

# Predicting Saltwater Intrusion into Coastal Aquifers Using Support Vector Regression Surrogate Models

F. Fa'al<sup>1</sup>, H. R. Ghafouri<sup>2</sup>, S. M. Ashrafi<sup>3</sup>

1. PhD Student, Dept. of Civil Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran  
(Corresponding Author) faalfateme@gmail.com
2. Prof., Dept. of Civil Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
3. Assist. Prof., Dept. of Civil Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

(Received Jan. 7, 2019 Accepted May 22, 2019)

#### To cite this article:

Fa'al, F., Ghafouri, H. R., Ashrafi, S.M. 2020. "Predicting saltwater intrusion into coastal aquifers using support vector regression surrogate models" Journal of Water and Wastewater, 31(2), 118-129. Doi: 10.22093/wwj.2019.161020.2803 (In Persian)

#### Abstract

The prediction of the intrusion of saline water into coastal aquifers as a result of changing the amount of groundwater extractions is a prerequisite for managing groundwater. This study investigates the capability of different types of Support Vector Regression (SVR) models to predict salinity concentrations at the selected well in the small coastal aquifer under different groundwater abstraction conditions. SVR models were trained and tested using input (random transient pumping from the production wells) derived from Latin Hypercube Sampling and output (salinity concentration at the selected well) datasets. The trained and tested models were then used to predict salinity concentrations at the selected well for new pumping datasets. The models ability for predicting and generalizing compared with commonly used artificial neural network (ANN) model was evaluated using different performance criteria. The results of the performance evaluation of the models showed that the predictive capability of the polynomial SVR model is superior to other models. Also, comparing different performance criteria for all SVR models, except for linear SVR model, proved their acceptable predictive performance. The prediction and generalisation ability of polynomial SVR, recommends using these models to connect to the optimization algorithm for a surrogate model based simulation-optimization approach in sustainable management of coastal aquifers.

**Keywords:** Coastal Aquifers, Saltwater Intrusion, Surrogate Models, Support Vector Regression, Prediction Capability.



مجله آب و فاضلاب، دوره ۳۱، شماره ۲، صفحه: ۱۲۹-۱۱۸

## پیش‌بینی پیشروی آب شور در آبخوان ساحلی با استفاده از ماشین بردار پشتیبان رگرسیونی به‌عنوان مدل جایگزین

فاطمه فعال<sup>۱</sup>، حمیدرضا غفوری<sup>۲</sup>، سیدمحمد اشرفی<sup>۳</sup>

۱- دانشجوی دکترا، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه شهید چمران اهواز، اهواز، ایران  
 (نویسنده مسئول) faalfateme@gmail.com

۲- استاد، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه شهید چمران اهواز، اهواز، ایران

۳- استادیار، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه شهید چمران اهواز، اهواز، ایران

(دریافت ۹۷/۱۰/۱۶ پذیرش ۹۸/۲/۳۱)

برای ارجاع به این مقاله به صورت زیر اقدام فرمایید:

فعال، ف.، غفوری، ح.، ر.، اشرفی، س. م.، ۱۳۹۹، "پیش‌بینی پیشروی آب شور در آبخوان ساحلی با استفاده از ماشین بردار پشتیبان رگرسیونی

به‌عنوان مدل جایگزین" مجله آب و فاضلاب، ۳۱(۲)، ۱۲۹-۱۱۸. Doi: 10.22093/wwj.2019.161020.2803

### چکیده

پیش‌بینی هجوم آب شور به آبخوان ساحلی در اثر تغییر میزان برداشت از آب زیرزمینی یکی از پیش‌نیازهای مدیریت آب زیرزمینی است. این پژوهش قابلیت انواع مختلف مدل‌های ماشین بردار پشتیبان رگرسیونی (SVR) را برای پیش‌بینی میزان شوری در چاه پایش منتخب در یک آبخوان ساحلی کوچک تحت شرایط مختلف برداشت آب زیرزمینی بررسی می‌کند. مدل‌های SVR با مجموعه‌های ورودی (پمپاژ گذرای تصادفی از چاه‌های آبده) حاصل از نمونه‌برداری ابرمکعب لاتین و خروجی (غلظت شوری در چاه پایش منتخب) آموزش یافته و آزمایش شدند. سپس، مدل‌های آموزش یافته و آزمایش شده برای پیش‌بینی میزان شوری در چاه پایش منتخب برای مجموعه داده‌های جدید پمپاژ استفاده شدند. قابلیت پیش‌بینی و تعمیم‌دهی عملی مدل‌ها در مقایسه با مدل رایج شبکه عصبی مصنوعی با استفاده از شاخص‌های ارزیابی مختلف بررسی شد. نتایج ارزیابی عملکرد مدل‌ها نشان داد که قابلیت پیش‌بینی مدل SVR چندجمله‌ای بر سایر مدل‌ها برتری دارد. همچنین همه مدل‌های SVR مورد بررسی به‌جز مدل SVR خطی، پیشروی آب شور را به‌طور قابل قبولی پیش‌بینی کردند. قابلیت پیش‌بینی و تعمیم‌دهی مدل‌های SVR چندجمله‌ای، استفاده از آنها را برای اتصال به الگوریتم بهینه‌سازی در شیوه بهینه‌سازی - شبیه‌سازی مبتنی بر مدل جانشین برای مدیریت آبخوان‌های ساحلی توصیه می‌کند.

**واژه‌های کلیدی:** آبخوان ساحلی، پیشروی آب شور، مدل جانشین، ماشین بردار پشتیبان رگرسیونی، قابلیت پیش‌بینی

### ۱- مقدمه

محیطی شود. بنابراین پایش و مدیریت دقیق سطح حد فاصل آب‌های شور و شیرین آبخوان ساحلی برای حفاظت منابع آب شیرین زیرزمینی بسیار ضروری است. اما مدیریت آبخوان‌های ساحلی در مقیاس منطقه‌ای برای کنترل پیشروی آب شور مسئله‌ای چالش برانگیز است که نیاز به انجام مکرر شبیه‌سازی‌های عددی

پیشروی آب شور به معنای نفوذ آب شور در آبخوان شیرین ساحلی است که می‌تواند باعث آلودگی منابع آب زیرزمینی شود (Chitsazan, 2014). پیشروی آب شور به‌علت برداشت بیش از حد و برنامه ریزی نشده از آبخوان‌های ساحلی می‌تواند مانع استفاده مفید از آبخوان‌ها و در نتیجه ایجاد افت بزرگ اقتصادی و زیست

جعبه سیاه آن، بار محاسباتی زیاد، و استعداد بیش‌برازش آن است. از سوی دیگر مدل برپایه فرمول صریح ریاضی استوار است که همان مدل‌های رگرسیون ساده هستند. گذشته از این GP برای یافتن ساختار بهینه به ارزیابی میلیون‌ها ساختار مختلف نیاز دارد که زمان آموزش وسیعی را می‌طلبد. محدودیت دیگر GP آمادگی برای همگرایی سریع در بهینه محلی و در نتیجه ناتوانی در یافتن پاسخ‌ها برای برخی از مسائل شامل توابع برازش چندقله‌ای (یعنی شامل تعداد زیادی قله یا تورفتگی) است (Pillay, 2004). فرامدل RBF نیز شبیه GP فرمول نسبتاً ساده‌ای دارد. کاربرد چنین مدل‌های ساده‌ای برای نمایش سیستم‌های پیچیده و بسیار غیرخطی آبخوان ساحلی ایده‌آل به نظر نمی‌رسد. استفاده از مدل‌های ماشین‌بردار پشتیبان رگرسیونی به‌عنوان مدل جایگزین برای شبیه‌سازی پیشروی آب شور بر برخی از محدودیت‌های ذکر شده غلبه می‌کند. یون و همکاران در سال ۲۰۱۱ نوسانات سطح آب زیرزمینی دو چاه پایش در یک آبخوان ساحلی را با استفاده از ANN و مدل ماشین‌بردار پشتیبان<sup>۴</sup> پیش‌بینی کردند. عملکرد مدل ANN در مراحل آموزشی و آزمایشی بهتر از مدل SVM بود. اما عملکرد کلی مدل SVM بهتر از مدل ANN در مرحله پیش‌بینی بود. مدل SVM برای ساختارهای ورودی و زمان‌های تأخیر متفاوت نسبت به مدل ANN قابلیت تعمیم‌دهی بهتری داشت (Yoon et al., 2011). یون و همکاران در سال ۲۰۱۷ نوسان سطح آب شور-آب شیرین را در جزیره‌ای در کره جنوبی توسط مدل‌های سری زمانی مبتنی بر ANN و SVM پیش‌بینی کردند. نتایج آنها نشان داد که مدل SVM دقیق‌تر از ANN برای منطقه مطالعاتی آنها بوده است (Yoon et al., 2017).

لال و داتا در سال ۲۰۱۸ مدل SVR با ابرپارامترهای مشخص را برای پیش‌بینی غلظت شوری در بخشی از یک آبخوان ساحلی تحت شرایط پمپاژ متغیر به کار بردند. آنها ۴۰۰ مجموعه داده پمپاژ ورودی را برای آموزش، ۱۰۰ داده برای اعتبارسنجی و ۲۰۰ داده برای پیش‌بینی به‌کار گرفتند. نتایج آنها نشان داد که مدل SVR قابلیت مناسبی برای تقریب فرایند پیشروی آب شور در مقایسه با مدل برنامه‌نویسی ژنتیک دارد (Lal and Datta, 2018).

کاسیفیتس و همکاران در سال ۲۰۱۹ روش رگرسیون فرایند

فرایندهای غیرخطی جریان و انتقال درون الگوریتم بهینه‌سازی دارد. یکی از روش‌های کاهش زمان محاسباتی ناشی از تکرارهای زمان‌بر بهینه‌سازی و شبیه‌سازی‌های گران‌عددی محیط فیزیکی پیچیده، ساختن مدل تقریبی کارآمد از مدل کامپیوتری بر پایه فیزیک اصلی است که امروزه از آن به‌عنوان مدل جانشین یا فرامدل نام برده می‌شود (Roy, 2018). هدف از فرامدل‌ها تخمین پاسخ‌های عددی زمان‌بر اصلی بر پایه انواع متغیرهای ورودی موردنظر در فرایند بهینه‌سازی و تحلیل حساسیت است. مدل‌های جانشین آموزش یافته و آزمایش شده با دقت و کارایی قابل قبول می‌توانند به‌عنوان شبیه‌سازهای تقریبی درون الگوریتم بهینه‌سازی استفاده شوند.

در پژوهش‌های پیشین پیشروی آب شور، انواع مختلفی از مدل‌های جانشین برای تقریب مدل شبیه‌سازی عددی استفاده شده‌اند. مدل‌های جانشین رایج مورد استفاده در مدل‌سازی پیشروی آب شور شامل شبکه عصبی مصنوعی<sup>۱</sup> (Bhattacharjya and Datta, 2005, Bhattacharjya et al., 2007, Dhar and Datta, 2009, Rao et al., 2004, Sreekanth and Datta, 2011) برنامه‌ریزی ژنتیک<sup>۲</sup> (GP) (Pillay, 2004, Sreekanth and Datta, 2010, Sreekanth and Datta, 2011) شعاعی<sup>۳</sup> (Christelis and Mantoglou, 2016, Papadopoulou et al., 2010)، شبکه عصبی پیمان‌های (Kourakos and Mantoglou, 2009, Sreekanth and Datta, 2010) استنتاج فازی (Roy and Datta, 2017c, Roy and Datta, 2017a) و رگرسیون انطباقی چند متغیره اسپلاین است (Roy and Datta, 2017b).

مدل‌های ANN آموزش یافته محبوب‌ترین انتخاب برای مدل جانشین در مدیریت آبخوان‌های ساحلی و مطالعات منابع آب بوده است (Razavi et al., 2012). فرامدل‌های موجود برای پیش‌بینی پیشروی آب شور در آبخوان‌های ساحلی دارای برخی معایب هستند. ساختار مدل‌های ANN (تعداد لایه‌های پنهان و تعداد نرون‌ها در لایه پنهان) نیاز به تثبیت به مقدار معینی دارند، که نیازمند ایجاد تعادل بین پیچیدگی و دقت مدل است (Sreekanth and Datta, 2011). معایب دیگر ANN طبیعت

<sup>1</sup> Artificial Neural Network (ANN)

<sup>2</sup> Genetic Programming (GP)

<sup>3</sup> Radial Basis Function (RBF)

<sup>4</sup> Support Vector Machines (SVM)

است. SEAWAT از MODFLOW (تفاضلات محدود) برای حل سیستم جریان و مدل انتقال سه بعدی مدولار MT3DMS برای حل معادلات انتقال مواد محلول استفاده می‌کند (Werner et al., 2013).

MT3DMS قادر به تلفیق پراکندگی<sup>۲</sup>، همرفت<sup>۳</sup> و واکنش‌های شیمیایی<sup>۴</sup> از گونه‌های شیمیایی محلول در سیستم‌های آب زیرزمینی است.

## ۲-۲- آبخوان ساحلی مورد استفاده

از آنجا که برای اجرای مدل شبیه‌سازی داده‌های کامل و جامعی از یک آبخوان مورد نیاز است، لذا در این پژوهش از یک آبخوان کوچک ساحلی مشخص استفاده شد.

اطلاعات مهم آبخوان ساحلی مورد نظر در جدول ۱ ارائه شده است. مرزهای این آبخوان ۲/۸۴ کیلومترمربعی عبارت‌اند از مرز بدون جریان در چپ نقشه، رودخانه به‌عنوان مرز با هد مشخص در سمت چپ پایین و مرز ساحلی با هد و غلظت مشخص. منطقه موردنظر عمدتاً شامل آبخوان سیلتی با حداکثر ضخامت ۱۰ متر در بالا و آبخوان رسی با حداکثر ضخامت ۲۰ متر در پایین است. این منطقه دارای ۳ چاه آبد<sup>۵</sup> با مقادیر پمپاژ متغیر است که درون آبخوان پایینی قرار دارند. نمای سه بعدی آبخوان با مکان چاه‌های

جدول ۱- مشخصات آبخوان ساحلی مورد استفاده

Table 1. The coastal aquifer properties

Parameter	Value
Hydraulic conductivity of upper aquifer	1.5 m/day
Hydraulic conductivity of lower aquifer	5 m/day
Recharge rate	0.001 m/day
Reference fluid density	1000 kg/m <sup>3</sup>
Density of seawater	1025 kg/m <sup>3</sup>
Concentration of seawater	19 g/L
Length of No-flow boundary	1785 m
Stream length	1230 m
Coastline length	4310 m
Number of active cells	16712

<sup>2</sup> Dispersion  
<sup>3</sup> Advection  
<sup>4</sup> Chemical reactions  
<sup>5</sup> Production Well (PW)

گاوسی را به‌عنوان مدل جانشین برای پیش‌بینی پیشروی آب شور در آبخوان ساده مستطیلی به‌کار بردند. روش آنها علی‌رغم زمان‌بر بودن کارایی بیشتری نسبت به روش‌های رگرسیون ماشین‌بردار پشتیبان و درخت‌های رگرسیونی داشت (Kopsiaftis et al., 2019).

در پژوهش حاضر قابلیت انواع مختلف مدل‌های SVR برای تقریب فرایند پیشروی آب شور در آبخوان ساحلی کوچک با اطلاعات مشخص بررسی و بهترین آنها از نظر قدرت تعمیم‌دهی و پیش‌بینی در مقایسه با مدل رایج ANN معرفی می‌شود. مدل ماشین‌بردار پشتیبان یکی از روش‌های یادگیری با نظارت است که هر دو کار رگرسیون و طبقه‌بندی را پشتیبانی می‌کند و از اصل کمینه‌سازی خطای ساختاری بهره می‌جوید (Tenne and Goh, 2010).

مدل‌های SVM با بسیاری مدل‌های تقریب دیگر به خوبی قابل مقایسه هستند زیرا به بهینه‌های محلی حساس نیستند، فرایند بهینه‌سازی آنها به ابعاد مسئله بستگی ندارد، و بیش‌برازش در آنها به ندرت اتفاق می‌افتد. SVM رگرسیونی برای ساخت مدل‌های تقریبی استفاده می‌شود. در این پژوهش به‌منظور دستیابی به مدل جانشین مؤثر و قابل اعتمادتر، انواع مختلف مدل‌های ماشین‌بردار پشتیبان رگرسیونی برای تقریب پاسخ مدل عددی شبیه‌سازی به برداشت‌های مختلف از آب زیرزمینی توسعه یافت و اجرا شد. سپس نتایج عملکرد پیش‌بینی مدل‌های SVR با استفاده از شاخص‌های ارزیابی مختلف شامل ریشه میانگین مربع خطا و ضریب کارایی نش-ساتکلیف<sup>۱</sup> و مربع ضریب همبستگی با نتایج حاصل از مدل رایج ANN مقایسه شد.

## ۲- روش پژوهش

### ۲-۱- ابزار شبیه‌سازی عددی جریان و انتقال آب زیرزمینی

در این پژوهش، پیشروی آب شور در آبخوان ساحلی با استفاده از کد عددی سه‌بعدی جریان و انتقال آب زیرزمینی SEAWAT شبیه‌سازی شد. مدل عددی SEAWAT اگرچه کاربردهای بسیاری از قبیل شبیه‌سازی ترکیبی جریان آب زیرزمینی و انتقال حرارت دارد به‌طور ویژه برای شبیه‌سازی پیشروی آب شور طراحی شده

<sup>1</sup> Nash-Sutcliffe (NS)

در این پژوهش ۶۰ نمونه آموزشی-آزمایشی<sup>۳</sup> و ۱۰۰ نمونه مجزای پیش‌بینی با مقادیر ۱۲ تایی پمپاژ گذرا با توزیع نمونه‌برداری یکنواخت با استفاده از LHS به دست آمد. حدود بالایی و پایینی این مقادیر به ترتیب ۱۰۰۰۰ و ۰ مترمکعب در روز می‌باشد.

#### ۲-۴- اجرای مدل شبیه‌سازی آبخوان ساحلی

در این پژوهش دوره مدیریتی ۷۳۰ روزه با ۴ گام زمانی مدیریتی در نظر گرفته شد. مجموعه ورودی به مدل شبیه‌سازی عددی شامل ۶۰۰ نمونه آموزشی-آزمایشی و ۱۰۰ نمونه پیش‌بینی با مقادیر ۱۲ تایی پمپاژ گذرا (۳ چاه تولید در ۴ گام زمانی مدیریتی) می‌شود که با استفاده از LHS به دست آمد. سپس مدل SEAWAT با استفاده از این مقادیر با طول گام زمانی اولیه ۱۰ روزه برای انتقال املاح اجرا شد. ۱۵ دوره تنش ۱۰ روزه برای مدل گذرای MODFLOW در نظر گرفته شد. هر شبیه‌سازی عددی در سیستم کامپیوتر با مشخصات Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz, 16 GB RAM به طور متوسط ۳۰ تا ۵۰ ثانیه برای همگرا شدن زمان می‌برد. بنابراین در مجموع ۷۰۰ مجموعه داده ورودی (مقادیر پمپاژ گذرا) و خروجی (غلظت نمک) با ۷۰۰ بار اجرای مدل شبیه‌سازی SEAWAT جمع‌آوری شد. سپس ۶۰۰ الگوی ورودی-خروجی برای آموزش و آزمایش مدل جانشین و ۱۰۰ مجموعه داده ورودی-خروجی مجزا برای ارزیابی قابلیت پیش‌بینی آن استفاده شدند.

#### ۲-۵- ماشین بردار پشتیبان رگرسیون

ماشین بردار پشتیبان در ابتدا به منظور طبقه‌بندی توسعه یافت و در سال ۱۹۹۵ توسط وپنیک به عنوان ابزاری جهت رگرسیون مطرح شد (Vapnik, 1995). SVR یک مورد خاص از ماشین‌های بردار پشتیبان است که برای ساخت مدل‌های تقریبی استفاده می‌شود. اکثر پژوهش‌های پیشین از SVM برای طبقه‌بندی و SVR برای پیش‌بینی تابع استفاده کرده‌اند. در واقع SVR به عنوان تعمیم مدل‌های درون‌یابی تابع شعاعی (RBF) شناخته می‌شود.

فرم پایه پیش‌بینی SVR، مجموع توابع پایه  $\psi^{(i)}$  با وزن  $w^{(i)}$ ، به اضافه یک عبارت پایه  $\mu$  است (Forrester et al., 2008)

آبده و چاه پایش<sup>۱</sup> منتخب (MW) در شکل ۱ نشان داده شده است.

#### ۲-۳- تولید مجموعه‌های تصادفی اولیه توسط نمونه‌برداری ابرمکعب لاتین<sup>۲</sup>

در این پژوهش با استفاده از زبان برنامه‌نویسی C++ مجموعه‌های تصادفی پمپاژ گذرا از طریق نمونه‌برداری ابرمکعب لاتین تولید شد. LHS شیوه نمونه‌برداری طبقه‌ای همگن است که از روش مونت کارلو برآمده است (Loh, 1996). با به‌کارگیری این شیوه، نقاط نمونه‌برداری در تمام فضای طراحی گسترش حداکثری می‌یابند. نمونه‌برداری ابرمکعب لاتین برای تولید نمونه‌ای با سایز n از مجموعه متغیرهای x شامل m متغیر بدین صورت عمل می‌کند که ابتدا دامنه هر متغیر به طور کامل به n فاصله هم‌احتمال تقسیم می‌شود و یک مقدار به طور تصادفی از هر فاصله انتخاب می‌شود. n مقدار به دست آمده برای متغیر اول به طور تصادفی بدون جایگزینی با n مقدار به دست آمده برای متغیر دوم جفت می‌شود. این n جفت نیز به طور تصادفی بدون جایگزینی با n مقدار از متغیر سوم ترکیب می‌شوند و n نمونه سه‌تایی تشکیل می‌دهند. این فرایند ادامه می‌یابد تا اینکه مجموعه n نمونه m تایی به صورت زیر شکل می‌یابد و به این ترتیب نمونه ابرمکعب لاتین ساخته می‌شود

$$x_i = [x_{i1}, x_{i2}, \dots, x_{im}] \quad i=1,2,\dots,n \quad (1)$$

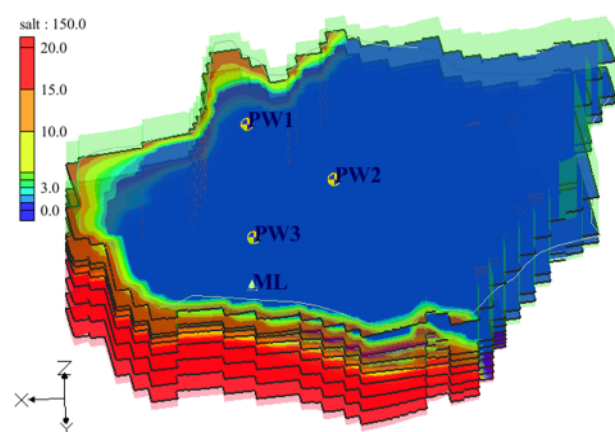


Fig. 1. Salinity concentration contours at the end of simulation period for one set of transient pumping from three production wells

شکل ۱- خطوط تراز غلظت شوری در انتهای دوره شبیه‌سازی برای یک مجموعه از مقادیر پمپاژ گذرا از چاه‌های آبده

<sup>1</sup> Monitoring Well (MW)

<sup>2</sup> Latin Hypercube Sampling (LHS)

<sup>3</sup> Training-testing

از تابع کرنل  $\psi$  برای حل مسائل غیر خطی استفاده می‌شود. با استفاده از تابع کرنل فضای ورودی به یک فضای چند بعدی نگاشته می‌شود که در این فضا یک ابر صفحه که مجموع کناره دو طرف و مجموع جریمه‌ها را به حداقل می‌رساند پیدا می‌شود

$$\hat{f}(x) = \mu + \sum_{i=1}^n (\alpha^{+(i)} - \alpha^{-i}) \psi(x^{(i)}, x^{(i)}) \quad (5)$$

متداول‌ترین توابع کرنل یا توابع پایه عبارت است از تابع کرنل چند جمله‌ای، تابع کرنل گاوسی و تابع کرنل خطی. تابع کرنل گاوسی یا RBF به علت واکنش‌های محدود و محلی در میان تمام دامنه محور  $x$  واقعی، محبوب‌ترین نوع کرنل مورد استفاده در SVMها بوده و به صورت زیر تعریف می‌شود

$$\psi(x^{(i)}, x^{(i)}) = \exp\left(\frac{-|x^{(i)} - x^{(i)}|^2}{\sigma^2}\right) \quad (6)$$

دقت در تخمین ابر پارامترهای  $C$ ،  $\varepsilon$  و  $\sigma$  بسیار مهم بوده و در کاهش خطای مسئله تأثیر به‌سزایی دارد. بنابراین برای دستیابی به دقت مطلوب از الگوریتم رایج بهینه‌سازی کمینه متوالی<sup>۳</sup> برای حل برنامه‌ریزی درجه دوم هنگام آموزش مدل‌های ماشین بردار پشتیبان استفاده شد (Cristianini and Shawe-Taylor, 2014).

واریانس کرنل گاوسی  $\sigma^2$  باید با کمینه‌سازی RMSE پیش‌بینی تنظیم شود. از آنجا که مقیاس  $y$  بر مقدار بهینه  $C$  تأثیر دارد، برای شروع  $y$  به مقادیر بین صفر و یک نرمال‌سازی می‌شود. بنابراین برای مطالعه حاضر سه نوع متداول مدل ماشین بردار پشتیبان با استفاده از محیط برنامه نویسی MATLAB (R2018a) ساخته شد: مدل رگرسیونی ماشین بردار پشتیبان خطی، مدل رگرسیونی ماشین بردار پشتیبان چندجمله‌ای، مدل رگرسیونی ماشین بردار پشتیبان گاوسی. شبه مدل رایج شبکه عصبی مصنوعی نیز برای ارزیابی کارایی مدل‌ها آموزش داده شد. پس از ساخت مدل‌ها توسط داده‌های آموزشی، اعتبارسنجی مدل‌ها با استفاده از داده‌های آزمایشی و مقایسه نتایج به دست آمده با نتایج حاصل از مدل شبه‌ساز اصلی توسط معیارهای ارزیابی مختلف انجام پذیرفت.

$$\hat{f}(x) = \mu + \sum_{i=1}^n w^{(i)} \psi(x, x^{(i)}) \quad (2)$$

پاسخ تمام نمونه‌های آموزشی حاصل از مدل SVR ساخته شده باید حداکثر به میزان  $\varepsilon$  از مقادیر تابع هدف اصلی  $y^{(i)}$  انحراف داشته و در عین حال مدل باید تا حد ممکن هموار<sup>۱</sup> باشد. بنابراین برای انجام یک پیش‌بینی با تعمیم خوب، باید تابع با حداکثر انحرافات  $\varepsilon$  از  $y$  و در عین حال حداقل پیچیدگی را پیدا کرد. با تعریف متغیرهای کمکی مثبت  $\xi^{-(i)}$  و  $\xi^{+(i)}$  مسئله بهینه‌سازی درجه دوم مقید به صورت زیر برقرار می‌شود

$$\text{Minimize } \frac{1}{2} |w|^2 + C \frac{1}{n} \sum_{i=1}^n (\xi^{+(i)} + \xi^{-(i)}) \quad (3)$$

$$\text{Subject to } \begin{cases} y^{(i)} - w \cdot x^{(i)} - \mu \leq \varepsilon + \xi^{+(i)} \\ w \cdot x^{(i)} + \mu - y^{(i)} \leq \varepsilon + \xi^{-(i)} \\ \xi^{+(i)}, \xi^{-(i)} \geq 0 \end{cases}$$

که در معادله بالا

ثابت  $C$  ضریبی است که میان پیچیدگی مدل و درجه خطای آن توازن برقرار می‌کند. به طوری که هرچه قدر میزان ضریب  $C$  بیشتر باشد، معیار هموار بودن تابع  $y^{(i)}$  تأثیر کمتری خواهد داشت. نقاطی که داخل تیوپ  $\varepsilon$  قرار دارند هیچ‌زانی ندارند، در حالی که زیان پیش‌بینی نقاط خارج به صورت خطی با نرخ  $C$  افزایش می‌یابد. پیش‌بینی SVR با استفاده از روش دوگان‌سازی استاندارد<sup>۲</sup> و محاسبه ضرایب لاگرانژ  $\alpha^{+(i)}, \alpha^{-i} \geq 0$  در حالت رگرسیون خطی به صورت زیر تعریف می‌شود

$$\hat{f}(x) = \mu + \sum_{i=1}^n (\alpha^{+(i)} - \alpha^{-i}) \psi(x^{(i)}, x) \quad (4)$$

که در معادله بالا نمونه‌هایی که ضرایب لاگرانژ  $(\alpha^{+(i)} - \alpha^{-i})$  غیر صفر دارند بردار پشتیبان نامیده می‌شوند. به منظور محاسبه  $w$  و  $\hat{f}(x)$  نیازی به استفاده از همه داده‌های آموزشی نیست و تنها از نمونه‌هایی که بردارهای پشتیبان را تشکیل می‌دهند، استفاده می‌شود.

<sup>1</sup> Flat

<sup>2</sup> Standard Dual Problem Method

<sup>3</sup> Sequential minimal optimization (SMO)



$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (C_n^O - C_n^P)^2} \quad (7)$$

$$NS = 1 - \frac{\sum_{n=1}^N (C_n^O - C_n^P)^2}{\sum_{n=1}^N (C_n^O - C^O)^2} \quad (8)$$

$$R^2 = \left( \frac{\sum_{n=1}^N (C_n^O - C^O)(C_n^P - C^P)}{\sqrt{\sum_{n=1}^N (C_n^O - C^O)^2} \sqrt{\sum_{n=1}^N (C_n^P - C^P)^2}} \right)^2 \quad (9)$$

که در این معادلات

$C_n^O$  و  $C_n^P$  به ترتیب مقادیر مشاهده شده (حاصل از مدل SEAWAT) و پیش‌بینی شده (حاصل از مدل جانشین SVR) غلظت آب دریا،  $C^O$  و  $C^P$  به ترتیب مقادیر میانگین غلظت آب شور مشاهده شده و پیش‌بینی شده؛ و  $N$  تعداد نقاط داده است (Roy and Datta, 2017a).

### ۳- نتایج و بحث

#### ۳-۱- عملکرد مراحل آموزشی و آزمایشی

نتایج ارزیابی عملکرد مدل‌های SVR در مراحل آموزشی و آزمایشی در جدول ۲ آورده شده است. RMSE ریشه میانگین مربعات خطا بین مقادیر مشاهده شده (حاصل از مدل SEAWAT) و پیش‌بینی شده یا خروجی از مدل جانشین را نشان می‌دهد. هر چه مقدار RMSE به صفر نزدیک‌تر باشد، دقت تخمین مدل بالاتر است. پارامتر آماری دیگر که توانایی مدل هیدرولوژیکی را برای پیش‌بینی رفتار سیستم ارزیابی می‌کند، NS است. هنگامی که مقدار NS مساوی با یک باشد توانایی پیش‌بینی مدل ۱۰۰ درصد کامل است.  $R^2$  مقدار همبستگی بین مقادیر غلظت آب شور واقعی و پیش‌بینی شده را نمایش می‌دهد. هر چه مقدار  $R^2$  به یک نزدیک‌تر باشد، دقت مدل جایگزین بالاتر است.

جدول ۲- نتایج ارزیابی مدل‌های جانشین مختلف در مراحل آموزش و آزمایشی

Table 2. The training and testing results of different types of surrogate models

Surrogate model	Runtime (s)	Training data			Testing data		
		$R^2$ (%)	NS	RMSE	$R^2$ (%)	NS	RMSE
Linear SVR	0.12	0.74	0.71	0.93	0.72	0.67	1.08
Polynomial SVR	5.26	0.98	0.98	0.22	0.98	0.98	0.24
Gaussian SVR	0.10	0.96	0.93	0.44	0.92	0.87	0.68
ANN	1.27	1.00	1.00	0.06	0.89	0.88	0.64

#### ۲-۶- ساخت مدل جانشین

مدل عددی جریان و انتقال آب زیرزمینی برای تولید مجموعه الگوهای ورودی-خروجی برای استفاده در مراحل ساخت و پیش‌بینی مدل جانشین استفاده شد. این مجموعه داده‌های تولید شده توسط مدل شبیه‌سازی اصلی به‌طور تصادفی بدون جایگزینی به داده‌های آموزشی و آزمایشی تقسیم شدند. مجموعه داده‌های آموزشی برای درک رابطه بین الگوهای پمپاژ-غلظت و ساخت مدل جانشین و مجموعه داده‌های آزمایشی برای آزمایش عملکرد مدل استفاده می‌شوند. برای انجام اعتبارسنجی متقابل از نسبت ۸۰ به ۲۰ برای تفکیک تصادفی بدون جایگزینی داده‌های آموزشی و آزمایشی در مرحله ساخت مدل جانشین استفاده شد (Lin et al., 2008).

بنابراین از ۶۰۰ مجموعه داده، ۴۸۰ نمونه برای آموزش و ۱۲۰ نمونه برای آزمایش مدل به‌کار گرفته شدند. هنگامی که مدل آموزش یافت و آزمایش شد نمونه‌های آزمایشی جداگانه‌ای از پمپاژ گذرا (۱۰۰ مجموعه داده حاصل از نمونه‌برداری ابرمکعب لاتین) برای پیش‌بینی مقادیر غلظت نمک در چاه‌های پایب به مدل‌ها وارد شد. غلظت خروجی از مدل شبیه‌سازی اصلی برای آموزش مدل جانشین استفاده شد در حالی که در مرحله پیش‌بینی به مدل جانشین وارد نشد. چهار نوع مدل جانشین توضیح داده شده برای پیش‌بینی غلظت شوری در چاه پایب منتخب استفاده شد. غلظت پیش‌بینی و غلظت حاصل از مدل شبیه‌سازی عددی برای ارزیابی قابلیت پیش‌بینی مدل‌های ایجاد شده مقایسه شد.

#### ۲-۷- معیارهای عملکرد مدل

عملکرد مدل جانشین شبیه‌ساز جریان و انتقال آب زیرزمینی با استفاده از شاخص‌های RMSE، NS و  $R^2$  ارزیابی شد. مقادیر کوچک‌تر RMSE و مقادیر بزرگ‌تر NS و  $R^2$  دقت بهتر تخمین مدل را نشان می‌دهند.

شکل ۲ مقادیر غلظت شوری واقعی و پیش‌بینی شده توسط انواع مختلف مدل‌های توسعه یافته و نمودار رگرسیون متناظر با آن‌ها را برای مجموعه داده‌های آموزشی و آزمایشی نشان می‌دهد. به منظور ارزیابی عملکرد مدل‌ها، مقادیر غلظت شبیه‌سازی شده توسط مدل عددی SEAWAT دقیق فرض شده و به عنوان مقادیر غلظت واقعی در نظر گرفته شده‌اند. مقدار غلظت پیش‌بینی شده همان خروجی حاصل از مدل‌های جانشین هستند. زمان همگرایی در مدل‌های SVR بسیار کمتر از مدل شبیه‌سازی اصلی است. فرایند آموزش و آزمایش مدل جانشین SVR تنها چند ثانیه برای کل داده‌ها طول می‌کشد در حالی که زمان همگرایی مدل شبیه‌سازی SEAWAT برای تنها هر نمونه داده ۳۰ تا ۵۰ ثانیه زمان می‌برد. قید زمان یکی از عوامل اصلی انتخاب مدل جانشین است که مدل SVR را به روشی کارآمد و کاربردی برای جانشینی مدل شبیه‌سازی اصلی در طی فرایند بهینه‌سازی تبدیل می‌کند. همان‌طور که در جدول ۲ مشاهده می‌شود مدل SVR گاوسی سریع‌تر از سایر مدل‌ها همگرا شده اما نتایج برتری ارائه نکرده است. مدل SVR چندجمله‌ای علی‌رغم دقت بیشتر، زمان پاسخ بیشتری نسبت به سایر مدل‌های ارائه شده داشته است.

### ۳-۲- قابلیت پیش‌بینی مدل SVR

عملکرد مدل‌های SVR برحسب دقت پیش‌بینی با استفاده از مجموعه داده‌های متفاوت برای پیش‌بینی ارزیابی شد. شکل ۳ نمودار پراکندگی غلظت شبیه‌سازی شده حاصل از مدل عددی را در برابر غلظت پیش‌بینی شده توسط مدل‌های جانشین مختلف برای مجموعه داده‌های جدید پیش‌بینی نشان می‌دهد. این شکل نشان می‌دهد که تخمین‌های SVR چندجمله‌ای به مقادیر غلظت شوری واقعی در چاه پایش خیلی نزدیک است. نمودارهای رگرسیون مقدار همبستگی بین مقادیر غلظت شوری واقعی و پیش‌بینی شده را برای تمامی مدل‌ها نشان می‌دهد. مقدار  $R^2$  نشان داده شده در شکل، همبستگی مثبت قوی مدل یادشده را بین مقادیر غلظت شوری واقعی و پیش‌بینی شده نشان می‌دهد. معادلات خطی نمودارهای رگرسیون نشان می‌دهند که مقادیر ضرایب  $a$  و  $b$  در معادله  $y=ax+b$  با مقدار  $R^2$  زیاد به ترتیب به ۱ و ۰ بسیار نزدیک می‌باشند. همان‌طور که در شکل ۳ مشاهده می‌شود قابلیت تعمیم‌دهی

همان‌طور که جدول ۲ نشان می‌دهد، عملکرد مدل‌های SVR در مراحل آموزش و آزمایش برای سه معیار ارزیابی روند مشابهی را نشان می‌دهد. مقدار RMSE داده‌های آموزشی و آزمایشی با استفاده از تابع کرنل گاوسی نسبت به SVR خطی کمتر شده و با استفاده از تابع کرنل چندجمله‌ای به صفر نزدیک می‌شود. در مدل شبکه عصبی پرسپترون چندلایه از پنج نرون در لایه پنهان آن استفاده شده است. مدل ANN هر چند کمترین مقدار RMSE را برای داده‌های آموزشی (۰/۰۲) داشته و تا حد ممکن بر داده‌های آموزشی انطباق یافته است، اما عملکرد نسبتاً ضعیف‌تری نسبت به مدل SVR چندجمله‌ای بر داده‌هایی که در هنگام آموزش استفاده نشده بودند، داشته است.

همان‌طور که در جدول ۲ مشاهده می‌شود مدل SVR چندجمله‌ای کمترین مقدار RMSE را در داده‌های آزمایشی و پیش‌بینی داشت. این نتیجه می‌تواند به علت استعداد کمتر بیش‌برازش مدل SVR نسبت به ANN باشد زیرا مدل SVR مبتنی بر اصل کمینه‌سازی خطای ساختاری است که نه فقط خطای تجربی بلکه پیچیدگی مدل را نیز کمینه می‌کند. هر چه پیچیدگی مدل SVR کمتر باشد، می‌تواند داده‌هایی که در آموزش استفاده نشده‌اند را تعمیم نماید (Chevalier et al., 2011). با مقایسه مقادیر NS و  $R^2$  مدل‌های مختلف مورد مطالعه مشاهده می‌شود که مقدار آن‌ها برای مدل SVR چندجمله‌ای بیشترین است. بنابراین این مدل بیشترین توانایی را برای پیش‌بینی در بین سایر مدل‌ها دارد. شو و واردا در سال ۲۰۰۸ در مسئله استفاده از سیستم استنتاج نوروفازی انطباقی برای تخمین وقوع سیل بیان کردند که دقت پیش‌بینی مدل قابل قبول فرض می‌شود اگر NS مقدار بزرگ‌تر یا مساوی با ۰/۸ داشته باشد (Shu and Ouarda, 2008). مقدار NS برای مدل SVR خطی کمتر از ۰/۷۱ است بنابراین مدل ساده SVR خطی دقت قابل قبولی برای مسئله مورد بررسی نداشته است. اما مقدار NS سایر مدل‌های مورد مطالعه در محدوده قابل قبول قرار داشته و این مدل‌ها می‌توانند برای پیش‌بینی پیشروی آب شور استفاده شوند. نتایج پیش‌بینی به‌دست آمده نشان می‌دهد که مدل SVR چندجمله‌ای بیشترین دقت پیش‌بینی را داشته و قادر به تقلید کارآمد مدل شبیه‌سازی عددی پیچیده جریان و انتقال آب زیرزمینی است.



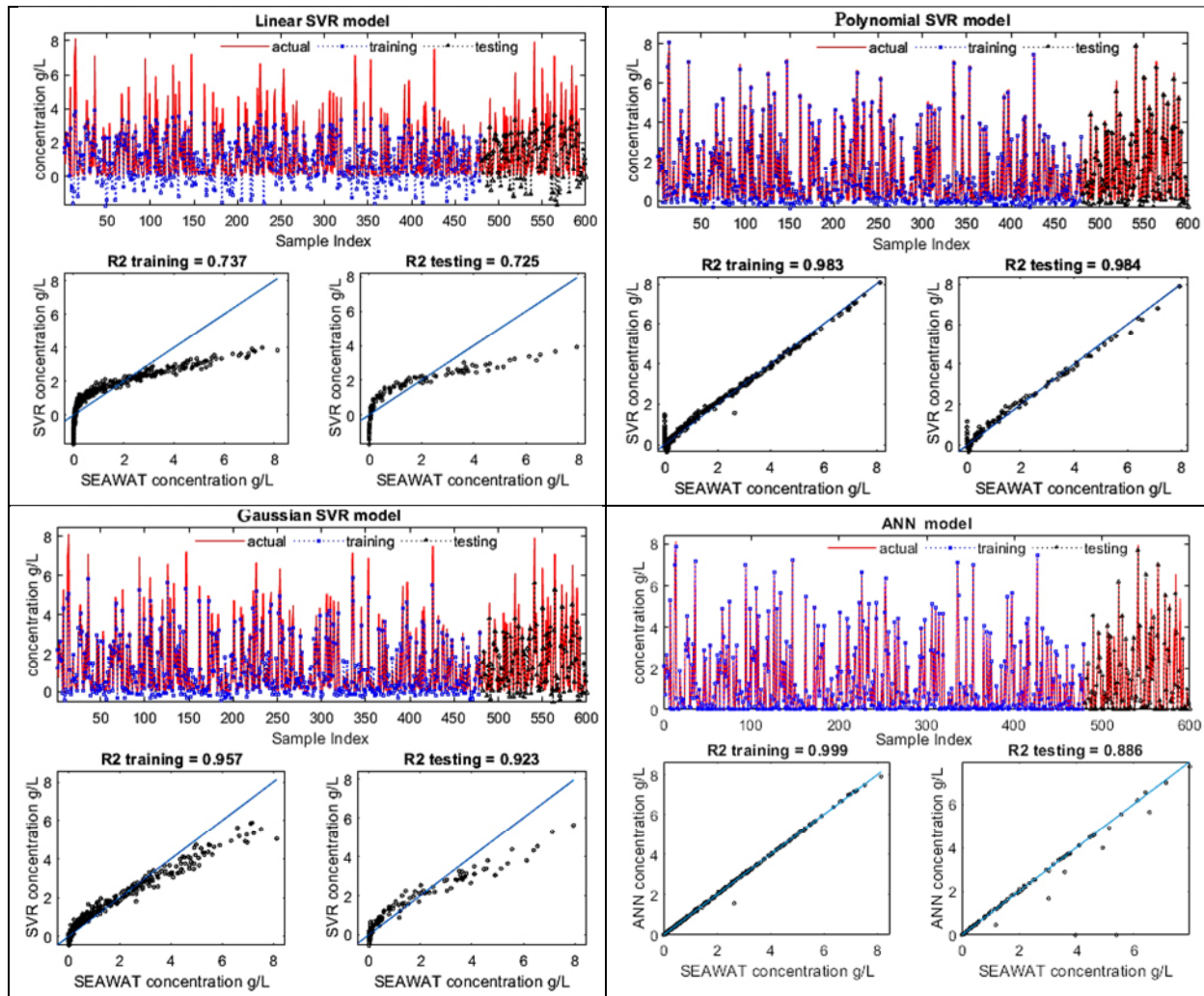


Fig. 2. The training and testing results of different types of surrogate models: linear SVR, polynomial SVR, gaussian SVR, ANN

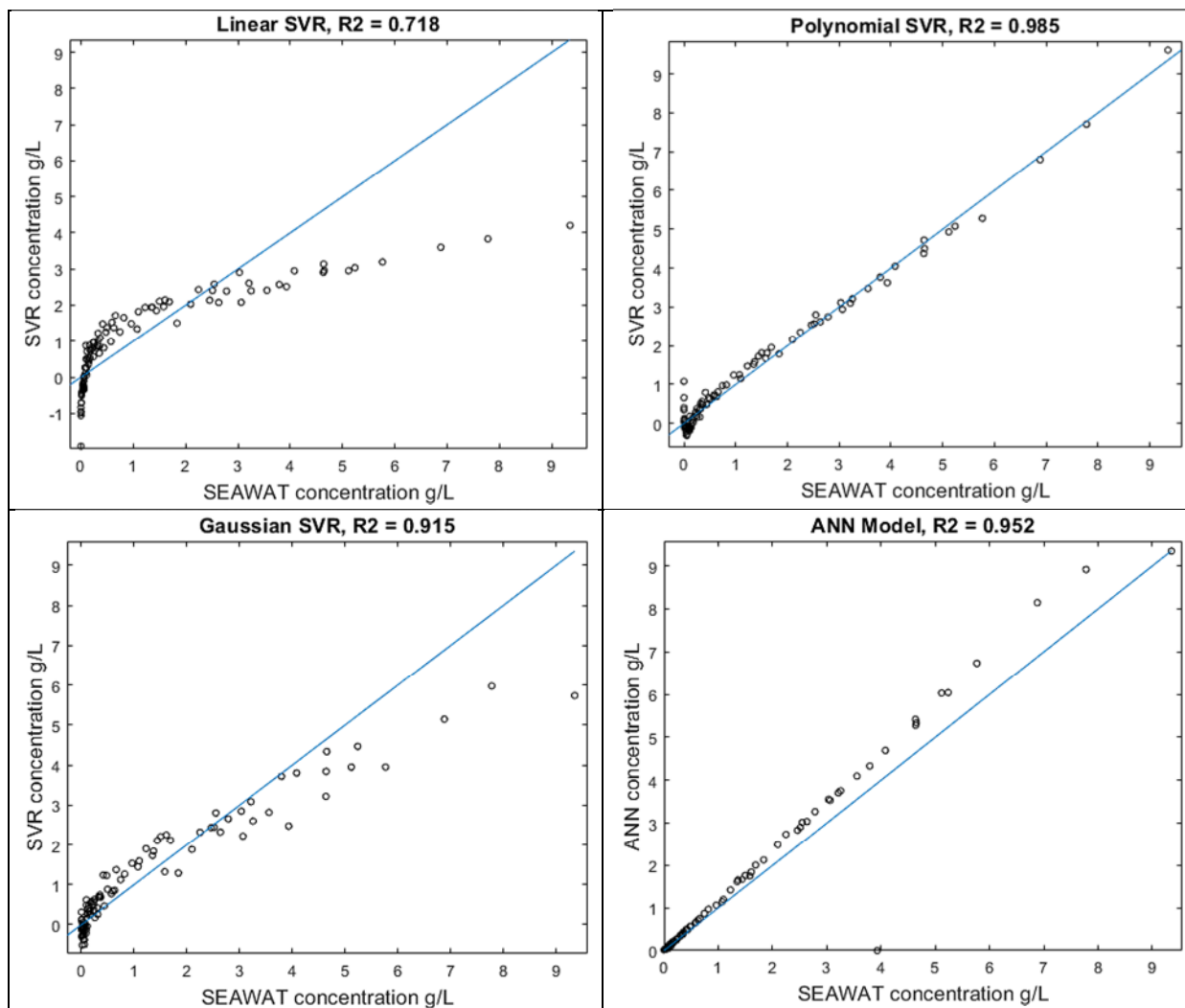
شکل ۲- نتایج مدل‌های جانشین مختلف در مراحل آموزشی و آزمایشی: SVR خطی، SVR چندجمله‌ای، SVR گاوسی، ANN

قبول مقادیر غلظت شوری توسط این مدل در مسئله مورد بررسی است.

مدل محبوب شبکه عصبی مصنوعی هرچند دارای مقدار NS قابل قبول برای داده‌های جدید پیش‌بینی است اما دقت کمتری نسبت به مدل SVR چندجمله‌ای داشته است. بنابراین مدل SVR چندجمله‌ای به خوبی آموزش دیده و می‌تواند برای پیش‌بینی غلظت شوری در چاه‌های پایش استفاده شود. قابلیت پیش‌بینی مدل‌های جانشین وابستگی کامل به ابعاد مسئله دارد و نیاز است بررسی‌های بیشتری برای استفاده از این گونه مدل‌ها در مسائل بزرگ مقیاس انجام شود.

مدل شبکه عصبی مصنوعی نسبت به مدل SVR چندجمله‌ای برای مجموعه داده‌های جدید کمتر بوده است.

نتایج پیش‌بینی مدل‌های مختلف در جدول ۳ نیز ارائه شده است. نتایج پیش‌بینی با نتایج آموزشی و آزمایشی برحسب برآورد خطاهای محاسبه شده همسو هستند. مقادیر  $RMSE$ ،  $NS$  و  $R^2$  برای مدل SVR چندجمله‌ای به ترتیب برابر  $0.23$ ،  $0.99$  و  $0.99$  هستند. مقدار  $RMSE$  مدل SVR خطی به علت استفاده از تابع خطی و برآزش ضعیف آن بیشتر از سایر مدل‌ها بوده و دقت پیش‌بینی قابل قبولی ندارد. اما مقدار  $NS$  برای مدل SVR چندجمله‌ای بیشتر از  $0.99$  است که نشانگر قابلیت پیش‌بینی قابل



**Fig. 3.** The forecasting results of different types of surrogate models: linear SVR, polynomial SVR, gaussian SVR, ANN

شکل ۳- نتایج مدل‌های جانشین مختلف برای داده‌های مجزای پیش‌بینی: SVR خطی، SVR چندجمله‌ای، SVR گاوسی، ANN

جدول ۳- نتایج ارزیابی مدل‌های جانشین مختلف در مرحله پیش‌بینی

**Table 3.** The forecasting results of different types of surrogate models

Surrogate model	Forecasting data		
	R <sup>2</sup> (%)	NS	RMSE
Linear SVR	0.72	0.68	1.07
Polynomial SVR	0.99	0.99	0.23
Gaussian SVR	0.91	0.88	0.64
ANN	0.95	0.94	0.51

#### ۴- نتیجه‌گیری

چارچوب شبیه‌سازی-بهینه‌سازی برای امکان‌پذیری محاسباتی مدل مدیریت بهینه به الگوریتم بهینه‌سازی اتصال یابد. در این پژوهش مدل‌های SVR مختلف برای پیش‌بینی فرایند پیشروی آب شور در

پژوهش حاضر نشان می‌دهد که مدل SVR چندجمله‌ای جانشین قابل قبولی برای مدل شبیه‌سازی عددی در مسئله پیش‌بینی پیشروی آب شور است. مدل جانشین آموزش یافته می‌تواند در

بهینه‌سازی را کاهش دهد. نتایج به‌دست آمده از این پژوهش اثبات می‌کند که مدل‌های SVR چندجمله‌ای می‌توانند با دقت بیشتری نسبت به مدل رایج شبکه عصبی مصنوعی برای مسائل پیش‌بینی پیشروی آب شور و مطالعات مدیریت آن در مسائل با ابعاد کوچک به‌کار گرفته شوند. در پژوهش‌های بعدی لازم است کارایی این مدل‌ها در مسائل با ابعاد متوسط و بزرگ مورد بررسی قرار گیرد و همچنین مدل‌های جانشین قابل اعتماد در قالب یک مسئله شبیه‌ساز-بهینه‌ساز برای مدیریت بهینه برداشت آب زیرزمینی از آبخوان ساحلی استفاده شود.

### ۵- قدردانی

به این وسیله از حمایت مالی معاونت پژوهش و فناوری دانشگاه شهید چمران اهواز در قالب پژوهانه (Grant No. SCU EC98.150) در انجام این پژوهش قدردانی می‌شود.

آبخوان ساحلی در واکنش به الگوهای برداشت متفاوت از چاه‌های آبدار ساخته شدند. نتایج ارزیابی عملکرد مدل‌های SVR در مقایسه با مدل رایج شبکه عصبی مصنوعی نشان داد که مدل SVR چندجمله‌ای بهترین عملکرد را در بین سایرین داشته است. مدل SVR چندجمله‌ای ابزار قوی و مؤثری برای پیش‌بینی قابل اعتماد و دقیق پیشروی آب شور به آبخوان ساحلی می‌باشد و می‌تواند به‌عنوان جانشین مدل پیچیده عددی در زمینه مطالعات مدیریت آبخوان ساحلی استفاده شود. به‌طور کلی نتایج ارزیابی پیشنهاد می‌کند که مدل‌های پیش‌بینی SVR چندجمله‌ای می‌توانند در مطالعات مدیریت منابع آب زیرزمینی به‌عنوان جانشین قابل قبول و کارآمدی برای مدل عددی پیچیده استفاده شوند. مزیت به‌کارگیری مدل جانشین SVR زمان آموزشی و آزمایشی کم آن نسبت به مدل شبیه‌سازی عددی اصلی است. استفاده از مدل‌های جانشین SVR می‌تواند به‌طور معنی‌داری بار محاسباتی مسائل

### References

- Bhattacharjya, R. K. & Datta, B. 2005. Optimal management of coastal aquifers using linked simulation optimization approach. *Water Resources Management*, 19, 295-320.
- Bhattacharjya, R. K., Datta, B. & Satish, M. G. 2007. Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers. *Journal of Hydrologic Engineering*, 12, 273-282.
- Chevalier, R. F., Hoogenboom, G., McClendon, R. W. & Paz, J. A. 2011. Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks. *Neural Comput and Applic*, 20, 151-159.
- Chitsazan, N. 2014. *Bayesian saltwater intrusion prediction and remediation design under uncertainty*. PhD Thesis, Louisiana State University, USA.
- Christelis, V. & Mantoglou, A. 2016. Pumping optimization of coastal aquifers assisted by adaptive metamodeling methods and radial basis functions. *Water Resources Management*, 30(15), 5845-5859.
- Cristianini, N. & Shawe-taylor, J. 2014. *An introduction to support vector machines and other Kernel-Based learning methods*, Cambridge University Press.
- Dhar, A. & Datta, B. 2009. Saltwater intrusion management of coastal aquifers. I: linked simulation-optimization. *Journal of Hydrologic Engineering*, 14, 1263-1272.
- Forrester, A., Sobester, A. & Keane, A. 2008. *Engineering design via surrogate modelling: a practical guide*, John Wiley & Sons. N.Y.
- Kopsiaftis, G., Protopapadakis, E., Voulodimos, A., Doulamis, N. & Mantoglou, A. 2019. Gaussian process regression tuned by bayesian optimization for seawater intrusion prediction. *Computational Intelligence and Neuroscience*, 10, 1-12.
- Kourakos, G. & Mantoglou, A. 2009. Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. *Advances in Water Resources*, 32 (4), 507-521.

- Lal, A. & Datta, B. 2018. Development and implementation of support vector machine regression surrogate models for predicting groundwater pumping-Induced saltwater intrusion into coastal aquifers. *Water Resources Management*, 32(7), 2405-25419.
- Lin, R. S., Horn, S. D., Hurdle, J. F. & Goldfarb-Rumyantzev, A. S. 2008. Single and multiple time-point prediction models in kidney transplant outcomes. *Journal of Biomedical Informatics*, 41, 944-952.
- Loh, W.-L. 1996. On Latin hypercube sampling. *The Annals of Statistics*, 24, 2058-2080.
- Papadopoulou, M. P., Nikolos, I. K. & Karatzas, G. P. 2010. Computational benefits using artificial intelligent methodologies for the solution of an environmental design problem: saltwater intrusion. *Water Science Technology*, 62, 1479-1490.
- Pillay, N. 2004. *An investigation into the use of genetic programming for the induction of novice procedural programming solution algorithms in intelligent programming tutors*. University of KwaZulu-Natal Durban.
- Rao, S., Sreenivasulu, V., Bhallamudi, S. M., Thandaveswara, B. & Sudheer, K. 2004. Planning groundwater development in coastal aquifers. *Hydrological Sciences Journal*, 49, 155-170.
- Razavi, S., Tolson, B. A. & Burn, D. H. 2012. Numerical assessment of metamodelling strategies in computationally intensive optimization. *Environmental Modelling Software*, 34, 67-86.
- Roy, D. K. 2018. *Development of a sustainable groundwater management strategy and sequential compliance monitoring to control saltwater intrusion in coastal aquifers*. PhD Thesis, James Cook University, Australia.
- Roy, D. K. & Datta, B. 2017a. Fuzzy C-mean clustering based inference system for saltwater intrusion processes prediction in coastal aquifers. *Water Resources Management*, 31, 355-376.
- Roy, D. K. & Datta, B. 2017b. Multivariate adaptive regression spline ensembles for management of multilayered coastal aquifers. *Journal of Hydrologic Engineering*, 22, 04017031.
- Roy, D. K. & Datta, B. 2017c. Optimal management of groundwater extraction to control saltwater intrusion in multi-layered coastal aquifers using ensembles of adaptive neuro-fuzzy inference system. *World Environmental and Water Resources Congress*. 139-150.
- Shu, C. & Ouarda, T. 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology*, 349, 31-43.
- Srekanth, J. & Datta, B. 2010. Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. *Journal of Hydrology*, 393, 245-256.
- Srekanth, J. & Datta, B. 2011. Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management. *Water Resources Management*, 25, 3201-3218.
- Tenne, Y. & Goh, C.-K. 2010. *Computational intelligence in expensive optimization problems*, Springer Science & Business Media, Switzerland.
- Vapnik, V. 1995. *The nature of statistical learning theory*, Springer, New York.
- Werner, A. D., Bakker, M., Post, V. E., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., et al. 2013. Seawater intrusion processes, investigation and management: recent advances and future challenges. *Advances in Water Resources*, 51, 3-26.
- Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O. & Lee, K.-K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology*, 396, 128-138.
- Yoon, H., KIM, Y., Ha, K., Lee, S.-H. & Kim, G.-P. 2017. Comparative evaluation of ANN- and SVM-time series models for predicting freshwater-saltwater interface fluctuations. *Water*, Doi: 36390/w9050323.

