مجله علوم آماری، بهار و تابستان ۱۳۹۷
جلد ۱۲، شماره ۱، ص ۲۳۹ – ۲۵۳
DOI: 10.29252/jss.12.1.239

تحلیل آمارشکل تپه‌های ماسه‌ای اردستان در حضور خطای اندازه‌گیری

نقی همتی دیارجان، موسی گل‌علی‌زاده

گروه آمار، دانشگاه تربیت مدرس

تاریخ دریافت: ۱۳۹۶/۳/۱۴ تاریخ آخرین بازنگری: ۱۳۹۷/۱/۲۶

چکیده: داده‌های شکل با توجه به تعدد منابع خطاها اغلب در معرض ابتلا به خطای اندازه‌گیری قرار دارند. نادیده گرفتن چنین خطایی در صورت وجود، باعث بروز مشکلات زیادی از قبیل اریبی برآوردگرها می‌شود. در این حالت برآوردگرهای حاصل از عدم دخالت خطای اندازه‌گیری برآوردگرهای ناپخته نامیده می‌شوند. برآوردگرهای ناپخته برای پارامترهای مقیاس و دوران در هنگام استفاده از انطباق پروکراستس داده‌های شکل دو بعدی، اریب هستند. برای تصحیح اریبی و بهبود برآوردگرهای ناپخته، در این مقاله روش‌های کالبیدن رگرسیون که از طریق بکارگیری مدل‌های رگرسیونی مختلط و توزیع نرمال مختلط حاصل می‌شود و همچنین روش امتیاز شرطی پیشنهاد می‌شود. به علاوه، با انجام شبیه‌سازی آماری عملکرد این روش‌ها مورد مطالعه قرار می‌گیرند. هم‌چنین، تحلیل آماری مربوط به شکل تپه‌های ماسه‌ای اردستان، با فرض وجود خطای اندازه‌گیری در مشاهدات، انجام می‌شود.

واژه‌های کلیدی: تحلیل پروکراستس، خطای اندازه‌گیری، توزیع نرمال مختلط، روش‌های تصحیح اریبی، داده‌های لندفرم.

۱ مقدمه

تحلیل آمارشکل شاخه جدیدی از آمار چند متغیره است که راجع به هندسه اشیاء بدون توجه به مکان، مقیاس و دوران آن‌ها بحث می‌کند. یکی از مرسوم‌ترین رویکرد در این تحلیل، مطالعات اشکال (پیکربندی‌های

آدرس الکترونیکی نویسنده مسئول مقاله: موسی گل‌علی‌زاده، golalizadeh@modares.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): 62H99، 62J05

هندسی [۱]) بر اساس تعدادی نقاط شاخص[۲] روی سطح اشیاء و قرار دادن مختصات دکارتی این نقاط در ماتریس پیکربندی [۳] است. به عنوان مثال پیکربندی هندسی در فضای $\mathbb{R}^m$ با L نقطه شاخص توسط ماتریس $\mathbf{X}_{L\times m}$ نمایش داده می‌شود. بنا به درایدن و ماردیا (۱۹۹۸) و ماردیا و همکاران (۱۹۷۹) انطباق دو شی $\mathbf{X}$ و $\mathbf{Y}$ توسط تحلیل پروکراستس عادی [۴] (OPA) صورت می‌گیرد.

با توجه به تعدد منابع خطا اطلاعات شکل نیز می‌تواند در معرض ابتلا به خطای اندازه‌گیری قرار گیرد. به عنوان مثال در عکس‌برداری‌های پزشکی، منابع خطای ممکن به جز خطای تصادفی ذاتی، شامل خطاهای ناشی از دستگاه اسکن و تنوع بین اپراتورها خواهد بود. با این حال معمولا خطای اندازه‌گیری در تحلیل‌های مرسوم آمار شکل در نظر گرفته نمی‌شود. فولر (۱۹۸۰) نشان داد که نادیده گرفتن خطای اندازه‌گیری در مدل رگرسیون خطی معمولی منجر به برآوردگر اریب، معروف به برآوردگر ناپخته [۵] خواهد شد. لذا تصحیح اریبی رویکردی است که در چنین مواقعی دنبال می‌شود. بیشتر مطالعات راجع به مدل‌های رگرسیونی در حضور خطای اندازه‌گیری تنها معطوف به مدل‌هایی با متغیرهای تصادفی حقیقی مقدار بوده است. به عنوان مثال، استفانسکی (۱۹۸۵) را ببینید. معمولا متغیرهای تصادفی دو بعدی مورد مطالعه در آمارشکل به صورت متغیرهای مختلط توصیف می‌شوند. لذا، بررسی تاثیر خطای اندازه‌گیری برای داده‌های دو بعدی آمارشکل و همچنین ارزیابی روش‌های متفاوت تصحیح اریبی در این حوزه از اهمیت زیادی برخوردار است که مقاله حاضر به این موضوع می‌پردازد.

کالیبدن رگرسیون [۶] یکی از روش‌های تصحیح اریبی برای مدل رگرسیونی تحت تأثیر خطای اندازه‌گیری است که از الگوریتم کالیبدن استفاده می‌کند. این الگوریتم به عنوان یک دیدگاه کلی توسط کارول و استفانسکی (۱۹۹۰) و گلسر (۱۹۹۰) پیشنهاد شد. به خاطر وجود مشکلاتی در شکل اولیه این روش، یک اصلاح از آن توسط کلایتون (۱۹۹۲) پیشنهاد شد. آرمسترانگ (۱۹۸۵) کالیبدن رگرسیون را برای مدل‌های خطی تعمیم‌یافته استفاده و فولر (۱۹۸۷) آن را برای مدل‌های خطی عمومی به کار برد. روش دیگر برای تصحیح اریبی، روش امتیاز شرطی [۷] است. این روش طوری طراحی شده است که استنباط‌های سازگار را تحت فرض‌های ناچیز مدل تضمین می‌کند (گلسر، ۱۹۸۱). ایده اصلی این روش برای به دست آوردن برآوردگرها مبتنی بر یک تابع امتیاز نااریب است که با شرطی کردن روی یک آماره بسنده برای پیشگوی

[1] Geometrical configurations
[2] Landmark
[3] Configuration matrix
[4] Procrustes Ordinary Analysis
[5] Naive
[6] Regression calibration
[7] Conditional score

مستعد خطا ساخته می‌شود. سازگاری برآوردگرها از این حقیقت که آن‌ها، برآوردگرهای $M-$ هستند (استفانسکی و بووس، ۲۰۰۲)، ناشی می‌شود. گلسر (۱۹۸۱) روش امتیاز شرطی را برای مدل‌های خطی تعمیم داد. استفانسکی و کارول (۱۹۸۷) این روش را برای مدل‌های خطای اندازه‌گیری خطی تعمیم‌یافته ساختاری و تابعی نیز در نظر گرفتند.

در بخش ۲ مدل خطای اندازه‌گیری برای داده‌های شکل معرفی و اریبی برآوردگر ناپخته بررسی می‌شود. هم‌چنین برای تصحیح و کاهش اریبی برآوردگر ناپخته، دو روش کالیبدن رگرسیون و امتیاز شرطی برای برآورد پارامترهای مدل رگرسیونی مختلط در حضور خطای اندازه‌گیری تشریح می‌شود. مطالعات شبیه‌سازی برای ارزیابی روش‌های برآورد در بخش ۳ آمده است و در نهایت، تحلیل یک مثال واقعی که مربوط به داده‌های لندفرم از تپه‌های ماسه‌ای اردستان است در بخش ۴ ارائه می‌شود.

۲ مدل خطای اندازه‌گیری برای داده‌های شکل دو بعدی

فرض کنید X و Y دو پیکربندی مورد علاقه، هرکدام شامل $L(\geq ۳)$ نقطه شاخص باشند. واضح است که با نمایش عدد مختلط، $\mathbf{X}$ و $\mathbf{Y}$ عناصری L بعدی درفضای مختلط $\mathbb{C}^L$ هستند. بنابراین می‌توان از توزیع‌های مختلط برای مدل‌بندی پیکربندی‌های دو بعدی استفاده کرد. بنا به گوودمن (۱۹۶۳) یکی از مهم‌ترین توزیع‌های آماری در این حالت توزیع نرمال مختلط است. خانواده متغیرهای تصادفی نرمال مختلط متشکل از متغیرهای تصادفی ای هستند که بخش‌های حقیقی و مجازی آن‌ها تواماً نرمال هستند.

فرض کنید $\mathbf{U} = (u_۱, \ldots, u_L)^T$ و $\mathbf{V} = (v_۱, \ldots, v_L)^T$ بردارهای تصادفی در $\mathbb{R}^L$ هستند طوری که $(\mathbf{U}, \mathbf{V})^T = (u_۱, \ldots, u_L, v_۱, \ldots, v_L)^T$ یک بردار تصادفی نرمال $۲L$ بعدی است. در این صورت بردار تصادفی مختلط $\mathbf{X} = \mathbf{U} + i\mathbf{V}$ از توزیع نرمال مختلط پیروی می‌کند. پیروی متغیر تصادفی $\mathbf{X}$ از توزیع نرمال مختلط برای حالتی که بخش‌های حقیقی و مجازی آن مستقل از هم هستند به صورت $\mathbf{X} \sim \mathrm{CN}(\mu, ۲\sigma^۲)$ نمایش داده می‌شود، که در آن μ میانگین $\mathbf{X}$ و $۲\sigma^۲$ واریانس این متغیر است. توجه شود که در این‌جا واریانس هر کدام از بخش‌های متغیر تصادفی مختلط $\mathbf{X}$ برابر $\sigma^۲$ در نظر گرفته شده است.

برای انطباق $\mathbf{X}$ بر روی $\mathbf{Y}$ از طریق OPA، مدل خطی

$$\mathbf{Y} = \beta_۰ ۱_L + \beta_۱ \mathbf{X} + \epsilon \qquad (۱)$$

مد نظر قرار می‌گیرد، که در آن β_0 پارامتر انتقال، β_1 پارامتر مقیاس و دوران، 1_L بردار ستونی Lبعدی از یک‌ها و ϵ بردار خطای تصادفی است. به علاوه، $\mathbf{X}$, $\mathbf{Y}$ و ϵ بردارهای ستونی Lبعدی هستند. با پیروی از درایدن و ماردیا (۱۹۹۸)، انطباق $\mathbf{X}$ بر روی $\mathbf{Y}$ می‌تواند به عنوان یک مسئله کم‌ترین توان‌های دوم فرمول‌بندی شود. قابل اشاره است که $\mathbf{X}$ و $\mathbf{Y}$ هر دو به عنوان داده‌های مشاهده شده در نظر گرفته می‌شوند، اما مدل (۱) فقط عبارت خطای تصادفی ϵ برای $\mathbf{Y}$ را شامل می‌شود و خطای اندازه‌گیری ممکن در $\mathbf{X}$ را نادیده می‌گیرد. در حضور خطای اندازه‌گیری برای پیکربندی $\mathbf{X}$، مسئله انطباق دو پیکربندی می‌تواند به صورت مدل خطی مختلط با خطای اندازه‌گیری به ازای $\ell = 1, \cdots, L$ به صورت

$$\begin{cases} Y_\ell = \beta_0 + \beta_1 X_\ell + \epsilon_\ell \\ W_\ell = X_\ell + U_\ell \end{cases} \qquad (۲)$$

نوشته شود، که در آن $\mathbf{X} = (X_1, \ldots, X_L)^T$ و $\mathbf{Y} = (Y_1, \ldots, Y_L)^T$ متعلق به فضای مختلط $\mathbb{C}^L$، $\mathbf{W} = (W_1, \ldots, W_L)^T \in \mathbb{C}^L$ مشاهده آلوده به خطای اندازه‌گیری، $\mathbf{X}$ مقدار واقعی مشاهده، $\mathbf{U} = (U_1, \ldots, U_L)^T \in \mathbb{C}^L$ خطای اندازه‌گیری و $\epsilon = (\epsilon_1, \ldots, \epsilon_L)^T \in \mathbb{C}^L$ عبارت خطای تصادفی برای مدل‌بندی پاسخ $\mathbf{Y}$ است.

فرض می‌شود بردار متغیرهای تصادفی $\mathbf{X}$، $\mathbf{U}$ و ϵ از یکدیگر مستقل‌اند و به ازای $\ell = 1, \cdots, L$ داریم: $\epsilon_l \sim CN(0, 2\sigma_\epsilon^2)$ و $U_\ell \sim CN(0, 2\sigma_u^2)$. همچنین برای سادگی فرض می‌شود، $Re(U_\ell)$ و $Im(U_\ell)$ ناهمبسته و U_ℓها مستقل‌اند. علاوه براین، $\Theta = (\beta, \sigma_\varepsilon^2)^T$ پارامترهای نامعلوم مدل هستند، که در آن $\beta = (\beta_0, \beta_1)^T$. در ادامه خواص برآوردگر ناپخته β حاصل از نادیده گرفتن خطای اندازه‌گیری تحت مدل (۲)، که در اصل یک روش رگرسیون معمولی است، تشریح می‌شود.

۱.۲ روش رگرسیون معمولی

از انطباق $\mathbf{X}$ روی $\mathbf{Y}$ با روش OPA، برآوردگر β به صورت $\widehat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1})^T = (\mathbf{X}_D^* \mathbf{X}_D)^{-1} \mathbf{X}_D^* \mathbf{Y}$ به دست می‌آید، که در آن $\mathbf{X}_D = [1_L \ \ \mathbf{X}]$ ماتریس طرح مختلط $2 \times L$بعدی و $\mathbf{X}_D^*$ ترانهاده مزدوج مختلط $\mathbf{X}_D$ است. در اکثر موارد پیکربندی واقعی $\mathbf{X}$ مشاهده نمی‌شود و به جای آن پیکربندی W مشاهده می‌شود. به عبارتی دیگر می‌توان نوشت: $\mathbf{W} = \mathbf{X} + \mathbf{U}$. حال تحت مدل (۲) برآوردگر ناپخته‌ی

β برای پیکربندهای مشاهده شده به صورت

$$\widehat{\beta}_{naive} = (\widehat{\beta}_{۰,W}, \widehat{\beta}_{۱,W})^T = (\mathbf{W}_D^* \mathbf{W}_D)^{-۱} \mathbf{W}_D^* \mathbf{Y} \qquad (۳)$$

است، که در آن $\mathbf{W}_D$ ماتریس طرح برای پیکربندی مشاهده شده است. برای مطالعه اثرات خطای اندازه‌گیری در روش OPA، فرض می‌شود نقاط شاخصی از توزیع نرمال مختلط برای $\mathbf{X}$ وجود دارند طوری که برای $\ell = ۱, \cdots, L$، $X_\ell \sim CN(\mu_x, ۲\sigma_x^۲)$. هم‌چنین فرض می‌شود متغیرهای تصادفی مختلط نرمال $\{\mathbf{X}, \mathbf{U}, \mathbf{W}\}$ دو به دو مستقل هستند.

لم ۱: (دو و همکاران، ۲۰۱۴) تحت فرض‌های نرمال مختلط برای X_ℓ، U_ℓ و ϵ_ℓ به ازای $\ell = ۱, \cdots, L$ اگر $\mu_X = (\mu_۱, \cdots, \mu_L)^T = \mu ۱_L$، آنگاه

$$E(\widehat{\beta}_{۰,naive}) = \beta_۰ + (۱ - \lambda)\mu\beta_۱, \quad E(\widehat{\beta}_{۱,naive}) = \lambda\beta_۱$$

که در آن $\lambda = \frac{\sigma_x^۲}{\sigma_x^۲ + \sigma_u^۲}$ نرخ قابلیت اعتماد [۸] نامیده می‌شود.

اگر λ بزرگ (نزدیک یک) باشد خطای اندازه‌گیری کوچک و اریبی برآوردگر ناپخته در قدر مطلق کوچک است. همچنین اگر μ_x به ازای تمام مولفه‌ها برابر صفر باشد، آنگاه $E(\hat{\beta}_{۰,naive}) = \beta_۰$ که نشان دهنده نااریبی $\hat{\beta}_{۰,naive}$ تحت فرض صفر بودن میانگین X است. چون $\lambda \in [۰, ۱]$، تساوی دوم در لم ۱ یک اثر کاهشی از خطای اندازه‌گیری را روی برآوردگر $\beta_۱$ نشان می‌دهد. در انطباق دو پیکربندی، این اثر کاهشی به معنی کم برآورد کردن پارامتر مقیاس است.

لم ۲: (دو و همکاران، ۲۰۱۴) تحت فرض های نرمال مختلط برای X_ℓ، U_ℓ و ϵ_ℓ به ازای $\ell = ۱, \cdots, L$ اگر $\mu_X = (\mu_۱, \cdots, \mu_L)^T \neq \mu ۱_L$، آنگاه

$$\begin{aligned}
E(\widehat{\beta}_{۰,naive}) &= \beta_۰ + \Delta\breve{\mu}_x(L - ۲ - \frac{۱}{L})\beta_۱ + o(\sigma_x^۲ + \sigma_u^۲).K, \\
E(\widehat{\beta}_{۱,naive}) &= \beta_۱\{۱ - \Delta L(۱ - \frac{۴}{L} - (\frac{L-۲}{L^۲})\frac{\sum_{\ell=۱}^{L}((\mu_\ell^{(R)})^۲ + (\mu_\ell^{(I)})^۲)}{\|\mu_x - \breve{\mu}_x ۱_L\|^۲})\} \\
&\quad + o(\sigma_x^۲ + \sigma_u^۲).K
\end{aligned}$$

[8] Reliability rate

که در آن $\breve{\mu}_x = \mathbf{1}_L^T \frac{\mu_x}{L}$، $\Delta = \frac{2\sigma_u^2}{\|\mu_x - \breve{\mu}_x \mathbf{1}_L\|^2}$ و $K = \frac{\beta_1 \sigma_u^2}{\sigma_x^2 + \sigma_u^2}$ و $\mu_\ell^{(R)}$ و $\mu_\ell^{(I)}$ به ترتیب بخش‌های حقیقی و مجازی μ_ℓ هستند. می‌توان ملاحظه کرد که هر چه پراکندگی μ_x زیاد شود، اریبی در برآوردگرهای ناپخته کمتر می‌شود.

۲.۲ روش کالیبدن رگرسیون برای تصحیح اریبی

روش کالیبدن رگرسیون برای هر مدل رگرسیونی به اندازه کافی تقریب دقیقی را فراهم می‌کند (کارول و همکاران، ۲۰۰۶). فرض کنید هدف مطالعه این است که پیکربندی $\mathbf{X}$ از طریق مدل رگرسیونی مختلط

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X} + \epsilon$$

روی پیکربندی $\mathbf{Y}$ منطبق شود، که در آن β_0 و β_1 پارامترهای مدل هستند. مدل رگرسیونی براساس پیکربندی‌های مشاهده شده $(\mathbf{Y}, \mathbf{W})^T$ به مدل رگرسیونی ناپخته معروف است و در این حالت ارتباط بین متغیرها معمولا به صورت

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{W} + \epsilon$$

نمایش داده می‌شود که برآوردهای حاصل از آن به برآوردهای اریب منجر خواهد شد. در این صورت برای تصحیح اریبی می‌توان الگوریتم کالیبدن رگرسیون را به صورت زیر بکار برد:

گام ۱: رگرسیون مختلط پیکربندی $\mathbf{X}$ روی پیکربندی $\mathbf{W}$ انجام گیرد.

گام ۲: پیکربندی $\mathbf{X}$ برآورد شده از مدل رگرسیونی مختلط به جای پیکربندی $\mathbf{X}$ مشاهده نشده جایگذاری شود و تحلیل استاندارد برای دستیابی به برآورد پارامترها صورت پذیرد.

گام ۳: خطای استاندارد در **گام ۱** تصحیح شود.

حال با فرض اینکه واریانس خطای اندازه‌گیری (σ_u^2) معلوم است، تابع کالیبدن یا بهترین تقریب خطی پیکربندی $\mathbf{X}$ به شرط پیکربندی $\mathbf{W}$ به صورت

$$E(\mathbf{X}|\mathbf{W}) \approx \hat{\mu}_w + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}(\mathbf{W} - \hat{\mu}_w)$$

به دست می‌آید، که در آن

$$\hat{\mu}_w = \sum_{\ell=1}^{L} W_\ell / L, \quad \sigma_x^2 = \sigma_w^2 - \sigma_u^2.$$

با توجه به مدل‌بندی مختلط داده‌های شکل دوبعدی روش کالیبدن رگرسیون در آمارشکل نیز قابل اجرا است. نحوه عملکرد این روش برای مدل رگرسیونی مختلط در مطالعه‌ای شبیه‌سازی بررسی می‌شود.

۳.۲ روش امتیاز شرطی

بنا به کارول و همکاران (۲۰۰۶) روش کالیبدن رگرسیون بطور تقریبی سازگار است ولی روش امتیاز شرطی کاملا سازگار، ساده‌تر و نیرومند است. برای تشریح نحوه عملکرد روش امتیاز شرطی برای داده‌های شکل دو بعدی، مدل (۲) را درنظر بگیرید. در این روش، فرضی درباره توزیع X صورت نمی‌گیرد. به پارامترهای β_0، β_1 و σ_ϵ^2 به چشم ثابت معلوم و به X_ℓ به چشم پارامتر نگاه می‌شود که در این صورت $\Delta_\ell = W_\ell + \frac{\overline{\beta_1} Y_\ell \sigma_u^2}{\sigma_\epsilon^2}$ به ازای $\ell = 1, \cdots, L$ یک آماره بسنده برای X_ℓ است که، در آن $\overline{\beta_1}$ مزدوج مختلط β_1 است. حال تحت فرض نرمال مختلط برای ϵ_ℓ و U_ℓ، دو گشتاور اول Y_ℓ به شرط Δ_ℓ به صورت

$$E(Y_\ell|\Delta_\ell, \Theta) = \frac{\beta_0 + \beta_1 \Delta_\ell}{1 + \frac{\|\beta_1\|^2 \sigma_u^2}{\sigma_\epsilon^2}}, \quad Var(Y_\ell|\Delta_\ell, \Theta) = \frac{2\sigma_\epsilon^2}{1 + \frac{\|\beta_1\|^2 \sigma_u^2}{\sigma_\epsilon^2}},$$

هستند. سپس تابع امتیاز مختلط به صورت

$$\psi_{cond}(Y_\ell, \Delta_\ell, \Theta) =$$
$$[Y_\ell - E(Y_\ell|\Delta_\ell, \Theta), \overline{\{Y_\ell - E(Y_\ell|\Delta_\ell, \Theta)\}} \Delta_\ell, (\frac{L-p}{L})\sigma_u^2 - \frac{||Y_\ell - E(Y_\ell|\Delta_\ell, \Theta)||^2}{\frac{Var(Y_\ell|\Delta_\ell, \Theta)}{\sigma_u^2}}]^T$$

تعریف می‌شود که با پیروی از لیندسی (۱۹۸۲) به تابع امتیاز شرطی معروف است. در اینجا، p تعداد پارامترهای Θ به جز σ_ϵ^2 و $\psi(Y_\ell, \Delta_\ell, \Theta)$ بردار امتیاز مختلط نااریب است. در این حالت، برآوردگر Θ از حل $\sum_{\ell=1}^{L} \psi(Y_\ell, \Delta_\ell, \Theta) = 0$ با استفاده از قضیه برآورد$-M$ (هابر، ۱۹۶۷) به دست می‌آید و در منابع علمی آماری به برآوردگر امتیاز شرطی معروف است. توجه شود که واریانس این برآوردگر بر اساس

الگوی ساختار واریانس ساندویچ که برای برآوردگرهای $M-$ توسط استفانسکی و بووس (۲۰۰۲) معرفی شد، قابل محاسبه است.

۳ مطالعه شبیه‌سازی

دراین بخش به کمک شبیه‌سازی مقایسه‌ای بین عملکرد برآوردگرهای کالیبدن رگرسیون، امتیاز شرطی و ناپخته در تصحیح اریبی صورت گیرد. توجه شود نمادهای Naive، CSE و RC که در ادامه مورد استفاده قرار می‌گیرند به ترتیب نماینده روش‌های ناپخته، امتیاز شرطی و کالیبدن رگرسیون هستند. در مطالعات انجام شده توسط دو و همکاران (۲۰۱۴)، برای تولید پیکربندی X، میانگین شکل $(\mu_{\mathbf{x}})$ از توزیع یکنواخت مختلط تولید شد. به عبارت دیگر، برای قسمت‌های حقیقی و مجازی $\mu_{\mathbf{x}}$ به ازای $\ell = ۱, \cdots, L$ توزیع $U(-۵, ۵)$ در نظر گرفته شد. آن‌ها، سپس $Re(\mathbf{X})$ و $Im(\mathbf{X})$ را به طور مستقل از $N(Re(\mu_{\mathbf{x}}), \frac{\sigma_x^۲}{۲})$ و $N(Im(\mu_{\mathbf{x}}), \frac{\sigma_x^۲}{۲})$ شبیه‌سازی کردند. هم‌چنین، آن‌ها $Re(\epsilon_\ell)$ و $Im(\epsilon_\ell)$ را به طور مستقل از $N(۰, \frac{\sigma_\varepsilon^۲}{۲})$ و به طور مشابه $Re(\mathbf{U}_\ell)$ و $Im(\mathbf{U}_\ell)$ را به صورت مستقل از $N(۰, \frac{\sigma_u^۲}{۲})$ و در آخر، پیکربندی‌های $\mathbf{W}$ و $\mathbf{Y}$ را براساس تساوی‌های $\mathbf{W} = \mathbf{X} + \mathbf{U}$ و $\mathbf{Y} = \beta_۰ + \beta_۱\mathbf{X} + \epsilon$ بدست آوردند. مقادیر واقعی پارامتر به صورت $\beta = (۱ + ۲i, ۲ + i)^T$ و $\sigma_\varepsilon^۲ = ۲$ فرض شدند. به علاوه، با تعداد نقاط شاخص $L = ۳۰$ عمل شبیه‌سازی را $M = ۱۰۰۰$ بار تکرار و سپس تحلیل‌های مورد نظر را انجام دادند. نتایج مطالعات دو و همکاران (۲۰۱۴) حاکی از این بود که برآوردهای روش CSE اریبی روش Naive را تا حد زیادی تصحیح می‌کند. اما روش RC برآوردهای ضعیفی حتی بدتر از روش Naive ارائه داده است. از آنجا که عملکرد روش RC در برآورد پارامترهای شکل در روش به کار رفته توسط آن‌ها مناسب به نظر نمی‌رسید، در این مقاله سعی شد رویکرد دیگری برای این منظور دنبال شود. انتظار ما آن است که ضعیف بودن روش RC بخاطر عدم جمع‌پذیری توزیع‌های یکنواخت و نرمال بوده است که در محاسبات خود را نشان داد. لذا، ترجیح دادیم برای تولید پیکربندی $\mathbf{X}$، میانگین شکل $(\mu_{\mathbf{x}})$ را از توزیع نرمال مختلط تولید کنیم. به طور دقیق‌تر، ابتدا قسمت‌های حقیقی و مجازی $\mu_{\mathbf{x}}$ به ازای $\ell = ۱, \cdots, L$ از توزیع نرمال با میانگین صفر و واریانس $\frac{۱۰۰}{۱۲}$ تولید و سپس $Re(\mathbf{X})$ و $Im(\mathbf{X})$ به طور مستقل از $N(Re(\mu_{\mathbf{x}}), \frac{\sigma_x^۲}{۲})$ و $N(Im(\mu_{\mathbf{x}}), \frac{\sigma_x^۲}{۲})$ شبیه‌سازی شدند. لازم به ذکر است که انتخاب واریانس $\frac{۱۰۰}{۱۲}$ برای μ_x به دلیل انجام مقایسه درست‌تر با روش بکار رفته در دو و همکاران (۲۰۱۴) صورت گرفته است.

به طور دقیق‌تر، با انتخاب توزیع یکنواخت برای بخش‌های حقیقی و مجازی μ_x در بازه $(-۵, ۵)$ میانگین این توزیع صفر و واریانس آن ۱۰۰/۱۲ خواهد بود. در نتیجه، انتظار می‌رود برای مقایسه منطقی‌تر واریانس متناظر برای توزیع نرمال مختلط نیز برابر همین مقدار در نظر گرفته شود. از این رو، $Re(\epsilon_\ell)$ و $Im(\epsilon_\ell)$ به طور مستقل از $N(۰, \frac{\sigma_\varepsilon^۲}{۲})$ و به علاوه، $Re(\mathbf{U}_\ell)$ و $Im(\mathbf{U}_\ell)$ به صورت مستقل از $N(۰, \frac{\sigma_u^۲}{۲})$ و در آخر، پیکربندی‌های $\mathbf{W}$ و $\mathbf{Y}$ براساس تساوی‌های $\mathbf{W} = \mathbf{X} + \mathbf{U}$ و $\mathbf{Y} = \beta_۰ + \beta_۱\mathbf{X} + \epsilon$ بدست آمدند. در اینجا نیز مقادیر واقعی پارامتر به صورت $\beta = (۱ + ۲i, ۲ + i)^T$ و $\sigma_\varepsilon^۲ = ۲$ در نظر گرفته شده‌اند. داده‌ها به ازای λهای مختلف با تعداد نقاط شاخص $L = ۳۰$ به ازای $M = ۱۰۰۰$ بار شبیه‌سازی شدند. نتایج حاصل از برآورد پارامترها به تفکیک روش‌ها و پارامترها در جدول ۱ آمده است. همان‌طور که ملاحظه می‌شود برآوردگرهای روش Naive، برآوردگرهایی اریب را ارائه خواهد کرد. از طرف دیگر و براساس انتظار، روش CSE اریبی برآوردگرهای روش Naive را تاحدود زیادی کاهش داده است. در واقع، چون این روش هیچ فرض توزیعی برای پیکربندی صحیح (X) در نظر نمی‌گیرد و به زبانی ساده‌تر، به فرض توزیع پیکربندی حساس نیست، عملکرد آن مشابه آن‌چیزی است که توسط دو و همکاران (۲۰۱۴) گزارش شد. اما، روش RC در اینجا خیلی بهتر از روش آن‌ها و حتی بهتر از روش CSE عمل کرده است. این موضوع گویای واقعیت مهمی به این ترتیب است که روش RC بسته به توزیع در نظر گرفته شده برای پیکربندی صحیح و همچنین میزان پراکندگی مولفه‌های میانگین شکل، می‌تواند عملکردهای متفاوتی داشته باشد. این نکته از آن جهت حائز اهمیت است که روش RC در ذات خود از ویژگی جمع‌پذیری متغیرهای میانگین و خطای پیکربندی شکل بهره می‌برد. لذا، توزیع مورد نظر برای این دو متغیر باید از یک خانواده باشند. چون توزیع نرمال مختلط تحت عمل جمع بسته است، توزیع حاصل برای پیکربندی‌های $\mathbf{X}$ نیز نرمال خواهند بود. در نتیجه روش RC که از ویژگی‌های رگرسیون استفاده می‌کند در روابط خود از کمیت‌هایی بهره می‌برد که بطور صحیحی بدست آمده‌اند. با این حال، مطالعه نظری بیشتر راجع به این موضوع ضروری است.

۴ تحلیل داده‌های تپه‌های ماسه‌ای اردستان

در این بخش مجموعه داده‌های لندفرم[9] یا زمین‌چهره مربوط به تپه‌های ماسه‌ای واقع در منطقه اردستان استان اصفهان مورد تحلیل قرار می‌گیرد. لندفرم در علوم جغرافیایی به شکل یا اشکال طبیعی یا فیزیکی

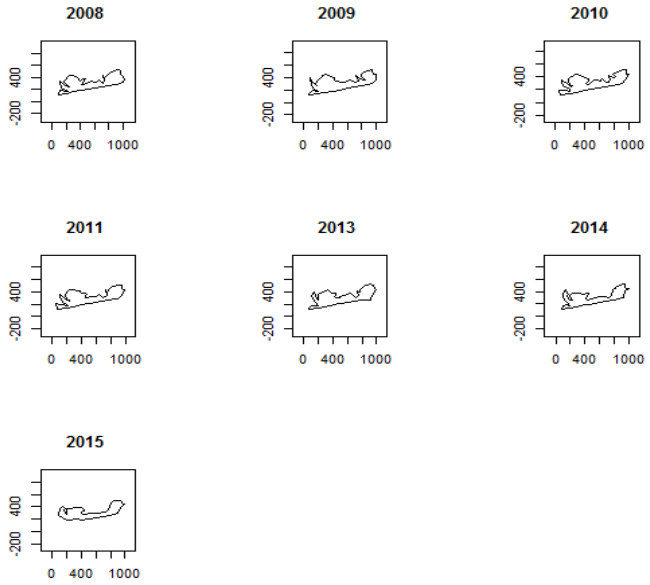
[9] Landform

جدول ۱. میانگین (انحراف معیار) برآورد پرامترها با میانگین‌های شکل نابرابر، $L = ۳۰$ و نقطه شاخص از توزیع نرمال مختلط

پارامتر	مقدار واقعی	روش برآورد	نرخ قابلیت اعتماد ۰٫۵		۰٫۸		۰٫۹۵	
$\beta_{۰۰}$	۱	Naive	۱٫۰۸۲	(۰٫۸۱۱)	۱٫۰۵۱	(۰٫۵۷۲)	۱٫۰۱۳	(۰٫۳۳۸)
		RC	۱٫۰۰۰	(۱٫۳۶۸)	۱٫۰۰۷	(۰٫۶۶۶)	۱٫۰۰۰	(۰٫۳۵۰)
		CSE	۱٫۲۲۲	(۱٫۷۸۵)	۱٫۱۲۳	(۰٫۸۶۳)	۰٫۹۹۱	(۰٫۳۸۵)
$\beta_{۰۱}$	۲	Naive	۲٫۰۶۰	(۰٫۷۷۰)	۲٫۰۳۲	(۰٫۵۶۱)	۱٫۹۹۷	(۰٫۳۰۵)
		RC	۱٫۹۸۶	(۱٫۳۰۳)	۲٫۰۰۴	(۰٫۶۵۶)	۱٫۹۹۲	(۰٫۳۱۵)
		CSE	۲٫۴۹۹	(۱٫۹۷۱)	۲٫۲۰۲	(۰٫۸۴۵)	۲٫۰۱۸	(۰٫۳۲۷)
$\beta_{۱۰}$	۲	Naive	۱٫۰۰۶	(۰٫۱۴۵)	۱٫۶۰۲	(۰٫۱۲۱)	۱٫۹۰۱	(۰٫۰۷۵)
		RC	۲٫۰۱۲	(۰٫۲۹۱)	۲٫۰۰۲	(۰٫۱۵۱)	۲٫۰۰۱	(۰٫۰۷۹)
		CSE	۱٫۷۸۵	(۰٫۳۳۴)	۱٫۸۳۹	(۰٫۱۵۲)	۱٫۸۹۰	(۰٫۰۸۰)
$\beta_{۱۱}$	۱	Naive	۰٫۴۹۸	(۰٫۱۵۱)	۰٫۷۹۷	(۰٫۱۲۱)	۰٫۸۵۲	(۰٫۰۷۰)
		RC	۰٫۸۹۷	(۰٫۳۰۳)	۰٫۸۹۶	(۰٫۱۵۱)	۱٫۰۰۳	(۰٫۰۷۴)
		CSE	۰٫۹۰۷	(۰٫۲۹۴)	۰٫۸۷۳	(۰٫۱۵۴)	۰٫۸۹۷	(۰٫۰۷۶)
$\sigma_{\varepsilon}^{۲}$	۲	Naive	۴۹٫۵۴۶	(۸٫۴۰۷)	۲۰٫۳۸۵	(۳٫۸۷۷)	۶٫۰۸۷	(۱٫۱۹۵)
		RC	۴۹٫۵۴۶	(۸٫۴۰۷)	۲۰٫۳۸۵	(۳٫۸۷۷)	۶٫۰۸۷	(۱٫۱۹۵)
		CSE	۱۰٫۸۲۹	(۱۲٫۰۸۴)	۳٫۸۳۶	(۳٫۸۹۰)	۱٫۸۰۵	(۱٫۲۴۶)

سطح زمین مانند دشت‌ها، دره‌ها، فلات‌ها، کوه‌ها و پدیده‌هایی که در سطح زمین ایجاد شده و عوامل درونی و یا بیرونی در ایجاد آن نقش داشته و دارند، گفته می‌شود (رابرت و همکاران، ۲۰۰۰). محدوده گسترش لندفرم‌های بادی در سال‌های ۲۰۰۸ تا ۲۰۱۵ به جز لندفرم سال ۲۰۱۲ که در دسترس نیست، با تعداد نقاط نابرابر رقومی شده است. به عنوان مثال، محدوده گسترش لندفرم‌های بادی در سال‌های ۲۰۰۸ و ۲۰۰۹ به ترتیب با ۱۰۲۲۶ و ۹۶۰۶ نقطه رقومی شدند. با توجه به اینکه معمولا تحلیل‌های آماری شکل براساس اشکالی با تعداد نقاط شاخص یکسان صورت می‌گیرد، لذا در هر لندفرم ۵۲ نقطه شاخص متناظر انتخاب شد. این نقاط به گونه‌ای انتخاب شدند که بتوان از طریق آن‌ها به طور مناسبی اشکال اولیه لندفرم‌ها را بازسازی کرد. نمایش هندسی این اشکال در سال‌های مختلف در شکل ۱ آمده است.

هدف اصلی تحلیل داده‌های لندفرم، برازش مدل‌های بررسی شده در این مقاله با استفاده از روش‌های OPA ناپخته، کالبیدن رگرسیون و امتیاز شرطی است. از آنجایی که اطلاعات اخذ شده مبتنی بر ۵۲ نقطه شاخص از پیکربندی‌های لندفرم دقیق نیست، فرض آلوده بودن این نقاط شاخص به خطای اندازه‌گیری معقول بنظر می‌رسد. دلیل این امر ناشی از جمع‌آوری اطلاعات از طریق داده‌های ثبت شده توسط ماهواره‌های جغرافیایی و در شرایط آب و هوایی متفاوت است. به عبارتی دیگر، واضح است که اندازه‌گیری دقیق نقاط شاخص برای هرکدام از لندفرم‌ها ممکن نبوده و در عوض باید قبول کرد که محقق جغرافیایی برای ثبت هر یک از لندفرم‌ها دچار خطای اندازه‌گیری می‌شود. پیکربندی لندفرم در دو سال



شکل ۱. نمایش هندسی لندفرم‌های تپه‌های ماسه‌ای اردستان اصفهان در طی سال‌های ۲۰۰۸ تا ۲۰۱۵.

متوالی به عنوان $\mathbf{W}$ و $\mathbf{Y}$ در نظر گرفته و انطباق $\mathbf{W}$ بر روی $\mathbf{Y}$ در سال‌های ۲۰۰۸ تا ۲۰۱۵ با اجرای روش‌های OPA ناپخته، کالبیدن رگرسیون و امتیاز شرطی انجام می‌شود. از آنجایی که در هر سال فقط داده‌های شکل مربوط به یک لندفرم موجود است، برآورد واریانس خطای اندازه‌گیری امکان‌پذیر نیست. لذا، تحلیل‌های مورد نظر به ازای دو سطح نرخ قابلیت اعتماد ($\lambda = ۰/۵, ۰/۹$) انجام شده است. از طرفی دیگر، چون پارامتر انتقال در تحلیل پروکراستس اشکال صفر برآورد می‌شود (به دلیل مرکزی کردن پیکربندی‌ها)، فقط برآوردهای پارامترهای $\beta_۱ = (\beta_{۱۰}, \beta_{۱۱})^T$ و $\sigma^۲_\varepsilon$ مدنظر قرار گرفته‌اند. نتایج مربوط به برآورد این پارامترها براساس انطباق لندفرم سال ۲۰۰۸ بر روی لندفرم سال ۲۰۰۹ در جدول ۲ آمده است. لندفرم سایر سال‌ها نیز به صورت متوالی بر روی هم منطبق شده و نتایج بدست آمده است. اما، به دلیل زیاد بودن حجم مطالب از ارائه همه نتایج خودداری شده و فقط برآوردهای مقیاس حاصل از انطباق لندفرم کلیه سال‌ها با روش‌های Naive، RC و CSE در شکل ۲ رسم شده است.

همانطور که در جدول ۲ ملاحظه می‌شود، روش Naive به ازای دو سطح λ برای هر پارامتر تنها یک مقدار را ارائه کرده است. چون داده‌های لندفرم آلوده به خطای اندازه‌گیری هستند، برآوردهای روش Naive اریب خواهند بود و لذا، نباید به مقادیر حاصل از این روش اعتماد کرد. عملکرد روش‌های تصحیح

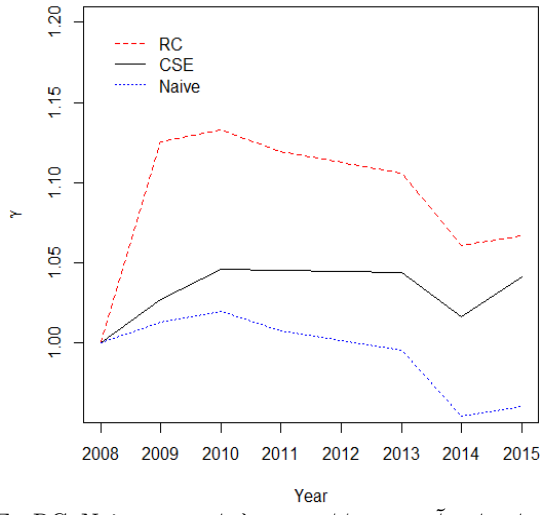
اریبی در اینجا نیز همانند مطالعات شبیه‌سازی است. روش RC به ازای دو سطح λ برای پارامترهای β_0 و β_1 نتایج متفاوتی را ارائه داده است. اما، برآورد پارامتر σ_ε^2 با این روش دقیقا مشابه روش Naive شده است. روش CSE برای پارامترهای β_0 و β_1 علاوه بر اینکه به ازای λهای متفاوت نتایج تقریباً یکسانی را ارائه می‌دهد، خطای برآورد پارامترها را نیز بدست می‌دهد که این امر برای روش‌های دیگر کاهش اریبی ممکن نیست. به عبارتی دیگر، دو روش Naive و RC به دلیل نداشتن تکرار در داده‌های لندفرم هر سال، مقادیری برای برآورد خطای استاندارد ارائه نمی‌کنند. با نگاه به برآوردهای پارامتر σ_ε^2 به نظر می‌رسد که روش CSE نسبت به دو روش دیگر که نتایج یکسانی دارند، ضعیف‌تر عمل کرده است. اتفاق چنین امری می‌تواند یک ضعف برای روش CSE تلقی شود. اما به طور کلی همانگونه که در مطالعات شبیه‌سازی ملاحظه شده است، این روش برای تصحیح اریبی ناشی از خطای اندازه‌گیری در تحلیل شکل، عملکرد مناسبی داشته است. بنابراین، مطالعه عمیق‌تر این داده‌ها و تعدیل برخی روش‌های تصحیح اریبی مطالعه‌شده در این مقاله می‌تواند موضوع تحقیقات آتی باشد.

جدول ۲. برآورد پارامترها حاصل از انطباق داده‌های لندفرم

پارامتر	روش برآورد	نرخ قابلیت اعتماد ۰٫۵	نرخ قابلیت اعتماد ۰٫۹
β_{10}	Naive	−۰٫۲۱۱ (−)	−۰٫۲۱۱ (−)
	RC	−۰٫۴۲۱ (−)	−۰٫۲۳۴ (−)
	CSE	−۰٫۲۱۴ (۰٫۰۶۵)	−۰٫۲۱۴ (۰٫۰۱۸)
β_{11}	Naive	۰٫۹۹۱ (−)	۰٫۹۹۱ (−)
	RC	۱٫۹۸۱ (−)	۱٫۱۰۱ (−)
	CSE	−۱٫۰۰۶ (۰٫۰۱۱)	−۱٫۰۰۵ (۰٫۰۱۷)
σ_ε^2	Naive	۷۵۱٫۹۲۵ (−)	۷۵۱٫۹۲۵ (−)
	RC	۷۵۱٫۹۲۵ (−)	۷۵۱٫۹۲۵ (−)
	CSE	۹۴۶٫۷۸۵ (۲۰۰۴۸۲)	۹۴۴٫۲۱۱ (۹۰۷۱)

در ادامه برای ارائه تصویر روشنی از اثرات خطای اندازه‌گیری، برآوردهای پارامتر مقیاس ($\gamma = \|\widehat{\beta}_1\|$) حاصل از انطباق متوالی لندفرم سال‌های ۲۰۰۸ تا ۲۰۱۵ با سه روش مقایسه شده‌اند. شکل ۲ برآوردهای پارامتر مقیاس با سه روش به ازای $\lambda = ۰٫۹$ را نشان می‌دهد. همانگونه که در این شکل مشاهده می‌شود برآوردهای مقیاس از روش Naive در زیر دو نمودار دیگر قرار گرفته است. چون داده‌ها آلوده به خطا هستند، کم برآوردی روش Naive برای پارامتر مقیاس مشهود است. نمودار برآوردهای مقیاس با روش CSE در بین دو نمودار قرار دارد. با توجه به اینکه مقدار λ برابر ۰٫۹ فرض شده است، طبیعی است

که برآوردهای پارامترهای مقیاس حاصل از روش CSE نزدیک به برآوردهای پارامترهای مقیاس روش Naive باشد که نشانگر این حقیقت است که روش CSE برآوردهای نزدیک به واقعیتی را ارائه کرده است. با توجه به نتایج حاصل از شبیه‌سازی در حالت میانگین‌های شکل نابرابر، می‌توان گفت برآوردهای این روش برآوردهای نامناسبی برای پارامترهای مقیاس ارائه می‌کنند.



شکل ۲. مقیاس‌های برآورد شده به ازای $\lambda = ۰/۹$ با سه روش Naive، RC و CSE .

بحث و نتیجه‌گیری

در این مقاله، مدل خطای اندازه‌گیری برای انطباق پروکراستس بر اساس مدل‌بندی مختلط داده‌های شکل تعمیم داده شده است. اثرات نامطلوب برآوردگرهای ناپخته پارامترهای انتقال، مقیاس و دوران برای داده‌های شکل دو بعدی توصیف شدند. مشاهده شد که در صورت وجود خطای اندازه‌گیری برآوردگرهای ناپخته برای پارامترهای انتقال، مقیاس و دوران اریب هستند. بزرگی و شدت این اریبی به چندین پارامتر کنترل از جمله نرخ قابلیت اعتماد، پراکندگی μ_x و تعداد نقاط شاخص بستگی دارد که در این بین نرخ قابلیت اعتماد، تأثیرگذارترین پارامتر است. به زبانی ساده، به هر اندازه نرخ قابلیت اعتماد کوچکتر باشد، اریبی برآوردگرها بزرگتر خواهد بود. برای تصحیح اریبی برآوردگرهای ناپخته دو روش کالیبدن رگرسیون و

امتیاز شرطی مورد مطالعه قرار گرفتند. این روش‌ها برای مدل‌های رگرسیونی خطی با متغیرهای حقیقی مقدار بسیار کارا و مناسب هستند. در این مقاله، همین روش‌ها برای مسئله انطباق پروکراستس داده‌های شکل دو بعدی یا به طور معادل برای مدل رگرسیونی مختلط مورد استفاده قرار گرفتند. در مطالعه شبیه‌سازی نشان داده شد که روش کالیبدن رگرسیون بسته به اینکه پیکربندی صحیح دارای چه توزیعی و پراکندگی مولفه‌های میانگین به چه صورت باشد، می‌تواند عملکردهای متفاوتی داشته باشد. اما روش امتیاز شرطی به فرض توزیعی حساس نیست و به طور کلی عملکرد بهتری نسبت به روش کالیبدن رگرسیون در برآورد پارامترهای مدل رگرسیونی مختلط در حضور خطای اندازه‌گیری دارد.

برای تحقیقات آتی در این حوزه از آمارشکل می‌توان از روش‌های دیگری که برای تصحیح اریبی در مدل‌های رگرسیونی مختلط وجود دارد، استفاده کرد. مطالعه مدل خطای اندازه‌گیری با داده‌های گمشده و بررسی خطای اندازه‌گیری برای داده‌های طولی شکل از دیگر تحقیقاتی است که می‌تواند در آینده مدنظر قرار گیرند.

تقدیر و تشکر

نویسندگان از داوران و ویراستار محترم مجله به خاطر صرف وقت و پیشنهادهای ارزنده که باعث بهبودی مقاله شد کمال تشکر و قدردانی را دارند.

مراجع

Armstrong, B. (1985), Measurement Error in Generalized Linear Models, *Communications in Statistics*, **14**, 529–544.

Carroll, R. J., and Stefanski, L. A. (1990), Approximate Quasilikelihood Estimation in Models with Surrogate Predictors, *Journal of the American Statistical Association*, **85**, 652–663.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models*, Chapman and Hall, Boca Raton.

Clayton, D. G. (1992), Models for the Analysis of Cohort and Case-Control Studies with Inaccurately Measured Exposures, In J. H. Dwyer, M. Feinleib, P. Lipsert, P., et al. (Eds.), *Statistical Models for Longitudinal Studies of Health*, 301-331, New York

Dryden, I. L. and Mardia, K. V. (1998), *Statistical Shape Analysis*, John Wiley and Sons, Chichester.

Du, J., Dryden, I. L., and Huang, X. (2014), Size and Shape Analysis of Error-Prone Shape Data, *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2014.908779.

Fuller, W. A. (1980), Properties of Some Estimators for the Errors-in-Variables Model, *The Annals of Statistics*, **28**, 407-422.

Gleser, L. J. (1981), Estimation in a Multivariate "Errors in Variables" Regression Model: Large Sample Results, *The Annals of Statistics*, **9**, 24–44.

Gleser, L. J. (1990), Improvements of the Naive Approach to Estimation in Nonlinear Errors-in-Variables Regression Models, In P. J. Brown and W. A. Fuller (Eds.) *Statistical Analysis of Measurement Error Models and Application*, American Mathematics Society, Providence.

Goodman, N. R. (1963), Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction), *The Annals of Mathematical Statistics*, **34**, 152–177.

Huber, P. J. (1967), The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions, *Proceedings of the 5th Berkeley Symposium*, **1**, 221–233.

Lindsay, B. (1982), Conditional Score Functions: Some Optimality Results, *Biometrika*, **69**, 503-512.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*. Academic Press, London.

Robert A. McMillan, David H. McNabb and R. Keith Jones (2000). Automated Landform Classification Using DEMs: A Conceptual Framework for a Multi-Level, Hierarchy of Hydrologically and Geomorphologically Oriented Physiographic Mapping Units, In *Proceedings of the 4th International Conference on Integrating GIS and Environmental, Modeling: Problems, Prospects and Research Needs. Banff, Alberta, Canada.*

Stefanski, L. A. (1985), The Effects of Measurement Error on Parameter Estimation, *Biometrika*, **72**, 583–592.

Stefanski, L. A. and Carroll, R. J. (1987), Conditional Scores and Optimal Scores for Generalized Linear Measurement-Error Models, *Biometrika*, **74**, 703–716.

Stefanski, L. A., and Boos, D. D. (2002), The Calculus of M-Estimation, *The American Statistician*, **56**, 29-38.