

ارائه تکنیک داده‌های پانلی در برآورد تبخیر تعرق مرجع ماهانه

پویا شیرازی^۱، کامران داوری^{۲*}، بیژن قهرمان^۳ و مهدی جباری نوقابی^۴

تاریخ دریافت: ۱۳۹۷/۱۱/۶ تاریخ پذیرش: ۱۳۹۷/۱۱/۲۷

چکیده

یکی از مهمترین اجزای چرخه هیدرولوژیکی، تبخیر-تعرق می‌باشد و برآورد آن برای بسیاری از مطالعات نظیر مدیریت منابع آب، شبیه‌سازی عملکرد محصول و برنامه‌ریزی آبیاری حائز اهمیت است. هدف از این پژوهش، برآورد تبخیر-تعرق مرجع با استفاده از روش داده‌های پانلی است. روش داده‌های پانلی یک تحلیل چند متغیره بوده که در آن تحلیل رگرسیون در هر دو بعد مکانی و زمانی صورت می‌گیرد. این مطالعه با استفاده از داده‌های هواشناسی ۹ ایستگاه سینوپتیک خراسان رضوی در بازه زمانی ۲۰۱۰-۱۹۷۱ انجام شد. داده‌های موجود به صورت کاملاً تصادفی به دو مجموعه تقسیم شدند، ۷۵ درصد برای توسعه مدل و ۲۵ درصد باقیمانده برای ارزیابی مدل اختصاص یافت. برای توسعه مدل میانگین ماهیانه دمای هوا و میانگین ماهانه سرعت باد به عنوان متغیرهای مستقل و مجموع ماهانه تبخیر-تعرق مرجع به عنوان متغیر مستقل در نظر گرفته شد. نتایج نشان داد که تصریح اثرهای ثابت دوطرفه از سایر تصریح‌ها مناسبتر است. شاخص‌های آماری (جذر میانگین مربعات خطای ۹/۸۵ میلی‌متر در ماه، میانگین قدرمطلق خطا ۷/۳۸ میلی‌متر در ماه و ضریب تعیین ۰/۹۹) کارآمد بودن این مدل را آشکار ساخت. همچنین، در مقایسه مدل داده‌های پانلی با مدل رگرسیون حداقل مربعات معمولی و روش هارگریوز-سامانی، مدل فوق دارای توانایی بالاتری در برآورد تبخیر-تعرق مرجع می‌باشد.

واژه‌های کلیدی: پنمن-مانتیس فائو، تصریح اثرهای ثابت، روش داده‌های پانلی

مقدمه

پارامترهای مورد نیاز از یک سو و عدم اندازه‌گیری برخی از پارامترها از سوی دیگر است، که برآورد تبخیر-تعرق مرجع را در برخی از مناطق با مشکل مواجه کرده است (فرسادنیا و همکاران، ۱۳۹۴). بر خلاف روش‌های نظری، اطلاعات مورد نیاز در روش‌های تجربی معمولاً زیر مجموعه‌ای از پارامترهای هواشناسی (معمولاً دما، باد) می‌باشند. استفاده از روش‌های تجربی ساده است، اما ماهیت تجربی این روش‌ها باعث شده نتایج حاصله مختص به مکان و شرایط اقلیمی باشد که معادله تجربی در آن توسعه یافته است. از روش‌های تجربی متعددی برای برآورد تبخیر-تعرق مرجع استفاده شده است، از جمله می‌توان به روش شبکه عصبی مصنوعی (شرفان و قربانی، ۱۳۹۳؛ Yassin et al., 2016; Slavisa et al., 2003) ، منطق فازی (Lin et al., 2014; Shiri et al., 2013) ، الگوریتم ژنتیک (Eslamian et al., 2012; Ozgur and Aytac, 2010) و رگرسیون چند متغیره (Karbasi, 2018; Feng et al., 2017; Kisi, 2011) اشاره کرد.

اصطلاح «داده‌های پانلی» به پشته کردن داده‌های مشاهده‌ای مقاطع مختلف در طول دوره آماری آن‌ها اشاره دارد. به عبارت دیگر، به مجموعه‌ای از داده‌ها گفته می‌شود که براساس آن، مشاهدات به وسیله تعداد زیادی از متغیرهای مقطعی، در طول یک دوره زمانی مشخص مورد بررسی قرار گرفته باشند. بنابراین، داده‌های پانلی در بر

یکی از مهمترین اجزای چرخه هیدرولوژیکی، تبخیر-تعرق می‌باشد و برآورد آن برای بسیاری از مطالعات نظیر بیان هیدرولوژیکی آب، طراحی و مدیریت سیستم‌های آبیاری، برنامه‌ریزی و مدیریت منابع آب و غیره حائز اهمیت است (وزیری و همکاران، ۱۳۸۷). روش‌های متعددی جهت برآورد مستقیم و غیر مستقیم تبخیر-تعرق مرجع ارائه شده است. روش‌های مستقیم (مانند روش لایسیمتر) نیازمند صرف وقت، هزینه و دقت بالا می‌باشند از این‌رو تمرکز بیشتر بر روی تکنیک‌های غیر مستقیم در برآورد تبخیر-تعرق مرجع می‌باشد. روش‌های غیر مستقیم به دو دسته نظری و تجربی تقسیم می‌شوند. از روش‌های نظری می‌توان به روش پنمن-مانتیس فائو اشاره کرد که بر اساس فرآیندهای فیزیکی بنا نهاده شده است (وزیری و همکاران، ۱۳۸۷). مشکل اساسی روش‌های نظری، تعدد

۱- دانشجوی دکتری گروه مهندسی علوم آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد

۲- استاد گروه مهندسی علوم آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد

۳- استاد گروه مهندسی علوم آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد

۴- استادیار گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

(Email: k.davary@um.ac.ir)

*- نویسنده مسئول:

پژوهشی در استان خراسان رضوی به پیش‌بینی درجه حرارت پروفیل خاک با استفاده از مدل‌های داده‌های پانلی پرداختند. داده‌های مورد نیاز این تحقیق (میانگین روزانه دمای خاک) طی ۹ سال (۲۰۰۱ تا ۲۰۰۹) در ۶ عمق مختلف خاک (۵، ۱۰، ۲۰، ۳۰، ۵۰ و ۱۰۰ سانتیمتر) و از ۱۰ ایستگاه هواشناسی انتخاب گردید. در این تحقیق میانگین دمای هوا روزانه، بارندگی روز قبل و میانگین دمای هوای ۷ روز گذشته به عنوان پارامتر ورودی به منظور پیش‌بینی میانگین دمای خاک روز بعد در مدل‌های داده‌های پانلی استفاده شدند. نتایج تحقیق نشان داد که مدل‌های داده‌های پانلی در مقایسه با نتایج مدل‌های رگرسیون خطی کلاسیک، از شاخص‌های عملکردی بهتری برخوردار بودند (Mahabbati et al., 2017).

با بررسی پژوهش‌های انجام شده مشخص شد اکثر روش‌های تجربی، در برآورد تبخیر-تعرق مرجع به صورت تک ایستگاهی عمل می‌نمایند. از آنجا که پارامترهای تبخیر-تعرق مرجع در بعد مکان (ایستگاه‌های مختلف) و بعد زمان (طول دوره آماری) ثابت نیستند، برآورد تبخیر-تعرق مرجع براساس تنها یکی از این دو بعد، منطقی بنظر نمی‌رسد. در این راستا، به منظور ارائه تکنیکی جدید، هدف اصلی این پژوهش بررسی قابلیت روش داده‌های پانلی در برآورد تبخیر-تعرق مرجع می‌باشد.

مواد و روش‌ها

مدل رگرسیون داده‌های پانلی

مدل رگرسیون داده‌های پانلی یک تحلیل چند متغیره بوده که در آن تحلیل رگرسیون در هر دو بعد مکانی و زمانی صورت می‌گیرد. از جمله مزایای این روش نسبت به مدل‌هایی که تنها بعد مکانی یا زمانی را در نظر می‌گیرند می‌توان به موارد زیر اشاره کرد: ۱- تعداد مشاهدات و داده‌ها بیشتر بوده در نتیجه باعث افزایش درجه آزادی و کاهش همخطی می‌شود، ۲- با این روش می‌توان اثراتی را شناسایی و اندازه‌گیری کرد که در مدل‌هایی که تنها بعد مکانی یا زمانی را در نظر می‌گیرند قابل شناسایی نیست. مدل رگرسیون داده‌های پانلی در حالت کلی به صورت زیر بیان می‌شود (Pesaran, 2015; Baltagi, 2008; Hsiao, 2003):

$$y_{it} = \alpha + \beta x_{it} + u_{it} \quad \begin{matrix} i = 1, 2, \dots, N \\ t = 1, 2, \dots, T \end{matrix} \quad (1)$$

که در آن i : بعد مکان، t : بعد زمان، N : تعداد مقاطع، T : طول سری زمانی هر یک از مقاطع، y : بردار متغیر وابسته، x : ماتریس متغیرهای مستقل، α : عرض از مبدا، β : ماتریس ضرایب متغیرهای مستقل و u : جزء خطای مدل می‌باشد. جزء خطای مدل را می‌توان به صورت زیر بیان کرد:

$$u_{it} = \mu_i + \lambda_t + v_{it} \quad (2)$$

که در آن μ_i : اثر مکانی غیر قابل مشاهده، λ_t : اثر زمانی غیر قابل

گیرنده هر دو جنبه داده‌های سری زمانی و سری مقطعی می‌باشد. بررسی منابع نشان می‌دهد که با وجود استفاده گسترده از روش داده‌های پانلی در علم اقتصاد (Baltagi, 2008; Lee and Yu, 2010; Suzuki, 2017; Maurseth, 2018) کاربرد این روش در حوضه علوم کشاورزی و محیطی بسیار جوان است. ایزدی و همکاران پژوهشی را با هدف پیش‌بینی سطح آب‌های زیرزمینی در دشت نیشابور با استفاده از مدل داده‌های پانلی انجام دادند. برای انجام این پژوهش در ابتدا، چاه‌های مشاهده‌ای موجود در دشت نیشابور، با توجه به رفتار نوسان آن‌ها و با استفاده از روش Ward خوشه بندی شدند که منجر به تشکیل شش منطقه گردید. سپس برای هر منطقه، یک خوشه مشاهدات به عنوان نماینده آن انتخاب شد و برای هر منطقه، مقادیر بارش و دمای ماهانه به عنوان متغیرهای مستقل با روش رگرسیون گام به گام تعیین شد. در نهایت، عملکرد تصریح‌های مختلف مدل داده‌های پانلی مانند اثرات ثابت و اثرات تصادفی مورد بررسی قرار گرفتند. نتایج نشان داد که تصریح اثرات ثابت دو طرفه، دارای برتری بهتری نسبت به سایر تصریح‌ها بوده است. علاوه بر این، نتایج حاصل از مدل داده‌های پانلی با نتایج یک مدل شبکه عصبی مصنوعی مقایسه گردید و در نهایت برتری مدل داده‌های پانلی بر مدل شبکه عصبی مصنوعی اثبات گردید (Izady et al., 2012). کواکبی و همکاران (۱۳۹۳) در پژوهشی عوامل موثر بر وقوع خشکسالی را با استفاده از روش داده‌های پانلی در استان خراسان رضوی تعیین کردند. آن‌ها ابتدا به پایش سالانه و فصلی خشکسالی ۱۰ ایستگاه سینوپتیک استان خراسان رضوی در طی یک دوره ۲۴ ساله با دو شاخص SPI و RDI پرداختند. سپس پارامترهای موثر بر خشکسالی سالانه و فصلی با روش داده‌های پانلی و شاخص‌های SPI و RDI تعیین شدند. تحلیل به روش داده‌های پانلی نشان داد که تمام شش پارامتر متوسط بیشینه و کمینه دما، ساعات آفتابی، درصد رطوبت نسبی و متوسط سرعت باد در ارتفاع ۲ متر که برای محاسبه شاخص RDI مورد نیاز است، برای محاسبه خشکسالی سالانه و فصلی ۱۰ ایستگاه مورد مطالعه نیاز نیست و تنها برخی از این پارامترها بسته به مقیاس زمانی مورد نظر نیاز است. پارامتر متوسط دمای کمینه در مقیاس‌های زمانی سالانه و فصل تابستان، پارامتر درصد رطوبت نسبی در مقیاس‌های زمانی سالانه، فصل بهار و زمستان به عنوان پارامترهای موثر پیش‌بینی انتخاب شدند. همچنین بارش نیز تنها عامل موثر بر خشکسالی با توجه به شاخص‌های RDI و SPI در مقیاس‌های زمانی سالانه و فصلی می‌باشد (کواکبی و همکاران، ۱۳۹۳). هادی قنوت و همکاران (۱۳۹۳) کاربرد مدل داده‌های پانلی را در پیش‌بینی تغییرات کلر آب زیرزمینی بررسی کردند. نتایج آن‌ها نشان داد که تصریح اثرهای تصادفی یک طرفه بهترین عملکرد را در میان تصریح‌های مدل داده‌های پانلی در پیش‌بینی کلر آب زیرزمینی داشته است (هادی قنوت و همکاران، ۱۳۹۴). محبتی و همکاران در

مشاهده و u_{it} : باقیمانده‌ها می‌باشد. ماهیت و شرایط μ_i و λ_i موجب تقسیم‌بندی مدل رگرسیون داده‌های پانلی به سه روش می‌شود (مهرگان و اشرف‌زاده، ۱۳۸۷؛ Baltagi, 2008):

تصریح اثرهای مشترک: در صورتی که μ_i و λ_i فقط شامل یک جمله ثابت باشند، مدل به صورت اثرهای مشترک در می‌آید که به آن مدل داده‌های تلفیقی نیز گفته می‌شود. این مدل با روش رگرسیون حداقل مربعات عادی قابل برآورد است.

$$ET_0 = 0.0023 \frac{R_a}{\lambda} \left(\frac{T_{\max} + T_{\min}}{2} + 17.8 \right) \sqrt{T_{\max} - T_{\min}} \quad (4)$$

که در آن ET_0 : میزان تبخیر-تعرق مرجع (میلی‌متر در روز)، T_{\max} : دمای بیشینه هوا در ارتفاع دو متری (درجه سانتی‌گراد)، T_{\min} : دمای کمینه هوا در ارتفاع دو متری (درجه سانتی‌گراد)، R_a : تابش فرا زمینی (مگاژول بر متر مربع در روز) و γ : ثابت سایکرومتری (کیلو پاسکال بر درجه سانتی‌گراد) می‌باشد.

تصریح اثرهای ثابت: اگر μ_i و λ_i قابل مشاهده نباشند اما با x_{it} همبستگی داشته باشند، به صورت یک عرض از مبدا ثابت برای هر مقطع و هر زمان در می‌آیند. اصطلاح ثابت به آن معنا است که عرض از مبدا برای هر مقطع در طی زمان و عرض از مبدا برای هر زمان در طول مقاطع تغییر نمی‌کنند. این مدل را با روش رگرسیون حداقل مربعات با متغیرهای مجازی برآورد می‌کنند.

منطقه مورد مطالعه و داده‌ها

استان خراسان رضوی با مساحتی بالغ بر ۱۲۷۶۰۰ کیلومتر مربع در شمال شرقی ایران قرار گرفته و حدود ۷/۷ درصد از مساحت کشور را شامل می‌شود. همانطور که شکل ۱ نشان داده شده است، این استان بین ۳۳/۵ تا ۳۸ درجه عرض شمالی و ۵۶ تا ۶۱ درجه طول شمالی قرار دارد.

تصریح اثرهای تصادفی: اگر μ_i و λ_i با x_{it} همبستگی نداشته باشند و به صورت یک عامل تصادفی برای هر مقطع و زمان بروز کنند، آن را مدل اثرهای تصادفی گویند. در واقع مدل تصادفی راهی است برای برخورد با این واقعیت که T مشاهده روی N مکان، شبیه مشاهدات روی NT واحد مختلف نیست. این مدل با روش حداقل مربعات تعمیم یافته برآورد می‌شود.

در این پژوهش از داده‌های میانگین ماهیانه دمای هوا، دمای کمینه هوا، دمای بیشینه هوا، رطوبت نسبی، سرعت باد و مجموع ماهیانه بارندگی و ساعات آفتابی ۹ ایستگاه هواشناسی سینوپتیک استان خراسان رضوی، ایستگاه‌های تربت جام، تربت حیدریه، درگز، سبزوار، سرخس، کاشمر، گلکان، مشهد و نیشابور با طبقه‌بندی اقلیمی کوپن-گایگر (BSK) (اقلیم سرد و نیمه خشک با تابستان‌های گرم و زمستان‌های سرد که میزان بارندگی کمتر از تبخیر-تعرق پتانسیل می‌باشد)، استفاده شده است. در جدول ۱ مشخصات جغرافیایی و اقلیمی ایستگاه‌های سینوپتیک مورد مطالعه ارائه شده است.

اگر تعداد مشاهدات T در همه ایستگاه‌ها با یکدیگر برابر باشد، مدل داده‌های پانلی از نوع متوازن است، و اگر تعداد مشاهدات T در هر ایستگاه متفاوت باشد، مدل داده‌های پانلی نامتوازن می‌باشد (Baltagi, 2008).

روش‌های محاسبه تبخیر-تعرق مرجع

در این پژوهش تبخیر-تعرق مرجع بر اساس روش‌های پنمن-مانتیس فائو (Allen et al., 1998) و هارگریوز-سامانی (Samani, 2000) برای همه ایستگاه‌ها محاسبه گردید و با توجه به توصیه کمیسیون بین‌المللی آبیاری و زهکشی (ICID) و سازمان خواربار و کشاورزی ملل متحد (FAO) روش پنمن-مانتیس فائو به عنوان روش مرجع برای ارزیابی مدل در نظر گرفته شد (وزیری و همکاران، ۱۳۸۷). معادله روش پنمن-مانتیس فائو برای محاسبه تبخیر-تعرق مرجع به صورت زیر بیان شده است:

کل داده‌های موجود به صورت کاملاً تصادفی به دو مجموعه تقسیم شدند، ۷۵ درصد برای توسعه مدل و ۲۵ درصد باقیمانده برای ارزیابی مدل اختصاص یافت. تمام محاسبات با استفاده از زبان برنامه نویسی R صورت گرفته است (Croissant and Millo, 2018).

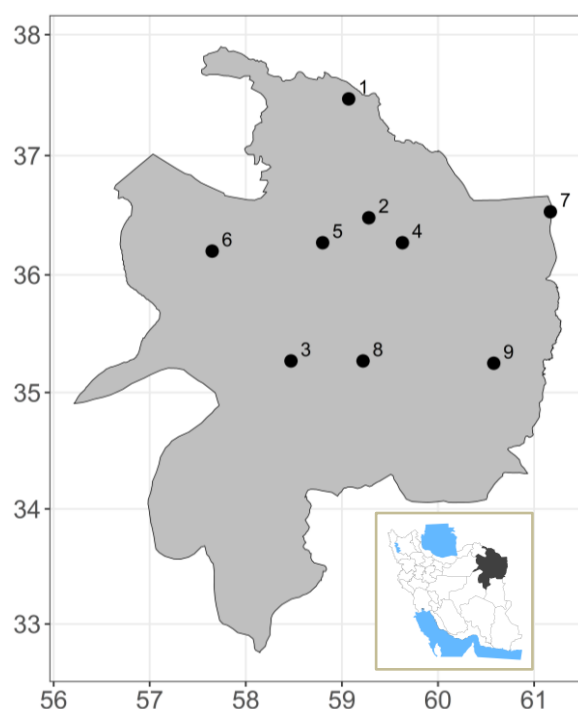
$$ET_0 = \frac{900}{\Delta + \gamma(1 + 0.34u_2)} \left(0.408\Delta(R_n - G) + \gamma \frac{R_n}{T + 273} u_2 (e_s - e_a) \right) \quad (3)$$

ارزیابی خطا و شاخص‌های آماری

کارایی مدل در مراحل توسعه و ارزیابی با استفاده از شاخص‌های آماری زیر مورد بررسی قرار گرفت:

خطای جذر میانگین مربعات (RMSE)، تفاوت میان مقدار پیش‌بینی شده و مشاهده شده می‌باشد. رنج تغییرات خطای جذر میانگین مربعات بین صفر تا مثبت بینهایت بوده و مقادیر نزدیک به صفر نشان دهنده کارایی بهتر مدل است.

که در آن ET_0 : میزان تبخیر-تعرق مرجع (میلی‌متر در روز)، R_n : تابش خالص در سطح پوشش گیاهی (مگاژول بر متر مربع در روز)، G : شار حرارتی خاک (مگاژول بر متر مربع در روز)، T : میانگین دمای هوا در ارتفاع دو متری (درجه سانتی‌گراد)، u_2 : سرعت باد در ارتفاع دو متری (متر بر ثانیه)، e_s : فشار بخار اشباع (کیلو پاسکال)، e_a : فشار بخار واقعی (کیلو پاسکال)، Δ : شیب منحنی فشار بخار آب اشباع



شکل ۱- موقعیت جغرافیایی استان خراسان رضوی. دایره‌های مشکی پررنگ به همراه شماره آن، نشان دهنده موقعیت ۹ ایستگاه سینوپتیک منتخب می‌باشند.

جدول ۱- مشخصات اقلیمی و جغرافیایی ایستگاه‌های مورد مطالعه

| ایستگاه | طول جغرافیایی | عرض جغرافیایی | ارتفاع از سطح دریا (متر) | دوره آماری | متوسط سالانه دمای هوا (درجه سانتی‌گراد) | متوسط سالانه سرعت باد (متر بر ثانیه) | متوسط سالانه بارندگی (میلی‌متر) | طبقه‌بندی اقلیمی کوپن-گایگر |
|-------------|---------------|---------------|--------------------------|------------|---|--------------------------------------|---------------------------------|-----------------------------|
| تربت جام | ۶۰/۵۸ | ۳۵/۲۵ | ۹۵۰/۴ | ۲۰۱۰-۱۹۹۳ | ۱۵/۷ | ۴/۱ | ۱۷۲/۸ | Bsk |
| تربت حیدریه | ۵۹/۲۲ | ۳۵/۲۷ | ۱۴۵۱ | ۲۰۱۰-۱۹۷۱ | ۱۴/۳ | ۲/۱ | ۲۷۴/۴ | Bsk |
| درگز | ۵۹/۰۷ | ۳۷/۴۷ | ۵۱۴ | ۲۰۰۹-۲۰۰۸ | ۱۶/۵ | ۲/۹ | ۱۷۷/۷ | Bsk |
| سبزوار | ۵۷/۶۵ | ۳۶/۲۰ | ۹۷۲ | ۲۰۱۰-۱۹۷۱ | ۱۷/۹ | ۲/۸ | ۲۰۰/۴ | Bsk |
| سرخس | ۶۱/۱۷ | ۳۶/۵۳ | ۲۳۵ | ۲۰۱۰-۱۹۸۴ | ۱۸/۱ | ۲/۲ | ۱۸۲/۹ | Bsk |
| کاشمر | ۵۸/۴۷ | ۳۵/۲۰ | ۱۱۱۰ | ۲۰۱۰-۱۹۸۶ | ۱۷/۹ | ۱/۵ | ۱۹۴/۸ | Bsk |
| گلمکان | ۵۹/۲۸ | ۳۶/۴۸ | ۱۱۷۶ | ۲۰۱۰-۱۹۸۷ | ۱۳/۳ | ۳/۲ | ۲۰۹/۶ | Bsk |
| مشهد | ۵۹/۶۳ | ۳۶/۲۷ | ۹۹۹/۲ | ۲۰۱۰-۱۹۷۱ | ۱۴/۶ | ۲/۰ | ۲۶۰/۴ | Bsk |
| نیشابور | ۵۸/۸۰ | ۳۶/۲۷ | ۱۲۱۳ | ۲۰۱۰-۱۹۹۱ | ۱۴/۴ | ۱/۳ | ۲۳۸/۳ | Bsk |

کوچک نشان دهنده کارایی بهتر مدل است.

$$RSR = \frac{\sqrt{\sum_{i=1}^N (O_i - S_i)^2}}{\sqrt{\sum_{i=1}^N (O_i - \bar{S})^2}} \quad (۶)$$

میانگین قدرمطلق خطا (MAE)، یک شاخص آماری برای نشان

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - O_i)^2} \quad (۵)$$

نسبت خطای جذر میانگین مربعات نسبی به انحراف از معیار داده‌های مشاهداتی (RSR)، یک شاخص آماری نرمال شده می‌باشد. رنج تغییرات خطای نسبی بین صفر تا مثبت بینهایت بوده و مقادیر

تغرق مرجع محاسبه شده با روش پنمن-مانتیس فائو دارند، استفاده گردد. به این منظور ماتریس همبستگی بین پارامترهای هواشناسی و تبخیر-تغرق مرجع روش پنمن-مانتیس فائو محاسبه شد (شکل ۲). با توجه به همبستگی بالای میانگین ماهیانه دمای کمینه هوا، میانگین ماهیانه دمای بیشینه هوا و میانگین ماهیانه دمای هوا با تبخیر-تغرق مرجع، باید متغیر دما انتخاب گردد، ولی به دلیل جلوگیری از همخطی چندگانه متغیرهای مستقل، تنها میانگین ماهیانه دمای هوا انتخاب گردید. پارامتر مجموع ماهیانه ساعات آفتابی همبستگی بالایی با تبخیر-تغرق مرجع دارد، ولی به دلیل اینکه این پارامتر دارای بیشترین داده گمشده (حدود ۲۰ درصد) در بین ایستگاه‌های مورد بررسی می‌باشد، انتخاب نمی‌گردد. ۲- متغیرهایی انتخاب گردند که جزء ساده‌ترین داده‌های هواشناسی باشند و در تمام ایستگاه‌های هواشناسی اندازه‌گیری گردند. از بین پارامترهای میانگین ماهیانه رطوبت نسبی، میانگین ماهیانه سرعت باد و مجموع ماهیانه بارندگی، به دلیل در دسترس بودن بیشتر، میانگین ماهیانه سرعت باد انتخاب گردید.

دادن فاصله بین مقادیر مشاهده شده و پیش‌بینی شده است با این فرض که جهت این تفاضل مد نظر نمی‌باشد. رنج تغییرات میانگین قدرمطلق خطا بین صفر تا مثبت بینهایت بوده و مقادیر نزدیک به صفر نشان دهنده کارایی بهتر مدل است.

$$MAE = \frac{1}{N} \sum_{i=1}^N |S_i - O_i| \quad (7)$$

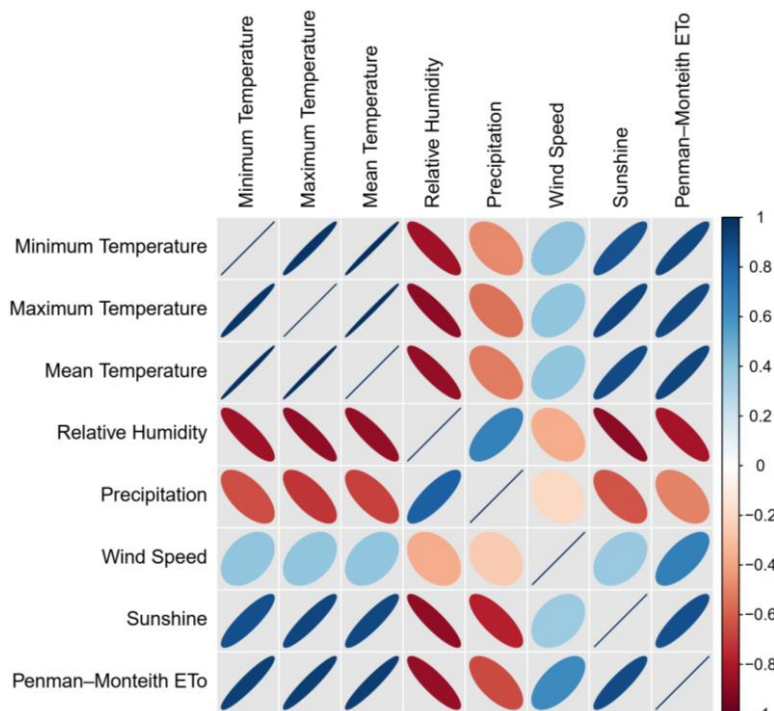
ضریب تعیین (R^2)، بیانگر میزان احتمال همبستگی میان مقادیر مشاهده شده و پیش‌بینی شده می‌باشد. رنج تغییرات این شاخص بین صفر تا یک بوده و مقادیر نزدیک به یک نشان دهنده کارایی بهتر مدل است.

که در آن‌ها N: تعداد مشاهدات، O_i : داده‌های مشاهده شده و S_i : داده‌های شبیه‌سازی شده می‌باشد.

نتایج و بحث

انتخاب ورودی‌های مدل

به منظور تعیین ورودی‌های مدل داده‌های پانلی، دو شرط در نظر گرفته شده است. ۱- از متغیرهایی که حداکثر همبستگی را با تبخیر-



شکل ۲- ماتریس ضرایب همبستگی بین پارامترهای هواشناسی و تبخیر-تغرق مرجع

پارامترها برای پیش‌بینی تبخیر-تغرق مرجع در بسیاری از مراجع گزارش شده است (Kisi, 2011; Eslamian et al., 2012; Yassin et al., 2016). این رابطه را می‌توان در مدل داده‌های پانلی به صورت زیر بیان کرد:

توسعه و ارزیابی مدل

برای توسعه مدل میانگین ماهیانه دمای هوا و میانگین ماهانه سرعت باد به عنوان متغیرهای مستقل و مجموع ماهانه تبخیر-تغرق مرجع به عنوان متغیر مستقل در نظر گرفته شد. استفاده از این

قدرمطلق خطا و جذر میانگین مربعات خطا (به ترتیب MAE: ۲/۰۳ و ۳/۱۸۶، RMSE: ۳/۶۳ و ۵/۰۵) و ایستگاه‌های مشهد و تربت جام دارای بیشترین میانگین قدرمطلق خطا و جذر میانگین مربعات خطا (به ترتیب MAE: ۶/۶۲ و ۶/۱۶، RMSE: ۸/۱۴ و ۸/۰۷) می‌باشند. همچنین ایستگاه درگز دارای کمترین نسبت خطای جذر میانگین مربعات نسبی به انحراف از معیار داده‌های مشاهداتی (RSR: ۰/۰۴) و ایستگاه‌های مشهد و گلکان دارای بیشترین نسبت خطای جذر میانگین مربعات نسبی به انحراف از معیار داده‌های مشاهداتی (به ترتیب RSR: ۰/۱ و ۰/۱) می‌باشند. در مرحله ارزیابی مدل داده‌های پانلی تبخیر-تعرق مرجع ماهانه نیز، ایستگاه‌های تربت حیدریه و نیشابور دارای کمترین میانگین قدرمطلق خطا و جذر میانگین مربعات خطا (به ترتیب MAE: ۵/۱۵ و ۵/۲، RMSE: ۶/۶۳ و ۷/۳۲) و ایستگاه‌های درگز و سبزوار دارای بیشترین میانگین قدرمطلق خطا و جذر میانگین مربعات خطا (به ترتیب MAE: ۱۰/۵۳ و ۸/۹۱، RMSE: ۱۳/۳۷ و ۱۲/۴۶) می‌باشند. همچنین ایستگاه تربت حیدریه دارای کمترین نسبت خطای جذر میانگین مربعات نسبی به انحراف از معیار داده‌های مشاهداتی (RSR: ۰/۰۷) و ایستگاه‌های درگز و سرخس دارای بیشترین نسبت خطای جذر میانگین مربعات نسبی به انحراف از معیار داده‌های مشاهداتی (به ترتیب RSR: ۰/۱۲ و ۰/۱۲) می‌باشند. در مجموع برای تمامی ایستگاه‌ها، در مرحله توسعه، مدل داده‌های پانلی دارای میانگین قدرمطلق خطا ۵/۴۶ میلی‌متر در ماه و جذر میانگین مربعات خطا ۷/۲۵ میلی‌متر در ماه و در مرحله ارزیابی، مدل داده‌های پانلی دارای میانگین قدرمطلق خطا ۷/۳۸ میلی‌متر در ماه و جذر میانگین مربعات خطا ۹/۸۵ میلی‌متر در ماه می‌باشد که نشان دهنده عملکرد بسیار خوب مدل در برآورد تبخیر-تعرق ماهانه می‌باشد.

مقایسه نتایج مدل داده‌های پانلی با روش هارگریوز-سامانی

نتایج نشان داد که مدل داده‌های پانلی - اثرهای ثابت دو طرفه توانایی لازم برای برآورد تبخیر و تعرق مرجع با حداقل خطا را دارد. در این بخش کارایی این مدل در مقایسه با روش رگرسیون حداقل مربعات معمولی و روش هارگریوز-سامانی مورد بررسی قرار می‌گیرد. در شکل ۳ نمودار تیلور به منظور بررسی و تحلیل مقادیر انحراف معیار، ضریب تعیین و جذر میانگین مربعات خطا بین داده‌های محاسبه شده توسط روش پنمن-مانتیس فائو (به عنوان مرجع) و داده‌های شبیه‌سازی شده توسط مدل داده‌های پانلی، مدل رگرسیون حداقل مربعات معمولی و روش هارگریوز-سامانی رسم گردیده است. لازم به ذکر است که در نمودار تیلور فاصله طولی از مبدا مختصات نشان دهنده انحراف معیار، خط چین‌های شعاعی نشان دهنده ضریب تعیین و خط چین‌های قطاعی نشان دهنده مقادیر جذر میانگین

$$ET_{it} = \alpha + \beta_1 T_{it} + \beta_2 W_{it} + \mu_i + \lambda_t + u_{it} \quad (7)$$

که در آن ET_{it} ، تبخیر-تعرق مرجع (میلی‌متر در ماه) برای ایستگاه i و زمان t ، α ، عرض از مبدا کلی (میلی‌متر در ماه)، β_1 و β_2 ، ضریب متغیرهای مستقل دمای هوا و سرعت باد، μ_i و λ_t ، اثرات مکانی و زمانی غیر قابل مشاهده، u_{it} ، باقیمانده‌ها، T_{it} و W_{it} ، به ترتیب میانگین ماهیانه دمای هوا (درجه سانتی‌گراد) و میانگین ماهانه سرعت باد (متر بر ثانیه) برای ایستگاه i و زمان t می‌باشد. علاوه بر این، از یک مدل رگرسیون حداقل مربعات معمولی نیز برای برآورد تبخیر-تعرق مرجع استفاده شد. در این مدل رابطه متغیرهای مستقل و متغیر وابسته به صورت زیر می‌باشد:

$$ET = a + b_1 T + b_2 W + u \quad (8)$$

که در آن a ، عرض از مبدا، b_1 و b_2 ، ضریب متغیرهای مستقل میانگین ماهیانه دمای هوا و میانگین ماهانه سرعت باد و u ، باقیمانده‌ها می‌باشد.

در مرحله بعد، تصریح‌های اثرهای مشترک، اثرهای ثابت دوطرفه و اثرهای تصادفی دوطرفه توسعه داده شدند. برای انتخاب بهترین تصریح از آزمون‌های چاو و هاسمن استفاده شد و نتایج آن در جدول ۲ نشان داده شده است. در ابتدا باید مشخص شود که آیا μ_i و λ_t به صورت یک عرض از مبدا ثابت برای هر مقطع و هر زمان در می‌آیند (تصریح اثرهای ثابت) یا μ_i و λ_t فقط شامل یک جمله ثابت می‌باشند (تصریح اثرهای مشترک). برای اینکار آزمون چاو به کار برده شده است. نتیجه آزمون چاو بیانگر رد شدن فرضیه صفر و وجود ناهماهنگی در ایستگاه‌های مورد مطالعه در سطح معنی‌داری ۱ درصد است که بیانگر خوب بودن روش داده‌های پانلی برای برآورد مدل است. آزمون هاسمن نیز برای تشخیص استفاده از تصریح اثرهای ثابت یا اثرهای تصادفی به کار برده می‌شود. با توجه به نتیجه این آزمون، فرضیه استفاده از تصریح اثرهای تصادفی پذیرفته نمی‌شود و باید تصریح اثرهای ثابت به کار برده شود. بنابراین، در این پژوهش تصریح اثرهای ثابت دوطرفه به عنوان مدل مناسب انتخاب گردید.

جدول ۲- مقادیر محاسبه شده آزمون‌های چاو و هاسمن

| آزمون | مقدار محاسبه شده | p-value | فرض صفر |
|-------|------------------|---------|----------------|
| چاو | ۱۳/۱۴ | ۰/۰۰۱ | روش اثر مشترک |
| هاسمن | ۴۱/۰۷ | ۰/۰۰۱ | روش اثر تصادفی |

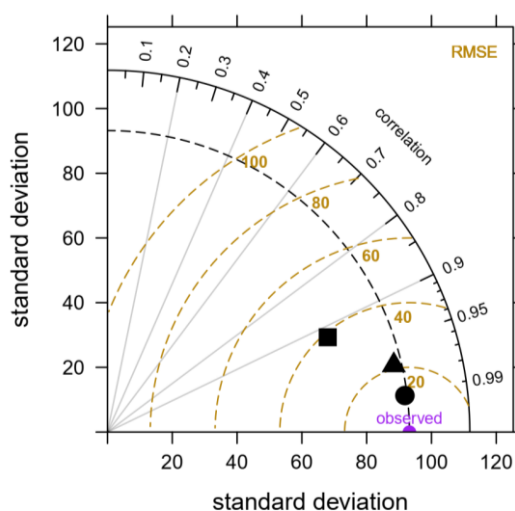
در جدول ۳ شاخص‌های آماری میانگین قدرمطلق خطا (MAE)، جذر میانگین مربعات خطا (RMSE) و نسبت خطای جذر میانگین مربعات نسبی به انحراف از معیار داده‌های مشاهداتی (RSR) برای همه ایستگاه‌ها در مرحله توسعه و ارزیابی مدل مورد مقایسه قرار گرفته است. در مرحله توسعه مدل داده‌های پانلی تبخیر-تعرق مرجع ماهانه، ایستگاه‌های درگز و تربت حیدریه دارای کمترین میانگین

میانگین مربعات خطا می‌باشد. همانطور که مشخص است، مدل داده‌های پانلی به خوبی توانسته است مقدار تبخیر-تعرق مرجع را، نسبت به دو روش دیگر، برآورد کند.

مربعات خطا می‌باشد که با بزرگ شدن قطاع دایره مقدار پارامتر مذکور بیشتر می‌شود. به عبارت دیگر هر نقطه بر روی نمودار تیلور نشان دهنده همزمان سه پارامتر انحراف معیار، ضریب تعیین و جذر

جدول ۳- خطاهای مدل داده‌های پانلی تبخیر-تعرق مرجع ماهانه در مراحل توسعه و ارزیابی مدل در ایستگاه‌های مختلف. واحد MAE و RMSE، ساعت می‌باشد.

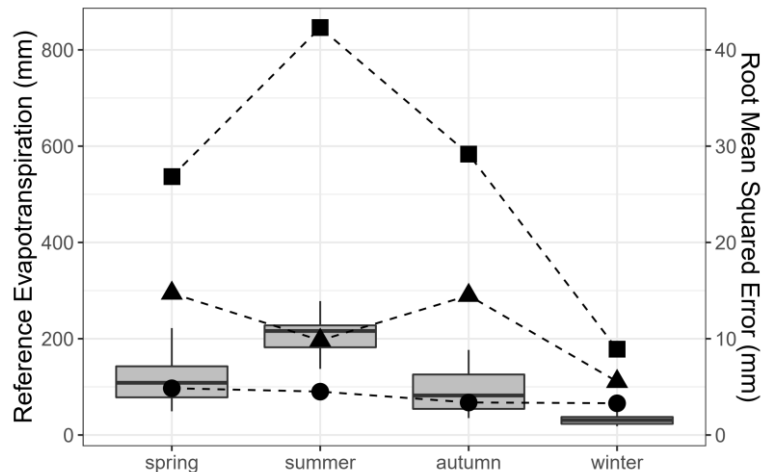
| ایستگاه | مرحله توسعه مدل | | | مرحله ارزیابی مدل | | |
|----------------|-----------------|------|------|-------------------|-------|-------|
| | RSR | RMSE | MAE | RSR | RMSE | MAE |
| تربت جام | 0/07 | 8/07 | 6/16 | 0/09 | 10/64 | 8/38 |
| تربت حیدریه | 0/06 | 5/05 | 3/86 | 0/07 | 6/63 | 5/15 |
| درگز | 0/04 | 3/63 | 2/03 | 0/12 | 13/34 | 10/53 |
| سبزوار | 0/07 | 7/17 | 5/59 | 0/11 | 12/46 | 8/91 |
| سرخس | 0/09 | 8/06 | 6/15 | 0/12 | 10/72 | 7/87 |
| کاشمر | 0/09 | 6/82 | 5/01 | 0/12 | 9/22 | 7/02 |
| گلمکان | 0/1 | 8/3 | 6/47 | 0/12 | 10/8 | 8/48 |
| مشهد | 0/1 | 8/14 | 6/62 | 0/11 | 10/12 | 8/15 |
| نیشابور | 0/07 | 5/16 | 4/03 | 0/1 | 7/32 | 5/2 |
| همه ایستگاه‌ها | 0/08 | 7/25 | 5/46 | 0/1 | 9/85 | 7/38 |



شکل ۳- نمودار تیلور مقایسه تبخیر-تعرق مرجع محاسبه شده با روش پنمن-مانتیس فائو (به عنوان مرجع) با روش مدل داده‌های پانلی (دایره مشکی)، مدل رگرسیون حداقل مربعات معمولی (مثلث مشکی) و روش هارگریوز-سامانی (مربع مشکی) در مرحله توسعه مدل.

RMSE های سه مدل در فصل زمستان کمترین و در فصل تابستان بیشترین می‌باشد. نتایج بیانگر عملکرد بسیار خوب مدل داده‌های پانلی است، که این توانایی می‌تواند ناشی از در نظر گرفتن همزمان اثر مکانی و اثر زمانی در مدل باشد.

در شکل ۴ نمودار جعبه‌ای مقادیر تبخیر-تعرق مرجع ماهانه محاسبه شده با روش پنمن-مانتیس فائو در ایستگاه نیشابور و RMSE مدل‌های داده‌های پانلی، روش رگرسیون حداقل مربعات معمولی و روش هارگریوز-سامانی برای هر فصل در مرحله توسعه مدل نشان داده شده است. همانطور که مشخص است، تفاوت بین



شکل ۴- نمودار جعبه‌ای مقادیر تبخیر-تعرق مرجع ماهانه محاسبه شده با روش پنمن-ماتیس فائو (به عنوان مرجع) در ایستگاه نیشابور و مدل‌های داده‌های پانلی (خطچین با دایره مشکی)، روش رگرسیون حداقل مربعات معمولی (خطچین با مثلث مشکی) و روش هارگریوز-سامانی (خطچین با مربع مشکی) برای هر فصل در مرحله توسعه مدل.

نتیجه گیری

در این پژوهش، کاربرد مدل داده‌های پانلی به عنوان تکنیکی مفید و کارآمد در برآورد تبخیر-تعرق مرجع مورد بررسی قرار گرفت. یکی از مهمترین فواید این مدل، برآورد تبخیر-تعرق مرجع در چندین ایستگاه به صورت همزمان و با طول دوره‌های آماری متفاوت می‌باشد. به عبارت دیگر، تجزیه و تحلیل داده‌های پانلی، تحلیل رگرسیون در هر دو بعد مکانی و زمانی را ارائه می‌دهد. در این مطالعه، تصریح اثرهای ثابت دو طرفه مناسبترین تصریح برای برآورد تبخیر-تعرق مرجع بر اساس آزمون‌های چاو و هاسمن انتخاب گردید. مدل داده‌های پانلی، جذر میانگین مربعات خطای ۹/۸۵ میلی‌متر در ماه، میانگین قدرمطلق خطا ۷/۳۸ میلی‌متر در ماه و ضریب تعیین ۰/۹۹ را نشان داد. مقایسه مدل داده‌های پانلی با مدل رگرسیون حداقل مربعات معمولی و روش هارگریوز-سامانی تاکید دارد که، مدل فوق دارای توانایی بالایی در برآورد تبخیر-تعرق مرجع می‌باشد. کاربرد مدل داده‌های پانلی در مدیریت منابع آب جدید می‌باشد و با توجه به کاربرد موفق این مدل در پژوهش حاضر، مدل داده‌های پانلی در بخش‌های دیگر مدیریت منابع آب نیز پیشنهاد می‌گردد.

منابع

- تکنیک پیش‌بینی غیر-نظارت شونده در برآورد تبخیر-تعرق گیاه مرجع. تحقیقات منابع آب ایران. ۱۱: ۳۱-۴۲.
- کواکبی، غ.، موسوی بایگی، م.، مساعدی، ا.، جباری نوقابی، م.، ۱۳۹۳. تعیین عوامل موثر بر وقوع خشکسالی با تحلیل داده‌های پانلی (مطالعه موردی استان خراسان رضوی). نشریه آب و خاک (علوم و صنایع کشاورزی) ۶: ۱۲۹۸-۱۳۱۰.
- مهرگان، ن.، اشرفزاده، ح. ر. ۱۳۸۷. اقتصادسنجی پانل دیتا. چاپ اول دانشگاه تهران، موسسه تحقیقات تعاون، تهران.
- وزیری، ژ.، سلامت، ع.، انتصاری، م.، مسچی، م.، حیدری، ن.، دهقانی سانج، ح.، ۱۳۸۷. تبخیر و تعرق (دستورالعمل محاسبه آب مورد نیاز گیاهان). انتشارات کمیته ملی آبیاری و زهکشی ایران، تهران.
- هادی قنوت، غ.، شهیدی، ع.، خاشعی سیوکی، ع.، هاشمی، ر.، ۱۳۹۴. کاربرد مدل داده‌های ترکیبی در برآورد غلظت کلر آبخوان دزفول. آب و فاضلاب ۵: ۴۸-۵۵.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56.
- Baltagi, B., 2008. Econometric analysis of panel data. Wiley.
- Croissant, Y., Millo, G., 2018. Panel data econometrics with R. Wiley.
- Eslamian, S.S., Gohari, S.A., Zareian, M.J., Firoozfar, A., 2012. Estimating Penman-Monteith reference evapotranspiration using artificial neural networks

- شریفان، ح.، قربانی، خ.، ۱۳۹۳. بهبود برآورد تبخیر-تعرق پتانسیل با استفاده از ضریب اصلاحی به کمک مدل درخت تصمیم M5.
- نشریه آبیاری و زهکشی ایران. ۱: ۵۳-۶۱.
- فرسادنی، ف.، زحمتی، س.، قهرمان، ب.، مقدم‌نیا، ع. ۱۳۹۴. ارائه

- Maurseth, P.B., 2018. The effect of the Internet on economic growth: Counter-evidence from cross-country panel data. *Economics Letters*. 172: 74-77.
- Ozgur, K., Aytac, G., 2010. Evapotranspiration modeling using linear genetic programming technique. *Journal of Irrigation and Drainage Engineering*. 136. 10: 715-723.
- Pesaran, M.H., 2015. Time series and panel data econometrics. Oxford University Press.
- Samani, Z. 2000. Estimating solar radiation and evapotranspiration using minimum climatological data. *Journal of Irrigation and Drainage Engineering*, 126. 4: 265-267.
- Shiri, J., Nazemi, A.H., Sadraddini, A.A., Landaras, G., Kisi, O., Fard, A.F., Marti, P., 2013. Global cross-station assessment of neuro-fuzzy models for estimating daily reference evapotranspiration. *Journal of Hydrology*. 480: 46-57.
- Slavisa, T., Branimir, T., Miomir, S., 2003. Forecasting of reference evapotranspiration by artificial neural networks. *Journal of Irrigation and Drainage Engineering*. 129. 6: 454-457.
- Suzuki, T., 2017. Effects of the global economic crisis on FDI inflow in eastern European economies: a panel data analysis. *Economics of European Crises and Emerging Markets*. 10: 63-91.
- Yassin, M.A., Alazba, A.A., Mattar, M.A., 2016. Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agricultural Water Management*. 163: 110-124.
- and genetic algorithm: a case study. *Arabian Journal for Science and Engineering*. 37. 4: 935-944.
- 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.
- Hsiao, C. 2003. Analysis of panel data. Cambridge University Press.
- Izady, A., Davary, K., Alizadeh, A., Ghahraman, B., Sadeghi, M., Moghaddamnia, A., 2012. Application of "panel-data" modeling to predict groundwater levels in the Neishaboor Plain, Iran. *Hydrogeology Journal*. 20. 3: 435-447.
- Karbasi, M., 2018. Forecasting of multi-step ahead reference evapotranspiration using wavelet- gaussian process regression model. *Water Resources Management*. 32. 3: 1035-1052.
- Kisi, O., 2011. Evapotranspiration modeling using a wavelet regression model. *Irrigation Science*. 29. 3: 241-252.
- Lee, L.F., Yu, J., 2010. Some recent developments in spatial panel data models. *Regional Science and Urban Economics*. 40. 5: 255-271.
- Lin, C.H., Chao, C., Cheng, K.H., 2014. Study for regional evapotranspiration by fuzzy inference system. *Advanced Materials Research*. 838-841: 1776-1779.
- Mahabbati, A., Izady, A., Mousavi Baygi, M., Davary, K., Hasheminia, S.M., 2017. Daily soil temperature modeling using "panel-data" concept. *Journal of Applied Statistics*. 44. 8: 1385-1401.

Using Panel Data Technique to Estimation Reference Evapotranspiration

P. Shirazi¹, K. Davary^{2*}, B. Ghahraman³, M. Jabbari Nooghabi⁴

Recived: Jan.26, 2019

Accepted: Feb.16, 2019

Abstract

Evapotranspiration is the most important part of the hydrological cycle, which plays a key role in water resource management, crop yield simulation, and irrigation scheduling. The purpose of this research was to estimate the reference evapotranspiration using 'panel-data' models. Panel-data multivariate analysis endows regression analysis with both spatial and temporal dimensions. This study was carried out using weather data of 9 synoptic stations of Khorasan Razavi during 1971-2000. Data were divided randomly into two sub-sets, 75% for model development and 25% for model evaluation. The panel-data models were developed using the monthly mean air temperature and monthly mean wind speed as inputs in order to estimate monthly reference evapotranspiration. The results indicated that the two-way fixed effects models were superior. The statistical index (RMSE = 9.85, MAE = 7.38 and $R^2 = 0.99$) revealed the effectiveness of this model. In addition, these results were compared with the results of ordinary least squares regression and Hargreaves-Samani equation which showed the superiority of the panel-data models.

Keywords: FAO Penman-Monteith, Panel Data Models, Specification Fixed Effects

1- Ph.D. Student, Water Science Engineering Department, Ferdowsi University of Mashhad
2- Professor, Water Science Engineering Department, Ferdowsi University of Mashhad
3- Professor, Water Science Engineering Department, Ferdowsi University of Mashhad
4- Assistant Professor, Statistics Department, Ferdowsi University of Mashhad
(*- Corresponding Author Email: k.davary@um.ac.ir)