نشریه هواشناسی کشاورزی
(وابسته به انجمن مهندسی آبیاری و آب ایران)
جلد ۱۰، شماره ۲، پاییز و زمستان ۱۴۰۱، صص.۳۶-۲۸
مقاله علمی پژوهشی

Journal of Agricultural Meteorology
(Iranian Soc. of Water and Irrig. Eng.)
Vol. 10, No. 2, Autumn & Winter 2023, pp. 28-36
DOI: 10.22125/agmj.2022.315265.1126

مقایسه کارآیی روش‌های هوشمند و آماری در بازسازی داده‌های ساعت آفتابی (مطالعه موردی: شرق حوضه دریاچه ارومیه)

وحید مونس‌خواه[۱]، محمد خالدی علمداری*[۱]، معین هادی[۱]، سعید صمدیان فرد[۲]

تاریخ دریافت: ۱۴۰۰/۰۸/۲۵
تاریخ پذیرش: ۱۴۰۱/۰۳/۲۳

چکیده

یکی از متغیرهای هواشناسی که در مطالعات اقلیمی و برآورد تبخیرتعرق اهمیت زیادی داشته و عموماً دارای خلأهای آماری نسبتاً زیادی می‌باشد، ساعات آفتابی است. در پژوهش حاضر به‌منظور بازسازی داده‌های این کمیت در ایستگاه‌های تبریز، سراب، سهند و مراغه در دوره آماری ۱۳۶۹ تا ۱۳۹۸ از روش‌های هوشمند رگرسیون ماشین بردار پشتیبان (SVR)، شبکه‌های عصبی مصنوعی (ANN) و جنگل‌های تصادفی (RF) و روش‌های آماری شامل نسبت نرمال (NR)، مختصات جغرافیایی (GC) و ضریب همبستگی وزنی (CCW) استفاده شده‌است. برای ارزیابی و مقایسه نتایج از شاخص‌های ضریب همبستگی، جذر میانگین مربعات خطا، میانگین انحرافات مطلق و دیاگرام تیلور استفاده گردید. نتایج نشان داد که در حالت کلی، روش‌های ANN و مختصات جغرافیایی به ترتیب در بین روش‌های هوشمند و آماری، بالاترین دقت را در بازسازی داده‌های ساعات آفتابی دارند. در ایستگاه‌های تبریز و سهند، روش مختصات جغرافیایی به‌ترتیب با RMSE معادل ۱/۰۴ و ۱/۱۳ ساعت، در سراب روش SVR با RMSE معادل ۱/۵۸ ساعت و در مراغه روش نسبت نرمال با RMSE معادل ۱/۴۵ ساعت، بالاترین دقت را در بازسازی داده‌های ساعات آفتابی دارند. همچنین روش RF، کمترین دقت را بازسازی داده‌های ساعت آفتابی از خود نشان داد. به عنوان یک نتیجه کلی چنین می‌توان بیان نمود که در ایستگاه‌های تبریز، سراب و سهند، هر دو دسته روش‌های هوشمند و آماری دقت تقریباً مشابهی دارند ولی در ایستگاه مراغه، روش‌های آماری برآوردهای دقیق‌تری در بازسازی داده‌های ساعات آفتابی دارند.

واژه‌های کلیدی: خلاء آماری، حوضه دریاچه ارومیه، دیاگرام تیلور، ساعات آفتابی

مقدمه

برآورد داده‌های گم شده، به‌عنوان اولین مرحله در مطالعات هیدرولوژیکی و اقلیم‌شناسی شناخته می‌شود. داده‌های گمشده به دلایلی مانند خرابی موقت دستگاه اندازه‌گیری یا تعویض ادوات، قطعی ارتباط، عدم قرائت توسط کارشناس، تغییر محل اندازه‌گیری، تغییر اشخاص قرائت کننده و پالایش داده‌ها (حذف داده‌های پرت توسط سازمان هواشناسی) ایجاد می‌گردد (Hasanpour and Dinpashoh, 2012). روش‌های مختلفی به منظور بازسازی و برآورد داده‌های گم شده وجود دارند که از جمله آن‌ها می‌توان به انواع روش‌های آماری و هوشمند اشاره کرد. امروزه استفاده از روش‌های هوش مصنوعی در بازسازی داده‌های گم شده هیدرولوژیکی و اقلیمی بسیار گسترش یافته است. به عنوان مثال Naghidi et al., (2010) برای تخمین داده‌های گم شده دبی ماهانه حوزه آبخیز کارون بزرگ، Coulibaly and Evora (2007) برای برآورد داده‌های گم شده هواشناسی و Tabari and Talaee (2015) برای بازسازی داده‌های کیفی رودخانه مارون از روش‌های هوشمند استفاده کردند. ساعات آفتابی یکی از مهم‌ترین متغیرهای اقلیمی است که برآورد دقیق آن در مطالعات هواشناسی، هیدرولوژی و کشاورزی از جمله برآورد نیاز آبی حائز اهمیت است. این در حالی است که بررسی داده‌های هواشناسی در ایستگاه‌های سینوپتیک واقع در حوضه

[۱] دانشجوی دکتری آبیاری و زهکشی، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
(*نویسنده مسئول: m.khaledi.a@gmail.com)

[۲] استادیار آبیاری و زهکشی، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

نحوه ارجاع مقاله:
مونس‌خواه، و.، خالدی علمداری، م.، هادی، م.، صمدیان فرد، س. ۱۴۰۱. مقایسه کارآیی روش‌های هوشمند و آماری در بازسازی داده‌های ساعت آفتابی (مطالعه موردی: شرق حوضه دریاچه ارومیه). نشریه هواشناسی کشاورزی، ۱۰(۲): ۳۶-۲۸. DOI: 10.22125/agmj.2022.315265.1126
Mouneskhah, V., Khaledi Alamdari, M., Hadi, M., Samadianfard, S. 2023. Comparison of the efficiency of intelligent and statistical methods in the reconstruction of sunshine hours data (Case study: East of Urmia Lake basin). Journal of Agricultural Meteorology, 10(2): 28-36. DOI: 10.22125/agmj.2022.315265.1126

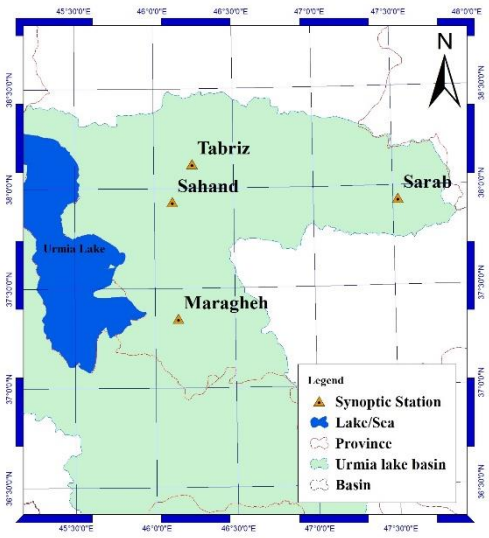
DOI: 10.22125/agmj.2022.315265.1126

دریاچه ارومیه، نشان می‌دهد که ساعات آفتابی یکی از داده‌هایی است که تعداد داده گم شده زیادی دارد. Fooladmand (2012) با استفاده از داده‌های ماهانه دماهای کمینه، بیشینه و متوسط و رطوبت نسبی، ساعات آفتابی ماهانه را در استان فارس برآورد نمود. نتایج نشان داد که معادلات به‌دست آمده برای تخمین ساعات آفتابی از دقت بالایی برای تخمین ET_0 ماهانه با استفاده از روش FAO-PM برخوردار است. Hasanpour and Dinpashoh (2012) یازده روش مبتنی بر هوش مصنوعی و کلاسیک را برای تخمین داده‌های اقلیمی گم شده در سه ناحیه اقلیمی ایران مورد بررسی قرار دادند. ایشان گزارش کردند که روش‌های مبتنی بر هوش مصنوعی دقت بیش‌تری در تخمین داده‌های گم شده دارند. Armanuos et al., (2020) نیز با بررسی ۲۱ روش کلاسیک بر روی اطلاعات ۳۴ ساله ۱۵ ایستگاه در منطقه اتیوپی به این نتیجه رسیدند که روش‌های نسبت نرمال، عکس فاصله وزن‌دار، رگرسیون خطی چندگانه، ضریب همبستگی وزنی و میانگین حسابی قابل اعتمادترین روش‌ها برای برآورد داده‌های گم شده بارش بوده و از این بین نسبت نرمال با بیشترین همبستگی و کمترین خطا نسبت به سایر روش‌ها قابل اعتمادتر می‌باشد. Bayat and Mirlatifi, (2009) با استفاده از مدل‌های رگرسیونی و شبکه‌های عصبی مصنوعی، تابش کل خورشیدی روزانه را در دو ایستگاه هواشناسی کرج (اقلیم خشک) و شیراز (اقلیم نیمه‌خشک) برآورد کردند. ایشان گزارش کردند که مدل شبکه عصبی مصنوعی با ورودی ساعات آفتابی حداکثر و تابش فرازمینی روزانه و ساعات آفتابی اندازه‌گیری شده، با ضریب همبستگی ۰/۹۴ و RMSE معادل ۲/۳۴ مگاژول بر مترمربع در روز بهترین نتیجه را ارائه داد. Behranget al., (2010) نیز طی پژوهشی گزارش کردند که استفاده از شبکه عصبی در مقایسه با معادلات تجربی، سبب بهبود نتایج تخمین تابش خورشیدی می‌گردد. Piri et al., (2013) در مطالعه‌ای به منظور مدل‌سازی تابش خورشیدی در ایستگاه‌های زاهدان و بجنورد، گزارش کردند که مدل نروفازی برآورد بهتری نسبت به روش‌های تجربی در برآورد تابش دارد. Karbasi (2016) با استفاده از دو نوع شبکه عصبی مصنوعی پرسپترون چند لایه و تابع پایه شعاعی و همچنین داده‌های هواشناسی ایستگاه هدف و ایستگاه‌های مجاور به بازسازی داده‌های مجموع ماهیانه ساعات آفتابی ایستگاه هواشناسی زنجان اقدام نمود. نتایج نشان داد که می‌توان با استفاده از داده‌های هواشناسی ایستگاه هدف و ایستگاه‌های مجاور، مجموع ماهیانه ساعات آفتابی را با دقت بالایی بازسازی کرد. نتایج سناریوهای مختلف اعمال شده نشان داد که در صورتی که صرفاً از داده‌های هواشناسی ایستگاه هدف استفاده شود، می‌توان با پارامترهای هواشناسی حداقل و حداکثر دما، رطوبت نسبی متوسط، تابش فرازمینی و تعداد روزهای صاف، ابری و نیمه‌ابری با RMSE معادل ۱۶/۷۹ ساعت و خطای متوسط ۶/۴۴ درصد، مجموع ماهیانه ساعات آفتابی را تخمین زد. بهترین نتیجه نیز زمانی حاصل شد که از هر دو سری داده هواشناسی ایستگاه هدف و ایستگاه‌های مجاور استفاده گردید. Sharifi et al., (2021) با استفاده از داده‌های تابش خورشیدی، ساعات آفتابی و دمای هوا، توانایی مدل‌های هوشمند را در برآورد تابش خورشیدی ماهانه ایستگاه تبریز مورد ارزیابی قرار داده و گزارش کردند که شبکه عصبی مصنوعی، بهترین مدل برای برآورد تابش خورشیدی ماهانه است. تا کنون مطالعات متعددی در مورد برآورد داده‌های گم‌شده به‌خصوص داده‌های دما و بارش صورت گرفته است. با این حال بازسازی داده‌های ساعات آفتابی کمتر مورد توجه پژوهشگران قرار گرفته است. با توجه به اینکه ساعات آفتابی یکی از متغیرهای اساسی مورد نیاز به‌منظور برآورد نیاز آبی گیاهان است و با علم به این‌که این متغیر داده‌های گم‌شده نسبتاً زیادی دارد، در پژوهش حاضر ارزیابی کارآیی روش‌های هوشمند و آماری در بازسازی ساعات آفتابی در شرق حوضه دریاچه ارومیه مورد توجه قرار گرفته است.

مواد و روش‌ها

منطقه مورد مطالعه

محدوده مطالعاتی در پژوهش حاضر، شرق حوضه آبریز دریاچه ارومیه واقع در منطقه شمال‌غرب ایران می‌باشد. این مطالعه با استفاده از داده‌های ساعات آفتابی ایستگاه‌های سینوپتیک منتخب در شرق حوضه دریاچه ارومیه شامل تبریز، سراب، سهند و مراغه انجام گرفت. موقعیت جغرافیایی ایستگاه‌های مورد مطالعه در شکل ۱ نشان داده شده است. برای اطمینان از توزیع نرمال داده‌های مورد استفاده، از آزمون کولموگروف–اسمیرنوف استفاده شد.



شکل۱- موقعیت جغرافیایی منطقه و ایستگاه‌های مطالعاتی

Figure 1- Location of the studied area

به‌منظور بازسازی داده‌های ساعات آفتابی در ایستگاه‌های منتخب واقع در شرق حوضه دریاچه ارومیه شامل ایستگاه‌های تبریز، سراب، سهند و مراغه در یک بازه زمانی ۳۰ ساله بین سال‌های ۱۳۶۹ تا ۱۳۹۸، از داده‌های روزانه ساعات آفتابی صرفاً در روزهایی که در هر چهار ایستگاه مورد مطالعه داده وجود داشت، استفاده شد. همچنین، ۷۵ درصد داده‌ها برای واسنجی و ۲۵ درصد باقیمانده برای صحت‌سنجی روش‌های مذکور مورد استفاده قرار گرفتند. برای بازسازی داده‌های ساعات آفتابی در هریک از ایستگاه‌ها، تمامی داده‌های یک ایستگاه حذف نموده و آن ایستگاه به‌عنوان ایستگاه هدف برای بازسازی داده‌ها تعیین گردید. در ادامه با استفاده از داده‌های سایر ایستگاه‌ها و با کاربرد روش‌های مورد مطالعه، بازسازی داده‌ها به انجام رسید. در جدول ۱، برخی از مشخصات هواشناسی و اقلیمی بلندمدت ایستگاه‌های مورد مطالعه ارائه شده است.

جدول ۱- مشخصات اقلیمی ایستگاه‌های سینوپتیک مورد مطالعه در شرق حوضه دریاچه ارومیه

Table 1- Characteristics of the studied synoptic stations in the east of Lake Urmia basin.

Station	Elevation above sea level (m)	Average Temperature (°C)	Rain (mm)	Average sunshine hours (hr)
Tabriz	1361	13.1	254	7.9
Sarab	1682	8.8	239	8
Sahand	1641	12.2	225	7.8
Maraghe	1344	13.3	277	8.2

رگرسیون ماشین بردار پشتیبان (SVR) [1]

ماشین بردار پشتیبان یکی از روش‌های یادگیری است که بر مبنای تئوری یادگیری آماری در سال ۱۹۹۲ میلادی معرفی شده است (Boser et al., 1992). گسترش ماشین بردار پشتیبان بر اساس رگرسیون نیز در سال ۱۹۹۵ به نتیجه رسید (Vapnik, 1995). ماشین بردار پشتیبان مبتنی بر کمینه کردن ساختاری ریسک می‌باشد که از نظریه آموزش آماری گرفته شده است (Vapnik, 1998). مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده مدل طبقه‌بندی ماشین بردار پشتیبان و مدل رگرسیون بردار پشتیبان تقسیم‌بندی می‌شوند. مدل رگرسیون بردار پشتیبان در حل مسائل پیش‌بینی کاربرد دارد. در یک مدل رگرسیونی SVR لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز مشخص می‌شود (معادله ۱).

$$y=f(x)+noise \qquad (۱)$$

بنابراین موضوع اصلی پیدا کردن فرم تابع f است که بتواند به صورت صحیح موارد جدیدی را که SVR تاکنون تجربه نکرده است، پیش‌بینی کند. این تابع به وسیله آموزش مدل SVR بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرآیندی به منظور بهینه‌سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVR شناخته شده است که عبارتند از مدل‌های رگرسیونی SVR نوع اول که به مدل‌های v- SVR مشهورند و مدل‌های رگرسیونی SVR نوع دوم که به مدل‌های ε – SVR مشهورند. در این مطالعه مدل ε – SVR به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده گردید. برای این مدل، تابع خطا به صورت معادله ۲ تعریف می‌شود.

$$\frac{1}{2}W^TW+C\sum_{i=1}^{N}\xi_i+C\sum_{i=1}^{N}\xi_i^* \qquad (۲)$$

تابع خطای مذکور لازم است که با توجه به محدودیت‌های زیر کمینه گردد (معادله ۳).

$$W^T\phi(X_i)+b-y_i\leq \varepsilon+\xi_i^* \qquad (۳)$$

$$y_i-W^T\phi(X_i)-b\leq \varepsilon+\xi_i^* \qquad \xi_i.\xi_i^*\geq 0 \ , \ i=1.\cdots.N$$

[1] Support Vector Regression

DOI: 10.22125/agmj.2022.315265.1126

که در آن C ثابت گنجایش، W بردار ضرایب، W^T ترانهاده بردار ضرایب، ξ_i و ξ_i^* ضرایب کمبود، b ضریبی ثابت، N الگوی آموزش مدل و ϕ تابع کرنل است (Vapnik, 1998). از بین همه توابع کرنل موجود در این روش، بهترین تابع بر حسب کمترین خطا انتخاب شد.

شبکه‌های عصبی مصنوعی (ANN)[1]

شبکه‌های عصبی مصنوعی از شبیه‌سازی شبکه‌های عصبی موجودات زنده الهام گرفته شده‌اند که به‌عنوان ابزاری قدرتمند دارای الگوی پردازش اطلاعات هستند (Menhaj, 2005). تکنیک شبکه عصبی مصنوعی از جمله روش‌های هوشمند است که به‌طور گسترده‌ای در مدل‌سازی و پیش‌بینی فرآیندهای هیدرولوژیکی مورد استفاده قرار گرفته است. این شبکه‌ها از نورون‌ها تشکیل می‌شوند که در گروه‌هایی به نام لایه قرار گرفته و از طریق اتصالات وزن‌دار به یکدیگر متصل می‌شوند. هر ساختار ساده شبکه از سه لایه تشکیل شده است: لایه ورودی، لایه پنهان و لایه خروجی. وقتی داده‌های ورودی به لایه ورودی وارد می‌شوند، از طریق شبکه عصبی عبور کرده و در لایه میانی بر روی آن‌ها پردازش انجام می‌شود تا زمانی که خروجی در لایه خروجی به‌دست آید. هر نورون از طریق اتصالات وزنی ورودی‌های زیادی را از سلول‌های عصبی دیگر دریافت می‌کند. این ورودی‌های وزنی جمع شده و یک تابع انتقالی را ایجاد می‌کنند که در نهایت خروجی نهایی نورون را تولید می‌کند (Talebizadeh et al., 2009). با توجه به این‌که شبکه عصبی مصنوعی به اطلاعات دقیق در مورد روند فیزیکی حاکم بر سیستم‌ها نیاز ندارد، به‌طور مؤثری برای مدل‌سازی فرآیندهای هیدرولوژیکی پیچیده استفاده می‌شود. شبکه‌های عصبی مصنوعی انواع مختلفی دارند که متداول‌ترین آن‌ها پرسپترون چندلایه (MLP)[2] می‌باشد که در این مطالعه از این مدل استفاده شده است. مدل MLP توسط سلول‌های عصبی ساده‌ای به نام پرسپترون تشکیل می‌شود (Kuan and White, 1994). پرسپترون با ایجاد یک ترکیب خطی با توجه به وزن ورودی خود و سپس تعیین خروجی از طریق یک تابع انتقال غیرخطی، یک خروجی منفرد از چندین ورودی را محاسبه می‌کند که خروجی آن به‌صورت معادله ۴ تعریف می‌گردد.

$$y=f(\sum_{i=1}^{n} w_i x_i+b) \quad (۴)$$

در این معادله، w_i بردار وزن، x_i بردار ورودی (i=1,2…n)، b بایاس، f تابع انتقال و y خروجی می‌باشد.

جنگل‌های تصادفی (RF) [3]

روش جنگل‌های تصادفی را اولین بار بریمن در سال ۲۰۰۱ با توسعه درخت‌های تصمیم، به عنوان یک تکنیک جدید ارائه داده است که پیش‌بینی چندین الگوریتم منفرد را با هم با استفاده از قوانین مبتنی ترکیب می‌کند. این روش در بین روش‌های درختی، تکنیک نسبتاً پیچیده‌ای است که به منظور افزایش دقت مدل در آن چندین درخت تصمیم آموزش داده می‌شود. نتیجه حاصل پیش‌بینی گروهی از درختان تصمیم است (Breiman, 2001). اصول کلی تکنیک‌های آموزش گروهی بر پایه این فرض است که دقت آن‌ها از دیگر الگوریتم‌های آموزشی بالاتر است، چون ترکیبی از چند مدل پیش‌بینی، دقیق‌تر از یک مدل می‌باشد و گروه‌ها قدرت مجموعه‌های منفرد و منحصر به فرد از طبقه‌ها را بیشتر می‌کنند، در حالی که هم‌زمان نقاط ضعف طبقه‌ها را کاهش می‌دهند (Kotsiantis and, Pintelas, 2004). در یک طبقه‌بندی مبتنی بر جنگل‌های تصادفی، دو پارامتر توسط کاربر تعیین می‌گردد: اندازه یک زیرمجموعه تصادفی از ویژگی‌ها (M) و تعداد درخت‌ها (T). انتخاب پارامتر M بر روی نرخ خطای نهایی مؤثر است. با افزایش M، هم وابستگی بین درخت‌ها و هم صحت و دقت طبقه‌بندی تک درخت در جنگل افزایش می‌یابد. نرخ خطا با وابستگی متناسب بوده، اما با صحت طبقه‌بندی نسبت عکس دارد. معمولاً، مقدار M برابر جذر تعداد ویژگی‌ها در نظر گرفته می‌شود. با توجه به سریع بودن و عدم بیش-برازش روش جنگل تصادفی، تعداد درخت‌ها می‌تواند تا حد امکان زیاد باشد. اما با توجه به محدودیت حافظه ماشین، معمولاً از چندصد تا چندهزار تا انتخاب می‌شود (Jin, 2012). به طور کلی روش جنگل‌های تصادفی برای طبقه‌بندی به این صورت است که در ابتدا T نمونه بوت استرپ[4] از داده آموزشی انتخاب می‌شود و سپس از هر نمونه بوت استرپ β یک درخت طبقه‌بندی و رگرسیون[5] (CART) هرس نشده ایجاد می‌شود که برای انشعاب در هر گره CART، تنها یک متغیر انتخاب شده به‌صورت تصادفی

[1] Artificial Neural Networks
[2] Multi-Layer Perceptron
[3] Random Forests
[4] Boot Strap
[5] Classification and Regression Tree

DOI: 10.22125/agmj.2022.315265.1126

استفاده می‌شود. در نهایت خروجی طبقه‌بندی بر اساس یک نتیجه میانگین، از پیش‌بینی‌های تمام درخت‌های منفرد آموزش دیده به‌دست می‌آید. یک مجموعه داده بوت استرپ، مجموعه‌ای از نقاط انتخابی به طور تصادفی است که از مجموعه داده آموزشی انتخاب شده است.

مختصات جغرافیایی[1] (روش گرافیکی)

از جمله روش‌های مورد استفاده برای بازسازی داده‌های گم شده، روش مختصات جغرافیایی یا روش گرافیکی می‌باشد. در این روش پس از تعیین موقعیت ایستگاه‌های منطقه مورد مطالعه روی نقشه توپوگرافی که با استفاده از مختصات جغرافیایی آن‌ها صورت می‌پذیرد، ایستگاه هدف برای بازسازی داده به‌عنوان مرکز مختصات جغرافیایی قرار داده شده و مختصات هر یک از ایستگاه‌های اطراف آن نسبت به این مختصات جغرافیایی تعیین می‌گردد. بدیهی است که ایستگاه‌های نزدیک‌تر به ایستگاه مدنظر سهم بیشتری در بازسازی آن خواهند داشت؛ لذا لازم است که ضریب وزنی بزرگ‌تری به آن اختصاص داده شود. این ضریب وزنی با استفاده از معادله ۵ محاسبه می‌گردد.

$$W = \frac{1}{x^2+y^2} \qquad (۵)$$

که در آن x و y به‌ترتیب طول و عرض مختصاتی ایستگاه می‌باشد. در نهایت داده‌های گم‌شده در ایستگاه هدف با استفاده از معادله ۶ محاسبه می‌شود.

$$N_x = \frac{\sum_{i=1}^{n} W_i N_i}{\sum_{i=1}^{n} W_i} \qquad (۶)$$

که در آن N_x مقدار برآورد شده داده گم‌شده در ایستگاه x، N_i مقدار داده موجود در ایستگاه i و n معرف تعداد ایستگاه‌هایی است که برای برآورد داده گم‌شده، از داده‌های آن‌ها استفاده شده است.

نسبت نرمال[2]

روش نسبت نرمال ابتدا توسط Paulhus and Kohler (1952) برای تخمین داده‌های گم شده بارندگی به‌کار رفت و در ادامه توسط Young (1992) اصلاح شد. این روش عمدتاً به میانگین نسبت داده‌های بین ایستگاه‌های شاهد و ایستگاه هدف بستگی دارد. در این روش ساعات آفتابی در ایستگاه هدف متناسب است با نسبت میانگین ساعات آفتابی در ایستگاه هدف به میانگین ساعت آفتابی در ایستگاه‌های شاهد ضربدر ساعات آفتابی همزمان ایستگاه شاهد که از طریق معادله ۷ محاسبه می‌گردد.

$$N_x = \frac{1}{n}\sum_{i=1}^{n}\frac{\bar{N}_x}{\bar{N}_i}N_i \qquad (۷)$$

که در آن $\bar{N}_x$ میانگین داده‌های ساعات آفتابی در ایستگاه هدف، $\bar{N}_i$ میانگین داده‌های ساعات آفتابی در ایستگاه شاهد i ام و N_i داده‌های ساعات آفتابی در ایستگاه i ام می‌باشند.

ضریب همبستگی وزنی[3]

در این روش به‌منظور برآورد داده گم شده در ایستگاه هدف، از ضرایب همبستگی ایستگاه‌های شاهد استفاده می‌شود. کارآیی این روش به قدرت همبستگی بین ایستگاه هدف و ایستگاه‌های اطراف بستگی دارد. برای برآورد داده گم شده با استفاده از این روش از معادله ۸ استفاده می‌شود (Teegavarapu and Chandramouli, 2005).

$$N_x = \sum_{i=1}^{n}\left(\frac{(r_i)}{\sum_{i=1}^{n}(r_i)}\right)N_i \qquad (۸)$$

که در آن r_i ضریب همبستگی پیرسون بین داده‌های ساعات آفتابی ایستگاه هدف و ایستگاه شاهد i می‌باشد.

معیارهای ارزیابی نتایج

برای ارزیابی دقت و کارآیی روش‌های هوشمند و روش‌های آماری در بازسازی داده‌های ساعات آفتابی، نتایج به‌دست آمده با مقادیر واقعی مقایسه شدند. برای این‌منظور از شاخص‌های آماری ضریب همبستگی (r)، جذر میانگین مربعات خطا (RMSE)، میانگین انحراف مطلق (MAD) برای تعیین میزان همبستگی بین مقادیر ساعات آفتابی واقعی و مقادیر برآورد شده و نیز تعیین مقادیر خطای بازسازی داده‌های ساعات آفتابی استفاده شد (معادله‌های ۹ تا ۱۱).

$$r = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}\right) \qquad (۹)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}} \qquad (۱۰)$$

$$MAD = \frac{\sum_{i=1}^{n}|x_i-y_i|}{n} \qquad (۱۱)$$

در این روابط x_i و y_i به‌ترتیب i امین داده مشاهداتی و برآورد شده، $\bar{x}$ و $\bar{y}$ به ترتیب میانگین داده‌های مشاهداتی و برآورد شده و n طول سری زمانی داده‌ها است. علاوه بر این، دیاگرام Taylor (2001) برای تحلیل دقت روش‌های مورد استفاده در بازسازی داده‌ها به‌کار گرفته شد. دیاگرام

[1] Geographical Coordinates (GC)
[2] Normal Ratio (NR)
[3] Correlation Coefficient Weighted (CCW)

DOI: 10.22125/agmj.2022.315265.1126

تیلور، راه‌حلی گرافیکی برای ارزیابی دقت داده‌های پیش‌بینی شده با به تصویر کشیدن همزمان پارامترهای آماری می‌باشد. در دیاگرام مذکور، هر نقطه بیانگر عملکرد روش متناظر بوده و هرچه نقاط متناظر روش‌ها به نقطه داده‌های مشاهداتی در مختصات قطبی نزدیک‌تر باشد، نشان‌دهنده دقت بالاتر و خطای کمتر آن روش می‌باشد (Gleckler et al., 2008).

نتایج و بحث

در ابتدا به‌منظور بررسی میزان همبستگی داده‌های ساعات آفتابی در ایستگاه‌های مورد مطالعه، ضرایب همبستگی ساده پیرسون داده‌های ساعات آفتابی محاسبه شد که نتایج در قالب ماتریس همبستگی در جدول ۱ ارائه شده است. بر این اساس، بیشترین و کمترین میزان همبستگی معادل ۰/۹۵ و ۰/۸۷ به‌ترتیب بین داده‌های تبریز با سهند و سراب با سهند مشاهد شد.

جدول ۲- ماتریس ضرایب همبستگی داده‌های ساعات آفتابی

Table 2- Correlation coefficients matrix of sunshine hours data.

	Tabriz	Sarab	Sahand	Maragheh
Tabriz	1	0.90	0.95	0.91
Sarab		1	0.87	0.89
Sahand			1	0.91
Margheh				1

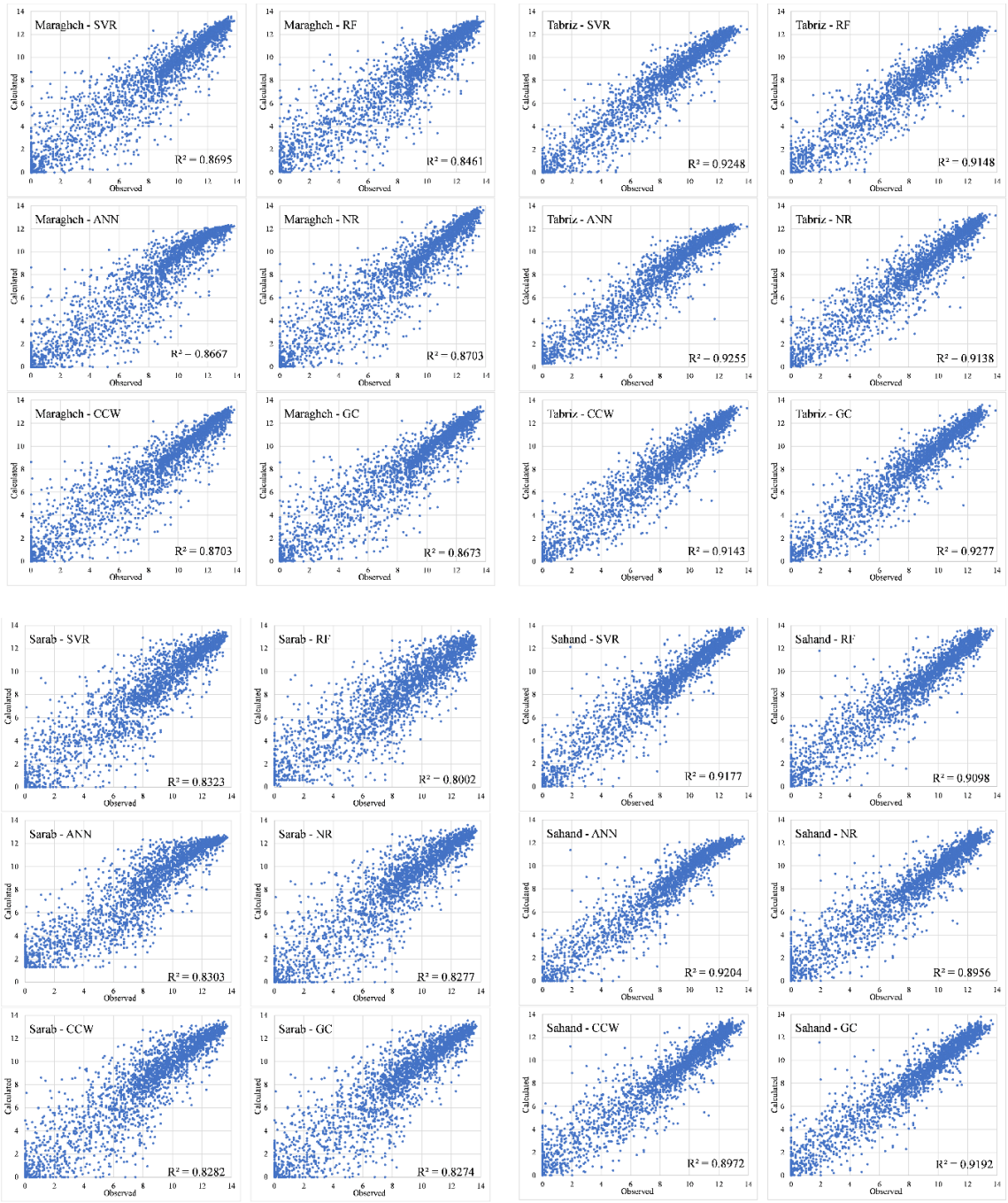
بعد از انجام محاسبات مورد نیاز، مقادیر آماره‌های مورد استفاده به‌منظور مقایسه نتایج حاصل از روش‌های هوشمند در بازسازی داده‌های ساعات آفتابی و ارزیابی روش‌ها در جدول ۳ ارائه گردید. نتایج نشان داد که در ایستگاه‌های تبریز و سهند روش شبکه عصبی مصنوعی، به‌ترتیب با مقادیر RMSE معادل ۱/۰۶ ساعت و ۱/۱۴ ساعت و مقادیر MAD معادل ۰/۷۵ ساعت و ۰/۷۷ ساعت، بیشترین دقت و کارآیی را در بازسازی داده‌های ساعات آفتابی دارند. نتایج به‌دست آمده از پژوهش Bayat K, Mirlatifi (2009) نیز نشان داد که مدل شبکه عصبی مصنوعی بهترین نتیجه را در برآورد تابش کل خورشیدی روزانه حاصل می‌کند. Sharifi et al., (2021) نیز گزارش کردند که شبکه عصبی مصنوعی، بهترین مدل برای برآورد تابش کل خورشیدی ماهانه در ایستگاه تبریز است. در ایستگاه‌های سراب و مراغه نیز روش رگرسیون ماشین بردار پشتیبان، به‌ترتیب با مقادیر RMSE معادل ۱/۵۸ ساعت و ۱/۴۶ ساعت و مقادیر MAD معادل ۱/۱۱ ساعت و ۰/۹۶ ساعت، دقیق‌ترین روش در بین روش‌های مورد بررسی می‌باشند (جدول ۳). در نقطه مقابل نیز با توجه به مقادیر شاخص‌های خطای مورد بررسی، در تمام ایستگاه‌های مورد مطالعه، روش جنگل‌های تصادفی کمترین دقت و بیشترین خطا را در بازسازی داده‌های ساعات آفتابی دارد. همچنین در شکل ۲، مقادیر ساعات آفتابی مشاهداتی و محاسباتی ایستگاه‌های مورد مطالعه از طریق هریک از روش‌های مورد استفاده نیز ارائه شده است. تراکم بیشتر نقاط حول خط نیمساز، حاکی از دقت بالای برآورد داده می‌باشد.

جدول ۳- مقادیر خطای روش‌های هوشمند در بازسازی ساعات آفتابی

Table 3- Error-values of intelligent methods in reconstruction sunshine hours

Station	Method	R	RMSE (hr)	MAD (hr)
Tabriz	**ANN**	**0.96**	**1.06**	**0.75**
	RF	**0.96**	**1.17**	**0.84**
	SVR	**0.96**	**1.10**	**0.76**
Sarab	**ANN**	**0.91**	**1.58**	**1.17**
	RF	**0.89**	**1.71**	**1.27**
	SVR	**0.91**	**1.58**	**1.11**
Sahand	**ANN**	**0.96**	**1.14**	**0.77**
	RF	**0.95**	**1.26**	**0.87**
	SVR	**0.96**	**1.23**	**0.83**
Maragheh	**ANN**	**0.93**	**1.56**	**1.09**
	RF	**0.92**	**1.59**	**1.08**
	SVR	**0.93**	**1.46**	**0.96**

در جدول ۴ نیز مقادیر شاخص‌های مورد بررسی جهت ارزیابی روش‌های آماری در برآورد ساعات آفتابی ارائه شده است. بر این اساس در ایستگاه‌های تبریز و سهند، روش مختصات جغرافیایی به‌ترتیب با مقادیر RMSE معادل ۱/۰۴ ساعت و ۱/۱۳ ساعت و مقادیر MAD معادل ۰/۷۲ ساعت و ۰/۷۵ ساعت، بهترین برآورد را در بازسازی داده‌های ساعات آفتابی دارند. در ایستگاه مراغه نیز روش نسبت نرمال با مقادیر RMSE معادل ۱/۴۵ ساعت و MAD معادل ۰/۹۶ ساعت، بیشترین دقت و کارآیی را در بازسازی داده‌های ساعات آفتابی دارد. همچنین نتایج نشان داد که در ایستگاه سراب، هر سه روش مورد استفاده دقت یکسانی دارند (جدول ۴). در نقطه مقابل نیز، در ایستگاه‌های تبریز و سهند، روش‌های نسبت نرمال و ضریب همبستگی وزنی دقت مشابهی در برآورد ساعات آفتابی دارند. در ایستگاه مراغه نیز روش مختصات جغرافیایی، کمترین دقت را در بازسازی داده‌های ساعات آفتابی دارد (جدول ۴).



شکل ۲- مقایسه مقادیر مشاهداتی و محاسباتی داده‌های ساعات آفتابی

Figure 2- Comparison of observational and computational values of sunshine hours data

این در حالی است که نتایج پژوهش Khansari et el., (2018) نشان داد که روش مختصات جغرافیایی بهترین روش برای برآورد داده‌های گم شده ساعات آفتابی در ایستگاه مراغه می‌باشد. به‌نظر می‌رسد دلیل اصلی مغایرت نتایج، تفاوت در ایستگاه‌های شاهد جهت بازسازی داده‌ها و نیز بازه زمانی مورد بررسی باشد. ایشان از داده‌های ۵ ساله ایستگاه‌های تبریز، مهاباد، ارومیه و میانه برای بازسازی داده‌های ساعات آفتابی در مراغه استفاده نمودند. همچنین، دیاگرام تیلور جهت بررسی و تحلیل مقادیر همبستگی و انحراف معیار بین داده‌های مشاهداتی، روش‌های هوشمند و روابط آماری مورد استفاده جهت بازسازی داده‌ها (مرحله صحت‌سنجی) رسم شد (شکل ۳). در دیاگرام تیلور، فاصله شعاعی از نقطه مشاهداتی (نقطه سبزرنگ) نشان‌دهنده مقدار جذر میانگین مربعات خطای روش‌های مورد مطالعه می‌باشد.

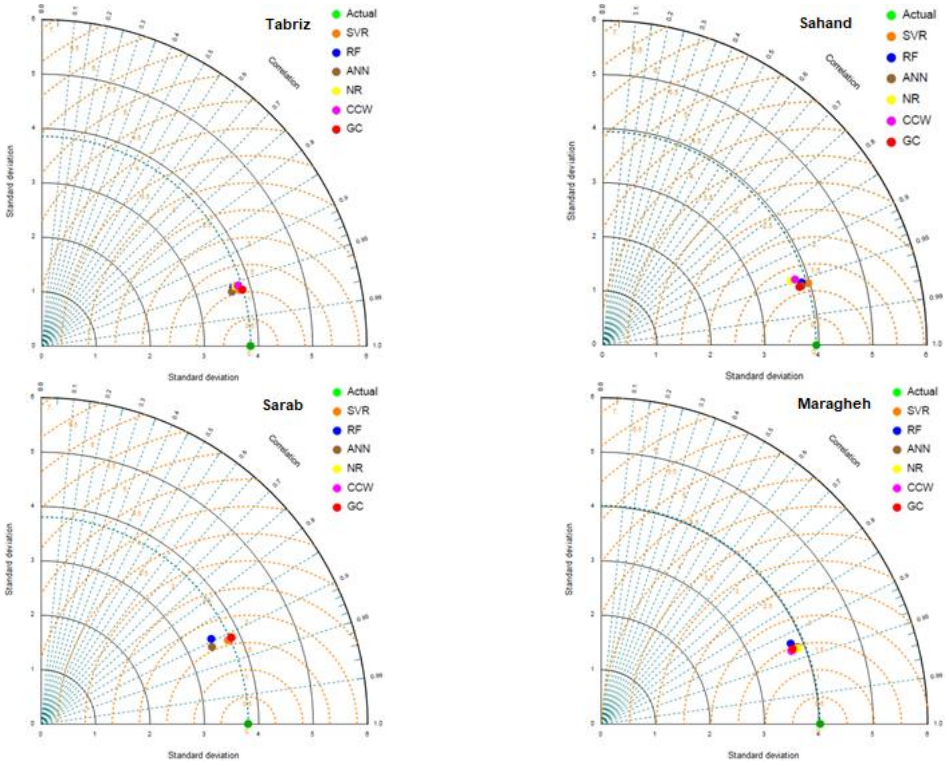
DOI: 10.22125/agmj.2022.315265.1126

جدول ۴- مقادیر خطای روش‌های آماری در بازسازی ساعات آفتابی

Table 4- Error-values of statistical methods in reconstruction sunshine hours

Station	Method	R	RMSE (hr)	MAD (hr)
Tabriz	NR	0.96	1.13	0.77
	GC	0.96	1.04	0.92
	CCW	0.96	1.14	0.77
Sarab	NR	0.91	1.63	1.13
	GC	0.91	1.62	1.13
	CCW	0.91	1.62	1.13
Sahand	NR	0.95	1.28	0.86
	GC	0.96	1.13	0.75
	CCW	0.95	1.28	0.86
Maragheh	NR	0.93	1.45	0.96
	GC	0.93	1.49	1.01
	CCW	0.93	1.47	1.00

بر این اساس در ایستگاه‌های تبریز و سهند، روش مختصات جغرافیایی، در ایستگاه سراب، روش رگرسیون ماشین بردار پشتیبان و در ایستگاه مراغه، روش نسبت نرمال، بر مبنای فاصله شعاعی کمتر تا نقطه مشاهداتی (نقطه سبز رنگ)، بازسازی دقیق‌تری از مقادیر ساعات آفتابی داشته‌اند. در نقطه مقابل نیز، در سه ایستگاه تبریز، سراب و مراغه، روش جنگل‌های تصادفی و در ایستگاه سهند، روش نسبت نرمال، خطای بیشتری در بازسازی داده‌های ساعات آفتابی داشتند (شکل ۳). همچنین مقایسه نتایج حاصل از روش‌های هوشمند و آماری در بازسازی داده‌های ساعات آفتابی نشان داد که به‌طور کلی در ایستگاه‌های تبریز، سراب و سهند، هر دو دسته روش‌های هوشمند و آماری دقت تقریباً مشابهی دارند. اما در ایستگاه مراغه، روش‌های آماری در مقایسه با هوشمند برآوردهای دقیق‌تری در بازسازی داده‌های ساعات آفتابی دارند.



شکل ۳- دیاگرام تیلور روش‌های هوشمند و روش‌های آماری در بازسازی داده‌های ساعات آفتابی

Figure 3- Taylor diagram of Intelligent methods and statistical methods in reconstruction sunshine hours

نتیجه‌گیری

استفاده از داده‌های صحیح و پیوسته، شرط اولیه انجام مطالعات هیدرولوژیکی است. در ثبت داده‌های ساعات آفتابی، به عنوان یکی از داده‌های اصلی برآورد تبخیرتعرق و نیاز آبی گیاهان، خلأهای زیادی وجود دارد. لذا در پژوهش حاضر، بازسازی داده‌های ساعات آفتابی با استفاده از روش‌های هوشمند و روش‌های آماری در ایستگاه‌های تبریز، سراب، سهند و مراغه واقع در شرق حوضه دریاچه ارومیه مدنظر قرار گرفت. نتایج نشان داد که در بین روش‌های هوشمند، روش شبکه عصبی مصنوعی و در بین روش‌های آماری، روش مختصات جغرافیایی بالاترین دقت را در بازسازی داده‌های ساعات آفتابی دارند. به‌طور کلی در بین تمام روش‌های مورد بررسی نیز، روش مختصات جغرافیایی بیشترین دقت و روش جنگل‌های تصادفی،

کمترین دقت را برآورد و بازسازی داده‌های ساعت آفتابی دارد. همچنین با توجه به مقادیر خطای تقریباً مشابه روش‌های هوشمند و آماری در ایستگاه‌های تبریز، سراب و سهند، تفاوت مشهودی بین روش‌های هوشمند و آماری وجود ندارد، اما در ایستگاه مراغه، روش‌های آماری مورد استفاده برای بازسازی داده‌های ساعات آفتابی، دقت بالاتری در مقایسه با روش‌های هوشمند دارند.

منابع

Armanuos, AM., Al-Ansari, N., Yaseen, ZM. 2020. Cross assessment of twenty-one different methods for missing precipitation data estimation. Atmosphere, 11(4):1-34.

Bayat, K., Mirlatifi, SM. 2009. Estimation of global solar radiation using regression and artificial neural network models. Journal of agricultural sciences and natural resources, 16(3): 270-280. (In Farsi)

Behrang, MA., Assareh, E., Ghanbarzadeh, A. and Noghrehabadi, AR. 2010. The potential of different artifcial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data, Solar Energy, 84: 1468–1480.

Boser, BE., Guyon, IM., Vapnik, VN. 1992. A training algorithm for optimal margin classifiers. In: D.Haussler, editor, 5th Annual ACM Workshop on COLT. Pittsburgh, PA, 144-152.

Breiman, L. 2001. Random forests. Machine Learning, 45(1): 5-32.

Coulibaly, PND., Evora B. 2007. Comparison of neural network methods for infilling missing daily weather records. Journal of hydrology, 341: 27-41.

Fooladmand, HR. 2012. Comparing reference evapotranspiration using actual and estimated sunshine hours in south of Iran. African Journal of Agricultural Research 7(7): 1164-1169.

Gleckler, PJ., Taylor, KE., Doutriaux, C. 2008. Performance metrics for climate models. Journal of Geophysical Research: Atmospheres, 113(D6): 1-20.

Hasanpour, M., Dinpashoh, Y. 2012. Evaluation of efficiency of different estimation methods for missing climatological data. Stochastic Environment Research and Risk Assessment, 26:59–71.

Jin, J. 2012. A random forest based method for urban land cover classification using LiDAR data and aerial imagery. MSc Thesis, University of Waterloo.

Karbasi, M. 2016. Reconstruction of missing data of monthly total sunshine hours using artificial neural networks. Iranian journal of irrigation and drainage, 10(5):570-580. (In Farsi)

Khansari, S., Rezayi, A., Shiri, J., dashti, Sh., hatamimaleki H. 2018. Reconstruction of missing data in meteorological variables used to estimate daily evapotranspiration. The Second National Conference on Climatology of Iran, Mashhad, Iran. (In Farsi)

Kotsiantis, S., Pintelas, P. Combining bagging and boosting. 2004. International Journal of Computational Intelligence, 1(4): 324-33.

Kuan, CM., White, H. 1994. Artificial neural networks: An econometric perspective. Econometric Reviews, 13: 1-91.

Menhaj, MB. 2005. Computational intelligence-volume I: Fundamentals of neural networks. Tehran: AmirKabir University. (In Farsi)

Naghidi, R., Shayannezhad, M., Sadati Nejad, SJ. 2010. Comparison of different methods for estimating of monthly discharge missing data in grand Karoon River basin. Journal of Watershed Management Research, 1(1): 59-71.

Paulhus, JLH., Kohler, MA. 1952. Interpolation of missing precipitation records. Monthly Weather Review, 80:129–133.

Piri, J., Ansari, H., Farid-Hosseini, A. 2013. Modeling ground-reached solar radiation using ANFIS and empirical models (Case of study: Zahedan and Bojnourd stations). Iranian Journal of Energy, 16(3): 37-58. (In Farsi)

Sharifi, S., Rezaverdinejad, V., Nourani, V., Behmanesh, J. 2021. Evaluation of the capability of intelligent models in estimating monthly global solar radiation. Water and Soil Science, 31(2): 13-26. (In Farsi)

Tabari, H., Talaee, PH. 2015. Reconstruction of river water quality missing data using artificial neural networks. Water Quality Research Journal of Canada, 50(4): 326-335.

Talebizadeh, M., Morid, S., Ayyoubzadeh, SA., Ghasemzadeh, M. 2009. Uncertainty analysis in sediment load modeling using ANN and SWAT model. Water Resources Management, 24 (9): 1747-1761.

Taylor, KE. 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research. Atmospheres, 106: 7183-7192.

Teegavarapu, RSV., Chandramouli, V. 2005 Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Journal of Hydrology, 312: 191–206.

Vapnik, VN. 1995. The Nature of statistical learning theory. Springer, New York.

Vapnik, VN. 1998. Statistical learning theory. Wiley, New York.

Young, KC. 1992. A Three-way model for interpolating for monthly precipitation values. Monthly Weather Review, 120: 2561–2569.

Journal of Agricultural Meteorology
(Iranian Soc. of Water and Irrig. Eng.)
Vol. 10, No. 2, Autumn & Winter 2023, pp. 28-36
DOI: 10.22125/agmj.2022.315265.1126

نشریه هواشناسی کشاورزی
(وابسته به انجمن مهندسی آبیاری و آب ایران)
جلد ۱۰، شماره ۲، پاییز و زمستان ۱۴۰۱، صص.۳۶-۲۸
مقاله علمی پژوهشی

Comparison of the efficiency of intelligent and statistical methods in the reconstruction of sunshine hours data (Case study: East of Urmia Lake basin)

V. Mouneskhah[1], M. Khaledi Alamdari[1], M. Hadi[1], S. Samadianfard[2]

Received: 16/11/2021
Accepted: 13/06/2022

Abstract

One of the climate variables with relatively large gaps in observation and significant importance in estimation of evapotranspiration is sunshine hours. In the present study, in order to reconstruction the sunshine hour data of several selected stations in Tabriz province, Iran namely,Tabriz, Sarab, Sahand and Maragheh during the period of 1990 to 2019, skill of intelligent approaches of SVR, ANN and RF was compared with statistical methods of normal ratio, geographical coordinates and weight correlation coefficient. Statistical indices of R, RMSE, MAD and Taylor diagrams were used for evaluation of comparisons. The obtained results showed that ANN and geographical coordinate methods have the highest accuracy in reconstruction sunshine hours among the selected intelligent and statistical methods, respectively. In Tabriz and Sahand stations, the geographical coordinate method with RMSE of 1.04 and 1.13 hours, respectively, in the Sarab station SVR with RMSE of 1.58 hours and in Maragheh station the normal ratio method with RMSE of 1.45 hours showed the highest accuracy in generating sunshine hours. Besides, RF method had the lowest accuracy in reconstruction of sunshine hours data. It can be concluded that in Tabriz, Sarab and Sahand stations, both types of intelligent and statistical methods have almost same accuracy, but in Maragheh station, statistical methods provided slightly better estimations.

Keywords: Data gaps, Sunshine hours, Taylor diagram, Urmia Lake basin

[1] Ph.D Candidate of Irrigation and Drainage, Department of Water Engineering, Tabriz University, Tabriz, Iran (*Coresponding Author Email Address: m.khaledi.a@gmail.com)
[2] Assistant Professor, Department of Water Engineering, Tabriz University, Tabriz, Iran

نحوه ارجاع مقاله:

مونس‌خواه، و.، خالدی علمداری، م.، هادی، م.، صمدیان فرد، س. ۱۴۰۱. مقایسه کارآیی روش‌های هوشمند و آماری در بازسازی داده‌های ساعت آفتابی (مطالعه موردی: شرق حوضه دریاچه ارومیه). نشریه هواشناسی کشاورزی، ۱۰(۲): ۳۶-۲۸. DOI: 10.22125/agmj.2022.315265.1126

Mouneskhah, V., Khaledi Alamdari, M., Hadi, M., Samadianfard, S. 2023. Comparison of the efficiency of intelligent and statistical methods in the reconstruction of sunshine hours data (Case study: East of Urmia Lake basin). Journal of Agricultural Meteorology, 10(2): 28-36. DOI: 10.22125/agmj.2022.315265.1126