

پیش‌بینی سری‌های زمانی کیفیت آب رودخانه سفیدرود با استفاده از مدل‌های خطی تصادفی

بهزاد جمالی و کیومرث ابراهیمی**

* نگارنده مسئول، نشانی: کرج، پردیس کشاورزی و منابع طبیعی، دانشکده مهندسی و فناوری کشاورزی، گروه مهندسی آبیاری و آبادانی.

کد پستی: ۳۱۵۸۷-۷۷۸۷۱، تلفن: ۱۱۰ - ۲۲۲۶۱۸۱ (۰۲۶۱)، پیام‌نگار: Ebrahimik@ut.ac.ir

** به‌ترتیب دانشجوی کارشناسی ارشد مهندسی منابع آب و استادیار گروه مهندسی آبیاری و آبادانی دانشگاه تهران

تاریخ دریافت: ۱۰/۱/۸۹؛ تاریخ پذیرش: ۹۰/۴/۱۲

چکیده

پیش‌بینی کیفیت آب رودخانه‌ها به‌منظور مدیریت مناسب حوضه آنها ضروری است، تا بتوان برای کنترل مقدار آلاینده‌ها و رساندن آنها به حد مجاز گام‌هایی برداشت. در مقاله حاضر، قابلیت پیش‌بینی سری‌های زمانی پارامترهای هدایت الکتریکی و کلر ایستگاه آستانه از رودخانه سفیدرود با استفاده از مدل‌های خطی تصادفی بررسی شده است. به‌منظور پیش‌بینی فصلی سری‌های زمانی پارامترهای مذکور، از مدل خودهمبسته میانگین متحرک فصلی انباشته (SARIMA) استفاده شد. برای برازش این مدل از سری‌های زمانی بین سال‌های ۱۳۷۰ تا ۱۳۸۱، و به‌منظور صحت‌سنجی مدل از داده‌های سال‌های ۱۳۸۲ تا ۱۳۸۴ استفاده شد. ابتدا روند و ایستایی سری‌های زمانی هدایت الکتریکی و کلر به‌ترتیب با آزمون‌های من-کندال بررسی و پس از آن الگوهای فصلی ۱۲ ماهه پس از تفاضل‌گیری مرتبه اول حذف شد. مدل‌سازی یک سری زمانی به‌طور کلی شامل سه مرحله است: شناسایی مدل، برآورد پارامترهای مدل، و کنترل تشخیصی. در مرحله شناسایی، با استفاده از توابع خودهمبسته و خودهمبسته جزئی مدل‌های مختلف SARIMA شناسایی و مدلی که کمترین مقدار ضریب آکائیک را داشت به‌عنوان بهترین مدل برازش داده شده انتخاب شد. پارامترهای مدل با استفاده از روش تخمین حداقل مربعات باقیمانده‌ها برآورد شد. در مرحله کنترل تشخیصی، مشاهده شد که باقیمانده مدل‌ها مستقل، همسان واریانس و دارای توزیع نرمال است. سری‌های زمانی هدایت الکتریکی و کلر برای سال‌های ۱۳۸۲ تا ۱۳۸۴ توسط مدل‌های انتخاب شده پیش‌بینی شد. به‌منظور صحت‌سنجی مدل، میانگین و واریانس داده‌های مشاهده‌ای و پیش‌بینی شده مقایسه شدند و اختلاف معنی‌داری بین آنها مشاهده نشد. ریشه خطای مربعات متوسط برای پارامترهای هدایت الکتریکی و کلر به‌ترتیب برابر ۲۷۸/۹ و ۲/۲۲ برآورد شد. به این ترتیب با اطمینان بالا می‌توان روش مدل‌سازی SARIMA را برای پیش‌بینی مقادیر هدایت الکتریکی و غلظت کلر در ایستگاه آستانه از رودخانه سفیدرود پیشنهاد کرد.

واژه‌های کلیدی

سری زمانی، کلر، من-کندال، هدایت الکتریکی، SARIMA

مقدمه

بعضی از آنها آلودگی را حتی با چشم می‌توان دید. حوضه آبریز سفیدرود در شمال غرب ایران و در محل تلاقی رشته کوه‌های البرز، زاگرس قرار دارد و دارای دو سرشاخه اصلی است. سرشاخه شاهرود در جهت جنوب شرقی - شمال غربی و سرشاخه قزل اوزن در جهت جنوب

گسترش مراکز جمعیتی و فعالیت‌های کشاورزی، صنعتی، و معدنی در حوضه آبریز رودخانه‌ها بدون دقت کافی به مسائل زیست محیطی سبب توسعه آلودگی در بسیاری از رودخانه‌های کشور شده است، تا آنجا که در

آبی است. در این بازه، چندین سرشاخه با دبی‌های قابل توجه به رودخانه تخلیه می‌شوند که تغییراتی را در دبی رودخانه سفیدرود ایجاد می‌کنند.

در منطقه پایاب سد سفیدرود، پارامتر شوری نقشی مهم در کاهش تولید محصولات زراعی دارد. در اکثر گیاهان، آنیون کلر موجود در منابع آب و خاک عامل اصلی کاهش رشد و عملکرد محصولات است. کرباسی و شهبازی (Karbasi & Shahbazi, 2008) با استفاده از تکنیک آماری تحلیل عاملی نشان دادند که پارامترهای کلر (Cl) و هدایت الکتریکی (EC) از مهمترین عوامل مؤثر در کیفیت آب رودخانه‌های استان گیلان هستند. مشخصات آماری پارامترهای EC و Cl مربوط به دوره زمانی ۱۳۷۰ تا ۱۳۸۴ از ایستگاه‌های هیدرومتری آستانه و رودبار (شکل ۱) در جدول (۱) ارائه شده است. غلظت کلر در طول رودخانه تغییر زیادی ندارد. شبیه‌سازی پارامتر هدایت الکتریکی در طول رودخانه سفیدرود در ۴ ماه از سال ۱۳۸۷ به کمک مدل واسنجی شده QUAL2K نشان می‌دهد که این پارامتر تحت تأثیر منابع گسترده در طول رودخانه سفیدرود نیست.

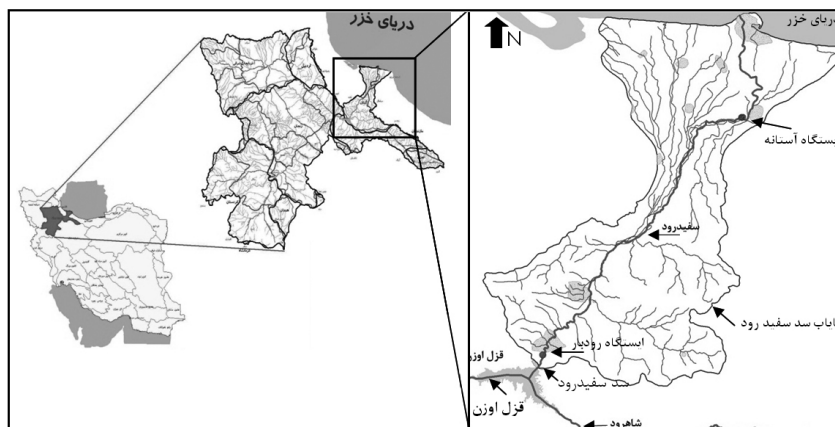
پیش‌بینی هدایت الکتریکی و غلظت کلر می‌تواند نقشی مهم در مدیریت و برنامه‌ریزی موضوع کاهش تأثیرات آلودگی سفیدرود داشته باشد. در این خصوص می‌توان از مدل‌های خودهمبسته میانگین متحرک^۱ (ARMA)، خودهمبسته میانگین متحرک انباشته^۲ (ARIMA)، روش‌های تشخیص الگو^۳، و مدل‌های فیزیکی استفاده کرد.

به شمال جریان دارد (شکل ۱). محدوده مورد مطالعه در این تحقیق، پایاب سد سفیدرود است با مساحتی حدود ۲۸۳۵ کیلومتر مربع که از بعد از سد سفیدرود شروع می‌شود و پس از پیمودن حدود ۱۱۵ کیلومتر و عبور از شهر رودبار، رستم آباد، توتکابن، آستانه اشرفیه، و کياشهر به دریای خزر می‌ریزد. جریان آب تنظیمی از سد سفیدرود در محدوده شبکه آبیاری سفیدرود توزیع می‌شود و شالیزارهای زیردست را آبیاری می‌کند. حدود ۸۰ درصد آب مورد نیاز ۲۳۰۰۰۰ هکتار از شالیزارهای استان گیلان از شبکه آبیاری سفیدرود تأمین می‌شود. به‌طور کلی بازدیدهای میدانی، سنجش‌ها، و ارزیابی‌ها نشان می‌دهد که رودخانه سفیدرود با مشکلات متعددی روبرو است مانند: تخلیه فاضلاب‌های خام شهری و صنعتی، ورود زه‌آب‌های کشاورزی، برداشت مصالح رودخانه‌ای، فرسایش، و منابع آلاینده طبیعی. بسیاری از بازه‌های رودخانه بالقوه آسیب پذیر است، ولی شدت آلودگی تنها در برخی از بازه‌های این محدوده از رودخانه بالا است و اولویت بالایی نیز برای اجرای برنامه‌های کنترل و کاهش آلودگی دارند. این بازه‌ها اغلب در محدوده مراکز جمعیتی و صنعتی قرار دارند، بنابراین میزان مخاطرات ناشی از آلودگی آنها بالا است.

در طول رودخانه سفیدرود، ۴۲ آبادی، هر یک با فاصله‌ای کمتر از ۱۰۰۰ متر نسبت به محور و در دو طرف رودخانه قرار دارند و منابع آلاینده نقطه‌ای زیادی در طول رودخانه به ثبت رسیده است. کاربری اراضی زیردست سد در مجاور این رودخانه، علاوه بر مسکونی پراکنده، از کیلومتر صفر تا ۱۸ مرتع و از کیلومتر ۱۸ تا ۱۱۵ زراعت

1-Autoregressive Moving Average
3-Pattern recognition

2-Autoregressive Integrated Moving Average



شکل ۱- موقعیت کلی محدوده مورد مطالعه

جدول ۱- پارامترهای آماری EC و CI در ایستگاه‌های هیدرومتری رودبار و آستانه

| پارامتر | ایستگاه | دوره | کمینه | بیشینه | میانگین | انحراف معیار |
|---------------------------|---------|-----------|-------|--------|---------|--------------|
| EC | رودبار | ۱۳۷۰-۱۳۸۴ | ۴۰۰ | ۲۳۷۴ | ۱۲۷۲ | ۴۱۳ |
| (میکروزیمنس بر سانتی‌متر) | آستانه | ۱۳۷۰-۱۳۸۴ | ۲۵۲ | ۲۳۳۶ | ۱۱۶۸ | ۴۱۵/۶ |
| CI | رودبار | ۱۳۷۰-۱۳۸۴ | ۰/۳۹ | ۱۴/۵ | ۶/۰۴ | ۲/۸۲ |
| (میلی‌گرم بر لیتر) | آستانه | ۱۳۷۰-۱۳۸۴ | ۲ | ۱۵/۶ | ۵/۷۸ | ۲/۸۵ |

یکی از روش‌های پیش‌بینی فرآیندهای پویا و تصادفی در سری‌های زمانی ایستا، استفاده از تحلیل سری زمانی است. به‌منظور پیش‌بینی کیفیت آب رودخانه با روش‌های تصادفی در تجزیه و تحلیل سری زمانی، محققان مطالعات زیادی کرده‌اند. احمد و همکاران (Ahmad *et al.*, 2001) برای پیش‌بینی کیفیت آب رودخانه گنگ در هند به‌منظور مدیریت صحیح حوضه آن رودخانه از مدل ARIMA استفاده کردند و نتیجه گرفتند که در زمان مناسب می‌توان برای کاهش آلاینده‌ها و رساندن آن به حد مجاز استانداردها گام‌هایی برداشت. کورونچ و همکاران (Kurunc *et al.*, 2005) با اندازه‌گیری پتاسیم، سدیم، کلر، منیزیم، کلسیم، سولفات، دما، هدایت الکتریکی، pH، و SAR، عملکرد دو روش SARIMA و توماس - فیرینگ را در پیش‌بینی کیفیت آب و جریان رودخانه در ایستگاه پایش دوروجاسو واقع در رودخانه سیرلماک ترکیه مطالعه کردند. راگاوون و فرناندز (Ragavan & Fernandez,

2006) با استفاده از مدل پارامتری تنظیمی فصلی $ARIMA^1$ در نرم‌افزار SAS، روند طولانی مدت کیفیت آب چند رودخانه را که به‌طور تصادفی انتخاب شدند پیش‌بینی کردند. بن‌یحیی و همکاران (Benyahya *et al.*, 2007) سری زمانی هفتگی حداکثر دمای رودخانه Deschutes را با استفاده از مدل‌های خودهمبسته (AR) و خودهمبسته دوره‌ای (PAR) مدل‌سازی کردند. حسن‌زاده و همکاران (Hassanzadeh *et al.*, 2008) تجزیه و تحلیل سری زمانی، توزیع فراوانی، و پیش‌بینی غلظت SO_2 را برای دوره ۲۰۰۵-۲۰۰۰ در پنج ایستگاه هیدرومتری شهر تهران با استفاده از مدل ARMA بررسی کردند. دوردو (Durdu, 2009) غلظت بور را در ۵ ایستگاه هیدرومتری رودخانه بویوک مندرس^۲ ترکیه با استفاده از مدل‌های ARIMA و SARIMA به مدت ۳۶ ماه پیش‌بینی کرد.

هدف از تحقیق حاضر، توسعه یک مدل خطی تصادفی برای پیش‌بینی میزان هدایت الکتریکی و غلظت

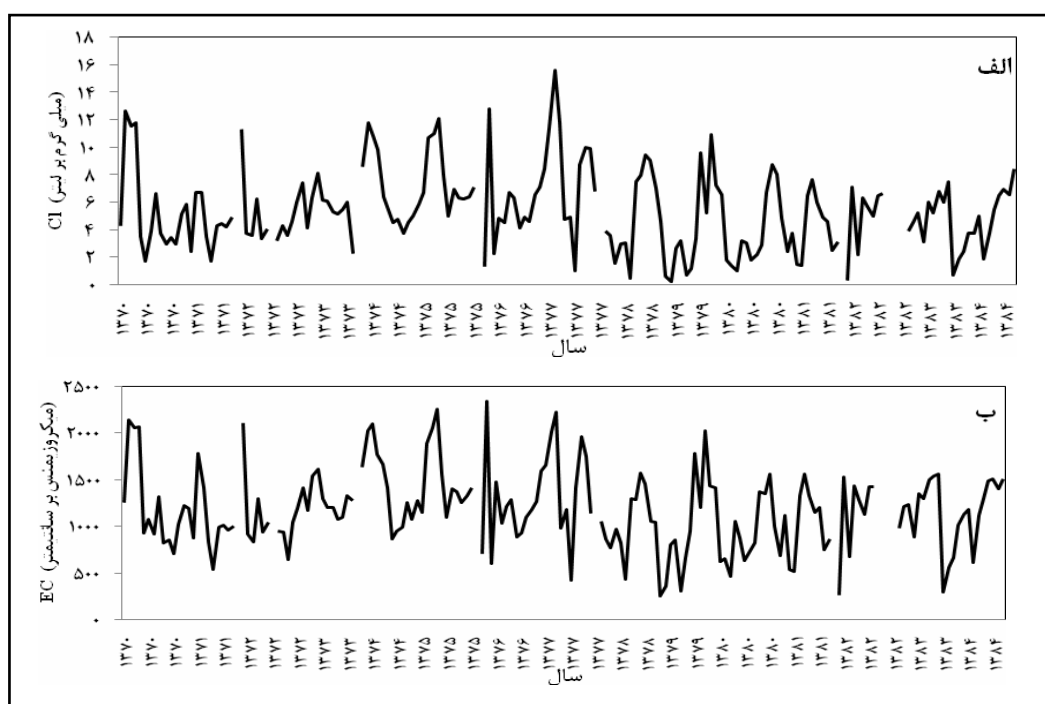
داده‌های ثبت نشده هدایت الکتریکی و کلر ایستگاه آستانه با استفاده از روش میان‌یابی (راگوان و فرناندز) در نرم‌افزار SAS میان‌یابی شد. برای بررسی روند سری زمانی هدایت الکتریکی و کلر از آزمون من-کندال^۱ غیر فصلی استفاده شد. این آزمون غیر پارامتری برای تشخیص روند در داده‌های سری زمانی است و یکی از مزایای آن این است که داده‌ها نیاز به مطابقت با هیچ توزیع خاصی ندارند. مزیت دوم این آزمون حساسیت کم به پرش‌های کوچک در سری‌های زمانی ناهمگن است (Durdu, 2009). این آزمون را ابتدا Mann در سال ۱۹۴۵ ارائه و پس از آن Kendall در سال ۱۹۶۶ توسعه داد. این آزمون نیز مانند دیگر آزمون‌های آماری بر مبنای مقایسه فرض صفر و یک است و در نهایت در مورد پذیرش یا رد فرض صفر تصمیم‌گیری می‌کند.

فرض صفر این آزمون مبتنی بر تصادفی بودن و فقدان روند در سری داده‌هاست و پذیرش فرض یک (رد فرض صفر) دال بر وجود روند در سری داده‌هاست.

کلر ایستگاه آستانه رودخانه سفیدرود است. از بین دو ایستگاه هیدرومتری آستانه و رودبار، ایستگاه آستانه به این دلیل بررسی شد که داده‌های ثبت نشده در آن کمتر بود (شکل ۲). مقایسه میانگین ماهانه پارامترهای هدایت الکتریکی و کلر در ایستگاه‌های آستانه و رودبار نشان می‌دهد که مقادیر میانگین این دو پارامتر در ماه‌های مختلف در سطح ۵ درصد اختلاف معنی‌داری ندارند. بنابراین، پیش‌بینی این دو پارامتر در ایستگاه آستانه می‌تواند برآورد خوبی نیز برای ایستگاه رودبار باشد.

مواد و روش‌ها

داده‌های مورد نیاز این تحقیق شامل داده‌های سری زمانی هدایت الکتریکی و کلر ایستگاه آستانه و رودبار از سازمان تحقیقات منابع آب ایران تهیه شد. شکل ۲ نمودارهای سری زمانی هدایت الکتریکی و کلر را در ایستگاه آستانه برای سال‌های ۱۳۷۰ تا ۱۳۸۴ نشان می‌دهد. بریدگی‌های این نمودارها نشان دهنده دوره‌هایی است که در آنها داده‌ها ثبت نشده است.



شکل ۲- الف: نمودار سری زمانی کلر؛ ب: سری زمانی هدایت الکتریکی ایستگاه آستانه

Z_t جزء استوکستیک سری زمانی، B عملگر انتقال پسرو^۲، و $\varphi(B)$ و $\theta(B)$ به ترتیب چند جمله‌ای‌های از مرتبه p و q هستند (Durdu, 2009). رابطه^۲ ۲ و ۳ به ترتیب مدل $AR(p)$ و $MA(q)$ را نشان می‌دهند.

$$\varphi(B)\nabla^d z_t = \theta(B)a_t \quad (1)$$

$$\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) \quad (2)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \quad (3)$$

باکس و همکاران (Box et al., 1994) مدل $SARIMA$ را برای سری‌های زمانی فصلی توسعه دادند. در صورتی که در یک سری زمانی رفتار تناوبی در فواصل زمانی مشخص (s) مشاهده شود، این سری زمانی دارای تناوب فصلی است و برای مدل‌سازی آن از مدل $SARIMA$ استفاده می‌شود.

مدل $SARIMA$ به صورت $ARIMA(p, d, q)(P, D, Q)s$ نشان داده می‌شود که در آن (p, d, q) بخش غیرفصلی مدل و $(P, D, Q)s$ بخش فصلی مدل است. رابطه^۴ ۴، مدل $SARIMA$ را نشان می‌دهد که در آن: P تعداد جملات خود رگرسیون فصلی، Q تعداد جملات میانگین متحرک فصلی، D تعداد دفعات تفاضل‌گیری مرتبه اول فصلی، s طول فصل، ϕ پارامتر فصلی مدل AR از مرتبه P ، و θ پارامتر فصلی مدل MA از مرتبه Q است (Durdu, 2009).

$$\varphi_p(B)\phi_P(B^s)\nabla^d\nabla_x^D z_t = \theta_q(B)\theta_Q(B^s)a_t \quad (4)$$

همانطور که اشاره شد، استفاده از تفاضل در سری زمانی یکی از روش‌های ایستا کردن آن سری است. یک سری زمانی بعد از تفاضل به صورت رابطه^۵ ۵ بیان می‌شود

به منظور پیش‌بینی سری‌های زمانی پارامترهای هدایت الکتریکی و کلر رودخانه سفیدرود از مدل خودهمبسته میانگین متحرک فصلی انباشته^۱ ($SARIMA$) استفاده شد. دوازده سال (۱۳۷۰ تا ۱۳۸۱) از سری زمانی هدایت الکتریکی و کلر ایستگاه آستانه برای انتخاب بهترین مدل و ۳ سال (۱۳۸۲ تا ۱۳۸۴) برای صحت‌سنجی آن استفاده شد. مدل‌های AR در ترکیب با مدل‌های MA ، مدل عمومی سری زمانی $ARMA$ را تولید می‌کنند. مدل $ARMA$ برای سری‌های زمانی ایستا کاربرد دارد. باکس و جنکینز (Box & Jenkins, 1976) مدل $ARIMA$ را برای سری‌های زمانی غیر ایستا توسعه دادند. با d مرتبه تفاضل‌گیری مرتبه یک، سری‌های زمانی غیر ایستا به سری‌های زمانی ایستا تبدیل می‌شوند. اگر یک سری زمانی پس از d مرتبه تفاضل‌گیری مرتبه اول ایستا و بعد توسط فرآیند $ARMA(p, q)$ مدل‌سازی شود، در این صورت سری زمانی اصلی، خودهمبسته میانگین متحرک فصلی $ARIMA(p, d, q)$ خواهد بود که در آن p تعداد جملات خود رگرسیون، q تعداد جملات میانگین متحرک و d تعداد دفعات تفاضل‌گیری مرتبه اول برای ایستا شدن سری زمانی است. هدف این مدل، شناسایی و تعیین یک مدل آماری است که بتوان آن را مدل تولیدکننده^۶ داده‌های نمونه واقعی از فرآیند تصادفی تعبیر کرد. در صورتی که از مدل $ARIMA$ برای پیش‌بینی استفاده شود ضرورت دارد ویژگی‌های آن در طی زمان ثابت باشد. بنابراین، دلیل نیاز به سری زمانی ایستا این است که هر مدلی که از این سری‌ها به دست آید می‌توان ثابت دانست و مبنایی معتبر برای پیش‌بینی‌ها به شمار آورد (Pokorny, 1987). معادله عمومی مدل $ARIMA(p, d, q)$ را می‌توان به صورت رابطه^۱ ۱ (Box & Jenkins, 1976) ارائه داد که در آن p, d, q مقادیر صفر و یا مثبت هستند (Mishra & Desai, 2005). در این رابطه،

معیار اطلاعات آکائیک^۴ (AIC) استفاده می‌شود (Akaike, 1974). از میان تمام مدل‌ها، معمولاً باقیمانده^۵ مدلی که کمترین مقدار را دارد، بیشترین شباهت را به خطای تصادفی (اختلال سفید)^۶ خواهد داشت (Makridakis et al., 2003). بنابراین، مدلی که کمترین مقدار AIC را دارد به‌عنوان مدل اولیه انتخاب شد. رابطه ریاضی محاسبه AIC به‌صورت رابطه ۷ است، که در آن:

$$m = (p + q + P + Q)$$

تعداد پارامترهای به‌کار رفته در مدل، و RSS مجموع مربعات باقیمانده‌هاست.

$$AIC = n * \left(\ln \left(\frac{2 * \pi * RSS}{n} \right) + 1 \right) + 2 * m \quad (7)$$

پس از تعیین مدل ARIMA، پارامترهای این مدل با روش تخمین حداقل مربعات برآورد شد. روش تخمین حداقل مربعات بر مبنای یافتن تخمین‌گرها با هدف حداقل‌سازی مجموع مربع تفاوت‌ها بین مقادیر مشاهده شده و مقادیر تخمینی بنا شده است. به‌منظور یافتن حداقل مجموع مربعات خطا باید تمام مشتقات جزئی آن نسبت به پارامترهای تخمینی برابر صفر باشد (Malmir, 2006).

در مرحله کنترل تشخیصی، باقیمانده‌های مدل بررسی می‌شود تا درستی مدل و پارامترهای تعیین شده در مرحله قبل ثابت شود. برای بررسی اینکه باقیمانده‌ها مستقل، واریانس همسان^۷، و سری‌ها دارای توزیع نرمال هستند روش‌های مختلفی وجود دارد. یک آزمون ساده برای بررسی استقلال آن است که باید باقیمانده‌های حاصل از مدل اختلال سفید باشند. به عبارت دیگر باقیمانده‌ها باید سری‌هایی باشند که به‌صورت تصادفی از یک توزیع نرمال با میانگین صفر و واریانس ثابت انتخاب شده باشند. وگرنه باید مدل را رد و از ابتدا (مرحله اول)

که در آن، x سری زمانی و s نشانگر تأخیر زمانی در تفاضل است. برای رسیدن به ایستایی، تفاضل‌ها می‌تواند چند بار تکرار شود. برای مثال، برای دو بار تفاضل‌گیری و تأخیر زمانی ۱۲، رابطه ۵ به‌صورت رابطه ۶ بیان می‌شود که در آن w سری جدید حاصل از تفاضل است.

$$u_t = x_t - x_{t-s} \quad (5)$$

$$W_t = u_t - u_{t-12} = (x_t - x_{t-12}) - (x_{t-12} - x_{t-24}) \quad (6)$$

مدل‌سازی سری زمانی شامل ۳ مرحله است: تشخیص، تخمین مدل، و کنترل تشخیصی^۱ (Box & Modares, 2007), (Mishra & Desai, Jenkins, 1976), (Durdu, 2009), (2005)) در مرحله تشخیص، با استفاده از توابع خودهمبسته^۲ (ACF) و خودهمبسته جزئی^۳ (PACF)، مقادیر d و p شناسایی می‌شوند. مرحله تشخیص شامل دو مرحله است: ۱- در صورت لزوم تفاضل‌گیری مرتبه اول؛ ۲- سم توابع ACF و PACF برای تشخیص مقادیر p و q . برای بررسی وجود خودهمبستگی در سری زمانی، ابزار آماری مناسبی است. مقدار ACF در تأخیر صفر برابر ۱ است و همواره در بازه ۱- تا ۱ قرار می‌گیرد. به‌طور کلی، آزمون خودهمبستگی باید برای تأخیر ۱ تا $\frac{n}{4}$ انجام شود که در آن n تعداد کل مشاهدات است. تأخیرهای بیشتر به لحاظ آماری غیرقابل اطمینان هستند (Box & Jenkins, 1976). PACF تأثیر خودهمبستگی در تأخیرهای کوچک‌تر را از همبستگی تأخیرهای بزرگ‌تر حذف می‌کند. به سخی دیگر همبستگی جزئی عبارت است از همبستگی بین Y_t و Y_{t-k} ، پس از حذف تأثیر k تعداد Y میانی. با در نظر گرفتن نمودارهای ACF و PACF سری‌های زمانی هدایت الکتریکی و کلر، مدل‌های مختلف ARIMA برای انتخاب مدل مناسب شناسایی شد. برای انتخاب بهترین مدل از

1- Diagnostic cheking

3- Partial Auto Correlation Function

5- Residuals

7- Homoscedastic

2- Auto Correlation Function

4- Akaike Information Criterion

6- White noise

نشان می‌دهد که باقیمانده‌ها به‌طور مساوی در اطراف میانگین توزیع شده‌اند. برای صحت‌سنجی مدل، از آزمون‌های Z و F به‌ترتیب برای مقایسه میانگین و واریانس داده‌های پیش‌بینی شده و داده‌های مشاهداتی سری زمانی از سال ۱۳۸۲ تا ۱۳۸۴ استفاده شد. همچنین ریشه خطای مربعات متوسط برای هر دو پارامتر از رابطه ۷ برآورد شد:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [z(x_i) - z'(x_i)]^2} \quad (9)$$

که در آن،

$z(x_i)$ = مقدار ثبت شده؛ $z'(x_i)$ = مقدار پیش‌بینی شده؛ N = تعداد داده‌های پیش‌بینی شده؛ و $RMSE$ = ریشه خطای مربعات متوسط است (Karamouz & Araghinejad, 2005), (Ahmad et al., 2001).

نتایج و بحث

بررسی روند و ایستایی

در بررسی روند سری زمانی هدایت الکتریکی و کلر براساس آزمون ناپارامتری من-کندال فرض صفر که مبنی بر عدم وجود روند در سری‌های زمانی مورد بررسی است، در سطح ۵ درصد رد نشد و در نتیجه تفاضل‌گیری غیر فصلی ضروری تشخیص داده نشد. از آنجا که در این تحقیق مدل‌سازی در مقیاس ماهانه مد نظر بوده است، حالت تناوبی ماهانه نیز بررسی شد. شکل ۳ به وضوح وجود تناوب را در میانگین‌های ماهانه مقادیر هدایت الکتریکی و کلر نشان می‌دهد. توابع ACF و $PACF$ این سری‌های زمانی در چند تأخیر اول مضرب ۱۲ تفاوت معنی‌داری با صفر دارد که نشان دهنده دوره تناوب ۱۲ و در نتیجه فقدان ایستایی در سری‌های زمانی مورد بررسی

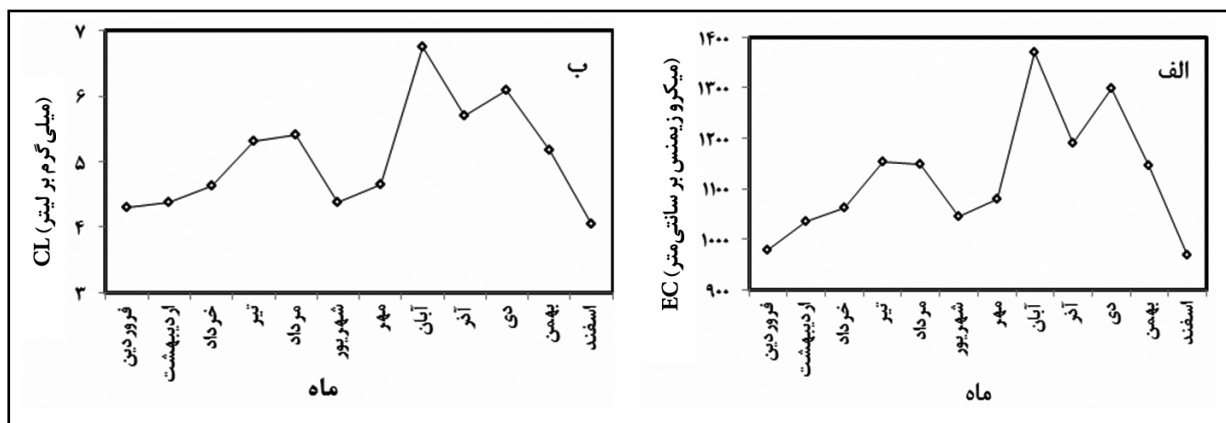
شروع کرد. برای آزمون اول بررسی استقلال باقیمانده‌ها، ACF و $PACF$ باقیمانده‌ها بررسی شد. در صورتی که توابع خودهمبسته و خودهمبسته جزئی باقیمانده‌ها اختلاف معنی‌داری با صفر نداشته باشند، بین باقیمانده‌ها همبستگی وجود ندارد و در نتیجه باقیمانده‌ها مستقل هستند. آزمون لجانگ-باکس^۱ (LB) به‌عنوان آزمون دوم برای بررسی استقلال باقیمانده‌ها به‌کار رفت. به‌منظور تست فرض صفر (اختلال سفید بودن)، آزمون برای تعداد مختلف کل خودهمبستگی‌ها در تأخیرهای پی در پی انجام شد. آزمون LB با رابطه ۸ تعریف می‌شود (Ljung-Box, 1978):

$$LB = n(n+2) \sum_{k=1}^l \frac{r_k^2}{n-k} \quad (8)$$

که در آن،

l = تعداد کل تأخیرهای خودهمبستگی؛ n = تعداد نمونه؛ و r_k = خودهمبستگی باقیمانده‌ها، در تأخیر k است. آزمون LB دارای توزیع χ^2 دو (با درجه آزادی l است. اگر LB محاسبه شده از مقدار بحرانی به‌دست آمده از جدول χ^2 در سطح معین بزرگ‌تر باشد فرض صفر رد می‌شود. برای آزمون نرمال بودن باقیمانده‌ها از رسم هیستوگرام باقیمانده‌ها، رسم احتمال نرمال باقیمانده‌ها و آزمون کلموگروف-اسمیرنف^۲ ($K-S$) باقیمانده استفاده شد. آزمون $K-S$ یک آزمون ناپارامتری برای بررسی برازش داده‌ها بر یک توزیع تئوری است (Durdu, 2009). برای شرح کامل این آزمون به مرجع (Karamouz & Araghinejad, 2005) مراجعه شود. به‌منظور بررسی هم‌واریانس باقیمانده‌ها، مقادیر باقیمانده‌ها در مقابل مقادیر برازش داده شده رسم شد. این نمودار، پراکندگی تصادفی را در اطراف صفر نشان می‌دهد و به سخنی دیگر

است. تفاضل فصلی مرتبه اول این سری‌ها در سطح ۵ درصد نشان داد که سری‌های زمانی فوق ایستا هستند. به این ترتیب، کلیه مدل‌های $ARIMA(p, 0, q)(P, I, Q)_{12}$ در مرحله تشخیص مدل بررسی شدند.



شکل ۳- الف: میانگین ماهانه هدایت الکتریکی، و ب: میانگین ماهانه غلظت کلر در ایستگاه آستانه

تشخیص مدل

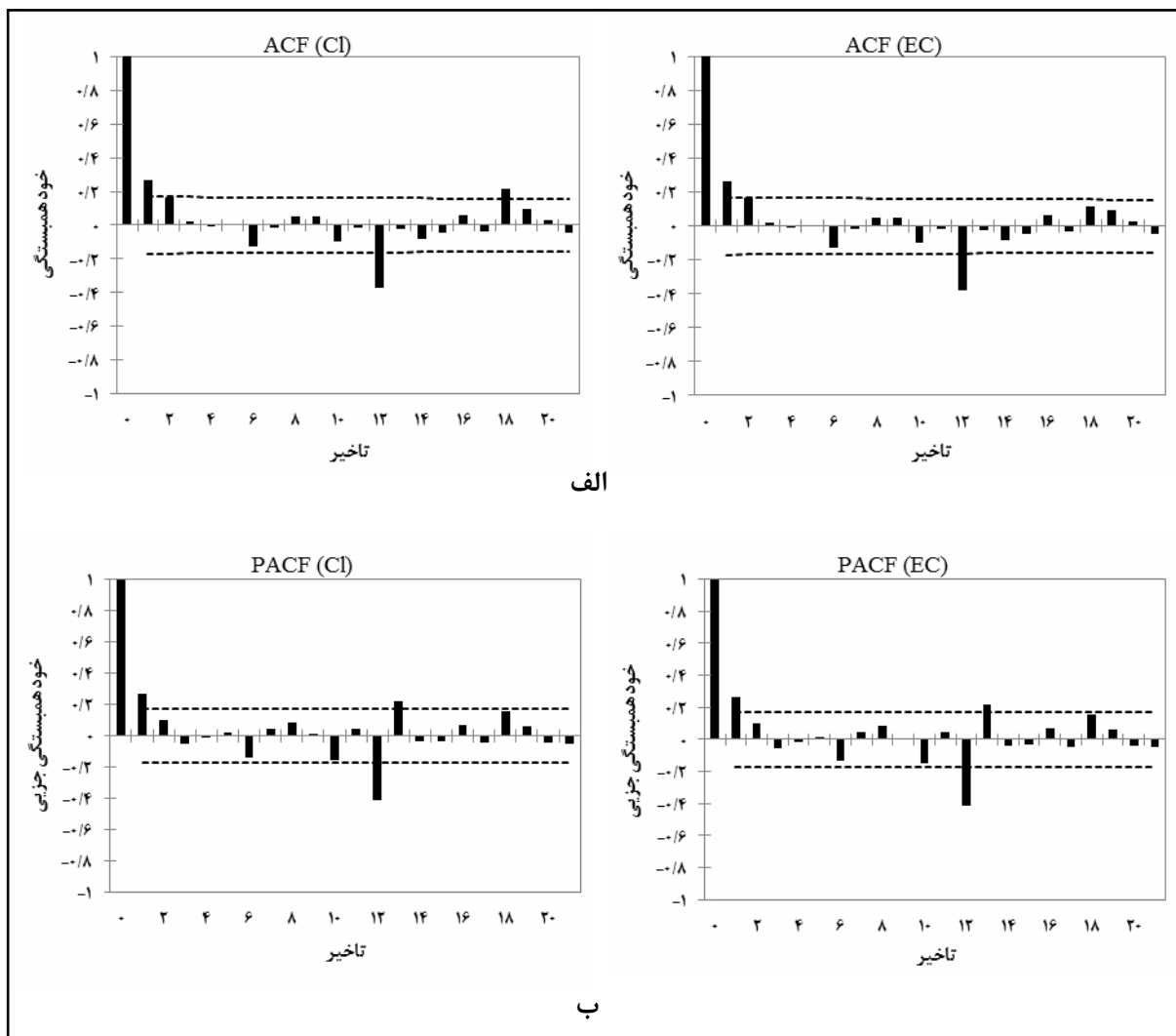
با توجه به این توابع مدل‌های مختلف SARIMA شناسایی شد. از بین مدل‌های مناسب پیش‌بینی شده، مدل‌های فصلی و غیر فصلی که کمترین ضریب AIC را داشتند انتخاب شدند. جدول ۲ مدل‌های انتخاب شده و ضریب AIC متناظر را نشان می‌دهد.

مدل‌های $ARIMA(2,0,1)(1,1,2)_{12}$ و $ARIMA(2,0,1)(1,1,1)_{12}$ که کمترین مقدار AIC را دارند هستند به ترتیب برای پارامترهای کلر و هدایت الکتریکی انتخاب شدند.

شکل ۴. (الف و ب) به ترتیب توابع ACF و PACF تفاضل فصلی مرتبه اول سری زمانی هدایت الکتریکی و کلر را نشان می‌دهد. ACF برای هر دو پارامتر در وقفه‌های ۱، ۲ و ۱۲ معنی‌دار است (شکل ۴. الف). در تمامی دیگر وقفه‌ها این همبستگی‌ها اختلاف معنی‌داری با صفر ندارند. PACF نیز برای هر دو پارامتر در وقفه‌های ۱، ۱۲ و ۱۳ معنی‌دار هستند (شکل ۴. ب). توابع ACF و PACF وجود تناوب فصلی ۱۲ ماهه را به روشنی نشان می‌دهند.

جدول ۲- معیار AIC برای خلاصه‌ای از بهترین مدل‌های انتخاب شده

| AIC | مدل | پارامتر |
|-------|----------------------------|----------------|
| ۶۸۵/۴ | $ARIMA(2,0,1)(1,1,2)_{12}$ | کلر |
| ۶۹۳/۰ | $ARIMA(2,0,1)(1,1,1)_{12}$ | |
| ۶۹۵/۳ | $ARIMA(1,0,1)(1,1,2)_{12}$ | |
| ۶۹۰/۸ | $ARIMA(2,0,1)(1,1,1)_{12}$ | |
| ۸۷۶/۷ | $ARIMA(2,0,1)(1,1,1)_{12}$ | هدایت الکتریکی |
| ۸۸۵/۱ | $ARIMA(2,0,1)(1,1,2)_{12}$ | |
| ۸۸۹/۰ | $ARIMA(1,0,1)(1,1,2)_{12}$ | |
| ۸۸۱/۶ | $ARIMA(2,0,1)(1,1,1)_{12}$ | |



شکل ۴- الف: ACF؛ ب: PACF تفاضل فصلی مرتبه اول سری زمانی کلر و هدایت الکتریکی

تخمین پارامترها

درصد خطا در این جدول ارائه شده است. خطای استاندارد پارامترهای مدل در مقایسه با مقادیر پارامترها کوچک هستند. مقادیر پارامترها در بازه اطمینان ۹۵ درصد قرار دارند و از این رو اختلاف پارامترهای محاسبه شده از صفر، معنی‌دار است.

مقادیر پارامترهای مدل ناشی از برازش مدل‌های $ARIMA(2,0,1)(1,1,2)_{12}$ برای پارامتر کلر و $ARIMA(2,0,1)(1,1,1)_{12}$ برای پارامتر هدایت الکتریکی در جدول ۳ نشان داده شده است. مقدار پارامترهای مدل، خطای استاندارد مربوط به آنها، و حدود اطمینان ۹۵

جدول ۳- پارامترهای مدل‌های انتخاب شده برای پیش‌بینی هدایت الکتریکی و کلر

| پارامتر | پارامترهای مدل | مقدار پارامتر | خطای استاندارد | حدود اطمینان ۹۵ درصد خطای استاندارد | | |
|----------------|----------------|---------------|----------------|-------------------------------------|----------|--|
| | | | | حد بالا | حد پایین | |
| | ثابت | -۰/۰۵۵ | ۰/۰۰ | -۰/۰۵۵ | -۰/۰۵۵ | |
| | AR(1) | ۱/۱۸۲ | ۰/۰۰ | ۱/۱۸۲ | ۱/۱۸۲ | |
| | AR(2) | -۰/۳۷۵ | ۰/۰۰ | -۰/۳۷۵ | -۰/۳۷۵ | |
| کلر | SAR(1) | -۰/۴۰۷ | ۰/۰۰ | -۰/۴۰۷ | -۰/۴۰۷ | |
| | MA(1) | -۰/۶۷۴ | ۰/۰۰ | -۰/۶۷۴ | -۰/۶۷۴ | |
| | SMA(1) | -۰/۳۷۱ | ۰/۰۰ | -۰/۳۷۱ | -۰/۳۷۱ | |
| | SMA(2) | -۰/۲۹۵ | ۰/۰۰ | -۰/۲۹۵ | -۰/۲۹۵ | |
| | ثابت | -۱۶/۲۲۵ | ۱/۹۲۷ | -۱۲/۴۴۸ | -۲۰/۰۰۲ | |
| | AR(1) | ۱/۲۹۰ | ۰/۰۱۸ | ۱/۶۳۷ | ۰/۹۴۳ | |
| | AR(2) | -۰/۳۳۳ | ۰/۰۱۱ | -۰/۱۱۰ | -۰/۵۵۶ | |
| هدایت الکتریکی | SAR(1) | ۰/۱۰۶ | ۰/۰۱۴ | ۰/۳۸۴ | -۰/۱۷۲ | |
| | MA(1) | -۰/۹۱۶ | ۰/۰۱۴ | -۰/۶۳۵ | -۱/۱۹۷ | |
| | SMA(1) | -۰/۷۹۶ | ۰/۰۱۳ | -۰/۵۴۶ | -۱/۰۴۷ | |

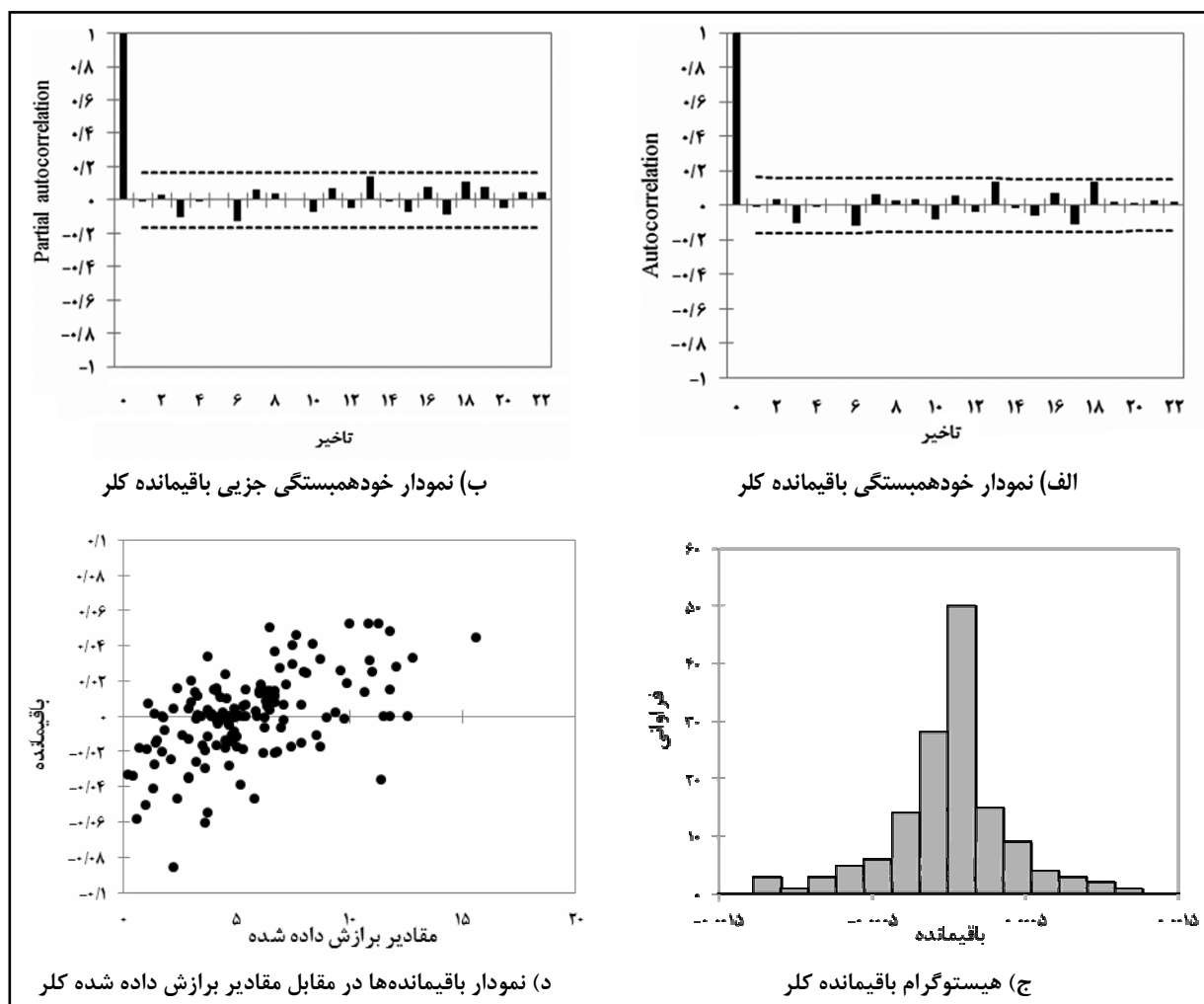
کنترل تشخیصی

نتایج بررسی نرمال بودن باقیمانده‌ها با آزمون K-S در جدول ۴ نشان می‌دهد که باقیمانده سری‌های زمانی هدایت الکتریکی و کلر در سطح ۵ درصد از توزیع نرمال تبعیت می‌کند. با رسم هیستوگرام باقیمانده‌های هدایت الکتریکی و کلر مشاهده می‌شود این هیستوگرام‌ها از فرضیات توزیع نرمال تبعیت می‌کنند. شکل ۵ ج، هیستوگرام باقیمانده کلر را نشان می‌دهد. توزیع فراوانی باقیمانده‌ها متقارن و به شکل توزیع نرمال است. برای بررسی هم‌وابستگی بودن باقیمانده‌ها، نمودار باقیمانده‌ها در مقابل داده‌های برازش شده رسم شد. شکل ۵ د نمودار باقیمانده کلر را در مقابل داده‌های برازش شده این پارامتر نشان می‌دهد. در این شکل مشاهده می‌شود که باقیمانده‌ها به‌طور مساوی در اطراف صفر پراکنده شده‌اند.

توابع ACF و PACF باقیمانده پارامترهای هدایت الکتریکی و کلر رسم و مشاهده شد که در سطح ۵ درصد اختلاف معنی‌داری با صفر ندارند؛ می‌توان نتیجه گرفت که باقیمانده پارامترهای فوق مستقل هستند. شکل ۵ (الف و ب) به ترتیب ACF و PACF باقیمانده‌های کلر را نشان می‌دهد. در تمام تأخیرها مقادیر ACF و PACF بین خطوطی قرار دارند که نشان دهنده حدود اطمینان ۹۵ درصد است. به‌عنوان آزمون دوم برای بررسی استقلال باقیمانده‌ها، از آزمون LB استفاده شد. همانطور که در جدول (۴) آورده شده است، مقادیر LB محاسبه شده از مقادیر χ^2 در سطح ۵ درصد و درجه آزادی مربوط کمتر و در نتیجه باقیمانده‌های محاسبه شده در این مدل‌ها اختلال سفید (مستقل) هستند.

جدول ۴- آزمون لجانگ - باکس و K-S برای باقیمانده‌های سری زمانی کلر و هدایت الکتریکی

| پارامتر | مدل | آزمون K-S | | | |
|----------------|-----------------------------------|--------------|------------|-------|------------|
| | | سطح معناداری | P محاسباتی | LB | درجه آزادی |
| کلر | ARIMA(2,0,1)(1,1,2) ₁₂ | ۵ درصد | ۰/۰۶۳ | ۳۴/۴۲ | ۳۰ |
| هدایت الکتریکی | ARIMA(2,0,1)(1,1,2) ₁₂ | ۵ درصد | ۰/۰۷۶ | ۳۸/۸۶ | ۳۰ |



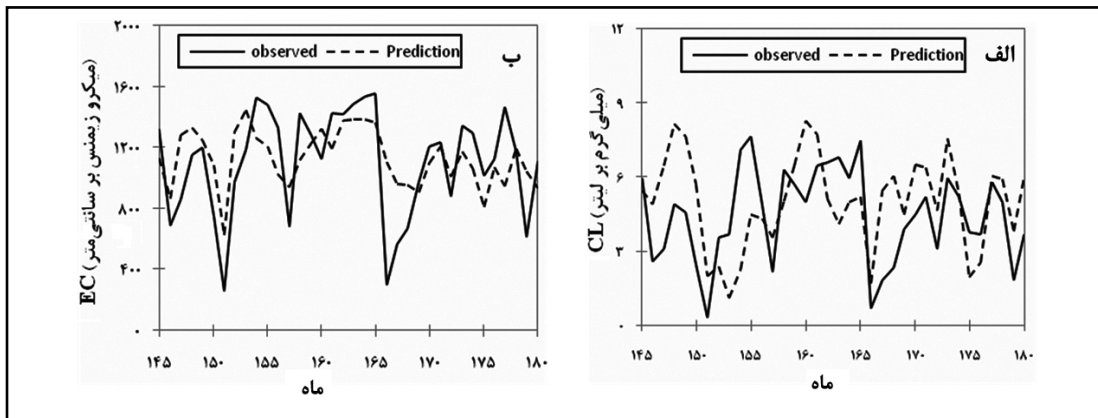
شکل ۵- کنترل تشخیصی پارامتر کلر

پیش‌بینی پارامترها

نتایج آزمون‌های Z و F به ترتیب برای مقایسه میانگین و واریانس سری زمانی پیش‌بینی شده هدایت الکتریکی و کلر، و داده‌های مشاهداتی برای ۳ سال ۱۳۸۲ تا ۱۳۸۴ در جدول (۵) ارائه شده است. مقادیر $RMSE$ نیز برای این دو پارامتر در این جدول موجود است. Z محاسباتی بین مقادیر بحرانی Z جدول $(1/96) \pm$ برای آزمون دوطرفه در سطح ۵ درصد قرار دارد. همچنین مقادیر F محاسباتی داده‌های مشاهداتی و پیش‌بینی شده کوچک‌تر از F بحرانی جدول است. بنابراین می‌توان نتیجه گرفت که از لحاظ آماری اختلاف معنی‌داری بین داده‌های

مشاهداتی و پیش‌بینی شده وجود ندارد. با توجه به اینکه مقادیر پارامتر هدایت الکتریکی از لحاظ مقدار بزرگ هستند، مقدار $RMSE$ برای این پارامتر بزرگ‌تر از پارامتر کلر است. شکل ۶ نمودار داده‌های مشاهداتی و پیش‌بینی شده توسط بهترین مدل‌های $SARIMA$ را نشان می‌دهد.

همانطور که از شکل مشاهده می‌شود، میان رفتار سری زمانی پیش‌بینی شده هدایت الکتریکی و کلر و داده‌های مشاهداتی تطابق قابل قبولی وجود دارد. مقادیر پیش‌بینی شده و مشاهداتی دارای روند افزایش و کاهش همزمان هستند.



شکل ۶- مقایسه نمودار داده‌های پیش‌بینی و مشاهده شده هدایت الکتریکی و کلر

جدول ۵- مقایسه آماری مقادیر مشاهده و پیش‌بینی شده هدایت الکتریکی و کلر

| پارامتر | میانگین مشاهده‌ای | میانگین پیش‌بینی | Z محاسباتی | واریانس مشاهده‌ای | واریانس پیش‌بینی | F محاسباتی | F بحرانی | RMSE |
|--|-------------------|------------------|------------|-------------------|------------------|------------|----------|-------|
| کلر (میلی‌گرم بر لیتر) | ۴/۴۴ | ۴/۸۳ | -۰/۷۴۲ | ۳/۵۸ | ۶/۴۴ | ۰/۵۵۷ | ۱/۹۶۱ | ۲۷۸/۹ |
| هدایت الکتریکی (میکروزیمنس بر سانتی‌متر) | ۱۱۰۰/۳ | ۰/۶۰۱ | ۰/۶۰۱ | ۲۳۸۷۰/۲ | ۲۲۰۸۶/۴ | ۱/۰۸۱ | ۱/۹۶۱ | ۲/۲۲ |

بور در ۵ ایستگاه هیدرومتری رودخانه بویوک مندرس مناسب هستند.

نتیجه‌گیری

با توجه به اهمیت سفیدرود در استان گیلان و تأثیر زیاد پارامترهای هدایت الکتریکی و کلر بر کیفیت آب این رودخانه، استفاده از یک روش مناسب به منظور پیش‌بینی مقادیر این پارامترها، بهره‌برداری بهینه از این رودخانه را ممکن می‌سازد. هدف از این تحقیق، توسعه مدل‌های SARIMA (Box et al., 1994) به منظور پیش‌بینی سری‌های زمانی پارامترهای هدایت الکتریکی و غلظت کلر ایستگاه آستانه سفیدرود است. داده‌های سری زمانی پارامترهای هدایت الکتریکی و غلظت کلر دارای تناوب فصلی ۱۲ ماهه و فاقد مؤلفه روند هستند. وجود تناوب فصلی با توجه به تغییرات دبی و دمای آب در طول سال توجیه پذیر است. مدل نهایی انتخاب شده برای سری زمانی داده‌های هدایت الکتریکی و کلر به ترتیب

با توجه به نتایج به دست آمده از این تحقیق می‌توان شبیه‌سازی زمانی پارامترهای هدایت الکتریکی و کلر را در ایستگاه آستانه از رودخانه سفیدرود به روش SARIMA با اطمینان توصیه کرد. نتایج همسو و مشابه نیز بر اساس روش به کار برده شده در این تحقیق برای مطالعات موردی دیگر به دست آمده است که از مهمترین آنها می‌توان به این موارد اشاره کرد: احمد و همکاران (Ahmad et al., 2001) پیش‌بینی پارامترهای کیفیت آب رودخانه را با استفاده از مدل SARIMA توصیه کردند. نتایج تحقیق کورونچ و همکاران (Kurunç et al., 2005) نشان می‌دهد که مدل ARIMA برای پیش‌بینی پارامترهای هدایت الکتریکی، pH، و کلر در رودخانه سیرلماک ترکیه بسیار مناسب است. راگاو و فرناندز (Ragavan & Fernandez, 2006) کارایی مدل‌های SARIMA را در پیش‌بینی دراز مدت روند کیفیت آب رودخانه نشان دادند. همچنین دوردو (Durdu, 2009) نشان داد مدل‌های ARIMA و SARIMA برای پیش‌بینی ۳۶ ماهه غلظت

پیش‌بینی سریع داده‌ها وجود دارد، از این مدل‌ها استفاده کرد.

قدردانی

از دانشگاه تهران و شرکت مدیریت منابع آب به دلیل تأمین امکانات و داده‌های لازم جهت انجام این تحقیق و تهیه مقالات مربوطه قدردانی می‌شود.

ARIMA(2,0,1)(1,1,2)₁₂ و ARIMA(2,0,1)(1,1,1)₁₂ هستند. باقیمانده^۲ این مدل‌ها تصادفی، واریانس همسان و دارای توزیع نرمال هستند. پیش‌بینی‌های این مدل‌ها از لحاظ آماری با مقادیر ثبت شده اختلاف معنی‌داری ندارند. با توجه به اینکه تعیین پارامترهای مدل‌های تصادفی خطی زمان‌بر نیست می‌توان در مواردی که نیاز به

مراجع

- Ahmad, S., Khan, I.H. and Parida, B.P. 2001. Performance of stochastic approaches for forecasting river water quality. *Water Res.* 35, 4261–4266.
- Akaike, H. 1974. A new look at statistical model identification. *IEEE Transactions on Automated Circuits.* 19, 716–723.
- Benyahya, L., Hilaire, A.S., Quarda, B.M.J.T., Bobee, B. and Nedushan, B.A. 2007. Modeling of water temperatures based on stochastic approaches: Case study of the Deschutes river. *J. Environ. Eng. Sci.* 6, 437–448.
- Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control.* San Francisco. CA: Holden Day, 575 p.
- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control.* 3rd Ed. Prentice Hall, Englewood Cliffs Inc., New Jersey. 598 p.
- Durdu, O. 2009. Stochastic approaches for time series forecasting of boron: A case study of Western Turkey. *Environ. Monit. Assess.* 687–701.
- Hassanzadeh, S., Hosseinibalam, F. and Alizadeh, R. 2008. Statistical models and time series forecasting of sulfur dioxide: A case study Tehran. *Environ. Monit. Assess.* 155, 149–155.
- Karamouz, M. and Araghinejad, Sh. 2005. *Advanced Hydrology.* Amirkabir University of Technology. Ch. 5, 256–257. (in Farsi)
- Karbasi, A. and Shahbazi, A. 2008. Analysing water quality of rivers in Gilan province. *Advanced Environ Res. J.* 5, 2–3. (in Farsi)
- Kurunç, A., Yürekli, K. and Çevik, O. 2005. Performance of two stochastic approaches for forecasting water quality and streamflow data from Yesilirmak river, Turkey. *Environ Modell Softw.* 20, 1195–1200.
- Ljung, G.M. and Box, G.E.P. 1978. On a measure of a lack of fit in time series models. *Biometrika.* 65 (2): 297–303.
- Makridakis, S., Wheelwright, S.C. and Hyndman, R. 2003. *Forecasting Methods and Applications.* Singapore: Wiley (ASIA).
- Malmir, M. 2006. Prediction of low flow time series in rivers. M. Sc. Thesis. Faculty of Agriculture. Theran University. Karaj. Iran. (in Farsi)
- Mishra, A.K. and Desai, V.R. 2005. Drought forecasting using stochastic models. *Stoch. Environ. Res. Risk Assess.* 19, 326–339.
- Mishra, A.K. and Desai, V.R. 2006. Drought forecasting using feed-forward recursive neural network. *Ecol. Model.* 198, 127–138.
- Modarres, R. 2007. Streamflow drought time series forecasting. *Stoch. Environ. Res. Risk Assess.* 21, 223–233.
- Pokorny, M. 1987. *An Introduction to Econometrics.* Oxford: Basil Blackwell Ltd., New York.
- Ragavan, A.J. and Fernandez, G.C. 2006. Modeling water quality trend in long term time series. *Proceedings of SAS Users Group Meeting (SUGI 31).* March 2006, San Francisco, USA. paper 205–31.

Time Series Forecasting of Sefidrood River Water Quality Using Linear Stochastic Models

B. Jamali and K. Ebrahimi*

* Corresponding Author: Assistant Professor of Water Resources Engineering, Department of Irrigation and Reclamation Engineering, University of Tehran, IRAN. Email: EbrahimiK@ut.ac.ir

Received: 1 January 2011, Accepted: 3 July 2011

In the management of river basins, prediction of river water quality is essential to maintain water quality within standard limits. This study performed a time-series analysis of the prediction of chlorine concentration and electrical conductivity time series data for the period of 1991-2005 from Sefidrood River in northern Iran. The seasonal prediction of chlorine and electrical conductivity time series data was done using the linear stochastic model known as the multiplicative seasonal autoregressive integrated moving average (SARIMA). Initially, Mann-Kendall and Box-Pierce tests were used to identify the trend and stationary state of the time series, respectively. The results showed that there was no significant trend in these time series, but that 12-month seasonal patterns were observed. As a result, seasonal patterns were removed from both time series data using first-order differencing. SARIMA modeling was performed in three steps: model identification, parameter estimation and diagnostic checking. Different models of SARIMA were identified according to the ACF and PACF time series results and the model with the minimum AIC criterion was selected. For parameter estimation, model parameters were estimated using a least squares optimization algorithm that minimized the residual sum of squares. The results of diagnostic checking then indicated that the residuals were independent, normally distributed and homoscedastic. The selected SARIMA model was then used to predict chlorine concentration and electrical conductivity time series data for 2003-2005. There was a good unanimity between the predicted and observed data. For model verification, the mean and variance of the predicted and observed data were compared. The RMSE for Cl and EC were 2.2 and 278.9, respectively. The results showed that there was no significant difference between the predicted and observed time series data. This study showed that the SARIMA model can be used reliably to predict chlorine concentration and electrical conductivity time series data in Sefidrood River.

Keywords: Electrical conductivity, Mann-Kendall, SARIMA, time series