UNIVERSITAS GADJAH MADA
DEPARTEMEN TEKNIK SIPIL DAN LINGKUNGAN
PRODI MAGISTER TEKNIK SIPIL

Statistika Teknik

Analisis Frekuensi

Analisis Frekuensi

- Diambilkan dari paparan pada
  - The Second ASEAN Natural Disaster Conference
    - University of Yangon, Myanmar
    - 29-30 September 2014

2nd ANDC

Yangon, Myanmar

29-30 September 2014

I s t i a r t o

Master of Engineering in Natural Disaster Management

Universitas Gadjah Mada, INDONESIA

Frequency analysis on extreme hydrologic data

Needs

- Engineers (hydraulics, hydrology) are frequently required to derive design values on hydrologic events, such as flow discharge or rainfall depth (precipitation)
  - Statistical method applied to time series data is a common practice
    - Fit those data to theoretical probability distributions
    - Use cumulative distribution function of the selected probability distribution to predict magnitude of hydrologic events
    - Define the design value(s) based on the magnitude of hydrologic event having particular probability of occurence

this method is known as frequency analysi

Frequency Analysis

Time series (annual or partial time series of hydrologic event data)

Fitting of the time series to theoretical distributions

Predicting probability of occurrence of the hydrologic events

Define design values of the hydrologic events

Probability Distributions

Common probability distributions applicable to hydrologic events

- Theoretical Probability Distributions
  - Gumbel
  - Log Normal
  - Log Pearson Type III
  - Normal

Probability Distributions

- The theoretical probability distributions
  - their pdf and cdf have complex expression so that they are not readily solvable
  - graphical method is a practical solution to fit data to the theoretical distributions

Ven te Chow *et al*. (1988) developed a transformation coordinate such that the cdf of a theoretical distribution shows as a straight line

→

fitting data to theoretical distribution and predicting probability of occurrence can be easily carried out

Transformed probability coordinate

- Chow *et al*. (1988)

$$y_T = \bar{y} + K_T \cdot s_Y$$

  - y_T is the hydrologic magnitude at *T*-year return period,
  - y is the average value of the data,
  - K_T is a frequency factor, and
  - s_Y is the standard deviation of the data.

Transformed probability coordinate

- Chow *et al.* (1988)

$$y_T = \bar{y} + K_T \cdot s$$

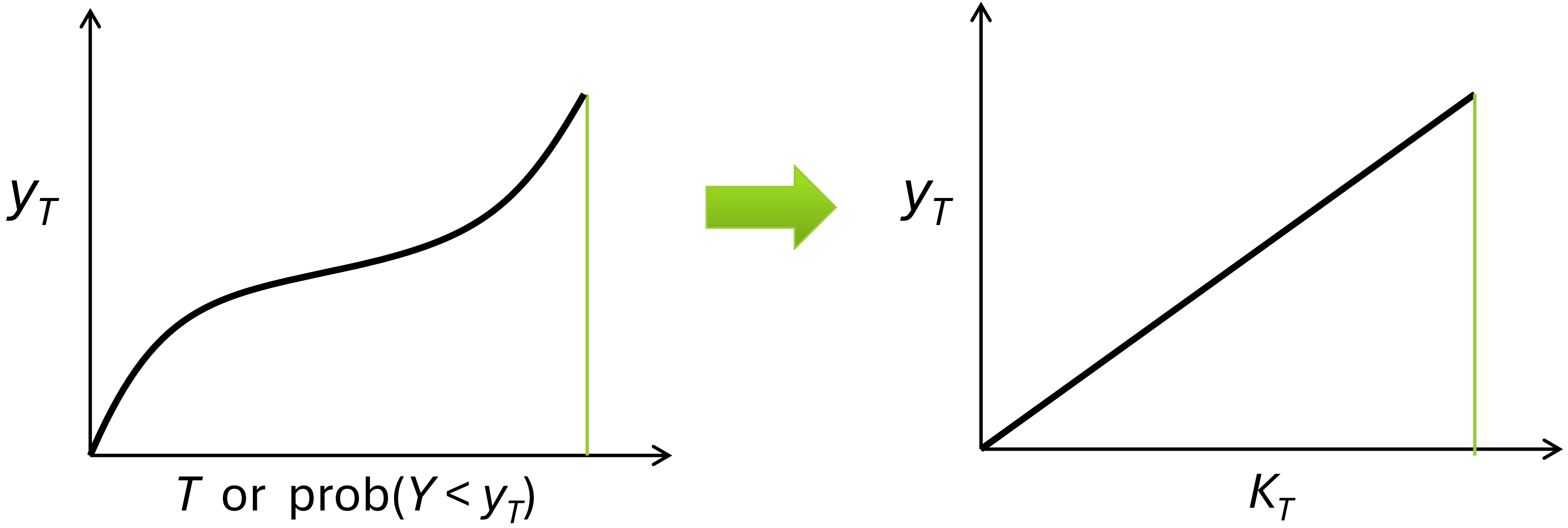
$\bar{y}$, s_Y } constant → plot of (y_T vs K_T) is a straight line

↓ defined by data

K_T is function of T and distribution

$$\text{prob}(Y < y_T) = 1 - \frac{1}{T} \Leftrightarrow T = \frac{1}{1 - \text{prob}(Y < y_T)}$$

Normal Distribution



Normal Distribution

$$K_T = z$$

$$z = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}$$

$$w = \begin{cases} \left[\ln\left(\frac{1}{p^2}\right)\right]^{\frac{1}{2}} & 0 < p = \frac{1}{T} < 0.5 \\ \left[\ln\left(\frac{1}{(1-p)^2}\right)\right]^{\frac{1}{2}} & p = \frac{1}{T} > 0.5 \end{cases}$$

Probability Paper

- Nowadays
  - Frequency analysis can easily be carried out by spreadsheet computer application
    - Unfortunately, spreadsheet cannot provide data plot the way probability paper does
    - Most of frequency analysis application program do not provide data plot at all
    - Lost of insight into data pattern (distribution)
    - Lost of ‘physical’ meaning of the computed values

Probability Paper

- In the past
  - Probability papers were available in the market
    - Normal, Log Normal, Gumbel distributions
    - Not available for Log Pearson Type III since K_T changes with data
      - Log Pearson Type III was plotted on Log Normal paper
      - It shows as a curve, not a straight line
  - Data were manually plotted on the paper
    - Scatter data plot and theoretical line give good insight into the distribution of the data
    - Fitting the data to the theoretical line can easily be performed

Probability Paper

- AProb_4E
  - A computer application on frequency analysis that has data plotting capability
  - Can be freely downloaded from website https://istiarto.staff.ugm.ac.id/

https://istiarto.staff.ugm.ac.id/index.php/aprob/

Probability Paper

- Nowadays
  - There is no (or at least difficult to find) computer application on frequency analysis having data plot capability
    - Commercial application program may exist (?)
    - Practising engineers and students might be not able to afford for commercial application program

→ Need for frequency analysis application program, which is a free-ware and capable to produce graphical plot of the data on probability paper

AProb_4E

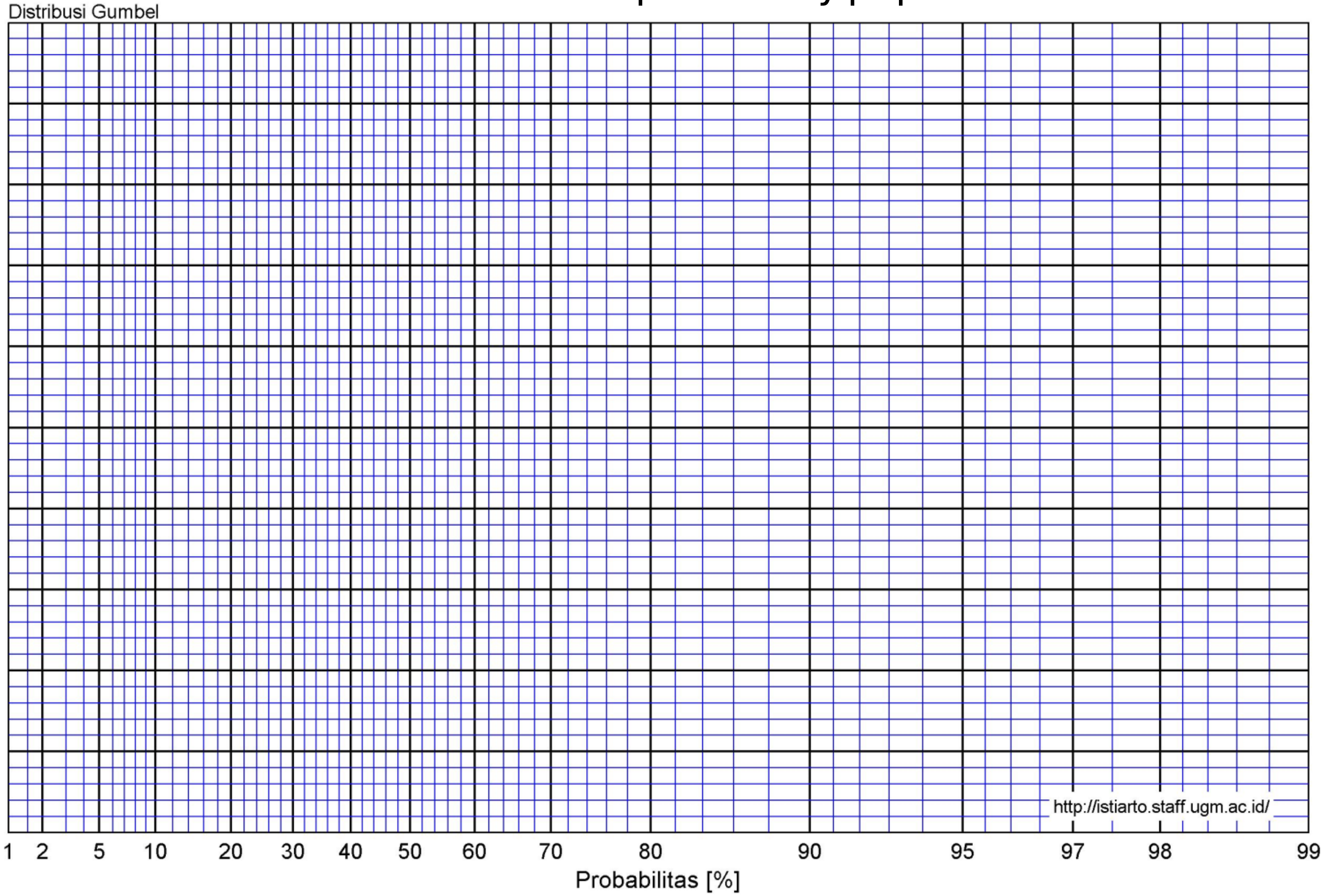
- Plotting position of data values on probability paper

$$\text{prob}(Y < y_m) = \frac{m}{n+1}$$

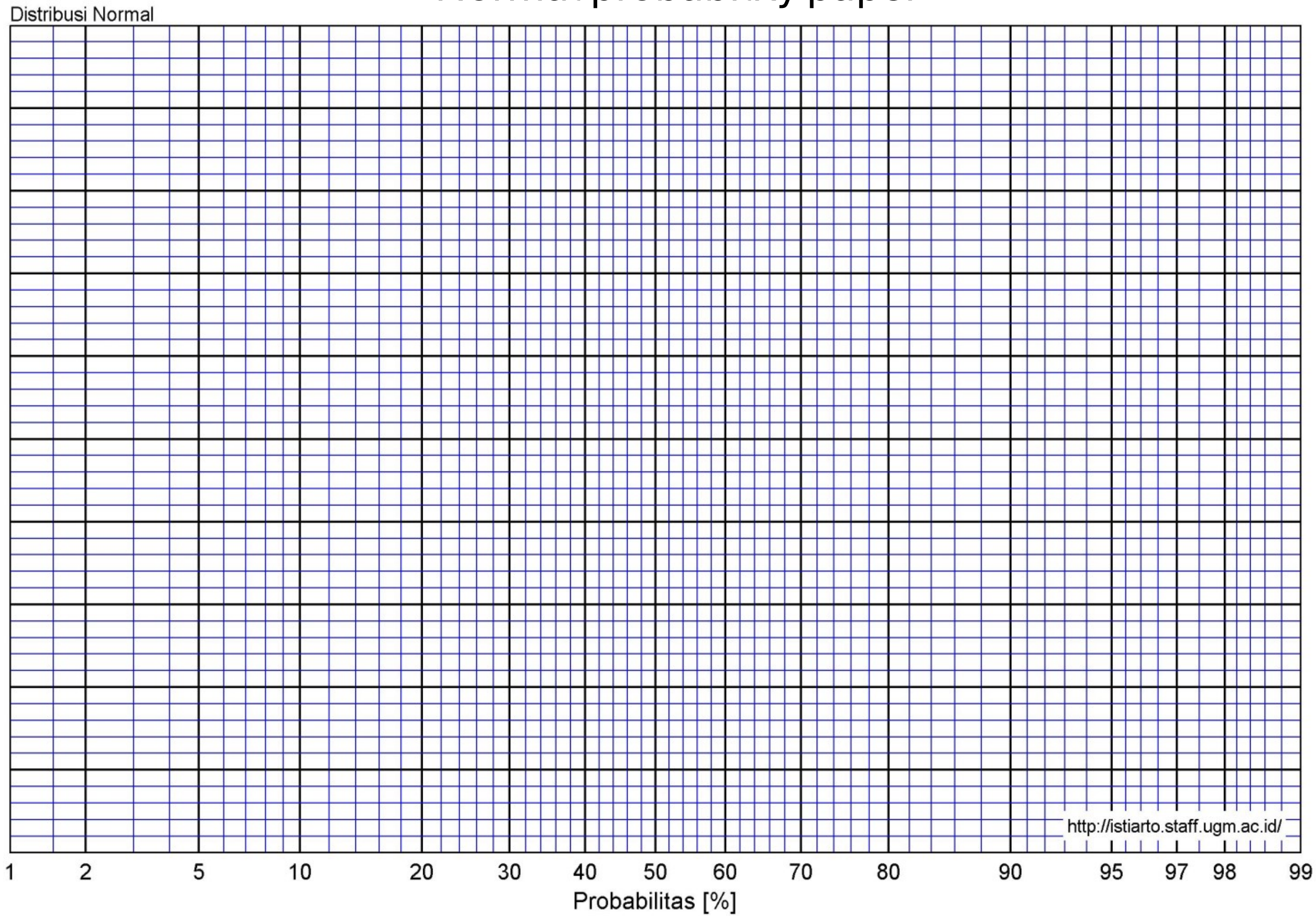
  - m is the rank of the data being sorted in ascending order
  - n is the number of data

- This is known as Weibull formula of plotting time series data

Gumbel probability paper



Normal probability paper



Judul Data :	**Debit Maksimum DPS Citarum**	
Cara Urut Data :	b	Ket. : B = urutan debit besar ke kecil ; K = kecil ke besar

UJI CHI-SQUARE

Jumlah kelas :	5	Ket. : Jumlah kelas yang dikehendaki untuk uji Chi-Kuadrat
onfidence Interval :	0.05	Ket. : Derajat Ketidak-percayaan yang diinginkan

KALA-ULANG

Jumlah kasus :	8	Ket. : Jumlah probabilitas yang dikehendaki

Tahun	Debit (m^3/dt)	Probabilitas
1918	244.00	.900
1919	217.00	.500
1920	285.00	.200
1921	261.00	.100
1922	295.00	.050
1923	252.00	.020
1924	275.00	.010
1925	204.00	.001
1926	208.00	
1927	194.00	
1928	256.00	
1929	207.00	
1930	354.00	
1931	445.00	
1932	350.00	
1933	336.00	
1934	328.00	
1973	269.00	
1974	323.00	

PROSES

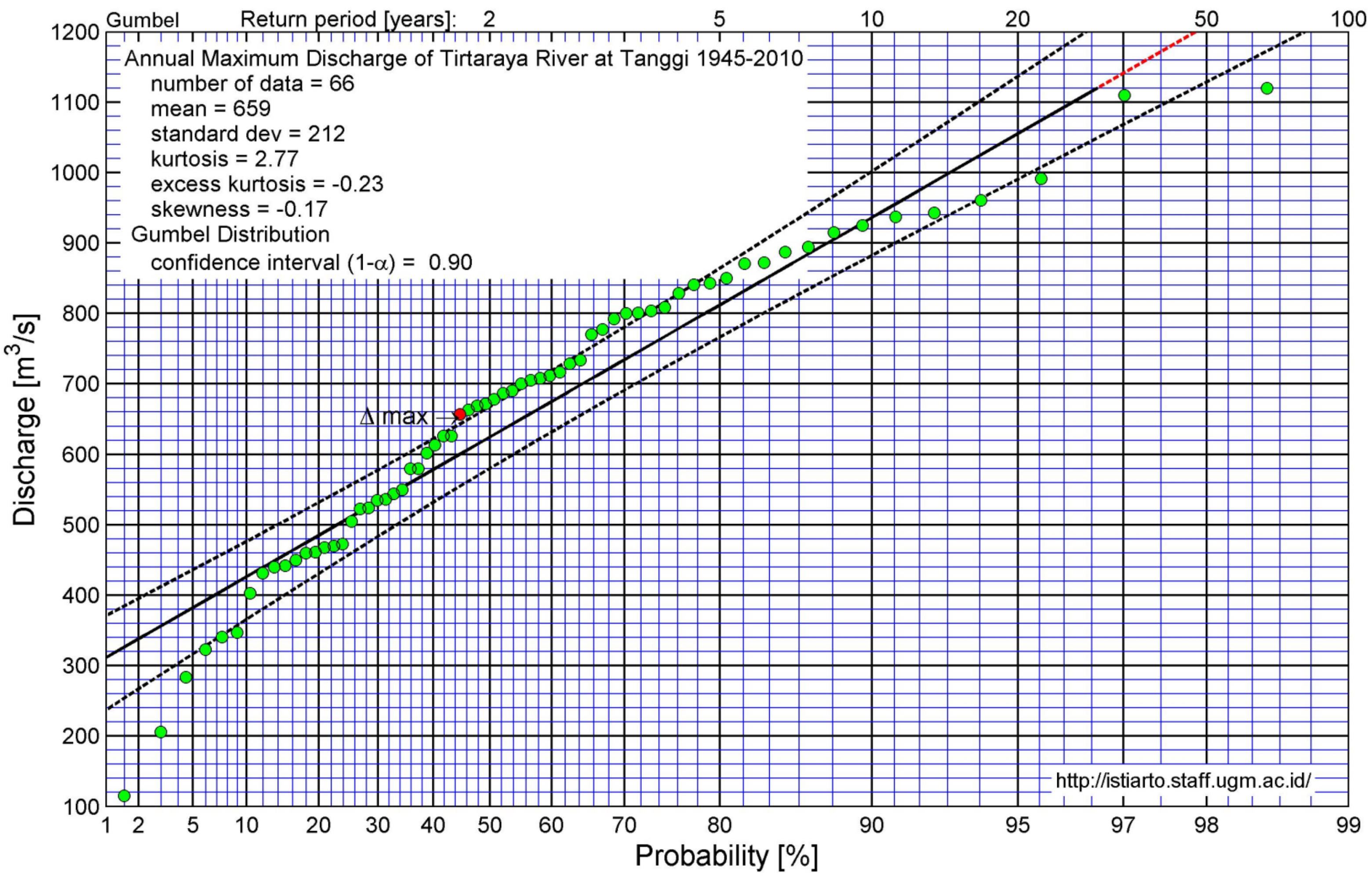
Example of frequency analysis performed by using spreadsheet application/computer program

KALA-ULANG Debit Maksimum DPS Citarum

P(x >= Xm)	T	Karakteristik Debit (m^3/dt) Menurut Probabilitasnya							
Probabilitas	Kala-Ulang	NORMAL		LOG-NORMAL		GUMBEL		LOG-PEARSON III	
		X_T	K_T	X_T	K_T	X_T	K_T	X_T	K_T
0.9	1.1	214.997	-1.282	219.417	-1.202	225.065	-1.100	219.250	-1.286
0.5	2.	286.200	0.000	281.092	-0.092	277.073	-0.164	281.433	0.006
0.2	5.	332.961	0.842	330.748	0.802	326.173	0.719	330.862	0.843
0.1	10.	357.403	1.282	360.103	1.330	358.682	1.305	359.818	1.277
0.05	20.	377.588	1.645	386.299	1.802	389.865	1.866	385.496	1.634
0.02	50.	400.306	2.054	418.069	2.373	430.228	2.592	416.438	2.034
0.01	100.	415.452	2.326	440.687	2.781	460.475	3.137	438.336	2.299
0.001	1,000.	457.894	3.090	510.803	4.043	560.420	4.936	505.553	3.037

Ket : 1. $X_T = \mu + K_T \cdot \sigma$
2. Menurut Uji Chi-Kuadrat, yang terbaik menggunakan distribusi NORMAL
3. Sedangkan menurut Uji Smirnov-Kolmogorov, yang terbaik menggunakan distribusi NORMAL
4. Hitungan dilakukan dengan menggunakan rumus dalam buku 'Applied Hidrology', 1988, Ven Te Chow, et. al.

Example of frequency analysis performed by using spreadsheet application/computer program



Gumbel
Return period [years]: 2 5 10 20 50 100
Annual Maximum Discharge of Tirtaraya River at Tanggi 1945-2010
number of data = 66
mean = 659
standard dev = 212
kurtosis = 2.77
excess kurtosis = -0.23
skewness = -0.17
Gumbel Distribution
confidence interval (1-α) = 0.90
Δ max
Discharge [m3/s]
1200 1100 1000 900 800 700 600 500 400 300 200 100
1 2 5 10 20 30 40 50 60 70 80 90 95 97 98 99
Probability [%]
http://istiarto.staff.ugm.ac.id/

Hidrologi | i s t i a r t o

istiarto.staff.ugm.ac.id/index.php/artikel/hidrologi/

SIA UGM MTPBA | internet JTSL UGM | Online hyd...iguel Ponce | What Is My ...s Since 1999 | KenRockwell...ter Pictures | OSU Stream...rial – Home | e-Procurement PLN

cPanel Login | Hidrologi | i s t i a r t o

i s t i a r t o

Jurusan Teknik Sipil dan Lingkungan
Fakultas Teknik UGM

Home | Courses | Kuliah | HEC-RAS | **Artikel** | Profil | Galeri | Kirim Pesan

Hidrologi

1. Analisis Frekuensi
   - Frequency Analysis on Hydrologic Data (AProb_4E), in English
   - Analisis Data Hidrologi Ekstrem (AProb_3, AProb_31)
   - Analisis Data Hidrologi Ekstrem (AProb_2, AProb_21)
   - Analisis Frekuensi Data Hidrologi Ekstrem
2. Memplotkan Data pada Kertas Probabilitas
   - Plot Data pada Kertas Probabilitas (PProb_4)

Cari di situs ini

Search

Terbaru

- Site Visit ke PT Bukit Asam Tanjungenim SumSel
- Penyelesaian Numeris Persamaan Konveksi-Difusi
- Penyelesaian Numeris Persamaan Difusi
- Kursus HEC-RAS Juli 2014

AProb_4E

- Confidence interval (1 – α)
  - An interval within which the true value (which is unkonwn) can reasonably be expected to lie.
  - The size of the interval depends on the confidence level (1 – α)
- The estimate of event value for a particular return period, y_T
  - An upper and lower limits are specified by adjustment of K_T

$$U_{T,\alpha} = \bar{Y} + s_Y \cdot K_{T,\alpha}^U$$

$$L_{T,\alpha} = \bar{Y} + s_Y \cdot K_{T,\alpha}^L$$

upper and lower confidence limit factors for normally distributed data; these are determined using the noncentral *t* distribution

AProb_4E

- Goodness of fit test
  - Smirnov-Kolmogorov test
  - Chi-square test

  - Both tests are applied with confidence level of $(1 – \alpha) = 0.90$

Goodness of Fit Test

- Smirnov-Kolmogorov Test

$$\Delta_{max} = \max\left|\text{prob}(Y < y) - \underline{\text{prob}}(Y < y)\right|$$

→ rejected if $\Delta_{max} < D_c$

D_c is critical value according to the Smirnov-Kolmogorov table

- according to the observed data
- according to the distribution being tested

Goodness of Fit Test

- Chi-square Test

$$\chi_c^2 = \sum_{i=1}^{k} \left(\frac{O_i - E_i}{E_i} \right)$$

→ rejected if $\chi_c^2 > \chi^2_{1-\alpha/2,k-p-1}$

p is the number of parameters estimated from the data

O_i the observed relative frequency in the *i*th class interval
E_i the expected relative frequency, according to the distribution being tested, in the *i*th class interval
k the number of class intervals

Goodness of Fit Test

- Chi-square Test

$$\chi_c^2 = \sum_{i=1}^{k} \frac{n[f_s(x_i) - p(x_i)]^2}{p(x_i)}$$

→ rejected if $\chi_c^2 > \chi_{1-\alpha/2,k-p-1}^2$

p is the number of parameters estimated from the data

k is the number of class intervals

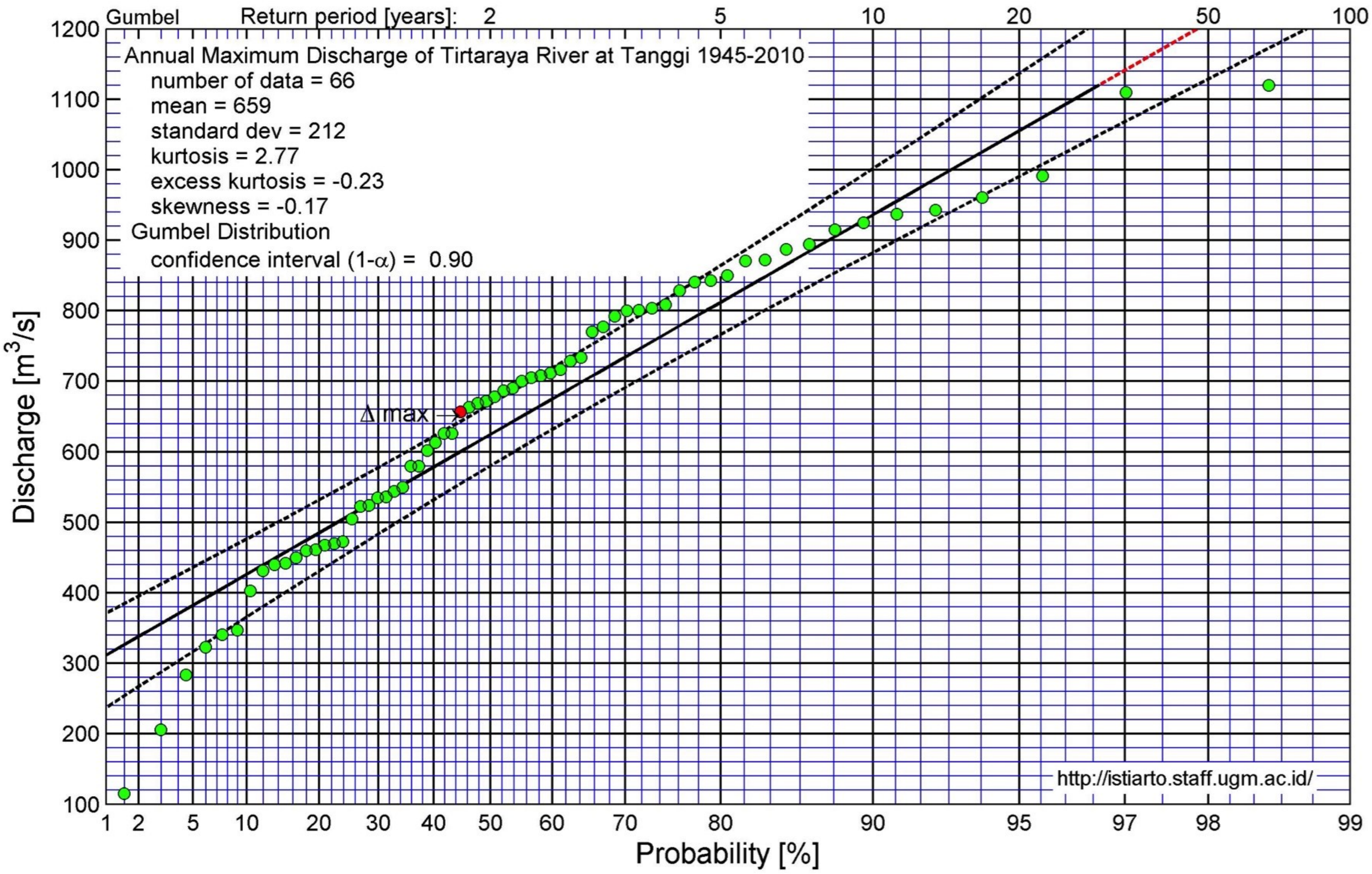
theroretical probability according to distribution being tested

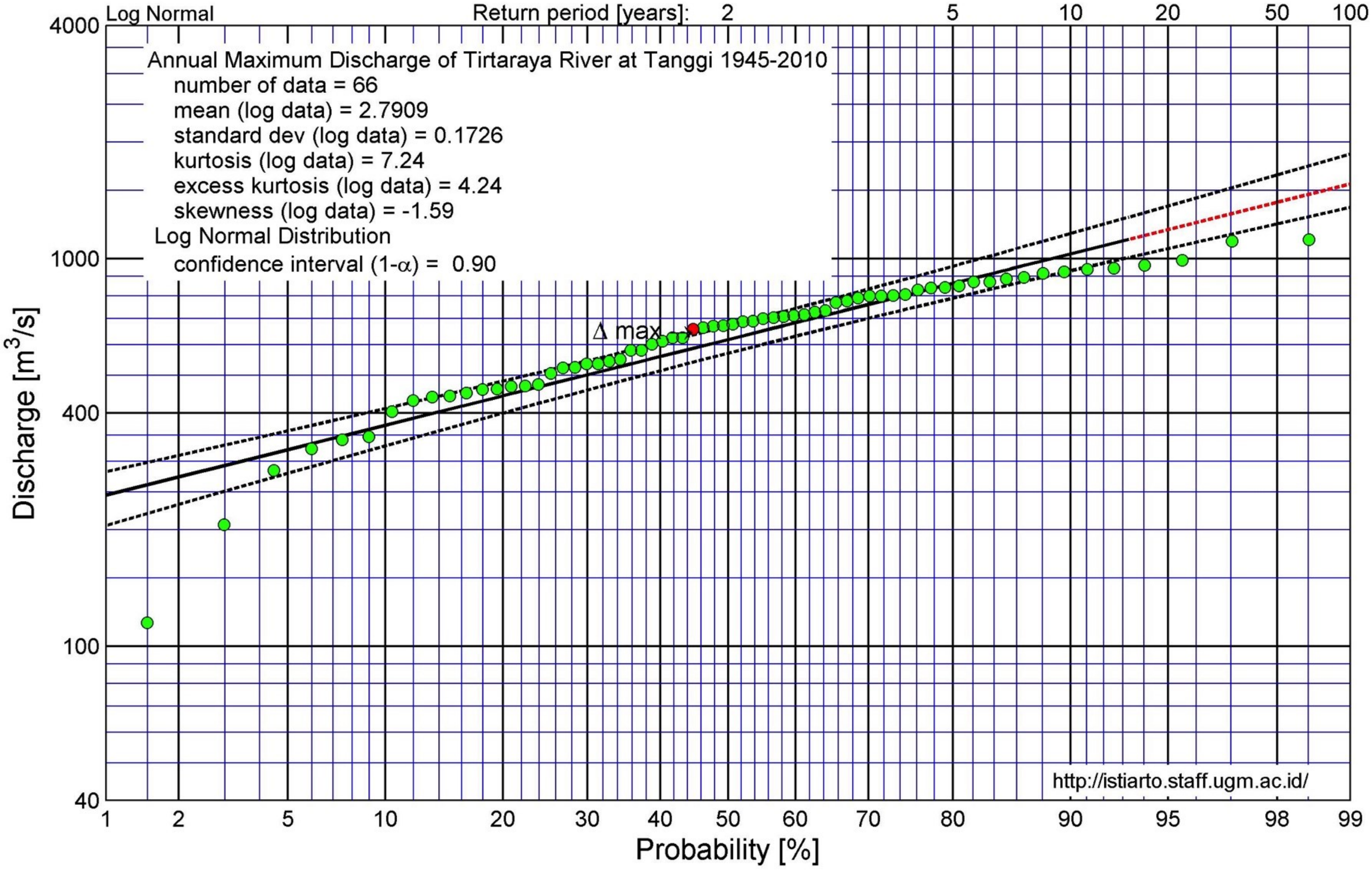
observed relative frequency

Result of AProb_4E in text format

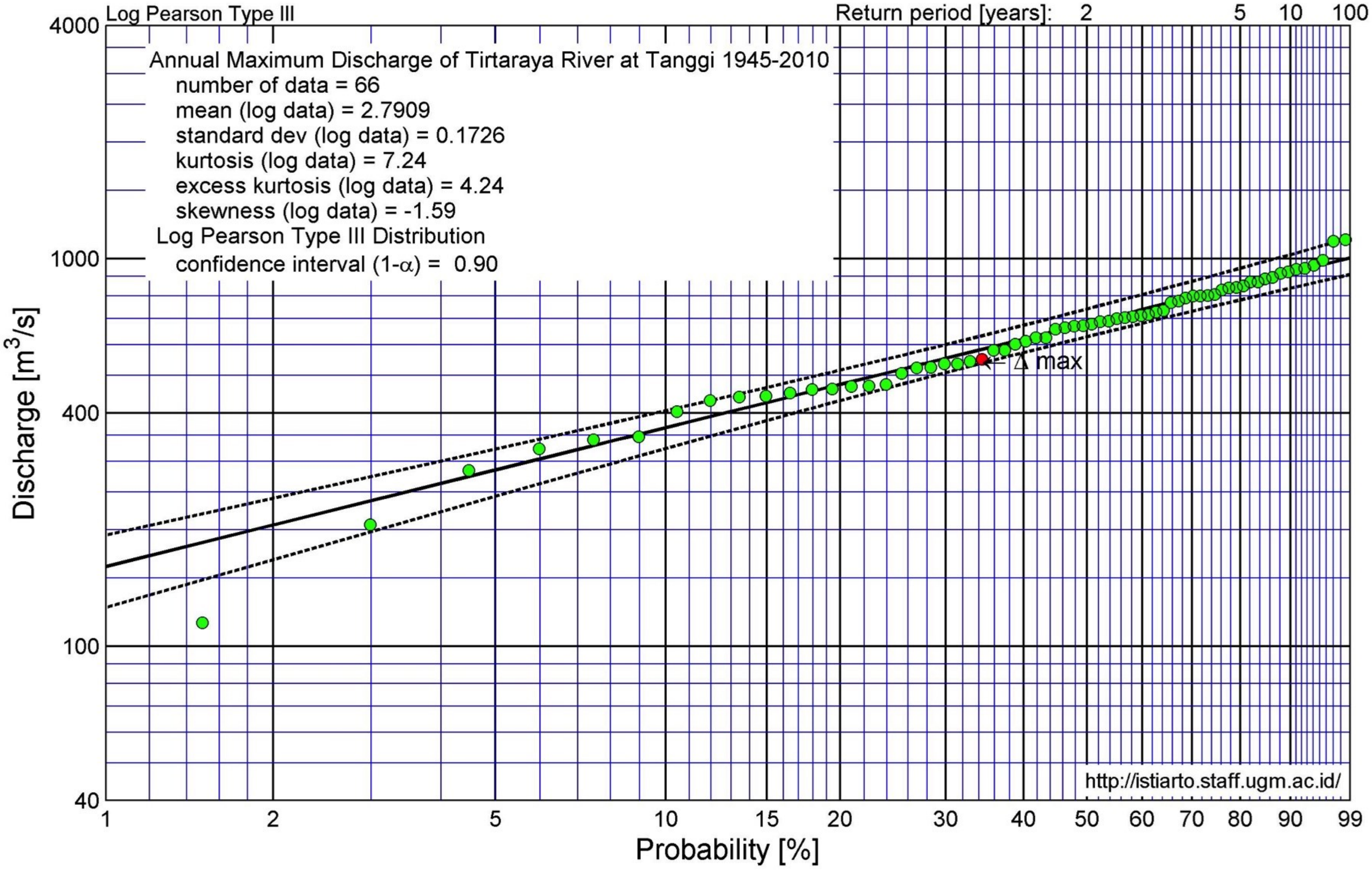
Tirtaraya Discharge resume.txt

```
Statistics of data
--> number of data   : 66
--> minimum          : 115
--> maximum          : 1120
--> mean             : 659.409091
--> standard dev     : 212.143274
--> kurtosis         : 2.772395
--> excess kurtosis  : -0.227605
--> skewness         : -0.167274

Statistics of log data
--> number of data   : 66
--> minimum          : 2.060698
--> maximum          : 3.049218
--> mean             : 2.790866
--> standard dev     : 0.172613
--> kurtosis         : 7.239843
--> excess kurtosis  : 4.239843
--> skewness         : -1.587294

Goodness of fit test, \alpha =  0.10 (confidence level 1-\alpha) =  0.90
                   Gumbel        Log Normal    Log Pearson III     Normal
Smirnov-Kolmogorov    OK                OK                 OK          OK
Discrepancy (max)  0.118             0.114              0.047       0.049
Chi-square            OK              fail               fail          OK
Chi-square (max)   6.545            13.758             13.758       3.152

Estimates on values according to various return periods [years]
Return period     Gumbel        Log Normal    Log Pearson III     Normal
      2              625               618                681         659
      5              812               863                854         838
     10              936              1030                919         931
     20             1060              1190                959        1010
     50             1210              1400                990        1100
    100             1320              1560               1000        1150
    200             1440              1720               1010        1210
    500             1590              1940               1020        1270
```


Gumbel
Return period [years]: 2 5 10 20 50 100
Annual Maximum Discharge of Tirtaraya River at Tanggi 1945-2010
number of data = 66
mean = 659
standard dev = 212
kurtosis = 2.77
excess kurtosis = -0.23
skewness = -0.17
Gumbel Distribution
confidence interval (1-α) = 0.90
Δmax
Discharge [m3/s]
1200 1100 1000 900 800 700 600 500 400 300 200 100
1 2 5 10 20 30 40 50 60 70 80 90 95 97 98 99
Probability [%]
http://istiarto.staff.ugm.ac.id/



Log Normal
Return period [years]: 2 5 10 20 50 100
Annual Maximum Discharge of Tirtaraya River at Tanggi 1945-2010
number of data = 66
mean (log data) = 2.7909
standard dev (log data) = 0.1726
kurtosis (log data) = 7.24
excess kurtosis (log data) = 4.24
skewness (log data) = -1.59
Log Normal Distribution
confidence interval (1-α) = 0.90
Δ max
Discharge [m³/s]
4000
1000
400
100
40
1 2 5 10 20 30 40 50 60 70 80 90 95 98 99
Probability [%]
http://istiarto.staff.ugm.ac.id/


Log Pearson Type III
Return period [years]: 2 5 10 100
Annual Maximum Discharge of Tirtaraya River at Tanggi 1945-2010
number of data = 66
mean (log data) = 2.7909
standard dev (log data) = 0.1726
kurtosis (log data) = 7.24
excess kurtosis (log data) = 4.24
skewness (log data) = -1.59
Log Pearson Type III Distribution
confidence interval (1-α) = 0.90
Δ max
Discharge [m³/s]
4000
1000
400
100
40
1 2 5 10 15 20 30 40 50 60 70 80 90 99
Probability [%]
http://istiarto.staff.ugm.ac.id/

AProb_4E

- Improvement
  - Feature that will be added: test for outliers

$$y_H = \bar{Y} + K_n \cdot s_Y$$
$$y_L = \bar{Y} - K_n \cdot s_Y$$

high and low outlier thresholds

in log unit

if logarithms of the data are greater than y_H or less than y_L → they are considered as outliers

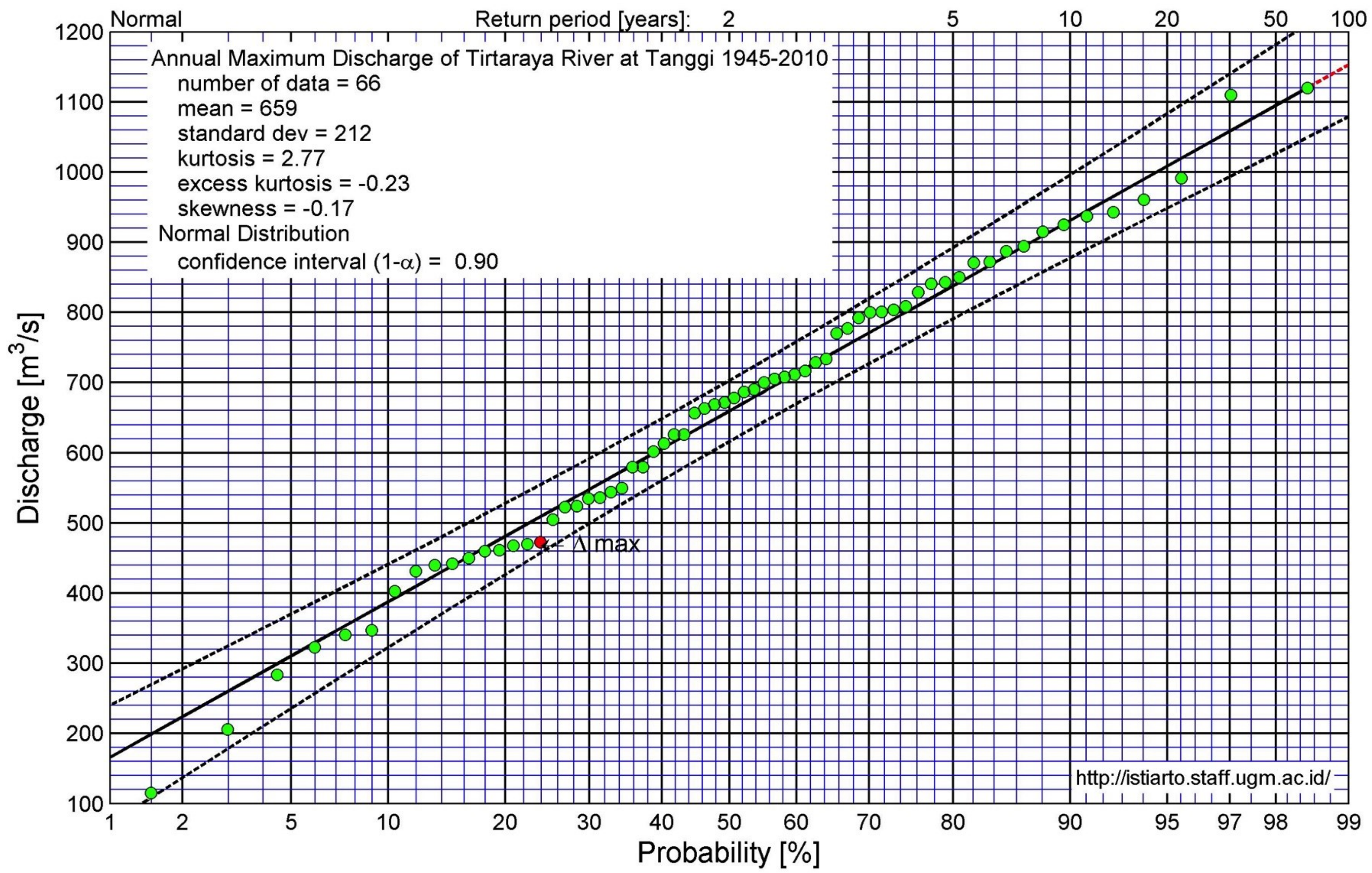

Normal
Return period [years]: 2 5 10 20 50 100
Annual Maximum Discharge of Tirtaraya River at Tanggi 1945-2010
number of data = 66
mean = 659
standard dev = 212
kurtosis = 2.77
excess kurtosis = -0.23
skewness = -0.17
Normal Distribution
confidence interval (1-α) = 0.90
Δ max
Discharge [m3/s]
Probability [%]
http://istiarto.staff.ugm.ac.id/

TABLE 12.5.3

Outlier test K_n values

Sample size n	K_n	Sample size n	K_n	Sample size n	K_n	Sample size n	K_n
10	2.036	24	2.467	38	2.661	60	2.837
11	2.088	25	2.486	39	2.671	65	2.866
12	2.134	26	2.502	40	2.682	70	2.893
13	2.175	27	2.519	41	2.692	75	2.917
14	2.213	28	2.534	42	2.700	80	2.940
15	2.247	29	2.549	43	2.710	85	2.961
16	2.279	30	2.563	44	2.719	90	2.981
17	2.309	31	2.577	45	2.727	95	3.000
18	2.335	32	2.591	46	2.736	100	3.017
19	2.361	33	2.604	47	2.744	110	3.049
20	2.385	34	2.616	48	2.753	120	3.078
21	2.408	35	2.628	49	2.760	130	3.104
22	2.429	36	2.639	50	2.768	140	3.129
23	2.448	37	2.650	55	2.804		

Source: U.S. Water Resources Council, 1981. This table contains one-sided 10-percent significance level K_n values for the normal distribution.

Screenshot

Outliers

- Transformasikan data debit ke data logaritma debit, $y = \log x$
- Jika nilai logaritma debit banjir maksimum dalam sampel melampaui y_H, maka debit itu adalah outlier
  - Bandingkan debit banjir itu dengan data banjir historis dan informasi banjir di lokasi sekitar
    - Jika ada catatan yang menunjukkan bahwa debit outlier tsb adalah debit maksimum dalam periode yang panjang, maka debit outlier tsb dimasukkan sbg catatan historis banjir dan dikeluarkan dari analisis frekuensi
    - Jika tidak ada informasi banjir di lokasi sekitar, maka debit outlier tsb tetap diikutkan dalam analisis frekuensi.

Outliers

- Transformasikan data debit ke data logaritma debit, $y = \log x$
- Jika nilai logaritma debit banjir dalam sampel lebih kecil daripada y_L, maka debit itu adalah outlier
  - Debit oulier itu dikeluarkan dari analisis frekuensi

References

- Chow, Ven Te, Maidment, David R., Mays, Larry W., 1988, *Applied Hydrology*, McGraw-Hill, Inc., New York.
- Haan, C.T., 1982, *Statistical Methods in Hydrology*, 1st Ed., 3rd Printing, The Iowa State Univ. Press, Ames, Iowa, USA.

Terima kasih