.poursaeid.ac@gmail.com)

## Abstract

### Background and Objectives

Artificial intelligence models as powerful methods in modeling complex nonlinear problems have a significant ability, and this has been proven in numerous articles. Artificial intelligence has been used in various issues, including engineering, medicine, etc. The success of these methods compared to analytical and numerical methods, their easiness, speed, and accuracy. Today, one of the challenges of human life is the issues related to water resources management. This study has investigated the performance of artificial intelligence and regression models in water resources problems. Various researches have been done in the case of modeling and parametric analysis of water resources. However, this study used artificial intelligence (Learning Machine) models to simulate water's qualitative and quantitative parameters. The models used in this study are Self-Adapting Extreme Learning Machine (SAELM), Least Square Support Vector Machine (LSSVM), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Multiple Linear Regression (MLR). Due to the growing global population, one of the most critical challenges is access to safe drinking water. Due to its location in the semi-arid region and low rainfall in our country, Iran, this danger is felt more than ever. One of the serious issues is the salinity leakage into groundwater resources. In this study, an attempt has been made to simulate the leakage of salinity dynamic flow into the freshwater resources of the coastal aquifer, using artificial intelligence and statistical models.

### Methodology

The study area in this work is Mighan Wetland in Markazi province. Annual rainfall occurs in small amounts in this area. According to the statistical results, the maximum and minimum rainfall values range from 461 mm in the northeast to 208 mm in the center of the Arak plain. In this study, qualitative and quantitative parameters: water salinity, total dissolved solids (TDS), chloride ion (Cl), sampling time (t), electrical conductivity (EC), salinity, and groundwater level (GWL) were simulated. This work used ANFIS, LSSVM, SAELM, and MLR models for simulation. In this study, data from 173 months of sampling were used. 80% of the sample size was used for training and 20% for testing models. After simulation and obtaining the results, the Wilson Score method performed the uncertainty analysis without continuity correction. In this method, the prediction error ( $e_i$ ), the mean prediction error (Mean), and the standard deviation of the error are ( $S_e$ ). If the mean error value in predicting the target variable is positive, the model's performance is Over Estimated (OS). Also, if the average value of the model error is negative, the model's performance



is Under Estimated (US). Moreover, the Uncertainty Analysis results with a significance of 5% were obtained. Also, five approaches measured the performance of the models. The proposed approaches were: 1) Evaluation of prediction by accuracy chart, 2) Performance evaluation by mathematical indices, 3) Performance evaluation by Uncertainty Analysis by Wilson Score method without continuity correction, 4) Accuracy evaluation by error distribution charts, and 5) Performance evaluation by discrepancy rate (DR) charts. Finally, all the results are given at the end of each section, respectively.

## Findings

The simulation was performed using artificial intelligence and regression models. The simulation results showed higher accuracy of artificial intelligence models. The results showed that different models were successful in predicting water parameters. Approach 1- According to the prediction accuracy charts, 16 charts were drawn, and the most accurate models are depicted in Figures 4 to 7. After modeling, the results showed that the most accurate models in simulating groundwater parameters were the SAELM model in GWL simulation. According to the results, the SAELM model in GWL and EC simulation, LSSVM in TDS simulation, and MLR in Salinity simulation were superior models. Approach 2- According to the performance measurement indices, the results showed that the SAELM model was the best in simulating parameters (EC) and (GWL). Then the LSSVM model was also the most accurate in modeling the (TDS) parameter. MLR was the best model in (Salinity) parameter simulation. Approach 3- Uncertainty analysis was performed based on the Wilson score method. The performance of the models in the simulation showed that the SAELM model's performance was underestimated, and other superior models in the simulation had overestimated performance. Approach 4- The best accuracy was assigned to SAELM and MLR models based on the error distribution diagrams. Approach 5- Based on the DR diagrams, SAELM and MLR models were the most accurate models in the simulation.

## Conclusion

In all five performance approaches, the top models were introduced as follows: According to the performance measurement indices, each model successfully estimated one of the quantitative and qualitative parameters. Finally, study results showed that the SAELM was the best model for simulating EC and GWL parameters. It had the least computational error. Then LSSVM was also the most accurate model in estimating and modeling the TDS parameter. Also, MLR was the best model in Salinity parameter simulation. The uncertainty analysis results showed that the performance of the SAELM was underestimated, and other superior models were overestimated. The best accuracy was assigned to SAELM and MLR models based on the DR diagrams. SAELM and MLR models were the most accurate models based on the error distribution charts. Finally, it should be noted that based on the simulation results, the following predictions are possible:

- The graphs gradually showed more insignificant changes in estimating GWL changes by the SAELM model. However, these changes will continue their downward procedure.
- Based on the prediction diagrams, EC changes are almost like a time series and will continue similar to past behavior in the future changes. These changes will also occur in less time.
- Based on the estimation of the TDS changes, the LSSVM charts behave like time series and seem the future behavior is similar to past time series. The changes occur in the same range.
- Based on the estimation of Salinity changes, the MLR showed the changes of this parameter within a higher range. This tolerance occurred with higher values than before.

**Keywords:** ANFIS, Artificial intelligence, LSSVM, MLR, SAELM, Uncertainty analysis.

## مقاله پژوهشی

### شبیه‌سازی با استفاده از ماشین یادگیری و رگرسیون چندگانه خطی در مهندسی هیدرولیک

مجتبی پورسعید<sup>۱\*</sup>، امیرحسین پورسعید<sup>۲</sup>، سعید شعبانلو<sup>۳</sup>

تاریخ دریافت: ۱۴۰۰/۰۷/۲۹

تاریخ پذیرش: ۱۴۰۱/۰۷/۲۷

تاریخ ویرایش: ۱۴۰۱/۰۷/۲۶

تاریخ انتشار آنلاین: ۱۴۰۲/۱۰/۰۱

۱-دانش آموخته دکترای مهندسی عمران آب، معاونت امور نظام فنی و اجرایی، سازمان مدیریت و برنامه‌ریزی لرستان.

۲-دانشجوی دکترای مهندسی برق قدرت، دانشکده مهندسی برق، دانشگاه لرستان.

۳-دانشیار گروه مهندسی علوم آب، دانشگاه آزاد اسلامی واحد کرمانشاه.

\*مسئول مکاتبات، پست الکترونیکی: mojtaba.poursaeid.ac@gmail.com

#### چکیده

هوش مصنوعی یکی از روشهای ابداعی نسبتاً جدید در تحلیل و شبیه‌سازی پدیده‌های طبیعی است. مدل‌های هوش مصنوعی به عنوان روش‌های قدرتمند در مدل‌سازی مسائل غیرخطی و پیچیده توانایی قابل توجهی داشته است. تحقیقات مختلفی در شاخه مدل‌سازی و تحلیل پارامتریک منابع آب انجام گرفته است. ولی در این مطالعه از ۴ مدل هوش مصنوعی برای شبیه‌سازی کیفی و کمی آب دریاچه میقان واقع در شهرستان اراک استان مرکزی استفاده شد. مدل‌های بکار گرفته شده در این مطالعه عبارتند از: مدل ماشین یادگیری نیرومند خودتطبیق<sup>۱</sup> (SAELM)، مدل حداقل مربعات رگرسیون بردار پشتیبان<sup>۲</sup> (LSSVM)، مدل شبکه‌های عصبی-فازی<sup>۳</sup> (ANFIS) و مدل آماری رگرسیون خطی چندگانه<sup>۴</sup> (MLR) که برای پیش‌بینی تغییرات پارامترهای هیدروژئولوژیکی استفاده شد. در این مطالعه پارامترهای کل جامدات محلول<sup>۵</sup> (TDS)، هدایت الکتریکی<sup>۶</sup> (EC)، شوری<sup>۷</sup> و سطح آب زیرزمینی<sup>۸</sup> (GWL) شبیه‌سازی شدند. نتایج نشان داد که مدل‌های مختلف در زمینه پیش‌بینی منابع آب به نوبه خود موفق بودند. همچنین با توجه به آمار شاخص‌های سنجش عملکرد، مدل SAELM دارای بالاترین دقت در شبیه‌سازی دو پارامتر GWL و EC، مدل LSSVM بالاترین دقت را در شبیه‌سازی TDS و مدل MLR نیز در تخمین تغییرات پارامتر Salinity به عنوان بهترین مدل انتخاب گردید. جهت بررسی جامع دقت مدل‌ها برای مدل‌های برتر در شبیه‌سازی، با پنج رویکرد عملکرد مدل‌ها مورد سنجش قرار گرفت. رویکردهای مورد نظر عبارت بودند از: ۱) سنجش با نمودار پیش‌بینی ۲) سنجش عملکرد با شاخص‌های ریاضی ۳) سنجش عملکرد به روش عدم قطعیت ویلسون ۴) سنجش دقت با نمودارهای توزیع خطا و ۵) سنجش عملکرد با نمودارهای نرخ اختلاف خطا. در پایان کلیه نتایج به ترتیب در پایان هر قسمت آورده شده است.

واژه‌های کلیدی: ماشین یادگیری، عدم قطعیت، منابع آب، هدایت الکتریکی، کل جامدات محلول، کیفیت آب، هوش مصنوعی

<sup>1</sup> Self-Adaptive Extreme Learning Machine

<sup>2</sup> Least Square Support Vector Machine

<sup>3</sup> Adaptive Neuro-Fuzzy Inference System

<sup>4</sup> Multiple Linear Regression

<sup>5</sup> Total Dissolved Solids

<sup>6</sup> Electrical Conductivity

<sup>7</sup> Salinity

<sup>8</sup> Groundwater Level

## مقدمه

پیشرفت علوم مهندسی عامل بسیاری از تکنولوژی‌های امروزی است که رفاه انسان‌ها را بیش از پیش گردانیده است. الگوریتم‌های فراابتکاری، هوش مصنوعی، یادگیری ماشین و شبکه‌های عصبی و غیره از جمله پیشرفت‌های اخیر در عرصه علوم ریاضی و مهندسی هستند که بسیاری از چالش‌های علمی انسان را در حل مسائل پیش‌رو مرتفع نموده‌اند. در زمینه هوش مصنوعی تحقیقات متعددی صورت گرفته است و بعد از مختصری با محوریت روش‌های هوش مصنوعی، به تعدادی از تحقیقات گذشته که در این زمینه انجام گرفته است، اشاره می‌شود. بانرجی و همکاران (۲۰۱۱) در مطالعه خود به شبیه‌سازی انتقال جبهه آب شور به منابع زیرزمینی با استفاده از شبکه عصبی پرداختند. در این کار، مدل با داده‌های دو سال آموزش داده شد و بعد از آن به تخمین پارامترهای کیفی در ۵ سال بعد پرداخته شد. کاستا کوردوبا و همکاران (۲۰۱۴) جهت پیش‌بینی سری‌های زمانی، نسبت به شبیه‌سازی جبهه شور اقدام کردند. نتایج برتری مدل مبتنی بر سری زمانی را برای تخمین و پیش‌بینی شوری و میزان تبادل با منابع زیرزمینی آبخوان ساحلی، تایید نمود. وانگ و همکاران (۲۰۱۹) به بررسی توانایی مدل‌های یادگیری عمیق LSTM برای تخمین میزان سرایت آلاینده‌های نقطه‌ای به رودخانه‌ها پرداختند. نتایج این تحقیق، توانایی مدل یادگیری عمیق را در ردیابی دقیق منابع نقطه‌ای آلاینده‌ها را تأیید نمود. الکایرن و همکاران (۲۰۱۹) بر روی عملکرد مدل‌های شبکه‌عصبی برای شبیه‌سازی کیفی منابع آب پرداختند. سپس دو روش شبکه‌های عصبی-بازگشتی BPNN و سری‌های زمانی ARIMA برای مدل‌سازی کیفی انتخاب و مدل‌سازی کیفی منابع آب بر این اساس انجام شد. همچنین پورسعید و همکاران (۲۰۲۰) به بررسی توانایی و عملکرد مدل‌های هوش مصنوعی در زمینه پیش‌بینی میزان تبادل شوری جبهه آب شور به منابع آب زیرزمینی ساحلی در دریاچه میقان اراک در استان

مرکزی ایران پرداختند. آنها در این مقاله با توسعه مدل‌های ELM و OSELM و ترکیب با بهینه‌ساز موجک WT به مدل‌های هیبرید WELM و WOSELM دست یافتند. جامعی و همکاران (۲۰۲۰) روی شبیه‌سازی و مدل‌سازی کیفی منابع آب کار کردند. آنها با استفاده از مدل برنامه‌ریزی ژنتیک (GP)، نسبت به تخمین تغییرات پارامتر (TDS) منابع آب پرداختند. نتایج نشان داد که مدل‌های WMGGP و W-GEP دقت مناسبی در تخمین و مدل‌سازی TDS داشته‌اند. همچنین رجایی و همکاران (۲۰۲۰)، در یک مقاله مروری محققان به مطالعه ۱۳ مقاله مرتبط با انواع مختلف مدل‌های هوش مصنوعی بکارگیری شده جهت شبیه‌سازی کیفی آب پرداختند. کدخدازاده و همکاران (۲۰۲۱) به مدل‌سازی کیفی منابع آب پرداخته و الگوریتم بهینه‌ساز ژنتیک را با مدل LSSVM، ترکیب نموده در مدل هیبرید LSSVM-GBO حاصل گردید. سپس شبیه‌سازی پارامتر EC و TDS آب انجام گردید. نوردین و همکاران (۲۰۲۱) به بررسی کاربرد هوش مصنوعی در مطالعات در زمینه کیفیت منابع آب پرداختند و نهایتاً توصیه‌هایی جهت بهبود شبیه‌سازی ارائه شد. در این تحقیق برای اولین بار در دریاچه میقان استان مرکزی، با استفاده از مدل‌های معتبر هوش مصنوعی از زیرمجموعه‌های ماشین‌های یادگیری و نیز شبکه‌های عصبی-فازی، بطور همزمان همراه با مدل ریاضی-آماري، جهت مدل‌سازی استفاده گردید. همچنین با ۵ رویکرد عددی و نموداری، به بررسی عملکرد هر مدل در شبیه‌سازی پرداخته شده است. در نتیجه‌گیری کلی، عملکرد هر مدل بطور جداگانه تشریح شده است.

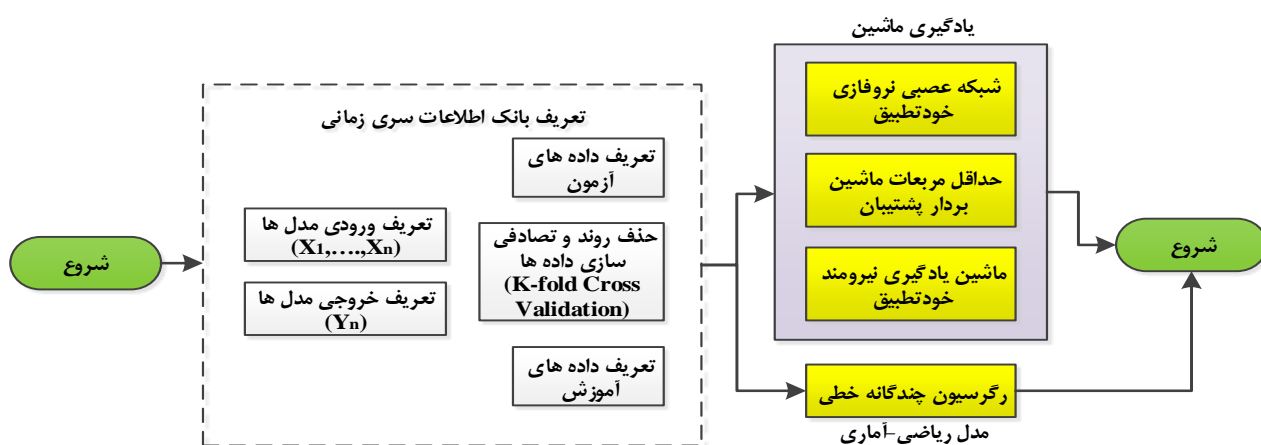
## مواد و روش‌ها

منطقه جغرافیایی که تحقیق حاضر در آن محل انجام گرفته است، حوزه آبریز دریاچه میقان در شهرستان اراک، واقع در استان مرکزی است. پارامترهای مورد بررسی کل جامدات محلول (TDS)، هدایت الکتریکی (EC)، شوری و سطح ایستابی آب زیرزمینی (GWL) بودند که

در این مطالعه از TDS، EC، شوری و GWL استفاده شد. شوری به صورت میزان انواع نمک موجود در حجم معینی از آب تعریف می شود. هرچند که EC به صورت میزان توانایی آب در رسانایی جریان الکتریکی قابل بیان است. همچنین TDS نیز به عنوان میزان املاح موجود در آب تعریف می گردد که این املاح می توانند از انواع کربنات های کلسیم و منیزیم، یون های کلسیم و منیزیم باشد.

با مدل های هوش مصنوعی و آماری شبیه سازی گردید. داده های بدست آمده در منطقه مربوط به نمونه برداری در مدت زمانی ۵ سال از محل تعداد ۱۷ چاه مشاهداتی حفر شده در ناحیه ویسمه، در مجاورت دریاچه مذکور بود.

### پارامترهای کیفی آب



شکل ۱- فلوچارت مراحل مختلف تحقیق.

بطوریکه بردار  $H$  در نقش خروجی ها، بردار  $\lambda$  در نقش بردار وزن ها برای هر ورودی می باشد. این دو بردار به عنوان عامل اتصال دهنده لایه مخفی شبکه و لایه خروجی آن، مقادیر بایاس را بصورت تصادفی مقداردهی می کنند و بردار  $f(x)$  نیز نقش بردار خروجی را دارد. بر همین اساس معادلات شماره ۲، ۳ و ۴ قابل استنتاج است:

$$\text{Weight vector: } \lambda = [\lambda_1, \dots, \lambda_m]^T \quad [2]$$

$$\text{Output vector: } H(x) = [h_1(x), \dots, h_m(x)]$$

$$H = \begin{bmatrix} h(x_1; c_1, f_1) & \dots & h(x_1; c_k, f_k) \\ \vdots & \ddots & \vdots \\ h(x_n; c_1, f_1) & \dots & h(x_n; c_k, f_k) \end{bmatrix}_{n \times k} \quad [3]$$

$$\lambda = H^\dagger f \quad [4]$$

در پایان لازم به تذکر است که  $H^\dagger$  ماتریس معکوس مور-پنروز ماتریس  $H$  است. در حقیقت در این روش با استفاده تکنیک حل یک رابطه خطی، از یک عملیات طرفین

مدل ماشین یادگیری نیرومند خودتطبیق (SAELM) اولین بار شخصی به نام هوانگ در سال ۲۰۰۴ مبنای این مدل را ارائه کرد (هوانگ و همکاران ۲۰۰۴). این مدل یکی از زیرمجموعه ماشین های یادگیری است و باتوجه به تک لایه بودن، قدرت و سرعت و دقت آن نسبت به بسیاری از روش های همتای خود اعم از شبکه های عصبی و ماشین های یادگیری مورد تایید محققین قرار گرفته است (هوانگ و همکاران ۲۰۱۲). در این مدل یادگیری، اگر  $n$  را تعداد نرون لایه مخفی فرض کنیم، بر اساس تعریف ریاضی، شبکه تک لایه پیشخور به صورت زیر قابل نوشتن است (پورسعید و همکاران ۲۰۲۲):

$$f_m(x) = \sum_{i=1}^m \lambda_i h_i(x) = H(x)\lambda \quad [1]$$

لازم بذکر است که  $\lambda_i$  بعنوان نویز شبکه تعریف می‌گردد. سپس برای هر  $x_i$ ، خروجی عبارتست از مجموعه وزن‌دار از  $n$  تابع هسته<sup>۹</sup> که متغیر مرکزی توابع هسته موردنظر با استفاده از  $x_i$  آموزش مشخص می‌گردد.  $\alpha_i$  ها در این معادلات همان ضرایب لاگرانژی<sup>۱۰</sup> هستند. حال با ترکیب ماتریس کرنل‌ها و توابع کرنل، هم چنین با داشتن مقادیر بردار  $a$ ، بردار  $Y$  بصورت ماتریسی معادله را نوشته و نتیجه را به شکل رابطه (۶) خواهیم داشت. توجه داشته باشید رابطه مذکور در حقیقت همان فرم ماتریسی مسئله بهینه سازی درجه ۲ است که تغییر شکل به حالت ماتریسی داده است:

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & \Omega_{11} + \frac{1}{\lambda} & \dots & \Omega_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \Omega_{n1} & \dots & \Omega_{nn} + \frac{1}{\lambda} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad [V]$$

و در معادله فوق  $\Phi_{ij}$  ماتریس هسته‌ها و  $\omega(x_i, x_j)$  توابع کرنل می‌باشند که بصورت زیر نوشته خواهند شد:

$$\Phi_{i,j} = \omega(x_i, x_j) \quad , \quad \Phi_{i,j} = \varphi^T(x_i) \cdot \varphi(x_j) \quad [A]$$

$, \quad i, j = 1, \dots, n$

#### شبکه عصبی فازی تطبیقی (ANFIS)

شبکه‌های عصبی مصنوعی نرو-فازی یا به اختصار ANFIS، نوعی از شبکه‌های عصبی پیشخور هستند که بر اساس منطق فازی شبیه سازی می‌کنند. در این نوع از شبکه‌ها از دو نوع سیستم استنباطی مبتنی بر منطق فازی

وسطین ساده بهره می‌گیریم تا بر اساس آن بردار وزن-ها که همان  $\lambda$  است را در طرف چپ تساوی، همچنین  $f$  و  $H^\dagger$  را در طرف مقابل قرار داده تا مجهولات را به دست آوریم.

#### مدل حداقل مربعات ماشین بردار پشتیبان (LSSVM)

این مدل، از زیرمجموعه مدل‌های ماشین بردار پشتیبان است که دچار تغییراتی شده است. این تغییرات باعث خاصیت خودتطبیقی شده که به تبع توان تنظیم پارامترهای ثابت در مدل را دارد (سپانکوویچ و سنکار ۲۰۰۹). مدل LSSVM براساس مینیمم‌سازی ریسک-ساختاری عمل می‌کند (کریستیان و تیلور ۲۰۰۰، دیباک و همکاران ۲۰۰۱).

در تئوری این مدل، براساس مسئله برنامه‌ریزی درجه ۲ معادله حاصل می‌شود که ثابت‌های مدل در آن تعیین شده و توسط یک بهینه ساز مانند الگوریتم ژنتیک GA بهینه می‌شوند. برای بیان ریاضی یک مدل LSSVM، اگر  $x_i$  و  $y_i$  را ورودی‌ها و خروجی‌ها در نظر بگیریم، سپس تابع رگرسیونی غیرخطی به شکل زیر خواهد بود (والیان و هاروات ۲۰۰۷):

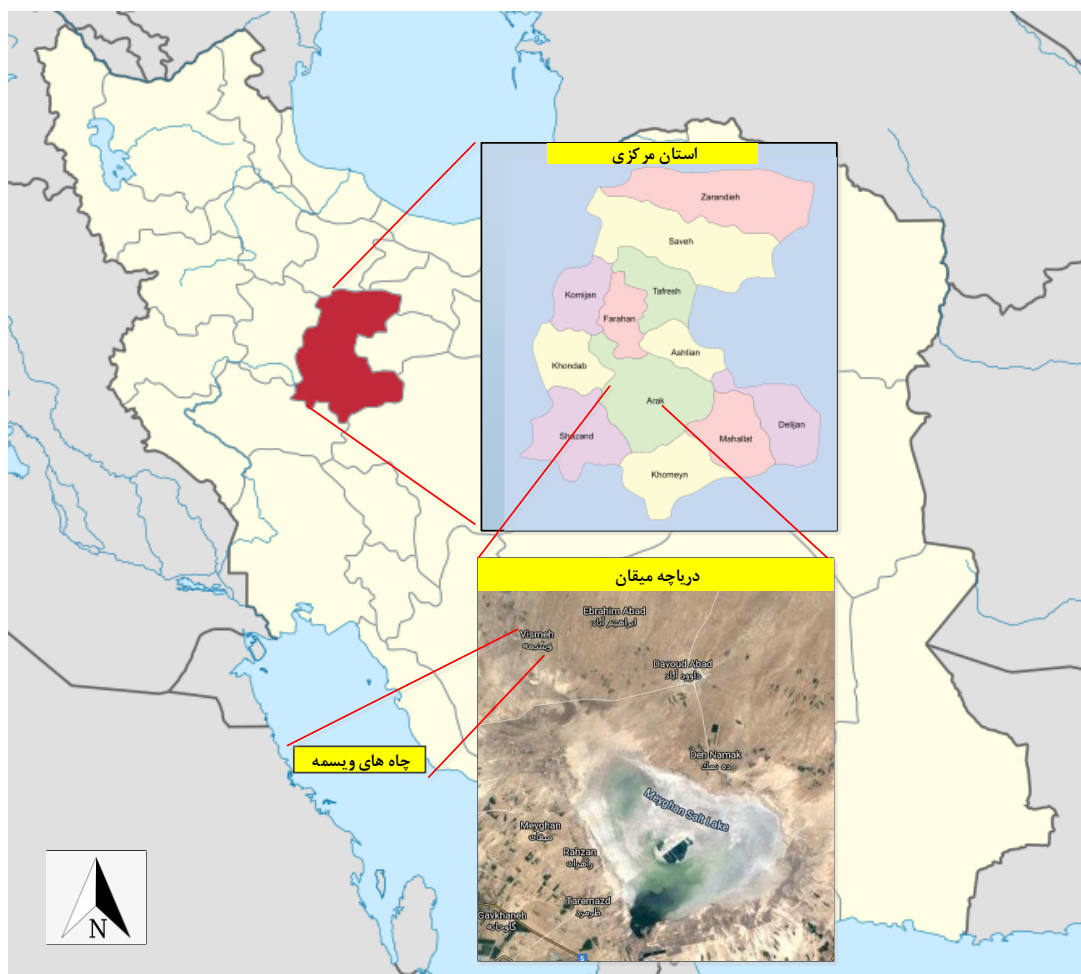
$$f(x) = \sum_{i=1}^k w_i \varphi_i(x) + b = w^T \cdot \varphi(x_i) + b \quad [5]$$

که در اینجا  $w$  بردار وزن،  $b$  بایاس و  $\varphi$  ترسیم‌گرهای غیرخطی که در فضای ویژگی عمل نمایند هستند و لذا با کمینه کردن مسئله بهینه سازی درجه (۲)، مسئله رگرسیون غیرخطی بصورت زیر قابل حل است:

$$Min_{w,e,b} \left\{ \Omega(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \right\} \quad [6]$$

<sup>10</sup> Lagrangian Multipliers

<sup>9</sup> Kernel function



شکل ۲- محدوده جغرافیایی تالاب میقان (منبع: Wikimedia & GoogleMap).

لایه اول: این لایه توابع عضویت را با استفاده از قوانین گاوسی برای هر گره از شبکه بوسیله نقاط فازی شده فراهم می آورد. لایه دوم: در این لایه کلیه مقادیر فازی شده نقاط شبکه از اپراتورها رد می گردد.  $O_i$  و  $D_i$  به عنوان پارامترهای عضویت و نیز  $\mu_{i_1}(x)$  و  $\mu_{i_2}(x)$  هم به عنوان پارامترهای پیشین لایه اول می باشند.

لایه سوم: در این لایه کلیه مقادیر بر اساس فرمول ذیل نرمالیزه می شوند.

لایه چهارم: در این لایه مقدار تابع خطی هر گره محاسبه شده و مقادیر خطا نیز با استفاده از روش محاسبه خطا در شبکه های عصبی انتشار عقبگرد به دست می آید.

FIS استفاده می گردد (الشفاعی و همکاران ۲۰۰۶): سیستم استنباطی ممدانی<sup>۱۱</sup> و سیستم استنباطی تاکاگی-سوگنو<sup>۱۲</sup>. در این شبکه ها حداقل دو ورودی  $D_1$  و  $D_2$  وجود دارد که برای شبکه مبتنی بر سیستم استنباط فازی TS-FIS و دو اصل شرطی اگر-آنگاه<sup>۱۳</sup> برای هر خروجی  $O$  خواهد بود (آرورا و کشاری ۲۰۲۱). شبکه های نروفازی در یک لایه ورودی و ۵ لایه دیگر خلاصه می گردد که این شبکه می تواند به نوعی بیانگر یک شبکه عصبی چند لایه باشد.

لایه صفر: لایه ورودی با  $n$  ورودی.

<sup>13</sup> If-then

<sup>11</sup> Mamdani

<sup>12</sup> Tagaki-Sugeno

لایه پنجم: در این لایه مجموع مقادیر خروجی هر گره از لایه چهارم محاسبه می شود.

## نتایج و بحث

### ارزیابی عملکرد با نمودارهای پیش‌بینی

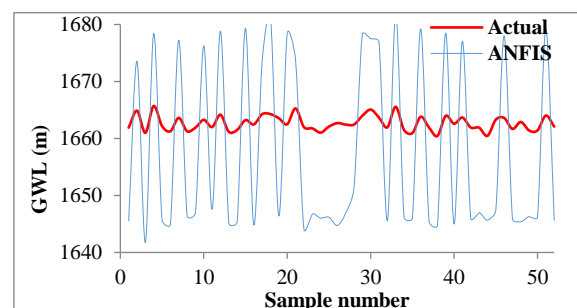
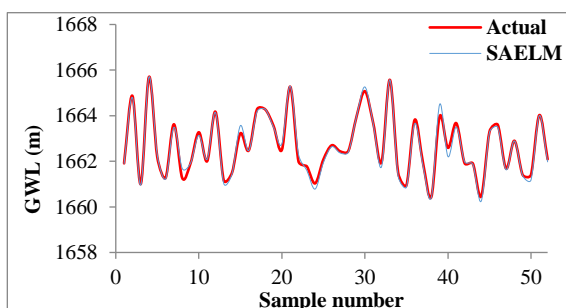
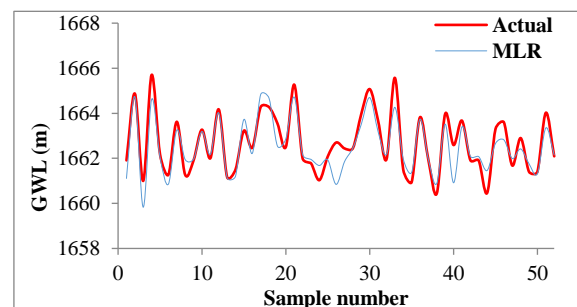
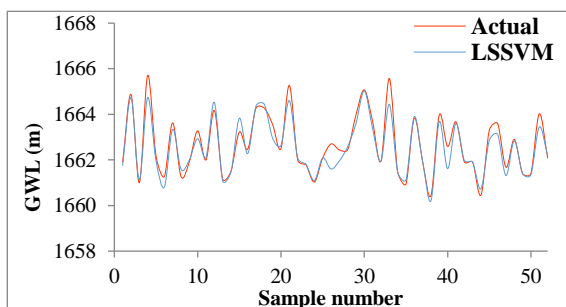
مقادیر پارامترهای کیفی و کمی آب به ترتیب شوری، EC، TDS و GWL به عنوان ورودی برای هر مدل‌سازی بطور جداگانه در مدل‌ها وارد و شبیه‌سازی انجام گرفت. در نمودارها محور قائم مقادیر شبیه‌سازی شده و محور افقی تعداد آزمون‌های نمونه‌گیری را نشان می‌دهند. در نهایت بر اساس نمودارهای سنجش دقت تخمین، با توجه به بهترین تطبیق و کمترین اختلاف بین نمودارهای مربوط به مقادیر مشاهداتی (نمونه‌برداری) با مقادیر شبیه‌سازی شده، مدل SAELM بعنوان بهترین و برترین مدل در شبیه‌سازی پارامتر GWL تعیین می‌گردد.

### مدل رگرسیونی چندگانه خطی (MLR)

روش‌های رگرسیون خطی چندگانه روش‌هایی مبتنی بر محاسبات آماری می‌باشند. در این روش‌ها جهت بررسی روابط متغیر ورودی و چند متغیر هدف قابل استفاده هستند. تعریف ریاضی این مدل بصورت ذیل می‌باشد:

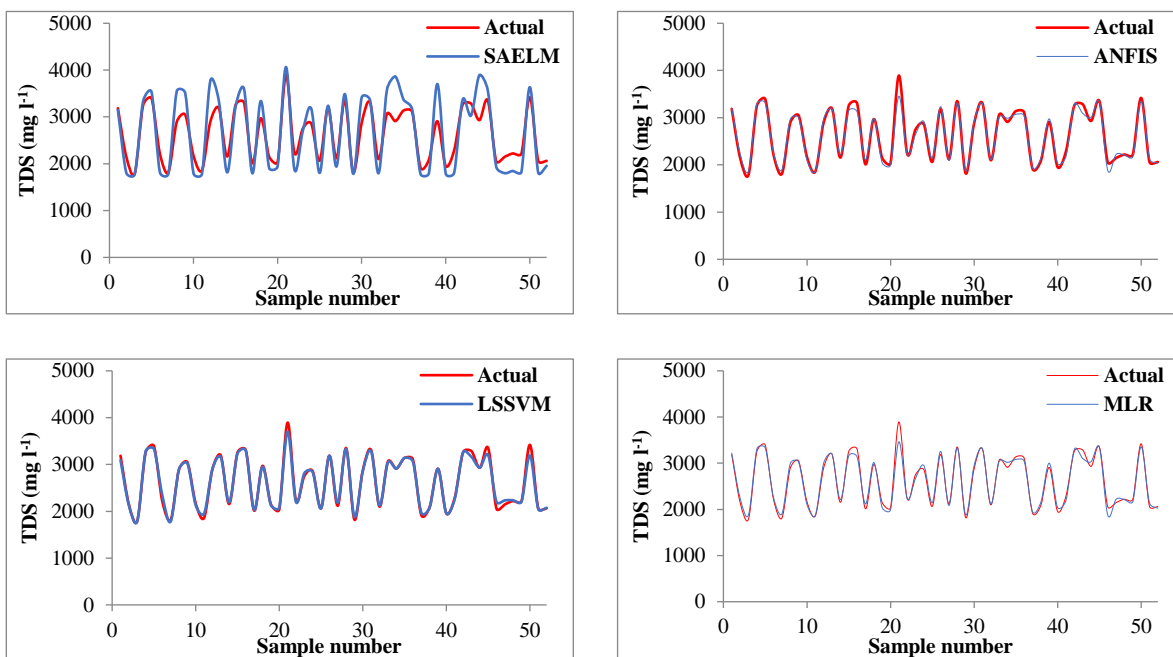
$$y_i = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + \zeta_i \quad [9]$$

که در این معادله  $y_i$  متغیر هدف،  $x_m$  متغیرهای چندگانه ورودی مدل،  $a_m$  ضرایب رگرسیونی و  $\zeta$  خطای تصادفی در مدل رگرسیونی است (مصطفی و عبدو ۲۰۱۲)

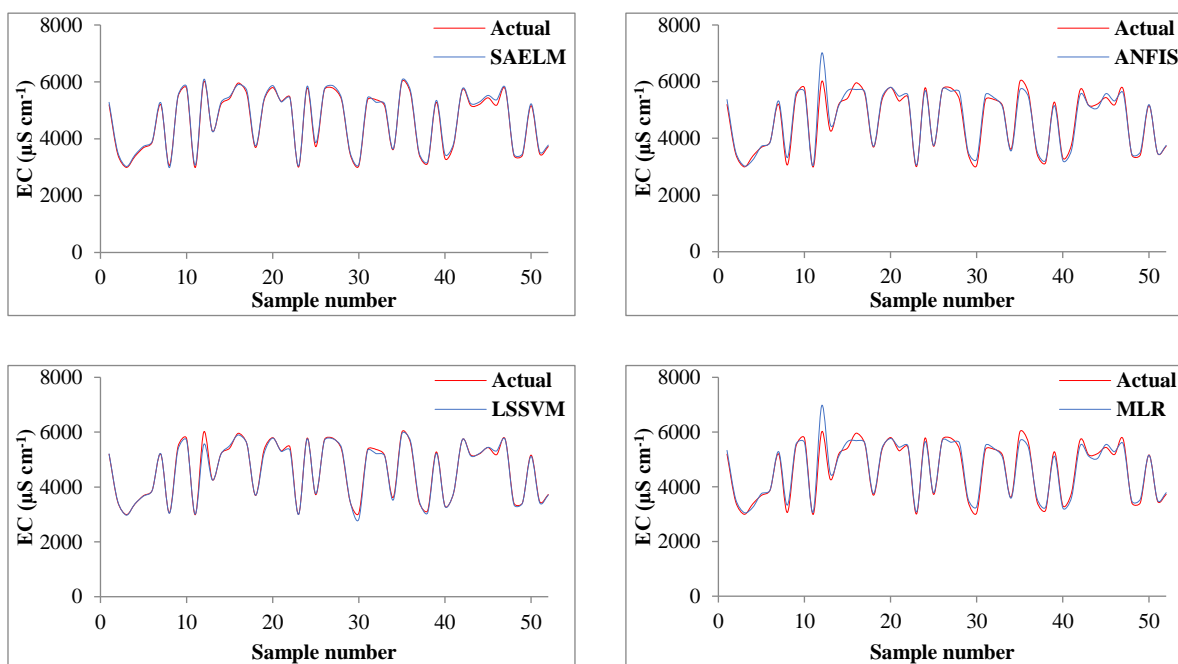


شکل ۳- نمودارهای سنجش دقت مدل‌ها در تخمین GWL.

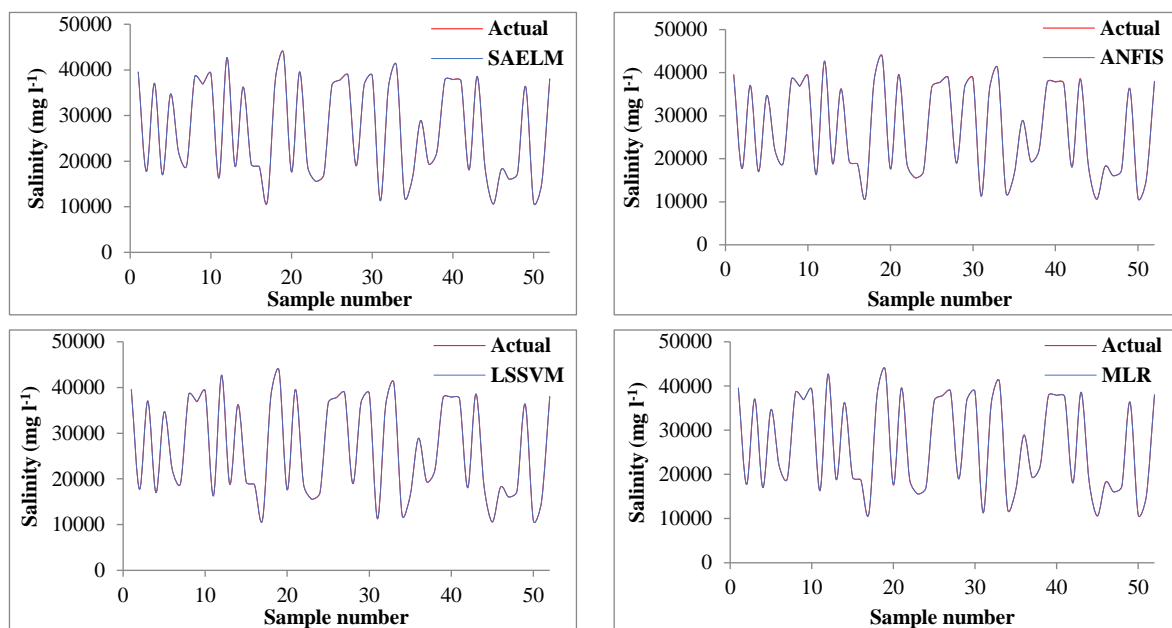




شکل ۴- نمودارهای سنجش دقت مدل‌ها در تخمین TDS.



شکل ۵- نمودارهای سنجش دقت مدل‌ها در تخمین EC.



شکل ۶- نمودارهای سنجش دقت مدل‌ها در تخمین شوری.

جدول ۱- نتایج معیارهای ارزیابی عملکرد مدل‌ها.

پارامتر	مدل	R	MAPE	RMSE
Salinity	SAELM	۰/۹۹۸۹	۰/۰۷۵۳۳	۲۴/۸۹۱۹
	LSSVM	۰/۹۹۹۹	۱/۱۹E-۰۴	۰/۰۳۳۵۸
	ANFIS	۰/۹۹۸۸	۰/۱۲۳۶	۵۱/۸۴
	MLR	۰/۸۸۹۷	۱/۲۷۹E-۱۴	۵/۲۸E-۱۲
EC	SAELM	۰/۹۹۹۸	۱/۳۴۹۸	۷۰/۸۹۵
	LSSVM	۰/۹۹۴۵	۱/۶۲۱۱	۹۶/۱۶۱۶
	ANFIS	۰/۹۸۷۵	۲/۶۶۹۰	۱۹۱/۳۵۱۷
	MLR	۰/۹۸۲۹	۲/۷۹E۰۰	۱۹۰/۲۵۴۴
TDS	SAELM	۰/۹۵۷۹	۱۲/۷۲۹	۴۳۳/۹۸۴۹
	LSSVM	۰/۹۹۴۹	۱/۶۱۱۶	۶۶/۵۵۶
	ANFIS	۰/۹۸۶۹	۲/۶۶	۹۹/۸۳۰۵
	MLR	۰/۹۸۵۸	۲/۷۸۸	۹۶/۲۶۳
GWL	SAELM	۰/۹۹۸۹	۰/۰۰۸۸	۰/۱۳۵۸
	LSSVM	۰/۹۶	۰/۰۱۴۹	۰/۲۶۰۹
	ANFIS	۰/۸۳۵۷	۷/۹۷	۱۴/۵۷۵
	MLR	۰/۸۸۹	۰/۱۲۷۹	۰/۵۹۵۹

\*اعداد پررنگ نتایج مدل‌های برتر هستند

با توجه به شاخص های ریاضی سنجش عملکرد، با توجه نتایج پررنگ مندرج در جدول، مشخص گردید که در شبیه سازی دو پارامتر EC و GWL، مدل SAELM بهترین مدل بوده است. همچنین در شبیه سازی پارامتر TDS مدل LSSVM به عنوان برترین مدل شناخته شد. مدل MLR در شبیه سازی شوری دقیق ترین نتایج را ارائه دادند.

### تحلیل عدم قطعیت به روش امتیاز ویلسون<sup>۱</sup>

بعد از شبیه سازی مدل های هوش مصنوعی و رگرسیونی و به منظور بررسی میزان خطا در پیش بینی مجهولات، میزان عدم قطعیت هر مدل قابل محاسبه بوده و بر اساس آن می توان عملکرد هر مدل را در تخمین مقادیر هدف مورد ارزیابی و سنجش قرار داد. در تحقیق

حاضر به روش تحلیل عدم قطعیت ویلسون، برای مدل های کاربردی موجود این محاسبات انجام گردید. قبل از نتایج تحلیل، پارامترهای محاسباتی تحلیل بصورت زیر معرفی می گردد: میزان خطای تخمین  $\delta_i$ ، میانگین خطا  $\bar{\delta}_i$  و انحراف استاندارد مقادیر خطا  $S_i$ ، که در جدول زیر نتایج تحلیل عدم قطعیت به روش ویلسون (بدون تصحیح) با مقادیر باند اطمینان (WUB) درج شده است. مقادیر مذکور با احتساب سطح معناداری ۵٪ محاسبه گردید که باند اطمینان به مقدار  $\pm 1.64S_i$  را ایجاد خواهد کرد. همچنین بر اساس مقادیر انحراف استاندارد خطا  $S_i$  برای هر مدل، در صورتی که مقدار آن مثبت باشد مدل دارای عملکرد بیش برآوردی<sup>۲</sup> و در صورت منفی بودن آن مدل دارای عملکرد کم برآوردی<sup>۳</sup> خواهد بود (پورسعید و همکاران ۲۰۲۱).

جدول ۲- نتایج تحلیل عدم قطعیت ویلسون.

نام مدل	باند اطمینان	میانگین خطای تخمین	انحراف استاندارد	سطح اطمینان ۹۵٪		عملکرد مدل
				حداقل باند	حداکثر باند	
SAELM(GWL)	۰/۰۳۵۳۲	-۰/۰۱۲۰۳	۰/۱۷۹۸۶	-۰/۰۴۳۶	۰/۲۰۸۷	کم برآوردی
SAELM(EC)	۹/۱۱۴۵	۹۳۲۲/۲۷۱۵	۹۴۲۳/۲۱	۹۳۰۲/۳۶۳۳	۹۳۰۰/۹۲۲۷	بیش برآوردی
LSSVM(TDS)	۱۳/۰۱۴۸	۵۲۵۳/۲۰۸۲	۵۲۶۹/۸۸۳	۵۱۴۰/۰۰۹۲	۵۳۷۱/۱۶۱۵	بیش برآوردی
MLR(Salinity)	۱/۲۵۹E-۱۲	۵۳۸۰۱/۱۷۸۷	۵۷۹۸۹/۵۱	۵۳۸۹۹/۲۸۱	۵۳۹۹۹/۱۸۹	بیش برآوردی

نسبت اختلاف خطا به اختصار DR، برای تعیین میزان دقت مدل های برتر در شبیه سازی و مدل سازی بکار می رود. طبق تعریف، این نسبت برابر با حاصل تقسیم مقادیر شبیه سازی شده به مقادیر مشاهداتی است. بعد از شناسایی مدل های برتر در شبیه سازی، با توجه به تعریف ریاضی نسبت اختلاف، دقت بالای SAELM و MLR اثبات گردید.

بر اساس نتایج جدول ۱، عملکرد کم برآوردی مدل SAELM و نیز برای سایر مدل های برتر در شبیه سازی عملکرد بیش برآوردی تعیین گردید. همچنین بر اساس این تحلیل، مدل SAELM با حداقل مقدار میانگین خطای تخمین برابر با ۰/۰۱۲۰۳ به عنوان بهترین و دقیق ترین مدل انتخاب شد.

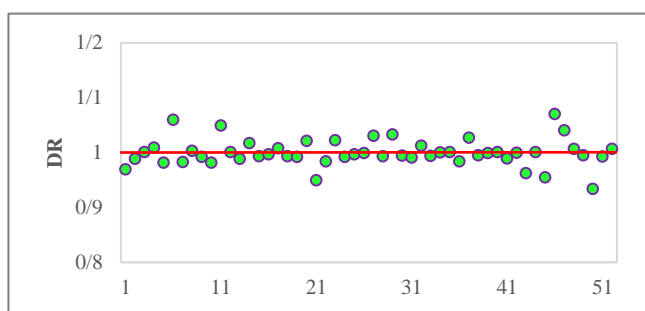
### نمودار نسبت اختلاف خطا<sup>۴</sup>

<sup>3</sup> Underestimated

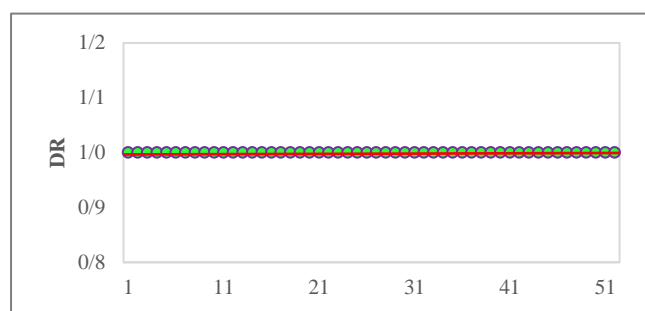
<sup>1</sup> Wilson Score Method Uncertainty Analysis

<sup>4</sup> Discrepancy ratio

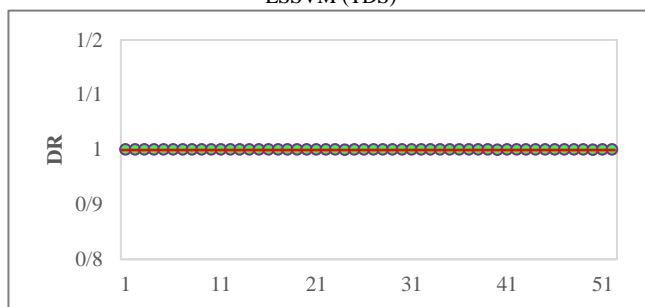
<sup>2</sup> Overestimated



LSSVM (TDS)



MLR (Salinity)



SAELM (GWL)



SAELM (EC)

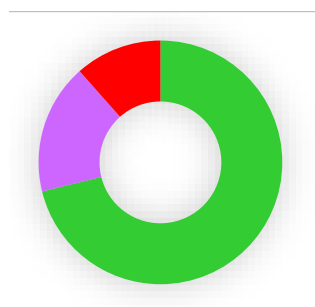
شکل ۸- نمودارهای نرخ اختلاف خطا برای مدل‌های برتر در شبیه‌سازی.

خطاهای شبیه‌سازی که برابر با حاصل اختلاف مقادیر شبیه‌سازی شده هر مدل با مقدار مشاهداتی متناظر با آن می‌باشد، در بازه‌های مشخصی از خطای شبیه‌سازی تعیین می‌شود و سپس در قالب نمودارهای توزیع خطا به تصویر کشیده می‌شود (پورسعید و همکاران ۲۰۲۱).

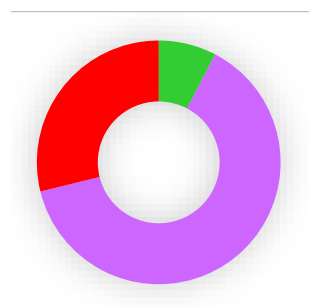
زیرا تمرکز و تعدد هرچه بیشتر نقاط روی نمودار به خط افقی  $DR=1$ ، دقت بالای مدل‌های موردنظر را نشان می‌دهد (پورسعید و همکاران ۲۰۲۱).

### نمودار توزیع خطا

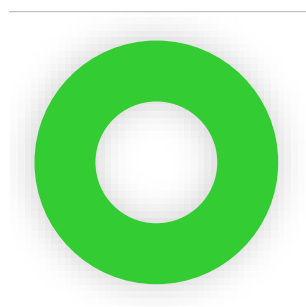
نمودارهای توزیع خطای یکی دیگر از رویکردهای بررسی دقت و توزیع خطای مدل‌سازی می‌باشد. در این روش



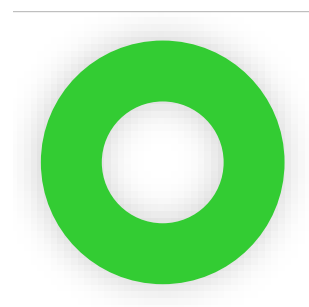
LSSVM(TDS)



SAELM(EC)



SAELM(GWL)



MLR(Salinity)

درصد خطا: کمتر از ۱% (سبز)، بین ۱ تا ۵% (بنفش)، بزرگتر از ۵% (قرمز)

شکل ۹- نمودارهای توزیع خطای مدل‌های برتر در شبیه‌سازی.

## نتیجه گیری کلی

در تحقیق حاضر از خصوصیات کمی و کیفی آبها طبق نمودارهای دقت پیش بینی، تعداد ۱۶ نمودار ترسیم گردید که دقیق ترین مدلها در شکل های ۴ تا ۷ به تصویر کشیده شده است. طبق معیارهای سنجش عملکرد هر مدل در زمینه تخمین یکی از پارامترهای کمی و کیفی آب موفق بود. در نهایت نتایج بررسی های کمی نشان داد که مدل SAELM در شبیه سازی پارامترهای EC و GWL برترین مدل و با کمترین خطای محاسباتی بود. همچنین مدل LSSVM در تخمین و مدل سازی پارامتر TDS دقیق ترین مدل بود. در شبیه سازی پارامتر شوری نیز مدل MLR برترین مدل بود. بعد از شبیه سازی بر اساس روش امتیاز ویلسون تحلیل عدم قطعیت انجام گرفت. عملکرد مدلها را در شبیه سازی بر اساس این تحلیل نشان داد که عملکرد مدل SAELM بصورت کم برآورد و سایر مدل های برتر بیش برآورد تعیین گردید. بر اساس نمودارهای توزیع خطا، بهترین دقت به مدل های SAELM و MLR اختصاص یافت. ۵- بر اساس نسبت اختلاف خطا، مدل های SAELM و MLR دقیق ترین مدلها در شبیه سازی برآورد گردیدند. در نهایت باید ذکر نمود که بر اساس بررسی نتایج شبیه سازی مدل های منتخب و برتر در تخمین پارامترهای فوق به ترتیب پیش بینی های ذیل محتمل است:

- در تخمین تغییرات GWL توسط مدل برتر SAELM، به مرور دامنه ای کوچکتر و تغییراتی ناچیزتر را تجربه نمود، هرچند که به نظر می رسد نمودار سیر نزولی را ادامه خواهد داد.

- بر اساس نمودار تخمین میزان تغییرات EC مطابق با شبیه سازی مدل برتر SAELM، این پارامتر ماهیت تغییرات سری زمانی خود را تقریباً حفظ کرد، هرچند بر اساس نمودار پیش بینی به نظر می رسد در آینده به شرط رفتار مشابه با سری زمانی، تغییرات در بازه کمتری اتفاق خواهد افتاد.

- بر اساس نمودار تخمین میزان تغییرات TDS مطابق با شبیه سازی مدل برتر در این زمینه LSSVM، این پارامتر ماهیت تغییرات سری زمانی خود را تقریباً و با سیر نامنظم حفظ نمود و به نظر می رسد در آینده به شرط رفتار مشابه با سری زمانی، تغییرات در همین محدوده به وقوع بپیوندد.

- بر اساس نمودار تخمین میزان تغییرات شوری مطابق با شبیه سازی مدل برتر در این زمینه MLR، به نظر می رسد که تغییرات این پارامتر با دامنه بالاتری رخ بدهند و مقادیر حداقل و حداکثر این پارامتر با مقادیری بیش از گذشته رخ دهند.

## منابع مورد استفاده

- Arora S and Keshari AK, 2021. ANFIS-ARIMA modelling for scheming re-aeration of hydrologically altered rivers. *Journal of Hydrology* 601:126635.
- Banerjee P, Singh VS and Chattopadhyay K, 2011. Artificial neural network model as a potential alternative for groundwater salinity forecasting. *Journal of Hydrology* 398:212-220.
- Che Nordin NF, Mohd NS and Koting S, 2021. Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundwater Sustainable Development* 14:100643.
- Cristianini N and Shawe-Taylor J, 2000. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge.
- Cuest Cordoba GA, Tuhovčák L and Tauš M, 2014. Using artificial neural network models to assess water quality in water distribution networks. *Procedia Engineeng* 70:399-408.
- Dibike YB, Velickov S, Solomatine D and Abbott MB, 2001. Model induction with Support Vector Machines: Introduction and applications. *Journal of Computing in Civil Engineering* 15:208-216.
- El-Shafie A, Taha MR and Noureldin A, 2006. A neuro-fuzzy model for inflow forecasting of the Nile river

- at Aswan high dam. *Water Resources Management* 21:533–556.
- Huang GB, Zhou H, Ding X and Zhang R, 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Cybernetics* 42:513–529.
- Huang GB, Zhu QY and Siew CK, 2004. Extreme learning machine: A new learning scheme of feedforward neural networks. *IEEE Int Conf Neural Networks - Conference Proceedings* 2:985–990.
- Huang GB, Zhu QY and Siew CK, 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70:489–501.
- Jamei M, Ahmadianfar I, Chu X and Yaseen ZM, 2020. Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: New approach. *Journal of Hydrology* 589:125335.
- Kadkhodazadeh M and Farzin S, 2021. A novel LSSVM model integrated with GBO algorithm to assessment of water quality parameters. *Water Resources Management* 35:3939–3968. <https://doi.org/10.1007/s11269-021-02913-4> (In Persian with English abstract)
- Mustapha A and Abdu A, 2012. Application of principal component analysis & multiple regression models in surface water quality assessment. *Environmental Earth Sciences* 2:16–23
- Poursaeid M, Mastouri R, Shabanlou S and Najarchi M, 2020. Estimation of total dissolved solids, electrical conductivity, salinity and groundwater levels using novel learning machines. *Environmental Earth Sciences* 79:1–25. <https://doi.org/10.1007/S12665-020-09190-1>
- Poursaeid M, Mastouri R, Shabanlou S and Najarchi M, 2021. Modelling qualitative and quantitative parameters of groundwater using a new wavelet conjunction heuristic method: wavelet extreme learning machine versus wavelet neural networks. *Water and Environment Journal* 35:67–83. <https://doi.org/10.1111/WEJ.12595>
- Poursaeid M, Poursaeid AH and Shabanlou S, 2022. A comparative study of artificial intelligence models and a statistical method for groundwater level prediction. *Water Resources Management* 36:1499–1519. <https://doi.org/10.1007/S11269-022-03070-Y>
- Rajae T, Khani S and Ravansalar M, 2020. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems* 200:103978.
- Sapankevych N and Sankar R, 2009. Time series prediction using support vector machines: A survey. *IEEE Computational Intelligence Magazine* 4:24–38.
- Valyon J and Horvath G, 2007. Extended Least Squares LS-SVM. *World Academy of Science, Engineering and Technology* 36.
- Wang P, Yao J and Wang G, 2019. Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. *Science of Total Environment* 693:133440. <https://doi.org/10.1016/J.SCITOTENV.2019.07.246>