

رهیافت ترکیب وزنی مدل‌ها با روش‌های عکس واریانس و رگرسیون حداقل مربعات در تخمین

دمای نقطه شبنم

لاله پرویز*

دانشیار دانشکده کشاورزی دانشگاه شهید مدنی آذربایجان

(دریافت: ۱۴۰۰/۰۱/۱۳، بازنگری: ۱۴۰۰/۰۶/۱۴، پذیرش: ۱۴۰۰/۰۶/۲۰، انتشار آنلاین: ۱۴۰۰/۰۶/۲۰)

چکیده

بر آورد دمای نقطه شبنم با روشی کارآمد در بسیاری از زمینه‌ها مانند برنامه‌ریزی‌های بخش کشاورزی از جمله محافظت محصولات در برابر صدمات، مطالعات هواشناسی، هیدرولوژیکی و اکولوژیکی مفید می‌باشد. در این تحقیق از رهیافت ترکیب پیش‌بینی مدل‌های رگرسیون درختی، روش گروهی مدلسازی داده‌ها (GMDH) و روش تجربی جهت تخمین دمای نقطه شبنم در ایستگاه‌های رشت، یزد و ارومیه استفاده شد. داده‌های ورودی مدل‌های منفرد شامل داده‌های بیشینه سرعت باد، میانگین سرعت باد، دمای بیشینه، دمای کمینه، متوسط دما، متوسط رطوبت نسبی، کمینه و بیشینه رطوبت نسبی و فشار بخار اشباع بودند. رهیافت ترکیبی براساس رویکرد وزنی با روش‌های عکس واریانس و رگرسیون حداقل مربعات انجام گرفت. در مدل‌های منفرد، مدل GMDH نسبت به سایر مدل‌ها از کارایی بالایی برخوردار است به طوری که میزان کاهش RMSE از روش تجربی و رگرسیون درختی به مدل GMDH در ایستگاه رشت به ترتیب ۶۶/۶۶ و ۵۹/۴۵ درصد می‌باشد. رهیافت ترکیبی نسبت به مدل‌های انفرادی دقت عمل بالایی دارد و روش رگرسیون حداقل مربعات نسبت به روش عکس واریانس با معیارهای خطای متفاوت دارای خطای کمتری است، به طوری که محاسبه ضریب نش-ساتکلیف در ایستگاه‌های رشت، ارومیه و یزد به ترتیب ۰/۹۷، ۰/۹۶ و ۰/۸۷ می‌باشد. همچنین نوع معیار خطای مورد استفاده و توان تعریف شده در روش عکس واریانس در مقادیر پیش‌بینی موثر است و توان مناسب مربوط به خطا براساس داده‌های موجود ارائه شد. در بررسی تاثیر تنوع اقلیمی، ایستگاه رشت دارای کمترین خطا می‌باشد. همچنین عملکرد رهیافت ترکیبی با استفاده از مدل‌های منفرد قوی افزایش می‌یابد.

واژه‌های کلیدی: رهیافت ترکیب، عکس واریانس، رگرسیون حداقل مربعات، GMDH

مقدمه

که دمای نقطه شبنم با گذشت زمان به آرامی در حال افزایش است و می‌تواند یک متغیر مهم برای بررسی تغییرات طولانی مدت آب و هوایی باشد. هنگامی که دمای هوای سطحی به اندازه نقطه شبنم کاهش یابد، شبنم تشکیل می‌شود، بخصوص در مناطق خشک که بارندگی کمتری دارند شبنم می‌تواند برای بقاء گیاه ضروری باشد (Shank et al., 2008). بنابراین برآورد دمای نقطه شبنم با روشی کارآمد از اهمیت چشمگیری برخوردار است. در سال‌های گذشته روش‌های هوش مصنوعی در تخمین دمای نقطه شبنم موفقیت چشمگیری داشته‌اند (Amirmojahedi et al., 2016). تکنیک‌های شبکه عصبی و نروفازی در برآورد دمای نقطه شبنم در کره جنوبی مورد بررسی قرار گرفتند. داده‌های

دمای نقطه شبنم از پارامترهای کلیدی در مدل‌ها جهت برآورد تبخیر-تعرق، فشار واقعی بخار آب و یا رطوبت نسبی می‌باشد. همچنین ثابت شده است که تغییرات دمای نقطه شبنم ممکن است تاثیر قابل توجهی در تغییرات اقلیمی، در بررسی آسیب دیدگی محصول در هنگام یخزدگی و شناسایی باران و برف داشته باشد (Qaderi et al., 2020). همچنین می‌تواند در تعیین میزان رطوبت موجود در هوا مورد استفاده قرار گیرد، چرا که تخمین رطوبت نزدیک سطح از دیدگاه کشاورزی بسیار مهم است (Kisi et al., 2013). مدل‌های زیادی در زمینه‌های کشاورزی، اکولوژی، هیدرولوژی و اقلیم‌شناسی به دمای نقطه شبنم به عنوان یک متغیر ورودی نیاز دارند (Kim et al., 2015). راینسون (2000) نتیجه گرفت

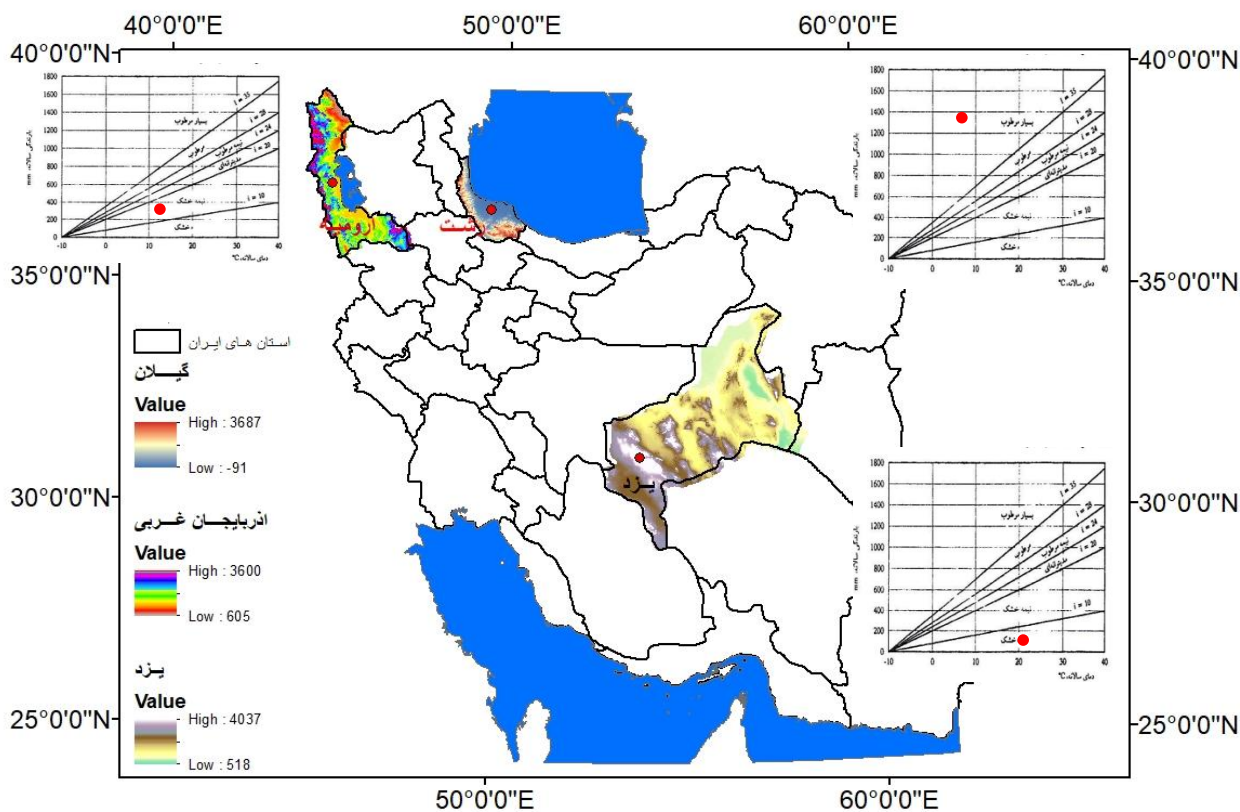
از رهیافت ترکیبی سه مدل جهت پیش‌بینی شاخص کیفیت هوا در چین استفاده شد که در مقایسه با مدل‌های منفرد دارای کارایی بالایی بود. همچنین روش عکس‌واریانس نسبت به خطای میانگین مربعات و میانگین وزنی ساده دقت بالای داشت (Song and Fu, 2020). از رهیافت ترکیبی ناهمگن چند دانه‌ای در پیش‌بینی قیمت نفت خام هم استفاده شد. مدل‌های منفرد مورد استفاده شامل رگرسیون خطی، شبکه عصبی مصنوعی و رگرسیون بردار پشتیبان بودند. رهیافت ترکیبی نسبت به مدل‌های منفرد دقت بالایی داشت، به عنوان نمونه کمترین مقدار RMSE در رگرسیون خطی و شبکه عصبی مصنوعی به ترتیب ۵/۴۲ و ۵/۸۹ و در رهیافت ترکیبی ۵/۰۳ بود (Wang et al., 2020). هدف این تحقیق تخمین دمای نقطه شبنم با رهیافت ترکیب پیش‌بینی مدل‌ها با استفاده از نه عنصر اقلیمی تعیین شده است. مدل‌های GMDH، رگرسیون درختی و روش تجربی به عنوان مدل‌های منفرد در نظر گرفته شدند. رهیافت ترکیب مدل‌ها براساس رویکرد وزنی با روش‌های عکس‌واریانس و رگرسیون حداقل مربعات انجام شده است و تاثیر تنوع اقلیمی در نتایج نیز مورد بررسی قرار می‌گیرد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

در این تحقیق جهت تخمین دمای نقطه شبنم از داده‌های هواشناسی ایستگاه‌های ارومیه، یزد و رشت در بازه زمانی ۱۹۹۱-۲۰۱۹ (اخذ شده از سازمان هواشناسی کل کشور) استفاده شد که موقعیت ایستگاه‌ها در شکل ۱ نشان داده شده است. انتخاب ایستگاه‌ها براساس بررسی تنوع اقلیمی بر نتایج مدل‌سازی است. براساس شاخص دومارتن اقلیم ایستگاه‌های رشت، ارومیه و یزد به ترتیب بسیار مرطوب، نیمه‌خشک و خشک طبقه‌بندی شده است و اقلیم نمای دومارتن نیز در شکل ۱ آورده شده است.

هواشناسی مورد استفاده شامل دمای هوا، ساعات آفتابی، سرعت باد، فشار بخار اشباع و رطوبت نسبی بودند. مدل‌های ANFIS و GRNM (Generalized Regression Neural Networks Model) (Kisi et al., 2013) در تخمین دمای نقطه شبنم، عملکرد مدل‌های شبکه عصبی و برنامه‌ریزی بیان ژن مورد ارزیابی قرار گرفتند. در این تحقیق برنامه‌ریزی بیان ژن نسبت به شبکه عصبی مصنوعی عملکرد قابل قبولی داشت (Shiri et al., 2014). از مدل‌هایی بر پایه روش گروهی مدل‌سازی داده‌ها (GMDH (Group Method of Data Handling) در تخمین دمای روزانه نقطه شبنم در استان کرمان استفاده شد. مدل‌های GMDH، مدل‌های توسعه‌یافته GMDH و دو مدل هیبرید بر پایه GMDH مورد استفاده قرار گرفت. نتایج نشان داد که مدل هیبرید در برآورد دمای نقطه شبنم دارای دقت بالایی است (Qaderi et al., 2020). با توجه به قابلیت مدل‌های بیان شده در پیش‌بینی دمای نقطه شبنم، رهیافتی برای افزایش کارایی مدل‌ها از اهمیت چشمگیری برخوردار است. ترکیب نتایج مدل‌ها یکی از رهیافت‌هایی است که می‌تواند صحت نتایج پیش‌بینی را افزایش دهد. ترکیب مدل‌ها جهت افزایش دقت پیش‌بینی بارش پاییزه با استفاده از دو شاخص SOI و NINO 3.4 در زیر حوضه سمیرم بکار گرفته شد. از روش وزن‌دهی رتبه‌بندی شده به مدل‌ها در ترکیب مدل‌ها استفاده شد. رهیافت ترکیبی نسبت به مدل‌های منفرد کارایی قابل ملاحظه‌ای داشت (مدرسی و همکاران، ۱۳۹۴). چندین طرح ترکیبی وزنی با استفاده از پنج مدل منفرد در پیش‌بینی ۹ سری زمانی مورد استفاده قرار گرفتند که برخی از آنها عبارتند از: روش‌هایی با پایه آماری، روش‌هایی بر پایه خطا، رگرسیون حداقل مربعات و روش‌های وزنی دهی دیفرانسیلی. برتری طرح‌های ترکیبی نسبت به مدل‌های انفرادی اثبات شده است. تاثیر تعیین دقیق وزن‌ها در رویکرد وزنی بر مقادیر پیش‌بینی از نتایج این تحقیق به شمار می‌رود (Adhikari and Agrawal, 2014).



شکل ۱: موقعیت ایستگاههای مطالعاتی در کشور و اقلیم نمای دومارتن ایستگاهها

۲-۲- رهیافت ترکیب پیش بینی مدلها

که $F: R^N \rightarrow R^N$ به صورت تابع ترکیب خطی و $W = [w_1, w_2, \dots, w_n] \in R^N$ بردار وزنی در نظر گرفته می شود. بنابراین طرح ترکیبی با معادله ۲ بیان می شود.

$$\hat{y}_k = w_1 \hat{y}_k^{(1)} + w_2 \hat{y}_k^{(2)} + \dots + w_n \hat{y}_k^{(n)} = \sum_{i=1}^n w_i \hat{y}_k^{(i)} \quad (2)$$

($\forall k = 1, 2, \dots, N$)

\hat{y}_k : مقادیر پیش بینی شده حاصل از مدل های مختلف، w : وزن های محاسبه شده (Adhikari and Agrawal, 2014).

جهت تعیین وزن ها روش های مختلفی وجود دارد که در ادامه به دو مورد اشاره می شود.

روش ترکیبی سعی در بهبود دقت پیش بینی کلی دارد. به طور کلی مدل های ترکیبی از چندین مدل منفرد برای پیش بینی متغیر مورد نظر استفاده می کنند و سپس معیارهای ترکیبی مناسبی (اختصاص وزن) را برای ادغام این مدل های منفرد برای مدل ترکیبی که ساخته می شود، اعمال میکنند. روش ترکیبی با استخراج اطلاعات نمونه از یک مدل منفرد دقت پیش بینی را افزایش می دهد (Song and Fu 2020). در یک روش ترکیبی خطی، پیش بینی ترکیبی از طریق یک تابع خطی از پیش بینی های منفرد محاسبه می شود. اگر $Y = [y_1, y_2, \dots, y_N]^T \in R^N$ مجموعه داده های سری زمانی واقعی که پیش بینی های آنها از n مدل مختلف بدست خواهد آمد، در نظر گرفته شود، یک ترکیب خطی از n مدل پیش بینی بصورت معادله ۱ بیان می شود.

$$\hat{Y} = F(f_1(Y), f_2(Y), \dots, f_n(Y), W) \quad (1)$$

$$e_i = \sum_{t=1}^N (y_{it} - \hat{y}_{it})^2 \quad i = 1, 2, \dots, n \quad (۴)$$

۲-۲-۲-۲ روش رگرسیون حداقل مربعات (LSR) (Least Square Regression)

در این روش تعیین وزن روش ترکیبی براساس کمینه‌سازی مجموع مربعات خطا (SSE) است. فرم کلی ترکیب خطی بیان شده را می‌توان بصورت ماتریسی (رابطه ۵) نوشت و مجموع مربعات خطای محاسبه شده براساس رابطه ۶ محاسبه می‌شود.

$$\hat{Y} = UW, \quad U = \begin{bmatrix} \hat{y}_1^{(1)} & \hat{y}_1^{(2)} & \dots & \hat{y}_1^{(n)} \\ \hat{y}_2^{(1)} & \hat{y}_2^{(2)} & \dots & \hat{y}_2^{(n)} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \hat{y}_N^{(1)} & \hat{y}_N^{(2)} & \dots & \hat{y}_N^{(n)} \end{bmatrix}, \quad w = [w_1, w_2, \dots, w] \quad (۵)$$

$$SSE = \sum_{k=1}^N (y_k - \hat{y}_k)^2 = (Y - UW)^T (Y - UW) = Y^T Y - 2W^T U^T Y + W^T U^T U W \quad (۶)$$

دمای نقطه شبنم استفاده شد و نحوه پیش‌بینی در شکل ۲ نشان داده شده است.

۲-۳-۱-۱ مدل GMDH

این الگوریتم اولین بار توسط Ivakhnenko (1966) معرفی شد. هدف اصلی آن شناسایی روابط مربوط به سیستم‌های چند بعدی غیر خطی بزرگ، تقریب و پیش‌بینی آنها بود.

۲-۲-۱-۱ روش عکس واریانس (Inverse variance - IV)

در این روش در ابتدا خطای پیش‌بینی مدل منفرد براساس یک معیار خطا مانند مجموع مربعات خطا SSE (Sum of Squared Error) محاسبه می‌شود. سپس براساس معیار خطای محاسبه شده، وزن کم به خطای بیشتر مدل منفرد و وزن بیشتر به خطای کمتر مدل منفرد تعلق می‌گیرد. بنابراین نحوه محاسبه وزن‌ها براساس رابطه ۳ است.

$$w_i = \frac{e_i^{-1}}{\sum_{i=1}^n e_i^{-1}}, \quad \sum_{i=1}^n w_i = 1 \quad (۳)$$

ε: نمایانگر معیار خطای مدل منفرد است که بصورت‌های مختلف تعریف می‌شود مانند مجموع مربعات خطا (رابطه ۴).

با کمینه‌سازی مجموع مربعات خطا با توجه به بردار وزن‌ها، بردار وزنی براساس رابطه ۷ محاسبه می‌شود.

$$W = (U^T U)^{-1} U^T Y \quad (۷)$$

۲-۳-۲ مدل‌های انفرادی

در این تحقیق از سه نوع مدلسازی به‌عنوان مدل‌های منفرد یعنی GMDH، رگرسیون درختی و رابطه تجربی در تخمین

در مقابل مقادیر عددی محاسبه می‌شود. اساس روش براساس تقسیم فضای X به مجموعه‌های جداگانه $A_1, \dots, A_j, \dots, A_z$ است. این تقسیم‌بندی را می‌توان با تقسیم‌های متوالی باینری متوالی براساس پیش‌بینی‌های مختلف بدست آورد، به طوری که گروه‌های حاصل از نظر متغیر پاسخ Y یکدست باشند. رگرسیون درختی یک مقدار متوسط $E(Y|X \in A_j)$ را فراهم می‌کند. نتیجه نهایی برگ‌ها و نقاط تقسیم باینری، گره نامیده می‌شوند و هر گره شامل چگونگی تقسیم است. برخی از مراحل مدلسازی را می‌توان به این صورت بیان کرد:

- ۱- شروع با گره ریشه و انجام تقسیم‌های ممکن برای هر یک از پیش‌بینی‌ها
- ۲- انتخاب بهترین تقسیم با استفاده از معیارهای برازش نکویی تقسیم و تقسیم دسته داده‌ها به گره‌های فرزند چپ و راست
- ۳- اگر الگوریتم مراحل فوق را برای هر یک از گره‌ها (در انتها نیستند) تکرار کند و درخت توسعه یابد، به خاطر خاصیت بازگشتی است (Samitas and Armenatzoglou, 2014).

GMDH ساختار بهینه را در فضای توابع چند جمله‌ای جستجو می‌کند که به عنوان یک شبکه چند جمله‌ای چند لایه‌ای تحقق می‌یابد. Ivakhnenko چند جمله‌ای کلموگروف-گابور را بکار برد که ثابت می‌کند هر تابع $y=f(x)$ را می‌توان با چند جمله‌ای نامحدود VKG (Volterra-Kolmogorov-Gabor) بازنمایی کرد.

(۸)

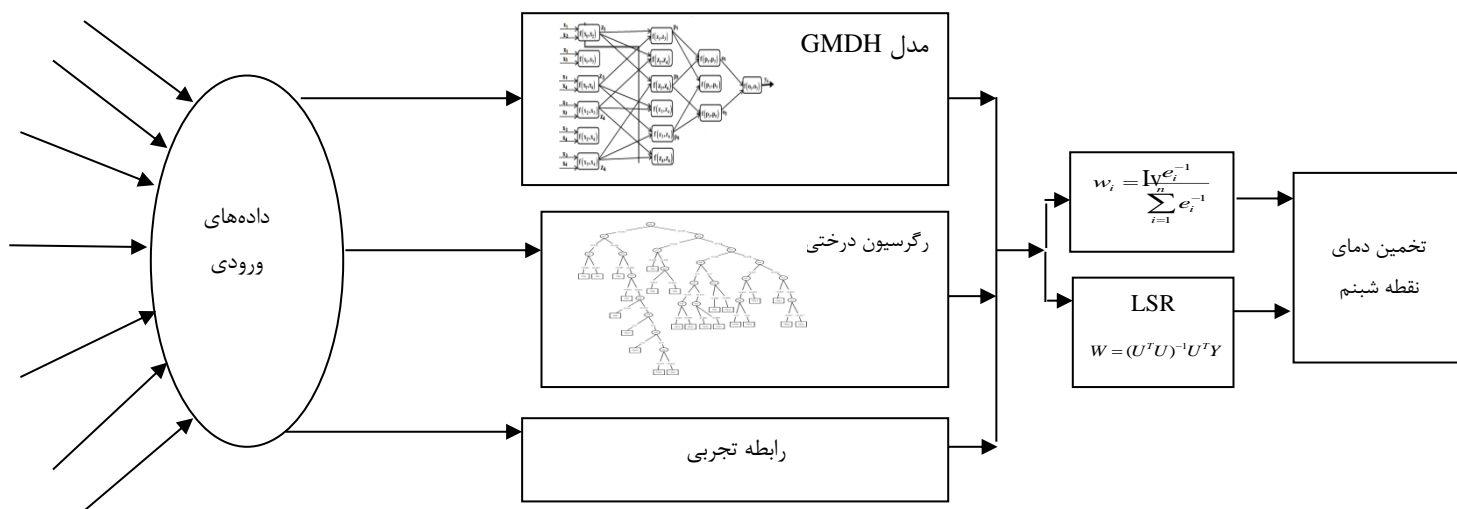
$$y_n = a_0 = \sum_{i=1}^M a_i x_i + \sum_{i=1}^M \sum_{j=1}^M a_{ij} x_i x_j + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M a_{ijk} x_i x_j x_k \dots$$

X : بردار ورودی، A : بردار ضرایب.

در الگوریتم GMDH سری VKG با چند جمله‌ای‌های مرتبه دوم با استفاده از جفت متغیرها تخمین زده می‌شود. شبکه را می‌توان از عناصر ساده چند جمله‌ای و تاخیری ساخت (Onwubolu et al., 2008).

۲-۳-۲- رگرسیون درختی

Breiman et al., (1984) برای اولین بار این رویکرد را زمینه های آماری وارد کردند. در مورد بخش طبقه‌بندی این نوع رویکرد مساله کلاس یا دسته مطرح است، ولی در مسایل رگرسیون درختی نیازی به تخصیص کلاس نیست و



شکل ۲: نحوه ترکیب پیش‌بینی مدل‌های منفرد جهت برآورد دمای نقطه شبنم

$$d = 1 - \frac{\sum_{i=1}^N |F_i - O_i|}{\sum_{i=1}^N (|F_i - \bar{O}| + |O_i - \bar{O}|)} \quad (15)$$

F: داده‌های پیش‌بینی شده

O: داده‌های مشاهداتی

N: تعداد داده‌ها جهت ارزیابی مدل

کمینه مقادیر MAPE، MSE، RMSE نشان‌دهنده مدلی کارآمد است. مقادیر نزدیک به یک ضریب نش-ساتکلیف و d بیانگر عملکرد قابل قبول مدل است. مقادیر بیشتر از یک GMER نشان‌دهنده بیش برآورد داده‌های پیش‌بینی شده است (Wang et al., 2020- Danandeh Mehr 2018, Chen et al., 2021).

نتایج

۳-۱- بررسی رابطه بین داده‌ها و دمای نقطه شبنم

در این تحقیق از داده‌های بیشینه سرعت باد، میانگین سرعت باد، دمای بیشینه، دمای کمینه، متوسط دما، متوسط رطوبت نسبی، کمینه و بیشینه رطوبت نسبی و فشار بخار اشباع در تخمین دمای نقطه شبنم به عنوان داده‌های ورودی استفاده شد. سال‌های ۱۹۹۱-۲۰۱۰ و ۲۰۱۱ تا ۲۰۱۹ به ترتیب به عنوان دوره‌های واسنجی و صحت‌سنجی در نظر گرفته شدند (تقسیم‌بندی داده‌ها براساس نسبت ۷۰-۳۰ می‌باشد). مقادیر ضرایب همبستگی بین دمای نقطه شبنم و داده‌های هواشناسی در شکل ۳ آورده شده است.

۳-۳-۲- روش تجربی

دمای نقطه شبنم را با استفاده از داده‌های رطوبت نسبی و متوسط دما، می‌توان بصورت تجربی با استفاده از رابطه ۹ بدست آورد.

$$T_d = (112 + 0.9T)(RH)^{0.125} + (0.1T - 112) \quad (9)$$

RH: رطوبت نسبی، T: دمای هوا (علیزاده، ۱۳۸۷).

۴-۲- معیارهای ارزیابی

در این تحقیق از معیارهای مختلفی از جمله:

RMSE (Root Mean Square Error)

MAPE (Mean Square Error)

GMER (Geometric Mean Error Ratio)

MAPE (Mean Absolute Percentage Error)

ضریب همسانی تعدیل شده (d) و ضریب

نش-ساتکلیف (Nash) برای ارزیابی عملکرد مدل‌های منفرد و رهیافت ترکیب پیش‌بینی مدل‌ها استفاده شد (معادلات ۱۰-۱۵).

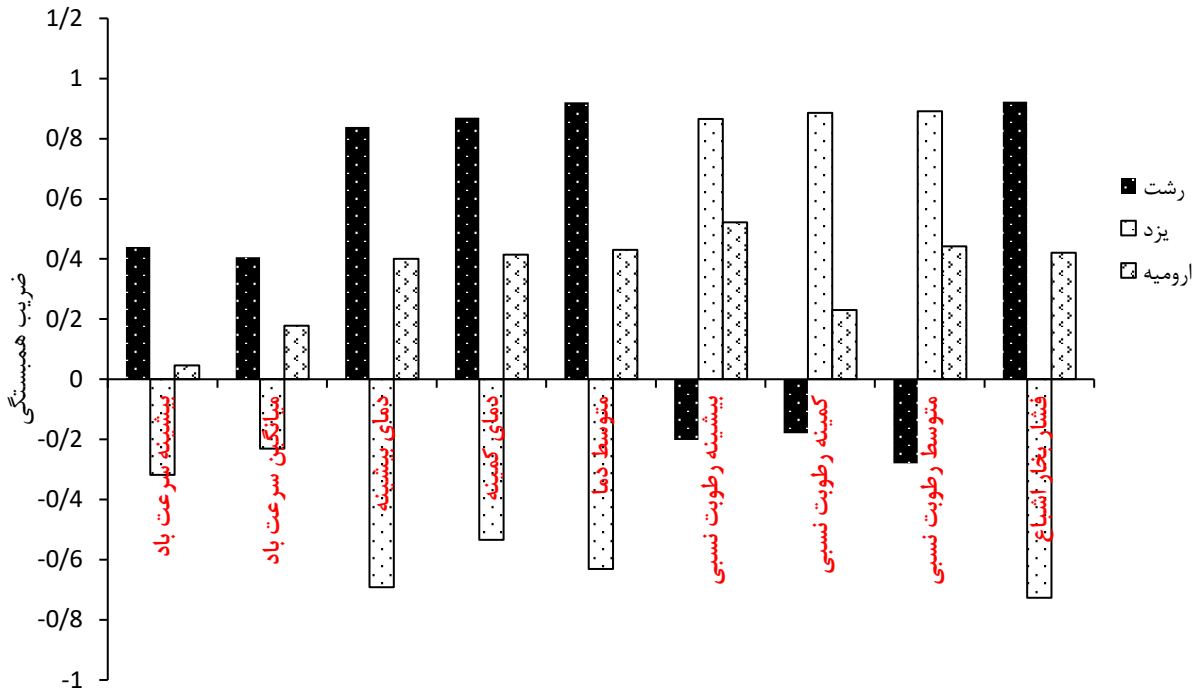
$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{O_i - F_i}{O_i} \right| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - F_i)^2} \quad (11)$$

$$Nash = 1 - \frac{\sum_{i=1}^N (F_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (12)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (O_i - F_i)^2 \quad (13)$$

$$GMER = \exp\left[\frac{1}{n} \sum_{i=1}^N \ln\left(\frac{F_i}{O_i}\right)\right] \quad (14)$$



شکل ۳. مقادیر ضرایب همبستگی دمای نقطه شبنم با داده‌های هواشناسی در ایستگاه‌های مطالعاتی

همبستگی بالا سرعت باد مشاهده نشد. در تحقیق Kisi et al., (2013) نیز سرعت باد به عنوان داده‌ای با تاثیر کم در برآورد دمای نقطه شبنم معرفی شد.

۲-۳- بررسی عملکرد مدل‌های انفرادی

در ادامه جهت تعیین دمای نقطه شبنم با استفاده از داده‌های هواشناسی از مدل‌های GMDH، رگرسیون درختی و روش تجربی استفاده شد که پارامترهای مدل‌های GMDH و رگرسیون درختی در جدول ۱ آورده شده است. قابل ذکر است پارامترهای موجود در جدول از تحلیل حساسیت بدست آمده‌اند.

بر اساس شکل ۳ در ایستگاه رشت بیشینه مقدار ضرایب همبستگی مربوط به فشار بخار اشباع، متوسط دما، دمای کمینه و دمای بیشینه، در ایستگاه یزد مربوط به متوسط رطوبت نسبی، کمینه رطوبت نسبی، بیشینه رطوبت نسبی و فشار بخار اشباع، در ایستگاه ارومیه مربوط به بیشینه رطوبت نسبی، متوسط رطوبت نسبی، متوسط دما و فشار بخار اشباع بوده است. به طور کلی می‌توان گفت که فشار بخار اشباع، دما و رطوبت نسبی به‌عنوان داده‌های موثر در تخمین دمای نقطه شبنم می‌باشند. در تحقیق جرعه نوش و سپاسخواه (۱۳۹۷) در تعیین نقطه شبنم روزانه از مقادیر دما و رطوبت نسبی نیز استفاده شده است. در مورد داده‌هایی با ضرایب

جدول ۱: پارامترهای مدل‌های GMDH و رگرسیون درختی در برآورد دمای نقطه شبنم

| مدل | پارامتر | رشت | ارومیه | یزد |
|---------------|---------------------------------|-----|--------|-----|
| GMDH | بیشینه تعداد نرون ها در یک لایه | ۲ | ۶ | ۲ |
| | بیشینه تعداد لایه ها | ۲ | ۳ | ۵ |
| | فشار انتخاب | ۰/۸ | ۰/۷ | ۰/۹ |
| رگرسیون درختی | کمینه تعداد گره‌ها برای تقسیم | ۷ | ۵ | ۵ |
| | کمینه تعداد برگ‌ها | ۲ | ۲ | ۲ |

تعیین دقیق پارامترهای مدل از اهمیت چشمگیری در افزایش کارایی مدل برخوردار است، به عنوان نمونه در ایستگاه یزد در مدل GMDH درصد کاهش RMSE از مدل با پارامترهای بیشینه تعداد نرون‌ها در یک لایه، بیشینه تعداد لایه‌ها و فشار انتخاب برابر با ۰/۷، ۳، ۲ به ۰/۹، ۵، ۲ برابر ۳۶/۷۹ درصد و در رگرسیون درختی درصد کاهش

از RMSE از مدل با پارامترهای کمینه تعداد گره‌ها و کمینه تعداد برگ‌ها برابر با ۵، ۶ به ۲، ۵ برابر با ۲۵/۷۵ درصد بوده است. بعد از تخمین پارامترهای مدل، عملکرد مدل‌های GMDH، رگرسیون درختی و روش تجربی با استفاده از برخی معیارهای ارزیابی مورد بررسی قرار گرفت که نتایج در جدول ۲ آورده شده است.

جدول ۲: مقایسه عملکرد مدل‌های GMDH رگرسیون درختی و روش تجربی در برآورد دمای نقطه شبنم

| نام ایستگاه | مدل‌های مورد استفاده | MSE | RMSE | MAPE | Nash | d |
|-------------|----------------------|-------|------|-------|-------|------|
| رشت | روش تجربی | ۰/۲۰۷ | ۰/۴۵ | ۰/۰۳۳ | ۰/۲۳ | ۰/۵۶ |
| | GMDH | ۰/۰۲۴ | ۰/۱۵ | ۰/۰۰۹ | ۰/۹ | ۰/۸۴ |
| | رگرسیون درختی | ۰/۱۳ | ۰/۳۷ | ۰/۰۲ | ۰/۴۸ | ۰/۵۸ |
| ارومیه | روش تجربی | ۲/۰۰۳ | ۱/۴۱ | ۰/۶۱ | -۲/۷۸ | ۰/۳ |
| | GMDH | ۰/۰۶۹ | ۰/۲۶ | ۰/۱ | ۰/۸۶ | ۰/۸۱ |
| | رگرسیون درختی | ۰/۳۶ | ۰/۶ | ۰/۲۳ | ۰/۳۱۱ | ۰/۶۲ |
| یزد | روش تجربی | ۱۴/۳۵ | ۳/۷۸ | ۱/۶۵ | -۷/۲۱ | ۰/۲۷ |
| | GMDH | ۰/۶۷ | ۰/۸۱ | ۰/۴۷ | ۰/۶۱ | ۰/۷۳ |
| | رگرسیون درختی | ۰/۹۷ | ۰/۹۸ | ۰/۴۶ | ۰/۴۴ | ۰/۵۳ |

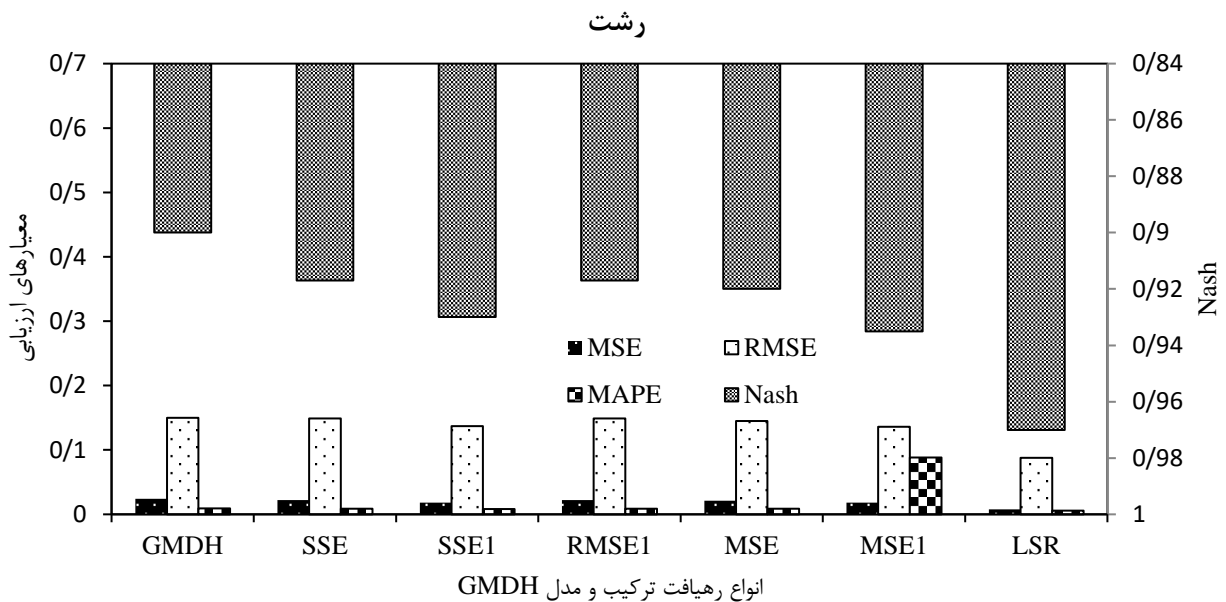
براساس نتایج جدول ۲، مقادیر آماره‌ها در صورت استفاده از مدل GMDH در بهترین حالت و با استفاده از روش تجربی در ضعیف‌ترین حالت از نظر عملکرد قرار دارند، به عنوان نمونه در ایستگاه رشت میزان کاهش MSE، RMSE، MAPE از روش تجربی به مدل GMDH به ترتیب برابر با ۸۸/۴، ۶۶/۶۶، ۷۲/۷۲ درصد و میزان افزایش d از رابطه تجربی به مدل GMDH برابر با ۵۰ درصد است. همچنین میزان کاهش MSE، RMSE، MAPE از رگرسیون درختی به مدل GMDH به ترتیب برابر با ۸۱/۵۳، ۵۹/۴۵ و ۵۵ درصد و میزان افزایش d در این حالت برابر با ۴۴/۸۲ درصد می‌باشد. در ایستگاه ارومیه میزان کاهش MSE، RMSE، MAPE از رابطه تجربی به مدل GMDH به ترتیب برابر با ۹۶/۵۵، ۸۱/۵۶ و ۸۳/۶ درصد و از رابطه تجربی به رگرسیون درختی به ترتیب برابر با ۸۲/۰۲ و ۲۹/۴۴ و ۶۲/۵۷ درصد بود که میزان درصد کاهش از رابطه تجربی به مدل GMDH بیشتر از

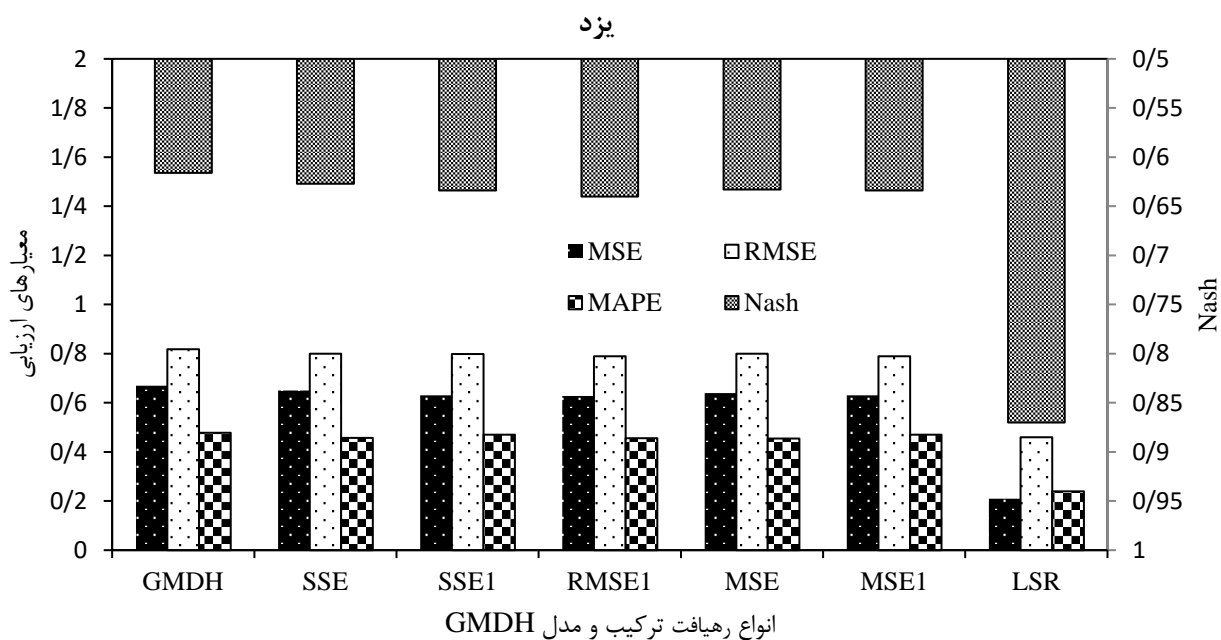
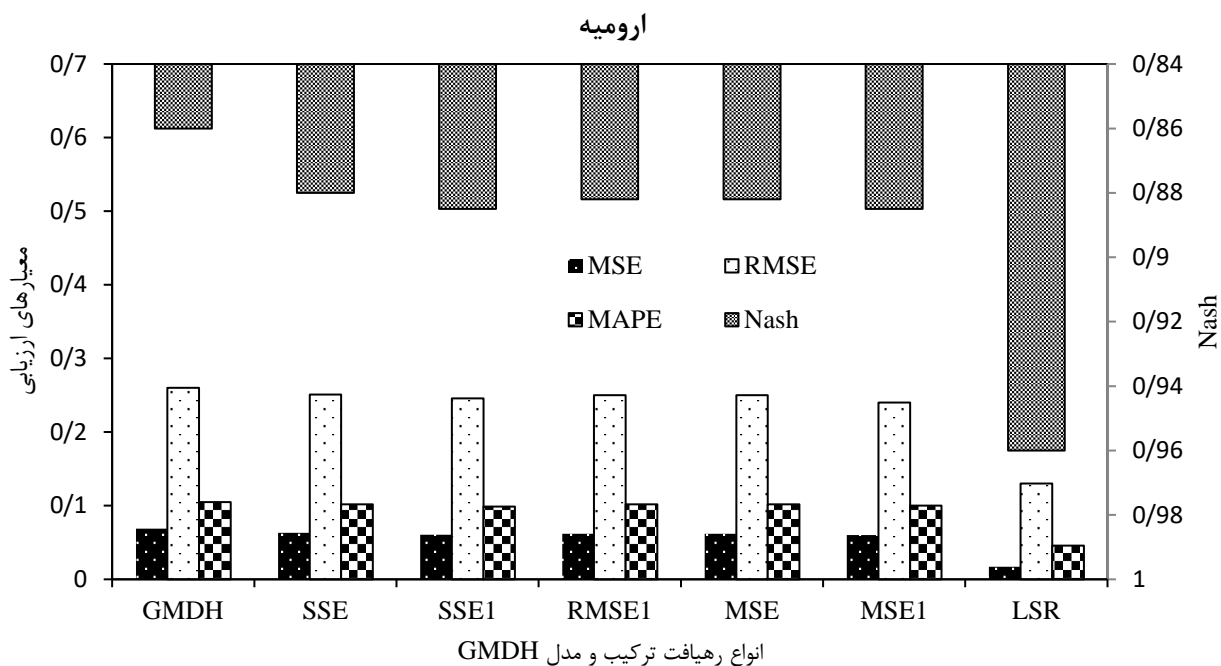
رابطه تجربی به رگرسیون درختی است. در یزد و ارومیه نیز میزان افزایش d از رگرسیون درختی به GMDH به ترتیب برابر با ۳۷/۷۳ و ۶۴/۳۰ درصد بود. بنابراین مدل GMDH نسبت به رگرسیون درختی و رابطه تجربی دارای عملکرد بهتری است. اگر مدل GMDH به عنوان مدل مبنا در نظر گرفته شود و در نتایج تاثیر تنوع اقلیمی مورد بررسی قرار گیرد، می‌توان گفت که ایستگاه‌های رشت و یزد به ترتیب دارای کمینه و بیشینه خطا (بیشینه و کمینه مقادیر ضریب نش و d) می‌باشند. میزان GMER مربوط به مدل GMDH در ایستگاه‌های رشت، ارومیه و یزد به ترتیب برابر با ۱/۰۰۶۳، ۱/۰۹ و ۱/۱ بود که بیانگر بیش‌برآورد در پیش‌بینی‌های مدل مورد استفاده است. در این حالت نیز مقدار GMER ایستگاه رشت کمترین مقدار را نسبت به سایر ایستگاه‌ها به خود اختصاص داده است.

۳-۳- بررسی عملکرد رهیافت ترکیبی

قابل ذکر است این مقایسه نسبت به بهترین مدل انتخاب شده مرحله قبل یعنی مدل GMDH انجام گرفته است. همچنین معیارهای خطای مورد استفاده در روش عکسواریناس شامل SSE ، $RMSE$ و MSE بودند. در شکل ۴ منظور از اندیس ۱ کنار معیار خطاها، نمایانگر بکارگیری توان منفی دو خطا در معادله ۳ است.

در ادامه تاثیر استفاده از رهیافت ترکیب مدل‌ها در نتایج پیش‌بینی مورد بررسی قرار می‌گیرد که روش واریانس معکوس (با معیارهای خطای متفاوت) و رگرسیون حداقل مربعات از روش‌های مربوط به رهیافت ترکیب پیش‌بینی مدل‌ها بودند و نتایج در شکل ۴ آورده شده است.





شکل ۴: نتایج بررسی تاثیر رهیافت ترکیب مدل‌ها با روش‌های مختلف در پیش‌بینی دمای نقطه شبنم

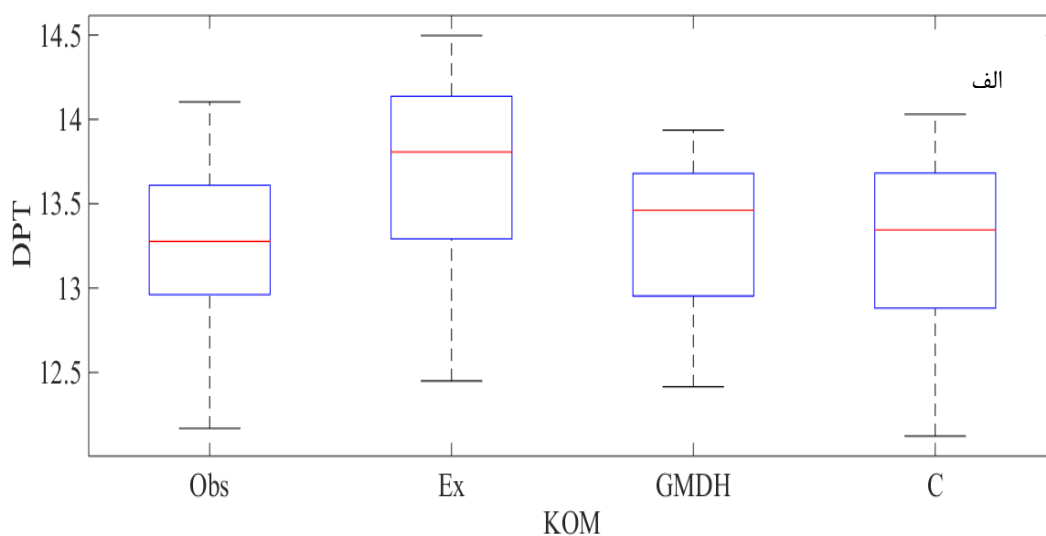
میزان کاهش MSE در ایستگاه رشت از مدل GMDH به روش‌های SSE، SSE1، RMSE، MSE، MSE1 و LSR به ترتیب ۸/۳۳، ۲۵، ۵/۳۳، ۲۵/۸، ۱۲ و ۶۷/۹۱ درصد میزان کاهش RMSE در ایستگاه ارومیه از مدل GMDH به

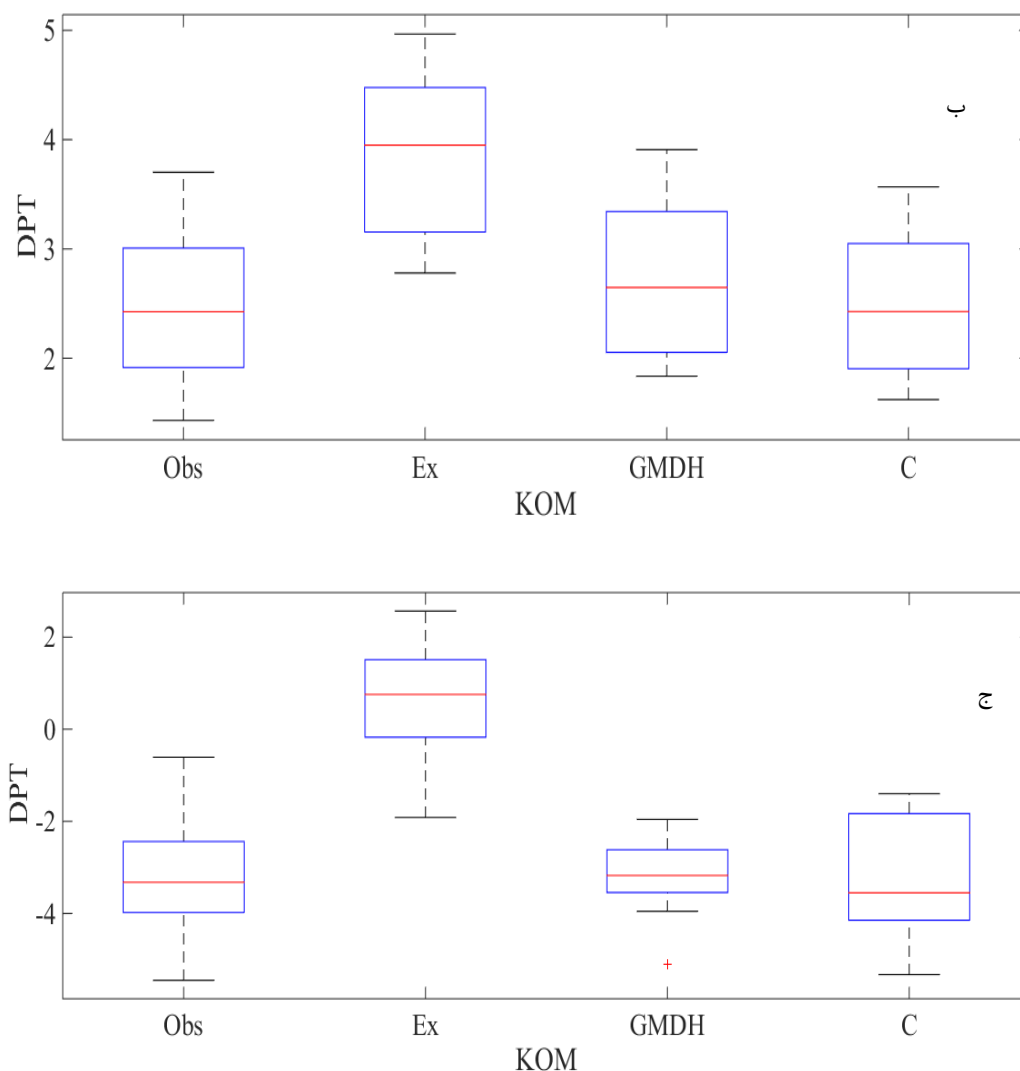
با توجه به شکل ۴ دو نکته مشهود است:

- ۱- تاثیر مثبت رهیافت ترکیبی در برآورد دقیق دمای نقطه شبنم
- ۲- افزایش دقت عملکرد روش رگرسیون حداقل مربعات نسبت به عکس‌واریانس.

در سه ایستگاه بیشتر از یک بود که بیانگر بیش برآورد در مقادیر پیش‌بینی است. در رهیافت ترکیبی نیز ایستگاه رشت نسبت به سایر ایستگاه‌ها وضعیت بهتری داشت، به طوری که مقادیر ضریب نش در روش رگرسیون حداقل مربعات در ایستگاه‌های رشت، ارومیه و یزد به ترتیب ۰/۹۷، ۰/۹۶ و ۰/۸۷ و مقدار ضریب d به ترتیب برابر با ۰/۹۲، ۰/۹۱ و ۰/۸۴ بود. میزان ضریب همبستگی بر نمودار پراکنش داده‌های مشاهداتی و پیش‌بینی در مدل‌های منفرد GMDH، رگرسیون درختی و رهیافت ترکیبی LSR در ایستگاه رشت به ترتیب برابر با ۰/۹۴۳، ۰/۵۴۵ و ۰/۹۷، در ایستگاه ارومیه به ترتیب ۰/۹۶، ۰/۵۲ و ۰/۹۶ و در ایستگاه یزد به ترتیب ۰/۶۳، ۰/۶۲ و ۰/۸۸ بود. بیشینه و کمینه مقدار دمای نقطه شبنم در ایستگاه رشت به ترتیب در سال‌های ۲۰۱۲ و ۲۰۱۱، در ایستگاه ارومیه در سال‌های ۲۰۱۲ و ۲۰۱۶ و در ایستگاه یزد در سال‌های ۲۰۱۹ و ۲۰۱۷ اتفاق افتاده است که این مساله در رهیافت ترکیبی LSR هم حفظ شده است. در ادامه نمودار جعبه ای داده‌های دمای نقطه شبنم مربوط به داده‌های مشاهداتی و پیش‌بینی با رهیافت ترکیبی و بهترین (GMDH) و ضعیف‌ترین (روش تجربی) مدل منفرد در طی دور صحت‌سنجی در شکل ۵ رسم شده است (در شکل منظور از DPT دمای نقطه شبنم به درجه سانتی‌گراد و KOM نماد انواع مدل‌ها است).

روش‌های SSE، SSE1، RMSE، MSE، MSE1 و LSR به ترتیب ۳۸/۴۶، ۸۴/۳، ۸۴/۵، ۶۹/۳، ۷/۳ و ۵۰ درصد، میزان کاهش MAPE در ایستگاه یزد از مدل GMDH به روش‌های SSE، SSE1، RMSE، MSE، MSE1 و LSR به ترتیب ۴/۳۹، ۱/۶۷، ۴/۶، ۴/۸، ۱/۶۷ و ۴۹/۷۹، میزان افزایش ضریب نش در ایستگاه رشت از مدل GMDH به روش‌های SSE، SSE1، RMSE، MSE، MSE1 و LSR به ترتیب ۲، ۲/۳، ۲، ۳/۲، ۸۸/۳ و ۷/۷ درصد است. با توجه به نتایج در رویکرد وزنی نسبت به مدل‌های منفرد بهبود آماره‌ها مشاهده شد. برتری عملکرد رهیافت ترکیب مدل‌ها نسبت به مدل‌های منفرد در تحقیق مدرسی و همکاران (۱۳۹۴)، Song and Fu 2020، Adhikari and Agrawal (2014) و Graeber and Kleine (2013) نیز مشاهده شد. در تحقیق Song and Fu 2020 در توان معادله ۳ در صورت و مخرج بجای منفی ۱ از منفی ۰/۵ استفاده شد که این مساله در این تحقیق مورد ارزیابی قرار گرفت ولی بهبودی در نتایج وجود نداشت. بنابراین بجای توان ۰/۵- توان ۲- مورد ارزیابی قرار گرفت که توان ۲- نسبت به توان ۱- از عملکرد بهتری برخوردار بود. در واقع در توان منفی دو سهم یا وزن پیش‌بینی‌های مربوط به مدل GMDH افزایش می‌یابد. همچنین نوع معیار خطا در نتایج رهیافت ترکیبی عکس واریانس موثر است به عنوان نمونه استفاده از معیار RMSE بهبودی در نتایج نداشت. مقادیر GMER در روش رگرسیون حداقل مربعات



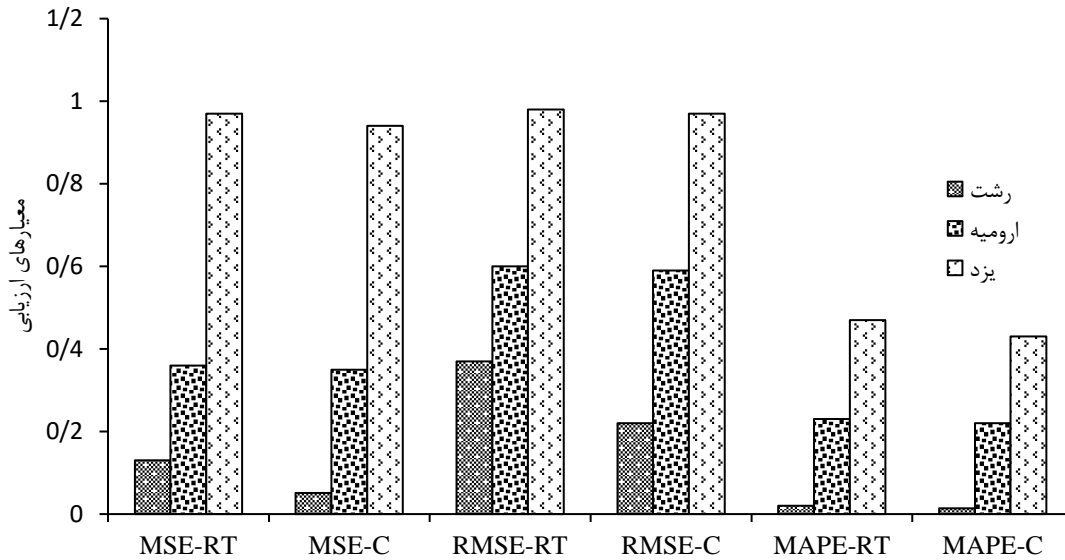


شکل ۵: نمودار مقادیر دمای نقطه شبنم؛ مشاهداتی (Obs) و پیش‌بینی با روش رگرسیون حداقل مربعات (C).

مدل GMDH و مدل تجربی (Ex) در ایستگاه‌های الف: رشت، ب: ارومیه و ج: یزد

تجربی از نظر میانه و خطوط Whikeres با داده‌های مشاهداتی تفاوت وجود دارد. در بررسی دیگر از ترکیب دو رابطه تجربی و رگرسیون درختی با حذف مدل کارآمد (GMDH) استفاده شد که در این حالت روش عکس‌واریانس با معیار مجموع مربعات خطا بکار گرفته شد و نتایج در شکل ۶ آورده شده است (در این شکل RT نماد رگرسیون درختی و C نماد ترکیب است).

در شکل ۵ نمودار جعبه‌ای داده‌های مشاهداتی و روش رگرسیون حداقل مربعات دارای تشابه زیادی هستند. میانه باکس پلات‌های مشاهداتی و رگرسیون حداقل مربعات بخصوص در ایستگاه‌های رشت و ارومیه تطابق زیادی دارند. در مورد خطوط Whikeres (مناطق مربوط به چارک اول و سوم) نیز این مساله صحت دارد. در مورد باکسپلات روش



شکل ۶: مقایسه رهیافت ترکیبی دو رابطه تجربی و رگرسیون درختی نسبت به عملکرد رگرسیون درختی

بر آورد دمای نقطه شبنم با روشی کارآمد در تحقیقات بخش کشاورزی، هواشناسی و اکولوژی از اهمیت چشمگیری برخوردار است. در سال‌های اخیر استفاده از تکنیک‌های هوش مصنوعی در این زمینه کاربرد زیادی داشتند، ولی در این تحقیق سعی در استفاده از رهیافت ترکیب پیش‌بینی مدل-ها براساس رویکرد وزنی است. مدل‌های مورد استفاده در رهیافت ترکیبی شامل مدل‌های GMDH، رگرسیون درختی و روش تجربی هستند. در مقایسه عملکرد سه مدل در پیش‌بینی دمای نقطه شبنم، مدل GMDH دارای عملکرد خوبی بود، چرا که این نوع مدل با در نظر گرفتن مجموعه‌ای از نرون‌ها و ترکیب چند جمله‌ای‌های درجه دوم دارای تقریب خوبی است. استفاده از رهیافت ترکیبی میزان دقت در مقادیر پیش‌بینی را افزایش می‌دهد چرا که مبنای عملکرد رهیافت ترکیبی استخراج اطلاعات از مدل‌های منفرد و ترکیب آنها با روشی کارآمد و دقیق است. تعیین وزن با روشی کارآمد در رهیافت ترکیبی با رویکرد وزنی از اهمیت چشمگیری برخوردار است که روش‌های عکس واریانس و رگرسیون حداقل مربعات به این منظور استفاده شدند. در این تحقیق در روش عکس واریانس از معیارهای خطای گوناگونی استفاده شد. برخی از معیارهای خطا مانند RMSE از کارایی لازم در افزایش دقت رهیافت ترکیب

بر اساس نتایج شکل ۶ رهیافت ترکیب پیش‌بینی مدل‌ها، دقت پیش‌بینی را نسبت به رگرسیون درختی و روش تجربی افزایش داده است، میزان کاهش RMSE از رگرسیون درختی به رهیافت ترکیبی در ایستگاه‌های رشت، ارومیه و یزد به ترتیب ۴۰/۵۴، ۱/۶۶ و ۱/۰۲ درصد و میزان کاهش IRMSE از رابطه تجربی به رهیافت ترکیبی در ایستگاه‌های رشت، ارومیه و یزد به ترتیب ۵۱/۱۱، ۵۸/۱۵ و ۷۴/۳۳ درصد بود. مساله‌ای که در اینجا باید به آن اشاره کرد میزان کاهش خطا از رهیافت ترکیبی (رابطه تجربی + رگرسیون درختی) به رهیافت ترکیبی (رابطه تجربی + رگرسیون درختی + GMDH) است، به طوری که درصد کاهش RMSE در ایستگاه‌های رشت، ارومیه و یزد از رهیافت دو مدلی به سه مدلی ۳۲/۲۷، ۵۷/۴۵ و ۱۷/۵۲ درصد بود. بنابراین استفاده از مدلی کارآمد در رهیافت ترکیبی میزان دقت پیش‌بینی را افزایش می‌دهد. در حالت دیگر از ترکیب رگرسیون درختی و GMDH استفاده شد که در این حالت مقادیر RMSE برای ایستگاه‌های رشت، ارومیه و یزد به ترتیب برابر با ۰/۱۴، ۰/۲۵ و ۰/۸ بوده است که نتایج مشابه حالت سه مدلی است و این مساله توانایی رهیافت ترکیبی با استفاده از مدلی کارآمد را در این فرآیند نشان می‌دهد.

نتیجه‌گیری کلی

بینی بارش پاییزه، نشریه هواشناسی کشاورزی، دوره ۳، شماره ۲، صفحات ۱ تا ۱۳.

- 4-Adhikari, R. and R.K. Agrawal, 2014, Performance evaluation of weights selection schemes for near combination of multiple forecasts, *Artificial Intelligence Review*, Vol. 42, pp.529-54
- 5-Amirmojahedi, M., K. Mohammadi, S. Shamshirband, A.S. Danesh, A. Mostafaeipour, and A. Kamsin, 2016, A hybrid computational intelligence method for predicting dew point temperature, *Environmental Earth Sciences*, Vol. 75, pp.1-12
- 6-Breiman, L., J. Friedman, R. Olshen and C. Stone. 1984, *Classification and regression trees*, Wadsworth and Brooks, Monterey
- 7-Chen, W., H. Xu, Z. Chen and M. Jiang, 2021, A novel method for time series prediction based on error decomposition and nonlinear combination of forecasters, *Neurocomputing*, Vol. 426, pp.85-103
- 8-Danandeh Mehr, A., 2018, An improved gene expression programming model for streamflow forecasting in intermittent streams, *Journal of hydrology*, Vol. 563, pp.669-678
- 9-Graeber, D. and A. Kleine, 2013, The combination of forecasts in the trading of electricity from renewable energy sources, *Journal of Business Economics*, Vol. 83, pp.409-435
- 10-Ivakhnenko, A.G., 1971, Polynomial theory of complex systems, *IEEE transactions on Systems, Man, and Cybernetics*, (4), pp.364-378
- 11-Kim, S., V.P. Singh, C.J. Lee and Y. Seo, 2015, Modeling the physical dynamics of daily dew point temperature using soft computing techniques, *KSCE Journal of Civil Engineering*, Vol. 19, pp.1930-1940
- 12-Kisi, O., S. Kim and J. Shiri, 2013, Estimation of dew point temperature using neuro-fuzzy and neural network techniques, *Theoretical and Applied Climatology*, Vol. 114, pp.365-373
- 13-Onwubolu, G.C., P. Buryan and F. Lemke, 2008, Modeling tool wear in end-milling using enhanced GMDH learning networks, *The International Journal of Advanced Manufacturing Technology*, Vol. 39, pp.1080-1092
- 14-Qaderi, K., B. Bakhtiari, M.R. Madadi and Z. Afzali-Gorouh, 2020, Evaluating GMDH-based models to predict daily dew point temperature (case study of Kerman province),

برخوردار نبود و بین عملکرد سایر معیارهای خط مانند SSE و MSE اختلاف فاحشی مشاهده نشد. همچنین توانی که در معادله عکس‌واریناس تعریف می‌شود، هم در دقت عمل مدل سهم مهمی دارد به طوری که توان ۲- در این تحقیق و با داده‌های موجود مقادیر خطا را در روند پیش‌بینی کاهش داد. در صورت استفاده از توانمندی ۲، وزن مربوط به مدل GMDH افزایش می‌یابد. بنابراین استفاده از معیار خطای جامع و توان مناسب، دقت روش عکس‌واریناس را افزایش خواهد داد. به‌طور کلی روش رگرسیون حداقل مربعات نسبت به روش عکس‌واریناس دارای کارایی بهتری بود. در بررسی تاثیر تنوع اقلیمی، ایستگاه‌های رشت و ارومیه نسبت به ایستگاه یزد عملکرد بهتری داشتند. در بررسی دیگر از رهیافت ترکیبی دو مدلی (بدون مدل GMDH) در پیش‌بینی دمای نقطه شبنم استفاده شد که مقادیر خطا نسبت به دو مدل استفاده شده کاهش داشت، ولی نسبت به رهیافت ترکیبی سه مدلی (با حضور GMDH) خطا افزایش پیدا کرد. در واقع افزایش دقت در رهیافت ترکیبی تابع روش مورد استفاده جهت تعیین وزن و نوع مدل مورد استفاده به‌عنوان مدل منفرد است. به‌طور کلی رهیافت ترکیبی با اتخاذ روش مناسب وزن‌دهی نسبت به مدل انفرادی دارای برآوردهای دقیق‌تری است.

سپاسگزاری

بدین وسیله از سازمان هواشناسی کل کشور به دلیل تامین داده‌ها جهت تهیه مقاله تشکر و قدردانی می‌شود.

منابع

- ۱- جرعه نوش، م. ه. و ع. سپاسخواه، ۱۳۹۷، پیش‌بینی سرمادگی با تخمین دمای حداقل روزانه در مناطق نیمه خشک، تحقیقات کشاورزی ایران، دوره ۳۷، شماره ۱، صفحات ۱۹ تا ۳۲.
- ۲- علیزاده، ا.، ۱۳۸۷، اصول هیدرولوژی کاربردی، انتشارات دانشگاه فردوسی مشهد، صفحه ۱۲۶.
- ۳- مدرسی، ف.، ش. عراقی نژاد و ک. ابراهیمی، ۱۳۹۴، ارزیابی راهبرد ترکیب مدل‌ها در افزایش دقت پیش

- Meteorology and Atmospheric Physics, Vol.132, pp.667-682
- 15-Robinson, P.J., 2000, Temporal trends in United States dew point temperatures, International Journal of Climatology: A Journal of the Royal Meteorological Society, Vol.20, pp.985-1002
- 16-Samitas, A. and A. Armenatzoglou, 2014, Regression tree model versus Markov regime switching: a comparison for electricity spot price modelling and forecasting, Operational Research, Vol.14, pp.319-340
- 17-Shank, D.B., G. Hoogenboom and R.W. McClendon, 2008, Dewpoint temperature prediction using artificial neural networks, Journal of applied meteorology and climatology, Vol.47, pp.1757-1769
- 18-Shank, D.B., R.W. McClendon, J. Paz and G. Hoogenboom, 2008, Ensemble artificial neural networks for prediction of dew point temperature, Applied Artificial Intelligence, Vol. 22, pp.523-542
- 19-Shiri, J., S. Kim and O. Kisi, 2014, Estimation of daily dew point temperature using genetic programming and neural networks approaches, Hydrology Research, Vol.45, pp.165-181
- 20-Song, C. and X. Fu, 2020, Research on different weight combination in air quality forecasting models, Journal of Cleaner Production, Vol.261, pp.1-16
- 21-Wang, J., H. Zhou, T. Hong, X. Li and S. Wang, 2020, A multi-granularity heterogeneous combination approach to crude oil price forecasting, Energy Economics, Vol.91, pp.1-9

Weight combinations approach of models with inverse variance and least square regression methods for dew point temperature estimation

L. Parviz^{*1}

1. Associate Professor, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran

(Received: 02 Apr 2021, Revised: 05 Sep 2021, Accepted: 11 Sep 2021, Published online: 11 Sep 2021)

Abstract

Dew point temperature estimation with proper method is useful in many fields such as agricultural planning including crop protection to the damages, meteorological, hydrological and ecological studies. In this study, the forecast combination approach of regression tree, Group Method of Data Handling (GMDH) and experimental method were applied to forecast dew point temperature in Rasht, Yazd and Urmia stations. The input variables of individual models were the maximum wind speed, mean wind speed, maximum, minimum and mean temperature, mean relative humidity, maximum and minimum relative humidity and saturation vapor pressure. The used methods of weight combination approach were inverse variance and least square regression. In the individual models, GMDH is more efficient than other models, so that the RMSE decreasing from empirical and regression tree to GMDH model in Rasht station is 66.66% and 59.45%, respectively. The combination approach is more accurate rather than the individual models and least square regression method has less error than the inverse variance with different error criteria, so that Nash-Sutcliffe coefficient in Rash, Urmia and Yazd stations is 0.97, 0.96 and 0.87, respectively. Also, the kind of error criteria and defined power in the inverse variance method is effective on forecasting values and the proper power basis on available data was proposed. In order to investigate the impact of climate diversity, Rasht station has the least error. The use of robust individual models will also increase the ability of forecast combination approach.

Key words: Combination approach, Inverse variance, Least squares regression, GMDH

* Corresponding author:

Email Address: Laleh_parviz@yahoo.com