



Universitas Gadjah Mada
Fakultas Teknik
Departemen Teknik Sipil dan Lingkungan
Prodi Magister Teknik Sipil

Statistika Teknik

Analisis Data *Time Series*

Analisis Data *Time Series*

- Acuan
 - Haan, C.T., 1982, *Statistical Methods in Hydrology*, 1st Ed., 3rd Printing, The Iowa State Univ. Press, Ames, Iowa, USA.
 - Chapter 14, pp 275-288

Data Time Series

■ *Time series data*

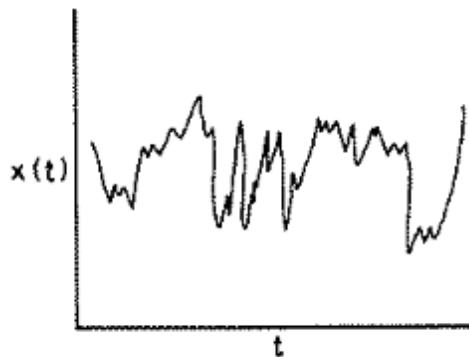
- Data yang diperoleh dari operasi (observasi, pengukuran, eksperimen) urut menurut waktu
- Data *time series* berupa
 - hasil observasi atau pengukuran pada waktu-waktu tertentu (diskret)
 - hasil perata-rataan pada suatu selang waktu
 - hasil observasi atau pengukuran secara menerus (kontinu)
- Sekumpulan *time series* adalah himpunan dari sejumlah *time series* hasil pengukuran variabel yang sama
 - Time series tunggal disebut realisasi
 - Kelompok *time series* beranggota sejumlah realisasi

Data Time Series

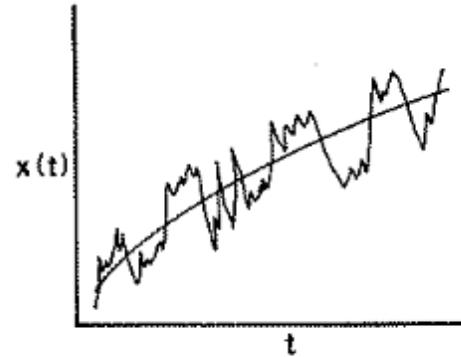
- Data *time series* dapat berasal atau bersusun dari
 - peristiwa atau kejadian yang bersifat deterministik
 - peristiwa atau kejadian yang bersifat stokastik
 - campuran peristiwa atau kejadian deterministik+stokastik
- Data *time series* hidrologi
 - umumnya berupa data komponen stokastik yang disuperposisikan pada data komponen deterministik
 - contoh
 - temperatur udara harian menunjukkan pola musiman (komponen deterministik) dan perubahan atau fluktuasi dari pola musiman, yang bersifat random (acak)

Data Time Series

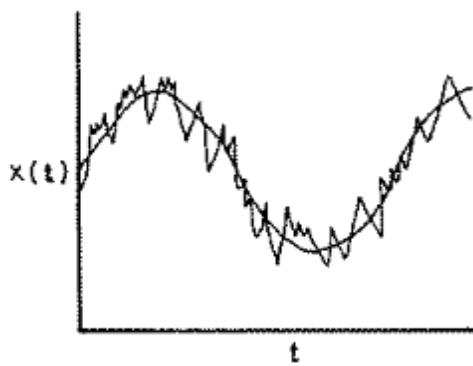




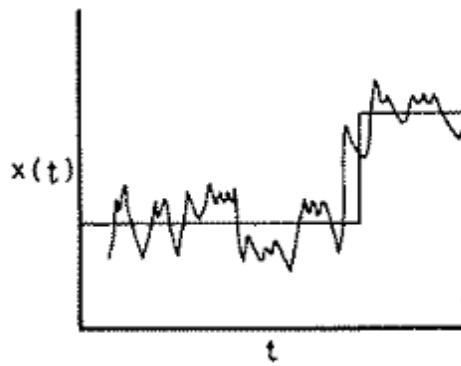
a. STOCHASTIC



b. STOCHASTIC + TREND



c. STOCHASTIC + PERIODIC



d. STOCHASTIC + JUMP

Contoh data *time series* yang terdiri dari komponen stokastik dan deterministik

Data Time Series Deterministik

- Pola, kecenderungan (*trend*)
 - Perubahan DAS yang berlangsung selama beberapa tahun → memunculkan perubahan pola debit aliran permukaan
 - Perubahan lingkungan secara alamiah dan perlahaan atau perubahan lingkungan akibat ulah manusia dapat menimbulkan perubahan pola data time series
- Loncatan (*jump*)
 - Bencana alam (gempa, kebakaran hutan)
 - Pembendungan aliran sungai oleh dam
- Periodik
 - Faktor astronomis
 - Periodik yang bersifat tahunan, bulanan, mingguan

Skala Waktu

■ Diskret

- Data yang diperoleh dari pengamatan atau pengukuran pada waktu-waktu tertentu yang dipisahkan menurut waktu Δt
- Data yang diperoleh dari pengamatan nilai atau variabel yang merupakan fungsi waktu, yang terjadi pada waktu Δt
 - hujan rerata bulanan ($\Delta t = 1$ bulan)
 - debit puncak tahunan ($\Delta t = 1$ tahun)
 - hujan harian ($\Delta t = 1$ hari)

■ Kontinu

- Data yang diperoleh dari pengamatan atau pengukuran secara menerus (kontinu)
 - muka air dari AWLR
 - curah hujan dari ARR
- Walaupun data kontinu, tetapi dalam analisis, data dibaca pada waktu-waktu tertentu
 - curah hujan dibaca per selang waktu tertentu, misal setiap 5 menit
 - curah hujan dibaca pada data puncak, selang waktu antar data tidak beraturan

AWLR dan ARR



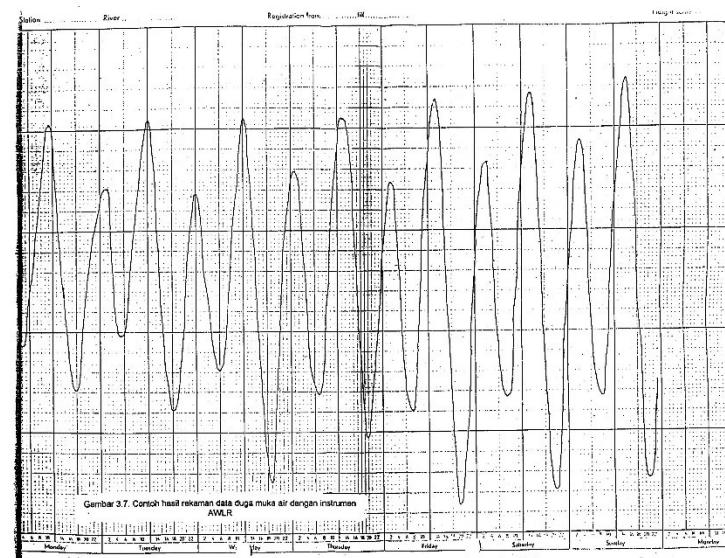
Automatic Water Level Recorder, AWLR



Automatic Rainfall Recorder, ARR

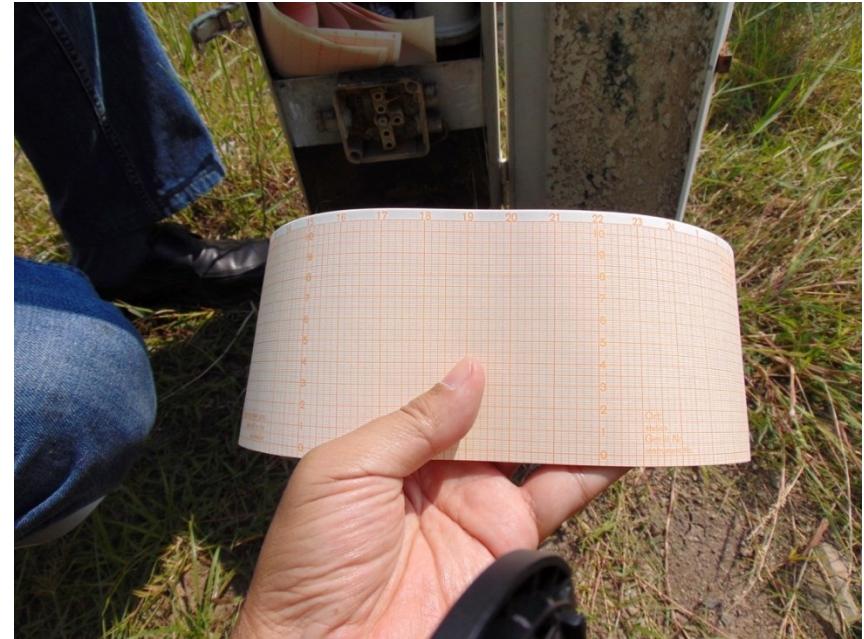


Automatic Water Level Recorder, AWLR



Untuk analisis, data dibaca setiap jam ($\Delta t = 1$ jam) atau setiap muka air ekstrem (pasang tertinggi dan surut terendah)

Data Curah Hujan, ARR



Automatic Rainfall Recorder, ARR

Skala Waktu

- Yang dibahas dalam bab ini
 - Selang waktu konstan, Δt konstan
 - Data *time series* tidak selalu merupakan fungsi waktu, namun dapat pula data hasil pengamatan atau pengukuran dalam fungsi yang lain, misal fungsi jarak/ruang (*spatial*)
 - data lebar sungai di setiap tumpang lintang
 - lebar sungai adalah variabel random
 - jarak tumpang lintang adalah variabel ruang
- Variabel random dalam data *time series*
 - Variabel random kontinu
 - kedalaman (volume) hujan per hari
 - Variabel random diskreit
 - hari hujan (1) dan hari tidak hujan (0) per hari

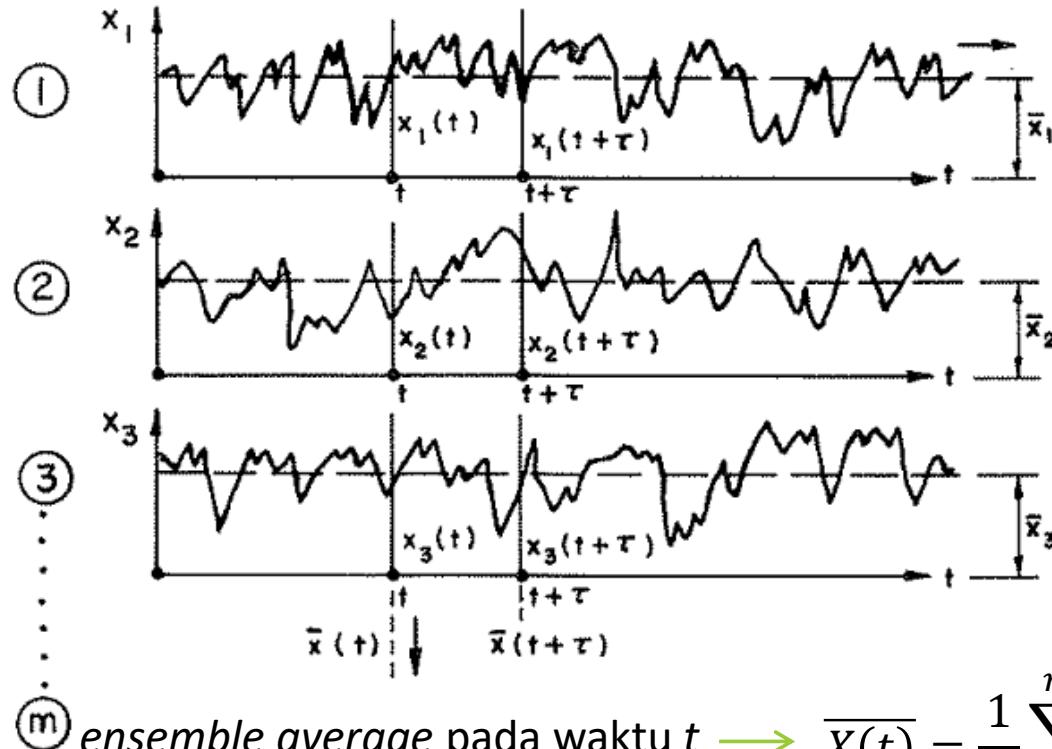
Proses Stokastik

- Proses stokastik: $X(t)$
 - pdf $X(t)$: $p(x; t) \rightarrow$ perilaku probabilistik $X(t)$ pada waktu t
- Jika sifat-sifat suatu *time series* tidak berubah terhadap waktu, maka *time series* tersebut disebut proses permanen (*stationary*)
 - *Time series* permanen: $p(x; t_1) = p(x; t_2), t_1 \neq t_2$
 - *Time series* tak-permanen: $p(x; t_1) \neq p(x; t_2)$

Proses Stokastik

- Sifat-sifat *time series* dapat diperoleh dari atau didasarkan pada
 - realisasi tunggal selama suatu selang waktu → dikenal sebagai *time average properties*
 - beberapa realisasi pada waktu tertentu → dikenal sebagai *ensemble properties*
- Apabila *time average properties* = *ensemble properties*, maka *time series* tsb memiliki sifat *ergodic*

Proses Stokastik



time average properties
 realisasi ke- i selama selang
 waktu 0 s.d. T

$$\bar{X}_i = \frac{1}{T} \int_0^T X_i(t) dt$$

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_i(t_j)$$



ensemble average pada waktu t $\rightarrow \bar{X}(t) = \frac{1}{m} \sum_{i=1}^m X_i(t)$ $\bar{X}(t) = \int_{-\infty}^{\infty} X(t)p(x; t)dx$

Proses Stokastik

$\overline{X(t)} = \overline{X(t + \tau)}$, $\forall t$ dan τ  proses stokastik bersifat permanen untuk nilai rerata
(*stationary in the mean, first-order stationary*)

$$\text{cov}(X(t), X(t + \tau)) = \frac{1}{m} \sum_{i=1}^m (X_i(t) - \overline{X(t)})(X_i(t + \tau) - \overline{X(t + \tau)})$$

 kovarian $X(t)$ dan $X(t+\tau)$
jika $\tau = 0 \rightarrow$ varian *time series*

Jika proses stokastik memiliki sifat permanen (*stationary*) untuk nilai rerata dan kovarian, maka *time series* tsb memiliki sifat *second-order stationary*

Proses Stokastik

- Untuk *time series* yang bersifat *ergodic*
 - nilai rerata waktu (*time average mean*) sama dengan nilai rerata bersama (*ensemble average*)
 - hal di atas berlaku pula untuk nilai-nilai rerata waktu yang lain, tidak hanya *mean*
- Oleh karena itu
 - sifat-sifat proses random (stokastik) permanen dapat diukur dari data historis tunggal (realisasi tunggal)
 - kadang, *data time series* realisasi tunggal dipecah menjadi beberapa *time series* pendek

Proses Stokastik

- Untuk suatu proses random dengan realisasi tunggal, $i = 1$

$$\bar{X}_i = \frac{1}{T} \int_0^T X_i(t) dt \quad \text{nilai rerata, realisasi tunggal, variabel random kontinu}$$

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^{0_n} X_i(t_j) \quad \text{nilai rerata, realisasi tunggal, variabel random diskret}$$

$$\text{cov}(X_i(t), X_i(t + \tau)) = \frac{1}{T - \tau} \int_0^{T-\tau} (X_i(t) - \bar{X}_i)(X_i(t + \tau) - \bar{X}_i) dt \quad \text{var random kontinu}$$

$$\text{cov}(X_i(t), X_i(t + \tau)) = \frac{1}{n - 1} \sum_{j=1}^{0_n} (X_i(t_j) - \bar{X}_i)(X_i(t_j + \tau) - \bar{X}_i) \quad \text{var random diskret}$$

Autokorelasi

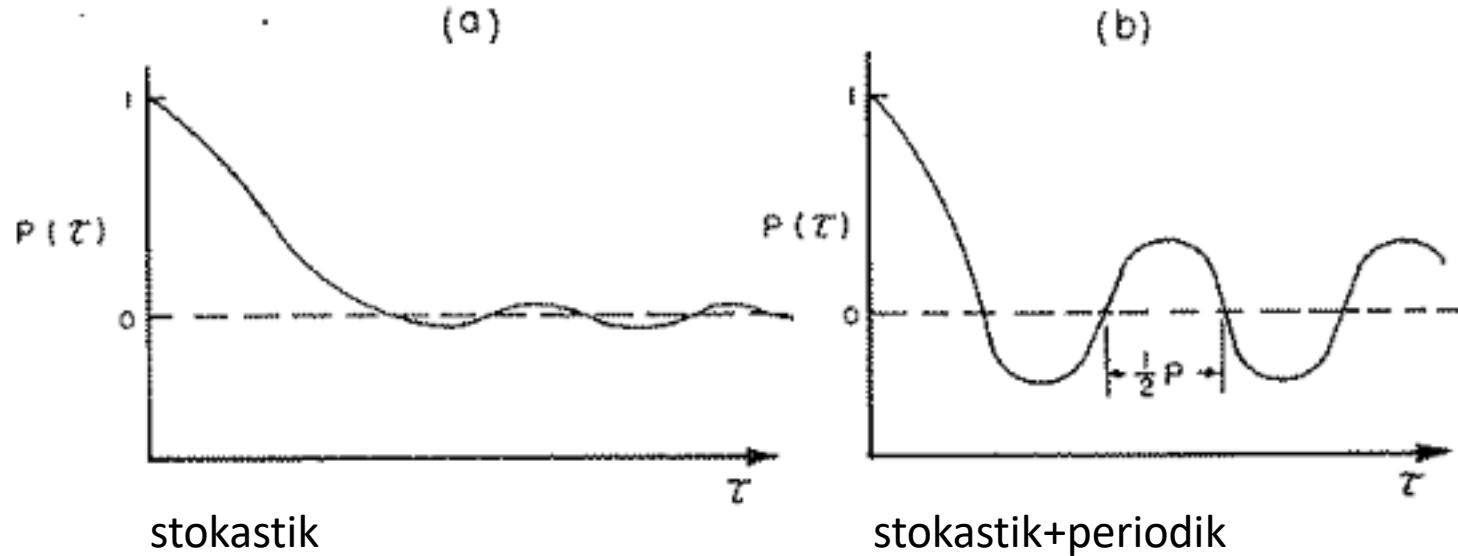
- Yang dibahas adalah *time series* ergodic, sehingga hanya diperlukan realisasi tunggal, $i = 1$ saja
- Autokorelasi (*autocorrelation*), $\rho(\tau)$

$$\rho(\tau) = \frac{\text{cov}(X(t), X(t + \tau))}{\text{var}(X(t))}$$

- Untuk $\tau = 0$, maka $\rho(\tau) = 1$ karena $\text{cov}(X(t), X(t + \tau)) = \text{var}(X(t))$
- Jika τ kecil, maka $\rho(\tau)$ positif
Jika τ bertambah besar, maka $\rho(\tau)$ negatif

Autokorelasi

- Plot fungsi autokorelasi vs selang waktu τ disebut koreogram



Autokorelasi

- Koreogram
 - berguna untuk mengetahui jika data yang berurutan tersebut *independent*
 - jika koreogram menunjukkan adanya korelasi yang kuat antara $X(t)$ dan $X(t+\tau)$, maka data tidak *independent*
- Autokorelasi dengan demikian
 - menunjukkan “memori” proses stokastik
 - jika $\rho(\tau) = 0$, proses dikatakan tidak memiliki memori terhadap kejadian sebelum $t - \tau$
 - pada prinsipnya untuk sebagian besar proses random, $\rho(\tau)$ haruslah sama dengan nol untuk τ besar
 - jika $\rho(\tau)$ untuk τ besar menunjukkan suatu pola yang tidak sama dengan nol, maka hal ini mengindikasikan suatu komponen deterministik

Autokorelasi

- Untuk skala waktu diskrit, fungsi autokorelasi menjadi $\rho(k)$, k adalah jumlah selang waktu yang memisahkan $X(t)$ dan $X(t+\tau)$
- Hubungan antara τ dan k

$$\tau = k\Delta t$$

- Δt adalah panjang selang waktu, misal 1 hari, 1 bulan, 1 tahun, dsb.

Autokorelasi

- Jika $\rho(k) = 0$ untuk semua $k \neq 0$
 - proses disebut sebuah proses random murni
 - hal ini menunjukkan data saling *linearly independent*
- Jika $\rho(k) \neq 0$ untuk sejumlah $k \neq 0$
 - data yang terpisah $k \Delta t$ adalah *dependent*
 - proses disebut sebuah proses random
- Jika sebuah time series bersifat tak permanen (*nonstationary*)
 - $\rho(k) \neq 0$ untuk semua $k \neq 0$ karena adanya komponen deterministik
 - Jika komponen deterministik tidak dihilangkan terlebih dulu, maka kita tidak dapat menentukan sampai seberapa jauh $\rho(k) \neq 0$ akan dipengaruhi oleh komponen deterministik

Analisis Spektral

- Autokorelasi
 - *Time series* dalam domain waktu
- Analisis spektral
 - *Time series* dalam domain frekuensi
- *Time series*
 - Sampel dari suatu populasi yang dicirikan oleh keragaman dalam suatu spektrum frekuensi kontinu
 - Sampel random dari suatu proses menurut waktu, *temporal* (atau ruang, *spatial*) yang tersusun dari oskilasi semua frekuensi yang mungkin terjadi

Analisis Spektral

■ Analisis spektral

- Spektrum keragaman (*a variance spectrum*) → membagi keragaman (*variance*) menjadi sejumlah rentang frekuensi
- Variabel yang umumnya dipakai dalam analisis adalah kerapatan spektral (*spectral density*)
- *Spectral density*
 - jumlah varian per rentang frekuensi

■ Beberapa istilah, variabel

- Frekuensi, $f [T^{-1}]$
- Frekuensi sudut (*angular frequency*), $\omega [\text{rad } T^{-1}]$
- Periode, T atau $p [T]$
- *Spectral density*, S



$$\omega = \frac{2\pi}{T} = \frac{2\pi}{p} = 2\pi f$$
$$S(f) = \frac{S(\omega)}{2\pi}$$

Analisis Spektral

- Hubungan antara fungsi kerapatan spektral dan fungsi autokorelasi

$$S(t) = \int_{-\infty}^{\infty} \rho(\tau) \exp(-i2\pi f\tau) d\tau = 2 \int_0^{\infty} \rho(\tau) \cos(2\pi f\tau) d\tau$$

- Transformasi Fourier

$$\rho(\tau) = \int_{-\infty}^{\infty} S(f) \cos(2\pi f\tau) df$$

- Untuk $\tau = 0$, $\rho(0) = 1$ dan $\cos(0) = 1$ yang menunjukkan bahwa:

$$\int_{-\infty}^{\infty} S(f) df = 1$$

$S(f)$ dapat dipandang sebagai *probability density function* (pdf) yang memberikan kontribusi terhadap varian tak berdimensi (*normalized variance*) dalam rentang frekuensi dari f_1 s.d. f_2 .

Analisis Spektral

- Kontribusi terhadap varian yang diberikan oleh $S(f)$

$$\int_{f_1}^f S(f) df$$

- Jika autokorelasi dihitung sbg $\text{Cov}(X(t), X(t+\tau))$, maka $\rho(0) = \text{Var}(X(t))$

Ingin: autokorelasi
$$\rho(\tau) = \frac{\text{cov}(X(t), X(t + \tau))}{\text{var}(X(t))}$$

Analisis Spektral

- Data time series hidrologi
 - Umumnya berupa tabel variabel sebagai fungsi waktu
 - Selang waktu adalah diskret, bukan kontinu
 - Oleh karena itu
 - Spektral harus disesuaikan untuk mengakomodasikan sejumlah diskret frekuensi; kepada frekuensi inilah varian akan didistribusikan
 - Untuk data yang diukur pada selang waktu Δt seragam
 - Oskilasi data yang memiliki frekuensi tertinggi yang dapat memberikan informasi mengenai data tsb adalah oskilasi data yang memiliki frekuensi sbb:

$$f_N = \frac{1}{2\Delta t} \quad \text{Frekuensi Nyquist}$$

Analisis Spektral

- Kerapatan spektral sampel

- dihitung dengan memakai $r(k)$ dan integrasi persamaan $S(f)$

$$\hat{S}'(f) = \Delta t \left[r(0) + 2 \sum_{k=1}^{m-1} r(k) \cos(2\pi kf\Delta t) + r(m) \cos(2\pi mf\Delta t) \right]$$

- m adalah jumlah maksimum *lag* korelasi {jumlah maksimum selang waktu untuk menghitung $r(k)$ }
 - m sebaiknya tidak melebihi 10% s.d. 25% jumlah data pada sampel
 - persamaan di atas dipakai untuk menghitung kerapatan spektral sampel untuk frekuensi:

$$f = \frac{k f_N}{m}$$

Analisis Spektral

- Nilai kerapatan spektral sampel tersebut perlu dihaluskan
 - Nilai estimasi kerapatan spektral adalah:

$$\hat{S}(0) = 0.5 \left(\hat{S}'(0) + \hat{S}'(f_N/m) \right)$$

$$\hat{S}(k f_N/m) = 0.25 \hat{S}'((k-1) f_N/m) + 0.5 \hat{S}'(k f_N/m) 0.25 \hat{S}'((k+1) f_N/m)$$

$$k = 1, 2, \dots, m-1$$

$$\hat{S}(f_N) = 0.5 [\hat{S}'((m-1) f_N/m) + \hat{S}'(f_N)]$$

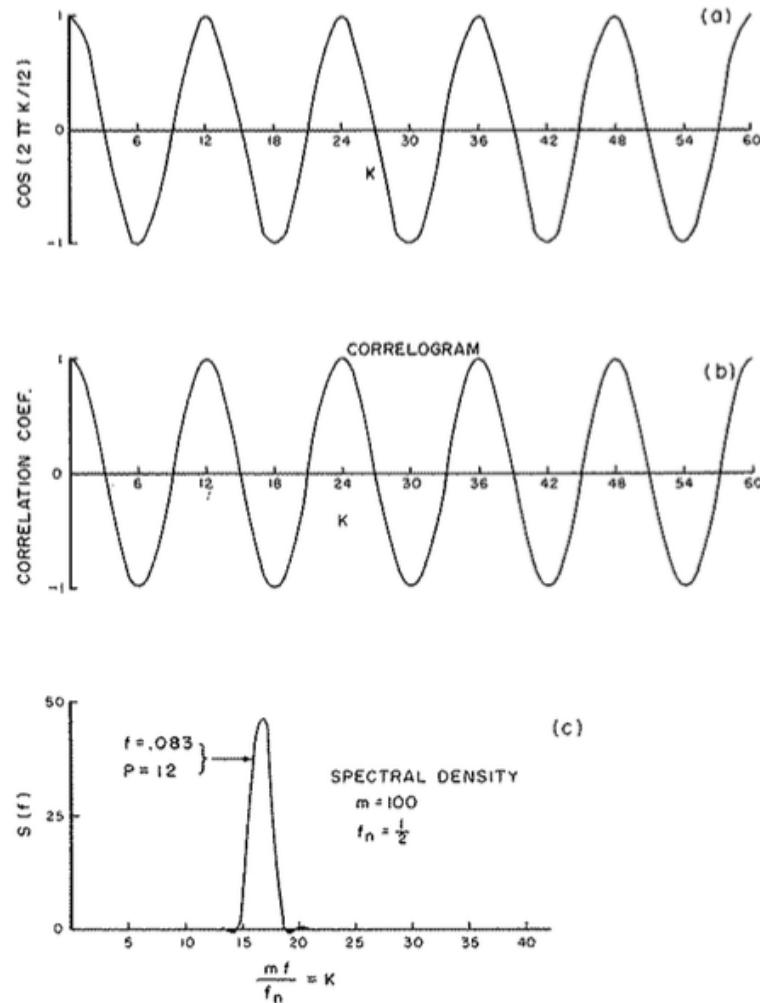


Fig. 14.4. (a) $\cos(2\pi k)/12$ and its correlogram (b) and spectral density (c).

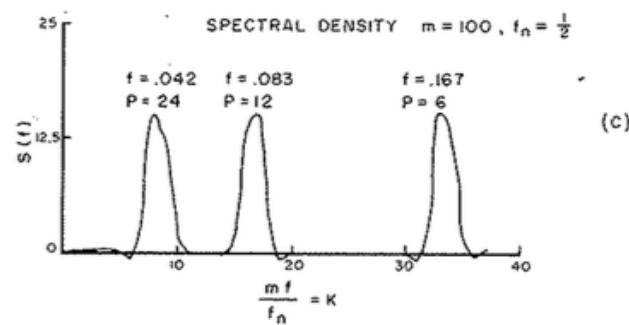
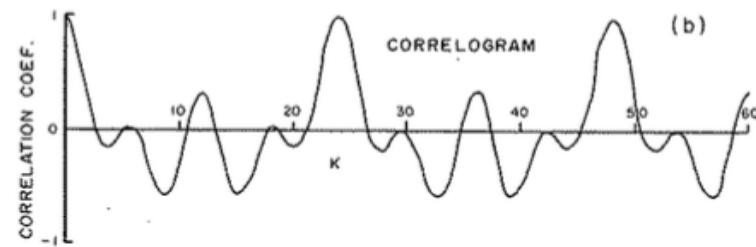
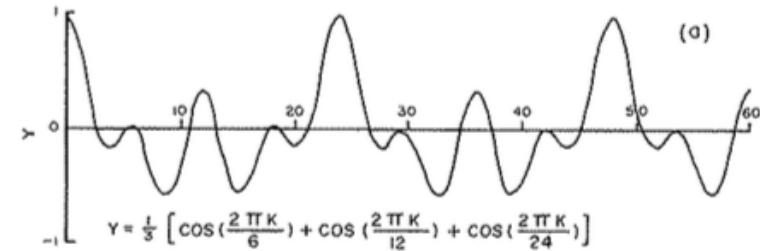


Fig. 14.5. (a) Sum of 3 cosines and its correlogram (b) and spectral density (c).

Table 2.1. Peak discharges (cfs), Kentucky River, near Salvisa, Kentucky
 (McCabe 1962)

1895	47,300	1917	111,000	1939	84,300
96	54,400	18	71,700	40	45,000
97	87,200	19	96,100	41	28,400
98	65,700	20	92,500	42	46,000
99	91,500	21	34,100	43	80,400
1900	53,500	22	69,000	44	55,000
01	67,800	23	73,400	45	72,900
02	70,000	24	99,100	46	71,200
03	66,900	25	79,200	47	46,800
04	34,700	26	62,600	48	84,100
05	58,000	27	93,700	49	61,300
06	47,000	28	68,700	50	87,100
07	66,300	29	80,100	51	70,500
08	80,900	30	32,300	52	77,700
09	80,000	31	43,100	53	44,200
10	52,300	32	77,000	54	20,600
11	58,000	33	53,600	55	85,000
12	67,200	34	70,800	56	82,900
13	115,000	35	89,400	57	88,700
14	46,100	36	62,600	58	60,200
15	52,400	37	112,000	59	40,300
16	94,300	38	44,000	60	50,500

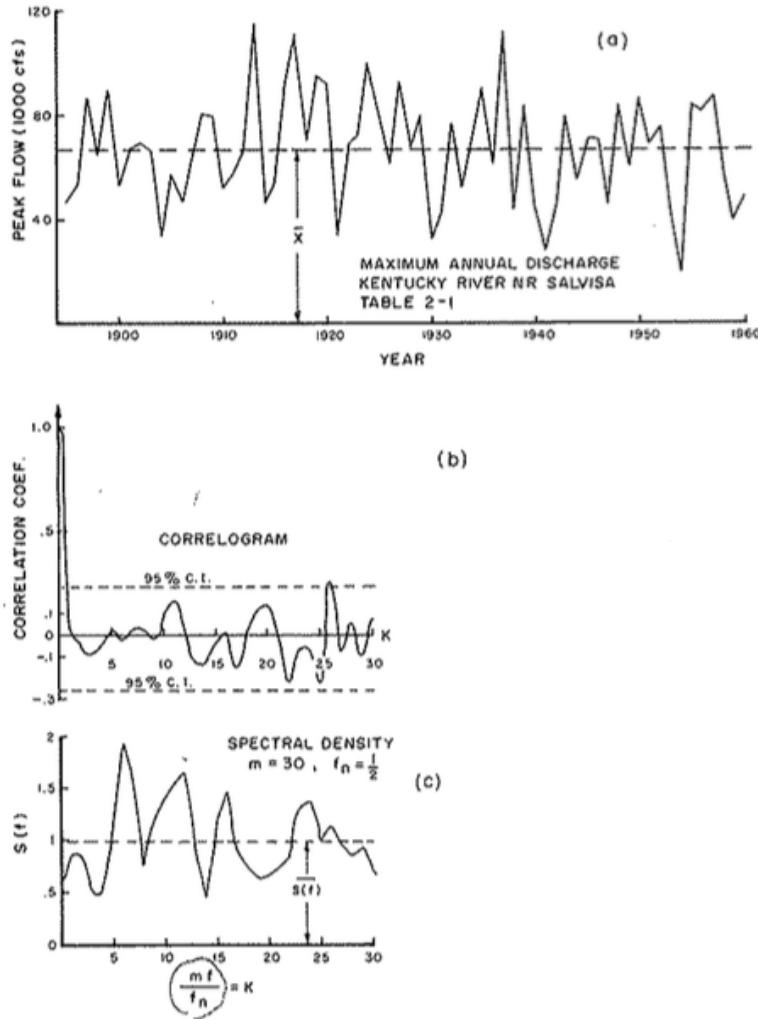


Fig. 14.6. (a) Kentucky River peak flows and its correlogram (b) and spectral density (c).

Monthly Runoff¹ Cave Creek near Fort Spring, Kentucky

Water

Year	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May ²	Jun	Jul	Aug	Sep	Total
1953	2	5	19	240	86	416	147	354	31	18	7	1	1326
1954	0	2	4	54	22	40	139	35	8	7	6	14	331
1955	2	4	30	73	463	579	59	197	55	24	28	3	1517
1956	4	6	13	59	637	469	192	28	32	64	38	8	1550
1957	7	10	172	308	325	103	392	68	24	6	5	2	1422
1958	3	106	432	200	221	117	235	236	19	369	170	12	2120
1959	6	9	17	270	195	112	102	24	24	5	4	2	770
1960	3	36	269	219	313	291	68	19	364	138	14	30	1764
1961	12	52	79	204	295	532	476	414	159	48	18	4	2291
1962	2	6	76	346	401	508	330	79	96	30	8	7	1889
1963	39	141	124	150	146	548	52	25	14	29	11	3	1282
1964	1	4	3	87	173	788	45	21	11	8	2	16	1158
1965	15	7	347	276	230	449	146	31	8	5	1	2	1517
1966	4	2	2	48	281	79	202	332	25	14	41	11	1040
1967	7	119	357	97	161	466	50	476	33	14	15	7	1802
1968	9	38	271	135	98	425	238	199	91	29	75	16	1625
1969	14	22	112	278	216	73	237	74	40	27	66	17	1177
1970	7	25	91	130	389	291	568	206	38	14	6	27	1792

1. The entries should be divided by 100 to get monthly flow in inches. 2 = .02 inches, 240 = 2.40 inches, etc. Data from U.S.G.S. Water Supply Papers.

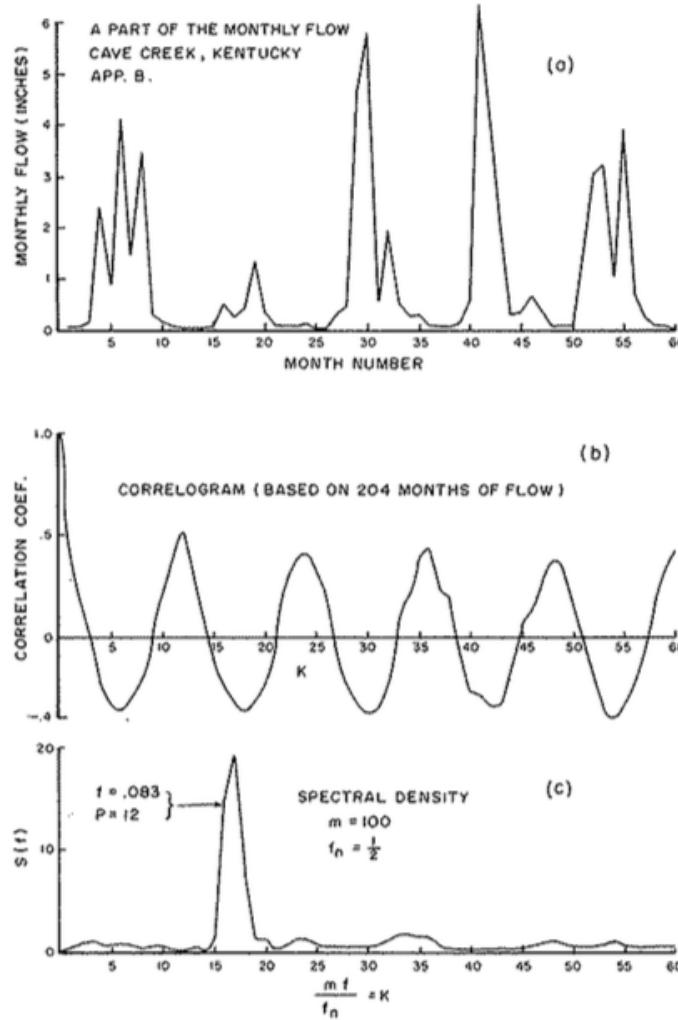


Fig. 14.7. (a) Cave Creek monthly runoff and its correlogram (b) and spectral density (c).

Terima kasih