

برآورد پارامترهای هیدرودینامیکی LNAPL در آبخوان دشت ری با استفاده از مدل برنامه‌نویسی بیان ژن

کمال خدایی^۱، زهرا بوسلیک^{۲*}، علی اکبر شهسواری^۳، ناعمه فایضی^۴، بنیامین رضازاده^۴

چکیده

برآورد پارامترهای هیدرودینامیکی مایعات فاز غیرآبی سبک (Light Non-Aqueous Phase Liquid: LNAPL) از ارکان اساسی مطالعات انتقال آلودگی در آبخوان‌های نفتی و عاملی کلیدی در طراحی و مدیریت پروژه‌های پاکسازی محسوب می‌شود. با این حال، اندازه‌گیری مستقیم این پارامترها از طریق روش‌های صحرایی و آزمایشگاهی، معمولاً پرهزینه، زمان‌بر و در بسیاری موارد با محدودیت‌های اجرایی همراه است. در پژوهش حاضر، کارایی مدل برنامه‌ریزی بیان ژن (Gene Expression Programming: GEP) در تخمین پارامترهای هیدرودینامیکی LNAPL در محدوده صنعتی ری مورد ارزیابی قرار گرفت. داده‌های ماهانه ضخامت LNAPL، تراز سطح آب زیرزمینی، تراز سطح و قابلیت انتقال LNAPL حاصل از مدل LDRM Distribution and Recovery (Model) برای سه حلقه چاه نماینده طی دوره آماری ۱۳۸۵ تا ۱۳۹۹ گردآوری و مدل‌سازی با استفاده از نرم‌افزار GeneXproTools انجام شد. عملکرد مدل با شاخص‌های R^2 و RMSE ارزیابی شد. نتایج نشان داد مدل در تخمین تراز سطح LNAPL و قابلیت انتقال از دقت بالایی برخوردار است، به گونه‌ای که در برخی موارد ضریب تعیین به بیش از ۰/۹۹ رسید. در مقابل، عملکرد مدل در تخمین هدایت هیدرولیکی مطلوب نبود که علت آن به محدود بودن دامنه تغییرات این پارامتر و یکنواختی نسبی رسوبات آبخوان باز می‌گردد. نوآوری پژوهش در به کارگیری روش GEP برای استخراج روابط ریاضی صریح و قابل استفاده در برآورد پارامترهای هیدرودینامیکی آبخوان ری است. یافته‌ها ضمن تأکید بر کارایی بالای مدل GEP در برآورد پارامترهای پویا، ضرورت توجه به محدودیت‌های داده‌های ورودی در مدل‌های داده‌محور را آشکار می‌سازد.

واژه‌های کلیدی: آبخوان‌های آلوده، پارامترهای هیدرودینامیکی، LNAPL، مدل GEP.

پذیرفته شده برای چاپ
فصلنامه علمی
استناد

^۱ دانشیار گروه زمین‌شناسی محیطی، پژوهشکده علوم پایه کاربردی جهاد دانشگاهی شهید بهشتی.

^{۲*} استادیار گروه زمین‌شناسی محیطی، پژوهشکده علوم پایه کاربردی جهاد دانشگاهی شهید بهشتی. ایمیل:

zahra.boosalik@yahoo.com

^۳ استادیار گروه زمین‌شناسی محیطی، پژوهشکده علوم پایه کاربردی جهاد دانشگاهی شهید بهشتی.

^۴ گروه زمین‌شناسی محیطی، پژوهشکده علوم پایه کاربردی جهاد دانشگاهی شهید بهشتی.

مقدمه

آلودگی منابع آب زیرزمینی به هیدروکربن‌های نفتی، به‌ویژه ترکیبات فاز مایع غیرآبی سبک (LNAPLs)، یکی از بحرانی‌ترین چالش‌های ژئومحیط‌زیستی در مجتمع‌های صنعتی و پالایشگاهی سراسر جهان محسوب می‌شود (Lari et al., 2018). این آلاینده‌ها به دلیل چگالی کمتر از آب و انحلال‌پذیری محدود، پس از نفوذ به زیرسطح، در زون غیراشباع و در امتداد نوسانات سطح ایستابی تجمع یافته و یک فاز مجزا و متحرک تشکیل می‌دهند (Tomlinson et al., 2017). مدیریت ریسک و طراحی فرآیندهای اصلاح صحرائی^۱ در این محیط‌ها، نیازمند درک دقیق و کمی از فرآیندهای انتقال جرم و هیدرودینامیک جریان چندفازی است (Mercer & Cohen, 1990). در این راستا، تعیین دقیق پارامترهای هیدرودینامیکی کلیدی نظیر هدایت هیدرولیکی (K)، قابلیت انتقال آبخوان (T) و پویایی تراز سطح LNAPL، پیش‌شرط اساسی برای شبیه‌سازی عددی انتقال آلودگی و طراحی بهینه سامانه‌های بازیابی فعال (مانند سیستم‌های پمپاژ دوگانه) به‌شمار می‌رود (Charbeneau et al., 2000).

با وجود اهمیت بنیادین، اندازه‌گیری مستقیم این پارامترها با استفاده از روش‌های کلاسیک هیدروژئولوژیکی، نظیر آزمون پمپاژ^۲ یا اسلاگ تست^۳ و آزمایش‌های آزمایشگاهی، با محدودیت‌های شدیدی مواجه است. این روش‌ها علاوه بر تحمیل هزینه‌های سنگین مالی و زمانی، به دلیل ماهیت غیرخطی جریان چندفازی و ناهمگنی ذاتی ساختار زمین‌شناختی، اغلب نتایجی موضعی ارائه می‌دهند که قابلیت تعمیم در مقیاس صحرائی را ندارند (Butler, 2020). علاوه بر این، حضور فاز آلاینده در منافذ خاک، روابط مشخصه فشار موئینگی-اشباع (S-Pc) را به‌شدت تحت تأثیر قرار داده و کارایی فرضیات خطی روش‌های سنتی را در برآورد پارامترهای هیدرولیکی مؤثر به حداقل می‌رساند. از این‌رو، توسعه رویکردهای داده‌محور که قادر باشند با کمترین نیاز به آزمون‌های مخرب میدانی، برآوردی کالیبره و قابل اطمینان ارائه دهند، ضرورتی اجتناب‌ناپذیر در مهندسی آب‌های زیرزمینی است.

در دو دهه اخیر، الگوریتم‌های یادگیری ماشین (ML) و هوش مصنوعی (AI) به عنوان ابزارهایی کارآمد در شبیه‌سازی فرآیندهای پیچیده هیدروژئولوژیکی مطرح شده‌اند (Rajaei et al., 2019). رویکردهایی نظیر شبکه‌های عصبی مصنوعی (ANN)، سیستم‌های استنتاج فازی-عصبی (ANFIS) و ماشین‌های بردار پشتیبان (SVM) به‌طور گسترده در مدل‌سازی سطح ایستابی و هیدروشیمی به‌کار گرفته شده‌اند (Chen et al., 2021; Mahmoud et al., 2023). با این وجود، بزرگترین محدودیت ساختاری این مدل‌ها، ماهیت «جعبه‌سیاه»^۴ آن‌هاست؛ چرا که عدم شفافیت در ساختار درونی، مانع از درک فیزیکی فرآیند مدل‌سازی و استخراج معادلات ریاضی کاربردی برای استفاده‌های مهندسی می‌شود (Kisi et al., 2021). برای رفع این نقیصه، مدل برنامه‌ریزی بیان ژن (GEP) به عنوان یک الگوریتم تکاملی پیشرفته بر پایه ژنوم دوگانه، توانسته است با ادغام مزایای الگوریتم ژنتیک (GA) و برنامه‌ریزی ژنتیکی (GP)، علاوه بر قدرت بالا در تقریب توابع پیچیده، خروجی‌هایی در قالب ساختارهای ریاضی صریح و تفسیرپذیر ارائه دهد.

¹ Remediation

² Pumping Test

³ Slug Test

⁴ Black-box

مرور مطالعات پیشین نشان می‌دهد که مدل‌های هوشمند در حوزه‌های مختلف هیدرولوژی کاربرد یافته‌اند. برای مثال، استفاده از ترکیب GA با مدل اجزای محدود برای تخمین پارامترهای هیدرولیکی (Rastogi, 2001 & Prasad) و کاربرد GEP در تخمین نوسانات سطح آب زیرزمینی (Kisi, 2011 & Shiri) و تبخیر و تعرق (Guven, 2012 & Traore) گزارش شده است. همچنین، پژوهش‌های متعددی بر پیش‌بینی نوسانات سطح ایستابی (Abareshi et al., 2014) و غلظت آلاینده‌ها (Sadeghi Aghdam et al., 2016) تمرکز داشته‌اند. در سال‌های اخیر نیز، توسعه مدل‌های ترکیبی (Hybrid AI) برای برآورد هدایت هیدرولیکی در آبخوان‌های آلوده (Salehi et al., 2023) و مدل‌سازی داده‌محور انتقال LNAPL (Kumar et al., 2024) گام‌های مهمی در این مسیر محسوب می‌شوند. با این حال، تحلیل انتقادی ادبیات تحقیق نشان می‌دهد که بخش عمده مطالعات یا بر پیش‌بینی پارامترهای عمومی آبخوان‌های بکر متمرکز بوده‌اند، یا در مدل‌های مبتنی بر GEP، تمرکز اصلی صرفاً بر کاهش خطای پیش‌بینی بوده و از اهمیت استخراج روابط ریاضی صریح و بررسی تأثیر توزیع آماری داده‌ها بر پایداری ساختار مدل (به‌ویژه در مواجهه با داده‌های با تغییرپذیری مکانی اندک که منجر به بیش‌برازش یا Overfitting می‌شود) مغفول مانده است. به‌ویژه، پژوهش‌های متمرکز بر برآورد همزمان پارامترهای هیدرودینامیکی فاز LNAPL در سایت‌های آلوده واقعی با استفاده از مدل‌های تفسیرپذیر مانند GEP، بسیار محدود است. این خلأ مطالعاتی، فرصتی برای ارائه یک چارچوب مدل‌سازی جدید فراهم می‌آورد. پژوهش حاضر با هدف پر کردن این شکاف‌های دانشی، به ارزیابی جامع کارایی مدل GEP در برآورد همزمان سه پارامتر کلیدی شامل هدایت هیدرولیکی (K)، قابلیت انتقال (T) و تراز سطح LNAPL در یک سایت صنعتی پیچیده می‌پردازد. محدوده مطالعاتی، آبخوان صنعتی ری در جنوب دشت تهران است (شکل ۱)؛ منطقه‌ای که به دلیل استقرار پالایشگاه بزرگ تهران و چندین شرکت صنعتی، بیش از نیم قرن با نشت مستمر ترکیبات نفتی و تشکیل لایه‌های ضخیم LNAPL دست‌به‌گریبان است. این آبخوان از نوع آبرفتی با لایه‌بندی چندلایه بوده و به دلیل پمپاژ گسترده صنعتی، دارای نوسانات شدید تراز آب و انحراف گرادیان هیدرولیکی به سمت غرب است که پویایی جریان آلاینده را به شدت غیرخطی کرده است. نوآوری‌های اصلی این مطالعه در سه محور تعریف می‌شود:

۱. کاربرد هدفمند GEP در حوزه تخصصی: نخستین تلاش برای به‌کارگیری مدل GEP جهت برآورد همزمان پارامترهای هیدرودینامیکی فاز LNAPL در یک آبخوان آلوده واقعی.
۲. ارائه روابط ریاضی صریح و کاربردی: جایگزینی مدل‌های جعبه‌سیاه با معادلات ریاضی تفسیرپذیر و مستقل برای هر چاه، جهت تسهیل مدیریت مهندسی و کاهش هزینه‌های محاسباتی.
۳. ارزیابی جامع پارامترهای چندگانه: بررسی توانمندی مدل در مواجهه با پارامترهایی با ماهیت‌های فیزیکی متفاوت (پویای سطح LNAPL، ساختاری K و ترکیبی T) و تحلیل پایداری مدل در شرایط تغییرپذیری داده‌ای محدود.

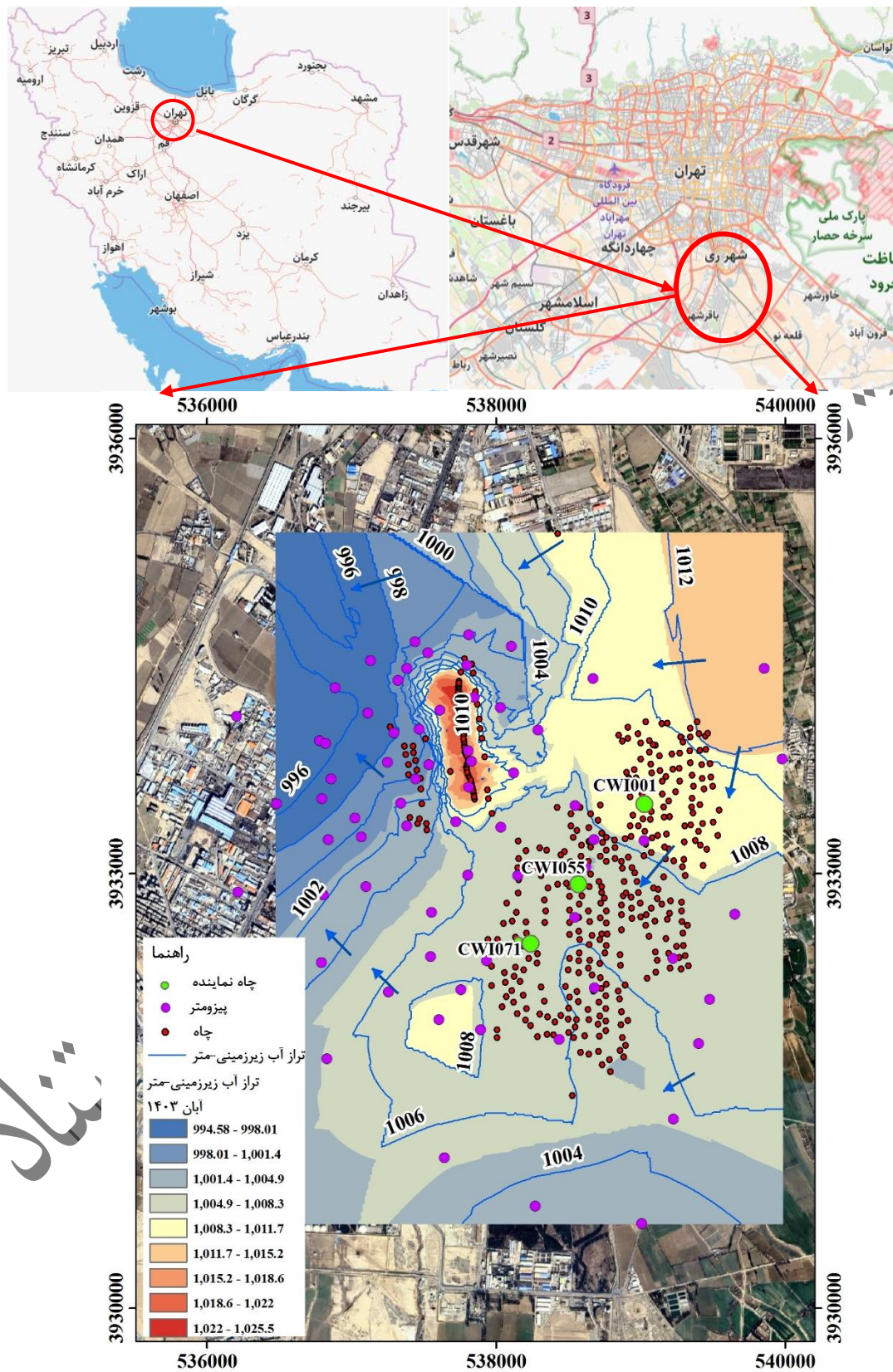


Fig 1- Location of the study area

شکل ۱- موقعیت منطقه مورد مطالعه

مواد و روش‌ها

داده‌های مورد استفاده: در محدوده صنعتی ری، شبکه پایشی متشکل از حدود ۴۰۰ حلقه چاه مشاهده‌ای/پیزومتری به‌منظور بررسی وضعیت آبخوان و رفتار آلودگی نفتی استقرار یافته است. با توجه به تفاوت‌های هیدروژئولوژیکی موجود در منطقه، چاه‌های پایش بر اساس ویژگی‌های هیدرولیکی، مشخصات لیتولوژیکی و هندسه لایه‌های آبخوان به سه گروه اصلی با رفتار هیدرودینامیکی مشابه تقسیم‌بندی شدند. به‌منظور توسعه مدل‌های هوشمند، از میان چاه‌های موجود، سه حلقه چاه شاخص شامل CWI001، CWI055 و CWI071 که دارای کامل‌ترین و پیوسته‌ترین داده‌های ثبت‌شده بودند، به‌عنوان نماینده سه گروه مذکور انتخاب شدند (هر چاه به عنوان نماینده یک گروه از چاه‌ها است). انتخاب این چاه‌ها این امکان را فراهم می‌سازد که نتایج حاصل از مدل‌سازی، با رعایت حدود اعتبار، به سایر چاه‌های دارای ویژگی‌های مشابه در هر گروه تعمیم داده شود. دوره آماری مورد مطالعه شامل داده‌های ماهانه ۱۴ ساله از سال ۱۳۸۵ تا ۱۳۹۹ است. متغیرهای مورد استفاده در این پژوهش شامل تراز آب زیرزمینی (WT)، تراز سطح LNAPL (OL)، عمق سطح آب (D)، ضخامت ظاهری LNAPL (OT)، قابلیت انتقال LNAPL (LT) و هدایت هیدرولیکی آبخوان (K) بود. لازم به ذکر است که داده‌های مربوط به LT از نتایج مدل LDRM استخراج شده‌اند. در این پژوهش مدل GEP با استفاده از داده‌های حاصل از LDRM، روابط جایگزین و صریحی را برای تخمین پارامترهای هدف توسعه می‌دهد؛ به بیان دیگر، هدف صرفاً بازتولید مکانیکی خروجی یک مدل دیگر نیست، بلکه استخراج روابط کاربردی و کم‌هزینه‌ای است که بتوانند با استفاده از متغیرهای ورودی در دسترس، رفتار پارامترهای هدف را تقریب بزنند. مشخصات آماری داده‌های مورد استفاده، شامل میانگین، کمینه و بیشینه متغیرها، برای دوره‌های آموزش و آزمون در جدول ۱ ارائه شده است. برای توسعه و ارزیابی مدل‌ها، داده‌های هر چاه به دو بخش تقسیم شدند؛ به‌طوری‌که داده‌های ۷ سال نخست (۱۳۸۵ تا ۱۳۹۲) برای آموزش/واسنجی مدل و داده‌های ۷ سال دوم (۱۳۹۲ تا ۱۳۹۹) برای آزمون و صحت‌سنجی آن مورد استفاده قرار گرفتند. مبنای تقسیم‌بندی داده‌ها، حفظ ترتیب زمانی مشاهدات و ارزیابی توان پیش‌بینی مدل در مواجهه با داده‌های دیده‌نشده بوده است؛ از این رو، داده‌های اولیه برای آموزش و داده‌های بعدی برای آزمون در نظر گرفته شدند. این نوع تفکیک، نسبت به تقسیم تصادفی، برای سری‌های زمانی و داده‌های فرایندی واقع‌بینانه‌تر است؛ زیرا شرایط واقعی پیش‌بینی را بهتر بازنمایی می‌کند. برای اطمینان از اینکه نتایج صرفاً محصول یک تفکیک خاص نیستند، چند آرایش مختلف تقسیم داده نیز طی پژوهش بررسی شد و روند کلی عملکرد مدل پایدار باقی ماند.

پیش‌پردازش داده‌ها: از آنجا که متغیرهای مورد مطالعه دارای واحدها و دامنه‌های عددی متفاوتی هستند، پیش از ورود به فرآیند مدل‌سازی، کلیه داده‌ها نرمال‌سازی شدند. این اقدام به‌منظور جلوگیری از غلبه متغیرهای با مقادیر عددی بزرگ‌تر بر فرآیند یادگیری، افزایش سرعت همگرایی الگوریتم و بهبود دقت پیش‌بینی انجام شد.

مدل‌سازی با استفاده از برنامه‌نویسی بیان ژن: به‌منظور استخراج روابط غیرخطی بین متغیرهای هیدرودینامیکی آبخوان



و پارامترهای مرتبط با LNAPL، از الگوریتم برنامه‌نویسی بیان ژن (Gene Expression Programming: GEP) استفاده شد. این الگوریتم یکی از روش‌های تکاملی است که با ترکیب قابلیت‌های الگوریتم ژنتیک (GA) و برنامه‌نویسی ژنتیک (GP)، امکان استخراج روابط ریاضی صریح و قابل تفسیر را فراهم می‌کند. در روش GEP، ساختارهای ریاضی به صورت کروموزوم‌های خطی با طول ثابت کدگذاری شده و سپس به درختان بیانی تبدیل می‌شوند. هر کروموزوم در این روش از یک یا چند ژن تشکیل شده است و هر ژن دارای دو بخش «سر» (Head) و «دم» (Tail) است. بخش سر می‌تواند شامل توابع و متغیرها باشد، در حالی که بخش دم تنها از متغیرها و ثابت‌ها تشکیل می‌شود. طول بخش دم بر اساس طول سر و حداکثر تعداد آرگومان توابع مورد استفاده از رابطه زیر به دست می‌آید:

$$t = h(n - 1) + 1$$

معادله ۱:

که در آن، t طول دم، h طول سر و n بیشترین تعداد آرگومان توابع موجود در مجموعه تابع‌ها است.

در این پژوهش، پیاده‌سازی مدل GEP در محیط نرم‌افزار 5 GeneXproTools انجام شد. انتخاب پارامترهای ساختاری مدل، از جمله طول سر کروموزوم و تعداد ژن‌ها، بر اساس بررسی اولیه، تحلیل حساسیت و نیز با بهره‌گیری از نتایج مطالعات پیشین (Abareshi et al., 2014; Sadeghi Aghdam et al., 2016; Kumar et al., 2024) انجام گرفت تا تعادل مناسبی میان دقت پیش‌بینی و پیچیدگی روابط استخراجی برقرار شود. بر این اساس، طول سر کروموزوم برابر ۸ و تعداد ژن‌ها برابر ۳ در نظر گرفته شد.

مراحل اجرای مدل GEP: فرآیند توسعه مدل GEP در این پژوهش در شش گام اصلی انجام شد:

تعریف متغیرهای ورودی و خروجی برای هر سناریوی مدل‌سازی؛

انتخاب توابع ریاضی و عملگرهای جبری مورد استفاده در ساختار روابط؛

تعیین تابع برازش (Fitness Function) برای هدایت فرآیند تکامل؛

تنظیم معماری کروموزوم‌ها شامل طول سر کروموزوم و تعداد ژن‌ها؛

انتخاب تابع پیونددهنده برای ترکیب خروجی ژن‌ها؛

تعیین نرخ عملگرهای ژنتیکی شامل جهش، جابه‌جایی و نوترکیبی.

پس از اعمال تنظیمات اولیه، ترکیب‌های مختلفی از ساختار مدل و پارامترهای کنترلی به صورت آزمون و خطا بررسی شد تا

مناسب‌ترین توپولوژی برای هر سناریوی شبیه‌سازی تعیین شود.

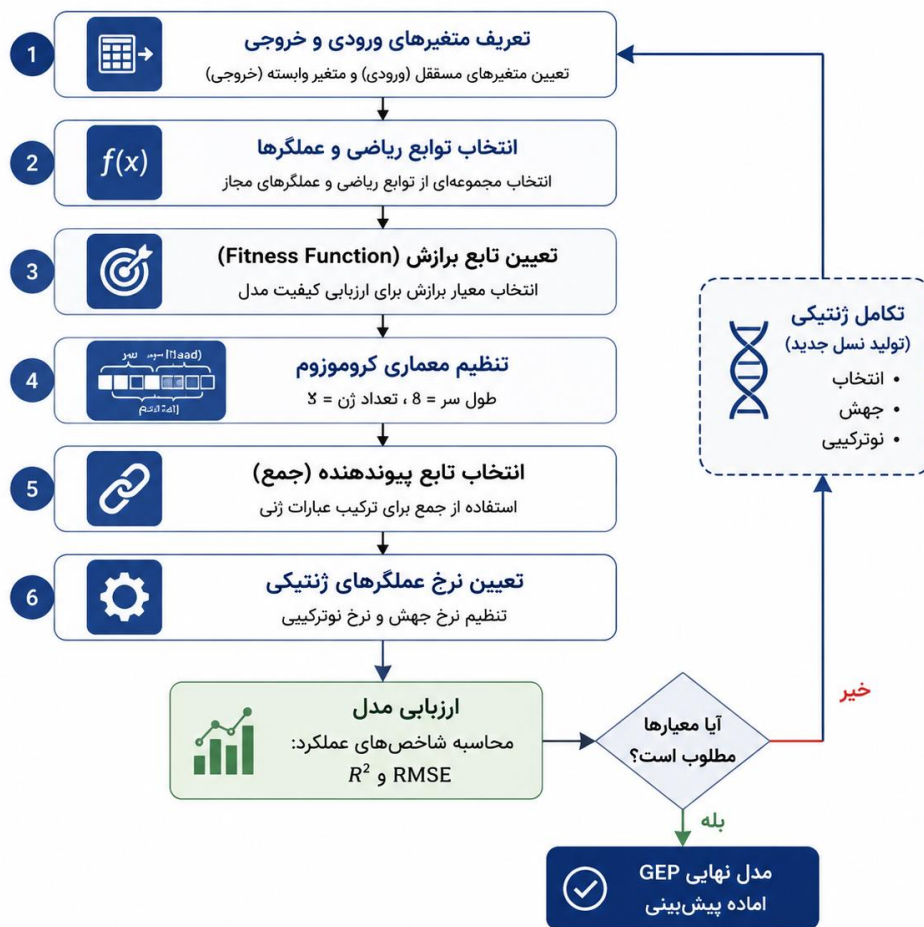


Fig 2- Staged development model

شکل ۲- مراحل توسعه مدل GEP

سناریوهای مدل‌سازی

با توجه به روابط فیزیکی حاکم بر رفتار LNAPL و پارامترهای هیدرودینامیکی آبخوان، سه سناریوی اصلی برای مدل‌سازی در هر یک از چاه‌های منتخب تعریف شد. در این سناریوها، متغیرهای ورودی بر اساس میزان اثرگذاری بر پارامتر هدف انتخاب شدند. سناریوهای مورد بررسی در جدول ۲ خلاصه شده‌اند.

سناریوی (۱) شبیه‌سازی قابلیت انتقال LNAPL: ورودی‌ها: OT و K، خروجی: LT

سناریوی (۲) شبیه‌سازی تراز سطح LNAPL: ورودی‌ها: WT، OT و K، خروجی: OL

سناریوی (۳) شبیه‌سازی هدایت هیدرولیکی آبخوان: ورودی‌ها: OT و LT، خروجی: K

در هر سناریو، مدل به صورت مجزا برای هر یک از سه چاه منتخب اجرا شد تا برای هر چاه و هر پارامتر هدف، یک رابطه مستقل استخراج شود.

معیارهای ارزیابی عملکرد مدل: به منظور ارزیابی دقت و کارایی مدل‌های توسعه یافته در مراحل آموزش و آزمون، از دو شاخص

آماري ضريب تعيين (R^2) و ریشه میانگین مربعات خطا (RMSE) استفاده شد. ضريب تعيين نشان دهنده میزان تطابق بین مقادير پیش‌بینی شده و مقادير مشاهده شده است و هرچه مقدار آن به ۱ نزدیک‌تر باشد، عملکرد مدل مطلوب‌تر خواهد بود. همچنین، شاخص RMSE میزان انحراف عددی پیش‌بینی‌ها از مقادير واقعی را نشان می‌دهد و مقادير کمتر آن بیانگر دقت بالاتر مدل است.

Table 1- Statistical parameters of the data under study

جدول ۱- پارامترهای آماری داده‌های مورد مطالعه

Well	Period	Statistical parameter	LT (m)	OL (m)	WT (m)	D (m)	OT (m ² /s)	K (m/day)
CWI001	Training	AVR	0.71	1010.95	1006.2	17.18	4.76	0.4
		MAX	3.59	1017.86	1011.6	21.33	14.62	0.4
		MIN	0.02	1005.27	1002	11.8	0.49	0.4
	Test	AVR	0.7	1010.31	1006.3	17.08	4.02	0.4
		MAX	2.11	1014.09	1009.9	20.92	9.7	0.4
		MIN	0.03	1006.5	1002.4	13.49	0.55	0.4
	Period	AVR	0.7	1010.71	1006.2	17.14	4.49	0.4
		MAX	3.59	1017.86	1011.6	21.33	14.62	0.4
		MIN	0.02	1005.27	1002	11.8	0.49	0.4
CWI055	Training	AVR	0.34	1005.38	1003.1	18.28	2.31	0.3
		MAX	1.61	1008.34	1005.9	21.72	6.48	0.4
		MIN	0	1002.56	999.61	15.45	0	0.2
	Test	AVR	0.09	1006.04	1005.3	16.08	0.78	0.3
		MAX	0.46	1009.07	1008.7	19.66	2.5	0.4
		MIN	0	1003.72	1001.7	12.61	0.35	0.2
	Period	AVR	0.21	1005.72	1004.2	17.15	1.53	0.3
		MAX	1.61	1009.07	1008.7	21.72	6.48	0.4
		MIN	0	1002.56	999.61	12.61	0	0.2
CWI071	Training	AVR	0.14	1004.79	1003.3	17.36	1.51	0.4
		MAX	0.4	1006.33	1005.2	19.74	2.76	0.4
		MIN	0.01	1002.25	1000.9	15.44	0.33	0.4
	Test	AVR	0.07	1006.98	1006	14.66	1.01	0.4
		MAX	0.33	1010.39	1009.8	18.06	2.11	0.4
		MIN	0	1004.63	1002.6	10.79	0.49	0.4
	Period	AVR	0.1	1005.91	1004.7	15.94	1.22	0.4
		MAX	0.4	1010.39	1009.8	19.74	2.76	0.4
		MIN	0	1002.25	1000.9	10.79	0	0.4

WT: سطح ایستابی، OL: ارتفاع سطح LNAPL، D: عمق سطح آب، OT: ضخامت LNAPL، LT: قابلیت انتقال و K: هدایت هیدرولیکی پارامترهای هدف و پارامترهای ورودی معرفی شده برای هر شبیه‌سازی در جدول ۲ ارائه گردیدند.

Table 2- Inputs and outputs of GEP models performed

جدول ۲- ورودی و خروجی‌های مدل‌های GEP انجام شده

Model	Inputs	Target Parameter
LNAPL Transmissivity Simulation	LNAPL thickness, LNAPL thickness, and hydraulic conductivity	LNAPL Transmissivity
LNAPL Thickness Simulation	Water table elevation, LNAPL thickness, and hydraulic conductivity	LNAPL Level
Hydraulic Conductivity Simulation	LNAPL thickness, LNAPL thickness, and LNAPL transmissivity	Hydraulic Conductivity

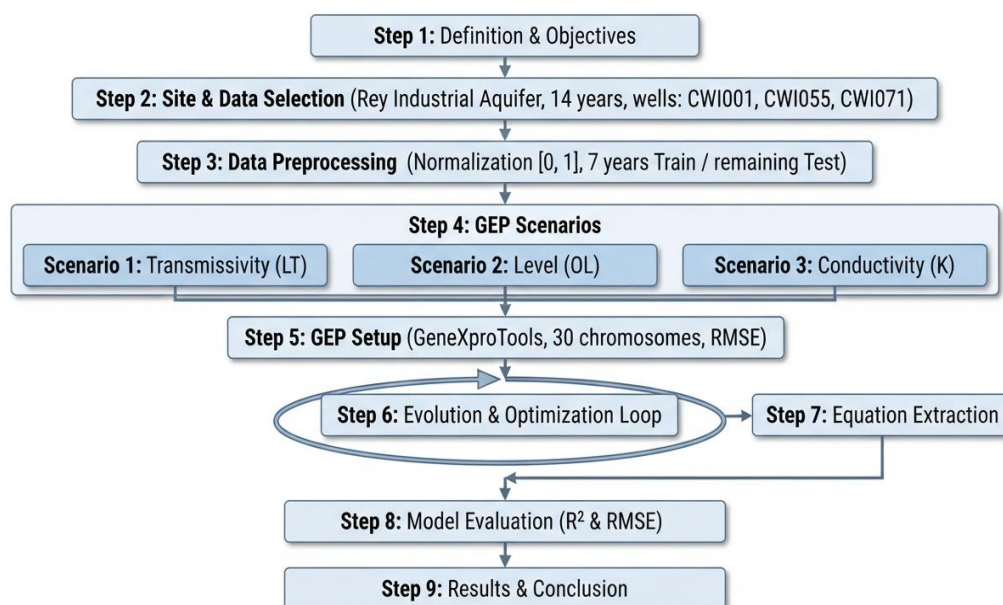


Fig 3- The flowchart of the study

شکل ۳-فلوچارت مطالعه

نتایج و تحلیل نتایج

ارزیابی مدل GEP در برآورد قابلیت انتقال LNAPL

به‌منظور برآورد میزان قابلیت انتقال LNAPL با استفاده از مدل برنامه‌نویسی بیان ژن (GEP)، پارامترهای «هدایت هیدرولیکی» و «ضخامت LNAPL» به‌عنوان متغیرهای ورودی و «قابلیت انتقال LNAPL» به‌عنوان متغیر خروجی (هدف) تعیین شدند. مشخصات عملگرهای ژنتیکی به‌کاررفته در توسعه مدل در جدول ۳ ارائه شده است. در این مدل، بر اساس تست‌های اولیه و رجوع به الگوهای متداول در مطالعات پیشین (Abareshi et al., 2014; Sadeghi Aghdam et al., 2016; Kumar et al., 2024)،

تعداد کروموزوم‌ها برابر با ۳۰، طول سر کروموزوم برابر با ۸ و تعداد ژن‌ها در هر کروموزوم برابر با ۳ در نظر گرفته شد. جهت دستیابی به بهینه‌ترین ساختار رابطه ریاضی، مجموعه‌های متفاوتی از توابع و عملگرها مورد آزمون قرار گرفتند و در هر گام، شاخص‌های آماری مدل ارزیابی شدند. به عنوان نمونه، سه مورد از توابع تست شده برای چاه CWI001 به همراه بهینه‌ترین توابع برازش یافته برای چاه‌های CWI055 و CWI071 در جدول ۳ (در سطرهای خاکستری رنگ) نمایش داده شده است. بررسی عملکرد مدل‌ها نشان داد که در هر سه چاه منتخب، استفاده از مجموعه چهار عملگر اصلی ریاضی (شامل جمع، تفریق، ضرب و تقسیم) منجر به بهینه‌ترین نتایج و بالاترین میزان انطباق می‌گردد. برای نمونه در چاه CWI001، مقدار ضریب تعیین (R^2) برای مجموعه‌های دوم و سوم توابع آزمایش شده به ترتیب برابر با ۰/۷۰ و ۰/۸ حاصل شد. جدول ۴ مقادیر شاخص‌های آماری مختلف را جهت ارزیابی همبستگی میان داده‌های تخمین زده شده توسط مدل و داده‌های محاسباتی نشان می‌دهد. بر اساس نتایج، کمترین مقدار R^2 در مدل‌های نهایی برابر با ۰/۹۴ و بیشترین مقدار آن برابر با ۰/۹۹ به دست آمده است که این امر نشان دهنده انطباق و برازش بسیار بالای داده‌های محاسباتی و مشاهداتی است. با توجه به این شاخص‌های عملکردی که حاصل به کارگیری توابع برازش بهینه در تخمین قابلیت انتقال چاه‌های منطقه است، می‌توان دریافت که مدل GEP از کارایی و دقت فوق العاده‌ای در برآورد و شبیه سازی پارامتر قابلیت انتقال LNAPL برخوردار است.

Table 3- Values of genetic operators used in the GEP model

جدول ۳- مقادیر اپراتورهای ژنتیکی به کار رفته در مدل GEP

Well	Set of functions used	Chrom. num.	Chrom. head size	Gene num.	Mutation rate	RNC rate	One-point crossover rate	Two-point crossover rate	Gene crossover rate	Gene displacement	IS displacement
CWI001	+ , - , × , /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005
	+ , - , × , / , ln x ,	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005
	ex + , - , × , / + , - , , x3 , ln x	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005
CWI055	+ , - , × , /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005
CWI071	+ , - , × , /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005

Table 4- Various statistical indices to estimate the transmissibility of CWI001 well

جدول ۴- شاخص‌های آماری مختلف برای تخمین قابلیت انتقال چاه CWI001

Function	CWI001		CWI055		CWI071	
	Training	Test	Training	Test	Training	Test
RMSE	0.19	0.26	0.12	0.05	0.02	0.05
R^2	0.95	0.94	0.99	0.94	0.96	0.95
Fitness-RMSE	841.7	852.2	891.27	943.978	983.492	966.029

از بررسی شکل ۴ نیز مشاهده می‌شود که مقادیر قابلیت انتقال برآورد شده توسط مدل GEP در حد قابل قبولی به مقادیر واقعی قابلیت نزدیک بوده و همبستگی خوبی را نشان می‌دهد این امر خود نشان‌دهنده کارایی خوب مدل GEP در برآورد میزان قابلیت انتقال می‌باشد.

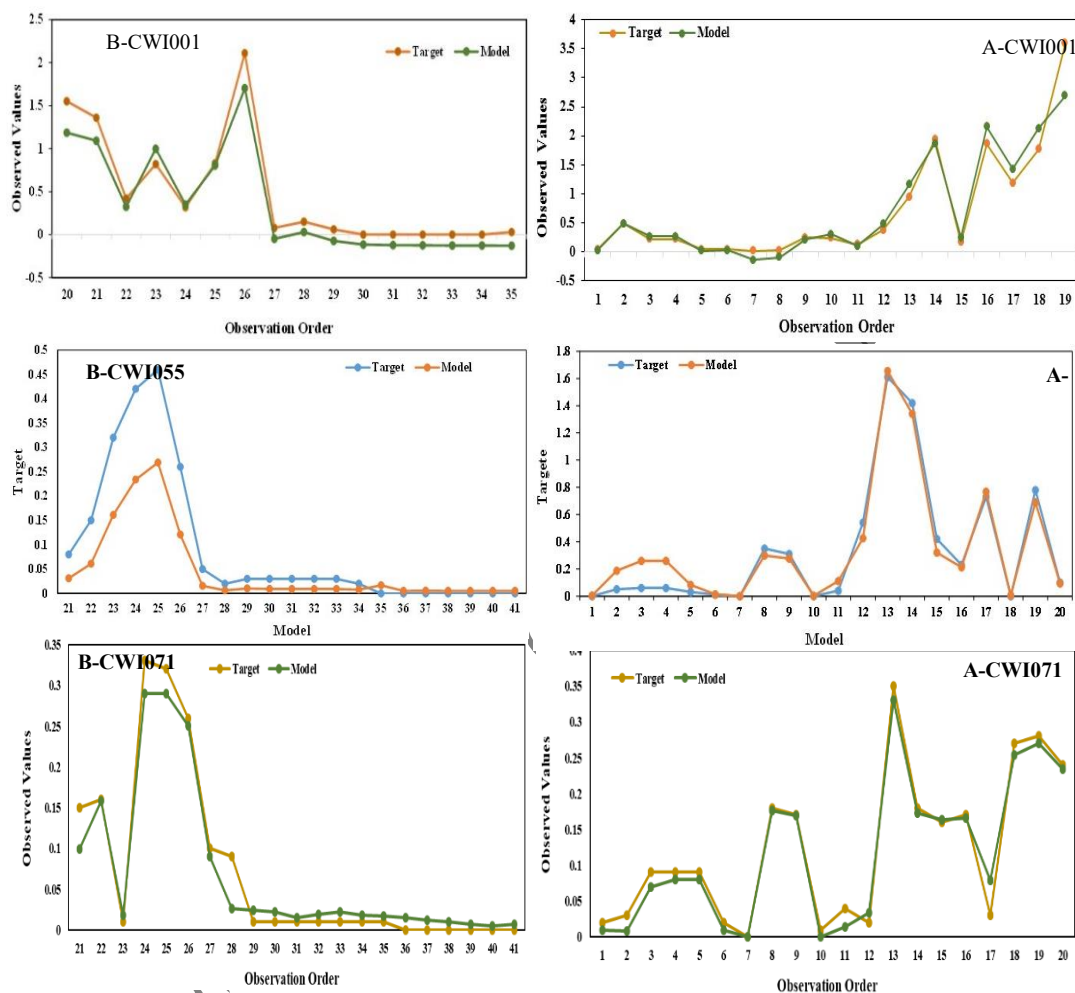


Fig 4- Time series graph of transfer capability in the most optimal execution mode

شکل ۴- نمودار سری زمانی قابلیت انتقال در بهینه‌ترین حالت اجرای مدل

یکی از مزایای اصلی مدل GEP در مقایسه با سایر مدل‌های هوش مصنوعی، ارائه بیان ریاضی برای فرآیند مورد مطالعه است. بر این اساس، معادله کلی بدست آمده برای پیش‌بینی قابلیت انتقال LNAPL برای هر کدام از چاه‌های منتخب مطابق معادله‌های جدول ۵ می‌باشد. این معادله حاصل جمع، روابط استخراجی برای سه ژنی می‌باشد که در مرحله آغازین تنظیم شد. لازم به ذکر است این روابط به صورت مطلق برای کل آبخوان و همه شرایط زمین‌شناسی قابل تعمیم در نظر گرفته نمی‌شوند، بلکه در چارچوب گروه‌های چاه با مشخصات هیدرودینامیکی مشابه و در محدوده شرایط داده‌های مورد استفاده قابل بهره‌برداری هستند. اهمیت این روابط در آن است که برای هر گروه نماینده، یک ابزار سریع و کم‌هزینه برای تخمین پارامترهای هدف فراهم می‌کند و می‌تواند

در غربالگری اولیه، تحلیل‌های مدیریتی و کاهش نیاز به اجرای مکرر مدل‌های پیچیده‌تر مورد استفاده قرار گیرند.

Table 5- Relationships derived from each subtree to predict the transferability

جدول ۵- روابط حاصل از هر زیر درخت برای پیش‌بینی قابلیت انتقال

Well	Equation
CWI001	$\text{Transmissivity} = \left(\left((-0.43 - k^2) \right) \times \left(\frac{(k+0.1)}{T} \right) \right) + 0.83 + \frac{(OT+K)}{-8.75} \left(\frac{1}{OT} \right) \left((K+OT) - (OT+9.1) \right) + \left(\frac{OT}{9.8} \right) \left(\frac{1}{OT} + 1 \right) / 1.76K$
CWI055	$\text{Transmissivity} = \frac{\left(\left(\frac{dk}{-6.5} \right) \times d(K) \right) + (d(OT) \times d(K))}{(11.8) - d(OT)} + \frac{\left(\frac{6.7 + d(K)}{1.98} \right) + (1.98 + d(OT))}{(14.43) + (1.98d(K))} + \frac{(d(K)^2) - 8.6}{(14.43) + (1.98d(K))} \times (d(OT) - d(K))$
CWI071	$\text{Transmissivity} = \left(d(K) \times \frac{((d(K) - d(OT)) \times d(OT))}{(-8.97) - d(K)} \right) + \left(\left((-0.18 + d(K)) - \frac{(d(OT) + d(K)) \times d(OT)}{-8.2 + d(K)} \right) \times \left((d(K)^2) \times (d(K) - d(K)) \right) \right) + \left(\frac{8.2}{d(K)^2} + \left(-\frac{8.2}{d(K)} \right) \right)$

ارزیابی مدل GEP در برآورد هدایت هیدرولیکی

به منظور برآورد میزان هدایت هیدرولیکی با استفاده از مدل GEP، پارامترهای «قابلیت انتقال» و «ضخامت LNAPL» به عنوان متغیرهای ورودی و «هدایت هیدرولیکی» به عنوان متغیر خروجی تعریف شدند. تنظیمات مربوط به عملگرهای ژنتیکی مورد استفاده برای چاه CWI055 در جدول ۶ ارائه شده است. در این مدل نیز مشابه سناریوهای پیشین، تعداد کروموزوم‌ها ۳۰، طول سر کروموزوم ۸ و تعداد ژن‌ها برابر با ۳ در نظر گرفته شد. همچنین پس از ارزیابی مجموعه‌های مختلف توابع، ترکیب چهار عملگر اصلی ریاضی (جمع، تفریق، ضرب و تقسیم) به عنوان ساختار پایه انتخاب گردید (جدول ۶). جدول ۷ نشان‌دهنده شاخص‌های آماری حاصل از ارزیابی همبستگی میان مقادیر تخمینی مدل و داده‌های محاسباتی است. تحلیل این شاخص‌ها نشان می‌دهد که مدل GEP در برآورد و شبیه‌سازی پارامتر هدایت هیدرولیکی در محدوده مورد مطالعه از کارایی و دقت لازم برخوردار نیست. علت این امر را می‌توان به یکنواختی نسبی رسوبات و عدم وجود تغییرات مکانی و زمانی مشهود در پارامتر هدایت هیدرولیکی نسبت داد؛ به طوری که به دلیل ثبات بودن تقریبی این پارامتر در دو چاه CWI001 و CWI071، امکان اجرای مدل و برقراری رابطه معنادار عملاً میسر نگردید. جهت اطمینان از عدم کارایی مدل در شبیه‌سازی این پارامتر، مدل‌سازی با ترکیب‌های متفاوتی از متغیرهای ورودی مجدداً تکرار شد، اما در هیچ‌یک از حالات، نتیجه بهینه‌ای حاصل نگردید. مطابق با شکل ۵، مقادیر هدایت هیدرولیکی

برآورد شده توسط مدل GEP انطباق ضعیفی با مقادیر واقعی نشان می‌دهد، به طوری که ضریب تعیین (R^2) برای داده‌های آموزش و آزمون به ترتیب برابر با ۰/۲۶ و ۰/۷ به دست آمد. با توجه به عدم برازش مناسب و خطای بالای مدل، معادله مستخرج از آن فاقد اعتبار علمی لازم تشخیص داده شد و لذا از ارائه آن در این بخش صرف نظر گردید.

Table 6- Values of genetic operators used in the GEP model

جدول ۶- مقادیر اپراتورهای ژنتیکی به کار رفته در مدل GEP

Parameter Description	Set of Functions Used	Chromosome Number	Chromosome Head Size	Gene Number	Mutation Rate	RNC Rate	One-Point Crossover Rate	Two-Point Crossover Rate	Gene Crossover Rate	Gene Translocation Rate	IS Translocation Rate
Parameter setting	+, -, *, /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005

Table 7- Results of selecting different fitting functions of well CWI055 model

جدول ۷- نتایج انتخاب تابع‌های برازش مختلف مدل چاه CWI055

Function	RMSE	R^2
Training	0.07	0.26
Test	0.05	0.8

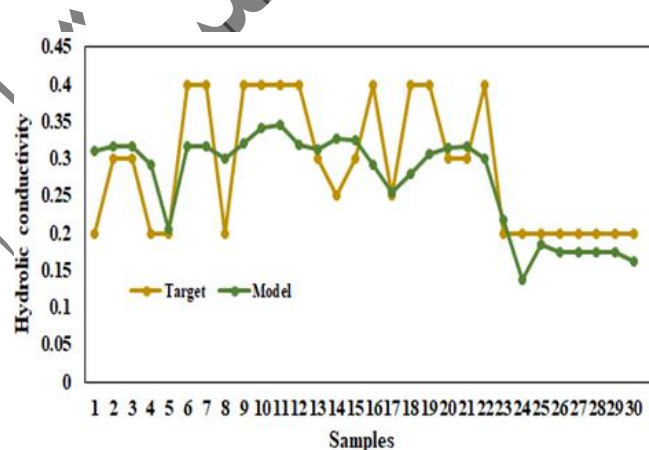


Fig 5- Time series of hydraulic conductivity of well CWI055 in the most optimal of running model

شکل ۵- سری زمانی هدایت هیدرولیکی چاه CWI055 در بهینه‌ترین اجرای مدل

این عدم موفقیت عمدتاً ناشی از یکنواختی بافت رسوبات و محدود بودن دامنه تغییرات این پارامتر در آبخوان است. به بیان دقیق‌تر، از آنجا که مدل‌های داده‌محور برای استخراج الگو، نیازمند پراکندگی (Variance) کافی در داده‌ها هستند، واریانس بسیار اندک پارامتر هدایت هیدرولیکی در منطقه، مانع از یادگیری معنادار مدل GEP شده است. این یافته روش‌شناختی بیانگر آن است که کارایی

مدل GEP در تخمین پارامترهایی با تغییرات مکانی-زمانی محدود، با چالش مواجه خواهد بود و پیشنهاد می‌شود در چنین شرایطی از داده‌های با وضوح مکانی بالاتر یا روش‌های جایگزین استفاده گردد.

ارزیابی مدل GEP در برآورد ارتفاع LNAPL

به منظور برآورد ارتفاع سطح LNAPL با استفاده از مدل برنامه‌ریزی بیان ژن (GEP)، پارامترهای هدایت هیدرولیکی، ضخامت LNAPL و ارتفاع سطح ایستابی به عنوان متغیرهای ورودی و ارتفاع سطح LNAPL به عنوان متغیر خروجی تعریف شدند. عملگرهای ژنتیکی به کار رفته در این مدل‌ها در جدول ۸ ارائه شده است. در این بخش نیز، مشابه مدل‌های توسعه یافته پیشین، تعداد کروموزوم‌ها ۳۰، طول سر کروموزوم ۸ و تعداد ژن‌ها ۳ در نظر گرفته شد. همچنین، برای ساخت روابط ریاضی مدل از مجموعه چهار عملگر اصلی شامل جمع، تفریق، ضرب و تقسیم استفاده گردید. جدول ۹ شاخص‌های آماری مربوط به میزان انطباق بین داده‌های تخمین زده شده توسط مدل و داده‌های محاسباتی را نشان می‌دهد. بر اساس نتایج، کمترین مقدار ضریب تعیین R^2 در این مدل‌ها ۰/۹۹ و بیشترین مقدار آن برابر با ۱ به دست آمد که بیانگر برازش بسیار مطلوب و همخوانی بالای داده‌های محاسباتی و مشاهداتی است. با توجه به این شاخص‌ها، می‌توان نتیجه گرفت که مدل GEP از توانایی بسیار مناسبی در برآورد و شبیه‌سازی پارامتر ارتفاع سطح LNAPL برخوردار است. همچنین، بررسی شکل ۶ نشان می‌دهد که مقادیر برآورد شده ارتفاع سطح LNAPL توسط مدل GEP، تطابق بسیار خوبی با مقادیر اندازه‌گیری شده داشته و از همبستگی بالایی برخوردار است. معادلات نهایی به دست آمده برای پیش‌بینی ارتفاع سطح LNAPL در چاه CWI001 در جدول ۱۰ ارائه شده است. این معادلات از ترکیب روابط استخراج شده برای سه ژن تعریف شده در مرحله تنظیم اولیه مدل حاصل شده‌اند.

Table 8- Values of genetic operators used in the GEP model

جدول ۸- مقادیر اپراتورهای ژنتیکی به کار رفته در مدل GEP

Well	Parameter Description	Set of Functions	Chromosome Number	Chromosome Head Size	Gene Number	Mutation Rate	RNC Rate	One-Point Crossover	Two-Point Crossover	Gene Crossover	Gene Translocation	IS Translocation
CWI001	Parameter setting	+, -, ×, /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005
CWI055	Parameter setting	+, -, ×, /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005
CWI071	Parameter setting	+, -, ×, /	30	8	3	0.001	0.002	0.002	0.002	0.002	0.002	0.005

Table 9- Results of selecting different fitting functions in the CWI001 well model

جدول ۹- نتایج تابع‌های برازش مختلف در مدل چاه CWI001

	CWI055		CWI055		CWI071	
Function	Training	Test	Training	Test	Training	Test
RMSE	0.0046	0.0046	0	0.00004	0.0006	0.0007
R ²	1	1	0.999	1	0.999	1
Best Fitness	995.4	995.4	0.999	1000	999.357	966.029

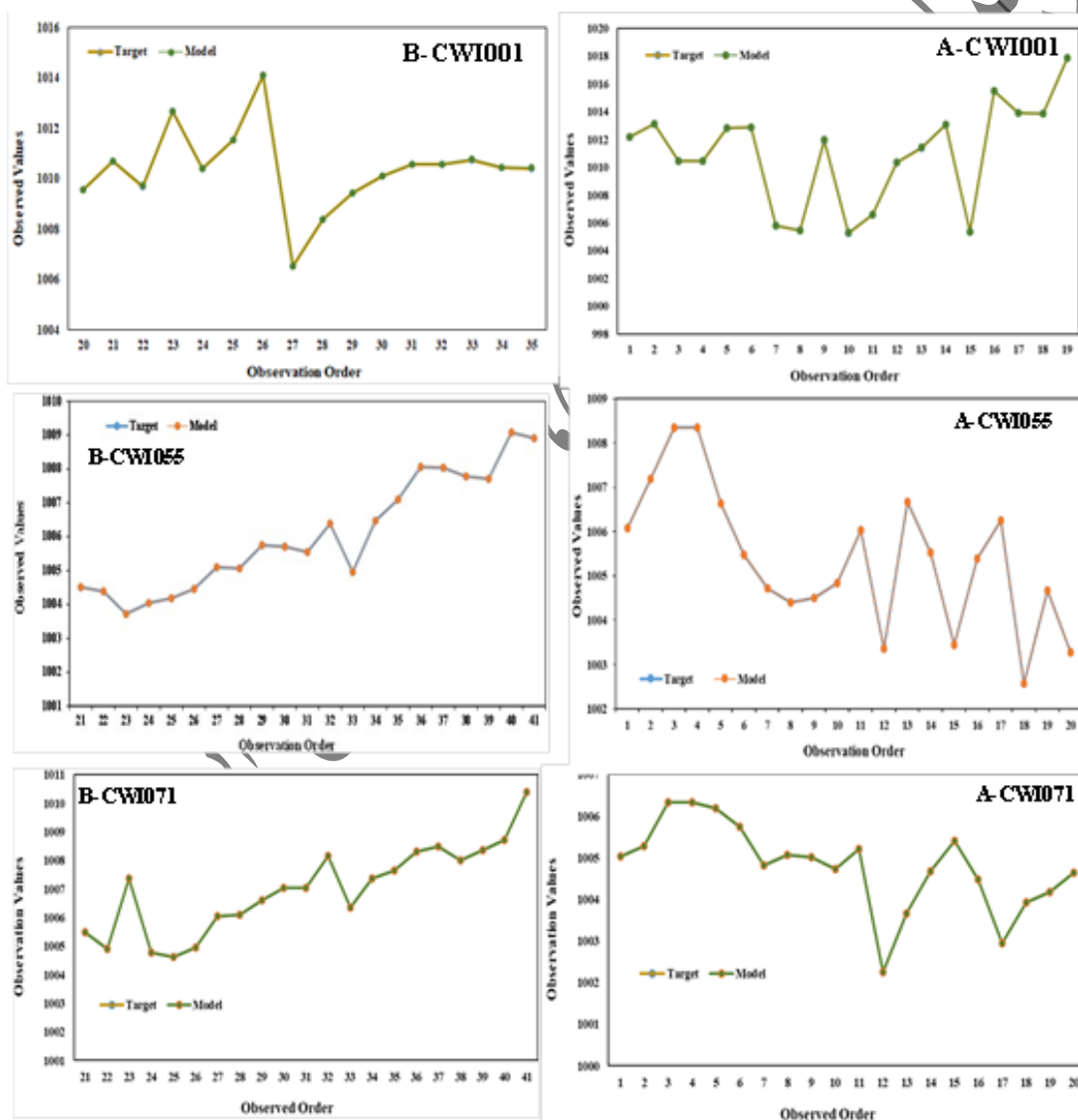


Fig 6- Time series graph of LNAPL elevation in the most optimal running models

شکل ۶- سری زمانی ارتفاع LNAPL در بهینه‌ترین اجرای مدل‌ها

Table 10- Relationships obtained from each subtree to predict the transfer level of wells

جدول ۱۰- روابط حاصل از هر زیر درخت برای پیش‌بینی سطح چاه‌ها

Well	Equation
	LNAPL LEVEL=
CWI001	$d(OT)+d(WT)+((((d(K)\times(-8.7))-8.88)\times(-2.53\times d(K)))\times(\frac{-1.8}{8.02}\times(-2.53+d(K))))$
	LNAPL LEVEL=
CWI055	$\left(d(OT)+\left(d(K)\times\left(\frac{d(k)}{(d(WT)-d(OT))+d(WT)}\right)\right)\right)+\left(\frac{d(ot)}{((d(WT)-d(K))-d(WT))\times(-8.62)}\times d(K)\right)+d(WT)$
	LNAPL LEVEL=
CWI071	$\left(\left(\left(\frac{3.3}{d(WT)}\right)\right)-(-0.7)\right)+((d(OT)-2.9)))+25.67-(d(WT)-\left(\frac{d(WT)}{\frac{d(K)}{\frac{d(WT)}{d(K)}}-d(WT)}\right)-\left(\frac{-0.51}{d(WT)}\right))$

بر اساس محاسبه‌ها و مدل‌سازی‌های انجام شده و با توجه به نتایج (شکل‌های ۷ و ۸) مدل‌سازی پارامترهای مورد مطالعه برای سه چاه موجود در منطقه، مدل GEP دارای توانایی قابل توجهی در تخمین تراز سطح LNAPL و در مرتبه بعدی قابلیت انتقال LNAPL بوده و نتایج قابل اطمینانی برای این پارامترها ارائه می‌دهد.

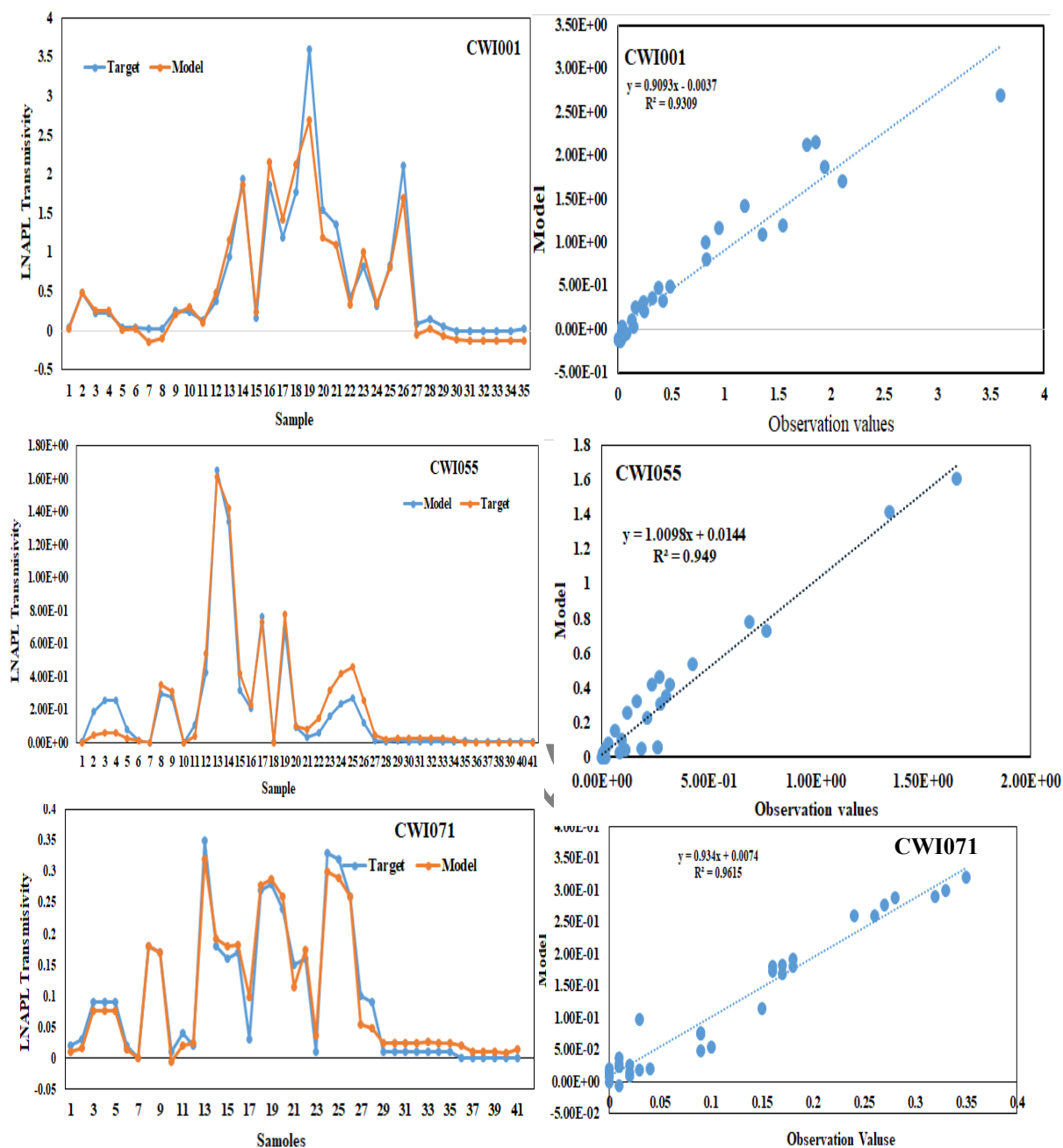


Fig 7- Time series graphs of LNAPL transfer capability of wells in the most optimal execution

شکل ۷- نمودارهای سری زمانی قابلیت انتقال LNAPL بهینه ترین حالت

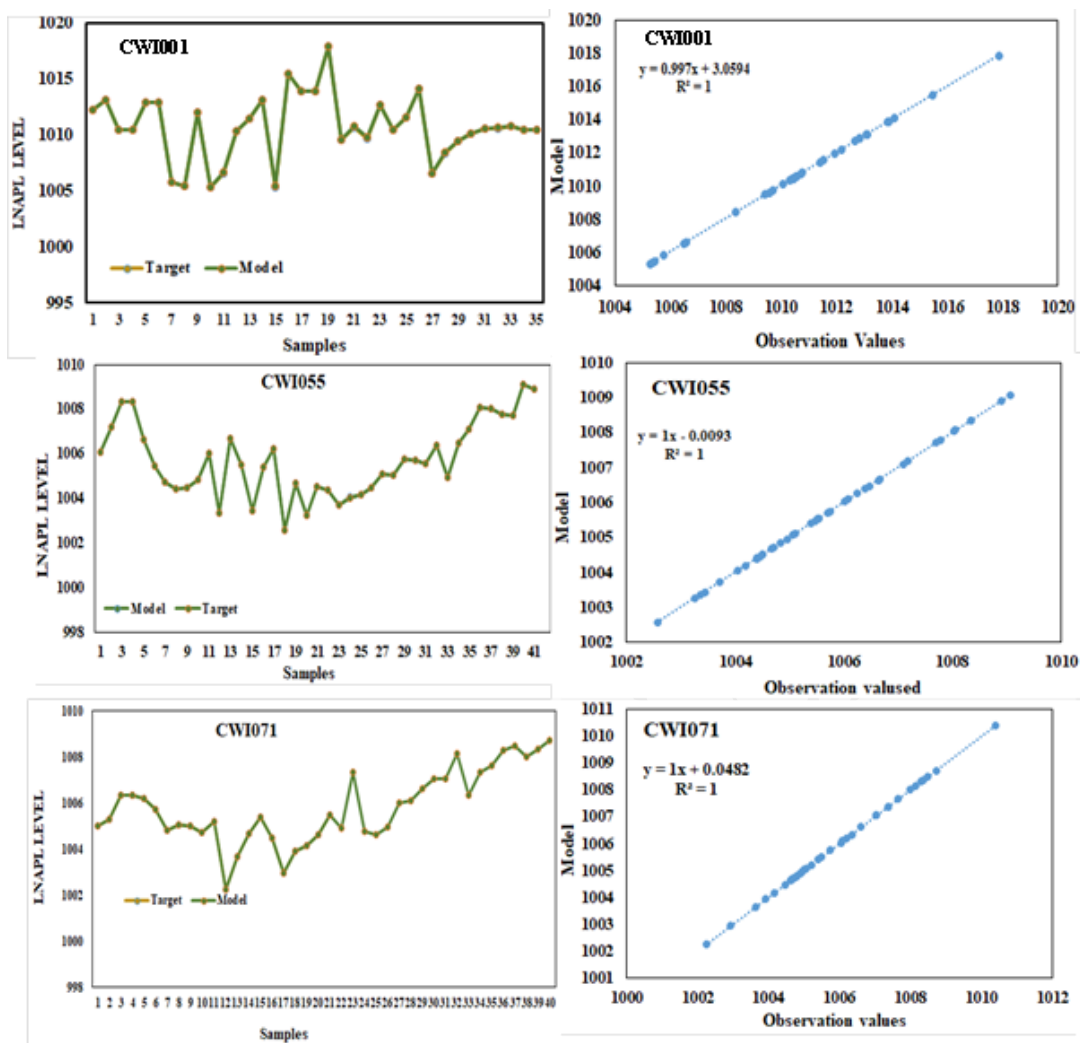


Fig 8- Time series graphs of LNAPL level in the most optimal execution mode

شکل ۸- نمودارهای سری زمانی سطح LNAPL بهینه‌ترین اجرای مدل

مجموعه یافته‌های این پژوهش نشان می‌دهد که مدل GEP می‌تواند به عنوان یک ابزار کمکی و مکمل در کنار روش‌های سنتی صحرایی در آبخوان‌های صنعتی مورد استفاده قرار گیرد. موفقیت این مدل در تخمین پارامترهای پویا و وابسته به زمان مانند تراز سطح LNAPL و قابلیت انتقال، آن را به گزینه‌ای مناسب برای شبیه‌سازی رفتار کوتاه‌مدت و میان‌مدت آلودگی تبدیل می‌کند. در مقابل، ناکامی مدل در تخمین پارامترهای ایستا مانند هدایت هیدرولیکی که وابستگی شدیدی به بافت زمین‌شناسی دارند، نشان‌دهنده لزوم توجه به ماهیت پارامترها در انتخاب روش مدل‌سازی است. این امر به مدیران محیط‌زیست کمک می‌کند تا با دیدگاه بازتری از نتایج مدل‌های هوش مصنوعی استفاده نمایند و آن‌ها را با داده‌های میدانی تکمیل کنند.

محدودیت‌های پژوهش

این پژوهش با وجود ارائه نتایج قابل اتکا برای برخی پارامترهای هدف، دارای محدودیت‌هایی است که می‌تواند در تفسیر نتایج و تعمیم‌پذیری آن‌ها اثرگذار باشد:

محدودیت دامنه تغییرات هدایت هیدرولیکی (K): در داده‌های مورد استفاده، مقدار K در چاه‌ها تقریباً ثابت یا با تغییرات بسیار محدود گزارش شده است. از آنجا که مدل‌های داده‌محور برای یادگیری الگو نیازمند تغییرپذیری کافی در داده‌ها هستند، یکنواختی K باعث کاهش توان مدل در استخراج رابطه معنادار شده و در نهایت معادله به‌دست‌آمده برای K فاقد اعتبار لازم تشخیص داده شد. تراز LNAPL معمولاً ارتباط مستقیم‌تر و پویاتر با متغیرهای زمانی مورد استفاده در مدل، مانند ضخامت LNAPL و نوسانات سطح آب، دارد؛ در حالی که هدایت هیدرولیکی بیشتر تابع ویژگی‌های ذاتی محیط متخلخل، بافت رسوبات و ناهمگنی‌های موضعی آبخوان است و لزوماً از طریق داده‌های زمانی محدود، با دقت بالا قابل استنتاج نیست.

وابستگی بخشی از داده‌ها به خروجی مدل LDRM: در این مطالعه، برخی پارامترها (از جمله قابلیت انتقال) از مدل LDRM استخراج شده‌اند. بنابراین، بخشی از عدم قطعیت نتایج می‌تواند ناشی از عدم قطعیت‌ها و مفروضات مدل پایه (LDRM) باشد و لزوماً فقط به عملکرد GEP مربوط نباشد.

محدودیت در تنوع متغیرهای ورودی و اطلاعات زمین‌شناسی: متغیرهای ورودی عمدتاً شامل ضخامت LNAPL، تراز آب، تراز LNAPL، قابلیت انتقال و هدایت هیدرولیکی بوده‌اند. نبود برخی متغیرهای مؤثر مانند ویژگی‌های دانه‌بندی/لایه‌بندی، تخلخل مؤثر، ناهمسانگردی، تغییرات مکانی لیتولوژی، نرخ برداشت/پمپاژ و شرایط مرزی می‌تواند باعث شود روابط استخراج‌شده، همه عوامل فیزیکی کنترل‌کننده را پوشش ندهند.

وابستگی نتایج به تنظیمات و فرآیند سعی و خطا (Feature Selection و Topology): مطابق یافته‌های مطالعه، عملکرد مدل به شدت به انتخاب ترکیب ورودی‌ها و تنظیمات معماری (طول سر، تعداد ژن‌ها، عملگرها و نرخ‌های ژنتیکی) وابسته است؛ لذا امکان دارد با تنظیمات دیگر، روابط متفاوتی به‌دست آید. این موضوع می‌تواند بازتولیدپذیری نتایج را بدون گزارش کامل تنظیمات و روند کالیبراسیون دشوارتر کند.

خلاصه و جمع‌بندی

این پژوهش نشان داد که مدل برنامه‌نویسی بیان ژن (GEP) به‌عنوان یک روش هوش مصنوعی کارآمد، قادر است پارامترهای هیدرودینامیکی متغیر با زمان (مانند تراز سطح LNAPL و قابلیت انتقال) را با دقت بسیار بالا ($R^2 > 0.99$) در آبخوان صنعتی ری شبیه‌سازی کند. این موفقیت، پتانسیل بالای GEP را در کاهش هزینه‌ها و زمان مطالعات پایش و پاکسازی آشکار می‌سازد. با این حال، کارایی این روش در تخمین پارامترهای ساختاری آبخوان که تغییرپذیری مکانی اندکی دارند (مانند هدایت هیدرولیکی) به‌طور قابل توجهی کاهش می‌یابد؛ یافته‌ای که نشان می‌دهد مدل‌های داده‌محور نمی‌توانند به‌طور کامل جایگزین روش‌های صحرایی شده

و باید به عنوان ابزاری مکمل در کنار آن‌ها به کار روند. نتایج همچنین تأیید نمود که دقت و کارایی مدل GEP به شدت به پیکربندی و ترکیب متغیرهای ورودی وابسته است؛ به گونه‌ای که کوچک‌ترین تغییر در ساختار ورودی‌ها می‌تواند به نتایج متفاوتی منجر شود. از این رو، دست‌یابی به بهینه‌ترین توپولوژی مدل، مستلزم انجام فرآیند سعی و خطا در انتخاب ترکیب بهینه ورودی‌ها (Feature Selection) است تا تعادل میان پیچیدگی مدل و دقت خروجی‌ها برقرار گردد.

تشکر

بدین وسیله از پژوهشکده علوم پایه کاربردی جهاد دانشگاهی واحد شهید بهشتی و شرکت پالایش نفت تهران جهت همکاری و حمایت‌های ارزشمند در اجرای این پژوهش کمال تشکر و قدردانی را داریم.

منابع

- Abareshi F, Meftah Halghi M, Sanikhani H, Dehghani AA (2014) Comparison of three intelligence techniques for predicting water table depth fluctuations (Case study: Zarringol plain). Journal of Water and Soil Conservation, Vol. 21(1), 163-180.
- Charbeneau RJ, Johns RT, Lake LW, McAdams III, MJ (2000) Free-product recovery of petroleum hydrocarbon liquids. Groundwater Monitoring & Remediation, 20(3), pp.147-158.
- Chen Y, Li B, Singh VP, et al (2021) Machine learning approaches for groundwater level prediction: A review, Journal of Hydrology, 603, 126900.
- Kisi O, Parmar KS, Soni K (2021) Evolutionary computing techniques in groundwater modeling and prediction: State-of-the-art review, Water, 13(9), 1234.
- Kumar A, Yadav S, Singh R (2024) Data-driven modeling of LNAPL migration in heterogeneous aquifers using machine learning algorithms, Journal of Contaminant Hydrology, 260, 104273.
- Lari KS, Johnston CD, Rayner JL, Davis GB (2018) Field-scale multi-phase LNAPL remediation: Validating a new computational framework against sequential field pilot trials. Journal of hazardous
- Mercer JW, Cohen RM (1990) A review of immiscible fluids in the subsurface: Properties, models, characterization and remediation. Journal of contaminant hydrology, 6(2), pp.107-163.
- Mahmood W, Mohammed AS, Asteris PG, Ahmed H (2023) Soft computing technics to predict the early-age compressive strength of flowable ordinary Portland cement: W. Mahmood et al. Soft Computing, 27(6), pp.3133-3150.

Prasad KL, Rastogi AK (2001) Estimating net aquifer recharge and zonal hydraulic conductivity values for Mahi Right Bank Canal project area, India by genetic algorithm. *Journal of Hydrology*, 243(3-4), pp.149-161.

Rajaei T, Ebrahimi H, Nourani V (2019) A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, 572, pp.336-351.

Sadeghi Aghdam F, Nadiri AA, Asghari Moghaddam A, Armanfar F (2017) Application of Genetic Expression Programming (GEP) to produce Intelligence Committee Fuzzy Logic model (CFLM) to predict arsenic concentration in water resources of the Sahand Dam Basin. *Iranian Water Research Journal*, 11(4), pp.85-96.

Salehi M, Abdolazadeh M, Javadi AA (2023) Hybrid AI models for estimation of hydraulic conductivity in contaminated aquifers. *Hydrogeology Journal*, 31, pp.1871–1888.

Shiri J, Kisi O (2011) Comparison of genetic programming with neurofuzzy systems for predicting short-term water table depth fluctuations. *Computers & Geosciences*, 37, pp.1692-1701.

Tomlinson DW, Rivett MO, Wealthall GP, Sweeney REH (2017) Understanding complex LNAPL sites: Illustrated handbook of LNAPL transport and fate in the subsurface. *Journal of Environmental Management*, 204(Pt 2), pp.748-756.

Traore S, Guven A (2012) Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. *Water Resources Management*, 26, pp.4367-4380.

Estimation of LNAPL hydrodynamic parameters in the Rey Plain aquifer using gene expression programming modeling

Kamal Khodaei¹, Zahra Boosalik^{*2}, Aliakbar Shahsavari³, Naeme Faezi⁴, Benyamin Rezazadeh⁴

¹ Associate Professor, Department of Environmental Geology, Research Institute for Applied Basic Sciences, Shahid Beheshti ACECR.

² Assistant Professor, Department of Environmental Geology, Research Institute for Applied Basic Sciences, Shahid Beheshti ACECR. Email: zahra.boosalik@yahoo.com

⁴ Associate Professor, Department of Environmental Geology, Research Institute for Applied Basic Sciences, Shahid Beheshti ACECR.

³ Department of Environmental Geology, Research Institute for Applied Basic Sciences, Shahid Beheshti ACECR.

Abstract

Estimating the hydrodynamic parameters of Light Non-Aqueous Phase Liquid (LNAPL) is a fundamental pillar of contaminant transport studies in hydrocarbon-contaminated aquifers and a key factor in the design and management of remediation projects. However, direct measurement of these parameters through field and laboratory methods is typically costly, time-consuming, and often constrained by operational limitations. In this study, the efficacy of the Gene Expression Programming (GEP) model in estimating LNAPL hydrodynamic parameters within the Ray industrial zone was evaluated. Monthly data regarding LNAPL thickness, groundwater levels, and LNAPL levels and transmissivity—derived from the LNAPL Distribution and Recovery Model (LDRM)—were collected for three representative wells over the statistical period of 2006–2020. Modeling was conducted using GeneXproTools 5 software, and performance was assessed using R2 and RMSE metrics. Results indicated that the model demonstrates high precision in estimating LNAPL levels and transmissivity, with the coefficient of determination (R2) exceeding 0.99 in some instances. Conversely, the model's performance in estimating hydraulic conductivity was suboptimal, which is attributed to the limited range of variations in this parameter and the relative uniformity of the aquifer sediments. The innovation of this research lies in applying the GEP method to derive explicit, practical mathematical relationships for estimating hydrodynamic parameters in the Ray aquifer. While highlighting the high efficacy of the GEP model in estimating dynamic parameters, the findings underscore the necessity of accounting for input data limitations in data-driven modeling.

Keywords: Contaminated aquifers, Hydrodynamic parameters, LNAPL, GEP model.