

تجزیه و تحلیل داده‌های مقادیر فرین دمای شهر تبریز

علی اکبر حیدری^{۱*}^۱ استادیار، گروه آمار، دانشگاه تبریز

(دریافت: ۱۴۰۰/۰۳/۱۶، بازنگری: ۱۴۰۰/۰۷/۰۳، پذیرش: ۱۴۰۰/۰۸/۱۶، انتشار آنلاین: ۱۴۰۰/۰۸/۱۶)

چکیده

مدل‌بندی داده‌های مقادیر فرین، یعنی بیشترین و کمترین مقادیر پدیده‌های طبیعی به وسیله فرمول‌های آماری، پیش‌بینی رفتار اینگونه داده‌ها در آینده را امکان‌پذیر می‌سازد. یکی از روش‌های تحلیل داده‌های مقادیر فرین، روش فزونی‌ها است و در آن تمامی فزونی‌های بالای یک سرحد بزرگ، با استفاده از توزیع پارتوی تعمیم یافته مدل‌بندی می‌شوند. در این مقاله، با استفاده از این روش، به هر کدام از داده‌های مقادیر زیاد و مقادیر کم دمای ایستگاه هواشناسی تبریز در بازه زمانی ۱۳۳۰ تا ۱۳۹۹ مدل آماری مناسب برازش داده شده است. سپس پارامترهای مدل‌ها، برآورد شده و نیکویی برازش آنها مورد ارزیابی قرار گرفته است. در پایان برخی پیش‌بینی‌های لازم مانند میانگین دوره‌های بازگشت مقادیر کمینه و بیشینه دمای شهر تبریز برای مدت زمانهای مختلف، انجام شده است. از جمله نتایج این تحقیق این است که با اطمینان ۹۵ درصد، یک کران بالا برای بیشینه دمای هوای تبریز برابر با ۲۲/۳۷- درجه سانتیگراد با میانگین دوره بازگشت ۲۰ ساله و یک کران پایین برای کمینه دمای هوای تبریز برابر با ۲۲/۳۷- درجه سانتیگراد با میانگین دوره بازگشت ۲۰ ساله است.

کلمات کلیدی: دمای بیشینه، دمای کمینه، توزیع پارتوی تعمیم یافته، پیش‌بینی، دوره بازگشت

۱ مقدمه

می‌شود. بنابراین ۵۰ مقدار از مجموعه داده‌ها در نظر گرفته می‌شود و با استفاده از قضایای حدی، یک توزیع پارامتری با عنوان توزیع مقادیر فرین تعمیم یافته (General Extreme Distribution) به داده‌ها برازش داده می‌شود. مثال‌هایی از این روش در مقالات اسمیت (۱۹۸۹)، روتزن و تجویدی (۱۹۹۷)، ونوری و همکاران (۲۰۰۸)، رجبی و مدرس (۲۰۰۸)، کولی و ساین (۲۰۱۰)، دیویسن و غلامرضایی (۲۰۱۲)، چنگک و همکاران (۲۰۱۴)، آسلجسونگ و همکاران (۲۰۱۷)، اشلوگل و لاه (۲۰۱۷) و انصاری اصفه و همکاران (۲۰۲۰) ملاحظه می‌شود. وقتی مقادیر بزرگ متعددی در یک دوره وجود داشته باشد، استفاده از ماکزیمای هر دوره ممکن است موجب از دست دادن اطلاعات زیادی گردد. در چنین مواردی از روش دیگری استفاده می‌شود. این روش، به روش «فزونی‌ها»

چگونگی رفتار داده‌های مقادیر فرین پدیده‌های طبیعی از اهمیت بسیار زیادی برخوردار است. یکی از این پدیده‌های مهم دما است. مقادیر زیاد دما به همراه مقادیر کم بارندگی باعث خشکسالی و مقادیر کم دما باعث بروز یخبندان می‌شوند. بنابراین اگر بتوانیم رفتار اینگونه داده‌ها را به وسیله فرمول‌های آماری بیان کنیم، می‌توانیم چگونگی رفتار آنها در آینده را پیش‌بینی نماییم. معمولاً دو روش برای مدل‌بندی مقادیر فرین، به کار برده می‌شود. یکی از این روش‌ها که به روش «ماکزیمای» معروف است. بر اساس توزیع توام مجانبی آماره ترتیبی بالایی است و هنگامی به داده‌ها برازش داده می‌شود که داده‌ها مشتمل بر مجموعه‌ای از ماکزیمایها باشند. مثلاً اگر بیشینه دمای روزانه یک نقطه، در طول ۵۰ سال در دسترس باشند، در این روش، بیشینه دمای هر سال ماکزیمای آن سال نامیده

مشاهده شده باشند.

در این صورت $z_i = \text{Max}\{x_{i,1}, \dots, x_{i,m}\}$ ماکزیمای دوره i ام نامیده می‌شود.

تعریف ۴- اگر X_1, X_2, X_3, \dots دنباله‌ای از متغیرهای تصادفی با تابع توزیع مشترک F باشند، اولین زمان فزونی از u به صورت $\tau_1 = \min\{m: X_m > u\}$ تعریف می‌شود. به همین ترتیب، r امین زمان فزونی با

$$\tau_r = \min\{m > \tau_{r-1}: X_m > u\}$$

تعریف می‌شود. گاهی برای تاکید بر وابستگی به سرحد u ، به جای τ_r از $\tau_{r,u}$ استفاده می‌شود. چون

$$\begin{aligned} P\{\tau_1 = k\} \\ &= p\{X_1 \leq u, \dots, X_{k-1} \leq u, X_k > u\} \\ &= p(1-p)^{k-1}, \quad k = 1, 2, 3, \dots \end{aligned}$$

بنابراین زمان اولین فزونی از u به صورت هندسی با پارامتر $p = 1 - F(u)$ توزیع شده است و میانگین آن به صورت $E(\tau_{1,u}) = \frac{1}{p}$ است.

برای متغیر تصادفی مستقل و هم توزیع $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \dots$ که دوره‌های بازگشت (Return periods) نامیده می‌شوند، مشاهده می‌شود که برای اعداد صحیح و مثبت $k(1), \dots, k(r)$

$$\begin{aligned} P\{\tau_1 = k(1), \tau_2 - \tau_1 = k(2), \dots, \tau_r - \tau_{r-1} \\ &= k(r)\} \\ &= P\{\tau_1 = k(1), \tau_2 = k(1) + k(2), \dots, \tau_r \\ &= k(1) + \dots + k(r)\} \\ &= P\{X_1 \leq u, \dots, X_{k(1)-1} \leq u, X_{k(1)} > u, X_{k(1)+1} \\ &\leq u, \dots, X_{k(1)+\dots+k(r)} > u\} \\ &= \prod_{j \leq r} P\{\tau_1 = k(j)\} \end{aligned}$$

و همچنین $E(\tau_1) = E(\tau_{r+1} - \tau_r)$ یعنی میانگین دوره‌های بازگشت باهم برابرند. به همین جهت در حالت کلی مقدار $E(\tau_{1,u}) = \frac{1}{p}$ میانگین دوره‌های بازگشت نامیده می‌شود.

تعریف ۵: سرحدی مانند u را که در آن میانگین اولین زمان فزونی از u برابر T است با $u(T)$ نمایش داده و سرحد T -سال (T-Year Threshold) نامیده می‌شود.

بنابراین سرحد T -سال، جواب معادله است،

$$E(\tau_{1,u}) = \frac{1}{1 - F(u)} = T$$

پس $F(u(T)) = 1 - \frac{1}{T}$ سرحد T -سال،

معروف است و در آن تمامی داده‌های منتخب بیش از یک سرحد (peak over threshold) که معمولاً از پیش تعیین شده است از داده‌ها استخراج شده و به آن‌ها یک مدل پارامتری با عنوان توزیع پارتوی تعمیم یافته (Generalized Pareto distribution) توزیع پارتوی تعمیم یافته (Generalized Pareto distribution) داده می‌شود. به این ترتیب به طور مثال، در مثال بیشینه دمای روزانه، ممکن است در یک سال، چند داده وجود داشته باشد که از سرحد داده شده بزرگتر باشند و همگی آنها انتخاب شوند و در بعضی سال‌ها ممکن است اصلاً چنین داده‌ای وجود نداشته باشد و از آن سال هیچ داده‌ای انتخاب نشود.

مثال‌هایی از کاربرد این روش به این صورت است که هاسکینگ و والیس (۱۹۸۷) آن را برای مدل‌بندی سیل‌های سالیانه رودخانه نید انگلستان، گرمشاو (۱۹۹۳) برای مدل‌بندی داده‌های مقاومت کششی یک نمونه تصادفی از رشته‌های قالی، روتزن و تجویدی (۱۹۹۷) برای مطالعه کولاک‌های شدید در سوئد و لئونارد و همکاران (۲۰۰۸) برای مدل‌بندی بارش و طوفان در آمریکا به کار بردند. مثال‌های دیگری از این روش در مقالات لانگوسیس و همکاران (۲۰۱۶)، کانگ و سانگ (۲۰۱۷)، کومار و همکاران (۲۰۲۰)، مارتینز و همکاران (۲۰۲۰)، شارپ و خوارز (۲۰۲۱)، یسی و همکاران (۲۰۲۱) و نگاتسوکا و بالاکریشان (۲۰۲۱) ملاحظه می‌شود. تاکنون در ایران از این روش برای مدل‌بندی داده‌های هواشناسی استفاده نشده است. برای ارائه فرمولها و مدل‌بندی داده‌ها، ابتدا تعاریف زیر را که برگرفته از کتاب لیدبتر و همکاران (۱۹۸۳) است، بیان می‌کنیم:

تعریف ۱- برای مشاهدات x_1, \dots, x_n و y سرحد (Threshold) از پیش تعیین شده u ، هر x_i ای که از u بیشتر باشد، یک فزونی (Exceedance) نامیده می‌شود.

تعریف ۲- برای هر فزونی y_i ، اندازه $y_i - u$ مقدار فزونی (Excesses) بالای u نامیده می‌شود. تعریف ۳- فرض کنید سری داده‌ها در n دوره متوالی با اندازه m به صورت

$$\dots, x_{i,1}, \dots, x_{i,m}, x_{n,1}, \dots, x_{n,m}, x_{1,1}, \dots, x_{1,m}$$

نوشت، که در آن‌ها μ و σ به ترتیب پارامترهای مکانی و مقیاس هستند و γ پارامتر شکل نامیده می‌شود. با توجه به این

که $\lim_{\gamma \rightarrow 0} (1 + \gamma x)^{\frac{1}{\gamma}} = \exp(x)$ ملاحظه می‌شود

$\lim_{\gamma \rightarrow 0} G_{\gamma}(x) = G_0(x)$. به عنوان حالت‌های خاص:

۱- G_0 تابع توزیع گامبل است.

۲- اگر $\gamma > 0$ آنگاه G_{γ} تابع توزیع فرشه (Fréchet) است.

۳- اگر $\gamma < 0$ آنگاه G_{γ} تابع توزیع وایبل (Weibull) است.

۲-۱- توزیع پارتوی تعمیم یافته

این روش مدل‌بندی که در مقاله حاضر مورد مطالعه قرار خواهد گرفت، اولین بار توسط پیکندز (۱۹۷۵) معرفی شد. نمونه‌های دیگری از این مدل‌بندی را می‌توان در مقالات دیویسن (۱۹۸۲) و دیویسن و اسمیت (۱۹۹۰) مشاهده کرد. پیکندز (۱۹۷۵) ثابت کرد که داده‌های بیشتر از یک سرحد از پیش تعیین شده، دارای توزیع پارتوی تعمیم یافته (GPD)، با فرم کلی توابع توزیع زیر است:

$$W_{k,\sigma}(x) = \begin{cases} 1 - \exp\left(-\frac{x}{\sigma}\right) & k = 0, \quad \sigma > 0 \\ 1 - \left(1 - \frac{kx}{\sigma}\right)^{\frac{1}{k}} & k \neq 0, \quad \sigma > 0 \end{cases}$$

که در آن σ پارامتر مقیاس و k پارامتر شکل است. توجه شود که برای $k \leq 0$ ، $x > 0$ و برای $k > 0$ ، x محدوده x به صورت

$$\lim_{k \rightarrow 0} W_{k,\sigma}(x) = W_{0,\sigma}(x) \text{ در اینجا } 0 < x < \frac{\sigma}{k}$$

تابع چگالی متناظر با آن‌ها به صورت زیر است:

$$w_{k,\sigma}(x) = \begin{cases} \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) & k = 0, \sigma > 0 \\ \frac{1}{\sigma} \left(1 - \frac{kx}{\sigma}\right)^{\frac{1}{k}-1} & k \neq 0, \sigma > 0 \end{cases}$$

برای مقادیر به‌خصوصی از k تابع توزیع GP به برخی از توزیع‌های شناخته شده و معروف تبدیل می‌شود.

- هنگامی که $k = 0$ است، تابع توزیع پارتوی تعمیم یافته به تابع توزیع نمایی $EXP(\sigma)$ تبدیل می‌شود.

با استفاده از رابطه

$$u(T) = F^{-1}\left(1 - \frac{1}{T}\right)$$

بدست می‌آید که برابر با چندک $\left(1 - \frac{1}{T}\right)$ ام تابع توزیع F است.

با استفاده از تعریف $u(T)$ مشاهده می‌شود که

$$E(\tau_{1,u(T)}) = T \text{ و } \text{Var}(\tau_{1,u(T)}) = T(1-T)$$

. بنابراین سرحد T - سال نشان‌دهنده آن است که میانگین تعداد

فزونی‌های بالای $u(T)$ در زمان سپری شده به اندازه T برابر یک

است. یکی از اهداف اصلی تحلیل مقادیر فرین، برآورد سرحد

T - ساله است.

۱-۱- توزیع مقادیر فرین تعمیم یافته

روش کلاسیک تحلیل مقادیر فرین، مبتنی بر توزیع‌های حدی ماکزیمای نمونه‌ای است که توسط فیشر و تیپت (۱۹۲۸) ارائه شده است. فرض کنید X_1, X_2, \dots, X_n دنباله‌ای از متغیرهای تصادفی دو به دو مستقل با تابع توزیع مشترک $F(x)$ باشند و برای $n \in \mathbb{N}$ ، $M_n = \text{Max}(X_1, X_2, \dots, X_n)$ ، فرض کنید دنباله‌هایی از اعداد حقیقی مانند a_n و b_n وجود دارند به گونه‌ای که

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) \lim_{n \rightarrow \infty} F^n(a_n x + b_n) \quad (1-1) \\ = G(x)$$

که در آن $G(x)$ یک توزیع ناتباهمیده (Nondegenerate) است

(فیشر و تیپت ۱۹۲۸)، در این صورت گوئیم $F(x)$ در دامنه

جذب $G(x)$ قرار دارد و با نماد $F \in D(G)$ نمایش داده می‌شود.

یک نتیجه اصلی از قضیه مقادیر فرین کلاسیک بیانگر آن است

که $G(x)$ بایستی یک تابع توزیع مقادیر فرین تعمیم یافته باشد.

تمام توزیع‌های حدی ممکن در رابطه (۱.۱)، یعنی توزیع‌های

مقادیر فرین را می‌توان به صورت خانواده پارامتری

$$G_{\gamma,\mu,\sigma} = \exp\left(-\left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{1}{\gamma}}\right), \gamma \neq 0$$

و

$$G_{0,\mu,\sigma} = \exp\left(-e^{-\frac{(x-\mu)}{\sigma}}\right)$$

۱-۲- برآوردگرهای ماکزیمم درست‌نمایی

شود. در این صورت، تعیین برآوردگرهای ماکزیمم درست‌نمایی GPD یک مسئله بهینه‌سازی بر روی فضای محدود شده

$$A = \{k < 0, \sigma > 0\} \cup \left\{0 < k \leq 1, \frac{\sigma}{k} > x_{(n)}\right\}$$

است. پس برای تعیین برآوردگر ماکزیمم درست‌نمایی، بایستی دو مقدار برای (k, σ) مورد بررسی قرار گیرد. اولین مورد

$$L(k, \sigma; x) = \begin{cases} -n \log \sigma + \left(\frac{1}{k} - 1\right) \sum_{i=1}^n \log \left(1 - \frac{kx_i}{\sigma}\right), & k \neq 0 \\ n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^n x_i, & k = 0 \end{cases}$$

ماکزیمم محلی لگاریتم درست‌نمایی در فضای A است و دومین مورد کرانه‌های A برای حالت $k = 1$ است.

ماکزیمم محلی روی A : بردار مشتق لگاریتم درست‌نمایی را در نظر بگیرید. جواب معادلات

$$\begin{cases} \frac{\partial L(k, \sigma; x)}{\partial k} = 0 \\ \frac{\partial L(k, \sigma; x)}{\partial \sigma} = 0 \end{cases}$$

را می‌توان به صورت

$$\begin{cases} n(\hat{k} - 1) = \sum_{i=1}^n \ln \left(1 - \frac{\hat{k}x_i}{\hat{\sigma}}\right) + (\hat{k} - 1) \sum_{i=1}^n \left(1 - \frac{\hat{k}x_i}{\hat{\sigma}}\right)^{-1} \\ n = (\hat{k} - 1) \sum_{i=1}^n \left(1 - \frac{\hat{k}x_i}{\hat{\sigma}}\right)^{-1} \end{cases}$$

یا

$$\begin{cases} \left(1 + \frac{1}{n} \sum_{i=1}^n \ln \left(1 - \frac{\hat{k}x_i}{\hat{\sigma}}\right)\right) \left(\frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\hat{k}x_i}{\hat{\sigma}}\right)^{-1}\right) = 1 \\ \hat{k} = -\frac{1}{n} \sum_{i=1}^n \ln \left(1 - \frac{\hat{k}x_i}{\hat{\sigma}}\right) \end{cases}$$

نوشت. برآوردهای ماکزیمم درست‌نمایی بایستی به صورت عددی محاسبه شوند؛ زیرا آماره‌های بسنده مینیمال، آماره‌های ترتیبی هستند و معادلات درست‌نمایی غیرخطی به هیچ وجه ساده تر نمی‌شوند. هاسکینگ و والیس (۱۹۸۷) برای پیدا کردن برآورد ماکزیمم درست‌نمایی حالت تعدیل شده‌ای از الگوریتم نیوتن رافسون را بکار بستند. گریمشاوا (۱۹۹۳) در این زمینه الگوریتم کاملی را ارائه داده است.

– هنگامی که $k = 1$ است، تابع توزیع پارتوی تعمیم یافته به تابع توزیع یکنواخت $UNIF(0, \sigma)$ تبدیل می‌شود.

– هنگامی که $k < 0$ است، تابع توزیع پارتوی تعمیم یافته به تابع توزیع پارتو $Pareto(k, \sigma)$ تبدیل می‌شود.

۲- برآورد پارامترهای توزیع پارتوی تعمیم یافته

در این بخش، در پی برآورد پارامترهای مقیاس و شکل توزیع پارتوی تعمیم یافته هستیم. متأسفانه برآورد این پارامترها در حالت کلی مساله ساده ای نیست. یکی از مشهورترین روشهای برآورد، همان برآورد بیشینه درست‌نمایی (Maximum Likelihood Estimation-MLE) است.

این برآوردگرها علی‌رغم اینکه برای $k < 1/2$ بطور مجانبی نرمال و کارا هستند، اما برای $1/2 < k < 1$ ، برای نرمال و کارایی مجانبی بودن مشکل دارند و برای $k > 1$ اصلاً فاقد این خصوصیات هستند. علاوه بر این، برای $k \leq 1/2$ ، گشتاورهای مرتبه دوم و بالاتر، موجود نیستند، لذا برآوردهای روش گشتاوری (Method of Moment-MOM) و برآوردهای گشتاورهای وزنی – احتمالی (Probability Weighted Moment-PWM) موجود نیستند. ولی این روش‌ها در برخی فواصل، برآوردهای بهتری از بیشینه درست‌نمایی ارائه می‌کنند. روش دیگری که به ازای همه مقادیر پارامتری قابل محاسبه است، روش صدک‌های اصلی (Elemental Percentile Method-EPM) نامید می‌شود که توسط کاستیو و هادی (۱۹۹۷) معرفی شده است. متأسفانه هیچکدام از این روش‌ها به طور یکنواخت برای تمام مقادیر پارامترها بهترین روش نیستند. لذا پژوهشگر باید با توجه به شرایط مسأله، بهترین روش را انتخاب کند. در این بخش ضمن معرفی این برآوردگرها، نقاط ضعف و قوت آن‌ها نیز برای مقادیر مختلف پارامترها مورد بررسی قرار می‌گیرد.

۲-۲- برآوردگرهای گشتاوری

اگر $k > -\frac{1}{r}$ باشد، r امین گشتاور تابع توزیع پارتوی تعمیم یافته با پارامترهای σ و k موجود است. لذا اگر متغیر تصادفی X دارای توزیع پارتوی تعمیم یافته با $1 + rk > 0$ باشد، آنگاه

$$E(X) = \frac{\sigma}{(1+k)} \quad \text{و} \quad \text{Var}(X) = \frac{\sigma^2}{(1+k)^2(1+2k)}$$

و برآوردگرهای گشتاوری σ و k به صورت

$$\hat{\sigma} = \frac{1}{2} \bar{x} \left(\frac{\bar{x}^2}{S^2 + 1} \right), \quad \hat{k} = \frac{1}{2} \left(\frac{\bar{x}^2}{S^2 - 1} \right)$$

خواهند بود که در آن \bar{x} و S^2 میانگین و واریانس نمونه هستند.

۲-۳- برآوردگرهای گشتاورهای وزنی - احتمالی

گشتاورهای وزنی - احتمالی متغیر تصادفی X با تابع توزیع F که توسط هاسکینگ و همکاران (۱۹۸۵) ارائه شده‌اند، کمیت‌های

$$M_{p,r,s} = E(X^p (F(X))^r (1 - F(X))^s)$$

هستند، که در آن s و p و r اعدادی حقیقی هستند. برای

توزیع GP استفاده از گشتاورهای

$$\alpha_s = M_{1,0,s} = \frac{E(X(1 - F(X))^s)}{\sigma} = \frac{1}{(s+1)(s+1+k)}$$

که برای $k > -1$ وجود دارند، مناسب هستند و پارامترهای آن توسط هاسکینگ و والیس (۱۹۸۷)، به صورت

$$\sigma = \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1} - 2 \quad (۲-۱)$$

$$k = \frac{\alpha_0}{\alpha_0 - 2\alpha_1} - 2$$

محاسبه شده‌اند. برآوردگرهای گشتاورهای وزنی - احتمالی $\hat{\sigma}$ و \hat{k} با جایگزین کردن برآوردگرهای حاصل از یک نمونه مشاهده شده به اندازه n ، به جای α_0 و α_1 در رابطه (۲-۱) به

دست می‌آیند. دو برآورد برای α_r ، $r = 0, 1$ عبارتند از

$$\alpha_r = n^{-1} \sum_{j=1}^n \frac{(n-j)(n-j-1)\dots(n-j-r+1)}{(n-1)(n-2)\dots(n-r)} x_{j:n}$$

و

$$\tilde{\alpha}_r = n^{-1} \sum_{j=1}^n (1 - p_{j:n})^r x_{j:n}$$

که در آن $x_{1:n} \leq \dots \leq x_{n:n}$ نمونه‌های مرتب شده هستند و

$p_{j:n} = \frac{(j-\gamma)}{(n+\delta)}$ است به طوری که در آن γ و δ ثابت‌های

مناسبی هستند. چندک‌های توزیع GP برحسب پارامترها

عبارتند از

$$x(F) = \frac{\sigma[1 - (1 - F)^k]}{k} \quad k \neq 0$$

$$= -\sigma \log(1 - F) \quad k = 0 \quad (۲-۲)$$

برآورد چندک $\hat{x}(F)$ به وسیله جایگزینی برآوردگرهای $\hat{\sigma}$ و \hat{k} در رابطه (۲-۲) تعریف می‌شود.

۲-۴- روش صدک‌های اصلی

روش صدک‌های اصلی، رویه‌ای دو مرحله‌ای است که توسط کاستیو و هادی (۱۹۹۷) برای GPD مورد استفاده قرار گرفته است. در این روش در مرحله اول، برآوردگرهایی مقدماتی برای پارامترها محاسبه می‌شوند و با ترکیب آن‌ها در مرحله دوم برآوردگرهایی برای همه پارامترها به دست می‌آید. این روش برای همه مقادیر پارامترها کار می‌کند و هنگامی برآوردگرهای دیگر وجود نداشته باشند یا با معیارهای مشاهده شده ناسازگار باشند، تنها روش انتخابی مناسب است.

۲-۴-۱- برآوردهای اولیه

فرض کنید $x_{i:n}$ و $x_{j:n}$ دو آماره ترتیبی مجزا در یک نمونه تصادفی به اندازه n از $F(x; k, \sigma)$ باشند. در این صورت، با تعریف

$$p_{i:n} = \frac{i}{n+1}$$

و $C_i = \ln(1 - p_{i:n}) < 0$ در نظر گرفتن معادله

$$C_i \ln\left(1 - \frac{x_{j:n}}{\delta}\right) = C_j \ln\left(1 - \frac{x_{i:n}}{\delta}\right) \quad (۲-۳)$$

برآوردگر

$$\hat{k}(i, j) = \frac{\ln\left(1 - \frac{x_{i:n}}{\delta(i, j)}\right)}{C_i} \quad (۲-۴)$$

$$\hat{k}_{EPM} = \text{median}\{\hat{k}(1,2), \hat{k}(1,3), \dots, \hat{k}(n-1, n)\}$$

$$\hat{\sigma}_{EPM} = \text{median}\{\hat{\sigma}(1,2), \hat{\sigma}(1,3), \dots, \hat{\sigma}(n-1, n)\}$$

که در آنها $\text{median}\{y_1, y_2, \dots, y_n\}$ میانه $\{y_1, y_2, \dots, y_n\}$ می‌باشند.

۳ تحلیل داده‌های فرین دمای تبریز

در این بخش، با توجه به مطالب ذکر شده در بخش‌های قبلی، داده‌های مقادیر فرین دمای شهر تبریز مورد تجزیه و تحلیل قرار می‌گیرند، به این ترتیب که ابتدا به «فزون‌ها» مدل مناسب توزیع پارتوی تعمیم یافته برازش داده، سپس نیکویی برازش آن‌ها پارتوی مورد ارزیابی قرار گرفته و در نهایت برخی پیش‌بینی‌ها انجام می‌شوند. برای ارزیابی نیکویی برازش از نمودارهای تابع چگالی، تابع چندک و qq - پلات استفاده کرده‌ایم. در روش فزون‌ها، یکی از نکات بسیار مهم، انتخاب سرحد است. یکی از روش‌های متداول این است که ابتدا به روش ماکریم، توزیع GEV را به داده‌ها برازش داده، سپس پارامتر میانگین (μ) آن را به عنوان سرحد برای روش فزون‌ها قرار می‌دهند. همچنین با این کار دیگر نیازی به پارامتر مکانی نیست و توزیع پارتوی تعمیم یافته را به صورت $GP(k, \sigma)$ نمایش می‌دهند (اسمیت (۱۹۸۴)).

در این مقاله نیز برای انتخاب سرحد از همین تکنیک استفاده شده است. محاسبات توسط نرم‌افزار R و پکیج‌های $(fExtremes)$ ، $(CompRandFld)$ ، $(extRemes)$ موجود در آن انجام شده است و سرانجام برخی پیش‌بینی‌ها راجع به میانگین دوره بازگشت را برای بیشینه و کمینه دمای شهر تبریز با محاسبه سرحد T - سال برای ۱۰، ۲۰، ۵۰، ۷۵، ۱۰۰ و ۲۰۰ سال به انجام رسانده‌ایم.

لازم به ذکر است که در این مقاله، منظور از داده‌های دمای شهر تبریز، داده‌هایی است که در ایستگاه هواشناسی تبریز ثبت شده است.

برای k به دست می‌آید. سپس می‌توان برآوردگر σ را به صورت:

$$\hat{\sigma}(i, j) = \hat{k}(i, j) \hat{\delta}(i, j) \quad (۲-۵)$$

محاسبه نمود. اکنون یک الگوریتم برای تعیین برآوردهای k و σ ارائه می‌نماییم.

الگوریتم اول

دو آماره ترتیبی $x_{i:n} < x_{j:n}$ را انتخاب و C_j و C_i را محاسبه کنید و قرار دهید

$$d = C_j x_{i:n} - C_i x_{j:n}$$

۱- اگر $d = 0$ باشد، آن‌گاه قرار دهید $\hat{\delta}(i, j) = \pm\infty$ و $\hat{k}(i, j) = 0$ و به مرحله ۵ بروید؛ در غیر این صورت به مرحله ۳ بروید.

۲- مقدار $\delta_0 = \frac{x_{i:n} x_{j:n} (C_j - C_i)}{d}$ را محاسبه کنید. اگر $\delta_0 > 0$ ، آنگاه $\delta_0 > x_{j:n}$. بنابراین روش نصف کردن را روی فاصله $[x_{j:n}, \delta_0]$ بکار برید تا جایی برای $(2-2)$ مانند $\hat{\delta}(i, j)$ به دست آورده و به مرحله ۵ بروید؛ در غیر این صورت به مرحله ۴ بروید.

۳- برای حل $(2-2)$ و به دست آوردن $\hat{\delta}(i, j)$ از روش نصف کردن بر روی $[\delta_0, 0]$ استفاده کنید.

۴- برای محاسبه $\hat{\delta}(i, j)$ و $\hat{k}(i, j)$ با استفاده از $(2-4)$ و $(2-5)$ از $\hat{\delta}(i, j)$ استفاده کنید.

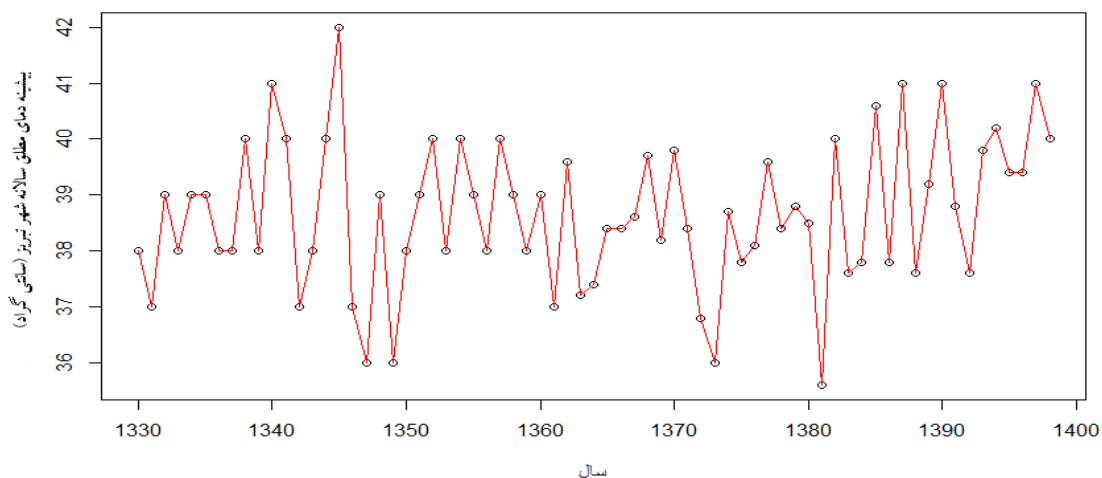
۲-۴-۲- برآورد نهایی

برآوردهای نهایی با استفاده از الگوریتم زیر به دست می‌آیند:

الگوریتم دوم

۱- الگوریتم دوم را برای محاسبه $\hat{\delta}(i, j)$ و $\hat{k}(i, j)$ همه زوج‌های مجزای $x_{i:n} < x_{j:n}$ بکار برید.

۲- میانه‌های هر کدام از مجموعه‌های فوق را برای به دست آوردن برآوردهایی از k و δ بکار برید. یعنی:



شکل ۱. بیشینه دمای سالیانه شهر تبریز از سال ۱۳۳۰ تا ۱۳۹۹

۳-۱-۱- بیشینه دمای شهر تبریز

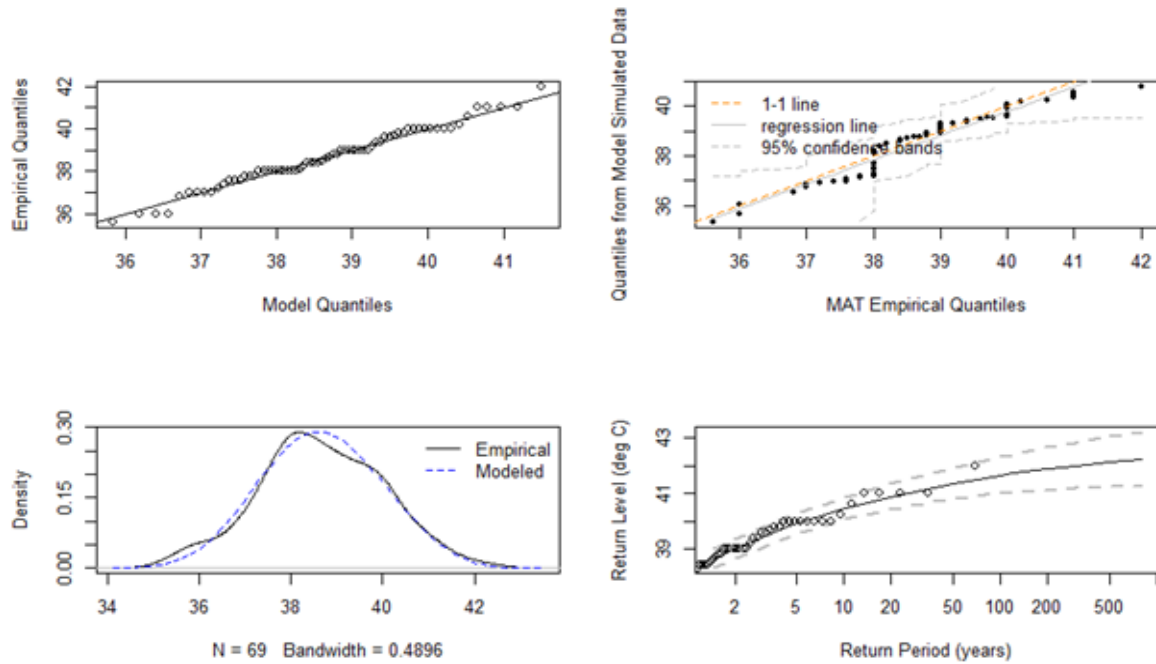
برای این عامل، بیشینه دمای ماهانه شهر تبریز بر حسب سانتی گراد از فروردین سال ۱۳۳۰ تا دی سال ۱۳۹۹ را از آمار سازمان هواشناسی در اختیار داشتیم.

۳-۱-۱-۱- برازش توزیع GEV به بیشینه دمای سالانه

در این قسمت، ماکزیمای سالیانه دمای شهر تبریز را از سال ۱۳۳۰ تا سال ۱۳۹۹ در نظر گرفته ایم به این صورت که در طول هر سال از بین بیشینه دمای ماهانه آن سال، ماکزیمم گرفتیم که رفتار آن در شکل ۱، نشان داده شده است. همان گونه که انتظار داشتیم، با گرفتن ماکزیمای سالیانه، همبستگی داده ها از بین رفته و تقریباً از هم مستقل هستند. جدول ۱، پارامترهای مربوط به توزیع مقادیر فرین بیشینه دمای شهر تبریز را که به روش های ماکزیمم درست نمایی، گشتاوری و گشتاوری وزنی - احتمالی برآورد شده اند را نشان می دهد که نتایج برآورد ماکزیمم درست نمایی و گشتاوری وزنی - احتمالی بسیار نزدیک به هم است. با توجه به جدول فوق به نظر می رسد استفاده از برآورد بیشینه درست نمایی و برآورد گشتاوری وزنی - احتمالی نتایج خوبی را به همراه دارد که البته به خاطر خاصیت پایایی،

برآوردگرهای ماکزیمم درست نمایی ارجح هستند. با توجه به جدول فوق به نظر می رسد استفاده از برآورد بیشینه درست نمایی و برآورد گشتاوری وزنی - احتمالی نتایج خوبی را به همراه دارد که البته به خاطر خاصیت پایایی، ماکزیمم درست نمایی ارجح هستند. بنابراین، برای این عامل، توزیع مقادیر فرین تعمیم یافته با $\sigma = 1.3228$ ، $\mu = 38.1913$ و $\gamma = -0.2746$ به دست آمد. همچنین آزمون $H_0: \gamma = 0$ (فرض پیروی از توزیع گامبل) را انجام دادیم که $P_value = 0.0016$ شده و نشان از رد این فرض است. شکل ۲، نشان دهنده برازش خوب مدل به داده هاست. جدول ۱. برآورد پارامترهای GEV بیشینه دمای شهر تبریز

پارامتر	μ	σ	γ
بیشینه درست نمایی	۳۸/۱۹	۱/۳۲	-۰/۲۷
گشتاوری	۳۷/۰۶	۱/۳۹	-۰/۵۹
گشتاوری وزنی - احتمالی	۳۸/۱۸	۱/۳۳	-۰/۲۷

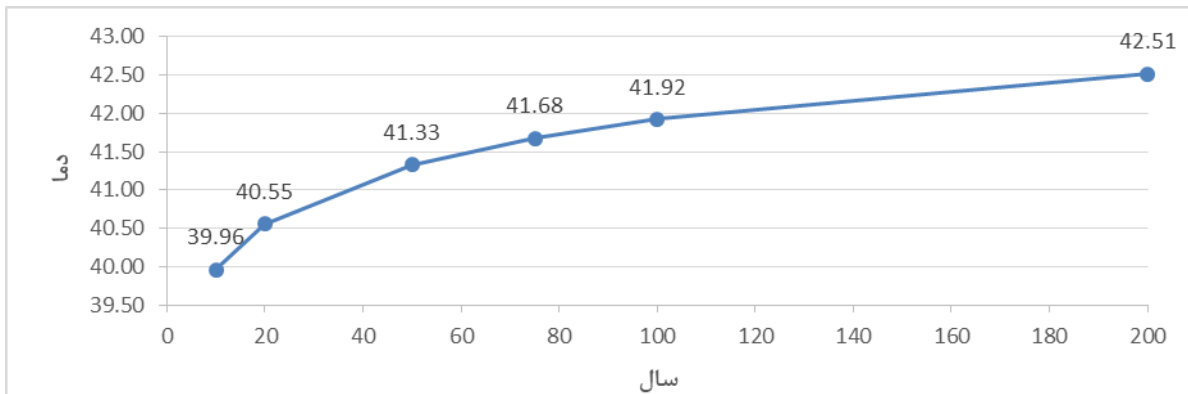


شکل ۲. نمودارهای QQ-plot (سمت راست بالا)، بیشینه دمای سالانه (MAT) با فاصله اطمینان 95٪ (سمت چپ بالا)، نمودار تابع چگالی (سمت راست پایین) و همچنین نمودار دوره بازگشت (سمت راست پایین) مدل برازش داده شده توزیع GEV به مقادیر فرین بیشینه دمای تبریز

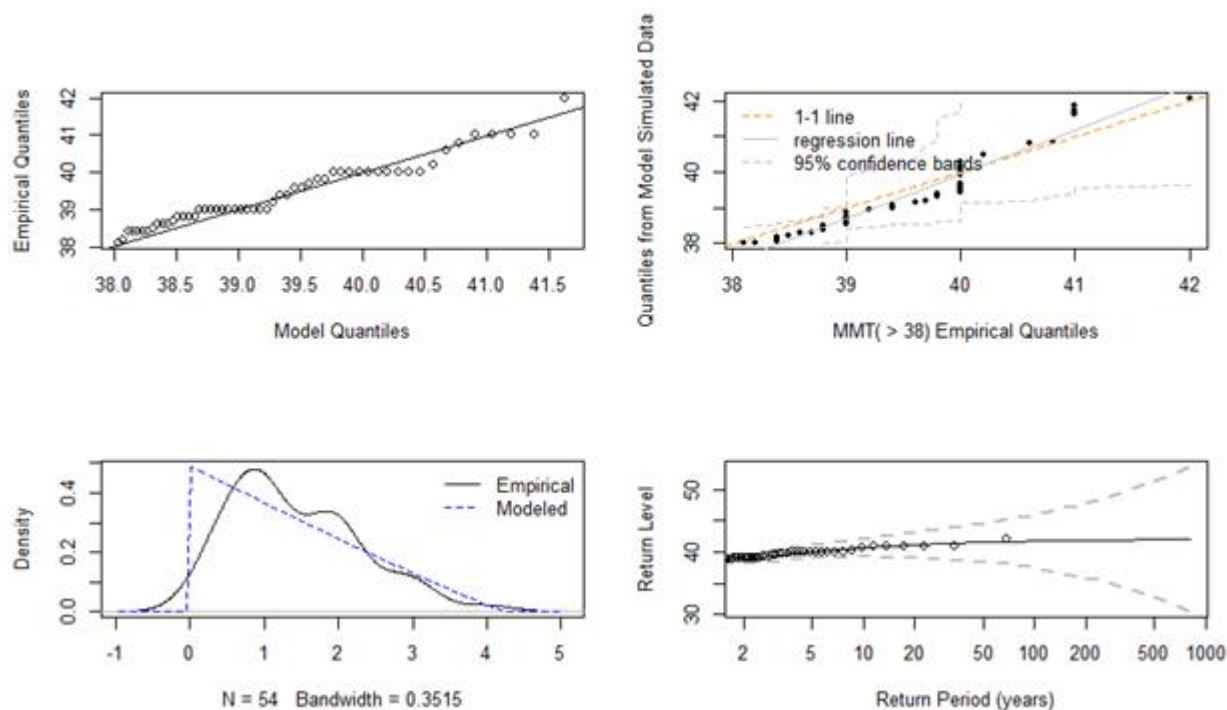
گشتاورها وجود نداشتند، لذا با استفاده از روش صدکهای اصلی، به آنها توزیع پارتوی تعمیم یافته با $\sigma = 2.03882$ و $H_0: k = 0$ انجام آزمون شد. با انجام آزمون $k = -0.48$ (فرض پیروی از توزیع نمایی)، $P_value = 0.0003$ به دست آمد که نشان‌دهنده رد فرض فوق است. شکل ۳، نشان‌دهنده برازش خوب مدل به داده‌هاست.

۲-۱-۳ برازش توزیع GP به بیشینه دمای سالیانه شهر تبریز

با انتخاب سرحد ۳۸ برای داده‌های بیشینه دمای ماهانه شهر تبریز، در بین کل داده‌ها، تعداد ۵۴ ماه که بیشینه دمای آنها بیش از این سرحد بودند انتخاب شدند. دربرازش توزیع پارتوی تعمیم یافته به این داده‌ها، MLE قابل محاسبه نبود و



شکل ۴. سرحد T ساله بیشینه دمای سالیانه شهر تبریز



شکل ۳. نمودارهای QQ-plot (سمت راست بالا)، بیشینه دمای سالانه (MAT) با فاصله اطمینان ۹۵٪ (سمت چپ بالا)، نمودار تابع چگالی (سمت راست پایین) و همچنین نمودار دوره بازگشت (سمت راست پایین) مدل برازش داده شده GPD به مقادیر فرین بیشینه دمای تبریز

جدول ۲. سرحد T ساله بیشینه دمای سالیانه شهر تبریز

ردیف	سطح دوره بازگشت	برآورد	فاصله اطمینان ۹۵٪
۱	۱۰ سال	۳۹/۹۶	(۳۸/۵۴ و ۴۱/۵۷)
۲	۲۰ سال	۴۰/۵۵	(۳۸/۵۴ و ۴۲/۵۶)
۳	۵۰ سال	۴۱/۳۳	(۳۸/۴۴ و ۴۴/۶۰)
۴	۷۵ سال	۴۱/۶۸	(۳۷/۹۵ و ۴۵/۴۱)
۵	۱۰۰ سال	۴۱/۹۲	(۳۷/۲۵ و ۴۶/۱۹)
۶	۲۰۰ سال	۴۲/۵۱	(۳۶/۵۶ و ۴۸/۴۶)

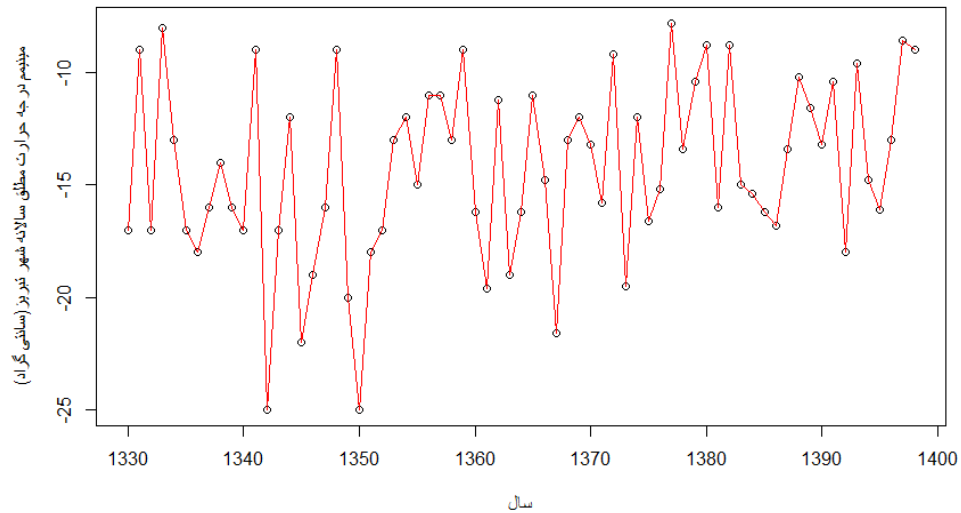
با توجه به این جدول مثلاً به طور متوسط هر ۲۰ سال یک بار بیشینه دما از ۴۰/۵۵ درجه بیشتر می شود و یک فاصله اطمینان ۹۵ درصدی برای بیشینه دمای ۲۰ ساله برابر با فاصله (۳۸/۵۴ و ۴۲/۵۶) است. همچنین یک کران بالا با اطمینان ۹۹ درصد برای دمای بیشینه، ۴۶/۱۹ درجه سانتیگراد با دوره بازگشت ۱۰۰ ساله است.

۳-۱-۳- برخی پیش بینی ها راجع به بیشینه دمای سالیانه شهر تبریز

برای انجام پیش بینی ها، ما مقادیر سرحد T ساله را با توجه به مدل پارتوی تعمیم یافته برازش داده شده به داده ها انجام داده ایم. همان گونه که قبلاً نیز اشاره شد، این مقادیر در واقع چندک (1 - $\frac{1}{T}$) توزیع هستند. پس اگر بخواهیم یک کران بالای ۹۵ درصدی برای دمای بیشینه بدست آوریم، داریم:

$$\left(1 - \frac{1}{T}\right) = 0.95$$

که نتیجه می دهد که: $T = 20$. پس سرحد ۲۰ سال در واقع یک کران بالا با اطمینان ۹۵ درصد برای داده های فرین است. به همین ترتیب سرحد ۱۰۰ سال در واقع یک کران بالا با اطمینان ۹۹ درصد برای داده های فرین است. جدول ۲، نشان دهنده مقادیر سرحد T ساله بیشینه دمای سالیانه شهر تبریز است.



شکل ۵. کمینه دمای سالیانه شهر تبریز از سال ۱۳۳۰ تا سال ۱۳۹۹

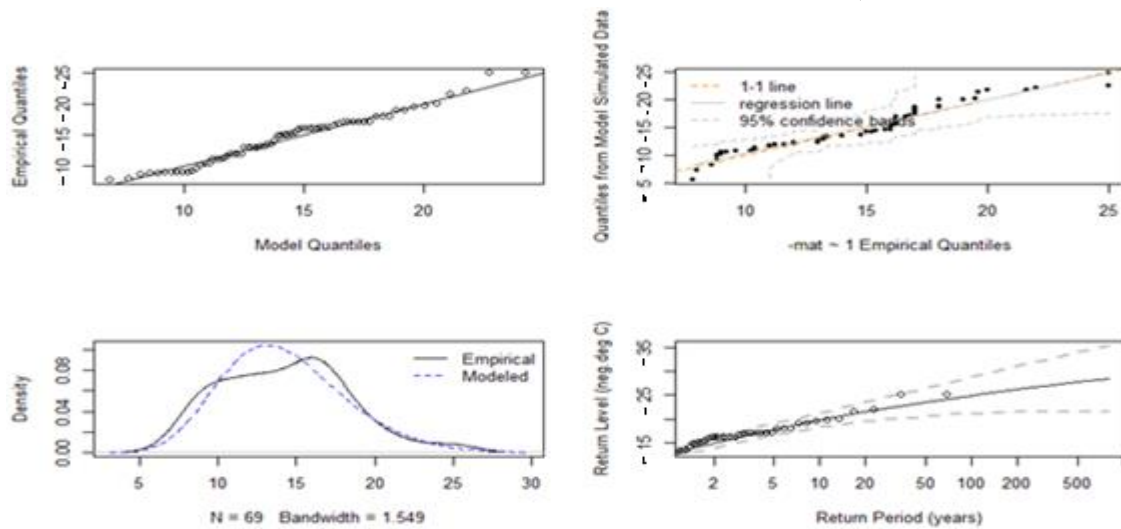
۳-۲- کمینه دمای سالیانه شهر تبریز

برای شهری مانند تبریز که جزو شهرهای سردسیر ایران است، اطلاع از توزیع مقادیر بسیار پایین دما برای کارهایی از قبیل پیش‌بینی مصرف گاز، اهمیت بسیاری دارد. برای این عامل، کمینه دمای ماهیانه شهر تبریز برحسب سانتی‌گراد را از فروردین ۱۳۳۰ تا دی ۱۳۹۹ در اختیار داشته‌ایم.

۳-۲-۱- برازش توزیع *GEV* به کمینه دمای سالیانه

شهر تبریز

در این قسمت، مینمای سالیانه دمای شهر تبریز را از سال ۱۳۳۰ تا سال ۱۳۹۹ در نظر گرفته‌ایم که رفتار آن در شکل ۵، نشان داده شده است.



شکل ۶. نمودارهای QQ-plot (سمت راست بالا)، کمینه دمای سالانه (*MAT*) با فاصله اطمینان ۹۵ درصد (سمت چپ بالا)، نمودار تابع چگالی (سمت راست پایین) و همچنین نمودار دوره بازگشت (سمت راست پایین) مدل برازش داده شده توزیع *GEV* به مقادیر فرین کمینه دمای تبریز

جدول ۳، پارامترهای مربوط به توزیع مقادیر فرین کمینه دمای شهر تبریز که به روش‌های بیشینه درست‌نمایی، گشتاوری و

جدول ۳. برآورد پارامترهای GEV کمینه دمای شهر تبریز

پارامتر	μ	σ	γ
بیشینه درست‌نمایی	-۱۵/۱۷	۴/۳۹۶	-۰/۵۶
گشتاوری	-۱۶/۱۲	۳/۱۲	-۱/۹۴
گشتاوری وزنی - احتمالی	-۱۵/۵۱	۴/۲۳	-۰/۳۹

گشتاوری وزنی - احتمالی برآورد شده‌اند را نشان می‌دهد که نتایج برآورد ماکزیمم درست‌نمایی و گشتاوری وزنی - احتمالی همانند نتایج جدول ۱، بسیار نزدیک به هم است. مجدداً ترجیح ما استفاده از برآورد ماکزیمم درست‌نمایی است. لذا برای این عامل توزیع مقادیر فرین تعمیم یافته با $\sigma = 4.3963$ ، $\mu = -15.172$ و $\gamma = -0.5553$ را در نظر گرفتیم. ما همچنین آزمون $H_0: \gamma = 0$ (فرض پیروی از توزیع گامبل) را انجام داده‌ایم که $P_value = 0.00003$ شده است و نشان می‌دهد که این فرض رد می‌شود. شکل ۶، نشان‌دهنده برازش خوب مدل به داده‌هاست.

جدول ۴. سر حد T ساله کمینه دمای سالیانه شهر تبریز

ردیف	فاصله اطمینان ۹۵٪	برآورد	سطح دوره بازگشت
۱	(-۲۰/۶۰ و -۳۲/۳۵)	-۱۹/۳۸	۱۰ سال
۲	(-۲۲/۳۷ و -۱۹/۳۳)	-۲۰/۹۱	۲۰ سال
۳	(-۲۳/۹۳ و -۲۰/۶۵)	-۲۲/۸۰	۵۰ سال
۴	(-۲۶/۰۳ و -۲۱/۲۳)	-۲۳/۶۳	۷۵ سال
۵	(-۲۶/۸۱ و -۲۱/۶۵)	-۲۵/۰۱	۱۰۰ سال
۶	(-۲۸/۶۸ و -۲۲/۶۳)	-۲۵/۶۶	۲۰۰ سال

- vulnerability analysis of road network. Transportation Research Part C: Emerging Technologies, Volume 120, pp 1-30
2. Cheng, L. AghaKouchak, A. Gilleland, E. Katz, R. W. (2014), non-stationary extreme value analysis in a changing climate. *Clim. Change*, 127, pp. 353-369.
 3. Cooley, D. and Sain, S. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of agricultural, biological, and environmental statistics*, 15(3):381-302.
 4. Davison, A. C. (1983), Modeling excesses over High Thresholds, with an Application. In *Statistical and Applications*, ed Tiago de Oliveira, NATO ASI Series, Dordrecht: Reidel, pp. 361-382
 5. Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 368, pages 581-608. The Royal Society.
 7. Davison, A.C. and Smith, R.L. (1990), Models for Exceedances over High Thresholds, *J. R. Statist. Soc. B* 52, No. 3, 393-332.
 8. Fisher, R.A. and Tippett, L. H. C. (1928), Limiting forms of the Frequency Distribution of the Large or Smallest Member of Sample. *Proc. Cambridge Phil. Soc.* 24, 180-190
 9. Grimshaw, S. D. (1993), Computing Maximum Likelihood Estimation for the Generalized Pareto Distribution, *Technometrics*, Vol. 35 No. 2, 185-191.
 10. Hosking, J. (1985). Algorithm as 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(3):301-310.
 11. Hosking, J.R.M. and Wallis. J.R.(1987), Parameter and Quantile Estimation for the Generalized Pareto Distribution, *Technometrics*, 29, 339-349.
 12. Hosking, J.R.M., Wallis. J.R. and Wood, E. F. (1985), Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. *Technometrics*, 27, 251-261.
 13. Joe, H. (1987), Estimation of Quantiles of the Maximum of N Observations, *Biometrika*, 74, 347-354.
 14. Kang, S.; Song, J. (2017) Parameter and quantile estimation for the generalized Pareto

۴ بحث و نتیجه گیری

در این مقاله ضمن بیان دو روش معمول تحلیل مقادیر فرین یعنی استفاده از توزیعیهای مقادیر فرین تعمیم یافته و پارتوی تعمیم یافته و برآورد پارامترهای آنها، داده‌های مقادیر فرین بیشینه و کمینه دمای شهر تبریز را مورد تجزیه و تحلیل قرار دادیم. به این ترتیب که ابتدا به روش ماکزیمیا، توزیع GEV را به داده‌ها برازش داده، سپس پارامتر میانگین (μ) آن را به عنوان سرحد برای روش «فزونیه‌ها» قرار داده و توزیع پارتوی تعمیم یافته را به فزونیه‌های بیشتر (کمتر) از آن سرحد برازش دادیم. سپس نیکویی برازش آن‌ها را مورد ارزیابی قرار دادیم که نمودارها نشان از برازش خوب مدل‌ها به داده‌ها داشت. در نهایت برخی پیش‌بینی‌ها راجع به میانگین دوره بازگشت را برای بیشینه و کمینه دمای شهر تبریز با محاسبه سرحد T - سال برای ۱۰، ۲۰، ۵۰، ۷۵، ۱۰۰ و ۲۰۰ سال انجام دادیم. محاسبات نشان می‌دهند که بطور مثال، با اطمینان ۹۵ درصد، کران بالا برای بیشینه دمای هوای تبریز برابر با ۴۲/۵۶ درجه سانتیگراد با میانگین دوره بازگشت ۲۰ ساله و کران پایین برای کمینه دمای هوای تبریز برابر با ۲۲/۳۷ - درجه سانتیگراد با میانگین دوره بازگشت ۲۰ ساله است. همچنین یک کران بالا با اطمینان ۹۹ درصد برای دمای بیشینه، ۴۶/۱۹ درجه سانتیگراد با دوره بازگشت ۱۰۰ ساله و یک کران پایین با اطمینان ۹۹ درصد برای دمای کمینه، ۲۶/۸۱ - درجه سانتیگراد با دوره بازگشت ۱۰۰ ساله است.

تقدیر و تشکر

بدینوسیله از اداره کل هواشناسی استان آذربایجان شرقی که داده‌های این تحقیق را در اختیار قرار دادند، تشکر و قدردانی می‌شود.

۵ منابع

1. Ansari Esfeh, M., Kattan, L., Lam, W.H.K., Ansari Esfe, R., Salari, M. (2020), Compound generalized extreme value distribution for modeling the effects of monthly and seasonal variation on the extreme travel delays for

23. Rajabi, M.R., Modarres, R. (2008), Extreme Value Frequency Analysis of Wind Data for Isfahan, Iran. *Journal of Wind Engineering and Industrial Aerodynamics* 96, 78-87.
24. Rootzen, H. and Tajvidai, N. (1997), Extreme Value Statistics and Wind Storm Losses: A Case Study, *Scandinavian Actuarial Journal*, 1: 70-94.
25. Schlögl, M. Laaha, G. (2017), Extreme weather exposure identification for road networks—a comparative assessment of statistical methods. *Nat. Hazards Earth. Sci.*, 17, pp. 515-531.
26. Sharpe, M., Juárez, M. A., (2021), Estimation of the Pareto and related distributions – A reference-intrinsic approach, *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2021.1916826
27. Smith, R. L. (1984), Threshold Method for Sample Extremes. In *Statistical Extremes and Applications*,
28. ed, T. Tiago de Oliveira, Dordrecht: Reidal, 261-638.
29. Smith, R. L. (1985), Maximum Likelihood Estimation in a class of Nonregular Cases. *Biometrika*, 72, 67-90
30. Smith, R. L. (1987), Estimating Tails of Probability Distributions. *The Annals of Statistics*, 15, 1173-1207.
31. Wenuri, H. Xu, S. Naji, S. (2008), Evaluation of GEV Model for Frequency Analysis of Annual Maximum Water Levels in the Coast of United States. *Ocean Engineering*. 35, 1132-1147.
32. Yi, H., Liang P., Dabao, Zh., Zifeng, Zh., (2021), Risk Analysis via Generalized Pareto Distributions. *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2021.1874390.
- distribution in peaks over threshold framework. *J. Korean Stat. Soc.*, 46, 487–501.
15. Kumar, D., Nassar, M., Malik, M.R. (2020), Estimation of the Location and Scale Parameters of Generalized Pareto Distribution Based on Progressively Type-II Censored Order Statistics. *Ann. Data. Sci.* <https://doi.org/10.1007/s40745-020-00266-0>
16. Langousis, A., Mamalakis, A., Puliga, M., Deidda, R. (2016), Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resource. Res.*, 52, 2659–2681.
17. Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983), *Extremes and Related Properties of Random*
18. *Sequences and Processes*. Springer-Verlag, New York Inc.
19. Leonard, M. Metcalfe, A. Lambert, M. (2008), Frequency Analysis of Rainfall and Stream Flows Extremes According for Seasonal and Climatic Partitions. *Journal of Hydrology*, 348, 135-147.
20. Martins, A.L.A., Liska, G.R., Beijo, L.A. (2020), Generalized Pareto distribution applied to the analysis of maximum rainfall events in Uruguaiiana, RS, Brazil. *SN Appl. Sci.* 2, 1479. <https://doi.org/10.1007/s42452-020-03199-8>
21. Nagatsuka, H., Balakrishnan, N. (2021), Efficient likelihood-based inference for the generalized Pareto distribution. *Ann Inst Stat Math.* <https://doi.org/10.1007/s10463-020-00782-z>
22. Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131.

Data analysis on the extreme values of temperature of Tabriz city

Ali Akbar Heydari^{*1}

1. Assistant Professor, Department of Statistics, University of Tabriz, Tabriz, Iran

(Received: 06 Jun 2021, Revised: 25 Sep 2021, Accepted: 07 Nov 2021, Published online: 07 Nov 2021)

Abstract

Modeling data of extreme values, i.e., the highest and lowest values of natural phenomena by statistical formulas, makes it possible to predict the behavior of such data in the future. One of the methods of analyzing extreme values is the exceedances method, in which all the exceedances over a high threshold are modeled using the generalized Pareto distribution. In this paper, using this method, the statistical model is fitted to the data of high values and low values of temperature of Tabriz city in the period of 1951 to 2020, the parameters of the model are estimated, the goodness of fit of the model is evaluated, and finally some necessary predictions such as the mean return periods of minimum and maximum values of the temperature of Tabriz have done for different periods of time. One of the results of this study is that with 95% confidence, the upper limit for the maximum temperature of Tabriz is equal to 41.11°C with an average return period of 20 years and the lower limit for the minimum temperature of Tabriz is equal to -20.91°C with the average return period of 20 years.

Keywords: Maximum temperature, Minimum temperature, Generalized Pareto distribution, Prediction, Return period.

* Corresponding author:
Email Address: heydari@tabrizu.ac.ir