

کاربرد مدل داده‌های ترکیبی در برآورد غلظت کلر آبخوان دزفول

غزاله هادی‌قنات^۱، علی شهیدی^۲، عباس خاشعی سیوکی^۲، سید رضا هاشمی^۲

۱- دانشجوی کارشناسی ارشد گروه مهندسی آب، دانشگاه بیرجند
(نویسنده مسئول) ۳۳۲۲۵۸۵ (۰۶۱۱) hadighanavat@gmail.com
۲- استادیار گروه مهندسی آب، دانشگاه بیرجند

(دریافت ۹۲/۷/۵ پذیرش ۹۳/۲/۲۸)

چکیده

آب‌های زیرزمینی در مناطق خشک و نیمه‌خشک مانند ایران به دلیل کم هزینه و در دسترس بودن اهمیت زیادی دارند. با توجه به کمبود مطالعات در بخش کیفیت منابع آب زیرزمینی در مقایسه با کمیت این منابع، این پژوهش با هدف پیش‌بینی تغییرات کلر آب زیرزمینی در دشت دزفول در استان خوزستان صورت گرفت. داده‌های ترکیبی مدلی رگرسیونی با در نظر گرفتن متغیرها در واحدهای مختلف و در طی زمان، امکان پیش‌بینی کیفیت آب را به‌طور توأم در چندین چاه فراهم می‌آورد. به این منظور، در محدوده مورد مطالعه، پارامترهای هواشناسی بارندگی و تبخیر و تعرق پتانسیل و پارامترهای کیفی EC، سدیم، کلسیم و منیزیم برای تخمین کلر در ده چاه انتخابی به‌صورت فصلی در یک دوره هشت ساله جمع‌آوری شد. در مرحله بعد، انواع مدل‌های داده‌های ترکیبی شامل اثر مشترک، ثابت و تصادفی بر روی داده‌های موجود برازش داده شد. نتایج نشان داد که مدل داده‌های ترکیبی با اثرات تصادفی بهترین نتیجه را برای پیش‌بینی کیفیت (کلر) آب زیرزمینی داشته است. معیارهای عملکرد ($RMSE=2.445$ ، $R^2=0.96$) نیز بیانگر دقت مدل است.

واژه‌های کلیدی: داده‌های ترکیبی، کیفیت آب زیرزمینی، مدل‌سازی، دشت دزفول

۱- مقدمه

در دهه‌های اخیر، توسعه و رشد روز افزون صنعت و کشاورزی، افزایش جمعیت، وقوع خشکسالی و تأثیر عمده‌امان مطلوب این توسعه و رشد بر روی محیط‌زیست، باعث شده توجه فزاینده‌ای به منابع آب معطوف شود. با توجه به این که از میان این منابع، آب زیرزمینی در مناطق خشک و نیمه‌خشک نقش یک مخزن مطمئن آب شیرین را ایفا می‌نماید از اهمیت قابل توجهی برخوردار است. اما مدیریت و بهره‌برداری بهینه این منابع بدون شناخت آنها امکان‌پذیر نیست. اکثر مطالعات متعدد انجام شده بر روی منابع آب زیرزمینی بر شناخت کمیت آب تمرکز دارد و به کیفیت این منابع کمتر پرداخته شده است.

به‌طور معمول برای مدل‌سازی کیفیت آب زیرزمینی از روش‌های زمین‌آماری و هوش مصنوعی استفاده می‌شود. غفوری و همکاران به ارزیابی زمین‌آماری کیفیت آب زیرزمینی دشت داراب استان فارس پرداخته و به این نتیجه رسیده‌اند که برای تمام شاخص‌های کیفی مورد مطالعه آب زیرزمینی، روش کوکره‌جینگ بهتر از سایر روش‌ها، شبیه‌سازی را انجام می‌دهد [۱]. غلامی و

همکاران به بررسی روش رگرسیون چند متغیره و شبکه عصبی^۱ در شبیه‌سازی شوری آب زیرزمینی در سواحل استان مازندران پرداخته‌اند. نتایج بیانگر آن است که کارایی شبکه عصبی مصنوعی در برآورد شوری آب زیرزمینی بیشتر از روش رگرسیون چند متغیره است [۲]. قمیشیون و همکاران به بررسی تغییرات مکانی کیفیت آب زیرزمینی دشت سمنان-سرخه با استفاده از روش‌های زمین‌آماری پرداخته‌اند که روش کوکره‌جینگ برای عوامل کیفی دارای دقت بالاتری نسبت به دیگر روش‌های میان‌یابی است [۳]. محمدی و همکاران به بررسی تغییرات مکانی-زمانی سطح آب زیرزمینی دشت کرمان با استفاده از روش زمین‌آماری مناسب (طی دوره آماری ۱۰ ساله ۱۳۸۵-۱۳۷۵) پرداخته‌اند [۴]. احمددار و همکاران به تخمین میزان فلوراید با استفاده از شبکه عصبی مصنوعی و عکس فاصله با استفاده از پارامترهای کیفی معمول در حوضه موماندیار هند پرداخته‌اند. نتایج پژوهش پایین بودن همبستگی همه پارامترهای مورد نظر را نشان داده به‌طوری که بالاترین همبستگی تنها ۰/۳۷ بوده است. علاوه بر این، پارامتر pH،

¹ Artificial Neural Network (ANN)

کلرید، سولفات و کلسیم مؤثرتر از هشت پارامتر دیگر بر فلوراید است. کلرید پارامتری است که حداکثر همبستگی و حساسیت را با فلوراید به خود اختصاص می‌دهد [۵].

آل محالوی و همکاران کاربرد شبکه عصبی مصنوعی را در پیش‌بینی نیترات آب زیرزمینی در مناطق روستایی و کشاورزی مورد بررسی قرار داده‌اند. نتایج بیانگر آن بوده که بهترین شبکه برای تخمین نیترات آب زیرزمینی، MLP^۱ است [۶]. بانرجی و همکاران مدل شبکه عصبی را به‌عنوان جایگزین بالقوه برای پیش‌بینی شوری آب زیرزمینی بررسی نموده‌اند. نتایج بیانگر آن است که شبکه عصبی پیشنهادی به‌عنوان جایگزین ساده‌تر و دقیق‌تر نسبت به روش‌های عددی ظاهر شده است [۷]. فرهنگ و همکاران به مقایسه روش کریجینگ، شبکه عصبی و ANFIS برای مدل‌سازی توزیع زمانی و مکانی آلودگی آب‌های زیرزمینی پرداخته‌اند، که مدل ANFIS در مقایسه با دو مدل دیگر در تخمین توزیع مکانی و زمانی کیفیت آب زیرزمینی توانمندتر عمل نموده است [۸]. چو و همکاران به پیش‌بینی پتانسیل آلودگی آرسنیک آب زیرزمینی با استفاده از شبکه عصبی مصنوعی در کامبودیا و تایلند پرداخته‌اند. نتایج گویای آن است که مدل PC-ANN که یکی از چهار مدل مورد بررسی در این تحقیق است، با بهترین دقت میزان آرسنیک را پیش‌بینی می‌کند [۹]. زارع ایبانه و همکاران به ارزیابی خوردگی آب‌های زیرزمینی در استان همدان با استفاده از ANFIS پرداخته‌اند. نتایج بیانگر توانمندی مدل ANFIS در تعیین پتانسیل خوردگی آب زیرزمینی در منطقه مورد مطالعه است [۱۰]. زارع ایبانه و همکاران به تخمین کلرید در آب زیرزمینی با اندازه‌گیری هدایت الکتریکی پرداخته‌اند. نتایج نشان داده که سه مدل رگرسیون خطی، چند جمله‌ای و نمایی برای داده‌های ۳۳۱۵ نمونه معنی‌دار و ۹۹ درصد میزان کلرید را تبیین می‌نماید، به‌علاوه دیگر مدل‌های مورد استفاده برای مجموعه داده‌های دیگر، رابطه خوبی میان کلرید و هدایت الکتریکی، به جز در مواردی که هدایت الکتریکی کمتر از 3 dS.m^{-1} است، را نشان می‌دهد [۱۱].

در این میان داده‌های ترکیبی به‌دلیل در نظر گرفتن واحدهای مقطعی در طی زمان، توانایی زیادی در برآورد روابط بین متغیرهای مؤثر بر کیفیت آب زیرزمینی چندین چاه توام را داراست. در این روش علاوه بر در نظر گرفتن مکان، زمان نیز در مدل‌سازی در نظر گرفته می‌شود. از مزایای این مدل می‌توان به افزایش حجم نمونه و افزایش کارایی اشاره نمود.

بررسی منابع نشان می‌دهد که مدل داده‌های ترکیبی^۲ به‌طور عمده در علم اقتصاد کاربرد دارد. بر این اساس صالح‌نیا و فلاحی

به بررسی تأثیر عوامل اقلیمی و اقتصادی بر عملکرد گندم آبی پرداخته‌اند. نتایج گویای آن است که در بین متغیرهای مورد بررسی، عملکرد گندم نسبت به درجه روز سرمای حساسیت بیشتری نشان می‌دهد و الگوی اثرات تصادفی یک جانبه برای تخمین تابع عملکرد گندم مناسب است [۱۲]. علیجانی و همکاران اثر درجه حرارت و بارندگی بر عملکرد گندم آبی ایران را بررسی نموده‌اند. نتایج نشان می‌دهد که تغییرات عملکرد گندم استان‌ها نسبت به بارندگی واکنش بیشتری نشان می‌دهد [۱۳]. مدل داده‌های ترکیبی در مدیریت منابع آب مدلی نوین است. از معهود تحقیقات انجام شده می‌توان به موارد زیر اشاره کرد: فلاحی و همکاران به ارزیابی عوامل مؤثر بر مصرف آب شرب خانوار و پیش‌بینی تقاضای آن با روش داده‌های تابلویی (ترکیبی) پرداخته‌اند. نتایج نشان می‌دهد که اگر قیمت آب برای سال‌های آینده افزایش نیابد، با افزایش درآمد، سرانه مصرف آب خانوار افزایش می‌یابد. اما در صورت افزایش تعرفه‌ها در سال‌های آینده، مصرف سرانه آب خانوار در شبانه روز کاهش خواهد یافت [۱۴]. آربوش و باربران تأثیر قیمت بر نیاز آب شهری با داده‌های ترکیبی پویا را بررسی نموده‌اند. نتایج بیانگر آن است که قیمت، یک ابزار تأثیر محدود است [۱۵]. ایزدی و همکاران کاربرد مدل داده‌های ترکیبی در پیش‌بینی سطح آب زیرزمینی را بررسی کرده‌اند. نتایج پژوهش، کارایی بالای مدل و برتری آن نسبت به شبکه عصبی مصنوعی را نشان می‌دهد [۱۶].

با بررسی تحقیقات انجام شده مشخص می‌شود اکثر مدل‌های مورد استفاده در شبیه‌سازی کیفیت آب زیرزمینی تنها بر پایه مکان و یا زمان عمل می‌نمایند اما از آن جا که کیفیت آب زیرزمینی در دو بعد زمان و مکان در حرکت است و پارامترهای آن در زمان و مکان ثابت نیست، پس مدل‌سازی بر اساس تنها یکی از این دو بعد، روند درستی به نظر نمی‌رسد، در نتیجه یا مدل‌های پیشین باید تغییر یابد تا به‌صورت زمانی و مکانی مدل‌سازی نمایند و یا از مدل‌های نوین برای مدل‌سازی پارامترهای کیفی آب زیرزمینی استفاده شود. در این راستا، مدل داده‌های ترکیبی را می‌توان به‌عنوان یک مدل با رویکرد زمانی و مکانی معرفی نمود. هدف از این تحقیق نشان دادن توانمندی این روش در مدل‌سازی و تخمین پارامترهای کیفیت آب زیرزمینی بود.

۲- مواد و روش‌ها

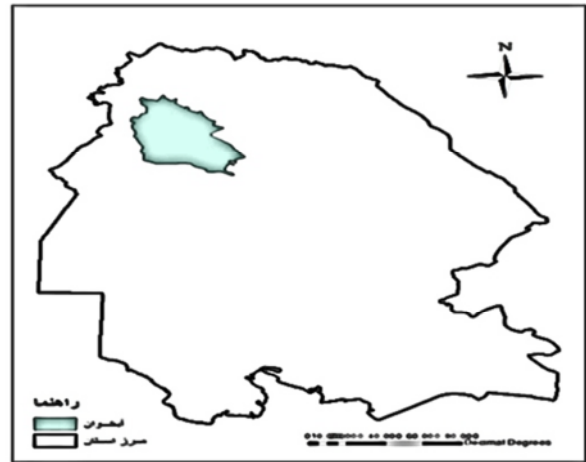
۲-۱- منطقه مورد مطالعه

دشت دزفول- اندیمشک در شمال جلگه خوزستان و در 150 کیلومتری شمال اهواز و در طول‌های جغرافیایی 45° تا 47° تا 48° درجه شرقی و عرض‌های 32° تا 33° شمالی قرار گرفته است.

¹ Multilayer Perceptrons

² Panel data

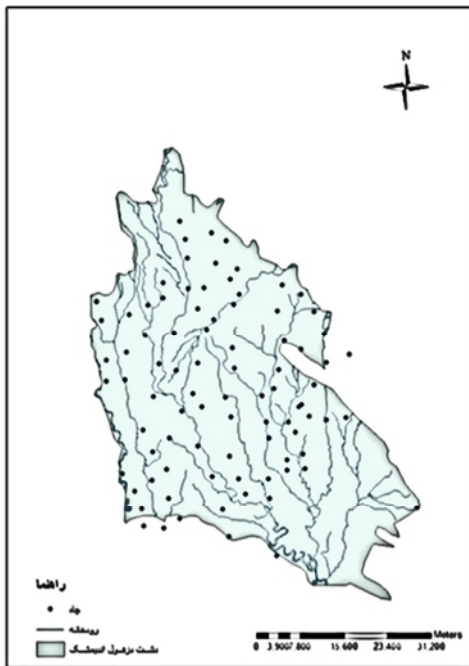
جایگاه دشت دزفول اندیمشک در استان خوزستان در شکل ۱ نشان داده شده است.



شکل ۱- جایگاه دشت دزفول- اندیمشک در استان خوزستان

هواشناسی استان خوزستان دریافت شد. پس از بررسی داده‌های مربوط به چاه‌ها، ۸۷ چاه از منابع انتخابی دارای دوره آماری کامل تر تشخیص داده شد. جایگاه چاه‌ها در دشت دزفول اندیمشک در شکل ۲ آورده شده است.

داده‌های هواشناسی نیز باید برای تطبیق با داده‌های کیفی، با روش میانگین‌گیری به داده‌های فصلی تبدیل شوند. پس از آن با استفاده از روش عکس فاصله و داده‌های مربوط به سه ایستگاه موجود در منطقه، داده‌های هواشناسی برای چاه‌های منتخب که وارد مدل می‌شوند، به دست می‌آیند. در جدول ۱ مشخصات ایستگاه‌های هواشناسی مورد استفاده درج شده است.



شکل ۲- جایگاه چاه‌های موجود در دشت دزفول اندیمشک

بلندترین نقطه دشت در شمال منطقه حدود ۱۵۰ متر و کمترین آن در قسمت جنوب دشت حدود ۶۰ متر نسبت به سطح دریا ارتفاع دارد. شیب عمومی دشت از شمال به سمت جنوب بوده و برآمدگی‌هایی در حوالی هفت تپه دشت را از قسمت جنوبی جلگه خوزستان جدا می‌سازد.

زمین‌شناسی حوضه شامل سازندهای آجاجاری و بخش لهبری و کنگلومرای بختیاری است. به علت حضور کنگلومرای بختیاری در قسمت بالا و تخریب و فرسایش و حمل مواد این سازند، قسمت اعظم دشت شامل مواد درشت‌دانه و قلوهای بوده که به تدریج در قسمت‌های جنوبی به رسوبات دانه‌ریز تبدیل می‌شوند. سازند بختیاری به دلیل نفوذپذیری و تخلخل کافی، مهم‌ترین سازند به لحاظ تغذیه آبخوان دشت است.

حوضه دزفول- اندیمشک در اقلیم گرم و خشک قرار گرفته است. میزان متوسط بارندگی در سطح حوضه ۳۱۰ میلی‌متر در سال و متوسط تبخیر ۲۱۹ میلی‌متر در سال است.

۲-۲- جمع‌آوری و تجزیه و تحلیل داده‌ها

متغیرهای ورودی برای تعیین کلر با استفاده از این مدل رگرسیونی، شامل پارامترهای هواشناسی بارندگی، تبخیر و پارامتر کیفی EC به عنوان معیار شوری است. بارندگی شاخص تغذیه آب زیرزمینی و تبخیر به عنوان نمایی از تبخیر از سطح دشت در نظر گرفته می‌شود. پارامترهای کیفی به صورت فصلی در یک دوره هشت ساله (۱۳۸۹-۱۳۸۲) و برای ۱۰۵ چاه انتخابی از سازمان آب و برق خوزستان و پارامترهای هواشناسی به صورت ماهانه از سه ایستگاه شوشتر، حسینیه و دزفول در دوره هشت ساله مذکور از سازمان

جدول ۱- مشخصات ایستگاه‌های هواشناسی مورد استفاده

ردیف	نام ایستگاه	طول جغرافیایی شرقی	عرض جغرافیایی شمالی	ارتفاع (متر)
۱	شوشتر	۴۸-۵۰	۳۲-۰۳	۶۷
۲	دزفول	۴۸-۲۳	۳۲-۲۴	۱۴۳
۳	حسینیه	۴۸-۱۶	۳۲-۴۰	۳۵۴

روش تحقیق به این صورت است که با استفاده از تحلیل خوشه‌ای، چاه‌های موجود در منطقه خوشه‌بندی می‌شود. خوشه‌بندی به این دلیل است که استفاده از همه چاه‌ها وقت‌گیر بوده و می‌توان با خوشه‌بندی رفتارهای یکسان یک مجموعه چاه را در قالب یک

چاه تحلیل نمود. در خوشه‌بندی گروه‌هایی با حداکثر تشابه درون گروهی و حداکثر تفاوت بین گروهی تعیین می‌شوند که مناطق همگن را بر اساس پارامترهای تعیین شده، به دست می‌آورد. خوشه‌بندی به روش وارد^۱ که یکی از پرکاربردترین روش‌های خوشه‌بندی در هیدروژئولوژی است با استفاده از نرم‌افزار SPSS19 انجام می‌شود. پارامترهای مورد استفاده برای خوشه‌بندی طول و عرض جغرافیایی، میانگین ۵ ساله (۱۳۹۰-۱۳۸۶) یون‌های سولفات، بی‌کربنات، سدیم، منیزیم، کلسیم و کلر است.

قبل از برآورد مدل، باید مانایی^۲ متغیرهای مدل بررسی شود. به این دلیل که اگر متغیرهای مورد استفاده در برآورد ضرایب مدل ناماننا باشند، در عین حال که ممکن است هیچ رابطه مفهومی بین متغیرهای مدل وجود نداشته باشد، ضریب تبیین به دست آمده ممکن است به‌طور کاذبی بالا باشد. برای بررسی مانایی می‌توان از آزمون‌های مانایی و یا ریشه واحد^۳ استفاده کرد. آزمون‌های لوین، لیو و چو (LLC)^۴، ایم، پسران و شین (IPS)^۵، فیشر-دیکی فولر تعمیم یافته (ADF)^۶ و فیشر-فلپش پرون (PP)^۷ از مهم‌ترین آزمون‌های ریشه واحد داده‌های ترکیبی هستند که با نرم‌افزار Eviews7 قابل بررسی هستند.

در صورتی که متغیرها ناماننا باشند، باید از هم‌جمعی (هم‌انباشتگی)^۸ متغیرهای مدل اطمینان حاصل شود. مفهوم هم‌جمعی (هم‌انباشتگی) این است که اگر یک سری زمانی نظیر Y به صورت I(d)^۹ و X نیز به صورت I(d) (مقدار d در هر دو متغیر یکسان است) باشد، دو سری می‌توانند هم‌انباشته باشند. لازم به توضیح است که اگر از سری زمانی d مرتبه تفاضل گرفته شود و این سری تفاضل گرفته شده مانا باشد، آنگاه سری زمانی اصلی انباشته از مرتبه I(d) است [۱۷]. در چنین مواردی رگرسیون بر روی مقادیر دو متغیر معنی‌دار است یعنی رگرسیون دیگر ساختگی نیست و هیچ‌گونه اطلاعات بلندمدتی از دست نمی‌رود. این موضوع برخلاف نتیجه حاصل از کاربرد تفاضل‌های مرتبه اول که اطلاعات بلندمدت را از دست می‌داند، است. در واقع هم‌انباشتگی به این معناست که علی‌رغم اینکه سری‌ها به تنهایی غیرمانا هستند، ولی ترکیب خطی از دو یا چند سری زمانی (غیر مانا) می‌تواند مانا باشد. از آزمون‌های هم‌انباشتگی داده‌های ترکیبی می‌توان به آزمون

کائو^{۱۰} و آزمون پدرونی^{۱۱} اشاره نمود [۱۸].

پس از تعیین خوشه‌ها و بررسی مانایی و هم‌انباشتگی با استفاده از نماینده هر خوشه مدل‌های مختلف داده‌های ترکیبی برآزش داده می‌شود. در مرحله بعد با استفاده از آزمون‌های چو^{۱۲}، بروش-پاگان^{۱۳} و هاسمن^{۱۴} بهترین مدل داده‌های ترکیبی انتخاب می‌شود. در نهایت برای بررسی عملکرد مدل از معیارهای عملکرد متفاوتی همچون ضریب تبیین (R²) و جذر متوسط مربع خطا (RMSE) استفاده می‌شود.

۲-۳- تئوری مدل

داده‌های آماری به سه دسته داده‌های زمانی، مکانی و ترکیبی تقسیم می‌شوند. در داده‌های ترکیبی، متغیرها هم در میان مقاطع و هم در طول زمان اندازه‌گیری می‌شوند. از جمله مزایای این روش نسبت به مدل‌هایی که تنها مقاطع و یا زمان را می‌بینند می‌توان به موارد زیر اشاره کرد [۱۹]:

- تعداد مشاهدات و داده‌ها بیشتر بوده و باعث می‌شود اعتماد به برآوردها بیشتر شود.

- به محققان تجربی اجازه می‌دهد مدل‌های پیشرفته‌تری را تبیین کرده و آزمون کنند، که فرضیه‌های مقیدکننده کمتری در برداشته باشد.

- زیاد بودن تعداد مشاهدات، مسئله همخطی را تا حدود بسیاری حل می‌کند.

- با این مجموعه داده‌ها می‌توان اثراتی را شناسایی و اندازه‌گیری کرد که در داده‌های مقطعی محض و یا سری زمانی خالص قابل شناسایی نیست.

- استفاده از این روش، اریب برآورد را از بین می‌برد یا کم می‌کند. فرم کلی معادله رگرسیونی داده‌های ترکیبی به صورت زیر است

$$Y_{it} = \alpha + \beta X_{it} + U_{it} \quad (1)$$

$$U_{it} = \mu_i + v_{it} \quad (2)$$

که در این رابطه

i مکان، t زمان، Y_{it} متغیر وابسته، α ثابت معادله، β ضرایب معادله، X_{it} متغیر مستقل، U_{it} جمله خطا، v_{it} خطای باقیمانده (مستقل از مکان و زمان تغییر می‌کند)، μ_i اثر فردی غیرقابل مشاهده (اثر مربوط به هر چاه است که با متغیر ورودی بیان

1 Ward

2 Stationary

3 Unit Root Test

4 Levin, Lin and Chu

5 Im, Pesaran and Shin

6 Augmented Dicky -fuller

7 Philips-Peron

8 Cointegration

9 Integrated of Order

10 Kao

11 Pedroni

12 Chow Test

13 The Breusch-Pagan LM Test

14 Hausman Test

نمی‌شود) است.

ماهیت و شرایط μ_i موجب تقسیم‌بندی داده‌های ترکیبی به سه روش می‌شود:

۱- روش اثر مشترک^۱: در صورتی که μ_i فقط شامل یک جمله ثابت باشد، مدل به صورت اثرات مشترک درمی‌آید. در بعضی از مراجع این مدل را جزو داده‌های ترکیبی به حساب نمی‌آورند و به عنوان مدل متفاوتی به نام داده‌های تلفیقی^۲ معرفی می‌نمایند. این مدل با روش رگرسیون حداقل مربعات عادی^۳ قابل برآورد است.

۲- روش اثر ثابت^۴: اگر μ_i قابل مشاهده نباشد اما با X_{it} همبستگی داشته باشد، به صورت یک عرض از مبدا ثابت برای هر مقطع در می‌آید. اصطلاح ثابت به آن معنا است که این عرض از مبدا در طی زمان تغییر نمی‌کند. مدل اثر ثابت وقتی مدل مناسبی است که توجه ما معطوف به N مکان در یک مجموعه از مکان‌ها باشد. به نوعی این مدل فقط برای مطالعه خصوصیات فردی مکان‌های مورد مطالعه، قابل کاربرد است و برای مکان‌های خارج از نمونه قابل کاربرد نیست. این مدل را با روش رگرسیون حداقل مربعات با متغیرهای مجازی^۵ برآورد می‌کنند.

۳- روش اثر تصادفی^۶: اگر μ_i با X_{it} همبسته نباشد و به صورت یک عامل تصادفی برای هر مقطع بروز کند، روش را اثر تصادفی گویند. مدل تصادفی در صورتی تصریح مناسبی خواهد بود که N مکان به طور تصادفی از یک جمعیت بزرگ استخراج شود. در واقع مدل تصادفی راهی است برای برخورد با این واقعیت که T مشاهده روی N مکان شبیه مشاهدات روی NT واحد مختلف نیست. این مدل با روش حداقل مربعات تعمیم یافته^۷ برآورد می‌شود.

ابتدا الگوی مورد نظر به سه روش بالا با استفاده از نرم‌افزار Eviews7 برآورد می‌شود. پس از آن باید این موضوع بررسی شود که کدامیک از سه مدل برای منطقه مطالعاتی مناسب‌تر است. برای تعیین روش مناسب از آزمون‌های چو، LM، بروش-پاگان، هاسمن استفاده می‌شود. آزمون چو برای انتخاب میان روش اثرات ثابت و مشترک استفاده می‌شود. فرض H_0 در این آزمون این است که کلیه واحدها دارای عرض از مبدا مشترک می‌باشند (روش اثرات مشترک). این فرضیه با توجه به آماره F مورد آزمون قرار می‌گیرد

$$F(n-1, nT-n-k) = \frac{(R_f^2 - R_c^2)/(n-1)}{(1 - R_f^2)/(nT-n-k)} \quad (3)$$

¹ Common Effect

² Pooled Data

³ Ordinary Least Square (OLS)

⁴ Fix Effect Model

⁵ Least Square Dummy Variables (LSDV)

⁶ Random Effect Model

⁷ Generalized Least Square (GLS)

که در این رابطه

n تعداد مقاطع، T تعداد مشاهدات سری زمانی، k تعداد متغیر مستقل، R_f^2 و R_c^2 به ترتیب ضریب تبیین روش اثرات ثابت و مشترک است. در صورتی که P -value مربوط به این آماره بیش از $0/05$ باشد، فرض H_0 رد نمی‌شود.

آزمون LM بوش-پاگان برای انتخاب بین دو روش اثرات مشترک و تصادفی با استفاده از باقیمانده مدل اثرات مشترک قابل محاسبه است

$$LM = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n \left(\sum_{t=1}^T e_{it} \right)^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right] \quad (4)$$

که در این رابطه

e باقیمانده مقادیر مشاهده شده و پیش‌بینی شده در مدل اثرات مشترک است. بقیه پارامترها مشابه رابطه ۳ است. این آماره دارای توزیع کی دو (χ^2) با درجه آزادی یک است. چنانچه P -value این آماره کمتر از $0/05$ باشد، فرض H_0 رد می‌شود که به این معناست که تفاوت بین مقاطع تصادفی است.

آماره هاسمن برای مدل‌هایی با k پارامتر دارای توزیع χ^2 با k (تعداد متغیرهای مستقل) درجه آزادی به روش ماتریسی محاسبه می‌شود و چنانچه آماره هاسمن محاسبه شده کمتر از آماره بحرانی باشد، به این مفهوم است که فرض H_0 رد نمی‌شود. یعنی این فرضیه که اثرات فردی با سایر متغیرهای توضیحی همبستگی دارد، رد نمی‌شود. در پایان برای بررسی عملکرد مدل‌های برازش داده شده از معیارهای عملکرد متفاوتی مثل ضریب تبیین^۸ و ریشه متوسط خطای مربعات^۹ استفاده می‌شود.

۳- نتایج و بحث

چاه‌های موجود در دشت با تحلیل خوشه‌ای وارد خوشه‌بندی شد. نمودار درختی حاصل از این خوشه‌بندی در سطح چاه‌ها به ۱۰ خوشه تقسیم شد. جایگاه نماینده هر خوشه در شکل ۳ نشان داده شده است.

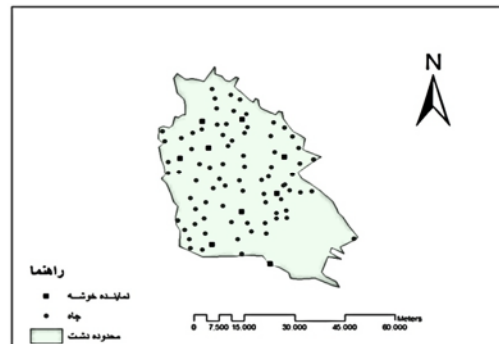
با استفاده از ۱۰ چاه منتخب در منطقه، مانایی متغیرهای موجود در مدل با آزمون‌های فیشر-دیکی فولر تعمیم یافته (ADF) سنجیده شد. نتایج این دو آزمون در جدول ۲ آورده شده است. با اتکا به این که مقادیر P -value مربوط به هر دو آزمون برای همه

⁸ Coefficient of Determination

⁹ Root Mean Square Error (RMSE)

متغیرها کمتر از ۰/۰۵ بود، فرض صفر مبنی بر وجود ریشه واحد رد شد و در نتیجه تمام متغیرها مانا بودند. با توجه به مانایی متغیرها دیگر نیازی به انجام آزمون هم‌جمعیتی نیست.

با تعیین نماینده هر خوشه، داده‌های هواشناسی برای هر چاه با روش عکس فاصله به دست آمد که در نتیجه ورودی‌ها مشخص شد.



شکل ۳- نمایی از نماینده خوشه‌ها

جدول ۲- آزمون مانایی متغیرها

متغیر	آزمون	P-value
بارندگی	دیکی - فولر تعمیم یافته	۰/۰۰۰۰
تبخیر	دیکی - فولر تعمیم یافته	۰/۰۰۰۰
کلرید	دیکی - فولر تعمیم یافته	۰/۰۰۰۰
EC	دیکی - فولر تعمیم یافته	۰/۰۰۰۰

با استفاده از آزمون دیکی - فولر تعمیم یافته، مانایی برای متغیرهای هواشناسی و کیفی مربوط به ده چاه انتخابی بررسی شد. نتایج این آزمون در جدول ۲ درج شده است.

با توجه به میزان P-value های موجود در جدول که همگی از میزان بحرانی ۰/۰۵ کمتر هستند، فرض صفر مبنی بر وجود ریشه واحد رد می‌شود و تمام پارامترهای ورودی در مدل مانا هستند. بر این اساس با توجه به اینکه پارامترهای مدل، ایستا و بر روی یک طول موج یکسان هستند، دیگر نیازی به انجام آزمون‌های هم‌جمعیتی نیست چون رگرسیون کاذبی حاصل نمی‌شود.

در مرحله بعد سه مدل اثر مشترک، ثابت و تصادفی بر روی متغیرها برای دوره آماری ۸۹-۸۲ برآزش شد. سپس باید مشخص شود که کدام مدل دارای اعتبار است. برای تعیین اعتبار از آزمون‌های هاسمن و چو استفاده شد. نتایج این دو آزمون در جدول ۳ آورده شده است. در این جدول P-value مربوط به آزمون چو از میزان بحرانی ۰/۰۵ کوچک‌تر شده که در نتیجه فرض صفر این آزمون مبنی بر اعتبار مدل اثر مشترک رد شده و از میان مدل‌های اثر ثابت و مشترک، مدل اثر ثابت معتبر شناخته شد. اما P-value

مربوط به آزمون هاسمن از میزان بحرانی ۰/۰۵ بزرگ‌تر است که در نتیجه فرض صفر این آزمون مبنی بر اعتبار مدل اثر تصادفی رد نشد. پس از بین مدل‌های اثر ثابت و تصادفی نیز مدل اثر تصادفی معتبر شناخته شد.

نتایج برازش مدل داده‌های ترکیبی با اثر تصادفی در جدول ۴ آورده شده است.

جدول ۳- اعتبارسنجی مدل‌های داده‌های ترکیبی

آزمون	P-value
چو	۰/۰۰۰۰
هاسمن	۰/۸۹۶۱

جدول ۴- ضرایب مدل اثر تصادفی

متغیر	ضرایب	P-value
ثابت معادله	-۱/۳۲۶۲	۰/۳۷۰۶
EC	۵/۷۰۳۱	۰/۰۰۰۰
بارندگی	۰/۰۰۴	۰/۶۱۶۳
تبخیر	-۰/۰۰۲۶	۰/۰۷۷۵

P-value موجود در جدول ۴ بیانگر معنی‌دار بودن ضرایب هر متغیر ورودی است. در صورتی که P-value بزرگ‌تر از ۰/۰۵ باشد در سطح معنی‌دار ۹۵ درصد فرض صفر مبنی بر عدم معنی‌دار بودن ضرایب رد نمی‌شود و ضریب مربوط به آن متغیر ورودی از نظر آماری معنی‌دار نمی‌شود. بنابراین در جدول ۴ مربوط به مدل اثر تصادفی تنها متغیر ورودی معنی‌دار EC است.

اکنون که مدل اثر تصادفی پذیرفته شد متغیر تبخیر از ورودی‌ها حذف و مجدداً این مدل بر روی داده‌ها برازش داده شد. ضرایب جدید در جدول ۵ آورده شده است.

جدول ۵- ضرایب مدل اثر تصادفی با دو ورودی

متغیر	ضرایب	P-value
ثابت معادله	-۲/۲۴۷۹	۰/۰۸۳۴
EC	۵/۷۲۲۶	۰/۰۰۰۰
بارندگی	۰/۰۱۴۹	۰/۰۰۳۶

برای تعیین میزان اثر بارندگی در تخمین کلر، به‌طور امتحانی این پارامتر نیز از ورودی‌ها حذف و مدل تنها بر اساس EC برازش شد. جدول ۶ نتایج این مدل جدید را نشان می‌دهد.

جدول ۶- مدل اثر تصادفی با یک ورودی

متغیر	ضرایب	P-value
ثابت معادله	-۱/۷۶۷	۰/۱۴۸۵
EC	۵/۶۶۴۹	۰/۰۰۰۰

برای بررسی دقت دو مدل اثر تصادفی قابل قبول از پارامترهای عملکرد R^2 و RMSE استفاده شد. نتایج در جدول ۷ آورده شده است.

جدول ۷- پارامترهای عملکرد مدل داده‌های ترکیبی

مدل داده‌های ترکیبی	R^2	RMSE
مدل اثر تصادفی با یک ورودی EC	۰/۹۶	۲/۴۴۵
مدل اثر تصادفی با دو ورودی بارندگی و EC	۰/۹۶۲	۲/۴۸۷

با توجه به R^2 های بیان شده در جدول ۷ دقت مدل اثر تصادفی با دو ورودی بالاتر از مدل اثر تصادفی با یک ورودی است اما در مقابل، RMSE مربوط به مدل اثر تصادفی با یک ورودی عدد کوچک تری را به خود اختصاص داده است. البته نکته قابل توجه، اختلاف پایین مربوط به هر دو پارامتر عملکرد در دو مدل اثر تصادفی است.

۴- نتیجه‌گیری

با توجه به نتایج این تحقیق هر دو مدل داده‌های ترکیبی دارای عملکرد قوی هستند اما از میان این دو مدل برحسب دو ضریب عملکرد R^2 و RMSE، داده‌های ورودی کمتر، مدل داده‌های ترکیبی

با یک ورودی EC بهتر است. این مدل نه تنها به ورودی کمتری نیاز دارد بلکه ورودی آن نیازی به پردازش پیش از ورود به مدل ندارد.

مدل داده‌های ترکیبی مدلی نوپا در مدیریت منابع آب است. با توجه به این تحقیق هر دو مدل با عملکرد بسیار خوبی توان تخمین پارامترهای کیفی در آب زیرزمینی را دارا هستند. رد پارامتر تبخیر توسط مدل داده‌های ترکیبی با سه ورودی با توجه به خصوصیات دشت دز فول- اندیمشک قابل پذیرش است. زیرا تبخیر تنها از ۲۰ درصد سطح دشت صورت می‌گیرد، یعنی ارتفاع آب کمتر از ۵ متر است. اما در مدل داده‌های ترکیبی با دو ورودی بارندگی و EC، نکته جالب ضریب مثبت بارندگی است که با توجه به اینکه منطقه توسط رودخانه‌های دز و کهنک تغذیه می‌شود و این دو رودخانه نقش بسزایی را در تغذیه دشت ایفا می‌نمایند، ضریب مثبت در این معادله دور از انتظار و بیانگر نقش بالای بارندگی در تغذیه و تأثیر بالای مقادیر کمی بر کیفیت آب دشت دز فول- اندیمشک دارد.

با استفاده از این مدل می‌توان هزینه و زمان آزمایش‌های کیفی در بخش آب زیرزمینی را کاهش داد و از پارامتر EC که دارای روش اندازه‌گیری کم هزینه‌ای است استفاده و پارامتر حائز اهمیت کلر را تخمین زد. در انتها توصیه می‌شود با نظر به جدید بودن این روش، کاربرد آن در مسائل دیگر مدل‌سازی و در مناطق دیگر برای مدل‌سازی کیفیت آب زیرزمینی مورد آزمون قرار گیرد.

۵- مراجع

- Ghafouri, V., Malekpour, N., and Mardani, A. (2011). "Geostatistical evaluation of groundwater quality plain Darab in Fars Province." *J. of Soil and Water Conservation*, 1(2), 81-94. (In Persian)
- Gholami, V., Derakhshan, S.H., and Dervari, Z. (2012). "Multiple regression analysis and artificial neural networks (ANN) in the simulation of groundwater salinity in the coastal of Mazandaran Province. *J. of Water Research in Agriculture*, 26(3), 355-365.
- Ghomeshion, M., Malekian, A., Hoseini, Kh., Gharachelo, S., and Khamoushi, M.R. (2012). "A survey on spatial variations of groundwater quality in Semnan/Sorkheh plain using geostatistical techniques." *Iranian J. of Range and Desert Research*, 19 (3), 535-545. (In Persian)
- Mohamadi, S., Salajegheh, A., Mahdavi, M., and Bagheri, R. (2012). "An investigation on spatial and temporal variations of groundwater level in Kerman plain using suitable geostatistical method (During a 10-year period)." *Iranian J. of Range and Desert Research*, 19 (1), 60-71. (In Persian)
- Ahmad Dar, I., Sankar, K., Ahmad Dar, M., and Majumder, M. (2012). "Fluorid contamination_ Artificial neural network modeling and inverse distance weighting approach." *J. of Water Sciences*, 25 (2), 165-182.
- Al-Mahallawi, K., Mania, J., Hani, A., and Shahrour, I. (2011). "Using of neural network for the prediction of nitrate groundwater contamination in rural and agricultural areas." *Environ. Earth Sci.*, 65, 917-928.
- Banerjee, P., Singh, V.S., Chattopadhyay, K., Chandra, P.C., and Singh, B. (2011). "Artificial neural network as potential alternative for groundwater salinity forecasting." *J. of Hydrology*, 398, 212-220.

8. Farahmand, A. R., Manshouri, M., Liaghat, A., and Sedghi, H. (2010). "Comparison of kriging, ANN and ANFIS model for spatial and temporal distribution modeling of groundwater contaminants." *J. of Food, Agriculture and Environment*, 8, 1146-1155.
9. Hwa Cho, K., Sthiannopkao, S., Pachepsky, Y.A., Kim, K.W., and Kin, J.H. (2011). "Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network." *Water Research*, 45, 5535-5544.
10. Zare Abyaneh, H., Bayat Varkeshi, M., and Mohammadi, K. (2011). "Assessment of groundwater corrosivity in Hamedan Province, Iran using an adaptive neuro_ Fuzzy inference system and (ANFIS)." *Geosciences Journal*, 15 (5), 433-439. (In Persian)
11. Zare Abyaneh, H., Nazemi, A. H., Neyshabouri, M. R., Mohammadi, K., and Majzoobi, G. H. (2005). "Chloride estimation in groundwater from conductivity measurement." *Tarim Bilimeri Dergisi*, 11 (1), 110-114.
12. Salehnia, N., and Falahi, M.A. (2011). "Evaluating eco-climatic variables on wheat yield using panel data model." *J. of Water and Soil*, 24 (2), 375-384.
13. Alijani, F., Karbasi, A., and Mozafari, M. (2011). "Effects of temperature and rainfall on the yield of irrigated wheat in Iran." *Agricultural Economic and Development*, 19(76), 143-166.
14. Falahi, M.A., Ansari, H., and Moghaddas, S. (2011). "Evaluating effective factors on household water consumption and forecasting demands: Panel data." *J. of Water and Wastewater*, 23-4(84), 78-87. (In Persian)
15. Arbues, F., and Barberan, R. (2004). "Price impact on residential water demand: A dynamic panel data approach." *Water Resources Research*, 40 (11), doi. 10.1029/2004 WR00-3092.
16. Izady, A., Davary, K., Alizadeh, A., Ghahraman, B., and Sadeghi, M. (2012). "Application of Panel data modeling to predict groundwater level in Neishaboer plain, Iran." *Hydrogeology Journal*, 20, 435-477.
17. Gujarati, D.(2012). *Basic econometrics*, University of Tehran Press, Tehran. (In Persian)
18. Aflatooni, A., and Nikbakht, L. (2010). *Applied econometric research in accounting, financial management and economic sciences*, 1st Ed., Termeh of Publication, Tehran. (In Persian)
19. Ashrafzadeh, H.R., and Mehregan, N. (2010). *Panel data econometrics*, 1st Ed., Cooperative Research Institute of Tech. Pub., Tehran university. (In Persian)