

بررسی عملکرد مدل‌های داده مینا در شبیه سازی افق‌های زمانی سطح آب زیرزمینی با استفاده از روش تلفیقی آزمون گاما و الگوریتم ژنتیک (GA-GT)

علی میرعربی

دانشجوی دکترا، گروه زمین شناسی معدنی و آب، دانشکده علوم زمین، دانشگاه شهید بهشتی

حمیدرضا ناصری

استاد گروه زمین شناسی، دانشکده علوم زمین، دانشگاه شهید بهشتی

محمد نخعی

دانشیار گروه زمین شناسی، دانشکده علوم زمین دانشگاه خوارزمی

فرشاد علیجانی

استادیار گروه زمین شناسی معدنی و آب، دانشکده علوم زمین، دانشگاه شهید بهشتی

تاریخ دریافت: ۱۳۹۷/۰۲/۰۳ تاریخ پذیرش: ۱۳۹۷/۱۱/۰۹

H-nassery@sbu.ac.ir

چکیده

به منظور مدیریت پایدار منابع آب زیرزمینی، تعیین رفتار سطح آب زیرزمینی ضروری است. نوسانات سطح آب زیرزمینی فرآیندی غیرخطی و پیچیده است که مدل‌های هوش محاسباتی داده مینا قادر هستند بدون تقریب و ساده‌سازی به مدل‌سازی آن بپردازند. در این مطالعه میزان دقت و کارایی هر یک از مدل‌های داده مینا هوش مصنوعی پرسپترون چند لایه (MLP)، استنتاجی فازی عصبی (ANFIS) و رگرسیون بردار پشتیبان (SVR) در شبیه‌سازی سه افق زمانی پیش‌رو (t+2, t+1 و t+3) سطح آب زیرزمینی دشت هشتگرد مورد ارزیابی قرار گرفت. در این راستا از قابلیت‌های روش آزمون گاما و تلفیق آن با الگوریتم ژنتیک (GA-GT) بمنظور انتخاب ترکیب ورودی بهینه و نیز روش M-Test در تعیین طول بهینه داده‌های آموزش مدل استفاده شد. نتایج مشخص کرد که دقت مدل‌های داده مینا در شبیه‌سازی افق زمانی t+1 سطح آب زیرزمینی بیشتر از افق زمانی t+2 و t+3 است. بمنظور تعیین میزان عملکرد و کارایی مدل‌ها، نتایج براساس شاخص DDR مورد تحلیل قرار گرفت که محاسبات نشان داد مدل ANFIS در شبیه‌سازی افق زمانی اول بهترین عملکرد و کارایی را داشته است. همچنین استنباط شد که مدل MLP و ANFIS در شبیه‌سازی افق‌های ماهانه نزدیک‌تر و SVR در افق‌های ماهانه دورتر کاربرد بهتری دارند.

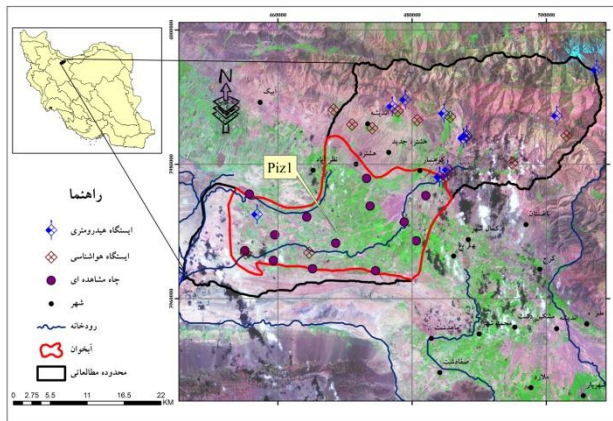
کلمات کلیدی: سطح آب زیرزمینی، گاما تست، شبکه عصبی مصنوعی، رگرسیون بردار پشتیبان، الگوریتم ژنتیک

مقدمه

پیچیده بیان شود. از روش‌های پرکاربردی که می‌توان با داده‌های سری زمانی موجود و کمترین متغیر ممکن به بررسی روابط غیرخطی بین پارامترهای موثر بر نوسانات سطح آب زیرزمینی و در نتیجه شبیه‌سازی رفتار پیچیده آبخوان پرداخت، روش داده مینا مبتنی بر هوش محاسباتی است. روش داده مینا تلاش دارد بدون در نظر گرفتن ساختار درونی فرآیندهای فیزیکی یک رابطه مستقیم بین داده‌های ورودی و خروجی ارائه کند. شبکه‌های عصبی مصنوعی (ANN)، سیستم استنتاجی فازی عصبی (ANFIS)، ماشین بردار پشتیبان (SVR) و برنامه ریزی ژن (GP) از مهمترین هوش‌های محاسباتی برای شبیه‌سازی آب زیرزمینی می‌باشند. در طی سالیان اخیر تحقیقات متعددی در ارتباط با شبیه‌سازی نوسانات سطح آب زیرزمینی با استفاده از روش هوش‌های محاسباتی در نقاط مختلف دنیا انجام شده است. (Nourani and Mousavi 2016; Suryanarayana et al. 2014; Shirmohammadi et al. 2013; Shiri et al. 2013; Adamowski and Chan 2011; Banerjee et al. 2011; Daliakopoulos et al. 2005; Nayak et al. 2004; Nourani et al. 2008; Srekanth et al. 2009; Choubin et al. 2017)

آب زیرزمینی یکی از مهمترین منابع تامین آب بویژه در مناطق نیمه خشک و خشک محسوب می‌شود. کاهش نزولات جوی و خشکسالی دهه اخیر و عدم رعایت مدیریت پایدار و تعادل بین مصارف و منابع منجر شده است که منابع آب زیرزمینی با تهدیدهایی از قبیل بهره‌برداری بیش از حد، کاهش و تخریب کیفیت و فرونشست آبخوان مواجه باشد. مدیریت منابع آب-زیرزمینی یکی از ارکان اساسی در حفاظت کمی و کیفی آب زیرزمینی است و جهت برنامه‌ریزی‌های بلندمدت و استفاده بیشتر و بهتر از پتانسیل‌های آبی دشت‌ها از اهمیت و حساسیت بسیار زیادی برخوردار است. شبیه‌سازی رفتار آبخوان و پیش‌بینی نوسانات سطح آب زیرزمینی از الزامات یک مدیریت صحیح و پایدار منابع آب زیرزمینی می‌باشد. اثر متقابل فاکتورهای هیدروژئولوژی و هیدرولوژی از قبیل پارامترهای هیدرولیکی، میزان تخلیه و تغذیه آب زیرزمینی، بارندگی، جریان آب سطحی، دما، تبخیر و عوامل دیگر سبب شده است تا نوسانات آب زیرزمینی به صورت پدیده‌ای غیرخطی و

هواشناسی کریم آباد، هیدرومتری فشنند و منبع انتخابی حاجی شاه که دارای آمار دقیق و صحیح می باشند، جمع آوری گردید.



شکل ۱. موقعیت منطقه مورد مطالعه.

به منظور بررسی میزان همبستگی متغیرهای مورد اشاره و حذف متغیرهای همبسته از شاخص آماری تورم واریانس (VIF) استفاده گردید. در این روش همبستگی هر متغیر با سایر متغیرها محاسبه شده و در نهایت متغیرهایی که دارای ضریب VIF بالا هستند به عنوان متغیرهای ورودی که همبسته می باشند شناخته شده و از فرآیندهای مدل سازی حذف می شوند. مقدار آستانه حذف متغیرهای چند خطی ورودی برای VIF عدد ۵ در نظر گرفته شده است. شاخص تورم واریانس بصورت زیر محاسبه می شود:

(رابطه ۱)

$$VIF = \frac{1}{1 - R^2}$$

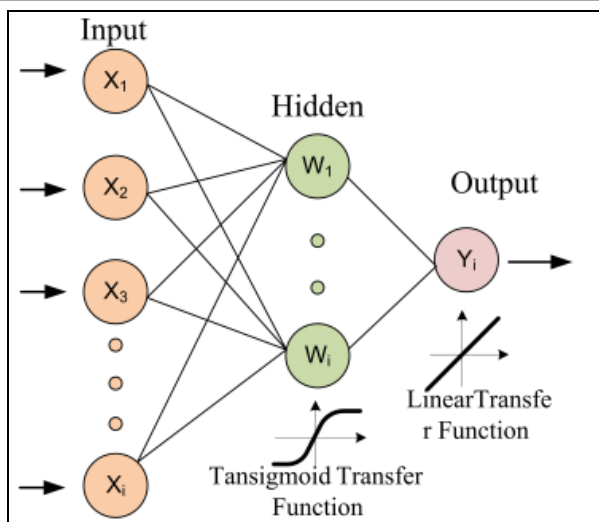
در (جدول ۱) نتایج محاسبه شاخص VIF از طریق نرم افزار SPSS برای متغیرهای جمع آوری شده ارائه شده است که متغیر دما و تبخیر با مقدار VIF بین ۷/۰۰۹ و ۶/۹۳ بیشترین وابستگی را با سایر متغیرها دارند و از متغیرهای ورودی حذف می شوند. در نهایت مدل های داده مبنای پیش-بینی نوسانات سطح آب زیرزمینی با ۴ متغیر ورودی شامل سطح آب زیرزمینی (H) برحسب متر، بارش کل ماهانه (P) برحسب میلی متر، متوسط ماهانه دبی رودخانه (R) برحسب مترمکعب بر ثانیه و نرخ پمپاژ ماهانه (Pw) برحسب لیتر بر ثانیه به عنوان متغیرهای ورودی مناسب در مدل ها تعیین شد که مشخصات آماری آن ها در (جدول ۲) ارائه شده است. نظر به این که اثر مقدار برخی از متغیرها در ماه های گذشته در سازی سطح آب زیرزمینی بسیار مشهود بوده و ضروری است تعداد تاخیرهای زمانی مورد نیاز به طور مناسب تعیین گردد، لذا برای متغیرهای سطح آب زیرزمینی، بارندگی و دبی رودخانه حداکثر سه تاخیر زمانی در نظر گرفته شد و در نهایت ۱۳ متغیر شامل (H_t, H_{t-1}, H_{t-2}, H_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3}, R_t, R_{t-1}, R_{t-2}, R_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}) به عنوان ورودی های مدل داده مبنای استفاده شد.

یکی از فرایندهای پیچیده و مهم در فرآیند مدل سازی با هوش های محاسباتی، پیش پردازش پارامترهای ورودی جهت کاهش ابعاد داده ها و انتخاب ترکیب مناسب می باشد. علیرغم وجود تحقیقات فراوان برای مدل سازی آب زیرزمینی با استفاده از هوش های محاسباتی، هم چنان در این باره سوالات اساسی در خصوص میزان اهمیت هر یک از داده های ورودی، بهترین نوع ترکیب ورودی به مدل و تعداد داده های مورد نیاز برای آموزش و آزمون مدل وجود دارد (Remesan and Mathew, 2015). بررسی تحقیقات گذشته نشان می دهد که در اکثر مدل های هوش محاسباتی بکار گرفته شده در آب زیرزمینی، از روش های پیش پردازش پارامترهای ورودی استفاده نشده است بلکه بهترین ترکیب ورودی از میان ترکیبات محدود و با استفاده از روش سعی و خطا انتخاب شده است (Choubin et al. 2017). روش های مختلفی برای پیش پردازش داده های ورودی و کاهش ابعاد داده ها و انتخاب ترکیب بهینه وجود دارد که از جمله آن ها می توان به آزمون گاما (GT)، تحلیل مولفه اصلی (PCA)، تئوری انتروپی (TE) و الگوریتم ژنتیک (GA) اشاره کرد. هر یک از این روش ها مزایا و محدودیت های خاص خود را نیز دارند و یک روش معمول و قاعده مند برای یافتن بهترین ترکیب از میان پارامترهای مختلف ورودی دیده نمی شود. این مقاله قصد دارد ضمن مقایسه و کاربرد سه مدل رایج داده مبنای (SVR, ANFIS, MLP) در شبیه سازی افق های زمانی (ماهانه) مختلف سطح آب زیرزمینی، از تلفیق روش های آزمون گاما و الگوریتم آموزشی ژنتیک بمنظور تعیین بهترین ترکیب از میان پارامترهای مختلف ورودی و نیز از روش آزمون M-test جهت تعیین طول دسته داده های ورودی در مرحله آموزش برای ایجاد یک مدل هموار استفاده نماید.

موارد و روش پژوهش

منطقه مورد مطالعه و داده های مورد استفاده

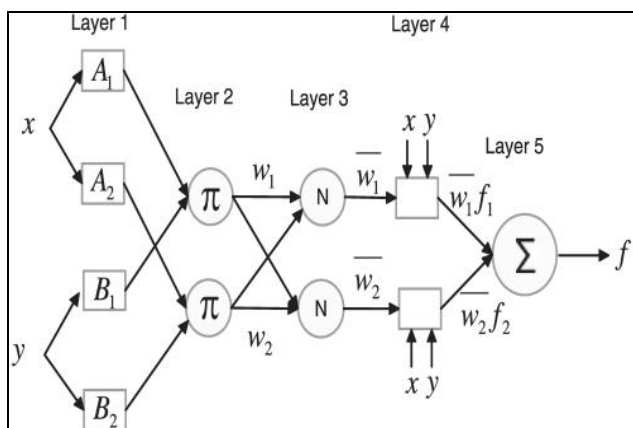
دشت هشتگرد که در شرق محدوده مطالعاتی تهران-کرج قرار دارد جزء مناطق نیمه خشک محسوب می شود. وسعت پهنه آبخوان آبرفتی دشت هشتگرد ۹۱۴/۵ کیلومترمربع است. متوسط بارندگی ۲۰ ساله در این دشت حدود ۲۸۰ میلی متر در سال می باشد. دشت هشتگرد از رسوبات آبرفتی کواترنز شامل ماسه دانه ریز و رس ماسه ای با ضخامت ۳۵ الی ۱۴۰ متر تشکیل شده است. این دشت دارای ۱۳ چاه مشاهده ای است که به صورت ماهانه داده برداری می شود (شکل ۱). براساس اطلاعات و تحلیل چاه های مشاهده ای روند عمومی جریان آب زیرزمینی شمال شرقی به جنوب غربی می باشد. برداشت بی رویه و نرخ شدید افت در آبخوان هشتگرد منجر به رخداد فرونشست نیز شده است (گزارش ممنوعیت دشت، ۱۳۹۵). به منظور شبیه سازی نوسانات سطح آب زیرزمینی از چاه مشاهده ای (Piz 1) که در مرکز دشت حفر شده و نوسانات سطح آب زیرزمینی در آن تحت تاثیر بسیاری از فاکتورهای هیدروژئولوژی و هیدرولوژی می باشد، به عنوان نمایانگر آبخوان مورد مطالعه انتخاب شد. در بررسی اولیه، سری زمانی متغیرهایی همچون سطح آب زیرزمینی (H)، بارندگی (P)، میزان دبی رودخانه (R)، میزان برداشت چاه های بهره برداری (Q)، دما (T) و تبخیر (E_v) در یک بازه زمانی ۱۶ ساله (سال آبی ۸۰-۱۳۷۹ تا ۹۵-۱۳۹۴) از ایستگاه باران سنجی و



شکل ۲. ساختار شاماتیکی مدل شبکه عصبی پرسپترون چند لایه.

سیستم استنتاجی فازی عصبی

سیستم استنتاجی فازی عصبی (ANFIS) تلفیق الگوریتم‌های یادگیری شبکه عصبی و منطق فازی است که با هدف ایجاد رابطه بین فضای ورودی و خروجی استفاده می‌شود و قابلیت خوبی در آموزش، ساخت و طبقه بندی دارد (نخعی و ودیعتی، ۱۳۹۱). این مدل شامل پنج لایه است که به ترتیب شامل گره‌های ورودی، گره‌های قاعده، گره‌های متوسط، گره‌های نتیجه و گره‌های خروجی است که به طور مستقیم با هم ارتباط دارند (شکل ۳). در این مدل از الگوریتم پس انتشار خطا با حداقل کردن میانگین مربعات خطا بین خروجی شبکه و خروجی جهت یادگیری مدل استفاده می‌شود. بهره‌گیری از دانش متخصص و ایجاد قوانین فازی از اطلاعات عددی و در نتیجه تبدیل پیچیده هوش بشری به سیستم‌های فازی از مزیت‌های ANFIS است. سیستم‌های استنتاج فازی شامل سه نوع ممدانی، سوگنو و سوکاموتو می‌باشند که سیستم سوگنو در محاسبات عملکرد بهتری دارد و از خروجی قطعی برخوردار می‌باشد (کیا و همکاران، ۱۳۹۴، کامرانی و رضایی، ۱۳۹۴).



شکل ۳. نمایی از ساختار استنتاجی فازی عصبی

جدول ۱. محاسبات شاخص VIF.

Parameters	Unstandardize Coefficients		Standardize Coefficients	t	Sig.	Colinearity Statistics	
	B	S _{sq. Error}	Beta			Tolerance	VIF
H	1198.9	3.483		344.2	0		
P	0.013	0.023	0.05	0.579	0.563	0.593	1.669
R	-0.614	0.575	-0.083	-1.068	0.287	0.734	1.363
T	-0.037	0.116	-0.057	-0.32	0.749	0.143	7.009
Ev	0.003	0.008	0.067	0.382	0.703	0.144	6.939
Q	-0.58	0.095	-0.41	-6.084	0	0.979	1.021

جدول ۲. مشخصات آماری متغیرهای ورودی منتخب.

پارامتر آماری	متغیر ورودی			
	H	P	R	Q
تعداد	۱۹۲	۱۹۲	۱۹۲	۱۹۲
میانگین	۱۱۷۸.۶۴	۱۹.۶۹	۰.۳۵	۳۵.۱۲
کمینه	۱۱۶۸.۳۲	۰	۰.۰۰	۲۳
بیشینه	۱۱۹۲.۲۵	۱۴۹	۴.۹۲	۴۶
انحراف معیار	۶.۲۰	۲۳.۳۴	۰.۸۴	۴.۳۸

شبکه عصبی پرسپترون چند لایه

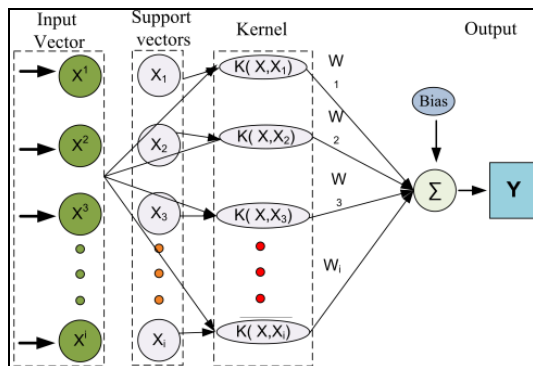
شبکه عصبی مصنوعی به‌عنوان نسل ششم محاسبه سیستمی نوین جهت تجزیه و تحلیل داده‌هاست که قادر است به سادگی ارتباطات پنهان میان داده‌ها حتی داده‌هایی که دارای ارتباط غیرخطی، توزیعی، موازی، غیررقومی (آنالوگ) و غیرالگوریتمی هستند با قابلیت فراگیری و طراحی خود کشف کند. از جمله مزایای شبکه عصبی می‌توان به قابلیت تقریب زنی هر نوع تابع، قابلیت تطبیق، سادگی کار با آن و عدم نیاز به درک عمیق از روابط بین متغیرها و تابع، اشاره کرد (نخعی و صابری، ۱۳۹۱).

انواع مختلفی از شبکه عصبی وجود دارد که متداول‌ترین نوع شبکه عصبی در پیش بینی و حل مسائل غیرخطی شبکه پرسپترون چند لایه (MLP) است که در این مقاله نیز استفاده شده است. مدل MLP از یک لایه ورودی، یک یا چند لایه مخفی و لایه خروجی تشکیل می‌شود (شکل ۲). بهترین الگوریتم جهت آموزش شبکه در زمینه منابع آبی الگوریتم لوبنبرگ-مارکوت می‌باشد (Yu et al., 2017). از موضوعات بسیار مهم در طراحی شبکه عصبی تعداد گره میانی می‌باشد. حداکثر مجاز تعداد گره میانی از طریق (رابطه ۲) ارائه شد (Huang, 2003). در این رابطه N^0 تعداد گره استفاده شده خروجی و N^i تعداد گره استفاده شده در لایه ورودی می‌باشد که برابر با تعداد متغیرهای استفاده شده در ترکیب ورودی است.

(رابطه ۲)

$$N^H = 2 \sqrt{(N^0 + 2)N^i}$$

رگرسیون بردار پشتیبان



شکل ۴. ساختار شماتیک مدل رگرسیون بردار ماشین

گاما

آزمون گاما یک مدل بهینه‌سازی غیرخطی است که به کمک آن می‌توان ترکیب مناسب از بین پارامترهای ورودی مختلف برای مدل‌سازی و ایجاد یک مدل هموار را بررسی کرد (بسالت پور و همکاران، ۱۳۹۲). برای اولین بار آزمون گاما توسط (Koncar 1997) ارائه و در سال‌های بعد توسط بسیاری از محققین مورد استفاده قرار گرفت (Durrant, 2001; Choubin and malekian, 2017;). در تکنیک GT هر ترکیب ورودی شامل ردیفی از اعداد صفر و یک که ماسک گفته می‌شود، می‌باشد که عدد یک نشان‌دهنده مشارکت آن پارامتر در ترکیب ورودی و عدد صفر نشان‌دهنده عدم شرکت آن است (Remesan and Mathew, 2015). به-عنوان مثال در یک ترکیب ورودی پنج متغیره ماسک ۱۰۱۱۱ بیانگر این است که پارامتر دوم در این ماسک در ترکیب ورودی شرکت نمی‌کند و کنار گذاشته می‌شود. جزئیات روش آزمون گاما در مراجعی همچون Durrant (2001) ارائه شده است. نکته قابل اشاره این است که انتخاب بهترین ماسک بعنوان ترکیب ورودی در آزمون گاما براساس محاسبه پارامترهای آماره گاما (Γ) و شاخص V_{ratio} انجام می‌شود. جهت تعیین بهترین ماسک نیاز است تمام ترکیبات معنادار (ماسک‌ها) از داده‌های ورودی به-صورت دستی برای محاسبه آماره گاما و شاخص V_{ratio} ایجاد شود. هرچه شاخص V_{ratio} و آماره گاما (Γ) یک ماسک کمتر و به صفر نزدیک شود بیانگر توانایی آن ماسک جهت استفاده در شبیه‌سازی مدل می‌باشد و هرگاه به یک نزدیک شوند نمایانگر یک خطای تصادفی بسیار بزرگ است و ایجاد مدل براساس ماسک مدنظر، نامناسب و ناکارآمد است (قبائی سوق و همکاران، ۱۳۸۹).

معیار ارزیابی خطا

در این تحقیق از ویژگی‌های آماری میانگین قدرمطلق خطا (MAE)، جذر میانگین مربعات خطا (RMSE) و ضریب همبستگی (R) و نش-ساتکلیف (NSE) مطابق با روابط زیر به منظور ارزیابی کارایی و خطای هر شبکه و توانایی آن برای پیشگویی دقیق استفاده شد. (رابطه ۷)

$$MAE = \frac{1}{N} \sum_{i=1}^N |H_{Pi} - H_{Oi}|$$

مدل‌های ماشین بردار پشتیبان (SVM) به دو گروه عمده طبقه‌بندی ماشین بردار پشتیبان (CSVM) و رگرسیون ماشین بردار پشتیبان (SVR) تقسیم بندی می‌شوند (Kecman, 2005). از طبقه‌بندی ماشین بردار پشتیبان جهت خوشه‌بندی داده‌هایی که ماهیت رده‌بندی دارند، استفاده می‌شود. مدل رگرسیون ماشین بردار پشتیبان در پیش‌بینی مسائل کاربرد دارد. در مدل SVR منظور از رگرسیون به‌دست آوردن ابرصفحه است که بر داده‌های مورد نظر برازش داده می‌شود و همواره در میان تمام ابرصفحاتی که داده‌ها را تفکیک می‌کنند، آن ابرصفحه را بر می‌گزینند که بیشترین تفکیک پذیری یا بیشترین اندازه حاشیه را میان داده‌های رده‌های مختلف حاصل کنند (رمضانی و ذونعمت، ۱۳۹۶). مدل رگرسیون ماشین بردار پشتیبان به صورت ضرب داخلی $f(x) = W^T \cdot \phi(x) + b$ که بردار ضرایب، b ثابت مشخصه‌های تابع رگرسیونی (بایاس) و $\phi(x)$ ضرب داخلی متغیرها و $f(x)$ که خود تابعی از چند متغیر مستقل x است، ارائه می‌شود و بهینه رگرسیون به وسیله مینیمم تابع زیر بیان می‌شود.

(رابطه ۳)

$$\text{minimize: } \frac{1}{2} \|W\|^2 + C(\sum_{i=1}^N (\xi_i + \xi_i^*))$$

(رابطه ۴)

$$\text{subject to } \begin{cases} f(x) - d_i \leq \varepsilon + \xi_i^*, & i=1,2,\dots,N \\ d_i - f(x) \leq \varepsilon + \xi_i, & i=1,2,\dots,N \\ \xi_i, \xi_i^*, & i=1,2,3,\dots,N \end{cases}$$

که $\|W\|$ نرم بردار ضرایب و کمبود ξ_i و ξ_i^* متغیرهای کمبود کمکی، ε مقدار خطای مجاز و C پارامتر ضریب تعادل پیچیدگی و N تعداد نمونه‌ها می‌باشد. در نهایت، تابع رگرسیون بردار پشتیبان را می‌توان به‌صورت زیر بازنویسی کرد.

(رابطه ۵)

$$f(x) = \sum_{i=1}^I (\alpha_i - \alpha_i^*) k(x_i, x) + b$$

که α_i و α_i^* ضرایب لاگرانژ و $k(x_i, x)$ ضرب داخلی متغیرها می‌باشد. به-دلیل پیچیده بودن محاسبه $k(x_i, x)$ از توابع کرنل مختلف از جمله خطی، درجه دوم، شعاعی و چند جمله‌ای در رگرسیون ماشین بردار پشتیبان استفاده می‌شود. معمولاً تابع کرنل شعاعی برای پیش‌بینی عملکرد بهتری دارد. معادله این تابع کرنل به صورت زیر است.

(رابطه ۶)

$$k(x_i, x) = \exp(-\gamma \|x - x_i\|^2)$$

در ساخت یک مدل کارآمد ماشین بردار پشتیبان با کرنل شعاعی، پارامترهای تابع کرنل (γ) و ظرفیت تابع (C) باید به طور دقیق برآورد شوند (Kecman, 2005). اصول انتخاب C و γ توسط کاربر و براساس اطلاعات قبلی و یا تجربه شخصی و به‌صورت سعی و خطا انتخاب می‌گردد. ساختار شماتیکی از SVR در (شکل ۴) ارائه شده است.

(رابطه ۸)

$$f_1(M) = \begin{cases} 1 - \frac{1}{1-10V_{Ratio}} & \text{if } V_{Ratio} < 0 \\ 2 - \frac{2}{1+V_{Ratio}} & \text{if } V_{Ratio} \geq 0 \end{cases}$$

(رابطه ۱۲)

(رابطه ۱۳)

$$f_2(M) = 1 - \frac{1}{1 + \left| \frac{\text{gradient}(M)}{\text{range of output}} \right|}$$

(رابطه ۱۴)

$$f_3(M) = \frac{\text{Active}(M)}{\text{Length}(M)}$$

در این رابطه M تعداد ترکیب معنادار (ماسک) و W وزن‌های از پیش تعیین شده می‌باشند. باید دقت کرد که مجموع مقادیر W_1, W_2, W_3 باید مساوی یا کمی بیشتر از یک در نظر گرفته شود (Marquez and Coimbra, 2011). مقادیر پارامترهای استفاده شده در الگوریتم ژنتیک جهت مینیمم کردن تابع فوق و استخراج بهترین ماسک با کمترین مقدار آماره گاما و V_{ratio} در (جدول ۳) ارائه شده است.

جدول ۳. مقادیر تعیین شده پارامترهای الگوریتم ژنتیک و تابع هدف.

پارامتر	مقدار	
الگوریتم ژنتیک	اندازه جمعیت	۱۰۰
	نرخ جمعیت	۰.۵
	نرخ جهش	۰.۰۵
	زمان (دقیقه)	۵
مقادیر تابع هدف	W1	۱
	W2	۰.۱
	W3	۰.۱

با مینیمم کردن تابع هدف در الگوریتم ژنتیک، بهترین ترکیب ورودی جهت شبیه‌سازی افق‌های زمانی (ماهانه) اول تا سوم سطح آب زیرزمینی بدست آمد که نتایج در (جدول ۴) ارائه شده است. طبق این جدول، آماره گاما (Γ) و شاخص V_{ratio} کمترین مقدار ممکن در بین ترکیبات معنادار را دارد و نزدیک به صفر می‌باشند. بر این اساس بهترین ترکیب ورودی در شبیه‌سازی افق زمانی اول و دوم با تعداد ۹ متغیر و در افق زمانی سوم بهترین ترکیب ورودی با هفت متغیر بدست آمد.

جدول ۴. نتایج اجرای روش GA-GT و ترکیب بهینه ورودی.

افق زمانی	پارامترهای مدل تلفیقی GT-GA		
	Mask*	گاما	مقدار V_{ratio}
t+1	111010101011	28×10^{-7}	0.028
t+2	101111000111	66×10^{-6}	0.002
t+3	1010110100011	18×10^{-4}	0.042

* Mask [$H_t, H_{t-1}, H_{t-2}, H_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3}, R_t, R_{t-1}, R_{t-2}, R_{t-3}, Q_t$]

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (H_{Pi} - H_{Oi})^2}{N}}$$

(رابطه ۹)

$$R = \frac{(\sum_{i=1}^N (H_{Oi} - \bar{H}_O)(H_{Pi} - \bar{H}_P))^2}{\sum_{i=1}^N (H_{Oi} - \bar{H}_O)^2 \sum_{i=1}^N (H_{Pi} - \bar{H}_P)^2}$$

(رابطه ۱۰)

$$NSE = 1 - \frac{\sum_{i=1}^N (H_{Oi} - H_{Pi})^2}{\sum_{i=1}^N (H_{Oi} - \bar{H}_O)^2}$$

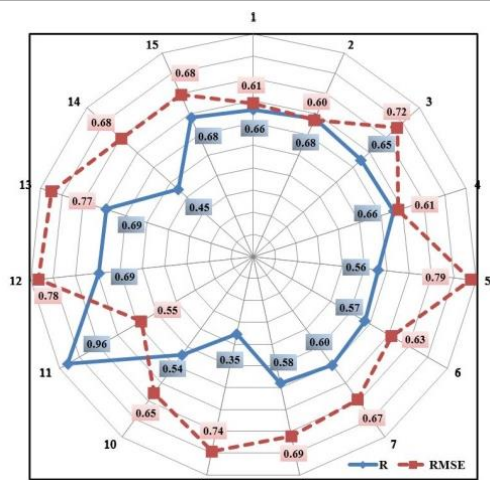
در روابطه فوق H_{Oi} نتایج مشاهداتی، H_{Pi} نتایج محاسباتی و N تعداد کل مشاهدات است. بهینه ترین جواب برای مدل هنگامی ایجاد خواهد شد که RMSE و MAE به سمت صفر و R و NSE به سمت یک میل کند.

نتایج و بحث

یکی از عوامل ایجاد خطا در مدل داده مبنا نبود رابطه هموار بین مجموعه داده‌های ورودی و خروجی می‌باشد و نیاز است ترکیب ورودی بهینه تعیین شود (قبایی سوق و همکاران، ۱۳۸۹). یکی از روش‌های مناسب در تعیین ترکیب ورودی بهینه، آزمون گاما (GT) است. در آزمون گاما اگر تعداد پارامترهای ورودی زیاد باشند، تعداد فراوانی ترکیب معنادار جهت محاسبه آماره گاما و انتخاب ترکیب مناسب وجود خواهد داشت. به طور کلی اگر N پارامتر ورودی بر وقوع پدیده‌ها موثر باشند، تعداد $2^N - 1$ ترکیب معنادار از پارامترهای ورودی بوجود می‌آید. در این مقاله با ۱۳ پارامتر ورودی و وجود بیش از ۸۱۹۰ ترکیب معنادار، بررسی تک تک ترکیبات (ماسک) ایجاد شده و محاسبه مقدار آماره گاما برای یافتن بهترین ترکیب کاری بسیار وقت‌گیر و غیرممکن است. بنابراین هنگامی که عوامل موثر بر پدیده‌ها بطور قابل ملاحظه‌ای زیاد باشد، ضروری است از روش‌های تلفیقی برای تعیین اهمیت پارامترهای ورودی و بهترین ترکیب از میان ترکیبات معنادار استفاده کرد. لذا بعد از نرمال کردن پارامترهای ورودی، از قابلیت‌های الگوریتم آموزش ژنتیک و تلفیق آن با آزمون گاما (GA-GT) استفاده شد تا مدل بطور خودکار و با کمترین زمان تمامی ترکیبات معنادار را مورد ارزیابی قرار دهد و بهترین ترکیب ورودی با کمترین مقدار آماره گاما (Γ) را جستجو و معرفی کند. الگوریتم ژنتیک به عنوان یک ابزار بهینه-ساز دارای خاصیت تکاملی است و چنانچه خطای موجود در جواب به دست آمده پذیرفته نشود، الگوریتم تا رسیدن به جواب قانع کننده ادامه می‌یابد. این الگوریتم با انجام عملیات مشخص ژنتیکی نیاز به پارامترهایی همچون انتخاب و اندازه جمعیت، نرخ ترکیب و جهش به سمت تکامل دارد که می-بایست به الگوریتم معرفی شود. در این مقاله جهت انتخاب بهترین ترکیب ورودی (ماسک) با استفاده از الگوریتم آموزش ژنتیک از مینیمم کردن تابع زیر استفاده شد (Wilson et al., 2004).

(رابطه ۱)

$$f(M) = 1 - ((w_1 f_1(M)) + (w_2 f_2(M)) + (w_3 f_3(M)))$$



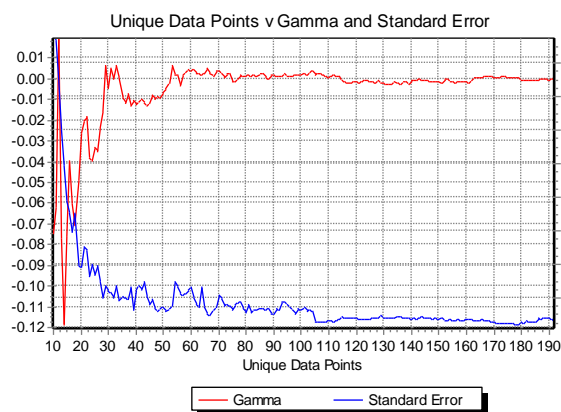
شکل ۶. نمودار راداری مقادیر RMSE و R حداکثرگره میانی ممکن در مدل MLP در افق یک ماه پیش رو.

در شبیه‌سازی سطح آب زیرزمینی با مدل ANFIS از سیستم استنتاج فازی عصبی از نوع GenFIS2 که از الگوریتم خوشه بندی کاهشی بهره می‌گیرد، استفاده شد. ساختارهای مختلف از طریق روش سعی و خطا با به کار بردن شعاع‌های تأثیر متفاوت اجرا شد که با توجه به نتایج حاصله شعاع بهینه برای داشتن بهترین ساختار سیستم استنتاج تطبیقی فازی-عصبی مقدار ۰/۵۲ می‌باشد که قادر است سطح آب زیرزمینی در گام زمانی اول را با R برابر با ۰/۹۷ و RMSE برابر با ۰/۴ شبیه سازی کند. در طراحی ساختار مدل SVR همانطور که قبلاً گفته شد از تابع RBF و انتخاب پارامترهای بهینه C, γ از روش سعی و خطا استفاده شد. به این صورت است که با ثابت در نظر گرفتن یک پارامتر، مقدار پارامتر دیگر را تغییر داده تا براساس کمترین مقدار MSE بهترین مقدار پارامتر انتخاب شود. ساختارهای بدست آمده برای مدل‌های MLP, ANFIS و SVR در (جدول ۶) ارائه شده است. با کدنویسی در نرم افزار MATLAB شبیه‌سازی سطح آب زیرزمینی هشتگرد در سه افق زمانی پیش رو با هر یک از مدل‌های مذکور انجام گرفت که نتایج در (جدول ۷) نشان داده شده است.

براساس (جدول ۷)، در مرحله آزمون، هر یک از مدل‌های MLP, ANFIS و SVR دارای همبستگی بالا و میانگین مربعات خطای پایین با داده‌های واقعی می‌باشند که بیانگر قابلیت مناسب این مدل‌ها در پیش بینی سطح آب زیرزمینی در هر کدام از سه افق زمانی پیش رو ($t+1, t+2, t+3$) است. همچنین شبیه‌سازی سطح آب زیرزمینی در افق زمانی اول هر سه مدل دقیق‌تر از افق‌های زمانی دوم و سوم می‌باشد. به‌طوری‌که به‌عنوان مثال در مدل MLP ضریب همبستگی با داده‌های واقعی سطح آب زیرزمینی از افق-های زمانی اول تا سوم به ترتیب از مقدار ۰/۹۶ تا ۰/۷۱ کاهش و مقدار میانگین مربعات خطا به ترتیب از مقدار ۰/۵۵ تا ۱/۴۴ افزایش می‌یابد.

در اولین افق زمانی پیش‌رو ($t+1$)، مدل ANFIS در مرحله آزمون با داشتن ضریب همبستگی بالای ۹۷ درصد و کمترین میانگین مربعات خطا ($RMSE < 0.4$) و مقدار NSE بیشتر از ۰/۹۳، دقت بالایی در پیش‌بینی دارد. مدل‌های SVR و MLP هر چند بیش از ۹۶ درصد با داده‌های واقعی همبستگی دارند، اما از مقدار NSE کمتر ($NSE = 0.88$) و مقدار میانگین

از روش M-test برای تعیین طول بهینه داده‌های ورودی مورد نیاز برای آموزش مدل استفاده شد. آزمون M-test در واقع نموداری است که تغییرات مقدار آماره گاما و خطای استاندارد شده در مقابل تعداد کل داده‌های ورودی را نشان می‌دهد. در نقطه‌ای که نمودارهای مقدار آماره گاما و خطای استاندارد شده تقریباً به حالت افقی و پایدار برسد آن نقطه تعیین کننده تعداد نقاط کافی جهت استفاده در مرحله آموزش مدل می‌باشد (Remesan and Mathew, 2015). شکل ۵ نتایج آزمون M-test را با استفاده از متغیرهای استفاده شده در ترکیب منتخب برای گام زمانی اول نشان می‌دهد. طبق این شکل، نمودار آماره گاما از نقطه ۱۲۰ به بعد تقریباً هر دو نمودار هموار می‌شوند و نشان‌دهنده کفایت طول داده‌ها در مرحله آموزش در این نقطه است. با استفاده از آزمون M-test برای طول بهینه داده‌های ورودی برای مدل شبیه‌سازی گام‌های زمانی دوم و سوم نیز تعیین و ارائه شد (جدول ۵).



شکل ۵. نمودار تغییرات آماره گاما (M-test) در گام زمانی اول.

جدول ۵. طول بهینه دسته داده های آموزش مدل به روش M-test.

پارامتر	افق زمانی پیش رو		
	t+1	t+2	t+3
طول بهینه دسته داده ورودی	۱۲۰	۱۴۵	۱۴۰
درصد	۶۳٪	۷۶٪	۷۳٪

با تعیین ترکیب بهینه ورودی و طول بهینه داده‌های ورودی، شبیه سازی سه افق زمانی پیش رو سطح آب زیرزمینی دشت هشتگرد شامل ($t+1, t+2, t+3$) با استفاده از مدل‌های ANFIS, MLP و SVR صورت گرفت. در طراحی شبکه عصبی MLP از الگوریتم لونیگ-مارکوت جهت آموزش و از (رابطه ۲) جهت حداکثرگره میانی استفاده شد. با توجه به تعداد متغیرهای ورودی در ترکیب بهینه تعداد حداکثر ۱۵ گره میانی در افق زمانی اول و دوم و حداکثر ۱۲ گره میانی در افق زمانی سوم پیش رو مشخص و با روش سعی و خطا با حداقل شدن مقدار خطا (RMSE) و حداکثر شدن ضریب تبیین (R) مناسب‌ترین گره تعیین شد که شکل ۶ نمودار راداری مقادیر RMSE و R مدل در گره‌های مختلف در افق یک ماه پیش‌رو را نشان داده است.

باشد بیانگر بیشتر بودن مقدار محاسبه شده در قیاس با داده واقعی است و مدل برآورد بالاتری نسبت به واقعیت ارائه می‌دهد. به همین ترتیب اگر شاخص کمتر از صفر باشد مدل برآورد کمتری از واقعیت دارد. پارامترهای آماری شاخص DDR برای هر یک از مدل‌های اجرا شده در مرحله آزمون محاسبه و در (جدول ۸) ارائه شده است. میانگین مقادیر شاخص DDR نشان می‌دهد که مدل ANFIS در افق زمانی اول پیش رو و مدل SVR در افق زمانی اول تا سوم پیش رو دارای مقدار مثبت هستند و در پیش بینی سطح ایستایی برآورد بالاتری نسبت به سطح ایستایی واقعی دارند. میانگین شاخص DDR در افق‌های سه گانه مدل MLP و افق دوم و سوم پیش رو در مدل ANFIS مقدار منفی بوده و مبین برآورد کمتر مدل از داده‌های واقعی است. کمترین مقدار انحراف معیار متعلق به مدل ANFIS در افق زمانی اول پیش رو است. برای درک بیشتر و بهره‌مندی گرافیکی از این شاخص در ارزیابی عملکرد مدل‌ها، از توزیع گوسین (نرمال) استفاده گردید (Noori et al., 2011). بدین منظور مقادیر DDR در مرحله آزمون مدل‌ها در افق‌های زمانی مختلف محاسبه و استاندارد گردید (S_{DDR}). سپس با استفاده از تابع گوسین مقادیر S_{DDR} نرمال‌سازی و ترسیم شدند. هر مدلی که S_{DDR} آن در تابع گوسین از مقادیر بیشترین برخوردار باشد یعنی نمودار زنگوله‌ای آن بالاتر، کشیده‌تر و مقادیر S_{DDR} متمرکز در تقاطع محور عمودی و افقی باشد، بیانگر عملکرد و کارایی بالاتر آن مدل می‌باشد. بر این اساس عملکرد هر یک از مدل‌های ANFIS, MLP و SVR در پیش‌بینی افق-های ماهانه اول تا سوم سطح آب زیرزمینی مورد ارزیابی قرار گرفت و نمودارهای مربوط در (شکل ۷ تا ۹) ترسیم گردید. نمودارهای ترسیم شده از مقادیر نرمال شده S_{DDR} مدل MLP (شکل ۷) نشان می‌دهد که مدل MLP در شبیه‌سازی افق زمانی اول سطح آب زیرزمینی دشت هشترگرد کارایی و عملکرد بهتر و در افق زمانی دوم عملکرد پائین‌تری داشته است. در مدل ANFIS نمودارهای نرمال شده مقادیر S_{DDR} (شکل ۸) بیان می‌دارد که این مدل در افق زمانی اول شبیه‌سازی سطح آب زیرزمینی کارایی و عملکرد بهتری نسبت به افق زمانی دوم و سوم داشته است و با افزایش افق زمانی کارایی مدل ANFIS کاهش یافته است. شکل ۹ نشان می‌دهد که مدل SVR در شبیه‌سازی سطح آب زیرزمینی در افق زمانی دوم کارایی بالاتری نسبت به شبیه‌سازی افق زمانی سوم و اول داشته است. همچنین مدل SVR در افق زمانی اول عملکرد و کارایی ضعیفی از خود نشان داده است.

در ادامه بمنظور مقایسه میزان کارایی مدل‌ها با یکدیگر، حداکثر مقدار نرمال شده S_{DDR} هر یک از مدل‌ها استخراج و در (شکل ۱۰) ارائه گردید. بر این اساس مدل ANFIS در شبیه‌سازی افق زمانی اول پیش‌رو سطح آب زیرزمینی با حداکثر مقدار نرمال شده S_{DDR} برابر با ۲/۴۳ بهترین عملکرد و کارایی در بین مدل‌های اجرا شده در تمامی افق‌های زمانی مورد مطالعه داشته است. مدل MLP در شبیه‌سازی افق زمانی اول و مدل SVR در شبیه‌سازی افق زمانی دوم سطح آب زیرزمینی به ترتیب با حداکثر مقدار نرمال شده S_{DDR} برابر ۲/۱۸ و ۲/۱۴ در رتبه‌های بعدی قرار گرفته‌اند. مدل MLP در شبیه‌سازی افق زمانی دوم و مدل ANFIS در شبیه‌سازی افق زمانی سوم سطح آب زیرزمینی به ترتیب با حداکثر مقدار نرمال شده S_{DDR} برابر ۱/۵۱ و ۱/۶ ضعیف‌ترین کارایی و عملکرد را داشته‌اند. این

مربعات خطا بیشتری ($RMSE > 0.54$) برخوردار هستند که دقت کمتری در شبیه‌سازی سطح آب زیرزمینی اولین افق زمانی در مقایسه با مدل ANFIS دارند. دقت شبیه‌سازی مدل‌های ANFIS, MLP و SVR در افق زمانی دوم و سوم پیش‌رو ($t+2, t+3$) مشابه بوده بطوریکه مدل ANFIS نتایج قابل قبولی نسبت به مدل SVR و MLP داشته‌است.

جدول ۶ ساختار بهینه طراحی شده مدل‌ها در افق زمانی پیش رو

افق زمانی پیش رو	علامت	مدل داده‌مبنا		
		MLP	ANFIS	SVR
		(Input- Hidden- Output)	R	(C-γ)
اول	t+1	(۹-۱۱-۱)	۰.۵۲	(۵۶۷-۷۳۴)
دوم	t+2	(۹-۷-۱)	۰.۵۷	(۵۳۶-۵۹۳)
سوم	t+3	(۷-۶-۱)	۰.۶۵	(۴۷۶-۷۶۳)

جدول ۷. نتایج مرحله آموزش و آزمون مدل‌های داده مبنا در پیش‌بینی افق‌های زمانی مختلف.

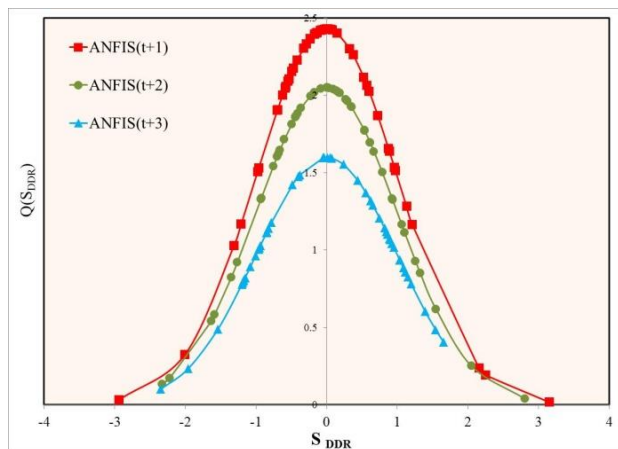
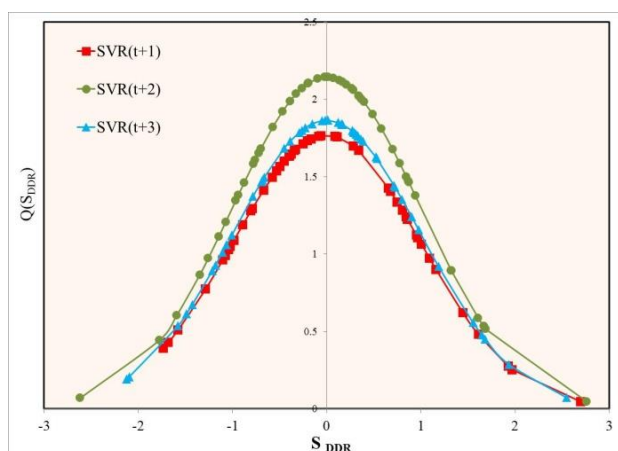
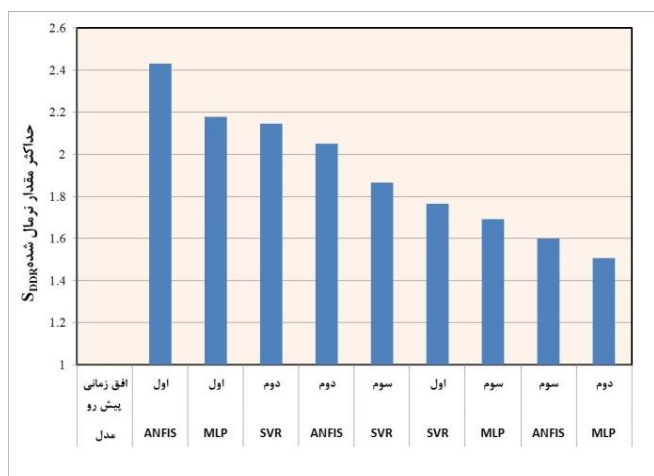
مدل	افق زمانی پیش رو	مرحله	معیار ارزیابی خطا			
			R	RMSE	MAE	NSE
MLP	t+1	آموزش	۰.۹۹	۰.۴۶	۰.۳۳	۰.۹۹
		آزمون	۰.۹۶	۰.۵۵	۰.۴۱	۰.۸۸
	t+2	آموزش	۰.۹۹	۰.۷۴	۰.۵۶	۰.۹۷
		آزمون	۰.۸۷	۰.۹۶	۰.۷۲	۰.۶۴
	t+3	آموزش	۰.۹۳	۰.۹۶	۰.۷۴	۰.۹۷
		آزمون	۰.۷۱	۱.۴۴	۱.۲۳	۰.۲۰
ANFIS	t+1	آموزش	۰.۹۹	۰.۲۴	۰.۱۶	۰.۹۹
		آزمون	۰.۹۷	۰.۴۰	۰.۲۹	۰.۹۳
	t+2	آموزش	۰.۹۹	۰.۴۷	۰.۳۴	۰.۹۹
		آزمون	۰.۹۱	۰.۸۶	۰.۶۹	۰.۷۰
	t+3	آموزش	۰.۹۸	۰.۹۲	۰.۶۹	۰.۹۶
		آزمون	۰.۸۹	۰.۸۱	۰.۷۰	۰.۷۵
SVR	t+1	آموزش	۰.۹۹	۰.۳۶	۰.۲۳	۰.۹۹
		آزمون	۰.۹۸	۰.۵۴	۰.۴۲	۰.۸۸
	t+2	آموزش	۰.۹۹	۰.۷۸	۰.۵۹	۰.۹۷
		آزمون	۰.۹۱	۰.۹۲	۰.۷۵	۰.۶۶
	t+3	آموزش	۰.۹۷	۰.۹۲	۰.۷۴	۰.۹۵
		آزمون	۰.۸۹	۰.۹۶	۰.۷۹	۰.۶۵

از آنجائیکه معیارهای ارزیابی RMSE, NSE و R^2 بیانگر متوسط مقدار خطای مدل بوده و نمی‌توانند نحوه رفتار و توزیع خطای مدل را مشخص کنند، بنابراین در ارزیابی میزان عملکرد مدل‌ها چندان کارایی ندارند. در واقع نحوه توزیع خطا می‌تواند در ارزیابی کارایی و عملکرد شبیه‌سازی مدل بسیار موثر باشد (Nourani et al., 2016). بنابراین در این بخش، برای اولین بار در مطالعات شبیه‌سازی سطح آب زیرزمینی از شاخص به‌نام DDR^{10} به منظور بررسی کارایی و عملکرد مدل‌های داده مبنا و نحوه توزیع تغییرات خطا در افق‌های زمانی پیش‌رو استفاده شد. شاخص DDR مطابق رابطه ۱۵ بدست می‌آید (Noori et al., 2011).

(رابطه ۱۵)

$$DDR = \left(\frac{H_{Pi}}{H_{Oi}} \right) - 1$$

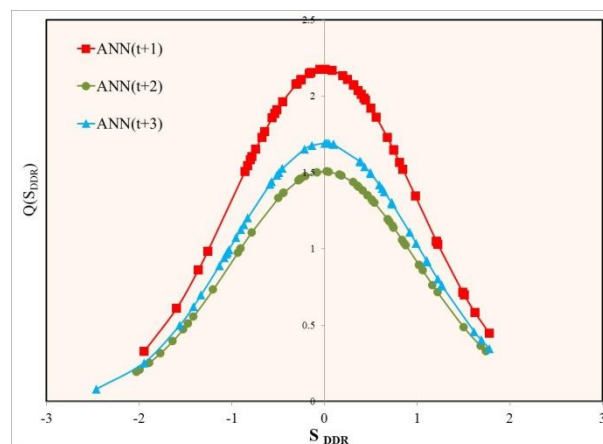
که در این رابطه H_{Pi} مقدار پیش‌بینی شده در مدل و H_{Oi} مقدار واقعی می‌باشد. اگر شاخص DDR برابر با صفر باشد، مقدار محاسبه شده در مدل دقیقاً برابر با مقدار واقعی می‌باشد. اگر شاخص DDR بیشتر از صفر

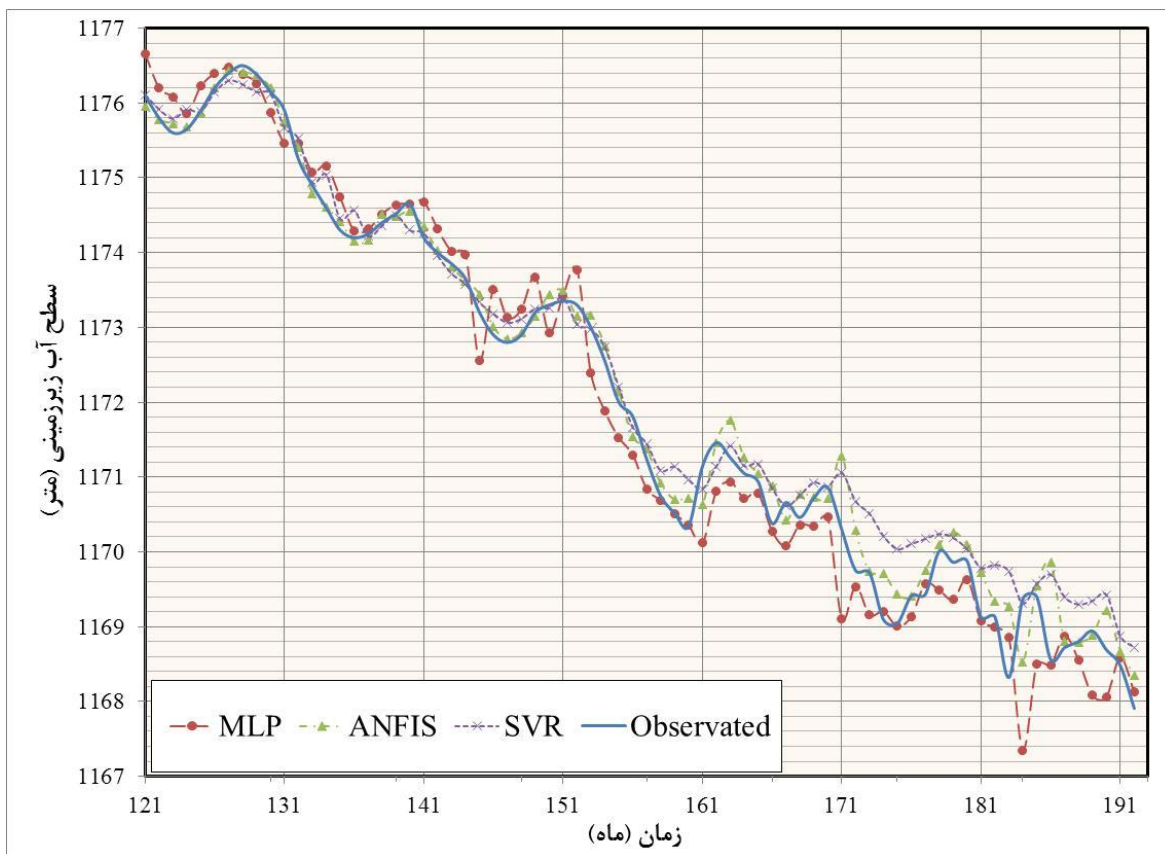

 شکل ۸. توزیع نرمال شده S_{DDR} مدل ANFIS در سه افق زمانی پیش‌رو.

 شکل ۹. توزیع نرمال شده S_{DDR} مدل SVR در سه افق زمانی پیش‌رو.

 شکل ۱۰. نمودار حداکثر مقدار نرمال شده S_{DDR} هر یک از مدل‌ها.

محاسبات نشان داد که عملکرد و کارایی مدل ANFIS و MLP در مقایسه با مدل SVR در شبیه‌سازی افق زمانی اول سطح آب زیرزمینی بهتر است اما هر چه افق زمانی جهت شبیه‌سازی افزایش یابد عملکرد SVR در مقایسه با دو مدل دیگر ارتقا پیدا می‌کند. در واقع با افزایش افق‌های زمانی پیش‌بینی مدل SVR عملکرد بهتری نسبت به MLP و ANFIS خواهد داشت که علت این امر را می‌توان به ساختار SVR و وجود تابع کرنل شعاعی در هسته پردازش آن دانست که امکان فراگیری کامل اطلاعات آب زیرزمینی اعم از فرآیندهای پیچیده و غیرخطی و ارتباط زمانی آنها را فراهم می‌کند. در (شکل ۱۱) مقادیر پیش‌بینی سطح ایستابی در مدل‌های ANFIS, MLP و SVR در گام زمانی اول در مرحله آزمون در مقایسه با داده‌های مشاهده‌ای را نشان می‌دهد.

جدول ۸. پارامتر آماری محاسبه شاخص DDR .

مدل داده مبنا	گام زمانی	$DDR (10^{-3})$			انحراف معیار
		بیشینه	کمینه	میانگین	
MLP	t+1	۰.۹۰	-۰.۵۰	۰.۰۱	۰.۴۰
	t+2	۰.۸۰	-۰.۷۰	۰.۰۱	۰.۵۳
	t+3	۱.۲۰	-۱.۲۰	۰.۲۰	۰.۶۰
ANFIS	t+1	۰.۷۹	-۰.۷۹	۰.۰۳	۰.۳۱
	t+2	۰.۵۰	-۰.۶۴	-۰.۰۴	۰.۴۵
	t+3	۰.۵۵	-۰.۵۳	۰.۰۹	۰.۵۶
SVR	t+1	۰.۸۰	-۰.۲۰	۰.۱۰	۰.۲۰
	t+2	۱.۰۰	-۰.۵۰	۰.۱۱	۰.۴۰
	t+3	۱.۶۰	-۰.۳۰	۰.۴۰	۰.۴۴


 شکل ۷. توزیع نرمال شده S_{DDR} مدل MLP در سه افق زمانی پیش‌رو.



شکل ۱۱. مقایسه سطح ایستابی مشاهده‌ای و شبیه‌سازی شده افق زمانی اول (t+1) مدل MLP, ANFIS و SVR در مرحله آزمون.

خلاصه و نتیجه گیری

امروزه در زمینه مدل‌سازی و پیش‌بینی سطح آب زیرزمینی که یک فرایند پیچیده و غیرخطی و متأثر از پارامترهای مختلف هیدروژئولوژی و هیدرولوژی و با عدم قطعیت می‌باشد، استفاده از روش‌های دانش مینا و عددی که دسترسی به تمام پارامترهای مورد نیاز آن‌ها مشکل و یا اندازه-گیری آن‌ها محتاج صرف هزینه و زمان زیادی می‌باشند، کمتر مورد توجه قرار گرفته و به ازای آن از مدل‌های هوش محاسباتی داده مینا که از دقت و اعتبار بالایی برخوردار هستند و پارامترهای ورودی کمتر و در دسترس‌تر نیاز دارند، مورد اقبال قرار گرفته است. در این پژوهش میزان دقت و کارایی سه مدل هوش محاسباتی داده مینا MLP, ANFIS و SVR در پیش‌بینی سه افق ماهانه سطح آب زیرزمینی دشت هشتگرد (t+1, t+2, t+3) مورد بررسی قرار گرفت. در ابتدا سری زمانی متغیرهایی که می‌توانند در نوسانات سطح آب زیرزمینی موثر باشند در یک بازه زمانی ۱۶ ساله (سال آبی ۸۰-۱۳۷۹ تا ۹۵-۱۳۹۴) جمع‌آوری گردید و سپس با بهره‌گیری از روش تحلیل تورم واریانس (VIF) متغیرهای همبسته شناسایی و حذف شدند و در نهایت متغیرهای سطح آب زیرزمینی (H) برحسب متر، بارش کل ماهانه (P) برحسب میلی‌متر، متوسط ماهانه دبی رودخانه (R) برحسب مترمکعب بر ثانیه و نرخ پمپاژ ماهانه (Pw) برحسب لیتر بر ثانیه تعیین شدند. با اعمال حداکثر سه تاخیر زمانی برای متغیرهای سطح آب زیرزمینی، بارندگی و دبی

رودخانه در نهایت ۱۳ متغیر شامل $(H_t, H_{t-1}, H_{t-2}, H_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_t, R_{t-1}, R_{t-2}, R_{t-3})$ به‌عنوان ورودی‌های مدل داده مینا استفاده شد. یکی از اقدامات مهم در مدل‌های داده مینا پیش پردازش داده‌های ورودی قبل از مدل‌سازی است که سعی شد با استفاده از روش الگوریتم آموزشی ژنتیک و تلفیق آن با آزمون گاما (GA-GT) بهینه‌ترین ترکیب از پارامترهای ورودی تعیین شود. بر این اساس بهترین ترکیب ورودی در افق زمانی اول و دوم با تعداد ۹ متغیر و در افق زمانی سوم بهترین ترکیب ورودی با هفت متغیر بدست آمد. در ادامه با بهره‌گیری از روش M-test طول بهینه داده‌های ورودی برای آموزش مدل‌ها تعیین شد. دقت مدل‌ها با معیارهای ارزیابی $RMSE, MAE, NSE$ و R^2 مورد ارزیابی قرار گرفت و نتایج مشخص کرد که هر سه مدل MLP, ANFIS و SVR در شبیه‌سازی سطح آب زیرزمینی از دقت مناسبی برخوردار هستند. در هر سه مدل، شبیه‌سازی ماه اول (t+1) در قیاس با ماه‌های دوم و سوم (t+2, t+3) دارای دقت بالاتری می‌باشد.

برای اولین بار در مطالعات شبیه‌سازی آب زیرزمینی، عملکرد و کارایی مدل‌ها از طریق شاخص DDR و ترسیم نمودار تابع گوسین (نرمال) مقادیر نرمال شده مورد ارزیابی قرار گرفت. بررسی مدل‌ها در نمودار گوسین نرمال شده نشان داد که مدل ANFIS در شبیه‌سازی افق زمانی اول سطح آب زیرزمینی بهترین عملکرد و کارایی در بین مدل‌های اجرا شده و تمامی افق-های زمانی مورد مطالعه داشته است. مدل MLP در شبیه‌سازی افق زمانی

SVR عملکرد بهتری نسبت به MLP و ANFIS خواهد داشت. لذا قابل استنباط است که مدل MLP و ANFIS در شبیه سازی افق های ماهانه نزدیکتر و SVR در افق های ماهانه دورتر قابلیت کاربرد بالاتری دارند.

دوم و مدل ANFIS در شبیه سازی افق زمانی سوم پیش رو سطح آب زیرزمینی ضعیف ترین کارایی و عملکرد را داشته اند. به طور کلی هر چه افق زمانی جهت شبیه سازی افزایش یابد عملکرد SVR در مقایسه با دو مدل دیگر ارتقا پیدا می کند. در واقع با افزایش افق های زمانی شبیه سازی مدل

منابع

- سالت پور.ع.الف، حاج عباسی.م.ع.، ایوبی.ش.۱۳۹۲، استفاده از آزمون گاما برای انتخاب ورودی های بهینه در مدل سازی مقاومت برشی خاک با استفاده از شبکه های عصبی مصنوعی، مجله پژوهش های حفاظت آب و خاک، جلد بیستم، شماره اول.
- رمضانی چرمهینه.ع.، ذونعمت کرمانی.م.، ۱۳۹۶، بررسی کارایی روش های رگرسیون بردار پشتیبان و شبکه عصبی پرسپترون چند لایه و رگرسیون خطی چند متغیره بمنظور پیش بینی تراز سطح آب زیرزمینی، پژوهشنامه مدیریت حوزه آبخیز، سال هشتم، شماره ۱۵.
- شرکت آب منطقه ای البرز، ۱۳۹۵، گزارش ممنوعیت دشت هشتگرد، شرکت مدیریت منابع آب ایران، وزارت نیرو.
- کامرانی.ص.، رضایی.م.، ۱۳۹۴، ارزیابی کیفی آب زیرزمینی در دشت لنجان با روش فازی تاکاگی - سوگنو، مجله زمین شناسی کاربردی پیشرفته، دوره ۵، شماره ۱۶، ص ۲۲-۲۴.
- کیا.ع. عمادی.ع.ر.، فضل اولی.ر.، ۱۳۹۴، بررسی امکان کاربرد سیستم استنتاج فازی - عصبی تطبیقی (ANFIS) در برآورد بار رسوب معلق بابلرود، پژوهشنامه مدیریت حوزه آبخیز، سال ششم، شماره ۳.
- قبائی سوق.م.، مساعدی.الف.، حسام.م.، هزارجریبی.الف.، ۱۳۸۹، ارزیابی تأثیر پیش پردازش پارامترهای ورودی به شبکه عصبی مصنوعی با استفاده از روشهای رگرسیون گام به گام و گاما تست به منظور تخمین سریعتر تبخیر و تعرق روزانه، نشریه آب و خاک، جلد ۲۴، شماره ۳، ص ۵۷-۶۹.
- نخعی.م.، صابری.الف.، ۱۳۹۱، پیش بینی نوسانات سطح آب زیرزمینی دشت قروه با استفاده از شبکه عصبی - موجکی و مقایسه آن با مدل عددی، مجله زمین شناسی کاربردی پیشرفته، دوره ۲، شماره ۴، ص ۴۷-۵۸.
- نخعی.م.، ودیعی.م.، ۱۳۹۱، کاربرد مدل استنتاج فازی در ارزیابی کیفیت آب قنات جهت مصارف شرب و کشاورزی (مطالعه موردی: استان تهران)، مجله زمین شناسی کاربردی پیشرفته، دوره ۲، شماره ۶، ص ۴۴-۵۲.
- Adamowski.J., Chan.HF., 2011, A Wavelet Neural Network Conjunction Model For Groundwater Level Forecasting, *Journal of Hydrology*, No:407, P:28-40
- Choubin.B., Malekian.A., 2017, Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions, *Environ Earth Science* No:17, P:670-80
- Daliakopoulos.I.N., Coulibaly.P., Tsanis.I.K., 2005. Groundwater level forecasting using artificial neural networks, *Journal Hydrology*, Vol:309, No:1-4, P:229-240
- Durrant.P.J., 2001, Wingamma A Non-Linear Data Analysis and Modelling Tool With Applications To Flood Prediction, PhD Thesis, Department of Computer Science, Cardiff University, Wales, UK
- Huang.G.B., 2003, Learning Capability and Storage Capacity of Two-Hidden-Layer Feed Forward Networks, *IEEE Transactions on Neural Networks*, No:14, p:274-281
- Kecman.V., 2005, Support Vector Machines: An Introduction, Theory and Applications, Springer-Verlag, New York. No:4, P:1-48
- Koncar.N., 1997, Optimisation Methodologies for Direct Inverse Neurocontrol, PhD Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London.
- Marquez. R., Coimbra .C.F.M., 2011, Forecasting of Global And Direct Solar Irradiance Using Stochastic Learning Methods Ground Experiments and The NWS Database, *Solar Energy*, No:85, P:746-56
- Moghaddam., A., Nadiri. A., 2008, An ANN-Based Model For Spatiotemporal Groundwater Level Forecasting, *Hydrology Process*, No: 22, P:5054-5066
- Nayak.P., Sudheer. K., Rangan.D., Ramasastri. K., 2004, A Neuro-Fuzzy Computing Technique For Modeling Hydrological Time Series, *Journal Hydrology*, No: 291, P:52-66
- Noori.R., Karbassi.A.R., Moghaddamnia.A., Zokaei-Ashtiani.M., Farokhnia.A., 2011, Assessment Of Input Variables Determination On The SVM Model Performance Using PCA, Gamma Test And Forward Selection Techniques For Monthly Stream Flow Prediction, *Journal of Hydrology*, No: 401, P:177-189
- Nourani.V., Mousavi.S., 2016, Spatiotemporal Groundwater Level Modeling Using Hybrid Artificial Intelligence-Meshless Method, *Journal of Hydrology*, No:536, P:10-25
- Remesan.R., Mathew.J., 2015, Hydrological Data-Driven Models (A Case Study Approach), Springer International Publishing Switzerland, 249P
- .B., Vafakhah.M., Moosavi.V., Moghaddamnia.A., 2016, Application of Several Data-Driven Techniques For Predicting Groundwater Level. *Water Resources Management*, No: 27, P:419-432

- Wilson.I., Jones.A., Jenkins.D., Ware.J., 2004, Predicting Housing Value: Genetic Algorithm Attribute Selection and Dependence Modeling Utilizing The Gamma Test, Applications of Artificial Intelligence in Finance and Economics, No:19, P:243-275
- Yu.H., Wen.X., Feng.Q., Deo.R.C, Si.J., Wu.M., 2018, Comparative Study of Hybrid-Wavelet Artificial Intelligence Models For Monthly Groundwater Depth Forecasting In Extreme Arid Regions, Northwest China. Water Resources Management, No:32, P:301-323
- Banerjee.P., Singh.V.S., Chattopadhyay.K., Chandra.P.C., Singh.B., 2011, Artificial neural network model as a potential alternative for groundwater salinity forecasting, Journal of Hydrology, No:398, P: 212-220
- Suryanarayana.C., Sudheer.C., Mahmood. V., Panigrahi.B., 2014, An Integrated Wavelet-Support Vector Machine For Groundwater Level Prediction In Visakhapatnam, India, Neurocomputing, No:145, P:324-335