

## PENGGUNAAN TABURAN PARETO UMUM DALAM MENGANALISIS NILAI EKSTRIM BANJIR MENGGUNAKAN SIRI ALIRAN PUNCAK MELEBIHI PARAS

ANI SHABRI\*

**Abstrak.** Siri banjir tahunan maksimum (*Annual Maximum*, AM) merupakan pendekatan yang begitu terkenal dalam analisis frekuensi banjir. Siri puncak melebihi paras (*peaks over threshold*, POT) telah digunakan sebagai alternatif kepada siri banjir tahunan maksimum. Masalah utama dalam pendekatan POT adalah berkaitan pemilihan paras yang sesuai. Dalam kajian ini, kesan perubahan paras bagi siri POT ke atas nilai anggaran dikaji. Model POT dengan andaian bahawa bilangan puncak melebihi paras bertabur secara Poisson dan magnitud puncak melebihi paras tertabur secara Pareto Umum (*General Pareto Distribution*, GPD) dibincangkan. Parameter taburan GPD dianggar menggunakan kaedah kebarangkalian pemberat momen (*Probability Weighted Moment*, PWM) untuk paras yang diketahui. Perbandingan kesesuaian model POT dan model AM dalam menganggarkan nilai hujung atas taburan dibuat. Hasil kajian menunjukkan bahawa apabila paras siri POT boleh disesuaikan oleh taburan Pareto dengan proses Poisson, model POT didapati dapat menghasilkan anggaran nilai hujung atas taburan lebih baik berbanding model aliran maksimum.

**Kata kunci:** Siri puncak melebihi paras, proses poisson, taburan pareto umum, GEV, hujung atas taburan

**Abstract.** Annual maximum flood series remains the most popular approach to flood frequency analysis. Peaks over threshold series have been used as an alternative to annual maximum series. One specific difficulty of the POT approach is the selection of the threshold level. In this study the effect of raising the threshold of the POT series on heavy-tailed distributions estimation is investigated. The POT model described by the generalized Pareto distribution for peak magnitudes with the Poisson process for the occurrence of peaks is discussed. Estimation of the GPD parameters by the method of probability weighted moment (PWM) is formulated for known thresholds. A comparison of the efficiencies of the POT and AM models in heavy-tailed distributions is made. The result showed that when the threshold of POT series can be fitted by GPD with the Poisson process, the POT model is more efficient than the annual maximum (AM) model in estimating the highest extreme value.

**Keywords:** Peaks over threshold, poisson process, pareto distribution, GEV, heavy-tailed distributions

### 1.0 Pengenalan

Pengetahuan mendalam terhadap kejadian ekstrim seperti banjir pada masa hadapan begitu penting kepada para jurutera dalam merancang dan mereka bentuk kerja-kerja hidraulik seperti dalam mereka bentuk jambatan, benteng, pembentung, struktur kawalan banjir dan sebagainya [5]. Analisis frekuensi banjir merupakan satu kaedah

\* Jabatan Matematik, Fakulti Sains UTM, 81310 Skudai, Johor. e-mail: ani\_sabri@hotmail.com

yang sering digunakan oleh jurutera untuk meramal banjir pada tempoh ulangan yang tertentu dengan menggunakan taburan kebarangkalian.

Analisis frekuensi banjir selalunya dilakukan dengan memadankan taburan kebarangkalian dengan sampel aliran tahunan maksimum. Sampel aliran tahunan maksimum terdiri daripada siri aliran maksimum setiap tahun iaitu hanya satu cerapan bagi setiap tahun. Kaedah alternatif terdiri daripada kesemua puncak siri masa yang melebihi paras yang diberi. Kelebihan kaedah ini adalah bilangan cerapan bergantung kepada paras yang ditentukan iaitu nilai cerapan tidak semestinya satu untuk setiap tahun.

Masalah utama dalam menggunakan kaedah POT adalah menentukan nilai paras bagi menghasilkan model statistik yang sesuai. Sehingga kini tiada satu garis panduan ataupun ujian statistik dalam memilih paras yang sesuai, walaupun banyak penyelidikan telah dilakukan. Banyak kajian-kajian yang lepas mengatakan bahawa paras yang dipilih sepatutnya cukup tinggi supaya bilangan berlakunya banjir tahunan tertabur secara Poisson dan magnitud banjir melebihi paras tertabur secara Eksponen [3 dan 6].

Cunnane [3] dan Van Montfort [10] telah membuat kajian dalam pemilihan paras dan mendapati bahawa taburan eksponen tidak selalunya sesuai dalam memodelkan magnitud banjir.

Kajian-kajian terkini banyak memberi tumpuan terhadap taburan Pareto umum dalam memodelkan aliran banjir melebihi paras, dengan taburan eksponen adalah kes khusus bagi taburan GPD [1, 3, 4, 6, 7, 8, 9 dan 10].

Banyak kajian mendapati bahawa pemilihan taburan Poisson bagi pemilihan paras yang sesuai didapati adalah sukar dan taburan-taburan lain seperti taburan binomial negatif dan taburan binomial perlu juga dipertimbangkan [7].

Kajian yang dilakukan oleh Birikundavyi dan Rousselle [1] terhadap 66 stesen sungai di Kanada mendapati bahawa paras yang terbaik bagi memenuhi proses Poisson adalah paras yang menghasilkan min bilangan banjir tahunan di sekitar 1.7 ( $\lambda \approx 1.7$ ).

Objektif kajian ini bertujuan untuk menentukan paras aliran yang terbaik bagi menganggarkan magnitud banjir. Taburan GPD dan gabungan antara taburan GPD dengan proses Poisson akan dibincangkan. Parameter bagi taburan GPD dianggar menggunakan PWM. Kesan kenaikan paras bagi siri aliran banjir melebihi paras juga dipertimbangkan. Dalam kes ini kajian yang dilakukan oleh Birikundavyi dan Rousselle [1] dengan  $\lambda \approx 1.7$  juga dikaji sama ada ia boleh dijadikan panduan dalam memilih paras yang sesuai. Untuk menentukan kaedah yang menghasilkan anggaran kuantil terbaik, model aliran maksimum taburan GEV digunakan sebagai perbandingan.

## 2.0 TABURAN KEBARANGKALIAN

### 2.1 Taburan Kebarangkalian Pareto Umum

Taburan Pareto umum boleh ditulis sebagai

$$G(x) = \begin{cases} 1 - \left(1 - \frac{k}{\alpha} x\right)^{1/k}, & k \neq 0 \\ 1 - \exp\left(-\frac{x}{\alpha}\right), & k = 0 \end{cases} \quad (1)$$

dengan  $\alpha$  dan  $k$  adalah parameter skala dan bentuk masing-masing [4]. Bagi  $k = 0$ , taburan GPD menjadi taburan eksponen manakala bagi  $k > 0$  dan  $-\infty < x < \infty$  taburan GPD menjadi taburan Pareto.

Wang [11] mencadangkan bahawa nilai anggaran banjir pada  $T$  tahun boleh diperolehi berdasarkan persamaan (1) dan ditambah dengan nilai paras  $X_0$ . Nilai anggaran ini tidak mempertimbangkan sebarang taburan untuk mewakili bilangan banjir melebihi paras. Nilai anggaran ini boleh ditulis sebagai

$$x_T = X_0 + \frac{\alpha}{k} (1 - (1 - G)^k), \quad k \neq 0 \quad (2a)$$

$$x_T = X_0 + \alpha \log(1 - G), \quad k = 0 \quad (2b)$$

dengan  $G$  boleh diungkapkan sebagai fungsi bagi tempoh ulangan  $T$  dan ianya boleh ditulis sebagai

$$G = 1 - \frac{1}{T} \quad (3)$$

## 2.2 Gabungan antara Taburan Poisson dan Taburan Pareto Umum (PD-GPD)

Fungsi taburan kebarangkalian  $F(x)$  boleh diperolehi menggunakan teorem kebarangkalian keseluruhan (*total probability theorem*) dan ia boleh ditulis sebagai [7]

$$F(x) = \sum_{j=0}^{\infty} P(v = j)(G(x))^j \quad (4)$$

dengan andaian bahawa bilangan berlakunya banjir tahunan melebihi paras dalam selang masa  $(0, t)$  tertabur secara Poisson [1]:

$$P(v) = \frac{e^{-\lambda t} (\lambda t)^v}{v!}, \quad v = 1, 2, 3, \dots \quad (5)$$

dengan min dan varians diberikan oleh  $\lambda t$ . Dalam kes ini, andaian bahawa nilai magnitud puncak aliran banjir melebihi paras adalah tertabur dengan taburan GPD.

Gabungan antara taburan Poisson dan taburan GPD berdasarkan persamaan (4) menghasilkan ungkapan

$$F(x) = \text{eksp}[-\lambda t(1 - G(x))] \quad (6)$$

Dengan menetapkan  $t = 1$  dan memasukkan nilai  $G(x)$  ke dalam persamaan (6), maka yang berikut diperoleh

$$F(x) = \text{eksp}\left(-\lambda\left(1 - \frac{k}{\alpha}x\right)^{1/k}\right) \quad (7)$$

Berdasarkan persamaan (7), anggaran banjir pada  $T$  tahun setelah ditambah dengan nilai paras  $X_0$  boleh ditulis sebagai

$$x_T = X_0 + \frac{\alpha}{k}\left(1 - \left(-\frac{\ln F}{\lambda}\right)^k\right) \quad (8)$$

dengan  $\lambda$  adalah min bilangan puncak melebihi paras dalam setahun dan  $F$  boleh diungkapkan sebagai fungsi bagi tempoh ulangan  $T$  iaitu  $F = 1-1/T$ .

### 2.3 Anggaran Parameter

Terdapat dua cara bagi menganggar parameter taburan GPD menggunakan data aliran melebihi paras iaitu sama ada nilai puncak ditetapkan atau tidak. Jika paras  $X_0$  ditetapkan, bilangan data dan magnitud aliran banjir digunakan untuk menganggarkan  $\alpha$  dan  $k$ . Ungkapan bagi parameter  $\lambda$ ,  $\alpha$  dan  $k$  bagi taburan GPD dalam sebutan PWM diberikan oleh [11]

$$k = \frac{4\beta_1 - 3\beta_0 + X_0}{\beta_0 - 2\beta_1}, \quad \alpha = (\beta_0 - X_0)(1 + k) \quad (9)$$

dengan  $\beta_j = \int_0^1 X(F)F^j dF$  adalah PWM,  $X(F)$  adalah kuantil banjir atau song-sangan

bagi  $F(X)$  dan  $F$  adalah fungsi taburan GPD.

Jika paras  $X_0$  tidak ditetapkan, parameter  $X_0$ ,  $\alpha$  dan  $k$  mestilah dianggarkan bersamasama. Ungkapan untuk parameter ini dalam sebutan PWM diberikan oleh

$$k = \frac{9\beta_2 - 10\beta_1 + 2\beta_0}{2\beta_1 - 3\beta_2} \quad (10a)$$

$$\alpha = (2\beta_1 - \beta_0)(k + 1)(k + 2) \quad (10b)$$

$$X_0 = \beta_0 - \frac{\alpha}{k + 1} \quad (10c)$$

Bagi sampel rawak bersaiz  $m$ , anggaran saksama bagi  $\beta_j$  diperolehi daripada susunan tertib sampel  $x_1 \leq x_2 \leq \dots \leq x_m$  menggunakan statistik berikut.

$$\hat{\beta}_j = \sum_{i=1}^m \frac{(i-1)(i-2)\dots(i-j)}{(m-1)(m-2)\dots(m-j)} x_i \quad (11)$$

Statistik ini digunakan dalam persamaan (9) dan (10) untuk mendapatkan anggaran parameter bagi taburan GPD.

### 3.0 Pengujian Kesesuaian Taburan Poisson

Ujian  $\chi^2$  digunakan untuk menentukan kesesuaian model taburan Poisson dengan bilangan puncak melebihi paras. Cunnane [3] mencadangkan penggunaan indeks ubahan yang ditakrifkan sebagai:

$$I = \frac{V}{E} \quad (12)$$

dengan  $E$  dan  $V$  adalah masing-masing min dan varians bilangan puncak melebihi paras dalam  $N$  tahun. Ujian statistik yang berpadanan kepada indeks ubahan adalah

$$d = \frac{(N-1)V}{E} \quad (13)$$

Statistik  $d$  tertabur secara taburan  $\chi^2$  dengan  $N-1$  darjah kebebasan. Taburan Poisson ditolak sekiranya nilai  $d$  kiraan berada di luar julat  $(\chi_{\alpha/2}^2, \chi_{1-\alpha/2}^2)$  dengan aras keertian  $\alpha$  [7].

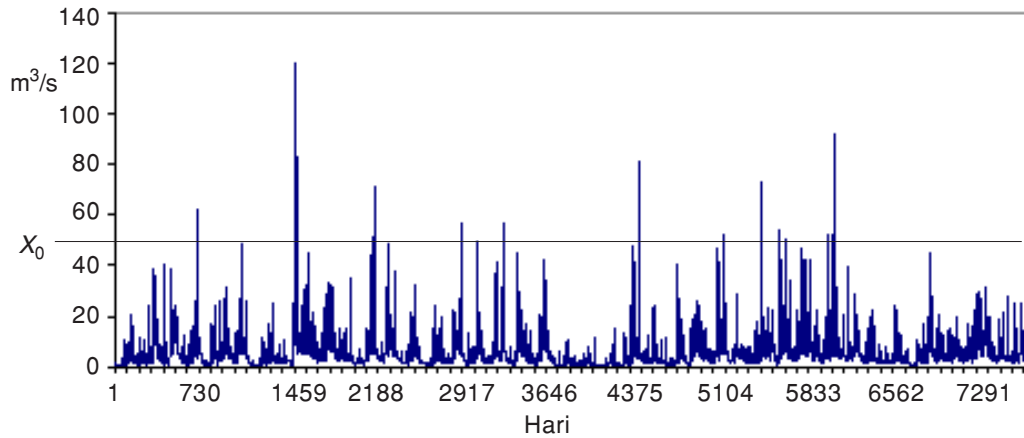
### 4.0 Analisis Kajian

Data aliran harian sungai Melaka digunakan sebagai sampel kajian ini. Data ini diperolehi daripada Jabatan Parit dan Saliran di Ampang, Selangor. Saiz sampel adalah 21 tahun (1980-2000) dengan nilai minimum, maksimum, min dan varians ketinggian adalah  $0.1 \text{ m}^3/\text{s}$ ,  $119.9 \text{ m}^3/\text{s}$ ,  $6.389 \text{ m}^3/\text{s}$  dan  $46.772 \text{ m}^3/\text{s}$  masing-masing. Bentuk siri masa data kajian dan paras ditunjukkan dalam Rajah 1.

Bagi menentukan kesan perubahan paras terhadap kesesuaian data bagi taburan GPD dan proses Poisson, sebanyak 10 paras dipilih. Ujian indeks ubahan digunakan untuk menguji kesesuaian model Poisson manakala ujian Kolmogorov-Smirnov (K-S) digunakan untuk menguji kesesuaian model GPD, dan PD-GPD.

Sebagai contoh pengiraan, bagi sungai Melaka pada paras  $36 \text{ m}^3/\text{s}$ , didapati nilai min,  $E = 3.143$  dan varians,  $V = 6.629$ . Ujian indeks ubahan memberikan nilai

$$d = \frac{(N-1)V}{E} = \frac{(21-1)(6.629)}{3.143} = 42.182$$



**Rajah 1** Plot siri masa bagi data aliran Sungai Melaka (1980-2000)

Hipotesis bagi proses bilangan cerapan melebihi paras tertabur secara taburan Poisson ditolak disebabkan nilai  $d$  berada di luar nilai kritikal  $\chi_{1-\alpha/2}^2 7.434$  dan  $\chi_{\alpha/2}^2 = 39.997$  pada aras keertian 1% dan darjah kebebasan 20.

Ujian K-S boleh ditulis sebagai

$$D = \text{maks} |F(x_i) - F_0(x_i)|$$

dengan  $F(x_i)$  adalah fungsi ketumpatan kumulatif bagi taburan GPD atau PD-GPD dan  $F_0(x_i) = \frac{i}{n} \cdot F_0(x_i)$  adalah kedudukan memplot,  $n$  adalah saiz sampel dan  $i$  adalah susunan tertib dengan  $i = 1$  menunjukkan nilai sampel terkecil.

Menggunakan paras yang sama, anggaran parameter bagi taburan GPD menggunakan kaedah PWM adalah  $\xi = 36$ ,  $\alpha = 9.791$  dan  $k = -0.226$ . Ujian K-S memberikan nilai  $D = 0.088$  dan nilai kritikal  $D_{0.05, n=66} = 0.201$ . Nilai  $D$  kiraan didapati kurang daripada nilai kritikal  $D_{0.05, n=66}$ . Ini menunjukkan bahawa hipotesis taburan GPD sesuai untuk dipadankan dengan data.

Manakala ujian K-S terhadap taburan PD-GPD memberikan nilai  $D = 0.359$ . Taburan PD-GPD ditolak disebabkan nilai  $D$  kiraan melebihi nilai kritikal pada aras keertian 1%.

Hasil pengiraan yang sama untuk nilai paras yang berlainan dapat ditunjukkan dalam Jadual 1.

Jadual 1 menunjukkan bahawa tinggi paras berkadar songsang dengan bilangan puncak melebihi paras, iaitu semakin meningkat paras bilangan puncak melebihi paras semakin berkurangan. Hasil ujian statistik khi-kuasa dua menunjukkan bahawa bilangan data melebihi paras tertabur secara Poisson iaitu pada paras,  $X_0$  melebihi 42  $\text{m}^3/\text{s}$ . Ujian Kolmogorov-Smirnov menunjukkan bahawa hampir kesemua data tertabur secara GPD kecuali pada paras 60  $\text{m}^3/\text{s}$  dan 70  $\text{m}^3/\text{s}$ . Taburan PD-GPD didapati sesuai pada paras 42  $\text{m}^3/\text{s}$ , 44  $\text{m}^3/\text{s}$  dan 48  $\text{m}^3/\text{s}$  pada aras keertian 1%.

**Jadual 1** Ciri-ciri data aliran di atas puncak. Hasil pengujian khi-kuasa dua dan ujian Kolmogorov-Smirnov

No	Paras $X_0(m^3/s)$	Bilangan Puncak,	Min $\lambda$	Ujian $\chi^2$ Poisson	Taburan GPD	Taburan PD-GPD
1	4	3252	813	P*	D	D*
2	18	376	17.9	P*	D	D*
3	26	149	7.1	P*	D	D*
4	30	106	5.0	P*	D	D*
5	36	66	3.1	P*	D	D*
6	42	35	1.7	P*	D	D*
7	44	31	1.5	P*	D	D*
8	48	23	1.1	P*	D	D*
9	60	11	0.5	P*	D	D*
10	70	7	0.3	P*	D	D*

P\* dan D\* ditolak pada aras keertian 1%

#### 4.1 Perbandingan Anggaran Nilai Banjir

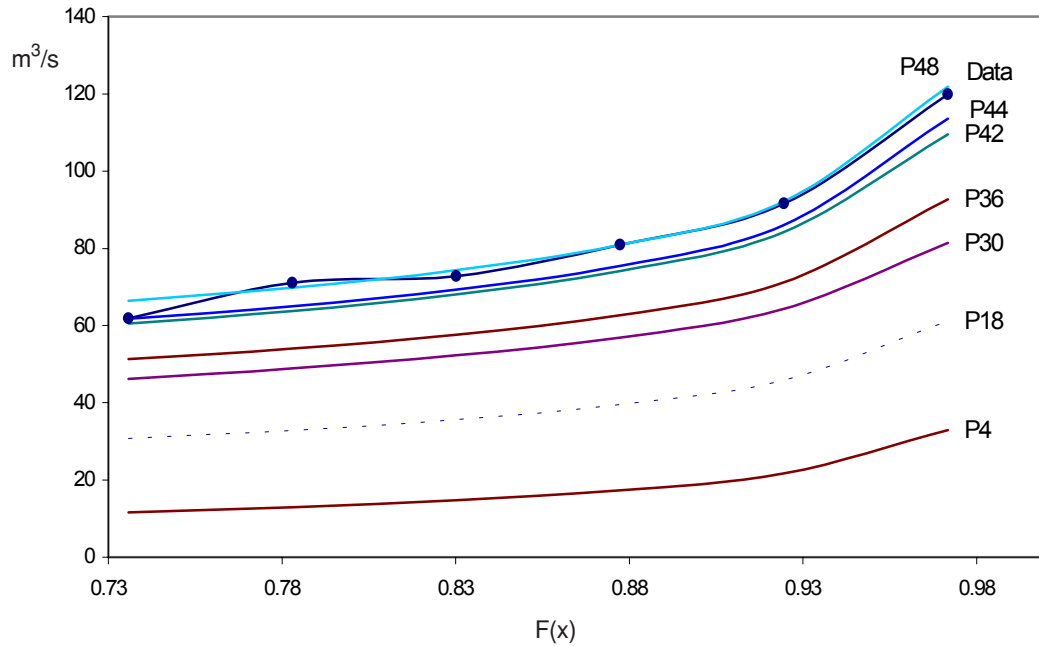
Pada paras aliran banjir yang dipilih, nilai anggaran banjir dikira berdasarkan kesesuaian model yang diuji. Berdasarkan Jadual 1, taburan GPD digunakan untuk menganggarkan banjir pada kesemua paras kecuali pada paras 60 m<sup>3</sup>/s dan 70 m<sup>3</sup>/s.

Rajah 2 menunjukkan perbandingan nilai anggaran banjir bagi taburan GPD menggunakan paras,  $X_0 = 18, 30, 36, 42, 44$  dan  $48 \text{ m}^3/\text{s}$  dan cerapan data. Dalam kajian ini hanya nilai hujung kanan taburan iaitu  $F(x) > 0.7$  iaitu bersamaan tempoh ulangan  $T > 3$  tahun dipertimbangkan. Tempoh ini begitu penting kepada jurutera dalam merancang dan mereka bentuk struktur hidrolik.

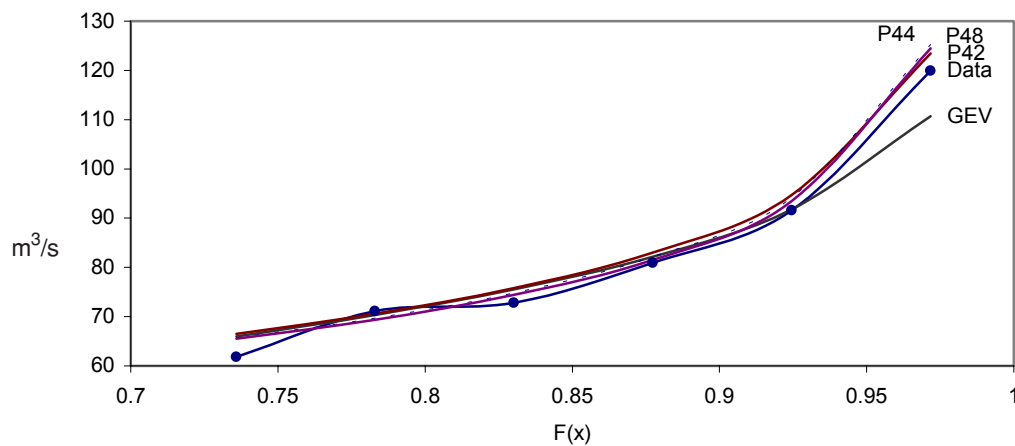
Berdasarkan Rajah 2, didapati nilai anggaran banjir menggunakan taburan GPD bersandar kepada nilai paras iaitu semakin tinggi paras semakin tinggi nilai anggaran yang diperolehi. Nilai paras yang menghasilkan anggaran yang terbaik yang menghampiri dengan cerapan data adalah nilai paras pada sekitar  $42 \text{ m}^3/\text{s}$  dan  $48 \text{ m}^3/\text{s}$ .

Perbandingan antara anggaran banjir menggunakan taburan PD-GPD, GEV dan cerapan data dapat ditunjukkan dalam Rajah 3. Taburan aliran tahunan maksimum taburan Ekstrim Umum (*General Extreme Value Distribution*, GEV) digunakan sebagai perbandingan. Taburan ini sering digunakan dalam hidrologi dan telah dijadikan sebagai taburan piawai di United Kingdom dan Perancis [2]. Parameter bagi taburan GEV dianggarkan menggunakan kaedah PWM berdasarkan data aliran tahunan maksimum sungai Melaka. Ujian kesesuaian Kolmogorov-Smirnov menunjukkan bahawa taburan GEV sesuai dipadankan dengan data pada aras keertian 1%.

Berdasarkan Rajah 3 didapati bahawa nilai anggaran banjir pada ketiga-tiga paras menggunakan taburan PD-GPD adalah hampir sama dan nilai anggarannya begitu hampir dengan data cerapan untuk kesemua nilai  $F(x)$  yang dipilih. Anggaran nilai



**Rajah 2** Perbandingan taburan GPA menggunakan paras berlainan



**Rajah 3** Perbandingan antara taburan PD-GPA dan GEV

banjir menggunakan taburan GEV didapati menghampiri cerapan untuk nilai  $F(x)$  yang kecil dan semakin berjauhan apabila nilai  $F(x)$  semakin bertambah.

## 5.0 KESIMPULAN

Dalam kajian ini taburan GPD dan gabungan antara taburan Poisson dan GPD digunakan dalam menganalisis data aliran sungai di negeri Melaka. Masalah utama



dalam menggunakan POT model adalah memilih paras yang optimum bagi menghasilkan nilai anggaran yang terbaik.

Hasil kajian menggunakan taburan GPD dan PD-GPD menunjukkan bahawa nilai paras terbaik adalah nilai paras pada sekitar  $42 \text{ m}^3/\text{s}$  dan  $48 \text{ m}^3/\text{s}$ . Nilai paras ini didapati cukup tinggi dan ia memenuhi syarat bahawa bilangan puncak melebihi paras tertabur secara Poisson dengan nilai puncak tertabur secara GPD. Hasil kajian ini adalah bertepatan dengan kajian yang dibuat oleh Wang [11] dan Madsen [6]. Jika dilihat paras  $42 \text{ m}^3/\text{s}$  dan  $44 \text{ m}^3/\text{s}$  adalah paras yang menghasilkan nilai  $\lambda$  di sekitar  $1.7 \text{ m}^3/\text{s}$ . Ia juga bertepatan dengan hasil kajian yang dibuat oleh Birikundavyi dan Rousselle [1].

Hasil kajian juga menunjukkan bahawa apabila paras yang dipilih terlalu rendah, ia tidak dapat memenuhi syarat taburan PD-GPD dan nilai anggaran banjir yang dianggarkan tidak sesuai. Manakala jika paras terlalu tinggi, hanya beberapa cerapan sahaja yang melebihi paras dan ia didapati tidak sesuai untuk di padankan dengan sebarang taburan.

Dalam kajian ini didapati pemilihan nilai paras memainkan peranan yang penting dalam menghasilkan anggaran banjir yang terbaik. Oleh itu kajian selanjutnya menggunakan lebih banyak data dan model-model lain perlu dilakukan bagi menentukan paras yang optimum.

## RUJUKAN

- [1] Birikundavyi, S. dan J. Rousselle. 1997. "Use of Partial Duration Series For Single-Station And Regional Analysis of Floods". *Journal of Hydrologic Engineering*. 2: 68-74.
- [2] Cunnane, C. 1989. Statistical Distributions For Flood Frequency Analysis. *Operational Hydrology Report No.33*. WMO-No. 718.
- [3] Cunnane, C. 1979. "A Note on the Poisson Assumption in Partial Duration Series Models". *Water Resour. Res.* 15(2): 489-494.
- [4] Heckert N. A., E. Simiu, Fellow, ASCE dan T. Whalen. 1998. "Estimates of Hurricane Wind Speeds By "Peaks Over Threshold" Method". *Journal of Structural Engineering*. 124: 445-449.
- [5] Lang, M., Ouarda, T. B. M. J., dan Bobee, B. 1999. "Towards Operational Guidelines for Over-Threshold Modelling". *Journal of Hydrology*. 225: 103-117.
- [6] Madsen H, P. F. Rasmussen, dan D. Rosbjerg. 1997. "Comparison of Annual Maximum Series and Partial Duration Series Methods For Modelling Extreme Hydrology Events. 1. At-Site Modelling". *Water Resour. Res.* 33(4): 747-757.
- [7] Onoz B., dan M. Bayazit. 2001. "Effect of The Occurrence Process of the Peaks Over Threshold on The Flood Estimates". *Journal of Hydrology*. 244: 86-96.
- [8] Rosbjerg D, H. Madsen, dan P. F. Rasmussen 1985. "Prediction in Partial Duration Series with Generalized Pareto Distributed Exceedances". *Water Resour. Res.* 28(11): 3001-301.
- [9] Soares C. G. dan M. Scotto. 2001. "Modelling Uncertainty in Long-Term Predictions of Significant Wave Height". *Ocean Engineering*. 28: 329-342.
- [10] Van Montfort, M. A. J., dan J. V. Witter. 1985. "Testing Exponentiality Against Generalized Pareto Distribution". *Journal of Hydrology*. 78: 305-315.
- [11] Wang, Q, J. 1991. "The POT Model Described by the Generalized Pareto Distribution with Poisson Arrival Rate". *Journal of Hydrology*. 129: 263-280.