Archive of SID.ir



نشریه علمی پژوهشی مهندسی آبیاری و آب ایران
سال دوازدهم. شماره چهل و هشت. تابستان ۱۴۰۱

# مدلسازی تبخیر ماهانه با استفاده از روش‌های منفرد و هیبریدی-موجک داده‌کاوی در حوضه‌های آبریز ایران با تنوع اقلیمی

علیرضا عمادی*[۱]، سروین زمان‌زاد قویدل[۲]، سینا فاضلی[۳]، سهیلا زارعی[۳]، علی رشید نیقی[۴]

تاریخ ارسال:۱۳۹۹/۰۵/۲۰
تاریخ پذیرش:۱۳۹۹/۱۰/۱۳

مقاله پژوهشی برگرفته از طرح دانشگاه علوم کشاورزی و منابع طبیعی ساری می‌باشد.

## چکیده

تبخیر به عنوان یکی از پارامترهای طبیعی، همواره مورد توجه محققین بوده است. در این پژوهش، متغیر تبخیر ماهانه با استفاده از روش‌های شبکه عصبی مصنوعی، سامانه استنتاجی فازی-عصبی تطبیقی و برنامه‌ریزی بیان ژن و ترکیب روش‌های مذکور با تئوری موجک، در دو اقلیم متفاوت ایران مدلسازی شد. بدین منظور، داده‌های هواشناسی بارش، رطوبت نسبی، دمای میانگین، دمای بیشینه، دمای کمینه و سرعت باد، در طول دوره آماری ۱۳۹۷-۱۳۸۴ مربوط به دو حوضه آبریز دریاچه ارومیه و گاوخونی به کار گرفته شد. در این مطالعه، اثر فصلی و نویززدایی داده‌ها اعمال شد. دقت روش‌های مورد مطالعه بر اساس شاخص‌های آماری ضریب همبستگی (R)، ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب کارایی نش-ساتکلیف (NSE) مورد بررسی قرار گرفت. نتایج حاکی از این است که در دو اقلیم مختلف روش‌های هیبرید-موجک برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی منفرد به ترتیب دارای بالاترین و ضعیف‌ترین عملکرد در میان سایر مدل‌های داده کاوی به کار رفته در این تحقیق هستند. مدل هیبریدی موجک-برنامه‌ریزی بیان ژن با مقدار RMSE برابر با ۲۰/۸۷۰ و ۱۵۶/۸۸۴ میلی‌متر به ترتیب برای ایستگاه‌های تازه‌کند در حوضه آبریز دریاچه ارومیه و کوهپایه در حوضه آبریز گاوخونی عملکرد بالاتری را داشته است. همچنین، نتایج نشان داد که تأثیر به کارگیری ضریب فصلی و نویززدایی داده‌ها در ارتقاء عملکرد مدل‌ها قابل توجه است. بر اساس نتایج، عملکرد مدل‌ها در حوضه آبریز دریاچه ارومیه با اقلیم Dsa بهتر بوده است. همچنین، روش‌های داده‌کاوی هیبریدی را می‌توان به عنوان جایگزین مناسبی برای روش‌های قدیمی معرفی نمود.

**واژه‌های کلیدی: اقلیم، تبخیر، داده‌کاوی، موجک**

[*۱] دانشیار گروه مهندسی آب. دانشکده مهندسی زراعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران (۰۹۱۲۳۹۳۱۳۱۱) (emadia355@yahoo.com, a.emadi@sanru.ac.ir) **نویسنده مسئول**

[۲] مدیر پروژه و عضو هیات مدیره شرکت مهندسین مشاور دانشوران عمران آب، ارومیه، ایران (۰۹۱۴۱۰۶۶۴۰۹) (snzghavidel@ut.ac.ir)

[۳] دانشجوی دوره دکتری مهندسی منابع آب، گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی کرج، دانشگاه تهران، کرج، ایران (۰۹۱۸۳۳۰۹۲۱۵، ۰۹۱۸۸۷۱۸۸۲۶) (s.fazeli@ut.ac.ir، soheila_zareie@yahoo.com)

[۴] پژوهشگر پسا دکتری، دانشگاه مینه ستو آمریکا، مینه ستو، آمریکا (niaghi@umn.edu)

Archive of SID.ir

Archive of SID.ir



## مقدمه

کمبود آب یکی از مهمترین و بزرگترین چالش‌های قرن آینده است (Genanu et al., 2017). کشور ما در منطقه‌ای از جهان قرار دارد که میزان متوسط بارش در آن یک سوم متوسط جهانی است و تلفات تبخیر نقش مهمی در بیلان آب مخازن سدها دارد. تاکنون مطالعات متعددی با استفاده از روش‌های داده‌کاوی برای مدلسازی و پیش‌بینی پدیده‌های مختلف هیدرولوژیکی و هواشناسی از جمله کیفیت آب، میزان بار رسوب رودخانه، تبخیر از سطح آزاد آب، تبخیر و تعرق پتانسیل گیاه مرجع، بارندگی، رواناب، دبی سیل ورودی به مخازن و ... انجام شده است (Montaseri et al., 2018; Adnan et al., 2019).

با این وجود به کارگیری روش‌های منفرد و هیبرید موجک-شبکه عصبی مصنوعی[1]، سامانه استنتاجی عصبی-فازی تطبیقی[2] (مدل دسته‌بندی تفریقی) و برنامه‌ریزی بیان ژن[3] در زمینه مدلسازی پارامترهای هواشناسی از جمله تبخیر در شرایط مختلف اقلیمی با اعمال ضریب فصلی و نویززدایی داده‌های هواشناسی به طور محدود در پژوهش‌های مختلف سیستم‌های آبی و اقلیمی گزارش شده است (Singh et al., 2019).

صمدیان‌فر و همکاران (۱۳۹۷) از داده‌های هواشناسی دما، رطوبت نسبی، سرعت باد، تبخیر و تابش خورشیدی به منظور برآورد و شبیه‌سازی مقادیر تبخیر از تشت در ایستگاه‌های هواشناسی تبریز و جلفا استفاده کردند. آنها روش‌های هوشمند برنامه‌ریزی ژنتیک، رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی را به کار گرفتند. نتایج نشان داد که شبکه عصبی مصنوعی از دقت بالاتری نسبت به سایر روش‌ها برخوردار است. Tabari et al., (2010)در مطالعه خود برای تخمین میزان تبخیر از تشتک در ناحیه‌ای با اقلیم نیمه‌خشک در ایران، مدل‌های ANN و رگرسیونی را به کار برده‌اند که نتایج، برتر بودن مدل ANNرا نسبت به مدل رگرسیونی نشان می‌دهد. Shirsath and Singh (2010) نتایج مدل‌های ANN، رگرسیون خطی چندگانه (MLR)، Penman ، Priestley-Taylor و SS را برای برآورد تبخیر روزانه در هند مقایسه کردند و دریافتند که مدل ANN تخمین بهتری نسبت به سایر مدل‌ها دارد. Traore and Guven (2012) از مدل برنامه‌ریزی بیان ژن جهت تخمین تبخیر و تعرق در کشور ساحلی بورکینافاسو[4] استفاده نمودند. آن‌ها از داده‌های هواشناسی شش ایستگاه واقع در نواحی مختلف کشور استفاده کردند و جهت مدل‌سازی، ترکیبات ورودی مختلف داده‌های هواشناسی را به کار بردند. نتایج به دست آمده حاکی از دقت بالای روش برنامه‌ریزی در فرآیند مدل‌سازی تبخیر و تعرق بود. Shiri et al., (2012) با مقایسه روش برنامه‌ریزی ژنتیک و سامانه استنتاجی عصبی-فازی تطبیقی جهت مدلسازی تبخیر تعرق مرجع روزانه در کشور اسپانیا به این نتیجه رسیدند که عملکرد برنامه‌ریزی ژنتیک در این زمینه بهتر از سامانه استنتاجی عصبی-فازی می‌باشد. Malik and Kumar (2015) برای پیش‌بینی میزان تبخیر تشت در ایستگاه‌های مختلف هند، رگرسیون خطی چندگانه (MLR) و CANFIS را به کار بردند. نتایج حاکی از این بود که عملکرد مدل‌های ANN در مقایسه با سایر مدل‌ها بهتر است. Tamta et al., (2018) با استفاده از روش‌های ANN و CANFIS به تخمین میزان تبخیر روزانه در هاوالباغ[5]، هند پرداختند. نتایج ایشان حاکی از عملکرد مناسب هر دو روش جهت تخمین میزان تبخیر روزانه می‌باشد. همچنین ایشان به عملکرد برتر ANN در مقایسه با CANFIS اشاره نمودند. Patle et al., (2019) با استفاده از روش‌های رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN) به تخمین

[1] Artificial Neural Networks
[2] Adaptive-Network-based Fuzzy Inference Systems
[3] Gene-Expression Programming
[4] Burkina Faso
[5] Hawalbagh

Archive of SID.ir

Archive of SID.ir



تبخیر ماهانه در هندوستان پرداختند. نتایج ایشان نشان داد که دما، سرعت باد و ساعت آفتابی با میزان تبخیر همبستگی مثبت دارند، در حالی که رطوبت نسبی دارای همبستگی منفی می‌باشد. همچنین نتایج حاکی از عملکرد بهتر مدل‌های ANN نسبت به مدل‌های MLR برای تخمین میزان تبخیر ماهانه می‌باشد. Aparajita et al., (2019) با استفاده از مدل‌های شبکه عصبی مصنوعی (ANN) و مدل ترکیبی-موجک شبکه عصبی مصنوعی (WANN) به مدل‌سازی تبخیر ماهانه پرداختند. نتایج ایشان نشان داد که مقادیر تبخیر برآورد شده توسط مدل WANN دقیق‌تر از مقادیر برآوردی در مدل ANN می‌باشد. Malik et al., (2020) با استفاده از مدل‌های MM-ANN، MGGP، MARS، SVM و M5Tree میزان تبخیر از تشت در هند را مورد بررسی قرار دادند. بر اساس نتایج روش‌های شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن به عنوان کارآمدترین روش‌های تخمین تبخیر از تشت معرفی شدند. Araghi et al., (2010) با استفاده از روش‌های MLP، ANN و ANFIS و ترکیب این روش‌ها با تئوری موجک به تخمین تبخیر-تعرق مرجع در سه اقلیم مختلف ایران پرداختند. نتایج ایشان نشان داد که تعداد زیاد سطوح تجزیه سبب کاهش عملکرد مدل‌ها می‌شود. همچنین نتایج حاکی از عملکرد قابل قبول روش‌های داده‌کاوی جهت تخمین تبخیر-تعرق مرجع می‌باشد.

نبود آمار کافی در سنوات گذشته و احتمال خطا در اندازه‌گیری تشت تبخیر در پروژه‌های تحقیقاتی، محققان را بر آن داشته تا از روش‌های داده‌کاوی به جای روش‌های تجربی استفاده کنند. بنابراین در این پژوهش، جهت محاسبه مقدار تبخیر ماهانه از مدل‌های شبکه عصبی مصنوعی، سامانه استنتاجی فازی-عصبی تطبیقی و برنامه‌ریزی بیان ژن و همچنین هیبرید این روش‌ها با تئوری موجک استفاده شده است. هدف از پژوهش حاضر یافتن بهترین و مناسب‌ترین ترکیب پارامترهای هواشناسی به عنوان ورودی روش‌های داده‌کاوی و انتخاب مدل برتر جهت تخمین تبخیر ماهانه در حوضه‌های آبریز با شرایط اقلیمی مختلف با اعمال ضریب فصلی و نویززدایی داده‌ها می‌باشد. همچنین حوضه‌های آبریز منتخب جزء مهم‌ترین حوضه‌های تأمین آب شرب، کشاورزی و صنعت می‌باشند که در سال‌های اخیر شدیداً تحت تاثیر تغییرات اقلیمی بوده‌اند. از نماد تغییرات اقلیمی در این دو حوضه می‌توان به تغییرات مقدار آب در دریاچه ارومیه و زاینده‌رود اشاره کرد.

## مواد و روش‌ها

### منطقه مورد مطالعه و داده‌ها

با توجه به طبقه‌بندی اقلیمی کوپن گیجر[1]، دو حوضه آبریز دریاچه ارومیه و گاوخونی، به ترتیب دارای دو اقلیم مختلف سرد و خشک با تابستان گرم (Dsa[2]) و بیابانی سرد و خشک (Bsk[3]) هستند و در این تحقیق به منظور تخمین و بررسی تبخیر ماهانه انتخاب شده‌اند. حوضه دریاچه ارومیه در شمال‌غرب و حوضه گاوخونی در قسمت مرکزی ایران با مساحت به ترتیب ۵۱۸۷۶ و ۴۱۵۵۲/۳ کیلومترمربع واقع شده‌اند. موقعیت مکانی ایستگاه‌های هواشناسی منتخب تازه‌کند واقع در حوضه دریاچه ارومیه و کوهپایه در حوضه گاوخونی، در شکل ۱ نشان داده شده است. در این مطالعه، از داده‌های بارش ماهانه، دماهای میانگین، کمینه و بیشینه، رطوبت اندازه‌گیری شده در ساعات ۱۲ و ۱۸ و سرعت باد ماهانه برای تخمین تبخیر ماهانه در طول دوره آماری ۱۳۹۷-۱۳۸۴ ایستگاه‌های منتخب استفاده شده است. معیار انتخاب این پارامترها معنی‌دار بودن ارتباط همبستگی پیرسون آن‌ها با متغیر تبخیر در سطح معنی‌داری ۰/۰۵ می‌باشد. مشخصات ایستگاه‌های مورد مطالعه در جدول ۱ و همچنین مشخصات آماری داده‌های مورد استفاده در ایستگاه‌های منتخب در جدول ۲ آورده شده است. ضریب تغییرات تبخیر ماهانه برای ایستگاه

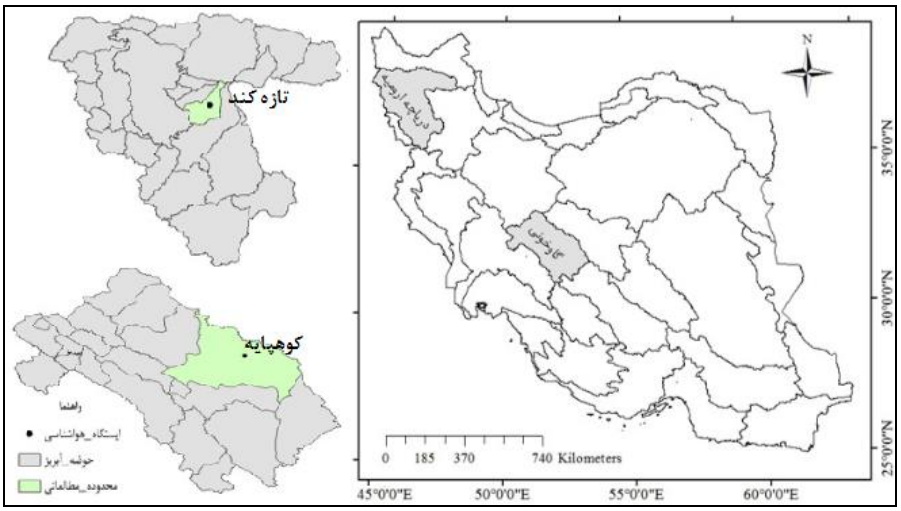
[1] Köppen-Geiger climate classification
[2] Snow dry, hot summer
[3] Arid Steppe cold

Archive of SID.ir

Archive of SID.ir



تازه‌کند و کوهپایه به ترتیب ۰/۷۶۹ و ۱/۰۲۱ می‌باشد که نشان‌دهنده تغییرات بیشتر میزان تبخیر در اقلیم بیابانی سرد و خشک می‌باشد.




**شکل (۱): موقعیت جغرافیایی ایستگاه‌های هواشناسی منتخب**

**جدول (۱): مشخصات ایستگاه‌های مورد مطالعه**

| نام ایستگاه | محدوده مطالعاتی | حوضه آبریز | اقلیم | ارتفاع از سطح دریا | طول جغرافیایی | عرض جغرافیایی |
|---|---|---|---|---|---|---|
| تازه‌کند | مراغه | دریاچه ارومیه | سرد و خشک با تابستان گرم | ۱۶۰۰ | ۴۶° ۱۶′ | ۳۷° ۲۸′ |
| کوهپایه | کوهپایه-سگز | گاوخونی | بیابانی سرد و خشک | ۱۸۰۰ | ۵۲° ۲۶′ | ۳۲° ۴۳′ |

**جدول (۲): مشخصات آماری متغیرهای به کارگرفته شده پژوهش**

| نام ایستگاه | متغیر | بارش | دمای میانگین | دمای کمینه | دمای بیشینه | سرعت باد | نمناکی ۱۸ | نمناکی ۱۲ | تبخیر |
|---|---|---|---|---|---|---|---|---|---|
| | | P | $T_{mean}$ | $T_{min}$ | $T_{max}$ | W | Nam18 | Nam12 | Evp |
| | واحد | mm | $^{\circ}C$ | $^{\circ}C$ | $^{\circ}C$ | km/hr | % | % | mm |
| تازه‌کند | میانگین | ۱۸/۴۲۰ | ۱۲/۸۶۸ | ۵/۰۹۰ | ۲۰/۴۱۹ | ۳۲۶۷/۶۲۵ | ۶۲/۳۳۰ | ۶۱/۰۵۱ | ۱۲۷/۰۶۲ |
| | انحراف معیار | ۲۳/۱۲۸ | ۸/۴۷۷ | ۷/۵۵۹ | ۹/۶۳۵ | ۳۰۹۸/۶۴۹ | ۱۴/۹۱۴ | ۱۲/۹۷۷ | ۹۷/۷۲۷ |
| | ضریب تغییرات | ۱/۲۵۶ | ۰/۶۵۹ | ۱/۴۸۵ | ۰/۴۷۲ | ۰/۹۴۸ | ۰/۲۳۹ | ۰/۲۱۳ | ۰/۷۶۹ |
| کوهپایه | میانگین | ۱۱/۰۲۵ | ۱۴/۹۳۶ | ۷/۰۲۱ | ۲۲/۵۴۱ | ۴۱۹۰/۳۹۰ | ۳۵/۱۱۳ | ۲۸/۸۱۵ | ۲۴۵/۸۱۴ |
| | انحراف معیار | ۱۵/۰۰۴ | ۹/۲۲۱ | ۸/۹۰۷ | ۹/۸۱۱ | ۱۴۲۴/۱۰۲ | ۱۶/۶۹۳ | ۱۲/۶۷۴ | ۲۵۱/۰۱۷ |
| | ضریب تغییرات | ۱/۳۶۱ | ۰/۶۱۷ | ۱/۲۶۸ | ۰/۴۳۵ | ۰/۳۴۰ | ۰/۴۷۵ | ۰/۴۴۰ | ۱/۰۲۱ |

### شبکه عصبی مصنوعی[1] (ANN)

ANN همانند مغز انسان شامل لایه‌هایی از نرون‌های متصل به هم می‌باشد. نرون‌ها مجموعه‌ای از ورودی‌ها و وزن‌ها را دریافت و با اعمال توابع ریاضی، خروجی‌ها را به صورت مجموعه‌ای از فعال‌سازها یا همان توابع محرک‌ها، تولید می‌کنند. ساختار شبکه‌های عصبی مصنوعی به صورت لایه‌ای است و از یک لایه ورودی، یک لایه خروجی و یک یا چند لایه پنهان (میانی) تشکیل شده است. هر لایه شامل تعدادی گره، نرون یا واحد است که با وزن‌های متفاوت به صورت شبکه‌ای به هم متصل می‌شوند (Montaseri et al., 2018). در این مطالعه،

[1] Artificial Neural Networks

Archive of SID.ir

Archive of SID.ir



از الگوریتم لونبرگ-مارگارت جهت آموزش مدل استفاده شده است (Emadi et al., 2021). از مزایا و دلایل به کارگیری روش شبکه عصبی مصنوعی با الگوریتم آموزشی لونبرگ-مارگارت در این مطالعه می‌توان به این نکته اشاره کرد که الگوریتم مذکور جهت رسیدن به سرعت آموزشی مرتبه دوم بدون محاسبه ماتریس هسین طراحی شده است. ماتریس ژاکوبین به کارگرفته شده در ساختار این الگوریتم می‌تواند از طریق روش انتشار برگشتی استاندارد محاسبه گردد که پیچیدگی کمتری نسبت به محاسبه ماتریس هسین داشته و همچنین سریعتر و دقیق‌تر است.

### سامانه استنتاجی فازی-عصبی تطبیقی[1] (ANFIS)

ترکیب سیستم‌های فازی که مبتنی بر قواعد منطقی بوده، و روش شبکه‌های عصبی مصنوعی چند لایه که توان استخراج دانش از اطلاعات عددی را دارند، منجر به ارائه سیستم استنتاج تطبیقی فازی-عصبی شده است (Jang, 1933). این مدل توسط Jang (1993) بسط یافت. در این روش ترکیبی، بخش فازی، رابطه بین متغیرهای ورودی و خروجی را ایجاد کرده و پارامترهای مربوط به توابع عضویت بخش فازی به وسیله شبکه عصبی بهینه می‌شود. از یک ساختار ANFIS که از مجموعه ای از قواعد اگر-آنگاه فازی نوع تاکاگی-سوگنو (TSK) (فقط برای یک قانون) تشکیل گردیده است، می‌توان جهت مدلسازی و نگاشت داده‌های ورودی-خروجی استفاده کرد. اگر یک سیستم استنتاج فازی با دو متغیر ورودی x و y و یک متغیر خروجی $f$ در نظر گرفته شود، مدل فازی مرتبه اول سوگنو، که شامل دو قانون فازی IF-THEN(اگر-آنگاه) می‌باشد، که به صورت زیر است:

قانون اول: اگر x برابر $A_1$ و y برابر $B_1$ باشد آنگاه

$$f_1 = p_1A_1 + q_1B_1 + r_1 \quad (1)$$

قانون دوم: اگر x برابر $A_2$ و y برابر $B_2$ باشد آنگاه

$$f_2 = p_2A_2 + q_2B_2 + r_2 \quad (2)$$

در اینجا $f$ متغیر خروجی می‌باشد که از طریق متوسط وزن‌دار به دست می‌آید. تعریف معمولی از این مدل شناسایی یک تابع $\hat{f}$ (به طوری که تقریباً بتواند بجای تابع اصلی $f$ مورد استفاده قرار گیرد) می‌باشد (Emadi et al., 2021).

### برنامه‌ریزی بیان ژن[2] (GEP)

برنامه‌ریزی بیان ژن، توسعه یافته برنامه‌ریزی ژنتیک، که آن هم شاخه‌ای از الگوریتم ژنتیک است (Holland, 1975)، می‌باشد. ابتدا الگوریتم GEP، با کمک راه‌حل‌ها شروع به تولید جمعیت اولیه بوسیلهٔ فرآیند تصادفی و یا با استفاده از مقداری اطلاعات درباره مسئله می‌کند. کروموزوم‌ها که به صورت بیان درختی (ETs) نشان داده می‌شوند، مطابق با یک تابع برازش ارزیابی می‌گردند تا میزان مناسب بودن یک راه‌حل برای مسئله تعیین گردد. برای ارزیابی تابع برازش از پردازش تعدادی نمونه از مسئله هدف (موارد برازش)، استفاده می‌گردد. شرط توقف روند تکامل این است که کیفیت رضایت‌بخش از یک راه‌حل پیدا شود و یا نسل‌ها به تعداد معینی برسد (Emadi et al., 2021). اما اگر هیچکدام از این شرایط توقف یافت نشود، بهترین راه‌حل از نسل حاضر نگه داشته می‌شود و بقیه راه‌حل‌ها به فرآیندی گزینشی واگذار می‌شوند. این عمل به معنای نخبه‌گزینی است و بر اساس آن بهترین افراد شانس بهتری برای تولید فرزندان دارند. بنابراین، با پیش رفتن نسل به جلو و تکرار روند انتخاب بهترین افراد، انتظار می‌رود که کیفیت جمعیت نیز به طور متوسط بهبود یابد (Lopes and Weinert, 2004).

[1] Adaptive-Network-based Fuzzy Inference Systems

[2] Gene-Expression Programming

Archive of SID.ir

Archive of SID.ir



### تبدیل موجک[1]

توابعی هستند که از صفر شروع شده و به صورت نوسانی به صفر ختم می‌شوند. این توابع یک روش داده محور[2] را تشکیل می‌دهند که "موجک" نامیده می‌شود. توابعی موجود در موجک قادر هستند داده‌ها را به مقادیر فرکانسی مختلف آن جداسازی کنند (Emadi et al., 2021). هر مؤلفه‌ای از داده به وسیله تفکیک‌پذیری متناسب با مقیاس آن ارزیابی می‌شود. مقیاس‌های پایین، جزئیاتی از سیگنال که سریع تغییر می‌کنند را نشان داده و مقیاس‌های بالا، تغییرات کند فرکانسی اجزا را نشان می‌دهند (Kucuk and Agiralioglu, 2006). لذا، می‌توان گفت تبدیل موجک ابزاری ریاضیاتی است که برای استخراج اطلاعات از انواع مختلف داده‌ها، از طریق تجزیه و تحلیل انحرافات محلی در سری‌های زمانی، به کار می‌رود (Emadi et al., 2021). به عبارت دیگر از طریق تجزیه محتوای فرکانسی سیگنال‌ها در سری‌های زمانی و اشکال، نوسانات زمانی مناسب از سیگنال‌های غیر ثابت و موقت استخراج شده و به این ترتیب داده‌ها را تجزیه و تحلیل می‌کند. که این نشان می‌دهد موجک پویایی فرکانس پایین سری‌های زمانی غیرثابت را به خوبی تجزیه کرده، نویز را نشان داده، و ابزاری قوی در تعیین الگوهای فصلی و تغییر واریانس است (Popoola, 2007).

تبدیل موجک پیوسته و گسسته دو نوع کلی تبدیل موجک هستند. تبدیل موجک گسسته که داده‌های اصلی را به باندها تجزیه می‌کند در ارتقای پیش‌بینی‌ها مؤثر است. در این پژوهش سطح تجزیه موجک بر اساس رابطه (۳) مشخص شد، که در آن N طول دوره آماری و L سطح تجزیه را نشان می‌دهد. از آنجاییکه تعداد نمونه‌ها (داده‌ها) در هر دو ایستگاه تازه‌کند و کوهپایه ۱۵۶ می‌باشد، لذا تجزیه موجک در سطح ۱ انجام گردید.

(۳)

$$L = \mathrm{int}[\log(N)]$$

توانایی‌های منحصربفرد روش موجک موجب شده است که هیبرید آن با روش‌های دیگر توسط محققین توسعه داده شود. در این مطالعه، روش موجک با هر سه روش ANFIS ANN، و GEP هیبرید و نتایج آنها با هم مقایسه شد. ساختار مدل‌های به کارگرفته شده در شکل‌های ۲ تا ۴ نشان داده شده است. انتخاب تعداد نرون‌های بهینه در روش شبکه عصبی در بازه ۱ تا ۱۵ و شعاع دسته در روش سامانه استنتاجی فازی عصبی تطبیقی در بازه ۰/۸-۰/۲ بوده است.

### معیارهای ارزیابی دقت مدل‌ها

به منظور ارزیابی مدل‌ها، از شاخص‌های آماری ضریب همبستگی[3] (R)، ریشه میانگین مربعات خطا[4] (RMSE)، میانگین خطای مطلق[5] (MAE) و ضریب کارایی نش-ساتکلیف[6] (NSE) مطابق روابط (۴) تا (۷) استفاده شده است.

(۴)

$$R = \left[ \frac{\sum_{i=1}^{N} (Evp_{io} - \overline{Evp}_o)(Evp_{ie} - \overline{Evp}_e)}{\sqrt{\sum_{i=1}^{N} (Evp_{io} - \overline{Evp}_o)^2 \sum_{i=1}^{N} (Evp_{ie} - \overline{Evp}_e)^2}} \right]$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Evp_{io} - Evp_{ie})^2}$$

(۵)

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |Evp_{io} - Evp_{ie}|$$

(۶)

$$NSE = 1 - \frac{(Evp_{ie} - Evp_{io})^2}{(Evp_{io} - \overline{Evp_{io}})^2} \qquad (۷)$$

[1] Wavelet Transition
[2] Data Driven

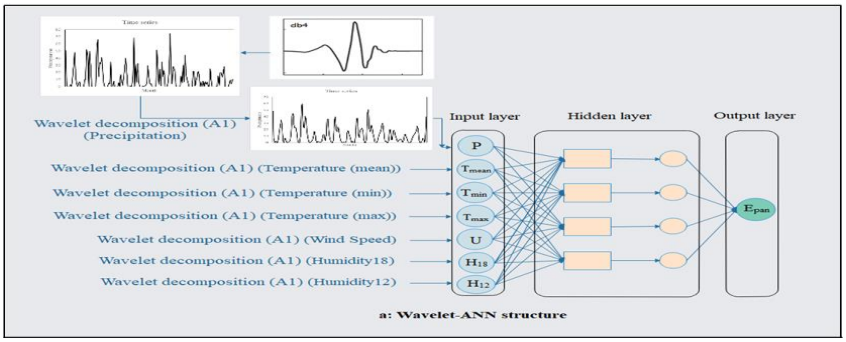
[3] Correlation Coefficient
[4] Root Mean Square Error
[5] Mean Absolute Error
[6] Nash–Sutcliffe Model Efficiency Coefficient

Archive of SID.ir

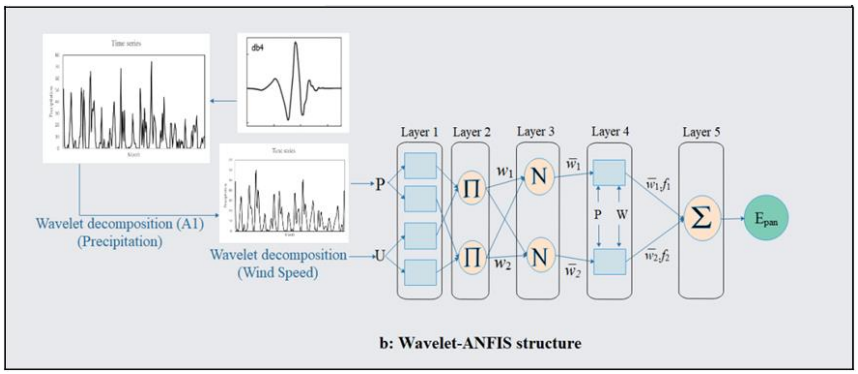
Archive of SID.ir



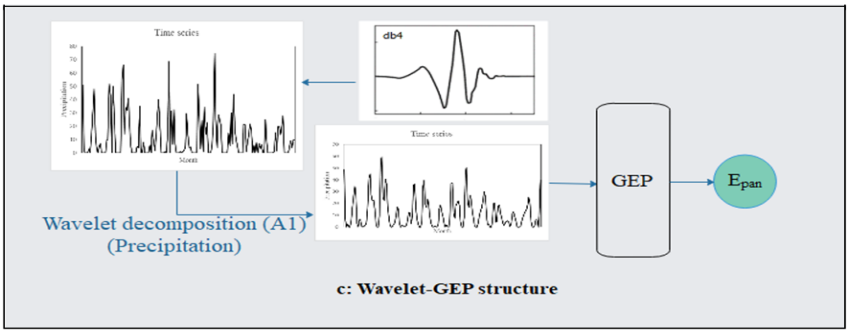
در روابط (۴) تا (۷) $Evp_{io}$ و $Evp_{ie}$ مقادیر تبخیر مشاهداتی و شبیه‌سازی شده و N تعداد کل مشاهدات است. $\overline{Evp_o}$ و $\overline{Evp_e}$ میانگین مقادیر تبخیر مشاهداتی و شبیه‌سازی شده است. بهترین مدل، مدلی خواهد بود که مقدارهای چهار معیار برای آن به ترتیب، یک، نزدیک به صفر، صفر و یک باشد




شکل (۲): ساختار مدل هیبرید-موجک شبکه عصبی مصنوعی




شکل (۳): ساختار مدل هیبرید-موجک سیستم استنتاج فازی عصبی تطبیقی




شکل (۴): ساختار مدل هیبرید-موجک برنامه‌ریزی بیان ژن

Archive of SID.ir

Archive of SID.ir



## بحث و نتایج

در ابتدا پارامترهای هواشناسی موردنظر انتخاب شدند. معیار انتخاب پارامترهای به کار رفته در این پژوهش معنی‌دار بودن ارتباط همبستگی پیرسون آنها با متغیر تبخیر در سطح ۰/۰۵ بود. سپس طول دوره آماری پارامترهای موردنظر تعیین شد. در نهایت مدل‌های منفرد شبکه عصبی مصنوعی، سامانه استنتاجی فازی عصبی تطبیقی و برنامه‌ریزی بیان ژن و هیبرید-موجک این مدل‌ها اجرا شدند.

### آنالیز موجک

آنالیز موجک توانایی تجزیه داده‌ها را به زیرسری-های اصلی و فرعی دارد. هدف از این کار جداسازی زیرسری‌های فرعی از اصلی و یا به عبارت دیگر نویززدایی داده‌ها می‌باشد. زیرا این اجزای فرعی داده‌ها منجر به ایجاد نویز و یا خطا در مدلسازی شده و عملکرد مدل‌های داده‌کاوی را در حالت هیبرید-موجک کاهش می‌دهد. لازم به ذکر است که استفاده از تئوری موجک و نویززدایی داده‌ها، ماهیت داده‌ها را تغییر نمی‌دهد بلکه صرفا نویزها و عدم قطعیت در پدیده‌های هیدرولوژیکی را از بین می‌برد. جدول ۳ نتایج آنالیز موجک به تفکیک زیرسری‌های اصلی (A1) و (D1) را برای متغیرهای به کارگرفته شده در دو ایستگاه تازه‌کند و کوهپایه که به ترتیب در حوضه آبریز دریاچه ارومیه و گاوخونی قرار دارد، نشان می-دهد. با توجه به نتایج مندرج در جدول ۳ مقادیر کمینه و بیشینه حاصل برای زیرسری اصلی و فرعی متغیر تبخیر به ترتیب برای دو ایستگاه تازه‌کند و کوهپایه (A1: ۱۴/۵۴۴-، ۲۸۳/۰۳۲+؛ D1: ۶۰/۲۶۱-، ۵۹/۰۳۶+) و (A1: ۳۶/۴۸۸-، ۹۱۶/۴۰۳+؛ D1: ۱۱۳/۶۵۱-، ۱۰۶/۶۹۹+) بدست آمده است

**جدول (۳): نتایج آنالیز موجک**

| نام ایستگاه | سری تجزیه | حداقل یا حداکثر | ضریب فصلی | بارش | دمای میانگین | دمای کمینه | دمای بیشینه | سرعت باد | نمناکی ۱۸ | نمناکی ۱۲ | تبخیر |
|---|---|---|---|---|---|---|---|---|---|---|---|
| تازه‌کند | A1 | حداقل | ۱/۴۸۹ | -۱۱/۵۰۲ | -۴/۳۶۵ | -۱۰/۵۳۲ | -۱/۲۶۴ | -۳۳۷۹/۳۹۸ | ۲۵/۲۲۴ | ۳۷/۰۵۷ | -۱۴/۵۴۴ |
| | | حداکثر | ۱۱/۹۹۵ | ۷۱/۲۵۰ | ۲۵/۴۸۱ | ۱۶/۷۷۸ | ۳۴/۷۰۴ | ۲۲۴۵۶/۵۹۳ | ۹۱/۴۳۳ | ۹۰/۶۵۱ | ۲۸۳/۰۳۲ |
| | D1 | حداقل | -۴/۲۳۱ | -۶۷/۸۱۵ | -۳/۶۵۴ | -۲/۸۶۸ | -۲/۹۲۵ | -۱۸۴۳۴/۷۶۰ | -۸/۸۴۱ | -۸/۷۷۶ | -۶۰/۲۶۱ |
| | | حداکثر | ۴/۰۷۶ | ۷۰/۲۵۰ | ۳/۱۸۳ | ۳/۲۱۷ | ۳/۲۱۴ | ۱۹۶۷۱/۴۰۷ | ۸/۷۷۶ | ۱۰/۶۴۲ | ۵۹/۰۳۶ |
| کوهپایه | A1 | حداقل | ۰/۹۶۸ | -۹/۱۱۵ | -۱/۱۵۴ | -۸/۴۶۱ | ۶/۱۲۳ | ۲۰۵۱/۳۹۱ | ۱۰/۶۰۲ | ۸/۹۲۳ | -۳۶/۴۸۸ |
| | | حداکثر | ۱۱/۱۵۶ | ۴۷/۷۹۰ | ۲۸/۹۶۰ | ۲۱/۰۱۵ | ۳۷/۴۸۰ | ۸۹۳۸/۷۷۶ | ۳۷/۸۸۱ | ۵۹/۲۵۲ | ۹۱۶/۴۰۳ |
| | D1 | حداقل | -۴/۲۳۱ | -۲۳/۸۸۹ | -۵/۰۰۰ | -۲/۴۲۴ | -۷/۷۸۶ | -۱۱۸۷/۳۷۶ | -۲۴/۶۳۰ | -۱۲/۴۸۷ | -۱۱۳/۶۵۱ |
| | | حداکثر | ۴/۰۷۶ | ۲۲/۹۲۱ | ۵/۷۲۶ | ۲/۸۵۴ | ۷/۸۰۵ | ۱۱۶۵/۷۸۰ | ۲۰/۷۸۰ | ۱۱/۱۴۹ | ۱۰۶/۶۹۹ |

### مدل‌های منفرد و هیبریدی-موجک

همان طور که در قسمت‌های قبل توضیح داده شد برای تخمین متغیر تبخیر ماهانه در دو ایستگاه تازه‌کند و کوهپایه از سه مدل منفرد شبکه عصبی مصنوعی، سامانه استنتاجی فازی عصبی تطبیقی و برنامه‌ریزی بیان ژن و سه مدل هیبرید-موجک آنها استفاده شد. در روش شبکه عصبی مصنوعی منفرد و هیبریدی-موجک آن برای تخمین میزان تبخیر ماهانه در ایستگاه‌های تازه‌کند و کوهپایه تعداد نرون در لایه پنهان به ترتیب (۳، ۴) و (۵، ۳) به دست آمد. همچنین، نوع توابع محرک در لایه پنهان برای روش منفرد و هیبریدی-موجکی شبکه عصبی در هر دو اقلیم (دریاچه ارومیه Dsa و حوضه آبریز گاوخونی Bsk)، تانژانت سیگموئید به دست آمد. در روش منفرد و هیبریدی-موجک سامانه استنتاجی فازی-عصبی تطبیقی به ترتیب برای اقلیم‌های Dsa و Bsk شعاع دسته (۰/۳۸، ۰/۳۵) و (۰/۴۵، ۰/۳۲) به دست آمد. نتایج و مشخصات مدل‌های به کار گرفته شده منفرد

Archive of SID.ir

Archive of SID.ir



ANN، ANFIS و GEP و همچنین هیبریدی-موجک WANN، WANFIS و WGEP توسط مقادیر R، RMSE، MAE و NSE در جدول ۴ آورده شده است. به طور کلی می‌توان گفت در این پژوهش مدل‌های هیبرید-موجکی نتایج بهتری را نسبت به مدل‌های منفرد ارائه داده‌اند. در میان تمامی مدل‌ها برای هر دو ایستگاه تازه‌کند و کوهپایه، مدل WGEP بهترین مدل و ANN ضعیف‌ترین مدل بوده است. در میان مدل‌های منفرد مدل GEP برای هر دو ایستگاه عملکرد بهتری نسبت به سایر مدل‌های منفرد داشته است هرچند که همین مدل GEP نسبت به مدل‌های هیبریدی نتایج ضعیف‌تری را ارائه داده است. بالاترین مقدار ضریب همبستگی و ضریب کارایی نش-ساتکلیف مربوط به ایستگاه تازه‌کند و مدل WGEP است و مقدار آن به ترتیب ۰/۹۷۶ و ۰/۹۵۳ به دست آمد. همچنین آماره‌های RMSE و MAE در این مدل کمترین مقدار را دارد که میزان آن به ترتیب ۲۰/۸۷۰ و ۱۵/۴۳۶ محاسبه شده است. کمترین مقدار ضریب همبستگی و ضریب کارایی نش-ساتکلیف و بیشترین مقدار RMSE و MAE مربوط به مدل ANN منفرد در ایستگاه کوهپایه است. به طور کلی بر اساس نتایج جدول ۴، مدل‌ها در حوضه دریاچه ارومیه با اقلیم Dsa عملکرد بهتری نسبت به مدل‌ها در حوضه گاوخونی با اقلیم Bsk داشته‌اند. به عبارت دیگر نتایج نشان می‌دهد که توانایی مدل‌ها برای شبیه‌سازی تبخیر در اقلیم خشک با تابستان گرم بیشتر بوده است. بر اساس نتایج مدل WGEP به عنوان بهترین مدل در هر دو ایستگاه معرفی شد. درصد بهبود عملکرد مدل WGEP در ایستگاه تازه‌کند نسبت به مدل GEP برمبنای RMSE، ۲۵ درصد و همچنین درصد بهبود عملکرد مدل WGEP نسبت به WANFIS (که رتبه دوم بهترین عملکرد بعد از WGEP را دارد) ۹ درصد می‌باشد. همچنین، در ایستگاه کوهپایه درصد بهبود مدل WGEP نسبت به مدل GEP، ۴ درصد و درصد بهبود مدل WGEP نسبت به مدل WANFIS، یک درصد می‌باشد. در شکل‌های ۵ و ۶ مقادیر مشاهداتی و تخمینی تبخیر ماهانه در ایستگاه‌های تازه‌کند و کوهپایه با استفاده از مدل‌های منفرد و هیبریدی-موجک نشان داده شده است. از آنجایی که روند تغییرات داده‌های مدل شده شباهت زیادی با روند تغییرات موجک مادر دابچیز نوع چهارم دارد، لذا این نوع موجک مادر می‌تواند بهترین عملکرد را در مدلسازی هیبریدی داشته باشد. بر اساس معادله برازش داده شده بر داده‌های مشاهداتی و برآوردی (y=ax)، نزدیک بودن مقدار ضریب a به عدد ۱ و یا نیمساز ربع اول نشان‌دهنده عملکرد بهتر مدل‌ها می‌باشد. مقادیر ضریب a برای مدل‌های شبکه عصبی مصنوعی، سامانه استنتاجی فازی عصبی تطبیقی و برنامه‌ریزی بیان ژن به ترتیب برای ایستگاه‌های تازه‌کند و کوهپایه (۰/۹۹۳، ۰/۹۶۷ ، ۱/۰۳۶) و (۰/۶۵۸، ۰/۶۵۷ ، ۰/۶۵۶) به دست آمد. مقادیر ضریب a برای مدل‌های هیبریدی-موجک شبکه عصبی مصنوعی، هیبریدی-موجک سامانه استنتاجی فازی عصبی تطبیقی و هیبریدی-موجک برنامه‌ریزی بیان ژن به ترتیب برای ایستگاه‌های تازه‌کند و کوهپایه (۰/۹۹۹، ۰/۹۷۱ ، ۱/۰۳۱) و (۰/۶۶۴، ۰/۶۶۲ ، ۰/۶۶۱) به دست آمده است. به عبارت دیگر، ضرایب مربوط به مدل‌های هیبریدی بزرگتر از ضرایب مربوط به مدل‌های منفرد می‌باشد. شکل‌های ۵ و ۶ نشان می‌دهد که مدل‌ها در حوضه دریاچه ارومیه با اقلیم Dsa عملکرد بهتری را از خود نشان داده‌اند. شکل‌های ۷ و ۸ مقایسه نتایج حاصل از مدل‌های منفرد و هیبرید-موجک تخمینی در ایستگاه‌های تازه‌کند و کوهپایه را نشان می‌دهد

Archive of SID.ir

Archive of SID.ir



**جدول (۴): نتایج و مشخصات مدل‌های به کار گرفته شده منفرد و هیبریدی-موجک در دوره آزمون**

| نام ایستگاه | تازه‌کند | | | |
|---|---|---|---|---|
| نام مدل | R | RMSE (mm) | MAE (mm) | NSE |
| ANN (tansig-tansig-۳) | **۰/۹۴۸** | **۳۱/۲۳۱** | **۲۱/۳۸۹** | **۰/۹۰۱** |
| ANFIS (۰/۳۸) | ۰/۹۵۳ | ۲۹/۰۱۵ | ۱۹/۸۳۳ | ۰/۹۱۴ |
| GEP | ۰/۹۵۶ | ۲۷/۹۴۱ | ۱۸/۱۸۷ | ۰/۹۲۰ |
| WANN (tansig-logsig-۴) | ۰/۹۶۸ | ۲۴/۴۴۶ | ۱۸/۴۴۰ | ۰/۹۳۶ |
| WANFIS (۰/۳۵) | ۰/۹۷۰ | ۲۳/۰۵۱ | ۱۶/۸۸۵ | ۰/۹۴۳ |
| WGEP | **۰/۹۷۶** | **۲۰/۸۷۰** | **۱۵/۴۳۶** | **۰/۹۵۳** |

**جدول (۴): نتایج و مشخصات مدل‌های به کار گرفته شده منفرد و هیبریدی-موجک در دوره آزمون**

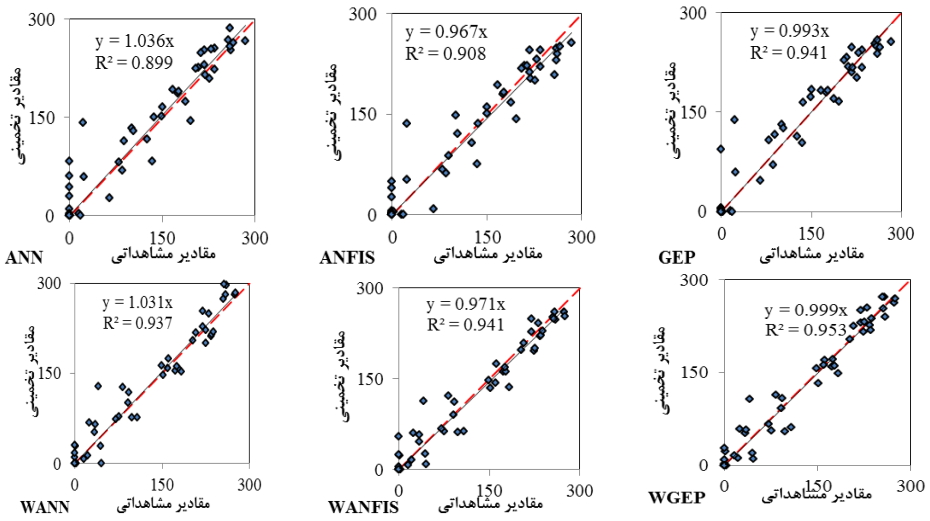
| نام ایستگاه | کوهپایه | | | |
|---|---|---|---|---|
| نام مدل | R | RMSE (mm) | MAE (mm) | NSE |
| ANN (tansig-tansig-۵) | **۰/۹۳۳** | **۱۶۶/۰۱۴** | **۱۱۷/۰۱۶** | **۰/۷۰۸** |
| ANFIS (۰/۴۵) | ۰/۹۳۶ | ۱۶۵/۷۹۱ | ۱۱۶/۴۰۲ | ۰/۷۰۸ |
| GEP | ۰/۹۴۴ | ۱۶۳/۶۰۹ | ۱۱۴/۴۶۶ | ۰/۷۱۶ |
| WANN (tansig-tansig-۳) | ۰/۹۴۹ | ۱۵۹/۹۳۱ | ۱۱۳/۱۰۲ | ۰/۷۲۳ |
| WANFIS (۰/۳۲) | ۰/۹۵۵ | ۱۵۸/۲۶۶ | ۱۱۱/۸۲۵ | ۰/۷۲۸ |
| WGEP | **۰/۹۵۸** | **۱۵۶/۸۸۴** | **۱۱۰/۹۹۴** | **۰/۷۳۳** |

شکل ۹ نمودار جعبه‌ای داده‌های تخمینی مدل هیبرید موجک برنامه‌ریزی بیان ژن برای دو ایستگاه تازه‌کند و کوهپایه در دو اقلیم مختلف را نشان می‌دهد. این نمودار درک بهتری از توزیع داده‌های تبخیر تخمینی را ارائه می‌دهد. نمودار جعبه‌ای یک روش استاندارد برای نمایش توزیع داده‌ها است که براساس شاخص‌های آماری کوچکترین مقدار چارک اول (۰/۲۵ درصد)، میانه ۵۰ درصد، چارک سوم (۰/۷۵ درصد) و بزرگترین مقدار ساخته شده است. این نمودار همچنین می‌تواند تقارن در داده‌ها را نشان دهد. با توجه به چگونگی پراکندگی داده‌ها ایستگاه کوهپایه در اقلیم Bsk بیشترین مقدار میانه و ایستگاه تازه‌کند واقع در اقلیم Dsa کمترین مقدار میانه را دارد.

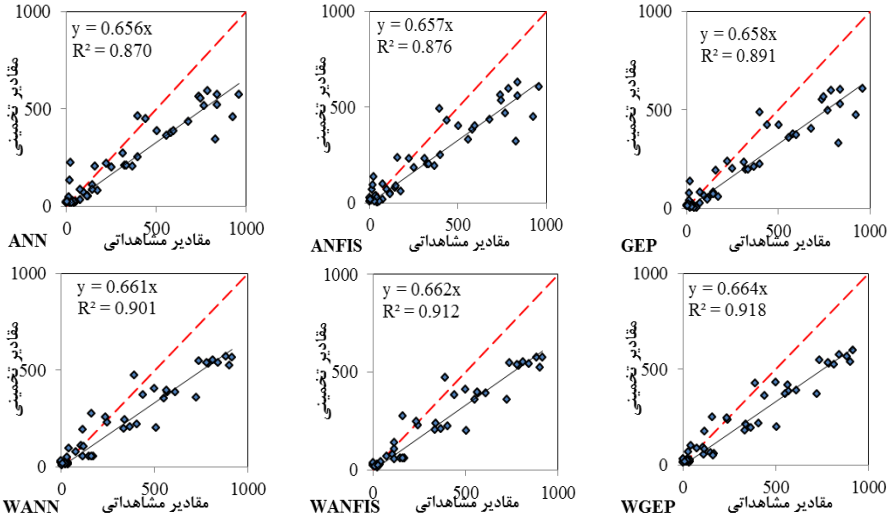
Archive of SID.ir

Archive of SID.ir

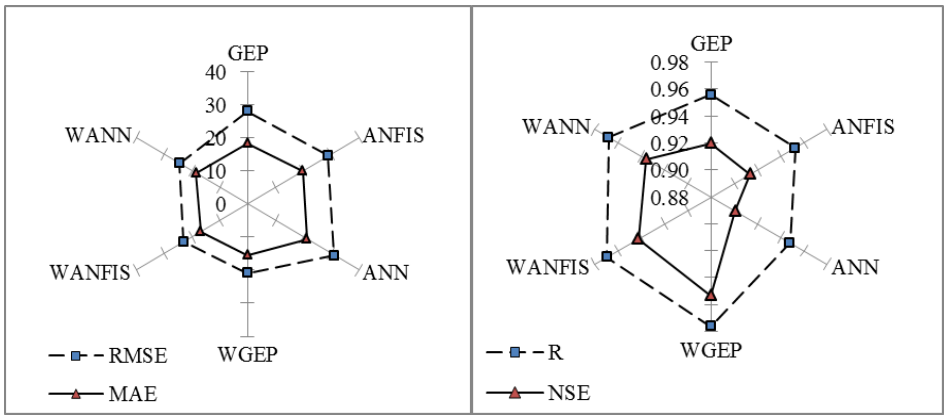





شکل (۵): مقادیر مشاهداتی و تخمینی تبخیر ماهانه در ایستگاه تازه‌کند با استفاده از مدل‌های منفرد و هیبریدی-موجک در دوره آزمون




شکل (۶): مقادیر مشاهداتی و تخمینی تبخیر ماهانه در ایستگاه کوهپایه با استفاده از مدل‌های منفرد و هیبریدی-موجک در دوره آزمون

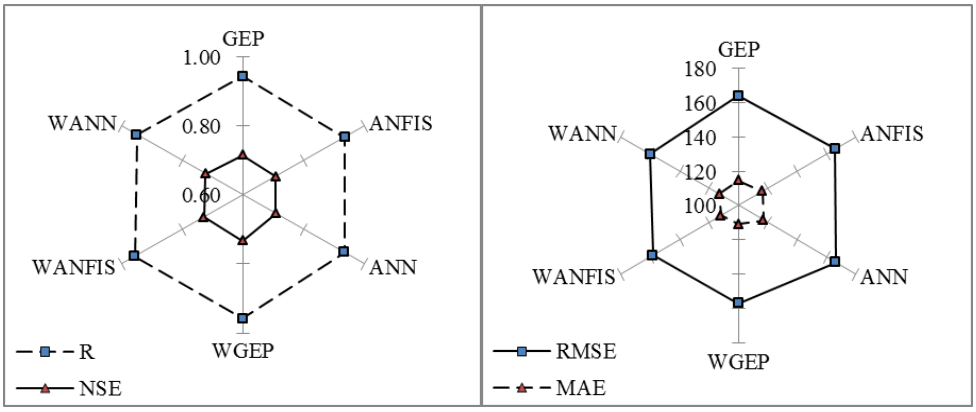



شکل (۷): مقایسه نتایج حاصل از مدل‌های تخمینی در ایستگاه تازه‌کند واقع در حوضه آبریز دریاچه ارومیه (دوره آزمون)

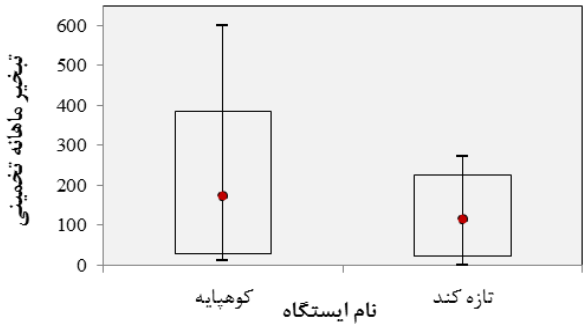
Archive of SID.ir

Archive of SID.ir






**شکل (۸): مقایسه نتایج حاصل از مدل‌های تخمینی در ایستگاه کوهپایه واقع در حوضه آبریز گاوخونی (دوره آزمون)**




**شکل (۹): توزیع داده‌های تخمینی مدل هیبرید-موجک برنامه‌ریزی بیان ژن (دوره آزمون)**

یکی از ویژگی‌های منحصر به فرد برنامه‌ریزی بیان ژن ارائه رابطه ریاضی کاربردی جهت استفاده از نتایج مدلسازی در زمان و مکان‌های مختلف می‌باشد. در جدول ۵ نتایج معادلات ریاضی مستخرج از مدل‌های بهینه برنامه‌ریزی بیان ژن در حالت منفرد و هیبرید-موجک در دو ایستگاه مورد مطالعه با شرایط اقلیمی مختلف آورده شده است. از این روابط می‌توان جهت تخمین میزان تبخیر ماهانه در مناطق هم اقلیم با مطالعات موردی این تحقیق، استفاده نمود. استفاده از روابط ریاضی به‌دست آمده می‌تواند در جلوگیری هدر رفت زمان و هزینه جهت اندازه‌گیری میزان تبخیر بسیار مؤثر واقع شود.

**جدول (۵): معادلات ریاضی مستخرج از برنامه‌ریزی بیان ژن در حالت منفرد و هیبرید-موجک**

| حوضه | ایستگاه | مدل | معادله |
|---|---|---|---|
| دریاچه ارومیه | تازه‌کند | منفرد | $Evp_{x,t}=\left(\tan^{-1}\left(\left(\cos\left(P_{x,t}\right)\right)-T\min_{x,t}\right)\right)+\left(3\times T\max_{x,t}+4.93\right)+\left(\left(2.88\times T\max_{x,t}\right)\times\left(\tan^{-1}\left(T\min_{x,t}\right)\right)\right)$ |
| | | هیبرید-موجک | $Evp_{x,t}=\left(T\max_{x,t}-M_{x,t}\right)+\left(\left(\sqrt[2]{\left(\left(T\max_{x,t}\right)+\left(4+T\max_{x,t}\right)\right)}\right)\times\left(T\max_{x,t}-7.33\right)\right)+\left(\frac{H12_{x,t}}{M_{x,t}}\right)$ |
| گاوخونی | کوهپایه | منفرد | $Evp_{x,t}=\left(T\max_{x,t}\times\left(\left(Cos\left(\sqrt[3]{P_{x,t}}\right)\right)^2\right)\right)+\left(\frac{Sin\left(T\min_{x,t}-2.16\right)}{Sin\left(T\max_{x,t}\right)}\right)+\left(T\min_{x,t}\times Tmean_{x,t}\right)+\left(2\times M_{x,t}\right)$ |
| | | هیبرید-موجک | $Evp_{x,t}=\left(H12_{x,t}\right)+\left(T\min_{x,t}\times Tmean_{x,t}\right)+\left(\left(-9.03+\left(\left(Cos\left(H12_{x,t}\right)\right)\times Tmean_{x,t}\right)\right)+\left(Tmean_{x,t}\right)\right)$ |

ساختار خاص و منحصر به فرد هر مدل عامل اصلی تفاوت عملکرد مدل‌ها است. با توجه به اینکه اساس روش برنامه‌ریزی بیان ژن تولید کروموزوم است، این روش نسبت به عملکرد سایر مدل‌ها دارای اولویت بالایی است. همچنین روش ANFIS که ترکیبی از قوانین فازی و ساختار نورون‌ها است، عملکرد نسبی بالایی در مقایسه با روش شبکه عصبی دارد، زیرا شبکه عصبی مصنوعی فقط ساختار نورون‌ها

Archive of SID.ir

Archive of SID.ir



را شامل می‌شود. تجزیه و تحلیل موجک با از بین بردن نویز داده‌ها و امکان برقراری روابط غیرخطی پیچیده در ساختار آن، عملکرد مدل‌های منفرد را در کلیه روش‌ها و اقلیم‌های مختلف ارتقا می‌بخشد. شرایط آب و هوایی فقط بر میزان عملکرد مدل‌ها تأثیر می‌گذارد و نه بر اولویت آن. با این حال نتایج سه روش منفرد و هیبرید-موجک آنها می‌تواند برای برآورد متغیر تبخیر ماهانه در حوضه ها با انواع آب و هوایی Dsa و Bsk قابل قبول باشد. از طرف دیگر، در تحقیقات فعلی، استفاده از ضریب فصلی برای هر دو نوع اقلیمی Dsa و Bsk یکی دیگر از عواملی است که نتایج را برای برآورد تبخیر ماهانه قابل قبول و قوی می‌کند. جدول ۶ رتبه‌بندی فعلی، مدل‌های به کار گرفته شده به لحاظ عملکرد آنها را نشان می‌دهد. نتایج حاکی از آن است که در دو اقلیم مختلف، روش‌های هیبرید-موجک برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی منفرد به ترتیب بالاترین و ضعیف‌ترین عملکرد در میان سایر مدل‌های داده کاوی داشتند

.

**جدول (۶): رتبه‌بندی مدل‌های به کار گرفته شده**

| حوضه آبریز | گاوخونی | دریاچه ارومیه |
|---|---|---|
| نام ایستگاه | کوهپایه | تازه‌کند |
| رتبه‌بندی | | |
| ۱ | **WGEP** | **WGEP** |
| ۲ | WANFIS | WANFIS |
| ۳ | WANN | WANN |
| ۴ | GEP | GEP |
| ۵ | ANFIS | ANFIS |
| ۶ | ANN | ANN |

در این پژوهش مدل WGEP به‌عنوان بهترین مدل در هر دو اقلیم انتخاب شد. بر این اساس نتایج حاصل از شبیه‌سازی متغیر تبخیر توسط این مدل در هر دو ایستگاه تازه‌کند و کوهپایه به سه دسته ۳۰ درصد کمینه، ۴۰ درصد میانه و ۳۰ درصد بیشینه تقسیم‌بندی شد. این دسته‌بندی به همراه ضرایب همبستگی آن‌ها در جدول ۷ آمده است. نتایج حاکی از این بود که مدل هیبرید-موجک برنامه‌ریزی بیان ژن در تخمین ۴۰ درصد میانه در هر دو ایستگاه عملکرد بهتری داشته و ضریب همبستگی در این حالت برای هر دو ایستگاه تازه‌کند و کوهپایه به ترتیب ۰/۹۲۸ و ۰/۸۵۸ بوده است. همچنین، ضعیف‌ترین برآورد مربوط به ۳۰ درصد بیشینه در هر دو اقلیم با ضرایب همبستگی ۰/۶۰۶ و ۰/۸۴۰ بوده است. سپس، خطای نسبی تخمینی حاصل در بازه مذکور برای هر دو ایستگاه تازه‌کند و کوهپایه محاسبه شد و نتایج آن در جدول ۸ ذکر شده است. میانگین خطای نسبی در ایستگاه تازه‌کند و کوهپایه به ترتیب برابر با ۰/۰۵۵ و ۰/۳۴۶ می‌باشد

.

Archive of SID.ir

Archive of SID.ir



جدول (۷): بازه‌بندی نتایج حاصل از مدل هیبرید-موجک برنامه‌ریزی بیان ژن به همراه ضریب همبستگی آنها

| نام ایستگاه | بازه‌بندی | ضریب همبستگی |
|---|---|---|
| تازه‌کند | ۳۰% min | ۰/۸۰۴ |
| | ۴۰% mid | **۰/۹۲۸** |
| | ۳۰% max | ۰/۶۰۶ |
| کوهپایه | ۳۰% min | ۰/۸۵۰ |
| | ۴۰% mid | **۰/۸۵۸** |
| | ۳۰% max | ۰/۸۴۰ |

جدول (۸): خطای نسبی ۳۰ درصد بیشینه مدل هیبرید-موجک برنامه‌ریزی بیان ژن

| ردیف | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۱۰ | ۱۱ | ۱۲ | ۱۳ | ۱۴ | ۱۵ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| کوهپایه | ۰/۰۹ | ۰/۰۵ | ۰/۱۴ | ۰/۰۴۲ | ۰/۰۳ | ۰/۱۱ | ۰/۰۳ | ۰/۰۷ | ۰/۰۱ | ۰/۰۷ | ۰/۰۲ | ۰/۰۵ | ۰/۰۸ | ۰/۰۵ | ۰/۰۲ |
| تازه‌کند | ۰/۱۴ | ۰/۶۱ | ۰/۳۲ | ۰/۲۶ | ۰/۳۳ | ۰/۳۶ | ۰/۴۹ | ۰/۲۶ | ۰/۳۲ | ۰/۳۳ | ۰/۳۶ | ۰/۳۲ | ۰/۳۶ | ۰/۴۰ | ۰/۳۴ |

## نتیجه‌گیری

در این تحقیق از روش‌های شبکه عصبی مصنوعی، سامانه استنتاجی فازی-عصبی تطبیقی و برنامه‌ریزی بیان ژن و ترکیب این روش‌ها با تئوری موجک برای شبیه‌سازی مقادیر ماهانه تبخیر استفاده شد. بدین منظور، متغیرهای هواشناسی بارش ماهانه، رطوبت نسبی، دمای میانگین، بیشینه و کمینه ماهانه و سرعت باد ماهانه جهت مدلسازی به کار برده شد. نتایج حاصل از این پژوهش نشان داد که در دو اقلیم مختلف (Dsa و Bsk) روش‌های هیبرید-موجک برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی منفرد به ترتیب دارای بالاترین و ضعیف‌ترین عملکرد در میان سایر مدل‌های داده‌کاوی هستند. همچنین، نتایج نشان داد که مدل‌ها در حوضه دریاچه ارومیه با اقلیم Dsa عملکرد بهتری نسبت به حوضه گاوخونی با قلیم Bsk را از خود نشان داده‌اند. با این حال، نتایج سه روش منفرد و هیبرید-موجک آنها می‌تواند برای برآورد متغیر تبخیر ماهانه در حوضه‌ها با انواع آب و هوایی Dsa و Bsk قابل قبول باشد. در نهایت می‌توان اظهار داشت که روش‌های داده‌کاوی هیبریدی را می‌توان به عنوان جایگزین مناسبی برای روش‌های قدیمی معرفی نمود.

## تقدیر و تشکر

این مقاله مستخرج از طرح پژوهشی دانشگاهی می‌باشد. نویسندگان این مقاله از دانشگاه علوم کشاورزی و منابع طبیعی ساری (SANRU) به دلیل حمایت مالی، کمال تشکر را دارند (کد طرح: ۱۵-۱۳۹۹-۰۲).

## منابع

صمدیان‌فر، س.، س. هاشمی و م. ایزدیار. ۱۳۹۷. برآورد تبخیر روزانه از تشت تبخیر با استفاده از روش‌های یادگیری ماشینی. نشریه آبیاری و زهکشی ایران. شماره ۴، جلد ۱۲، ص ۱۰۱۵-۱۰۰۴.

Adnan, R. M., A. Malik, A. Kumar, K. S. Parmar and O. Kisi. 2019. Pan evaporation modeling by three different neurofuzzy intelligent systems using climatic inputs. Arab J Geosci, 12: 606–614.

Aparajita, S., A. R. S. Kumar, R. M. Singh, V. K. Tripathi, A. Thakur and P. Kumar. 2019. Monthly evaporation modelling using ANN and WANN model. Environment and Ecology, 37(3A): 787-793.

Archive of SID.ir

Archive of SID.ir



Araghi, A., J. Adamowski and C. J. Martinez. 2020. Comparison of wavelet-based hybrid models for the estimation of daily reference evapotranspiration in different climates. Journal of Water and Climate Change, 11(1): 39-53.

Emadi, A., S. Zamanzad-Ghavidel, S. Fazeli, S. Zarei and A. Rashid-Niaghi. 2021. Multivariate modeling of pan evaporation in monthly temporal resolution using a hybrid evolutionary data-driven method (case study: Urmia Lake and Gavkhouni basins). Environ Monit Assess, 193(6):355.

Genanu, M., T. Alamirew, G. Senay and M. Gebremichael. 2017. Remote Sensing Based Estimation of Evapo-Transpiration Using Selected Algorithms: The Case of Wonji Shoa Sugar Cane Estate, Ethiopia. International Journal of Sensors and Sensor Networks, 5(1): 1-13.

Holland, J.H. 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. MIT Press, 1992. First Published by University of Michigan Press, USA.

Kucuk, M. and N. Agiralioglu. 2006. Wavelet regression techniques for stream flow predictions, J. Appl. Stat. 33(9), 943–960.

Lopes, HS., and W.R. Weinert. 2004. EGIPSYS: An enhanced gene expression programming approach for symbolic regression problems. Applied Mathematics and Computer Science 14(3): 375-384.

Malik, A., and A. Kumar. 2015. Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. Water resources management, 29(6): 1859-1872.

Malik, A., A. Kumar, S. Kim, M. H. Kashani, V. Karimi, A. Sharafati and K. W. Chau. 2020. Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. Engineering Applications of Computational Fluid Mechanics, 14(1): 323-338.

Montaseri, M., S. Zamanzad-Ghavidel and H. Sanikhani. 2018. Water quality variations in different climates of Iran: toward modeling total dissolved solid using soft computing techniques. Stochastic environmental research and risk assessment, 32(8): 2253-2273.

Patle, G. T., M. Chettri and D. Jhajharia. 2019. Monthly pan evaporation modelling using multiple linear regression and artificial neural network techniques. Water Supply.

Popoola, A.O. 2007. Fuzzy-Wavelet Method for Time Series Analysis. University of Surrey. Department of Computing School of Electronics and Physical Sciences University of Surrey Guildford, Surrey GU2 7XH, UK.

Shiri, J., Ö. Kişi, G. Landeras, J. J. López, A. H. Nazemi and L. C. Stuyt. 2012. Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). Journal of Hydrology, 414: 302-316.

Shirsath, P. B., and A. K. Singh. 2010. A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. Water resources management, 24(8): 1571-1581.

Singh, A., R. M. A. R. Singh, A. Kumar, S. Kumar, Hanwat and V. K. Tripathi. 2019. Evaluation of soft computing and regression-based techniques for the estimation of evaporation. Journal of Water and Climate Change, 12(1): 32-43.

Tabari, H., S. Marofi and A.A. Savziparvar. 2010. Estimation of daily pan evaporation using artificial neural networks and multivariate non-linear regression. Irrigation Science, 28: 399-406.

Tamta, S., P. S. Kashyap and P. Kumar. 2018. Estimation of Evaporation in Hilly Area by Using Ann and Canfis System Based Models. Int. J. Curr. Microbiol. App. Sci, 7(1), 911-919.

Traore, S., and A. Guven. 2012. Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. Water resources management, 26(15): 4367-4380.

Archive of SID.ir

Archive of SID.ir



Zareie, S., and A. Fakherifard. 2014. Assessment of temporal relationship between meteorological and hydrological droughts by GP and ANFIS models in Sofi-Chi basin. Water Resources Engineering Journal, 7(21): 37-50

.

Archive of SID.ir

Archive of SID.ir



# Modeling Of Monthly Evaporation Using Single and Hybrid-Wavelet Data-Driven Methods in Basins of Iran with Climate Variety

**Alireza Emadi[1]*; Sarvin Zamanzad-Ghavidel[2]; Sina Fazeli[3]; Soheila Zareie[3]; and Ali Rashid-Niaghi[4]**

## Abstract

Evaporation as one of the natural parameters has always been considered by researchers. In this study, the monthly evaporation variable was modeled in two different climates of Iran using artificial neural network, adaptive fuzzy-neural inference system and gene expression programming methods and combining these methods with wavelet theory. For this purpose, meteorological data of precipitation, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed were used during the statistical period of 1384-1397 related to the two catchments of Urmia Lake and Gavkhouni. In this study, the seasonal effect and data noise reduction were applied. The accuracy of the studied methods was evaluated based on statistical correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE) and nash-sutcliffe efficiency (NSE). The results show that in two different climates, the wavelet-hybrid gene expression programming and the single artificial neural network have the highest and weakest performance, respectively, among other data mining models used in this study. The hybrid wavelet-gene expression programming model with RMSE value of 20.870 and 156.884 had higher performance for Tazehkand station in Urmia Lake catchment area and Kuhpayeh catchment in Gavkhouni catchment area, respectively. Also, the results showed that the effect of seasonal factor utilization and data noise reduction in model performance improvement is significant. Based on the results of the models performance Urmia Lake catchment area with Dsa climate has been better. However, hybrid data mining methods can be introduced as a good alternative to the old methods.

**Keywords: Climate, Evaporation, Data Mining, Wavelet**

[1] *Associate professor, Department of Water Engineering, Faculty of Agricultural Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. 09123931311, emadia355@yahoo.com, a.emadi@sanru.ac.ir.
[2] Project manager and member of the director board of Daneshvaran Omran Ab Consulting Engineering Company, Urmia, Iran. 09141066409, snzghavidel@ut.ac.ir.
[3] Ph.D student, Department of Irrigation & Reclamation Engineering, Faculty of Agriculture Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Tehran, Iran. 09183309215, 09188718826, s.fazeli@ut.ac.ir, soheila_zareie@yahoo.com.
[4] Postdoctoral Researcher, University of Minnesota, Minnesota, United States. niaghi@umn.edu.

Archive of SID.ir

Archive of SID.ir



Research Paper

# Modeling Of Monthly Evaporation Using Single and Hybrid-Wavelet Data-Driven Methods in Basins of Iran with Climate Variety

Alireza Emadi [1*], Sarvin Zamanzad-Ghavidel [2], Sina Fazeli [3], Soheila Zareie [3], Ali Rashid-Niaghi [4]

[1] *Associate professor, Department of Water Engineering, Faculty of Agricultural Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. 09123931311, emadia355@yahoo.com, a.emadi@sanru.ac.ir.
[2] Project manager and member of the director board of Daneshvaran Omran Ab Consulting Engineering Company, Urmia, Iran. 09141066409, snzghavidel@ut.ac.ir.
[3] Ph.D student, Department of Irrigation & Reclamation Engineering, Faculty of Agriculture Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Tehran, Iran. 09183309215, 09188718826, s.fazeli@ut.ac.ir, soheila_zareie@yahoo.com.
[4] Postdoctoral Researcher, University of Minnesota, Minnesota, United States. niaghi@umn.edu.

10.22125/IWE.2021.243211.1401.

Received:
**August 10, 2020**
Accepted:
**January 2, 2021**
Available online:
**June.01.2022**

**Keywords:**
**Climate, Evaporation, Data Mining, Wavelet**

## Abstract

Evaporation as one of the natural parameters has always been considered by researchers. In this study, the monthly evaporation variable was modeled in two different climates of Iran using artificial neural network, adaptive fuzzy-neural inference system and gene expression programming methods and combining these methods with wavelet theory. For this purpose, meteorological data of precipitation, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed were used during the statistical period of 1384-1397 related to the two catchments of Urmia Lake and Gavkhouni. In this study, the seasonal effect and data noise reduction were applied. The accuracy of the studied methods was evaluated based on statistical correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE) and nash-sutcliffe efficiency (NSE). The results show that in two different climates, the wavelet-hybrid gene expression programming and the single artificial neural network have the highest and weakest performance, respectively, among other data mining models used in this study. The hybrid wavelet-gene expression programming model with RMSE value of 20.870 and 156.884 had higher performance for Tazehkand station in Urmia Lake catchment area and Kuhpayeh catchment in Gavkhouni catchment area, respectively. Also, the results showed that the effect of seasonal factor utilization and data noise reduction in model performance improvement is significant. Based on the results of the models performance Urmia Lake catchment area with Dsa climate has been better. However, hybrid data

* **Corresponding Author: Alireza Emadi**
**Address**: Associate professor, Department of Water Engineering, Faculty of Agricultural Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.
**Email**: emadia355@yahoo.com , a.emadi@sanru.ac.ir
**Tel:** 09123931311



Archive of SID.ir

Archive of SID.ir



mining methods can be introduced as a good alternative to the old methods.

## 1. Introduction

Evaporation is a key component of the climate change process that its accurate estimation can improve the management of water resources and the monitoring network. Many climatic parameters such as air temperature, precipitation, relative humidity, sunlight, and wind speed affect the rate of evaporation. There are direct and indirect methods for calculating and predicting evaporation (Shiri et al., 2012). The multiplicity of required data and the difficulty in estimating meteorological variables discourage researchers from using indirect methods.

## 2. Materials and Methods

In this study, meteorological data of precipitation, relative humidity, mean temperature, minimum temperature, maximum temperature, and wind speed, during the statistical period of 2005-2018 in Tazehkand and Kuhpayeh stations, located in Urmia Lake and Gavkhoni catchments with the area of 876.51 and 552.41 square kilometers were used for the assessment of monthly Evp. According to the Köppen-Geiger climate classification, Urmia Lake and Gavkhoni catchments have Dsa and Bsk climate. Both catchments have a great impact on industrial, agricultural, and livestock activities in northwestern and central of Iran.

To analyze the suitability of the selected catchments, the distribution of stations was evaluated using F-computational values at the 99% probability level. Pearson correlation between data (precipitation, temperature, humidity with two time scales, wind speed, and pan evaporation) was calculated using SPSS software at a significance level of 99%. The monthly evaporation variable was modeled in two different climates of Iran using artificial neural network, adaptive fuzzy-neural inference system and gene expression programming methods and combining them with wavelet theory. In an artificial neural network using mathematical functions, the outputs are generated as a set of actuators (transfer functions). Adaptive fuzzy-neural inference system is a multilayer transmission network with input-output structure that uses neural network learning algorithms and fuzzy logic to transfer inputs to output. Gene expression programming, using a genetic algorithm, selects a population of individuals and applies genetic modifications using one or more genetic agents. Wavelet transform is also a mathematical tool that uses different types of data to extract information (Aparajita et al., 2019). Due to the high process similarity of the DWT (db4) tool with the data set of the present study, the db4 approach was chosen.

## 3. Results

In this study, the seasonal effect and data de-noising were applied and the wavelet was combined with three methods of ANN, GEP, and ANFIS and their results were compared. The accuracy of the studied methods was evaluated based on the statistical indices of correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), and Nash-Sutcliffe efficiency coefficient (NSE). Mathematical relation is applied to use modeling results at different time and space scales. The hybrid model of wavelet- gene expression programming with RMSE values equal to 20.870 and 156.884 mm for Tazehkand and Kuhpayeh stations in the Urmia Lake and Gavkhooni catchments had a higher performance, respectively. Also, the results showed that the effect of using seasonal coefficient and noise reduction of data is significant on improving the performance of models. In this study, WGEP model was selected as the best model in both climates. The estimation results of the evaporation variable by the mentioned model in both stations were divided into three categories of 30% minimum, 40% medium, and 30% maximum.

## 4. Discussion and Conclusion

Evaporation is a very important parameter in agricultural research, hydrology and soil and water protection. The results of this study showed that in two different climates (Dsa and Bsk) hybrid-wavelet methods of gene expression programming and single artificial neural network have the highest and



Archive of SID.ir

Archive of SID.ir



weakest performance among other data mining models, respectively. The results also showed that the models in the Urmia Lake basin with Dsa climate showed better performance than the Gavkhoni basin with Bsk climate. However, the results of their three single and hybrid-wavelet methods can be acceptable for estimating the monthly evaporation variable in catchments with different Dsa and Bsk climates. Finally, it can be stated that hybrid data mining methods can be introduced as a suitable alternative to older methods.

## 5. Six important references


1. Adnan, R. M., A. Malik, A. Kumar, K. S. Parmar and O. Kisi. 2019. Pan evaporation modeling by three different neurofuzzy intelligent systems using climatic inputs. Arab J Geosci, 12: 606–614.
2. Aparajita, S., A. R. S. Kumar, R. M. Singh, V. K. Tripathi, T. Aradhana and K. Pushpendra. 2019. Monthly evaporation modelling using ANN and WANN model. Environment and Ecology, 37(3A): 787-793.
3. Araghi, A., J. Adamowski and C. J. Martinez. 2020. Comparison of wavelet-based hybrid models for the estimation of daily reference evapotranspiration in different climates. Journal of Water and Climate Change, 11(1): 39-53.
4. Tamta, S., P. S. Kashyap and P. Kumar. 2018. Estimation of Evaporation in Hilly Area by Using Ann and Canfis System Based Models. Int. J. Curr. Microbiol. App. Sci, 7(1): 911-919.
5. Malik, A., and A. Kumar. 2015. Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. Water resources management, 29(6): 1859-1872.
6. Malik, A., A. Kumar, S. Kim, M. H. Kashani, V. Karimi, A. Sharafati and K. W. Chau. 2020. Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. Engineering Applications of Computational Fluid Mechanics, 14(1): 323-338.


## Conflict of Interest

Authors declared no conflict of interest.


## Acknowledgments

We are grateful to the Sari Agricultural Science and Natural Resources University (SANRU) financed this research (Grant No. 02–1399-15).




Archive of SID.ir