Workshop on Hydrology

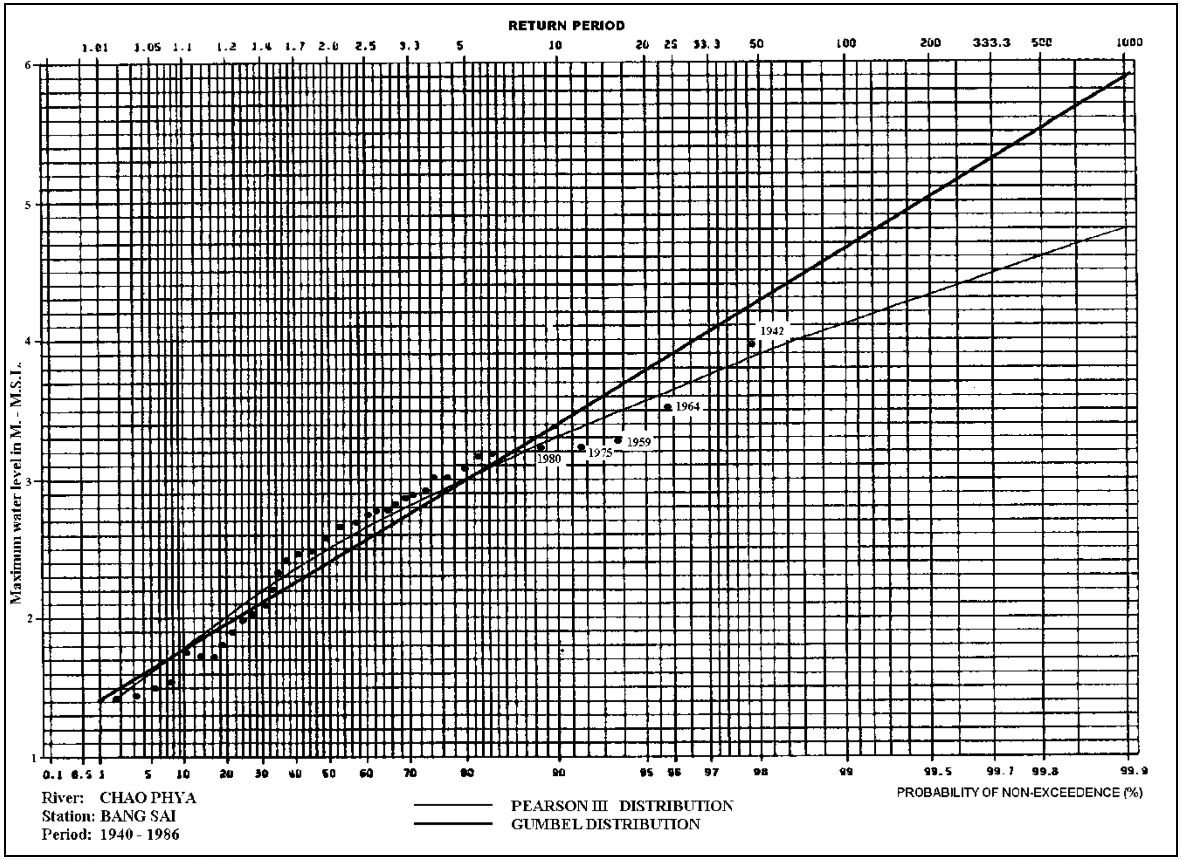
Dr. Pieter J.M. de Laat
Associate Professor in Land and Water Development
UNESCO-IHE Institute for Water Education
E-mail: p.delaat@unesco-ihe.org

Workshop on Hydrology

Contents

- Analysis of Extremes
- Data Analysis
- Composition of a Rating Curve
- Rainfall-Runoff Modelling

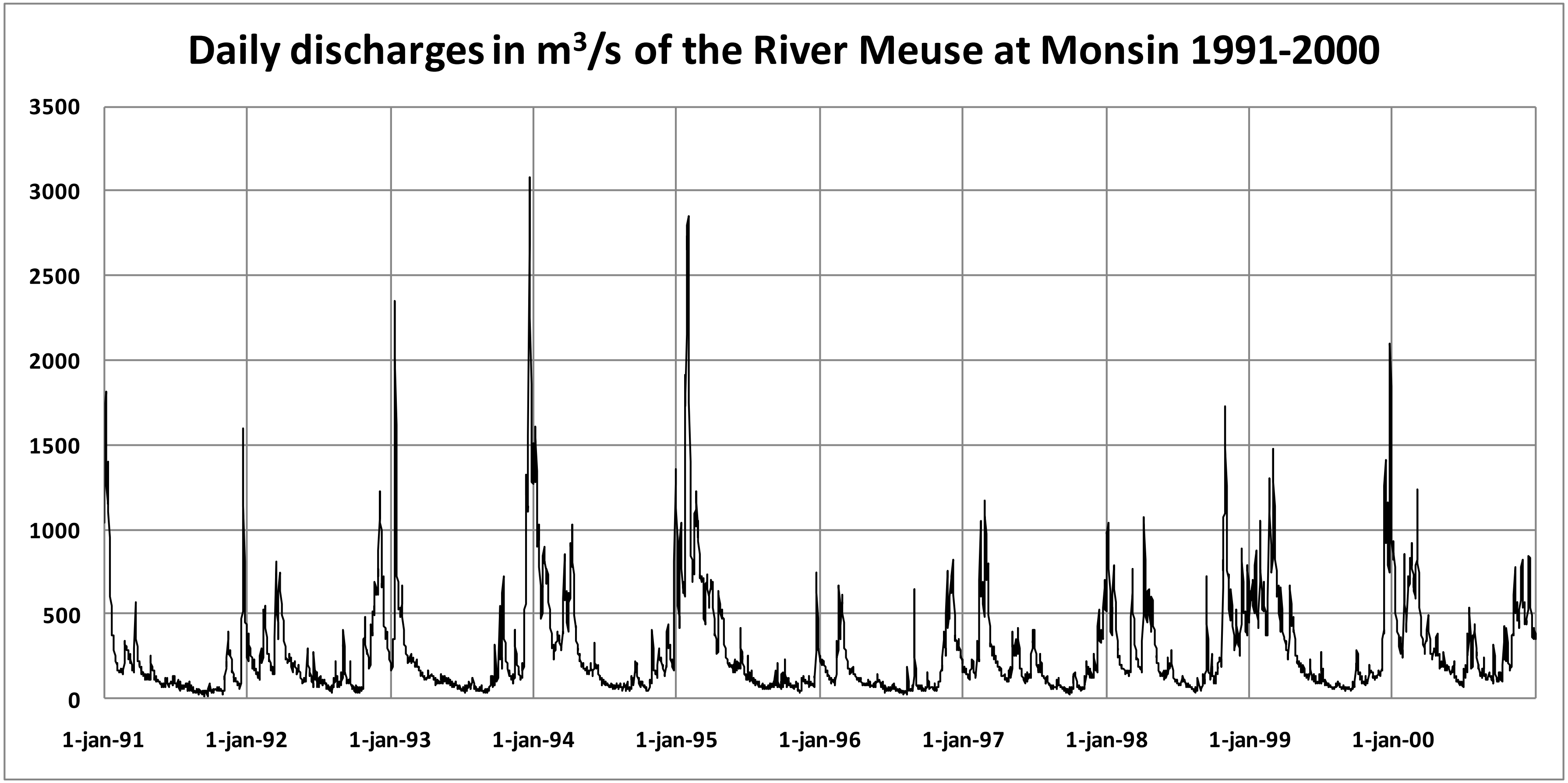
Analysis of Extremes



Analysis of extreme events

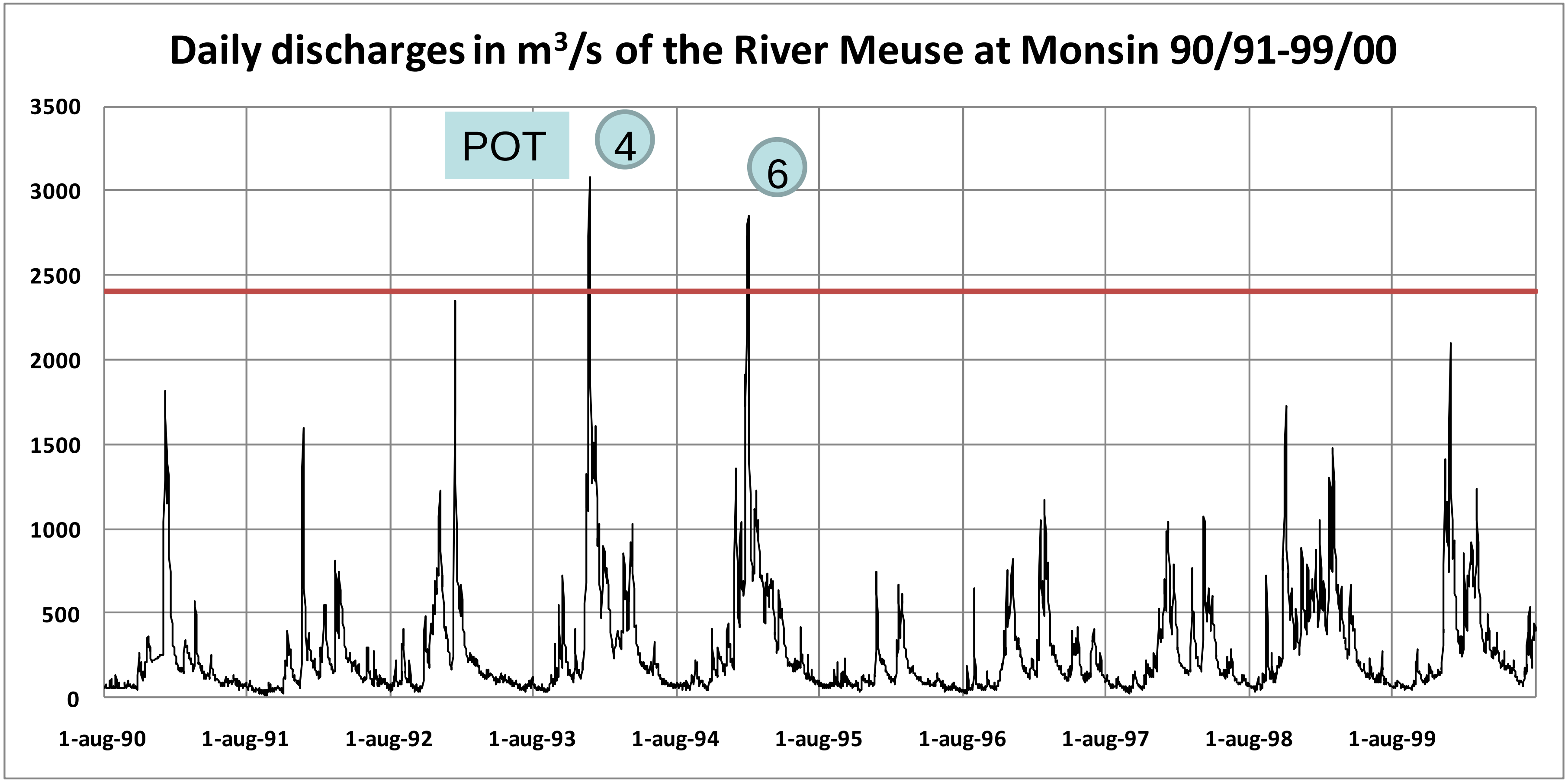
- Use long time series of daily values (20 years is the minimum)
- Select from each year the day with the highest (or lowest) value
- Make sure that the annual extremes are independent of each other (do not belong to the same extreme event)
- Use of *water years* or *hydrological years* often ensures the independency of annual extremes
- The use of POT (Peak Over Threshold) often results in values that are not independent

For the analysis of extreme discharges of the River Meuse
the use of calendar years does not result
in a set of independent annual extremes



Choose Hydrological Year: 1 August – 31 July
The annual extremes are independent

The 10 extremes larger than 2400 m^3/s (POT values) are not independent



Numerical example Given data: Annual maximum daily rainfall of 10 years (N = 10)

Year	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
mm/d	56	52	60	70	34	30	44	48	40	38

Rank values in descending order $T = \frac{1}{P}$

m Rank	Rainfall amount (mm)	p Probability exceedence	T Return period
1	70	0.09	11.0
2	60	0.18	5.5
3	56	0.27	3.7
4	52	0.36	2.8
5	48	0.45	2.2
6	44	0.54	1.8
7	40	0.64	1.6
8	38	0.73	1.4
9	34	0.82	1.2
10	30	0.91	1.1

Estimate probability of exceedance:

$$\cancel{p = \frac{m}{N}}$$

Weibull:

$$p = \frac{m}{N+1}$$

Gringerton:

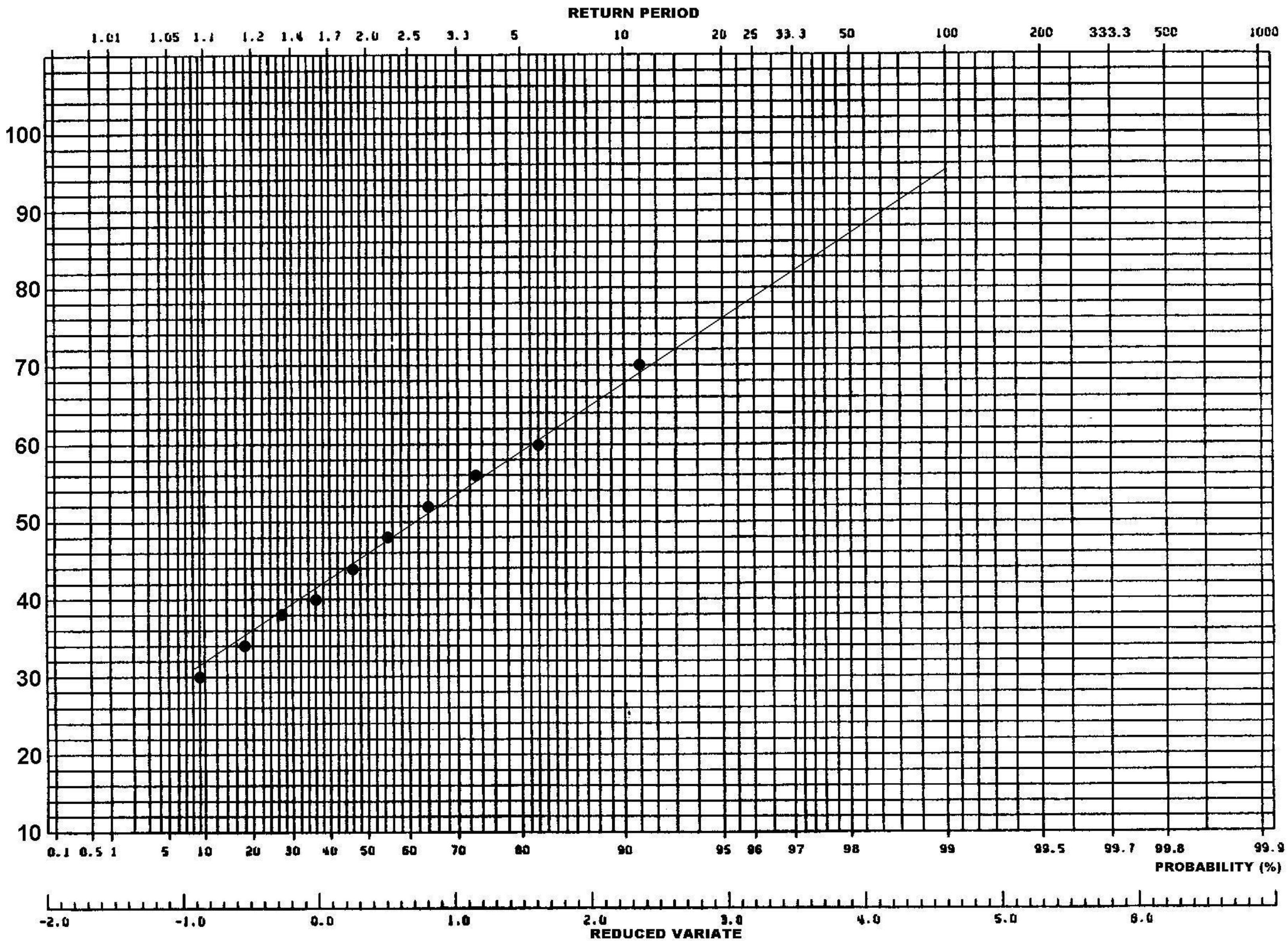
$$p = \frac{m - 0.44}{N + 0.12}$$

Plot on special (Gumbel) paper

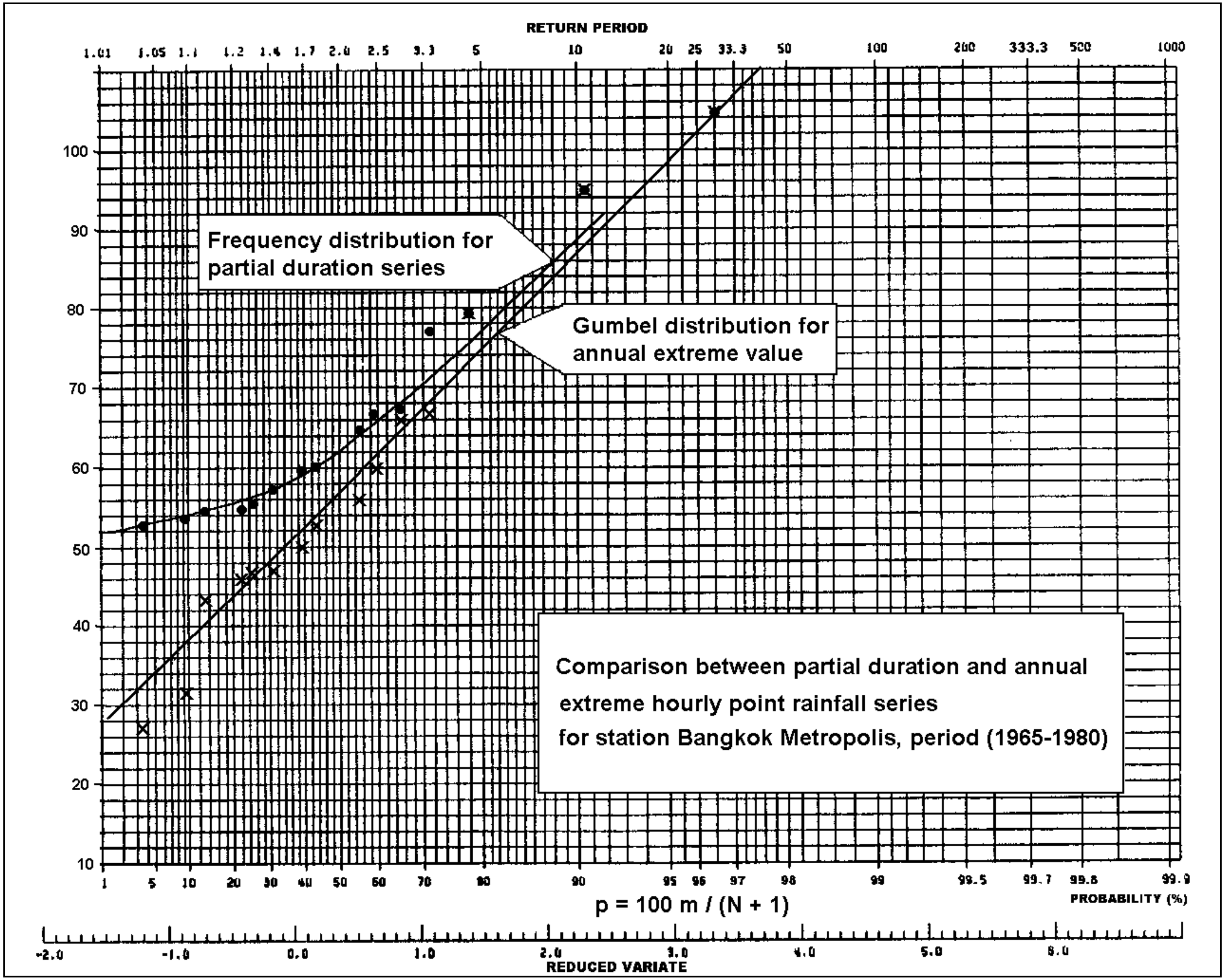
Rank	Rainfall amount (mm)	p Probability exceedence	T Return period	Log T	q = 1 - p Probability of non-exc.	y Reduced Variate
1	70	0.09	11.0	1.041	0.91	2.351
2	60	0.18	5.5	0.740	0.82	1.606
3	56	0.27	3.7	0.564	0.73	1.144
4	52	0.36	2.8	0.439	0.64	0.794
5	48	0.45	2.2	0.342	0.55	0.501
6	44	0.54	1.8	0.263	0.46	0.238
7	40	0.64	1.6	0.196	0.36	-0.012
8	38	0.73	1.4	0.138	0.27	-0.262
9	34	0.82	1.2	0.087	0.18	-0.533
10	30	0.91	1.1	0.041	0.09	-0.875

Plot on linear paper or in spreadsheet
where $y = -\ln(-\ln(1-p))$

Example of Annual Extremes plotted on extreme value paper

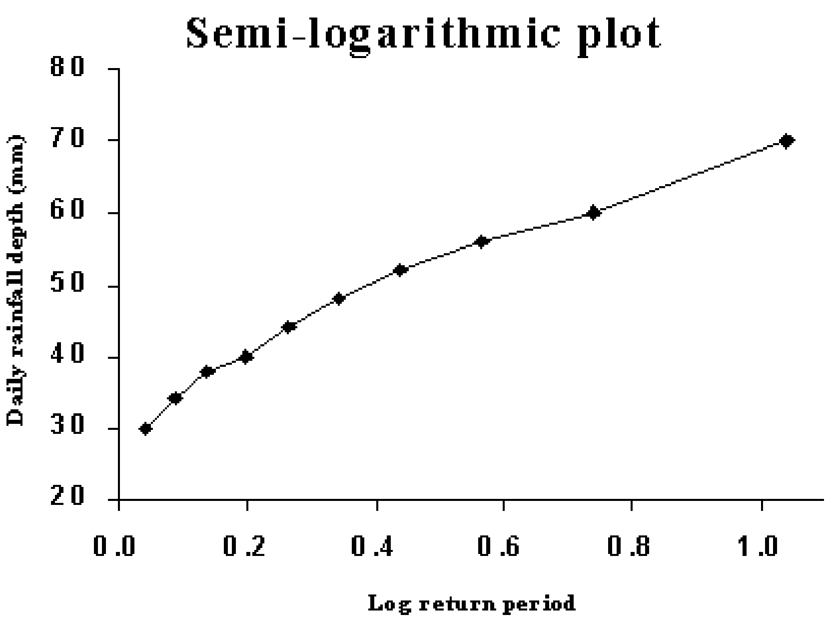


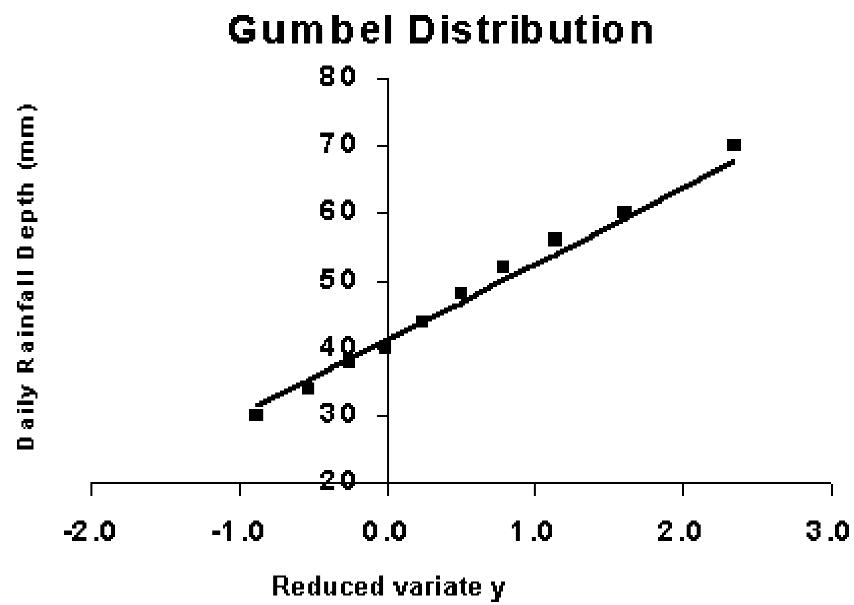
Example of Annual Extremes and POT plotted on extreme value paper



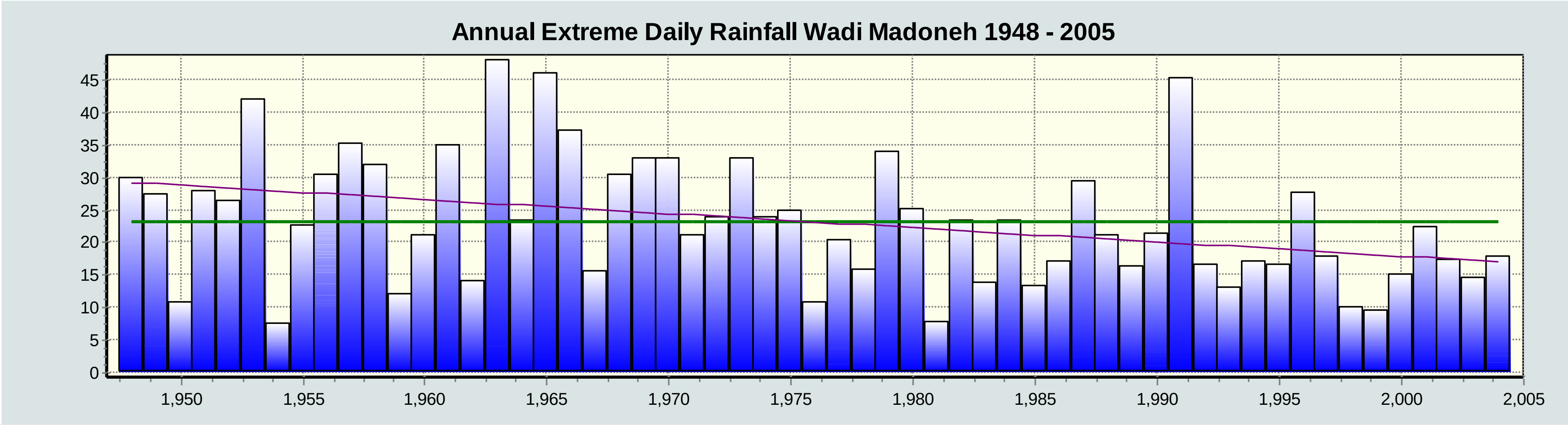
Apart from the Gumbel and the logarithmic distributions there are more extreme value distributions. Try different distributions to find the one that best fits the data. A normal distribution is generally not suitable to fit extreme rainfall or runoff data.

It is generally acceptable to extrapolate up to twice the length of the record. So, if you have 50 years of data, the extreme event to be exceeded once in 100 years can be extrapolated.





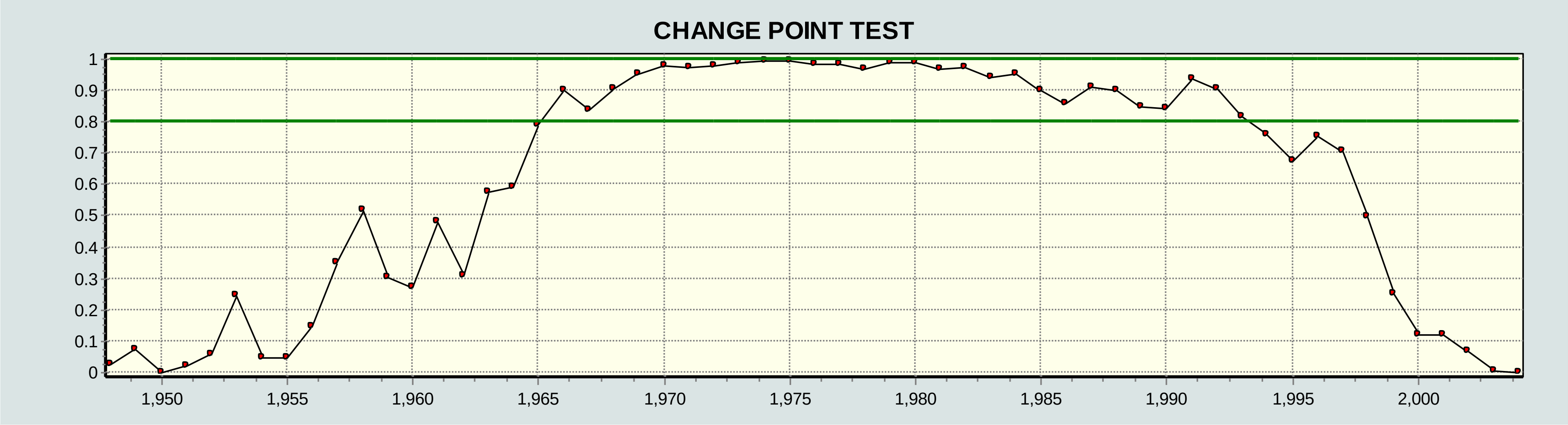
Annual Extreme daily rainfall
Wadi Madoneh 1948 – 2004



Trend is statistically significant

Time series not suitable for extreme value analysis

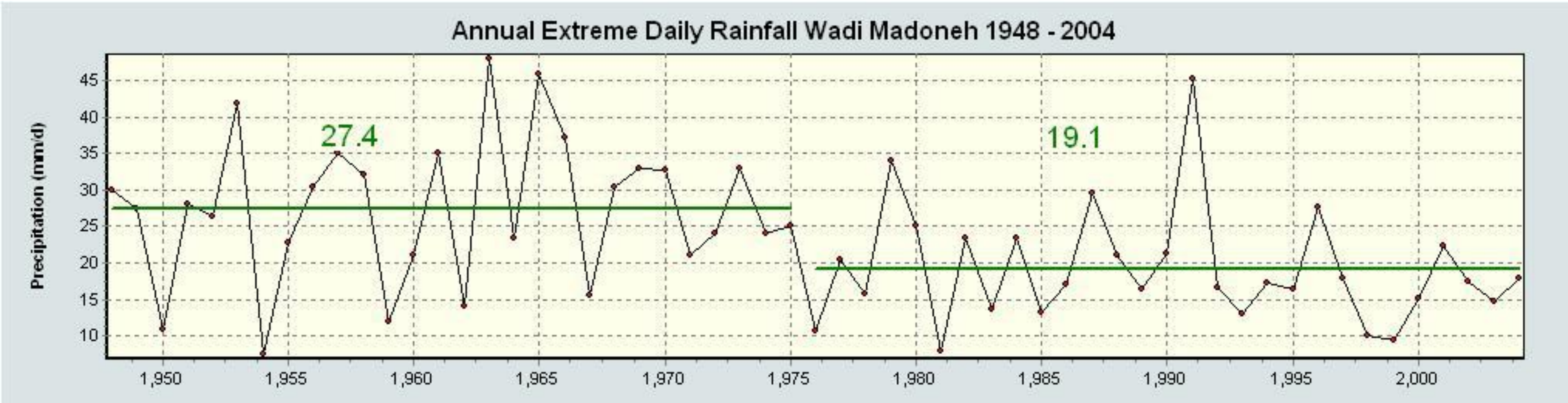
Annual Extreme daily rainfall Wadi Madoneh 1948 - 2004



Pettitt - test

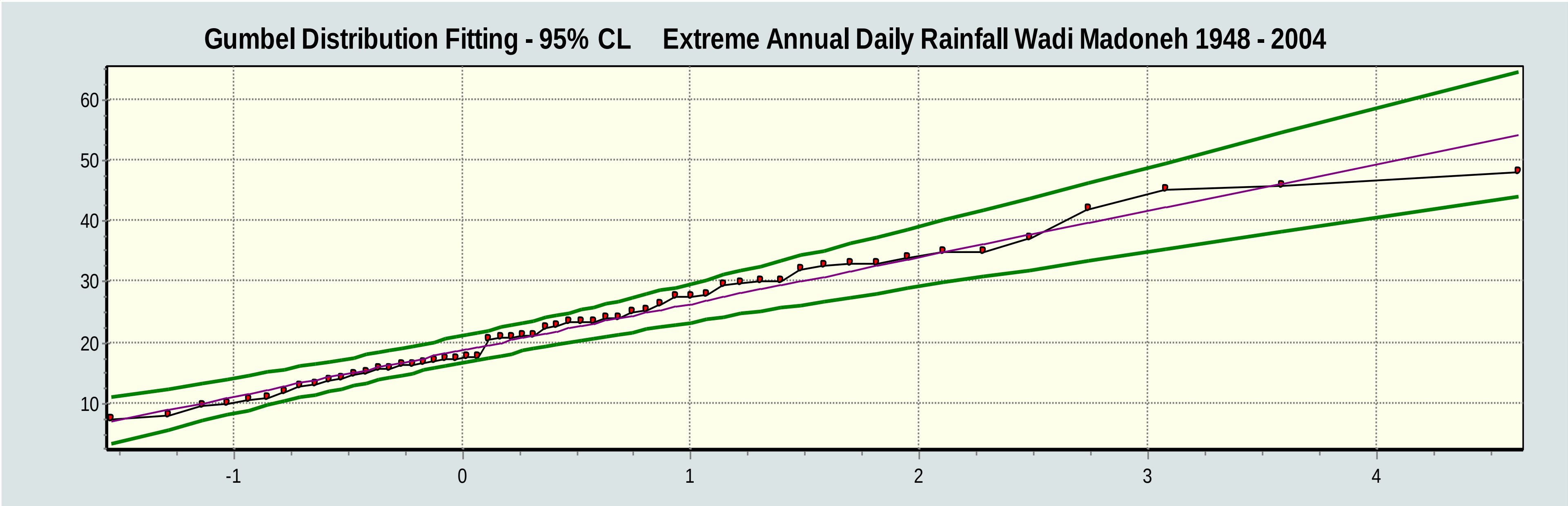
Change point in 1975 with a probability of 99.6 %

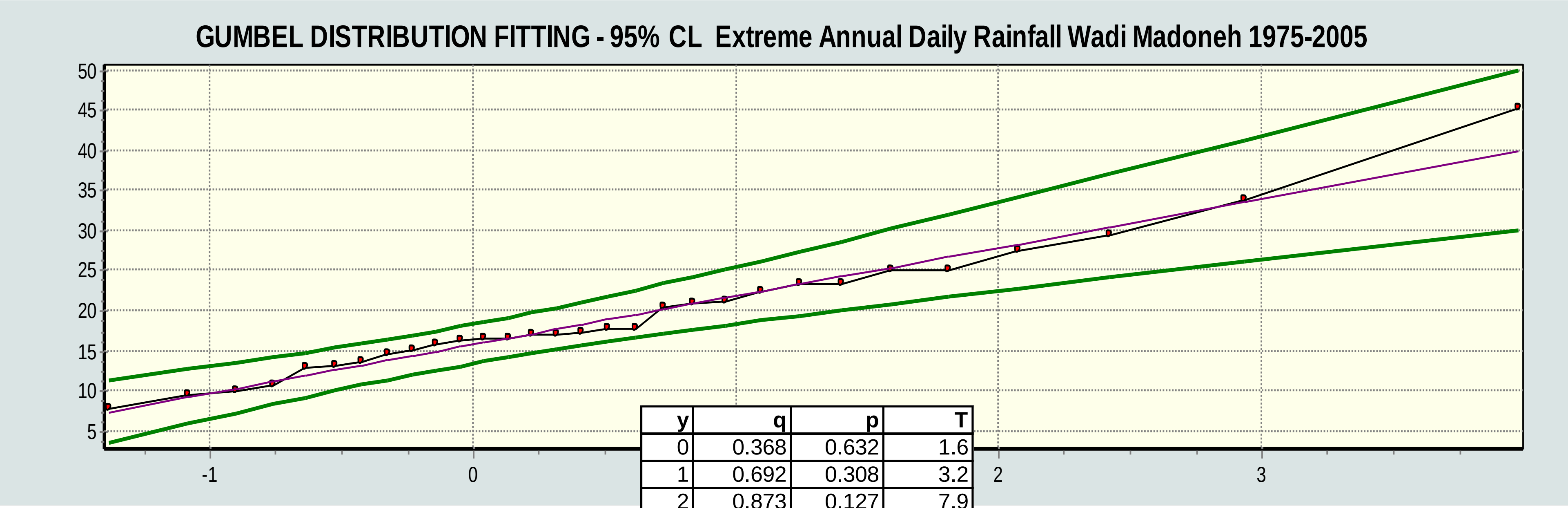
Annual Extreme daily rainfall Wadi Madoneh 1948 - 2004



Split record test

Mean is significantly unstable





y	q	p	T
0	0.368	0.632	1.6
1	0.692	0.308	3.2
2	0.873	0.127	7.9
3	0.951	0.049	20.6
4	0.982	0.018	55.1
5	0.993	0.007	148.9
6	0.998	0.002	403.9
7	0.999	0.001	1097.1

$$1-p=q=exp(-exp(-y))$$

$$y=-ln(-ln(1-p))$$

Assumptions of Frequency Analysis

1. **All data points are correct and precisely measured**
   - When analyzing discharges, be aware of the uncertainty of extreme data
2. **Independent events: extremes are not part of the same event**
   - Carefully check the data set; plot the whole record, in particular all events of the POT series
   - Carefully choose the hydrological year, and even so, check the independence in climates that have an even distribution of events over the year
3. **Random sample: Every value in the population has equal chance of being included in the sample**
4. **The hydrological regime has remained static during the complete time period of the record**
   - No climate change, and for an analysis of peak discharges: no land use change, no changes in the river channels, no change in the flood water management etc. in the catchment (often not the case for long records!)
5. **All extremes originate from the same statistical population (homogeneity)**
   - Extreme rainfall events may be generated by different rain bringing mechanisms. The same applies to floods: different flood generating mechanisms (e.g. rain storms, snow melt, snow-on-ice etc.) might cause floods with different frequencies/recurrence intervals

Procedure for analysis of extremes

- Determine for each year the minimum and maximum value
- Delete years with missing values (for which minimum value equals -1)
- Rank annual extremes in descending order
- Compute the probability of exceedence p (plotting position) with the equation of Weibull $p = 1/(N+1)$ or Gringerton $p = (m-0.44)/(N+0.12)$
- Compute the logarithm of the return period ($T = 1/p$)
- Plot the annual extremes vs log T
- Compute the reduced variable $y = -ln(-ln(1-p))$
- Plot the annual extremes vs y
- Compute the extreme value according to Gumbel X_{Gum}

- The Gumbel distribution (based on the method of moments) allows an extrapolation beyond the period of observation

Reduced variate y

$$y = -\ln(-\ln(1-1/T))$$

$$y = a\,(X_{Gum} - b)$$

$$b = \overline{X}_{ext} - s_{ext}\frac{y_N}{\sigma_N}$$

Find y_N and σ_N from appendix E

$$a = \frac{\sigma_N}{s_{ext}}$$

$$X_{Gum} = \overline{X}_{ext} + \frac{s_{ext}}{\sigma_N}\left(y - y_N\right)$$

- Plot the theoretical Gumbel distribution in the same chart.
- Rule of thumb: Do <u>not extrapolate</u> recurrence intervals <u>beyond twice the length of your data record</u>

Adding confidence limits

Compute the standard error of estimate SE_X in terms of y which for a Gumbel distribution follows from:

$$SE_X = \frac{s_{ext}}{\sqrt{N}}\left[1 + \frac{1.14}{\sigma_N}\left(y - y_N\right) + \frac{1.10}{\sigma_N^2}\left(y - y_N\right)^2\right]^{0.5}$$

Compute the degree of freedom from sample size N

$$\nu = N - 1$$

Find from Student's t-distribution the critical values t_c for 95% confidence interval (appendix B)

Assuming that the errors of the estimated extremes are normally distributed, the upper and lower limit of the confidence interval X_c are as follows related to the standardized values of t_c

$$\pm t_c = \frac{X_c - X_{Gum}}{SE_X}$$

So the confidence limits are computed from

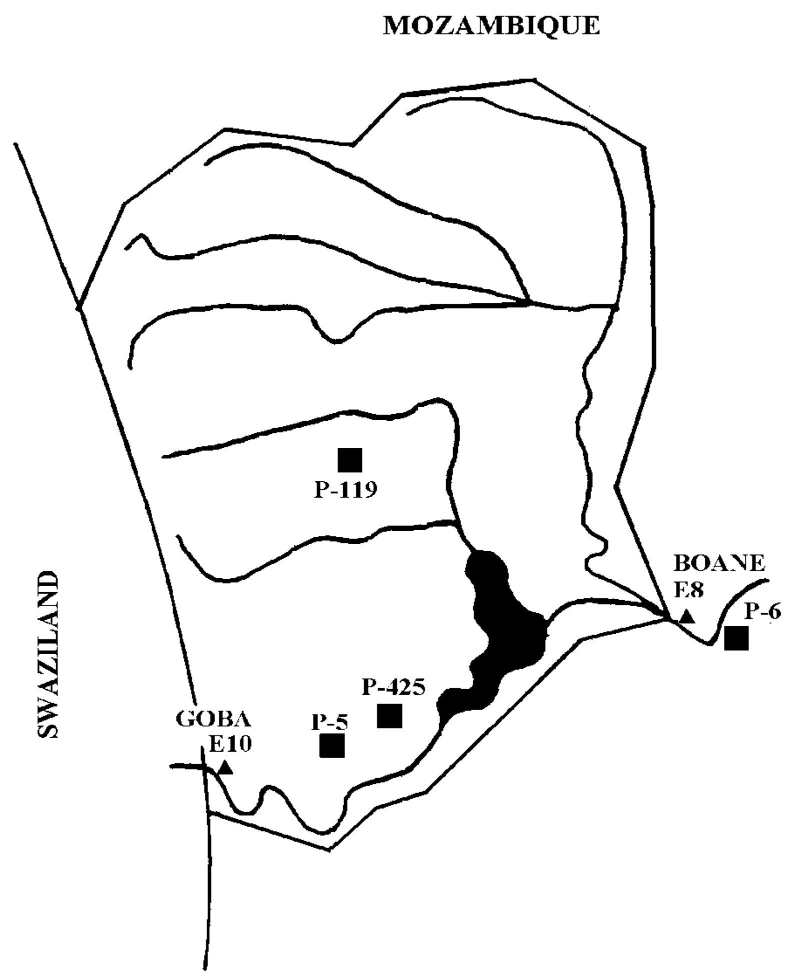
$$X_c = X_{Gum} \pm t_c \, SE_X$$

Data Analysis

- Tabular Comparison
- Data Completion through Linear Regression
- Double Mass Analysis
- Method of Cumulative Residuals

January	P425	P119	P5	P6
51/52		108.8	114.2	70.8
52/53		305.4	186.2	172.3
53/54		84.2	87.4	44.3
54/55		293.8	154.1	235.0
55/56		123.0	85.6	54.9
56/57		87.4	66.2	59.8
57/58		253.1	216.2	171.7
58/59	175.3	123.7	162.9	79.5
59/60	79.5	63.5	76.4	84.3
60/61	56.0	49.1	110.0	84.5
61/62	142.4	118.1	93.1	188.6
62/63	95.7	115.8	111.8	84.7
63/64	249.0	173.2	210.3	215.5
64/65	12.3	56.2	14.3	40.4
65/66	546.8	672.2	625.1	587.6
66/67	76.6	190.1	48.5	162.1
67/68	121.5	113.7	92.0	71.4
68/69	157.7	188.7	125.5	111.4
69/70	4.7	13.0	98.4	9.6
70/71	51.8	98.6	87.9	53.1
71/72	218.0	320.6	156.8	210.4
72/73	56.7	57.2	79.1	63.3
73/74	108.6	209.6	151.7	299.5
74/75	81.3	183.8	138.7	232.3
75/76	210.2	416.7	311.4	275.0
76/77	44.3	150.9	77.0	115.5
77/78	122.0	354.0	305.6	202.9
78/79	122.5	171.2	60.5	129.0
79/80	62.2	157.7	52.9	33.8
80/81		112.5		194.4
81/82		96.1		24.5
MIS	9	0	2	0
AVG	127.1	176.2	141.4	140.7
STDEV	114.0	133.3	116.3	114.9
MIN	4.7	13.0	14.3	9.6
MAX	546.8	672.2	625.1	587.6
P20	31.3	64.2	43.7	44.2
P80	222.8	288.2	239.0	237.2

Example of tabular comparison of monthly rainfall values of 4 stations (P5, P6, P119 and P425) in the Umbeluzi catchment in Mozambique

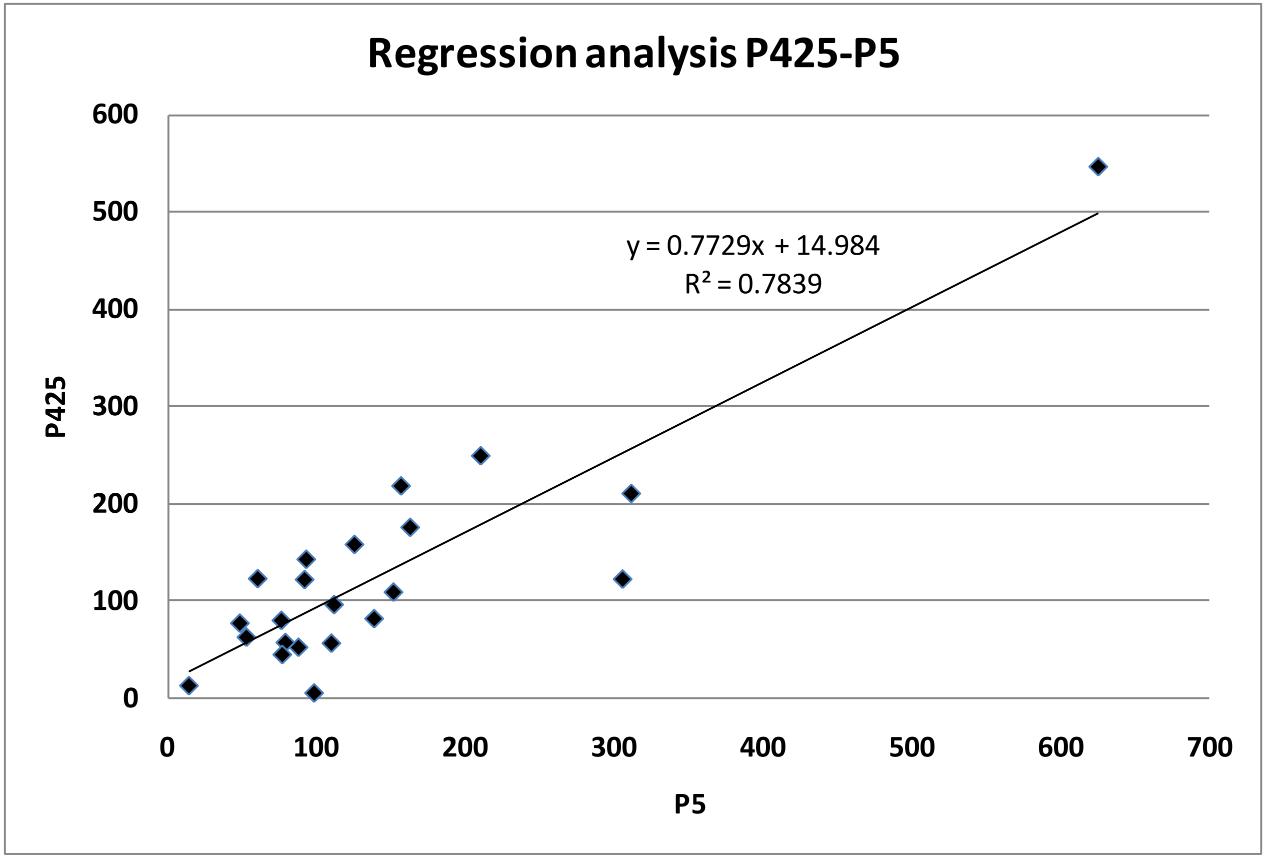


Filling in missing data through linear regression

January	P425	P5
58/59	175.3	162.9
59/60	79.5	76.4
60/61	56.0	110.0
61/62	142.4	93.1
62/63	95.7	111.8
63/64	249.0	210.3
64/65	12.3	14.3
65/66	546.8	625.1
66/67	76.6	48.5
67/68	121.5	92.0
68/69	157.7	125.5
69/70	4.7	98.4
70/71	51.8	87.9
71/72	218.0	156.8
72/73	56.7	79.1
73/74	108.6	151.7
74/75	81.3	138.7
75/76	210.2	311.4
76/77	44.3	77.0
77/78	122.0	305.6
78/79	122.5	60.5
79/80	62.2	52.9

$$Y = C_1 X + C_0$$

$$P_{425} = 0.7729 P_5 + 14.984$$



Filling in missing data through multiple linear regression

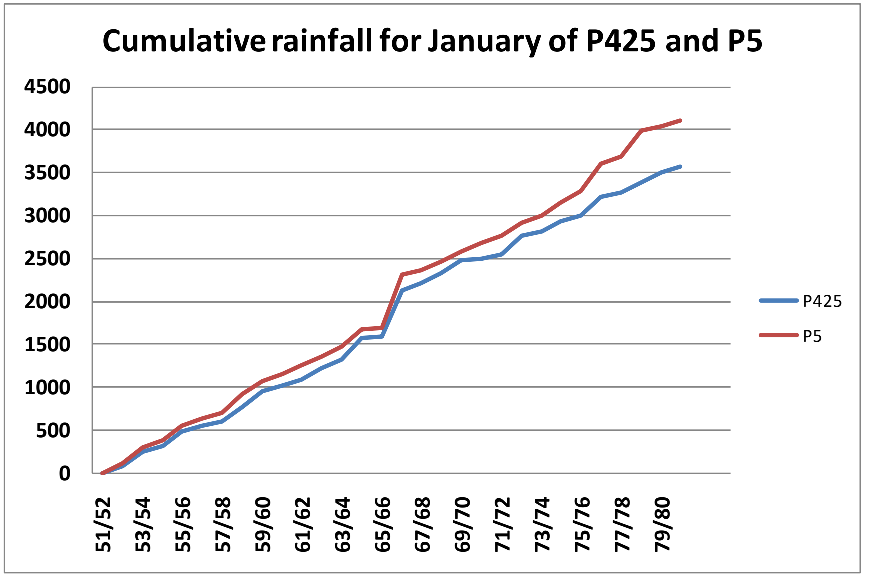
$$Y = C_0 + C_1 X_1 + C_2 X_2 + C_3 X_3$$

$$P_{425} = 3.8211 + 0.0918 P_{119} + 0.4495 P_5 + 0.2729 P_6$$

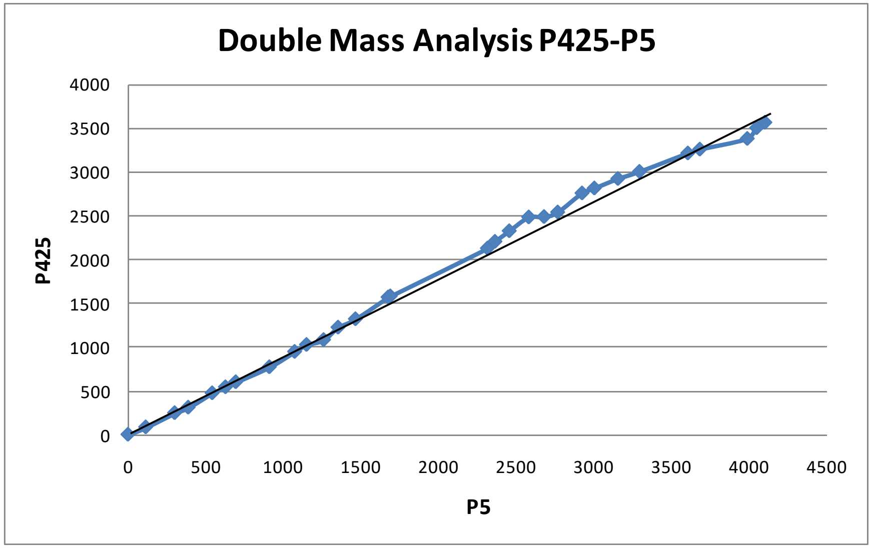
January	P425	P119	P5	P6
58/59	175.3	123.7	162.9	79.5
59/60	79.5	63.5	76.4	84.3
60/61	56.0	49.1	110.0	84.5
61/62	142.4	118.1	93.1	188.6
62/63	95.7	115.8	111.8	84.7
63/64	249.0	173.2	210.3	215.5
64/65	12.3	56.2	14.3	40.4
65/66	546.8	672.2	625.1	587.6
66/67	76.6	190.1	48.5	162.1
67/68	121.5	113.7	92.0	71.4
68/69	157.7	188.7	125.5	111.4
69/70	4.7	13.0	98.4	9.6
70/71	51.8	98.6	87.9	53.1
71/72	218.0	320.6	156.8	210.4
72/73	56.7	57.2	79.1	63.3
73/74	108.6	209.6	151.7	299.5
74/75	81.3	183.8	138.7	232.3
75/76	210.2	416.7	311.4	275.0
76/77	44.3	150.9	77.0	115.5
77/78	122.0	354.0	305.6	202.9
78/79	122.5	171.2	60.5	129.0
79/80	62.2	157.7	52.9	33.8

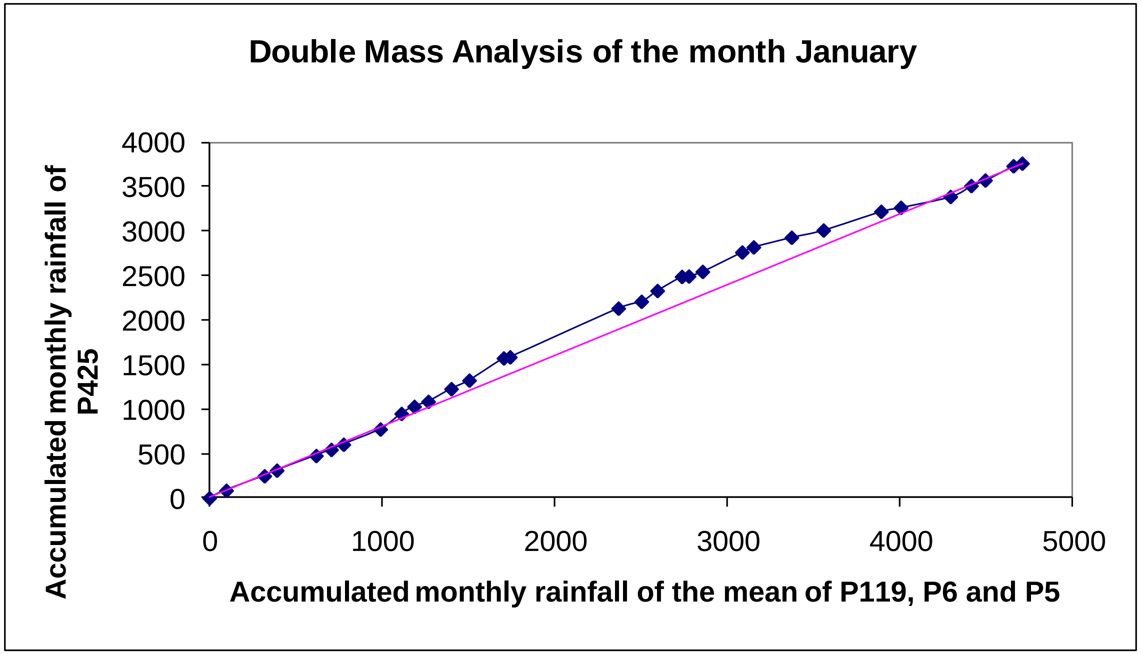
SUMMARY OUTPUT	
P425 and P119, P5, P6	
Regression Statistics	
Multiple R	0.9044524
R Square	0.8180341
Adjusted R Square	0.7877065
Standard Error	52.524354
Observations	22
ANOVA	
	df
Regression	3
Residual	18
Total	21
	Coefficients
Intercept	3.821143
X Variable 1	0.0918169
X Variable 2	0.4495039
X Variable 3	0.2729471

January	P425	P5	P425 sum	P5 sum
			0	0
51/52	84.5	114.2	84.5	114.2
52/53	162.6	186.2	247.1	300.4
53/54	62.9	87.4	310.0	387.8
54/55	164.2	154.1	474.2	541.9
55/56	68.6	85.6	542.8	627.5
56/57	57.9	66.2	600.7	693.7
57/58	171.1	216.2	771.8	909.9
58/59	175.3	162.9	947.1	1072.8
59/60	79.5	76.4	1026.6	1149.2
60/61	56.0	110.0	1082.6	1259.2
61/62	142.4	93.1	1225.0	1352.3
62/63	95.7	111.8	1320.7	1464.1
63/64	249.0	210.3	1569.7	1674.4
64/65	12.3	14.3	1582.0	1688.7
65/66	546.8	625.1	2128.8	2313.8
66/67	76.6	48.5	2205.4	2362.3
67/68	121.5	92.0	2326.9	2454.3
68/69	157.7	125.5	2484.6	2579.8
69/70	4.7	98.4	2489.3	2678.2
70/71	51.8	87.9	2541.1	2766.1
71/72	218.0	156.8	2759.1	2922.9
72/73	56.7	79.1	2815.8	3002.0
73/74	108.6	151.7	2924.4	3153.7
74/75	81.3	138.7	3005.7	3292.4
75/76	210.2	311.4	3215.9	3603.8
76/77	44.3	77.0	3260.2	3680.8
77/78	122.0	305.6	3382.2	3986.4
78/79	122.5	60.5	3504.7	4046.9
79/80	62.2	52.9	3566.9	4099.8

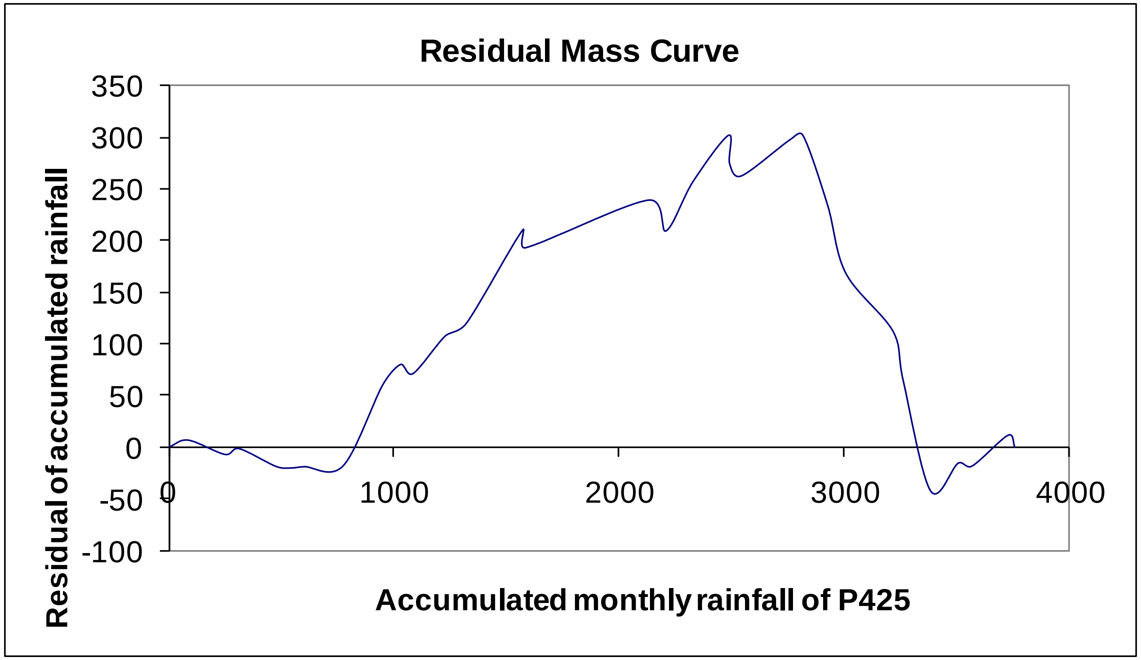


Double Mass Analysis



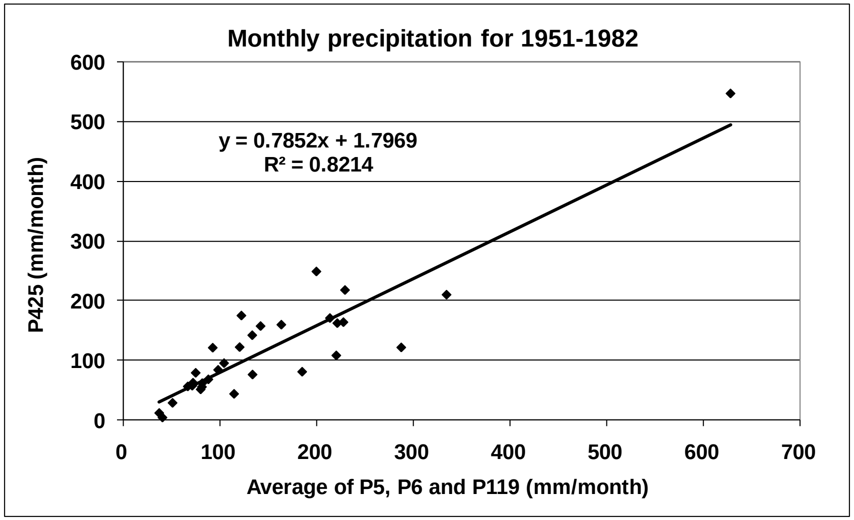


Double Mass Curve of station P425 against the mean of all other stations

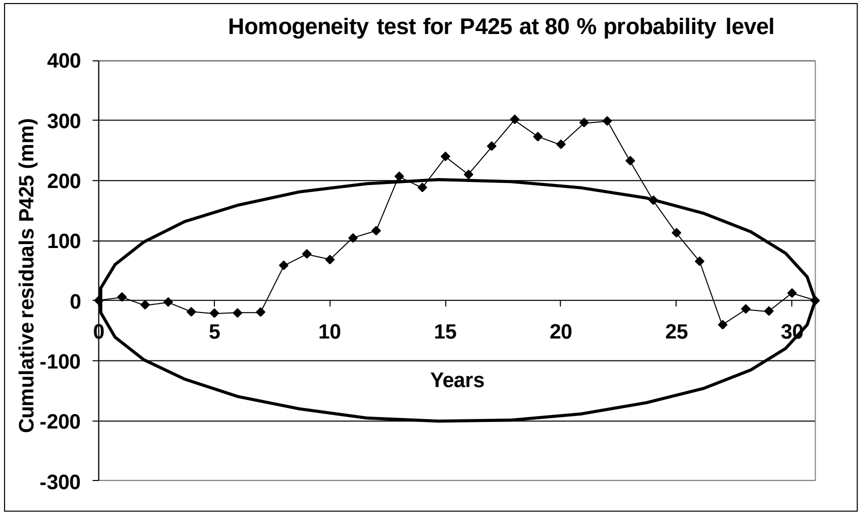


Deviation from average linear is the Residual Mass Curve

Method of Cumulative Residuals for testing the homogeneity of a time series

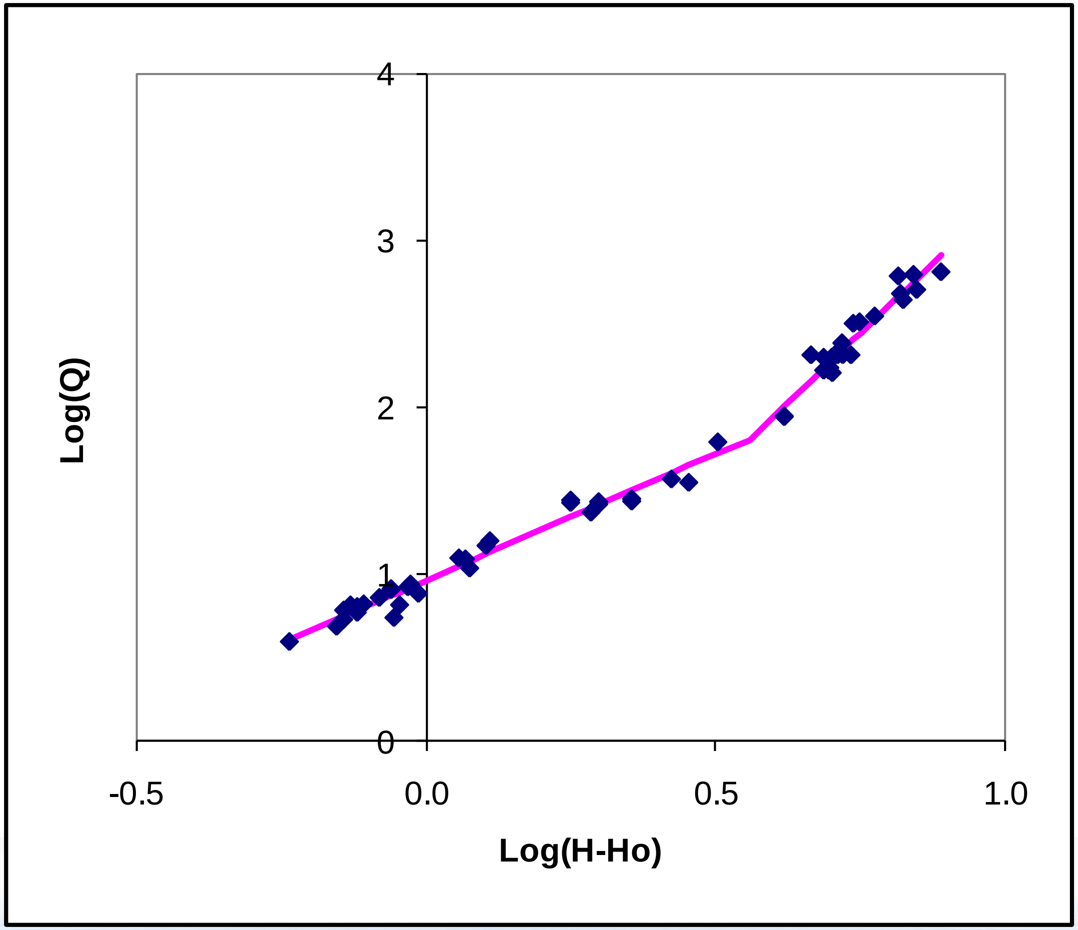


Method based on regression of monthly data between target station (P425) and average of other 3 stations.

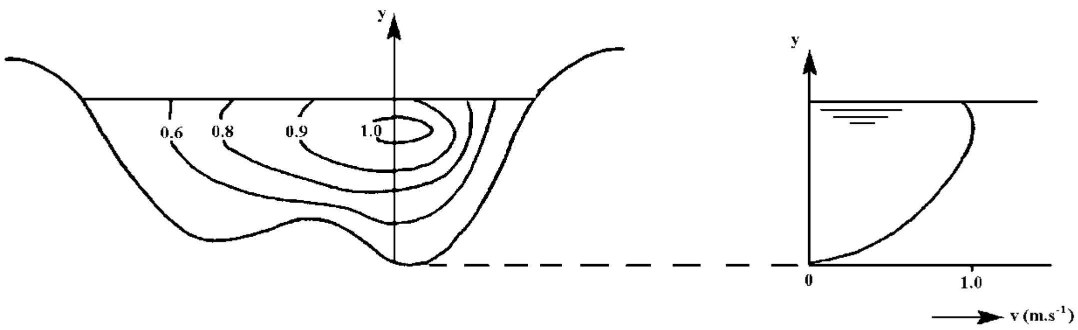


If the cumulated residuals of the monthly data and the regression line lay inside the ellipse, the time series of monthly values of P425 is considered homogeneous at 80% level of confidence

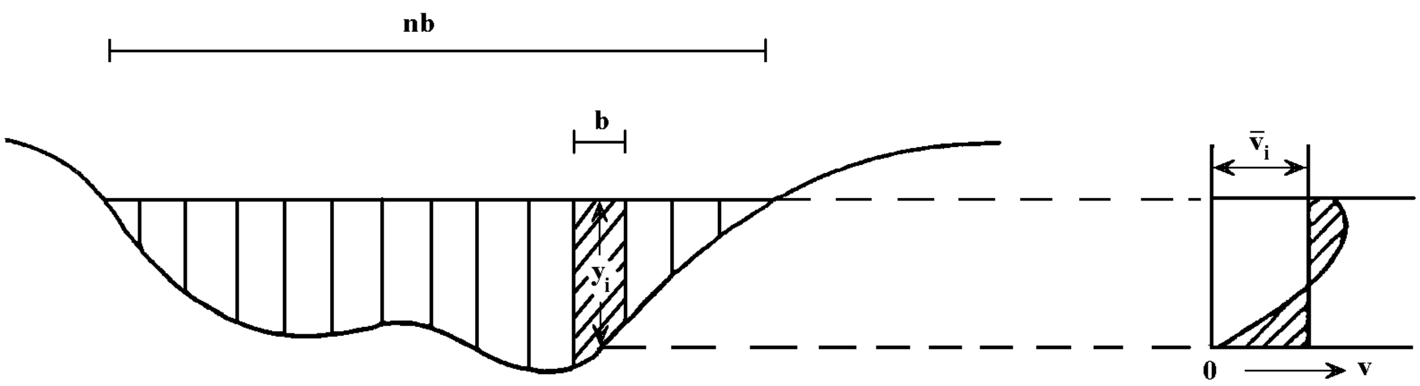
Composition of a Rating Curve



Discharge measurement: Velocity-area method



$$Q = \bar{v}\,A$$

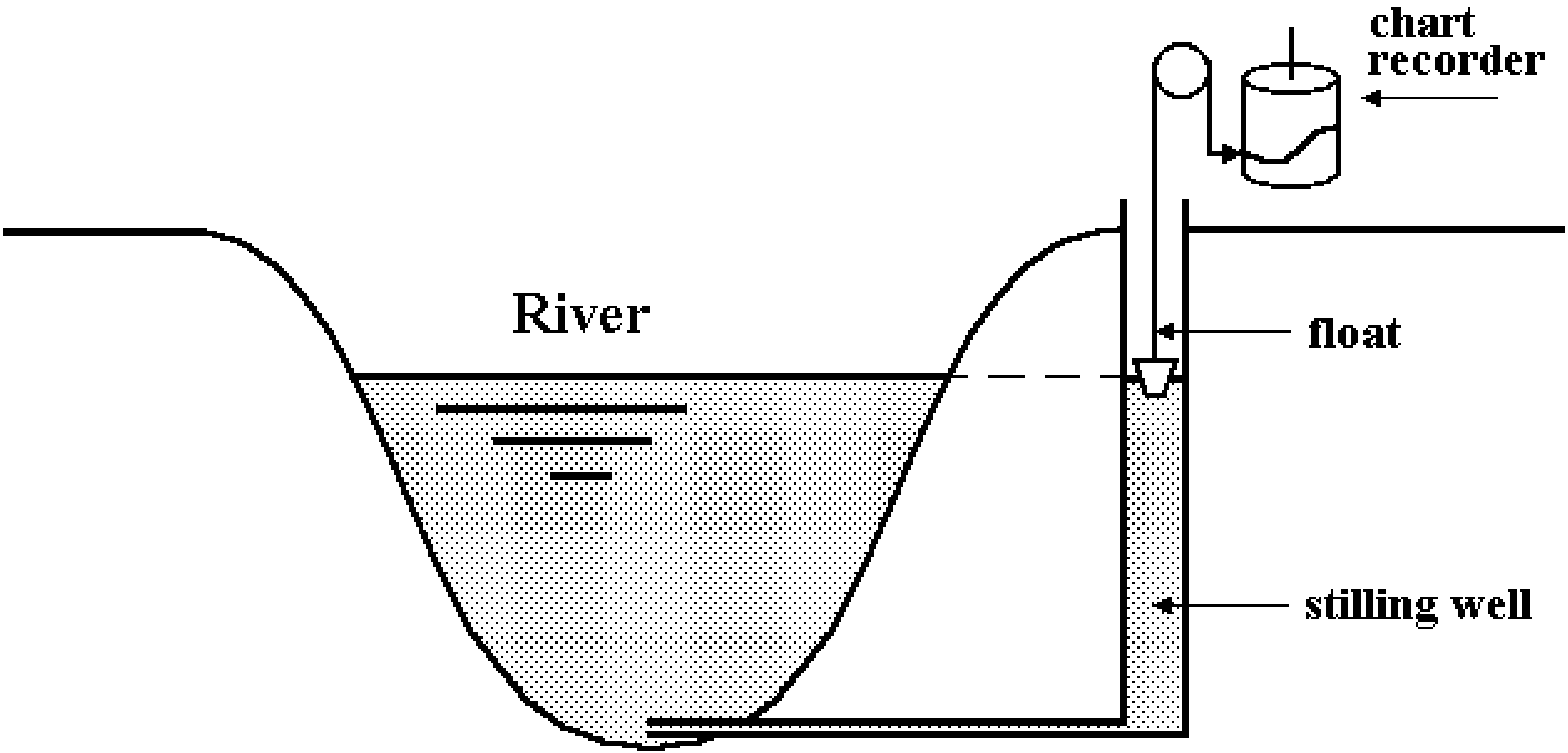


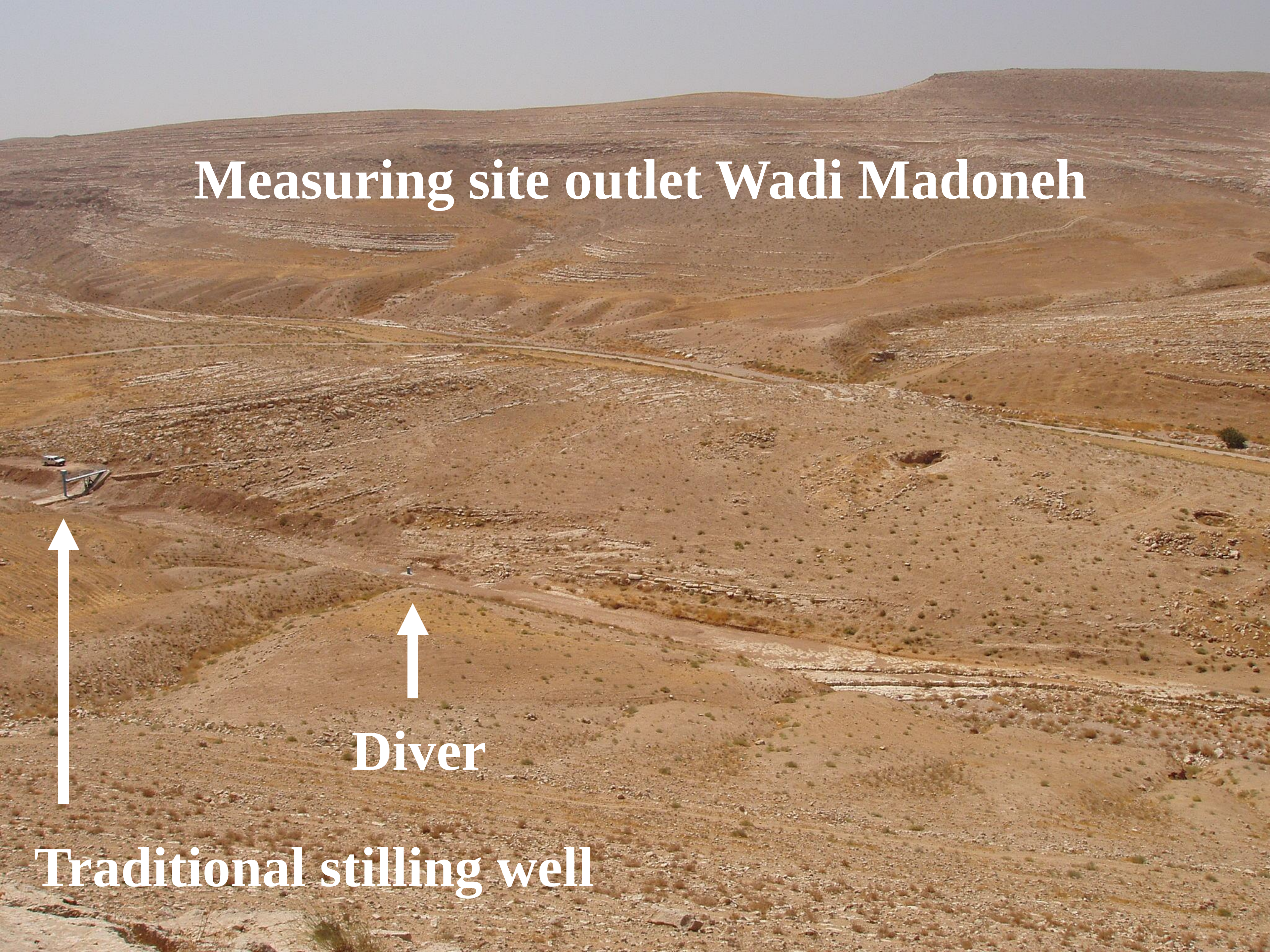
$$Q = \sum_{1}^{n} \bar{v}_i\,A_i$$

Velocity-area method

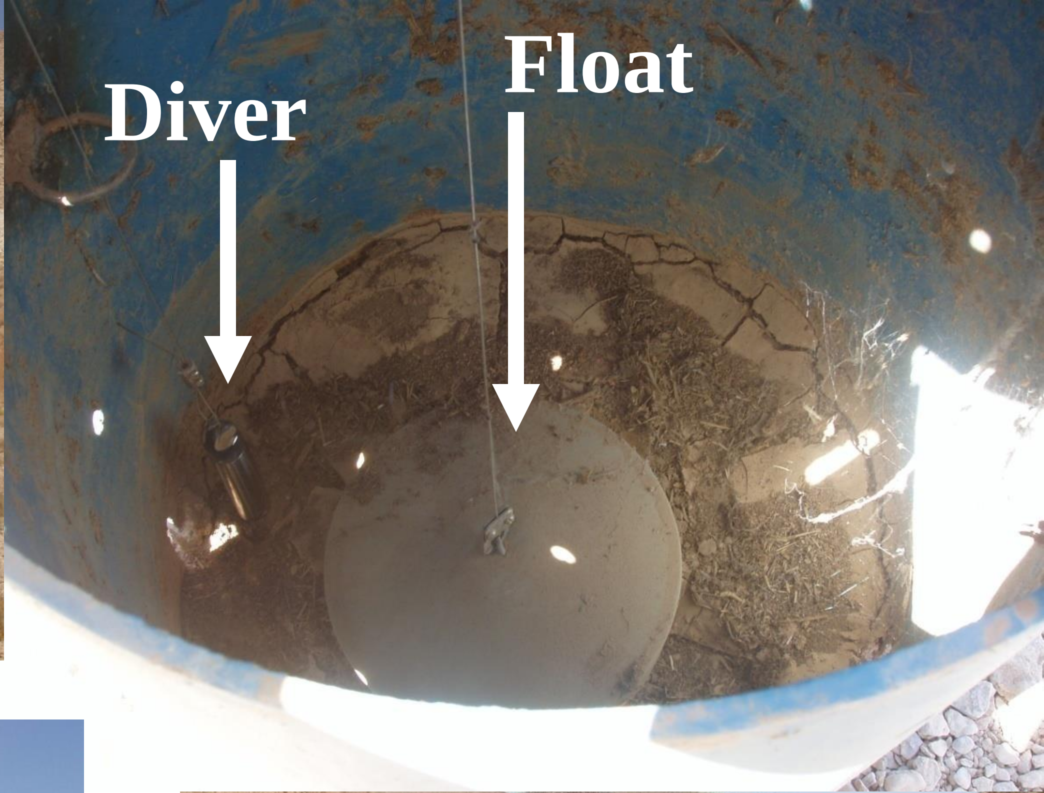
Current meter

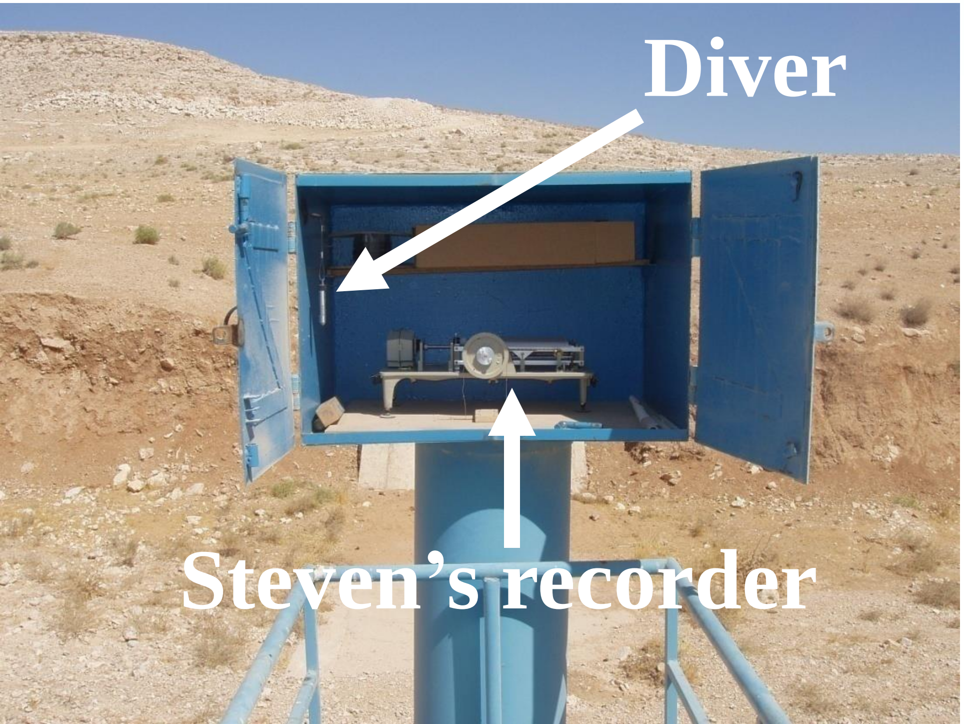
Traditional float-actuated recording gauge of river stage




Measuring site outlet Wadi Madoneh
Diver
Traditional stilling well


Traditional
stilling well


Diver
Float

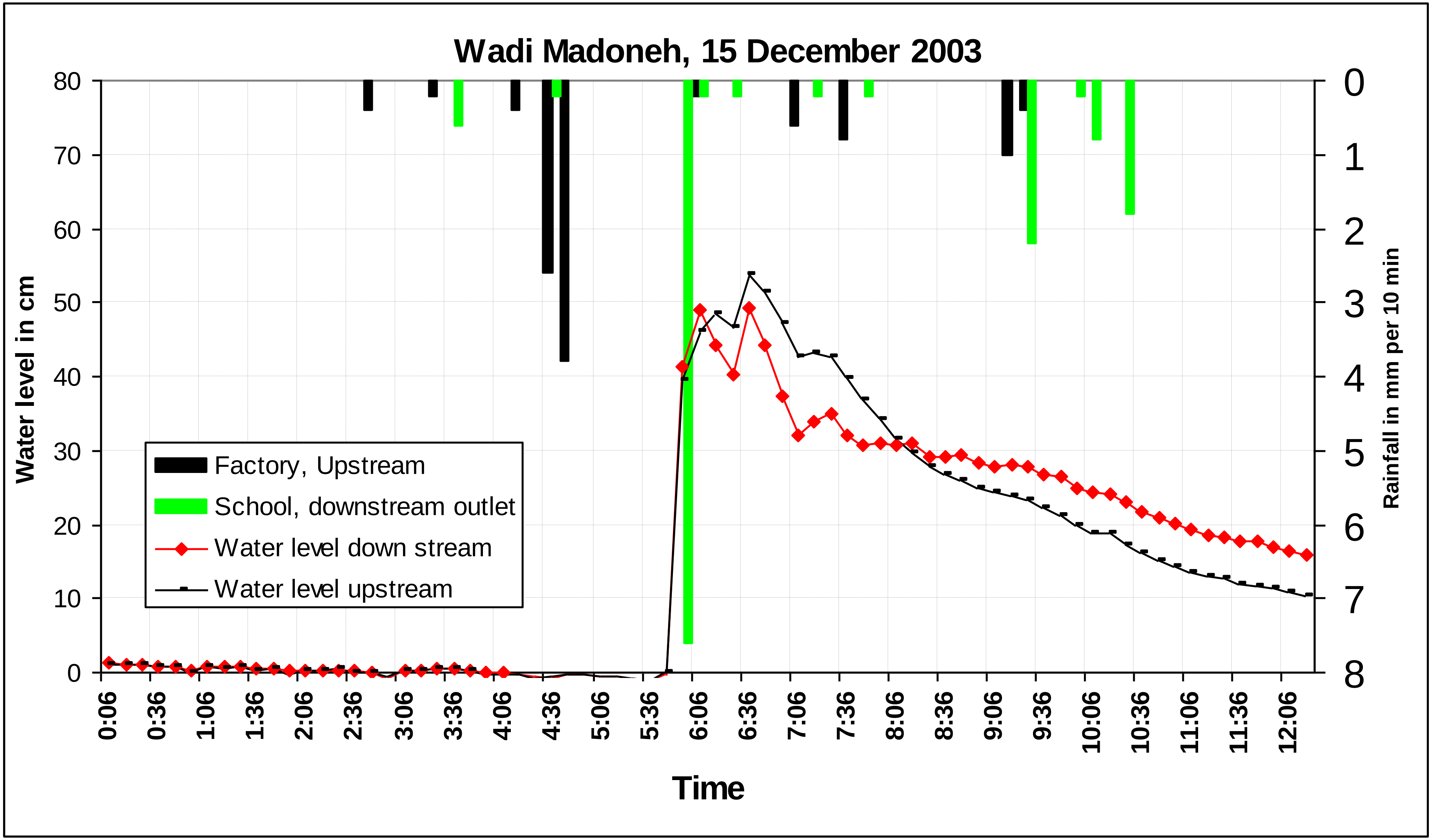

Diver
Steven’s recorder


Stilling well Diver

Wadi
Madoneh
Jordan
2008

New design water
level measuring
station

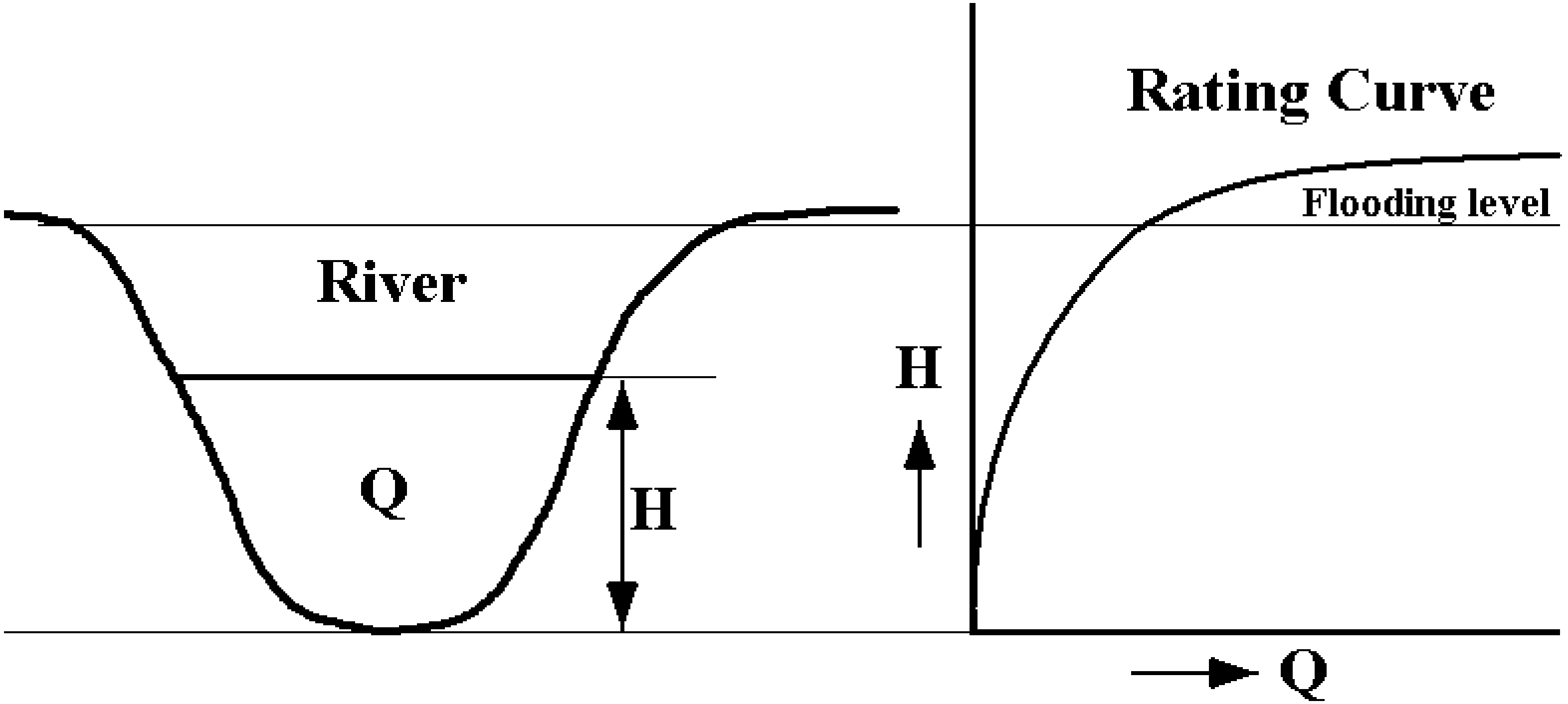
Rainfall-runoff event Wadi Madoneh (Jordan)



Rating Curve

Equation: $$Q = a\,H^{b}$$

General equation: $$Q = a\left(H - H_0\right)^{b}$$



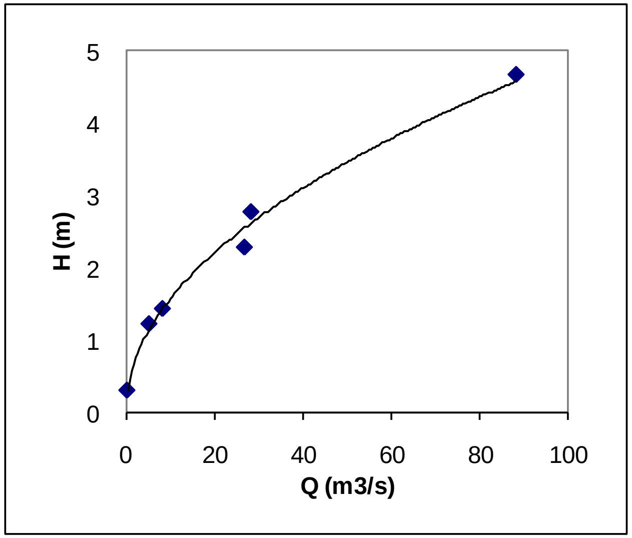
Rating Curve

$$Q = a\left(H - H_0\right)^b$$

$$\mathrm{Log}(Q) = \mathrm{Log}(a) + b\,\mathrm{Log}\left(H - H_0\right)$$

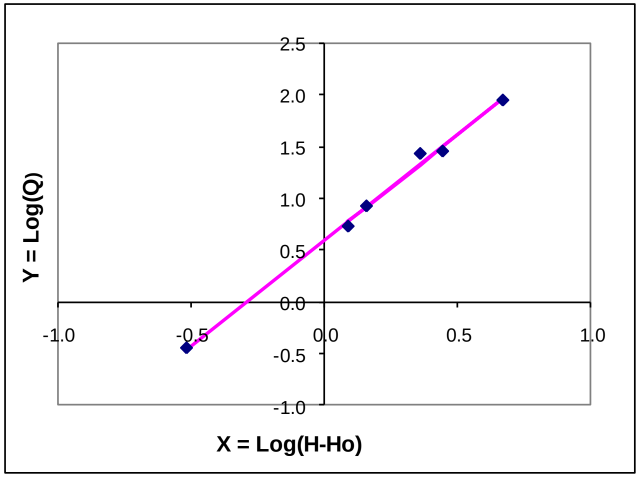
$$Y = C_0 + C_1 X$$

Find C_0 and C_1 from regression analysis
where Log(a) = C_0
b = C_1

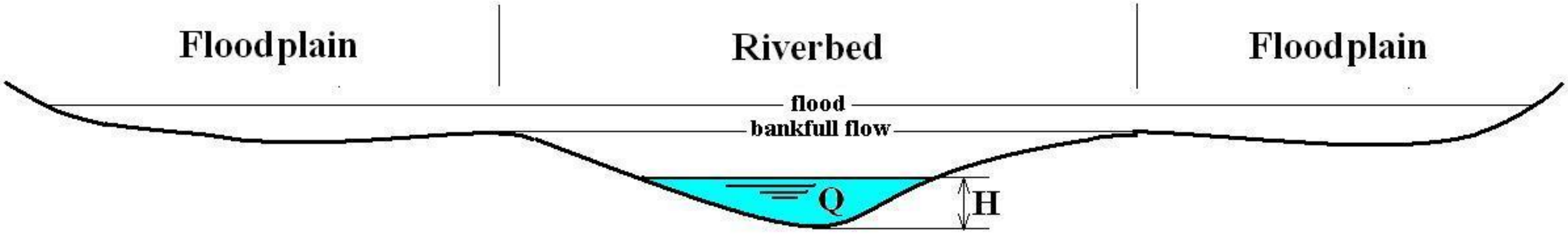


Example with $H_0 = 0$

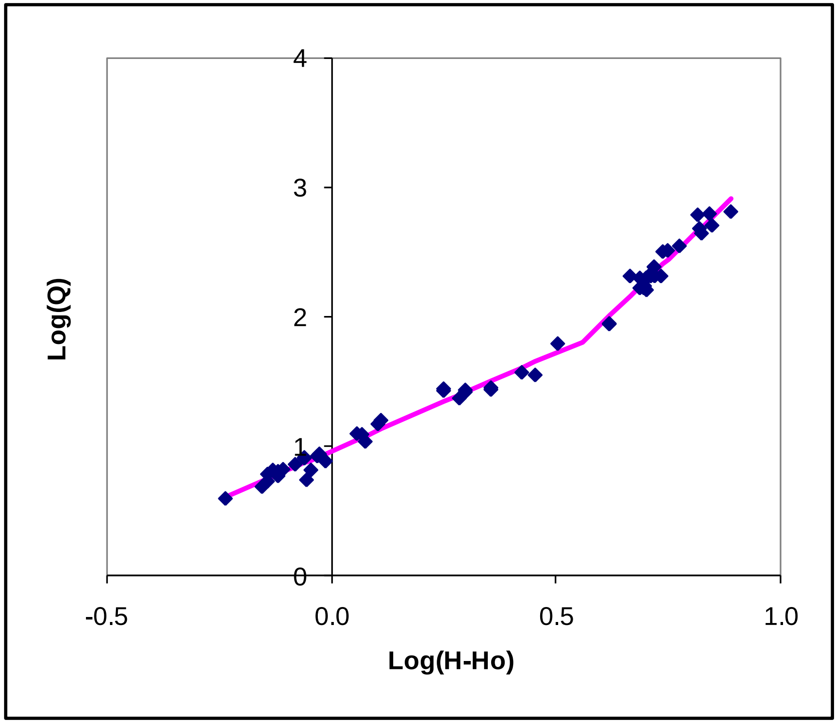
Date	H	Q	LOG(H-Ho)	LOG(Q)
3-jul-81	0.3	0.4	-0.523	-0.447
11-feb-81	4.7	88.3	0.669	1.946
18-mrt-81	2.8	28.4	0.442	1.453
27-mrt-81	2.3	26.9	0.358	1.430
30-apr-81	1.4	8.4	0.155	0.923
24-aug-81	1.2	5.3	0.086	0.728



Rating Curve

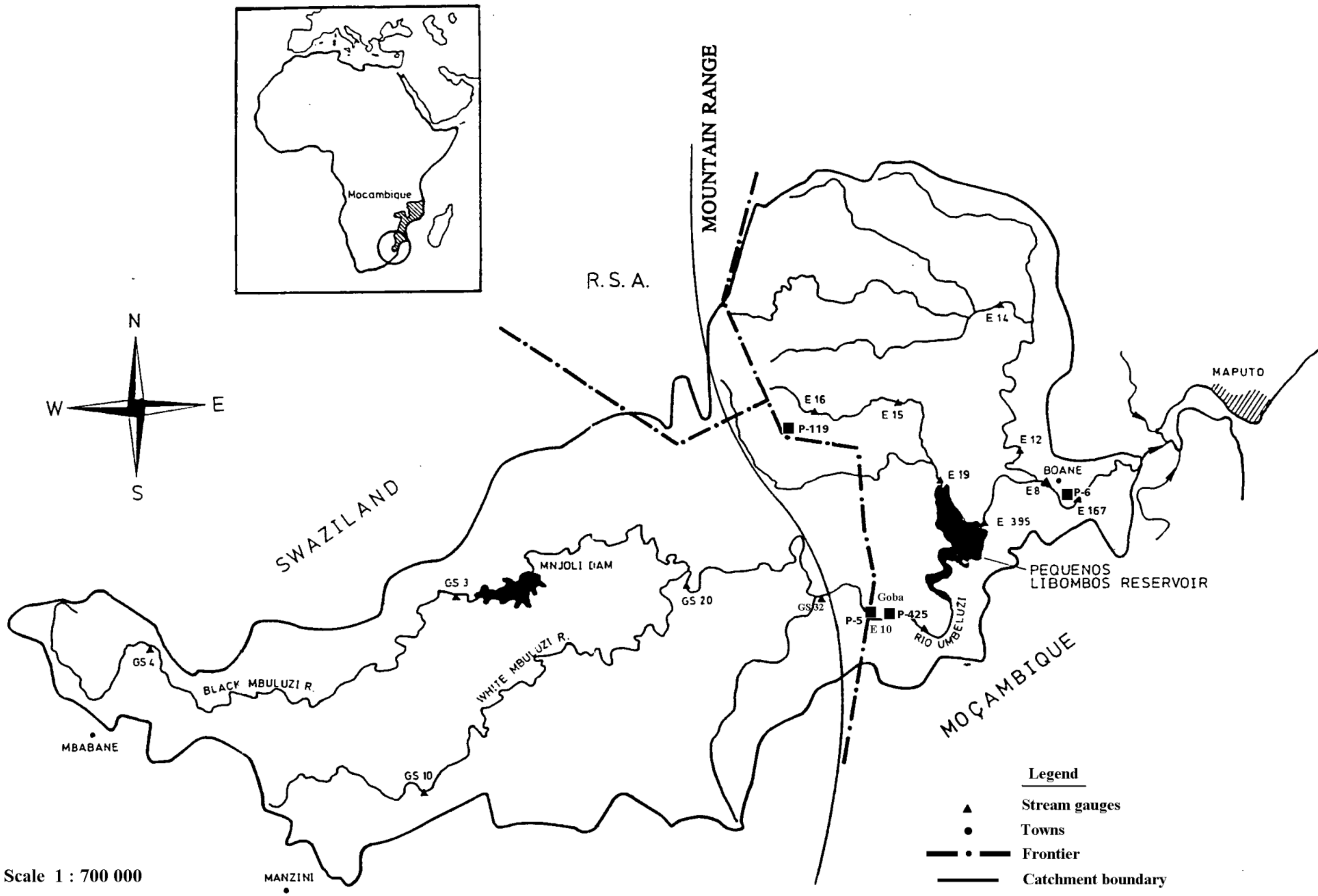


- Rating curve may change with time
- Rating curve applies over limited range of discharges
- Rating curve may be different for different ranges of discharges



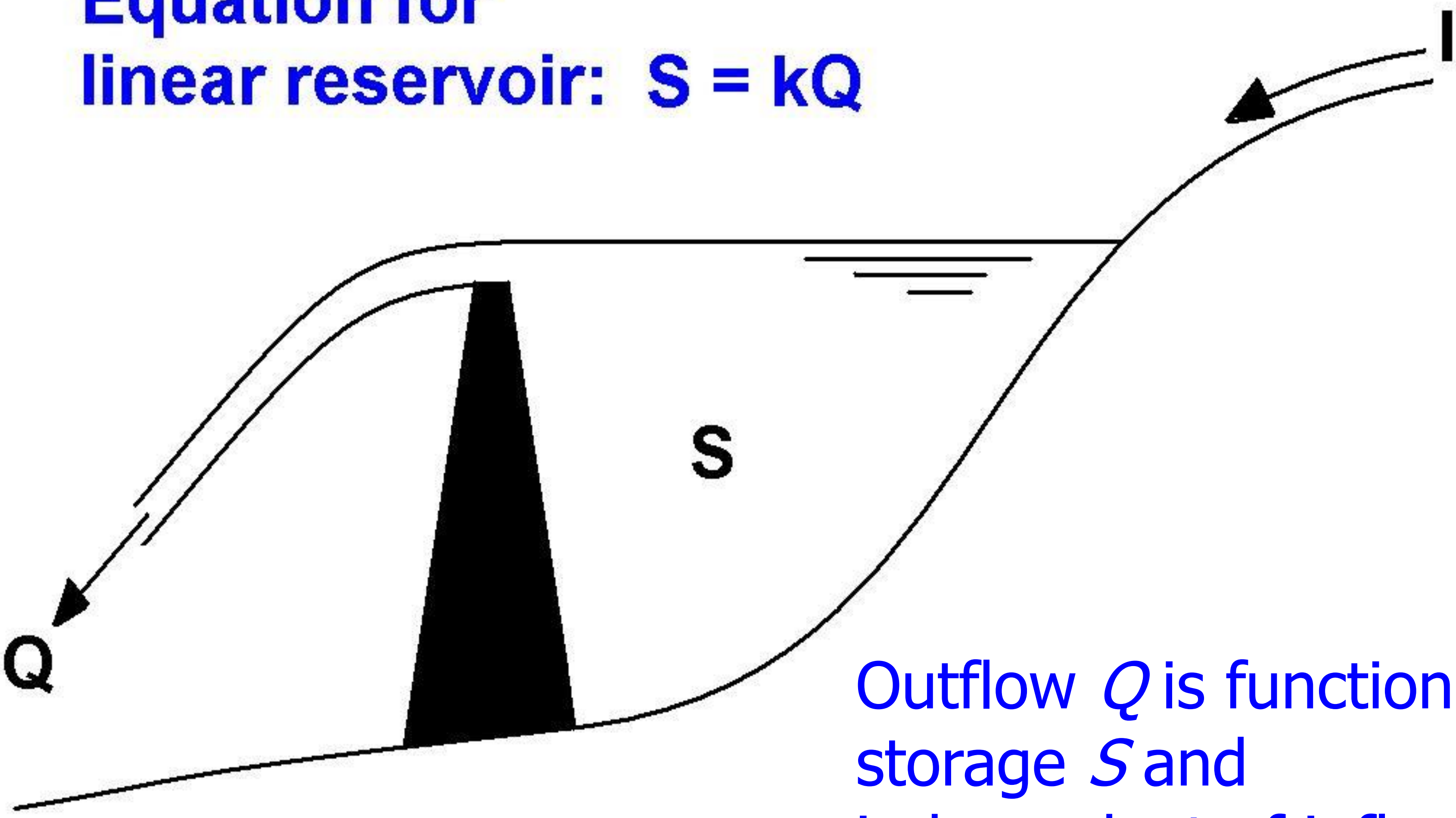
Rainfall-Runoff Modelling

- Description of Catchment
- Flood Routing
- Base Flow Separation
- Estimating Areal Rainfall
- Computation of the Φ-index
- Derivation of the Unit Hydrograph
- Predicting runoff by convoluting rainfall


Mocambique
R.S.A.
MOUNTAIN RANGE
N
W
E
S
SWAZILAND
MAPUTO
E 14
E 16
E 15
P-119
E 12
BOANE
E 19
E8
P-6
E 167
E 395
MNJOLI DAM
GS 3
GS 20
GS 32
Goba
P-5
P-425
E 10
RIO UMBELUZI
PEQUENOS
LIBOMBOS RESERVOIR
GS 4
BLACK MBULUZI R.
WHITE MBULUZI R.
MOÇAMBIQUE
MBABANE
GS 10
MANZINI
Scale 1 : 700 000
Legend
Stream gauges
Towns
Frontier
Catchment boundary
P-5
Rainfall station

Reservoir Routing

Equation for
linear reservoir: $S = kQ$



Outflow Q is function of storage S and independent of inflow I

Flood Routing (Muskingum)

Flow equation for a river reach:

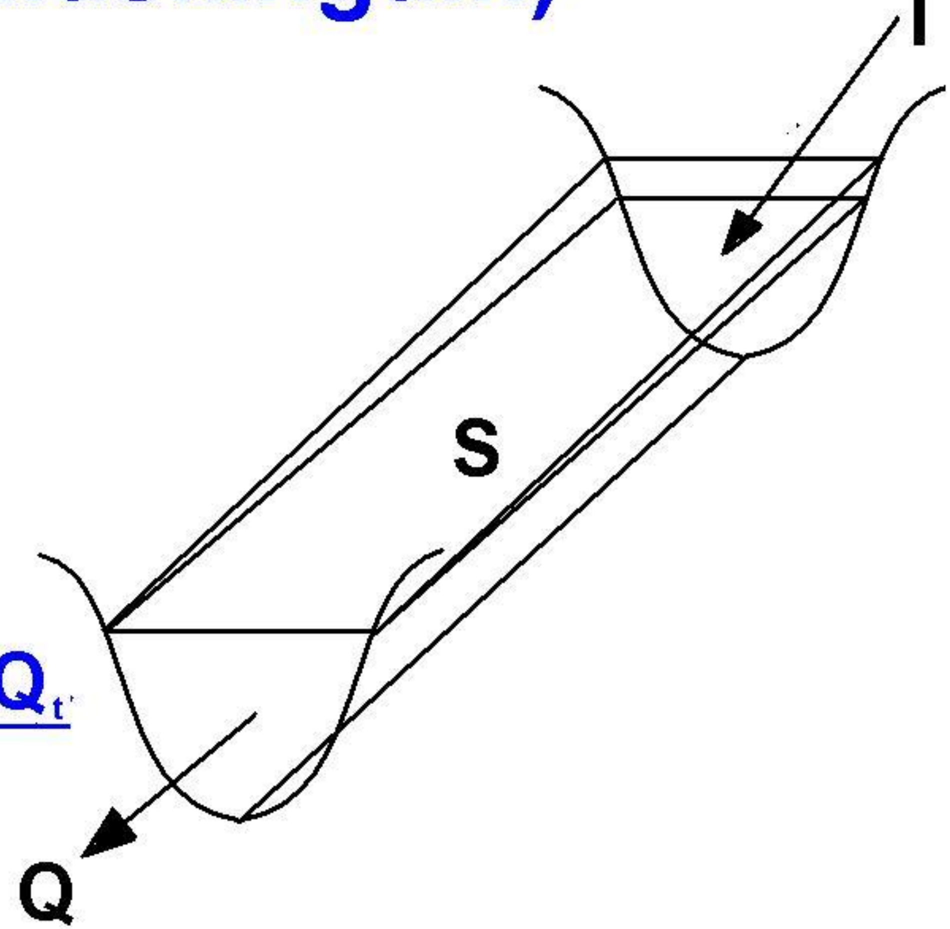
$$S = K[xI + (1-x)Q]$$

Balance equation:

$$\frac{S_{t+1} - S_t}{\Delta t} = \frac{I_{t+1} + I_t}{2} - \frac{Q_{t+1} + Q_t}{2}$$

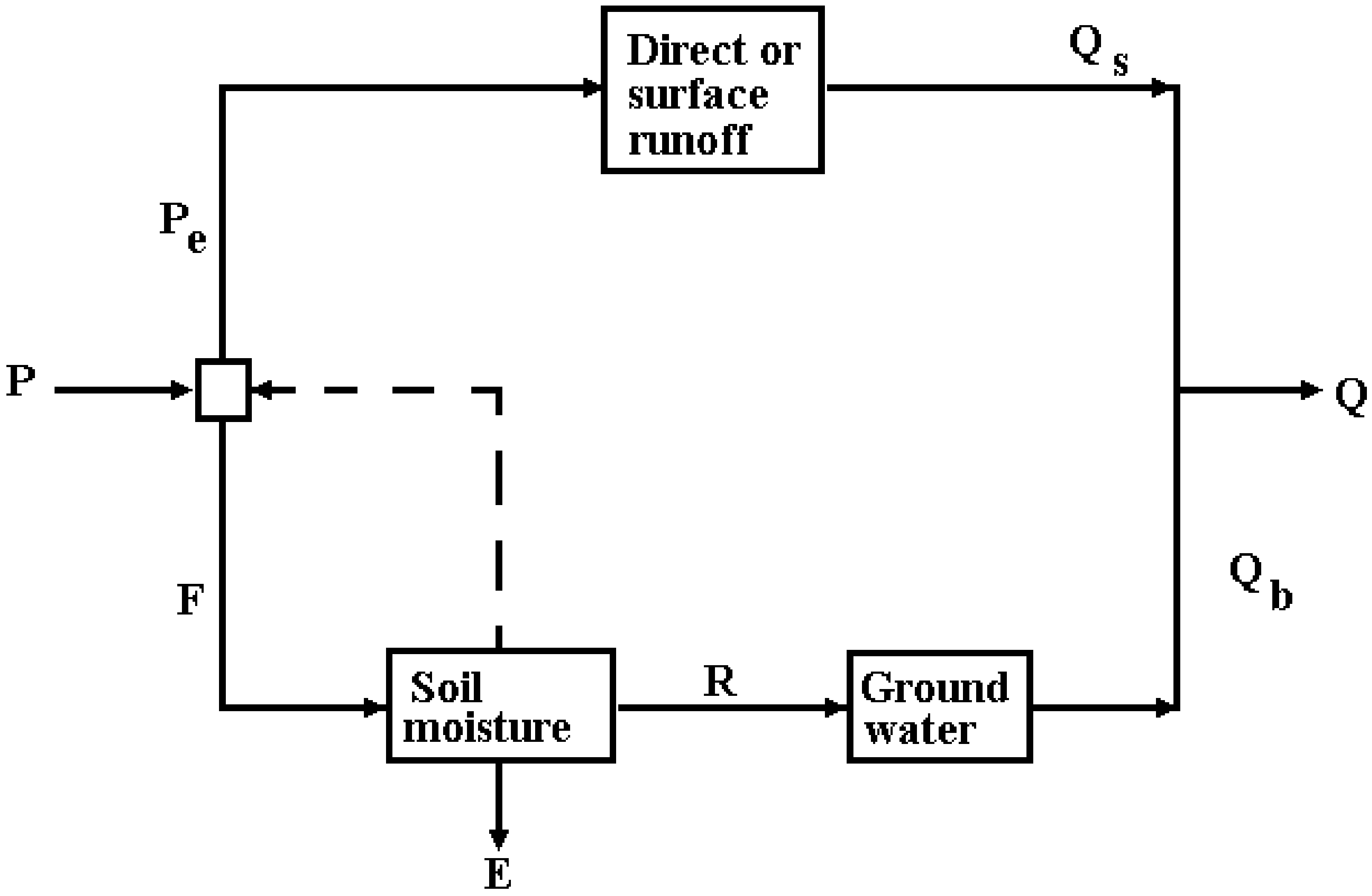
Routing equation:

$$Q_{t+1} = c_1 I_t + c_2 I_{t+1} + c_3 Q_t$$



Outflow *Q* is function of storage *S* and inflow *I*

Simplified catchment model (Dooge, 1973)



Base Flow Separation

Dry weather flow in a river can often be described similar to the depletion of a linear (groundwater) reservoir for which

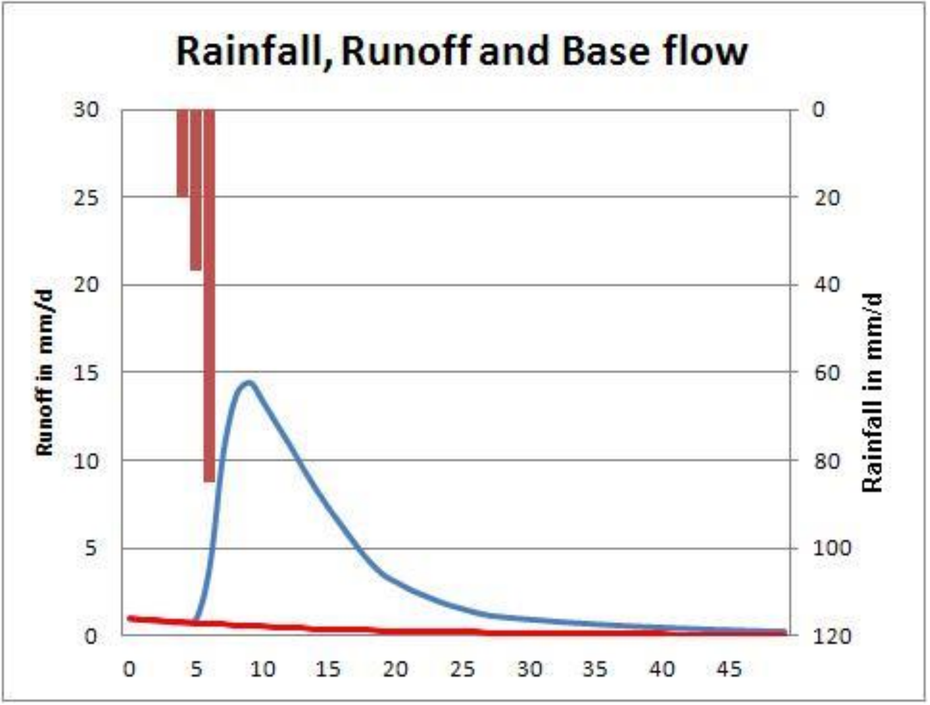
$$\boldsymbol{S = kQ}$$ where k is the reservoir parameter.

Combining this equation with the continuity equation gives

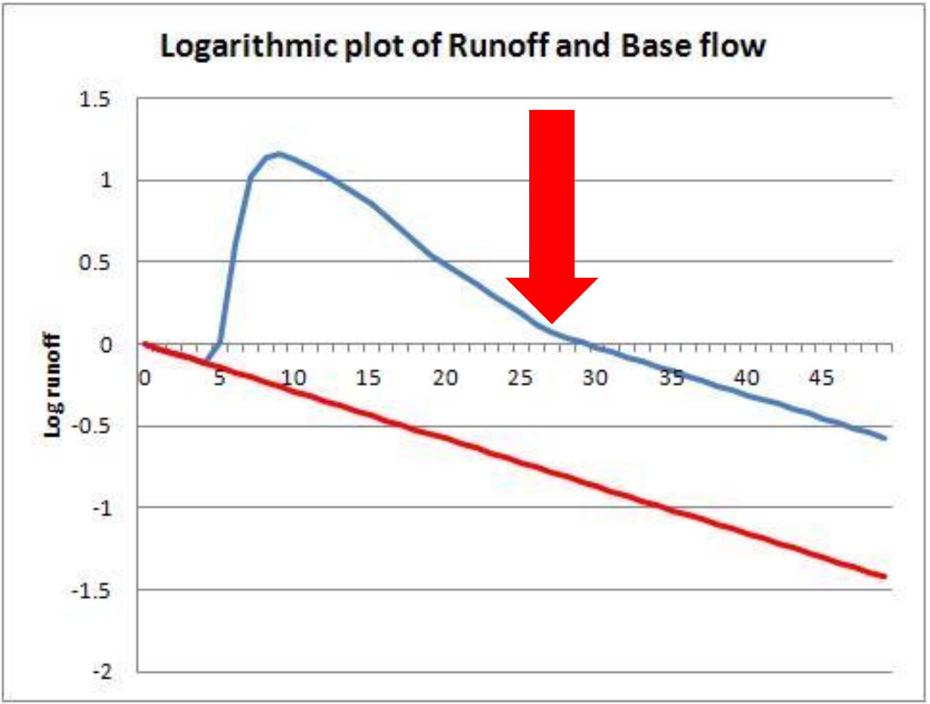
$$\boldsymbol{Q_t = Q_0\, e^{-\frac{t}{k}}}$$

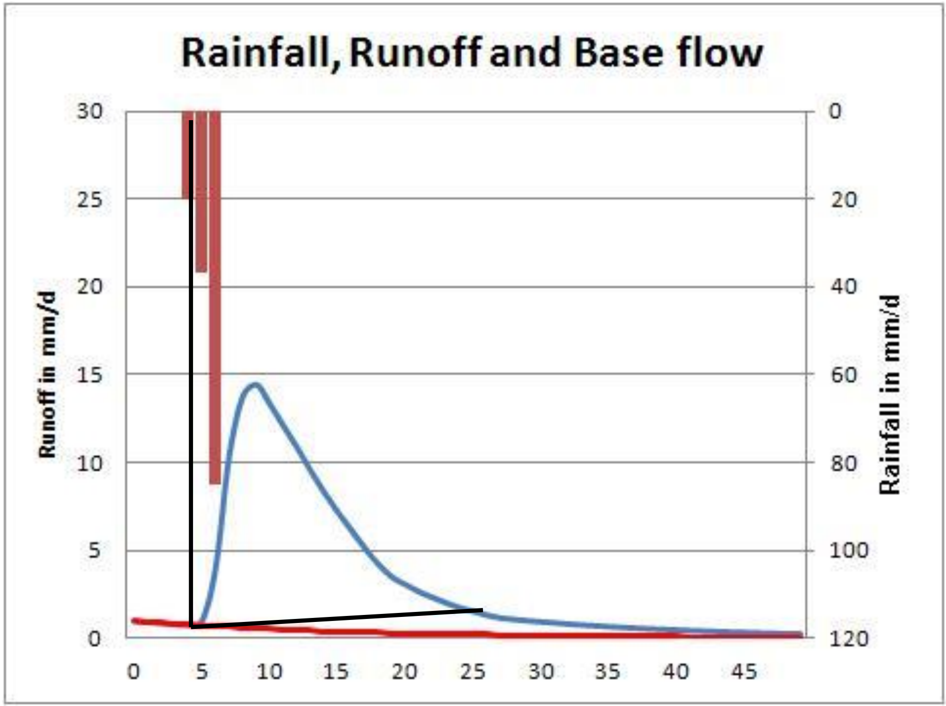
$$\boldsymbol{\ln Q_t = \ln Q_0 - \frac{t}{k}}$$

Hence, river flow Q plotted on a log-scale results in a straight line during dry weather flow, that is during the period that the flow in the river is sustained by groundwater outflow only.

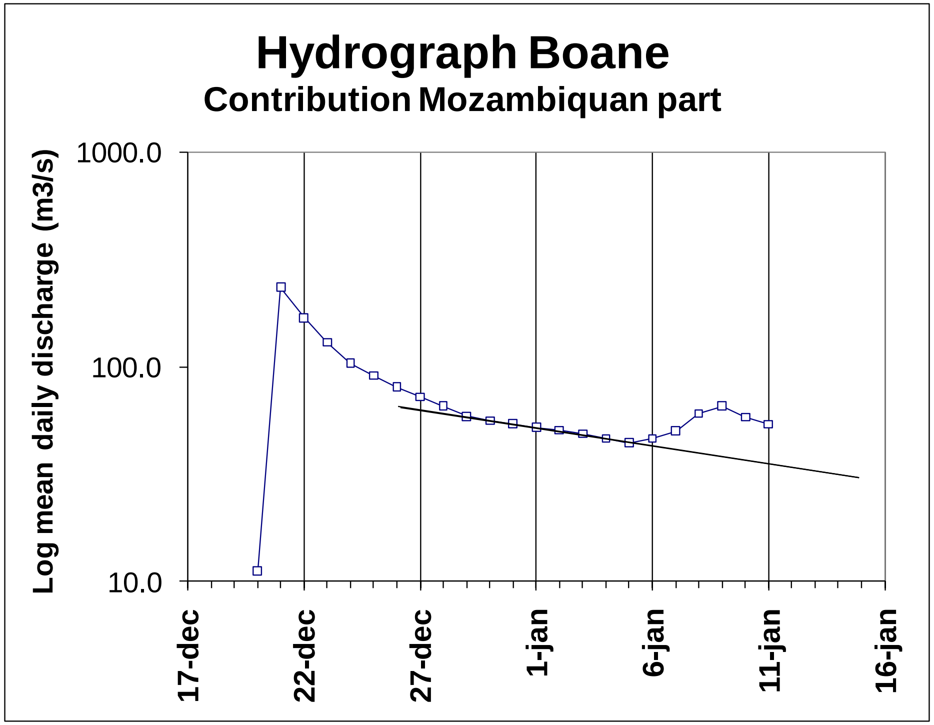


Separating Direct Runoff from the from the observed hydrograph



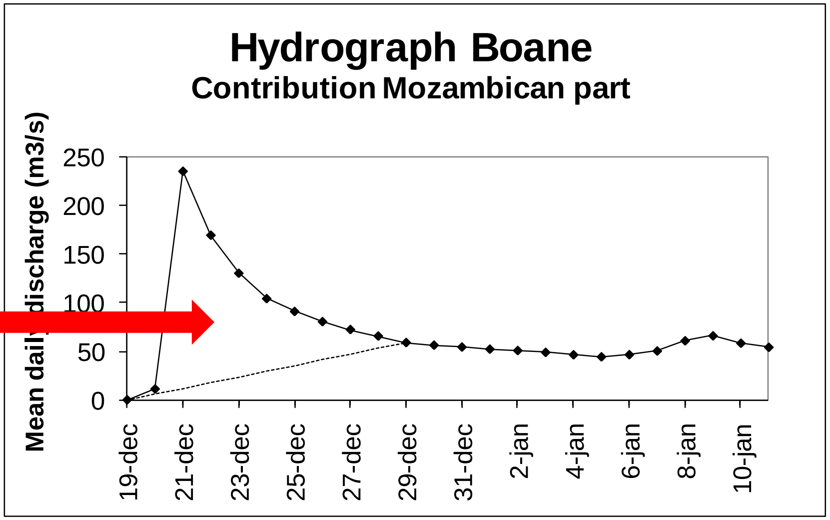


Log Q plotting shows that the depletion curve starts on 29 December



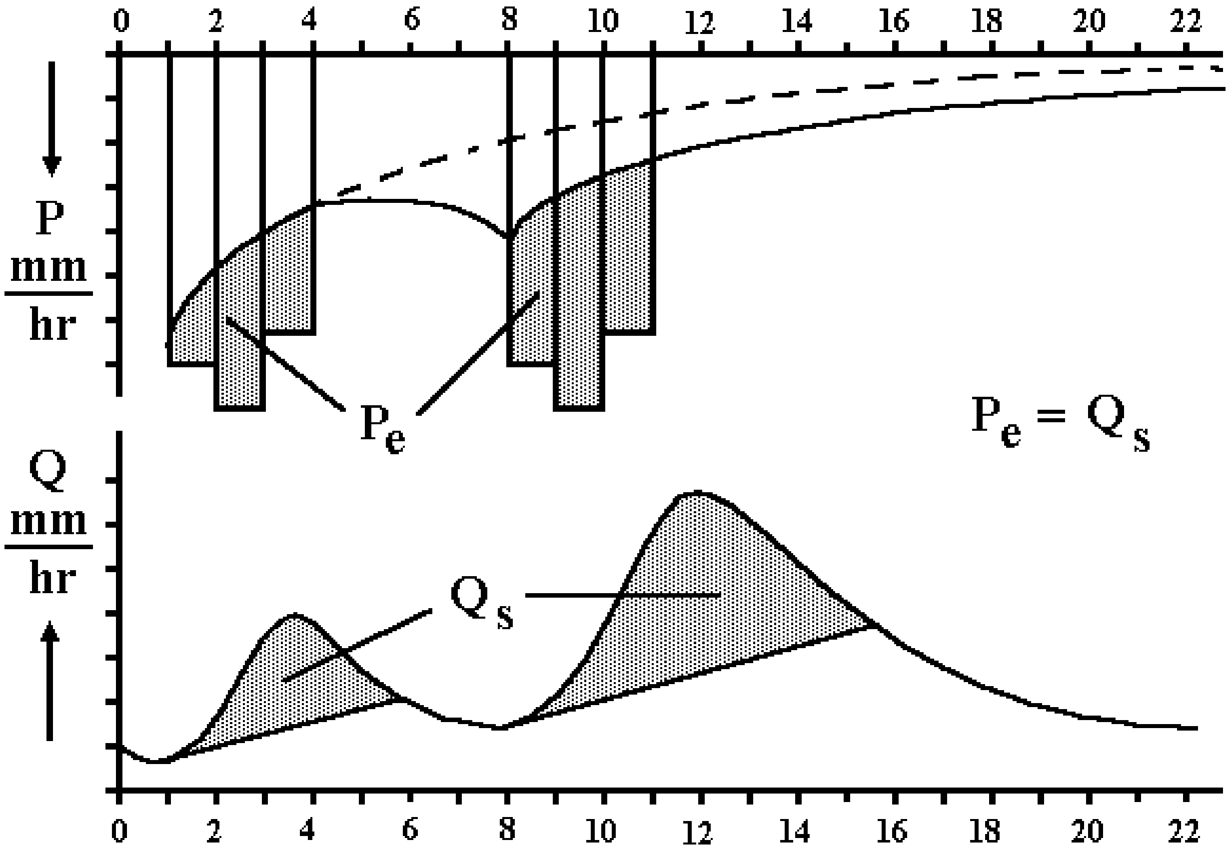
Base flow separation by straight line starts from beginning of storm until 29 December

Direct (or surface) runoff

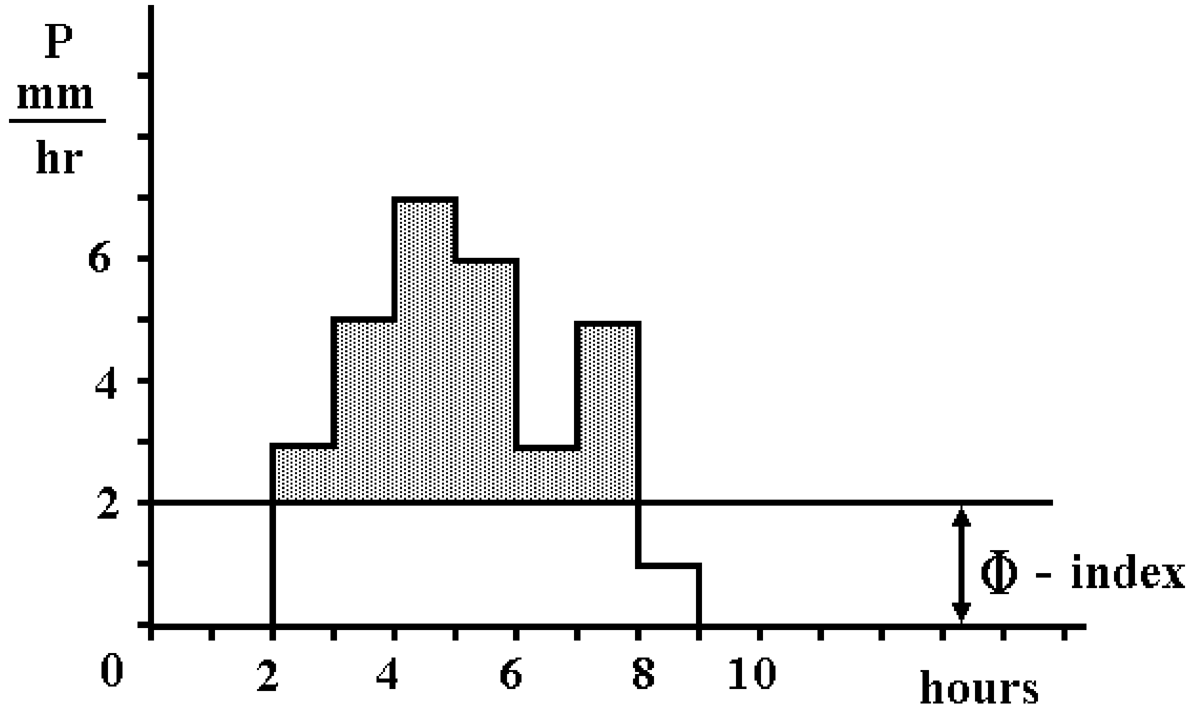


Estimating effective precipitation

$$P_e = P - F$$

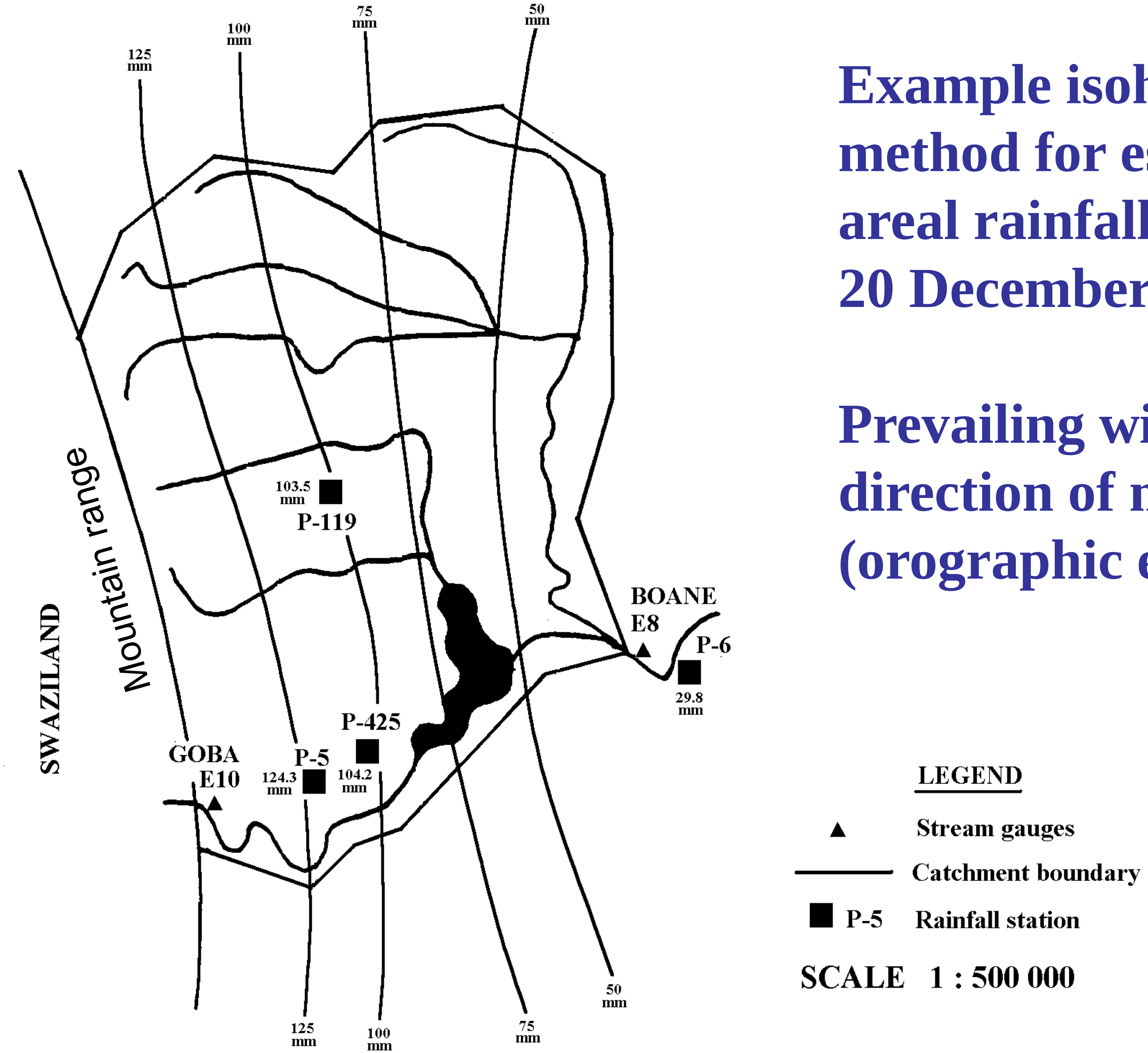


The 💣-index considers the loss rate to be constant



Find Φ-index such that Effective rainfall = Surface flow

	E10 GOBA Qm m3/s	E8 BOANE Qm m3/s	E8 Qmusk m3/s	E8 Qm-Qmusk m3/s		Baseflow m3/s	Surface flow m3/s	Areal rainfall mm	Effective rainfall mm
19-dec-73	12.7	17.1	17.1	0.00		0.00	0.00	40	0
20-dec-73	49.1	25.0	13.8	11.20	1	5.89	5.31	103	56
21-dec-73	135.0	278.0	42.9	235.10	2	11.78	223.32	62	15
22-dec-73	114.0	286.5	117.2	169.30	3	17.67	151.63	15	0
23-dec-73	104.0	244.7	114.5	130.20	4	23.56	106.64	0	0
24-dec-73	75.3	209.9	105.8	104.10	5	29.45	74.65	5	0
25-dec-73	57.5	172.0	81.1	90.90	6	35.34	55.56	11	0
26-dec-73	57.8	142.4	62.0	80.40	7	41.23	39.17		
27-dec-73	47.1	130.8	58.5	72.30	8	47.12	25.18	236	71
28-dec-73	36.0	114.8	49.2	65.60	9	53.01	12.59		
29-dec-73	33.5	97.4	38.5	58.90	10	58.90	0.00		
30-dec-73	33.5	90.7	34.5	56.20					
31-dec-73	28.2	88.0	33.7	54.30					
1-jan-74	18.1	81.3	29.2	52.10					
2-jan-74	23.5	71.0	20.3	50.70					
3-jan-74	23.4	71.7	22.9	48.80		Total =	6.0E+07	m3	
4-jan-74	30.1	69.6	23.3	46.30		Total =	71	mm	
5-jan-74	25.9	73.0	28.8	44.20					
6-jan-74	25.4	72.8	26.4	46.40					
7-jan-74	23.4	75.8	25.6	50.20				**Constant loss rate (Φ-index) =**	47
8-jan-74	20.7	84.5	23.8	60.70					
9-jan-74	19.8	87.1	21.3	65.80					
10-jan-74	19.1	78.3	20.1	58.20					
11-jan-74	17.3	73.3	19.3	54.00					

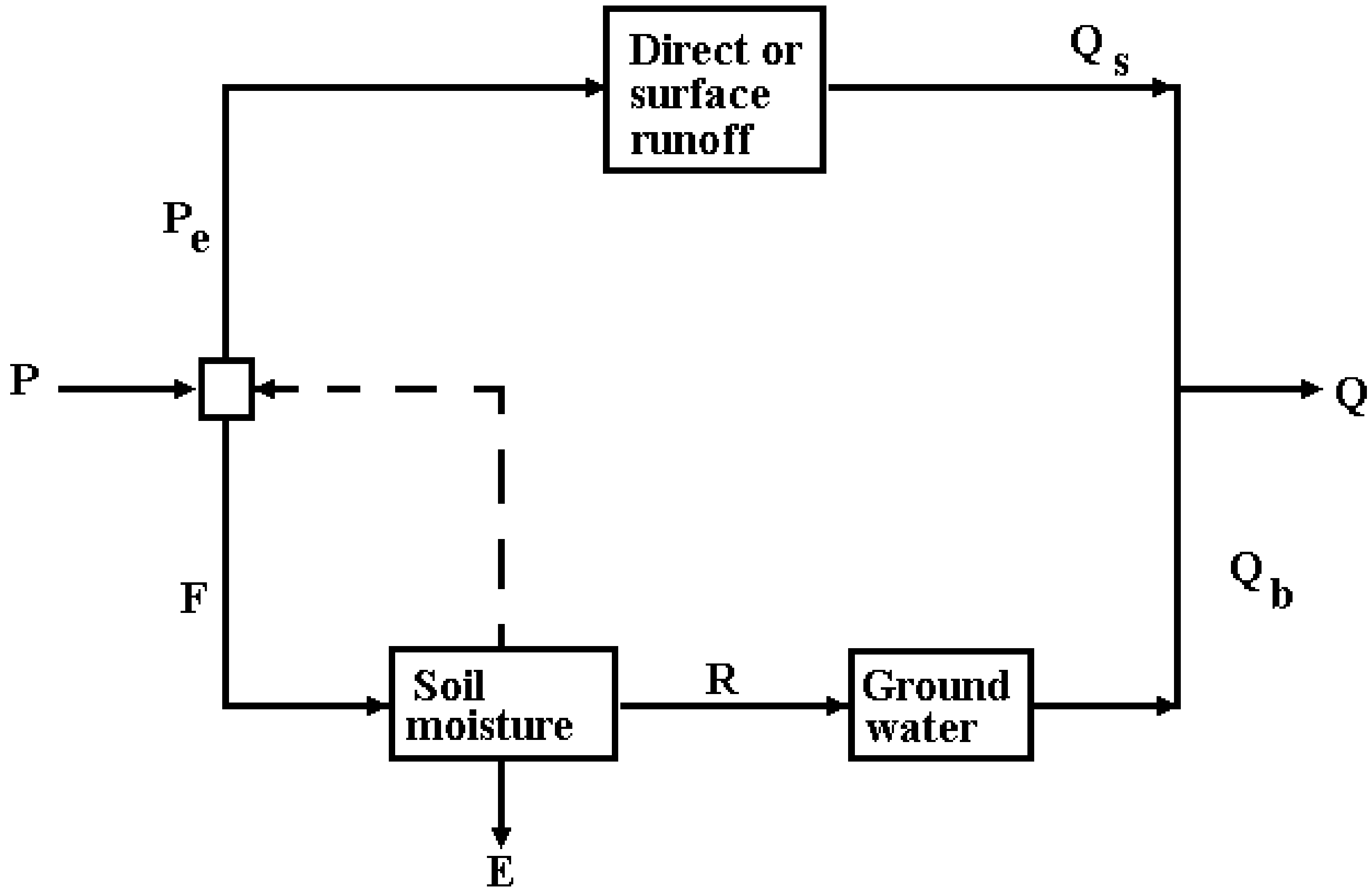


Example isohyetal method for estimating areal rainfall for 20 December 1973

Prevailing wind in direction of mountains (orographic effect)

Unit hydrograph theory

Simplified catchment model (Dooge, 1973)



The Unit Hydrograph (UH) is a *transfer function* that changes (transfers) effective precipitation P_e in surface runoff Q_s

It should be realized that

Surface runoff Q_s = Q – base flow

Losses = P – Q_s

Effective precipitation P_e = P –Losses

Hence $P_e = Q_s$

Conceptual modelling: the unit hydrograph method

Definition of UH: runoff of a catchment to a unit depth of *effective* rainfall (e.g. 1 mm) falling uniformly in space and time during a period of T (minutes, hours, days).

So, it is a lumped model, which limits its applications to catchments up to 500-1000 km^2.

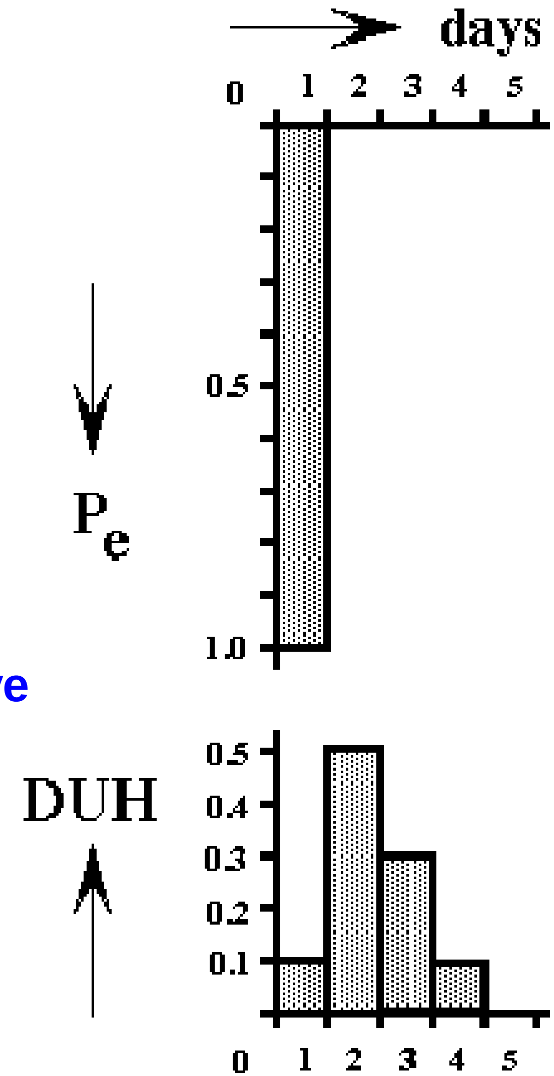
Example of Distribution Unit Hydrograph DUH

DUH ordinates result from 1 mm of effective precipitation P_e

Length DUH: 4 days

Memory of system: 3 days

Sum of DUH ≈ 1 (no losses)



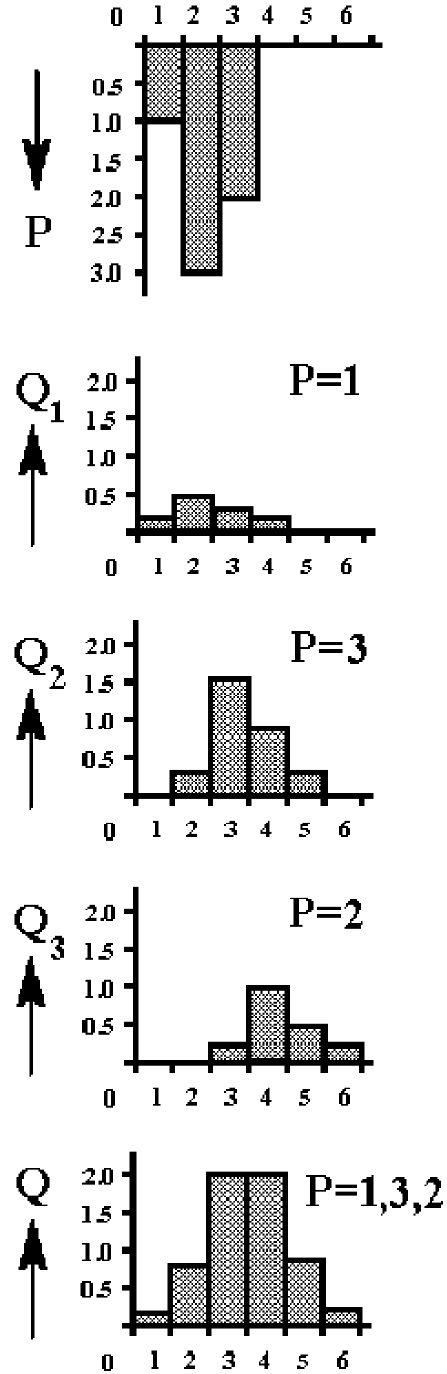
Assumptions for application of UH theory

1. **System is linear** (twice as much rainfall produces twice as much runoff

2. **System is time-invariant** (UH does not change with time)

3. **Principle of superposition applies** (runoff produced by rain on one day may be added to runoff produced by rain on the following day)

Time	1	2	3	4	5	6	7
DUH	0.1	0.5	0.3	0.1			
P	1	3	2				
Q1	0.1	0.5	0.3	0.1			
Q2		0.3	1.5	0.9	0.3		
Q3			0.2	1.0	0.6	0.2	
Q	0.1	0.8	2.0	2.0	0.9	0.2	0.0



Time	1	2	3	4	5	6	7
DUH	0.1	0.5	0.3	0.1			
P	1	3	2				
Q1	0.1	0.5	0.3	0.1			
Q2		0.3	1.5	0.9	0.3		
Q3			0.2	1.0	0.6	0.2	
Q	0.1	0.8	2.0	2.0	0.9	0.2	0.0

Solution by matrix inversion

Derivation of UH from given P and Q

$$
\begin{aligned}
Q_1 &= P_1U_1 \\
Q_2 &= P_2U_1 + P_1U_2 \\
Q_3 &= P_3U_1 + P_2U_2 + P_1U_3 \\
Q_4 &= 0 + P_3U_2 + P_2U_3 + P_1U_4 \\
Q_5 &= 0 + 0 + P_3U_3 + P_2U_4 \\
Q_6 &= 0 + 0 + 0 + P_3U_4
\end{aligned}
$$

$$
\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \end{pmatrix} = \begin{pmatrix} P_1 & 0 & 0 & 0 \\ P_2 & P_1 & 0 & 0 \\ P_3 & P_2 & P_1 & 0 \\ 0 & P_3 & P_2 & P_1 \\ 0 & 0 & P_3 & P_2 \\ 0 & 0 & 0 & P_3 \end{pmatrix} x \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}
$$

$$Y = c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4 \leftarrow$$

Solution by multiple linear regression

There are more equations than unknowns. Least squares solution of UH ordinates.

Split-record: Calibration and Validation

- **Calibration:** Part of the data set (P and Q) is used to derive the model parameters (e.g. UH)
- **Validation:** Another part of the data set is used to assess the performance of the model

Convolution is the computation of runoff from rainfall using the UH.

$$\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \end{pmatrix} = \begin{pmatrix} P_1 & 0 & 0 & 0 \\ P_2 & P_1 & 0 & 0 \\ P_3 & P_2 & P_1 & 0 \\ 0 & P_3 & P_2 & P_1 \\ 0 & 0 & P_3 & P_2 \\ 0 & 0 & 0 & P_3 \end{pmatrix} x \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}$$

Validation

Compare for the validation period the computed and observed discharge Q. Calculate e.g. R^2 (coefficient of determination) as a measure for the "goodness of fit".