



دانشگاه گوارز، ریزش آب و خاک

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و چهارم، شماره سوم، ۱۳۹۶
<http://jwsc.gau.ac.ir>

تأثیر تفاضل‌گیری در ایستایی و دقت مدل‌های سری زمانی در پیش‌بینی تراز سطح دریاچه

*مهسا خادمی^۱، حمید معینی^۱، حسین بنکداری^۲ و عیسی ابتهج^۳

^۱دانش‌آموخته کارشناسی‌ارشد گروه مهندسی عمران، دانشگاه رازی کرمانشاه،

^۲استاد گروه مهندسی عمران، دانشگاه رازی کرمانشاه، ^۳دانشجوی دکتری گروه مهندسی عمران، دانشگاه رازی کرمانشاه

تاریخ دریافت: ۹۵/۴/۳؛ تاریخ پذیرش: ۹۶/۶/۳۰

چکیده

سابقه و هدف: یکی از فرض‌های بسیار مهم در مدل‌سازی سری‌های زمانی ایستا بودن آن است. میزان ایستایی می‌تواند متفاوت باشد به طوری که در منابع مختلف ایستایی مرتبه اول، مرتبه دوم، قوی و اکید تعریف شده است. بنابراین در این پژوهش به بررسی تأثیر تفاضل‌گیری‌های فصلی، غیرفصلی و توأم بر میزان ایستایی سری زمانی پرداخته شد. همچنین تأثیر میزان ایستایی بر عملکرد مدل‌های ARMA، ARIMA و SARIMA در مدل‌سازی و پیش‌بینی سری زمانی تراز ماهانه سطح دریاچه از جنبه‌های مختلف بررسی گردید.

مواد و روش‌ها: بدین‌منظور از ۹۶ داده ماهانه اندازه‌گیری‌شده از دریاچه میسنگان-هاژن واقع در مرز کشورهای آمریکا و کانادا استفاده شد. ۷۶ سال ابتدایی این داده‌ها برای دوره واسنجی و ۲۰ سال انتهایی برای اعتبارسنجی در نظر گرفته شد. ابتدا به کمک آزمون‌های من-کندال فصلی و فیشر وجود اجزاء روند و دوره در سری بررسی شد. این دو جز اصلی‌ترین عوامل ناپایستاکنده سری زمانی هستند. سپس از تفاضل‌گیری‌های فصلی، غیرفصلی و هر دو استفاده شد و نتایج با داده‌های بدون تفاضل‌گیری مقایسه شد. به‌منظور بررسی میزان ایستایی سری‌های به‌دست آمده نیز از نمودار ACF و آزمون دیکی-فولر تعمیم‌یافته استفاده شد. نوع و تعداد پارامترهای مورد نیاز در مدل‌ها نیز با استفاده از نمودار ACF برای هر کدام از این حالات تعیین گردید. سپس هر کدام از سری‌ها با استفاده از مدل مناسب خود، مدل‌سازی و پیش‌بینی شدند.

یافته‌ها: بررسی‌ها نشان داد که هیچ‌گونه روند و تناوبی در داده‌ها وجود ندارد و سری زمانی ایستا است. با این حال استفاده از تفاضل‌گیری‌های فصلی و توأم میزان ایستایی را بیش‌تر می‌کند. اما تفاضل‌گیری غیرفصلی سری را ناپایستا می‌کند. استفاده هم‌زمان از تفاضل‌گیری فصلی و غیرفصلی دارای بیش‌ترین تأثیر در میزان ایستا شدن تراز سطح دریاچه است. مطابق با نمودار ACF، استفاده از تفاضل‌گیری توأم باعث می‌شود که به استفاده از پارامترهای فصلی در مدل احتیاج پیدا شود. در صورتی که در دیگر حالت‌ها این‌گونه نیست. بنابراین سری بدون تفاضل‌گیری با مدل ARMA، سری تفاضل‌گیری فصلی شده با مدل ARIMA و سری تفاضل‌گیری توأم شده با مدل SARIMA مدل‌سازی گردید. نتایج نشان داد که هنگام استفاده از تفاضل‌گیری توأم، تعداد مدل‌های موردنیاز برای دستیابی به

* مسئول مکاتبه: mahsa.khademi88@yahoo.com

دقیق‌ترین پیش‌بینی به اندازه بسیار زیادی کاهش می‌یابد. به طوری که بدون تفاضل‌گیری به مدل ARMA نیاز بود که این میزان هنگام استفاده از تفاضل‌گیری‌های فصلی و غیرفصلی به ۶۴ مدل SARIMA کاهش یافت. از طرف دیگر با استفاده از تعداد پارامترهای بسیار کم‌تر (۲ پارامتر) در مدل SARIMA نتیجه‌ای مشابه و حتی بهتر از مدل ARMA با تعداد ۲۱ پارامتر به دست آمد.

نتیجه‌گیری: نتایج نشان داد که ایستاسازی هرچه بیشتر تر از ماهانه دریاچه که به خودی خود ایستاست، تعداد مدل‌ها و تعداد پارامترهای موردنیاز مدل‌ها را برای دستیابی به بهترین نتیجه به اندازه زیادی کاهش می‌دهد. بدین‌منظور تفاضل‌گیری توأم بیشتر از سایر روش‌ها سری موردنظر را ایستا نمود.

واژه‌های کلیدی: پیش‌بینی، سری زمانی، سطح دریاچه، SARIMA، ARIMA

مقدمه

تأمین آب موردنیاز کشاورزی، صنعت و آشامیدنی نیازمند برنامه‌ریزی‌های دقیق و درازمدت است. با توجه به مشکلاتی مانند خشکسالی‌ها، کمبود بارندگی و مصرف بی‌رویه آب‌های زیرزمینی در سال‌های اخیر، اهمیت این موضوع بیش‌تر شده است. از این‌رو پیش‌بینی متغیرهای هیدرولوژیکی به‌منظور مدیریت بهره‌برداری از منابع آب امری ضروری است که در این بین تراز سطح دریاچه‌ها به‌عنوان یک میراث طبیعی اهمیت ویژه‌ای دارد. بدین‌منظور استفاده از روش‌هایی که با صرف کم‌ترین وقت و هزینه به دقیق‌ترین نتیجه برسد بسیار دارای اهمیت است.

از گذشته تا به امروز پژوهشگران زیادی به مطالعه سری‌های زمانی هیدرولوژیکی در زمینه مدل‌سازی و پیش‌بینی پرداخته‌اند (۴، ۷، ۱۲، ۱۳، ۲۵، ۲۶، ۲۷، ۳۲، ۳۳، ۳۴ و ۳۵). مطالعه تراز سطح دریاچه‌ها نیز به‌طور خاص مورد استقبال بسیاری قرار گرفته است (۵، ۱۵ و ۲۴). برای مثال کیسی و همکاران (۲۰۱۵) به مطالعه سطح آب دریاچه ارومیه با استفاده از چند روش مختلف پرداخت و در آخر نتایج مدل‌ها را به‌منظور انتخاب بهترین مدل با یکدیگر مقایسه کرد (۱۴).

به‌منظور مدل‌سازی و پیش‌بینی سری‌های زمانی مدل‌های مختلفی وجود دارد. از جمله آن‌ها می‌توان

مدل‌های خود همبسته با میانگین متحرک (ARMA)، خود همبسته با میانگین متحرک تجمعی (ARIMA) و خود همبسته با میانگین متحرک تجمعی فصلی (SARIMA) را نام برد. توسط این مدل‌ها بسیاری از مطالعات هیدرولوژیکی از جمله پیش‌بینی دبی جریان (۱۸ و ۳۰)، سطح آب زیرزمینی (۲۰)، میزان بارش (۲۸)، دمای هوا (۲)، تبخیر و تعرق (۱)، خشکسالی (۱۷) و پارامترهای اقلیمی (۱۱) صورت گرفته است. در این بین مطالعه سطح دریاچه نیز مورد توجه بوده است. خطیبی و همکاران (۲۰۱۴) به پیش‌بینی تراز سطح شش دریاچه با استفاده از مدل‌های مختلف از جمله مدل SARIMA پرداخت و نتایج را با یکدیگر مقایسه کرد (۱۰). شفاپی و کیسی (۲۰۱۵) نوسانات سطح آب دریاچه را به کمک تبدیل موجک تجزیه و سپس با مدل SARIMA مدل‌سازی کردند. مقایسه نتایج با نتایج مدل SARIMA بدون استفاده از تبدیل، بیانگر بهبود دقت این مدل بود (۲۳).

یکی از مهم‌ترین فرضیات مدل‌سازی با استفاده از روش‌های مذکور، ایستا بودن سری زمانی است (۳). اگر یک سری زمانی نایستا باشد روش‌هایی وجود دارد که می‌توان آن‌ها را ایستا نمود. از جمله آن‌ها می‌توان شناسایی و حذف مستقیم عوامل نایستایی (۱۶)، استفاده از توابع تبدیل و تفاضل‌گیری را نام

برد. یکی از کاربردهای تفاضل‌گیری حذف یا کاهش اثر عوامل ناپایستایی مثل مؤلفه‌های روند و تناوب است. این مؤلفه‌ها را می‌توان به ترتیب با تفاضل‌گیری‌های غیرفصلی و فصلی از بین برد (۲۲). اما سوال اینجاست که آیا در صورتی که این مؤلفه‌ها در سری وجود نداشته باشند باز هم تفاضل‌گیری‌های فصلی و غیرفصلی می‌توانند باعث ایستا شدن سری شوند؟ و این‌که اگر سری زمانی خود ایستا باشد، آیا این عمل می‌تواند ایستایی سری را قوی‌تر کند؟ قوی‌تر شدن ایستایی سری در نتایج پیش‌بینی چه تأثیری دارد؟

در این پژوهش انواع تفاضل‌گیری‌های فصلی، غیرفصلی و توأم (هم فصلی و هم غیرفصلی) بر روی داده‌های مربوط به تراز ماهانه دریاچه می‌شیکان-هاژن به طول ۹۶ سال انجام می‌شود. سپس عملکرد آن‌ها در ایستاسازی سری بررسی می‌گردد. همچنین تأثیر این عمل بر تعداد مدل‌های موردنیاز و پارامترهای موردنیاز برای آن‌ها سنجیده خواهد شد. در آخر دقت پیش‌بینی توسط مدل‌ها با یکدیگر مقایسه می‌شود. در این بین روشی برای انتخاب دقیق‌ترین مدل نیز ارائه می‌گردد.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه: دریاچه‌های می‌شیکان، هاژن، سوپریور، اُنتاریو و اِریه به نام دریاچه‌های بزرگ شناخته می‌شوند (شکل ۱). در حدود ۵۹ درصد از این حوضه آبریز در ایالات متحده و ۴۱ درصد در کشور کانادا قرار دارد. جمعیت آن در حدود ۳۳ میلیون نفر است که گردشگری، کشاورزی و تولید محصولات صنعتی از جمله فعالیت‌های تجاری آنان است (۳۱). طول حوضه از شمال تا جنوب در حدود ۱۱۲۵ کیلومتر و از غرب به شرق ۱۴۵۰ کیلومتر است و مساحتی در حدود ۴۰۲

کیلومترمربع را پوشش می‌دهد. این مجموعه یکی از بزرگ‌ترین منابع آب شیرین ایالات متحده است. در این حوضه دریاچه می‌شیکان به وسیله یک کانال عمیق به دریاچه هاژن متصل شده است. بنابراین تمام شرایط هیدرولیکی آن‌ها (از جمله تراز سطح) مشابه است. به همین دلیل این دو به‌عنوان یک دریاچه به‌نام می‌شیکان-هاژن شناخته شده‌اند. این دریاچه در موقعیت جغرافیایی ۸۱ تا ۸۹ درجه غربی و ۴۱ تا ۴۶ درجه شمالی واقع شده است. داده‌های مورد استفاده مربوط به تراز ماهانه سطح همین دریاچه از سال ۱۹۱۸ تا ۲۰۱۳ می‌باشد که در ایستگاه هیدرومتری ساحل هاربور اندازه‌گیری شده‌اند (شکل ۲) که ۷۶ سال ابتدایی آن برای دوره واسنجی و ۲۰ سال انتهایی آن برای دوره اعتبارسنجی (دوره پیش‌بینی) در نظر گرفته شد. با توجه به طولانی بودن سری مورد مطالعه، نتایج این مطالعه از اطمینان بالاتری برخوردار خواهد بود. از آن‌جا که در ایران برای هیچ دریاچه‌ای داده‌هایی چنین طولانی وجود ندارد، می‌توان نتایج این پژوهش را برای مناطق مشابه در داخل ایران به‌کار برد.

تعریف ایستایی سری‌های زمانی: ایستایی به‌طور فراوان تحت عنوان ایستایی ضعیف، ایستایی کوواریانس و ایستایی مرتبه دوم تعریف شده است. در هر حال بدون در نظر گرفتن شرایط بیش‌تر اگر در یک سری میانگین ثابت، واریانس ثابت و کوواریانس در دوره‌های مشخص نیز ثابت باشد (رابطه‌های ۱ تا ۴) به آن سری ایستا گفته می‌شود (۳). همچنین اگر برای هر انتخاب نقاط زمانی t_1 تا t_k و برای هر تأخیر زمانی h ، توزیع توأم $x(t_1)$ تا $x(t_n)$ مانند توزیع توأم $x(t_1+h)$ تا $x(t_n+h)$ باشد، سری زمانی ایستای اکید خواهد بود (۱۹).

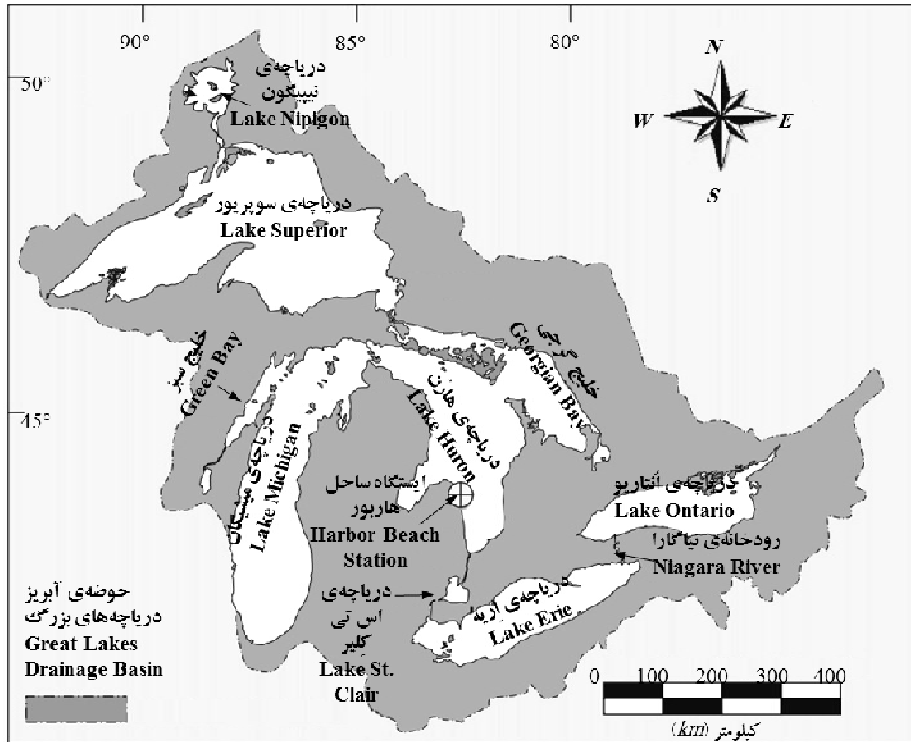
که در آن، $x(t)$ سری زمانی، E و Var توابع امید ریاضی و واریانس، γ_x تابع اتوکواریانس سری زمانی، m_1 و m_2 اعداد ثابت و $Z = \{\pm 1, \pm 2, \pm 3, \dots\}$ اعداد صحیح به صورت هستند.

$$E(|x(t)|^2) < \infty, t \in Z \quad (۱)$$

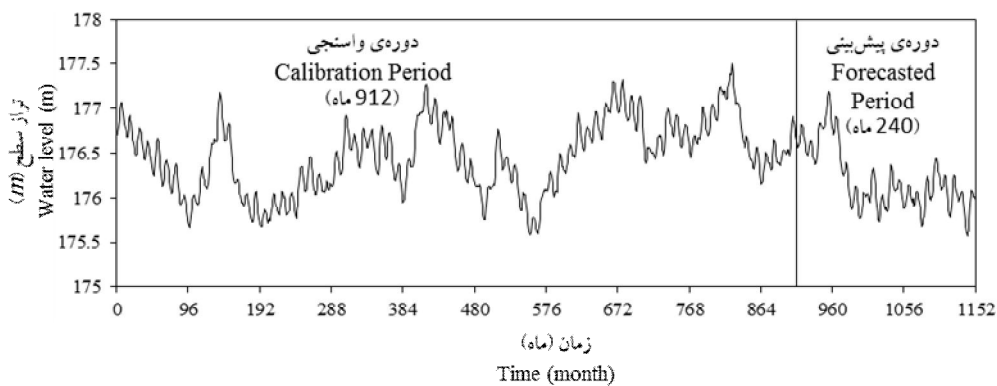
$$E(x(t)) = m_1, t \in Z \quad (۲)$$

$$Var(x(t)) = m_2, t \in Z \quad (۳)$$

$$\gamma_x(r, s) = \gamma_x(r+t, s+t), r, s, t \in Z \quad (۴)$$



شکل ۱- حوضه آبریز دریاچه‌های بزرگ.
Figure 1. Great Lakes watershed.



شکل ۲- سری زمانی تراز ماهانه دریاچه میشیگان- هارن.
Figure 2. Lake Michigan-Huron monthly time series level.

$$\text{Var}(M_K) = \sum_{k=1}^{\omega} \frac{n_k(n_k-1)(2n_k+5)}{18} + 2 \sum_{i=1}^{\omega-1} \sum_{j=i+1}^{\omega} \sigma_{ij} \quad (8)$$

که در آن، M_K آماره آزمون من- کندال فصلی، S_k آماره من- کندال برای مقادیر سری در فصل k ام، sgn تابع علامت، ω تعداد فصل‌های موجود در هر سال، n_k برابر n امین داده از فصل k ام و σ_{ij} کوواریانس آماره آزمون در فصول i و j است.

با تعریف $u_{M_K} = M_K \text{var}(M_K)^{-0.5}$ در صورتی که شرط زیر برقرار باشد، در سطح معنی‌داری α مؤلفه روند در سری وجود نخواهد داشت:

$$u_{\frac{\alpha}{2}} < u_{M_K} < u_{1-\frac{\alpha}{2}} \quad (9)$$

که در آن، $u_{\frac{\alpha}{2}}$ و $u_{1-\frac{\alpha}{2}}$ چندانک $\frac{\alpha}{2}$ و $1-\frac{\alpha}{2}$ (که با P_{\min} و P_{\max} نشان داده می‌شوند) از توزیع نرمال هستند. مقدار α نیز، $0/05$ در نظر گرفته شد.

به منظور تعیین مؤلفه تناوبی در سری زمانی تراز ماهانه دریاچه نیز از آزمون فیشر به صورت زیر استفاده شد (9):

$$F^* = \frac{N(N-2)(\alpha_k^2 + \beta_k^2)}{4 \left(\sum_{k=1}^h (x(t) - \alpha_k \cos(\omega_k t) - \beta_k \sin(\omega_k t)) \right)} \quad (10)$$

$$F^* \geq F(2, N-2) \quad (11)$$

که در آن، F^* آماره آزمون فیشر، α_k و β_k ضرایب بسط سری فوریه، ω_k فرکانس زاویه‌ای، N تعداد داده‌ها و F مقدار بحرانی در توزیع F در سطح معنی‌داری در نظر گرفته شده است که بر اساس جدول این توزیع، در سطح معنی‌داری 5% و برای تعداد داده‌های بیش‌تر از 1000 این مقدار برابر 3 می‌باشد.

شناسایی عوامل نایستایی: عوامل مختلفی در سری زمانی می‌توانند وجود داشته باشند که منجر به نایستایی شدن سری می‌گردند. از جمله مهم‌ترین این عوامل که در سری‌های هیدرولوژیکی بیش‌ترین تأثیر را دارند، مؤلفه‌های روند $(T(t))$ و تناوب $(P(t))$ می‌باشند (16). روند به افزایش یا کاهش در خصوصیات آماری سری مثل میانگین و انحراف معیار نسبت به زمان گفته می‌شود. تناوب نیز به تغییرات دوره‌ای همین خصوصیات با یک دوره تناوب مشخص گفته می‌شود. برای بازه‌های درون‌سالی مثل سری‌های ماهانه و روزانه، تناوبی بودن سری‌های زمانی هیدرولوژیکی معمولاً قابل توجه خواهد بود. روند و تناوب اجزاء قطعی سری‌های زمانی هستند که در صورت شناسایی آن‌ها، آنچه باقی خواهد ماند مؤلفه تصادفی $(S(t))$ خواهد بود. بنابراین سری‌های زمانی هیدرولوژیکی از جمله تراز سطح دریاچه $(x(t))$ معمولاً به صورت زیر هستند:

$$x(t) = T(t) + P(t) + S(t) \quad (5)$$

آگاهی از وجود یا عدم وجود عوامل نایستایی مثل مؤلفه‌های روند و تناوبی کمک زیادی در تعیین نوع تفاضل‌گیری به منظور ایستاسازی سری می‌کند. برای مثال تأثیر این دو مؤلفه بر نایستایی شدن سری را می‌توان به ترتیب با تفاضل‌گیری غیرفصلی و تفاضل‌گیری فصلی با گام ω (که برابر با دوره تناوب است) از بین برد. برای تعیین عوامل نایستایی آزمون‌های مختلفی وجود دارد. در اینجا از آزمون من- کندال فصلی برای تعیین روند در داده‌ها استفاده گردید (8). این آزمون به صورت زیر است:

$$M_K = \sum_{k=1}^{\omega} (S_k - \text{sgn}(S_k)) \quad (6)$$

$$S_k = \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k-1} \text{sgn}(x(ki) - x(kj)) \quad (7)$$

که در آن، Δ عملگر تفاضلی مرتبه اول، α ، β_1 و β_2 مقدار ثابت و ضریب روند خطی و درجه دو، ϕ_1 ضریب آمین مرتبه خودهمبستگی، p حداکثر مرتبه خودهمبستگی در نظر گرفته شده، τ آماره آزمون دیکی- فولر تعمیم یافته، $\hat{\gamma}$ ضریب (ریشه) تخمین زده شده، $\sigma_{\hat{\gamma}}$ خطای استاندارد ضریب تخمین زده شده، H_0 فرض صفر مبنی بر وجود ریشه واحد و H_1 فرض جایگزین مبنی بر عدم وجود ریشه واحد است.

با محاسبه مقدار τ در صورتی که این آماره از مقدار بحرانی در توزیع دیکی- فولر کوچک تر باشد، فرض صفر رد خواهد شد. در نتیجه ریشه واحدی وجود ندارد و سری ایستا خواهد بود. هر کدام از مقادیر α ، β_1 و β_2 را می توان صفر در نظر گرفت. در این صورت حالات مختلفی از جمله بدون ثابت و روند، فقط ثابت، ثابت و روند خطی، ثابت و روند خطی و روند درجه دو مورد آزمون قرار خواهد گرفت. ضعف این آزمون در این است که اثر مؤلفه تناوبی بر نایست شدن سری را در نظر نمی گیرد. بنابراین برای سری های فصلی استفاده از دیگر آزمون ها در کنار آن ضروری خواهد بود.

با رسم نمودار همبستگی نگار نیز می توان به صورت شهودی ایستا بودن سری زمانی را تعیین کرد. این نمودار مقادیر تابع خودهمبستگی (ACF) را به ازای گام های زمانی مختلف نشان می دهد. در صورتی که سری زمانی ایستا باشد، این نمودار بلافاصله به صورت نمایی یا نوسانی به صفر میرا خواهد شد (۳). برای تشخیص این موضوع تعداد گام های لازم برابر یک چهارم کل داده ها می باشد. در این نمودار مقدار ضریب خودهمبستگی داده ها (ρ) برای گام های مختلف رسم خواهد شد. مقدار این ضریب برای گام h به صورت زیر به دست می آید:

برای تعیین دوره تناوب داده ها (ω) به منظور تفاضل گیری فصلی، تحلیل طیفی داده ها با رسم نمودار طیفی صورت گرفت. این نمودار چگالی طیفی واریانس (I_k) نسبت به فرکانس های مختلف (ω_k) را نشان می دهد (رابطه های ۱۲ و ۱۳). به این ترتیب ω در تفاضل گیری فصلی برابر گامی از این نمودار خواهد بود که دارای نقطه اوج معنی دار باشد.

$$\omega_k = \frac{2\pi k}{N}, \quad k = 1, 2, \dots, h \quad (12)$$

$$I_k = \frac{N}{2} (\alpha_k^2 + \beta_k^2), \quad k = 1, 2, \dots, h \quad (13)$$

آزمون های ایستایی: ایستا بودن سری های زمانی یک فرض بسیار مهم در مدل سازی و پیش بینی آن ها است. اطمینان از این که آیا بعد از تفاضل گیری سری ایستا شده باشد را می توان توسط روش های مختلف بررسی کرد. بدین منظور از آزمون ایستایی دیکی- فولر تعمیم یافته (ADF) استفاده شد (۲۱). این آزمون، ایستایی را به طور مستقیم تعیین نمی کند بلکه براساس وجود یا عدم وجود ریشه واحد عمل می کند. عدم وجود ریشه واحد بیانگر ایستا بودن سری خواهد بود. همچنین این آزمون روند خطی و درجه دو را نیز در خود جای می دهد. بدین منظور مدل زیر در نظر گرفته می شود:

$$\Delta x(t) = \alpha + \beta_1 t + \beta_2 t^2 + \gamma x(t-1) + \phi_1 \Delta x(t-1) + \dots + \phi_{p-1} \Delta x(t-p+1) + \varepsilon(t) \quad (14)$$

$$\tau = \frac{\hat{\gamma}}{\sigma_{\hat{\gamma}}} \quad (15)$$

$$H_0: \gamma = 0 \quad (16)$$

$$H_1: \gamma < 0 \quad (17)$$

θ پارامتر میانگین متحرک غیرفصلی، Φ پارامتر خودهمبسته فصلی، Θ پارامتر میانگین متحرک فصلی، ω دوره تناوب، B عملگر تفاضلی به صورت $B(x_t) = x_{t-1}$ ، $(1-B)^D$ برابر D امین تفاضل گیری فصلی با گام ω و $(1-B)^d$ برابر d امین تفاضل گیری غیرفصلی می باشد.

حداکثر مرتبه پارامترهای مدل های مذکور $(p, q, P$ و $Q)$ را می توان از روی نمودار ACF شناسایی کرد. به این ترتیب که حداکثر مرتبه مورد نیاز برای پارامترهای خودهمبسته و میانگین متحرک غیرفصلی $(p$ و $q)$ برابر تعداد گام هایی از نمودار ACF سری ایستا شده است که دارای مقادیر بزرگی هستند. حداکثر مرتبه مورد نیاز برای پارامترهای فصلی $(P$ و $Q)$ نیز برابر تعداد گام های دارای مقادیر بزرگ از همین نمودار است که به صورت فصلی به ازای هر دوره تناوب رسم شده باشد (۶). به این ترتیب با ضرب تعداد مرتبه های مورد نیاز پارامترها در یکدیگر می توان دریافت که چه تعداد مدل وجود دارد که دقیق ترین مدل در بین آنها است. اگر مرتبه های مورد نیاز برای پارامترهای هر مدل به صورت $P = \{0, 1, \dots, n\}$ ، $q = \{0, 1, \dots, m\}$ ، $p = \{0, 1, \dots, l\}$ و $Q = \{0, 1, \dots, o\}$ باشند، آن گاه تعداد مدل های مورد نیاز (NOM) به صورت زیر به دست می آید:

$$NOM_{ARMA} = (l+1)(m+1) \quad (26)$$

$$NOM_{ARIMA} = (l+1)(m+1) \quad (27)$$

$$NOM_{SARIMA} = (l+1)(m+1)(n+1)(o+1) \quad (28)$$

انتخاب دقیق ترین مدل: رویکرد مرسوم در انتخاب مدل مناسب به این صورت است که با شناسایی مرتبه های مورد نیاز پارامترهای مدل، به تدریج تعداد پارامترهای مدل افزایش داده می شود. در همین حال،

$$\rho(h) = \frac{\sum_{t=h}^N (x(t) - \bar{x})(x(t-h) - \bar{x})}{\sum_{t=h}^N (x(t) - \bar{x})^2} \quad (18)$$

$$, h=1, 2, \dots, N/\xi$$

مدل های سری های زمانی: از بین مدل های سری های زمانی از مدل های ARMA، ARIMA و SARIMA به منظور پیش بینی تراز آب دریاچه استفاده شد. این مدل ها با روابط زیر بیان می شوند:

ARMA (p, q) :

$$\varphi(B)x(t) = \theta(B)\varepsilon(t) \quad (19)$$

ARIMA (p, d, q) $(l, D, 0)_\omega$:

$$\varphi(B)(1-B^\omega)^D(1-B)^d x(t) = \theta(B)\varepsilon(t) \quad (20)$$

SARIMA (p, d, q) $(P, D, Q)_\omega$:

$$\Phi(B^\omega)\phi(B)(1-B^\omega)^D(1-B)^d x(t) = \Theta(B^\omega)\vartheta(B)\varepsilon(t) \quad (21)$$

$$\Phi(B^\omega) = (1 - \phi_1 B^\omega - \phi_2 B^{2\omega} - \dots - \phi_p B^{p\omega}) \quad (22)$$

$$\varphi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \quad (23)$$

$$\Theta(B^\omega) = (1 - \theta_1 B^\omega - \theta_2 B^{2\omega} - \dots - \theta_q B^{q\omega}) \quad (24)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \quad (25)$$

که در آن، $x(t)$ سری تراز آب دریاچه، ε_t سری باقی مانده ها، p مرتبه پارامتر خودهمبسته غیرفصلی، q مرتبه پارامتر میانگین متحرک غیرفصلی، P مرتبه پارامتر خودهمبسته فصلی، Q مرتبه پارامتر میانگین متحرک فصلی، φ پارامتر خودهمبسته غیرفصلی،

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x(t) - \hat{x}(t))^2} \quad (۳۲)$$

$$CRM = \frac{\sum_{t=1}^N x(t) - \sum_{t=1}^N \hat{x}(t)}{\sum_{t=1}^N x(t)} \quad (۳۳)$$

که در آن، $x(t)$ و $\hat{x}(t)$ ترازهای اندازه‌گیری شده و پیش‌بینی شده دریاچه، \bar{x} و $\bar{\hat{x}}$ میانگین ترازهای اندازه‌گیری شده و پیش‌بینی شده دریاچه، t ترتیب ماه‌ها و N تعداد کل ماه‌ها در دوره اعتبارسنجی است.

آزمون استقلال باقی‌مانده‌ها: پس از انتخاب دقیق‌ترین مدل، مستقل بودن باقی‌مانده‌های آن نیز بررسی گردید. در این‌جا برای بررسی استقلال باقی‌مانده‌ها آزمون والد-ولفویتس به‌کار برده شد (۲۹). انجام این آزمون برای تمام وقایع هیدرولوژیکی که به‌صورت سری زمانی ثبت شده باشند امکان‌پذیر است. برای انجام این آزمون ابتدا به هر داده کوچک‌تر و بزرگ‌تر از میانگین به‌ترتیب مشخصه‌های a و b داده می‌شود. سپس آماره این آزمون به‌صورت محاسبه می‌گردد:

$$u_R = \frac{R - E_R}{S_d} \quad (۳۴)$$

$$E_R = \frac{n_a + n_b + 2n_a n_b}{n_a + n_b} \quad (۳۵)$$

$$S_d = \sqrt{\frac{2n_a n_b (2n_a n_b - n_a - n_b)}{(n_a + n_b)^2 (n_a + n_b - 1)}} \quad (۳۶)$$

که در آن، u_R آماره آزمون والد-ولفویتس، R تعداد دنباله‌های به‌دست آمده، E_R تعداد دنباله‌های قابل انتظار، S_d انحراف معیار دنباله‌ها، n_a تعداد داده‌های بزرگ‌تر از میانگین و n_b تعداد داده‌های کوچک‌تر از میانگین هستند.

عملکرد مدل با استفاده از یک معیار ارزیابی که تعداد پارامترهای آن را نیز در نظر می‌گیرد (مثل معیار اطلاعات آکائیکه (AIC) یا معیار اطلاعات بیزی (BIC)) مورد سنجش قرار می‌گیرد. افزایش تعداد پارامترها تا جایی ادامه پیدا می‌کند که مقدار معیار موردنظر بهبود یابد اما این روش لزوماً دقیق‌ترین مدل را ارائه نمی‌دهد. زیرا ممکن است به‌دلیل وجود حداقل نسبی در مقادیر معیارهای خطا، مدلی با تعداد پارامتر بیشتر وجود داشته باشد که دارای دقت بیشتری است. بنابراین در این پژوهش به‌منظور دستیابی به دقیق‌ترین مدل، تمام مدل‌های مورد نیاز به‌دست آمده از رابطه‌های ۲۶ تا ۲۸ بر روی سری برازش داده شد. سپس توسط تمام آن مدل‌ها مقدار تراز ماهانه سطح دریاچه در دوره اعتبارسنجی پیش‌بینی شد. سپس بر اساس نتایج چند معیار ارزیابی که فارق از تعداد پارامترهای مدل و تنها براساس دقت مدل عمل می‌کنند، بهترین مدل انتخاب گردید. در اینجا معیارهای ضریب تعیین (R^2)، دلالت واریانس (VAF)، جذر میانگین مربعات خطای نسبی (RMSRE)، جذر میانگین مربعات خطا (RMSE) و ضریب جرم باقی‌مانده‌ها (CRM) به‌کار برده شدند. بهترین مقدار برای R^2 و VAF برابر ۱۰۰ و برای دیگر معیارها برابر صفر می‌باشد. این معیارها به‌صورت زیر می‌باشند:

$$R^2 = \frac{\left(\sum_{t=1}^N (x(t) - \bar{x})(\hat{x}(t) - \bar{\hat{x}}) \right)^2}{\sum_{t=1}^N (x(t) - \bar{x})^2 \sum_{t=1}^N (\hat{x}(t) - \bar{\hat{x}})^2} \times 100 \quad (۲۹)$$

$$VAF = \left(1 - \frac{\text{var}(x(t) - \hat{x}(t))}{\text{var}(x(t))} \right) \times 100 \quad (۳۰)$$

$$RMSRE = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(\frac{x(t) - \hat{x}(t)}{x(t)} \right)^2} \times 100 \quad (۳۱)$$

نتایج و بحث

قبل از مدل‌سازی سری زمانی تراز ماهانه سطح دریاچه به‌منظور بررسی ایستایی، وجود عوامل نایستایی از جمله روند و تناوب در آن مورد آزمون قرار گرفتند. به‌منظور بررسی وجود روند از آزمون من-کندال فصلی استفاده شد. نتایج این آزمون در جدول ۱ ارائه شده است. مشاهده می‌شود که مقدار آماره آزمون بین حدود بحرانی نشان داده شده در سطح معنی‌داری ۰.۵٪ قرار دارد. همچنین مشاهده می‌شود که مقدار احتمال تجمعی این آماره از هر دو طرف توزیع نرمال بین مقادیر حداقل و حداکثر در همین سطح معنی‌داری قرار دارد. به این ترتیب در این سری هیچ‌گونه روند صعودی و روند نزولی وجود ندارد.

در صورت برقراری شرط زیر سری باقی‌مانده‌ها در سطح معنی‌داری α مستقل خواهد بود:

$$u_{\frac{\alpha}{2}} < u_R < u_{1-\frac{\alpha}{2}} \quad (37)$$

احتمال متناظر با u_R را با P_{u_R} نشان داده می‌شود که در صورت برقراری رابطه ۳۷ این مقدار بزرگ‌تر از مقدار α خواهد بود. چنانچه $u_R < u_{\frac{\alpha}{2}}$ داده‌ها ممکن است دارای مؤلفه روند و چنانچه $u_R > u_{1-\frac{\alpha}{2}}$ داده‌ها ممکن است دارای مؤلفه تناوبی باشند. در نتیجه سری باقی‌مانده‌ها تصادفی نخواهد بود.

جدول ۱- نتایج آزمون من-کندال فصلی برای تعیین روند.

Table 1. Seasonal Mann-Kendall test results to determine trends.

| احتمال بحرانی The probability of critical | احتمال آزمون Test probability | آماره بحرانی The critical test | آماره آزمون Test statistics |
|--|----------------------------------|-----------------------------------|--------------------------------|
| $P_{\max} = 97.5$ | $P_{u_{MK}} = 22.3$ | $u_{0.975} = 1.96$ | $u_{MK} = 0.76$ |
| $P_{\min} = 2.5$ | $P_{u_{MK}} = 77.7$ | $u_{0.05} = -1.96$ | $u_{MK} = 0.76$ |

و بیش‌ترین مقدار به‌دست آمده برای آماره‌های این آزمون را نشان می‌دهد. مشاهده می‌شود که مقادیر به‌دست آمده کوچک‌تر از مقدار بحرانی در نظر گرفته شده هستند. همچنین احتمال متناظر با این مقادیر در توزیع فیشر از مقدار احتمال بحرانی بزرگ‌تر است. بنابراین با وجود ماهانه بودن سری داده‌ها، مؤلفه تناوبی در آن وجود ندارد.

علاوه بر روند، مؤلفه تناوبی نیز عامل نایستا شدن سری است. بنابراین وجود این مؤلفه در سری زمانی تراز ماهانه سطح دریاچه با استفاده از آزمون فیشر بررسی شد. به این ترتیب که آماره این آزمون برای ۵۷۶ هارمونی از فرکانس پایه $(1/N)$ محاسبه شد. سپس مقادیر به‌دست آمده با مقدار بحرانی در سطح معنی‌داری ۰.۵٪ مقایسه گردید. جدول ۲ مقادیر کم‌ترین

جدول ۲- نتایج آزمون فیشر برای تعیین ترم تناوبی.

Table 2. Fisher's test results to determine the periodic semester.

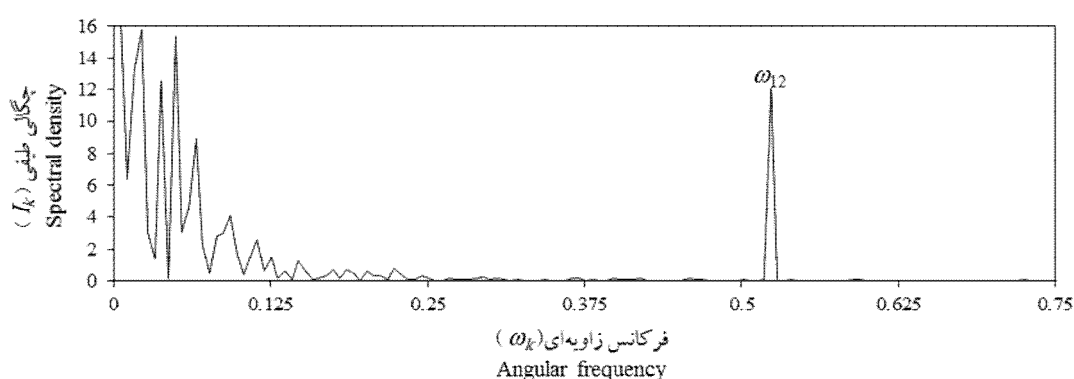
| حد بحرانی Critical | حداکثر Max | حداقل Min | آماره آزمون Test statistics |
|------------------------|-------------------------|-------------------------|--------------------------------|
| $F(2,1150) = 3.000$ | $F_{\max}^* = 0.105$ | $F_{\min}^* = 0.000$ | احتمال آماره (%) |
| $P_{F(2,1150)} = 5.00$ | $P_{F_{\max}^*} = 0.90$ | $P_{F_{\min}^*} = 1.00$ | (%) The probability statistic |

همراه تفاضل‌گیری غیرفصلی (d) و تفاضل‌گیری توأم (هم غیرفصلی و هم فصلی) بر روی داده‌ها انجام داد. بدین‌منظور سری اصلی به سه سری جدید تبدیل می‌شود. شکل ۴ این سری‌ها را نشان می‌دهد. با مشاهده سری اصلی (شکل ۱) نمی‌شد با اطمینان عدم وجود روند صعودی و نزولی را نتیجه گرفت. اما مشاهده می‌شود که نمودار سری‌های تفاضل‌گیری شده کاملاً افقی هستند. در نتیجه روندی در آن‌ها وجود ندارد. به هر حال از روی این نمودارها نمی‌توان درباره وجود یا عدم وجود تناوب در آن‌ها حرفی زد.

هر چند مشخص شد که روند و تناوب به‌عنوان دو عامل مهم نایستایی در سری تراز ماهانه سطح دریاچه وجود ندارد، اما باز هم نمی‌توان با اطمینان گفت که سری ایستا است. زیرا ممکن است دیگر عوامل نایستایی هم‌چنان وجود داشته باشند. بنابراین برای اطمینان از ایستا بودن سری‌های اصلی و تفاضل‌گیری‌شده آزمون دیکی - فولر تعمیم‌یافته به‌کار برده شد. جدول ۳ نتایج این آزمون را نشان می‌دهد.

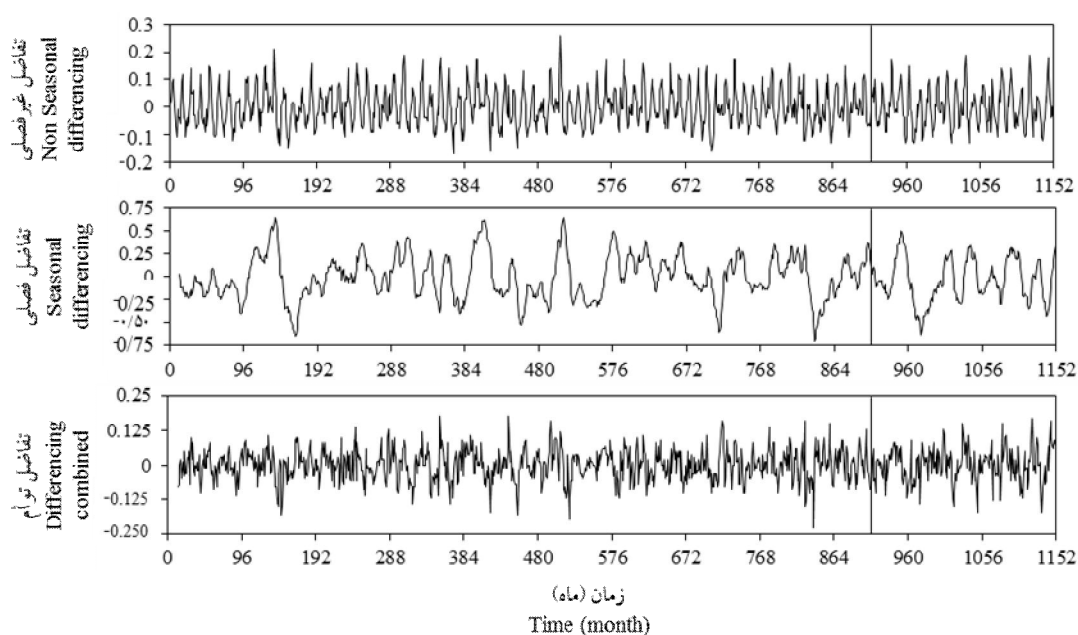
یکی از اهداف این پژوهش بررسی تأثیر انواع تفاضل‌گیری بر میزان ایستایی تراز ماهانه سطح دریاچه است. در هنگام استفاده از تفاضل‌گیری فصلی، گام تفاضل‌گیری باید مناسب باشد تا نتیجه مطلوب که همان ایستا شدن سری است، حاصل شود. در سری‌هایی که دارای مؤلفه تناوبی هستند، این گام برابر دوره تناوب است. اما مشاهده شد که در سری مورد مطالعه، مؤلفه تناوبی وجود ندارد. بنابراین به‌منظور شناسایی مناسب‌ترین گام تفاضل‌گیری از تحلیل طیفی بهره گرفته شد. به این ترتیب نمودار طیفی داده‌ها رسم گردید (شکل ۳). مشاهده می‌شود که این نمودار در فرکانس زاویه‌ای معادل با تأخیر ۱۲ ماه، دارای نقطه اوج می‌باشد. پس می‌توان دریافت که مناسب‌ترین گام برای تفاضل‌گیری فصلی ۱۲ ماه است. البته از آنجایی که داده‌ها ماهانه هستند این نتیجه مورد انتظار بود. اما از طرف دیگر داده‌ها تناوب معنی‌داری نداشتند. بنابراین تحلیل طیفی برای رسیدن به چنین نتیجه‌ای ضرورت یافته بود.

با مشخص شدن گام مناسب (ω) برای تفاضل‌گیری فصلی (D)، می‌توان این عمل را به



شکل ۳- نمودار طیفی سری زمانی در فضای فرکانس.

Figure 3. Frequency spectral time series graph in space.



شکل ۴- سری‌های زمانی تفاضل‌گیری شده.

Figure 4. Time series by differencing.

جدول ۳- نتایج آزمون ایستایی دیکی - فولر تعمیم یافته.

Table 3. Dickey-Fuller test results table.

| $d = 1$ $D = 1$ | $d = 0$ $D = 1$ | $d = 1$ $D = 0$ | $d = 0$ $D = 0$ | حد بحرانی Critical | نوع آزمون Type of test |
|--------------------|--------------------|--------------------|--------------------|-----------------------|--|
| -17.61 | -7.13 | -6.27 | -0.24 | -1.94 | بدون ثابت و روند Without constant and Rand |
| -17.60 | -7.13 | -6.27 | -4.33 | -2.87 | فقط ثابت Only Constant |
| -17.60 | -7.13 | -6.27 | -4.35 | -1.64 | ثابت و روند خطی Constant and linear trend |
| -17.59 | -7.14 | -6.27 | -4.93 | -1.64 | ثابت و روندهای خطی و درجه دو Constant linear and quadratic trends |

می‌شود که مقادیر آماره‌ها برای تفاضل‌گیری توأم، فصلی، غیرفصلی و سری اصلی به ترتیب اختلاف بیش‌تری با حدود بحرانی دارند. بنابراین می‌توان دریافت که چهار سری مذکور، به ترتیب دارای ایستایی بیش‌تری هستند.

از یک طرف آزمون دیکی - فولر تعمیم یافته برای تعیین ایستایی سری‌های فصلی ضعیف عمل می‌کند.

در آن نتایج برای چهار حالت ایستایی ارائه شده است. مشاهده می‌شود که برای حالات فقط ثابت، ثابت و روند خطی و ثابت و روندهای خطی و درجه دو، مقدار آماره آزمون برای سری اصلی و انواع سری‌های تفاضل‌گیری شده کم‌تر از حد بحرانی است. بنابراین تمام این سری‌ها از نظر وجود روندهای خطی و درجه دو ایستا هستند. همچنین مشاهده

باعث کاهش این خودهمبستگی تا ۱۱ گام شده است. تفاضل‌گیری توأم نیز این مقدار را به ۳ گام غیرفصلی و یک گام فصلی با تأخیر ۱۲ کاهش داده است. در نتیجه تفاضل‌گیری در سری‌هایی که خود ایستا هستند نیز می‌تواند میزان ایستایی را قوی‌تر کند.

از روی همبستگی‌نگارها علاوه بر ایستایی، بیش‌ترین تعداد پارامترهای موردنیاز برای مدل‌های سری‌های زمانی نیز تعیین گردید. این تعداد که بر اساس گام‌های با مقادیر بزرگ از این نمودار به‌دست آمد در جدول ۴ ارائه شده است. در این جدول همچنین تعداد مدل‌های موردنیاز برای دستیابی به دقیق‌ترین مدل ارائه شده است. مشاهده می‌شود که برای سری اصلی به تعداد بسیار زیادی مدل احتیاج است. در صورتی که تفاضل‌گیری‌های فصلی و توأم این مقدار را به اندازه قابل‌توجهی کاهش داده‌اند. با این حال تفاضل‌گیری توأم به کم‌تر از نصف تعداد مدل‌های تفاضل‌گیری فصلی احتیاج دارد. بنابراین ایستایی بیش‌تر سری به کمک تفاضل‌گیری این مزیت را هم دارد که با تعداد کم‌تری مدل می‌توان به دقیق‌ترین مدل دست یافت. از آن‌جایی هم که سری حاصل از تفاضل‌گیری غیرفصلی نایستا است، نمی‌توان آن را مدل‌سازی کرد.

از طرف دیگر هر چند معلوم شد که داده‌ها تناوبی نیستند، اما نتایج تحلیل طیفی نشان داد که از آن‌جایی که به‌صورت ماهانه اندازه‌گیری شده‌اند، ممکن است آثار فصلی بودن در آن‌ها وجود داشته باشد. همچنین در مورد فصلی بودن داده‌های تفاضل‌گیری شده اطلاعاتی در دسترس نیست. بنابراین به‌منظور اطمینان بیش‌تر با رسم همبستگی‌نگار داده‌ها نیز ایستا بودن انواع سری‌های اصلی و تفاضل‌گیری‌شده بررسی شد که در شکل ۵ ارائه شده‌اند. مشاهده می‌شود که سری‌های اصلی، تفاضل‌گیری‌شده فصلی و توأم پس از چند گام به صفر نزدیک شده و داخل حدود اطمینان ۹۵٪ قرار گرفته‌اند. بنابراین در این مرحله ایستا بودن آن‌ها محرز می‌شود. اما سری تفاضل‌گیری‌شده غیرفصلی نامیرا و نوسانی است. بنابراین این سری نایستا و دارای تناوب است. نتیجه‌ای که آزمون دیکی-فولر در شناسایی آن ناتوان بود. از این موضوع می‌توان نتیجه گرفت که تفاضل‌گیری مرتبه اول همیشه باعث ایستا شدن سری نخواهد شد. بنابراین همواره بعد از این عمل باید ایستایی مورد بررسی قرار گیرد. از سوی دیگر مشاهده می‌شود که هر چند سری اصلی خود ایستا است اما تا ۳۷ گام دارای خودهمبستگی زیادی می‌باشد. تفاضل‌گیری فصلی

جدول ۴- مرتبه‌های پارامترها و تعداد مدل‌های مورد نیاز.

Table 4. Order parameter and the number of models needed.

| NOM | Q | P | q | P | نوع سری Type of series |
|------|-----|-----|-----|-----|--|
| 1444 | 0 | 0 | 37 | 37 | سری اصلی The Original Series |
| 144 | 0 | 0 | 11 | 11 | تفاضل‌گیری فصلی Seasonal differencing |
| 64 | 1 | 1 | 3 | 3 | تفاضل‌گیری توأم Differencing combined |

۲۰ تا ۳۳) برای نتایج تمامی مدل‌ها محاسبه و دقیق‌ترین مدل براساس همین معیارها انتخاب گردید. دقیق‌ترین مدل‌ها به همراه نتایج معیارهای ارزیابی در دوره اعتبارسنجی در جدول ۵ ارائه شده‌اند. مشاهده می‌شود که دقت هر سه مدل تقریباً مشابه است و مدل SARIMA به مقداری ناچیزی دقیق‌تر از دیگر مدل‌هاست. اما باید توجه داشت که این مدل تنها با ۲ پارامتر به این دقت رسیده است، در حالی که مدل‌های ARMA و ARIMA به ترتیب با ۲۱ و ۲۰ پارامتر به این دقت رسیده‌اند. این نیز بیانگر مزیت دیگر ایستایی بیشتر حاصل از تفاضل‌گیری توأم است.

موضوع دیگری که از جدول ۴ می‌توان دریافت، نوع مدل مناسب برای مدل‌سازی و پیش‌بینی است. از آنجایی که سری اصلی بدون تفاضل‌گیری و پارامترهای فصلی است، مدل ARMA برای مدل‌سازی آن مناسب است. به این ترتیب مدل مناسب برای سری حاصل از تفاضل‌گیری فصلی ARIMA و برای تفاضل‌گیری توأم SARIMA می‌باشد. بنابراین سری‌های مذکور توسط همه مدل‌های ارائه شده در جدول ۴ مدل‌سازی و سپس توسط مدل‌های به‌دست آمده، در دوره اعتبارسنجی، پیش‌بینی شدند. سپس معیارهای ارزیابی (رابطه‌های

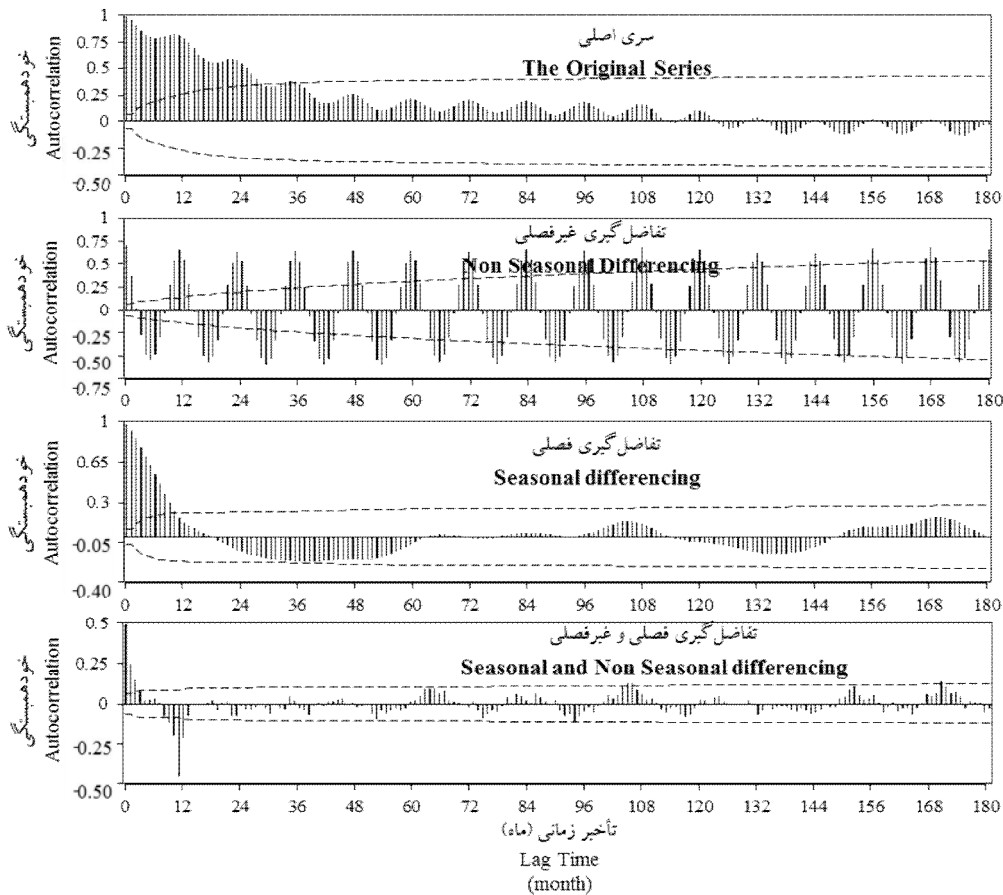
جدول ۵- نتایج معیارهای ارزیابی دقیق‌ترین مدل‌ها.

Table 5. The results of the detailed evaluation criteria most models.

| SARIMA(1,1,0)(0,1,1) ₁₂ | ARIMA(10,0,10)(0,1,0) ₁₂ | ARMA(13,8) | معیار ارزیابی Evaluation Criteria |
|------------------------------------|-------------------------------------|------------|--------------------------------------|
| 98.943 | 98.823 | 98.936 | $(\%) R^2$ |
| 98.92 | 98.8 | 98.91 | $(\%) VAF$ |
| 0.000204 | 0.000215 | 0.000205 | $(\%) RMSRE$ |
| 0.0359 | 0.0379 | 0.0361 | RMSE |
| 0.000 | 0.000 | 0.000 | CRM |

سه مدل اکو استیت نتورک (ESN)، شبکه عصبی برگشت‌کننده (RNN) و شبکه عصبی بیزی (BNN) به پیش‌بینی تراز سطح این دریاچه پرداخت که مقدار RMSE برای این سه مدل به ترتیب برابر ۰/۰۴، ۰/۰۵ و ۰/۱۲ به‌دست آمد (۳۷). با مقایسه تمام این مقادیر با مقادیر به‌دست آمده در جدول ۵ می‌توان دریافت که عملکرد مدل‌های ARMA، ARIMA و SARIMA از تمام این مدل‌های جدیدتر و پیچیده‌تر برای داده‌هایی با شرایط این منطقه بهتر خواهد بود. این در حالی است که این مدل‌ها از مزایای سادگی و سرعت بالاتر نیز برخوردارند.

با مقایسه نتایج این پژوهش با پژوهش‌های دیگر می‌توان دریافت که مدل‌های خطی و ساده‌تر به‌کار رفته در اینجا، برای دریاچه‌ای با این ویژگی‌ها، دقیق‌تر از مدل‌های غیرخطی و پیچیده و جدیدتر هستند. برای مثال آلتونکاینک (۲۰۱۴) از چهار مدل منطق فازی (FL)، پرسپترون چندلایه (MLP)، تبدیل موجک- فازی (Wavelet-Fuzzy) و تبدیل موجک- چندلایه (Wavelet-multilayer) در پیش‌بینی تراز سطح همین دریاچه بهره برد. نتایج او نشان داده که مقدار معیار RMSE برای هر یک از این چهار مدل به ترتیب برابر ۰/۱۶۶، ۰/۱۶۵، ۰/۱۱۲ و ۰/۱۱۸ می‌باشد (۳۶). همچنین کولیالی (۲۰۱۰) با استفاده از



شکل ۵- همبستگی نگارهای انواع سری‌ها.
Figure 5. Graphical correlation of Series.

است که مقدار احتمال متناظر با آماره آزمون برای هر سه مدل بزرگتر از سطح معنی‌داری ۰.۰۵ است. بنابراین باقی‌مانده‌های آن‌ها مدل مستقل است و اعتبارشان تأیید می‌گردد.

اعتبار مدل‌های به‌دست آمده زمانی تأیید می‌شود که باقی‌مانده‌های آن‌ها مستقل (تصادفی) باشند. بنابراین آزمون تصادفی بودن والد- ولفوویتس بر روی باقی‌مانده‌های دقیق‌ترین مدل‌ها صورت گرفت. نتایج این آزمون در جدول ۶ ارائه شده است. واضح

جدول ۶- نتایج آزمون والد- ولفوویتس برای تعیین تصادفی بودن باقی‌مانده‌های مدل‌ها.

Table 6. Wald-Wolfowitz test results to determine the randomness of the residuals models.

| SARIMA(1,1,0)(0,1,1) ₁₂ | ARIMA(10,0,10)(0,1,0) | ARMA(13,8) | نام پارامتر Parameter name |
|------------------------------------|-----------------------|------------|-------------------------------|
| 443 | 439 | 422 | R |
| -0.39769 | -0.78603 | -0.89867 | u_R |
| 0.690858 | 0.431852 | 0.36883 | $(\%) P_{u_R}$ |

نتیجه گیری

سری زمانی تراز ماهانه سطح دریاچه میشیگان- هارژن به خودی خود فاقد مؤلفه‌های روند و تناوبی و در نتیجه ایستا و آماده مدل‌سازی بود. با این حال انجام انواع تفاضل‌گیری بر روی آن صورت گرفت که مهم‌ترین نتایج به شرح زیر است:

۱- تفاضل‌گیری فصلی و توأم میزان ایستایی سری را بیش‌تر و تفاضل‌گیری غیرفصلی سری را نایستا کرد. در این بین تفاضل‌گیری توأم بیش‌ترین ایستایی را نتیجه داد.

۲- ایستایی بیش‌تر داده‌ها دستیابی به دقیق‌ترین مدل را آسان کرد به این ترتیب که به تعداد بسیار کم‌تری مدل نیاز بود.

۳- تفاضل‌گیری توأم این مزیت را هم داشت که علاوه بر نیاز به کم‌ترین تعداد مدل، با پارامترهای بسیار کم‌تری به دقیق‌ترین نتیجه رسید.

۴- مدل‌های قدیمی و ساده استوکستیک برای پیش‌بینی تراز سطح دریاچه با ویژگی‌های ۱- ایستا، ۲- فاقد روند، تناوب و پرش و ۳- ماهانه بهتر از مدل‌های جدیدتر و پیچیده‌تر است.

منابع

1. Azad Talatapeh, N., Behmanesh, J., and Montasari, M. 2013. Predicting Potential Evapotranspiration Using Time Series Models (Case study: Urmia). *J. Water Soil.* 27: 1. 213-223. (In Persian)
2. Bolyani, Y., Fazelnia, G., and Bayat, A. 2012. Analysis and modeling annual temperature of Shiraz using ARIMA model. *Geographic Space.* 12: 38. 127-144. (In Persian)
3. Brockwell, P.J., and Davis, R.A. 1991. *Time series: theory and methods.* Second edition, Springer Science & Business Media, NY, 577p.
4. Chebaane, M., Salas, J.D., and Boes, D.C. 1995. Product periodic autoregressive processes for modeling intermittent monthly stream flows. *J. Water Resour. Res.* 31: 6. 1513-1518.
5. Çimen, M., and Kisi, O. 2009. Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *J. Hydrol.* 378: 3-4. 253-262.
6. Cryer, J.D., and Chan, K.S. 2008. *Time Series Analysis With Applications in R.* Second Ed., Springer, NY, 491p.
7. Giri, A., and Singh, N.B. 2014. Comparison of Artificial Neural Network Algorithm for Water Quality Prediction of River Ganga. *Environ. Res. J.* 8: 2. 55-63.
8. Hirsch, R.M., and Slack, J.R. 1984. A nonparametric trend test for seasonal data with serial dependence. *J. Water Resour. Res.* 20: 6. 727-732.
9. Kashyap, R.L., and Ramachandra Rao, A. 1976. *Dynamic stochastic models from empirical data.* Academic press, NY, 352p.
10. Khatibi, R., Ghorbani, M., Naghipour, L., Jothiprakash, V., Fathima, T., and Fazelifard, M. 2014. Inter-comparison of time series models of lake levels predicted by several modeling strategies. *J. Hydrol.* 511: 530-545.
11. Khazaei, M., and Mirzaei, M. 2014. Forecasting the climatic variables using time series analysis of Zohre catchment. *Sci. J. Manage. Syst.* 14: 34. 233-250. (In Persian)
12. Kisi, Ö. 2004. River flow modeling using artificial neural networks. *J. Hydrol. Engin.* 9: 1. 60-63.
13. Kisi, O., and Cigizoglu, H.K. 2007. Comparison of different ANN techniques in river flow prediction. *J. Civil Engin. Environ. Syst.* 24: 3. 211-231.
14. Kisi, O., Shiri, J., Karimi, S., Shamsirband, Sh., Motamedi, Sh., Petkovic, D., and Hashim, R. 2015. A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *J. Appl. Math. Comp.* 270: 731-743.
15. Makarynska, D., and Makarynsky, O. 2008. Predicting sea-level variations at the Cocos (Keeling) Islands with artificial neural networks. *J. Comp. Geosci.* 34: 12. 1910-1917.

16. Marco, J.B., Harboe, R., and Salas, J.D. 1993. Stochastic hydrology and its use in water resources systems simulation and optimization. Springer Science & Business Media, Peniscola, Spain, 483p.
17. Maroofi, S., Khotar, B., Sadeghifar, M., Parsafar, N., and Ildormi, A. 2014. Forecasting the drought using SARIMA time series and SPI index in the central region of the Hamedan province. 28: 1. 213-235. (In Persian)
18. Omid, R., Radmanesh, F., and Zarei, H. 2013. River flow prediction using stochastic models. The First National Conference on Challenges on Water Resources and Agricultural, 13th February, Khorasgan Branch of Islamic Azad university, Iran, 8p. (In Persian)
19. Peña, D., Tiao, G.C., and Tsay, R.S. 2011. A course in time series analysis. John Wiley & Sons, INC, NY, 460p.
20. Poormohammadi, S., Malekinezhad, H., and Poorshareyati, R. 2013. Comparison of ANN and time series appropriately in prediction of ground water table (Case study: Bakhtegan basin). J. Water Soil Cons. 20: 4. 251-262. (In Persian)
21. Said, S.E., and Dickey, D.A. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika. 71: 3. 599-607.
22. Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L. 1980. Applied modeling of hydrologic time series. Water Resources Publication, Colorado, 484p.
23. Shafaei, M., and Kisi, O. 2015. Lake Level Forecasting Using Wavelet-SVR, Wavelet-ANFIS and Wavelet-ARMA Conjunction Models. J. Water Resour. Manage. 30: 1. 79-97.
24. Shamim, M.A., Hassan, M., Ahmad, S., and Zeeshan, M. 2015. A comparison of Artificial Neural Networks (ANN) and Local Linear Regression (LLR) techniques for predicting monthly reservoir levels. KSCE J. Civil Engin. 8p. DOI: 10.1007/s12205-015-0298-z.
25. Sharma, N., Zakaullah, M., Tiwari, H., and Kumar, D. 2015. Runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed. Modeling Earth Systems Environment 1 (23), 8p. DOI: 10.1007/s40808-015-0027-0.
26. St-Hilaire, A., Ouarda, T.B., Bargaoui, Z., Daigle, A., and Bilodeau, L. 2012. Daily river water temperature forecast model with a k-nearest neighbour approach. J. Hydrol. Proc. 26: 9. 1302-1310.
27. Tao, P.C., and Delleur, J.W. 1976. Seasonal and nonseasonal ARMA models in hydrology. J. Hydrol. Div. 102: 10. 1541-1559.
28. Veisipoor, H., Samakoosh, J.M., Sahneh, B., and Yousofi, Y. 2010. Analysis prediction the precipitation and temperature using time series models (ARIMA). Geography. 4: 12. 63-70. (In Persian)
29. Wald, A.B., and Wolfowitz, J.A. 1943. An exact test for randomness in the non-parametric case based on serial correlation. The Annals of Mathematical Statistics. 14: 4. 378-88.
30. Wang, W.C., Chau, K.W., Xu, D.M., and Chen, X.Y. 2015. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. J. Water Resour. Manage. 29: 8. 2655-2675.
31. Wilcox, D.A., Thompson, T.A., Booth, R.K., and Nicholas, J.A. 2007. Water-level variability and water availability. J. Great Lakes. Geological Survey Circular. 1311, U.S., 25p.
32. Jabbari Gharabagh, S., Rezaei, H., and Mohammadnezhad, B. 2015. Comparison of reconstructed phase space and chaotic behavior of Nazloochay river flow at different temporal scales. J. Water Soil Cons. 22: 5. 135-151. (In Persian)
33. Rajaei, T., and Ebrahimi, H. 2015. Application of wavelet-neural network model for forecasting of groundwater level time series with non-stationary and nonlinear characteristics. J. Water Soil Cons. 22: 5. 99-115. (In Persian)
34. Ahmadi, F., Dinpazhooh, Y., Fakherifard, A., Khalili, K., and Darbandi, S. 2015. Comparing Nonlinear Time Series Models and Genetic Programming for Daily River Flow Forecasting (Case study: Barandouz-Chai River). J. Water Soil Cons. 22: 1. 151-169. (In Persian)

35. Rajaei, T., and Broomand, A. 2016. Prediction of Monthly Dissolved Oxygen Using Wavelet and Artificial Neural Network Combined Model. J. Water Soil Cons. 22: 6. 153-169. (In Persian)
36. Altunkaynak, A. 2014. Predicting water level fluctuations in Lake Michigan-Huron using wavelet-expert system methods. Water resources management. 28: 8. 2293-2314.
37. Coulibaly, P. 2010. Reservoir computing approach to Great Lakes water level forecasting. J. Hydrol. 381: 1. 76-88.



The effect of differencing in stationary and accuracy of time series in predicting of lake level

***M. Khademi¹, H. Moeeni¹, H. Bonakdari² and I. Ebtehaj³**

¹M.Sc. Graduate, Dept. of Civil Engineering, Razi University of Kermanshah,

²Professor, Dept. of Civil Engineering, Razi University of Kermanshah,

³Ph.D. Student, Dept. of Civil Engineering, Razi University of Kermanshah

Received: 06/23/2016; Accepted: 09/21/2017

Abstract

Background and Objectives: One of the most important assumptions in the modeling of time series, it is to be stationary. The amount of stationary can be various, so that different definitions exist such as first order and second order stationary and strong and strict stationary. Therefore, this study cover the effect of differencing on the stationary value as well as the precision of the ARMA, ARIMA and SARIMA models in the modeling and monthly prediction of time series.

Materials and Methods: For this purpose, 96 years data of lake level, which is monthly measurement related to Michigan-Huron Lake on the border of United States and Canada, are used. The 76-years of primary utilized for training and the rest of 20-years are used for validation. Firstly, the existing of the trend and period components in the time-series were assessed using Fischer and man-Kendal tests. These two components are the main factors in the appearance of non-stationary in time series. Therefore seasonal differencing, non-seasonal differencing and both of them at same time were measured and their results were compared by non- differencing data. To assessment of achieved time-series differencing, the ACF diagram and generalized Dicky-Fouler test were utilized. The type and amount of required parameters in different models were determined by ACF diagram. Then, each of series was modeled and predicted using appropriate model. The results indicated that there is not a certain trend and period in series. However, the using of seasonal differencing increased the stationary but non-seasonal differencing lead to non-stationary of these time series. The most increasing in stationary was indicated by using of seasonal and non-seasonal differencing. Due to ACF diagram, using both of differencing results in use of seasonal parameters in model. Therefore, series without differencing with ARMA model and series with seasonal differencing with SARIMA are modeled.

Results: The investigations showed the concurrent using of seasonal and non-seasonal differencing has the most impact on the rate of getting stationary alignment of the Lake in compare with other methods. As a result, the numbers of model needed to achieve the most accurate predictions were reduced in large scale. In such a way in non-differencing situation, 1444 model of ARMA were needed that this amount in situation of seasonal differencing and non-seasonal differencing were reduced in 64 models of SARIMA. On the other hand, by reducing much more number of parameters (two parameters) in SARIMA model, similar result is even better than ARMA model with 21 parameters.

Conclusion: The results showed that the more making stationary of monthly lake level which itself is stationary, reduces the number of models and the number of model's parameters needed to achieve the best outcome too much. For this purpose, combined differencing made the series stationary more than the other methods.

Keywords: Forecast, Time series, Lake level, ARIMA, SARIMA

* Corresponding Author; Email: mahsa.khademi88@yahoo.com