

Penjelmaan Box-Cox dan Penggunaannya dalam Analisis Frekuensi Banjir (*Box-Cox Transform and Its Application in Flood Frequency Analysis*)

Ani Shabri

Jabatan Matematik, Fakulti Sains, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor
e-mail: ani@utm.my

Abstrak Dalam kertas ini, tiga kaedah digunakan untuk menganggar parameter Box-Cox bagi menjelmakan data kepada taburan normal. Kaedah yang dicadangkan adalah berdasarkan parameter yang dapat mengoptimumkan pengujian statistik Anderson Darling, Shipro Wilk dan min ralat kuasa dua. Kesesuaian kaedah ini digabungkan dengan taburan normal dinilai dan dibandingkan dengan taburan yang begitu popular dalam analisis frekuensi banjir iaitu taburan General Extreme Value (GEV). Keupayaan taburan ini dinilai menggunakan data aliran tahunan banjir di seluruh stesen di Johor. Analisis kajian menunjukkan taburan normal setelah data dinormalkan menggunakan parameter Box-Cox memberikan kesesuaian yang sebaik dan setanding dengan taburan GEV dan ianya dicadangkan sesuai digunakan untuk kajian selanjutnya.

Katakunci Penjelmaan Box-Cox; analisis frekuensi banjir; pengujian statistik; L-momen.

Abstract This paper studies three methods for estimating the Box-Cox parameter used to transform data to normality. The proposed methods are based on finding the parameter values that optimizing the Anderson-Darling, Shipro Wilk and relative mean square error tests for normality. The performance of the these methods combined with normal distribution are evaluated and compared with the most applied distributions for flood frequency analysis, namely generalized extreme value (GEV). The performances of these distributions are tested using the annual maximum flood data of Johor stations. The results show that the normal distribution after transformed data to normality performed as well as the GEV distribution and they are recommended for further analysis.

Keywords Box-Cox Transform; flood frequency analysis; test statistics; L-momen.

1 Pengenalan

Mengenalpasti taburan statistik yang benar-benar sesuai dipadankan dengan set data hidrologi menjadi masalah utama yang dihadapi oleh jurutera terutamanya apabila set data bersaiz kecil dan terhad. Masalah yang wujud adalah berkaitan dengan mengenalpasti taburan statistik yang sebenar untuk dipadankan dengan data sampel disebabkan taburan sebenar populasi tidak diketahui.

Terdapat banyak taburan yang dicadangkan seperti taburan lognormal 3 (LN3), Pearson-III (P3), Log-Pearson III (LP3), generalized extreme value (GEV), Extreme Value Type 1

(EV1), Wakeby, Generalized Pareto (GP) and Generalized-Logistic (GL) untuk dipadankan dengan data aliran tahunan maksimum. Taburan LP3 telah dicadangkan dalam analisis banjir di Australia dan Amerika Syarikat. Laporan kajian banjir yang dilakukan oleh British National Environment Research Council (NERC) terhadap data dari 30 stesen di UK dengan data melebihi 30 tahun, didapati taburan GEV sesuai untuk dipadankan dengan data aliran banjir di Great Britain (Raynal dan Salas [11]).

Ahmad et al. [1] telah menguji keupayaan taburan Log-Logistik dengan taburan GEV, P3 dan LN3. Taburan GL didapati lebih baik dalam menganggarkan dalam menganggarkan banjir pada nilai tempoh ulangan yang tinggi.

Haktanir [5] telah membandingkan beberapa taburan menggunakan data simulasi berdasarkan taburan LP3, GEV dan Wakeby dengan saiz sampel 30 dan 100 tahun. Hanya anggaran banjir pada tempoh ulangan 100, 1000 dan 10 000 tahun dipertimbangkan. Beliau mendapati bahawa taburan LN3 adalah terbaik berbanding taburan LN2, EV1, LP3 dan P3.

Onoz dan Bayazit [8] telah menggunakan data dari 19 stesen dengan saiz sampel melebihi 60 tahun dari beberapa buah negara dianalisis menggunakan tujuh taburan. Hasil kajian menggunakan ujian statistik Kolmogorov-Smirnov, pekali korelasi, Anderson Darling dan rajah L-momen didapati taburan GEV adalah yang terbaik.

Vogel et al. [14] telah menganalisa data banjir dari 61 stesen di Australia dengan saiz data melebihi 20 tahun dan 383 stesen di USA dengan saiz data melebihi 30 tahun. Gambarajah L-momen digunakan untuk membandingkan antara taburan. Hasil kajian didapati bahawa taburan GP adalah yang terbaik di Australia manakala taburan LN2, LN3, GEV dan LP3 di USA.

Vogel et al. [15] menggunakan 383 data aliran banjir di USA. Purata saiz sampel adalah disekitar 50 tahun. Rajah L-momen juga digunakan untuk menguji kesesuaian model. Hasil yang dijalankan didapati taburan LN2, LN3, LP3 dan GEV adalah yang sesuai.

Oleh kerana belum ada kesepakatan ahli statistik atau hidrologi bagi menentukan taburan statistik yang terbaik untuk dijadikan taburan piawai dalam analisis frekuensi banjir, oleh itu dalam kajian ini kami cuba melihat dan mengkaji penjelmaan Box-Cox taburan normal samada sesuai digunakan. Tiga kaedah digunakan untuk menganggar parameter Box-Cox bagi menjemakan data kepada normal. Kesemua kaedah adalah berdasarkan kepada pengoptimuman pengujian statistik iaitu statistik Shipro-Wilk dan Anderson dan min ralat kuasa dua. Ketiga-tiga kaedah ini digunakan dalam analisis frekuensi banjir dan kekuatan kaedah ini dibanding dan dinilai dengan taburan GEV.

2 Penjelmaan Box-Cox

Penjelmaan Box-Cox telah diperkenalkan oleh G.E.P. Box dan D. R. Box dalam tahun 1964. Penjelmaan ini didapati berupaya untuk menjelmakan sebarang data tak normal kepada yang data menghampiri taburan normal (Thyer et al. [13]). Semenjak penjelmaan ini diperkenalkan ianya didapati begitu meluas digunakan tertutamanya dalam siri masa (Spitzer [12], Park [9]; Thyer et al. [13]; Grimaldi [4]), logik kabur, analisis multivarit (Freeman & Modarres [3]) dan lain-lain bidang. Penjelmaan ini sesuai untuk nilai data positif dan satu pemalar perlu ditambah jika terdapat set cerapan yang mengandungi satu atau lebih nilai negatif.

Penjelmaan Box-Cox diberikan oleh

$$y = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(x), & \lambda = 0 \end{cases} \quad (1)$$

dengan λ adalah parameter. Penjelmaan ini memberikan penjelamaan skala cerapan asalan x kepada skala y supaya data tertabur secara normal.

2.1 Taburan Normal

Taburan normal merupakan taburan yang terpenting dalam hampir keseluruhan bidang statistik. Fungsi ketumpatan kebarangkalian bagi taburan normal diberikan oleh

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \quad (2)$$

dengan y adalah penjelmaan data, μ dan σ adalah min dan sisihan piawai bagi y . Fungsi taburan kumulatif bagi taburan normal diberikan oleh

$$F(u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (3)$$

Kuantil bagi taburan normal diberikan oleh

$$\hat{x}_i = (\lambda y_i + 1)^{1/\lambda} \quad (4)$$

dengan $y_i = \mu + \sigma z_i$ dan z_i adalah pembolehubah normal piawai yang berpadanan dengan cerapan ke- i (Rao & Hamed [10]).

2.2 Anggaran bagi Parameter Taburan Normal

Anggaran parameter bagi taburan normal μ dan σ berdasarkan kaedah kebolehhadiah maksimum dan kaedah momen adalah sama (Rao & Hamed [10]) dan boleh ditulis sebagai

$$\mu = \frac{1}{n} \sum_{i=1}^n y_i \quad (5)$$

dan

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2} \quad (6)$$

3 Pengujian Statistik

Terdapat banyak jenis pengujian statistik yang digunakan untuk menguji kenormalan data. Sebahagian daripada pengujian statistik didapati mudah tetapi hanya sesuai untuk sesuatu keadaan.

Pengujian statistik traditional seperti Shapiro-Wilk (SW) merupakan salah satu pengujian statistik yang paling berupaya dalam kajian literature dalam menguji kenormalan data (Zhang & Wu [16]), terutama untuk sampel baersaiz kecil ($n \leq 50$). Pengujian statistik lain seperti statistik Anderson Darling, tidak sebaik statistik SW tetapi terbaik berbanding statistik berdasarkan fungsi taburan empirik seperti statistik Cramer-von Mises (CR), Watson (W) dan Kolmogorov Smirnov (KS).

3.1 Statistik Anderson-Darling

Pengujian statistik Anderson-Darling ditakrifkan sebagai

$$AD = n \int_{-\infty}^{\infty} \frac{\{F_n(x) - F(x)\}^2}{F(x)\{1 - F(x)\}} dF(x) = - \sum_{i=1}^n \frac{2i - 1}{n} \cdot [\ln(F(X_{(i)})) + \ln(1 - F(X_{(n+1-i)}))] - n. \tag{7}$$

$F(x)$ adalah fungsi ketumpatan bagi taburan normal bagi pembolehubah rawak x diperolehi dari persamaan (3).

3.2 Pengujian Statistik Shapiro-Wilk

Pengujian statistik Shapiro-Wilk (SW) merupakan pengujian statistik yang begitu diminati dan meluas digunakan untuk menguji kenormalan data (Bai & Chen, [2]). Pengujian statistik SW ditakrifkan sebagai

$$SW = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{8}$$

dengan $x_{(i)}$ adalah statistik tertib, $a' = (a_1, \dots, a_n) = m'V^{-1} (m'V^{-2}m)^{-1/2}$, m dan V adalah min vektor dan matrik kovarians bagi statistik tertib bagi taburan normal piawai diperolehi dari persamaan (4).

3.3 Min Ralat Relatif Kuasa Dua

Min ralat relatif kuasa dua (MRK) sering digunakan dalam hidrologi untuk menguji dan menentukan kesesuaian data cerapan dengan cerapan yang dijangkan berdasarkan kaedah yang berlainan. MRK ditakrifkan sebagai

$$MRK = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \hat{x}_i}{x_i}\right)^2 \tag{9}$$

dengan x_i adalah nilai cerapan dan \hat{x}_i adalah nilai jangkaan diperolehi dari persamaan (4).

3.4 Anggaran bagi Parameter Box-Cox

Tiga kaedah pengujian statistik AD, SW dan MRK digunakan untuk menentukan parameter λ . Nilai statistik ini digunakan sebagai fungsi objektif dan masing-masingnya ditakrifkan sebagai

(1) Statistik Anderson-Darling

$$F_1 = AD$$

(2) Statistik Shipro-Wilk

$$F_2 = SW$$

(3) Min Ralat Relatif Kuasa Dua

$$F_3 = MRK$$

Kaedah leleran mudah digunakan dengan nilai awalan $\lambda_0 = -1.0$ ditetapkan dan dimasukkan kedalam fungsi objektif F_1 , F_2 dan F_3 . Nilai awalan λ_0 kemudiannya ditambah dengan 0.0011 dan nilai fungsi objektif ini dikira. Proses ini diulangi sehingga nilai $\lambda = 1.0$. Parameter λ yang terbaik dipilih apabila fungsi objektif bagi statistik (1), (2) dan (3) menghasilkan nilai yang optimum dan penjelmaan data tertabur secara normal.

4 Taburan GEV

Taburan Nilai Ekstim Umum “General Extreme Value (GEV)” diperkenalkan oleh Jenkinson dalam tahun 1955. Taburan ini begitu meluas digunakan dalam kejadian ekstim seperti banjir, kemarau, paras air laut ekstim, ribut. Taburan GEV telah dijadikan taburan piawai di United Kingdom dalam analisis banjir, analisis frekuensi curahan hujan di United States dan analisis ombak laut (Martins & Stedinger [7]).

Fungsi ketumpatan kebarangkalian (fkk) bagi taburan GEV diberikan oleh

$$f(x) = \frac{1}{\alpha} \left[1 - k \left(\frac{x - \xi}{\alpha} \right) \right]^{1/k-1} e^{-[1 - k(\frac{x-\xi}{\alpha})]^{1/k}} \quad (10)$$

dengan parameter lokasi ξ , α dan k adalah parameter lokasi, skala dan bentuk masing-masing. Fungsi taburan kumulatif bagi taburan GEV boleh ditulis sebagai

$$\begin{aligned} F(x) &= \exp \left[- \left\{ 1 - \frac{k}{\alpha} (x - \xi)^{1/k} \right\} \right] \quad k \neq 0 \\ &= \exp \left[- \exp \left\{ - \left(\frac{x - \xi}{\alpha} \right) \right\} \right] \quad k = 0 \end{aligned} \quad (11)$$

Julat bagi pembolehubah rawak x bergantung kepada tanda bagi parameter k . Bila k negatif ($\gamma > 1.1396$) pembolehubah x berada dalam julat $\xi + \alpha/k < x < \infty$ yang mana ia sesuai untuk analisis frekuensi banjir. Kuantil bagi taburan GEV diberikan oleh

$$x(F) = \xi + \frac{\alpha}{k} \left[1 - [-\ln \{F(x)\}]^k \right] \quad (12)$$

Parameter taburan GEV begitu popular dianggarkan menggunakan kaedah L-momen (Hosking [6]).

5 Analisis Data

Data aliran sungai di negeri Johor yang diperolehi daripada Jabatan Perparitan dan Saliran Malaysia digunakan dalam kajian ini. Maklumat yang berkaitan dengan data ditunjukkan dalam Jadual 1. Terdapat 10 stesen saiz data antara 10 hingga 42 tahun dengan min aliran banjir bagi setiap stesen adalah berbeza iaitu dalam julat 2.4 - 6490 m^3/s , keluasan kawasan tadahan 1.84 - 3130 km^2 , pekali variasi (CV) 0.452 - 1.285, pekali kepencongan (Cs) -0.102 - 2.847 dan pekali kurtosis (Ck) 2.4-11.856. Pekali kepencongan dan pekali kurtosis pencong yang ditunjukkan dalam Jadual 1 menunjukkan ada data dalam stesen kajian tertabur secara normal. Bagi menentukan wujudnya data yang tertabur secara normal, pengujian statistik Shipro-Wilk digunakan. Nilai aras keertian p untuk pengujian SW dapat ditunjukkan dalam Jadual 1.

Jadual 1: Ciri-Ciri Aliran Tahunan Maksimum Sungai Di Negeri Johor

| Bil | No. Stesen | Tempoh Data | Saiz Sampel n | Luas Kawasan Tadahan | Min (m^3/s) | Pekali Variasi Cv | Pekali Kepencongan Cs | Pekali Kurtosis Ck | Nilai p Statistik SW |
|-----|------------|-------------|---------------|----------------------|-----------------|-------------------|-----------------------|--------------------|----------------------|
| 1 | 1732401 | 1981-1996 | 16 | 1.84 | 1.21 | 0.587 | 1.076 | 4.068 | 0.0600* |
| 2 | 1737451 | 1965-2002 | 38 | 1130.00 | 238.95 | 0.693 | 1.372 | 4.442 | < 0.0001 |
| 3 | 1836401 | 1978-1992 | 15 | 209.00 | 53.53 | 0.960 | 2.477 | 9.822 | < 0.0001 |
| 4 | 1836402 | 1977-2002 | 26 | 624.00 | 96.73 | 0.617 | 1.885 | 6.824 | < 0.0001 |
| 5 | 2130401 | 1978-1990 | 10 | 286.00 | 45.91 | 0.722 | -0.102 | 2.490 | 0.3828* |
| 6 | 2130422 | 1971-2002 | 30 | 350.00 | 35.29 | 0.617 | 0.373 | 2.667 | 0.3391* |
| 7 | 2235401 | 1978-1995 | 25 | 587.00 | 444.44 | 0.711 | 1.301 | 4.114 | 0.0011 |
| 8 | 2237471 | 1961-1994 | 42 | 207.00 | 693.34 | 1.124 | 2.847 | 10.927 | < 0.0001 |
| 9 | 2527411 | 1961-1988 | 37 | 3130.00 | 208.83 | 0.452 | 1.736 | 7.210 | 0.0001 |
| 10 | 2528414 | 1961-1984 | 37 | 658.00 | 279.05 | 1.285 | 2.759 | 11.856 | < 0.0001 |

* Tertabur Secara Normal pada Aras Keertian 5%

Hasil pengujian statistik SW yang ditunjukkan dalam Jadual 1 menunjukkan bahawa hanya 3 stesen (stesen 1, 5 dan 6) didapati tertabur secara normal iaitu menghasilkan nilai p lebih besar dari aras keertian 5%.

5.1 Anggaran Parameter Penjelmaan Box-Cox

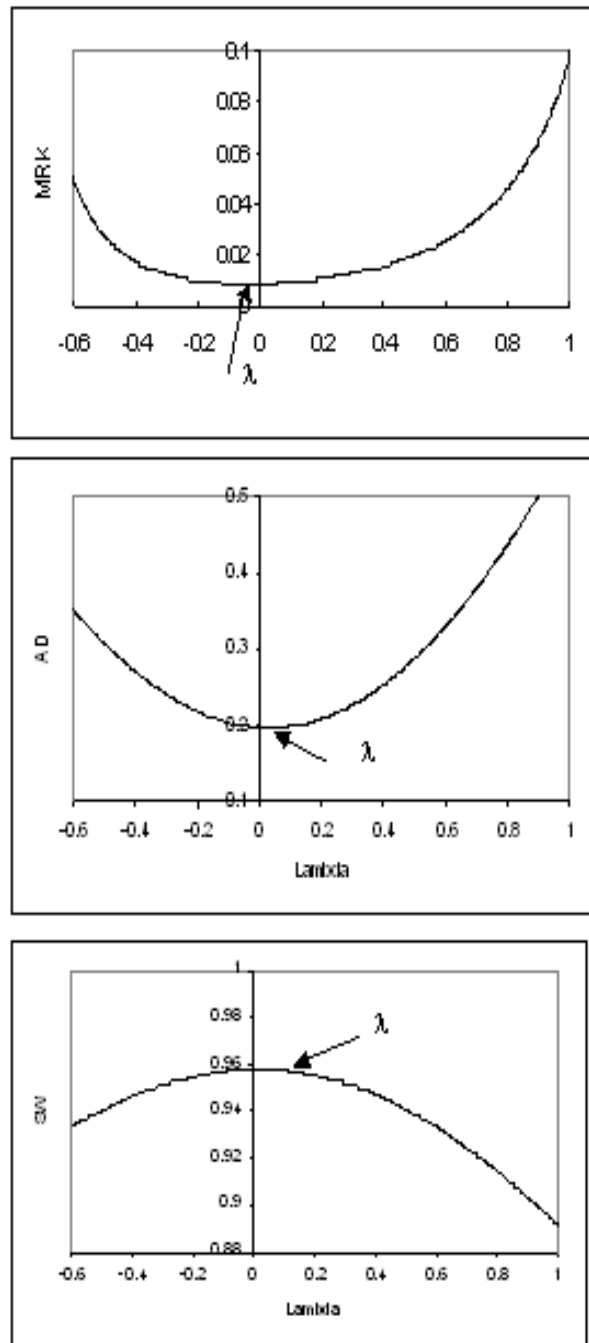
Kesemua data aliran tahunan maksimum di Johor dijelmakan menggunakan penjelmaan Box-Cox dengan λ bermula dengan -1.0 dan ditambah 0.001 sehingga $\lambda = +1$. Nilai λ ditentukan apabila ianya dapat mengoptimumkan fungsi objektif. Rajah 1 menunjukkan fungsi objektif yang optimum untuk kaedah statistik SW, AD dan MRK untuk stesen 1.

Jadual 2 menunjukkan nilai λ bagi setiap fungsi objektif yang optimum dan pengujian kenormalan data yang telah dijelmakan dengan menggunakan pengujian statistik SW dan AD untuk kesemua stesen.

Jadual 2 menunjukkan bahawa didapati parameter Box-Cox (λ) adalah positif jika data asal tertabur secara normal dan negatif jika data asal tidak tertabur normal berdasarkan kaedah pengujian statistik AD dan SW. Hampir kesemua data setelah penjelmaan Box-Cox dilakukan tertaburan secara normal kecuali untuk stesen 5 berdasarkan kaedah MRK.

5.2 Kesesuaian Penjelmaan Box-Cox Berbanding Taburan GEV

Parameter penjelmaan Box-Cox yang optimum berdasarkan statistik AD, SW dan MRK yang diperolehi dalam Jadual 2 digunakan untuk menganggar kuantil taburan normal



Rajah 1: Keluk bagi Fungsi Objektif Kaedah MRK, AD dan SW bagi Menentukan Parameter Penjelmaan Box-Cox yang Optimum

Jadual 2: Pengoptimuman Fungsi Objektif dan Pengujian Kenormalan Data

| Bil | No. Stesen | n | MRK | | | Anderson Darling | | Shipro-Wilk | |
|-----|------------|----|-------------------|------------------|------------------|-------------------|------------------|-------------------|------------------|
| | | | Parameter Box-Cox | Nilai p Stat. AD | Nilai p Stat. SW | Parameter Box-Cox | Nilai p Stat. AD | Parameter Box-Cox | Nilai p Stat. SW |
| 1 | 1732401 | 16 | -0.0391 | 0.8639 | 0.6271 | 0.0350 | 0.8690 | 0.0130 | 0.6299 |
| 2 | 1737451 | 38 | -0.1788 | 0.0859 | 0.0485 | -0.2290 | 0.0856 | -0.2730 | 0.0511 |
| 3 | 1836401 | 15 | -0.6294 | 0.2348 | 0.4270 | -0.7460 | 0.2570 | -0.5590 | 0.4386 |
| 4 | 1836402 | 26 | -0.5546 | 0.8629 | 0.7897 | -0.7680 | 0.8800 | -0.7240 | 0.8250 |
| 5 | 2130401 | 10 | -0.3944 | 0.0007 | 0.0009 | 0.9700 | 0.4710 | 0.9810 | 0.3829 |
| 6 | 2130422 | 30 | 0.4108 | 0.3948 | 0.3460 | 0.7500 | 0.7330 | 0.6950 | 0.5571 |
| 7 | 2235401 | 25 | 0.0225 | 0.4078 | 0.5640 | -0.1300 | 0.4790 | -0.0530 | 0.5862 |
| 8 | 2237471 | 42 | -0.4547 | 0.2686 | 0.3171 | -0.5810 | 0.3180 | -0.5590 | 0.3646 |
| 9 | 2527411 | 37 | -0.2091 | 0.8000 | 0.7472 | -0.2290 | 0.7500 | -0.2950 | 0.8230 |
| 10 | 2528414 | 37 | -0.2525 | 0.2565 | 0.3925 | -0.4050 | 0.3767 | -0.3500 | 0.4531 |

berdasarkan pers. (4). Kekuatan ketiga-tiga kaedah ini kemudiannya dinilai dan dibandingkan dengan taburan GEV dan taburan normal berdasarkan min punca ralat kuasa dua seperti yang ditunjukkan dalam Jadual 3.

Data yang dipadankan dengan taburan normal didapati menghasilkan anggaran yang paling tidak sesuai berbanding taburan GEV dan penjelmaan Box-Cox taburan normal yang lain. Penjelmaan Box-Cox taburan normal berdasarkan statistik AD memberikan nilai min ralat kuasa dua yang lebih rendah berbanding taburan GEV untuk 4 stesen, statistik SW untuk 5 stesen dan MRK untuk 6 stesen.

Kecekapan penganggaran, ϕ digunakan untuk mengukur kecekapan kaedah penjelmaan Box-Cox berdasarkan statistik AD, SW dan MRK berbanding taburan GEV dalam memodelkan aliran tahunan maksimum di Johor. Kecekapan anggaran, ϕ ditakrifkan sebagai

$$\phi = \frac{\text{MRK Penjelmaan Box-Cox}}{\text{MRK Taburan GEV}}$$

Nilai ϕ kurang daripada satu mencadangkan bahawa penjelmaan Box-Cox lebih baik berbanding taburan GEV. Hasil kajian kecekapan penganggaran, ϕ untuk seluruh stesen di Johor dapat ditunjukkan dalam Jadual 3.

Jadual 3: Perbandingan Kaedah Penjelmaan Box-Cox dan Taburan Normal Berdasarkan Min Ralat Relatif Kuasa Dua dan Kecekapan Anggaran ϕ

| Bil | No. Stesen | n | Min Punca Ralat Kuasa Dua | | | | | Kecekapan Anggaran ϕ | | | | |
|-----|------------|----|---------------------------|----------------|----------------|--------------|--------------|---------------------------|----------------|----------------|--------------|--------------|
| | | | Taburan GEV | Taburan Normal | Penjelmaan MRK | Taburan AD | Taburan SW | Taburan GEV | Taburan Normal | Penjelmaan MRK | Taburan AD | Taburan SW |
| 1 | 1732401 | 16 | 0.011 | 0.098 | 0.009 | 0.009 | 0.009 | 1.000 | 8.909 | 0.803 | 0.819 | 0.812 |
| 2 | 1737451 | 38 | 0.021 | 0.313 | 0.016 | 0.016 | 0.017 | 1.000 | 14.905 | 0.762 | 0.762 | 0.810 |
| 3 | 1836401 | 15 | 0.020 | 1.219 | 0.016 | 0.024 | 0.017 | 1.000 | 60.950 | 0.800 | 1.200 | 0.850 |
| 4 | 1836402 | 26 | 0.003 | 0.174 | 0.003 | 0.008 | 0.006 | 1.000 | 54.803 | 1.091 | 2.437 | 1.808 |
| 5 | 2130401 | 10 | 6.276 | 3.533 | 5.067 | 2.908 | 3.212 | 1.000 | 0.563 | 0.807 | 0.463 | 0.512 |
| 6 | 2130422 | 30 | 0.162 | 0.374 | 0.029 | 0.029 | 0.050 | 1.000 | 2.309 | 0.179 | 0.364 | 0.309 |
| 7 | 2235401 | 25 | 0.014 | 0.516 | 0.014 | 0.017 | 0.015 | 1.000 | 36.857 | 1.000 | 1.214 | 1.071 |
| 8 | 2237471 | 42 | 0.015 | 2.803 | 0.016 | 0.033 | 0.025 | 1.000 | 186.867 | 1.067 | 2.200 | 1.667 |
| 9 | 2527411 | 37 | 0.003 | 0.045 | 0.003 | 0.003 | 0.003 | 1.000 | 13.649 | 1.019 | 1.021 | 1.039 |
| 10 | 2528414 | 37 | 0.034 | 10.203 | 0.024 | 0.095 | 0.037 | 1.000 | 300.088 | 0.706 | 2.794 | 1.088 |

Berdasarkan Jadual 3 didapati statistik AD memberikan ϕ lebih besar dari 1 untuk 5 stesen, statistik SW untuk 3 stesen dan MRK untuk 2 stesen. Ini menunjukkan kaedah MRK adalah kaedah yang terbaik diikuti oleh statistik SW dan AD dalam menganggar parameter penjelmaan Box-Cox untuk digunakan dalam analisis frekuensi banjir.

6 Kesimpulan

Ketepatan menganggar magnitud banjir berpadanan tempoh ulangan berlakunya banjir ekstrim begitu penting kepada para hidrologi. Kaedah yang biasa dalam analisis frekuensi banjir adalah dengan memadankan sampel data banjir dengan taburan yang diketahui. Bagaimanapun, dalam situasi sebenar, fungsi taburan populasi sebenar tidak diketahui. Oleh itu satu alternatif penjelmaan Box-Cox dikaji dan dinilai dengan taburan GEV yang merupakan taburan yang terkenal dalam memodelkan data aliran banjir.

Data aliran tahunan maksimum dari negeri Johor dipilih sebagai data kajian. Analisis kajian terhadap data aliran tahunan maksimum di Johor, didapati sebahagian daripada data kajian tertabur secara normal berdasarkan pengujian statistik Shipro-Wilk. Tiga kaedah pengujian statistik Anderson Darling, Shipro-Wilk dan min ralat kuasa dua digunakan untuk menentukan parameter Box-Cox λ yang optimum dan penjelmaan data tertabur secara normal. Selepas penjelmaan Box-Cox dilakukan didapati kesemua data didapati tertabur secara normal .

Taburan normal tanpa penjelmaan Box-Cox tidak sesuai digunakan untuk memodelkan data banjir walaupun ianya didapati tertabur secara normal. Hasil kajian perbandingan antara tiga kaedah, kaedah pengujian statistik berdasarkan MRK didapati adalah kaedah yang terbaik berbanding statistik SW dan AD dalam memodelkan penjelmaan Box-Cox taburan normal. Walaubagaimanapun secara keseluruhannya didapati kesemua kaedah penjelmaan Box-Cox taburan Normal didapati menghasilkan keputusan yang hampir sama dan ianya didapati setanding dengan taburan GEV. Ini menunjukkan penjelmaan Box-Cox taburan normal boleh dijadikan salah satu taburan alternatif yang sesuai digunakan dalam analisis frekuensi walaupun taburan sebenar populasi tak diketahui.

Rujukan

- [1] M.I. Ahmad, C.D. Sinclair & A. Werritty, *Log-Logistic flood frequency analysis*, Journal of Hydrology, 98(1986), 205-224.
- [2] Z.D. Bai & L. Chen, *Weighted W Test for Normality and Asymptotics: A Revisit of Chen-Shapiro Test for Normality*, Journal of Statistical Planning and Inference, 113(2003), 485-503.
- [3] J. Freeman & R. Modarres, *Efficiency of test for independence after Box-Cox Transformation*, Journal of Multivariate Analysis, 2004.
- [4] S. Grimaldi, *Linear Parametric Models Applied to Daily Hydrological Series*, Journal of Hydrologic Engineering, 9(2004), 383-391.
- [5] T. Haktanir, *Comparisons of Various Flood Frequency Distributions using Annual Flood peaks Data of Rivers in Anatolia*, Journal of Hydrology, 136(1992), 1-31.
- [6] J.R.M. Hosking, *L-Moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics*, J.R. Stat. Soc., Ser. B52(2), 1990, 105-124.
- [7] E.S. Martins & J.R. Stedinger, *Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrolic Data*, Water Resources Research, 36(3), 737-744, 2000.

- [8] B. Onoz & M. Baryazit, *Best-fit Distribution of Largest Available Flood Samples*, Journal of Hydrology, 167(1995), 195-208.
- [9] H.H. Park, *Analysis and Prediction of Walleye Pollock (Theragra Chalcogramma) Landings in Korea by Time Series Analysis*, Fisheries Research, 38(1998), 1-7.
- [10] A.R. Rao & H.H. Hamed *Flood Frequency Analysis*, CRC Press, 2000.
- [11] J.A. Raynal & J.D.Salas, *Estimation Procedures For The Type 1 Extreme Value Distribution*, Journal of Hydrology, 87(1986), 315-336.
- [12] J.J. Spitzer, *A Monte Carlo Investigation of the Box-Cox Transformation in Small Samples*, Journal of the American Statistical Association, 73(1978), 488-495.
- [13] M. Thyer, G. Kuczera & Q.J. Wang, *Quantifying Parameter Uncertainty in Stochastic Models Using The Box-Cox Transformation*, Journal of Hydrology, 265(2002), 246-257.
- [14] R.M. Vogel, T.A. McMahon & F.H.S. Chiew, *Floodflow frequency model selection in Australia*, Journal of Hydrology 146(1993), 421-449.
- [15] R.M. Vogel, Jr. W.O. Thomas, & T.A. McMahon, *Flood-flow frequency selection in Southwestern United States*, Journal Water Resources Planning Management, 119(3)(1993), 353-366.
- [16] J. Zhang & Y. Wu, *Likelihood-Ratio Tests for Normality*, Computational Statistics & Data Analysis 49(2005), 709-721.