



การวิเคราะห์ทางสถิติในงานอุทกวิทยา

ในบทนี้ เป็นการทบทวนหลักการเบื้องต้น ซึ่งได้รวบรวมจากหนังสือสถิติ ความน่าจะเป็น และสถิติในงานอุทกวิทยาเล่มต่าง ๆ [13-23] เกี่ยวกับการวิเคราะห์ทางสถิติในงานอุทกวิทยาโดยจะกล่าวถึง ทฤษฎีความน่าจะเป็นและสถิติเบื้องต้น, ลักษณะของขบวนการทางสถิติ, อนุกรมทางอุทกวิทยา, เทคนิคที่ใช้วิเคราะห์โครงสร้างของอนุกรม ซึ่งได้แก่ วิธีปรับข้อมูลให้เรียบ (Smoothing Method) การวิเคราะห์ความสัมพันธ์ในตัวเอง (Auto-correlation Analysis) และการวิเคราะห์ทางสเปกตรอล (Spectral Analysis) ซึ่ง 2 วิธีหลังนี้เป็นเทคนิคที่สำคัญในการวิเคราะห์ทางสถิติ

2.1 ทฤษฎีความน่าจะเป็นและสถิติเบื้องต้น

ปรากฏการณ์ทางอุทกวิทยาที่เกิดขึ้นมักมีความไม่แน่นอน การที่จะคาดหมายปรากฏการณ์เหล่านี้ว่าเป็นวิทยาศาสตร์นั้น สามารถใช้ทฤษฎีความน่าจะเป็นและสถิติมาวิเคราะห์พฤติกรรมต่าง ๆ ซึ่งเริ่มตั้งแต่ การเก็บรวบรวมข้อมูล การสรุปข้อมูลให้อยู่ในรูปความน่าจะเป็นและสถิติ การแปลความหมายของข้อมูลที่สรุปนั้น เพื่อให้เข้าใจลักษณะของปรากฏการณ์ที่เกิดขึ้น ตลอดจนทำนายพฤติกรรมที่มีโอกาสจะเกิดขึ้นในอนาคต สำหรับทฤษฎีที่จะกล่าวในหัวข้อนี้เป็นการทบทวนโดยสังเขป ประกอบด้วย กฎความน่าจะเป็น การกระจายความน่าจะเป็น ทารามิเตอร์ทางสถิติที่ใช้บอกคุณสมบัติการกระจายความน่าจะเป็น และชนิดของการกระจายความน่าจะเป็น

2.1.1 กฎความน่าจะเป็น

ความน่าจะเป็น (Probability) ขึ้นกับสถิติของการทดลอง/สำรวจ เช่น เหตุการณ์ E_1 เป็นเหตุการณ์ที่เกิดขึ้นหลังจากการทดลอง/สำรวจ N ครั้ง ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 ระหว่างการทดลอง/สำรวจ กับจำนวนครั้งที่ทดลอง/สำรวจทั้งหมด ดังสมการ 2-1

$$P(E_1) = \frac{n_1}{N} \quad (2-1)$$

โดย	$P(E_1)$	เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1
	n_1	เป็นจำนวนครั้งที่เกิดเหตุการณ์ E_1 ระหว่างการทดลอง/สำรวจ
	N	เป็นจำนวนครั้งที่ทดลอง/สำรวจทั้งหมด

คุณสมบัติของความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ สรุปเป็นกฎของความน่าจะเป็น ดังนี้

- 1) ความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ต้องไม่เป็นลบและไม่เกินหนึ่ง

$$0 < P(E_i) < 1 \quad (2-2)$$

- 2) ผลรวมของความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ทุกเหตุการณ์คือหนึ่ง

$$\sum_i P(E_i) = 1 \quad (2-3)$$

- 3) ในกรณีที่เหตุการณ์ 2 เหตุการณ์ E_1 และ E_2 เป็นเหตุการณ์ที่ไม่เป็นอิสระต่อกัน (Dependent Events)

ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2 เท่ากับผลรวมของความน่าจะเป็นที่จะเกิดแต่ละเหตุการณ์ ลบความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad (2-4)$$

โดย $P(E_1 \cup E_2)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 หรือ E_2

$P(E_1 \cap E_2)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2

- 4) ในกรณีที่เหตุการณ์ 2 เหตุการณ์ E_1 และ E_2 เป็นเหตุการณ์ที่เป็นอิสระต่อกัน (Independent Events)

ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 หรือ E_2 เท่ากับผลรวมของความน่าจะเป็นที่จะเกิดแต่ละเหตุการณ์

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \quad (2-5)$$

- 5) ในกรณีที่เหตุการณ์ E_1 เกิดขึ้นก่อนเหตุการณ์ E_2

ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 เมื่อกำหนดค่าให้เกิดเหตุการณ์ E_1 (Conditional probability) เท่ากับ อัตราส่วนของความน่าจะเป็นที่จะเกิด

เหตุการณ์ E_1 และ E_2 กับความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \quad (2-6)$$

โดย $P(E_2|E_1)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 หลังจากเกิดเหตุการณ์ E_1

ดังนั้น $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2|E_1) \quad (2-7)$

ถ้าเหตุการณ์ E_1 และ E_2 เป็นอิสระต่อกัน (Independent Events) ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 เมื่อกำหนดให้เกิดเหตุการณ์ E_1 เท่ากับ ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 ดังนั้นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2 เท่ากับผลคูณของความน่าจะเป็นที่จะเกิดเหตุการณ์ทั้ง 2 เหตุการณ์

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) \quad (2-8)$$

สำหรับปรากฏการณ์ทางอุทกวิทยานั้น มักแสดงความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ด้วยค่ารอบปี (Return Period) ซึ่งเป็นส่วนกลับของความน่าจะเป็นที่จะเกิดเหตุการณ์นั้น ๆ ดังสมการ 2-9

$$T = \frac{1}{P(E_1)} \quad (2-9)$$

โดย T เป็น ค่ารอบปี

ความหมายของค่ารอบปีนั้น มักเข้าใจสับสน เช่น ค่ารอบปีเท่ากับ 10 ไม่ได้หมายถึง ในระยะเวลา 10 ปีจะเกิดเหตุการณ์นี้ครั้งหนึ่ง แต่มีความหมายคือ ถ้าพิจารณาระยะเวลาแล้วจะได้ว่าเหตุการณ์นั้นเกิดขึ้นโดยเฉลี่ย 1 ครั้งในระยะเวลา 10 ปี หรือกล่าวอีกนัยหนึ่งว่าเหตุการณ์นั้นมีความน่าจะเป็นที่จะเกิดในหนึ่งปีเท่ากับ $1/10$ โดยทั่วไปสามารถสรุปความสัมพันธ์ของความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ กับค่ารอบปีได้ดังนี้

1) ความน่าจะเป็นที่จะเกิดเหตุการณ์ E ในแต่ละปี

$$P(E) = \frac{1}{T} \quad (2-10)$$

2) ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ E ในแต่ละปี

$$P(\bar{E}) = 1 - P(E) = 1 - \frac{1}{T} \quad (2-11)$$

3) ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ E ในเวลา n ปีติดกัน

$$[P(\bar{E})]^n = \left(1 - \frac{1}{T}\right)^n \quad (2-12)$$

4) ความน่าจะเป็นที่จะเกิดเหตุการณ์ E อย่างน้อย 1 ครั้ง ในเวลา n ปี ซึ่งเรียกอีกอย่างหนึ่งว่าความเสี่ยง (Risk)

$$R_n(E) = 1 - [P(\bar{E})]^n = 1 - \left(1 - \frac{1}{T}\right)^n \quad (2-13)$$

ตาราง 2-1 แสดงค่ารอบปีสัมพันธ์กับความเสี่ยงและระยะเวลาที่พิจารณา ตัวอย่างเช่น ต้องการหาค่ารอบปีที่ควรใช้ในการออกแบบระบบระบายน้ำ โดยยอมให้ความเสี่ยง 20% ที่จะเกิดน้ำท่วม 5 ปีต่อมา จากตารางนี้จะได้ว่าค่ารอบปีที่ควรใช้ในการออกแบบคือ 22.9 ปี เป็นต้น

2.1.2 การกระจายความน่าจะเป็น

การกระจายความน่าจะเป็น (Probability distribution) เป็นความสัมพันธ์ระหว่างความน่าจะเป็นของตัวแปร กับค่าของตัวแปรนั้น ๆ ใช้สัญลักษณ์ $P(X=x)$ โดย X แทนตัวแปรสุ่ม และ x แทนค่าของตัวแปร $P(X=x)$ นี้ อาจเขียนแทนด้วย $P(x)$ เช่น ตัวแปรที่มีค่าที่เป็นไปได้คือ 0 และ 1 โดยมีความน่าจะเป็นที่จะมีค่าเป็น 0 เท่ากับ 0.3 และความน่าจะเป็นที่จะมีค่าเป็น 1 เท่ากับ 0.7 สามารถเขียนแทนด้วย $P(X=0) = 0.3$ และ $P(X=1) = 0.7$ หรือแทนด้วย $P(0) = 0.3$ และ $P(1) = 0.7$ เป็นต้น

ในการนำเสนอการกระจายความน่าจะเป็นนี้มักนิยมใช้กราฟ ทั้งในรูปของความน่าจะเป็น (Probability density) $P(x)$ หรือ $f(x)$ และความน่าจะเป็นสะสม (Cumulative probability) $P(X \leq x)$ หรือ $F(x)$ ซึ่งมีความสัมพันธ์ดังนี้

$$P(X \leq x') = F(x') = \sum_i p(x_i) \quad \text{โดย } x_i \leq x' \quad (2-14)$$

$$\text{หรือ } P(X \leq x') = F(x') = \int_{-\infty}^{x'} p(x) dx = \int_{-\infty}^{x'} f(x) dx \quad (2-15)$$

ตาราง 2-1 ค่ารอบปีสัมพันธ์กับค่าความเสี่ยงและระยะเวลาที่พิจารณา (อ้างอิง 13)

Risk (%)	Expected Design Life (years)							
	2	5	10	15	20	25	50	100
75	2.0	4.1	7.7	11.3	14.9	18.5	36.6	72.6
50	3.4	7.7	14.9	22.1	29.4	36.6	72.6	144.8
40	4.4	10.3	20.1	29.9	39.7	49.4	98.4	196.3
30	6.1	14.5	28.5	42.6	56.6	70.6	140.7	280.9
25	7.5	17.9	35.3	52.6	70.0	87.4	174.3	348.1
20	9.5	22.9	45.3	67.7	90.1	112.5	224.6	448.6
15	12.8	31.3	62.0	92.8	123.6	154.3	308.2	615.8
10	19.5	48.0	95.4	142.9	190.3	237.8	475.1	949.6
5	39.5	98.0	195.5	292.9	390.4	487.9	975.3	1950.1
2	99.5	248.0	495.5	743.0	990.5	1238.0	2475.4	4950.3
1	199.5	498.0	995.5	1493.0	1990.5	2488.0	4975.5	9950.4

015766

ในรูป 2-1 ก แสดงการกระจายความน่าจะเป็นของตัวแปรที่ไม่ต่อเนื่องและรูป 2-1 ข แสดงการกระจายความน่าจะเป็นของตัวแปรที่ต่อเนื่อง

2.1.3 พารามิเตอร์ทางสถิติที่ใช้บอกคุณสมบัติการกระจายความน่าจะเป็น

คุณสมบัติการกระจายความน่าจะเป็นสามารถอธิบายได้ด้วยพารามิเตอร์ทางสถิติ ซึ่งในการหาค่าพารามิเตอร์เหล่านี้ มักใช้ค่าพารามิเตอร์ของตัวแปรสุ่มตัวอย่าง เพื่อประมาณค่าพารามิเตอร์ของตัวแปรทั้งหมด พารามิเตอร์ที่สำคัญมีดังนี้

- 1) พารามิเตอร์ที่ใช้วัดศูนย์กลางของการกระจาย โดยทั่วไปใช้ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) เขียนแทนด้วย μ สำหรับตัวแปรทั้งหมด และ \bar{X} สำหรับตัวแปรสุ่มตัวอย่าง สามารถหาค่าได้จากสมการ 2-16 และ 2-17

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i \quad (2-16)$$

โดย N เป็นจำนวนตัวแปรทั้งหมด

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2-17)$$

โดย n เป็นจำนวนตัวแปรสุ่มตัวอย่าง

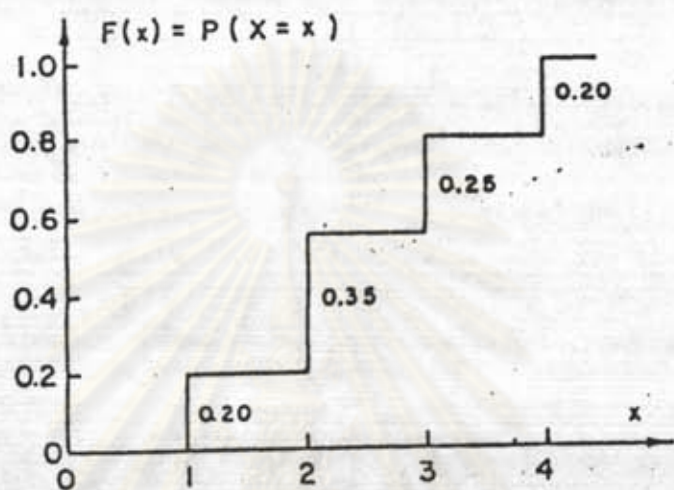
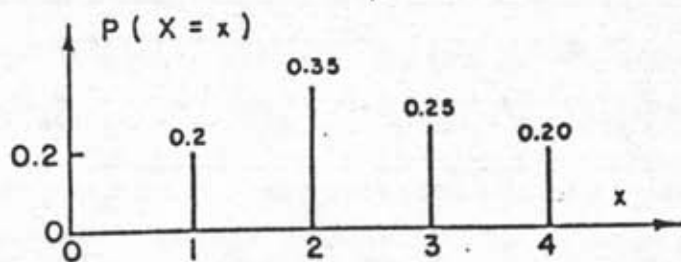
ส่วนพารามิเตอร์อื่นที่ใช้วัดศูนย์กลางนี้ได้แก่ ค่ามัธยฐาน (Median) เป็นค่าของตัวแปรที่แบ่งการกระจายออกเป็น 2 ส่วน แต่ละส่วนมีความน่าจะเป็นเท่ากับ 0.5 ค่าฐานนิยม (Mode) เป็นค่าของตัวแปรที่มีความน่าจะเป็นสูงสุด

- 2) พารามิเตอร์ที่ใช้วัดความผันแปรของการกระจาย โดยทั่วไปใช้ค่าความแปรปรวน (Variance) เขียนแทนด้วย σ^2 สำหรับตัวแปรทั้งหมด และ S^2 สำหรับตัวแปรสุ่มตัวอย่างสามารถหาค่าได้จากสมการ 2-18 และ 2-19

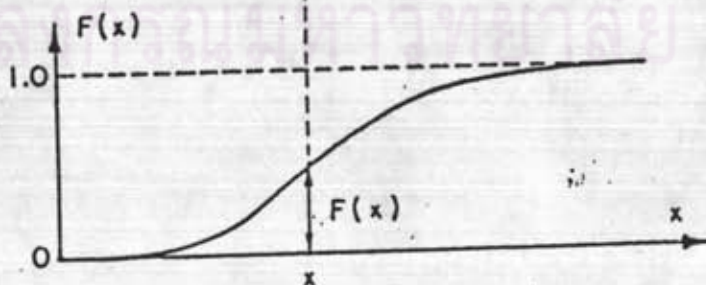
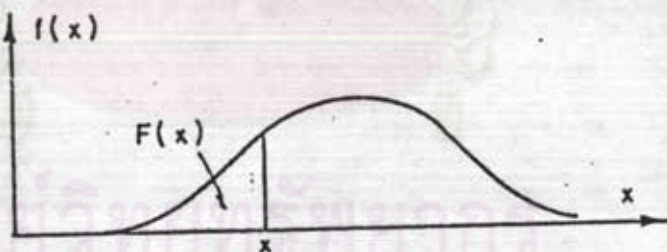
$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 \quad (2-18)$$

โดย N เป็นจำนวนตัวแปรทั้งหมด

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2-19)$$



ก) การกระจายความน่าจะเป็นของตัวแปรไม่ต่อเนื่อง



ข) การกระจายความน่าจะเป็นตัวแปรต่อเนื่อง

รูป 2-1 การกระจายความน่าจะเป็นของตัวแปร (อ้างถึง 17)

n เป็นจำนวนตัวแปรสุ่มตัวอย่าง

ส่วนทวารามิเตอร์อื่นที่ใช้วัดความผันแปรได้แก่ ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็นค่ารากที่ 2 ของความแปรปรวนเขียนแทนด้วย σ สำหรับตัวแปรทั้งหมด และ S สำหรับตัวแปรสุ่มตัวอย่าง ค่าพิสัย (Range) เป็นค่าความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของตัวแปร ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variation) เป็นอัตราส่วนระหว่าง σ กับ μ หรือ S กับ \bar{X} เขียนแทนด้วย C_v ใช้สำหรับเปรียบเทียบความผันแปรของการกระจายหลาย ๆ ชุด

- 3) ทวารามิเตอร์ที่ใช้วัดความสมมาตรของการกระจาย โดยทั่วไปใช้ค่าความเบ้ (Skewness) เขียนแทนด้วย α สำหรับตัวแปรทั้งหมด และ a สำหรับตัวแปรสุ่มตัวอย่าง สามารถหาค่าได้จากสมการ 2-20 และ 2-21

$$\alpha = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^3 \quad (2-20)$$

โดย N เป็นจำนวนตัวแปรทั้งหมด

$$a = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (X_i - \bar{X})^3 \quad (2-21)$$

โดย n เป็นจำนวนตัวแปรสุ่มตัวอย่าง

ส่วนทวารามิเตอร์อื่นที่ใช้วัดความสมมาตรได้แก่ สัมประสิทธิ์ความเบ้ (Coefficient of Skewness) เป็นอัตราส่วนระหว่าง α กับ σ^3 หรือ a กับ S^3 เขียนแทนด้วย C_s ถ้า $C_s = 0$ แสดงว่า การกระจายความน่าจะเป็นสมมาตรรอบค่าเฉลี่ย ถ้าค่า C_s เป็นบวก แสดงว่าการกระจายนี้จะเบ้ซ้าย และถ้า C_s เป็นลบ แสดงว่าการกระจายนี้จะเบ้ขวา

ตาราง 2-2 แสดงค่าบ้านรายปีของเมือง Anniston, Los Angeles และ Richmond จะเห็นได้ว่า 1) บ้านรายปีที่เมือง Anniston มีค่าเฉลี่ยสูง, ความผันแปรของข้อมูลปานกลาง และการกระจายของข้อมูลเบ้ขวา 2) บ้านรายปีที่เมือง Los Angeles มีค่าเฉลี่ยต่ำ, ความผันแปรสูง และเบ้ซ้ายเล็กน้อย และ 3) บ้านรายปีที่เมือง Richmond มีค่าเฉลี่ยปานกลาง, ความผันแปรต่ำ และเบ้ขวาเล็กน้อย

ตาราง 2-2 ค่าสถิติน้ำฝนรายปีของเมือง Anniston, Los Angeles และ Richmond

Year	Annual rainfall (in.)		
	Anniston, Ala.	Los Angeles, Calif.	Richmond, Va.
1928	48	9	43
1927	49	19	44
1926	55	19	38
1925	98	9	31
1924	43	8	47
1923	53	6	49
1922	56	15	52
1921	47	20	31
1920	69	11	51
1919	57	9	40
1918	61	18	41
1917	64	8	43
1916	99	23	37
1915	54	17	36
1914	40	23	34
1913	47	17	38
1912	58	10	36
1911	44	18	37
1910	44	5	43
1909	64	24	34
1908	44	19	53
1907	51	15	49
1906	71	21	47

Solution

Parameter	Anniston	Los Angeles	Richmond
Mean, \bar{x}	57.2 in.	14.9 in.	41.5 in.
Standard deviation, s	15.5 in.	5.9 in.	6.7 in.
Coefficient of variance, $C_v = s/\bar{x}$	0.27	0.40	0.16
Coefficient of skewness, $C_s = s/\bar{x}^3$	1.69	-0.15	0.16

2.1.4 ชนิดของการกระจายความน่าจะเป็นมาตรฐาน

ในการอธิบายปรากฏการณ์ทางอุทกวิทยา สามารถใช้การกระจายความน่าจะเป็นมาตรฐาน (Standard probability distribution) ซึ่งมีอยู่หลายทฤษฎี ดังตาราง 2-3 แสดงทฤษฎีการกระจายความน่าจะเป็นมาตรฐานชนิดต่าง ๆ ส่วนหนึ่งที่นิยมใช้กันมาก โดยแสดงฟังก์ชันการกระจายความน่าจะเป็น ค่าเฉลี่ย และความแปรปรวนของฟังก์ชัน การเลือกใช้ทฤษฎีเหล่านี้ให้เหมาะสมนั้น จะต้องมีการทดสอบความเหมาะสมในเชิงสถิติก่อน ซึ่งรายละเอียดของทฤษฎีและการทดสอบนั้นสามารถศึกษาได้จากหนังสือสถิติทั่วไป

2.2 ลักษณะของขบวนการทางสโตแคสติก

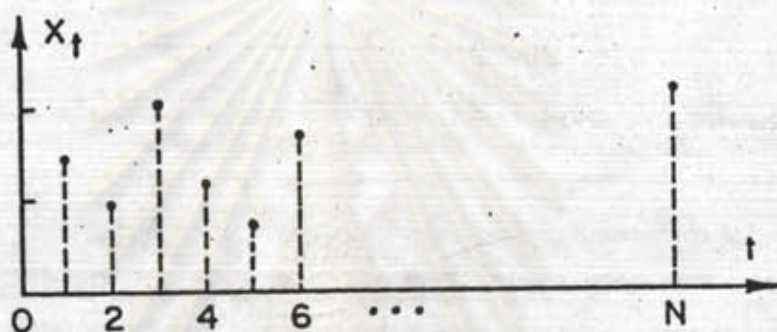
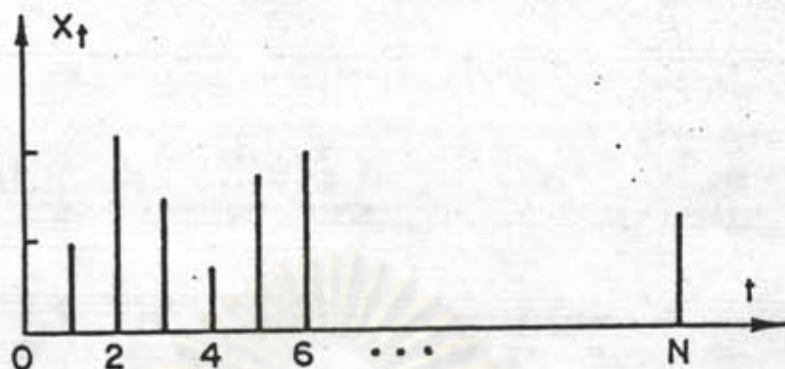
ในหัวข้อที่แล้วได้กล่าวถึงทฤษฎีความน่าจะเป็นและสถิติ ซึ่งใช้ในการหาโอกาสเกิดของเหตุการณ์ที่ไม่วุ่นวาย โดยตัวแปร X ของเหตุการณ์นี้เรียกว่า ตัวแปรที่ไม่วุ่นวาย (Random Variable) อนุกรม X_1, X_2, X_3, \dots ที่เกิดจากตัวแปร X นี้ เรียกว่า อนุกรมทางความน่าจะเป็น (Probabilistic Series) กลไกที่ทำให้เกิดการเปลี่ยนแปลงค่าตัวแปร X นี้ เรียกว่า ขบวนการทางความน่าจะเป็น (Probabilistic Process)

ในทางตรงข้าม ถ้าตัวแปร X นั้นสามารถทำนายผล X อย่างแน่นอนได้ ตัวแปรนี้เรียกว่าตัวแปรดีเทอร์มิเนติก (Deterministic Variable) อนุกรม X_1, X_2, X_3, \dots ที่เกิดจากตัวแปร X นี้เรียกว่า อนุกรมทางดีเทอร์มิเนติก (Deterministic Series) กลไกที่ทำให้เกิดการเปลี่ยนแปลงค่าตัวแปร X นี้ เรียกว่าขบวนการทางดีเทอร์มิเนติก (Deterministic Process)

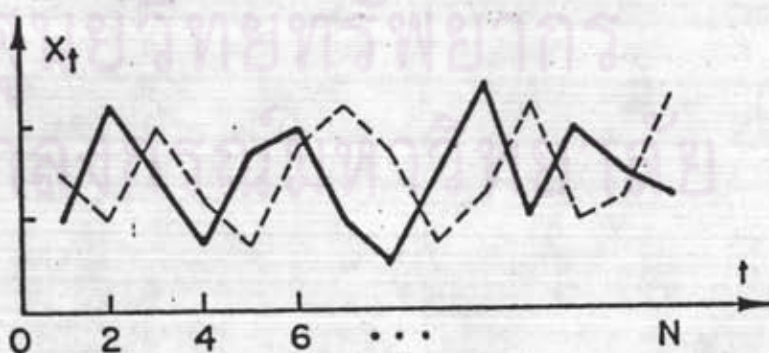
สำหรับตัวแปร X ที่ไม่สามารถทำนายอย่างแน่นอน และจะมีส่วนที่สัมพันธ์กับเหตุการณ์ที่เกิดขึ้นแล้ว ตัวแปร X นี้เรียกว่า ตัวแปรสโตแคสติก (Stochastic Variable) อนุกรม X_1, X_2, X_3, \dots ที่เกิดจากตัวแปร X นี้ เรียกว่า อนุกรมทางสโตแคสติก (Stochastic Series) กลไกที่ทำให้เกิดการเปลี่ยนแปลงค่าตัวแปร X นี้ เรียกว่าขบวนการทางสโตแคสติก (Stochastic Process) ในรูป 2-2 แสดงตัวอย่างขบวนการทางสโตแคสติก 2 ขบวนการเขียนในรูปกราฟแท่ง ซึ่งสามารถเขียนแทนให้อยู่ในรูปกราฟเส้นแบบต่อเนื่อง ทั้ง ๆ ที่ข้อมูลไม่ต่อเนื่อง เพื่อให้ง่ายต่อการพิจารณาลำดับเหตุการณ์มากขึ้น ดังในรูป 2-2 ข

ตาราง 2-3 ทฤษฎีการกระจายความน่าจะเป็นมาตรฐานชนิดต่าง ๆ (อ้างอิง 13)

Distribution of Random Variable X	Probability Distribution Function	Range	Mean \bar{x} or μ	Variance s^2 or σ^2
Binomial	$P(x) = \frac{n!}{x!(n-x)!} p^x(1-p)^{n-x}$	$0 \leq x < n$	np	$np(1-p)$
Poisson	$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$0 \leq x \leq \dots$	λ	λ
Uniform	$f(x) = \frac{1}{b-a}$	$a \leq x \leq b$	$(b+a)/2$	$\frac{(b-a)^2}{12}$
Exponential	$f(x) = \frac{1}{a} e^{-x/a}$	$0 \leq x \leq \infty$	a	a^2
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$	$-\infty \leq x \leq \infty$	μ	σ^2
Lognormal ($y = \ln x$)	$f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp[-(y - \mu_y)^2/2\sigma_y^2]$	$-\infty \leq y \leq \infty$ ($0 \leq x \leq \infty$)	μ_y	σ_y^2
Gamma	$f(x) = \frac{x^\alpha e^{-x/\beta}}{\beta^{\alpha+1} \Gamma(\alpha+1)}$	$0 \leq x \leq \infty$	$\beta(\alpha+1)$	$\beta^2(\alpha+1)$
Extreme value	$f(x) = \alpha \exp\{-\alpha(x-u) - e^{-\alpha(x-u)}\}$	$-\infty \leq x \leq \infty$	$u + \frac{0.5772}{\alpha}$	$\frac{\pi^2}{6\alpha^2}$



ก) ตัวอย่างขบวนการสโตแคสติก 2 ขบวนการในรูปแบบกราฟแท่ง



ข) ตัวอย่างขบวนการสโตแคสติก 2 ขบวนการในรูปแบบกราฟเส้น

คุณสมบัติของขบวนการทางสโตแคสติกนี้ ขึ้นกับความสัมพันธ์ของค่าตัวแปร $X_1, X_2, X_3 \dots$ สามารถอธิบายด้วยพารามิเตอร์ทางสถิติ ซึ่งได้แก่ ค่าเฉลี่ย (Mean), ความแปรปรวน (Variance), ค่าฟังก์ชันความสัมพันธ์ในตัวเอง (Autocorrelation Function) และค่าความหนาแน่นของสเปกตรัม (Spectrum Density) ซึ่ง 2 ค่าหลังนี้จะกล่าวรายละเอียดการวิเคราะห์ในหัวข้อ 2.4

ขบวนการทางสโตแคสติก สามารถเรียกชื่อตามความสมดุลง (Stationary) ได้หลายแบบคือ ขบวนการทางสโตแคสติกที่มีค่าเฉลี่ยคงที่ไม่ขึ้นกับเวลานั้น เรียกว่าขบวนการทางสโตแคสติกที่สมดุลงสำหรับค่าเฉลี่ย (Stationary in Mean) ในทำนองเดียวกัน ขบวนการทางสโตแคสติกที่มีค่าความแปรปรวนคงที่ไม่ขึ้นกับเวลานั้น เรียกว่า ขบวนการทางสโตแคสติกที่สมดุลงสำหรับค่าความแปรปรวน (Stationary in Variance) สำหรับขบวนการทางสโตแคสติกที่สมดุลงทั้ง 2 อย่างเรียกว่า สมดุลงอย่างอ่อน (Weak Stationary) ถ้ามีคุณสมบัติทางสถิติอย่างอื่นรวมทั้ง 2 อย่างนี้สมดุลงไม่ขึ้นกับเวลา เรียกว่า สมดุลงอย่างมาก (Strong Stationary) แต่ถ้าคุณสมบัติทางสถิติทุกตัวขึ้นกับเวลาจะเรียกว่าไม่สมดุลง (Nonstationary)

2.3 อนุกรมทางอุทกวิทยา

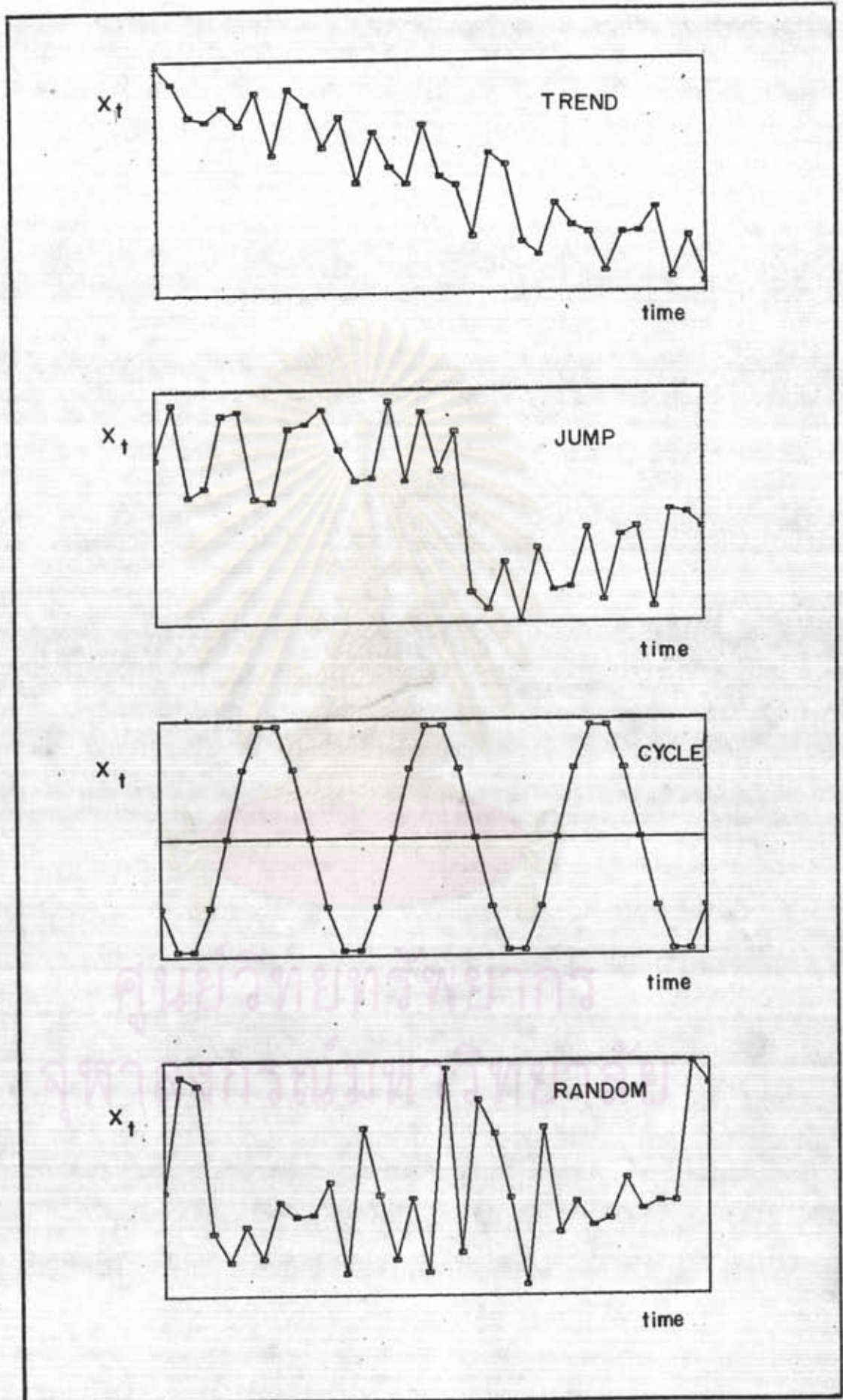
อนุกรมทางอุทกวิทยา (Hydrologic Series) เป็นอนุกรมที่เกิดจากตัวแปรทางอุทกวิทยา อนุกรมนี้สามารถแบ่งตามชนิดของตัวแปรได้เป็น อนุกรมที่ตัวแปรขึ้นกับเวลา ซึ่งเรียกว่า อนุกรมเวลา (Time Series) เช่น ปริมาณน้ำฝนรายปีที่สถานีวัดน้ำฝนแห่งหนึ่ง อนุกรมที่ตัวแปรขึ้นกับระยะทาง ซึ่งเรียกว่าอนุกรมระยะทาง (Line Series) เช่น ความลึกของแม่น้ำตามความยาวของแม่น้ำ และอนุกรมที่ขึ้นกับจำนวนครั้งที่เกิดปรากฏการณ์ที่สนใจในระยะเวลาหรือพื้นที่ที่พิจารณา ซึ่งเรียกว่าอนุกรมจำนวนครั้ง (Counting Series) เช่น จำนวนวันฝนตกในแต่ละปีของสถานีวัดน้ำฝนแห่งหนึ่งที่กล่าวมาทั้ง 3 ชนิดนี้เป็นอนุกรมที่พิจารณาตัวแปรใน 1 มิติ ยังมีอนุกรมทางอุทกวิทยาที่อาจพิจารณาตัวแปรในหลายมิติอีก แต่อนุกรมทางอุทกวิทยาที่เกิดขึ้นส่วนใหญ่จะเป็นอนุกรมเวลา

ในการพิจารณาลักษณะของอนุกรมทางอุทกวิทยา ซึ่งเป็นอนุกรมเวลานี้ สามารถแยกส่วนประกอบของอนุกรมออก เป็นดังนี้

อนุกรมเวลาหนึ่ง = ส่วนของแนวโน้ม + ส่วนฤดูกาล + ส่วนวงจร + ส่วนสโตแคสติก
+ ส่วนที่ไม่แน่นอน

ในรูป 2-3 แสดงตัวอย่างของอนุกรมเวลาที่มีส่วนประกอบต่าง ๆ กัน รายละเอียดของ
ส่วนประกอบต่าง ๆ มีดังต่อไปนี้

- 1) ส่วนที่เปลี่ยนแปลงอย่างมีแนวโน้มหรืออย่างทันทีทันใด (Trend of Jump Component) เป็นส่วนของอนุกรมทางอุทกวิทยา ที่เกิดการเปลี่ยนแปลงค่าของตัวแปร ความล่าช้า เหตุการณ์ โดยไม่คำนึงถึงความผันแปรระยะสั้น มักเกิดจากสาเหตุของ องค์ประกอบอื่นเปลี่ยนแปลง เช่น ปริมาณน้ำใช้ของคนในกรุงเทพมหานครมากขึ้น เรื่อย ๆ เนื่องจากคนในกรุงเทพมหานครมีมากขึ้น การเปลี่ยนแปลงนี้มี 2 ลักษณะ คือการเปลี่ยนแปลงอย่างมีแนวโน้ม (Trend) และการเปลี่ยนแปลงอย่างทันทีทันใด (Jump) สำหรับการวิเคราะห์การเปลี่ยนแปลงทั้ง 2 ลักษณะทำได้โดยเขียนกราฟ ของข้อมูลในอดีตตามลำดับเหตุการณ์ แล้วสังเกตการเปลี่ยนแปลงของข้อมูล ถ้า ข้อมูลมีความผันแปรมาก ก็ต้องทำการปรับข้อมูลให้เรียบมากขึ้น ซึ่งกล่าว รายละเอียดในหัวข้อ 2.4.1 นอกจากการสังเกตแล้ว ยังสามารถใช้สมการทาง คณิตศาสตร์แบบต่าง ๆ เช่น สมการเส้นตรง สมการเลขยกกำลัง เพื่อศึกษาการ เปลี่ยนแปลงอย่างมีแนวโน้มของอนุกรมได้
- 2) ส่วนที่เป็นฤดูกาล (Seasonal Component) เป็นส่วนของอนุกรมทางอุทกวิทยา ที่เกิดขึ้นในฤดูกาลประจำปี มักเป็นปรากฏการณ์ที่มีผลมาจากการโคจรของ โลกรอบดวงอาทิตย์ในหนึ่งปี คลอจนสาเหตุอื่น ๆ ที่เกิดขึ้นในลักษณะฤดูกาล เช่น ปริมาณน้ำฝนมากในช่วงฤดูฝน คือ เดือน พค. - ตค. และปริมาณน้ำฝนน้อย ในฤดูแล้ง คือ เดือน พย. - เมย. สำหรับการวิเคราะห์ผลของฤดูกาล อาจทำได้โดยการหาค่าเฉลี่ยแต่ละฤดูกาล เช่น มีข้อมูลน้ำฝนรายเดือน 30 ปี ก็เอาค่าเฉลี่ยของน้ำฝนเดือน มค. จนถึง ตค. ใน 30 ปีนั้น มาเป็นตัวแทนผลของ ฤดูกาลเพื่อขจัดออกจากอนุกรมที่พิจารณาเป็นต้น แต่ถ้าข้อมูลที่สนใจในช่วงเวลา เป็นปี เช่นน้ำฝนรายปี ผลของฤดูกาลก็ไม่จำเป็นต้องนำมาพิจารณา
- 3) ส่วนที่เป็นวงจร (Cyclicality Component) เป็นส่วนของอนุกรมทางอุทกวิทยา ที่มีความสำคัญมาก หมายถึงรอบของเหตุการณ์ที่เกิดขึ้น โดยมากวงจรนี้มีผลจาก การโคจรของดวงดาว เช่น โลก ดวงจันทร์ และดวงอาทิตย์ เป็นต้น ซึ่งการ



รูป 2-3 อุนกรรมของการเปลี่ยนแปลงในรูปแบบต่าง ๆ

โคจรเป็นวง ดังนั้นวงจرتึ้นฐานที่สำคัญทางอุทกวิทยามักประกอบด้วย 1 วัน, 7 วัน, 15 วัน, 30 วัน, 1 ปี และหลาย ๆ ปี แล้วแต่ขบวนการที่เกิดขึ้นมีองค์ประกอบใดที่มีอิทธิพลต่อขบวนการ สำหรับการวิเคราะห์ห้วงจรอาจทำได้โดยใช้การวิเคราะห์ทางสเปกตรอล (Spectral Analysis) ซึ่งกล่าวรายละเอียดในหัวข้อ 2.4.3

- 4) ส่วนที่เป็นสโตแคสติก (Stochastic Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่ตัวแปรไม่สามารถทำนายหรือคาดการณ์อย่างแน่นอนได้ แต่มีส่วนที่สัมพันธ์กับเหตุการณ์ที่เคยเกิดขึ้น มักจะพิจารณาหลังจากวิเคราะห์ส่วนเปลี่ยนแปลงที่สามารถทำนายได้อย่างแน่นอน เช่น แนวโน้มฤดูกาล และวงจร สำหรับการวิเคราะห์ส่วนที่เป็นสโตแคสติก ทำได้โดยใช้แบบจำลองทางสโตแคสติก เช่น AR, MA, ARIMA ซึ่งในการเลือกแบบจำลองนี้ ทำโดยเปรียบเทียบค่าความสัมพันธ์ในตัวเอง (Autocorrelation function) ของอนุกรมนี้กับแบบจำลอง จากนั้นก็ประมาณค่าพารามิเตอร์ที่จะใช้ในแบบจำลองตามความเหมาะสม ซึ่งการวิเคราะห์ความสัมพันธ์ในตัวเองนี้กล่าวรายละเอียดในหัวข้อ 2.4.2
- 5) ส่วนที่ไม่แน่นอน (Random Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่ไม่สามารถทำนายได้ด้วยองค์ประกอบข้างต้น สามารถวิเคราะห์โดยใช้ทฤษฎีความน่าจะเป็น เพื่อหาฟังก์ชันการกระจายความน่าจะเป็นที่เหมาะสมกับข้อมูลที่พิจารณา จากนั้นจึงหาความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ดังกล่าวในหัวข้อ 2.1

2.4 เทคนิคที่ใช้วิเคราะห์โครงสร้างของอนุกรม

การวิเคราะห์โครงสร้างของอนุกรม ใช้สำหรับศึกษาพฤติกรรม และส่วนประกอบของอนุกรม ซึ่งเป็นประโยชน์ต่อการอธิบายคุณสมบัติของอนุกรมนั้น ๆ และการเข้าใจลักษณะทางกายภาพที่มีผลต่อโครงสร้างของอนุกรมนั้น ๆ โดยเทคนิคที่ใช้วิเคราะห์ที่จะกล่าวถึงมีวิธีปรับข้อมูลให้เรียบ (Smoothing Method) การวิเคราะห์ความสัมพันธ์ในตัวเอง (Autocorrelation Analysis) และการวิเคราะห์ทางสเปกตรอล (Spectral Analysis)

2.4.1 วิธีปรับข้อมูลให้เรียบ (Smoothing Method)

อนุกรมทางอุทกวิทยามักมีความผันแปรมากในการวิเคราะห์ เพื่อหาแนวโน้มและพฤติกรรมของข้อมูลนั้น จึงจำเป็นต้องปรับข้อมูลให้เรียบมากขึ้น ซึ่งทำได้โดยแบ่งข้อมูลออกเป็นช่วง ๆ แล้วหาค่าเฉลี่ยทีละช่วงต่อเนื่องกันไป (Moving Average) ดังสมการ 2-22 เป็นการหาค่าเฉลี่ยช่วงละ N ข้อมูล หรือเขียนแทนด้วย $MA(N)$

$$C_j = \sum_{i=j-(N-1)/2}^{j+(N-1)/2} W_k X_i \quad (2-22)$$

โดย C_j = ค่าเฉลี่ยที่ j เป็นตัวแทนข้อมูลที่ j ซึ่ง

$$j = (N+1)/2, (N+3)/2, \dots, T-(N-1)/2$$

X_i = ค่าตัวแปรของอนุกรม ซึ่ง $i = 1, 2, \dots, T$

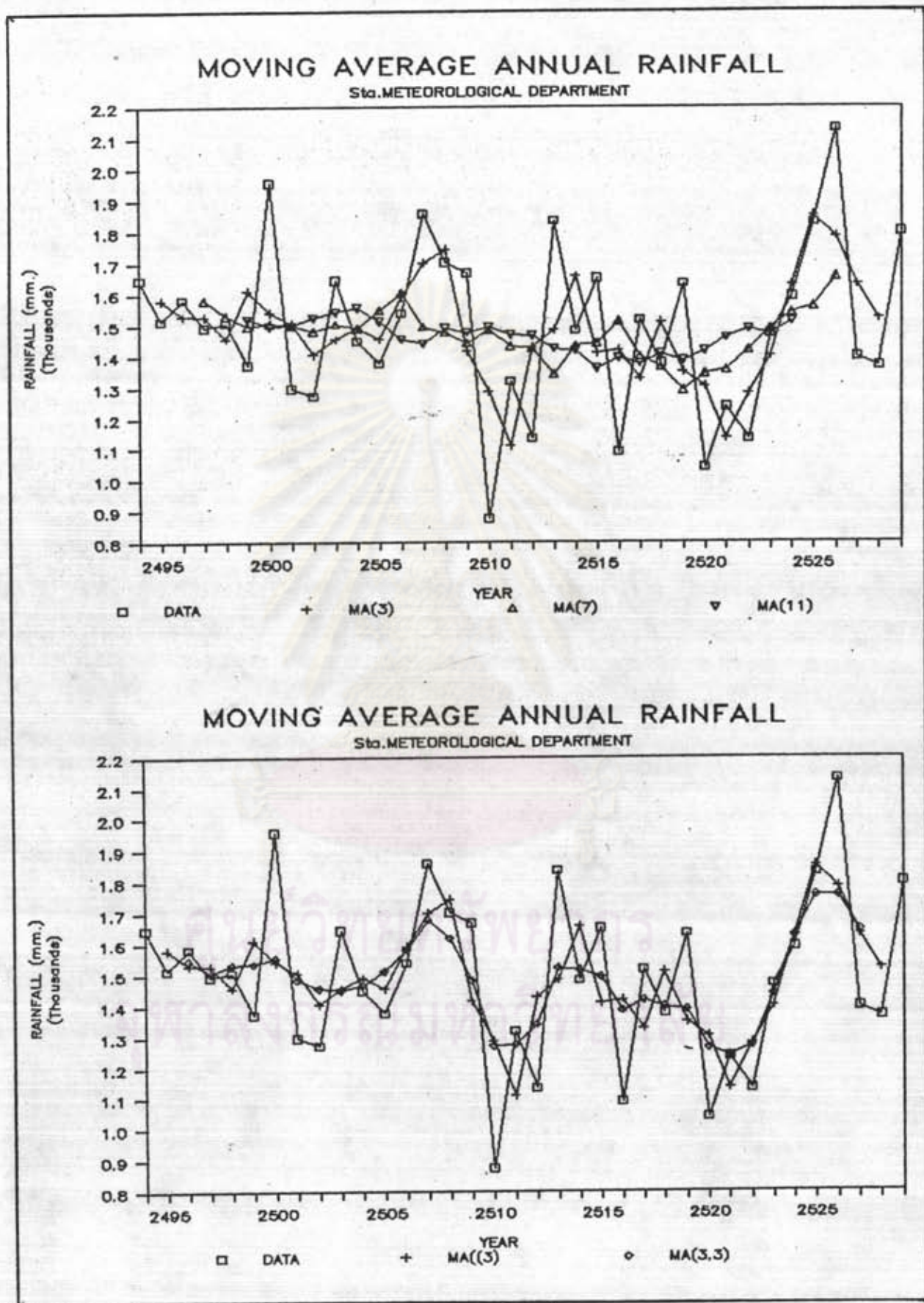
T = ความยาวของอนุกรมทั้งหมด

N = ช่วงของข้อมูลที่พิจารณา มักจะให้เป็นเลขคี่ เพื่อให้ j เป็นจำนวนเต็ม

W_k = น้ำหนักของข้อมูลมีผลรวมเท่ากับหนึ่ง

$$\text{คือ } \sum_{k=1}^N W_k = 1 \text{ ถ้า } W_k \text{ คงที่จะได้ } W_k = 1/N$$

การเลือกค่า W_k ให้เหมาะสมขึ้นกับจุดประสงค์และลักษณะของข้อมูลที่พิจารณา แต่ส่วนใหญ่ มักใช้ W_k คงที่เพื่อสะดวกในการคำนวณ และช่วงของข้อมูลที่พิจารณา N จะมากหรือน้อยขึ้นกับจุดประสงค์ที่ใช้ในการวิเคราะห์ กล่าวคือ ถ้า N มีค่ามากก็จะสามารถปรับกราฟที่พิจารณาให้เรียบมากขึ้น แต่ความสามารถในการเป็นตัวแทนข้อมูลจะลดลง แต่ถ้า N มีค่าน้อยก็จะปรับกราฟที่พิจารณาให้เรียบได้น้อยลง แต่ความสามารถในการเป็นตัวแทนข้อมูลจะเพิ่มขึ้น ในบางกรณีการหาค่าเฉลี่ยเพียงครั้งเดียว ยังไม่สามารถปรับข้อมูลให้เหมาะสม จึงอาจทำการหาค่าเฉลี่ยซ้ำอีก (Double Moving Average) เขียนแทนด้วย $MA(N,N)$ ซึ่งจะเห็นว่า W_k ที่ใช้ในการคำนวณไม่คงที่ เช่น $MA(3,3)$ มีค่า W_k เป็น $1/9, 2/9, 3/9, 2/9, 1/9$ ตามลำดับ ในรูป 2-4 แสดงตัวอย่างค่าเฉลี่ยเคลื่อนที่ที่ N ต่าง ๆ กับของปริมาณน้ำฝนรายปีที่กรมอุตุนิยมวิทยา



รูป 2-4 ตัวอย่างค่าเฉลี่ยเคลื่อนที่ของปริมาณน้ำฝนรายปีที่กรมอุตุนิยมวิทยา

2.4.2 การวิเคราะห์ความสัมพันธ์ในตัวเอง (Autocorretation Analysis)

การวิเคราะห์ความสัมพันธ์ในตัวเอง เป็นเทคนิคที่ใช้สำหรับหาความสัมพันธ์และรูปแบบของอนุกรม เพื่อเป็นแนวทางในการเลือกแบบจำลองทางสถิติ โดยค่าหึ่งขึ้นความสัมพันธ์ในตัวเอง เป็นค่าความสัมพันธ์ระหว่างสัมประสิทธิ์ของความสัมพันธ์ในตัวเอง (Autocorrelation Coefficient) กับช่วงเวลาถัดไป k (lag time) ซึ่งค่าสัมประสิทธิ์ความสัมพันธ์ในตัวเองใช้วัดความสัมพันธ์เชิงเส้นของตัวแปร X_t กับ X_{t+k} เขียนแทนด้วย ρ_k สำหรับตัวแปรทั้งหมด และ r_k สำหรับตัวแปรตัวอย่าง จาก [18] หาค่าได้โดย

$$\rho_k = \frac{\sum_{t=1}^{N-k} (X_t - \mu)(X_{t+k} - \mu)}{\sum_{t=1}^N (X_t - \mu)^2} \quad (2-23)$$

$$r_k = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} \quad (2-24)$$

ในกรณีที่อนุกรมไม่สมคลุ้ย ค่า r_k จากสมการ 2-24 ใช้ประมาณ ρ_k ได้ไม่แม่นยำ โดยเฉพาะอนุกรมทางอุทกวิทยาที่เป็นอนุกรมเปิด ไม่มีจุดเริ่มต้นและจุดสิ้นสุด (Open Series) มักใช้สมการ 2-25 ในการประมาณ ρ_k แทน

$$r_k = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X}_t)(X_{t+k} - \bar{X}_{t+k})}{\left(\sum_{t=1}^{N-k} (X_t - \bar{X}_t)^2 \sum_{t=1}^{N-k} (X_{t+k} - \bar{X}_{t+k})^2 \right)^{1/2}} \quad (2-25)$$

โดย \bar{X}_t เป็นค่าเฉลี่ย $N-k$ ค่าแรกคือ X_1, X_2, \dots, X_{N-k}
 \bar{X}_{t+k} เป็นค่าเฉลี่ย $N-k$ ค่าหลังคือ $X_{k+1}, X_{k+2}, \dots, X_N$

ทั้งสมการ 2-24 และ 2-25 ถ้า $k = 0$ จะได้ $r_k = 1$ และ r_k อยู่ระหว่าง -1 ถึง 1 สำหรับการเลือก k นั้น ไม่ควรให้ k มีค่ามากนัก เพราะถ้า k มากจะทำให้ $N-k$ น้อยลง ดังนั้นความแม่นยำของ r_k ก็น้อยลง

เมื่อนำค่าสัมประสิทธิ์ของความสัมพันธ์ในตัวเองไปเขียนกราฟสัมพันธ์กับ k จะได้กราฟซึ่งเรียกว่ากราฟความสัมพันธ์ (Correlogram) หรือฟังก์ชันความสัมพันธ์ในตัวเอง (Autocorrelation Function) ดังในรูป 2-5 แสดงตัวอย่างกราฟความสัมพันธ์ โดยทั่วไปนิยม

แสดงกราฟความสัมพันธ์ด้วยกราฟเส้นมากกว่ากราฟแท่ง เพราะสามารถแสดงลำดับเหตุการณ์ได้ชัดเจนกว่า

จากกราฟความสัมพันธ์ ถ้าตัวแปรไม่เป็นอิสระต่อกัน (Dependent Variables) ค่า r_k จะมีค่าเข้าใกล้ 1 ถ้าตัวแปรเป็นอิสระต่อกัน (Independent Variables) ค่า r_k จะมีค่าเป็น 0 หรือมีค่าใกล้ 0 โดยอยู่ในขอบเขตที่ยอมรับได้ (Confidence Interval) ดังสมการ 2-26 ก และ ข

$$r_k (95\%) = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k} \quad (2-26 \text{ ก})$$

โดย $r_k (95\%)$ เป็นค่าขอบเขตของ r_k ที่ 95% ของข้อมูล N ตัว

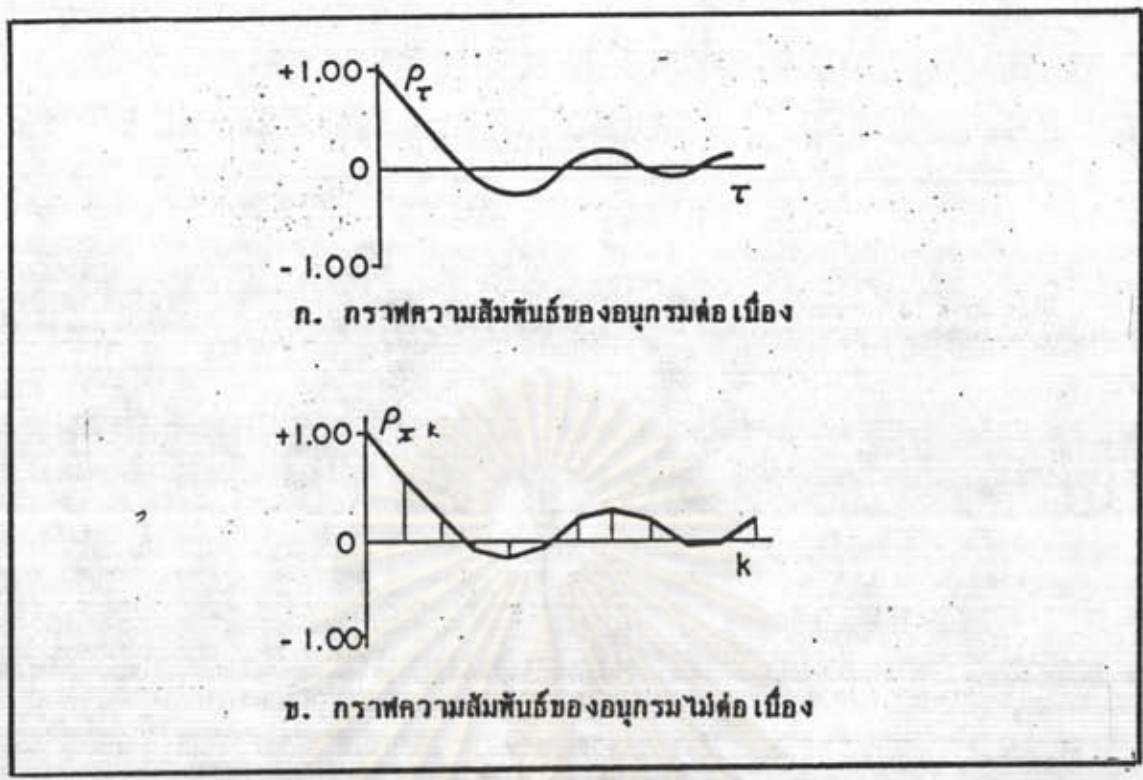
$$r_k (99\%) = \frac{-1 \pm 2.326\sqrt{N-k-1}}{N-k} \quad (2-26 \text{ ข})$$

โดย $r_k (99\%)$ เป็นค่าขอบเขตของ r_k ที่ 99% ของข้อมูล N ตัว

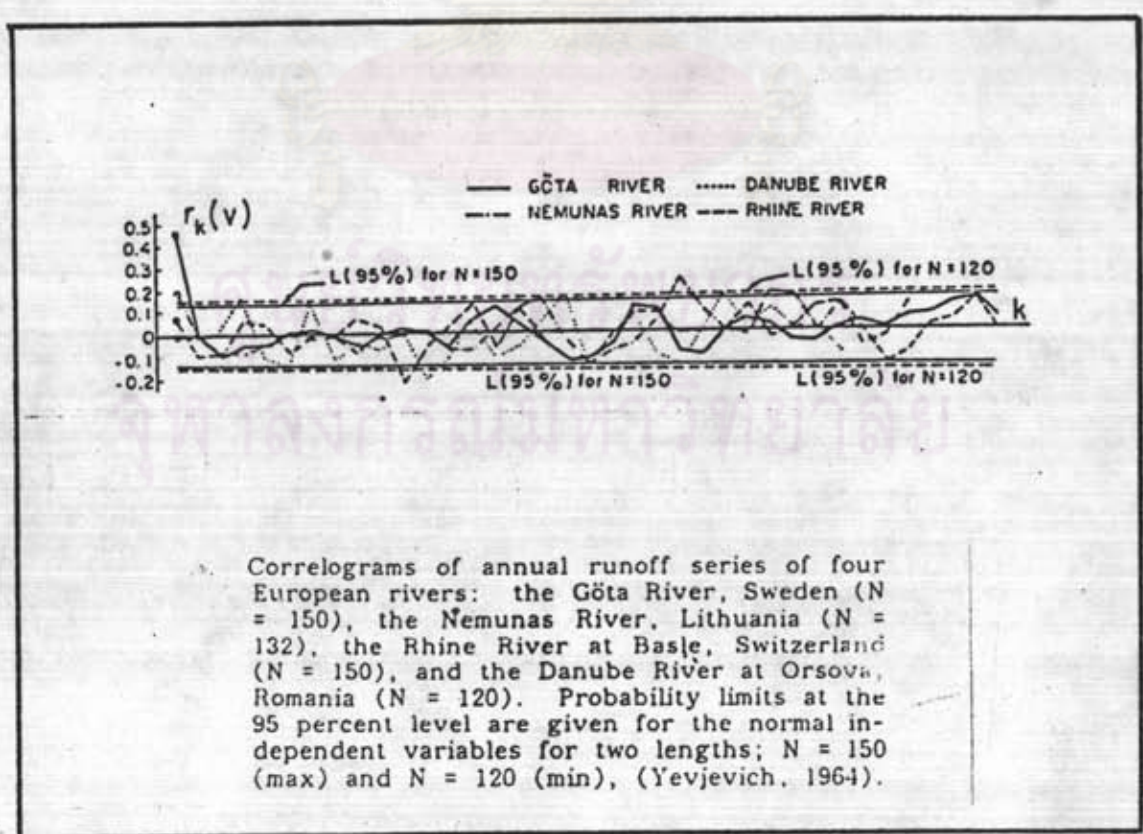
รูป 2-6 แสดงกราฟความสัมพันธ์ r_k ซึ่งหาจากสมการ 2-25 ของน้ำท่ารายปีสำหรับแม่น้ำ 4 สายในยุโรป และขอบเขตหาจากสมการ 2-26 ก สำหรับ $N = 150$ และ $N = 120$ ค่า r_1 ของแม่น้ำไรน์ (Rhine) และดานูบ (Danube) อยู่ในขอบเขต 95% ส่วนค่า r_1 ของแม่น้ำโกต้า (Gota) และ เนมูนีส (Nemunies) อยู่นอกขอบเขตนี้ [18]

2.4.3 การวิเคราะห์ทางสเปกตรอล (Spectral Analysis)

การวิเคราะห์ทางสเปกตรอล เป็นเทคนิคที่ใช้สำหรับศึกษาโครงสร้างของอนุกรมถิกวิธีหนึ่ง โดยเฉพาะใช้หาคุณสมบัติของวงจร (Cyclicity) การวิเคราะห์นี้มี 2 แบบคือการวิเคราะห์ค่าสเปกตรอลเชิงเส้น (Line Spectral Analysis) และการวิเคราะห์ค่าความหนาแน่นของสเปกตรอล (Variance Density Spectral Analysis)



รูป 2-5 ลักษณะของกราฟความสัมพันธ์ (Correlogram) (อ้างถึง 17)



รูป 2-6 กราฟความสัมพันธ์ (Correlogram) ของน้ำท่ารายปี
 ที่แม่น้ำ 4 สายในยุโรป (อ้างถึง 18)

2.4.3.1 การวิเคราะห์ค่าสเปกตรอลเชิงเส้น

การวิเคราะห์ค่าสเปกตรอลเชิงเส้น ใช้หาคุณสมบัติอนุกรมที่มีวงจรมอดูลค่าความถี่เชิงมุม (Angular frequency) ทั้งหมดที่เป็นไปได้ จากนั้นประมาณค่าแอมพลิจูดและเฟสของแต่ละความถี่เชิงมุมนั้นจากอนุกรม แล้วแทนในสมการ 2-27

$$\mu_t = \mu_x + \sum_{j=1}^m [A_j \cos \lambda_j t + B_j \sin \lambda_j t] \quad (2-27)$$

$$A_j = \frac{2}{N} \sum_{t=1}^N x_t \cos \lambda_j t, \quad j = 1, 2, \dots, m \quad (2-28)$$

$$B_j = \frac{2}{N} \sum_{t=1}^N x_t \sin \lambda_j t, \quad j = 1, 2, \dots, m \quad (2-29)$$

$$C_j = \frac{1}{2} [A_j^2 + B_j^2]^{1/2} \quad (2-30)$$

$$\phi_j = \tan^{-1} [B_j/A_j] \quad (2-31)$$

- โดย μ_t เป็นอนุกรมของส่วนที่เป็นวงจรมอดูลค่าความถี่เชิงมุม
- μ_x เป็นค่าเฉลี่ยของอนุกรมประมาณได้ด้วย \bar{x}
- A_j, B_j เป็นค่าสัมประสิทธิ์ฟูเรียร์ (Fourier Coefficient)
- C_j เป็นค่าแอมพลิจูด ซึ่งค่า $C_j^2/2$ เป็นค่าความแปรปรวนที่ความถี่นั้นๆ หรือเรียกว่าค่าสเปกตรอลเชิงเส้น
- ϕ_j เป็นค่าเฟส
- N เป็นจำนวนข้อมูล
- m เป็นจำนวนความถี่ที่เป็นไปได้ซึ่งขึ้นกับ N คือ ถ้า N เป็นเลขคู่ m มีค่า $N/2-1$ ถ้า N เป็นเลขคี่ m มีค่า $(N-1)/2$
- λ_j เป็นความถี่เชิงมุม มีค่า $2\pi f_j$
- f_j เป็นความถี่ของวงจรมอดูลค่าความถี่ มีค่า j/N

ในทางปฏิบัติ ไม่นิยมใช้ j ทั้ง m ค่า แต่จะเลือกใช้ j ที่มีความสำคัญเพียงบางตัวเท่านั้น ดังในรูป 2-7 แสดงการวิเคราะห์สเปกตรอลเชิงเส้นของข้อมูล 36 ค่า โดยรูป 2-7ก เป็นข้อมูล รูป 2-7ข เป็นค่าสเปกตรอลเชิงเส้นที่ความถี่ต่าง ๆ หรือเรียกกราฟนี้ว่ากราฟวงจร (Periodogram) รูป 2-7 ค แสดงเปรียบเทียบข้อมูลกับอนุกรมที่สร้างขึ้นจากความถี่ 5 ค่า [22]

สำหรับการเลือกความถี่ที่สำคัญสามารถพิจารณาจากกราฟวงจร แล้วเลือกค่าความถี่ที่ค่าสเปกตรอลเชิงเส้นสูง ๆ หรือสามารถพิจารณาจากกราฟวงจรสะสม (Cumulative Periodogram) ซึ่งเกิดจากการเขียนกราฟระหว่าง P_i กับ i ตามสมการ

$$P_i = \frac{\sum_{j=1}^i C_j^2 / 2}{S^2} \quad (2-32)$$

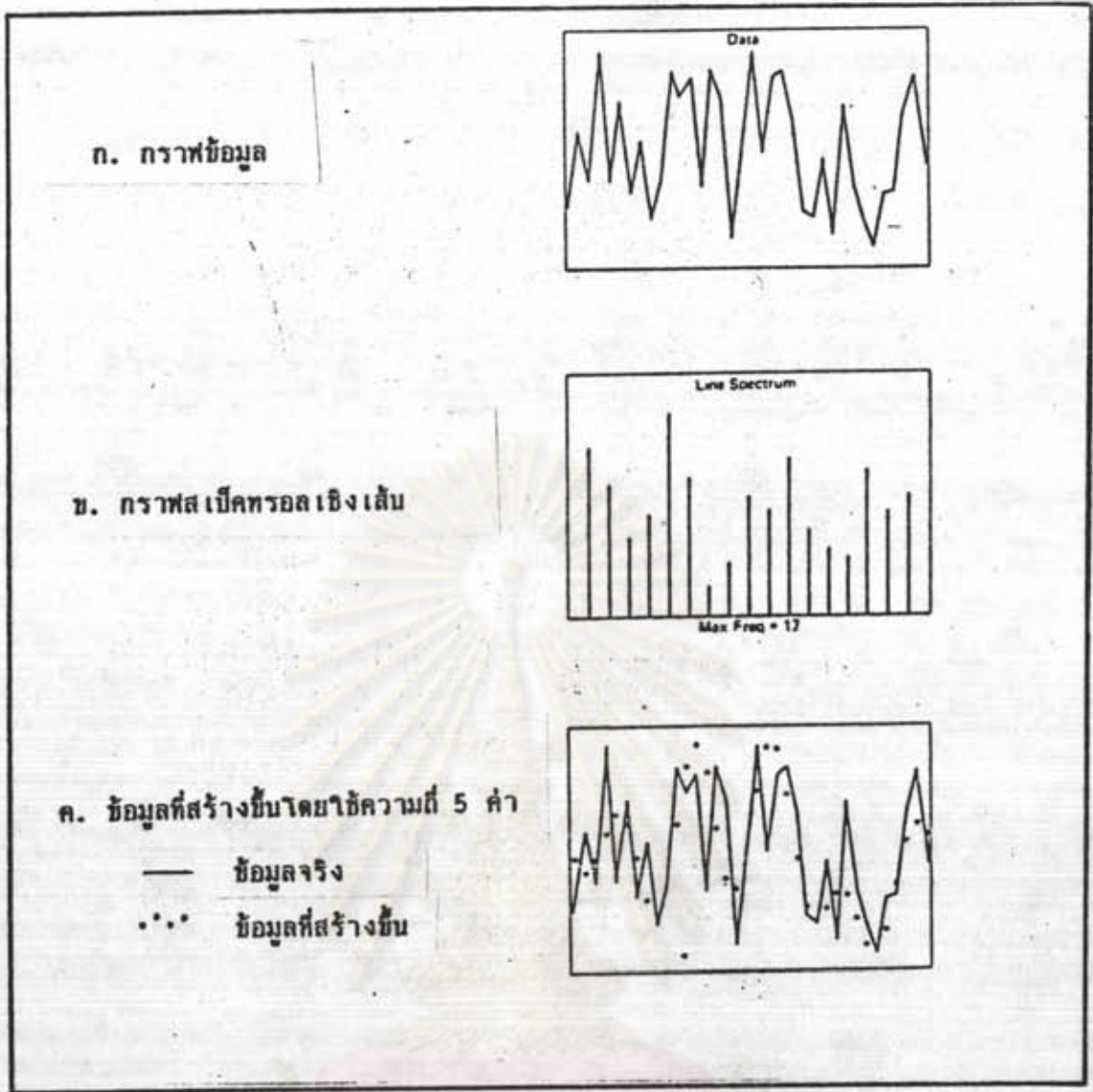
โดย P_i เป็นค่ากราฟวงจรสะสม
 $C_j^2 / 2$ เป็นค่าสเปกตรอลเชิงเส้น
 S^2 เป็นค่าความแปรปรวนของข้อมูลที่พิจารณา
 i เป็นจำนวนเต็ม มีค่า $1, 2, \dots, m$

โดยทั่วไปกราฟนี้จะประกอบด้วย 2 ส่วน ส่วนแรกเป็นส่วนที่เป็นวงจร ค่า P_i จะเพิ่มอย่างรวดเร็วเมื่อเพิ่ม i อีกส่วนเป็นส่วนที่ไม่แน่นอน ค่า P_i จะเพิ่มขึ้นอย่างช้า ๆ เมื่อเพิ่ม i จุดที่เกิดจากการปรับกราฟทั้ง 2 ส่วนให้เรียบต่อกัน จะเป็นจำนวนความถี่ของวงจรที่สำคัญตามต้องการ ดังแสดงในรูป 2-8 เป็นกราฟวงจรสะสมที่มีความถี่ของวงจรเท่ากับ h^* ถ้าอนุกรมมีตัวแปรที่เป็นอิสระต่อกัน กราฟนี้จะเป็นเส้นตรงเชื่อมโยงระหว่างจุด $(0,0)$ กับ $(m,1)$ หรืออาจจะขึ้นลงรอบเส้นตรงนี้ แต่ยังคงอยู่ในขอบเขตที่เชื่อถือได้ ซึ่งตามการทดสอบของโคลโมโกรอฟ-สไมร์นอฟ (Kolmogorov-Smirnov) [23] จะให้

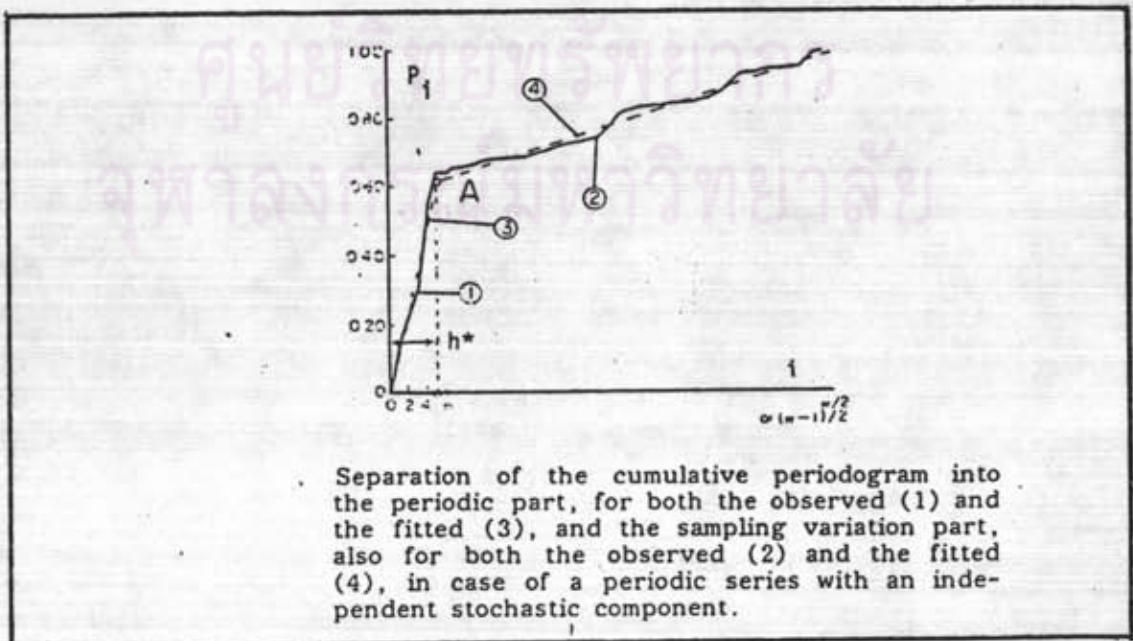
$$\Delta P_i (95\%) = \frac{+1.36}{\sqrt{m}} \quad (2-33ก)$$

$$\Delta P_i (75\%) = \frac{+1.02}{\sqrt{m}} \quad (2-33ข)$$

ซึ่ง $\Delta P_i (95\%)$ และ $\Delta P_i (75\%)$ เป็นค่าขอบเขตที่ผลต่างของ P_i ของข้อมูลกับ P_i ตามทฤษฎี มีความเชื่อมั่น 95% และ 75% ตามลำดับ



รูป 2-7 การวิเคราะห์สเปกตรัมเชิงเส้นของข้อมูลตัวอย่าง (อ้างอิง 22)



รูป 2-8 กราฟวงจรสะสม (Cumulative Periodogram) ของอนุกรมที่มีวงจร (อ้างอิง 18)

2.4.3.2 การวิเคราะห์ค่าความหนาแน่นของสเปกตรอล

การวิเคราะห์สเปกตรอลเชิงเส้นที่กล่าวมานั้นใช้ความถี่เชิงมุมที่ไม่ต่อเนื่อง ส่วนการวิเคราะห์ค่าความหนาแน่นของสเปกตรอลจะใช้ความถี่เชิงมุมอย่างต่อเนื่อง ซึ่งสามารถใช้กับอนุกรมทางอุทกวิทยาที่ต่อเนื่องได้ และยังประยุกต์ใช้กับอนุกรมทางอุทกวิทยาที่ไม่ต่อเนื่องได้ดีกว่าการวิเคราะห์สเปกตรอลเชิงเส้น เพราะสามารถประมาณค่าความถี่พื้นฐานและแอมพลิจูดของส่วนที่เป็นวงจรมีได้

จากสมการ 2-27 สามารถแปลงให้อยู่ในรูปเลขยกกำลังด้วยจำนวนเชิงซ้อนได้ คือ

$$f(t) = \mu_T - \mu_X = \sum_{j=1}^n C_j e^{i\lambda_j t} \quad (2-34)$$

$$C_j = \frac{2}{T} \int_0^T f(t) e^{-i\lambda_j t} dt \quad (2-35)$$

โดย i เป็นจำนวนเชิงซ้อน $= \sqrt{-1}$ และ T เป็นความยาวของอนุกรมแทนค่า C_j จากสมการ 2-35 ในสมการ 2-34 จะได้

$$f(t) = \mu_T - \mu_X = \sum_{j=1}^n \left[\frac{2}{T} \int_0^T f(t) e^{-i\lambda_j t} dt \right] e^{i\lambda_j t} \quad (2-36)$$

เขียนให้อยู่ในรูปฟูเรียร์อินทิกรัล (Fourier Integral) และฟูเรียร์ทรานส์ฟอร์ม (Fourier Transform) จะได้สมการ 2-37 และ 2-38 ตามลำดับ

$$f(t) = \frac{1}{\pi} \int_0^\infty \phi(\lambda) e^{i\lambda t} d\lambda \quad (2-37)$$

$$\phi(\lambda) = \int_0^\infty f(t) e^{i\lambda t} dt \quad (2-38)$$

ซึ่ง $\phi(\lambda)$ เป็นฟังก์ชันความหนาแน่นของสเปกตรอลสามารถอธิบายในรูปจำนวนเชิงซ้อนได้เป็น

$$\phi(\lambda) = \alpha(\lambda) + i\beta(\lambda) \quad (2-39)$$

$$\alpha(\lambda) = \int_0^\infty f(t) \cos \lambda t dt \quad (2-40)$$

$$\beta(\lambda) = \int_0^\infty f(t) \sin \lambda t dt \quad (2-41)$$

โดย $\alpha(\lambda)$, $\beta(\lambda)$ เป็นความหนาแน่นของ A_j และ B_j ที่แต่ละ λ

โดยทั่วไปนิยามพิจารณาความหนาแน่นของสเปกตรอลสำหรับความแปรปรวนหนึ่งหน่วย
 $u(\lambda) = \phi(\lambda)/\sigma_x^2$ ซึ่งฟังก์ชันนี้มีความสัมพันธ์กับฟังก์ชันความสัมพัทธ์ในตัวเอง $\rho(\tau)$
 สามารถเขียนสมการท่อนองเดียวกับสมการ 2-37 และ 2-38 ได้เป็น

$$u(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho(\tau) e^{i\lambda\tau} d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho(\lambda) \cos \lambda\tau d\tau \quad (2-42)$$

$$\rho(\tau) = \int_{-\infty}^{\infty} u(\lambda) e^{i\tau\lambda} d\lambda = \int_{-\infty}^{\infty} u(\lambda) \cos \tau\lambda d\lambda \quad (2-43)$$

เนื่องจากฟังก์ชันความสัมพัทธ์ในตัวเองสมมาตรรอบแกน y และใช้ λ อยู่ระหว่าง 0 ถึง ∞
 τ อยู่ระหว่าง 0 ถึง ∞ สามารถแทนในสมการ 2-42 และ 2-43 ได้

$$\gamma(\lambda) = \frac{2}{\pi} \int_0^{\infty} \rho(\tau) \cos \lambda\tau d\tau \quad (2-44)$$

$$\rho(\tau) = \int_0^{\infty} \gamma(\lambda) \cos \tau\lambda d\lambda \quad (2-45)$$

โดย $\gamma(\lambda)$ มีค่าเท่ากับ $2u(\lambda)$

ดังนั้นสามารถเขียนสมการ 2-27 เป็นฟังก์ชันต่อเนื่องได้เป็น

$$\mu_{\tau} = \mu_x + \sigma_x^2 \int_0^{\infty} \alpha(\lambda) \cos \tau\lambda d\lambda + \sigma_x^2 \int_0^{\infty} \beta(\lambda) \sin \tau\lambda d\lambda \quad (2-46)$$

ซึ่ง $\alpha(\lambda)$ และ $\beta(\lambda)$ เป็นฟังก์ชันต่อเนื่องประมาณได้จาก $a(\lambda)$ และ $b(\lambda)$
 ดังสมการ 2-40 และ 2-41 โดยใช้ความยาวของอนุกรม T ค่า $\alpha(\lambda)$ และ $\beta(\lambda)$
 นั้นสามารถเขียนให้สัมพันธ์กับ $\gamma(\lambda)$ ได้เป็น

$$\gamma(\lambda) = \frac{1}{2} [\alpha^2(\lambda) + \beta^2(\lambda)]^{1/2} \quad (2-47)$$

จากสมการ 2-44 และ 2-45 แปลงให้อยู่ในรูปของความถี่ได้เป็น

$$\gamma(f) = 4 \int_0^{\infty} \rho(\tau) \cos 2\pi f\tau d\tau \quad (2-48)$$

$$\rho(\tau) = \int_0^{\infty} \gamma(f) \cos 2\pi f\tau df \quad (2-49)$$

ถ้าใช้กับอนุกรมที่ไม่ต่อเนื่อง สมการ 2-44 และ 2-48 จะกลายเป็น

$$\gamma(\lambda) = \frac{1}{\pi} [1 + 2 \sum_{k=1}^{\infty} \rho(k) \cos \lambda k] \quad (2-50)$$

$$\gamma(f) = 2 [1 + 2 \sum_{k=1}^{\infty} \rho(k) \cos 2\pi f k] \quad (2-51)$$

ซึ่งค่าผลรวมทางขวามือของ 2-50 และ 2-51 เป็น 0 สำหรับค่าแปรที่เป็นอิสระต่อกัน จะได้ $\gamma(\lambda) = 1/\pi$ และ $\gamma(f) = 2$ โดย $0 < \lambda < \pi$ และ $0 < f < 0.5$

ในการประมาณค่า $\gamma(f)$ จากตัวอย่างซึ่งค่อนข้างจะผันแปรสูง จึงควรมีการปรับให้เรียบ (Smoothing) โดยสามารถคำนวณหาจาก

$$g(f) = 2 [1 + 2 \sum_{k=1}^m D(k) \rho(k) \cos 2\pi f k] \quad (2-52)$$

ซึ่งค่า m เป็น lag time มากที่สุดที่เป็นไปได้ของข้อมูลตัวอย่างมี N ข้อมูลจะได้ m มีค่า $N-1$ แต่โดยทั่วไปใช้ $m < N/4$ เพื่อให้เกิดความแม่นยำในการประมาณ และค่า $D(k)$ เป็นตัวปรับค่าฟังก์ชันตามเวลาให้เรียบมากขึ้น โดยทั่วไปนิยมปรับค่าฟังก์ชันตามความถี่ให้เรียบมากขึ้น โดยใช้ $S(f)$ ตามสมการ

$$S(f_j) = \frac{1}{4} g(f_{j-1}) + \frac{1}{2} g(f_j) + \frac{1}{4} g(f_{j+1}) \quad (2-53)$$

ซึ่ง $g(f)$ หาจากสมการ 2-51 ค่า $S(f_j)$ ในสมการนี้แปลงกลับเป็นตัวปรับฟังก์ชันตามเวลาได้เป็น

$$D(k) = (1 + \cos 2\pi f k) / 2 \quad (2-54)$$

ปกติ $S(f)$ ในสมการ 2-50 จะไม่ติดลบ อย่างไรก็ตาม ในการคำนวณอาจมีความคลาดเคลื่อนได้บ้าง ทำให้บางครั้งค่า $S(f)$ ติดลบ จึงมักแทนค่าติดลบนี้ด้วยศูนย์

เมื่อนำค่า $S(f)$ นี้ ไปเขียนกราฟสัมพันธ์กับความถี่ จะได้ฟังก์ชันความหนาแน่นของสเปกตรอล ดังในรูป 2-9 ค แสดงค่าความหนาแน่นของสเปกตรอล สำหรับข้อมูลน้ำฝนรายเดือนที่ Hachita, New Mexico ซึ่งหาจากกราฟความสัมพันธ์ในรูป 2-9 ข จากกราฟนี้สามารถหาความถี่พื้นฐานของวงจรต่าง ๆ ได้ โดยพิจารณาความถี่ที่ทำให้ค่าความหนาแน่นของสเปกตรอลสูง ๆ [17]

ถ้าข้อมูลมีความไม่แน่นอนมาก (Random Variable) กราฟความหนาแน่นของสเปกตรอลนี้ จะมีค่า $S(f) = 2$ หรือมีค่าขึ้นลงรอบ ๆ ค่านี้ ดังในรูป 2-10 โดยอยู่ในขอบเขตที่เชื่อถือได้ ซึ่งตามการทดสอบแบบไคส์แควร์ (Chi-square) [17] จะได้ ดังในรูป 2-10

$$P[S(f) > \frac{1 - \chi_{\alpha}^2(u)}{u} \gamma(f)] = \frac{\alpha}{2} \quad (2-55 \text{ ก})$$

$$P[S(f) \leq \frac{\chi_{\alpha}^2(u)}{u} \gamma(f)] = 1 - \frac{\alpha}{2} \quad (2-55 \text{ ข})$$

โดย u เป็นระดับของความเป็นอิสระมีค่า $\frac{8}{3} \frac{N}{m}$ ถึง $\frac{2N}{m}$

N เป็นจำนวนข้อมูลทั้งหมด

m เป็น lag time มากที่สุดที่พิจารณา

$\chi_{\alpha}^2(u)$ เป็นค่าจากตารางการแจกแจงไคส์แควร์

α เป็นค่าระดับความน่าจะเป็น

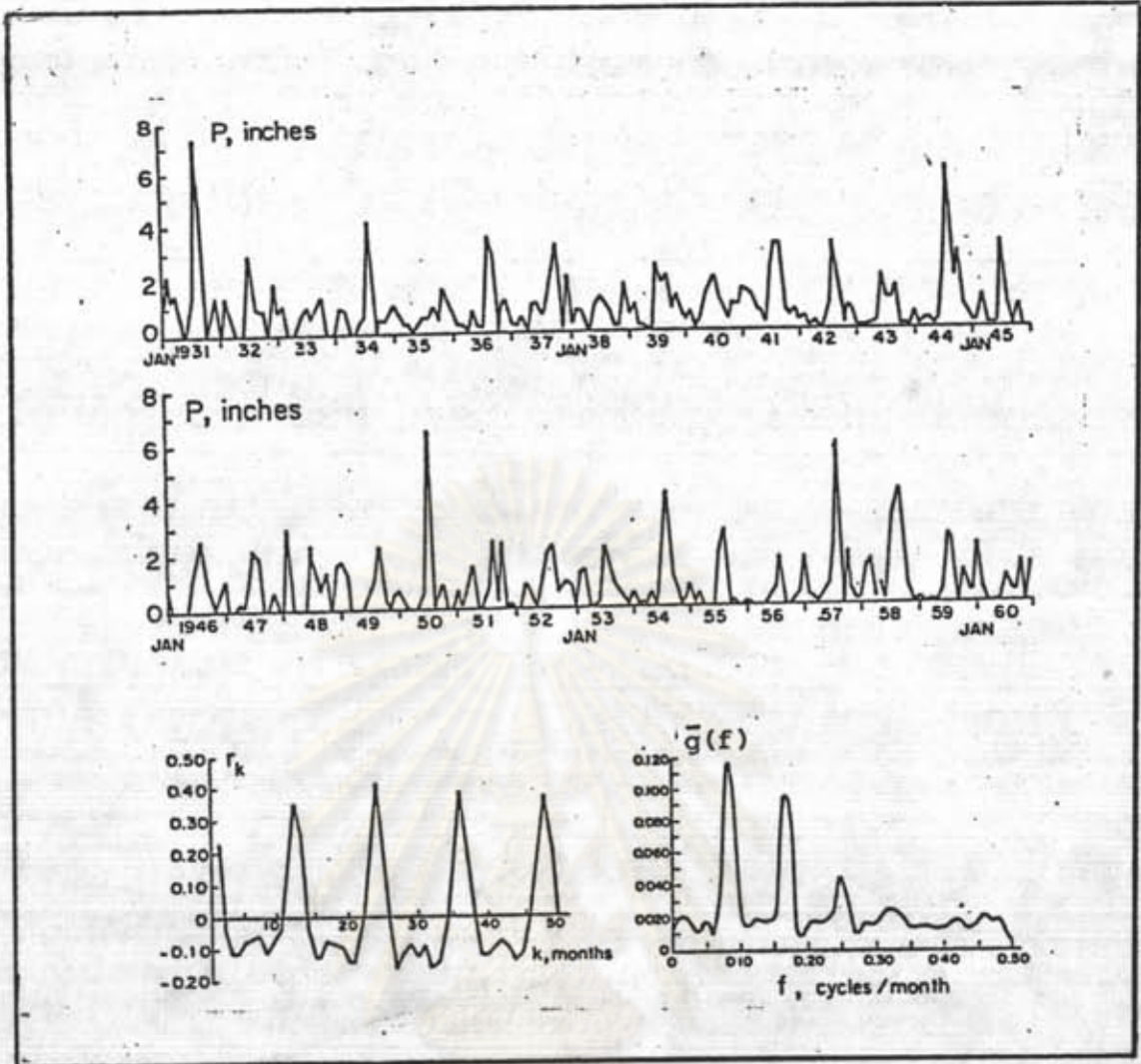
$\gamma(f)$ มีค่าเป็น 2

$\frac{1 - \chi_{\alpha}^2(u)}{u} \gamma(f)$ เป็นขอบเขตระดับล่าง

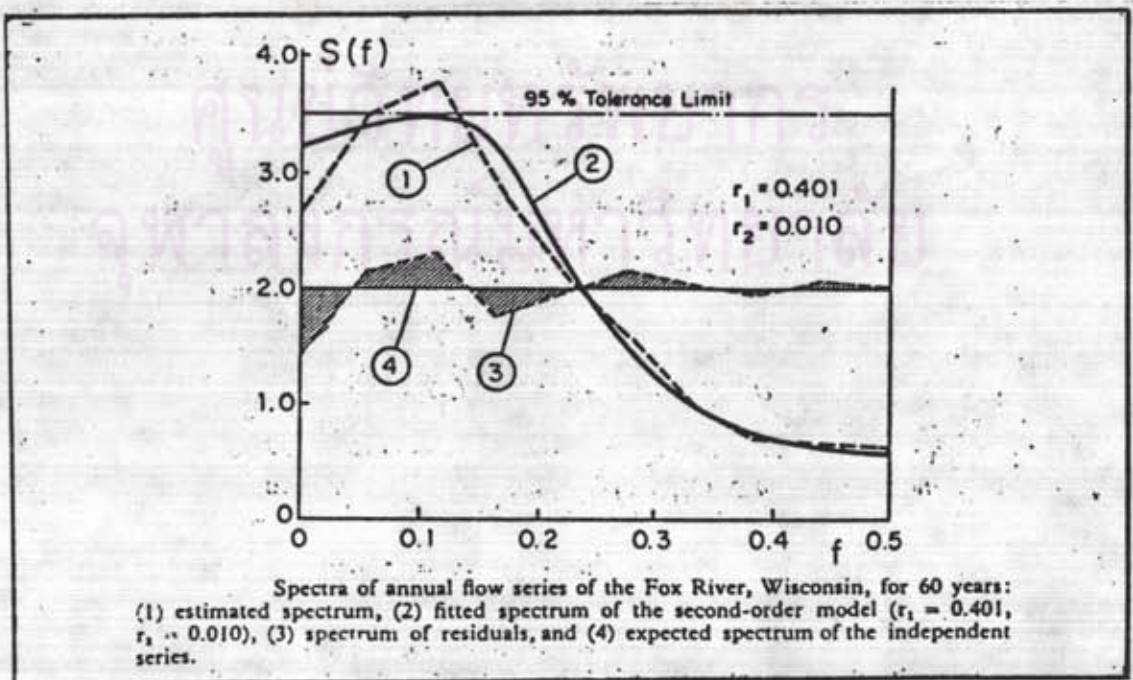
$\frac{\chi_{\alpha}^2(u)}{u} \gamma(f)$ เป็นขอบเขตระดับบน

สำหรับค่าความหนาแน่นของสเปกตรอลที่ $f = 0$ และ $f = 0.5$ จะมีค่ามากกว่าค่าความหนาแน่นของสเปกตรอลที่ f อื่น ๆ ดังนั้นค่าขอบเขตต้องเพิ่มขึ้น 2 เท่าของค่าปกติ ซึ่งในทางปฏิบัติมีความสำคัญมากเพราะค่า $g(f_0)$ และ $g(f_{0.5})$ นี้ มักมีค่าสูงกว่าขอบเขตของ f อื่น ๆ [17:106]

จุฬาลงกรณ์มหาวิทยาลัย



รูป 2-9 ชุดข้อมูลน้ำฝนรายเดือนที่ Hachita, New Mexico (1931-1960) พร้อมทั้งกราฟความสัมพันธ์ และความหนาแน่นทางสเปกตรอล (อ้างอิง 17)



รูป 2-10 ความหนาแน่นทางสเปกตรอลของปริมาณน้ำท่ารายปีที่แม่น้ำ Fox, Wisconsin (อ้างอิง 17)