

Two Non-Parametric Trend Tests Applicable for Long-Memory Processes (Case Study: Mashhad Air Temperature)

B. Ghahraman^{*}

Abstract

Trend investigation is one of the key characteristics of time series. However, all common trend tests (e.g. Kendall and Mann-Kendall) are based on stationary time series and with this assumption that there is no long-memory in the time series. Based on both classical Kendall and Mann-Kendall tests, it was shown in this study that there was a significant (p -value <0.001) increase in annual Mashhad temperature (with a record length of 127 years spanning 1885-2011). However, considering the time series as a long-memory process (fractional Gaussian normal with Hurst exponent of 0.92), standard deviation of the test statistics were increased by a factor of greater than 6. As a result, the increasing trend in temperature was not accepted at common significant levels of 0.01 and 0.05. Some regression equations were developed for correction of standard deviations of Kendall and Mann-Kendall trend tests as a function of record length and Hurst exponent, for the first time. The results remained identical for incomplete time series. A stochastic method, based on frequency factor of Chow, was proposed for data filling and showed that the results are stable.

Keywords: Hurst exponent, Kendall test, Mann-Kendall tests, Stochastic generation.

Received: October 13, 2012

Accepted: July 7, 2013

استفاده از دو آزمون ناپارامتریک برای تشخیص روند در یک سری زمانی دارای حافظه (مطالعه موردی: درجه حرارت سالانه مشهد)

بیژن قهرمان^{*۱}

چکیده

بررسی روند زمانی در یک سری زمانی یکی از ویژگی‌های مهم آن به‌شمار می‌آید. با این حال تمامی آزمون‌های متداول تعیین روند (مثلاً کندال و من-کندال) بر اساس فرض ایستا بودن سری زمانی و حافظه‌دار نبودن آن بنا شده است. هر دو آزمون کندال و من-کندال در حالت کلاسیک نشان دادند که سری زمانی درجه حرارت سالانه مشهد (به طول ۱۲۷ سال از ۱۸۸۵ تا ۲۰۱۱) دارای روند افزایشی معنی‌دار (p -مقدار کوچک تر از ۰/۰۰۱) است در حالی که با توجه به مفهوم حافظه بلند مدت در داده‌ها (فرآیند اغتشاش نرمال جزئی؛ FGN با نمایه هرست برابر با ۰/۹۲)، انحراف معیار آماره‌های آزمون افزون بر ۶ برابر بیش‌تر شد و در نتیجه هیچ‌کدام از این دو آزمون معنی‌داری روند افزایشی را در سطوح معنی‌داری متداول ۰/۰۱ و ۰/۰۵ تأیید نمودند. روابط رگرسیونی برای تصحیح انحراف معیار در دو آزمون ناپارامتری روند کندال و من-کندال به‌عنوان تابعی از نمایه هرست و طول دوره آماری برای اولین مرتبه به‌دست آمد. نشان داده شد که نتایج در شرایط عدم تکمیل داده‌ها یکسان باقی می‌ماند. روش جدید استوکاستیکی بر پایه مفهوم عامل فراوانی چاو پیشنهاد شد و نشان داده شد که نتایج پایدار باقی می‌ماند.

کلمات کلیدی: آزمون کندال، آزمون من-کندال، شبیه‌سازی استوکاستیکی، نمایه هرست.

تاریخ دریافت مقاله: ۲۲ مهر ۱۳۹۱

تاریخ پذیرش مقاله: ۱۶ تیر ۱۳۹۲

1- Professor of Irrigation, College of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran. Email: bijan.ghahraman@yahoo.com

*- Corresponding Author

۱- استاد دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران

*- نویسنده مسئول

برآورد انحراف معیار آماره‌های آزمون‌های روند در سری‌های زمانی دارای حافظه ارائه نشده است. هدف این مقاله استخراج و ارائه چنین روابطی برای دو آزمون ناپارامتریک کندال و من-کندال می‌باشد.

روش‌های استوکاستیک به‌طور کل و روش‌های غیرپارامتریک مورد اشاره به داده‌های طولانی مدت نیاز دارند. پیشینه آماربرداری در ایران طولانی نیست ولی تنها ۵ ایستگاه هواشناسی در ایران وجود دارد (مشهد، تهران، اصفهان، بوشهر و جاسک) که از پیش از جنگ جهانی دوم نیز دارای داده‌های درجه حرارت و باران در مقیاس ماهانه می‌باشد. برای انجام این پژوهش از داده‌های درجه حرارت ایستگاه مشهد استفاده شد ولی خلاءهای آماری در داده‌های این ایستگاه وجود دارد. استفاده از روش‌های همبستگی برای تکمیل داده‌ها کاملاً متداول است ولی بیش‌تر این خلاءها مربوط به سال‌های پیش از ۱۹۵۰ است که در آن سال‌ها ایستگاه‌های دارای آمار و در فاصله کم نسبت به ایستگاه مشهد وجود ندارد. باین‌حال در چنین شرایطی امکان استفاده از ایستگاه‌های با فاصله بیش‌تر مطرح می‌شود. مثلاً (Cornes (2010) از همبستگی درجه حرارت در ایستگاه‌های پاریس، لندن و برلین استفاده کرد. در برخی از مواقع در برقراری چنین همبستگی‌هایی به‌جای استفاده از متغیرهای همسان، از پدیده‌های بزرگ مقیاس اقلیمی استفاده می‌شود. (Katz et al., 2003; Lintner and Chiang, 2005; Roozbahani and Zahraie, 2005). در برخی مواقع نیز برای تکمیل خلاءها از ویژگی‌های داخلی سری زمانی استفاده می‌شود. (Ghahraman and Ahmadi (2007) داده‌های کمبود بارندگی در ایستگاه مشهد را با روش‌های زمین آماری و برازش چند جمله‌ای بر میانگین‌های متحرک و خلیلی و بذرافشان (۱۳۸۷) آن‌ها را با روش خودهمبستگی تکمیل کردند.

۲- مواد و روش‌ها

درجه حرارت ماهانه مشهد در سال‌های پیش از ۱۹۴۱ در محل سفارت‌خانه سابق آمریکا در مشهد و در سال‌های پس از ۱۹۵۰ در محل ایستگاه همدید واقع در فرودگاه مشهد اندازه‌گیری می‌شود. داده‌ها از سه منبع شامل سایت Word Weather Record، سایت کاوش گر اقلیمی KNMI در هلند و آمارنامه سازمان هواشناسی کشور در بازه زمانی ۱۸۸۵ تا ۲۰۱۱ می‌باشد. کمبودهای داده‌ها به‌وسیله روش‌های متعارف همبستگی با سایر ایستگاه‌ها (مثلاً در ترکمنستان و قزاقستان) و روش استفاده از ویژگی‌های سری زمانی (Ghahraman and Ahmadi, 2007) تکمیل گردید. در نگرشی دیگر، خلاءهای آماری توسط روش استوکاستیکی تکمیل گردید. این روش بر پایه رابطه فراوانی (Chow, 1954) و به‌صورت

در علم آمار، سری زمانی به توالی داده‌ها به ترتیب زمان اشاره دارد. این مفهوم در علوم آب (مثلاً هیدرولوژی، هیدروکلیماتولوژی و هواشناسی) کاربردهای عملی فراوانی یافته است. بررسی وجود روند زمانی در داده‌های یک ایستگاه یکی از جنبه‌های مهم سری زمانی به شمار می‌آید. برای این منظور روش‌های پارامتریک و ناپارامتریک فراوانی در کتب آماری و هیدرولوژی ارائه شده است. با این‌حال قدرت این آزمون‌ها با یکدیگر یکسان نمی‌باشد مثلاً (Onoz and Bayazit (2003) با شبیه‌سازی مونت کارلو نشان دادند که برای توزیع‌های چوله، قدرت آزمون پارامتریک t از آزمون ناپارامتریک من-کندال کم‌تر است. در نتیجه نمی‌توان انتظار داشت که آزمون‌های مختلف روند منجر به نتایج یکسانی شوند. (Ghahraman (2006) در بررسی روند درجه حرارت سالانه در ۳۴ ایستگاه همدید در ایران نشان داد که در حالی که نتایج دو آزمون t و من-کندال مشابه بودند ولی آزمون والد-ولفوویتز با آن‌ها سازگار نبود. کاربردهای فراوانی برای دو آزمون ناپارامتریک روند کندال (برخی مواقع به آن رگرسیون سن-تیل گفته می‌شود) و من-کندال در علوم آب گزارش شده است. (Obeysekera et al. (2011) برای بررسی روند باران و درجه حرارت در جنوب فلوریدا و Safari (2012) برای بررسی روند درجه حرارت در رواندا از هر دو آزمون کندال و من-کندال و (Olofintoye and Sule (2010) برای بررسی روند باران و درجه حرارت در چند ایستگاه در نیجریه تنها از آزمون من-کندال استفاده کردند. با این‌حال تمامی این آزمون‌ها بر اساس فرض ایستایی و نداشتن حافظه بلند مدت داده‌ها استوار بوده و انحراف معیار آماره‌های آن‌ها تنها به طول دوره آماری بستگی دارد (Kottegoda, 1980) را برای آزمون کندال و Salas, 1992) برای آزمون من-کندال).

(Hurst (1951) برای اولین بار حافظه‌دار بودن سری زمانی هیدرولوژیکی را معرفی کرد. پس از معرفی مفهوم فرآیندهای خود-متشابه در قالبی نظری (Kolmogrov, 1941)، ماندلبروت در سال ۱۹۶۵ فرآیند اغتشاش نرمال جزئی (FGN¹) را معرفی کرد (Mandelbrot, 1971) که پایه‌ای برای نظریه‌های فرکتال و آشوب به شمار می‌آید. پس از وی سایر پژوهش‌گران (Koutsoyiannis, 2002) نیز به این مفهوم پرداخته و نشان دادند که مؤلفه‌های هیدرولوژیکی بیش‌تری از فرآیند FGN پیروی می‌کنند. در چنین شرایطی انحراف معیار آماره‌های این آزمون‌ها نه تنها به طول دوره آماری بلکه به برخی ویژگی‌های سری زمانی، که بیانگر حافظه آن باشند، بستگی خواهد داشت. در حد مطالعات کتابخانه‌ای این تحقیق، تا به حال هیچ رابطه ریاضی مشخص برای

$$s(k) \approx c_k(H) k^H \sigma \quad (4)$$

که در آن $c_k(H)$ تابعی از H بوده و با رابطه زیر داده می‌شود:

$$s(k) \approx c_k(H) = \sqrt{\frac{n/k - (n/k)^{2H-1}}{n/k - 1/2}} \quad (5)$$

محاسبه توأمان σ و H در یک فرآیند تکراری انجام می‌شود. در ابتدا $c_k(H)$ را برابر با ۱ قرار داده و با برقراری رگرسیون خطی بین $\ln[s^{(k)}]$ و $\ln[k]$ و σ برآورد می‌شود. در گام بعد پس از محاسبه $c_k(H)$ از روی H ، با برقراری رگرسیون بین $\ln[s^{(k)}]$ و $\ln[c_k(H)k]$ مجدداً σ و H برآورد می‌شود. این فرآیند تا رسیدن به همگرایی تکرار می‌شود.

Koutsoyiannis (2002) مفهوم حافظه‌دار بودن یک سری زمانی با H معین را با ترکیب چند (مثلاً ۳) مدل ساده خودهمبسته (مارکوف) با مقیاس‌های زمانی متفاوت تقریب زد و نشان داد که در این صورت ساختار تابع خودهمبستگی آن نیز حفظ می‌شود:

$$\rho_j = (1 - c_1 - c_2) \rho_j + c_1 \varphi_j + c_2 \xi_j \quad (6)$$

که در آن ρ ، φ و ξ ضرایب خودهمبستگی با تأخیر ۱ به ترتیب برای مقیاس‌های زمانی کوتاه، متوسط و بلند و c_1 و c_2 دو عدد ثابت هستند که در نامساوی $c_1 + c_2 < 1$ صدق می‌کنند. میانگین‌های این سه مدل مارکوف به ترتیب $(1 - \rho)\mu$ ، $(1 - \varphi)\mu$ و $(1 - \xi)\mu$ میانگین سری پایه X است و پراش‌های آن‌ها به ترتیب $(1 - c_1 - c_2)\gamma_0$ ، $c_1\gamma_0$ و $c_2\gamma_0$ پراش سری پایه X است) می‌باشد. بر این اساس Koutsoyiannis (2002) نگرش "نوسانات در چند مقیاس زمانی متفاوت" را برای تولید استوکاستیکی سری زمانی با حفظ نمایه هرست (فرآیند FGN) پیشنهاد کرد. سه تا از پارامترهای معادله (۶) با روابط زیر تقریب زده می‌شود (Koutsoyiannis, 2002) که دقت آن‌ها برای سری‌های زمانی متداول در هیدرولوژی مناسب است:

$$\begin{cases} \rho = 1.52(H - 0.5)^{1.32} \\ \varphi = 0.953 - 7.69(1 - H)^{3.85} \\ \xi = \begin{cases} 0.932 + 0.08H, & H \leq 0.76 \\ 0.993 + 0.007H, & H > 0.76 \end{cases} \end{cases} \quad (7)$$

در حالی که وی برای دو پارامتر دیگر c_1 و c_2 از دو تأخیر دلخواه (مثلاً ۲ و ۱۰۰) و حل دستگاه دو معادله و دو مجهول استفاده کرد. به سادگی می‌توان نشان داد که این انتخاب دلخواه منجر به ناپایداری در جواب‌ها برای c_1 و c_2 می‌شود. به‌عنوان جایگزین و بر مبنای مفهوم معادلات نرمال (Kovacs, 1987)، می‌توان معادلات رگرسیونی برحسب H را به‌دست آورد:

$X = \bar{X} + KS$ است که در آن X_T مقدار واقعه با دوره بازگشت T سال، \bar{X} و S به ترتیب میانگین و انحراف معیار داده‌ها و K عامل فراوانی است که به دوره بازگشت و برخی ویژگی‌های توزیع فراوانی داده‌ها بستگی دارد (Kite, 1977 و Bobee and Ashkar, 1991). بر مبنای داده‌های موجود، عامل فراوانی برای هر ماه و در سال‌های مختلف محاسبه و کمینه و بیشینه آن در هر ماه استخراج شد. سپس فرض شد که عامل فراوانی در مواقع کمبود داده در بین مقادیر کمینه و بیشینه قرار داشته باشد. با چنین فرضی از تولید داده‌های یاغی جلوگیری می‌شود گرچه چنین داده‌هایی در بررسی روند اشکالی را به‌وجود نمی‌آورند. سپس اعداد تصادفی با توزیع یکنواخت استاندارد تولید و با توجه به کمینه و بیشینه عامل فراوانی در ماه معین، مقدار استوکاستیکی عامل فراوانی تولید شده و با توجه به میانگین و انحراف معیار محاسبه شده از روی داده‌های واقعی، داده مفقود برآورد گردید. پس از تکمیل داده‌های ماهانه، درجه حرارت سالانه محاسبه گردید.

۱-۲- فرآیندهای دارای حافظه

فرآیند FGN فرآیندی است که در رابطه

$$(Z_i^{(k)} - k\mu) \hat{=} \left(\frac{k}{l}\right)^H (Z_j^{(l)} - l\mu) \quad (1)$$

صدق کنند. در این رابطه علامت $\hat{=}$ به مفهوم هم‌توزیع بودن و H نمایه (یا ضریب) هرست (Hurst, 1951) می‌باشد. $Z_i^{(k)}$ فرآیند تصادفی تجمیم شده از متغیر پایه X (بدون مؤلفه چرخه‌ای؛ مثلاً سالانه) با گام زمانی k (بزرگ‌تر از ۱) می‌باشد. این معادله برای تمامی اعداد صحیح i و j (به مفهوم ایستا بودن فرآیند) و تمامی گام‌های زمانی k و l صادق است. در ساده‌ترین حالت ($i=j=l=1$)، معادله (۲) خاصیت خود متشابه بودن انحراف معیار را نشان می‌دهد:

$$s^{(k)} = \sigma [Z_i^{(k)}] = k^H \sigma \quad (2)$$

که از روی آن می‌توان به سادگی نمایه H را به‌دست آورد (Koutsoyiannis, 2002). یکی دیگر از ویژگی‌های فرآیند FGN، وابسته بودن تابع خودهمبستگی متغیر پایه به H و وابسته به مقیاس نبودن متغیرهای تجمیم یافته به آن می‌باشد:

$$\rho_j^{(k)} = \rho_j = (1/2) \left\{ (j+1)^{2H} + (j-1)^{2H} \right\} - j^{2H}, \quad (3) \\ j > 0$$

با این حال انحراف معیار سری زمانی که از FGN پیروی می‌کند به H بستگی دارد (Beran, 1994) که مقدار آن نیز باید به‌دست آید. در این شرایط معادله (۲) به‌صورت زیر تقریب زده می‌شود (Koutsoyiannis, 2003):

متفاوت"، ۱۰۰ سری تولید گردید. سپس انحراف معیار ۱۰۰ مقدار S محاسبه و آزمون معنی‌داری روند بررسی گردید.

۳- نتایج و بحث

۳-۱- تحلیل داده‌ها

داده‌ها با روش رگرسیون و تحلیل داخلی تکمیل شد. در جدول ۱ برخی ویژگی‌های آماری درجه حرارت ماهانه و سالانه مشاهد ارائه شده است. بررسی جدول نشان می‌دهد که مقادیر حدی (به جز یک مورد) بدون تغییر باقی مانده‌اند. تفاوت‌ها به‌طور کلی کم‌تر از ۵٪ است.

شکل ۱ سری زمانی درجه حرارت سالانه مشاهد را به همراه با میانگین متحرک ۵ و ۲۵ ساله نشان می‌دهد. روند افزایشی درجه حرارت در هر سه مقیاس زمانی به چشم می‌خورد. با این حال شیب روند به زمان بستگی داشته و از حوالی سال ۱۹۸۰ بر مقدار آن افزوده شده است. پژوهش‌گران متعددی بدون در نظر گرفتن داده‌های تاریخی، روند افزایشی درجه حرارت مشاهد را نشان داده‌اند (Ghahraman, 2006; Asmari et al., 2011؛ براتی و طرقي، ۱۳۷۸؛ داودی و همکاران، ۱۳۸۹؛ علیزاده و همکاران، ۱۳۸۹).

شکل ۲ تغییرات لگاریتم انحراف معیار را در مقابل لگاریتم گام زمانی (رابطه ۲) نشان می‌دهد. خطی بودن لگاریتم این دو متغیر در مقابل هم، نشان می‌دهد که درجه حرارت سالانه مشاهد از فرآیند FGN پیروی می‌کند. شیب این خط (۰/۸۲۸۱) به بیانی ساده، برابر با نمایه هرست می‌باشد. انحراف از خط مستقیم در گام‌های زمانی بزرگ‌تر ناشی از کم شدن طول دوره سری تجمیع شده می‌باشد. علاوه بر سری زمانی درجه حرارت مشاهد، فرآیندهای هیدروکلیماتولوژی و هیدرولوژیکی دیگری نیز وجود دارند که نمایه هرست برای آن‌ها از ۰/۵ بزرگ‌تر است (یعنی سری زمانی از فرآیند FGN پیروی می‌کند). برای نمونه برخی از این فرایندها در جدول ۲ ارائه شده است.

۳-۲- تحلیل روند

آماره τ و انحراف معیار سری زمانی دمای ایستگاه مشاهد، به‌ترتیب ۰/۳۳۶۱ و ۰/۰۶۰۰ به‌دست آمد. از این‌رو بر مبنای آزمون کلاسیک کندال فرض H_0 مبنی بر عدم وجود روند در داده‌ها در هر سطح معنی‌داری (p -مقدار برابر با $10^{-4} \times 1/0.479$ است) رد می‌شود. با این‌حال همان‌گونه که پیش‌تر بیان شد این مقدار انحراف معیار تنها برای فرآیندهای تصادفی محض ($H=0.5$) صادق است

$$\begin{cases} c_1 = f_1(H) \\ c_2 = f_2(H) \end{cases} \quad (8)$$

۲-۲- آزمون روند کندال

آماره τ و پراش آن در آزمون روند کندال به‌صورت زیر است (Kottegoda, 1980):

$$\tau = \frac{4p}{n(n-1)} - 1, \quad \text{var}(\tau) = \frac{2(2n+5)}{9n(n-1)} \quad (9)$$

که در آن p تعداد جفت مشاهداتی ($x_j, x_i; j>i$) است که در آن نامساوی $x_j < x_i$ (اگر روند نزولی مورد بررسی باشد) یا $x_j > x_i$ (اگر روند افزایشی مورد بررسی باشد) برقرار باشد. τ برای سری تصادفی از توزیع نرمال و با میانگین 0 و پراش $\text{var}(\tau)$ پیروی می‌کند. در حالی که این مقدار پراش برای سری زمانی که از فرآیند FGN پیروی کند درست نمی‌باشد. هیچ رابطه شناخته شده‌ای برای چنین شرایطی وجود ندارد و برای برآورد آن تنها می‌توان از شبیه‌سازی استوکاستیک استفاده کرد. برای این منظور با استفاده از نگرش اصلاح شده "نوسانات در چند مقیاس زمانی متفاوت"، ۱۰۰ سری تولید و برای هرکدام مقدار τ محاسبه گردید. سپس انحراف معیار ۱۰۰ مقدار τ محاسبه و آزمون معنی‌داری روند بررسی گردید.

۳-۳- آزمون روند من-کندال

آماره S در آزمون روند من-کندال و پراش آن به‌صورت زیر است (Salas, 1992):

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{sgn}(x_j - x_i), \quad (10)$$

$$\text{var}(S) = \frac{1}{18} \{N(N-1)(2N+5) - c\}$$

که در آن $\text{sgn}(u)$ تابع علامت است و مقدار آن برحسب مثبت، صفر یا منفی بودن u به‌ترتیب برابر با ۱، ۰ یا -۱ خواهد بود و c ضریب تصحیح ناشی از وجود داده‌های جفتی بوده و با $c = \sum_{p=1}^q t_p(t_p-1)(2t_p+5)$ داده می‌شود (q تعداد گروه‌های جفتی و t_p تعداد مشاهدات در p -امین گروه می‌باشد). همان‌گونه که برای آزمون کندال بیان شد، این مقدار پراش برای سری زمانی که از فرآیند FGN پیروی کند ($H>0.5$) درست نمی‌باشد ولی هیچ رابطه شناخته شده‌ای برای چنین شرایطی وجود ندارد و برای برآورد آن تنها می‌توان از شبیه‌سازی استوکاستیک استفاده کرد. آماره آزمون من-کندال با رابطه زیر داده می‌شود:

$$Z_S = \{S - \text{sgn}(S)\} / \sigma_Z \quad (11)$$

این آماره از توزیع نرمال استاندارد پیروی می‌کند. در این‌جا نیز با استفاده از نگرش اصلاح شده "نوسانات در چند مقیاس زمانی

(Beran, 1994) در حالی که مقدار H برای سری زمانی درجه حرارت سالانه مشهد به بیانی ساده و غیر دقیق، ۰/۸۲۸۱ می‌باشد.

جدول ۱- برخی ویژگی‌های آماری درجه حرارت ماهانه مشهد (سانتی‌گراد)*

ماه	کمینه	بیشینه	میانگین	انحراف معیار
ژانویه	-۷/۱	۷/۳	۱/۳۱	۲/۹۶
	(-۷/۱)	(۷/۳)	(۱/۱۴)	(۲/۹۵)
فوریه	-۸/۵	۹/۰	۳/۴۸	۲/۹۸
	(-۸/۵)	(۹/۰)	(۳/۲۸)	(۲/۸۵)
مارس	۳/۷	۱۴/۹	۸/۲۳	۲/۱۱
	(۳/۶)	(۱۴/۹)	(۸/۲۷)	(۲/۰۸)
آوریل	۱۰/۷	۱۹/۴	۱۴/۳۴	۱/۸۴
	(۱۰/۷)	(۱۹/۴)	(۱۴/۳۹)	(۱/۷۲)
مه	۱۵/۱	۲۴/۴	۱۹/۴۶	۱/۶۸
	(۱۵/۱)	(۲۴/۴)	(۱۹/۵۲)	(۱/۵۶)
ژوئن	۲۰/۴	۲۸/۲	۲۳/۹۹	۱/۵۵
	(۲۰/۴)	(۲۸/۲)	(۲۳/۸۷)	(۱/۴۶)
ژوئیه	۲۳/۲	۲۸/۹	۲۶/۰۲	۱/۳۱
	(۲۲/۷)	(۲۸/۹)	(۲۵/۸۳)	(۱/۲۸)
اوت	۲۰/۷	۲۷/۹	۲۴/۳۲	۱/۵۹
	(۲۰/۷)	(۲۷/۹)	(۲۴/۱۹)	(۱/۴۸)
سپتامبر	۱۶/۴	۲۳/۳	۱۹/۸۳	۱/۴۶
	(۱۶/۴)	(۲۳/۳)	(۱۹/۷۱)	(۱/۳۹)
اکتبر	۱۰/۰	۲۰/۰	۱۴/۲۱	۱/۷۹
	(۱۰/۰)	(۲۰/۰)	(۱۴/۱۶)	(۱/۷۴)
نوامبر	۳/۲	۱۲/۳	۸/۳۶	۱/۹۷
	(۳/۲)	(۱۲/۳)	(۸/۲۶)	(۱/۸۷)
دسامبر	-۳/۷	۸/۴	۳/۸۱	۲/۳۴
	(-۳/۷)	(۸/۴)	(۳/۶۳)	(۲/۲۸)
سالانه	۱۱/۰	۱۷/۰	۱۴/۰۳	۱/۱۶
	(۱۱/۰)	(۱۷/۰)	(۱۳/۸۶)	(۱/۰۷)

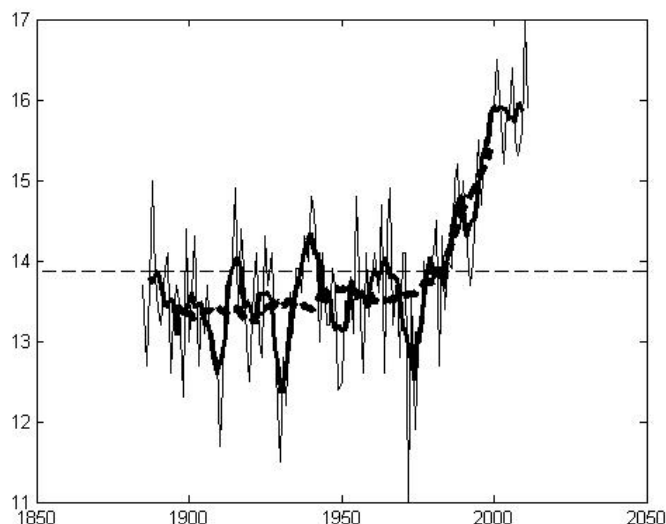
* اعداد داخل پرانتز مربوط سری‌های زمانی پس از بازسازی (یعنی ۱۸۸۵ تا ۲۰۱۱) و اعداد خارج از پرانتز مربوط به دوره زمانی ۱۹۵۱ تا ۲۰۱۱ می‌باشد

شکل ۳ فرآیند همگرا شدن برآورد H (رابطه‌های ۴ و ۵) را نشان می‌دهد. مقدار H در اولین تکرار ناشی از به‌کارگیری نادقیق رابطه (۲) است که در آن انحراف معیار ثابت و مستقل از H فرض شده

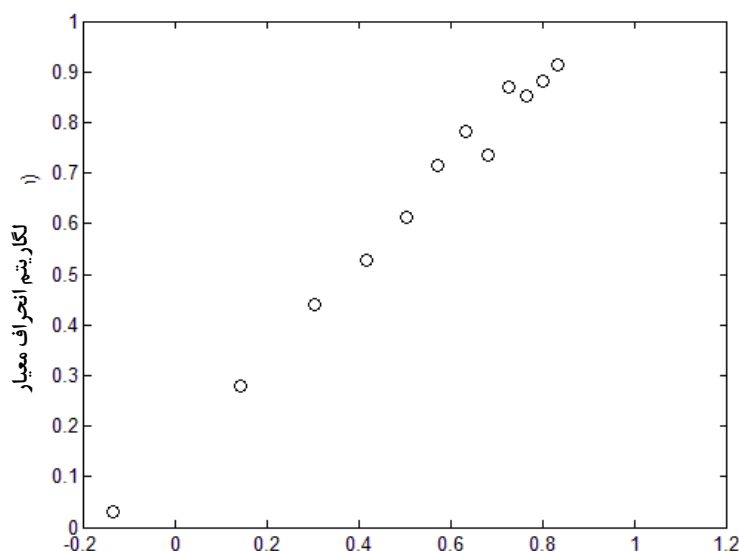
است. ولی با آگاهی از این که انحراف معیار نیز به‌طور متقابل به H وابسته است (Beran, 1994)، امکان برآورد توأم آن نیز وجود دارد. فرآیند همگرا شدن انحراف معیار مشابه شکل ۳ بود و در این‌جا نشان داده نشد.

با داشتن H ، ابتدا پارامترهای رابطه (۶) به‌دست آمد و سپس بر اساس نگرش اصلاح شده "نوسانات در چند مقیاس زمانی متفاوت"، سری استوکاستیکی درجه حرارت سالانه مشهد تولید گردید به‌گونه‌ای که در آن‌ها نمایه هرست حفظ گردد. هر کدام از این ۱۰۰ سری مصداق‌هایی برای سری زمانی درجه حرارت سالانه‌ی مشهد به‌شمار می‌آیند. با این حال نباید گمان کرد که این سری‌های تولید شده دقیقاً نسخه یکسان از سری تاریخی باشند (برای مثال شکل ۴-۷ در صفحه ۱۴۲ از کتاب برن (Beran, 1998) نشان می‌دهد که چنین مصداق‌هایی می‌توانند به آسانی با سری نایستا اشتباه گرفته شوند). برای هر کدام از این سری‌ها، آماره τ (رابطه ۹) محاسبه و سپس انحراف معیار آن از روی ۱۰۰ مقدار به‌دست آمد. این مقدار ۰/۴ بود که حدود ۶/۵ برابر مقدار نظیر برای سری تصادفی است. در نتیجه p -مقدار برای آزمون کندال ۰/۲ بوده، روند افزایشی درجه حرارت سالانه مشهد در سطوح معنی‌داری ۱ و ۵ درصد معنی‌دار نبوده: رفتار مشاهده شده ناشی از رفتار طبیعی یک فرآیند FGN می‌باشد.

این نتیجه برخلاف نتایج گزارش شده پیشینیان (Ghahraman, 2006؛ Asmari et al., 2011؛ براتی و طرقي، ۱۳۷۸؛ داودی و همکاران، ۱۳۸۹؛ علیزاده و همکاران، ۱۳۸۹) است. آن‌ها نه تنها داده‌های تاریخی را در نظر نگرفتند بلکه به حافظه‌دار بودن درجه حرارت نیز توجه نکردند. با این حال در نظر نگرفتن داده‌های تاریخی نتیجه معنی‌داری را تغییر نمی‌دهد زیرا در این بازه روند افزایشی داده‌ها ناچیز است (شکل ۱). در حالی که مشکل اصلی، در نظر نگرفتن حافظه سری زمانی است. برای مثال داده‌های حداقل سطح آب در رودخانه نیل (Beran, 1994) و میانگین متحرک ۲۵ ساله آن در شکل ۴ ترسیم شده است. بازه‌های زمانی a تا e در این شکل به روشنی روند مشخص افزایشی یا کاهشی را از خود نشان می‌دهند. طول زمانی این بازه‌ها به‌ترتیب ۱۳۷، ۸۶، ۱۳۲، ۱۲۹ و ۱۰۷ سال است. هر کدام از این بازه‌ها هم از نقطه نظر جهت کلی روند و هم از نقطه نظر طول دوره زمانی به‌خوبی با شکل ۱ همخوانی دارد. بنابراین پرداختن به تنهایی به یکی از این بازه‌ها و در نظر نگرفتن حافظه سری زمانی (پرداختن به سایر بازه‌ها با روندی مخالف) منجر به نتیجه‌گیری‌هایی نادرست خواهد شد.



شکل ۱- سری زمانی سالانه (خط ممتد نازک)، میانگین‌های متحرک ۵ (ممتد ضخیم) و ۲۵ (منقطع ضخیم) سالانه و میانگین دراز مدت درجه حرارت سالانه مشهد (منقطع افقی)



شکل ۲- انحراف معیار سری‌های زمانی تجمع شده در مقابل گام زمانی برای سری زمانی درجه حرارت سالانه مشهد

جدول ۲- نمایه هرست برای برخی فرآیندهای هیدروکلیماتولوژیکی و هیدرولوژیکی

H	طول دوره آماري- سال	مرجع داده‌ها	نوع داده
۰/۷۵	۱۹۹۰	(a)	قطر استاندارد شده تنه درختان در آبراهه ماموت در یوتا- ایالات متحده آمریکا
۰/۸۸	۹۹۲	Jones et al. (1998a,b)	آنومالی درجه حرارت نیم‌کره شمالی
۰/۸۵	۶۶۳	Beran, 1994	حداقل سالانه سطح آب رودخانه نیل در مصر
۰/۵۹	۳۳۱	Biondi et al. (2001)	PDO^2 تطویل شده بر مبنای قطر تنه درختان
۰/۷۹	۲۳۲	ftp.cru.uea.ac.uk	متوسط درجه حرارت سالانه پاریس
۰/۸۸	۱۱۳	(b)	نمایه نوسان دهه‌ای پاسیفیک (PDO)
۰/۷۹	۹۱	Koutsoyiannis (2003)	روان آب سالانه رودخانه بوئوتیکاس کفپاس در یونان
۰/۶۱	۶۳	(c)	نمایه نوسان آتلانتیک شمالی (NAO^3)

(a) <ftp://ftp.ngdc.noaa.gov/paleo/treering/chronologies/asciifiles/usawest/ut509.crn>

(b) <http://jisao.washington.edu/pdo/PDO.latest>

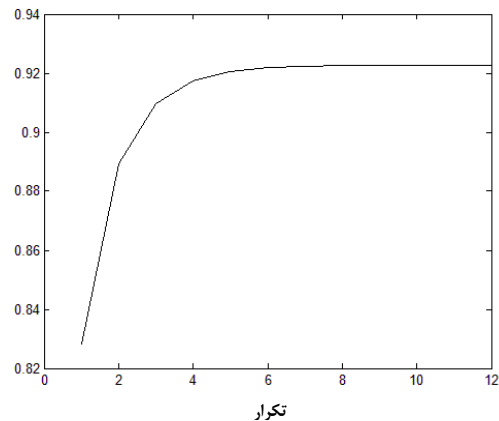
(c) <http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii.table>

می‌باشد. می‌توان این پارامتر را $R_S^{[\tau]}$ (انحراف معیار نسبی τ ; مشابه مفهوم فاکتور تورم واریانس) نامید. به منظور بررسی رفتار $f(N, H)$ در ابتدا N را ثابت نگاه داشته و به ازاء مقادیر مختلف H (از 0.5 تا 0.99 با گام‌های 0.1) انحراف معیار τ را محاسبه کردیم. نتایج، وابستگی قوی خطی $R_S^{[\tau]} = a + bH$ را نشان داد. مثلاً برای $N=100$ رابطه زیر ($r=0.985$) به دست آمد:

$$f(100, H) = -3.1865 + 9.5580H \quad (13)$$

مقدار $R_S^{[\tau]}$ در $H=0.5$ بزرگ‌تر از 1 بود و مقدار برآوردی آن توسط رابطه (13) $1/6$ بود (شکل 5). در حالی که انتظار می‌رفت مقدار آن برابر با 1 باشد. با این حال می‌توان به این نکته توجه کرد که اولاً این مقدار ایستا نمی‌باشد و مقدار آن با تکرار تولید استوکاستیکی تغییر می‌کند و ثانیاً هدف این مقاله پرداختن به فرآیندهای FGN است و بنابراین انحراف در $H=0.5$ اهمیتی نخواهد داشت (مقدار H برای درجه حرارت سالانه مشهد 0.92 می‌باشد). سپس این روش برای N های مختلف (از 50 تا 250 با گام 5) تکرار گردید. هر دو پارامتر a و b رفتاری خطی و نظام‌مند با $\ln N$ از خود نشان دادند (ضریب همبستگی برای برآورد a و b به ترتیب برابر با 0.98 و 0.99 بود). در نتیجه رابطه زیر به‌عنوان رابطه‌ای جامع برای تصحیح انحراف معیار τ در آزمون روند کندال پیشنهاد می‌شود:

$$R_S^{[\tau]} = f(N, H) = 8.6223 - 2.6455 \ln N + (-19.1718 + 6.4021 \ln N)H \quad (14)$$



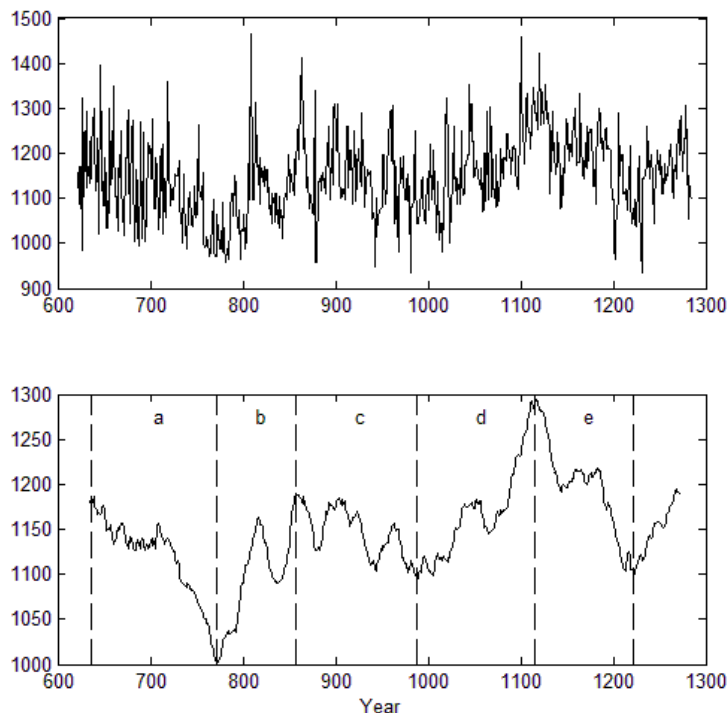
شکل ۳- همگرا شدن برآورد نمایه هرست برای سری زمانی درجه حرارت سالانه مشهد

در روش کلاسیک آزمون کندال ($H=0.5$)، انحراف استاندارد τ تنها به N (طول دوره آماری) بستگی دارد (معادله ۹). در حالی که با درک ماهیت فرآیند FGN مقدار آن باید به H به عنوان مشخصه‌ای کلیدی در این فرآیند نیز بستگی داشته باشد. در این جا ما روش بالا برای محاسبه انحراف معیار τ را توسط تولید استوکاستیکی برای N و H معین تعمیم می‌دهیم.

انحراف معیار τ به صورت زیر در نظر گرفته شد:

$$S_\tau = S_\tau^{[classic]} \times f(N, H) \quad (12)$$

که در آن $S_\tau^{[classic]}$ انحراف معیار τ در روش کلاسیک (رابطه ۹) و $f(N, H)$ تابعی نامعلوم از طول دوره آماری (N) و نمایه هرست (H)



شکل ۴- سری زمانی حداقل سالانه سطح آب در رودخانه نیل (بالا) و میانگین متحرک ۲۵ ساله آن (پایین)

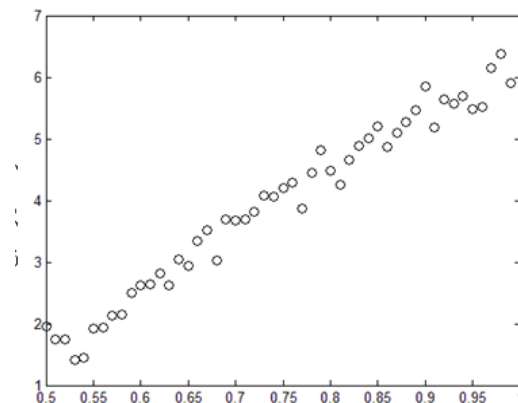
می‌شود و از این رو منجر به اضافه شدن مقداری نااطمینانی به سری زمانی می‌شود افزون بر این که در سال‌های پیش از ۱۹۵۰ فاصله ایستگاه‌هایی که در رگرسیون مشارکت می‌کنند نسبتاً زیاد است. در این بخش، خلاءهای آماری تکمیل نگردید و سال‌های بدون آمار صرفاً از سری زمانی حذف گردید. با این کار گرچه بر مبنای اصول آمار کلاسیک تصادفی بودن سری نقض می‌شود ولی با توجه به هدف این مقاله که بررسی روند در داده‌ها است، با اتخاذ یک فرض ضمنی اشکالی را به وجود نمی‌آورد. فرض ضمنی در این جا این است که داده‌ها در سال‌های حذف شده در بین دو سال i و i بزرگ‌تر از مقدار داده در سال i نباشد. گرچه این فرض تا اندازه‌ای محدود کننده است ولی با توجه به زیاد نبودن درصد داده‌های مفقود نسبت به کل طول دوره سری زمانی، مقدار p در آزمون کندال (تعداد جفت مشاهداتی که در آن نامساوی $x_i > x_{i+1}$ برقرار باشد) و آماره S در آزمون روند من-کندال (رابطه ۱۰) تغییر چندانی نمی‌کنند. مقادیر p - مقدار برای هر کدام از آزمون‌های کندال و من-کندال و برای دو حالت کلاسیک و در نظر گرفتن فرآیند FGN در جدول ۳ آورده شده است. در نتیجه بر مبنای حالت کلاسیک، فرض H_0 مبنی بر عدم وجود روند در داده‌ها در هر سطح معنی‌داری و برای هر دو آزمون کندال و من-کندال رد می‌شود در حالی که در سطوح معنی‌داری متداول $0.1/0$ و $0.05/0$ در هیچ‌کدام از دو آزمون روند دلیلی برای رد کردن فرض H_0 مبنی بر عدم وجود روند در داده‌ها وجود ندارد. نتایج به دست آمده در حالت تکمیل داده‌ها به روش استوکاستیکی نیز مشابه بود (جدول ۳).

جدول ۳- p- مقدار برای آزمون‌های روند کندال و من-کندال در دو حالت عدم تکمیل خلاءهای آماری و تکمیل داده‌ها به روش استوکاستیکی برای درجه حرارت سالانه مشهد

فرآیند FGN	آزمون کلاسیک	آزمون روند	حالت
۰/۱۱۵۰	$3/5 \times 10^{-10}$	کندال	عدم تکمیل خلاءها ($H=0/91$)
۰/۵۲۶۰	$7/7 \times 10^{-11}$	من-کندال	
۰/۲۳۰۷	$1/6 \times 10^{-6}$	کندال	تکمیل داده‌ها به روش استوکاستیکی ($H=0/90$)
۰/۴۶۰۵	$2/95 \times 10^{-6}$	من-کندال	

۴- نتیجه گیری

هر دو آزمون کندال و من-کندال در حالت کلاسیک نشان دادند که سری زمانی درجه حرارت سالانه مشهد به طول ۱۲۷ سال دارای روند افزایشی معنی‌دار (p - مقدار کوچک‌تر از $0.001/0$) است در حالی که با توجه به مفهوم حافظه بلند مدت در داده‌ها (فرآیند اغتشاش نرمال جزئی؛ FGN با نمایه هرست برابر با $0.92/0$)، انحراف



شکل ۵- وابستگی انحراف معیار نسبی σ به نمایه‌ی هرست (H) برای سری زمانی به طول ۱۰۰

مجدداً یادآوری می‌شود که اعداد تجربی ثابت در معادله (۱۴) یکتا نمی‌باشد و مقادیر آن با هر بار تکرار در فرآیند تولید استوکاستیکی تغییر می‌کند.

نتایج در آزمون روند من-کندال مشابه بود. آماره من-کندال و انحراف معیار آن در نگرش کلاسیک (رابطه ۱۰) به ترتیب 2925 و $(479/85)^2$ به دست آمد. از این رو آماره آزمون Z_S (رابطه ۱۱) برابر با $6/09$ می‌باشد. در نتیجه بر مبنای آزمون کلاسیک من-کندال فرض H_0 مبنی بر عدم وجود روند در داده‌ها در هر سطح معنی‌داری (p - مقدار برابر با $10^{-9}/1$ است) رد می‌شود. در حالی که 100 بار تولید استوکاستیکی از سری زمانی، انحراف استاندارد آماره S را به $3125/8$ ($6/5$ برابر بیش‌تر) افزایش داد که متناظر با p - مقداری برابر با $0.35/0$ می‌باشد. در نتیجه در سطوح معنی‌داری متداول $0.1/0$ و $0.05/0$ دلیلی برای رد کردن فرض H_0 مبنی بر عدم وجود روند در داده‌ها وجود ندارد. مشابه با آنچه برای تعمیم آزمون کندال بیان شد، مفهوم انحراف معیار نسبی S را در نظر گرفتیم و رابطه زیر برای برآورد آن به دست آمد:

$$R_S^{[r]} = f(N, H) = 8.7677 - 2.6749 \ln N N + (-19.2591 + 6.4147 \ln N) H \quad (15)$$

مقدار ضریب تصحیح در ازاء $H=0/5$ با این رابطه نیز بزرگ‌تر از 1 به دست می‌آید که بر مبنای آنچه برای آزمون کندال بیان شد برای فرآیندهای FGN که در آنها $H > 0/5$ است اشکالی را به وجود نمی‌آورد.

۳-۳- تأثیر عدم تکمیل خلاءهای آماری و روش تکمیل استوکاستیکی

تکمیل خلاءهای آماری عمدتاً توسط روش‌های رگرسیونی انجام

Biondi F, Gershunov A, Cayan DR (2001) North Pacific decadal climate variability since 1661. *J. Climate* 14:5-10.

Bobee B, Ashkar F (1999) The Gamma family and derived distributions applied in hydrology. Water Resources Publications, 203p.

Chow VT (1954) The log-probability law and its engineering applications. Proc. ASCE 80, Paper no. 536:1-25.

Cornes RC (2010) Early meteorological data from London and Paris: Extending the North Atlantic oscillation series. PhD Dissertation, University of East Anglia. URL: http://www.cru.uea.ac.uk/cru/pubs/thesis/2010-cornes/Master_withlinks.pdf.

Ghahraman B (2006) Time trend in the mean annual temperature of Iran. *Turk. J. Agric. Forest.* 30:439-448.

Ghahraman B, Ahmadi F (2007) Application of geostatistics in time series: Mashhad annual rainfall. *Iran-Watershed Manage. Sci. Engin.* 1(1):7-15.

Hurst HE (1951) Long term storage capacities of reservoirs. *Transac ASCE* 116:776-808.

Jones PD, Briffa KR, Barnett TP, Tett SFB (1998a) Millennial temperature reconstructions, IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series #1998-039, NOAA/NGDC Paleoclimatology Program, Boulder, Colorado, USA (ftp.ngdc.noaa.gov/paleo/contributions_by_author/jones1998/)

Jones PD, Briffa KR, Barnett TP, Tett SFB (1998b) High-resolution paleoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. *Holocene* 8(4):455-471.

Katz RW, Parlange MB, Tebaldi C (2003) Stochastic modelling of the effects of large-scale circulation on daily weather in the southeastern US. *Clim. Change* 60:189-216.

Kite GW (1977) Frequency and risk analyses in hydrology. Water Resources Publications, 224p.

Kolmogrov AN (1941) Local structure of turbulence in fluid for very large Reynolds numbers. Translation in *Turbulence*. S.K. Friedlander and L. Topper (eds.), 1961, Interscience Publishers, New York:151-155.

Kottegoda NT (1980) Stochastic water resources technology. MacMillan Press, London, UK.

Koutsoyiannis D (2002) The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci.-J.* 47(4):573-595.

معیار آماره‌های آزمون افزون بر ۶ برابر بیش‌تر شد و در نتیجه هیچ‌کدام از این دو آزمون، معنی‌داری روند افزایشی را در سطوح معنی‌داری متداول ۰/۰۱ و ۰/۰۵ تأیید نمودند. روابط رگرسیونی برای تصحیح انحراف معیار برای دو آزمون ناپارامتری روند کندال و من-کندال به‌عنوان تابعی از نمایه هرست و طول دوره آماری برای اولین مرتبه به‌دست آمد. همواره تکمیل داده‌ها با نااطمینانی همراه است. برای هدف بررسی روند نشان داده شد که عدم تکمیل داده‌ها نتایج یکسانی به‌بار می‌آورد. روش جدید استوکاستیکی بر پایه مفهوم عامل فراوانی چاو پیشنهاد شد و نشان داده شد که نتایج پایدار باقی می‌ماند.

۵- قدردانی

این تحقیق در زمان حضور در دانشگاه مانی‌توبا و در قالب فرصت مطالعاتی اعطا شده به نویسنده توسط دانشگاه فردوسی مشهد انجام گردید که بدین‌وسیله تشکر می‌شود.

پی‌نوشت‌ها

- 1- FGN; Fractional Gaussian Noise
- 2- PDO; Pacific Decadal Oscillation
- 3- NAO; North Atlantic Oscillation

۶- مراجع

براتی غ، طرقی ج (۱۳۷۸) تعیین روند تغییرات دما و بارش شهر مشهد طی دوره ۱۹۵۱-۹۴. تحقیقات جغرافیایی، شماره ۵۴ و ۵۵: ۱۵۱-۱۶۵.

خلیلی ع، بذرافشان ج (۱۳۸۷) ارزیابی مخاطره تداوم خشک‌سالی با استفاده از داده‌های بارندگی سالانه قرن گذشته در ایستگاه‌های قدیم ایران. *مجله ژئوفیزیک ایران*، جلد ۲، شماره ۱۳: ۲-۲۳.

داودی م، محمدی ح م، بای ن (۱۳۸۹) تجزیه و تحلیل و پیش‌بینی برخی عناصر اقلیمی مشهد. *نیوار*، شماره ۷۰ و ۷۱: ۳۵-۴۶.

علیزاده ا، سیاری ن، حسامی کرمانی م ر، بنایان اول م، فریدحسینی ع (۱۳۸۹) بررسی پتانسیل اثرات تغییر اقلیمی بر منابع و مصارف آب کشاورزی (مطالعه موردی: حوضه آبریز کشف رود). *آب و خاک*، جلد ۲۴، شماره ۴: ۸۱۵-۸۳۵.

Asmari M, Abbasi F, Arabshahi H (2011) Preliminary time series analysis of Mashhad air temperatures (1961-2005). *Int. J. Sci. Adv. Technol.* 1(6):73-78.

Beran J (1994) Statistics for long-memory processes, Vol. 61 of Monographs on Statistics and Applied Probability, Chapman and Hall, New York, USA.

- Olofintoye OO, Sule BF (2010) Impact of global warming on the rainfall and temperature in the Niger Delta of Nigeria. USEP: J. Res. Inf. Civil Engin. 7(2):33-48.
- Onoz B, Bayazit M (2003) The power of statistical tests for trend detection. Turk. J. Engin. Environ. Sci. 27:247-251.
- Safari B (2012) Trend analysis of the mean annual temperature in Rwanda during the last fifty two years. J. Environ. Protec. 3:538-551.
- Salas JD (1992) Analysis and Modeling of Hydrologic Time Series. In: D.R. Maidment (Ed.), Handbook of Hydrology. McGraw Hill Book Company, U.S.A.
- Zahraie B, Roozbahani A (2011) SST clustering for winter precipitation prediction in southeast of Iran: Comparison between modified K-means and genetic algorithm-based clustering methods. Expert Sys. Applic. 38(5):1287-1304.
- Koutsoyiannis D (2003) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol. Sci.-J. 48(1):3-24.
- Kovacs G (1987) Chapter 9: Distribution of hydrologic parameters over area and along rivers, In: O. Starosolszky (ed.), Applied Surface Hydrology. Water Resources Publications. Littleton, Colorado, 80161 USA, 418-502.
- Lintner BR , Chiang JCH (2005) Reorganization of tropical climate during El Niño: a weak temperature gradient approach. J. Climate 18(24):5312-5329.
- Mandelbrot BB (1971) A fast fractional Gaussian noise generator. Water Resour. Res. 7(3):543-553.
- Obeysekera J, Park J, Irizarry-Ortiz M, Trimble P, Barnes J, VanArman J, Said W, Gadzinski E (2011) Past and projected trends in climate and sea level for South Florida. Interdepartmental Climate Change Group, South Florida Water Management District, West Palm Beach, Florida, Hydrologic and Environmental Systems Modeling Technical Report. July 5.