



Technical Note

Assessment of A Non-Parametric Method for Hydrologic Data Disaggregation in Space-Time Based On K-Nearest Neighbors (Case Study: Three Sub-Basins in West of Urmia Lake)

M. M. Bateni^{1*} and M. Montaseri²

Abstract

Stochastic parametric disaggregation models that maintain spatial and temporal correlation are widely used in hydrology. To avoid high complexity and large number of parameters, which imposes a significant amount of uncertainty to the results, the use of non-parametric disaggregation methods has been widely suggested by researchers as an alternative. Among the non-parametric modeling methods, the K-nearest neighbors method proposed by Prairie et al. gains strong mathematical basis and inherent simplicity. In our work, the modified disaggregation approach of the K-nearest neighbors method is used for temporal and spatial disaggregation of rainfall and river flow values and the method performance is evaluated. The exploited flow and rainfall data corresponded to three stations in three sub-basins located at west of Lake Urmia. The total amount of annual rainfall and flow of the three stations were generated using stochastic lag -1 autoregressive model (AR (1)). Using the non-parametric disaggregation model, the generated annual values were disaggregated into three sub-basins. The annual values for each sub-basin were then disaggregated into different months. Comparing statistics of disaggregated data with those of historical data, showed the good performance of the proposed disaggregation model and its ability to disaggregate streamflow and rainfall data.

Keywords: Non-Parametric Models, K-Nearest Neighbor, Disaggregation Models, Hydrological System, Synthetic Time Series.

Received: March 11, 2017

Accepted: July 21, 2017

یادداشت فنی

ارزیابی یک روش توزیع ناپارامتری برای داده‌های هیدرولوژیک در مکان-زمان مبتنی بر K نزدیک‌ترین همسایه (مطالعه موردی: سه زیرحوضه غرب دریاچه ارومیه)

محمد مهدی باطنی^{۱*} و مجید منتصری^۲

چکیده

در هیدرولوژی، مدل‌های آماری توزیع داده جهت توزیع جریان و بارندگی در مکان و زمان با حفظ وابستگی زمانی و مکانی به طور گسترده‌ای استفاده می‌شوند. از بین این مدل‌ها، مدل‌های توزیع داده ناپارامتری به دلیل سادگی و عدم نیاز به تعداد پارامترهای زیاد اخیراً بسیار مورد استقبال محققین قرار گرفته‌اند. در این تحقیق روش توزیع ناپارامتری ال-شمارا اصلاح شده به عنوان یک روش K نزدیک‌ترین همسایه معرفی و سپس عملکرد آن در توزیع زمانی و مکانی مقادیر جریان رودخانه‌ای و بارش در سه حوضه آبریز در غرب دریاچه ارومیه مورد استفاده قرار گرفته است. بدین منظور ابتدا مقادیر مجموع جریان و بارندگی سالیانه سامانه هیدرولوژیک مورد نظر با استفاده از مدل آماری خود بازگشتی مرتبه اول تولید گردید. در ادامه براساس مدل توزیع ناپارامتری ال-شمارا اصلاح شده، ابتدا داده‌های سالیانه تولیدی از مدل خود بازگشتی مرتبه اول در سه حوضه آبریز توزیع شد. پس از آن، مقادیر سالیانه در هر حوضه به ماه‌های مختلف سال توزیع گردید. مقایسه پارامترهای آماری داده‌های تاریخی و تولیدی در دو سطح سالیانه و ماهیانه نشان می‌دهد که روش توزیع ناپارامتری ال-شمارا بعنوان یک روش K نزدیک‌ترین همسایه اصلاح شده، در بازتولید همبستگی‌های زمانی بهتر از همبستگی‌های مکانی عمل کرده است. همچنین این روش آمارگان تاریخی در سطح سالانه را به خوبی بازتولید کرده، در صورتی که عملکرد آن در بازتولید آمارگان تاریخی در سطح ماهانه ضعیف‌تر است.

کلمات کلیدی: مدل ناپارامتری، K نزدیک‌ترین همسایه، مدل توزیع داده،

سامانه هیدرولوژیک، سری‌های زمانی مصنوعی.

تاریخ دریافت مقاله: ۹۵/۱۲/۲۱

تاریخ پذیرش مقاله: ۹۶/۴/۳۰

1- PhD Candidate, Water Engineering, Urmia University, Urmia, Iran. Email: mm.bateni@urmia.ac.ir

2- Associate Professor, Water Engineering, Urmia University, Urmia, Iran.

*- Corresponding Author

۱- کاندیدای دکترای مهندسی منابع آب، دانشگاه ارومیه، ارومیه.

۲- دانشیار گروه مهندسی آب، دانشگاه ارومیه.

*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان تابستان ۱۳۹۷ امکانپذیر است.

خصوصیات موجود در داده‌ها از جمله در شرایط غیرنرمال بودن تابع توزیع احتمالی آن‌ها دارد. این روش معیابی دارد که در موارد زیر خلاصه شده‌اند (۱) ممکن است مقادیر منفی تولید شود (۲) محدوده تغییرات تولید شده توسط این مدل در یک جهت محدود است و (۳) واریانس داده‌های تولیدی کمتر از مقدار تاریخی است (Salas and Lee, 2010).

از آنجا که استفاده از برآوردگر چگالی هسته‌ای در مسائل با ابعاد زیاد محدودیت‌های فراوانی دارد، Prairie et al. (2007) روش توزیع ناپارامتری لال - شارما را با جایگزین کردن برآوردگر K نزدیک‌ترین همسایه به جای چگالی هسته‌ای، در حوضه آبریز رودخانه کلرادو آمریکا برای توزیع مکانی و زمانی جریان رودخانه مورد استفاده قرار دادند. این محققان ابتدا اقدام به تولید داده جریان کل سالانه در حوضه آبریز (شاخه اصلی پایین دست) کردند و سپس با استفاده از روش توزیع ناپارامتری مذکور اقدام به توزیع مکانی (توزیع در زیر حوضه‌ها) و زمانی (توزیع جریان سالانه در فصل‌های مختلف سال) داده‌ها کردند.

روش توزیع داده پیرا و همکاران، علاوه بر ویژگی‌های عمومی روش‌های توزیع ناپارامتری به مراتب از روش‌های توزیع ناپارامتری دیگر ساده‌تر و قابل فهم بوده و به واسطه پایه ریاضی آسان و منطقی آن دارای ظرفیت کاربردی بالایی است. این روش دارای عملکرد قابل قبولی در توزیع داده‌های غیرگوسی یا غیر نرمال بوده و قادر به تولید مقادیر حدی بزرگتر و کوچکتر (بی‌سابقه) در سری تاریخی نیز هست اما همچنان از همان معایب روش توزیع داده لال - شارما رنج می‌برد. لذا در این تحقیق از روش توزیع داده پیرا و همکاران که به گونه‌ای اصلاح شده که مقادیر منفی تولید نکند، استفاده گردید و عملکرد این روش توزیع ناپارامتری به عنوان یک روش K نزدیک‌ترین همسایه در توزیع زمانی و مکانی مقادیر جریان رودخانه‌ای و بارش مورد ارزیابی قرار گرفته است. بدین منظور از داده‌های جریان و بارندگی سه حوضه آبریز مجاور با داشتن شرایط هیدرولوژیکی همسان در غرب دریاچه ارومیه استفاده شده است. ابتدا مقادیر مجموع جریان و بارندگی سالیانه سامانه هیدرولوژیک مورد نظر با استفاده از مدل آماری خود بازگشتی مرتبه اول تولید گردیده است. در ادامه براساس مدل توزیع ناپارامتری لال - شارما اصلاح شده، داده‌های مجموع تولیدی از مدل خود بازگشتی مرتبه اول در سه حوضه آبریز در سطح سالیانه توزیع شده و سپس مقادیر سالیانه تولید شده در هر حوضه به ماه‌های مختلف سال توزیع گردیده است. در نهایت، پارامترهای آماری داده‌های تولیدی به دست آمده از مدل با داده‌های تاریخی جهت ارزیابی عملکرد مدل مورد مقایسه قرار گرفته است.

تحلیل اکثر سامانه‌های منابع آب از جمله سامانه مخازن ذخیره، پیش خشکسالی و غیره براساس داده‌های جریان یا بارندگی ماهانه انجام می‌پذیرد، لذا انتخاب مدل آماری ماهانه برای یک تحلیل دقیق‌تر از رفتار و عملکرد واقعی چنین سامانه‌هایی دارای اهمیت ویژه‌ای است. در هیدرولوژی برای تولید داده معمولاً از مدل‌های سالانه استفاده شده و سپس این داده‌های سالانه به وسیله مدل‌های توزیع^۱ در مقیاس ماهانه، توزیع می‌شوند (Montaseri and Heydari, 2016). به علاوه از این مدل‌ها به عنوان مناسب‌ترین ابزار توزیع داده‌ها در مکان (در سامانه‌های چندگانه) نیز استفاده می‌شود (Salas, 2006). این مدل‌ها برای توزیع داده‌های تولیدی تجمعی جریان یا بارندگی در مکان (توزیع جریان کل رودخانه به جریان در سرشاخه‌های آن) یا زمان (توزیع بارندگی سالانه به بارندگی ماهانه) مورد استفاده قرار می‌گیرند. مدل‌های توزیع در دو دسته کلی ناپارامتری^۲ و پارامتری^۳ طبقه‌بندی می‌شوند. در روش‌های ناپارامتری داده‌های ماهانه مصنوعی مستقیماً به وسیله وزن‌دهی و نمونه‌گیری مجدد از داده‌های ثبت شده تاریخی تولید می‌شوند (Prairie et al., 2007). لذا در این مدل‌ها از عدم قطعیت برآورد پارامترهای مدل و انتخاب توزیع احتمالی داده‌های تاریخی اجتناب می‌گردد (Acharya and Ryu, 2014).

مدل‌های پارامتریک از یک رابطه ریاضی که معمولاً یک رابطه خطی بوده، تبعیت کرده و عمدتاً بسیار پیچیده‌تر از مدل‌های ناپارامتری هستند. به همین جهت اکثر محققین ترجیح می‌دهند از روش‌های ناپارامتری که ساختار تجربی داده‌های مشاهداتی را حفظ می‌کنند، استفاده کنند. در حالی که استفاده از مدل‌های پارامتریک نیازمند فرضیاتی در مورد توزیع احتمالی حاشیه‌ای و تغییرات ساختاری مکانی و زمانی متغیرها می‌باشد (Vogel and Shallcross, 1996). به علاوه این مدل‌ها به دلیل تعداد زیاد پارامترها برای مسائل با ابعاد کم و سری داده‌های طولانی مدت مناسب‌تر هستند و در سامانه‌های منابع آب با ابعاد بزرگتر، سال‌های آماری کم و با به‌کارگیری تبدیل‌هایی مانند لگاریتمی، توانی و امثال آن برای داده‌های با توزیع غیرنرمال دارای عدم قطعیت قابل توجهی خواهند بود (Nowak et al., 2010; Ganji et al., 2007).

از پرکاربردترین مدل‌های توزیع ناپارامتری، روش برآورد چگالی هسته‌ای^۴ می‌باشد که Lall and Sharma (1996) معرفی کردند. Tarboton et al. (1998) مبنای کاربرد و به‌کارگیری این روش توزیعی را به طور دقیق بیان نمودند. روش لال - شارما یک روش ناپارامتری توزیع داده است که توانایی قابل توجهی در حفظ و بازتولید

۲- روش تحقیق

۲-۱- منطقه مورد مطالعه و داده‌های مورد استفاده

برای اطمینان از تصادفی بودن داده‌های مورد استفاده در این مطالعه از آزمون گردش‌ها^۵ استفاده گردید. در صورتی که آماره این آزمون (Z) در سطح معنی‌دار مشخص (مثلاً ۵ درصد) در محدوده مقادیر مجاز تئوریک قرار نگیرد، فرضیه H_0 یعنی تصادفی بودن داده‌ها رد و فرضیه H_1 یعنی تصادفی نبودن داده‌ها تأیید خواهد شد (Kottegoda and Rosso, 2008). نتایج حاصل از آزمون مذکور بیانگر تصادفی بودن داده‌های تاریخی بارندگی و جریان در سطح معنی‌دار ۵ درصد در ایستگاه‌های باران‌سنجی و آب‌سنجی منتخب است. همچنین به منظور بررسی ایستایی و وجود روند در داده‌های تاریخی از آزمون ضریب همبستگی رتبه بزرگی^۶ استفاده گردید (Dahmen and Hall, 1990). مقادیر آماره آزمون فوق (Z') نیز در محدوده مجاز آن در سطح معنی‌دار ۵ درصد برای داده‌های تاریخی بارندگی و جریان در ایستگاه‌های منتخب قرار گرفت که بیانگر ایستایی داده‌های تاریخی بارندگی و جریان در تمام این ایستگاه‌ها است. از آنجا که داده‌های مورد نیاز برای مدل‌سازی خود بازگشتی مرتبه اول الزاماً می‌بایست نرمال باشند، در ابتدا اقدام به برازش داده‌های جریان و بارندگی سالانه با شش توزیع احتمال رایج (یعنی نرمال، لاگ نرمال دو و سه پارامتری، پیرسون نوع سوم، لاگ پیرسون نوع سوم و گامبل) و انتخاب توزیع مناسب برای داده‌های بارندگی و جریان سالانه با استفاده از آزمون PPCC^۷ شد (Vogel, 1986). همچنین توزیع مناسب برای داده‌های مجموع بارندگی و جریان سالانه نیز با استفاده از همین آزمون مشخص گردید. نتایج حاصل نشان داد که داده‌های جریان و بارندگی سالانه در ایستگاه‌های باران‌سنجی و آب‌سنجی با توزیع احتمال پیرسون نوع سوم (۴ مورد) و لاگ نرمال سه پارامتری (۲ موردی) تطبیق بالاتری دارند. پس از برازش توزیع مناسب، داده‌ها از توزیع اصلی خود به توزیع نرمال منتقل شدند. به منظور توضیح بیشتر در مورد توابع انتقال از توزیع‌های پیرسون نوع سوم و لاگ نرمال سه پارامتری به توزیع نرمال به (Kite (1988) مراجعه کنید.

در این مطالعه، سه حوضه آبریز نازلوچای، شهرچای و باراندوزچای واقع شده در غرب دریاچه ارومیه دارای شرایط اقلیمی، مورفولوژی و هیدرولوژیک مشابه و نیز رابطه همبستگی بالا مابین رخدادهای اقلیمی و هیدرولوژیک مانند بارندگی و جریان رودخانه‌ای به عنوان مناطق مطالعاتی انتخاب شده‌اند. حوضه‌های آبریز مذکور در مجاورت هم از طرف مرزهای غربی کشور به شرق تا دریاچه ارومیه کشیده شده و رواناب‌های سطحی آن‌ها بوسیله رودخانه‌های سه گانه نازلوچای، شهرچای و باراندوزچای زهکشی می‌شوند (شکل ۱). رودخانه‌های مذکور از کوه‌های مرزی غرب کشور ایران سرچشمه گرفته و پس از تأمین آب شرب و آب کشاورزی مسیر خود از جمله شهر ارومیه و دشت وسیع ارومیه در نهایت به دریاچه ارومیه تخلیه می‌شوند. در شکل ۱ موقعیت جغرافیایی حوضه‌های آبریز و رودخانه‌های سه گانه مذکور نشان داده شده است. ایستگاه‌های هواشناسی و آب‌سنجی مختلفی در حوضه‌های آبریز نازلوچای، شهرچای و باراندوزچای توسط وزارت نیرو تجهیز و مورد بهره‌برداری قرار می‌گیرند. در این مطالعه در هر حوضه آبریز، یک ایستگاه باران‌سنجی و یک ایستگاه آب‌سنجی در بالادست رودخانه که معرف رژیم طبیعی رودخانه باشد به عنوان ایستگاه‌های مبنای هواشناسی و هیدرولوژی انتخاب شده و مورد استفاده قرار گرفته است (شکل ۱). ایستگاه‌های ششگانه منتخب (سه ایستگاه باران‌سنجی و سه ایستگاه آب‌سنجی) دارای دوره آماری مشترک ۴۰ ساله (۱۳۵۱ الی ۱۳۹۰) برای ایستگاه‌های باران‌سنجی (آباجالو، بند و دیزج) و دوره آماری ۵۱ ساله (۱۳۴۰ الی ۱۳۹۰) برای ایستگاه‌های آب‌سنجی (تپیک، میرآباد و هاشم‌آباد) هستند. در این مطالعه، از داده‌های هیدرولوژیکی در مقیاس زمانی سالانه و ماهانه استفاده شده است. مشخصات آماری داده‌های سالانه بارندگی و جریان رودخانه ایستگاه‌های مورد مطالعه در جدول ۱ ارائه شده است.

Table 1- Basic statistical properties of stramflow and precipitation data for selected stations

جدول ۱- مشخصات آماری داده‌های سالانه تاریخی برای ایستگاه‌های باران‌سنجی و آب‌سنجی منتخب

Station type	Station name	Subbasin	Altitude (m)	Mean (mm or m ³ /s)	Coefficient of variation (CV) %	Lag-1 serial correlation coefficient	Skewness coefficient
Rain gauge	Abajaloo	Nazloochai	1290	269.9	36	0.02	1.45
	Band	Shahrchai	1390	409.3	27	-0.03	0.58
	Dizaj	Barandoozchai	1325	343.2	31	0.10	0.76
Streamflow gauge	Tapik	Nazloochai	1450	372.9	43	0.29	0.94
	Mirabaad	Shahrchai	1497	136.6	40	0.18	0.88
	Hashemabaad	Barandoozchai	1478	262.6	36	0.12	0.94

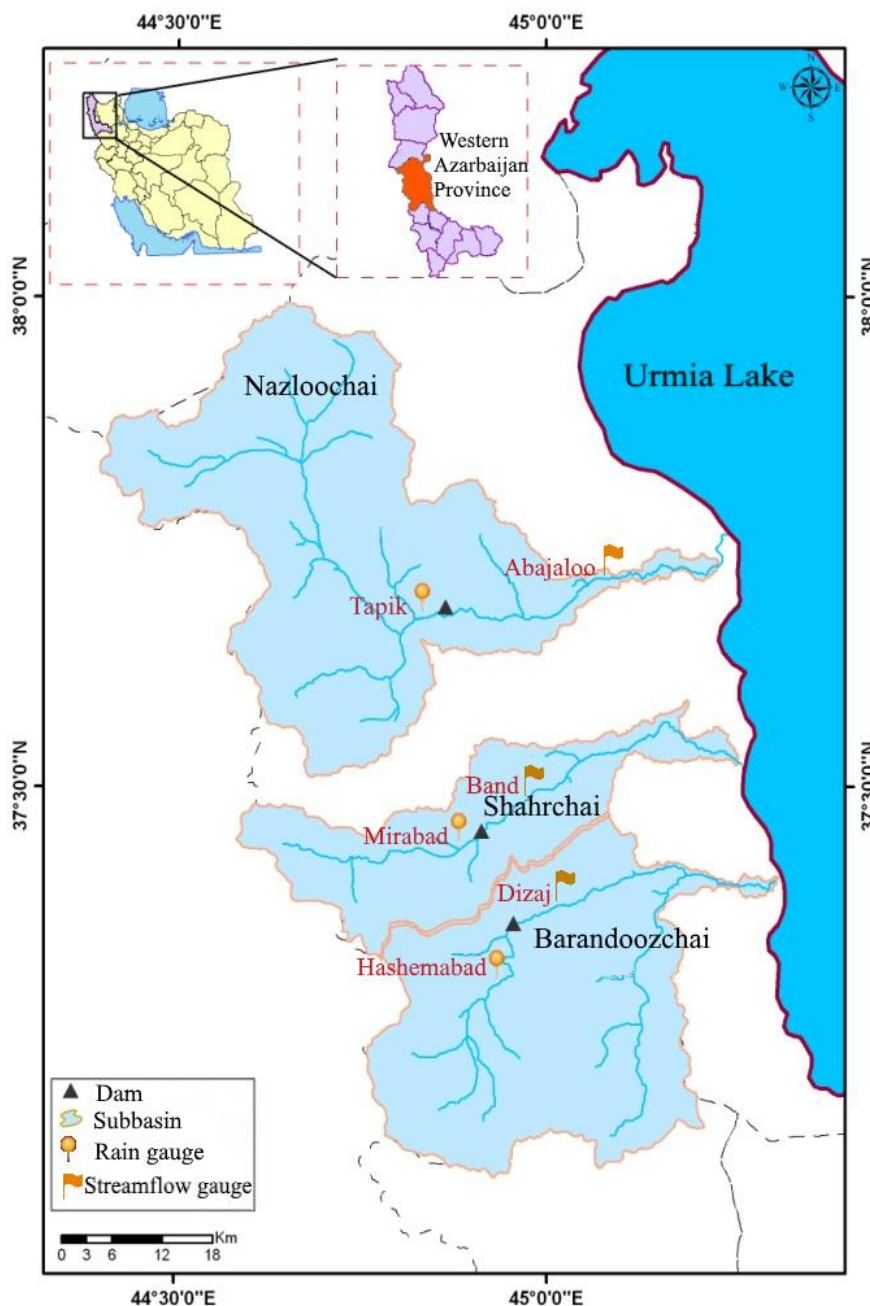


Fig. 1- Locations of selected sub-basins and precipitation and streamflow gauge stations in west of Urmia lake basin

شکل ۱- موقعیت جغرافیایی حوضه‌های آبریز غرب دریاچه ارومیه و ایستگاه‌های باران‌سنجی و آب‌سنجی مورد استفاده

شده است. (Prairie et al., 2007). الگوریتم این روش در زیر با جزئیات ارائه

۲-۲- روش توزیع داده ناپارامتری اصلاح شده Prairie et al. (2007)

۲-۲-۱- مراحل الگوریتم توزیع

در این روش، ابتدا نزدیک‌ترین همسایه‌های متغیر مجموع ورودی بر اساس فاصله اقلیدسی تعیین شده و همسایه‌ها بر اساس یک تابع وزن رتبه‌بندی می‌شوند. سپس، همسایه‌ای با تابع وزنی حداکثر، باز نمونه‌گیری شده و آنگاه داده مجموع یا همان داده ورودی مدل توزیعی با الگویی شبیه به این همسایه در مکان-زمان توزیع می‌شود

۱- داده‌های تاریخی توزیع شده طوری در ماتریس X قرار می‌گیرند که داده‌های توزیع شده در ردیف‌ها و داده‌های مجموع در ستون‌ها باشند. این ماتریس دارای d ردیف است که برابر تعداد فصول یا

ایستگاه‌های منفرد بوده و N ستون دارد که برابر تعداد سال‌های مشاهداتی است.

۲- X توسط ماتریس دوران یک‌معامده^۱ (R) به ماتریس Y انتقال می‌یابد،

$$Y = RX \quad (2)$$

ماتریس دوران مربعی R با ابعاد d به گونه‌ای محاسبه می‌شود که خاصیت جمع‌پذیری داده‌های توزیع شده تضمین شود. بر همین اساس، ماتریس انتقال یافته Y را که دارای ابعادی مشابه ماتریس X است، می‌توان به صورت (U, Z') نشان داد که U شامل d-1 سطر اول از داده‌های توزیع شده و Z'=Z/√d است که Z همان داده‌های مجموع تاریخی است.

۳- داده مجموع (به عنوان مثال جریان سالانه) از مدل مناسب تولید داده تولید می‌شود (Z_{sim}). این داده نیز بر √d تقسیم می‌شود تا به فضای داده‌های تاریخی ماتریس Y انتقال یابد. حاصل این کار با Z'_{sim} نشان داده می‌شود.

۴- K نزدیک‌ترین همسایه داده مجموع تولید شده با تعیین فاصله داده‌های تاریخی با داده مجموع تولید شده در فضای انتقال یافته تعیین می‌شود. به این همسایه‌ها وزنی برابر با مقدار زیر تعلق می‌گیرد:

$$W(k) = \frac{1}{k \sum_{i=1}^K \frac{1}{i}} \quad k=1,2,\dots,K \quad (3)$$

این تابع وزنی، وزن بیشتری به همسایه‌های نزدیک‌تر و وزن کمتری به همسایه‌های دورتر می‌دهد (Lall and Sharma, 1996). با مجموع کردن K بردار متناظر نزدیک‌ترین همسایه که هر یک در وزنی براساس رابطه (۳) ضرب شده‌اند، مقدار مطلوب بردار توزیعی انتقال یافته (u*) به دست می‌آید.

به طور معمول از دو روش تجربی و روش اعتبارسنجی متقابل تعمیم یافته (GCV)^۹ برای تعیین مقدار بهینه پارامتر K استفاده می‌شود. اختلاف محسوسی بین استفاده از روش تجربی و GCV در تعیین تعداد همسایگان (K) مشاهده نشده است (Lall and Sharma, 1996)؛ Yates et al., 2003؛ Rajagopalan and Lall, 1999؛ Prairie et al., 2006). از این رو در این مطالعه نیز از روش تجربی برای تعیین تعداد همسایه‌ها استفاده شد. در روش تجربی معمولاً پارامتر K را برابر نزدیک‌ترین عدد صحیح به √N در نظر می‌گیرند که N برابر تعداد داده‌های مجموع مشاهداتی است.

۵- بردار توزیعی انتقال یافته کامل (y*) با اضافه کردن Z'_{sim} به u* به صورت y* = (u* Z'_{sim}) ساخته می‌شود.

۶- گام نهایی بازگشت به فضای اصلی با استفاده از معکوس چرخش R است:

$$x^* = R^T y^* \quad (4)$$

که x* داده‌های توزیع شده هستند که مجموعشان برابر Z_{sim} خواهد بود.

۷- در صورتی که داده‌های تولیدی از الگوریتم مذکور منفی باشند، داده‌های منفی برابر صفر در نظر گرفته شده و به اندازه مجموع مقادیر منفی به سهمی متناسب با بزرگی هر مقدار مثبت از هر یک از این مقادیر کم می‌شود تا مجموع مقادیر توزیع شده ثابت باشد.

مراحل ۱ تا ۶ شامل روش پیشنهادی پیرا و همکاران و مرحله هفتم دربرگیرنده اصلاحی است که به منظور اجتناب از تولید مقادیر منفی به روش اضافه شده است.

هر چند در این روش از داده‌های تاریخی استفاده شده و از آن‌ها باز نمونه‌گیری می‌شود، اما در مراحل ۵ و ۶ امکان شبیه‌سازی و تولید مقادیر توزیع شده‌ای که قبلاً مشاهده نشده‌اند فراهم می‌آید. جزییات به دست آوردن ماتریس چرخش R در قالب یک مثال عددی در پیوست آمده است.

۲-۳- ارزیابی روش توزیع داده

عملکرد روش توزیعی K نزدیک‌ترین همسایه ارائه شده براساس توانایی آن در بازتولید آماره‌های داده‌های تاریخی سنجیده می‌شود. پروسه ارزیابی عملکرد در مطالعه حاضر همانند بسیاری از مطالعات دیگر مانند (Maheepala and Perera, 1996)؛ Savic et al., 1989؛ Montaseri and McMahon and Mein, 1986؛ Adeloye, مقایسه‌های ذیل را شامل می‌شود.

- پارامترهای آماری پایه یعنی میانگین، انحراف معیار، ضریب چولگی و خود همبستگی با تاخیر یکساله داده‌های تولیدی و تاریخی در دو سطح سالانه و ماهانه

- ساختمان همبستگی داده‌های ماهانه یعنی همبستگی بین ماه‌های مختلف داده‌های تاریخی و تولیدی

- همبستگی مابین داده‌های ماهانه در ماه‌های مختلف با داده‌های سالانه تولیدی و تاریخی

- منحنی فراوانی^{۱۰} براساس داده‌های تولیدی و تاریخی

به منظور بررسی عملکرد روش توزیعی ارائه شده ابتدا مقادیر مجموع داده‌های سالانه در ۳ ایستگاه باران‌سنجی به صورت یک سری زمانی و مقادیر مجموع داده‌های سالانه در ۳ ایستگاه آب‌سنجی نیز به صورت یک سری زمانی در نظر گرفته شدند. پس از برآورد خصوصیات آماری داده‌های تاریخی این دو سری، اقدام به برازش مدل اتورگرسیو مرتبه اول برای هر سری شد و داده مصنوعی در ۱۰۰۰ نمونه با طولی برابر

۳-۱- نتایج در سطح ماهانه و سالانه

به منظور بررسی عملکرد روش توزیعی پیشنهادی در بازتولید میانگین داده‌های تاریخی، میانگین داده‌های تاریخی و تولیدی برای ایستگاه‌های منتخب در جدول ۲ آمده است. همانطور که از جدول ۲ مشخص است، میانگین داده‌ها در اکثر ماه‌ها برای هر یک از ایستگاه‌های باران‌سنجی با دقت قابل قبولی توسط این رویکرد توزیع داده، باز تولید شده است. این حالت برای ایستگاه‌های آب‌سنجی برقرار نیست. در سطح سالانه نیز روش توزیعی عملکرد قابل قبولی در مورد میانگین داده‌ها دارد و برای همه ایستگاه‌ها میانگین سالانه به خوبی توسط این رویکرد توزیع داده، باز تولید شده است.

جدول ۳ انحراف معیار داده‌های تاریخی و تولیدی را برای ایستگاه‌های منتخب نشان می‌دهد. برخلاف میانگین، انحراف معیار در ایستگاه‌های باران‌سنجی در سطح ماهانه به خوبی بازتولید نشده در حالی که در ایستگاه‌های آب‌سنجی این پارامتر در سطح ماهانه به خوبی بازتولید شده است. به علاوه در سطح سالانه در همه ایستگاه‌ها انحراف معیار به خوبی بازتولید شده است.

همانطور که از جدول ۲ مشخص است در منطقه مورد مطالعه ماه‌های فروردین تا تیر ماه‌های پرآب و ماه‌های آبان تا اردیبهشت ماه‌های پرباران هستند. بر اساس شکل ۲، در ماه‌های کم آب هر چند فاصله میان چارکی برای چولگی داده‌های شبیه‌سازی مشتمل بر مقادیر تاریخی می‌گردد؛ اما از آنجا که محدوده مقادیر چولگی برای داده‌های شبیه‌سازی شده نسبتاً وسیع است می‌توان گفت که در این ماه‌ها چولگی به خوبی بازتولید نمی‌شود.

دوره آماری داده‌های تاریخی تولید گردید. سپس با استفاده از روش توزیعی K نزدیک‌ترین همسایه اقدام به توزیع مکانی این داده‌ها گردید. داده‌های حاصل از این مرحله در سطح سالانه بوده و در گام بعدی این داده‌های سالانه ایستگاه‌ها نیز با همان روش توزیعی K نزدیک‌ترین همسایه بین ماه‌های مختلف سال توزیع شدند.

۳- نتایج و تحلیل نتایج

در مطالعه حاضر عملکرد مدل توزیع ناپارامتری ل-ل - شارما اصلاح شده در سامانه‌های منابع آب چندگانه مورد ارزیابی قرار گرفته و کیفیت بازتولید آماره‌های داده‌های توزیعی در سطح سالانه و ماهانه به تفکیک مورد بررسی و بحث قرار گرفته است.

نتایج مطالعه در جداول ۲ تا ۴ و اشکال ۲ تا ۵ نمایش داده شده‌اند. در جداول و نمودارهای ابتدایی از مقادیر متوسط آمارگان برای ۱۰۰۰ نمونه داده تولیدی به همراه داده‌های تاریخی استفاده شده است. در نمودارهای جعبه‌ای (اشکال ۴ و ۵)، مقادیر میانه، چارک بالا، چارک پایین و چندک‌های ۵ و ۹۵ درصد پارامتر مربوط به سری‌های شبیه‌سازی شده (۱۰۰۰ سری) برای هر ماه و برای سال به صورت خطوط افقی در نمودارها نمایش داده شده‌اند. دامنه واقع بین دو چارک بالا و پایین که دامنه میان چارکی نامیده می‌شود، در برگزیده ۵۰ درصد میانی از مقادیر مربوط و دامنه واقع بین چندک‌های ۵ و ۹۵ درصد در برگزیده ۹۰ درصد داده‌ها است. چنانچه آماره تاریخی مورد نظر در محدوده میان چارکی نمودار جعبه‌ای حاصل از داده‌های شبیه‌سازی شده باشد، عملکرد مدل در مورد آن آماره مطلوب ارزیابی می‌شود.

Table 2- Mean of historical data for selected stations (values in braces are for generated data)
جدول ۲- میانگین داده‌های ایستگاه‌های منتخب (مقادیر میانگین برای داده‌های تولیدی درون پرانتز آمده است)

Month (Persian calendar)	Streamflow gauges			Rain gauges		
	Tapik	Mirabaad	Hashemabaad	Abajaloo	Band	Dizaj
Mehr	8.20 (19.35)	2.30 (10.85)	7.11(16.15)	13.73(19.04)	19.65(24.86)	15.90(30.53)
Aban	12.62 (9.63)	3.84(11.42)	12.68(19.13)	32.98(50.96)	45.65(61.24)	37.85(45.39)
Azar	4.61(9.44)	3.69(11.47)	13.70(20.37)	26.15 (20.61)	48.40(48.32)	40.03(47.39)
Dey	4.23(9.37)	3.18(11.35)	12.13 (19.82)	20.10(21.13)	38.80(37.38)	33.03(36.12)
Bahman	4.13(9.37)	3.37(11.42)	11.82 (19.66)	21.45(21.57)	43.28(43.56)	35.23(36.61)
Esfand	9.47(10.24)	7.12 (14.05)	16.48 (21.80)	32.50(26.32)	54.90(45.77)	45.48 (38.21)
Farvardin	27.64(13.34)	24.81(24.39)	37.43 (33.51)	47.68(42.61)	70.23(60.33)	57.50(40.40)
Ordibehesht	53.86(21.77)	46.04 (32.19)	60.17(43.18)	49.48(36.62)	54.1 (43.44)	53.53(38.18)
Khordad	49.69(15.94)	44.76(27.19)	49.85 (31.43)	15.10(12.61)	19.53(14.38)	14.73(11.60)
Tir	22.39(10.88)	16.79(13.94)	25.55(19.09)	5.40(7.91)	7.65(11.79)	5.50(10.23)
Mordad	8.66(8.98)	5.33 (10.82)	10.34 (13.87)	2.45(5.92)	3.13(8.74)	2.30(5.95)
Shahrivar	5.51(9.03)	2.43 (10.22)	5.39 (13.15)	2.95(5.70)	3.98(8.61)	2.18(5.48)
Annual	161.89(136.10)	163.65(189.31)	262.65(271.15)	269.95(271.00)	409.27(408.44)	343.23(346.10)

Table 3- Standard deviation of historical data for selected stations (values in braces are for generated data)
جدول ۳- انحراف معیار داده‌های ایستگاه‌های منتخب (مقادیر انحراف معیار برای داده‌های تولیدی درون پرانتز آمده است)

Month (Persian calendar)	Streamflow gauges			Rain gauges		
	Tapik	Mirabaad	Hashemabaad	Abajaloo	Band	Dizaj
Mehr	4.99(19.35)	0.88(0.98)	3.99(6.26)	16.73(4.57)	24.47(10.08)	24.81(10.83)
Aban	12.62 (22.06)	0.83(0.97)	6.05(7.48)	32.17(12.53)	38.25(10.37)	34.86(12.13)
Azar	12.47 (24.15)	0.63(0.98)	4.90(7.40)	17.98(14.57)	29.70(10.71)	24.78(10.73)
Dey	12.47(23.54)	0.63(0.98)	3.52(6.78)	17.98(8.87)	22.38(16.01)	24.40(14.20)
Bahman	12.32(23.62)	0.68(0.98)	3.78(6.33)	22.96(8.65)	30.13(10.08)	28.68(10.66)
Esfand	19.44(27.60)	0.61(0.98)	7.73(7.00)	23.33(8.84)	29.44(15.26)	29.53(10.04)
Farvardin	59.29(45.36)	0.64(0.93)	15.51(8.00)	27.37(10.15)	38.30(16.40)	33.96(22.35)
Ordibehesht	112.64(60.79)	0.79(0.97)	24.89(11.21)	31.71(16.74)	34.15(19.33)	37.15(14.99)
Khordad	78.09(38.35)	0.67(0.97)	24.08(11.43)	14.49(11.80)	19.33(11.05)	14.43(11.03)
Tir	27.97(21.93)	0.67(0.98)	15.62(8.73)	9.94(8.12)	12.03(8.89)	8.73(7.52)
Mordad	10.39(17.30)	0.66(0.98)	8.62(6.55)	3.76(8.24)	6.31(6.58)	6.26(7.19)
Shahrivar	7.00(16.60)	0.68(0.98)	3.07(6.01)	6.40(7.13)	7.00(6.12)	5.37(7.07)
Annual	372.89(340.65)	0.75(0.98)	95.00(92.17)	95.39(88.57)	111.10(109.99)	104.87(91.59)

سطح ماهانه، مقادیر چولگی برای همه ایستگاه‌های آب‌سنجی و باران‌سنجی در سطح سالانه نسبتاً قابل قبول بازتولید شده‌اند. از شکل ۳ می‌توان دریافت که ضریب خود همبستگی با تاخیر یک ساله برای هیچ یک از ایستگاه‌ها در سطح ماهانه به خوبی بازتولید نشده است.

در ماه‌های پر آب غالباً چولگی به خوبی بازتولید شده است؛ هر چند در مواردی این شاخص در داده‌های شبیه‌سازی شده کمتر برآورد شده است. در سطح ماهانه در همه ایستگاه‌های باران‌سنجی چولگی در داده‌های شبیه‌سازی شده با مقدار تاریخی آن مطابقت ندارد. برخلاف

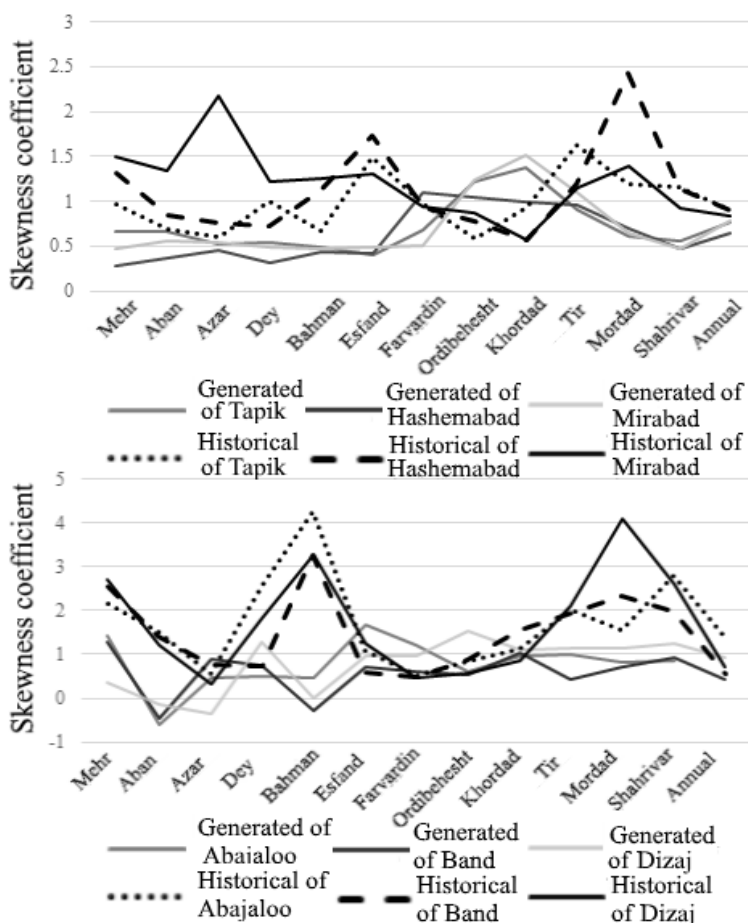


Fig. 2- Skewness coefficients of monthly and annual streamflow and rain data

شکل ۲- نمودار چولگی برای داده‌های ماهانه و سالانه بارش و جریان

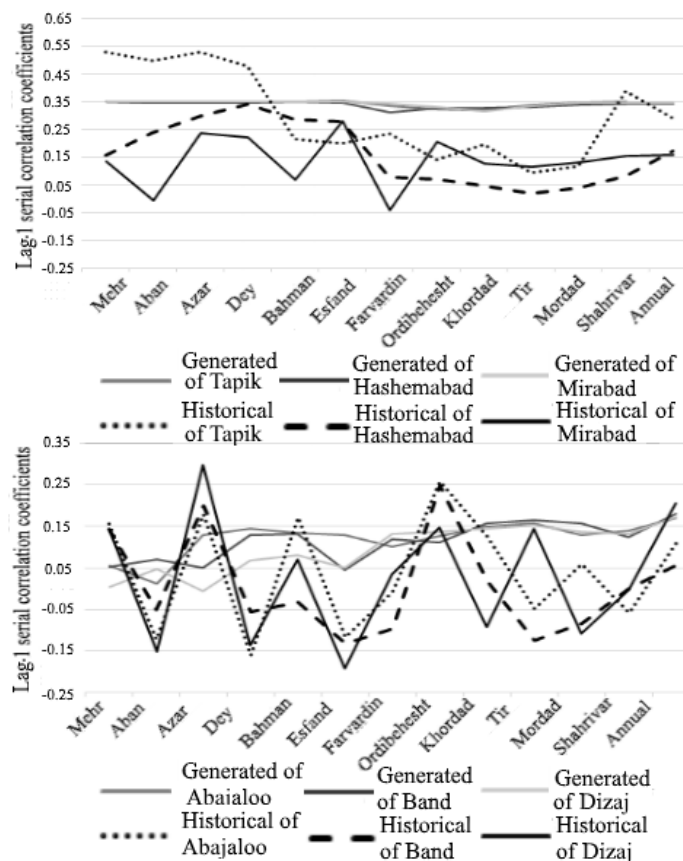


Fig. 3- Lag-1 serial correlation coefficients of monthly and annual streamflow and rain data
شکل ۳- نمودار خود همبستگی با تأخیر یک برای داده‌های ماهانه و سالانه بارش و جریان

آب‌سنجی بیشتر است (چه در سطح سالانه و چه در سطح ماهانه). همبستگی میان ایستگاه‌های باران‌سنجی بیشتر از همبستگی میان ایستگاه‌های آب‌سنجی است.

در سطح ماهانه نیز همبستگی میان ایستگاه‌های مختلف کم و بیش مطابق داده‌های سالانه است. در تمام ماه‌ها همبستگی مکانی میان داده‌های ماهانه سه ایستگاه باران‌سنجی برای داده‌های شبیه‌سازی شده توزیعی کمتر از داده‌های تاریخی است. در مورد همبستگی مکانی داده‌های سالانه این سه ایستگاه نیز کم برآوردی دیده می‌شود. برای ایستگاه‌های آب‌سنجی به طور کلی همبستگی مکانی به خوبی در داده‌های شبیه‌سازی شده بازتولید نشده است. مقادیر گزارش شده برای داده‌های تولیدی در این جدول مقادیر میانگین‌گیری شده از ۱۰۰۰ سری تولید شده هستند.

۴- خلاصه و جمع‌بندی

در این مطالعه از چارچوب توزیع K نزدیک‌ترین همسایه ارائه شده توسط Prairie et al (2007) که برای اجتناب از تولید مقادیر منفی

همانند سطح ماهانه، در سطح سالانه ضریب خود همبستگی با تأخیر یک ساله برای بیشتر ایستگاه‌ها به خوبی بازتولید نشده است.

هر چند در هیچ یک از ایستگاه‌ها انطباق کاملی بین تابع توزیع داده‌های شبیه‌سازی شده توزیعی و تاریخی وجود ندارد؛ اما به طور کلی می‌توان گفت که تابع توزیع برای داده‌های توزیع شده باران‌سنجی نسبت به آب‌سنجی مطابقت بیشتری با داده‌های تاریخی دارد. به عنوان نمونه، این تابع برای یک ایستگاه باران‌سنجی و یک ایستگاه آب‌سنجی برای یک ماه پرآب و یک ماه کم آب در شکل‌های ۴ و ۵ آورده شده است که در آنها آمارگان داده‌های تاریخی به صورت نقاط توپر نشان داده شده است.

۳-۲- همبستگی مکانی

در جدول ۴ همبستگی مابین داده‌های ایستگاه‌های منتخب گزارش شده است. همبستگی مابین داده‌های سالانه ایستگاه‌های منتخب مقادیر بالایی را نشان می‌دهد (بالاتر از ۰/۷). به طور کلی این مقادیر از همبستگی مابین ایستگاه‌های باران‌سنجی و مابین ایستگاه‌های

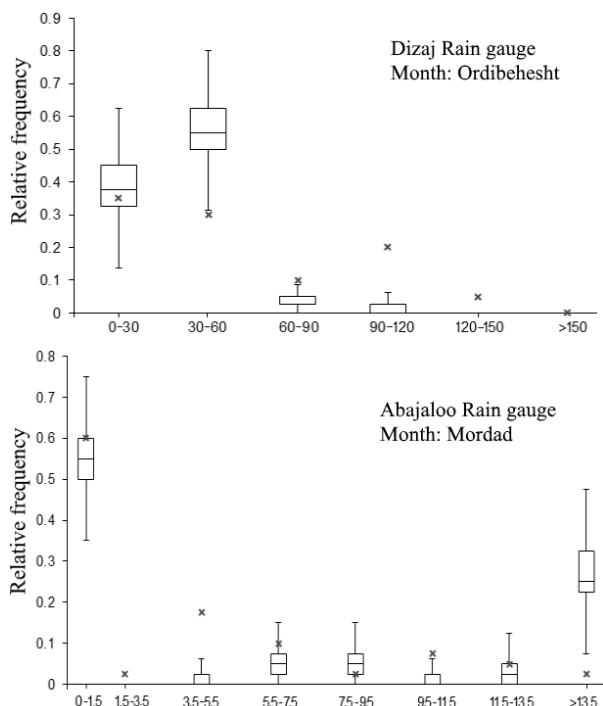


Fig. 5- Box plot of generated rain data (historical data shown by multiplication sign)

شکل ۵- نمودار جعبه‌ای تابع توزیع برای داده‌های بارش شبیه‌سازی شده (آمارگان داده‌های تاریخی به صورت علامت ضربدر نشان داده شده است)

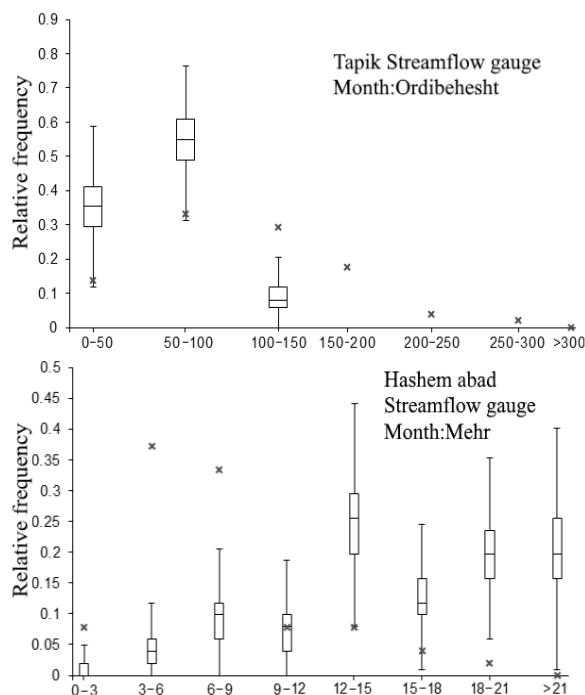


Fig. 4- Box plot of generated streamflow data (historical data shown by multiplication sign)

شکل ۴- نمودار جعبه‌ای تابع توزیع برای داده‌های جریان شبیه‌سازی شده (آمارگان داده‌های تاریخی به صورت علامت ضربدر نشان داده شده است)

به طور کلی روش مذکور توانایی حفظ خصوصیات توزیعی داده‌های هیدرولوژیک را دارد. به علاوه، این چارچوب برای شبیه‌سازی و توزیع داده‌های هیدرولوژیک در مکان/زمان ساده، قوی، و ممسک است. هر چند محدودیت‌هایی مانند عدم حفظ همبستگی ماه آخر سال با ماه اول سال بعد هم در روش توزیع داده پیشنهاد شده وجود دارد که می‌بایست مد نظر باشند.

اصلاح شده بود، برای توزیع مکانی و زمانی داده‌های بارش و جریان رودخانه استفاده گردید.

همانطور که پیش‌تر به صورت مفصل بحث شد، بازتولید خصوصیات آماری داده‌های تاریخی توسط داده‌های توزیع شده هدف نهایی استفاده از روش‌های توزیع داده است.

Table 4- Spatial correlation among the selected stations (values in braces are for generated data)
جدول ۴- همبستگی مکانی مابین داده‌های ایستگاه‌های منتخب (مقادیر همبستگی برای داده‌های تولیدی درون پرانتز آمده است)

Month (Persian calendar)	Streamflow gauges			Rain gauges		
	Tapik- Hashemabaad	Mirabaad- Hashemabaad	Tapik- Mirabaad	Dizaj-Abajaloo	Dizaj-Band	Band- Abajaloo
Mehr	0.56(0.96)	0.88(0.98)	0.60(0.99)	0.84(-0.04)	0.93(-0.44)	0.88(0.32)
Aban	0.69(0.96)	0.83(0.97)	0.71(0.99)	0.86(-0.09)	0.92(0.56)	0.90(-0.22)
Azar	0.62(0.96)	0.63(0.98)	0.54(0.99)	0.84(0.00)	0.90(0.40)	0.78(0.63)
Dey	0.66(0.96)	0.63(0.98)	0.54(0.99)	0.75(0.76)	0.75(0.58)	0.73(0.70)
Bahman	0.54(0.97)	0.68(0.98)	0.61(0.99)	0.91(0.70)	0.92(0.75)	0.88(0.76)
Esfand	0.84(0.97)	0.61(0.98)	0.75(0.99)	0.87(0.56)	0.89(0.82)	0.87(0.56)
Farvardin	0.75(0.93)	0.64(0.93)	0.74(0.97)	0.86(0.67)	0.92(0.72)	0.91(0.82)
Ordibehesht	0.50(0.95)	0.79(0.97)	0.66(0.98)	0.73(0.75)	0.87(0.84)	0.81(0.61)
Khordad	0.59(0.96)	0.67(0.97)	0.87(0.98)	0.44(0.90)	0.76(0.88)	0.49(0.89)
Tir	0.70(0.96)	0.67(0.98)	0.85(0.98)	0.81(0.92)	0.81(0.91)	0.93(0.89)
Mordad	0.75(0.97)	0.66(0.98)	0.85(0.99)	0.44(0.85)	0.45(0.86)	0.40(0.78)
Shahrivar	0.71(0.97)	0.68(0.98)	0.75(0.99)	0.87(0.85)	0.86(0.80)	0.82(0.73)
Annual	0.74(0.97)	0.75(0.98)	0.86(0.99)	0.83(0.95)	0.90(0.98)	0.87(0.96)

- Journal of Water Resources Planning and Management 128(6):456-461
- Montaseri M, Heydari J (2016) A comparison among the performance of the stochastic models in generating the monthly streamflow and rainfall data. *Iran-Water Resources Research* 11(3):69-84
- Nowak K, Prairie J, Rajagopalan B, Lall U (2010) A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. *Water resources research* 46(8):1-13
- Prairie JR, Rajagopalan B, Fulp TJ, Zagona EA (2006) Modified K-NN model for stochastic streamflow simulation. *Journal of Hydrologic Engineering* 11(4):371-378
- Prairie J, Rajagopalan B, Lall U, Fulp T (2007) A stochastic nonparametric technique for space- time disaggregation of streamflows. *Water resources research* 43(3):1-10
- Rajagopalan B, Lall U (1999) A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water resources research* 35(10):3089-3101
- Salas J, Sveinsson O, Lane W, Frevert D (2006) Stochastic streamflow simulation using SAMS-2003. *Journal of irrigation and drainage engineering* 132(2):112-122
- Salas JD, Lee T (2010) Nonparametric simulation of single-site seasonal streamflows. *Journal of Hydrologic Engineering* 15(4):284-296
- Savic DA, Burn DH, Zrinji Z (1989) A comparison of streamflow generation models for reservoir capacity yield analysis. *Journal of the American Water Resources Association* 25(5):977-983
- Tarboton DG, Sharma A, Lall U (1998) Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water resources research* 34(1):107-119
- Vogel RM (1986) The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses. *Water resources research* 22(4):587-590
- Vogel RM, Shallcross AL (1996) The moving blocks bootstrap versus parametric time series models. *Water resources research* 32(6):1875-1882
- Yates D, Gangopadhyay S, Rajagopalan B, Strzepek K (2003) A technique for generating regional climate scenarios using a nearest neighbor algorithm. *Water resources research* 39(7):1-15
- در مجموع می‌توان گفت که این روش هم برای داده‌های بارش و هم برای داده‌های جریان رودخانه قادر به بازتولید داده با خصوصیات آماری مشابه داده‌های تاریخی است. بر همین اساس و با توجه به سادگی محاسباتی، این روش توزیع داده ناپارامتری برای توزیع مکانی و زمانی داده‌های هیدرولوژیک توصیه می‌شود.
- پی‌نوشت‌ها**
- 1 -Disaggregation Models
 - 2 -Parametric
 - 3- Non-Parametric
 - 4 -Kernel Density Estimation
 - 5-Runs Test
 - 6-Rank Order Correlation Coefficient Test
 - 7-Probability Plot Correlation Coefficient Test
 - 8-Orthonormal
 - 9- Generalized Cross Validation (GCV)
 - 10- Frequency Curve
- ۵- مراجع**
- Acharya A, Ryu JH (2013) Simple method for streamflow disaggregation. *Journal of Hydrologic Engineering* 19(3):509-519
- Dahmen E, Hall MJ (1990) Screening of hydrological data: tests for stationarity and relative consistency. *International Institute for Land Reclamation and Improvement*, 58p
- Ganji A, Khalili D, Homayoun-far M (2007) Impact of uncertainty on risk indices in reservoir operation. *Iran-Water Resources Research* 2(3):13-26
- Kite GW (1988) Frequency and risk analysis in hydrology. *Water Resources Publications*, Fort Collins, CO, 257 p
- Kottegoda NT, Rosso R (2008) Applied statistics for civil and environmental engineers. *Blackwell Chichester, UK*, 736 p
- Lall U, Sharma A (1996) A nearest neighbor bootstrap for resampling hydrologic time series. *Water resources research* 32(3):679-693
- Maheepala S, Perera B (1996) Monthly hydrologic data generation by disaggregation. *Journal of hydrology* 178(1-4):277-291
- McMahon TA, Mein RG (1986) River and reservoir yield. *Water Resources Publications*, Fort Collins, CO, USA
- Montaseri M, Adeloje AJ (2002) Effects of integrated planning on capacity-yield-performance functions.