

کاربرد الگوی کمانک تطبیقی چندمتغیره در الگوبندی و تبیین درازمدت شوری و باقی مانده خشک آب رودخانه

مطالعه موردی: حوضه‌های آبریز استان خراسان جنوبی

مجید جانفدا^۱، حجت رضائی پزند^۲

چکیده

شوری (Ec) و باقی مانده خشک (TDS) از عوامل کیفی آب‌های شرب، کشاورزی و صنعت محسوب می‌شوند. ابتدا رابطه بین دبی-شوری و دبی- TDS (توابع سنجه) نمونه‌های برداشت شده با برازش الگوهای رگرسیونی نمایی، توانی و شبکه عصبی انجام می‌شود. سپس تحلیل دراز مدت Ec و TDS آب رودخانه (مقیاس روز، ماه و سال) با الگوی انتخابی صورت می‌گیرد. نمونه‌های برداشتی اغلب از جریان عادی و سیلاب‌های کوچک است. سیلاب‌های بزرگ و کم آبی‌های شدید رودخانه اثرات مهمی در تحلیل و تبیین نوسانات Ec و TDS دارند. لذا نیاز به تخمین Ec و TDS این سیلاب‌ها و کم آبی‌هاست (برون‌یابی). الگوهای رایج در برون‌یابی به دلیل کران دار بودن شوری و باقیمانده خشک خوب عمل نمی‌کنند. هدف مقاله حاضر به کارگیری الگوی جدید کمانک تطبیقی چند متغیره^۳ (MARS) برای حل این مشکل است. حوضه‌های آبریز خراسان جنوبی نیز برای مطالعه موردی انتخاب شد. MARS یک الگوی رگرسیون تکه‌ای خطی (یا غیرخطی) کمانکی است که در برون‌یابی و رعایت کران‌ها عملکرد بهتری دارد. سه الگوی نمایی، توانی و MARS (با نردبان تبدیل) بر آمار دبی- Ec و دبی- TDS ایستگاه‌های استان خراسان جنوبی برازش داده شد. دو دیدگاه آماری و فیزیکی این الگوها معیار انتخاب الگوی برتر است که MARS برتری خود را در این دو حالت نشان داد. نتایج الگوهای برازشی در دو حالت درون‌یابی و برون‌یابی با هم مقایسه شد. برون‌یابی الگوی MARS اعداد معقول و برون‌یابی سایر الگوها اعداد بسیار بزرگ (حتی بی‌نهایت) ارائه می‌کنند. تحلیل‌ها نشان داد که الگوی MARS با نردبان تبدیل می‌تواند جایگزین خوبی برای الگو بندی منحنی سنجه Ec و TDS باشد.

کلید واژه‌ها: شوری آب، باقی مانده خشک، MARS، نمایی، توانی، خراسان جنوبی.

مقدمه:

رودخانه‌ها به عنوان مهم‌ترین منابع تأمین و انتقال آب مصرفی بخش‌های صنعت، کشاورزی و شهری اهمیت خاصی دارند. عبور از بسترها و مناطق مختلف سبب ایجاد، نوسانات کیفی زیاد در آب رودخانه‌ها می‌شود. تخمین خصوصیات کیفی آب رودخانه‌ها یکی از نیازهای اصلی مدیریت، برنامه‌ریزی و بهره‌برداری از منابع آب‌های سطحی است. شوری^۱ (Ec) و باقی‌مانده خشک^۲ (TDS) آب از جمله شاخص‌های مهم کیفی آب محسوب می‌شوند (لو، ۱۹۹۲). روش‌های مرسوم تبیین درازمدت این شاخص‌های کیفی در مقیاس روز، ماه و سال با تحلیل و الگوبندی نمونه‌های برداشت شده دبی- Ec و دبی-TDS و تعمیم نتایج آن به دبی درازمدت روزانه رودخانه و دبی‌های ماهانه و سالانه حوضه آبریز رودخانه است. بنابراین نمونه‌های برداشتی (مشاهدات) باید میزان Ec و TDS درازمدت جریان رودخانه را تبیین کنند. اما در عمل، نمونه‌ها اغلب اندک و عمدتاً در دبی‌های عادی و سیلاب‌های کوچک برداشت، سپس الگوهای مناسب بر این نقاط برازش و مقدار Ec و TDS درازمدت با این الگوها محاسبه می‌شود. از طرفی دیگر سیلاب‌های بزرگ و کم‌آبی‌های کوچک نیز مکرر در حوضه‌های آبریز رخ می‌دهند که سهم بسزایی در تبیین نوسانات این دو عامل دارند و نمونه Ec و TDS آن‌ها در اختیار نیست (آرشیو سازمان مدیریت منابع آب وزارت نیرو). معمولاً دبی‌های مربوط به این سیلاب‌ها خارج دامنه نمونه‌های اندازه‌گیری است. لذا تبیین درازمدت میزان Ec و TDS رودخانه با نمونه‌های محدود مشاهده‌ای دقت مناسبی ندارد. سیلاب‌های بزرگ و کم‌آبی‌های کوچک از نظر علم آمار به داده پرت^۳ معروفاند (بارنت و لويس، ۱۹۹۴ و سرمد، ۲۰۱۲). بنابراین الگوی برازشی بر نمونه رسوب باید Ec و TDS مربوط به این نقاط پرت را نیز به خوبی برآورد یا به عبارت دیگر آن‌ها را برون‌یابی کند. لذا عملکرد الگوی برازشی به دو حالت درون‌یابی (دامنه دبی‌های نمونه) و برون‌یابی (خارج داده‌های نمونه) تقسیم می‌شود.

Ec و TDS از دیدگاه کیفیت آب کران‌دار است. زیرا این دو نمی‌تواند از یک حداقل و حداکثری پیشی گیرند. رفتار منحنی شوری در جریان عادی رودخانه و سیلاب‌های کوچک، تقریباً از یک منحنی S وارون (پله‌ای) مشابه شکل (۲) پیروی می‌کند. زیرا شوری با افزایش دبی کاهش می‌یابد و این کاهش دارای یک سقف است. این نشان از وجود یک توزیع آمیخته^۴ دارد (وایزبرگ، ۲۰۰۵). این رفتار در TDS (به صورت معکوس) نیز مشاهده می‌شود (شکل ۳). بنابراین باید الگوئی بر داده‌های دبی- Ec (منحنی سنجه شوری) و دبی- TDS (منحنی سنجه TDS) برازش داد که این ویژگی‌ها را بهتر تبیین کند. اغلب محققین و پژوهشگران برای این منظور از الگوهای جدول توزیع فراوانی، منحنی‌های نمایی و توانی (منحنی سنجه)، برازش چشمی، شبکه عصبی و رگرسیون تک‌های استفاده می‌کنند (آرشیو سازمان مدیریت منابع آب وزارت نیرو، فریدمن، ۲۰۰۹). این الگوها برازش خوبی در دامنه داده‌ها یا مشاهدات دارند. اما در برون‌یابی (خارج دامنه داده‌ها) خوب عمل نمی‌کنند. زیرا نیکوئی برازش آن‌ها مربوط به دامنه داده‌هاست (درون‌یابی) نه در برون‌یابی. مثلاً الگوهای نمایی یا توانی در برون‌یابی شوری (TDS) در دبی‌های بالا با مساله بیش برآورد و در کم‌آبی‌ها با مساله کم‌برآورد مواجه است (شیتز، ۲۰۰۹). زیرا این عوامل به ازای دبی‌های پایه و جریان‌های عادی رشد نمایی یا توانی دارد اما در سیلاب‌های بزرگ و کم‌آبی‌های کوچک کران‌دار بوده و نمی‌توانند از حداکثر و حداقلی فراتر رودند (شیتز، ۲۰۰۹). لذا الگوهایی برای منحنی سنجه مناسب است که در درون‌یابی‌ها و برون‌یابی‌ها محدودیت فیزیکی (کران‌ها) را لحاظ نماید. پیش‌بینی (برآورد) شوری آب در سال‌های اخیر با شبکه عصبی مصنوعی

1- Electrical conductivity (Umhos/cm)

2- Total Dissolved Solids

3- Outliers

4- Mixtures distribution

نیز انجام و نتایج رضایت بخشی در درون‌یابی ارائه داده است. پالانی^۱ و همکاران (۲۰۰۸)، از شبکه‌های عصبی برای پیش‌بینی مشخصات کیفی آب‌های ساحلی سنگاپور بهره گرفتند. آن‌ها الگوی خود را برای ارزیابی و پیش‌بینی پارامترهای کیفی آب در هر نقطه در محدوده مورد مطالعه با توجه به پارامترهای اندازه‌گیری شده در دیگر نقاط محاسبه نمودند. چپوس^۲ و چانگ^۳ (۲۰۰۶) برای شبیه‌سازی مشخصات کیفی مخزن سد شیهمن^۴ بر روی رودخانه تاهان^۵ در تایوان از شبکه‌های عصبی مصنوعی بهره گرفتند. کرمی و همکاران (۱۳۸۵)، شبکه عصبی مصنوعی را برای کیفیت آب رودخانه کارون به کار بردند. اسدالله فردی و همکاران (۱۳۸۷)، الگوهای شبکه عصبی را برای الگو بندی و پیش‌بینی شوری و باقی مانده خشک (درون‌یابی) آب رودخانه حوضه آبریز آجی چای برازش و تحلیل کردند. این محققین گام‌های موثری را در زمینه پیش‌بینی TDS نیز برداشته‌اند. اسدالله فردی و همکاران (۱۳۸۸)، TDS رودخانه تلخه رود را با استفاده از شبکه عصبی مقادیر پیش‌بینی کردند. قاسم پور و همکاران (۱۳۸۹) نیز به تبیین رابطه دبی- TDS ایستگاه‌های بلده و رزن رودخانه نور پرداختند. قبادیان و مشایخی (۱۳۸۷)، از مدل شبکه عصبی به منظور پیش‌بینی پارامترهای کیفی آب رودخانه قره‌سو در محل ایستگاه قرباغستان در نزدیکی شهر کرمانشاه استفاده نمودند. بختیاری و همکاران (۱۳۸۶)، برای ارزیابی کیفی رودخانه کرخه از شبکه‌های عصبی استفاده کردند. محجوبی و تجریشی (۱۳۸۹) به مقایسه عملکرد شبکه عصبی و درخت تصمیم در پیش‌بینی TDS رودخانه کارون پرداخته‌اند.

Ec و TDS با منحنی‌های تکه‌ای نیز قابل برآورد است. زیرا رفتار منحنی سنجه آن‌ها به صورت منحنی‌های آمیخته و S شکل است (آرشیو سازمان مدیریت منابع آب وزارت نیرو). این برآورد می‌تواند دقت تخمین Ec و TDS آب را به ویژه در برون‌یابی افزایش دهد. روش تکه‌ای از توزیع‌های آمیخته استفاده و چند منحنی را در نقاط گره به هم پیوند می‌زند. تعیین نقاط گره در این روش با سعی و خطا انجام و بر فرضیات خاص خود استوار است (وایزبرگ، ۲۰۰۵، شیتز ۲۰۰۹، ریان و همکاران، ۲۰۰۲). رگرسیون کمانک (اسپیلین) تطبیقی چند متغیره (MARS) نسخه پیشرفته تکه‌ای خطی است که از توابع آینه‌ای معروف به چوب‌هاکی استفاده می‌کند (فریدمن، ۱۹۹۱، هاستی و همکاران، ۲۰۰۹). MARS بر داده‌های دبی-SAR برازش و با الگوهای مرسوم مقایسه شد که نتایج MARS جواب رضایت بخشی در برون‌یابی ارائه داد (رضایی پزند و جانفدا، ۱۳۹۰). الگوهای مذکور رفتار منحنی سنجه Ec و TDS را در دامنه نمونه‌های برداشتی (درون‌یابی) به خوبی تبیین می‌کند. زیرا این الگوها رگرسیون خطی و غیرخطی هستند که در درون‌یابی جواب مساعدی می‌دهند. الگوهای رگرسیونی معمولی در برون‌یابی توصیه نمی‌شوند. زیرا بازه اطمینان آن‌ها سریعاً گسترده و دقت برون‌یابی کاهش می‌یابد (وایزبرگ، ۲۰۰۵ و شیتز، ۲۰۰۹).

هدف و نوع‌آوری این مقاله به‌کارگیری الگوی رگرسیونی کمانک (اسپیلین) تطبیقی چند متغیره (MARS) با نردبان تبدیل برای الگو بندی منحنی سنجه Ec و TDS است. MARS در هر دو حالت برون‌یابی و درون‌یابی خوب عمل کند. این الگو می‌تواند رفتار دراز مدت این دو پدیده را بهتر تبیین کند. الگوی MARS بر آمار مشاهده‌ای دبی

Ec و دبی-TDS ایستگاه‌های آب سنجی استان خراسان جنوبی برازش داده شد. سپس عملکرد این الگو با الگوهای رایج (نمایی و توانی) در دو حالت درون‌یابی و برون‌یابی و با معیارهای فیزیکی و آماری ارزیابی شده است.

مواد و روش‌ها

منطقه مورد مطالعه و داده‌ها

منطقه مورد مطالعه ایستگاه‌های آب سنجی واقع در استان خراسان جنوبی است. مساحت این استان بالغ بر ۹۵۳۸۵ کیلومتر مربع است. ایستگاه‌های آب سنجی فعال این استان شامل: افین، فرخی، خونیک، منصورآباد و دربند است. تنها ایستگاه منصورآباد آمار دبی- Ec و دبی-TDS مناسبی ندارد. سایر ایستگاه‌ها آمار نسبتاً کامل و قابل تحلیل دارند. این ایستگاه‌های زیر نظر شرکت آب منطقه‌ای خراسان جنوبی قرار دارند (شرکت آب منطقه‌ای خراسان جنوبی).

رگرسیون MARS

بنیان ریاضی

فریدمن^۱ (۱۹۹۱) رگرسیون کمانکی تطبیقی چند متغیره (MARS) را معرفی کرده است. این روش به طور گسترده در داده‌کاوی و الگوسازی به کار می‌رود. می‌توان آن را تعمیمی از رگرسیون خطی گام به گام و یا شکل اصلاح شده درخت رگرسیون^۲ انگاشت (هاستی و همکاران، ۲۰۰۹). قابلیت‌های این روش متناسب با افزایش تعداد متغیرهای ورودی بهتر نمایان می‌شود (فریدمن، ۱۹۹۱، هاستی و همکاران، ۲۰۰۹). حالت یک متغیره برای تحلیل داده‌های Ec و TDS مناسب است. بنیان ریاضی این روش به اختصار مطابق زیر است.

فرض کنید $(x_1, y_1), \dots, (x_n, y_n)$ مجموعه مشاهدات باشد. روش MARS مقادیر متغیر پاسخ (Y_i) را با استفاده از الگویی خاص که بر توابعی به نام توابع مبنا (آینه‌ای یا چوب‌هاکی) استوار است، برآورد می‌کند. این توابع مطابق رابطه (۱) بیان می‌شوند.

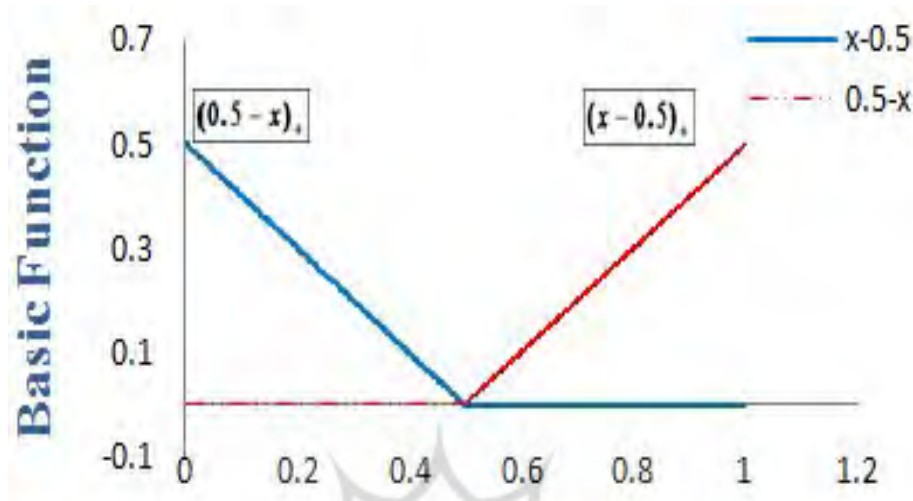
$$h_1(x) = (x-t)_+ = \begin{cases} x-t & x > t \\ 0 & x \leq t \end{cases} \quad h_2(x) = (t-x)_+ = \begin{cases} t-x & x < t \\ 0 & x \geq t \end{cases} \quad (1)$$

t ثابت گره و نماد "+" به معنای بخش مثبت است. گره t در عمل یکی از مشاهدات متغیر X است. دو تابع $h_1(x)$ و $h_2(x)$ هر یک به تنهایی تابع چوب‌هاکی و هر دو با هم زوج منعکس یافته^۳ (تابع آینه‌ای) در گره t هستند. شکل (۱) نمودار این توابع را در $t = 0/5$ نشان می‌دهد.

1-Fridman

2-Classification and Regression Tree

3-Reflected Pair



شکل ۱: نمودار تابع پایه

مجموعه C (رابطه ۲) را شامل همه این زوج توابع منعکس یافته در نظر می‌گیریم که $\{x_1, x_2, \dots, x_n\}$ مقادیر مشاهده شده X هستند.

$$\forall t \in \{x_1, x_2, \dots, x_n\}, C = \{(X - t)_+, (t - X)_+\} \quad (2)$$

ساختار کلی الگوی رگرسیون یک متغیره MARS برای برآورد Y مطابق رابطه (۳) است که h_i یکی از توابع مجموعه C است.

$$Y = f_{\gamma_m}(x) = b_0 + \sum_{i=1}^{\gamma_m} b_i h_i(X) \quad (3)$$

ضرایب $\{b_0, b_1, \dots, b_{\gamma_m}\}$ با کمینه‌سازی مجموع مربعات خطا برآورد می‌شوند.

الگوریتم رگرسیون یک متغیره MARS

پیکربندی الگوی MARS، مشابه رگرسیون خطی چند متغیره گام به گام است. توابع موجود در مجموعه C (رابطه ۲) جانشین X می‌شود. این الگو از سه مرحله مطابق الف) آرایش (ایجاد الگوی اولیه به روش پیش‌رو)، ب) پیرایش (اصلاح الگوی اولیه به روش پس‌رو) و پ) گزینش (انتخاب الگوی بهینه رگرسیون خطی یک متغیره MARS) تشکیل می‌شود (فریدمن، ۱۹۹۱). در طی مراحل آرایش و پیرایش الگوهای مختلفی بر اساس توابع مبنا بر داده‌ها برازش و مجموع مربعات خطا (SSE) اندازه‌گیری می‌شود. الگوی نهایی در مرحله گزینش انتخاب می‌شود. معیار تشخیص در این مرحله، استفاده از اعتبار متقابل تعمیم یافته^۱ (GCV) برای الگوهای مرحله پیرایش است. GCV برای ژامین الگو مطابق رابطه (۴) است. m_j تعداد توابع پایه در الگوی ژام. SSE_j مجموع مربعات خطای الگوی ژام، γ پارامتر هموارسازی (عددی بین ۲ تا ۴) و n تعداد مشاهدات است. الگوی نهایی انتخابی با رگرسیون

MARS دارای کمترین مقدار GCV است. وجود هر تابع پایه در الگو معادل یک گره است. این الگو می‌تواند چند گره‌ای نیز باشد (فریدمن، ۱۹۹۱).

$$GCV_j = \frac{SSE_j}{(1 - \frac{vm_j}{n})} \quad j = 1, 2, \dots, 2n - 2 \quad (4)$$

تبدیلات در رگرسیون

نردبان تبدیل روشی برای تبدیل متغیر پاسخ یا متغیر پیش‌بین است. هدف آن تثبیت واریانس خطا و تقویت الگو است. نردبان تبدیل اغلب در بازه [۲ و -۲] انجام می‌شود. تبدیل صفر معادل تبدیل لگاریتمی و تبدیلات منفی توابع معکوس است (ویزبرگ، ۲۰۰۵ و شیتز، ۲۰۰۹).

نرم‌افزار

نرم‌افزار "R.2.13.0" و برنامه نویسی در آن به منظور انجام محاسبات استفاده شده است. بسته‌های نرم‌افزاری earth، plotrix، plotmo، leaps و mda در تحلیل رگرسیون MARS استفاده شده است (وب سایت نرم‌افزار).

نتایج و بحث

الگوهای رایج نمایی، توانی و شبکه عصبی رفتار منحنی سنجه Ec و TDS را در درونیابی به خوبی تبیین می‌کند. این الگوها در برون‌یابی توصیه نمی‌شوند. زیرا بازه اطمینان سریعاً گسترده و دقت برون‌یابی کاهش می‌یابد. بنابراین تبیین درازمدت Ec و TDS در مقیاس روز، ماه و سال دقت مناسب ندارند.

هدف این مقاله معرفی یک الگوی جدید (MARS) برای ارتقای منحنی سنجه دبلی - Ec و دبلی - TDS در برون‌یابی و درونیابی است تا دقت برآورد دراز مدت این دو پدیده را افزایش دهد. MARS با الگوهای نمایی، توانی مقایسه شده است. داده‌های مورد مطالعه دبلی - Ec و دبلی - TDS ایستگاه‌های افین، فرخی، خونیک، دربند و منصورآباد (استان خراسان جنوبی) است. الگوی انتخابی باید از دو دیدگاه فیزیکی و آماری، همچنین در درونیابی و برون‌یابی خوب عمل کند. بررسی آمار دبلی روزانه حوضه‌ها نشان داد که دبلی‌های مربوط به نمونه‌های برداشتی Ec و TDS کمتر از دبلی‌های سیلابی رخ داده در آمار روزانه است. غربال داده‌ها انجام که منجر به حذف تعدادی از داده‌های برخی از ایستگاه‌های آب سنجی شد. الگوهای انتخابی بر داده‌های غربال شده ایستگاه‌های فوق برآزش داده شد (جدول ۱). دو الگوی نمایی و توانی در اغلب ایستگاه‌ها در حالت درونیابی پاسخ نسبتاً مساعدی می‌دهند. اما این دو الگو در برون‌یابی رفتار بدی از خود نشان می‌دهند (جدول ۴). به همین دلیل در اغلب ایستگاه‌ها رد شدند. شرح تمام مراحل برای کلیه ایستگاه‌ها پرحجم است. لذا ایستگاه خونیک برای الگوی دبلی - شوری و ایستگاه دربند برای الگوی دبلی - TDS به عنوان نمونه انتخاب که خلاصه تحلیل‌ها مطابق زیر است.

الف - تحلیل منحنی سنجه Ec ایستگاه خونیک: نمونه‌های برداشتی ۴۷ عدد در فاصله زمانی ۱۳۵۰ تا ۱۳۸۴ است. گستردگی زمانی نمونه‌ها، نسبتاً خوب و بیانگر کیفیت آب رودخانه است (جدول ۲). نمودار پراکنش شوری - دبلی (شکل ۲) یک رابطه تکه خطی یا غیرخطی را نشان می‌دهد. الگوهای نمایی، توانی و MARS بر داده‌ها برآزش داده شد که خلاصه نتایج مطابق زیر است.

الگوی نمایی: ضریب تعیین 0.13 و مجموع مربعات خطا $SSE = 65605247$ است (جدول ۴). این الگو در درون یابی و برون یابی بسیار نامعقول است (جدول ۳). مثلاً شوری مربوط به دبی‌های کمتر از ۳ مترمکعب بر ثانیه عدد ثابت 6210 میکروموس و برای دبی بیش از ۱۰ مترمکعب بر ثانیه مساوی منهای بی‌نهایت بوده که نامعقول است.

الگوی توانی (لگاریتمی): ضریب تعیین 0.59 و $SSE = 96899964$ است (جدول ۴). این الگو در برون یابی خوب عمل نمی‌کند (جدول ۳). مثلاً شوری مربوط به دبی‌های بیش از ۱۰ مترمکعب (برون یابی) عددی صفر است. این شوری از نظر فیزیکی پذیرفته نیست.

الگوی رگرسیون MARS: الگوهای مختلف MARS با نردبان تبدیل برازش داده شد تا بتوان دامنه وسیع‌تری از الگوها را برای درون یابی و برون یابی به دست آورد. بررسی‌های اولیه نشان داد که مناسب‌ترین تبدیلات تمامی ایستگاه‌ها در بازه $[-1, 0]$ است. تبدیلات به گونه‌ای انتخاب شدند که الگوی نهایی انتخاب شده بتواند هر دو دیدگاه آماری و فیزیکی را در درون یابی و برون یابی رعایت کند. ملاحظات فیزیکی و آماری الگوی انتخابی را با تبدیل $x^{-0.4}$ ترجیح داد (جدول ۲-الف). ضریب تعیین اصلاح شده 0.73 ، $SSE = 42645908$ و $GCV = 13199955$ است (جدول ۱-الف). رفتار باقی‌مانده‌ها (آسیب‌شناسی) و نیکویی برازش (شکل ۲ و ۴) موید مناسب بودن الگو است. الگوی نهایی دوگره‌ای در نقاط $1/741$ و $3/081$ قرار دارد. نتایج فوق مناسب بودن الگوی MARS را از دیدگاه آماری تایید می‌کند. شوری درون یابی و برون یابی شده با الگوی MARS (جدول ۳) نشان از رفتار معقول فیزیکی این الگو دارد. برون یابی Ec مربوط به دبی‌های 0.1 (کم آبی‌های شدید) و 500 مترمکعب بر ثانیه (سیلاب‌های بزرگ) به ترتیب $10300/4$ و $869/1$ میکروموس پیش‌بینی می‌شود که همخوانی خوبی با نمونه‌های مشاهداتی دارد.

مقایسه الگوهای برازشی: جدول (۳) مقادیر برآوردی شوری آب را با الگوهای MARS، نمایی و لگاریتمی در دو حالت درون یابی و برون یابی نشان می‌دهد. دبی داده‌های مشاهده‌ای با حداقل 0.2 و حداکثر $33/32$ مترمکعب بر ثانیه به ترتیب دارای شوری‌های 7500 و 1590 میکروموس بر سانتی‌متر است. این اعداد نشان می‌دهند که مقدار شوری در دبی‌های بزرگ (عمدتاً سیلاب‌ها) کمتر از 1500 میکروموس است. این موضوع فقط تطابق خوبی با الگوی MARS دارد (جدول ۳). دبی‌های 0.1 ، 1 ، 10 و 100 مترمکعب بر ثانیه برای مقایسه شوری‌های برآوردی سه الگوی MARS، نمایی و توانی (لگاریتمی) انتخاب شد. شوری‌های برآوردی در جدول (۳) ارائه شده‌اند. این شوری‌ها برای الگوی MARS به ترتیب: $10074/1$ ، $4727/1$ ، 2214 و $899/4$ میکروموس است. این شوری‌ها برای الگوی نمایی به ترتیب: 6210 ، 6210 و $-\infty$ و برای الگوی لگاریتمی به ترتیب: $5974/4$ ، $5500/7$ ، $0/0$ و $0/0$ میکروموس است. مقایسه این اعداد نشان می‌دهد که دو الگوی نمایی و لگاریتمی جواب‌های نامناسبی در این برون یابی (حتی در درون یابی) ارائه می‌کنند. اما الگوی MARS در دو حالت درون یابی و برون یابی از دقت خوبی برخوردار است. بنابراین الگوی MARS با تبدیل نردبانی مناسب می‌تواند شوری آب رودخانه را به خوبی الگوبندی و شوری دراز مدت آب رودخانه را به طور معقول تبیین کند.

ب- تحلیل منحنی سنج TDS ایستگاه دربند: نمونه‌های برداشتی حوضه دربند 55 عدد بوده که در فاصله زمانی 1350 تا 1384 صورت گرفته است. گستردگی زمانی نمونه‌ها، نسبتاً خوب و بیانگر کیفیت آب رودخانه است

(جدول ۲). داده‌های جدول (۲) برای بررسی رابطه دبی-TDS استفاده شده است. نمودار پراکنش داده‌ها (شکل ۳) یک رابطه تکه خطی یا غیرخطی را نشان می‌دهد. سه الگوی نمایی، توانی و MARS (با نردبان تبدیل) انتخاب و بر داده‌ها برازش داده شد. خلاصه نتایج مطابق زیر است.

الگوی نمایی: ضریب تعیین $0/6$ و مجموع مربعات خطا $SSE = 1787030$ است (جدول ۴). این الگو در درون‌یابی و برون‌یابی بسیار نامعقول است (جدول ۳). مثلاً شوری مربوط به دبی‌های بیش از ۱۰ مترمکعب (برون‌یابی) عددی صفر است. این شوری از نظر فیزیکی پذیرفته نیست.

الگوی توانی (لگاریتمی): ضریب تعیین $0/69$ و $SSE = 1219446$ است (جدول ۴). این الگو در برون‌یابی خوب عمل نمی‌کند (جدول ۳). مثلاً شوری مربوط به دبی‌های بیش از ۱۰ مترمکعب را $-\infty$ برآورد می‌کند که از نظر فیزیکی پذیرفته نیست.

الگوی رگرسیون MARS: الگوی MARS با نردبان تبدیل بر داده‌ها برازش داده شد. بررسی‌های اولیه نشان داد که مناسب‌ترین تبدیلات تمامی حالات در بازه $[-1,0]$ است. تبدیلات به گونه‌ای انتخاب شدند که الگوی نهایی انتخاب شده بتواند هر دو دیدگاه آماری و فیزیکی را در درون‌یابی و برون‌یابی رعایت کند. ملاحظات فیزیکی و آماری الگوی انتخابی را با تبدیل $x^{-0/7}$ ترجیح داد (جدول ۱-ب). زیرا این تبدیل در دو حالت درون‌یابی و برون‌یابی از نظر فیزیکی جواب معقول‌تری می‌دهد. ضریب تعیین اصلاح شده $R^2 = 0/716$ ، $SSE = 1159997$ و $GCV = 26089$ است (جدول ۱-ب). این موارد نشان از خوبی نیکویی برازش دارد. رفتار باقی‌مانده‌ها و نمودار نکویی برازش (شکل ۲و۴) موید مناسب بودن الگوست. الگوی نهایی یک گره‌ای است. گره در نقطه $2/0852$ قرار دارد. موارد فوق مناسب بودن الگوی MARS را از دیدگاه آماری تایید می‌کند. TDS درون‌یابی و برون‌یابی شده با الگوی MARS (جدول ۳) نشان از تبیین فیزیکی مناسب این الگو دارد. برون‌یابی تا دبی‌های ۵۰۰ مترمکعب بر ثانیه انجام که نتایج در جدول (۳) آمده است. TDS مربوط به دبی‌های $0/01$ (کم‌آبی‌های شدید) و ۵۰۰ مترمکعب بر ثانیه (سیلاب‌های بزرگ) به ترتیب $1309/6$ و $284/7$ است که همخوانی خوبی نیز با نمونه‌های مشاهداتی دارد.

مقایسه الگوهای برازشی: جدول (۳) مقادیر برآوردی TDS را با MARS، نمایی و لگاریتمی را در دو حالت درون‌یابی و برون‌یابی نشان می‌دهد. دبی داده‌های مشاهده‌ای با حداقل $0/01$ و حداکثر $4/16$ مترمکعب بر ثانیه به ترتیب دارای TDSهای 1247 و 456 است. کمترین TDS مشاهده‌ای 308 به ترتیب مربوط به دبی $1/03$ است. این اعداد نشان می‌دهند که مقدار TDS در دبی‌های بزرگ (عمدتاً سیلاب‌ها) کمتر از ۵۰۰ است. این موضوع فقط تطابق خوبی با الگوی MARS دارد (جدول ۳). دبی‌های $0/01$ ، 1 و 100 مترمکعب بر ثانیه برای مقایسه TDSهای برآوردی سه الگوی MARS، نمایی و لگاریتمی انتخاب شد. TDSهای برآوردی در جدول (۳) ارائه شده‌اند. این TDSها برای الگوی MARS به ترتیب: $1309/6$ ، $617/7$ ، $347/7$ و $284/7$ است. این TDSها برای الگوی نمایی به ترتیب: 647 ، 785 و $-\infty$ و برای الگوی لگاریتمی به ترتیب: 706 ، 434 و $0/0$ است. مقایسه این اعداد نشان می‌دهد که دو الگوی نمایی و لگاریتمی جواب‌های نامناسبی در این برون‌یابی (حتی در درون‌یابی) ارائه می‌کنند. اما الگوی MARS در دو حالت درون‌یابی و برون‌یابی از دقت خوبی برخوردار است. بنابراین الگوی MARS با تبدیل نردبانی مناسب می‌تواند TDS آب رودخانه را به خوبی در دراز مدت و به طور معقول الگوبندی و تبیین کند.

نتیجه گیری

روش‌های مرسوم برآورد Ec و TDS استفاده از الگوهای نمایی و لگاریتمی است. این الگوها برای درون‌یابی مناسب اما در برون‌یابی‌ها خوب عمل نمی‌کنند (جدول ۳). سه الگوی نمایی، توانی و MARS (با نردبان تبدیل) بر داده‌ها (پس از غربال) آب سنجی پنج ایستگاه آب سنجی در استان خراسان جنوبی برازش داده شد. دو دیدگاه فیزیکی و آماری نتایج تحلیل ایستگاه نشان می‌دهد که الگوهای نمایی، توانی در برون‌یابی‌ها خوب عمل نمی‌کنند (جدول ۴). تحلیل و مقایسه الگوی MARS با الگوهای فوق به دو صورت آماری و فیزیکی انجام شد. تحلیل آماری با معیارهای برازش الگو مانند: ضریب تعیین، مجموع مربعات خطا، اعتبار متقابل تعمیم یافته (GCV)، رفتار خطاها و غیره انجام شد (جدول ۱ تا ۴). بررسی فیزیکی در دو حالت درون‌یابی و برون‌یابی با در نظر گرفتن کران‌های Ec و TDS حوضه‌های آبریز انجام شد (جدول ۳). الگویی مناسب است که بتواند برون‌یابی و درون‌یابی را به طور همزمان بهتر تبیین کند. دو الگوی نمایی و لگاریتمی در برون‌یابی عملکرد مناسبی نداشتند. الگوی مبتنی بر روش رگرسیون کمانک تطبیقی چند متغیره (MARS) در مقایسه با دو الگوی قبل در دو حالت درون‌یابی و برون‌یابی عملکردی مناسب دارد (جدول ۳ و ۴). جداول (۳ و ۴) نشان از برتری الگوی MARS دارد. زیرا این الگو علاوه بر نیکویی برازش آماری، دیدگاه فیزیکی Ec و TDS آب را نیز رعایت و در برون‌یابی عملکرد مناسبی دارد (اشکال ۲ تا ۵). لذا کاربرد رگرسیون MARS با نردبان تبدیل برای الگوی بندی Ec و TDS توصیه می‌شود.

جدول ۱- الف: مشخصات الگوهای برازشی بر شوری ایستگاه‌های آب سنجی استان خراسان جنوبی

ایستگاه	الگو	تعداد گره	GCV	SSE	R ^۲
افین	$y = 4067.177 - 4671.255 \times h(x^{-0/25} - 1.88622) - 2224.290 \times h(1.88622 - x^{-0/25}) + 9474.191 \times h(x^{-0/25} - 2.06495)$	۲	206874	4633986	0.58
فرخی	$y = 7716.5907585.089 \times h(1.17565x^{-0/15}) + 6481.137 \times h(x^{-0/15} - 1.32919)$	2	2169164	49416270	0.65
خونیک	$y = 7821.523 - 1364.018 \times h(x^{-0/4} - 1.7411) - 4175.383 \times h(1.7411 - x^{-0/4}) + 2628.087 \times h(x^{-0/4} - 3.08134)$	۲	1319995	43645908	0.74
در بند	$y = 1732.3491 - 657.8436 \times h(2.18218 - x^{-0/5})$	1	52354	2521744	0.75
منصورآباد	$y = 3210.5722 - 302.8849 \times h(\ln(x)) + 7.6009$	1	476994	1907977	0.5

جدول ۱- ب: مشخصات الگوهای برازشی بر TDS ایستگاه‌های آب سنجی استان خراسان جنوبی

ایستگاه	مدل	تعداد	GCV	SSE	R ^۲
افین	$y = 2704.1210 - 4381.5613 \times h(x^{-0/3} - 2.43126) - 969.3923 \times h(2.43126 - x^{-0/3}) + 4895.4557 \times h(x^{-0/3} - 2.53639)$	۲	100884	2259795	0.55
	$y = 5253.809 - 2754.222 \times h(1.46505 - x^{-0/3}) + 1218.707 \times h(x^{-0/3} - 1.76674)$	۲	903266	21459406	0.68
خونیک	$y = 4743.084 - 2838.506 \times h(1.6245 - x^{-0/35}) - 2998.173 \times h(x^{-0/35} - 2.32282) + 4345.961 \times h(x^{-0/35} - 2.67699)$	۳	603661	19960197	۳۰.7
دربند	$y = 983.74696 + 14.14513 \times h(x^{-0/7} - 2.08522) - 337.32215 \times h(2.08522 - x^{-0/7})$	1	۱26089.	1159997	0.716
منصورآباد	$y = 7.8819321 - 0.2395742 \times h(x) + 7.6009$	1	۸0.2029	0.811917	0.68

جدول ۲- نمونه‌های مشاهده‌ای و برآوردی شوری ایستگاه خونیک و TDS ایستگاه دربند با استفاده از روش MARS

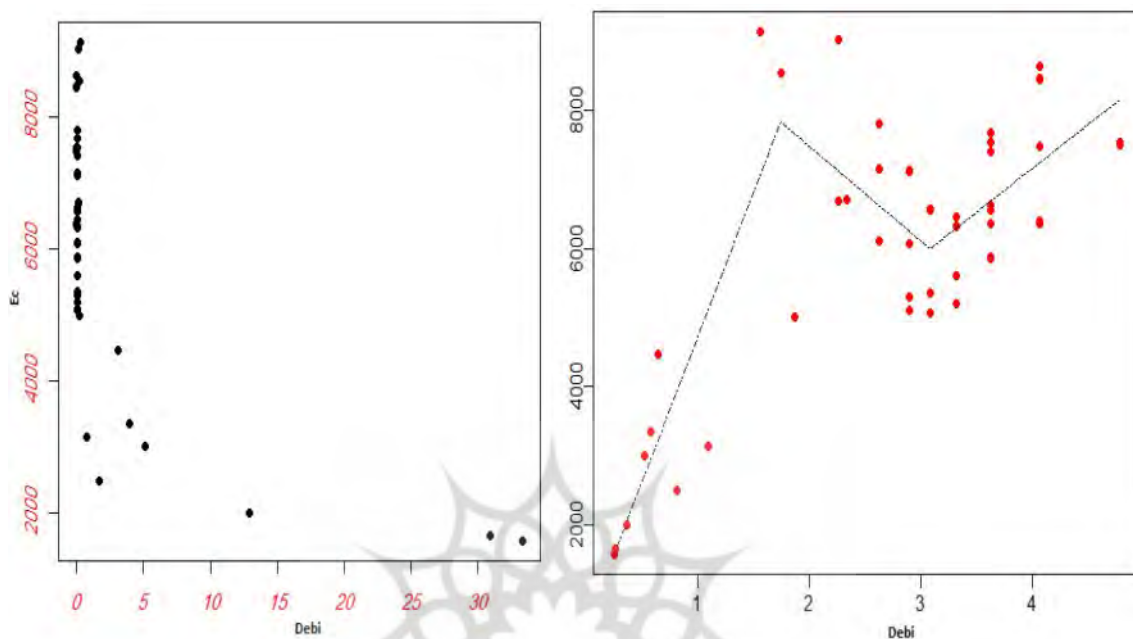
ایستگاه خونیک						ایستگاه دربند					
دبی	مشاهده‌ای	برآوردی	دبی	مشاهده‌ای	برآوردی	دبی	مشاهده‌ای	برآوردی	دبی	مشاهده‌ای	برآوردی
0.25	8540	7205.68	0.09	6100	7205.68	0.25	8540	7821.52	0.06	5350	5993.41
0.33	9120	6574.8	0.04	6350	6622.2	0.33	9120	7057.41	0.21	5000	7649.98
0.03	8620	7246.11	0.04	5850	6622.2	0.03	8620	7237.9	0.05	5200	6288.09
0.03	8440	7246.11	0.04	5875	6622.2	0.03	8440	7237.9	0.07	5300	6244.74
0.03	8450	7246.11	0.07	6075	6383.95	0.03	8450	7237.9	0.07	5100	6244.74
0.13	9020	7205.68	1.68	2500	3890.18	0.13	9020	7111.5	33.32	1590	1578.86
0.04	7670	6622.2	0.8	3150	4927.24	0.04	7670	6679.25	30.94	1657	1609.76
0.04	7530	6622.2	4	3350	2974.31	0.04	7530	6679.25	12.94	2016	2051.18
0.03	7480	7246.11	0.06	5075	5842.97	0.03	7480	7237.9	0.05	6450	6288.09
0.02	7500	8239.28	0.06	5350	5842.97	0.02	7500	8142.87	0.06	6555	5993.41
0.02	7530	8239.28	0.21	5000	7205.68	0.02	7530	8142.87	0.05	6325	6288.09
0.04	7400	6622.2	0.05	5200	6179.71	0.04	7400	6679.25	0.05	5600	6288.09
0.13	6685	7205.68	0.07	7125	6383.95	0.09	6100	6622.67	0.04	6555	6679.25
0.03	6355	7246.11	0.09	7800	7205.68	0.04	6350	6679.25	0.07	7115	6244.74
0.03	6390	7246.11	0.09	7150	7205.68	0.04	5850	6679.25	5.11	3010	2726.1

جدول شماره ۳: برآورد Ec ایستگاه خونیک و TDS ایستگاه دربند با الگوهای مختلف در حالت درون یابی و برون یابی

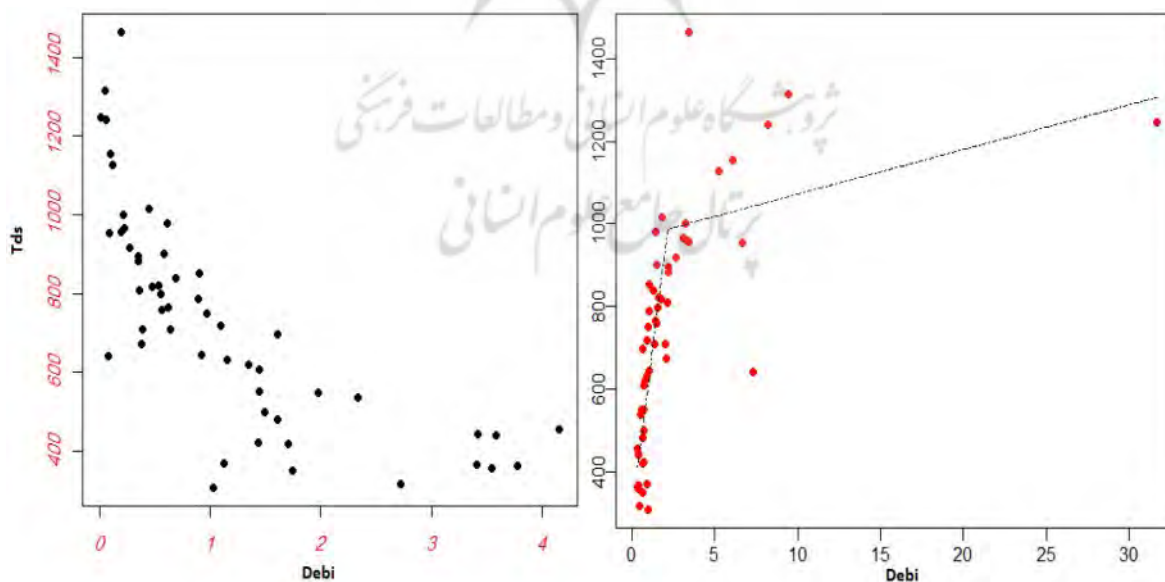
دبی	درون یابی Ec ایستگاه خونیک			درون یابی TDS ایستگاه دربند			برون یابی Ec ایستگاه خونیک			برون یابی TDS ایستگاه دربند			
	MARS	نمایی	لگارتیمی	MARS	نمایی	لگارتیمی	دبی	MARS	نمایی	لگارتیمی	MARS	نمایی	لگارتیمی
0.01	10074.1	6210	5974.423	1309.6	785	706	4	2949.9	$10^{-8.71}$	447.6613	408.2	$10^{-5.69}$	$10^{-1.65}$
0.02	8142.9	6210	5971.491	1173	784	703	5	2745.1	$10^{-4.82}$	4.795577	389.7	$10^{-3.15}$	$10^{-4.0910}$
0.03	7237.9	6210	5968.531	1118.9	784	701	10	2214	$-\infty$	0	347.7	$-\infty$	0
0.04	6679.2	6210	5965.543	1088.9	784	699	15	1965.1	$-\infty$	0	331	$-\infty$	0
0.05	6288.1	6210	5962.526	1069.4	783	697	20	1811.5	$-\infty$	0	321.8	$-\infty$	0
0.1	6770.2	6210	5947.006	1025.1	782	687	30	1622.9	$-\infty$	0	311.5	$-\infty$	0
0.5	6061.2	6210	5792.738	828.3	758	588	50	1424.9	$-\infty$	0	302.2	$-\infty$	0
1	4727.1	6210	5500.766	617.7	647	434	100	1213.5	$-\infty$	0	293.8	$-\infty$	0
2	3716.1	6210	4388.714	488	-17100	115	200	1053.3	$-\infty$	0	288.6	$-\infty$	0
3	3242.4	6210	2375.265	436.7	$10^{-5.84}$	3.08	500	899.4	$-\infty$	0	284.7	$-\infty$	0

جدول شماره ۴- مقایسه معیارهای آماری الگوهای برازشی

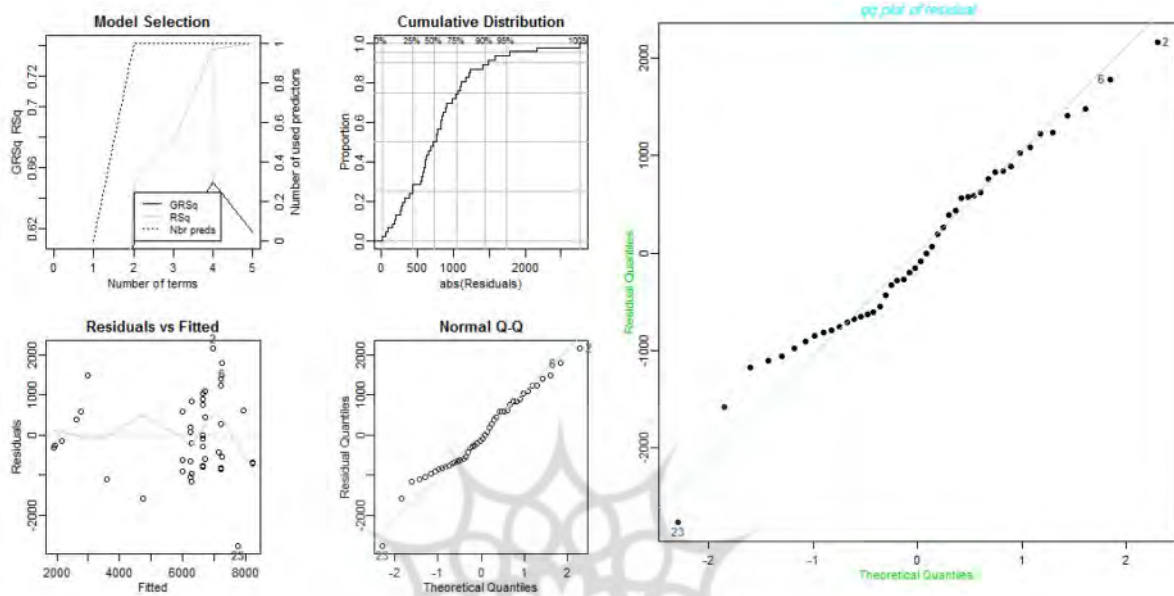
ایستگاه	الگو	رابطه دبی - شوری			رابطه دبی - TDS		
		مجموع مربعات خطا	ضریب تبیین	ضریب تبیین	مجموع مربعات خطا	ضریب تبیین	ضریب تبیین
شوری	MARS	SSE 4633986	R^2 0.58	SSE 2259795	R^2 0.55		
	نمایی	SSE 3602252	R^2 0.902	SSE 3400786	R^2 0.3029		
	توانی	SSE 4827056	R^2 0.57	SSE 2703897	R^2 0.5841		
شوری	MARS	SSE 49416270	R^2 0.653	SSE 21459406	R^2 0.68		
	نمایی	SSE 50343675	R^2 0.6349	SSE 20683560	R^2 0.677		
	توانی	SSE 73748160	R^2 0.6388	SSE 34319241	R^2 0.6174		
خونیک	MARS	SSE 43645908	R^2 0.734	SSE 19960197	R^2 0.707		
	نمایی	SSE 65605247	R^2 0.1297	SSE 1219446	R^2 0.1271		
	توانی	SSE 96899964	R^2 0.5984	SSE 1787030	R^2 0.605		
دربند	MARS	SSE 2521744	R^2 0.75	SSE 1159997	R^2 0.716		
	نمایی	SSE 2962208	R^2 0.7013	SSE 1219446	R^2 0.6963		
	توانی	SSE 4595847	R^2 0.5739	SSE 1787030	R^2 0.6097		
منصور آباد	MARS	SSE 1907977	R^2 0.5	SSE 0.811917	R^2 0.68		
	نمایی	SSE 1907977	R^2 0.5307	SSE 0.811917	R^2 0.6326		
	توانی	SSE 381165	R^2 0.9058	SSE 0.0820504	R^2 0.9629		



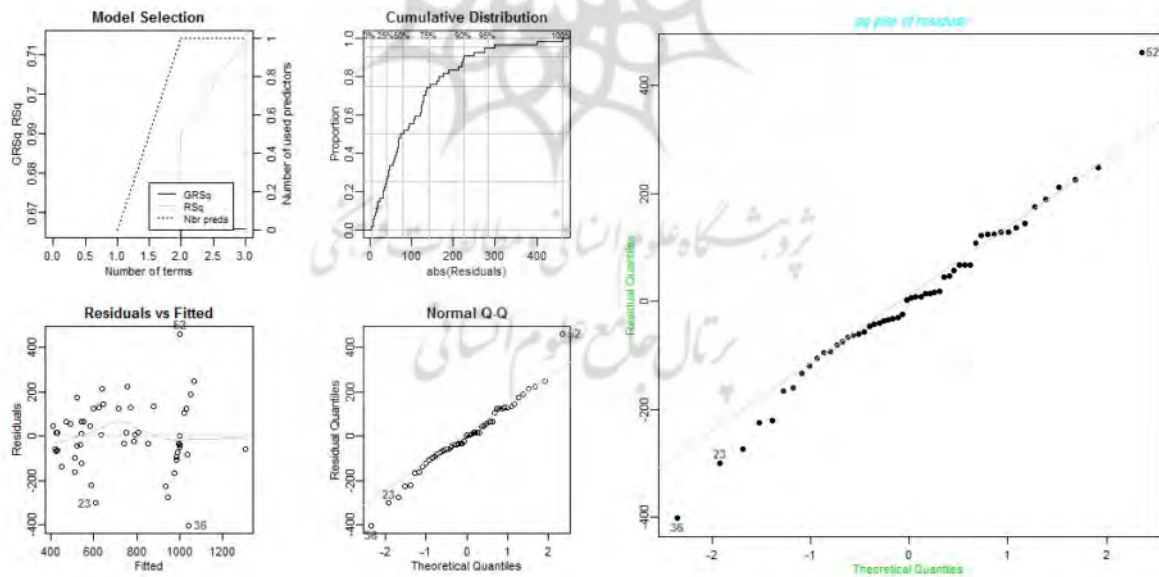
شکل ۲: نمودار پراکنش دبی - شوری ایستگاه خونیک (سمت چپ)، برازش الگوی MARS (سمت راست)



شکل ۳: نمودار پراکنش دبی - TDS ایستگاه دربند (سمت چپ)، برازش الگوی MARS (سمت راست)



شکل ۴: نمودارهای آسیب‌شناسی برازش الگوی مارس بر داده‌های شوری ایستگاه خونیک



شکل ۵: نمودارهای آسیب‌شناسی برازش الگوی مارس بر داده‌های TDS ایستگاه دربند

منابع:

۱. اسدالله فردی، غلامرضا، قنبری، علی، کنعانی، شهیر (۱۳۸۷)، پیش بینی شوری با استفاده از شبکه های عصبی مصنوعی مطالعه موردی. حوضه آبریز رودخانه آجی چای. چهارمین کنگره ملی مهندسی عمران. دانشگاه تهران.
۲. اسدالله فردی، غلامرضا، تکلیفی، آیدین و قنبری، علی، (۱۳۸۸)، استفاده از روش شبکه عصبی مصنوعی در پیش بینی مقادیر TDS در رودخانه تلخه رود. سومین همایش تخصصی مهندسی محیط زیست.
۳. آرشیو سازمان مدیریت منابع آب وزارت نیرو.
۴. بختیاری، مرتضی، کاشفی پوردزفولی، سید محمود، عزیزی مبصر، جوانشیر، (۱۳۸۶)، استفاده از شبکه های عصبی در ارزیابی کیفی رودخانه کرخه. ششمین کنفرانس هیدرولیک ایران.
۵. رضائی پزند حجت و جانفدا مجید، (۱۳۹۰)، کاربرد رگرسیون مارس در برآورد نسبت جذب سدیم مثال کاربردی: خراسان شمالی. حوضه آبریز دربند سملقان. چهارمین کنفرانس مدیریت منابع آب ایران. دانشگاه امیر کبیر.
۶. قاسمیپور، لیلا، مصطفوی، رویا و دانشور، نرگس، (۱۳۸۹)، تعیین رابطه بین مواد جامد در آب TDS با دبی در محدوده ایستگاه های هیدرومتری بلده و رزن رودخانه نور. اولین همایش ملی مدیریت منابع آب اراضی ساحلی.
۷. قبادیان، رسول، مشایخی، حسن، ۱۳۸۷، شبیه سازی پارامترهای کیفی آب رودخانه قره سو در محل ایستگاه قرباغستان با استفاده از شبکه عصبی مصنوعی، سومین کنفرانس مدیریت منابع آب.
۸. کرمی، مهدی (۱۳۸۵)، پیش بینی کیفیت آب رودخانه کارون با استفاده از شبکه های عصبی مصنوعی. هفتمین سمینار بین المللی مهندسی رودخانه، دانشگاه اهواز.
۹. محجوبی، عماد و تجریشی، مسعود، ۱۳۸۹، مقایسه عملکرد الگوریتم های شبکه عصبی مصنوعی و درختان تصمیم گیری در پیش بینی تغییرات شوری آب رودخانه ها - مطالعه موردی: رودخانه کارون، چهارمین همایش و نمایشگاه تخصصی مهندسی محیط زیست.

10. Lo. Shuh-shiaw. (1992). "Glossary of hydrology", Water resources Pub.
11. Barnet. V. Lewis. T. (1994). "Outliers in statistical data", John Wiley & Sons.
12. Sarmad, M. (2012), "Detecting Outliers in Data from Exponential Distribution", Heber International Conference on Applications of Mathematics and Statistics.
13. Weisberg. S. (2005). "Applied Linear Regression, John Wiley & Sons", INC. PUB.
14. Freedman, D. A. (2009), "Statistical Models: Theory and Practice", Cambridge.
15. Sheather. S. J. (2009). "A modern approach to regression with R", Springer.
16. Ryan. S.E. et al. (2002), "difining phase of bedload transport using piecewise regression", J. Earth Surf. Process, Landforms, 27, 971-999.
17. Friedman. J.H. (1991), "Multivariate adaptive regression splines (with discussion)". J. Annals of Statistics, 19, 1-141.
18. Hastie, T, et al, (2009), "The elements of Statistical Learning Data Mining, Inference and Prediction", 2th Ed., Univ. Press.
19. Palani S. et al, (2008), "An ANN Application for Water Quality Forecasting", J. of Marine Pollution Bulletin.
20. Chavez P., Chang F., (2006), "Simulation of Multiple Water Quality Parameters Using Artificial Neural Networks", 7th International Conference on Hydro information, 2006
21. www.R-project.com.