

مقاله علمی- پژوهشی

ارزیابی کارایی درخت تصمیم در ترکیب با تبدیل موجک به منظور پیش‌بینی نوسانات سطح آب‌زیرزمینی (مطالعه موردی: دشت کرمان - باغین)

هانیه رستمی‌نژاد^۱، سجاد شهابی^{۲*}، محمدرضا مددی^۳

تاریخ دریافت: ۱۴۰۱/۱۰/۰۳ تاریخ پذیرش: ۱۴۰۱/۱۱/۲۱

چکیده

پایش سطوح آب‌های زیرزمینی و برآورد دقیق نوسانات آن در دوره‌های پیش‌رو به‌خصوص در مناطق خشک و نیمه‌خشک ضرورت دارد. با توجه به قابلیت مدل‌های مبتنی بر هوش مصنوعی در مدل‌سازی پدیده‌های هیدرولوژیکی، در این پژوهش از مدل درخت تصمیم M5P در ترکیب با تبدیل موجک به منظور پیش‌بینی نوسانات سطح آب‌زیرزمینی دشت کرمان - باغین استفاده شده است. جهت توسعه مدل ترکیبی موجک-درخت تصمیم (W-M5P)، خروجی‌های تبدیل موجک به‌عنوان ورودی بر M5P اعمال می‌شوند. برای ارزیابی کارایی مدل ترکیبی در مقایسه با مدل منفرد، از چندین معیار از جمله ضریب همبستگی (R)، شاخص توافق (I_a) و شاخص پراکندگی (SI) استفاده شد. نتایج نشان داد، به‌رغم اینکه ورودی مدل ترکیبی، صرفاً داده‌های هواشناسی ایستگاه سینوپتیک بوده و از سطح آب در دوره‌های پیشین استفاده نشده، با این وجود WM5P کارایی بالایی در مدل‌سازی نوسانات سطح آب‌زیرزمینی در مقایسه با مدل منفرد ارائه نموده است. به‌گونه‌ای که مدل WM5P برای افق پیش‌بینی سه ماه با موجک Coif4 و سطح تجزیه شش، مقدار SI را از ۰/۶۳۹۴ به ۰/۱۸۱ کاهش و هم‌زمان I_a را از ۰/۶۸۹۸ به ۰/۹۹۹۸ افزایش داده است. براین اساس، انتخاب موجک کوپفلت با مرتبه ۴ و سطوح تجزیه ۵ و ۶ در مدل ترکیبی، کاراترین مدل در برآورد سطح آب‌زیرزمینی دشت کرمان - باغین می‌باشد.

واژه‌های کلیدی: پیش‌پردازش داده، سطح ایستایی، مدل‌سازی، هوش مصنوعی، یادگیری ماشین

مقدمه

آب‌زیرزمینی مورد استفاده قرار گرفته‌اند. این روش‌ها اغلب دارای محدودیت‌هایی همچون نیاز به حجم بالای داده یا پارامترهای ورودی می‌باشند که کمبود این داده‌ها منجر به خطای زیاد در پیش‌بینی‌ها (دقت پایین) می‌شود.

در سال‌های اخیر روش‌های مبتنی بر هوش مصنوعی^۲ (AI) به دلیل سادگی و نتایج قابل‌قبولی که ارائه می‌کنند، به‌طور گسترده‌ای در مدل‌سازی آب‌زیرزمینی مورد استفاده قرار گرفته‌اند (Rajaei et al., 2019). این روش‌ها شامل انواع مدل‌های یادگیری ماشین^۳ (ML)، شبکه عصبی مصنوعی^۴ (ANN)، سیستم استنتاج فازی عصبی- تطبیقی^۵ (ANFIS)، ماشین بردار پشتیبان^۶ (SVM)، رگرسیون بردار پشتیبان^۷ (SVR) و شبکه بیزین^۸ (BN) می‌باشند. به‌علت عملکرد مناسب این مدل‌ها، از آن‌ها در مسائل مختلف مهندسی و مدیریت منابع آب، پیش‌بینی‌های هیدرولوژیکی، تحلیل ریسک و تجزیه و تحلیل عدم قطعیت، استفاده شده است (Zyoud et al., 2016; Milan et al., 2018; Shariat et al., 2019; Sharafati et al., 2020). به‌دلیل محدودیت‌های عملی، دسترسی به داده‌ها یا حل معادلات پیچیده، مدل‌های هوش مصنوعی (AI) همراه با سایر

بهره‌برداری بیش از حد از منابع آب‌زیرزمینی، اکثر آبخوان‌های دشت‌های ایران را به‌خصوص در مناطق خشک و نیمه‌خشک کشور به‌طور قابل ملاحظه با کسری مخزن مواجه کرده و سطح آب‌زیرزمینی^۱ (GWL) در این مناطق به سرعت در حال کاهش است. برای آگاهی از وضعیت این منابع ارزشمند در آینده و برنامه‌ریزی و مدیریت پایدار آن‌ها، مدل‌سازی آبخوان‌ها از اهمیت بالایی برخوردار است. تاکنون روش‌های متعدد و رویکردهای متنوعی بر پایه انواع مدل‌های مفهومی، فیزیکی، عددی و آماری برای پیش‌بینی سطح

۱- دانشجوی کارشناسی‌ارشد گروه مهندسی آب، دانشکده مهندسی عمران و نقشه‌برداری، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران
۲- استادیار گروه مهندسی آب، دانشکده مهندسی عمران و نقشه‌برداری، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران
۳- دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه جیرفت، جیرفت، ایران
(*) نویسنده مسئول: Email: s.shahabi@kgut.ac.ir

مدل‌هایی که به آن‌ها اشاره شد، برای شبیه‌سازی سطوح آب‌های زیرزمینی در نظر گرفته شده‌اند و مطالعاتی نظیر (Rajaei et al., 2022; Tao et al., 2019) پس از مرور بر پژوهش‌های متعدد، توانایی بالای آن‌ها را در این زمینه گزارش کرده‌اند. با وجود عملکرد مطلوبی که مدل‌های AI در شبیه‌سازی پدیده‌های هیدرولوژیکی نشان داده‌اند، به علت دشواری درک دینامیک و تأثیرات متغیرهای مختلف بر خروجی شبیه‌سازی مورد انتقاد قرار گرفته‌اند (Solomatine and Xue, 2004). از این‌رو، مدل‌هایی همچون درخت تصمیم M5 با قابلیت رفع این موانع و ضعف‌ها به همراه کاربرد موفقیت‌آمیزی که در مدل‌سازی متغیرهای هیدرولوژیکی از جمله برآورد نوسانات سطح آب‌زیرزمینی دارند، توصیه می‌شود. در ادامه نمونه‌ای از پژوهش‌ها در زمینه بررسی قابلیت و قدرت مدل ذکر شده در تخمین سطح آب‌زیرزمینی شرح داده شده است.

در مطالعه‌ای پیش‌بینی سطح آب‌زیرزمینی با بهره‌مندی از توسعه درخت مدل M5 تنها با استفاده از سطوح آب‌زیرزمینی تاریخی از یک چاه پایش آب‌زیرزمینی مورد بحث و تحلیل قرار گرفت. نتایج حاصل از مدل‌سازی با حداقل داده، همبستگی زیادی را میان سطح آب اندازه‌گیری شده و پیش‌بینی شده نشان داد. بنابراین، این ادعا مطرح شد که مدل M5 می‌تواند ابزاری مفید و روشی امیدوارکننده برای پیش‌بینی در این زمینه باشد (Nalarajan and Mohandas, 2015). از مقایسه میان نتایج حاصل از اجرای سه مدل داده محور ANN، SVM و درخت تصمیم M5 در پیش‌بینی کوتاه‌مدت سطح آب‌زیرزمینی مشخص شد، روش‌های یادشده توانایی بهبود روند پیش‌بینی سطح آب‌زیرزمینی را دارند و مدل SVM ضعیف‌ترین و درخت تصمیم M5 بهترین عملکرد را داشت (Huang and Tian, 2015).

بررسی دقیق‌تر دیگر مطالعات انجام شده در این حیطه نشان داد، علی‌رغم انعطاف‌پذیری و سودمندی مدل‌های یادگیری ماشین در پیش‌بینی و مدل‌سازی فرآیندهای هیدرولوژیکی که به‌طور کلی عملکرد موفق و قابل قبولی داشته‌اند، هنگامی که سطح آب‌زیرزمینی تغییرات چندمقیاسی را نشان می‌دهد، نتایج آن‌ها از دقت کافی برخوردار نبوده و با وجود تمام مزایا، در صورتی که داده‌های ورودی آن غیر ایستا باشند، نتایج رضایت‌بخشی ارائه نمی‌دهند، این در حالی است که داده‌های هیدرولوژیکی اغلب غیرایستا می‌باشند. در نتیجه، زمانی که پدیده‌های هیدرولوژیکی بسیار غیرثابت هستند، این مدل‌ها ممکن است نتوانند سری‌های زمانی هیدرولوژیکی را به‌درستی شبیه‌سازی کنند (Nourani et al., 2014). در سال‌های اخیر روش تبدیل موجک^{۱۱} (WT) به‌عنوان یک ابزار قدرتمند در تجزیه و تحلیل سری‌های زمانی و رفع موانع و نقص‌ها در مدل‌سازی پدیده‌های هیدرولوژیکی با روش‌های یادگیری ماشین، کاربرد فراوانی یافته است. شماری از محققین در مطالعات خود از فرآیند پیش‌پردازشگر تبدیل

موجک در ترکیب با مدل‌های یادگیری ماشین برای بهینه شدن پاسخ‌ها استفاده کرده و تأثیر مطلوب آن را بر نتایج ارزیابی نمودند (Shahabi et al., 2016; Azarpira and Shahabi, 2021; Bahmani and Ouarda, 2021). تبدیل موجک، برای پیش‌پردازش سری‌های زمانی به قصد افزایش دقت مدل‌های مبتنی بر هوش مصنوعی پیشنهاد شده است (Solgi, et al., 2017). تبدیل موجک به هیدرولوژیست‌ها کمک می‌کند سیگنال زمان را به سیگنال‌های فرعی تجزیه کنند و این امکان را برای مدل‌ها فراهم می‌کند تا اطلاعاتی با سطوح تفکیک متفاوت دریافت کنند (Nourani et al., 2009). مهم‌ترین و بارزترین خصوصیت تبدیل موجک به‌عنوان یک پیشرفت در پردازش سیگنال این است که می‌تواند به‌طور قابل اعتمادی کاستی‌های مدل هوش مصنوعی را در برخورد با رفتار غیرثابت سیگنال‌ها برطرف کند (Grossmann and Morlet, 1984). کاربرد این تکنیک پیش‌پردازشگر همراه با روش‌های محاسبات نرم در زمینه مدل‌سازی نوسانات تراز آب‌زیرزمینی، نتایج را به‌گونه‌ای موفقیت‌آمیز بهبود داده است. از جمله در پژوهشی، دو مدل یادگیری ماشین شامل اسپلاین رگرسیون تطبیقی چند متغیره^{۱۲} (MARS) و درختان مدل^{۱۳} (M5(MT)) در ترکیب با موجک برای افزایش دقت پیش‌بینی نوسانات سطح آب‌زیرزمینی با مدل‌های ساده خود مقایسه شدند. نتایج این مطالعه صحت برتری مدل‌های ترکیبی نسبت به همتای ساده خود را تأیید کرد. همچنین مدل ترکیبی اسپلاین رگرسیون تطبیقی چند متغیره-موجک (W-MARS) پیش‌بینی دقیق‌تری را از سطح آب‌زیرزمینی ارائه داد (Rezaie-balf et al., 2017). دو مدل هیبریدی برنامه‌ریزی بیان ژن-موجک^{۱۴} (W-GEP) و درخت تصمیم M5-موجک (W-M5) به‌منظور شبیه‌سازی در سه چاه مشاهده‌ای توسعه داده شد. برای تجزیه سیگنال‌های زمانی، توابع موجک هار(Haar)، کویفلت^{۱۵} مرتبه یک (Coif1)، سیملت^{۱۶} از مرتبه سه (Sym3)، دوشی^{۱۷} مرتبه دو و چهار (db2 و db4) که به‌طور گسترده در مطالعات هیدرولوژیکی کاربرد دارند، استفاده شدند. این مطالعه نشان داد که مدل‌های هیبریدی W-GEP و W-M5 عملکرد بهبود یافته‌ای را نسبت به مدل‌های ساده خود نشان می‌دهند. همچنین گزارش شده است که انتخاب یک سطح تجزیه مناسب به‌طور قابل توجهی بر دقت مدل ترکیبی تأثیر می‌گذارد (Bahmani et al., 2020). سامانی و همکاران از ترکیب پیش-پردازنده موجک (W) با مدل‌های یادگیری ماشین (شامل ANN، ANFIS، روش دسته‌بندی گروهی داده‌ها^{۱۸} (GMDH) و رویکرد حداقل مربعات ماشین بردار پشتیبان^{۱۹} (LSSVM)) برای پیش‌بینی تغییرات سطح ایستایی دشت زرن-ساوه استفاده کردند. نتیجه این مطالعه نشان داد که مدل ترکیبی W-LSSVM بهترین عملکرد را در میان مدل‌های منفرد و ترکیبی مورد بررسی داشته است (Samani et al., 2022). ودیعی و همکاران کارایی مدل‌های ANN، منطق

مطالعه‌ای که صرفاً با استفاده از پارامترهای هواشناسی، سطح آب‌زیرزمینی را مدل کرده باشد، ملاحظه نشد. این رویکرد جدید به‌طور ویژه در مناطقی که اطلاعات اندازه‌گیری شده سطح آب‌زیرزمینی وجود ندارد یا با محدودیت مواجه باشد اهمیت بالایی دارد. بر این اساس در این پژوهش، صرفاً با استفاده از پارامترهای هواشناسی مدلی برای تخمین سطح آب‌زیرزمینی و بررسی عملکرد آن در یک تا سه ماه آینده ارائه شده است. مبنای انتخاب این پارامترها مطالعه تحقیقات گذشته و بررسی همبستگی میان ۱۴ پارامتر هواشناسی در دسترس طی دوره مطالعاتی با سطح آب‌زیرزمینی و در نظر گرفتن شرایط فیزیکی و اقلیمی محدوده مورد مطالعه بوده است. شایان ذکر است، نظر به افت شدید سطح آب‌زیرزمینی دشت کرمان- باغین در چند دهه گذشته و اهمیت استراتژیک آن در تأمین مصارف بخش‌های شرب، کشاورزی و صنعت، برآورد نوسانات تراز آب‌زیرزمینی در آینده برای اخذ تصمیم‌های اصولی، مدیریت برنامه ریزی‌ها برای برداشت از این منابع و آمادگی هرچه بیشتر جامعه در مقابل کمبودها و خشکسالی‌ها، لازم و اجتناب‌ناپذیر است. بر اساس دانش نویسندگان، سابقه‌ای از انجام تحقیقات مشابه در این منطقه وجود ندارد و این پژوهش در نوع خود برای اولین بار در این دشت استراتژیک صورت می‌گیرد. از این رو، پژوهش حاضر در نظر دارد از تکنیک تبدیل موجک برای پیش‌پردازش سری‌های زمانی ورودی به مدل M5P استفاده نماید. بدین ترتیب یک مدل ترکیبی W-M5P با هدف دستیابی به دقیق‌ترین نتایج (نزدیکی مقادیر مشاهداتی و پیش‌بینی شده) ارائه می‌شود. قابلیت مدل توسعه داده شده در پیش-بینی سطح آب‌زیرزمینی با پارامترهای هواشناسی مورد بررسی قرار می‌گیرد. همچنین میزان تأثیر انتخاب سطوح تجزیه و توابع موجک مختلف بر نتایج بررسی می‌گردد.

مواد و روش تحقیق

الگوریتم درخت تصمیم M5P

الگوریتم M5P (Wang and Witten, 1996) نسخه توسعه یافته الگوریتم M5 است که در ابتدا توسط کوینلان^{۳۰} (Quinlan, 1992) پیشنهاد شد. ویژگی‌های مثبت این مدل شامل نیاز به حجم کم داده‌های ورودی (Nalarajan and Mohandas, 2015)، دقت بالا (Huang and Tian, 2015; Rezaie-balf et al., 2017)، سادگی (Sattari et al., 2018) و عملکرد بالا (Javadinejad et al., 2020) می‌باشد. الگوریتم M5P رابطه میان پارامترهای ورودی و خروجی را ارزیابی می‌کند و از چهار مرحله اصلی تشکیل شده است. در مرحله اول، فضای ورودی به چند فضای فرعی تقسیم می‌شود تا یک درخت ساخته شود. تغییرپذیری درون زیرفضا از ریشه تا گره با استفاده از معیار تقسیم به حداقل می‌رسد. انحراف استاندارد مقادیری

فازی^{۲۰} (FL)، ANFIS، و LSSVM را برای پیش‌بینی سطح آب‌زیرزمینی دشت کرج-تهران مورد بررسی قرار دادند. آن‌ها ترکیبی از متغیرهای هواشناسی و هیدرولوژیکی را به‌عنوان پارامترهای ورودی این مدل‌ها در نظر گرفتند. نتایج این پژوهش حاکی از برتری نسبی مدل ANFIS به‌خصوص در پیش‌بینی یک‌ماهه و دو‌ماهه سطح آب‌زیرزمینی می‌باشد (Vadiati et al., 2022). عبدالله‌زاده و اشمالز به ارزیابی مدل‌های برنامه‌ریزی ژنتیک^{۲۱} (GP) و SVR در ترکیب با پیش‌پردازش موجک در پیش‌بینی نوسانات سطح آب پرداختند. شاخص‌های آماری نشان دادند که مدل ترکیبی W-SVR بهترین عملکرد را در مقایسه با سایر مدل‌ها از خود نشان داد. در مطالعه دیگری که توسط فام و همکاران انجام شد، عملکرد هفت مدل مختلف یادگیری ماشین شامل درخت تصادفی^{۲۲} (RT)، جنگل تصادفی^{۲۳} (RF)، الگوریتم یادگیری درخت^{۲۴} (DS)، مدل درخت SVM، M5T، رگرسیون خطی وزنی محلی^{۲۵} (LWLR) و هرس خطای کاهش یافته درخت^{۲۶} (REP Tree) برای پیش‌بینی سطح ایستابی مورد ارزیابی قرار گرفت. در این مطالعه از داده‌های هواشناسی و مقادیر سطح آب‌زیرزمینی دو چاه در شمال بنگلادش به‌عنوان پارامترهای ورودی مدل‌های مذکور استفاده شد. نتایج این مطالعه نشان داد که مدل‌های مبتنی بر RT و RF از عملکرد بالاتری برخوردار بودند (Pham et al., 2022). وی و همکاران از دو مدل SVM و RF به‌صورت منفرد و در ترکیب با پیش‌پردازش موجک (هفت نوع موجک) برای پیش‌بینی سطح آب‌زیرزمینی استفاده کردند و گزارش دادند که مدل‌های ترکیبی مورد استفاده با دقت بالایی توانستند نوسانات ایستابی را پیش‌بینی نمایند (Wei et al., 2022).

شناخت متغیرهای مؤثر و کاهش عدم قطعیت داده‌های ورودی، کارایی تکنیک‌های پیش‌بینی را افزایش می‌دهد. تحقیقات روی متغیرهای مؤثر بر سطح آب‌های زیرزمینی نشان‌دهنده تنوع و گستردگی این متغیرها، به‌ویژه در مقیاس زمانی است (Kardan Moghaddam et al., 2021). متغیرهای ورودی از جمله سطح آب‌زیرزمینی (GWL)، بارش (P)، دمای کمینه^{۲۷} (T_{min})، دمای بیشینه^{۲۸} (T_{max}) و رطوبت نسبی^{۲۹} (RH) (Sreekanth et al., 2009)، سطح آب‌زیرزمینی (GWL)، بارش (P)، دما^{۳۰} (T) و تبخیر^{۳۱} (E) (Rakhshandehroo et al., 2012)، سطح آب‌زیرزمینی (GWL) (Nalarajan and Mohandas, 2015)، سطح آب‌زیرزمینی (GWL)، تبخیر و تعرق^{۳۲} (ET) و دبی رودخانه^{۳۳} (Q) (Yue et al., 2018)، سطح آب‌زیرزمینی (GWL)، بارش (P)، دما (T)، میانگین سرعت باد^{۳۴} (ffm) و رطوبت نسبی (RH) (Üneş et al., 2019)، در مطالعات قبلی برای پیش‌بینی سطح آب‌زیرزمینی استفاده شده است. مبتنی بر دانش و مرور نویسندگان، در تحقیقات پیشین، مدل‌سازی و پیش‌بینی سطح آب‌زیرزمینی همواره با GWL به‌عنوان تنها پارامتر ورودی و یا در کنار پارامترهای هواشناسی صورت گرفته و

تحت مقیاس‌های چندزمانی تعیین می‌کند. الگوریتم \log_{10}^n که n طول سری داده‌ها است برای انتخاب سطح تجزیه توصیه می‌شود (Bahmani and Ouarda, 2021). هرگاه یک سری زمانی تا سطح m تجزیه شود، $m+1$ زیرسری ایجاد شده که یکی از آن‌ها زیرسری تقریب می‌باشد که دربرگیرنده مفاهیم اساسی سری است و m زیرسری باقیمانده جزئیات سری اصلی هستند (Misiti et al., 1996). به‌طور کلی، به سبب اینکه سری‌های زمانی هیدرولوژیکی دارای داده‌های پرت و غیر ایستا هستند و این امر موجب افزایش پیچیدگی و زمان بر شدن فرآیند محاسبات می‌شود، کاربرد تکنیک تبدیل موجک برای پیش‌پردازش داده‌های سری ضرورت دارد.

مدل ترکیبی موجک-درخت M5P

از آنجایی که مشخص شده است، مدل‌های هوش مصنوعی دارای محدودیت‌هایی مانند فرآیندهای غیرخطی و غیرثابت می‌باشند، برخی از رویکردهای مدل‌سازی ترکیبی که شامل پیش‌پردازش داده‌ها و یا ترکیب تکنیک‌های مختلف هوش مصنوعی هستند، در سال‌های اخیر برای افزایش قابلیت‌های هوش مصنوعی توسعه یافته‌اند. تجزیه و تحلیل موجک نمونه‌ای برای پیش‌پردازش داده‌ها است، که در مدل‌سازی GWL کاربرد فراوانی داشته است. مدل‌های ترکیبی شامل چند تکنیک منفرد و یا الگوریتم‌های بهینه‌سازی هستند که به‌طور معمول با در نظر گرفتن دقت مدل‌سازی، به مراتب قابل اعتمادتر از مدل‌های منفرد هستند (Ardabili et al., 2020).

از این‌رو در این پژوهش، از تجزیه و تحلیل موجک برای حذف نویز، فشردگی و تجزیه سری‌های زمانی داده‌های ورودی برای ایجاد مدل ترکیبی موجک-درخت تصمیم M5P با هدف برآورد سطح آبرزیرزمینی استفاده شده است و عملکرد مدل WM5P متأثر از فرآیند پیش‌پردازشگر خواهد بود. در این مدل ترکیبی، سری‌های زمانی ورودی پس از انتخاب نوع موجک مادر (بر اساس بررسی شباهت و تطبیق بصری بیشتر با سری خام) و مقیاس زمانی (سطح تجزیه) مناسب که اثربخشی به‌سزایی بر نتیجه‌گیری نهایی دارند، تجزیه شده و به مدل منفرد M5P داده می‌شوند. نمایی شماتیک از مدل ترکیبی در شکل ۱ نشان داده شده است.

معیار ارزیابی کارایی مدل‌ها

برای بررسی و اعتبارسنجی کارایی مدل‌ها از شاخص‌های آماری نظیر ضریب همبستگی^{۳۷} (R)، جذر میانگین مربعات خطا^{۳۸} (RMSE)، خطای مطلق میانگین^{۳۹} (MAE)، ضریب کارایی ناش-ساتکلیف^{۴۰} (NSE)، شاخص توافق^{۴۱} (I_a) و شاخص پراکندگی^{۴۲} (SI) که به‌وسیله روابط زیر محاسبه می‌شوند، استفاده شده است:

$$R = \frac{\sum_{i=1}^n (F_i - \bar{F}_i)(O_i - \bar{O}_i)}{\sqrt{\sum_{i=1}^n (F_i - \bar{F}_i)^2 \sum_{i=1}^n (O_i - \bar{O}_i)^2}} \quad (2)$$

که به آن گره می‌رسد برای اندازه‌گیری تغییرپذیری استفاده می‌شود. ساخت درخت با استفاده از ضریب کاهش انحراف استاندارد (SDR) انجام می‌شود که کاهش خطای مورد انتظار در گره را به‌شرح زیر به حداکثر می‌رساند:

$$SDR^{*g} = sd(S) - \sum_{i=1}^n \frac{Si}{S} \times sd(Si) \quad (1)$$

که در آن S مجموعه‌ای از رکوردهای داده‌ای است که به گره می‌رسد، Si مجموعه‌های حاصل از تقسیم گره بر اساس یک ویژگی مشخص و sd انحراف معیار است. پس از ساخت درخت، در مرحله دوم، یک مدل رگرسیون خطی در هر یک از فضاهای فرعی با استفاده از داده‌های مرتبط با آن زیرفضا توسعه می‌یابد. سپس، یک تکنیک هرس برای غلبه بر مشکل برازش بیش از حد اعمال می‌شود. به این معنی که عمل هرس، شاخه‌هایی از درخت را که دقت مدل را افزایش نمی‌دهد، حذف (هرس) می‌کند. با این حال، فرآیند هرس می‌تواند باعث ناپوستگی شدید بین مدل‌های خطی مجاور شود. برای جبران این مشکل در مرحله نهایی، فرآیند هموارسازی انجام می‌شود. در این فرآیند، تمام مدل‌ها از برگ تا ریشه با هم ترکیب می‌شود تا مدل نهایی برگ را ایجاد کند و مقدار پیش‌بینی شده برگ هنگام بازگشت به ریشه فیلتر می‌شود (Wang and Witten, 1996). برای آشنایی و کسب اطلاعات بیشتر در مورد M5P و M5 می‌توان به (Quinlan, 1992; Wang and Witten, 1996, Solomatine and Xue, 2004) مراجعه کرد.

تبدیل موجک

موجک یک تحلیل طیفی وابسته به زمان است که سری‌های زمانی را در فضای فرکانس زمانی باز می‌کند تا توصیفی در مقیاس زمانی از فرآیندها و روابط آن‌ها ارائه دهد (Daubechies, 1990). تبدیل موجک به‌منظور تجزیه فرکانس‌های یک سری زمانی مورد استفاده قرار می‌گیرد و از ویژگی‌های بارز آن، توانایی به‌دست آوردن اطلاعات در زمان، تناوب و موقعیت به‌طور همزمان می‌باشد (Misiti et al., 1996). تحلیل و پیش‌بینی سری‌های زمانی هیدرولوژیکی یک رویکرد مؤثر برای تعیین تغییرپذیری فرآیندهای هیدرولوژیکی و برآورد مقادیر آینده می‌باشد (Nourani et al., 2014). موجک‌ها نوعی تابع ریاضی (شکل موج) هستند که با میانگین حدود صفر، ممکن است در مدت زمان کوتاهی نوسان و فروپاشی کنند (Rezaie-balf et al., 2017). دو موضوع انتخاب موجک مادر و مقیاس زمانی مناسب، اساس تمام تحلیل‌های موجک و از مسائل کلیدی می‌باشد که بر پیش‌پردازش و مدل‌سازی داده به کمک موجک تأثیر می‌گذارد (Tao et al., 2022). انتخاب مقیاس زمانی مناسب موضوع مهمی برای تحلیل موجک و به‌ویژه برای مدل‌سازی موجک است. زیرا به طور مستقیم دقت ویژگی‌های شناسایی شده در یک سری زمانی را

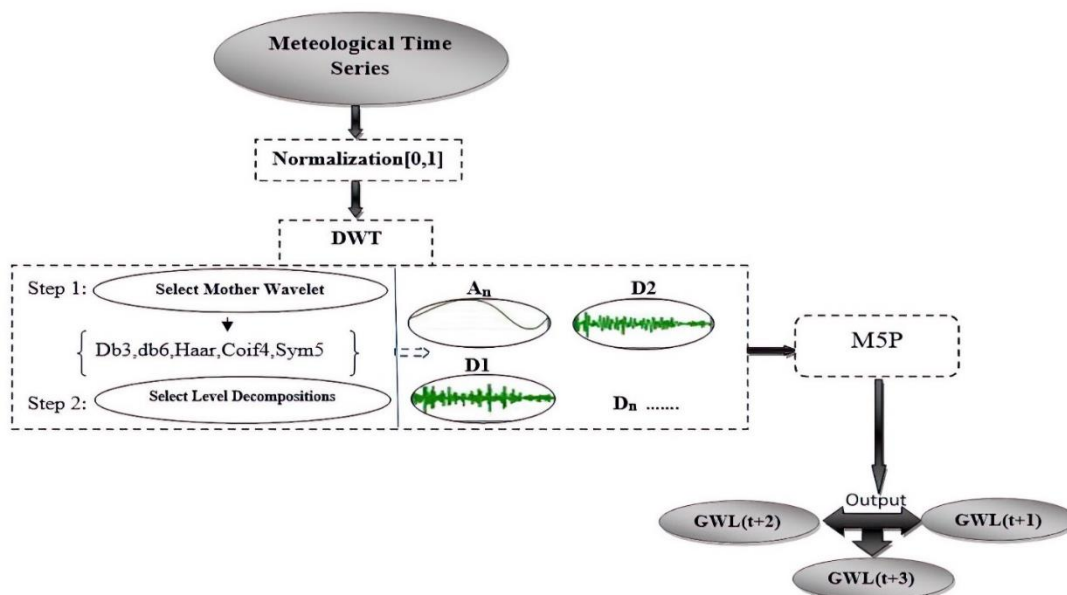
$$I_a = 1 - \frac{\sum_{i=1}^n (O_i - F_i)^2}{\sum_{i=1}^n (|F_i - \bar{O}_i| + |O_i - \bar{O}_i|)^2} \quad (۶)$$

$$SI = \frac{RMSE}{\bar{O}} \quad (۷)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (F_i - O_i)^2}{n}} \quad (۳)$$

$$MAE = \frac{\sum_{i=1}^n |F_i - O_i|}{n} \quad (۴)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - F_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \quad (۵)$$



شکل ۱- شماتیک مدل ترکیبی WM5P برای پیش‌بینی سطح آب‌یرزمینی در سه ماه آینده

در روابط ذکرشده، O_i و F_i به ترتیب مقادیر ماهانه پیش‌بینی شده و مشاهده شده سطح آب‌یرزمینی، \bar{O}_i و \bar{F}_i نیز میانگین ماهانه مقادیر مشاهداتی و پیش‌بینی شده و n تعداد کل داده‌ها می‌باشد. مقادیر شاخص‌های RMSE و MAE و SI هر چه کوچک‌تر و به صفر نزدیک باشند، مدل مدنظر در پیش‌بینی قوی‌تر است. هر اندازه مقادیر ضریب همبستگی R به عدد یک متمایل‌تر باشد، نتایج پیش‌بینی و مقادیر مشاهداتی همبستگی بالاتری دارند. شاخص NSE (معیاری بی بعد) نیز هر چه به یک نزدیک‌تر باشد، مدل دقیق‌تر است و مقادیر کمتر از صفر عملکرد ضعیف و عدم قابلیت مدل را نشان می‌دهد. همچنین هرچه شاخص I_a (بی بعد) به یک نزدیک‌تر باشد، دقت مدل در برآورد سطح آب‌یرزمینی بالاتر است.

تهیه اطلاعات موردنیاز و اجرای روش‌ها

تحلیل و آماده‌سازی پارامترهای ورودی و خروجی مدل

در این مطالعه با هدف پیش‌بینی نوسانات سطح آب‌یرزمینی در یک، دو و سه ماه آینده در دشت کرمان-باغین، داده‌های سطح آب‌یرزمینی ۳۲ حلقه چاه پیژومتری به صورت ماهانه از سال آبی ۸۱-۱۳۸۰ تا ۹۹-۱۳۹۸ از شرکت مدیریت منابع آب ایران تهیه شد. سطح آب‌یرزمینی در حدود نیمی از این چاه‌ها، در شماری از ماه‌ها و سال‌ها ثبت نشده و آمار ناقص داشته که برای رفع این مشکل، داده‌های چاه ناقص به کمک آمار چاه‌های مجاور خود با روش تفاضل و نسبت‌ها و درون‌یابی-برون‌یابی بازسازی شد. لازم به ذکر است، میانگین GWL در کل منطقه طی دوره مدنظر، با استفاده از روش چندضلعی‌های تیسن محاسبه گردید.

علاوه بر آن، از میانگین ماهانه هشت پارامتر هواشناسی شامل بارش (P)، متوسط دما (T)، حداقل دما (t_{min})، حداقل رطوبت نسبی (u_{min})، دمای نقطه شبنم (t_{dm})، سرعت باد (ffm)، فشار متوسط

در روابط ذکرشده، O_i و F_i به ترتیب مقادیر ماهانه پیش‌بینی شده و مشاهده شده سطح آب‌یرزمینی، \bar{O}_i و \bar{F}_i نیز میانگین ماهانه مقادیر مشاهداتی و پیش‌بینی شده و n تعداد کل داده‌ها می‌باشد. مقادیر شاخص‌های RMSE و MAE و SI هر چه کوچک‌تر و به صفر نزدیک باشند، مدل مدنظر در پیش‌بینی قوی‌تر است. هر اندازه مقادیر ضریب همبستگی R به عدد یک متمایل‌تر باشد، نتایج پیش‌بینی و مقادیر مشاهداتی همبستگی بالاتری دارند. شاخص NSE (معیاری بی بعد) نیز هر چه به یک نزدیک‌تر باشد، مدل دقیق‌تر است و مقادیر کمتر از صفر عملکرد ضعیف و عدم قابلیت مدل را نشان می‌دهد. همچنین هرچه شاخص I_a (بی بعد) به یک نزدیک‌تر باشد، دقت مدل در برآورد سطح آب‌یرزمینی بالاتر است.

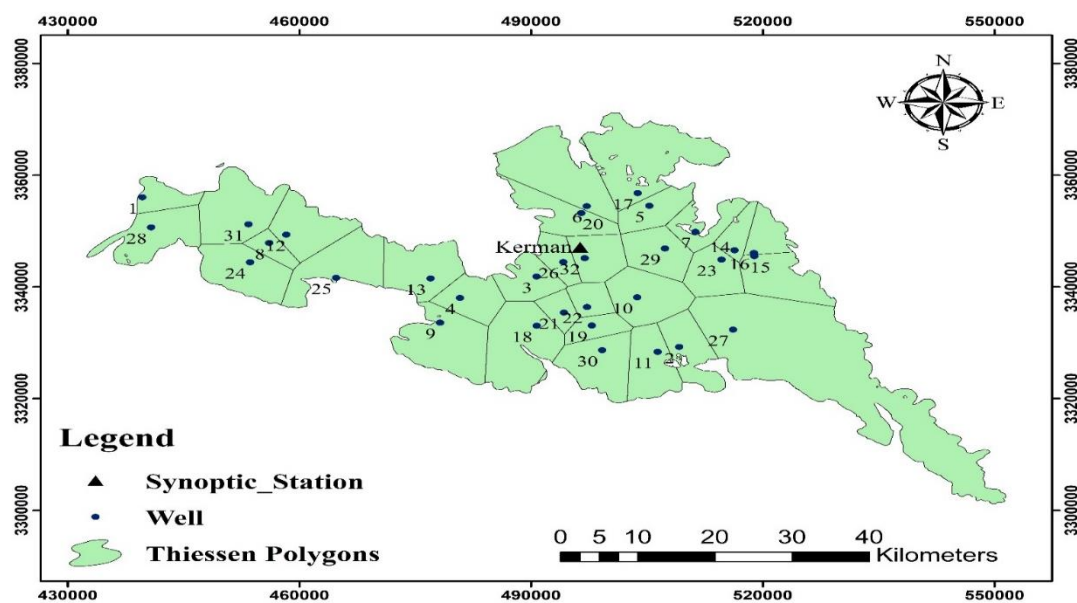
منطقه مورد مطالعه

محدوده مطالعاتی کرمان-باغین یکی از ۱۲ محدوده مطالعاتی حوضه آبریز درانجیر و ساغند بوده و در جنوب این حوضه آبریز قرار دارد. مساحت محدوده مطالعاتی ۵۴۰۴ کیلومتر مربع می‌باشد که ۲۵۵۸ کیلومتر مربع آن دشت و باقی آن، یعنی ۲۸۴۵ کیلومتر مربع، ارتفاعات می‌باشد. این محدوده مطالعاتی بین طول جغرافیایی ۱۸' - ۵۶° تا ۳۵' - ۵۷° شرقی و عرض جغرافیایی ۴۶' -

فرآیندها بر سطح آب‌زیرزمینی تأثیرگذار باشد. اول اینکه دمای هوا بر میزان مصرف آب تأثیر می‌گذارد. اگرچه دما ممکن است مستقیماً بر آبخوان تأثیر نگذارد، اما در مناطق خشک و نیمه‌خشک، همچون بخش‌های وسیعی از جغرافیای ایران، مصرف آب کشاورزی و خانگی به آب‌های زیرزمینی وابسته است و با توجه به دمای اتمسفر افزایش یا کاهش می‌یابد. در نتیجه دمای هوا می‌تواند به‌طور غیرمستقیم بر نوسانات آب‌زیرزمینی تأثیر بگذارد. دوم، دمای هوا ممکن است بر تبخیر در لایه بالایی خاک و همچنین سرعت تعرق گیاه تأثیرگذار باشد. بنابراین تغذیه مجدد آبخوان از بارش را کاهش می‌دهد. در اینجا نیز دمای هوا از طریق تأثیر آن بر تبخیر و تعرق، تأثیر غیرمستقیم بر سطح آب‌زیرزمینی دارد (Rezaie-balf et al., 2017). با مشاهده شکل ۲، موقعیت قرارگیری چاه‌ها و ایستگاه سینوپتیک با چندضلعی‌های تیسن مشخص می‌شود. در ادامه خصوصیات آماری سری زمانی پارامترهای هواشناسی در جدول ۱ ذکر شده است.

ایستگاه (P_{om}) و تعداد ساعات آفتابی (sshn) ثبت شده در ایستگاه سینوپتیک کرمان طی دوره ۱۹ساله استفاده شده است. انتخاب شماری از متغیرها برای مدل‌سازی با تکیه بر دانش محققین و همین‌طور بر اساس مطالعات پیشین با بررسی میزان همبستگی متغیرها با متغیر خروجی برای یافتن بهترین متغیرهای ورودی به مدل و انتخاب متغیرهای با مقادیر بالاتر ضریب همبستگی که از ۰ تا ۱ تغییر می‌کند، می‌باشد (Rajaei et al., 2019). این تکنیک به‌طور قابل قبولی توسط فرناندو و کر (۲۰۰۳) مورد استفاده قرار گرفت. از این‌رو، همبستگی میان تمام ۱۴ سری زمانی پارامترهای هواشناسی در دسترس با سطح آب‌زیرزمینی طول دوره مدنظر محاسبه شده و پارامترهای دارای مقادیر بالاتر همبستگی که نشان‌دهنده تطابق بهتر متغیر با سطح آب است، انتخاب شده‌اند.

برای نمونه، متغیرهایی نظیر بارش و دما به این دلیل انتخاب می‌شوند که هر دو به‌طور مستقیم بر سطح آب‌زیرزمینی تأثیر می‌گذارند. در واقع، دمای هوا ممکن است از طریق تعدادی از



شکل ۲- چندضلعی‌های تیسن منطقه مورد مطالعه همراه با موقعیت چاه‌های پیزومتریک و ایستگاه هواشناسی

جدول ۱- ویژگی‌های آماری پارامترهای هواشناسی ایستگاه سینوپتیک کرمان (ماهانه)

پارامتر	بارش (mm)	دمای حداقل (C°)	دمای متوسط (C°)	سرعت باد (m/s)	دمای نقطه شبنم (C°)	حداقل رطوبت نسبی (%)	تعداد ساعات آفتابی	فشار متوسط ایستگاه
حداقل	۰	-۵/۷۸۷	۱/۷۵۲	۱/۲۱۸	-۱۶/۱۷۵	۱	۱۶۰/۴	۸۱۸/۷۱۴
حداکثر	۹۰/۶۱	۲۰/۳۷۴	۲۹/۱۸۷	۴/۹۹۶	۳/۹۸۲	۱۵	۳۸۶/۷	۸۲۹/۶۶۶
میانگین	۹/۵۹۴	۷/۹۷۷	۱۶/۸۶۴	۲/۹۶۹	-۴/۱۱۶	۵/۱۱۱	۲۸۲/۱۷۵	۸۲۴/۶۸۹
انحراف معیار	۱۵/۲۸۷	۷/۴۰۵	۷/۷۵۳	۰/۷۵۸	۴/۱۲۶	۲/۲۷۳	۵۴/۰۱۳	۲/۷۰۲
چولگی	۲/۳۲۸	-۰/۰۵۷	-۰/۰۸	-۰/۲۲۶	-۰/۴۴۹	۰/۸۸۷	-۰/۱۳۷	-۰/۲۷۵

پیش‌پردازش داده‌ها و شیوه مدل‌سازی

با توجه به غیریکنواخت بودن سری زمانی متغیرهای ورودی و سطح آب‌ریز زمینی مورد استفاده در پژوهش، از رابطه ۸ برای نرمال‌سازی آن‌ها استفاده شده است. نظر به این که تقسیم داده‌ها به بخش‌های آموزش و آزمایش به‌عنوان ورودی به مدل از مسائل حائز اهمیت می‌باشد (Lallahem et al., 2005)، درصدهای مختلفی از داده‌ها برای ورودی به مدل (۸۰٪-۲۰٪، ۷۵٪-۲۵٪، و ۷۰٪-۳۰٪) مورد آزمون و خطا قرار گرفت و در پایان مناسب‌ترین تقسیم یعنی ۷۵٪ داده‌ها برای آموزش و ۲۵٪ باقی‌مانده برای آزمایش به مدل M5P داده شد. همچنین برای ارتقاء دقت مدل M5P و بهبود عملکرد آن، سری‌های زمانی ورودی نرمال شده، با استفاده از کد تجزیه با توابع موجک مادر مختلف و سطوح تجزیه مختلف به زیرسری‌هایی شامل تقریب (A) و جزئیات (D) تجزیه شدند که ضریب تقریب نشان دهنده وضوح بالا و روندهای هموار در سری‌ها و ضریب جزئیات اطلاعاتی جزئی از روندهای موجود در زیرسری‌ها را به نمایش می‌گذارد. باتوجه به اهمیت و تأثیر انتخاب موجک مادر در ارتقاء و تفاوت در نتایج ارائه شده، برای غلبه بر این مشکل، شباهت

بین سیگنال اصلی و موجک مادر در انتخاب نوع موجک مادر در نظر گرفته می‌شود. در این جا ۴ موجک که مشخصات آن‌ها در جدول ۲ ارائه شده است، با نام‌های دوشی مرتبه ۳ و ۶ (db3 و db6)، که این نوع موجک مادر به‌طور گسترده در مطالعات هیدرولوژیکی مورد استفاده قرار می‌گیرد، کوئفلت مرتبه ۴ (coif4) و سیملت مرتبه ۵ (sym5) به‌دلیل شباهت با سری زمانی اصلی و موجک هار (Haar) به‌سبب ماهیت تاریخی و مقایسه ساده میان مدل منفرد و ترکیبی به‌کار گرفته شده است. علاوه بر این انتخاب سطح تجزیه مناسب بر کارایی مدل‌های ترکیبی اثرگذار است. در این پژوهش از سطوح تجزیه ۳ تا ۶ استفاده شد. در پایان نتایج به‌دست آمده از مدل منفرد M5P و ترکیبی WM5P با شاخص‌های آماری برای ارزیابی عملکرد و اعتبارسنجی مدل‌ها استفاده شده‌اند.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (8)$$

که در آن، X_{min} کوچکترین داده در سری زمانی، X_{max} بزرگترین داده و X مقدار داده مورد نظر می‌باشد.

جدول ۲- موجک‌های مورد استفاده در تحقیق همراه با خصوصیات آن‌ها

موجک مادر	خصوصیات
Haar	✓ موجک‌های متقارن و غیر پیوسته هستند.
	✓ قدیمی‌ترین نوع موجک می‌باشد که شبیه یک تابع گام است.
	✓ کاربرد در سیگنال‌هایی که تغییرات شدید دارند، به‌دلیل گستره نسبتاً باریکی که انرژی آن روی آن توزیع می‌شود.
خانواده Daubechies (db3 و db6)	✓ موجک مرتبه اول خانواده Daubechies به‌عنوان موجک هار نیز شناخته می‌شود.
	✓ مرتبه توابع Daubechies تعداد لحظه‌های ناپدید شدن یا تعداد گشتاورهای صفر تابع موجک را نشان می‌دهد.
	✓ موجک‌های متقارن و گسسته هستند.
Coif4	✓ موجک‌های متقارن فشرده‌ای هستند.
	✓ موجک‌های کوئفلت متقارن‌تر هستند.
	✓ ممان‌های ناپدید شدن بیشتری نسبت به موجک‌های Daubechies دارند.
Sym5	✓ Symlet ها نیز موجک‌های متعامد و فشرده‌ای هستند.
	✓ تقریباً متقارن هستند و کمترین عدم تقارن را دارند.
	✓ ویژگی‌های symlet تقریباً شبیه موجک‌های db است.

بحث و تحلیل نتایج

نتایج مدل منفرد M5P

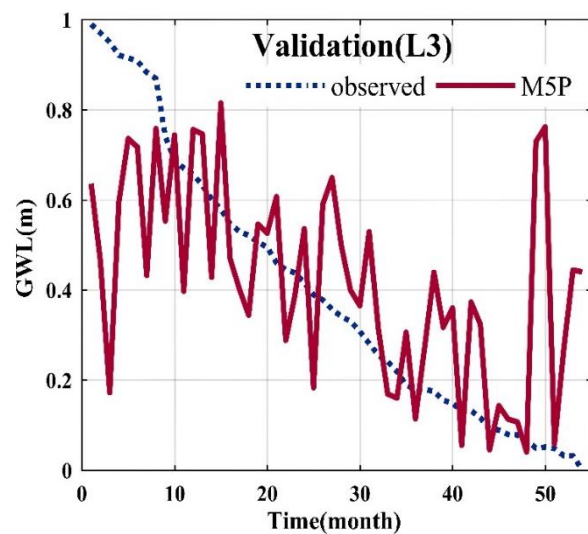
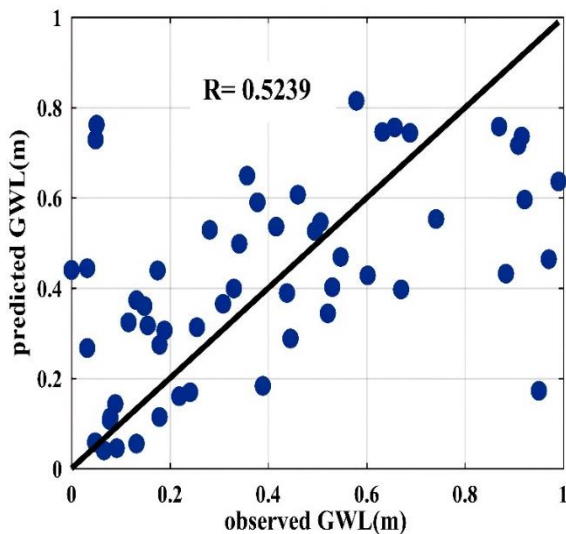
در ابتدای این پژوهش، سری‌های زمانی هشت پارامتر هواشناسی به‌عنوان ورودی و سطح آب‌ریز زمینی به‌عنوان خروجی، به مدل M5P معرفی شد. باتوجه به نتایج حاصل از مدل M5P که برای داده‌های آموزش و اعتبارسنجی افق‌های پیش‌بینی یک، دو و سه

ماهه در جدول ۳ ارائه شده، عملکرد مدل منفرد در هر ماه نسبت به ماه قبل از خود هر چند با اختلاف اندک افزایش داشته که مقادیر پیش‌بینی شده به‌تدریج به مقادیر مشاهداتی منطبق شده است. مقادیر شاخص‌های R و MAE از ماه اول تا سوم به‌ترتیب از ۰/۵۱۱۴ و ۰/۱۹۹۹ و ۰/۵۲۳۹ و ۰/۱۹۴۷ رسیده و بیانگر این موضوع می‌باشد که در زمان‌های پیش‌بینی طولانی‌تر، عدم قطعیت‌ها کاهش یافته و دقت مدل در تخمین سطح آب‌ریز زمینی بهبود می‌یابد. نتایج

پیش‌بینی مدل M5P و مقادیر مشاهده شده سه ماه آینده در شکل ۳ نشان داده شده است. همچنین نمودار پراکندگی این مدل با $R=0.5239$ در شکل ۴ رسم شده است. در این شکل‌ها سطح آب در مقیاس نرمال شده ترسیم شده است.

جدول ۳- نتایج حاصل از مدل M5P در افق‌های پیش‌بینی یک تا سه ماه

اعتبارسنجی						آموزش						مدل	افق پیش‌بینی
R	MAE	RMSE	I_a	NSE	SI	R	MAE	RMSE	I_a	NSE	SI		
۰/۵۱۱۴	۰/۱۹۹۹	۰/۲۵۹۷	۰/۶۸۹۸	۰/۲۳۶۱	۰/۶۳۹۴	۰/۷۱۷۶	۰/۱۵۴۴	۰/۲۰۶۲	۰/۸۱۶۴	۰/۵۱۴۷	۰/۴۵۶۴	M5P	یک ماه
۰/۵۲۱۱	۰/۱۹۵۲	۰/۲۶۱۷	۰/۷۱۰۵	۰/۲۲۱۸	۰/۶۵۳۲	۰/۷۶۳۷	۰/۱۳۹۸	۰/۱۹	۰/۸۴۷۶	۰/۵۸۱۲	۰/۴۲۲۰	M5P	دو ماه
۰/۵۲۳۹	۰/۱۹۴۷	۰/۲۶۰۷	۰/۷۱۳۲	۰/۲۲۵۲	۰/۶۵۶۲	۰/۷۷۱۲	۰/۱۳۷۶	۰/۱۸۶	۰/۸۵۳۷	۰/۵۹۲۸	۰/۴۱۵۱	M5P	سه ماه



شکل ۳- توزیع مقادیر پیش‌بینی شده و مقادیر مشاهداتی ماه سوم شکل ۴- نمودار پراکندگی نتایج پیش‌بینی سطح آب زیرزمینی ماه سوم

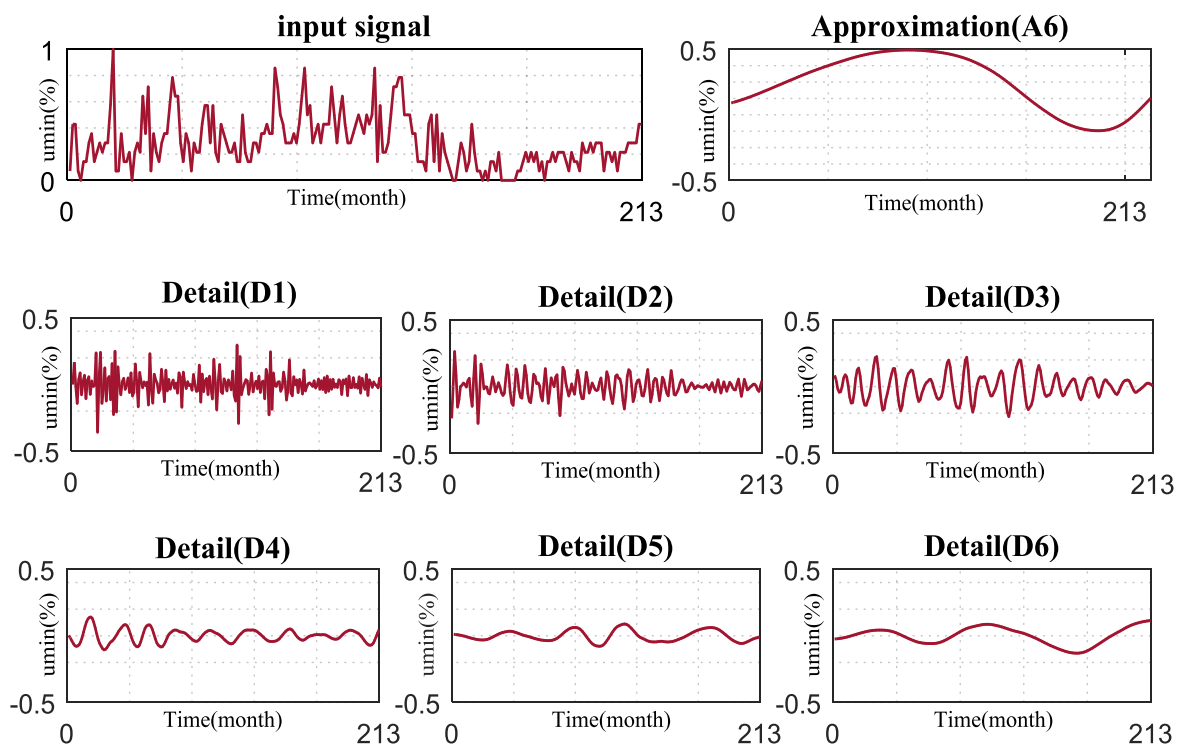
را ثابت می‌کند. با بررسی شاخص‌های مختلف در تحلیل قابلیت تعمیم مدل، موجک Haar با شش سطح تجزیه در ماه اول و دوم و Symlet با پنج سطح تجزیه در ماه سوم، نسبت به موجک‌های انتخابی دیگر ضعیف‌تر است. به‌طور کلی سطوح تجزیه پنج و شش بیشترین تأثیر را در افزایش قابلیت پیش‌بینی با مدل‌های ترکیبی در هر سه ماه دارند. ساختمان مدل ترکیبی برگزیده که در ماه اول دارای پنج سطح تجزیه و ماه‌های دوم و سوم دارای شش سطح تجزیه می‌باشد، بیانگر این موضوع می‌باشد که سطح بالای تجزیه مطلقاً مطلوب‌ترین و کارآمدترین نتایج را ارائه نمی‌دهد.

ارتباط بسیار نزدیک توزیع نتایج پیش‌بینی داده‌های آزمایش مدل WM5P با مقادیر مشاهداتی سطح آب زیرزمینی در ماه‌های مختلف و همین‌طور نمودارهای پراکندگی با مقدار ضریب همبستگی محاسبه شده برای داده‌های آزمایش در شکل ۶ به نمایش درآمده است.

نتایج مدل ترکیبی موجک-M5P (WM5P)

با به‌کارگیری موجک‌های مادر و مقیاس‌های زمانی مختلف برای تجزیه سیگنال‌های هواشناسی ورودی به سیگنال‌های فرعی A و D و اعمال آن‌ها بر مدل M5P، مدل ترکیبی WM5P به‌دست آمده و برترین نتایج حاصل از اجرای مدل در هر ماه، در جدول ۴ ارائه شده است. به‌عنوان نمونه، نمودارهای تجزیه سری زمانی حداقل رطوبت نسبی (سیگنال ورودی) به زیرسری‌های شامل یک تقریب و شش جزئیات در ماه سوم با موجک دوبشی مرتبه ششم و تا شش سطح تجزیه در شکل ۵ نمایش داده شده است.

با توجه به جدول ۴ می‌توان نتیجه گرفت که پیش‌پردازش داده‌ها با استفاده از موجک منجر به بهبود دقت مدل منفرد شده است. M5P همراه با موجک Coif4 در هر سه ماه بهترین عملکرد را از خود نشان داده و به این معنی است که از نظر ساختار شباهت بیشتری با شکل ظاهری سری زمانی مادر داشته و برتری آن نسبت به دیگر موجک‌ها



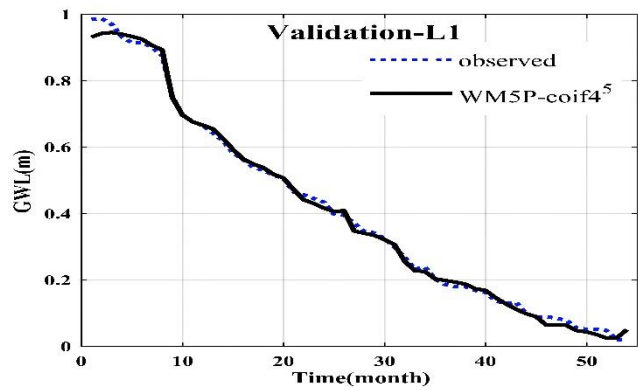
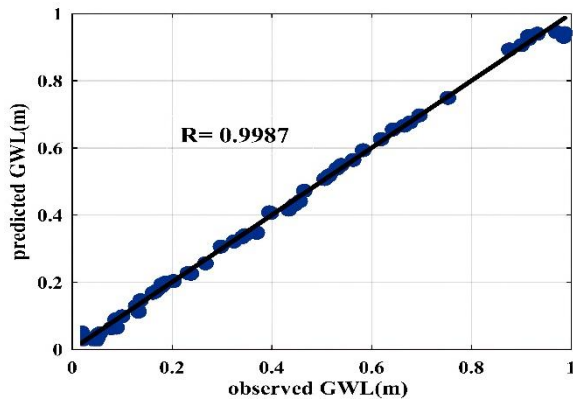
شکل ۵- تجزیه حداقل رطوبت نسبی به‌عنوان سیگنال ورودی با شش سطح با استفاده از موجک db6 در ماه سوم

جدول ۴- نتایج مدل ترکیبی WMSP برای افق‌های پیش‌بینی یک تا سه ماه*

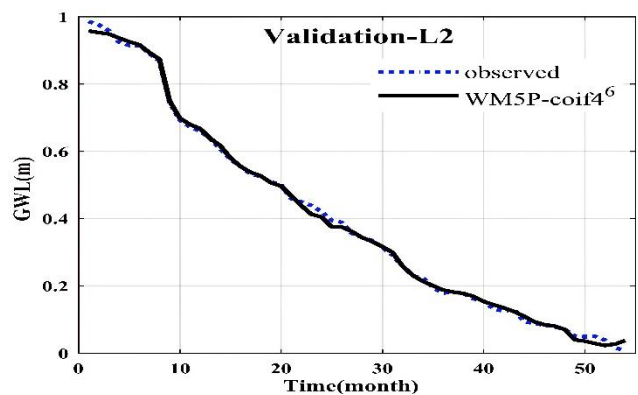
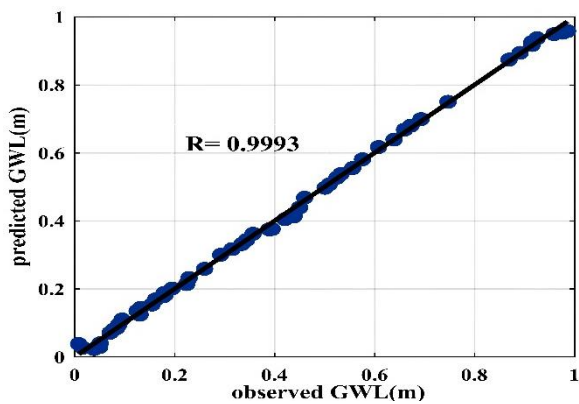
اعتبارسنجی						آموزش						مدل	افق پیش‌بینی
R	MAE	RMSE	I _a	NSE	SI	R	MAE	RMSE	I _a	NSE	SI		
۰/۹۹۲۴	۰/۰۲۹۱	۰/۰۳۷۵	۰/۹۹۵۹	۰/۹۸۴۲	۰/۰۹۲	۰/۹۹۱۸	۰/۰۲۹۸	۰/۰۳۸۴	۰/۹۹۵۷	۰/۹۸۳۳	۰/۰۷۷	M5P-Haar ⁶	یک ماه
۰/۹۹۶۳	۰/۰۲۲۷	۰/۰۲۸	۰/۹۹۷۷	۰/۹۹۱۰	۰/۰۶۸۹	۰/۹۹۶۷	۰/۰۱۸۵	۰/۰۲۴۶	۰/۹۹۸۲	۰/۹۹۳۱	۰/۰۵۴۴	M5P-db6 ⁶	
۰/۹۹۶۳	۰/۰۲۱۷	۰/۰۲۳۳	۰/۹۹۷۸	۰/۹۹۱۶	۰/۰۶۷۲	۰/۹۹۶۳	۰/۰۲۰۷	۰/۰۲۶۳	۰/۹۹۸۰	۰/۹۹۲۱	۰/۰۵۸۲	M5P-Sym5 ⁶	
۰/۹۹۶۹	۰/۰۱۷۲	۰/۰۲۳۹	۰/۹۹۸۳	۰/۹۹۳۵	۰/۰۵۸۸	۰/۹۹۸۵	۰/۰۱۳۵	۰/۰۱۶۵	۰/۹۹۹۲	۰/۹۹۶۹	۰/۰۳۶۵	M5P-db3 ⁶	
۰/۹۹۸۷	۰/۰۱۲	۰/۰۱۶	۰/۹۹۹۹	۰/۹۹۷	۰/۰۳۸	۰/۹۹۸۹	۰/۰۰۹۳	۰/۰۱۴۱	۰/۹۹۹۴	۰/۹۹۷۷	۰/۰۳۱۲	M5P-Coif4 ⁵	
۰/۹۹۱۱	۰/۰۳۱۶	۰/۰۴۰۵	۰/۹۹۵۱	۰/۹۸۱۴	۰/۱۰۱۱	۰/۹۹۱۲	۰/۰۳۰۶	۰/۰۳۹۶	۰/۹۹۵۳	۰/۹۸۱۹	۰/۰۸۷۹	M5P-Haar ⁶	دو ماه
۰/۹۹۲۵	۰/۰۲۲۲	۰/۰۳۶۳	۰/۹۹۶۲	۰/۹۸۵۰	۰/۰۹۰۶	۰/۹۹۵۴	۰/۰۲۱۷	۰/۰۲۸۷	۰/۹۹۷۵	۰/۹۹۰۴	۰/۰۶۳۷	M5P-Sym5 ⁵	
۰/۹۹۸۹	۰/۰۱۱۹	۰/۰۱۴۷	۰/۹۹۹۴	۰/۹۹۷۵	۰/۰۳۶۷	۰/۹۹۸۹	۰/۰۱۱۳	۰/۰۱۴	۰/۹۹۹۴	۰/۹۹۷۷	۰/۰۳۱۱	M5P-db6 ⁶	
۰/۹۹۹۲	۰/۰۰۹۸	۰/۰۱۲۴	۰/۹۹۹۶	۰/۹۹۸۳	۰/۰۳۰۹	۰/۹۹۹۱	۰/۰۱۰۶	۰/۰۱۲۸	۰/۹۹۹۵	۰/۹۹۸۱	۰/۰۲۸۴	M5P-db3 ⁶	
۰/۹۹۹۳	۰/۰۰۸۴	۰/۰۱۱	۰/۹۹۹۶	۰/۹۹۸۶	۰/۰۲۷۷	۰/۹۹۹۴	۰/۰۰۷۲	۰/۰۱	۰/۹۹۹۷	۰/۹۹۸۸	۰/۰۲۲۲	M5P-Coif4 ⁶	
۰/۹۸۳	۰/۰۲۰۶	۰/۰۵۴۹	۰/۹۹۱۰	۰/۹۹۵۶	۰/۱۳۸۲	۰/۹۹۷۹	۰/۰۱۵۵	۰/۰۱۹۵	۰/۹۹۸۹	۰/۹۹۵۵	۰/۰۴۳۵	M5P-Sym5 ⁶	سه ماه
۰/۹۸۹۹	۰/۰۳۲۹	۰/۰۴۲۹	۰/۹۹۴۵	۰/۹۷۸۹	۰/۱۰۸	۰/۹۹۱۲	۰/۰۳۰۶	۰/۰۳۹۳	۰/۹۹۵۳	۰/۹۸۱۸	۰/۰۸۷۷	M5P-Haar ⁶	
۰/۹۹۹۴	۰/۰۰۸۵	۰/۰۱۰۸	۰/۹۹۹۷	۰/۹۹۸۷	۰/۰۲۷۲	۰/۹۹۹۳	۰/۰۰۹۱	۰/۰۱۱۲	۰/۹۹۹۶	۰/۹۹۸۵	۰/۰۲۵	M5P-db3 ⁶	
۰/۹۹۹۷	۰/۰۰۶۱	۰/۰۰۷۸	۰/۹۹۹۸	۰/۹۹۹۳	۰/۰۱۹۶	۰/۹۹۹۶	۰/۰۰۶۱	۰/۰۰۷۹	۰/۹۹۹۸	۰/۹۹۹۲	۰/۰۱۷۶	M5P-db6 ⁶	
۰/۹۹۹۷	۰/۰۰۵۴	۰/۰۰۷۲	۰/۹۹۹۸	۰/۹۹۹۴	۰/۰۱۸۱	۰/۹۹۹۴	۰/۰۰۷۳	۰/۰۰۹۸	۰/۹۹۹۷	۰/۹۹۸۸	۰/۰۲۱۹	M5P-Coif4 ⁶	

*در جدول، اندیس‌های توان، تعداد سطوح تجزیه در هر یک از مدل‌های ترکیبی می‌باشد. همچنین ردیف‌های پررنگ و زیرخط‌دار کاراترین مدل ترکیبی هم‌راه را نشان می‌دهد

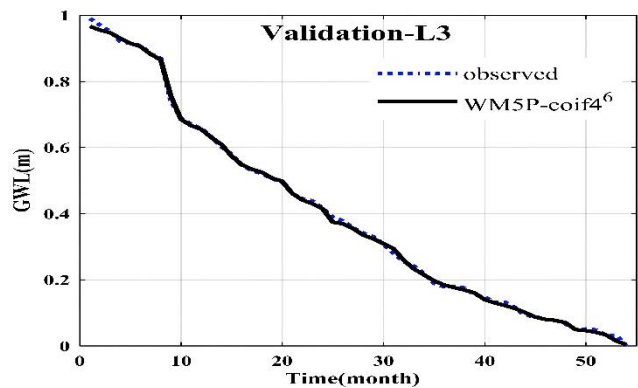
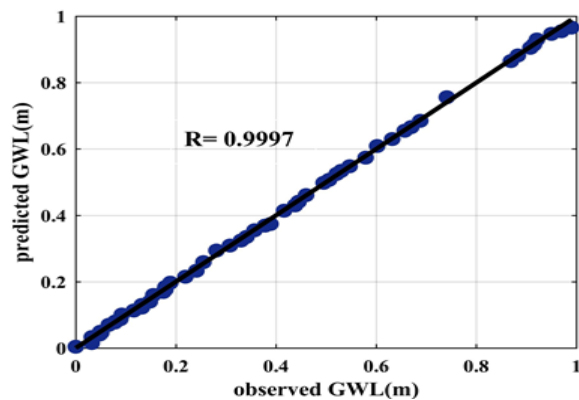
(الف)



(ب)



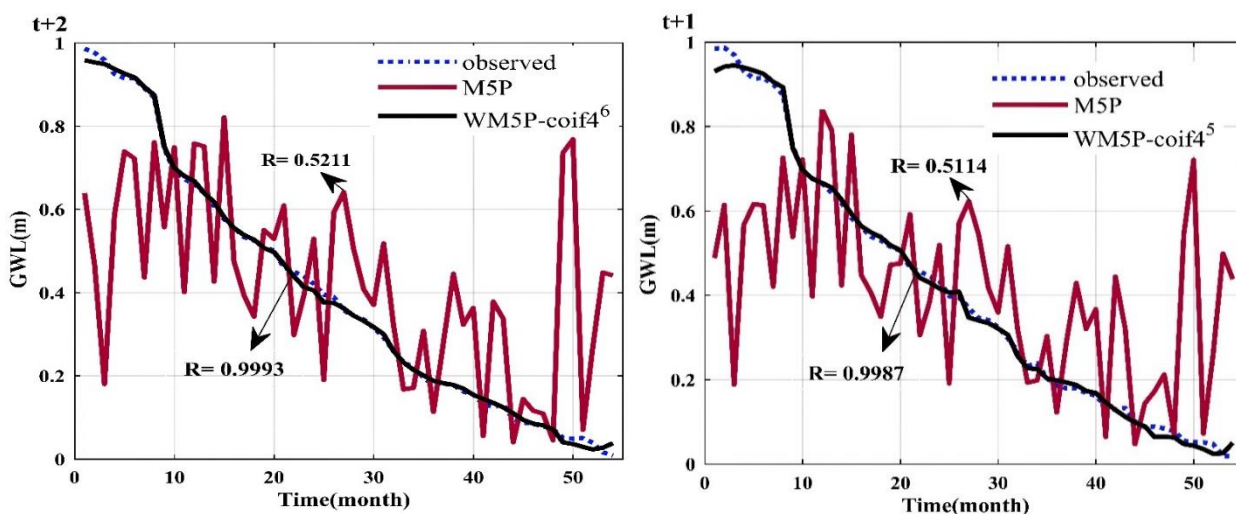
(ج)



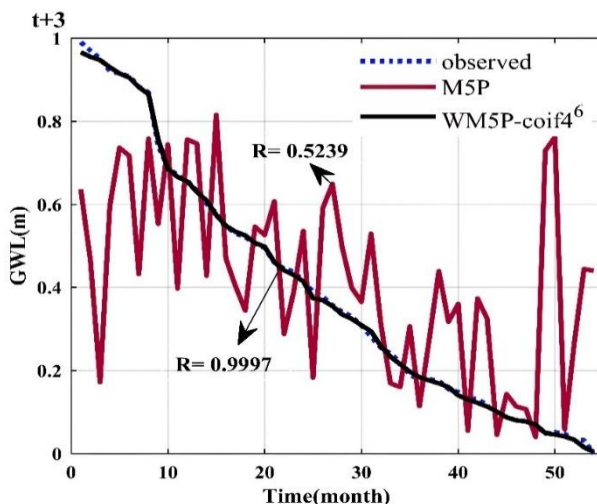
شکل ۶- سطح آب زیرزمینی پیش‌بینی شده با مدل ترکیبی برتر در یک، دو و سه ماه آینده (به ترتیب الف-ج) همراه با نمودارهای پراکندگی (داده‌های آزمایش)

ترکیبی در هر ماه با مقادیر واقعی برای مجموعه داده‌های اعتبارسنجی در شکل‌های ۷ تا ۹ ارائه شده است. تمامی محورهای قائم در این شکل‌ها، مقادیر نرمال شده سطح آب زیرزمینی هستند.

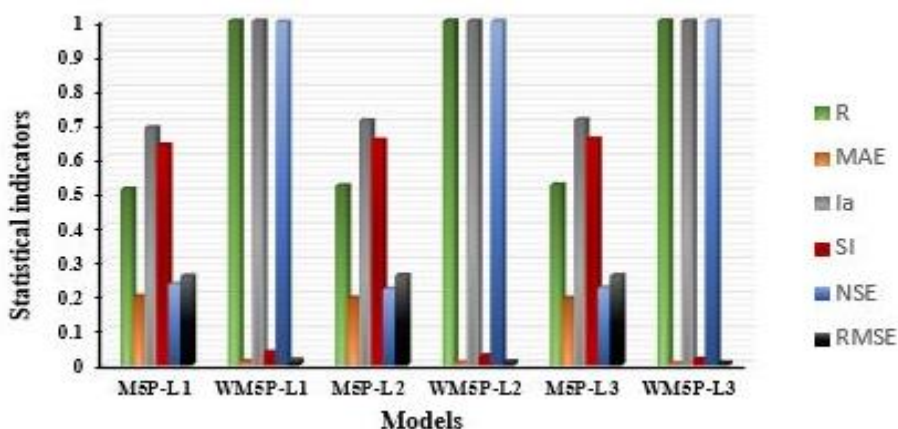
مقایسه نتایج مدل منفرد M5P و ترکیبی WM5P و تأثیر پیش‌پردازش بر نتایج پیش‌بینی برای روشن شدن توانایی مدل منفرد و ترکیبی در شبیه‌سازی، مقایسه مقادیر تخمین زده شده توسط مدل‌های منفرد و بهترین مدل



شکل ۷- عملکرد مدل‌ها در برآورد GWL ماه اول (داده‌های آزمایش) شکل ۸- عملکرد مدل‌ها در برآورد GWL ماه دوم (داده‌های آزمایش)



شکل ۹- عملکرد مدل‌ها در برآورد GWL ماه سوم (داده‌های آزمایش)



شکل ۱۰- شاخص‌های ارزیابی عملکرد مدل منفرد و مدل ترکیبی WM5P برتر در هر ماه

مقایسه مدل‌های منفرد و ترکیبی استفاده شد. مشاهده شد که کاربرد تبدیل موجک در پیش‌پردازش داده‌ها، تا حد زیادی توانایی مدل را افزایش داده و در زمان‌های پیش‌بینی طولانی‌تر خطا و عدم قطعیت‌ها کاهش یافته و مقادیر پیش‌بینی شده به مقادیر واقعی بسیار نزدیک شده و در نهایت عملکرد مدل بهبود پیدا کرده است. نتایج حاصل از تحلیل و مقایسه میان رویکرد ترکیبی موجک-درخت تصمیم (M5P)(WM5P) با M5P نشان داد که مدل WM5P ایزاری قدرتمندتر در برآورد سطح آب‌زیرزمینی در این منطقه است. همچنین صرفاً استفاده از پارامترهای هواشناسی به‌عنوان ورودی مدل با فرض عدم وجود اطلاعات تاریخی سطح آب‌زیرزمینی در محدوده مطالعاتی که وجه تمایز این تحقیق با تحقیقات دیگر در این زمینه می‌باشد، مدلی موفقیت آمیز و رضایت‌بخش تهیه شد که قادر است با قابلیت اطمینان بالا و دقتی مطلوب به پیش‌بینی سطح آب‌زیرزمینی در یک تا سه ماه آتی بپردازد. تأثیر انتخاب نوع توابع موجک و تطابق بصری آن با سری مادر و انتخاب سطح تجزیه مناسب در بهینه‌سازی مدل‌های ترکیبی و دستیابی به بهترین نتایج ثابت گردید. موجک Coiflet مرتبه چهار (Coif4) با ۵ سطح تجزیه با مقادیر شاخص‌های $MAE=0/012$ و $NSE=0/997$ ، موجک Coiflet مرتبه چهار با ۶ سطح تجزیه دارای $MAE=0/0084$ و $NSE=0/9986$ و موجک Coiflet مرتبه چهار با ۶ سطح تجزیه دارای $MAE=0/0054$ و $NSE=0/9994$ به ترتیب مدل‌های ترکیبی بهینه در ماه اول، دوم و سوم می‌باشند و با استناد به این نتایج می‌توان گفت، همیشه سطح بالای تجزیه برای بهینه‌سازی مدل‌ها مفید نمی‌باشد و باید مدلی با موجک مادر و سطح تجزیه کارآمد شناسایی و انتخاب گردد. همچنین توصیه می‌شود در تحقیقات آینده، توانایی رویکردهای پیش‌پردازشگر دیگر همچون الگوریتم تجزیه حالت تجربی (EMD) و سناریوهای ورودی با تأخیرهای زمانی مختلف در تخمین سطح آب‌زیرزمینی بررسی گردد.

در شکل ۱۰ مقادیر معیارهای ارزیابی عملکرد برای مدل منفرد و ترکیبی برگزیده در هر کدام از ماه‌های پیش‌بینی، نمایش داده شده است. تغییرات قابل ملاحظه در ارتقاء شاخص‌ها، حاکی از تأثیر تجزیه سری‌های زمانی بر کارایی مدل پیش‌بینی و موفقیت در اعمال تکنیک تبدیل موجک می‌باشد. از طرفی، با افزایش زمان پیش‌بینی، عملکرد مدل در هر ماه نسبت به ماه قبل از خود بهبود یافته است. بر اساس یافته‌های مدل‌سازی، مدل WM5P، به‌علت پیش‌پردازش سری پارامترهای ورودی به مدل با قابلیت اعتماد بالاتری در قیاس با اجرای مدل بدون عملیات پیش‌پردازش سری‌های زمانی ورودی به تخمین سطح آب‌زیرزمینی پرداخته است. برای مثال، با توجه به ارزیابی عملکرد موجک $coif4$ با $Le=5$ در ماه اول برای مجموعه داده‌های اعتبارسنجی، مشخص شد که مدل ترکیبی یاد شده دارای مقادیر $R=0/9987$ ، $I_a=0/999$ و $NSE=0/997$ بالاتر نسبت به مدل منفرد با $R=0/5114$ ، $I_a=0/6898$ و $NSE=0/2361$ و همچنین شاخص‌های $SI=0/038$ ، $MAE=0/012$ و $RMSE=0/016$ کوچکتر از همتای منفرد با $MAE=0/1999$ و $RMSE=0/2597$ ، $SI=0/6394$ بوده است. مطابق شکل‌های ۹ تا ۱۱، با مقایسه بصری قدرت پیش‌بینی هر کدام از رویکردها، مدل‌های ترکیبی منتخب در مقایسه با مدل منفرد منجر به تناسب بهتر با مقادیر مشاهداتی شده است.

نتیجه‌گیری

به‌منظور شبیه‌سازی سطح آب‌زیرزمینی و برآورد وضعیت آتی نوسانات این منابع ارزشمند در دشت کرمان-باغین، از مدل منفرد درخت تصمیم M5P استفاده شد. همچنین با بهره‌مندی از عمل پیش‌پردازشگر تبدیل موجک و تعیین بهترین موجک‌های مادر و سطح تجزیه، سری‌های زمانی ورودی به زیرسری‌هایی تجزیه شد. در پایان شاخص‌های ارزیابی برای بررسی صحت و اعتبار مدل‌ها و

پی‌نوشت‌ها

- 1- Groundwater Level
- 2- Artificial Intelligence
- 3- Machine Learning
- 4- Artificial Neural Network
- 5- Adaptive Neuro-Fuzzy Inference Systems
- 6- Support Vector Machine
- 7- Support Vector Regression
- 8- Bayesian Network
- 9- Reyhanli
- 10- M5P-Tree
- 11- Wavelet Transform
- 12- Multivariate Adaptive Regression Splines
- 13- Model Trees
- 32- Evapo-Transpiration
- 33- River Discharge
- 34- Average Wind Speed
- 35- Quinlan
- 36- Standard Deviation Reduction
- 37- Correlation Coefficient
- 38- Root Mean Square Error
- 39- Mean Absolute Error
- 40- Nash-Sutcliffe model efficiency coefficient
- 41- Index of Agreement
- 42- Scatter Index
- 43- Fernando and Kerr

- 14- Wavelet- Gene Expression Programming
- 15- Coiflet
- 16- Symlet 17- Daubechies
- 18- Group method of data handling
- 19- Least square support vector machine
- 20- Fuzzy logic
- 21- Genetic programming
- 22- Random tree

منابع

- Abdolhazadeh, M. and Schmalz B. 2022. Assessment of wavelet-SVR and wavelet-GP models in predicting the groundwater level using areal precipitation and consumption data. *Hydrological Sciences Journal*. 67(7): 1026-1039.
- Ardabili, S., Mosavi, A. and Várkonyi-Kóczy AR. 2020. Advances in machine learning modeling reviewing hybrid and ensemble methods. In *International Conference on Global Research and Education*. Springer, Cham: 215-227.
- Azarpira, F. and Shahabi, S. 2021. Evaluating the capability of hybrid data-driven approaches to forecast monthly streamflow using hydrometric and meteorological variables. *Journal of Hydroinformatics*. 23(6): 1165-1181.
- Bahmani, R. and Ouarda, T. B. 2021. Groundwater level modeling with hybrid artificial intelligence techniques. *Journal of Hydrology*. 595: 125659.
- Bahmani, R., Solgi, A. and Ouarda, T. B. 2020. Groundwater level simulation using gene expression programming and M5 model tree combined with wavelet transform. *Hydrological Sciences Journal*. 65(8): 1430-1442.
- Daubechies, I. 1990. The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*. 36(5):961-1005.
- Grossmann, A. and Morlet, J. 1984. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*. 15(4): 723-736.
- Huang M. and Tian, Y. 2015. Prediction of groundwater level for sustainable water management in an arid basin using data-driven models. *International Conference on Sustainable Energy and Environmental Engineering 2015 Oct*. Atlantis Press: 134-137.
- Javadinejad, S., Dara, R. and Jafary F. 2020. Modelling groundwater level fluctuation in an Indian coastal aquifer. *Water SA*. 46(4): 665-671.
- Lallahem, S., Mania, J., Hani, A. and Najjar, Y. 2005. On the use of neural networks to evaluate groundwater levels in fractured media. *Journal of hydrology*. 307(1-4): 92-111.
- Milan, S. G., Roozbahani, A. and Banihabib, M. E. 2018. Fuzzy optimization model and fuzzy inference system for conjunctive use of surface and groundwater resources. *Journal of hydrology*. 566: 421-434.
- Kardan Moghaddam, H., Ghordoyee, Milan, S., Kayhomayoon, Z. and Arya Azar, N. 2021. The prediction of aquifer groundwater level based on spatial clustering approach using machine learning. *Environmental Monitoring and Assessment*. 193(4): 1-20.
- Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.-M. 1996. *Wavelet toolbox*. The MathWorks Inc., Natick, MA. 15: 21.
- Nalarajan, N. A. and Mohandas, C. 2015. Groundwater level prediction using M5 model trees. *Journal of the Institution of Engineers (India): Series A*. 96(1): 57-62.
- Nourani, V., Baghana, A. H., Adamowski, J. and Kisi, O. 2014. Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *Journal of Hydrology*. 514: 358-377.
- Nourani, V., Komasi, M. and Mano, A. 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water resources management*. 23(14): 2877-2894.
- Pham, Q. B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A. R. M. T., Talukdar, S., Nguyen, X. C., Ahmed, A.N. and Anh, D. T. 2022. Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Computing and Applications*. 34(13): 10751-10773.
- Quinlan, J. R. 1992. *Learning with continuous classes*. 5th Australian joint conference on artificial intelligence, World Scientific.
- Rajaei, T., Ebrahimi, H. and Nourani, V. 2019. A review of the artificial intelligence methods in groundwater level modeling. *Journal of hydrology*. 572: 336-351.
- Rakhshandehroo, G. R., Vaghefi, M. and Aghbolaghi, M. A. 2012. Forecasting groundwater level in Shiraz plain using artificial neural networks. *Arabian Journal for Science and Engineering*. 37(7): 1871-1883.

- Rezaie-balf, M., Naganna, S. R., Ghaemi, A. and Deka, P. C. 2017. Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *Journal of hydrology*. 553: 356-373.
- Samani, S., Vadiati, M., Nejatijahromi, Z., Etebari, B., and Kisi, O. 2022. Groundwater level response identification by hybrid wavelet-machine learning conjunction models using meteorological data. *Environmental Science and Pollution Research*, 1-22.
- Sattari, M. T., Mirabbasi, R., Sushab, R. S. and Abraham, J. 2018. Prediction of groundwater level in Ardebil plain using support vector regression and M5 tree model. *Groundwater*. 56(4): 636-646.
- Sharafati, A., Asadollah, S. B. H. S. and Neshat, A. 2020. A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. *Journal of Hydrology*. 591: 125468.
- Shariat, R., Roozbahani, A. and Ebrahimian, A. 2019. Risk analysis of urban stormwater infrastructure systems using fuzzy spatial multi-criteria decision making. *Science of the Total Environment*. 647: 1468-1477.
- Shahabi, S., Khanjani, M. J. and Hessami Kermani, M. 2016. Hybrid wavelet-GMDH model to forecast significant wave height. *Water Science and Technology: Water Supply*. 16(2): 453-459.
- Sreekanth, P., Geethanjali, N., Sreedevi, P., Ahmed, S., Kumar, N. R. and Jayanthi, P. K. 2009. Forecasting groundwater level using artificial neural networks. *Current science*: 933-939.
- Solgi, A., Pourhaghi, A., Bahmani, R. and Zarei, H. 2017. Pre-processing data using wavelet transform and PCA based on support vector regression and gene expression programming for river flow simulation. *Journal of Earth System Science*. 126(5): 1-17.
- Solomatine, D. P. and Xue, Y. 2004. M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China. *Journal of Hydrologic Engineering*. 9(6): 491-501.
- Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Salim, H., Sungwon, K., Sulaiman, SO., Tan, M.L., Sa'adi, Z., Mehr, AD., Allawi, MF. 2022. Groundwater level prediction using machine learning models: a comprehensive review. *Neurocomputing*.
- Üneş, F., Maruf, A. G. and Taşar, B. 2019. Ground Water Level Estimation for Dörtöyl region in HATAY. *International Journal of Environment, Agriculture and Biotechnology*. 4(3): 859-864.
- Vadiati, M., Rajabi Yami, Z., Eskandari, E., Nakhaei, M., and Kisi, O. 2022. Application of artificial intelligence models for prediction of groundwater level fluctuations: Case study (Tehran-Karaj alluvial aquifer). *Environmental Monitoring and Assessment*, 194(9), 619.
- Wang, Y. and Witten, I. H. 1996. Induction of model trees for predicting continuous classes. *Computer Science Working Papers*.
- Wei, A., Chen, Y., Li, D., Zhang, X., Wu, T., and Li, H. 2022. Prediction of groundwater level using the hybrid model combining wavelet transform and machine learning algorithms. *Earth Science Informatics*. 15(3): 1951-1962.
- Yue, Q., Zhang, F. and Guo, P. 2018. Optimization-based agricultural water-saving potential analysis in Minqin County, Gansu Province China. *Water*. 10(9): 1125.
- Zyoud, S. H., Kaufmann, L. G., Shaheen, H., Samhan, S. and Fuchs-Hanusch, D. 2016. A framework for water loss management in developing countries under fuzzy environment: Integration of Fuzzy AHP with Fuzzy TOPSIS. *Expert Systems with Applications*. 61: 86-105.

Evaluating Decision Tree Efficiency in Combination with Wavelet Transform to Predict Groundwater Level Fluctuation

H. Rostaminezhad Dolatabad¹, S. Shahabi^{2*}, M.R. Madadi³

Received: Dec.24, 2022

Accepted: Feb.10, 2023

Abstract

Accurate monitoring of groundwater levels and estimating their fluctuations in the future is of importance, especially in arid and semi-arid areas. Regarding the high capabilities of AI-based models in the modeling of hydrologic phenomena, this research used the MP5 decision tree, in combination with the wavelet transform, to predict groundwater level fluctuations of the Kerman-Bagheyn plane. To develop the wavelet-decision tree (W-M5p) hybrid model, the wavelet transformation outputs were exported to the MP5 as inputs. Several statistical criteria, including coefficient of correlation (R), agreement index (I_a), and scattering index (SI), were used to evaluate the performance of the hybrid model compared to the single model. The results indicated that, even when the inputs of the hybrid model includes only the meteorological data from a synoptic station (the water level of previous periods were not used in the analysis), that the performance of the WM5P was superior to the single model in the prediction of groundwater fluctuations. The WM5P model with three months of forecast horizon with the Coif4 wavelet and decomposition level of 6 reduced the SI value from 0.6394 to 0.0181 and, at the same time, increased the I_a from 0.6898 to 0.9998. Consequently, the Coiflet4 with decomposition levels of 5 and 6 was the most efficient wavelet in the hybrid model for reliable estimation of the Kerman-Bagheyn plane groundwater level.

Keywords: Artificial intelligence, Data preprocessing, Machine learning, Modeling, Water table

1- MSc Student, Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman, Iran

2- Assistant Professor, Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman, Iran

3- Associate Professor, Department of Water Engineering, Faculty of Agriculture, University of Jiroft, Jiroft, Iran

(*-Corresponding Author Email: s.shahabi@kgut.ac.ir)