دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشنامه مدیریت حوزه آبخیز

برآورد غلظت رسوب معلق روزانه با استفاده از شبکه عصبی مصنوعی و خوشه‌بندی داده‌ها به روش نگاشت خود سازمان‌ده (مطالعه موردی: ایستگاه هیدرومتری سیرا - رودخانه کرج)

محمودرضا طباطبائی[۱]، کریم سلیمانی[۲]، محمود حبیب نژاد روشن[۲] و عطاالله کاویان[۳]

۱- دانشجوی دکتری، دانشگاه علوم کشاورزی و منابع طبیعی ساری، (نویسنده مسوول: taba1345@hotmail.com)
۲ و ۳- استاد و دانشیار، دانشگاه علوم کشاورزی و منابع طبیعی ساری

تاریخ دریافت: ۹۳/۲/۱۲ تاریخ پذیرش: ۹۳/۶/۴

چکیده

امروزه برآورد دقیق بار رسوب معلق رودخانه‌ای از جنبه‌های مختلف مهندسی منابع آب، مسائل زیست‌محیطی و کیفیت آب از اهمیت ویژه‌ای برخوردار است. در این راستا، مدل‌های هیدرولوژیکی حوزه، به‌دلیل عوامل متعدد تاثیرگذار ثابت و متغیر، کارایی مناسبی در برآورد میزان رسوب معلق از خود نشان نداده‌اند. همچنین اغلب مطالعات شبیه‌سازی برآورد رسوب معلق، تنها بر مبنای دبی جریان خروجی حوزه استوار است که نتایج حاصله نیز، گواه بر عدم کارآیی مطلوب آنها است. این در حالی است که عوامل تاثیرگذاری همچون نوع بارش، فصل سال و شکل هیدروگراف جریان که نقش عمده‌ای در این فرآیند ایفا می‌نمایند در شبیه‌سازی برآورد میزان رسوب معلق نادیده گرفته شده‌اند. در پژوهش حاضر، از روش شبکه عصبی پرسپترون چند لایه و داده‌های آب و هواشناسی (دبی و غلظت رسوب معلق روزانه جریان، متوسط بارش و دمای روزانه) حوزه آبخیز سد کرج در یک دوره زمانی ۳۰ ساله (۱۳۶۰ تا ۱۳۹۰) به‌منظور برآورد غلظت رسوب معلق روزانه ایستگاه هیدرومتری سیرا استفاده شده است. در این روش، با توجه به نقش تغییرات فصلی و وضعیت جریان در تولید و انتقال رسوب حوزه، ابتدا بر اساس سه متغیر رژیم بارش، وضعیت هیدروگراف جریان و نوع رواناب حاصل از بارش، داده‌های مورد استفاده به ۵ گروه تفکیک و سپس برای هر گروه، مدل جداگانه‌ای طراحی گردید. همچنین به‌منظور افزایش قدرت تعمیم‌دهی مدل‌ها، از شبکه عصبی نگاشت خود سازمان‌ده (SOM) جهت خوشه‌بندی و از شاخص سیلهوت، در تعیین تعداد بهینه خوشه‌ها استفاده شد. نتایج پژوهش نشان داد که استفاده از متغیرهای بارش و دمای روزانه، به‌همراه دبی جریان و تفکیک زمانی داده‌ها، نقش مهمی در افزایش دقت برآورد رسوب رودخانه داشته است. در این رابطه، بیشترین خطای محاسبه شده در بین مدل‌ها زمانی است که برای تمامی فصول سال، تنها از یک مدل واحد، جهت برازش به داده‌ها استفاده می‌گردد. نتایج این پژوهش می‌تواند به‌عنوان الگوئی مناسب در برآورد رسوب معلق سایر رودخانه‌های کشور مورد استفاده قرار گیرد.

واژه‌های کلیدی: خوشه‌بندی، رسوب معلق، رودخانه کرج، سیرا، شبکه عصبی، نگاشت خود سازمان‌ده

مقدمه

بار رسوب معلق رودخانه‌ای که حدود ۷۵ تا ۹۵ درصد کل رسوب رودخانه را شامل می‌گردد، از جنبه‌های مختلف (نظیر مهندسی منابع آب، مسائل زیست‌محیطی، کیفیت آب و غیره) حائز اهمیت بوده و می‌تواند به‌عنوان شاخصی از وضعیت فرسایش خاک و شرایط اکولوژیک حوزه در نظر گرفته شود (۲۴). همچنین طراحی بهینه و عملکرد مناسب سازه‌های منابع آب، نظیر مخزن، سد و کانال، نیازمند تخمین دقیق از بار رسوبی رودخانه است (۹،۱). بطورکلی، روابط بین پارامترهای کیفیت آب رودخانه و فرآیندهای فیزیکی، ژئوشیمیائی و بیولوژیکی انجام شده بین منابع حوزه (خاک، پوشش‌گیاهی، زمین‌شناسی، کاربری اراضی و ...)، متغیرهای هواشناسی (دما، بارش، ذوب برف و ...)، متغیر هیدرولوژیکی رودخانه (دبی) و همچنین دخالت‌های انسانی، اغلب بسیار پیچیده، غیرقطعی و غیرخطی بوده بنحوی که درک کامل آنها را غیرممکن می‌سازد (۱۱). در این شرایط، استفاده از هوش محاسباتی (نظیر شبکه‌های عصبی مصنوعی، سیستم استنتاج فازی-عصبی و ...) ابزار مناسبی در شبیه‌سازی و برآورد متغیرهای کیفی آب رودخانه نظیر بار رسوب معلق محسوب می‌گردد. در این ارتباط، در دهه اخیر، محققین بدنبال استفاده از بکارگیری این شیوه‌های مدل‌سازی در زمینه برآورد بار رسوب معلق رودخانه‌ها بوده‌اند. اولکی و همکاران (۲۳)، با استفاده از مدل‌های نروفازی تطبیقی، منحنی سنجه رسوب، شبکه عصبی و رگرسیون چندمتغیره، مقدار رسوب معلق روزانه رودخانه گدیز[1] در ترکیه را برآورد نمودند. ورودی‌های مدل، شامل داده‌های دبی جریان (روزانه و یک روز قبل)، بارش (روزانه و یک روز قبل) و مقدار رسوب روزانه بود. مقایسه نتایج مدل‌ها نشان داد که روش‌های مبتنی بر هوش محاسباتی (نروفازی و شبکه عصبی)، برآوردهای دقیق‌تری نسبت به روش‌های رگرسیون و منحنی سنجه رسوب داشته‌اند. مصطفی و همکاران (۱۷)، به منظور برآورد دبی رسوب معلق رودخانه پری[2] در مالزی، از یک شبکه عصبی پرسپترون چند لایه استفاده نمودند. آنها به منظور آموزش شبکه عصبی، از چهار روش مختلف کاهش شیب[3]، کاهش شیب با مومنتوم[4]، گرادیان توام[5] و لونبرگ مارکواردت[6] (LM) بهره گرفتند. نتایج نشان داد که روش‌های LM و گرادیان توام، عملکرد بهتری نسبت به سایر روش‌های آموزش داشته است. با این حال، روش LM، به لحاظ سرعت زمانی، بسیار سریع‌تر از روش گرادیان توام سبب همگرائی شبکه شده است. کاکائی لفدانی و همکاران (۶)، توانائی مدل‌های شبکه عصبی و ماشین بردار پشتیبان در برآورد رسوب معلق روزانه رودخانه دویرج واقع در غرب ایران را مورد بررسی قرار دادند. آنها از داده‌های بارش و دبی جریان به‌عنوان ورودی و از دبی رسوب به‌عنوان خروجی استفاده نمودند. بهترین ورودی‌ها برای هر دو مدل با استفاده از الگوریتم ژنتیک و آزمون گاما تعیین گردید. نتایج، حاکی از برتری مدل‌های یاد شده نسبت به مدل‌های رگرسیونی بود. ملسی و همکاران (۱۵)، از مدل شبکه عصبی پرسپترون چند

1- Gediz 2- Pary 3- Gradient Descent
4- Gradient Descentwith Momentum 5- Scaled Conjugate Gradient 6- Levenberg Marquardt (LM)

لایه به‌منظور پیش‌بینی مقدار رسوب معلق سه رودخانه بزرگ در آمریکا (می‌سی‌سی‌پی[۱]، میسوری[۲] و ریوگراند[۳]) استفاده نمودند. داده‌های ورودی مدل، دبی‌های جریان (روزانه و روز ما قبل)، بارش روزانه و مقدار رسوب روز قبل در نظر گرفته شد. نتایج نشان داد که دقت استفاده از شبکه عصبی در برآورد بار رسوب رودخانه‌ها بیشتر از روش‌های رگرسیونی (خطی و غیرخطی) و مدل ARIMA بوده است. رجائی و همکاران (۱۹) در تحقیقی، کارآیی مدل‌های شبکه عصبی مصنوعی، مدل ترکیبی موجک- شبکه عصبی، رگرسیون چندمتغیره و منحنی سنجه رسوب را جهت برآورد رسوب معلق روزانه در رودخانه آیووا[۴] ایالات متحده آمریکا مورد بررسی قرار دادند. نتایج تحقیق، کارآیی بهتر مدل ترکیبی موجک- شبکه عصبی را نسبت به مدل‌های دیگر نشان داد. دمیرسی و بالتاسی (۳) در تحقیقی در ایستگاه ساکری‌منتو فری‌پورت[۵] در آمریکا، از روش منطق فازی، رگرسیونی و منحنی سنجه رسوب جهت برآورد مقدار غلظت رسوب معلق استفاده نمودند. در این تحقیق از داده‌های دبی جریان، غلظت رسوب معلق و دمای آب که بطور پیوسته در مدت زمان ۵ سال تهیه شده بود جهت برآورد مقدار غلظت رسوب معلق استفاده گردید. نتایج تحقیق حاکی از برتری دقت مدل فازی نسبت به سایر روش‌ها بود. زو و همکاران (۲۴)، با بکارگیری شبکه عصبی و استفاده از داده‌های هواشناسی (متوسط بارش و دمای ماهانه) و متوسط دبی جریان ماهانه، دبی رسوب رودخانه لانگ‌چوآنگ‌جیانگ[۶] در حوزه یانگ تسه در چین را برآورد نمودند. نتایج این تحقیق، عملکرد بهتر این روش را نسبت به روش‌های رگرسیونی مورد استفاده، نشان داد.

خوشه‌بندی و نمونه‌گیری از آنها، نقش مهمی در ساخت مجموعه داده‌های همگن و مشابه (نظیر مجموعه داده‌های واسنجی[۷]، اعتبارسنجی[۸] و آزمون[۹]) جهت استفاده در مدل‌های داده مبنا[۱۰] (نظیر مدل‌های رگرسیون، شبکه‌های عصبی، نروفازی) دارند. عدم استفاده از داده‌های مشابه و همگن در سه بخش یاد شده، تاثیرات بسیار مستقیم در میزان دقت و کارائی نهائی مدل‌های طراحی شده داشته، سبب کاهش قدرت تعمیم‌دهی آنها خواهد شد (۱۴). متاسفانه در مدل‌سازی با استفاده از شبکه‌های عصبی و سایر روش‌های داده مبنا، کمتر به این مسئله توجه شده و مدل‌سازی عمدتاً در راستای طراحی بهینه شبکه‌های عصبی نظیر انتخاب تعداد لایه‌های پنهان، تعداد نورون‌ها، مقایسه روش‌های آموزش شبکه و غیره بوده در حالیکه، تمامی موارد یاد شده می‌تواند به‌طور قابل ملاحظه‌ای تحت شعاع تعداد و نوع داده‌های استفاده شده در سه مجموعه یاد شده قرار گیرد (۲۲،۲). در این راستا در تحقیق حاضر، به‌منظور طبقه‌بندی داده‌ها در سه دسته مشابه و همگن، از روش خوشه‌بندی نگاشت خود سازمان‌ده SOM[۱۱] استفاده شده است. در این رابطه، نور و همکاران (۱۸) به‌منظور برآورد دبی جریان روزانه، مقدار رسوب معلق و فسفر موجود در رودخانه‌های

1- Mississippi
2- Missouri
3- Rio Grande
4- Iowa
5- Sacremento Freeport
6-ngchuanjiang
7- Calibration
8- Cross Validation
9- Test
10- Data Driven
11- Self-Organizing Map

دو حوزه جنگلی در شمال کانادا، از شبکه عصبی مصنوعی و خوشه‌بندی به روش SOM استفاده نمودند. در تحقیقی مشابه، لی و همکاران (۱۱)، از SOM و شبکه عصبی در برآورد میزان ازت خارج شده از ۵ حوزه جنگلی در شمال کانادا استفاده نمودند. بوودن و همکاران (۲) به‌منظور پیش‌بینی میزان نمک موجود در آب رودخانه موری[1] در جنوب استرالیا، از SOM و شبکه عصبی استفاده کردند. نتایج تحقیق، برتری کارائی مدل شبکه عصبی با داده‌های خوشه‌بندی شده در مقایسه با زمانی که داده‌ها به‌صورت تصادفی طبقه‌بندی می‌شوند را نشان داد. در رابطه با نمونه‌گیری از داده‌های خوشه‌بندی شده به روش SOM، روش‌های مختلفی از سوی محققین پیشنهاد گردیده که مبنای کلی این روش‌ها بر اساس نمونه‌گیری تصادفی یکنواخت از خوشه‌ها می‌باشد و تعداد نمونه‌ها، بر اساس سه قاعده تخصیص برابر، تخصیص متناسب و تخصیص Neyman[2] محاسبه می‌شود (۱۴). در رابطه با تاثیر تغییرات فصلی بر مقدار رسوب معلق رودخانه‌ها، تحقیقات انجام شده نشان می‌دهد که مقدار رسوب معلق رودخانه‌ای، تحت اثر رژیم بارش، شرایط فصلی و بطور کلی شرایط هیدرواقلیمی حوزه بوده و دارای تغییرات و نوسانات شدید در طی ماه‌ها و فصول مختلف دوره آماری است (۲۰). در این رابطه مساعدی و همکاران (۱۶)، جهت تعیین مدل‌های بهینه منحنی‌سنجه‌ی رسوب در سه ایستگاه هیدرومتری تمر، گنبد و قزاقلی (واقع در رودخانه گرگانرود)، از معیارهای هیدرولوژیکی و اقلیمی (نظیر وضعیت هیدروگراف، کلاسه‌بندی دبی و زمان اندازه‌گیری جریان) جهت تفکیک داده‌ها استفاده نمودند. نتایج آنها نشان داد که زمان نمونه‌برداری رسوب و شرایط هیدروگراف جریان، نقش مهمی در انتخاب مدل بهینه دارد. ذرتی‌پور و همکاران (۲۵)، کلاسه‌بندی داده‌ها به فصول مختلف (سیلابی، خشک و تر) را سبب افزایش دقت برآورد مدل منحنی سنجه‌ی رسوب در ایستگاه گلینک حوزه آبخیز طالقان می‌دانند. قورقی و همکاران (۴)، مدل‌های مختلفی از منحنی سنجه‌ی رسوب را برای ایستگاه‌های هیدرومتری واقع در رودخانه‌های تلوار و چم‌شور (حوزه سفیدرود در استان کردستان) ارائه می‌نمایند که مبنای اصلی آنها تفکیک داده‌ها بر اساس شرایط اقلیمی و هیدرولوژیکی حوزه بوده و نتیجه‌گیری می‌نمایند که مدل‌سازی برآورد بار رسوب معلق رودخانه‌ای بر اساس تفکیک داده‌ها، سبب همگنی داده و افزایش دقت برآورد مقدار رسوب می‌گردد. با توجه به آنچه گفته شد، اهداف و نوآوری‌های این تحقیق را می‌توان به شرح ذیل خلاصه نمود:

الف- برآورد غلظت رسوب معلق روزانه ایستگاه هیدرومتری سیرا (واقع در رودخانه کرج) با طراحی و ساخت مدل‌های مختلف شبکه عصبی بر پایه تفکیک زمانی داده‌های حوزه.

ب- بررسی نقش بکارگیری متغیرهای بارش و دما، در مقدار کارائی مدل‌ها. در تحقیقات مشابه، دما ندرتاً استفاده شده است.

ج- تعیین موثرترین ترکیب متغیرهای ورودی مدل‌ها.

1- Murray

2- Neyman Allocation

د- تهیه مجموعه‌های مشابه و همگن از داده‌ها با بکارگیری خوشه‌بندی به روش SOM و شاخص سیلهوت.

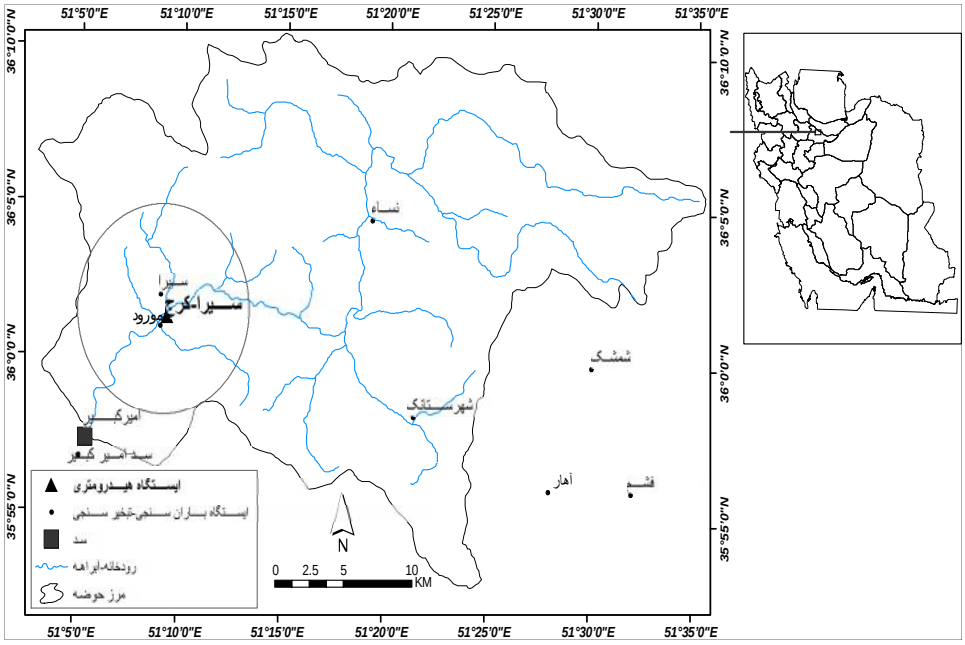
پیاده‌سازی و استفاده از خوشه‌بندی به روش SOM، از دیگر نوآوری‌های این تحقیق بوده که در سایر تحقیقات مرتبط با برآورد رسوب معلق، بندرت مورد توجه قرار گرفته است.

مواد و روش‌ها

در این پژوهش، به‌منظور طراحی و کدنویسی مدل‌های شبکه عصبی، پیاده‌سازی الگوریتم‌های خوشه‌بندی و شاخص‌های تعیین تعداد بهینه خوشه‌ها، از نرم‌افزار MATLAB نسخه ۷/۱۱ و جهت انجام تحلیل‌های آماری از نرم‌افزار SPSS نسخه ۱۹ و همچنین نرم‌افزار MATLAB استفاده شده است.

منطقه مورد مطالعه و داده‌های مورد استفاده

منطقه مورد مطالعه در حوزه آبخیز رودخانه کرج در مختصات جغرافیائی '۳۵ °۵۱ تا '۳ °۵۱ طول شرقی و '۱۱ °۳۶ تا '۵۳ °۳۵ عرض شمالی در فاصله ۳۰ تا ۶۰ کیلومتری شمال و شمال غرب استان تهران واقع می‌باشد (شکل ۱).



شکل ۱- موقعیت حوزه آبخیز کرج و ایستگاه هیدرومتری سیرا

مساحت حوزه (مساحت حوزه بالا دست ایستگاه سیرا) ۶۵۰۶۳ هکتار و ارتفاع متوسط آن، ۲۸۲۷ متر از سطح دریا می‌باشد. حوزه آبخیز سد کرج از شمال به حوزه آبخیز چالوس، هراز، طالقان رود، از غرب به حوزه آبخیز کردان و از شرق به حوزه‌های رودخانه‌های منتهی به تهران و رودخانه جاجرود محدود می‌شود. کلیه سرشاخه‌های رودخانه کرج از ارتفاعات البرز سرچشمه می‌گیرند. خاک‌های منطقه، دارای ضخامتی

متغیر بوده و شامل خاک‌های جوان و تکامل نیافته آنتی‌سول تا خاک‌های با تکامل متوسط اینسپتی‌سول با مواد مادری مختلف می‌باشند (۲۱). آمار مورد استفاده در این تحقیق، شامل داده‌های آب‌سنجی (دبی لحظه‌ای، دبی متوسط روزانه و غلظت رسوب روزانه) ایستگاه هیدرومتری سیرا و داده‌های بارش و دمای روزانه ایستگاه‌های باران‌سنجی و تبخیرسنجی تماب در یک دوره زمانی ۳۰ ساله (۱۳۶۰ تا ۱۳۹۰) به تعداد ۶۱۰ رکورد اطلاعاتی می‌باشد. خصوصیات آماری داده‌های مورد استفاده در این مدت، در جدول (۱) ارائه شده‌است. همانطور که از اطلاعات آماری جدول ۱ استنباط می‌شود، غلظت رسوب دارای چولگی و ضریب تغییرات زیاد بوده و تغییرات بین حداقل و حداکثر آن بسیار زیاد می‌باشد. این موضوع، به همراه سایر آماره‌های محاسبه شده، حکایت از پیچیدگی مدل‌سازی برآورد رسوب معلق رودخانه دارد.

جدول ۱- ویژگی‌های آماری داده‌های مورد استفاده (۱۳۶۰ تا ۱۳۹۰)

متغیر / خصوصیت آماری	دبی لحظه‌ای (m^3)	غلظت رسوب معلق (mg/l)	متوسط بارش روزانه (mm)	متوسط دمای روزانه (°c)
حداکثر	۱۳۶/۱۷	۱۱۲۰۱/۶۷	۶۶/۵۶	۲۲/۶۵
حداقل	۲/۶۳	۲	۰	-۲۵/۷۸
میانگین	۱۷/۱۴	۴۹۳/۶۴	۵/۰۱	۰/۲۴
انحراف معیار	۱۶/۹۱	۱۱۳۷/۴۷	۱۰/۲	۹/۰۶
چولگی	۲/۲۷	۵/۰۱	۷/۰۸	-۰/۱۸
ضریب تغییرات	۰/۹۹	۲/۳۰	۲/۰۳	۰/۳۷

شبکه‌های عصبی مصنوعی

در این تحقیق، از دو نوع شبکه عصبی با ناظر (پرسپترون چند لایه) و بدون ناظر (نگاشت خود سازمان‌ده) به ترتیب جهت مدل‌سازی رسوب و خوشه‌بندی داده‌ها استفاده شده که در ذیل به اختصار به آنها اشاره می‌شود:

الف- شبکه عصبی پرسپترون رو به جلو[1]

این شبکه، از عناصر عملیاتی ساده‌ای به نام نورون ساخته شده که به صورت موازی در کنار هم، عمل می‌کنند. شبکه‌های عصبی چند لایه، از یک لایه ورودی (در ارتباط با داده‌های ورودی)، یک یا چند لایه پنهان (جهت سازماندهی نورون‌ها) و یک لایه خروجی (در رابطه با داده خروجی) تشکیل می‌شوند. عملکرد شبکه عصبی، از طریق نحوه اتصال بین اجزاء، با تنظیم مقادیر هر اتصال که به نام وزن اتصال بیان می‌شود، تعیین می‌گردد. یکی از انواع شبکه‌های عصبی پرکاربرد در هیدرولوژی و منابع آب، شبکه‌های عصبی پرسپترون چند لایه رو به جلو با الگوی آموزش پس انتشار[2] خطا است (۲۳). در این نوع از شبکه‌های عصبی، جهت جریان داده، از لایه ورودی به سمت لایه پنهان و از لایه پنهان به سمت لایه

1- Feed-forwardMulti-layerPerceptron (FFMLP)

2- Back Propagation

خروجی بوده و از این نظر بـه آنهـا شـبکه‌هـای عصبی رو به جلو یـا پیشـخور گفتـه مـی‌شـود (۲۲). به منظور آموزش شـبکه عصـبی، مقـدار خطا در جهت بیشترین شیب تابع خطا محاسبه شده و این مقدار، به لایـه‌هـای قبـل (لایـه یـا لایه‌های پنهـان) فرسـتاده شـده تـا بـا تنظـیم مجـدد مقـادیر وزن نـورون‌هـا، مقـدار خطـا را کاهش دهند (قانون دلتا) (۲۲) رابطه‌ی ۱:

$$w_{ij}^{new} = w_{ij}^{old} - \eta \frac{\partial E}{\partial wij} \qquad (۱)$$

که در آن: w_{ij}^{old} و w_{ij}^{new}، به‌ترتیـب وزن بـین نورون‌های i و j قبل و بعد از یک تکرار معـین، η نرخ یادگیری و E تابع خطا می‌باشد.

آمـوزش شـبکه و کـاهش خطـا تـا ایجـاد همگرائی در شبکه ادامه مـی‌یابـد. شـبکه‌هـای عصـبی مـی‌تواننـد دارای چنـدین لایـه پنهـان باشند با این وجود، تحقیقات انجام شـده نشـان می‌دهد که شبکه‌های عصـبی پیشـخور، بـا دارا بودن یک لایه پنهان، قادر بـه تقریـب زدن هـر نـوع تـابع غیرخطـی مـی‌باشـند (۵). در ایـن تحقیق، از یک شبکه عصبی پرسپترون سه لایه (یک لایه ورودی، یک لایه پنهـان و یـک لایـه خروجی) استفاده شده و تعداد بهینه نورون‌ها با سعی و خطا تعیین گردید. همچنین بـه‌منظـور آمـــوزش شـــبکه عصـــبی از روش لـــونبرگ مارکواردت به‌دلیل کارائی و همگرائی سـریع‌تـر در آمـوزش شـبکه و اثبـات کارامـدی آن در تخمـین رسـوب رودخانـه‌ای (۲۳،۹) اسـتفاده گردید. توابع فعال سـازی در نـورون‌هـای لایـه پنهان و لایه خروجی، بـه ترتیـب سـیگوئید یـا لوگ سیگموئید و خطی در نظر گرفته شد.

ب- شبکه عصبی نگاشت خود سازمان‌ده (SOM)

شبکه عصبیSOM، یک شبکه عصبی مصنوعی غیرنظارتی بوده و الگوریتم آموزش آن، به صورت رقابتی و بدون ناظر انجام می‌شود. ساختار SOM، از یک لایه ورودی و یک لایه خروجی (لایه‌کوهنن) تشکیل می‌شود (۱۰). در این ساختار، نورون‌های لایه ورودی، محل ارتباط داده‌های ورودی با شبکه بوده و به ازاء هر متغیر ورودی، یک نورون در این لایه وجود دارد (به‌عنوان مثال دبی، بارش و...). لایه خروجی، ماتریس یا شبکه‌ای (عموماً یک شبکه دو بعدی) از نورون‌ها را تشکیل داده بنحوی که هر نورون این شبکه، به کلیه نورون‌های لایه ورودی متصل بوده ولی به نورون‌های دیگر این لایه متصل نمی‌باشد (۷). فرآیند آموزش در شبکه SOM از سه مرحله رقابت[1]، همکاری[2] و تطبیق[3] تشکیل می‌گردد. در فاز رقابت، با معرفی یک داده به شبکه، فاصله اقلیدسی این داده نسبت به نورون‌های لایه خروجی محاسبه شده و هر نورون از لایه خروجی که فاصله کمتری را نسبت به آن داشته باشد به‌عنوان نورون برنده یا شبیه‌ترین نورون به بردار ورودی BMU[4] انتخاب می‌گردد.

در شبکه SOM، مقدار فاصله اقلیدسی مطابق با رابطه‌ی ۲ محاسبه می‌شود:

$$j = 1, 2, \dots, M \qquad (۲)$$

$$D_j = |x - w_j| = \left[\sum_{i=1}^{N} (x_i - w_{ji})^2\right]^{\frac{1}{2}},$$

که در آن: D_j، فاصله نورون jام لایه خروجی از بردار ورودی x ($X= (x_i;\ i=1,2,3,..,N) \in R^n$)،

1- Competitive Phase
2- Co-operative Phase
3- Adaptive Phase
4- Best Matching Unit (BMU)

N، تعداد متغیرهای بردار ورودی، M، تعداد نورون‌های لایه خروجی، W_{ji}، وزن نورون خروجی $(W_{ji}; j=1,2, \ldots, M; i=1,2,\ldots,N)$ و علامت ||.||، نشان‌دهنده فاصله می‌باشد (۲).

در ابتدا وزن نورون‌های لایه خروجی به صورت تصادفی تعریف شده و در طی فرایند آموزش و در طول زمان، به مقادیر متغیرهای بردارهای ورودی بیشتر شبیه می‌گردد. پس از آنکه BMU مشخص گردید، وزن‌های آن و وزن‌های دیگر نورون‌های همسایه آن، برحسب مقدار فاصله‌ای که از BMU دارند (فاز همکاری) بر طبق رابطه‌ی ۳ به هنگام می‌شوند (فاز تطبیق).

$$wji(t+1) = wji(t) + \theta(t) \cdot \eta(t) \cdot [xi(t) - wji(t)] \quad (۳)$$

که در آن: t، نماینده زمان، $\theta(t)$، تابعی است که فاصله نورون‌های همسایه BMU را به نسبتی از همسایگی تبدیل می‌کند و $\eta(t)$، نرخ یادگیری است.

ارزیابی خوشه‌ها و روش نمونه‌گیری از آنها

جهت ساخت سه مجموعه داده همگن و مشابه (مجموعه داده‌های آموزش، اعتبارسنجی و آزمون)، تعیین تعداد بهینه خوشه‌ها و نحوه نمونه‌گیری از آنها حائز اهمیت می‌باشد. در این تحقیق، جهت تعیین تعداد بهینه خوشه‌ها، از ضریب سیلهوت[1] به دلیل کارائی و ساده‌گی پیاده‌سازی آن در نرم‌افزار MATLAB استفاده شده است.

با استفاده از این ضریب، مقدار درجه عضویت هر داده نسبت به خوشه‌ای که به آن تعلق گرفته است اندازه‌گیری می‌شود (۸) رابطه‌ی ۴:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad , -1 \le s(i) \le 1 \quad (۴)$$

که در آن: s(i)، ضریبی است که مقدار درجه عضویت یک داده (مثلاً داده iام) به خوشه‌ای که به آن تعلق گرفته (مثلاً خوشه A) را پس از عملیات خوشه‌بندی نشان می‌دهد، a(i)، متوسط عدم شباهت (فاصله) داده iام از سایر داده‌های قرار گرفته در خوشه A و b(i)، کوچک‌ترین معدل فاصله داده iام در خوشه A تا سایر داده‌ها در هر خوشه‌ای مثل E که متفاوت از خوشه A باشد را بیان می‌کند.

در عمل، ابتدا با به کمک رابطه‌ی ۴ ضریب سیلهوت برای هر داده محاسبه و سپس با استفاده از رابطه‌ی ۵، متوسط ضریب سیلهوت $\bar{s}(k)$ به ازاء تمامی داده‌های موجود (n) در هر یک از خوشه‌ها (k) محاسبه می‌گردد.

$$\bar{s}(k) = 1/n \sum_{i=1}^{n} s(i) \quad (۵)$$

در این رابطه، کافمن و روسو جدول راهنمای زیر را (جدول ۲) ارائه نمودند (۸).

جدول ۲- راهنمای تفسیر ضریب سیلهوت

ضریب سیلهوت $k(\bar{s})$	تفسیر
۱- ۰/۷۵	خوشه‌بندی داده قوی است
۰/۷۵-۰/۵	خوشه‌بندی داده معقولانه است
۰/۵-۰/۲۵	خوشه‌بندی داده ضعیف است
کمتر از ۰/۲۵	خو شه‌بندی نشده

1- silhouette Coefficient

به‌منظور نمونه‌گیری از داده‌های خوشه‌بندی شده به روش SOM، از روش تخصیص متناسب استفاده شده است. در این روش، تعداد نمونه‌ها متناسب با اندازه خوشه تغییر کرده بنحوی که با افزایش اندازه هر خوشه، تعداد نمونه‌گیری از آن خوشه افزایش یافته و بالعکس (۱۴)، رابطه‌ی ۶:

$$nh = n\frac{Nh}{\sum_{j=1}^{H} Nj} \qquad (۶)$$

که در آن: nh، تعداد نمونه گرفته شده از خوشه h، n، تعداد داده مورد نیاز، Nh تعداد داده‌ها در خوشه h، H، تعدادخوشه‌ها و Nj تعداد داده در سایر خوشه‌ها می‌باشد.

در این تحقیق از ۷۰٪ داده‌ها برای آموزش، ۱۵٪ داده‌ها برای اعتبارسنجی و مابقی داده‌ها برای آزمون استفاده شده است.

تحلیل آماری داده‌های حاصل از خوشه‌بندی

به‌منظور بررسی همگنی و مشابهت توزیع داده‌ها در سه مجموعه آموزش، اعتبارسنجی و آزمون، علاوه بر مقایسه پارامترهای آماری (میانگین، انحراف معیار، چولگی ...) از آزمون ناپارامتری دو نمونه‌ای کولموگروف- اسمیرنوف[1] (KS) نیز استفاده گردید. در این آزمون، فرض H_0 و H_1، به‌ترتیب یکسان بودن و یکسان نبودن توزیع داده‌ها در دو جمعیت را بیان می‌نماید. در آزمون مربوطه، با محاسبه آماره آزمون (Dc) از رابطه ۷، و مقایسه آن با مقدار آماره بحرانی که با توجه به تعداد داده و سطح خطا (مثلاً ۱٪ = α) از جدول تعیین می‌گردد (D_t)، می‌توان یکسان بودن توزیع جمعیت داده‌ها را در سه مجموعه، به صورت دو به دو با یکدیگر مقایسه نمود.

$$D_c = \mathrm{Max}\left|\frac{F(n_{i1})}{n_1} - \frac{F(n_{i2})}{n_2}\right| \qquad (۷)$$

که در آن: $F(n_{i1})$ و $F(n_{i2})$، فراوانی تجمعی متغیر x در دو مجموعه مورد مقایسه، n1 و n2 تعداد داده‌ها در دو مجموعه و D_c، آماره آزمون، حداکثر قدر مطلق اختلاف فراوانی تجمعی نسبی دو مجموعه می‌باشد (۱۳). چنانچه آماره محاسبه شده آزمون (D_c) از آماره جدول (D_t)، کوچک‌تر باشد، فرض H_0 آزمون، تائید می‌شود. در این تحقیق از نرم‌افزار MATLAB جهت انجام این آزمون استفاده شده است.

استانداردسازی داده‌ها

استاندارد نمودن داده‌ها، به منظور بی‌بعد نمودن داده‌ها در محاسبات فاصله (در عملیات خوشه‌بندی) و یا به‌منظور جلوگیری از کوچک‌شدن بیش از حد وزن‌های تخصیص یافته به نورون‌ها (در مدل‌های شبکه عصبی) می‌باشد. در این تحقیق با توجه به استفاده از توابع محرک سیگموئید و تانژانت هایپربولیک به‌ترتیب از رابطه‌های ۸ و ۹ جهت استاندارد نمودن داده‌ها در بازه‌های [۰/۹- ۰/۱] و [۰/۹- ۰/۹-] استفاده شده است.

$$z = 0.1 + 0.8 * \frac{X_i - X_{imin}}{X_{imax} - X_{imin}} \qquad (۸)$$

$$z = \left(\frac{1.8(X_i - X_{imin})}{X_{imax} - X_{imin}}\right) - 0.9 \qquad (۹)$$

که در آن: Z، متغیر استاندارد شده X_i، X_i، متغیر اولیه، X_{imin} و X_{imax} به‌ترتیب مقادیر کمینه و بیشینه متغیر X_i می‌باشند.

متغیرهای ورودی و خروجی

در آموزش مدل‌های شبکه عصبی، از داده‌های دبی لحظه‌ای (Q_i)، دبی متوسط روزانه (Q_t)، متوسط بارش روزانه (P_t) و دمای متوسط

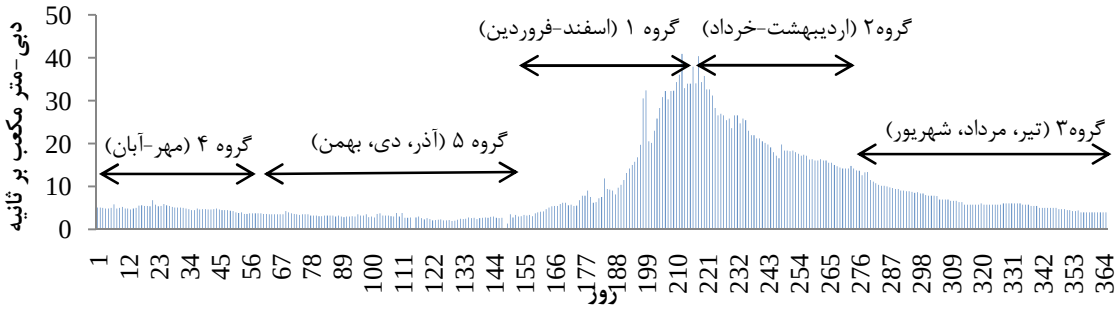
1- Two-Sample Kolmogoro-Smirnov Test

روزانه (T_t) به‌عنوان متغیرهای ورودی (تخمین‌گرهای مدل[1]) و از غلظت رسوب معلق (SSC_t) به‌عنوان متغیر خروجی استفاده شده است. همچنین به‌منظور افزایش دقت مدل‌سازی غلظت رسوب معلق روزانه، علاوه بر داده‌های روز جاری (t)، از داده‌های تا ۵ روز ماقبل آن (t-۵) نیز استفاده گردید. به دلیل تغییرات کم دما درمدت ۵ روز و جلوگیری از معرفی داده‌های مشابه به شبکه عصبی (جلوگیری از بیش برازش)، فقط از دمای متوسط روزانه (T_t) برای هر الگوی داده استفاده شده است.

روش مدل‌سازی و اسامی مدل‌ها

با توجه به تغییرات شدید غلظت رسوب معلق رودخانه در فصول مختلف سال (به دلیل نقش عوامل هیدرواقلیمی حوزه) و به منظور بکارگیری داده‌های همگن در مدل‌سازی رسوب، ابتدا بر اساس سه عامل ۱-رژیم بارش (برفی، بارانی، برفی-بارانی) ۲- رواناب حاصل از نوع بارش (۱۲) و تغییرات کلی هیدروگراف سالیانه جریان (شاخه صعودی یا نزولی) (شکل۲)، ماه‌های مختلف سال در طول دوره آماری به ۵ گروه تفکیک و سپس برای داده‌های هر گروه، مدل شبکه عصبی جداگانه‌ای (مدل‌های ANN-5 تا ANN-1) مطابق جدول ۳ ساخته شد. لازم به ذکر است که الگوی تغییرات دبی روزانه رودخانه کرج (واقع در ایستگاه سیرا)، در سال‌های آماری این تحقیق، تقریباً مشابه با الگوی نشان داده شده در شکل ۲ می‌باشد. پس از تفکیک گروه‌های اصلی، داده‌های مورد نیاز مدل‌ها با انجام عملیات خوشه‌بندی و نمونه‌گیری از داده‌های هر گروه، تهیه گردید. به‌منظور بررسی تاثیر ترکیبات مختلف متغیرهای ورودی در کارائی مدل‌سازی رسوب معلق، در هر گروه، مدل‌های مختلفی از شبکه عصبی (بر مبنای ترکیب متغیرهای ورودی) طراحی، آموزش و مورد آزمون قرار گرفت و سپس با توجه به شاخص‌های ارزیابی کارائی مدل‌ها، از میان آنها، مدل بهینه انتخاب شد. مدل‌هائی که ورودی آنها دبی جریان، بارش و دما می‌باشند با اسامی ANN-NM-QPT و مدل‌هائی که از دبی جریان به تنهائی استفاده می‌نمایند با اسامی ANN-NM-Q مشخص شده‌اند. در این اسامی، حروف NM، شماره گروه می‌باشد. به عنوان مثال، مدل ANN-2-QPT، مدل شبکه عصبی مربوط به گروه ۲ بوده که از متغیرهای دبی جریان، بارش و دمای روزانه در ساخت آن استفاده شده است. به‌منظور بررسی تاثیر گروه‌بندی داده‌ها در کارائی برآورد رسوب معلق، یک مدل شبکه عصبی واحد (ANN-ALL) نیز برای کلیه فصول سال تهیه شد. داده‌های مورد استفاده در بخش‌های آموزش و آزمون این مدل، از داده‌های مورد استفاده در بخش‌های آموزش و آزمون مدل‌های ۵ گانه قبل انتخاب گردید. پس از آن، مقدار خطای مدل با مقدار خطای مدل‌های ۵ گانه مقایسه شد. همانند مدل‌های ۵ گانه در این مدل نیز، جهت بررسی اثر متغیرهای بارش و دما در برآورد رسوب معلق، یک مدل با نام ANN-ALL-QPT و به منظور بررسی اثر دبی جریان به تنهائی، مدل دیگری به نام ANN-ALL-Q طراحی، آموزش و مورد آزمون قرار گرفت.

1- Model Estimator



شکل ۲- الگوی تغییرات دبی روزانه رودخانه کرج واقع در ایستگاه سیرا (۱۳۶۱-۱۳۶۰) و گروه‌بندی داده‌ها

جدول ۳- تفکیک زمانی داده‌ها و اسامی مدل‌ها

نام مدل	گروه داده‌ها و اسامی ماه‌ها	رژیم بارش	وضعیت کلی هیدروگراف جریان	نوع روان‌آب	تعداد داده
ANN-1	گروه ۱- اسفند، فروردین	باران- برف	شاخه صعودی	روان‌آب بارش همراه با ذوب برف	۱۵۵
ANN-2	گروه ۲- اردیبهشت، خرداد	باران	شاخه نزولی	روان‌آب بارش همراه با ذوب برف	۱۳۳
ANN-3	گروه ۳- تیر، مرداد، شهریور	باران	تغییر چندانی ندارد	دبی پایه	۱۱۳
ANN-4	گروه ۴- مهر، آبان	باران	تغییر چندانی ندارد	روان‌آب بارش بدون ذوب برف	۸۸
ANN-5	گروه ۵ - آذر، دی، بهمن	برف-ب اران	تغییر چندانی ندارد	روان‌آب زیادی مشاهده نمی‌شود	۱۲۱
ANN-ALL	تمامی ماه‌های سال	همه رژیم‌ها	تمامی شرایط	همه نوع روان‌آب	۶۱۰

ارزیابی کارائی مدل‌ها

به‌منظور ارزیابی نتایج گرفته شده از مدل‌های شبکه عصبی و مقایسه آنها با داده‌های رسوب مشاهداتی (داده‌های مجموعه آزمون)، اندازه‌گیری مقدار خطا و ترسیمات گرافیکی انجام گردید. همچنین نمودار پراکنش[1] داده‌های مشاهده‌ای، با داده‌های محاسباتی مدل ترسیم، و معادله خط رگرسیون و ضریب تبیین[2] (R^2) بهترین خط برازش تعیین گردید (رابطه‌ی ۱۰). به منظور بررسیمیزان خطای محاسباتی مدل‌ها، از شاخص‌های ریشه میانگین مربعات خطا[3] (RMSE)، میانگین قدر مطلق خطا[4] (MAE) و معیار ناش- ساتکلیف[5] (NS) به‌ترتیب رابطه‌های ۱۱، ۱۲ و ۱۳ استفاده شد.

$$R^2 = \left[\frac{\sum_{j=1}^{n}(S_O-\bar{S}_O)(S_M-\bar{S}_M)}{\sqrt{\sum_{i=1}^{n}(S_O-\bar{S}_O)^2\sum_{i=1}^{n}(S_M-\bar{S}_M)^2}}\right]^2 \quad (۱۰)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_M - S_O)^2} \quad (۱۱)$$

$$MAE = \frac{\sum_{i=1}^{n}|(S_O-S_M)|}{n} \quad (۱۲)$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(S_M-S_O)^2}{\sum_{i=1}^{n}(S_O-\bar{S}_O)^2} \quad (۱۳)$$

در رابطه‌های فوق، S_o و S_M به‌ترتیب غلظت رسوب معلق مشاهده‌ای و برآورد شده، $\bar{S}_O$، میانگین غلظت رسوب مشاهده‌ای، $\bar{S}_M$،

1- Scatter Plot
2- Coefficient of Determination
3- Nash-Sutcliffe
4- Root Mean Square error
5- Mean Absolute Error

میانگین غلظت رسوب برآورد شده و n، تعداد داده‌ها می‌باشد.

نتایج و بحث

نتایج خوشه‌بندی و تحلیل آماری داده‌ها

نتایج تعداد بهینه خوشه‌ها برای هر یک از مدل‌های شبکه عصبی محاسبه و در جدول ۴ بیان شده است. همچنین نتایج پارامترهای آماری و آزمون‌های ناپارامتری دو نمونه‌ای KS روی سه مجموعه آموزش، اعتبارسنجی و آزمون مدل‌ها محاسبه که با توجه به کثرت جداول آماری تنها به ذکر نتایج یکی از آنها (نتایج آزمون KS در مدل ANN-2-QPT) اکتفا شده است (جدول ۵). نتایج این جدول، نشان‌دهنده همگنی و توزیع یکسان داده‌ها (تائید فرض H_0 آزمون) در سه مجموعه یاد شده در این مدل می‌باشد. به عبارت دیگر، داده‌های واسنجی مدل ANN-2-QPT (داده‌های مجموعه آموزش)، به‌نحوی انتخاب شده‌اند که معرف و نماینده داده‌ها در کل دوره آماری این مدل (داده ماه‌های گروه ۲) باشند.

از نتایج فوق می‌توان نتیجه‌گیری نمود که داده‌های مورد استفاده در واسنجی مدل‌ها (داده‌های مجموعه آموزش) به‌نحوی انتخاب شده‌اند که معرف و نماینده داده‌ها در کل دوره آماری باشند. این مسئله، قدرت تعمیم‌دهی مدل‌های ساخته شده را افزایش می‌دهد.

نتایج مدل‌سازی

نتایج طراحی، آموزش و ارزیابی مدل‌های مختلف شبکه عصبی در جدول ۶ خلاصه شده است. برای هر مدل، بهترین ترکیب متغیرهای ورودی و همچنین ساختار شبکه عصبی و نوع تابع محرک آن، به همراه شاخص‌های ارزیابی کارائی مدل بیان گردیده است. به عنوان مثال جهت برآورد غلظت رسوب معلق (SSC_t) در مدل ANN-2-QPT، بهترین ترکیب داده‌های ورودی، داده‌های دبی لحظه‌ای جریان (Q_i)، دبی متوسط روزانه (Q_t)، دبی متوسط با تاخیر یک روزه (Q_{t-1})، دمای متوسط روزانه (T_t)، بارش متوسط روزانه (P_t) و بارش متوسط با تاخیر یک روزه (P_{t-1}) بوده است. همچنین، شاخص‌های ارزیابی کارائی مدل‌ها، با توجه به مجموعه داده‌های آزمون هر مدل، محاسبه شده‌اند. همانطور که نتایج نشان می‌دهد، در کلیه مدل‌ها، اضافه نمودن متغیرهای دما و بارش سبب افزایش کارائی آنها شده است. از سوی دیگر، کارائی استفاده از چند مدل ساده (مدل‌های ANN-1 تا ANN-5)، به مراتب بیشتر از زمانی است که تنها از یک مدل واحد (مدل ANN-ALL)، جهت برآورد رسوب معلق استفاده می‌شود.

جدول ۴- تعداد بهینه خوشه‌ها

نام مدل	ضریب سیلهوت $\overline{s(k)}$	تعداد بهینه خوشه‌ها
ANN-1	۰/۷	۳
ANN-2	۰/۷۷	۳
ANN-3	۰/۸۶	۴
ANN-4	۰/۸۷	۴
ANN-5	۰/۹۳	۱۲
ANN-ALL	۰/۸۲	۳

جدول ۵- نتایج آزمون ناپارامتری دو نمونه‌ای KS روی داده‌های سه مجموعه آموزش، اعتبارسنجی و آزمون در مدل ANN-2-QPT

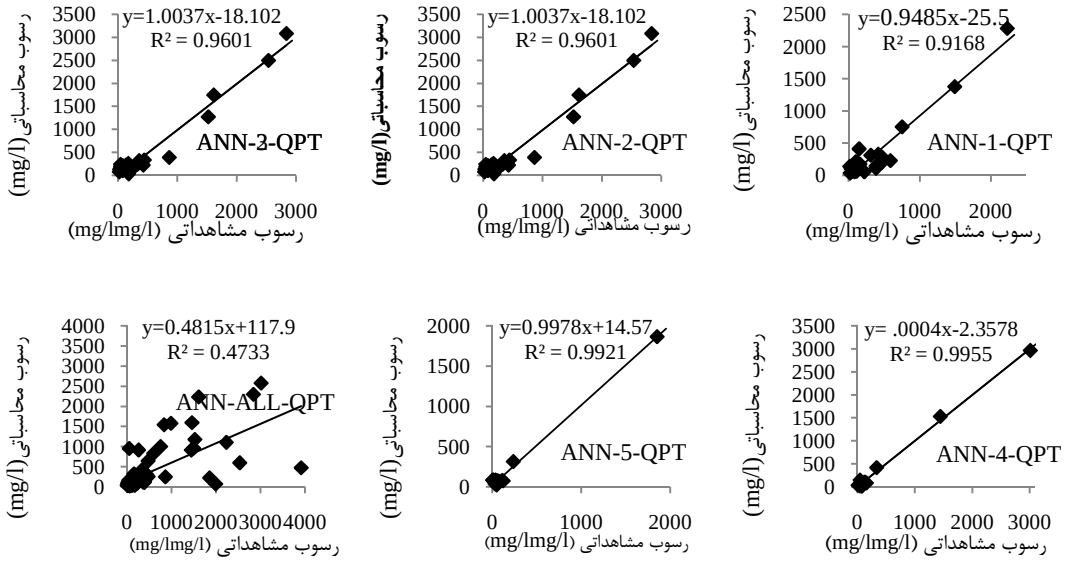
متغیر	مجموعه‌های مورد مقایسه	P-value	آماره D_c (محاسباتی)	آماره D_t (جدول)
دبی لحظه‌ای (Q_t)	آموزش- اعتبارسنجی	۰/۴۶	۰/۱۸	۰/۳۵**
دبی لحظه‌ای (Q_t)	آموزش- آزمون	۰/۶۲	۰/۱۸	۰/۴**
دبی لحظه‌ای (Q_t)	اعتبارسنجی- آزمون	۰/۳۳	۰/۲۷	۰/۴۸**
دبی متوسط روزانه (Q_t)	آموزش- اعتبارسنجی	۰/۷۱	۰/۱۵	۰/۳۵**
دبی متوسط روزانه (Q_t)	آموزش- آزمون	۰/۶۲	۰/۱۸	۰/۴**
دبی متوسط روزانه (Q_t)	اعتبارسنجی- آزمون	۰/۵۱	۰/۲۳	۰/۴۸**
دبی متوسط با تاخیر یک روزه (Q_{t-1})	آموزش- اعتبارسنجی	۰/۵۵	۰/۱۷	۰/۳۵**
دبی متوسط با تاخیر یک روزه (Q_{t-1})	آموزش- آزمون	۰/۷۳	۰/۱۷	۰/۴**
دبی متوسط با تاخیر یک روزه (Q_{t-1})	اعتبارسنجی-آزمون	۰/۸	۰/۱۸	۰/۴۸**
دمای متوسط روزانه (T_t)	آموزش- اعتبارسنجی	۰/۲۵	۰/۲۲	۰/۳۵**
دمای متوسط روزانه (T_t)	آموزش- آزمون	۰/۳	۰/۲۳	۰/۴**
دمای متوسط روزانه (T_t)	اعتبارسنجی- آزمون	۰/۶۱	۰/۲۱	۰/۴۸**
بارش متوسط روزانه (P_t)	آموزش- اعتبارسنجی	۰/۲۹	۰/۲۱	۰/۳۵**
بارش متوسط روزانه (P_t)	آموزش- آزمون	۰/۹۶	۰/۱۲	۰/۴**
بارش متوسط روزانه (P_t)	اعتبارسنجی- آزمون	۰/۱۹	۰/۳۱	۰/۴۸**
بارش متوسط با تاخیر یک روزه (P_{t-1})	آموزش- اعتبارسنجی	۰/۴	۰/۱۹	۰/۳۵**
بارش متوسط با تاخیر یک روزه (P_{t-1})	آموزش- آزمون	۰/۹۷	۰/۱۲	۰/۴**
بارش متوسط با تاخیر یک روزه (P_{t-1})	اعتبارسنجی- آزمون	۰/۷۳	۰/۱۹	۰/۴۸**
غلظت متوسط رسوب معلق (SSC_t)	آموزش- اعتبارسنجی	۰/۷۹	۰/۱۴	۰/۳۵**
غلظت متوسط رسوب معلق (SSC_t)	آموزش- آزمون	۰/۸۱	۰/۱۵	۰/۴**
غلظت متوسط رسوب معلق (SSC_t)	اعتبارسنجی- آزمون	۱	۰/۱	۰/۴۸**

**: معنی‌داری در سطح اطمینان ۹۹ درصد یا خطای یک درصد (α = ٪۱)

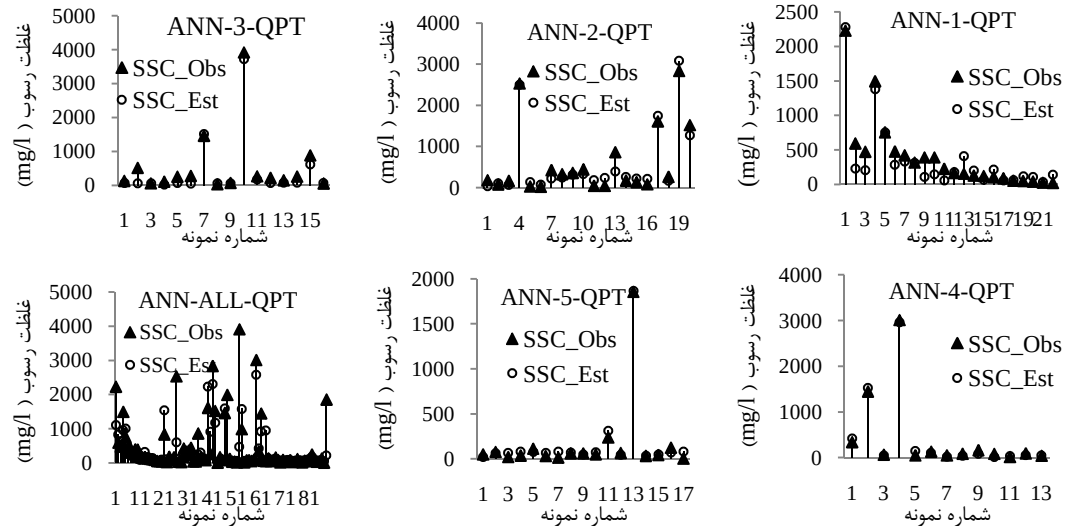
جدول ۶- ساختار و نتایج ارزیابی مدل‌های مختلف شبکه عصبی در تخمین غلظت رسوب معلق

نام مدل	بهترین ترکیب متغیرهای ورودی	تابع محرک	ساختار شبکه	شاخص‌های ارزیابی کارائی مدل (با داده‌های آزمون)			
				R^2	MAE (mg/l)	RMSE (mg/l)	NS
ANN-1-QPT	Q_i,Q_{t-1},P, P_{t-1}, T	Logsig	۱:۸:۱	۰/۹۱	۱۱۷/۸۸	۱۵۶/۹۳	۰/۹۰
ANN-1-Q	Q_i, Q_{t-1}, Q_{t-2}, Q_{t-3}	Logsig	۱:۱۲:۱	۰/۵۱	۲۰۶/۷۴	۴۱۲/۲۹	۰/۳۶
ANN-2-QPT	Q_i, Q_t, Q_{t-1}, P, P_{t-1}, T	Logsig	۱:۱۱:۱	۰/۹۶	۱۳۷/۲۰	۱۶۹/۳۷	۰/۹۶
ANN-2-Q	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Logsig	۱:۸:۱	۰/۹۵	۲۱۲/۸۶	۲۸۱/۸۲	۰/۸۸
ANN-3-QPT	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3}	Logsig	۱:۱۹:۱	۰/۹۸	۱۲۹/۶۸	۱۷۳/۶۳	۰/۹۶
ANN-3-Q	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	Logsig	۱:۱۲:۱	۰/۹۱	۱۹۹/۹۷	۳۲۷/۰۵	۰/۸۷
ANN-4-QPT	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, P_t, P_{t-1}, P_{t-2}, P_{t-3}, T	Logsig	۱:۱۳:۱	۰/۹۹	۴۸/۰۶	۵۵/۸۷	۰/۹۹
ANN-4-Q	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	Logsig	۱:۵:۱	۰/۹۹	۵۳/۶۱	۶۷/۷۰	۰/۹۹
ANN-5-QPT	Q_i, Q_t, Q_{t-1}, Q_{t-2}, P_t, P_{t-1}, P_{t-2}, T	Logsig	۱:۱۱:۱	۰/۹۹	۳۳/۰۴	۴۰/۴۷	۰/۹۹
ANN-5-Q	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Logsig	۱:۱۲:۱	۰/۳۶	۱۳۷/۶۶	۴۱۵/۰۵	۰/۰۲
ANN-ALL-QPT	Q_i, Q_t, Q_{t-1}, P, P_{t-1}, T	Logsig	۱:۱۱:۱	۰/۴۷	۲۴۶/۳۲	۵۶۲/۹۴	۰/۴۵
ANN-ALL-Q	Q_i, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}	Logsig	۱:۱۲:۱	۰/۳۹	۲۶۸/۹۴	۶۰۸/۵۳	۰/۳۵

شکل ۳، نمودار پراکنش غلظت رسوب معلق مشاهده‌ای و محاسباتی داده‌های آزمون مدل‌های مختلف (مدل‌هائی که از بارش و دما به همراه دبی جریان به عنوان ورودی استفاده می‌کنند) را نشان می‌دهد. همانطور که مشاهده می‌شود، خط رگرسیون در مدل‌های ANN-1-QPT تا ANN-5-QPT در مقایسه با مدل ANN-ALL-QPT، برازش بهتری به داده‌ها داشته است. در شکل ۴، مدل‌های مبتنی بر تفکیک داده (مدل‌های ANN-1 تا ANN-5) به خوبی توانسته‌اند مقادیر رسوب در غلظت‌های کم و زیاد را برآورد نمایند، این در حالی است که مدل ANN-ALL در برآورد صحیح مقادیر رسوب چندان موفق نبوده است. در مجموع می‌توان گفت که طبقه‌بندی زمانی داده‌ها بر اساس شرایط هیدرواقلیمی حوزه و بکارگیری متغیرهای دما و بارش به همراه دبی جریان، سبب افزایش دقت برآورد غلظت رسوب معلق شده است. نتایج گرفته شده در این تحقیق، در رابطه با مدل‌سازی رسوب معلق با بکارگیری متغیرهای بارش و دما، با نتایج دیگر محققین نظیر اولکی و همکاران (۲۳)، کاکائی لفدانی و همکاران (۶)، زو و همکاران (۲۴)، کیسی و شیری (۹) و ملسی و همکاران (۱۵) مطابقت می‌نماید. همچنین نتایج گرفته شده در زمینه همگن‌سازی داده‌ها با استفاده از طبقه‌بندی داده‌ها بر اساس شرایط هیدرواقلیمی حوزه، با نتایج مساعدی و همکاران (۱۶)، ذرتی پور و همکاران (۲۵) و قورقی و همکاران (۴) هم‌سو می‌باشد.



شکل ۳- نمودار نقطه‌ای نتایج حاصل از شبیه‌سازی غلظت رسوب معلق توسط مدل‌های شبکه عصبی در مقابل مقادیر مشاهداتی (در مرحله آزمون)



شکل ۴- نتایج حاصل از شبیه‌سازی غلظت رسوب معلق توسط مدل‌های شبکه عصبی در مقابل مقادیر مشاهداتی (در مرحله آزمون)

برآورد هرچه دقیق‌تر مقدار رسوب معلق رودخانه‌ها، به دلیل نقش منفی آنها در کاهش شاخص‌های کیفی آب، انتقال آلودگی، کاهش ظرفیت مخازن و کانال‌ها از اهمیت ویژه‌ای برخوردار است. در اغلب تحقیقات انجام شده در زمینه شبیه‌سازی رسوب معلق رودخانه‌ای، از دبی جریان به‌عنوان تنها متغیر تخمین‌گر مقدار رسوب استفاده می‌شود و این در حالی است که دبی جریان به تنهائی نمی‌تواند واریانس رسوب رودخانه را به خوبی در فصول مختلف سال تبیین نماید. نتایج این تحقیق نشان داد که استفاده از دبی جریان به همراه متغیرهای دما و

بارش روزانه (به‌عنوان متغیرهای تاثیرگذار در فرسایش و رسوبدهی حوزه)، می‌تواند کارائی مدل‌سازی را به مقدار زیادی افزایش دهد. همچنین به دلیل نقش تغییرات فصلی و شرایط آب و هوائی در فرسایش و رسوبدهی حوزه‌ها، تفکیک زمانی و همگن‌سازی داده‌ها (بر اساس رژیم بارش، وضعیت هیدروگراف جریان، نوع رواناب حاصل از بارش) قبل از مدل‌سازی، نقش مهمی در دقت برآورد رسوب رودخانه دارد. در این رابطه، استفاده از چند مدل ساده که بر اساس معیارهای فوق‌الذکر تدوین شده‌اند، بهتر از یک مدل واحد (که در آن هیچگونه تفکیکی در داده‌ها انجام نشده است) می‌تواند مقدار رسوب معلق رودخانه را برآورد نماید. نکته مهم دیگری که در مدل‌سازی متغیرهای محیطی (نظیر رسوب معلق) بایستی به آن توجه گردد، انتخاب داده‌های مناسب جهت واسنجی مدل‌ها می‌باشد. این داده‌ها بایستی به گونه‌ای انتخاب شوند که ضمن آنکه معرف داده‌ها در کل دوره آماری هستند، با دیگر مجموعه داده‌ها (نظیر مجموعه داده‌های اعتبارسنجی و آزمون)، مشابه و دارای توزیع یکسان باشند. در این راستا، شبکه عصبی SOM، ابزار مناسبی در خوشه‌بندی داده‌ها و ساخت مجموعه داده‌های مشابه و همگن می‌باشد. در مجموع، نتایج این تحقیق و ساختار مدل‌سازی بکار رفته در آن، می‌تواند به عنوان الگوئی در برآورد و یا پیش‌بینی بسیاری از متغیرهای محیطی حوزه‌ها از جمله دبی، تبخیر و تعرق، متغیرهای کیفی آب رودخانه‌ها (نظیر نیترات، فسفات) و غیره مورد استفاده قرار گیرد.

منابع

1. Azamathulla, H.M., Y.C. Cuan, A.A. Ghani and C.K. Chang. 2013. Suspended sediment load prediction of river systems: GEP approach. Arabian Journal of Geosciences, 6(9): 3469-3480.
2. Bowden, G.J., H.R. Maier and G.C. Dandy. 2002. Optimal division of data for neural network models in water resources applications. Water Resources Research, 38(2): 1-2.
3. Demirci, M. and A. Baltaci. 2013. Prediction of suspended sediment in river using fuzzy logic and multilinear regression approaches. Neural Computing and Applications, 23(1): 145-151.
4. Ghorghi, J.H., M. Habibnezhad, G. Vahabzadeh and A. Khaledi Darvishan. 2012. Efficiency of different data separation methods to increase the accuracy of sediment rating curve- A case study in part of the Sefidrood watershed. Irrigation & Water Engineering, 2(7): 97-111. (In Persian)
5. Hornik, K., M. Stinchcombe and H. White. 1989. Multilayer feedforward networks are universal approximators. Neural networks, 2(5): 359-366.
6. Kakaei Lafdani, E., A. Moghaddam Nia and A. Ahmadi. 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. Journal of Hydrology, 478(25): 50-62.

7. Kalteh, A.M., P. Hjorth and R. Berndtsson. 2008. Review of the self-organizing map (SOM) approach in water resources: analysis, modeling and application. Environmental Modeling & Software, 23(7): 835-845.
8. Kaufman, L. and P.J. Rousseeuw 2009. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc., 342 pp.
9. Kisi, O. and J. Shiri. 2012. River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. Computers & Geosciences, 43: 73-82.
10. Kohonen, T. 1982. Analysis of a simple self-organizing process. Biological cybernetics, 44(2): 135-140.
11. Li, X., M.H. Nour, D.W. Smith and E.E. Prepasc. 2010. Neural networks modeling of nitrogen export: model development and application to unmonitored boreal forest watersheds. Environmental Technology, 31(5): 495-510.
12. Management of watershed management of Jihad-e-Tehran providence. 1997. Detailed-Executive plan of watershed management of Shahrestanak watershed (Karaj dam). Hydrology and Meteorology report, 49 pp. (In Persian)
13. Mansourfar, K. 2008. Advanced to Statistical Methods Using Applied Software. Tehran, Iran, Tehran University Publishing, 459 pp. (In Persian)
14. May, R.J., H.R. Maier and G.C. Dandy. 2010. Data splitting for artificial neural networks using SOM-based stratified sampling. Neural Networks, 23(2): 283-294.
15. Melesse, A.M., S. Ahmad, M.E. McClain, X. Wang and Y.H. Lim. 2011. Suspended sediment load prediction of river systems: An artificial neural network approach. Agricultural Water Management, 98(5): 855-866.
16. Mosaedi, A., A. Mohammadi Ostadkelayeh, N.A. Najafi and F. Yaghmaei. 2006. Optimization of the relations between flow discharge and suspended sediment discharge in selected hydrometric stations of Gorganroud river. Iranian Journal of Natural Resources, 59(2): 331-342. (In Persian)
17. Mustafa, M.R., R.B. Rezaur, S. Saiedi and M.H. Isa. 2012. River suspended sediment prediction using various multilayer perceptron neural network training algorithms-a case study in Malaysia. Water resources management, 26(7): 1879-1897.
18. Nour, M., D. Smith, M. El-Din and E. Prepas. 2006. Neural networks modeling of stream flow, phosphorus, and suspended solids: application to the Canadian Boreal forest. Water Science & Technology, 53(10): 91-99.
19. Rajaee, T., V. Nourani, M. Zounemat-Kermani and O. Kisi. 2010. River suspended sediment load prediction: Application of ANN and wavelet conjunction model. Journal of Hydrologic Engineering, 16(8): 613-627.
20. Rodríguez-Blanco, M.L., M.M. Taboada-Castro, L. Palleiro-Suárez and M.T. Taboada-Castro. 2010. Temporal changes in suspended sediment transport in an Atlantic catchment, NW Spain. Geomorphology, 123(1): 181-188.
21. Salajegheh, A. and A. Fathabadi. 2009. Estimation of the suspended sediment loud of karaj river using fuzzy logic and neural networks. Journal of Range and Watershed Management, 62(2): 271-282. (In Persian)
22. Tayfur, G. 2012. Soft Computing in Water Resources Engineering - Artifical Neural Networks, Fuzzy Logic and Genetic Algorithms. WIT Press, Southampton, England, UK, 267 pp.

23. Ulke, A., G. Tayfur and S. Ozkul. 2009. Predicting suspended sediment loads and missing data for Gediz river, Turkey. Journal of Hydrologic Engineering, 14(9): 954-965.
24. Zhu, Y.M., X.X. Lu and Y. Zhou. 2007. Suspended sediment flux modeling with artificial neural network: An example of the Longchuanjiang River in the Upper Yangtze Catchment, China. Geomorphology, 84(1): 111-125.
25. Zoratipour, A., M. Mahdavi, S.K. Sigaroudi, A. Salajegheh and N.S. Almaali. 2009. Assessment of the effect of classification on the improved estimation of suspended sediment load using hydrological methods (case study: Taleghan basin). Iranian Journal of Natural Resources, 61(4): 809-819. (In Persian)

Estimation of Daily Suspended Sediment Concentration using Artificial Neural Networks and Data Clustering by Self-Organizing Map (Case Study: Sierra Hydrometry Station- Karaj Dam Watershed)

Mahmoodreza Tabatabaei[1], Karim Solaimani[2], Mahmoud Habibnejad Roshan[2] and Ataollah Kavian[3]

1- PhD Student, Sari Agricultural Sciences and Natural Resources University
(Corresponding author: taba1345@hotmail.com)
2 and 3- Professor and Associate Professor, Sari Agricultural Sciences and Natural Resources University
Received: May, 2 2014 Accepted: August, 26 2014

Abstract

Nowadays, the accurate estimation of rivers suspended sediment load (SSL), from various aspects, such as water resources engineering, environmental issues, water quality and so on is important. In this regard, because of various roles of fixed and dynamic variables of watersheds, the watershed hydrological models have not showen a proper efficiency in statimation of SSL. Also, the most SSL studies are based on only flow discharge variable whereas the results of the present study have proved that the efficiency of these modeles is very poor. On the other hands, the parameters such as rainfall type, year seasons and flow hydrograph shape have important role in watershed sediment yield that were ignored in the most SSL simulations. In the present study, multi layers perceptron neural network and hydro-meteorological data (daily flow discharge, suspended sediment concentration, daily rainfall and temperature) of Karaj dam watershed in a 30-year period (1981 to 2011) were used to estimate daily suspended sediment concentration of Sierra station. Due to the role of seasonal changes and flow conditions in sediment yield and sediment transport of the watershed, based on rainfall regime, hydrograph condition and runoff type, the data used in this study were first seperated into 5 groups and then for each group, a separate model was designed. In order to increase the generalization ability of the neural network models, self-organizing map (SOM) and Silhouette coefficient were used for data clustering and determination of the optimal number of clusters respectively. The research results showed that the use of daily precipitation and temperature variables along with flow discharge and data separating based on watershed time and hydro climatic conditions has had an important role in increasing the accurate estimation of the river sediment. In this regard, among the models, the maximum calculated error is when only a single model is used for all year seasons. The research results and its used structure can be used as a pattern to estimate and to forecast many environmental variables of watersheds.

Keywords: Clustering, Karaj River, Neural Networks, Self-Organizing Map, Suspended Sediment