

استفاده از روش یادگیری ماشین به منظور برآورد تبخیر و تعرق (مطالعه موردی: استان سمنان)

چکیده

تبخیر و تعرق (ET) در مناطق خشک و نیمه‌خشک از اهمیت زیادی برخوردار بوده و برآورد دقیق آن در برنامه‌ریزی و مدیریت شیوه‌های آبیاری حائز اهمیت است. استان سمنان دارای تعداد محدودی ایستگاه سینوپتیک و تبخیرسنجی است که کمبود داده‌های تبخیر و تعرق سبب شده که برآورد مکانی آن با مشکل مواجه گردد. در این مطالعه از محصول تبخیر و تعرق حاصل از داده‌های بازتحلیل ERA5-Land به همراه متغیرهای کمکی ارتفاع و دما برای برآورد تبخیر و تعرق در منطقه مورد مطالعه استفاده شد. همچنین به منظور برقراری ارتباط بین متغیرهای کمکی و داده‌های تبخیر و تعرق از روش یادگیری ماشین جنگل تصادفی (RF) استفاده شد و نقشه تبخیر و تعرق در منطقه مورد مطالعه با استفاده از مدل RF تهیه شد. دقت مدل RF در برآورد تبخیر و تعرق با استفاده از چهار معیار آماری شامل ضریب همبستگی (r)، مقدار اریبی (BIAS)، میانگین ریشه مربعات خطا (RMSE) و شاخص KGE مورد ارزیابی قرار گرفت. نتایج مرحله اعتبارسنجی، کارایی بالای مدل RF را نشان داد ($R^2 = 0.95$ ، $BIAS = -4.1$ ، $RMSE = 98.6$ و $KGE = 0.92$). همچنین مشخص شد عملکرد مدل RF در برآورد تبخیر و تعرق با استفاده از متغیرهای ورودی به وابستگی خطای مدل (BIAS) به توپوگرافی بستگی دارد و متغیر ارتفاع عامل مهمی در برآورد تبخیر و تعرق محسوب می‌گردد. نتایج این مطالعه نشان داد که استفاده از داده‌کاوی و پردازش در محیط برنامه‌نویسی R، در مناطقی با تعداد محدود ایستگاه هواشناسی، برآورد دقیق میزان تبخیر و تعرق در مناطق خشک و نیمه‌خشک را ممکن می‌سازد.

واژه‌های کلیدی: ارتفاع، دما، ERA5-Land مدل جنگل تصادفی (RF)

Using Machine Learning Method to Estimate Evapotranspiration (Case Study: Semnan Province)

Abstract

Evapotranspiration (ET) plays a crucial role in arid and semi-arid regions, and its precise estimation of ET is essential for effective irrigation planning and management. Semnan province faces challenges due to a scarcity of synoptic and evaporation stations, making spatial estimation of ET difficult. This study utilized the evapotranspiration product from the ERA5-Land reanalysis dataset, in conjunction with auxiliary variables such as elevation and temperature, to estimate ET in the study area. The Random Forest (RF) model was employed to establish the relationship between auxiliary variables and ET data, resulting in the creation of an ET map using the RF model. The accuracy of the RF model in estimating ET was assessed against observational data using four statistical criteria: R^2 , BIAS, RMSE, and KGE. The validation results demonstrated the high efficiency of the RF model ($R^2 = 0.95$, $BIAS = -4.1$, $RMSE = 98.6$, and $KGE = 0.92$). It was observed that the RF model's performance in ET estimation is influenced by the relationship between model error (BIAS) and topography, with elevation playing a significant role in ET estimation accuracy. This study underscores the effectiveness of utilizing data mining and processing techniques within the R programming environment to accurately estimate ET in regions with limited weather stations, particularly in arid and semi-arid areas. By leveraging these methods, it becomes possible to enhance the estimation of evapotranspiration in such challenging environments.

Keywords: Elevation, ERA5-Land, Random Forest Model (RF), Temperature

تبخیر و تعرق (ET) فرایند خروج آب از سطح خاک و گیاه است که تحت تأثیر دما، رطوبت، تابش خورشیدی و سرعت باد است. تبخیر و تعرق یکی از مهم‌ترین مؤلفه‌ها در چرخه هیدرولوژیکی و مدیریت آب در کشاورزی، برآورد آب موردنیاز و بررسی تأثیرات تغییر اقلیم است. همچنین این متغیر در بیلان آبی حوضه، و مدیریت منابع آب و نیز در فعالیت‌هایی مانند مدیریت ریسک سیلاب یا خشکسالی از اهمیت بالایی برخوردار است (Kisi, 2016; Krajewski et al., 2017). در این زمینه تعداد محدود ایستگاه‌های اندازه‌گیری تبخیر و تعرق سبب می‌شود که پهنه‌بندی این متغیر به‌خصوص در مناطقی با مساحت زیاد با مشکلات زیادی همراه باشد. از طرف دیگر کمبود ایستگاه‌ها در برخی مناطق موجب می‌شود که استفاده از معادلات تجربی سبب افزایش خطا در برآورد تبخیر و تعرق در این مناطق شود. برای حل این مشکل، استفاده از متغیرهای کمکی از قبیل داده‌های باز تحلیل شده که متغیرهای اقلیمی از جمله تبخیر و تعرق را به‌صورت شبکه‌ای در دسترس قرار می‌دهد، به‌طور گسترده‌ای مورد توجه قرار گرفته است (Crossett et al., 2020). مرکز اروپایی پیش‌بینی میان‌مدت وضع هوا^۲ که در چارچوب خدمات تغییر اقلیم کوپرنیک کمیسیون اروپا قرار دارد، در حال تولید مجموعه داده‌های جهانی پیشرفته شده برای بخش زمین از نسل پنجم تحلیل مجدد اروپایی^۳ است. مزیت اصلی استفاده از داده‌های بازتحلیل شده ERA5-Land وضوح افقی بالای داده‌ها است که در سطح جهانی به ۹ کیلومتر می‌رسد. داده‌های بازتحلیل شده در بسیاری از مطالعات بدون اصلاح اریبی مورد استفاده قرار می‌گیرند؛ اما نتایج بسیاری از مطالعات نشان داده که این داده‌ها در کشور با خطا همراه بود و استفاده مستقیم از داده‌ها ممکن است که سبب خطا در برآورد خصوصیات اقلیمی مورد بررسی شوند.

استفاده از متغیرهای کمکی به‌منظور پهنه‌بندی خصوصیات اقلیمی معمولاً با کمک روش‌های یادگیری ماشین انجام می‌گیرد. به‌عنوان مثال امینی و همکاران (۱۴۰۱) از متغیرهای کمکی ارتفاع، دمای سطح زمین، شاخص پوشش سطح و داده‌های شبکه‌ای بارش و مدل RF برای برآورد و پهنه‌بندی بارش استان سمنان استفاده کردند. از طرف دیگر پارامتر اقلیمی تبخیر و تعرق دارای پیچیدگی‌های زیادی بوده و به متغیرهای اقلیمی بسیاری وابسته می‌باشد. این عامل سبب می‌شود که محققین برای برآورد و پهنه‌بندی تبخیر و تعرق از روش‌های یادگیری ماشین کمک بگیرند (Torres et al., 2011). از این میان، جنگل تصادفی (RF) یکی از تکنیک‌های یادگیری ماشینی جدید و یک ابزار مدل‌سازی غیرخطی است که به‌طور گسترده‌ای برای پیش‌بینی میزان تبخیر و تعرق در مناطق مختلف آب‌وهوایی استفاده می‌شود (Chen et al., 2020; Wang et al., 2021). این مدل که به دلیل سادگی و توانایی انجام وظایف طبقه‌بندی و رگرسیون شناخته شده است، امکان استفاده از چندین متغیر کمکی برای برآورد تبخیر و تعرق را فراهم می‌کند (Hengel et al., 2018). در این زمینه توپوگرافی به‌عنوان یک متغیر کمکی، عامل مهمی بر رطوبت خاک و توزیع مجدد آب اشیاع شده^۴ در نظر گرفته می‌شود. این موضوع اهمیت مدل‌سازی فرایندهای هیدرولوژیکی تحت تأثیر توپوگرافی در تهیه نقشه دقیق تبخیر و تعرق را نشان می‌دهد (Chen et al., 2005).

پیشینه پژوهش

در پژوهشی، Feng et al. (2017) از الگوریتم RF برای پیش‌بینی تبخیر و تعرق روزانه در جنوب غربی چین استفاده کردند و نتایج حاصل از آن را با روش شبکه عصبی مبتنی بر رگرسیون عمومی (GRNN) مورد مقایسه قرار دادند. نتایج نشان داد که هر دو روش یادگیری نتایج قابل‌قبولی در مورد تبخیر و تعرق روزانه ارائه می‌کنند؛ اگرچه نتایج حاصل از مدل RF به نسبت GRNN عملکرد بهتری

1. Evapotranspiration

2. ECMWF

3. ERA5-Land

4. Saturated water redistribution

داشت. (Wu et al. (2020) کاربرد مدل RF برای شبیه‌سازی تبخیر و تعرق مرجع روزانه در دو ایستگاه واقع در ناحیه خشک و نیمه‌خشک شمال غرب چین را مورد ارزیابی قرار دادند. نتایج نشان داد مدل RF عملکرد خوبی در پیش‌بینی تبخیر و تعرق برای مناطق نیمه‌خشک با داده‌های محدود دارد. (Abed et al. (2022) از روش یادگیری ماشین RF و دو تکنیک یادگیری عمیق، یعنی شبکه عصبی پیچشی (CNN) و شبکه عصبی عمیق (DNN) برای برآورد دقیق میزان تبخیر و تعرق ماهیانه استفاده کردند. آن‌ها داده‌های حاصل از چهار ایستگاه هواشناسی را در مالزی طی بازه زمانی ۲۰۱۹-۲۰۲۰ برای ارزیابی مدل‌ها مورد استفاده قرار دادند و نتایج نشان داد سه مدل ML توسعه‌یافته در این مطالعه بهتر از مدل‌های تجربی عمل کرده و می‌توانند به طور قابل توجهی باعث افزایش دقت برآورد تبخیر و تعرق ماهانه شوند. (Rashid Niaghi et al. (2021) عملکرد مدل‌های ML شامل برنامه‌ریزی بیان ژن (GEP)، ماشین بردار پشتیبان (SVM)، رگرسیون خطی چندگانه (LR) و جنگل تصادفی (RF) را در برآورد تبخیر و تعرق پتانسیل با استفاده از چهار شاخص آماری با یکدیگر مقایسه کردند. نتایج این محققین نشان داد که مدل RF دارای کمترین مقادیر شاخص پراکندگی (SI) و میانگین ریشه مربعات خطا (RMSE) بوده و بهترین عملکرد را به صورت مکانی یا محلی دارد. (Shiri, (2018) مدل RF جفت شده را با الگوریتم موچک برای برآورد تبخیر و تعرق در بخش جنوبی ایران مورد ارزیابی قرار داد. نتایج حاکی از آن بود که مدل RF جفت شده در مقایسه با RF معمولی از کارایی بالاتری برخوردار است. (Hameed et al. (2021) در تحقیقی کاربرد روش‌های یادگیری ماشین را در پیش‌بینی تبخیر و تعرق در ترکیه مورد بررسی قرار دادند. برای این منظور، از بیست ترکیب ورودی شامل پارامترهای هیدرولوژیکی و جغرافیایی و سه مدل شامل رگرسیون خطی چندگانه (MLR)، جنگل تصادفی (RF) و ماشین یادگیری افراطی (ELM) استفاده شد. نتایج این مطالعه نشان داد که مدل ELM بر اساس معادله پنمن-مانتیت در مقایسه با سایر مدل‌ها مانند MLR و RF نتایج دقیق‌تری ارائه می‌دهد. (Ahmadi et al. (2021) از روش‌های یادگیری ماشین برای برآورد تبخیر و تعرق مرجع ماهانه شش ایستگاه در ایران استفاده کردند. آن‌ها از برنامه‌ریزی بیان ژن (GEP) و رگرسیون بردار پشتیبان (SVR) به عنوان مدل‌های مستقل استفاده کردند. سپس یک مدل ترکیبی جدید از طریق جفت کردن SVR کلاسیک با یک الگوریتم بهینه‌سازی (SVR-IWD) معرفی شد. نتایج نشان داد که مدل هیبریدی پیشنهادی SVR-IWD عملکرد بهتری نسبت به روش SVR در برآورد تبخیر و تعرق ماهانه دارد.

تاکنون تحقیقات متعددی در خصوص برآورد تبخیر و تعرق در سطح جهان و نیز کشور ایران صورت گرفته که از متغیرهای اقلیمی برای برآورد تبخیر و تعرق در ایستگاه استفاده شده است. اما مطالعات کمی از متغیرهای کمکی که به راحتی در اختیار می‌باشند برای پهنه‌بندی و مدل‌سازی تبخیر و تعرق استفاده کرده‌اند. استان سمنان با وسعتی گسترده دارای تعداد محدودی ایستگاه سینوپتیک و تبخیرسنجی است که پراکنش مناسبی نداشته است؛ این عامل سبب می‌شود که در مناطق عمده کشاورزی، برآورد تبخیر و تعرق با خطای زیادی همراه باشد. استفاده از مدل‌های یادگیری ماشین و استفاده از متغیرهای کمکی که به راحتی در اختیار هستند، تا حدودی مشکل در برآورد تبخیر و تعرق را در استان برطرف می‌کند؛ لذا هدف اصلی این پژوهش تهیه نقشه دقیق تبخیر و تعرق سالانه در استان سمنان با استفاده از مدل RF می‌باشد. همچنین بررسی ارتباط بین متغیرهای کمکی و داده‌های تبخیر و تعرق از دیگر اهداف این مطالعه می‌باشد.

روش‌شناسی پژوهش

منطقه مورد مطالعه

استان سمنان دارای مساحتی برابر با ۹۶۸۱۶ کیلومتر مربع است که حدود ۵/۹ درصد از مساحت کل کشور را شامل می‌شود. مختصات جغرافیایی استان از ۱۵° ۳۴' تا ۲۰° ۳۷' عرض شمالی و از ۵۰° ۵۱' تا ۵۷° ۰۳' طول شرقی متغیر است. ارتفاع استان از شمال به جنوب کاهش می‌یابد؛ به گونه‌ای که اختلاف ارتفاع تا بیش از ۳۰۰۰ متر قابل مشاهده است. متوسط ارتفاع از سطح دریا در این استان

برابر ۱۰۶۷ متر و گستره جغرافیایی آن را می‌توان به سه قسمت کوهستانی، کوهپایه‌ای و نواحی پست دشت کویر تقسیم کرد. اقلیم غالب استان گرم و خشک است و میزان متوسط بارندگی سالانه آن، ۱۳۶ میلی‌متر است. متوسط درجه حرارت سالانه، حداکثر مطلق و حداقل مطلق به ترتیب ۱۷/۰۱، ۴۳/۵ و ۸/۴- درجه سانتی‌گراد در بازه زمانی ۲۰ ساله است.

در پژوهش حاضر، داده‌های تبخیر و تعرق اندازه‌گیری شده (مشاهداتی) از ۶ ایستگاه سینوپتیک اداره کل هواشناسی و ۵ ایستگاه تبخیر سنجی شرکت آب منطقه‌ای استان سمنان طی سال‌های ۱۹۹۳ تا ۲۰۲۰ اخذ شد (شکل ۱). با توجه به تعداد اندک ایستگاه‌های اندازه‌گیری تبخیر و تعرق در استان سمنان، از متغیرهای کمکی دما، ارتفاع و محصول تبخیر و تعرق حاصل از داده‌های ERA5-Land برای برآورد تبخیر و تعرق با استفاده از مدل RF استفاده شد. در ادامه هر یک از متغیرهای کمکی مورد استفاده با جزئیات مورد ارزیابی قرار گرفته است.

محصول تبخیر و تعرق حاصل از داده‌های بازتحلیل ERA5_land

در این مطالعه در ابتدا داده‌های تبخیر و تعرق ERA5-Land از سال ۱۹۹۳ تا ۲۰۲۰ با دقت مکانی ۹ کیلومتر در مقیاس زمانی ماهانه در سامانه گوگل ارث انجین مورد ارزیابی و مجموع داده‌های ماهانه هر سال به‌عنوان داده‌های تبخیر و تعرق سالانه محاسبه و دانلود شد. سپس داده‌ها با استفاده از روش نمونه‌گیری مجدد^۱ به داده‌های با قدرت تفکیک مکانی ۵ کیلومتر تبدیل شدند. استفاده از این داده‌ها، شبکه‌ای از داده‌های تبخیر و تعرق را در سطح استان به دست می‌دهد که نبود یا کمبود داده اندازه‌گیری شده در نقاط فاقد ایستگاه را جبران می‌کند. از آنجاکه داده‌های بازتحلیل به‌صورت شبکه‌ای بوده و داده‌های ایستگاهی به‌صورت نقطه‌ای است؛ مقادیر شبکه‌ای محصول تبخیر و تعرق داده‌های بازتحلیل ERA5-Land در محل مختصات ایستگاه‌ها استخراج و بدین ترتیب در مختصات هر ایستگاه، داده بازتحلیل تبخیر و تعرق برای ورود به مدل یادگیری ماشین (RF) به دست آمد.

داده‌های دمای اندازه‌گیری شده و بازتحلیل شده حاصل از ERA5-Land

متغیرهای کمکی مورد استفاده در این مطالعه برای پهنه‌بندی تبخیر و تعرق، شامل داده‌های ERA5-Land، ارتفاع و دما می‌باشند که داده‌های ارتفاع و تبخیر و تعرق ERA5-land به‌صورت شبکه‌ای در اختیار می‌باشند. از طرف دیگر وجود داده‌های دما اهمیت بسیار زیادی در برآورد تبخیر و تعرق داشته که خطا در برآورد دما سبب خطای زیادی در برآورد تبخیر و تعرق خواهد شد؛ لذا در مطالعه حاضر از داده‌های دمای ERA5-Land استفاده شد، اما در ابتدا این داده‌ها اصلاح اریبی شد و سپس داده‌های دمای اصلاح شده به‌عنوان متغیر کمکی دما در برآورد تبخیر و تعرق استفاده شدند. به‌منظور اصلاح اریبی داده‌های دما و تهیه نقشه دمای استان در سال‌های مختلف (۱۹۹۳-۲۰۲۰) از دو سری داده شامل داده‌های اندازه‌گیری شده و نیز بازتحلیل شده به شرح زیر استفاده شد:

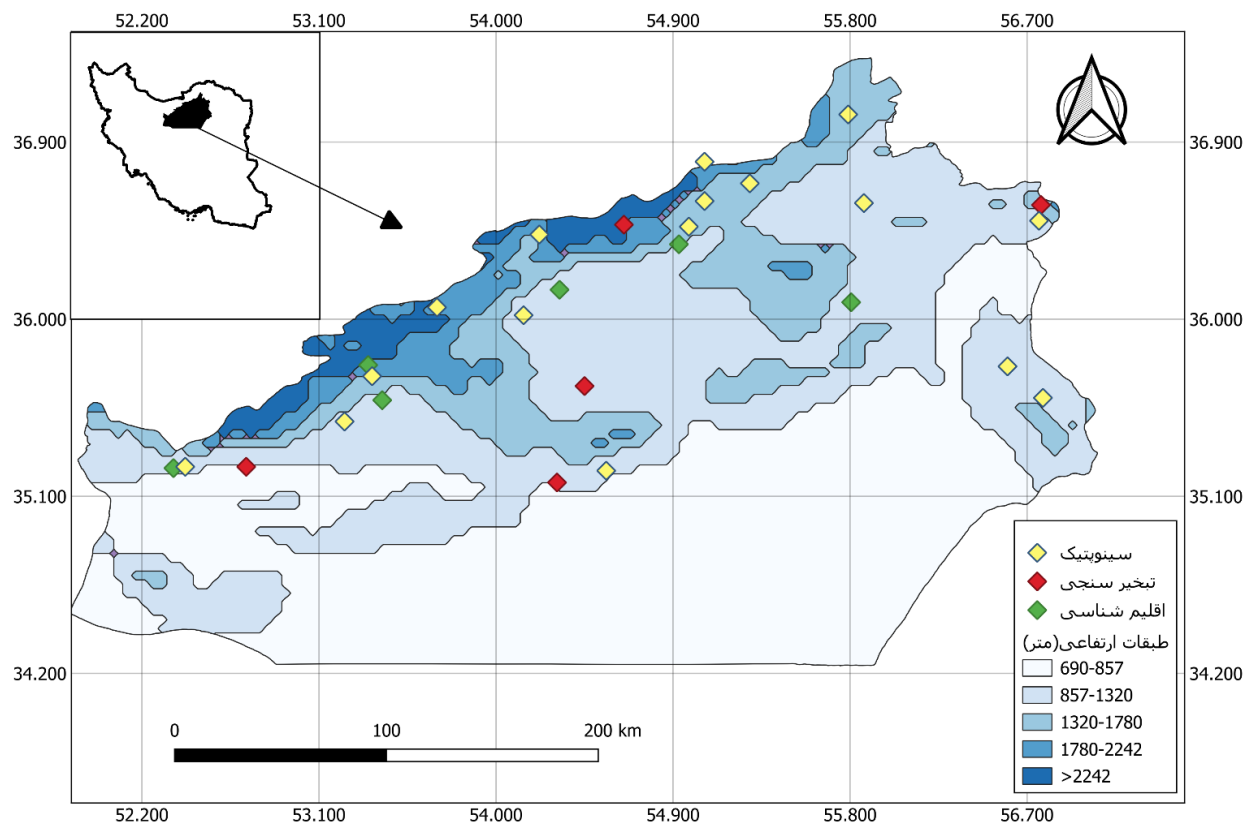
در ابتدا، داده‌های دمای سالانه ۹ ایستگاه سینوپتیک و ۱۶ ایستگاه اقلیم‌شناسی در دوره آماری (۱۹۹۳-۲۰۲۰) اخذ شد. به دلیل عدم پراکنش مناسب ایستگاه‌های زمینی در بخش‌هایی از استان و نیز کمبود و یا نبود داده اندازه‌گیری شده دما در برخی سال‌های موردنظر، از داده‌های بازتحلیل شده دمای ERA5-Land نیز استفاده شد. بدین منظور، داده‌های دمای بازتحلیل شده سالانه ERA5-Land طی سال‌های ۱۹۹۳ تا ۲۰۲۰ در سامانه گوگل ارث ارزیابی و سپس با قدرت تفکیک مکانی ۹ کیلومتر دانلود شدند. در ابتدا برای یکسان‌سازی این داده‌ها با داده‌های تبخیر و تعرق ERA5-Land، به داده‌های با قدرت تفکیک مکانی ۵ کیلومتر تبدیل گردید. در گام بعد اصلاح اریبی داده‌های دمای سالانه حاصل از ERA5-Land به روش نگاشت چندک (QM) در محیط نرم‌افزاری R توسط بسته

1. Resampling
2. Quantile Mapping(QM)

qmap انجام شد. نگاشت چندک یک روش ناپارامتری بوده و بدون فرضی اولیه برای تمامی توزیع‌های آماری قابل استفاده است. در این گام، مقادیر داده‌های دمای باز تحلیل شده بر اساس داده‌های اندازه‌گیری شده ایستگاهی به روش نگاشت چندک اصلاح آریبی شد.

داده‌های کمکی ارتفاع

باتوجه به اینکه مقادیر تبخیر و تعرق متأثر از تغییرات دما و ارتفاع در استان است، در این مطالعه از ارتفاع نیز به‌عنوان یک متغیر کمکی در برآورد میزان تبخیر و تعرق استفاده شد. بدین منظور، از نقشه مدل رقومی ارتفاع (DEM) حاصل از داده‌های SRTM با قدرت تفکیک مکانی ۹۰ متر استفاده شد. برای تعیین ارتفاع ایستگاه‌ها (نقاط دارای داده اندازه‌گیری) و نقاط فاقد داده‌های اندازه‌گیری تبخیر و تعرق و برای یکسان کردن قدرت تفکیک مکانی متغیرهای کمکی مورداستفاده، مدل رقومی ارتفاع با اندازه سلول برابر با ۹۰ متر با استفاده از روش نمونه‌گیری مجدد به نقشه با قدرت تفکیک پنج کیلومتر تغییر یافت. موقعیت مکانی ایستگاه‌های هواشناسی به همراه الگوی توزیع مکانی طبقات ارتفاعی در شکل ۱ نشان داده شده است.



شکل ۱. موقعیت ایستگاه‌های مورد بررسی و طبقات ارتفاعی در استان سمنان

مدل سازی مقادیر تبخیر و تعرق با استفاده از مدل جنگل تصادفی (RF)

الگوریتم جنگل تصادفی (RF) یکی از روش‌های یادگیری ماشین است که به دلیل ساختار ساده و دقت بالا، به طور گسترده‌ای در زمینه علوم زمین مورد استفاده قرار گرفته است (Guo et al., 2021). در این روش نیازی به نرمال‌سازی داده‌های ورودی نیست (Li et al., 2022). برای برآورد تبخیر و تعرق و تهیه نقشه تبخیر و تعرق استان، متغیرهای کمکی شامل داده‌های دمای سالانه بازتحلیل و اصلاح اریبی شده، متغیر ارتفاع و نیز داده‌های سالانه بازتحلیل شده تبخیر و تعرق که همبستگی معنی‌داری با داده‌های تبخیر و تعرق ایستگاه‌ها داشتند انتخاب شده و به‌عنوان متغیرهای کمکی وارد فاز مدل‌سازی به روش جنگل تصادفی شدند. به‌عبارت‌دیگر، از متغیرهای محیطی مذکور استفاده شد تا مدل RF در نقاط فاقد داده، مقادیر تبخیر و تعرق سالانه در استان را برآورد نماید. مدل RF در محیط برنامه‌نویسی R و با استفاده از بسته caret اجرا و نقشه تبخیر و تعرق استان در دوره آماری (۱۹۹۳-۲۰۲۰) تهیه و خروجی گرفته شد.

شاخص‌های ارزیابی عملکرد مدل

برای ارزیابی عملکرد روش RF در برآورد تبخیر و تعرق، از شاخص ضریب همبستگی (r)، BIAS، RMSE و KGE استفاده گردید. این چهار معیار به صورت زیر محاسبه می‌شوند:

$$r = \frac{\sum_{i=1}^n (ET_{pi} - \overline{ET_{pi}})(ET_{oi} - \overline{ET_{oi}})}{\sqrt{\sum_{i=1}^n (ET_{pi} - \overline{ET_{pi}})^2} \sqrt{\sum_{i=1}^n (ET_{oi} - \overline{ET_{oi}})^2}} \quad \text{رابطه (۱)}$$

$$BIAS = \frac{1}{n} \sum_{i=1}^n ET_{pi} - ET_{oi} \quad \text{رابطه (۲)}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (ET_{pi} - ET_{oi})^2} \quad \text{رابطه (۳)}$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad \text{رابطه (۴)}$$

جایی که در رابطه (۱) ضریب همبستگی پیرسون بین داده‌های برآورد شده و مشاهداتی است. BIAS انحراف سیستماتیک از مقدار واقعی بر حسب میلی‌متر (رابطه ۲)، RMSE میانگین ریشه مربعات خطا بر حسب میلی‌متر (رابطه ۳) و KGE معیار ارزیابی ارائه شده توسط Gupta et al. (2009) برای مقایسه متغیرهای برآورد شده و مشاهداتی است (رابطه ۴). ET_{pi} مقادیر برآورد شده تبخیر و تعرق سالانه بر حسب میلی‌متر و ET_{oi} مجموع تبخیر و تعرق مشاهداتی سالانه بر حسب میلی‌متر است. α نسبت انحراف معیار مقادیر برآورد شده به انحراف معیار مقادیر مشاهداتی است. β نسبت میانگین مقادیر برآورد شده به میانگین مقادیر مشاهداتی و n تعداد کل سال‌ها را نشان می‌دهد. در معیار بازبینی شده کلینگ - گوپتا بجای آلفا از گاما استفاده می‌شود که نسبت ضریب تغییرات مقادیر برآورد شده به ضریب تغییرات مقادیر مشاهداتی است (Kling et al., 2012). این آماره‌ها در گام‌های زمانی سالانه برای دوره ۱۹۹۳-۲۰۲۰ با استفاده از بسته نرم‌افزاری الحاقی XLSTAT محاسبه شدند.

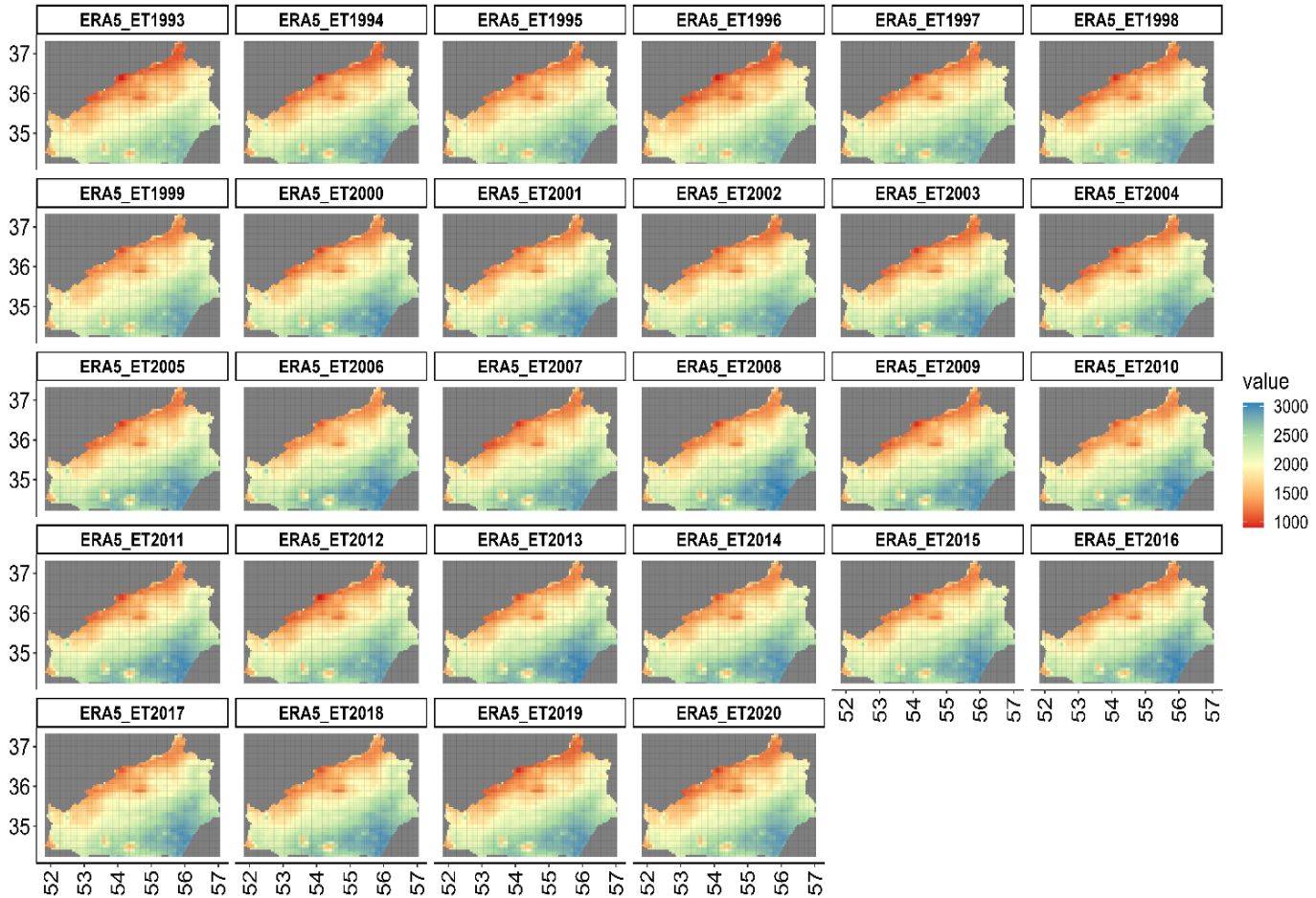
خلاصه آماری متغیرهای کمکی مورد استفاده

در این مطالعه، برخی از متغیرهای مؤثر در برآورد تبخیر و تعرق مانند ارتفاع، دمای سالانه و تبخیر و تعرق حاصل از داده‌های بازتخلیـل شده ERA5-Land در استان سمنان مورد ارزیابی قرار گرفت و ارتباط این متغیرها با داده‌های تبخیر و تعرق مشاهداتی مدل‌سازی گردید. خلاصه آماری متغیرهای کمکی مورد استفاده در جدول ۱ نشان داده شده است. مقدار تبخیر و تعرق داده‌های بازتخلیـل ERA5-Land بین ۱۰۴۲ تا ۲۸۷۷ میلی‌متر در سال در دوره آماری (۱۹۹۳-۲۰۲۰) متغیر بود. مقادیر دمای سالانه در استان بین ۸/۶ تا ۲۲/۸ درجه سانتی‌گراد در دوره آماری (۱۹۹۳-۲۰۲۰) بوده و تغییرات ارتفاع در منطقه مطالعاتی بین ۶۹۰ تا ۳۸۸۵ متر بود که از شمال به جنوب کاهش می‌یابد. ضریب تغییرات (cv) متغیرهای ارتفاع، تبخیر و تعرق ERA5-Land و دما به ترتیب برابر با ۴۱، ۲۲ و ۲۳ درصد به دست آمد. همان‌طور که نتایج ضریب تغییرات نیز نشان می‌دهد کمترین درصد ضریب تغییرات مربوط به متغیر تبخیر و تعرق حاصل از ERA5-Land و بیشترین آن مربوط به متغیر ارتفاع است. ضریب تغییرات نشان‌دهنده دامنه تغییرپذیری یک متغیر بوده که هر چه مقدار این ضریب کمتر باشد نشان‌دهنده پراکنش یکنواخت‌تر ویژگی مورد مطالعه است. باتوجه به دامنه مقادیر ضریب تغییرات بر اساس معیار ارائه شده توسط Wilding, (1985) اگر درصد ضریب تغییرات کمتر از ۱۵ درصد باشد، تغییرپذیری کم، اگر این ضریب بین ۱۵ تا ۳۵ درصد باشد، تغییرپذیری متوسط و اگر ضریب تغییرات بیشتر از ۳۵ درصد باشد، تغییرپذیری ویژگی زیاد است؛ بنابراین متغیرهای متوسط دمای سالانه و تبخیر و تعرق حاصل از داده‌های بازتخلیـل شده ERA5-Land دارای تغییرپذیری متوسط و متغیر کمکی ارتفاع دارای تغییرپذیری زیاد بود که نشان‌دهنده تنوع زیاد متغیر ارتفاع در استان سمنان است. مناطق جنوبی استان دارای اراضی مسطح و با شیب کم بوده که به کویر مرکزی ایران مرتبط می‌شوند. این مناطق کمترین میزان ارتفاع را داشته و مقدار تبخیر و تعرق در این مناطق بالاتر از ۲۵۰۰ میلی‌متر در سال است.

جدول ۱. آماره‌های توصیفی متغیرهای کمکی مؤثر در برآورد تبخیر و تعرق استان سمنان (۱۹۹۳-۲۰۲۰)

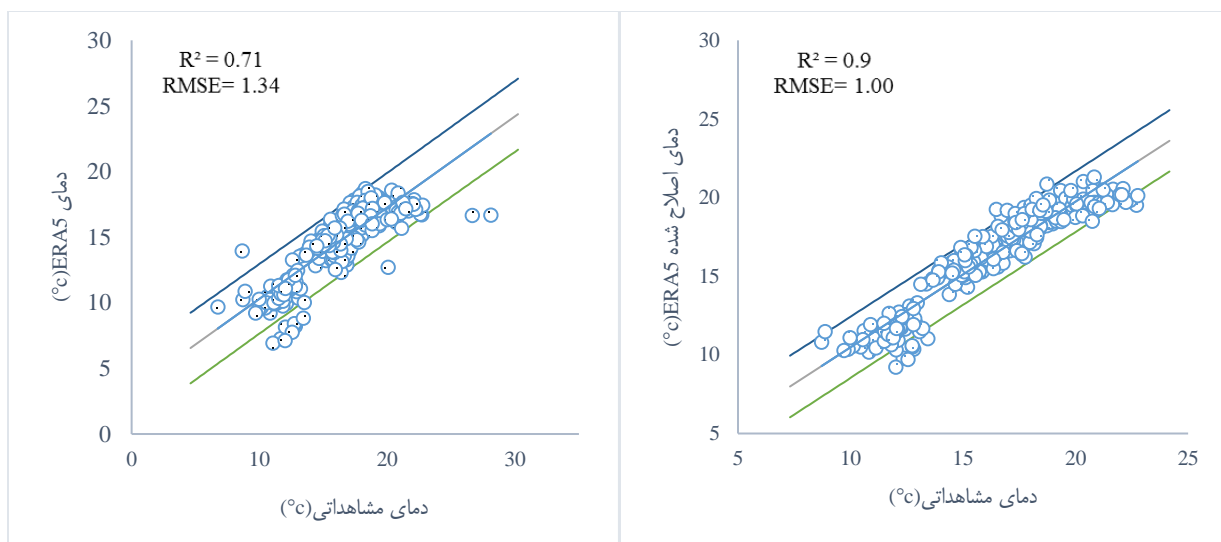
متغیرها	کمینه	بیشینه	میانگین	انحراف معیار	ضریب تغییرات
تبخیر و تعرق ERA5-Land (mm)	۱۰۴۲	۲۸۷۷	۲۰۱۶	۴۴۸	٪۲۲
دمای سالانه (°C)	۸/۶	۲۲/۸	۱۸/۳	۴/۲	٪۲۳
مدل رقومی ارتفاع (m)	۶۹۰	۳۸۸۵	۱۰۵۷	۴۳۷	٪۴۱

پراکنش تبخیر و تعرق بر اساس داده‌های ERA5-Land طی سال‌های (۱۹۹۳-۲۰۲۰) در شکل ۲ نشان داده شده است. بر اساس داده‌های تبخیر و تعرق حاصل از داده‌های بازتخلیـل شده ERA5-Land، در قسمت‌های شمالی استان میزان تبخیر و تعرق کم و در قسمت‌های میانی استان به طرف شمال، مقدار متوسطی از تبخیر و تعرق مشاهده شد که مربوط به ارتفاعات استان بود. همچنین، در بیشتر مناطق شرقی و جنوبی که دارای ارتفاع کم و منتهی به بخش‌های مرکزی و کویری استان است، حداکثر مقدار تبخیر و تعرق مشاهده شد.



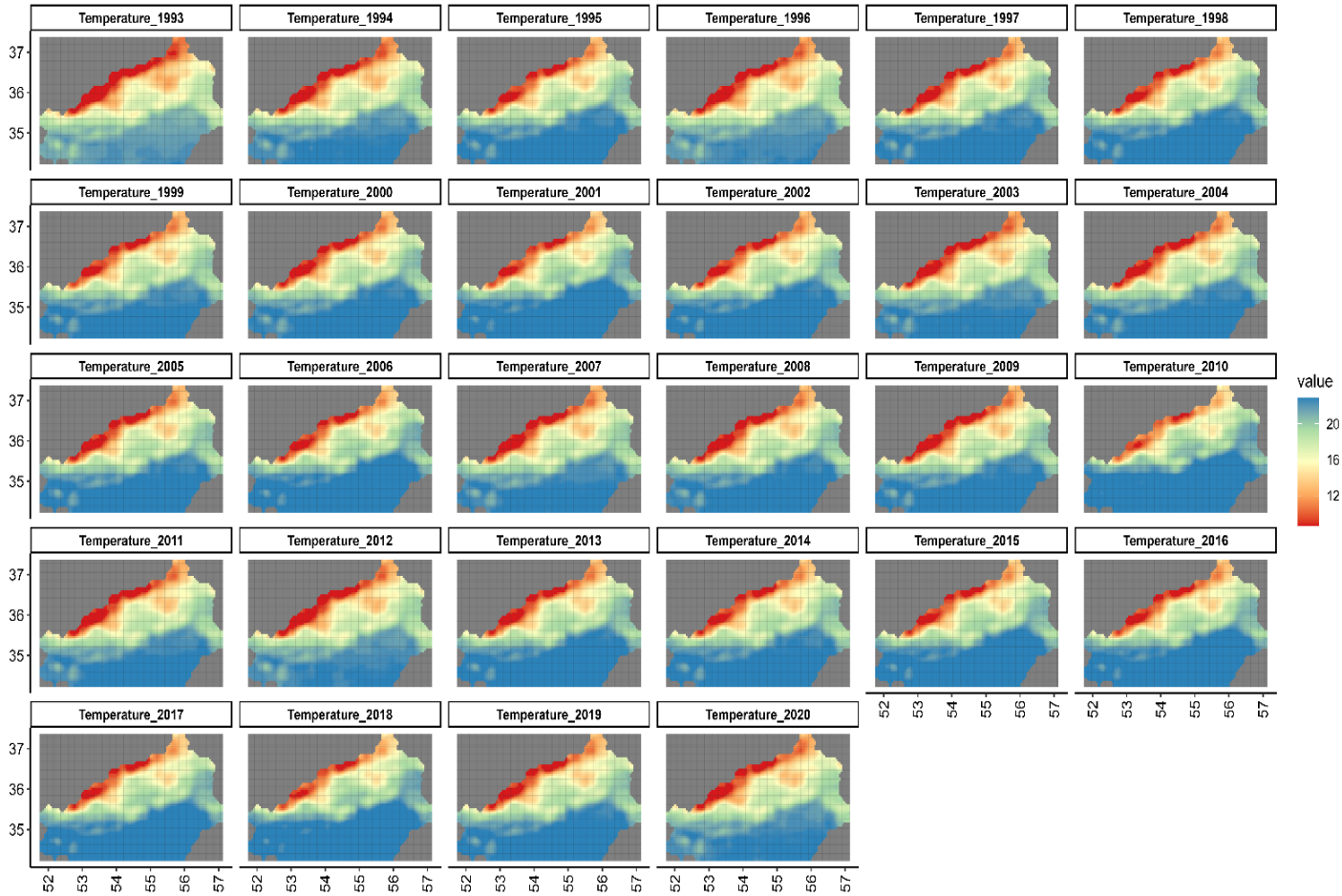
شکل ۲. الگوی توزیع مکانی تبخیر و تعرق حاصل از داده‌های باز تحلیل شده ERA5-Land در دوره آماری (۱۹۹۳-۲۰۲۰)

همان‌طور که بیان شد دما یک متغیر بسیار مهم در برآورد تبخیر و تعرق می‌باشد، لذا اصلاح آریبی داده‌های شبکه‌ای دما اهمیت بسیار زیادی دارد. نتایج ضریب تبیین (R^2) و مقادیر میانگین ریشه مربعات خطا (RMSE) برای داده‌های سالانه دمای بازتحلیل شده ERA5-Land قبل و بعد از اصلاح آریبی در شکل ۳ نشان داده شده است. مقدار R^2 بین داده‌های اندازه‌گیری شده دما و داده‌های بازتحلیل ERA5-Land قبل و بعد از اصلاح آریبی به ترتیب برابر با $0/71$ و $0/9$ متغیر بود. همچنین میزان RMSE در برآورد دما قبل از اصلاح آریبی برابر با $1/34$ درجه سانتیگراد بود که بعد از اصلاح آریبی مقدار آن به 1 کاهش یافت. این نتایج نشان داد که داده‌های دمای سالانه بازتحلیل شده حاصل از ERA5-Land بعد از اصلاح آریبی به‌وسیله روش نگاشت چندک، همبستگی بیشتری با داده‌های مشاهداتی نشان داد. به‌طوری‌که این داده‌ها پس از اصلاح آریبی به خط $1:1$ نزدیکتر بوده و در یک راستا قرار گرفته‌اند.



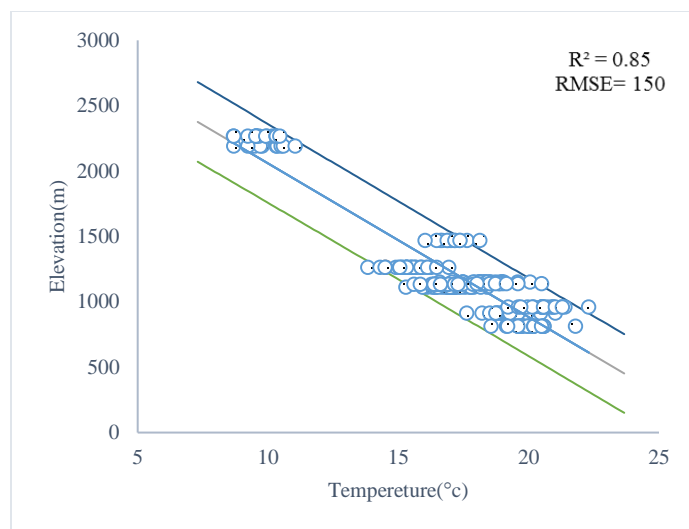
شکل ۳. ارتباط بین دمای مشاهداتی و دمای ERA5-Land (نمودار راست)؛ دمای مشاهداتی و دمای اصلاح شده ERA5-Land (نمودار چپ) در دوره ۱۹۹۳-۲۰۲۰

الگوی توزیع مکانی دمای سالانه حاصل از داده‌های بازتحلیل شده ERA5-Land پس از اصلاح اریبی برای دوره آماری (۱۹۹۳-۲۰۲۰) در شکل ۴ نشان داده شده است. بر اساس نتایج، در قسمت‌های شمالی استان میزان دما کمینه بوده و در قسمت‌های مرکزی، متوسطی از دما مشاهده شد که مربوط به ارتفاعات میانی منطقه مطالعاتی است. همچنین در بیشتر مناطق شرقی و جنوبی که دارای ارتفاع کم بوده و مناطق کویری را شامل می‌شوند، حداکثر مقدار دمای سالانه مشاهده گردید (شکل ۴).



شکل ۴. الگوی توزیع مکانی داده‌های سالانه دمای اصلاح‌اریبی شده ERA5-Land در دوره آماری (۱۹۹۳-۲۰۲۰)

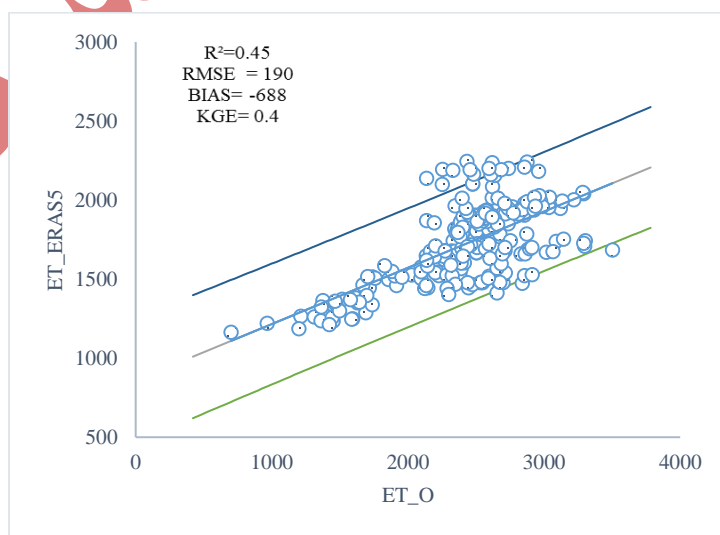
بررسی همبستگی بین دمای سالانه ERA5-Land بعد از اصلاح‌اریبی و نتایج حاصل از متغیر ارتفاع در منطقه مورد مطالعه نشان داد ارتباط منفی و معنی‌داری بین دما و متغیر ارتفاع وجود دارد. کمترین میانگین دمای سالانه برآورد شده مربوط به ایستگاه‌های واقع شده در طبقات ارتفاعی بیش از ۲۰۰۰ متر است و در طبقات ارتفاعی میانی (۱۰۰۰ تا ۱۵۰۰ متر)، میانگین تغییرات دمای سالانه بین ۱۴ تا حدود ۲۰ درجه سانتی‌گراد به دست آمد. در این مطالعه کمترین میانگین دمای سالانه (۸/۶ درجه سانتی‌گراد) مربوط به ایستگاه مچن است و بیشترین آن (۲۲/۸ درجه سانتی‌گراد) مربوط به ایستگاه آرادان می‌باشد که در طبقه ارتفاعی کمتر از ۸۵۰ متر قرار دارد (شکل ۵).



شکل ۵. ارتباط بین متغیر ارتفاع و دمای سالانه ERA5-Land پس از اصلاح اریبی در دوره آماری (۱۹۹۳-۲۰۲۰)

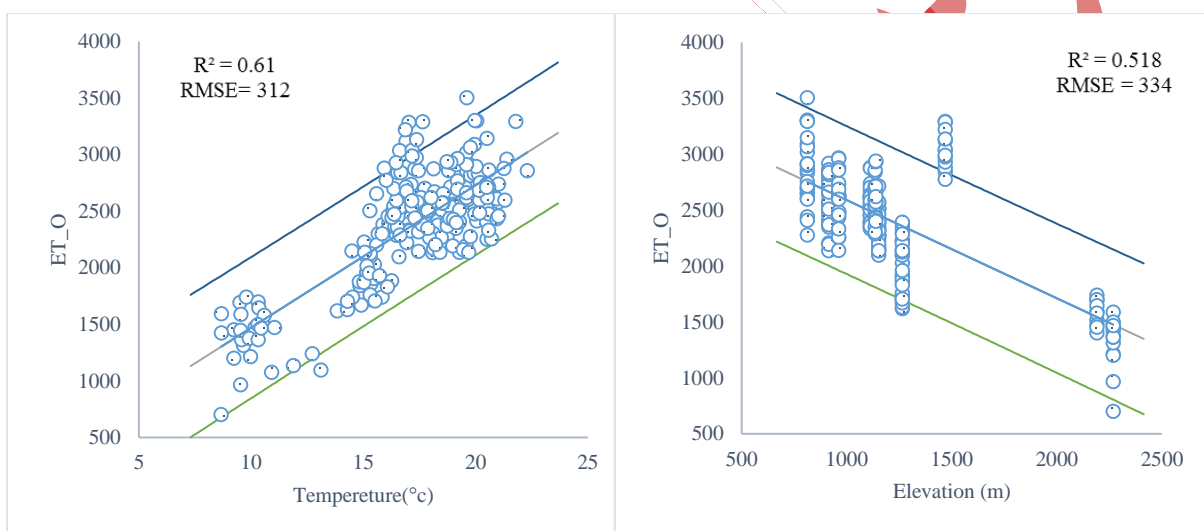
بررسی ارتباط بین متغیرهای کمکی و مقادیر تبخیر و تعرق اندازه‌گیری شده

نتایج نشان داد که ضریب همبستگی پیرسون بین داده‌های تبخیر و تعرق حاصل از ERA5-Land و تبخیر و تعرق ایستگاه‌های زمینی، مثبت و برابر با $r = 0.67^*$ است که نشان‌دهنده ارتباط معنی‌دار این داده‌ها در برآورد تبخیر و تعرق در مناطق فاقد ایستگاه‌های زمینی است (شکل ۶). همچنین مقدار آماره RMSE برابر با ۱۹۰ میلی‌متر در سال محاسبه شد که نشان می‌دهد که استفاده از این داده‌ها به‌تنهایی به‌طور متوسط ۱۹۰ میلی‌متر خطا در برآورد تبخیر و تعرق سالانه منطقه مورد مطالعه خواهد داشت. اگرچه مقادیر KGE نتایج قابل قبول را برای این داده‌ها نشان می‌دهد؛ اما نتایج خطای مدل (BIAS) نشان داد که داده‌های ERA5-Land میزان واقعی تبخیر را کمتر از مقدار واقعی برآورد می‌کنند؛ لذا استفاده مستقیم از داده‌ها سبب خطا در برآورد تبخیر و تعرق می‌شود (شکل ۶).



شکل ۶. ارتباط بین تبخیر و تعرق مشاهداتی و تبخیر و تعرق حاصل از داده‌های باز تحلیل شده ERA5-Land در دوره آماری (۱۹۹۳-۲۰۲۰)

شکل ۷ نشان می‌دهد که مقادیر ضریب تعیین بین تبخیر و تعرق مشاهداتی و ارتفاع بیش از ۰/۵ می‌باشد که نشان‌دهنده ارتباط معنی‌داری بین مقادیر این دو متغیر است ($R^2=0/51$). همچنین، مقدار ضریب همبستگی بین متغیر دمای سالانه ERA5-Land پس از اصلاح آریبی با داده‌های تبخیر و تعرق مشاهداتی مثبت بود ($r=0/78^*$) که نشان می‌دهد ارتباط مستقیمی بین دمای سالانه و تبخیر و تعرق وجود دارد، به طوری که با افزایش دمای سالانه، میزان تبخیر و تعرق نیز در استان افزایش می‌یابد. با توجه به توپوگرافی منطقه مطالعاتی، مناطق کوهستانی در بخش شمالی استان سمنان دارای کمترین دمای سالانه و تبخیر و تعرق هستند، در حالی که در بخش‌های جنوبی استان با افزایش دمای سالانه، تبخیر و تعرق سالانه نیز افزایش می‌یابد. کاهش دما در استان معمولاً با افزایش ارتفاع همراه است؛ بنابراین می‌توان انتظار داشت که با افزایش ارتفاع در بخش‌های شمالی استان، هر دو متغیر دما و تبخیر و تعرق کاهش یابند. همبستگی منفی و معنی‌دار بین ارتفاع نقاط مشاهداتی و مقادیر سالانه تبخیر و تعرق مشاهداتی ($r=-0/72^*$) موید همین مطلب است که می‌توان آن را ناشی از تأثیر متغیر ارتفاع بر تبخیر و تعرق دانست (شکل ۷).

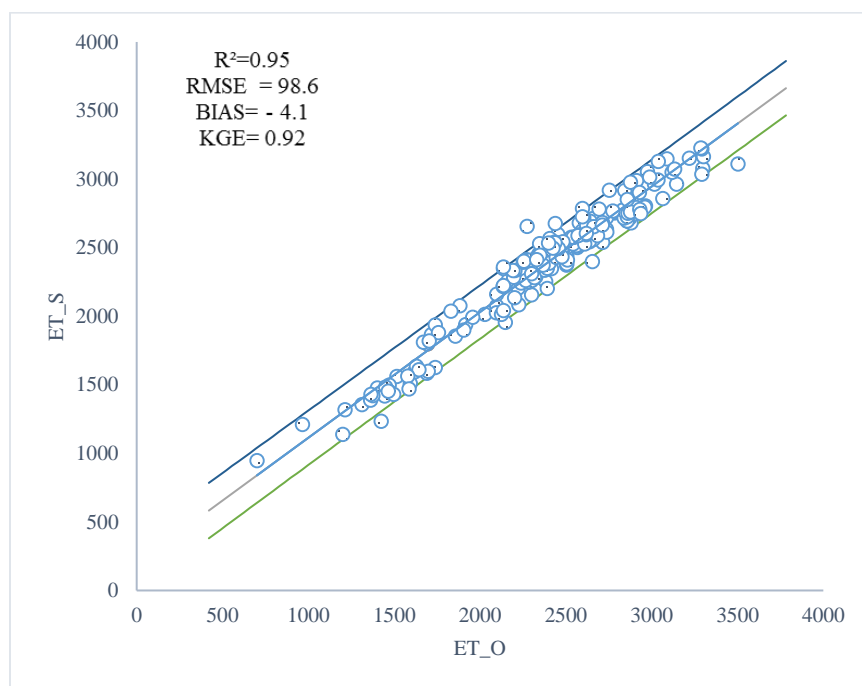


شکل ۷. ارتباط ET مشاهداتی با ارتفاع (نمودار چپ)؛ و ET مشاهداتی با دمای سالانه ERA5-Land پس از اصلاح آریبی (نمودار راست) در دوره آماری (۱۹۹۳-۲۰۲۰)

ارزیابی مدل RF و تهیه نقشه تبخیر و تعرق در استان سمنان

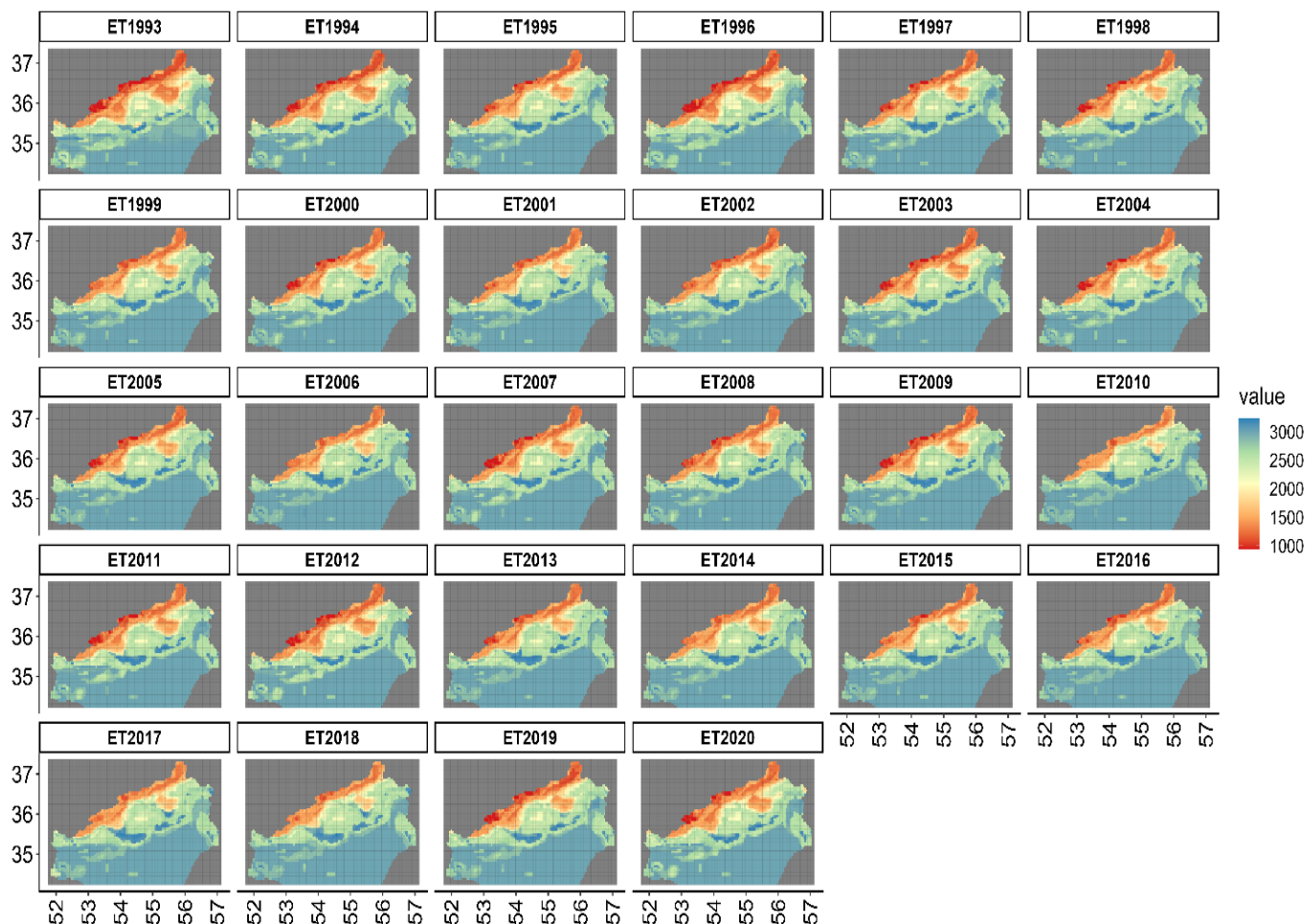
نمودار پراکنندگی داده‌های برآورد شده تبخیر و تعرق سالانه با استفاده از مدل RF در برابر داده‌های مشاهداتی در مقیاس زمانی سالانه (۱۹۹۳-۲۰۲۰) در شکل ۸ نشان داده شده است. با توجه به ضریب تعیین و KGE بالای بین مقادیر سالانه ET مشاهداتی و ET برآورد شده ($R^2=0/95$ ، $KGE=0/92$)، نتایج نشان داد که مدل RF به خوبی توانسته است مقادیر تبخیر و تعرق سالانه را در منطقه مورد مطالعه برآورد نماید. همچنین میانگین ریشه مربعات خطا (RMSE) ۹۸/۶ محاسبه شد. به عبارت دیگر، حدود ۹۹ میلی‌متر خطا در برآورد تبخیر و تعرق سالانه ایستگاه‌های مورد مطالعه با استفاده از مدل RF وجود داشت. که نشان از دقت بالای این مدل در برآورد تبخیر و تعرق استان است (شکل ۸). داده‌های تبخیر و تعرق برآورد شده با استفاده از مدل RF در تمامی دامنه داده به خط ۱:۱ نزدیک بوده و در یک راستا قرار گرفته‌اند. این نتایج نشان می‌دهد که استفاده از متغیرهای کمکی به همراه یک روش یادگیری ماشین، قادر به برآورد تبخیر و تعرق در تمامی مقادیر تبخیر و تعرق و در اقلیم‌های مختلف خواهد شد. مقدار خطای مدل (BIAS) برای مقادیر برآورد شده

تبخیر و تعرق با استفاده از مدل RF ۴/۱- بود که بیانگر کارایی بالای مدل RF در برآورد تبخیر و تعرق سالانه در استان است. این نتایج نشان می‌دهد که متغیرهای کمکی ارتفاع، دما و ET حاصل از داده‌های بازتحلیل شده ERA5-Land در صورت استفاده در یک مدل غیرخطی مانند مدل RF به‌خوبی قادر به برآورد متغیری با تغییرپذیری مکانی بالا مانند تبخیر و تعرق در مناطق فاقد داده اندازه‌گیری هستند.



شکل ۸. ارتباط بین تبخیر و تعرق اندازه‌گیری شده و برآورد شده حاصل از مدل RF در دوره آماری (۱۹۹۳-۲۰۲۰)

پس از بررسی کارایی و دقت مدل RF در برآورد تبخیر و تعرق سالانه در مناطق فاقد داده‌های مشاهده‌ای، نقشه تبخیر و تعرق سالانه در استان با قدرت تفکیک مکانی ۵ کیلومتر برای دوره آماری (۱۹۹۳-۲۰۲۰) تهیه شد (شکل ۹). حداکثر مقادیر تبخیر و تعرق بر اساس تبخیر و تعرق برآورد شده ۳۲۲۶ میلی‌متر و حداقل آن حدود ۹۴۶ میلی‌متر در سال برآورد شد. در حالی که بر اساس داده‌های بازتحلیل شده ERA5-Land میزان تبخیر و تعرق بین ۱۱۶۵ تا ۲۲۴۶ میلی‌متر در سال به دست آمد. تغییرات مکانی تفاوت ۲۳ الی ۴۳ درصدی را با داده‌های بازتحلیل ERA5-Land در مناطق مختلف استان نشان می‌دهد. بر اساس داده‌های تبخیر و تعرق به دست آمده از مدل RF در قسمت‌های شمالی استان میزان تبخیر و تعرق کم و در قسمت‌های میانی استان به طرف شمال، متوسطی از تبخیر و تعرق مشاهده شد که مربوط به ارتفاعات متوسط منطقه بود و در بیشتر مناطق شرقی و جنوبی که منطبق بر مناطق کم‌ارتفاع و کویری استان بود، حداکثر مقدار تبخیر و تعرق مشاهده گردید (شکل ۹).



شکل ۹. نقشه تبخیر و تعرق تهیه شده استان سمنان با استفاده از متغیرهای کمکی و مدل RF برای دوره آماری (۱۹۹۳-۲۰۲۰)

شاخص‌های آماری مدل RF به تفکیک ایستگاه‌های مطالعاتی

در جدول ۲ نتایج آماره‌های مورد استفاده در برآورد تبخیر و تعرق سالانه به تفکیک ایستگاه‌های اندازه‌گیری شده استان سمنان با مدل RF برای دوره ۱۹۹۳-۲۰۲۰ نشان داده شده است. تجزیه و تحلیل مستقل ایستگاه‌های مورد بررسی این امکان را می‌دهد که عملکرد مدل RF در برآورد تبخیر و تعرق را به کمک متغیرهای ورودی ارزیابی نمائیم. با بررسی مقادیر شاخص‌های آماری مندرج در جدول ۲ می‌توان گفت که مدل RF در برآورد تبخیر و تعرق در استان سمنان از کارایی بالایی برخوردار است. مقادیر ضریب همبستگی پیرسون در سطح معنی‌داری ۵ درصد در تمامی ایستگاه‌ها، بین ۰/۷ تا ۰/۹۶ متغیر بود. همچنین، مقدار میانگین ریشه مربعات خطا (RMSE) در ایستگاه‌های مورد بررسی در استان سمنان بین ۱۷/۲۱ تا ۸۲/۵۹ متغیر است. در تمامی ایستگاه‌های مورد بررسی مقادیر KGE بین ۰/۴۵ تا ۰/۷۴ به دست آمد که این نتایج عملکرد خوب مدل RF در برآورد تبخیر و تعرق در تمامی ایستگاه‌های مطالعاتی بر اساس معیار ارزیابی کلینگ - گوپتا ارائه می‌دهد (جدول ۲). البته بدترین عملکرد این معادلات برای برآورد سالانه تبخیر و تعرق در ایستگاه شه‌میرزاد با (KGE=۰/۴۵) و خطای مدل (BAIS=-۲۸/۰۲) به دست آمد (جدول ۲).

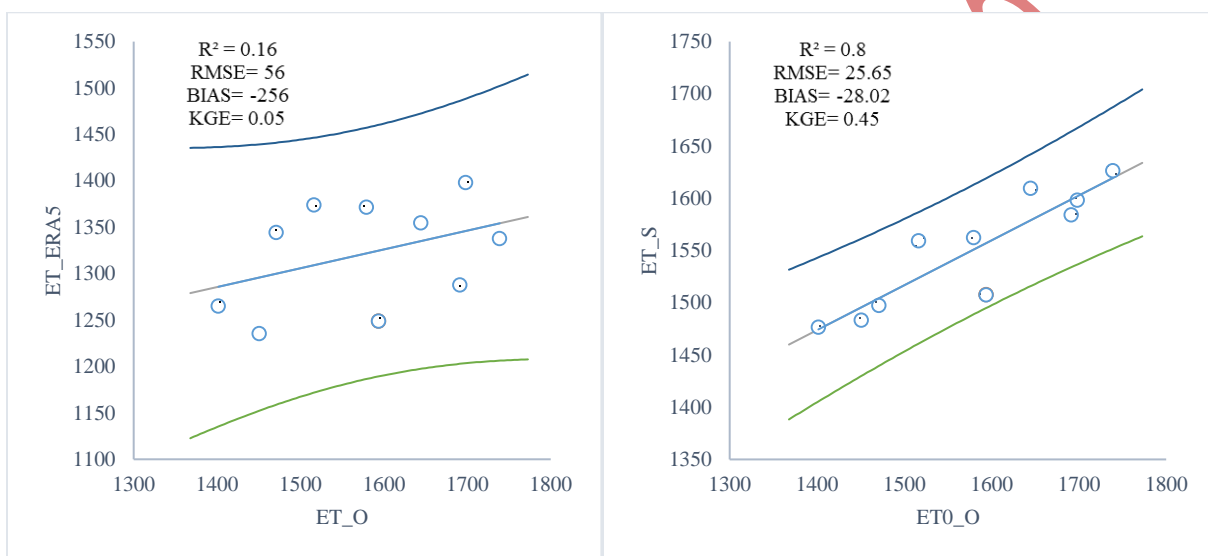
جدول ۲. آماره محاسباتی سالانه برای تبخیر و تعرق با مدل RF به تفکیک ایستگاه‌های مورد مطالعه

ایستگاه	ارتفاع	طول	عرض	R ²	RMSE	KGE	BIAS
شاهرود	۱۲۶۴	۵۴/۹۲	۳۶/۳۸	۰/۸۱	۷۶/۹۶	۰/۶۵	-۱/۹۴
بیارجمند	۱۱۱۲	۵۵/۸۰	۳۶/۰۸	۰/۷۵	۴۱/۰۱	۰/۴۹	-۷/۹۴
سمنان	۱۱۵۲	۵۳/۴۲	۳۵/۵۸	۰/۸۲	۴۵/۹۰	۰/۶۵	۲/۲۵
گرمسار	۹۱۴	۵۲/۳۶	۳۵/۲۴	۰/۸۵	۶۲/۱۶	۰/۶۶	-۱/۲۷
دامغان	۱۱۳۵	۵۴/۳۲	۳۶/۱۴	۰/۷۰	۶۴/۷۱	۰/۵۸	-۱۶/۰۲
شهمیرزاد	۲۱۹۰	۵۳/۳۵	۳۵/۷۶	۰/۸۰	۲۵/۶۵	۰/۴۵	-۲۸/۰۲
آرادان	۸۱۲	۵۲/۷۳	۳۵/۲۵	۰/۷۸	۸۲/۵۹	۰/۵۱	-۹/۲۲
مجن	۲۲۶۸	۵۴/۶۵	۳۶/۴۸	۰/۷۹۰	۷۱/۹۶	۰/۵۷	۱۹/۰۶۲
فرومد	۱۴۶۹	۵۶/۷۷	۳۶/۵۸	۰/۸۴	۵۲/۹۵	۰/۷۴	-۳/۱۶۸
حسینان	۹۶۱	۵۴/۳۱	۳۵/۱۷	۰/۸۴	۵۸/۶۵	۰/۶۰	-۰/۶۷۶
فرات	۱۱۳۹	۵۴/۴۵	۳۵/۶۶	۰/۹۶	۱۷/۲۱	۰/۴۹	-۷/۹۱

بحث

در شکل ۱۰ نتایج شاخصه‌های آماری (مقادیر مشاهداتی در مقابل مقادیر برآورد شده) با مدل RF و ERA5-Land برای ایستگاه هواشناسی شهمیرزاد با ارتفاع تقریبی ۲۲۰۰ متر نشان داده شده است. با توجه به شکل ۱۰ در این ایستگاه مدل RF دقت بیشتری (با مقدار RMSE کمتر و شاخص KGE بهتر) نسبت به داده‌های تبخیر و تعرق باز تحلیل شده ERA5-Land قبل از اصلاح اریبی را نشان می‌دهد. در ایستگاه شهمیرزاد مقادیر میانگین ریشه مربعات خطا (RMSE) از حدود ۲۶ میلی‌متر در سال برای مدل RF تا ۵۶ میلی‌متر در سال برای ERA5-Land به دست آمده است (شکل ۱۰). اما نتایج RMSE به تنهایی برای نشان دادن بهترین روش برای برآورد تبخیر و تعرق کافی نیست. در واقع این موضوع اهمیت تجزیه و تحلیل مشترک دیگر شاخص‌های آماری را تقویت می‌کند. همچنین بر اساس شکل ۱۰ مقادیر KGE برای شبیه‌سازی داده‌های تبخیر و تعرق مدل RF و تبخیر و تعرق داده‌های باز تحلیل شده ERA5-Land در این ایستگاه، در محدوده قابل قبول ($-0.41 < KGE \leq 1$) قرار می‌گیرد (Knoben et al., 2019). با این حال KGE یک الگوی فضایی پیچیده‌تر در ارتباط با خطای مدل (BIAS) در این ایستگاه‌ها ارائه می‌دهد. در واقع میزان خطای مدل (BIAS) یک معیار مهم در مدل‌سازی متغیرهای اقلیمی و هیدرولوژیکی است، زیرا اطلاعاتی را در مورد اینکه آیا مقادیر مدل‌سازی شده در یک حوضه یا منطقه به خوبی پیش‌بینی شده است را ارائه می‌دهد. میزان خطای داده‌های باز تحلیل شده ERA5-Land در برآورد تبخیر و تعرق در مقایسه با داده‌های مشاهداتی عدد بالای ۲۵۶- به دست آمد. در صورتی که میزان خطای مدل RF در برآورد تبخیر و تعرق در مقایسه با داده‌های مشاهداتی عدد ۲۸/۰۲- به دست آمد (شکل ۱۰). این میزان از خطای کمتر مدل RF در برآورد تبخیر و تعرق در ایستگاه شهمیرزاد (واقع در مناطق کوهستانی) به این دلیل است که مدل RF بر خلاف ERA5-Land یک درون‌یابی فضایی را اعمال می‌کند که پیش‌بینی‌های خود را با متغیرهای کمکی که تأثیر مستقیم بر تبخیر و تعرق دارند، مانند متغیرهای توپوگرافی انجام می‌دهد (Strong et al., 2017). این موضوع زمانی اهمیت می‌یابد که طبق نتایج به دست آمده متغیر ارتفاع در رده با تغییرپذیری زیاد بر اساس شاخص Wilding, (1985) در کل استان قرار می‌گیرد (جدول ۱). این ارزیابی، وابستگی عملکرد به موقعیت مکانی در برآورد داده‌های تبخیر و تعرق مدل RF در مقایسه با تبخیر و تعرق داده‌های باز تحلیل شده ERA5-Land را نشان می‌دهد. به عبارت دیگر اگرچه وضوح فضایی بالاتر ERA5-Land می‌تواند در مناطقی با توپوگرافی پیچیده مفید باشد، اما هیچ بهبودی در رابطه با تبخیر و تعرق ناشی از توپوگرافی حاصل نشده است که با نتایج تحقیقات Zandler et al. (2019) و Hamm et al. (2020) نیز مطابقت دارد. همچنین متغیر

ارتفاع بر متغیر دما نیز تأثیرگذار است و نشان‌دهنده همبستگی منفی و معنی‌دار بین این دو متغیر است (شکل ۵). در واقع با افزایش ارتفاع در استان سمنان، دما کاهش و تبخیر و تعرق هم کاهش پیدا می‌کند. تخمین بیش از حد خطای مدل داده‌های بازتحلیل شده ERA5-Land در سایر مناطق جهان در تحقیقات دیگران نیز بیان شده است (Amjad et al., 2020; Crossett et al., 2020; Bandhauer et al., 2022). از این منظر، به دلیل استفاده از متغیرهای کمکی در مدل جنگل تصادفی RF و به دلیل روشی که روندهای سیستماتیک داده‌های نمونه را بیش‌برآورد یا کم برآورد ایجاد نمی‌کنند (Wu et al., 2019; Wang et al., 2019) عملکرد مدل RF برای درون‌یابی داده‌های سالانه تبخیر و تعرق قابل توجه است؛ همان‌طور که در تحقیقات (Feng et al., 2017; Wu et al., 2019; Wang et al., 2019) و Silva et al. (2019) نیز نشان داده شده است.



شکل ۱۰. ارتباط ET مشاهداتی و شبیه‌سازی مدل RF (نمودار چپ)؛ و ET مشاهداتی و شبیه‌سازی ERA5-Land (نمودار راست) در ایستگاه شه‌میرزاد

نتیجه‌گیری

ارزیابی نتایج تطابق خوبی بین داده‌های مشاهداتی و برآورد شده مدل RF نشان داد ($R^2 = 0.95$ ، $BIAS = -4.1$ ، $RMSE = 98.6$ و $KGE = 0.92$). استفاده از داده‌های کاوی و پردازش در محیط برنامه‌نویسی R، در مناطقی با تعداد محدود ایستگاه هواشناسی، دقت مناسبی در برآورد تبخیر و تعرق در مناطق خشک و نیمه‌خشک ارائه می‌دهد. همچنین می‌تواند به برنامه‌های مدیریت کارآمد آب و چندین زمینه تحقیقاتی، از جمله هواشناسی، کشاورزی و اکولوژی کمک نماید. ذکر این نکته ضروری است اگرچه عوامل کمکی مختلفی برای پیش‌بینی متغیر تبخیر و تعرق در مدل RF در نظر گرفته شده است؛ اما مطالعات در این زمینه نشان داده است که متغیرهایی مانند تشعشعات خورشیدی نیز می‌تواند سبب افزایش دقت در برآورد تبخیر و تعرق شود (Brust et al., 2021; Cunha et al., 2021; Wang et al., 2021). اما معمولاً این اطلاعات به‌صورت شبکه‌ای در دسترس نبوده است؛ بنابراین استفاده از این متغیرها به‌عنوان متغیرهای کمکی امکان‌پذیر نمی‌باشد. تجزیه و تحلیل مستقل ایستگاه‌های مورد بررسی، وابستگی عملکرد خوب مدل RF به موقعیت مکانی را نیز نشان داد. بررسی ارتباط بین داده‌های برآورد شده با داده‌های مشاهداتی نشان داد عملکرد مدل RF در برآورد تبخیر و تعرق به کمک متغیرهای ورودی به وابستگی خطای مدل (BIAS) به توپوگرافی بستگی دارد و متغیر ارتفاع عامل مهمی در تصحیح سوگیری سیستماتیک داده‌ها به‌منظور دقت در برآورد تبخیر و تعرق محسوب می‌گردد. در واقع خطای مدل (BIAS) مجموعه داده‌های برآوردی

ERA5-Land به‌عنوان فاکتور مهم در تشخیص صحت ارزیابی KGE مدل، مطابق با الگوریتم جنگل تصادفی RF تصحیح شده است. داده‌های بازتحلیل ERA5-Land، مناطقی از اراضی کوهستانی بالادست استان را که دارای تبخیر و تعرق کمتری هستند با خطای زیادی برآورد کرده‌اند. به‌عبارت‌دیگر داده‌های بازتحلیل تبخیر و تعرق ERA5-Land با داشتن قدرت تفکیک زمانی و مکانی بالا برای مناطق عاری از ایستگاه تبخیرسنجی استفاده می‌گردد، اما مقدار مطلق آنها معمولاً بیانگر تبخیر و تعرق یک منطقه نبوده و مقادیر تبخیر و تعرق حاصل از این‌گونه داده‌ها با خطا همراه است. در بسیاری از مطالعات از محصول تبخیر و تعرق داده‌های بازتحلیل شده ERA5-Land در مدل‌سازی هیدرولوژیکی استفاده می‌شود. اما رویکرد در نظر گرفته شده در این مدل از یک محدودیت موجود در بسیاری از مطالعات رنج می‌برد: عدم قطعیت مرتبط با مشاهدات؛ همچنین کیفیت این محصول فضایی، به‌شدت به شرایط توپوگرافی وابسته است (Herrera et al., 2019). براین‌اساس محصول تبخیر و تعرق داده‌های بازتحلیل ERA5-Land، قبل از استفاده در مطالعات هیدرولوژیکی، کشاورزی و فرسایش باید واسنجی شده و سپس مورد استفاده قرار گیرند. نتایج حاصل از پژوهش حاضر می‌تواند به‌عنوان مبنایی برای برآورد تبخیر و تعرق با استفاده از متغیرهای کمکی در مناطق خشک و نیمه‌خشک قرار گیرد.

پیشنهادها

در پژوهش‌های آینده لازم است داده‌های تبخیر و تعرق حاصل از سایر محصولات بازتحلیل شده و نیز چندین روش یادگیری ماشین برای برآورد دقیق‌تر تبخیر و تعرق در دوره زمانی پایه مورد استفاده قرار گیرد. همچنین استفاده از دامنه وسیع‌تری از داده‌ها از مناطق همجوار استان با شرایط اقلیمی مشابه، می‌تواند به کاهش میزان خطا در برآورد تبخیر و تعرق در منطقه مورد مطالعه کمک نماید.

سپاس‌گزاری

این مقاله از طرح پسادکتری دانشگاه سمنان با شماره قرارداد ۲۳۵۰۰ استخراج گردیده است.

منابع

امینی، الهام؛ ذوالفقاری، علی‌اصغر؛ کابلی، سید حسن و رحیمی، محمد (۱۴۰۱). برآورد نقشه‌های فرسایشی و بارش در مناطقی با ایستگاه باران‌سنجی محدود (مطالعه موردی: استان سمنان). *مجله تحقیقات آب و خاک ایران*، ۵۳(۹)، ۲۰۴۴-۲۰۲۷.

REFERENCES

- Abed, M., Imteaz, M.A., Ahmed, AN., & Huang YF. (2022). Modelling monthly pan evaporation utilising Random Forest and deep learning algorithms. *Scientific Reports*, 12(1), 1-29. <https://dx.doi.org/10.1038/s41598-022-17263-3>.
- Ahmadi, F., Mehdizadeh, S., Mohammadi, B., Bao Pham, Q., Doan, T.N.C., & Vo, N.D. (2021). Application of an artificial intelligence technique enhanced with intelligent water drops for monthly reference evapotranspiration estimation. *Agricultural Water Management*, 244, 106622. <https://doi.org/10.1016/j.agwat.2020.106622>.
- Amini, E., Zolfaghari, A.A., kaboli, S.H., & Rahimi, M. (2022). Estimation of Rainfall Erosivity Map in Areas with Limited Number of Rainfall Station (Case study: Semnan Province). *Iranian Journal of Soil and Water Research*, 53(9), 2027-2044. <https://doi.org/10.22059/IJSWR.2022.343710.669279>. (In Persian).
- Amjad, M., Yilmaz, M.T., Yucel, I., & Yilmaz K.K. (2020). Performance Evaluation of Satellite- and Model-Based Precipitation Products over Varying Climate and Complex Topography. *Journal of Hydrology*, 584, 124707. <https://doi.org/10.1016/j.jhydrol.2020.124707>.

- Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Baserud, L., Izsak, B., Szentes, O., Tveito, O.E., & Frei, C. (2022). Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. *International Journal of Climatology*, 42(1), 727-747. <https://doi.org/10.1002/joc.7269>.
- Brust, C., Kimball, J.S., Maneta, M.P., Jencso, K., He, M., & Reichle, R.H. (2021). Using SMAP Level-4 soil moisture to constrain MOD16 evapotranspiration over the contiguous USA. *Remote Sensing of Environment*, 255(1), 112277. <https://doi.org/10.1016/j.rse.2020.112277>.
- Chen, J.M., Chen, X., Ju, W., Geng, X., & Zhang, X. (2005). Distributed hydrological model for mapping evapotranspiration using remote sensing inputs. *Journal of Hydrology*, 305(1-4), 15-39. <https://doi.org/10.1016/j.jhydrol.2004.08.029>.
- Chen, Z., Sun, S., Wang, Y., Wang, Q., & Zhang, X. (2020). Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation. *Computers and Electronics in Agriculture*, 169, 105206. <https://doi.org/10.1016/j.compag.2019.105206>.
- Crossett, C.C., Betts, A.K., Dupigny-Giroux, L.A.L., & Bomblies, A. (2020). Evaluation of daily precipitation from the ERA5 global reanalysis against GHCN observations in the northeastern United States. *Climate*, 8(12), 1-14. <https://doi.org/10.3390/cli8120148>.
- Cunha, A.C., Filho, L.R.A.G., Tanaka, A.A., Goes, B.C., & Putti, F.F. (2021). Influence of the estimated global solar radiation on the reference evapotranspiration obtained through the penman-monteith FAO 56 method. *Agricultural Water Management*, 243, 106491. <https://doi.org/10.1016/j.agwat.2020.106491>.
- Feng, Y., Cui, N., Gong, D., Zhang, Q., & Zhao, L. (2017). Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management*, 193, 163-173. <https://doi.org/10.1016/j.agwat.2017.08.003>.
- Guo, B., Zhang, D., Pei, L., Su, Y., Wang, X., Bian, Y., Zhang, D., Yao, W., Zhou, Z., & Guo, L. (2021). Estimating PM2.5 concentrations via random forest method using satellite, auxiliary, and ground-level station dataset at multiple temporal scales across China in 2017. *Science of The Total Environment*, 778, 146288. <https://doi.org/10.1016/j.scitotenv.2021.146288>.
- Gupta, H.V., Kling, H., Yilmaz, K., & Martinez, G. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2): 80-91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hameed, M.M., AlOmar, M.K., Razali, S.F.M., Khalaf, A.M.K., Baniya, W.J., Sharafati, A., & AlSaadi, M.A. (2021). Application of Artificial Intelligence Models for Evapotranspiration Prediction along the Southern Coast of Turkey. *Hindawi*, 8850243. <https://doi.org/10.1155/2021/8850243>.
- Hamm, A., Arndt, A., Kolbe, C., Wang, X., Thies, B., Boyko, O., Reggiani, P., Scherer, D., Bendix, J., & Schneider, C. (2020). Intercomparison of gridded precipitation datasets over a sub-region of the Central Himalaya and the southwestern Tibetan plateau. *Water*, 12(11), 1-23. <https://doi.org/10.3390/w12113271>.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., & Graler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ (Environmental Science)*, 6, 1-49. <https://doi.org/10.7287/peerj.preprints.26693v2>.
- Herrera, S., Cardoso, R.M., Soares, P.M., Espirito-Santo, F., Viterbo, P., & Gutierrez, J.M. (2019). A new gridded dataset of daily precipitation and temperatures over Iberia. *Earth System Science Data*, 11(4), 1947-1956. <https://doi.org/10.5194/essd-11-1947-2019>.
- Kisi, O. (2016). Modeling reference evapotranspiration using three different heuristic regression approaches. *Agricultural Water Management*, 169, 162-172. <https://doi.org/10.1016/j.agwat.2016.02.026>.
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425, 264-277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>.

- Knoben, W.J.M., Freer, J.E., & Woods, R.A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23 (10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>.
- Krajewski, W.F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilla, R., Niemeier, J., Quintero, F., Seo, B.C., Small, S.J., Webber, L.J., & Young, N.C. (2017). Real-time flood forecasting and information system for the state of Iowa. *Bulletin of the American Meteorological Society*, 98(3), 539–554. <https://doi.org/10.1175/BAMS-D-15-00243.1>.
- Li, Y., Wang, W., Wang, G., & Tan, Q. (2022). Actual evapotranspiration estimation over the Tuojiang River Basin based on a hybrid CNN-RF model. *Journal of Hydrology*, 610, 127788. <https://doi.org/10.1016/j.jhydrol.2022.127788>
- Rashid Niaghi, A., Hassanijalilian, O., & Shiri, J. (2021). Estimation of Reference Evapotranspiration Using Spatial and Temporal Machine Learning Approaches. *Hydrology*, 8(1), 1-15. <https://doi.org/10.3390/hydrology8010025>.
- Shiri, J. (2018). Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology. *Journal of Hydrology*, 561, 737–750. <https://doi.org/10.1016/j.jhydrol.2018.04.042>.
- Silva Junior, J.C., Medeiros, V., Garrozi, C., Montenegro, A., & Goncalves, G.E. (2019). Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian' Northeast. *Computers and Electronics in Agriculture*, 166, 105017. <https://doi.org/10.1016/j.compag.2019.105017>.
- Strong, C., Khatri, K.B., Kochanski, A.K., Lewis, C.S., & Allen, L.N. (2017). Reference evapotranspiration from coarse-scale and dynamically downscaled data in complex terrain: Sensitivity to interpolation and resolution. *Journal of Hydrology*, 548, 406–418. <https://doi.org/10.1016/j.jhydrol.2017.02.045>.
- Torres, A.F., Walker, W.R., & McKee, M. (2011). Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agricultural Water Management*, 98(4), 553-562. <https://doi.org/10.1016/j.agwat.2010.10.012>.
- Wang, S., Lian, J., Peng, Y., Hu, B., & Chen, H. (2019). Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. *Agricultural Water Management*, 221, 220–230. <https://doi.org/10.1016/j.agwat.2019.03.027>.
- Wang, Y., Zhang, Y., Yu, X., Jia, G., Liu, Z., Sun, L., Zheng, P., & Zhu, X. (2021). Grassland soil moisture fluctuation and its relationship with evapotranspiration. *Ecological Indicators*, 131, 108196. <https://doi.org/10.1016/j.ecolind.2021.108196>.
- Wilding, L.G. (1985, 30 Nov–1 Dec). *Soil spatial variability: Its documentation, accommodation and implication to soil surveys*. Paper presented at the Proceedings of a workshop of the ISSS and the SSSA, Las Vegas, USA.
- Wu, L., Peng, Y., Fan, J., & Wang, Y. (2019). Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data. *Hydrology Research*, 50(6), 1730–1750. <https://doi.org/10.2166/nh.2019.060>.
- Wu, M., Feng, Q., Wen, X., Deo, R., Yin, Z., Yang, L., & Sheng, D. (2020). Random forest predictive model development with uncertainty analysis capability for the estimation of evapotranspiration in an arid oasis region. *Hydrology Research*, 51(4), 648–665. <https://doi.org/10.2166/nh.2020.012>.
- Zandler, H., Haag, I., & Samimi, C. (2019). Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions. *Scientific Reports*, 9(1), 1-15. <https://doi.org/10.1038/s41598-019-51666-z>.

Using Machine Learning Method to Estimate Evapotranspiration (Case Study: Semnan Province)

EXTENDED ABSTRACT

Background and purpose:

Evapotranspiration (ET) plays a crucial role in the hydrological cycle and water management within the agricultural sector, aiding in the estimation of necessary water resources and understanding the impacts of climate change. Additionally, precise estimation of evapotranspiration is vital for effective irrigation planning and management strategies. The objective of this research is to develop a precise annual ET map of Semnan province using the RF model. Furthermore, the study aims to explore the relationship between predicted data and observational data to assess the significance of auxiliary variables in evapotranspiration estimation.

Materials and methods:

Semnan province, spanning 96,816 km², faces challenges in estimating evapotranspiration (ET) due to a limited number of synoptic and evaporation stations. To address this issue, an alternative data-driven model with minimal data requirements was employed. During the statistical period of 1993-2020, 6 synoptic stations and 5 evaporation stations were utilized for evapotranspiration estimation. Initially, the study utilized the evapotranspiration product from the ERA5-Land reanalysis dataset, along with auxiliary variables such as Digital Elevation Model (DEM) and temperature, to estimate evapotranspiration in the study area. Additionally, the Random Forest (RF) method was employed to establish the connection between auxiliary variables and evapotranspiration data, leading to the creation of an evapotranspiration map using the RF model. Subsequently, the model's accuracy in estimating ET was evaluated by comparing it to observational data using four statistical criteria: Pearson correlation coefficient (R^2), root mean square error (RMSE), BIAS, and Kling-Gupta efficiency (KGE).

Findings:

The results of the validation phase demonstrated the high efficiency of the RF model, with values of $R^2 = 0.95$, BIAS = -4.1, RMSE = 98.6, and KGE = 0.92. Analysis of the ET data generated by the RF model revealed lower ET levels in the northern regions of the province. In the central to northern areas, average ET levels were observed, correlating with the region's average elevation. Conversely, the highest ET levels were recorded in most of the eastern and southern desert regions of the province. Furthermore, independent analysis of the monitored stations highlighted the RF model's reliance on topography. The elevation variable emerged as a crucial factor in systematically correcting model errors to enhance ET estimation accuracy. These findings underscored the effectiveness of auxiliary variables in a non-linear model like the RF model for accurately estimating complex and spatially variable characteristics such as ET. Notably, the model error (BIAS) of the ERA5-Land reanalysis dataset was rectified as a significant factor in determining model accuracy through the RF algorithm. Unlike the RF model, ERA5-Land employs spatial interpolation, enhancing predictions by incorporating covariates that directly impact ET, such as topographic variables.

Conclusion:

The results indicated that using RF model with different number of input variables to prepare ET map and processing in R programming environment provides proper accuracy in ET estimation for arid and semi-arid regions.

Keywords: *Elevation, ERA5-Land, Random Forest Model (RF), Temperature*