

ارزیابی مدل‌های تجربی و هوش مصنوعی در برآورد تبخیر- تعرق مرجع (مطالعه موردی: ایستگاه بروجرد)

یاسر سبزواری^۱ و مهری سعیدی نیا^{۲*}

(تاریخ دریافت: ۱۳۹۹/۲/۱؛ تاریخ پذیرش: ۱۳۹۹/۸/۲۷)

چکیده

فائوپنمن مانتیث روشی مینا، برای برآورد تبخیر- تعرق مرجع است. در خیلی از موارد دسترسی به همه اطلاعات مورد نیاز، مشکل است، لذا جایگزینی مدل‌های با ورودی‌های اولیه کم و دقت مناسب ضرورت می‌یابد. هدف از این پژوهش بررسی قابلیت مدل‌های تجربی، برنامه‌ریزی بیان ژن، رگرسیون گام‌به‌گام و شبکه‌بیزین در برآورد تبخیر- تعرق مرجع است. برای اطلاعات ورودی مدل از اطلاعات روزانه ایستگاه سینوپتیک بروجرد در بازه زمانی ۱۳۹۶-۱۳۷۵ استفاده شد. بر اساس همبستگی بین پارامترهای ورودی و خروجی، شش الگوی ورودی برای مدل‌سازی تعیین شد. نتایج نشان داد از بین مدل‌های تجربی، مدل کیمبرلی‌پنمن دارای عملکرد بهتری است. برنامه‌ریزی بیان ژن با الگوی چهارم و عملگرهای پیش‌فرض مدل، دارای $R^2=0/98$ و $RMSE=0/9$ ، شبکه‌بیزین با الگوی ششم، دارای $R^2=0/91$ و $RMSE=1/01$ و رگرسیون گام‌به‌گام با الگوی ششم دارای $R^2=0/91$ و $RMSE=0/9$ در مرحله آموزش دقیق‌ترین الگوها هستند. مقایسه عملکرد مدل‌ها حاکی از برتری مدل برنامه‌ریزی بیان ژن نسبت به بقیه مدل‌ها بود به طوری که دارای متوسط قدر مطلق خطای نسبی (AARE) به میزان ۰/۱۲ و نسبت میانگین (MR) به میزان ۰/۹۴ بود. نتایج به‌دست آمده نشان داد که برنامه‌ریزی بیان ژن دارای توانایی قابل قبولی در تخمین تبخیر- تعرق مرجع تحت شرایط آب‌وهوایی بروجرد بوده و می‌تواند به‌عنوان یک مدل مناسب ارائه شود.

واژه‌های کلیدی: فائوپنمن مانتیث، تبخیر- تعرق مرجع، شبکه‌بیزین، برنامه‌ریزی بیان ژن

۱- گروه مهندسی آب، دانشکده کشاورزی، دانشگاه صنعتی اصفهان

۲- گروه مهندسی آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان

*: مسئول مکاتبات: پست الکترونیکی: Saeedinia.m@lu.ac.ir

مقدمه

یکی از عوامل مؤثر بر نیاز آبی گیاهان تبخیر- تعرق مرجع (ET_0) (Reference Evapotranspiration) بوده که یکی از مهم‌ترین اجزای چرخه هیدرولوژی و تابع متغیرهای اقلیمی از جمله: سرعت باد در نزدیکی سطح زمین، دمای هوا، تابش خورشیدی و رطوبت نسبی است. محاسبه ET_0 به صورت مستقیم با استفاده از لایسیمتر صورت می‌گیرد که معمولاً هزینه بر و زمان‌بر است (۱۲، ۱۶ و ۲۸). لذا بیشتر از روش‌های غیرمستقیم استفاده می‌شود. بنابراین ET_0 را می‌توان با استفاده از مدل‌های آب و هواشناسی که بر پایه معادلات فیزیکی، یا رابطه تجربی متغیرهای هواشناسی هستند، برآورد و شبیه‌سازی کرد. سازمان خواروبار جهانی (FAO) (Food and Agriculture Organization) روش فائو- پنمن-مانتیث (PMF-56) را به عنوان معادله استاندارد توصیه کرده است که ET_0 را به صورت روزانه و ماهانه برآورد می‌کند (۳). این روش به تعدادی از متغیرهای اقلیمی نظیر: حداقل و حداکثر دمای هوا، تابش خورشیدی، رطوبت نسبی، ساعات آفتابی و بسیاری از پارامترهای دیگر نیاز دارد. با وجود این، این متغیرهای هواشناسی به خصوص در کشورهای در حال توسعه، یا در دسترس نبوده یا معمولاً غیرقابل اعتماد هستند (۴). بنابراین ارائه روشی که بتواند با داده‌های ورودی کمتر، دقت مناسبی را به همراه داشته باشد، الزامی است. در سال‌های اخیر، روش‌های هوش مصنوعی (AI) (Artificial Intelligence) با داده‌های هواشناسی محدود، به طور موفقیت‌آمیزی در شبیه‌سازی ET_0 مورد استفاده قرار گرفته‌اند. از جمله روش‌های هوش مصنوعی به کار گرفته شده برای شبیه‌سازی ET_0 می‌توان به شبکه بیزین (۶) و برنامه‌ریزی بیان ژن (۱۱ و ۱۶) اشاره کرد. حسینی و همکاران (۱۰) به ارزیابی و تحلیل حساسیت روش‌های مختلف تخمین تبخیر- تعرق مرجع روزانه در یک اقلیم خشک سرد پرداختند. نتایج مشخص کرد که شبکه عصبی مصنوعی عملکرد مناسب‌تری نسبت به مدل‌های تجربی و مدل رگرسیون گام به گام دارد. یاسین و همکاران (۲۷) به ارزیابی شبکه عصبی

مصنوعی و برنامه‌ریزی بیان ژن برای شبیه‌سازی تبخیر تعرق مرجع در اقلیم خشک پرداختند. نتایج نشان داد مدل شبکه عصبی مصنوعی با الگوی شماره ۸ (شامل همه پارامترها) با $R^2=0/998$ و $RMSE=0/2$ mm/day دارای بهترین دقت است. شیری (۲۱) به ارزیابی مدل‌های فائو- پنمن-مانتیث ۵۶، تجربی، نیمه‌تجربی و برنامه‌ریزی بیان ژن در شبیه‌سازی تبخیر تعرق مرجع روزانه برای مناطق فراخشک ایران پرداخت. نتایج حاکی از برتری مدل برنامه‌ریزی بیان ژن نسبت به روش‌های تجربی و نیمه‌تجربی است. مهدی‌زاده و همکاران (۱۸) با استفاده از مدل‌های رگرسیون چندمتغیره تطبیقی، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان، تبخیر تعرق مرجع ماهانه را شبیه‌سازی کردند. نتایج نشان داد که روش رگرسیون چندمتغیره تطبیقی و ماشین بردار پشتیبان با هسته توابع پایه شعاعی به طور کلی بهتر از برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان با توابع چندجمله‌ای مدل‌سازی را انجام می‌دهند. مقایسه دقت معادلات تجربی و روش‌های محاسبات نرم نشان داد که عملکرد رگرسیون چندمتغیره تطبیقی و ماشین بردار پشتیبان با هسته توابع پایه شعاعی بهتر از معادلات تجربی است. متر (۱۷) با استفاده از برنامه‌ریزی بیان ژن به مدل‌سازی تبخیر تعرق مرجع ماهانه در مصر پرداخت. نتایج نشان داد که برنامه‌ریزی بیان ژن با ورودی‌های بیشینه و کمینه درجه حرارت، رطوبت نسبی و سرعت باد با $RMSE=0/426$ میلی‌متر در روز، $R^2=0/963$ برای مرحله آزمون و $RMSE=0/43$ میلی‌متر در روز، $R^2=0/962$ برای مرحله آزمون دارای بیشترین دقت در برآورد ET_0 بوده است. سان و همکاران (۲۳) در پژوهشی به بهبود شبیه‌سازی تبخیر- تعرق با استفاده از مدل میانگین شبکه بیزین با مدل توازن انرژی سطحی در چین پرداختند. نتایج نشان داد که مدل میانگین شبکه بیزین نسبت به چهار مدل توازن انرژی سطحی تبخیر- تعرق را در دو اقلیم بهتر پیش‌بینی کرده است. ولی‌پور و همکاران (۲۵) در تحقیقی با تکنیک یادگیری ماشین با تکنولوژی ترکیبی داده، به مدل‌سازی تبخیر تعرق در اقلیم‌های مختلف پرداختند. نتایج این تحقیق نشان داد که مدل

به‌منظور برآورد ET_0 از روش‌های تعریف شده در نرم‌افزار RefET-۳/۲ استفاده شد. در این نرم‌افزار روش‌های محاسباتی با توجه به نوع داده‌های ورودی (بیشینه و کمینه درجه حرارت، درجه حرارت متوسط روزانه، بیشینه و کمینه رطوبت نسبی، رطوبت نسبی متوسط روزانه، سرعت باد، میزان بارش، ارتفاع و مختصات جغرافیایی ایستگاه) شامل ۷ روش ترکیبی بر پایه پنمن، دو روش دمایی، یک روش تشعشعی و سه روش ترکیبی تشعشعی- دمایی است. در این پژوهش از دو روش ترکیبی فائوپنمن-مانتیث (۳) و کیمبرلی-پنمن (۲۶)، دو روش دمایی بلانسی-کریدل-فائو ۲۴ (۵) و هارگریوز-سامانی (۹)، روش تشعشعی فائو ۲۴ (۱۱) و سه روش تشعشعی دمایی شامل معادلات پرستلی-تیلور (۱۹)، مکینک (۱۵) و تورک (۲۴) استفاده شد. روش محاسبه هرکدام از روش‌ها در جدول ۱ آورده شده است.

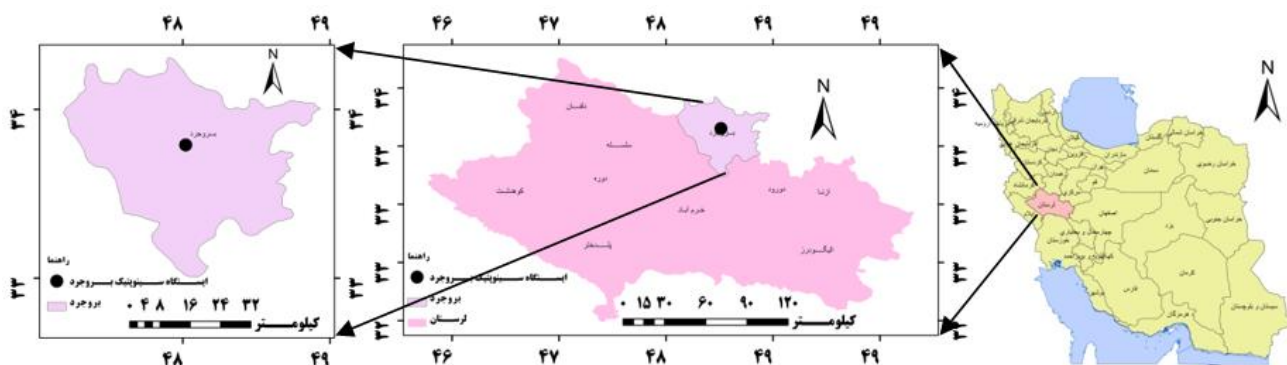
در این روابط: ET_0 : تبخیر- تعرق گیاه مرجع، برحسب میلی‌متر بر روز، R_n : تشعشع خالص خورشیدی رسیده به زمین (موج کوتاه) برحسب مگاژول بر متر مربع بر روز که مقدار آن برای عرض‌های جغرافیایی مختلف برای هر ماه به دست می‌آید. G : شار گرمایی خاک، برحسب مگاژول بر متر مربع بر روز Δ : شیب نمودار فشار بخار اشباع نسبت به دما، برحسب کیلو پاسکال بر درجه سلسیوس؛ ثابت سایکرومتری، برحسب کیلو پاسکال بر درجه سلسیوس e_a ؛ کمبود فشار بخار در ارتفاع ۲ متری، برحسب کیلو پاسکال WS : میانگین روزانه سرعت باد در ارتفاع ۲ متری از سطح زمین، برحسب متر بر ثانیه T_{mean} : میانگین روزانه دمای هوا در ارتفاع ۲ متری از سطح زمین، برحسب درجه سلسیوس. T_D : میانگین دامنه تغییرات دمای روزانه (اختلاف حداکثر و حداقل دما) در دوره ۲۲ ساله (۱۳۷۵-۱۳۹۶). a_w و b_w : ضرایب تجربی هستند که بنابر پیشنهاد اولیه پنمن در صورتی که سرعت باد برحسب متر بر ثانیه باشد مقادیر عددی آنها به ترتیب ۱ و ۰/۵۳۷ است. k_w : ضریب ثابت برابر ۶/۴۳ برحسب (میلی‌متر بر روز).

برنامه‌ریزی بیان ژن یک ابزار قدرتمند در پیش‌بینی تبخیر- تعرق در اقلیم نیمه‌خشک و مدیترانه‌ای است اما در اقلیم‌های بسیار مرطوب و بعضی مناطق خشک باید با احتیاط مورد استفاده قرار گیرد. براساس مطالعات صورت گرفته، ضرورت مدیریت بهینه منابع آب و افزایش تولید محصولات کشاورزی، برای تضمین امنیت غذایی در همه کشورهای حائز اهمیت است. بنابراین، ET_0 باید به‌طور دقیقی برآورد شود. نقص یا نبود اطلاعات هواشناسی، مشکلی تأثیرگذار در برآورد ET_0 است. بنابراین ET_0 با استفاده از اطلاعات ورودی کمتر و دقت بالا برآورد و مدل‌سازی شود. از این‌رو هدف این پژوهش ارزیابی مدل‌های تجربی در برآورد ET_0 و بررسی دقت مدل‌های رگرسیون گام به گام، شبکه بیزین (BN) و برنامه‌ریزی بیان ژن (GEP) در برآورد ET_0 و معرفی مدل برتر در برآورد ET_0 در شرایط آب و هوایی بروجرد است.

مواد و روش‌ها

معرفی منطقه مورد مطالعه و اطلاعات مورد استفاده

بروجرد در ارتفاع ۱۶۲۰ متری از سطح دریا و $33^{\circ} 9'$ درجه شمالی و $48^{\circ} 8'$ درجه شرقی واقع است. میزان بارش در شهر بروجرد تقریباً برابر ۵۸۸ میلی‌متر در سال است که در فصل سرما بارش‌ها بیشتر به صورت برف است. آمار و اطلاعات هواشناسی مورد نیاز برای محاسبه تبخیر- تعرق گیاه مرجع در این پژوهش، از ایستگاه سینوپتیک بروجرد با مختصات جغرافیایی طول ۴۸ درجه و ۴۵ دقیقه، عرض ۳۳ درجه و ۳۵ دقیقه و ارتفاع از سطح دریا ۱۶۲۹، در یک دوره زمانی ۲۲ ساله (۱۳۷۵-۱۳۹۶) دریافت شد. شکل ۱ نشان‌دهنده موقعیت دشت بروجرد و ایستگاه سینوپتیک بروجرد است. اطلاعات هواشناسی مورد استفاده شامل روزانه: درجه حرارت حداکثر و درجه حرارت حداقل $(T_{max}$ و $T_{min})$ [C]، رطوبت نسبی حداکثر و رطوبت حداقل $(Rh_{max}$ و $Rh_{min})$ ، تابش خورشیدی (N) $[MJ m^{-2} d^{-1}]$ و سرعت باد در ارتفاع دو متری (WS) $[ms^{-1}]$ به‌منظور مقایسه $[mmd^{-1}]$ با ET_0 روش فائوپنمن-مانتیث است.



شکل ۱. موقعیت جغرافیایی بروجرده در استان و کشور

جدول ۱. مدل‌های تجربی برآورد تبخیر- تعرق مرجع

ردیف	نام روش محاسبه ET_0	شکل معادله
۱	هارگریوز سامانی (۹)	$ET_0 = 0.0023R_n(T_{mean} + 17.8)T_D^{0.5}$
۲	بلانی کریدل فائو ۲۴ (۵)	$ET_0 = a + b[P(0.46T_{mean} + 8.13)]$
۳	مکینک (۱۵)	$ET_0 = 0.61 \frac{\Delta}{\Delta + \gamma} \cdot \frac{R_s}{2.45} - 0.12$
۴	تورک-۱۹۶۱ (۲۴)	$ET_0 = a_T \times 0.013 \frac{T_{mean}}{T_{mean} + 15} \times \frac{23.8856R_s + 50}{\lambda}$
۵	پریستلی تیلور (۱۹)	$ET_0 = 1.26 \frac{\Delta}{\Delta + \gamma} \cdot \frac{R_n - G}{\lambda}$
۶	معادله اصلاح شده پنمن توسط کیمبرلی (۲۶)	$ET_0 = \frac{\Delta}{\Delta + \gamma} (R_n - G) + k_w \frac{\gamma}{\gamma + \Delta} (a_w + b_w u_2) (e_s - e_o) / \lambda$
۷	تابشی فائو (۱۱)	$ET_0 = a + b \left[\frac{\Delta}{\Delta + \gamma} R_s \right]$
۸	فائوپنمن مانتیث (۳)	$ET_0 = \frac{[0.408(R_n - G) + [(900\gamma)/(T + 273)] \times U_2 \times (e_a - e_d)]}{\Delta + \gamma \times (1 + 0.34U_2)}$

ورودی مدل‌سازی، از رگرسیون گام به گام استفاده شد. ترکیب‌های ورودی مدل‌سازی در جدول ۳ آورده شده است. بر اساس همبستگی بین پارامترهای اقلیمی مختلف و ET_0 و میزان اثرگذاری آنها بر ET_0 با استفاده از رگرسیون گام به گام و ترتیب ورود این پارامترها در مدل‌سازی، ۶ الگوی ورودی برای ارزیابی ترکیب‌های مختلف تعیین شد.

مدل رگرسیون گام به گام

با این روش می‌توان به‌طور همزمان به تحلیل و بررسی چند متغیر مختلف پرداخت. رگرسیون گام به گام در واقع، رابطه بین چند متغیر پیش‌بین با متغیر پاسخ مورد نظر را بیان می‌کند.

در این پژوهش اطلاعات مورد استفاده متشکل از دو دسته داده‌های آموزش و داده‌های آزمون هستند. داده‌های آموزش در هنگام آموزش مدل مورد استفاده قرار می‌گیرند، به این ترتیب که مقدار خطا به‌زای این داده‌ها به‌عنوان شاخص عملکرد مدل‌ها در نظر گرفته می‌شود. داده‌های آزمون پس از اتمام فرایند آموزش و برای ارزیابی عملکرد مدل‌ها در تعمیم نتایج، مورد استفاده قرار می‌گیرند. داده‌های آموزش شامل ۷۰ درصد داده‌های مورد استفاده و داده‌های آزمون شامل ۳۰ درصد باقیمانده داده‌های مورد استفاده است. جدول ۲، نشان‌دهنده پارامترهای آماری داده‌های مورد استفاده است. برای بررسی اثرگذاری متغیرهای هواشناسی بر ET_0 و تعیین الگوهای

جدول ۲. آمار توصیفی داده‌های مورد استفاده

متغیر	چولگی	کشیدگی	حداکثر	حداقل	انحراف معیار	میانگین	
Tmax	-۱/۰۹	-۰/۱۸	۴۱/۴	-۸/۴	۱۱/۳۲	۲۱/۴۸	داده‌های آموزش
Tmin	-۰/۷۱	-۰/۲۱	۲۷	-۲۱/۴	۸/۳۲	۸/۳۱	
RHmax	-۱/۴	۰/۰۸	۱۰۰	۱۲	۲۴/۵۲	۵۸	
RHmin	۰/۶۳	۱/۱۴	۹۴	۱	۱۸/۴	۲۶/۵۴	
WS	۱/۵۷	۰/۹۶	۱۳/۸۷	۰	۱/۸۷	۳/۳۶	
N	-۰/۱۳	-۰/۹۴	۱۳/۸	۰	۳/۶۳	۸/۴۹	
ET0	-۱/۱۳	۰/۲۶	۱۳/۱	۰/۳۸	۲/۹۶	۴/۹۶	داده‌های آزمون
Tmax	-۱/۱۵	-۰/۱۳	۴۱	-۱۰	۱۱/۱۱	۲۱/۶۸	
Tmin	-۰/۸۵	-۰/۱۳	۲۶/۲	-۲۰/۴	۸/۱۷	۸/۵۴	
RHmax	-۱/۳۹	۰/۰۸	۱۰۰	۱۵	۲۴/۵	۵۷/۷	
RHmin	۰/۵۲	۱/۰۸	۹۰	۰	۱۸/۰۷	۲۶/۳۵	
WS	۱/۵۴	۰/۹۸	۱۳	۰	۱/۸۸	۳/۳۵	
N	-۰/۳۱	-۰/۸۶	۱۳	۰	۳/۶۷	۸/۳۷	
ET0	-۱/۴۳	۷/۷۴	۹۷/۲	۰/۳۹	۳/۵۵	۴/۹۹	

جدول ۳. الگوهای ورودی شبکه بیزین و برنامه‌ریزی بیان ژن

ردیف	کد الگو	الگوی ورودی مدل
۱	M1	$ET0=f\{Tmax\}$
۲	M2	$ET0=f\{Tmax, WS\}$
۳	M3	$ET0=f\{Tmax, WS, N\}$
۴	M4	$ET0=f\{Tmax, WS, N, Tmin\}$
۵	M5	$ET0=f\{Tmax, WS, N, Tmin, Rhmin\}$
۶	M6	$ET0=f\{Tmax, WS, N, Tmin, Rhmin, Rhmax\}$

داشته باشد.

در صورت نقض این دو فرض معادله رگرسیونی را نمی‌توان تخمین زد. مدل رگرسیونی به شکل ماتریسی را می‌توان به صورت معادله زیر نشان داد:

$$Y = X\beta + e \quad (1)$$

که β ماتریس ضرایب رگرسیونی که ضرایب نامشخصی هستند که در حقیقت، مسئول برآورد پارامتر پاسخ هستند، e ماتریس خطای برازش و Y نیز ماتریس پاسخ است. با حل معادله (۱)

متغیرهای پیش‌بین، متغیرهایی هستند که با بهره‌گیری از آنها متغیر دیگر پیش‌بینی شده و متغیر پاسخ متغیری است که مورد پیش‌بینی قرار می‌گیرد. این گونه مدل‌ها دارای فرضیاتی هستند. فرض‌هایی که رگرسیون گام به گام را از رگرسیون ساده متمایز می‌کند این است که:

(۱) تعداد متغیرهای پیش‌بین (مستقل) در رگرسیون باید کمتر از تعداد مشاهدات باشد.

(۲) همبستگی خطی کامل بین متغیرهای پیش‌بین و پاسخ وجود

برحسب β خواهیم داشت:

$$\beta = (X'X)^{-1}(X'Y) \quad (2)$$

که در رابطه بالا، X ترانهاده ماتریس X است. برای محاسبه معکوس $(X'X)$ لازم است متغیرهای مستقل همبستگی زیادی نداشته باشند، زیرا در این صورت ماتریس $(X'X)$ را نمی‌توان معکوس کرد و باعث افزایش خطا در اثر گرد کردن داده‌ها و محاسبات می‌شود. در این مطالعه برای انجام تحلیل رگرسیون گام به گام از نرم‌افزار SPSS-24 استفاده شد.

برنامه‌ریزی بیان ژن

برنامه‌ریزی بیان ژن (GEP)، توسعه یافته روش‌های برنامه‌ریزی ژنتیک (GP) و الگوریتم ژنتیک (GA) است که در سال ۱۹۹۹ توسط فریرا (۸) بر اساس تئوری داروین ابداع شد. حالت‌های مختلف ترکیب‌های ژن‌ها را ژنوتیپ (در این مطالعه ترکیب‌های مختلف از پارامترهای هواشناسی) و نتایج حاصل از حالت‌های مختلف ترکیب‌های ژن‌ها را فنوتیپ (در این مطالعه حالت‌های حاصل از ترکیب‌های مختلف از پارامترهای هواشناسی برای پارامتر ET_0) گویند. در روش برنامه‌ریزی بیان ژن، کروموزوم‌های خطی و ساده با طول ثابت، مشابه با الگوریتم ژنتیک و ساختارهای شاخه‌ای با اندازه و شکل‌های متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک ترکیب می‌شوند. در این روش ژنوتیپ و فنوتیپ از هم جدا شده و سیستم قادر خواهد بود از تمام مزایای تکاملی بهره‌گیرد. با وجود اینکه فنوتیپ در GEP مشابه ساختار شاخه‌ای GP است، اما ساختار شاخه‌ای در GEP که بیان درختی نیز نامیده می‌شود، بیانگر تمام ژنوم‌های مستقل است. در مجموع، در GEP بهسازی در یک ساختار خطی اتفاق افتاده و سپس به صورت ساختار درختی بیان می‌شود. این امر سبب خواهد شد که تنها ژنوم اصلاح شده به نسل بعد انتقال یابد و نیازی به ساختارهای سنگین تکثیر و جهش نباشد. برنامه‌ریزی بیان ژن برخلاف الگوریتم ژنتیک روی ساختار درختی رابطه‌ها به جای سلسله ارقام دودویی

عمل می‌کند که این ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در رابطه‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند (۸).

در این پژوهش برای اجرای مدل GEP از نرم‌افزار Genexprotools 5 استفاده شد. علاوه بر چهار عملگر اصلی (F1)، $\{+, -, *, /\}$ ، حالت‌هایی بر اساس عملگرهای ریاضی پیش‌فرض برنامه (F2)، $\{\sqrt{\quad}, \sqrt[3]{\quad}, \text{Exp}, \text{Ln}, \text{Sin}, \text{Cos}, \text{Atan}\}$ نیز در نظر گرفته شد (۱۲ و ۱۳). در ساخت درخت تجزیه و تولید جمعیت اولیه، از راه‌حلی‌هایی که از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در روابط) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند، استفاده شد.

در ابتدای اجرای مدل GEP، باید تعداد ژن‌ها و کروموزوم‌ها مشخص شود. با توجه به مطالعات انجام شده قبلی، بهترین جمعیت از ۳۰ تا ۵۰ کروموزوم و پاسخ مناسب از انتخاب ۲ یا ۳ ژن حاصل می‌شود (۸). بنابراین برای جمعیت اولیه در این پژوهش از ۳۰ کروموزوم و از ۳ ژن در ترکیب کروموزوم استفاده شد. همچنین تابع پیوند (+) به علت برتری بر دیگر توابع انتخاب شد. حداکثر تکرار تابع برازش هم ۱۰۰۰ تعیین شد. مقادیر نرخ عملگرهای ژنتیکی ترانهش درج متوالی و ترانهش ریشه درج متوالی برابر ۰/۱، نرخ جهش و نرخ وارون‌سازی همان مقادیر پیش‌فرض مدل و مقدار مجموع نرخ سه عملگر ترکیب تک نقطه‌ای، ترکیب دو نقطه‌ای و ترکیب ژن، طبق نظر فریرا (۸) برابر ۰/۷ تعیین شد.

شبکه بیزین

شبکه بیزین مدلی گرافیکی احتمالی است (۱۴) که یک نمایش نموداری از اثرات متقابل متغیرهایی است که باید مدل‌بندی شوند، به طوری که علاوه بر کیفیت رابطه بین متغیرهای مسئله، کمیت ارتباط بین این متغیرها را نیز به نمایش می‌گذارد (۷). این روش بر مبنای محاسبات قانون بیز است. قانون بیز به صورت رابطه زیر نشان داده شده است:

بالاترین امتیاز به عنوان ساختار مورد نظر انتخاب می‌شود (۲۰). دو الگوریتم PC و NPT با توجه به سادگی، در شبکه‌های بیزین برای آموزش ساختار مدل‌سازی بیشتر مورد استفاده قرار می‌گیرند. در این تحقیق برای آموزش در شبکه‌های بیزین، از الگوریتم PC استفاده شد. دلیل انتخاب این الگوریتم این است که مدل شبکه بیزین اغلب برای مدل‌سازی داده‌های کیفی استفاده می‌شود. با توجه به ماهیت کمی بودن داده‌های تحقیق، باید از الگوریتمی که توانایی مدل‌سازی داده‌های کمی را دارد استفاده شود که الگوریتم PC دارای این قابلیت است. در این الگوریتم، مجموعه‌ای از استقلال و وابستگی عبارات شرطی توسط آزمون‌های آماری استخراج می‌شود که در آن استقلال شرطی بین هر جفت متغیر وجود نداشته و جهت کمان‌ها بر اساس استقلال شرطی تعیین می‌شود. شبکه بیزین علاوه بر پیش‌بینی متغیر هدف با استفاده از وضعیت متغیرهای ورودی، قادر است با در دست داشتن وضعیت متغیر پیش‌بینی شونده، میزان تأثیر هر یک از متغیرهای ورودی را بر خروجی‌های مدل تعیین کند (۱). در این تحقیق برای آموزش پارامتری شبکه، با تنظیم دو متغیر سطح معنی‌داری (Significance Level) و حداکثر اندازه مجاورت (Max Adjacency Size) با توجه به تأثیر پارامترها بر یکدیگر استفاده شد. برای اجرای شبکه بیزین از نرم‌افزار GeNie 2 استفاده شد (۲).

معیارهای ارزیابی

به منظور ارزیابی دقت و کارایی مدل‌ها، از معیارهای ضریب تبیین (R^2)، ریشه میانگین مربعات خطا (Root Mean Square Error) (RMSE)، متوسط قدرمطلق خطای نسبی (Average Absolute Relative Error) (AARE) و نسبت میانگین (Mean Ratio) MR، ((۱۳ و ۲۲) به صورت زیر استفاده شد:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2} \quad (5)$$

$$P(b|a) = P(a|b) \times \frac{P(b)}{P(a)} \quad (3)$$

که در رابطه فوق $P(a)$: احتمال وقوع پیشامد a ، $P(b)$: احتمال وقوع پیشامد b ، $P(b|a)$: احتمال وقوع پیشامد b به شرطی که a اتفاق افتاده باشد و $P(a|b)$: احتمال وقوع پیشامد a به شرطی که b اتفاق افتاده باشد، هستند.

شبکه بیزین متشکل از گره‌ها، که متغیرهای مکانیزم مدیریتی هستند، بندها، که روابط منتج بین متغیرها را به وجود می‌آورند و احتمالات، است. شبکه بیزین ساختاری گرافیکی است که متغیرها به صورت گره و روابط علت و معلولی آنها با استفاده از پیکان موسوم به Edge مشخص می‌شود. چنانچه در گراف گره‌هایی باشد که قبل از آنها، گره دیگری در گراف وجود داشته باشد، با توزیع احتمال شرطی تعریف می‌شوند، در غیر این صورت با احتمال آغازین بیان می‌شوند. پایین‌ترین بخش در شبکه بیزین از طریق قانون احتمال کل حاصل می‌شود. احتمالات مرتبط با بخش‌های بالایی شبکه بر اساس قانون بیز هستند. در حالت کلی، اگر گره x دارای والد نباشد گره غیرمشروط و در غیر این صورت گره مشروط است (۲).

در مورد ساختار شبکه بیزین اگر مجموعه‌ای شامل n متغیر به صورت $\{X_1, X_2, X_3, \dots, X_n\}$ وجود داشته باشد که هر X_i یک متغیر تصادفی باشد و گره‌های والد این متغیر با $parent(X_i)$ نشان داده شوند، احتمال هر رویداد مانند $X=(x_1, x_2, x_3, \dots, x_n)$ به صورت زیر محاسبه می‌شود:

$$P(x_1 + x_2 + \dots + x_n) = \prod P(x_i | parents(x_i)) \quad (4)$$

الگوریتم‌های یادگیری شبکه بیزین به الگوریتم‌های یادگیری بر مبنای محدودیت و بر مبنای امتیاز، طبقه‌بندی می‌شوند. دسته اول بر اساس استقلال شرطی و وابستگی بین متغیرها، توسط آزمون‌های آماری به دست می‌آید (مانند الگوریتم‌های PC (Path condition) و NPT (Necessary Path Condition)). در روش‌های یادگیری بر مبنای امتیاز نیز همه روابط ممکن بین گره‌ها ارزیابی شده و یک ساختار با

بلانی کرایدل با مقدار $4/42$ (mm.day^{-1}) و مدل تشعشعی فائو با مقدار $5/7$ (mm.day^{-1}) بودند. بر همین اساس و با در نظر گرفتن شاخص RMSE، در محل پژوهش به ترتیب مدل پرستلی تیلور با $8/72$ (mm.day^{-1})، مدل مکینک با $8/4$ (mm.day^{-1})، هارگریوزسامانی با $7/42$ (mm.day^{-1}) و مدل تورک با مقدار $7/1$ (mm.day^{-1}) در مقایسه با مدل مرجع فائوپنمن مانتیث ضعیف‌ترین نتایج را داشتند. با توجه به شاخص MR، مدل کیمبرلی پنمن با ۲ درصد نسبت به مدل فائوپنمن مانتیث، مقدار تبخیر و تعرق مرجع روزانه را کمتر برآورد کرده است. همچنین مدل پرستلی تیلور با ۹۱ درصد، مدل مکینک با ۸۳ درصد، هارگریوزسامانی با ۵۹ درصد، مدل تورک با ۳۹ درصد، مدل بلانی کرایدل با ۳۲ درصد و تشعشعی فائو با ۳۰ درصد نسبت به مدل فائوپنمن مانتیث، مقدار تبخیر و تعرق مرجع روزانه را بیشتر برآورد کردند. بر اساس شاخص AARE، بدترین مدل‌ها به ترتیب، مدل پرستلی تیلور با $1/63$ ، مدل مکینک با $1/54$ و هارگریوزسامانی با $1/23$ متوسط قدرمطلق خطای نسبی در برآورد تبخیر و تعرق مرجع و در مقابل مدل کیمبرلی پنمن با $0/52$ ، مدل تشعشعی فائو با $0/83$ ، مدل بلانی کرایدل با $0/86$ و مدل تورک با $0/97$ متوسط قدرمطلق خطای نسبی، مناسب‌ترین مدل در برآورد تبخیر و تعرق مرجع نسبت به مدل فائوپنمن مانتیث هستند. بر اساس ضریب تبیین، مدل‌های بلانی کرایدل، هارگریوزسامانی، تشعشعی فائو و مکینک مناسب‌ترین مدل‌ها معرفی شدند و در مقابل مدل‌های کیمبرلی پنمن، پرستلی تیلور و تورک به‌عنوان نامناسب‌ترین مدل‌ها شناخته شدند. در یک جمع‌بندی کلی و با در نظر گرفتن همه شاخص‌های ارزیابی در محل پژوهش نسبت به مدل فائوپنمن مانتیث، ترتیب مناسب بودن مدل‌ها این‌گونه است: مدل کیمبرلی پنمن، تشعشعی فائو، بلانی کرایدل، تورک، هارگریوزسامانی، مکینک و پرستلی تیلور.

رگرسیون گام به گام

برای بررسی تأثیر پارامترهای ورودی مدل شامل: درجه

$$R^2 = \frac{\left(\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \right)^2}{\left(\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2 \right)}, \quad 0 \leq R^2 \leq 1 \quad (6)$$

$$AARE = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - y_i|}{y_i} \quad (7)$$

$$MR = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{y_i} \quad (8)$$

در روابط بالا، X_i و Y_i به ترتیب مقادیر محاسباتی و مشاهداتی در گام زمانی i ام، N تعداد گام‌های زمانی، \bar{X} و \bar{Y} نیز به ترتیب میانگین مقادیر مشاهداتی و محاسباتی است. شاخص ضریب تبیین نشانگر دقت تخمین است که مقداری همواره بین ۱- و ۱+ دارد که علامت + و - بیانگر همبستگی مثبت یا منفی است و هرچه به ۱ نزدیک باشد نشان از همبستگی بیشتر بین مقادیر مشاهداتی و محاسباتی دارد. آماره ریشه میانگین مربعات همواره مثبت بوده و بهترین حالت عملکرد زمانی است که مقدار به صفر نزدیک شود. مقدار مثبت آماره متوسط قدرمطلق خطای نسبی بیانگر آن است که مدل مورد بررسی تبخیر- تعرق مرجع را بیشتر برآورد کرده و مقدار منفی آن بیانگر آن است که مدل مقدار تبخیر- تعرق را کمتر برآورد می‌کند و هرچه مقدار آن به صفر نزدیک باشد حاکی از دقت بالاتر است. شاخص نسبت میانگین به منظور بیش‌برآوردی یا کم‌برآوردی مدل‌ها نسبت به مقدار اندازه‌گیری شده، مورد استفاده قرار گرفت.

نتایج و بحث

ارزیابی مدل‌های تجربی برآورد تبخیر و تعرق مرجع

جدول ۴ دربردارنده مقادیر شاخص‌های $RMSE$ ، R^2 ، AARE و MR مدل‌های پرستلی تیلور، هارگریوزسامانی، بلانی کرایدل، مدل تشعشعی فائو، کیمبرلی پنمن، تورک و مکینک، نسبت به مدل مرجع فائوپنمن مانتیث است. بر اساس شاخص $RMSE$ مناسب‌ترین مدل‌ها، به ترتیب کیمبرلی پنمن با $2/77$ (mm.day^{-1})، مدل

جدول ۴. ارزیابی عملکرد هفت مدل برآورد تبخیر و تعرق مرجع نسبت به مدل فانو-پنمن-مانتیت

ردیف	مدل	R ²	MR	AARE	(mm.day ⁻¹) RMSE
۱	پریستلی - تیلور	۰/۸۴	۱/۹۱	۱/۶۳	۸/۷۲
۲	هارگریوز- سامانی	۰/۹۵	۱/۵۹	۱/۲۳	۷/۴۲
۳	بلانی کریدل	۰/۹۶	۱/۳۲	۰/۸۶	۴/۴۲
۴	تشعشعی فانو	۰/۹۵	۱/۳	۰/۸۳	۵/۷
۵	کیمبرلی- پنمن	۰/۸۲	۰/۹۸	۰/۵۲	۲/۷۷
۶	تورک	۰/۸۸	۱/۳۹	۰/۹۷	۷/۱
۷	مکینک	۰/۹۴	۱/۸۳	۱/۵۴	۸/۴

جدول ۵. مقدار ضرایب دقت مدل و پارامترهای رگرسیون گام به گام

خطای معیار شبیه‌سازی	ضریب تعیین تصحیح شده	ضریب تعیین	ضریب همبستگی چندگانه
۰/۹۰۳	۰/۹۰۷	۰/۹۰۷	۰/۹۵۲
سطح معنی‌داری	t	ضرایب استاندارد شده	ضرایب استاندارد نشده
		Beta	B
۰	-۴/۴۱۳	۰/۱۲۵	-۰/۵۵۳ (Constant)
۰	۱۳/۶۴۳	۰/۱۸۷	۰/۰۶۷ (Tmin)
۰	۲۹/۹۹۴	۰/۵۲	۰/۱۳۶ (Tmax)
۰	۱/۴۹۲	۰/۰۱۳	۰/۰۰۲ (Rhmin)
۰	-۱۴/۹۹۴	-۰/۱۳۱	-۰/۰۱۶ (Rhmax)
۰	۵۹/۸۶۷	۰/۲۸۳	۰/۴۴۸ (WS)
۰	۲۹/۵۳	۰/۲۰۱	۰/۱۶۴ (Sun)

مقادیر بتا و سطح معنی‌داری است. با مشاهده سطح معنی‌داری حاصل شده می‌توان گفت همه پارامترها در سطح یک درصد معنی‌دار بوده و نشان‌دهنده این موضوع است که پارامترهای ورودی همگی بر متغیر تبخیر- تعرق اثرگذار بوده و در پیش‌بینی مورد استفاده قرار می‌گیرند. مقایسه ضریب استاندارد شده به دست آمده برای پارامترهای ورودی نشان می‌دهد که بیشینه درجه حرارت با ضریب ۰/۵۲ دارای بیشترین تأثیر بر تبخیر- تعرق مرجع است و پیش‌بینی کننده قوی‌تری نسبت به بقیه متغیرهاست. بعد از آن به ترتیب سرعت باد، ساعات آفتابی، درجه حرارت حداقل، رطوبت نسبی حداکثر و حداقل دارای

حرارت حداکثر، درجه حرارت حداقل، رطوبت نسبی حداکثر، رطوبت حداقل، ساعات آفتابی و سرعت باد در برآورد تبخیر- تعرق مرجع و تعیین الگوی ورودی مدل‌های BN و GEP، تحلیل رگرسیونی گام به گام بین این پارامترها صورت گرفت که نتایج در جدول ۵ نشان داده شده است: در این جدول، مقدار ضریب همبستگی چندگانه (R)، ضریب تعیین (R²) و ضریب تعیین تصحیح شده (R² Adjusted) و ضرایب رگرسیون نشان داده شده است. در این مطالعه مقدار R² هم برابر ۰/۹۰۷ است که نشان‌دهنده کارایی و دقت مناسب مدل در پیش‌بینی تبخیر- تعرق مرجع است. یکی از بخش‌های مهم جدول (۵)

جدول ۶. روابط رگرسیونی بین تبخیر- تعرق مرجع و پارامترهای هواشناسی

شاخص ارزیابی		رابطه
آموزش	آزمون	
$R^2=0/80$ و $RMSE=3/11$	$R^2=0/73$ و $RMSE=3/71$	$ET_0 = -0/553 + 0/136 T_{max}$
$R^2=0/84$ و $RMSE=1/9$	$R^2=0/75$ و $RMSE=2/78$	$ET_0 = -0/553 + 0/136 T_{max} + 0/448 WS$
$R^2=0/89$ و $RMSE=1/27$	$R^2=0/79$ و $RMSE=2/38$	$ET_0 = -0/553 + 0/136 T_{max} + 0/448 WS + 0/164 Sun$
$R^2=0/90$ و $RMSE=1/29$	$R^2=0/80$ و $RMSE=2/40$	$ET_0 = -0/553 + 0/136 T_{max} + 0/448 WS + 0/164 Sun + 0/067 T_{min}$
$R^2=0/90$ و $RMSE=1/34$	$R^2=0/80$ و $RMSE=2/42$	$ET_0 = -0/553 + 0/136 T_{max} + 0/448 WS + 0/164 Sun + 0/067 T_{min} + 0/02 Rh_{min}$
$R^2=0/91$ و $RMSE=0/9$	$R^2=0/82$ و $RMSE=2/21$	$ET_0 = -0/553 + 0/136 T_{max} + 0/448 WS + 0/164 Sun + 0/067 T_{min} + 0/02 Rh_{min} - 0/016 Rh_{max}$

در این جدول، ET_0 : تبخیر و تعرق گیاه مرجع، T_{max} : درجه حرارت حداکثر، WS : سرعت باد، Sun : تعداد ساعات آفتابی، T_{min} : درجه حرارت حداقل، Rh_{min} : رطوبت نسبی حداقل و Rh_{max} : رطوبت نسبی حداکثر هستند.

ساختار ترکیبی شماره ۴ شامل T_{min} ، T_{max} و N و WS با عملگرهای پیش فرض مدل، با مقدار $RMSE=2/2$ و $R^2=0/91$ در مرحله آزمون و $RMSE=0/9$ و $R^2=0/98$ در مرحله آموزش، دارای کمترین خطا و بیشترین دقت بوده و توانسته است میزان ET_0 را با دقت مناسبی برآورد کند. در مقابل ترکیب شماره ۱ شامل T_{max} با چهار عملگر اصلی با مقدار $RMSE=2/42$ و $R^2=0/84$ در مرحله آزمون و $RMSE=1/3$ و $R^2=0/90$ در مرحله آموزش، دارای کمترین دقت بوده است. متر (۱۷) در پژوهش خود الگوی شامل T_{min} ، T_{max} و Rh و WS را با $RMSE=0/43$ و $R^2=0/99$ در مرحله آزمون و $RMSE=0/426$ و $R^2=0/99$ در مرحله آموزش و یاسین و همکاران (۲۷) در پژوهش خود الگوی شامل T_{min} ، T_{max} ، Rh_{min} ، Rh_{max} و N و WS با $RMSE=1/14$ و $R^2=0/95$ در مرحله آزمون و $RMSE=1/12$ و $R^2=0/95$ در مرحله آموزش را به عنوان مناسب ترین مدل انتخاب کردند.

شکل (۲- الف) مقادیر مشاهداتی و محاسباتی تبخیر- تعرق مرحله آموزش مدل را در طول دوره آماری نشان می دهد. همان طور که مشاهده می شود مدل توانسته است روند تغییرات ET_0 را به طور مناسبی پیش بینی کند. در شکل (۲- ب) مقادیر مشاهداتی در مقابل مقادیر محاسباتی آموزش مدل نسبت به نیمساز ناحیه اول ترسیم شده اند. بر اساس نمودار پراکنش (قسمت ب) مشاهده می شود که عملکرد مدل مناسب بوده

بیشترین تأثیر در پیش بینی و برآورد تبخیر- تعرق هستند. حسینی و همکاران (۱۰) نتیجه گرفتند که بیشینه درجه حرارت و سرعت باد بیشترین تأثیر را بر تبخیر- تعرق مرجع داشته است که با نتایج پژوهش حاضر مطابقت دارد. مقدار t (چه مثبت و چه منفی) اگر بزرگ تر از $1/96$ باشد و مقدار سطح معنی داری اگر کوچک تر از $0/01$ باشد نشان می دهد متغیر پیش بین بر متغیر پاسخ تأثیر معنی دار دارد. با توجه به مقادیر مربوط به هر شش متغیر، همه متغیرها دارای تأثیر معنی دار هستند.

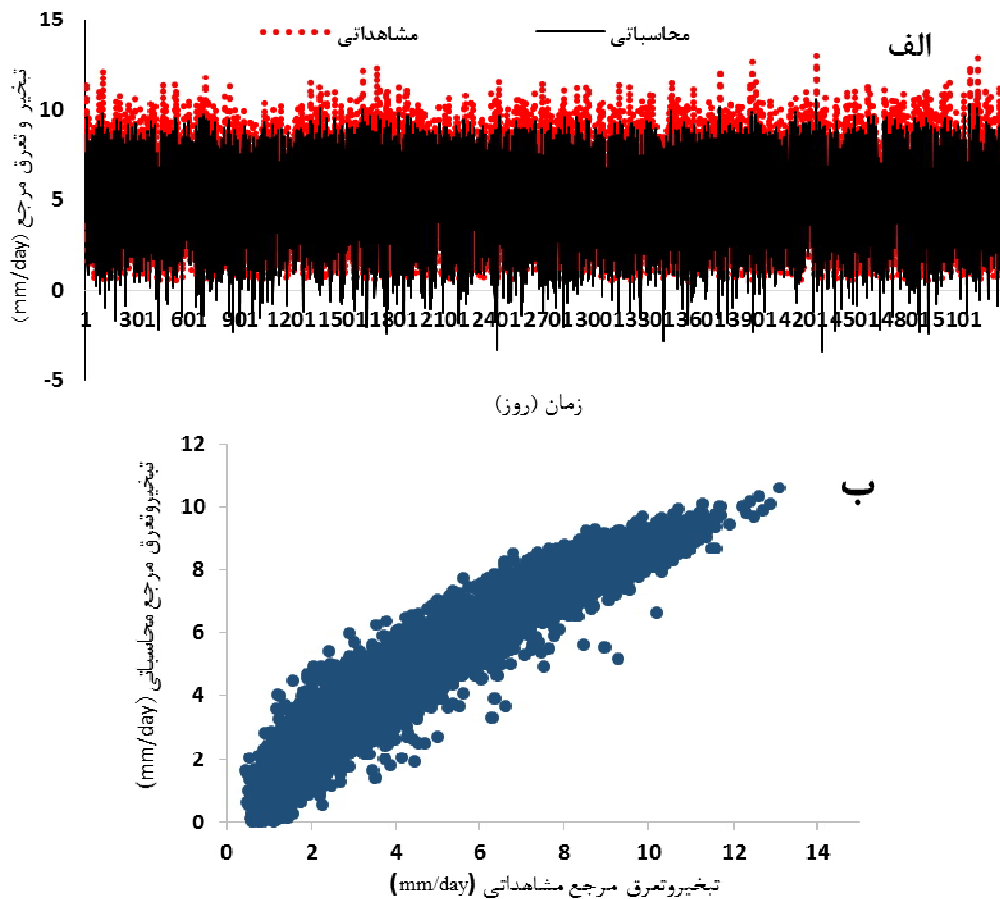
جدول ۶ نیز نشان دهنده رابطه رگرسیونی بین پارامترهای ورودی و تبخیر- تعرق مرجع بر اساس شش الگوی ورودی است. بر اساس شاخص های ارزیابی، الگوی شماره ۶ با $RMSE=2/21$ و $R^2=0/91$ در مرحله آموزش و $RMSE=0/9$ و $R^2=0/82$ در مرحله آزمون به عنوان مناسب ترین الگو معرفی شد.

مدل سازی تبخیر- تعرق مرجع به وسیله GEP

نتایج حاصل از اجرای مدل GEP برای داده های آموزش و آزمون در شش الگوی ورودی به همراه دو نوع عملگر ریاضی، در جدول ۷ ارائه داده شده است. همان طور که مشاهده می شود، برنامه ریزی بیان ژن توانسته است در تمام سناریوها عملکرد و دقت مناسبی از خود ارائه دهد. در بین سناریوهای انجام شده،

جدول ۷. روش برنامه‌ریزی بیان ژن با استفاده از دو مجموعه عملگر ریاضی

ردیف	الگو	عملگر	آموزش		آزمون	
			RMSE(mm/day)	R ²	RMSE(mm/day)	R ²
۱	M1	F1	۳/۱	۹۰/۰	۴۲/۲	۸۴/۰
		F2	۲۴/۱	۹۲/۰	۲/۳۵	۸۶/۰
۲	M2	F1	۱/۱	۹۶/۰	۳۱/۲	۸۸/۰
		F2	۹۲/۰	۹۷/۰	۲۲/۲	۹۱/۰
۳	M3	F1	۳/۱	۹۱/۰	۴۲/۲	۸۴/۰
		F2	۱۴/۱	۹۵/۰	۳۱/۲	۸۸/۰
۴	M4	F1	۱۲/۱	۹۶/۰	۳۲/۲	۸۸/۰
		F2	۹/۰	۹۸/۰	۲/۲	۹۱/۰
۵	M5	F1	۰۳/۱	۹۷/۰	۲۶/۲	۹۲/۰
		F2	۰۴/۱	۹۶/۰	۲۶/۲	۹۰/۰
۶	M6	F1	۲۵/۱	۹۴/۰	۳۸/۲	۸۵/۰
		F2	۹۳/۰	۹۵/۰	۲۲/۲	۹۱/۰



شکل ۲. نمودار حاصل از مقادیر بهینه مدل برنامه‌ریزی بیان ژن برای داده‌های ثبت شده مرحله آموزش: (الف) مقادیر محاسباتی و مشاهداتی نسبت به زمان و (ب) نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

جدول ۸. نتایج حاصل از مدل شبکه بیزین

آزمون		آموزش		الگو	ردیف
RMSE(mm/day)	R2	RMSE(mm/day)	R2		
۴۲/۲	۸۳/۰	۳/۱	۸۱/۰	M1	۱
۸۳/۲	۸۶/۰	۹۵/۱	۸۸/۰	M2	۲
۳۳/۳	۸۸/۰	۶۴/۲	۸۹/۰	M3	۳
۰۳/۳	۸۹/۰	۲۷/۲	۹۰/۰	M4	۴
۰۱/۳	۹۱/۰	۲/۲	۹۰/۰	M5	۵
۲۶/۲	۹۳/۰	۰۱/۱	۹۱/۰	M6	۶

مدل سازی تبخیر- تعرق مرجع به وسیله شبکه بیزین

در جدول ۸ نتایج حاصل از شبکه بیزین آورده شده است. همان طور که مشاهده می شود ساختارهای ترکیبی عملکرد بهتری را نسبت به ساختارهای ساده تر نشان می دهند به طوری که ساختار ترکیبی شماره ۶ شامل همه متغیرهای هواشناسی با $R^2=0/93$ و $RMSE=2/26$ (mm/day) در مرحله آزمون و $R^2=0/91$ و $RMSE=1/01$ (mm/day) در مرحله آموزش عملکرد بهتری نسبت به سایر ساختارها داشته و توانسته است میزان تبخیر- تعرق مرجع منطقه مورد مطالعه را با دقت مناسبی شبیه سازی کند. در مقابل ساختار شماره ۳ با $R^2=0/88$ و $RMSE=3/33$ (mm/day) در مرحله آزمون و $R^2=0/89$ و $RMSE=2/64$ (mm/day) در مرحله آموزش دارای بدترین عملکرد بوده است.

در شکل ۳ ساختار شبکه بیزین حاصل از الگوی منتخب نشان داده شده است. هدف از این روش استخراج ارتباط تبخیر- تعرق مرجع و پارامترهای تأثیرگذار بر آن است، که در نهایت منجر به ارائه رابطه ۱۰ شد.

$$ET0=0.133635*T_{max}+0.445122*WS \quad (10)$$

$$0.161776*SolarRadiation+0.0685841*T_{min}-$$

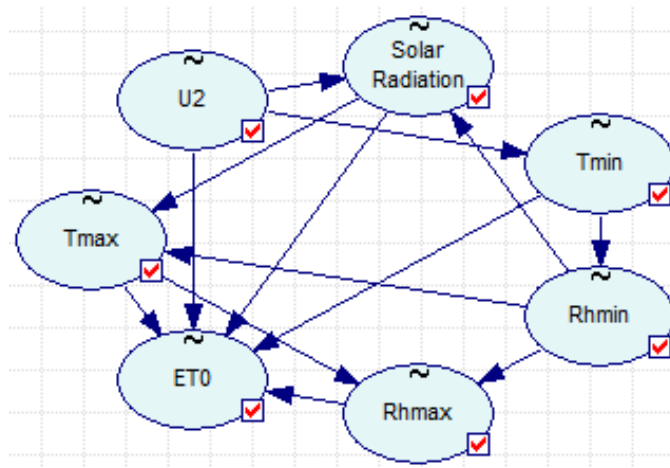
$$0.0151305*Rh_{max}+Normal(-0.471033,0.902415)$$

در شکل ۴- الف تغییرات مقادیر مشاهداتی و محاسباتی تبخیر- تعرق مرجع مرحله آموزش نسبت به زمان نشان داده شده

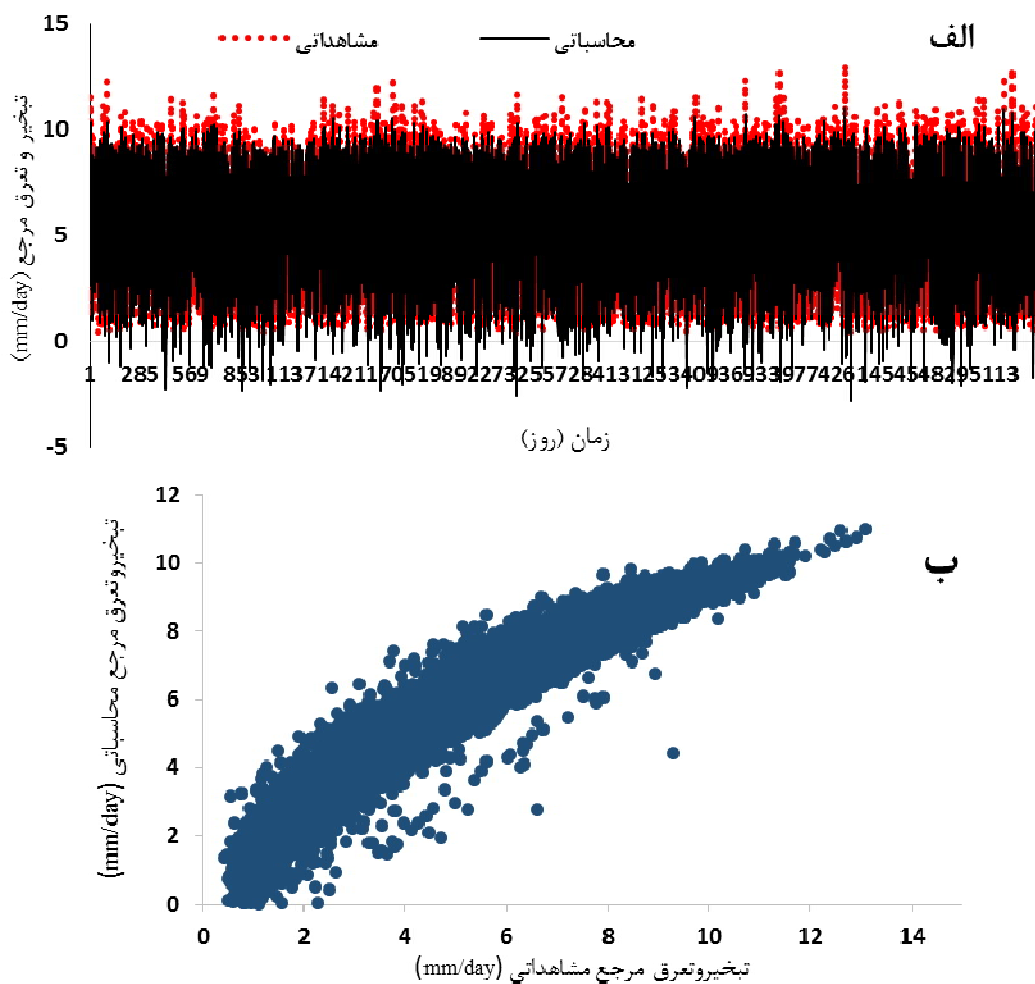
به طوری که در اکثر موارد، مقادیر محاسباتی و پیش بینی شده روی خط نیمساز ($X=Y$) واقع شده که بیانگر نزدیک بودن مقادیر مشاهداتی و محاسباتی است. علاوه بر آن مشاهده می شود که مدل در برآورد مقادیر حداقل عملکرد مناسبی از خود نشان داده به طوری که مقادیر تخمینی نزدیک به مقادیر واقعی هستند. ولی در برآورد برخی از مقادیر حداکثر از دقت آن کاسته شده است. در بیان این نتایج می توان اظهار کرد معادله حاصل از برنامه ریزی بیان ژن، از ترکیب تصادفی مجموع ترمینالها (متغیرها) و توابع به دست آمده است. بنابراین از مجموعه عملگرهای انتخابی از جمله \sin , \cos , \ln و ... برای استخراج رابطه ای استفاده می شود که ممکن است رابطه بین ورودی ها و خروجی های آن خطی باشد، که این امر باعث کاهش دقت مدل شده است.

پس از ارزیابی و صحت سنجی مدل، فرم ریاضی مدل استخراج شده با توجه به تابع پیوند جمع بین ژن ها، برای پیش بینی تبخیر- تعرق مرجع در ایستگاه مطالعاتی، با قرار دادن مقادیر ثابت و پارامترهایی که در ساختار بیان درختی مدل GEP به کار رفته، تعیین شد. رابطه ۹، نشان دهنده رابطه استخراج شده از GEP برای پیش بینی تبخیر- تعرق مرجع است:

$$ET0 = \operatorname{atan}(-4.77) + \operatorname{atan}(\operatorname{atan}(\operatorname{atan}(WS) - \operatorname{tanh}(\operatorname{gepMax}2((T_{max} - N), -10.01)))) + (((N + (WS - N))/2) + (((WS + T_{min})/2) + ((T_{max} + N)/2))/2) \quad (9)$$



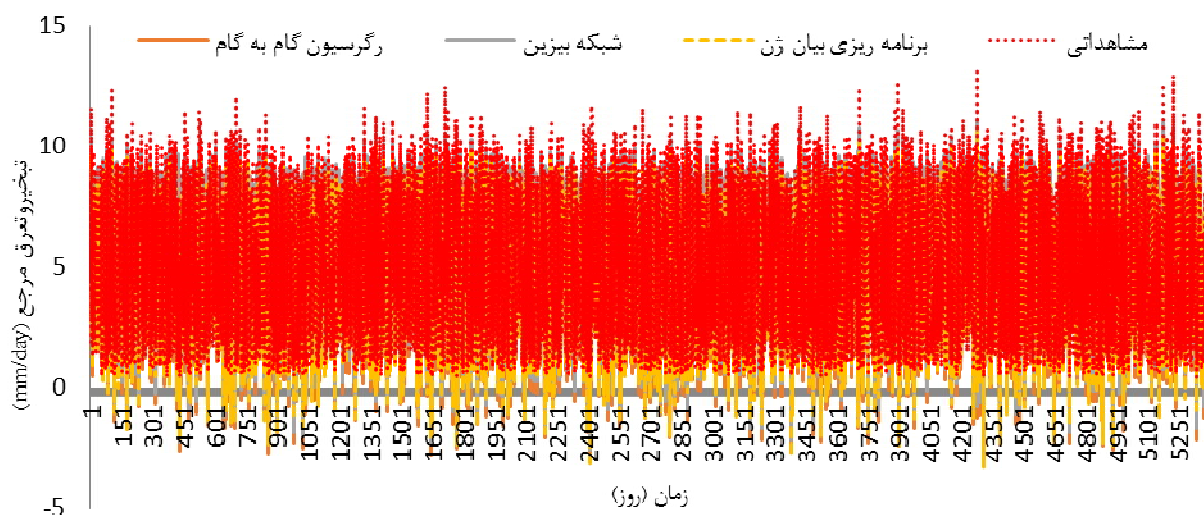
شکل ۳. ساختار شبکه بیزین به کار رفته برای شبیه‌سازی



شکل ۴. نمودار حاصل از مقادیر بهینه مدل شبکه بیزین برای داده‌های ثبت شده مرحله آموزش (الف) مقادیر محاسباتی و مشاهداتی نسبت به زمان و (ب) نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

جدول ۹. مقادیر شاخص‌های خطاسنجی مدل‌های برنامه‌ریزی بیان ژن، شبکه بیزین و رگرسیون گام به گام

نام مدل	RMSE (mm/day)	R ²	AARE	MR
برنامه‌ریزی بیان ژن	۰/۹	۰/۹۸	۰/۱۲	۰/۹۴
شبکه بیزین	۱/۰۱	۰/۹۱	۰/۲	۰/۹۰
رگرسیون گام به گام	۰/۹	۰/۹۱	۰/۱۴	۰/۹۲



شکل ۵. نمودار مقادیر محاسباتی و مشاهداتی مدل‌های مختلف نسبت به زمان

رگرسیون گام به گام در مرحله آموزش است. بر اساس این جدول و مقادیر شاخص‌ها، مدل برنامه‌ریزی بیان ژن به دلیل دارای بودن خطای کمتر ($RMSE=0/9$ و $AARE=0/12$) و دقت بیشتر ($R^2=0/98$) و کمترین کم‌برآوردی ($MR=0/94$) نسبت به مدل فائوپنمن‌مانتیت دارای بهترین عملکرد نسبت به سایر مدل‌ها بوده است و توانسته قابلیت بهتری را در مدل‌سازی ET_0 از خود نشان دهد.

در شکل ۵ نتایج بهینه به دست آمده از مدل‌ها نسبت به زمان نشان داده شده است. با توجه به این شکل مشاهده می‌شود که برنامه‌ریزی بیان ژن در برآورد مقادیر کمینه و بیشینه نسبت به مدل رگرسیون گام به گام و شبکه بیزین عملکرد بهتری داشته و توانسته است مقادیر تخمینی را نزدیک به مقادیر مشاهداتی پیش‌بینی کند. در مقابل شبکه بیزین در تخمین مقادیر کمینه عملکرد خوبی نداشته و نتوانسته است دقت مناسبی را ارائه دهد.

شکل ۴- ب نمودار پراکنش مقادیر تخمینی و مقادیر محاسباتی نسبت به نیمساز ربع اول را نشان می‌دهد. همان‌طور که در این شکل مشاهده می‌شود بیشتر مقادیر در اطراف خط نیمساز پراکنده شده‌اند که نشان از نزدیک بودن مقادیر محاسباتی به مقادیر مشاهداتی است. ولی در خصوص مقادیر بیشینه مدل عملکرد نامناسبی داشته و می‌توان اظهار کرد که مدل شبکه بیزین در تخمین مقادیر بیشینه ضعیف عمل کرده و مقادیر را کمتر از مقدار واقعی تخمین زده است که در شکل قابل ملاحظه است.

ارزیابی عملکرد مدل‌ها

در این بخش به ارزیابی دقت و عملکرد مدل‌های مورد استفاده در این مطالعه پرداخته می‌شود. جدول ۹ نشان‌دهنده شاخص‌های خطاسنجی مدل‌های برنامه‌ریزی بیان ژن، شبکه بیزین و

نتیجه گیری

تخمین مقادیر خروجی است که تعداد داده‌ها برای آموزش شبکه بیشتر شده و شبکه به‌خوبی تعمیم داده شود. همچنین مدل‌های مبتنی بر برنامه‌ریزی بیان ژن راه‌حلی را ارائه می‌کنند که می‌توان رابطه بین متغیرهای ورودی و خروجی را مشخص کرد. گفتنی است که در خصوص ترکیب‌هایی با متغیرهای کمتر نیز مدل‌ها جواب‌های خوبی ارائه کردند که می‌تواند در ایستگاه‌هایی با نواقص آماری در برآورد متغیر مورد نظر مورد استفاده قرار گیرند. نتایج حاصل از معیارهای ارزیابی نشان داد مدل برنامه‌ریزی بیان ژن دقت بالایی نسبت به مدل‌های تجربی، رگرسیون گام به گام و شبکه بیزین از خود نشان داده است. در پایان پیشنهاد می‌شود، از روش‌های جدیدتر مانند تئوری بازی‌ها در تعیین متغیرهای مؤثر و الگوهای ورودی مدل‌سازی و از روش‌های نوین هوشمند برای مدل‌سازی تبخیر- تعرق مرجع استفاده و با مدل‌های قدیمی‌تر مقایسه شود.

در این پژوهش عملکرد مدل‌های تجربی، رگرسیون گام به گام، برنامه‌ریزی بیان ژن و شبکه بیزین برای برآورد تبخیر- تعرق مرجع با استفاده از داده‌های روزانه ایستگاه بروجرد طی دوره زمانی ۱۳۷۵ تا ۱۳۹۶ مورد بررسی قرار گرفت. تبخیر- تعرق مرجع پیش‌بینی شده در مدل‌های مذکور، بر اساس معیارهای ارزیابی و تبخیر- تعرق مرجع مشاهداتی مورد مقایسه قرار گرفت. نتایج نشان داد از بین مدل‌های تجربی، مدل کیمبرلی‌پنمن در ایستگاه مطالعاتی دارای عملکرد بهتری است. نتایج به‌دست آمده از اجرای مدل‌ها با ساختارهای ورودی مختلف نشان داد که در هر سه مدل مورد بررسی، ساختارهای ترکیبی از عملکرد بهتری برخوردار هستند که نشان‌دهنده انتخاب مناسب پارامترهای ورودی و تأثیرگذار بودن آنها در مدل‌سازی بوده است. از طرفی افزایش تعداد متغیرهای مؤثر در ورودی موجب گسترش حافظه مدل برای

منابع مورد استفاده

1. Aguilera, P. A., A. Fernández, R. Fernández, R. Rumí and A. Salmerón. 2011. Bayesian networks in environmental modelling. *Environmental Modelling & Software* 26(12): 1376-1388.
2. Akhoni Pourhoseini, F. and A. Asadi. 2017. Application of Bayesian Network and Support Vector Machine in predicting of water table changes (case study: Ardabil plain). *Iran-Watershed Management Science & Engineering* 11(36): 33-42.
3. Allen, R. G., L. S. Pereira, D. Raes and M. Smith. 1998. Crop evapotranspiration-guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO, Rome* 300(9): D05109.
4. Almorox, J., V. H. Quej and P. Martí. 2015. Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology* 528: 514-522.
5. Blaney, H. F. and W. D. Criddle. 1950. Determining Water Requirements in Irrigated Areas from Climatological and Irrigation Data; Soil Conservation Service Technical Paper 96; Soil Conservation Service, US Department of Agriculture: Washington, DC, USA.
6. Chen, D., X. Wang, S. Liu, Y. Wang, Z. Gao, L. Zhang and X. Wei. 2015. Using Bayesian analysis to compare the performance of three evapotranspiration models for rainfed jujube (*Ziziphus jujuba* Mill.) plantations in the Loess Plateau. *Agricultural Water Management* 159: 341-357.
7. Davies, P. E. 2007. Bayesian decision networks for management of high conservation assets (national water Initiative-Australian government water fund. Report 6/6 Report to the Conservation of Freshwater Ecosystem Values Project, Water Resources Division, Department of Primary Industries and Water).
8. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*.
9. Hargreaves, G. L. and Z. A. Samani. 1985. Reference crop evapotranspiration from temperature. *Journal of Applied Research in Water Sciences* 1: 96-99.
10. Hosseini S., N. Ganji Khoramdeln and A. H. Khalat Abadi Farahani. 2015. Evaluation and sensitivity analysis of different methods of estimating daily reference evapotranspiration in a cold climate. *Journal of Applied Research in Water Sciences* 1(2): 29-40. (In Farsi).
11. Khoshhal, J., H. Zareabyane, A. R. Joshani and M. Khazae. 2017. Evaluation of different potential

- evapotranspiration methods with pan-FAO model in East and Northeast of Iran. *Journal of Natural Geography* 8(28): 1-16. (In Farsi).
12. Kool, D., N. Agam, N. Lazarovitch, J. L. Heitman, T. J. Sauer and A. Ben-Gal. 2014. A review of approaches for evapotranspiration partitioning. *Agricultural and Forest Meteorology* 184: 56-70.
 13. Legates, D. R. and G. J. McCabe. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35(1): 233-241.
 14. Lerner, U., R. Parr, D. Koller and G. Biswas. 2000. Bayesian fault detection and diagnosis in dynamic systems. In *Aaai/laai* 531-537.
 15. Makkink, G. F. 1957. Testing the Penman formula by means of lysimeters. *Journal of Institution Water Engineering* 11: 277-288.
 16. Martí, P., P. González-Altozano, R. López-Urrea, L. A. Mancha and J. Shiri. 2015. Modeling reference evapotranspiration with calculated targets. Assessment and implications. *Agricultural Water Management* 149: 81-90.
 17. Mattar, M. A. 2018. Using gene expression programming in monthly reference evapotranspiration modeling: a case study in Egypt. *Agricultural Water Management* 198: 28-38.
 18. Mehdizadeh, S., J. Behmanesh and K. Khalili. 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture* 139: 103-114.
 19. Priestley, C. H. B. and R. J. Taylor. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review* 100(2): 81-92.
 20. SadeghiHesar, A, H. Tabatabaee and M. Jalali. 2012. Monthly rainfall forecasting using Bayesian belief networks. *International Research Journal of Applied and Basic Sciences* 3: 2226.
 21. Shiri, J. 2017. Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran. *Agricultural Water Management* 188: 101-114.
 22. Shiri, J., A. H. Nazemi, A. A. Sadraddini, G. Landeras, O. Kisi, A. F. Fard and P. Marti. 2014. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Computers and Electronics in Agriculture* 108: 230-241.
 23. Sun, H., Y. Yang, R. Wu, D. Gui, J. Xue, Y. Liu and D. Yan. 2019. Improving estimation of cropland evapotranspiration by the Bayesian model averaging method with surface energy balance models. *Atmosphere*, 10(4): 188.
 24. Turc, L. 1961. Estimation of irrigation water requirements, potential evapotranspiration: A simple climatic formula evolved up to date. *Annals Agronomiques* 12: 13-49.
 25. Valipour, M., G. Sefidkouhi, M. Ali, M. Raeini-Sarjaz and S. M. Guzman. 2019. A hybrid data-driven machine learning technique for evapotranspiration modeling in various climates. *Atmosphere* 10(6): 311.
 26. Wright, J. L. and M. E. Jensen. 1972. Peak water requirements of crops in Southern Idaho. *Journal of Irrigation and Drainage Engineering ASCE* 96(1): 193-201.
 27. Yassin, M. A., A. A. Alazba and M. A. Mattar. 2016. Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agricultural Water Management* 163: 110-124.
 28. Zhang, B. Z., Y. Liu, D. Xu, N. N. Zhao, B. Lei, R. D. Rosa, P. Paredes, T. A. Paco and L. S. Pereira. 2013. The dual crop coefficient approach to estimate and partitioning evapotranspiration of the winter wheat-summer maize crop sequence in North China Plain. *Irrigation Science* 31(6): 1303-1316.

Evaluation of Experimental Models and Artificial Intelligence in Estimation of Reference Evapotranspiration (Case Study: Boroujerd Station)

Y. Sabzevari¹ and M. Saeidnia^{2*}

(Received: April 20-2020 ; Accepted: November 17-2020)

Abstract

The FAO Penman-Monteith is a baseline method to estimate reference evapotranspiration. In many cases, it is difficult to access all data, so replacing simpler models with lower input data and appropriate accuracy is necessary. The purpose of this study is to investigate the capability of the experimental models, gene expression programming, stepwise regression, and Bayesian network in estimating reference evapotranspiration. In this research, daily information of the Boroujerd synoptic station in the period of 1996 -2017 was used as model inputs. Based on the correlation between input and output parameters, six input patterns were determined for modeling. The results showed that the Kimberly-Penman model has the best performance among the experimental models. Gene expression programming with fourth pattern and Default Model Operators ($R^2 = 0.98$ and $RMSE = 0.9$), Bayesian Network with sixth pattern ($R^2=0.91$ and $RMSE = 1.01$), and stepwise regression with sixth pattern have the most accurate patterns at $R^2 = 0.91$ and $RMSE = 0.9$ in the training stage. Comparison of the performance of the three models showed that the gene expression programming model was superior to the other two models with the Average Absolute Relative Error (AARE) of 0.12 and the Mean Ratio (MR) of 0.94. The results showed that gene expression programming had an acceptable ability to estimate reference evapotranspiration under the weather conditions of Boroujerd and could be introduced as a suitable model.

Keywords: FAO-Penman-Monteith, Reference evapotranspiration, Bayesian network, Gene expression programming

1. Department of Water Engineering, Faculty of Agriculture, Isfahan University of Technology, Isfahan, Iran.

2. Department of Water Engineering, Faculty of Agriculture and Natural Resources, Lorestan University, Lorestan, Iran.

Corresponding author, Email: Saeedinia.m@lu.ac.ir