

## ارزیابی عملکرد رهیافت توسعه یافته‌ی نزدیکترین همسایه در شبیه سازی داده های هواشناسی روزانه

مهدی قمقما<sup>۱\*</sup>، نوذر قهرمان<sup>۲</sup> و شهاب عراقی نژاد<sup>۳</sup><sup>۱</sup> دانشجوی کارشناسی ارشد و <sup>۲،۳</sup> استادیاران، دانشکده مهندسی و فناوری کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

(تاریخ دریافت: ۱۳۸۹/۴/۶ - تاریخ تصویب: ۱۳۸۹/۱/۱۹)

## چکیده

مولدهای هواشناسی (Weather Generators) با هدف تطویل سری اطلاعات انواع متغیرهای وضع هوا اعم از بارش، دما و رطوبت نسبی، برای ارتقاء فهم و درک از عملکرد هر سیستمی که اقلیم عامل تاثیرگذار بر آن باشد، توسعه یافته‌اند. الگوریتم‌های متفاوتی از این مولدها در دو نوع کلی پارامتری و ناپارامتری تا به امروز ارائه شده‌اند. در این مطالعه کارایی مولد ناپارامتری k نزدیکترین همسایه با قابلیت برونیابی داده‌ها در سری مصنوعی، برای چندین ایستگاه شامل ایستگاه‌های تهران، مشهد، قزوین، بوشهر، تبریز و رشت با آمار قابل قبول ۴۵ سال (۲۰۰۵-۱۹۶۱) ارزیابی و برای بیان برتری نسبی این روش‌ها نسبت به روش‌های پارامتری، نتایج آن با خروجی مولد پارامتری LARS-WG مقایسه شده است. نتایج حاصله نشان داد که روش ناپارامتری به کارگرفته شده در این مطالعه در شبیه‌سازی اکثر پارامترهای سری مشاهده شده، نسبت به روش پارامتری مطمئن‌تر عمل می‌نماید. با این وجود در شبیه‌سازی طول دوره‌های درازمدت تر و خشک، مولد LARS-WG بهتر عمل می‌کند که البته اختلاف این دو مولد ناچیز می‌باشد.

واژه‌های کلیدی: رهیافت ناپارامتری، دوره‌های تر و خشک، برونیابی، مولد هواشناسی

## مقدمه

مبرم به سری اطلاعات هواشناسی وجود دارد. با این وجود در اکثر مناطق دنیا یا دسترسی به آمار هواشناسی وجود ندارد و یا طول دوره‌ی آماری موجود کوتاه می‌باشد. لذا به دلیل اهمیت دسترسی به این اطلاعات، از چند دهه‌ی اخیر روش‌هایی ابداع و توسعه یافته است که کمابیش قادر به شبیه‌سازی مناسب این داده‌ها براساس ویژگی‌های معینی که مقتبس از سری داده‌های موجود هستند، می‌باشد. همچنین در کنار کفایت طول دوره‌ی آماری در مطالعات، کیفیت این اطلاعات حائز اهمیت است. در واقع به علت غیرخطی بودن ساختار مدل‌های فرآیند پایه، و اثر به سزای پیشگویی‌های مختلف مرتبط با وقایع اقلیمی در مطالعات زراعی و هیدرولوژیکی، تغییرات کوچک در ورودی این مطالعات ممکن است منجر به تغییرات فاحش در خروجی آنها گردد (Semenov, 2008). لذا دو عامل کفایت و کیفیت داده‌های هواشناسی در انتخاب روش‌های شبیه‌سازی آنها دارای اهمیت است.

روش‌های تولید داده با نام مولدهای هواشناسی (Weather Generators) از دهه‌ی ۱۹۶۰ میلادی کاربردی شده‌اند و ساختار غالب این روش‌ها ترکیبی از معادلات عددی، آماری و تجربی می‌باشد که در نهایت هدفی جز شبیه‌سازی سری اطلاعات مصنوعی آب و هوا با ویژگی‌های معین منطبق بر مشاهدات ندارند. جزء مشترک در همه‌ی این روش‌ها، جزء تصادفی می‌باشد، که برحسب نوع روش یا مستقیماً عدد تولید

یکی از دستاوردهای علمی مرتبط با پیرامون بشر در طول چند دهه‌ی اخیر، شبیه‌سازی روابط حاکم بر فرآیندهای پیچیده محیطی می‌باشد که معلول تاثیر بسزا و مهم آن بر زندگی انسان است. مدل‌های مختلف اکولوژیکی، هیدرولوژیکی، زراعی، فرسایش، حفاظت آب و خاک و غیره از جمله شبیه‌سازهای محیطی می‌باشند که دارای ساختار ریاضی بوده و چون بر مبنای وقوع کنش‌ها و واکنش‌های خطی یا غیرخطی مختلف بین اجزای محیط هستند، به آنها مدل‌های فرآیند پایه (Process-based) می‌گویند (Semenov, 2008; Bannayan and Hoogenboom, 2008). با اجرای این مدل‌ها می‌توان براساس ورودی‌های معینی، نزدیک به آنچه که در طبیعت رخ می‌دهد، خروجی معینی دریافت نمود.

در کنار همه‌ی اجزای تاثیرگذار بر این فرآیندها، عامل اقلیم از اهمیت خاصی برخوردار می‌باشد، بنابراین دسترسی به سری اطلاعات داده‌های هواشناسی برای اجرای این مدل‌ها دارای اهمیت است. همچنین در بررسی و تحلیل وقایع حدی آب و هوایی مانند دوره‌های ممتد خشک و تر، دوره‌های گرما و یخبندان نیز که منجر به پیش‌بینی‌های هواشناسی می‌گردد، نیاز

\*پست الکترونیک نویسنده مسئول: mghamghami@yahoo.com

Lall et al. (2008) Bannayan and Hoogenboom اشاره نمود. (1996) رهیافتی ناپارامتری برای شبیه‌سازی دوره‌های تر و خشک و مقدار بارندگی به کار بردند که در آن، بسته به معنی‌داری یا عدم معنی‌داری همبستگی پیاپی دوره‌های تر و خشک و همبستگی متقاطع بارندگی کل دوره تر به میزان بارش روزانه، چگالی‌های مستقل یا شرطی به روش ناپارامتری برآورد هسته‌ای چگالی محاسبه شده و از این طریق، شبیه‌سازی متغیرهای مورد نظر انجام می‌شود. (1997) Rajagopalan et al. از روش مذکور برای شبیه‌سازی بارش روزانه و برای سایر متغیرها اعم از دما و تابش با پذیرش معنی‌داری همبستگی پیاپی در این متغیرها و همبستگی متقاطع با عامل بارش، اقدام به برآورد تابع چگالی احتمالی شرطی به روش ناپارامتری برای شبیه‌سازی سایر متغیرها نمودند. Brandsma and Buishand (2001, 1998) کاربرد یک فرآیند بازنمونه‌گیری نزدیکترین همسایه را برای چندین ایستگاه به طور مجزا تشریح کردند و در مطالعه‌ای دیگر قابلیت شبیه‌سازی همزمان برای چندین ایستگاه در یک حوضه‌ی آبریز را در روش مذکور ایجاد نمودند. Lall and Sharma (1999) از رهیافت ناپارامتری مبتنی بر نزدیکترین همسایه برای شبیه‌سازی وقوع دوره‌های تر و خشک و مقدار بارش روز تر استفاده نمودند. (1999) Rajagopalan and Lall نیز کارایی این روش را برای شبیه‌سازی انواع متغیرهای هواشناسی در قبال یک مدل خودهمبسته مرتبه ۱ برتر ارزیابی کردند. (2002) Sharma and O'Neill داده‌های بارش ماهانه را با روش ناپارامتری برآورد هسته‌ای شبیه‌سازی نمودند. بدین صورت که بارندگی هر ماه را وابسته به میزان بارندگی ماه قبل و مجموع بارندگی ۱۲ ماه قبل دانستند. (2006) Mehrotra et al. کارایی سه رهیافت چندمکانی تولید داده‌های بارش، شامل روش برازش توزیع نمایی ترکیبی (روش ویلکز)، زنجیرمارکف پنهان و k نزدیکترین همسایه را آزمون نموده و نتایج آنها روش ویلکز و KNN را در یک سطح کارایی اثبات نمود. Bannayan and Hoogenboom (2008) از روش KNN برای ارائه‌ی پیشگویی‌های هواشناسی استفاده نمودند. روش نزدیک‌ترین همسایه‌ی مورد استفاده در این مطالعات، از نوع کلاسیک می‌باشد که با وجود شبیه‌سازی مناسب داده‌ها، قادر به تولید مقادیر و یا بردارهای جدید (که در سری تاریخی مشاهده نشوند) نمی‌باشد. اهمیت این موضوع در شبیه‌سازی وقایع حدی آب و هوا (Extreme Events) می‌باشد، چرا که ممکن است پدیده‌ای دارای دوره بازگشت بالا باشد و در سری تاریخی مشاهده نگردد. اما در سری‌های طولانی‌تر داده‌ها حتماً به وقوع می‌پیوندند و باید توسط مدل شبیه‌ساز ساخته شود. در مطالعه‌ای

شده استفاده می‌گردد، و یا متغیر نرمال استاندارد متناظر با آن، به دلیل وجود همین بخش تصادفی، به این نوع مدل‌سازی، مدل‌سازی استوکاستیک (Stochastic Modeling) نیز می‌گویند. در مدل‌سازی متغیرهای هواشناسی به طور کل از دو نوع تکنیک پارامتری و ناپارامتری استفاده شده است، که منجر به توسعه‌ی دو رهیافت پارامتری و ناپارامتری در ساختار مولدهای هواشناسی می‌گردد. مولدهای با ساختار پارامتری قدمت بیشتری نسبت به انواع ناپارامتری دارند و ویژگی مشترک همه آنها پذیرش الگوهای آماری خاص در بحث تراکم داده‌ها می‌باشد. مشهود است که بخشی از خطا با پذیرش شرط تبعیت از نوعی تابع آماری خاص، در ابتدای کار وارد روند شبیه‌سازی می‌شود. چرا که اولاً ممکن است تابع احتمالاتی پیش‌فرض لزوماً بهترین منحنی بیانگر تراکم داده‌ها در سری نباشد و ثانیاً ممکن است یک تابع توزیع برای منطقه‌ای خاص بهترین منحنی باشد اما برای مناطق دیگر از لحاظ معیارهای خطا نامطلوب باشد. به همین دلیل اکثر روش‌های پارامتری برای مناطق خاصی ارائه شده‌اند و باید برای مناطق دیگر واسنجی شوند. از طرفی دیگر این سوال پیش می‌آید که آیا می‌توان در روند مدل‌سازی سری داده‌های هواشناسی و هیدرولوژیکی، از روش‌هایی استفاده نمود که در آن وجود پیش‌فرض‌ها به حداقل برسد و یا به عبارتی دیگر شبیه‌سازی داده‌ها مستقیماً از طریق خود سری مشاهده شده صورت گیرد؟ (Lall, 1995; Ashrafzadeh, 2005). پاسخ محققین به این سوال منجر به ابداع و توسعه روش‌های ناپارامتری تولید داده طی دو دهه اخیر گردیده است. روش‌های ناپارامتری تولید داده‌های هواشناسی برخلاف انواع پارامتری آن، صحت هیچ‌گونه فرضی را از قبل در شبیه‌سازی داده‌ها دخیل نمی‌کنند. در این نوع مولدها، همه یا بخشی از داده‌های موجود در سری مشاهده شده، در شبیه‌سازی مقدار یا بردار داده‌ها مشارکت می‌نمایند (Lall and Sharma, 1999). در واقع می‌توان گفت که هر چه ساختار مولد هواشناسی از تکنیک‌های پارامتری فاصله گرفته و به تکنیک‌های ناپارامتری نزدیکتر باشد، انتقال واقعیات موجود در سری مشاهده شده به سری شبیه‌سازی شده به طور نسبی بیشتر می‌گردد. همچنین این روش‌ها می‌توانند در هر منطقه بهترین یا نزدیک به بهترین جواب را داشته باشند. در مورد مطالعات صورت گرفته در زمینه روش‌های ناپارامتری تولید داده‌های هواشناسی می‌توان به کار Lall et al. (1996) Rajagopalan et al. (1996) Brandsma and Buishand (1998, 2001) Lall and Sharma (1999)، (1999) Rajagopalan and Lall Sharma and O'Neill (1999)، (2002) Mehrotra et al. (2006) Sharif and Burn (2006) و

می‌گردد، سپس فرض می‌گردد که وقایع موجود در آن بازه از توزیع یکنواخت تبعیت می‌نماید، بنابراین عددی تصادفی در بازه  $[0, 1]$  تولید شده و برابر با تابع توزیع تجمعی یکنواخت در نظر گرفته می‌شود و واقعه‌ی معادل این احتمال تجمعی به عنوان داده تولید شده، فرض می‌گردد. همچنین انتخاب بازه‌ها بر اساس ویژگی‌های موجود در داده‌های هواشناسی می‌باشد. برای طول سری‌های تر و خشک و ارتفاع بارندگی، اندازه فواصل تدریجاً با افزایش  $i$  زیاد می‌شود. چرا که مقادیر کوچک طول سری‌ها و میزان بارندگی، اغلب بخش وسیعی از مجموعه داده‌ها را در بر می‌گیرند و این نوع انتخاب از توان تفکیک اندک برای این مقادیر جلوگیری می‌نماید. در مدل‌سازی داده‌های درجه حرارت روزانه نیز از توزیع آماری نرمال مشروط بر خشک یا تر بودن روز معین استفاده می‌شود و پارامترهای روزانه توزیع برای هر ماه مجزا با تابع سری فوریه ۳ هارمونیک هموار می‌گردد.

#### تشریح ساختار مدل KNN

الگوریتم نزدیکترین همسایه شامل انتخاب تعدادی مشخص از بردار داده‌ها و سپس بازنمونه‌گیری تصادفی از آن مجموعه برای شبیه‌سازی دوره‌ی زمانی متعاقب یک دوره‌ی معین می‌باشد. در ابتدای کار با تعریف یک پنجره متحرک حول روز  $t$ ، همه همسایگانی که می‌توانند در شبیه‌سازی بردار روز  $t+1$  مشارکت داشته باشند تعیین می‌گردند. سپس با تعیین نزدیکترین همسایگان از طریق برآورد فاصله ماهالانبیس (اقتباس از Davis, 1986) و برازش یک توزیع احتمالاتی گسسته یا توزیع ناپارامتری بر آنها، بردار روز متعاقب یکی از همسایگان به روش تصادفی انتخاب و وارد مرحله‌ی آشفتنگی می‌شود. روند گام به گام شبیه‌سازی به شرح زیر است (Sharif and Burn, 2006):

فراخوانی ماتریس داده‌ها از فایل هواشناسی که سطور ماتریس، طول دوره‌ی آماری روزانه و ستون‌های آن متغیرهای هدف می‌باشند که در این مطالعه ۴ متغیر شامل دمای کمینه و بیشینه، رطوبت نسبی و بارندگی مورد نظر است ( $p=4$ ):

$$X_{i,p} = [x_{i,1} \quad x_{i,2} \quad \dots \quad x_{i,p}]$$

- تعیین عرض پنجره متحرک و تولید ماتریس همسایگان روز  $t$  ( $L$ )، بنابراین این ماتریس دارای ابعاد  $L * p$  می‌باشد.

- محاسبه ماتریس کوواریانس  $C_t$  برای هر روز بر اساس ماتریس همسایگان

- انتخاب اولیه مقادیر بردار (فرضاً یک ژانویه) و شروع فرآیند شبیه‌سازی ( $t = 1$ )

- محاسبه فاصله‌ی ماهالانبیس (Davis, 1986) براساس دو بردار  $X_t$  (بردار روز معین)

و  $X_m$  که  $L : m = 1$  (بردارهای همسایه):

که توسط Sharif and Burn (2006) صورت گرفت، قابلیت در مدل به کار رفت که نقص مذکور را برطرف می‌نماید و به عبارتی روش توسعه‌یافته‌ی آنها قابلیت برون‌یابی داده‌های مشاهده شده را دارا می‌باشد. در این مطالعه نیز عملکرد رهیافت ناپارامتری توسعه‌یافته‌ی  $k$  نزدیکترین همسایه (Sharif and Burn, 2006) با قابلیت برون‌یابی داده‌ها در سری ساختگی، تشریح و برای چندین ایستگاه با آمار مشاهداتی قابل قبول با هدف شبیه‌سازی چهار متغیر بارندگی، دمای بیشینه، دمای کمینه و رطوبت نسبی در مقیاس روزانه اجرا گردیده است و برای درک بهتر از عملکرد روش مذکور، برخی نتایج آن با خروجی مولد LARS-WG مقایسه شده است.

#### روش انجام تحقیق

به طور کل روش‌های  $k$  نزدیکترین همسایه، زیرمجموعه‌ی روش‌های ناپارامتری خودگردان ساز (Bootstrapping) می‌باشند. این نام به روند بازنمونه‌گیری با جایگزینی از مجموعه‌ی مشاهده شده منتسب است. نمونه‌گیری مکرر در روش KNN از همسایگان روز معینی صورت می‌گیرد که قرار است بردار روز متعاقب آن شبیه‌سازی شود و معیار آن نیز تعیین فواصل هر کدام از همسایگان با روز معین می‌باشد. روش به کار رفته در این تحقیق علاوه بر دارا بودن ویژگی الگوریتم مشابه با همه روش‌های نزدیکترین همسایه در روند شبیه‌سازی، دارای توانایی تولید مقادیری خارج از سری مشاهده شده می‌باشد که با اضافه شدن زیرروال Perturbation در مدل، این قابلیت عملیاتی شده است.

#### تشریح مولد LARS-WG

مدل LARS-WG از سری مولدهای پارامتری ترتیبی تشریح شده توسط Racsco et al. (1991) و Semenov and Barrow (1997) می‌باشد. این مولد در ساختار خود از توزیع‌های نیمه‌تجربی و نرمال برای مدل‌سازی متغیرهای هواشناسی استفاده می‌کند و به همین دلیل در برخی از منابع، از این مولد به عنوان روش نیمه پارامتری یاد شده است. طول سری‌های روزانه‌ی تر و خشک، مقدار بارش روزانه با توزیع نیمه تجربی و دماهای بیشینه و کمینه با توزیع نرمال از سری داده‌های تاریخی مدل می‌شوند. یک توزیع نیمه تجربی یک هیستوگرام با ۱۰ بازه می‌باشد  $\{a_{i-1}, a_i; h_i; i = 2, \dots, 10\}$  به ترتیب کران‌ای پایین و بالای بازه  $i$  ام و  $h$  تعداد وقایع از داده‌ای مشاهده شده در بازه معین می‌اشد. به طور کلی روند تولید با یک مولد نیمه تجربی بدین صورت می‌اشد که در ابتدا به طور تصادفی یک بازه، براساس فراوانی نسبی وقایع موجود در هر بازه به عنوان احتمالات آستانه‌ی یا احتمالات بحرانی انتخاب

$$\lambda^{new} = -x_{\varphi,t+1} / (z_{\alpha} \times \sigma_{\varphi}) \quad (۴)$$

که اندیس ۴ اشاره به متغیر بارندگی دارد. از مقدار عرض‌باند جدید در رابطه (۳) برای شبیه‌سازی استفاده می‌شود. بردار شبیه‌سازی شده به عنوان ورودی مرحله ۵ مدنظر است و مراحل ۵ تا ۱۱ برای تولید سری با طول مکفی تکرار می‌گردد.

#### تعیین پارامترهای مولد KNN

در این مطالعه عرض پنجره متحرک ۷ در نظر گرفته شد که با توجه به طول آماری ۴۵ سال، طول بردار همسایگان برای هر متغیر برابر با  $674 = 45 \times 15 - 1$  می‌باشد. همچنین از روش سعی و خطا برای تعیین تعداد نزدیکترین همسایه‌ها استفاده می‌شود. اما Sharma and Lall (1996) و Sharif and Burn (2006, 2007) از روش جذر طول همسایگان ( $L$ ) استفاده نمودند که  $k = \sqrt{L}$ . در این روش نیز ما به این طریق بسنده نمودیم. همچنین برای برآورد عرض باند بهینه از روش ارائه شده توسط Silverman (1986) استفاده شده است که مقدار آن در حدود ۰/۵ برآورد گردید. احتمال آستانه‌ای برای به حداقل رساندن مقادیر منفی بارندگی نیز با روش سعی و خطا مقادیر ۰/۰۶ و ۰/۱ حسب ایستگاه‌های مختلف برآورد گردید.

#### مطالعه موردی

مشخصات جغرافیایی ایستگاه‌های مطالعاتی در جدول (۱) آمده است. طول دوره آماری تاریخی، ۴۵ سال (۱۹۶۱-۲۰۰۵) می‌باشد و برای بررسی عملکرد رهیافت مورد بحث، ۱۰۰ سال داده روزانه شبیه‌سازی شد.

جدول ۱- مشخصات جغرافیایی ایستگاه‌های مطالعاتی

ایستگاه‌ها	مشخصات جغرافیایی				
	قزوین	تبریز	رشت	بوشهر	مشهد
طول	۵۰/۰۱	۴۶/۲۹	۴۹/۵۹	۵۰/۸۳	۵۹/۵۹
عرض	۲۶/۲۶	۳۸/۰۸	۳۷/۲۷	۲۸/۹۷	۳۶/۳۰
ارتفاع	۱۲۹۰	۱۳۹۵	۱	۹	۱۰۲۲

#### نتایج و بحث

برای ارزیابی مولد ناپارامتری KNN آزمون‌های متعددی انجام شد. از خط نیمساز برای بررسی چشمی کارایی روش‌های مورد مطالعه استفاده شده است. آزمون T-Student برای بررسی معنی‌داری آماری اختلافات میانگین‌ها و آزمون F برای بررسی معنی‌داری اختلافات واریانس‌ها استفاده شده است. مولد ناپارامتری مورد بحث قادر به شبیه‌سازی همزمان ۴ متغیر مقدار بارش، دمای بیشینه و کمینه و رطوبت نسبی و مولد پارامتری مورد بحث قادر به شبیه‌سازی متغیرهای بارش، دمای بیشینه و

$$D_m = \sqrt{(X_t - X_m)C_t^{-1}(X_t - X_m)^T} \quad (۱)$$

صعودی نمودن این فواصل و تعیین  $k$  تعداد از نزدیکترین بردارها به بردار روز معین ( $k: j=1$ ).

یک توزیع احتمالاتی گسسته را بر این مقادیر همسایه برازش داده می‌شود. به طوری که هرچه مشابهت بین دو بردار اصلی و همسایه بیشتر باشد براساس توزیع مذکور وزن ( $W_i$ ) بیشتری هم به خود می‌گیرد. همچنین احتمال جمع‌ی هر همسایه ( $P_j$ ) محاسبه می‌گردد (Sharif and Burn, 2006):

$$W_i = \frac{1/j}{\sum_{i=1}^k 1/j}, P_j = \sum_{i=1}^j W_i \quad (۲)$$

تولید یک عدد تصادفی ( $r$ ) بین صفر و یک و بر اساس توزیع یکنواخت. اگر  $r < p_j$  باشد روز متناظر با اندیس  $j$  از سری زمانی اصلی انتخاب و بردار روز متعاقب آن به عنوان بردار  $X_{t+1}$  لحاظ می‌شود.

بردار تولید شده می‌تواند به کار گرفته شود اما برای تولید مقادیری خارج از سری مشاهداتی، به طور حتم باید زیرروالی به روش اضافه شود. در ابتدا یک توزیع ناپارامتری برآوردگر هسته‌ای (Sharma et al., 1997; Sharma and O'Neill, 2002; Sharif and Burn, 2006) بر نزدیکترین همسایگان برازش می‌یابد که شامل تخمین انحراف معیار شرطی ( $\sigma_j$ ) و پارامتر عرض باند ( $\lambda$ ) می‌باشد.

با تولید تصادفی متغیر  $Z_{t+1}$  از توزیع نرمال استاندارد با میانگین صفر و انحراف معیار یک مقدار جدید برای روز متعاقب روز  $t$  از رابطه زیر به دست می‌آید:

$$y_{i,t+1} = x_{i,t+1} + \lambda \sigma_i Z_{t+1} \quad (۳)$$

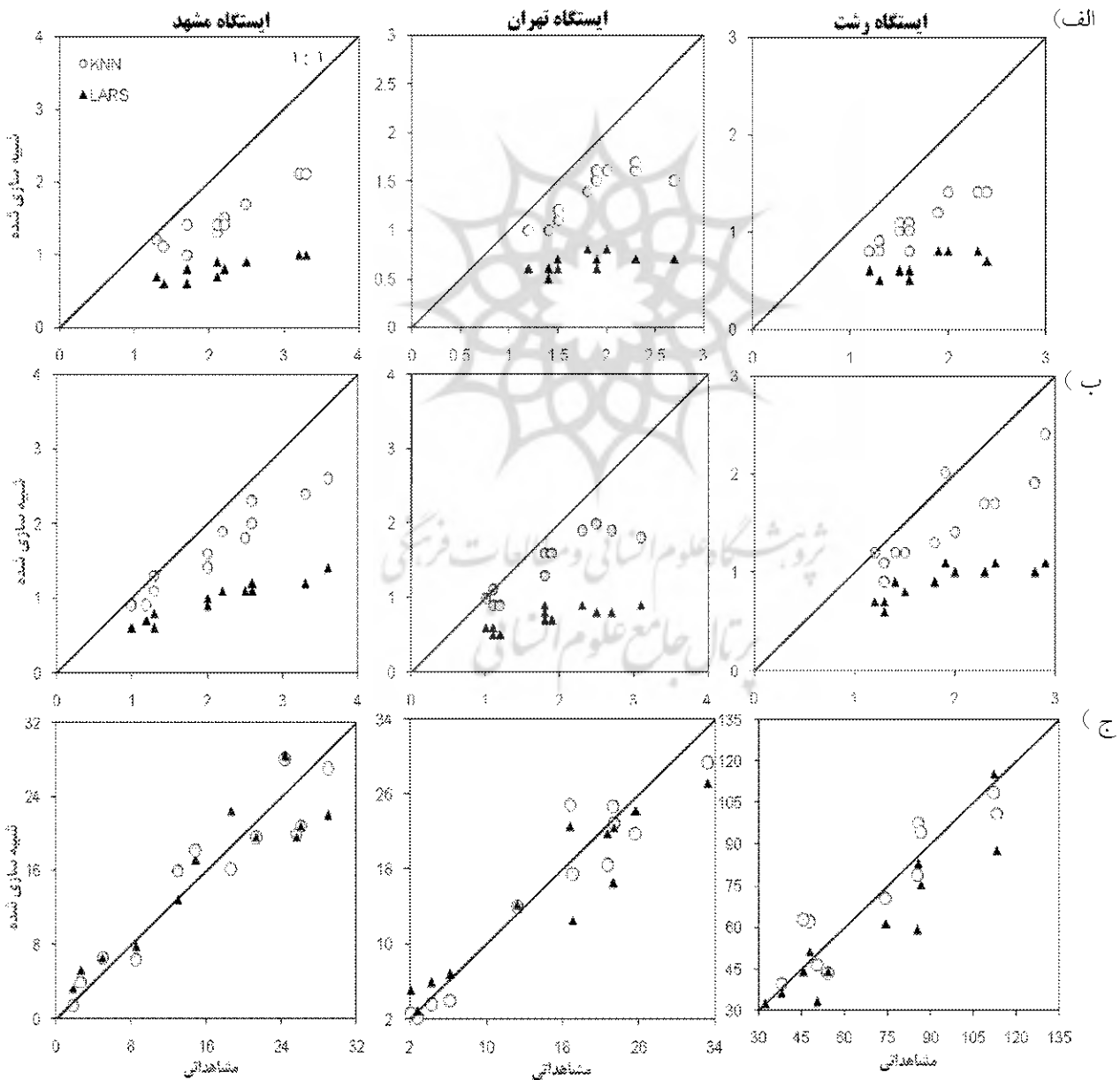
که  $i$  شاخص متغیر و  $\lambda$  نیز مقدار شبیه‌سازی شده برای روز  $t+1$  با اعمال قابلیت آشفستگی در داده‌ها می‌باشد.

با توجه به اینکه مقادیر بارش مرزی هستند، اجرای مراحل ۹ و ۱۰ ممکن است منجر به تولید مقادیر منفی بارش شود که غیرقابل قبول می‌باشد. در برخی از مولدها این مقادیر صفر در نظر گرفته می‌شوند که البته ممکن است منجر به بیش برآوردگی در مقادیر ماهانه این متغیر شود (Sharif and Burn, 2006). برای غلبه بر این مساله باید احتمال وقوع مقادیر منفی به حداقل برسد که از طریق تعریف یک مقدار احتمال آستانه‌ای ( $\alpha$ ) و تغییر مقدار پارامتر عرض باند اعمال می‌شود.  $\alpha$  بزرگترین احتمال منفی شدن مقدار مورد نظر ما می‌باشد که با روش سعی و خطا به دست می‌آید و متغیر نرمال استاندارد متناظر با آن  $Z_{\alpha}$  محسوب می‌شود. به این ترتیب مقدار جدید  $\lambda$  از طریق زیر به دست می‌آید:

ماهانه دما و رطوبت) و گاما (بر داده‌های ماهانه بارندگی) مشاهده گردید که خروجی مولد KNN تطابق بیشتری از نظر فراوانی نسبی و احتمال وقوع پدیده‌ها به سری‌های مشاهده‌ای دارد و نیز اثر پدیده کم برآوردگی واریانس‌های ماهانه در مورد دو مولد در شکل (۴) مشاهده می‌شود. برای بیان مناسب‌تر تطابق احتمال وقوع پدیده‌ها در سری‌های مصنوعی با سری‌های مشاهده‌ای در این مطالعه از تستی به نام LEPS SCORE استفاده شد. این تست که بیانی از میزان خطای شبیه‌سازی شده یا پیش‌بینی شده در فضای احتمالاتی است، مهارت کلی روش مورد نظر را ارزیابی می‌نماید (Araghinejad, 2006; Ward and Folland, 1991; Potts et al., 1996).

کمینه می‌باشد. هر دو مولد در شبیه‌سازی میانگین‌های ماهانه به خصوص برای متغیرهای دما و همچنین KNN برای رطوبت نسبی به خوبی عمل می‌نمایند و در همه موارد اختلافات در سطح ۵ درصد معنی‌دار نمی‌باشد (شکل ۲). نیز این نتیجه در مورد بارندگی در بیشتر موارد حاصل شده است.

اما در مورد واریانس‌های ماهانه برای متغیرهای دما و رطوبت، در هر دو مولد پدیده کم برآوردگی مشاهده شد که البته نتایج مولد KNN قابل اطمینان‌تر و به واقعیت نزدیک‌تر می‌باشد (شکل ۱). در مورد پارامترهای روزانه نیز روش ناپارامتری، مقادیر میانگین و واریانس‌های روزانه را به خوبی تولید نموده و البته در پارامترهای روزانه بارندگی پراکنش بیشتری مشهود است (شکل ۳). با برازش توابع توزیع تجمعی نرمال (بر داده‌های



شکل ۱- بررسی میزان تطابق انحراف معیارهای ماهانه (الف) دمای کمینه، (ب) دمای بیشینه و (ج) بارندگی

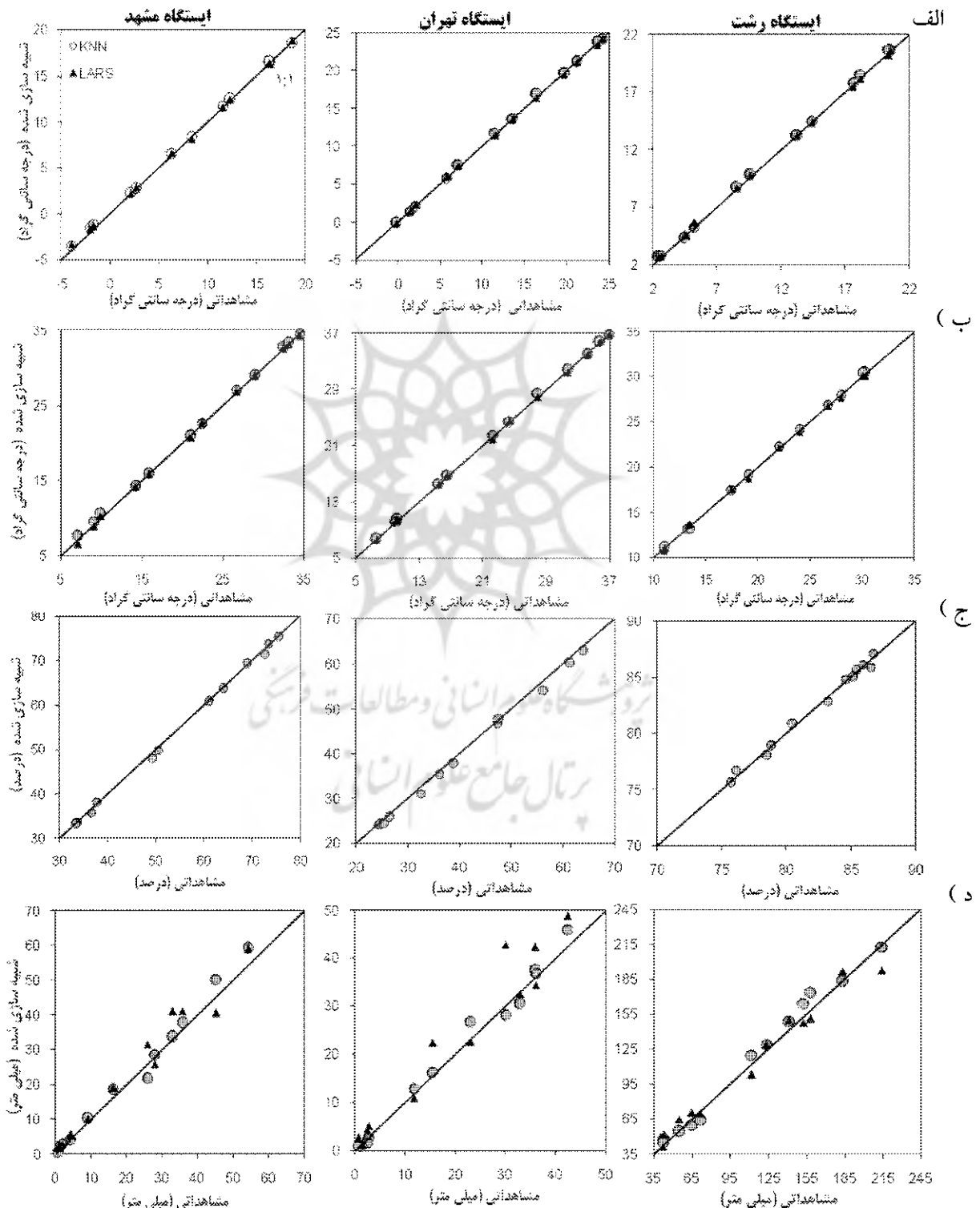
می‌باشد، بدین معنا که هرچه تطابق احتمالاتی بیشتر باشد، این مقدار به ۲ نزدیکتر است و بالعکس. در نهایت مهارت کلی مدل موردنظر از رابطه زیر به دست می‌آید:

$$SK = \frac{\sum 100S}{\sum S_m} \quad (۶)$$

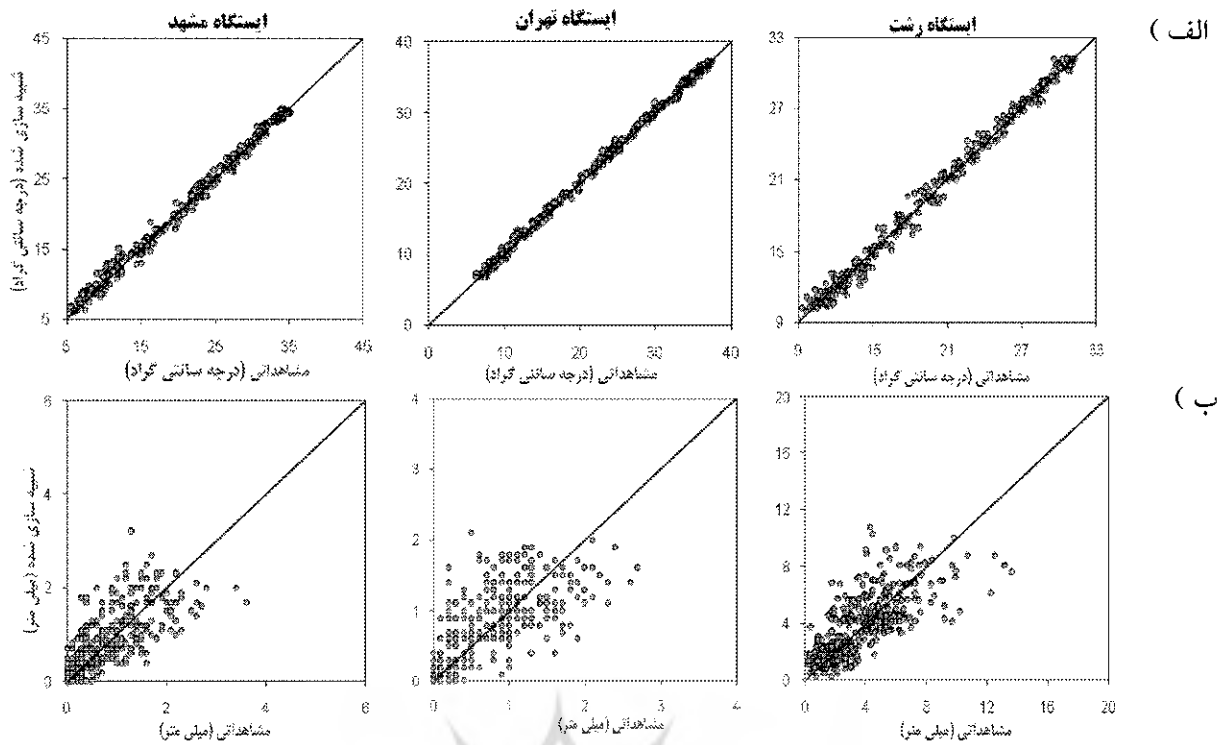
امتیازدهی در این روش به صورت زیر تعریف می‌شود:

$$S = 3(1 - |F_s - F_o| + F_s^y - F_s + F_o^y - F_o) - 1 \quad (۵)$$

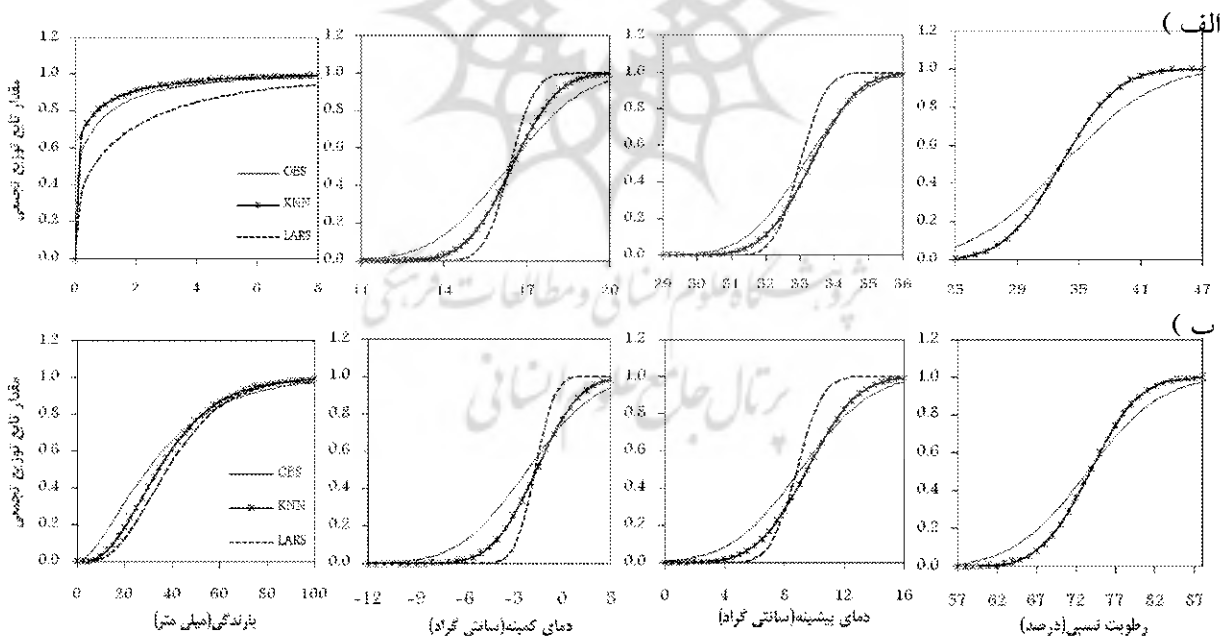
که  $F_o$ ,  $F_s$  به ترتیب مقادیر تابع توزیع تجمعی شبیه‌سازی شده و مشاهده شده در یک نقطه‌ی معین و  $S$  مقداری بین -۱ تا ۲



شکل ۲- بررسی میزان تطابق میانگین های ماهانه الف) دمای کمینه، ب) دمای بیشینه، ج) رطوبت نسبی و د) بارندگی



شکل ۳- بررسی میزان تطابق مانگین های روزانه مولد KNN الف) دمای، ب) بیشینه، ب) بارندگی



شکل ۴- نمودار تابع توزیع تجمعی مقادیر ماهانه برای ایستگاه مشهد الف) ماه اوت ، ب) ماه فوریه

۱۰۰- می شود، که هرچه به عدد ۱۰۰ نزدیک تر باشد، کارایی روش شبیه سازی یا پیش بین کننده بیشتر می باشد. در این مطالعه پس از شبیه سازی با دو مولد مورد نظر، با فرض تبعیت داده های ماهانه دما و رطوبت از تابع توزیع نرمال و تبعیت داده های ماهانه بارندگی از تابع توزیع گاما، این امتیاز دهی برای هر دو روش صورت گرفت که نتایج آن در جدول شماره (۲) برای

$S_m$  بستگی به مثبت یا منفی بودن مقدار  $S$  متناظر دارد. اگر  $S$  مثبت باشد برای محاسبه  $S_m$  در رابطه (۵) مقدار تابع توزیع تجمعی مشاهداتی را معادل شبیه سازی شده قرار می دهیم و برای  $S$  منفی، اگر مقدار CDF مشاهداتی بزرگتر از ۰/۵ باشد، در رابطه (۵) معادل صفر و در غیر این صورت معادل یک قرار می گیرد. طبق این روش، مقدار  $SK$  بین ۱۰۰ تا

ایستگاه تبریز به طور خاص آمده است. مشاهده شد که مولد KNN با بازسازی بهتر پارامترهای منحنی‌های آماری در اغلب موارد امتیاز بالاتری نسبت به مولد LARS دارا می‌باشد و مولد LARS نیز در برخی موارد در مورد متغیر بارندگی خوب عمل می‌نماید.

جدول ۲- توانایی کلی دو مولد در میزان تطابق منحنی‌های احتمال براساس آزمون LEPS SCORE

ماه	متغیر	فوریه	می	اوت	نوامبر
		LARS-WG	KNN-WG	LARS-WG	KNN-WG
دما	کمینه	۷۴/۵	۹۱	۸۶/۸	۸۹/۱
دما	بیشینه	۶۵/۶	۸۱/۵	۷۳/۸	۸۱/۹
رطوبت	نسبی	-	۷۸/۱	-	۸۴/۲
بارندگی		۹۵	۹۲	۹۲/۵	۹۴/۷

بازگشت‌های ۵۰ ساله یا ۱۰۰ ساله که احتمال وقوع آنها در سری آماری طولانی قریب می‌باشد. این قابلیت در مولد KNN با اضافه نمودن زیرروال Perturbation فراهم گردید. در جدول (۳) بزرگترین مقادیر روزهای خاصی از سال در طول دوره‌های تاریخی و تولیدی آمده است که مشاهده می‌گردد در اکثر موارد بزرگترین مقادیر روزانه بارندگی در روز معین سال در سری شبیه سازی شده بزرگتر از سری مشاهداتی می‌باشد.

خاصیت دیگر روش KNN توسعه یافته نسبت به نوع کلاسیک آن (Bannayan, M. and Hoogenboom, 2008)، توانایی آن در شبیه سازی داده‌هایی خارج از سری مشاهداتی می‌باشد و این مساله زمانی اهمیت پیدا می‌کند که تولید سری‌های با طول آماری بیشتر از مشاهداتی (مثلاً ۱۰۰ سال به بالا) مدنظر باشد. تحت این شرایط مولد باید قابلیت شبیه‌سازی وقایع با دوره بازگشت بالا را داشته باشد، یعنی مثلاً بارندگی‌های با دوره

جدول ۳- مقادیر روزانه بیشینه بارندگی در روزهای معینی از سال (میلی متر)

نام ایستگاه	شماره روز سال دوره آماری	۱	۳۱	۶۱	۹۱	۱۲۱	۱۵۱	۱۸۱	۲۱۱
تهران	طی ۴۵ سال	۱۷	۵	۴۹	۲۴	۲	۲	۰	۰/۶
تهران	طی ۱۰۰ سال	۱۷/۶	۳۹	۲۹/۹	۳۵/۵	۲۰	۲/۶	۰/۲	۲
تهران	طی ۴۵ سال	۴۴	۳۱	۵۰	۴۸	۱۰	۱۵	۲۲	۷/۱
تهران	طی ۱۰۰ سال	۵۰/۷	۶۱/۴	۳۴/۳	۴۶/۸	۶۹/۹	۲۴/۷	۶۷/۸	۹۵/۳

دما، بیشینه با کمینه و رطوبت با بارندگی می‌باشد. مثلاً مقدار ضریب همبستگی متقاطع بین دمای بیشینه و کمینه برای ماه نوامبر در سری مشاهده شده معادل ۰/۳۸، در سری مصنوعی KNN معادل ۰/۳۲ و در سری مصنوعی LARS-WG معادل ۰/۲- می‌باشد. نیز در مورد دما و بارندگی، این ضرایب در بیشتر ماه‌ها در سری مشاهده شده دارای معنی داری نبوده است. مشابه این نتایج در کار Sharif and Burn (2006) نیز مشهود می‌باشد.

**نتیجه گیری نهایی و پیشنهادات**

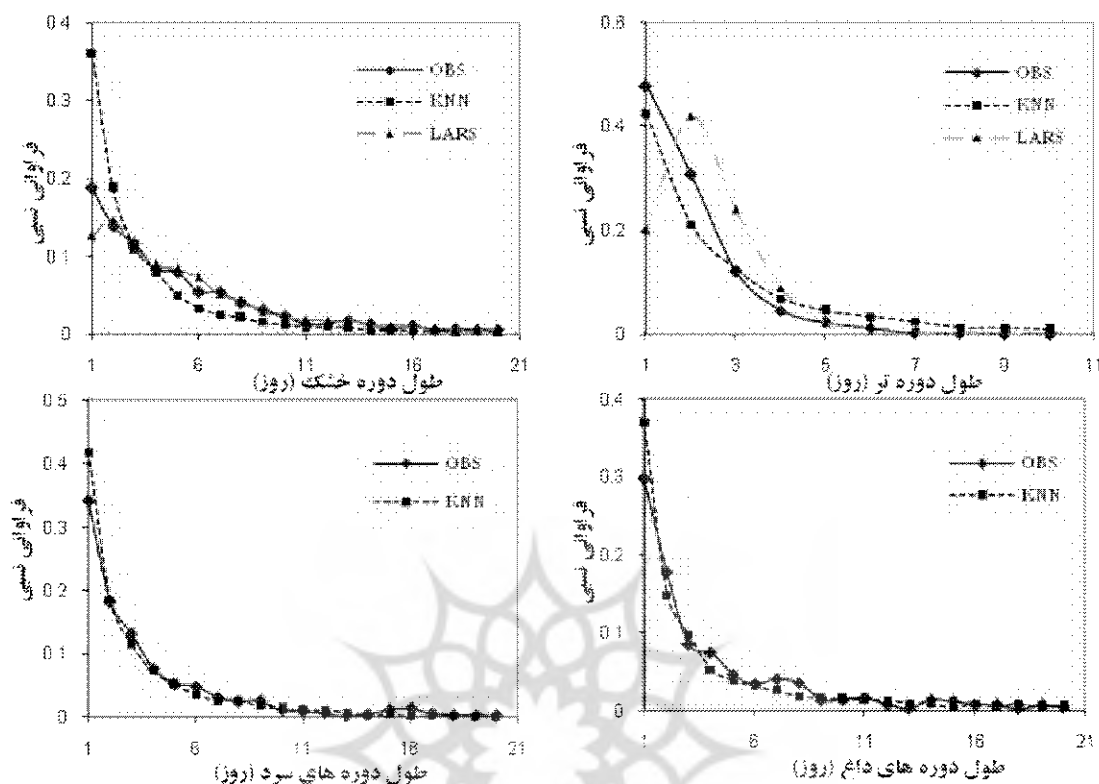
در مجموع مولد ناپارامتری توسعه یافته k نزدیکترین همسایه تقریباً قادر به شبیه‌سازی مناسب واقعیات موجود در سری مشاهده شده می‌باشد که دلیل اصلی آن عدم پذیرش پیش فرض‌هایی در ابتدای کار است. تولید همزمان و با کیفیت چندین متغیر و حفظ همبستگی‌های بین متغیرها از مزایای این

نکته قابل تامل دیگر که در اکثر مطالعات جای بحث دارد، شبیه‌سازی تداوم دوره‌های سرد - گرم و تر - خشک می‌باشد. شکل (۵) به خوبی نشان می‌دهد که مولد KNN قادر به شبیه سازی فراوانی وقوع طول دوره‌های گرم و سرد (به خصوص درازمدت) می‌باشد. همچنین در شبیه سازی فراوانی نسبی طول دوره‌های تر و خشک همان طور که در همان شکل نشان داده شده است، می‌بینیم که مولد LARS-WG در مورد دوره‌های درازمدت بهتر عمل می‌نماید که علت اساسی آن مدل سازی مجزای این پارامتر در روند تولید می‌باشد و اساساً برتری آن نسبت به روش‌های پارامتری نوع ریچاردسون نیز کارا بودن آن در این زمینه است. به هر حال اختلاف این دو روش در شبیه سازی ترسالی‌ها و خشکسالی‌ها در مقایسه با مشاهداتی ناچیز و قابل اغماض می‌باشد. همچنین دریافت شد که مولد KNN قادر به حفظ ضرایب همبستگی متقاطع متغیرها به خصوص بین



اساس مطالعات مختلف، برترین مولد پارامتری بوده است، دارای برتری نسبی است و برای استفاده در کاربری‌های گوناگون مولدهای هواشناسی، توصیه می‌گردد.

روش می‌باشد و باعث شده است که این مولد برخلاف مولدهای نوع پارامتری، یک مولد چندمتغیره (Multivariate) محسوب شود. همچنین این مولد در مقایسه با مولد LARS-WG که بر



شکل ۵ - فراوانی نسبی تداوم دوره های با ویژگی مشترک

روبرو می‌باشد. بنابراین باید گامی در جهت شبیه سازی همزمان دنباله‌های ساختگی برای چندین ایستگاه در سطح یک حوزه برداشت. لازمه توسعه چنین مولدی، نیل به تکنیکی در مولد می‌باشد که بتواند همبستگی‌های مکانی را در صورت معنی‌داری در شبیه سازی دخالت دهد که چنین مولدی به طور حتم یک مولد Multi-site محسوب می‌شود.

با این وجود این مولد، تنها قادر به شبیه سازی سری مصنوعی برای یک ایستگاه می‌باشد که یک مولد Single-site محسوب می‌گردد. در بسیاری از تحقیقات که مقیاس مطالعاتی در سطح یک حوزه آبریز می‌باشد، اغلب نیاز به دسترسی مناسب به سری‌های درازمدت هواشناسی در چندین ایستگاه وجود دارد که البته اجرای این مولد به صورت تک ایستگاهی با مشکل

REFERENCES

Araghinejad, S., Burn D. H., and Karamouz M. (2006). Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors, *Water Resour. Res.*, 42, W03431, doi:10.1029/2004WR003853.

Ashrafzadeh, A. (2005). *Streamflow simulation using nonparametric methods*. Ph. D. dissertation, University of Tehran. (In Farsi)

Bannayan, M., and Hoogenboom, G. (2008). Predicting realization of daily weather data for climate forecasts using the non-parametric nearest-neighbor re-sampling technique. *International Journal of Climatology*, 28 (10), 1357-1368.

Brandsma, T., and Buishand, T.A. (1998). Simulation of extreme precipitation in the Rhine basin by nearest neighbor resampling. *Hydrology and earth System Sciences.*, 2 (2-3), 195-209.

Buishand, T. A., and Brandsma, T. (2001). Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest neighbor resampling. *Water Resources*, 37 (11), 2761-2776.

Davis, J. (1986). *Statistics and data analysis in geology*. Wiley, New York.

Lall, U. (1995). *Nonparametric function estimation. Recent Hydrologic Applications*, us national reports, 1991-1994.

Lall, U., Rajagopalan, B., and Tarboton, D.G. (1996). A nonparametric wet/dry spell model for re-sampling daily precipitation. *Water Resources*, 32 (9), 2803-2823.

Mehrotra, R., Srikanthan, R., and Sharma, A. (2006). A comparison of three stochastic multi-site

- precipitation occurrence generators. *Journal of Hydrology*, 331 (1-2), 280-292.
- Potts, J.M., Folland, C.K., Jolliffe, I.T., and Sexton, D. (1996). Revised LEPS scores for assessing climate model simulations and long-range forecasts. *Journal of Climate*, 9, 34-53.
- Racsko, P., Szeidl, L., and Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological Modeling*, 57, 27-41.
- Rajagopalan, B., and Lall, U. (1999). A k-nearest neighbor simulator for daily precipitation and other weather variables. *Water Resources research*, 35, 3089-3101.
- Rajagopalan, B., Lall, U., and Tarboton, D. G. (1997). Evaluation of kernel density estimation methods for daily precipitation re-sampling. *Stochastic Hydrology and Hydraulics*, 11, 523-547.
- Rajagopalan, B., Lall, U., Tarboton, D. G., and Bowles, D. S. (1997). Multivariate nonparametric re-sampling scheme for generation of daily weather variable. *Stochastic Hydrology and Hydraulics*, 11, 65-93.
- Semenov, M. A. (2008). Simulation of extreme weather events by a stochastic weather generator. *Climate Research*, 35, 203-212.
- Semenov, M. A., and Barrow, E. M. (1997). Use of stochastic weather generator in the development of climate change scenarios. *Climatology*, 35, 397-414.
- Sharif, M., and Burn, D.H. (2006). Simulating climate change scenarios using an improved k nearest neighbor model. *Journal of Hydrology*, 325, 179-196.
- Sharma, A., and Lall, U. (1999). A nonparametric approach for daily rainfall simulation. *Mathematics and computer in simulation*, 48, 361-371.
- Sharma, A., and O'neil, R. (2002). A nonparametric approach for representing inter-annual dependence in monthly streamflow sequences. *Water Resources*, 38(7), 5-10.
- Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, New York.
- Ward, M. N., and Folland, C.K. (1991). Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology*, 11, 711-743.

