

## کاربرد برنامه‌ریزی بیان ژن (GEP) برای ایجاد مدل ترکیبی فازی (CFLM) به منظور پیش‌بینی غلظت آرسنیک منابع آبی در حوضه آبریز سد سهند

فریبا صادقی‌اقدم<sup>۱\*</sup>، عطاالله ندیری<sup>۲</sup>، اصغر اصغری‌مقدم<sup>۳</sup> و فریدون آرمانفر<sup>۴</sup>

### چکیده

مصرف آب‌های زیرزمینی آلوده به آرسنیک، منجر به بروز بیماری‌های متعدد و مرگ انسان‌ها می‌شود. در صورت آلودگی غیرنقطه‌ای و متأثر از زمین‌شناسی آب‌ها، به سادگی نمی‌توان مانع گسترش آن شد؛ لذا این نوع آلودگی‌ها می‌بایست به دقت بررسی شوند. گزارش‌های متعدد اخیر در منطقه مطالعاتی سد سهند، حاکی از وجود آنومالی آرسنیک با مقادیر بیش از استاندارد WHO (۰/۰۱ mg/L) است. با توجه به تحقیقات قبلی، مبنی بر مناسب نبودن مدل‌های خطی زمین‌آماری برای پیش‌بینی غلظت آرسنیک کل (III, V) در منطقه، از مدل‌های هوش مصنوعی همچون برنامه‌ریزی بیان ژنتیک (GEP) و منطق فازی استفاده شد که با الهام از طبیعت قادر به تخمین پارامترهای پدیده‌های طبیعی با دقت قابل توجهی نسبت به سایر روش‌ها هستند. برای تخمین غلظت‌های آرسنیک کل، از پارامترهای pH، سولفات، نیترات، فلئوئورید، آهن و آرسنیک نمونه‌ها به‌عنوان ورودی مدل‌های فازی ممدانی (MFL)، لارسن (LFL) و سوگنو (SFL) استفاده شد. با توجه به مناسب بودن و تشابه نتایج سه مدل فازی و به منظور استفاده همزمان مزایای هر سه مدل، از برنامه‌ریزی بیان ژنتیک برای تولید مدل ترکیبی نتایج سه مدل منفرد فازی استفاده شد. با توجه به مزایای برنامه‌ریزی بیان ژنتیک و نتایج مراحل آموزش و آزمایش مدل که به ترتیب با ضریب تبیین  $R^2$  برابر ۰/۹۶۷ و ۰/۹۲۴ و مقادیر RMSE برابر ۰/۰۷۲ و ۰/۰۹۶ است، مدل مذکور قادر به ارائه مدل ترکیبی با دقت بیشتری از سه مدل منفرد فازی ارائه شده است.

**واژه‌های کلیدی:** آرسنیک، برنامه‌ریزی بیان ژنتیک، سد سهند، فازی، مدل ترکیبی، هوش مصنوعی.

**ارجاع:** صادقی‌اقدم ف. ندیری ع. اصغری‌مقدم ا. و آرمانفر ف. کاربرد برنامه‌ریزی بیان ژن (GEP) برای ایجاد مدل ترکیبی فازی (CFLM) به منظور پیش‌بینی غلظت آرسنیک منابع آبی در حوضه آبریز سد سهند. مجله پژوهش آب ایران. ۲۷: ۸۵-۹۶.

۱- دانشجوی دکتری هیدروژئولوژی، گروه علوم زمین، دانشکده علوم طبیعی، دانشگاه تبریز.

۲- استادیار گروه علوم زمین، دانشکده علوم طبیعی، دانشگاه تبریز.

۳- استاد گروه علوم زمین، دانشکده علوم طبیعی، دانشگاه تبریز.

۴- کارشناس ارشد مهندسی محیط‌زیست، سازمان آب منطقه‌ای آذربایجان شرقی، تبریز.

\* نویسنده مسئول: [Fariba\\_Sadeghi@tabrizu.ac.ir](mailto:Fariba_Sadeghi@tabrizu.ac.ir)

تاریخ دریافت: ۱۳۹۵/۱۰/۰۵ تاریخ پذیرش: ۱۳۹۵/۰۶/۲۸

## مقدمه

شناسایی و پایش کیفیت منابع آبی سدها به عنوان یکی از گام‌های اصلی مدیریت کیفیت منابع آب، جایگاه ویژه‌ای دارد. امروزه اکثر آب‌های طبیعی در معرض آلودگی قرار گرفته یا آلوده شده‌اند؛ از این‌رو لازمه هرگونه اقدام برای کنترل و کاهش آلاینده‌های آبی و تأثیرات آن‌ها، آگاهی کامل از نحوه توزیع و پراکندگی آلاینده‌های موجود در سطح بوده که در اختیار داشتن چنین اطلاعاتی فقط از طریق ایستگاه‌های سنجش آلودگی توزیع یافته در سطح منطقه مورد مطالعه و انجام آنالیزهای مختلف امکان‌پذیر است.

عناصر سنگین با غلظت اندک در سنگ، خاک، آب، گیاهان، هوا و حتی در بدن انسان یافت می‌شوند. آرسنیک به دلیل سمی بودن بسیار بالا به عنوان یکی از مهم‌ترین آلاینده‌ها محسوب می‌شود (دراگون، ۲۰۰۶) که در اثر پدیده‌های انسان‌زاد (باندسچو و همکاران، ۲۰۱۱)، زمین‌زاد (چیپیان و همکاران، ۲۰۱۲) و زیست‌زاد (ماهیماراجا و همکاران، ۲۰۰۵)، غلظت آن افزایش یابد و باعث آلودگی آب‌های سطحی و زیرزمینی می‌شود. آلودگی طبیعی آب‌ها، ناشی از زمین‌شناسی را به سادگی نمی‌توان از بین برد و یا مانع گسترش آن شد؛ لذا این نوع آلودگی‌ها می‌بایست به دقت بررسی شوند (سو و همکاران، ۲۰۱۶) و در صورت وجود آنومالی‌های<sup>۱</sup> خطرناک آلوده‌کننده در آب‌ها، در چگونگی مصرف آن تجدید نظر صورت گیرد؛ زیرا جان عده بی‌شماری در جهان به دلیل مصرف آب آلوده در معرض خطر قرار می‌گیرد. مصرف آب آلوده منجر به بروز بیماری‌های متعدد از جمله انواع زخم‌های پوستی، مشکلات کارکرد اندام‌های داخلی، سرطان، اختلال در DNA و RNA و مرگ انسان می‌شود (انور و همکاران، ۲۰۱۱). به عنوان مثال در طی چندین سال اخیر غلظت آرسنیک تأثیرات سوئی بر سلامت مردم شهر هشتروود و روستاهای تابعه داشته است (مسافری و همکاران، ۲۰۰۸).

مطالعات پیشین در منطقه سد سهند شامل مطالعات هیدروشیمیایی و ژئوشیمیایی (ندیری و همکاران، ۱۳۹۰؛ ۲۰۱۳ الف؛ صادقی‌اقدم، ۱۳۹۱) منجر به شناسایی آنومالی آرسنیک و مکان‌های مرتبط با آن شد. با استفاده از روش‌های آماری و مدل‌سازی زمین‌آماري اقدام به

برآورد مقادیر آرسنیک در منطقه گردید که نتایج بسیار دقیقی از مدل‌های خطی زمین‌آماري به دست نیامد (صادقی‌اقدم، ۱۳۹۱). دلیل این امر را می‌توان در آبخوان‌های تکه‌تکه سطحی و غیرمرتبط با هم دانست که می‌توان گفت روند عمومی نامشخص و غیرخطی را دنبال می‌کنند. همچنین با توجه به برخی محدودیت‌های مدل‌های عددی و فیزیکی مانند وقت‌گیر بودن، پیچیده بودن، پرهزینه بودن و نیاز داشتن به داده‌های زیاد، روش‌های دیگری مانند روش‌های مختلف هوش مصنوعی برای پیش‌بینی پارامترهای هیدروژئولوژیکی و هیدروژئوشیمیایی مانند آرسنیک ارائه شده است.

مدل‌های هوشمند در مطالعات متعددی به منظور داده‌کاوی یا تخمین پارامترهای کیفی مورد استفاده قرار می‌گیرد (ستاری و همکاران، ۲۰۱۶؛ مندل و همکاران، ۲۰۱۵؛ فلاح و همکاران، ۲۰۱۳). مطالعات فراوانی نیز درباره پیش‌بینی پارامترهای هیدروژئولوژیکی با استفاده از مدل‌های هوش مصنوعی از جمله فازی انجام شده است (الاتونجی و همکاران، ۲۰۱۱) که به عنوان یک تقریب‌گر جهانی می‌توانند در پیش‌بینی بررسی مسائل غیرخطی مفید باشند (اصغری مقدم و همکاران، ۱۳۸۸). در پیش‌بینی پارامترها عموماً بیش از یک مدل هوش مصنوعی قابل استفاده است که می‌توانند نتایج مشابهی حاصل کنند. هر کدام از مدل‌ها دارای عدم قطعیت و مزایای مربوط به خود هستند و نتایج هر یک از مدل‌ها در برخی نقاط بر مدل دیگر ارجحیت دارند و انتخاب یک مدل به عنوان مدل برتر منطقی به نظر نمی‌رسد؛ لذا بهترین کار استفاده هم‌زمان از هر سه مدل برای تجمیع مزایای آن‌ها است (چن و لین، ۲۰۰۶). همچنین ندیری و همکاران (۲۰۱۳ب) توصیه کرده‌اند که به جای مدل خطی هوش مصنوعی، می‌توان از تجمیع مدل‌های هوش مصنوعی استفاده کرد. به این ترتیب مدل‌های چندگانه برای پیش‌بینی چنین پارامترهایی توسعه یافته‌اند (کدخدایی و همکاران، ۲۰۰۹؛ ندیری و همکاران، ۲۰۱۴، ۲۰۱۳ ب؛ تایفور و همکاران، ۲۰۱۴؛ ندیری و همکاران، ۲۰۱۵). اساس تئوری این روش استفاده ترکیبی از نتایج هر یک از روش‌ها برای رسیدن به یک نتیجه کلی و بهتر است. به این ترتیب، می‌توان از مزیت انفرادی هر یک از این مدل‌ها بهره برد (لبنی و همکاران، ۲۰۱۰).

خزر می‌باشد. اقلیم منطقه با استفاده از تقسیم‌بندی اقلیم‌نمای آمبرژه و با توجه به داده‌های آب و هوایی با متوسط میزان بارش ۳۵۰ میلی‌متر و متوسط درجه‌حرارت سالانه  $+11^{\circ}\text{C}$ ، جزو مناطق نیمه‌خشک سرد است. این منطقه از لحاظ زمین‌شناسی مشابه با زون البرز-آذربایجان (آقناتی، ۱۳۸۳) و جزو مناطق آتشفشانی-رسوبی جوان است. سد سهند در پایین‌دست محل تلاقی دو رودخانه آملالو و قرنقو احداث گردیده است. این سد با حجم مخزنی ۱۶۵ میلیون مترمکعب، تأمین‌کننده منابع آب کشاورزی، صنعت و شرب منطقه می‌باشد.

در محدوده مطالعاتی سازندهای به سن میوسن تا کواترنری وجود دارد. مهم‌ترین سازندهای منطقه مطابق با شکل ۱ شامل نهشته‌های قرمز زیرین و فوقانی و نهشته‌های آذرآواری و گدازه‌های توده آتشفشانی سهند هستند که کیفیت آب‌های منطقه را تحت تأثیر قرار می‌دهند. مطالعات و بررسی‌های صحرایی منجر به شناسایی چندین گسلش در سازندهای میوسن شد که این گسل‌ها آب‌های زیرزمینی و سطحی را به یکدیگر مرتبط می‌کنند. طی مطالعه مقاطع نازک و صیقلی سولفیدهای آرسنیک در شکستگی‌ها و درز و شکاف‌های سنگ آهک، مارل و کنگلومرا ماسه‌سنگ و سازند قرمز مشاهده شد و بیشترین غلظت‌های آرسنیک در نمونه‌های آنالیز شده از آب چشمه‌های مرتبط با گسل‌ها شناسایی گردید. تغلیظ اپی‌ترمال عنصر آرسنیک در شکستگی‌های سازند میوسن منشأ عمده این آنومالی شناسایی شد (ندیری و همکاران، ۱۳۹۰؛ ۲۰۱۳ ب).

از نظر هیدروژئولوژیکی در محدوده مطالعاتی سفره آب زیرزمینی به‌صورت وسیع و یکپارچه نبوده و از تعدادی سفره‌های ناهمگن پچ‌مانند تشکیل شده است. شرایط هیدروژئولوژیکی محدوده مطالعاتی در ارتباط مستقیم با سازندهای زمین‌شناسی و زمین‌ساخت منطقه است. آب‌های سطحی و زیرزمینی منطقه مرتبط با یکدیگر هستند، به‌طوری که آب‌های سطحی ناشی از بارش جوی و ذوب برف دامنه‌های سهند ضمن نفوذ در زمین به‌صورت جریانات زیرسطحی به آب‌های زیرزمینی می‌پیوندند و آب‌های زیرزمینی در بسیاری از مناطق به واسطه زمین‌ساخت منطقه و از طریق چشمه‌ها به سطح راه می‌یابند.

برنامه‌ریزی ژنتیک (GP)<sup>۱</sup> از جمله روش‌های الهام گرفته از طبیعت است که در حل مسائل دقیق و پیچیده از آن استفاده می‌شود. تا به حال مطالعات مختلفی با استفاده از روش‌های هوش مصنوعی صورت گرفته که برتری روش‌های GP را به اثبات رسانده است. پژوهشگران زیادی از GP برای مطالعه فرآیندهای مختلف از جمله پیش‌بینی و شبیه‌سازی سطح آب زیرزمینی (فلاح و همکاران، ۲۰۱۳)، حمل رسوب رودخانه (آیتک و کیشی، ۲۰۰۸)، تخمین داده‌های ناقص (استوریکار و دئو، ۲۰۰۸)، تعیین هیدروگراف واحد (رابونال و همکاران، ۲۰۰۷)، تعیین دبی روزانه (گوون، ۲۰۰۹)، پیش‌بینی جریان (شعیب و همکاران، ۲۰۱۵)، مدل‌سازی فرآیند بارش- رواناب (جایوردنا و همکاران، ۲۰۰۵) و پیش‌بینی کوتاه‌مدت و بلندمدت دمای هوا (کیشی و همکاران، ۲۰۱۱)، تخمین تبخیر و تعرق مرجع در اقلیم خشک (یاسین و همکاران، ۲۰۱۶) سود جستند.

در این مطالعه از روش‌های فازی ممدانی (MFL)<sup>۲</sup>، لارسن (LFL)<sup>۳</sup> و سوگنو (SFL)<sup>۴</sup> برای پیش‌بینی و بررسی مقادیر آرسنیک کل در منطقه مطالعاتی استفاده شد. با توجه به کارایی تقریباً مشابه این مدل‌ها و به منظور استفاده هم‌زمان از هر سه مدل منفرد، مدل منطق فازی مرکب (CFLM)<sup>۵</sup> با استفاده از برنامه‌ریزی بیان ژنتیک با توجه به کاربرد پارامترهای بهینه در مدل‌سازی ایجاد گردید.

## مواد و روش‌ها

### منطقه مطالعاتی

منطقه مطالعاتی به مساحت حدود ۳۸۴ کیلومتر مربع در استان آذربایجان شرقی، در غرب شهرستان هشترود و در محدوده‌ای به مختصات UTM،  $X=644000$  تا  $X=677000$  طول شرقی و  $Y=4134000$  تا  $Y=4155000$  عرض شمالی واقع شده است. موقعیت مکانی حاصل از پیوند عناصر واحدهای توپوگرافی باعث شکل‌یابی منطقه به صورت چاله‌ای بسته شده است که تنها به سمت شرق و دره قزل‌اوزن باز می‌شود. محدوده مطالعاتی از نظر هیدروژئولوژیکی جزوی از حوضه آبریز سفیدرود دریای

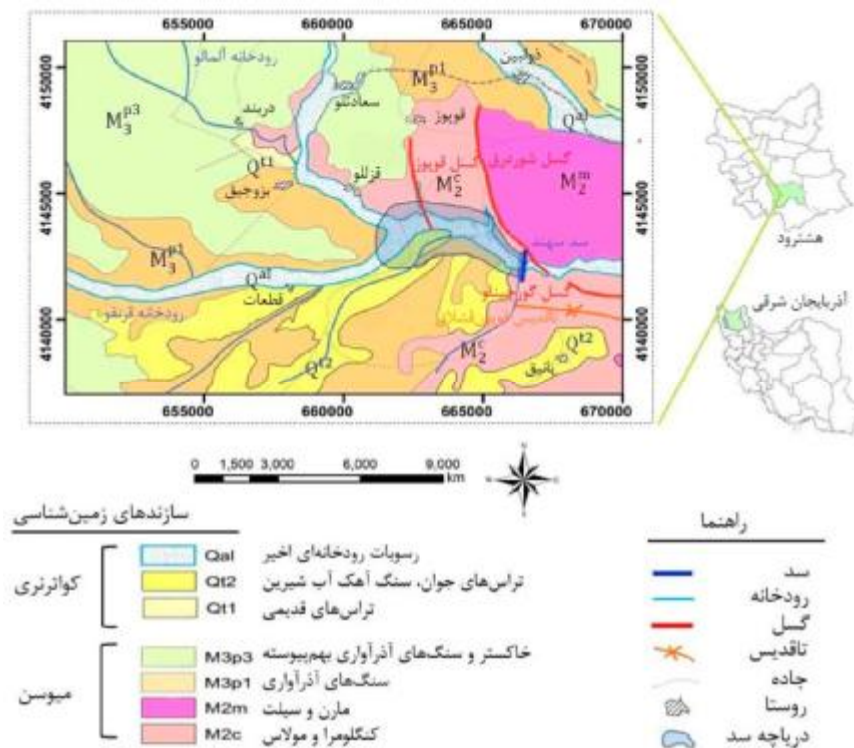
1- Genetic Programming (GP)

2- Mamdani Fuzzy Logic (MFL)

3- Larsen Fuzzy Logic (LFL)

4- Sugeno Fuzzy Logic (SFL)

5- Committee Fuzzy Logic Model (CFLM)



شکل ۱- نقشه زمین‌شناسی محدوده مطالعاتی

سولفات، نیتрат، فلئورید، آهن و آرسنیک است (جدول ۲). بر این اساس عناصر فلئوراید و نیترات همبستگی مثبت و سایر پارامترها همبستگی منفی با عنصر آرسنیک نشان می‌دهند. در مقالات متعددی نیز ارتباط بین آرسنیک و پارامترهای اسیدیته (ویلسون و همکاران، ۲۰۰۸)، سولفات (انور و همکاران، ۲۰۱۱؛ چیبیان و همکاران، ۲۰۱۲)، نیترات (دهقانی و عباس‌نژاد، ۲۰۱۱)، فلئورید (چوهان و بلورا، ۲۰۱۰) و آهن (ارملند و استلز، ۲۰۰۳) مشهود است. تعداد شصت دسته از این داده‌ها در مرحله آموزش و بیست دسته داده در مرحله آزمایش برای تخمین آرسنیک به عنوان ورودی مدل‌های فازی استفاده شد.

### تجزیه و تحلیل داده‌ها

برای پیش‌برد این مطالعه از میان ۶۴۰ دسته داده نمونه‌برداری شده مربوط به عناصر اصلی، فرعی، فلزات سنگین و پارامترهای هدایت‌الکتریکی و pH که طی سال‌های مختلف توسط دانشگاه تبریز (آقایی، ۱۳۸۸؛ صادقی‌اقدم، ۱۳۹۱) در قالب پایان‌نامه کارشناسی ارشد و طرح تحقیقاتی (ندیری، ۱۳۹۰) انجام پذیرفته بود، تعداد ۸۰ دسته داده کامل به منظور اجرای سه مدل فازی ممدانی، لارسن و سوگنو مورد استفاده گرفت که خصوصیات آماری آنها در جدول ۱ آورده شده است. هر دسته داده انتخابی شامل پارامترهایی با بیشترین همبستگی با عنصر آرسنیک ( $<0/3$ ) شامل اسیدیته،

جدول ۱- خصوصیات آماری پارامترهای ورودی در مدل‌های فازی

پارامتر	کشیدگی	کج‌شدگی	واریانس	انحراف معیار	میانگین	بیشینه	کمینه
pH	۲/۹۲	۰/۲۴	۰/۲۵	۰/۵۰	۷/۶۸	۹/۶۸	۶/۳۸
SO <sub>4</sub> <sup>2-</sup> (mg/L)	۱۱/۶۳	۳/۱۹	۵۵۳۷۳۲۲/۵۷	۷۴۴/۱۳	۳۱۷/۸۶	۳۹۰۰/۰۰	۲/۰۰
NO <sub>3</sub> <sup>-</sup> -N(mg/L)	۱۶/۰۶	۳/۷۲	۴۷/۲۷	۶/۸۸	۴/۵۷	۴۲/۸۷	۰/۱۰
F (mg/L)	۳/۷۱	۱/۵۱	۰/۲۲	۰/۴۷	۱/۸۲	۲/۵۷	۰/۰۶
Fe(II, III) (mg/L)	۲۷/۴۶	۵/۰۹	۱۲/۴۰	۳/۵۲	۱/۰۶	۲۱/۳۵	۰/۰۰
As(III, V) (mg/L)	۸/۰۱	۲/۶۱	۰/۱۴	۰/۳۷	۰/۲۲۹	۲/۰۰۹	ND*

\*ND: Not Detected

جدول ۲- همبستگی بین پارامترهای مختلف با استفاده از ضریب همبستگی پیرسون

پارامتر	Pb	Zn	Mn	Al	Fe(II,III)	As(III,V)	SiO <sub>2</sub>	F	PO <sub>4</sub>	NO <sub>3</sub>	SO <sub>4</sub>	Cl	HCO <sub>3</sub>	CO <sub>3</sub>	K	Na	Mg	Ca	pH	EC
EC	-.۰۲	.۰۵	.۰۶	-.۰۲	.۰۶۶	-.۰۳	.۰۱۹	.۰۱۴	.۰۰	-.۰۱	-.۰۳۲	.۰۹۹	.۰۶۳	-.۰۱۳	.۰۵۴	.۰۹۹	.۰۹۷	.۰۹۰	.۰۴۱	۱
pH	.۰۱۲	-.۰۱	-.۰۴۷	-.۰۱	-.۰۲۹	-.۰۴۲	.۰۱۶	-.۰۳۸	-.۰۹	-.۰۲۱	-.۰۲۸	-.۰۴۰	-.۰۶۲	.۰۷۴	-.۰۴۲	-.۰۴۲	-.۰۴۵	-.۰۴۳		۱
Ca	-.۰۱۷	.۰۹۵	.۰۵۰	.۰۱۹	.۰۵۹	.۰۰	-.۰۱۵	.۰۲۲	.۰۰۵	.۰۱۵	.۰۳۹	.۰۸۷	.۰۶۳	-.۰۲۴	.۰۴۰	.۰۸۸	.۰۹۲			۱
Mg	-.۰۱۱	-.۰۰۸	.۰۵۲	-.۰۱۵	.۰۴۹	.۰۶۶	-.۰۱۵	.۰۲۸	.۰۱۰	.۰۱۳	.۰۴۶	.۰۹۳	.۰۴۴	-.۰۲۳	.۰۵۰	.۰۹۶				۱
Na	-.۰۰۲	.۰۰۵	.۰۶۳	-.۰۲۰	.۰۶۹	-.۰۰۳	-.۰۲۰	.۰۱۱	-.۰۱	.۰۱۰	.۰۳۱	.۰۹۹	.۰۷۱	-.۰۱۲	.۰۶۰					۱
K	-.۰۱۰	-.۰۱۵	.۰۶۷	-.۰۲۰	.۰۵۵	.۰۰۷	-.۰۱۵	.۰۱۷	-.۰۰	-.۰۱	.۰۲۸	.۰۵۶	.۰۵۰	.۰۰۱						۱
CO <sub>3</sub>	.۰۰۷	-.۰۱۳	-.۰۲۱	-.۰۱۱	.۰۳۵	-.۰۱۷	-.۰۲۴	-.۰۱۹	-.۰۰۹	-.۰۲۰	-.۰۱۴	-.۰۱۱	-.۰۲۸							۱
HCO <sub>3</sub>	.۰۰۵	.۰۲۳	.۰۳۱	-.۰۰۱	-.۰۱۹	.۰۰۷	-.۰۱۸	.۰۲۲	.۰۰۶	.۰۰۹	.۰۶۱	.۰۵۹								۱
Cl	.۰۰۵	.۰۱۱	.۰۶۸	-.۰۲۲	.۰۷۲	-.۰۰۴	-.۰۲۰	.۰۰۹	-.۰۰۱	.۰۱۱	-.۰۱۹									۱
SO <sub>4</sub>	-.۰۱۱	-.۰۰۷	.۰۲۹	-.۰۱۰	.۰۳۶	-.۰۳۸	-.۰۰۳	.۰۳۱	-.۰۰۱	.۰۰										۱
NO <sub>3</sub>	-.۰۳۱	-.۰۰	-.۰۰۵	.۰۰۶	-.۰۰۴	-.۰۳۲	.۰۱۸	.۰۴۳	.۰۰۲											۱
PO <sub>4</sub>	.۰۰۸	-.۰۰۶	-.۰۳۲	-.۰۰۹	-.۰۰۹	.۰۰	.۰۰۷	-.۰۲۹												۱
F	.۰۱	.۰۹۰	-.۰۰۵	.۰۳۹	-.۰۱۳	-.۰۳۸	.۰۱۶													۱
SiO <sub>2</sub>	-.۰۳۳	-.۰۳۱	.۰۳۰	.۰۰	-.۰۰	.۰۱۹														۱
As(III,V)	.۰۱۰	-.۰۱۶	.۰۲۴	.۰۲۴	-.۰۲۲															۱
Fe(II,III)	-.۰۱۷	-.۰۰	.۰۸۹	.۰۰۸																۱
Al	-.۰۱۶	.۰۰۵	.۰۰۹																	۱
Mn	-.۰۲۰	-.۰۱۰																		۱
Zn	.۰۵۴																			۱
Pb																				۱

## منطق فازی (FL)

دست می‌آید. دو روش دیگر به‌وسیله توابع عضویت تعریف می‌گردند و از طریق روش دسته‌بندی FCM<sup>۱</sup> به‌دست می‌آیند. فهم این نوع سیستم با استفاده از متغیرهای زبانی آسان‌تر است؛ ولی خروجی آن نیاز به فازی‌زدایی دارد. تفاوت روش ممدانی با لارسن نیز در نوع عملگرهای فازی استفاده شده است. عملگر مرحله جمع<sup>۲</sup> هر دو روش max و ولی عملگر مربوط به مرحله استلزام<sup>۳</sup> روش ممدانی (min) and و روش لارسن product است (دمیکو و کلیر، ۲۰۰۴).

در این مطالعه از روش‌های فازی ممدانی، لارسن و سوگنو با استفاده از کدنویسی در محیط نرم‌افزار متلب برای پیش‌بینی و بررسی مقادیر آرسنیک کل استفاده شده است. به‌دلیل مشابهت نتایج سه مدل فازی و مناسب بودن آنها، خروجی مدل‌های فازی از برنامه‌ریزی بیان ژن برای ایجاد مدل ترکیبی بین آن‌ها استفاده شد.

اساس تئوری فازی را برای اولین بار لطفی‌زاده (۱۹۶۵) معرفی کرد. در مدل‌های پیچیده با درک مشکل، مسائل وابسته به استدلال و استنباط بشری، غلبه بر عدم قطعیت داده‌های هیدرولوژیکی که دارای خطای ذاتی و انسانی هستند، می‌توان به‌منظور تخمین پارامتر و کاهش خطا از روش‌های فازی بهره برد (ندیری و همکاران، ۲۰۱۴).

هر سیستم فازی شامل سه بخش اصلی شامل مرحله فازی کردن داده‌ها به‌وسیله تعریف تابع عضویت، مرحله ایجاد ارتباط بین ورودی و خروجی به‌وسیله سری قوانین (if-then) و در نهایت مرحله جمع نتایج سیستم و غیرفازی‌سازی است که به‌وسیله عملگرهای فازی، شامل and (به صورت کمینه‌سازی و وزن‌دهی)، or (به صورت بیشینه‌سازی)، و not جمع نتایج صورت گرفته و غیرفازی‌سازی انجام می‌شود.

مدل فازی به سه روش فازی مختلف ممدانی (۱۹۷۵؛ ۱۹۷۶؛ ۱۹۷۷)، لارسن (۱۹۸۰) و سوگنو (۱۹۸۵) قابل اجرا است. تفاوت روش سوگنو با دو روش دیگر در خروجی ثابت یا رابطه خطی روش سوگنو است. این روش توسط روش دسته‌بندی کاهشی و با مقادیر (۱ و ۰) به

1- Fuzzy C-Means  
2- Aggregation  
3- Implication

### برنامه‌ریزی بیان ژنتیک (GEP)<sup>1</sup>

برنامه‌ریزی ژنتیک (GP) که اولین بار توسط کرامر (۱۹۸۵) معرفی گردید و توسط کوزا (۱۹۹۲) گسترش یافت، جزو الگوریتم‌های تکاملی و زیررده‌ای از روش‌های جست‌وجوی تصادفی محسوب می‌شود. GEP مشابه با GA (الگوریتم ژنتیک) و GP می‌باشد و تفاوت بین این سه در ماهیت افراد است (فریرا، ۱۹۹۹). در GA طبیعت افراد به صورت ردیف‌های علامت‌دار خطی به صورت رشته‌های بیت با طول ثابت (کروموزوم) و بر اساس سیستم ارقام دوتایی عمل می‌کند. در GP افراد ماهیت غیرخطی با طول و شکل متفاوت و به صورت درخت تجزیه<sup>۲</sup> هستند و در GEP افراد به صورت ردیف‌های علامت‌دار با طول ثابت (کروموزوم) رمزگذاری می‌شوند که سپس به صورت غیرخطی و با شکل و اندازه متفاوت به وسیله درخت بیان نمایان می‌گردند (فررا، ۲۰۰۱). شکل‌گیری ساختار کروموزوم‌ها باعث تشکیل ژن‌ها و ایجاد درخت بیان می‌شود. ساختمان هر ژن به وسیله سر و دم آن کنترل می‌شود و این سازمان عملکردی و ساختاری ژن‌ها GEP را برای ایجاد برنامه معتبر ضمانت می‌کند که حد اصلاح کروموزوم‌ها در آن اهمیت ندارد. رمزگشایی اطلاعات مجموعه‌ای از کدها و قوانین را آشکار می‌کند. کدهای ژنتیکی بسیار ساده بوده و ارتباط تک به تک آنها در بین علایمی از کروموزوم‌ها به صورت توابع و پایانه‌ها آشکار می‌شوند. قوانین نیز بسیار ساده هستند که سازماندهی محلی توابع و پایانه‌های درخت بیان و نوع ارتباط بین زیردرختان را بر عهده دارند؛ بنابراین در GEP دو بیان شامل ژن و درخت بیان وجود دارد و به آن بیان Karva گفته می‌شود. در ساختار درخت هر شاخه از یک مجموعه پایانه (متغیرهای مستقل مسأله و متغیرهای حالت سیستم، اعداد ثابت و تصادفی) و مجموعه توابع (عملگر حسابی و توابع مثلثاتی اصلی) تشکیل می‌شود (برلی و همکاران، ۲۰۰۶) که در سر ژن قرار می‌گیرند؛ ولی در قسمت پایانه فقط توابع اصلی قرار می‌گیرند (فررا، ۲۰۰۱).

الگوریتم مورد نظر اقدام به تعریف تابع هدف در قالب معیارهای کیفی می‌کند، سپس تابع یاد شده را برای اندازه‌گیری و مقایسه روش‌های حل مختلف در فرآیند گام

به گام تصحیح ساختار داده‌ها به کار می‌گیرد و در نهایت روش حل مناسب را ارائه می‌کند. مراحل اجرای برنامه‌ریزی ژنتیک به گونه‌ای است که ابتدا با تولید کروموزوم و ایجاد جمعیت اولیه<sup>۳</sup> به وسیله ترکیب تصادفی پایانه‌ها و توابع شروع می‌شود. ارزیابی افراد جمعیت (ژن) به وسیله تابع برازش صورت می‌پذیرد که شاخصی برای سنجش دقت مدل و توانایی مدل برای حل مسأله است و به این ترتیب ژن‌های موثر انتخاب می‌شوند. تعداد مناسبی از جمعیت تولید شده به صورت احتمالی انتخاب می‌شوند؛ به گونه‌ای که معیار انتخاب احتمالی فرد با برازش بهتر به فرد نامرغوب است؛ ولی صرفاً به معنی حذف فرد نامرغوب نیست.

وقتی عملگرهای ریاضی و پایانه‌ها دست به دست هم عمل کنند؛ امکان تقلید تکاملی امکان پذیر می‌شود. در هر بار تولید جمعیت جدید (فرزندان) از فرمول‌ها یکی از عملگرهای ژنتیکی<sup>۴</sup> انتخابی عبور یا تکرار<sup>۵</sup>، تقاطع<sup>۶</sup>، جهش<sup>۷</sup> و کپی یا تولیدمثل<sup>۸</sup> روی تعداد مشخصی از جمعیت انتخابی استفاده می‌شود و جمعیت جدید تولید و مدل به وسیله تابع برازش ارزیابی می‌شود.

در مراحل بعد پارامترهای کنترل‌کننده اجرای مدل (مقادیر مؤلفه‌های عددی و متغیرهای کیفی به کار برده شده برای کنترل اجرای مدل) و معیارهای توقف اجرای برنامه می‌بایست تعیین شود که این معیار متناسب با تکرار فرآیند روی فرزندان برای ارائه نتایج اجرای بهتر برنامه (تعداد، تولید جمعیت جدید و مقدار برازش فرمول‌ها) است. هر فرآیند جست‌وجو با تولید تصادفی یک سری از درختان انجام می‌شود. تابع هدف<sup>۹</sup> هر درخت معمولاً با یک معیار خطا ارزیابی می‌شود. درخت با مقادیر تابعی بهتر با استفاده از روش‌های الیتیسیم<sup>۱۰</sup> (نگهداری کروموزوم بهتر در هر نسل جدید)، چرخ رولت<sup>۱۱</sup>، تورنومنت<sup>۱۲</sup> یا روش ارزش‌دهی<sup>۱۳</sup> انتخاب می‌گردد. عملگرهای دیگر به وسیله انشعابات دیگر تأثیر می‌گذارد و

3- Generation

4- Operators

5- Replication

6- Crossover

7- Mutation

8- Reproduction

9- Objective function

10- Elitism

11- Roulette wheel

12- Tournament

13- Ranking method

1- Gene Expression Programming

2- Parse trees

گردید. بر این اساس داده‌ها به هفت دسته تقسیم و هفت قانون اگر- آنگاه تعیین شد. توابع عضویت استفاده شده برای مدل‌سازی فازی مقادیر آرسنیک تابع گوسی بود که به داده‌های دسته‌بندی شده برازش داده شد. تابع عضویت خروجی مدل سوگنو، تابع خطی است که بر اساس ورودی‌ها ساخته می‌شود؛ برای مثال خروجی اولین تابع عضویت از شش ورودی مطابق معادله (۴) ایجاد می‌شود:

$$\text{Output}_{\text{MFI}} = C_1 \times \text{pH} + C_2 \times \text{SO}_4^{2-} + C_3 \times (\text{NO}_3^- - \text{N}) + C_4 \times \text{F} + C_5 \times \text{Fe(II,III)} + C_6 \times \text{As(III,V)} + C_7 \quad (4)$$

در معادله فوق پارامترهای به ترتیب  $C_1, C_2, C_3, C_4, C_5, C_6$  و  $C_7$  ضریب ورودی‌های  $\text{pH}, \text{SO}_4^{2-}, \text{NO}_3^- - \text{N}, \text{F}, \text{Fe(II,III)}$  و  $\text{As(III,V)}$  مدل‌های ممدانی و لارسن هستند. پارامتر ثابت هر معادله است. این پارامترها با برآورد خطی حداقل مربعات به دست می‌آید. با این توصیف هر تابع عضویت خروجی هفت پارامتر خواهد داشت. در مدل ممدانی و لارسن از روش دسته‌بندی FCM استفاده شد. در این روش نیز تعداد دسته‌های بهینه بر اساس کمترین مقدار RMSE که به ترتیب برابر بود با ۰/۱۱، ۰/۱۲ میلی‌گرم بر لیتر، ۱۲ دسته تعیین شد. در این دو روش نیز توابع عضویت ورودی و خروجی تابع گوسی بود. مقادیر ضریب تبیینی نیز برای مدل‌های ممدانی و لارسن در مرحله آموزش، برابر با ۰/۹۴ و ۰/۹۱ بود. نتایج این دو مدل نیز در جدول ۳ ارائه شده است. با وجود این که هر سه مدل فازی نتایج قابل قبول دارند، ولی مدل ممدانی نتایج نسبتاً بهتری نسبت به دو مدل دیگر دارد. چون هر یک از این سه مدل مزیت‌های خاص خود را دارد، برای استفاده از مزیت‌های هر سه مدل در یک زمان از مدل مرکب استفاده شد.

جدول ۳- نتایج مدل‌های فازی در مراحل آموزش و آزمایش

مراحل	مدل			معیار ارزشیابی
	SFL	MFL	LFL	
آموزش	۰/۱۰	۰/۱۱	۰/۱۲	RMSE (mg/L)
	۰/۹۳	۰/۹۴	۰/۹۱	R <sup>2</sup>
آزمایش	۰/۱۵	۰/۱۲	۰/۳۱	RMSE (mg/L)
	۰/۸۶	۰/۹۱	۰/۸۸	R <sup>2</sup>

### تخمین آرسنیک با استفاده از مدل GEP

در این روش انتخاب مناسب داده‌های ورودی به‌عنوان داده‌های آموزشی دارای اهمیت بالایی است؛ زیرا انتخاب

زیر درختان دیگر حاصل می‌شود. این فرآیند تا زمان حصول حداکثر تعداد تکرار یا تابع خطای معین ادامه می‌یابد.

توابع برازش با هدف بیشینه‌سازی برازش و کاهش خطای تخمین به‌صورت شاخص ضریب تبیین ( $R^2$ ) و جذر میانگین مربعات خطا (RMSE) و ضریب نش- ساتکلیف (NS) مطابق معادلات (۱)، (۲) و (۳) استفاده شده است که برای تعیین دقت و ارزیابی مدل مورد استفاده قرار می‌گیرد.

$$R^2 = \left( \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (x_i - \bar{y})^2}} \right)^2 \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2)$$

$$\text{NS} = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

که در آن  $x$  مقادیر مشاهداتی،  $y$  مقادیر محاسباتی،  $\bar{x}$  میانگین مشاهداتی،  $\bar{y}$  میانگین محاسباتی،  $i$  شماره داده و  $n$  تعداد کل داده‌ها است.

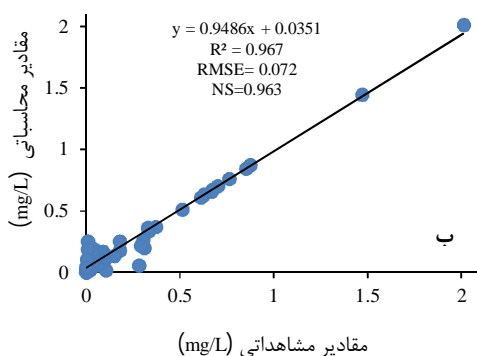
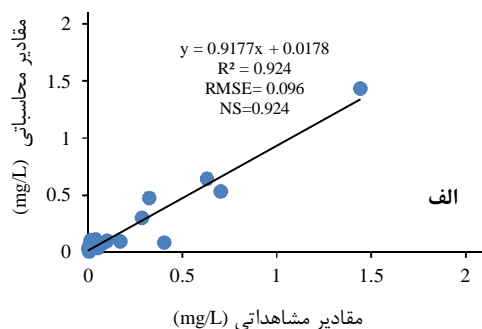
در این تحقیق سعی شده با استفاده از برنامه‌ریزی بیان ژنتیک و کاربرد نرم‌افزار GeneXproTools (v. 4.0)، مدل منطق فازی مرکب به‌منظور پیش‌بینی مقادیر غلظت آرسنیک ایجاد شود و دقت مدل نسبت به سه مدل فازی منفرد مورد ارزیابی قرار گیرد.

### نتایج و بحث

همان‌گونه که در قسمت روش‌شناسی اشاره شد، در این تحقیق از سه روش فازی ممدانی، لارسن و سوگنو برای پیش‌بینی آرسنیک استفاده شد. اولین مرحله در این روش، دسته‌بندی داده‌ها و تعیین توابع عضویت است که در این پژوهش از روش کاهشی برای مدل سوگنو استفاده شد که روشی کارا و مفید در دسته‌بندی داده‌ها و تعداد قوانین در پژوهش‌های اخیر بوده است (جو، ۱۹۹۴؛ ندیری و همکاران، ۲۰۱۴). پارامتر اصلی در این روش دسته‌بندی برای تعیین دسته‌ها و قوانین اگر- آنگاه شعاع دسته‌بندی با مقادیر بین صفر و یک است. کاهش این پارامتر باعث افزایش دسته‌ها و قوانین و افزایش آن به سمت یک باعث کاهش دسته‌ها و قوانین خواهد شد (جو، ۱۹۹۴). بازده مدل فازی بستگی به تعیین شعاع بهینه دسته‌بندی دارد که مقدار بهینه آن بر اساس کمترین RMSE، ۰/۴ تعیین

این ترتیب رابطه (۵) بین ورودی‌ها و خروجی مدل حاصل شد.

$$\text{Output} = \sum \text{varTemp} = [(d_2+d_1)-d_1] + [d_2-d_2)/d_0] + [(d_1-d_2) \times ((d_2+d_2)-(d_2 \times d_2))] = [d_2] + [(d_1-d_2) \times (2d_2-d_2^2)] \quad (5)$$



شکل ۲- مقایسه نتایج مشاهده‌ای و محاسباتی برنامه‌ریزی ژنتیک برای تابع F<sub>1</sub>، مرحله آموزش (الف) و مرحله آزمایش (ب)

جدول ۴- مقایسه حالات مختلف انتخاب عملگرها برنامه

توابع	آموزش		آزمایش	
	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>
F <sub>1</sub>	۰/۰۹۶	۰/۹۲۴	۰/۰۷۲	۰/۹۶۷
F <sub>2</sub>	۰/۱۰۶	۰/۹۲۶	۰/۰۷۷	۰/۹۶۲
F <sub>3</sub>	۰/۱۳۱	۰/۸۹۸	۰/۰۷۴	۰/۹۶۳

جمعیت اولیه تصادفی به منظور آموزش ماهیت حاکم بر پدیده در پیچیدگی، حافظه درگیر و دقت الگوها تأثیرگذار است. از داده‌های خروجی سه مدل فازی به عنوان داده‌های ورودی در برنامه‌ریزی ژنتیک استفاده شد. در این مطالعه پارامترهای اجرایی مدل به‌وسیله تکرار و با استفاده از معیار تابع برازش بهینه شده‌اند. تعداد ۶۰ داده آموزشی و ۲۰ داده آزمایشی به‌گونه‌ای انتخاب شد که کمینه و بیشینه داده‌ها در مرحله آموزش وارد گردند. تولید جمعیت اولیه با انتخاب ۲۰ کروموزوم اولیه و با ویژگی اندازه سر یا طول هر راس<sup>۱</sup> برابر ۷، تعداد ۳ ژن و ۲ ثابت در هر ژن برای برنامه انجام شد. عملگر ریاضی جمع‌کننده برای تابع پیوند بین زیرشاخه‌ها انتخاب گردید. همچنین برای مقایسه نتایج برنامه و انتخاب بهینه‌ترین حالت برای اجرای مدل از سه مجموعه تابع با عملگرهای اصلی مختلف استفاده شد. مطابق نتایج جدول ۴ تابع F<sub>1</sub> شامل چهار عملگر اصلی (+, -, ×, ÷) به‌عنوان توابع اصلی برای برنامه و بهترین حالت برازش نسبت به توابع F<sub>2</sub> و F<sub>3</sub> (پیش‌فرض برنامه) انتخاب شد.

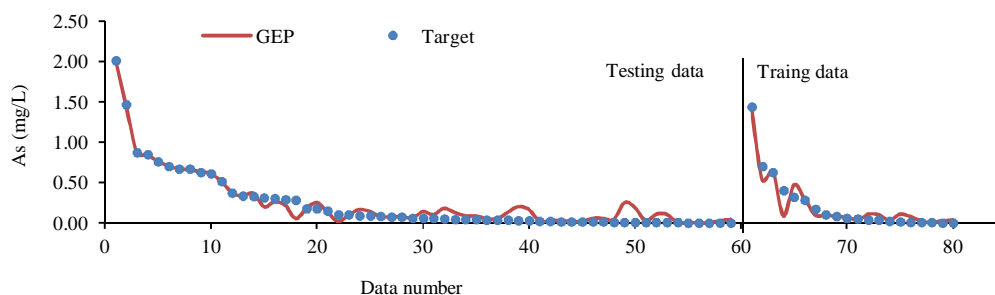
برای ارزیابی میزان خطای مدل از معیارهای ارزیابی R<sup>2</sup>، RMSE و NS استفاده گردید که نتایج مدل در مرحله آموزش و آزمایش برای تابع F<sub>1</sub> مطابق شکل ۲ به‌ترتیب با R<sup>2</sup> برابر ۰/۹۷ و ۰/۹۲ ارائه شده است. بدین ترتیب از مدل برنامه‌ریزی ژنتیک با دقت بالایی برای ترکیب مدل‌های فازی استفاده شد که نمودار مقادیر آرسنیک مشاهده‌ای و محاسباتی در مراحل مربوط به آزمایش و آموزش در شکل ۳ قابل مشاهده است.

علاوه بر تعیین پارامترهای کنترل‌کننده اجرای مدل برای تابع F<sub>1</sub> مطابق جدول ۵، باید معیار توقف اجرای برنامه تعیین شود. به این ترتیب برنامه تا زمان رسیدن به تعداد جمعیت مد نظر برابر ۱۰۰۰ ادامه می‌یابد که با حداکثر تولید ۹۳۵ و با اندازه برنامه ۲۱ به نتیجه مناسب رسید. رابطه بین ورودی و خروجی برنامه‌ریزی بیان ژنتیک به صورت زیر ارائه شد که مطابق آن پارامترهای d<sub>0</sub>، d<sub>1</sub> و d<sub>2</sub> به‌ترتیب مربوط به خروجی مدل فازی نوع لارسن، ممدانی و سوگنو است که به‌عنوان ورودی‌های مدل GEP در نظر گرفته شده بودند. در شکل ۴ درخت بیان ژن و روابط حاصل شده برای هر زیر درخت قابل مشاهده است که به

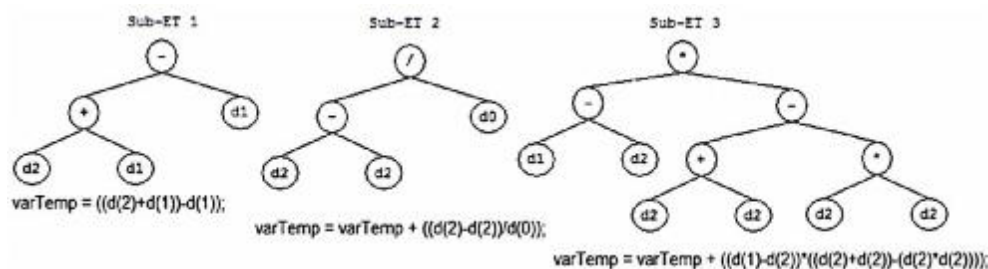


جدول ۵- مشخصات پارامترهای برنامه‌ریزی بیان ژنتیک استفاده شده برای تابع F1

مقادیر	پارامترهای برنامه‌ریزی بیان ژنتیک
۰/۰۴۴	میزان جهش (Mutation rate)
۰/۱	میزان وارون‌سازی (Inversion rate)
۰/۱	میزان ترانهش درجه متوالی (IS transposition rate)
۰/۱	میزان ترانهش ریشه درجه متوالی (RIS transposition rate)
۰/۳	میزان ترکیب تک نقطه (One poin recombination)
۰/۳	میزان ترکیب دو نقطه (Two poin recombination)
۰/۱	ترکیب ژن (Gene recombination)
۰/۱	ترانهش ژن (Gene transposition)
۰/۱	جهش توالی الحاقی ریشه (RNC mutation)



شکل ۳- مقادیر آرسنیک مشاهداتی و محاسباتی مدل برنامه‌ریزی ژنتیک برای تابع F1 در مراحل آزمایش و آموزش



شکل ۴- درخت بیان و روابط حاصل شده برای هر زیر درخت

### نتیجه‌گیری

روش تکرار با استفاده از توابع برازش و به‌منظور بیشینه‌سازی برازش و کاهش خطای تخمین انتخاب شد. مدل ارائه شده با توجه به عدم پاسخگویی مناسب مدل‌های آماری (صادقی‌اقدم، ۱۳۹۱)، راه‌حل صریح و با دقت بالایی را به‌منظور برآورد دقیقی از مقادیر آرسنیک در منطقه مطالعاتی ارائه می‌کند.

در این پژوهش، مقایسه نتایج مدل منطق فازی مرکب حاصل از ترکیب مدل‌های فازی ممدانی، لارسن و سوگنو که توسط برنامه‌ریزی بیان ژنتیک ایجاد شد، نشان از برتری مدل ترکیبی ارائه شده نسبت به هر یک از مدل‌های فازی منفرد دارد. نتایج مطالعات پیشین (چن و لین، ۲۰۰۶؛ ندیری و همکاران، ۲۰۱۳ ب) نیز که با کاربرد مدل‌های ترکیبی به‌منظور تجمیع مزایای مدل‌های منفرد انجام شده، دلیل بر درستی روش استفاده شده در این مطالعه است. بر مبنای این مدل می‌توان ساختار درختی و رابطه بین متغیرهای ورودی (سه مدل فازی) و متغیر خروجی (غلظت آرسنیک) را تعیین کرد. پارامترهای بهینه و عملگرهای برنامه‌ریزی بیان ژنتیک با

### منابع

- اصغری‌مقدم الف. ندیری ع. فیجانی الف. ۱۳۸۸. استفاده از مدل‌های شبکه عصبی و مصنوعی و زمین آمار برای پیش‌بینی مکانی غلظت فلئورید. مجله دانش آب- خاک. ۱۹/۱(۲): ۱۲۹-۱۴۵.
- آقناباتی س.ع. ۱۳۸۳. زمین‌شناسی ایران. سازمان

13. Chiu S. 1994. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*. 2: 267-278.
14. Chouhan S. and Flora S. J. S. 2010. Arsenic and fluoride: Two major ground water pollutants. *Indian journal of experimental biology*. 48(7): 666-678.
15. Cramer N. L. 1985. A representation for the adaptive generation of simple sequential programs. In J. J. Grefenstette, ed., *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*, Erlbaum. Carnegie-Mellon University, Pittsburgh, Pa.
16. Dehghani M. and Abbasnejad A. 2011. Cadmium, Arsenic, Lead and Nitrate Pollution in the Groundwater of Anar Plain. *Journal of Environmental Studies*. 36(56).
17. Demicco R. V. and Klir G. 2004. *Fuzzy Logic in Geology*. Elsevier Science (USA). 347 p.
18. Dragon K. 2006. Application of factor analysis to study contamination of a semi-confined aquifer (Wielkopolska Buried Valley aquifer, Poland). *Journal of Hydrology*. 331: 272-279.
19. Fallah-Mehdipour E. Bozorg Haddad O. and Marin M. A. 2013. Prediction and simulation of monthly groundwater levels by genetic programming. *Journal of Hydro-environment Research*. 7: 253-260.
20. Ferreira C. 2001. Gene Expression Programming: a New Adaptive Algorithm for Solving Problems. *Complex Systems*. 13: 87-129.
21. Guven A. 2009. Linear genetic programming for time-series modeling of daily flow rate. *Earth System Science*. 118(2): 157-173.
22. Jayawardena A. W. Muttill N. and Fernando T. M. K. G. 2005. Rainfall-Runoff Modelling using Genetic Programming. In Zerger, A. and Argent, R.M. (eds.) *MODSIM 2005 International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 1841-1847. ISBN: 0- 9758400-2-9.
23. Kadhodaie-Ilkhchi A. Rezaee M. R. and Rahimpour-Bonab H. 2009. A committee neural network for prediction of normalized oil content from well log data, An example from South Pars Gas Field, Persian Gulf. *Journal of Petroleum Science and Engineering*. 65: 23-32.
24. Kisi O. Shiri J. and Nazemi A. H. 2011. A Wavelet-Genetic Programming Model for Predicting Short-Term and Long-Term Air Temperatures. *Scienceline Publicat ion. Journal of Civil Engineering and Urbanism*.  
زمین‌شناسی و اکتشافات معدنی کشور. ۵۸۶ ص.  
۳. آقای ح. ۱۳۸۸. بررسی هیدروژئوشیمی محدوده غرب هشت‌رود با تأکید بر آنومالی‌های احتمالی آرسنیک در منطقه. پایان‌نامه کارشناسی ارشد. گروه زمین‌شناسی. دانشگاه تبریز. ۱۵۷ ص.  
۴. صادقی‌ا قدم ف. ۱۳۹۱. بررسی تغییرات زمانی و مکانی کیفیت منابع آب ورودی به سد سهند هشت‌رود با تأکید بر آنومالی آرسنیک. پایان‌نامه کارشناسی ارشد. گروه زمین‌شناسی. دانشگاه تبریز. ۱۶۱ ص.  
۵. ندیری ع. ۱۳۹۰. بررسی تغییرات هیدروژئوشیمیایی آب انباشته شده در سد سهند و منابع آب مرتبط با آن در حوضه. وزارت نیرو، شرکت آب منطقه‌ای استان آذربایجان شرقی.  
۶. ندیری ع. اصغری‌مقدم الف. صادقی‌ا قدم ف و آقای ح. ۱۳۹۰. بررسی آنومالی آرسنیک موجود در منابع آب سد سهند. *مجله محیط‌شناسی*. ۳۸(۳): ۷۴-۶۱.
7. Anawar H. M. Akai J. Mihaljevic M. Sikder A. M. Ahmed G. Tareq S. M. and Rahman M. M. 2011. Arsenic Contamination in Groundwater of Bangladesh: Perspectives on Geochemical, Microbial and Anthropogenic Issues, *Water*. 3:1050-1076.
8. Aytek A. and Kisi O. 2008. A genetic programming approach to suspended sediment modeling. *J Hydrol*. 351: 288-298.
9. Borelli A. De Falco I. Della C. A. Nicodemi M. and Trautteur G. 2006. Performance of Genetic Programming to Extract the Trend in Noisy Data Series. *Physica A*. 370: 104-108.
10. Bundschuh J. Litter M. I. Parvez F. Román-Ross G. Nicolli H. B. Jean J. Sh. Liu Ch. Lopez D. Armienta M. A. Guilherme L.R.G. Cuevas A. G. Cornejo L. Cumbal L. and Toujaguez L. R. 2011. One century of arsenic exposure in Latin America: A review of history and occurrence from 14 countries. *Science of the Total Environment*. 429: 2-35.
11. Chen C. H. and Lin Z. S. 2006. A committee machine with empirical formulas for permeability prediction. *Computers and Geosciences*. 32: 485-496.
12. Chiban M. Zerbet M. Carja G. and Sinan F. 2012. Application of low-cost adsorbents for Arsenic removal: A review. *Journal of Environmental Chemistry and Ecotoxicology*. 4(5):91-102.

- of Sahand Dam of Hashtrud, Iran. ASCE Texas Section Fall Conference and Centennial Celebration.
38. Nadiri A. Hassan M. M. and Asadi S. 2015. Supervised Intelligence Committee Machine to Evaluate Field Performance of Photocatalytic Asphalt Pavement for Ambient Air Purification Transportation Research Record: Journal of the Transportation Research Board. 2528: 96-105.
  39. Olatunji S. O. Selamat A. and Abdulraheem A. 2011. Modeling the permeability of carbonate reservoir using type-2 fuzzy logic systems. Computers in Industry. 62: 147-163.
  40. Oremland R. S. and Stolz J. F. 2003. The ecology of Arsenic. Berlin. Heidelberg. Science. 300: 939-944.
  41. Rabunal J. R. Puertas J. Suarez J. and Rivero D. 2007. Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. Hydrological Processes. 21:476-485.
  42. Sattari M. T. Joudi A. R. and Kusiak A. 2016. Estimation of Water Quality Parameters With Data-Driven Model. American Water Works Association, Journal AWWA. 108(4): E232-E239.
  43. Shoaib M. Shamseldin A. S. Melville B. W. and Muneer Khan M. 2015. Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. Journal of Hydrology. 527: 326-344.
  44. Su C. Zhu Y. Abbas Z. and Huq Md. E. 2016. Sources and controls for elevated arsenic concentrations in groundwater of Datong Basin, Northern China. Environmental Earth Sciences. 75(7): 1-13.
  45. Sugeno M. 1985. Industrial Application of Fuzzy Control. Elsevier Science Pub. Co., New York, USA. 278 p.
  46. Tayfur G. Nadiri A. A. and Asghari Moghaddam A. 2014. Supervised intelligent committee machine method for hydraulic conductivity estimation. Water resources management. 28(4): 1173-1184.
  47. Ustoorikar K. and Deo M. C. 2008. Filling up gaps in wave data with genetic programming. Marine Structures. 21: 177-195.
  48. Wilson J. Schreier H. and Brown S. 2008. Arsenic in groundwater in the White Rock-Surrey-Langley Area. Fraser Health Authority, Abbotsford. 40 p.
  49. Yassin M. A. Alazba A. A. and Mattar M. A. 2016. Artificial neural networks versus 1(1): 25-37.
  25. Koza J. R. 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection, Cambridge, MA. The MIT Press. 840 p.
  26. Labani M. M. Kadkhodaie-Ilkhchi A. and Salahshoor K. 2010. Estimation of NMR log parameters from conventional well log data using a committee machine with intelligent systems, A case study from the Iranian part of the South Pars gas field. Persian Gulf Basin: Journal of Petroleum Science and Engineering. 72: 175-185.
  27. Larsen P. M. 1980. Industrial applications of fuzzy logic control: International Journal of Man-Machine Studies. 12: 3-10.
  28. Mahimairaja S. Bolan N. S. Adriano D. C. and Robinson B. 2005. Arsenic contamination and its risk management in complex environmental settings. Advances in Agronomy. Elsevier Inc. Advances in Agronomy. 86: 1-82.
  29. Mamdani E. H. 1976. Advances in the linguistic synthesis of fuzzy controllers. International Journal of Man-Machine Studies. 8(6): 669-678.
  30. Mamdani E. H. 1977. Application of fuzzy logic to approximate reasoning using linguistic synthesis computers. IEEE Transactions. C-26(12): 1182-1191.
  31. Mamdani E. H. and Assilian S. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies. 7: 1-13.
  32. Mandal S. Mahapatra S. S. Adhikari S. and Patel R. K. 2015. Modeling of Arsenic (III) Removal by Evolutionary Genetic Programming and Least Square Support Vector Machine Models. Environ. Process. 2: 145-172.
  33. Moghaddam A. 2013b. Supervised committee machine with artificial intelligence for prediction of fluoride concentration. Journal of Hydroinformatics. 15(4): 1474-1490.
  34. Mosaferi M. Yunesian M. Dastgiri S. Mesdaghinia A. and Esmailnasab N. 2008. Prevalence of skin lesions and exposure to Arsenic in drinking water in Iran. Science of the Total Environment. 390: 69-76.
  35. Nadiri A. A. Chitsazan N. Tsai F. T. C. and Moghaddam A. A. 2014. Bayesian Artificial Intelligence Model Averaging for Hydraulic Conductivity Estimation. Journal of Hydrologic Engineerin. 19: 520-532
  36. Nadiri A. A. Fijani E. Tsai F. T. C. and Asghari.
  37. Nadiri A. A. Sadeghi Aghdam F. Asghari Moghaddam A. and Asadi S. 2013a. Arsenic Anomalies Occurrence in Water Resources

- gene expression programming for estimating reference evapotranspiration in arid climate. *Agricultural Water Management*. 163: 110-124.
50. Zadeh L. A. 1965. Fuzzy sets. *Information and control*. 8: 338-353.