

## پیش‌بینی بلندمدت جریان رودخانه با استفاده از روش تحلیل طیف تکین در حوضه کرخه

فرید فروغی<sup>۱\*</sup>، شهاب عراقی نژاد<sup>۲</sup>

۱. دانشجوی دکتری، گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران و مربی دانشگاه

شیراز، دانشکده کشاورزی و منابع طبیعی داراب، شیراز

۲. دانشیار، گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج

(تاریخ دریافت: ۱۳۹۵/۶/۱۳ - تاریخ تصویب: ۱۳۹۵/۷/۱۸)

### چکیده

آبدهی رودخانه یکی از مهمترین داده‌های ورودی حوضه آبریز، برای استفاده در مدل‌های هیدرولوژیکی می‌باشد. در ده‌های گذشته روش‌های مختلفی برای تجزیه و تحلیل و پیش‌بینی متغیرهای فیزیکی استفاده شده است. یکی از آنها، روش آماری تحلیل طیف تکین (SSA) می‌باشد. SSA از روش‌های مورد استفاده در مدل‌سازی فرایندهای مختلف آماری است و اخیراً استفاده از آن در رشته‌های مختلف مهندسی از جمله منابع آب به منظور رفع اجزاء تصادفی موجود در سری‌های زمانی گسترش یافته است. هدف اصلی از این تحقیق پیش‌بینی مقادیر جریان رودخانه‌های حوضه کرخه با استفاده از روش SSA می‌باشد. ابتدا ایستگاه‌های شاخص در حوضه کرخه (پنج ایستگاه) برای این کار انتخاب گردید. دوره پراپی ایستگاه‌ها تعیین شد. برای مدل‌سازی به روش SSA از نرم‌افزار آماری CaterpillarSSA و R استفاده گردید. ۷۰ درصد داده‌ها برای واسنجی و ۳۰ درصد داده‌ها برای صحت‌سنجی در نظر گرفته شد. ابتدا از روش SSA برای پردازش اولیه داده‌ها و رفع نوفه موجود در سری‌های زمانی جریان رودخانه استفاده شد. در مرحله بعد از الگوریتم بازگشتی روش SSA برای ساخت مدل پیش‌بینی آبدهی رودخانه در ایستگاه‌های حوضه کرخه استفاده شد. برای بررسی عملکرد مدل از معیار جذر میانگین مربعات خطا نرمال شده، میانگین قدرمطلق خطای نسبی و ضریب همبستگی استفاده گردید. بر این اساس در مرحله واسنجی بیشترین و کمترین مقدار آماره NRMSE به ترتیب برابر ۰/۴۳ و ۰/۳۰ (همین‌طور آماره MARE به ترتیب برابر ۰/۳۶ و ۰/۲۷) برای ایستگاه‌های پل‌چهر و چم انجیر به دست آمد. در مرحله صحت‌سنجی بیشترین مقدار آماره NRMSE و MARE به ترتیب برابر ۰/۴۷ و ۰/۵ و برای ایستگاه‌های پل‌چهر بود. کمترین مقدار آماره NRMSE برای ایستگاه پل دختر و چم انجیر نزدیک به هم و برابر ۰/۳ و ۰/۳۱ و کمترین مقدار آماره MARE برای ایستگاه چم انجیر و پل دختر، نزدیک به هم و برابر ۰/۲۹ و ۰/۳۰ بود. در نهایت بهترین و ضعیف‌ترین نتیجه در دو مرحله واسنجی و صحت‌سنجی به ترتیب برای ایستگاه‌های چم انجیر و پل‌چهر به دست آمد. بر اساس نتایج این تحقیق می‌توان با به‌کارگیری روش SSA مقادیر جریان رودخانه را با دقت مناسب پیش‌بینی نمود.

**واژه‌های کلیدی:** پیش‌بینی بلندمدت، آبدهی، جریان رودخانه، تحلیل طیف تکین، حوضه کرخه.

### مقدمه

روش تحلیل طیف تکین<sup>۱</sup> (SSA) یکی از روش‌های تحلیل سری‌های زمانی است که با توجه به توانمندی آن مورد استقبال زیادی قرار گرفته است و روز به روز بر گستره کاربرد آن در حوضه‌های مختلف افزوده می‌شود. SSA یکی از روش‌های ناپارامتری نسبتاً جدید و قدرتمند برای تحلیل سری زمانی است که دارای قابلیت فراوانی است و می‌توان از آن در کاهش سطح نوفه<sup>۲</sup> و مدل‌سازی بهره گرفت. گرایش کاربران مختلف به

استفاده از روش SSA به‌عنوان ابزاری نیرومند در تحلیل سری‌های زمانی، حاصل عدم وابستگی به پیش‌فرض‌های محدودکننده‌ای است که در اغلب روش‌ها وجود دارد (مانند خطی بودن، نرمال بودن، ...) و در عین حال توانمندی آن در بررسی مسائل متنوع مانند پیش‌بینی و پالودن است. SSA مبتنی بر مباحث جبر ماتریس‌ها، سری‌های زمانی کلاسیک، آمار چند متغیره، هندسه چند متغیره، سیستم‌های پویا و پردازش سیگنال است (Hassani, 2007). اغلب محققینی که با SSA سر و کار دارند پیدایش آن را به مقاله‌هایی از Broomhead and King (1986) و Broomhead et al. (1987) نسبت می‌دهند هر چند در برخی منابع به نوشته‌هایی پیش از

\* نویسنده مسئول : foroughifarid@gmail.com

1. Singular Spectrum Analysis  
2. Noise

زمانی نیز به گونه‌ای می‌باشد که اغلب بارندگی‌های مناطق خشک و نیمه‌خشک به صورت رگبار در چند روز محدود از سال می‌بارد و دامنه نوسانات آن در طول یک دوره آماری بسیار بالاست (Hajibigloo *et al.*, 2013). در چنین شرایطی برآورد صحیح آبدهی رودخانه جهت برنامه‌ریزی در مدیریت منابع آب، پیش‌بینی تداوم خشکسالی‌ها و برنامه‌ریزی بهره‌برداری از مخزن از اهمیت زیادی برخوردار است. با توجه به اینکه تاکنون در ایران از این روش برای پیش‌بینی آبدهی استفاده نشده است و یک روش نسبتاً جدیدی می‌باشد و با توجه به قابلیت‌های بالای این روش ضرورت استفاده از این روش مشخص می‌گردد. لذا در این تحقیق پیش‌بینی دوره پرآبی (دی تا خرداد) جریان رودخانه-های حوضه کرخه در غرب ایران با استفاده از روش SSA مدنظر قرار گرفت.

## مواد و روش‌ها

### موقعیت جغرافیایی منطقه و ایستگاه‌های آب‌سنجی

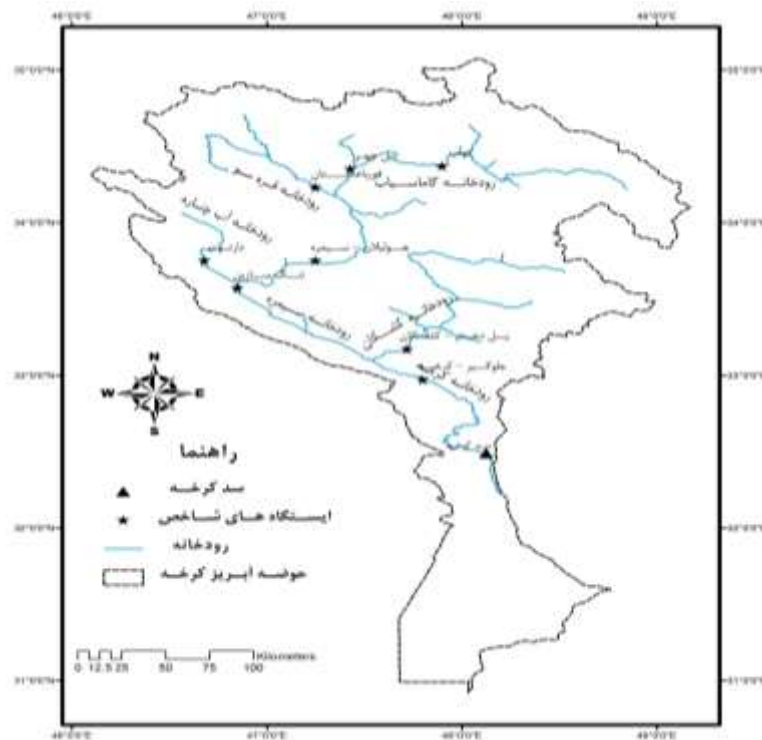
حوضه کرخه با مساحت ۵۱۶۴۰ کیلومترمربع در جنوب غرب ایران در محدوده ۵۸' ۳۰° تا ۵۶' ۳۴° عرض شمالی و ۴۶° ۰۶' تا ۴۹° ۱۰' طول شرقی قرار دارد (شکل ۱). ارتفاع حوضه از ۳ متر در جنوب (دشت آزادگان) تا ۳۶۴۵ متر در کوه‌های کارین (حد شرقی آن با حوضه دز) متغیر است. رودخانه کرخه ۹۰۰ کیلومتر طول دارد و سومین رودخانه بزرگ کشور بر اساس متوسط آبدهی سالانه (۸/۵ میلیارد مترمکعب) است. اقلیم حوضه کرخه خشک و نیمه‌خشک با زمستان ملایم و مرطوب و تابستان گرم و خشک است. متوسط بارش حوضه از ۱۵۰ میلی-متر در جلگه‌های خشک جنوبی تا ۷۵۰ میلی-متر در کوه‌های شمالی متغیر است و به سبب کوهستانی بودن حوضه جریان رودخانه‌های آن از بارش و برف نشأت می‌گیرد (Jamab Consulting Engineers, 2006).

برای انجام این تحقیق ایستگاه‌های شاخص در حوضه کرخه شامل جلوگیر، پل‌دختر، قورباغستان، چم انجیر، پل‌چهر انتخاب گردید که مشخصات این ایستگاه‌ها در جدول (۱) نشان داده شده است. مقادیر آبدهی این ایستگاه‌ها از شرکت مدیریت منابع آب ایران تهیه گردید. دوره پرآبی ایستگاه‌های حوضه کرخه در ماه‌های ژانویه تا جون می‌باشد که پیش‌بینی بلندمدت میانگین جریان رودخانه در این دوره مدنظر می‌باشد. برای نمونه دوره پرآبی برای ایستگاه چم انجیر در شکل (۲) رسم گردیده است.

این مانند Basilevsky *et al.* و Kumaresan and Tufts (1980) و (1979) و حتی قبل‌تر اشاره شده است. با این وجود نسخه‌های معرفی شده توسط Broomhead and King (1986) نقطه عطفی در کاربردهای SSA است. پس از تحقیق‌های Broomhead and King (1986)، روش SSA جذابیت بیشتری پیدا کرد که منجر به کاربرد آن در حوضه‌های مختلف علوم شد. Vautard and Ghil (1989) بیان می‌کند که SSA اطلاعات درون سری‌های زمانی کوتاه‌مدت نوفه‌دار را با دقت بالا استخراج می‌کند. Lisi *et al.* (1995) در استخراج مؤلفه‌های مهم از سری‌های زمانی شاخص نوسانات جنوبی، از SSA استفاده کردند. Golyandina *et al.* (2001) مبنای نظری و کاربردی SSA را تشریح نمود. Sivapragasam *et al.* (2001) از روش SSA برای پردازش اولیه داده‌های بارندگی و رواناب استفاده کردند. نتایج این محققین نشان داد که استفاده از SSA باعث عملکرد بسیار بهتر مدل‌های پیش‌بینی بارندگی و رواناب گردید. Wu *et al.* (2010) نیز از این روش برای پردازش اولیه داده‌های بارندگی استفاده کردند و نتیجه‌گیری نمودند که پردازش اولیه داده‌ها با این روش، عملکرد مدل‌های مورد استفاده آنها را به طور قابل‌ملاحظه‌ای بهبود بخشید. Hassani and Thomakos (2007) و Hassani *et al.* (2010) روش SSA را به طور مفصل تشریح نمود و نتایج پیش‌بینی این روش را با روش‌های ساریمای، الگوریتم ای‌آرای آر<sup>۲</sup> و الگوریتم حالت وینتر<sup>۳</sup> مقایسه نمود و گزارش کرد که روش SSA با دقت بسیار زیاد نسبت به سایر روش‌های ذکر شده قادر به پیش‌بینی است. Hassani *et al.* (2011) در خصوص تفکیک‌پذیری سیگنال و نوفه و طول پنجره در روش SSA به-طور تحلیلی بررسی کردند. Hassani *et al.* (2010) از روش SSA برای کاهش سطح نوفه در روش رگرسیون خطی استفاده کردند. Marques *et al.* (2006) از روش SSA برای استخراج اجزای مهم از نوفه سری‌های زمانی هیدرولوژیکی بارش، رواناب و دمای سطح آب استفاده نمود و در مرحله بعد از الگوریتم بازگشتی SSA برای پیش‌بینی سری‌های زمانی استفاده نمود.

متوسط بارندگی در ایران حدود ۲۵۰ میلی‌متر است که کمتر از یک سوم متوسط بارندگی در جهان می‌باشد که این مقدار از توزیع زمانی و مکانی نامناسبی برخوردار است به طوری که ۷۳ درصد از سطح کشور بارندگی کمتر از ۲۵۰ میلی‌متر، ۲۶ درصد بیشتر از ۲۵۰ میلی‌متر و سطح باقی‌مانده ۱ درصد نیز بالای ۱۰۰۰ میلی‌متر بارندگی دارد. بارندگی از نظر توزیع

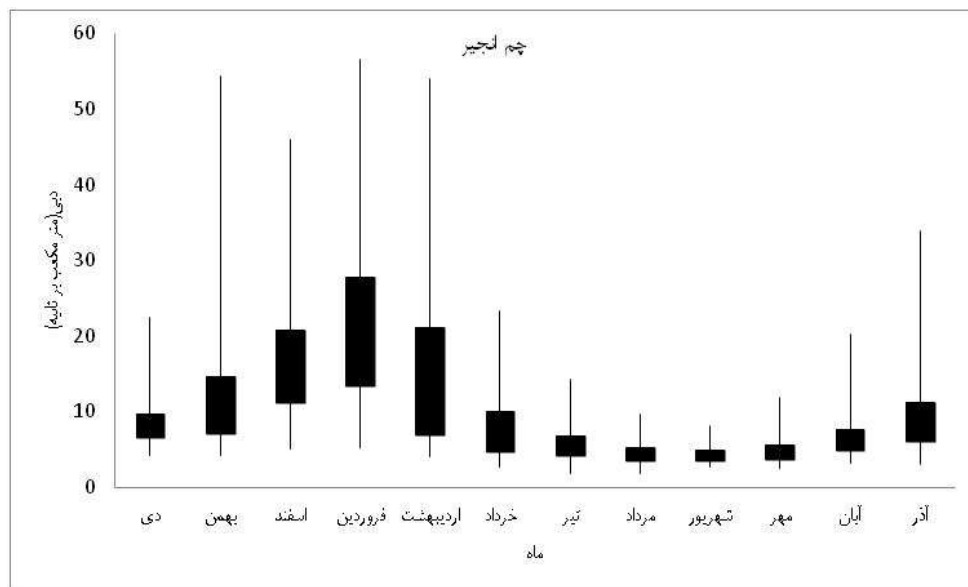
1. Sarima  
2. Arar Algorithm  
3. Holt-Winter Algorithm



شکل ۱. حوضه کرخه و ایستگاه‌های آب‌سنجی واقع در آن

جدول ۱.۱ ایستگاه‌های آب‌سنجی واقع در حوضه کرخه

ایستگاه	رودخانه	طول جغرافیایی	عرض جغرافیایی	مساحت (km <sup>2</sup> )	ارتفاع از سطح دریا (m)	متوسط آبدهی سالانه (mcm)
چمانجیر	خرم‌آباد	۴۸°۱۴'۵۶"	۳۳°۲۶'۴۴"	۱۵۹۰	۱۱۴۰	۱۰/۱
جلوگیر	کرخه	۴۷°۴۸'۰۰"	۳۲°۵۸'۰۰"	۳۹۳۸۰	۴۵۰	۱۴۴/۳
پلدختر	کشکان	۴۷°۴۳'۰۶"	۳۳°۰۹'۴۳"	۹۱۴۰	۶۵۰	۴۸/۵
قورباغستان	قره‌سو	۴۷°۱۵'۰۰"	۳۴°۱۴'۰۰"	۵۳۷۰	۱۲۶۸	۲۰/۶
پل‌چهر	گاماسیاب	۴۷°۲۶'۰۰"	۳۴°۲۰'۰۰"	۱۰۸۶	۱۲۸۰	۳۱/۸



شکل ۲. منحنی‌های توزیع دبی برای احتمالات کمترین، بیشترین، ۲۵ و ۷۵ درصد برای ایستگاه چمانجیر

نظریه تحلیل طیف تکین (SSA)

بردار ویژه  $E_X$  مقادیر ویژه  $\Lambda_X$  و ماتریس تأخیر هم‌تغییری  $C_X$  از رابطه زیر به دست می‌آید:

$$SVD(C_X) = SVD(XX^T) \quad (\text{رابطه ۳})$$

$$C_X = E_X \Lambda_X E_X^T \quad (\text{رابطه ۴})$$

اگر طرفین رابطه فوق را ابتدا از سمت چپ در  $E_X^T$  و سپس از سمت راست در  $E_X$  ضرب شود رابطه زیر به دست می‌آید:

$$C_X = E_X \Lambda_X E_X^T \quad (\text{رابطه ۵})$$

ماتریس  $\Lambda_X$  حاوی مقادیر ویژه ماتریس  $C_X$  می‌باشد.

درآیه‌های ماتریس  $\Lambda_X$  را با  $\lambda_1, \dots, \lambda_M$  که به صورت ترتیبی کاهشی است ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq 0$ ) نشان داده می‌شود. بردارهای  $E_1, \dots, E_M$  بردار ویژه متناظر با مقادیر ویژه ماتریس  $C_X$  می‌باشد اگر  $d = \max(i, \lambda_i > 0)$  باشد و  $V_i = X^T E_i / \sqrt{\lambda_i}$  در نظر گرفته شود آنگاه تجزیه مقدار تکین ماتریس مسیر  $X$  به صورت زیر نوشته می‌شود:

$$X = X_1 + \dots + X_d \quad (\text{رابطه ۶})$$

به طوری که  $X_i = \sqrt{\lambda_i} E_i V_i^T$ , ( $i = 1, \dots, d$ ) می‌باشد و رتبه ماتریس  $X_i$  برابر با یک است بنابراین این ماتریس‌ها، ماتریس‌های پایه هستند. مقادیر  $X_i$  را عناصر ماتریس مسیر  $X$  می‌نامیم. مجموعه  $(\sqrt{\lambda_i}, E_i, V_i)$  را  $i$  امین سه‌تایی ویژه<sup>۱</sup> تجزیه مقدار تکین در رابطه (۶) گویند. ریشه دوم مقادیر ویژه به نام مقادیر تکین نامیده می‌شود و دنباله  $\{\lambda_k^{1/2} : 1 \leq k \leq M\}$  طیف تکین نامیده می‌شود. مؤلفه اصلی<sup>۲</sup> PC که با  $A^k(i)$  نشان داده می‌شود برای هر بردار ویژه  $E^k(j)$  از تصویر کردن ماتریس مسیر  $X$ ، بر آن بردار ویژه به شرح زیر به دست می‌آید:

$$A^k(i) = \sum_{j=1}^M X_{(i+j-1)} E^k(j), 1 \leq i \leq K \quad (\text{رابطه ۷})$$

که زاندیس زمان می‌باشد.

ج) گام سوم گروه‌بندی

گام گروه‌بندی متناظر با افراز ماتریس‌های پایه به چندین گروه و جمع ماتریس‌ها درون هر گروه است. هنگامی که بسط (۶) به دست آمد، رویه گروه‌بندی، مجموعه نشانگرهای  $\{1, \dots, d\}$  را به  $m$  زیرمجموعه مجزا  $I_1, \dots, I_m$  افراز می‌کند. فرض کنیم  $I_i = \{i_1, \dots, i_p\}$  مجموعه نشانگرها  $i_1, \dots, i_p$  باشد، در این صورت ماتریس  $X_{I_i}$  متناظر با گروه  $I_i$  به صورت  $X_{I_i} = X_{i_1} + \dots + X_{i_p}$  تعریف می‌شود. این ماتریس‌ها برای تمامی  $I_1, \dots, I_m$  محاسبه می‌شود و در نتیجه بسط (۶) به تجزیه زیر منجر می‌شود.

روش SSA برای افزایش نسبت سیگنال به نوفه استفاده می‌گردد. روش SSA از توالی بردارها در محدوده فاز استفاده می‌کند که مانند استفاده از یک پنجره متحرک روی سری زمانی است که تصاویر متفاوتی از سری را برداشت می‌کند. الگوریتم استفاده از این روش شامل دو مرحله تجزیه (شامل دو گام تعبیه کردن و تجزیه کردن) و بازسازی (شامل دو گام گروه‌بندی و میانگین‌گیری قطری) به شرح زیر می‌باشد.

الف) گام اول تعبیه کردن

فرض کنید که  $F_N = (f_1, \dots, f_N)$  یک سری زمانی با طول  $N > 2$  است. ابتدا سری زمانی اصلی را به دنباله‌ای از بردارهای چندبعدی تبدیل نموده و ماتریس اولیه را تشکیل می‌دهیم که این فرایند را تعبیه کردن می‌نامیم. این مرحله می‌تواند به عنوان یک نگاشت در نظر گرفته شود که سری زمانی یک‌بعدی موردنظر را به سری زمانی چندبعدی  $X_1, \dots, X_K$  بردارهای  $X_i = (f_i, \dots, f_{i+L+1})'$  تبدیل می‌کند. طول ماتریس اولیه  $X$  را با  $L$  نشان می‌دهیم و آن را طول پنجره می‌نامیم. مقدار  $L$  یک عدد صحیحی است و  $1 < L < N$  می‌باشد.  $K = N - L + 1$  فرض نموده و  $K$  بردار تأخیر  $L$  بعدی را به صورت ماتریس مسیر زیر در نظر می‌گیریم:

$$X = (X_1, \dots, X_K)' = \begin{bmatrix} f_1 & f_2 & \dots & f_k \\ f_2 & f_3 & \dots & f_{k+1} \\ \vdots & \vdots & & \vdots \\ f_L & f_{L+1} & \dots & f_N \end{bmatrix} \quad (\text{رابطه ۸})$$

بردارهای ماتریس  $X_i$  را بردارهای  $L$  تأخیر (یا به طور ساده بردارهای تأخیر) می‌نامند. تنها پارامتر این مرحله طول پنجره است. ثابت می‌شود که انتخاب بهینه برای طول پنجره وقتی است که  $L$  با توجه به تعداد داده‌های هر ایستگاه برابر با میانه در نظر گرفت شود (Hassani et al., 2011; Menezes et al., 2015). در رابطه فوق  $X$  به عنوان یک مجموعه داده چند متغیره با  $L$  مشخصه و  $K$  مشاهده در نظر گرفته می‌شود. واضح است که  $x_{ij} = f_{i+j-1}$  است. ماتریس مسیر  $X$ ، یک ماتریس هنگل است یعنی عناصر مساوی روی قطرهای فرعی دارد.

ب) گام دوم تجزیه مقدار تکین (SVD)

در این قسمت مقدار تکین ماتریس مسیر  $X$  محاسبه شده و به صورت مجموعه ماتریس‌های یک رتبه‌ای دو به دو متعامد پایه-ای معرفی می‌شود. برای این کار ابتدا ماتریس کوواریانس تأخیر  $C_X = XX^T$  را که ماتریسی  $L \times L$  است محاسبه می‌شود:

$$C_X = X^T X = \begin{bmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,L} \\ C_{2,1} & C_{2,2} & \dots & C_{2,L} \\ \vdots & \vdots & & \vdots \\ C_{L,1} & C_{L,2} & \dots & C_{L,L} \end{bmatrix} \quad (\text{رابطه ۹})$$

1. Eigen Triple  
2. Principle Component

(رابطه ۱۰)

$$g_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m,k-m+1}^* & 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+1}^* & L^* \leq k \leq K^* \\ \frac{1}{N-K+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1}^* & K^* \leq k \leq N \end{cases}$$

الگوریتم پیش‌بینی بازگشتی<sup>۱</sup> (RSSA) در روش SSA

پیش‌بینی زمانی می‌تواند انجام شود که مدل مشخصی ایجاد شده باشد. در پیش‌بینی SSA، این مدل‌ها می‌توانند به کمک روابط بازگشتی خطی<sup>۲</sup> (LRF) توصیف شوند. گوییم سری  $F_N$  اختلاف بعد نابیشتر از  $d$  ( $f \dim(F_N) \leq d$ ) دارد اگر  $0 \leq d \leq N-1$  و اعداد  $a_1, \dots, a_d$  طوری وجود داشته باشد که رابطه زیر برقرار باشد آنگاه این رابطه، رابطه بازگشت خطی (LRF) نامیده می‌شود:

(رابطه ۱۱)

$$f_{i+d} = \sum_{k=1}^d a_k f_{i+d-k}, \quad 1 \leq i \leq N-d, \quad a_d \neq 0$$

با توجه به اینکه الگوریتم پیش‌بینی مبتنی بر یک رابطه بازگشتی است لذا عنوان RSSA برای نامیدن آن در نظر گرفته شده است. اساس این روش مبنی بر آن است که اگر زیر فضای  $r$  بعدی  $U_r$ ،  $span\{U_1, \dots, U_r\}$  از  $R^L$  یک فضای قائم نباشد ( $U_i$ ها بردارهای ویژه چپ ماتریس مسیر  $x$  هستند) آنگاه هر بردار در این زیرفضا در یک رابطه بازگشتی صدق می‌کند به طوری که آخرین مؤلفه آن را می‌توان به صورت یک ترکیب خطی منحصره‌فرد از سایر مؤلفه‌ها نوشت (Danilov, 1997). این رابطه بازگشتی در SSA به معادله بازگشتی خطی (LRF) مشهور است. بنابراین هر یک از گروه‌های ایجاد شده در مرحله گروه‌بندی را با فرض وجود شرایط بالا می‌توان به کمک این رابطه بازگشتی ادامه داد. حال اگر سری زمانی به گونه‌ای باشد که بتوان چنین رابطه بازگشتی را برای آن توصیف کرد آنگاه  $M$  جمله نهایی سری به کمک یک رابطه بازگشتی منحصره‌فرد به صورت زیر برآورد می‌شود:

$$g_i = \begin{cases} \tilde{f}_i & i = 1, \dots, N \\ \sum_{j=1}^{L-1} a_j g_{i-j} & i = N+1, \dots, N+M \end{cases} \quad (رابطه ۱۲)$$

در این رابطه بردار  $R = (a_1, \dots, a_{L-1})$  از رابطه زیر

محاسبه می‌شود:

$$R = \frac{1}{1 - (\pi_1^2 + \dots + \pi_r^2)} \sum_{i=1}^r \pi_i p_i^{\vee} \quad (رابطه ۱۳)$$

$$X = \overbrace{X_1 + \dots + X_d}^{SVD} = \overbrace{X_{I_1} + X_{I_2} + \dots + X_{I_m}}^{Grouping} \quad (رابطه ۸)$$

رویه انتخاب مجموعه  $I_1, \dots, I_m$  گروه‌بندی ویژه سه‌گانه نامیده می‌شود. در ساده‌ترین حالت که  $m=2$  است، گروه اول همان مؤلفه اصلی (سیگنال) و گروه باقی‌مانده، نوفه در نظر گرفته می‌شود. در این حالت  $r$  تا از بزرگ‌ترین مقادیر تکیین و بردارهای ویژه متناظر برای تقریب سری اصلی انتخاب شده و بقیه مقادیر تکیین به عنوان مؤلفه نوفه در نظر گرفته می‌شود.

(د) گام چهارم میانگین‌گیری قطری (هنکل‌سازی)

یکی از ویژگی‌های ماتریس هنکلی که در SSA استفاده می‌شود آن است که رابطه یک به یکی با سری اصلی دارد. اما ماتریس-هایی که در مرحله‌ی گروه‌بندی به دست آمده است دارای خاصیت هنکلی نیستند و لذا نمی‌توان تقریبی یکتا از سری به دست آورد. روشی بهینه برای حل این مشکل استفاده از میانگین‌گیری قطری است (Golyandina et al., 2001). هدف از میانگین‌گیری قطری، تبدیل یک ماتریس به صورت ماتریس هنکل است که این ماتریس می‌تواند به سری زمانی برگردانده شود. یعنی ماتریس به دست آمده از رابطه (۸) را به سری جدید به طول  $N$  تبدیل می‌کند. اگر  $y_{ij}$  درآیه‌های ماتریس  $Y$  باشد.  $k$  امین جمله از سری حاصل با میانگین‌گیری  $y_{ij}$  روی همه  $i$  و  $j$  به دست می‌آید که  $i+j = k+1$  است. این رویه میانگین-گیری قطری یا هنکل‌سازی ماتریس  $Y$  نامیده می‌شود. نتیجه هنکل‌سازی ماتریس  $Y$  ماتریس هنکل  $HY$  است. با استفاده از رویه هنکل‌سازی روی همه مؤلفه‌های رابطه (۸)، بسط دیگری به صورت زیر (رابطه ۹) به دست می‌آید که  $k=1, \dots, m$  است.

$$X = \tilde{X}_{I_1} + \dots + \tilde{X}_{I_m}, \quad HX_{I_k} = \tilde{X}_{I_k} \quad (رابطه ۹)$$

یک گروه‌بندی خوب منجر به تجزیه رابطه (۸) می‌شود که در آن ماتریس حاصل  $X_{I_k}$  تقریباً همه هنکل هستند. با توجه به انتخاب مناسب  $r$  تا از  $m$  زیر سری، سری بازسازی شده از مرحله میانگین‌گیری قطری، یک سری با سطح نوفه اندک می‌باشد و لذا می‌توان از آن برای مدل‌سازی یا پیش‌بینی، که در بخش بعد به آن اشاره می‌گردد، استفاده کرد. بیان ریاضی مطالب شرح داده شده را می‌توان در زیر خلاصه کرد:

فرض کنیم که  $Y$  یک ماتریس  $L \times K$  با درآیه‌های  $y_{ij}$  که  $1 \leq i \leq L, 1 \leq j \leq K$  باشد. همچنین شرایط زیر نیز برقرار باشد:

$$N = L + K - 1, \quad K^* = \max(L, K), \quad L^* = \min(L, K)$$

اگر  $L < K$  باشد آنگاه  $y_{ij}^* = y_{ij}$  در غیر این صورت  $y_{ij}^* = y_{ji}$  می‌باشد. میانگین‌گیری قطری، ماتریس  $Y$  را به وسیله رابطه زیر به سری‌های  $g_1, \dots, g_n$  تبدیل می‌کند.

استفاده گردید. ابتدا از روش SSA برای پردازش اولیه داده‌ها و رفع نوفه موجود در سری‌های زمانی جریان رودخانه استفاده شد. سپس از الگوریتم بازگشتی روش SSA برای ساخت مدل پیش‌بینی آینده رودخانه در ایستگاه‌های حوضه کرخه استفاده گردید. برای پیش‌بینی جریان رودخانه با استفاده از نرم‌افزار R یک مدل کامپیوتری تهیه شد. مدل ارائه شده برای داده‌های دوره واسنجی و صحت‌سنجی به‌کاربرده شد و عملکرد آن نیز مورد بررسی قرار گرفت.

## یافته‌ها و بحث

### انتخاب طول پنجره در روش SSA

تنها پارامتر در مرحله تجزیه، طول پنجره  $L$ ، است. با انتخاب میانه داده‌ها برای مقدار  $L$ ، روش SSA جواب بهینه را می‌دهد (Hassani et al., 2011). بنابراین طول پنجره با توجه به تعداد داده‌های هر ایستگاه برابر با میانه انتخاب گردید که نتایج آن برای ایستگاه‌های مختلف حوضه کرخه در جدول (۲) ارائه شده است. علاوه بر آن برای بررسی بیشتر این مسئله مقادیر قدرمطلق  $w$ -همبستگی که نشان‌دهنده همبستگی بین مقادیر ویژه می‌باشد، به ازای طول پنجره‌های مختلف در جدول (۲) آورده شده است. به‌عنوان مثال برای ایستگاه چم انجیر چون تعداد داده‌های موردبررسی ۳۹ است، طول پنجره برابر  $L=20$  انتخاب گردید. همان‌طور که مشاهده می‌شود با انتخاب طول پنجره برابر با مقدار ۲۰، مقدار قدرمطلق  $w$ -همبستگی برای این طول پنجره کمترین مقدار را دارد. یعنی کمترین همبستگی بین سیگنال و نوفه در حالتی است که با طول پنجره ۲۰، سری زمانی را بازسازی کنیم.

### بررسی و تعیین نقطه برش و گروه‌بندی در روش SSA

با انتخاب نقطه برش برای یک سری زمانی می‌توان سیگنال را از نوفه تفکیک نمود. این کار به چند روش انجام می‌شود که در اینجا از ترکیب سه روش ۱- بررسی شکاف‌ها در طیف مقادیر ویژه ۲- بررسی ماتریس مقادیر قدر مطلق  $w$ -همبستگی در مقیاس ۲۰ درجه خاکستری از سفید تا سیاه ۳- بررسی نمودار مؤلفه اصلی ویژه سه گانه‌ها استفاده می‌گردد تا نقطه برش تعیین شود. سپس از نتایج آن در مرحله گروه‌بندی استفاده می‌گردد.

### بررسی شکاف‌ها در طیف مقادیر ویژه

سری نوفه معمولاً دنباله‌ای از مقادیر تکین، که به کندی کاهش می‌یابند، را تولید می‌کند. به بیانی دیگر در صورتی که مقادیر تکین به‌صورت نزولی مرتب و رسم شوند، می‌توان دو قطعه

که در آن  $\{p_1, \dots, p_r\}$  یک پایه متعامد یکه برای زیرفضای  $L_r$  است.  $p_i^T$  برداری است شامل  $L-1$  مؤلفه اول  $P_i$ ،  $\pi_i$  مؤلفه آخر بردار  $P_i$  و  $r$  مقداری است که در مرحله گروه‌بندی تعیین شده است. فرمول بازگشتی فوق در اصل پیش‌بینی سری زمانی تا  $M$  مرحله جلوتر است. نکته‌ای که در این شیوه حائز اهمیت است، عدم وابستگی پیش‌بینی به یک الگوی از پیش تعیین شده است. و تنها با استفاده از روابط موجود بین داده‌ها مقدار پیش‌بینی به دست می‌آید.

### معیارهای بررسی عملکرد مدل

#### ضریب همبستگی خطی

برای بیان همبستگی خطی بین مقادیر پیش‌بینی شده و مشاهده شده از ضریب همبستگی خطی استفاده می‌شود مقدار صفر این آماره برای عدم همبستگی و مقدار یک آن برای همبستگی کامل (برابری کامل مقادیر مشاهده شده و پیش‌بینی شده) می‌باشد. هر چقدر که مقدار این آماره به یک نزدیک‌تر باشد نشان‌دهنده این است که مدل از دقت بیشتری برای پیش‌بینی برخوردار است.

#### جذر میانگین مربعات خطای نرمال شده

این آماره برای سنجش دقت مدل استفاده می‌شود که از رابطه زیر به دست می‌آید:

$$NRMSE = \frac{1}{O_a} \sqrt{\frac{\sum_{i=1}^N (O-P)^2}{N}} \quad (\text{رابطه ۱۴})$$

در معادله فوق  $O$  و  $O_a$  به ترتیب داده مشاهده شده و دامنه آن،  $P$  داده پیش‌بینی شده،  $N$  تعداد داده‌ها و  $NRMSE$  جذر میانگین مربعات خطای نرمال شده است. مقدار  $NRMSE$  بدون بعد می‌باشد.

#### ث میانگین قدر مطلق خطای نسبی<sup>۱</sup> (MARE)

آماره متوسط خطای نسبی از رابطه زیر به دست می‌آید. مقدار کم این آماره نشان‌دهنده عملکرد بهتر مدل می‌باشد.

$$MARE = \frac{\sum_{i=1}^N \frac{|O-P|}{O}}{N} \quad (\text{رابطه ۱۵})$$

مدل‌سازی و پیش‌بینی بلندمدت (دوره پرآبی) جریان رودخانه مقادیر آینده ایستگاه‌های موردنظر در حوضه کرخه از سازمان مدیریت منابع آب ایران تهیه شد. برای این کار داده‌ها به دو دسته واسنجی (۷۰ درصد) و صحت‌سنجی (۳۰ درصد) تقسیم گردید.

برای مدل‌سازی و پیش‌بینی جریان رودخانه در حوضه کرخه به روش SSA از نرم‌افزار آماری CaterpillarSSA و R

درجهٔ خاکستری، از سفید تا سیاه متناظر با مقادیر قدرمطلق همبستگی‌ها (از صفر تا یک)، نشان می‌دهد. با توجه به شکل (۴) در ابتدای نمودار، مؤلفه اول با مؤلفه‌های دیگر همبستگی ندارد درحالی‌که مؤلفهٔ دوم و سوم با هم همبستگی دارند. از مؤلفه سوم به بعد همبستگی بین مؤلفه‌ها با رنگ خاکستری مشخص می‌شوند. بنابراین با توجه به اینکه  $w$ -همبستگی بین مؤلفه‌های ۱-۳ و سایر مؤلفه‌ها (که بین خود این‌ها همبستگی وجود دارد) تقریباً ناچیز است می‌توان مؤلفه‌های ۱-۳ را در یک گروه (سیگنال) و سایر مؤلفه‌ها را در گروه نوفه قرار داد. همچنین همبستگی زیادی بین مؤلفه‌های دوم و سوم است، بنابراین می‌توانند باهم بیانگر ویژگی دوره باشند. بر اساس این اطلاعات، سه ویژه سه‌گانه برای بازسازی سری اصلی انتخاب می‌کنیم و بقیه را به‌عنوان نوفه در نظر می‌گیریم. علاوه بر آن مؤلفه‌های ۸ و ۱۲ همبستگی کمی با بقیه مؤلفه‌ها دارند لذا می‌توان از آنها برای ترکیب در قسمت روند (همراه مؤلفه اول) برای تشخیص بهتر روند استفاده کرد. در قسمت بعد با استفاده از مؤلفه‌های اصلی می‌توان تشخیص داد که کدامیک برای ترکیب در روند بهتر است.

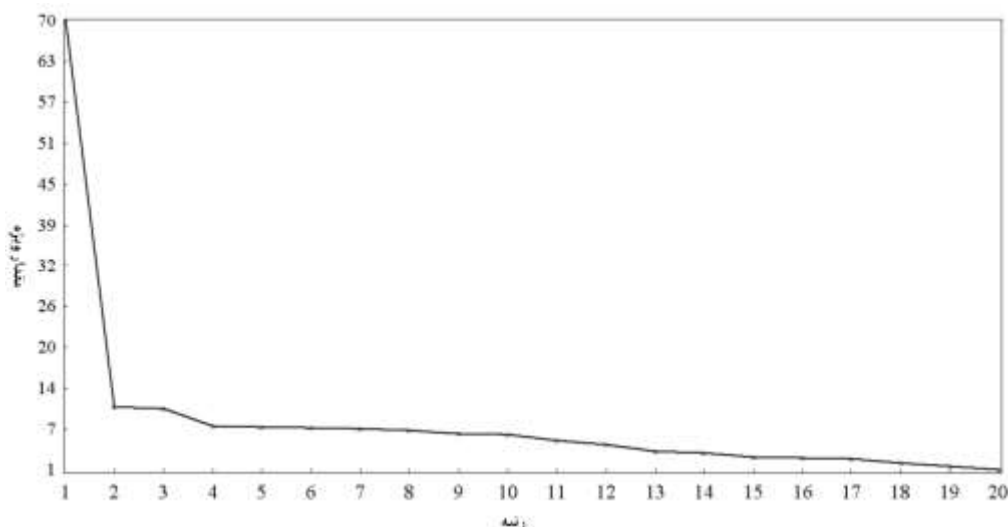
متفاوت را در شکل حاصل تشخیص داد. یکی با شیب تند که نمایشگر سیگنال است و دیگری بخش هموارتر که نمایشگر نوفه در سری‌های زمانی می‌باشد. به‌عنوان مثال برای ایستگاه چم-انجیر با انتخاب طول پنجره  $L=20$  و اجرای گام‌های اول و دوم (انجام SVD ی ماتریس مسیر  $X$ )، تعداد ۲۰ مقدار ویژه به‌دست می‌آید که بر اساس سهم خود در تجزیه مرتب شده‌اند. شکل (۳) نمودار نیمه لگاریتمی این ۲۰ مقادیر ویژه را نشان می‌دهد. با توجه به شکل (۳) افت قابل ملاحظه‌ای در مقادیر ویژه حول مؤلفهٔ ۳ (نقطه برش اولیه) به بعد اتفاق می‌افتد (منحنی هموارتر می‌گردد) که می‌توان آن را به‌عنوان شروع سطح نوفه تعبیر کرد. همچنین یک زوج آشکار (زوج ویژه سه‌گانه ۲-۳)، تقریباً با مقادیر تکین مساوی، متناظر با مؤلفهٔ فصلی سری، مربوط به دورهٔ معینی است که این موضوع در مراحل بعدی بیشتر بررسی خواهد شد.

بررسی ماتریس مقادیر قدر مطلق  $w$ -همبستگی

در ادامه مقادیر ماتریس قدرمطلق  $w$ -همبستگی‌ها بررسی می‌گردد. شکل (۴) مقادیر قدرمطلق  $w$ -همبستگی‌ها را برای ایستگاه چم انجیر و برای ۲۰ مؤلفهٔ بازسازی‌شده در مقیاس ۲۰

جدول ۲. مقادیر قدر مطلق  $w$ -همبستگی به ازای طول پنجره‌های مختلف برای ایستگاه‌های واقع در حوضه کرخه

ایستگاه	تعداد داده موردبررسی (N)							طول پنجره (L)			
	۵	۶	۱۰	۱۴	۱۵	۱۸	۱۹	۲۰			
جلوگیر	۰/۱۰۱	۰/۰۹۹	۰/۱۲۴	۰/۰۶۵	۰/۰۶۲	۰/۰۵۳	۰/۰۴۷	-			
پل دختر	۰/۱۵۱	۰/۱۳۸	۰/۱۰۵	۰/۰۷۳	۰/۰۷۲	۰/۰۴۴	۰/۰۴۳	۰/۰۴۱			
قورباغستان	۰/۰۹۵	۰/۰۸۲	۰/۰۸۵	۰/۰۶۵	۰/۰۶۳	۰/۰۴۸	۰/۰۵۴	۰/۰۴۴			
چم‌انجیر	۰/۰۷۲	۰/۰۹۸	۰/۰۷۳	۰/۰۴۱	۰/۰۲۹	۰/۰۱۴۵	۰/۰۱۴۹	۰/۰۱۲			
پل چهر	۰/۰۲۳	۰/۰۱۷	۰/۰۱۴	۰/۰۰۷	۰/۰۰۶	۰/۰۰۵	۰/۰۰۳	۰/۰۰۲			



شکل ۳. نمودار نیمه لگاریتمی ۲۰ مقدار ویژه برای ایستگاه چم‌انجیر

بررسی نمودار مؤلفه اصلی ویژه سه گانه‌ها

برای نهایی شدن تعیین نقطه برش، نمودار مؤلفه‌های اصلی (که به صورت سری زمانی نشان داده شده است) ۹ ویژه سه گانه (ویژه سه گانه‌های ۱-۹) به همراه ویژه سه گانه ۱۲ برای ایستگاه چمنجیر در شکل (۵) نشان داده شده است.

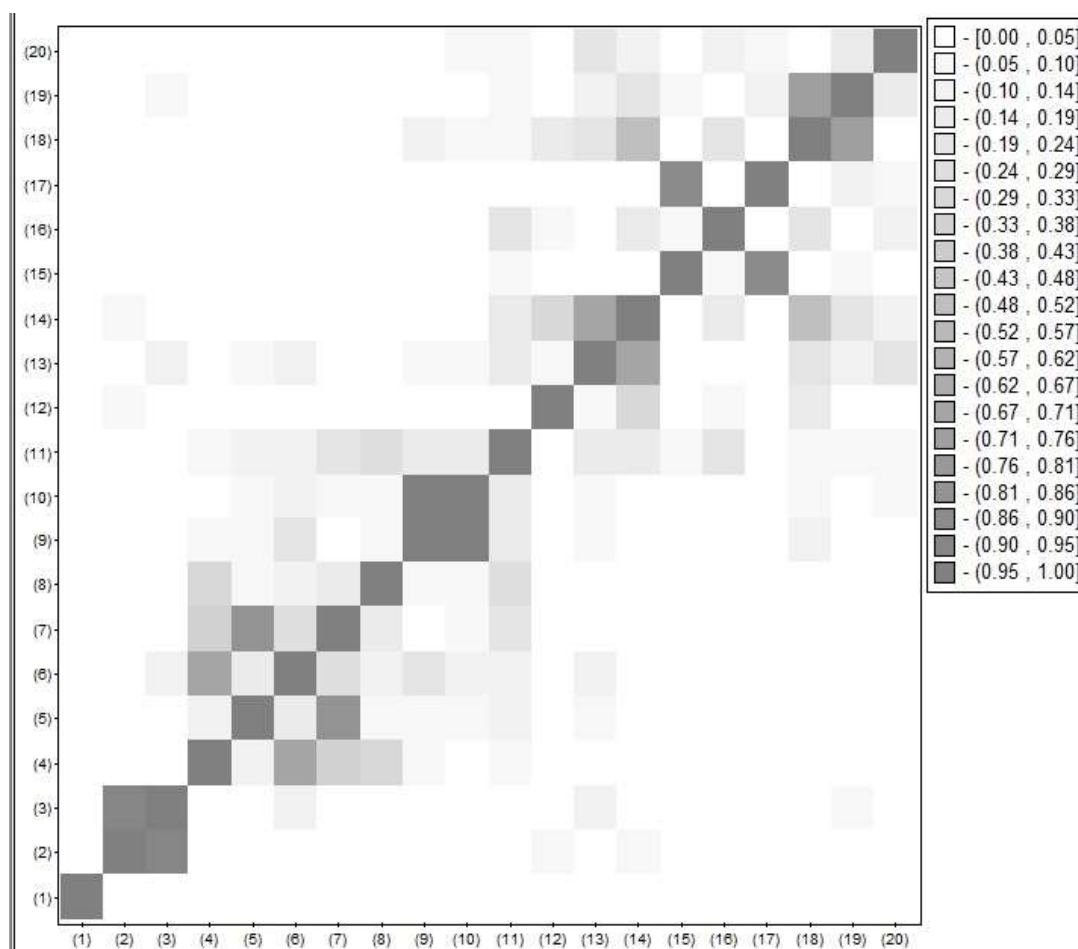
در اینجا دو رده وجود دارد:

۱. مؤلفه‌هایی که در رده سیگنال در نظر گرفته می‌شوند که تغییرات آرامی دارند.

۲. مؤلفه‌هایی که در رده نوفه قرار می‌گیرند که تغییرات زیادی دارند.

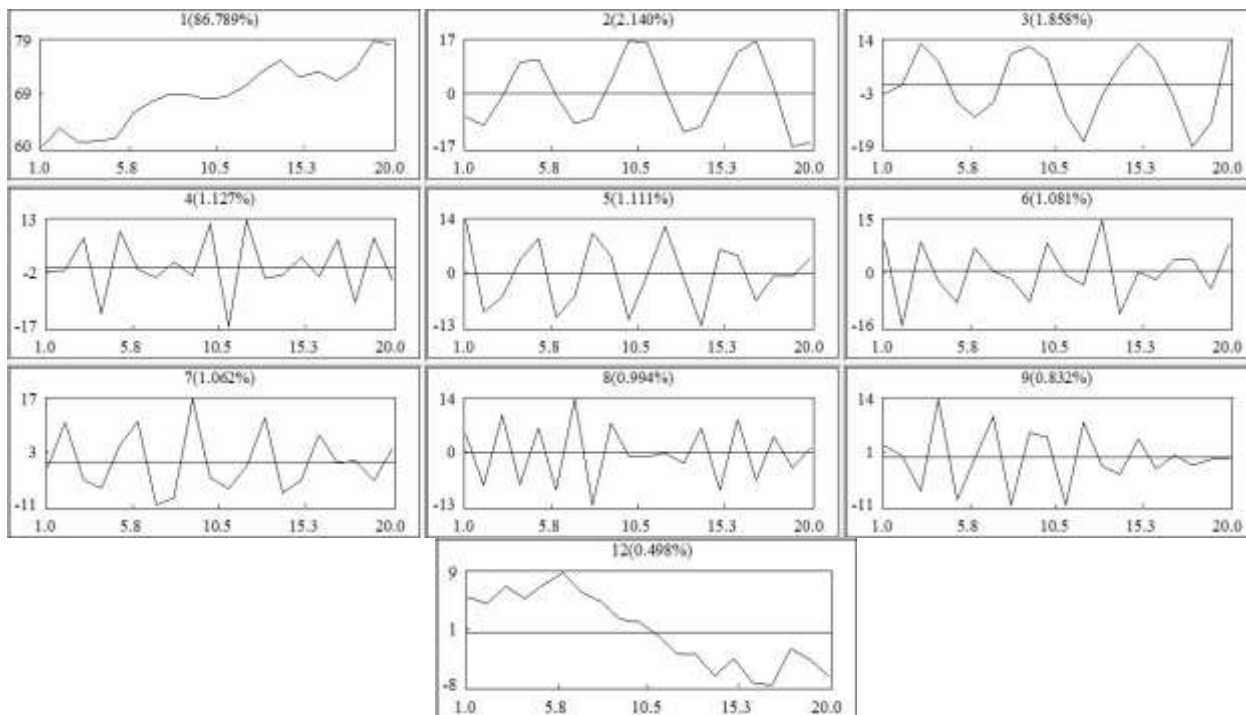
بنابراین واضح است که ویژه سه گانه ۱، ۲، ۳ و ۱۲ می‌تواند به عنوان قسمت سیگنال در نظر گرفته شود در حالی که بقیه ویژه سه گانه‌ها به رده نوفه تعلق دارند. بنابراین ویژه سه گانه ۱ روند را نشان می‌دهد و ویژه سه گانه ۱۲ نیز می‌تواند در بهتر نشان داده شدن روند به کار رود. ویژه سه گانه ۲-۳ دوره را نشان می‌دهد. همچنین مشخص است که مؤلفه ۸ نوفه است.

شکل (۶-الف) روند استخراج شده را که از ویژه سه گانه ۱ و ۱۲ به دست آمده است به همراه منحنی سری زمانی اولیه برای ایستگاه چمنجیر را نشان می‌دهد. همان طوری که مشخص است روند تقریباً صعودی است. شکل (۶-ب) مؤلفه فصلی به دست آمده از ویژه سه گانه‌های ۲ و ۳ به همراه منحنی سری زمانی اولیه برای ایستگاه چمنجیر را نشان می‌دهد. واضح است که شکل (۶-ب) الگوی یکسانی از دوره‌ای بودن را در سری زمانی اصلی نشان می‌دهد که با گذشت زمان این دوره مشخص تر می‌شود. بنابراین برای بازسازی سیگنال بهتر است که از ویژه سه گانه‌های ۱-۳ به همراه ویژه سه گانه ۱۲ استفاده شود و بقیه مؤلفه‌های ویژه سه گانه را می‌توان به عنوان نوفه در نظر گرفت. بنابراین برای ایستگاه چمنجیر محل نقطه برش مقدار ویژه ۳ در نظر گرفته می‌شود. برای تعیین محل نقطه برش برای سری زمانی آینده سایر ایستگاه‌های حوضه کرخه باید همین روش را انجام داد. این کار برای سایر ایستگاه‌ها انجام گردید و نتایج آن در جدول (۳ الف) ارائه شده است.

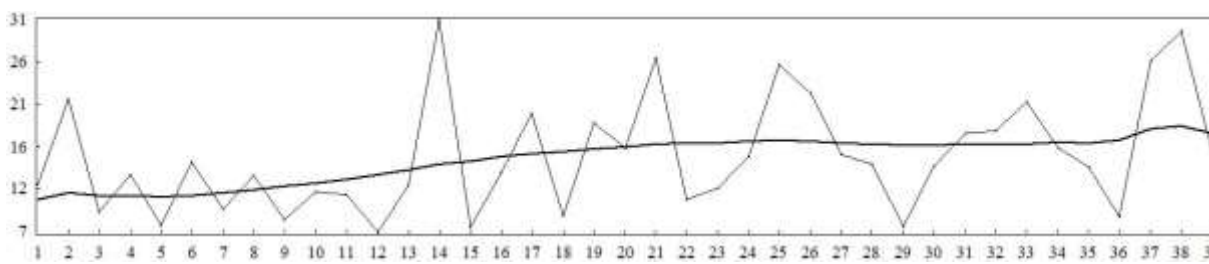


شکل ۴. مقادیر ماتریس قدرمطلق W-همبستگی در مقیاس ۲۰ درجه خاکستری برای ۲۰ مؤلفه بازسازی شده برای ایستگاه چمنجیر.

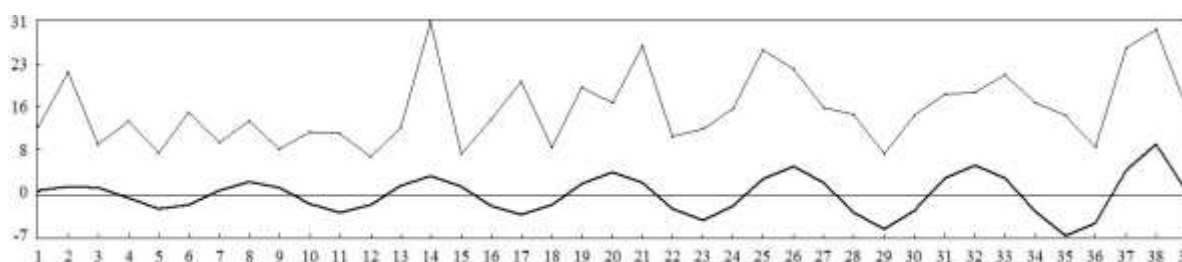




شکل ۵. مقادیر مؤلفه‌های اصلی ۹ ویژه سه‌گانه (ویژه سه‌گانه ۹-۱) به همراه ویژه سه‌گانه ۱۲، محور xها مکان مشاهده (۱ تا ۲۰ طول پنجره) و محور yها مقادیر مؤلفه اصلی برای ایستگاه چمانجیر.



شکل ۶-الف. منحنی سری زمانی اولیه و روند استخراج‌شده از ویژه سه‌گانه ۱ برای ایستگاه چمانجیر، محور افقی سال و محور عمودی آبدهی mcm.



شکل ۶-ب. منحنی سری زمانی اولیه و مؤلفه فصلی استخراج‌شده از ویژه سه‌گانه‌های ۲ و ۳ برای ایستگاه چمانجیر، محور افقی سال و محور عمودی آبدهی

بررسی سری بازسازی شده (با ویژه سه‌گانه ۳-۱ به همراه ۱۲) و باقی‌مانده‌ها (با ویژه سه‌گانه‌های ۲۰-۴ به جز ۱۲) برابر ۰/۰۱۲ است که کمترین مقدار آن به‌ازای نقاط برش مختلف می‌باشد یعنی کمترین همبستگی بین سیگنال و نوفه حالتی است که از ۳ مقدار ویژه (به‌همراه ۱۲) برای بازسازی سری استفاده کنیم. این مسئله مؤید این مطلب است که گروه‌بندی قابل قبول است. همین روش برای تعیین گروه‌بندی سایر ایستگاه‌ها استفاده گردید که نتایج آن در جدول (۳ ب) ارائه گردیده است.

بعد از تعیین نقطه برش برای بررسی بیشتر فرایند گروه‌بندی ویژه سه‌گانه‌ها از مقادیر عددی قدرمطلق  $w$ -همبستگی بین سری بازسازی شده و باقی‌مانده‌ها استفاده گردید. مقادیر قدرمطلق  $w$ -همبستگی بین سری بازسازی شده و باقی‌مانده‌ها برای ایستگاه‌های مختلف در حوضه کرخ به‌ازای مقادیر مختلف نقطه برش در جدول (۳ ب) ارائه گردیده است. به‌عنوان مثال برای ایستگاه چمانجیر مقادیر عددی قدرمطلق  $w$ -همبستگی

## پیش‌بینی بلندمدت (دوره پراپی) جریان رودخانه

### بررسی عملکرد مدل پیش‌بینی جریان

برای تعیین مقادیر طول پنجره و نقطه برش و گروه‌بندی مطابق روش توضیح داده‌شده در قسمت قبل عمل گردید. مدل ارائه‌شده برای داده‌های دوره واسنجی و صحت‌سنجی به‌کاربرده شد و عملکرد آن نیز مورد بررسی قرار گرفت که به شرح زیر می‌باشد.

مدل ارائه‌شده برای پیش‌بینی جریان رودخانه در حوضه کرخه برای ایستگاه‌های موردنظر شامل جلوگیر، پل‌دختر، قورباغستان، چم‌انجیر و پل‌چهر (حرف اول نام ایستگاه‌ها به لاتین، JPGCP) مورد استفاده قرار گرفت. برای بررسی عملکرد مدل از معیارهای ضریب همبستگی خطی، جذر متوسط مربعات خطای نرمال شده (NRMSE) و متوسط خطای نسبی (MARE) استفاده گردید؛ که نتایج مرحله واسنجی و صحت‌سنجی به ترتیب در جدول (۴ الف و ۴ ب) ارائه شده است.

در مرحله صحت‌سنجی با استفاده از مقدار طول پنجره و نقطه برش مشخص‌شده در مرحله واسنجی، مدل موردنظر اجرا گردید. در مرحله واسنجی که نتایج آن در جدول (۴ الف) ارائه شده است، مقادیر آماره NRMSE برای ایستگاه‌های JPGCP به‌ترتیب برابر ۰/۳۵، ۰/۳۲، ۰/۳۳، ۰/۳ و ۰/۴۳ بود. بیشترین و کمترین مقدار این آماره به‌ترتیب برای ایستگاه‌های پل‌چهر و چم‌انجیر به‌دست آمد. مقدار میانگین قدر مطلق خطای نسبی MARE برای ایستگاه‌های JPGCP به‌ترتیب برابر ۰/۳، ۰/۳، ۰/۳، ۰/۲۷ و ۰/۳۵ می‌باشد. بیشترین و کمترین مقدار آن به‌ترتیب برای ایستگاه‌های پل‌چهر و چم‌انجیر به‌دست آمد. مقادیر ضریب همبستگی برای ایستگاه‌های JPGCP به‌ترتیب برابر ۰/۶۳، ۰/۷۲، ۰/۶۵، ۰/۶۹ و ۰/۵۸ به‌دست آمد؛ که کمترین مقدار آن برای ایستگاه پل‌چهر و بیشترین مقدار آن برای ایستگاه پل‌دختر بود. نتیجه اجرای مدل برای مرحله واسنجی برای تمامی ایستگاه‌ها به‌جز پل‌چهر در یک محدوده است و اختلاف زیادی با هم ندارد.

در مرحله صحت‌سنجی که نتایج آن در جدول (۴ ب) ارائه شده است، مقادیر آماره NRMSE برای ایستگاه‌های JPGCP به‌ترتیب برابر ۰/۳۵، ۰/۳، ۰/۴۴، ۰/۳۱ و ۰/۴۷ بود. بیشترین و کمترین مقدار این آماره به‌ترتیب برای ایستگاه‌های پل‌چهر و پل‌دختر به‌دست آمد. مقدار میانگین قدرمطلق

خطای نسبی MARE برای ایستگاه‌های JPGCP به‌ترتیب برابر ۰/۳۸، ۰/۳، ۰/۴، ۰/۲۹ و ۰/۵ می‌باشد. بیشترین و کمترین

مقدار آن به‌ترتیب برای ایستگاه‌های پل‌چهر و پل‌دختر به‌دست آمد. مقادیر ضریب همبستگی برای ایستگاه‌های JPGCP به‌ترتیب برابر ۰/۵۵، ۰/۶۱، ۰/۵۸، ۰/۶۳ و ۰/۵۱ به‌دست آمد؛ که کمترین و بیشترین مقدار آن به‌ترتیب برای ایستگاه پل‌چهر و چم‌انجیر بود مقدار ضریب همبستگی برای ایستگاه چم‌انجیر و پل‌دختر نزدیک به هم به‌دست آمد. نتیجه اجرای مدل برای مرحله صحت‌سنجی برای ایستگاه‌های چم‌انجیر و پل‌دختر نسبت به بقیه ایستگاه‌ها بهتر بود. شکل (۷) مقادیر جریان رودخانه مشاهده‌ای و پیش‌بینی‌شده نسبت به خط یک به یک در مرحله واسنجی برای ایستگاه چم‌انجیر را نشان می‌دهد. شکل (۸) همین نمودار را برای ایستگاه چم‌انجیر و در مرحله صحت‌سنجی را نشان می‌دهد. شکل (۹) نیز مقادیر جریان رودخانه مشاهده‌ای و پیش‌بینی‌شده در دو مرحله واسنجی و صحت‌سنجی را برای ایستگاه چم‌انجیر در طول دوره آماری را نشان می‌دهد و ملاحظه می‌گردد که مقادیر جریان مشاهده‌ای و پیش‌بینی‌شده با یکدیگر مطابقت دارند.

جدول ۳ الف. مقادیر نقطه برش و مقادیر ویژه برای بازسازی سیگنال برای ایستگاه‌های حوضه کرخه بر اساس مقادیر مؤلفه اصلی.

ایستگاه	شماره ویژه سه‌گانه برای بازسازی سیگنال	نقطه برش
جلوگیر	۱، ۲ و ۳	۳
پل‌دختر	۱، ۲ و ۳	۳
قورباغستان	۱، ۲ و ۸	۲
پل‌چهر	۱ و ۱۱	۱
چم‌انجیر	۱، ۲، ۳ و ۱۲	۳

جدول ۳ ب. مقادیر W همبستگی برای تعیین نقطه برش برای ایستگاه‌های حوضه کرخه

ایستگاه	شماره ویژه مقدار برای انتخاب نقطه برش				
	۱	۲	۳	۴	۵
جلوگیر	۰/۴۷	۰/۴۳	۰/۳۸	۰/۷۲	۰/۸۰
پل‌دختر	۰/۴۱	۰/۳۲	۰/۳۲	۰/۶۶	۰/۸۶
قورباغستان	-	۰/۴۴	۰/۴۷	۰/۵۹	۰/۴۹
پل‌چهر	۰/۰۳۳	۰/۴۴	۰/۶۹	۰/۶۰	۰/۲۹
چم‌انجیر	-	۰/۳۵	۰/۱۲	۰/۳۹	۰/۶۶

جدول ۴ الف. نتایج اجرای مدل برای مرحله واسنجی واسنجی

معیار ارزیابی	جلوگیر	پل‌دختر	قورباغستان	چم‌انجیر	پل‌چهر
NRMSE	۰/۳۵	۰/۳۲	۰/۳۳	۰/۳	۰/۴۳
MARE	۰/۳	۰/۳	۰/۳	۰/۲۷	۰/۳۵
R	۰/۶۳	۰/۷۲	۰/۶۵	۰/۶۹	۰/۵۸

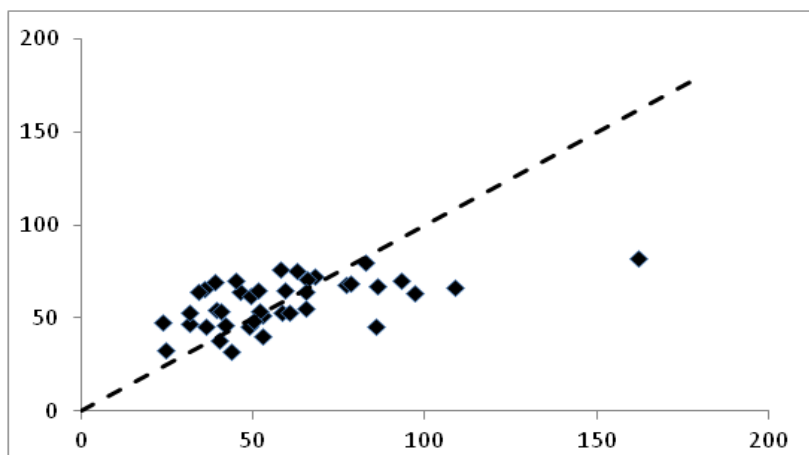
مخزن سد شاهچراغی استفاده نمود و گزارش کرد که مقادیر RMSE، R و MARE برای بهترین مدل شبکه عصبی دینامیک با داده‌های آینده به ترتیب برابر ۰/۴، ۰/۳۵ و ۰/۱۷ بود. Akbarinia (2012) یک مدل شبکه عصبی برای پیش‌بینی مقادیر آینده ماهانه (فروردین ماه) رودخانه کرخه ارائه کرد. به عنوان مثال مقادیر RMSE، R و MARE برای ایستگاه پل دختر به ترتیب برابر ۰/۳۶، ۰/۷۴/۴ و ۵۱/۹ بود. همین آماره‌ها برای ایستگاه قورباغستان به ترتیب برابر ۵۰/۹، ۰/۶۲ و ۵۱/۷ به دست آمد. از میان موارد ذکر شده (Jamali (2009) و Akbarinia (2012) مدل پیش‌بینی آینده ماهانه را ارائه کردند و Meidani (2012) مدل پیش‌بینی آینده برای دوره پربابی (۱۰ بهمن تا ۳۱ خرداد) را ارائه داد. همواره انتظار می‌رود که هر چقدر که افق پیش‌بینی افزایش یابد دقت مدل نیز کاهش یابد. با مقایسه نتایج این تحقیق با سایرین می‌توان گفت که با به‌کارگیری روش SSA می‌توان مقادیر جریان رودخانه را با دقت مناسب پیش‌بینی نمود.

جدول ۴ ب. نتایج اجرای مدل برای مرحله صحت‌سنجی

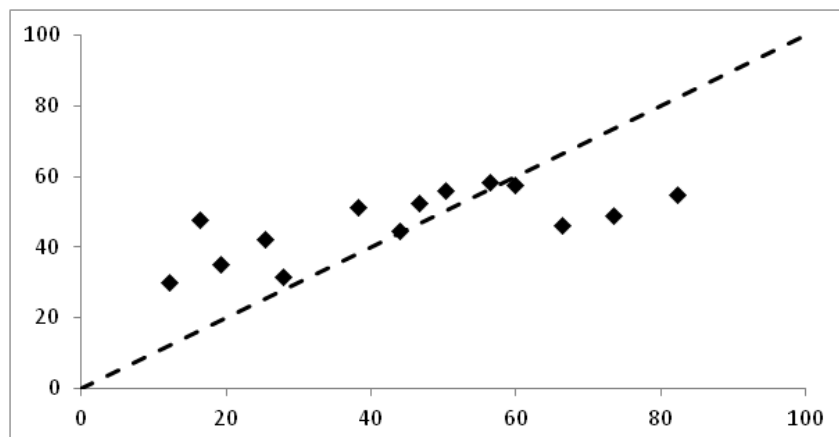
معیار ارزیابی	صحت‌سنجی				
	نام ایستگاه				
NRMSE	پل‌چهر	چم‌انجیر	قورباغستان	پل‌دختر	جلوگیر
MARE	۰/۴۷	۰/۳۱	۰/۴۴	۰/۳	۰/۳۵
R	۰/۵	۰/۲۹	۰/۴	۰/۳	۰/۳۸
	۰/۵۱	۰/۶۳	۰/۵۸	۰/۶۱	۰/۵۵

#### مقایسه نتایج این تحقیق با نتایج دیگران

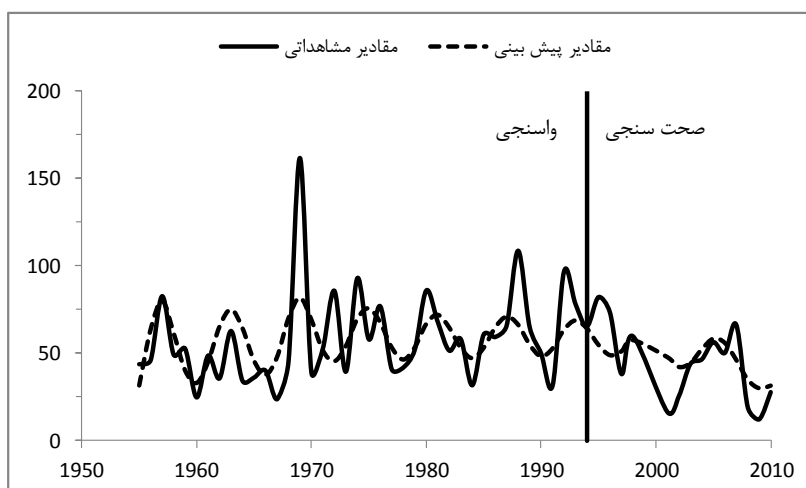
Meidani (2012) از مدل K نزدیک‌ترین همسایه با ورودی شاخص اقلیمی و آینده برای پیش‌بینی جریان رودخانه‌های کارون و دز استفاده نمود. برای ارزیابی دقت مدل از آماره‌های جذر میانگین مربعات خطا (RMSE) و ضریب همبستگی (R) استفاده نمود. او گزارش کرد که مقادیر RMSE برای کارون و دز به ترتیب برابر ۱۸۸/۷۳ و ۱۷۰/۸۲ و مقادیر R برای کارون و دز به ترتیب برابر ۰/۳۵ و ۰/۳۸ بود. Jamali (2009) یک مدل شبکه عصبی دینامیک برای پیش‌بینی جریان ماهانه ورودی به



شکل ۷. نمودار مقایسه داده‌های مشاهده‌ای با داده‌های پیش‌بینی در مرحله واسنجی برای ایستگاه چم‌انجیر



شکل ۸. نمودار مقایسه داده‌های مشاهده‌ای با داده‌های پیش‌بینی در مرحله صحت‌سنجی برای ایستگاه چم‌انجیر



شکل ۹. نمودار مقایسه داده‌های مشاهده‌ای با داده‌های پیش‌بینی در دو مرحله واسنجی و صحت‌سنجی برای ایستگاه چم‌انجیر

## نتیجه‌گیری

در این مقاله از روش SSA برای مدل‌سازی سری‌های زمانی آبدهی رودخانه‌های حوضه کرخه در دوره پربابی استفاده شد. برای مدل‌سازی به روش SSA از نرم‌افزار آماری CaterpillarSSA و R استفاده گردید. طول پنجره با توجه به تعداد داده‌های هر ایستگاه برابر با میانه انتخاب گردید. علاوه بر آن برای بررسی بیشتر این مسئله مقادیر قدرمطلق  $w$ -همبستگی که نشان‌دهنده همبستگی بین مقادیر ویژه می‌باشد، به ازای طول پنجره‌های مختلف محاسبه گردید که مشخص گردید کمترین همبستگی بین سیگنال و نوفه در حالتی است که طول پنجره با توجه به تعداد داده‌های هر ایستگاه برابر با میانه باشد. برای تعیین محل دقیق نقطه برش برای سری زمانی آبدهی، از سه روش ۱-بررسی شکاف‌ها در طیف مقادیر ویژه ۲-بررسی ماتریس مقادیر قدر مطلق  $w$ -همبستگی در مقیاس ۲۰ درجه خاکستری از سفید تا سیاه ۳-بررسی نمودار مؤلفه اصلی ویژه سه‌گانه‌ها، استفاده گردید و از نتایج آن در مرحله گروه-بندی استفاده شد. بعد از تعیین نقطه برش برای بررسی بیشتر فرایند گروه‌بندی ویژه سه‌گانه‌ها، از مقادیر عددی قدرمطلق  $w$ -همبستگی بین سری بازسازی شده و باقی‌مانده‌ها استفاده گردید و مشاهده گردید که کمترین مقدار آن به‌ازای نقاط برش محاسبه شده می‌باشد که مؤید این مطلب است که گروه‌بندی

قابل قبول بوده است. پس از آنکه از روش SSA برای پردازش اولیه داده‌ها و رفع نوفه موجود در سری‌های زمانی جریان رودخانه استفاده گردید. از الگوریتم بازگشتی روش SSA برای ساخت مدل پیش‌بینی آبدهی رودخانه در ایستگاه‌های حوضه کرخه استفاده شد. مدل ارائه شده برای داده‌های دوره واسنجی و صحت‌سنجی به‌کاربرده شد و عملکرد آن نیز بررسی گردید. بر این اساس در مرحله واسنجی بیشترین و کمترین مقدار آماره NRMSE به ترتیب برابر ۰/۴۳ و ۰/۳۰ (همین‌طور آماره MARE به ترتیب برابر ۰/۳۶ و ۰/۲۷) برای ایستگاه‌های پل‌چهر و چم‌انجیر به‌دست آمد. در مرحله صحت‌سنجی بیشترین مقدار آماره NRMSE و MARE به ترتیب برابر ۰/۴۷ و ۰/۵ برای ایستگاه‌های پل‌دختر و چم‌انجیر نزدیک به هم و برابر ۰/۳ و ۰/۳۱ و کمترین مقدار آماره MARE برای ایستگاه چم‌انجیر و پل‌دختر، نزدیک به هم و برابر ۰/۲۹ و ۰/۳۰ بود. در نهایت بهترین و ضعیف‌ترین نتیجه در دو مرحله واسنجی و صحت‌سنجی به-ترتیب برای ایستگاه‌های چم‌انجیر و پل‌چهر به‌دست آمد. در نهایت نتایج این مدل با نتایج سایر محققین مقایسه گردید. بر اساس نتایج این تحقیق می‌توان با به‌کارگیری روش SSA مقادیر جریان رودخانه را با دقت مناسب پیش‌بینی نمود.

## REFERENCES

- Akbarinia, M. (2012). Long Lead Stream flow Forecasting using data-driven models case study: Karkheh river MSc. thesis, *Irrigation and reclamation engineering group*, Tehran University, Karaj. (In Farsi)
- Basilevsky, A., Derek, P., and Hum, J. (1979). Karhunen-Loeve analysis of historical time series with an application to plantation births in Jamaica. *Journal of the American Statistical Associations*, 74: 284-290.
- Broomhead, D. S., and King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D*, 20: 217-236.
- Broomhead, D. S., King, G. P., and Pike, E. R. (1987). Singular spectrum analysis with application to dynamical systems. *Noise and Fractal, IOP*

*Publication, Bristol.*

- Danilov, D. (1997). Principal components in time series forecast. *Journal of Computational and Graphical Statistics*, 6:112-121.
- Golyandina, N., Nekrutkin, V., and Zhiglovsky, A. (2001). Analysis of time series structure: SSA and related techniques. *Chapman & Hall/CRC*.
- Hajibigloo, M., Ghezelsifloo, A. A., and Alimirzaei, H. (2013). Discussion and forecast monthly average rainfall techniques using SARIMA (case study: pluviometry station Babaaman Bojnourd). *Journal of Irrigation Science and Engineering*, 36 (3): 41-54. (In Farsi)
- Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5(2007): 239-257.
- Hassani, H., Mahmoudvand, R., and Yarmohammadi, M. (2010). Filtering and denoising in linear regression analysis. *Fluctuation and Noise Letters*, 9 (4): 343-358.
- Hassani, H. Mahmoudvand, R. and Zokaei, M. (2011). Separability and window length in singular spectrum analysis. *C. R. Acad. Sci. Paris, Ser. I*, 349: 987-990.
- Hassani, H., and Thomakos, D. (2010). A review on singular spectrum analysis for economic and financial time series, *Statistics and Its Interface*. 3(3): 377-397.
- Jamab Consulting Engineers. (2006). Water balance report of Karkheh River basin area: Preliminary analysis, *Ministry of Energy*, Tehran. Iran. (In Farsi)
- Jamali, F. S., (2009). An artificial neural network model for Shahcheraghi reservoir inflow forecasting using snow cover area data. MSc. thesis, *Irrigation and drainage engineering group*, Tehran University, Pakdasht. (In Farsi)
- Kumaresan, R., and Tufts, D. W. (1980). Data-adaptive principal component signal processing. *In Proc. of the 19th IEEE Conf. on Decision and Control*, 19: 949-954.
- Lisi, F. Nicolis, O., and Sandri, M. (1995). Combining singular-spectrum analysis and neural networks for time series forecasting. *Neural Processing Letters*, 2 (4): 6-10.
- Marques, C. A. F., Ferreira, J. A., Rocha, A., Castanheira, J. M., Melo-Goncalves, P., Vaz, N., and Dias, J. M. (2006). Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth*, 31:1172-1179.
- Meidani, E. (2012). Long lead streamflow forecasting using statistical methods: case study of Karoon and Dez rivers. MSc. thesis, *Irrigation and reclamation engineering group*, Tehran University, Karaj. (In Farsi)
- Menezes, M. L., Souza, R. C., and Moreira Pessanha, J. F. (2015). Electricity consumption forecasting using singular spectrum analysis, *Dyna. rev. fac. nac. minas*, 82 (190): 138-146.
- Sivapragasam, C., Liong, S. Y., and Pasha, M. F. K. (2001). Rainfall and runoff forecasting with SSA-SVM approach. *Journal of Hydroinformatics* 3 (7): 141-152.
- Vautard, R., and M. Ghil. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, 35: 395-424.
- Wu, C. L., Chau, K. W., and Fan, C. (2010). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, 389: 146-167.