مجله مدل‌سازی پیشرفته ریاضی، دوره ۷، شماره ۱، بهار و تابستان ۱۳۹۶

مدل‌سازی داده‌های مهندسی آب با استفاده از روش رگرسیون فازی استوار کمترین مربعات پیراسته

جلال چاچی[1] و مهدی روزبه

گروه آمار، دانشگاه سمنان

تاریخ دریافت: ۱۳۹۵/۳/۲۹ تاریخ پذیرش: ۱۳۹۶/۴/۳۰

چکیده: روش‌های برآوردیابی پارامترهای مدل‌های رگرسیون فازی کمترین مربعات خطا حساسیت (بسیار) زیادی نسبت به داده‌های پرت دارند. اغلب روش‌های موجود برآوردیابی پارامترهای این مدل‌ها با رویکرد کمترین مربعات خطا، تحت تأثیر داده‌های پرت، برآوردهایی نامناسب، دور از انتظار و با خطای زیاد ارائه می‌دهند. لذا در این مطالعه یک مدل رگرسیون فازی استوار کمترین مربعات پیراسته برای مدل‌سازی متغیرهای ورودی حقیقی-مقدار و متغیر خروجی فازی-مقدار معرفی خواهد شد. در این رویکرد، تابع هدف در برآوردیابی پارامترهای مدل به‌گونه‌ای ساختاربندی می‌شود که مجموع h تا از کوچک‌ترین توان دوم باقیمانده‌های مرتب‌شده کمینه شوند. این روش دارای الگوریتمی است که با جستجو در مجموعه مشاهدات به برآورد بهترین پارامترهای مدل بر اساس ترکیب‌های مختلف انتخاب h مشاهده خوب از مجموعه n تایی مشاهدات، می‌پردازد. این موضوع باعث کاهش تأثیر مشاهدات پرت در فرآیند برآوردیابی پارامترهای مدل می‌شود. در انتها کاربرد روش پیشنهادی این مقاله در مدل‌سازی داده‌های واقعی در مهندسی آب (آب‌شناسی) که اغلب شامل مشاهدات پرت هستند، موردبررسی و مطالعه قرار می‌گیرد. ازاین‌رو، در این مطالعه به مقایسه بین روش پیشنهاد شده در این مقاله و روش متداول رگرسیون کمترین مربعات فازی که در آن مشاهدات پرت و مشاهدات خوب تأثیر یکسانی در برآوردیابی پارامترهای مدل دارند، پرداخته می‌شود. نتایج تجربی این مطالعه کاربردی برتری برازش بهتر روش پیشنهادی بر این داده‌ها را در مقایسه با روش متداول رگرسیون فازی کمترین مربعات خطا نشان می‌دهد. همچنین روش پیشنهاد شده در این مقاله مشاهدات پرتی را که تأثیر نامطلوبی در برآوردیابی پارامترها داشته‌اند را مشخص نموده است.

واژه‌های کلیدی: رگرسیون فازی، رگرسیون کمترین مربعات پیراسته، داده پرت، دبی معلق.

رده‌بندی ریاضی (۲۰۱۰): ۰۳E۷۲، ۶۲G۰۸.

۱- آدرس الکترونیکی نویسنده مسئول مقاله: jchachi@semnan.ac.ir

۱- مقدمه

بررسی داده‌های پرت در هر تجزیه و تحلیل آماری، به‌خصوص در مدل‌سازی‌های مبتنی بر شیوه کمترین مربعات خطا (ازجمله مدل‌های رگرسیون خطی و یا غیرخطی)، از اهمیت بسیاری برخوردار است. در مدل‌های رگرسیون مبتنی بر شیوه کمترین مربعات خطا (شامل رویکردهای آماری و یا فازی) مشاهدات پرت در نمونه (حتی به تعداد اندک) تأثیر مخرب و نامناسبی بر برآوردهای به‌دست آمده دارند. این مشاهدات باعث می‌شوند که مقادیر برآوردهای به‌دست آمده دور از انتظار، غیرقابل اطمینان و دارای خطای زیاد باشند. ازاین‌جهت به‌منظور برآورد استوار پارامترها، مجموعه مشاهدات نمونه باید با هدف شناسایی نقاط پرت موردبررسی و تحلیل قرار گیرد [۱ و ۲]. در صورت مشاهده نقطه (یا نقاط) پرت در مجموعه داده‌ها برآوردگرهای مبتنی بر شیوه کمترین مربعات خطا از دقت کافی در برآورد پارامترها برخوردار نمی‌باشند. برای رفع این اشکال برآوردگرهای استوار که مبتنی بر روش‌های برآوردیابی استوار هستند، جایگزین بسیار مناسبی برای برآورد پارامترها در حضور مشاهدات پرت می‌باشند [۳]. روش‌های برآوردیابی استوار اغلب به روش‌هایی گفته می‌شود که در حضور مشاهدات پرت در نمونه (و حتی بدون حذف نقطه یا نقاط پرت) برآوردگرهای دقیق و استواری برای پارامترهای مدل ارائه دهند [۴]. این‌گونه روش‌ها همواره باید دو از دو جهت زیر موردتوجه قرار گیرند [۵]:

۱. روش برآوردیابی توانایی شناسایی داده‌های پرت را داشته باشد،

۲. برآوردیابی استوار پارامترهای مدل در حضور مشاهدات پرت (بدون حذف آن‌ها از نمونه) انجام شود.

تاکنون به‌طور عمده می‌توان گفت که بررسی داده‌های پرت در روش‌های برآوردیابی استوار مدل‌های رگرسیون فازی از دو جهت بالا موردتحقیق و مطالعه قرار گرفته است [۵-۷]. از طرفی توجه کنید که در چارچوب مدل‌سازی رگرسیونی در محیط فازی (متغیرهای ورودی-خروجی فازی) داده‌های پرت تنوع و گسترگی رخداد بیشتری نسبت به حالت‌های مدل‌سازی رگرسیون آماری دارند [۸-۱۲]. در این حالت یک داده پرت می‌تواند به‌صورت حالات زیر رخ دهد:

۱. بر حسب مشاهدات متغیر حقیقی-مقدار ورودی باشد،

۲. بر حسب مشاهدات مراکز و/یا پهناهای متغیر فازی-مقدار خروجی باشد،

۳. بر حسب مدل رگرسیونی باشد (اما نه بر حسب مشاهدات متغیرهای ورودی و یا خروجی)،

۴. بر حسب مشاهدات متغیرهای ورودی و یا خروجی باشد (اما نه بر حسب مدل رگرسیونی)،

۵. بر حسب ترکیبی از موارد بالا باشد و دیگر اینکه در مدل رگرسیون فازی مقادیر متغیر ورودی نیز فازی باشد که این خود می‌تواند بر حسب مشاهدات مراکز و/یا پهناهای آن، مقادیری پرت داشته باشد.

ازاین‌رو با توجه به مطالب بالا، گستردگی و تنوع رخداد مشاهدات پرت در مطالعات مدل‌سازی رگرسیون در محیط فازی باعث معرفی روش‌ها و رویکردهای مختلف و متنوعی در برآوردیابی استوار پارامترهای مدل‌های رگرسیون فازی شده است. به‌طور عمده این روش‌ها را می‌توان به دو رویکرد اصلی زیر دسته‌بندی نمود:

الف- رویکردهای مبتنی بر برآوردیابی استوار پارامترهای مدل رگرسیون فازی (اغلب با حضور داده‌های پرت): این‌گونه روش‌ها که اخیراً بسیار موردتوجه قرار گرفته‌اند، عمدتاً بر مبنای استفاده از روش‌های استوار کلاسیک در محیط فازی هستند. در این‌گونه روش‌ها اغلب تابع هدف، که پارامترهای بهینه مدل از طریق کمینه کردن (و یا برخی حالات بیشینه کردن آن) به‌دست می‌آیند، به گونه انتخاب می‌شود که نسبت به مشاهدات پرت حساسیت نداشته باشد (یا به عبارتی نسبت به داده‌های پرت استوار باشد) [۵-۷].

ب- رویکردهای مبتنی بر تشخیص و شناسایی داده‌های پرت: این‌گونه روش‌ها عمدتاً مبتنی بر نمایش تابعی داده‌ها (به‌عنوان مثال نمودار پراکنش داده‌ها) و/یا روش‌های مبتنی بر تجزیه و تحلیل داده‌های فازی هستند [۱۳-۱۵].

در این مقاله به معرفی مدل رگرسیون فازی استوار کمترین مربعات پیراسته برای داده‌های ورودی دقیق (حقیقی-مقدار) و خروجی فازی پرداخته می‌شود. همچنین فرض می‌شود مجموعه مشاهدات شامل داده‌های پرت هستند و روش‌های مدل‌سازی مبتنی بر رویکرد کمترین مربعات خطا در مدل‌سازی این داده‌ها مناسب نمی‌باشند. روش پیشنهادی به‌گونه‌ای عمل می‌کند که در حضور داده‌های پرت، برآوردهای استواری برای پارامترهای مدل ارائه می‌دهد. روش برآوردیابی بهینه پارامترهای مدل رگرسیون فازی بر مبنای روش استوار کمترین مربعات پیراسته است. در انتها نتایج تجربی مدل پیشنهاد شده در مدل‌سازی داده‌های واقعی در مهندسی آب برتری این روش را در مقایسه با چند روش معمول و متداول رگرسیون فازی کمترین مربعات خطا را نشان می‌دهد.

مطالب این مقاله به‌صورت زیر تدوین شده است. در بخش ۲، برخی از مفاهیم و تعاریف موردنیاز از مجموعه‌های فازی بیان می‌شوند. در بخش ۳، روش برآوردیابی مدل رگرسیون استوار فازی کمترین مربعات پیراسته بیان می‌شود. در بخش ۴، با استفاده از داده‌های واقعی در مهندسی آب به مقایسه بین روش پیشنهادی و چند روش مطرح دیگر در رگرسیون فازی پرداخته می‌شود. در

انتها به بحث و نتیجه‌گیری و بحث پیرامون رویکردهای آتی در مدل‌سازی رگرسیونی در محیط فازی پرداخته می‌شود.

۲- مجموعه‌ها و حساب اعداد فازی

در این مقاله فرض کنید مجموعه اعداد حقیقی $\mathbb{R}$، مجموعه مرجع باشد. مجموعه فازی $\tilde{A}$ از $\mathbb{R}$ با تابع عضویت $\tilde{A}(x):\mathbb{R}\to[0,1]$ مشخص می‌شود [۱۶]. α-برش مجموعه فازی $\tilde{A}$ برای هر $\alpha\in(0,1]$ به‌صورت مجموعه معمولی $A_\alpha=\{x\in\mathbb{R}:\tilde{A}(x)\geq\alpha\}$ تعریف می‌شود و A_0 بستار مجموعه $\{x\in\mathbb{R}:\tilde{A}(x)>0\}$ است. مجموعه فازی $\tilde{A}$ از $\mathbb{R}$ را یک عدد فازی گوییم هرگاه برای هر $\alpha\in(0,1]$، مجموعه‌های $A_\alpha=\{x\in\mathbb{R}:\tilde{A}(x)\geq\alpha\}$ ناتهی، بسته و کران‌دار باشند. یک رده خاص از مجموعه‌های فازی در $\mathbb{R}$، اعداد فازی LR هستند. یک عدد فازی LR به‌صورت $\tilde{N}=(n,l,r)_{LR}$ نشان داده می‌شود که در آن $n\in\mathbb{R}$، مرکز، $l\in\mathbb{R}^+$ و $r\in\mathbb{R}^+$ به ترتیب پهناهای چپ و راست عدد فازی و $L:\mathbb{R}^+\to[0,1]$ و $R:\mathbb{R}^+\to[0,1]$ به‌ترتیب توابع شکل نزولی چپ و راست، با شرط $L(0)=R(0)=1$ هستند. تابع عضویت و α-برش عدد فازی $\tilde{N}=(n,l,r)_{LR}$ به‌صورت زیر تعریف می‌شود [۱۶]

$$\tilde{N}(x)=\begin{cases} L(\frac{n-x}{l}) & \text{if} \quad x\leq n, \\ R(\frac{x-n}{r}) & \text{if} \quad x\geq n. \end{cases}$$

$$\begin{aligned} N_\alpha &=[n-L^{-1}(\alpha)l,n+R^{-1}(\alpha)r], \quad \alpha\in[0,1], \\ &=[N^l_\alpha,N^u_\alpha]. \end{aligned}$$

عدد فازی $\tilde{N}=(n,l,r)_{LR}$ با شرایط $L=R$ و $l=r=\lambda$ یک عدد فازی LR متقارن نامیده می‌شود و به‌صورت $\tilde{N}=(n,\lambda)_L$ نشان داده می‌شود. اکنون فرض کنید $\tilde{N}=(n,l_n,r_n)_{LR}$ و $\tilde{M}=(m,l_m,r_m)_{LR}$ دو عدد فازی LR باشند و $\lambda\in\mathbb{R}$ یک عدد حقیقی باشد. در این صورت حساب اعداد فازی به‌صورت زیر تعریف می‌شود

$$\lambda\otimes\tilde{M}=\begin{cases} (\lambda m,\lambda l_m,\lambda r_m)_{LR} & \text{if} \quad \lambda>0, \\ \mathcal{I}_{\{0\}} & \text{if} \quad \lambda=0, \\ (\lambda m,|\lambda|r_m,|\lambda|l_m)_{RL} & \text{if} \quad \lambda<0, \end{cases}$$

$$\lambda \oplus \tilde{M} = (\lambda + m, l_m, r_m)_{LR},$$
$$\tilde{M} \oplus \tilde{N} = (m+n, l_m + l_n, r_m + r_n)_{LR},$$

که در آن $\mathcal{I}_{\{0\}}$ تابع نشانگر مجموعه معمولی صفر است [۱۶].

۳- رگرسیون فازی استوار کمترین مربعات پیراسته

در مدل‌های رگرسیون فازی مشاهدات پرت تأثیر نامناسب و مخربی بر برآوردگرهای مبتنی بر رویکرد کمترین مربعات خطا دارند. لذا به‌منظور شناسایی مشاهدات پرت، مجموعه داده‌ها باید به‌دقت موردبررسی و مطالعه قرار گیرد. در صورت وجود مشاهدات پرت در مجموعه داده‌ها، مناسب است که از برآوردگرهای استوار در برآورد پارامترهای مدل استفاده شود. در برخی از روش‌های برآوردیابی استوار، مشاهدات پرت شناسایی شده از مجموعه داده‌ها کنار گذاشته می-شوند و سپس از یک رویکرد متداول برآوردیابی (غیر استوار) در برآورد پارامترهای مدل استفاده می‌شود؛ اما همان‌گونه که در بخش مقدمه بیان شد، روش‌های برآوردیابی استوار باید در حضور مشاهدات پرت به برآوردیابی پارامترهای مدل بپردازند. ازاین‌رو، در این بخش با استفاده از رویکرد کمترین مربعات پیراسته به مدل‌سازی مجموعه مشاهدات زیر

$$(\tilde{y}_1, \mathbf{x}_1), \ldots, (\tilde{y}_n, \mathbf{x}_n)$$

مربوط به یک یا چند متغیر مستقل و یک متغیر وابسته پرداخته می‌شود. در این مشاهدات $\tilde{y}_i = (y_i, l, r)_{LR}$ و $\mathbf{x}_i = [x_{0i}, x_{1i}, \ldots, x_{ki}] \in \mathbb{R}^{k+1}$ $(i = 1, \ldots, n; k < n; x_{0i} = 1)$ به ترتیب i-امین مشاهدات متغیرهای وابسته و مستقل هستند. بدون اینکه از کلیت مسئله کاسته شود فرض می‌شود مشاهدات مربوط به متغیرهای مستقل نامنفی هستند. در صورت منفی بودن مشاهدات متغیرهای مستقل با استفاده از روابط خطی آن‌ها را نامنفی می‌کنیم. بدیهی است که تغییر خطی مقادیر متغیرها تأثیری در مدل بهینه برآورد شده ندارد. اکنون به‌منظور مدل‌سازی داده‌های بیان شده، مدل رگرسیون فازی زیر در نظر گرفته می‌شود

$$\begin{aligned}(y, G(l), G(r))_{LR} &= \tilde{\beta}_0 \oplus (\tilde{\beta}_1 \otimes x_1) \oplus \ldots \oplus (\tilde{\beta}_k \otimes x_k) \\ &= (\beta_0, \gamma_0, \delta_0)_{LR} \oplus ((\beta_1, \gamma_1, \delta_1)_{LR} \otimes x_1) \oplus \ldots \oplus ((\beta_k, \gamma_k, \delta_k)_{LR} \otimes x_k) \\ &= (\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \gamma_0 + \gamma_1 x_1 + \ldots + \gamma_k x_k, \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)_{LR}\end{aligned}$$

که در آن $G: (0, \infty) \to \mathbb{R}$ تابعی معکوس‌پذیر است (نقش این تابع در ادامه توضیح داده می‌شود [۹]). در انتها همان‌گونه که در قضیه ۱ بیان می‌شود، پس از برآورد بهینه پارامترها و با توجه به معکوس‌پذیری تابع $G: (0, \infty) \to \mathbb{R}$، مدل رگرسیون فازی بالا به‌صورت زیر نوشته می‌شود

$$\begin{aligned}\hat{\tilde{y}}_{LTS} &= (\hat{y}, \hat{l}, \hat{r})_{LR} \\ &= (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k, G^{-1}(\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \ldots + \hat{\gamma}_k x_k), \\ &\quad G^{-1}(\hat{\delta}_0 + \hat{\delta}_1 x_1 + \ldots + \hat{\delta}_k x_k))_{LR}\end{aligned}$$

در ادامه پارامترهای مدل رگرسیون فازی به‌گونه‌ای برآورد می‌شوند که داده‌های پرت تأثیر مخربی در روش برآوردیابی پارامترها نداشته باشند. به عبارتی شیوه برآورد بردار پارامترهای

$$\tilde{\boldsymbol{\beta}} = [\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_k]' = [(\beta_0, \gamma_0, \delta_0)_{LR}, (\beta_1, \gamma_1, \delta_1)_{LR}, \ldots, (\beta_k, \gamma_k, \delta_k)_{LR}]'$$

به‌گونه‌ای است که بر طبق یک فاصله بین اعداد فازی مجموع توان دوم باقیمانده‌ها (خطاها) بین مقادیر مشاهده شده متغیر وابسته و مقادیر برآورد شده آن توسط مدل، کمینه شود. باقیمانده‌ها یا خطاها که به‌صورت فاصله یا تفاضل بین مقادیر مشاهده شده متغیر وابسته و مقادیر برآورد شده آن تعریف می‌شوند نقشی بسیار مهم و اساسی در برآورد پارامترهای یک مدل رگرسیون (کلاسیک و یا فازی) دارند. با مطالعه و تجزیه و تحلیل آن‌ها اطلاعات بسیار مهمی از قبیل پرت بودن یک مشاهده می‌توان استخراج کرد. از طرفی مشاهدات پرت در یک مدل رگرسیون (کلاسیک و یا فازی) اغلب باقیمانده بزرگی دارند. در ادامه به‌منظور محاسبه باقیمانده‌ها (خطاها) و همچنین ساختاردهی مسئله بهینه‌سازی که پارامترهای بهینه مدل از آن به‌دست می‌آیند، از فاصله زیر بین دو عدد فازی $\tilde{M} = (m, l_m, r_m)_{LR}$ و $\tilde{N} = (n, l_n, r_n)_{LR}$ استفاده می‌شود [۱۳]

$$D^2(\tilde{M}, \tilde{N}) = (m-n)^2 + c_l (l_m - l_n)^2 + c_r (r_m - r_n)^2$$

که در آن $c_l = \int_0^1 \left(L^{-1}(\alpha)\right)^2 d\alpha$ و $c_r = \int_0^1 \left(R^{-1}(\alpha)\right)^2 d\alpha$. برای مثال، فاصله بین دو عدد فازی مثلثی $\tilde{M} = (m, l_m, r_m)_T$ و $\tilde{N} = (n, l_n, r_n)_T$ به‌صورت زیر محاسبه می‌شود

$$D^2(\tilde{M}, \tilde{N}) = (m-n)^2 + \frac{1}{3}(l_m - l_n)^2 + \frac{1}{3}(r_m - r_n)^2$$

که در آن $c_r = c_l = \int_0^1 (1-\alpha)^2 d\alpha = \frac{1}{3}$.

در ادامه برای برآورد استوار پارامترهای فازی مدل، مسئله بهینه‌سازی باید به‌گونه‌ای ساختاربندی شود که پس از شناسایی داده‌های پرت (که اغلب باقیمانده‌های بزرگی دارند) تأثیر مخرب آن‌ها بر پارامترهای برآورد شده کاهش یابد. لذا به‌منظور از بین بردن تأثیر مشاهدات پرت در تابع خطا، از رویکرد کمترین مربعات خطای پیراسته در ساختار تابع هدف استفاده می‌شود. در ادامه ساختار

مسئله بهینه‌سازی که پارامترهای بهینه مدل از آن به‌دست می‌آیند، در دو قسمت تابع هدف و قیدهای مربوط به آن، بررسی می‌شوند.

تابع هدف: گستردگی و تنوع رخداد مشاهدات پرت در مطالعات مدل‌سازی رگرسیون در محیط فازی در مقایسه با حالت کلاسیک بسیار بیشتر است [۵]. همان‌گونه که بیان شد، در یک مدل رگرسیون فازی با مقادیر متغیر ورودی حقیقی-مقدار و مقادیر متغیر خروجی فازی-مقدار، مشاهدات پرت می‌توانند برای مقادیر متغیر ورودی و/یا مقادیر مراکز و/یا پهناهای متغیر خروجی فازی، ثبت شوند. اکنون چون می‌خواهیم داده‌های پرت تأثیر مخربی در برآورد پارامترهای مدل نداشته باشند، لذا تابع هدف باید به‌گونه‌ای ساختاربندی شود که در حالت کلی اگر مشاهده‌ای پرت بر حسب مقادیر متغیر ورودی و/یا مقادیر مراکز و/یا مقادیر پهناهای متغیر خروجی فازی وجود داشت، تأثیر مخرب آن‌ها را بر پارامترهای برآورد شده بی‌اثر نماید. برای این منظور، تابع هدف سه قسمتی زیر (که قسمت‌های آن به ترتیب مشاهدات پرت بر حسب مقادیر متغیر ورودی، مقادیر مراکز و مقادیر پهناهای متغیر خروجی فازی را بی‌اثر می‌کند) در برآورد پارامترهای مدل بکار برده می‌شود

$$SSE = SSE_{۱} + SSE_{۲} + SSE_{۳}$$

که در آن

$$\begin{aligned}
SSE_{۱} &= \sum_{i=۱}^{h_{۱}} \varepsilon_{(i)}^{۲},\ \varepsilon_{(۱)}^{۲} \le \varepsilon_{(۲)}^{۲} \le \ldots \le \varepsilon_{(h_{۱})}^{۲} \le \ldots \le \varepsilon_{(n)}^{۲},\ \varepsilon_{i}^{۲} \\
&= [y_i - (\beta_{\circ} + \beta_{۱} x_{۱i} + \cdots + \beta_k x_{ki})]^{۲} \\
SSE_{۲} &= c_l \sum_{i=۱}^{h_{۲}} \epsilon_{(i)}^{۲},\ \epsilon_{(۱)}^{۲} \le \epsilon_{(۲)}^{۲} \le \ldots \le \epsilon_{(h_{۲})}^{۲} \le \ldots \le \epsilon_{(n)}^{۲},\ \epsilon_{i}^{۲} \\
&= [G(l_i) - (\gamma_{\circ} + \gamma_{۱} x_{۱i} + \cdots + \gamma_k x_{ki})]^{۲} \\
SSE_{۳} &= c_r \sum_{i=۱}^{h_{۳}} \zeta_{(i)}^{۲},\ \zeta_{(۱)}^{۲} \le \zeta_{(۲)}^{۲} \le \ldots \le \zeta_{(h_{۳})}^{۲} \le \ldots \le \zeta_{(n)}^{۲},\ \zeta_{i}^{۲} \\
&= [G(r_i) - (\delta_{\circ} + \delta_{۱} x_{۱i} + \cdots + \delta_k x_{ki})]^{۲}
\end{aligned}$$

و $h_{۱}$، $h_{۲}$ و $h_{۳}$ به ترتیب تعداد مشاهدات خوب (غیر پرت) بر حسب مقادیر متغیر ورودی، مقادیر مراکز و مقادیر پهناهای متغیر خروجی فازی را مشخص می‌کنند [۱-۳].

قیدها: در مسئله بهینه‌سازی که در ادامه ساخته می‌شود، قیدها باید با توجه به مقادیری که پارامترهای مدل می‌توانند اختیار کنند تعیین شوند. لذا پارامترهای فازی مدل باید به‌گونه‌ای

برآورد شوند که دارای پهناهای برآورد شده نامنفی باشند. برای این کار می‌توان به روش‌های زیر عمل کرد:

۱. شرط نامنفی بودن پهناها، یعنی $\delta_j \in \mathbb{R}^+$ و $\gamma_j \in \mathbb{R}^+$ در مسئله بهینه‌سازی وارد شود [۷، ۸ و ۱۰].

۲. شرط‌های $\delta_j \in \mathbb{R}$ ، $\gamma_j \in \mathbb{R}$ ، $\widehat{l_i} \in \mathbb{R}^+$ و $\widehat{r_i} \in \mathbb{R}^+$ در مسئله بهینه‌سازی وارد شود. توجه شود که در این حالت در نهایت مقداری که برای پهناهای متغیر پاسخ برآورد می‌شود نامنفی است، اگرچه که ممکن است برخی از مقادیر برآورد شده برای δ_jها و/یا γ_jها منفی باشند [۷ و ۱۷]. ازآنجاکه دامنه مقادیر پارامترها در این روش، شامل دامنه مقادیر پارامترها در روش ۱ است، لذا پارامترهای برآورد شده در این روش در مقایسه با روش ۱ مقدار تابع هدف کمتری نتیجه می‌دهند.

۳. فرارو و همکاران [۹] با در نظر گرفتن تابع معکوس‌پذیر $G:(\circ,\infty) \to \mathbb{R}$ از پهناهای متغیر پاسخ، برآوردهایی نامنفی برای پهناهای متغیر پاسخ به‌دست آوردند.

در ادامه، با الهام از روش‌های ۲ و ۳ در بالا، روش زیر در شکل‌گیری قیدهای مسئله بهینه‌سازی در نظر گرفته می‌شود. ابتدا تابع معکوس‌پذیر $G:(\circ,\infty) \to \mathbb{R}$ از پهناهای متغیر پاسخ در نظر گرفته می‌شود (یک انتخاب مناسب برای تابع $G(t)$ تابع $Log(t)$ است). سپس با در نظر گرفتن مقادیر $G(l_i)$ و $G(r_i)$ در محاسبه تابع هدف به برآورد مقادیر $\delta_j \in \mathbb{R}$ ،$\gamma_j \in \mathbb{R}$ در مسئله بهینه‌سازی پرداخته می‌شود. اکنون بدون توجه به علامت مقادیر برآورد شده برای δ_jها و γ_j ها (ممکن است برخی از مقادیر برآورد شده برای δ_jها و/یا γ_jها منفی به‌دست آیند) برآوردهای زیر

$$\hat{y} = \hat{\beta}_\circ + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$
$$\widehat{G}(l) = \hat{\gamma}_\circ + \hat{\gamma}_1 x_1 + \ldots + \hat{\gamma}_k x_k$$
$$\widehat{G}(r) = \hat{\delta}_\circ + \hat{\delta}_1 x_1 + \ldots + \hat{\delta}_k x_k$$

به‌دست می‌آیند. در انتها چون تابع $G(.)$ معکوس‌پذیر است، معادلات زیر نتیجه می‌شوند

$$\hat{y} = \hat{\beta}_\circ + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$
$$\hat{l} = G^{-1}\left(\hat{\gamma}_\circ + \hat{\gamma}_1 x_1 + \ldots + \hat{\gamma}_k x_k\right)$$
$$\hat{r} = G^{-1}\left(\hat{\delta}_\circ + \hat{\delta}_1 x_1 + \ldots + \hat{\delta}_k x_k\right)$$

توجه کنید که چون $G^{-1}:\mathbb{R}\to\mathbb{R}^{+}$ لذا برآوردهایی که برای پهناهای چپ و راست متغیر فازی خروجی به‌دست می‌آیند همواره مقادیری نامنفی دارند.

قضیه ۱: در مدل‌سازی مجموعه مشاهدات $(\tilde{y}_1,\mathbf{x}_1),\ldots,(\tilde{y}_n,\mathbf{x}_n)$ پارامترهای مدل رگرسیون فازی زیر

$$\begin{aligned}\tilde{y}_{LTS} &= (y,l,r)_{LR}\\ &= (\beta_0+\beta_1 x_1+\ldots+\beta_k x_k, G^{-1}(\gamma_0+\gamma_1 x_1+\ldots+\gamma_k x_k),\\ &\quad G^{-1}(\delta_0+\delta_1 x_1+\ldots+\delta_k x_k))_{LR}\end{aligned}$$

با استفاده از رویکرد کمترین مربعات پیراسته از مسئله بهینه‌سازی زیر به‌دست می‌آید

$$\begin{aligned}&Min \quad SSE = SSE_1+SSE_2+SSE_3\\ &s.t. \quad \beta_j\in\mathbb{R} \quad \gamma_j\in\mathbb{R} \quad \delta_j\in\mathbb{R} \quad j=0,1,\ldots,k\end{aligned}$$

در اجرای مسئله بهینه‌سازی بالا از نرم‌افزار R استفاده شده است [۱۸]. در نهایت پس از برآورد بهینه پارامترهای مدل مطابق قضیه بالا، مدل بهینه رگرسیون فازی به‌صورت زیر نوشته می‌شود

$$\begin{aligned}\hat{\tilde{y}}_{LTS} &= (\hat{y},\hat{l},\hat{r})_{LR}\\ &= (\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_k x_k, G^{-1}(\hat{\gamma}_0+\hat{\gamma}_1 x_1+\ldots+\hat{\gamma}_k x_k),\\ &\quad G^{-1}(\hat{\delta}_0+\hat{\delta}_1 x_1+\ldots+\hat{\delta}_k x_k))_{LR}\end{aligned}$$

نکته ۱: مدل رگرسیون فازی کمترین مربعات معمولی حالت خاصی از مدل رگرسیون فازی کمترین مربعات پیراسته (قضیه ۱) است. اگر در تعریف تابع هدف قضیه ۱ قرار دهیم $h_1=h_2=h_3=n$، مدل رگرسیون فازی کمترین مربعات پیراسته مطابق قضیه زیر تبدیل به یک مدل رگرسیون فازی کمترین مربعات متداول می‌شود.

قضیه ۲: در مدل‌سازی مجموعه مشاهدات $(\tilde{y}_1,\mathbf{x}_1),\ldots,(\tilde{y}_n,\mathbf{x}_n)$ پارامترهای مدل رگرسیون فازی زیر

$$\begin{aligned}\tilde{y}_{LS} &= (y,l,r)_{LR}\\ &= (\beta_0+\beta_1 x_1+\ldots+\beta_k x_k, G^{-1}(\gamma_0+\gamma_1 x_1+\ldots+\gamma_k x_k), G^{-1}(\delta_0+\delta_1 x_1+\ldots+\delta_k x_k))_{LR}\end{aligned}$$

با استفاده از رویکرد کمترین مربعات از مسئله بهینه‌سازی زیر به‌دست می‌آید

$$\begin{aligned}&Min \quad SSE = SSE_1+SSE_2+SSE_3\\ &s.t. \quad \beta_j\in\mathbb{R} \quad \gamma_j\in\mathbb{R} \quad \delta_j\in\mathbb{R}, \quad j=0,1,\ldots,k\end{aligned}$$

که در آن

$$SSE_1=\sum_{i=1}^{n}\varepsilon_i^2,\qquad \varepsilon_i^2=[y_i-(\beta_0+\beta_1x_{1i}+\cdots+\beta_kx_{ki})]^2$$
$$SSE_2=c_l\sum_{i=1}^{n}\epsilon_i^2,\qquad \epsilon_i^2=[G(l_i)-(\gamma_0+\gamma_1x_{1i}+\cdots+\gamma_kx_{ki})]^2$$
$$SSE_3=c_r\sum_{i=1}^{n}\zeta_i^2,\qquad \zeta_i^2=[G(r_i)-(\delta_0+\delta_1x_{1i}+\cdots+\delta_kx_{ki})]^2$$

در اجرای مسئله بهینه‌سازی بالا از نرم‌افزار R استفاده شده است [۱۸]. در نهایت پس از برآورد بهینه پارامترهای مدل مطابق قضیه بالا، مدل بهینه رگرسیون فازی به‌صورت زیر نوشته می‌شود

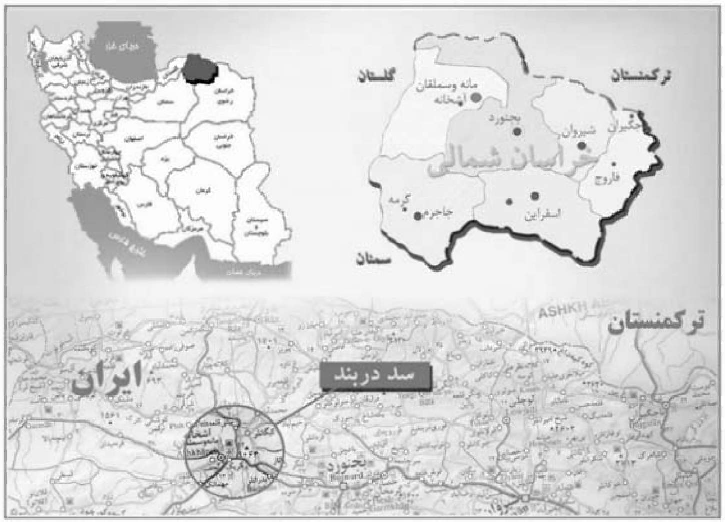
$$\begin{aligned}\hat{\tilde{y}}_{LS}&=(\hat{y},\hat{l},\hat{r})_{LR}\\&=(\hat{\beta}_0+\hat{\beta}_1x_1+\ldots+\hat{\beta}_kx_k,G^{-1}(\hat{\gamma}_0+\hat{\gamma}_1x_1+\ldots+\hat{\gamma}_kx_k).\\&\quad G^{-1}(\hat{\delta}_0+\hat{\delta}_1x_1+\ldots+\hat{\delta}_kx_k))_{LR}\end{aligned}$$

۴- مطالعه کاربردی در مهندسی آب

یکی از مسائل مهم در مهندسی آب اندازه‌گیری دبی یا بار معلق و سرعت آب در حوزه‌های آبریز است. در مطالعات کاربردی برآورد دبی بر اساس سرعت آب اهمیت زیادی دارد. لذا بر اساس مطالعه‌ای در دربند، واقع در خراسان شمالی (شکل ۱)، با استفاده از ابزارهای استاندارد، برخی از خصوصیات آب‌شناسی حوزه‌های آبریز منطقه ثبت گردید. در این خصوص از ایستگاه‌های مختلف اندازه‌گیری به تعداد ۵۱ جفت مشاهده مربوط به دبی و سرعت آب حوزه‌های آبریز منطقه جمع‌آوری شد. در این مشاهدات در سال‌های پرباران که اغلب با سیل همراه است، مشاهداتی به‌عنوان دورافتاده یا پرت ثبت می‌شوند. لذا در مدل‌سازی این مجموعه مشاهدات بهتر است که از روش‌های استوار در مدل‌سازی رگرسیون فازی استفاده کنیم. از طرفی در مواقع سیل قادر به اندازه‌گیری دقیق بار معلق رودخانه‌ها نیستم و ازاین‌رو ماهیت این متغیر را به‌صورت نادقیق و فازی ثبت کرده‌ایم. بنا به محدودیت‌های آزمایشگاهی و ماهیت متغیر پاسخ، این مشاهدات به‌صورت اعداد فازی مثلثی متقارن گزارش شده‌اند (جدول ۱) [۶، ۱۱ و ۱۲]. نمودارهای پراکنش مراکز و پهناهای متغیر وابسته فازی در مقابل مقادیر متغیر مستقل به‌ترتیب در شکل‌های ۲ و ۳ رسم شده‌اند. در ادامه با استفاده از رگرسیون فازی استوار کمترین مربعات پیراسته به برآورد متغیر پاسخ بر اساس مقادیر متغیر مستقل پرداخته می‌شود. این مدل رگرسیون فازی استوار به‌صورت زیر به‌دست می‌آید:

$$\widehat{y}_{LTS} = \left(۴/۷۴۵ + ۰/۳۹۷x , \exp\{ ۰/۴۶۸ + ۰/۰۳۹x \} \right)_T$$

خلاصه‌ای از نتایج این مدل در جدول ۲ گزارش شده است که در ادامه مورد تحلیل و بررسی قرار خواهند گرفت.



شکل (۱): موقعیت جغرافیایی محل گردآوری و مطالعه میدانی داده‌های مهندسی آب

مطالعه مقایسه‌ای: در این قسمت مدل رگرسیون فازی استوار کمترین مربعات پیراسته که توسط قضیه ۱ معرفی شد، با مدل رگرسیون فازی کمترین مربعات که توسط قضیه ۲ معرفی می‌شود، مورد مقایسه قرار خواهد گرفت. اکنون با استفاده از قضیه ۲، مدل رگرسیون فازی کمترین مربعات برای برآورد متغیر پاسخ بر اساس مقادیر متغیر مستقل به‌صورت زیر به‌دست می‌آید:

$$\widehat{y}_{LS} = \left(۵/۶۰۵ + ۰/۲۰۵x , \exp\{ ۰/۴۷۷ + ۰/۰۴۱x \} \right)_T$$

به‌منظور مقایسه مدل‌های کمترین مربعات $\widehat{y}_{LS}$ و کمترین مربعات پیراسته $\widehat{y}_{LTS}$، جزئیات این مدل‌ها شامل مقادیر خطای برازش و مقادیر معیارهای نیکویی برازش در جدول ۲ گزارش‌شده است.

جدول (۱): مشاهدات متغیر وابسته فازی و مقادیر متغیر مستقل

No.	x	(y,l)	No.	x	(y,l)	No.	x	(y,l)
۱	۲/۹۸	(۱/۹، ۴/۹۰)	۱۸	۲/۰۱	(۱/۶، ۵/۳۶)	۳۵	۲/۶۸	(۱/۹، ۵/۳۰)
۲	۵/۵۶	(۲/۰، ۷/۲۳)	۱۹	۲/۵۳	(۲/۱، ۸/۴۷)	۳۶	۸/۵۳	(۲/۴، ۷/۹۹)
۳	۳/۵۹	(۱/۷، ۵/۵۳)	۲۰	۱۰/۲۰	(۲/۳، ۷/۹۶)	۳۷	۱۶۴	(۲/۱، ۶/۷۲)
۴	۱/۹۱	(۱/۸، ۴/۶۳)	۲۱	۱۰/۲۰	(۲/۳، ۵/۷۳)	۳۸	۲/۷۳	(۲/۳، ۷/۸۴)
۵	۶/۶۷	(۲/۴، ۸/۰۳)	۲۲	۱۰/۲۰	(۲/۴، ۵/۳۷)	۳۹	۹/۰۰	(۲/۸، ۹/۶۲)
۶	۱۰/۲۵	(۲/۳، ۷/۴۱)	۲۳	۱۰/۲۵	(۲/۳، ۷/۹۶)	۴۰	۲/۴۳	(۱/۸، ۵/۸۱)
۷	۱۱/۱۲	(۲/۳، ۷/۹۰)	۲۴	۱۰/۵۳	(۲/۳، ۵/۱۴)	۴۱	۴/۹۷	(۲/۰، ۶/۸۰)
۸	۴/۷۶	(۲/۲، ۶/۵۷)	۲۵	۴/۲۸	(۱/۹، ۷/۱۵)	۴۲	۴/۴۶	(۱/۸، ۶/۲۶)
۹	۱/۳۶	(۱/۸، ۶/۳۱)	۲۶	۵/۰۳	(۱/۹، ۸/۱۰)	۴۳	۹/۶۸	(۲/۵، ۷/۷۹)
۱۰	۲/۳۶	(۱/۶، ۴/۹۸)	۲۷	۷/۲۰	(۲/۴، ۸/۰۴)	۴۴	۳/۰۵	(۱/۶، ۴/۸۳)
۱۱	۳/۱۱	(۱/۹، ۶/۳۴)	۲۸	۷/۲۰	(۲/۶، ۷/۵۶)	۴۵	۲/۳۲	(۱/۹، ۵/۳۷)
۱۲	۱/۵۴	(۱/۶، ۶/۲۶)	۲۹	۷/۲۰	(۱/۹، ۸/۲۹)	۴۶	۶/۵۳	(۲/۶، ۷/۸۳)
۱۳	۱/۹۸	(۱/۶، ۵/۱۴)	۳۰	۸/۷۵	(۲/۷، ۸/۹۶)	۴۷	۲/۴۵	(۱/۶، ۶/۹۰)
۱۴	۳/۹۷	(۱/۸، ۵/۵۱)	۳۱	۶/۷۱	(۲/۲، ۵/۵۰)	۴۸	۱/۵۸	(۱/۵، ۷/۶۶)
۱۵	۵/۲۰	(۲/۰، ۷/۸۶)	۳۲	۲/۰۴	(۱/۹، ۴/۶۹)	۴۹	۲/۲۵	(۱/۵، ۴/۹۰)
۱۶	۶/۹۴	(۱/۹، ۷/۲۴)	۳۳	۲/۹۲	(۱/۸، ۵/۹۵)	۵۰	۴/۲۴	(۱/۸، ۶/۹۱)
۱۷	۴/۱۶	(۱/۹، ۵/۹۸)	۳۴	۳/۱۶	(۱/۹، ۶/۶۸)	۵۱	۹/۹۷	(۲/۳، ۷/۸۸)

مقایسه بر اساس مقدار خطاهای برازش: طبق نتایج این جدول مقدار SSE برای مدل رگرسیون فازی کمترین مربعات پیراسته ۳۵/۲۸ است که بسیار کمتر از ۶۳/۴۷، مقدار SSE برای مدل رگرسیون فازی کمترین مربعات است. کاهش مقدار خطای مدل رگرسیون فازی استوار کمترین مربعات پیراسته در مقایسه با مدل رگرسیون فازی کمترین مربعات به این دلیل است که روش رگرسیون فازی استوار کمترین مربعات پیراسته تعداد ۴ داده پرت را در مقادیر مراکز متغیر وابسته فازی و تعداد ۴ داده پرت را در مقادیر پهناهای متغیر وابسته فازی شناسایی کرده است و تأثیر آن‌ها را در برآورد پارامترهای مدل از بین برده است. نمودارهای پراکنش مراکز و پهناهای متغیر وابسته فازی در مقابل مقادیر متغیر مستقل به‌ترتیب در شکل‌های ۲ و ۳ رسم

شده‌اند. در این دو شکل مشاهداتی که با دایره نشان داد شده‌اند مقادیری از داده‌ها هستند که توسط روش کمترین مربعات پیراسته به‌عنوان مشاهدات پرت شناسایی شده‌اند.

جدول (۲): مقایسه نیکویی برازش و مقادیر خطا در مدل‌های کمترین مربعات $\hat{\tilde{y}}_{LS}$ و کمترین مربعات پیراسته $\hat{\tilde{y}}_{LTS}$

	کمترین مربعات $\hat{\tilde{y}}_{LS}$	کمترین مربعات پیراسته $\hat{\tilde{y}}_{LTS}$
SSE_1	۶۲/۰۱	۳۵/۰۸
SSE_2	۰/۷۳	۰/۱۰
SSE_3	۰/۷۳	۰/۱۰
SSE	۶۳/۴۷	۳۵/۲۸
MSM	۰/۴۹	۱/۴۰
MAE_1	۱/۴۹	۱/۴۰
MAE_2	۰/۷۴	۰/۷۲
h_1	۵۱	۴۷
h_2	۵۱	۴۷
h_3	۵۱	۴۷

مقایسه بر اساس معیارهای نیکویی برازش: برای مقایسه مدل‌های رگرسیون فازی از معیارهای نیکویی برازش نیز می‌توان استفاده نمود. تاکنون معیارهای نیکویی برازش مختلفی برای مقایسه برازش مدل‌های رگرسیون فازی معرفی شده است [۸، ۱۱ و ۱۲]. در این زمینه و به‌منظور مقایسه نیکویی برازش مدل پیشنهاد شده در این مقاله با رویکرد رگرسیون فازی کمترین مربعات از سه معیار زیر استفاده می‌شود:

$$MSM=\frac{1}{n}\sum_{i=1}^{n}\frac{\int \min\{\tilde{y}_i(t),\hat{\tilde{y}}_i(t)\}dt}{\int \max\{\tilde{y}_i(t),\hat{\tilde{y}}_i(t)\}dt},$$

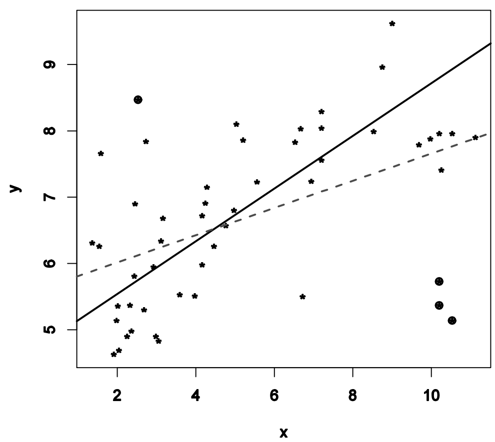
$$MAE_1=\frac{1}{n}\sum_{i=1}^{n}\int|\tilde{y}_i(t)-\hat{\tilde{y}}_i(t)|dt,$$

$$MAE_2=\frac{1}{n}\sum_{i=1}^{n}\frac{\int|\tilde{y}_i(t)-\hat{\tilde{y}}_i(t)|dt}{\int\tilde{y}_i(t)dt}.$$

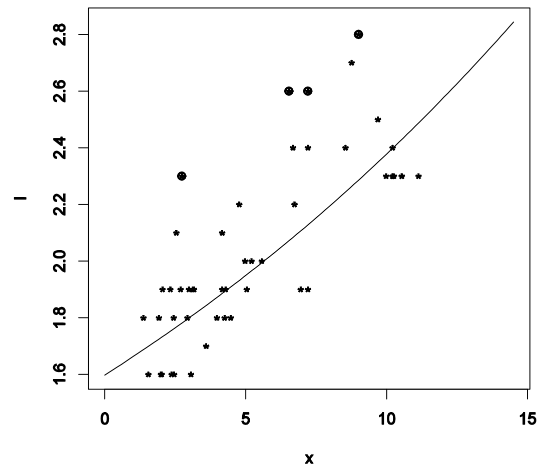
که در آن MSM میانگین اندازه‌های تشابه بین مقادیر مشاهده شده متغیر پاسخ و مقادیر برآورد شده متغیر پاسخ توسط مدل است و MAE_1 و MAE_2 میانگین قدرمطلق خطاها بین

مقادیر مشاهده شده متغیر پاسخ و مقادیر برآورد شده متغیر پاسخ توسط مدل هستند. مقادیر متناظر با سه معیار MSM، $MAE_۱$ و $MAE_۲$ برای دو مدل رگرسیون فازی در جدول ۲ بیان شده است. طبق این مقادیر، MSM برای مدل رگرسیون فازی کمترین مربعات پیراسته ۰/۵۱ است که کمی بیشتر از مقدار ۰/۴۹ مدل رگرسیون فازی کمترین مربعات است. همچنین مقادیر $MAE_۱$ و $MAE_۲$ برای مدل رگرسیون فازی کمترین مربعات پیراسته به‌ترتیب برابر ۱/۴۰ و ۰/۷۲ است که کمی کمتر از مقادیر ۱/۴۹ و ۰/۷۴ برای مدل رگرسیون فازی کمترین مربعات است.

توجه کنید که روش‌های برآوردیابی مبتنی بر رویکرد کمترین مربعات پیراسته نقطه شکست بالایی دارند، یعنی هر چه تعداد داده‌های پرت بیشتر و همچنین مقادیر مشاهده شده آن‌ها دورتر (فرین) باشد این‌گونه مدل‌ها بهتر عمل می‌کنند. لذا در مطالعات شبیه‌سازی شده به‌راحتی می-توان نتایجی به‌دست آورد که برتری قابل‌ملاحظه مدل رگرسیون فازی کمترین مربعات پیراسته را نتیجه گرفت.



شکل (۲): برازش مراکز متغیر وابسته فازی در مقابل مقادیر متغیر مستقل. مشاهداتی که با دایره نشان داده شده‌اند مقادیری از داده‌ها هستند که توسط روش کمترین مربعات پیراسته (خط پر) به‌عنوان مشاهدات پرت شناسایی شده‌اند.



شکل (۳): برازش پهناهای متغیر وابسته فازی در مقابل مقادیر متغیر مستقل. مشاهداتی که با دایره نشان داده شده‌اند مقادیری از داده‌ها هستند که توسط روش کمترین مربعات پیراسته (خط پر) به‌عنوان مشاهدات پرت شناسایی شده‌اند.

۵- بحث و نتیجه‌گیری

برآوردگرهای استوار که مبتنی بر روش‌های برآوردیابی استوار هستند، جایگزین بسیار مناسبی برای برآورد کمترین مربعات خطای پارامترها در حضور مشاهدات پرت می‌باشند. لذا در این مقاله یک روش برآوردیابی استوار مبتنی بر روش کمترین مربعات پیراسته برای برآورد پارامترهای یک مدل رگرسیون فازی معرفی شد که در حضور مشاهدات نامطلوب یا پرت در نمونه (بدون حذف نقطه یا نقاط پرت) برآوردگرهای بهتری در مقایسه با برآوردگرهای مبتنی بر روش کمترین مربعات فازی برای پارامترهای مدل به‌دست آورد. روش برآوردیابی پیشنهاد شده توانایی بی‌اثر ساختن مشاهداتی که دارای مقادیر خطاهای نسبتاً بزرگی هستند (که می‌توان این مشاهدات را داده‌های پرت یا نامطلوب نامید) را دارد و به برآورد پارامترهای مدل رگرسیون فازی در حضور این مشاهدات (بدون حذف آن‌ها از نمونه) می‌پردازد. در انتها در یک مطالعه مقایسه‌ای، رویکرد رگرسیون فازی پیشنهاد شده در این مقاله با رویکرد معروف و متداول رگرسیون فازی کمترین مربعات مورد مقایسه و بررسی قرار گرفت و نیکویی برازش این مدل‌ها با سه معیار MSM، $MAE_۱$ و $MAE_۲$ محاسبه شد. بر اساس نتایج عددی و مدل‌سازی یک مجموعه داده در محیط

فازی، برتری مدل رگرسیون فازی کمترین مربعات پیراسته در برازش به این مجموعه از داده‌ها نتیجه شد.

سپاسگزاری

نویسندگان مقاله از جناب آقای مهندس رضایی پژند در سازمان سدسازی و مهار آب خراسان رضوی کمال تشکر و قدردانی را بابت انتشار و در اختیار قرار دادن داده‌های واقعی بکار رفته در مثال مهندسی آب در این مقاله را دارند. همچنین از زحمات و پیگیری‌های مستمر سردبیر محترم مجله و نظرات ارزشمند داوران محترم که باعث بهبودی و اصلاح هرچه بیشتر مقاله گردید کمال تشکر و قدردانی را دارند.

منابع

[1] Huber, P. and Ronchetti, E.M. (2009). *Robust Statistics*, 2ed. Wiley, Hoboken, NJ.

[2] Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, Wiley, Hoboken, NJ.

[3] Andersen, R. (2007). *Modern Methods for Robust Regression*, Sage: Thousand Oaks, CA.

[4] Roozbeh, M. (2016). Robust ridge estimator in restricted semi-parametric regression models, *Journal Multivariate Analysis*, **147**, 127-144.

[5] D'Urso, P., Massari, R. and Santoro, A. (2011). Robust fuzzy regression analysis, Information Sciences, **181**, 4154-4174.

[6] Chachi, J. and Roozbeh, M. (2017). A fuzzy robust regression approach applied to bedload transport data, *Communications in Statistics-Simulation and Computation*, **46**, 1703-1714.

[7] Chachi, J., Taheri, S.M., Fattahi, S. and Hosseini Ravandi, S.A. (2017). Two robust fuzzy regression models and their applications to predict imperfections of cotton yarn, *Journal of Textiles and Polymers*, In Press.

[8] Arefi, M. and Taheri, S.M. (2015). Least squares regression based on Atanassov's intuitionistic fuzzy inputs-outputs and Atanassov's intuitionistic fuzzy parameters, *IEEE Transactions on Fuzzy Systems,* **23**, 1142-1154.

[9] Ferraro, M.B., Coppi, R., Gonzalez-Rodriguez, G. and Colubi, A. (2010). A linear regression model for imprecise response, *International Journal of Approximate Reasoning*, 51, 759-770.

[10] Chachi, J. and Taheri, S.M. (2016). Multiple fuzzy regression model for fuzzy input-output data, *Iranian Journal of Fuzzy Systems*, **13**, 63-78.

[11] Chachi, J., Taheri, S.M. and Arghami, N.R. (2014). A hybrid fuzzy regression model and its application in hydrology engineering, *Applied Soft Computing*, **25**, 149-158.

[12] Chachi, J., Taheri, S.M. and Rezaei Pazhand, H. (2016). Suspended load estimation using L_1-Fuzzy regression, L_2-Fuzzy regression and MARS-Fuzzy regression models, *Hydrological Sciences Journal*, **61**, 1489-1502.

[13] Coppi, R., D'Urso, P., Giordani, P. and Santoro, A. (2006). Least squares estimation of a linear regression model with *LR* fuzzy response, *Computational Statistics and Data Analysis*, **51**, 267-286.

[14] Hung, W.L. and Yang, M.S. (2006). An omission approach for detecting outliers in fuzzy regressions models, *Fuzzy Sets and Systems*, **157**, 3109-3122.

[15] Peters, G. (1994). Fuzzy linear regression with fuzzy intervals, Fuzzy Sets and Systems 63, 45-55.

[16] Zimmermann, H.J. (2001). *Fuzzy Set Theory and Its Applications*, 4th ed., Kluwer Nihoff, Boston.

[17] Chang, P.T. and Lee, S. (1994). Fuzzy linear regression with spreads unrestricted in sign, *Computers and Mathematics with Applications,* **28**, 61-70.

[18] Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. 2nd ed., Sage Publications: Thousand Oaks, CA.

Modeling Hydrology Data Using a Robust Least Trimmed Squares Fuzzy Regression Approach

Jalal Chachi, Mahdi Roozbeh

Department of Statistics, Semnan University, Semnan, Iran

Abstract

Estimation methods of parameters of fuzzy least-squares regression have very sensitivity to unusual data (e.g. outliers). In the presence of outliers, most of the existing estimation methods of parameters of this kind of models using least-squares approach provide unexpected and unreliable estimators with amounts of errors. Therefore, in this paper a robust least trimmed squares fuzzy regression model is described for modeling for crisp input-fuzzy output variables. In this approach, the constructed target function in model parameter estimation problem in such a way which minimizes the sum of the h smallest squared residuals. This method has an algorithm that estimates the optimal values of the parameters based on different selected combinations of h good observations of the data set of size n . Therefore, this method has the ability of reducing the effects of such a data in estimation of the parameters of the model. Finally, the investigated fuzzy regression model is applied and studied to modeling real-world data set in hydrology which sometimes contains outlier points. In this regard, a comparison study between the proposed method and ordinary least squares fuzzy regression method is considered. The comparison results of the applied study reveal that for this particular data set the proposed method performs better fitting than the well-known ordinary fuzzy least-squares regression model. Also the proposed method identified the points that have bad effect on estimation problem of the parameters.

Keywords: Fuzzy regression, Least trimmed squares regression; Outlier, Debi, Suspended load.

Mathematics Subject Classification (2010): 03E72, 62G08.