



مدل‌بندی تابعی بارندگی ایران بر اساس دما و رطوبت

سید محمدابراهیم حسینی نسب[†]، نصیبه خیرالله‌زاده[‡] و نورالله تازی که میاندره[‡]

[†] دانشگاه شهید بهشتی

[‡] دانشگاه تربیت مدرس

چکیده. اخیراً تحلیل داده‌های تابعی به دلیل دسترسی آسان‌تر به داده‌هایی که ذاتاً به صورت تابع هستند پیشرفت قابل توجهی داشته است. مدل‌بندی بارندگی بر اساس دما و رطوبت با در نظر گرفتن ماهیت واقعی این پدیده‌ها که تابعی پیوسته از زمان هستند تا کنون به درستی در ایران انجام نشده است. داده‌های مربوط معمولاً به صورت روزانه یا ماهانه (گسسته) جمع‌آوری می‌شوند. اما اگر این داده‌های گسسته به صورت برداری فرض شوند و از روش‌های چندمتغیره برای تحلیل آن‌ها استفاده شود، مشکلات زیادی از جمله وجود بی‌نهایت جواب برای معادلات نرمال منتج از آن در رگرسیون وجود دارد.

با توجه به این که داده‌های گسسته‌ای اولیه برای بررسی باید با استفاده از روش مناسب به توابع پیوسته تبدیل شوند، به دلیل سادگی، رویکرد توابع پایه برای این کار مورد استفاده قرار می‌گیرد که خود به عنوان روشی برای کاهش بُعد داده‌های تابعی مطرح می‌باشد. در این مقاله، مشکلاتی که در نتیجه‌ای استفاده از روش‌های رگرسیون چندگانه به جای به کارگیری روش‌های رگرسیون تابعی رخ می‌دهند بحث می‌شود. در ادامه، با در نظر گرفتن مدل رگرسیون خطی تابعی ساده که در آن فقط یک متغیر مستقل تابعی (دما یا رطوبت نسبی) وجود دارد و متغیر پاسخ آن (لگاریتم مقدار بارندگی) اسکالر یا تابعی است ضرایب را برآورد می‌کنیم. پس از آن با توسعه‌ی مدل فوق و با وارد کردن هر دو متغیر مستقل تابعی به طور هم‌زمان به مدل، پارامترهای آن را برآورد می‌کنیم. در انتها نیز، به تفسیر و ارزیابی مدل‌های به دست آمده می‌پردازیم.

واژگان کلیدی. تحلیل داده‌های تابعی؛ رگرسیون تابعی؛ هموارسازی؛ توابع پایه.

۱ مقدمه

رگرسیون خطی چندگانه در زمینه‌های گوناگونی مانند علوم زیستی، رفتاری و اجتماعی برای مدل‌بندی روابط ممکن بین متغیرها استفاده می‌شود. از جمله دلایل محبوبیت این شاخه از آمار، سادگی معادله‌ی نهایی می‌باشد که برای برآورد روابط پیچیده‌ی بین متغیرهای مستقل و وابسته به دست می‌آید. به علاوه از آن برای اهداف دیگری از جمله خلاصه‌سازی داده‌ها، پیش‌بینی مقدار متغیر وابسته و برآورد ضرایب رگرسیونی استفاده می‌گردد. با توسعه‌ی مفاهیم رگرسیون برای حالتی که در آن داده‌ها تابعی هستند، اهداف فوق در این‌جا نیز قابل حصول می‌باشند.

در حالی که در رگرسیون خطی چندگانه متغیرهای مستقل و وابسته اسکالر هستند، در حالت تابعی حد اقل یکی از متغیرهای مستقل یا متغیرهای وابسته، تابعی از متغیری دیگر (معمولاً زمان) می‌باشد. با توجه به این که امکان استفاده از روش‌های کلاسیک برای برآورد ضرایب رگرسیون وقتی که مشاهدات تابعی هستند وجود ندارد و نیز با بررسی این داده‌ها در قالب داده‌های چندمتغیره بسیاری از مزایای تحلیل از بین می‌رود، بنا بر این ضرورت تحلیل داده‌های تابعی با در نظر گرفتن ماهیت اصلی آن‌ها بیش از پیش مشخص می‌شود (تازیکه میاندره و دیگران، ۱۳۸۷). برای مثال فرض کنید $Temp_{ij}$ مقدار مشاهده شده‌ی دما در ایستگاه هواشناسی i ام در روز j ام باشد. می‌خواهیم $\log Prec_i$ را بر اساس دما پیش‌بینی کنیم. برای این منظور مدل زیر را در نظر می‌گیریم:

$$\log Prec_i = \alpha + \sum_{j=1}^{365} \beta_j Temp_{ij} + e_i \quad i = 1, 2, \dots, 35.$$

این مدل شامل ۳۵ معادله و ۳۶۶ مجهول است. حتی اگر ماتریس ضرایب پُرتبه باشد دستگاه بی‌نهایت جواب دارد که هر کدام از جواب‌ها یک مدل مناسب برای داده‌های مشاهده شده را نتیجه می‌دهد. بنا بر این در عمل به دلیل یکتا نبودن جواب، استفاده از چنین روشی سودمند نیست (رمسی و سیلورمن، ۲۰۰۵). برای جزییات بیشتر و همچنین یک مرور کلی راجع به پیشینه‌ی رگرسیون خطی تابعی و کاربردهای آن می‌توان به حسینی‌نسب و دیگران (۱۳۸۷) رجوع کرد.

در بخش ۲ این مقاله، با استفاده از توابع پایه، هر تابع را با تعداد متناهی از توابع پایه تقریب می‌زنیم و بدین ترتیب بُعد را کاهش می‌دهیم. سپس در بخش ۳، پارامترهای رگرسیون خطی تابعی ساده را وقتی متغیر پاسخ اسکالر و تابعی است، برآورد خواهیم کرد. در بخش ۴، پارامترهای مدل را با دو متغیر مستقل وقتی که متغیر پاسخ اسکالر یا تابعی باشد برآورد خواهیم کرد. در بخش ۵، بر اساس مدل‌های ذکر شده در

بخش‌های قبلی به تحلیل داده‌های واقعی خواهیم پرداخت. بر اساس این داده‌ها، سعی می‌کنیم تا ارتباط بین مقدار بارندگی، دما و رطوبت را با استفاده از مفاهیم رگرسیون تابعی مدل‌بندی کنیم.

۲ هموارسازی با استفاده از توابع پایه

در این مقاله فرض می‌کنیم تمامی توابع مورد بررسی به فضای تابعی $L^2([0, T])$ متعلق هستند. از این رو، می‌توان هر یک از مشاهدات را به صورت ترکیب خطی شمارا از توابع پایه نوشت. در عمل معمولاً با انتخاب تعداد متناهی از جملات این ترکیب خطی، هر تابع را تقریب می‌زنیم. به عنوان مثال، اگر $\{\theta_k\}$ مجموعه‌ای از توابع پایه باشند و فرض کنیم θ یک بردار از توابع پایه به طول k_β باشد، آنگاه برای تابع $\beta(\cdot)$ می‌توان نوشت

$$(1) \quad \beta(t) \approx \sum_{k=1}^{k_\beta} b_k \theta_k(t) \quad \text{یا} \quad \beta(t) \approx \theta' b,$$

که در آن $b = (b_1, \dots, b_{k_\beta})'$ می‌باشد. به طور مشابه، چنین بسطی را برای مشاهدات مستقل X_i به صورت زیر داریم:

$$(2) \quad X_i(t) \approx \sum_{k=1}^{k_x} c_{ik} \psi_k(t) \quad \text{یا} \quad X(t) \approx C \psi(t),$$

که در آن C ماتریس $N \times k_x$ از ضرایب c_{ik} است. غیر از توابع یک بُعدی می‌توان رویه‌ی فوق را برای توابع دو یا چند بُعدی نیز تعریف نمود. تابع دو بُعدی $\beta(s, t)$ را در نظر بگیرید. این تابع را می‌توان با استفاده از مجموعه‌ای از توابع پایه $\{\theta_l\}$ و $\{\eta_k\}$ بسط داد، به طوری که اگر η یک بردار از توابع پایه به طول k_1 و θ یک بردار از توابع پایه به طول k_2 باشند، آنگاه برای تابع $\beta(\cdot, \cdot)$ داریم

$$(3) \quad \beta(s, t) \approx \sum_{k=1}^{k_1} \sum_{l=1}^{k_2} b_{kl} \theta_l(s) \eta_k(t) = \theta'(s) B \eta(t).$$

لازم به ذکر است که مقادیر k_β, k_x, k_1, k_2 (تعداد جملات بسط توابع پایه) می‌تواند بر اساس بینش تحلیل‌گر انتخاب شود، یا برای انتخاب آن‌ها می‌توان از روش‌های متداول دیگر مانند اعتبارسنجی متقابل استفاده نمود (هال و حسینی نسب، ۲۰۰۶).

۳ برآورد پارامترها در مدل رگرسیون خطی تابعی ساده

همان‌طور که در مقدمه اشاره شد، در حالت تابعی، رگرسیون خطی با یک متغیر مستقل تابعی، را در دو حالت می‌توان بررسی نمود. در ادامه، برآورد ضرایب رگرسیونی برای این دو حالت ارائه می‌شود.

۳/۱ حالت اول: متغیر پاسخ اسکالر

مدل رگرسیون خطی تابعی ساده به صورت

$$(۴) \quad y_i = \alpha + \int_0^T X_i(t)\beta(t)dt + \epsilon_i \quad 1 \leq i \leq N,$$

تعریف می‌شود که در آن y_i و ϵ_i اسکالر هستند، α و $\beta(\cdot)$ غیر تصادفی و زوج‌های $(X_1, \epsilon_1), \dots, (X_n, \epsilon_n)$ مستقل و هم‌توزیع‌اند. علاوه بر آن، توابع تصادفی X_i مستقل از خطاهای ϵ_i هستند (برای جزئیات بیشتر به باب‌روفسکی (۲۰۰۹) مراجعه شود). همچنین فرض می‌کنیم مقدار تابع X روی فاصله $[0, T]$ مشاهده شود.

برآورد تابعی $\beta(\cdot)$ مسئله‌ای با بُعد بی‌نهایت است. بنا بر این، در رگرسیون خطی تابعی، حل این مسئله مستلزم استفاده از روش‌های هموارسازی یا نظم می‌باشد تا به وسیله‌ی آن بتوانیم بُعد را کاهش دهیم. این مورد، وجه تمایز تحلیل رگرسیون تابعی در مقایسه با تحلیل رگرسیون خطی کلاسیک است (هال و حسینی‌نسب، ۲۰۰۹). هال و حسینی‌نسب (۲۰۰۶) نیز روشی را برای برآورد ضرایب رگرسیونی مدل (۴) ارائه کردند و سپس نرخ همگرایی آن را بدست آوردند. اگر بردار α را به طول N از عناصر مشابه α در نظر بگیریم و \hat{y} را بردار برآورد مشاهدات تعریف کنیم، آن‌گاه برای برآورد مدل (۴) با استفاده از (۱) و (۲) داریم

$$(۵) \quad \hat{y} = \alpha + \int_0^T X(t)\beta(t)dt = \alpha + \int_0^T C\psi(t)\theta'(t)'bdt = \alpha + CJ_{\psi\theta}b,$$

که در آن اندازه‌ی ماتریس $J_{\psi\theta}$ و $k_x \times k_\beta$ و $\int \psi(t)\theta'(t)dt$ می‌باشد. اگر

$$\zeta = (\alpha, b_1, \dots, b_k)'$$

بردار ضرایب رگرسیونی تابعی و $[1, CJ_{\psi\theta}] Z_{N \times (k_\beta + 1)} =$ آن‌گاه (۵) را

می‌توان به صورت $\hat{y} = Z\hat{\zeta}$ نمایش داد. بنا بر این برآورد کم‌ترین توان‌های دوم پارامتر ζ به صورت

$$(۶) \quad Z'Z\hat{\zeta} = Z'y,$$

می‌باشد که در آن $y = (y_1, \dots, y_N)'$. با توجه به این که می‌توان با وارد کردن عبارت تاوان همواری در ملاک کم‌ترین توان‌های دوم برآورد به دست آمده را بهبود بخشید، بنا بر این برآوردهای جدید را برای این حالت به دست می‌آوریم. در این حالت باید ضرایب رگرسیون $\beta(\cdot)$ و α طوری برآورد شوند که عبارت

$$\begin{aligned} \text{PENSS}_{\lambda}(\alpha, \beta) &= \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \int \{L\beta(t)\}^2 dt \\ &= (y - \hat{y})'(y - \hat{y}) + \lambda b'R_1 b \\ (۷) \quad &= (y - Z\zeta)'(y - Z\zeta) + \lambda \zeta'R\zeta, \end{aligned}$$

مینیم شود، که در آن L یک عملگر دیفرانسیل خطی است. همچنین λ را می‌توان به طور اختیاری یا از طریق اعتبار سنجی متقابل انتخاب کرد. به علاوه $R_1 = \int (L\theta(t))(L\theta(t))' dt$ و R به صورت

$$R = \begin{pmatrix} \circ & \circ \\ \circ & R_1 \end{pmatrix},$$

می‌باشد. با مشتق‌گیری از عبارت (۷) نسبت به ζ و مساوی صفر قرار دادن حاصل، داریم

$$(۸) \quad (Z'Z + \lambda R)\hat{\zeta} = Z'y.$$

۳/۲ حالت دوم: متغیر پاسخ تابعی

مدل رگرسیون خطی تابعی ساده وقتی که متغیر پاسخ تابعی باشد به صورت

$$(۹) \quad y_i(t) = \alpha(t) + \int_0^T X_i(s)\beta(s, t)ds + \epsilon_i(t) \quad 1 \leq i \leq N,$$

نمایش داده می‌شود که در آن $y_i(\cdot)$ و $\epsilon_i(\cdot)$ تصادفی و تابعی هستند، توابع $\alpha(t)$ و $\beta(s, t)$ غیر تصادفی و زوج‌های $(X_1(\cdot), \epsilon_1(\cdot)), \dots, (X_n(\cdot), \epsilon_n(\cdot))$ مستقل و هم‌توزیع‌اند. علاوه بر آن، توابع تصادفی

$X_i(\cdot)$ مستقل از خطاهای $\epsilon_i(\cdot)$ هستند. همچنین فرض می‌کنیم مقدار توابع $X_i(\cdot)$ و $y_i(\cdot)$ روی فاصله‌ی $[0, T]$ مشاهده شود.

برای برآورد توابع $\alpha(\cdot)$ و $\beta(\cdot, \cdot)$ از روش‌های هموارسازی یا روش‌های نظم استفاده می‌کنیم. برای این منظور، فرض کنید مشاهدات به صورت $(X_1(\cdot), y_1(\cdot)), \dots, (X_N(\cdot), y_N(\cdot))$ باشند که در آن توابع $X_i(\cdot)$ متغیر مستقل و توابع $y_i(\cdot)$ متغیر وابسته باشند. آنگاه با استفاده از (۱)، (۲) و (۳) برای توابع $X_i(s), y_i(t), \alpha(t), \beta(s, t)$ می‌توان به ترتیب آن‌ها را بر حسب مجموعه‌ای از توابع پایه $\{\eta_k\}, \{\theta_k\}$ و $\{\eta_k\}, \{\theta_k\}$ به طول k_1, k_2, k_3 و (k_1, k_2) نوشت. بنا بر این

$$(10) \quad y_i(t) \approx \sum_{k=1}^{k_1} d_{ik} \eta_k(t) \quad \text{یا} \quad \mathbf{y}(t) \approx \mathbf{D} \boldsymbol{\eta}(t),$$

که در آن \mathbf{D} ماتریس $N \times k_1$ از ضرایب d_{ik} می‌باشد. حال برای برآورد مدل (۹) با استفاده از (۱) و (۳) نتیجه می‌شود

$$\begin{aligned} \hat{\mathbf{y}}(t) &= \boldsymbol{\alpha}(t) + \int_0^T \mathbf{X}(s) \beta(s, t) ds \\ &= \mathbf{a}' \boldsymbol{\eta}(t) + \int_0^T \mathbf{C} \boldsymbol{\theta}(s) \boldsymbol{\theta}'(s) \mathbf{B} \boldsymbol{\eta}(t) ds \\ &= \mathbf{a}' \boldsymbol{\eta}(t) + \mathbf{C} \mathbf{J}_{\theta\theta} \mathbf{B} \boldsymbol{\eta}(t), \end{aligned}$$

که در آن اندازه‌ی ماتریس $\mathbf{J}_{\theta\theta}$ ، $k_2 \times k_2$ و $\mathbf{J}_{\theta\theta} = \int \boldsymbol{\theta}(s) \boldsymbol{\theta}'(s) ds$ می‌باشد. به علاوه $\boldsymbol{\zeta} = (\mathbf{a}, \mathbf{B}')' = (\alpha, \dots, \alpha)$ و $\boldsymbol{\alpha}_{N \times 1} = (\alpha, \dots, \alpha)$ ماتریس $N \times k_2$ از ضرایب c_{ik} می‌باشد. اگر $\boldsymbol{\zeta} = (\mathbf{a}, \mathbf{B}')'$ بردار ضرایب رگرسیونی تابعی و $\mathbf{Z}_{N \times (k_2+1)} = [\mathbf{1}, \mathbf{C} \mathbf{J}_{\theta\theta}]$ ماتریس ضرایب باشد، آنگاه مدل (۹) به صورت $\hat{\mathbf{y}} = \mathbf{Z} \boldsymbol{\zeta} \boldsymbol{\eta}$ نوشته می‌شود. بنا بر این معادله‌ی کم‌ترین توان‌های دوم برای برآورد بردار $\boldsymbol{\zeta}$ ، که از مینیم کردن $\int_0^T (\mathbf{y}(t) - \hat{\mathbf{y}}(t))' (\mathbf{y}(t) - \hat{\mathbf{y}}(t)) dt$ به دست می‌آید برابر است با

$$(11) \quad \mathbf{Z}' \mathbf{Z} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta} = \mathbf{Z}' \mathbf{D} \mathbf{J}_{\eta\eta},$$

که در آن $\mathbf{J}_{\eta\eta} = \int \boldsymbol{\eta}(t) \boldsymbol{\eta}'(t) dt$ است.

برای به دست آوردن فرم بسته‌ای برای $\boldsymbol{\zeta}$ از حاصل ضرب کرونگر استفاده می‌کنیم. اگر $\text{vec}(\mathbf{B})$

بردارى باشد که از روی هم قرار دادن ستون‌های ماتریس B حاصل شود، آن‌گاه

$$\text{vec}(ABC') = (C \otimes A)\text{vec}(B).$$

از این رو، معادله‌ی (۱۱) به صورت زیر بازنویسی می‌شود:

$$(12) \quad (J_{\eta\eta} \otimes Z'Z)\text{vec}(\zeta) = (J_{\eta\eta} \otimes Z')\text{vec}(D).$$

اگر بخواهیم در برآورد ضرایب رگرسیون از روش‌های هموارسازی استفاده نماییم، باید دو تاوان همواری را به کار ببریم که یکی تغییرات β نسبت به s و دیگری تغییرات β نسبت به t را نشان می‌دهد. در این جا $\text{PEN}_s(\beta)$ تغییرات β نسبت به s و $\text{PEN}_t(\beta)$ تغییرات β نسبت به t را نشان می‌دهد. از این رو، ضرایب رگرسیون $\alpha(t)$ و $\beta(s, t)$ باید طوری برآورد شوند که عبارت زیر مینیمم شود. بنا بر این

$$(13) \quad \begin{aligned} \text{PENSS}_\lambda(\alpha, \beta) &= \int \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\}' \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\} dt \\ &+ \lambda_s \text{PEN}_s(\beta) + \lambda_t \text{PEN}_t(\beta), \end{aligned}$$

که در آن

$$(14) \quad \begin{aligned} \text{PEN}_s(\beta) &= \int \int (L_s \beta(s, t))' ds dt \\ &= \int \int \{L_s \theta'(s) B \eta(t)\} \{L_s \theta'(s) B \eta(t)\}' ds dt \\ &= \int \int \{L_s \theta'(s) B \eta(t) \eta'(t) B' L_s \theta(s)\} ds dt \\ &= \int \int \text{tr}\{B \eta(t) \eta'(t) B' (L_s \theta(s)) (L_s \theta'(s))\} ds dt \\ &= \int \text{tr}\{B \eta(t) \eta'(t) B' R_\circ\} dt \\ &= \text{tr}(B J_{\eta\eta} B' R_\circ) \\ &= \text{tr}(B' R_\circ B J_{\eta\eta}), \end{aligned}$$

و به طور مشابه

$$\begin{aligned}
 \text{PEN}_t(\beta) &= \int \int \{L_t \beta(s, t)\}' ds dt \\
 &= \text{tr}(\mathbf{B} \mathbf{J}_{\theta\theta} \mathbf{B}' \mathbf{S}_\circ) \\
 (15) \quad &= \text{tr}(\mathbf{B}' \mathbf{S}_\circ \mathbf{B} \mathbf{J}_{\theta\theta}),
 \end{aligned}$$

که در عبارات بالا $\mathbf{R}_\circ = \int (L_s \boldsymbol{\theta}(s))(L_s \boldsymbol{\theta}(s))' ds$ ماتریس متقارن با اندازه‌ی $k_2 \times k_2$ ، $\mathbf{S}_\circ = \int (L_t \boldsymbol{\eta}(t))(L_t \boldsymbol{\eta}(t))' dt$ ماتریس متقارن با اندازه‌ی $k_1 \times k_1$ ، L_t و L_s عملگر دیفرانسیل خطی می‌باشند. با جایگذاری روابط به دست آمده در رابطه‌ی (۱۳) داریم

$$\begin{aligned}
 \text{PENSS}_\lambda(\alpha, \beta) &= \int \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\}' \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\} dt \\
 &\quad + \lambda_s \text{tr}(\mathbf{B}' \mathbf{R}_\circ \mathbf{B} \mathbf{J}_{\eta\eta}) + \lambda_t \text{tr}(\mathbf{B}' \mathbf{S}_\circ \mathbf{B} \mathbf{J}_{\theta\theta}) \\
 &= \int \{\mathbf{y}(t) - \mathbf{Z} \boldsymbol{\zeta} \boldsymbol{\eta}(t)\}' \{\mathbf{y}(t) - \mathbf{Z} \boldsymbol{\zeta} \boldsymbol{\eta}(t)\} dt \\
 (16) \quad &\quad + \lambda_s \text{tr}(\boldsymbol{\zeta}' \mathbf{R} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta}) + \lambda_t \text{tr}(\boldsymbol{\zeta}' \mathbf{S} \boldsymbol{\zeta} \mathbf{J}_{\theta\theta}),
 \end{aligned}$$

که در آن \mathbf{R} و \mathbf{S} به صورت زیر تعریف می‌شوند:

$$\mathbf{R}_{(k_2+1) \times (k_2+1)} = \begin{pmatrix} \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \mathbf{R}_\circ \end{pmatrix}, \quad \mathbf{S}_{(k_1+1) \times (k_1+1)} = \begin{pmatrix} \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \mathbf{S}_\circ \end{pmatrix}.$$

با مشتق‌گیری از عبارت (۱۶) نسبت به $\boldsymbol{\zeta}$ و مساوی صفر قرار دادن حاصل، داریم

$$\mathbf{Z}' \mathbf{Z} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta} + \lambda_s \mathbf{R} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta} + \lambda_t \mathbf{S} \boldsymbol{\zeta} \mathbf{J}_{\theta\theta} = \mathbf{Z}' \int \mathbf{y}(t) \boldsymbol{\eta}'(t) dt = \mathbf{Z}' \mathbf{D} \mathbf{J}_{\eta\eta}.$$

مشابه با آنچه در رابطه‌ی (۱۲) انجام شد، فرم بسته‌ی $\boldsymbol{\zeta}$ را می‌توان به صورت زیر نمایش داد:

$$(17) \quad \text{vec}(\boldsymbol{\zeta}) = (\mathbf{J}_{\eta\eta} \otimes \mathbf{Z}' \mathbf{Z} + \lambda_s \mathbf{J}_{\eta\eta} \otimes \mathbf{R} + \lambda_t \mathbf{J}_{\theta\theta} \otimes \mathbf{S})^{-1} (\mathbf{J}_{\eta\eta} \otimes \mathbf{Z}') \text{vec}(\mathbf{D}),$$

که در آن $\text{vec}(\boldsymbol{\zeta})$ به طول $1 \times (k_1 + 1) k_2$ است.

۴ برآورد پارامترها در مدل رگرسیون خطی تابعی با دو متغیر مستقل

در این بخش با توسعه‌ی مدل‌های (۴) و (۹) هنگامی که دو متغیر مستقل تابعی در مدل وجود دارد، مشابه با آنچه در بخش قبل بررسی شد، می‌توان بر حسب این که متغیر پاسخ اسکالر یا تابعی است، دو حالت زیر را در نظر گرفت و ضرایب رگرسیونی مربوط به آن‌ها را برآورد کرد.

۴/۱ حالت اول: متغیر پاسخ اسکالر

مدل رگرسیون خطی تابعی با دو متغیر مستقل $X(t)$ و $Z(t)$ به صورت

$$(۱۸) \quad y_i = \alpha + \int_0^T X_i(t)\beta(t)dt + \int_0^T Z_i(t)\gamma(t)dt + \epsilon_i \quad 1 \leq i \leq N,$$

است که در آن α ، y_i و ϵ_i اسکالر هستند. α ، $\beta(\cdot)$ و $\gamma(\cdot)$ غیر تصادفی و زوج‌های $(X_1(\cdot), \epsilon_1(\cdot))$ ، $(X_N(\cdot), \epsilon_N(\cdot))$ ، $(Z_1(\cdot), \epsilon_1(\cdot))$ ، $(Z_N(\cdot), \epsilon_N(\cdot))$ ، ... مستقل و هم‌توزیع‌اند. علاوه بر آن، توابع تصادفی $X_i(\cdot)$ و $Z_i(\cdot)$ مستقل از خطاهای ϵ_i هستند.

برآورد تابعی $\beta(\cdot)$ و $\gamma(\cdot)$ باید با استفاده از روش‌های هموارسازی یا نظم صورت گیرد. همانند قبل این روش‌ها ما را قادر می‌سازد تا بُعد را کاهش دهیم. برای این منظور توابع $\beta(\cdot)$ ، $\gamma(\cdot)$ و متغیرهای مستقل $X(\cdot)$ و $Z(\cdot)$ را به ترتیب بر حسب مجموعه‌ی توابع پایه‌ی $\{\theta_k(t)\}$ ، $\{\psi_k(t)\}$ و $\{\varphi_k\}$ با طول‌های k_β ، k_γ ، k_x و k_z با استفاده از روابط (۱) و (۲) بسط می‌دهیم. اگر فرض کنیم α برداری به طول N با عناصر تکراری α باشد، آنگاه با جایگذاری روابط بالا در رابطه‌ی (۱۸) داریم

$$\begin{aligned} \hat{y} &= \alpha + \int_0^T X(t)\beta(t)dt + \int_0^T Z(t)\gamma(t)dt \\ &= \alpha + \int_0^T C\psi(t)\theta' b dt + \int_0^T E\varphi(t)\vartheta' d dt \\ &= \alpha + C J_{\psi\theta} b + E J_{\varphi\vartheta} d, \end{aligned}$$

که در آن $b = (b_1, \dots, b_{k_\beta})'$ و $d = (d_1, \dots, d_{k_\gamma})'$ ، به علاوه C ماتریس $N \times k_x$ از ضرایب c_{ik} و E ماتریس $N \times k_z$ از ضرایب e_{ik} می‌باشند. ماتریس‌های $J_{\psi\theta}$ و $J_{\varphi\vartheta}$ به ترتیب با اندازه‌های

می‌باشند که از روابط زیر به دست می‌آیند:

$$J_{\psi\theta} = \int \psi(t)\theta(t)dt, \quad J_{\varphi\vartheta} = \int \varphi(t)\theta(t)dt.$$

با فرض این که بردار ضرایب رگرسیونی تابعی برابر $\zeta = (\alpha, b_1, \dots, b_{k_\beta}, d_1, \dots, d_{k_\gamma})'$ و ماتریس ضرایب $Z = [1, C J_{\psi\theta}, E J_{\varphi\vartheta}]$ با ابعاد $N \times (k_\beta + k_\gamma + 1)$ باشد، اگر بخواهیم در برآورد ضرایب رگرسیون از روش‌های هموارسازی استفاده نماییم در این حالت باید عبارت

$$\begin{aligned} \text{PENSS}_\lambda(\alpha, \beta, \gamma) &= \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda_1 \int [L\beta(t)]^2 dt + \lambda_2 \int [L\gamma(t)]^2 dt \\ &= (y - \hat{y})'(y - \hat{y}) + \lambda_1 b' R_1 b + \lambda_2 d' R_2 d, \\ &= (y - \hat{y})'(y - \hat{y}) + \zeta' R \zeta, \end{aligned}$$

را نسبت به α ، $\beta(\cdot)$ و $\gamma(\cdot)$ مینیمم کنیم که در آن L یک عملگر دیفرانسیل خطی است. همچنین λ_1 و λ_2 را می‌توان به طور اختیاری یا از طریق اعتبارسنجی متقابل انتخاب کرد. به‌علاوه $R_1 = \int (L\theta(t))(L\theta(t))' dt$ و $R_2 = \int (L\vartheta(t))(L\vartheta(t))' dt$ می‌باشد که R به‌صورت زیر تعریف می‌شود:

$$R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda_1 R_1 & 0 \\ 0 & 0 & \lambda_2 R_2 \end{pmatrix}.$$

با مشتق‌گیری عبارت قبل نسبت به ζ و سپس ساده کردن عبارت حاصل، داریم

$$(Z'Z + R)\hat{\zeta} = Z'y. \quad (19)$$

۴/۲ حالت دوم: متغیر پاسخ تابعی

مدل رگرسیون خطی تابعی با دو متغیر مستقل $X(t)$ و $Z(t)$ وقتی که متغیر پاسخ $y(t)$ تابعی باشد به صورت زیر است:

$$(۲۰) \quad y_i(t) = \alpha(t) + \int_0^T X_i(s)\beta(s,t)ds + \int_0^T Z_i(s)\gamma(s,t)ds + \epsilon_i(t) \quad 1 \leq i \leq N,$$

که در آن $\alpha(t)$ ، $y_i(t)$ و $\epsilon_i(t)$ تابعی هستند، توابع $\alpha(t)$ ، $\beta(s,t)$ و $\gamma(s,t)$ غیر تصادفی و زوج های $(X_1(\cdot), \epsilon_1(\cdot))$ ، \dots ، $(X_N(\cdot), \epsilon_N(\cdot))$ و $(Z_1(\cdot), \epsilon_1(\cdot))$ ، \dots ، $(Z_N(\cdot), \epsilon_N(\cdot))$ مستقل و هم توزیع اند. علاوه بر آن، توابع تصادفی $X_i(\cdot)$ و $Z_i(\cdot)$ مستقل از خطاهای $\epsilon_i(\cdot)$ هستند.

برای برآورد توابعی ضرایب $\alpha(\cdot)$ ، $\beta(\cdot, \cdot)$ و $\gamma(\cdot, \cdot)$ و نیز به منظور کاهش بُعد از روش های هموارسازی استفاده می کنیم. با استفاده از رابطه ی (۳)، ضرایب رگرسیون تابعی $\beta(\cdot, \cdot)$ و $\gamma(\cdot, \cdot)$ را بر حسب مجموعه ای از توابع پایه ی متعامد یکه ی $\{\theta_l(s)\}$ و $\{\eta_k(t)\}$ به طول های k_1 و k_2 بسط دهیم. به علاوه توابع $\alpha(t)$ ، $X(s)$ و $Z(s)$ را به ترتیب بر حسب مجموعه ای از توابع پایه ی متعامد یکه ی $\{\eta_k\}$ ، $\{\theta_k\}$ و $\{\theta_k\}$ به طول های k_1 ، k_2 و k_2 با استفاده از روابط (۱) و (۲) می نویسیم. حال مدل به صورت زیر بیان می شود:

$$\begin{aligned} \hat{y}(t) &= \alpha(t) + \int_0^T X(s)\beta(s,t)ds + \int_0^T Z(s)\gamma(s,t)ds \\ &= a' \eta(t) + \int_0^T E \theta(s) \theta'(s) B \eta(t) ds + \int_0^T F \theta(s) \theta'(s) C \eta(t) ds \\ &= a' \eta(t) + E J_{\theta\theta} B \eta(t) + F J_{\theta\theta} C \eta(t), \end{aligned}$$

که در آن یک ماتریس از ضرایب b_{kl} ، یک ماتریس از ضرایب c_{kl} ، یک ماتریس از ضرایب $C_{k_2 \times k_1}$ ، یک ماتریس E ، یک ماتریس F از ضرایب e_{ik} ، ماتریسی $N \times k_2$ از ضرایب f_{ik} و $a = (a_1, \dots, a_{k_1})'$ است. به علاوه ماتریس $J_{\theta\theta} = \int \theta(s) \theta'(s) ds$ در رابطه ی قبل با اندازه ی $k_2 \times k_2$ می باشد.

اگر فرض کنیم بردار ضرایب رگرسیون تابعی به صورت $\zeta = (a', B', C')'$ و اندازه ی ماتریس ضرایب $Z = [1, E J_{\theta\theta}, F J_{\theta\theta}]$ ، $N \times (2k_2 + 1)$ باشد و بخواهیم در برآورد ضرایب رگرسیون از روش های هموارسازی استفاده نماییم، آن گاه باید چهار تاوان همواری را به کار ببریم که دو تا از آن ها تغییرات $\beta(s, t)$ نسبت به s و t و دوتای دیگر تغییرات $\gamma(s, t)$ نسبت به s و t را نشان می دهد. $PEN_s(\beta)$

تغییرات $\beta(s, t)$ نسبت به s ، $\text{PEN}_t(\beta)$ تغییرات $\beta(s, t)$ نسبت به t ، $\text{PEN}_s(\gamma)$ تغییرات $\gamma(s, t)$ نسبت به s و $\text{PEN}_s(\gamma)$ تغییرات $\gamma(s, t)$ نسبت به t را نشان می‌دهد و به روش مشابه از مدل‌های (۱۴) و (۱۵) به دست می‌آیند. در این حالت باید ضرایب رگرسیون $\alpha(t)$ ، $\beta(s, t)$ و $\gamma(s, t)$ طوری برآورد شوند که عبارت زیر مینیمم شود.

$$\begin{aligned} \text{PENSS}_\lambda(\alpha, \beta, \gamma) &= \int \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\}' \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\} dt + \lambda_{s_1} \text{PEN}_s(\beta) \\ &\quad + \lambda_{t_1} \text{PEN}_t(\beta) + \lambda_{s_2} \text{PEN}_s(\gamma) + \lambda_{t_2} \text{PEN}_t(\gamma) \\ &= \int \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\}' \{\mathbf{y}(t) - \hat{\mathbf{y}}(t)\} dt + \lambda_{s_1} \text{tr}(\mathbf{B}' \mathbf{R}_0 \mathbf{B} \mathbf{J}_{\eta\eta}) \\ &\quad + \lambda_{t_1} \text{tr}(\mathbf{B}' \mathbf{S}_0 \mathbf{B} \mathbf{J}_{\theta\theta}) + \lambda_{s_2} \text{tr}(\mathbf{C}' \mathbf{R}_0 \mathbf{C} \mathbf{J}_{\eta\eta}) + \lambda_{t_2} \text{tr}(\mathbf{C}' \mathbf{S}_0 \mathbf{C} \mathbf{J}_{\theta\theta}) \\ &= \int \{\mathbf{y}(t) - \mathbf{Z} \boldsymbol{\zeta} \eta(t)\}' \{\mathbf{y}(t) - \mathbf{Z} \boldsymbol{\zeta} \eta(t)\} dt \\ &\quad + \text{tr}(\boldsymbol{\zeta}' \mathbf{R} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta}) + \text{tr}(\boldsymbol{\zeta}' \mathbf{S} \boldsymbol{\zeta} \mathbf{J}_{\theta\theta}), \end{aligned} \quad (21)$$

که در آن \mathbf{R} و \mathbf{S} به صورت زیر است:

$$\mathbf{R} = \begin{pmatrix} \mathbf{O}' & \mathbf{O}' & \mathbf{O}' \\ \mathbf{O} & \lambda_{s_1} \mathbf{R}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \lambda_{s_2} \mathbf{R}_0 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{O}' & \mathbf{O}' & \mathbf{O}' \\ \mathbf{O} & \lambda_{t_1} \mathbf{S}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \lambda_{t_2} \mathbf{S}_0 \end{pmatrix}.$$

با مشتق‌گیری از عبارت (۲۱) نسبت به $\boldsymbol{\zeta}$ و مساوی صفر قرار دادن حاصل، داریم

$$\mathbf{Z}' \mathbf{Z} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta} + \mathbf{R} \boldsymbol{\zeta} \mathbf{J}_{\eta\eta} + \mathbf{S} \boldsymbol{\zeta} \mathbf{J}_{\theta\theta} = \mathbf{Z} \int \mathbf{y}(t) \eta'(t) dt = \mathbf{Z}' \mathbf{D} \mathbf{J}_{\eta\eta}.$$

برای به دست آوردن فرم بسته‌ای برای $\boldsymbol{\zeta}$ از حاصل ضرب کرونکر استفاده می‌کنیم،

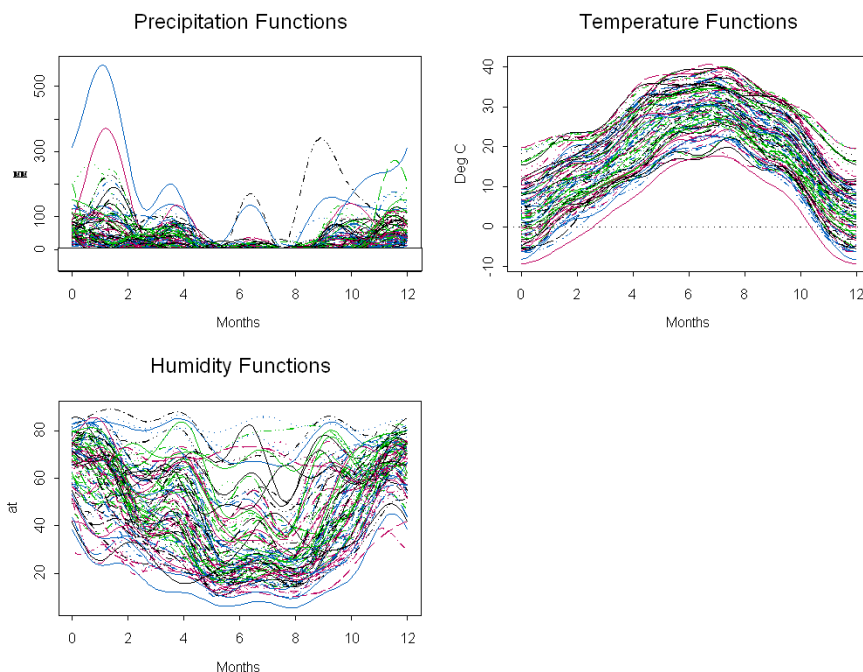
$$(22) \quad \text{vec}(\boldsymbol{\zeta}) = (\mathbf{J}_{\eta\eta} \otimes \mathbf{Z}' \mathbf{Z} + \mathbf{J}_{\eta\eta} \otimes \mathbf{R} + \mathbf{J}_{\theta\theta} \otimes \mathbf{S})^{-1} = (\mathbf{J}_{\eta\eta} \otimes \mathbf{Z}') \text{vec}(\mathbf{D}),$$

که در آن $\text{vec}(\boldsymbol{\zeta})$ برداری به طول $k_1(k_2 + 1)$ است.

۵ تحلیل داده‌های واقعی

داده‌هایی که در این مقاله مورد تحلیل قرار گرفته‌اند، مربوط به ۱۰۲ ایستگاه هواشناسی ایران در سال ۲۰۰۶ می‌باشد و داده‌های بارندگی، دما و رطوبت مربوط به آن‌ها مورد تحلیل قرار گرفته است. در این مجموعه داده، برای هر ایستگاه ۱۲ عدد گزارش شده است که همان میانگین ماهانه دما، رطوبت یا مقدار بارندگی در آن منطقه می‌باشند.

در ابتدا باید منحنی‌های مناسب طوری به مجموعه داده‌ها برازش شوند تا خطای اندازه‌گیری تا حد ممکن پالایه شود. به دلیل ماهیت دوره‌ای (تناوبی) پدیده‌های بارندگی، دما و رطوبت، برای برازش منحنی به مجموعه داده‌ها از بسط توابع قبل در قالب سری‌های فوریه استفاده نموده‌ایم. در نهایت استفاده از یک سری فوریه با ۱۳ جمله برازش مناسبی به داده‌ها ایجاد می‌نماید. نمودار منحنی‌های برازش داده شده به داده‌های اولیه مورد بررسی در شکل ۱ آمده است (برای جزئیات بیشتر در مورد این داده‌ها به تازیکه میاندره و حسینی نسب (۱۳۸۷) مراجعه کنید). در انتها متذکر می‌شویم که در تمامی مدل‌های رگرسیونی که

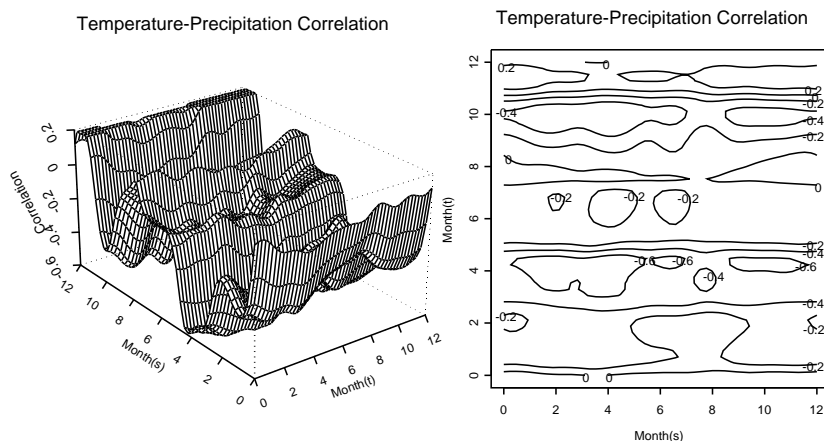


شکل ۱. برازش منحنی به داده‌های دما، میزان رطوبت و میزان بارندگی مربوط به ۱۰۲ ایستگاه هواشناسی در ایران با استفاده از سری‌های فوریه

در ادامه محاسبه می‌شود، به دلیل آن که مقدار کل بارندگی مربوط به مناطق اقلیمی متفاوتی از سراسر ایران می‌باشد مقدار آن در میان ایستگاه‌های هواشناسی با تغییرات زیادی همراه است. بنا بر این، از لگاریتم آن به عنوان متغیر وابسته استفاده می‌کنیم.

۵/۱ مدل‌بندی بارندگی بر اساس دما

شکل ۲ همبستگی بین دما و میزان بارندگی را نشان می‌دهد. در نمودار سمت راست، منحنی تراز مربوط به همبستگی بین دما و بارندگی که بر حسب زمان (ماه) بیان شده است نشان داده می‌شود. بر اساس این نمودار، همبستگی دما و میزان بارندگی در اکثر ماه‌ها منفی می‌باشد. بیش‌ترین همبستگی بین دما و بارندگی در ماه آوریل (اواسط فروردین تا اواسط اردیبهشت) و تقریباً برابر -0.6 می‌باشد. به علاوه همبستگی بین دما و بارندگی در ماه آگوست (اواسط مرداد تا اواسط شهریور) برابر صفر است. در دو قسمت زیر مقدار لگاریتم بارندگی را برای حالت‌هایی که اسکالر یا تابعی باشد بررسی و بر حسب دما مدل‌بندی می‌کنیم. برای این منظور مقدار متغیر مستقل دما در زمان t را با $\text{Temp}(t)$ نمایش می‌دهیم. به علاوه متغیر وابسته بارندگی را در دو حالت اسکالر و تابعی به ترتیب با y و $y(t)$ نشان می‌دهیم.



شکل ۲. تابع همبستگی مقاطع میزان بارندگی-دما و نمودار خطوط تراز مربوط به آن

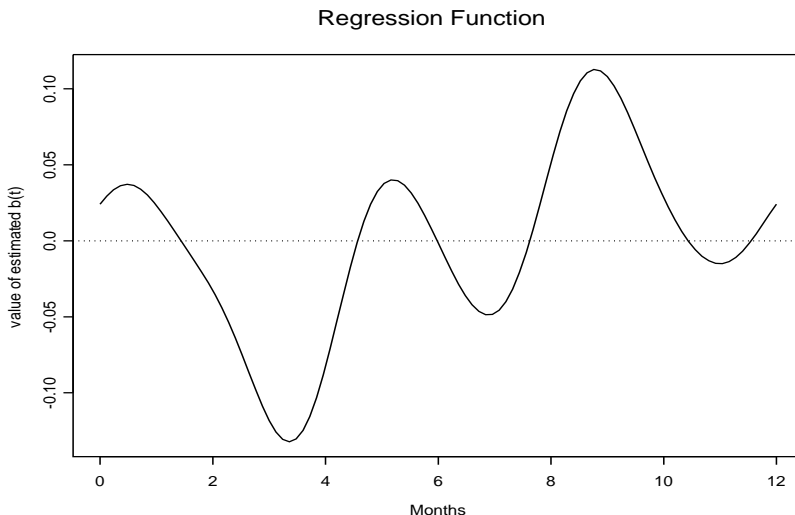
۵/۱/۱ متغیر وابسته y ، اسکالر (مقدار لگاریتم کل بارندگی سالیانه) باشد.

با توجه به این که متغیر مستقل و وابسته به ترتیب تابعی و اسکالر هستند، در نتیجه مدل (۴) را بررسی می‌کنیم و بنا بر این ضرایب رگرسیونی $\hat{\gamma}$ را می‌توان از فرمول (۸) به دست آورد.

شکل ۳ شیب برآورد شده‌ی مدل را نشان می‌دهد. این منحنی وزن مثبت بزرگ‌تر را به اواخر جولای تا اواسط اکتبر (اوایل مرداد تا اواخر مهر) و اواسط نوامبر تا اواسط ژانویه (اواخر آبان تا اواخر دی) اختصاص می‌دهد. در مقابل این منحنی وزن‌های منفی کوچک را به اواسط ژانویه تا اواسط آوریل (اواخر دی تا اواخر فروردین) و اوایل ژوئن تا اواخر جولای (اواسط خرداد تا اوایل مرداد) اختصاص می‌دهد. همچنین بر اساس برآوردهای $\hat{\alpha}$ و $\hat{\beta}$ ، می‌توان میزان بارندگی را بر اساس مدل برآورد شده‌ی

$$(23) \quad \widehat{\log \text{Prec}_i} = \hat{\alpha} + \int_0^T \text{Temp}_i(t) \hat{\beta}(t) dt \quad 1 \leq i \leq N,$$

پیش‌بینی کرد. در شکل ۴ لگاریتم مقادیر مشاهده‌شده در برابر مقادیر پیش‌بینی‌شده‌ی لگاریتم مقدار بارندگی ترسیم شده است. این نمودار گویای این مطلب است که مقادیر پیش‌بینی‌شده بر اساس مدل (۲۳) نسبتاً قابل قبول می‌باشند. لازم به ذکر است که مقدار ضریب تعیین برای مدل (۲۳) برابر $R^2 = 0.757$ می‌باشد.



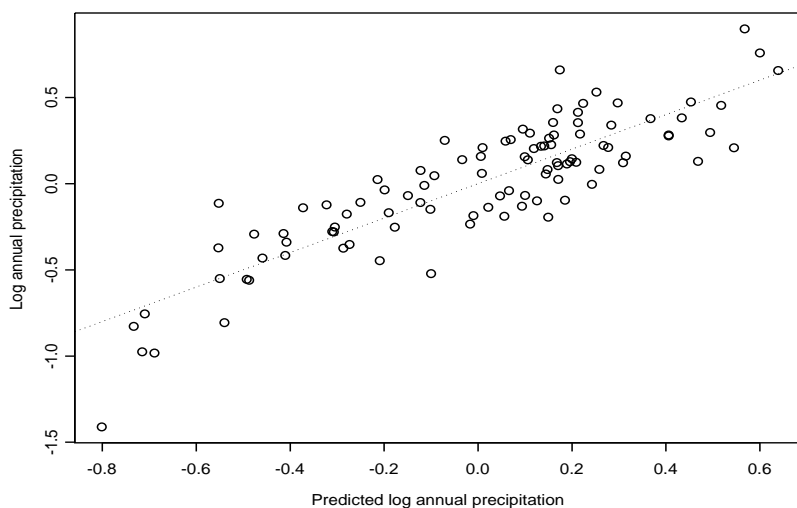
شکل ۳. شیب برآورد شده‌ی رگرسیون وقتی که متغیر وابسته، لگاریتم کل بارندگی سالانه و متغیر مستقل، دما است. نقطه‌ی ثابت $\hat{\alpha} = -0.1312152$ برآورد شده است.

۵/۱/۲ متغیر وابسته y ، تابعی (مقدار لگاریتم بارندگی در طول زمان) باشد.

مدل مورد بررسی در این بخش، مدل (۹) می‌باشد. بنابراین برآورد ضرایب رگرسیونی $\hat{\gamma}(0)$ را از فرمول (۱۷) می‌توان به دست آورد.

نمودارهای (a) و (b) در شکل ۵ به ترتیب نمایش سه‌بعدی و منحنی‌های خطوط تراز شیب برآورد شده‌ی رگرسیون خطی تابعی را نشان می‌دهند. بر اساس این نمودارها، در ماه سپتامبر (اواسط شهریور تا اواسط مهر) بیش‌ترین تأثیر دما بر میزان بارندگی در اواخر ماه سپتامبر (اوایل مهر) اتفاق افتاده است. همان‌طور که مشاهده می‌شود تأثیر دما بر میزان بارندگی در اکثر ماه‌ها معکوس (دارای وزن منفی) می‌باشد. روش‌های گوناگونی برای ارزیابی مدل‌های خطی تابعی وجود دارد، که از آن جمله می‌توان به معیار توان دوم همبستگی اشاره نمود. این معیار به‌صورت زیر تعریف می‌شود:

$$R^2(t) = \frac{1 - \sum_i \{\hat{y}_i(t) - y_i(t)\}^2}{\sum_i \{y_i(t) - \bar{y}(t)\}^2}.$$

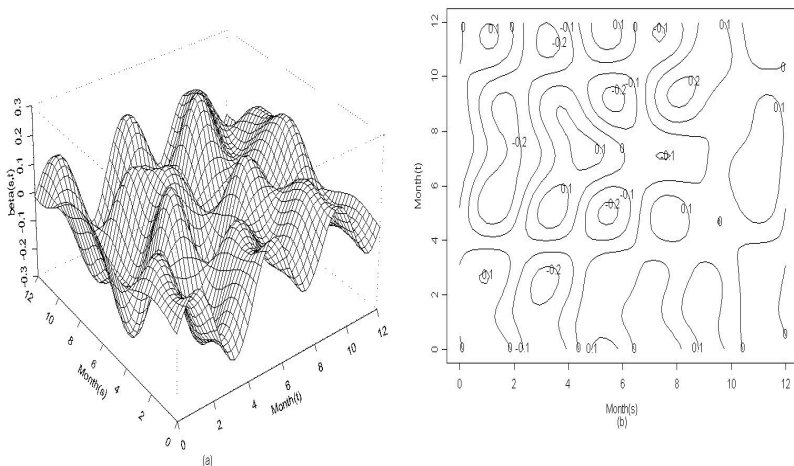


شکل ۴. نمودار مشاهده‌شده (لگاریتم مقدار بارندگی ایران در سال ۲۰۰۶) در برابر مقادیر پیش‌بینی‌شده وقتی متغیر مستقل دما باشد. خط راست روی نمودار، نشان‌دهنده‌ی مقدار مشاهده‌شده در مقابل مقدار پیش‌بینی‌شده می‌باشد.

اگر بخواهیم از معیار بالا فقط یک عدد به دست آوریم، می‌توان از متوسط تابع $R^2(t)$ روی زمان استفاده نمود. اما استفاده از تابع $R^2(t)$ جزئیات بیش‌تری را در مورد برازش بیان می‌کند. این تابع در نمودار (a) در شکل ۶ ترسیم شده است. مقدار این ضریب در ماه‌های چهارم تا ششم (اواسط فروردین تا اواسط مرداد) و ماه‌های نهم تا یازدهم (اواسط مهر تا اواسط آذر) افزایش می‌یابد که این گویای این مطلب است که تغییرات بارندگی در این دوره‌ی زمانی به خوبی به وسیله‌ی دما بیان می‌شود. روش دیگر برای ارزیابی مدل، استفاده از معیار زیر است:

$$F - \text{ratio}(t) = \frac{\sum_i \{\hat{y}_i(t) - \bar{y}(t)\}^2 / (k_e - 1)}{\sum_i \{y_i(t) - \hat{y}_i(t)\}^2 / (n - k_e)}.$$

در کسر بالا، $k_e - 1$ درجه‌ی آزادی مربوط به مجموع توان‌های دوم رگرسیون یعنی $\sum_i \{\hat{y}_i(t) - \bar{y}(t)\}^2$ و $n - k_e$ درجه‌ی آزادی مربوط به توان‌های دوم خطا یعنی $\sum_i (y_i(t) - \hat{y}_i(t))^2$ است. لازم به ذکر است که در این مقاله $k_e = 7$ و $n = 102$. این تابع در نمودار (b) در شکل ۶ برای ارزیابی برازش داده‌های لگاریتم مقدار بارندگی ترسیم شده است. در این نمودار، مقدار $F - \text{ratio}(t)$ برای هر نقطه‌ی زمانی t بالای خط $F(6, 95, 0.05)$ قرار دارد و این نشان‌دهنده‌ی این مطلب است که دما در پیش‌بینی مقدار لگاریتم بارندگی در سطح معنی‌داری ۹۵ درصد تأثیرگذار است. توجه شود که بیش‌ترین مقدار این

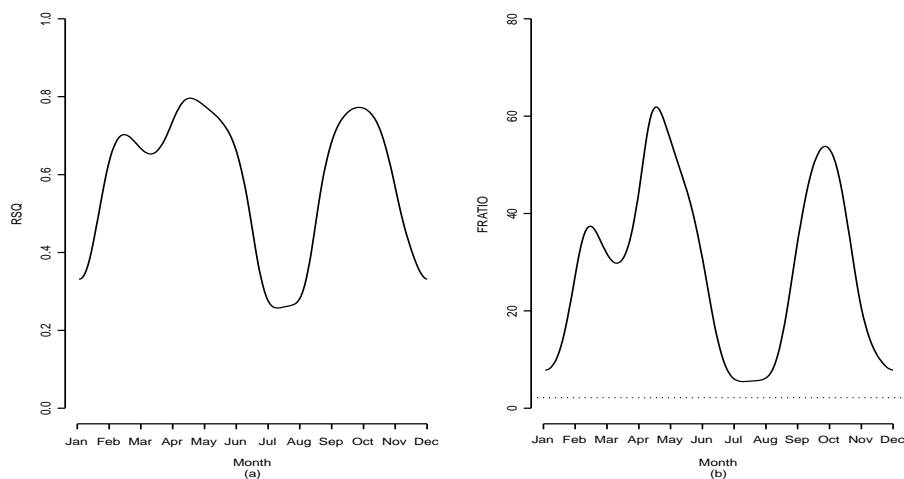


آماره در ماه‌های آوریل تا ژوئن (اواسط فروردین تا اواسط مرداد) و ماه‌های سپتامبر تا نوامبر (اواسط شهریور تا اواسط آذر) می‌باشد و کم‌ترین مقدار آن در ماه جولای (اواسط مرداد تا اواسط شهریور) مشاهده می‌شود.

۵/۲ مدل‌بندی بارندگی بر اساس رطوبت نسبی

از پدیده‌های گوناگون دیگری به جز دما نیز می‌توان برای پیش‌بینی میزان بارندگی استفاده نمود که در این بخش رطوبت نسبی را مورد توجه قرار می‌دهیم. برای این منظور از متغیر مستقل تابعی رطوبت نسبی یعنی $Humid(t)$ برای پیش‌بینی متغیر وابسته‌ی لگاریتم بارندگی در دو حالت اسکالر و تابعی استفاده می‌شود.

شکل ۷ همبستگی بین رطوبت و میزان بارندگی را آشکار می‌سازد. در این شکل نمودار سه‌بعدی و خطوط تراز برای نمایش این همبستگی ترسیم شده است و از آن‌ها نتایج کلی زیر قابل بیان است: میزان همبستگی بین میزان رطوبت و مقدار بارندگی در اکثر ماه‌ها مثبت می‌باشد. رطوبت نسبی همبستگی زیادی (تقریباً برابر ۰/۶) با میزان بارندگی در ماه‌های آوریل (اواسط فروردین تا اواسط اردیبهشت) و سپتامبر (اواسط شهریور تا اواسط مهر) دارد. همچنین رطوبت با میزان بارندگی در ماه آگوست (اواسط شهریور تا اواسط مهر) همبستگی ندارد.



شکل ۶. نمودار (a) نمایش سه‌بعدی و نمودار (b) منحنی‌های خطوط تراز برآورد تابعی $\beta(s, t)$ با استفاده از ۷ پایه‌ی تابعی برای هر شناسه‌ی وقتی متغیر مستقل، دما باشد.

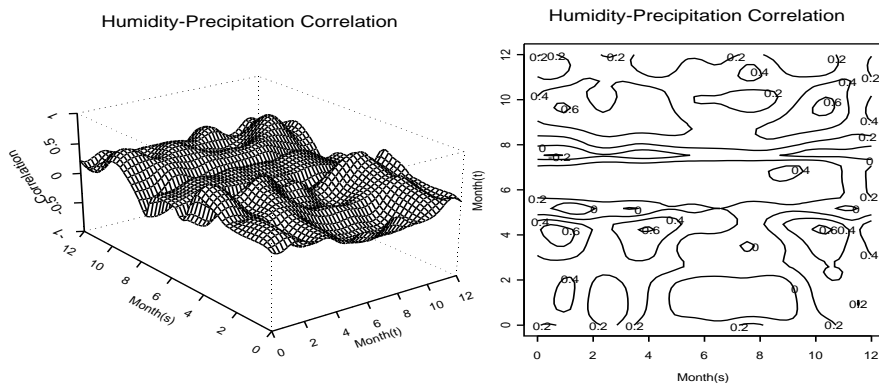
۵/۲/۱ متغیر وابسته y ، اسکالر و مقدار لگاریتم کل بارندگی سالیانه باشد.

با توجه به این که متغیرهای مستقل و وابسته به ترتیب تابعی و اسکالر هستند، از این رو مدل (۴) مدلی مناسب برای مدل بندی مورد نظر می باشد. در نتیجه ضرایب $\hat{\alpha}$ را می توان از فرمول (۸) به دست آورد.

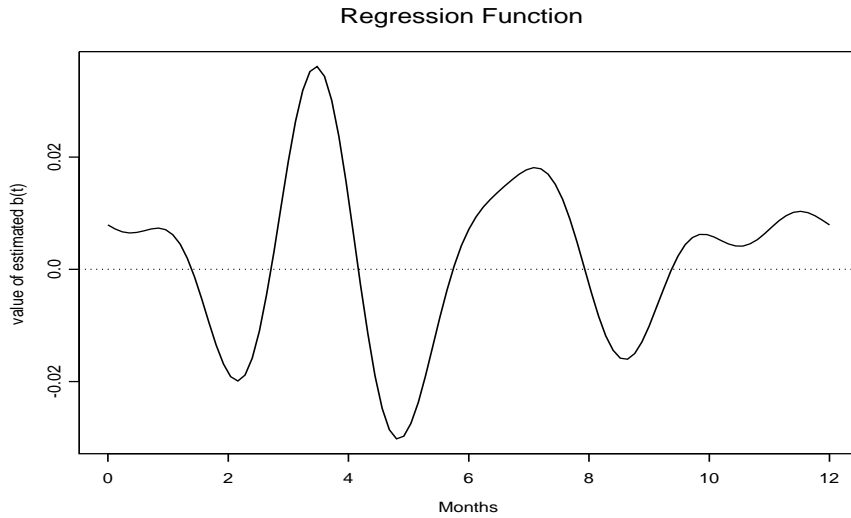
همچنان که شکل ۸ نشان می دهد شیب برآورد شده ی رگرسیون خطی تابعی، وزن های بیشتر را به اواخر فوریه تا اوایل آوریل (اوایل اسفند تا اواسط فروردین) و اواخر می تا آگوست (اواسط خرداد تا اواسط مرداد) اختصاص می دهد. علاوه بر آن وزن های کمتر متناظر با اوایل آوریل تا اواخر می (اواسط فروردین تا اواسط خرداد) و اواسط ژانویه تا اواخر فوریه (اواخر دی تا اواسط اسفند) می باشند. همچنین بر اساس برآوردهای $\hat{\alpha}$ و $\hat{\beta}(\cdot)$ ، می توان میزان بارندگی را بر اساس مدل برآورد شده ی

$$(24) \quad \widehat{\log \text{Prec}_i} = \hat{\alpha} + \int_0^T \text{Humid}_i(t) \hat{\beta}(t) dt \quad 1 \leq i \leq N,$$

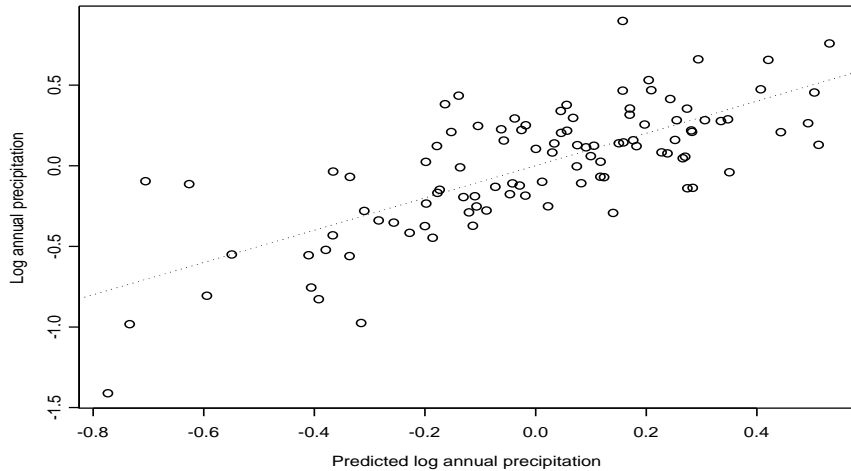
پیش بینی کرد. شکل ۹ لگاریتم مقادیر مشاهده شده در برابر مقادیر پیش بینی شده ی لگاریتم مقدار بارندگی را نشان می دهد. این نمودار بیان می کند مقادیر پیش بینی شده بر اساس مدل (۲۴) نسبتاً قابل قبول می باشند و ضریب تعیین برای آن $R^2 = 0.68$ است.



شکل ۷. تابع همبستگی متقاطع میزان بارندگی-رطوبت و نمودار خطوط تراز مربوط به آن



شکل ۸. شیب برآورد شده‌ی رگرسیون وقتی که متغیر وابسته، لگاریتم کل بارندگی سالانه و متغیر مستقل میزان رطوبت است. نقطه‌ی ثابت $\hat{\alpha} = -۱/۶۳۵۸۳۸$ برآورد شده است.



شکل ۹. نمودار مشاهده‌شده (لگاریتم مقدار بارندگی ایران در سال ۲۰۰۶) در برابر مقادیر پیش‌بینی‌شده وقتی که متغیر مستقل، رطوبت باشد. خط راست روی نمودار، نشان‌دهنده‌ی مقدار پیش‌بینی‌شده می‌باشد.

۵/۲/۲ متغیر وابسته y ، تابعی (مقدار لگاریتم بارندگی در لحظه t) باشد

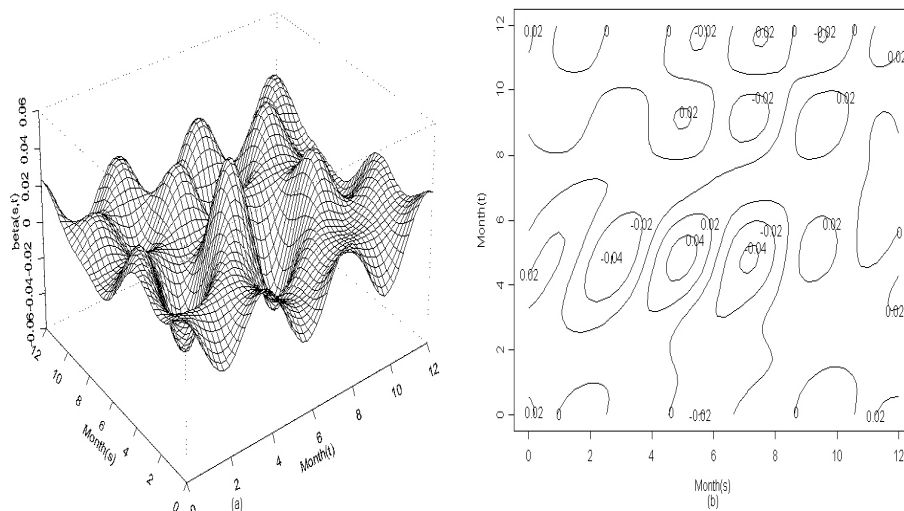
در عمل پیش‌بینی مقدار بارندگی در هر زمان از سال به جای محاسبه‌ی یک مقدار مطلوب‌تر می‌باشد. از این رو، در ادامه لگاریتم بارندگی را در هر نقطه از طول سال و به صورت یک منحنی، پیش‌بینی می‌کنیم. برای این منظور با تعریف متغیر مستقل تابعی رطوبت یعنی $Humid(t)$ و متغیر وابسته مقدار لگاریتم بارندگی یعنی $y(t)$ مدل (۹) را در نظر می‌گیریم. بر این اساس از فرمول (۱۷) می‌توان $\hat{y}(0)$ را برآورد کرد.

نمودارهای (a) و (b) در شکل ۱۰ به ترتیب نمایش سه بُعد نما و منحنی‌های خطوط تراز شیب برآورد شده‌ی رگرسیون خطی تابعی را نشان می‌دهند. بر اساس این نمودارها، بیش‌ترین تأثیر میزان رطوبت در ماه ژوئن (اواسط اردیبهشت تا اواسط خرداد) بر میزان بارندگی در ماه ژوئن (اواسط اردیبهشت تا اواسط خرداد) می‌باشد. بیش‌ترین ارتباط میزان رطوبت و بارندگی در ماه آوریل (اواسط فروردین تا اواسط اردیبهشت) است که معکوس می‌باشد. همان‌طور که مشاهده می‌شود تأثیر رطوبت روی بارندگی در ماه دسامبر (اواسط آذر تا اواسط دی) روند دوره‌ای دارد. تابع R^2 در نمودار (a) در شکل ۱۱ ترسیم شده است. در این نمودار بیش‌ترین مقادیر به ترتیب در ماه‌های چهارم تا ششم (اواسط فروردین تا اواسط مرداد)، ماه‌های نهم تا یازدهم (اواسط مهر تا اواسط آذر) و فوریه (اواسط بهمن تا اواسط اسفند) سال مشاهده می‌شود. به عبارت دیگر لگاریتم بارندگی در این دوره‌ی زمانی به خوبی به وسیله‌ی رطوبت بیان می‌شود.

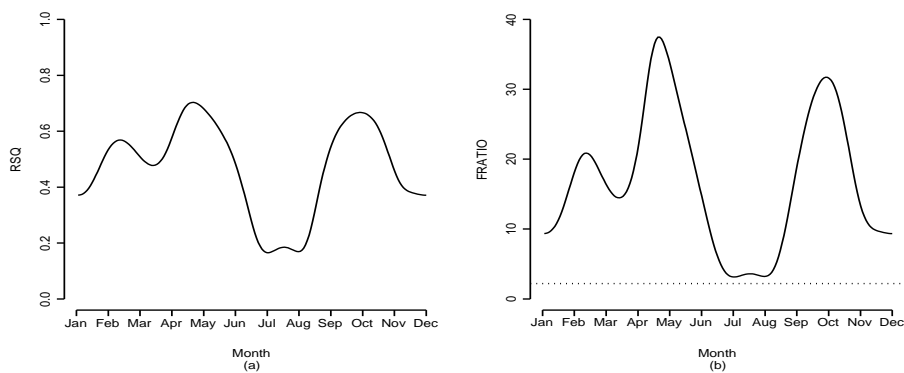
نمودار (b) در شکل ۱۱، $F - ratio$ برای برازش داده‌های لگاریتم مقدار بارندگی را نشان می‌دهد. در این نمودار مقدار $F - ratio(t)$ برای هر t بالای خط $F(6, 95, 0.05)$ قرار دارد و بنا بر این رطوبت در سطح معنی‌داری ۹۵ درصد از لحاظ آماری بر مقدار بارندگی اثر معنی‌داری دارد. به علاوه با بررسی این نتایج مشاهده می‌شود تأثیر رطوبت روی لگاریتم بارندگی در ماه‌های آوریل تا ژوئن (اواسط فروردین تا اواسط مرداد) و ماه‌های سپتامبر تا نوامبر (اواسط شهریور تا اواسط آذر) با اطمینان بیش‌تری پذیرفته می‌شود و کم‌ترین تأثیر رطوبت روی بارندگی در ماه جولای (اواسط مرداد تا اواسط شهریور) است.

۵/۳ مدل‌بندی بارندگی بر اساس دما و رطوبت

در بخش‌های قبلی با استفاده از رطوبت نسبی و دما، مدل‌بندی میزان بارندگی در کل سال یا تابع بارندگی بررسی شد. حال مدل‌بندی اثر توأم دو پدیده‌ی دما و رطوبت نسبی را مورد توجه قرار می‌دهیم و به تفسیر نتایج می‌پردازیم. در ادامه متغیرهای مستقل تابعی دما و میزان رطوبت در زمان t به ترتیب با $Temp(t)$ و $Humid(t)$ نشان می‌دهیم و در دو حالت اسکالر یعنی y و تابعی یعنی $y(t)$ پدیده‌ی بارندگی را مدل‌بندی می‌کنیم و نتایج حاصل را تفسیر می‌کنیم.



شکل ۱۰. نمودار (a) نمایش سه‌بعدی و نمودار (b) منحنی‌های تراز برآورد تابعی $\beta(s, t)$ با استفاده از پایه تابعی برای هر شناسه وقتی که متغیر مستقل رطوبت باشد.



شکل ۱۱. نمودار (a) تابع RSQ نسبت تغییرات بیان‌شده از لگاریتم مقدار بارندگی به‌وسیله‌ی مدل (۹) و نمودار (b) تابع F - ratio برای پیش‌بینی لگاریتم بارندگی بر اساس رطوبت. خط افقی $F(۶, ۹۵, ۰/۰۵)$ را نشان می‌دهد.

۵/۳/۱ وقتی که متغیر وابسته y ، اسکالر (مقدار لگاریتم کل بارندگی سالیانه) باشد.

با توجه به این که برآورد مقدار y در طول سال مورد توجه می باشد، مدل (۱۸) مدل مناسب برای تحلیل می باشد و برآورد ضرایب تابعی $\hat{\gamma}$ را می توان از فرمول (۱۹) به دست آورد.

شکل ۱۲ شیب برآورد شده ی رگرسیون وقتی که متغیر وابسته، لگاریتم کل بارندگی سالانه و متغیر مستقل، دما و میزان رطوبت است، را نشان می دهد. شیب برآورد شده ی رگرسیون خطی تابعی برای دما، وزن مثبت بزرگتری را به ترتیب به اواخر جولای تا اوایل اکتبر (اوایل مرداد تا اواسط مهر) و اوایل می تا اوایل ژوئن (اواسط اردیبهشت تا اواسط خرداد) اختصاص می دهد. علاوه بر آن وزن منفی کوچکتر برای اواسط فوریه تا اوایل می (اواخر بهمن تا اواسط اردیبهشت) و اوایل ژوئن تا اواسط جولای (اواسط خرداد تا اواخر تیر) در نظر گرفته می شود. همچنین شیب برآورد شده ی رگرسیون خطی تابعی برای رطوبت، وزن مثبت بیشتری را در اوایل ژوئن تا اوایل آگوست (اواسط خرداد تا اواسط مرداد) و اواسط نوامبر تا اوایل ژانویه (اواخر آبان تا اواسط دی) نشان می دهد. علاوه بر آن وزن منفی کوچکتر متناظر با اوایل آگوست تا اواسط سپتامبر (اواسط مرداد تا اواخر شهریور) و اواخر مارس تا اواسط ژوئن (اوایل فروردین تا اواخر خرداد) می باشند. همچنین بر اساس برآوردهای $\hat{\alpha}(\cdot)$ ، $\hat{\beta}(\cdot)$ و $\hat{\gamma}(\cdot)$ می توان میزان بارندگی را بر اساس مدل برآورد شده ی

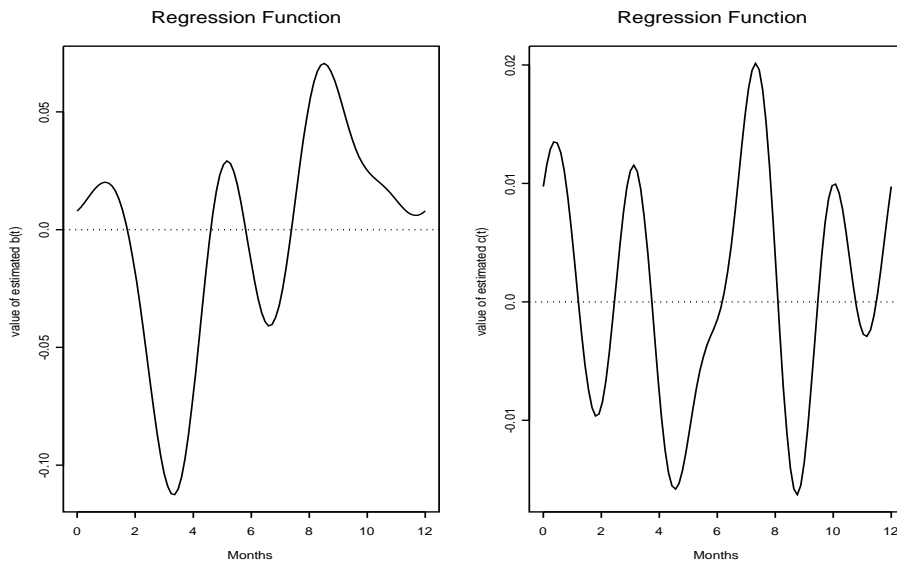
$$(۲۵) \quad \widehat{\log \text{Prec}}_i = \hat{\alpha} + \int_0^T \text{Temp}_i(t) \hat{\beta}(t) dt + \int_0^T \text{Humid}_i(t) \hat{\gamma}(t) dt \quad 1 \leq i \leq N,$$

پیش بینی کرد. شکل ۱۳ لگاریتم مقادیر مشاهده شده را در برابر مقادیر پیش بینی شده ی لگاریتم مقدار بارندگی نشان می دهد. این نمودار بیان می کند مقادیر پیش بینی شده بر اساس مدل (۲۵) نسبتاً قابل قبول می باشند. لازم به ذکر است ضریب تعیین در این حالت برابر $R^2 = 0.861$ می باشد.

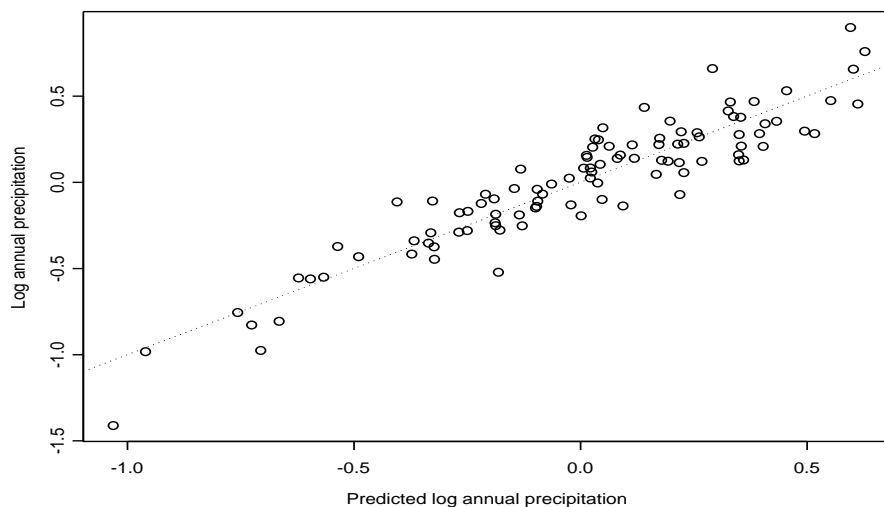
۵/۳/۲ متغیر وابسته y ، تابعی (مقدار لگاریتم بارندگی در زمان t) باشد.

در قسمت قبل مقدار لگاریتم بارندگی سالانه (اسکالر) را با استفاده از رطوبت و دما پیش بینی کردیم. با این حال بهتر است مقدار لگاریتم بارندگی را در طول سال و به صورت یک منحنی گزارش کنیم. بر این اساس در ادامه با تعریف $y(t)$ به عنوان تابع مقدار لگاریتم بارندگی مدل (۲۰) قابل تعریف می باشد و بنا بر این برآورد ضرایب $\hat{\gamma}$ را با استفاده از فرمول (۲۲) می توان محاسبه نمود.

نمودارهای (a) و (c) در شکل ۱۴ به ترتیب نمایش سه بُعد نما و منحنی خطوط تراز دما و نمودارهای



شکل ۱۲. شیب برآورد شده‌ی رگرسیون وقتی که متغیر وابسته، لگاریتم کل بارندگی سالانه و متغیر مستقل دما و میزان رطوبت است. نقطه‌ی ثابت $\hat{\alpha} = -0.2700388$ برآورد شده است.



شکل ۱۳. نمودار مشاهده‌شده (لگاریتم مقدار بارندگی ایران در سال ۲۰۰۶) در برابر مقادیر پیش‌بین‌شده وقتی متغیر مستقل دما و رطوبت باشد. خط راست روی نمودار، نشان‌دهنده‌ی مقدار مشاهده‌شده در مقابل مقادیر پیش‌بینی‌شده می‌باشد.

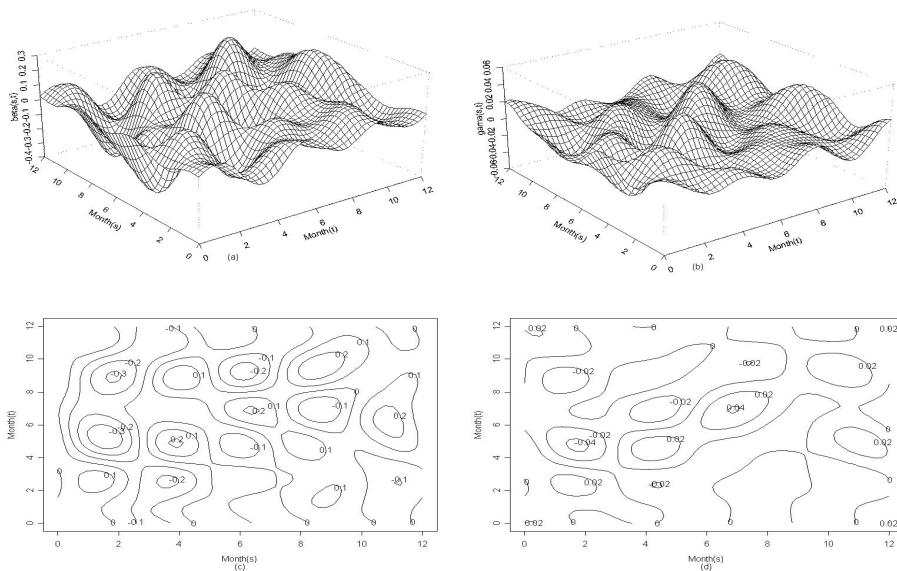
(b) و (d) در شکل ۱۴ به ترتیب نمایش سه بُعد نما و منحنی خطوط تراز رطوبت شیب برآورد شده‌ی رگرسیون خطی تابعی را نشان می‌دهند. بر اساس این نمودارها، تأثیر معکوس دما در ماه فوریه (اواسط بهمن تا اواسط اسفند) بر بارندگی در ماه سپتامبر (اواسط شهریور تا اواسط مهر) و آوریل (اواسط فروردین تا اواسط اردیبهشت) بیش‌تر می‌باشد. به علاوه دما در طول سال روی لگاریتم بارندگی در دسامبر (اواسط آذر تا اواسط دی) تأثیر چندانی ندارد. بیش‌ترین تأثیر رطوبت در ماه ژوئن (اواسط خرداد تا اواسط مهر) روی بارندگی در ماه ژوئن (اواسط خرداد تا اواسط مهر) می‌باشد. رطوبت روی مقدار بارندگی در اواخر سال تأثیری ندارد. ارتباط رطوبت در اوایل ماه فوریه (اواسط بهمن) بر مقدار بارندگی در آوریل (اوایل فروردین تا اوایل اردیبهشت) به صورت معکوس مشاهده می‌شود.

با بررسی تابع R^2 ، نمودار (a) در شکل ۱۵، به عنوان یک معیار برای اندازه‌گیری کیفیت مدل مشاهده می‌شود که مقدار آن در ماه‌های دوم تا اواسط ششم (اواسط بهمن تا اواسط تیر) و ماه‌های نهم تا یازدهم (اواسط شهریور تا اواسط آذر) سال بیش‌تر است و در نتیجه لگاریتم بارندگی در این دوره‌ی زمانی به خوبی به وسیله‌ی دما و رطوبت بیان می‌شود.

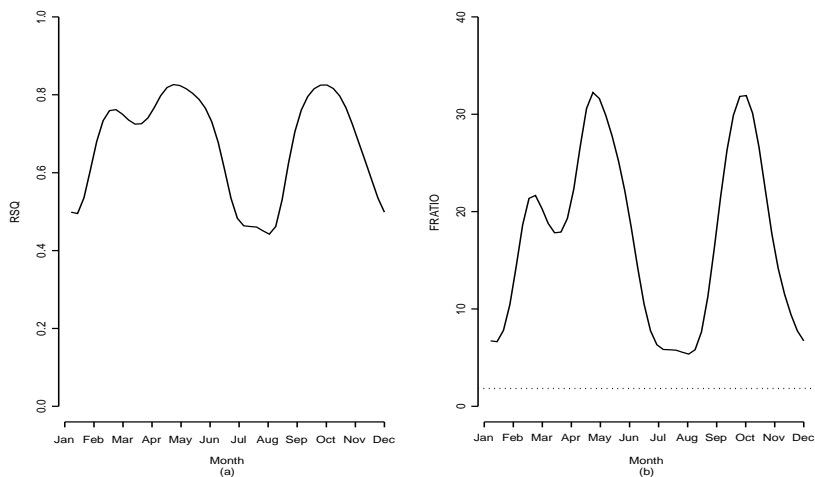
نمودار (b) در شکل ۱۵، $F - \text{ratio}$ را برای برازش داده‌های لگاریتم مقدار بارندگی نشان می‌دهد. در این نمودار، مقدار $F - \text{ratio}(t)$ در طول سال بالای خط $F(13, 88, 0.05)$ قرار دارد و این نشان‌دهنده‌ی این مطلب است که دما و رطوبت بر مقدار بارندگی اثر معنی‌داری دارند. بیش‌ترین مقدار این آماره به ترتیب در ماه‌های آوریل تا ژوئن (اواسط فروردین تا اواسط مرداد) و ماه‌های سپتامبر تا نوامبر (اواسط شهریور تا اواسط آذر) و کم‌ترین مقدار آن در ماه جولای (اواسط مرداد تا اواسط شهریور) می‌باشد.

۶ بحث و نتیجه‌گیری

در این مقاله مقدار بارندگی بر اساس دما یا رطوبت نسبی، با استفاده از مفاهیم رگرسیون خطی تابعی در دو حالت مجزا وقتی که متغیر وابسته (مقدار بارندگی) اسکالر یا تابعی باشد مدل‌بندی شده است. ابتدا با در نظر گرفتن هر یک از متغیرهای مستقل تابعی (دما یا رطوبت نسبی) در مدل به طور جداگانه، ضرایب مدل رگرسیونی برای هر یک از دو حالت (متغیر وابسته اسکالر یا تابعی) برآورد و با محاسبه‌ی معیارهای RSQ و $F - \text{ratio}$ ، مدل مورد ارزیابی قرار گرفت. در مرحله‌ی بعد، مدل قبل را با وارد کردن هم‌زمان هر دو متغیر مستقل دما و رطوبت نسبی توسعه دادیم. در هر یک از این حالت‌ها معلوم گردید که ملاک ارزیابی RSQ نسبت به حالتی که فقط یک متغیر مستقل در مدل وجود داشت، افزایش یافته است. به علاوه مقدار $F - \text{ratio}$ مشاهده‌شده نیز برای حالتی که دو متغیر مستقل به طور هم‌زمان در مدل قرار داشته‌اند، بزرگ‌تر است. در انتها می‌توان بیان نمود که نه فقط حضور هر یک از متغیرهای مستقل در مدل‌بندی معنی‌دار



شکل ۱۴. نمودارهای (a) و (b) به ترتیب نمایش سه بُعدی و منحنی خطوط تراز برآورد تابعی $\beta(s, t)$ با استفاده از ۷ پایه‌ی تابعی برای هر شناسه وقتی متغیر مستقل دما باشد را نمایش می‌دهد. به‌علاوه نمودارهای (b) و (d) نیز به طور مشابه برای متغیر مستقل رطوبت نسبی ترسیم شده‌اند.



شکل ۱۵. نمودار (a) تابع RSQ نسبت تغییرات بیان‌شده از لگاریتم مقدار بارندگی به‌وسیله‌ی مدل (۲۰) و نمودار (b) تابع F - ratio برای پیش‌بینی لگاریتم بارندگی بر اساس دما و رطوبت. خط افقی $F(13, 88, 0.05)$ را نشان می‌دهد.

است، بلکه با وارد کردن هر دو متغیر مستقل در مدل، پیش بینی تغییرات بارندگی بر اساس دما و رطوبت نسبی با صحت بیش تری همراه است.

مرجع ها

- تازیکه میاندره، ن. و حسینی نسب، م.ا. (۱۳۸۷). تحلیل تابعی دما و بارندگی در ایران با استفاده از مؤلفه های اصلی تابعی، پذیرفته شده در مجله ی پژوهش های آماری ایران.
- تازیکه میاندره، ن.؛ خیرالله زاده، ن.؛ موقری، ه. و حسینی نسب، م.ا. (۱۳۸۷). تحلیل داده های تابعی: تعمیمی از تحلیل داده های چندمتغیره. پذیرفته شده در مجله ی اندیشه ی آماری.
- حسینی نسب، م.ا.؛ خیرالله زاده، ن.؛ تازیکه میاندره، ن. و موقری، ه. (۱۳۸۷). رگرسیون خطی تابعی، ارسال شده به مجله ی اندیشه ی آماری.

Bobrowski, A. (2009). *Functional Analysis for Probability and Stochastic Processes*. Cambridge University Press, New York.

Hall, P. and Hosseini-nasab, M. (2006). On properties of functional principal components analysis. *J. Roy. Statist. Soc. Ser. B: Stat. Methodol* **68**, 109-126.

Hall, P. and Hosseini-nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. Accepted by *Math. Proc. cambridge philos. soc.*

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, 2nd edition*. Springer, New York.

سید محمدابراهیم حسینی نسب

گروه آمار، دانشکده ی علوم ریاضی،

دانشگاه شهید بهشتی،

تهران، ایران.

رایانشانی: m.hosseininasab@sbu.ac.ir

نصیبه خیرالله زاده

گروه آمار، دانشکده ی علوم ریاضی،

دانشگاه تربیت مدرس،

تهران، ایران.

رایانشانی:

نورالله تازیکه میاندره

گروه آمار، دانشکده ی علوم ریاضی،

دانشگاه تربیت مدرس،

تهران، ایران.

رایانشانی: