

مدلسازی و پیش‌بینی اکسیژن مورد نیاز بیولوژیکی (BOD) با استفاده از ترکیب ماشین بردار پشتیبان با تبدیل موجک

اباذر سلگی^{۱*} - امیر پورحقی^۲ - حیدر زارعی^۳ - هادی انصاری^۴

تاریخ دریافت: ۱۳۹۴/۱۰/۰۹

تاریخ پذیرش: ۱۳۹۵/۰۲/۰۶

چکیده

آلودگی‌های شیمیایی آب‌های سطحی یکی از موضوعات جدی است که کیفیت این‌گونه آب‌ها را تهدید می‌کند. این مطلب برای آب‌هایی که به طور مستقیم به مصارف زندگی بشر می‌رسند اهمیتی چند برابر بخشیده است. یکی از پارامترهایی مهمی که برای سنجش آلودگی آب استفاده می‌شود شاخص BOD می‌باشد. در این مطالعه، توانایی مدل ماشین بردار پشتیبان (SVM) به منظور مدل‌سازی و پیش‌بینی اکسیژن مورد نیاز بیولوژیکی (BOD) در رودخانه کارون واقع در غرب کشور ایران مورد ارزیابی قرار گرفت. به منظور بررسی مدل‌ها به صورت ترکیبی، از تبدیل موجک استفاده شد. بعد از تجزیه پارامترها با تبدیل موجک، با استفاده از روش تجزیه به مولفه‌های اصلی (PCA) مولفه‌های مهم تعیین شدند. سپس از این مولفه‌های مهم به عنوان ورودی به مدل ماشین بردار پشتیبان استفاده شد تا مدل ترکیبی ماشین بردار پشتیبان-موجک (WSVM) حاصل گردید. جهت انجام این تحقیق از سری زمانی ماهانه BOD رودخانه کارون در ایستگاه ملائانی و متغیرهای کمکی اکسیژن محلول (DO)، جریان رودخانه و دمای ماهانه در یک دوره ۱۳ ساله (۱۳۹۳-۱۳۸۱) استفاده شد. نتایج بدست آمده حاکی از آن بود که مدل SVM دارای ضریب تبیین ۰/۸۴ و جذر میانگین مربعات خطای ۰/۳۳۸ (میلی گرم بر لیتر) می‌باشد و اعمال تبدیل موجک روی داده‌های ورودی مدل باعث بهبود نتایج تا ضریب تبیین ۰/۹۴ و جذر میانگین مربعات خطای ۰/۲۱۰ (میلی گرم بر لیتر) شد. بنابراین ترکیب ماشین بردار پشتیبان با تبدیل موجک یک ایده جدید برای پیش‌بینی مقدار BOD رودخانه کارون می‌باشد. در پایان مقدار BOD برای یک دوره شش ماهه با استفاده از مدل WSVM پیش‌بینی شد.

واژه‌های کلیدی: پیش‌بینی BOD، PCA، مدل ترکیبی، رودخانه کارون

مقدمه

بیولوژیکی در آب است. روش‌های متداولی که برای تعیین BOD مورد استفاده قرار می‌گیرد، اغلب روش‌هایی بسیار مشکل و البته همراه با خطاهای اندازه‌گیری است. مبنای این روش‌ها، تعیین فاکتورهای پیچیده مانند اکسیژن مورد نیاز برای تنفس میکروارگانیسم‌ها در یک نمونه و اکسیداسیون آمونیاک است (۳). با این حال، از آنجایی که متغیرهای زیادی بر روی پارامترهای کیفیت آب مؤثر هستند و یک رابطه غیرخطی و پیچیده بین آنها برقرار است، روش‌های معمول به خوبی نمی‌توانند مسئله مدیریت کیفی منابع آب را حل کنند (۴ و ۵). چندین سال است که روش‌های هوش مصنوعی برای پیش‌بینی سری‌های زمانی غیرخطی استفاده می‌شود و عملکرد خوبی هم از آنها گزارش شده است (۶ و ۷). به تازگی، تبدیل موجک که یک روش پردازشی است، عملکرد خوبی از خود نشان داده است و در مدل‌سازی‌های هیدرولوژیکی به طور گسترده مورد استفاده قرار گرفته است (۸، ۹ و ۱۰). در زیر به برخی از مطالعات صورت گرفته در زمینه استفاده از BOD اشاره می‌شود:

اکراتوس و همکاران (۱۱)، یک مدل شبکه عصبی مصنوعی و

استریت و فلیس (۱) از نخستین کسانی بودند که مدل‌سازی کیفیت آب رودخانه و بحث اندازه‌گیری پارامترهای BOD و DO را انجام دادند (۱). BOD، مقدار اکسیژن مورد نیاز میکروارگانیسم‌های هوازی موجود در نمونه آب برای اکسیداسیون مواد آلی است که با توجه به استانداردهای بین‌المللی در دمای ۲۰ درجه سانتیگراد اندازه‌گیری شده و اندازه‌گیری آن ۵ روز به طول می‌انجامد (۲).

BOD یک اصطلاح برای مقدار اکسیژن مورد نیاز برای تجزیه بیولوژیکی یک ماده آلی در یک نمونه آب می‌باشد. بنابراین اندازه‌گیری BOD به عنوان پایه آشکارسازی مواد آلی قابل تجزیه

۱، ۲ و ۳- به ترتیب دانشجویان دکتری مهندسی منابع آب و استادیار گروه هیدرولوژی و منابع آب، دانشکده مهندسی علوم آب، دانشگاه شهید چمران اهواز (*- نویسنده مسئول: Email: A-solgi@phdstu.scu.ac.ir)
۴- دانشجوی دکتری مهندسی منابع آب، دانشکده کشاورزی، دانشگاه بوعلی سینا
DOI: 10.22067/jsw.v31i1.52080

برآورد نیاز BOD در رودخانه سورما^۸ در بنگلادش استفاده کردند. نتایج نشان داد که مدل انفیس می‌تواند پیش‌بینی کیفیت آب را با دقت بالایی انجام دهد. دینگ و همکاران (۱۹) از سیستم‌های فازی برای پیش‌بینی پارامترهای کیفی DO، COD، دمای آب و EC استفاده کرد. نتایج آن‌ها نشان داد که مدل به کار برده شده نتایج بهتری نسبت به مدل‌های موجود از خود نشان می‌دهند. علیایی و همکاران (۲۰) شبیه‌سازی تغییرات شاخص‌های BOD و DO را برای رودخانه دره مرادیگ همدان با استفاده از شبکه عصبی مصنوعی مورد بررسی قرار دادند. نتایج حاکی از کارایی بالای شبکه عصبی در تخمین شاخص‌های کیفی مذکور داشت.

پارما و بهاردواج (۲۱) آنالیز پارامترهای کیفیت آب را با استفاده از تابع موجک Db^۹ با درجه تجزیه ۵، در ر هاتنی کوند^{۱۰} واقع بر رودخانه یامونا در هند مورد بررسی قرار دادند. علیزاده و همکاران (۹) استفاده از مدل ترکیبی شبکه عصبی مصنوعی - موجک را برای پیش‌بینی پارامترهای کیفی در خلیج هیلو، اقیانوس آرام^{۱۱} انجام دادند. نتایج این بررسی نشان داد که مدل شبکه عصبی مصنوعی - موجک عملکرد بهتری نسبت به مدل شبکه عصبی مصنوعی دارد.

همان‌طوری که در قسمت بالا اشاره شد تحقیقات گسترده‌ای در دنیا در زمینه استفاده از مدل‌های شبکه عصبی مصنوعی و سیستم استنتاج فازی عصبی تطبیقی جهت پیش‌بینی BOD ارائه شده است ولی در رابطه با ماشین بردار پشتیبان هنوز مطالعات گسترده‌ای صورت پذیرفته است که این امر باعث شد در این مطالعه از توانایی مدل ماشین بردار پشتیبان به منظور پیش‌بینی پارامتر ماهانه BOD براساس داده‌های موجود دما، جریان رودخانه، DO و BOD استفاده شود. با توجه به رفتار دینامیکی سیستم‌های هیدرولوژیکی و به منظور افزایش دقت در پیش‌بینی، از آنالیز موجک به عنوان مدل ترکیبی استفاده شد. با توجه به مطالعاتی که تاکنون در زمینه کاربرد تبدیل موجک صورت پذیرفته است و در این مطالعه به برخی از آن‌ها اشاره گردید، مشخص شده است که استفاده از تبدیل موجک به دلیل پیش‌پردازی که روی داده‌های ورودی قبل از ورود به مدل‌ها انجام می‌گیرد باعث افزایش دقت مدل می‌شود به همین دلیل در این تحقیق از تبدیل موجک کمک گرفته شده است. همچنین به دلیل اینکه بعد از استفاده از تبدیل موجک هر پارامتر ورودی به تعدادی زیرسیگنال (که بسته به سطح تجزیه به زیرسیگنال‌های مختلفی تبدیل می‌شوند) تبدیل می‌شوند، از روش تجزیه به مولفه‌های اصلی^{۱۲} (PCA) جهت انتخاب زیرسیگنال‌های مهم استفاده شد. در آخر به منظور بررسی عملکرد مدل‌ها، از معیارهای مختلفی برای

همچنین یک سری معادلات طراحی برای پیش‌بینی حذف BOD و COD در جریان زیر سطحی در تالاب‌های مصنوعی ارائه دادند. نتایج مدل شبکه عصبی مصنوعی نسبت به مقادیر مشاهده‌ای دارای همبستگی خوبی بوده است. همچنین نتایج نشان داد که حذف COD به شدت به حذف BOD وابسته است و یک معادله برای پیش‌بینی حذف COD نیز ارائه شد. نجاه و همکاران (۱۲) توانایی بالای شبکه‌های عصبی را در تخمین شاخص‌های کیفی آب رودخانه جوهر مالزی و برآورد مقدار EC و کل جامدات محلول^۲ (TDS) و کدورت در این حوضه را مورد تأکید قرار دادند.

اسدالله‌فردی و همکاران (۱۳) توانایی دو مدل شبکه عصبی مصنوعی و شبکه عصبی بازگشتی^۳ را در تخمین شاخص کیفی آب رودخانه تلخه‌رود در آذربایجان شرقی برای برآورد پارامتر TDS مورد بررسی قرار دادند. ون و همکاران (۱۴) اکسیژن محلول رودخانه هیبه^۴ چین را با استفاده از شبکه عصبی مصنوعی پیش‌بینی کردند. نتایج آن‌ها حاکی از دقت و توانایی بالای مدل پیشنهادی برای پیش‌بینی و شبیه‌سازی پارامتر اکسیژن محلول بود.

پارما و بهاردواج (۱۵) تجزیه و تحلیل آماری و موجک پارامترهای کیفی آب، تقاضای شیمیایی اکسیژن (COD)، نیاز بیوشیمیایی به اکسیژن (BOD)، اکسیژن محلول (DO)، دمای آب (WT)، آمونیاک آزاد (AMM)، نیتروژن کج‌لال کل (TKN)^۵، کلیفرم مجموع (TC)، کلیفرم مدفوعی (FC) و pH جریان رودخانه یامونا^۶ در هند را برای داده‌های ماهانه یک دوره ۱۰ ساله مورد بررسی قرار دادند در این مطالعه از تابع موجک Db5 استفاده شد. در یک تحقیق به صورت مروری روش‌های مورد استفاده برای نیاز BOD ارائه شده است. در این مطالعه چگونگی ارزیابی و روش‌های اندازه‌گیری BOD ارائه شده است (۱۶). نجاه و همکاران (۱۷) از مدل سیستم استنتاج فازی عصبی تطبیقی برای پیش‌بینی غلظت DO استفاده کرد. در این مطالعه از پارامترهای دما، pH، نترات (NO3) و غلظت نیتروژن آمونیاکی^۷ (NH3-NL) برای افزایش دقت مدل استفاده شد. در آخر عملکرد مدل سیستم استنتاج فازی عصبی تطبیقی با مدل شبکه عصبی مصنوعی مقایسه شد. نتایج نشان داد مدل سیستم استنتاج فازی عصبی تطبیقی عملکرد بهتری به ویژه در وقایع بزرگ از خود نشان می‌دهد.

احمد و شاه (۱۸) از سیستم استنتاج فازی عصبی تطبیقی برای

8- Surma

9- Daubechies

10- HathniKund

11- Hilo Bay, Pacific Ocean

12- Principal Component Analysis

1- Johor

2- Total Dissolved Solids

3- Recurrent Neural Network

4- Heihe

5- Total Kjeldahl Nitrogen

6- Yamuna

7- Ammoniacal Nitrogen Concentration

ارزیابی استفاده شده است.

در جدول (۱) ارائه شده است.

به علت اینکه وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت می‌شود از روش استانداردسازی داده‌ها استفاده شده است که این کار مانع کوچک شدن بیش از حد وزن‌ها و سبب جلوگیری از اشباع زود هنگام نرون‌ها می‌گردد. با روش استانداردسازی هر عدد تبدیل به عددی بین ۰ تا ۱ می‌شود (۲۲). با توجه به پیشنهاد سلگی (۲۳) از رابطه زیر برای نرمال‌سازی (استاندارد سازی) استفاده شده است.

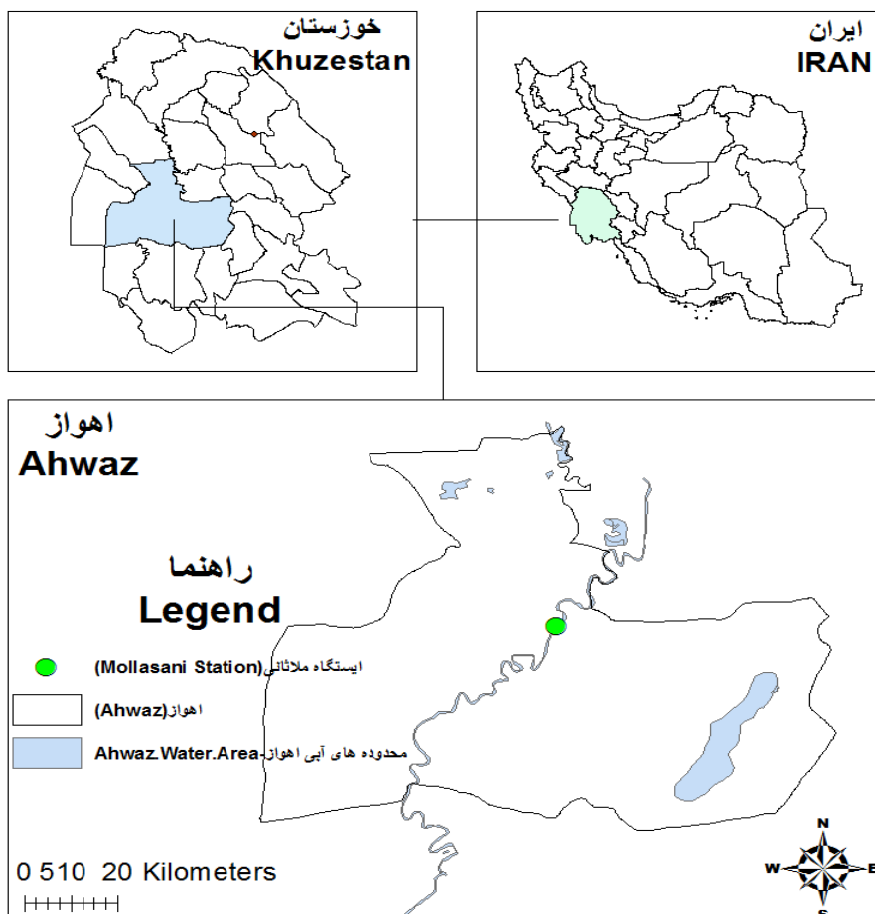
$$y = 0.5 + \left(0.5 \times \frac{x - \bar{x}}{x_{\max} - x_{\min}}\right) \quad (1)$$

که در این روابط X داده مورد نظر، \bar{X} میانگین داده‌ها، X_{\max} حداکثر داده‌ها، X_{\min} حداقل داده‌ها و y داده استاندارد شده می‌باشد. در این تحقیق از ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد برای شبیه‌سازی استفاده شده است.

مواد و روش‌ها

منطقه مورد مطالعه

رودخانه کارون از رشته‌کوه زاگرس سرچشمه گرفته و پس از عبور از مسیر کوهستانی در نزدیکی شهر گتوند به دشت خوزستان وارد می‌شود. رودخانه کارون در شمال شوشتر به دو شاخه تقسیم می‌شود که در جنوب شوشتر با یکدیگر پیوند می‌یابند. در این مسیر پر از پیچ و خم آب مورد نیاز شرب، کشاورزی و صنعت ده‌ها شهر و روستا، هزاران هکتار اراضی کشاورزی، تولید انرژی برق‌آبی، چندین طرح پرورش ماهی، طرح‌ها و کارخانه‌های صنعتی را تأمین می‌کند. در این مطالعه از اطلاعات ایستگاه ملاثانی با مختصات جغرافیایی ۴۸ درجه و ۵۲ دقیقه و ۴۰ ثانیه طول شرقی و ۳۱ درجه و ۳۵ دقیقه و ۰۱ ثانیه عرض شمالی، که در ارتفاع ۱۷ متری واقع شده است، استفاده شده است (شکل (۱)). پارامترهای استفاده شده در این مطالعه



شکل ۱- موقعیت رودخانه کارون در استان خوزستان و ایران
Figure 1- Location of Karun River in Khuzestan Province and Iran

جدول ۱- پارامترهای ماهانه مورد استفاده در این تحقیق

Table 1- Monthly Parameters used in study.

نام ایستگاه	پارامترهای استفاده شده	میانگین	حداقل	حداکثر
NameStation	Parameters used	Average	Mminimum	Maximum
ملاتانی Molasani	دما Temperatur(°C)	21.00	12.0	5.00
	اکسیژن محلول Dissolved Oxygen (mg/liter)	7.19	5.0	9.72
	جریان Flow (m ³ /s)	403.03	111.6	1626.63
	اکسیژن مورد نیاز بیولوژیکی Biological Oxygen Demand (mg/liter)	2.71	1.14	5.22

ماشین بردار پشتیبان^۱ (SVM)

در سال ۱۹۹۲ ماشین‌های بردار پشتیبان توسط وپنیک^۲ که یک ریاضیدان روس بود، ارائه گردید. این روش بر پایه نظریه یادگیری آماری^۳ بنا گردیده است (۲۴). در ساده‌ترین فرم آن یعنی SVM خطی، SVM عبارتست از یک ابر صفحه^۴ که مجموعه‌ی نمونه‌های مثبت و منفی را با حداکثر فاصله^۵ از هم جدا نموده است. جداکننده‌ای مناسب است که بیشترین فاصله را از هر دو دسته از داده‌ها داشته باشد، به عبارت دیگر سبب شود بیشترین حاشیه ریسک ایجاد شود. بنابراین در مورد این ماشین که خروجی آن برچسب گروه‌ها (در اینجا ۱- و +۱) می‌باشد، هدف بدست آوردن ماکزیمم فاصله بین دسته هاست. این حالت به داشتن ماکزیمم حاشیه تعبیر می‌شود. در ارتباط با SVM می‌توان مسئله طبقه‌بندی داده‌ها را در چند حالت مختلف بررسی نمود:

استفاده از ماشین بردار پشتیبان خطی^۶ و وجود دو دسته داده:

الف) داده‌ها دقیقاً در دو دسته قرار گرفته باشند، یعنی جدایی‌ناپذیر باشند^۷.

ب) داده‌های دو دسته قابل تفکیک به دو دسته جدا از هم نباشند^۸.

ماشین بردار پشتیبان غیرخطی^۹

ماشین بردار پشتیبان برای تفکیک‌های چند کلاسه^{۱۰}

حالت دوم یعنی ماشین بردار غیرخطی کاربرد بیشتری داشته به همین دلیل توضیحاتی در مورد آن ارائه می‌شود.

ماشین بردار پشتیبان غیرخطی، ابعاد مسأله را از طریق توابع کرنل^{۱۱} تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسأله را داشته باشد. معمولاً از چهار نوع کرنل زیر استفاده می‌شود.

$$k(x, y) = (x \cdot y + C)^d \quad (۲)$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

$$k(x, y) = \tanh(x \cdot y + \theta)$$

$$k(x, y) = (x \cdot y)$$

این کرنل‌ها به ترتیب کرنل چندجمله‌ای^{۱۲}، کرنل گاوسی یا

RBF، کرنل تانژانت هیپربولیک یا سیگموئید^{۱۳} و کرنل خطی^{۱۴}

می‌باشند (۲۵). کرنل RBF تنها یک پارامتر گاما دارد که با g مشخص می‌شود. برای این پارامتر یک بار مقدار پیش فرض $1/k$ را (تعداد

ویژگی‌هاست) استفاده می‌شود (به عبارتی هیچ مقداری را برای گاما به مدل وارد نمی‌شود) همچنین مقادیر $1/k^2$ و 1 را نیز تست می‌کنند. در

مورد کرنل سیگموئید با توجه به مسائلی که معمولاً از نظر همگرایی به جواب پیش می‌آید، تنها همان پیش فرض یعنی مقدار ثابت صفر و

گاما $1/k$ استفاده می‌شود یعنی این کرنل فاقد پارامتر است. کرنل

خطی هیچ پارامتری ندارد و بنابراین تنها یک حالت اجرا دارد. کرنل چندجمله‌ای دارای دو پارامتر است. پارامتر اول درجه چندجمله‌ای

10- Multiclass SVM

11- Kernel

12- Polynomial

13- Sigmoidal

14- Linear

1- Support Vector Machine

2- Vapnik

3- Statistical Learning

4- Hyperplane

5- MaximumMargin

6- Linear SVM

7- Separable Data

8- Nonseparable Data

9- Non-linear SVM

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (5)$$

در نهایت ضرایب موجک در هر نقطه از سیگنال (b) و برای هر مقدار از مقیاس (a) با رابطه ۶ قابل محاسبه می‌باشد (۲۹).

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi\left(\frac{t-b}{a}\right) f(t) dt \quad (6)$$

که در رابطه ۶ a کار مقیاس و b کار انتقال تابع را انجام می‌دهند. به ازای مقادیر مختلف a و b مقدار T بدست می‌آید هر موقع T بیشترین مقدار مثبت را داشته باشد بیشترین انطباق را دارد. به ازای T برابر صفر انطباقی وجود ندارد و به ازای T های منفی انطباق عکس یا بیشترین تفاوت وجود دارد. انتقال در تبدیل موجکی نشان دهنده موقعیت پنجره است. در این روش نیز یک تابع پنجره در نظر گرفته شده و در موج مورد نظر ضرب می‌شود. بر خلاف تبدیل قطعه‌ای، در تبدیل پیوسته موجکی پهنای تابع پنجره متغیر می‌باشد، به طوری که برای هر مؤلفه موج می‌توان عرض مناسبی را انتخاب نمود. این موضوع مهمترین خصوصیت تبدیل موجکی است (۸). مقیاس به طور ساده به معنای کشیده شدن یا فشرده شدن موجک می‌باشد. همانگونه که در نقشه‌ها مقیاس بزرگ متناظر با نشان ندادن جزئیات کلی است، در اینجا نیز مقیاس بزرگ متناظر با نشان ندادن جزئیات موج، و مقیاس کوچک متناظر با نشان دادن جزئیات موج است. به طور مشابه در جملات فرکانسی، فرکانس پائین (مقیاس بزرگ) متناظر با اطلاعات کلی موج می‌باشد که معمولاً اندازه موج داخلی است، در حالی که فرکانس‌های بزرگ متناظر با جزئیات اطلاعات پوشیده در موج است که معمولاً در یک زمان کوتاه به طول می‌انجامد. همچنین انتقال موجک به طور ساده به مفهوم به تأخیر انداختن یا جلو انداختن موجک و بیان کننده موقعیت موجک بر روی محور زمان است.

در انتخاب موجک مادر می‌بایست شرط‌های زیر وجود داشته باشد (۳۰).

- ۱- مساحت زیر منحنی آن صفر باشد. ۲- مقدار میانگین آن از هر دو سمت به صفر میل کند. ۳- همان‌های آن هم صفر باشد. این شرط اگر وجود نداشته باشد اشکالی ندارد.
- برای آشنایی بیشتر با این مبحث می‌توان به منبع (۳۰) مراجعه کرد.

است و با d مشخص می‌شود، به ازای مقادیر ۱، ۲، ۳ و ۴ مسائل حل می‌شوند. زمانی که C برابر صفر می‌باشد کرنل همگن نامیده می‌شود، برای پارامتر دوم نیز که عدد ثابت جمع شونده است و با r مشخص می‌شود، سه حالت ۰، ۱- و ۱+ استفاده می‌شود.

جهت کسب اطلاعات بیشتر در مورد روش SVM به منابع (۲۶) و (۲۷) مراجعه شود.

تبدیل موجک^۱

تئوری موجکی یکی از روش‌های علم ریاضی است که ایده اصلی آن برگرفته از تبدیل فوریه است که در قرن نوزدهم مطرح شده است ولی استفاده از آن قدمتی ۱۰ ساله دارد. مفهوم کلی موجک‌ها به صورت تئوری کنونی توسط مورلت و تیمی در مرکز تحقیقات فیزیک نظری مارسل زیر نظر آلکس گراسمن در فرانسه ارائه شد (۲۸). تبدیل موجک یکی از تبدیل‌های ریاضی کارآمد در زمینه پردازش سیگنال است. موجک‌ها توابع ریاضی هستند که شکل مقیاس-زمان از سری های زمانی و روابط آن‌ها جهت آنالیز سری‌های زمانی که شامل متغیرها و غیر ثابت‌ها می‌باشد را ارائه می‌دهد. تحلیل موجکی^۲ استفاده از فاصله‌های زمانی طولانی مدت را برای اطلاعات دارای فرکانس پائین و تناوب‌های کوتاه‌تر را برای اطلاعات دارای فرکانس بالا ارائه می‌دهد. تحلیل موجکی قادر به نمایش جنبه‌های مختلف داده‌های متفاوت، نقاط شکست و ناپیوستگی‌ها می‌باشد که ممکن است دیگر روش‌های آنالیز سیگنال آنها را نشان ندهند. تابع موجک، تابعی است که دو ویژگی مهم نوسانی بودن و کوتاه مدت بودن را دارا می‌باشد. $\psi(x)$ تابع موجک است اگر و فقط اگر تبدیل فوریه آن، شرط زیر (رابطه ۳) را ارضا کند (۲۹).

$$\int_{-\infty}^{+\infty} \frac{|\psi(x)|}{|\omega|^2} d\omega < +\infty \quad (3)$$

این شرط با عنوان شرط پذیرفتگی^۳ برای موجک $\psi(x)$ شناخته می‌شود. رابطه فوق را می‌توان معادل با رابطه ۴ دانست:

$$\psi(0) = \int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (4)$$

این ویژگی تابع با میانگین صفر، چنان محدودکننده نبوده و توابع بسیاری را می‌توان براساس آن تابع موجک نامید. $\psi(x)$ تابع موجک مادر است که توابع مورد استفاده در تحلیل، با دو عمل ریاضی انتقال^۴ و مقیاس^۵ در طول سیگنال مورد تحلیل، تغییر اندازه و تغییر محل می‌یابند.

- 1- Wavelet Transform
- 2- Wavelet Analysis
- 3- Admissibility
- 4- Translation
- 5- Scale

- 6- Admissibility conditions
- 7- Compact support
- 8- Regularity conditions

قرار دارد. این شاخص از رابطه (۷) به دست می‌آید که در این رابطه r_{ij} ضریب همبستگی بین متغیرهای i و j از ضریب همبستگی جزیی بین آن‌هاست (۳۱).

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2} \quad (7)$$

جهت آشنایی بیشتر با این مبحث به منبع (۳۱) مراجعه شود.

ساختارهای مختلف برای مدلسازی

برای اجرای مدل ماشین بردار پشتیبان با توجه به پارامترهای ورودی ۷ نوع ترکیب مختلف مطابق جدول (۲) مورد بررسی قرار گرفت.

جدول ۲- ترکیب‌های مختلف مدل ماشین بردار پشتیبان

Table 2- Different Combinations of SVM Model

ترکیب Combination	ورودی Input	خروجی Output
1	$BOD_{t-2}, BOD_{t-1}, BOD_t$	BOD_{t+1}
2	T_{t-2}, T_{t-1}, T_t	BOD_{t+1}
3	DO_{t-2}, DO_{t-1}, DO_t	BOD_{t+1}
4	Q_{t-2}, Q_{t-1}, Q_t	BOD_{t+1}
5	T_t, DO_t, Q_t, BOD_t	BOD_{t+1}
6	$T_{t-1}, T_t, DO_{t-1}, DO_t, Q_{t-1}, Q_t, BOD_{t-1}, BOD_t$	BOD_{t+1}
7	$T_{t-2}, T_{t-1}, T_t, DO_{t-2}, DO_{t-1}, DO_t, Q_{t-2}, Q_{t-1}, Q_t, BOD_{t-2}, BOD_{t-1}, BOD_t$	BOD_{t+1}

موجک‌ها شامل: موجک Haar، موجک Db2، موجک Db7 و موجک Sym4 با درجات مختلف تجزیه و به ازای پارامترهای مختلف مدل ماشین بردار پشتیبان، برنامه اجرا شد. جهت بدست آوردن درجه تجزیه مناسب برای توابع موجک از رابطه پیشنهادی زیر ارائه شد.

$$L = \text{Int}[\log(n)] \quad (8)$$

در این رابطه L درجه تجزیه پیشنهادی و N تعداد سری زمانی اعداد می‌باشد (۸). در این پژوهش با $N=156$ ، $L=2$ می‌باشد که برای دقت بیشتر، درجات تجزیه ۱ تا ۳ مورد بررسی قرار گرفت.

برای ارزیابی مدل‌ها از سه معیار، ضریب تعیین، جذرمیانگین مربعات خطا و معیار آکائیک استفاده شد. در زیر روابط این معیارها ارائه شده است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_i - T_o)^2}{N}} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (T_o - T)^2}{\sum_{i=1}^N (T_o - \bar{T})^2} \quad (10)$$

$$AIC = m \times \ln(RMSE) + 2(Npar) \quad (11)$$

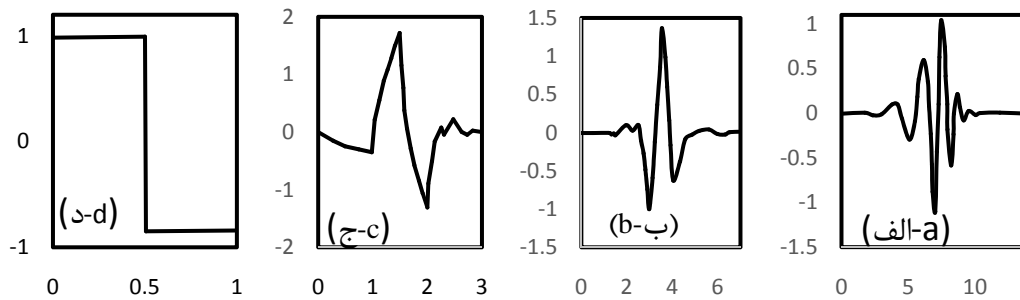
N : تعداد داده‌ها، T_o : داده‌های مشاهداتی، \bar{T} : میانگین داده‌های مشاهداتی، T_i : داده‌های محاسباتی، m : تعداد پارامترها، $Npar$: تعداد داده‌های آموزش دیده می‌باشد.

تجزیه به مولفه‌های اصلی (PCA)

تجزیه به مولفه‌های اصلی (PCA) نوعی از تجزیه و تحلیل آماری است که تعداد کمتری از عوامل را به نام مولفه‌های اصلی از میان عوامل اولیه گزینش می‌کند، به طوری که تعدادی از اطلاعات کم اهمیت حذف می‌شود. در صورتی که فاکتور تست KMO مربوط به این روش کمتر از ۰/۵ باشد، داده‌ها برای تجزیه و تحلیل عوامل اصلی مناسب نخواهد بود و اگر مقدار آن بین ۰/۵ تا ۰/۶۹ باشد باید با احتیاط بیشتر به تجزیه و تحلیل عوامل پرداخت. اما در صورتی که مقدار آن بزرگتر از ۰/۷ باشد همبستگی‌های موجود در بین داده‌ها برای تجزیه و تحلیل مناسب خواهد بود و می‌توان از این روش استفاده کرد (۳۱). شاخص KMO مورد استفاده در دامنه صفر تا یک

که پارامترهای BOD_t, Q_t, DO_t, T_t به ترتیب دما، اکسیژن محلول، جریان و اکسیژن مورد نیاز بیولوژیکی در دوره زمانی ماهانه و به ترتیب دما، DO، جریان و BOD در دوره‌های زمانی گذشته و BOD_{t+1} اکسیژن مورد نیاز بیولوژیکی در دوره آتی می‌باشد.

برای دستیابی به نتایج حاصل از مدل ترکیبی ماشین بردار پشتیبان-موجک برای مدل‌سازی ابتدا با استفاده از تبدیل موجک، پارامترهای ورودی به زیرسیگنال‌هایی تجزیه شده سپس زیر سیگنال‌های تجزیه شده با استفاده از روش تجزیه و تحلیل مولفه‌های اصلیمورد بررسی قرار گرفتند و مولفه‌های مهم به عنوان ورودی به مدل ماشین بردار پشتیبان وارد شد تا مدل ترکیبی ماشین بردار پشتیبان-موجک حاصل گردید. پس از اجرای متعدد این برنامه و در حالت‌های خاص و مقایسه آن‌ها با یکدیگر نتایج به دست آمد. یکی از نکات مهم و اساسی در مورد انتخاب موجک‌های مادر، طبیعت رخداد پدیده و نوع سری زمانی آن می‌باشد. لذا الگویی‌هایی از توابع موجک مادر که بتوانند به لحاظ شکل هندسی به خوبی بر منحنی سری زمانی منطبق شوند بهتر می‌توانند عمل نگاهت را انجام دهند و نتایج حاصل نیز بهتر خواهد بود. در این تحقیق با توجه به آزمایش موجک‌های مادر مختلف و توجه به نکته گفته شده در بالا ۴ تابع موجک زیر انتخاب شدند (شکل ۲). بدین منظور ابتدا برای هر یک از انواع



شکل ۲- الف) تابع موجک db7، ب) تابع موجک sym4، ج) تابع موجک db2، د) تابع موجک Haar
Figure 2-a) Wavelet db7, b) Wavelet sym4, c) Wavelet db2, d) Wavelet haar

داده‌های دما، جریان، DO و BOD استفاده شده است عملکرد مدل نسبت به زمانی که از تک تک پارامترها استفاده شده بود بهتر شد. به همین دلیل در ترکیب‌های ۶ و ۷ باز از همه پارامترها ولی با زمان تأخیرهای یک و دو ماهه استفاده شد تا بتوان مقایسه بهتری جهت دستیابی به ترکیب و ساختار برتر بدست آورد.

در ترکیب ۶ از همه پارامترها با یک زمان تأخیر استفاده شد. در این ترکیب، ساختاری که از تابع کرنل RBF استفاده شده بود با ضریب همبستگی ۰/۹۱۸ و جذر میانگین مربعات خطای شبیه‌سازی ۰/۳۳۸ بهترین عملکرد را داشت. در ترکیب ۷، ساختاری که از کرنل RBF استفاده شده بود با ضریب همبستگی ۰/۹۱۶ و جذر میانگین مربعات خطای شبیه‌سازی ۰/۳۵۶ بهترین عملکرد را داشت. در جدول (۳) بهترین نتیجه هر یک از ترکیب‌های هفت‌گانه ارائه شده است. براساس این جدول ترکیب ۶ دارای بهترین عملکرد برای مدل ماشین بردار پشتیبان بوده است و بعد از آن ترکیب ۷ نتیجه بهتری داشته یعنی اینکه افزایش تعداد تأخیرهای زمانی بیشتر از یک گام باعث کاهش عملکرد مدل شده است. شکل (۳) مقایسه مقادیر مشاهداتی را با مقادیر برآوردی برای ساختار برتر روش ماشین بردار پشتیبان در مراحل شبیه‌سازی نشان می‌دهند. با توجه به این شکل عملکرد مدل مناسب بوده است. همان‌طوری که از بررسی معیارهای ارزیابی برای بهترین ساختار هر یک از ترکیب‌های هفت‌گانه مشخص شد کرنل RBF نسبت به دیگر کرنل‌ها دارای عملکرد بهتری بوده است.

در مدل‌های ترکیبی سیگنال‌های اصلی پارامترهای ورودی با استفاده از تبدیل موجک به زیرسیگنال‌هایی تجزیه شده است. این زیرسیگنال‌ها توسط کدنویسی در نرم افزار متلب به عدد تبدیل شد، سپس با روش تجزیه به مولفه‌های اصلی مولفه‌های مهم تعیین شدند و به عنوان ورودی به مدل ماشین بردار پشتیبان مورد استفاده قرار گرفتند تا مدل ترکیبی ماشین بردار پشتیبان-موجک ایجاد شود. در این مدل‌سازی علاوه بر بررسی توابع موجک مختلف، پارامترهای مورد نیاز مدل ماشین بردار پشتیبان نیز بهینه شد تا بهترین عملکرد مدل ترکیبی حاصل شود.

ضریب R^2 میزان انطباق داده‌هایی که توسط مدل به وجود می‌آید و داده‌های واقعی را نشان می‌دهد و هر چقدر این ضریب به عدد یک نزدیکتر باشد بیانگر این است که این انطباق بهتر صورت گرفته و میزان خطای حاصله کمتر می‌باشد لذا در همه حالات برای بررسی میزان تأثیرگذاری هر پارامتر از جمله نوع تابع موجک، توابع انتقال، قوانین آموزش، توابع عضویت و درجه تجزیه موجک بر روی عملکرد مدل‌ها از این عامل استفاده شده است. همچنین پارامتر RMSE جذر میانگین مربع خطاهای داده‌های محاسباتی و مشاهداتی را بیان می‌کند. واضح است که هر چه مقدار این عدد کمتر باشد به تبع آموزش و شبیه‌سازی داده‌ها بهتر صورت گرفته است. در مورد AIC معیار آکائیک نیز می‌توان گفت که هر چه ضریب آکائیک مربوط به مدل کمتر باشد آن مدل عملکرد بهتری دارد. ضریب آکائیک موقعی که کمتر است ناشی از دو عامل می‌باشد یکی خطای مدل و دیگری تعداد پارامترهاست پس معیار خوبی برای بررسی مدل‌ها می‌باشد.

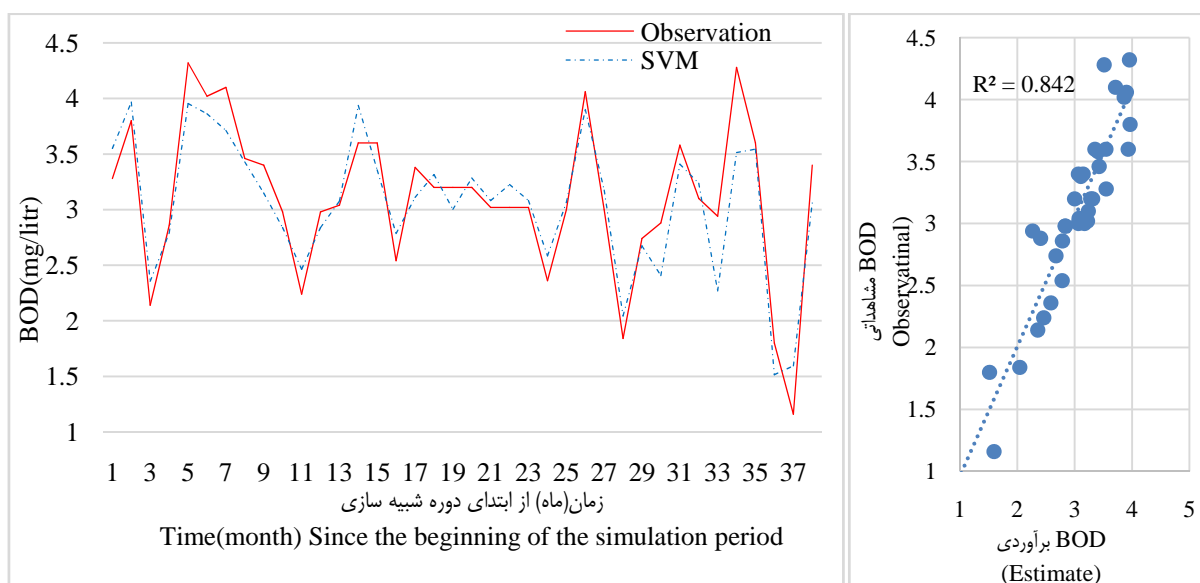
نتایج و بحث

ساختارهای مختلف در هر یک از ترکیب‌های مختلف مدل ماشین بردار پشتیبان به این صورت که هر یک از ساختارها به ازای کرنل، و دیگر پارامترهای مختلف مدل مورد بررسی قرار گرفت. نتایج حاصل از بررسی ترکیب‌های مختلف، به خصوص ترکیب شماره ۱ که فقط از داده‌های BOD استفاده شده است، نشان داد بهترین عملکرد مربوط به کرنل RBF با ضریب همبستگی ۰/۷۷۳ و جذر میانگین مربعات خطای ۰/۵۷۹ می‌باشد. در ترکیب‌های ۲ و ۳ که به ترتیب از داده‌های دما و DO استفاده شده است عملکرد مدل با توجه به ضریب همبستگی و جذرمیانگین مربعات خطای بدست آمده ضعیف بوده است. در ترکیب ۴ هم که از داده‌های جریان استفاده شد عملکرد مدل باز ضعیف بوده و این نشان‌دهنده این مطلب می‌باشد که فقط استفاده از پارامترهای BOD، دما، DO و یا جریان به هیچ وجه مناسب نبوده و در نتیجه باید از ترکیب‌هایی که به صورت ترکیبی از پارامترهای ذکر شده می‌باشد استفاده نمود. در ترکیب ۵ که از

جدول ۳- بهترین ساختار مدل ماشین بردار پشتیبان

Table 3-The Best Structure in SVM Model

ترکیب Combination	پارامترهای مدل Model Parameters					ضریب همبستگی R		جذر میانگین مربعات خطا RMSE	
	Kernel	Gama	Sin2	B	P-Degree	آموزش	تست	آموزش	تست
						Train	Test	Train	Test
1	RBF	10.01	456.6	0.176	-----	0.932	0.773	0.0387	0.0579
2	Lin	1.402	-----	4.5218	-----	0.807	0.608	0.0646	0.0832
3	Lin	5.432	-----	10.028	-----	0.890	0.718	0.0506	0.0628
4	Poly	0.031	0.532	0.498	3	0.866	0.746	0.0523	0.0575
5	Poly	0.701	789.8	12.06	3	0.931	0.903	0.0383	0.0373
6	RBF	1.241	1289	3.125	-----	0.956	0.918	0.0307	0.0338
7	RBF	17.04	10.37	0.345	-----	0.954	0.916	0.0314	0.0356



شکل ۳- مقایسه نتایج مدل ماشین بردار پشتیبان با مقادیر مشاهده‌ای، مرحله شبیه‌سازی.

Figure 3- Compare the results of the SVM model with actual values, the SimulationStep

می‌توان از آن برای پیش‌بینی مقدار BOD استفاده نمود. نکته مهمی که بایستی بدان توجه داشت این است که در روش ترکیبی، استفاده از درجات تجزیه بالای سری زمانی BOD تأثیر چندانی در نتایج مطلوب نشان نداد و فقط زمان اجرای برنامه را طولانی‌تر ساخت و این به خاطر این بود که هر چه درجه تجزیه بیشتر شود سیگنال به همان مقدار به بسامدهای بالاگذر و پایین‌گذر تجزیه می‌شود و دقت کار افزایش می‌یابد، اما از مرحله‌ای به بعد مقادیر بسامدها تقریباً ثابت باقی می‌مانند. به همین دلیل باید درجه تجزیه مناسب انتخاب گردد و همیشه افزایش درجات مناسب نخواهد بود.

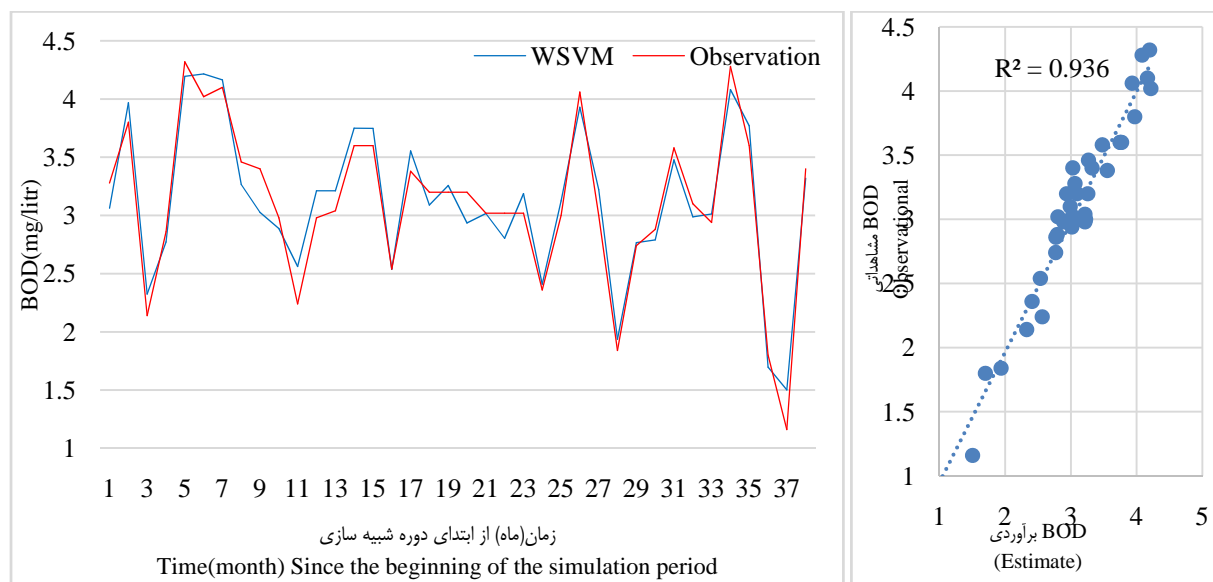
نتایج حاصل از مدل‌سازی با مدل ترکیبی ماشین بردار پشتیبان-موجک به ازای توابع موجک مختلف برای درجه تجزیه‌های مختلف در جدول (۴) ارائه شده است. با توجه به جدول (۴) مشاهده می‌شود که درجه تجزیه ۲ دارای بهترین عملکرد بوده است در این ساختار برتر از کرنل RBF با ضریب همبستگی ۰/۹۷ و جذرمیانگین مربعات خطای شبیه‌سازی ۰/۰۲۱۰ استفاده شده است. نتایج حاصل از مدل-سازی این ساختار برتر برای داده‌های شبیه‌سازی در شکل‌های (۴) ارائه شده است.

با بررسی شکل (۴) این نتیجه حاصل می‌شود که عملکرد مدل در مرحله شبیه‌سازی مناسب بوده و از این مدل می‌توان جهت پیش‌بینی مقادیر آبی استفاده نمود. با توجه به جدول (۴) در تمامی سطوح تجزیه، برای مدل ماشین بردار پشتیبان-موجک، تابع موجک Db2 عملکرد بهتری داشته و به عنوان موجک برتر شناخته می‌شود و

جدول ۴- بهترین ساختار مدل ماشین بردار پشتیبان-موجک

Table 4- The Best Structure in WSVMModel

نوع موجک مادر Mother Wavelet Type	درجه تجزیه Decomposition Level	پارامترهای مدل Model Parameters					ضریب همبستگی R		جزر میانگین مربعات خطا RMSE	
		Kernel	Gama	Sin2	B	P- Degree	آموزش Train	شبیه‌سازی Simulate	آموزش Train	شبیه‌سازی Simulate
Db2	1	RBF	3.399	6.129	3.875	----	0.97	0.96	0.0264	0.0245
	2	RBF	1.564	1.098	-0.801	----	0.98	0.97	0.0188	0.0210
	3	Lin	0.867	----	0.1216	----	0.98	0.97	0.0208	0.0240
Db7	1	Lin	7120	----	0.916	----	0.97	0.95	0.0255	0.0264
	2	RBF	0.543	1.327	-0.924	----	0.97	0.96	0.0236	0.0243
	3	Poly	1295	0.121	10.81	3	0.97	0.95	0.0274	0.0260
Sym4	1	Lin	3.187	----	3.101	----	0.97	0.92	0.0242	0.0322
	2	RBF	3.234	7.890	0.129	----	0.97	0.94	0.0245	0.0329
	3	Poly	1232	1.645	-0.830	3	0.97	0.93	0.0275	0.0303
Haar	1	Poly	31.45	1.021	1.651	4	0.98	0.95	0.0243	0.0273
	2	RBF	0.483	2.743	6.855	----	0.97	0.96	0.0254	0.0257
	3	Lin	1.857	----	3.406	----	0.98	0.95	0.0204	0.0255



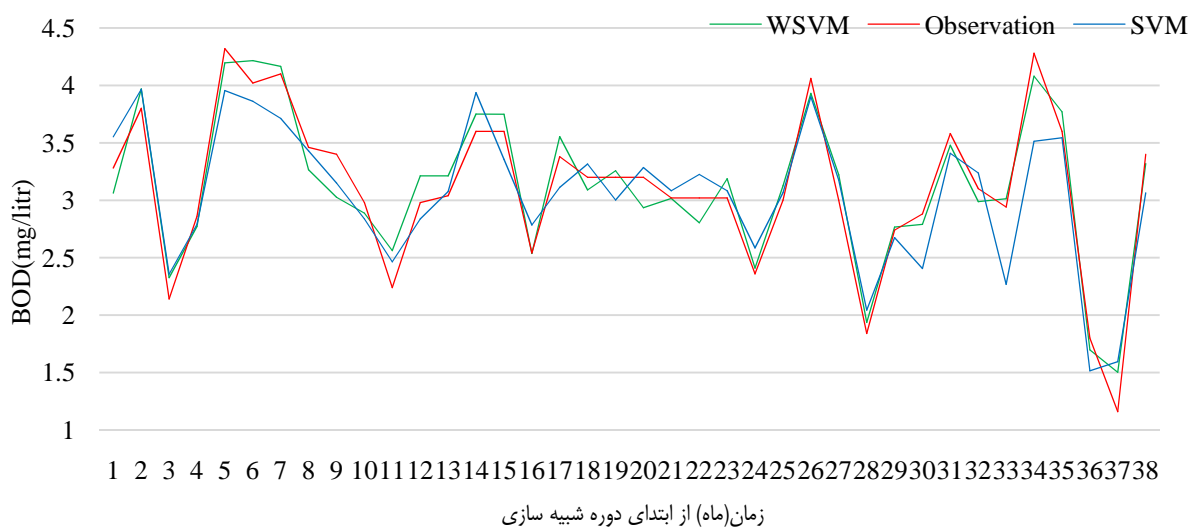
شکل ۴- مقایسه نتایج مدل ماشین بردار پشتیبان-موجک با مقادیر مشاهده‌ای، مرحله شبیه‌سازی.
Figure 4- Compare the results of the WSVM model with actual values, the Simulation Step

بهتری است. شکل (۵) نتایج حاصل از مدل‌های استفاده شده را در مرحله شبیه‌سازی نشان می‌دهد. با توجه به این شکل معلوم است که عملکرد مدل ترکیبی ماشین بردار پشتیبان-موجک بهتر بوده چون مقادیر برآوردی با این مدل به مقادیر مشاهده‌ای نزدیک‌تر می‌باشد. بنابراین مدل ترکیبی نسبت به مدل ماشین بردار پشتیبان بهتر بوده و این به دلیل این است که قبل از ورود داده‌ها به مدل روی داده‌ها پردازش صورت گرفته و داده‌ها دسته‌بندی شده و با فرکانس‌های مختلف وارد مدل ماشین بردار پشتیبان شده است. این موضوع را می‌توان از روی مقایسه مقادیری که برای ضریب همبستگی چه در مرحله آموزش و چه در مرحله شبیه‌سازی داده‌ها بدست می‌آید، یافت.

با توجه به جدول و نتایج حاصل مشخص شد که تجزیه مرتبه بالا اگر چه دقت محاسبات در آموزش مدل را بالا می‌برد ولی نتیجه معکوس روی نتایج داده‌های شبیه‌سازی شده دارد. بنابراین نیاز به استفاده از درجات بالای تجزیه نمی‌باشد زیرا فرکانس‌های BOD در مقیاس زمانی معین ماهانه می‌باشد که با انتخاب حداکثر درجه تجزیه این مهم بدست می‌آید.

با بررسی ساختارهای برتر دو مدل ماشین بردار پشتیبان و ماشین بردار پشتیبان-موجک مشخص می‌شود که مدل ترکیبی ماشین بردار پشتیبان-موجک دارای عملکرد بهتری بوده و از نظر بررسی تمامی معیارهای ارزیابی نسبت به ماشین بردار پشتیبان دارای عملکرد

مقایسه نتایج حاصل از مدل‌های مختلف در جدول (۵) ارائه شده است.



شکل ۵- مقایسه مدل‌های استفاده شده در این تحقیق

Figure 5-Comparison of models used in this study

بهتری داشته است. براساس معیار آکائیک که در آن هر مدلی که معیار آکائیک کمتری داشته باشد مدل برتر شناخته می‌شود لذا براساس این معیار مشاهده می‌شود که باز مدل ترکیبی ماشین بردار پشتیبان-موجک دارای معیار آکائیک کمتر بوده و به عنوان مدل برتر شناخته می‌شود. با بررسی ساختارهای مختلف مشخص شد که از بین کرنل‌های مختلف، کرنل نوع RBF دارای بهترین عملکرد بوده است.

با توجه به جدول فوق و مقایسه مدل‌های ماشین بردار پشتیبان و ماشین بردار پشتیبان-موجک مشاهده می‌شود که تبدیل موجک حدود ۵ درصد افزایش ضریب همبستگی را موجب شده است. همچنین با توجه به پارامتر جذر میانگین مربعات خطا نیز مشاهده می‌شود که مدل ترکیبی ماشین بردار پشتیبان-موجک دارای جذر میانگین مربعات خطای شبیه‌سازی کمتر بوده و در نتیجه عملکرد

جدول ۵- مقایسه مدل‌های استفاده شده در این تحقیق

Table 5- Comparison of models used in this study

مدل Model	ضریب همبستگی R		جذر میانگین مربعات خطا RMSE		معیار اطلاعاتی آکائیک AIC	
	آموزش Train	تست Test	آموزش Train	تست Test	آموزش Train	تست Test
SVM	0.956	0.918	0.0307	0.0338	202.13	48.90
WSVM	0.984	0.968	0.0188	0.0210	182.32	29.65

بوده و این به دلیل بارندگی و تغییرات دبی رودخانه می‌باشد، چون با تغییرات بارندگی و تغییرات دبی مقدار BOD نیز دچار تغییرات می‌شود.

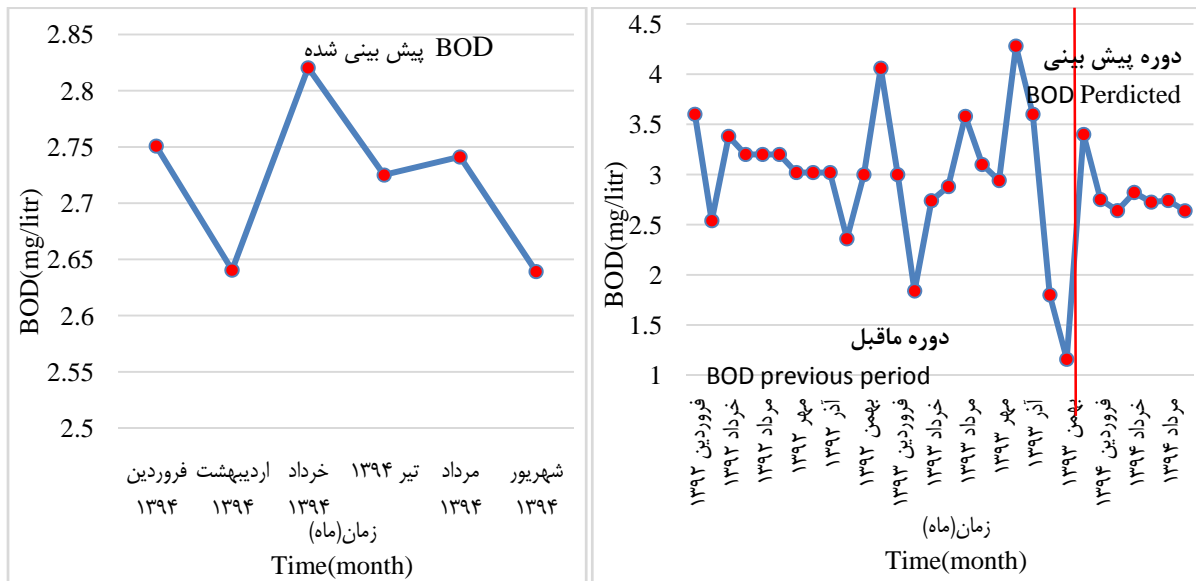
با توجه به اینکه مطالعه‌ای مانند این مطالعه مبنی بر استفاده از ترکیب SVM و موجک برای پیش‌بینی BOD صورت نگرفته است امکان مقایسه نتایج وجود نداشته است ولی در موارد مشابه که از مدل‌های دیگر استفاده شده است می‌توان به این مطلب اشاره کرد که

پیش‌بینی BOD برای یک دوره شش ماهه با مدل برتر

بعد از بررسی مدل‌های مختلف مشخص شد که مدل ترکیبی ماشین بردار پشتیبان-موجک مدل برتر بوده به همین دلیل با تعیین بهینه پارامترهای این مدل اقدام به پیش‌بینی BOD شد. نتایج پیش‌بینی BOD از فروردین ۱۳۹۴ تا شهریور ۱۳۹۴ در شکل (۶) ارائه شده است. برای بررسی روند تغییر BOD در شکل (۶) مقادیر BOD از فروردین ۱۳۹۲ به بعد آورده شده است. بررسی شکل (۶) نشان می‌دهد که فصول پاییز و زمستان نوسانات مقدار BOD بیشتر

مدل به خاطر استفاده از تبدیل موجک همخوانی دارد.

نتایج این مطالعه با نتایج عزیزاده و همکاران (۹) مبنی بر افزایش دقت



شکل ۶- پیش‌بینی و بررسی روند BOD (از فروردین ۱۳۹۲ تا شهریور ۱۳۹۶)
Figure 6-The Forecast and Examine BOD Trend (of April 2013 to September 2015)

زمانی رفتار آن پدیده را پیش‌بینی نمود.

یکی از نکات مهم و اساسی در مورد انتخاب موجک مادر، طبیعت رخداد پدیده و نوع سری زمانی آن می‌باشد. لذا الگوهایی از توابع موجک مادر که بتوانند به لحاظ شکل هندسی به خوبی بر منحنی سری زمانی منطبق شوند بهتر می‌توانند عمل نگاهت را انجام دهند و به تبع نتایج بهتری را در بر خواهند داشت. در این تحقیق با توجه به آزمایش موجک‌های مختلف و توجه به نکته بالا، نتایج حاکی از آن است که موجک‌های Db2 و Db7 دارای جواب‌های مناسبی می‌باشند. ضریب همبستگی برای داده‌های آموزشی و شبیه‌سازی در مدل ترکیبی نسبت به مدل ماشین بردار پشتیبان بیشتر بوده که این موضوع بیانگر دقت روش ترکیبی می‌باشد. مقایسه نتایج حاصل از مدل‌سازی ماهانه حاکی از این است که استفاده از تبدیل موجک در حالت ماهانه، باعث افزایش ۵ درصدی عملکرد شده است.

بررسی ساختارهای مختلف و آنالیز حساسیت نشان داد که از بین پارامترهای استفاده شده در این تحقیق پارامتر BOD دارای بیشترین اهمیت و سپس پارامترهای جریان، DO و دما در مرحله بعدی اهمیت قرار دارند. یعنی اینکه BOD بیشترین تأثیر و دما کمترین تأثیر را دارند. همچنین از بین کرنل‌های مختلف کرنل نوع RBF دارای بهترین عملکرد بوده است.

نتیجه‌گیری کلی

در این تحقیق از مدل ماشین بردار پشتیبان برای پیش‌بینی BOD در رودخانه کارون در دوره زمانی ماهانه استفاده شد. بعد از استفاده از تبدیل موجک و تجزیه نمودن سیگنال اصلی به زیرسیگنال‌هایی، از روش تجزیه به مولفه‌های اصلی (PCA) استفاده شد و مولفه‌های مهم به عنوان ورودی به مدل ماشین بردار پشتیبان استفاده شد تا مدل ترکیبی ماشین بردار پشتیبان-موجک حاصل گردد. با بررسی و استفاده از تمامی پارامترهای مؤثر از توابع موجک گرفته تا کرنل‌ها و دیگر پارامترهای مدل ماشین بردار پشتیبان سعی شد مدل‌سازی در بهترین وضعیت صورت گرفته تا بتوان مقایسه‌ها در بهترین حالت ممکن انجام شوند. ترکیب مدل ماشین بردار پشتیبان با آنالیز موجک یک ایده نو برای به وجود آمدن شبکه‌هایی توانمندتر می‌باشد پردازش داده‌های سری زمانی با آنالیز موجک باعث می‌شود که نقش و میزان تأثیر هر پارامتر در خروجی مدل ماشین بردار پشتیبان به طور واضح بررسی گردد. همچنین تبدیل داده‌ها از حوزه زمان به حوزه فرکانسی با استفاده از تبدیل موجک این امکان را فراهم می‌آورد که در بررسی روند رخداد یک پدیده به این مهم که در هر بازه زمانی چه فرکانس‌هایی رخ می‌دهد دست یافت، و با استفاده از این مهم دوره بازگشت رخداد یک پدیده را به دست آورد و الگوی

1. Streeter H.W., Phelps E.B. 1925. A study of the pollution and natural purification of the Ohio River, III. Factors concerned in the phenomenon of oxidation and reaeration. Bulletin 146 Public Health Service, Washington, DC, USA.
2. Liu L., Deng L., Yong D., Dong S. 2011. Native biofilm cultured under controllable condition and used in mediated method for BOD measurement. *Talanta*, 84:895-9.
3. Kunwar P., Singh A., Amrita M., Gunja J. 2009. Artificial neural network modeling of the river water quality-A case study. *Ecol Model*, 220(8):88-95.
4. Xiang S.L., Liu Z.M., Ma L.P. 2006. Study of multivariate linear regression analysis model for ground water quality prediction. *Guizhou Science*, 24:60-2.
5. Wu H.J., Lin Z.Y., Guo S.L. 2000. The application of artificial neural networks in the resources and environment. *Resour Environ Yangtze Basin*, 9:237-41.
6. Kisi O. 2008. Stream flow forecasting using neuro-wavelet technique. *Hydrological Processes*, 22: 4142-52.
7. Nourani V., Kisi Ö., Komasi M. 2011. Two hybrid Artificial Intelligence approaches for modeling rainfall-runoff process. *Journal of Hydrology*, 402:41-59.
8. Nourani V., Komasi M., Mano A. 2009. A Multivariate ANN-Wavelet Approach for Rainfall-Runoff Modeling. *Water Resource Management*, 23:2877-94.
9. Alizadeh M.J., Kavianpour M.R. 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine Pollution Bulletin*, 98(1-2):171-8.
10. Solgi A., Nourani V., Pourhaghi A. 2014. Forecasting Daily Precipitation Using Hybrid Model of Wavelet-Artificial Neural Network and Comparison with Adaptive Neuro Fuzzy Inference System (Case Study: Verayneh Station, Nahavand). *Advances in Civil Engineering*, 1-12.
11. Christos S.A., John N.E.P., Vassilios A.T. 2008. An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands. *Chemical Engineering Journal*, 143(1-3):96-110.
12. Najah A., Elshafie A., Karim O., Jaffar O. 2009. Prediction of Johor river water quality parameters using artificial neural networks. *European Journal of Scientific Research*, 28(4):22-35.
13. Asadollahfardi G., Taklif A., and Ghanbari A. 2012. Application of Artificial Neural Network to Predict TDS in Talkheh Rud River. *Journal of Irrigation and Drainage Engineering (ASCE)*, 138(20):363-370.
14. Wen X., Fang J., Diao M., Zhang C. 2013. Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China. *Environmental monitoring and assessment*, 185(5):4361-71.
15. Parmar K.S., Bhardwaj R. 2013. Wavelet and statistical analysis of river water quality parameters. *Applied Mathematics and Computation*, 219:10172-82.
16. Jouanneau S., Recoules L., Durand M.J., Boukabache A., Picot V., Primault Y. 2014. Methods for assessing biochemical oxygen demand (BOD): A review. *Water Research*, 49:62-82.
17. Najah A., Shafie A.E., Kari O.A., Amr H., Shafie E. 2014. Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environmental Science and Pollution Research*, 21(3):1658-70.
18. Ahmed A.A.M., Shah S.M.A. 2015. Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. *Journal of King Saud University - Engineering Sciences*.
19. Deng W., Wang G., Zhang X. 2015. A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting. *Chemometrics and Intelligent Laboratory Systems*, 149(Part A):39-49.
20. Olyaie E., Banejad H., Samadi M.T., Rahmani A.R., and Saghi M.H. 2010. Performance Evaluation of Artificial Neural Networks for Predicting Rivers Water Quality Indices (BOD and DO) in Hamadan Morad Beik River. *water and soil science*, 20.1 (3): 200-210 (in Persian with English abstract).
21. Parmar K.S., Bhardwaj R. 2012. Analysis of Water Parameters Using Daubechies Wavelet (Level 5) (Db5). *American Journal of Mathematics and Statistics*, 2(3): 57-63.
22. Riad S., Mania J., Bouchaou L., Najjar Y. 2004. Rainfall-runoff model using an artificial neural network approach. *Mathematical and Computer Modelling*, 40(7-8): 839-46.
23. Solgi A. 2014. Stream flow forecasting using combined Neural Network Wavelet model and comparison with Adaptive Neuro Fuzzy Inference System and Artificial Neural Network methods (Case Study: Gamasyab River, Nahavand). MSc. Shahid Chamran University of Ahvaz, Iran (in Persian with English abstract).
24. Cortes C., Vapnik V. 1995. Support-Vector Networks. *Machine Learning*, 20:273-95.
25. Sujay Raghavendra N., Paresh Chandra D. 2014. Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19(0):372-86.
26. Schölkopf, B., A. Smola and K.R. Müller. 2005. Kernel principal component analysis. *Lecture Notes in Computer Science*. 1327: 583-588.
27. Schölkopf, B. and A.J. Smola. 2002. *Learning with Kernels*. MIT Press, Cambridge, MA.

28. Grossmann A., Morlet J. 1984. Decomposition of hardy function into square integrable wavelets of constant shape. *SIMA Journal on Mathematical Analysis (SIMA)*, 5:723-36.
29. Mallat S.G. 1998. A wavelet tour of signal processing. 2, editor: San Diego. 557 p.
30. Fofola-Georgiou E., Kumar P. 1995. Wavelet in geophysics. New York: Academic Press.
31. Hutcheson G., and Nick S. 1999. The multivariate social scientist: Introductory statistics using generalized linear models. Thousand Oaks, CA, Sage Publications.

Modeling and Forecast Biological Oxygen Demand (BOD) using Combination Support Vector Machine with Wavelet Transform

A. Solgi^{*1} - A. Pourhaghi² - H. Zarei³ - H. Ansari⁴

Received:30-12-2015

Accepted:25-04-2016

Introduction: Chemical pollution of surface water is one of the serious issues that threaten the quality of water. This would be more important when the surface waters used for human drinking supply. One of the key parameters used to measure water pollution is BOD. Because many variables affect the water quality parameters and a complex nonlinear relationship between them is established conventional methods can not solve the problem of quality management of water resources. For years, the Artificial Intelligence methods were used for prediction of nonlinear time series and a good performance of them has been reported. Recently, the wavelet transform that is a signal processing method, has shown good performance in hydrological modeling and is widely used. Extensive research has been globally provided in use of Artificial Neural Network and Adaptive Neural Fuzzy Inference System models to forecast the BOD. But support vector machine has not yet been extensively studied. For this purpose, in this study the ability of support vector machine to predict the monthly BOD parameter based on the available data, temperature, river flow, DO and BOD was evaluated.

Materials and Methods: SVM was introduced in 1992 by Vapnik that was a Russian mathematician. This method has been built based on the statistical learning theory. In recent years the use of SVM, is highly taken into consideration. SVM was used in applications such as handwriting recognition, face recognition and has good results. Linear SVM is simplest type of SVM, consists of a hyperplane that dataset of positive and negative is separated with maximum distance. The suitable separator has maximum distance from every one of two dataset. So about this machine that its output groups label (here -1 to +1), the aim is to obtain the maximum distance between categories. This is interpreted to have a maximum margin. Wavelet transform is one of methods in the mathematical science that its main idea was given from Fourier transform that was introduced in the nineteenth-century. Overall, concept of wavelet transform for current theory was presented by Morlet and a team under the supervision of Alex Grossman at the Research Center for Theoretical Physics Marcel in France. After the parameters decomposition using wavelet analysis and using principal component analysis (PCA), the main components were determined. These components are then used as input to the support vector machine model to obtain a hybrid model of Wavelet-SVM (WSVM). For this study, a series of monthly of BOD in Karun River in Molasani station and auxiliary variables dissolved oxygen (DO), temperature and monthly river flow in a 13 years period (2002-2014) were used.

Results and Discussion: To run the SVM model, seven different combinations were evaluated. Combination 6 which was contained of 4 parameters including BOD, dissolved oxygen (DO), temperature and monthly river flow with a time lag have best performance. The best structure had RMSE equal to 0.0338 and the coefficient of determination equal to 0.84. For achieving the results of the WSVM, the wavelet transform and input parameters were decomposed to sub-signal, then this sub-signals were studied with Principal component analysis (PCA) method and important components were entered as inputs to SVM model to obtain the hybrid model WSVM. After numerous run this program in certain modes and compare them with each other, the results was obtained. One of the key points about the choice of the mother wavelet is the time series. So, the patterns of the mother wavelet functions that can better adapt to diagram curved of time series can do the mappings operation and therefore will have better results. In this study, according to different wavelet tests and according to the above note, four types of mother wavelet functions Haar, Db2, Db7 and Sym3 were selected.

Conclusions: Compare the results of the monthly modeling indicate that the use of wavelet transforms can increase the performance about 5%. Different structures and sensitivity analysis showed that the most important parameter which used in this study was parameter BOD, and then flow, DO and temperature were important. This means that the most effective BOD and temperature with minimum impact. Also between different kernels

1, 2 and 3- Ph.D. Students and Assistant Professor, Department of Water Resources Engineering, Faculty of Water Sciences, ShahidChamran University of Ahvaz, Iran, Respectively

(*- Corresponding Author Email: A-solgi@phdstu.scu.ac.ir)

4- Ph.D. Student, Department of Water Resources Engineering, Faculty of Agriculture, Bu-Ali SinaUniversity, Hamedan, Iran

types, RBF kernel showed the best performance. So, combined wavelet with support vector machine is a new idea to predict BOD value in the Karun River.

Keyword: Forecast BOD, Hybrid Model, PCA, Karun River