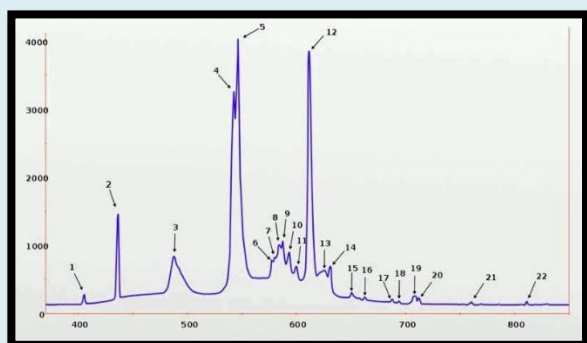
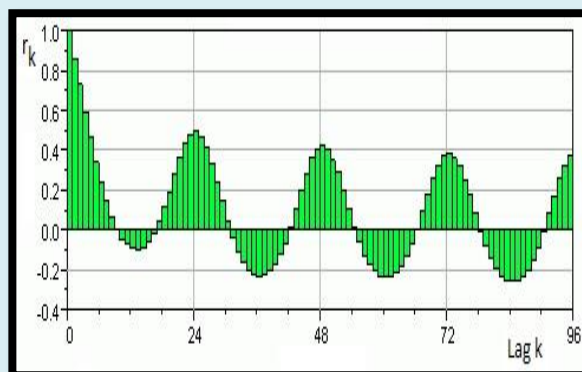
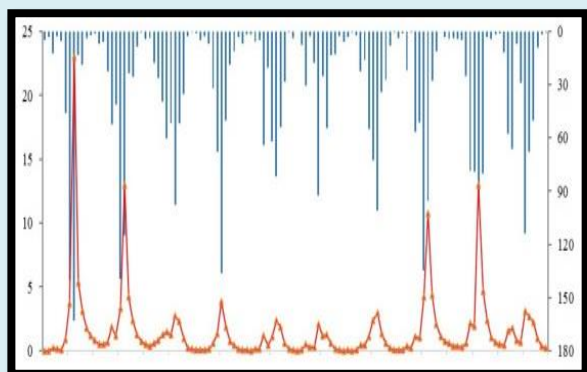


เอกสารคำสอน 02207651

การวิเคราะห์ทางสโตแคสติกสำหรับระบบชลประทาน STOCHASTIC ANALYSIS OF IRRIGATION SYSTEM



$$z_t = \sum_{j=1}^p \phi_j z_{t-j} - \sum_{j=0}^q \theta_j \varepsilon_{t-j}; \theta_0 = -1$$
$$\Phi(B^\omega)\phi(B)(1-B^\omega)^D(1-B)^d x_t = \Theta(B^\omega)\theta(B)\varepsilon_t$$
$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \\ \vdots \\ Z_t^n \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & \dots & a^{1n} \\ a^{21} & a^{22} & \dots & a^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a^{n1} & a^{n2} & \dots & a^{nn} \end{bmatrix} \begin{bmatrix} Z_{t-1}^1 \\ Z_{t-1}^2 \\ \vdots \\ Z_{t-1}^n \end{bmatrix} + \begin{bmatrix} b^{11} & b^{12} & \dots & b^{1n} \\ b^{21} & b^{22} & \dots & b^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b^{n1} & b^{n2} & \dots & b^{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \vdots \\ \varepsilon_t^n \end{bmatrix}$$

โดย
รองศาสตราจารย์ ดร.วรารุญ วุฒิมณีชัย
ภาควิชาวิศวกรรมชลประทาน
คณะวิศวกรรมศาสตร์ กำแพงแสน
มหาวิทยาลัยเกษตรศาสตร์

สิงหาคม 2563

คำนำ

ผู้เขียนตั้งใจเขียนเอกสารชุดนี้ขึ้นเพื่อใช้เป็นเอกสารคำสอนวิชา 02207651 การวิเคราะห์ทางสโตแคสติกของระบบชลประทาน ซึ่งเป็นวิชาในหลักสูตรวิศวกรรมศาสตรดุษฎีบัณฑิต ของภาควิชาวิศวกรรมชลประทาน คณะวิศวกรรมศาสตร์ กำแพงแสน มหาวิทยาลัยเกษตรศาสตร์ เนื้อหาส่วนใหญ่แปลและเรียบเรียงจากตำรา Applied Modeling of Hydrologic Time Series ซึ่งแต่งโดย Professor Dr. Jose D. Salas และคณะ จาก Colorado State University จากประสบการณ์ในการสอนวิชานี้ในช่วง 5 ปีที่ผ่านมา พบว่านิสิตที่ไม่มีพื้นฐานที่ดีเกี่ยวกับ “ความน่าจะเป็นและสถิติทางวิศวกรรม” มาก่อน จะมีปัญหาในการอ่านตำราต่างประเทศ ที่มีการอธิบายที่มาของสูตร หลักเกณฑ์ วิธีการและตัวอย่างต่างๆ อย่างรวบรัด ดังนั้นเอกสารชุดนี้จึงเน้นการอธิบายรายละเอียดหลักเกณฑ์และวิธีการต่างๆเพิ่มเติม เพื่อให้ผู้อ่านเข้าใจหลักและวิธีการที่กล่าวถึงในตำราของ Professor Dr. Jose D. Salas ได้ง่ายขึ้น

เอกสารคำสอนชุดนี้มี 10 บท 365 หน้า เนื้อหาของเอกสารคำสอนประกอบด้วย รายละเอียดกระบวนการและรูปแบบทางสโตแคสติก คุณลักษณะของอนุกรมเวลาทางอุทกวิทยา หลักการและเทคนิคทางสถิติสำหรับการจำลองอนุกรมเวลา แบบจำลองทางสโตแคสติกแบบต่างๆ เช่น แบบจำลอง Autoregressive, Autoregressive and Moving Average, Autoregressive Integrated Moving Average, Multivariate และ Disaggregation และข้อพิจารณาในการประยุกต์ใช้แบบจำลองสโตแคสติกในการจำลองอนุกรมเวลา การใช้แบบจำลอง Markov Chain สำหรับการจำลองตัวแปรสุ่มแบบต่อเนื่อง นอกจากนี้ยังได้เพิ่มรายละเอียดเกี่ยวกับทฤษฎีความน่าจะเป็น และการวิเคราะห์ความเสี่ยงไว้ในภาคผนวกท้ายเล่ม

ผู้เขียนหวังเป็นอย่างยิ่งว่าเอกสารคำสอนชุดนี้จะมีประโยชน์ต่อนิสิตและนักศึกษาที่เรียนวิชาการวิเคราะห์ทางสโตแคสติกทางวิศวกรรมชลประทานและอุทกวิทยา

รองศาสตราจารย์ ดร. วราวุธ วุฒิวณิชย์

ภาควิชาวิศวกรรมชลประทาน

คณะวิศวกรรมศาสตร์ กำแพงแสน

มหาวิทยาลัยเกษตรศาสตร์

สิงหาคม 2563

สารบัญ		หน้า
	หน้าปก	
	คำนำ	(i)
	สารบัญ	(ii)
บทที่	หัวข้อเรื่อง	
1	บทนำ	1-1
	1.1 คำนำ (Introduction)	1-1
	1.2 กระบวนการสโตแคสติก (Stochastic Process)	1-2
	1.3 รูปแบบของกระบวนการสโตแคสติก (Types of Stochastic Process)	1-2
	1.3.1 กระบวนการสโตแคสติกแบบ Discrete vs. Continuous	1-2
	1.3.2 Independent(Purely Random) vs. Dependent Process	1-5
	1.3.3 Stationary vs. Non-stationary Process	1-7
	1.4 องค์ประกอบของอนุกรมเวลา (Components of Time Series)	1-9
	1.5 คุณสมบัติพื้นฐานของกระบวนการสโตแคสติก	1-14
	1.6 แบบจำลองอนุกรมเวลาแบบง่าย (Simple Time Series Model)	1-16
	1.7 ค่าสถิติหลักของแบบจำลองสโตแคสติกแบบง่าย	1-19
	1.8 แบบจำลองสโตแคสติก	1-19
	1.9 Systematic Approach สำหรับการจำลองทางอุทกวิทยา	1-20
	1.10 การประยุกต์การจำลองอนุกรมเวลาทางอุทกวิทยา	1-22
	1.11 รูปแบบจำลองอนุกรมเวลาที่จะกล่าวถึงในบทต่างๆ	1-23
	1.12 เอกสารอ้างอิง	1-27
	1.13 แบบฝึกหัดท้ายบท	1-27
2	คุณลักษณะของอนุกรมเวลาทางอุทกวิทยา	2-1
	2.1 ชนิดของอนุกรมเวลา (Types of Time Series)	2-1
	2.2 ช่วงเวลาของอนุกรมเวลา (Time Interval)	2-1
	2.3 คุณสมบัติทั่วไปของอนุกรมเวลาทางอุทกวิทยา	2-2
	2.4 สหสัมพันธ์ในตัวเอง (Autocorrelation)	2-7
	2.5 Spectral Analysis	2-7

2.5.1	Spectral Density Function Versus Autocorrelation Function	2-8
2.5.2	วิธีการคำนวณหาค่าประมาณของ SDF S_f	2-9
2.6	คุณสมบัติทางสถิติพื้นฐานของอนุกรมเวลา	2-12
2.7	คุณสมบัติที่เกี่ยวกับภาวะความแห้งแล้งของอนุกรมเวลา	2-14
2.7.1	Run Length	2-14
2.7.2	Run Sum	2-15
2.8	Storage Related Characteristics of Time Series	2-15
2.9	การทดสอบ Non-homogeneity และ Inconsistency ของอนุกรมเวลาทางอุทกวิทยา	2-17
2.10	คุณสมบัติของอนุกรมเวลารายปี (Characteristics of Annual Time Series)	2-19
2.10.1	Correlogram Test	2-20
2.10.2	สรุปคุณสมบัติของอนุกรมเวลารายปี	2-22
2.11	คุณสมบัติของอนุกรมเวลาแบบ Periodic	2-23
2.12	คุณสมบัติของอนุกรมเวลาแบบ Multivariate	2-25
2.13	เอกสารอ้างอิง	2-28
2.14	แบบฝึกหัดท้ายบท	2-28
3	หลักการและเทคนิคทางสถิติสำหรับการจำลองอนุกรมเวลา	3-1
3.1	เทคนิคการประมาณค่าพารามิเตอร์เบื้องต้น (Basic Estimation Techniques)	3-1
3.1.1	วิธีการตรวจสอบความเหมาะสมของการประมาณค่าพารามิเตอร์	3-1
3.1.2	ค่าประมาณแบบ Unbiased (Unbiased Estimator)	3-2
3.1.3	Efficient Estimator	3-2
3.2	Method of Moments	3-2
3.3	Method of Least Square Estimate	3-3
3.4	Method of Maximum Likelihood	3-4
3.5	การประมาณค่าพารามิเตอร์โดยวิธี Regionalization	3-6
3.6	การแปลงข้อมูลให้เป็นการแจกแจงปกติ (Normalization)	3-6
3.7	การหาพารามิเตอร์แบบ Periodic โดยใช้ Fourier Series	3-7
3.7.1	วิธีการประมาณค่าแบบ Non-Parametric Approach	3-8
3.7.2	การประมาณค่าพารามิเตอร์โดยวิธี Fourier Series	3-8
3.7.3	การประมาณค่า Fourier Series Coefficients	3-8
3.7.4	Cumulative Periodogram Test (P_c)	3-9

3.8	การประมาณค่าพารามิเตอร์ของแบบจำลองหลายตัวแปร	3-12
3.8.1	วิธีการถอด Square Root ของเมทริกซ์	3-12
3.9	การทดสอบความเหมาะสมของแบบจำลอง (Goodness of Fit Test)	3-13
3.9.1	การทดสอบความเป็นอิสระ (Test of Independence)	3-13
3.9.2	การทดสอบความเป็นปกติ (Test of Normality)	3-16
3.9.3	การทดสอบจำนวนพารามิเตอร์ที่พอเหมาะ (Parsimony of Parameters)	3-17
3.10	การสังเคราะห์ข้อมูลและการพยากรณ์ (Generation and Forecasting)	3-17
3.10.1	การสังเคราะห์ข้อมูล (Generation of Synthetic Samples)	3-17
3.10.2	การพยากรณ์ (Forecasting)	3-20
3.11	เอกสารอ้างอิง	3-21
3.12	แบบฝึกหัดท้ายบท	3-22
4	การจำลองอนุกรมเวลาด้วย แบบจำลอง Autoregressive	4-1
4.1	รายละเอียดของแบบจำลอง AR	4-1
4.1.1	สมการคณิตศาสตร์ของแบบจำลอง AR	4-1
4.1.2	Properties of AR Models with Constant Parameters	4-3
4.1.3	Properties of AR Model with Periodic Parameters	4-7
4.2	แบบจำลอง AR สำหรับอนุกรมเวลารายปี (AR Modeling of Annual Time Series)	4-8
4.2.1	แบบจำลอง AR	4-8
4.2.2	การประมาณค่าพารามิเตอร์ของแบบจำลอง Annual AR	4-9
4.2.3	วิธีการทดสอบความเหมาะสมสำหรับแบบจำลอง Annual AR	4-12
4.2.4	การสังเคราะห์ข้อมูลใหม่โดยแบบจำลอง Annual AR	4-13
4.3	Periodic AR Model	4-23
4.3.1	AR with Constant Coefficient	4-23
4.3.2	AR with Periodic Coefficient	4-23
4.3.3	การประมาณค่าพารามิเตอร์สำหรับ Periodic AR Models	4-24
4.3.4	ความน่าเชื่อถือของพารามิเตอร์แบบ Periodic	4-25
4.3.5	ความเหมาะสมของแบบจำลอง Periodic AR	4-26
4.3.6	การสังเคราะห์ข้อมูลโดยใช้แบบจำลอง Periodic AR	4-27
4.3.7	ตัวอย่างการจำลองอนุกรมเวลาแบบ Periodic ด้วยแบบจำลอง AR	4-28
4.4	เอกสารอ้างอิง	4-39

4.5	ภาคผนวก	4-40
5	การจำลองอนุกรมเวลาด้วยแบบจำลอง Autoregressive and Moving Average	5-1
5.1	คำนำ (Introduction)	5-1
5.2	รูปแบบทางคณิตศาสตร์ของแบบจำลอง ARMA(p,q)	5-1
5.3	คุณสมบัติของ MA(q) (Properties of MA(q))	5-1
5.3.1	แบบจำลอง MA(q)	5-1
5.3.2	แบบจำลอง MA(1) จะมีลักษณะเป็น One Step Long Memory	5-2
5.3.3	Residual ของ MA(1)	5-2
5.3.4	Invertible Conditions	5-2
5.3.5	Partial Autocorrelation Function (PACF) ของ MA(q)	5-3
5.3.6	คุณสมบัติทั่วไปของ ARMA(p, q)	5-3
5.4	การสร้างแบบจำลอง ARMA สำหรับอนุกรมเวลารายปี	5-11
5.4.1	แบบจำลอง Annual ARMA	5-11
5.4.2	ขั้นตอนการจำลอง	5-11
5.4.3	การทดสอบความเหมาะสมของแบบจำลอง Annual ARMA(p, q)	5-16
5.5	การสังเคราะห์ข้อมูล (Generation)	5-16
5.6	การพยากรณ์ด้วยแบบจำลอง Annual ARMA	5-17
5.6.1	หลักการพยากรณ์ทางสโตแคสติก	5-17
5.6.2	คุณสมบัติของ $E(\mathcal{E}_{t+h})$	5-17
5.6.3	One step Ahead Forecasting ($L = 1$)	5-18
5.6.4	Forecasted Function	5-18
5.6.5	Forecast Error	5-19
5.7	การจำลองอนุกรมเวลารายปีด้วยแบบจำลอง ARMA(p,q)	5-19
5.8	ตัวอย่างการจำลองอนุกรมเวลาปริมาณน้ำท่ารายปีด้วยแบบจำลอง ARMA(p,q)	5-21
5.9	การจำลองด้วย ARMA สำหรับอนุกรมเวลาแบบ Periodic	5-45
5.9.1	ขั้นตอนการจำลอง	5-45
5.9.2	ตัวอย่างการจำลองอนุกรมเวลาแบบ Periodic ด้วย ARMA with Constant Coefficients	5-46
5.9.3	ตัวอย่างการจำลองอนุกรมเวลาแบบ Periodic ด้วย ARMA with Periodic Coefficients	5-51

5.10	เอกสารอ้างอิง	5-54
5.11	ภาคผนวก	5-55
6	การจำลองอนุกรมเวลาด้วยแบบจำลอง Autoregressive Integrated Moving Average	6-1
6.1	คำนำ	6-1
6.2	รูปแบบของแบบจำลอง ARIMA	6-1
6.3	วิธีการ Differencing	6-2
6.4	หลักการ Infinite Variance Concept	6-3
6.5	แบบจำลอง ARIMA (p, 1, q)	6-3
6.6	ฟังก์ชันการพยากรณ์ (Forecasted Function) ของแบบจำลอง ARIMA(p,1,q) สำหรับ Lead Time เท่ากับ L	6-6
6.7	แบบจำลอง Multiplicative ARIMA สำหรับอนุกรมเวลาแบบ Periodic	6-7
6.7.1	Seasonal Differencing	6-7
6.7.2	Simple Seasonal ARIMA(P,1,Q) ω	6-7
6.7.3	General Simple Seasonal ARIMA(P,D,Q) ω	6-8
6.7.4	ARIMA(0,1,Q) ω หรือ IMA(1,Q) ω	6-8
6.7.5	ARIMA (P,1,0) ω	6-8
6.7.6	การพิต 1 st Order Seasonal Difference ของอนุกรมเวลารายเดือนด้วยแบบจำลอง AR(1)	6-9
6.7.7	การพิต 1 st Order Seasonal Difference ของอนุกรมเวลารายเดือนด้วยแบบจำลอง MA(1)	6-9
6.7.8	แบบจำลอง Multiplicative ARIMA(1, 0, 0)x(0, 1, 1) $_{12}$	6-9
6.7.9	แบบจำลอง ARIMA(p,0,q)x(P, 1, Q) $_{12}$	6-10
6.7.10	แบบจำลอง Multiplicative ARIMA(p, d, q)x(P, D, Q) $_{12}$	6-10
6.8	การประมาณค่าพารามิเตอร์ของ Simple ARIMA(p,d,q)	6-11
6.9	การตรวจสอบ Goodness of Fit ของ ARIMA(p,d,q)	6-11
6.10	ขั้นตอนการจำลองด้วย Simple ARIMA	6-11
6.11	ตัวอย่างการสร้างแบบจำลอง Simple ARIMA	6-12
6.12	ตัวอย่างการสร้างแบบจำลอง Multiplicative ARIMA	6-16
6.13	ข้อจำกัดที่ต้องพิจารณาในการนำแบบจำลอง Multiplicative ARIMA ไปประยุกต์ใช้	6-32
6.14	การเปรียบเทียบและข้อจำกัดของแบบจำลอง ARMA และ ARIMA	6-34

6.15	เอกสารอ้างอิง	6-41
6.16	ภาคผนวก	6-42
7	การจำลองอนุกรมเวลาทางอุทกวิทยาด้วยแบบจำลองหลายตัวแปร	7-1
7.1	คำนำ	7-1
7.2	แบบจำลองหลายตัวแปรลำดับที่หนึ่ง (Multivariate Lag One)	7-1
7.3	Lag k Correlation Matrix	7-2
7.4	การหาพารามิเตอร์ของแบบจำลองหลายตัวแปร (Multivariate Model)	7-3
7.4.1	การหาค่าพารามิเตอร์ของ Multivariate AR(1)	7-3
7.4.2	การหาค่าพารามิเตอร์ของ Multivariate AR(2)	7-5
7.5	การสังเคราะห์ข้อมูลใหม่ (Generation of Synthetic Data)	7-6
7.6	Approximate Multivariate ARMA(p,q)	7-7
7.7	การทดสอบความเหมาะสมของแบบจำลองรายตัวแปรรายปี	7-8
7.8	ตัวอย่างการจำลองอนุกรมเวลารายปีด้วย Multivariate AR(p)	7-8
7.8.1	การวิเคราะห์เบื้องต้นและการระบุรูปแบบของแบบจำลอง	7-9
7.8.2	การประมาณค่าพารามิเตอร์	7-9
7.8.3	การทดสอบ Goodness of Fit ของแบบจำลอง	7-12
7.8.4	Optional Tests of the Model	7-14
7.8.5	Reliability ของพารามิเตอร์แบบจำลอง	7-15
7.9	ข้อจำกัดของ Multivariate Annual Modeling	7-16
7.10	การประยุกต์ใช้แบบจำลองแบบหลายตัวแปรกับอนุกรมเวลารายปี (Annual Multivariate Model) ในทางปฏิบัติ	7-16
7.11	การจำลองแบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic	7-17
7.11.1	การประยุกต์ใช้แบบจำลอง AR และ ARMA แบบ Constant พารามิเตอร์กับกรณีหลายตัวแปร	7-17
7.11.2	แบบจำลอง AR หลายตัวแปรแบบ Periodic พารามิเตอร์	7-18
7.11.3	แบบจำลอง ARMA หลายตัวแปรแบบ Periodic พารามิเตอร์	7-19
7.12	ตัวอย่างการจำลองแบบหลายตัวแปรกับอนุกรมเวลา Periodic	7-20
7.13	ข้อจำกัดของการจำลองแบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic	7-35
7.14	การประยุกต์แบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic	7-35
7.15	เอกสารอ้างอิง	7-35

8	การจำลองอนุกรมเวลาด้วยแบบจำลองแบบแยกส่วน	8-1
	8.1 คำนำ	8-1
	8.2 แบบจำลองแยกส่วนทั่วไป (General Disaggregation Models)	8-1
	8.3 แบบจำลองแยกส่วนตามเวลาแบบ Single Site	8-3
	8.4 แบบจำลองแยกส่วนตามเวลาแบบ Multisite	8-4
	8.5 แบบจำลองแยกส่วนตามพื้นที่ (Spatial Disaggregation Model)	8-5
	8.6 แบบจำลองแยกส่วนตามพื้นที่และตามเวลา	8-6
	8.7 การประมาณค่าพารามิเตอร์ของแบบจำลอง	8-6
	8.7.1 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาพื้นฐาน	8-6
	8.7.2 ขั้นตอนการคำนวณหาพารามิเตอร์ (Calculation Procedure)	8-7
	8.7.3 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบขยาย	8-10
	8.7.4 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบย่อ	8-11
	8.7.5 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามพื้นที่แบบขยายกับ	8-12
	กรณี Multisite	
	8.7.6 วิธีการคำนวณค่า Variance-Covariance Matrices สำหรับแบบจำลองแยกส่วน	8-13
	ตามเวลาแบบย่อ (Condensed Model)	
	8.7.7 วิธีการคำนวณค่า Variance-Covariance Matrices สำหรับแบบจำลองแยกส่วน	8-13
	ตามพื้นที่ (Spatial)	
	8.8 การทดสอบความเหมาะสมของแบบจำลอง (Goodness of Fit Test)	8-14
	8.9 ตัวอย่างการจำลองแบบแยกส่วน	8-14
	8.10 ประโยชน์และข้อจำกัดของการจำลองแยกส่วน	8-19
	8.11 ตัวอย่างการประยุกต์ใช้แบบจำลองแยกส่วน	8-21
	8.12 เอกสารอ้างอิง	8-22
	8.13 ภาคผนวก	8-23
9	ข้อพิจารณาในการประยุกต์ใช้แบบจำลองสโตแคสติก	9-1
	9.1 การเตรียมข้อมูลอนุกรมเวลา (Pretreatment of Historical Data)	9-1
	9.1.1 การแปลความหมายข้อมูล (Data Compilation)	9-1
	9.1.2 การเติมข้อมูลที่ขาดหายและการต่อขยายข้อมูล (Data Fill-in and Extension)	9-2
	9.1.3 การปรับข้อมูลให้เป็นไปตามสภาพธรรมชาติ	9-2

9.2 การเลือกแบบจำลอง (Model Selection)	9-3
9.2.1 สิ่งสำคัญเกี่ยวกับสถิติของอนุกรมเวลาตัวอย่าง	9-3
9.2.2 การรักษาคุณสมบัติทางสถิติของอนุกรมเวลาตัวอย่าง	9-3
9.2.3 วัตถุประสงค์ในการสังเคราะห์ข้อมูล (Propose for Generation)	9-4
9.2.4 ความอ่อนไหวหรือความไวของผลลัพธ์ (Sensitivity of Results) การวิเคราะห์ ทั้งลุ่มน้ำ (Reginald Analysis)	9-4
9.2.5 การวิเคราะห์ทั้งลุ่มน้ำ	9-4
9.3 การประยุกต์ใช้แบบจำลอง (Model Application)	9-5
9.3.1 การหาขนาดอ่างเก็บน้ำ (Reservoir Sizing Studies)	9-5
9.3.2 การศึกษาการปฏิบัติการอ่างเก็บน้ำ (Reservoir Operation Studies)	9-5
9.3.3 การศึกษาลุ่มน้ำ (Basin-Wide Studies)	9-6
9.4 ข้อจำกัดของแบบจำลอง (Model Limitations)	9-6
9.4.1 Short-and Long-Term Persistence	9-6
9.4.2 ข้อจำกัดของแบบจำลองแบบ (Periodic)	9-7
9.4.3 ข้อจำกัดของแบบจำลอง Disaggregation	9-7
9.4.4 สรุป	9-7
9.5 เอกสารอ้างอิง	9-8
10 Markov Chain Modelling	10-1
10.1 Definition	10-1
10.2 States of the system and theirs probabilities	10-1
10.3 Conditional probability and transition matrix	10-2
10.4 Markov chain model	10-4
10.4.1 First order Markov chain model	10-4
10.4.2 n th order Markov chain model	10-6
10.5 A Simple Weather Markov Chain Model	10-6
10.5.1 Predicting the weather condition	10-7
10.5.2 Steady state of the weather condition	10-8
10.6 The Cola Example	10-10
10.7 Classification of State of Markov Chain	10-15
10.7.1 State Classification - Population distribution example	10-15

10.7.2 State Classification - Coin tossing example	10-17
10.8 Steady State Probabilities and Mean First Passage Times	10-20
10.9 Types of States in a Markov Chain	10-21
10.10 References	10-25
10.11 Exercises	10-26

Appendix

Appendix A: Random Variable and Probability Distribution Functions	Appendix A-1
A.1 Random Variable	Appendix A-1
A.2 Probability Distribution Functions of Random Variable	Appendix A-1
A.3 Properties of Random Variables and Related Theory	Appendix A-2
A.3.1 คุณสมบัติสำคัญที่ควรทราบเกี่ยวกับตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)	Appendix A-5
A.3.2 คุณสมบัติสำคัญที่ควรทราบเกี่ยวกับตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)	Appendix A-5
A.3.3 การแจกแจงความน่าจะเป็นร่วม (Joint Probability Distribution)	Appendix A-5
A.3.4 การแจกแจงแบบมีเงื่อนไข (Conditional Probability Distributions)	Appendix A-6
A.3.5 การเป็นอิสระของตัวแปรสุ่ม (Independence)	Appendix A-6
A.4 ทฤษฎีของ Bayes (Bayes' Theorem)	Appendix A-7
A.5 การคาดคะเนทางคณิตศาสตร์ (Mathematical Expectation)	Appendix A-7
A.6 การหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มหรือฟังก์ชันของตัวแปรสุ่ม	Appendix A-7
A.7 การหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มที่เป็นฟังก์ชันของตัวแปรสุ่มร่วมกัน 2 ตัว	Appendix A-8
A.8 การหาความแปรปรวน (Variance) ของตัวแปรสุ่มและฟังก์ชันของตัวแปรสุ่ม	Appendix A-9
A.9 กฎเกี่ยวกับ Variance	Appendix A-10
A.10 ความแปรปรวนร่วมกันของตัวแปรสุ่ม (Covariance)	Appendix A-10
A.11 สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)	Appendix A-11
A.12 สัมประสิทธิ์ความเบ้ (Skewness Coefficient)	Appendix A-11
A.13 สัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Autocorrelation Coefficient)	Appendix A-11
A.14 แบบฝึกหัด	Appendix A-12
Appendix B: Review Basic Matrix Operation	Appendix B-1
Appendix C: First Order Analysis of Uncertainty	Appendix C-1
C.1 นิยาม First Order Analysis of Uncertainty	Appendix C-1
C.2 Taylor Series Expansion	Appendix C-1
C.3 การวิเคราะห์ความไม่แน่นอนอันดับที่ 1 ของสมการ Manning	Appendix C-2

C.4 ตัวอย่างการใช้ FAU วิเคราะห์ความเสี่ยง

Appendix C-4

C.5 ความเสี่ยง (Risk)

Appendix C-5

บทที่ 1

บทนำ

INTRODUCTION

1.1 คำนำ (Introduction)

กระบวนการแบบ Deterministic คือกระบวนการที่สามารถคาดการณ์ผลลัพธ์ได้อย่างได้อย่างถูกต้องแน่นอน ยกตัวอย่างเช่น ถ้าคุณฝากเงินธนาคาร 1 ล้านบาทวันนี้ โดยมีอัตราดอกเบี้ย 10% คุณจะสามารถคำนวณได้ว่าคุณจะมีเงินเท่าใดอีก 5 ปีข้างหน้า ซึ่งแตกต่างจากกระบวนการแบบสโตแคสติก(Stochastic) ซึ่งคุณจะไม่สามารถคาดเดาได้อย่างแม่นยำว่าผลลัพธ์จะออกมาเป็นอย่างไร เพียงแต่สามารถที่จะอธิบายความน่าจะเป็นของผลลัพธ์นั้นๆได้โดยใช้ทฤษฎีความน่าจะเป็น ยกตัวอย่าง คุณมาสามารถจะบอกแน่นอนว่าอัตราดอกเบี้ยอีก 5 ปีข้างหน้าจะมีค่าเท่าไร แต่สามารถบอกได้ว่ามีโอกาสมากหรือน้อยเท่าใดที่อัตราดอกเบี้ยจะสูงหรือต่ำกว่า 10 %

ในการทำงานร่วมกัน การคำนวณออกแบบในทางวิศวกรรม เช่นการออกแบบคานและเสา ใช้กระบวนการทาง Deterministic เพื่อให้การออกแบบทำได้ง่าย โดยการนำผลการวิเคราะห์น้ำหนักบรรทุกไปคำนวณหาแรงเฉือนและโมเมนต์ตามสูตร กำหนดความสามารถรับแรงของคอนกรีตและเหล็ก ก็สามารถออกแบบหาขนาดและขนาดเหล็กเสริมของคานและเสาได้ตามข้อกำหนดในการออกแบบที่กำหนดให้ แต่ในความเป็นจริงแล้วค่าน้ำหนักบรรทุกและความสามารถในการรับแรงของวัสดุจริงไม่ใช่ตัวแปรแบบ Deterministic ไม่สามารถคาดการณ์ค่าจริงที่จะเกิดขึ้นได้ ดังนั้นในการออกแบบทางวิศวกรรมจึงต้องมีการกำหนดค่าความปลอดภัย (Safety Factor) ในการออกแบบเพื่อรองรับความไม่แน่นอนที่อาจเกิดขึ้น เช่นการกำหนดให้ใช้ f_c เท่ากับ $0.45f_c'$ หรือ f_s เท่ากับ $0.5f_y$ เป็นต้น

การออกแบบทางวิศวกรรมชลประทานในปัจจุบัน บางส่วนยังคำนวณออกแบบตามหลักการแบบ Deterministic แบบง่ายๆ โดยไม่ได้มีการพิจารณาถึงข้อเท็จจริงว่าตัวแปรที่มีผลต่อการออกแบบส่วนใหญ่เป็นตัวแปรสโตแคสติก เช่นการคำนวณออกแบบขนาดอ่างเก็บน้ำโดยใช้ข้อมูลปริมาณน้ำฝน-น้ำท่าในอดีตและความต้องการน้ำที่วิเคราะห์ได้ โดยไม่ทราบว่าขนาดอ่างเก็บน้ำที่ออกแบบนั้นมีความเสี่ยงที่จะเกิดขนาดน้ำมากน้อยเท่าใด มีค่าสัมประสิทธิ์ความปลอดภัยสูงหรือต่ำเท่าใด มากน้อยเท่าใด ดังนั้นการวิเคราะห์และการจำลองทางสโตแคสติกจึงเป็นวิชาที่นักศึกษาชั้นค้ำ เพื่อใช้เป็นเครื่องมือในการสังเคราะห์ข้อมูลใหม่ ซึ่งจะมีประโยชน์ในการจำลองสถานการณ์ของอ่างเก็บน้ำที่ออกแบบว่ามีความเสี่ยงต่อการเกิดกรขนาดน้ำเท่าใด ควรต้องเพื่อค่าความปลอดภัยมากน้อยเท่าใด เท่าที่พบมีการใช้โปรแกรมสำเร็จรูปเพื่อการสังเคราะห์ข้อมูล เพื่อนำไปใช้ในการจำลองสถานการณ์กันบ้าง แต่การใช้ยังจำกัด และผู้ใช้ส่วนใหญ่ยังขาดความรู้ความเข้าใจหลักทฤษฎีทางสโตแคสติก โปรแกรมสำเร็จรูปที่นิยมใช้ได้แก่ HEC-4

1.2 กระบวนการสุ่ม (Stochastic Process)

ถ้าให้ X คือตัวแปรสุ่ม (Random Variable) โดยมี $f(x)$ เป็นฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Distribution Function) และ X มีค่าอยู่ระหว่างค่าขีดจำกัดล่างและขีดจำกัดบน หรือ $x_{lower} < x < x_{upper}$.

เซตของค่าสังเกตของ $x_t = \{x_1, x_2, \dots, x_n\}$ เมื่อ t คือมิติของเวลาหรือระยะทางหรืออื่นๆ

ถ้า t คือเวลา

x_t คืออนุกรมเวลา (Time Series)

เซตของค่าสังเกตที่มาจากกระบวนการสุ่มเรียกกว่า Realization of Stochastic Process

กระบวนการสุ่ม คือ เซตของตัวแปรสุ่ม $\{X_t: X_1, X_2, \dots, X_n\}$

$X_1 =$ ตัวแปรสุ่ม ซึ่งมี $f(x_1)$ เป็นฟังก์ชันการแจกแจงความน่าจะเป็น

$X_2 =$ ตัวแปรสุ่ม โดยมี $f(x_2)$ เป็นฟังก์ชันการแจกแจงความน่าจะเป็น

.

$X_n =$ ตัวแปรสุ่ม โดยมี $f(x_n)$ เป็นฟังก์ชันการแจกแจงความน่าจะเป็น

$f(x_1, x_2, \dots, x_n) =$ ฟังก์ชันการแจกแจงความน่าจะเป็นร่วมของหลายตัวแปรสุ่ม (Joint Probability Distribution Function)

แบบจำลองซึ่งอธิบายโครงสร้างความน่าจะเป็นของเซตของค่าสังเกต $\{x_1, x_2, \dots, x_n\}$ เรียกว่าแบบจำลองสุ่ม (Stochastic Model)

1.3 รูปแบบของกระบวนการสุ่ม (Types of Stochastic Process)

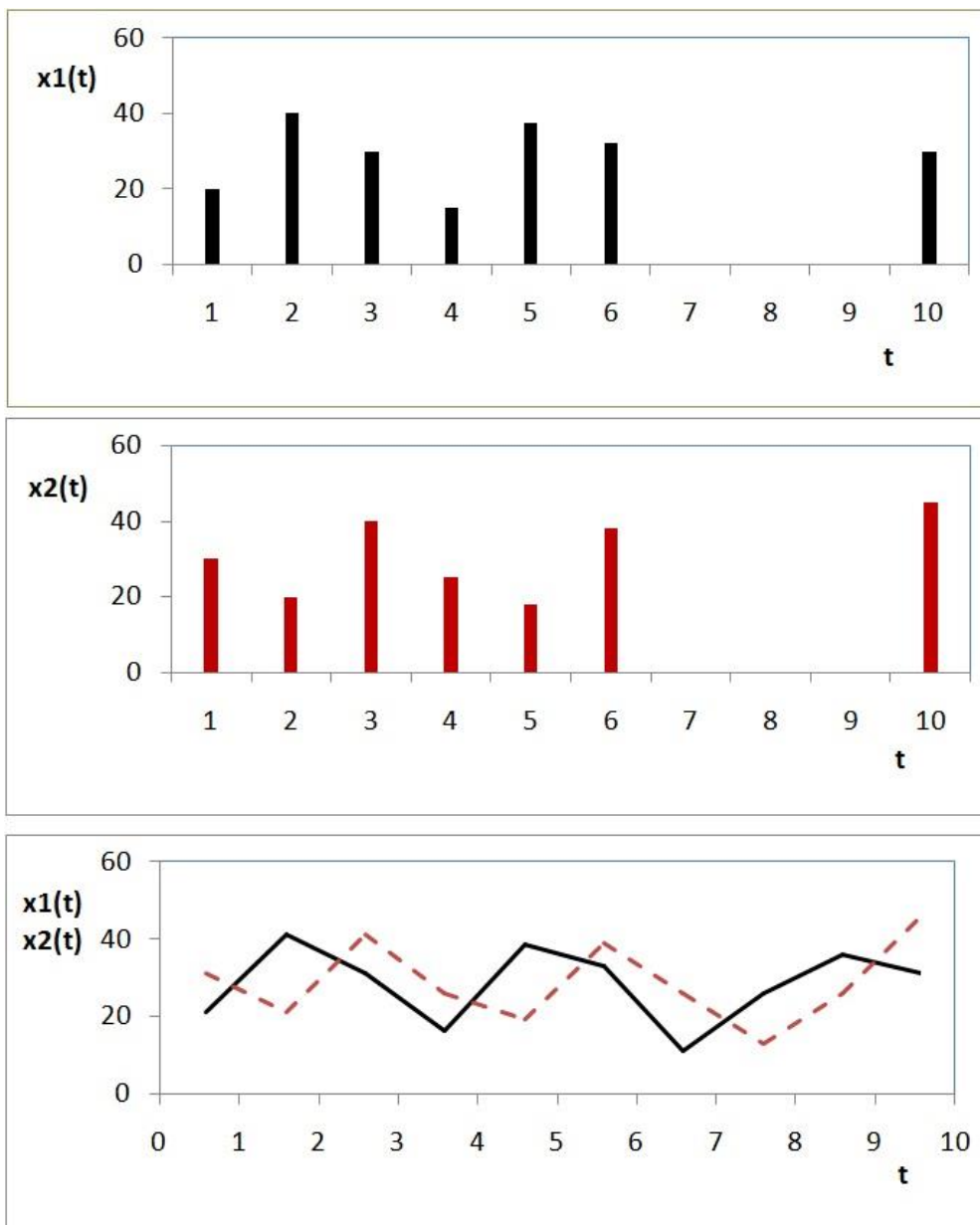
รูปแบบของกระบวนการสุ่มสามารถแบ่งออกได้เป็นหลายลักษณะดังนี้

- Discrete vs. Continuous Process
- Independent vs. Dependent Process
- Stationary vs. Non-stationary Process

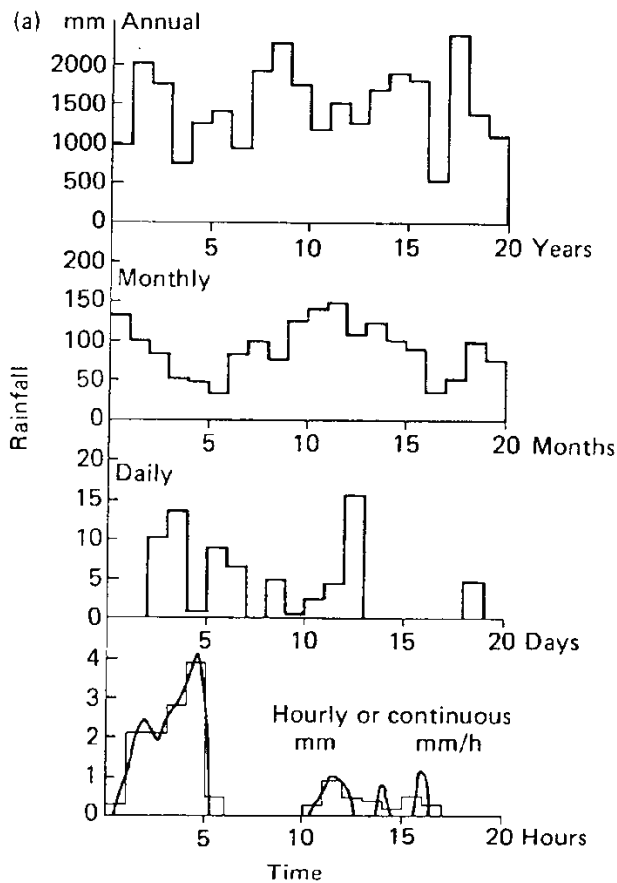
1.3.1 กระบวนการสุ่มแบบ Discrete vs. Continuous

Discrete Process คือกระบวนการสุ่มที่มีค่าไม่ต่อเนื่อง เช่นจำนวนวันที่ฝนตกในแต่ละปี (No. of Rainy Days Per Year) ขณะที่ Continuous Process คือกระบวนการสุ่มที่มีค่าต่อเนื่อง เช่น

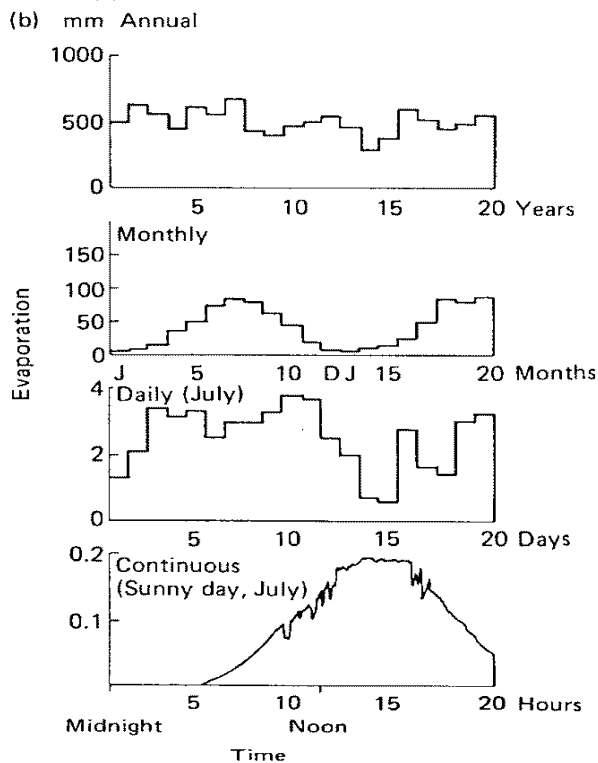
ปริมาณฝนรายปีเป็นต้น โดยทั่วไปข้อมูลอุทกวิทยาจะมีการบันทึกในช่วงเวลาที่ไม่ต่อเนื่อง (Discrete Time) เช่น รายปี รายเดือน รายวัน แต่สามารถนำมาพล็อตเป็นกราฟได้ทั้งแบบต่อเนื่องและไม่ต่อเนื่อง ดังรูปที่ 1.1 และปริมาณฝน น้ำท่า การระเหยที่มีการบันทึกในช่วงเวลาไม่ว่าจะเป็นรายปี รายเดือน รายวันหรือรายชั่วโมง ดังรูปที่ 1.2 ถือว่าเป็นข้อมูลที่มาจากระบวนการสโตแคสติกแบบต่อเนื่อง (Continuous Process)



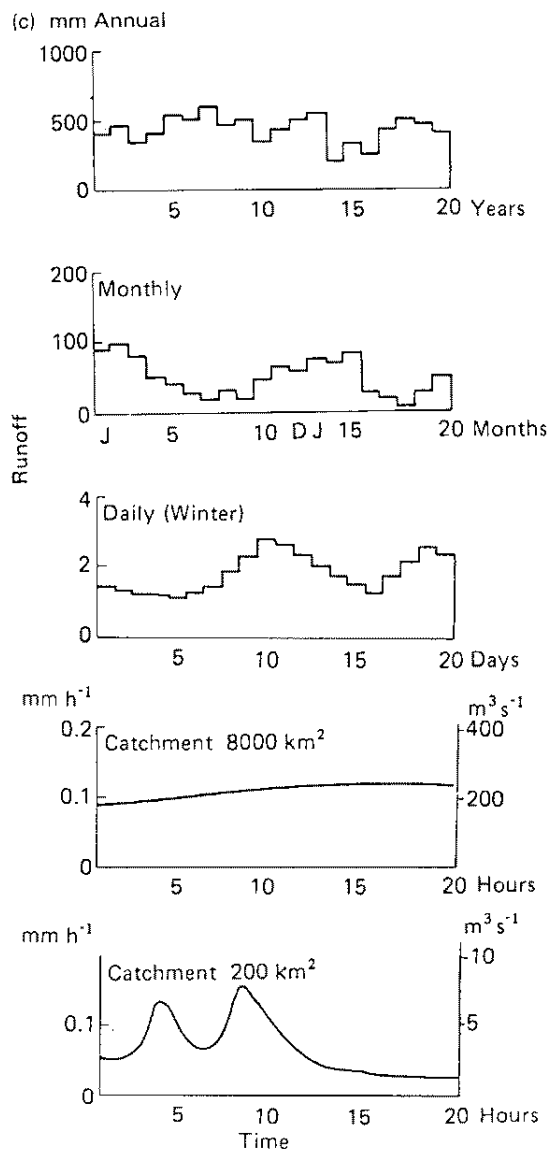
รูปที่ 1.1 Two realizations of stochastic processes
(a) plotted at discrete points in time (b) plotted as continuous lines.



(a) Rainfall Time Series



(b) Evaporation Time Series



(c) Runoff Time Series

รูปที่ 1.2 Different Time Series of Rainfall, Evaporation and Runoff

1.3.2 Independent (Purely Random) vs. Dependent Process

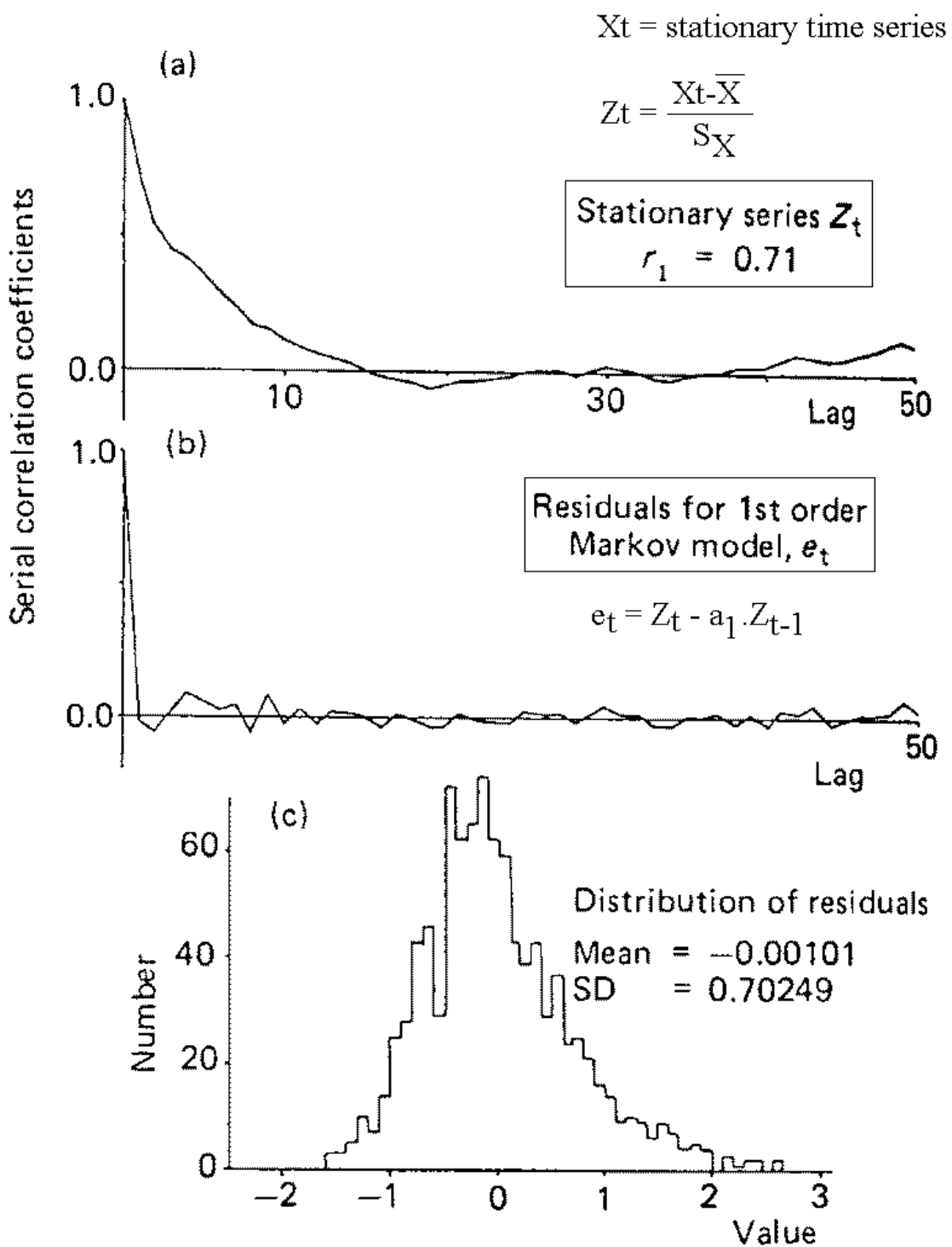
กระบวนการสโตแคสติกแบบอิสระ (Independent) และแบบไม่อิสระ (Dependent) จะพิจารณาได้จากคุณสมบัติของฟังก์ชันการแจกแจงความน่าจะเป็นร่วมของหลายตัวแปรสุ่ม (Joint Probability Distribution Function) ดังนี้

ถ้า $f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n)$ หรือฟังก์ชันการแจกแจงความน่าจะเป็นร่วมของหลายตัวแปรสุ่ม เท่ากับผลคูณของ Marginal Distribution แสดงว่าเป็น Independent Process

ถ้า $f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_1x_2) \dots f(x_n|x_1x_2 \dots x_{n-1})$ หรือฟังก์ชันการแจกแจงความน่าจะเป็นร่วมของหลายตัวแปรสุ่ม เท่ากับผลคูณของ Conditional Distribution แสดงว่าเป็น Dependent Process ซึ่งกรณีของ Dependent Process สามารถแบ่งออกได้เป็น 2 แบบคือ

- Serially Dependence Stochastic Process (Time Dependence) หรือกระบวนการทางสโตแคสติกที่ขึ้นอยู่กับเวลา
- Spatially Dependence Stochastic Process (Area Dependence) หรือกระบวนการทางสโตแคสติกที่ขึ้นอยู่กับพื้นที่

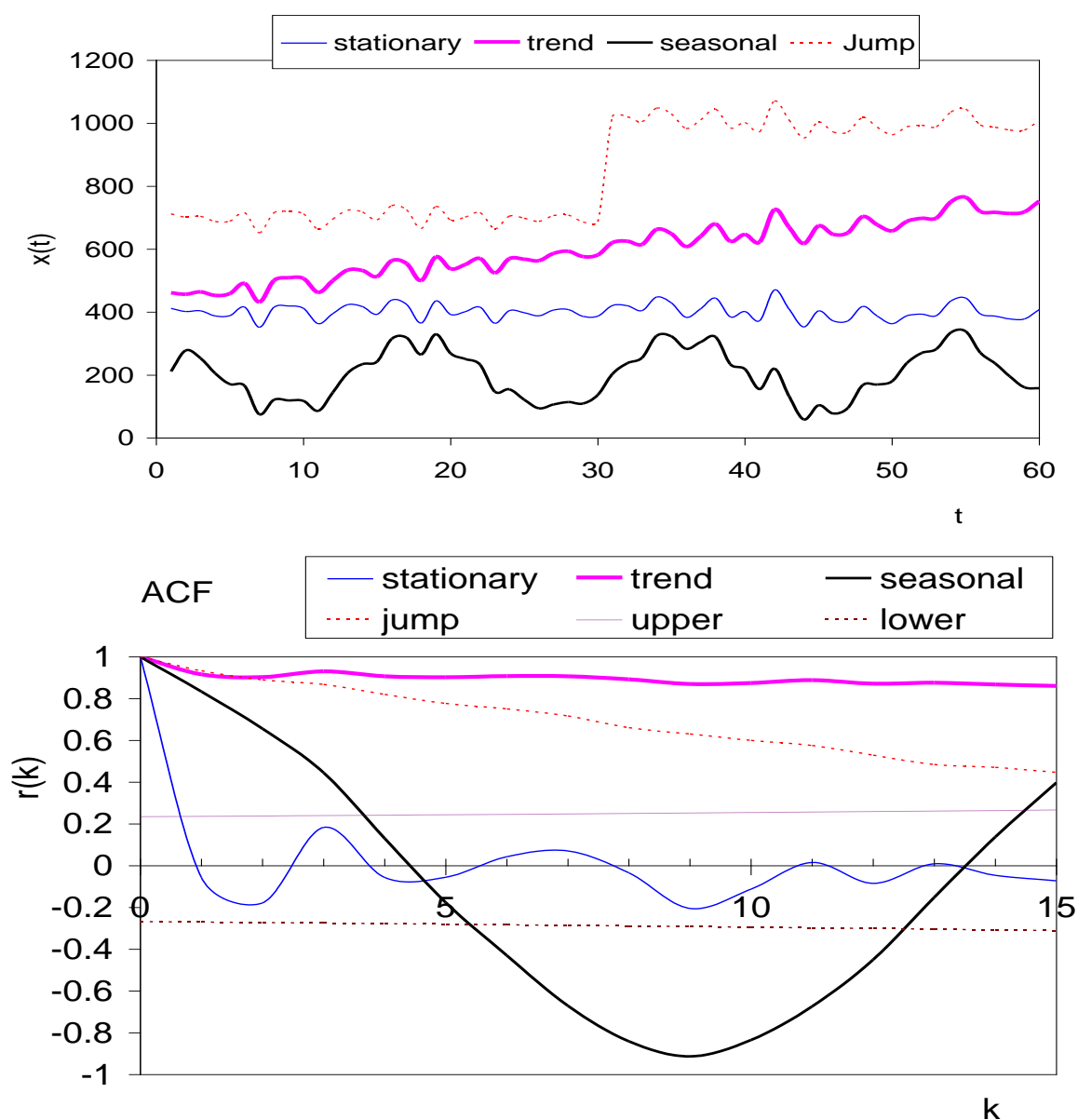
คุณสมบัติทางสถิติที่สามารถแสดงให้เห็นว่าอนุกรมเวลาเป็นแบบอิสระหรือไม่ คือฟังก์ชันสหสัมพันธ์ในตัวเอง หรือ Autocorrelation Function (ACF) ตัวอย่าง ACF ของ Stationary Dependent and Independent Series แสดงอยู่ในรูปที่ 1.3



รูปที่ 1.3 Autocorrelation Function of Dependent Series (z_t) and Independent Series (ϵ_t)

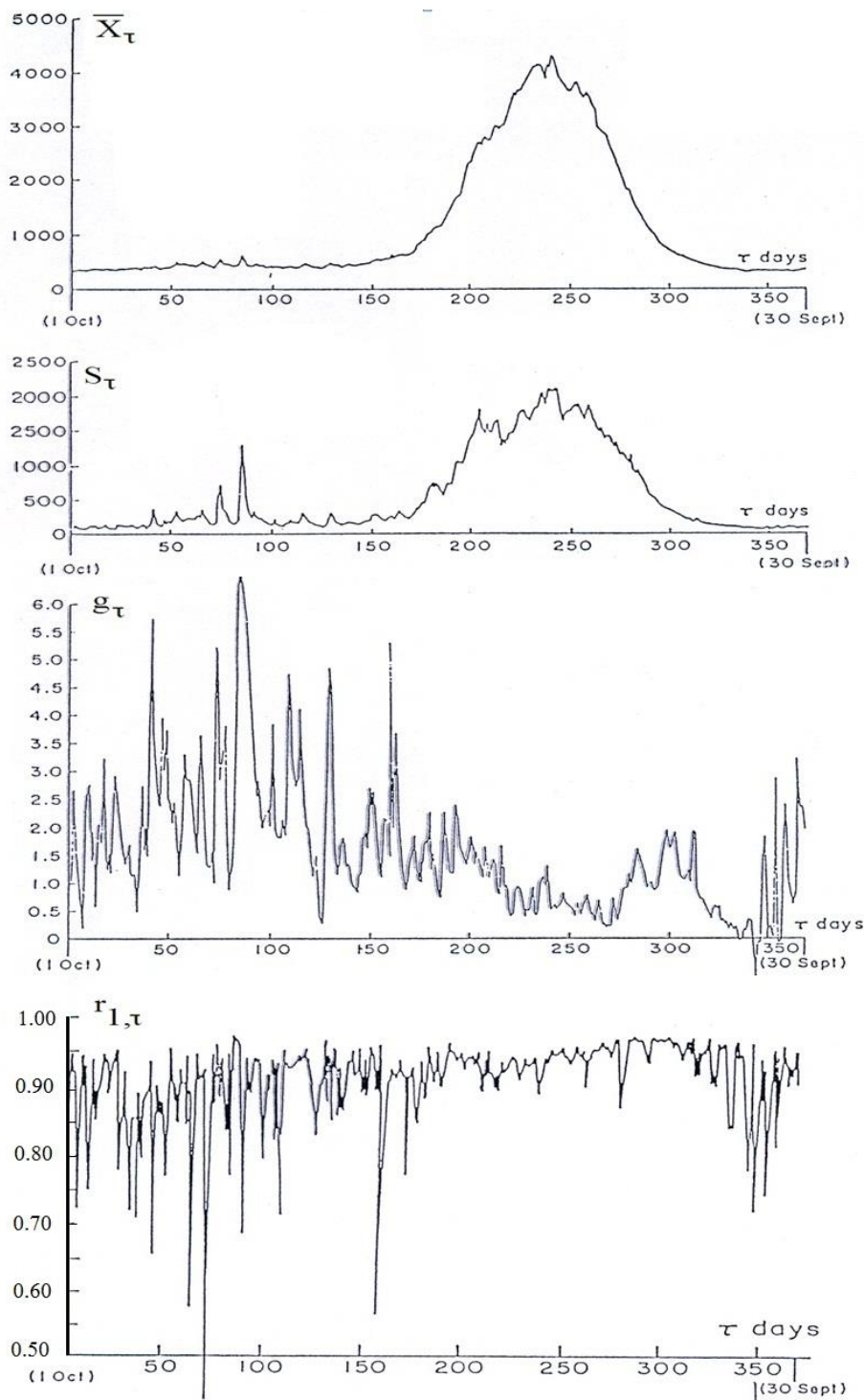
1.3.3 Stationary vs. Non-stationary Process

กระบวนการแบบ Stationary คือกระบวนการสโตแคสติกซึ่งพารามิเตอร์เช่น μ , σ^2 , ρ ไม่เปลี่ยนแปลงไปตามเวลา ในทางกลับกันกระบวนการแบบ Non-stationary คือกระบวนการซึ่งพารามิเตอร์เปลี่ยนไปตามเวลา ปรากฏการณ์เช่น Trend, Jump หรือ Seasonal Variations คือองค์ประกอบของ Non-stationary ที่สามารถเห็นได้ในอนุกรมเวลา ดังรูปที่ 1.4



รูปที่ 1.4 Stationary and Non-stationary Time Series.

ค่าประมาณของพารามิเตอร์ของปริมาณการไหลของน้ำรายวันในแม่น้ำ Boise แสดงให้เห็นว่า ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และ Lag 1 ACF เปลี่ยนแปลงไปตามเวลา หรือเป็น Non-stationary Process ดังรูปที่ 1.5



รูปที่ 1.5 Statistical estimates of daily flows of 40 years long sample of Boise river near Twin Springs, Idaho, USA.

1.4 องค์ประกอบของอนุกรมเวลา (Components of Time Series)

อนุกรมเวลาประกอบด้วยองค์ประกอบที่สำคัญ 4 องค์ประกอบคือ Trend, Periodic, Catastrophic และ Random ดังรูปที่ 1.6 และสามารถเขียนเป็นสมการอนุกรมเวลาได้ดังสมการที่ 1.1

$$Y(t) = Y_1(t) + Y_2(t) + Y_3(t) + Y_4(t) \quad [1.1]$$

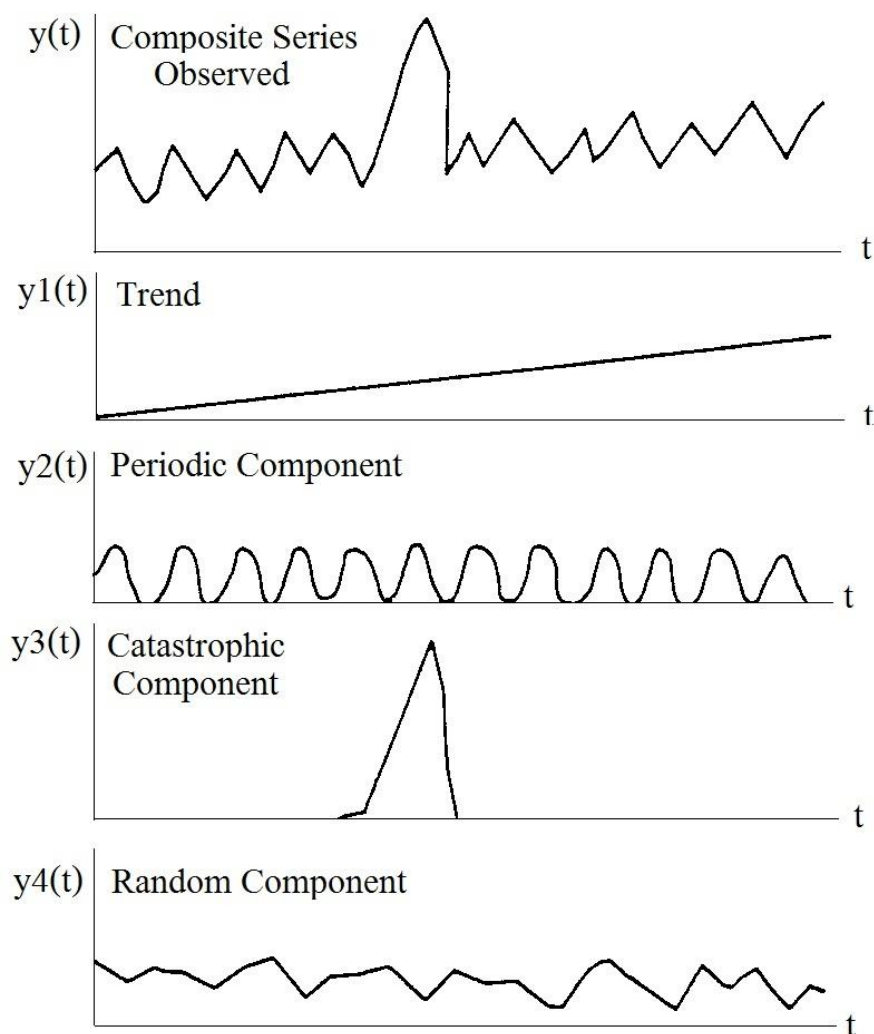
เมื่อ $Y(t)$ = อนุกรมเวลา

$Y_1(t)$ = Trend

$Y_2(t)$ = Periodic Component

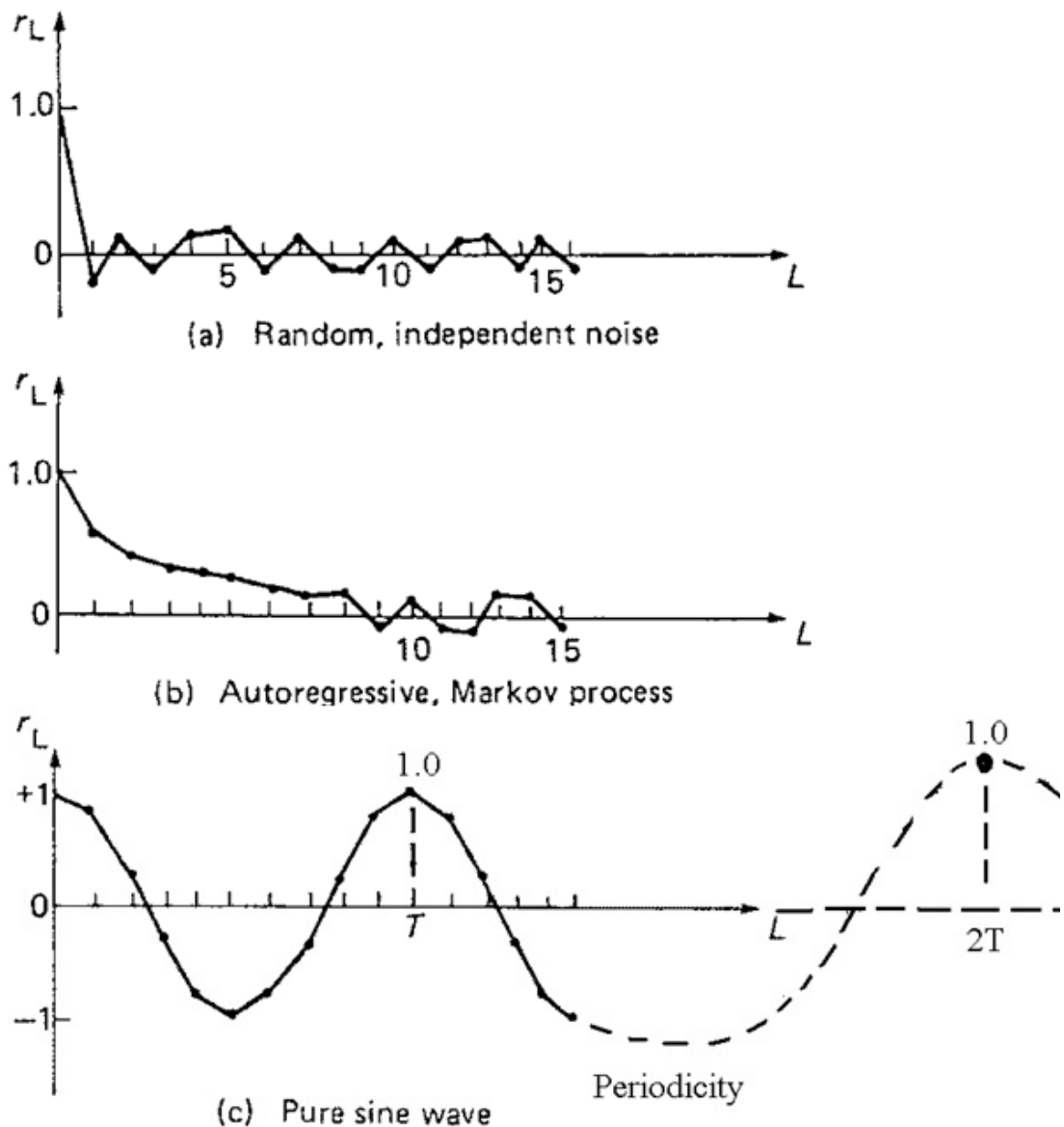
$Y_3(t)$ = Catastrophic Component

$Y_4(t)$ = Random Component

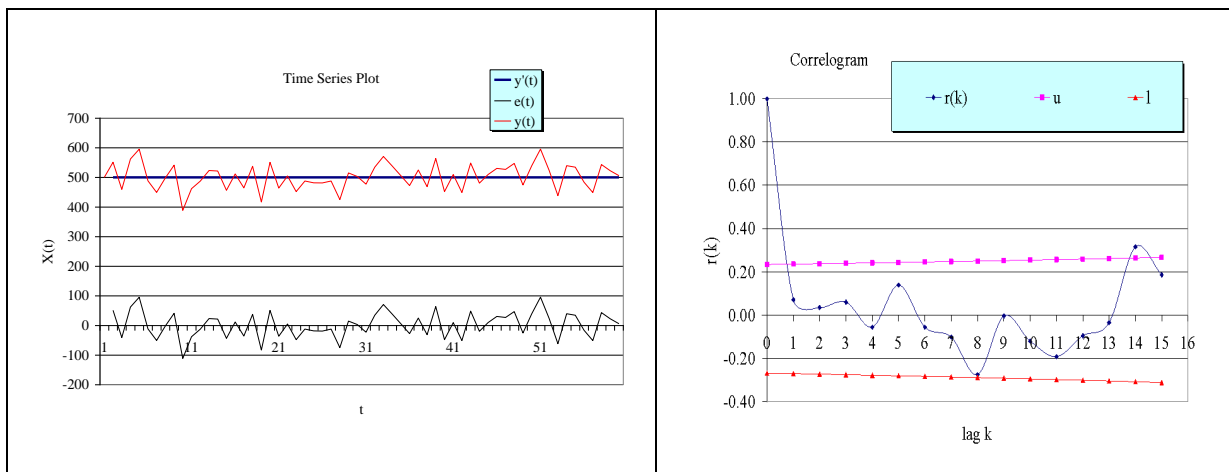


รูปที่ 1.6 Components of Time Series

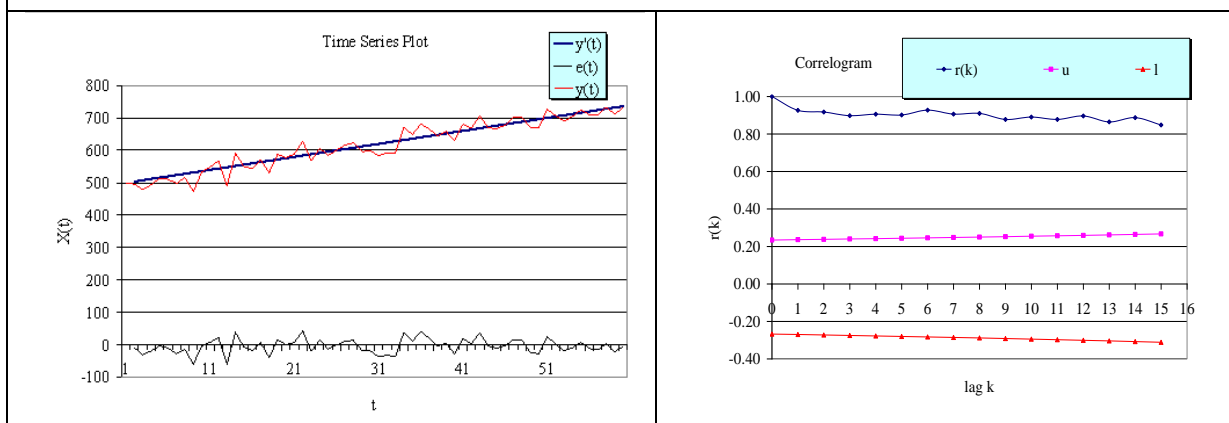
Autocorrelation Function (ACF) หรือเรียกอีกอย่างหนึ่งคือ Correlogram ของกระบวนการสโตแคสติกแบบต่างๆแสดงอยู่ในรูปที่ 1.7 ส่วนรูปที่ 1.8 แสดงการพล็อตเปรียบเทียบระหว่างอนุกรมเวลาและ ACF ของอนุกรมเวลาแบบต่างๆ ACF คือคุณสมบัติที่สำคัญของ Periodic Time Series เมื่อวิเคราะห์อนุกรมเวลาใน Time Domain ส่วน Spectral Density Function คือคุณสมบัติที่สำคัญของ Periodic Time Series เมื่อวิเคราะห์อนุกรมเวลาใน Frequency Domain รูปที่ 1.9 แสดง ACF และ Spectral Density Function ของ Periodic Time Series ที่มี การ Remove Periodic Components ที่มี Frequency ต่างๆ



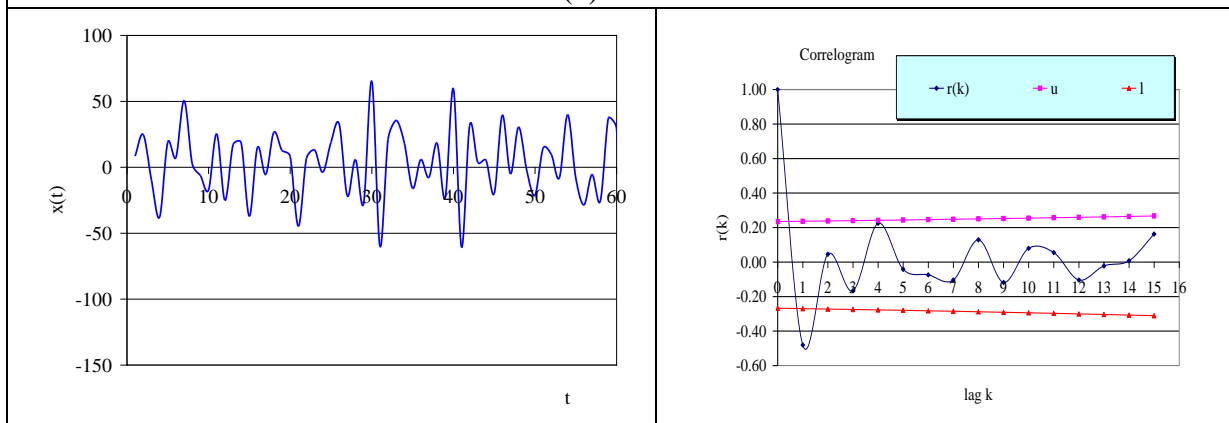
รูปที่ 1.7 Correlogram or ACF Showing Independent, Markov Process and Periodicity



(a) Stationarity

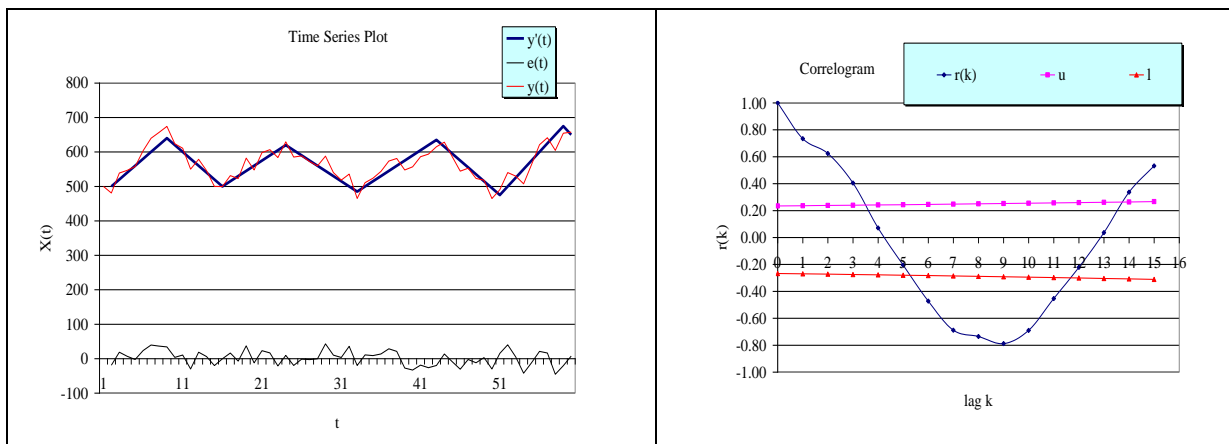


(b) Trend

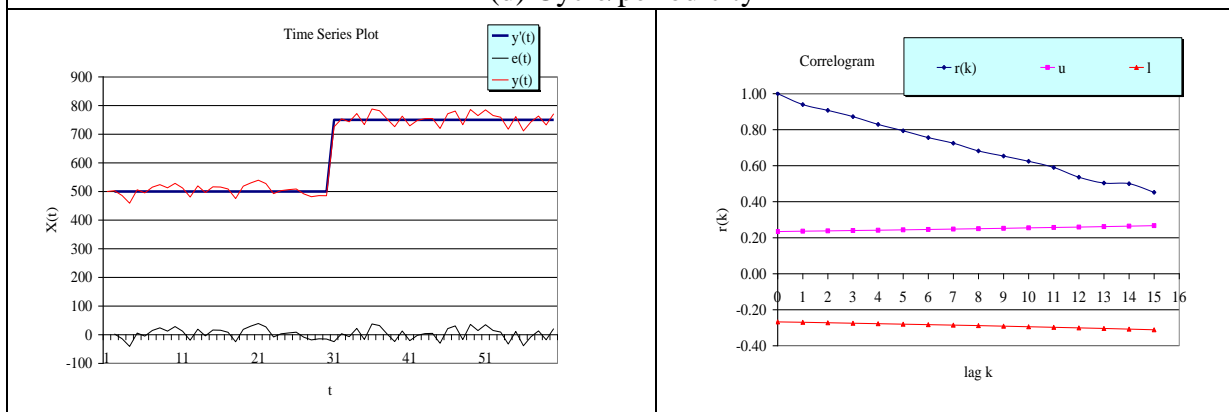


(c) After Removing Trend From (b) by Differencing

รูปที่ 1.8 Characteristics of Time Series and Correlogram for Different Types of Time Series



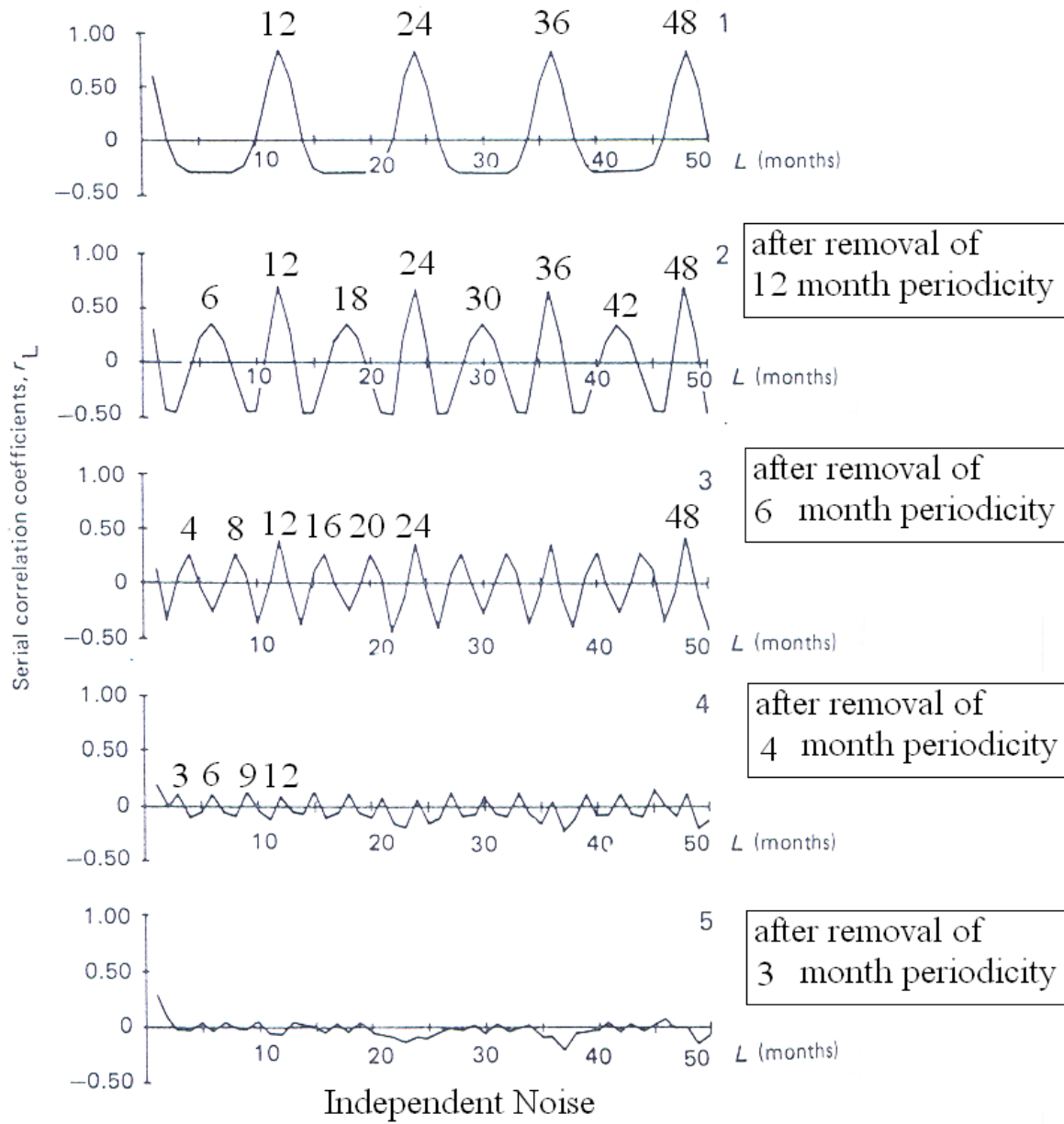
(d) Cycle/periodicity



(d) Jump

รูปที่ 1.8(ต่อ) Characteristics of Time Series and Correlogram for Different Types of Time Series

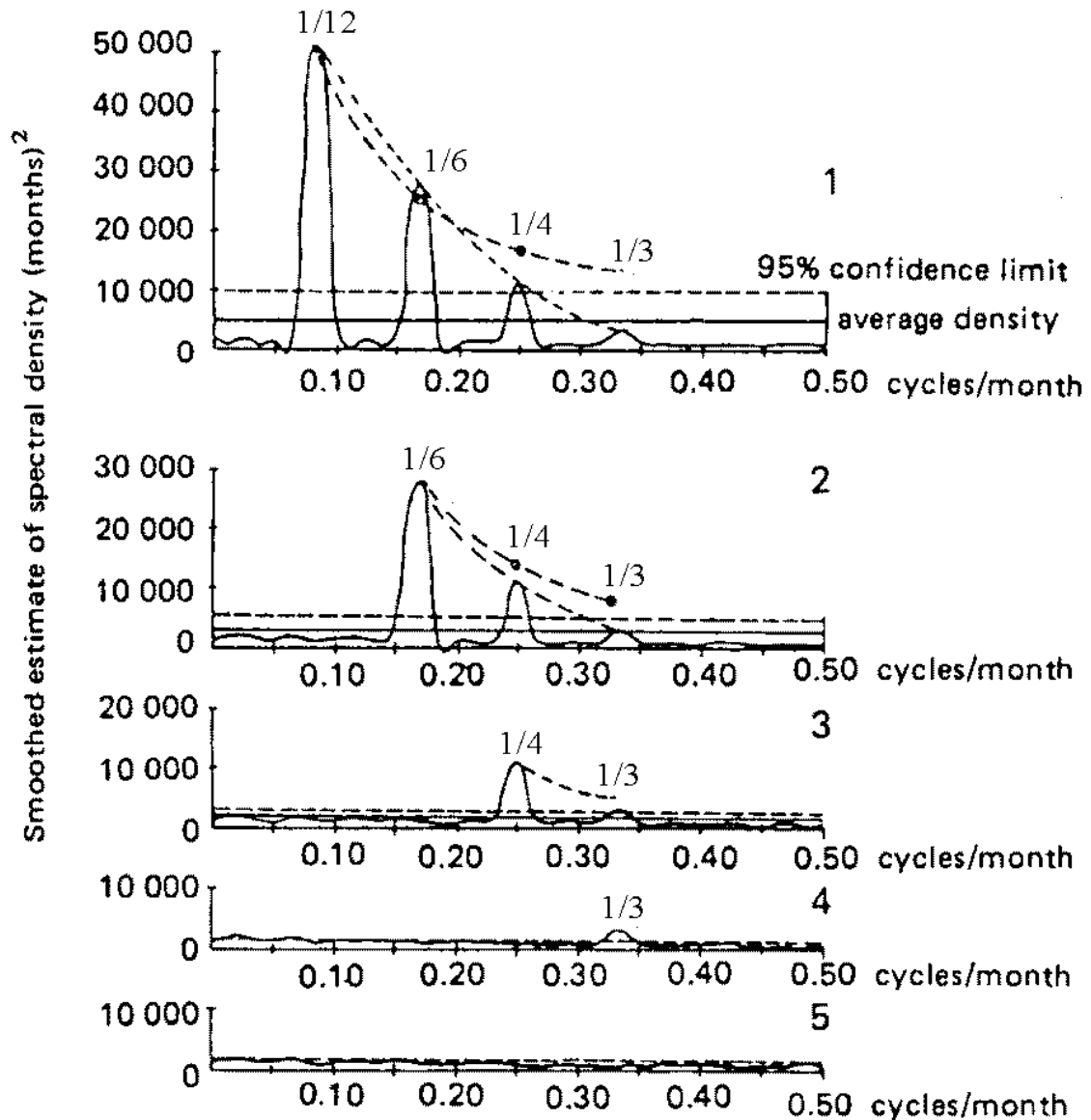
(a) Correlogram



(a) ACF After Removing Periodicity (Time domain)

รูปที่ 1.9 ACF and Spectral Density Function after Removing Periodicity

(b) Variance spectrum



(b) Spectral Density Function after Removing Periodicity (Frequency domain)

รูปที่ 1.9(ต่อ) ACF and Spectral Density Function after Removing Periodicity

1.5 คุณสมบัติพื้นฐานของกระบวนการสโตแคสติก (Basic Properties of Stochastic Process)

ให้ $\{X_t\}$ = เซตของตัวแปรสุ่มหรือ Stochastic Process

$$E(X_t) = \mu_x : t = 1, 2, \dots, N \tag{1.2}$$

$$VAR(X_t) = \sigma_x^2 : t = 1, 2, \dots, N \tag{1.3}$$

$$COV(X_t, X_{t-k}) = \sigma_{x_t, x_{t-k}}; t = 1, 2, \dots, N \quad [1.4]$$

เมื่อ $\sigma_{x_t, x_{t-k}}$ คือ Lag k Autocovariance

$$\rho_k = \frac{COV(X_t, X_{t-k})}{\sqrt{VAR(X_t)VAR(X_{t-k})}} \quad [1.5]$$

เมื่อ ρ_k คือ Lag k ACF ดังสมการ (2.5a, 2.5b) of Salas et al.(1980)

ACF หรือ Correlogram หรือ Serial Correlation Coefficients คือพารามิเตอร์ที่ใช้วัดความสัมพันธ์เชิงเส้น (Linear Dependence) ของกระบวนการสโตแคสติก

ตามที่ได้กล่าวมาแล้ว กระบวนการสโตแคสติกแบบ Stationary คือกระบวนการที่พารามิเตอร์ μ, σ^2, ρ ไม่เปลี่ยนแปลงไปตามเวลา ซึ่งสามารถแบ่งออกได้เป็น 2 แบบตามคุณสมบัติของพารามิเตอร์ คือ First Order Stationary และ Second Order Stationary ดังตารางที่ 1.1

ตารางที่ 1.1 Different Types of Stationary Processes

First order stationary process	$\mu_x = \mu = constant$
Second order stationary process	$\mu_x = \mu = constant$ and $\sigma_x^2 = \sigma^2 = constant$
- Weak stationary (weak sense)second order stationary
- Strong stationary (strict sense)second order stationary + other parameters such as correlation coefficient are not varied with time.

กระบวนการสโตแคสติกอาจเป็น Stationary เมื่อพิจารณาจากพารามิเตอร์ตัวหนึ่ง แต่อาจเป็น Non-stationary เมื่อพิจารณาจากพารามิเตอร์ตัวอื่น

สรุป กระบวนการสโตแคสติกสามารถแบ่งออกได้เป็นประเภทต่างๆดังตารางที่ 1.2

ตารางที่ 1.2 Classification of Stochastic Processes

Deterministic	Vs.	Stochastic
Discrete	Vs.	Continuous
Independent	Vs.	Dependent
Stationary	Vs.	Non-stationary

1.6 แบบจำลองอนุกรมเวลาแบบง่าย (Simple Time Series Model)

ให้ $x_t =$ Independent Time Series with $f(x;\theta)$

$$x_t = \mu + \sigma \varepsilon_t ; \quad t = 1, 2, \dots \dots \dots \quad [1.6]$$

$\varepsilon_t = \frac{x_t - \mu}{\sigma}$ เป็น Independent Series ซึ่งมี Mean = 0 and Standard Deviation = 1

$\mu, \sigma =$ Constant Parameters (do not vary with time)

แบบจำลองจะเป็น Stationary ถ้าพารามิเตอร์มีค่าคงที่

$x_t =$ Dependent Time Series

$$\varepsilon_t = \frac{x_t - \mu}{\sigma}$$

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} - \xi_t \quad [1.7]$$

$\varepsilon_t =$ Dependent Series with Mean = 0 ; Standard Deviation = 1

$\xi_t =$ Independent Series with Mean = 0 and Variance = $(1 - \phi^2)$

สมการ 1.7 คือรูปแบบจำลองสโตแคสติกแบบ Autoregressive ซึ่งเป็นแบบจำลองแบบง่าย ดู

Yevjevich (1963), Thomas (1965), Fiering (1967)

แบบจำลองสโตแคสติกที่ซับซ้อนมากขึ้นคือ Autoregressive and Moving Average (ARMA) ดู O' Connor (1976)

Joint Physical and Statistical Analysis ของแบบจำลอง

AR(1)-First Order Autoregressive Model

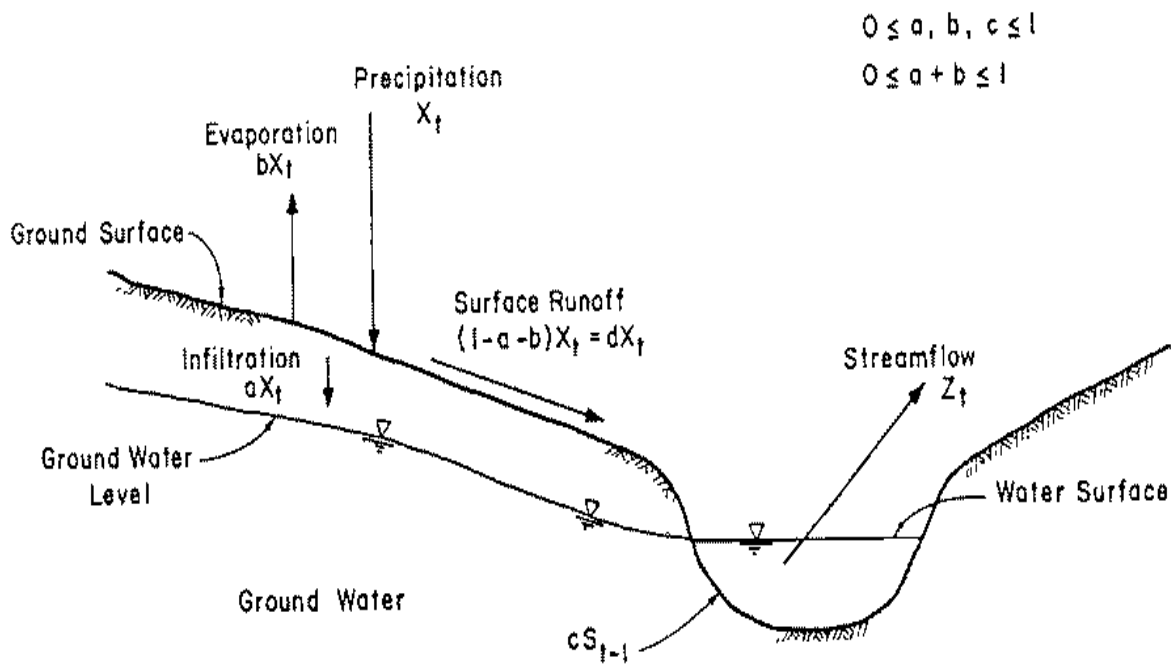
ให้ $Y_t =$ Time Dependent Annual Runoff Series

$$Y_t - \bar{Y} = \phi_1 (Y_t - \bar{Y}) + \varepsilon_t$$

$\phi_1 =$ Autoregressive Coefficient.

$\varepsilon_t =$ Independent Statistical Component or White Noise.

Fiering (1968) ได้พิสูจน์ให้เห็นว่าอัตราการไหลของน้ำในทางน้ำสามารถอธิบายได้ด้วยแบบจำลอง ARMA(1,1) ดังรูปที่ 1.10



รูปที่ 1.10 Conceptual Representation of the Precipitation-Stream Flow Process
(Salas and Smith. 1980a)

แบบจำลอง Streamflow

$$Z_t = cS_{t-1} + dX_t \quad [1.8]$$

เมื่อ

$Z_t = \text{Annual Stream Flow}$

$c.S_{t-1} = \text{Groundwater Contribution}$

$d.X_t = \text{Surface Runoff}$

จากสมการ Continuity Equation สำหรับ Groundwater Storage

$$S_t = S_{t-1} + aX_t - cS_{t-1} \quad [1.9]$$

$$S_t = (1 - c)S_{t-1} + aX_t \quad [1.10]$$

จากสมการ [1.8] และ [1.10] จะได้

$$Z_t = (1 - c)Z_t + dX_t - [d(1 - c) - ac]X_{t-1} \quad [1.11]$$

เมื่อ ปริมาณฝนรายปี (X_t) คือ Independent Series

แบบจำลอง ARMA (1,1)

$$Z_t = \phi_1 Z_{t-1} - \theta_0 \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$\theta_0 = -1$
 Independent (Moving Average)
 Dependent (Autoregressive)

พิสูจน์

จากสมการที่ [1.8]

$$Z_t = cS_{t-1} + dX_t \quad [1.12]$$

$$Z_{t-1} = cS_{t-2} + dX_{t-1} \quad [1.13]$$

$$[1.13] \times (1-c)$$

$$(1-c)Z_{t-1} = c(1-c)S_{t-2} + d(1-c)X_{t-1} \quad [1.14]$$

$$[1.12] - [1.14]$$

$$Z_t - (1-c)Z_{t-1} = cS_{t-1} - c(1-c)S_{t-2} + dX_t - d(1-c)X_{t-1} \quad [1.15]$$

$$Z_t = (1-c)Z_{t-1} + \{cS_{t-1} - c(1-c)S_{t-2}\} + \{dX_t - d(1-c)X_{t-1}\} \quad [1.16]$$

จากสมการที่ [1.10]

$$S_t = (1-c)S_{t-1} + aX_t$$

$$S_{t-1} = (1-c)S_{t-2} + aX_{t-1}$$

$$cS_{t-1} = c(1-c)S_{t-2} + acX_{t-1}$$

$$cS_{t-1} - c(1-c)S_{t-2} = acX_{t-1} \quad [1.17]$$

แทนค่าสมการ [1.17] ใน [1.16]

$$Z_t = (1-c)Z_{t-1} + acX_{t-1} + \{dX_t - d(1-c)X_{t-1}\}$$

$$Z_t = (1-c)Z_{t-1} + dX_t + \{ac - d(1-c)\}X_{t-1} \quad [1.18]$$

$$Z_t = \phi_1 Z_{t-1} - \theta_0 \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad [1.19]$$

สมการ [1.18] เป็น ARMA (1,1) เมื่อ $\theta_0 = -1$

แบบจำลองสโตแคสติกคือแบบจำลองที่สามารถสร้างอนุกรมเวลาที่มีคุณสมบัติทางสถิติคล้ายคุณสมบัติทางสถิติของอนุกรมเวลาในอดีต แต่ไม่ได้หมายความว่าต้องสร้างอนุกรมเวลาที่เหมือนอนุกรมเวลาในอดีตทุกประการ เช่นการโคลน (Cloning)

1.7 ค่าสถิติหลักของแบบจำลองสโตแคสติกแบบง่าย (Main Statistics of Simple Stochastic Model)

- (1) Mean (μ).....less uncertain
- (2) Standard Deviation (σ).....less uncertain
- (3) Skewness Coefficient (γ).....highly uncertain
- (4) Autocorrelation Coefficient (ρ_k).....very uncertain for small N
- (5) Other statistics of hydrologic time series
 - range (storage capacity)
 - run (drought)
 - rescaled range

1.8 แบบจำลองสโตแคสติก (Stochastic Models)

ในเอกสารชุดนี้จะกล่าวถึงแบบจำลองของอนุกรมเวลาทั้ง Annual Time Series และ Periodic Time Series ดังแสดงในตารางที่ 1.3

ตารางที่ 1.3 Important Stochastic Models to be included in this Handouts

(1) AR(p)	Chapter 4
(2) ARMA(p,q)	Chapter 5
(3) ARIMA(p,d,q), Seasonal ARIMA(P,D,Q) _w , Multiplicative ARIMA(p,d,q)x(P,D,Q) _w	Chapter 6
(4) Multivariate AR(p)	Chapter 7
(5) Disaggregate	Chapter 8

ข้อจำกัดของแบบจำลอง (Limitations of Model)

- (1) ไม่สามารถ Reproduce Short Term Dependence
- (2) ไม่สามารถ Reproduce Long Term Dependence
- (3) การประมาณค่าพารามิเตอร์มีความยุ่งยาก
- (4) มีข้อจำกัดในการ Generate ข้อมูลใหม่จำนวนมากๆ
- (5) ขาดพื้นฐานทางด้านกายภาพ (Physical Basis)

(6) มีพารามิเตอร์มากเกินไป

1.9 Systematic Approach สำหรับการจำลองอนุกรมเวลาทางอุทกวิทยา

Systematic Approach สำหรับการจำลองอนุกรมเวลาทางอุทกวิทยาประกอบด้วย 6 ขั้นตอนดังนี้ (ดูรูปที่ 1.11)

(1) การหาส่วนประกอบของแบบจำลอง (Identification of Model Composition)

- a. Univariate หรือ Multivariate Model
- b. Combination of Univariate and Disaggregation
- c. Combination of Multivariate and Disaggregation
- d. ส่วนประกอบของแบบจำลองขึ้นอยู่กับคุณลักษณะของระบบแหล่งน้ำ คุณลักษณะของอนุกรมเวลา และความเห็นของผู้สร้างแบบจำลอง (Modeler's Input)
- e. ยกตัวอย่าง เช่นถ้าต้องการตั้งเคราะห์อนุกรมเวลาของปริมาณน้ำที่ไหลเข้าระบบอ่างเก็บน้ำรายเดือน ทางเลือกในการสร้างแบบจำลองอาจเป็น (1) จำลองปริมาณน้ำที่ไหลเข้าระบบอ่างเก็บน้ำรายเดือนด้วย Multivariate Model หรือ (2) จำลองปริมาณน้ำที่ไหลเข้าระบบอ่างเก็บน้ำรายปีด้วย Multivariate Model แล้วใช้ Disaggregation Model แปลงปริมาณน้ำที่ไหลเข้าระบบอ่างเก็บน้ำรายปีเป็นรายเดือน หรือ (3) ใช้ Univariate Model จำลองปริมาณน้ำที่ไหลเข้าแต่ละอ่างเก็บน้ำรายเดือนแบบเป็นอิสระแก่กัน กรณีนี้จะใช้เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอ่างเก็บน้ำ (Cross Correlation Coefficient) ไม่มีนัยสำคัญทางสถิติ

(2) การเลือกประเภทแบบจำลอง (Selection of Model Type)

- a. AR
- b. ARMA - เหมาะกับอนุกรมเวลาซึ่ง ACF ที่ค่าลดลงอย่างช้าๆ (Low Decaying Correlation)
- c. ARIMA
- d. อื่นๆ

(3) การหารูปลักษณะของแบบจำลอง (Identification of Model Form)

- a. หาลำดับ (Orders) ของแบบจำลอง เช่นค่า p และ q
- b. ตรวจสอบว่าอนุกรมเวลามีลักษณะเบ้ (Skewed)หรือไม่ ถ้าเบ้ ต้องตรวจสอบต่อไปว่าอนุกรมเวลาเป็นแบบ Constant หรือ Periodic ถ้าเป็นแบบ Periodic ต้องดูว่าควรใช้ Fourier Series วิเคราะห์หรือไม่

(4) การประมาณค่าพารามิเตอร์แบบจำลอง (Estimation of Model Parameters)

- a. Method of moments

b. Method of maximum likelihood

(5) การทดสอบความเหมาะสมของแบบจำลอง (Testing Goodness of Fit of the Model)

a. ตรวจสอบ (Verify) ว่าแบบจำลองเป็นไปตามสมมติฐานที่กำหนดหรือไม่ เช่น Residuals (ζ_t)

เป็นอิสระและมีการแจกแจงแบบปกติ (Independence and Normality)

b. ตรวจสอบ (Verify) ว่าแบบจำลองเป็นตัวแทนที่เหมาะสมของอนุกรมเวลาของค่าสังเกต

(Historical Time Series)

- ตรวจสอบ Model Correlogram เปรียบเทียบกับ Historical Correlogram

- ตรวจสอบค่าสถิติของอนุกรมเวลาที่สังเคราะห์ขึ้นแล้วเปรียบเทียบกับค่าสถิติของอนุกรมเวลาของค่าสังเกต เช่น ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ ค่าสหสัมพันธ์ ค่าสถิติที่เกี่ยวข้องกับปริมาณน้ำในอ่างเก็บน้ำ (Storage Related Statistics) และค่าสถิติที่เกี่ยวข้องกับความแห้งแล้ง (Drought Related Statistics)

c. ถ้าการตรวจสอบความเหมาะสมไม่ผ่านเกณฑ์ในข้อ a และ b ต้องปรับเปลี่ยนรูปลักษณะและชนิดของแบบจำลอง

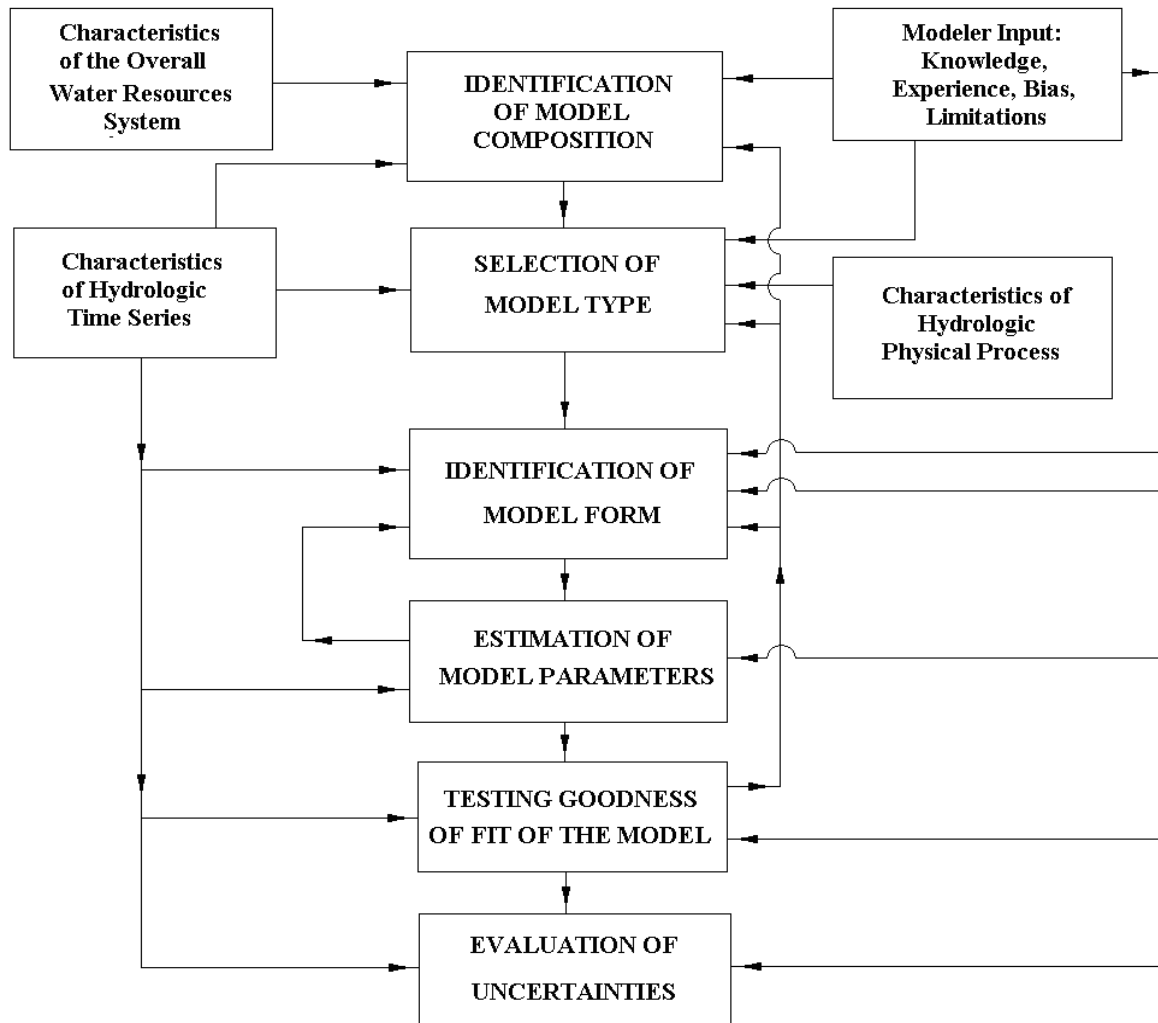
(6) การประเมินความไม่แน่นอน (Evaluation of Uncertainties)

a. ความไม่แน่นอนของแบบจำลอง (Model Uncertainty)

- ทดสอบความแตกต่างระหว่างค่าสถิติที่คำนวณจากอนุกรมเวลาที่สังเคราะห์ขึ้นจากแบบจำลองทางเลือกรูปแบบต่างๆ

b. ความไม่แน่นอนของแบบพารามิเตอร์ (Parameter Uncertainty)

- หากการแจกแจงความน่าจะเป็นของพารามิเตอร์ และใช้แบบจำลองกับพารามิเตอร์ที่สุ่มจากการแจกแจงนั้น



รูปที่ 1.11 Systematic Approach for Time Series Modeling

1.10 การประยุกต์การจำลองอนุกรมเวลาทางอุทกวิทยา (Application of Time Series Modeling in Hydrology)

(1) การสังเคราะห์อนุกรมเวลาใหม่ (Generation of Synthetic Time Series) สำหรับ

- Reservoir Sizing
- Risk of Failure
- Planning Capacity Expansion

(2) การพยากรณ์ (Forecasting)

- Short Term Planning of Reservoir Operation.
- Real Time Operation
- Operation of River Basin

- Planning the Operation During Ongoing Drought.

1.11 สรุปแบบจำลองอนุกรมเวลาที่จะกล่าวถึงในบทต่างๆ

บทที่ 4 แบบจำลอง Autoregressive

Annual AR(p)

$$z_t = \sum_{j=1}^p \phi_j z_{t-j} + \varepsilon_t$$

$$\varepsilon_t = \sigma_\varepsilon \xi_t$$

เมื่อ

z_t = Time Dependent Series with Normal Distribution and $E(z_t) = 0$; $\text{Var}(z_t) = 1$

ε_t = Independent Normal Variable with Mean = 0 and Variance = σ_ε^2

ϕ_1, \dots, ϕ_p = Autoregressive Coefficients

Periodic AR(p)

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} + \sigma_{\varepsilon,\tau} \xi_{v,\tau}; \tau = 1, \dots, \omega$$

$z_{v,\tau}$ = Time Dependent Normal Variables with Mean = 0 and Variance = 1

$\sigma_{\varepsilon,\tau}$ = Independent Normal Variable with Mean = 0 and Variance = $\sigma_{\varepsilon,\tau}^2$

$\phi_{1,\tau}, \dots, \phi_{p,\tau}$ = Periodic Autoregressive Coefficients

บทที่ 5 แบบจำลอง Autoregressive and Moving Average

Annual ARMA(p,q)

$$z_t = \sum_{j=1}^p \phi_j z_{t-j} - \sum_{j=0}^q \theta_j \varepsilon_{t-j}; \theta_0 = -1$$

Periodic ARMA(p,q)

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} + \varepsilon_{v,\tau} - \sum_{j=1}^q \theta_{j,\tau} \varepsilon_{v,\tau-j}; \theta_0 = -1, \tau = 1, \dots, \omega$$

บทที่ 6 แบบจำลอง Autoregressive Integrated Moving Average

Condensed Form ของ Multiplicative ARIMA(p,d,q)x(P,D,Q)_w

$$\Phi(B^\omega)\phi(B)(1-B^\omega)^D(1-B)^d x_t = \Theta(B^\omega)\theta(B)\varepsilon_t$$

$Bx_t = x_{t-1}$: Backward Operator

$$B^n x_t = x_{t-n}$$

$(1-B)x_t = x_t - x_{t-1}$; 1st Order Simple Differencing

$(1-B)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$; 2st Order Simple Differencing

$(1-B)^d x_t = d^{\text{th}}$ Order Simple Difference

$(1-B^\omega)x_t = x_t - x_{t-\omega}$; 1st Order Seasonal Differencing

$(1-B^\omega)^2 x_t = x_t - 2x_{t-\omega} + x_{t-2\omega}$; 2st Order Seasonal Differencing

$(1-B^\omega)^D x_t = D^{\text{th}}$ Order Seasonal Differencing

$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p)$ = Autoregressive Operator

$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q)$ = Moving average Operator

$\Phi(B^\omega) = (1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} \dots - \Phi_p B^{p\omega})$ = Seasonal Autoregressive Operator

$\Theta(B^\omega) = (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega})$ = Seasonal Moving Average Operator

Simple Non-periodic ARIMA(p,d,q)

$$\phi(B)(1-B)^d x_t = \theta(B)\varepsilon_t$$

Non-periodic ARIMA(1,1,1)

$$(1 - \phi_1 B)(1 - B)x_t = (1 - \theta_1 B)\varepsilon_t$$

$$(1 - \phi_1 B - B + \phi_1 B^2)x_t = (1 - \theta_1 B)\varepsilon_t$$

$$x_t = (1 + \phi_1)x_{t-1} - \phi_1 x_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Periodic ARMA(P,D,Q)_w

$$\Phi(B^\omega)(1-B^\omega)^D x_t = \Theta(B^\omega)\varepsilon_t$$

Periodic ARIMA(1,1,1)

$$(1 - \Phi_1 B^\omega)(1 - B^\omega)x_t = (1 - \Theta_1 B^\omega)\varepsilon_t$$

$$(1 - \Phi_1 B^\omega - B^\omega + \Phi_1 B^{2\omega})x_t = (1 - \Theta_1 B^\omega)\varepsilon_t$$

$$x_t = (1 + \Phi_1)x_{t-\omega} - \Phi_1 x_{t-2\omega} + \varepsilon_t - \Theta_1 \varepsilon_{t-\omega}$$

แบบจำลอง Multiplicative ARIMA(1,1,1)x(1,1,1)₁₂ กรณี $\omega=12$

$$\Phi(B^{12})\phi(B)(1-B^{12})(1-B)x_t = \theta(B^{12})\theta(B)\varepsilon_t$$

$$x_t = (1+\phi_1)x_{t-1} - \phi_1 x_{t-2} + (1+\Phi_1)x_{t-12} - (1+\phi_1 + \Phi_1 + \Phi_1\phi_1)x_{t-13} + (\phi_1 + \Phi_1\phi_1)x_{t-14} -$$

$$\Phi_1 x_{t-14} + (\Phi_1 + \Phi_1\phi_1)x_{t-25} - \Phi_1\phi_1 x_{t-26} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1\varepsilon_{t-12} + \theta_1\theta_1\varepsilon_{t-13}$$

บทที่ 7 แบบจำลองหลายตัวแปร (Multivariate)

Multivariate AR(1)

$$Z_t = A_1 Z_{t-1} + B\varepsilon_t$$

Expanded Form

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \\ \vdots \\ Z_t^n \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & \dots & a^{1n} \\ a^{21} & a^{22} & \dots & a^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a^{n1} & a^{n2} & \dots & a^{nn} \end{bmatrix} \begin{bmatrix} Z_{t-1}^1 \\ Z_{t-1}^2 \\ \vdots \\ Z_{t-1}^n \end{bmatrix} + \begin{bmatrix} b^{11} & b^{12} & \dots & b^{1n} \\ b^{21} & b^{22} & \dots & b^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b^{n1} & b^{n2} & \dots & b^{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \vdots \\ \varepsilon_t^n \end{bmatrix}$$

$$\text{เมื่อ } Z_t = \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ \vdots \\ Z_t^n \end{bmatrix} = \text{Column Matrix ขนาด } [n \times 1]$$

$$A_1 = \begin{bmatrix} a^{11} & a^{12} & \dots & a^{1n} \\ a^{21} & a^{22} & \dots & a^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a^{n1} & a^{n2} & \dots & a^{nn} \end{bmatrix} \dots [n \times n]$$

$$B_1 = \begin{bmatrix} b^{11} & b^{12} & \dots & b^{1n} \\ b^{21} & b^{22} & \dots & b^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b^{n1} & b^{n2} & \dots & b^{nn} \end{bmatrix} \dots [n \times n]$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \vdots \\ \varepsilon_t^n \end{bmatrix} \dots [n \times 1]$$

ε_t = Vector of independent normally distributed random variables with mean = 0 and standard deviation = 1 (uncorrelated in time and space)

Multivariate AR(2)

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B\varepsilon_t$$

Periodic Multivariate AR(1)

$$Z_{v,\tau} = A_{1,\tau} Z_{v,\tau-1} + B_\tau \varepsilon_{v,\tau}$$

$A_{1,\tau} = (n \times n)$ Periodic Coefficient Matrix of Period τ

$B_{\tau} = (n \times n)$ Periodic Coefficient Matrix of Period τ

$\underline{\varepsilon}_{v,\tau} = (n \times 1)$ Vector of Independent Normally Distributed Random Variable

บทที่ 8 แบบจำลองแยกส่วน (Disaggregation)

แบบจำลองพื้นฐาน (Basic Model)

$$Y = AX + B\underline{\varepsilon}$$

แบบจำลองแบบขยาย (Extended Model)

$$Y = AX + B\underline{\varepsilon} + CZ$$

Y =Column Vector ของค่ารายเดือนของปีปัจจุบัน ซึ่งมีขนาด (12×1)

X =ค่าปัจจุบันของอนุกรมเวลารายปี (1×1)

Z = Column Vector ของค่ารายเดือนจำนวน w' เดือนของปีก่อนหน้านั้น ซึ่งมีขนาด $(w' \times 1)$

$\underline{\varepsilon}$ = Column Vector ของค่าปัจจุบันของ Completely Random Series หรือ Stochastic Term
ซึ่งเมทริกซ์มีขนาด (12×1)

A =พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด (12×1)

B =พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด (12×12)

C =พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(12 \times w')$

กรณีที่จำนวนสถานีเท่า n จะสามารถเขียนแบบจำลองแยกส่วนแบบขยายเต็มรูปได้ ดังนี้

$$\begin{bmatrix} y_{v,1}^{(1)} \\ y_{v,2}^{(1)} \\ \vdots \\ y_{v,12}^{(1)} \\ y_{v,1}^{(2)} \\ y_{v,2}^{(2)} \\ \vdots \\ y_{v,12}^{(2)} \\ \vdots \\ y_{v,1}^{(n)} \\ y_{v,2}^{(n)} \\ \vdots \\ y_{v,12}^{(n)} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{12n,1} & a_{12n,2} & \dots & a_{12n,n} \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1,12n} \\ b_{21} & b_{22} & \dots & b_{2,12n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ b_{12n,1} & b_{12n,2} & \dots & b_{12n,12n} \end{bmatrix} \begin{bmatrix} \varepsilon_1^{(1)} \\ \varepsilon_2^{(1)} \\ \vdots \\ \varepsilon_{12}^{(1)} \\ \varepsilon_1^{(2)} \\ \varepsilon_2^{(2)} \\ \vdots \\ \varepsilon_{12}^{(2)} \\ \vdots \\ \varepsilon_1^{(n)} \\ \varepsilon_2^{(n)} \\ \vdots \\ \varepsilon_{12}^{(n)} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1,12n} \\ c_{21} & c_{22} & \dots & c_{2,12n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_{12n,1} & c_{12n,2} & \dots & c_{12n,12n} \end{bmatrix} \begin{bmatrix} y_{v-1,1}^{(1)} \\ y_{v-1,2}^{(1)} \\ \vdots \\ y_{v-1,12}^{(1)} \\ y_{v-1,1}^{(2)} \\ y_{v-1,2}^{(2)} \\ \vdots \\ y_{v-1,12}^{(2)} \\ \vdots \\ y_{v-1,1}^{(n)} \\ y_{v-1,2}^{(n)} \\ \vdots \\ y_{v-1,12}^{(n)} \end{bmatrix}$$

แบบจำลองแบบย่อ ใช้จำลองที่ละ 1 เดือน

$$Y_{\tau} = A_{\tau}X + B_{\tau}\underline{\varepsilon} + C_{\tau}Y_{\tau-1}$$

เมื่อ τ =เดือนปัจจุบัน เมื่อ $\tau=1$ ถึง 12 ($w=12$)

Y_{τ} =ค่าน้ำท่าของเดือนปัจจุบัน τ

X =ค่าน้ำท่าของปีปัจจุบัน

$Y_{\tau-1}$ = ค่าน้ำท่าของเดือนก่อนหน้านั้น $\tau-1$

$\underline{\varepsilon}$ = Column Vector ของค่าปัจจุบันของ Completely Random Series หรือ Stochastic Term

A_{τ} , B_{τ} , C_{τ} =ค่าพารามิเตอร์ของเดือน τ

1.12 เอกสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series.**
Water Resources Publications. USA. pp.484

1.13 แบบฝึกหัดท้ายบท

From the conceptual representation of the precipitation- stream flow process in Chapter 1, the stream flow process in a very simplified form, the groundwater recharge is assumed equal to the soil infiltration, can be represented by ARMA(1,1) model.

1. Let the precipitation-stream flow process be more complicated as shown in Figure 1. The groundwater recharge is a fraction of the soil water but there is no interflow. Please derive the stochastic model for the stream flow.

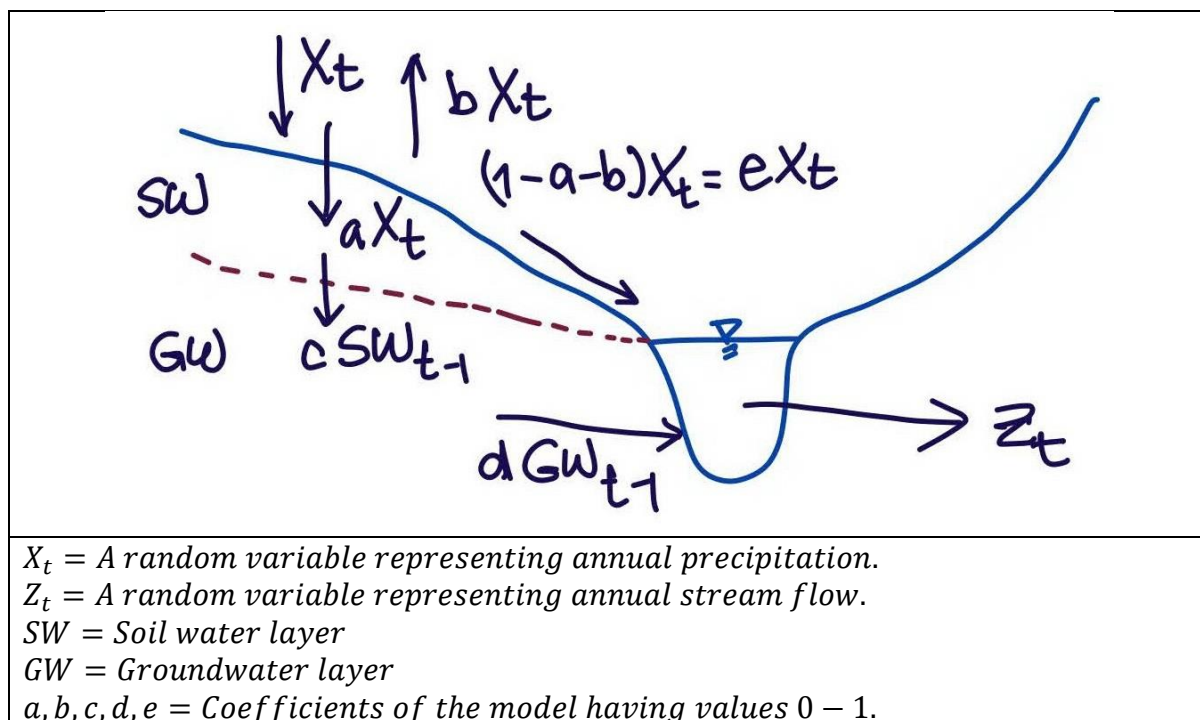


Figure 1

2. If the precipitation - stream flow process has the interflow term as shown in Figure 2, Please derive the stochastic model for the stream flow.

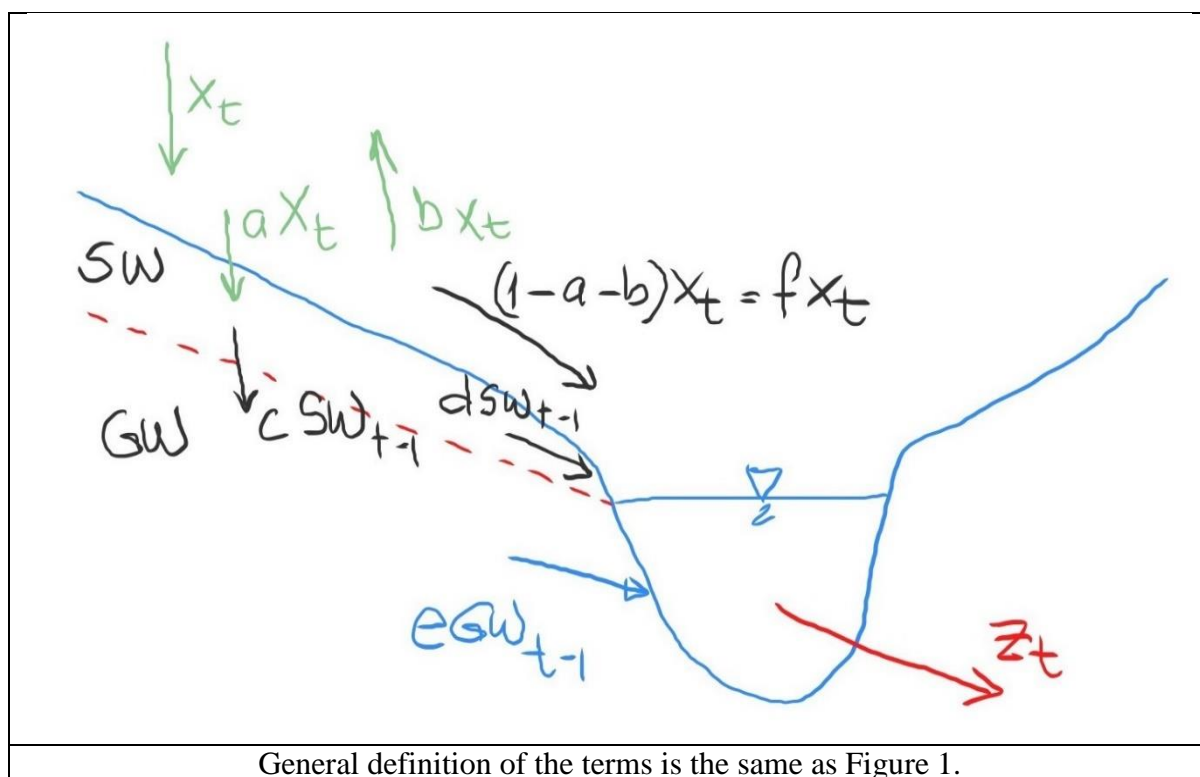


Figure 2

ภาคผนวก

APPENDIX A1.1 DEFINITIONS, TERMS AND NOTATION

Some definitions and terms are given herein so that the reader has less difficulty in following the material presented in this book. Also the notation has been carefully selected to make the presentation of equations and models clear and following as much as feasible, the standard notation commonly encountered in published literature.

DEFINITIONS AND TERMS

Normalization	The operation by which a time series is transformed into normal.
Standardization	The operation by which a time series with a given mean and standard deviation is converted into a series with mean zero and standard deviation one.
Independent series	Time series which does not have any dependence in time or in space.
Independent stochastic component	Time series which is independent in time and identically distributed.
White noise	Same as independent stochastic component but normally distributed.
Periodic series	Time series with periodic components or periodic parameters.
Seasonal series	Time series with time intervals that are a fraction of the year.
Historical series	Time series measured in the past.
Original series or data	Series or data available before any analysis is made.
AR model	Autoregressive model.
ARMA model	Autoregressive moving average model.
ARIMA model	Autoregressive integrated moving average model.
Empirical distribution	Frequency distribution of data (no reference to any certain probability distribution function).

Normal (0,1)	Normal distribution with mean zero and variance one.
Population	Theoretical, true or known distribution, parameter or statistical property.
Sample	Observed, assumed or generated data of a limited size.
Estimate	Distribution, parameter or any statistical property estimated from a sample.

NOTATION

<u>Item</u>	<u>Description</u>
Population parameter	lc. Greek letter and Roman letters. Example: ϕ , θ , b , A
Estimated parameter	lc. Greek and Roman letters with caret. Example: $\hat{\phi}$, $\hat{\theta}$, \hat{b} , \hat{A} , \bar{y}
Normal transformation	$Y = g(X)$. Example: $Y = \log(X)$
Inverse normal transformation	$X = g^{-1}(Y)$. Example: $X = \log^{-1}(Y) = \text{antilog}(Y)$
Logarithms	$\log_e(X)$, $\ln(X)$: Base e $\log_{10}(X)$: Base 10 $\log(X)$: Any base
Original data	X , x
Normalized data	Y , y
Standardized data	Z , z
Stochastic component	ε_t , ξ_t or independent series
Variance	$\text{Var}(X)$, σ^2 , s^2 , C_0
Standard deviation	σ , s
Covariance	$\text{Cov}(X, Y)$, S_{XY}
Covariance matrix of parameters $\hat{\phi}$ and $\hat{\theta}$	$V(\hat{\phi}, \hat{\theta})$

Autocovariance function	c_k
Autocorrelation function	ACF
Population autocorrelation function	$\rho_k, \rho_k(x), \rho_{k,\tau}$
Sample autocorrelation function	$r_k, r_k(x), r_{k,\tau}$
Partial autocorrelation function	$\phi_k(k), \text{PACF}$
Spectral density function	$g(f)$
Univariate series	$x_t, \varepsilon_t, x_t^{(i)}$
Multivariate series	$X_t, \underline{\varepsilon}_t$
Generated series	\hat{x}_t, \hat{X}_t
A converges to B	$A \rightarrow B$
R proportional to n	$R \sim n$
A equivalent to B	$A \leftrightarrow B$
Inverse of matrix A	A^{-1}
Transpose of matrix A	A^T
Derivative of S respect to ϕ	$\frac{\partial S}{\partial \phi}, \nabla_{\phi} S$

APPENDIX A1.2 ELEMENTARY STATISTICAL PRINCIPLES

Elementary statistical principles used in various chapters of this book are briefly reviewed herein. The purpose is a handy access to some basic definitions and properties. For a formal presentation on this subject the reader is referred to standard textbooks on probability and statistics such as those by Benjamin and Cornell (1970), or Mood et al. (1973).

Random Variable

It is a variable whose outcomes (values) are governed by chance. Its values can not be predicted with certainty but only in probability terms. Random variables can be discrete or continuous. Discrete random variables take on values only at discrete (specified) points, while continuous random

variables can take on any value on the real axis or any value between two boundaries values.

Probability Distribution Function

It is a function that defines the probability associated with a random variable. It is also called the probability law of a random variable. For instance, if a random variable X can take on only the values 0 and 1 with probabilities 0.3 and 0.7, respectively, then the probability distribution function or probability law is $p(X=0) = 0.3$ and $P(X=1) = 0.7$. The probability distribution function of discrete random variables may be represented by: the probability mass function (PMF) and the cumulative distribution function (CDF). For instance, for the discrete random variable $X=1, X=2, X=3$ and $X=4$, the PMF may be $P(X=1) = 0.2$, $P(X=2) = 0.35$, $P(X=3) = 0.25$ and $P(X=4) = 0.2$. The corresponding CDF is $P(X \leq 1) = 0.20$, $P(X \leq 2) = 0.20 + 0.35 = 0.55$, $P(X \leq 3) = 0.20 + 0.35 + 0.25 = 0.80$ and $P(X \leq 4) = 0.20 + 0.35 + 0.25 + 0.20 = 1.00$. These two functions are plotted in Fig. A1.1.

Similarly, the probability distribution function of a continuous random variable is represented by: the probability

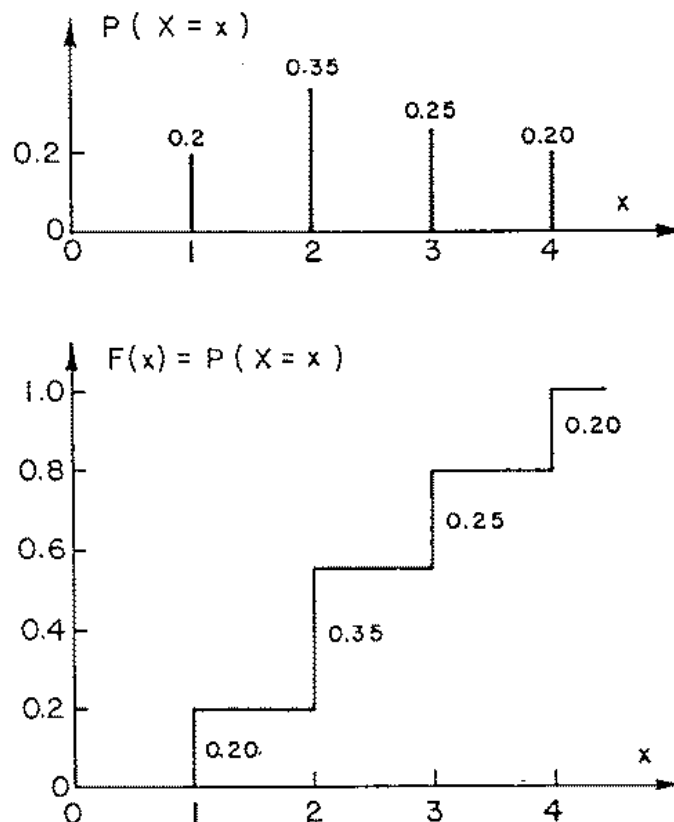


Figure A1.1. PMF and CDF of a discrete random variable.

density function (PDF), and the cumulative distribution function (CDF). The PDF of a random variable X , usually denoted by $f(x)$, serves to determine probabilities by integration. That is, the probability of X to be between x_1 and x_2 is obtained by $P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx$. On the other hand, the CDF, denoted by $F(x)$ is the cumulative probability from $-\infty$ up to x . That is

$$F(x) = \int_{-\infty}^x f(x) dx \quad (\text{A1.1})$$

Figure A1.2 shows schematically the PDF and CDF of a continuous random variable.

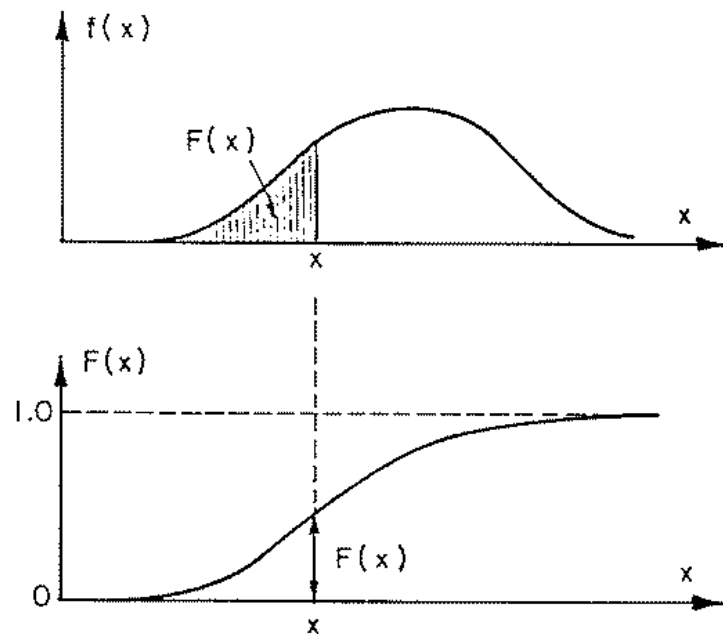


Figure A1.2. PDF and CDF of a continuous random variable.

Several types of probability distribution functions are used for discrete and continuous random variables. Examples of such functions are shown in Table A1.1.

Joint Random Variables and Probability Distribution Functions

Two or more random variables considered simultaneously are called joint random variables. The probability distribution of such variables are called joint probability distribution functions. For instance, the joint PDF of the continuous variables X and Y are denoted by $f(x,y)$ and the joint CDF by

Table A1.1. Examples of Probability Distribution Functions

Name	Function
<u>DISCRETE</u>	
Uniform	$P(X=x) = \frac{1}{n} ,$ $x = 1, 2, \dots, n$
Binomial	$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} ,$ $x = 0, 1, 2, \dots, n$
Poisson	$P(X=x) = \frac{e^{-v} v^x}{x!} ,$ $x = 0, 1, \dots, \infty$
Geometric	$P(X=x) = (1-p)^{x-1} p ,$ $x = 1, 2, 3, \dots$
<u>CONTINUOUS</u>	
Uniform	$f(x) = \frac{1}{(b-a)} ,$ $a < x < b$
Normal	$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\} ,$ $-\infty < x < \infty$
Lognormal -2	$f(x) = \frac{1}{\sqrt{2\pi} \sigma_y x} \exp \left\{ -\frac{1}{2} \left(\frac{\ln x - \mu_y}{\sigma_y} \right)^2 \right\} ,$ $0 < x < \infty$
Gamma-2	$f(x) = \frac{\alpha^k x^{\beta-1} e^{-\alpha x}}{\Gamma(\beta)} ,$ $0 < x < \infty$

by $F(x,y)$. As in Eq. (A1.1) the relation between $f(x,y)$ and $F(x,y)$ can be written as

$$F(x,y) = \int_{-\infty}^x \int_{-\infty}^y f(x,y) dx dy \quad (\text{A1.2})$$

Independent and Dependent Random Variables

Two continuous random variables X and Y are independent if $f(x,y) = f(x) \cdot f(y)$ or $F(x,y) = F(x) \cdot F(y)$. Otherwise they are dependent variables.

Expectation and Population Moments

The expectation $E(X)$ of a continuous random variable X is defined by the integral

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \mu \quad (\text{A1.3})$$

It is called the expected value or the mean of X . It is also called the first moment of X around zero. The r -th moment of X around zero is

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx \quad (\text{A1.4})$$

On the other hand, the r -th moment around the mean μ is

$$E[(X-\mu)^r] = \int_{-\infty}^{\infty} (x-\mu)^r f(x) dx \quad (\text{A1.5})$$

The variance of the variable x is the second moment of X around μ , or

$$E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \text{Var}(X) \quad (\text{A1.6})$$

The variance $\text{Var}(X)$ is also written as

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2 \quad (\text{A1.7})$$

Covariance is the linear relation between two random variables X and Y . For continuous variables the covariance is defined by

$$\begin{aligned} \text{Cov}(X, Y) &= E\{(X - \mu_x)(Y - \mu_y)\} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) \, dx dy \end{aligned} \quad (\text{A1.8})$$

The covariance is also written as

$$\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y) \quad (\text{A1.9})$$

Similar definitions as the ones given above exist for discrete random variables.

Sample

A sample is a set of observations x_1, x_2, \dots, x_N obtained from a known or unknown probability distribution function. For instance, if we take 20 numbers from a table of standardized normal random numbers they constitute a (random) sample of size $N = 20$ obtained from the known normal distribution. On the other hand, 20 years of annual streamflows also constitute a sample of size $N = 20$ but in this case the distribution is not known.

Frequency Distribution

If the sample x_1, \dots, x_N is ordered from the lowest to the largest, and an empirical probability is assigned to each ordered value, the array of ordered values and corresponding probabilities is called the frequency distribution. The ordered sample may be also arranged in groups or class intervals. In this case the frequency distribution is given by the typical histogram.

Sample Moments

The r -th moment around zero of a sample x_1, \dots, x_N is defined by

$$m_r = \frac{1}{N} \sum_{i=1}^N x_i^r \quad (\text{A1.10})$$

For $r = 1$, m_1 of Eq. (A1.10) is the sample mean \bar{X} . The r -th moment also can be defined around the sample mean \bar{X} as

$$m'_r = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^r \quad (\text{A1.11})$$

Equations (A1.10) and (A1.11) are the basic equations from which the commonly used statistical characteristics of samples are derived.

APPENDIX A1.3 ELEMENTARY MATRIX DEFINITIONS AND COMPUTATIONS

Matrix Definitions

A matrix is constituted by elements arranged in a given number of rows and given number of columns. For instance, a matrix A with 2 rows and 3 columns is represented as

$$A_{(2 \times 3)} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

The order of a matrix is given by the number of rows and columns. For instance, a $m \times n$ matrix has m rows and n columns. A matrix with only one row is called a row matrix and a matrix with only one column is called a column matrix. A square matrix has the same number of rows and columns. A diagonal matrix is a square matrix whose off diagonal elements are zero. The identity matrix I, is a diagonal matrix with elements equal to one. A lower triangular matrix has all elements above the diagonal equal to zero.

Addition, Subtraction and Multiplication of Matrices

Assume the (2×2) matrices A and B. The matrix $C = A \pm B$ is equal to

$$C = A \pm B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \pm \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} \end{bmatrix}$$

Thus, to add or to subtract matrices, they must be of the same order.

A matrix A($m \times n$) can be multiplied with another matrix B if this second matrix is of order ($n \times q$). For instance, let us multiply A(2×3) and B(3×1). We obtain

$$C = A \times B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}_{(2 \times 3)} \cdot \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}_{(3 \times 1)} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \end{bmatrix}_{(2 \times 1)}$$

A is positive semidefinite if the inequalities in eq. (A1.15) are replaced by \geq signs.

บทที่ 2

คุณลักษณะของอนุกรมเวลาทางอุทกวิทยา

CHARACTERISTICS OF HYDROLOGIC TIME SERIES

2.1 ชนิดของอนุกรมเวลา (Types of Time Series)

ในการวิเคราะห์ทางสถิติจะสามารแบ่งอนุกรมเวลาตามมิติได้เป็น 2 แบบคือ อนุกรมเวลา 1 มิติ (Unidimensional) และอนุกรมเวลาหลายมิติ (Multidimensional) ดังตารางที่ 2.1 อนุกรมเวลาแบบ 1 มิติคืออนุกรมเวลาของค่าสังเกต 1 ตัวแปรที่มีการบันทึก ณ ที่ใดที่หนึ่ง เช่นปริมาณฝนรายปีที่สถานี ก ส่วนอนุกรมเวลาหลายมิติคืออนุกรมเวลาของค่าสังเกตของหลายตัวแปรที่มีการบันทึก ณ ที่ใดที่หนึ่ง เช่นคุณภาพน้ำ DO, EC, pH ที่สถานี ข (Multivariate) หรืออนุกรมเวลาของค่าสังเกตของตัวแปรประเภทเดียวกันที่มีการบันทึก ณ หลายจุดพร้อมกัน (Multisite หรือ Multipoint) เช่น ปริมาณฝนรายปีที่สถานี ก, ข, ค, ง เป็นต้น ในการวิเคราะห์อนุกรมเวลาจะวิเคราะห์แบบ 1 มิติหรือหลายมิติ ขึ้นอยู่กับค่าสหสัมพันธ์ระหว่างอนุกรมเวลา (Cross Correlation) หรือระหว่างตัวแปรนั้นๆ เป็นเกณฑ์ ถ้าอนุกรมเวลาแบบมีสหสัมพันธ์สูง ก็ควรทำการวิเคราะห์แบบหลายมิติ ถ้าไม่มีสหสัมพันธ์ก็ควรวิเคราะห์แบบ 1 มิติ

ตารางที่ 2.1 การแบ่งชนิดของอนุกรมเวลาในการวิเคราะห์สถิติ

Unidimensional (Univariate)	Multidimensional (Multivariate)
- time	- multisite
- line	- multipoint
- counting series (no. of rainy days)	- multivariate
For example	For example
- flow at a gaging station - average rainfall over an area - aggregated flow over watershed	- flow at several points - different kinds of series at one point (several water quality parameters at one point)such as BOD, COD, DO, ...
No cross correlation exists.	Cross correlation exists.

2.2 ช่วงเวลาของอนุกรมเวลา (Time Interval)

อนุกรมเวลาทางอุทกวิทยาจะมีช่วงเวลาในการเก็บข้อมูลหลายระดับขึ้นอยู่กับความต้องการใช้งานเป็นหลัก ถ้าเก็บต่อเนื่องตลอดเวลา เช่นเก็บด้วยเครื่องเก็บและบันทึกข้อมูลอัตโนมัติ (Automatic Recording Machine) จะได้ข้อมูลแบบต่อเนื่องเพื่อใช้งานเฉพาะด้านเช่นการศึกษาความรุนแรงของพายุ นอกจากนี้อาจมีการเก็บและบันทึกข้อมูลเป็นรายชั่วโมง รายวัน รายสัปดาห์ รายเดือนหรือรายปี ช่วงเวลาของอนุกรมเวลาจะมีผลต่อการแสดงอิทธิพลขององค์ประกอบต่างๆของอนุกรมเวลาโดยเฉพาะอย่างยิ่ง Periodic Component ดังตารางที่ 2.2

ตารางที่ 2.2 การแบ่งช่วงเวลาและอิทธิพลขององค์ประกอบของอนุกรมเวลา

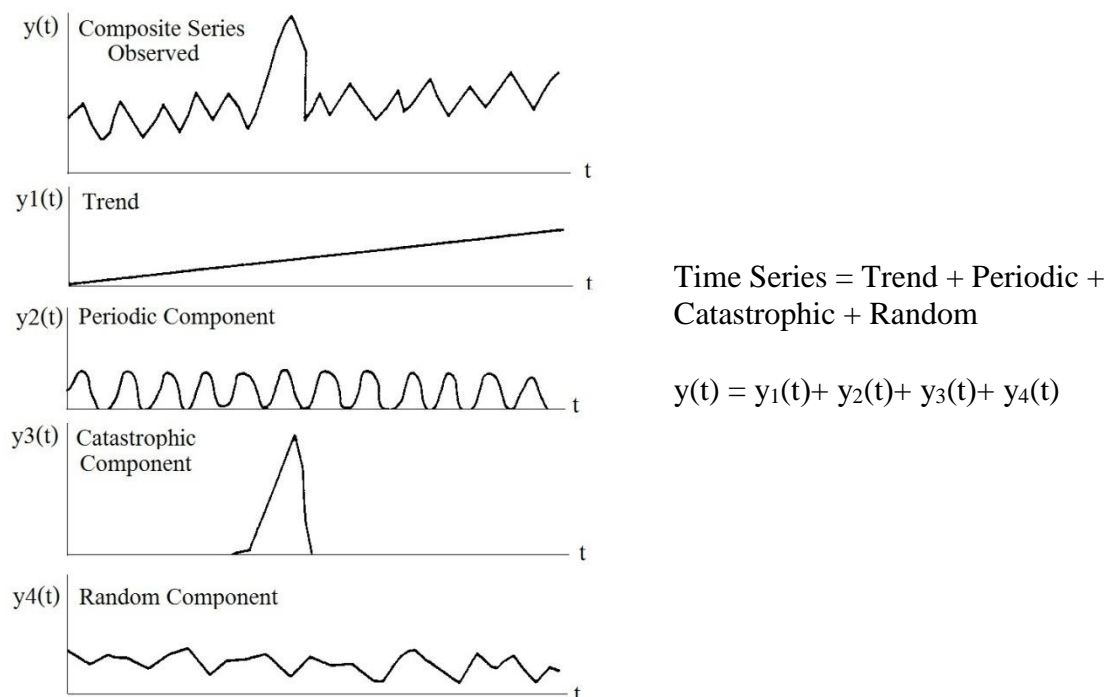
Time Interval	Components of Time Series
0 (Continuous Series)	All Components + Randomness
Hourly	Daily / Annual Cycles + Randomness
Day, week, month	Annual Cycle + Randomness
Annual	No cycle + Randomness

2.3 คุณสมบัติทั่วไปของอนุกรมเวลาทางอุทกวิทยา (General Properties of Hydrologic Time Series)

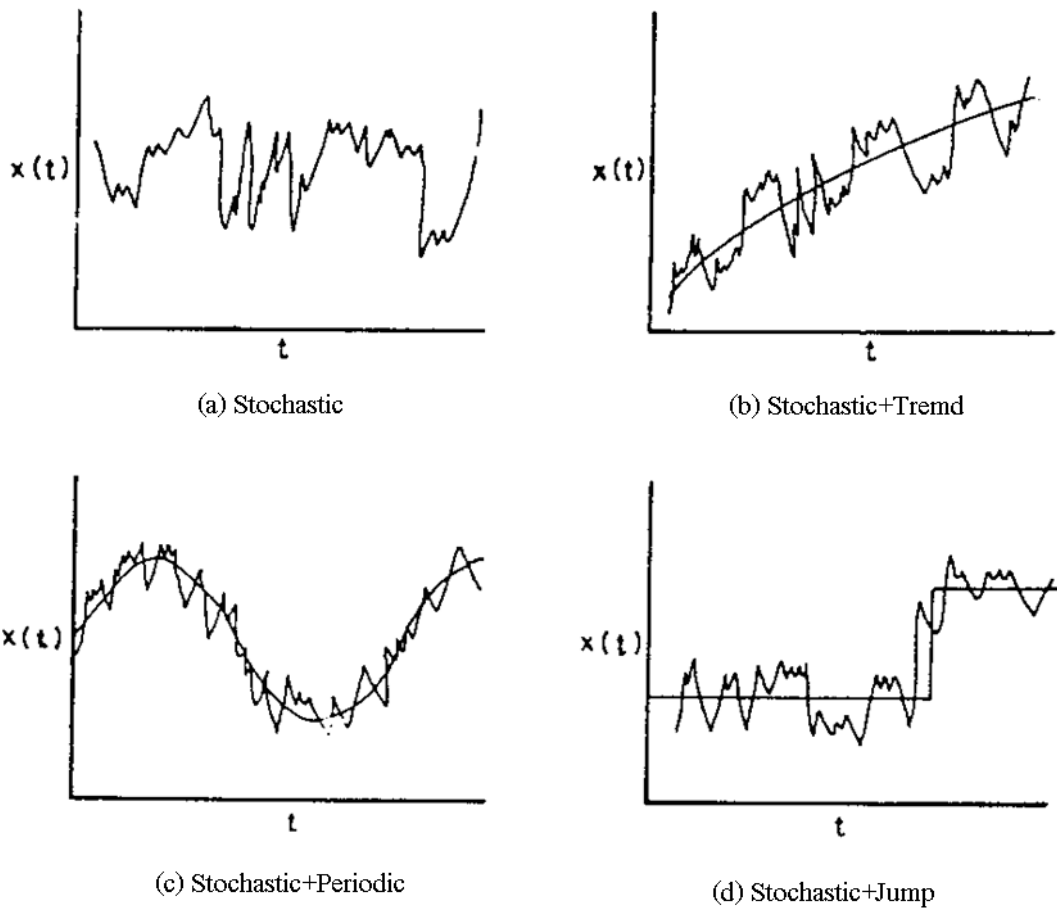
ตามที่กล่าวไว้ในบทที่ 1 อนุกรมเวลาทางอุทกวิทยาประกอบด้วย 4 องค์ประกอบที่สำคัญ คือ

- (1) Over year trend and other deterministic changes (trend + jump)
- (2) Cycle or periodic change of day and year
- (3) Almost periodic such as tidal
- (4) Stochastic or randomness

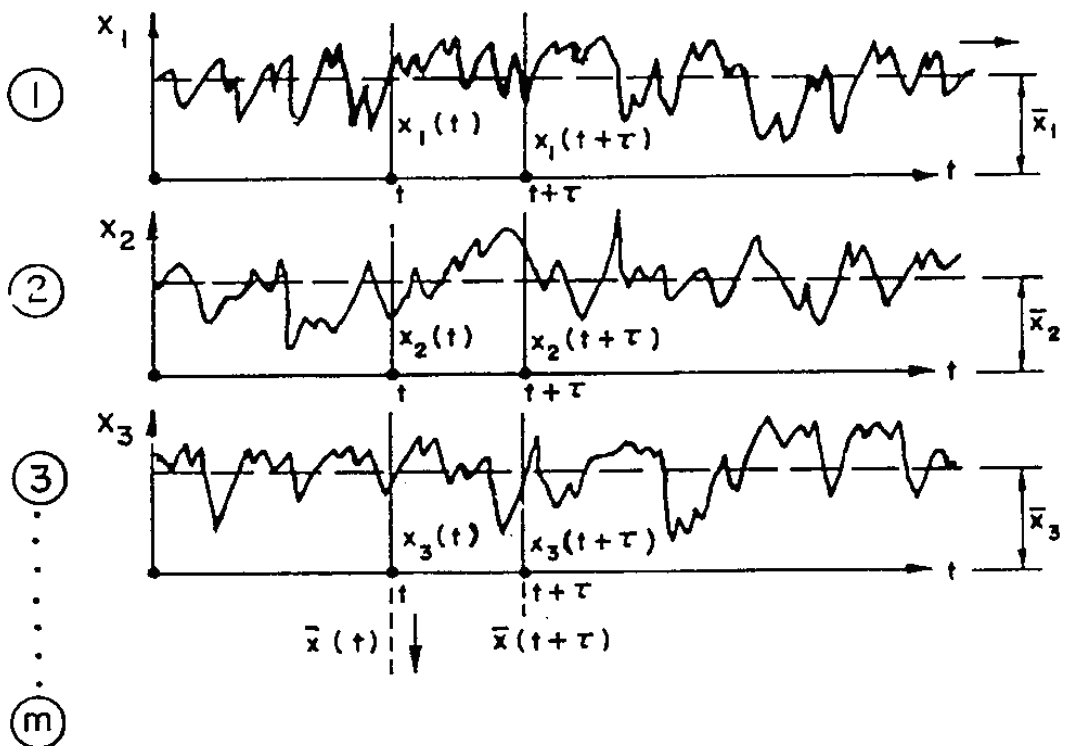
รูปที่ 2.1 แสดงกราฟอนุกรมเวลาซึ่งมีองค์ประกอบต่างๆ และการพล็อตแยกแต่ละองค์ประกอบของอนุกรมเวลา รูปที่ 2.2 แสดงองค์ประกอบทาง Stochastic และ Deterministic ในอนุกรมเวลา รูปที่ 2.3 แสดงอนุกรมเวลาแบบต่างๆ เปรียบเทียบกัน



รูปที่ 2.1 Time series components

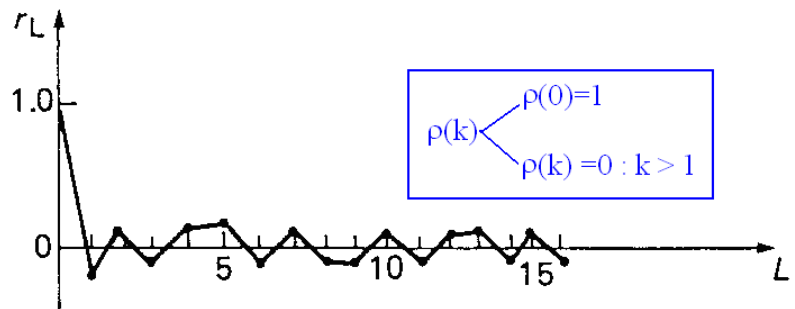


รูปที่ 2.2 Time Series Containing Stochastic and Several Types of Deterministic Components

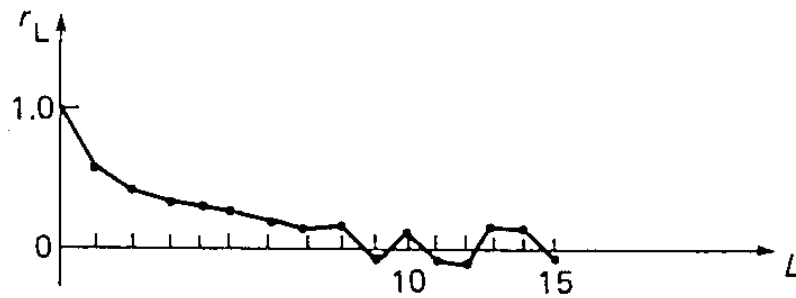


รูปที่ 2.3 Several Realizations of Stochastic Process

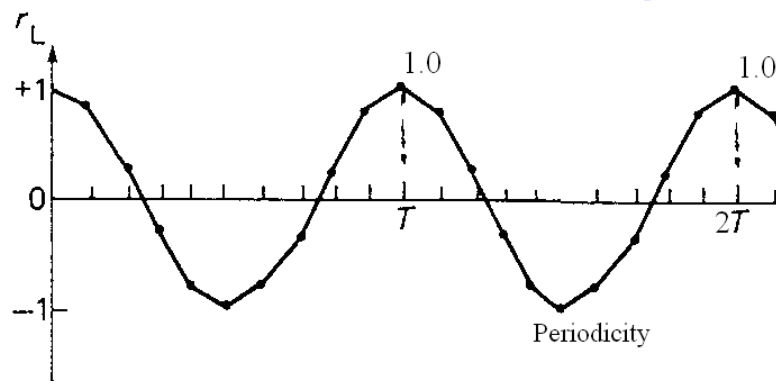
รูปที่ 2.4 แสดง Correlogram ของอนุกรมเวลาที่มีโครงสร้างแบบ Independent Noise, Autoregressive และแบบ Periodic รูปที่ 2.5 แสดง Correlogram ของอนุกรมเวลาที่มี Periodicity แบบหลายความถี่ รูปที่ 2.6 แสดง Correlogram ของอนุกรมเวลา (z_t) และ Residuals (e_t) ของ Autoregressive Model ลำดับที่ 1



(a) Random, independent noise

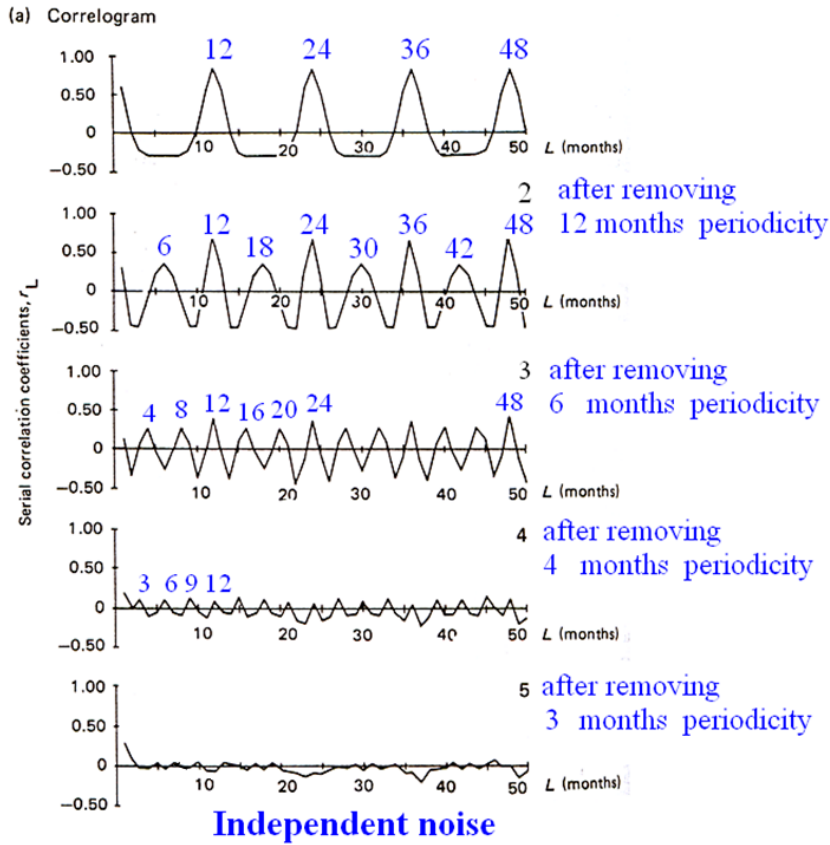


(b) Autoregressive, Markov process (Dependent)



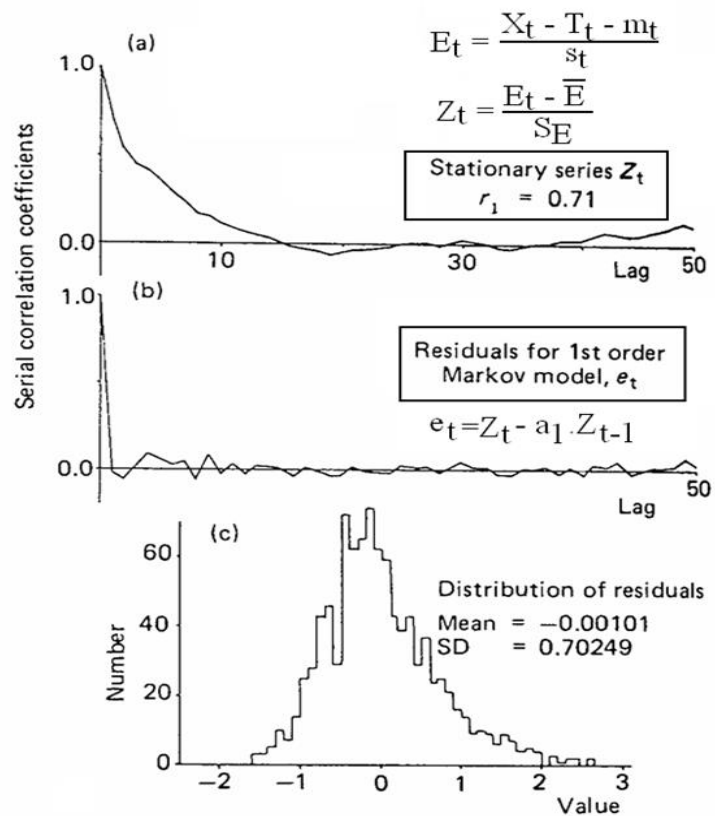
(c) Pure sine wave

รูปที่ 2.4 Correlogram



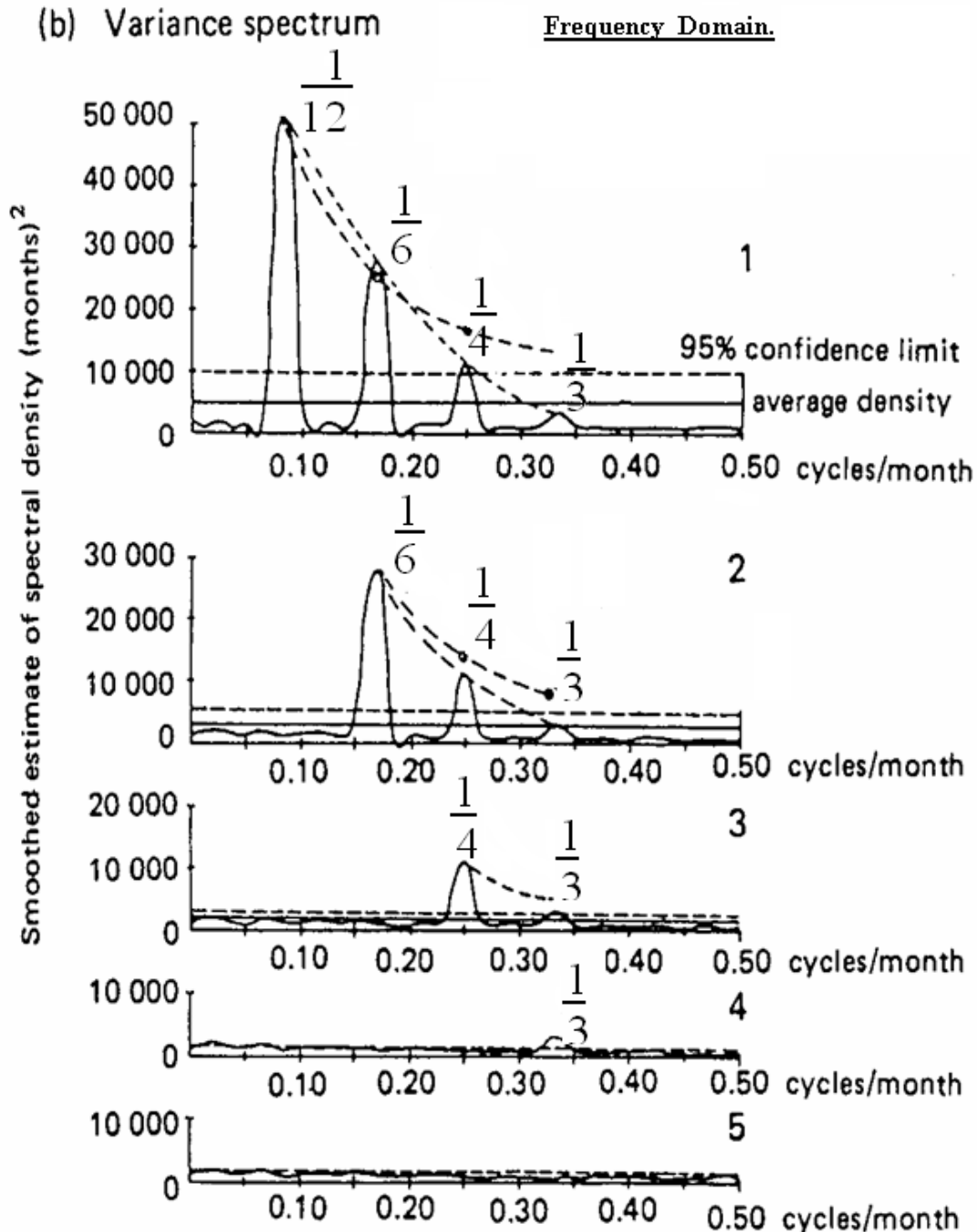
$$E_t = \frac{X_t + T_t - m_t}{S_t}$$

รูปที่ 2.5 Periodic Correlogram



รูปที่ 2.6 River Thames at Teddington Weir (82 years of monthly flows)

รูปที่ 2.7 แสดง Spectral Density Function ของอนุกรมเวลาที่ใช้คำนวณ Correlogram ในรูปที่ 2.5 ซึ่งมีหลาย Periodicity Spectral Density Function คือการวิเคราะห์ห้อนุกรมเวลาบน Frequency Domain ขณะที่ Correlogram คือการวิเคราะห์ห้อนุกรมเวลาบน Time Domain ซึ่งจะเห็นได้ว่าผลการวิเคราะห์ Spectral Density Function มีข้อดีคือจะทำให้เห็นอิทธิพลของ Periodic ของความถี่



รูปที่ 2.7 Periodic components of 30 years runoff record, ELK river at Clark, Colorado, USA.

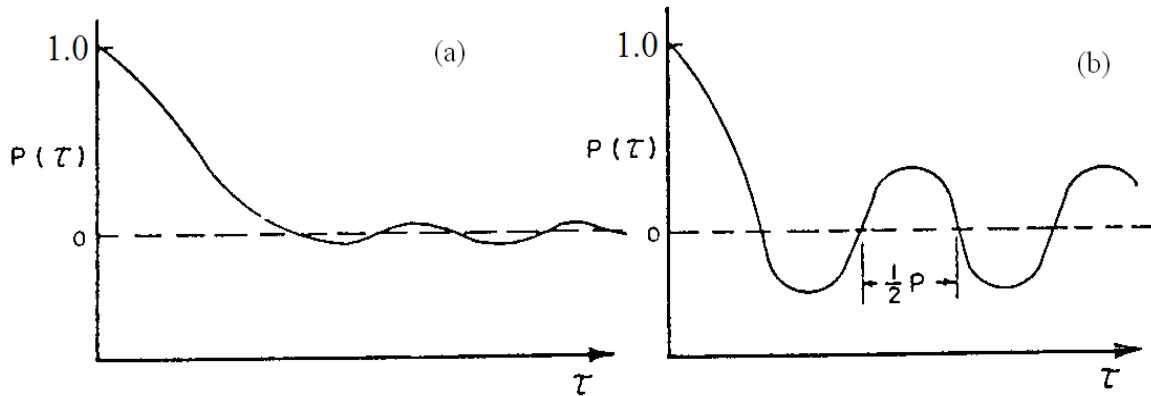
12, 6, 4 and 3-month periods present (2) 12-month period removed (3) 12 and 6-month periods removed (4) 12, 6 and 4-month periods removed (5) all periods removed (Reproduced from L.A. Roesner & V.N. Yevjevich (1966) Mathematical Models for time series of monthly precipitation and runoff, hydrological paper no. 15, Colorado State University)

2.4 สหสัมพันธ์ในตัวเอง (Autocorrelation)

สหสัมพันธ์ในตัวเอง คือการวิเคราะห์ห้อนุกรมเวลาใน Time Domain เพื่อหว่านอนุกรมเวลานั้นๆ มีองค์ประกอบทาง Stochastic และ Deterministic อะไรบ้าง รูปที่ 2.8 แสดง Correlogram ของอนุกรมเวลาที่มาจากรandom Process เปรียบเทียบกับอนุกรมเวลาที่มาจากรandom Process Superimposed on a Periodic Process Autocorrelation Coefficient ที่ Lag ต่างๆ สามารถคำนวณได้จากสมการ 2.1 หรือ 2.2

$$\rho(\tau) = \frac{COV(x(t), x(t+\tau))}{VAR(x(t))} \quad [2.1]$$

$$\rho(\tau) = \frac{COV(x(t), x(t+\tau))}{\sqrt{VAR(x(t))VAR(x(t+\tau))}} \quad [2.2]$$



รูปที่ 2.8 Typical Correlogram (a) Random Process (b) Random Process Superimposed on a Periodic Process

2.5 Spectral Analysis

Periodicities สามารถหาได้โดยการวิเคราะห์ห้อนุกรมเวลาบน Frequency Domain หรือการหา Spectral Density Function (SDF)

$$\text{Spectral Density Function} = S(f) = \frac{\text{Variance}}{\text{Interval of Frequency}} \quad [2.3]$$

เมื่อ

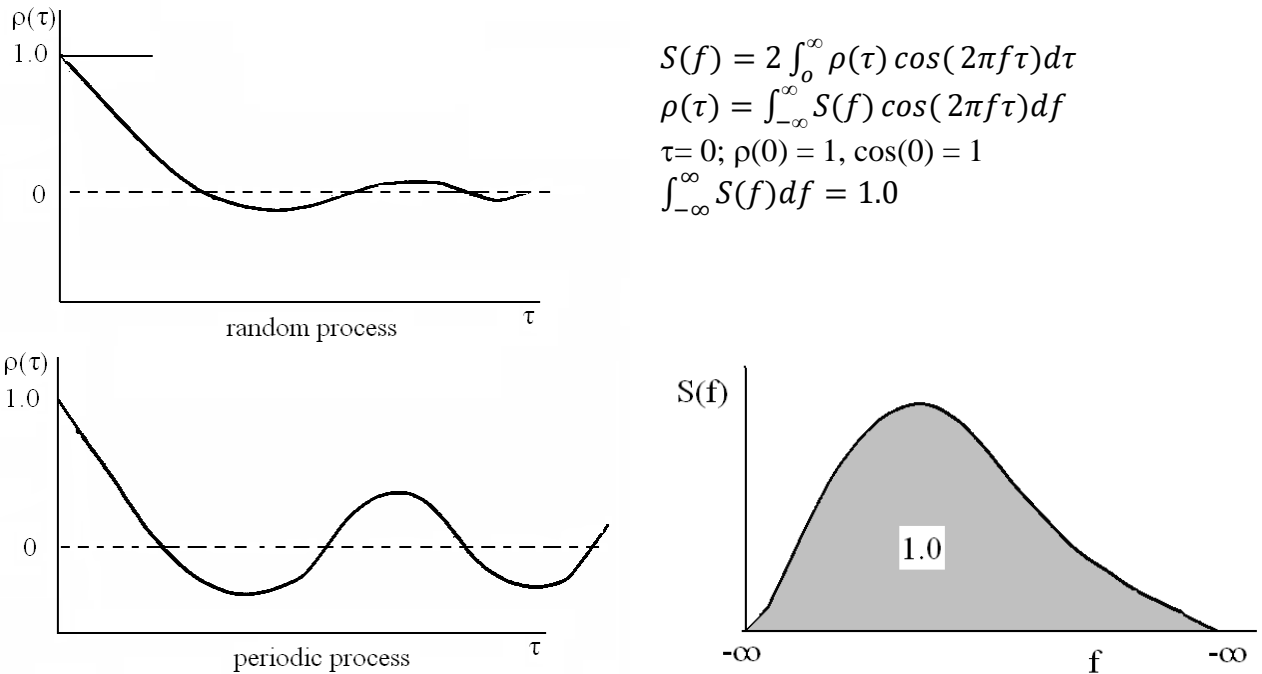
f = Frequency (cycle/unit time)

w = Angular Frequency (radians/unit time) = $2\pi f$

$$S(w) = S(2\pi f) = 2\pi S(f) \quad [2.4]$$

2.5.1 Spectral Density Function Versus Autocorrelation Function

SDF และ ACF มีความสัมพันธ์กันดังรูปที่ 2.9 และสมการ 2.5 และ 2.6



รูปที่ 2.9 ความสัมพันธ์ระหว่าง SDF และ ACF

$$S(f) = 2 \int_0^\infty \rho(\tau) \cos(2\pi f\tau) d\tau \quad [2.5]$$

$$\rho(\tau) = \int_{-\infty}^\infty S(f) \cos(2\pi f\tau) df \quad [2.6]$$

ให้ $\hat{S}(f)$ คือค่าประมาณของ $S(f)$ ในสมการที่ 2.5 ซึ่งจะหาได้จาก Numerical Integration ของสมการ 2.5 จะได้ Sample SDF $\hat{S}'(f)$ ดังสมการ 2.6 หลังจากนั้นจึงนำค่า $\hat{S}'(f)$ ไปหาค่าเฉลี่ยจะได้ ค่า $\hat{S}(f)$ ที่ต้องการ

$$\hat{S}'(f) = 2 \left[\frac{r(0) \cos(0) + r(1) \cos(2\pi f\Delta t)}{2} \Delta t + \frac{r(1) \cos(2\pi f\Delta t) + r(2) \cos(2\pi f2\Delta t)}{2} \Delta t + \dots + \frac{r(m-1) \cos(2\pi f(m-1)\Delta t) + r(m) \cos(2\pi fm\Delta t)}{2} \Delta t \right]$$

$$\hat{S}(f) = \Delta t \left[r(0) + 2 \sum_{k=1}^{m-1} r(k) \cos(2\pi fk\Delta t) + r(m) \cos(2\pi fm\Delta t) \right] \quad [2.7]$$

ถ้า $f=1$

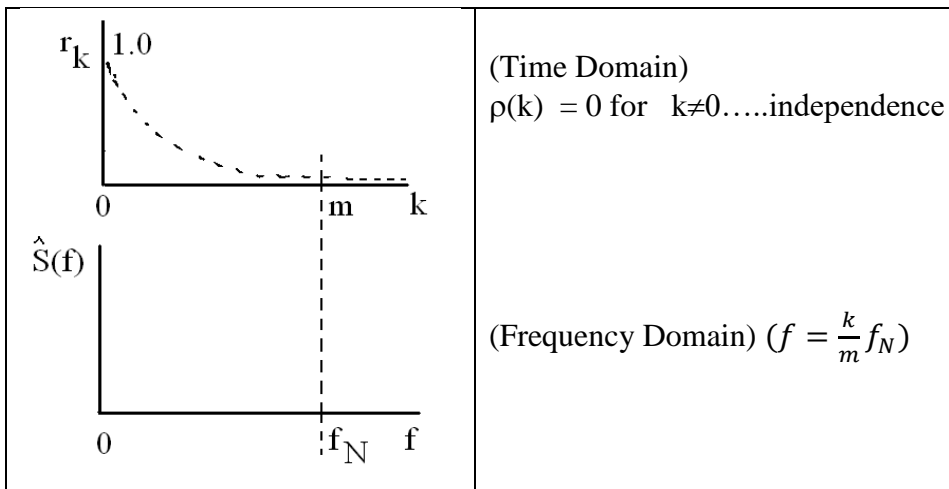
$$\hat{S}'(1) = \Delta t \left[r(0) + 2 \sum_{k=1}^{m-1} r(k) \cos(2\pi k\Delta t) + r(m) \cos(2\pi m\Delta t) \right] \quad [2.8]$$

$$f_N = \frac{1}{2\Delta t} \quad (\text{Nyquist frequency}) \quad [2.9]$$

$$f = \frac{kf_N}{m} \quad [2.10]$$

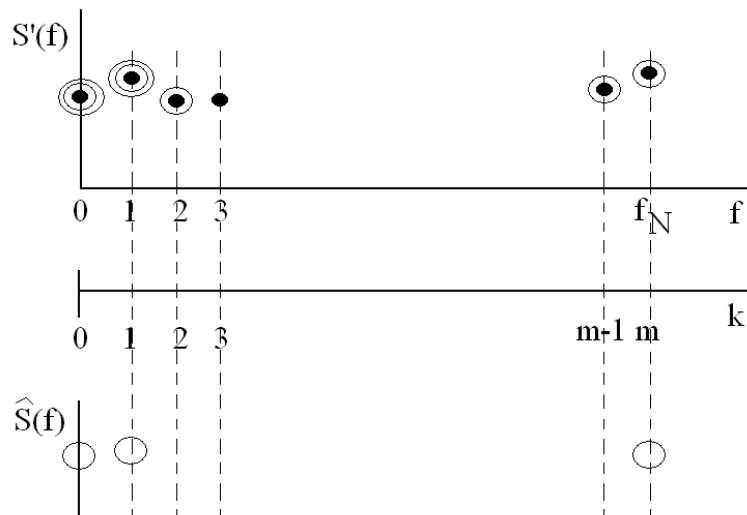
$$m = \text{Max. Lag} \approx \frac{N}{4}$$

$$k = \text{lag } k$$



2.5.2 วิธีการคำนวณหาค่าประมาณของ SDF $\hat{S}(f)$

ค่าประมาณของ SDF $\hat{S}(f)$ จะหาได้โดยการหาค่าเฉลี่ยของ $S'(f)$ ซึ่งคำนวณจากค่า k ตั้งแต่ 0 ถึง m (Maximum Lag) ดังนี้ (ดูรูปที่ 2.10)



รูปที่ 2.10 วิธีการคำนวณหาค่าประมาณของ SDF $\hat{S}(f)$

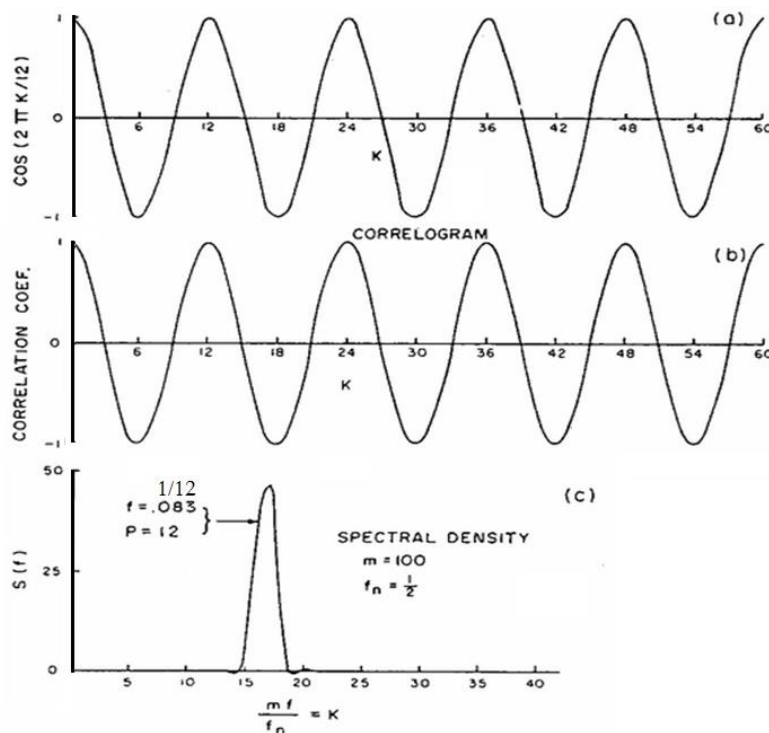
$k = 0,$ $f = kf_N/m = 0$	$\hat{S}(0) = 0.5[\hat{S}'(0) + \hat{S}'(\frac{f_N}{m})]$ [2.11]
$0 < k < m,$ $f = kf_N/m$	$\hat{S}(\frac{kf_N}{m}) = 0.25\hat{S}'((k-1)\frac{f_N}{m}) + 0.5\hat{S}'(\frac{kf_N}{m}) + 0.25\hat{S}'((k+1)\frac{f_N}{m});$ for $k = 1, 2, \dots, m-1$ [2.12]
$k = m, f = f_N$	$\hat{S}(f_N) = 0.5[\hat{S}'((m-1)\frac{f_N}{m}) + \hat{S}'(f_N)]$ [2.13]

ตัวอย่างการคำนวณหา ACF และ SDF สำหรับอนุกรมเวลาแบบ Periodic แบบ Cosine ซึ่งมีความถี่เท่ากับ $1/12 \{y_t = \cos(2\pi k/12)\}$ และแบบค่าเฉลี่ยของ 3 Cosine ซึ่งมีความถี่ เท่ากับ $1/6, 1/12$ และ $1/24$

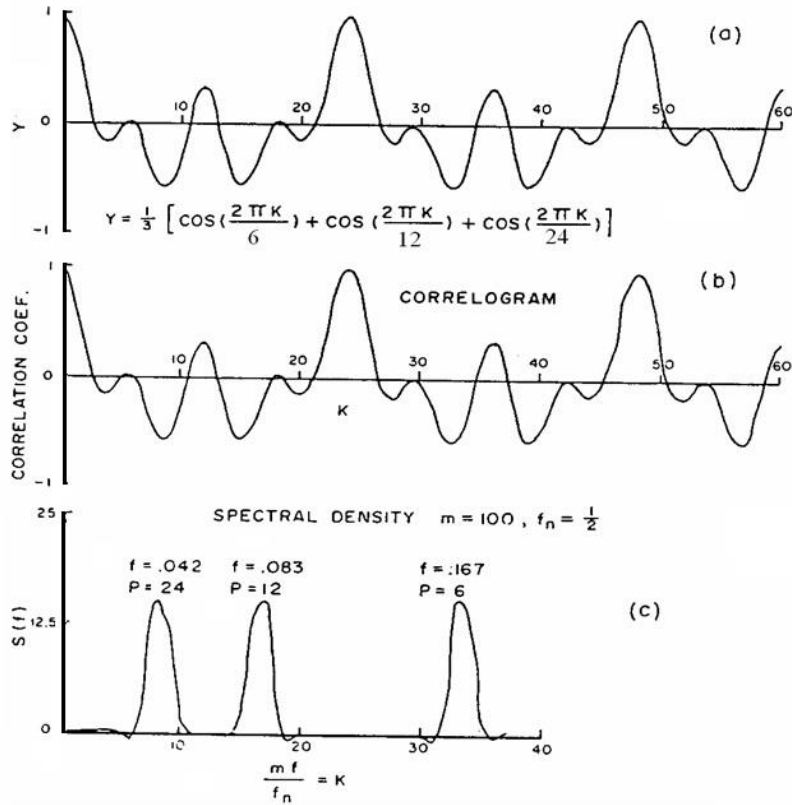
$$y_t = \frac{\cos(2\pi k/6) + \cos(2\pi k/12) + \cos(2\pi k/24)}{3}$$

แสดงอยู่ในรูปที่ 2.11 และ 2.12 ตามลำดับ

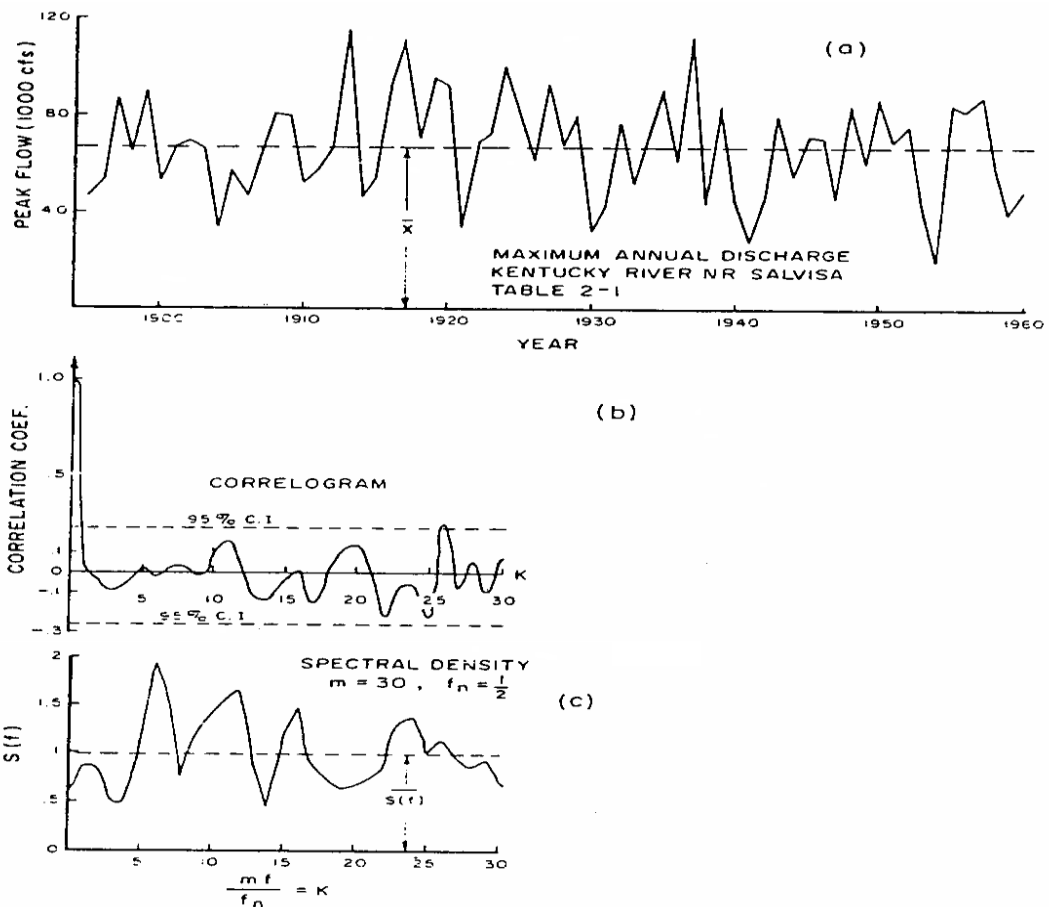
ตัวอย่างการคำนวณหา ACF และ SDF สำหรับอนุกรมเวลาของปริมาณการไหลสูงสุดของ Kentucky River แสดงอยู่ในรูปที่ 2.13 และตัวอย่างการคำนวณหา ACF และ SDF สำหรับอนุกรมเวลาของปริมาณน้ำท่ารายเดือนของ Cave Creek แสดงอยู่ในรูปที่ 2.14



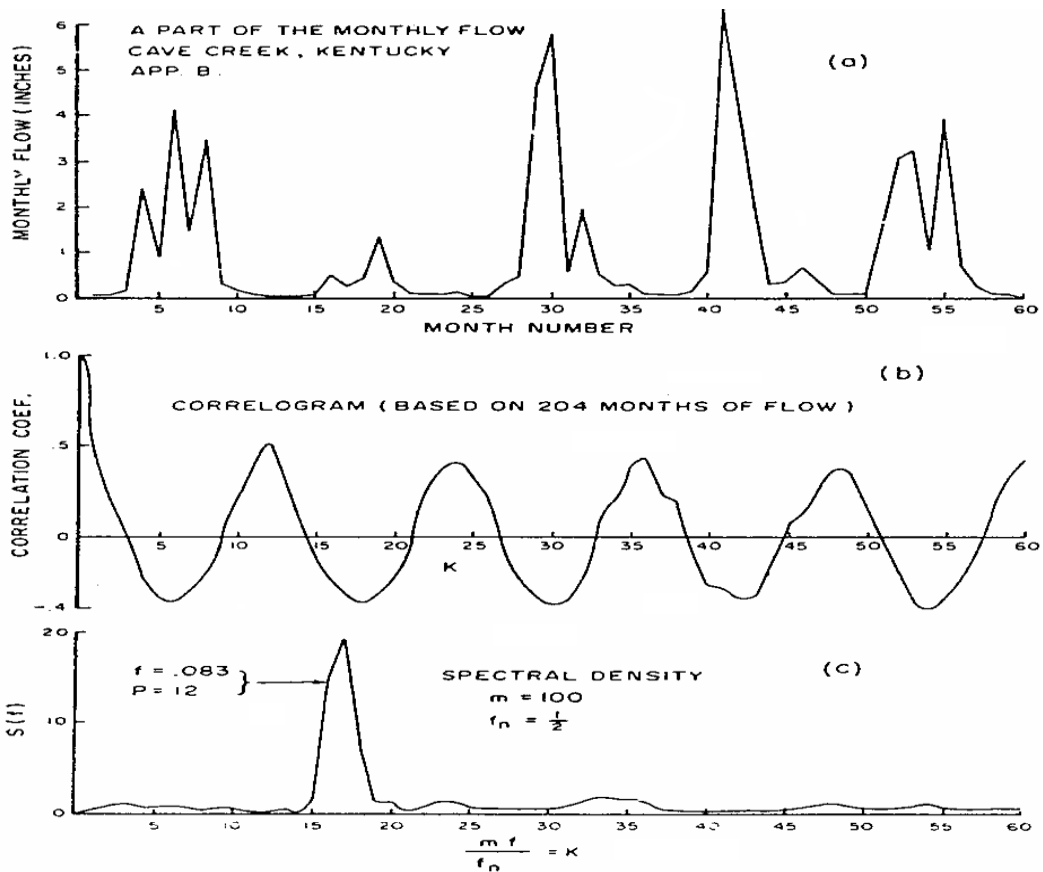
รูปที่ 2.11 (a) $\cos(2\pi k)/12$ (b) its correlogram (c) spectral density



รูปที่ 2.12 (a) Sum of 3 cosines (b) its correlogram (c) spectral density



รูปที่ 2.13 (a) Kentucky River Peak Flows and its Correlogram (b) Spectral Density



รูปที่ 2.14 Cave Creek Monthly Runoff (b) its Correlation (c) Spectral Density

2.6 คุณสมบัติทางสถิติพื้นฐานของอนุกรมเวลา (Basic Statistical Characteristics of Time Series)

ค่าสถิติพื้นฐานของอนุกรมเวลาได้แก่ ค่าเฉลี่ย ค่าความแปรปรวน ค่าสัมประสิทธิ์ความเบ้ (Skewness Coefficient) Autocovariance และ Autocorrelation ดังสมการที่ 2.14 – 2.18

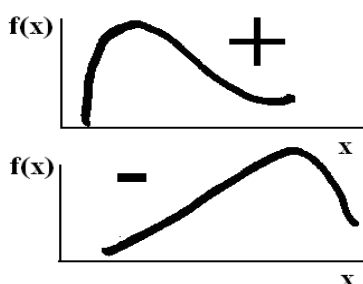
$$\bar{x} = \sum_{t=1}^N \frac{x_t}{N} \sim \mu \tag{2.14}$$

$$s^2 = \sum_{t=1}^N \frac{(x_t - \bar{x})^2}{N-1} \sim \sigma^2 \tag{2.15}$$

$$g = \frac{N \sum_{t=1}^N (x_t - \bar{x})^3}{(N-1)(N-2)s^3} \sim \gamma \tag{2.16}$$

$$c_k = \frac{1}{N-k-1} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{2.17}$$

เมื่อ $0 \leq k \leq N$; $k = \text{lag}$



การแจกแจงความน่าจะเป็นอาจมีค่าสัมประสิทธิ์ความเบ้เป็นบวกหรือลบก็ได้ ถ้ามีค่าเป็นบวกเรียกว่าเบ้ขวา ถ้ามีค่าเป็นลบเรียกว่าเบ้ซ้าย ดังรูป

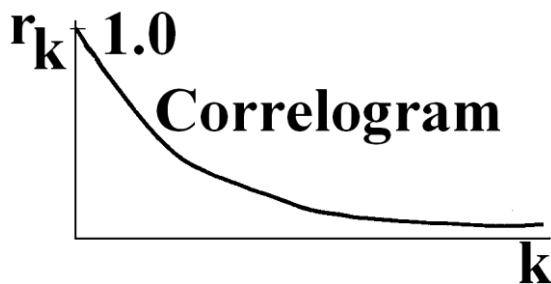
- $g = 0$ (Normal)
- $g > 0$ (Right Skewed)
- $g < 0$ (Left Skewed)

Lag k Autocovariance แสดงถึงระดับ Linear Auto-dependence หรือ Self Dependence ระหว่างอนุกรม

เวลา

$$r_k = \frac{C_k}{C_0} = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N-k} (x_t - \bar{x})^2} \approx \rho_k \quad [2.18]$$

= Lag k Autocorrelation coefficient = Serial Correlation Coefficient
= Autocorrelation Function (ACF)



$r_0=1$
 $r_1 \approx \rho_1 = 1^{\text{st}} \text{ Serial Correlation Coefficient}$

Alternative for calculating r_k

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x}_t)(x_{t+k} - \bar{x}_{t+k})}{\sqrt{\sum_{t=1}^{N-k} (x_t - \bar{x}_t)^2 \sum_{t=1}^{N-k} (x_{t+k} - \bar{x}_{t+k})^2}} \quad [2.19]$$

Complex Characteristics of Periodic Time Series(Nonstationary)

$X_{v,t}$ = periodic time series

v = year ; τ = time interval within year, $\tau=1,..w$ for monthly time series $w=12$

Periodic or Seasonal statistics

$$\bar{x}_\tau = \frac{1}{N} \sum_{v=1}^N x_{v,\tau} \approx \mu_\tau; \tau = 1, \dots, w \quad [2.20]$$

$$s_\tau^2 = \frac{1}{N-1} \sum_{v=1}^N (x_{v,\tau} - \bar{x}_\tau)^2 \approx \sigma_\tau^2 \quad [2.21]$$

$$g_\tau = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau} - \bar{x}_\tau)^3}{(N-1)(N-2)s_\tau^3} \approx \gamma_\tau \quad [2.22]$$

$$r_{k,\tau} = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau} - \bar{x}_\tau)(x_{v,\tau-k} - \bar{x}_{\tau-k})}{s_\tau s_{\tau-k}} \approx \rho_{k,\tau} \quad [2.23]$$

If $\tau-k < 1$ i.e. $\tau-k=12-12=0$; $\omega=12$

$$r_{k,\tau} = \frac{\frac{1}{N-1} \sum_{v=1}^{N-1} (x_{v,\tau} - \bar{x}_\tau)(x_{v-1,\omega+\tau-k} - \bar{x}_{\omega+\tau-k})}{s_\tau s_{\omega+\tau-k}} \quad [2.24]$$

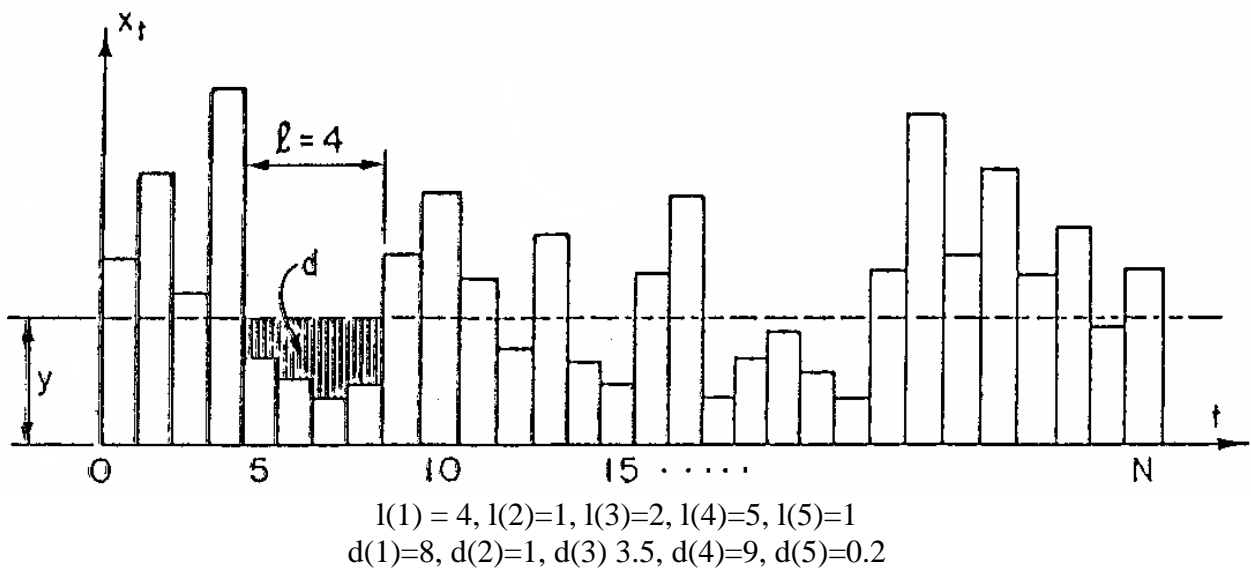
2.7 คุณสมบัติที่เกี่ยวกับภาวะความแห้งแล้งของอนุกรมเวลา (Drought Related Characteristics of Time Series)

กำหนดให้อนุกรมเวลาคือปริมาณน้ำต้นทุนในแต่ละเดือน $\{x_t: x_1, x_2, \dots, x_N\}$ และความต้องการน้ำรายเดือนซึ่งกำหนดให้มีค่าคงที่เท่ากับ y สถานการณ์น้ำในแต่ละเดือนจะแบ่งออกได้เป็น 2 แบบ คือ เดือนที่มีน้ำมากพอ (Surplus) และเดือนที่ขาดน้ำ (Deficit) ดังนี้

$$x_t < y ; \text{Deficit หรือ Negative Run}$$

$$x_t > y ; \text{Surplus หรือ Positive Run}$$

คุณสมบัติอนุกรมเวลาที่เกี่ยวข้องกับความแห้งแล้ง คือ Negative Run ซึ่งมีคุณสมบัติที่สำคัญ 2 ตัวคือ Run Length และ Run Sum ซึ่งสามารถหาได้ดังรูปที่ 2.15



รูปที่ 2.15 Drought Related Characteristics of Time Series (run-length and run-sum)

2.7.1 Run Length

Run Length ($l(j)$) หมายถึงจำนวนเดือนที่มีการขาดแคลนน้ำติดต่อกัน จากรูปที่ 2.15 จะเห็นได้ว่าการขาดน้ำ 5 ครั้ง คือ $l(1)=4, l(2)=1, l(3)=2, l(4)=5$ และ $l(5)=1$ Run Length เป็นตัวแปรสุ่ม ซึ่งมีค่าสถิติที่สำคัญคือ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐานและค่าพิสัยสูงสุด ซึ่งสามารถคำนวณหาได้จากสมการ

$$\bar{l}_N = \frac{1}{M} \sum_{j=1}^M l(j) = \text{Mean Run Length} \tag{2.25}$$

เมื่อ $j =$ run no. , the maximum run= M

$l(j)=$ Run Length j

$$s_N(l) = \sqrt{\frac{1}{M-1} \sum_{j=1}^m (l(j) - \bar{l}_N)^2} \tag{2.26}$$

$$l_N^* = \text{Max}(l(1), \dots, l(M)) \tag{2.27}$$

สำหรับค่า N และ y ใดๆ

$\bar{I}_N, s_N(l), I_N^*$ คือตัวแปรสุ่ม

ถ้า N และ y เพิ่มขึ้น จะมีผลทำให้ ค่าสถิติ $\bar{I}_N, s_N(l), I_N^*$ เพิ่มขึ้นด้วย

2.7.2 Run Sum, d(j)

Run Sum (d(j)) หมายถึงผลรวมปริมาณการขาดน้ำในช่วง Run Length (l(j)) ใดๆ จากรูปที่ 2.15 จะเห็นได้ว่าการขาดน้ำ 5 ครั้ง ปริมาณการขาดน้ำในแต่ละช่วงมีค่าเท่า d(1)=8, d(2)=1, d(3)=3.5, d(4)=9 และ d(5)=0.2 Run Sum เป็นตัวแปรสุ่มเหมือน Run Length มีค่าสถิติที่สำคัญคือ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐานและค่าพิสัยสูงสุด ค่า Run Sum และค่าสถิติสามารถคำนวณหาได้จากสมการ

$$d(j) = \sum_{t=t_j}^{t_j+l(j)-1} (y - X_t) \tag{2.28}$$

จากรูปที่ 2.15, ที่ j=1, t₁=5, l(1)=4

$$d(1) = \sum_{t=t_1}^{t_1+l(1)-1} (y - X_t)$$

$$d(1) = \sum_{t=5}^8 (y - X_t) = 8$$

$$\bar{d}_N = \frac{1}{M} \sum_{j=1}^M d(j) \tag{2.29}$$

$$S_N(d) = \sqrt{\frac{1}{M-1} \sum_{j=1}^M (d(j) - \bar{d}_N)^2} \tag{2.30}$$

$$d_N^* = \text{Max}(d(1), \dots, d(M)) \tag{2.31}$$

2.8 Storage Related Characteristics of Time Series

คุณสมบัติของอนุกรมเวลาที่เกี่ยวข้องกับการเก็บกักน้ำ (Storage) คือ Partial Sums ถ้ากำหนดให้ อนุกรมของน้ำทำคือ {x₁, x₂, , x_N} จะสามารถคำนวณหา Partial Sums(S_i)ได้จากสมการ

$$S_i = S_{i-1} + X_i - \bar{X}_N; i = 1, \dots, N \tag{2.32}$$

เมื่อ

$S_i = \text{Partial Sums (Use capital letter S)}$

$$S_0 = 0$$

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$$

$$s_N = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X}_N)^2} \quad (\text{Use small letter s})$$

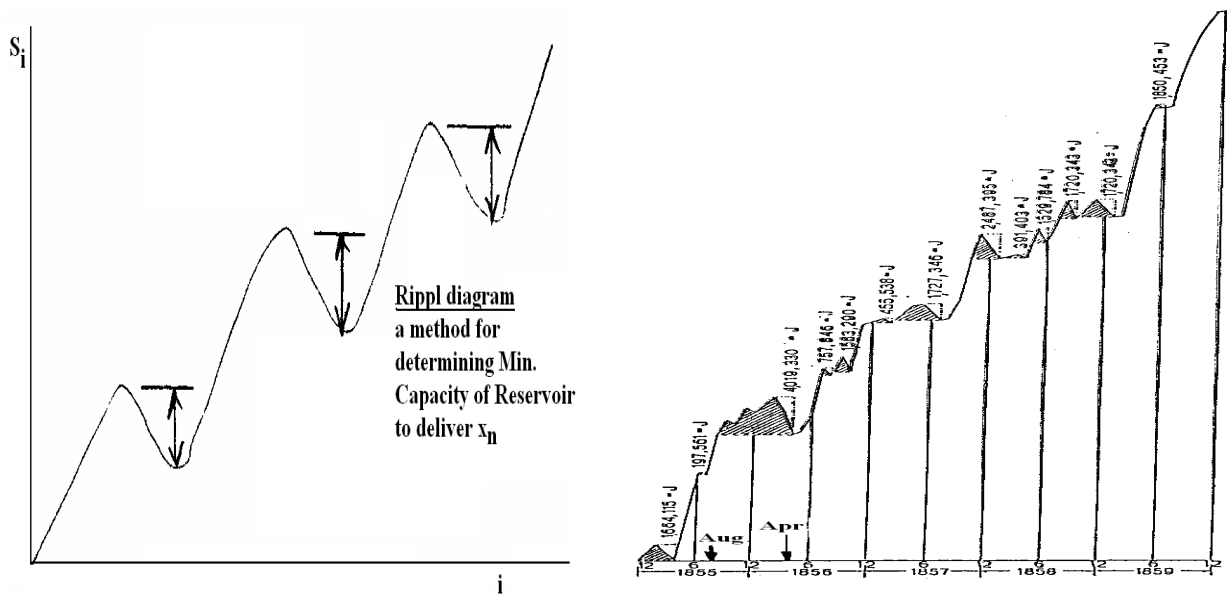
เมื่อทราบ S_i จะสามารถคำนวณหา Range of Partial Sums ได้จากสมการ

$$R_N = \frac{1}{S_N} [Max(S_0, S_1, \dots, S_N) - Min(S_0, S_1, \dots, S_N)] \quad [2.33]$$

เมื่อ

R_N = Range (Rescaled Range) of Partial Sums

แนวคิดของ Partial Sums จะสามารถนำไปใช้ในการขนาดของอ่างเก็บน้ำได้ เช่นวิธีหาขนาดอ่างเก็บน้ำของ Rippl ดังรูปที่ 2.16

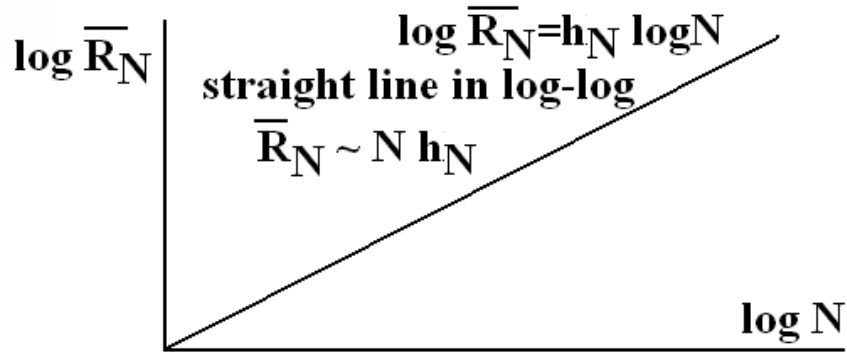


รูปที่ 2.16 Rippl Diagram for Determining Minimum Capacity of Reservoir to deliver \bar{X}_N

ถ้ามีอนุกรมเวลา M ชุด แต่ละชุดมีขนาด N จะสามารถคำนวณหาค่า $R_N(j)$ เมื่อ $j=1, \dots, M$ ได้ ถ้านำ $R_N(j)$ ไปหาค่าเฉลี่ย จะได้ ค่า Mean Range ดังสมการ

$$\text{Mean Range, } \bar{R}_N = \sum_{j=1}^M \frac{R_N(j)}{M} \quad [2.34]$$

เมื่อนำ \bar{R}_N ไปพล็อตเทียบกับ N ในกระดาษ Log-Log จะได้กราฟเส้นตรงที่เรียกว่า Hurst Phenomena ดังรูปที่ 2.17



Hurst Phenomena

ให้ h_N = ความลาดชันของเส้นตรง ซึ่งจะแปรผันตามค่า N ดังนี้

ถ้า $h_N > \frac{1}{2}$ อย่างรวดเร็วแสดงว่าอนุกรมเวลาเป็นแบบ Short Memory

ถ้า $h_N < \frac{1}{2}$ อย่างช้าๆแสดงว่าอนุกรมเวลาเป็นแบบ Long Memory

ถ้า $h_N = \frac{1}{2}$ แสดงว่าอนุกรมเวลาเป็นแบบ Infinite Memory

2.9 การทดสอบ Non-homogeneity และ Inconsistency ของอนุกรมเวลาทางอุทกวิทยา

Non-homogeneity และ Inconsistency ของอนุกรมเวลาทางอุทกวิทยา เช่น Trend หรือ Jump (Slippage) ทางสถิติสามารถทำได้ดังนี้

(1) Fitting Trend Function

$$X_t = b_0 + b_1 t + b_2 t^2 + \dots + b_m t^m \quad [2.35]$$

ทำการทดสอบว่า $b_0, b_1, \dots, b_m \neq 0$

(2) การทดสอบผลต่างระหว่างค่าเฉลี่ยโดยใช้ t-Test

$$H_0: \mu_1 - \mu_2 = 0 \quad [2.36]$$

$$H_1: \mu_1 - \mu_2 \neq 0 \quad [2.37]$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_P \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad [2.38]$$

$$S_P = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}} \quad [2.39]$$

$$\text{ถ้า } |t| > t_{\alpha/2, (N_1 + N_2 - 2)} \quad \dots \dots \dots \text{Reject } H_0 \quad [2.40]$$

(3) การทดสอบสัดส่วนค่าความแปรปรวนโดยใช้ f-Test ดังรูปที่ 2.18

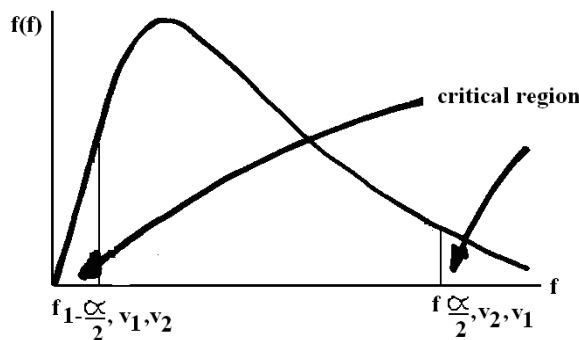
$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \tag{2.41}$$

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \tag{2.42}$$

$$f = \frac{S_1^2}{S_2^2} \tag{2.43}$$

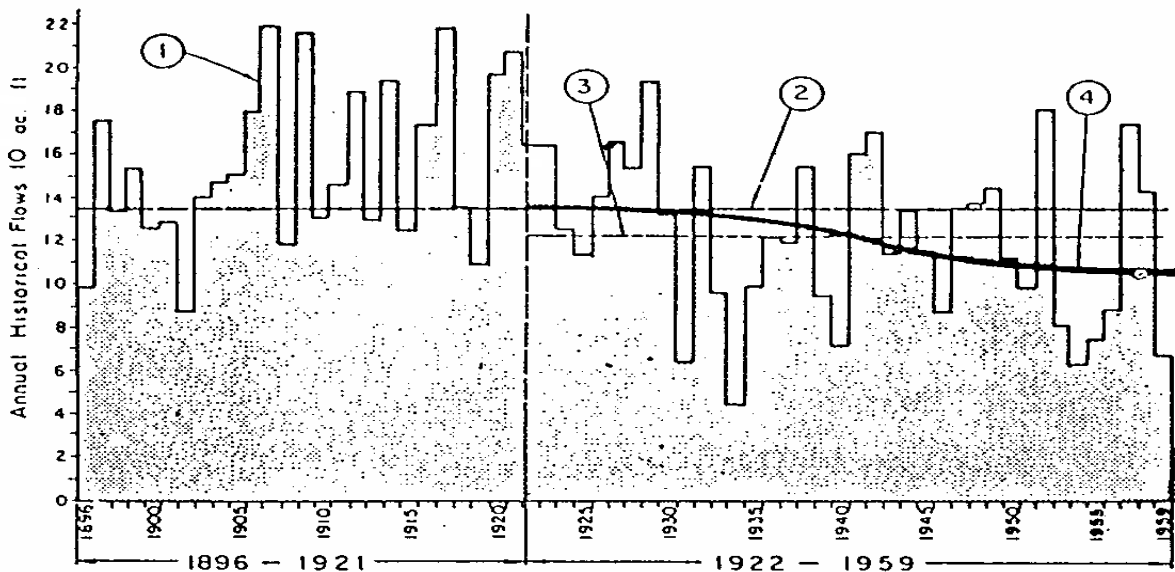
$$f > f_{\frac{\alpha}{2}, v_1, v_2} \quad \dots\dots\dots \text{Reject } H_0 \tag{2.44}$$

$$f < f_{1-\frac{\alpha}{2}, v_1, v_2} \quad \text{or} \quad \frac{1}{f_{\frac{\alpha}{2}, v_2, v_1}} \quad \dots\dots\dots \text{Accept } H_0 \tag{2.45}$$



รูปที่ 2.18 f-test

อนุกรมเวลาของปริมาณการไหลของน้ำรายปีของแม่น้ำโคโลราโด ที่ Lee Ferry ระหว่าง 1856-1959 แสดงการเกิด Jump และ Trend ดังรูปที่ 2.19 ตัวอย่างการใช้ t-Test ทดสอบการเปลี่ยนแปลงค่าเฉลี่ยของ Net Basin Supply (NBS) รายปีของ Great Lakes แสดงอยู่ในตารางที่ 2.3



รูปที่ 2.19 Annual flow series of Colorado river at Lee Ferry, Arizona, 1896-1959 (1) historical annual flow record with 1856-1921 period data estimated from other stations by correlation (2) arithmetic mean for period 1896-1959 (64 years) (3) arithmetic mean for historical observed period 1922-1959 and (4) approximate trend in the period 1922-1959 (after Yevjevich, 1972)

ตารางที่ 2.3 Example of Testing for Changes in the Mean of the Annual NBS of the Great Lakes System.

Lake	Subsample Sizes		Statistic From t-Tables (95%)	t Computed	Change in the Mean
	N ₁	N ₂			
Ontario	36	33	2.0	0.299	No
Erie	36	33	2.0	0.635	No
Superior	36	33	2.0	1.525	No
Michigan	36	33	2.0	0.866	No
St. Clair	43	26	2.0	4.477	Yes

2.10 คุณสมบัติของอนุกรมเวลารายปี (Characteristics of Annual Time Series)

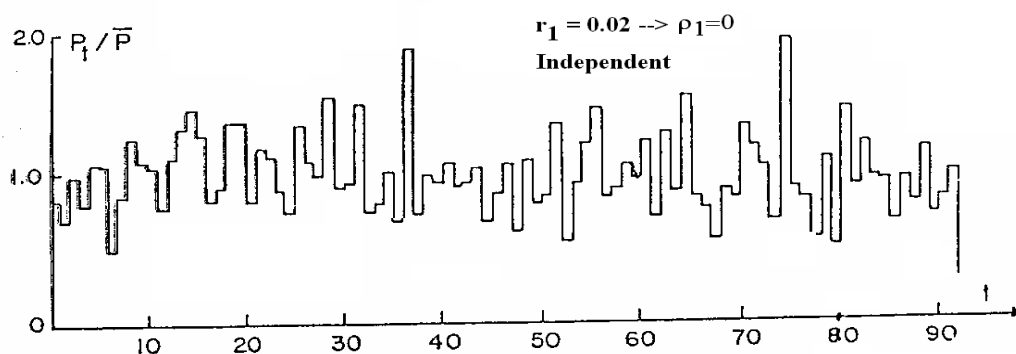
ผลการศึกษาอนุกรมเวลาของฝนรายปีจาก 1411 สถานีในสหรัฐอเมริกาพบว่าอนุกรมเวลาทั้งหมดมีลักษณะเป็น Independent Stationary Process

Independent หมายความว่าฝนปีหนึ่งไม่ได้ขึ้นอยู่กับฝนปีก่อนหน้านั้น หรือ Correlogram แสดงว่าไม่มี Serial Correlation

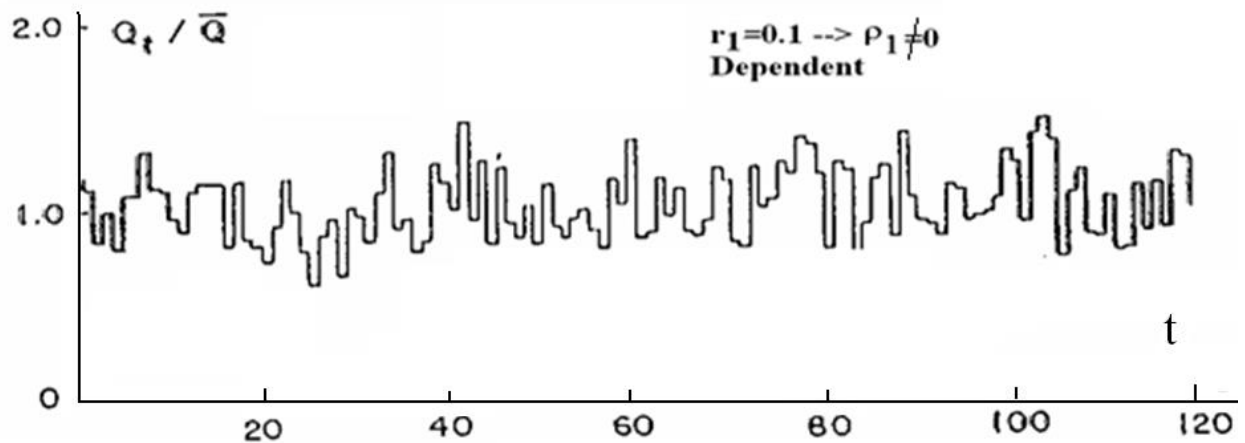
Stationary หมายความว่าพารามิเตอร์ของอนุกรมเวลาไม่เปลี่ยนแปลงไปตาม

รูปที่ 2.20 แสดงอนุกรมเวลาของฝนรายปีที่ Ft. Collins โคโลราโดมีลักษณะเป็นอนุกรมเวลาแบบ Independent เนื่องจาก r_1 มีค่าน้อยมาก ผลการทดสอบเชื่อได้ว่า $\rho_1 = 0$

รูปที่ 2.21 แสดงอนุกรมเวลาของน้ำท่ารายปีของแม่น้ำ Danube ที่ Orsova รูมาเนีย ซึ่งมีลักษณะเป็นอนุกรมเวลาแบบ Dependent เนื่องจาก r_1 มีค่าเท่ากับ 0.1 ผลการทดสอบเชื่อได้ว่า $\rho_1 \neq 0$



รูปที่ 2.20 Annual precipitation series at Ft. Collins, Colorado, for period 1887-1978 (92 years), in modular coefficients P_t/\bar{P} with P_t = annual values and $\bar{P} = 14.57$ inches = annual mean.



รูปที่ 2.21 Annual runoff series of Danube river at Orsova, Romania, for 120 years (1837-1957) in modular coefficients, Q_t / \bar{Q} with Q_t = annual value and \bar{Q} = annual mean.

วิธีการทดสอบความเป็นอิสระ (Test of Independence)

- (1) Correlogram (Time Domain)
- (2) Spectral Density (Frequency Domain)

2.10.1 Correlogram Test

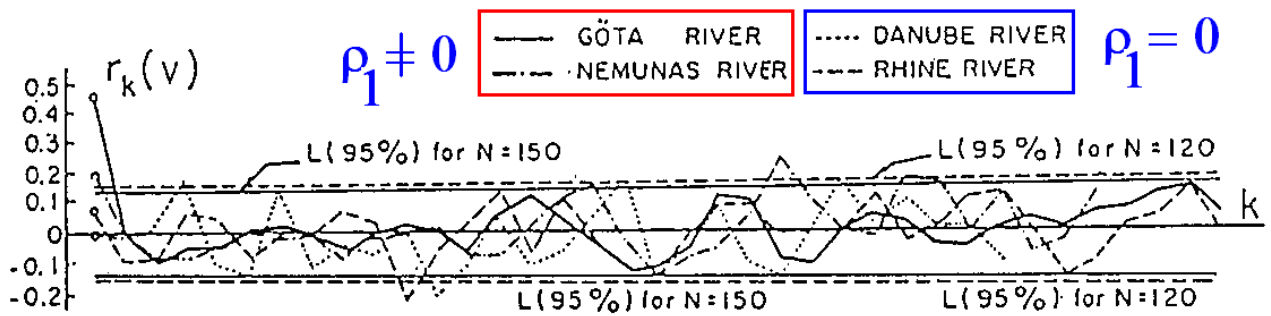
ถ้า Correlogram ที่ $k > 0$ มีค่าอยู่นอก Probability Limit ดังสมการ 2.46 และ 2.47 แสดงว่าเป็นอนุกรมเวลาแบบ Independent ในทางตรงกันข้าม ถ้า Correlogram ที่ $k > 0$ มีค่าอยู่ใน Probability Limit แสดงว่าเป็นอนุกรมเวลาแบบ Dependent

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k} \quad [2.46]$$

$$r_k(99\%) = \frac{-1 \pm 2.326\sqrt{N-k-1}}{N-k} \quad [2.47]$$

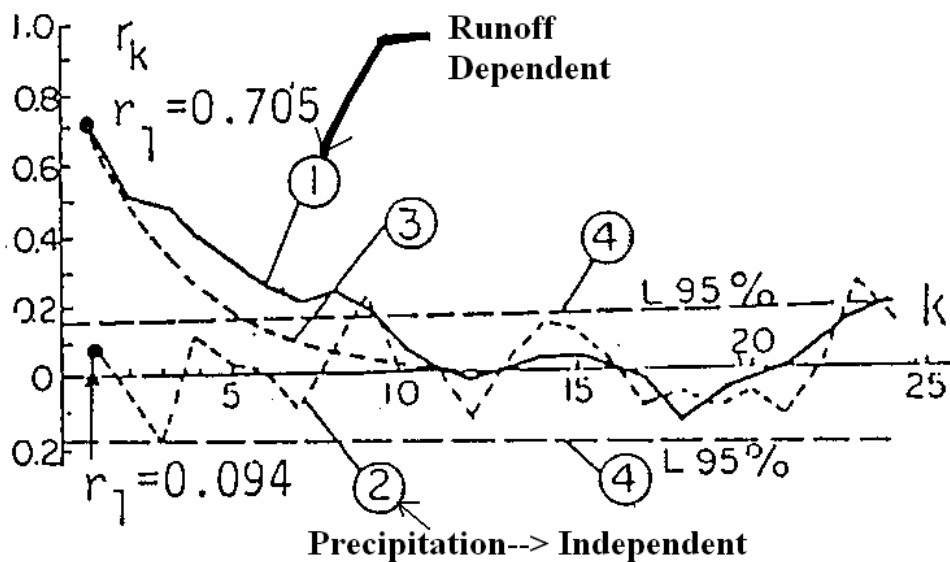
ตัวอย่าง Correlogram ของอนุกรมเวลาของน้ำท่ารายปีของแม่น้ำ 4 สายในยุโรป คือ แม่น้ำ Gota ในสวีเดน แม่น้ำ Nemunas ใน Lithuania แม่น้ำ Rhine ใน Basle, Switzerland และแม่น้ำ Danube ใน Orsova โรมาเนีย แสดงอยู่ในรูปที่ 2.22 พบว่าอนุกรมเวลาของน้ำท่ารายปีของแม่น้ำ Rhine และ Danube เป็นอนุกรมเวลาแบบ Independent ขณะที่อนุกรมเวลาของน้ำท่ารายปีของแม่น้ำ Gota และ Nemunas เป็นอนุกรมเวลาแบบ Dependent

$r_k < 5\%$ is out of limits



รูปที่ 2.22 Correlograms of annual runoff series of 4 European rivers; Gota river in Sweden (N=150), Nemunas river in Lithuania (N=132), Rhine river at Basle, Switzerland (N=150), Danube river at Orsova, Romania (N=120). Probability limits at the 95% level are given for the normal independent variables for two lengths, N=150(max) and N=120 (min), (Yevjavich, 1964)

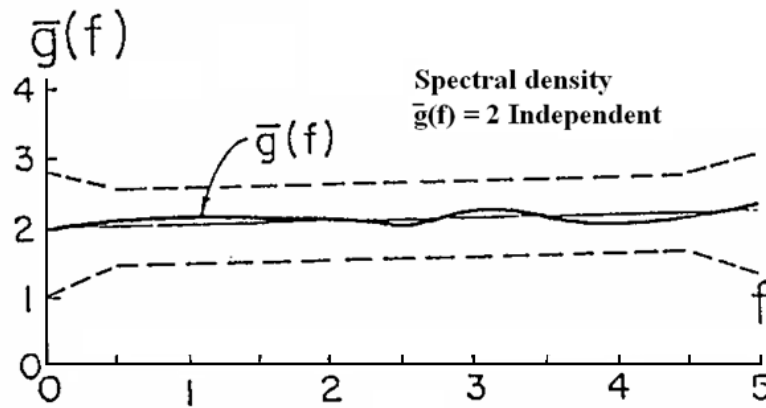
รูปที่ 2.23 แสดง Correlogram ของอนุกรมเวลาของน้ำท่ารายปี ฝนรายปี 1st Autoregressive Model ของน้ำท่า และ 95% Probability Limits ซึ่งแสดงให้เห็นว่าอนุกรมเวลาของน้ำท่าเป็น Dependent ขณะที่อนุกรมเวลาของฝนเป็น Independent



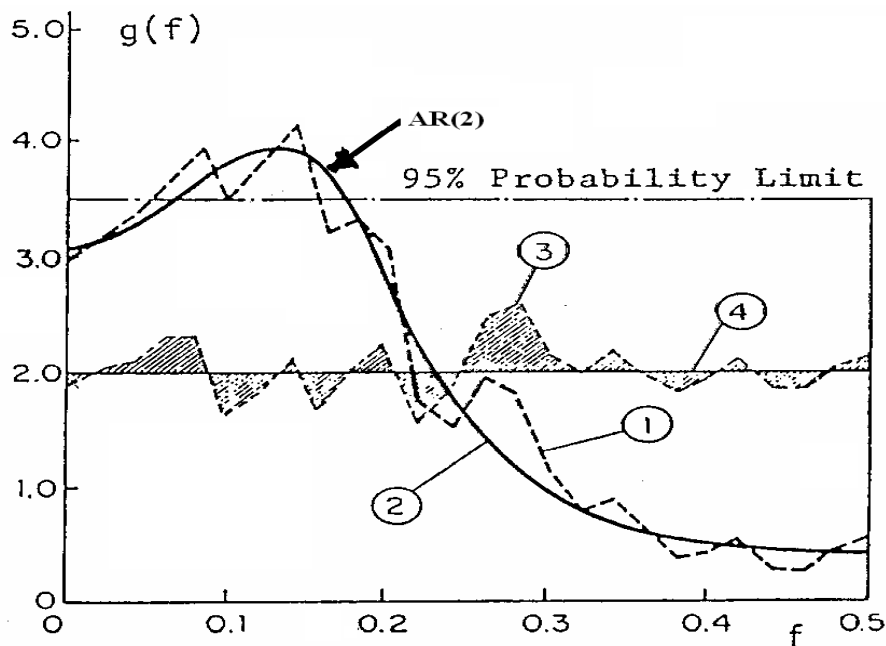
รูปที่ 2.23 Correlograms of St.Lawrence river at Ogdensburg, NY: (1) r_k of annual runoff series, (2) r_k of annual effective precipitation series, (3) correlogram of first order autoregressive model, (4) probability limits at 95% level for normal independent variables with N=97.

รูปที่ 2.24 แสดง Spectral Density ของอนุกรมเวลาของฝนรายปีในภาคตะวันออกเฉียงเหนือของประเทศสหรัฐอเมริกา ซึ่งมีค่า Spectral Density ประมาณ 2 แสดงว่าเป็นอนุกรมเวลาแบบบิสระ รูปที่ 2.25 แสดง Spectral

Density ของอนุกรมเวลาของน้ำท่ารายปีของแม่น้ำ Gota ใน Vanersburg ประเทศ Sweden เปรียบเทียบกับ Fitted Spectral Density ของแบบจำลอง AR(2) Spectral Density ของ residuals และ Expected spectrum ของอนุกรมเวลาแบบอิสระ



รูปที่ 2.24 Average variance spectrum of 231 annual (homogeneous) precipitation series of northwest area of USA (area between longitudes 94 and 85 and latitude 63.5 and Canadian-USA border).



รูปที่ 2.25 Spectra of annual flow series of Gota river at Vanersburg, Sweden, for 150 years; (1) estimated spectrum (2) fitted spectrum of second autoregressive model (3) spectrum of residuals and (4) expected spectrum of independent series.

2.10.2 สรุปคุณสมบัติของอนุกรมเวลารายปี (Summary of Annual Series Characteristics)

- (1) อนุกรมเวลารายปีของฝน น้ำท่า และอัตราการระเหยถือว่าเป็นเกิดจากกระบวนการสโตแคสติก
- (2) อนุกรมเวลารายปีมีโครงสร้างแบบ Dependent เนื่องจากกระบวนการธรณีฟิสิกส์ของน้ำในลุ่มน้ำ

- (3) อนุกรมเวลายิ่งระยะยาว (Long Time Series) ยิ่งจะมี Non-homogeneity เกิดขึ้นทั้งนี้เนื่องจากคนและธรรมชาติ และยังมี Systematic Error หรือ Inconsistency เกิดมากขึ้น
- (4) อาจมีการเปลี่ยนแปลงคุณสมบัติทางสถิติของอนุกรมเวลาซึ่งไม่ได้เกิดจาก Non-homogeneity และ Inconsistency แต่เกิดจากการเปลี่ยนแปลงทางภูมิศาสตร์ในพื้นที่หรือของภูมิภาคนั้นๆ

2.11 คุณสมบัติของอนุกรมเวลาแบบ Periodic

อนุกรมเวลาแบบ Periodic เช่นรายฤดูกาล รายเดือน รายสัปดาห์และรายวัน อาจมีคุณสมบัติทางสถิติคล้ายหรือต่างจากอนุกรมเวลารายปีได้ ส่วนที่ต่างคืออนุกรมเวลาแบบ Periodic จะมีค่าสถิติแบบ Periodic ส่วนที่เหมือนคือโครงสร้างแบบ Dependent ซึ่งมีค่าคงที่ $(\rho_{k,\tau})$ ดังนั้นแบบจำลอง AR จึงสามารถทำได้ทั้งแบบ Constant Coefficients และแบบ Periodic Coefficients

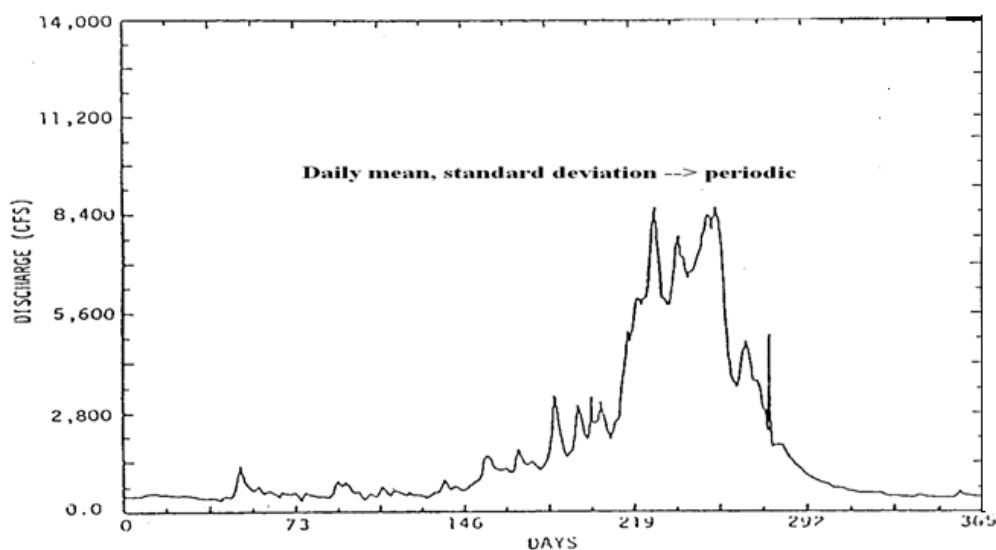
ถ้า $x_{v,\tau}$ คืออนุกรมเวลาแบบ Periodic สามารถแปลงเป็นอนุกรมเวลาแบบมาตรฐานได้จากสมการ 2.48

$$Z_{v,\tau} = \frac{x_{v,\tau} - \bar{x}_\tau}{S_\tau} \quad [2.48]$$

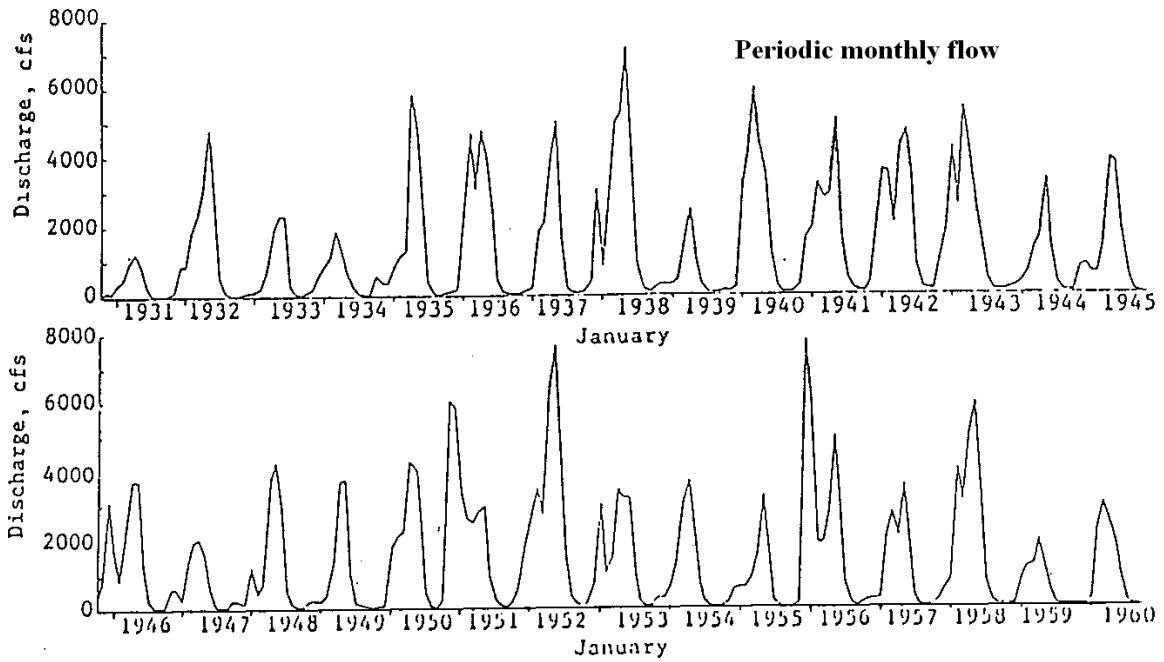
เมื่อ $z_{v,\tau}$ คือ Standardized Series ที่กำจัดค่า \bar{x}_τ และ S_τ แล้ว

ค่า Periodic Skewness สามารถกำจัดได้โดยการ Taking Logarithmic

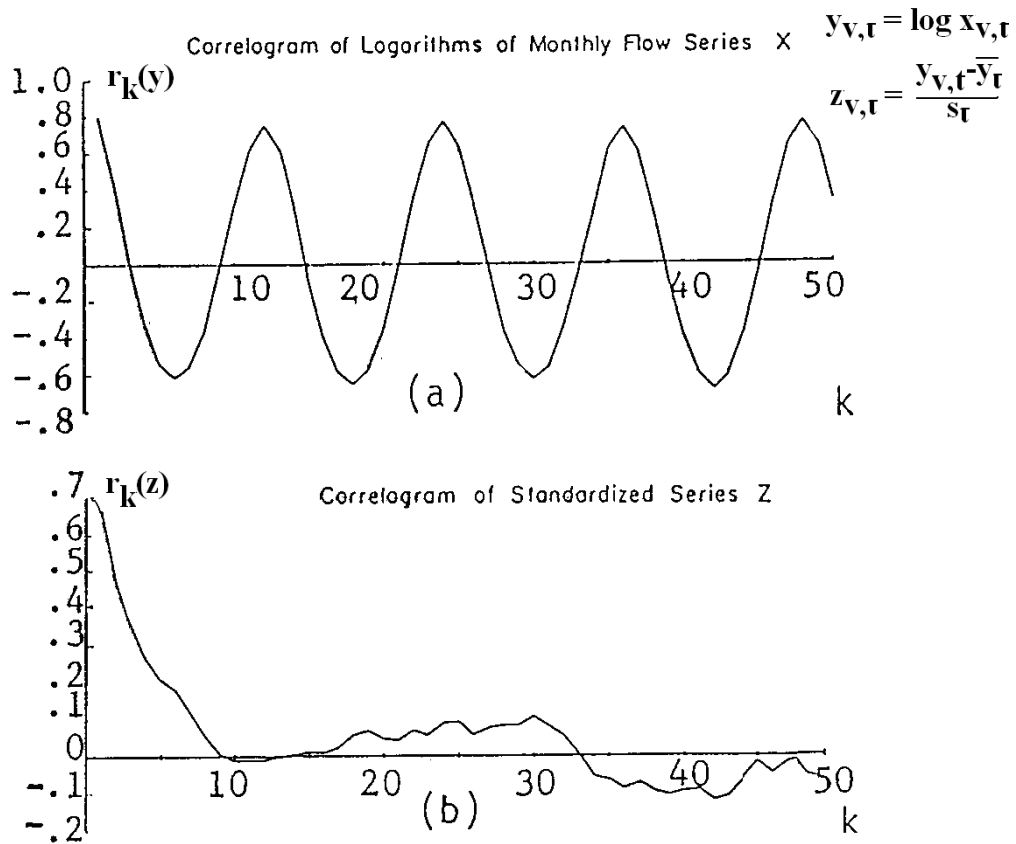
ตัวอย่างอนุกรมเวลาแบบ Periodic รายวันของอัตราการไหลของน้ำในแม่น้ำ Boise รัฐ Idaho แสดงอยู่ในรูปที่ 2.26 อนุกรมเวลาของอัตราการไหลของน้ำรายเดือนที่สถานี 11B.402 Middle Fork ของ American River ใกล้เมือง Auburn รัฐ California แสดงในรูปที่ 2.27 ส่วน Correlograms ของ Logarithms ของ Original Monthly Flow (y_t Series) และ ของ z_t Series ของสถานี 11B.402 แสดงอยู่ในรูปที่ 2.28



รูปที่ 2.26 Daily flows of Boise river, Idaho, USA for year 1921.

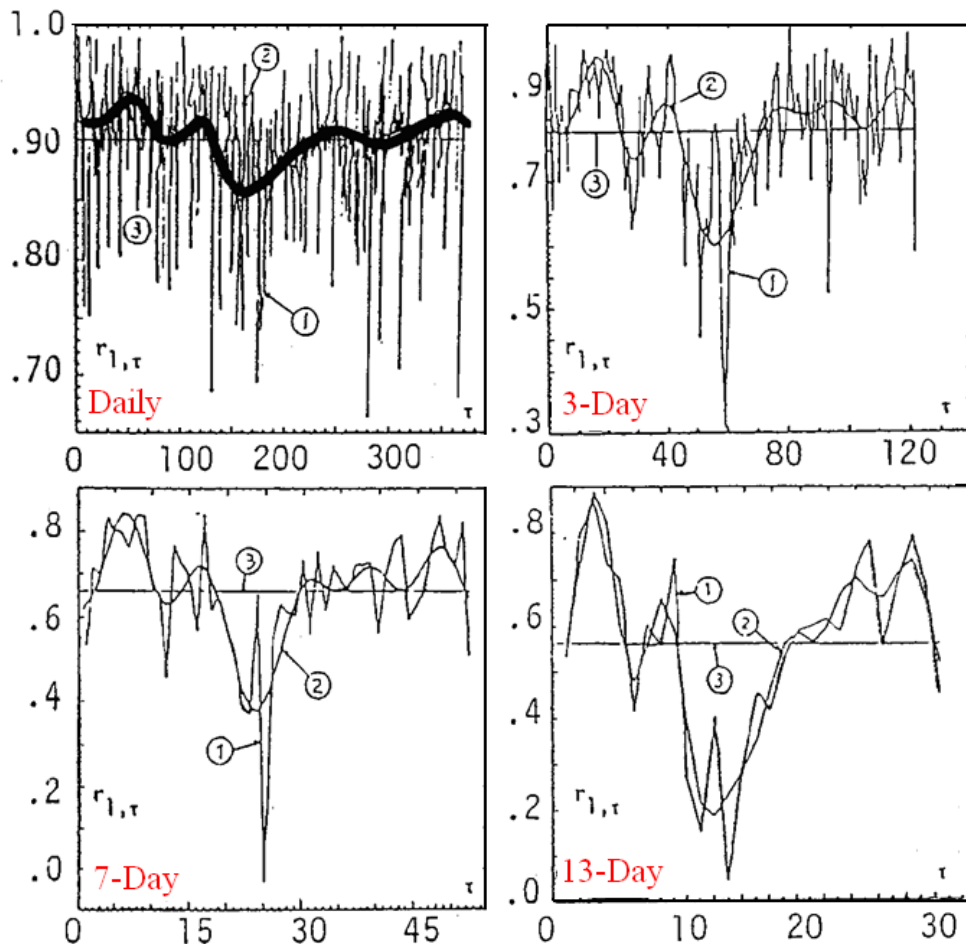


รูปที่ 2.27 Monthly river flows for station 11B.402, Middle Fork of American river near Auburn, California for period 1931-1960 (Roesner and Yevjevich, 1966)



รูปที่ 2.28 Correlograms of (a) logarithms of original monthly river flows (b) series on log-domain after removing periodic mean and periodic standard deviation for station 11B.402, Middle Fork of American river near Auburn, California (Roesner and Yevjevich, 1966)

Periodic Correlograms ของอนุกรมรายวัน ราย 3 วัน ราย 7 วัน และราย 13 วัน ของ Log Series ของแม่น้ำ Tioga แสดงอยู่ในรูปที่ 2.29



รูปที่ 2.29 Periodic correlograms of daily, 3-day, 7-day and 13-day series for log-transformed flows of Tioga river (1) $r(1,t)$ computed from eq. 2.11, (2) smoothed $r(1,t)$ using Fourier (trigonometric) series and (3) mean of $r(1,t)$ (Tao et al., 1976)

2.12 คุณสมบัติของอนุกรมเวลาแบบ Multivariate (Characteristics of Multivariate Time Series)

Lag k Autocorrelation และ Lag k Cross Correlation ของอนุกรมเวลาแบบ Multivariate สามารถคำนวณได้จากสมการ 2.49

$$r_k^{ij} = \frac{\sum_{t=1}^{N-k} (x_t^{(i)} - \bar{x}_t^{(i)})(x_{t+k}^{(j)} - \bar{x}_{t+k}^{(j)})}{\sqrt{(x_t^{(i)} - \bar{x}_t^{(i)})^2 (x_{t+k}^{(j)} - \bar{x}_{t+k}^{(j)})^2}} \quad [2.49]$$

Correlation Matrix ของ n อนุกรมเวลา แสดงอยู่ในสมการ 2.50

$$\hat{M}_k = \begin{bmatrix} r_k^{11} & r_k^{12} & \dots & r_k^{1n} \\ r_k^{21} & r_k^{22} & \dots & r_k^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_k^{n1} & r_k^{n2} & \dots & r_k^{nn} \end{bmatrix} \quad [2.50]$$

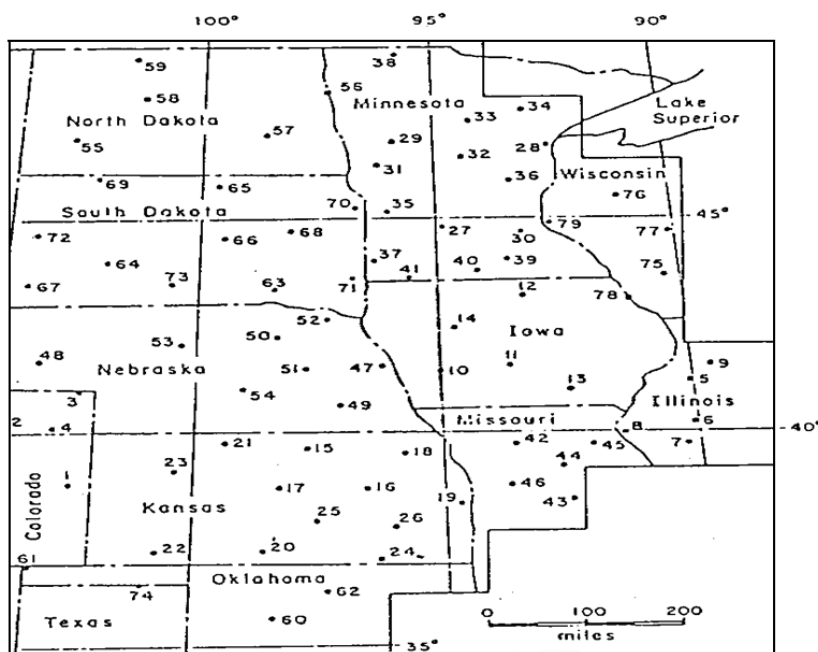
Periodic Lag k Cross-correlation สามารถคำนวณได้จากสมการ 2.51 และ 2.52

$$r_{k,\tau}^{ij} = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau}^{(i)} - \bar{x}_\tau^{(i)})(x_{v,\tau-k}^{(j)} - \bar{x}_{\tau-k}^{(j)})}{s_\tau^{(i)} s_{\tau-k}^{(j)}} \quad [2.51]$$

ถ้า $\tau - k < 1$

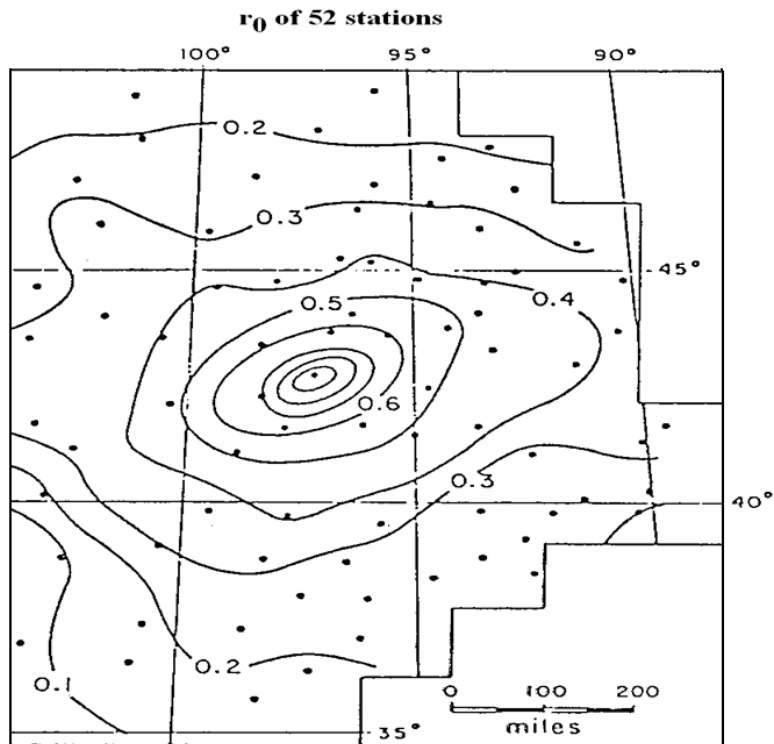
$$r_{k,\tau}^{ij} = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau}^{(i)} - \bar{x}_\tau^{(i)})(x_{v-1,\omega+\tau-k}^{(j)} - \bar{x}_{\omega+\tau-k}^{(j)})}{s_\tau^{(i)} s_{\omega+\tau-k}^{(j)}} \quad [2.52]$$

จากข้อมูลฝน 79 สถานีใน Upper Great Plains ของสหรัฐอเมริกา ในรูปที่ 2.30 สามารถนำมาสร้าง Isocorrelogram ของ 52 สถานีซึ่งมีความสัมพันธ์กับสถานีอื่นๆ ได้ดังรูปที่ 2.31 กราฟแสดงความสัมพันธ์ระหว่าง Lag-zero Cross Correlation Coefficient (r_0) กับ Interstation Distance (d) แสดงอยู่ในรูปที่ 3.32 และ Isolines ของฝนรายเดือนที่สังเคราะห์ขึ้นใหม่ 30 ปี แสดงอยู่ในรูปที่ 2.33

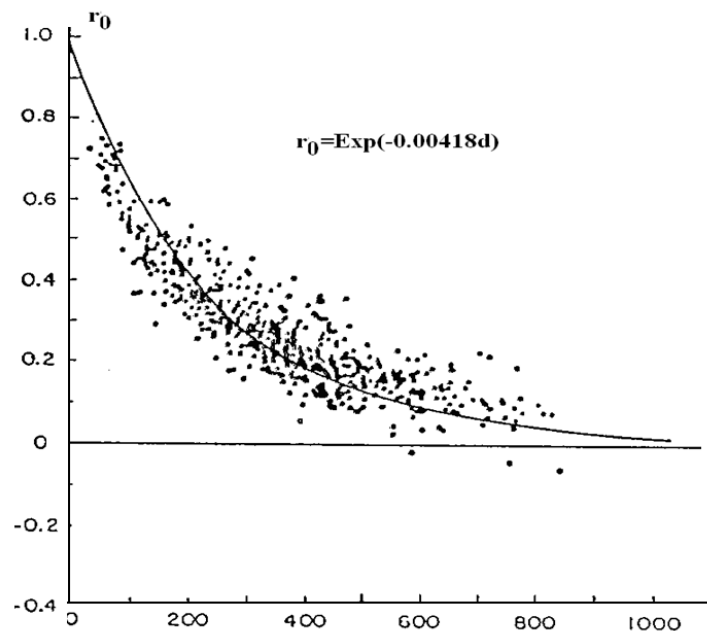


- $r_0 = \exp(-0.00418d)$
- mean, standard deviation of multivariate varies with latitude, longitude, altitude)
- regional characteristics

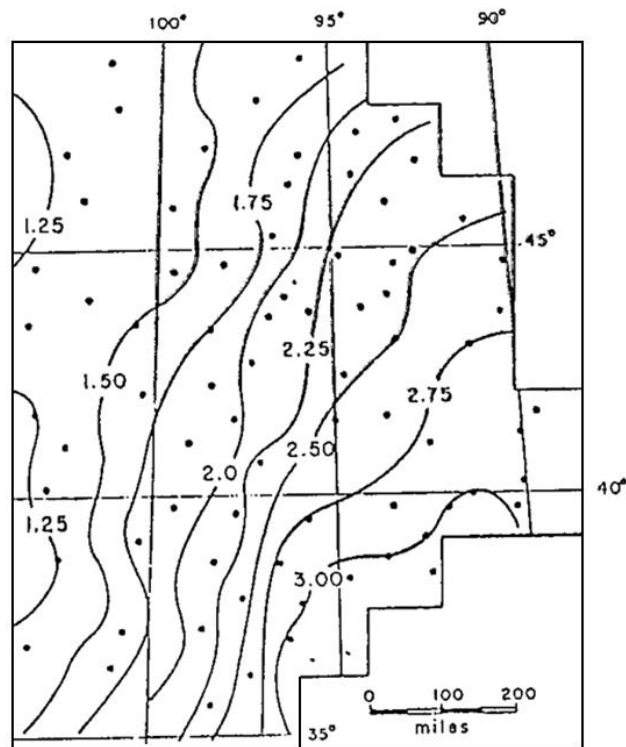
รูปที่ 2.30 Study area and location of 79 stations of precipitation in Upper Great Plains of USA (after Tase, 1976)



รูปที่ 2.31 Isocorrelogram patterns for series of station 52, as correlated with all other station series for residual independent series, monthly precipitation in upper Great plain of USA (after Tase, 1976)



รูปที่ 2.32 Lag-zero cross correlation coefficient r_0 versus interstation distance (d) and fitted function $r = \exp(-0.00418d)$ for stochastic component of monthly precipitation in Upper Great Plains of USA (after Tase, 1976)



รูปที่ 2.33 Isolines of 30 year generation monthly mean for precipitation of Upper Great Plains.

2.13 เอกสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series.**
 Water Resources Publications. USA. pp.484

2.14 แบบฝึกหัดท้ายบท

From the given periodic time series $x(t)$ for $t=1$ to 144 in Table 2.4, you are assigned to do the following tasks:

- (1) determine Autocorrelation Function (ACF) for the maximum lag of 60.
- (2) determine the Spectral Density Function (SDF).
- (3) identify the periodicity in the time series from ACF and SDF.
- (4) remove the periodicity from the given time series and determine ACF and SDF after removing the periodicity.
- (5) Plot ACF and SDF for all cases.

Table 2.4 Periodicity time series with 144 periods, $x(t)$

Year	Month	t	$x(t)$
1	1	1	21.8320
	2	2	22.1980
	3	3	22.9051
	4	4	21.9051
	5	5	21.2979
	6	6	22.2979
	7	7	21.1730
	8	8	19.1730
	9	9	18.4659

	10	10	17.0999
	11	11	17.0000
	12	12	19.0000
2	1	13	18.9001
	2	14	17.5341
	3	15	16.8270
	4	16	14.8270
	5	17	13.7021
	6	18	14.7021
	7	19	14.0949
	8	20	13.0949
	9	21	13.8020
	10	22	14.1680
	11	23	16.0000
	12	24	20.0000
3	1	25	21.8320
	2	26	22.1980
	3	27	22.9051
	4	28	21.9051
	5	29	21.2979
	6	30	22.2979
	7	31	21.1730
	8	32	19.1730
	9	33	18.4659
	10	34	17.0999
	11	35	17.0000
	12	36	19.0000
4	1	37	18.9001
	2	38	17.5341
	3	39	16.8270
	4	40	14.8270
	5	41	13.7021
	6	42	14.7021
	7	43	14.0949
	8	44	13.0949
	9	45	13.8020
	10	46	14.1680
	11	47	16.0000
	12	48	20.0000
5	1	49	21.8320
	2	50	22.1980
	3	51	22.9051
	4	52	21.9051
	5	53	21.2979
	6	54	22.2979
	7	55	21.1730
	8	56	19.1730
	9	57	18.4659
	10	58	17.0999

	11	59	17.0000
	12	60	19.0000
6	1	61	18.9001
	2	62	17.5341
	3	63	16.8270
	4	64	14.8270
	5	65	13.7021
	6	66	14.7021
	7	67	14.0949
	8	68	13.0949
	9	69	13.8020
	10	70	14.1680
	11	71	16.0000
	12	72	20.0000
7	1	73	21.8320
	2	74	22.1980
	3	75	22.9051
	4	76	21.9051
	5	77	21.2979
	6	78	22.2979
	7	79	21.1730
	8	80	19.1730
	9	81	18.4659
	10	82	17.0999
	11	83	17.0000
	12	84	19.0000
8	1	85	18.9001
	2	86	17.5341
	3	87	16.8270
	4	88	14.8270
	5	89	13.7021
	6	90	14.7021
	7	91	14.0949
	8	92	13.0949
	9	93	13.8020
	10	94	14.1680
	11	95	16.0000
	12	96	20.0000
9	1	97	21.8320
	2	98	22.1980
	3	99	22.9051
	4	100	21.9051
	5	101	21.2979
	6	102	22.2979
	7	103	21.1730
	8	104	19.1730
	9	105	18.4659
	10	106	17.0999
	11	107	17.0000

	12	108	19.0000
10	1	109	18.9001
	2	110	17.5341
	3	111	16.8270
	4	112	14.8270
	5	113	13.7021
	6	114	14.7021
	7	115	14.0949
	8	116	13.0949
	9	117	13.8020
	10	118	14.1680
	11	119	16.0000
	12	120	20.0000
11	1	121	21.8320
	2	122	22.1980
	3	123	22.9051

หลักการและเทคนิคทางสถิติสำหรับการจำลองอนุกรมเวลา

STATISTIC PRINCIPLES AND TECHNIQUES FOR TIME SERIES MODELING

เนื้อหาที่จะกล่าวถึงในบทที่ 3

- การประมาณค่าพารามิเตอร์ของแบบจำลอง (Estimating Model Parameters)
- การแปลงอนุกรมเวลาให้มีการแจกแจงแบบปกติ (Transforming Skewed Variables into Normal Variables)
- การประมาณค่าพารามิเตอร์แบบ Periodic ด้วย Fourier Series (Estimating Periodic Parameters by Fourier Series)
- การใช้เมทริกซ์ในการประมาณพารามิเตอร์ของแบบจำลองแบบหลายตัวแปร (Matrix Approach for Estimating Parameter of Multivariate Models)
- การทดสอบความเหมาะสมของแบบจำลอง (Goodness of Fit Test of Model)
- การรักษาคุณสมบัติทางสถิติของตัวอย่างและ การจำกัดจำนวนพารามิเตอร์ (Preservation of Historical Statistics and Parsimony of Parameters)
- การสังเคราะห์ข้อมูลใหม่และการพยากรณ์ (Generation and Forecasting)

3.1 เทคนิคการประมาณค่าพารามิเตอร์เบื้องต้น (Basic Estimation Techniques)

เทคนิคการประมาณค่าพารามิเตอร์เบื้องต้นมี 3 วิธีคือ

- (1) Method of Moment
- (2) Method of Least Square
- (3) Method of Maximum Likelihood

3.1.1 วิธีการตรวจสอบความเหมาะสมของการประมาณค่าพารามิเตอร์ (Goodness of Fit of Estimators)

การตรวจสอบความเหมาะสมในการประมาณค่าพารามิเตอร์ จะพิจารณาจากเกณฑ์ดังต่อไปนี้

- (1) Bias or Unbias
- (2) Efficiency; Efficiency Estimator คือค่าประมาณที่มีค่า Mean Square Error (MSE) น้อยที่สุด หรือ Minimum MSE)

- (3) Consistency หมายความว่าถ้า N เพิ่มขึ้น ค่าประมาณจะเข้าใกล้ค่าพารามิเตอร์มากขึ้น หรือ $\hat{\alpha}$ มีค่าเข้าใกล้ α
- (4) Sufficiency หมายความว่ามีการใช้ข้อมูลมากที่สุด (Maximum Use of the Information Contained in the Data)

3.1.2 ค่าประมาณแบบ Unbiased (Unbiased Estimator)

ให้ $\hat{\alpha}$ = ตัวประมาณค่า (Estimator)

α = พารามิเตอร์ (Parameters)

ถ้า $E(\hat{\alpha}) = \alpha$ แสดงว่า $\hat{\alpha}$ คือ Unbiased Estimator

ถ้า $E(\hat{\alpha}) \neq \alpha$ แสดงว่า $\hat{\alpha}$ คือ Biased estimator

$$Bias = \alpha - E(\hat{\alpha}) \quad [3.1]$$

3.1.3 Efficient Estimator

ค่าประมาณที่มีประสิทธิภาพหมายถึงค่าประมาณที่มีค่า Mean Square Error (MSE) ต่ำสุด จากสมการที่ 3.1

$$Bias = \alpha - E(\hat{\alpha})$$

$$Estimation Error = \alpha - \hat{\alpha} \quad [3.2]$$

$$\begin{aligned} MSE &= E(\alpha - \hat{\alpha})^2 \\ &= E[(\alpha - E(\hat{\alpha})) - (\hat{\alpha} - E(\hat{\alpha}))]^2 \\ &= E(\alpha - E(\hat{\alpha}))^2 + E(\hat{\alpha} - E(\hat{\alpha}))^2 - 2E[(\alpha - E(\hat{\alpha}))(\hat{\alpha} - E(\hat{\alpha}))] \\ &= (\alpha - E(\hat{\alpha}))^2 + Var(\hat{\alpha}) - 2(\alpha - E(\hat{\alpha}))E(\hat{\alpha} - E(\hat{\alpha})); E(\hat{\alpha} - E(\hat{\alpha})) = 0 \\ &= Var(\hat{\alpha}) + (\alpha - E(\hat{\alpha}))^2 \\ &= Var(\hat{\alpha}) + Bias^2 \end{aligned} \quad [3.3]$$

∴ MSE จะมีค่าต่ำสุด ถ้า เป็น Unbiased Estimator และ Variance ของ $\hat{\alpha}$ มีค่าต่ำสุด

3.2 Method of Moments

ถ้า X = ตัวแปรสุ่ม ซึ่งมีฟังก์ชันการแจกแจงเท่ากับ $f(x)$

$$\begin{aligned} E(X^r) &= r^{\text{th}} \text{ Population Moment of } X \\ &= \int x^r f(x) dx \end{aligned} \quad [3.4]$$

$$\begin{aligned} M_r &= r^{\text{th}} \text{ Sample Moment of } X \\ &= \frac{1}{N} \sum_{i=1}^N x_i^r \end{aligned} \quad [3.5]$$

หาพารามิเตอร์โดยให้สมการ 3.4 เท่ากับสมการ 3.5 หรือ $M_r = E(X^r)$

ยกตัวอย่างแบบจำลองอย่างง่าย

$$x_t = a + bz_t \tag{3.6}$$

$z_t =$ Independent Random Variable with $E(Z) = 0$ and $E(Z^2)=1$

$$E(X) = E(a+bZ) = a \tag{3.7}$$

$$E(X^2) = E(a+bZ)^2 = E(a^2+2abZ+b^2Z^2) = a^2+b^2 \tag{3.8}$$

$$M_1 = \frac{1}{N} \sum_{i=1}^N X_i$$

$$M_2 = \frac{1}{N} \sum_{i=1}^N X_i^2 = \hat{a}^2 + \hat{b}^2$$

$$\hat{a} = \frac{1}{N} \sum X_i ; \text{ เนื่องจาก } E(X) = a \tag{3.9}$$

$$\hat{a}^2 + \hat{b}^2 = \frac{1}{N} \sum X_i^2 ; \text{ เนื่องจาก } a^2+b^2 = E(X^2)$$

$$\hat{b} = \sqrt{\frac{1}{N} \sum X_i^2 - \hat{a}^2} \tag{3.10}$$

การประมาณพารามิเตอร์ด้วยวิธีโมเมนต์จะได้ค่าประมาณที่ Biased ยกเว้นค่าเฉลี่ยหรือประมาณ โดยใช้โมเมนต์ที่ 1 กรณีที่ค่า Biased ต้องมีการปรับแก้ให้เป็น Unbiased ดังตารางที่ 3.1

ตารางที่ 3.1 การปรับแก้ค่าพารามิเตอร์ให้เป็น Unbiased

Estimator	Unbiased	Biased
s^2	$= \frac{1}{N-1} \sum (X_i - \bar{X})^2$	$= \frac{1}{N} \sum (X_i - \bar{X})^2$
g	$= \frac{N}{(N-1)(N-2)} \cdot \sum \frac{(X_i - \bar{X})^3}{S^3}$	$= \sum \frac{(X_i - \bar{X})^3}{S^3}$

ค่าประมาณจะ Asymptotically Efficient หรือ Unbiased + Minimum MSE เมื่อ $N \rightarrow \infty$

3.3 Method of Least Square Estimate

อนุกรมตัวอย่าง $\{y_t: y_1, y_2, \dots, y_N\}$

แบบจำลอง $y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}; \alpha_1, \alpha_2, \dots, \alpha_p) + \varepsilon_t$

พารามิเตอร์แบบจำลอง $\alpha_1, \alpha_2, \dots, \alpha_p$

Sample Estimates $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p$

$\varepsilon_t =$ Residuals or Error Series with $E(\varepsilon_t) = 0$ and Variance $= \sigma_\varepsilon^2$

$$SSE = \sum \varepsilon_t^2 = \sum (y_t - f(y_{t-1}, y_{t-2}, \dots, y_{t-p}; \hat{\alpha}_1, \dots, \hat{\alpha}_p))^2$$

$$SSE = \sum (y_t - \hat{y}_t)^2 \tag{3.11}$$

$$\frac{\partial}{\partial \hat{\alpha}_1} \sum (y_t - \hat{y}_t)^2 = 0$$

$$\frac{\partial}{\partial \hat{\alpha}_2} \sum (y_t - \hat{y}_t)^2 = 0$$

$$\frac{\partial}{\partial \hat{\alpha}_p} \sum (y_t - \hat{y}_t)^2 = 0$$

Least Square Estimate ถือว่าเป็น Efficient Estimate (Min.MSE)

ตัวอย่างที่ 3.1 Least Square Estimate ของแบบจำลอง AR(1)

$$\text{AR(1) Model: } y_t = \hat{\phi}_1 y_{t-1} + \varepsilon_t \quad [3.12]$$

$$E(Y_t) = 0$$

$$E(\varepsilon_t) = 0$$

$$\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$$

กำหนดอนุกรมเวลาตัวอย่างคือ $\{y_1, y_2, \dots, y_N\}$

$$\sum_{t=1}^N \varepsilon_t^2 = \sum_{t=1}^N (y_t - \hat{\phi}_1 y_{t-1})^2$$

$$= \sum_{t=2}^N (y_t - \hat{\phi}_1 y_{t-1})^2$$

เนื่องจากไม่มีค่า $y_{t-1} = y_0$ ในอนุกรมเวลาตัวอย่าง

$$\frac{\partial \sum_{t=2}^N (y_t - \hat{\phi}_1 y_{t-1})^2}{\partial \hat{\phi}_1} = 0$$

$$\sum_{t=2}^N 2(y_t - \hat{\phi}_1 y_{t-1})(-y_{t-1}) = 0$$

$$\sum_{t=2}^N (-y_t \cdot y_{t-1} + \hat{\phi}_1 y_{t-1}^2) = 0$$

$$\sum_{t=2}^N (-y_t \cdot y_{t-1}) + \hat{\phi}_1 \sum_{t=2}^N y_{t-1}^2 = 0$$

$$\hat{\phi}_1 = \frac{\sum_{t=2}^N (-y_t \cdot y_{t-1})}{\sum_{t=2}^N y_{t-1}^2} \quad [3.13]$$

3.4 Method of Maximum Likelihood

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}; \alpha_1, \alpha_2, \dots, \alpha_p) + \varepsilon_t \quad [3.14]$$

Sample Series: $(y_t; y_1, y_2, \dots, y_N)$

$$\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)$$

$\varepsilon_t =$ Error Time Series which is Independent Series

$$E(\varepsilon_t) = 0$$

$$L(.) = \text{Likelihood Function} = f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N; \underline{\alpha}) \quad [3.15]$$

$$= f(\varepsilon_1; \underline{\alpha}) f(\varepsilon_2; \underline{\alpha}) \dots f(\varepsilon_N; \underline{\alpha}); \text{ Since } \varepsilon_t = \text{Independent Series}$$

$$= \prod_{t=1}^N f(\varepsilon_t; \underline{\alpha})$$

$LL(.) = \text{LogLikelihoodFunction}$

$$= \ln(\prod_{t=1}^N f(\varepsilon_t; \underline{\alpha}))$$

$$= \sum_{t=1}^N \ln(f(\varepsilon_t; \underline{\alpha})) \quad [3.16]$$

ตามหลัก Maximum Likelihood Estimate ตัวประมาณค่าที่ดีที่สุดคือตัวประมาณค่าที่ทำให้ค่า $LL(.)$ มีค่าสูงสุด ดังนั้นตามหลัก Calculus

$$\frac{\partial}{\partial \hat{\alpha}_1} LL(.) = 0$$

$$\frac{\partial}{\partial \hat{\alpha}_2} LL(.) = 0$$

$$\frac{\partial}{\partial \hat{\alpha}_p} LL(.) = 0$$

Maximum Likelihood Estimates มีคุณสมบัติที่สำคัญคือ Asymptotically Efficient, Consistent และ Sufficient Estimates

ตัวอย่างที่ 3.2 การหาพารามิเตอร์ของแบบจำลอง AR(1) โดยวิธี Maximum Likelihood

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

ε_t = Random Error Term ซึ่งมีลักษณะเป็น Independent and Normally Distributed ซึ่งมี

$$E(\varepsilon_t) = 0 \text{ และ } Var(\varepsilon_t) = \sigma_\varepsilon^2$$

$$L(.) = f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N; \phi_1)$$

$$= \prod_{t=1}^N f(\varepsilon_t; \phi_1)$$

$$= \prod_{t=1}^N \frac{1}{\sqrt{2\pi\sigma_\varepsilon}} e^{-\frac{1}{2}\left(\frac{\varepsilon_t}{\sigma_\varepsilon}\right)^2}$$

$$= \frac{1}{(\sqrt{2\pi\sigma_\varepsilon})^N} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^N \varepsilon_t^2}$$

$$LL(.) = \ln\left\{ \frac{1}{(\sqrt{2\pi\sigma_\varepsilon})^N} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^N \varepsilon_t^2} \right\}$$

$$= -N \ln(\sqrt{2\pi\sigma_\varepsilon}) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^N \varepsilon_t^2$$

$$= -N \ln(\sqrt{2\pi\sigma_\varepsilon}) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^N (y_t - \hat{\phi}_1 y_{t-1})^2$$

$$\frac{\partial LL(.)}{\partial \hat{\phi}_1} = -\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^N 2(y_t - \hat{\phi}_1 y_{t-1})(-y_{t-1}) = 0$$

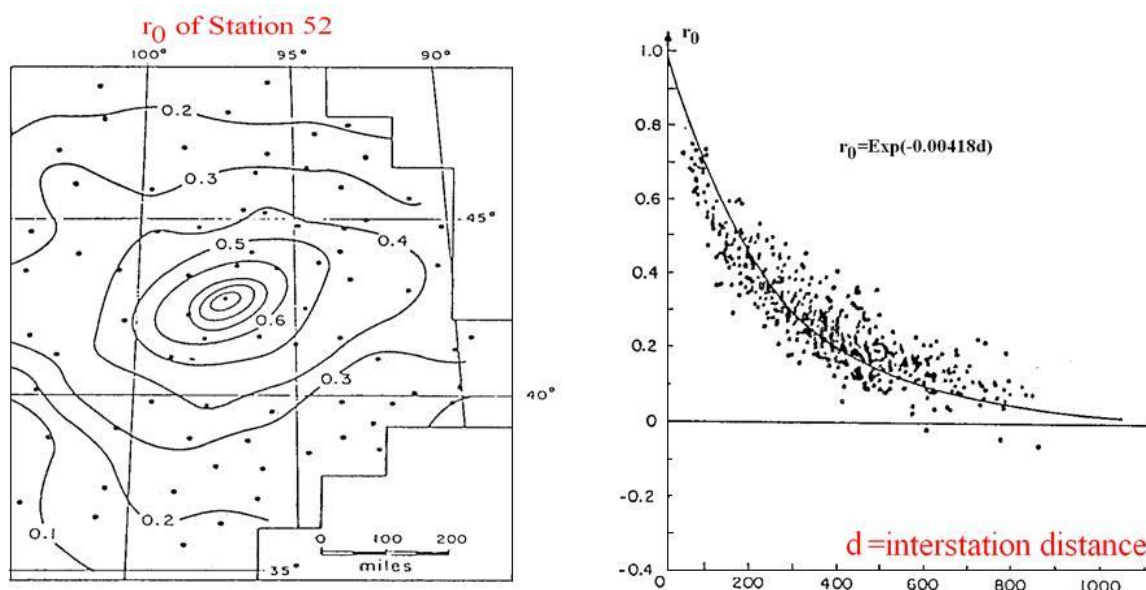
$$\sum_{t=2}^N (y_t y_{t-1} - \hat{\phi}_1 y_{t-1}^2) = 0$$

$$\hat{\phi}_1 = \frac{\sum_{t=2}^N y_t y_{t-1}}{\sum_{t=2}^N y_{t-1}^2}$$

เนื่องจากอนุกรมเวลาตัวอย่างไม่มีค่า y_0 ดังนั้นจึงไม่สามารถหาค่า $\varepsilon_1 = y_1 - \phi_1 y_0$ ได้

3.5 การประมาณค่าพารามิเตอร์โดยวิธี Regionalization (Parameter Estimation by Regionalization)

ในกรณีที่ไม่มีข้อมูล(Ungaged) หรือข้อมูลสั้น(Short Record) จะไม่สามารถใช้วิธีการประมาณค่าพารามิเตอร์ตามวิธีที่กล่าวมาแล้ว ในกรณีนี้จะนิยมใช้วิธีหาสมการแสดงความสัมพันธ์ระหว่างพารามิเตอร์ของแบบจำลองและพารามิเตอร์ของกลุ่มน้ำ เช่นพื้นที่หรือระยะทาง ซึ่งเรียกว่าวิธีการประมาณพารามิเตอร์แบบ Regionalization ยกตัวอย่างเช่น Lag Zero Cross Correlation Coefficient (r_0^{ij}) ของอนุกรมเวลาฝนรายเดือนของ 52 สถานีกับสถานีอื่นๆ อีก 79 สถานี มีความสัมพันธ์ดังรูปที่ 3.1



(a) Isocorrelation (lag zero cross correlation) pattern of monthly precipitation of Station 52 and other 79 stations series in the Upper Plain of USA.

(b) $r_0 = \text{Exp}\{-0.00418d\}$

r_0 = Lag 0 Cross Correlation Coefficient

รูปที่ 3.1 การหาพารามิเตอร์โดยวิธี Regionalization

3.6 การแปลงข้อมูลให้เป็นการแจกแจงปกติ (Normalization)

การแปลงข้อมูลที่มีการแจกแจงแบบเบ้ (Non-normal) x_t ให้เป็นการแจกแจงปกติ (Normal) y_t มีหลายวิธี ดังต่อไปนี้

กรณีอนุกรมเวลารายปี (Annual Series)

(1) การแปลงด้วยฟังก์ชัน Log

ถ้า x_t คือฝนหรือน้ำท่าซึ่งมีการแจกแจงแบบ Log Normal 2 (2 พารามิเตอร์)

$y_t = \text{Log}(x_t)$ จะมีการแจกแจงแบบปกติ $N(y; \mu_y, \sigma_y)$

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}; Y \text{ มีการแจกแจงแบบ Normal} \quad [3.18]$$

$$f(x) = \frac{1}{\sqrt{2\pi \cdot x \cdot \sigma_x}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu_y}{\sigma_y}\right)^2}; X \text{ มีการแจกแจงแบบ Log Normal} \quad [3.19]$$

และถ้า $z_t = \frac{\log(x_t - \mu_y)}{\sigma_y} \sim N(z; 0, 1)$; Z มีการแจกแจงแบบ Standard Normal

ถ้า x_t มีการแจกแจงแบบ Log Normal 3 (3 พารามิเตอร์).

$y_t = \log(x_t - c)$ จะมีการแจกแจงแบบปกติ $N(y; \mu_y, \sigma_y)$ โดยมี c เป็น Lower Bound

(2) การแปลงด้วย Square Root

ถ้า x_t มีการแจกแจงแบบ Gamma 2 (2 พารามิเตอร์)

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad [3.20]$$

$y_t = \sqrt{x_t}$ จะมีการแจกแจงแบบปกติ

ถ้า x_t มีการแจกแจงแบบ Gamma 3 (3 พารามิเตอร์)

$y_t = \sqrt{x_t} - c$ จะมีการแจกแจงแบบปกติ $N(y; \mu_y, \sigma_y)$ โดยมี c เป็น Lower Bound

(3) การแปลงด้วย Power Function

$$y = a(x-c)^b \quad [3.21]$$

b = ค่ายกกำลังซึ่งมีค่าคงที่ เช่น $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$

กรณีอนุกรมแบบ Periodic

มี 2 ทางเลือกในการแปลงข้อมูล

ทางเลือกที่ 1: Normalize $x_{v,\tau}$ ก่อนกำจัด μ_τ และ σ_τ

ทางเลือกที่ 2: ก่อนกำจัด μ_τ และ σ_τ ก่อนการ Normalization

ข้อเสียของทางเลือกที่ 1 คือ μ_τ และ σ_τ ของ Normalized Series ($y_{v,\tau}$) ถูกบิดเบือน (Distorted)

ข้อเสียของทางเลือกที่ 2 คือ $y_{v,\tau} = \frac{x_{v,\tau} - \mu_\tau}{\sigma_\tau}$ มีค่าทั้งบวกและลบ ซึ่งไม่สามารถ Take Log ได้

ฟังก์ชันในการแปลงอนุกรมเวลาแบบ Periodic ทั่วไป

$$y_{v,\tau} = a_\tau (x_{v,\tau} + c_\tau)^{b_\tau} \quad [3.22]$$

เมื่อ $c_\tau =$ Periodic Lower Bound $a_\tau, b_\tau =$ Other Parameters

$$\text{หรือ } y_{v,\tau} = \log(x_{v,\tau} + c_\tau) \quad ; \tau = 1, \dots, w \quad [3.23]$$

กรณีอนุกรมเวลาแบบ Periodic ควรใช้ฟังก์ชันการแปลงเดียวกันในทุกคาบ τ

3.7 การหาพารามิเตอร์แบบ Periodic โดยใช้ Fourier Series

$y_{v,\tau} =$ Normal Periodic Time Series

$$y_{v,\tau} = \mu_\tau + \sigma_\tau z_{v,\sigma}$$

$z_{v,\tau} =$ Standardized Normal Series ซึ่งอาจเป็นอนุกรมเวลาแบบ Dependent ที่มีค่า Correlation แบบคงที่ หรือ Periodic ก็ได้

AR(1) แบบ Periodic Coefficients

$$z_{v,\tau} = \phi_{1,\tau} z_{v,\tau-1} + \varepsilon_{v,\tau} \quad [3.24]$$

$\phi_{1,\tau}$ คือ Periodic Coefficient = $f(\rho_{k,\tau})$

กรณี AR(1): $\phi_{1,\tau} = \rho_{1,\tau}$ [3.25]

ถ้า $\mu_\tau, \sigma_\tau, \rho_{k,\tau}$ สามารถประมาณค่าได้จากอนุกรมตัวอย่าง

3.7.1 วิธีการประมาณค่าแบบ Non-Parametric Approach

ตามวิธี Non-Periodic จะสามารถประมาณค่าพารามิเตอร์ได้ดังนี้

$$\begin{aligned}\bar{x}_\tau &\sim \mu_\tau \\ s_\tau^2 &\sim \sigma_\tau^2 \\ g_\tau &\sim \gamma_\tau \\ r_{k,\tau} &\sim \rho_{k,\tau} \\ u_\tau &\sim v_\tau\end{aligned}$$

เมื่อ u คือสัญลักษณ์ของตัวประมาณค่าแบบ Non-parametric และ v คือพารามิเตอร์ตามลำดับ

3.7.2 การประมาณค่าพารามิเตอร์โดยวิธี Fourier Series

$$\bar{y}_\tau = \frac{1}{N} \sum_{v=1}^N y_{v,\tau}; \tau = 1, \dots, w \quad [3.26]$$

$$e_\tau = \mu_\tau - \bar{y}_\tau; \tau = 1, \dots, w \quad [3.27]$$

เมื่อ e_τ คือค่าความผิดพลาดในการประมาณค่า Mean โดยวิธี Non-parametric

กรณีที่ N น้อย e_τ จะมีค่ามาก

กรณีที่ N มาก พังค์ชันของ \bar{y}_τ จะ Smooth กว่าและ e_τ จะมีค่าน้อยกว่า

วิธี Fourier Series จะให้ค่าประมาณที่ Smooth กว่าวิธี Non-parametric

3.7.3 การประมาณค่า Fourier Series Coefficients

กำหนดให้ u_τ เป็นตัวประมาณค่าแบบ Non-parametric ของพารามิเตอร์ (v_τ)

ให้ \hat{v}_τ Fourier Series Estimates of v_τ

$$\hat{v}_\tau = \bar{u} + \sum_{j=1}^h \left\{ A_j \cos\left(\frac{2\pi j\tau}{\omega}\right) + B_j \sin\left(\frac{2\pi j\tau}{\omega}\right) \right\}; \tau = 1, \dots, w \quad [3.28]$$

$$\bar{u} = \frac{1}{\omega} \sum_{\tau=1}^{\omega} u_\tau \quad [3.29]$$

$A_j, B_j =$ Fourier Series Coefficients ($j=1, \dots, h$)

$j =$ Harmonics j

$h =$ Total No. of Harmonics and the Last Harmonic

$h = \frac{\omega}{2}$ ถ้า $\omega =$ เลขคู่ ยกตัวอย่าง $h = 12/2 = 6$

$h = \frac{\omega-1}{2}$ ถ้า $\omega =$ เลขคี่ ยกตัวอย่าง $h = (365-1)/2 = 182$

$$A_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} u_\tau \cos\left(\frac{2\pi j\tau}{\omega}\right); j = 1, \dots, h \quad [3.30]$$

$$B_j = \frac{2}{\omega} \sum_{\tau=1}^{\omega} u_\tau \sin\left(\frac{2\pi j\tau}{\omega}\right); j = 1, \dots, h \quad [3.31]$$

ถ้า $\omega =$ เลขคู่

$$A_h = \frac{1}{\omega} \sum_{\tau=1}^{\omega} u_\tau \cos\left(\frac{2\pi h\tau}{\omega}\right) \quad [3.32]$$

$$B_h = 0 \quad [3.33]$$

ถ้าใช้ทุก Harmonic ในการคำนวณ \hat{v}_τ จะได้ว่า $\hat{v}_\tau = u_\tau$

ดังนั้น ในการประมาณค่าพารามิเตอร์จะใช้เพียง Significant Harmonics (h^*) ซึ่งน้อยกว่า h

Cumulative Periodogram คือวิธีการหาค่า Significant Harmonics

3.7.4 Cumulative Periodogram Test (P_i)

$$MSD(u) = \frac{1}{\omega} \sum_{\tau=1}^{\omega} (u_{\tau} - \bar{u})^2 = \text{Variance of } u \quad [3.34]$$

เมื่อ MSD(u) = Mean Square Deviation of u หรือ Total Variation of u

$$MSD(j) = \frac{1}{2} (A_j^2 + B_j^2) ; j = 1, \dots, \dots, h \quad [3.35]$$

= Mean Squared Deviation of Each Harmonic (j)

$$MSD(u) = \sum_{j=1}^h MSD(j) ; \text{ All } h \text{ harmonics} \quad [3.36]$$

$$P_i = \sum_{j=1}^i \frac{MSD(h_j)}{MSD(u)} \quad [3.37]$$

เมื่อ P_i = Cumulative Periodogram

h_j = Significant Harmonic ที่ h_j

ขั้นตอนการคำนวณหา $\sum_{j=1}^i MSD(h_j)$ แสดงอยู่ในตารางที่ 3.2 เมื่อนำ P_i ไปพล็อตกราฟจะได้

Cumulative Periodogram ซึ่งสามารถแบ่งออกเป็น 2 ส่วน Periodic Part และ Sampling Variation Part ดังรูปที่ 3.2 จุดแบ่งระหว่างส่วนที่เป็น Periodic และ Sampling Variation คือ Significant Harmonics ถ้า Sampling Variation Part เป็นเส้นตรงดังรูปที่ 3.2 แสดงว่าอนุกรมเวลาเมื่อกำจัดส่วนที่เป็น Periodic ออกไปแล้วจะมีลักษณะเป็น Independent แต่ถ้า Sampling Variation Part เป็นเส้นโค้งดังรูปที่ 3.3 แสดงว่าอนุกรมเวลาเมื่อกำจัดส่วนที่เป็น Periodic ออกไปแล้วจะมีลักษณะเป็น Dependent

ตารางที่ 3.2 Calculation of Cumulative Periodogram

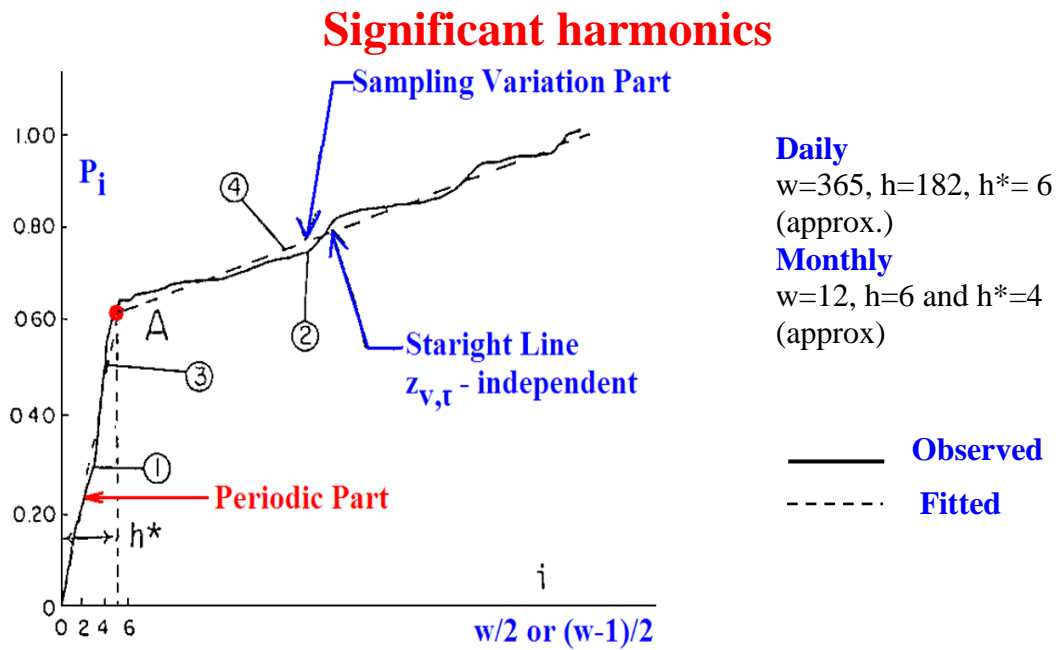
Harmonic	MSD	Rearranged MSD*	Rearranged Order Harmonic	Cumulative MSD(h _j) เมื่อ j=1 to i
j	MSD(j)	MSD(h _j)	i	$\sum_{j=1}^i MSD(h_j)$
1	MSD(1)	MSD(h ₁)	1	$\sum_{j=1}^{i=1} MSD(h_j) = MSD(h_1)$
2	MSD(2)	MSD(h ₂)	2	$\sum_{j=1}^{i=2} MSD(h_j) = MSD(h_1) + MSD(h_2)$
.	.	.	i	$\sum_{j=1}^i MSD(h_j)$
h	MSD(h)	MSD(h _h)	h	
Total	MSD(u)	MSD(u)		

*Re-arranged from most significant harmonics [highest MSD(j)] to least significant harmonics

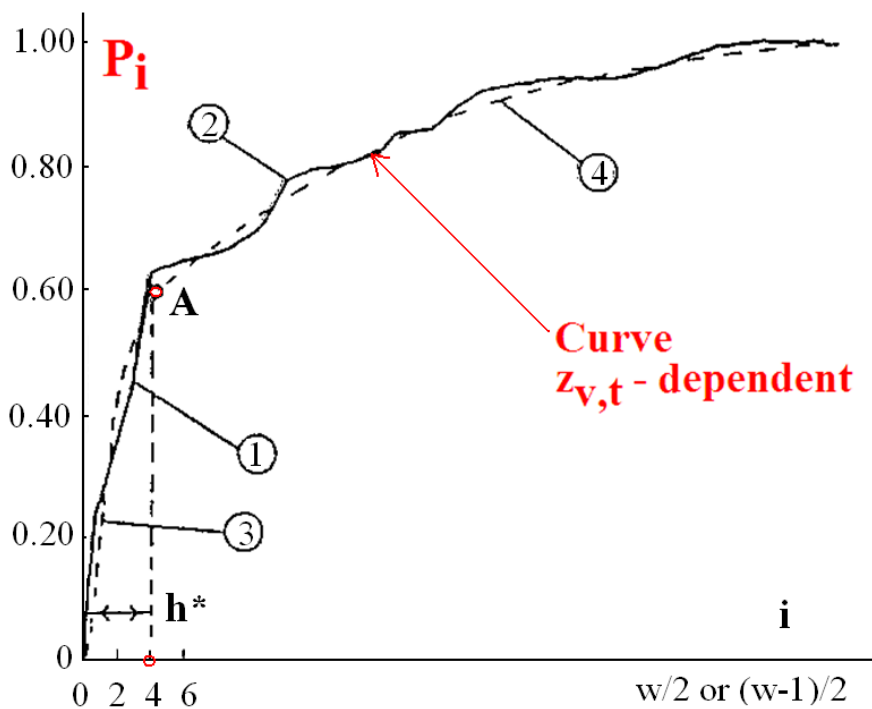
h₁ = most significant harmonic [highest MSD(j)]

h_h = least significant harmonic [lowest MSD(j)]

$$P_i = \frac{\sum_{j=1}^i MSD(h_j)}{MSD(u)}$$

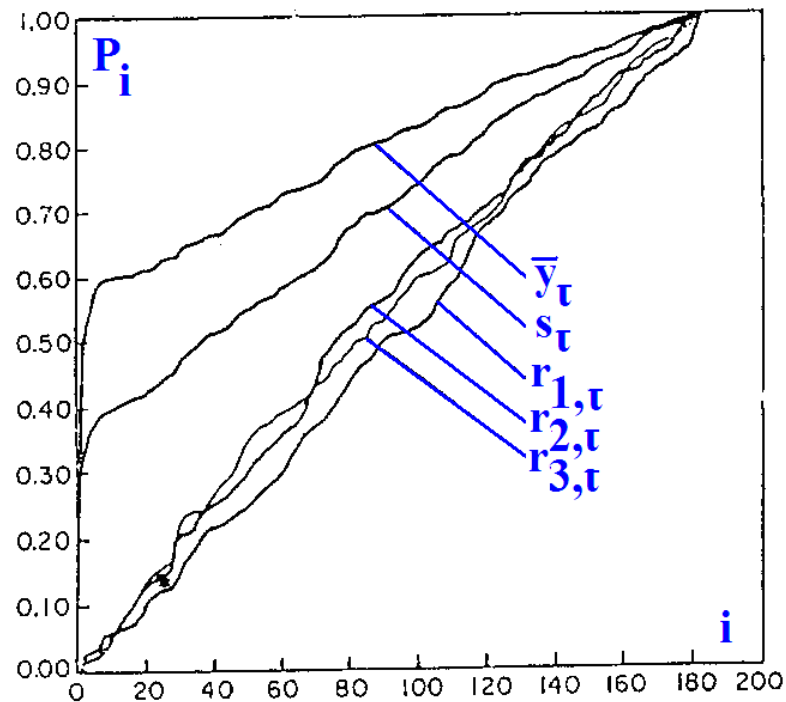


รูปที่ 3.2 Separation of the Cumulative Periodogram into the periodic part for both the observed (1) and the fitted (3) and the sampling variation part, also for both the observed (2) and the fitted (4) in case of a periodic time series with an independent stochastic component.

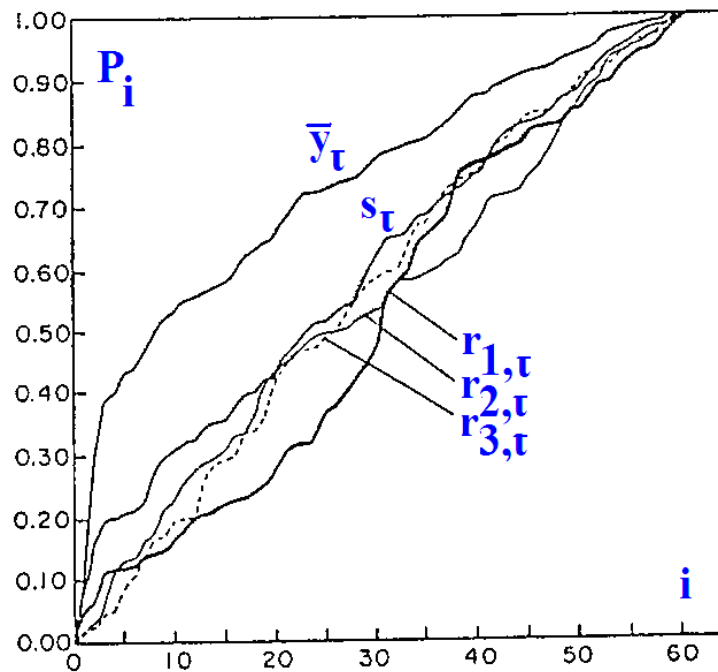


รูปที่ 3.3 Separation of the Cumulative Periodogram into the periodic part, observed (1) and fitted (3), and the sampling variation part, observed (2) and fitted(4), in case of a periodic series with an autoregressive stochastic component.

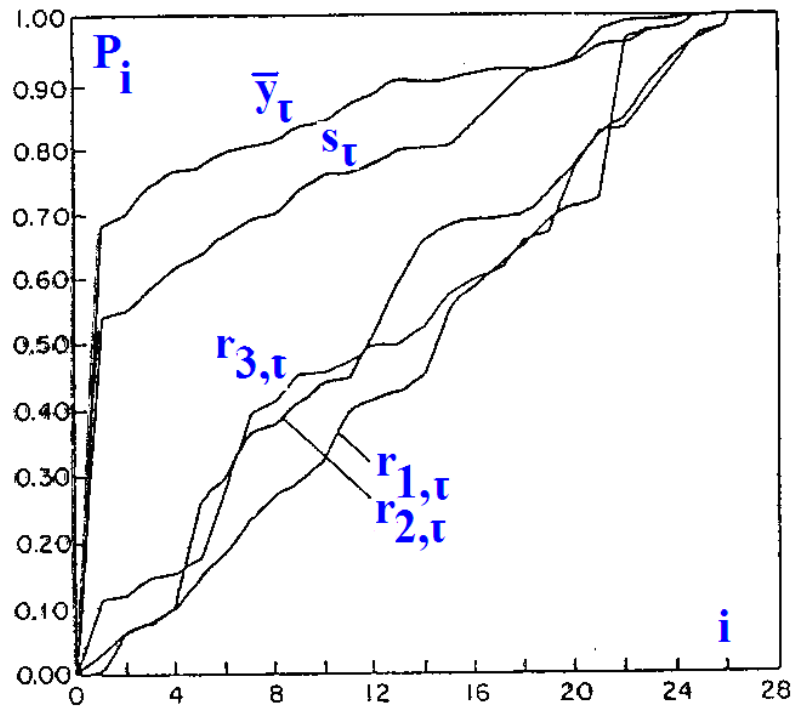
ตัวอย่าง Cumulative Periodogram ของพารามิเตอร์แบบต่างๆของฝนและน้ำท่าแสดงอยู่ในรูปที่ 3.4-3.7 กรณีที่จุดแบ่งระหว่าง Periodic Part และ Sampling Variation Part ไม่ชัดเจนจะหา Significant Harmonic(h^*) ได้ที่ 0.9-0.95 ของ Explained Variance (P_i) ดังรูปที่ 3.7



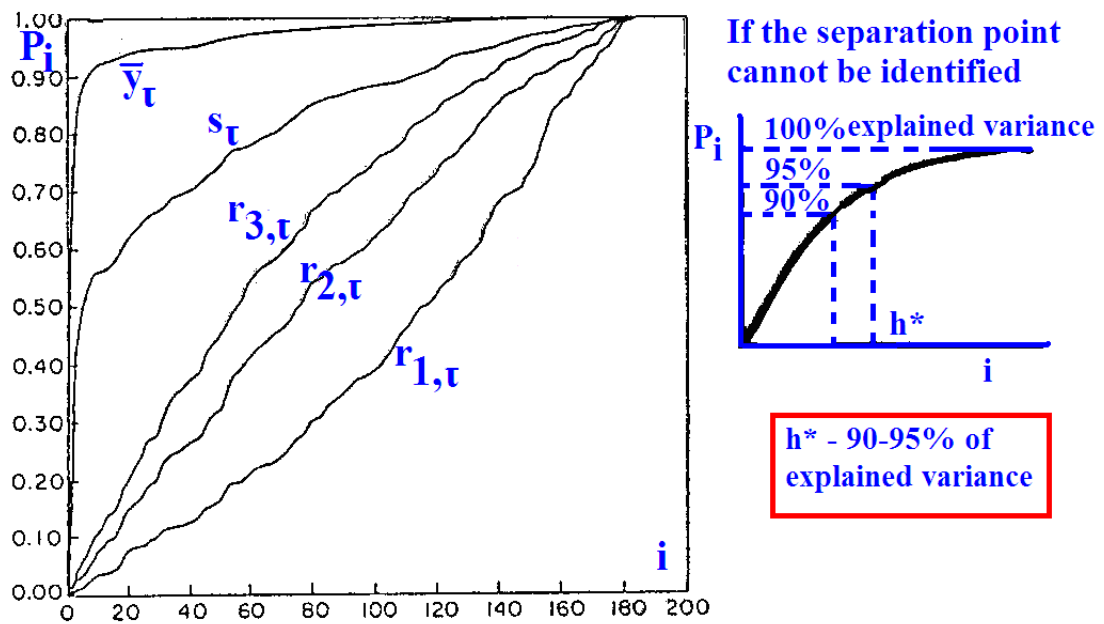
รูปที่ 3.4 Cumulative Periodogram of five parameters of daily precipitation series, Ft. Collins, Colorado



รูปที่ 3.5 Cumulative Periodogram of five parameters of 3-day precipitation series, Austin, Texas



รูปที่ 3.6 Cumulative periodogram of five parameters of 7-day precipitation series, Ames, Iowa



รูปที่ 3.7 Cumulative periodogram of five parameters of daily flow series of Tioga river

3.8 การประมาณค่าพารามิเตอร์ของแบบจำลองหลายตัวแปร (Multivariate Model Parameter Estimation)

3.8.1 วิธีการถอด Square Root ของเมทริกซ์

$$BB^T = D$$

[3.38]

ถ้ากำหนดค่า Matrix D จงหาค่า Matrix B

ผลการคำนวณจะให้ผลลัพธ์แบบ Unique Solution ก็ต่อเมื่อ

B = Lower Triangular Matrix

และ $D =$ Positive Definite หรือ Positive Semi-definite

(1) Positive Definite

$$\begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{bmatrix} > 0$$

(2) Positive Semi-definite

$$\begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{bmatrix} \geq 0$$

Graybill(1969) เสนอแนะวิธีในการหาค่า B เมื่อ D เป็น Positive Definite ดังนี้

$$i \begin{bmatrix} & j \\ 1 & & & \\ 1 & 2 & & \\ 1 & 3 & 2 & \\ 1 & 3 & 3 & 2 \end{bmatrix} \begin{aligned} (1) & b^{ij} = \frac{d^{ij}}{b^{jj}} \text{ for } j = 1, i = 1, \dots, n & [3.39] \\ (2) & b^{ij} = \sqrt{d^{ij} - \sum_{k=1}^{j-1} (b^{jk})^2} \text{ for } j = 2, \dots, n; i = j & [3.40] \\ (3) & b^{ij} = \frac{[d^{ij} - \sum_{k=1}^{j-1} b^{jk} b^{ik}]}{b^{jj}} \text{ for } j = 2, \dots, n-1; i = j+1 & [3.41] \end{aligned}$$

$n =$ size of square matrix B and D

Lane (1979) เสนอแนะวิธีในการหาค่า Matrix B เมื่อ $D =$ Positive หรือ Positive Semi-definite ดังนี้

$$i \begin{bmatrix} & j \\ x & & & \\ x & x & & \\ x & x & x & \\ x & x & x & x \end{bmatrix} \begin{aligned} (1) & b^{ki} = 0 \text{ for all } k < i \\ (2) & b^{ki} \neq 0 \text{ for all } k \geq i \\ \text{เมื่อ} & d^{ii} - \sum_{j < i} (b^{ij})^2 \leq 0 \\ \text{และ} & b^{ki} = \frac{d^{ki} - \sum_{j < i} b^{ij} b^{kj}}{\sqrt{d^{ii} - \sum_{j < i} (b^{ij})^2}} \text{ for all } k \geq i & [3.42] \end{aligned}$$

$$\text{เมื่อ } d^{ii} - \sum_{j < i} (b^{ij})^2 > 0 \quad [3.43]$$

เริ่มการคำนวณจากทุกค่าของ Matrix ในคอลัมน์ 1 แล้วจึงคำนวณค่า Matrix ในคอลัมน์ 2, 3, 4,.....

ตามลำดับ

3.9 การทดสอบความเหมาะสมของแบบจำลอง (Goodness of Fit Test)

สมมติฐานหลักของแบบจำลองคือ \mathcal{E}_t เป็นตัวแปรอิสระและมีการแจกแจงแบบปกติ (Independent and Normally Distributed) จึงต้องมีการทดสอบความเป็นอิสระของตัวแปร \mathcal{E}_t และทดสอบว่าตัวแปร \mathcal{E}_t มีการแจกแจงแบบปกติ

3.9.1 การทดสอบความเป็นอิสระ (Test of Independence)

การทดสอบความเป็นอิสระมีหลายวิธีดังนี้

(1) Anderson Test

วิธีนี้จะเริ่มจากการคำนวณหา Probability Limit ดังสมการ [3.44] ถ้า ACF ของ \mathcal{E}_t ที่ Lag มากกว่า 0 มีค่าอยู่ใน Probability Limit แสดงว่าตัวแปร \mathcal{E}_t เป็นตัวแปรอิสระ

$$r_k(\alpha) = \frac{-1 \pm Z_{\alpha/2} \sqrt{N-k-1}}{N-k} \quad [3.44]$$

$z_{\alpha/2}$ คือ z-score ที่ Exceedence Probability เท่ากับ $\alpha/2$

ถ้า r_k ตกในช่วง $\pm r_k(\alpha)$ แสดงว่าเป็นแบบอิสระ

(2) Porte Manteau Lack of Fit Test

- กรณี Constant Coefficient

ตรวจสอบโดยใช้ Porte Manteau Statistic ดังสมการ [3.45]

$$Q = (N - d) \sum_{k=1}^L r_k^2(\varepsilon) \quad [3.45]$$

ถ้า $Q < \chi_{L-p-q}^2$ แสดงว่าแบบจำลอง Adequate หรือ $\varepsilon_t = \text{Independent}$

$Q = \text{Porte Manteau Statistics}$

$L = \text{Max. Lag} = 0.1N - 0.3N$

$D = \text{No. of Differences}$ สำหรับ ARIMA(p,d,q)

$\chi_{L-p-q}^2 = \text{Chi-square Statistic}$ ที่ α Probability และ $L-p-q$ Degree of Freedom.

- Periodic Coefficient

กรณีของอนุกรมเวลาแบบ Periodic จะใช้ค่า Modified Porte Manteau Statistics (Q_1) ดังสมการ [3.46]

ในการทดสอบ

$$Q_1 = N \sum_{k=1}^L \sum_{\tau=1}^{\omega} [r_{k,\tau}(\varepsilon)]^2 \quad [3.46]$$

$\omega = \text{No. of Periods per Year}$

$N = \text{No. of Years of Record}$

(3) Cumulative Periodogram Test

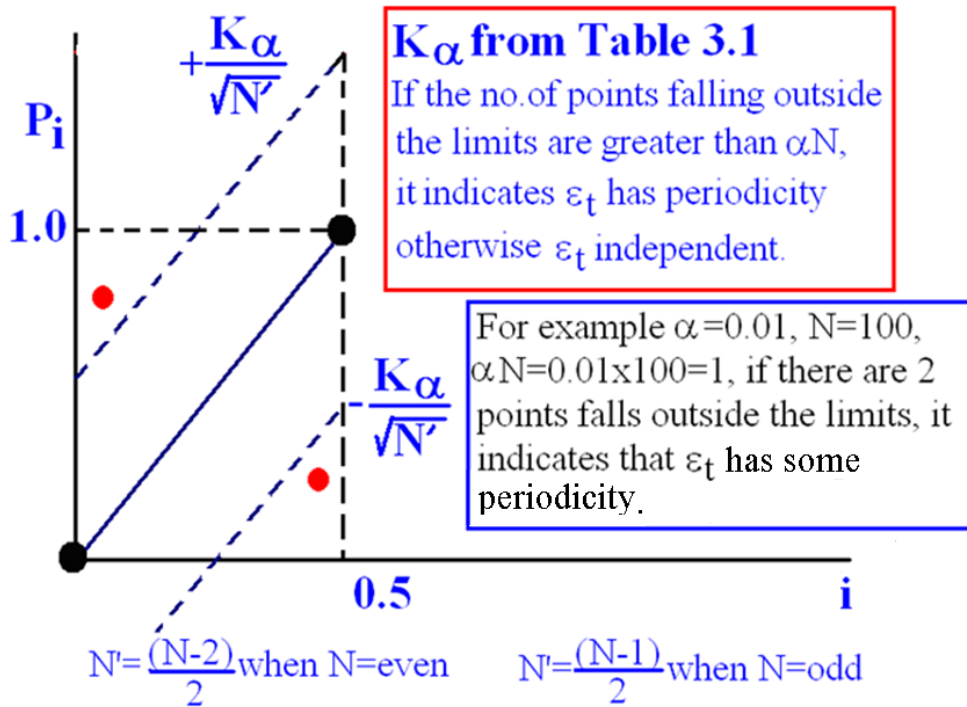
วิธีการทดสอบความเป็นอิสระโดยใช้ Cumulative Periodogram ได้กล่าวมาแล้วในหัวข้อ 3.7.4

หลักการทดสอบแบบง่ายคือการพล็อต P_i เทียบกับ i สำหรับอนุกรมเวลาแบบ Independent ค่า P_i จะกระจายอยู่รอบๆเส้นตรงที่เชื่อมจุด (0,0) และ (0.5, 1.0) ถ้าจำนวน P_i ที่คำนวณได้กระจายอยู่ภายนอก Confident Interval $\pm \frac{K\alpha}{\sqrt{N\omega}}$ ของเส้นตรงที่เชื่อมจุด (0,0) และ (0.5, 1.0) มากเกินกว่า αN แสดงว่า ε_t ยังคงมีองค์ประกอบที่เป็น Periodicity อยู่ Periodicity ที่มี Harmonic h_j ใน ε_t จะทำให้ $\text{MSD}(h_j)$ มีค่ามากและมีผลทำให้ค่าที่พล็อตเบี่ยงเบนจากเส้นตรงมากขึ้น ค่า $K\alpha$ คือ ค่า Smirnov-Kolmogorov Statistics ในตารางที่ 3.3

$$P_i = \sum_{j=1}^i \frac{\text{MSD}(h_j)}{\sigma_{\varepsilon}^2}; \quad i = 1, \dots, \frac{N}{2} \text{ or } \frac{N-1}{2} \quad [3.47]$$

เมื่อ $\sigma_{\varepsilon}^2 = \text{Total Variance} = \text{MSD}(u)$

$$\text{MSD}(h_j) = \frac{2}{N^2} \left[\left(\sum_{t=1}^N \varepsilon_t \cos(2\pi h_j t) \right)^2 + \left(\sum_{t=1}^N \varepsilon_t \sin(2\pi h_j t) \right)^2 \right]; \quad h_j = \frac{j}{N} \quad [3.48]$$



รูปที่ 3.8 การทดสอบความเป็นอิสระโดยใช้ Cumulative Periodogram

ตารางที่ 3.3 Smirnov-Kolmogorov Statistics, K_α

α	0.01	0.05	0.1	0.2	0.25
K_α	1.63	1.36	1.22	1.07	1.02

(1) Test of Independence in Space

ใช้หลักการทดสอบสมมติฐานว่า Cross Correlation Coefficients (ρ^{ij}) เท่ากับหรือต่างจาก 0 ถ้า N มีค่ามากจะทดสอบโดยใช้ Z-Score ดังสมการ [3.49] ถ้า N มีค่ามากจะทดสอบโดยใช้ Confident Limit ดังสมการ [3.50]

$$H_0 : \rho^{ij} = 0$$

$$H_1 : \rho^{ij} \neq 0$$

เมื่อ N มีค่ามาก

$$\frac{-Z_\alpha}{\sqrt{N}} < r_0^{ij} < \frac{+Z_\alpha}{\sqrt{N}} \tag{3.49}$$

ถ้า N มีค่าน้อย

$$r_\beta^l < r_0^{ij} < r_\beta^u \dots \text{Accept } H_0 : \rho^{ij} = 0$$

$$r_\beta = \frac{(1+\rho) - (1-\rho) \exp(2Z_\beta/\sqrt{N-3})}{(1+\rho) + (1-\rho) \exp(2Z_\beta/\sqrt{N-3})} \tag{3.50}$$

$$\beta = \frac{\alpha}{2} = \text{Upper limit } \left(\frac{\alpha}{2} = 0.05, 0.025 \right)$$

$$\beta = 1 - \frac{\alpha}{2} = \text{Lower limit}$$

3.9.2 การทดสอบความเป็นปกติ (Test of Normality)

การทดสอบความเป็นปกติคือการทดสอบว่าอนุกรมเวลาหรือตัวแปรมีการแจกแจงแบบปกติ มีวิธีที่นิยมใช้โดยทั่วไป 3 วิธีคือ

- (1) วิธีกราฟ โดยการพล็อต Empirical Distribution บนกระดาษกราฟการแจกแจงปกติ (Normal Probability Paper) ถ้ากราฟที่พล็อตเป็นเส้นตรง แสดงว่าอนุกรมเวลาที่มีการแจกแจงแบบปกติ
- (2) การทดสอบแบบ χ^2 -Test
- (3) การทดสอบความเบ้ (Skewness Test) การแจกแจงแบบปกติจะมีค่า $\gamma = 0$

ถ้า $-Z_{\frac{\alpha}{2}}\sqrt{\frac{6}{N}} < \gamma < Z_{\frac{\alpha}{2}}\sqrt{\frac{6}{N}}$ แสดงว่า $\gamma=0$ หรือมีการแจกแจงแบบปกติ ถ้า $N > 150$

ถ้า $N < 150$ ใช้ค่า $\gamma_{\alpha,N}$ จากตารางที่ 3.4

ถ้า $\hat{\gamma} < \gamma_{\alpha,N}$ แสดงว่า $\gamma=0$ หรือมีการแจกแจงแบบปกติ (Normal Distribution)

การทดสอบแบบ $\chi^2 - Test$

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} \quad [3.51]$$

O_i = Observed Frequency

e_i = Expected Frequency

ถ้า $\chi^2 < \chi_{\alpha, (k-2)}^2$ แสดงว่ายอมรับ $H_0 : x_t \sim N(x_t; \bar{x}, s)$

Table 3.4 Table for skewness test for normality for sample size less than 150
(after Snedecor and Cochran, 1967, p552)

N	γ_{α}		N	γ_{α}	
	α			α	
	0.02	0.10		0.02	0.10
25	1.061	0.711	70	0.673	0.459
30	0.986	0.662	80	0.631	0.432
35	0.923	0.621	90	0.596	0.409
40	0.870	0.587	100	0.567	0.389
45	0.825	0.558	125	0.508	0.350
50	0.787	0.534	150	0.464	0.321
60	0.723	0.492	175	0.430	0.298

3.9.3 การทดสอบจำนวนพารามิเตอร์ที่เหมาะสม (Parsimony of Parameters)

ในการสร้างแบบจำลองทางสถิติ มีหลักสำคัญที่ต้องยึดไว้คือการรักษาคุณสมบัติทางสถิติของข้อมูล (Preservation of Statistics) ซึ่งมีข้อเท็จจริงดังนี้ ถ้าแบบจำลองมีจำนวนพารามิเตอร์มาก คุณสมบัติทางสถิติของอนุกรมเวลาที่สังเคราะห์ขึ้นมา (Generated Data) จะมีค่าเข้าใกล้คุณสมบัติทางสถิติของอนุกรมเวลาตัวอย่าง (Historical Sample) ถ้าจำนวนพารามิเตอร์เท่ากับจำนวนข้อมูลตัวอย่าง (No. of Sample Size) อนุกรมเวลาที่สังเคราะห์ขึ้นมาจะเหมือนกับอนุกรมเวลาตัวอย่างทุกประการ ดังนั้นโดยปกติในการสร้างแบบจำลอง จะกำหนดให้มีจำนวนพารามิเตอร์ไม่มากเกินไป โดยจะเลือกใช้เฉพาะพารามิเตอร์ที่มีความแปรปรวนน้อยและมีความสำคัญต่อแบบจำลองเท่านั้น เช่น ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน และ Serial Correlation Coefficients

(1) Parsimony

ตามหลักการของ Parsimony ที่ต้องจำกัดจำนวนพารามิเตอร์ไม่ให้มากเกินไป จะกำหนดให้ดัชนี Parsimony (Index of Parsimony) ต้องมากกว่าหรือเท่ากับ 15 ดังสมการที่ 3.52

$$\delta = \frac{N}{k} = \text{Index of Parsimony} \geq 15 \quad [3.52]$$

เมื่อ $N =$ Sample Size

$k =$ จำนวนพารามิเตอร์

$N = 30 \times 12 = 360$ เดือน

$$k = \frac{N}{\delta} = \frac{360}{15} = 24 \text{ (จำนวนพารามิเตอร์ควรน้อยกว่า 24)}$$

โดยทั่วไป แบบจำลองรายเดือนจะมีจำนวนพารามิเตอร์ = 36 {12 μ , 12 σ , 12 ρ }

(2) Akaike Information Criteria (AIC)

AIC คือเกณฑ์ที่สำคัญที่ใช้ทดสอบ Parsimony of Parameter สำหรับแบบจำลอง ARMA(p,q) จะสามารถคำนวณหาค่า AIC ได้ดังสมการ 3.53

$$\text{AIC}(p, q) = N \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q) \quad [3.53]$$

$N =$ Sample Size

$\hat{\sigma}_\varepsilon^2 =$ Max. Likelihood Estimate ของ σ_ε^2

ตามเกณฑ์ AIC แบบจำลองที่เหมาะสมคือแบบจำลองที่มี AIC ต่ำสุด

3.10 การสังเคราะห์ข้อมูลและการพยากรณ์ (Generation and Forecasting)

3.10.1 การสังเคราะห์ข้อมูล (Generation of Synthetic Samples)

ตามหลักการที่กล่าวมาแล้ว การสร้างแบบจำลองอนุกรมเวลาทางอุทกวิทยาก็คือการสร้างแบบจำลองของ White Noise หรือ Independent Stochastic Component ซึ่งเรียกกันทั่วไปว่า Whitening Process โดยทั่วไปจะสมมติว่า ε คือ White Noise ซึ่งเป็น Independent Stationary Normal Variable.

การสังเคราะห์ข้อมูลอนุกรมเวลาจะเริ่มจากการสังเคราะห์ Independent Normal Variables ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 หลังจากนั้นจึงเพิ่ม Time and Spatial Dependence Structure และ Periodic Components เข้าไป

(1) Univariate Time Series

ยกตัวอย่างการสังเคราะห์ข้อมูลโดยใช้แบบจำลอง AR(1)

$$z_t = \phi_1 z_{t-1} + \varepsilon_t \quad [3.54]$$

เมื่อ

z_t = Dependent Normal Variable ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

$$Z_t = \frac{y_t - \hat{\mu}_y}{\hat{\sigma}_y} \quad [3.55]$$

ϕ_1 = Autoregressive Coefficient

สมมติให้ $y_t = \log(x_t)$ = Normalized Variable with Mean μ_y and Variance σ_y^2

ε_t = Independent Normal Variable with Mean Zero and Variance σ_ε^2

$$\varepsilon_t = \sigma_\varepsilon \xi_t \quad [3.56]$$

$$\sigma_\varepsilon^2 = 1 - \phi_1^2; (\sigma_z^2 = 1) \quad [3.57]$$

สามารถหา σ_ε^2 ได้โดยการ Taking Variance ของ ε_t ของแบบจำลอง of AR(1)

ξ_t = Independent Normal Variable with Mean Zero and Variance One

ขั้นตอนการสังเคราะห์อนุกรมเวลา

(1) Generation of ξ_1

(2) สมมติให้ $z_0 = \mu_z = 0$

(3) $z_1 = 0 + \sigma_\varepsilon \xi_1$

(4) $y_1 = \hat{\mu}_y + \hat{\sigma}_y z_1$

(5) $x_1 = \text{Antilog}(y_1)$

(6) ทำขั้นที่ 1 ถึง 5 สำหรับ $t = 2, 3, \dots, N$

(7) สามารถสังเคราะห์อนุกรมเวลาที่มีขนาด N ได้หลายชุดตามที่ต้องการ แต่ในการสังเคราะห์อนุกรมเวลาต้องมีการตรวจสอบว่าอนุกรมเวลาที่สังเคราะห์ขึ้นใหม่สามารถรักษาคูสมบัตินทางสถิติของอนุกรมเวลาในอดีตได้

ขั้นตอนการสังเคราะห์ White Noise (ξ_t) ด้วย Excel

(1) สังเคราะห์ Uniform Random Number โดยใช้ฟังก์ชัน RAND()

(2) สังเคราะห์ Standard Normal Variate โดยใช้ฟังก์ชัน NORMSINV ให้ผลลัพธ์จากฟังก์ชัน RAND() คือ Input สำหรับฟังก์ชัน NORMSINV

- (3) ในการสังเคราะห์ Random Number ด้วยคอมพิวเตอร์จะไม่ได้ Pure Random Number แต่ได้ Pseudo-Random ซึ่งมักมี Initial Biases ดังนั้นจึงจำเป็นต้องมีการสังเคราะห์ข้อมูลเพื่อไว้ตัดทิ้ง ซึ่งเรียกว่า Warm Up Length

อนุกรมเวลาที่สังเคราะห์ขึ้นจะมีประโยชน์ในการวิเคราะห์ความเสี่ยง (Risk) ในการวางแผน ออกแบบ และการปฏิบัติการเกี่ยวกับระบบแหล่งน้ำ

(2) Multivariate Time Series

การสังเคราะห์อนุกรมเวลาแบบหลายตัวแปรที่สามารถรักษาคุณสมบัติเกี่ยวกับ Time และ Cross Dependent Structures ของอนุกรมเวลาหลายตัวแปรมี 3 แนวทาง ดังนี้

แนวทางที่ 1: ใช้ Linear Regression Analysis

ถ้ามีจำนวนตัวแปรน้อยกว่า 5 ($n \leq 5$)

$$\xi_t^{(j)} = \alpha_0 + \alpha_1 \xi_t^{(1)} + \dots + \alpha_{j-1} \xi_t^{(j-1)} + \eta_t^{(j)} \quad [3.58]$$

เมื่อ

$\xi_t^{(j)}$ = Independent Normal Random Variable (White Noise) with Mean Zero and Variance

One ของอนุกรมเวลา j เมื่อ $j=2,3,\dots,n$

$\alpha_0, \alpha_1, \dots, \alpha_{j-1}$ = Regression Parameters

$\eta_t^{(j)}$ = Independent Normal Random Variable with Mean Zero and Variance = $\sigma_n^{2(j)}$

เริ่มจากการสังเคราะห์ $\xi_t^{(1)}, \eta_t^{(2)}$ แล้วจึงคำนวณหา $\xi_t^{(2)}, \xi_t^{(3)}, \dots, \xi_t^{(j)}$ เพื่อรักษา Cross-correlation Structure (เมื่อเพิ่มค่า $\xi_t^{(j)}$ ใหม่เข้าไปในสมการ จะต้องคำนวณค่าพารามิเตอร์ α_j ใหม่) หลังจากนั้นจึงแปลง $\xi_t^{(j)}$ เป็น $\hat{z}_t^{(j)}, \hat{y}_t^{(j)}, \hat{x}_t^{(j)}$ ตามลำดับ

ตัวอย่างการใช้สมการ Linear Regression ของ $\xi_t^{(j)}$ เมื่อ $j=1,\dots,5$ เพื่อการสังเคราะห์อนุกรมเวลารายเดือนของ Net Basin Water Supply ของ Lake St.Clair $\xi_t^{(5)}$ และ Lakes Ontario, Erie, Superior and Michigan-Huron แสดงอยู่ในสมการ 3.59 (Yevjevich, 1975)

$$\xi_t^{(5)} = -0.00065 + 0.1258\xi_t^{(1)} + 0.16721\xi_t^{(2)} - 0.04994\xi_t^{(3)} + 0.19315\xi_t^{(4)} + \eta_t^{(5)} \quad [3.59]$$

แนวทางที่ 2: ใช้ Multivariate Matrix Approach

โดยวิธีนี้จะสังเคราะห์อนุกรมเวลาทั้ง j ชุด พร้อมกันแทนที่จะสังเคราะห์ทีละชุดเหมือนแนวทางที่ 1 รายละเอียดของวิธีนี้จะกล่าวถึงในบทที่

$$Z_t = A_1 Z_{t-1} + B \varepsilon_t \quad \dots \dots \dots \text{Multivariate AR(1)} \quad [3.60]$$

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B \varepsilon_t \quad \dots \dots \dots \text{Multivariate AR(2)} \quad [3.61]$$

เมื่อ

$Z_t = (n \times 1)$ Vector Elements $z_t^{(i)}$; i = Series i

$A_1, A_2, B = (nxn)$ Matrix Parameters

$\epsilon_t = (nx1)$ Vector of Independent, Normally Distributed Random Variables with Mean Zero and Variance One

แนวทางที่ 3: Dissaggregation Approach

วิธีนี้เริ่มจากการสังเคราะห์ 1 อนุกรมเวลา เช่น อนุกรมเวลาของ 1 สถานี หรืออนุกรมเวลารายปี หลังจากนั้น จึงสังเคราะห์อนุกรมเวลาของสถานีอื่นจากอนุกรมเวลาของสถานีแรกที่สังเคราะห์ขึ้นตามวิธีการ Disaggregation หรือสังเคราะห์อนุกรมเวลารายเดือนจากอนุกรมเวลารายปีที่สังเคราะห์ขึ้น การสังเคราะห์อนุกรมเวลาต้องรักษา Time and Space Dependent Structure รายละเอียดของวิธีนี้จะกล่าวถึงในบทที่ 8

$$Y=AX+B\epsilon \quad [3.62]$$

เมื่อ $Y =$ Column Matrix of Seasonal Values ซึ่งผลรวมของ Y เท่ากับ X เมทริกซ์ Y มีขนาดเท่ากับ

$[(w*k)x1]$, $w =$ No.of Seasons, $k =$ No.of Time Series

$X =$ Annual Value ซึ่งเป็นเมทริกซ์ขนาด $(kx1)$

$A =$ Matrix Parameters of Size $[(w*k) x k]$

$B =$ Matrix Parameters of Size $[(w*k) x (w*k)]$

$\epsilon =$ Vector of Independent, Normally Distributed Random Variables with Mean Zero

and Variance One with Size $[(w*k) x 1]$

3.10.2 การพยากรณ์ (Forecasting)

การสังเคราะห์ข้อมูลคือกระบวนการในการสร้างอนุกรมเวลาขึ้นมาใหม่ ให้มีขนาดเท่ากับอนุกรมเวลาตัวอย่าง $(t=1, 2, \dots, N)$ ส่วนการพยากรณ์ต่างจากการสังเคราะห์อนุกรมเวลาตรงที่ การพยากรณ์คือการคาดการณ์ค่าของอนุกรมเวลาล่วงหน้าว่าจะมีค่าเท่าใดในอนาคตที่เวลา $t+1$ จากค่าเหตุการณ์ในปัจจุบันที่เวลา t ที่ทราบแล้ว ดังรูปที่ 3.9

ตัวอย่างการพยากรณ์ด้วย แบบจำลอง AR(1)

$$z_t = \phi_1 z_{t-1} + \epsilon_t \quad [3.63]$$

$$z_{t+1} = \phi_1 z_t + \epsilon_{t+1} \quad [3.64]$$

กำหนดให้ Expected Value of z_{t+1} เมื่อทราบค่า z_t เท่ากับ $[z_{t+1}]$

$$[z_{t+1}] = Z_t(1) = E(z_{t+1}|z_t) \quad [3.65]$$

$[z_t|z_{t-1}] = z_t$ เนื่องจากที่เวลา t จะทราบค่า z_{t-1}

และ $[\epsilon_{t+1}] = 0$

ดังนั้น

$$[z_{t+1}] = \phi_1 z_t \quad \text{หรือ} \quad z_t(1) = \phi_1 z_t \quad [3.66]$$

$$[z_{t+2}] = \phi_1 [z_{t+1}] \quad \text{หรือ} \quad z_t(2) = \phi_1 z_t(1) \quad [3.67]$$

$$[z_{t+3}] = \phi_1 [z_{t+2}] \quad \text{หรือ} \quad z_t(3) = \phi_1 z_t(2) \quad [3.68]$$

เทคนิคการพยากรณ์ในลักษณะนี้เรียกว่าการพยากรณ์ล่วงหน้า 1 คาบ หรือ One Step Ahead Forecasting ซึ่งมี Lead Time $L=1$

ตัวอย่างการพยากรณ์ด้วยแบบจำลอง ARMA(1,1) ซึ่งมี Lead Time L

$$z_t = \phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad [3.69]$$

$$z_{t+L} = \phi_1 z_{t+L-1} + \varepsilon_{t+L} - \theta_1 \varepsilon_{t+L-1} \quad [3.70]$$

$$[z_{t+L}] = \phi_1 [z_{t+L-1}] + [\varepsilon_{t+L}] - \theta_1 [\varepsilon_{t+L-1}] \quad [3.71]$$

ถ้า $L=1$

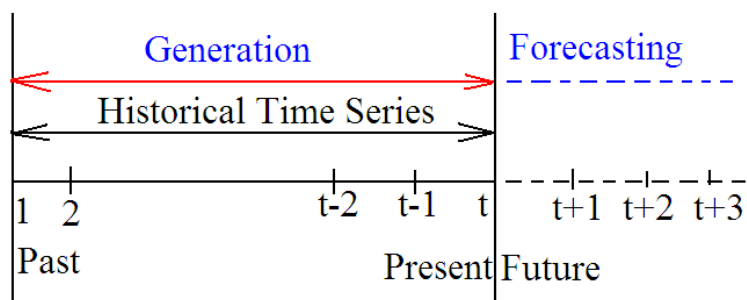
$$[z_{t+1}] = \phi_1 [z_{t+1-1}] + [\varepsilon_{t+1}] - \theta_1 [\varepsilon_{t+1-1}]$$

$$[z_{t+1}] = \phi_1 z_t - \theta_1 \varepsilon_t ; [z_t] = z_t, [\varepsilon_{t+1}] = 0, [\varepsilon_t] = \varepsilon_t$$

$$[z_{t+2}] = \phi_1 [z_{t+1}]; [\varepsilon_{t+1}] = 0$$

$$[z_{t+3}] = \phi_1 [z_{t+2}]$$

$$[z_{t+L}] = \phi_1 [z_{t+L-1}] \quad \text{เมื่อ} \quad L \geq 2$$



รูปที่ 3.9 ข้อแตกต่างระหว่างการสังเคราะห์ข้อมูลและการพยากรณ์

3.11 เอกสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series**. Water Resources Publications. USA. pp.484

3.12 แบบฝึกหัดท้ายบท

- (1) From the 3 monthly time series ($y_{v,\tau}$) in Table 2.4 of Assignment in Chapter 2, calculate and plot fourier series estimates of the following parameters and compare with the non-parametric estimates similar to the example as shown in Figure 1 and 2.
- Mean
 - Standard deviation
 - Lag k autocorrelation coefficients for each series for k=1,2,3

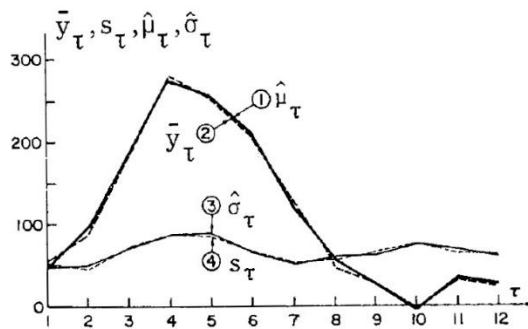


Figure 4.9.

(1) Fitted periodic mean $\hat{\mu}_\tau$, (2) periodic mean \bar{y}_τ , (3) fitted periodic standard deviation $\hat{\sigma}_\tau$ and (4) standard deviation s_τ , for the NBS of Lake Michigan-Huron.

$\bar{y}_\tau, s_\tau = non - parametric estimates$ and $\hat{\mu}_\tau, \hat{\sigma}_\tau = Fourier series estimates$

Figure 1

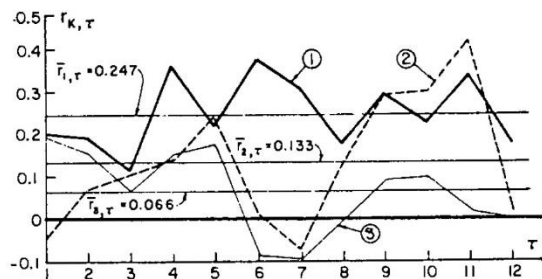


Figure 4.10.

Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized series $z_{v,\tau}$ for the NBS of Lake Michigan-Huron.

$$z_{v,\tau} = \frac{y_{v,\tau} - \bar{y}_\tau}{s_\tau}$$

Figure 2

การจำลองอนุกรมเวลาด้วยแบบจำลอง AUTOREGRESSIVE AUTOREGRESSIVE MODELING

แบบจำลอง Autoregressive หรือเรียกสั้นๆว่าแบบจำลอง AR เป็นแบบจำลองสโตแคสติกที่เป็นที่นิยมตั้งแต่ปี 1960 เป็นต้นมา แบบจำลอง AR ปกตินิยมใช้ในการจำลองทั้งอนุกรมเวลารายปีและอนุกรมเวลาแบบ Periodic เนื่องเป็นแบบจำลองที่เข้าใจง่าย ปรมาจารย์ที่มีส่วนสำคัญในการพัฒนาแบบจำลอง AR ได้แก่

- Thomas and Fiering (1962)
- Yevjevich (1963)
- Box and Jenkins (1970)

4.1 รายละเอียดของแบบจำลอง AR

แบบจำลอง AR สามารถแบ่งออกได้เป็น 2 แบบคือ

- AR Model with Constant Parameters สำหรับอนุกรมเวลารายปี
- AR Model with Periodic Parameters สำหรับอนุกรมเวลาแบบ Periodic

4.1.1 สมการคณิตศาสตร์ของแบบจำลอง AR (Mathematical Formulation of AR Models)

(1) AR Models with Constant Parameters

ให้ y_t = Stationary Normal Time Series with μ, σ^2 and Autoregressive (Markovian) Correlation (~Serial Correlation)

AR(p); Box and Jenkins (1970)

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$$

$$y_t = \mu + \sum_{j=1}^p \phi_j(y_{t-j} - \mu) + \varepsilon_t \quad [4.1]$$

เมื่อ

y_t = Time Dependent Series (Variable) with Normal Distribution and

$$E(y_t) = \mu; \text{Var}(y_t) = \sigma^2$$

ε_t = Independent Normal Variable with Mean = 0 and Variance = σ_ε^2

ϕ_1, \dots, ϕ_p = Autoregressive Coefficients

พารามิเตอร์ของแบบจำลองในสมการที่ 4.1 คือ $\{\mu, \sigma^2, \phi_1, \dots, \phi_p, \sigma_\varepsilon^2\}$ ซึ่งสามารถประมาณได้จากตัวอย่าง σ^2 และ σ_ε^2 มีสหสัมพันธ์ซึ่งกันและกัน

ตารางที่ 4.1 แสดงแบบจำลอง AR(p) ซึ่งเขียนในรูปแบบที่แตกต่างกัน จากสมการที่ 4.1 จะสามารถเขียนแบบจำลอง AR(1) หรือ First Order Autoregressive Model หรือ First Order Markov Model ได้ดังสมการ

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t \quad [4.2]$$

แบบจำลอง AR(1) มีพารามิเตอร์ 4 ตัวคือ $\mu, \sigma^2, \phi_1, \sigma_\varepsilon^2$

ตารางที่ 4.1 แบบจำลอง AR(p) ในรูปแบบต่างๆ ที่เป็นที่ยอมรับใช้กัน

No.	Forms of the AR(p) Modal	Parameters	Reference
1	$y_t = \mu + \sum_{j=1}^P \phi_j (y_t - \mu) + \sigma(1 - R^2)^{\frac{1}{2}} \xi_t$	$\mu, \sigma^2, \phi_1, \dots, \phi_p, R^2$ ($\sigma_\xi^2 = 1$)	<u>Fiering and Jackson (1971); Beard (1967)</u>
2	$y_t = \mu + \sigma z_t$ $z_t = \sum_{j=1}^p \phi_j z_{t-j} + \varepsilon_t \text{ or}$ $z_t = \sum_{j=1}^p \phi_j z_{t-j} + \sigma_\varepsilon \xi_t$ $\varepsilon_t = \sigma_\varepsilon \xi_t; E(\varepsilon_t) = 0; Var(\varepsilon_t) = \sigma_\varepsilon^2$	$\mu, \sigma^2, \phi_1, \dots, \phi_p, \sigma_\varepsilon^2$ ($\sigma_\xi^2 = 1$)	<u>Yevjevich(1972)</u>
3a	$y_t = \mu + \sum_{j=1}^P \phi_j (y_t - \mu) + \varepsilon_t \text{ or}$ $y_t = \mu + \sum_{j=1}^P \phi_j (y_t - \mu) + \sigma_\varepsilon \xi_t$	$\mu, \phi_1, \dots, \phi_p, \sigma_\varepsilon^2$	<u>Box and Jenkins (1970)</u>
3b	$y_t = \mu + z_t$ $z_t = \sum_{j=1}^P \phi_j z_{t-j} + \varepsilon_t \text{ or}$ $z_t = \sum_{j=1}^P \phi_j z_{t-j} + \sigma_\varepsilon \xi_t$	$\mu, \phi_1, \dots, \phi_p, \sigma_\varepsilon^2$ ($\sigma_\xi^2 = 1$)	<u>Box and Jenkins (1970)</u>
	$y_t =$ Stationary Normal Time Series with μ, σ^2 $\varepsilon_t =$ Independent Normal Variable with $\mu = 0$ and σ_ε^2 $\phi_1, \dots, \phi_p =$ Autoregressive Coefficients $\{\mu, \sigma^2, \phi_1, \dots, \phi_p, \sigma_\varepsilon^2\}$ are Parameters of the Model.		

(2) Periodic AR Model (Periodic Mean and Variance)

- Constant Autoregressive Parameter (ϕ_1, \dots, ϕ_p)

$$y_{v,\tau} = \mu_\tau + \sigma_\tau z_{v,\tau} \quad [4.3]$$

$$z_{v,\tau} = \sum_{j=1}^p \phi_j z_{v,\tau-j} + \varepsilon_{v,\tau} \quad [4.4]$$

$y_{v,\tau}$ = Dependent Normal Variables with μ_τ, σ_τ^2

$z_{v,\tau}$ = Dependent Normal Variables with Mean = 0 and Variance = 1

$\varepsilon_{v,\tau}$ = Independent Normal Variable with Mean = 0 and Variance = $\sigma_{\varepsilon,\tau}^2$

พารามิเตอร์ของแบบจำลอง $\{\mu_\tau, \sigma_\tau^2, \phi_1, \dots, \phi_p, \sigma_{\varepsilon,\tau}^2; \tau = 1, \dots, \omega\}$

- Periodic Autoregressive Parameter ($\phi_{j,\tau}$) $j = 1, \dots, p$ and $\tau = 1, \dots, \omega$

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} + \sigma_{\varepsilon,\tau} \xi_{v,\tau} \quad [4.5]$$

พารามิเตอร์ของแบบจำลอง $\{\mu_\tau, \sigma_\tau^2, \phi_{1,\tau}, \dots, \phi_{p,\tau}, \sigma_{\varepsilon,\tau}^2; \tau = 1, \dots, \omega\}$

4.1.2 Properties of AR Models with Constant Parameters

แบบจำลอง AR แบบพารามิเตอร์คงที่มีพารามิเตอร์ดังนี้

$E(y_t) = \mu$ $E(\varepsilon_t) = 0$ $Var(y_t) = \sigma^2$ $Var(\varepsilon_t) = \sigma_\varepsilon^2$	พารามิเตอร์เหล่านี้สามารถประมาณได้จากอนุกรมเวลาตัวอย่าง
--	---

$$\sigma_\varepsilon^2 = \sigma^2 (1 - \sum_{j=1}^p \phi_j \rho_j) \quad [4.6]$$

$\phi_j = j^{\text{th}}$ Autoregressive Coefficients

$\rho_j = \text{Lag } j \text{ Autocorrelation Coefficients of } y_t$

$$\text{AR}(1) : \sigma_\varepsilon^2 = \sigma^2 (1 - \phi_1 \rho_1) \quad [4.7]$$

$$\text{AR}(2) : \sigma_\varepsilon^2 = \sigma^2 (1 - \phi_1 \rho_1 - \phi_2 \rho_2) \quad [4.8]$$

(1) Autocorrelation Function (ρ_k) คือพรรณนิตที่แสดง Dependent Structure ของอนุกรมเวลา มีชื่อเรียกได้หลายแบบดังนี้

- ACF
- Lag k Serial Correlation
- Yule-Walker Equation (Model Correlogram)

$$\rho_k = \sum_{j=1}^p \phi_j \rho_{k-j}; k > 0 \quad [4.9]$$

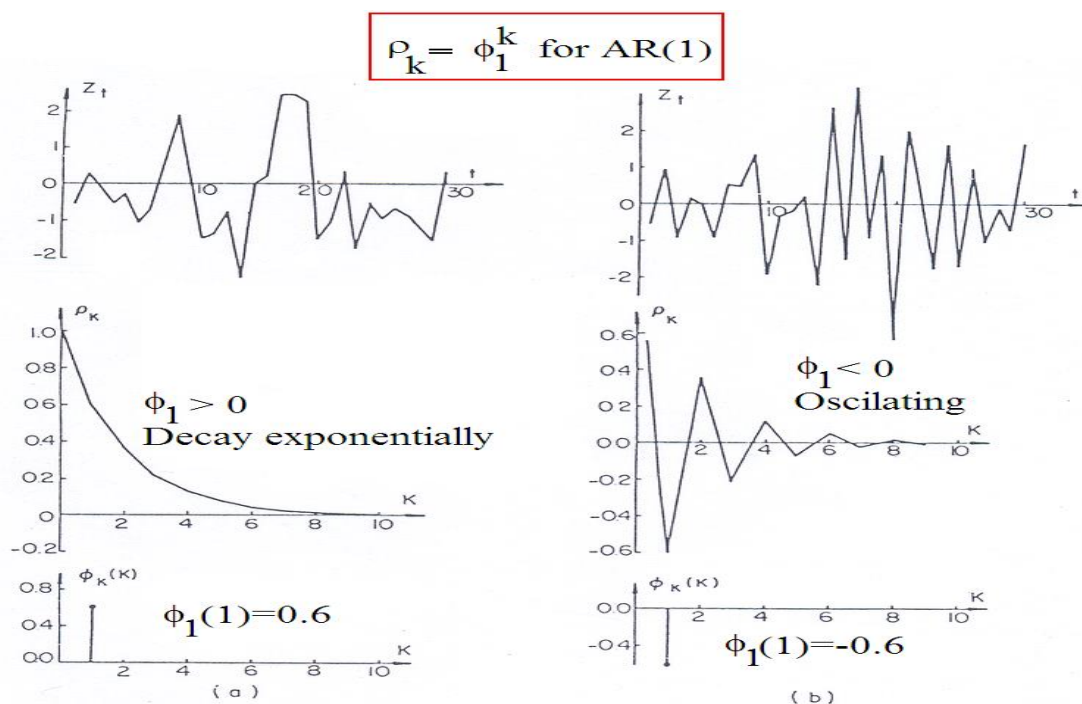
วิธีการพิสูจน์สมการ Yule-Walker [4.9] ทำได้ดังนี้

$$\text{จากสมการ[4.1], } y_t - \mu = \sum_{j=1}^p \phi_j (y_{t-j} - \mu) + \varepsilon_t \quad [4.10]$$

คุณสมบัติที่ [4.10] ด้วย $(y_{t-k} - \mu)$ แล้ว Take $E(\cdot)$ หลังจากนั้นหารด้วย $Var(y_t)$ หรือ σ^2 จะได้สมการ Yule-Walker ดังสมการ [4.9] สมการ Yule-Walker หรือ ACF มีประโยชน์สำหรับ (1) หา Order ของแบบจำลอง และ (2) เปรียบเทียบ Sample Correlogram กับ Model Correlogram สำหรับ AR(1) Model Correlogram แสดงอยู่ในตารางที่ 4.2 รูปที่ 4.1 แสดงอนุกรมเวลา ACF และ Partial ACF ของ AR(1) ซึ่งมีพารามิเตอร์ต่างๆ

ตารางที่ 4.2 Model Correlogram สำหรับ AR(1) : $\rho_k = \phi_1 \rho_{k-1} ; k > 0$

k	$\rho_k = \phi_1 \rho_{k-1}$	หมายเหตุ
1	$\rho_1 = \phi_1$	$\rho_0 = 1$
2	$\rho_2 = \phi_1^2$	
.		
k	$\rho_k = \phi_1^k$	



รูปที่ 4.1 The time series z_t , autocorrelation functions ρ_k and partial autocorrelation function $\phi_k(k)$ for AR(1) models with parameters (a) $\mu = 0, \sigma_\varepsilon^2 = 1$ and $\phi_1 = 0.60$ and (b) $\mu = 0, \sigma_\varepsilon^2 = 1$ and $\phi_1 = -0.60$

$$\text{AR}(1) : \quad \sigma_{\varepsilon}^2 = \sigma^2(1 - \phi_1^2) \quad \dots\dots\text{Biased} \quad [4.11]$$

$$\sigma_{\varepsilon}^2 = \frac{N}{N-1} \sigma^2(1 - \phi_1^2) \quad \dots\dots\text{Unbiased} \quad [4.12]$$

$$\text{AR}(2) : \text{ For } k=1 \quad ; \quad \rho_1 = \frac{\phi_1}{1-\phi_2} \quad [4.13]$$

$$k=2 \quad ; \quad \rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2 \quad [4.14]$$

(see appendix)

$$\sigma_{\varepsilon}^2 = \sigma^2 \frac{(1+\phi_2)}{(1-\phi_2)} [(1 - \phi_1^2)^2 - \phi_2^2] \quad \dots\text{Biased} \quad [4.15]$$

$$\sigma_{\varepsilon}^2 = \frac{N}{(N-2)} \sigma^2 \frac{(1+\phi_2)}{(1-\phi_2)} [(1 - \phi_1^2)^2 - \phi_2^2] \quad \dots\text{Unbiased} \quad [4.16]$$

$$\phi_1 = \rho_1 \frac{(1-\rho_2)}{(1-\rho_1^2)}$$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$$

จงพิสูจน์สมการที่ 4.13 และ 4.14 ของแบบจำลอง AR(2) จากสมการ Yule-Walker [4.9]

สำหรับ AR(2);

$$\begin{aligned} \text{ถ้า } k=1, \rho_1 &= \phi_1\rho_{1-1} + \phi_2\rho_{1-2} \\ &= \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1 + \phi_2\rho_1 \\ \rho_1 &= \frac{\phi_1}{1-\phi_2} \end{aligned}$$

$$\begin{aligned} \text{ถ้า } k=2, \rho_2 &= \phi_1\rho_{2-1} - \phi_2\rho_{2-2} = \phi_1\rho_1 - \phi_2 \\ &= \frac{\phi_1^2}{1-\phi_2} - \phi_2 \end{aligned}$$

(2) PACF (Partial Autocorrelation Function or Partial Correlogram) $\phi_k(k)$

คือกรณีที่แสดง Dependent Structure ของอนุกรมเวลาเช่นเดียวกับ ACF มีประโยชน์ในการหาชนิดและลำดับของแบบจำลอง PACF คือค่าสัมประสิทธิ์ Autoregressive ลำดับที่ k ของแบบจำลอง AR(p) เมื่อ p=k ดังสมการ

$$\rho_j = \phi_1(k)\rho_{j-1} + \phi_2(k)\rho_{j-2} + \dots + \phi_k(k)\rho_{j-k}; j = 1, \dots, k; k = p \quad [4.17]$$

$$\rho_j = \sum_{i=1}^k \phi_i(k)\rho_{j-i}; j = 1, \dots, k; k = p$$

$$\phi_j(k) = j^{\text{th}} \text{ autoregressive coefficient of AR}(k)$$

$$\phi(k) = k^{\text{th}} \text{ autoregressive coefficient of AR}(k) \Rightarrow \text{PACF}$$

สำหรับ AR(1); ถ้า k=1; $\phi_1(1) = \rho_1$

$$\phi_k(k) = 0, \text{ เมื่อ } k > 1$$

สำหรับ AR(2); ถ้า k=2; $\phi_2(2) = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$

$$\phi_1(1) = \rho_1$$

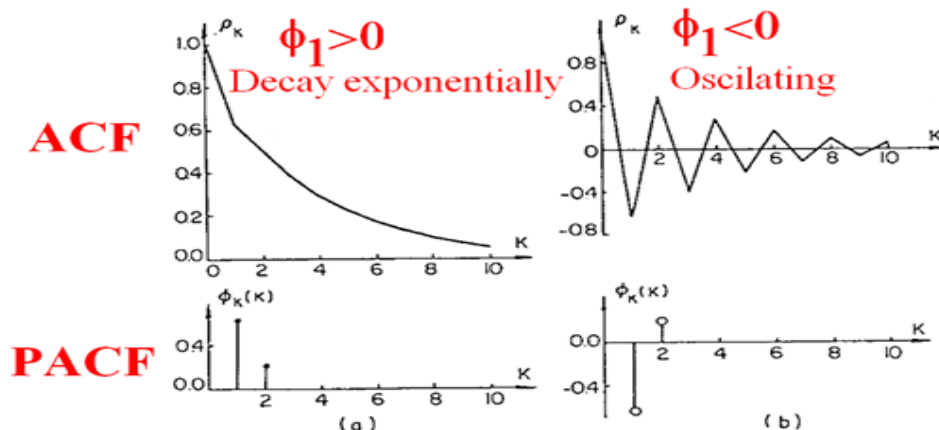
$$\phi_k(2) = 0, \text{ เมื่อ } k > 2$$

สำหรับ AR(p); $\phi_k(k) \neq 0, \text{ เมื่อ } k \leq p; \phi_k(k) = 0, \text{ เมื่อ } k > p$

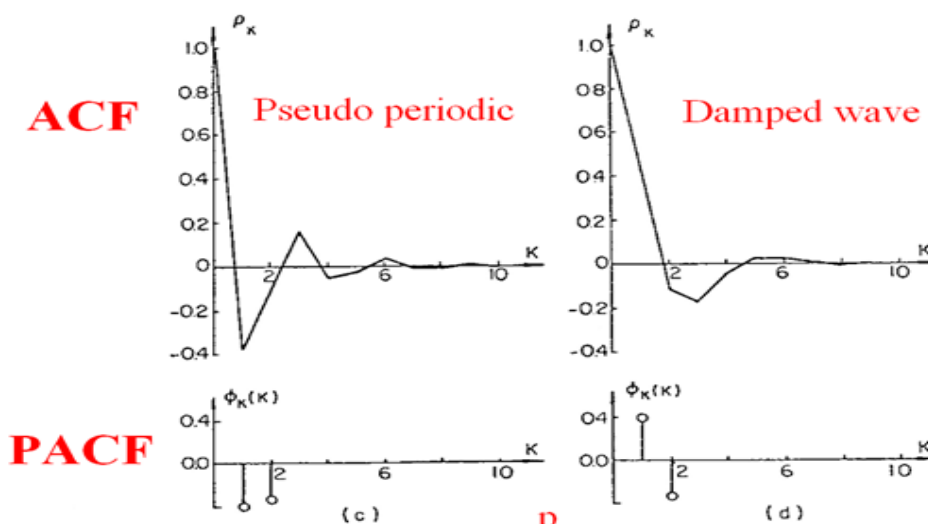
รูปที่ 4.2 แสดง ACF และ PACF ของ AR(2) ซึ่งมีพารามิเตอร์แตกต่างกัน 4 แบบ

ACF and PACF for AR(2)

$$\phi_1^2 + 4\phi_2 \geq 0$$



$$\phi_1^2 + 4\phi_2 < 0$$



$$\rho_k = \sum_{j=1}^p \phi_j \rho_{k-j}; k > 0$$

รูปที่ 4.2 ACF และ PACF ของ AR(2) ซึ่งมีพารามิเตอร์แตกต่างกัน 4 แบบ

(3) Parameters Stationary Conditions for AR(p) with Constant Parameter

Stationary Condition ของ AR(p) จะสามารถหาได้จากเกณฑ์ที่ว่า ถ้ารากของสมการ Characteristic ในสมการที่ 4.18 มีค่าอยู่ภายในวงกลมซึ่งมีรัศมี 1 หน่วย (Unit Circle) แสดงว่าพารามิเตอร์ของ AP(p) อยู่ใน Stationary Condition

$$u^p - \phi_1 u^{p-1} - \phi_2 u^{p-2} \dots \dots \phi_p = 0$$

$$|u_i| < 1 \text{ เมื่อ } i = 1, \dots, p \tag{4.18}$$

สำหรับแบบจำลอง AR(2) ดังสมการ: $z_t = 0.5z_{t-1} + 0.2z_{t-2} + \varepsilon_t$

Characteristic Equation คือ

$$\begin{aligned} u^2 - 0.5u - 0.2 &= 0 \\ |u_1| &= |0.762| < 1 \\ |u_2| &= |-0.261| < 1 \end{aligned}$$

Stationary Condition ของ AR(2) สามารถเขียนในรูปแบบอื่นๆ ได้ดังนี้

$\phi_1 + \phi_2 < 1$	$-1 < \rho_1 < 1$
$\phi_2 - \phi_1 < 1$	$-1 < \rho_2 < 1$
$-1 < \phi_2 < 1$	$\rho_1^2 < \frac{1 + \rho_2}{2}$

Stationary Condition ของ AR(1) คือ

$$\begin{aligned} u - \phi_1 &= 0 \\ u &= \phi_1 \\ -1 < \phi_1 < 1 &\text{ หรือ } -1 < \phi_2 < 1 \end{aligned}$$

4.1.3 Properties of AR Model with Periodic Parameters

แบบจำลอง PAR(p) ซึ่งมี Periodic Parameter คือ

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} + \varepsilon_{v,\tau} \quad [4.19]$$

เมื่อ

$$z_{v,\tau} = \left(\frac{y_{v,\tau} - \mu_\tau}{\sigma_\tau} \right)$$

$\phi_{j,\tau} = j^{\text{th}}$ Periodic Autoregressive Parameters สำหรับคาบที่ τ เมื่อ $\tau = 1, \dots, \omega$

$\varepsilon_{v,\tau} =$ Random Error ซึ่งมี Mean=0 และ Periodic Variance = $\sigma_{\varepsilon,\tau}^2$

ดังนั้นสมการที่ [4.18] สามารถเขียนในอีกรูปแบบหนึ่งได้ดังนี้

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} + \sigma_{\varepsilon,\tau} \xi_{v,\tau} \quad [4.20]$$

เมื่อ

$z_{v,\tau}$ คืออนุกรมปกติมาตรฐานที่มี Mean = 0 และ Variance = 1

$\xi_{v,\tau}$ คือ Random Error ที่มี Mean = 0 และ Variance = 1

$$\sigma_{\varepsilon,\tau}^2 = 1 - \sum_{j=1}^p \phi_{j,\tau} \rho_{j,\tau} \quad [4.21]$$

สำหรับ AR(1) Periodic Variance คือ

$$\sigma_{\varepsilon,\tau}^2 = 1 - \phi_{1,\tau} \rho_{1,\tau} \quad [4.22]$$

สำหรับ AR(2) Periodic Variance คือ

$$\sigma_{\varepsilon,\tau}^2 = 1 - \phi_{1,\tau} \rho_{1,\tau} - \phi_{2,\tau} \rho_{2,\tau} \quad [4.23]$$

Model Correlogram ของ Periodic AR(p) คือ

$$\rho_{k,\tau} = \sum_{j=1}^p \phi_{j,\tau} \rho_{|k-j|,\tau-l_j} ; k > 0 \quad [4.24]$$

เมื่อ

$$l_j = \min(k, j)$$

$$\rho_{0,\tau} = 1$$

สำหรับ AR(1); เมื่อ $k=1$; $\rho_{1,\tau} = \phi_{1,\tau} \rho_{0,\tau-1} = \phi_{1,\tau}$; $\rho_{0,\tau-1} = 1$

สำหรับ AR(2); เมื่อ $k=1$; $\rho_{1,\tau} = \phi_{1,\tau} \rho_{0,\tau-1} + \phi_{2,\tau} \rho_{1,\tau-1}$; $\rho_{0,\tau-1} = 1$

เมื่อ $k=2$; $\rho_{2,\tau} = \phi_{1,\tau} \rho_{1,\tau-1} + \phi_{2,\tau} \rho_{0,\tau-2}$; $\rho_{0,\tau-2} = 1$

4.2 แบบจำลอง AR สำหรับอนุกรมเวลารายปี (AR Modeling of Annual Time Series)

4.2.1 แบบจำลอง AR

ให้ x_t คืออนุกรมเวลาแบบ Non-normal ต้องแปลงให้เป็นอนุกรมเวลาแบบ Normal (y_t) โดยใช้ฟังก์ชันการแปลงที่เหมาะสม ดังนี้

$$y_t = g(x_t) \quad [4.25]$$

เมื่อ

$g(-)$ คือ ฟังก์ชันการแปลงให้เป็นอนุกรมเวลาแบบ Normal เช่น Log หรือ Square Root หรือ

Power Function

แบบจำลอง AR สำหรับอนุกรมเวลา y_t คือ

$$y_t = \mu + z_t \quad [4.26]$$

เมื่อ

y_t คืออนุกรมเวลาแบบ Normal ซึ่งมี Mean เท่ากับ μ และ Variance เท่ากับ σ^2

z_t คืออนุกรมเวลาแบบ Normal ซึ่งมี Mean กับ 0 และ Variance เท่ากับ 1 และมีรูปแบบจำลองแบบ

AR(p)

ถ้า z_t เป็น AR (0)

$$z_t = \varepsilon_t \quad [4.27]$$

ถ้า z_t เป็น AR (1)

$$z_t = \phi_1 z_{t-1} + \varepsilon_t \quad [4.28]$$

ถ้า z_t เป็น AR(2)

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \varepsilon_t \quad [4.29]$$

ถ้า z_t เป็น AR(p)

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + \varepsilon_t \quad [4.30]$$

4.2.2 การประมาณค่าพารามิเตอร์ของแบบจำลอง Annual AR (Parameter Estimation for Annual AR Models)

แบบจำลอง Annual AR(p) มีพารามิเตอร์เท่ากับ p+3 ตัวคือ $\mu, \sigma^2, \phi_1, \dots, \phi_p, \sigma_\varepsilon^2$ วิธีการประมาณค่าพารามิเตอร์ที่นิยมใช้มี 2 วิธีคือ โดยวิธี Moment และวิธี Maximum Likelihood

(1) การประมาณค่าพารามิเตอร์โดยวิธี Moment

$$\hat{\mu} = \bar{y} = \frac{1}{N} \sum_{t=1}^N y_t \quad [4.31]$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{(N-1)} \sum_{t=1}^N (y_t - \bar{y})^2 \quad [4.32]$$

พารามิเตอร์ ϕ_1, \dots, ϕ_p จะสามารถหาได้จากสมการ Yule-Walker หรือสมการ [4.9] โดยการประมาณค่าจากอนุกรมเวลาตัวอย่าง ดังนี้

$$r_k = \hat{\phi}_1 r_{k-1} + \hat{\phi}_2 r_{k-2} + \dots + \hat{\phi}_p r_{k-p}; k > 0 \quad [4.33]$$

เมื่อ r_1, \dots, r_k คือ Sample Autocorrelation Coefficients

กรณี AR(1), p=1 จะได้

$$\hat{\phi}_1 = r_1 \quad [4.34]$$

กรณี AR(2), p=2 จะได้

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2} \quad \text{และ} \quad \hat{\phi}_2 = \frac{(r_2-r_1^2)}{1-r_1^2} \quad [4.35]$$

ตัวประมาณค่าแบบ Unbiased ของ Variance of Error ($\hat{\sigma}_\varepsilon^2$) ของแบบจำลอง AR(p) จะหาได้จากสมการ

$$\hat{\sigma}_\varepsilon^2 = \frac{N\hat{\sigma}^2}{(N-p)} (1 - \sum_{j=1}^p \hat{\phi}_j r_j) \quad [4.36]$$

สำหรับ AR(1)

$$\hat{\sigma}_\varepsilon^2 = \frac{N\hat{\sigma}^2}{(N-1)} (1 - \hat{\phi}_1^2) \quad [4.37]$$

สำหรับ AR(2) จากสมการ [4.35] และสมการ [4.36] จะได้

$$\hat{\sigma}_\varepsilon^2 = \frac{N\hat{\sigma}^2}{(N-2)} \left(\frac{1+\hat{\phi}_2}{1-\hat{\phi}_2} \right) \{ (1 - \hat{\phi}_2^2) - \hat{\phi}_1^2 \} \quad [4.38]$$

การหาความน่าเชื่อถือของพารามิเตอร์ที่ประมาณได้ (Reliability of Estimated Parameters)

การหาความน่าเชื่อถือของพารามิเตอร์ที่ประมาณได้ คือการหาค่าช่วงความเชื่อมั่น (Confident Interval) ของพารามิเตอร์นั่นเอง

การหาช่วงความเชื่อมั่น $(1 - \alpha)$ ของ μ

$$\text{Var}(\bar{y}) = \frac{\hat{\sigma}^2}{N} \left[1 + \frac{2}{N} \sum_{k=1}^{N-1} (N-k) \hat{\rho}_k \right] \quad [4.39]$$

กรณี AR(1) $\hat{\rho}_k = \hat{\phi}_1^k$ จะได้ว่า

$$Var(\bar{y}) = \frac{\hat{\sigma}^2}{N^2(1-\hat{\phi}_1)^2} [(1-\hat{\phi}_1^2)N - 2\hat{\phi}_1(1-\hat{\phi}_1^N)] \quad [4.40]$$

จากตัวประมาณค่า \bar{y} และ $Var(\bar{y})$ จะหาช่วงความเชื่อมั่นของ $(1-\alpha)$ ของ μ ได้ดังนี้

$$(\bar{y} - t_{(N-1, 1-\alpha/2)} s(\bar{y})) < \mu < (\bar{y} + t_{(N-1, 1-\alpha/2)} s(\bar{y})) \quad [4.41]$$

เมื่อ

$t_{(N-1, 1-\alpha/2)}$ คือ ค่าสถิติ t ที่ Significant Level เท่ากับ α และ Degree of Freedom เท่ากับ N-1
 $s(\bar{y})$ คือ Standard Deviation ของ \bar{y} ซึ่งเท่ากับ $\sqrt{Var(\bar{y})}$

การหาช่วงความเชื่อมั่น $(1-\alpha)$ ของ σ^2

S^2 มีการแจกแจงแบบ Chi-Square

$$\chi^2 = \frac{(N-1)S^2}{\sigma^2} \quad [4.42]$$

เมื่อ χ^2 คือค่า Chi-Square ที่มี Degree of Freedom เท่ากับ (N-1)

ช่วงความเชื่อมั่นที่ $(1-\alpha)$ ของ σ^2 จะหาได้จากการแจกแจงแบบ Chi-Square ดังนี้

$$\begin{aligned} \chi_{1-\frac{\alpha}{2}, N-1}^2 &< \chi^2 < \chi_{\frac{\alpha}{2}, N-1}^2 \\ \chi_{1-\frac{\alpha}{2}, N-1}^2 &< \frac{(N-1)S^2}{\sigma^2} < \chi_{\frac{\alpha}{2}, N-1}^2 \\ \frac{(N-1)S^2}{\chi_{\frac{\alpha}{2}, N-1}^2} &< \sigma^2 < \frac{(N-1)S^2}{\chi_{1-\frac{\alpha}{2}, N-1}^2} \end{aligned} \quad [4.43]$$

การหาช่วงความเชื่อมั่น $(1-\alpha)$ ของ ϕ_p

Box และ Jenkins (1970) เสนอวิธีคำนวณหา Variance-Covariance Matrix ของ $\hat{\phi}_1, \dots, \hat{\phi}_p$ ดังนี้

สมการ

$$\underline{\hat{V}}(\underline{\hat{\phi}}) = (N-p)^{-1} (1 - \sum_{j=1}^p \hat{\phi}_j r_j) \begin{bmatrix} 1 & r_1 & r_2 & \dots & r_{p-1} \\ r_1 & 1 & r_1 & \dots & r_{p-2} \\ r_2 & r_1 & 1 & \dots & r_{p-3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & r_{p-3} & \dots & 1 \end{bmatrix}^{-1} \quad [4.44]$$

เมื่อ $\underline{\hat{V}}(\underline{\hat{\phi}})$ คือ Variance-Covariance Matrix ของ $\hat{\phi}_1, \dots, \hat{\phi}_p$

สำหรับ AR(1)

$$\underline{\hat{V}}(\underline{\hat{\phi}}) = Var(\hat{\phi}_1) = \frac{(1-\hat{\phi}_1^2)}{N-1} \quad [4.45]$$

สำหรับ AR(2)

$$Var(\hat{\phi}_1) = Var(\hat{\phi}_2) = \frac{(1-\hat{\phi}_2^2)}{N-2} \quad [4.46]$$

$$Cov(\hat{\phi}_1, \hat{\phi}_2) = \frac{-\hat{\phi}_1(1+\hat{\phi}_2)}{N-2} \quad [4.47]$$

ช่วงความเชื่อมั่นที่ $(1 - \alpha)$ ของ ϕ_j

$$(\hat{\phi}_j - u_{1-\alpha/2} s(\hat{\phi}_j)) < \phi_j < (\hat{\phi}_j + u_{1-\alpha/2} s(\hat{\phi}_j)) \quad [4.48]$$

เมื่อ

$u_{1-\alpha/2}$ คือ z-score หรือ Standardized Normal Distribution ที่ระดับนัยสำคัญ α

การช่วงความเชื่อมั่นที่ $(1 - \alpha)$ ของ σ_ε^2

$$\frac{N\hat{\sigma}^2}{(N-p)} (1 - \sum_{j=1}^p \phi_{j(lower)} r_j) < \sigma_\varepsilon^2 < \frac{N\hat{\sigma}^2}{(N-p)} (1 - \sum_{j=1}^p \phi_{j(upper)} r_j) \quad [4.49]$$

(2) วิธี Maximum Likelihood

Box และ Jenkins (1970) เสนอแนะว่าวิธีการคำนวณหาผลบวกของ Cross-Products ของอนุกรมเวลา z_t ดังสมการ

$$z_i z_j + z_{i+1} z_{j+1} + \dots + z_{N+1-i} z_{N+1-j} \quad [4.50]$$

และได้นิยามเทอม Sum of Cross-Products (D_{ij}) ไว้ดังนี้

$$D_{ij} = D_{ji} = \frac{N}{(N+2-i-j)} \sum_{l=0}^{N+1-i-j} z_{i+l} z_{j+l} \quad [4.51]$$

ค่าประมาณของ $\phi_1 \dots \phi_p$ โดยวิธี Maximum Likelihood จะหาได้โดยการแก้สมการ

$$D_{1j} = \hat{\phi}_1 D_{j2} + \hat{\phi}_2 D_{j3} + \dots + \hat{\phi}_p D_{j,p+1}; j = 2, \dots, p+1 \quad [4.52]$$

สำหรับ AR(1) จะได้

$$\hat{\phi}_1 = \frac{D_{12}}{D_{22}} \quad [4.53]$$

สำหรับ AR(2) จะได้

$$\hat{\phi}_1 = \frac{D_{12}D_{33} - D_{13}D_{23}}{D_{22}D_{33} - D_{23}^2} \quad \text{and} \quad \hat{\phi}_2 = \frac{D_{13}D_{22} - D_{12}D_{23}}{D_{22}D_{33} - D_{23}^2} \quad [4.54]$$

ค่าประมาณของ Variance ของ White Noise (σ_ε^2) คือ

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{(N-p)} (D_{11} - \sum_{j=1}^p \hat{\phi}_j D_{1,j+1}) \quad [4.55]$$

สำหรับ AR(1)

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{(N-1)} (D_{11} - \hat{\phi}_1 D_{12}) \quad [4.56]$$

สำหรับ AR(2)

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{(N-2)} (D_{11} - \hat{\phi}_1 D_{12} - \hat{\phi}_2 D_{13}) \quad [4.57]$$

การหาความน่าเชื่อถือของพารามิเตอร์ที่ประมาณได้ (Reliability of Estimated Parameters)

ช่วงความเชื่อมั่นของ μ ที่ $(1 - \alpha)$ ของ μ จะหาได้จากสมการ [4.41]

Variance-Covariance Matrix ของ $\hat{\phi}_1, \dots, \hat{\phi}_p$ คือ

$$\underline{\hat{V}}(\underline{\hat{\phi}}) = \hat{\sigma}_\varepsilon^2 D_p^{-1} \quad [4.58]$$

$$D_p = \begin{bmatrix} D_{22} & D_{23} & \cdot & D_{2,p+1} \\ D_{32} & D_{33} & \cdot & D_{3,p+1} \\ \cdot & \cdot & \cdot & \cdot \\ D_{p+1,2} & D_{p+1,3} & \cdot & D_{p+1,p+1} \end{bmatrix} \quad [4.59]$$

สำหรับ AR(1)

$$\underline{\hat{V}}(\underline{\hat{\phi}}) = \text{Var}(\hat{\phi}_1) = \frac{\hat{\sigma}_\varepsilon^2}{D_{22}} \quad [4.60]$$

สำหรับ AR(2)

$$\text{Var}(\hat{\phi}_1) = \frac{\hat{\sigma}_\varepsilon^2 D_{33}}{(D_{22}D_{33} - D_{23}^2)} ; \text{Var}(\hat{\phi}_2) = \frac{\hat{\sigma}_\varepsilon^2 D_{22}}{(D_{22}D_{33} - D_{23}^2)} \quad [4.61]$$

$$\text{Cov}(\hat{\phi}_1, \hat{\phi}_2) = \frac{-\hat{\sigma}_\varepsilon^2 D_{23}}{(D_{22}D_{33} - D_{23}^2)} \quad [4.62]$$

ช่วงความเชื่อมั่นที่ $(1 - \alpha)$ ของ ϕ_j คือ

$$\hat{\phi}_j - u_{\frac{\alpha}{2}} \sqrt{\text{Var}(\hat{\phi}_j)} < \phi_j < \hat{\phi}_j + u_{\frac{\alpha}{2}} \sqrt{\text{Var}(\hat{\phi}_j)} \quad [4.63]$$

ช่วงความเชื่อมั่นที่ $(1 - \alpha)$ ของ σ_ε^2 คือ

$$\frac{1}{(N-p)} (D_{11} - \sum_{j=1}^p \phi_{j(\text{lower})} D_{1,j+1}) < \sigma_\varepsilon^2 < \frac{1}{(N-p)} (D_{11} - \sum_{j=1}^p \phi_{j(\text{upper})} D_{1,j+1}) \quad [4.64]$$

4.2.3 วิธีการทดสอบความเหมาะสมสำหรับแบบจำลอง Annual AR

(1) การทดสอบสมมติฐานของแบบจำลอง (Test on Assumptions of Model)

สมมติฐานในการทดสอบแบบจำลองคือ White Noise (ε_t) เป็นตัวแปรอิสระ (Independent) และมีการแจกแจงแบบปกติ (Normal) จากสมการแบบจำลอง AR(p) จะหาค่าประมาณของ White Noise ได้ดังสมการ

$$\hat{\varepsilon}_t = z_t - \hat{\phi}_1 z_{t-1} - \cdots - \hat{\phi}_p z_{t-p} \quad [4.65]$$

การทดสอบว่า White Noise (ε_t) เป็นตัวแปรอิสระ จะให้ค่าสถิติ Porte Manteau (Q) ดังสมการ

$$Q = N \sum_{k=1}^L r_k^2(\hat{\varepsilon}) \quad [4.66]$$

เมื่อ

$L = \text{Maximum Lag}$ ซึ่งมีค่าประมาณ 10-30% ของขนาดอนุกรมตัวอย่าง (N)

ถ้า $Q < \chi_{\alpha, L-p}^2$ (ค่าสถิติ Chi-Square ซึ่งมีองศาเสรีเท่ากับ $L-p$) แสดงว่า White Noise (ε_t) เป็นตัวแปรอิสระ หรือ แบบจำลองเหมาะสม (Adequate) แต่ถ้า $Q \geq \chi_{L-p}^2$ แสดงว่าแบบจำลองไม่เหมาะสม (Inadequate) และต้องเลือกแบบจำลองใหม่โดยเพิ่มลำดับ (p) ของแบบจำลองเป็น $p+1$ เป็นต้น

ส่วนการทดสอบว่า White Noise (ε_t) มีการแจกแจงแบบปกติหรือไม่ทำได้ 3 วิธีคือ

- พล็อตกราฟ

- Chi-Square Test
- Smirnov-Komogorov Test (ดูหัวข้อ 3.9.2 ในบทที่ 3)

(2) การเปรียบเทียบ Historical และ Model Correlogram

(3) การทดสอบ Parsimony ของพารามิเตอร์

การทดสอบ Parsimony ของพารามิเตอร์จะใช้วิธี Akaike Information Criteria (AIC) ซึ่งเกณฑ์ในการทดสอบคือแบบจำลองที่มีค่า AIC ต่ำสุดจะมีความเหมาะสมด้าน Parsimony ของพารามิเตอร์มากกว่า

$$AIC(p) = N \ln(\hat{\sigma}_\varepsilon^2) + 2p \quad [4.67]$$

4.2.4 การสังเคราะห์ข้อมูลใหม่โดยแบบจำลอง Annual AR (Generation with Annual AR)

แบบจำลอง Annual AR(p)

$$\hat{z}_t = \hat{\phi}_1 \hat{z}_{t-1} + \dots + \hat{\phi}_p \hat{z}_{t-p} + \varepsilon_t \quad [4.68]$$

ε_t คือตัวแปรสุ่มอิสระซึ่งมีการแจกแจงแบบปกติ มี Mean=0 และ Variance= σ_ε^2

กำหนดให้ ξ_t คือตัวแปรสุ่มอิสระซึ่งมีการแจกแจงแบบปกติ มี Mean=0 และ Variance=1 หรือ

Standardized Normal Variate จึงสามารถเขียนสมการ [4.68] ได้ดังนี้

$$\hat{z}_t = \hat{\phi}_1 \hat{z}_{t-1} + \dots + \hat{\phi}_p \hat{z}_{t-p} + \hat{\sigma}_\varepsilon \xi_t \quad [4.69]$$

จากแบบจำลอง Annual AR(p) ในสมการที่ [4.69] กระบวนการในการสังเคราะห์ข้อมูลใหม่มีขั้นตอนดังต่อไปนี้

ขั้นที่ 1- เริ่มจากการสังเคราะห์ Standard Normal Variate (ξ_t) แล้วแทนค่าลงในสมการ [4.69] เพื่อคำนวณหาค่า \hat{z}_1 เนื่องจากทราบค่า z ที่เกิดขึ้นก่อนหน้านั้น คือ z_0, z_1, \dots, z_{p+1} โดยปกติจะสมมติให้มีค่าเท่ากับค่าเฉลี่ย หรือเท่ากับ 0 จะได้ว่า

$$\hat{z}_1 = \hat{\sigma}_\varepsilon \xi_1$$

สังเคราะห์ค่า ξ_2 เพื่อคำนวณหาค่า \hat{z}_2 ดังนี้

$$\hat{z}_2 = \hat{\phi}_1 \hat{z}_1 + \hat{\sigma}_\varepsilon \xi_2$$

ทำการสังเคราะห์ $\hat{z}_3, \hat{z}_4, \dots, \hat{z}_n$ ในทำนองกับวิธีการที่กล่าวแล้ว โดยทั่วไปจะการสังเคราะห์ข้อมูลใหม่เท่ากับ N' และกำหนดให้

$$N' = Ng + Nw \quad [4.70]$$

เมื่อ

Ng คือจำนวนข้อมูลที่ต้องการสังเคราะห์

Nw คือ Warm-up Length ซึ่งโดยปกติจะกำหนดค่า Nw เท่ากับ 50

ขั้นที่ 2- ตัดข้อมูล Nw ตัวแรกออกจากข้อมูล N' ที่สังเคราะห์ขึ้น เพื่อตัด Bias จากการใช้ค่า Pseudo Random Number จะได้อนุกรมเวลาที่สังเคราะห์ขึ้นใหม่ z_t เมื่อ t เท่ากับ $1, \dots, Ng$

ขั้นที่ 3-แปลงค่า z_t ที่สังเคราะห์ขึ้นมากลับเป็นค่า \hat{y}_t โดยใช้สมการการแปลงเช่นเดียวกับที่ใช้แปลง \hat{y}_t เป็น z_t เช่น

$$\hat{y}_t = \bar{y} + z_t \cdot \hat{\sigma}_\varepsilon \quad [4.71]$$

ขั้นที่ 4-แปลง \hat{y}_t กลับเป็น \hat{x}_t โดยใช้ Normalized Inverse Function ดังสมการ

$$\hat{x}_t = g^{-1}(\hat{y}_t) \quad [4.72]$$

ตัวอย่างการจำลองอนุกรมเวลารายปีด้วยแบบจำลอง AR

การจำลองอนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota ประเทศสวีเดน

$$A = 18,076 \text{ mile}^2 \text{ or } 47,439 \text{ km}^2, \bar{Q} = 19,000 \text{ cfs (538 cms.)}$$

อนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota มีจำนวน 171 ปี เริ่มจากปี 1807 แต่ในการจำลองจะเลือกใช้เฉพาะข้อมูลช่วง 50 ปี จากช่วง 1901-1950 ในการวิเคราะห์จะใช้ค่าอัตราส่วนระหว่างอัตราไหลต่อค่าเฉลี่ย ($Q_t/\bar{Q} = \text{Modular Coefficients}$) ดังแสดงในตารางที่ 4.3

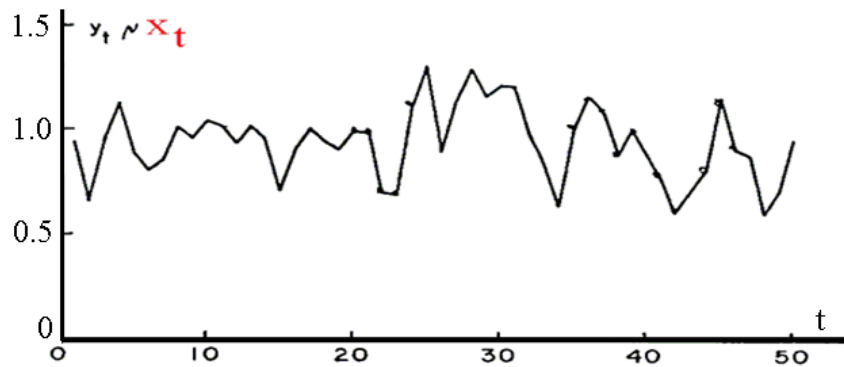
ตารางที่ 4.3 อนุกรมเวลาของน้ำท่าในรูป Modular Coefficients ของแม่น้ำ Gota

Years	1	2	3	4	5	6	7	8	9	10
01-10	0.935	0.662	0.950	1.121	0.880	0.802	0.856	1.080	0.959	1.345
11-20	1.153	0.929	1.158	0.957	0.705	0.905	1.000	0.948	0.907	0.991
21-30	0.994	0.701	0.692	1.086	1.306	0.895	1.149	1.297	1.168	1.218
31-40	1.209	0.974	0.834	0.638	0.991	1.198	1.091	0.892	1.020	0.869
41-50	0.772	0.606	0.739	0.813	1.173	0.916	0.880	0.601	0.720	0.955

[1] Preliminary Analysis and Model Identification

(1a) และ (1b) $\hat{\gamma} = -0.058 \sim 0$ (จากตารางที่ 3.4 $\gamma_{0.02,50} = 0.787, \gamma_{0.01,50} = 0.534$)

(1c) พล็อต x_t ในรูปที่ 4.3 ซึ่งแสดงให้เห็น Dependence Structure ระหว่าง Q_t



รูปที่ 4.3 อนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota สวีเดนระหว่าง 1901-1950 ในรูปของ Modular Coefficients

อนุกรมเวลาช่วงครึ่งหลังมีลักษณะ High Flow ตามด้วย High Flow และ Low Flow ตามด้วย Low Flow แสดงว่าอนุกรมเวลา x_t น่าจะเป็น AR(1) หรือไม่ก็ AR(2) ดังนั้นต้องมีการวิเคราะห์ Correlogram และ Partial Correlogram เพื่อหารูปแบบของแบบจำลอง

(1d) ใช้สมการ [4.71] คำนวณหา Sample Correlogram ดังแสดงรูปที่ 4.4 และตารางที่ 4.4

$$r_k(y) = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y}_t)(y_{t+k} - \bar{y}_{t+k})}{\left[\sum_{t=1}^{N-k} (y_t - \bar{y}_t)^2 \sum_{t=1}^{N-k} (y_{t+k} - \bar{y}_{t+k})^2 \right]^{1/2}} \quad [4.73]$$

เมื่อ $k = 1, \dots, 18 (\sim 0.3 N)$

Anderson(1941): $(1 - \alpha)$ Confidence Interval of $r_k(y)$

$$r_k(1 - \alpha) = \frac{-1 \pm u_{\alpha/2} \sqrt{N-k-1}}{N-k}$$

$$r_k(95\%) = \frac{-1 \pm 1.96 \sqrt{N-k-1}}{N-k}$$

$$r_k(99\%) = \frac{-1 \pm 2.576 \sqrt{N-k-1}}{N-k}$$

Sample Correlogram ในรูปที่ 4.4 แสดงว่า $r_1 \sim 0.4$ อยู่นอกช่วง 95% CI ขณะที่ r_2, r_3, \dots, r_{18} มีค่าแปรผันอยู่รอบค่า 0 หรืออยู่ภายใน 95% CI

Sample Correlogram มีลักษณะคล้าย Model Correlogram ของ AR(2) ในรูปที่ 4.2 (d) เมื่อ $\phi_1 = 0.5, \phi_2 = -0.3$ แสดงว่าแบบจำลอง AR(2) อาจจะเหมาะสม

Partial Autocorrelation Function $\phi_k(k)$

$$\rho_j = \phi_1(k)\rho_{j-1} + \phi_2(k)\rho_{j-2} + \dots + \phi_k(k)\rho_{j-k}; j = 1, \dots, k$$

$$\rho_j = -\rho_j$$

j	ρ_j
1	$\phi_1(k)\rho_0 + \phi_2(k)\rho_1 + \dots + \phi_k(k)\rho_{k-1} = \rho_1$
2	$\phi_1(k)\rho_1 + \phi_2(k)\rho_0 + \dots + \phi_k(k)\rho_{k-2} = \rho_2$
3	$\phi_1(k)\rho_2 + \phi_2(k)\rho_1 + \dots + \phi_k(k)\rho_{k-3} = \rho_3$
.	.
k	$\phi_1(k)\rho_{k-1} + \phi_2(k)\rho_{k-2} + \dots + \phi_k(k)\rho_0 = \rho_k$

ตารางที่ 4.4 Sample Correlogram $r_k(y)$ ของอนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota ประเทศสวีเดน

K	95 percent lower limit	$r_k(y)$	95 percent upper limit
1	-0.298	-0.397	0.257
2	-0.301	-0.013	0.259
3	-0.304	-0.004	0.262
4	-0.308	-0.007	0.264
5	-0.311	-0.108	0.267
6	-0.315	0.013	0.269
7	-0.319	0.103	0.272
8	-3.230	0.076	0.275
9	-0.327	-0.055	0.278
10	-0.331	-0.054	0.281
11	-0.335	-0.106	0.284
12	-0.340	-0.257	0.287
13	-0.345	-0.356	0.291
14	-0.350	-0.038	0.294
15	-0.355	0.06	0.298
16	-0.361	-0.016	0.302
17	-0.366	0.075	0.306
18	-0.372	-0.004	0.310

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \rho_{k-2} \\ \rho_2 & \rho_2 & 1 & \cdot & \rho_{k-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdot & 1 \end{bmatrix} \begin{bmatrix} \phi_1(k) \\ \phi_2(k) \\ \phi_3(k) \\ \cdot \\ \phi_k(k) \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \rho_k \end{bmatrix}$$

$$\text{ถ้า } k = 1: \phi_1(1) = \rho_1$$

$$\text{ถ้า } k = 2$$

$$j=1; \phi_1(2)\rho_0 + \phi_2(2)\rho_1 = \rho_1 \quad [4.74]$$

$$j=2; \phi_1(2)\rho_1 + \phi_2(2)\rho_0 = \rho_2 \quad [4.75]$$

$$\phi_1(2) = \frac{\rho_1(1-\rho_2)}{(1-\rho_1^2)}$$

$$\phi_2(2) = \frac{\rho_2 - \rho_1^2}{(1-\rho_1^2)}$$

$$\phi_k(k) = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_j(k-1)\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_j(k-1)\rho_j} \quad [4.76]$$

$$\phi_j(k) = \phi_j(k-1) - \phi_k(k)\phi_{k-j}(k-1) \quad [4.77]$$

$$\phi_2(2) = \frac{\rho_2 - \phi_1(1)\rho_1}{1 - \phi_1(1)\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}; (\phi_1(1) = \rho_1) \quad [4.78]$$

$$\begin{aligned} \phi_1(2) &= \phi_1(1) - \phi_2(2)\phi_1(1) = \rho_1 - \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \cdot \rho_1 \\ &= \frac{\rho_1 - \rho_1^3 - \rho_1\rho_2 + \rho_1^3}{1 - \rho_1^2} = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \end{aligned} \quad [4.79]$$

ถ้า $k=3$

$$k=1; \quad \hat{\phi}_1(1) = \hat{\rho}_1 = 0.397$$

$$k=2; \quad \hat{\phi}_2(2) = \frac{\hat{\rho}_2 - \hat{\phi}_1(1)\hat{\rho}_1}{1 - \hat{\rho}_1(1)\hat{\rho}_1} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{-0.13 - (0.397)^2}{1 - (0.397)^2} = -0.203$$

$$k=3; \quad \hat{\phi}_3(3) = \frac{\hat{\rho}_3 - \hat{\phi}_1(2)\hat{\rho}_2 - \hat{\phi}_2(2)\hat{\rho}_1}{1 - \hat{\phi}_1(2)\hat{\rho}_1 - \hat{\phi}_2(2)\hat{\rho}_2}$$

$$\hat{\phi}_1(2) = \hat{\phi}_1(1) - \hat{\phi}_2(2)\hat{\phi}_1(1)$$

$$= \hat{\phi}_1(1) \left(1 - \hat{\phi}_2(2) \right) = 0.397(1 + 0.203) = 0.4776$$

$$\hat{\phi}_3(3) = \frac{-0.004 - 0.4776(-0.013) + 0.203(0.397)}{1 - 0.4476(0.397) + 0.203(-0.013)} = \frac{0.0828}{0.8078} = 0.102$$

พิสัย Partial Correlogram

ถ้า $k = 2; j = 1, \dots, k - 1$

$$j=1; \quad \phi_1(2) = \phi_1(1) - \phi_2(2)\phi_1(1)$$

$$\phi_2(2) = \frac{\rho_2 - \phi_1(1)\rho_1}{1 - \phi_1(1)\rho_1}$$

$$\phi_1(1) = \rho_1$$

$$\phi_1(2) = \rho_1 - \phi_2(2)\rho_1$$

$$\phi_2(2) = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_1(2) = \rho_1 - \frac{(\rho_2 - \rho_1^2)}{(1 - \rho_1^2)} \rho_1$$

$$= \frac{\rho_1 - \cancel{\rho_1^2} - \rho_1 \rho_2 + \cancel{\rho_1^2}}{1 - \rho_1^2}$$

$$= \frac{\rho_1(1 - \rho_2)}{(1 - \rho_1^2)}$$

ถ้า $k = 3; j = 1, \dots, k - 1$

$$j=1; \quad \phi_1(3) = \phi_1(2) - \phi_3(3)\phi_2(2)$$

$$j=2; \quad \phi_2(3) = \phi_2(2) - \phi_3(3)\phi_1(2)$$

$$\phi_3(3) = \frac{\rho_3 - \phi_1(2)\rho_2 - \phi_2(2)\rho_1}{1 - \phi_1(2)\rho_1 - \phi_2(2)\rho_2}$$

$$\phi_3(3) = \frac{\rho_3 - \frac{\rho_1(1 - \rho_2)\rho_2}{1 - \rho_1^2} - \left(\frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \right) \rho_1}{1 - \rho_1 \frac{(1 - \rho_2)}{1 - \rho_1^2} \rho_1 - \frac{(\rho_2 - \rho_1^2)}{1 - \rho_1^2} \rho_2}$$

การคำนวณหา $(1 - \alpha)$ Confidence Interval ของ $\phi_k(k)$

$\phi_k(k)$ เมื่อ $k > p$ มีการแจกแจงแบบ Asymptotically Normal with Mean = 0 and Variance = $\frac{1}{N}$

$(1 - \alpha)$ Confidence Interval ของ $\phi_k(k)$ (Box and Jenkins, 1970)

$$\left\{ -\frac{u_\alpha}{\sqrt{N}}, +\frac{u_\alpha}{\sqrt{N}} \right\}$$

95% Confidence Limit ของ $\phi_k(k)$ เมื่อ $N = 50$

$$\text{Lower limit} = -\frac{u_{0.025}}{\sqrt{N}} = -\frac{1.96}{\sqrt{50}} = -0.277$$

$$\text{Upper limit} = +\frac{u_{0.025}}{\sqrt{N}} = +\frac{1.96}{\sqrt{50}} = 0.277$$

(1e) Sample Partial Correlogram

$$\hat{\phi}_k(k) = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_j(k-1)\hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_j(k-1)\hat{\rho}_j}$$

$$\phi_j(k) = \phi_j(k-1) - \phi_k(k)\phi_{k-j}(k-1)$$

$$k = 1, \dots, 18(0.3N)$$

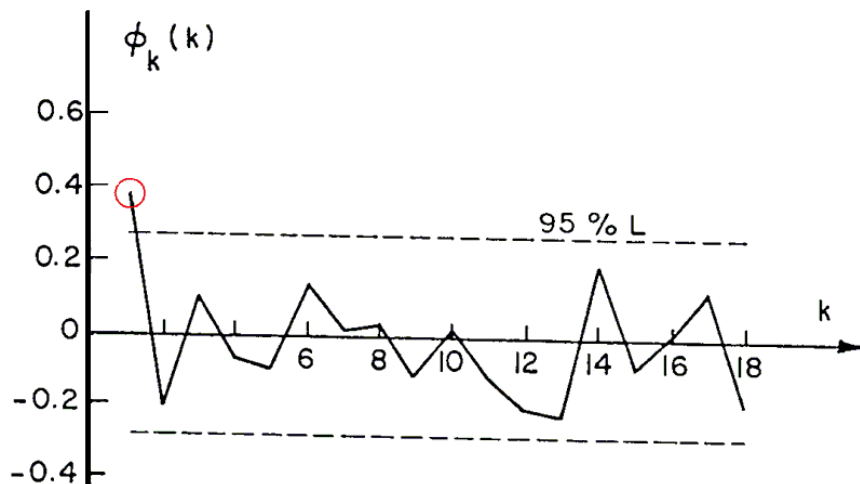
$$\phi(k) = 0 \text{ when } j > k$$

Sample Partial Correlogram $\hat{\phi}_k(k)$ และ 95% Confidence Limits ของอนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota แสดงอยู่ในตารางที่ 4.5 และรูปที่ 4.5 ซึ่งจะเห็นได้ว่ามีเพียง $\hat{\phi}_1(1)$ เท่านั้นที่อยู่นอก 95%

Confidence Limits แสดงว่า AR(1) อาจเหมาะสม

ตารางที่ 4.5 Partial Correlogram ของอนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota

K	95 percent lower limit	$\hat{\phi}_k(k)$	95 percent upper limit
1	-0.227	0.397	0.277
2	-0.227	-0.203	0.277
3	-0.227	0.102	0.277
4	-0.227	-0.061	0.277
5	-0.227	-0.100	0.277
6	-0.227	0.134	0.277
7	-0.227	0.019	0.277
8	-0.227	0.039	0.277
9	-0.227	-0.107	0.277
10	-0.227	0.011	0.277
11	-0.227	-0.117	0.277
12	-0.227	-0.202	0.277
13	-0.227	-0.226	0.277
14	-0.227	0.184	0.277
15	-0.227	-0.083	0.277
16	-0.227	0.001	0.277
17	-0.227	0.124	0.277
18	-0.227	0.203	0.277



รูปที่ 4.5 Partial Correlogram และ 95% Confidence Limits ของอนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Gota

(1f) กราฟอนุกรมเวลาแสดงว่าแบบจำลองอาจเป็น AR(1) หรือ AR(2)

Sample Correlogram แสดงว่าแบบจำลองคือ AR(2)

Partial Correlogram แสดงว่าแบบจำลองคือ AR(1)

ดังนั้นในขั้นนี้จะเลือกแบบจำลอง AR(2) สำหรับการวิเคราะห์ต่อไป

[2] การประมาณค่าพารามิเตอร์แบบจำลอง AR(2)

$$(2a) \quad \bar{y} = 0.9528$$

$$\hat{\sigma}^2 = 0.357$$

$$(2b) \quad \text{หา } \hat{z}_t = y_t - \bar{y}; (t = 1, \dots, 50)$$

$$\begin{cases} \hat{z}_1 = y_1 - 0.9528 = 0.935 - 0.9528 = -0.0178 \\ \hat{z}_2 = y_2 - 0.9528 = 0.662 - 0.9528 = -0.2908 \\ \hat{z}_{50} = y_{50} - 0.9528 = 0.955 - 0.9528 = 0.0022 \end{cases}$$

\hat{z}_t มี Mean = 0 และ Variance = $\hat{\sigma}^2 = 0.357$

(2c) Sample Correlogram $r_k(z) = r_k(y)$ ตามที่คำนวณในข้อ (1d) (ดูตารางที่ 4.4)

การประมาณค่าพารามิเตอร์ของ AR(2) โดยวิธี

(2d') สำหรับ $p = 2$; $D_{ij} = D_{ji}$ เมื่อ $i, j = 1, \dots, p+1 = 1, 2, 3$

(จากสมการที่ 4.62) $D_{11} = 2.079$

(จากสมการที่ 4.65) $D_{12} = 1.037 = D_{21}$

$$\hat{\phi}_1 = 0.542$$

$$\hat{\phi}_2 = -0.114$$

หมายเหตุ คำนวณด้วย VBA code

(2e') สำหรับ $p = 2$

$$\text{(จากสมการที่ 4.68)} \quad \hat{\sigma}_\varepsilon^2 = 0.0325 = \frac{1}{(N-2)} (D_{11} - \hat{\phi}_1 D_{12} - \hat{\phi}_2 D_{13})$$

(2f') สำหรับ $p = 2$

$$\begin{aligned} \text{(จากสมการที่ 4.24)} \quad \hat{\phi}_1 + \hat{\phi}_2 &= 0.542 - 0.114 = 0.428 < 1 \\ \hat{\phi}_2 - \hat{\phi}_1 &= -0.114 - 0.542 = -0.656 < 1 \\ -1 < \hat{\phi}_2 &= -0.114 < 1 \end{aligned}$$

แสดงว่าพารามิเตอร์ของแบบจำลองเป็นแบบ Stationary

[3] การทดสอบความเหมาะสมของแบบจำลอง (Tests of Goodness of Fit of Selected Model)

(3a) กำหนดหา $\hat{\varepsilon}_t$ โดยเริ่มจาก $t = 3, \dots, 50$

$$\begin{aligned} \hat{\varepsilon}_3 &= \hat{z}_3 - \hat{\phi}_1 \hat{z}_2 - \hat{\phi}_2 \hat{z}_1 \\ &= -0.0018 - 0.542(-0.2908) - (-0.114)(-0.0178) = 0.1538 \end{aligned}$$

(3b) $Q = N \sum_{k=1}^{L=12} r_k^2(\varepsilon) = 0.103$

$$Q < \chi_{0.95, (12-2)}^2 = 18.31$$

แสดงว่า $\hat{\varepsilon}_t$ เป็นอนุกรมหรือตัวแปรสุ่มแบบอิสระ (Independent)

(3c) $\gamma(\hat{\varepsilon}) = 0.298 < \gamma_{(0.10, 50)} = 0.534$ (ดูตารางที่ 3.4)

แสดงว่า $\hat{\varepsilon}_t$ มีการแจกแจงแบบปกติ (Normal)

(3d) ใช้ AIC สำหรับหาลำดับของแบบจำลอง AR

$$AIC(p, q) = N \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q)$$

$$\hat{\sigma}_\varepsilon^2 = 0.0301 \text{ โดยคำนวณจากค่า } \hat{\varepsilon}_t (t = 3, \dots, 50) \text{ ใน (3a)}$$

$$AIC(1) = 50 \ln(0.0301) + 2(1 + 0) = -173.15 \text{ (ค่าต่ำสุด)}$$

$$AIC(2) = 50 \ln(0.0301) + 2(2 + 0) = -172.36$$

$$AIC(3) = 50 \ln(0.0301) + 2(3 + 0) = -169.89$$

$$AIC(2) > AIC(1) \text{ แสดงว่า AR(1) เหมาะสมกว่า AR(2)}$$

เลือก AR(1) สำหรับการวิเคราะห์ โดยเริ่มขั้นที่ (2d') ใหม่

ทำซ้ำขั้นที่ (2d') ใหม่ เพื่อประมาณพารามิเตอร์ของแบบจำลอง AR(1) ใหม่ (2nd Iteration)

(2d') Maximum Likelihood Estimate

$$p = 1; \hat{\phi}_1 = \frac{D_{12}}{D_{22}} = 0.483; \text{ (Moment Estimate : } \hat{\phi}_1 = \hat{\rho}_1 = r_1 = 0.397)$$

$$(2e') \hat{\sigma}_\varepsilon^2 = \frac{1}{N-1} (D_{11} - \hat{\phi}_1 D_{12}) = \frac{1}{50-1} (2.079 - 0.483 \times 1.037) = 0.0322$$

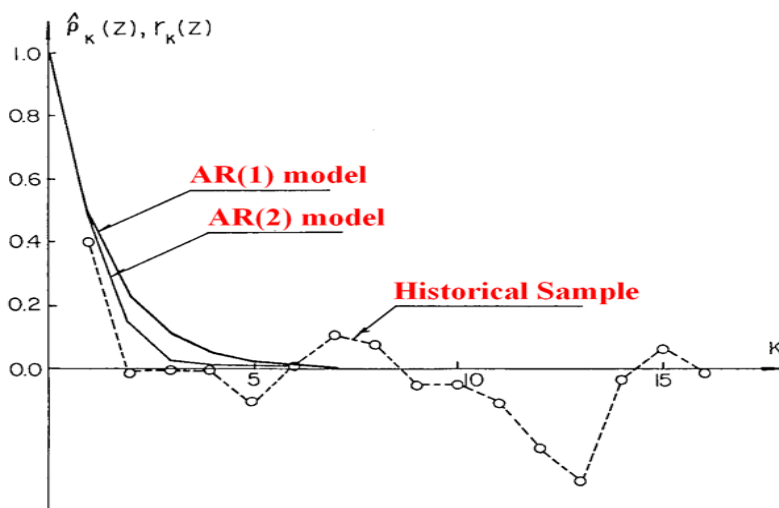
(2f') $-1 < \hat{\phi}_1 = 0.483 < 1$;แสดงว่าเป็นแบบ Stationary

การทดสอบความเหมาะสมของแบบจำลอง (2nd Iteration)

- (3a) $\hat{\varepsilon}_t = \hat{z}_t - \hat{\phi}_1 \hat{z}_{t-1}$
 $t = p + 1 = 2; \hat{\varepsilon}_2 = \hat{z}_2 - \hat{\phi}_1 \hat{z}_1$
 $= -0.2908 - 0.483(-0.0178) = -0.282$
 $\hat{\varepsilon}_3 = -0.0018 - 0.483(-0.2908) = 0.1387$
- (3b) $Q = N \sum_{k=1}^{L=12} r_k^2(\varepsilon) = 5.243; (L = 12)$
 $Q < \chi_{0.05, 12-1}^2 (= 19.68)$; แสดงว่า $\hat{\varepsilon}_t$ เป็นอนุกรมแบบอิสระ (Independent)
- (3c) $\hat{\nu}(\hat{\varepsilon}) = 0.162 < \gamma_{0.10, 50} = 0.539$; แสดงว่า $\hat{\varepsilon}_t$ มีการแจกแจงแบบปกติ (Normal)
- (3d) $AIC(0) = -166.63$
 $AIC(1) = -173.15 \dots \text{Minimum}$
 $AIC(2) = -172.36$
เลือกแบบจำลอง AR(1)
- (3e) $\rho_k = \hat{\phi}_1^k$ เมื่อ $k > 0$

รูปที่ 4.7 แสดง Sample Correlogram $r_k(z)$ และ Model Fitted Correlogram $\hat{\rho}_k(z)$ ของแบบจำลอง AR(1) และ AR(2) Model Correlogram ของ AR(2) มีลักษณะใกล้เคียงกับ Sample Correlogram มากกว่า Model Correlogram ของ AR(1)

ดังนั้น ตามผลการวิเคราะห์นี้ ควรเลือกแบบจำลอง AR(2) เพื่อการวิเคราะห์ เนื่องจาก $AIC(2) > AIC(1)$ เพียงเล็กน้อยเท่านั้น แต่อย่างไรก็ตามจะใช้แบบจำลอง AR(1) ในการวิเคราะห์ต่อไป



รูปที่ 4.6 การเปรียบเทียบ Sample Correlogram $r_k(z)$ ของอนุกรม z_t และ Model Correlogram $\hat{\rho}_k(z)$ ของแบบจำลอง AR(1) และ AR(2) สำหรับอนุกรมน้ำท่ารายปี ของแม่น้ำ Gota ประเทศสวีเดน

[4] Optional Tests of Model

ตัวอย่างการสังเคราะห์ข้อมูลใหม่

$$\hat{y} = 0.9528$$

$$\begin{aligned}\hat{\sigma}_\varepsilon &= \sqrt{0.0322} = 0.179 \\ \hat{\phi}_1 &= 0.483 \\ \hat{y}_t &= \hat{z}_t + 0.9528 \\ \hat{z}_t &= 0.483\hat{z}_{t-1} + 0.179\xi_t\end{aligned}$$

สุ่มเลือก ξ_t จากตาราง Standardized Normal Random Variables แล้วจึงคำนวณหา \hat{z}_t และ \hat{y}_t ดังตารางที่ 4.6

ตารางที่ 4.6 ตัวอย่างการสังเคราะห์ข้อมูล

t	ξ_t	\hat{z}_t	\hat{y}_t
1	0.414	0.074	1.027
2	-1.288	-0.195	0.758
3	1.019	0.088	1.041
4	0.617	0.153	0.106
5	-0.290	0.022	0.975
6	1.969	0.363	1.316
7	-0.661	0.057	1.010
8	0.595	0.134	1.087
9	-0.652	-0.052	0.901
10	0.906	0.137	1.090

คำนวณหาค่าสถิติจากข้อมูลที่สังเคราะห์ขึ้นใหม่ แล้วนำไปเปรียบเทียบกับค่าสถิติจากตัวอย่างเช่น $\hat{\mu}$, $\hat{\sigma}$ และ r_k

[5] ความน่าเชื่อถือของพารามิเตอร์ของแบบจำลอง (Reliability of Model Parameters)

(5a) จากสมการ [4.40]

$$\begin{aligned}Var(\bar{y}) &= \frac{\hat{\sigma}^2}{N^2(1-\hat{\phi}_1)^2} [(1-\hat{\phi}_1^2)N - 2\hat{\phi}_1(1-\hat{\phi}_1^N)] \\ \bar{y} &= 0.9528 \\ \hat{\sigma}^2 &= 0.0357 \\ \hat{\phi}_1 &= 0.483 \\ Var(\bar{y}) &= \frac{0.0357}{50^2(1-0.483)^2} [(1-0.483^2)50 - 2 \times 0.483^2(1-0.483^{50})] \\ &= 0.001996 \\ s(\bar{y}) &= 0.0477\end{aligned}$$

(5b) $t_{0.025,49} = 2.010$

$$\begin{aligned}t_{\left(\frac{\alpha}{2}, N-1\right)}s(\bar{y}) &= 2.010 \times 0.0447 = 0.0898 \\ 0.9528 - 0.0898 &< \mu < 0.9528 + 0.0898 \\ 0.8629 &< \mu < 1.0426\end{aligned}$$

(5c) จากสมการ 4.43

$$\frac{(N-1)S^2}{\chi_{\frac{\alpha}{2}, N-1}^2} < \sigma^2 < \frac{(N-1)S^2}{\chi_{1-\frac{\alpha}{2}, N-1}^2}$$

$$\chi_{0.025, 49}^2 = 70.2224 \text{ และ } \chi_{0.975, 49}^2 = 31.5549$$

$$0.0249 < \sigma^2 < 0.0554$$

(5d) กรณี AR(1)

(5d') Maximum Likelihood

$$\text{Var}(\hat{\phi}_1) = \frac{\hat{\sigma}_\varepsilon^2}{D_{22}} = \frac{0.0322}{2.147} = 0.015$$

$$s(\hat{\phi}_1) = \sqrt{0.015} = 0.1225$$

(5e') 95% Confidence Interval of ϕ_1

$$\{0.483 - 1.96(0.1225) < \phi_1 < 0.483 + 1.96(0.1225)\}$$

$$\{0.243 < \phi_1 < 0.723\}$$

(5f') 95% CI of σ_ε^2

$$\sigma_\varepsilon^2(\text{Lower}) = \frac{1}{N} (D_{11} - \hat{\phi}_1(\text{Lower})D_{12})$$

$$= \frac{1}{49} (2.079 - 0.723 \times 1.037) = 0.0271$$

$$\sigma_\varepsilon^2(\text{Upper}) = \frac{1}{49} (2.079 - 0.243 \times 1.037) = 0.0373$$

4.3 Periodic AR Model

ให้

$$y_{v,\tau} = g_\tau(x_{v,\tau})$$

$g_\tau =$ Normal Transform Function

$$y_{v,\tau} = \mu_\tau + \sigma_\tau z_{v,\tau}$$

$z_{v,\tau} =$ AR with Constant or Periodic Coefficient

4.3.1 AR with Constant Coefficient

แบบจำลอง Periodic AR with Constant Coefficient จะเริ่มจากการแปลง $z_{v,\tau}$ ให้เป็น z_t เมื่อ $t = (v-1)\omega + \tau$ และหลังแปลง $z_{v,\tau}$ ให้เป็น z_t แล้วจะทำการจำลอง z_t เหมือนการจำลองอนุกรมเวลารายปี

แบบจำลอง AR (p)

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t \quad [4.80]$$

ซึ่งมีพารามิเตอร์คือ $\{g_\tau(\cdot), \mu_\tau, \sigma_\tau, \phi_1, \dots, \phi_p, \sigma_\varepsilon; \tau = 1, \dots, \omega\}$

4.3.2 AR with Periodic Coefficient

$$AR(0): z_{v,\tau} = \varepsilon_{v,\tau} \quad [4.81]$$

$$AR(1): z_{v,\tau} = \phi_{1,\tau} z_{v,\tau-1} + \sigma_{\varepsilon,\tau} \xi_{v,\tau} \quad [4.82]$$

$$AR(p): z_{v,\tau} = \phi_{1,\tau} z_{v,\tau-1} + \dots + \phi_{p,\tau} z_{v,\tau-p} + \sigma_{\varepsilon,\tau} \xi_{v,\tau} \quad [4.83]$$

ซึ่งมีพารามิเตอร์คือ $\{g_{\tau}(\cdot), \mu_{\tau}, \sigma_{\tau}, \phi_{1,\tau}, \dots, \phi_{p,\tau}, \sigma_{\varepsilon,\tau}; \tau = 1, \dots, \omega\}$

$\phi_{j,\tau}$ = Periodic Autoregressive Coefficient; $j = 1, \dots, p$

$\sigma_{\varepsilon,\tau}^2$ = Periodic Variance of Residuals

4.3.3 การประมาณค่าพารามิเตอร์สำหรับ Periodic AR Models

เริ่มจากการประมาณ Periodic Mean และ Periodic Standard Deviation ดังนี้

ขั้นที่ 1 สำหรับ $\omega = 2, 3, 4, \dots, 12$

$$\hat{\mu}_{\tau} = \bar{y}_{\tau} = \frac{1}{N} \sum_{v=1}^N y_{v,\tau}; \tau = 1, \dots, \omega \quad [4.84]$$

$$\hat{\sigma}_{\tau} = S_{\tau} = \left[\frac{1}{(N-1)} \sum_{v=1}^N (y_{v,\tau} - \bar{y}_{\tau})^2 \right]^{\frac{1}{2}}; \tau = 1, \dots, \omega \quad [4.85]$$

ถ้า $\omega \geq 12$ ควรประมาณค่าโดยใช้ Fourier Series ในขั้นที่ 2

ขั้นที่ 2 การประมาณค่าโดยใช้ Fourier Series

$$\hat{\mu}_{\tau} = \bar{y}_{\tau} + \sum_{j=1}^{h^*(\bar{y})} [Ah_j^*(\bar{y}) \cdot \cos(2\pi h_j(\bar{y})\tau/\omega) + Bh_j^*(\bar{y}) \cdot \sin(2\pi h_j(\bar{y})\tau/\omega)] \quad [4.86]$$

$; \tau = 1, \dots, \omega$

$$\bar{y} = \sum_{\tau=1}^{\omega} \frac{\bar{y}_{\tau}}{\omega}$$

$Ah_j^*(\bar{y}), Bh_j^*(\bar{y}) = h_j^{\text{th}}$ Harmonic Fourier Coefficient for Mean

$h_j^*(\bar{y}) = j^{\text{th}}$ Significant Harmonic for Mean

$h^*(\bar{y}) =$ No. of Significant Harmonic from Total Harmonics of $\frac{\omega}{2}$ or $\frac{\omega-1}{2}$

$$\hat{\sigma}_{\tau} = \bar{S} + \sum_{j=1}^{h^*(s)} [Ah_j^*(s) \cdot \cos(2\pi h_j(s)\tau/\omega) + Bh_j^*(s) \cdot \sin(2\pi h_j(s)\tau/\omega)] \quad [4.87]$$

$; \tau = 1, \dots, \omega$

$$\bar{S} = \sum_{\tau=1}^{\omega} \frac{S_{\tau}}{\omega}$$

$Ah_j^*(s), Bh_j^*(s) = h_j^{\text{th}}$ Harmonic Fourier Coefficients for Standard Deviation

$h_j(s) = j^{\text{th}}$ Significant Harmonic for Standard Deviation

$h^*(s) =$ No. of Significant Harmonic

การประมาณค่า $\phi_{1,\tau}, \dots, \phi_{p,\tau}$ และ $\sigma_{\varepsilon,\tau}^2$

$$\text{จาก } z_{v,\tau} = \frac{y_{v,\tau} - \hat{\mu}_{\tau}}{s_{\tau}}; v = 1, \dots, N; \tau = 1, \dots, \omega \quad [4.88]$$

คำนวณหา $r_{k,\tau}$ หรือ $\hat{\rho}_{k,\tau}$ (Sample Periodic Correlation Coefficient) ของ $z_{v,\tau}$

AR(1)

$$\hat{\phi}_{1,\tau} = \hat{\rho}_{1,\tau}; \tau = 1, \dots, \omega \quad [4.89]$$

AR(2)

$$\begin{aligned} \hat{\phi}_{1,\tau} &= \frac{\hat{\rho}_{1,\tau} - \hat{\rho}_{1,\tau-1} \hat{\rho}_{2,\tau}}{1 - \hat{\rho}_{1,\tau-1}}; \tau = 1, \dots, \omega \\ \hat{\phi}_{2,\tau} &= \frac{\hat{\rho}_{2,\tau} - \hat{\rho}_{1,\tau-1} \hat{\rho}_{1,\tau-1}}{1 - \hat{\rho}_{1,\tau-1}^2}; \tau = 1, \dots, \omega \end{aligned} \quad [4.90]$$

คำนวณหา $\hat{\rho}_{k,\tau}$ โดยวิธี Fourier Series ($k = 1, \dots, p$)

$$\begin{aligned} \hat{\rho}_{k,\tau} &= \hat{\rho}_k + \sum_{j=1}^{h_k^*(r)} [A^* h_{jk}(r) \cdot \text{Cos}(2\pi h_{jk}(r)\tau/\omega) + B h_{jk}^*(r) \cdot \text{Sin}\left(\frac{2\pi h_{jk}(r)\tau}{\omega}\right)] \\ & \quad ; k = 1, \dots, p; \tau = 1, \dots, \omega \end{aligned} \quad [4.91]$$

$$\bar{\rho}_k = \sum_{\tau=1}^{\omega} \frac{r_{k,\tau}}{\omega}$$

$A h_{jk}(r), B h_{jk}(r) = h_{jk}^{\text{th}}$ Harmonic Fourier Coefficients for $r_{k,\tau}$

$h_{jk}(r) = j^{\text{th}}$ Significant Harmonic

$h_k^*(r) = \text{No. of Significant Harmonics}$

AR(p)

$$\sigma_{\varepsilon,\tau}^2 = 1 - \sum_{j=1}^p \hat{\phi}_{j,\tau} \cdot \hat{\rho}_{j,\tau}; \tau = 1, \dots, \omega \quad [4.92]$$

AR(1)

$$\sigma_{\varepsilon,\tau}^2 = 1 - \hat{\phi}_{1,\tau}^2; \tau = 1, \dots, \omega \quad [4.93]$$

4.3.4 ความน่าเชื่อถือของพารามิเตอร์แบบ Periodic (Reliability of Estimated Periodic

Parameters)

การประมาณ (Approximation) ช่วงความน่าเชื่อถือ (Confident Intervals) ของค่าพารามิเตอร์

$$\text{Var}(\hat{\mu}_\tau) = \frac{\hat{\sigma}_\tau^2}{N}; \tau = 1, \dots, \omega \quad [4.94]$$

$(1 - \alpha)\%$ Confidence Intervals ของ μ_τ

$$\hat{\mu}_\tau - t_{(N-1, \frac{\alpha}{2})} s(\hat{\mu}_\tau) < \mu_\tau < \hat{\mu}_\tau + t_{(N-1, \frac{\alpha}{2})} s(\hat{\mu}_\tau); \tau = 1, \dots, \omega \quad [4.95]$$

$(1 - \alpha)\%$ Confidence Intervals ของ σ_τ^2

$$\frac{(N-1)\hat{\sigma}_\tau^2}{\chi_{\frac{\alpha}{2}, N-1}^2} < \sigma_\tau^2 < \frac{(N-1)\hat{\sigma}_\tau^2}{\chi_{1-\frac{\alpha}{2}, N-1}^2}; \tau = 1, \dots, \omega \quad [4.96]$$

$(1 - \alpha)\%$ Confidence Intervals ของ $\rho_{k,\tau}$ สามารถหาโดยการแก้สมการ [4.24]

$$\rho_{k,\tau} = \sum_{j=1}^p \phi_{j,\tau} \rho_{|k-j|, \tau-lj}; k > 0$$

$$lj = \text{Min}(k, j)$$

โดยการแทนค่า $\rho_{k,\tau}$ ในสมการข้างบนด้วย ρ_β กำหนดให้ $\beta = \alpha/2$ สำหรับ $\rho_{k,\tau(\text{upper})}$ และ $\beta=1-\alpha/2$ สำหรับ $\rho_{k,\tau(\text{lower})}$ เมื่อ $j = 1, \dots, p$ และ $\tau = 1, \dots, \omega$

$$\rho_\beta = \frac{(1+\hat{\rho}_{k,\tau}) - (1-\hat{\rho}_{k,\tau}) \text{Exp}\left(\frac{2z_\beta}{\sqrt{N-3}}\right)}{(1+\hat{\rho}_{k,\tau}) + (1-\hat{\rho}_{k,\tau}) \text{Exp}\left(\frac{2z_\beta}{\sqrt{N-3}}\right)} \quad [4.97]$$

เมื่อ z_β = Standard Normal Variate or Z_β

$\beta = \frac{\alpha}{2}$ สำหรับ Upper Confidence Limit ของ $\rho_{k,\tau}$

$\beta = 1 - \frac{\alpha}{2}$ สำหรับ Lower Confidence Limit ของ $\rho_{k,\tau}$

$(1 - \alpha)\%$ ช่วงความเชื่อมั่นของ $\sigma_{\varepsilon,\tau}^2$ จะหาจากสมการ

$$\sigma_{\varepsilon,\tau(\text{upper})}^2 = 1 - \sum_{j=1}^P \phi_{j,\tau(\text{upper})} \hat{\rho}_{j,\tau}; \tau = 1, \dots, \omega$$

$$\sigma_{\varepsilon,\tau(\text{lower})}^2 = 1 - \sum_{j=1}^P \phi_{j,\tau(\text{lower})} \hat{\rho}_{j,\tau}; \tau = 1, \dots, \omega$$

ค่า $\phi_{j,\tau(\text{upper})}$ และ $\phi_{j,\tau(\text{lower})}$ คือ Upper และ Lower Confident Limits ของ $\phi_{j,\tau}$ ที่คำนวณได้ตามที่กล่าวไว้ในหัวข้อที่แล้ว

4.3.5 ความเหมาะสมของแบบจำลอง Periodic AR (Goodness of Fit for Periodic AR Models)

แบบจำลอง Periodic AR มีสมมติฐานว่า Residuals (ε_t or $\xi_{v,\tau}$) เป็นตัวแปรอิสระและมีการแจกแจงแบบปกติ (Independent and Normally Distributed) จึงต้องมีการทดสอบ Residual ว่าเป็นตัวแปรอิสระและมีการแจกแจงปกติ นอกจากนี้ควรมีการทดสอบเปรียบเทียบ Model Correlogram กับ Sample Correlogram

$$\xi_{v,\tau} = \frac{1}{\sigma_{\varepsilon\tau}} [z_{v,\tau} - (\hat{\phi}_{1,\tau} z_{v,\tau-1} + \dots + \hat{\phi}_{p,\tau} z_{v,\tau-p})] \quad [4.98]$$

ค่าแรกของ $\xi_{v,\tau}$ คือ $\xi_{1,p+1}$ เนื่องจาก $\tau - p = 1$ ดังนั้น $\tau = p + 1$

(A) การทดสอบความเป็นอิสระของ Residuals (Test of Independent Residuals)

การทดสอบความเป็นอิสระมี 2 แนวทางได้แก่

แนวทางที่ 1 กรณีที่วิเคราะห์จาก Periodic Residual Series ($\xi_{v,\tau}$) มี 2 วิธีคือ

(1) การทดสอบนัยสำคัญของ ACF ซึ่งมีขั้นตอนดังนี้

- กำหนดหา Periodic Lag 1 ACF หรือ $r_{1,\tau}(\xi)$; $\tau = 1, \dots, \omega$
- หาค่า Critical r จากสมการ

$$r_{\frac{\alpha}{2}} = \frac{t_{(N-2), \frac{\alpha}{2}}}{\sqrt{(N-2) + t_{(N-2), \frac{\alpha}{2}}^2}} \quad [4.99]$$

- ถ้า $|r_{1,\tau}(\xi)| < r_{\frac{\alpha}{2}}$ แสดงว่ายอมรับสมมติฐานหลัก H_0 ที่ว่า $\xi_{v,\tau}$ เป็นตัวแปรอิสระ

(2) การทดสอบโดยใช้ค่าสถิติ Modified Porte Manteau

ค่าสถิติ Modified Porte Manteau (Q_1) มีค่าดังสมการ

$$Q_1 = N \sum_{k=1}^L \sum_{\tau=1}^{\omega} [r_{k,\tau}(\xi)]^2 \quad [4.100]$$

ถ้า $Q_1 < \chi_{(L-p-q)}^2$ แสดงว่า $\xi_{v,\tau}$ เป็นตัวแปรอิสระ

แนวทางที่ 2 กรณีที่วิเคราะห์จาก Residual Series (ξ_t) มี 2 วิธีคือ

(1) Anderson Test ซึ่งมีขั้นตอนดังนี้

- คำนวณค่า ξ_t ; $t = 1, \dots, (N\omega - p)$
- คำนวณหา $r_k(\xi)$; $k=1, \dots, L$; $L=10-20\%$ ของ N
- หา Anderson Limit จากสมการ

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k} \quad [4.101]$$

$$r_k(99\%) = \frac{-1 \pm 2.576\sqrt{N-k-1}}{N-k} \quad [4.102]$$

(2) การทดสอบโดยใช้ค่าสถิติ Port Manteau

ค่าสถิติ Port Manteau (Q_1) มีค่าดังสมการ

$$Q = (N - d) \sum_{k=1}^L r_k^2(\xi) \quad [4.103]$$

ถ้า $Q < \chi_{(L-p-q)}^2$ แสดงว่า $\xi_{v,\tau}$ เป็นตัวแปรอิสระ

ถ้าผลการทดสอบพบว่า ξ ไม่เป็นอิสระ ให้เพิ่มลำดับ (Order) ของแบบจำลองจาก p เป็น $(p+1)$

(B) การทดสอบการแจกแจงแบบปกติ (Test of Normality)

การทดสอบว่า Residual มีการแจกแจงแบบปกติ สามารถทำได้ 2 วิธีคือ

(1) การทดสอบโดยวิธีกราฟ โดยการคำนวณหา Empirical Frequency ของ Residual แล้วนำไปพล็อตลงในกราฟการแจกแจงแบบปกติ (Normal Probability Paper) ถ้ากราฟเป็นเส้นตรงแสดงว่า Residual มีการแจกแจงแบบปกติ

(2) การทดสอบค่าสัมประสิทธิ์ความเบ้ (Skewness Coefficient, $\gamma(\xi)$) โดยใช้ค่าสถิติ Smirnov-Kolmogorov

(C) การทดสอบความเหมาะสมโดนการเปรียบเทียบ Correlogram ของ Historical, Model และ

Generated Data

(1) การเปรียบเทียบ Historical และ Model Correlograms

$$r_{k,\tau}(\xi) \text{ vs. } \rho_{k,\tau}$$

(2) การเปรียบเทียบ Historical และ Generated Correlograms

$$r_{k,\tau}(\xi) \text{ vs. } \hat{\rho}_{k,\tau}$$

4.3.6 การสังเคราะห์ข้อมูลโดยใช้แบบจำลอง Periodic AR

(Generation Using Periodic AR Models)

ให้ $x_{v,\tau}$ คือ อนุกรมเวลาเริ่มต้นซึ่งไม่ได้มีการแจกแจงแบบปกติ (Original Skewed Series)

ขั้นตอนการสังเคราะห์ข้อมูล

$$\hat{x}_{v,\tau} = \hat{g}_\tau^{-1}(\hat{y}_{v,\tau}) \quad [4.104]$$

$$\hat{y}_{v,\tau} = \hat{\mu}_\tau + \hat{\sigma}_\tau \hat{z}_{v,\tau} \quad [4.105]$$

(1) กรณีใช้แบบจำลอง AR with Constant Autoregressive Coefficient

$$\hat{z}_t = \hat{\phi} \hat{z}_{t-1} + \dots + \hat{\phi}_p \hat{z}_{t-p} + \hat{\sigma}_\varepsilon \xi_t \quad [4.106]$$

$$t = (v - 1)\omega + \tau$$

$$\xi_t = \text{Normal Variate with Mean} = 0 \text{ and Variance} = 1$$

(2) กรณีใช้แบบจำลอง AR with Periodic Autoregressive Coefficient

$$\hat{z}_{v,\tau} = \hat{\phi} \hat{z}_{v,\tau-1} + \dots + \hat{\phi}_{p,\tau} \hat{z}_{v,\tau-p} + \hat{\sigma}_{\varepsilon,\tau} \xi_{v,\tau} \quad [4.107]$$

$$\xi_{v,\tau} = \text{Normal Variate with Mean} = 0 \text{ and Variance} = 1$$

4.3.7 ตัวอย่างการจำลองอนุกรมเวลาแบบ Periodic ด้วยแบบจำลอง AR (Example of AR Modeling of Periodic Series)

การสร้างแบบจำลองสโตแคสติกสำหรับ Monthly Net Basin Supplies ของ Lakes Michigan – Huron ซึ่งมีรายละเอียดดังนี้

- Net Basin Supply (NBS) = Precipitation – Evaporation
- มีข้อมูลอนุกรมเวลาฝนรายเดือนเป็นเวลา 69 ปี (ดูภาคผนวก A7.5)
- Drainage Area = 97,400 ตารางไมล์
- พื้นที่ผิวน้ำ = 75,300 ตารางไมล์

(1) Preliminary Analysis and Model Identification

(1a) และ (1b) $x_{v,\tau} = \text{NBS}$ ซึ่งมีค่า – ถึง +

$$g = 0.6 \text{ (จากตารางที่ 3.4 } \gamma_{0.02, \text{large } N} = 0.430, \gamma_{0.1, \text{large } N} = 0.298)$$

เนื่องจาก $x_{v,\tau}$ บางค่ามีค่าติดลบ จึงไม่สามารถแปลงเป็นอนุกรมเวลาที่มีการแจกแจงแบบปกติ โดยใช้ฟังก์ชัน Logarithmic หรือ Power Function ได้ ดังนั้นในกรณีนี้ จึงทำการ Original Series และหาวิธีการจัดการค่าความเบ้โดยการวิเคราะห์หาการแจกแจงความน่าจะเป็นของ Residuals ที่หลัง

(1c) พล็อตอนุกรมเวลารายเดือน ($y_{v,\tau}$) ดังแสดงในรูปที่ 4.7 ซึ่งจะเห็นได้ว่าค่า NBS ช่วง Spring และ NBS บางค่าช่วง Summer มีค่าสูงกว่าค่า NBS ในฤดูอื่น แสดงถึง Periodicity ในอนุกรมเวลาที่เกิดขึ้นทุกปี ช่วงที่ NBS มีค่าน้อย ค่าส่วนเบี่ยงเบนมาตรฐานจะมีค่าสูงกว่าค่าเฉลี่ย

(1d) คำนวณหา \bar{y}_τ , s_τ และ $r_{k,\tau}$ เมื่อ $k = 1, 2, 3$ และ $\tau = 1, \dots, 12$

(1e) พล็อต \bar{y}_τ และ s_τ ในรูปที่ 4.8

เนื่องจาก $\omega = 12$ พอดี อาจไม่ต้องประมาณค่าโดยใช้ Fourier Series ก็ได้ แต่ในตัวอย่างนี้ต้องการแสดงวิธีการประมาณค่าโดยใช้ Fourier Series

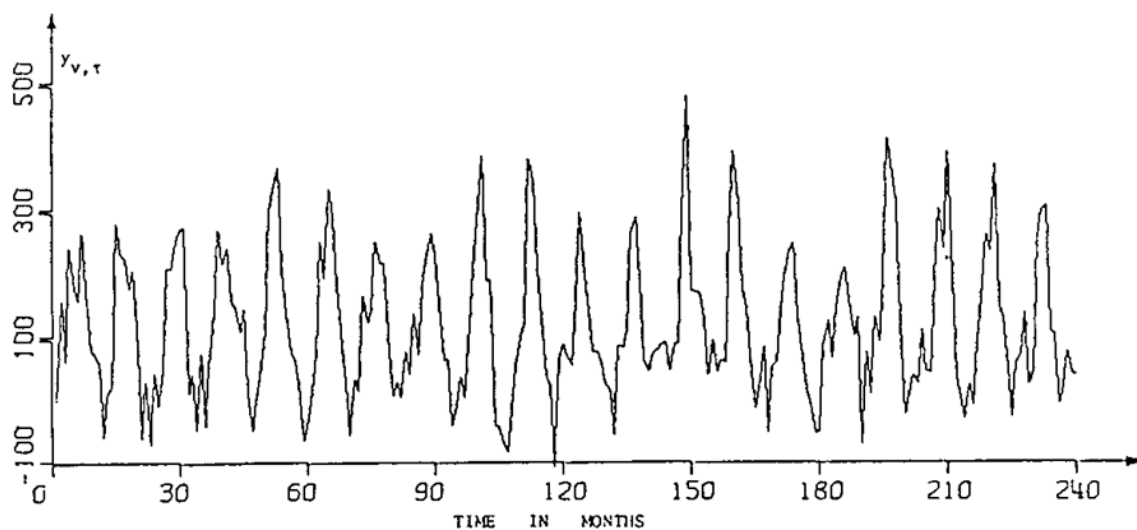
ให้ $\bar{y}_\tau = \mu_\tau$

$s_\tau = \sigma_\tau$

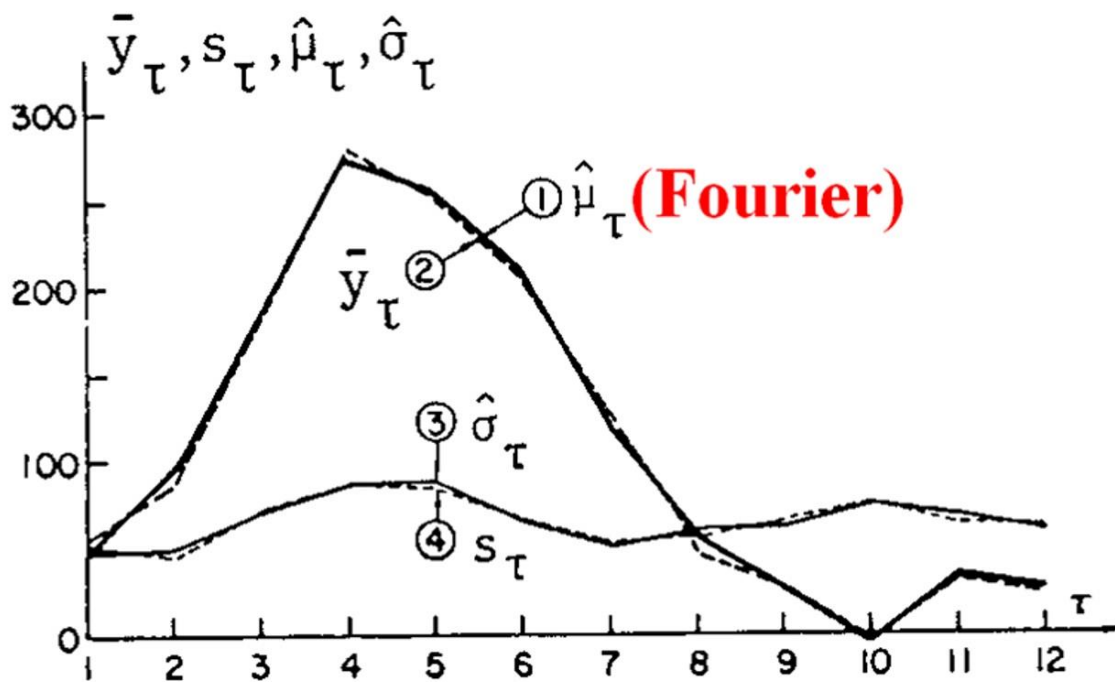
(1f) คำนวณหา $r_{k,\tau}$ สำหรับ $k = 1, 2, 3$ และ $\tau = 1, \dots, 12$ แล้วพล็อตในรูปที่ 4.9 ซึ่งจะเห็นได้ว่า $r_{1,\tau}$ ไม่เปลี่ยนแปลงอย่างมีนัยสำคัญในแต่ละเดือน

$r_{2,\tau}$ และ $r_{3,\tau}$ เปลี่ยนแปลงอย่างมีนัยสำคัญในแต่ละเดือน

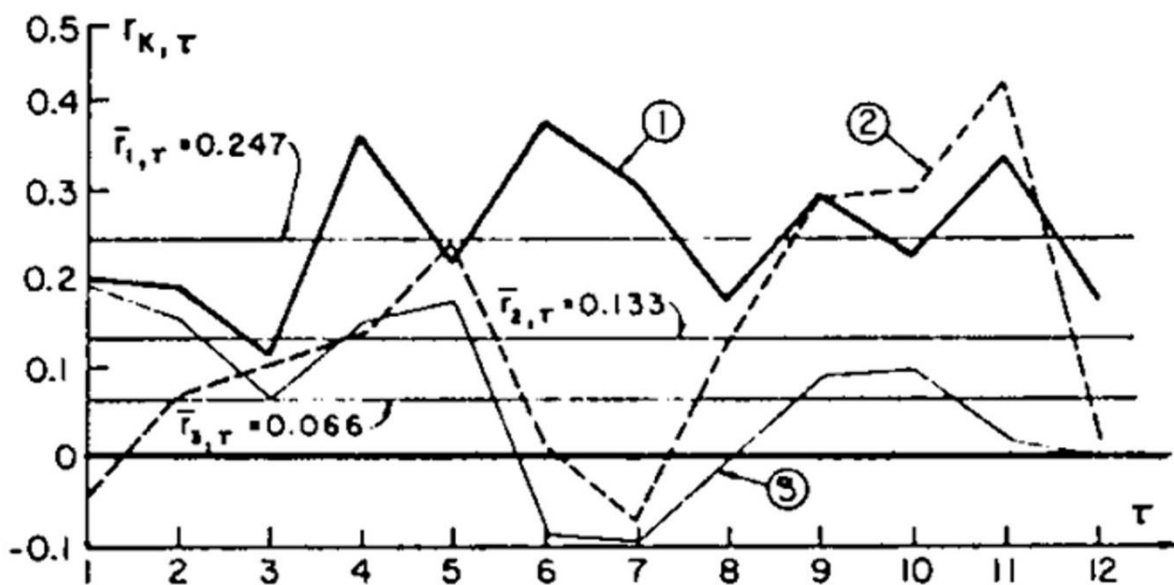
* เนื่องจากพิจารณาว่า $r_{1,\tau}$ มีความน่าเชื่อถือมากกว่า $r_{k,\tau}$ เมื่อ $k > 1$ จึงเลือกการจำลองโดยใช้แบบจำลอง AR Model with Constant Autoregressive Coefficients



รูปที่ 4.7 อนุกรมเวลา NBS รายเดือนของ Lake Michigan-Huron ช่วง 1990-1999



รูปที่ 4.8 (1) Fitted periodic mean $\hat{\mu}_\tau$ (2) periodic mean \bar{y}_τ (3) fitted periodic standard deviation $\hat{\sigma}_\tau$ (4) standard deviation s_τ for the NBS of Lake Michigan-Huron



รูปที่ 4.9 Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized $z_{v,\tau}$ series for the NBS of Lake Michigan-Huron

(2) การประมาณค่าพารามิเตอร์ (Estimation of Parameters)

(2a) การประมาณค่าพารามิเตอร์ โดย Fourier Series $\hat{\mu}_\tau, \hat{\sigma}_\tau$

(i) $\bar{y} = 109.787$

$s = 65.608$

$h = \frac{\omega}{2} = \frac{12}{2} = 6 = \text{Total No. of Harmonics}$

คำนวณหา $A_j(\cdot), B_j(\cdot)$ และ $MSD_j(\cdot)$ (หมายถึง Explained Variance) ของ $\hat{\mu}_\tau$ และ $\hat{\sigma}_\tau$ ดังแสดงในตารางที่ 4.7 และ 4.8

(ii) หาจำนวน Significant Harmonics โดยใช้ Cumulative Periodogram ดังรูปที่ 4.10 ซึ่งสามารถสรุปได้ว่า

$h^* \text{ of } \hat{\mu}_\tau = 2$

$h^* \text{ of } \hat{\sigma}_\tau = 3$

* พิจารณาร่วมกับผลการวิเคราะห์ค่า Significant Harmonics ของทะเลสาบอื่นๆใน Great Lakes จึงกำหนดให้ $h^* = 4$ และ Significant Harmonics คือ Harmonic ที่ 1, 2, 3 และ 5

(iii) ค่า $A_j^*(\cdot), B_j^*(\cdot)$ แสดงอยู่ในตารางที่ 4.7 และ 4.8

(iv) คำนวณค่า $\hat{\mu}_\tau$ และ $\hat{\sigma}_\tau$ ดังแสดงในรูปที่ 4.8 ในที่นี้จะแสดงเฉพาะตัวอย่างการคำนวณ $\hat{\mu}_1$ และ $\hat{\sigma}_1$

ตารางที่ 4.7 Fourier coefficients and Explained Variances for the Monthly Mean y_t of NBS of Lake Michigan-Huron.

Harmonic	Coefficient	Coefficient	Explained variance
J	$A_i(\bar{y})$	$B_i(\bar{y})$	
1	-93.848	84.036	90.916
2	8.191	-34.842	7.339
3	9.792	-0.353	0.550
4	-0.911	6.393	0.239
5	-8.567	-9.185	0.904
6	-2.140	0.000	0.206

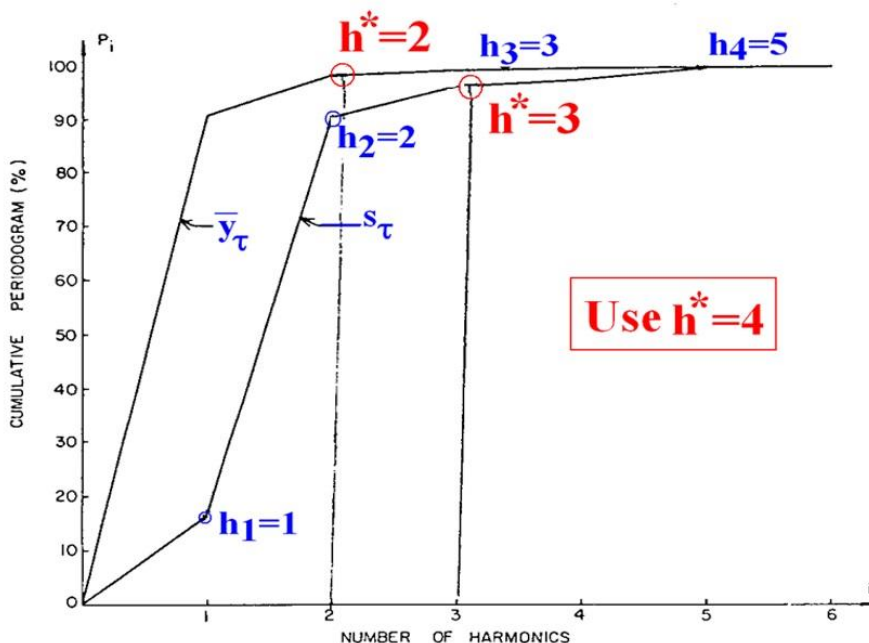
ตารางที่ 4.8 Fourier coefficients and Explained Variances for the Monthly Standard Deviation S_t of NBS of Lake Michigan-Huron

Harmonic	Coefficient	Coefficient	Explained variance
J	$A_i(\bar{y})$	$B_i(\bar{y})$	
1	-6.062	3.574	16.535
2	-2.364	-14.685	73.870
3	3.638	1.990	5.742
4	1.297	1.580	1.396
5	0.656	2.590	2.383
6	-0.333	0.000	0.037

$$\begin{aligned}\hat{\mu}_1 = & 109.787 + \left[-93.848 \cos\left(2\pi \times 1 \times \frac{1}{12}\right) + 84.036 \sin\left(2\pi \times 1 \times \frac{1}{12}\right) \right] \\ & + \left[8.191 \cos\left(2\pi \times 2 \times \frac{1}{12}\right) - 34.842 \sin\left(2\pi \times 2 \times \frac{1}{12}\right) \right] \\ & + \left[9.792 \cos\left(2\pi \times 3 \times \frac{1}{12}\right) - 0.353 \sin\left(2\pi \times 3 \times \frac{1}{12}\right) \right] \left[-8.567 \cos(2\pi \times 5 \times 1/12) \right] \\ & - 9.185 \sin(2\pi \times 5 \times 1/12) = 46.93\end{aligned}$$

Similarly for $\tau = 2$ Eq. (4.91) gives:

$$\begin{aligned}\hat{\mu}_2 = & 109.787 + \left[-93.848 \cos\left(2\pi \times 1 \times \frac{2}{12}\right) + 84.036 \sin\left(2\pi \times 1 \times \frac{2}{12}\right) \right] \\ & + \left[8.191 \cos(2\pi \times 2 \times 2/12) - 34.842 \sin(2\pi \times 2 \times 2/12) \right] \\ & + \left[9.792 \cos(2\pi \times 3 \times 2/12) - 0.353 \sin(2\pi \times 3 \times 2/12) \right] \\ & + \left[-8.567 \cos(2\pi \times 5 \times 2/12) \right] - 9.185 \sin(2\pi \times 5 \times 2/12) = 95.25\end{aligned}$$



รูปที่ 4.10 Cumulative periodogram for the mean \bar{y}_τ and standard deviation s_τ for the monthly NBS of Lake Michigan-Huron

(2b) การคำนวณหาค่าอนุกรมเวลา $z_{v,\tau}$

$$z_{1,1} = \frac{y_{1,1} - \hat{\mu}_1}{\hat{\sigma}_1} = \frac{-1.0 - 46.93}{50.96} = -0.94$$

$$z_{1,2} = \frac{y_{1,2} - \hat{\mu}_2}{\hat{\sigma}_2} = \frac{162 - 95.25}{48.58} = 1.374$$

$$z_{1,3} = -1.619$$

$$z_{1,4} = -0.364$$

.

$$z_{69,12} = 1.686$$

จำลองโดยแบบจำลอง AR with Constant Autoregressive Coefficients ตามที่สรุปในข้อ (1)

(3c)

(i) แปลง $z_{v,\tau}$ เป็น $z_t (t = 1, \dots, (v-1) * 12 + \tau, \dots, N\omega)$

$$N\omega = 69 \times 12 = 828$$

$$z_1 = -0.94$$

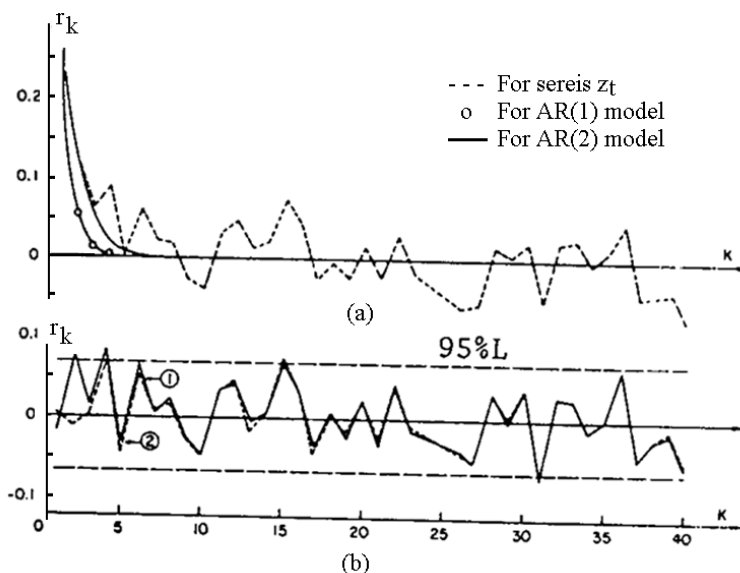
$$z_2 = 1.374$$

$$z_{828} = 1.686$$

(ii) คำนวณหา $r_k(z)$ เมื่อ $k = 1, \dots, 40$ ดังแสดงในรูปที่ 4.11(a) ซึ่งสรุปได้ว่า

$$r_1(z) = 0.229$$

$$r_2(z) = 0.135$$



รูปที่ 4.11 (a) Correlogram of the standardized series z_t and expected correlograms for the first and second order Markov models, (b) correlograms of the series ε_t after fitting the first (1) and second (2) order Markov models (Lake Michigan-Huron)

(iii) พล็อต Model Correlogram ของ AR(1) ซึ่งมี $\hat{\phi}_1 = 0.229$ และ AR(2) ซึ่งมี $\phi = 0.209$ และ $\hat{\phi}_2 = 0.087$ เปรียบเทียบกับ Sample Correlogram $r_k(z)$ ดังแสดงในรูปที่ 4.11(a) ซึ่งเห็นได้ว่าแบบจำลอง AR(2) ให้ค่าประมาณของ $r_k(z)$ ดีกว่า AR(1) ดังนั้นในขั้นนี้จึงสรุปได้ว่า AR(2) ดีกว่า AR(1) แต่ในการวิเคราะห์ควรต้องวิเคราะห์แบบจำลอง AR(3) และ ARMA(1,1) เปรียบเทียบกับ AR(1) และ AR(2) ด้วย

(iv) $N' = N\omega = 828$ (อนุกรมเวลารายเดือนมีข้อมูลยาว)

ใช้วิธี Moment เพื่อประมาณพารามิเตอร์ของ AR(2)

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{(1-r_1^2)} = \frac{0.229(1-0.135)}{(1-0.229^2)} = 0.209$$

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2} = \frac{0.135 - 0.229^2}{1 - 0.229^2} = 0.087$$

จากสมการที่ 4.38

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= \frac{N\hat{\sigma}^2(1+\hat{\phi}_2)}{(N-2)(1-\hat{\phi}_2)} \left[(1-\hat{\phi}_2)^2 - \hat{\phi}_1^2 \right] \\ &= \frac{828(1)(1+0.087)}{(828-2)(1-0.087)} [(1-0.087)^2 - 0.209^2] \\ &= 0.943\end{aligned}$$

(v) ตรวจสอบ Stationary Condition ของพารามิเตอร์โดยสมการ 4.18

$$\begin{aligned}\phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 < \phi_2 &< 1 \\ \hat{\phi}_1 + \hat{\phi}_2 &= 0.209 + 0.087 = 0.296 < 1 \\ \hat{\phi}_2 - \hat{\phi}_1 &= 0.087 - 0.209 = -0.122 < 1 \\ -1 < \phi_2 &= 0.087 < 1\end{aligned}$$

(3) การทดสอบความเหมาะสมของแบบจำลอง AR(2)(Test of Goodness of Fit of Selected Model)

(3a) จากสมการที่ 4.19 จะสามารถคำนวณหาอนุกรมเวลา $\hat{\varepsilon}_t$ ของ AR(2) ได้ดังนี้

$$\begin{aligned}\hat{\varepsilon}_3 &= z_3 - \hat{\phi}_1 z_2 - \hat{\phi}_2 z_1 \\ &= -1.619 - 0.209 \times 1.374 - 0.087 \times (-0.94) \\ &= -1.824 \\ \hat{\varepsilon}_4 &= z_4 - \hat{\phi}_1 z_3 - \hat{\phi}_2 z_2 \\ &= -0.364 - 0.209(-1.619) - 0.087(1.374) \\ &= -0.623 \\ \hat{\varepsilon}_{828} &= -1.623\end{aligned}$$

(3b) การทดสอบความเป็นอิสระของ Residuals

โดยใช้ Anderson Test of Correlogram

คำนวณหา $r_k(\hat{\varepsilon})$ โดยสมการ 2.5(b)

Then calculate 95% Probability Limits of $r_k(\hat{\varepsilon})$ by Eq.2.21a

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-K-1}}{N-K}$$

Correlogram ของ $\hat{\varepsilon}_t$ ของ AR(1) และ AR(2) และ 95% Probability Limits แสดงอยู่ในรูปที่

4.11(b)

$r_k(\hat{\varepsilon})$ ของ AR(2) อยู่ภายใน Probability Limits

$r_k(\hat{\varepsilon})$ ของ AR(1) มี 4 ค่าตกอยู่นอก Probability เกินกว่าค่าสูงสุดตามเกณฑ์ที่กำหนดคือ

$$(1 - 0.95)L = 0.05 \times 40 = 2 \text{ เมื่อ } L = \text{Max lag} = 40$$

จึงสรุปได้ว่าอนุกรมเวลา $\hat{\varepsilon}_t$ ของ AR(2) เป็นอิสระ

(3c) การทดสอบ Normality ของ $\hat{\varepsilon}_t$

ถ้า $\hat{\gamma}(\varepsilon) > 0$ แสดงว่า ε_t ไม่ได้มีการแจกแจงแบบปกติ
 เนื่องจาก $\hat{\gamma}(\varepsilon) > 0$ แต่ไม่ได้มีการแปลงเป็นการแจกแจงปกติ เนื่องจาก $x = NBS$ บางค่ามีค่า

ติดลบ

$$\hat{\sigma}_\varepsilon = \sqrt{0.943} = 0.971$$

$$\hat{\xi}_t = \frac{\varepsilon_t}{0.971} \quad (\text{Standardized})$$

ผลการวิเคราะห์การแจกแจงพบว่า PDF ของ $\hat{\xi}_t$ มีการแจกแจงแบบ Lognormal 3 พารามิเตอร์ โดยมีค่าพารามิเตอร์ดังนี้

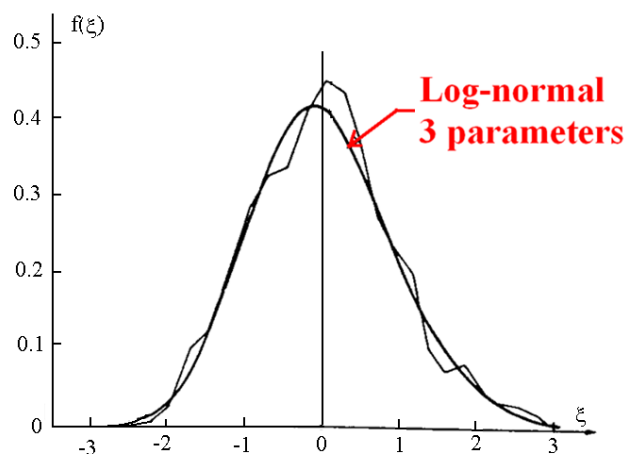
$$(\hat{\alpha} = 1.356; \hat{\beta} = 0.246; \hat{\xi}_0 = -4)$$

$$f(\xi) = \frac{1}{(\xi - \hat{\xi}_0)\beta\sqrt{2\pi}} \text{Exp} \left\{ -\frac{[\ln(\xi - \hat{\xi}_0) - \hat{\alpha}]^2}{2\hat{\beta}^2} \right\}$$

$$f(\xi) = \frac{1}{\sqrt{2\pi}(0.246)(\xi + 4.0)} \text{Exp} \left\{ -\frac{1}{2} \frac{[\ln(\xi + 4) - 1.356]^2}{0.246} \right\} \quad [4.108]$$

$\hat{\alpha} = \text{Mean}; \hat{\beta} = \text{Standard Deviation}$

Empirical Frequency และ Lognormal-3p แสดงอยู่ในรูปที่ 4.12



รูปที่ 4.12 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken line) of ξ_t variable of monthly mean NBS of Lake Michigan-Huron

(3d) การเปรียบเทียบ Sample correlogram $r_k(z)$ และ Residuals Correlogram $r_k(\hat{\varepsilon})$ ของ AR(1) และ AR(2) ในหัวข้อ (3b) แสดงว่า $r_k(\hat{\varepsilon})$ ของ AR(2) เหมาะสมกับ $r_k(z)$ มากกว่า $r_k(\hat{\varepsilon})$ ของ AR(1) และคาดว่า $r_k(\hat{\varepsilon})$ ของ AR(3) น่าจะเหมาะสมกับ $r_k(z)$ มากกว่า $r_k(\hat{\varepsilon})$ ของ AR(2) และน่าจะผ่านการทดสอบโดย Anderson Probability Limits

การทดสอบ Parsimony ของพารามิเตอร์โดยใช้ Akaike Information Criteria (AIC) พบว่า

$$AIC(1) = -41.34$$

$$AIC(2) = -44.59$$

$$AIC(3) = -41.72$$

$AIC(2)$ มีค่าต่ำสุด จึงสรุปอีกครั้งว่าควรจำลองโดยใช้แบบจำลอง AR(2)

(3e) Correlogram ของ AR(2) แสดงอยู่ในรูปที่ 4.11(a)

(4) Optional Tests of the Model

ทำการสังเคราะห์ห้อนุกรมเวลารายเดือนของ NBS ของ Lake Michigan-Huron โดยใช้แบบจำลอง AR(2) ตามสมการ 4.105 และ 4.106

$$\hat{y}_{v,\tau} = \hat{u}_\tau + \hat{\sigma}_\tau \cdot \hat{z}_{v,\tau} \quad [4.109]$$

$$\hat{z}_t = 0.209\hat{z}_{t-1} + 0.087\hat{z}_{t-2} + \sqrt{0.943}\xi_t \quad [4.110]$$

$\xi_t = \text{Lognormal 3 Distributed}$ โดยมีค่าพารามิเตอร์ดังนี้

$$(\hat{\alpha} = 1.356, \hat{\beta} = 0.246, \xi_0 = -4.0)$$

$$u = \frac{\ln(\xi_t - \xi_0) - \hat{\alpha}}{\hat{\beta}}$$

$$u = \frac{\ln(\xi_t - (-4)) - 1.356}{0.246} = \text{Standard Normal with mean}=0 \text{ and standard deviation}=1$$

$$\xi_t = -4 + \text{Exp}\{1.356 + 0.246u\} \quad [4.111]$$

\hat{u}_τ และ $\hat{\sigma}_\tau$ มีค่าดังตารางที่ 4.9

ตารางที่ 4.9 การคำนวณค่า \hat{u}_τ และ $\hat{\sigma}_\tau$

τ	1	2	3	4	5	6
\hat{u}_τ	46.93	95.25	176.80	277.60	254.98	210.49
$\hat{\sigma}_\tau$	50.96	48.58	72.15	86.70	88.03	65.01
τ	1	2	3	4	5	6
\hat{u}_τ	120.49	55.79	26.39	-5.87	33.13	25.35
$\hat{\sigma}_\tau$	52.45	59.06	63.80	72.31	66.25	61.48

การสังเคราะห์ (Generate) ค่า Standard Normal Random No. และ Log-normal 3 Transformed Random No. แสดงอยู่ในตารางที่ 4.10

ตารางที่ 4.10 การสังเคราะห์ Standard Normal Random No. และ Log-normal 3 Transformed Random No.

Order	u	ξ	Order	u	ξ	Order	u	ξ
1	0.414	0.297	7	-0.659	-0.700	13	1.195	1.207
2	-1.288	-1.173	8	0.595	0.492	14	-1.160	-1.083
3	1.019	0.986	9	-0.651	-0.694	15	-1.835	-1.529
4	0.616	0.515	10	0.906	0.849	16	-0.468	-0.541
5	-0.289	-0.386	11	0.678	0.585	17	0.68	-0.054
6	1.970	2.300	12	-1.175	-1.093	18	-0.595	-0.648

สมมติให้ $\hat{z}_0 = 0.0$ และ $\hat{z}_{-1} = 0.0$ จะสามารถสังเคราะห์ค่า \hat{z}_t ได้ดังนี้

$$\hat{z}_1 = 0.209(0.0) + 0.087(0.0) + 0.971(0.297) = 0.288$$

$$\hat{z}_2 = 0.209(0.288) + 0.087(0.) + 0.971(-1.173) = -1.079$$

$$\hat{z}_3 = 0.209(-1.079) + 0.087(0.288) + 0.971(0.986) = 0.757$$

$$\hat{z}_4 = 0.209(0.757) + 0.087(-1.079) + 0.971(0.515) = 0.564$$

$$\hat{z}_5 = 0.209(0.564) + 0.087(0.757) + 0.971(-0.386) = -0.191$$

เนื่องจากค่าแรกของ \hat{z}_t ที่สังเคราะห์ขึ้นมามี Bias เนื่องจากการสมมติค่า \hat{z}_0 และ \hat{z}_{-1} ดังนั้น

Fiering and Jackson (1971) จึงแนะนำว่าควรตัด 50 ค่าแรกที่สังเคราะห์ขึ้นมาทิ้ง

ตารางที่ 4.11 แสดงอนุกรมเวลา \hat{z}_t ที่ตัด 3 ค่าแรกทิ้ง หรือให้ $\hat{z}_t = \hat{z}_{t+3}$

ตารางที่ 4.11 อนุกรมเวลา \hat{z}_t ที่ตัด 3 ค่าแรกทิ้ง

t	\hat{z}_t	t	\hat{z}_t	t	\hat{z}_t
1	0.564	6	-0.607	11	-0.761
2	-0.191	7	0.733	12	-1.492
3	0.242	8	0.668	13	-0.903
4	-0.228	9	-0.858	14	-0.371
5	0.413	10	1.748	15	-0.785

หลังจากสังเคราะห์ค่า \hat{z}_t แล้ว จะสามารถคำนวณ Periodic Monthly Time Series ได้ดังนี้

$$\hat{y}_{1,1} = 46.93 + 50.96(0.564) = 75.67$$

$$\hat{y}_{1,2} = 95.25 + 48.58(-0.191) = 85.97$$

$$\hat{y}_{1,12} = 25.35 + 61.48(-1.492) = -66.38$$

$$\hat{y}_{2,1} = 46.93 + 50.96(-0.903) = 0.913$$

$$\hat{y}_{2,3} = 176.80 + 72.15(-0.785) = 120.16$$

รายละเอียดของ Optional Test จะแสดงอยู่ในหัวข้อ 4.3.5

(5) Reliability of Model Parameters

$$(5a) \quad s(\hat{\mu}_1) = \frac{\hat{\sigma}_1}{\sqrt{N}} = \frac{50.95}{\sqrt{69}} = 6.13$$

ค่า $s(\hat{\mu}_\tau)$ สำหรับ $\tau = 2, \dots, 12$ สามารถคำนวณได้ในลักษณะเดียวกัน

(5b) คำนวณหา 95% Confident Intervals ของ μ_τ จากสมการ 4.100

$$t_{68,0.025} = 2.0$$

$$46.93 - 2(6.13) < \mu_1 < 46.93 + 2(6.13)$$

$$34.67 < \mu_1 < 59.19$$

คำนวณหา 95% Confident Intervals ของ μ_τ ; $\tau = 2, \dots, 12$ ได้ในลักษณะเดียวกัน

(5c) คำนวณหา 95% Confident Intervals ของ σ_τ จากสมการ 4.101

$$\chi_{68,0.975}^2 = 47.1$$

$$\chi_{68,0.025}^2 = 92.66$$

$$\frac{\sqrt{68}(50.96)}{\sqrt{92.66}} < \sigma_1 < \frac{\sqrt{68}(50.96)}{\sqrt{47.11}}$$

$$43.65 < \sigma_1 < 61.23$$

คำนวณหา 95% Confident Intervals ของ $\sigma_\tau; \tau = 1, \dots, \omega$ ได้ในลักษณะเดียวกัน

(5d) เนื่องจากจำลองโดย AR(2) แบบ Constant Autoregressive Coefficients จึงสามารถคำนวณหา Variance ของ $\hat{\phi}_1$ โดยวิธี Moment ได้โดยใช้สมการ 4.46 ดังนี้

$$\begin{aligned} \text{Var}(\hat{\phi}_1) &= \text{Var}(\hat{\phi}_2) = \frac{1-\hat{\phi}_2^2}{N-2} \\ &= \frac{1-0.087^2}{828-2} = 0.001201 \\ s(\hat{\phi}_1) &= s(\hat{\phi}_2) = 0.035 \end{aligned}$$

และสามารถคำนวณหา $\text{Cov}(\hat{\phi}_1, \hat{\phi}_2)$ ได้จากสมการ 4.47

$$\begin{aligned} \text{Cov}(\hat{\phi}_1, \hat{\phi}_2) &= \frac{-\hat{\phi}_1(1+\hat{\phi}_2)}{N-2} \\ &= \frac{-0.209(1+0.087)}{828-2} \\ &= -0.000275 \\ \rho(\hat{\phi}_1, \hat{\phi}_2) &= \frac{\text{Cov}(\hat{\phi}_1, \hat{\phi}_2)}{\sqrt{\text{Var}(\hat{\phi}_1) \cdot \text{Var}(\hat{\phi}_2)}} \\ &= \frac{-0.000275}{0.001201} = -0.229 \end{aligned}$$

(5e) คำนวณหา 95% Confident Intervals ของ ϕ_1, ϕ_2 จากสมการ 4.48

$$\begin{aligned} u_{0.025} &= 1.96 \\ \hat{\phi}_j - u_{\frac{\alpha}{2}} s(\hat{\phi}_j) &< \phi_j < \hat{\phi}_j + u_{\frac{\alpha}{2}} s(\hat{\phi}_j) \\ 0.209 - 1.96(0.035) &< \phi_1 < 0.209 + 1.96(0.035) \\ 0.1404 &< \phi_1 < 0.2776 \\ 0.087 - 1.96(0.035) &< \phi_2 < 0.087 + 1.96(0.035) \\ 0.0184 &< \phi_2 < 0.1556 \end{aligned}$$

Option: หา Joint Confidence Region ของ ϕ_1, ϕ_2 โดยสมมติว่า ϕ_1, ϕ_2 มีการแจกแจงแบบ

Bivariate Normal Distribution โดยมีพารามิเตอร์ดังนี้

$$\begin{aligned} \mu(\hat{\phi}_1) &= 0.209 \\ s(\hat{\phi}_1) &= 0.035 \\ \mu(\hat{\phi}_2) &= 0.087 \\ s(\hat{\phi}_2) &= 0.035 \\ \rho(\hat{\phi}_1, \hat{\phi}_2) &= -0.229 \end{aligned}$$

(5f) คำนวณหา 95% Confident Intervals ของ σ_ε^2 จากสมการ 4.49

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{N\hat{\sigma}_\varepsilon^2(1+\hat{\phi}_2)}{(N-2)(1-\hat{\phi}_2)} \left\{ (1-\hat{\phi}_2)^2 - \hat{\phi}_1^2 \right\} \\ \text{Upper } \hat{\sigma}_\varepsilon^2 &= \frac{828(1+0.0184)}{(828-2)(1-0.0184)} \left\{ (1-0.0184)^2 - 0.1404^2 \right\} \\ &= 0.982 \quad (\text{คำนวณจาก Lower } \phi_j) \\ \text{Lower } \hat{\sigma}_\varepsilon^2 &= \frac{828(1+0.01556)}{(828-2)(1-0.01556)} \left\{ (1-0.01556)^2 - 0.2776^2 \right\} \\ &= 0.872 \quad (\text{คำนวณจาก Upper } \phi_j) \end{aligned}$$

4.4 เอกสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series.**

Water Resources Publications. USA. pp.484

APPENDIX A4.1 AUTOCORRELATION FUNCTION OF AR(p) MODELS

Section 4.1.2 indicated that the autocorrelation function ρ_k of AR models satisfies the difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k > 0 \quad (\text{A4.1})$$

where ϕ_1, \dots, ϕ_p are the autoregression coefficients of the AR(p) model of Eq. (4.1). For given or estimated set of ϕ coefficients, Eq. (A4.1) can be solved simultaneously to obtain $\rho_1, \dots, \rho_{p-1}$. Then for $k \geq p$, Eq. (A4.1) may be used recursively to find ρ_k . Equations (4.14), (4.15) and (4.16) give the autocorrelation coefficients ρ_k for the AR(1) and AR(2) models. A general computational procedure for determining ρ_k follows.

In general, for an autoregressive model of order p , the matrix solution for the $p-1$ autocorrelation coefficients $\rho_1, \dots, \rho_{p-1}$ of Eq. (A4.1) has the form (Salas and Smith, 1978)

$$\rho = A^{-1} [-\Phi] \quad (\text{A4.2})$$

where $\rho = [\rho_1, \rho_2, \dots, \rho_{p-1}]^T$, $[-\Phi] = [-\phi_1, -\phi_2, \dots, -\phi_{p-1}]^T$ and T denotes the transpose of the matrix. The elements of the matrix A may be obtained from

$$\text{diagonal} \quad a_{ij} = \phi_{2j} - 1 \quad ; \quad i = j = 1, \dots, (p-1) \quad (\text{A4.3})$$

$$\text{lower triangular} \quad a_{ij} = \phi_{i-j} + \phi_{i+j} \quad ; \quad i > 1, j < i \quad (\text{A4.4})$$

$$\text{and upper triangular} \quad a_{ij} = \phi_{i+j} \quad ; \quad i \geq 1, j > i \quad (\text{A4.5})$$

in which $\phi_i = 0$ for $i > p$. Once the $\rho_1, \dots, \rho_{p-1}$ autocorrelation coefficients are obtained from Eq. (A4.2), the coefficients ρ_k for $k \geq p$ are obtained recursively from Eq. (A4.1).

APPENDIX A4.2 PARTIAL AUTOCORRELATION FUNCTION OF AR(p) MODELS

The partial autocorrelation coefficient $\phi_k(k)$ in an AR process of order k is a measure of the linear association

between ρ_j and ρ_{j-k} for $j \leq k$. It is the k -th autoregressive coefficient and $\phi_k(k)$ for $k=1,2,\dots$ is the partial autocorrelation function.

The difference equation for an AR(k) model is (see Sec. 4.1.2)

$$\begin{aligned} \rho_j &= \phi_1(k)\rho_{j-1} + \phi_2(k)\rho_{j-2} + \dots + \phi_k(k)\rho_{j-k} ; \\ j &= 1, \dots, k \end{aligned} \quad (\text{A4.6})$$

where $\phi_j(k)$ is the j -th autoregressive coefficient of the AR(k) model. The partial autocorrelation is given by the last coefficient $\phi_k(k)$, $k = 1, 2, \dots$.

Equation (A4.6) constitutes the set of linear equations

$$\begin{array}{l} \mathbf{1} \quad \phi_1(k)\rho_0 + \phi_2(k)\rho_1 + \dots + \phi_k(k)\rho_{k-1} = \rho_1 \\ \mathbf{2} \quad \phi_1(k)\rho_1 + \phi_2(k)\rho_0 + \dots + \phi_k(k)\rho_{k-2} = \rho_2 \\ \mathbf{3} \quad \phi_1(k)\rho_2 + \phi_2(k)\rho_1 + \dots + \phi_k(k)\rho_{k-3} = \rho_3 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \mathbf{k} \quad \phi_1(k)\rho_{k-1} + \phi_2(k)\rho_{k-2} + \dots + \phi_k(k)\rho_0 = \rho_k \end{array}$$

which may be written as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1(k) \\ \phi_2(k) \\ \phi_3(k) \\ \cdot \\ \cdot \\ \phi_k(k) \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \cdot \\ \rho_k \end{bmatrix} \quad (\text{A4.7})$$

or

$$\begin{aligned} P_k \phi_k &= \rho_k \\ \phi_k &= P_k^{-1} \rho_k, \quad k = 1, 2, \dots \end{aligned} \quad (\text{A4.8})$$

Thus the partial autocorrelation function $\phi_k(k)$ is determined by successively applying Eq. (A4.8).

The partial autocorrelation function $\phi_k(k)$ may be also obtained recursively by means of Durbin's (1960) relations

$$\phi_1(1) = \rho_1, \quad \phi_1(2) = \frac{\rho_1(1 - \rho_2)}{(1 - \rho_1^2)}, \quad \phi_2(2) = \frac{\rho_2 - \rho_1^2}{(1 - \rho_1^2)}$$

$$\phi_k(k) = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_j(k-1) \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_j(k-1) \rho_j} \quad (\text{A4.9})$$

$$\phi_j(k) = \phi_j(k-1) - \phi_k(k) \phi_{k-j}(k-1)$$

To determine the partial autocorrelation function $\hat{\phi}_k(k)$ from a sample series z_1, \dots, z_N , first compute the sample autocorrelations r_k from Eq. (2.5b), then replace the ρ 's by the r 's in either Eq. (A4.8) or Eq. (A4.9).

On the hypothesis that the process is AR(p), the estimated $\hat{\phi}_k(k)$ for $k > p$ is asymptotically normal with mean zero and variance $1/N$. Hence, the $1 - \alpha$ probability limits for zero partial autocorrelation may be determined by (Box and Jenkins, 1970, p. 65 and 178).

$$\{-u_{1-\alpha/2}/\sqrt{N} \ ; \ +u_{1-\alpha/2}/\sqrt{N}\} \quad (\text{A4.10})$$

where $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution, N is the sample size and α is the probability level. The limits of expression (A4.10) may be used to give some guide as to whether theoretical partial autocorrelations are practically zero beyond a particular lag.

APPENDIX A4.3 ANNUAL FLOWS OF THE GÖTA RIVER, SWEDEN

The following table gives the annual flows of the Göta River near Sjötop-Vänersburg for the period 1901-1950. The data is in the form of modular values (actual annual flows divided by the mean) as given by Yevjevich (1963).

Years	0.935	0.662	0.950	1.121	0.880	0.802	0.856	1.080	0.959	1.345
01-10	0.935	0.662	0.950	1.121	0.880	0.802	0.856	1.080	0.959	1.345
11-20	1.153	0.929	1.158	0.957	0.705	0.905	1.000	0.948	0.907	0.991
21-30	0.994	0.701	0.692	1.086	1.306	0.895	1.149	1.297	1.168	1.218
31-40	1.209	0.974	0.834	0.638	0.991	1.198	1.091	0.892	1.020	0.869
41-50	0.772	0.606	0.739	0.813	1.173	0.916	0.880	0.601	0.720	0.955

**การจำลองอนุกรมเวลาด้วยแบบจำลอง AUTOREGRESSIVE AND MOVING AVERAGE
AUTOREGRESSIVE AND MOVING AVERAGE MODELLING**

5.1 คำนำ (Introduction)

แบบจำลอง Autoregressive and Moving Average ซึ่งเรียกสั้นๆ ว่า ARMA(p, q) คือ แบบจำลองสโตแคสติกที่พัฒนาต่อจากแบบจำลอง AR(p) ในบทที่ 4 โดยเพิ่มองค์ประกอบทางสโตแคสติกอีกองค์ประกอบหนึ่งคือ องค์ประกอบ Moving Average เข้าในแบบจำลอง ทำให้ ARMA(p, q) มีข้อดีเพิ่มขึ้นคือเหมาะสำหรับสามารถอธิบายปรากฏการณ์การไหลของน้ำที่มีลักษณะทั้ง Low Flow และ High Flow ผสมผสานกันไป องค์ประกอบ AR(p) และ MA(q) สามารถที่จะใช้อธิบายปรากฏการณ์ทางอุทกวิทยาที่แตกต่างกันดังนี้

AR(p) – ทำให้เกิด Low Flow/Baseflow หรือน้ำท่าในช่วงฤดูแล้ง ซึ่งเกิดจากอิทธิพลของน้ำใต้ดิน

MA(q) - ทำให้เกิด High Flow/Direct Runoff หรือน้ำท่าซึ่งเกิดจากฝน

ARMA(p, q) – จึงเป็นการผสมผสานของ 2 ปรากฏการณ์

ในหัวข้อ 1.6 ในบทที่ 1 ได้พิสูจน์ให้เห็นว่าสามารถใช้ ARMA(1,1) อธิบายปริมาณการไหลของน้ำในแม่น้ำได้ ดังนี้

$$\begin{aligned} Z_t &= cS_{t-1} + dX_t \\ S_t &= (1 - c)S_{t-1} + aX_t \\ Z_t &= (1 - c)Z_{t-1} + dX_t - [d(1 - c) - ac]X_{t-1} \dots \dots \dots ARMA(1,1) \end{aligned}$$

แบบจำลอง ARMA(p, q) มีข้อดีที่สำคัญคือ มีจำนวนพารามิเตอร์น้อยกว่าแบบจำลอง AR(p) เมื่อใช้แบบจำลองที่มีลำดับมาก จึงน่าสนใจมากกว่าเมื่อพิจารณาจากเกณฑ์ Parsimony

5.2 รูปแบบทางคณิตศาสตร์ของแบบจำลอง ARMA(p, q) (Mathematical Formula of ARMA)

$$\begin{aligned} AR(p): Z_t &= \sum_{j=1}^p \phi_j Z_{t-j} + \varepsilon_t \\ MA(q): Z_t &= -\sum_{j=0}^q \theta_j \varepsilon_{t-j}; \theta_0 = -1 \\ ARMA(p, q) &= AR(p) + MA(q) \\ Z_t &= \sum_{j=1}^p \phi_j Z_{t-j} - \sum_{j=0}^q \theta_j \varepsilon_{t-j} \end{aligned} \tag{5.1}$$

โดยมีพารามิเตอร์คือ $\{\mu, \sigma^2, \sigma_\varepsilon^2, \phi_j (j = 1, \dots, p), \theta_j (j = 0, \dots, q)\}$

5.3 คุณสมบัติของ MA(q) (Properties of MA(q))

5.3.1 แบบจำลอง MA(q)

$$\begin{aligned} Z_t &= -\sum_{j=0}^q \theta_j \varepsilon_{t-j}; \theta_0 = -1 \\ E(\varepsilon_t) &= 0 = E(Z_t) \\ Var(\varepsilon_t) &= E(\varepsilon_t)^2 = \sigma_\varepsilon^2 \\ E(Z_t \cdot Z_{t-k}) &= Cov(Z_t, Z_{t-k}) = \gamma_k \\ &= Lag k Autocovariance \text{ ของ } Z_t \end{aligned} \tag{5.2}$$

$$\gamma_k = E\left(-\sum_{j=0}^q \theta_j \varepsilon_{t-j}\right)\left(-\sum_{j=0}^q \theta_j \varepsilon_{t-k-j}\right) \quad [5.3]$$

$$\gamma_k = E(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})(\varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1} - \dots - \theta_q \varepsilon_{t-k-q})$$

$$\begin{aligned} \gamma_k &= \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}; k \leq q \\ &= 0 \quad ; k > q \end{aligned} \quad [5.4]$$

$$\text{ถ้า } k = 0 \quad ; \quad \gamma_0 = \text{Var}(Z_t) = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2 \quad [5.5]$$

$$\text{ACF} \quad ; \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sigma_\varepsilon^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{\sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2} \quad \text{เมื่อ } k \leq q \quad [5.6]$$

$$\rho_k = 0 \quad \text{เมื่อ } k > q \quad [5.7]$$

5.3.2 แบบจำลอง MA(1) จะมีลักษณะเป็น One Step Long Memory

$$z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad [5.8]$$

จากสมการ [5.4] ถ้า

$$\begin{aligned} q = 1, k = 1; \gamma_1 &= \sigma_\varepsilon^2 \sum_{j=0}^{1-1} \theta_j \theta_{j+k} = \sigma_\varepsilon^2 \theta_0 \theta_1 \\ &= -\sigma_\varepsilon^2 \theta_1 \end{aligned} \quad [5.9]$$

$$\begin{aligned} q = 1, k = 0; \gamma_0 &= \sigma_\varepsilon^2 \sum_{j=0}^1 \theta_j \theta_{j+k} \\ &= \sigma_\varepsilon^2 (\theta_0 \cdot \theta_0 + \theta_1 \cdot \theta_1) \\ &= (1 + \theta_1^2) \sigma_\varepsilon^2 \end{aligned} \quad [5.10]$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = -\frac{\theta_1}{(1+\theta_1^2)} \quad [5.11]$$

$$q = 1, k > 1, \rho_k = 0 \quad [5.12]$$

5.3.3 Residual ของ MA(1)

$$\begin{aligned} \varepsilon_t &= Z_t + \theta_1 \varepsilon_{t-1} \\ \varepsilon_{t-1} &= Z_{t-1} + \theta_1 \varepsilon_{t-2} \\ \varepsilon_t &= Z_t + \theta_1 (Z_{t-1} + \theta_1 \varepsilon_{t-2}) \\ &= Z_t + \theta_1 Z_{t-1} + \theta_1^2 \varepsilon_{t-2} \\ &= Z_t + \theta_1 Z_{t-1} + \theta_1^2 Z_{t-2} + \dots \end{aligned} \quad [5.13]$$

สมการจะ Converge เมื่อ $|\theta_1| < 1$

$$Z_t = -\theta_1 Z_{t-1} - \theta_1^2 Z_{t-2} + \dots + \varepsilon_t \quad [5.14]$$

จึงสรุปได้ว่า MA(1) คือ AR(∞) และมีลักษณะเป็น Long Memory

5.3.4 Invertible Conditions

$$|\theta_1| < 1$$

หรือ รากของสมการ Characteristics จะอยู่ในวงกลมที่มีรัศมีหนึ่งหน่วย (Unit Circle)

$$r - \theta_1 = 0 \quad [5.15]$$

กรณีของ MA(q)

$$r^q - \theta_1 r^{q-1} - \theta_2 r^{q-2} - \dots - \theta_q = 0 \quad [5.16]$$

ถ้ารากของสมการ [5.16] อยู่ใน Unit Circle แสดงว่า θ_j เป็นไปตาม Invertible Condition

กรณี MA(2) ซึ่งมี $\theta_1 = 0.6$ และ $\theta_2 = 0.2$

$$r^2 - 0.6r - 0.2 = 0$$

$$r_1 = 0.839$$

$$r_2 = -0.2385$$

5.3.5 Partial Autocorrelation Function (PACF) ของ MA(q)

PACF หรือ $\phi_k(k)$ ของ MA(q) ก็คือ ACF ของ AR(p)

ACF หรือ ρ_k ของ MA(q) ก็คือ PACF ของ AR(p)

5.3.6 คุณสมบัติทั่วไปของ ARMA(p, q)

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Lag k Cross-Covariance ของ z และ ε

$$\gamma_{z\varepsilon}(k) = Cov(z_{t-k}, \varepsilon_t)$$

$$\gamma_{z\varepsilon}(k) = E(z_{t-k}, \varepsilon_t)$$

$$= E[(\sum_{j=1}^p \phi_j Z_{t-k-j} - \sum_{j=0}^q \theta_j \varepsilon_{t-k-j}) \varepsilon_t]$$

$$= \sum_{j=1}^p \phi_j \gamma_{z\varepsilon}(k+j) - \sum_{j=0}^q \theta_j \gamma_{\varepsilon\varepsilon}(k+j); k \leq 0 \quad [5.17]$$

$$\gamma_{z\varepsilon}(k) = 0; k > 0 \quad [5.18]$$

(1) Lag k Autocovariance ของ ARMA(p, q)

$$E(z_t \cdot z_{t-k}) = \gamma_k$$

$$\gamma_k = \sum_{j=1}^p \phi_j \gamma_{k-j} - \sum_{j=0}^q \theta_j \gamma_{z\varepsilon}(k-j); k < q + 1 \text{ or } k \leq q$$

$$= \sum_{j=1}^p \phi_j \gamma_{k-j}; k \geq q + 1 \quad [5.19]$$

ถ้า $k=0$

$$\gamma_0 = E(z_t \cdot z_t) = Var(Z_t)$$

$$= \sigma_\varepsilon^2 + \sum_{j=1}^p \phi_j \gamma_j - \sum_{j=1}^q \theta_j \gamma_{z\varepsilon}(-j) \quad [5.20]$$

(2) Autocorrelation Function(ACF)

$$\frac{\gamma_k}{\gamma_0} = \rho_k = \sum_{j=1}^p \phi_j \rho_{k-j}; k \geq q + 1 \quad [5.21]$$

ACF ของ ARMA(p,q) จะแปรผันตามค่า ϕ_j และ θ_j ดังนี้

$\rho_j (j = 1, \dots, q)$ จะแปรผันตาม $\phi_j (j = 1, \dots, p)$ และ $\theta_j (j = 1, \dots, q)$

$\rho_j (j = q + 1, \dots, q + p + 1)$ จะแปรผันตาม $\phi_j (j = 1, \dots, p)$

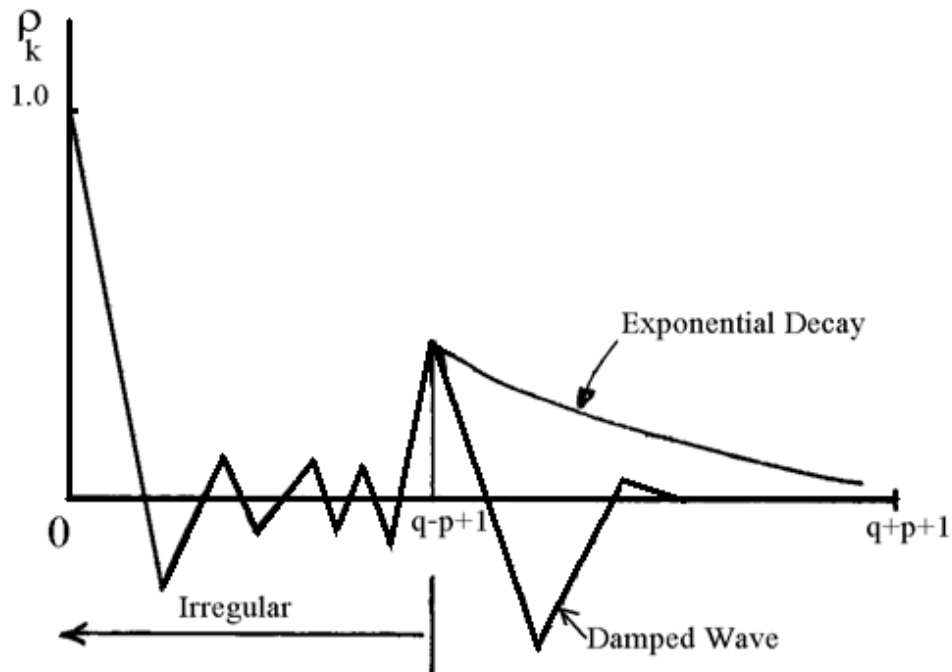
ซึ่งจะมีผลทำให้ ACF ของ ARMA(p,q) มีลักษณะที่แตกต่างกันดังรูปที่ 5.1 ถ้า $q-p \geq 0$ เช่น

ARMA(1,2) ACF จะมีลักษณะเป็น Irregular Shape ตามด้วย Exponential Decay หรือ Damped Wave แต่ถ้า

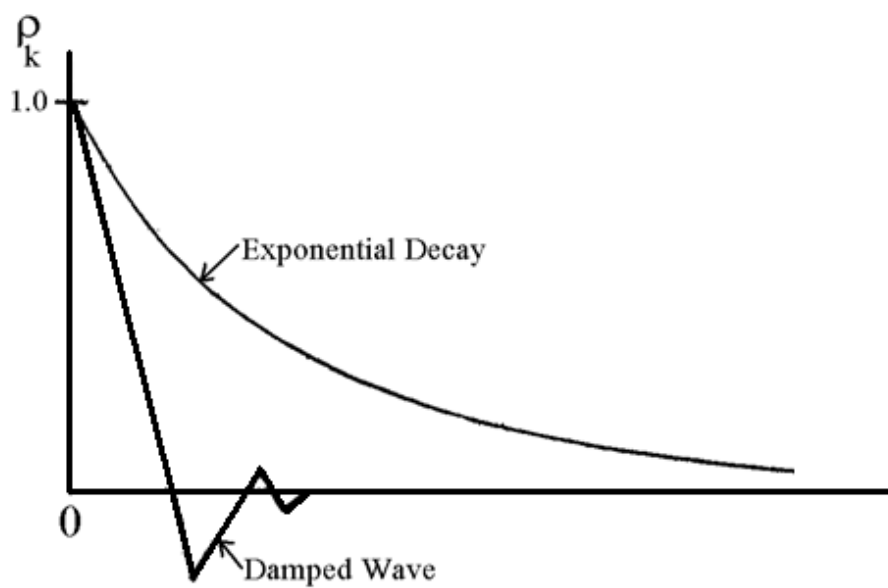
$q-p < 0$ เช่น ARMA(1,0) ACF จะมีรูปร่างเป็น Exponential Decay หรือ Damped Wave ดังรูปที่ 5.1

PACF ของ MA(p) แสดงอยู่ในรูปที่ 5.2 และ 5.3

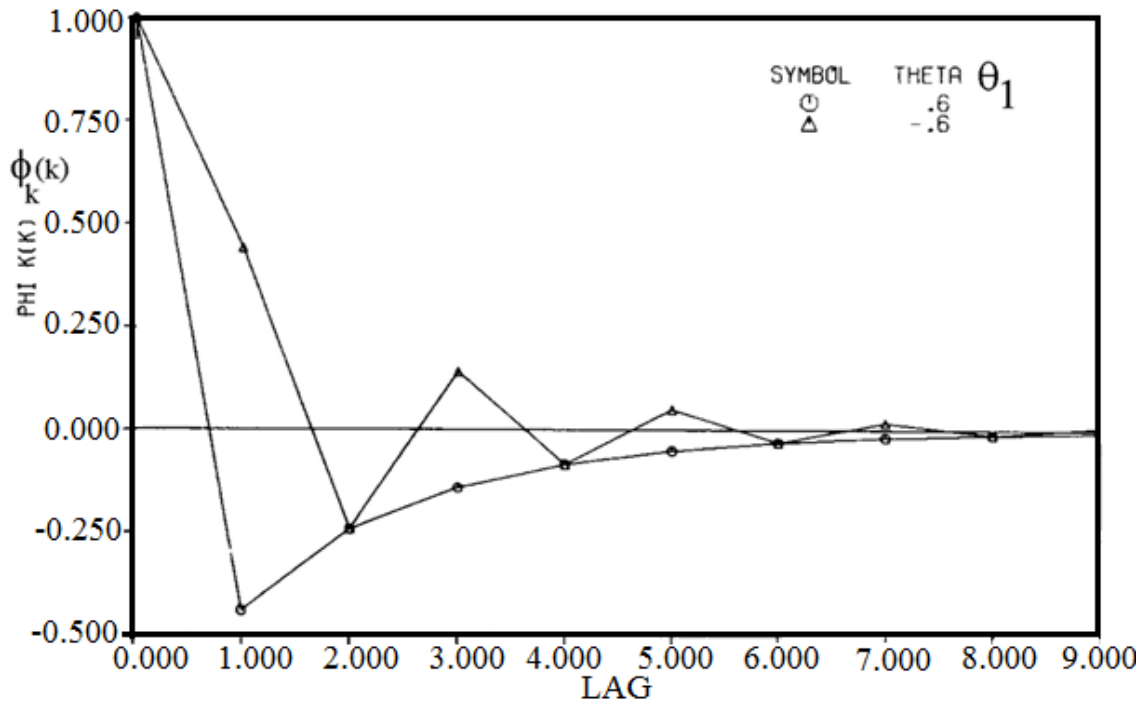
If $q-p \geq 0$ i.e. ARMA(1,1) or ARMA(1,2)



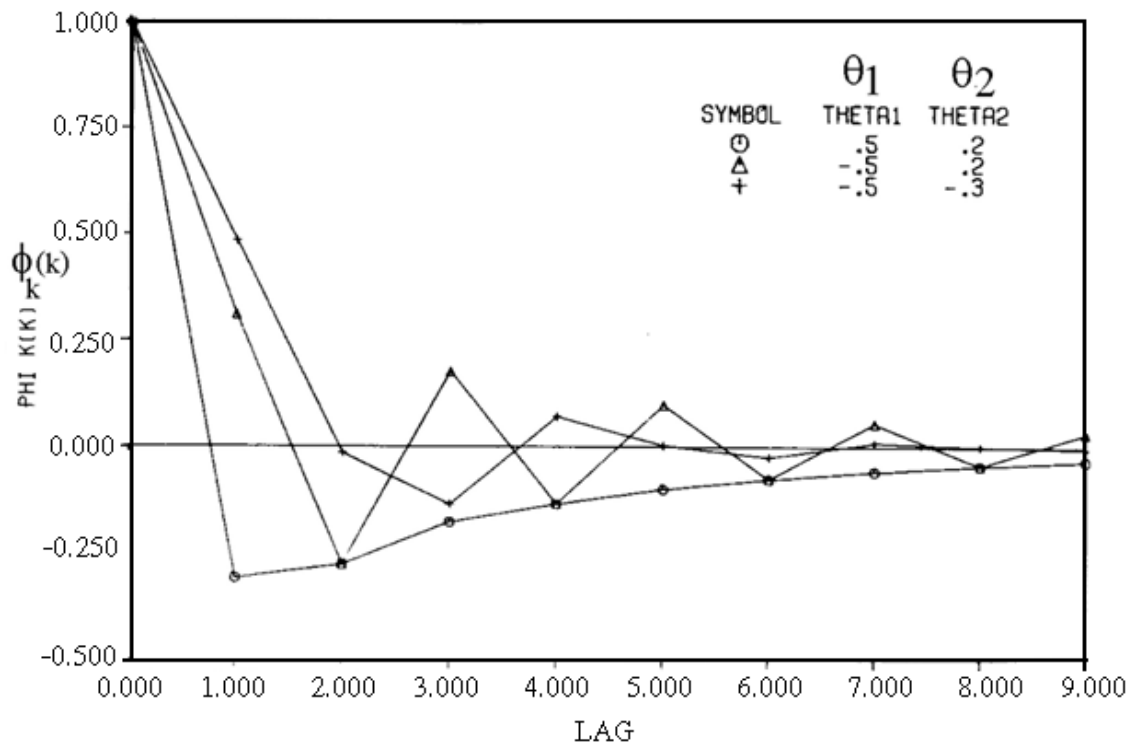
If $q-p < 0$ i.e. ARMA(1,0) or ARMA(2,1)



รูปที่ 5.1 ลักษณะของ ACF ของ ARMA(p,q) กรณีที่ p และ q มีค่าต่างกัน



รูปที่ 5.2 Theoretical Partial Autocorrelation for MA(1) Model



รูปที่ 5.3 Theoretical Partial Autocorrelation for MA(2) Model

Stationary and Invertibility

$$u^p - \phi_1 u^{p-1} - \phi_2 u^{p-2} - \dots - \phi_p = 0 \tag{5.22}$$

$$r^q - \theta_1 r^{q-1} - \theta_2 r^{q-2} - \dots - \theta_q = 0 \tag{5.23}$$

(3) คุณสมบัติพื้นฐานของ ARMA(1, 1)

$$\begin{aligned} Z_t &= \phi_1 Z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ Z_t &= \phi_1 (\phi_1 Z_{t-2} + \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= \phi_1^2 Z_{t-2} + \varepsilon_t + (\phi_1 - \theta_1) \varepsilon_{t-1} - \phi_1 \theta_1 \varepsilon_{t-2} \\ Z_t &= \phi_1^2 (\phi_1 Z_{t-3} + \varepsilon_{t-2} - \theta_1 \varepsilon_{t-3}) + \varepsilon_t + (\phi_1 - \theta_1) \varepsilon_{t-1} - \phi_1 \theta_1 \varepsilon_{t-2} \\ &= \phi_1^3 Z_{t-3} + \varepsilon_t + (\phi_1 - \theta_1) \varepsilon_{t-1} - \phi_1 (\phi_1 - \theta_1) \varepsilon_{t-2} - \phi_1^2 \theta_1 \varepsilon_{t-3} \\ &= \varepsilon_t + (\phi_1 - \theta_1) \varepsilon_{t-1} + \phi_1 (\phi_1 - \theta_1) \varepsilon_{t-2} + \phi_1^2 (\phi_1 - \theta_1) \varepsilon_{t-3} + \dots \end{aligned}$$

ดังนั้น จึงสามารถสรุปได้ว่า ARMA(1, 1) คือ MA(∞) ดังสมการ

$$z_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \tag{5.24}$$

สมการ [5.25] จะ Converged เมื่อ $|\theta_1| < 1$ หรือเมื่อกระบวนการสโตแคสติกเป็น Stationary ถ้า ARMA(1,1) มีพารามิเตอร์

$$\begin{aligned} \phi_1 &= 0.79 \\ \theta_1 &= 0.35 \\ \psi_i &= \phi_1^{i-1} (\phi_1 - \theta_1) \end{aligned} \tag{5.25}$$

i	1	2	3	4	5	6
ψ_i	0.44	0.35	0.27	0.22	0.17	0.14

ในทำนองเดียวกัน ARMA(1, 1) สามารถเขียนในรูปของ AR(∞) ดังสมการ

$$\begin{aligned} \varepsilon_t &= z_t - \phi_1 z_{t-1} + \theta_1 \varepsilon_{t-1} \\ &= z_t - \phi_1 z_{t-1} + \theta_1 (z_{t-1} - \phi_1 z_{t-2} + \theta_1 \varepsilon_{t-2}) \\ &= z_t - (\phi_1 - \theta_1) z_{t-1} - \theta_1 \phi_1 z_{t-2} + \theta_1^2 \varepsilon_{t-2} \\ z_t &= (\phi_1 - \theta_1) z_{t-1} + \theta_1 \phi_1 z_{t-2} - \theta_1^2 \varepsilon_{t-2} + \varepsilon_t \\ &= (\phi_1 - \theta_1) z_{t-1} + \theta_1 (\phi_1 - \theta_1) z_{t-2} + \theta_1^2 (\phi_1 - \theta_1) \varepsilon_{t-3} + \dots + \varepsilon_t \\ &= \pi_1 z_{t-1} + \pi_2 z_{t-2} + \pi_3 z_{t-3} + \dots + \varepsilon_t \end{aligned} \tag{5.26}$$

$$\pi_i = \theta_1^{i-1}(\phi_1 - \theta_1)$$

$$\text{ถ้า } \phi_1 = 0.79, \theta_1 = 0.35$$

i	1	2	3	4	5
π_i	0.44	0.15	0.05	0.02	0.01

สมการ [5.22] จะ Converged ถ้ากระบวนการสโตแคสติกเป็นแบบ Invertibility ถ้า $|\theta_1| \leq 1$
Admissible Region สำหรับ ϕ_1 และ θ_1 คือ

$$-1 < \phi_1 < 1 \text{ และ } -1 < \theta_1 < 1 \quad [5.27]$$

สำหรับกระบวนการสโตแคสติกทางอุทกวิทยา

$$0 < \phi_1 < 1, 0 < \theta_1 < 1 \text{ และ } \phi_1 > \theta_1 \quad [5.28]$$

(4) Variance ของ ARMA(1,1), γ_0

จากสมการ [5.20]

$$\gamma_0 = \sigma_\varepsilon^2 + \phi_1\gamma_1 - \theta_1\gamma_{z\varepsilon}(-1)$$

$$\begin{aligned} \gamma_{z\varepsilon}(-1) &= E(z_t \cdot \varepsilon_{t-1}) = E(\phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1}) \\ &= \phi_1 E(z_{t-1} \cdot \varepsilon_{t-1}) + E(\varepsilon_t \cdot \varepsilon_{t-1}) - \theta_1 E(\varepsilon_{t-1} \cdot \varepsilon_{t-1}) \\ &= \phi_1 \gamma_{z\varepsilon}(0) - \theta_1 \sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} \gamma_{z\varepsilon}(0) &= E(z_t \cdot \varepsilon_t) = E(\phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_t) \\ &= E(\phi_1 z_{t-1} \varepsilon_t) + E(\varepsilon_t \cdot \varepsilon_t) - E(\theta_1 \varepsilon_{t-1} \cdot \varepsilon_t) \\ &= \sigma_\varepsilon^2 \end{aligned} \quad [5.29]$$

$$\gamma_0 = \sigma_\varepsilon^2 + \theta_1\gamma_1 - \theta_1(\phi_1 - \theta_1)\sigma_\varepsilon^2 \quad [5.30]$$

จากสมการ [5.19] ถ้า $k = 1$

$$\begin{aligned} \gamma_1 &= \phi_1\gamma_0 - \theta_0\gamma_{z\varepsilon}(1) - \theta_1\gamma_{z\varepsilon}(0) \\ &= \phi_1\gamma_0 - \theta_1\sigma_\varepsilon^2 \end{aligned} \quad [5.31]$$

ถ้า $k \geq 2$

$$\begin{aligned} \gamma_2 &= \phi_1\gamma_1 - \theta_0\gamma_{z\varepsilon}(2) - \theta_1\gamma_{z\varepsilon}(1) \\ &= \phi_1\gamma_1 \\ \gamma_k &= \phi_1\gamma_{k-1} \end{aligned} \quad [5.32]$$

จากสมการ [5.30]

$$\gamma_0 - \phi_1\gamma_1 = [1 - \theta_1(\phi_1 - \theta_1)]\sigma_\varepsilon^2 \quad [5.33]$$

จากสมการ [5.31]

$$\phi_1\gamma_0 - \gamma_1 = \theta_1\sigma_\varepsilon^2 \quad [5.34]$$

แก้สมการ [5.33] และ [5.34] จะได้ค่า γ_0 และ γ_1 ดังนี้

$$\gamma_0 = \frac{(1+\theta_1^2-2\phi_1\theta_1)}{(1-\phi_1^2)}\sigma_\varepsilon^2 \quad [5.35]$$

$$\gamma_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{(1-\phi_1^2)}\sigma_\varepsilon^2 \quad [5.36]$$

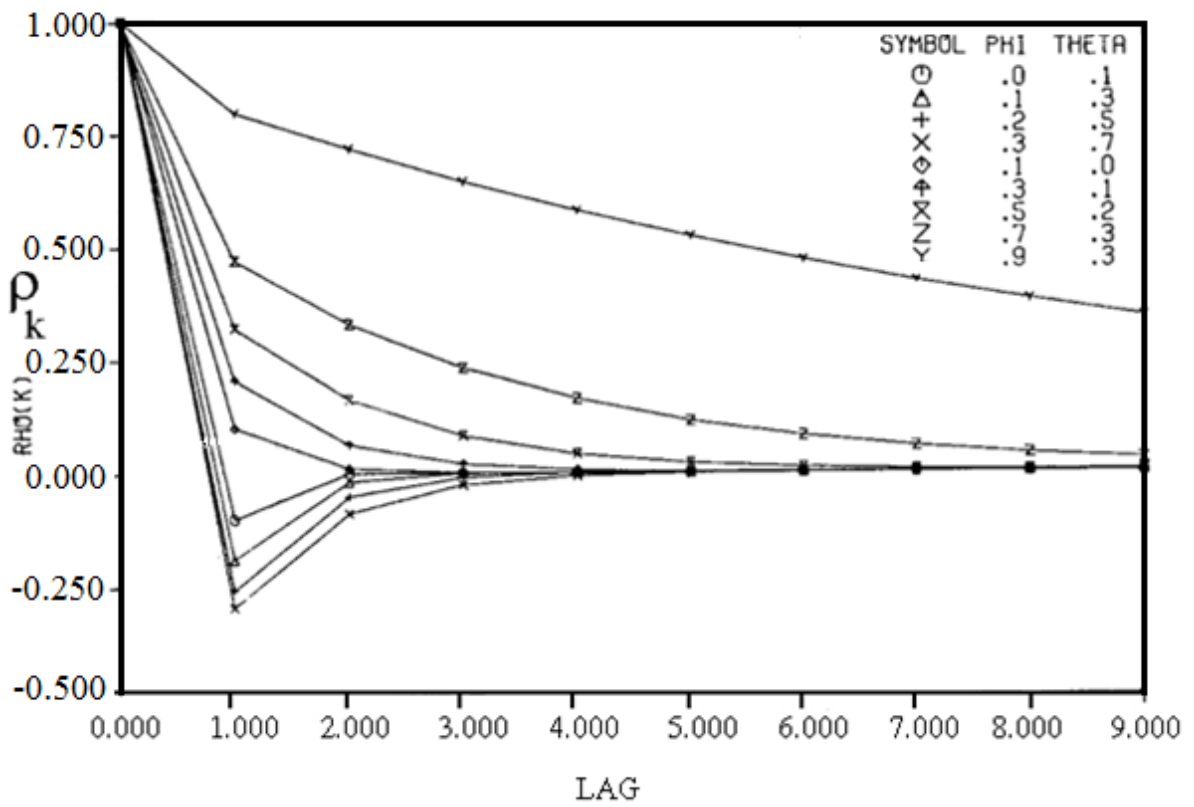
$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{(1+\theta_1^2-2\phi_1\theta_1)} \quad [5.37]$$

ถ้าหารสมการ [5.32] ด้วย γ_0 จะได้

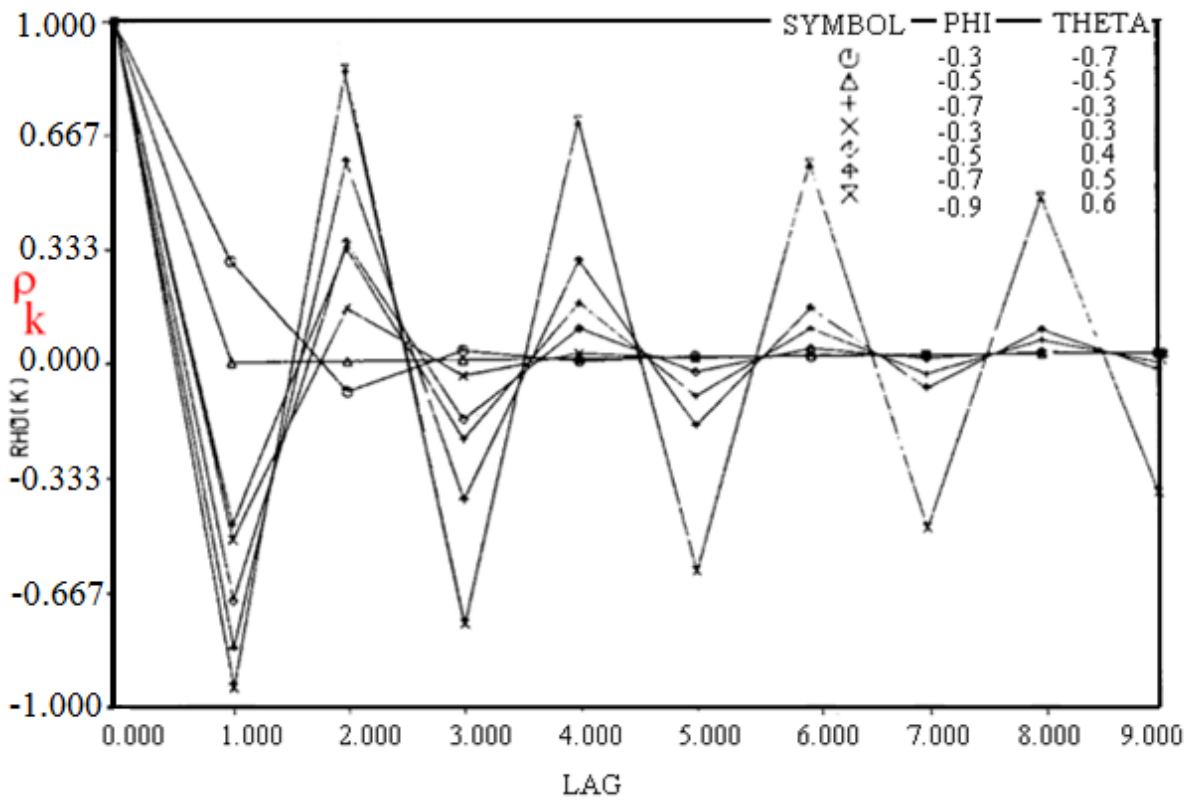
$$\rho_k = \phi_1\rho_{k-1}; k \geq 2 \quad [5.38]$$

กรณี ARMA(1,1) จะสามารถหาค่า ϕ_1 จากสมการ [5.38] เมื่อ $k=2$ แทนค่า ϕ_1 ลงในสมการ [5.37] จะสามารถหาค่า θ_1 ได้โดยวิธี Trial and Error และขั้นสุดท้ายแทนค่า ϕ_1 และ θ_1 ลงในสมการ [5.35] หรือ [5.36] เพื่อหาค่า σ_ε^2

กราฟแสดง ACF และ PACF ของ ARMA(1,1) กรณีที่ ϕ_1 และ θ_1 มีค่าต่างๆ กัน แสดงอยู่ในรูปที่ 5.4 และ 5.5 ตามลำดับ

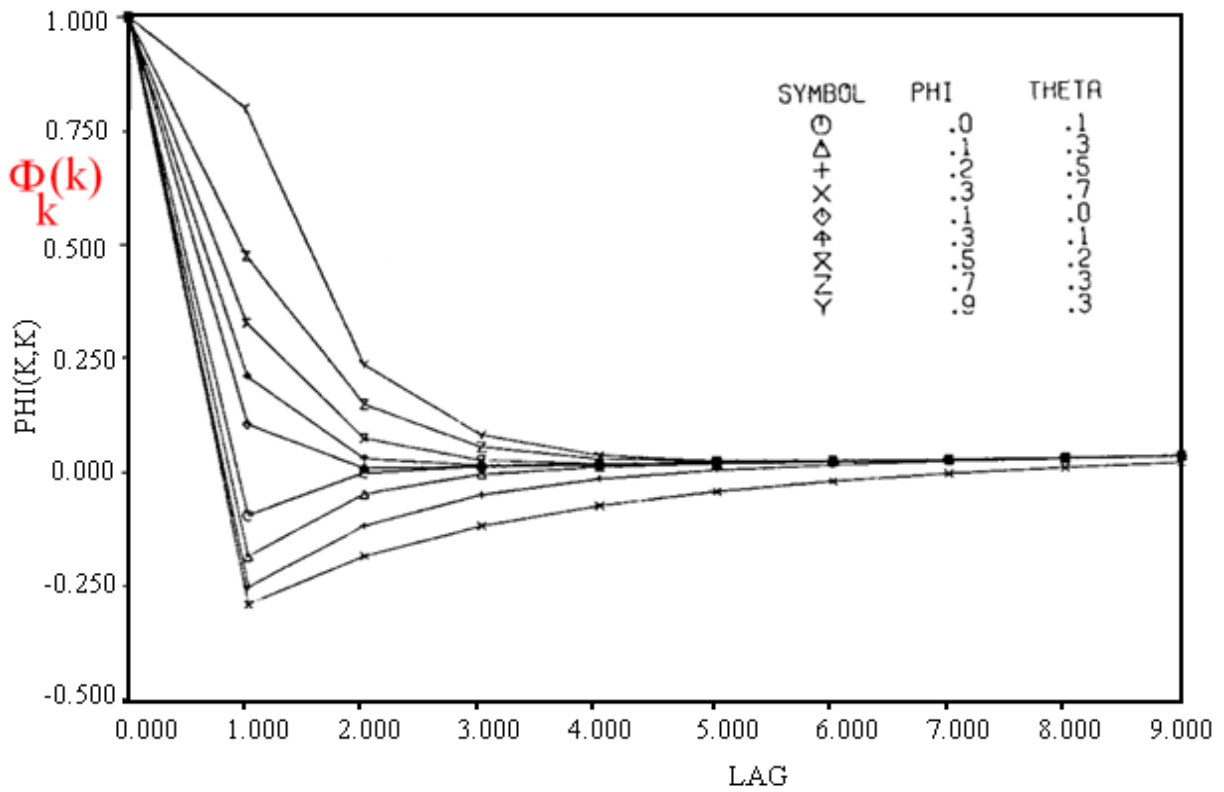


(a) ACF ของ ARMA(1.1) กรณีที่ ϕ_1 และ θ_1 มีค่าเป็นบวก

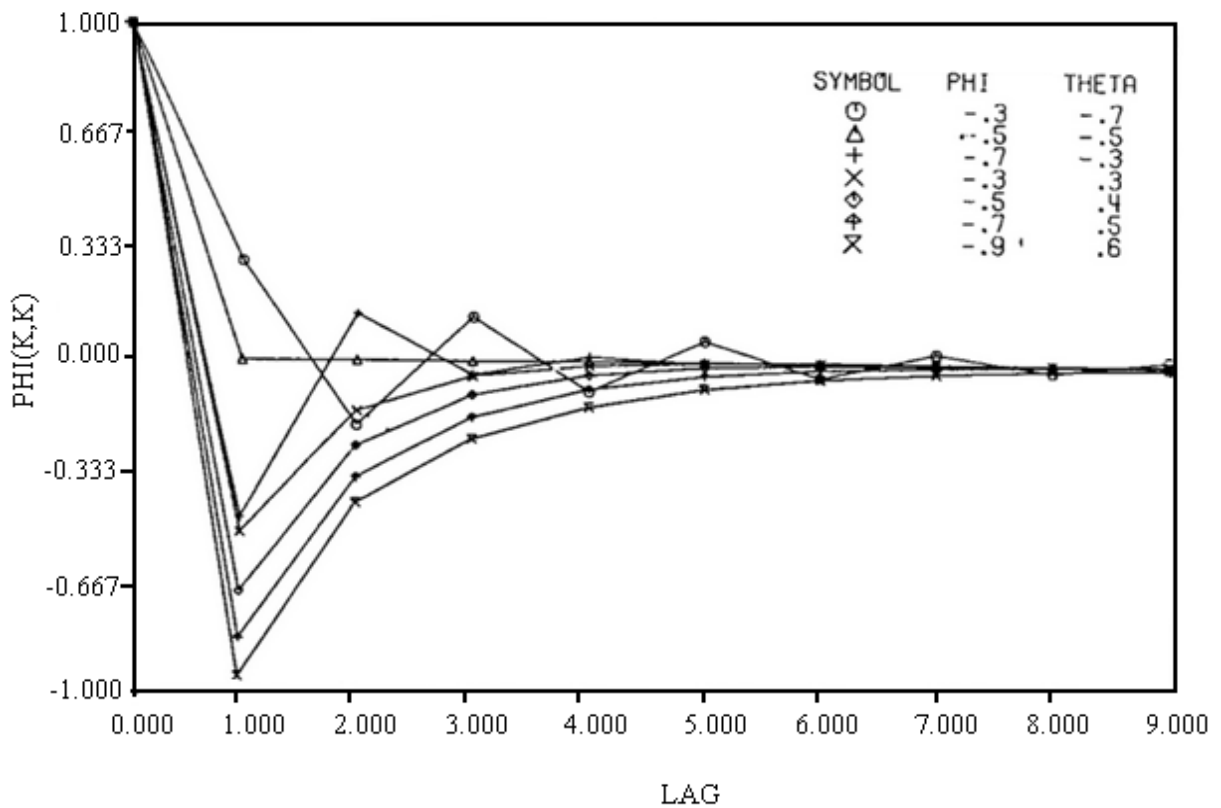


(b) ACF ของ ARMA(1.1) กรณีที่ ϕ_1 และ θ_1 มีค่าเป็นลบ

รูปที่ 5.4 Theoretical Autocorrelation for ARMA(1.1) Model กรณีที่ ϕ_1 และ θ_1 มีค่าต่างๆ กัน



(a) PACF ของ ARMA(1.1) กรณีที่ ϕ_1 และ θ_1 มีค่าเป็นบวก



(b) PACF ของ ARMA(1.1) กรณีที่ ϕ_1 และ θ_1 มีค่าเป็นลบ

รูปที่ 5.5 Theoretical Partial Autocorrelation for ARMA(1, 1) Model กรณีที่ ϕ_1 และ θ_1 มีค่าต่างๆ กัน

5.4 การสร้างแบบจำลอง ARMA สำหรับอนุกรมเวลารายปี

(ARMA Modeling of Annual Time Series)

5.4.1 แบบจำลอง Annual ARMA

สมมติให้ x_t คือ อนุกรมอุทกวิทยารายปี ซึ่งอาจมีการแจกแจงแบบปกติ แต่ถ้า x_t ไม่ได้มีการแจกแจงแบบต้องมีการแปลงให้เป็นอนุกรมเวลาแบบปกติ y_t โดยใช้ฟังก์ชันการแปลงที่เหมาะสม ขั้นตอนต่อไปคือการแปลง y_t เป็น z_t ดังสมการ

$$y_t = \mu + z_t \quad [5.39]$$

ทำการจำลองอนุกรมเวลา z_t ด้วยแบบจำลอง ARMA(p,q) ดังนี้

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad [5.40]$$

คุณสมบัติของแบบจำลอง AR, MA และ ARMA ตามที่กล่าวมาแล้วได้สรุปไว้ในตารางที่ 5.1

ตารางที่ 5.1 Identification Properties for AR, MA and ARMA Processes.

Process	Autocorrelation	Partial Autocorrelations
AR(p)	Infinite in extent, consists of damped exponentials and/or damped waves. Attenuates as $\rho_k = \sum_{j=1}^p \phi_j \rho_{k-j}$	Finite in extent, peaks at lags 1 through p then cuts off.
MA(q)	Finite in extent, peaks at lags 1 through q then cuts off.	Infinite in extent, consists of damped exponentials and/or damped waves.
ARMA(p, q)	Infinite in extent, first q-p lags: irregular then damped exponentials and/or damped waves. Attenuates as $\rho_k = \sum_{j=1}^p \phi_j \rho_{k-j}$ $(k \geq q + 1)$	Infinite in extent, first p-q lags irregular, then damped exponentials and/or damped waves.

5.4.2 ขั้นตอนการจำลอง

- การหารูปแบบแบบจำลอง (Identification)
- การประมาณค่าพารามิเตอร์ (Estimation)
- การตรวจสอบความเหมาะสม (Diagnostic Checking)

(1) การหารูปแบบแบบจำลอง (Identification)

การหารูปแบบแบบจำลอง จะพิจารณาจากกราฟอนุกรมเวลา ACF และ PACF (ดูคุณสมบัติของ ACF และ PACF ของแบบจำลองต่างๆในตารางที่ 5.1)

ACF (r_k vs. k ; $k = 1, \dots, \frac{N}{4}$)

95% Probability Limit คือ $\pm \frac{2}{\sqrt{N}}$

(2) การประมาณค่าพารามิเตอร์ (Estimation)

วิธีการประมาณค่าพารามิเตอร์ของ MA(q)

$$\begin{aligned} \text{จากสมการที่ 5.5} \quad \sigma_{\varepsilon}^2 &= \frac{\gamma_0}{\sum_{j=0}^q \theta_j^2} \\ &= \frac{\gamma_0}{1 + \theta_1^2 + \dots + \theta_q^2} \end{aligned} \quad [5.41]$$

จากสมการที่ 5.4 $\gamma_k = \sigma_{\varepsilon}^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}$; $k \leq q$ ($k = 0, 1, \dots, q$)

ถ้า $k = j$; $j = 0, \dots, q$; $k \leq q$

$$\begin{aligned} \gamma_j &= \sigma_{\varepsilon}^2 \sum_{i=0}^{q-j} \theta_i \theta_{i+j} \\ &= \sigma_{\varepsilon}^2 (\theta_0 \theta_j + \theta_1 \theta_{j+1} + \dots + \theta_{q-j} \theta_q) \\ \theta_j &= - \left(\frac{\gamma_j}{\sigma_{\varepsilon}^2} - \theta_1 \theta_{j+1} - \theta_2 \theta_{j+2} - \dots - \theta_{q-j} \theta_q \right) \end{aligned} \quad [5.42]$$

$c_0 \sim \gamma_0$ (Variance)

$c_j \sim \gamma_j$ (Autocovariance)

หา θ_j ($j = 1, \dots, q$) จากสมการ [5.41] และ [5.42] โดยการคำนวณแบบ Iterative ดังนี้

(1) หา C_0, \dots, C_j จากอนุกรมเวลาตัวอย่าง

(2) สมมติให้ $\hat{\theta}_j = 0$ ($j = 1, \dots, q$)

(3) คำนวณหา $\hat{\sigma}_{\varepsilon}^2$ จากสมการ [5.41]

$$\hat{\sigma}_{\varepsilon}^2 = \frac{C_0}{1 + \hat{\theta}_1^2 + \dots + \hat{\theta}_q^2}$$

(4) คำนวณหา θ_j จากสมการ [5.42]

$$\hat{\theta}_j = - \left(\frac{C_j}{\hat{\sigma}_{\varepsilon}^2} - \hat{\theta}_1 \hat{\theta}_{j+1} - \hat{\theta}_2 \hat{\theta}_{j+2} - \dots - \hat{\theta}_{q-j} \hat{\theta}_q \right)$$

(5) ตรวจสอบว่า $\hat{\theta}_j$ (คำนวณ) = $\hat{\theta}_j$ (สมมติ)

(6) ถ้าผลการคำนวณไม่เท่าให้กลับไปเริ่มคำนวณขั้นที่ (2) สมมติ $\hat{\theta}_j$ ($j = 1, \dots, q$) ใหม่

กรณี MA(1)

$$\left. \begin{aligned} \hat{\sigma}_{\varepsilon}^2 &= \frac{c_0}{1 + \hat{\theta}_1^2} \\ \hat{\theta}_1 &= - \frac{c_1}{\sigma_{\varepsilon}^2} \end{aligned} \right\} \text{ 2 สมการ 2 Unknown จะสามารถคำนวณหาค่า } \hat{\theta}_1, \sigma_{\varepsilon}^2 \text{ ได้}$$

กรณี MA(2)

$$\left. \begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{c_0}{1+\hat{\theta}_1^2+\hat{\theta}_2^2} \\ \hat{\theta}_2 &= -\frac{c_2}{\hat{\sigma}_\varepsilon^2} \\ \hat{\theta}_1 &= -\left(\frac{c_2}{\hat{\sigma}_\varepsilon^2} - \hat{\theta}_1\hat{\theta}_2\right) \end{aligned} \right\} \begin{array}{l} \text{จะสามารถคำนวณหาค่าพารามิเตอร์} \\ \hat{\sigma}_\varepsilon^2, \hat{\theta}_1, \hat{\theta}_2 \text{ ได้} \end{array}$$

กรณี ARMA(p, q)

การหาพารามิเตอร์ $\phi_j (j = 1, \dots, p)$ จากสมการ 5.19

กรณี $k > q+1$

$$Y_k = \sum_{j=1}^p \phi_j Y_{k-j}; \quad k \geq q+1$$

ประมาณค่าด้วย Sample Lag k Autocovariance

$$c_k = \sum_{j=1}^p \hat{\phi}_j c_{k-j}; \quad k \geq q+1 \quad [5.43]$$

สำหรับ ARMA(1, 1)

$$p=1, q=1, k \geq 2$$

$$k=2; c_2 = \hat{\phi}_1 c_1 \text{ จะได้ว่า } \hat{\phi}_1 = \frac{c_2}{c_1}$$

สำหรับ ARMA(2, 2)

$$p=2, q=2, k \geq 3$$

$$\left. \begin{aligned} k=3; c_3 &= \hat{\phi}_1 c_2 + \hat{\phi}_2 c_1 \\ k=4; c_4 &= \hat{\phi}_1 c_3 + \hat{\phi}_2 c_2 \end{aligned} \right\} \text{จะสามารถหาค่าพารามิเตอร์ } \hat{\phi}_1 \text{ และ } \hat{\phi}_2 \text{ ได้}$$

เมื่อทราบค่า $\hat{\phi}_j$ จะสามารถแปลง ARMA(p,q) เป็น MA(q) ได้ดังนี้

$$z'_t = z_t - \sum_{j=1}^p \hat{\phi}_j z_{t-j} \quad [5.44]$$

คำนวณหา $c'_j (j = 0, \dots, q)$ จากอนุกรมเวลา z'_t

หลังจากนั้นก็จะสามารถหา $\hat{\theta}_j$ ได้จากสมการ [5.41] และ [5.42] โดยวิธี Iterative ตามที่กล่าวมาแล้ว

ขั้นตอนการคำนวณแบบ Iterative

สมมติค่า $\hat{\theta}_j$ แล้วคำนวณหา

$$\hat{\sigma}_\varepsilon^2 = \frac{c_0}{\sum_{j=0}^q \hat{\theta}_j^2} \text{ และ}$$

$$\hat{\theta}_j = -\left(\frac{c_j}{\hat{\sigma}_\varepsilon^2} - \hat{\theta}_1 \hat{\theta}_{j+1} - \hat{\theta}_2 \hat{\theta}_{j+2} - \dots - \hat{\theta}_{q-j} \hat{\theta}_q\right)$$

ถ้า $\hat{\theta}_j$ (คำนวณ) = $\hat{\theta}_j$ (สมมติ) แสดงว่าค่าที่คำนวณได้ใช้ได้ จึงหยุดการคำนวณ

Efficient Estimate

วิธีการประมาณค่าที่ Efficient คือ Max. Likelihood หรือ LSE ถ้า ε_t มีการแจกแจงแบบปกติ

$$\varepsilon_t = z_t - \sum_{j=1}^p \hat{\phi}_j z_{t-j} + \sum_{j=1}^q \hat{\theta}_j \varepsilon_{t-1}$$

กำหนดให้ $SSE = S(\underline{\phi}, \underline{\theta}) = \sum_{t=1}^N \varepsilon_t^2$ [5.45]

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\underline{\phi}, \underline{\theta})}{N} = \text{Variance ของ } \varepsilon_t \quad [5.46]$$

สามารถใช้หลัก Optimization หาค่าพารามิเตอร์ได้ โดยการกำหนด Objective Function และ Constraints ดังนี้

$$\left. \begin{array}{l} \text{Min. } S(\underline{\phi}, \underline{\theta}) \\ \text{ST } -1 < \underline{\phi} < 1 \\ \quad -1 < \underline{\theta} < 1 \end{array} \right\} \text{ ใช้ Solver ใน Excel}$$

Excel มีเครื่องในการหาค่าตอบของ Optimization Model โดยการเรียกใช้ Solver

สำหรับ ARMA(1, 1): $z_t = \phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$

$$\varepsilon_t = z_t - \phi_1 z_{t-1} + \theta_1 \varepsilon_{t-1}$$

ถ้ากำหนดให้ $z_0 = 0$ และ $\varepsilon_0 = 0$ หรือ ($z_0 = \bar{z}, \varepsilon_0 = \bar{\varepsilon}$) จะได้

$$\varepsilon_1 = z_1$$

$$\varepsilon_2 = z_2 - \phi_1 z_1 + \theta_1 \varepsilon_1$$

⋮

$$\varepsilon_N = z_N - \phi_1 z_{N-1} + \theta_1 \varepsilon_{N-1}$$

จะสามารถหาค่า $SSE = S(\underline{\phi}, \underline{\theta}) = \sum_{t=1}^N \varepsilon_t^2$ เพื่อนำไปหาพารามิเตอร์ได้ โดยใช้ Solver ใน Excel

(3) ความน่าเชื่อถือของค่าประมาณพารามิเตอร์ (Reliability of Estimated Parameters)

Variance-Covariance Matrix ของพารามิเตอร์แสดงอยู่ในสมการ

$$v(\hat{\beta}) = \begin{bmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1 \hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_1 \hat{\beta}_{p+q}) \\ \text{Cov}(\hat{\beta}_2 \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_2 \hat{\beta}_{p+q}) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\hat{\beta}_{p+q} \hat{\beta}_1) & \text{Cov}(\hat{\beta}_{p+q} \hat{\beta}_2) & \dots & \text{Var}(\hat{\beta}_{p+q}) \end{bmatrix} \quad [5.47]$$

Box and Jenkins (1976) เสนอแนะวิธีการหา Variance-Covariance Matrix ดังสมการ

$$v(\hat{\beta}) \approx 2\hat{\sigma}_\varepsilon^2 \begin{bmatrix} \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_1^2} & \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_1 \partial \hat{\beta}_2} & \dots & \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_1 \partial \hat{\beta}_{p+q}} \\ \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_2 \partial \hat{\beta}_1} & \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_2^2} & \dots & \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_2 \partial \hat{\beta}_{p+q}} \\ \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_{p+q} \partial \hat{\beta}_1} & \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_{p+q} \partial \hat{\beta}_2} & \dots & \frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta}_{p+q}^2} \end{bmatrix}^{-1} \quad [5.48]$$

สำหรับแบบจำลอง ARMA(1, 1) Variance-Covariance Matrix จะอยู่ในรูปของสมการ

$$v(\hat{\beta}) = \begin{bmatrix} Var(\hat{\phi}_1) & Cov(\hat{\phi}_1 \hat{\theta}_1) \\ Cov(\hat{\phi}_1 \hat{\theta}_1) & Var(\hat{\theta}_1) \end{bmatrix} \cong 2\hat{\sigma}_\varepsilon^2 \begin{bmatrix} \frac{\partial^2 S(\hat{\phi}_1 \hat{\theta}_1)}{\partial \hat{\phi}_1^2} & \frac{\partial^2 S(\hat{\phi}_1 \hat{\theta}_1)}{\partial \hat{\phi}_1 \partial \hat{\theta}_1} \\ \frac{\partial^2 S(\hat{\phi}_1 \hat{\theta}_1)}{\partial \hat{\phi}_1 \partial \hat{\theta}_1} & \frac{\partial^2 S(\hat{\phi}_1 \hat{\theta}_1)}{\partial \hat{\theta}_1^2} \end{bmatrix}^{-1} \quad [5.49]$$

รูปแบบ Variance-Covariance Matrix สำหรับ ARMA(1,1) ของ Box and Jenkins(1976)

$$v(\hat{\phi}_1, \hat{\theta}_1) = \frac{1}{N} \frac{(1-\hat{\phi}_1 \hat{\theta}_1)}{(\hat{\phi}_1 - \hat{\theta}_1)^2} \begin{bmatrix} (1-\hat{\phi}_1^2)(1-\hat{\phi}_1 \hat{\theta}_1) & (1-\hat{\phi}_1^2)(1-\hat{\theta}_1^2) \\ (1-\hat{\phi}_1^2)(1-\hat{\theta}_1^2) & (1-\hat{\phi}_1 \hat{\theta}_1) \end{bmatrix} \quad [5.50]$$

และค่า Standard Errors คือ

$$\hat{\sigma}(\hat{\phi}_1) = \sqrt{\frac{1}{N} \frac{(1-\hat{\phi}_1 \hat{\theta}_1)^2 (1-\hat{\phi}_1^2)}{(\hat{\phi}_1 - \hat{\theta}_1)^2}} \quad [5.51]$$

และ
$$\hat{\sigma}(\hat{\theta}_1) = \sqrt{\frac{1}{N} \frac{(1-\hat{\phi}_1 \hat{\theta}_1)^2 (1-\hat{\theta}_1^2)}{(\hat{\phi}_1 - \hat{\theta}_1)^2}} \quad [5.52]$$

$$\left[\hat{\beta}_i - u_{\alpha/2} \hat{\sigma}(\hat{\beta}_i), \hat{\beta}_i + u_{\alpha/2} \hat{\sigma}(\hat{\beta}_i) \right] \quad [5.53]$$

Standard Error ของ $\hat{\sigma}_\varepsilon^2$

$$S(\beta) = S(\hat{\beta}) \left\{ 1 + \frac{\chi_{\alpha^2(p+q)}}{N} \right\} \quad [5.54]$$

(4) การประมาณค่าพารามิเตอร์แบบ Non-Linear

กำหนดให้ $S(\phi_i, \theta_i) = \sum_{t=1}^N e_t^2 =$ Sum of Square of Residual

จงหาพารามิเตอร์โดยวิธี Least Square

$$\frac{\partial S}{\partial \phi_i}(\phi_i, \theta_i) = 0 \quad \dots\dots\dots(\text{Linear}) \quad [5.55]$$

$$\frac{\partial S}{\partial \theta_i}(\phi_i, \theta_i) = 0 \quad \dots\dots\dots(\text{Non-Linear}) \quad [5.56]$$

ใช้ Linear Approximation of Taylor Series Expansion of ε_t คุรายละเอียดใน Box and Jenkins (1976)

5.4.3 การทดสอบความเหมาะสมของแบบจำลอง Annual ARMA(p, q)

$$z_t = \phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{5.57}$$

$$z_{t-1} = \phi_1 z_{t-2} + \varepsilon_{t+1} - \theta_1 \varepsilon_{t-2} \tag{5.58}$$

สมการที่ [5.58]-[5.57]

$$\begin{aligned} z_t - z_{t-1} &= \phi_1 z_{t-1} - \phi_1 z_{t-2} + \varepsilon_t - (1 + \theta_1) \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} \\ z_t &= (1 + \phi_1) z_{t-1} - \phi_1 z_{t-2} + \varepsilon_t - (1 + \theta_1) \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} \end{aligned} \tag{5.59}$$

สมการที่ 5.59 คือ ARMA(2, 2) ซึ่งไม่เป็นอิสระจากสมการ [5.57]

Corrected Residual Variance

$$(\hat{\sigma}_\varepsilon^2)_c = \frac{1}{N} S(\hat{\phi}, \hat{\theta}) \tag{5.60}$$

N = No. of Observations

n = No. of Parameters

(1) การทดสอบความเป็นอิสระของ Residuals ε_t โดยวิธี Porte Manteau Lack of Fit Test

$$Q = N \sum_{k=1}^L \rho_k^2(\hat{\varepsilon})$$

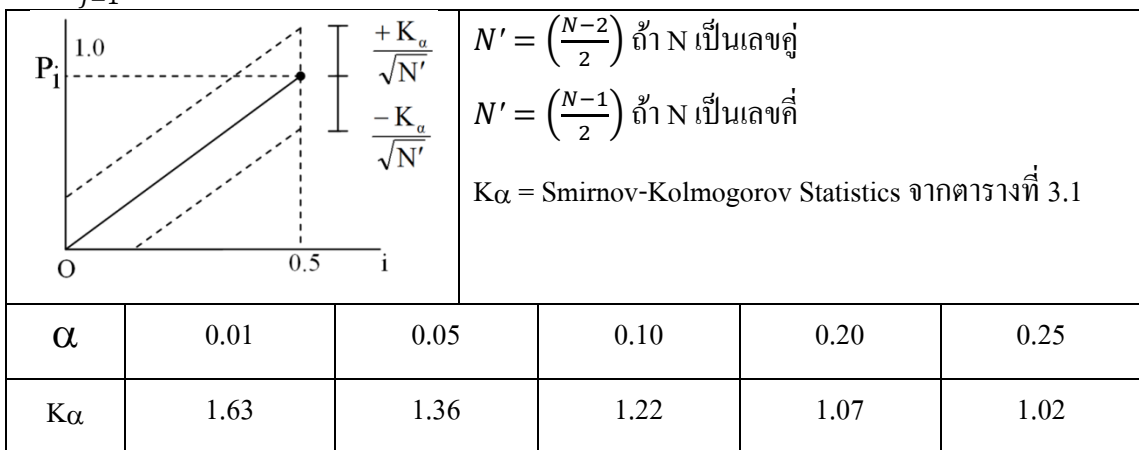
ถ้า $Q < \chi_{\alpha, (L-p-q)}^2$ แสดงว่า ε_t เป็นอนุกรมเวลาแบบ Independent

(2) การทดสอบความเป็นอิสระของ Residuals ε_t โดยวิธี Cumulative Periodogram และ Smirnov-

Kolmogorov Statistics

$$MSD(h_j) = \frac{2}{N^2} \left[\left(\sum_{t=1}^N \varepsilon_t \cos(2\pi h_j t) \right)^2 + \left(\sum_{t=1}^N \varepsilon_t \sin(2\pi h_j t) \right)^2 \right]$$

$$P_i = \sum_{j=1}^i \frac{MSD(h_j)}{\sigma_\varepsilon^2}; i = 1, \dots, \frac{N}{2}$$



(3) การทดสอบ Parsimony ของพารามิเตอร์

$$AIC(p, q) = N \ln(\text{MLE of } \hat{\sigma}_\varepsilon^2) + 2(p + q) \tag{5.61}$$

5.5 การสังเคราะห์ข้อมูล (Generation)

$$\hat{z}_t = \sum_{j=1}^p \hat{\phi}_j z_{t-j} + \varepsilon_t - \sum_{j=1}^q \hat{\theta}_j \varepsilon_{t-j} \tag{5.62}$$

จะต้องสังเคราะห์อนุกรมเวลาจำนวน $N_w + N_g$ เมื่อ $N_w = 50 \sim 100$

5.6 การพยากรณ์ด้วยแบบจำลอง Annual ARMA

5.6.1 หลักการพยากรณ์ทางสโตแคสติก

แนวคิดในการพยากรณ์ด้วยแบบจำลอง Annual ARMA แสดงอยู่ในรูปที่ 5.6

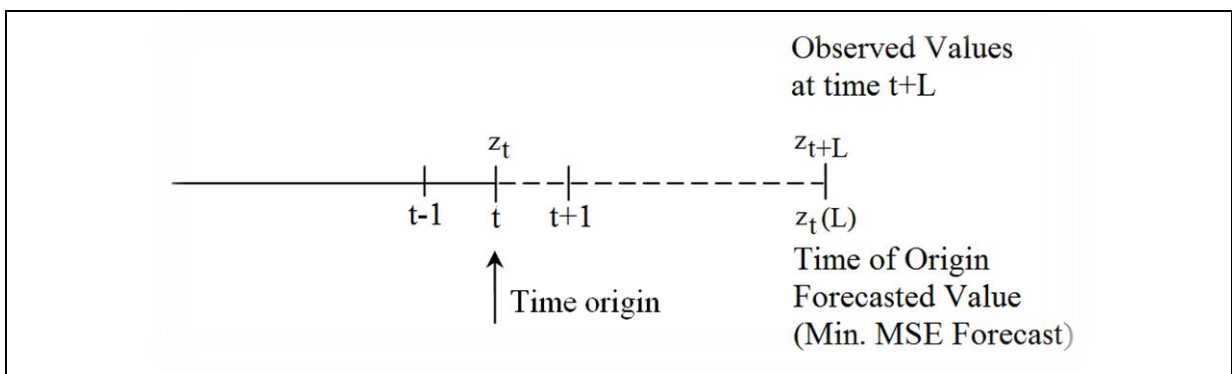
เมื่อ t = Time Origin of the Forecast

L = Lead Time ≥ 1

z_{t+L} = Value to be forecasted observation at time $t+L$

$z_t(L) = E(z_{t+L} | z_t, z_{t-1}, \dots) =$ Conditional Expectation

$z_t(L) = [z_t + L] =$ Forecasted Function = Min MSE Forecast



รูปที่ 5.6 แนวคิดในการพยากรณ์

แบบจำลอง ARMA(1,1) สามารถเขียนในรูปของ Infinite Weight of Sum of ϵ_t

$$z_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \quad \text{เมื่อ } \psi_j = \phi_1^{j-1} (\phi_1 - \theta_1) \tag{5.63}$$

$$z_{t+L} = \sum_{j=0}^{L-1} \psi_j \epsilon_{t+L-j} + \sum_{j=L}^{\infty} \psi_j \epsilon_{t+L-j} \tag{5.64}$$

เมื่อ

$\sum_{j=0}^{L-1} \psi_j \epsilon_{t+L-j}$ คือ Future Values ของ ϵ_{t+L-j} ยังไม่ทราบค่าที่เวลา t

ซึ่งเรียกว่า Forecasted Error หรือ $e_t(L)$

$\sum_{j=L}^{\infty} \psi_j \epsilon_{t+L-j}$ คือ Past and Present Values ของ ϵ_{t+L-j} ทราบค่าที่เวลา t

ซึ่งเรียกว่า Forecasted Function หรือ $z_t(L)$

$$e_t(L) = z_{t+L} - z_t(L) \tag{5.65}$$

= Forecast Error

5.6.2 คุณสมบัติของ $E(\epsilon_{t+j})$

$$E(\epsilon_{t+j}) = 0 ; j = 1, \dots, L$$

$$E(e_t(L)) = 0$$

$$z_{t+L} = \sum_{j=1}^P \psi_j z_{t+L-j} + \varepsilon_{t+L} - \sum_{j=1}^q \theta_j \varepsilon_{t+L-j}$$

Taking Conditional Expectation

$$[z_{t+L}] = \sum_{j=1}^P \phi_j [z_{t+L-j}] + [\varepsilon_{t+L}] - \sum_{j=1}^q \theta_j [\varepsilon_{t+L-j}] \quad [5.66]$$

$$\text{ถ้า } L-j \leq 0$$

$$[z_{t+L-j}] = z_{t+L-j} \quad [5.67]$$

$$[\varepsilon_{t+L-j}] = \varepsilon_{t+L-j} \quad [5.68]$$

$$\text{ถ้า } L-j > 0$$

$$[z_{t+L+j}] = z_t(L+j) \quad [5.69]$$

$$[\varepsilon_{t+L-j}] = 0 = E(e_t(L)) \quad [5.70]$$

5.6.3 One step Ahead Forecasting (L = 1)

$$[z_{t+1}] = z_t(1) = \sum_{j=1}^P \psi_j z_{t+1-j} + [\varepsilon_{t+1}] - \sum_{j=1}^q \theta_j \varepsilon_{t+1-j} \quad [5.71]$$

$$z_{t+1} = \sum_{j=1}^P \psi_j z_{t+1-j} + \varepsilon_{t+1} - \sum_{j=1}^q \theta_j \varepsilon_{t+1-j} \quad [5.72]$$

สมการ [5.72]-[5.71]

$$z_{t+1} - z_t(1) = \varepsilon_{t+1} = e_t(1) \quad [5.73]$$

5.6.4 Forecasted Function

ARMA(1, 1)

$$z_t(1) = \phi_1 z_t - \theta_1 \varepsilon_t$$

$$z_t(2) = \phi_1 z_t(1)$$

$$z_t(3) = \phi_1 z_t(2)$$

ARMA(2, 2)

$$z_t(1) = \phi_1 z_t + \phi_2 z_{t-1} - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1}$$

$$z_t(2) = \phi_1 z_t(1) + \phi_2 z_t - \theta_2 \varepsilon_t$$

$$z_t(3) = \phi_1 z_t(2) + \phi_2 z_t(1)$$

$$z_t(4) = \phi_1 z_t(3) + \phi_2 z_t(2)$$

AR(1)

$$z_t(1) = \phi_1 z_t + [\varepsilon_{t+1}]$$

$$z_t(2) = \phi_1 z_t(1)$$

$$z_t(3) = \phi_1 z_t(2)$$

AR(2)

$$z_t(1) = \phi_1 z_t + \phi_2 z_{t-1} [\varepsilon_{t+1}]$$

$$z_t(2) = \phi_1 z_t(1) + \phi_2 z_t$$

$$z_t(3) = \phi_1 z_t(2) + \phi_2 z_t(1)$$

5.6.5 Forecast Error

$$e_t(L) = \sum_{j=0}^{L-1} \psi_j \varepsilon_{t+L-j} \quad [5.74]$$

สำหรับ ARMA(1, 1)

$$\psi_j = (\phi_1 - \theta_1) \phi_1^{j-1} \quad [5.75]$$

Variance of Forecast Error

$$\begin{aligned} \text{var}(e_t(L)) &= E(e_t(L))^2 \\ &= E\left[\left(\sum_{j=0}^{L-1} \psi_j \varepsilon_{t+L-j}\right)^2\right] \\ &= E\left[\sum_{j=0}^{L-1} \psi_j^2 \varepsilon_{t+L-j}^2\right] \\ &= \sum_{j=0}^{L-1} \psi_j^2 E(\varepsilon_{t+L-j}^2) \\ &= \sum_{j=0}^{L-1} \psi_j^2 \sigma_\varepsilon^2 \end{aligned} \quad [5.76]$$

(1- α) Confident Limits of z_{t+L}

$$\begin{aligned} z_{t+L} &= z_t(L) \pm u_{\frac{\alpha}{2}} \hat{\sigma}_\varepsilon \sqrt{\sum_{j=0}^{L-1} \psi_j^2} \\ &= z_t(L) \pm u_{\frac{\alpha}{2}} \hat{\sigma}_\varepsilon \sqrt{1 + \sum_{j=1}^{L-1} \psi_j^2} \end{aligned} \quad [5.77]$$

5.7 การจำลองอนุกรมเวลารายปีด้วยแบบจำลอง ARMA(p,q)

ขั้นตอนการจำลอง

(1) กำหนดค่าสถิติ \bar{z}, s^2

(2) กำหนดค่า Lag k Autocovariance และ Lag k Autocorrelation และพล็อตกราฟ

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}); 0 \leq k \leq \frac{N}{4}$$

$$r_k = \frac{c_k}{c_0} = \hat{\rho}_k$$

(3) กำหนดหาค่า Partial Autoregressive Function จากสมการ

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdot & \rho_{k-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdot & 1 \end{bmatrix} \begin{bmatrix} \phi_1(k) \\ \phi_2(k) \\ \phi_3(k) \\ \cdot \\ \phi_k(k) \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \rho_k \end{bmatrix}$$

(4) หาค่าลำดับของ ARMA(p, q) จาก ACF และ PACF

(5) หาค่าประมาณเบื้องต้น (Initial Estimate) $\{\hat{\phi}_1, \dots, \hat{\phi}_p\}$ จากสมการ Yule Walker

$$c_k = \sum_{j=1}^p \hat{\phi}_j c_{k-j}; k \geq q + 1$$

กรณีที่ $\bar{z} \neq 0$

$$z_t = \sum_{j=1}^p \phi_j z_{t-j} - \sum_{j=0}^q \theta_j \varepsilon_{t-j} + \theta_{00}$$

$$\text{เมื่อ } \theta_{00} = \bar{z}(1 - \sum_{j=1}^p \phi_j) \quad [5.78]$$

(6) หาค่าประมาณเบื้องต้นของ $\theta_1, \dots, \theta_q$ จาก z'_t เมื่อ

$$z'_t = z_t - \sum_{j=1}^p \phi_j z_{t-j} \quad [5.79]$$

- กำหนดหาค่า c'_k (Lag k Autocovariance ของ z'_t) โดยวิธีของ Box and Jenkins (1976, p202)

$$c_j^{\square} = \sum_{i=1}^p \hat{\phi}_i^2 c_j + \sum_{i=1}^p (\hat{\phi}_0 \hat{\phi}_i + \hat{\phi}_1 \hat{\phi}_{i+1} + \dots + \hat{\phi}_{p-1} \hat{\phi}_p) d_j \quad [5.80]$$

$$d_j = c_{j+i} + c_{j-i}; j = 0, 1, \dots, q \quad [5.81]$$

$$\phi_0 = -1 \quad [5.82]$$

$$\hat{\sigma}_{\varepsilon}^2 = \frac{c_0^{\square}}{1 + \theta_1^2 + \dots + \theta_q^2} \quad [5.83]$$

$$\hat{\theta}_j = - \left(\frac{c_j^{\square}}{\hat{\sigma}_{\varepsilon}^2} - \hat{\theta}_1 \hat{\theta}_{j+1} - \hat{\theta}_2 \hat{\theta}_{j+2} - \dots - \hat{\theta}_{q-1} \hat{\theta}_q \right) \quad [5.84]$$

(7) Maximum Likelihood Estimate (MLE)

$$\varepsilon_j = 0, j = 1, \dots, \text{Max}(p, q)$$

ถ้า $p > q$

$$\varepsilon_{p+j} = z_{p+j} - \theta_{00} - \sum_{i=1}^p \hat{\phi}_i z_{p+j-i} + \sum_{i=1}^q \hat{\theta}_i \varepsilon_{p+j-i} \quad [5.85]$$

เมื่อ $j = 1, 2, \dots, N - p$

$$S(\underline{\phi}, \underline{\theta}) = \sum_{t=1}^N \varepsilon_t^2 = S \quad [5.86]$$

จงหาค่า $\underline{\phi}, \underline{\theta}$ ที่ทำให้ $S(\underline{\phi}, \underline{\theta})$ มีค่าต่ำสุด (Minimum)

$$\sigma_{\varepsilon}^2 = \frac{S}{N}$$

$\hat{\sigma}(\hat{\beta})$ = Diagonal Terms ของ Variance-Covariance Matrix

$$\hat{\beta}_i - u_{\frac{\alpha}{2}} \hat{\sigma}(\hat{\beta}_i) < \beta_i < \hat{\beta}_i + u_{\frac{\alpha}{2}} \hat{\sigma}(\hat{\beta}_i) \quad [5.87]$$

(8) Porte Manteau Lack of Fit Test

$$Q = N \sum_{k=1}^L [r_k(\varepsilon)]^2 ; \frac{N}{10} < L < \frac{N}{4} \quad [5.88]$$

ถ้า $Q < \chi_{\alpha, L-p-q}^2$; แสดงว่า ε_t เป็น Independent

(9) Akaike Test

$$AIC(p, q) = N \ln(\sigma_\varepsilon^2) + 2(p + q) \quad [5.89]$$

$\sigma_\varepsilon^2 = MLE$ ของ Error Variance

(10) Generation

$$\hat{z}_t = \sum_{j=1}^p \hat{\phi}_j z_{t-j} - \sum_{j=0}^q \hat{\theta}_j \varepsilon_{t-j} + \hat{\theta}_{00} \quad [5.90]$$

(11) Forecasting

ถ้า $L \leq q$

$$z_t(1) = \phi_1 z_t + \dots + \phi_p z_{t-p+1} - \theta_1 \varepsilon_t - \dots - \theta_q \varepsilon_{t-q} \quad [5.91]$$

$$z_t(2) = \phi_1 z_t(1) + \phi_2 z_t + \dots + \phi_p z_{t-p+2} - \theta_2 \varepsilon_t - \dots - \theta_q \varepsilon_{t-q+2} \quad [5.92]$$

$$z_t(L) = \phi_1 z_t(L-1) + \dots + \phi_{L-1} z_t(1) + \phi_L z_t + \dots + \phi_p z_{t-p+L} - \theta_L \varepsilon_t - \dots - \theta_q \varepsilon_{t-q+L} \quad [5.93]$$

ถ้า $L > q$

$$z_t(L) = \phi_1 z_t(L-1) + \dots + \phi_p z_t(L-p) \quad [5.94]$$

จงหา Ψ Weight จากสมการ [5.75]

$$\psi_j = (\phi_1 - \theta_1) \phi_1^{j-1}$$

$$Var[e_t(L)] = \sum_{j=0}^{L-1} \psi_j^2 \sigma_\varepsilon^2$$

$$z_{t+L}(\pm) = z_t(L) \pm u_{\frac{\alpha}{2}} \{1 + \sum_{j=0}^{L-1} \psi_j^2\} \hat{\sigma}_\varepsilon \quad [5.95]$$

5.8 ตัวอย่างการจำลองอนุกรมเวลาปริมาณน้ำท่ารายปีด้วยแบบจำลอง ARMA(p,q)

อนุกรมเวลาน้ำท่ารายปีของแม่น้ำ Niger ใน Koulicoro ออฟริกา มีข้อมูลระหว่าง 1906-1957 จำนวน 52 ปี ในการจำลองจะแปลงข้อมูลเป็นค่า Modular Coefficients และอนุกรมเวลา z_t ตารางที่ 5.2 โดยใช้ดังสมการ

$$u_t = \frac{\text{Annual Flow}}{\text{Mean Flow}} \quad [5.96]$$

$$z_t = \frac{u_t - \bar{u}}{s_u} \quad [5.97]$$

$$E(u_t) = 1$$

$$E(z_t) = 0$$

ขั้นที่ 1 $\bar{u} = 1$; $\bar{z} = -0.0001 \approx 0$

$$s_u^2 = 0.0575$$
 ; $s_z^2 = 0.9804 \approx 1$

ขั้นที่ 2 จงหาค่า $c_k, r_k, \hat{\phi}_k(k)$ ของ z_t ; $k = 1, \dots, 29$

$$\hat{\phi}_1(1) = r_1 = 0.53478$$

$$\hat{\phi}_2(2) = \frac{r_2 - r_1^2}{1 - r_1^2} = \frac{0.46292 - 0.53478^2}{1 - 0.53478^2} = 0.2478$$

$$\hat{\phi}_1(2) = \frac{r_1 - (1 - r_2)}{1 - r_1^2} = \frac{0.53478(1 - 0.46292)}{1 - 0.53478^2}$$

$$= 0.40226$$

$$\hat{\phi}_3(3) = \frac{r_3 - \hat{\phi}_1(2)r_2 - \hat{\phi}_2(2)r_1}{1 - \hat{\phi}_1(2)r_1 - \hat{\phi}_2(2)r_2} = 0.12096$$

ขั้นที่ 3 จากกราฟ r_k ในตารางที่ 5.2 และรูปที่ 5.7 จะสรุปว่าแบบจำลองที่เหมาะสมคือ AR(2) หรือ ARMA(1,1)

จากกราฟ $\hat{\phi}_k(k)$ ในตารางที่ 5.3 และรูปที่ 5.8 จะสรุปว่าแบบจำลองที่เหมาะสมคือ AR(1)

ทดลองจำลองโดยใช้ ARMA(1,1)

ขั้นที่ 4 ค่าประมาณเบื้องต้นของ (Initial Estimate) ของ $\hat{\phi}_1$ สำหรับ ARMA(1, 1)

$$c_2 = \hat{\phi}_1 c_1$$

$$\hat{\phi}_1 = \frac{c_2}{c_1} = \frac{0.45385}{0.52429} = 0.8656$$

ขั้นที่ 5 ค่าประมาณเบื้องต้นของ (Initial Estimate) ของ $\hat{\theta}_1$

$$z'_t = z_t - \hat{\phi}_1 z_{t-1}$$

$$c'_0 = (\hat{\phi}_1^2 c_0 + \hat{\phi}_1^2 c_0) + (\hat{\phi}_0 \hat{\phi}_1) d_0$$

$$c'_1 = (\hat{\phi}_0^2 c_1 + \hat{\phi}_1^2 c_1) + (\hat{\phi}_0 \hat{\phi}_1) d_1$$

ตารางที่ 5.2 Niger River at Koulicoro, Africa, Modular Coefficients (Annual Flow/Mean), Annual

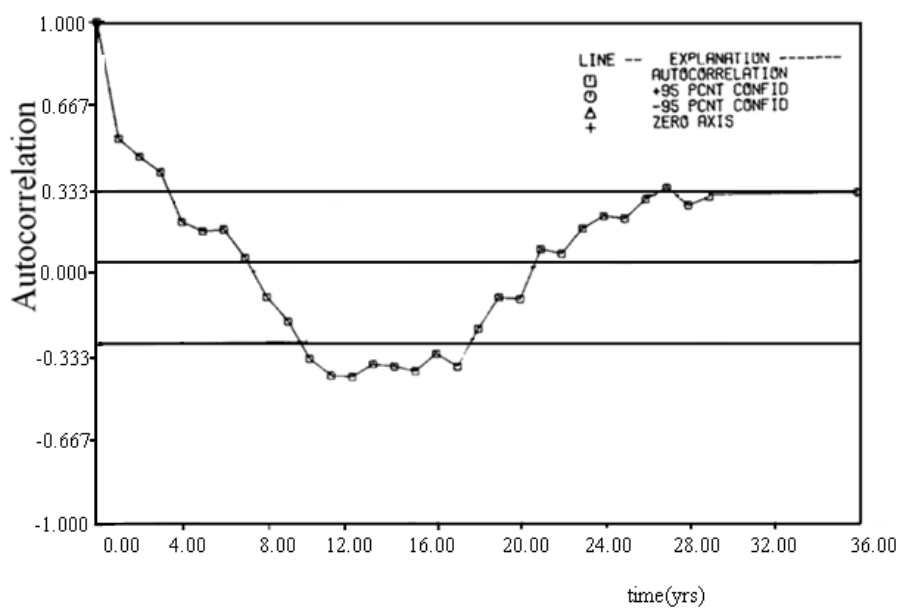
Flows (CFS) and Standardized Flows. 1906-1957 (after Yevjevich, 1963)

	MODULAR COEFFICIENTS (u_t)	ANNUAL FLOWS(x_t)	STANDARDIZED FLOWS(z_t)
1	0.73200	39792.98	-1.106432
2	0.78900	42891.62	-0.871109
3	1.27000	69039.74	1.114689
4	0.80300	43652.69	-0.81331
5	1.04300	56699.57	0.177524
6	0.84600	45990.25	-0.635785
7	0.52900	28757.50	-1.944512
8	0.60500	32889.01	-1.630748
9	0.90000	48925.80	-0.412848
10	0.89200	48490.90	-0.445875
11	0.95700	52024.43	-0.177524
12	1.03500	56264.67	0.144497
13	0.90200	49034.52	-0.404591
14	0.80300	43652.69	-0.81331
15	0.67700	36803.07	-1.333498
16	0.96100	52241.88	-0.161011
17	1.00800	54796.90	0.033028
18	1.42700	77574.57	1.762859
19	1.52900	83118.50	2.183964
20	1.09100	59308.94	0.375691
21	1.27600	69365.91	1.139459
22	1.40100	76161.16	1.655519
23	1.34100	72899.44	1.40781
24	1.31100	71268.58	1.283956
25	1.13200	61537.78	0.544959
26	1.15100	62570.66	0.623400
27	1.05700	57460.63	0.235323
28	0.94900	51589.54	-0.210552
29	0.93600	50882.83	-0.264222
30	1.12000	60885.44	0.495417
31	0.83200	45229.18	-0.693587
32	0.95600	51970.07	-0.181653
33	0.88300	48001.65	-0.483032
34	0.76100	41369.48	-0.986706
35	0.80400	43707.05	-0.809181
36	0.64500	35063.49	-1.465609
37	0.77200	41967.46	-0.941293
38	0.64600	35117.85	-1.461481
39	0.80200	43598.32	-0.817438
40	0.99600	54144.55	-0.016514
41	0.81700	44413.75	-0.755511
42	1.08500	58982.77	0.35092
43	0.89900	48871.44	-0.416976
44	0.98000	53274.76	-0.08257
45	1.38500	75291.37	1.589463
46	1.06000	57623.72	0.247709
47	1.28000	69583.36	1.155973
48	1.35200	73497.42	1.453224
49	1.33300	72464.55	1.374783
50	0.88300	48001.65	-0.483032
51	1.35500	73660.51	1.465609
Mean	1.0000	54362	-0.0001
Std.Dev.	0.2398	13167.569	0.9901

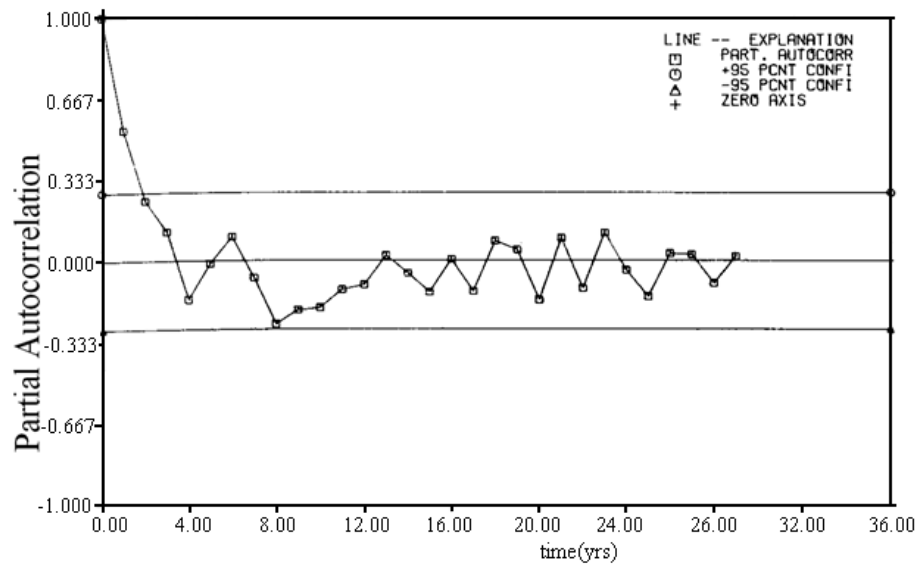
ตารางที่ 5.3 Auto-Covariance, Autocorrelation and Partial Autocorrelation of Standardized

Annual Flows, Niger River

LAG(k)	AUTOCOVARIANCE (c_k)	AUTOCORRELATION (r_k)	PARTIAL AUTOCORRELATION ($\hat{\rho}_k(k)$)
0	0.98039	1.000000	1.000000
1	0.52429	0.53478	0.53478
2	0.45385	0.46292	0.2478
3	0.39196	0.39980	0.12096
4	0.19723	0.20118	-0.1583
5	0.15919	0.16237	-0.01121
6	0.16544	0.16875	0.09781
7	0.52430	0.05348	-0.07003
8	-0.10133	-0.10335	-0.26015
9	-0.19676	-0.20069	-0.20307
10	-0.34234	-0.34919	-0.19383
11	-0.40911	-0.41729	-0.11954
12	-0.41522	-0.423520	-0.09932
13	-0.36725	-0.374600	0.02018
14	-0.37725	-0.384800	-0.05115
15	-0.39769	-0.405640	-0.13012
16	-0.32919	-0.335770	0.00465
17	-0.38113	-0.388750	-0.12676
18	-0.23566	-0.240370	0.08039
19	-0.11537	-0.117680	0.04421
20	-0.12348	-0.125950	-0.16125
21	0.06959	0.070980	0.0926
22	0.05039	0.051400	-0.11353
23	0.14841	0.151370	0.11483
24	0.19536	0.199270	-0.03871
25	0.18428	0.187970	-0.147950
26	0.25974	0.264940	0.02967
27	0.30171	0.307740	0.0255
28	0.23339	0.238050	-0.09234
29	0.26462	0.269910	0.01801



รูปที่ 5.7 Autocorrelation Function of Standardized Annual Flows, Niger River



รูปที่ 5.8 Partial Autocorrelation Function of Standardized Annual Flows, Niger River

เมื่อ $\hat{\phi}_0 = -1$

$$d_0 = c_1 + c_{-1} = 2c_1$$

$$d_1 = c_2 + c_0$$

$$c'_0 = c_0 + \hat{\phi}_1^2 c_0 - 2\hat{\phi}_1 c_1$$

$$c'_1 = c_1 + \hat{\phi}_1^2 c_1 - \hat{\phi}_1(c_2 + c_0)$$

$$c'_0 = (1 + 0.8656^2)0.98039 - 2(0.8656)(0.52429) = 0.8074$$

$$c'_1 = (1 + 0.8656^2)0.52429 - 0.8656(0.45385 + 0.98039) = -0.3244$$

$$\hat{\sigma}_\varepsilon^2 = \frac{c'_0}{1 + \hat{\theta}_1^2} = \frac{0.8074}{1 + \hat{\theta}_1^2}$$

และ $\hat{\theta}_1 = \frac{-c'_1}{\hat{\sigma}_\varepsilon^2} = \frac{0.3244}{\hat{\sigma}_\varepsilon^2}$

$$\hat{\theta}_1 = \frac{0.3244(1 + \hat{\theta}_1^2)}{0.8074}$$

$$0.3244\hat{\theta}_1^2 - 0.8074\hat{\theta}_1 + 0.3244 = 0$$

$$\hat{\theta}_1 = 1.9852, 0.5037$$

Invertibility $|\hat{\theta}_1| < 1$

$$\hat{\theta}_1 = 0.5037$$

$$\hat{\sigma}_\varepsilon^2 = \frac{0.8074}{1 + 0.5037^2} = 0.6440$$

$$\hat{\theta}_{00} = \bar{z}(1 - \hat{\phi}_1)$$

$$= -0.0001(1 - 0.8656)$$

$$= -0.000013 \approx 0$$

Preliminary ARMA(1, 1) Model

$$\bar{z} = -0.0001 \approx 0$$

$$z_t = 0.8656z_{t-1} + \varepsilon_t - 0.5037\varepsilon_{t-1}$$

$$\bar{u} = 1$$

$$\hat{\theta}_{oo} = \bar{u}(1 - \hat{\phi}_1) = 1(1 - 0.8656) = 0.1344$$

$$u_t = 0.1344 + 0.8656u_{t-1} + \varepsilon_t - 0.5037\varepsilon_{t-1}$$

การประมาณค่าพารามิเตอร์ของอนุกรมเวลา z_t โดย Fortran Library Function FTARPS* และ

FTMPS*

$$\hat{\phi}_1 = 0.865634$$

$$\hat{\theta}_1 = 0.503631$$

$$\hat{\theta}_{oo} = -0.000011$$

$$\hat{\sigma}_\varepsilon^2 = 0.643987$$

ขั้นที่ 6 การประมาณค่าพารามิเตอร์โดยวิธี MLE

$$\text{Min } S = \sum_{t=1}^N \varepsilon_t^2$$

$$ST. 0.05 < \hat{\phi}_1 < 1.0$$

$$0.05 < \hat{\theta}_1 < 1.0$$

$$\text{สมมติว่า } \hat{\phi}_1 = 0.8; \hat{\theta}_1 = 0.4$$

$$\varepsilon_1 = 0$$

$$\varepsilon_2 = z_2 - \hat{\phi}_1 z_1 + \hat{\theta}_1 \varepsilon_1 = -0.871109 - 0.8 \times (-1.106432) + 0 = 0.014$$

$$\begin{aligned} \varepsilon_3 &= z_3 - \hat{\phi}_1 z_2 + \hat{\theta}_1 \varepsilon_2 \\ &= 1.114639 - 0.8 \times (-0.871109) + 0.4 \times 0.014 = 1.18172 \end{aligned}$$

$$\varepsilon_4 = -0.9782$$

$$\varepsilon_t = z_t - \hat{\phi}_1 z_{t-1} + \hat{\theta}_1 \varepsilon_{t-1}$$

ค่า $\varepsilon_t, \varepsilon_t^2$ แสดงอยู่ในค่าตารางที่ 5.4 ผลการคำนวณได้ค่า $S = \sum_{t=1}^{51} \varepsilon_t^2 = 32.0923$

ในทำนองเดียวกัน ผลการคำนวณค่า S เมื่อ $\hat{\phi}_1$ และ $\hat{\theta}_1$ มีค่าระหว่าง 0.0500-0.9500 แสดงอยู่ในตารางที่ 5.5 ซึ่งสามารถสรุปได้ว่า S จะมีค่าต่ำสุด เท่ากับ 32.01 เมื่อ $\hat{\phi}_1 = 0.8, \hat{\theta}_1 = 0.35$

ทำการคำนวณอย่างละเอียดใหม่จากผลการคำนวณในตารางที่ 5.5 โดยกำหนดให้ $\hat{\phi}_1$ มีค่าระหว่าง 0.7550-0.8500 และ $\hat{\theta}_1$ มีค่าระหว่าง 0.30500-0.4000 ดังแสดงในตารางที่ 5.6 ซึ่งสรุปได้ว่า S จะมีค่าต่ำสุด เท่ากับ 32.01119 เมื่อ $\hat{\phi}_1 = 0.7900$ และ $\hat{\theta}_1 = 0.345$

ทำการคำนวณอย่างละเอียดซ้ำใหม่จากผลการคำนวณในตารางที่ 5.6 โดยกำหนดให้ $\hat{\phi}_1$ มีค่าระหว่าง 0.78550-0.79500 และ $\hat{\theta}_1$ มีค่าระหว่าง 0.34050-0.35000 ดังแสดงในตารางที่ 5.7 ซึ่งสรุปได้ว่า S จะมีค่าต่ำสุด เท่ากับ 32.010992 เมื่อ $\hat{\phi}_1 = 0.7905$ และ $\hat{\theta}_1 = 0.3480$

ตารางที่ 5.4 Residuals, Squares of residuals and Sum of squares of residuals for $\phi_1 = 0.8$ and $\theta_1 = 0.4$

t	<i>Residuals</i> (ε_t)	<i>Squares of Residuals</i> (ε_t^2)
1	0.0000	0.0000
2	0.0140	0.0002
3	1.8172	3.3022
4	-0.9782	0.9568
5	0.4369	0.1909
6	-0.6031	0.3637
7	-1.6771	2.8127
8	-0.7460	0.5565
9	0.5934	0.3521
10	0.1217	0.0148
11	0.2279	0.0519
12	0.3777	0.1426
13	-0.3691	0.1363
14	-0.6373	0.4061
15	0.9378	0.8794
16	0.5307	0.2816
17	0.3741	0.1400
18	1.8861	3.5573
19	1.5281	2.3351
20	-0.7602	0.5780
21	0.5348	0.2860
22	0.9579	0.9175
23	0.4665	0.2177
24	0.3443	0.1186
25	-0.3445	0.1187
26	0.0496	0.0025
27	-0.2435	0.0593
28	-0.4962	0.2462
29	-0.2943	0.0866
30	0.5891	0.3470
31	-0.8543	0.7298
32	0.0315	0.0010
33	-0.3251	0.1057
34	-0.7303	0.5334
35	-0.3119	0.0973
36	-0.9430	0.8893
37	-0.1460	0.0213
38	-0.7669	0.5881
39	0.0450	0.0020
40	0.6554	0.4296
41	-0.4801	0.2305
42	0.7633	0.5826
43	-0.3924	0.1540
44	0.0941	0.0088
45	1.6931	2.8667
46	-0.3466	0.1201
47	0.8192	0.6710
48	0.8561	0.7329
49	0.5546	0.3076
50	-1.3610	1.8523
51	1.3076	1.7099
Total		32.0923

ตารางที่ 5.5 Sum of Squares Surface - First Run by Program FTMXL, Values of S x 100

	$\hat{\theta}_1(Theta)$																			
$\hat{\phi}_1$	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000	0.5500	0.6000	0.6500	0.7000	0.7500	0.8000	0.8500	0.9000	0.9500	
0.0500	4228	5166	5524	5954	6470	7092	7846	8766	9698	11303	13067	15302	18159	21831	26537	32473	39658	47628	55272	71928
0.1000	4502	4861	5175	5555	6013	6567	7241	8065	9079	10341	11926	13935	16505	19807	24037	29366	35805	42916	49656	63701
0.1500	4351	4584	4856	5182	5591	6081	6678	7410	8315	9440	10356	12652	14950	17902	21682	26439	32174	38479	44376	56032
0.2000	4146	4333	4556	4852	5203	5832	6157	6903	7603	6501	9857	11452	13493	16116	19473	23693	28766	34317	39431	48923
0.2500	3556	4110	4305	4548	4250	5221	5679	6244	6945	7223	8928	10335	12135	14449	17410	21126	25581	30429	34922	42373
0.3000	3752	3913	4073	4276	4531	4848	5242	5731	6341	7105	8071	9201	10876	12902	15492	18738	22619	26816	30548	36382
0.3500	3553	3744	3870	4033	4247	4514	4848	5266	5790	6449	7284	8350	9716	11474	13720	16531	19830	23477	26609	30951
0.4000	3539	3602	3697	3925	3357	4217	4496	4848	5293	5855	6569	7482	6655	10165	12094	14504	17364	20413	23006	26079
0.4500	3451	3489	3553	3649	3782	3958	4166	4478	4243	5321	5924	6697	7692	8975	10613	12656	15071	17623	19739	21766
0.5000	3338	3400	3438	3504	3602	3738	3919	4155	4458	4349	5350	5995	6929	7904	9277	10989	13001	15108	16806	18013
0.5500	3331	3340	3352	3390	3455	3556	3694	3879	4122	4437	4846	5377	6064	6952	8088	9500	11154	12867	14210	14819
0.6000	3339	3307	3296	3308	3345	3411	3511	3650	3838	4087	4414	4341	5398	6119	7044	8192	9530	10901	11948	12184
0.6500	3352	3301	3269	3257	3269	3305	3370	3469	3608	3799	4053	4389	4830	5406	6145	7063	8128	9209	10022	10109
0.7000	3391	3322	3271	3239	3227	3237	3272	3335	3432	3571	3762	4019	4382	4812	5393	6115	6950	7792	8432	8593
0.7500	3455	3370	3302	3252	3219	3207	3215	3249	3309	3405	3542	3733	3992	4337	4786	5346	5995	6650	7177	7636
0.8000	3545	3446	3363	3296	3247	3214	3201	3209	3240	3299	3393	3530	3721	3981	4324	4757	5263	5782	6258	7239
0.8500	3550	3548	3453	3372	3308	3260	3229	3217	3224	3255	3315	3410	3549	3744	4008	4348	4753	5189	5673	-7401
0.9000	3501	3678	3572	3420	3405	3344	3300	3272	3252	3273	3308	3373	3476	3627	3838	4118	4467	4870	5425	8122
0.9500	3867	3835	3720	3920	3536	3466	3412	3374	3353	3351	3371	3419	3501	3628	3813	4069	4403	4825	5512	9402
	4158	4019	3897	3792	3701	3627	3567	3524	3493	3491	3506	3548	3625	3749	3934	4199	4563	5056	5934	11242
Remark:	S x 100 = 3201; Min S = 32.01 when $\hat{\phi}_1 = 0.8$, $\hat{\theta}_1 = 0.35$																			

ตารางที่ 5.6 Sum of Squares Surface - First Run by Program FTMXL, Values of $(S-32) \times 10^5$ (Refine the result from Table 5.5)

$\hat{\phi}_1$	$\hat{\theta}_1(Theta)$																			
	0.3050	0.3100	0.3150	0.3200	0.3250	0.3300	0.3350	0.3400	0.3450	0.3500	0.3550	0.3600	0.3650	0.3700	0.3750	0.3800	0.3850	0.3900	0.3950	0.4000
0.7550	5774	5629	5699	5986	6494	7224	8180	9364	10780	12430	14317	16445	18817	21437	24308	27434	30820	34468	38384	42572
0.7600	5011	4643	4487	4546	4821	5317	6034	6976	8146	9547	11181	13053	15164	17519	20122	22975	26083	29450	33079	36976
0.7650	4632	4045	3668	3502	3550	3814	4297	5002	5931	7087	8473	10092	11948	14044	16382	18968	21803	24893	28242	31852
0.7700	4638	3836	3241	2854	2678	2716	2969	3440	4133	5049	6191	7564	9168	11009	13089	15411	17980	20799	23872	27202
0.7750	5028	4015	3206	2602	2207	2022	2050	2292	2752	3433	4337	5467	6825	8416	10242	12307	14614	17167	19969	23025
0.7800	5802	4582	3563	2747	2137	1733	1529	1557	1790	2240	2909	3901	4919	6264	7842	9654	11704	13996	16534	19320
0.7850	6961	5538	4313	3239	2467	1849	1438	1236	1246	1469	1909	2568	3448	4554	5888	7452	9251	11288	13566	16089
0.7900	8505	6882	5455	4226	3197	2359	1746	1328	1119	1120	1335	1766	2415	3285	4380	5702	7255	9042	11066	13330
0.7950	10432	8614	6990	5560	4328	3295	2462	1833	1410	1194	1188	1395	1818	2458	3319	4404	5716	7258	9033	11045
0.8000	12745	10735	8317	7291	5859	4625	3583	2752	2119	1690	1469	1457	1657	2072	2704	3557	4633	5935	7467	9232
0.8050	15441	13244	11236	9417	7791	6359	5123	4004	3245	2608	2176	1950	1933	2127	2536	3161	4007	5075	6369	7893
0.8100	18522	16141	13947	11941	10124	8498	7066	5629	4790	3949	3310	2874	2645	2624	2814	3217	3838	4677	5739	7026
0.8150	21223	19427	17051	14860	12857	11042	9419	7988	6752	5712	4871	4231	3794	3562	3538	3725	4125	4741	5576	6633
0.8200	25838	23101	20547	18176	15590	13951	12180	19560	9132	7898	6659	6019	5379	4942	4709	4684	4869	5267	5880	6712
0.8250	30072	27163	24435	21888	19524	17344	15351	13546	11930	10505	9274	8239	7401	6763	6326	6095	6070	6255	6652	7264
0.8300	34691	31614	28716	25395	23458	21102	18931	16944	15145	13535	12116	10890	9859	9025	8390	7957	7728	7705	7891	8290
0.8350	39694	36453	33388	30501	27793	25265	22919	20756	18779	16988	15385	13973	12754	11729	10900	10270	9842	9617	9598	9788
0.8400	45082	41631	38454	35403	32526	29833	27316	24922	22820	20882	19081	17488	16085	14874	13857	13035	12413	11991	11772	11760
0.8450	50854	47296	43911	40700	37664	34805	32123	29620	27299	25159	23204	21435	19853	18460	17259	16252	15441	14827	14414	14204
0.8500	57010	53300	49761	46394	43200	40181	37333	24673	32126	29879	27754	25813	24057	22488	21109	19920	18125	18125	17523	17121
Remark: When $\hat{\phi}_1 = 0.79, \hat{\theta}_1 = 0.345$, the result is $(S - 32) \times 10^5 = 1119$. Thus $S = 30 + 1119/10^5$																				

ตารางที่ 5.7 Sum of Squares Surface - First Run by Program FTMXL, Values of (S-32.011) x 10⁷ (Refine the result from Table 5.6)

$\hat{\phi}_1$	$\hat{\theta}_1(Theta)$																			
	0.3405	0.3410	0.3415	0.3420	0.3425	0.3430	0.3435	0.3440	0.3445	0.3450	0.3455	0.3460	0.3465	0.3470	0.3475	0.3480	0.3485	0.3490	0.3495	0.3500
0.7855	11639	10770	10113	9667	9434	9413	9604	10008	10626	11456	12500	13758	15230	16916	18817	20933	23264	25810	28571	31549
0.7860	10893	9805	8928	8263	7809	7568	7538	7721	8117	8726	9548	10584	11833	13296	14974	16866	18973	21295	23832	26584
0.7665	10501	9254	9158	7273	6600	6138	5889	5851	6026	6414	7014	7828	8855	10096	11551	13220	15103	17201	19514	22042
0.7870	10643	9117	7803	-6699	5807	5126	4656	4398	4353	4519	4899	5491	6296	7315	8547	9994	11654	13529	15618	17922
0.7875	11138	9395	7862	6540	5429	4528	3839	3362	3096	3043	3202	3573	4157	4954	5964	7168	8626	10278	12144	14225
0.7880	12047	10087	8336	6796	5485	4347	3439	2743	2297	1984	1922	2073	2436	3012	3801	4803	6019	7449	9092	10950
0.7885	13370	11193	9225	7467	5919	4562	3456	2540	1836	1243	1062	992	1135	1490	2058	2839	3833	5041	6462	8097
0.7890	15107	12713	10528	8235	6783	5233	3889	2755	1832	1120	619	330	253	388	735	1295	2068	3054	4254	5667
0.7895	17258	14648	12246	10055	8072	6300	4738	3388	2245	1314	595	87	-209	-294	-167	172	724	1499	2467	3650
0.7900	19822	16996	14379	11971	9772	7783	6004	4434	3075	1926	989	262	-253	-557	-649	-530	-198	345	1103	2073
0.7905	28800	19759	16996	14203	11258	9682	7686	5899	4323	2957	1801	856	121	-401	-712	-812	-700	-276	160	909
0.7910	26192	22936	19329	17049	14410	11997	9784	7781	5928	4404	3031	1868	916	174	-355	-674	-781	-677	-360	168
0.7915	29998	26528	23265	20211	17365	14728	12288	10080	8070	6270	4680	3259	2129	1170	421	-115	-441	-556	-459	-150
0.7920	34218	30533	27057	23788	20727	17875	15231	12796	10570	8554	6747	5149	3762	2585	1619	863	318	-14	-136	-46
0.7925	30051	34953	31263	27780	24505	21438	18579	15329	13467	11255	9232	7418	5814	4420	3236	2263	1500	948	608	479
0.7930	42888	39787	39683	32187	23693	25417	22343	19478	16922	14374	12135	10105	2285	6674	5274	4083	3103	2333	1775	1427
0.7935	49339	45035	40919	37009	33307	29912	26524	23445	20574	17511	15457	13211	11175	9348	7731	6324	5127	4140	3363	2798
0.7940	55234	50998	46369	42245	38331	34623	31122	27823	24743	21866	19197	16735	14485	12442	10609	8985	7571	6367	5374	4591
0.7945	81522	56775	52234	47898	43771	39950	36135	32629	29330	26238	23355	20680	18213	15955	13906	12067	10437	9016	7806	6806
0.7950	68224	63255	58513	53996	49626	45493	41568	37846	34333	31028	27931	25042	22361	19888	17624	15569	13723	12087	10660	9443

Remark: When $\hat{\phi}_1 = 0.7905, \hat{\theta}_1 = 0.348$, the result is $(S - 32.011)x10^7 = -812$. Thus $S = 32.011 - \frac{812}{10^7} = 30.010992 \sim 32.011$

การหาค่า Standard Error ของพารามิเตอร์ ϕ_1 จากสมการ 5.51

$$\begin{aligned}\sigma(\hat{\phi}_1) &= \left[\frac{1}{N} \frac{(1-\hat{\phi}_1\hat{\theta}_1)^2(1-\hat{\phi}_1^2)}{(\hat{\phi}_1-\hat{\theta}_1)^2} \right]^{\frac{1}{2}} \\ &= \left[\frac{1}{51} \frac{(1-0.7905 \times 0.348)^2(1-0.7905^2)}{(0.7905-0.348)^2} \right]^{1/2} \\ &= 0.140\end{aligned}$$

การหาค่า Standard Error ของพารามิเตอร์ θ_1 จากสมการ 5.52

$$\begin{aligned}\sigma(\hat{\theta}_1) &= \left[\frac{1}{N} \frac{(1-\hat{\phi}_1\hat{\theta}_1)^2(1-\hat{\theta}_1^2)}{(\hat{\phi}_1-\hat{\theta}_1)^2} \right]^{\frac{1}{2}} \\ &= \left[\frac{1}{51} \frac{(1-0.7905 \times 0.348)^2(1-0.348^2)}{(0.7905-0.348)^2} \right]^{1/2} \\ &= 0.215\end{aligned}$$

ทางเลือกในการหาค่า Standard Error ของพารามิเตอร์ ϕ_1 และ θ_1 คือการหา Variance-Covariance Matrix ของพารามิเตอร์ ϕ_1 และ θ_1 ดังสมการที่ 5.49

$$V(\hat{\beta}) = 2\hat{\sigma}_\varepsilon^2 \begin{bmatrix} \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\phi}_1^2} & \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\phi}_1 \partial \hat{\theta}_1} \\ \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\phi}_1 \partial \hat{\theta}_1} & \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\theta}_1^2} \end{bmatrix}$$

(1) การหา Second Derivative ของ S เทียบกับ $\hat{\phi}_1$ โดยใช้ผลการคำนวณจากตารางที่ 5.7 ดังแสดงอยู่ในตารางที่ 5.8

ตารางที่ 5.8 การหา Second Derivative ของ S เทียบกับ $\hat{\phi}_1$ ที่ $\hat{\theta}_1 = 0.3480$

ที่ $\hat{\theta}_1 = 0.3480$				
	$\hat{\phi}_1$	S'	$\Delta S'$	$\Delta^2 S'$
	0.7895	172		
	0.7900	-530	172-(-530)=702	
ที่ $\hat{\phi}_1 = 0.005$	0.7905	-812	(-530+812)=282	702-282=440
	0.7910	-674	-812+674=-138	282+138=420
	0.7915	-115	-674+115=-559	-138+559=421

$$S' = (S - 32.011) \times 10^7$$

ค่า Second Difference ของ S' เทียบกับ $\hat{\phi}_1$ ที่ $\hat{\phi}_1 = 0.7905$ และ $\hat{\theta}_1 = 0.3480$ คือ 420 และค่า Second Derivative ของ S' เทียบกับ $\hat{\phi}_1$ คือ

$$\frac{\partial S^2}{\partial \phi_1^2} \approx \frac{420}{10^7 \times 0.0005^2} = 168$$

(2) หา Second Derivative ของ S เทียบกับ $\hat{\theta}_1$ โดยใช้ผลการคำนวณจากตารางที่ 5.7 ในทำนองเดียวกับ (1) ดังแสดงอยู่ในตารางที่ 5.9

ตารางที่ 5.9 การหา Second Derivative ของ S เทียบกับ $\hat{\theta}_1$ ที่ $\hat{\phi}_1 = 0.7905$

$\hat{\phi}_1 = 0.7905$				
	$\hat{\theta}_1$	S'	$\Delta S'$	$\Delta^2 S'$
	0.3470	-401		
	0.3475	-712	-401+712=311	
ที่ $\hat{\theta}_1 = 0.0005$	0.3480	-812	-712+812=100	311-100=211
	0.3485	-700	-812+700=-112	100+112=212
	0.349	-376	-700+376=-324	-112+324=212

$$S' = (S - 32.001) \times 10^7$$

ค่า Second Difference (S') เทียบกับ $\hat{\theta}_1$ ที่ $\hat{\phi}_1 = 0.7905$ และ $\hat{\theta}_1 = 0.3480$ คือ 212 และค่า

Second Derivative ของ S' เทียบกับ $\hat{\theta}_1$ คือ

$$\frac{\partial S^2}{\partial \theta_1^2} \approx \frac{212}{10^7 \times 0.0005^2} = 84.8$$

(3) หา Second Derivative ของ S เทียบกับ $\hat{\phi}_1$ และ $\hat{\theta}_1$ โดยใช้ผลการคำนวณจากตารางที่ 5.7 ในทำนองเดียวกับ (1) ดังแสดงอยู่ในตารางที่ 5.10

ตารางที่ 5.10 การหา Second Derivative ของ S เทียบกับ $\hat{\phi}_1$ และ $\hat{\theta}_1$

ϕ	θ											
	0.3470			0.3475			0.3480			0.3485		
	S'	$\Delta_\phi S'$	$\Delta^2_{\phi\theta} S'$	S'	$\Delta_\phi S'$	$\Delta^2_{\phi\theta} S'$	S'	$\Delta_\phi S'$	$\Delta^2_{\phi\theta} S'$	S'	$\Delta_\phi S'$	$\Delta^2_{\phi\theta} S'$
0.7895	-294			-167			172			742		
0.7900	-557	263	263-482=-219	-649	482	482-702=-220	-530	702	702-940=-238	-198	940	
0.7905	-401	-156	-219	-712	63	(-219)	-812	282	(-220)	-700	502	
0.7910	174	-575	-218	-355	-357	(-219)	-674	-138	(-219)	-781	81	
0.7915	1170	-996	-220	421	-776	-217	-115	-559	-219	-441	-340	
หมายเหตุ	$\Delta_\phi = \frac{\partial S}{\partial \phi_1} \quad \Delta^2_{\phi\theta} = \frac{\partial^2 S}{\partial \phi_1 \partial \theta_1}$											

$$S' = (S - 32.001) \times 10^7$$

ค่า Second Difference ของ S' เทียบกับ $\hat{\phi}_1$ และ $\hat{\theta}_1$ ที่ $\hat{\phi}_1 = 0.7905$ และ $\hat{\theta}_1 = 0.3480$ จะมีค่า

เท่ากับ $(-219-219-220-219)/4 = 219.25$

และค่า Second Derivative ของ S' เทียบกับ $\hat{\phi}_1$ และ $\hat{\theta}_1$ คือ

$$\frac{\partial S^2}{\partial \hat{\phi}_1 \hat{\theta}_1} \approx \frac{219.25}{10^7 \times 0.0005 \times 0.0005} = -87.7$$

(4) หา $\hat{\sigma}_\varepsilon^2$ สมการ 5.46

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= \frac{1}{N} S(\hat{\phi}_1, \hat{\theta}_1) \\ &= \frac{1}{51} \times 32.0109 = 0.6277\end{aligned}$$

(5) Variance-Covariance Matrix ของพารามิเตอร์

$$\begin{aligned}V(\hat{\beta}) &= 2\hat{\sigma}_\varepsilon^2 \begin{bmatrix} \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\phi}_1^2} & \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\phi}_1 \partial \hat{\theta}_1} \\ \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\phi}_1 \partial \hat{\theta}_1} & \frac{\partial^2 S(\hat{\phi}_1, \hat{\theta}_1)}{\partial \hat{\theta}_1^2} \end{bmatrix} \\ &= 2 \times 0.6277 \begin{bmatrix} 168 & -87.7 \\ -87.7 & 84.8 \end{bmatrix}^{-1} \\ &= 2 \times 0.6277 \begin{bmatrix} 0.01294 & 0.0134 \\ 0.0134 & 0.0256 \end{bmatrix} \\ &= \begin{bmatrix} \text{Var}(\hat{\phi}_1) & \text{Cov}(\hat{\phi}_1, \hat{\theta}_1) \\ \text{Cov}(\hat{\phi}_1, \hat{\theta}_1) & \text{Var}(\hat{\theta}_1) \end{bmatrix}\end{aligned}$$

$$\text{Var}(\hat{\phi}_1) = 2 \times 0.6277 \times 0.01294$$

$$\sigma(\hat{\phi}_1) = 0.127$$

$$\text{Var}(\hat{\theta}_1) = 2 \times 0.6277 \times 0.0256$$

$$\sigma(\hat{\theta}_1) = 0.179$$

ถ้าคำนวณ 95% CI จากสมการ 5.51 และ 5.52 จะได้

$$95\% \text{ CI. of } \phi_1 = 0.7905 \pm 1.96 \times 0.14$$

$$0.5161 < \phi_1 < 1.0649 \quad (\text{Stationary})$$

$$95\% \text{ CI. of } \theta_1 = 0.348 \pm 1.96 \times 0.215$$

$$-0.0734 < \theta_1 < 0.7594$$

ผลการคำนวณหาค่าพารามิเตอร์โดยโปรแกรม FTMXL

$$9 \text{ Iterations : } \hat{\phi}_1 = 0.791457; \hat{\theta}_1 = 0.348672$$

$$\hat{\theta}_{oo} = -0.000017; \hat{\sigma}_\varepsilon^2 = 0.627668$$

$$50 \text{ Iterations : } \hat{\phi}_1 = 0.790691; \hat{\theta}_1 = 0.348173$$

$$\hat{\theta}_{oo} = -0.000017; \hat{\sigma}_\varepsilon^2 = 0.627668$$

ผลการคำนวณค่า $\varepsilon_t, \varepsilon_t^2, \sum \varepsilon_t^2$ โดยโปรแกรม FTMAXL แสดงอยู่ในตารางที่ 5.11 คุณสมบัติทางสถิติของ Residual(ε_t) แสดงอยู่ในตารางที่ 5.12 และค่าพารามิเตอร์ของแบบจำลองสโตแคสติกแบบต่างๆ ที่คำนวณโดยโปรแกรม FTMAXL แสดงอยู่ในตารางที่ 5.13

ตารางที่ 5.11 Residuals, Squares of Residuals and Sums of Squares Obtained from IMSL Program FTMAXL

t (Rank)	ε_t (Residuals)	ε_t^2 (Square of Residuals)	$S = \sum \varepsilon_t^2$ (SS. Residuals)
1	0	0	0
2	0.004602	0.000021	0.000021
3	1.305755	3.260753	3.260774
4	-1.065905	1.136153	4.396927
5	0.449590	0.202131	4.599058
6	-0.619512	0.383795	4.982853
7	-1.657305	2.746660	7.729513
8	-0.669589	0.448349	8.177862
9	0.644370	0.415213	8.593075
10	0.105567	0.011144	8.604219
11	0.212192	0.045025	8.649245
12	0.359002	0.128883	8.778127
13	-0.393763	0.155049	8.933176
14	-0.630371	0.397367	9.330544
15	-0.909574	0.827324	10.157868
16	0.577270	0.333241	10.491109
17	0.361756	0.130867	10.621976
18	1.862870	3.470286	14.092262
19	1.438284	2.068661	16.160923
20	-0.851317	0.724740	16.885663
21	0.545302	0.297354	17.183017
22	0.943834	0.890823	18.073840
23	0.426643	0.182024	18.255865
24	0.318510	0.101448	18.357313
25	-0.360165	0.129719	18.487032
26	0.066525	0.004426	18.491458
27	-0.234859	0.055159	18.546616
28	-0.478673	0.229127	18.775744
29	-0.264462	0.069940	18.845684
30	0.612344	0.374965	19.220650
31	-0.872161	0.760665	19.981315
32	0.063208	0.003995	19.985310
33	-0.317206	0.100619	20.085929
34	-0.714991	0.511212	20.597141
35	-0.277526	0.077021	20.674162
36	-0.321925	0.849946	21.524108
37	-0.102758	0.010559	21.534667
38	-0.752300	0.565955	22.100622
39	0.076972	0.005925	22.106547
40	0.657309	0.432055	22.538602
41	-0.513239	0.263414	22.802016
42	0.769940	0.592808	23.394824
43	-0.426241	0.181681	23.576505
44	0.988480	0.009771	23.586276
45	1.689296	2.853721	26.439997
46	-0.421256	0.177457	26.617454
47	0.813059	0.661065	27.278519
48	0.821828	0.675401	27.953920
49	0.511183	0.261309	28.215229
50	-1.392861	1.940063	30.155292
51	1.362273	1.855787	32.011079

ตารางที่ 5.12 Statistics of the Residuals in Table 5.11

MEAN = 0.0539, VAR = 0.6248				
k Lag	$c_k(\varepsilon)$ Autocovar.	$r_k(\varepsilon)$ Autocorr.	$\sum r_k^2(\varepsilon)$ SSQ. Autocorr.	$\phi_k^{(k)}$ Partial Autocorr.
1	-0.02256	-0.03610	0.0013	-0.0361
2	0.04016	0.06428	0.00544	0.06306
3	0.07046	0.11278	0.01815	0.11785
4	-0.6901	-0.11046	0.03035	-0.10792
5	-0.03044	-0.04872	0.03273	-0.07369
6	0.06030	0.09651	0.04204	0.09839
7	0.06379	0.10211	0.05247	0.14944
8	-0.01173	-0.01877	0.05282	-0.02655
9	-0.01794	-0.02871	0.05365	-0.09669
10	-0.09876	-0.15807	0.07863	-0.18132
11	-0.12622	-0.20202	0.11944	-0.17457
12	-0.10522	-0.16842	0.14781	-0.16197
13	-0.03131	-0.05012	0.15032	-0.04769
14	-0.06449	-0.10323	0.16098	-0.11166
15	-0.07973	-0.12761	0.17726	-0.18308
16	-0.00033	-0.00052	0.17726	-0.04139
17	-0.12737	-0.20387	0.21883	-0.15880
18	-0.01558	-0.02494	0.21945	-0.00182
19	0.02623	0.04199	0.22121	0.04408
20	-0.08068	-0.12913	0.23789	-0.16143
21	0.12255	0.19615	0.27636	0.10352
22	-0.01591	-0.02547	0.27701	-0.09944
23	0.05219	0.08354	0.28399	0.04289
24	0.04297	0.06877	0.28872	-0.02014
25	-0.02497	-0.03997	0.29032	-0.17118
26	0.05075	0.08124	0.29692	-0.05581
27	0.0894	0.14309	0.31739	0.03417
28	-0.00531	-0.00850	0.31746	-0.12399
29	0.09309	0.14899	0.33966	0.03097

ตารางที่ 5.13 ค่าพารามิเตอร์ของแบบจำลองสโตแคสติกแบบต่างๆ ที่คำนวณโดยโปรแกรม FTMAXL

Model	$\hat{\theta}_{00}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\sigma}_\varepsilon^2$
AR(1)	-0.000036	0.5588	-	-	0.663424
AR(2)	-0.000028	0.4036	0.2534	-	0.622512
ARMA(1, 1)	-0.000017	0.7907	-	0.3482	0.627668
ARMA(2, 1)	-0.000017	0.7351	0.0586	0.3171	0.625062

การใช้ Solver ใน Excel หาค่าพารามิเตอร์โดยวิธี Optimization แสดงอยู่ในตารางที่ 5.14

ตารางที่ 5.14 การใช้ Solver หาค่าพารามิเตอร์ ARMA(1,1) โดยวิธี Non-linear Optimization

t	Z_t	ε_t
1.000	-1.1064	0.000
2.000	-0.8710	0.004
3.000	1.1148	1.805
4.000	-0.8132	-1.066
5.000	0.1776	0.449
6.000	-0.6357	-0.620
7.000	-1.9444	-1.658
8.000	-1.6307	-0.670
9.000	-0.4128	0.643
10.000	-0.4458	0.104
11.000	-0.1774	0.211
12.000	0.1446	0.358
13.000	-0.4045	-0.394
14.000	-0.8132	-0.631
15.000	-1.3334	-0.910
16.000	-0.1609	0.577
17.000	0.0331	0.361
18.000	1.7629	1.862
19.000	2.1840	1.439
20.000	0.3758	-0.850
21.000	1.1395	0.546
22.000	1.6556	0.945
23.000	1.4079	0.428
24.000	1.2840	0.320
25.000	0.5450	-0.359
26.000	0.6235	0.068
27.000	0.2354	-0.234
28.000	-0.2105	-0.478
29.000	-0.2641	-0.264
30.000	0.4955	0.612
31.000	-0.6935	-0.872
32.000	-0.1816	0.063
33.000	-0.4830	-0.317
34.000	-0.9866	-0.715
35.000	-0.8091	-0.278
36.000	-1.4655	-0.923
37.000	-0.9412	-0.104
38.000	-1.4614	-0.753
39.000	-0.8174	0.076
40.000	-0.0164	0.656
41.000	-0.7554	-0.514
42.000	0.3510	0.769
43.000	-0.4169	-0.427
44.000	-0.0825	0.099
45.000	1.5895	1.689
46.000	0.2478	-0.421
47.000	1.1561	0.814
48.000	1.4533	0.823
49.000	1.3749	0.512
50.000	-0.4830	-1.392
51.000	1.4657	1.363
Mean	0.0000	SSE = 32.01105
Standard Deviation	1.00000	

ARMA(1,1)

$$z_t = \phi_1 z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \varepsilon_{00}$$

$$\varepsilon_{00} = \bar{z}(1 - \phi_1)$$

$$\varepsilon_t = z_t - \phi_1 z_{t-1} + \theta_1 \varepsilon_{t-1} - \varepsilon_{00}$$

Non – linear optimization

Minimize $\sum_{t=1}^n \varepsilon_t^2$

st. $|\phi_1| < 1$
 $|\theta_1| < 1$

Solution

$$\phi_1 = 0.79067$$

$$\theta_1 = 0.34816$$

$$SSE = S(\underline{\phi}, \underline{\theta}) = \sum_{t=1}^n \varepsilon_t^2 = 32.01105$$

$$\sigma_\varepsilon^2 = \frac{S(\underline{\phi}, \underline{\theta})}{n - 1} = \frac{32.01105}{51 - 1} = 0.64022$$

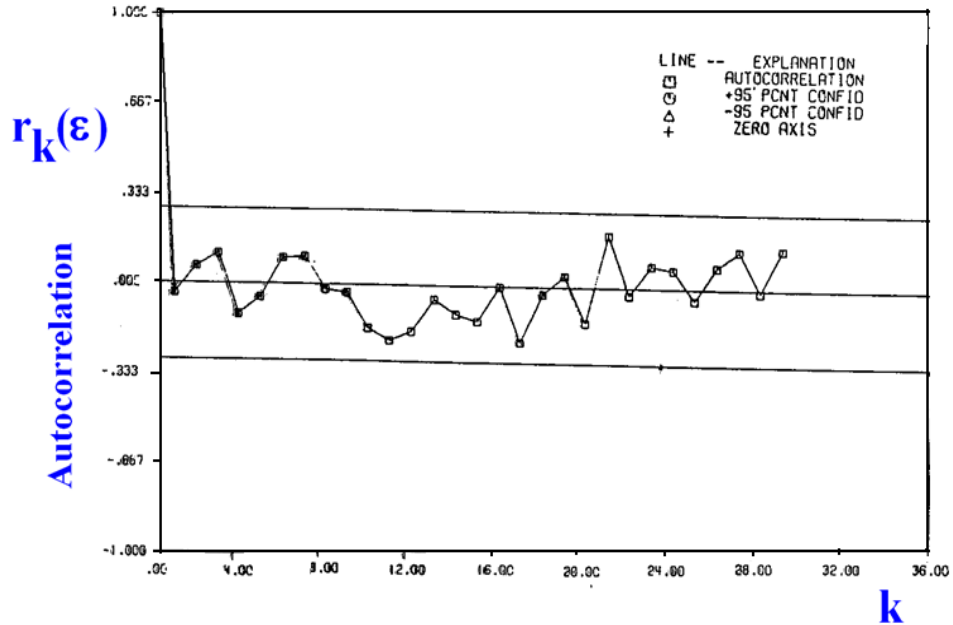
ขั้นที่ 7 Goodness of Fit

$$L = \frac{N}{10} + p + q = \frac{51}{10} + 1 + 1 = 7.1 \sim 8$$

$$Q = N \sum_{k=1}^{L=8} [r_k(\varepsilon)]^2 = 51 \times 0.33966 = 17.3227$$

$Q < \chi_{0.05, 29-1-1}^2 (= 40.113)$ แสดงว่า ε_t เป็น Independent ตามสมมติฐานที่ตั้งไว้

กราฟ ACF ของ ε_t แสดงอยู่ในรูปที่ 5.9



รูปที่ 5.9 Autocorrelation function of residuals

ขั้นที่ 8 การทดสอบ Parsimony ด้วย Akaike Information Criterion(AIC)

จากสมการที่ 5.61 $AIC(p, q) = N \ln(\text{MLE of } \hat{\sigma}_\varepsilon^2) + 2(p+q)$

แบบจำลอง	AIC(p,q)
AR(1)	$AIC(1,0) = 51 \times \ln(0.663424) + 2(1) = -18.9174$
AR(2)	$AIC(2,0) = 51 \times \ln(0.622512) + 2(2) = -20.1736$ต่ำสุด
ARMA(1,1)	$AIC(1,1) = 51 \times \ln(0.627668) + 2(2) = -19.7529$
ARMA(2,1)	$AIC(2,1) = 51 \times \ln(0.625062) + 2(3) = -17.9651$

แบบจำลอง AR(2) ให้ค่า AIC ต่ำสุด จึงถือว่ามีความเหมาะสมตามหลัก Parsimony และแบบจำลอง ARMA(1,1) ให้ค่า AIC สูงกว่า AR(2) เพียงเล็กน้อยจึงถือว่ามีความเหมาะสมใกล้เคียงกับ AR(2)

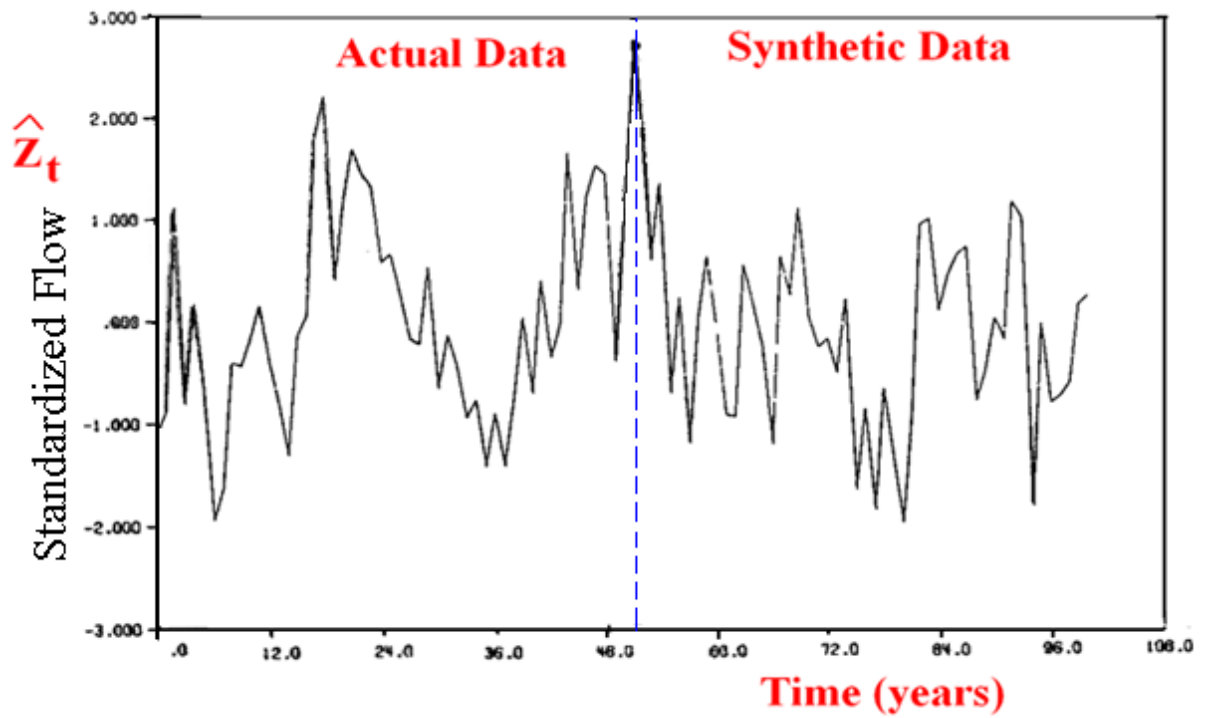
ขั้นที่ 9 การสังเคราะห์อนุกรมเวลา z_t โดยใช้แบบจำลอง ARMA(1,1) แสดงอยู่ในตารางที่ 5.15 กราฟแสดงเปรียบเทียบระหว่างอนุกรมเวลาที่สังเคราะห์ใหม่กับอนุกรมเวลาตัวอย่างของ z_t และ q_t แสดงอยู่ในรูปที่ 5.10 และ 5.11

ตารางที่ 5.15 Generation , Forecasting Weights , For 95 % Forecast Confidence Intervals Calculated for the parameters $\hat{\phi}_1 = 0.790691, \hat{\theta}_1 = 0.384173, \theta_{00} = -0.000017, \hat{\sigma}_\varepsilon^2 = 0.627668$ by Program FTCMP

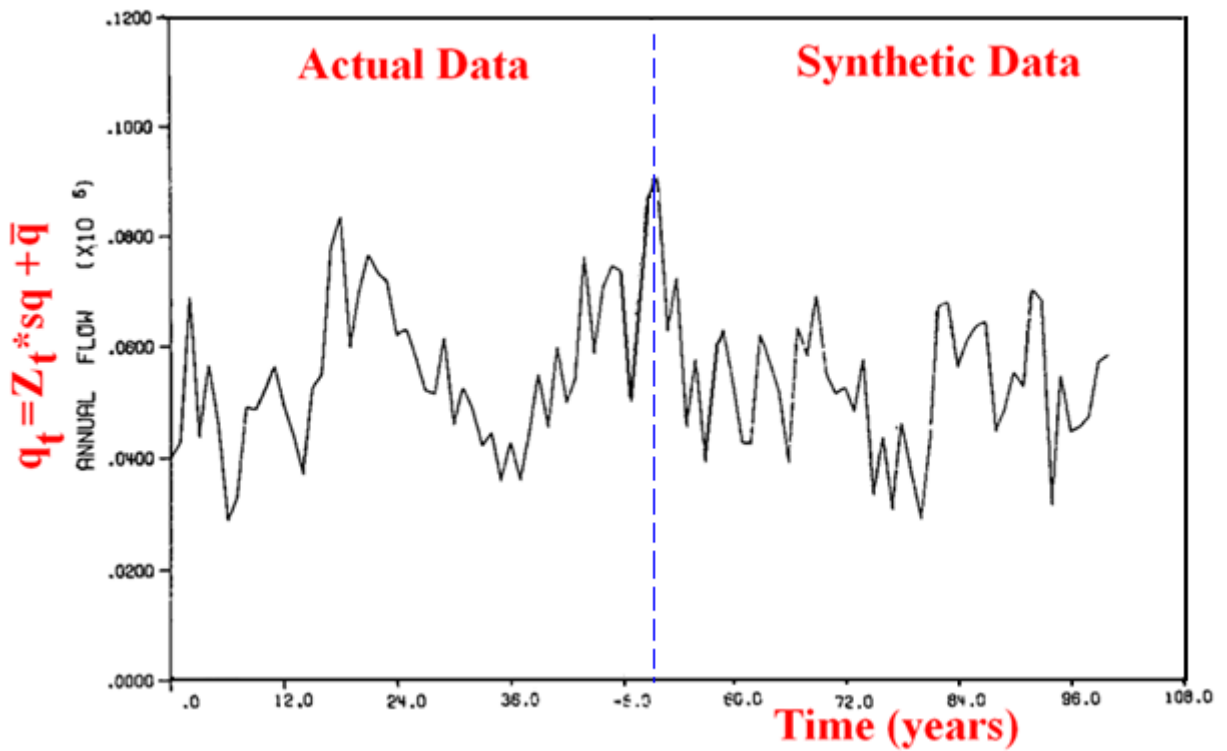
Lead	$\hat{Z}_t(\text{Gen.}) - \text{simulation}$	$\psi\text{-weights}$	$Z_t(L)$	Deviations
1	2.385685	0.442518	0.684271	1.540301
2	2.665586	0.349895	0.541030	1.684376
3	0.531424	0.276659	0.427770	1.768497
4	1.277439	0.218751	0.338217	1.819114
5	-0.775660	0.172965	0.267408	1.850055
6	0.161317	0.136762	0.211420	1.869140
7	-1.268832	0.108136	0.167151	1.880973
8	-0.432700	0.085502	0.132148	1.888333
9	0.560117	0.067606	0.104471	1.892920
10	-0.147956	0.053455	0.082587	1.895782
11	-0.992489	0.042267	0.065284	1.897569
12	-1.010478	0.033420	0.051603	1.898686
13	0.474262	0.026425	0.040785	1.899383
14	0.080840	0.020894	0.032231	1.899819
15	-0.328495	0.016521	0.025468	1.900092
16	-1.289060	0.013063	0.020120	1.900262
17	0.555338	0.010328	0.015892	1.900369
18	0.170012	0.008167	0.012549	1.900435
19	1.017159	0.006457	0.009905	1.900477
20	-0.059452	0.005106	0.007815	1.900503
21	-0.339637	0.004037	0.006162	1.900519
22	-0.262400	0.003192	0.004856	1.900530
23	-0.593587	0.002524	0.003822	1.900536
24	0.133808	0.001996	0.003005	1.900540
25	-1.737364	0.001578	0.002359	1.900542
26	-0.941933	0.001248	0.001849	1.900544
27	-1.930718	0.000967	0.001445	1.900545
28	-0.750438	0.000780	0.001125	1.900546
29	-1.445572	0.000617	0.000873	1.900546
30	-2.061807	0.000488	0.000673	1.900546
31	-1.016151	0.000386	0.000515	1.900546
32	0.835590	0.000305	0.000391	1.900546
33	0.890336	0.000241	0.000292	1.900546
34	-0.007729	0.000191	0.000214	1.900546
35	0.348259	0.000151	0.000152	1.900546
36	0.541926	0.000119	0.000103	1.900547
37	0.615442	0.000094	0.000065	1.900547
38	-0.885720	0.000075	0.000034	1.900547
39	-0.581482	0.000059	0.000010	1.900547
40	-0.079517	0.000047	-0.000009	1.900547
41	-0.287367	0.000037	-0.000024	1.900547
42	1.050647	0.000029	-0.000036	1.900547
43	0.893466	0.000023	-0.000045	1.900547
44	-1.913883	0.000018	-0.000053	1.900547
45	-0.134659	0.000014	-0.000059	1.900547
46	-0.905882	0.000011	-0.000063	1.900547
47	-0.842695	0.000009	-0.000067	1.900547
48	-0.713980	0.000007	-0.000070	1.900547
49	0.035985	0.000006	-0.000072	1.900547
50	0.126156	0.000004	-0.000074	1.900547

Remark: $Z_t = \hat{\phi}_1 Z_t(L-1) + \hat{\theta}_{00}, L \geq 2; Z_{51} = 1.46509, \varepsilon_{51} = 1.362273$

+



รูปที่ 5.10 Generation of z_t series , Niger River



รูปที่ 5.11 Actual and synthetic annual flows , Niger River

ขั้นที่ 10 การพยากรณ์ (Forecasting)

$$z_t(1) = \hat{\phi}_1 z_t - \hat{\theta}_1 \varepsilon_t + \hat{\theta}_{00}$$

$$z_t(2) = \hat{\phi}_1 z_t(1) + \hat{\theta}_{00}$$

$$z_t(L) = \hat{\phi}_1 z_t(L-1) + \hat{\theta}_{00}; L \geq 2$$

$$\hat{\phi}_1 = 0.790691; \hat{\theta}_1 = 0.348173; \hat{\theta}_{00} = -0.000017$$

$$z_{51} = 1.465609$$

$$\varepsilon_{51} = 1.362273$$

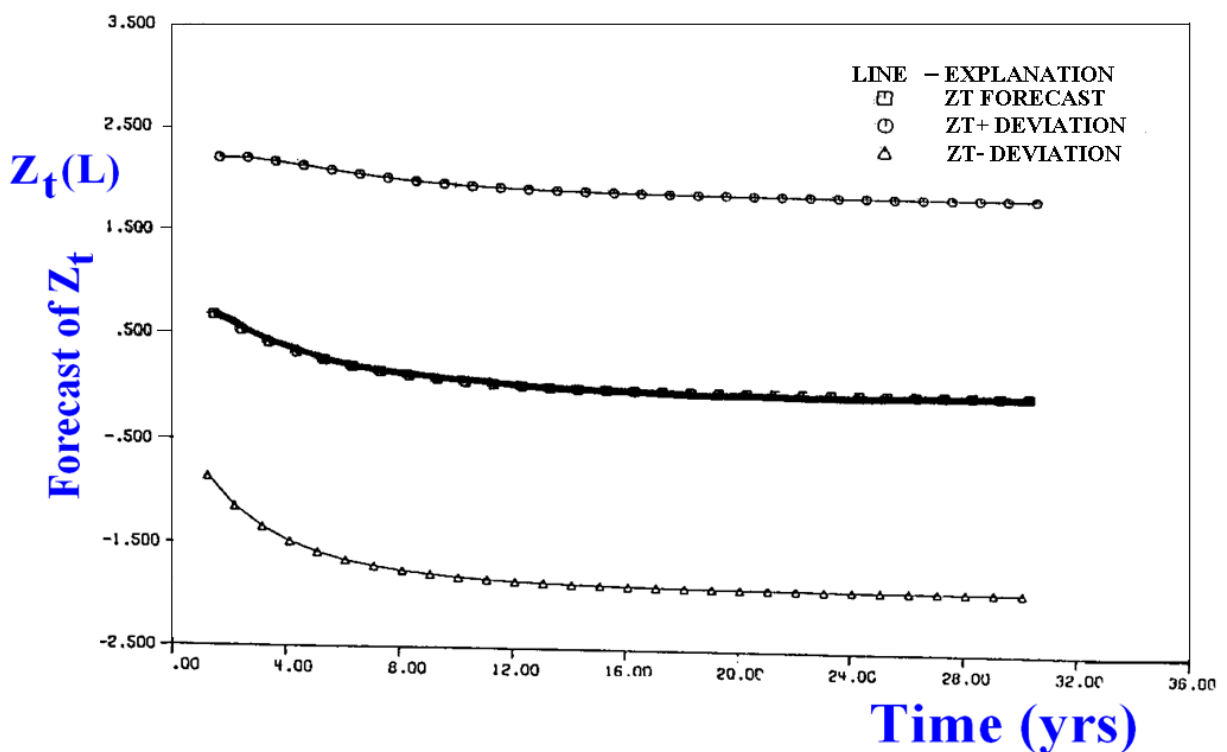
$$z_{51}(1) = 0.790691 \times 1.465609 - 0.348173 \times 1.362273 - 0.000017 = 0.6845$$

$$z_{51}(2) = 0.790691 \times 0.6845 - 0.000017 = 0.5412$$

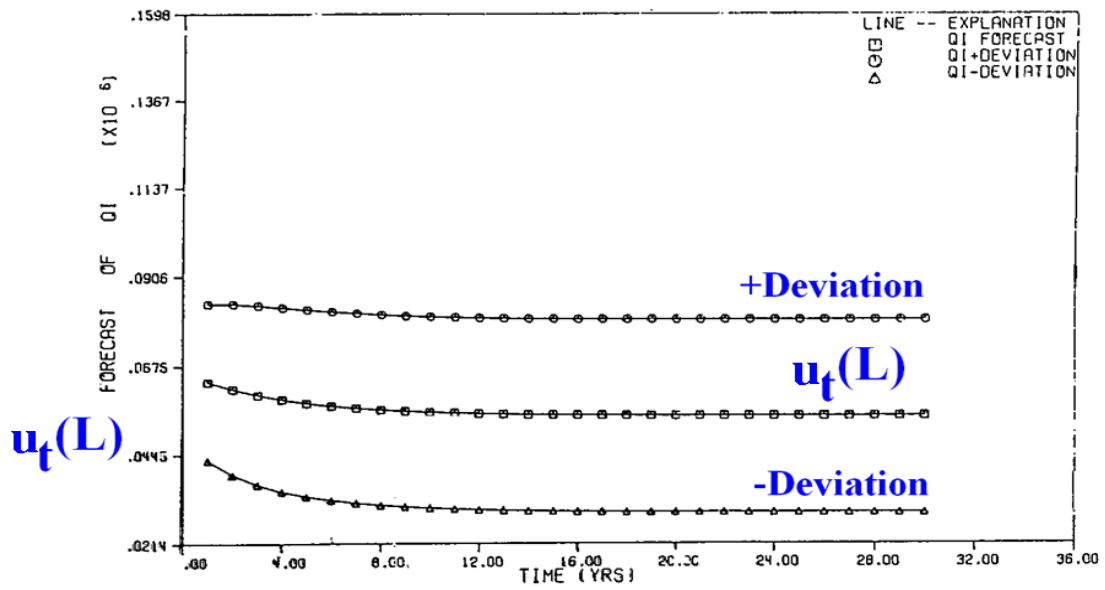
$$z_{51}(3) = 0.790691 \times 0.5412 - 0.000017 = 0.4279$$

$$u_t(L) = z_t(L) \times 13167.569 + 54362$$

ผลการพยากรณ์ $Z_{51}(L)$ ของแม่น้ำ Niger แสดงอยู่ในรูปที่ 5.12 และผลการพยากรณ์ค่า $q_t(L)$ ของแม่น้ำ Niger แสดงอยู่ในรูปที่ 5.13

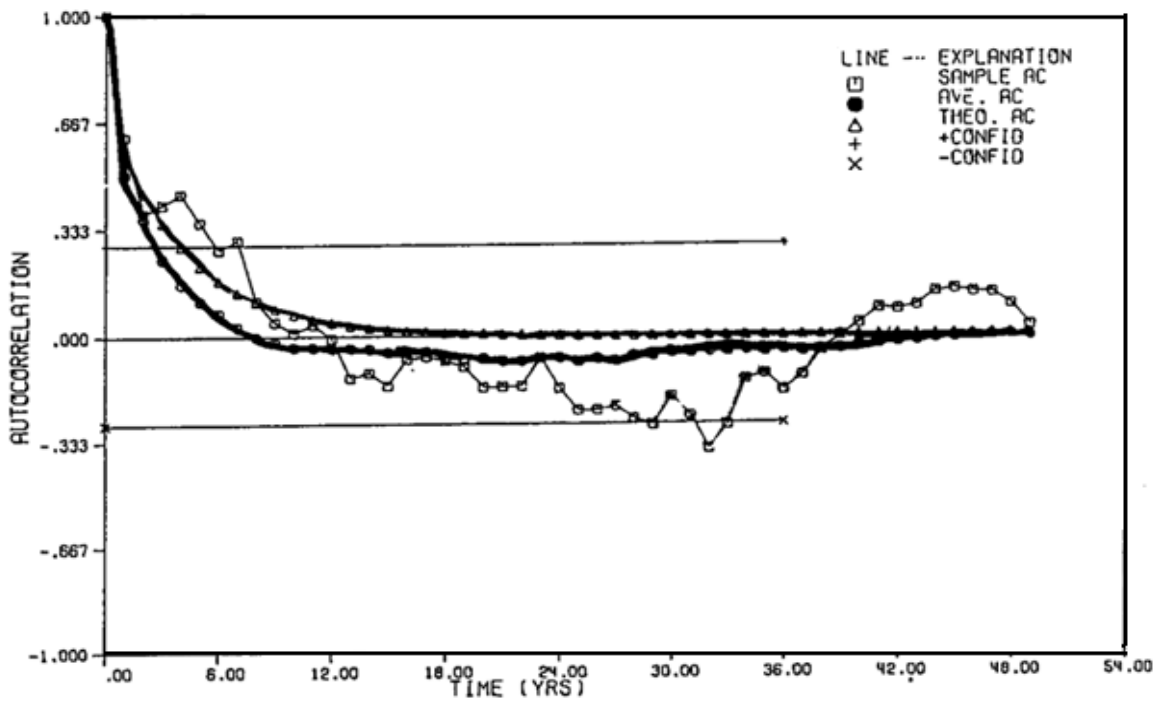


รูปที่ 5.12 Forecast of z_t Series , Niger River



รูปที่ 5.13 Forecasts of Annual Flows, Niger River

รูปที่ 5.14 แสดงการเปรียบเทียบค่า ACF ของแบบจำลอง ACF เล็กๆ ของอนุกรมเวลาที่สังเคราะห์ใหม่ 100 ชุด และ ACF ของอนุกรมเวลาตัวอย่าง



รูปที่ 5.14 Sample ACF Series, Average ACF of 100 Series and Theoretical ACF

การคำนวณ Ψ_j Weights จากสมการ 5.75

$$\psi_j = (\phi_1 - \theta_1)\phi_1^{j-1}$$

ตัวอย่างการคำนวณ Ψ_j

$$\psi_0 = 1$$

$$\psi_1 = (0.790691 - 0.348173)0.790691^0 = 0.44252$$

$$\psi_2 = (0.790691 - 0.348173)0.790691 = 0.34989$$

การคำนวณช่วงความเชื่อมั่น $(1-\alpha)$ CI ของ Min. MSE of Forecast, $\hat{z}_t(L)$ โดยสมการ 5.77

$$z_{t+L} = z_t(L) \pm u_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{1 + \sum_{j=1}^{L-1} \psi_j^2}$$

$$\pm u_{0.05/2} = \pm 1.96$$

$$\hat{\sigma}_\varepsilon^2 = 0.627668$$

$$\hat{\sigma}_\varepsilon = 0.79226$$

L	$\pm u_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{1 + \sum_{j=1}^{L-1} \psi_j^2}$
1	$= 1.96(0.79226)\sqrt{1 + 0} = 1.5528$
2	$= 1.96(0.79226)\sqrt{1 + \psi_1^2} = 1.96(0.79226)\sqrt{1 + 0.1958} = 1.6981$
3	$= 1.96\sqrt{1 + (0.1958 + 0.1224)}(0.79226) = 1.7829$

โปรแกรม FTCMIP สามารถคำนวณค่า $\pm u_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{1 + \sum_{j=1}^{L-1} \psi_j^2}$ ได้ดังแสดงในตารางที่ 5.11 ในหนังสือของ Salas *et. al.*(1980)

ช่วงความเชื่อมั่น 95% ของค่าพยากรณ์ $Z_{t+L}(\pm)$ ถูกพล็อตเปรียบเทียบกับค่าพยากรณ์ $Z_t(L)$ ดังแสดงอยู่ในรูปที่ 5.12 และช่วงความเชื่อมั่น 95% ของค่าพยากรณ์อัตราการใช้ไฟฟ้ารายปี $u_{t+L}(+)$ ถูกพล็อตเปรียบเทียบกับค่าพยากรณ์ $u_t(L)$ ดังแสดงอยู่ในรูปที่ 5.13 กำหนดให้

$$u_{t+L}(\pm) = Z_{t+L}(\pm) * S_u + \hat{u}$$

การพยากรณ์แบบ Real Time (Real Time Forecasting)

(1) ใช้ z_t เมื่อ $t = 1, \dots, 30$ (จากอนุกรมซึ่งมี $N = 51$) เพื่อคำนวณหาพารามิเตอร์แบบจำลองเบื้องต้น (Initial Model Parameters)

(2) คำนวณหาค่าพยากรณ์ (Forecasted Value) ที่

$$L = 1, t=30$$

$$L \leq q = 1$$

$$\begin{aligned} z_t(1) &= z_{t+1} - \varepsilon_{t+1} \\ &= (\hat{\phi}_1 z_t + \varepsilon_{t+1} - \hat{\theta}_1 \varepsilon_t) - \varepsilon_{t+1} \\ &= \hat{\phi}_1 z_t - \hat{\theta}_1 \varepsilon_t \end{aligned}$$

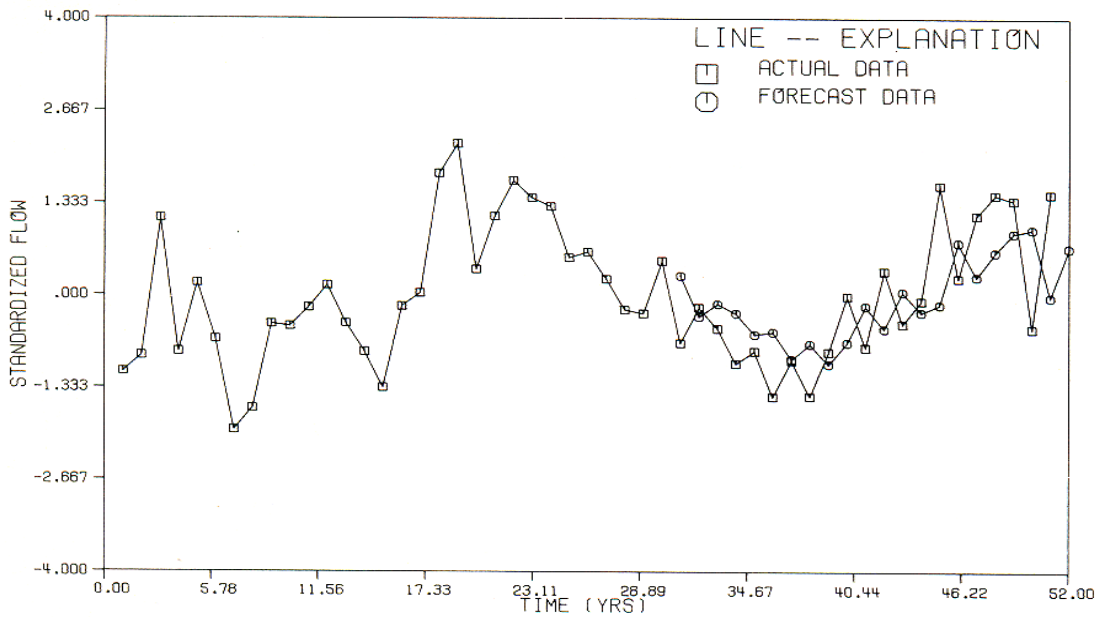
$$z_{30}(1) = \hat{\phi}_1 z_{30} - \hat{\theta}_1 \varepsilon_{30}$$

(3) เมื่อทราบค่า z_{31} แล้ว ทำการคำนวณหาพารามิเตอร์ใหม่จาก z_1, \dots, z_{31} หลังจากนั้นจึงหาค่าพยากรณ์ (Forecasted Value) ที่ $L = 1$ ดังนี้

$$z_{31}(1) = \hat{\phi}_1 z_{31} - \hat{\theta}_1 \varepsilon_{31}$$

หมายเหตุ: $\hat{\phi}_1$ และ $\hat{\theta}_1$ จะเปลี่ยนค่าตามอนุกรมเวลา z_t ที่เปลี่ยนไป

ผลการพยากรณ์แบบ Real Time ล่วงหน้า 1 คาบ หรือ $z_t(1)$ ที่ $t = 30, \dots, 51$ แสดงอยู่ในรูปที่ 5.15 โดยการเปรียบเทียบกับค่าจริง z_{t+1} ที่เกิดขึ้น และค่าพารามิเตอร์ในการพยากรณ์แบบ Real Time ซึ่งคำนวณโดยโปรแกรม FTCMP แสดงอยู่ในตารางที่ 5.16



รูปที่ 5.15 การพยากรณ์แบบ Real Time เปรียบเทียบกับค่าที่เกิดขึ้นจริง

ตารางที่ 5.16 ค่าพารามิเตอร์และผลการพยากรณ์แบบ Real Time ที่ $t = 30$ โดยโปรแกรม FTCMP

Year (L)	ARPS(ϕ_1)	PMAC(θ_{00})	PMAS(θ_1)	WNV	FORECAST $Z_t(L)$	ACTUAL Z_t	t
1	0.64709	0.22970	0.09899	0.61940	0.28173	-0.69358	31
2	0.66179	0.01374	0.14080	0.62997	-0.31199	-0.18165	32
3	0.66366	0.01133	0.14750	0.61098	-0.12819	-0.48303	33
4	0.67100	0.00593	0.15486	0.59659	-0.26454	-0.98671	34
5	0.67989	-0.00369	0.15102	0.59455	-0.56810	-0.80918	35
6	0.68569	-0.01079	0.14838	0.57942	-0.53231	-1.46561	
7	0.71851	-0.02085	0.16758	0.58689	-0.92385	-0.94129	
8	0.72000	-0.02731	0.16830	0.57117	-0.70370	-1.46148	
9	0.75393	-0.03283	0.19729	0.57056	-0.99393	-0.81744	
10	0.74869	-0.03794	0.19705	0.55679	-0.68424	-0.01651	
11	0.72065	-0.04123	0.16769	0.55346	-0.16035	-0.75551	
12	0.73932	-0.04234	0.20159	0.54845	-0.48652	0.35092	
13	0.73301	-0.04010	0.20883	0.55167	-0.04456	-0.41698	
14	0.74025	-0.04062	0.22786	0.54378	-0.24718	-0.08257	
15	0.74086	-0.04010	0.23010	0.53200	-0.13890	1.58946	
16	0.73325	-0.03093	0.21816	0.58499	0.76057	0.24771	
17	0.73393	-0.02875	0.24686	0.57723	0.26834	1.15597	
18	0.75955	-0.01952	0.27520	0.58031	0.62239	1.45322	
19	0.77424	-0.01111	0.27123	0.58100	0.89806	1.37478	
20	0.78550	-0.00432	0.27054	0.57252	0.95484	-0.48303	
21	0.73519	-0.00778	0.24702	0.60037	-0.02742	1.46561	
22	0.79067	-0.00002	0.34816	0.62767	0.68425		

ตัวอย่างการคำนวณการพยากรณ์แบบ Real Time

(1) กำหนดอนุกรมเวลา Z_1, \dots, Z_{30} ค่าพารามิเตอร์

$$\hat{\phi}_1 = 0.64709$$

$$\hat{\theta}_{00} = 0.02297$$

$$\hat{\theta}_1 = 0.09899$$

$$Z_{30} = 0.495417 \text{ (จากตารางที่ 5.2)}$$

$$\varepsilon_{30} = 0.612344 \text{ (จากตารางที่ 5.11)}$$

$$Z_{30}(1) = 0.64709 \times 0.495417 - 0.09899 \times 0.612344 + 0.02297 = 0.2829$$

$$Z_{30+1} = Z_{31} \text{ (จากตารางที่ 5.2)} = -0.693584$$

(2) จาก Z_1, \dots, Z_{31} ค่าพารามิเตอร์

$$z_{31}(1) = 0.66179 \times (-0.69358) - 0.14080(-0.872161) + 0.01374 = 0.32246$$

$$Z_{31+1} = Z_{32} = -0.181653$$

(3) จาก Z_1, \dots, Z_{32} ค่าพารามิเตอร์

$$Z_{32}(1) = 0.66366(-0.181653) - 0.1475(0.063208) + 0.01133 = -0.118549$$

$$Z_{32+1} = Z_{33} = -0.483032$$

เมื่อตรวจสอบ ACF ของอนุกรมเวลาสังเคราะห์ 100 ชุด แต่ละชุดมีความยาว 51 ปี เท่ากับอนุกรมเวลาตัวอย่าง พบว่าค่าเฉลี่ยของ ACF ของอนุกรมเวลา 100 ชุด ใกล้เคียงกับ ACF ของแบบจำลองและ ACF ของอนุกรมเวลาตัวอย่าง ดังรูปที่ 5.14

5.9 การจำลองด้วย ARMA สำหรับอนุกรมเวลาแบบ Periodic (ARMA Modeling of periodic Time Series)

$$\text{ให้ } y_{v,\tau} = f(x_{v,\tau})$$

$$z_{v,\tau} = \frac{y_{v,\tau} - \mu_\tau}{\sigma_\tau} \quad (\text{การกำจัด Within the Year Periodicity}) \quad [5.98]$$

กรณี ARMA (p, q) with Constant Coefficient (ACF Stationary)

$$z_t = \sum_{j=1}^p \phi_j z_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}; t = (v-1)\omega + \tau \quad [5.99]$$

กรณี ARMA (p, q) with Periodic Coefficient (ACF Non-stationary)

$$z_{v,\tau} = \sum_{j=1}^p \phi_{j,\tau} z_{v,\tau-j} + \varepsilon_{v,\tau} - \sum_{j=1}^q \theta_{j,\tau} \varepsilon_{v,\tau-j} \quad [5.100]$$

5.9.1 ขั้นตอนการจำลอง

(1) คำนวณหาค่า $r_{k,\tau}$ of $z_{v,\tau}$ จากสมการ 2.10

$$r_{k,\tau} = \frac{\frac{1}{N} \sum_{v=1}^N (x_{v,\tau} - \bar{x}_\tau)(x_{v,\tau-k} - \bar{x}_{\tau-k})}{S_\tau \cdot S_{\tau-k}}$$

(2) ประมาณค่า $\phi_{j,\tau}$ และ $\theta_{j,\tau}$ จากสมการ Yule-Walker

Initial Estimates สำหรับ ARMA(p,q)

$$\text{ARMA}(1, 0) : \hat{\phi}_{1,\tau} = r_{1,\tau} \quad [5.101]$$

$$\text{ARMA}(2, 0) : \hat{\phi}_{1,\tau} = \frac{r_{1,\tau} - r_{1,\tau-1} \cdot r_{2,\tau}}{1 - r_{1,\tau-1}^2} \quad [5.102]$$

$$\text{ARMA}(1, 1) : r_{2,\tau} = r_{1,\tau} \hat{\phi}_{1,\tau} \quad [5.103]$$

$$r_{1,\tau} = \frac{(1 - \hat{\theta}_{1,\tau} \hat{\phi}_{1,\tau})(\hat{\phi}_{1,\tau} - \hat{\theta}_{1,\tau})}{1 + \hat{\phi}_{1,\tau}^2 - 2\hat{\phi}_{1,\tau} \hat{\theta}_{1,\tau}} \quad [5.104]$$

Fourier Series Fit

$$\hat{\sigma}_{\varepsilon_\tau}^2 = \frac{(1 - \hat{\phi}_{1,\tau}^2)}{1 + \hat{\theta}_{1,\tau}^2 - 2\hat{\phi}_{1,\tau} \hat{\theta}_{1,\tau}} \quad [5.105]$$

(3) Goodness of Fit Test

Modified Porte Manteau (Tao and Dollar, 1976)

$$Q_1 = N \sum_{k=1}^L \sum_{\tau=1}^{\omega} r_{k,\tau}^2 \quad [5.106]$$

(4) Generation and forecasting

5.9.2 ตัวอย่างการจำลองอนุกรมเวลาแบบ Periodic ด้วย ARMA with Constant Coefficients

ให้ $x_{v,\tau}$ คืออนุกรมเวลารายของฝนรายเดือนที่สถานี Salamonia มลรัฐ Indiana (ดูตารางที่ 5.17)

$$v = 1, \dots, 38$$

$$\tau = 1, \dots, 12$$

ขั้นที่ 1 Normalization

$$y_{v,\tau} = \sqrt{x_{v,\tau}} \text{ (ดูตารางที่ 5.18)}$$

$$\bar{y}_1 = 1.5145; s_1 = 0.5660$$

.

.

$$\bar{y}_{12} = 1.6540; s_{12} = 0.5402$$

ขั้นที่ 2 Standardization

$$z_{v,\tau} = \frac{y_{v,\tau} - \bar{y}_\tau}{s_\tau}$$

$$z_{1,1} = \frac{\sqrt{3.920} - 1.5145}{0.5660} = 0.8222$$

ขั้นที่ 3 คำนวณหา ACF และ Partial ACF

$$\left. \begin{array}{l} r_k(z) \\ \hat{\phi}_k(k) \end{array} \right\} k = 1, \dots, 29 \text{ (ตารางที่ 5.19)}$$

ขั้นที่ 4 เลือกแบบจำลอง ARMA(1, 1) with Constant Coefficients

ขั้นที่ 5 การประมาณค่าพารามิเตอร์ $\hat{\phi}_1$ เบื้องต้น (Initial Estimates)

$$\hat{\phi}_1 = \frac{r_2}{r_1} = \frac{0.0179}{0.0288} = 0.62$$

ขั้นที่ 6 การประมาณค่าพารามิเตอร์ $\hat{\theta}_1$ เบื้องต้น (Initial Estimates)

$$\hat{\sigma}_\varepsilon^2 = \frac{c'_0}{1 + \hat{\theta}_1^2}$$

$$\hat{\theta}_1 = -\frac{c'_0}{\hat{\sigma}_\varepsilon^2}$$

ขั้นที่ 7 การประมาณค่าพารามิเตอร์โดยวิธี MLE

$$\text{Min } S(\hat{\phi}_1, \hat{\theta}_1) = \sum_{t=1}^N \varepsilon_t^2$$

$$|\hat{\phi}_1| < 1$$

$$|\hat{\theta}_1| < 1$$

Min S(0.9,0.9) (ตารางที่ 5.20)

ค่าประมาณสุดท้ายจากโปรแกรม FTMXL

$$\hat{\phi}_1 = 0.914; \hat{\theta}_1 = 0.961; \hat{\sigma}_\varepsilon^2 = 0.961$$

ขั้นที่ 8 การทดสอบว่า ε_t เป็นอนุกรมเวลาแบบ Independent โดยใช้ Modified Porte Manteau

$$Q_1(= 28.61) < \chi_{24-1-1,0.10}^2(= 30.8)$$

แสดงว่า ε_t เป็น Independent

ตารางที่ 5.17 Precipitation (inches) at Salamonia, Indiana.

YEAR	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1	3.920	4.910	3.850	8.210	2.690	4.050	3.680	1.460	4.020	4.340	5.070	7.970	54.170
2	3.900	2.710	7.960	7.160	2.150	7.100	6.260	9.690	0.360	1.570	2.010	4.840	55.710
3	2.680	1.230	3.510	1.470	1.040	4.070	1.360	1.220	7.910	1.920	5.870	4.960	37.240
4	0.320	2.200	2.480	3.990	0.830	9.800	2.560	7.000	5.230	1.780	3.450	2.260	41.900
5	1.660	2.820	1.760	1.960	1.380	4.110	4.600	0.800	1.220	2.600	0.990	2.780	26.680
6	7.490	4.500	3.130	7.750	2.440	1.110	4.950	4.470	4.440	0.820	4.600	2.960	48.660
7	5.880	1.260	2.880	2.410	1.970	7.480	1.660	4.700	3.290	9.380	3.180	3.240	47.330
8	0.500	3.370	1.970	4.440	4.250	5.680	5.570	1.310	5.140	5.480	1.680	1.060	40.450
9	0.770	1.390	1.370	2.780	4.430	1.800	6.930	4.620	6.290	1.530	2.000	1.530	35.440
10	1.060	4.240	2.460	1.860	0.580	0.520	2.730	0.660	5.960	4.590	3.310	1.630	29.600
11	6.560	1.490	2.360	2.710	3.270	4.080	2.680	4.990	6.650	4.330	3.280	1.530	43.930
12	1.580	5.330	4.430	0.650	1.630	6.410	3.710	4.300	1.620	3.050	0.430	2.810	35.950
13	1.940	1.080	1.790	1.070	3.790	5.900	5.850	4.600	0.510	3.060	7.730	4.000	41.320
14	0.250	2.990	3.370	1.750	6.720	4.400	4.650	3.700	7.790	3.660	2.360	5.970	47.610
15	3.180	4.160	2.270	2.100	4.630	3.560	2.190	8.510	3.530	5.690	3.630	0.560	44.010
16	2.940	4.580	3.350	4.460	0.380	1.850	8.720	5.150	5.320	5.950	4.000	2.980	49.680
17	1.980	2.290	1.120	3.070	3.960	6.670	6.040	3.940	3.740	0.520	2.130	2.160	37.620
18	2.360	7.800	5.340	0.300	3.220	5.940	1.510	3.980	9.580	3.300	6.780	2.720	52.840
19	4.440	1.090	4.930	2.560	6.850	3.790	5.870	4.990	6.380	3.520	3.910	3.460	51.790
20	2.530	6.640	1.550	7.300	0.000	6.180	3.130	2.910	4.430	1.650	2.670	3.070	42.060
21	1.510	0.000	5.800	3.840	3.250	5.210	3.210	2.670	3.500	2.440	1.620	3.740	36.790
22	1.580	2.170	5.320	4.730	0.730	3.380	3.580	3.990	2.230	7.090	1.310	0.160	36.270
23	1.140	1.090	2.110	2.920	2.970	1.730	3.530	1.950	2.590	1.360	5.510	3.960	30.860
24	3.860	1.230	3.070	2.290	8.370	5.640	2.390	2.910	3.780	4.800	1.200	5.830	45.370
25	5.710	3.270	0.750	2.080	7.350	3.470	4.140	3.340	4.890	3.790	3.400	2.250	44.440
26	2.660	1.560	2.460	3.900	3.730	1.620	6.790	6.340	8.010	4.030	2.830	2.950	46.880
27	2.860	7.550	6.410	2.370	0.630	2.060	3.000	4.850	4.450	7.730	3.760	6.440	52.110
28	1.200	2.860	1.140	8.290	3.200	2.930	1.310	3.770	2.960	2.750	3.010	1.500	34.920
29	4.000	2.390	3.840	2.600	3.610	0.970	1.590	3.970	9.230	4.150	2.410	0.970	39.730
30	2.450	2.910	1.940	1.890	3.590	6.260	3.550	8.330	4.020	4.160	4.270	0.950	44.320
31	2.800	3.720	3.930	4.290	6.170	4.020	0.790	4.460	2.660	4.110	2.820	2.240	42.010
32	2.290	1.960	2.410	1.320	0.790	0.080	1.980	2.790	2.730	5.860	2.580	1.590	26.380
33	0.160	1.260	0.910	1.850	2.410	4.530	6.980	1.950	8.010	2.700	2.180	2.050	34.990
34	0.510	2.550	5.590	2.900	4.590	3.340	4.770	2.250	2.910	7.030	2.610	7.430	46.480
35	3.360	1.320	1.250	5.070	4.860	1.870	7.400	2.530	1.240	4.270	4.290	2.870	40.330
36	1.760	3.600	3.620	1.440	1.480	3.980	4.650	6.280	3.570	2.830	4.390	1.010	38.610
37	3.890	4.960	5.690	1.810	0.520	5.080	3.870	9.530	1.650	7.200	2.940	4.140	51.280
38	1.340	4.330	5.450	7.510	1.200	1.760	4.320	4.190	5.800	5.410	3.130	2.180	46.620
MEAN	2.606	3.021	3.252	3.397	3.044	4.011	4.013	4.187	4.412	3.959	3.246	3.020	42.168
STDV	1.770	1.846	1.754	2.199	2.120	2.176	1.953	2.259	2.376	2.047	1.554	1.866	7.453

ตารางที่ 5.18 Standard Scores of Square Roots of Precipitation (inches) at Salamonia, Indiana.

YEAR	1	2	3	4	5	6	7	8	9	10	11	12
1	0.8222	1.0147	0.4603	1.9236	0.0329	0.1667	-0.0449	-1.3554	-0.0133	0.3082	1.1500	2.1642
2	0.8132	-0.0036	2.2267	1.5960	-0.2303	1.2422	1.1161	2.0352	-2.3261	-1.2494	-0.7589	1.0108
3	0.2165	-0.9638	0.2781	-0.9338	-0.9065	0.1749	-1.5410	-1.5401	1.3159	-1.0005	1.5418	1.0610
4	-1.6764	-0.2949	-0.3360	0.4234	-1.0712	2.0106	-0.6781	1.2036	0.4508	-1.0971	0.2475	-0.2788
5	-0.3995	0.0555	-0.8461	-0.6095	-0.6718	0.1912	0.4055	1.9142	1.4956	-0.5751	-1.7267	0.0248
6	2.1594	0.8457	0.0636	1.7828	-0.0853	-1.4147	0.5648	0.2574	0.1548	-1.9012	0.9053	0.1231
7	1.6083	-0.9398	-0.0847	-0.3460	-0.3253	1.3583	1.2979	0.3530	-0.3280	2.1455	0.0778	0.2703
8	-1.4265	0.3352	-0.6880	0.6129	0.6712	0.7783	0.8338	-1.4689	0.4182	0.7916	-1.0373	-1.1558
9	-1.1255	-0.8388	-1.1672	-0.1474	0.7366	-0.9396	1.3757	0.3200	0.8147	-1.2795	-0.7670	-0.7720
10	-0.8568	0.7346	-0.3491	-0.6721	-1.2975	-1.9630	-0.5741	-2.0602	0.7049	0.4192	0.1604	-0.6983
11	1.8492	-0.7643	-0.4153	-0.1839	0.2876	0.1790	-0.6043	0.4703	0.9312	0.3037	0.1415	-0.7720
12	-0.4550	1.1807	0.7535	-1.6360	-0.5174	1.0232	-0.0294	0.1851	-1.2186	-0.3237	-2.5032	0.0414
13	-0.2150	-1.0886	-0.8230	-1.2415	0.4974	0.8537	0.9504	0.3117	-2.1382	-0.3184	2.3601	0.6406
14	-1.7924	0.1447	0.2005	-0.7429	1.4749	0.3072	0.4286	-0.0821	1.2806	-0.0110	-0.4877	1.4613
15	0.4747	0.6997	-0.4761	-0.5246	0.8077	-0.0405	-0.9170	1.6868	-0.2210	0.8749	0.3571	-1.6764
16	0.3535	0.8793	0.1893	0.6211	-1.5173	-0.9091	2.0131	0.5335	0.4830	0.9760	0.5739	0.1339
17	-0.1898	-0.2412	-1.3978	-0.0008	0.5628	1.1071	1.0279	0.0272	-0.1303	-2.2472	-0.6634	-0.3411
18	0.0383	2.0462	1.1772	-2.0829	0.2666	0.8672	-1.4164	0.0451	1.7839	-0.1921	1.9562	-0.0087
19	1.0469	-1.0801	0.9912	-0.2638	1.5127	0.0584	0.9586	0.4703	0.8441	-0.0803	0.5221	0.3816
20	0.1344	1.6600	-1.0140	1.6410	-2.4509	0.9476	-0.3416	-0.4696	0.1509	-1.1902	-0.2639	0.1817
21	-0.5048	-2.9464	1.3776	0.3579	0.2793	0.6121	-0.2969	-0.5975	-0.2342	-0.6696	-1.0907	0.5182
22	-0.4550	-0.3131	1.1683	0.7300	-1.1570	-0.1202	-0.0971	0.0495	-0.8556	1.3952	-1.3843	-2.3212
23	-0.7894	-1.0801	-0.5872	-0.0757	0.1590	-0.9831	-0.1235	-1.0205	-0.6646	-1.4122	1.3690	0.6220
24	0.7953	-0.9638	0.0286	-0.4137	1.9304	0.7644	-0.7856	-0.4696	-0.1134	0.5101	-1.4967	1.4079
25	1.5459	0.2862	-1.7931	-0.5365	1.6548	-0.0801	0.1866	-0.2530	0.3263	0.0521	0.2166	-0.2850
26	0.2057	-0.7137	-0.3491	0.3842	0.4739	-1.0532	1.3225	0.9760	1.3451	0.1660	-0.1534	0.1177
27	0.3120	1.9655	1.6314	-0.3684	-1.2488	-0.7852	-0.4151	0.4141	0.1587	1.6158	0.4345	1.6359
28	-0.7404	0.0768	-1.3785	1.9477	0.2582	-0.3293	-1.5840	-0.0499	-0.4817	-0.4890	-0.0328	-0.7945
29	0.8577	-0.1828	0.4551	-0.2423	0.4265	-1.5280	-1.3525	0.0406	1.6873	0.2217	-0.4506	-1.2385
30	0.0896	0.1031	-0.7101	-0.6531	0.4185	0.9741	-0.1129	1.6316	-0.0133	0.2263	0.7259	-1.2574
31	0.2805	0.5015	0.5020	0.5509	1.3108	0.1544	-2.0927	0.2531	-0.0629	0.2032	-0.1602	-0.2912
32	-0.0022	-0.4437	-0.3820	-1.0436	-1.1048	-2.6858	-1.0617	-0.5329	-0.5939	0.9413	-0.3274	-0.7275
33	-1.9691	-0.9398	-1.6123	-0.6784	-0.0999	0.3579	1.3945	-1.0205	1.3451	-0.5174	-0.6245	-0.4113
34	-1.4141	-0.0918	1.2871	-0.0858	0.7936	-0.1382	0.4836	-0.8361	-0.5057	1.3740	-0.3061	1.9841
35	0.5627	-0.8926	-1.2750	0.8628	0.8877	-0.8970	1.5503	-0.6748	-1.4807	0.2765	0.7370	0.0743
36	-0.3320	0.4454	0.3380	-0.9553	-0.6085	0.1379	0.4286	0.5548	-0.2035	-0.4441	0.7919	-1.2013
37	0.8087	1.0348	1.3304	-0.7040	-1.3588	0.5649	0.0524	1.9892	-1.1993	1.4338	-0.0793	0.7048
38	-0.6306	0.7734	1.2259	1.7077	-0.7919	-0.9643	0.2736	0.1376	0.6506	0.7634	0.0456	-0.3285
MEAN	1.5145	1.6482	1.7382	1.7526	1.6184	1.9114	1.9409	1.9697	2.0131	1.9190	1.7493	1.6540
STDV	0.5660	0.5594	0.4864	0.5785	0.6603	0.6063	0.5027	0.5617	0.6075	0.5331	0.4368	0.5402

ตารางที่ 5.19 Autocorrelation and Partial Autocorrelation Functions, and Partial and Final Parameter Estimates for the Standard Scores of Square Roots of Precipitation (in) at Salomonina , Indians.

Lag	Autocorrelation	Partial Autocorrelation
0	1.0000	1.0000
1	0.0288	0.0288
2	0.0179	0.0171
3	-0.0517	-0.0528
4	0.0168	0.0196
5	-0.0199	-0.0192
6	0.0701	0.0682
7	0.1144	0.1137
8	0.0107	-0.0004
9	0.1394	0.1460
10	-0.0079	-0.0068
11	0.0279	0.0252
12	-0.0844	-0.0729
13	0.0145	-0.0048
14	-0.0250	-0.0299
15	0.0091	-0.0219
16	-0.0067	-0.0334
17	0.0546	0.0480
18	-0.0376	-0.0566
19	0.0052	0.0256
20	0.0090	0.0101
21	-0.0218	0.0001
22	-0.0556	-0.0515
23	-0.0505	-0.0420
24	-0.0699	-0.0836
25	0.0083	0.0248
26	0.0260	-0.0022
27	0.0190	0.0282
28	0.0390	0.0435
29	-0.0175	0.0082

Parameter Estimates ARMA(1,1)

	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\sigma}_\epsilon^2$	$S(\hat{\phi}_1, \hat{\theta}_1)$	Q^*
Preliminary	0.62	0.59	0.974	-	-
Sum of Squares Surface	0.9	0.9	-	439.0	-
FTMXL	0.914	0.961	0.961	-	28.60

ตารางที่ 5.20 Sum of Squares of Residuals of ARMA(1,1) for Standardized Monthly Rainfall Series
at Salamonina, Indiana.

THETA (θ)	PHI(ϕ)										
	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0.00	443.32	445.21	455.97	475.60	504.12	541.51	587.79	642.94	706.96	779.87	861.65
0.10	450.32	443.15	445.01	455.89	475.78	504.70	542.65	589.61	645.59	710.60	784.63
0.20	467.02	450.31	442.96	444.98	456.35	477.09	507.18	546.64	595.46	653.64	721.18
0.30	495.25	467.81	450.31	442.74	445.10	457.39	479.62	511.79	553.88	605.91	667.87
0.40	539.08	498.69	469.13	450.38	442.46	445.36	459.08	483.62	518.98	565.16	622.16
0.50	607.12	549.73	504.53	471.53	450.73	442.13	445.73	461.52	489.52	529.71	582.10
0.60	718.93	636.58	568.70	515.27	476.30	451.79	441.75	446.16	465.04	498.37	546.17
0.70	924.74	799.80	693.49	605.80	536.73	486.29	454.47	441.28	446.71	470.76	513.44
0.80	1383.05	1167.56	979.97	820.29	688.52	584.65	508.70	460.65	440.50	448.27	483.94
0.90	2849.93	2351.12	1910.04	1526.70	1201.08	933.20	723.05	570.64	475.95	439.00	459.78
1.00	14000.33	11400.10	9083.29	7049.91	5299.96	3833.44	2650.34	1750.67	1134.43	801.61	752.23

ขั้นที่ 9 การพยากรณ์ (Forecasting)

$$[z_{t+L}] = z_t(L) = \hat{\phi}_1[z_{t+L-1}] + \varepsilon_{t+L} - \hat{\theta}_1[\varepsilon_{t+L-1}]$$

$$z_t(1) = \hat{\phi}_1 z_t - \hat{\theta}_1 \varepsilon_t; [\varepsilon_{t+1}] = 0$$

$$z_t(2) = \hat{\phi}_1 z_t(1)$$

$$z_t(3) = \hat{\phi}_1 z_t(2)$$

.

$$z_t(L) = \hat{\phi}_1 z_t(L-1); L \geq 2$$

การคำนวณ Ψ -weights (สมการที่ 5.75)

$$\psi_0 = 1$$

$$\psi_1 = \hat{\phi}_1 - \hat{\theta}_1$$

$$\psi_2 = (\hat{\phi}_1 - \hat{\theta}_1)\hat{\phi}_1$$

.

$$\psi_j = (\hat{\phi}_1 - \hat{\theta}_1)\hat{\phi}_1^{j-1}$$

การคำนวณ Standard Error of Forecasts

$$\begin{aligned} \sigma(L) &= \sqrt{1 + \sum_{j=1}^{L-1} (\hat{\phi}_1 - \hat{\theta}_1)^2 \hat{\phi}_1^{2(j-1)} \hat{\sigma}_\varepsilon} \\ &= \sqrt{1 + \sum_{j=1}^{L-1} \psi_j^2 \hat{\sigma}_\varepsilon} \end{aligned}$$

$$z_{t+L}(\pm) = z_t(L) \pm u_{\alpha/2} \sigma(L)$$

5.9.3 ตัวอย่างการจำลองอนุกรมเวลาแบบ Periodic ด้วย ARMA with Periodic Coefficients

ให้ทำการจำลองอนุกรมเวลาน้ำท่ารายวันของแม่น้ำ Salomonیا ที่ Dora มลรัฐ Indiana

ขั้นที่ 1 Normalization โดยใช้ Log

$$y_{v,\tau} = \log(x_{v,\tau} - c_\tau)$$

ขั้นที่ 2 คำนวณหา $\hat{\mu}_\tau$ และ $\hat{\sigma}_\tau$ โดยใช้ Fourier Series ที่ 6 Harmonics (รูปที่ 5.16)

ขั้นที่ 3 คำนวณหา $r_{k,\tau}$ (สมการที่ 2.10) และ $\hat{\rho}_{k,\tau}$ โดยใช้ Fourier Series ที่ 6 Harmonics

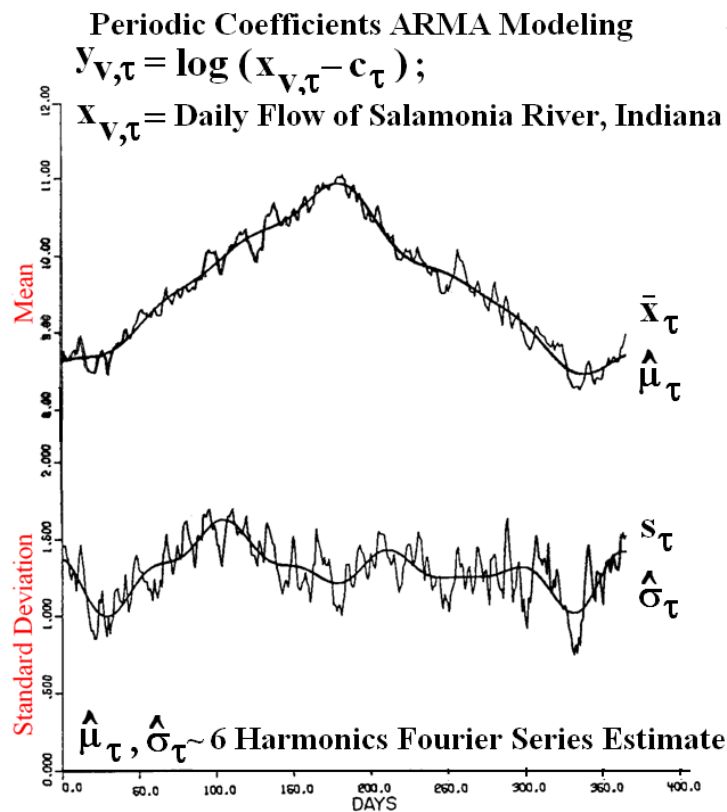
เมื่อ $k = 1, 2, 3$ (รูปที่ 5.17)

ขั้นที่ 4 และ 5 คำนวณหา $\hat{\phi}_{j,\tau}$ และ $\hat{\theta}_{j,\tau}$ (สมการที่ 5.103 และ 5.104) และ ประมาณค่าด้วย 6 Harmonics Fourier Series (รูปที่ 5.18)

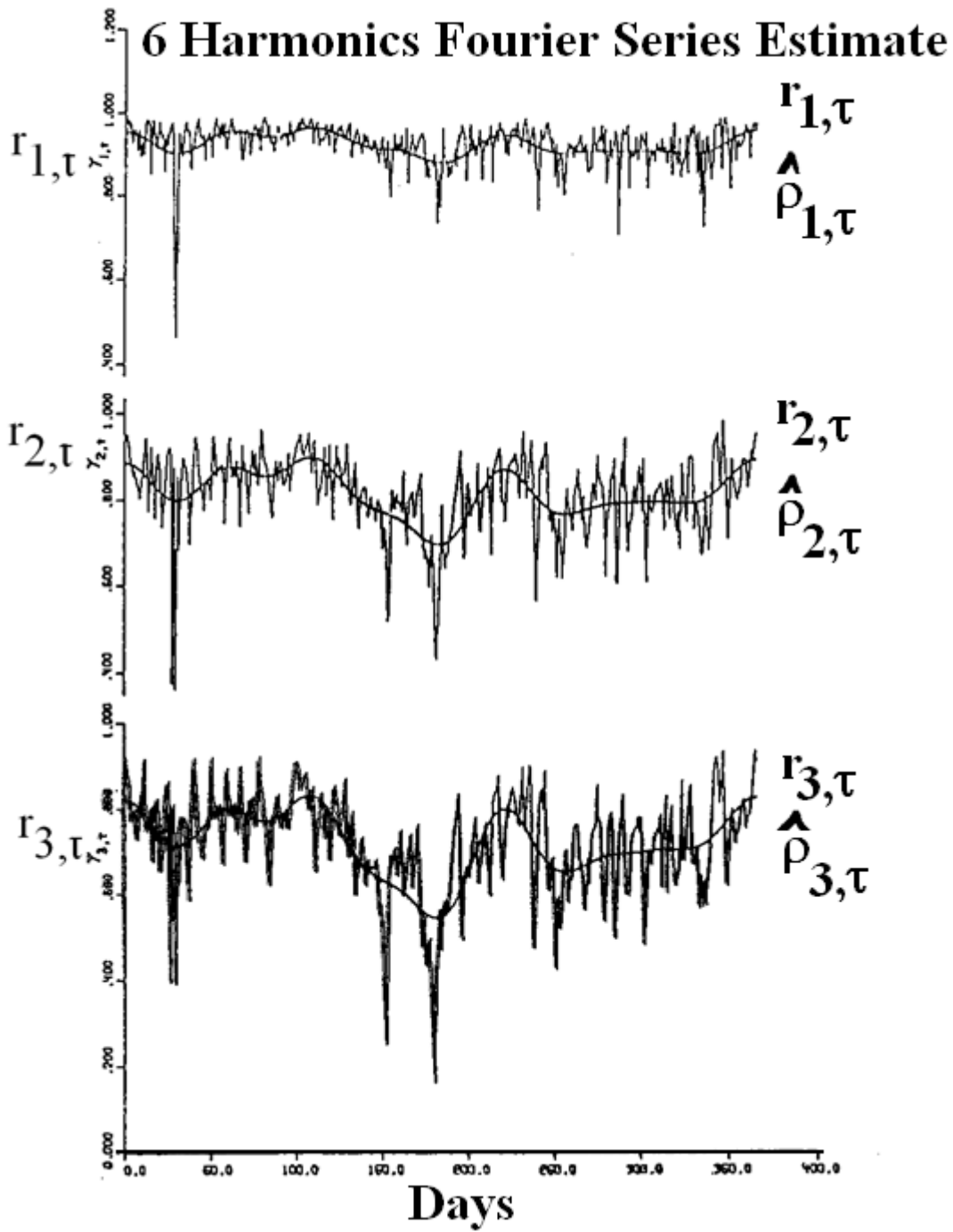
ขั้นที่ 6 คำนวณหา $r_{k,\tau}(\epsilon)$ และ Q_1 Statistics

ผลการคำนวณ Cumulative Periodogram ของ ϵ_t ของ ARMA(1, 1) แสดงอยู่ในรูปที่ 5.19 ซึ่งสามารถสรุปได้ว่า ϵ_t เป็น Independent

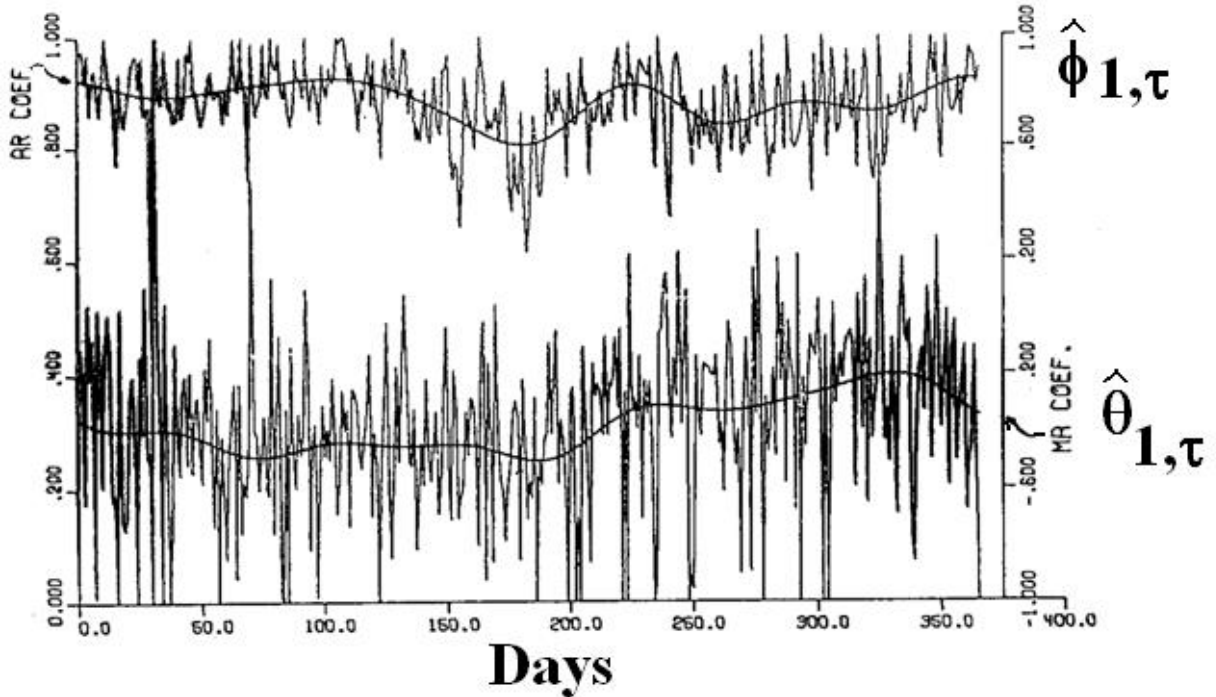
รูปที่ 5.20 แสดงการแจกแจงความน่าจะเป็นของ ϵ_t ในรูปของ Bi-grammar distribution



รูปที่ 5.16 Means and Standard Deviations of Logarithms of Daily Flows (Station 33245)

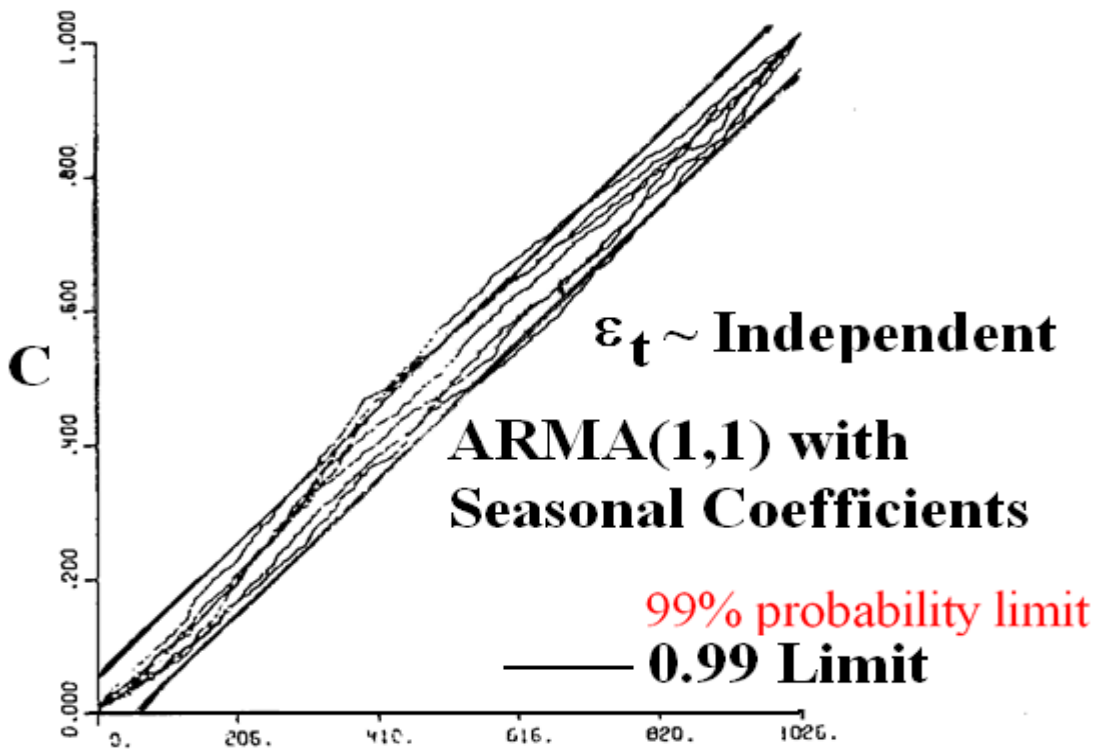


รูปที่ 5.17 Seasonal Serial Correlation Coefficients of Lags 1, 2 and 3 for Daily Flows (USGS Station 3-3245) (Logarithms).

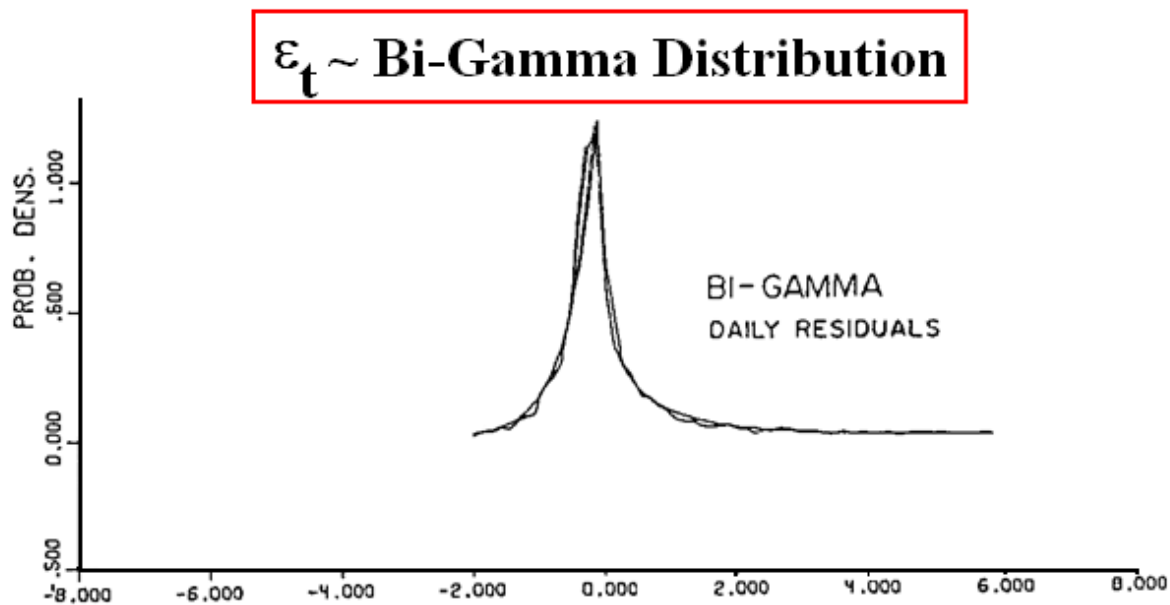


6 Harmonics Fourier Series Estimate

รูปที่ 5.18 Seasonal AR and MA Coefficients for Time Varying ARMA Model Applied to the Cyclically Standardized Daily Runoff Logarithms



รูปที่ 5.19 Cumulative Periodogram of Residuals in Time Varying ARMA(1, 1) Model



รูปที่ 5.20 Probability Distribution of Residuals

5.10 ออสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series**. Water Resources Publications. USA. pp.484

5.11 ภาคผนวก

(1) วิธีการคำนวณหาค่าของฟังก์ชันโดยวิธี Difference

ตัวอย่างการคำนวณหาค่า $f(19)$ โดยวิธี Difference จากค่าของฟังก์ชันในตาราง

x	11	17	21	23	31
f(x)	14.646	83.526	194,486	279.846	923.526

x	f(x)	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
11	14,646				
17	83,526	11,480			
21	194,486	27,740	1,626		
23	279,846	42,680	2,490	72	
31	923,526	80,460	3,778	92	1

$$\Delta f(x) = (83,526 - 14,646) / (17 - 11) = 11,480$$

$$\Delta^2 f(x) = (27,740 - 11,480) / (21 - 11) = 1,626$$

$$\Delta^3 f(x) = (2,490 - 1,626) / (23 - 11) = 72$$

$$\Delta^4 f(x) = (92 - 72) / (31 - 11) = 1$$

$$f(x) = 14,646 + 11,480(x-11) + 1,626(x-11)(x-17) + 72(x-17)(x-21) + 1(x-11)(x-17)(x-21)(x-23)$$

$$\begin{aligned} f(19) &= 14,646 + 11,480(8) + 1,626(8)(2) + 72(8)(2)(-2) + (8)(2)(-2)(-4) \\ &= 130,326 \end{aligned}$$

(2) จงหาค่าของฟังก์ชันโดยวิธี Finite Differences

ตัวอย่างการคำนวณ โดยใช้ Lagrange's Formula เพื่อหารูปแบบของฟังก์ชัน $f(x)$ ของค่าในตาราง

x	0	2	3	6
f(x)	659	705	729	804

ใช้ Lagrange's Formula

$$\begin{aligned} & \frac{f(x)}{x(x-2)(x-3)(x-6)} \\ &= \frac{659}{(0-2)(0-3)(0-6)} \left(\frac{1}{x-0} \right) + \frac{705}{(2-0)(2-3)(2-6)} \left(\frac{1}{x-2} \right) + \frac{729}{(3-0)(3-2)(3-6)} \left(\frac{1}{x-3} \right) + \frac{804}{(6-0)(6-2)(6-3)} \left(\frac{1}{x-6} \right) \\ &= \frac{659}{36x} + \frac{705}{8(x-2)} - \frac{729}{9(x-3)} + \frac{804}{72(x-6)} \\ &= \frac{-1,318(x-2)(x-3)(x-6) + 6,345x(x-3)(x-6) - 5,832x(x-2)(x-6) + 804x(x-2)(x-3)}{72x(x-2)(x-3)(x-6)} \end{aligned}$$

$$\begin{aligned} 72f(x) &= -1,318[(x^2 - 2x - 3x + 6)(x - 6)] + 6,345x[(x^2 - 3x)(x - 6)] \\ &\quad - 5,852[(x^2 - 2x)(x - 6)] + 80[(x^2 - 2x)(x - 3)] \end{aligned}$$

$$\begin{aligned} 72f(x) &= 1,318(x^3 - 11x^2 - 36x - 36) + 6,345(x^3 - 9x^2 + 18x) - 5,832(x^3 - 8x^2 + 12x) \\ &\quad + 804(x^3 - 5x^2 + 6x) \end{aligned}$$

$$\begin{aligned} 72f(x) &= 1,318x^3 + 14,498x^2 - 47,448x + 47,448 + 6,345x^3 - 57,105x^2 + 1,414,210x - \\ &\quad 5,832x^3 + 46,656x^2 - 69,984x + 804x^3 - 4,020x^2 + 4,824x \end{aligned}$$

$$72f(x) = -x^3 + 29x^2 + 1,602x + 47,448$$

$$f(x) = \frac{-x^3 + 29x^2 + 1,602x + 47,448}{72}$$

บทที่ 6

การจำลองอนุกรมเวลาด้วยแบบจำลอง AUTOREGRESSIVE INTEGRATED MOVING AVERAGE
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODELING

6.1 คำนำ

แบบจำลอง Autoregressive Integrated Moving Average หรือเรียกสั้นๆว่า ARIMA คือแบบจำลองสโตแคสติกที่สามารถจำลองอนุกรมเวลาซึ่งเป็น Nonstationary ได้โดยตรง โดยไม่ต้องแปลงอนุกรมเวลาให้เป็น Stationary ตามที่กล่าวมาในบทก่อนๆ จึงทำให้แบบจำลอง ARIMA มีข้อดีที่สำคัญคือมีพารามิเตอร์น้อยกว่าแบบจำลอง ARMA แต่มีข้อจำกัดในการใช้งานคือใช้ได้เฉพาะการพยากรณ์ ไม่สามารถใช้ในการสังเคราะห์ข้อมูลใหม่ (Generation) ได้ เพราะแบบจำลอง ARIMA มีค่าความแปรปรวนของค่าความผิดพลาด (Variance of Error) เข้าใกล้ ∞ แบบจำลอง ARIMA เป็นแบบจำลองที่เกี่ยวข้องกับแบบจำลอง ARMA ที่กล่าวถึงในบทที่ 5 ดังนี้คือ ถ้า x_t เป็นอนุกรมเวลาแบบ Nonstationary แล้วสามารถกำจัดส่วนที่เป็น Nonstationary ในอนุกรมเวลา x_t โดยใช้หลัก Differencing เช่น $u_t = x_t - x_{t-1}$ แล้วมีผลทำให้สามารถจำลองอนุกรมเวลา u_t ได้โดยใช้แบบจำลอง ARMA ได้ แสดงว่าสามารถจำลองอนุกรมเวลา x_t ได้โดยใช้แบบจำลอง ARIMA

ในบทนี้จะกล่าวถึงแบบจำลอง ARIMA รูปแบบต่างๆกัน ดังนี้

ARIMA(p,d,q)	Simple Non-periodic ARIMA หรือ No Seasonal ARIMA
ARIMA (P, D, Q) $_{\omega}$	Periodic ARIMA หรือ Seasonal ARIMA
ARIMA (p, d, q) \times (P, D, Q) $_{\omega}$	Multiplicative ARIMA ซึ่งเป็นการพิต Original Seasonal Series ด้วย ARIMA(P, D, Q) $_{\omega}$ และ พิต Residuals ด้วย ARIMA (p, d, q)

6.2 รูปแบบของแบบจำลอง ARIMA

Simple non-periodic ARIMA: ARIMA(p, d, q)	
$u_t = x_t - x_{t-1}$	1st Order Differencing
$u_t = \sum_{j=1}^p \phi_j u_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$	ARMA(1,1)
$x_t - x_{t-1} = \sum_{j=1}^p \phi_j (x_{t-j} - x_{t-j-1}) + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$	
$x_t = x_{t-1} + \sum_{j=1}^p \phi_j (x_{t-j} - x_{t-j-1}) + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$	ARIMA(p,1,q)

Periodic ARIMA or Seasonal ARIMA: ARIMA(P,D,Q)_w

$u_t = x_t - x_{t-\omega}$1st Order Seasonal Differencing

$u_t = \sum_{j=1}^P \Phi_j u_{t-j\omega} + \varepsilon_t - \sum_{j=1}^Q \theta_j \varepsilon_{t-j\omega}$ARMA(1,1)
--

$x_t = x_{t-\omega} + \sum_{j=1}^P \Phi_j (x_{t-j\omega} - x_{t-(j+1)\omega}) + \varepsilon_t - \sum_{j=1}^Q \theta_j \varepsilon_{t-j\omega}$...ARIMA(P,1,Q)
--

Multiplicative ARIMA

ARIMA(p, d, q)x(P, D, Q)_w

- Fit Original Seasonal Series with ARIMA(P, D, Q)_w

- Fit Residual with ARIMA(p, d, q)

Condensed Form $\forall \omega$ Multiplicative ARIMA(p,d,q)x(P,D,Q) _w
--

$\Phi(B^\omega) \phi(B)(1-B^\omega)^D(1-B)^d x_t = \Theta(B^\omega)\theta(B)\varepsilon_t$
--

$Bx_t = x_{t-1}$: Backward Operator

$B^n x_t = x_{t-n}$

$(1-B)x_t = x_t - x_{t-1}$; 1 st Order Simple Differencing
--

$(1-B)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$; 2 st Order Simple Differencing
--

$(1-B)^d x_t = d^{\text{th}}$ Order Simple Difference

$(1-B^\omega)x_t = x_t - x_{t-\omega}$; 1 st Order Seasonal Differencing
--

$(1-B^\omega)^2 x_t = x_t - 2x_{t-\omega} + x_{t-2\omega}$; 2 st Order Seasonal Differencing
--

$(1-B^\omega)^D x_t = D^{\text{th}}$ Order Seasonal Differencing
--

$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p)$ =Autoregressive Operator

$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q)$ =Moving average Operator

$\Phi(B^\omega) = (1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} \dots - \Phi_p B^{p\omega})$ =Seasonal Autoregressive Operator
--

$\Theta(B^\omega) = (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega})$ = Seasonal Moving Average Operator

6.3 วิธีการ Differencing

Differencing คือวิธีการในการกำจัดส่วนที่เป็น Nonstationary ออกจากอนุกรมเวลา ทำให้ได้อนุกรมเวลาแบบ Stationary เช่นสมมติให้ x_t คือ อนุกรมเวลาแบบ Nonstationary วิธีการ Differencing สามารถแบ่งออกเป็น 2 แนวทางคือ Simple Differencing สำหรับอนุกรมเวลาแบบ Nonperiodic และ Seasonal Differencing สำหรับอนุกรมเวลาแบบ Periodic ดังนี้

(1) Simple Difference

- First Order Difference

$$u_t = x_t - x_{t-1} \quad [6.1]$$

- Second Order Difference

$$w_t = u_t - u_{t-1} = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2} \quad [6.2]$$

(2) Seasonal Difference

- First Order Seasonal Difference

$$u_t = x_t - x_{t-\omega} \quad [6.3]$$

เมื่อ $\omega = \text{Maximum Period}$

$$Period = \frac{1}{Frequency} \quad [6.4]$$

- Second Order Seasonal Difference

$$w_t = u_t - u_{t-\omega} \quad [6.5]$$

$$w_t = (x_t - x_{t-\omega}) - (x_{t-\omega} - x_{t-2\omega}) = x_t - 2x_{t-\omega} + x_{t-2\omega} \quad [6.6]$$

6.4 หลักการ Infinite Variance Concept

$$\text{ให้ } x_t = (x_t - x_{t-1}) + (x_{t-1} - x_{t-2}) + (x_{t-2} - x_{t-3}) + \dots \rightarrow \infty$$

$$x_t = u_t + u_{t-1} + u_{t-2} + \dots \rightarrow \infty \quad [6.7]$$

$$x_t = \sum_{i=0}^{\infty} u_{t-i} \quad [6.8]$$

$$E(X_t) = E(\sum_{i=0}^{\infty} u_{t-i}) \quad [6.9]$$

$E(X_t)$ จะไม่ Converge เนื่องจากมี u_{t-i} จำนวน Infinite เทอม

$$Var(X_t) = Var(\sum_{i=0}^{\infty} u_{t-i}) = \infty \quad [6.10]$$

เนื่องจาก $Var(X_t)$ มีค่าอนันต์จึงทำให้แบบจำลอง ARIMA ไม่เหมาะที่จะนำมาใช้ในการสังเคราะห์ข้อมูลใหม่ แต่สามารถใช้ในการพยากรณ์ได้

6.5 แบบจำลอง ARIMA (p, 1, q)

ให้ $u_t = x_t - x_{t-1}$ คือ First Order Simple Difference ซึ่งสามารถจำลองด้วยแบบจำลอง ARMA(p,q) ดังนั้น จะสามารถจำลองอนุกรมเวลา x_t โดยตรงได้ด้วยแบบจำลอง ARIMA(p,1,q) ดังนี้

$$u_t = \sum_{j=1}^p \phi_j u_{t-j} - \sum_{j=0}^q \theta_j \varepsilon_{t-j}; \theta_0 = -1 \quad \dots \dots \text{ARMA}(p,q) \quad [6.11]$$

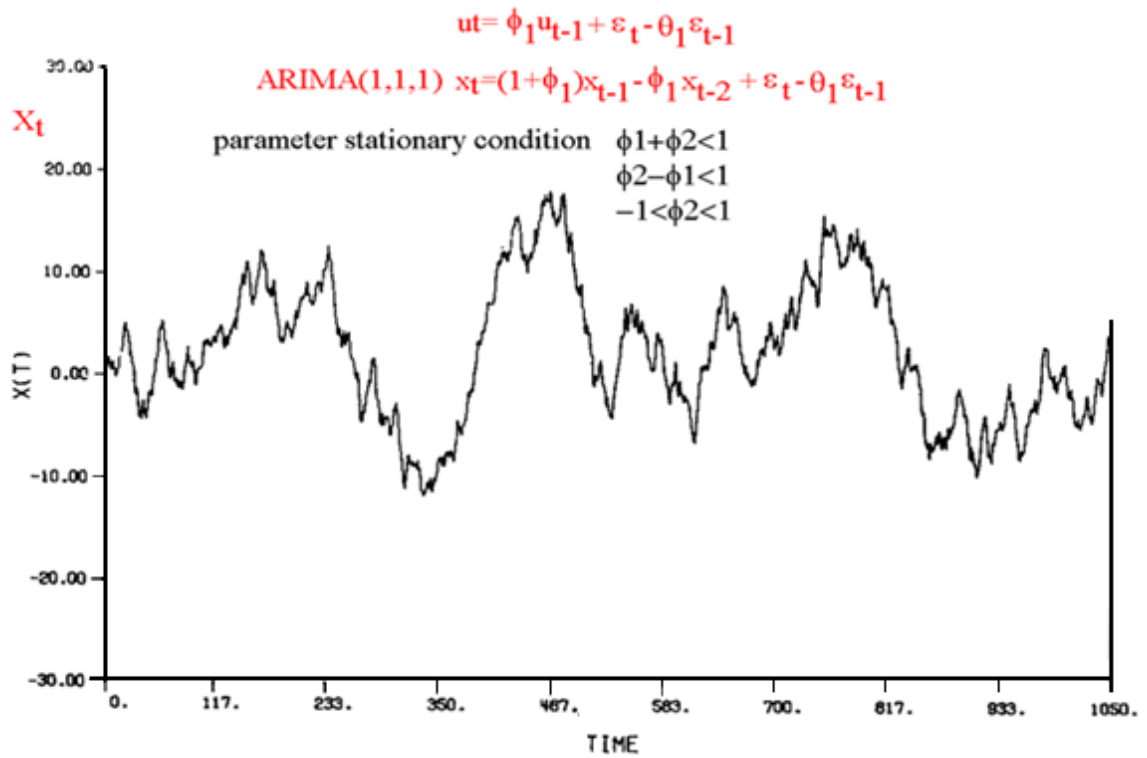
$$x_t - x_{t-1} = \sum_{j=1}^p \phi_j (x_{t-j} - x_{t-j-1}) - \sum_{j=0}^q \theta_j \varepsilon_{t-j}$$

$$x_t = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_{t-j} - \phi_p x_{t-p-1} - \sum_{j=0}^q \theta_j \varepsilon_{t-j}; \theta_0 = \phi_0 = -1; \quad \dots \dots \text{ARIMA}(p,1,q) \quad [6.12]$$

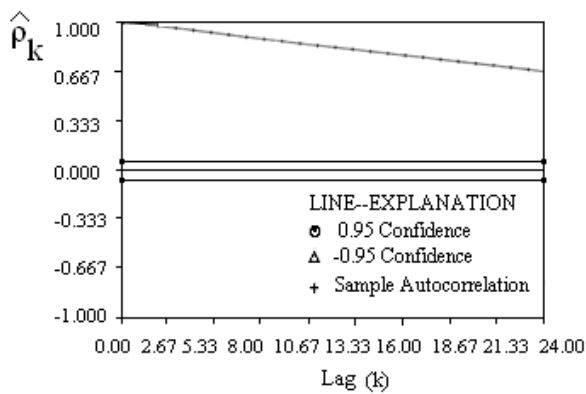
กรณี ARIMA(1,1,1)

$$x_t = (1 + \phi_1)x_{t-1} - \phi_1 x_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

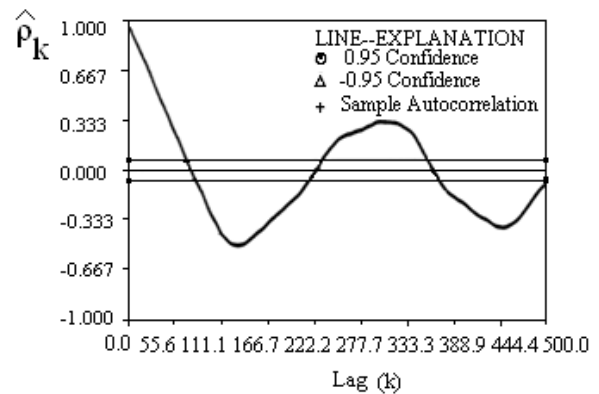
ตัวอย่างกราฟอนุกรมเวลาที่สังเคราะห์ขึ้นโดย ARIMA(1,1,1) ซึ่งมีค่าพารามิเตอร์ $\phi_1=0.5$, $\theta_1=0.3$ และ $\sigma_\varepsilon^2=0.6$ แสดงอยู่ในรูปที่ 6.1 Autocorrelation Function และ Spectral Density ของแบบจำลองเดียวกันแสดงอยู่ในรูปที่ 6.2 และ 6.3 ตามลำดับ



รูปที่ 6.1 อนุกรมเวลาที่สังเคราะห์ขึ้นจากแบบจำลอง ARIMA(1,1,1) ที่มีค่าพารามิเตอร์ $\phi_1=0.5$, $\theta_1=0.3$ และ $\sigma_\epsilon^2=0.6$

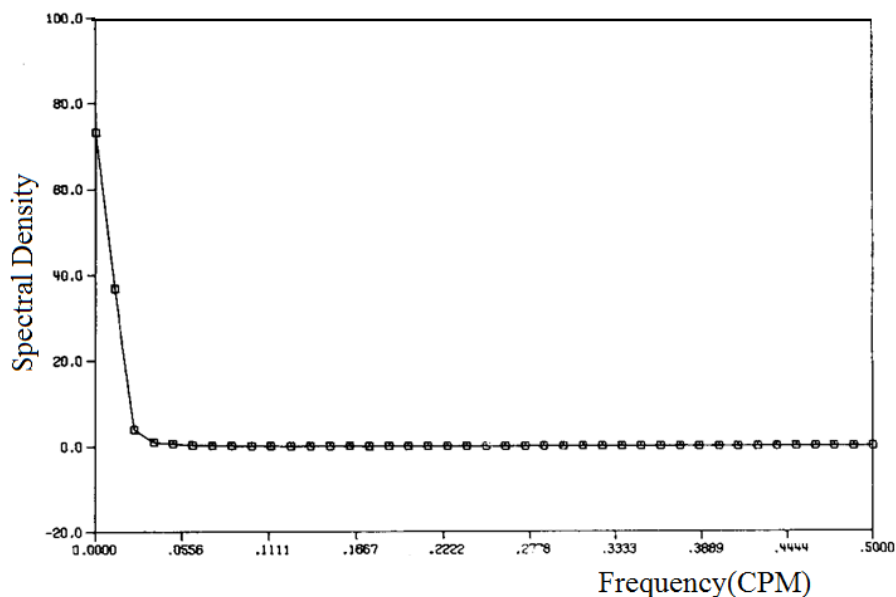


Lag 0-24



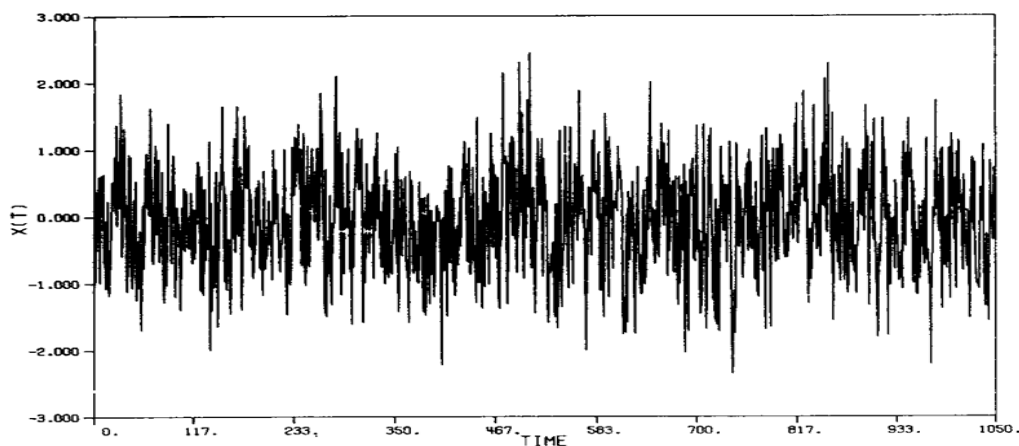
Lag 0-500

รูปที่ 6.2 Autocorrelation Function ของอนุกรมเวลาที่สังเคราะห์ขึ้นจากแบบจำลอง ARIMA(1,1,1) ที่มีค่าพารามิเตอร์ $\phi_1 = 0.5$, $\theta_1 = 0.3$ และ $\sigma_\epsilon^2 = 0.6$

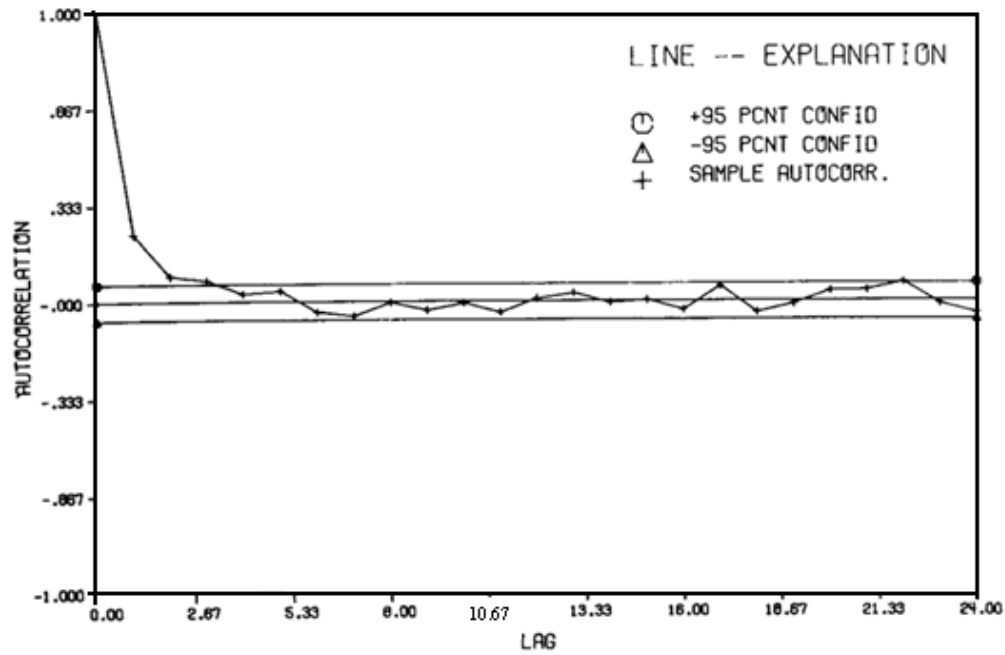


รูปที่ 6.3 Spectral Density ของอนุกรมเวลาที่สังเคราะห์ขึ้นจากแบบจำลอง ARIMA(1,1,1) ที่มีค่าพารามิเตอร์ $\phi_1 = 0.5$, $\theta_1 = 0.3$ และ $\sigma_\varepsilon^2 = 0.6$

ACF ในรูปที่ 6.2 มีค่าลดลงอย่างช้าๆ (Slow Decay) แสดงว่า x_t เป็นอนุกรมเวลาแบบ Nonstationary การกำจัด Nonstationary ออกจากอนุกรมเวลาทำได้โดยวิธี Differencing ผลการทำ 1st Differencing ของอนุกรมเวลาในรูปที่ 6.1 แสดงอยู่ในรูปที่ 6.4 และ ACF ของ 1st Differencing แสดงอยู่ในรูปที่ 6.5 ACF มีค่าลดลงเป็นศูนย์ที่ Lag 4 ซึ่งแสดงให้เห็นว่า 1st Differencing ของอนุกรมเวลา x_t เป็น Stationary แต่ถ้ายังไม่เป็น Stationary ให้ทำ Differencing โดยใช้ $d=2, 3, \dots$ จนกว่าจะเป็น Stationary



รูปที่ 6.4 1st Differencing ($u_t = x_t - x_{t-1}$) ของอนุกรมเวลาที่สังเคราะห์ขึ้นจากแบบจำลอง ARIMA(1,1,1) ที่มีค่าพารามิเตอร์ $\phi_1 = 0.5$, $\theta_1 = 0.3$ และ $\sigma_\varepsilon^2 = 0.6$



รูปที่ 6.5 Autocorrelation Function ของ 1st Differencing ($u_t = x_t - x_{t-1}$) ของอนุกรมเวลาที่
ตั้งตระหนักรู้ขึ้นจากแบบจำลอง ARIMA(1,1,1) ที่มีค่าพารามิเตอร์ $\phi_1 = 0.5$, $\theta_1 = 0.3$ และ $\sigma_\varepsilon^2 = 0.6$

6.6 ฟังก์ชันการพยากรณ์ (Forecasted Function) ของแบบจำลอง ARIMA(p,1,q) สำหรับ Lead Time เท่ากับ L

จากสมการ [6.12] ถ้า $t=t+L$ จะสามารถเขียนแบบจำลอง ARIMA(p,1,q) ได้ใหม่ดังนี้

$$x_{t+L} = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_{t+L-j} - \phi_p x_{t+L-p-1} - \sum_{j=0}^q \theta_j \varepsilon_{t+L-j} \quad [6.13]$$

จากสมการ [6.13] จะสามารถเขียนฟังก์ชันการพยากรณ์ได้ดังนี้

$$[x_{t+L}] = \sum_{j=1}^p (\phi_j - \phi_{j-1}) [x_{t+L-j}] - \phi [x_{t+L-p-1}] - \sum_{j=0}^q \theta_j [\varepsilon_{t+L-j}] \quad [6.14]$$

$$x_t(L) = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_t(L-j) - \phi_p x_t(L-p-1) - \sum_{j=0}^q \theta_j [\varepsilon_{t+L-j}] \quad [6.15]$$

ถ้า $j \geq L$	ถ้า $j < L$
$x_{t+L-j} = \text{Observed Value}$	$x_{t+L-j} = \text{Future Value}$
$[x_{t+L-j}] = x_t(L-j) = x_{t+L-j}$	$[x_{t+L-j}] = x_t(L-j) = \text{Forecasted Function.}$
$\varepsilon_{t+L-j} = \text{Error Value in the Past}$	$\varepsilon_{t+L-j} = \text{Future Value หรือ Random Value}$
$[\varepsilon_{t+L-j}] = \varepsilon_{t+L-j}$	$[\varepsilon_{t+L-j}] = 0$

จากข้อเท็จจริงในตารางข้างบนจะเขียนจะสามารถเขียนสมการ [6.15] ใหม่ได้ดังนี้

$$x_t(L) = \sum_{j=1}^{L-1} (\phi_j - \phi_{j-1}) x_t(L-j) + \sum_{j=L}^p (\phi_j - \phi_{j-1}) x_{t+L-j} - \phi_p x_{t+L-p-1} - \sum_{j=0}^{L-1} \theta_j [\varepsilon_{t+L-j}] - \sum_{j=L}^q \theta_j \varepsilon_{t+L-j} \quad [6.16]$$

และเนื่องจาก $\varepsilon_t(L-j) = 0$ เมื่อ $j < L$ จะได้ว่า

$$x_t(L) = \sum_{j=1}^{L-1} (\phi_j - \phi_{j-1}) x_t(L-j) + \sum_{j=L}^p (\phi_j - \phi_{j-1}) x_{t+L-j} - \phi_p x_{t+L-p-1} - \sum_{j=L}^q \theta_j \varepsilon_{t+L-j} \quad [6.17]$$

ถ้า $q < L$

$$x_t(L) = \sum_{j=1}^{L-1} (\phi_j - \phi_{j-1}) x_t(L-j) + \sum_{j=L}^p (\phi_j - \phi_{j-1}) x_{t+L-j} - \phi_p x_{t+L-p-1} \quad [6.18]$$

ถ้า $q < L$ และ $p = L-1 < L$

$$x_t(L) = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_t(L-j) - \phi_p x_{t+L-p-1} \quad [6.19]$$

ถ้า $q < L$ และ $p < L-1$

$$x_t(L) = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_t(L-j) - \phi_p x_t(L-p-1) \quad [6.20]$$

กรณี $L < p$ และ $L < q$	จากสมการที่ [6.15] $x_t(L) = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_t(L-j) - \phi_p x_t(L-p-1) - \sum_{j=0}^q \theta_j [\varepsilon_{t+L-j}]$
$L=1$	$x_t(1) = \sum_{j=1}^p (\phi_j - \phi_{j-1}) x_{t+1-j} - \phi_p x_{t-p} + \sum_{j=1}^q \theta_j \varepsilon_{t+1-j}$
$L=2$	$x_t(2) = (\phi_1 + 1) x_t(1) + \sum_{j=2}^p (\phi_j - \phi_{j-1}) x_{t+2-j} - \phi_p x_{t+1-p} + \sum_{j=2}^q \theta_j \varepsilon_{t+2-j}$
$L=3$	$x_t(3) = \sum_{j=1}^2 (\phi_j - \phi_{j-1}) x_t(3-j) + \sum_{j=3}^p (\phi_j - \phi_{j-1}) x_{t+2-j} - \phi_p x_{t+2-p} + \sum_{j=3}^q \theta_j \varepsilon_{t+2-j}$
$L=L$	$x_t(L) = \sum_{j=1}^{L-1} (\phi_j - \phi_{j-1}) x_t(L-j) + \sum_{j=L}^p (\phi_j - \phi_{j-1}) x_{t+L-j-1} - \phi_p x_{t+L-p-1} + \sum_{j=L}^q \theta_j \varepsilon_{t+2-j}$ [6.21]

6.7 แบบจำลอง Multiplicative ARIMA สำหรับอนุกรมเวลาแบบ Periodic

6.7.1 Seasonal Differencing

$$u_t = x_t - x_{t-\omega} \quad (1^{\text{st}} \text{ Order Seasonal Differencing}) \quad [6.22]$$

กรณีอนุกรมเวลารายเดือน $\omega = 12$

$$\begin{aligned} w_t &= u_t - u_{t-\omega} \quad (2^{\text{nd}} \text{ Order Seasonal Differencing}) \\ &= (x_t - x_{t-\omega}) - (x_{t-\omega} - x_{t-2\omega}) \\ &= x_t - 2x_{t-\omega} + x_{t-2\omega} \end{aligned} \quad [6.23]$$

$(1-B^\omega)^D x_t = D^{\text{th}} \text{ Order Seasonal Differencing}$

6.7.2 Simple Seasonal ARIMA(P,1,Q) ω

$$\begin{aligned} u_t &= x_t - x_{t-\omega} \\ u_t &= \sum_{j=1}^P \phi_j u_{t-j\omega} + \varepsilon_t - \sum_{j=1}^Q \theta_j \varepsilon_{t-j\omega} \end{aligned} \quad [6.24]$$

$$\begin{aligned}
 x_t - x_{t-\omega} &= \sum_{j=1}^P \Phi_j (x_{t-j\omega} - x_{t-(j+1)\omega}) + \varepsilon_t - \sum_{j=1}^Q \theta_j \varepsilon_{t-j\omega} \\
 x_t &= x_{t-\omega} + \sum_{j=1}^P \Phi_j (x_{t-j\omega} - x_{t-(j+1)\omega}) + \varepsilon_t - \sum_{j=1}^Q \theta_j \varepsilon_{t-j\omega}
 \end{aligned} \tag{6.25}$$

x_t มีแบบจำลองแบบ ARIMA(P,1,Q) ω

6.7.3 General Simple Seasonal ARIMA(P,D,Q) ω

แบบจำลอง ARIMA(P,D,Q) ω ในรูปย่อซึ่งเขียนในรูปของ Backward Operator

$$\Phi(B^\omega)(1-B^\omega)^D x_t = \Theta(B^\omega) \varepsilon_t \tag{6.26}$$

$(1-B^\omega)^D x_t = D^{\text{th}}$ Order Seasonal Differencing

$\Phi(B^\omega) = (1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} \dots - \Phi_P B^{P\omega}) =$ Seasonal Autoregressive Operator

$\Theta(B^\omega) = (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega}) =$ Seasonal Moving Average Operator

$(1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} \dots - \Phi_P B^{P\omega})(1-B^\omega)^D x_t = (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega}) \varepsilon_t$

ถ้า $D=1$

$(1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} \dots - \Phi_P B^{P\omega})(1-B^\omega) x_t = (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega}) \varepsilon_t$

$(1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} \dots - \Phi_P B^{P\omega} - B^\omega + \Phi_1 B^{2\omega} + \Phi_2 B^{3\omega} \dots + \Phi_{P-1} B^{P\omega} + \Phi_P B^{(P+1)\omega}) x_t$

$= (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega}) \varepsilon_t$

$(1 - B^\omega - \Phi_1 (B^\omega - B^{2\omega}) - \Phi_2 (B^{2\omega} - B^{3\omega}) \dots - \Phi_{P-1} (B^{(P-1)\omega} - B^{P\omega}) \dots - \Phi_P (B^{P\omega} - B^{(P+1)\omega})) x_t$

$= (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega}) \varepsilon_t$

$x_t = (B^\omega + \Phi_1 (B^\omega - B^{2\omega}) + \Phi_2 (B^{2\omega} - B^{3\omega}) \dots + \Phi_{P-1} (B^{(P-1)\omega} - B^{P\omega}) \dots + \Phi_P (B^{P\omega} - B^{(P+1)\omega})) x_t$

$+ (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} \dots - \Theta_Q B^{Q\omega}) \varepsilon_t \tag{6.27}$

สมการ [6.27] คือสมการแบบเต็มรูปของของสมการ [6.26]

6.7.4 ARIMA(0,1,Q) ω หรือ IMA(1,Q) ω

$u_t = x_t - x_{t-\omega}$

$u_t = \varepsilon_t - \theta_1 \varepsilon_{t-\omega} - \theta_2 \varepsilon_{t-2\omega} \dots - \theta_Q \varepsilon_{t-Q\omega} \tag{6.28}$

x_t มีแบบจำลองแบบ ARIMA(0,1,Q) ω

Autocorrelation Function ของ ARIMA(0,1,Q) ω

$\rho_{k\omega} = \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_Q \theta_{Q-k}}{1 + \theta_1^2 + \dots + \theta_Q^2}; k \leq Q\omega$

$= 0; k > Q\omega \tag{6.29}$

6.7.5 ARIMA (P,1,0) ω

$u_t = x_t - x_{t-\omega}$

$u_t = \Phi_1 u_{t-\omega} + \Phi_2 u_{t-2\omega} + \dots + \Phi_P u_{t-P\omega} + \varepsilon_t \tag{6.30}$

x_t มีแบบจำลองแบบ ARIMA(P,1,0) ω

Autocorrelation Function ของ ARIMA(P,1,0)₁₂

$$\rho_{k\omega} = \Phi_1 \rho_{(k-1)\omega} + \Phi_2 \rho_{(k-2)\omega} + \dots + \Phi_P \rho_{(k-P)\omega}; \quad k > 0 \quad [6.31]$$

6.7.6 การฟิต 1st Order Seasonal Difference ของอนุกรมเวลารายเดือนด้วยแบบจำลอง AR(1)

$$u_t = x_t - x_{t-12}$$

$$u_t = \Phi_1 u_{t-1} + \varepsilon_t \dots \text{AR}(1) \quad [6.32]$$

$$x_t - x_{t-12} = \Phi_1 (x_{t-1} - x_{t-13}) + \varepsilon_t$$

$$x_t = \Phi_1 x_{t-1} + x_{t-12} - \Phi_1 x_{t-13} + \varepsilon_t \quad [6.33]$$

สมการ [6.33] คือ ARIMA(1,1,0)₁₂ หรือ Nonstationary AR(13) ซึ่งมีค่า Autoregressive Coefficients ดังนี้

$$\phi_1 = \Phi_1$$

$$\phi_2 = \phi_3 = \dots = \phi_{11} = 0$$

$$\phi_{12} = 1$$

$$\phi_{13} = -\Phi_1$$

6.7.7 การฟิต 1st Order Seasonal Difference ของอนุกรมเวลารายเดือนด้วยแบบจำลอง MA(1)

$$u_t = x_t - x_{t-12}$$

$$u_t = \alpha_t - \theta_1 \alpha_{t-12} \dots \text{MA}(1) \quad [6.34]$$

$$x_t - x_{t-12} = \alpha_t - \theta_1 \alpha_{t-12}$$

$$x_t = x_{t-12} + \alpha_t - \theta_1 \alpha_{t-12} \quad [6.35]$$

สมการ [6.35] คือ ARIMA(1,1,0)₁₂ หรือ Nonstationary ARMA(12,12) ซึ่งมีค่า Autoregressive และ Moving Average Coefficients ดังนี้

$$\phi_1 = \phi_2 = \dots = \phi_{11} = 0$$

$$\phi_{12} = 1$$

$$\theta_1 = \theta_2 = \dots = \theta_{11} = 0$$

$$\theta_{12} = \theta_1$$

6.7.8 แบบจำลอง Multiplicative ARIMA(1, 0, 0)x(0, 1, 1)₁₂

$$u_t = x_t - x_{t-12}$$

ให้

$$u_t = \alpha_t - \theta_1 \alpha_{t-12} \dots \text{MA}(1) \quad [6.36]$$

$$x_t = x_{t-12} + \alpha_t - \theta_1 \alpha_{t-12} \dots \text{ARIMA}(0,1,1)_{12} \quad [6.37]$$

α_t คืออนุกรมเวลาของ Residual ใน MA(1) ซึ่งสามารถจำลองได้ด้วย AR(1)

ให้

$$\alpha_t = \phi_1 \alpha_{t-1} + \varepsilon_t \dots \text{AR}(1) \quad [6.38]$$

แทนค่า α_t และ α_{t-12} จากสมการ [6.38] ลงในสมการ [6.36]

$$\begin{aligned} u_t &= (\phi_1 \alpha_{t-1} + \varepsilon_t) - \theta_1 (\phi_1 \alpha_{t-13} + \varepsilon_{t-12}) \\ u_t &= \phi_1 (\alpha_{t-1} - \theta_1 \alpha_{t-13}) + \varepsilon_t - \theta_1 \varepsilon_{t-12} \end{aligned} \quad [6.39]$$

แทนค่า $u_{t-1} = \alpha_{t-1} - \theta_1 \alpha_{t-13}$ ลงในสมการ [6.39] จะได้

$$\begin{aligned} u_t &= \phi_1 u_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-12} \\ x_t - x_{t-12} &= \phi_1 (x_{t-1} - x_{t-13}) + \varepsilon_t - \theta_1 \varepsilon_{t-12} \\ x_t &= \phi_1 x_{t-1} + x_{t-12} - \phi_1 x_{t-13} + \varepsilon_t - \theta_1 \varepsilon_{t-12} \end{aligned} \quad [6.40]$$

สมการที่ [6.40] คือแบบจำลอง Multiplicative ARIMA(1, 0, 0)x(0, 1, 1)₁₂ ซึ่งเกิดจากการจำลอง 1st

Order Seasonal Difference (u_t) ซึ่งมี $\omega=12$ ด้วย MA(1) และจำลอง Residual (α_t) ด้วย AR(1)

6.7.9 แบบจำลอง ARIMA(p,0,q)x(P, 1, Q)₁₂

$u_t = 1^{\text{st}}$ Seasonal Difference ซึ่งมี $\omega=12$

$$\begin{aligned} u_t &= x_t - x_{t-12} \\ u_t &= \sum_{j=1}^P \phi_j u_{t-12j} + \alpha_t - \sum_{j=1}^Q \theta_j \alpha_{t-12j} \end{aligned} \quad [6.41]$$

ให้ Residual (α_t) ของ 6.40 มีแบบจำลองแบบ ARMA(p, q)

$$\alpha_t = \sum_{j=1}^p \phi_j \alpha_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad [6.42]$$

เมื่อแทนค่า α_t และ α_{t-12j} จากสมการ [6.42] และ $u_t = x_t - x_{t-12}$ ลงในสมการ [6.41]

$$\begin{aligned} x_t - x_{t-12} &= \sum_{j=1}^P \phi_j u_{t-12j} + \left(\sum_{j=1}^p \phi_j \alpha_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \right) - \\ &\quad \sum_{j=1}^Q \theta_j \left(\sum_{j=1}^p \phi_j \alpha_{t-(12+1)j} + \varepsilon_{t-12j} - \sum_{j=1}^q \theta_j \varepsilon_{t-(12+1)j} \right) \end{aligned}$$

$$x_t = x_{t-12} + \sum_{j=1}^P \phi_j u_{t-12j} + \left(\sum_{j=1}^p \phi_j \alpha_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \right) \quad [6.43]$$

จะสรุปได้ว่า x_t ในสมการ [6.43] มีแบบจำลองแบบ ARIMA (p,0, q) × (P, 1, Q)₁₂

6.7.10 แบบจำลอง Multiplicative ARIMA(p, d, q)x(P, D, Q)₁₂

แบบจำลอง Multiplicative ARIMA(1,1,1)x(1,1,1)₁₂ กรณี $\omega=12$

$$\Phi(B^{12})\phi(B)(1 - B^{12})(1 - B)x_t = \theta(B^{12})\theta(B)\varepsilon_t \quad [6.44]$$

$$(1 - \phi_1 B^{12})(1 - \phi B)(1 - B^{12})(1 - B)x_t = (1 - \theta_1 B^{12})(1 - \theta_1 B)\varepsilon_t$$

$$(1 - \phi_1 B^{12} - \phi_1 B + \phi_1 \phi_1 B^{13})(1 - B^{12} - B + B^{13})x_t = (1 - \theta_1 B^{12} - \theta_1 B + \theta_1 \theta_1 B^{13})\varepsilon_t$$

$$\{(1 - \phi_1 B^{12} - \phi_1 B + \phi_1 \phi_1 B^{13}) - (B^{12} - \phi_1 B^{24} - \phi_1 B^{13} + \phi_1 \phi_1 B^{25}) - (B - \phi_1 B^{13} - \phi_1 B^2 + \phi_1 \phi_1 B^{14}) + (B^{13} - \phi_1 B^{25} - \phi_1 B^{14} + \phi_1 \phi_1 B^{26})\} X_t = (1 - \theta_1 B^{12} - \theta_1 B + \theta_1 \theta_1 B^{13}) \varepsilon_t \phi$$

$$[1 - (1 + \phi_1)B + \phi_1 B^2 - (1 + \phi_1)B^{12} + (1 + \phi_1 + \Phi_1 + \Phi_1 \phi_1)B^{13} - (\phi_1 + \Phi_1 \phi_1)B^{14} + \Phi_1 B^{24} - (\Phi_1 + \Phi_1 \phi_1)B^{25} + \Phi_1 \phi_1 B^{26}] X_t = (1 - \theta_1 B^{12} - \theta_1 B + \theta_1 \theta_1 B^{13}) \varepsilon_t \phi$$

$$X_t = (1 + \phi_1)X_{t-1} - \phi_1 X_{t-2} + (1 + \Phi_1)X_{t-12} - (1 + \phi_1 + \Phi_1 + \Phi_1 \phi_1)X_{t-13} + (\phi_1 + \Phi_1 \phi_1)X_{t-14} - \Phi_1 X_{t-24} + (\Phi_1 + \Phi_1 \phi_1)X_{t-25} - \Phi_1 \phi_1 X_{t-26} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_1 \varepsilon_{t-12} + \theta_1 \theta_1 \varepsilon_{t-13} \quad [6.45]$$

6.8 การประมาณค่าพารามิเตอร์ของ Simple ARIMA(p,d,q)

หลังจากการทำ d^{th} -Differencing อนุกรมเวลา x_t (x_t เป็น ARIMA(p,d,q)) จนเป็น Stationary แล้ว จะ FIT Differenced Series ด้วยแบบจำลอง ARMA(p,q) จึงสามารถหาพารามิเตอร์ของแบบจำลอง ARMA(p,q) ตามวิธีที่กล่าวถึงในบทที่ 5

6.9 การตรวจสอบ Goodness of Fit ของ ARIMA(p,d,q)

ขั้นตอนและวิธีการทดสอบ Goodness of Fit ของ ARMA(p,q) ที่ FIT กับอนุกรมเวลาที่ผ่านมาการ Differencing จะเหมือนกับการทดสอบ Goodness of Fit ของ ARMA(p,q) ตามที่กล่าวไว้ในบทที่ 5 ซึ่งได้แก่

- การทดสอบสมมติฐานว่า Residual เป็นอิสระและมีการแจกแจงแบบปกติ
- การตรวจสอบเปรียบเทียบ Model Correlogram และ Correlogram ของอนุกรมเวลาที่สังเคราะห์ ขึ้นกับ Sample Correlogram
- การตรวจสอบ Parsimony ของพารามิเตอร์โดยใช้ Akaike Information Criteria

6.10 ขั้นตอนการจำลองด้วย Simple ARIMA (Procedure of Simple ARIMA Modeling)

ขั้นตอนการจำลองด้วย Simple ARIMA

- (1) ตรวจสอบ Normality ของอนุกรมเวลาตัวอย่าง ถ้าอนุกรมเวลาตัวอย่างไม่เป็น Normal ให้แปลงเป็น เป็น Normal
- (2) ทำการ Differencing โดยการพล็อตอนุกรมเวลาตัวอย่างที่ทำการ Normalized แล้ว ถ้าพบว่ามี Nonstationary เช่น Jump หรือ Trend หรือมีทั้งคู่ ให้ทำการ Differencing กรณีมี Jump หรือ Trend เพียงอย่างเดียวให้ทำ 1st Differencing กรณีที่มีทั้ง Jump และ Trend ให้ทำ 2nd Differencing

- (3) ตรวจสอบ ACF ถ้าพบว่ามีค่าลดลงอย่างช้า (Slow Decay) แสดงว่าต้องทำ Differencing การทำ Differencing ให้เริ่มจากลำดับที่ 1 ก่อน
- (4) การฟิตแบบจำลอง ARMA(p,q) กับอนุกรมเวลาที่ผ่านการ Differencing

6.11 ตัวอย่างการสร้างแบบจำลอง Simple ARIMA (Example of Simple ARIMA Modeling)

ให้จำลองฝนรายเดือนที่ Salmonia มลรัฐ Indiana (Station 12-7747) ดังตารางที่ 6.1 ด้วย ARIMA(1,1,1)

ขั้นที่ 1 ให้ x_t คือฝนรายเดือน ดังตารางที่ 5.1

$y_t = \sqrt{x_t}$ ซึ่งเป็น Normal Distribution โดยประมาณ

ขั้นที่ 2 ทำ 1st Differencing เพื่อกำจัด Nonstationarity ใน y_t ดังนี้

$$u_1 = \sqrt{4.910} - \sqrt{3.920} = 0.236$$

ผลการคำนวณค่า u_t อื่นๆ โดยโปรแกรม FTRDIF* ของ IMSL แสดงอยู่ในตารางที่ 6.2

ขั้นที่ 3 ทำการฟิต ARMA(1,1) กับ 1st Differenced Series (u_t) ตามขั้นตอนในบทที่ 5

(a) $\bar{u}_t = -0.0011$
 $s^2 = 0.6065$

(b) ACF และ PACF ของ u_t แสดงอยู่ในรูปที่ 6.6 และ 6.7 ตามลำดับ

95% Confident Interval ของ ACF และ PACF เท่ากับ

$$= \frac{\pm 2}{\sqrt{N}} = \frac{\pm 2}{\sqrt{455}} = \pm 0.094$$

จากรูปที่ 6.6 จะเห็นว่า ACF มีความสำคัญเฉพาะ Lag 1 เท่านั้น ส่วน PACF ในรูปที่ 6.7 มีความสำคัญช่วง Lag 1-10

(c) Identification

- ACF และ PACF แสดงให้เห็นว่า ARMA(1,1) น่าจะเหมาะกับอนุกรมเวลา u_t

(d) การประมาณค่าพารามิเตอร์ ϕ_1 ของ ARMA(1,1) โดยวิธีโมเมนต์

$$\hat{\phi}_1 = \frac{C_2}{C_1} = \frac{0.036415}{-0.48294} = -0.0754$$

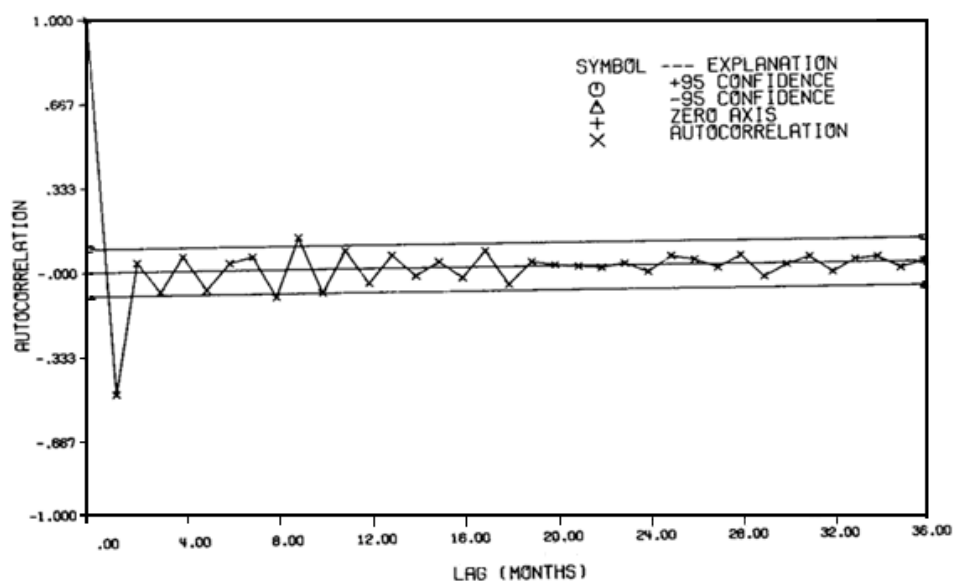
$$\hat{\theta}_{00} = \bar{u}_t(1 - \hat{\phi}_1) = -0.0011(1 + 0.0754) = -0.0012$$

ตารางที่ 6.1 Monthly Precipitation(inches) at Salanomia, Indian

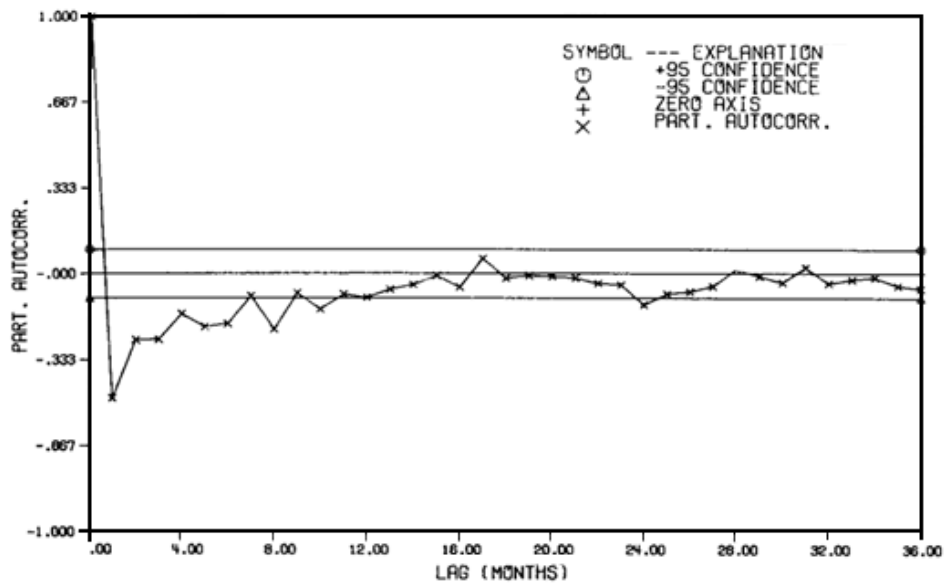
YEAR	Month												TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	
1	3.920	4.910	3.850	8.210	2.690	4.050	3.680	1.460	4.020	4.340	5.070	7.970	54.170
2	3.900	2.710	7.960	7.160	2.150	7.100	6.216	9.690	0.360	1.570	2.010	4.480	55.710
3	2.680	1.230	3.510	1.470	1.040	4.070	1.360	1.220	7.910	1.920	5.870	4.960	37.240
4	0.320	2.200	2.480	3.990	0.830	9.800	2.560	7.000	5.230	1.780	3.450	2.260	41.900
5	1.660	2.820	1.760	1.690	1.380	4.110	4.600	0.800	1.220	2.600	0.990	2.780	26.680
6	7.490	4.500	3.130	7.750	2.440	1.110	4.950	4.470	4.440	0.820	4.600	2.960	48.660
7	5.880	1.260	2.880	2.410	1.970	7.480	1.660	4.700	3.290	9.380	3.180	3.240	47.330
8	0.500	3.370	1.970	4.440	4.250	5.680	5.570	1.310	5.140	5.480	1.680	1.060	40.450
9	0.770	1.390	1.370	2.780	4.430	1.800	6.930	4.620	6.290	1.530	2.000	1.530	35.440
10	1.060	4.240	2.460	1.860	0.580	0.520	2.730	0.660	5.960	4.590	3.310	1.630	29.600
11	6.560	1.490	2.360	2.710	3.270	4.080	2.680	4.990	6.650	4.330	3.280	1.530	43.930
12	1.580	5.330	4.430	0.650	1.630	6.410	3.710	4.300	1.620	3.050	0.430	2.810	35.950
13	1.940	1.080	1.790	1.070	3.790	5.900	5.850	4.600	0.510	3.060	7.730	4.000	41.320
14	0.250	2.990	3.370	1.750	6.720	4.400	4.650	3.700	7.790	3.660	2.360	5.970	47.610
15	3.180	4.160	2.270	2.100	46.300	3.560	2.190	8.510	7.790	5.690	3.630	0.560	44.010
16	2.940	4.580	3.350	4.460	0.380	1.850	8.720	5.510	3.530	5.950	4.000	2.930	49.680
17	1.980	2.290	1.120	3.070	3.960	6.670	6.040	3.940	5.320	0.520	2.130	2.160	37.620
18	2.360	7.800	5.340	0.300	3.220	5.940	1.510	3.980	3.740	3.300	0.780	2.720	52.840
19	4.440	1.090	4.930	2.560	6.850	3.790	5.870	4.990	9.590	3.520	3.910	3.460	51.790
20	2.530	6.640	1.550	7.300	0.000	6.180	3.130	2.910	6.380	1.650	2.670	3.070	42.060
21	1.510	0.000	5.800	3.840	3.250	5.210	3.210	2.670	3.500	2.440	1.620	3.740	38.790
22	1.580	2.170	5.320	4.730	0.730	3.380	3.580	3.990	2.230	7.090	1.310	0.160	36.270
23	1.140	1.090	2.110	2.920	2.970	1.730	3.530	1.950	2.590	1.360	5.510	3.980	30.860
24	3.860	1.230	3.070	2.290	8.370	5.640	2.390	2.190	3.780	4.800	1.200	5.830	45.370
25	5.710	3.270	0.750	2.800	7.350	3.470	4.140	3.340	4.890	3.790	3.400	2.250	44.440
26	2.660	1.560	2.460	3.900	3.730	1.620	6.790	6.340	8.010	4.030	2.830	2.950	46.880
27	2.860	7.550	6.410	2.970	0.630	2.060	3.000	4.850	4.450	7.730	3.760	6.440	52.110
28	1.200	2.860	1.140	8.290	3.200	2.930	1.310	3.770	2.960	2.750	3.010	1.500	34.920
29	4.000	2.390	3.840	2.600	3.610	0.970	1.590	3.970	9.230	4.150	2.410	0.970	39.730
30	2.450	2.910	1.940	1.890	3.590	6.260	3.550	8.330	4.020	4.160	4.270	0.950	44.320
31	2.800	3.720	3.930	4.290	6.170	4.020	0.790	4.460	2.660	4.110	2.820	2.240	42.010
32	2.290	1.940	2.410	1.320	0.790	0.080	1.980	2.790	2.730	5.860	2.580	1.590	26.380
33	0.160	1.260	0.910	1.850	2.410	4.530	6.980	1.950	8.010	2.700	2.180	2.050	34.990
34	0.510	2.550	5.590	2.900	4.590	3.340	4.770	2.250	2.190	7.030	2.610	7.430	46.480
35	3.360	1.320	1.250	5.070	4.860	1.870	7.400	2.530	1.240	4.270	4.290	2.870	40.330
36	1.760	3.600	3.620	1.440	1.480	3.980	4.650	6.280	3.570	2.830	4.390	1.010	38.610
37	3.890	4.960	5.690	1.810	0.520	5.080	3.970	9.530	1.650	7.200	2.940	4.140	51.280
38	1.340	4.330	5.450	7.510	1.200	1.760	4.320	4.190	5.800	5.410	3.130	2.180	46.620
MEAN	2.606	3.021	3.252	3.397	3.044	4.011	4.013	4.187	4.412	3.959	3.246	3.020	42.168
STDV	1.770	1.846	1.754	2.199	2.120	2.176	1.953	2.259	2.376	2.047	1.554	1.866	7.453

ตารางที่ 6.2 1st Difference of Square Root of Monthly Precipitation at Salamonia, Indiana

Year	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
1		0.236	-0.254	0.903	-1.225	0.372	-0.094	-0.710	0.797	0.078	0.168	0.571
2	-0.848	-0.329	1.175	-0.146	-1.210	1.198	-0.163	0.611	-2.513	0.653	0.165	0.782
3	-0.563	-0.528	0.764	-0.661	-0.193	0.998	-0.851	-0.062	1.708	-1.427	1.037	-0.196
4	-0.1681	0.918	0.092	0.423	-1.086	2.219	-1.530	1.046	-0.359	-0.953	0.523	-0.354
5	-0.215	0.391	-0.353	0.073	-0.225	0.853	0.117	-1.250	0.210	0.508	-0.617	0.672
6	1.069	-0.615	-0.352	1.015	-1.222	-0.508	1.171	-0.111	-0.007	-1.202	1.239	-0.424
7	0.704	-1.302	0.575	-0.145	-0.149	1.331	-1.447	0.880	-0.354	1.249	-1.279	0.017
8	-1.093	1.129	-0.432	0.704	-0.046	0.322	-0.023	-1.216	1.123	0.074	-1.045	-0.267
9	-0.152	0.301	-0.009	0.497	-0.437	-0.763	1.291	-0.483	0.359	-1.271	0.177	-0.177
10	-0.207	1.030	-0.491	-0.205	-0.602	-0.040	0.931	-0.840	1.629	-0.299	-0.323	-0.543
11	1.235	-1.341	0.316	0.110	0.162	0.212	-0.383	0.597	0.345	-0.498	-0.270	-0.574
12	0.020	1.052	-0.204	-1.299	0.470	1.255	-0.606	0.148	-0.801	0.474	-1.091	1.021
13	-0.283	-0.354	0.299	-0.304	0.912	0.482	-0.010	-0.274	-1.431	1.035	1.031	-0.780
14	-1.500	1.229	0.107	-0.513	1.269	-0.495	0.059	-0.233	0.868	-0.878	-0.377	0.907
15	-0.650	0.256	-0.533	-0.058	0.703	-0.265	-0.407	1.437	-1.038	0.507	-0.480	-1.157
16	0.966	0.425	-0.310	0.282	-1.495	0.744	1.593	-0.684	0.037	0.133	-0.439	-0.274
17	-0.319	0.106	-0.455	0.694	0.238	0.593	-0.125	-0.473	-0.051	-1.213	0.738	0.010
18	0.057	1.257	-0.482	-1.763	1.247	0.643	-1.208	0.766	1.102	-1.280	0.787	-0.955
19	0.453	-1.063	1.176	-0.620	1.017	-0.670	0.476	-0.189	0.292	-0.650	0.101	-0.117
20	-0.270	0.986	-1.332	1.457	-2.702	2.486	-0.717	-0.063	0.399	-0.820	0.349	0.118
21	-0.523	-1.229	2.408	-0.449	-0.157	0.480	-0.491	-0.158	0.237	-0.309	-0.289	0.681
22	-0.677	0.216	0.833	-0.132	-1.320	0.984	0.054	0.105	-0.504	1.169	-1.518	-0.745
23	0.658	-0.024	0.409	0.256	0.015	-0.408	0.564	-0.482	0.213	-0.443	1.181	-0.357
24	-0.025	-0.858	0.643	-0.239	1.380	-0.518	-0.829	0.160	0.238	0.247	-1.055	1.319
25	-0.025	-0.531	-0.942	0.576	1.269	-0.848	0.172	-0.207	0.384	-0.265	-0.103	-0.344
26	0.131	-0.382	0.319	0.406	-0.044	-0.659	1.333	-0.088	0.312	-0.823	-0.325	0.035
27	-0.026	1.057	-0.216	-0.992	-0.746	0.642	0.297	0.470	-0.093	0.671	-0.841	0.599
28	-1.442	0.596	-0.623	1.812	-1.090	-0.077	-0.567	0.797	-0.221	-0.062	0.077	-0.510
29	0.755	-0.454	0.414	-0.347	0.288	-0.915	0.276	0.732	1.046	-1.001	-0.485	-0.588
30	0.580	0.141	-0.313	-0.018	0.520	0.607	-0.618	1.002	-0.881	0.035	0.027	-1.092
31	0.699	0.255	0.054	0.089	0.413	-0.479	-1.116	1.223	-0.481	0.388	-0.348	-0.183
32	0.017	-0.113	0.152	-0.404	-0.260	-0.606	1.124	0.263	-0.018	0.768	-0.315	-0.345
33	-0.881	0.722	-0.169	0.405	0.192	0.576	0.514	-1.246	1.434	-1.187	-0.167	-0.045
34	-0.718	0.883	0.767	-0.651	0.439	-0.315	0.356	-0.684	0.206	0.948	-1.038	1.110
35	-0.893	-0.684	-0.031	1.134	-0.047	-0.837	1.353	-1.130	-0.477	0.953	0.005	-0.377
36	-0.397	0.571	0.005	-0.703	0.017	0.778	0.161	0.350	-0.617	-0.207	0.413	-1.090
37	0.967	0.255	0.158	-1.040	-0.624	1.533	-0.287	1.120	-1.803	1.399	-0.889	0.320
38	0.877	0.923	0.254	0.405	-1.645	0.231	0.752	-0.032	0.361	-0.082	-0.557	-0.293



รูปที่ 6.6 ACF ของ 1st Differenced Square Root ของฝนรายเดือนที่ Salamonia มลรัฐ Indiana



รูปที่ 6.7 Partial ACF ของ 1st Differenced Square Root ของฝนรายเดือนที่ Salamonia มลรัฐ Indiana

(e) การประมาณค่าพารามิเตอร์ $\hat{\theta}_1$ และ $\hat{\sigma}_\varepsilon^2$ โดยวิธีโมเมนต์ด้วยโปรแกรม FTMP5 ของ IMSL

$$\hat{\theta}_1 = 0.5873$$

$$\hat{\sigma}_\varepsilon^2 = 0.4207$$

(f) การประมาณค่าพารามิเตอร์โดยวิธี MLE

ค่า SSE ของ ARIMA(1,1,1) แสดงอยู่ในตารางที่ 6.3 ซึ่งแสดงให้เห็นว่า

$$\hat{\phi}_1 = 0.0 \text{ และ } \hat{\theta}_1 = 0.9$$

$$\text{Min SSE} = 154.39$$

จาก Initial Estimate จะสามารถหา Final Estimate ได้โดยใช้โปรแกรม FTMXL ดังนี้

$$\hat{\phi}_1 = 0.0451$$

$$\hat{\theta}_{00} = -0.0010$$

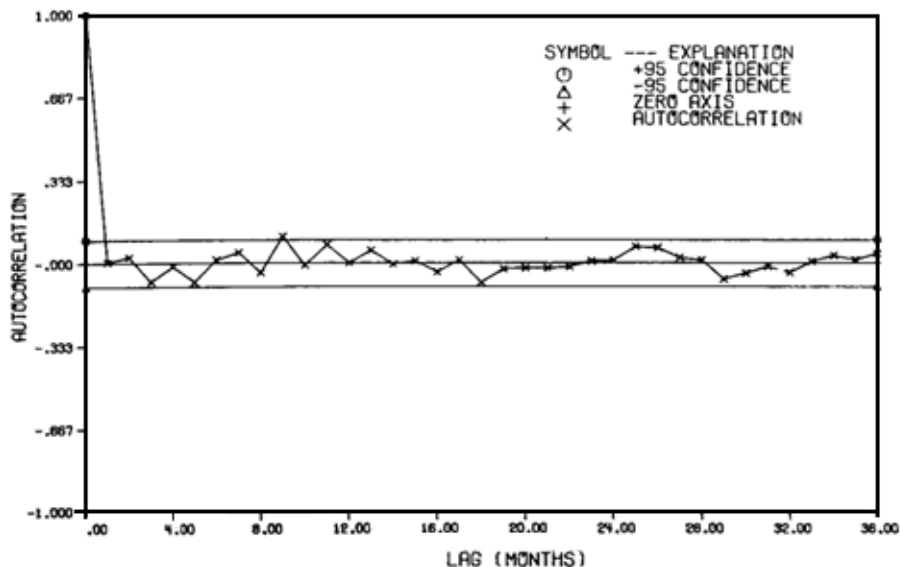
$\hat{\theta}_1 = 0.9593$ กรณีที่ $\hat{\theta}_1$ มีค่าเข้าใกล้ 1 อาจพิจารณาว่า ARIMA(1,1,1) ไม่เหมาะสมกับอนุกรมเวลาของฝนรายเดือนก็ได้

$$\hat{\sigma}_\varepsilon^2 = 0.3332$$

(g) การทดสอบ Goodness of Fit

Porte Manteau Lack of Fit Test

$Q = 0.0448 \times 455 = 20.06 < \chi_{22,0.05}^2 = 33.9$ แสดงว่าอนุกรมเวลาของ ε_t เป็น Independent และ ACF ของ ε_t ในรูปที่ 6.8 แสดงให้เห็นว่า $r_\varepsilon(k)$ ไม่มีนัยสำคัญ ซึ่งเป็นการยืนยันข้อสรุปที่ว่า ε_t เป็น Independent



รูปที่ 6.8 ACF ของ Residuals ของ 1st Differenced Square Root ของฝนรายเดือนที่ Salamonia มลรัฐ Indiana

ตารางที่ 6.3 Sum of Squares of Residuals of 1st Difference of Square Root of Monthly Rainfall Series at Salamonia, Indiana

THETA	PHI										
	-0.50	-0.40	-0.30	-0.20	-0.10	0.00	0.10	0.20	0.30	0.40	0.50
0.00	211.59	213.42	220.76	233.62	252.00	275.89	305.31	340.24	380.68	426.65	478.13
0.10	207.60	206.39	210.22	219.08	232.99	251.93	275.90	304.92	338.97	378.06	422.18
0.20	204.82	201.04	201.91	207.42	217.57	232.37	251.82	275.91	304.64	338.03	376.05
0.30	202.95	196.95	195.27	197.91	204.87	216.15	231.75	251.67	275.91	304.47	337.35
0.40	201.71	193.77	189.88	190.04	194.24	202.48	214.77	231.11	251.49	275.92	304.39
0.50	200.78	191.15	185.33	183.32	185.12	190.73	200.15	213.38	230.42	251.27	275.93
0.60	199.77	188.68	181.20	177.32	177.04	180.36	187.29	197.82	211.95	229.68	251.02
0.70	198.27	185.97	177.09	171.62	169.57	170.92	175.69	183.80	195.47	210.48	228.90
0.80	196.14	182.87	172.83	166.02	162.46	162.13	165.04	171.19	180.57	193.19	209.05
0.90	194.51	180.33	169.22	161.20	156.25	154.39	155.60	159.90	167.27	177.73	191.27
1.00	321.86	297.38	276.21	258.34	243.77	232.50	224.54	219.88	218.53	220.48	225.73

6.12 ตัวอย่างการสร้างแบบจำลอง Multiplicative ARIMA

Blue River ใกล้ White Cloud มลรัฐ Indiana มีพื้นที่ลุ่มน้ำ 461 ตารางไมล์ มีอนุกรมเวลาของน้ำท่ารายเดือน ($x_{v,\tau}$) ดังแสดงในตารางที่ 6.4

$$\text{ให้ } y_{v,\tau} = \text{Log}(x_{v,\tau}); v = 1, \dots, 38; \tau = 1, \dots, 12$$

$$z_{v,\tau} = \frac{y_{v,\tau} - \bar{y}_\tau}{s_\tau}$$

ตารางที่ 6.4 อนุกรมเวลาของน้ำท่ารายเดือนของ Blue River Near White Cloud, Indiana

Year	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
1	181	507	922	2776	1158	613	755	467	308	309	246	307
	146	538	1132	2676	1093	1819	2185	2163	214	112	56	107
	128	93	225	207	78	752	263	91	120	131	79	88
	46	23	58	385	208	2197	952	1569	648	201	115	73
5	55	92	128	293	503	709	1126	150	41	58	34	52
	407	761	273	6290	1088	396	659	1299	247	86	224	43
	260	114	563	631	1053	1780	1111	572	245	515	510	177
	28	52	63	763	2185	1973	2050	268	489	348	203	64
	30	28	26	173	1098	558	1770	453	182	67	38	24
10	19	45	74	210	141	71	421	94	261	213	88	51
	68	170	320	655	1061	1155	1106	453	816	304	87	43
	25	396	860	486	595	2549	680	723	414	218	67	24
	18	23	21	57	106	1132	1549	639	145	62	239	419
	66	53	191	667	1969	4299	1379	533	999	335	69	211
15	420	485	599	830	1898	1197	361	1656	425	160	307	44
	33	169	532	1415	426	275	2422	966	892	467	174	56
	38	65	99	353	891	1882	2277	396	98	363	98	79
	39	1209	1267	3178	1699	2334	575	423	1545	233	386	130
	249	86	922	4504	3404	1084	1781	973	885	198	138	314
20	117	1028	1095	2344	1990	2154	1609	375	179	109	61	105
	41	557	2066	1511	1633	1999	541	217	296	69	33	50
	22	33	193	1124	469	1214	815	1030	161	292	55	19
	15	22	24	123	170	303	384	189	181	45	71	249
	237	69	334	553	2155	2061	639	382	290	328	83	162
25	515	589	230	350	3149	1374	639	640	334	143	88	75
	45	36	109	527	1284	560	1979	1771	745	286	106	55
	37	1337	2417	797	527	861	824	944	19	879	800	384
	195	258	563	2664	1288	889	626	437	206	70	57	71
	69	137	520	725	1004	779	477	253	1882	771	85	42
30	24	39	56	162	603	2302	1368	2939	395	352	127	70
	49	140	616	1490	2019	1471	630	253	152	73	36	107
	70	104	110	216	200	2575	443	246	125	216	101	57
	18	20	18	112	151	4058	885	316	553	145	30	22
	14	28	281	583	1279	1153	902	248	86	302	57	541
35	247	116	99	1389	1557	725	1546	1068	137	89	101	36
	49	161	671	383	554	1349	543	1583	192	91	146	42
	51	140	1308	539	646	1237	1463	1511	428	121	216	75
	30	98	1028	2032	1033	376	1127	905	358	412	189	78

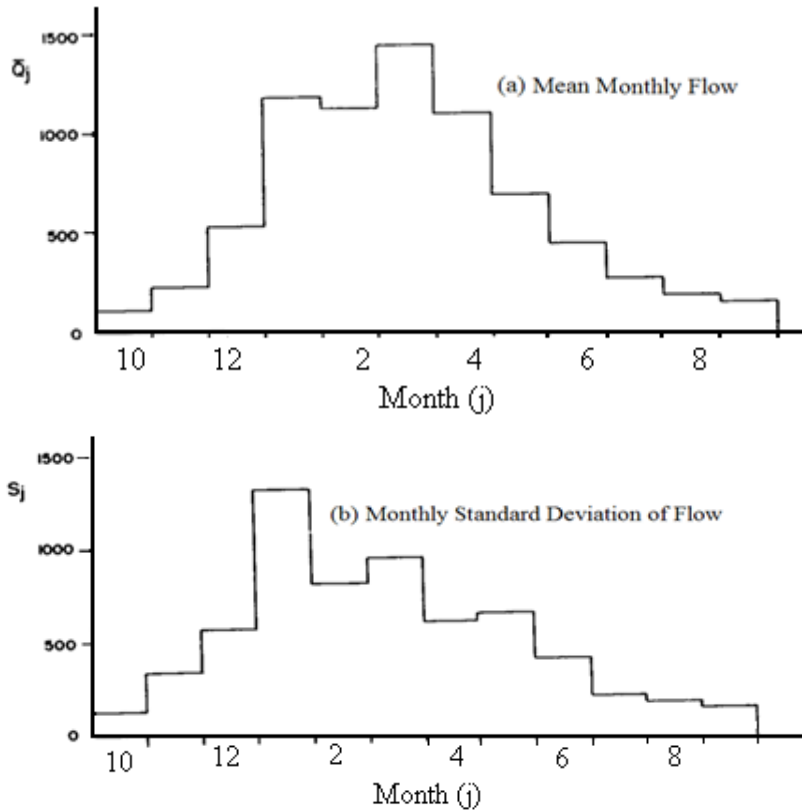
จากการศึกษารูปแบบของแบบจำลองสโตแคสติกของอนุกรมเวลาน้ำท่ารายเดือน ($z_{v,t}$) จำนวน 16 สถานีใน Indiana พบว่า AR(2) พัดกับข้อมูลทั้ง 16 สถานี ดังนั้นจึงเลือก AR(2) สำหรับการศึกษารูปแบบเปรียบเทียบกับ Multiplicative ARIMA ซึ่งมีจำนวนพารามิเตอร์น้อยกว่า AR(2) มี 27 พารามิเตอร์ 12 Mean, 12 Standard Deviation, ϕ_1 , ϕ_2 , σ_ϵ^2

ขั้นที่ 1 การแปลงข้อมูล(Transformation)

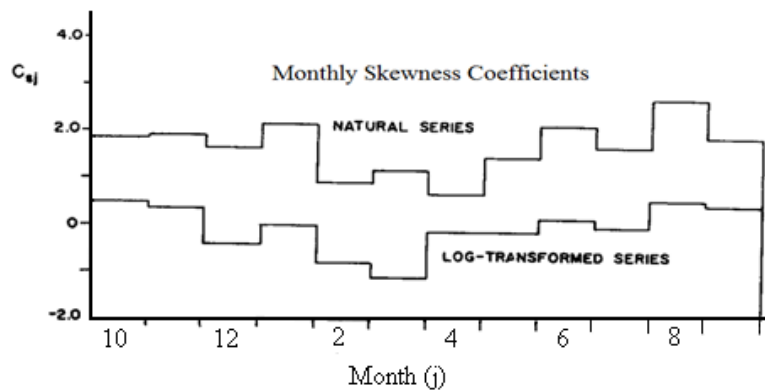
ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานรายเดือนของปริมาณน้ำท่าแสดงอยู่ในรูปที่ 6.9 และค่าสัมประสิทธิ์ความบิดเบี้ยวรายเดือนของลี้กของปริมาณน้ำท่าแสดงอยู่ในรูปที่ 6.10 ซึ่งแสดงให้เห็นว่า C_{st}

ของ $y_{v,\tau} = \text{Log}(x_{v,\tau})$ มีค่าเข้าใกล้ 0 และผลการพล็อต $x_{v,\tau}$ ในกระดาษกราฟ Log-Normal ในรูปที่ 6.11 พบว่ากราฟเป็นเส้นตรง แสดงว่า $x_{v,\tau}$ เป็น Log-Normal

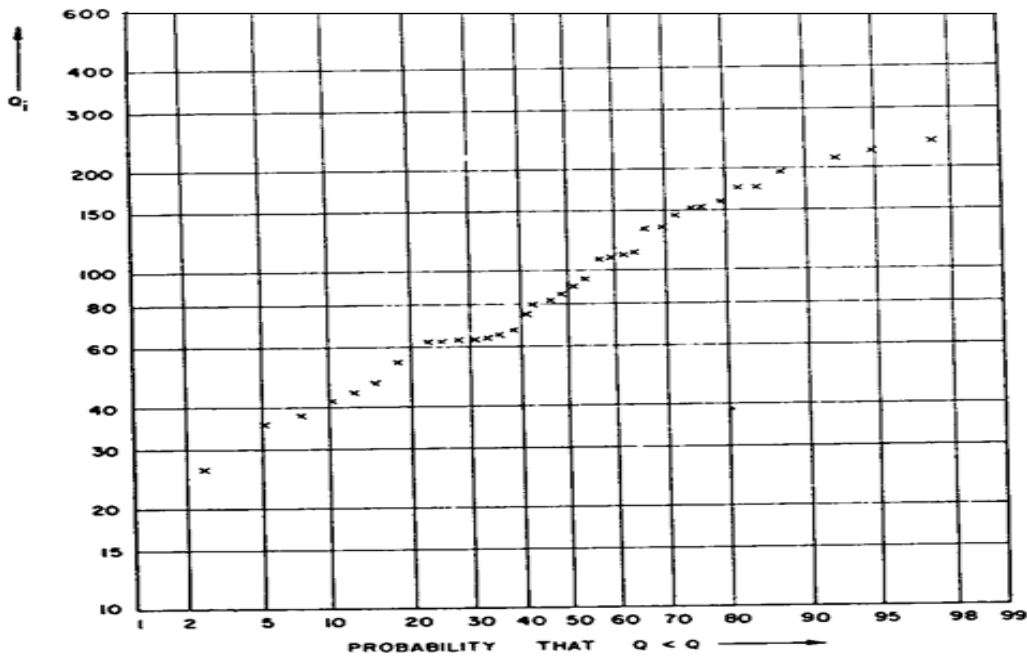
Correlogram ในรูปที่ 6.12 แสดงว่าอนุกรมเวลาเป็น Stationary แต่ r_j ไม่เท่ากับศูนย์ แสดงว่าอนุกรมเวลาเป็นแบบ Dependent



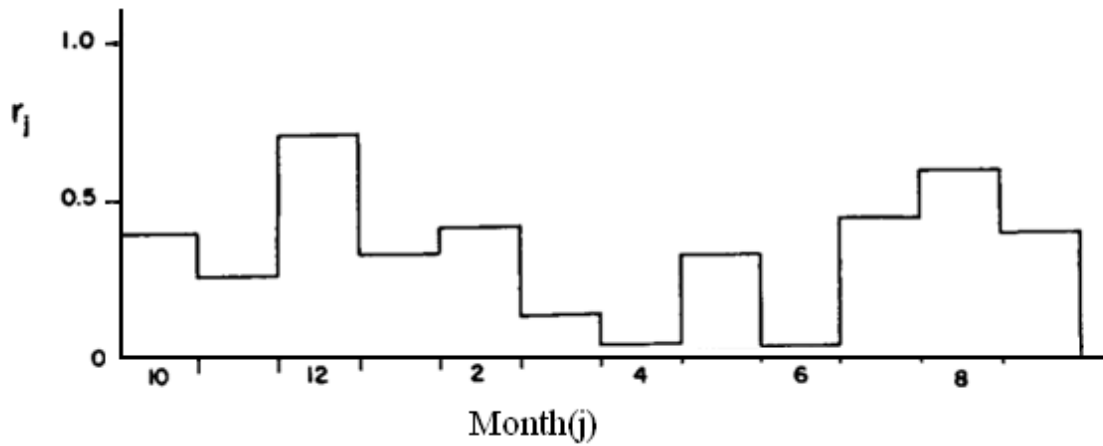
รูปที่ 6.9 (a) ค่าเฉลี่ยรายเดือน (b) ส่วนเบี่ยงเบนมาตรฐานรายเดือนของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana



รูปที่ 6.10 ค่าสัมประสิทธิ์ความเบ้ (Skewness Coefficients) รายเดือนของปริมาณการไหลของน้ำรายเดือน และของ Log-Transformed ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana



รูปที่ 6.11 Lognormal Probability Plot ของ Logarithms ของปริมาณการไหลของน้ำในเดือนเมษายนของ Blue River ใกล้อ White Cloud มลรัฐ Indiana



รูปที่ 6.12 ค่าสหสัมพันธ์ (Serial Correlations) รายเดือนของ Log-Transformed ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้อ White Cloud มลรัฐ Indiana

ขั้นที่ 2 Differencing และ Identification

ทำการ Differencing $y_{v,\tau} = \text{Log}(x_{v,\tau})$ แบบ Nonseasonal Differencing เมื่อ $d=0, 1, 2$ และ

Seasonal Differencing ($w=12$ months) เมื่อ $D=0, 1, 2$ ดังนี้

$$w_t = (1 - B)^d (1 - B^{12})^D y_t$$

$$y_t = \ln(x_t)$$

$$N = 456$$

ACF(r_k) ของอนุกรมเวลาที่ถูกรับ Differenced แสดงอยู่ในรูปที่ 6.13 จะเห็นได้จาก ACF ว่าเมื่อ $d=1$ และ $D=1$ จะสามารถกำจัด Cyclicity ได้หมดเมื่อเปรียบเทียบกับ กรณี $d=0$ และ $D=0$ จากรูปจะเห็นได้ว่า r_k ที่ $Lag=1, 2, 11, 12, 13$ มีค่าต่างจากศูนย์อย่างมีนัยสำคัญ และ Lag 1 และ 12 มีความสำคัญกว่า Lag อื่นๆ ซึ่งแสดงถึง Seasonal และ Nonseasonal Moving Average หรืออาจสรุปได้ว่าแบบจำลองที่เหมาะสมคือ

Multiplicative ARIMA(0,1,1)x(0,1,1)₁₂ หรือ

แต่เมื่อพิจารณาจาก Monthly Correlation ของ Nonseasonal Component แบบจำลองที่เหมาะสมคือ

Multiplicative ARIMA(1,0,0)x(0,1,1)₁₂ หรือ

Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂

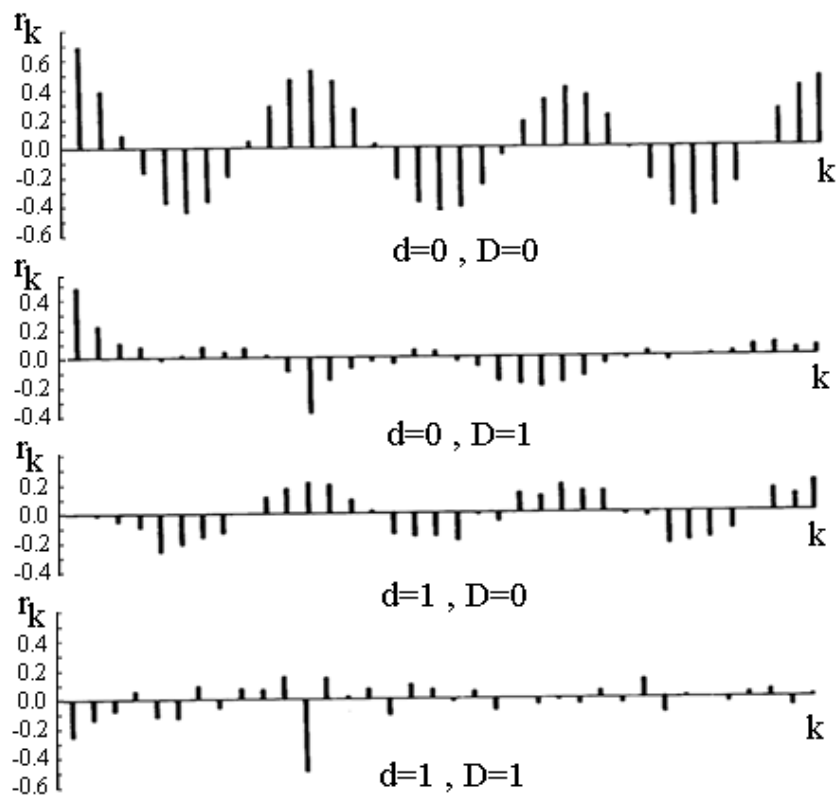
แต่จากประสบการณ์ของผู้จำลองจะเลือกวิเคราะห์ 2 แบบจำลองคือ

Multiplicative ARIMA(0,1,1)x(0,1,1)₁₂ และ Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂

ขั้นที่ 3 การหาพารามิเตอร์ของแบบจำลอง Multiplicative ARIMA(0,1,1)x(0,1,1)₁₂ โดยวิธี MLE

General Form ของ Multiplicative ARIMA(p,d,q)x(P,D,Q)_w

$$\phi(B)\Phi(B^w)(1-B)^d(1-B^w)^D y_t = \theta(B)\Theta(B^w)\varepsilon_t$$



รูปที่ 6.13 ค่าประมาณของ ACF ของ Differenced Series ของ Blue River ใกล้ White Cloud มลรัฐ

Indiana $w_t = (1-B)^d(1-B^{12})^D y_t$; $y_t = \ln(x_t)$, $N=456$

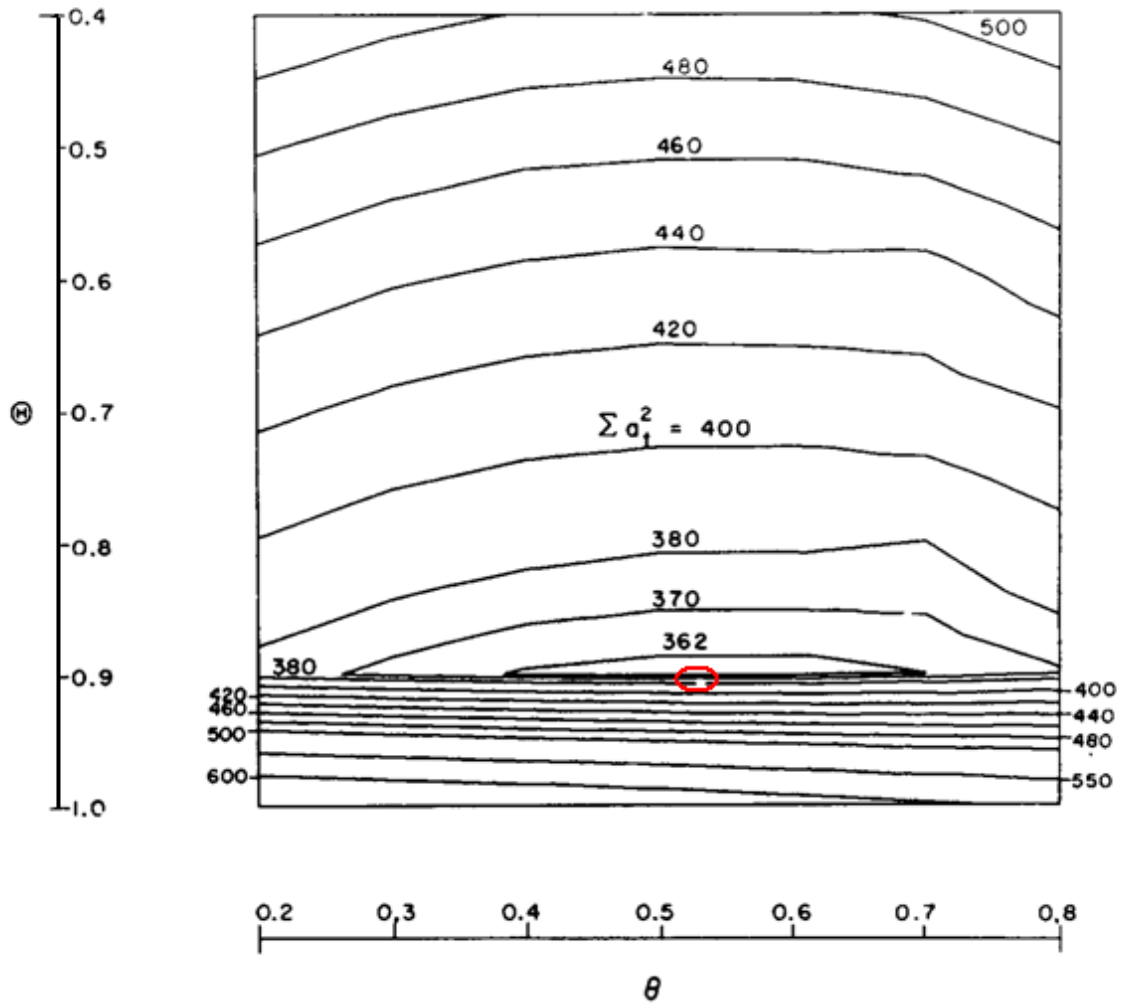
(3a) Multiplicative ARIMA(0,1,1)x(0,1,1)₁₂

$$(1 - B)(1 - B^{12})y_t = (1 - \theta_1 B)(1 - \theta_1 B^{12})\varepsilon_t$$

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_1 \varepsilon_{t-12} + \theta_1 \theta_1 \varepsilon_{t-13}$$

ผลการคำนวณ Sum Square Surface หรือ SSS = $\sum_{t=1}^N \varepsilon_t^2(\theta_1, \theta_1)$ เมื่อ $0 \leq \theta_1 \leq 0.8$ และ

$0.4 \leq \theta_1 \leq 1.0$ แสดงอยู่ในรูปที่ 6.14 ซึ่งพบว่า Minimum SSS อยู่ที่ $\theta_1=0.55$ และ $\Theta_1=0.9$



รูปที่ 6.14 Sum of Square Surface $\sum \varepsilon_t^2(\Theta_1, \theta_1)$ ของแบบจำลอง Multiplicative

ARIMA(0,1,1)x(0,1,1)₁₂ ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

(3b) การหาพารามิเตอร์ของแบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ โดยวิธี MLE

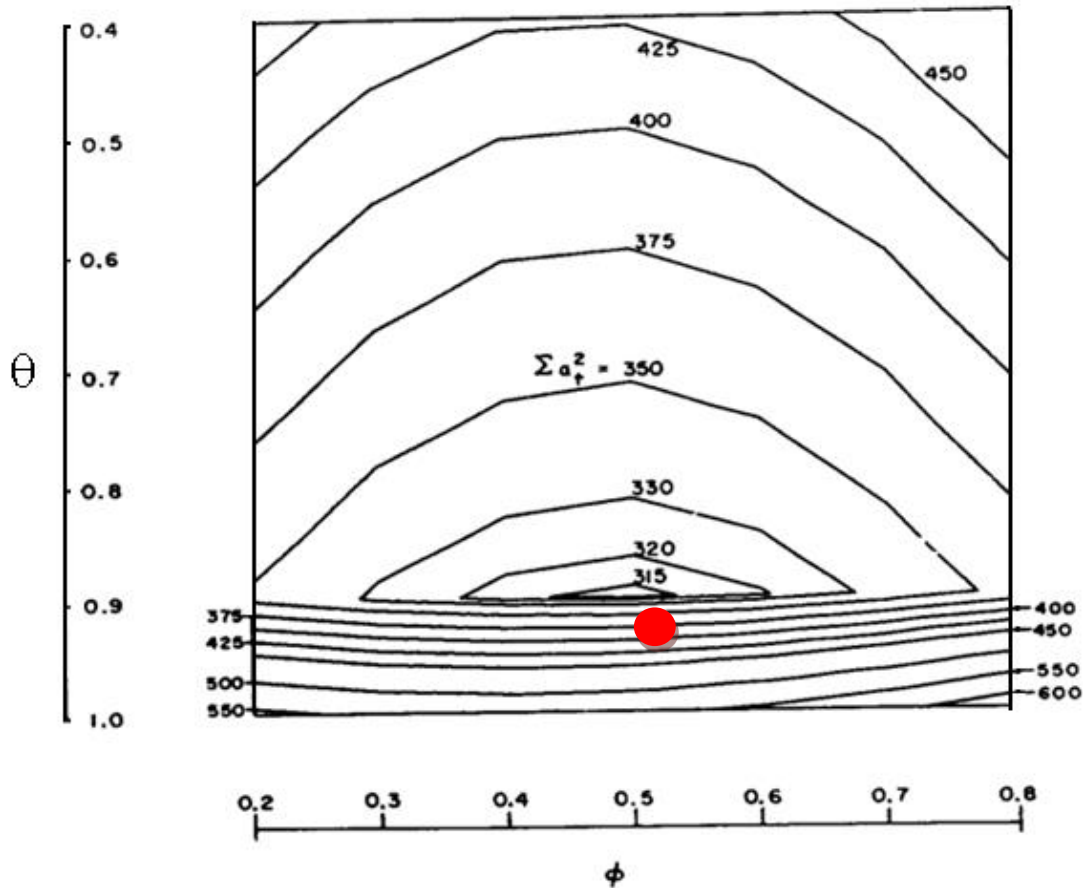
General Form ของ Multiplicative ARIMA(p,d,q)x(P,D,Q)_w

$$\phi(B)\phi(B^w)(1 - B)^d(1 - B^w)^D y_t = \theta(B)\theta(B^w)\varepsilon_t$$

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})y_t = (1 - \theta_1 B^{12})\varepsilon_t$$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + y_{t-12} - \phi_1 y_{t-13} - \phi_2 y_{t-14} + \varepsilon_t - \theta_1 \varepsilon_{t-12}$$

$SSS = \sum_{t=1}^N \varepsilon_t^2 (\phi_1, \phi_2, \theta_1)$ เมื่อ $0.2 \leq \phi_1 \leq 0.8, \phi_2 = 0.1$ และ $0.4 \leq \theta_1 \leq 1.0$ แสดงอยู่ในรูปที่ 6.15 ซึ่งพบว่า Minimum SSS อยู่ที่ $\phi_1=0.5, \phi_2=0.1$ และ $\theta_1=0.9$ แต่เมื่อเปลี่ยนค่า ϕ_2 เป็น 0 และ 0.2 SSS มีค่าไม่เปลี่ยนแปลง



รูปที่ 6.15 Sum of Square Surface $\sum \varepsilon_t^2(\phi_1, \phi_2, \theta_1)$ ของแบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ เมื่อ $\phi_2=0.1$ ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

(3c) ประมาณพารามิเตอร์แบบจำลองโดยวิธี **Nonlinear Iterative Estimation** ได้ค่าพารามิเตอร์ดัง

ตาราง

Multiplicative ARIMA(0,1,1)x(0,1,1) ₁₂		
Iterative	θ_1	Θ_1
Initial Estimate	0.55	0.9
Final Estimate	0.549±0.040	0.942±0.009

ตารางที่ 6.5 แสดงค่าพารามิเตอร์ของแบบจำลอง ARIMA(2,0,0)x(0,1,1)₁₂

ตารางที่ 6.5 Multiplicative ARIMA (2,0,0)x(0,1,1)₁₂ Applied to the Natural Logarithm of Monthly Flows, Blue River Near White Cloud, Indiana

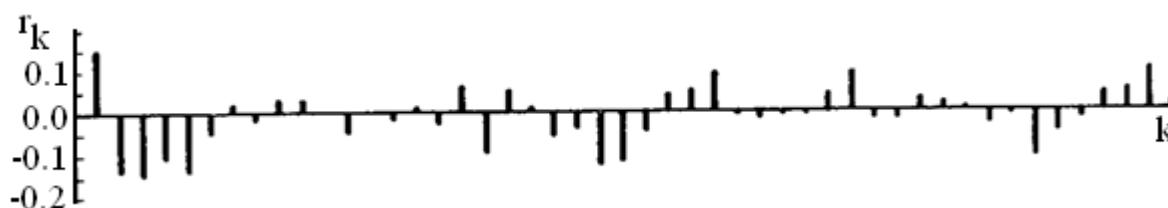
SUMMARY OF MODEL 1					
DATA - X = BLUE RIVER NEAR WHITE CLOUD			456 OBSERVATIONS		
DIFFERENCING ON X - 1 OF ORDER 12					
Model developed with transformed data = $\log(x(t)) = y(t)$					
Parameters				95 percent	
Number	Type	Order	Value	Lower limit	Upper limit
1	AUTOREGRESSIVE 1	1	5.55343E-01	4.59522E-01	6.51164E-01
2	AUTOREGRESSIVE 1	12	9.38650E-03	-0.8638E-01	1.05161E-01
3	MOVING AVERAGE 1	12	8.98802E-01	8.56597E-01	9.41007E-01
Other information and results					
Residual sum of squares = 3.28623E+02					
Residual mean square = 7.48572E-01					
Degree of freedom = 439					
No. of residuals = 442					
Residual standard error = 8.65201E-01					

(3d) Diagnostic Check

ACF ของ ϵ_t ของ ARIMA(0,1,1)x(0,1,1)₁₂ แสดงอยู่ในรูปที่ 6.16 ซึ่ง Lag 2-Lag 6 มีค่าติดลบและมีค่าต่างจากศูนย์อย่างมีนัยสำคัญ เมื่อ $StandardError = \pm \frac{1}{\sqrt{N}} = \pm \frac{1}{\sqrt{456}} = \pm 0.047$

$$Q = N \sum_{k=1}^{24} r_k^2(\epsilon) = 71.3 > \chi_{22,0.1}^2 = 30.8$$

ซึ่งสามารถสรุปได้ว่า ϵ_t ไม่ใช่ Random Series หรือเป็น Independent จึงไม่สามารถใช้ ARIMA(0,1,1)x(0,1,1)₁₂ ได้ ยิ่งกว่านั้น ACF ในรูปที่ 6.16 แสดงให้เห็นว่า Seasonal Component ถูกกำจัดหมด ดังนั้น Seasonal Part ของแบบจำลองจึงน่าจะใช้ได้ แต่ Nonseasonal Part ยังใช้ไม่ได้



รูปที่ 6.16 ACF ของ Residuals ของแบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ (Standard Error = $\pm 1/N^{1/2}$) ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

คำนวณหา ϵ_t ของ Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ดังแสดงในตารางที่ 6.6 ACF และ PACF ของ ϵ_t ของ Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ดังแสดงในตารางที่ 6.7 ซึ่งแสดงให้เห็นว่า ϵ_t เป็น Independent

Q ของ 30 Lag แรก = 17.30 < $\chi_{21,0.1}^2 (=29.6)$ ดังนั้นจึงยอมรับ ARIMA(2,0,0)x(0,1,1)₁₂

ตารางที่ 6.6 Residuals ของแบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

t	Fitted Value	Residual	Data Value
412	5.8713E+00	1.3651E+00	7.2363E+00
413	7.1448E+00	2.0571E-01	7.3505E+00
414	7.6022E+00	-1.0161E+00	6.5852E+00
415	6.3625E+00	9.8092E-01	7.3434E+00
416	6.4865E+00	4.8702E-01	6.9735E+00
417	6.1258E+00	-1.2059E+00	4.9200E+00
418	4.9399E+00	-4.5123E-01	4.4886E+00
419	3.9698E+00	6.4532E-01	4.6151E+00
420	4.6154E+00	-1.0319E+00	3.5835E+00
421	3.5533E+00	3.3856E-01	3.8918E+00
422	4.5199E+00	5.6155E-01	5.0814E+00
423	5.5856E+00	9.2314E-01	6.5088E+00
424	7.0752E+00	-1.1272E+00	5.9480E+00
425	6.4681E+00	-1.5097E-01	6.3172E+00
426	6.9134E+00	2.9367E-01	7.2071E+00
427	6.7969E+00	-4.9981E-01	6.2971E+00
428	5.9606E+00	1.4065E+00	7.3671E+00
429	6.2125E+00	-9.5504E-01	5.2575E+04
430	5.0853E+04	-5.7447E-01	4.5109E+00
431	4.0506E+00	9.3299E-01	4.9836E+00
432	4.7158E+00	-9.7815E-01	3.7377E+00
433	3.6766E+00	2.5524E-01	3.9318E+00
434	4.6003E+00	3.4130E-01	4.9416E+00
435	5.6018E+00	1.5744E+00	7.1753E+00
436	7.3305E+00	-1.0408E+00	6.2897E+00
437	6.6489E+04	-1.7807E-01	6.4708E+00
438	7.0317E+00	8.8749E-02	7.1204E+00
439	6.6996E+04	5.8860E-01	7.2882E+00
440	6.6525E+00	8.6801E-01	7.3205E+00
441	6.0993E+00	-4.0216E-02	6.0591E+00
442	5.4719E+04	-6.7615E-01	4.7958E+00
443	4.3108E+00	1.0645E+00	5.3753E+00
444	4.8370E+00	-5.1953E-01	4.3175E+00
445	4.0281E+00	-6.2689E-01	3.4012E+00
448	4.3456E+00	2.3932E-01	4.5850E+00
447	5.5581E+00	1.3773E+00	6.9354E+00
448	7.0881E+00	5.2869E-01	7.6168E+00
449	7.3656E+00	-4.2534E-01	6.9402E+00
450	7.3138E+00	-1.3842E+00	5.9296E+00
451	6.1023E+00	9.2503E-01	7.0273E+00
452	6.5840E+00	2.4390E-01	6.8079E+00
453	5.8082E+00	7.2378E-02	5.8805E+00
454	5.2995E+04	7.2150E-01	6.0210E+00
455	5.0973E+04	1.4448E-01	5.2417E+00
456	4.7218E+00	-3.6508E-01	4.3567E+00

ตารางที่ 6.7 ACF และ PACF ของ Residuals ของ Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ที่ประยุกต์ใช้
กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ
Indiana

1	AUTOCORRELATION FUNCTION DATA - THE ESTIMATED RESIDUALS - MODEL 1 ORIGINAL SERIES MEAN OF THE SERIES = -0.7050E-01 ST. DEV. OF SERIES = 8.60346E-01 NUMBER OF OBSERVATIONS = 442												442 OBSERVATIONS
	1 - 12	-0.01	-0.03	0.01	0.04	-0.04	0.02	0.07	-0.01	0.05	0.05	0.05	0.01
	ST.E.	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	13 - 24	0.06	0.00	0.04	-0.04	0.04	0.05	0.01	0.01	-0.06	0.02	-0.06	-0.05
	ST.E.	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	MEAN DIVIDED BY ST. ERROR = 1.72295E+00 TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE 1.73010E+01 SHOULD BE COMPARED WITH A CHI - SQUAR VARIABLE WITH 21 DEGREES OF FREEDOM												
1	THE ESTIMATED RESIDUALS - MODEL 1												
1	PARTIAL AUTOCORRELATIONS DATA - THE ESTIMATED RESIDUALS - MODEL 1 ORIGINAL SERIES MEAN OF THE SERIES = -0.7050E-01 ST. DEV. OF SERIES = 8.60346E-01 NUMBER OF OBSERVATIONS = 442												442 OBSERVATIONS
	1 - 12	-0.01	-0.03	0.01	0.04	-0.04	0.02	0.06	-0.01	0.05	0.05	0.05	0.02
	13 - 24	0.06	0.00	0.05	-0.04	0.03	0.04	-0.00	0.00	-0.08	0.01	-0.07	-0.07

(3e)การพยากรณ์(Forecasting)

แบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + y_{t-12} - \phi_1 y_{t-13} - \phi_2 y_{t-14} + \varepsilon_t - \theta_1 \varepsilon_{t-12}$$

ฟังก์ชันพยากรณ์ (Forecasting Function) ซึ่งมี Lead Time =L คือ

$$y_t(L) = [y_{t+L}] = \phi_1 [y_{t+L-1}] + \phi_2 [y_{t+L-2}] + [y_{t+L-12}] - \phi_1 [y_{t+L-13}] - \phi_2 [y_{t+L-14}] + [\varepsilon_{t+L}] - \theta_1 [\varepsilon_{t+L-12}]$$

กรณีที่ L มีค่าต่างๆ จะสามารถเขียนฟังก์ชันพยากรณ์ได้ดังนี้

$$L = 1 \quad : y_1(1) = \phi_1 y_t + \phi_2 y_{t-1} + y_{t-11} - \phi_1 y_{t-12} - \phi_2 y_{t-13} - \theta_1 \varepsilon_{t-11}$$

$$L = 2 \quad : y_t(2) = \phi_1 y_t(1) + \phi_2 y_t + y_{t-10} - \phi_1 y_{t-11} - \phi_2 y_{t-12} - \theta_1 \varepsilon_{t-10}$$

$$L = 3 \dots 12: y_t(L) = \phi_1 y_t(L-1) + \phi_2 y_t(L-2) + y_{t+L-12} - \phi_1 y_{t+L-13} - \phi_2 y_{t+L-14} - \theta_1 \varepsilon_{t+L-12}$$

$$L = 13 \quad : y_t(13) = \phi_1 y_t(12) + \phi_2 y_t(11) + y_t(1) - \phi_1 y_t - \phi_2 y_{t-1}$$

$$L = 14 \quad : y_t(14) = \phi_1 y_t(13) + \phi_2 y_t(12) + y_t(2) - \phi_1 y_t(1) - \phi_2 y_t$$

$$L = L \dots \quad : y_t(L) = \phi_1 y_t(L-1) + \phi_2 y_t(L-2) + y_t(L-12) - \phi_1 y_t(L-13) - \phi_2 y_t(L-14)$$

ตัวอย่างการหาค่าพยากรณ์ที่ t=456 และ L=1 โดยใช้ค่า y_t และ ε_t จากตารางที่ 6.6 จะสามารถหาค่า
พยากรณ์ $y_{456}(1)$ ได้ดังนี้

$$\begin{aligned}
 y_{456}(1) &= \phi_1 y_{456} + \phi_2 y_{455} + y_{445} - \phi_1 y_{444} - \phi_2 y_{443} - \theta_1 \varepsilon_{445} \\
 &= 0.55534 \times 4.3567 + 0.0093865 \times 5.2417 + 3.4012 - \\
 &\quad 0.55534 \times 4.3175 - 0.0093856 \times 5.3753 + 0.898802 \times \\
 &\quad 0.62689 = 3.9852
 \end{aligned}$$

ตัวอย่างการค่าพยากรณ์ที่ $t=456$ และ $L=14$ โดยใช้ค่า y_t และ ε_t จากตารางที่ 6.6 และค่าพยากรณ์ที่ $L < 14$ จากตารางที่ 6.8 จะสามารถหาค่าพยากรณ์ $y_{456}(14)$ ได้ดังนี้

$$\begin{aligned}
 y_{456}(14) &= \phi_1 y_{456}(13) + \phi_2 y_{456}(12) + y_{446}(2) - \phi_1 y_{456}(1) - \\
 &\quad \phi_2 y_{456} \\
 &= 0.55534 \times 3.9814 + 0.009386 \times 4.3596 + 4.6945 - \\
 &\quad 0.55534 \times 3.9852 - 0.009386 \times 4.3567 = 4.6945
 \end{aligned}$$

เมื่อเปรียบเทียบผลการพยากรณ์ที่เวลา $t=456$ ในตารางที่ 6.8 กับค่าเฉลี่ยรายเดือนพบว่า มีค่าใกล้เคียงกัน ดังนี้

L	$y_{456}(L)$	\bar{y}_L	% Difference
1	3.9852	4.1443	-3.84
2	4.6945	4.8189	-2.58
3	5.7638	5.5837	+3.23
4	6.492	6.503	-0.14
5	6.6869	6.6632	+0.36
6	7.0225	7.0079	+0.21
etc...			

การหาค่าความแปรปรวนของค่าความคลาดเคลื่อนในการพยากรณ์ (Variance of Forecast Error)

$$S_y^2(L) = (1 + \psi_1^2 + \dots + \psi_{L-1}^2) \sigma_\varepsilon^2 \quad [6.46]$$

เมื่อ

$$S_y^2(L) = \text{ความแปรปรวนของค่าความคลาดเคลื่อนในการพยากรณ์ที่ Lead Time=L}$$

$\psi_j = \psi$ -weight

จากอนุกรมเวลาแบบไม่มีที่สิ้นสุดของ Moving Average

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B) \varepsilon_t \quad [6.47]$$

เมื่อแทนค่า y_t ในสมการ [6.47] ลงในแบบจำลอง Multiplicative ARIMA(p,d,q)x(P,D,Q)_w ดัง

สมการ

$$\phi(B)\phi(B^w)(1-B)^d(1-B^w)^D y_t = \theta(B)\theta(B^w)\varepsilon_t$$

จะได้

$$\phi(B)\phi(B^w)(1-B)^d(1-B^w)^D\psi(B) = \theta(B)\theta(B^w) \quad [6.48]$$

จะสามารถคำนวณหา ψ_j ได้จากสมการ [6.48] โดยการเทียบค่าสัมประสิทธิ์ของ B ที่มีกำลังเท่ากัน สำหรับแบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ จะสามารถหา ψ_j ได้ดังนี้

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) &= 1 - \theta_1 B^{12} \\ \text{or} \\ (1 - \phi_1 B - \phi_2 B^2 - B^{12} + \phi_1 B^{13} + \phi_2 B^{14})(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) &= 1 - \theta_1 B^{12} \\ \text{Thus } \psi_0 &= 1 \\ \psi_1 &= \phi_1 = 0.555343 \\ \psi_2 &= \phi_1 \psi_1 = 0.555343 \times 0.555343 + 0.0093856 = 0.31779 \\ \psi_3 &= \phi_1 \psi_2 + \phi_2 \psi_1 = 0.555343 \times 0.31779 + 0.0093865 \times 0.555343 = \\ &= 0.1817 \\ &\cdot \\ &\cdot \\ &\cdot \\ \psi_{11} &= \phi_1 \psi_{10} + \phi_2 \psi_9 \\ \psi_{12} &= \phi_1 \psi_{11} + \phi_2 \psi_{10} + 1 - \theta_1 \\ \psi_{13} &= \phi_1 \psi_{12} + \phi_2 \psi_{11} + \psi_1 - \phi_1 \\ &= \phi_1 \psi_{12} + \phi_2 \psi_{11} \\ \psi_j &= \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \psi_{j-12} - \phi_1 \psi_{j-13} - \phi_2 \psi_{j-14}, \quad j \geq 14 \end{aligned}$$

ค่า ψ_j เมื่อ $j=1, \dots, 24$ แสดงอยู่ในตารางที่ 6.9

ค่า Standard Error ของค่าพยากรณ์เมื่อ $L=3$ จะเท่ากับ

$$\begin{aligned} S_y^2(3) &= (1 + \psi_1^2 + \psi_2^2)\sigma_\varepsilon^2 \\ S_y^2(3) &= (1 + 0.555343^2 + 0.31779239^2)0.865201^2 = (1.0271)^2 \\ S_y(3) &= 1.0271 \end{aligned}$$

ค่า Standard Error ในการพยากรณ์จะเพิ่มขึ้นอย่างช้าๆ เช่น

$$\begin{aligned} S_y(12) &= 1.048 \\ S_y(24) &= 1.054 \end{aligned}$$

ที่ $L=24$ ค่า Standard Error ในการพยากรณ์มีค่าเกือบคงที่ ซึ่งแสดงว่าค่าส่วนเบี่ยงเบนมาตรฐาน

ของปริมาณการไหลรายเดือนไม่มีผลต่อค่าพยากรณ์

ตัวอย่างการคำนวณหาช่วงความเชื่อมั่นที่ 95% ของค่าพยากรณ์ $y_{456}(3)$

$$95\% \text{ Confident Limit ของ } y_{456}(3) = y_{456}(3) + u_{0.025} S_y(3)$$

$$\text{Upper } y_{456}(3) = 5.7638 + 1.96 \times 1.0271 = 7.77$$

$$\text{Lower } y_{456}(3) = 5.7638 - 1.96 \times 1.0271 = 3.75$$

ผลการคำนวณช่วงความเชื่อมั่นที่ 95% ของค่าพยากรณ์ $y_{456}(L)$ เมื่อ $L=1, \dots, 24$ แสดงอยู่ในตารางที่

ตารางที่ 6.8 การพยากรณ์ Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

Period Ahead	Lower Confident Limit	Forecast	Upper Confident Limit	Actual If Known
1	2.2350823E+00	3.9351772E+00	5.6752916E+00	3.5335187E+00
2	2.7612828E+00	4.6945388E+00	6.6277865E+00	4.9272537E+00
3	3.7573237E+00	5.7637948E+00	7.7702959E+00	5.5134207E+00
4	4.4821553E+00	6.4919300E+00	8.5218240E+00	6.0008128E+00
5	4.6494835E+00	6.6888929E+00	8.7242373E+00	6.7845800E+00
6	4.9328070E+00	7.0224932E+00	9.0623794E+00	7.5288971E+00
7	4.7597591E+00	6.8004527E+00	8.8411463E+00	7.6503448E+00
8	4.4320301E+00	6.4723376E+00	8.5139451E+00	6.3731623E+00
9	3.5853961E+00	5.6273398E+00	7.6683835E+00	6.6243352E+00
10	3.1877122E+00	5.2387841E+00	7.2658560E+00	4.9344739E+00
11	2.6284551E+00	4.6895472E+00	6.7106024E+00	5.7235351E+00
12	2.3185524E+00	4.3398366E+00	6.4007207E+00	4.5432245E+00
13	1.9330280E+00	3.9314322E+00	6.0298384E+00	5.4071716E+00
14	2.6418266E+00	4.6924873E+00	6.7431478E+00	6.0112078E+00
15	3.7112215E+00	5.7826189E+00	7.8140182E+00	6.3317505E+00
16	4.4395733E+00	6.4913183E+00	8.5429577E+00	6.6137248E+00
17	4.8347308E+00	6.6834989E+00	8.7382171E+00	6.6137246E+00
18	4.9705298E+00	7.0222736E+00	9.0740176E+00	7.7159685E+00
19	4.7483747E+00	6.8003272E+00	8.6520799E+00	6.6908400E+00
20	4.4211606E+00	6.4729158E+00	8.5246710E+00	6.0344984E+00
21	3.5755427E+00	5.6272968E+00	7.6790548E+00	6.6598492E+00
22	3.1770042E+00	5.2287806E+00	7.2805170E+00	5.5535187E+00
23	2.617773E+00	4.6695338E+00	6.7212003E+00	5.3235187E+00
24	2.3070724E+00	4.3596289E+00	6.4113854E+00	5.0982304E+00
FORECAST RESULTS IN TERMS OF UNTRANSFORMED DATA				
1	9.92508023+00	5.3794824E+01	2.9157335E+02	
2	1.5820280E+01	1.0934845E+02	7.5580737E+02	
3	4.2833637E+01	3.1855488E+02	2.3691009E+03	
4	8.6574115E+01	6.5983514E+02	5.0232115E+03	
5	1.0452941E+02	9.0181900E+02	6.1505534E+03	
6	1.4593413E+02	1.1215795E+03	8.6246479E+03	
7	1.1671781E+02	3.9825387E+02	6.9129126E+03	
8	8.4101970E+01	6.4741504E+02	4.9837660E+03	
9	3.6100116E+01	2.7792181E+02	2.1399201E+03	
10	2.4232923E+01	1.8656682E+02	1.4363437E+03	
11	1.3352503E+01	1.0664944E+02	8.2108644E+02	
12	1.0160955E+01	7.8228698E+01	6.0227896E+02	
13	6.9103892E+00	5.3593736E+01	4.1564787E+02	
14	1.4308824E+01	1.0912420E+02	8.4822671E+02	
15	4.0903740E+01	3.1818083E+02	2.4750558E+03	
16	8.4747719E+01	5.5939206E+02	5.1304968E+03	
17	1.0300531E+02	8.0151111E+02	6.2367662E+03	
18	1.4410318E+02	1.1213332E+03	8.7256098E+03	
19	1.1541986E+02	8.9814109E+02	6.9980820E+03	
20	8.3192781E+01	6.4736857E+02	5.0375291E+03	
21	3.5713996E+01	2.7791040E+02	2.1625749E+03	
22	2.3974482E+01	1.8656144E+02	1.4517384E+03	
23	1.3705227E+01	1.0664801E+02	8.2588763E+02	
24	1.0053013E+01	7.8228098E+01	6.0873645E+02	

ตารางที่ 6.9 แบบจำลอง Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

WEIGHTS USED IN CALCULATING CONFIDENCE LIMITS AND UPDATING FORECASTS AFTER NEW OBSERVATION	
j	ψ_j
0	1.0000000E+00
1	5.5534304E-01
2	3.1779239E-01
3	1.8169652E-01
4	1.0388685E-01
5	5.9398334E-02
6	3.3961585E-02
7	1.9417872E-02
8	1.1102360E-02
9	6.3478843E-03
10	3.6294656E-03
11	2.0751829E-03
12	1.0238477E-01
13	5.6878146E-02
14	3.2547917E-02
15	1.8609146E-02
16	1.0639970E-02
17	6.0835082E-03
18	3.4783060E-03
19	1.9887558E-03
20	1.1370908E-03
21	6.5014292E-04
22	3.7172564E-04
23	2.1253781E-04
24	1.0131978E-01

(3f) Inverse Transformation

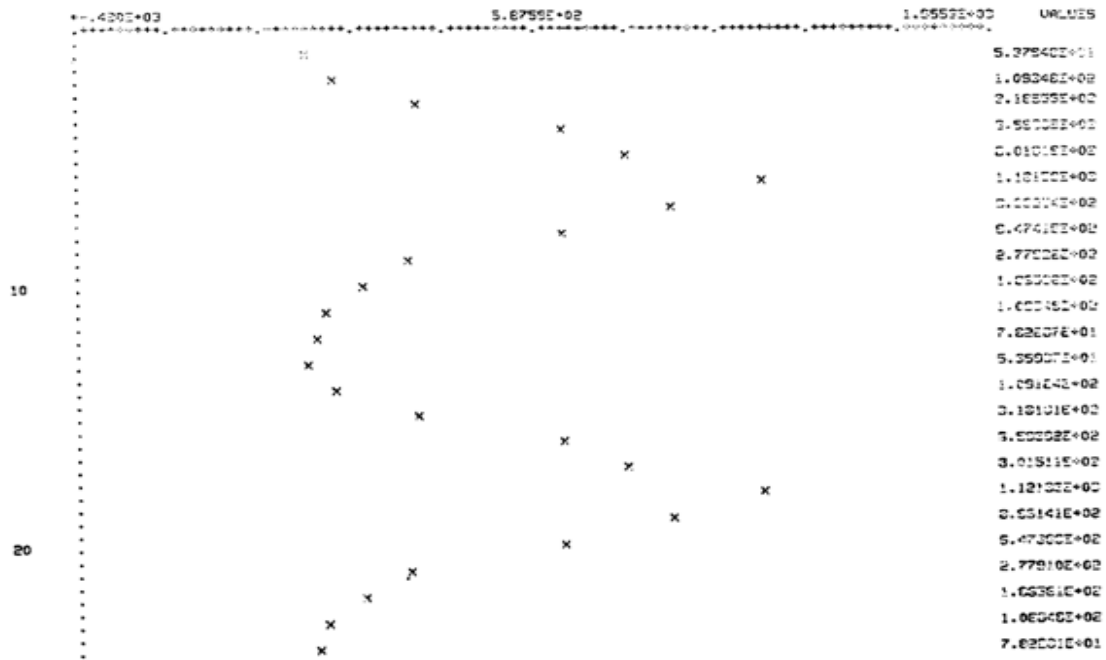
$$x_t(L) = \text{Exp}(y_t(L))$$

ตัวอย่างการ Inverse Transformation เพื่อหาค่า $x_t(L)$ จากค่า $y_t(L)$ ในตารางที่ 6.8 (ช่วงบน)

เมื่อ $L=1$ และ 2 แสดงอยู่ในตารางข้างล่าง

ค่าพยากรณ์ $y_t(L)$ จากตารางที่ 6.8			
L	Lower $y_{456}(L)$	$y_{456}(L)$	Upper $y_{456}(L)$
1	2.2950628	3.985177	5.6752916
2	2.7612926	4.6945396	6.6277865
$x_t(L) = \text{Exp}(y_t(L))$			
L	Lower $x_{456}(L)$	$x_{456}(L)$	Upper $x_{456}(L)$
1	9.925059286	53.79481	291.5733489
2	15.82027903	109.34845	755.8073382

ผลการทำ Inverse Transformation หรือ $x_{456}(L)$ เมื่อ $L=1, \dots, 24$ แสดงอยู่ในตารางที่ 6.8 (ช่วงล่าง) และถูกพล็อตในรูปที่ 6.17



รูปที่ 6.17 Graph of Forecasts at Origin, Graph Interval is $2.1360E+01$, $T=456$ with $(2,0,0) \times (0,1,1)_{12}$ Model Applied to Monthly Flow Natural Logarithms of the Blue River Near White Cloud, Indiana

(3g) Real Time Forecast

Real Time Forecast คือการพยากรณ์ที่มีการ Update เมื่อมีข้อมูลใหม่เพิ่มขึ้นสามารถทำได้ 2 แบบ
คือ

(1) การ Update โดยการประมาณค่าพารามิเตอร์ของแบบจำลองใหม่

(2) การ Update โดยใช้ค่าพารามิเตอร์เดิม

Forecast Update แบบที่ 2 จะสามารถทำได้ดังนี้

$$y_{t+1}(L) = y_t(L + 1) + \psi_L \varepsilon_{t+1} \quad [6.49]$$

เมื่อ

$y_{t+1}(L)$ = ค่าพยากรณ์ที่มีการ Update

$y_t(L + 1)$ = ค่าพยากรณ์ที่ Lead Time=L+1 จากฟังก์ชันการพยากรณ์เดิม

ψ_L = Ψ -weight ที่ Lead Time=L

ε_{t+1} = Forecast Error ที่เวลา t และ Lead Time=L+1 จากฟังก์ชันการพยากรณ์เดิม

$$\varepsilon_{t+1} = y_{t+1} - y_t(1)$$

แทนค่า ε_{t+1} ลงในสมการ [6.49] จะได้ Forecast Update ดังสมการ

$$y_{t+1}(L) = y_t(L + 1) + \psi_L (y_{t+1} - y_t(1)) \quad [6.50]$$

ดังนั้นเมื่อทราบค่า $y_{457} = \ln(36)$ จะสามารถหา Forecast Update ที่ L=1 สำหรับเวลา t=457 ได้
ดังนี้

$$y_{457}(1) = y_{456}(2) + \psi_1 (y_{457} - y_{456}(1))$$

$$y_{457}(1) = 4.6945 + 0.5553(\ln(36) - 3.9852) = 4.4714$$

$$y_{457}(2) = y_{456}(3) + \psi_2 (y_{457} - y_{456}(1))$$

$$y_{457}(2) = 5.7638 + 0.31779(\ln(36) - 3.9852) = 5.6361$$

ในทำนองเดียวกันเมื่อทราบ $y_{458} = \ln(138)$ จะสามารถหา Forecast Update ที่ L=1 สำหรับเวลา t=458 ได้ดังนี้

$$y_{458}(1) = y_{457}(2) + \psi_1 (y_{458} - y_{457}(1))$$

$$y_{458}(1) = 5.6361 + 0.5553(\ln(138) - 4.4714) = 5.8892$$

$$y_{458}(2) = y_{457}(3) + \psi_2 (y_{458} - y_{457}(1))$$

$$y_{458}(2) = 6.4190 + 0.31779(\ln(138) - 4.4715) = 6.5638$$

ค่า Update Forecast สำหรับ L=1 แสดงอยู่ในตารางที่ 6.10 และถูกพล็อตในรูปที่ 6.18 ค่า Inverse Transform ($x = \ln(y)$) ของค่า Update Forecast ถูกพล็อตในรูปที่ 6.19

ตารางที่ 6.10 Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ที่ประยุกต์ใช้กับ Natural Logarithms ของปริมาณการไหลของน้ำรายเดือนของ Blue River ใกล้ White Cloud มลรัฐ Indiana

UPDATED FORECASTS WITH 1 NEW							
PERIODS AHEAD OF PAST 436	UPDATED FORECAST	2 NEW OBSERVATIONS UPDATED FORECAST	3 NEW	4 NEW	5 NEW	ORIGINAL FORECAST	ACTUAL, IF KNOWN
2	4.4714314E+00		UPDATED FORECAST	UPDATED FORECAST	UPDATED FORECAST	4.5945393E+00	4.9272337E+00
3	5.6251503E+00	5.3392307E+00		UPDATED FORECAST	UPDATED FORECAST	5.7337943E+00	5.5134337E+00
4	6.4190101E+00	6.5538511E+00	6.3551354E+00		UPDATED FORECAST	6.4919900E+00	6.0083122E+00
5	6.6451559E+00	6.7275581E+00	6.6035316E+00	6.4162033E+00		6.6358829E+00	6.7845833E+00
6	6.5885354E+00	7.0459341E+00	6.9776937E+00	6.8576352E+00	7.0777704E+00	7.0224932E+00	7.5268071E+00
7	6.7658113E+00	6.3133333E+00	6.7743393E+00	6.7119144E+00	6.3321515E+00	3.3004527E+00	7.6505446E+00
8	6.4651238E+00	6.4803370E+00	6.4533432E+00	6.4223345E+00	6.4911157E+00	3.4720970E+00	6.8231632E+00
9	5.6228905E+00	5.6317303E+00	5.6139337E+00	5.5533553E+00	5.6377043E+00	5.6273393E+00	6.6340653E+00
10	5.2232344E+00	5.2312343E+00	5.2235337E+00	5.2122351E+00	5.2347104E+00	5.2237341E+00	4.9344733E+00
11	4.6539394E+00	4.6708325E+00	4.6709335E+00	4.6500352E+00	4.6725335E+00	4.6533472E+00	5.7235851E+00
12	4.3538030E+00	4.3304573E+00	4.3530715E+00	4.3542235E+00	4.3315739E+00	4.3553255E+00	4.5432843E+00
13	3.9403035E+00	3.8412543E+00	3.9333502E+00	3.0376913E+00	3.9413933E+00	3.5314322E+00	5.4071718E+00
14	4.6593417E+00	4.7163033E+00	4.7153255E+00	4.7142339E+00	4.7165709E+00	4.6524373E+00	6.0112578E+00
15	5.7495457E+00	5.7754703E+00	5.7355907E+00	5.7352724E+00	5.7376454E+00	5.7323155E+00	6.3347508E+00
16	6.4833433E+00	6.4933733E+00	6.4773016E+00	6.4413435E+00	6.4423237E+00	6.3910133E+00	6.6133342E+00
17	6.6322252E+00	6.6907057E+00	6.6734762E+00	6.0337730E+00	3.6975133E+00	3.3334333E+00	7.7155553E+00
18	7.0188301E+00	7.0245755E+00	7.0176553E+00	7.0034135E+00	7.0279333E+00	7.0322233E+00	6.6806423E+00
19	6.7939301E+00	6.8017033E+00	6.7977036E+00	6.7512533E+00	6.8035747E+00	5.3003272E+00	6.0844584E+00
20	6.4721170E+00	6.4737023E+00	6.4714159E+00	5.4577311E+00	6.4747724E+00	6.4729153E+00	6.0808423E+00
21	5.8538420E+00	5.6277433E+00	5.6254412E+00	5.6243344E+00	7.6233503E+00	5.3272333E+00	5.5863281E+00
22	5.2234955E+00	5.2230178E+00	5.2282703E+00	5.0270357E+00	3.2293376E+00	5.2237606E+00	5.5354328E+00
23	4.6593345E+00	4.6595333E+00	4.6322533E+00	4.6533547E+00	4.3553303E+00	4.6355333E+00	5.0985554E+00
24	4.3393455E+00	4.3397138E+00	4.3594333E+00	4.3590743E+00	4.3553273E+00	4.3335233E+00	
FORECAST RESULTS IN TERMS OF UNTRANSFORMED DATA							
2	3.7435233E+01						
3	2.3038142E+02	3.8113321E+02					
4	6.1339582E+02	7.0333333E+02	5.7544023E+02				
5	7.6904932E+02	8.8344300E+03	7.4139333E+02	6.1167674E+02			
6	1.0951377E+03	1.1482333E+03	1.0724454E+03	9.6037683E+02	1.1053227E+03		
7	3.3503402E+02	9.1035934E+02	3.7333917E+02	3.2214322E+02	2.7192750E+01		
8	6.4238593E+02	6.5240534E+02	6.3300312E+02	6.1545333E+02	6.5325343E+02		
9	2.7656522E+02	2.7314478E+02	2.7580443E+02	2.6553293E+02	2.3031743E+02		
10	1.8509074E+02	1.3703473E+02	1.3587478E+02	1.3350374E+02	1.3767474E+02		
11	1.0849403E+02	1.0620264E+02	1.0335798E+02	1.0384503E+02	1.0701143E+02		
12	7.8183520E+01	7.8232923E+01	7.3103322E+01	7.7306821E+01	7.3330401E+01		
13	5.1434488E+01	5.1483137E+01	5.1413953E+01	5.1300055E+01	5.1516017E+01		
14	1.0655952E+02	1.1175455E+02	1.1163752E+02	1.1152723E+02	1.1178343E+02		
15	3.1404326E+02	3.2228534E+02	3.1012974E+02	3.0990324E+02	3.1033223E+02		
16	6.5443130E+02	6.6423304E+02	6.5021403E+02	6.2753282E+02	6.2805538E+02		
17	7.8309305E+02	8.0439039E+02	7.5310500E+02	7.7558314E+02	3.1038083E+02		
18	1.1135333E+03	1.1240343E+03	1.1162003E+03	1.1035891E+03	1.1278933E+03		
19	8.8833719E+02	8.8337743E+02	3.8572314E+02	3.9003337E+02	9.0108231E+02		
20	6.4883163E+02	6.4737793E+02	6.4535233E+02	6.4402033E+02	6.4337162E+02		
21	2.7773350E+02	2.7303541E+02	2.7757213E+02	2.7708773E+02	2.7220533E+02		
22	1.8551273E+02	1.3360942E+02	1.2343958E+02	1.3524550E+02	1.8537471E+02		
23	1.0653209E+02	1.0698252E+02	1.0391812E+02	1.0354471E+02	1.0893302E+02		
24	7.8221442E+01	7.8234673E+01	7.3215590E+01	7.3184764E+01	7.3248321E+01		

6.13 ข้อจำกัดที่ต้องพิจารณาในการนำแบบจำลอง Multiplicative ARIMA ไปประยุกต์ใช้

(Limits to be Considered in Applications of Multiplicative ARIMA Models)

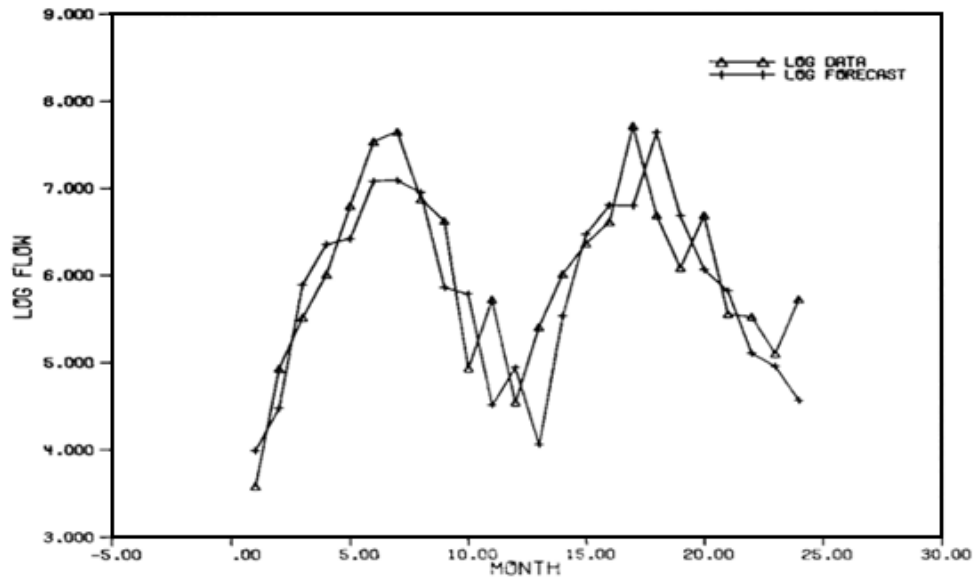
ค่าพยากรณ์ $y_t(L)$ จะฉายตัวแบบ Normal รอบๆ ค่าเฉลี่ยของ $y_t = \ln(x_t)$ โดยมีค่า Standard Error ในการพยากรณ์เท่ากับ $s_y(L)$ ค่าพยากรณ์แบบ Updated Forecast ในรูปของ $y_t(L)$ ในรูปที่

6.19 ให้ค่าถูกต้องแม่นยำกว่าค่าพยากรณ์ในรูปของ $x_t(L)$ ในรูปที่ 6.18 โดย \bar{x} , s_x และ \bar{y} , s_y มีความสัมพันธ์กันดังสมการ

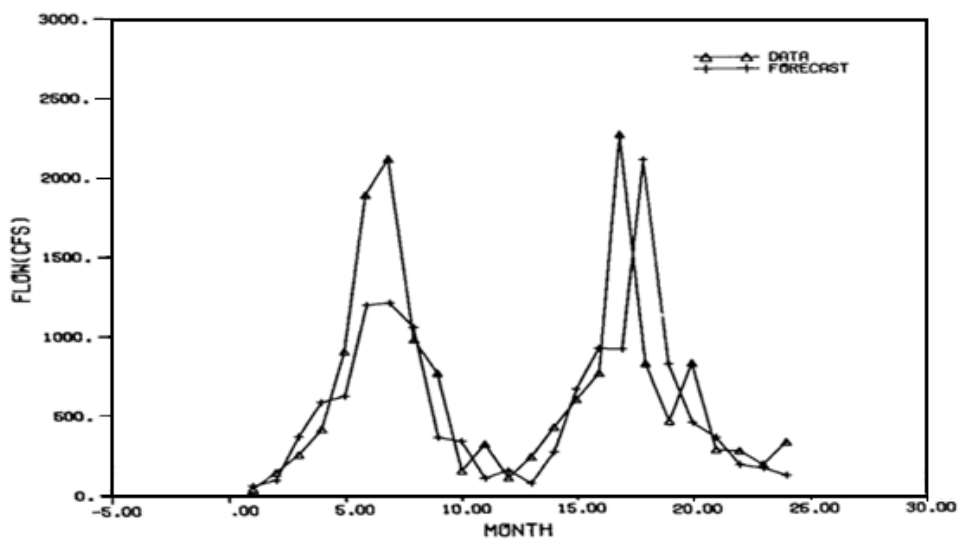
$$\bar{x} = \text{Exp}\left(\bar{y} + \frac{s_y^2}{2}\right) \quad [6.51]$$

$$s_x = \bar{x} \sqrt{\text{Exp}(s_y^2) - 1} \quad [6.52]$$

ดังนั้น จะสามารถหาค่าพยากรณ์ $x_t(L)$ และ Standard Error ของค่าพยากรณ์ได้จากสมการ



รูปที่ 6.18 Real Time Forecasts of Natural Logarithms of Monthly Flows, Blue River Near White Cloud, Indiana



รูปที่ 6.19 Real Time Forecasts of Monthly Flow, Blue River Near White Cloud, Indiana

$$x_t(L) = \text{Exp}\left(y_t(L) + \frac{s_y^2(L)}{2}\right) \quad [6.53]$$

$$s_t(L) = x_t(L) \sqrt{\text{Exp}(s_y^2(L)) - 1} \quad [6.54]$$

เป็นที่น่าสังเกตว่าค่าความคลาดเคลื่อนในการพยากรณ์ไม่ได้ขึ้นอยู่กับฤดูกาล ซึ่งเป็นผลทำให้เกิดความผิดพลาดในการพยากรณ์ค่า $x_t(L)$

ค่าความผิดพลาดที่กล่าวถึงสามารถแก้ไขได้โดยการแปลง $y_{v,\tau}$ ให้เป็นอนุกรมเวลาแบบ Stationary โดยวิธี Nonparametric Period Standardization (z_t) แล้วจึงฟิตแบบจำลองกับ z_t

$$z_t = \frac{y_{v,\tau} - \bar{y}_\tau}{s_\tau} \quad [6.55]$$

จากการศึกษาพบว่า 16 อนุกรมเวลาน้ำท่ารายเดือนผ่านการทดสอบความเหมาะสม

Wabash River ที่ Logansport, Indiana (Station 3-3290) เหมาะกับแบบจำลองแบบ AR(2) และมีค่าพารามิเตอร์เท่ากับ

$$\hat{\sigma}_\varepsilon^2 = 0.738, \hat{\phi}_1 = 0.468 \pm 0.043, \hat{\phi}_2 = 0.082 \pm 0.043$$

ค่าพยากรณ์โดยแบบจำลอง AR(2) ในรูปของ $y_t = \ln(x_t)$ แสดงอยู่ในรูปที่ 6.20 ซึ่งให้ผลใกล้เคียงกับ Multiplicative ARIMA(2,0,0)x(0,1,1)₁₂ ค่า Standard Error ในการพยากรณ์โดย AR(2) ในรูปของ z_t คือ

$$s_z(L) = \left(\sqrt{1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{L-1}^2}\right) \sigma_\varepsilon \quad [6.56]$$

ซึ่งทำให้ Standard Error ของ $y_t(L)$ เปลี่ยนแปลงไปตามฤดูกาลดังสมการ

$$s_y(t, L) = s_z(L) s_\tau \quad [6.57]$$

เมื่อ s_τ = Periodic Standard Deviation ของ y_t

โดยวิธีนี้ จึงทำให้การพยากรณ์ค่า x_t มีความถูกต้องมากขึ้นดังรูปที่ 6.21

6.14 การเปรียบเทียบและข้อจำกัดของแบบจำลอง ARMA และ ARIMA

(Comparison and Limitations of ARMA and ARIMA Models)

ข้อสรุปการฟิตแบบจำลอง ARMA และ ARIMA กับอนุกรมเวลาน้ำท่าและน้ำฝนรายเดือนแสดงอยู่ในตารางที่ 6.11 ซึ่งจะเห็นได้ว่าแบบจำลอง ARIMA มีจำนวนพารามิเตอร์น้อยกว่าแบบจำลอง ARMA แต่แบบจำลอง ARIMA มีข้อจำกัดที่สำคัญคือไม่เหมาะที่จะนำมาใช้ในการสังเคราะห์ข้อมูลใหม่ หรือการจำลอง (Simulation) ยกตัวอย่างแบบจำลอง ARIMA(0,1,0)₁₂

$$z_t = z_{t-12} + \varepsilon_t$$

ถ้าต้องการสังเคราะห์ค่าอนุกรมเวลาของเดือนมกราคม

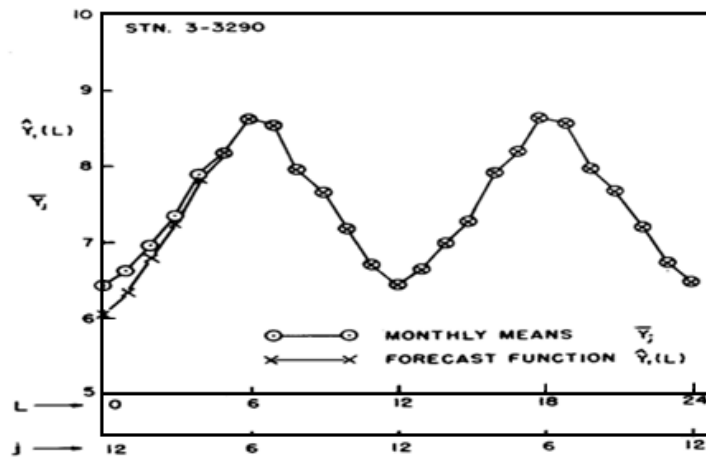
$$z_{13} = z_1 + \varepsilon_{13}$$

$$z_{25} = z_{13} + \varepsilon_{25} = z_1 + \varepsilon_{13} + \varepsilon_{25}$$

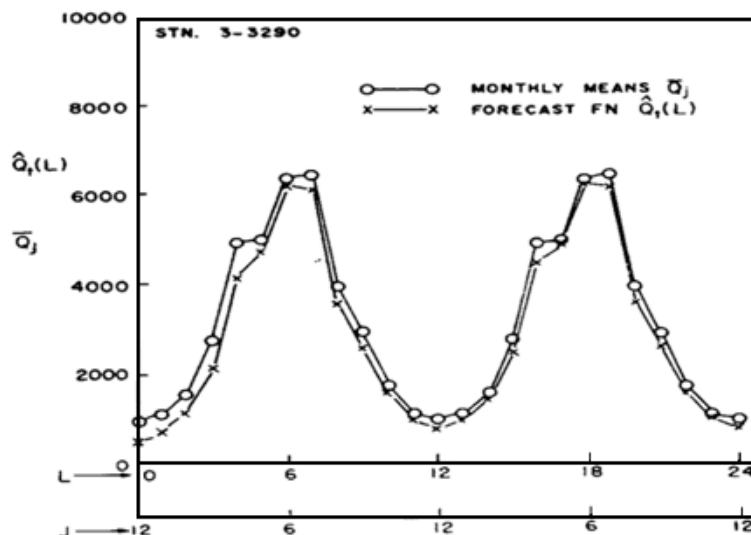
$$z_{37} = z_{25} + \varepsilon_{37} = z_1 + \varepsilon_{13} + \varepsilon_{25} + \varepsilon_{37}$$

$$z_{2T+1} = z_1 + \sum_{t=1}^T \varepsilon_{12t+1}$$

จากสมการข้างบนจะสามารถสรุปได้ว่า z_{2T+1} คือ Random Walk ของ T Independent Random Increments ซึ่งมี Mean เท่ากับ z_1 และ Variance เท่ากับ $T\sigma_\varepsilon^2$ ซึ่งจะเห็นได้ว่า Variance เพิ่มขึ้นตามค่า T และมีผลทำให้ค่าที่สังเคราะห์ใหม่มีค่าเพิ่มขึ้นตามเวลา จึงทำให้ ARIMA ไม่เหมาะที่จะนำมาใช้ในการสังเคราะห์ข้อมูลใหม่



รูปที่ 6.20 Forecasting Function for Logarithms of Monthly Flows of the Wabash River at Logansport, Indiana



รูปที่ 6.21 Forecasting Function for Monthly Flows of the Wabash River at Logansport, Indiana

ตารางที่ 6.11 Comparison of ARIMA Models for Monthly Data

Model	Parameters	%Passing*	Usage*	Model
Monthly flow logarithms-16 series	1,0,0	26	62.5	G & F
(McKerchar and Delleur , 1974)	2,0,0	27	100	G & F
	(2,0,0)x(0,1,1) ₁₂	4	100	F
Monthly Rainfall Square Root	0,0,0	1	47	G & F
15 series	1,0,1	27	100	G & F
(Delleur et al.,1976)	0,1,0	1	0	F
	1,1,1	3	40	F
	(1,1,1) ₁₂	3	80	F
	(1,0,0)x(1,1,1) ₁₂	4	100	F
* Porte Manteau test 90%				
** G = generation				
F = forecasting				

นอกจากนี้การทำ Seasonal Differencing อาจไม่เหมาะที่จะนำมาใช้ถ้าค่า Variance มีการเปลี่ยนแปลงตามฤดูกาล

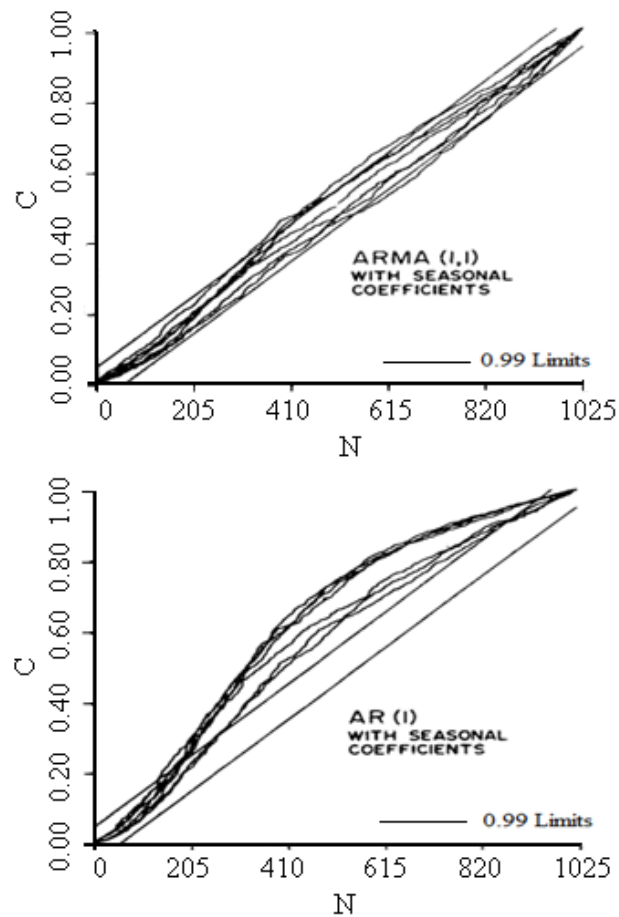
อย่างไรก็ตาม แบบจำลองแบบ Monthly Periodic Coefficients อาจมีปัญหาเรื่อง Parsimony เนื่องจากมีจำนวนพารามิเตอร์มากเกินไป

Monthly Periodic Coefficient AR(1) มีพารามิเตอร์มากกว่า Constant Coefficient AR(1) เท่ากับ 11 พารามิเตอร์

Monthly Periodic Coefficient AR(2) มีพารามิเตอร์มากกว่า Constant Coefficient AR(2) เท่ากับ 22 พารามิเตอร์

Seasonal Coefficient Model จะมีข้อดีสำหรับอนุกรมเวลารายสัปดาห์และรายวัน เช่น Weekly และ Daily Seasonal ARMA(1,1) มีพารามิเตอร์ 52 ตัว ถ้าใช้ 6 Significant Harmonics ในการคำนวณหา Mean และ Standard Deviation

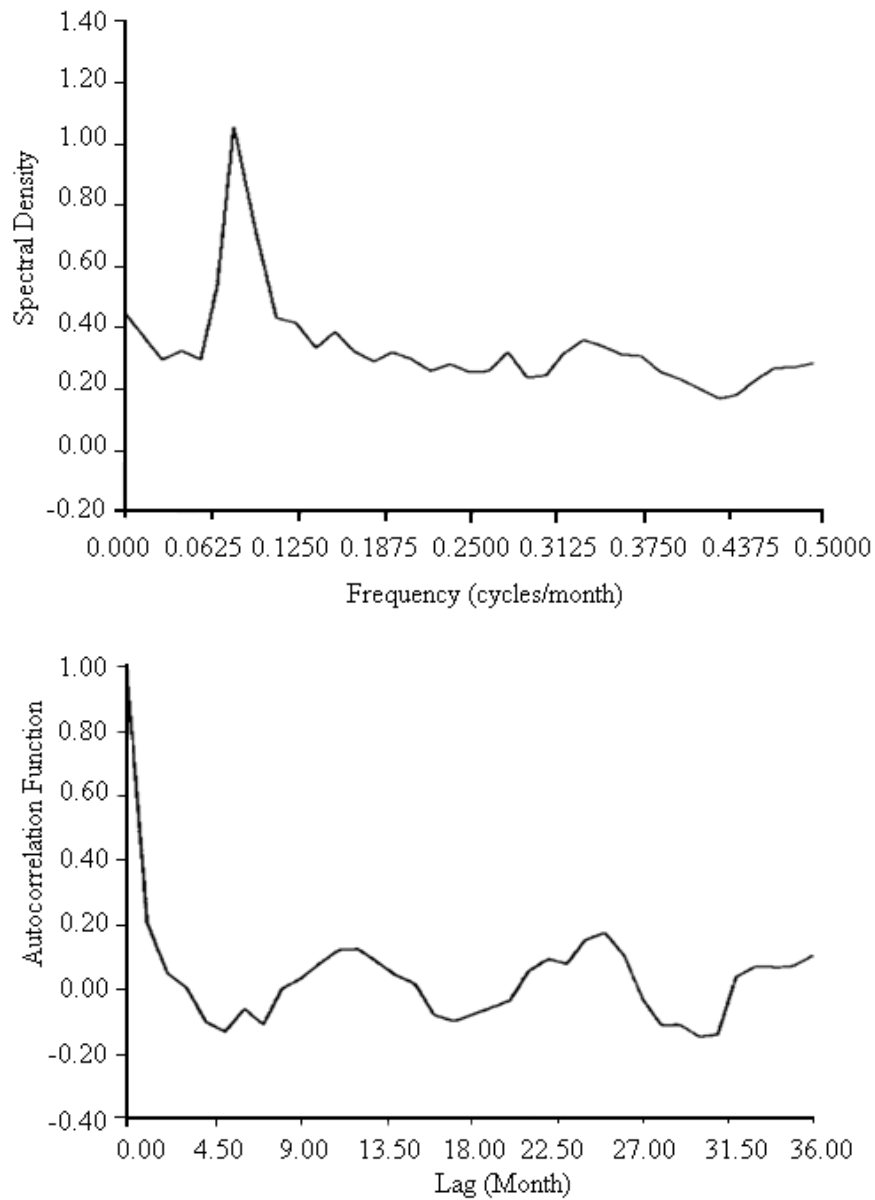
การทำ Nonparametric Standardization จะเพิ่มพารามิเตอร์ถึง $2 \times 365 = 730$ ตัว ทำให้ไม่สามารถนำมาใช้ได้ และยิ่งกว่านั้นจะเป็นการกำจัด Seasonality ของค่า Coefficients ตัวอย่างการสร้างแบบจำลองสำหรับ Logarithms ของอนุกรมเวลาน้ำท่ารายวันพบว่า Seasonal Coefficient ARMA(1,1) ผ่านการทดสอบ Residuals ด้วย Cumulative Periodogram ขณะที่ Seasonal Coefficient AR(1) ไม่ผ่านการทดสอบ ดังรูปที่ 6.22



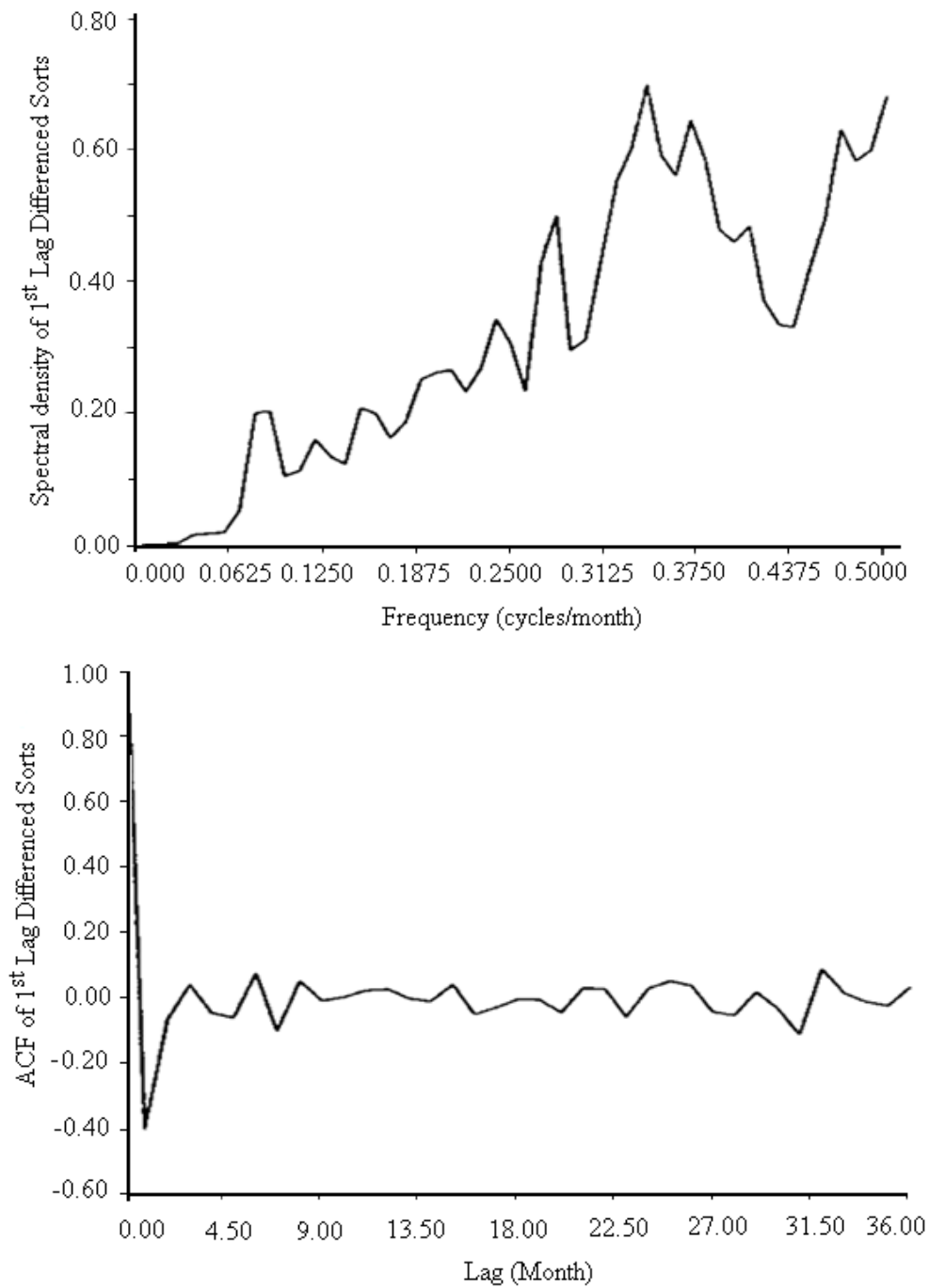
รูปที่ 6.22 Cumulative Periodiograms of Residuals of ARMA(1,1) Model Applied to Daily Flow Logarithms of Solomonie River at Dora, Indiana

ผลของ Differencing ต่อ Autocorrelation Function และ Spectrum

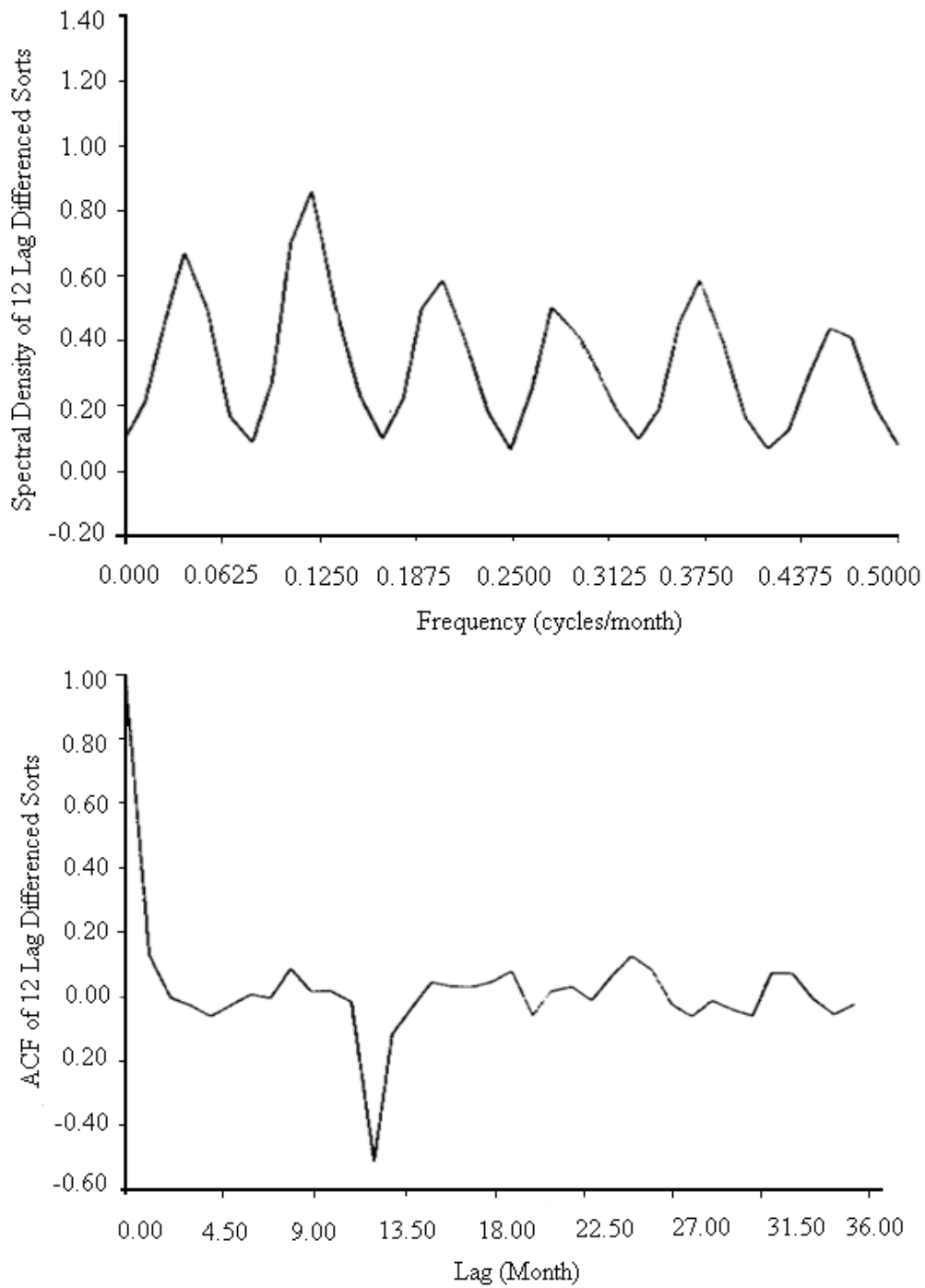
Kavvas และ Delleur (1975) ได้ศึกษาอิทธิพลของ Differencing ต่อ อนุกรมเวลารายเดือน โดยการ Rescaled อนุกรมเวลารายเดือนเพื่อให้มีค่าความแปรปรวนคงที่ โดยการหารค่าอนุกรมเวลาเริ่มต้นด้วยค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เมื่อเปรียบเทียบรูปที่ 6.23 และ 6.24 จะเห็นได้ว่า 1st Differencing ของอนุกรมเวลารายเดือนจะมีผลต่อ Spectral Density และ ACF รูปที่ 6.25 แสดงให้เห็นอิทธิพลของ 12th Month Differencing และรูปที่ 6.26 แสดงอิทธิพลของการ Standardization



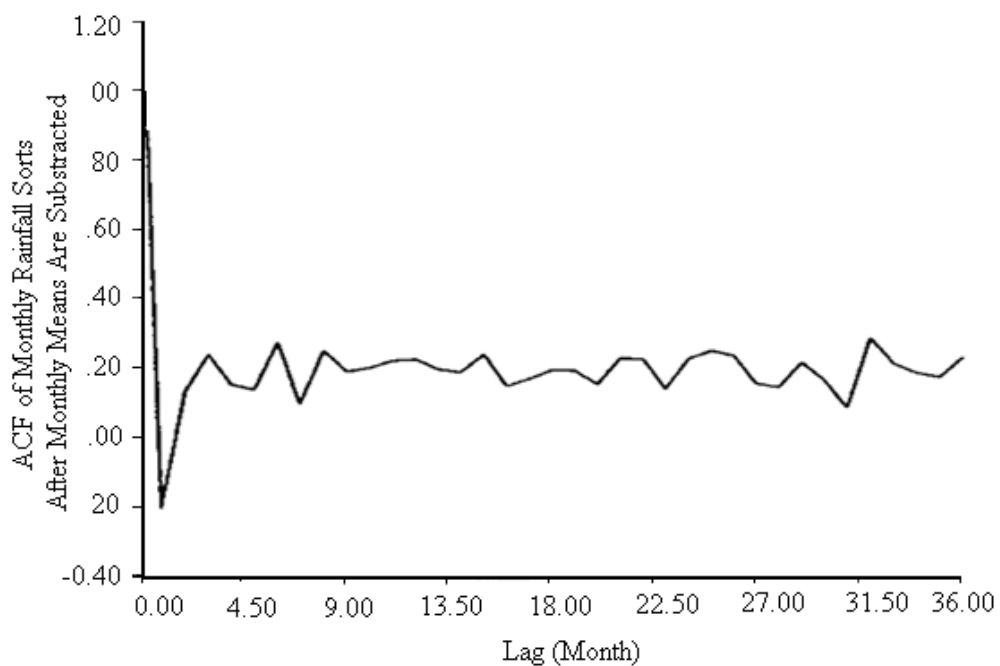
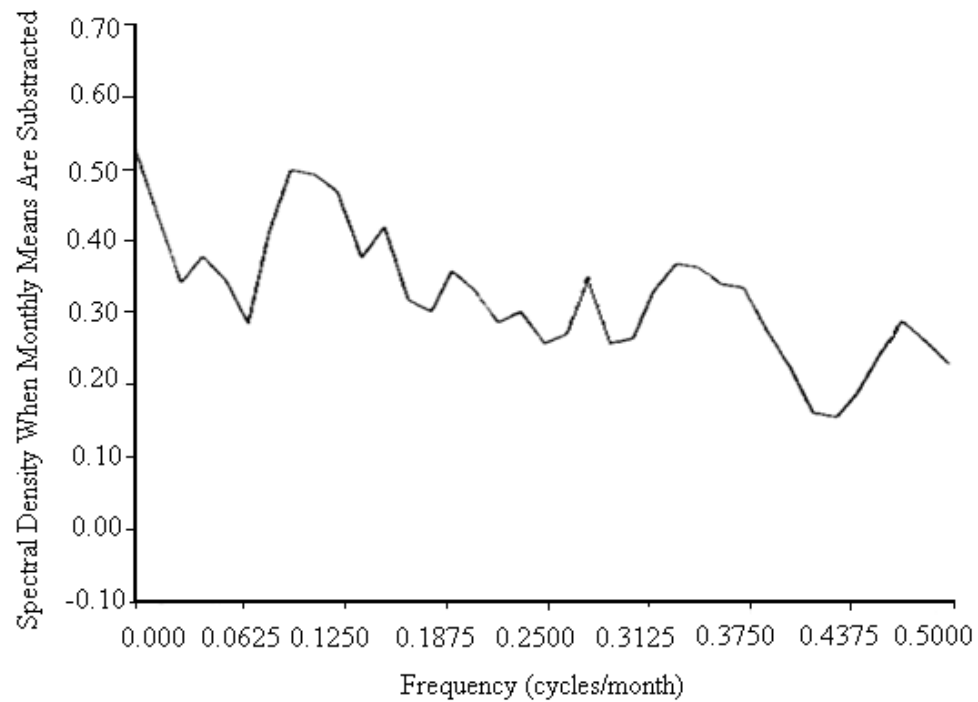
รูปที่ 6.23 Spectral Density and Autocorrelation Function of Rescaled Monthly Rainfall Square Roots of the Mississinewa River Basin above Marion, Indiana



รูปที่ 6.24 Spectral Density and Autocorrelation Function of Rescaled 1 Lag Differenced Monthly Rainfall Square Roots of the Mississinewa River Basin above Marion, Indiana



รูปที่ 6.25 Spectral Density and Autocorrelation Function of Rescaled 12 Lag Differenced Monthly Rainfall Square Roots of the Mississinewa River Basin above Marion, Indiana



รูปที่ 6.26 Spectral Density and Autocorrelation Function of Cyclicly Standardized Monthly Rainfall Square Roots of the Mississinewa River Basin above Marion, Indiana

6.15 เอกสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series.** Water Resources Publications. USA. pp.484

6.16 ภาคผนวก

APPENDIX A6.2. COMPUTER PROGRAM USED IN SIMPLE ARIMA MODELING, EXAMPLE 6.2.5

```

PROGRAM MAIN (INPUT,OUTPUT,TAPES=INPUT,TAPES=OUTPUT,TAPER,TAPES,TA  A   10
1PE3) A   20
REAL MMEAN(13),MSTDEV(13) A   30
DIMENSION XMAT(38,12), YTOTAL(39), TEMP(800), XTEMP(1000), HEADER(  A   40
18), PACU(50), ACU(50), AC(50), WKAREA(1000), XU(500), XCONF(4), YC  A   50
2ONF(4), GR(1000), A(1000), IND(8), PHI(11), THETA(11), SUMSQ(11,1  A   60
31) A   70
C A   80
C C A   90
C C A  100
C C A  110
DATA PHI/-.5,-.4,-.3,-.2,-.1,0.,.1,.2,.3,.4,.5/ A  120
DATA THETA/0.0,0.1,0.2,0.3,0.4,0.5,.6,.7,.8,.9,1.0/ A  130
C A  140
C C A  150
C C A  160
C C A  170
C C A  180
C C A  190
C C A  200
C C A  210
C C A  220
C C A  230
C C A  240
C C A  250
C C A  260
C C A  270
C C A  280
C C A  290
C C A  300
C C A  310
C C A  320
C C A  330
C C A  340
C C A  350
C C A  360
C C A  370
C C A  380
C C A  390
C C A  400
C C A  410
C C A  420
C C A  430
C C A  440
C C A  450
C C A  460
C C A  470
C C A  480
C C A  490
C C A  500
C C A  510
C C A  520
C C A  530
C C A  540
C C A  550
C C A  560
C C A  570
C C A  580
C C A  590
C C A  600
C C A  610
C C A  620
C C A  630
C C A  640
C C A  650
C C A  660
C C A  670
C C A  680
C C A  690
C C A  700
C C A  710
C C A  720
C C A  730
C C A  740
C C A  750
C C A  760
C C A  770
C C A  780
C C A  790
C C A  800
C C A  810
C C A  820
C C A  830
C C A  840
C C A  850
C C A  860
C C A  870
C C A  880
C C A  890
C C A  900
C C A  910
C C A  920
C C A  930
C C A  940
C C A  950
C C A  960
C C A  970
C C A  980
C C A  990
C C A 1000

```



```

C
C
C           IF (SUMSQ1(I5,I6).GT.SUMSQ1(ITHETA,IPHI)) GO TO 109
C           ITHETA=I5
C           IPHI=I6
C
C
C 109    CONTINUE
C 110    CONTINUE
C
C
C           WRITE THE SUM OF SQUARES SURFACE OUT TO TAPE 9
C
C           WRITE (9,125)
C           WRITE (9,126) (PHI(IT),IT=1,11)
C           WRITE (9,127) (THETA(IT),(SUMSQ1(IT,IT1),IT1=1,11),IT=1,11)
C
C
C           USE SUBROUTINE FTMXL TO ESTIMATE ARPS,PMAC,PMAS, AND COMPUTE
C           WNU. LEAVE RESIDUALS
C           IN THE FIRST IND(1) POSITIONS OF THE VECTOR #A#.
C
C
C           DATA IND/0,1,1,0,75,4,1,3/
C           IND(1)=NMONTHS
C
C
C           LET FINAL ESTIMATES OF ARPS AND PMAS FOR INPUT TO FTMXL BE THOSE
C           THAT LED TO THE MINIMUM SUMS OF SQUARES OF RESIDUALS GENERATED
C           ABOVE.
C
C           PMAS=THETA(ITHETA)
C           ARPS=PHI(IPHI)
C
C           IF PHI=1 SET PHI=.99 SINCE PHI=1 IS NOT DESIRABLE
C
C           IF (PMAS.EQ.1.0) PMAS=.999
C           WRITE (6,128) ARPS,PMAS
C
C
C           CALL FTMXL (XTEMP,IND,ARPS,PMAS,PMAC,WNU,GR,A,IER)
C           WRITE (6,129) ARPS,PMAS,PMAC,WNU
C
C
C           COMPUTE SUM OF SQUARES OF RESIDUALS
C
C           SSY=0.
C           DO 111 JH=1,NMONTHS
C               SSY1=A(JH)**2
C               SSY=SSY+SSY1
C               WRITE (6,130) A(JH),SSY1,SSY
C 111    CONTINUE
C
C           FIND AUTOCORRELATION AND PARTIAL AC OF RESIDUALS THAT ARE IN
C           THE #A# VECTOR AS OUTPUT OF FTMXL.
C
C
C           CALL FTAUTO (A,NMONTHS,24,24,7,AMEAN,ACU(1),ACU(2),AC(2),PACU(2),W
C           1KAREA)
C           AC(1)=1.
C           PACU(1)=1.
C
C
C           WRITE OUT AUTOCORRELATION AND PARTIAL AC. COMPUTE
C           SUMMATION((AUTOCORRELATION)**2).
C
C           SSQ=0.
C           DO 112 IH=1,25
C               IH1=IH-1

```

```

A 1370
A 1380
A 1390
A 1400
A 1410
A 1420
A 1430
A 1440
A 1450
A 1460
A 1470
A 1480
A 1490
A 1500
A 1510
A 1520
A 1530
A 1540
A 1550
A 1560
A 1570
A 1580
A 1590
A 1600
A 1610
A 1620
A 1630
A 1640
A 1650
A 1660
A 1670
A 1680
A 1690
A 1700
A 1710
A 1720
A 1730
A 1740
A 1750
A 1760
A 1770
A 1780
A 1790
A 1800
A 1810
A 1820
A 1830
A 1840
A 1850
A 1860
A 1870
A 1880
A 1890
A 1900
A 1910
A 1920
A 1930
A 1940
A 1950
A 1960
A 1970
A 1980
A 1990
A 2000
A 2010
A 2020
A 2030
A 2040

```

	WRITE (6,131) IH1,AC(IH),PACU(IH)	A 2050
	SSQ=SSQ+AC(IH)**2	A 2060
112	CONTINUE	A 2070
	SSQ=SSQ-1.	A 2080
	WRITE (6,132) SSQ	A 2090
	STOP	A 2100
C		A 2110
113	FORMAT (8A10)	A 2120
114	FORMAT (2X,8A10//10X, 35HORIGIAL DATA FROM SALAMONIA SERIES,//)	A 2130
115	FORMAT (12(1X,F5.3))	A 2140
116	FORMAT (20X, 31HLENGTH OF DIFFERENCED SERIES = ,I5)	A 2150
117	FORMAT (55X, 9HTABLE 6-1,//29X, 33HFIRST DIFFERENCE OF SQUARE ROO	A 2160
	ITS , 30HOF RAINFALL AT SALAMONIA, IND.,//,,,57X, 5HMONTH)	A 2170
118	FORMAT (2X, 4HYEAR,4X,12(3X,I2,4X)/)	A 2180
119	FORMAT (3X,I2,14X,11(F6.3,3X))	A 2190
120	FORMAT (3X,I2,5X,12(F6.3,3X))	A 2200
121	FORMAT (2X, 44HMEAN AND VARIANCE OF 1ST DIFF OF SQRT SERIES,2F15.7	A 2210
	1)	A 2220
122	FORMAT (5X,I2,10X,F10.4,10X,F10.4)	A 2230
123	FORMAT (//2X, 32HFTARPS ESTIMATE OF ARPS AND PMAC,2F15.6)	A 2240
124	FORMAT (//,2X, 30HFTMPS ESTIMATE OF PMAS AND WNU,2F15.6)	A 2250
125	FORMAT (9X,39X, 10HTABLE 6-2 ,//25X, 31HSUM OF SQUARES OF RESIDUAL	A 2260
	15 OF , 26H1ST DIFF. OF SQ. ROOT OF ,//,32X, 24HMONTHLY RAINFALL SE	A 2270
	2RIES , 18HAT SALAMONIA, IND.,//,52X, 3HPHI)	A 2280
126	FORMAT (5X, 5HTHETA,2X,11(3X,F5.2,1X)/)	A 2290
127	FORMAT (5X,F5.2,3X,11(F8.2,1X))	A 2300
128	FORMAT (//,2X, 42HVALUES OF ARPS AND PMAS LEADING TO MINIMUM, 29H	A 2310
	1SUMS OF SQUARES OF RESIDUALS,5X,2F15.6)	A 2320
129	FORMAT (2X, 34HFTMXL OUTPUT OF ARPS,PMAS,PMAC,WNU,4F15.6)	A 2330
130	FORMAT (3X, 30HRESIDUAL,SQUARE,SUM OF SQUARES,3F15.6)	A 2340
131	FORMAT (2X, 11HLAG,AC,PACU,I5,2F15.6)	A 2350
132	FORMAT (2X, 36HSUMMATION(AC**2)-NOT INCLUDING LAG 0,F15.6)	A 2360
C		A 2370
	END	A 2380

- *ILDEST - SET = 0 TO SUPPRESS LISTING OF DATA BY ESTIMATION ROUTINE. ANY OTHER VALUE DATA IS LISTED BY ESTIMATION ROUTINE
- *ILDFCA - SET = 0 TO SUPPRESS LISTING OF DATA BY FORECASTING ROUTINE. ANY OTHER VALUE DATA IS LISTED BY FORECASTING ROUTINE
- *INC - ARRAY CONTAINING (MFAC(1) + MFAC(3) + (2) NUMBERS OF EACH OF THE SPECIFIED TYPES OF PARAMETERS IN THE MODEL TO BE USED. MINIMUM SIZE IS (MFAC(1) + MFAC(3) + (2), MAXIMUM 25. (No. of ϕ_u ; μ , θ_o , θ_i 's)
- *IOD - ARRAY OF MINIMUM SIZE MFAC(2). CONTAINS THE MFAC(2) ORDERS OF DIFFERENCES OF EACH TYPE OF DIFFERENCE DESIRED IN TIME SERIES MODEL, I.E., THE VALUE OF S IN (1-B**S). MODEL WILL CONTAIN DIFFERENCE FACTORS (1 - B**IOD(1))**ND(1)*(1-B**IOD(2))**ND(2)*....
- *IOPA - ARRAY OF MINIMUM SIZE NP. CONTAINS ORDER OF EACH PARAMETER FROM LEFT TO RIGHT IN TIME SERIES MODEL TO BE USED (POWER OF B OPERATOR), order of each ϕ and θ
- *IPDEST - SET = 0 IF WISH TO SUPPRESS PLOTTING OF DATA BY ESTIMATION ROUTINE. ANY OTHER VALUE DATA IS PLOTTED
- *IPDFCA - SET = 0 TO SUPPRESS PLOTTING OF DATA BY FORECASTING ROUTINE. ANY OTHER VALUE DATA WILL BE PLOTTED
- IPDID - NO PURPOSE HERE
- *IPITER - SET = 0 IF PRINTED OUTPUT AFTER EACH ITERATION IN THE NON-LINEAR LEAST SQUARED ALGORITHM IS DESIRED. ANY OTHER VALUE SUPPRESSES THIS OUTPUT
- *IPRES - SET = 0 IF WISH TO SUPPRESS PLOT OF RESIDUALS. ANY OTHER VALUE RESIDUALS ARE PLOTTED
- *IPYYHR - SET = 0 IF PRINTING OF THE DATA, FITTED VALUES, AND THE RESIDUALS FROM THE FITTED MODEL IS DESIRED. ANY OTHER VALUE SUPPRESSES THIS OUTPUT

- *IWTPA - SET = 0 IF WISH TO SUPPRESS PLOTS OF RESIDUAL AUTOCORRELATIONS. ANY OTHER VALUE AUTOCORRELATIONS ARE PLOTTED

- IWTPF - SET = 0 TO SUPPRESS PLOTTING OF FORECASTS

- *MCSE - SET = 0 DO NOT WISH STANDARD ERRORS OF RESIDUAL AUTOCORRELATIONS. ANY OTHER VALUE STANDARD ERRORS ARE CALCULATED

- *MFAC - ARRAY OF SIZE 3. MFAC(1) = NO. OF AUTOREGRESSIVE FACTORS IN TIME SERIES MODEL, MFAC(2) = NO. OF DIFFERENCE FACTORS IN TIME SERIES MODEL, MFAC(3) = NO. OF MOVING AVERAGE FACTORS IN TIME SERIES MODEL

- *MIT - MAXIMUM NUMBER OF ITERATIONS FOR ESTIMATION. MAXIMUM ALLOWED IS 999

- *NAC - NUMBER OF AUTOCORRELATIONS OF RESIDUALS, MAXIMUM 150

- *NAPL - NUMBER OF RESIDUAL AUTOCORRELATIONS PRINTED PER LINE. BETWEEN 1 AND 12 INCLUSIVE, MAXIMUM 150

- *NCHI - THE NUMBER OF AUTOCORRELATIONS TO BE USED IN CALCULATING A CHI-SQUARE STATISTIC. SET LESS THAN OR = 0 IF NOT WANTED. MAXIMUM VALUE IS NAC

- *ND - ARRAY OF MINIMUM SIZE MFAC(2). CONTAINS THE MFAC(2) NUMBERS OF DIFFERENCES OF EACH TYPE OF DIFFERENCE DESIRED IN TIME SERIES MODEL. (SEE IOD)

- NDIFAC - NO PURPOSE HERE. IS SET = 0 BY A SUBROUTINE

- *NF - NUMBER OF FORECASTS DESIRED. BETWEEN 0 AND 300 INCLUSIVE

- *NOB - NUMBER OF OBSERVATIONS IN TIME SERIES TO BE MODELLED OR FORECAST

- NP - NUMBER OF PARAMETERS IN THE MODEL. MAXIMUM ALLOWED IS 50

- *NPAC - NUMBER OF PARTIAL AUTOCORRELATIONS OF RESIDUALS. MAXIMUM VALUE IS NAC
- NT - ARRAY OF MINIMUM SIZE NTO, CONTAINING FORECAST TIME ORIGINS
- *NTO - NUMBER OF TIME ORIGINS FOR FORECASTS, MAXIMUM 5. SOME GRAPH TITLES NOT GOOD IF EXCEEDS 9
- *NU - NUMBER OF NEW OBSERVATIONS READ IN FOR UPDATES OF FORECASTS, AND NUMBER OF UPDATES PRODUCED. MAXIMUM VALUE IS NF-1
- *PA - ARRAY OF MINIMUM SIZE NP. FOR ESTIMATION, CONTAINS INITIAL ESTIMATES OF PARAMETERS (NON-ZERO). FOR FORECASTING, CONTAINS PREVIOUSLY ESTIMATED PARAMETER VALUES (AUTOMATICALLY THERE IF ESTIMATION DONE FIRST)
- *SERIES - ARRAY OF SIZE 20. CONTAINS TITLE DESCRIBING DATA OR ANALYSIS
- *TLAM - DATA TRANSFORMATION PARAMETER. IF TLAM = 1.0, THE SERIES Z REMAINS AS READ IN. IF TLAM = 0, THE TRANSFORMED SERIES IS LN(Z+TM). FOR ANY OTHER VALUE OF TLAM, THE TRANSFORMED SERIES IS (Z+TM)**TLAM. (SEE TM)
- *TM - DATA TRANSFORMATION PARAMETER. IF TRANSFORMATION IS REQUESTED (SEE TLAM), THIS VALUE IS ADDED TO EACH VALUE OF THE SERIES Z BEFORE TRANSFORMATION IS MADE
- *Z - ARRAY FOR ORIGINAL TIME SERIES DATA. MINIMUM SIZE IS NOB+NF, MAXIMUM 500.
- *ZN - ARRAY FOR NEW OBSERVATIONS USED IN UPDATING FORECASTS. MINIMUM SIZE IS NU

(4) Reference:

"Computer program for Analysis of Univariate Time Series Models" by David J. Pack (1976).

APPENDIX A6.4. PROGRAM UNESTM AND DATA INPUT FOR MULTIPLICATIVE ARIMA MODELING, EXAMPLE 6.3.5

```

1.  000000B      PROGRAM UNESTM(INPUT,OUTPUT,TAPES=INPUT,TAPES=OUTPUT)
2.  002051B      DIMENSION PA(25),SCRATC(6000),RHO(150,1),STE(150,1),PHI(150,1)
3.  002051B      1,ZP(300,5),CLL(300,5),CUL(300,5),NT(5),U(300),ZN(25),FMZ(20)
4.  002051B      COMMON/TSRES/A(800)
5.  002051B      COMMON/TSOBS/Z(800)
6.  002051B      COMMON/TSPAR/MAX11,MBD,NP,NOB,MFAC(3),ND(30),IOD(30),INC(12)
7.  002051B      1,IOPA(50)
8.  002051B      COMMON/EPAR/EPS1,EPS2,MIT,NDIMS,TLAM,TH,ILDEST,IPDEST,IPRES
9.  002051B      1,IPITER,IPYYHR
10. 002051B      COMMON/FPAR/NT0,NU,ICI,ILDFCA,IPDFCA,IWTPF
11. 002051B      COMMON/IDPAR/NDIFAC,MD(5),JOD(5),NAPL,NCHI,MCSE,ILDID,IPDID,
12. 030155B      1MPKINT,IWTPA,ULAM,UM
13. 030164B      COMMON/NAMES/SERIES(20),TITL(20)
14. 030164B      COMMON/TSONE/C(200),CF(200),MPROB
15. 030174B      READ 2,NOB,IEYON,IFYON
16. 030204B      WRITE(6,2) NOB,IEYON,IFYON
17. 030204B      2 FORMAT(16I5)
18. 030215B      READ 3,(FMZ(I),I=1,20)
19. 030226B      WRITE(6,3) (FMZ(I),I=1,20)
20. 030236B      3 FORMAT(20A4)
21. 030246B      READ 5,FMZ) (Z(I),I=1,NOB)
22. 030246B      WRITE(6,FMZ) (Z(I),I=1,NOB)
23. 030250B      READ 3,(SERIES(I),I=1,20)
24. 030263B      WRITE(6,3) (SERIES(I),I=1,20)
25. 030275B      NMODEL=1
26. 030275B      DO 20 NPROB=1,NMODEL
27. 030275B      READ 6,TLAM,TH,(MFAC(I),I=1,3)
28. 030277B      WRITE(6,2) TLAM,TH,(MFAC(I),I=1,3)
29. 030313B      6 FORMAT(2F8.4,10I5)
30. 030327B      MAX=MFAC(2)
31. 030331B      IF(MAX.EQ.0) GO TO 8
32. 030342B      READ 2,(ND(I),IOD(I),I=1,MAX)
33. 030353B      WRITE(6,2) (ND(I),IOD(I),I=1,MAX)
34. 030353B      8 MAX=MFAC(1)+MFAC(3)+2
35. 030356B      READ 2,(INC(I),I=1,MAX)
36. 030362B      WRITE(6,2) (INC(I),I=1,MAX)
37. 030362B      NP=0
38. 030374B      DO 10 J=1,MAX
39. 030405B      10 NP=NP+INC(J)
40. 030416B      IF(NP.EQ.0) GO TO 13
41. 030427B      READ 2,(IOPA(I),I=1,NP)
42. 030427B      WRITE(6,2) (IOPA(I),I=1,NP)
43. 030430B      READ 7,(PA(I),I=1,NP)
44. 030444B      WRITE(6,7) (PA(I),I=1,NP)
45. 030460B      7 FORMAT(10F8.4)
46. 030463B      13 IF(IEYON.EQ.0) GO TO 15
47. 030476B      READ 6,EPS1,EPS2,MIT,ILDEST,IPDEST,IPRES,IPITER,IPYYHR
48. 030510B      WRITE(6,6) EPS1,EPS2,MIT,ILDEST,IPDEST,IPRES,IPITER,IPYYHR
49. 030523B      NDIMS=7*NP+NP*NP+2*NOB+NP*NOB
50. 030524B      READ 2,NAC,NPAC,MCSE,NAPL,IWTPA,NCHI
51. 030526B      WRITE(6,2) NAC,NPAC,MCSE,NAPL,IWTPA,NCHI
      CALL ESTIM(NPROB,NAC,NPAC,PA,RHO,STE,E,SH,PHI,SCRATC)
      IF(MAX11.EQ.(-1)) GO TO 20
      15 IF(IFYON.EQ.0) GO TO 20
      READ 2,NF,NT0,NU,ICI,ILDFCA,IPDFCA,IWTPF

```

```

1  52. 030541B      WRITE(6,2) NF,NTD,NU,ICI,ILDPCA,IPDFCA,IWTPF
    53. 030554B      READ 2,(NT(I),I=1,NTD)
    54. 030565B      WRITE(6,2) (NT(I),I=1,NTD)
    55. 030576B      IF(NU.EQ.0) GO TO 18
    56. 030577B      READ(5,FM2) (ZN(I),I=1,NU)
    57. 030610B      WRITE(6,FM2) (ZN(I),I=1,NU)
    58. 030621B      18 CALL FCAST(NPROB,NF,NT,PA,ZN,ZP,CLL,CUL,U,SCRATC(1),SCRATC(NOBB+1))
    59. 030636B      20 CONTINUE
    60. 030640B      STOP
    61. 030643B      END

```

456 1 1
(12F6.0)

181.	507.	922.	2776.	1158.	613.	755.	467.	308.	309.	246.	307.
146.	538.	1132.	2676.	1093.	1819.	2185.	2163.	214.	112.	56.	107.
128.	93.	225.	207.	78.	752.	263.	91.	120.	131.	79.	88.
46.	23.	58.	385.	208.	2197.	952.	1569.	648.	201.	115.	73.
55.	92.	128.	293.	503.	709.	1126.	150.	41.	58.	34.	52.
407.	761.	273.	6290.	1088.	396.	659.	1299.	247.	86.	224.	43.
260.	114.	563.	631.	1053.	1780.	1111.	572.	245.	515.	610.	177.
28.	52.	63.	763.	2185.	1973.	2050.	268.	489.	348.	203.	64.
30.	28.	26.	173.	1098.	558.	1770.	453.	182.	67.	38.	24.
19.	45.	74.	210.	141.	71.	421.	94.	261.	213.	88.	51.
68.	170.	320.	655.	1061.	1155.	1106.	453.	816.	304.	87.	43.
25.	396.	860.	486.	595.	2549.	680.	723.	414.	218.	67.	24.
18.	23.	21.	57.	106.	1132.	1549.	639.	145.	62.	239.	419.
66.	53.	191.	667.	1969.	4299.	1379.	533.	999.	335.	69.	211.
420.	485.	599.	830.	1898.	1197.	361.	1656.	425.	160.	307.	44.
33.	169.	532.	1415.	426.	275.	2422.	966.	892.	467.	174.	56.
39.	65.	99.	353.	891.	1882.	2277.	396.	98.	363.	98.	79.
38.	1209.	1267.	3178.	1699.	2334.	575.	423.	1545.	233.	386.	130.
249.	86.	922.	4504.	3404.	1084.	1781.	973.	885.	198.	138.	314.
117.	1028.	1095.	2344.	1990.	2154.	1609.	375.	179.	109.	61.	105.
41.	557.	2066.	1511.	1633.	1999.	641.	217.	296.	69.	33.	50.
22.	33.	193.	1124.	469.	1214.	815.	1030.	161.	292.	55.	19.
15.	22.	24.	123.	170.	303.	384.	189.	181.	45.	71.	249.
237.	69.	334.	553.	2155.	2061.	639.	382.	290.	328.	83.	162.
515.	589.	230.	350.	3149.	1374.	639.	640.	334.	143.	88.	75.
45.	36.	109.	527.	1284.	560.	1979.	1771.	745.	286.	106.	55.
37.	1337.	2417.	797.	527.	861.	824.	944.	193.	879.	800.	384.
195.	258.	563.	2664.	1288.	889.	626.	437.	206.	70.	57.	71.
69.	137.	520.	725.	1004.	779.	477.	253.	1882.	771.	85.	42.
24.	39.	56.	162.	603.	2302.	1368.	2939.	395.	352.	127.	70.
49.	140.	616.	1490.	2019.	1471.	630.	253.	152.	73.	36.	107.
70.	104.	110.	216.	200.	2575.	443.	246.	125.	216.	101.	57.
18.	20.	18.	112.	151.	4058.	885.	316.	553.	145.	30.	22.
14.	28.	281.	583.	1279.	1153.	902.	248.	86.	302.	57.	541.
247.	116.	99.	1389.	1557.	725.	1546.	1068.	137.	89.	101.	36.
49.	161.	671.	383.	554.	1349.	543.	1583.	192.	91.	146.	42.
51.	140.	1308.	539.	646.	1237.	1463.	1511.	428.	121.	216.	75.
30.	98.	1028.	2032.	1033.	376.	1127.	905.	358.	412.	189.	78.

BLUE RIVER NEAR WHITE CLOUD

0	0	1	1	1
1	12			
2	0	0	1	
1	2	12		
.5000	.1000	.9000		
.0040	.0040	96	0	0
24	24	1	12	1
24	1	23	4	0
456				
36.	138.	248.	407.	893.
223.	408.	581.	745.	2243.
				805.
				439.
				805.
				259.
				251.
				164.
				94.

บทที่ 7

การจำลองอนุกรมเวลาทางอุทกวิทยาด้วยแบบจำลองแบบหลายตัวแปร

MULTIVARIATE MODELING OF HYDROLOGIC TIME SERIES

7.1 คำนำ

แบบจำลองหลายตัวแปร (Multivariate Model) คือแบบจำลองที่สามารถจำลองอนุกรมเวลาหลายๆ ชุดพร้อมๆ กัน มีข้อดีคือสามารถรักษาคุณสมบัติด้านสหสัมพันธ์ทั้งตามเวลาและตามพื้นที่ (Temporal and Spatial Correlation) ของอนุกรมเวลาได้ แต่มีข้อเสียคือมีจำนวนพารามิเตอร์มาก ดังนั้น จึงมักใช้กับรูปแบบจำลองที่ไม่ซับซ้อนมาก เช่น AR(1) หรือ AR(2) เท่านั้น แต่แบบจำลอง AR สามารถรักษาคุณสมบัติด้าน Short Term Dependence เท่านั้น จึงไม่เหมาะกับอนุกรมเวลาที่มี Long Term Dependence

ในบทนี้จะกล่าวถึงการพัฒนาแบบจำลองหลายตัวแปรกับอนุกรมเวลาทั้งแบบรายปี และแบบ Periodic

7.2 แบบจำลองหลายตัวแปรลำดับที่หนึ่ง (Multivariate Lag One)

Matalas (1967) ได้เสนอแบบจำลองหลายตัวแปรลำดับที่หนึ่งตามแบบของ Markov (Multivariate Lag-One Markov Model) หรือ MAR(1) ซึ่งสามารถเขียนในรูปของเมทริกซ์ได้ดังสมการ

$$Z_t = A_1 Z_{t-1} + B \varepsilon_t \quad [7.1]$$

สมการ [7.1] สามารถเขียนแบบเต็มรูป (Expanded Form) ได้ดังนี้

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \\ \cdot \\ \cdot \\ Z_t^n \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & \cdot & \cdot & a^{1n} \\ a^{21} & a^{22} & \cdot & \cdot & a^{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a^{n1} & a^{n2} & \cdot & \cdot & a^{nn} \end{bmatrix} \begin{bmatrix} Z_{t-1}^1 \\ Z_{t-1}^2 \\ \cdot \\ \cdot \\ Z_{t-1}^n \end{bmatrix} + \begin{bmatrix} b^{11} & b^{12} & \cdot & \cdot & b^{1n} \\ b^{21} & b^{22} & \cdot & \cdot & b^{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b^{n1} & b^{n2} & \cdot & \cdot & b^{nn} \end{bmatrix} \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \cdot \\ \cdot \\ \varepsilon_t^n \end{bmatrix} \quad [7.2]$$

$$\text{เมื่อ } Z_t = \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ \cdot \\ \cdot \\ Z_t^n \end{bmatrix} = \text{Column Matrix ขนาด } [n \times 1]$$

$$A_1 = \begin{bmatrix} a^{11} & a^{12} & \cdot & a^{1n} \\ a^{21} & a^{22} & \cdot & a^{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a^{n1} & a^{n2} & \cdot & a^{nn} \end{bmatrix} \dots [n \times n] \quad B = \begin{bmatrix} b^{11} & b^{12} & \cdot & b^{1n} \\ b^{21} & b^{22} & \cdot & b^{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ b^{n1} & b^{n2} & \cdot & b^{nn} \end{bmatrix} \dots [n \times n]$$

$$\varepsilon_t = \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \cdot \\ \cdot \\ \varepsilon_t^n \end{bmatrix} \dots [n \times 1]$$

$\underline{\varepsilon}_t$ = Vector of independent normally distributed random variables with mean = 0 and standard deviation = 1 (uncorrelated in time and space)

โดยที่ $\underline{\varepsilon}_t$ = มีคุณสมบัติที่สำคัญคือ

$$E(\varepsilon_t^{(i)} \varepsilon_t^{(j)}) = 0 \text{ เมื่อ } i \neq j \tag{7.3}$$

$$E(\underline{\varepsilon}_t \underline{\varepsilon}_t^T) = I \tag{7.4}$$

7.3 Lag k Correlation Matrix

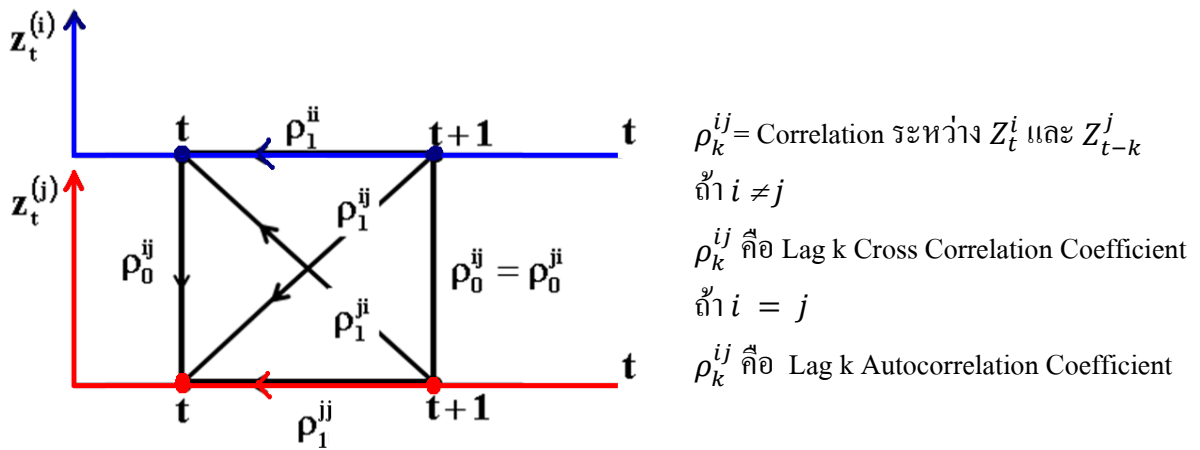
Lag k Correlation Matrix (M_k) สามารถหาได้จากสมการ

$$M_k = E(Z_t Z_{t-k}^T) = \begin{bmatrix} \rho_k^{11} & \rho_k^{12} & \cdot & \rho_k^{1n} \\ \rho_k^{21} & \rho_k^{22} & \cdot & \rho_k^{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \rho_k^{n1} & \rho_k^{n2} & \cdot & \rho_k^{nn} \end{bmatrix} = [\rho_k^{ij}] \tag{7.5}$$

M_k มีคุณสมบัติที่สำคัญคือ

$$M_{-k} = M_k^T \tag{7.6}$$

โดยที่ ρ_k^{ij} ใช้นิยามไว้ดังรูปที่ 7.1



รูปที่ 7.1 นิยามของ ρ_k^{ij}

สำหรับแบบจำลอง AR(1) จากสมการ (7.1)

$$Z_t = A_1 Z_{t-1} + B \underline{\varepsilon}_t$$

$$M_k = A_1 M_{k-1}; k > 0 \tag{7.7}$$

หรือ $M_k = A_1^k M_0; k > 0 \tag{7.8}$

สำหรับแบบจำลอง AR(2)

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B \underline{\varepsilon}_t \quad [7.9]$$

$$M_k = A_1 M_{k-1} + A_2 M_{k-2}; k > 0 \quad [7.10]$$

7.4 การหาพารามิเตอร์ของแบบจำลองหลายตัวแปร (Multivariate Model)

7.4.1 การหาค่าพารามิเตอร์ของ Multivariate AR(1)

กรณี AR(1) จะสามารถหาค่าพารามิเตอร์ได้จากสมการ [7.8] เมื่อ $k=1$

$$\begin{aligned} M_1 &= A_1 M_0 \\ \hat{A}_1 &= \hat{M}_1 \hat{M}_0^{-1} \end{aligned} \quad [7.11]$$

$$\hat{B} \hat{B}^T = \hat{M}_0 - \hat{A}_1 \hat{M}_1^T \quad [7.12]$$

เมื่อทราบ $\hat{B} \hat{B}^T$ จะสามารถ \hat{B} ได้โดยการหารากที่สอง (Square Root) ของเมตริก $\hat{B} \hat{B}^T$ ซึ่งจะกล่าวถึงรายละเอียดในหัวข้อถัดไป

การ Derive หาพารามิเตอร์ A_1 และ B ของแบบจำลอง AR(1)

$$Z_t = A_1 Z_{t-1} + B \underline{\varepsilon}_t$$

Post multiply สมการแบบจำลอง AR(1) ด้วย Z_{t-k}^T

$$Z_t \cdot Z_{t-k}^T = A_1 Z_{t-1} Z_{t-k}^T + B \cdot \underline{\varepsilon}_t Z_{t-k}^T$$

$$E(Z_t \cdot Z_{t-k}^T) = A_1 E(Z_{t-1} Z_{t-k}^T) + B E(\underline{\varepsilon}_t \cdot Z_{t-k}^T); E(\underline{\varepsilon}_t \cdot Z_{t-k}^T) = 0$$

$$M_k = A_1 M_{k-1}; k > 0$$

ถ้า $k=1$

$$M_1 = A_1 M_0$$

Thus

$$A_1 = M_1 M_0^{-1}$$

ในทำนองเดียวกัน Post multiply สมการแบบจำลอง AR(1) ด้วย Z_t^T

$$Z_t = A_1 Z_{t-1} + B \underline{\varepsilon}_t$$

$$Z_t Z_t^T = A_1 Z_{t-1} Z_t^T + B \underline{\varepsilon}_t Z_t^T$$

$$E(Z_t Z_t^T) = A_1 E(Z_{t-1} Z_t^T) + B E(\underline{\varepsilon}_t Z_t^T)$$

$$M_0 = A_1 M_{-1} + B E(\underline{\varepsilon}_t Z_t^T) \quad [7.13]$$

หาเทอมสุดท้ายของสมการที่ (7.13)

$$Z_t^T = (A_1 Z_{t-1} + B \underline{\varepsilon}_t)^T$$

Pre-multiply Z_t^T ด้วย $\underline{\varepsilon}_t$

$$\underline{\varepsilon}_t Z_t^T = \underline{\varepsilon}_t (A_1 Z_{t-1} + B \underline{\varepsilon}_t)^T$$

$$E(\underline{\varepsilon}_t Z_t^T) = E(\underline{\varepsilon}_t (A_1 Z_{t-1} + B \underline{\varepsilon}_t)^T)$$

$$E(\underline{\varepsilon}_t Z_t^T) = E(\underline{\varepsilon}_t Z_{t-1}^T Z A_1^T + \underline{\varepsilon}_t \underline{\varepsilon}_t^T B^T)$$

$$E(\underline{\varepsilon}_t Z_t^T) = E(\underline{\varepsilon}_t Z_{t-1}^T Z A_1^T) + E(\underline{\varepsilon}_t \underline{\varepsilon}_t^T) B^T$$

$$E(\underline{\varepsilon}_t Z_{t-1}^T Z A_1^T) = 0$$

$$E(\underline{\varepsilon}_t \underline{\varepsilon}_t^T) = I$$

Thus

$$\begin{aligned}
 E(\underline{\varepsilon}_t Z_t^T) &= B^T \\
 \text{แทนค่า } E(\underline{\varepsilon}_t Z_t^T) &= B^T \text{ ลงในสมการที่ [7.13]} \\
 M_0 &= A_1 M_1^T + B B^T \\
 B B^T &= M_0 - A_1 M_1^T
 \end{aligned} \tag{7.14}$$

การหาค่าเมตริกซ์ B โดยวิธี Square Root

ถ้า B = Lower Triangular Matrix

และ D = เป็นเมตริกแบบ Positive Definite นั่นคือ

$$|a_{11}| > 0; \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0; \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0$$

ให้ $B B^T = D$ [7.15]

จะสามารถหาค่าของเมตริก B ได้ดังนี้

$$b^{ij} = \frac{d^{ji}}{b^{jj}}; j = 1; i = 1, \dots, n \tag{7.16}$$

$$b^{i1} = \frac{d^{1i}}{b^{11}}; i = 1, \dots, n \tag{7.17}$$

จากสมการ [7.17] จะสามารถหาค่า b^{i1} ได้ดังนี้

$$\begin{bmatrix} b^{11} = \sqrt{d^{11}} \\ b^{21} = \frac{d^{12}}{b^{11}} \\ b^{31} = \frac{d^{13}}{b^{11}} \\ \cdot \\ b^{n1} = \frac{d^{1n}}{b^{11}} \end{bmatrix}$$

$$b^{ij} = \sqrt{d^{ij} - \sum_{k=1}^{j-1} (b^{jk})^2}; j = 2, \dots, n; i = j \tag{7.18}$$

จากสมการ [7.18] จะสามารถหาค่า b^{ij} เมื่อ $i = j$ ได้ดังนี้

$$\begin{aligned}
 b^{22} &= \sqrt{d^{22} - (b^{21})^2} \\
 b^{33} &= \sqrt{d^{33} - (b^{31})^2 - (b^{32})^2} \\
 b^{nn} &= \sqrt{d^{nn} - \sum_{k=1}^{n-1} (b^{nk})^2} \\
 b^{ij} &= \frac{[d^{ij} - \sum_{k=1}^{j-1} b^{jk} b^{ik}]}{b^{jj}}; j = 2, \dots, n-1; i = j+1, \dots, n
 \end{aligned} \tag{7.19}$$

จากสมการ [7.19] จะสามารถหาค่า b^{ij} เมื่อ $i = j + 1$ ได้ดังนี้

$$\begin{bmatrix} b^{32} = (d^{32} - b^{21}b^{31})/b^{22} \\ b^{42} = (d^{42} - b^{21}b^{41})/b^{22} \\ b^{n2} = (d^{n2} - b^{21}b^{n1})/b^{22} \end{bmatrix}$$

$$\begin{bmatrix} b^{43} = (d^{43} - b^{31}b^{41} - b^{32}b^{42})/b^{33} \\ b^{53} = (d^{53} - b^{31}b^{51} - b^{32}b^{52})/b^{33} \\ b^{n3} = (d^{n3} - b^{31}b^{n1} - b^{32}b^{n2})/b^{33} \end{bmatrix}$$

$$b^{n,n-1} = \frac{[d^{n,n-1} - \sum_{k=1}^{n-2} b^{n-1,k} b^{n,k}]}{b^{n-1,n-1}} \quad [7.20]$$

7.4.2 การหาค่าพารามิเตอร์ของ Multivariate AR(2)

Multivariate AR(2) มีลักษณะดังสมการ [7.9]

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B \underline{\varepsilon}_t$$

ค่าพารามิเตอร์ของ Multivariate AR(2) สามารถหาได้จากสมการดังต่อไปนี้

$$\hat{A}_1 = [\hat{M}_1 - \hat{M}_2 \hat{M}_0^{-1} \hat{M}_1^T] [\hat{M}_0 - \hat{M}_1 \hat{M}_0^{-1} \hat{M}_1^T]^{-1} \quad [7.21]$$

$$\hat{A}_2 = [\hat{M}_2 - \hat{M}_1 \hat{M}_0^{-1} \hat{M}_1^T] [\hat{M}_0 - \hat{M}_1 \hat{M}_0^{-1} \hat{M}_1^T]^{-1} \quad [7.22]$$

$$\hat{B} \hat{B}^T = \hat{M}_0 - [\hat{A}_1 \hat{M}_1^T + \hat{A}_2 \hat{M}_2^T] \quad [7.23]$$

การ Derive หาค่าพารามิเตอร์ของ Multivariate AR(2)

จากสมการ [7.9]

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B \underline{\varepsilon}_t$$

$$Z_t Z_{t-k}^T = A_1 Z_{t-1} Z_{t-k}^T + A_2 Z_{t-2} Z_{t-k}^T + B \underline{\varepsilon}_t Z_{t-k}^T$$

$$E(Z_t Z_{t-k}^T) = A_1 E(Z_{t-1} Z_{t-k}^T) + A_2 E(Z_{t-2} Z_{t-k}^T) + B E(\underline{\varepsilon}_t Z_{t-k}^T)$$

$$M_k = A_1 M_{k-1} + A_2 M_{k-2} \quad [7.24]$$

จากสมการ [7.24] เมื่อ $k=1$

$$M_1 = A_1 M_0 + A_2 M_{-1}$$

จากสมการ [7.6] เมื่อ $k=1$; $M_{-1} = M_1^T$

$$M_1 = A_1 M_0 + A_2 M_1^T$$

$$A_1 M_0 = M_1 - A_2 M_1^T \quad [7.25]$$

จากสมการ [7.24] เมื่อ $k=2$

$$M_2 = A_1 M_1 + A_2 M_0 \quad [7.26]$$

สมการ [7.26] $\times M_0^{-1}$

$$M_2 M_0^{-1} = A_1 M_1 M_0^{-1} + A_2$$

$$A_2 = M_2 M_0^{-1} - A_1 M_1 M_0^{-1} \quad [7.27]$$

แทนค่า A_2 ลงในสมการ [7.25]

$$A_1 M_0 = M_1 - (M_2 M_0^{-1} - A_1 M_1 M_0^{-1}) M_1^T$$

$$A_1 M_0 = M_1 - M_2 M_0^{-1} M_1^T + A_1 M_1 M_0^{-1} M_1^T$$

$$A_1 (M_0 - M_1 M_0^{-1} M_1^T) = (M_1 - M_2 M_0^{-1} M_1^T)$$

$$A_1 = (M_1 - M_2 M_0^{-1} M_1^T) (M_0 - M_1 M_0^{-1} M_1^T)^{-1} \quad [7.28]$$

จากสมการ [7.26] จะได้ว่า

$$A_2 M_0 = M_2 - A_1 M_1 \quad [7.29]$$

สมการ [7.25] $\times M_0^{-1}$

$$A_1 = M_1 M_0^{-1} - A_2 M_1^T M_0^{-1} \quad [7.30]$$

แทนค่า A_1 ลงในสมการ [7.29]

$$A_2 M_0 = M_2 - (M_1 M_0^{-1} - A_2 M_1^T M_0^{-1}) M_1$$

$$A_2 M_0 = M_2 - M_1 M_0^{-1} M_1 + A_2 M_1^T M_0^{-1} M_1$$

$$A_2 (M_0 - M_1^T M_0^{-1} M_1) = (M_2 - M_1 M_0^{-1} M_1)$$

$$A_2 = (M_2 - M_1 M_0^{-1} M_1) (M_0 - M_1^T M_0^{-1} M_1)^{-1} \quad [7.31]$$

Post-multiplying สมการที่ [7.9] ด้วย Z_t^T

$$Z_t \cdot Z_t^T = A_1 Z_{t-1} \cdot Z_t^T + A_2 Z_{t-2} \cdot Z_t^T + B \underline{\varepsilon}_t Z_t^T$$

$$E(Z_t \cdot Z_t^T) = A_1 E(Z_{t-1} \cdot Z_t^T) + A_2 E(Z_{t-2} \cdot Z_t^T) + BE(\underline{\varepsilon}_t Z_t^T)$$

$$M_0 = A_1 M_{-1} + A_2 M_{-2} + BE(\underline{\varepsilon}_t \cdot Z_t^T)$$

$$M_0 = A_1 M_1^T + A_2 M_2^T + BE(\underline{\varepsilon}_t \cdot Z_t^T) \quad [7.32]$$

$$E(\underline{\varepsilon}_t Z_t^T) = E(\underline{\varepsilon}_t (A_1 Z_{t-1} + A_2 Z_{t-2} + B \underline{\varepsilon}_t)^T)$$

จาก $(A + B)^T = (A^T + B^T)$ และ $(AB)^T = B^T A^T$

$$E(\underline{\varepsilon}_t Z_t^T) = E[\underline{\varepsilon}_t \{ (A_1 Z_{t-1})^T + (A_2 Z_{t-2})^T + (B \underline{\varepsilon}_t)^T \}]$$

$$E(\underline{\varepsilon}_t Z_t^T) = E[\underline{\varepsilon}_t \{ Z_{t-1}^T A_1^T + Z_{t-2}^T A_2^T + \varepsilon_t^T B^T \}]$$

เนื่องจาก $E(\underline{\varepsilon}_t Z_{t-1}^T A_1^T) = 0$ และ $E(\underline{\varepsilon}_t Z_{t-2}^T A_2^T) = 0$

$$E(\underline{\varepsilon}_t Z_t^T) = E(\underline{\varepsilon}_t \cdot \varepsilon_t^T) B^T$$

$$E(\underline{\varepsilon}_t \cdot \varepsilon_t^T) = I$$

$$E(\underline{\varepsilon}_t Z_t^T) = B^T$$

แทนค่า $E(\underline{\varepsilon}_t Z_t^T)$ ลงในสมการ [7.32]

$$M_0 = A_1 M_1^T + A_2 M_2^T + BB^T$$

$$BB^T = M_0 - A_1 M_1^T - A_2 M_2^T \quad [7.33]$$

7.5 การสังเคราะห์ข้อมูลใหม่ (Generation of Synthetic Data)

ขั้นตอนการสังเคราะห์อนุกรมเวลา

(1) สมมติ $\hat{Z}_0 = [0]$

- (2) สั้งเคราะห์ Standard Normal Variate $\underline{\varepsilon}_t$
 (3) สั้งเคราะห์ค่า \hat{Z}_t จากสมการ
 (4) $\hat{Z}_t = \hat{A}_1 Z_{t-1} + B \underline{\varepsilon}_t ; t = 1, \dots, N'$
 $N' = N_g + N_w$ [7.34]

เมื่อ N' = จำนวนข้อมูลที่ต้องสังเคราะห์

N_g = จำนวนข้อมูลที่ต้องการสังเคราะห์

N_w = จำนวนข้อมูลที่ต้องสังเคราะห์เพื่อไว้ตัดทิ้ง (Warm-up Length) ซึ่งโดยทั่วไปจะกำหนดให้เท่ากับ 50

7.6 Approximate Multivariate ARMA(p, q)

จากวิธีการสร้างแบบจำลอง ARMA(p, q) กรณีตัวแปรเดียวของอนุกรมเวลา Z_t ในบทที่ 5 จะสามารถนำมาประยุกต์กับกรณีหลายตัวแปรได้ดังนี้

$$Z_t^{(i)} = \sum_{j=1}^{p(i)} \phi_j^{(i)} Z_{t-j}^{(i)} + \underline{\varepsilon}_t^{(i)} - \sum_{j=1}^{q(i)} \theta_j^{(i)} \varepsilon_{t-j}^{(i)} \quad [7.35]$$

เมื่อ i คือจำนวนอนุกรมเวลา มีค่าเท่ากับ $1, \dots, n$

$\underline{\varepsilon}_t^{(i)}$ = ตัวแปรสุ่มแบบ Time Independent แต่ Space Dependent หรือ Residual Series หรือ Cross Series Dependence ซึ่งมี Mean = 0 และ Variance = $\sigma_\varepsilon^{2(i)}$

$$\text{ให้ } \underline{\varepsilon}'_t^{(i)} = \frac{\underline{\varepsilon}_t^{(i)}}{\sigma_\varepsilon^{(i)}} ; i = 1, \dots, n \quad [7.36]$$

$\underline{\varepsilon}'_t^{(i)}$ มี Mean = 0 และ Variance = 1

$$\underline{\varepsilon}'_t^{(i)} = B \underline{\xi}_t \quad [7.37]$$

$\underline{\xi}_t$ = ตัวแปรสุ่มซึ่ง Independent in Time and Space (Standard Normal Variate)

$$\hat{B} \hat{B}^T = \hat{M}_o \quad [7.38]$$

\hat{M}_o คือ Lag-zero Cross Correlation Matrix ของอนุกรมเวลา $Z_t^{(i)}$

การสังเคราะห์ข้อมูลด้วย Approximate Multivariate ARMA(p,q) มีขั้นตอนดังต่อไปนี้

- (1) สั้งเคราะห์ Standardized Independent Normal Random Variate

$$\xi_0^{(i)}, \xi_{-1}^{(i)}, \xi_{-2}^{(i)}, \dots, \xi_{-q(i)+1}^{(i)} ; i = 1, \dots, n$$

$$(2) \text{ หาค่า Residuals } \underline{\varepsilon}'_0^{(i)} = B \underline{\xi}_0$$

$$\underline{\varepsilon}'_{-1}^{(i)} = B \underline{\xi}_{-1}$$

.

$$\underline{\varepsilon}'_{-q(i)+1}^{(i)} = B \underline{\xi}_{-q(i)+1}$$

$$\text{หา } \underline{\varepsilon}'_j^{(i)} = \sigma_\varepsilon^{(i)} \underline{\varepsilon}'_j ; i = 1, \dots, n ; j = 0, 1, \dots, 1 - q$$

$$(3) \text{ หา } Z_t^{(i)} = \sum_{j=1}^{p(i)} \phi_j^{(i)} Z_{t-j}^{(i)} + \sum_{j=0}^{q(i)} \theta_j^{(i)} \varepsilon_{t-j}^{(i)}$$

ให้สังเคราะห์ข้อมูล $t = 1, \dots, N'$ เมื่อ $N' = N_g + N_w$

7.7 การทดสอบความเหมาะสมของแบบจำลองรายตัวแปรรายปี (Goodness of Fit Test for Multivariate Annual Models)

การทดสอบความเหมาะสมของแบบจำลองมีขั้นตอนดังนี้

- (1) เปรียบเทียบ Multivariate Correlation Matrices M_0 , M_1 และ M_2 ของอนุกรมเวลาตัวอย่างกับอนุกรมเวลาที่สังเคราะห์
- (2) เปรียบเทียบ Correlogram ของอนุกรมเวลาตัวอย่างกับอนุกรมเวลาที่สังเคราะห์ พร้อมตรวจสอบ Probability Limit ด้วยสมการ [3.49]

$$r_{\alpha/2} = \pm \frac{z_{\alpha/2}}{N}$$

- (3) ทดสอบ Independence in Time และ Space ของ Multivariate Residuals Series ($\underline{\varepsilon}_t$) หรือ ($\underline{\xi}_t$)

$$\underline{\varepsilon}_t = B^{-1}(Z_t - A_1 Z_{t-1}) \quad \text{สำหรับ AR(1)}$$

$$\underline{\varepsilon}_t = B^{-1}(Z_t - A_1 Z_{t-1} - A_2 Z_{t-2}) \quad \text{สำหรับ AR(2)}$$

$$\underline{\xi}_t = B^{-1} \underline{\varepsilon}_t$$

$$\underline{\varepsilon}'_t^{(i)} = \frac{\varepsilon_t^{(i)}}{\sigma_\varepsilon^{(i)}}; i = 1, \dots, n \quad (\varepsilon'_t = \text{Standardized } \varepsilon_t)$$

- (4) ทดสอบสมมติฐานการแจกแจงแบบปกติของ Multivariate Residuals Series ($\underline{\varepsilon}_t$) หรือ ($\underline{\xi}_t$)

7.8 ตัวอย่างการจำลองอนุกรมเวลารายปีด้วย Multivariate AR(p)

จงสร้างแบบจำลองแบบหลายตัวแปรสำหรับอนุกรมเวลาน้ำท่ารายปีของ 4 สถานี Strausburg, Antietam, Point of Rocks และ Cumberland ของลุ่มน้ำโปโตแมก ซึ่งมีข้อมูลระหว่างปี 1931-1960 ดังตารางที่ 7.1 กำหนดให้สถานี

Strausburg=1, Antietam=2, Point of Rocks=3 และ Cumberland=4

ตารางที่ 7.1 Annual Flows (cfs) at the Stations Strausburg, Antietam, Point of Rocks and Cumberland (1931-1960)

Station Strausburg (1)									
282	675	344	325	804	330	742	741	745	500
1047	490	241	592	567	366	998	461	430	545
610	494	769	984	767	226	790	378	232	402
Station Antietam (2)									
158	309	268	268	242	127	383	479	423	278
378	228	192	257	231	181	357	242	270	280
251	209	336	340	251	173	344	168	125	229
Station Points of Rocks (3)									
4642	10100	7767	7056	11350	4665	10840	11010	13030	8543
13210	7867	5220	8828	8925	6849	13480	6744	7317	9108
9002	8692	12670	13440	10670	4856	12700	6920	4920	6190
Station Cumberland (4)									
731	1314	1192	1344	1649	643	1237	1336	1609	1231
1440	1228	750	1172	1439	1134	1652	986	1087	1175
1113	1218	1574	1570	1356	760	1480	1060	852	799

7.8.1 การวิเคราะห์เบื้องต้นและการระบุรูปแบบของแบบจำลอง (Preliminary Analysis and Model Identification)

ผลการตรวจสอบการแจกแจงแบบปกติของอนุกรมเวลาพบว่า ค่าสัมประสิทธิ์ความเบ้ (Skewness Coefficients) ของอนุกรมเวลาของสถานีต่างๆมีค่าดังนี้

$\hat{\gamma}_1 = 0.388$	N=30 , จากตารางที่ 3.2 $\gamma_{0.1}(30) = 0.662$ $\gamma_{0.02}(30) = 0.986$
$\hat{\gamma}_2 = 0.486$	
$\hat{\gamma}_3 = 0.160$	
$\hat{\gamma}_4 = -0.278$	

จึงสามารถสรุปได้ว่าอนุกรมเวลาทั้ง 4 มีการแจกแจงแบบปกติ

พล็อตอนุกรมเวลาดังรูปที่ 7.2 ซึ่งไม่ได้แสดง Long-Term Memory และ Long-Term Dependence ผลการคำนวณค่า Lag One ACF พบว่ามีค่าน้อยมาก จึงเป็นการสนับสนุนข้อสรุปดังกล่าว

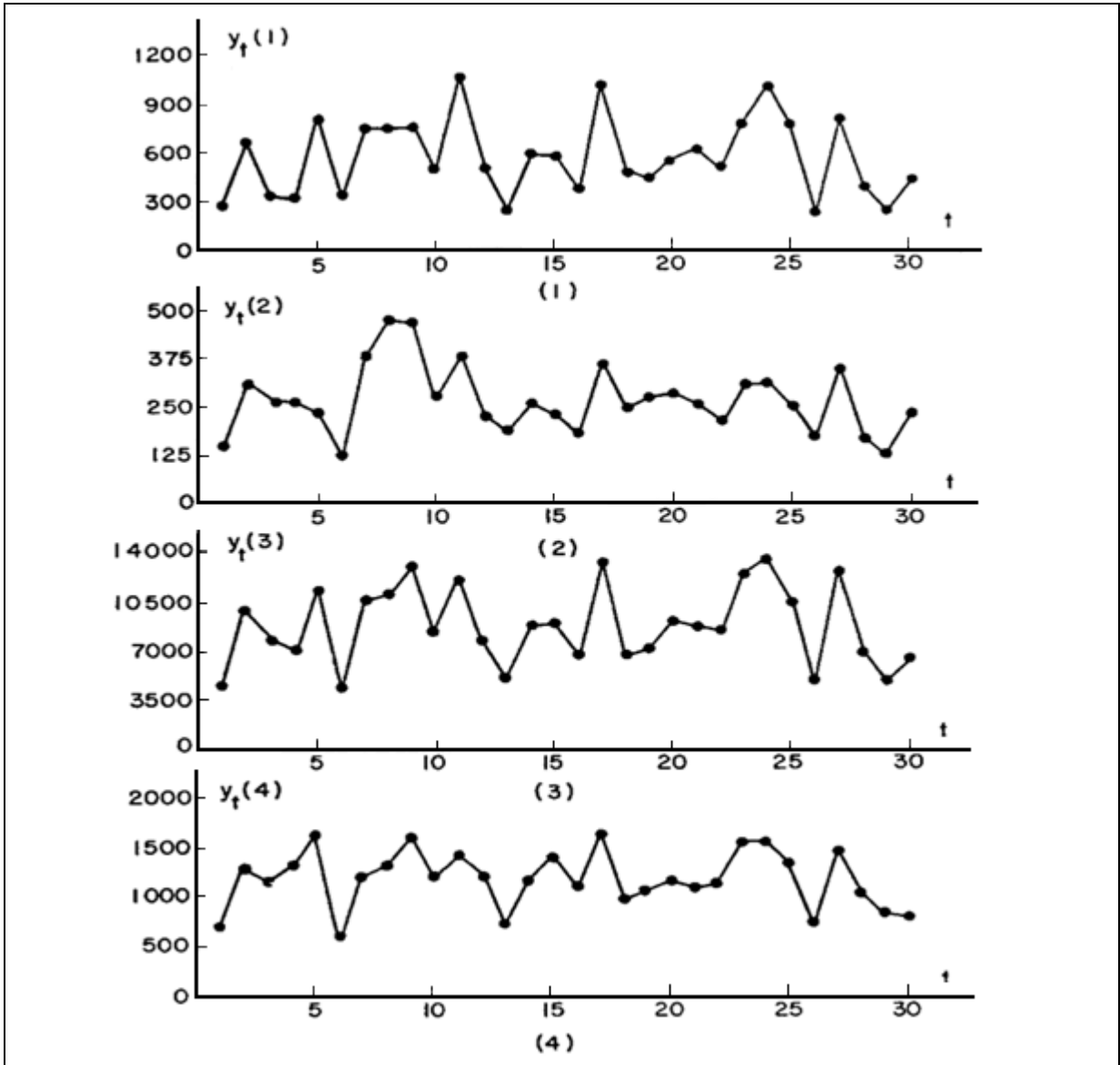
$r_1^{(1)} = -0.118$	เลือก Multivariate AR(1)
$r_1^{(2)} = -0.200$	
$r_1^{(3)} = -0.096$	
$r_1^{(4)} = -0.058$	

7.8.2 การประมาณค่าพารามิเตอร์

คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลาทั้ง 4 ดังแสดงอยู่ในตารางที่ 7.2

ตารางที่ 7.2 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลาทั้ง 4

อนุกรมเวลา (i)	Mean ($\hat{\mu}^{(i)}$)	Standard Deviation ($\hat{\sigma}^{(i)}$)
1	562.567	231.872
2	265.900	84.458
3	8,897.033	2,790.089
4	1,207.367	284.153



รูปที่ 7.2 Annual Flows (cfs) at the Stations (1)Struasburg (2)Antietam (3)Point of Rocks and (4)Cumberland (1931-1960)

คำนวณหา Standardized Series จากสมการ

$$z_t^{(i)} = \frac{y_t^{(i)} - \bar{\mu}^{(i)}}{\hat{\sigma}^{(i)}}$$

เช่น

$$z_1^{(1)} = \frac{282 - 562.567}{231.872} = -1.210$$

$$z_2^{(1)} = \frac{675 - 562.567}{231.872} = 0.485$$

$$z_{30}^{(1)} = \frac{402 - 562.567}{231.872} = -0.692$$

คำนวณหาค่า r_0^{ij} และ r_1^{ij} เมื่อ i และ $j = 1, 2, 3, 4$ จะได้ \hat{M}_0 และ \hat{M}_1 ดังนี้

$$\hat{M}_0 = \begin{bmatrix} 1.000 & 0.768 & 0.960 & 0.828 \\ 0.768 & 1.000 & 0.833 & 0.714 \\ 0.960 & 0.833 & 1.000 & 0.913 \\ 0.828 & 0.714 & 0.913 & 1.000 \end{bmatrix}$$

$$\hat{M}_1 = \begin{bmatrix} -0.118 & 0.067 & -0.079 & -0.110 \\ -0.129 & 0.200 & -0.119 & -0.242 \\ -0.148 & 0.121 & -0.096 & -0.147 \\ -0.118 & 0.208 & -0.040 & -0.058 \end{bmatrix}$$

$$r_k(\alpha) = \frac{-1 \pm u_{\alpha/2} \sqrt{N - k - 1}}{N - k}$$

$$r_1(0.95) = \frac{-1 \pm 1.96 \sqrt{30 - 1 - 1}}{30 - 1} = -0.39212, +0.32315$$

\hat{M}_1 แสดงให้เห็นว่า r_1^{ij} มีค่าน้อยมาก สูงสุดเท่ากับ 0.200 เท่านั้น แสดงว่าควรใช้เฉพาะ \hat{M}_0 ในการประมาณค่าพารามิเตอร์ อย่างไรก็ตามในตัวอย่างนี้จะแสดงวิธีการสร้างแบบจำลอง Multivariate AR(1) ต่อไป

คำนวณหา \hat{A}_1 และ $\hat{B}\hat{B}^T$ จากสมการ

$$A_1 = M_1 M_0^{-1}$$

$$B B^T = M_0 - A_1 M_1^T$$

$$\hat{A}_1 = \begin{bmatrix} -0.598 & 0.328 & 0.504 & -0.308 \\ -0.233 & 0.860 & -0.035 & -0.632 \\ -0.760 & 0.515 & 0.595 & -0.429 \\ -0.851 & 0.673 & 0.385 & -0.186 \end{bmatrix}$$

$$\hat{B}\hat{B}^T = \begin{bmatrix} 0.912 & 0.610 & 0.835 & 0.691 \\ 0.610 & 0.639 & 0.599 & 0.470 \\ 0.835 & 0.599 & 0.820 & 0.716 \\ 0.691 & 0.470 & 0.716 & 0.765 \end{bmatrix} = D$$

คำนวณหา \hat{B} โดยวิธี Square Root

$$\hat{b}^{11} = \frac{\hat{d}^{11}}{\hat{b}^{11}} = \frac{0.912}{\hat{b}^{11}}; \text{ or } \hat{b}^{11} = \sqrt{0.912} = 0.955$$

$$\hat{b}^{21} = \frac{\hat{d}^{12}}{\hat{b}^{11}} = \frac{0.610}{0.955} = 0.639$$

$$\hat{b}^{31} = \frac{\hat{d}^{13}}{\hat{b}^{11}} = \frac{0.835}{0.955} = 0.874$$

$$\hat{b}^{41} = \frac{\hat{d}^{14}}{\hat{b}^{11}} = \frac{0.691}{0.955} = 0.724$$

$$\hat{b}^{22} = \sqrt{\hat{d}^{22} - (\hat{b}^{21})^2} = \sqrt{0.639 - 0.639^2} = 0.481$$

$$\hat{b}^{32} = \frac{\hat{d}^{23} - \hat{b}^{21} \hat{b}^{31}}{\hat{b}^{22}} = \frac{0.599 - 0.639 \times 0.874}{0.481} = 0.084$$

$$\hat{B} = \begin{bmatrix} 0.955 & 0 & 0 & 0 \\ 0.639 & 0.481 & 0 & 0 \\ 0.874 & 0.084 & 0.222 & 0 \\ 0.714 & 0.061 & 0.367 & 0.325 \end{bmatrix}$$

7.8.3 การทดสอบ Goodness of Fit ของแบบจำลอง

(1) คำนวณหา Residual Matrix Series $\underline{\varepsilon}_t$ จากสมการ

$$\underline{\varepsilon}_t = \hat{B}^{-1}(Z_t - \hat{A}_1 Z_{t-1})$$

RESIDUAL SERIES NUMBER 1									
-1.266	0.454	-0.961	-1.472	0.946	-0.259	0.908	0.484	0.190	-0.749
2.066	-0.022	-1.272	-0.257	0.112	-0.472	2.086	-0.010	-0.617	-0.289
0.036	-0.245	1.032	1.882	1.317	-1.069	0.754	-0.941	-1.383	-0.964
RESIDUAL SERIES NUMBER 2									
-0.972	-0.144	1.188	1.419	-1.787	0.055	1.424	2.697	0.140	0.327
-0.258	-1.063	0.604	-1.181	-0.909	0.371	0.477	0.594	0.149	-0.217
-0.880	-1.080	1.472	0.077	-0.996	0.623	0.025	-0.963	-0.661	0.573
RESIDUAL SERIES NUMBER 3									
-1.517	-0.099	1.572	0.003	0.065	-0.944	-0.340	-1.722	1.513	-0.589
-1.801	0.351	-0.037	-0.928	0.417	1.005	-0.071	-1.094	-0.438	0.209
-0.690	1.423	2.314	-0.092	0.361	0.204	1.607	1.607	0.156	-0.952
RESIDUAL SERIES NUMBER 4									
-0.544	0.627	0.095	2.651	1.053	-1.271	0.093	0.550	-1.601	-0.129
-0.802	1.320	-1.477	0.156	2.316	0.505	0.848	0.498	0.504	-1.061
-0.940	-0.189	-0.243	-0.380	0.097	-0.465	-1.207	-0.174	-0.010	-1.091

(2) คำนวณหา Lag-Zero Cross Correlation Matrix $M_0(\varepsilon)$

$$M_0(\varepsilon) = \begin{bmatrix} 1.000 & -0.011 & -0.002 & 0.003 \\ -0.011 & 1.000 & -0.010 & 0.011 \\ -0.002 & -0.010 & 1.000 & 0.002 \\ 0.003 & 0.011 & 0.002 & 1.000 \end{bmatrix}$$

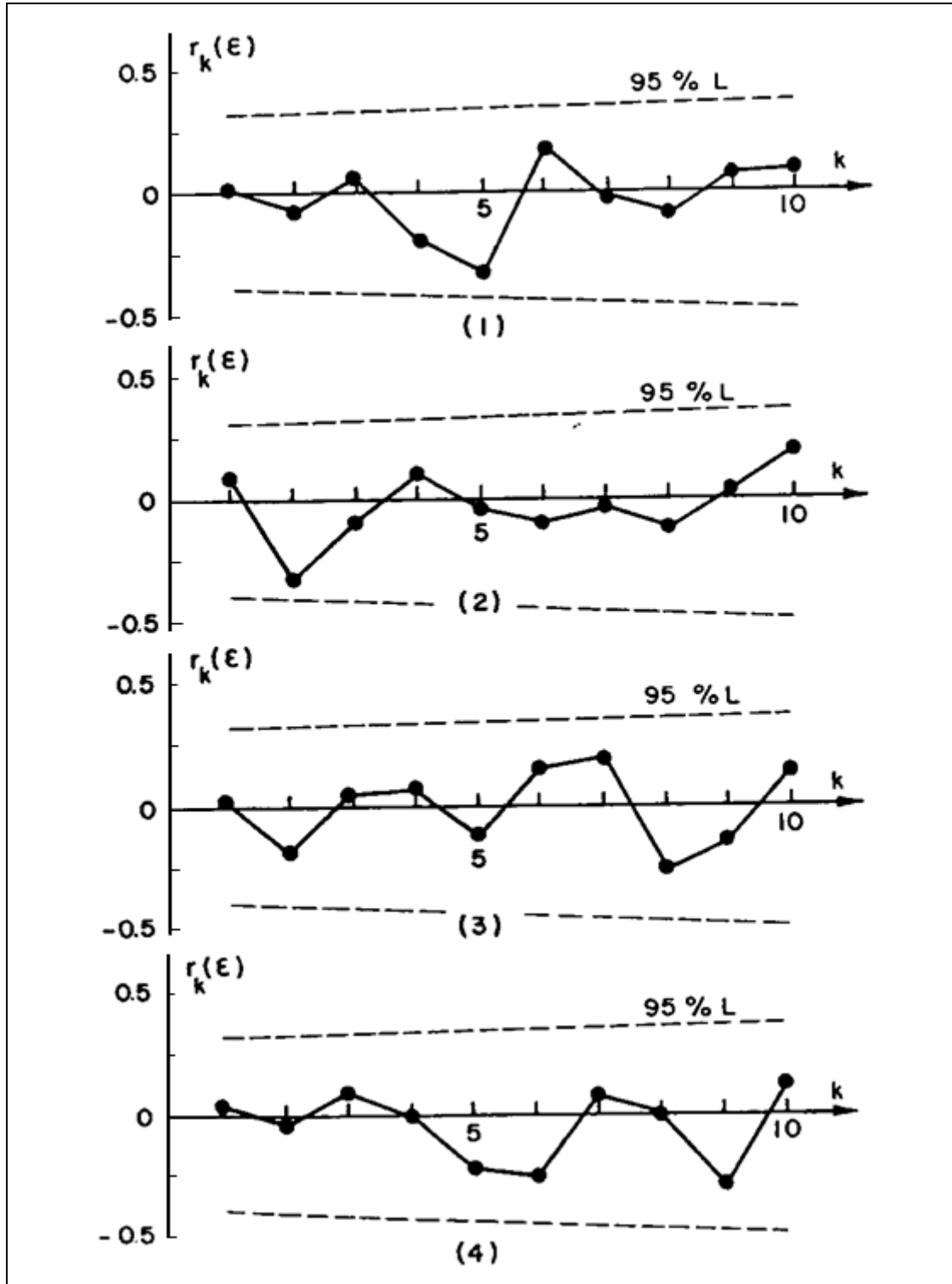
(3) Probability Limit ของ Lag-Zero Cross Correlation จะคำนวณจากสมการ

$$r(\alpha) = \pm \frac{u_{\alpha/2}}{\sqrt{N}}$$

$$r(0.05) = \pm \frac{1.96}{\sqrt{30}} = \pm 0.3578$$

(4) ค่า Lag-Zero Cross Correlation r_0^{ij} เมื่อ $i \neq j$ มีค่าอยู่ใน Probability Limit แสดงว่า $\underline{\varepsilon}_t$ เป็นอิสระและสรุปได้ว่า AR(1) เหมาะสมกับอนุกรมเวลาตัวอย่าง

(5) คำนวณหา Correlogram $r_k^{ii}(\varepsilon)$ ($i = 1, 2, 3, 4$) โดยใช้สมการ [2.5b] และคำนวณหา Probability Limit จากสมการ $r_k(\alpha) = \frac{-1 \pm u_{\alpha/2} \sqrt{N-k-1}}{N-k}$ ได้ผลดังรูปที่ 7.2



รูปที่ 7.3 Correlogram $r_k^{ii}(\varepsilon)$ ($i=1,2,3,4$) และ 95% Probability Limit

ค่า Correlation r_k^{ii} เมื่อ $i=1,2,3,4$ มีค่าอยู่ใน Probability Limit แสดงว่า ε_t^{ii} เป็นอิสระและสับสนุน
ข้อสรุปที่ว่า AR(1) เหมาะสมกับอนุกรมเวลาตัวอย่าง

ทดสอบว่า Residuals มีการแจกแจงแบบปกติ โดยใช้การทดสอบค่าสัมประสิทธิ์ความเบ้ ซึ่งผลการทดสอบแสดงอยู่ในตาราง และสามารถสรุปได้ว่า Residual Series มีการแจกแจงแบบปกติ

Residual Series (i)	Computed γ_i	Critical $\gamma(0.02)$	Hypothesis of Normality
1	0.541	0.986	Accepted
2	0.601	0.986	Accepted
3	0.305	0.986	Accepted
4	0.689	0.986	Accepted

7.8.4 Optional Tests of the Model

ใช้ Multivariate AR(1) ในการสังเคราะห์ห้อนุกรมเวลา แล้วทำการทดสอบความเหมาะสมตามวิธีการในหัวข้อ (3) นั่นคือการทดสอบความเป็นอิสระของ Residual กับพื้นที่โดยใช้ Lag-Zero Cross Correlation ($M_0(\epsilon)$) และทดสอบความเป็นอิสระของ Residual กับเวลาโดยใช้ Correlogram $r_k^{ii}(\epsilon)$

ตัวอย่างการสังเคราะห์ห้อนุกรมเวลา

$\hat{z}_t^{(i)} = \frac{\hat{y}_t^{(i)} - \hat{\mu}^{(i)}}{\hat{\sigma}^{(i)}}$	$\hat{y}_t^{(1)} = 562.6 + 231.9 \hat{z}_t^{(1)}$ $\hat{y}_t^{(2)} = 265.9 + 84.5 \hat{z}_t^{(2)}$ $\hat{y}_t^{(3)} = 8,897.0 + 2,790.1 \hat{z}_t^{(3)}$ $\hat{y}_t^{(4)} = 1,204.4 + 284.2 \hat{z}_t^{(4)}$
--	---

$$\hat{Z}_t = A_1 \hat{Z}_{t-1} + B \epsilon_t$$

ϵ_t = Independent Normal Random Number มี Mean=0 และ Standard Deviation=1

$$\hat{z}_t^{(1)} = -0.598 \hat{z}_{t-1}^{(1)} + 0.328 \hat{z}_{t-1}^{(2)} + 0.504 \hat{z}_{t-1}^{(3)} - 0.308 \hat{z}_{t-1}^{(4)} + 0.955 \epsilon_t^{(1)}$$

$$\hat{z}_t^{(2)} = -0.233 \hat{z}_{t-1}^{(1)} + 0.860 \hat{z}_{t-1}^{(2)} - 0.035 \hat{z}_{t-1}^{(3)} - 0.632 \hat{z}_{t-1}^{(4)} + 0.639 \epsilon_t^{(1)} + 0.481 \epsilon_t^{(2)}$$

$$\hat{z}_t^{(3)} = -0.760 \hat{z}_{t-1}^{(1)} + 0.515 \hat{z}_{t-1}^{(2)} + 0.595 \hat{z}_{t-1}^{(3)} - 0.429 \hat{z}_{t-1}^{(4)} + 0.874 \epsilon_t^{(1)} + 0.084 \epsilon_t^{(2)} + 0.222 \epsilon_t^{(3)}$$

$$\hat{z}_t^{(4)} = -0.851 \hat{z}_{t-1}^{(1)} + 0.673 \hat{z}_{t-1}^{(2)} + 0.385 \hat{z}_{t-1}^{(3)} - 0.186 \hat{z}_{t-1}^{(4)} + 0.724 \epsilon_t^{(1)} + 0.016 \epsilon_t^{(2)} + 0.367 \epsilon_t^{(3)} + 0.325 \epsilon_t^{(4)}$$

สังเคราะห์ค่า Standardized Normal Random Number $\epsilon_t^{(i)}$ เมื่อ t=1 ดังตาราง

t	(i)	$\epsilon_t^{(i)}$
1	1	-0.155
	2	0.420
	3	0.360
	4	-0.595

จากค่า $\epsilon_t^{(i)}$ ที่สังเคราะห์ได้ดังตารางข้างบน ถ้ากำหนดให้ $\hat{z}_0^{(i)} = 0$ เมื่อ i=1,2,3,4 จะสามารถสังเคราะห์หาค่า $\hat{z}_1^{(i)}$ ได้ดังนี้

$$\hat{z}_1^{(1)} = 0.955(-0.155) = -0.148$$

$$\begin{aligned}\hat{z}_t^{(2)} &= 0.639(-0.155) + 0.481(0.420) = 0.103 \\ \hat{z}_t^{(3)} &= 0.874(-0.155) + 0.084(0.420) + 0.222(0.360) = -0.020 \\ \hat{z}_t^{(4)} &= 0.724(-0.155) + 0.016(0.420) + 0.365(0.360) + 0.325(-0.525) = -0.144\end{aligned}$$

หลังจากนั้นจึงคำนวณหาค่า $\hat{y}_t^{(i)}$ ดังนี้

$$\begin{aligned}\hat{y}_t^{(1)} &= 562.6 + 231.9(-0.148) = 528.3 \\ \hat{y}_t^{(2)} &= 265.9 + 84.5(0.103) = 274.6 \\ \hat{y}_t^{(3)} &= 8,897.0 + 2,790.1(-0.020) = 8,841.2 \\ \hat{y}_t^{(4)} &= 1,204.4 + 284.2(-0.144) = 1,163.5\end{aligned}$$

ในการทำงานเดียวกันหาค่า Standardized Normal Random Number $\varepsilon_t^{(i)}$ เมื่อ $t=2$ ดังตาราง

t	(i)	$\varepsilon_t^{(i)}$
2	1	1.520
	2	-0.050
	3	-0.650
	4	0.265

จะสามารถหาค่า $\hat{z}_2^{(i)}$ และ $\hat{y}_2^{(i)}$ ได้ดังนี้

$$\begin{aligned}\hat{z}_2^{(1)} &= -0.598(-0.148) + 0.328(0.103) + 0.504(-0.020) - 0.308(-0.144) \\ &\quad + 0.955(1.520) = 1.608 \\ \hat{z}_2^{(2)} &= -0.233(-0.148) + 0.860(0.103) + 0.035(-0.020) - 0.632(-0.144) \\ &\quad + 0.639(1.520) + 0.481(-0.050) = 1.162 \\ \hat{z}_2^{(3)} &= -0.760(-0.148) + 0.515(0.103) + 0.595(-0.020) + 0.874(-0.144) \\ &\quad + 0.874(1.520) + 0.084(-0.050) + 0.222(-0.650) = 1.208 \\ \hat{z}_2^{(4)} &= -0.851(-0.148) + 0.673(0.103) + 0.385(-0.020) - 0.186(-0.144) \\ &\quad + 0.724(1.520) + 0.016(-0.050) + 0.367(-0.650) + 0.325(0.265) \\ &= 1.162 \\ \hat{y}_2^{(1)} &= 562.6 + 231.9(1.608) = 935.5 \\ \hat{y}_2^{(2)} &= 265.9 + 84.5(1.162) = 364.1 \\ \hat{y}_2^{(3)} &= 8,897.0 + 2,790.1(1.208) = 12,267.4 \\ \hat{y}_2^{(4)} &= 1,204.4 + 284.2(1.162) = 1,534.6\end{aligned}$$

7.8.5 Reliability ของพารามิเตอร์แบบจำลอง

คือการหาช่วงความเชื่อมั่นพารามิเตอร์ของแบบจำลองจากข้อมูลที่สังเคราะห์ขึ้นมา

7.9 ข้อจำกัดของ Multivariate Annual Modeling

แบบจำลอง Multivariate AR(1) มีคุณสมบัติที่คงค่า Mean, Standard Deviation, Lag-One Cross Correlation และ Lag-One Autocorrelation หรือเป็นแบบจำลองแบบ Short Memory จึงไม่สามารถคงค่า Long Term Persistence หรือ Long Term Dependence เช่น Drought หรือ Storage Related Characteristics ได้ ดังนั้น ถ้าต้องการแบบจำลองที่สามารถคงค่า Long Term Persistence ต้องใช้แบบจำลองอื่นเช่น ARMA(p,q)

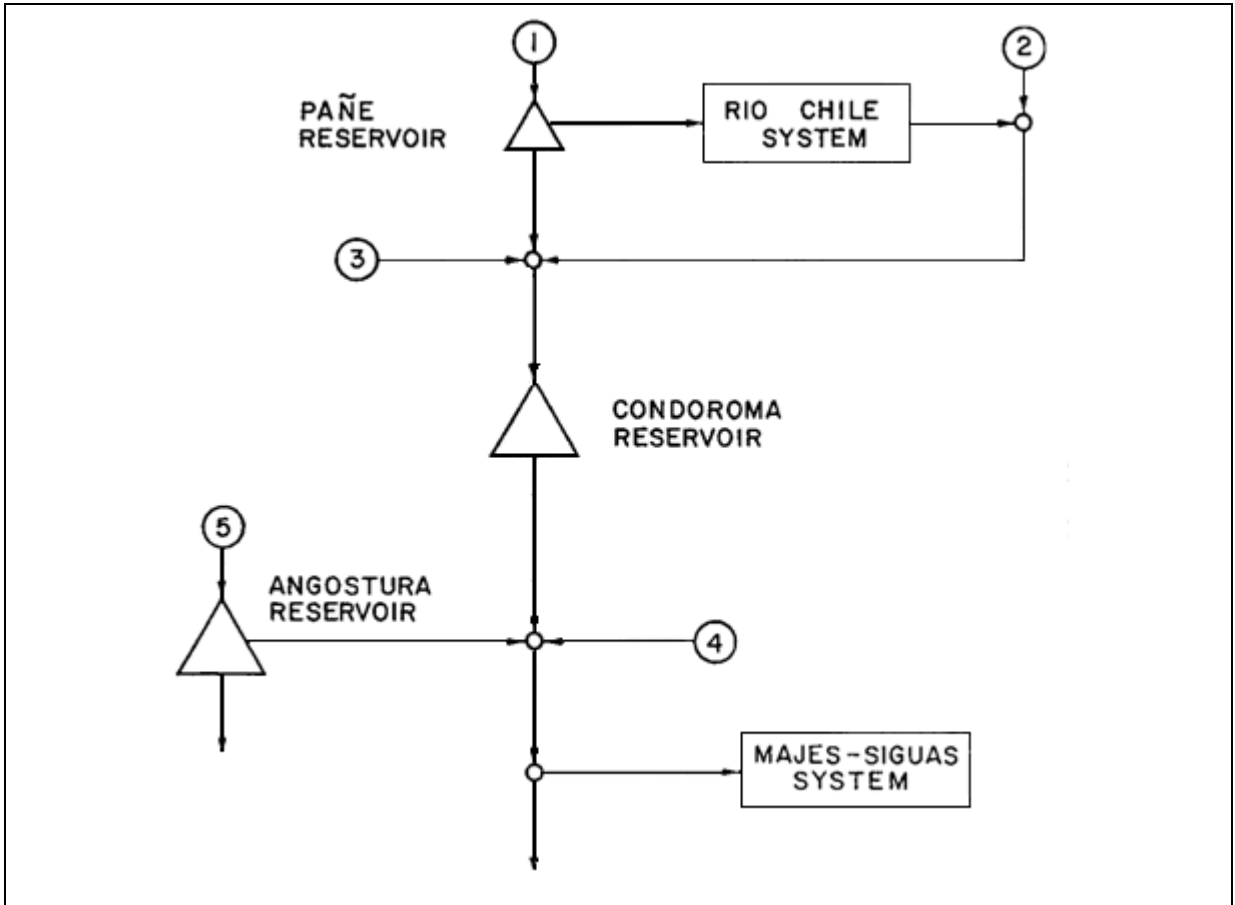
กรณีที่อนุกรมเวลาไม่ได้มีการแจกแจงแบบปกติ ทางเลือกในการจำลองมี 3 แนวทางคือ

- การแปลงอนุกรมเวลาแบบ Non-normal เป็น Normal โดยใช้ฟังก์ชันการแปลงที่เหมาะสม แล้วจึงทำการจำลองอนุกรมเวลาที่แปลงค่าแล้ว
- ทำการจำลองอนุกรมเวลาแบบ Non-normal แล้วหาฟังก์ชันการแจกแจงความน่าจะเป็นของ Uncorrelated Residuals
- หาความสัมพันธ์ระหว่างโมเมนต์ที่ 1 และ 2 ของอนุกรมเวลาแบบไม่ปกติ กับโมเมนต์ที่ 1 และ 2 ของอนุกรมเวลาแบบปกติ เพื่อจะได้คงค่าโมเมนต์ที่ 1 และ 2 ของอนุกรมเวลาแบบไม่ปกติ ตามวิธีที่กล่าวถึงในหัวข้อ 3.2.3

การหาค่าพารามิเตอร์ของแบบจำลองแบบหลายตัวแปร จำเป็นที่เมทริกซ์ $\hat{M}_0, \hat{B}\hat{B}^T$ ต้องเป็นเมทริกซ์แบบ Consistent (Positive Definite หรือ Positive Semi Definite) ถ้าเมทริกซ์ $\hat{M}_0, \hat{B}\hat{B}^T$ เป็นแบบ Inconsistent (ไม่เป็น Positive Definite หรือ Positive Semi Definite) จะไม่สามารถหาค่าพารามิเตอร์ได้ Crosby และ Maddock(1970) เสนอแนะวิธีการหาเมทริกซ์ $\hat{M}_0, \hat{B}\hat{B}^T$ ที่เป็นเมทริกซ์แบบ Consistent

7.10 การประยุกต์ใช้แบบจำลองแบบหลายตัวแปรกับอนุกรมเวลารายปี (Annual Multivariate Model) ในทางปฏิบัติ

การใช้แบบจำลองแบบหลายตัวแปรกับอนุกรมเวลารายปี ในการสังเคราะห์อนุกรมเวลาของหลายสถานี มีวัตถุประสงค์เพื่อการวางแผนระบบแหล่งน้ำซึ่งมีการใช้งานข้ามปี (Over-year Regulation) แต่กรณีระบบแหล่งน้ำแบบมีหลายอ่างเก็บน้ำ (Multireservoir) ที่มีการใช้งานทั้งในลักษณะข้ามปีและใช้งานเป็นฤดูกาลและจำเป็นต้องสังเคราะห์อนุกรมเวลารายฤดูกาล การสังเคราะห์อนุกรมเวลารายฤดูกาลทำได้ 2 แบบคือแบบที่ 1 การใช้แบบจำลองแบบหลายตัวแปรสังเคราะห์อนุกรมเวลารายฤดูกาล และแบบที่ 2 การจำลองโดยใช้แบบจำลอง Disaggregation แบบที่ 2 จะแบ่งการสังเคราะห์อนุกรมเวลาเป็น 2 ชั้นคือชั้นที่ 1 การสังเคราะห์อนุกรมเวลารายปีของหลายสถานี และชั้นที่ 2 คือการแตกอนุกรมเวลารายปีเป็นรายฤดูกาล (Disaggregation) รายละเอียดการใช้แบบจำลองแบบ Disaggregation จะได้กล่าวถึงในบทที่ 8 ตัวอย่างการสังเคราะห์อนุกรมเวลารายเดือนของ 5 สถานี ดังรูปที่ 7.4 สามารถทำได้ 2 แบบตามที่กล่าวมาแล้ว



รูปที่ 7.4 ระบบชลประทาน Majes-Siguas ในประเทศเปรูซึ่งต้องการสังเคราะห์อนุกรมเวลาของ 5 สถานี

7.11 การจำลองแบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic (Multivariate Modeling of Periodic Time Series)

ตามที่กล่าวมาแล้ว การจำลองแบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic สามารถทำได้ดังนี้
 ใช้แบบจำลอง AR แบบ Constant หรือ Periodic พารามิเตอร์
 ใช้แบบจำลอง ARMA แบบ Constant หรือ Periodic พารามิเตอร์

7.11.1 การประยุกต์ใช้แบบจำลอง AR และ ARMA แบบ Constant พารามิเตอร์กับกรณีหลายตัวแปร
 ตัวอย่างการประยุกต์ใช้แบบจำลอง AR(1), AR(2) และ ARMA(1,1) แบบ Constant พารามิเตอร์กับ
 กรณีหลายตัวแปร สามารถทำได้ดังนี้

ให้ $Z_t^{(i)} = Z_{v,\tau}^{(i)}$ [7.39]

เมื่อ $t = (v - 1)\omega + \tau$ และ $i = 1, \dots, n$

กรณี AR(1): $Z_t = A_1 Z_{t-1} + B \epsilon_t$

กรณี AR(2): $Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + B \epsilon_t$

กรณี ARMA(1,1): $Z_t = A_1 Z_{t-1} + B_0 \epsilon_t - B_1 \epsilon_{t-1}$

7.11.2 แบบจำลอง AR หลายตัวแปรแบบ Periodic พารามิเตอร์ (Multivariate AR Models with Periodic Parameters)

$$Y_{v,\tau} = \underline{\mu}_\tau + \underline{\sigma}_\tau Z_{v,\tau} \quad [7.40]$$

$$Y_{v,\tau} = \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \\ \vdots \\ y_{v,\tau}^{(n)} \end{bmatrix}; \quad \underline{\mu}_\tau = \begin{bmatrix} \mu_\tau^{(1)} \\ \mu_\tau^{(2)} \\ \vdots \\ \mu_\tau^{(n)} \end{bmatrix}$$

$$\underline{\sigma}_\tau = \begin{bmatrix} \sigma_\tau^{(1)} & 0 & \cdot & 0 \\ 0 & \sigma_\tau^{(2)} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \sigma_\tau^{(n)} \end{bmatrix} = \text{Diagonal Matrix}$$

$$Z_{v,\tau} = \begin{bmatrix} Z_{v,\tau}^{(1)} \\ Z_{v,\tau}^{(2)} \\ \vdots \\ Z_{v,\tau}^{(n)} \end{bmatrix}$$

แบบจำลอง AR(1) แบบ Periodic พารามิเตอร์

$$Z_{v,\tau} = A_{1,\tau} Z_{v,\tau-1} + B_\tau \underline{\varepsilon}_{v,\tau} \quad [7.41]$$

$A_{1,\tau} = (n \times n)$ Periodic Coefficient Matrix of Period τ

$B_\tau = (n \times n)$ Periodic Coefficient Matrix of Period τ

$\underline{\varepsilon}_{v,\tau} = (n \times 1)$ Vector of Independent Normally Distributed Random Variable

$$E(Z_{v,\tau}) = E(\underline{\varepsilon}_{v,\tau}) = 0 \quad [7.42]$$

$$\text{Var}(Z_{v,\tau}) = \text{Var}(\underline{\varepsilon}_{v,\tau}) = 1 \quad [7.43]$$

$$E(\underline{\varepsilon}_{v,\tau}^{(i)} \cdot \underline{\varepsilon}_{v,\tau}^{(j)}) = 0; i \neq j \quad [7.44]$$

$$E(\underline{\varepsilon}_{v,\tau} \cdot \underline{\varepsilon}_{v,\tau}^T) = I = \text{Identity Matrix} \quad [7.45]$$

การประมาณค่า Periodic พารามิเตอร์

$$Z_{v,\tau}^{(i)} = \frac{y_{v,\tau}^{(i)} - \mu_\tau^{(i)}}{\sigma_\tau^{(i)}} \quad [7.46]$$

$$\hat{A}_{1,\tau} = \hat{M}_{1,\tau} \hat{M}_{0,\tau-1}^{-1} \quad [7.47]$$

$$\hat{B}_\tau \hat{B}_\tau^T = \hat{M}_{0,\tau} - \hat{M}_{1,\tau} \hat{M}_{0,\tau-1}^{-1} \hat{M}_{1,\tau}^T \quad [7.48]$$

$\hat{M}_{0,\tau}, \hat{M}_{0,\tau-1}$ และ $\hat{M}_{1,\tau}^T = \text{Periodic Correlation Matrix}$

$$\hat{M}_{k,\tau} = \begin{bmatrix} r_{k,\tau}^{11} & r_{k,\tau}^{12} & \cdot & r_{k,\tau}^{1n} \\ r_{k,\tau}^{21} & r_{k,\tau}^{22} & \cdot & r_{k,\tau}^{2n} \\ \cdot & \cdot & \cdot & \cdot \\ r_{k,\tau}^{n1} & r_{k,\tau}^{n2} & \cdot & r_{k,\tau}^{nn} \end{bmatrix} \quad [7.49]$$

เมื่อ $k = 0, 1$ และ $\tau = 1, \dots, \omega$

$$r_{k,\tau}^{ij} = E \left[Z_{v,\tau}^{(i)} Z_{v,\tau-k}^{(j)} \right] = \text{Periodic Lag } k \text{ Correlation Matrix} \quad [7.50]$$

การสังเคราะห์อนุกรมเวลาด้วยแบบจำลองหลายตัวแปรแบบ Periodic พารามิเตอร์

Multivariate AR(1): $Z_{v,\tau} = A_{1,\tau} Z_{v,\tau-1} + B_{\tau} \underline{\varepsilon}_{v,\tau}$

- (1) สังเคราะห์ $\underline{\varepsilon}_{1,1}$
- (2) สมมติ $Z_{1,0} = [0]$
- (3) คำนวณ $Z_{1,1} = A_{1,1} Z_{1,0} + B_1 \underline{\varepsilon}_{1,1}$
- (4) สังเคราะห์ $\underline{\varepsilon}_{1,2}$
- (5) $Z_{1,2} = A_{1,2} Z_{1,1} + B_2 \underline{\varepsilon}_{1,2}$
- (6) ทำซ้ำขั้น 4 และ 5 จนกระทั่งได้ $\underline{\varepsilon}_{1,\omega}$ และ $Z_{1,\omega} = A_{1,\omega} Z_{1,\omega-1} + B_1 \underline{\varepsilon}_{1,\omega}$
- (7) สังเคราะห์ $\underline{\varepsilon}_{2,1}$
- (8) สมมติ $Z_{2,0} = [0] = [Z_{1,12}]$
- (9) $Z_{2,1} = A_{1,1} Z_{2,0} + B_1 \underline{\varepsilon}_{2,1}$
- .
- (10) $Z_{2,\omega} = A_{1,\omega} Z_{2,\omega-1} + B_{\omega} \underline{\varepsilon}_{2,\omega}$
- (11) ทำซ้ำจนกระทั่ง $V = N' = N_g + N_{\omega}$

7.11.3 แบบจำลอง ARMA หลายตัวแปรแบบ Periodic พารามิเตอร์ (Multivariate ARMA Models

with Periodic Parameters)

Multivariate ARMA(1,1)

$$Z_t = A_1 Z_{t-1} + B_0 \underline{\varepsilon}_t - B_1 \underline{\varepsilon}_{t-1} \quad [7.51]$$

การประมาณค่าพารามิเตอร์

$$\begin{aligned} Z_t \cdot Z_{t-k}^T &= A_1 Z_{t-1} \cdot Z_{t-k}^T + B_0 \underline{\varepsilon}_t \cdot Z_{t-k}^T - B_1 \underline{\varepsilon}_{t-1} Z_{t-k}^T \\ E(Z_t \cdot Z_{t-k}^T) &= A_1 E(Z_{t-1} \cdot Z_{t-k}^T) + B_0 E(\underline{\varepsilon}_t Z_{t-k}^T) - B_1 E(\underline{\varepsilon}_{t-1} Z_{t-k}^T) \\ M_k &= A_1 M_{k-1} + B_0 E(\underline{\varepsilon}_t Z_{t-k}^T) - B_1 E(\underline{\varepsilon}_{t-1} Z_{t-k}^T) \end{aligned}$$

ถ้า $k > q$ เช่น $k=2$ และ $q=1$ จะได้

$$M_2 = A_1 M_1 \quad [7.52]$$

ถ้า $k \leq q$ เช่น $k=1$ และ $q=1$

$$\begin{aligned} M_1 &= A_1 M_0 - B_1 E(\underline{\varepsilon}_{t-1} Z_{t-1}^T) \\ E(\underline{\varepsilon}_{t-1} Z_{t-1}^T) &= E[\underline{\varepsilon}_{t-1} (A_1 Z_{t-2} + B_0 \underline{\varepsilon}_{t-1} - B_1 \underline{\varepsilon}_{t-2})^T] \\ &= E(\underline{\varepsilon}_{t-1} Z_{t-2}^T A_1^T) + E(\underline{\varepsilon}_{t-1} \underline{\varepsilon}_{t-1}^T B_0^T) - E(\underline{\varepsilon}_{t-1} \underline{\varepsilon}_{t-2}^T B_1^T) \\ &= B_0^T \end{aligned}$$

$$M_1 = A_1 M_0 - B_1 B_0^T \quad [7.53]$$

จากสมการ [7.52] จะได้ $A_1 = M_2 M_1^{-1}$

แทนค่า A1 ลงไปใน [7.53] จะได้

$$B_1 B_0^T = M_2 M_1^{-1} M_0 - M_1 \quad [7.54]$$

$$B_0 B_0^T = M_0 \quad [7.55]$$

7.12 ตัวอย่างการจำลองแบบหลายตัวแปรกับอนุกรมเวลา Periodic

ระบบ Great Lakes ประกอบไปด้วยทะเลสาบ Superior, Michigan, Huron, Erie, Ontario และ St.Clair ทะเลสาบ Superior เป็นทะเลสาบที่ใหญ่และสำคัญที่สุดของระบบ Great Lakes ระบายน้ำผ่านแม่น้ำ St.Mary ลงสู่ทะเลสาบ Huron ทะเลสาบ Michigan และ Huron มีระดับน้ำเท่ากัน ดังนั้นในทางอุทกวิทยาจึงพิจารณาว่ารวมเป็นทะเลสาบเดียวกันโดยเรียกว่า ทะเลสาบ Michigan-Huron น้ำจากสองทะเลสาบนี้จะระบายผ่านแม่น้ำ St.Clair ลงสู่ทะเลสาบ St.Clair และหลังจากนั้นจะระบายผ่านแม่น้ำ Detroit ลงสู่แม่น้ำ Erie จุดทางออกของทะเลสาบ Erie คือแม่น้ำ Niagara ซึ่งระบายน้ำลงสู่ทะเลสาบ Ontario ซึ่งเป็นจุดต่ำสุดของ Great Lakes ต้องการศึกษาวางแผนการระบายน้ำของ Great Lakes จึงต้องสร้างแบบจำลองสโตแคสติกเพื่อใช้ในการสังเคราะห์อนุกรมเวลารายเดือนของ Net Basin Supplies (NBS) เนื่องจากขนาดของทะเลสาบ St.Clair มีขนาดเล็กมากเมื่อเทียบกับอีก 4 ทะเลสาบ ดังนั้นการจำลองจะแบ่งออกเป็น 2 ขั้นตอนคือ ขั้นที่ 1 ทำแบบจำลองแบบหลายตัวแปรของทะเลสาบ Superior, Michigan-Huron, Erie และ Ontario และขั้นที่ 2 จึงใช้ Multiple Linear Regression สังเคราะห์อนุกรมเวลาของ NBS ของทะเลสาบ St.Clair อนุกรมเวลา NBS รายเดือนของทะเลสาบทั้ง 4 แสดงอยู่ในตารางที่ 7.3(a)-(d) ในบทนี้จะกล่าวถึงเฉพาะ

ตารางที่ 7.3(a) Monthly Net Basin Supplies สำหรับทะเลสาบ Ontario (1900-1968)

29.0	58.0	38.0	106.0	34.0	30.0	28.0	10.0	-1.0	-2.0	32.0	23.0
24.0	4.0	78.0	153.0	54.0	64.0	26.0	18.0	9.0	-9.0	0.0	61.0
28.0	8.0	128.0	64.0	45.0	55.0	77.0	23.0	2.0	0.0	15.0	7.0
11.0	59.0	132.0	92.0	20.0	41.0	43.0	17.0	-1.0	20.0	-2.0	-9.0
11.0	58.0	106.0	125.0	82.0	56.0	43.0	27.0	7.0	18.0	-20.0	-10.0
22.0	-1.0	61.0	91.0	60.0	65.0	47.0	36.0	-3.0	17.0	6.0	30.0
57.0	27.0	34.0	67.0	37.0	56.0	34.0	3.0	-7.0	30.0	21.0	19.0
72.0	10.0	61.0	57.0	57.0	37.0	30.0	10.0	11.0	15.0	35.0	45.0
37.0	52.0	91.0	88.0	102.0	52.0	39.0	6.0	-7.0	0.0	-8.0	-13.0
30.0	50.0	41.0	113.0	106.0	27.0	39.0	11.0	-9.0	-7.0	11.0	9.0
23.0	29.0	84.0	76.0	64.0	30.0	23.0	20.0	8.0	3.0	8.0	0.0
25.0	40.0	42.0	76.0	43.0	44.0	14.0	4.0	11.0	8.0	10.0	39.0
24.0	26.0	48.0	143.0	99.0	56.0	25.0	16.0	22.0	14.0	35.0	38.0
110.0	1.0	111.0	82.0	51.0	34.0	12.0	11.0	-6.0	5.0	23.0	10.0
15.0	15.0	70.0	111.0	59.0	28.0	12.0	17.0	4.0	-5.0	-2.0	-5.0
40.0	62.0	36.0	44.0	31.0	27.0	25.0	56.0	14.0	3.0	-5.0	16.0
59.0	26.0	50.0	128.0	94.0	104.0	19.0	-6.0	-9.0	-5.0	1.0	4.0
7.0	30.0	88.0	102.0	34.0	76.0	54.0	2.0	-1.0	32.0	7.0	15.0
-2.0	60.0	96.0	86.0	44.0	39.0	26.0	6.0	22.0	33.0	17.0	35.0
29.0	26.0	62.0	92.0	118.0	40.0	26.0	8.0	-3.0	6.0	5.0	3.0
0.0	14.0	85.0	58.0	23.0	26.0	36.0	8.0	3.0	13.0	15.0	54.0
23.0	31.0	96.0	66.0	45.0	19.0	9.0	-11.0	-10.0	-2.0	17.0	13.0
4.0	47.0	72.0	130.0	44.0	62.0	37.0	5.0	2.0	-7.0	-5.0	-7.0
24.0	18.0	73.0	75.0	56.0	54.0	9.0	10.0	6.0	1.0	14.0	35.0
40.0	23.0	52.0	91.0	88.0	36.0	30.0	17.0	16.0	-2.0	-11.0	-5.0
11.0	75.0	103.0	56.0	38.0	29.0	25.0	1.0	27.0	11.0	49.0	31.0
19.0	28.0	59.0	122.0	55.0	45.0	20.0	32.0	36.0	34.0	63.0	27.0
32.0	39.0	95.0	39.0	60.0	37.0	37.0	-4.0	-1.0	13.0	63.0	72.0
65.0	49.0	58.0	97.0	50.0	56.0	42.0	25.0	-7.0	20.0	30.0	41.0
52.0	27.0	98.0	132.0	91.0	33.0	29.0	-4.0	-2.0	5.0	14.0	9.0
69.0	59.0	92.0	67.0	58.0	43.0	22.0	-6.0	-3.0	-19.0	-2.0	-5.0
7.0	17.0	49.0	54.0	62.0	35.0	24.0	-9.0	7.0	-3.0	10.0	23.0
81.0	54.0	44.0	108.0	55.0	25.0	29.0	18.0	-12.0	8.0	34.0	23.0
32.0	16.0	49.0	105.0	49.0	20.0	3.0	6.0	-8.0	-21.0	2.0	15.0
44.0	10.0	66.0	94.0	29.0	34.0	7.0	-20.0	27.0	-20.0	18.0	9.0
44.0	31.0	64.0	44.0	52.0	53.0	30.0	-10.0	-1.0	-6.0	9.0	13.0
18.0	24.0	164.0	107.0	40.0	23.0	6.0	1.0	11.0	9.0	14.0	21.0
93.0	49.0	28.0	94.0	67.0	40.0	11.0	9.0	-18.0	14.0	24.0	15.0
28.0	82.0	82.0	68.0	36.0	20.0	25.0	5.0	21.0	-7.0	0.0	5.0
21.0	43.0	71.0	114.0	43.0	26.0	16.0	8.0	-16.0	-4.0	-15.0	2.0
8.0	19.0	26.0	141.0	74.0	47.0	35.0	-7.0	4.0	-10.0	21.0	50.0
75.0	23.0	35.0	76.0	22.0	16.0	22.0	-6.0	3.0	9.0	14.0	24.0
18.0	29.0	118.0	84.0	64.0	33.0	25.0	1.0	10.0	5.0	28.0	36.0
46.0	54.0	90.0	87.0	158.0	63.0	27.0	17.0	-2.0	13.0	23.0	-7.0
12.0	28.0	52.0	90.0	62.0	53.0	21.0	-4.0	6.0	-13.0	1.0	15.0
22.0	31.0	124.0	89.0	100.0	56.0	39.0	4.0	36.0	54.0	37.0	43.0
31.0	40.0	80.0	29.0	49.0	38.0	12.0	1.0	6.0	22.0	16.0	22.0
66.0	41.0	75.0	139.0	108.0	113.0	61.0	12.0	-4.0	-6.0	7.0	11.0
16.0	40.0	110.0	88.0	72.0	42.0	22.0	3.0	-13.0	2.0	25.0	4.0
53.0	59.0	53.0	75.0	34.0	20.0	7.0	-16.0	3.0	0.0	-2.0	38.0
80.0	42.0	68.0	114.0	34.0	37.0	24.0	16.0	-8.0	17.0	34.0	45.0
52.0	68.0	103.0	142.0	51.0	47.0	45.0	1.0	4.0	0.0	27.0	40.0
63.0	49.0	79.0	104.0	71.0	30.0	15.0	2.0	1.0	-20.0	10.0	27.0
33.0	29.0	81.0	53.0	94.0	30.0	17.0	4.0	-2.0	-13.0	0.0	18.0
16.0	78.0	79.0	115.0	63.0	44.0	3.0	5.0	10.0	33.0	34.0	25.0
35.0	26.0	106.0	101.0	37.0	18.0	6.0	16.0	-11.0	67.0	17.0	5.0
21.0	29.0	77.0	128.0	104.0	37.0	22.0	22.0	0.0	-1.0	-4.0	29.0
29.0	37.0	57.0	63.0	45.0	46.0	22.0	-15.0	6.0	-22.0	9.0	34.0
12.0	21.0	54.0	75.0	37.0	35.0	17.0	3.0	21.0	-2.0	14.0	12.0
31.0	43.0	73.0	130.0	56.0	26.0	25.0	-1.0	-12.0	12.0	20.0	65.0
37.0	61.0	34.0	157.0	87.0	51.0	11.0	1.0	-21.0	-11.0	-7.0	-15.0
-6.0	37.0	68.0	85.0	61.0	51.0	22.0	1.0	-9.0	-17.0	4.0	7.0
15.0	26.0	56.0	82.0	38.0	17.0	6.0	14.0	1.0	8.0	19.0	8.0
0.0	7.0	69.0	93.0	59.0	16.0	10.0	15.0	-14.0	-12.0	21.0	13.0
21.0	15.0	79.0	82.0	47.0	23.0	12.0	10.0	-20.0	-17.0	-4.0	15.0
18.0	52.0	49.0	82.0	39.0	23.0	10.0	12.0	11.0	8.0	42.0	50.0
27.0	45.0	91.0	48.0	32.0	29.0	-19.0	3.0	-4.0	11.0	30.0	42.0
34.0	16.0	38.0	80.0	52.0	37.0	20.0	4.0	11.0	31.0	57.0	44.0
31.0	22.0	72.0	55.0	53.0	46.0	19.0	11.0	18.0	6.0	54.0	44.0

ตารางที่ 7.3(b) Monthly Net Basin Supplies สำหรับทะเลสาบ Erie (1900-1968)

-9.0	61.0	66.0	48.0	25.0	12.0	5.0	-4.0	-39.0	-32.0	-8.0	-15.0
10.0	-27.0	60.0	51.0	36.0	24.0	-7.0	-21.0	-41.0	-41.0	-21.0	0.0
-4.0	-18.0	67.0	42.0	45.0	48.0	73.0	-35.0	11.0	-17.0	-16.0	8.0
15.0	61.0	108.0	90.0	12.0	23.0	11.0	-5.0	-28.0	-33.0	-22.0	-42.0
32.0	21.0	153.0	84.0	48.0	18.0	11.0	-18.0	-21.0	-34.0	-33.0	-24.0
30.0	-8.0	57.0	59.0	65.0	55.0	2.0	-18.0	-18.0	-48.0	-4.0	-2.0
37.0	8.0	32.0	54.0	17.0	23.0	1.0	-4.0	25.0	-5.0	-3.0	53.0
92.0	1.0	68.0	19.0	46.0	48.0	9.0	-31.0	4.0	-6.0	-18.0	34.0
42.0	62.0	106.0	54.0	59.0	0.0	-2.0	-22.0	-37.0	-36.0	-67.0	-7.0
15.0	91.0	52.0	62.0	96.0	42.0	-9.0	-15.0	-42.0	-46.0	9.0	-6.0
0.0	37.0	80.0	66.0	40.0	4.0	8.0	-12.0	-18.0	-10.0	-11.0	-16.0
34.0	29.0	33.0	71.0	21.0	12.0	-23.0	-3.0	2.0	-12.0	-16.0	51.0
26.0	13.0	81.0	105.0	42.0	20.0	1.0	9.0	-3.0	-14.0	-38.0	-7.0
123.0	36.0	182.0	87.0	23.0	11.0	2.0	-27.0	-34.0	-31.0	32.0	-8.0
7.0	-1.0	59.0	74.0	94.0	7.0	-1.0	-9.0	-28.0	-35.0	-26.0	-24.0
28.0	65.0	7.0	14.0	36.0	20.0	37.0	26.0	12.0	-30.0	-26.0	25.0
85.0	25.0	90.0	74.0	72.0	54.0	-3.0	-41.0	-38.0	-37.0	-20.0	2.0
32.0	1.0	68.0	82.0	69.0	75.0	38.0	-24.0	-16.0	6.0	7.0	-5.0
-30.0	45.0	76.0	-28.0	32.0	4.0	-6.0	-20.0	-21.0	-7.0	-37.0	48.0
-11.0	-3.0	92.0	53.0	90.0	6.0	-14.0	-14.0	-29.0	-7.0	-32.0	-39.0
10.0	-25.0	54.0	95.0	31.0	33.0	10.0	-7.0	-30.0	-25.0	8.0	10.0
25.0	33.0	89.0	88.0	23.0	14.0	-16.0	-18.0	-27.0	-20.0	20.0	8.0
-13.0	24.0	79.0	101.0	56.0	14.0	-2.0	-21.0	-17.0	-45.0	-42.0	4.0
19.0	-4.0	75.0	46.0	61.0	12.0	-4.0	-38.0	0.0	-51.0	-18.0	81.0
24.0	27.0	59.0	84.0	54.0	60.0	4.0	-25.0	-3.0	-28.0	-36.0	13.0
-11.0	55.0	76.0	30.0	7.0	14.0	-1.0	-6.0	6.0	-32.0	22.0	-24.0
24.0	15.0	55.0	109.0	16.0	36.0	6.0	15.0	65.0	22.0	27.0	1.0
12.0	45.0	96.0	39.0	69.0	23.0	20.0	-27.0	-18.0	-25.0	52.0	58.0
28.0	35.0	36.0	54.0	22.0	74.0	22.0	-17.0	-45.0	-18.0	8.0	8.0
37.0	32.0	92.0	121.0	49.0	3.0	1.0	-41.0	-27.0	-24.0	2.0	34.0
121.0	50.0	62.0	47.0	6.0	23.0	-21.0	-30.0	-29.0	-36.0	-28.0	-1.0
13.0	19.0	44.0	49.0	32.0	26.0	6.0	-21.0	-11.0	-27.0	-5.0	36.0
93.0	37.0	26.0	49.0	57.0	12.0	9.0	-12.0	-35.0	-19.0	6.0	27.0
40.0	10.0	76.0	72.0	71.0	-1.0	-15.0	-23.0	-20.0	-34.0	-26.0	18.0
5.0	-8.0	28.0	64.0	10.0	9.0	-20.0	-16.0	-1.0	-55.0	-12.0	-10.0
15.0	-15.0	61.0	17.0	42.0	22.0	7.0	-19.0	-37.0	-21.0	-19.0	12.0
-45.0	11.0	110.0	48.0	17.0	1.0	-20.0	-28.0	-14.0	-20.0	-30.0	6.0
117.0	61.0	12.0	124.0	35.0	87.0	27.0	6.0	-46.0	-21.0	-24.0	20.0
10.0	101.0	78.0	56.0	35.0	26.0	16.0	-15.0	-15.0	-37.0	-21.0	-15.0
24.0	36.0	71.0	86.0	16.0	28.0	4.0	-24.0	-37.0	-29.0	-29.0	-22.0
-7.0	27.0	48.0	104.0	58.0	46.0	6.0	7.0	-18.0	-24.0	-20.0	54.0
18.0	9.0	30.0	44.0	23.0	19.0	-2.0	-25.0	-42.0	-16.0	-9.0	2.0
-7.0	57.0	104.0	70.0	55.0	24.0	17.0	-14.0	-21.0	-3.0	18.0	42.0
10.0	47.0	59.0	67.0	136.0	40.0	29.0	-24.0	-30.0	-22.0	-16.0	-21.0
-9.0	30.0	58.0	113.0	58.0	25.0	-29.0	-19.0	-9.0	-42.0	-15.0	-7.0
-11.0	37.0	116.0	60.0	80.0	64.0	7.0	-27.0	27.0	9.0	-3.0	-6.0
14.0	5.0	71.0	3.0	54.0	67.0	-6.0	-38.0	-26.0	-13.0	-19.0	2.0
43.0	-5.0	59.0	142.0	98.0	71.0	0.0	8.0	-41.0	-8.0	-40.0	30.0
-18.0	41.0	107.0	64.0	67.0	32.0	-9.0	-14.0	-38.0	-23.0	7.0	4.0
56.0	63.0	50.0	41.0	41.0	20.0	-2.0	-28.0	-26.0	-18.0	-12.0	33.0
137.0	67.0	105.0	93.0	26.0	26.0	11.0	-14.0	-15.0	-3.0	24.0	43.0
62.0	82.0	88.0	74.0	45.0	27.0	-7.0	-30.0	-38.0	-23.0	9.0	24.0
97.0	46.0	68.0	63.0	39.0	2.0	-21.0	-16.0	-36.0	-73.0	-8.0	17.0
84.0	14.0	55.0	34.0	61.0	15.0	-13.0	-19.0	-49.0	-41.0	-11.0	-15.0
17.0	62.0	76.0	106.0	8.0	10.0	-19.0	-16.0	-29.0	60.0	-1.0	19.0
28.0	45.0	90.0	63.0	12.0	0.0	-10.0	-6.0	-35.0	-5.0	-3.0	5.0
-36.0	47.0	103.0	79.0	105.0	28.0	15.0	24.0	-42.0	-23.0	-40.0	30.0
22.0	51.0	35.0	127.0	46.0	51.0	26.0	-27.0	-12.0	-27.0	-6.0	62.0
8.0	12.0	36.0	50.0	22.0	40.0	33.0	2.0	-1.0	-32.0	-10.0	-9.0
70.0	86.0	78.0	80.0	57.0	7.0	-3.0	-5.0	-38.0	-1.0	-1.0	57.0
42.0	44.0	25.0	70.0	53.0	46.0	-3.0	-11.0	-45.0	-61.0	-24.0	-49.0
-16.0	40.0	66.0	125.0	34.0	28.0	9.0	-4.0	-28.0	-50.0	-20.0	-21.0
-19.0	25.0	64.0	21.0	10.0	15.0	-19.0	-16.0	-32.0	-27.0	-2.0	-7.0
-19.0	-26.0	108.0	47.0	15.0	-1.0	-10.0	-21.0	-39.0	-35.0	-25.0	-20.0
19.0	9.0	75.0	82.0	23.0	15.0	-8.0	0.0	-41.0	-45.0	-25.0	10.0
31.0	49.0	67.0	55.0	30.0	7.0	-12.0	-13.0	-12.0	-24.0	-10.0	32.0
-14.0	35.0	50.0	42.0	27.0	24.0	-5.0	-5.0	-41.0	-48.0	33.0	65.0
4.0	11.0	54.0	62.0	47.0	27.0	-4.0	-22.0	-26.0	-8.0	8.0	46.0
38.0	29.0	56.0	32.0	64.0	30.0	4.0	-36.0	-32.0	-37.0	23.0	-5.0

ตารางที่ 7.3(c) Monthly Net Basin Supplies สำหรับทะเลสาบ Superior (1900-1968)

-22.0	44.0	-11.0	124.0	127.0	78.0	200.0	221.0	261.0	77.0	-7.0	-51.0
-39.0	-9.0	52.0	121.0	126.0	197.0	187.0	86.0	13.0	98.0	-7.0	-30.0
-35.0	11.0	40.0	129.0	167.0	171.0	108.0	75.0	77.0	58.0	55.0	-15.0
-20.0	-57.0	94.0	146.0	290.0	94.0	164.0	97.0	113.0	47.0	12.0	-79.0
-13.0	34.0	34.0	70.0	245.0	149.0	84.0	154.0	121.0	124.0	-31.0	-37.0
-58.0	37.0	101.0	108.0	197.0	183.0	155.0	137.0	140.0	25.0	54.0	-19.0
-20.0	26.0	-10.0	160.0	155.0	204.0	88.0	89.0	71.0	35.0	33.0	-22.0
7.0	20.0	67.0	54.0	241.0	163.0	117.0	189.0	142.0	20.0	-17.0	-87.0
-26.0	19.0	14.0	113.0	287.0	196.0	136.0	54.0	51.0	-3.0	-55.0	-10.0
-34.0	37.0	3.0	104.0	242.0	80.0	204.0	93.0	59.0	25.0	86.0	7.0
-17.0	-21.0	24.0	123.0	109.0	80.0	83.0	104.0	35.0	0.0	-2.0	-77.0
-12.0	-18.0	-3.0	65.0	221.0	174.0	210.0	130.0	96.0	11.0	-8.0	7.0
-28.0	20.0	36.0	189.0	161.0	165.0	98.0	136.0	52.0	54.0	-49.0	-2.0
-73.0	-7.0	125.0	145.0	203.0	120.0	174.0	108.0	127.0	63.0	691.0	-43.0
-31.0	1.0	-3.0	165.0	188.0	140.0	126.0	93.0	80.0	2.0	18.0	-100.0
50.0	-1.0	-35.0	131.0	154.0	240.0	125.0	76.0	177.0	92.0	93.0	22.0
20.0	2.0	36.0	277.0	268.0	216.0	115.0	104.0	149.0	51.0	-21.0	23.0
-63.0	19.0	84.0	91.0	158.0	168.0	100.0	124.0	62.0	21.0	-20.0	-43.0
-2.0	20.0	-13.0	115.0	218.0	163.0	113.0	103.0	40.0	103.0	63.0	24.0
-20.0	15.0	5.0	154.0	139.0	104.0	77.0	42.0	59.0	-10.0	73.0	-42.0
12.0	25.0	150.0	132.0	153.0	193.0	129.0	72.0	-8.0	51.0	-43.0	28.0
-66.0	-7.0	48.0	198.0	176.0	82.0	115.0	61.0	37.0	-22.0	-57.0	-74.0
-40.0	-5.0	49.0	171.0	189.0	137.0	126.0	67.0	32.0	-25.0	-44.0	-27.0
-28.0	-37.0	33.0	100.0	107.0	103.0	129.0	72.0	58.0	46.0	-2.0	-28.0
-34.0	-28.0	-3.0	127.0	76.0	94.0	120.0	148.0	76.0	29.0	-44.0	-69.0
-20.0	-6.0	45.0	110.0	110.0	143.0	111.0	66.0	81.0	-31.0	-41.0	-37.0
-19.0	-20.0	53.0	54.0	144.0	172.0	171.0	109.0	160.0	71.0	62.0	18.0
-12.0	38.0	108.0	169.0	272.0	169.0	184.0	57.0	72.0	50.0	11.0	-3.0
-2.0	4.0	56.0	168.0	194.0	216.0	167.0	154.0	138.0	148.0	-2.0	-35.0
5.0	38.0	88.0	124.0	125.0	113.0	152.0	38.0	88.0	36.0	-2.0	-25.0
-16.0	31.0	-3.0	97.0	176.0	209.0	144.0	28.0	54.0	15.0	5.0	-43.0
-25.0	-44.0	-15.0	81.0	139.0	134.0	117.0	34.0	89.0	69.0	70.0	-4.0
-8.0	25.0	-3.0	118.0	202.0	89.0	171.0	121.0	-22.0	4.0	5.0	-15.0
-18.0	27.0	-4.0	163.0	238.0	117.0	107.0	44.0	85.0	42.0	8.0	-24.0
20.0	6.0	47.0	117.0	203.0	121.0	108.0	64.0	170.0	55.0	102.0	-21.0
19.0	-15.0	97.0	153.0	140.0	194.0	175.0	71.0	42.0	84.0	10.0	-6.0
4.0	35.0	85.0	129.0	289.0	117.0	48.0	100.0	40.0	-14.0	-18.0	5.0
10.0	85.0	-4.0	190.0	221.0	89.0	173.0	113.0	32.0	46.0	38.0	-27.0
2.0	18.0	84.0	253.0	185.0	222.0	99.0	114.0	62.0	27.0	43.0	-7.0
28.0	26.0	56.0	145.0	250.0	246.0	133.0	119.0	45.0	0.0	-25.0	-56.0
-10.0	-20.0	-4.0	83.0	265.0	212.0	117.0	34.0	26.0	-3.0	34.0	-7.0
-36.0	-9.0	-14.0	285.0	150.0	159.0	103.0	89.0	198.0	106.0	2.0	-5.0
-16.0	-8.0	69.0	139.0	234.0	99.0	124.0	101.0	58.0	85.0	51.0	-34.0
-6.0	19.0	50.0	124.0	261.0	296.0	116.0	109.0	22.0	21.0	-8.0	-71.0
-31.0	-19.0	28.0	114.0	246.0	252.0	188.0	141.0	97.0	-34.0	52.0	-62.0
-2.0	52.0	149.0	185.0	134.0	130.0	118.0	118.0	70.0	0.0	25.0	-27.0
14.0	21.0	110.0	98.0	134.0	159.0	100.0	60.0	110.0	84.0	22.0	-26.0
-35.0	-7.0	12.0	199.0	217.0	293.0	86.0	103.0	66.0	12.0	-17.0	-44.0
-30.0	-19.0	41.0	272.0	89.0	93.0	103.0	120.0	0.0	-29.0	62.0	-25.0
3.0	-7.0	23.0	115.0	185.0	178.0	178.0	48.0	28.0	82.0	-3.0	-19.0
11.0	-6.0	59.0	150.0	354.0	224.0	189.0	121.0	75.0	90.0	67.0	-29.0
-25.0	83.0	80.0	237.0	195.0	194.0	120.0	152.0	143.0	79.0	265.0	9.0
7.0	-10.0	30.0	195.0	106.0	201.0	234.0	119.0	21.0	-90.0	5.0	-23.0
-3.0	21.0	60.0	159.0	250.0	234.0	158.0	129.0	23.0	-14.0	-8.0	-5.0
-11.0	30.0	3.0	226.0	267.0	203.0	85.0	46.0	53.0	16.0	16.0	-64.0
-38.0	10.0	47.0	201.0	140.0	98.0	124.0	104.0	30.0	66.0	52.0	-43.0
-1.0	-28.0	-15.0	114.0	212.0	124.0	162.0	78.0	36.0	1.0	-9.0	0.0
-65.0	22.0	48.0	186.0	127.0	175.0	112.0	51.0	59.0	-18.0	57.0	-30.0
-4.0	-8.0	11.0	102.0	90.0	158.0	153.0	101.0	88.0	8.0	40.0	-42.0
-24.0	1.0	34.0	110.0	244.0	138.0	92.0	202.0	136.0	62.0	-33.0	-17.0
2.0	-24.0	5.0	234.0	265.0	128.0	110.0	83.0	35.0	1.0	53.0	-71.0
-25.0	40.0	66.0	120.0	166.0	106.0	88.0	32.0	48.0	51.0	-4.0	-7.0
-51.0	44.0	30.0	94.0	217.0	97.0	85.0	110.0	72.0	-16.0	-41.0	-37.0
1.0	0.0	71.0	142.0	128.0	198.0	80.0	86.0	46.0	4.0	8.0	-66.0
15.0	-30.0	8.0	203.0	265.0	177.0	90.0	135.0	113.0	17.0	35.0	-12.0
-18.0	19.0	48.0	153.0	253.0	131.0	119.0	118.0	133.0	65.0	87.0	17.0
-4.0	14.0	100.0	137.0	193.0	112.0	108.0	140.0	3.0	76.0	9.0	4.0
14.0	18.0	48.0	237.0	105.0	185.0	101.0	106.0	-11.0	68.0	-5.0	-22.0
-19.0	-37.0	79.0	226.0	156.0	251.0	264.0	151.0	165.0	89.0	-32.0	41.0

ตารางที่ 7.3(d) Monthly Net Basin Supplies สำหรับทะเลสาบ Michigan-Huron (1900-1968)

-1.0	162.0	60.0	246.0	175.0	156.0	268.0	145.0	78.0	73.0	56.0	-61.0
10.0	21.0	284.0	230.0	223.0	175.0	208.0	79.0	-64.0	35.0	-74.0	48.0
-12.0	52.0	212.0	207.0	257.0	272.0	276.0	8.0	45.0	-52.0	78.0	-45.0
59.0	144.0	273.0	215.0	245.0	155.0	147.0	107.0	148.0	12.0	-51.0	4.0
44.0	115.0	304.0	334.0	370.0	207.0	134.0	76.0	64.0	24.0	-68.0	-40.0
1.0	79.0	254.0	190.0	337.0	286.0	101.0	82.0	27.0	-60.0	36.0	13.0
169.0	122.0	149.0	255.0	215.0	214.0	103.0	5.0	32.0	3.0	80.0	41.0
141.0	71.0	184.0	224.0	205.0	225.0	148.0	62.0	58.0	-44.0	-7.0	40.0
3.0	125.0	210.0	288.0	389.0	187.0	192.0	-43.0	-45.0	-71.0	-65.0	-12.0
57.0	106.0	126.0	334.0	343.0	207.0	126.0	36.0	20.0	-109.0	59.0	92.0
69.0	55.0	162.0	298.0	194.0	138.0	76.0	76.0	56.0	24.0	13.0	-60.0
91.0	82.0	114.0	266.0	293.0	187.0	61.0	47.0	78.0	83.0	92.0	95.0
47.0	93.0	90.0	299.0	484.0	170.0	170.0	164.0	125.0	39.0	98.0	44.0
66.0	59.0	272.0	395.0	306.0	172.0	143.0	43.0	-15.0	21.0	88.0	-55.0
57.0	64.0	146.0	199.0	240.0	251.0	124.0	66.0	15.0	-9.0	-56.0	-49.0
91.0	128.0	64.0	140.0	195.0	211.0	154.0	100.0	134.0	-73.0	80.0	7.0
135.0	90.0	194.0	416.0	354.0	311.0	117.0	-25.0	10.0	42.0	25.0	114.0
45.0	43.0	197.0	306.0	236.0	395.0	240.0	35.0	3.0	-32.0	28.0	-9.0
96.0	187.0	264.0	235.0	374.0	146.0	129.0	50.0	-30.0	63.0	76.0	141.0
25.0	50.0	260.0	302.0	309.0	106.0	104.0	-7.0	24.0	80.0	40.0	43.0
-10.0	67.0	305.0	307.0	157.0	223.0	141.0	75.0	65.0	-5.0	-20.0	81.0
74.0	34.0	258.0	383.0	147.0	135.0	49.0	73.0	55.0	-6.0	6.0	127.0
-7.0	164.0	223.0	461.0	270.0	217.0	214.0	4.0	46.0	-95.0	25.0	-34.0
5.0	40.0	191.0	308.0	306.0	208.0	108.0	59.0	83.0	10.0	-45.0	42.0
-7.0	106.0	196.0	285.0	291.0	202.0	172.0	190.0	22.0	-80.0	-21.0	-68.0
-1.0	93.0	183.0	150.0	71.0	195.0	121.0	-21.0	-7.0	-34.0	17.0	20.0
34.0	80.0	141.0	292.0	250.0	277.0	133.0	100.0	80.0	60.0	161.0	51.0
54.0	107.0	223.0	225.0	333.0	179.0	143.0	-38.0	52.0	3.0	91.0	50.0
51.0	99.0	185.0	407.0	267.0	269.0	188.0	131.0	48.0	198.0	133.0	85.0
124.0	47.0	317.0	474.0	401.0	225.0	124.0	29.0	-35.0	1.0	-8.0	-37.0
74.0	182.0	143.0	200.0	240.0	242.0	143.0	-24.0	-62.0	-71.0	-50.0	-24.0
-29.0	-9.0	121.0	132.0	192.0	156.0	62.0	-56.0	163.0	-25.0	144.0	-50.0
208.0	73.0	59.0	180.0	236.0	116.0	109.0	46.0	-58.0	24.0	-16.0	94.0
54.0	71.0	79.0	361.0	360.0	158.0	78.0	-57.0	-20.0	-22.0	-19.0	42.0
77.0	-28.0	126.0	276.0	150.0	166.0	49.0	-33.0	143.0	-55.0	131.0	54.0
29.0	114.0	192.0	174.0	137.0	242.0	110.0	8.0	0.0	-50.0	86.0	-39.0
86.0	94.0	178.0	195.0	258.0	112.0	36.0	71.0	80.0	23.0	-80.0	67.0
87.0	116.0	40.0	325.0	217.0	173.0	80.0	47.0	-1.0	-6.0	33.0	-32.0
119.0	230.0	310.0	260.0	244.0	216.0	111.0	81.0	29.0	-65.0	-22.0	15.0
36.0	124.0	122.0	247.0	235.0	257.0	80.0	91.0	-32.0	-51.0	-38.0	-22.0
-22.0	28.0	44.0	195.0	291.0	264.0	124.0	127.0	54.0	-36.0	72.0	92.0
71.0	46.0	44.0	293.0	164.0	120.0	89.0	-13.0	128.0	181.0	128.0	51.0
53.0	52.0	274.0	223.0	323.0	236.0	106.0	-12.0	55.0	28.0	54.0	53.0
68.0	149.0	241.0	276.0	363.0	398.0	170.0	84.0	-34.0	-36.0	59.0	-67.0
41.0	65.0	152.0	189.0	223.0	254.0	92.0	-9.0	77.0	-58.0	14.0	-34.0
6.0	67.0	221.0	230.0	304.0	288.0	135.0	41.0	82.0	29.0	70.0	23.0
113.0	108.0	279.0	146.0	209.0	202.0	59.0	-15.0	-5.0	-30.0	-50.0	15.0
45.0	47.0	107.0	447.0	395.0	279.0	152.0	53.0	32.0	19.0	-20.0	-89.0
-1.0	89.0	298.0	291.0	229.0	158.0	91.0	-8.0	-81.0	-136.0	134.0	-15.0
97.0	88.0	94.0	228.0	166.0	239.0	120.0	-49.0	-59.0	-41.0	-49.0	22.0
160.0	99.0	198.0	343.0	215.0	229.0	163.0	60.0	26.0	-17.0	4.0	45.0
105.0	141.0	228.0	490.0	200.0	183.0	223.0	102.0	18.0	151.0	110.0	90.0
179.0	83.0	192.0	386.0	208.0	198.0	233.0	100.0	-61.0	-193.0	63.0	60.0
56.0	107.0	228.0	261.0	230.0	246.0	112.0	67.0	-74.0	-37.0	-57.0	-17.0
-17.0	130.0	162.0	344.0	228.0	317.0	113.0	40.0	73.0	256.0	53.0	23.0
45.0	64.0	143.0	313.0	176.0	135.0	49.0	-19.0	-127.0	47.0	-13.0	-30.0
25.0	61.0	144.0	262.0	3178.0	179.0	170.0	132.0	-58.0	-34.0	-13.0	21.0
-13.0	64.0	109.0	239.0	251.0	238.0	180.0	-11.0	15.0	2.0	69.0	92.0
49.0	21.0	74.0	159.0	65.0	152.0	128.0	27.0	34.0	-36.0	-49.0	-43.0
20.0	84.0	181.0	391.0	261.0	111.0	92.0	170.0	30.0	78.0	92.0	114.0
102.0	112.0	85.0	423.0	495.0	264.0	197.0	100.0	1.0	-58.0	79.0	-65.0
-17.0	88.0	178.0	224.0	153.0	197.0	126.0	62.0	139.0	26.0	24.0	-2.0
96.0	105.0	174.0	237.0	234.0	150.0	74.0	64.0	-2.0	-20.0	-60.0	-45.0
0.0	77.0	221.0	221.0	220.0	118.0	118.0	64.0	0.0	-29.0	-32.0	-70.0
15.0	-2.0	100.0	240.0	211.0	111.0	118.0	32.0	32.0	-78.0	13.0	5.0
58.0	106.0	152.0	362.0	265.0	124.0	66.0	92.0	204.0	46.0	90.0	138.0
44.0	102.0	264.0	209.0	151.0	138.0	40.0	14.0	-82.0	-52.0	152.0	139.0
107.0	68.0	163.0	437.0	188.0	328.0	83.0	45.0	-15.0	57.0	84.0	152.0
67.0	167.0	123.0	275.0	77.0	315.0	100.0	129.0	16.0	29.0	30.0	129.0

(1) การวิเคราะห์เบื้องต้นและการระบุรูปแบบของแบบจำลอง (Preliminary Analysis and Model Identification)

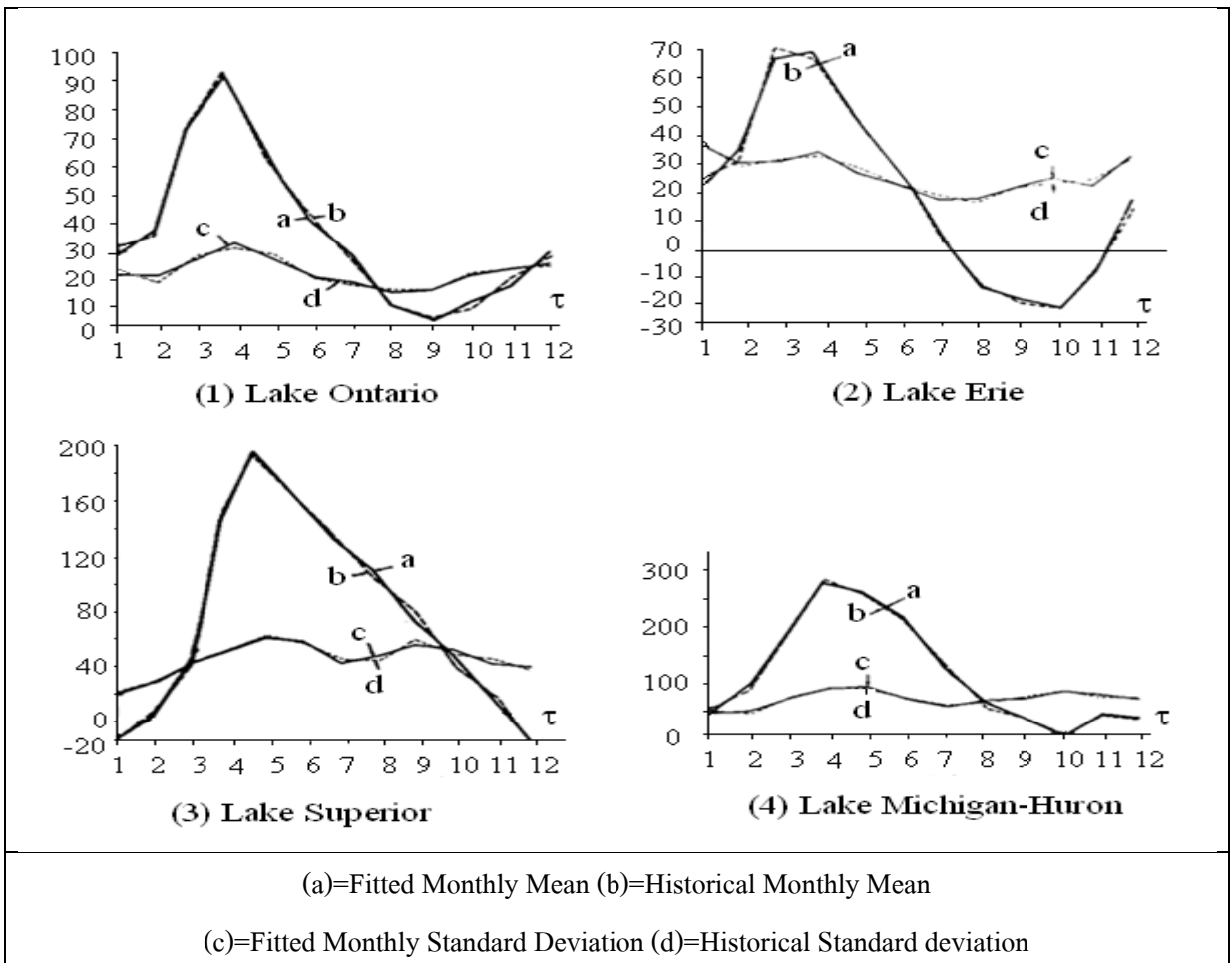
Identification)

ขั้นที่ 1a และ 1b ตรวจสอบ Normality โดยใช้ Smirnov-Kolmogorov Statistics จากตารางที่ 3.2 (N=69)

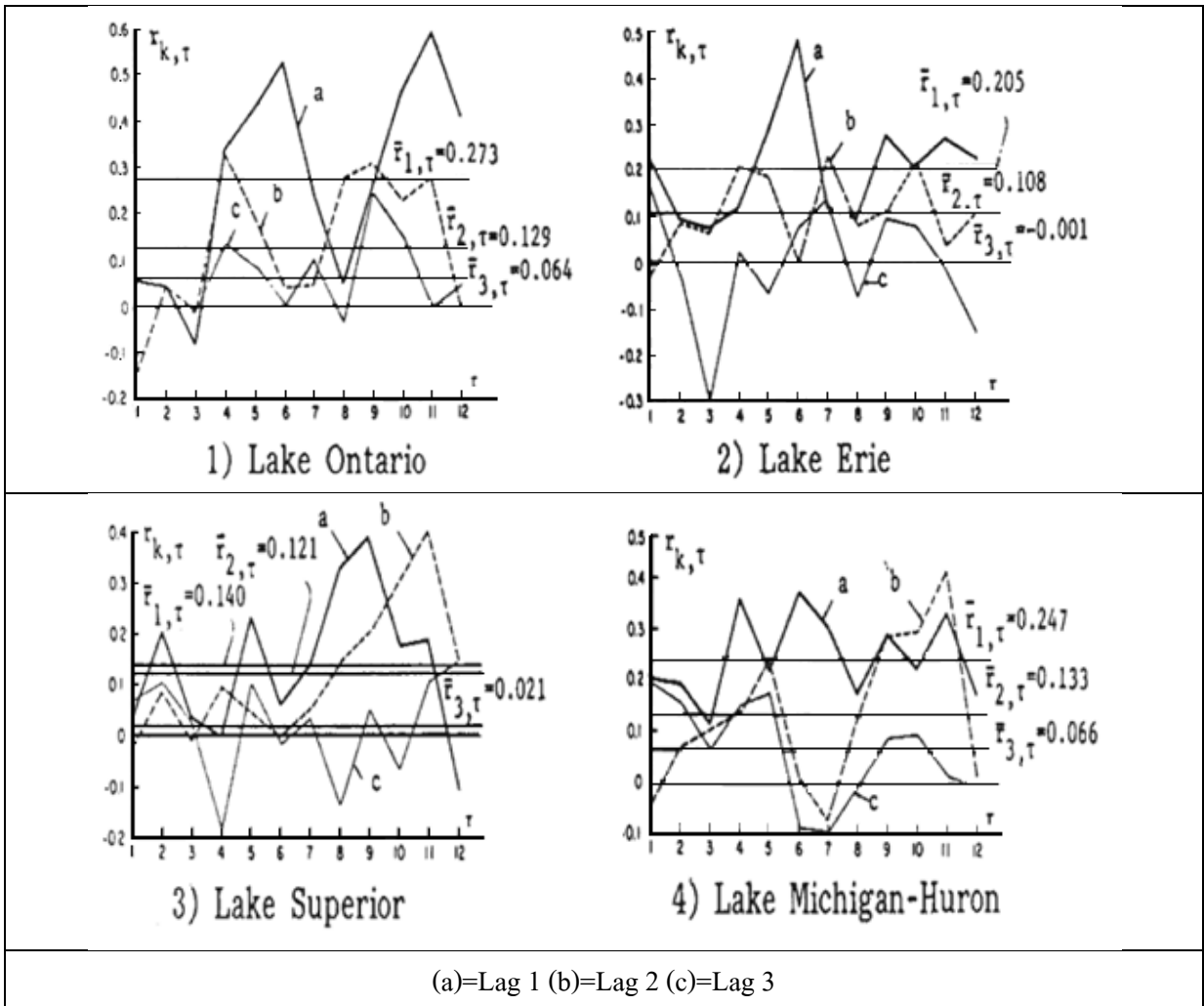
i	Lake	Computed γ_i	Critical $\gamma_{69}(0.10)$	Critical $\gamma_{69}(0.02)$	Hypothesis of Normality
1	Ontario	1.005	0.462	0.679	Rejected at $\alpha = 0.10$
2	Erie	0.648	0.462	0.679	Rejected at $\alpha = 0.10$
3	Superior	0.489	0.462	0.679	Rejected at $\alpha = 0.10$
4	Michigan-Huron	0.597	0.462	0.679	Rejected at $\alpha = 0.10$

ขั้นที่ 1c และ 1d จากตารางข้างบนค่าสามารถสรุปได้ว่าอนุกรมเวลามีค่าสัมประสิทธิ์ความเบ้เป็นบวกที่ $\alpha = 0.10$ จึงควรมีการแปลงอนุกรมเวลาให้เป็นปกติ แต่มีปัญหาคือค่า NBS บางค่ามีค่าติดลบทำให้ไม่สามารถแปลงค่าได้โดยใช้วิธีการแบบง่าย ๆ ดังนั้นในตัวอย่างนี้จึงวิเคราะห์ต่อโดยไม่แปลงอนุกรมเวลาให้เป็นปกติ

ขั้นที่ 1e คำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานรายเดือนของอนุกรมเวลา NBS ของทะเลสาบทั้ง 4 ได้ดังแสดงในรูปที่ 7.5 และ คำนวณหา $r_{k,\tau}^{(i)}$ เมื่อ $\tau=1,2,3$ ได้ดังแสดงในรูปที่ 7.6 ค่า $r_{k,\tau}^{(i)}$ ไม่ได้แสดงรูปแบบใดๆ บางเดือนมีค่าต่างจากศูนย์ บางเดือนไม่ต่างจากศูนย์ จึงพิจารณาว่า $r_{k,\tau}^{(i)}$ เป็นพารามิเตอร์ที่มีค่าคงที่



รูปที่ 7.5 Fitted and Historical Periodic Mean and Standard Deviation of Monthly NBS of 4 Lakes



รูปที่ 7.6 Monthly Correlation Coefficients ที่ Lag 1, 2, 3 ของทะเลสาบทั้ง 4

ขั้นที่ 1f ถึงแม้ว่าไม่จำเป็นต้องใช้ Fourier Series เพื่อหาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐาน เนื่องจาก $w=12$ แต่การหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานโดยใช้ Fourier Series ทำให้ลดจำนวนพารามิเตอร์ของแบบจำลอง

ขั้นที่ 2 การประมาณค่าพารามิเตอร์

ขั้นที่ 2a คำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานรายเดือนโดยวิธี Fourier Series ค่า Fourier Series Coefficients ของ 4 Harmonics แสดงอยู่ในตารางที่ 7.4 และค่าค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานรายเดือนที่คำนวณได้แสดงอยู่ในตารางที่ 7.5 ผลการฟิตด้วย Fourier Series ในรูปที่ 7.5 จะเห็นได้ว่าฟิตได้ดี และทำให้จำนวนพารามิเตอร์ของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานรายเดือนลดลงจาก 12 เป็น 8 ($4 A_j$ และ $4 B_j$)

ตารางที่ 7.4 Harmonics และ Corresponding Fourier Coefficients สำหรับการคำนวณ Fitted Monthly Means และ Fitted Monthly Standard Deviation ของ NBS ของ 4 ทะเลสาบ

Lake	Mean μ_τ			Standard Deviation σ_τ		
	Harmonics	A_j	B_j	Harmonics	A_j	B_j
Ontario	1	-13.162	32.586	1	-0.166	6.608
	2	-3.048	-10.522	2	0.024	-2.987
	3	6.398	-3.169	4	-0.312	1.238
	4	0.398	5.079	3	1.213	0.160
Erie	1	-11.602	41.503	1	3.732	7.761
	2	4.780	-2.611	3	0.721	2.681
	3	-1.371	-5.022	4	0.023	2.329
	4	4.250	2.751	5	-0.829	0.259
Superior	1	-94.327	-0.243	1	-11.270	-2.557
	2	5.407	-23.493	2	-2.006	-9.090
	3	2.930	10.799	4	2.404	-0.912
	4	-8.776	-1.776	3	-0.343	2.487
Michigan	1	-93.848	84.036	2	-2.364	-14.685
	2	8.191	-34.842	1	-6.062	3.574
	5	-8.567	-9.185	3	3.638	1.990
	3	9.792	-0.353	5	0.656	2.590

ตารางที่ 7.5 Fitted Monthly Means และ Fitted Monthly Standard Deviation ของ NBS ของ 4 ทะเลสาบ

Month	Lake Ontario		Lake Erie		Lake Superior		Lake Michigan Huron	
	$\hat{\mu}_\tau^{(1)}$	$\hat{\sigma}_\tau^{(1)}$	$\hat{\mu}_\tau^{(2)}$	$\hat{\sigma}_\tau^{(2)}$	$\hat{\mu}_\tau^{(3)}$	$\hat{\sigma}_\tau^{(3)}$	$\hat{\mu}_\tau^{(4)}$	$\hat{\sigma}_\tau^{(4)}$
1	29.248	21.813	20.580	37.200	-13.872	22.879	46.926	50.963
2	37.015	20.752	34.257	29.751	4.505	27.510	95.250	48.583
3	73.162	25.951	66.532	29.915	46.706	41.663	176.800	72.146
4	89.996	30.662	68.344	32.325	142.308	52.258	277.598	86.701
5	61.474	25.132	43.675	25.266	193.276	59.724	254.984	88.034
6	38.074	18.504	25.958	20.951	159.960	54.309	210.601	65.013
7	25.798	15.173	3.459	15.917	128.152	39.981	120.492	52.454
8	6.534	11.899	-17.006	15.297	105.113	42.523	55.786	59.561
9	1.652	13.055	-21.697	19.236	68.791	51.753	26.394	63.798
10	7.596	16.623	-24.165	20.791	42.542	46.104	-5.866	72.314
11	12.427	17.915	-11.767	19.782	8.540	37.787	33.129	66.254
12	24.546	20.599	13.394	28.200	-22.834	31.083	25.355	61.476

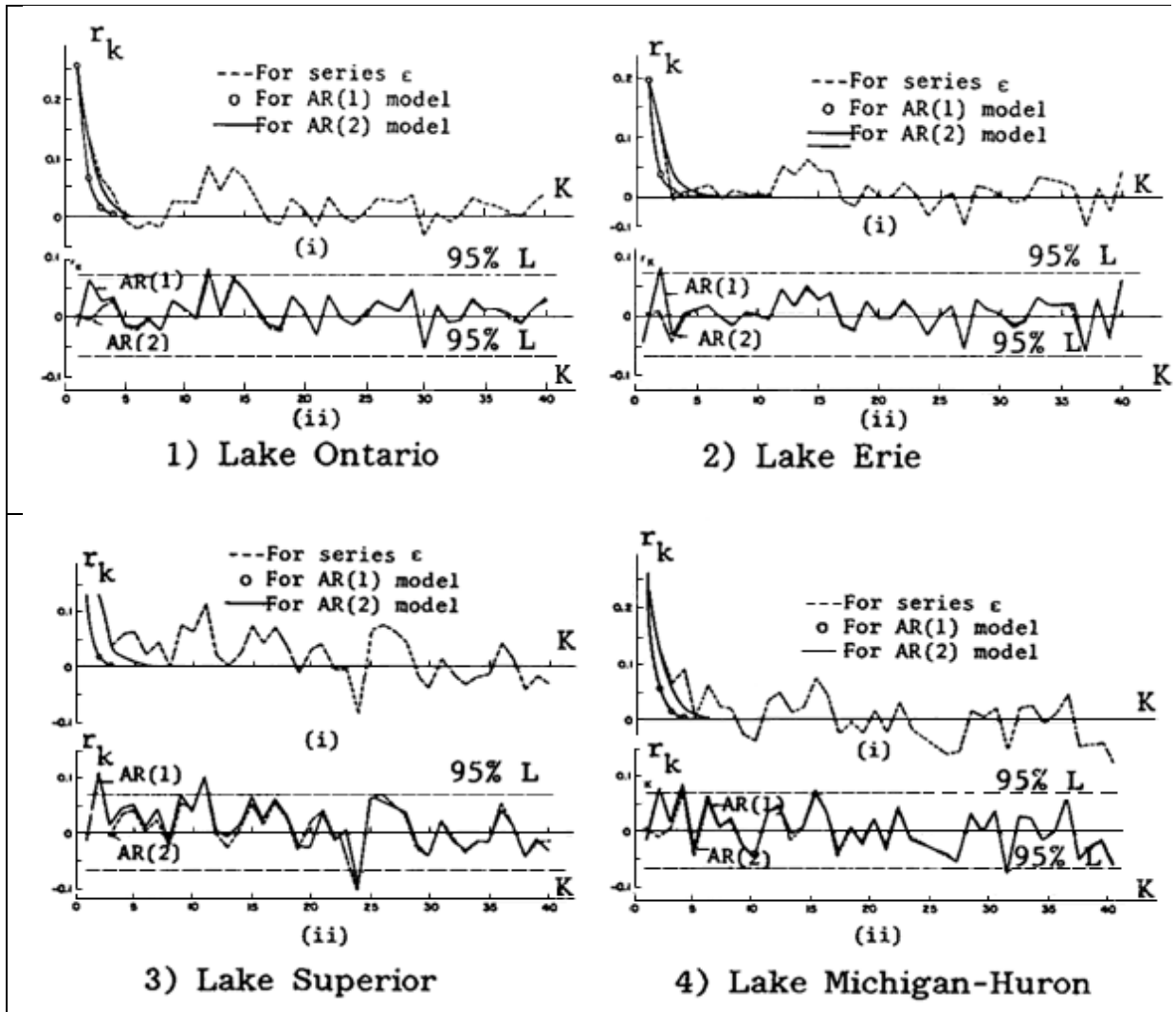
ขั้นที่ 2b ใช้ค่าประมาณแบบ Fourier Series ในตารางที่ 7.5 ในการแปลง $y_{v,\tau}^{(i)}$ เป็น $z_{v,\tau}^{(i)}$ โดยสมการ

$$z_{v,\tau}^{(i)} = \frac{y_{v,\tau}^{(i)} - \hat{\mu}_\tau^{(i)}}{\hat{\sigma}_\tau^{(i)}} \text{ เมื่อ } i=1,2,3,4$$

ตัวอย่างค่า $z_{v,\tau}^{(i)}$ ของทะเลสาบ Ontario ที่คำนวณได้

$z_{1,1}^{(1)} =$	0.011
$z_{1,2}^{(1)} =$	1.011
.	.
$z_{69,12}^{(1)} =$	-0.658

ขั้นที่ 2c แปลงค่า $z_{v,\tau}^{(i)}$ เป็น $z_t^{(i)}$ เมื่อ $t = (v - 1)12 + \tau$ แล้วคำนวณหา $r_k^{(i)}(z)$ เมื่อ $k=1,2,\dots,40$ และ $i=1,2,3,4$ โดยใช้สมการ [2.5b] ซึ่งเมื่อนำ $r_k^{(i)}(z)$ ไปพล็อตกราฟได้กราฟดังรูปที่ 7.7



รูปที่ 7.7 Historical Correlogram ของ $z_t^{(i)}$, AR(1) และ AR(2) Model Correlogram และ Residuals Correlogram ของ AR(1) และ AR(2)

Correlogram ของทะเลสาบ Ontario และ Erie แสดง Short Term Dependence ขณะที่ Correlogram ของทะเลสาบ Superior และ Michigan-Huron แสดง Longer Term Dependence จึงเสนอว่าควรใช้แบบจำลอง AR(1) หรือ AR(2) ในการจำลองอนุกรมเวลาของทะเลสาบ Ontario และ Erie และใช้แบบจำลอง ARMA ในการจำลองอนุกรมเวลาของทะเลสาบ Superior และ Michigan-Huron หรืออาจเริ่มจากการใช้แบบจำลอง สำหรับ Short Term Dependence เช่น AR(1) และ AR(2) กับทั้ง 4 ทะเลสาบก่อน แล้วดูว่าผลการทดสอบความเหมาะสมว่าใช้ได้หรือไม่ ถ้าได้ค่อยเปลี่ยนแบบจำลอง

ขั้นที่ 2d ทดลองใช้ AR(1) กับอนุกรมเวลาของแต่ละทะเลสาบ ได้ค่า Autoregressive Parameter ดังนี้

$$\hat{\phi}_1^{(1)} = 0.256$$

$$\hat{\phi}_1^{(2)} = 0.196$$

$$\hat{\phi}_1^{(3)} = 0.134$$

$$\hat{\phi}_1^{(4)} = 0.299$$

คำนวณ AR(1) Model Correlogram และ AR(1) Residuals Correlogram จาก $\hat{\varepsilon}_t^{(i)}$ เมื่อ $t=2$ ถึง 828 แล้วพล็อตกราฟดังรูปที่ 7.5 ซึ่งสรุปได้ว่า AR(1) Model Correlogram ต่างจาก Historical Correlogram และ AR(1) Residual Correlogram มีหลายค่าอยู่นอก Probability Limits จึงสรุปได้ว่า AR(1) ไม่เหมาะสม (Inadequate) จึงควรทดลองจำลองโดยใช้ AR(2)

ค่าพารามิเตอร์ของแบบจำลอง AR(2) แสดงอยู่ในตารางที่ 7.6

ตารางที่ 7.6 Serial Correlation Coefficients และ Autoregressive Coefficients ของแบบจำลอง AR(2)

(i)	Lake	$r_1^{(i)}$	$r_2^{(i)}$	$\hat{\phi}_1^{(i)}$	$\hat{\phi}_2^{(i)}$
1	Ontario	0.256	0.134	0.2375	0.0732
2	Erie	0.196	0.111	0.1809	0.0756
3	Superior	0.134	0.130	0.1183	0.1142
4	Michigan-Huron	0.229	0.135	0.2091	0.0870

ผลการเปรียบเทียบ AR(2) Model Correlogram กับ Historical Correlogram ในรูปที่ 7.7 พบว่าแบบจำลอง AR(2) พอดีกับอนุกรมเวลาของทะเลสาบ Ontario และ Erie พอดีกับอนุกรมเวลาของทะเลสาบ Michigan-Huron แต่ไม่พอดีกับอนุกรมเวลาของทะเลสาบ Superior

AR(2) Residual Correlogram ซึ่งคำนวณจาก $\hat{\varepsilon}_t^{(i)}$ เมื่อ $t=3$ ถึง 828 ดังรูปที่ 7.7 ให้ข้อสรุปที่น่าสนใจคือ AR(2) พอดีกับอนุกรมเวลาของทะเลสาบ Ontario และ Erie พอดีกับอนุกรมเวลาของทะเลสาบ Michigan-Huron แต่ไม่พอดีกับอนุกรมเวลาของทะเลสาบ Superior ดังนั้นอาจต้องลองทดสอบกับ AR(p) ที่ลำดับสูงกว่า 2 หรือ ARMA(p,q) อย่างไรก็ตามในตัวอย่างจะใช้แบบจำลอง AR(2) กับอนุกรมเวลาของทะเลสาบทั้ง 4

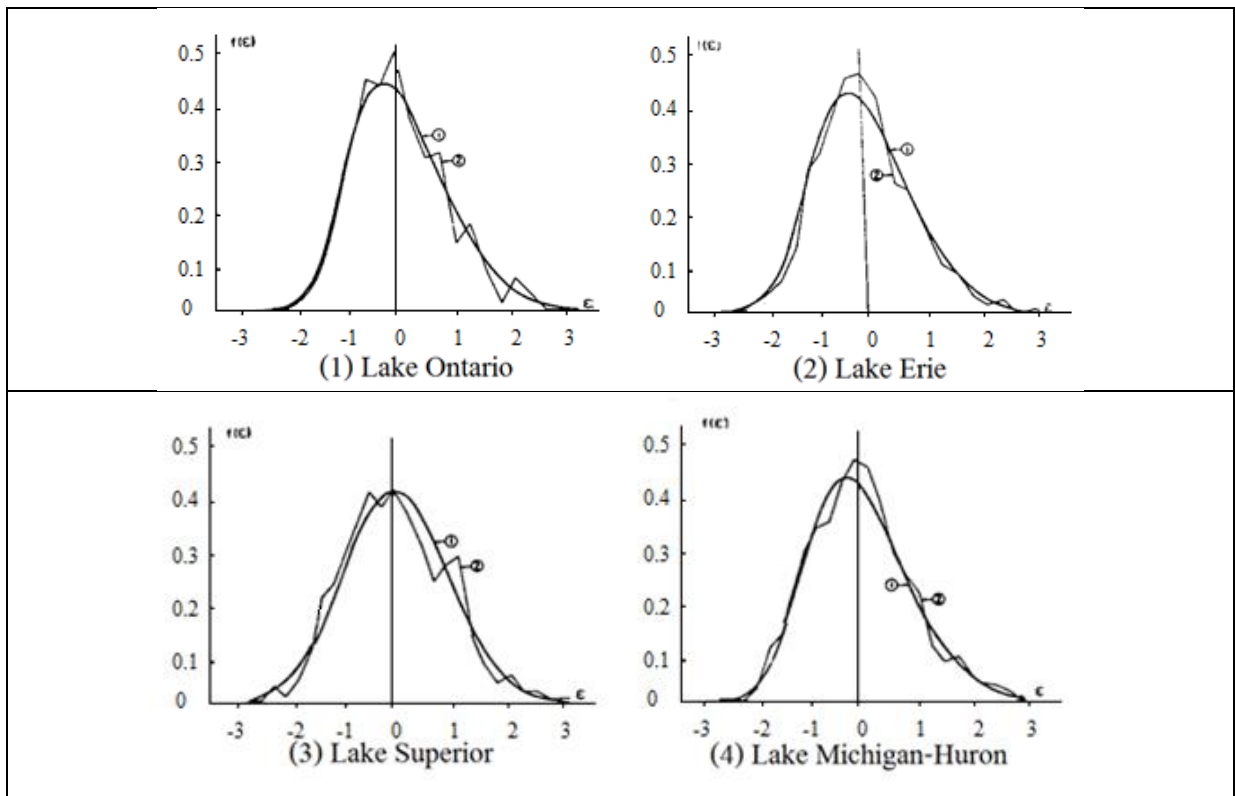
ขั้นที่ 2e เนื่องจากอนุกรมเวลาที่ใช้ในการจำลองไม่เป็นแบบ Normal จึงต้องวิเคราะห์หาฟังก์ชันการแจกแจงความน่าจะเป็นของ Residuals $\hat{\varepsilon}_t^{(i)}$ ซึ่งมี Mean=0 และ Standard Deviation=1 และพบว่า Log-Normal แบบ 3 พารามิเตอร์พอดีกับ Residuals

$$f(\varepsilon) = \frac{1}{(\varepsilon_t - \varepsilon_0)s\sqrt{2\pi}} \text{Exp} \left\{ \frac{-[\ln(\varepsilon_t - \varepsilon_0) - m]^2}{2s^2} \right\} \quad [7.56]$$

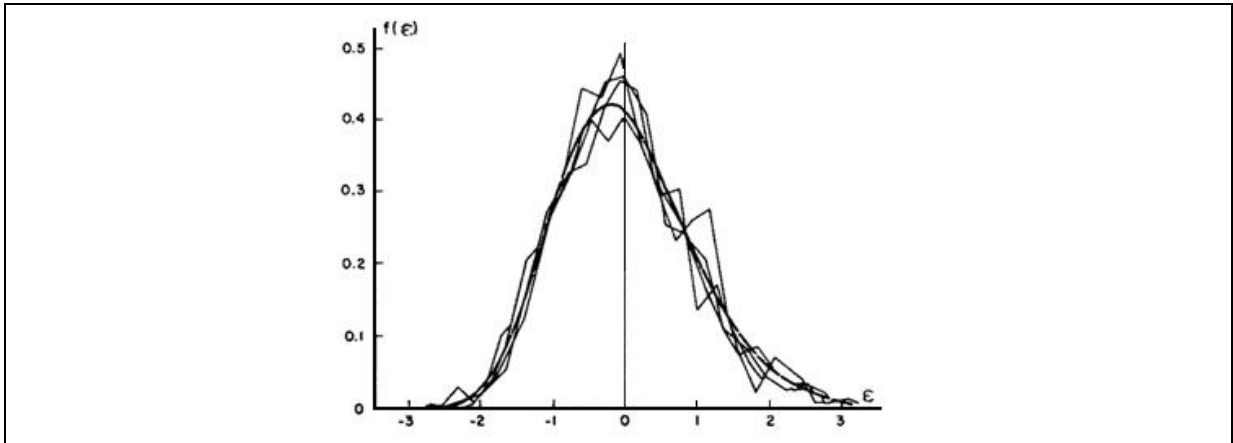
ค่าพารามิเตอร์ของ Log-Normal 3 พารามิเตอร์แสดงอยู่ในตารางที่ 7.7 Probability Density Function และ Empirical Frequency Distribution ของ Residuals แสดงอยู่ในรูปที่ 7.8 ซึ่งพบว่า Log-Normal 3 พารามิเตอร์ฟิตกับ Residual Series ได้ดี และมีการแจกแจงคล้ายกันดังรูปที่ 7.9

ตารางที่ 7.7 พารามิเตอร์ของ Log-Normal 3 พารามิเตอร์

(i)	Parameters			
	Lake	m	s	ϵ_0
1	Ontario	1.5969	0.1883	-5.0272
2	Erie	1.7454	0.1655	-5.8079
3	Superior	2.1801	0.1102	-8.9020
4	Michigan-Huron	1.7787	0.1608	-6.000
Four lakes combined		1.8016	0.1572	-6.1354



รูปที่ 7.8 Probability Density Function of Log-Normal 3 Parameters กับ Empirical Density ของ Residuals



รูปที่ 7.9 Empirical Density ของ Residuals ของ 4 ทะเลสาบเปรียบเทียบกับ Probability Density Function of Log-Normal 3 Parameters ซึ่งคำนวณโดยใช้ค่าพารามิเตอร์เฉลี่ยของ 4 ทะเลสาบ

การแปลงค่า Residuals ของ 4 ทะเลสาบ ที่เป็น Log-Normal 3 พารามิเตอร์ ให้เป็น Normal โดยใช้ค่าพารามิเตอร์เฉลี่ยของทะเลสาบทั้ง 4 จากตารางที่ 7.7 โดยสมการ

$$\varepsilon'_t{}^{(i)} = \frac{\ln(\varepsilon_t^{(i)} - \varepsilon_0) - m}{s} \tag{7.57}$$

เมื่อ $m = 1.8016$
 $s = 0.1572$
 $\varepsilon_0 = -6.1354$

เนื่องจาก Mean ของ $\varepsilon'_t{}^{(i)}$ ไม่เท่ากับ 0 และ Standard Deviation ของ $\varepsilon'_t{}^{(i)}$ ไม่เท่ากับ 1 จึงต้องมีการแปลงให้เป็นค่า Standard Normal โดยสมการ

$$\varepsilon''_t{}^{(i)} = \frac{\varepsilon'_t{}^{(i)} - \bar{\varepsilon}'^{(i)}}{s'^{(i)}} \tag{7.58}$$

$$\underline{\varepsilon}''_t = B \underline{\xi}_t \tag{7.59}$$

เมื่อ $\underline{\xi}_t$ คือ Variate Series Independent in Time and In Space ของ 4 ทะเลสาบ

เมทริกซ์ $\hat{M}_o = \hat{B} \hat{B}^T$ และ \hat{B} ของ $\varepsilon''_t{}^{(i)}$ แสดงอยู่ในตารางที่ 7.8 และ 7.9

ตารางที่ 7.8 Cross Correlation Matrix \hat{M}_o ของ $\varepsilon''_t{}^{(i)}$

Lake	Ontario (1)	Erie (2)	Superior (3)	Michigan-Huron (4)
Ontario (1)	1.000	0.608	0.246	0.541
Erie (2)	0.608	1.000	0.191	0.487
Superior (3)	0.246	0.191	1.000	0.444
Michigan-Huron (4)	0.541	0.487	0.444	1.000

ตารางที่ 7.9 ค่าประมาณของ \hat{B} ของ $\varepsilon_t^{(i)}$

Lake	Ontario (1)	Erie (2)	Superior (3)	Michigan-Huron (4)
Ontario (1)	1.000	0.000	0.000	0.000
Erie (2)	0.608	0.794	0.000	0.000
Superior (3)	0.246	0.051	0.968	0.000
Michigan-Huron (4)	0.541	0.199	0.310	0.756

ขั้นที่ 3 การทดสอบความเหมาะสม (Goodness of Fit) ของแบบจำลอง

โดยทั่วไป การทดสอบความเหมาะสมของแบบจำลองหลายตัวแปรคือการตรวจสอบว่า Residual Series ($\varepsilon_t^{(i)}$) ต้องเป็นอิสระทั้งด้านเวลาและพื้นที่ (Independence in Time and in Space) และ Residual Series ต้องมีการแจกแจงแบบปกติ แต่สำหรับตัวอย่างนี้ได้ระบุไว้ในหัวข้อ 7.12(1) ว่าจะจำลองอนุกรมเวลาที่ไม่ได้มีการแจกแจงปกติ ดังนั้นการทดสอบความเหมาะสมจึงพิจารณาเฉพาะความเป็นอิสระเท่านั้น Correlogram ของ $\varepsilon_t^{(i)}$ ในรูปที่ 7.7 สามารถสรุปได้ว่า Residual Series ($\varepsilon_t^{(i)}$) เป็นอิสระด้านเวลา แต่ยังไม่สามารถทดสอบความเป็นอิสระด้านพื้นที่ และจะใช้การทดสอบค่าสถิติของอนุกรมเวลาสังเคราะห์ที่เปรียบเทียบกับค่าสถิติอนุกรมเวลาตัวอย่างที่เรียกว่า Optional Test แทน ดังแสดงในขั้นที่ 4

ขั้นที่ 4 การทดสอบคุณสมบัติทางสถิติของอนุกรมเวลาสังเคราะห์ (Optional Test)

แบบจำลองหลายตัวแปรของอนุกรมเวลารายเดือนของ NBS ของ Great Lake คือ

$$y_{v,t}^{(i)} = \hat{\mu}_\tau^{(i)} - \hat{\sigma}_\tau^{(i)} z_{v,t}^{(i)}$$

ค่า $\hat{\mu}_\tau^{(i)}$ และ $\hat{\sigma}_\tau^{(i)}$ แสดงอยู่ในตารางที่ 7.5 และ

$$\begin{aligned}\hat{z}_t^{(1)} &= 0.2375 \hat{z}_{t-1}^{(1)} + 0.0732 \hat{z}_{t-2}^{(1)} + 0.964 \varepsilon_t^{(1)} \\ \hat{z}_t^{(2)} &= 0.1809 \hat{z}_{t-1}^{(2)} + 0.0756 \hat{z}_{t-2}^{(2)} + 0.978 \varepsilon_t^{(2)} \\ \hat{z}_t^{(3)} &= 0.1183 \hat{z}_{t-1}^{(3)} + 0.1142 \hat{z}_{t-2}^{(3)} + 0.984 \varepsilon_t^{(3)} \\ \hat{z}_t^{(4)} &= 0.2091 \hat{z}_{t-1}^{(4)} + 0.0870 \hat{z}_{t-2}^{(4)} + 0.970 \varepsilon_t^{(4)}\end{aligned}$$

และจากสมการ [7.56] จะได้

$$\varepsilon_t^{(i)} = -6.1354 + \text{Exp} \left(1.8016 + 0.1572 \varepsilon_t^{\prime(i)} \right); i = 1, \dots, 4$$

และจากสมการ [7.58] จะได้

$$\begin{aligned}\varepsilon_t^{\prime(1)} &= 1.000 \xi_t^{(1)} \\ \varepsilon_t^{\prime(2)} &= 0.608 \xi_t^{(1)} + 0.794 \xi_t^{(2)} \\ \varepsilon_t^{\prime(3)} &= 0.246 \xi_t^{(1)} + 0.051 \xi_t^{(2)} + 0.968 \xi_t^{(3)} \\ \varepsilon_t^{\prime(4)} &= 0.541 \xi_t^{(1)} + 0.199 \xi_t^{(2)} + 0.310 \xi_t^{(3)} + 0.756 \xi_t^{(4)}\end{aligned}$$

เมื่อ $\xi_t^{(i)}$ คือ Independent Normal Random Variables ซึ่งมี Mean=0 และ Variance=1 และ $i=1,2,3,4$

ใช้แบบจำลองหลายตัวแปรนี้ สังเคราะห์อนุกรมเวลารายเดือนของ NBS ของ 4 ทะเลสาบ จำนวน 20 ชุด แต่ละชุดมีความยาว 50 ปี นำมาคำนวณหาค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน และค่าสัมประสิทธิ์ความเบ้ ดังแสดงในตารางที่ 7.10 (Lake Erie) และ 7.11 (Lake Superior)

ผลการเปรียบเทียบค่าสถิติจากอนุกรมเวลาตัวอย่างและอนุกรมเวลาสังเคราะห์ พบว่าค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลาสังเคราะห์ของทะเลสาบ Erie และ Superior มีคุณสมบัติคล้ายคลึงกับค่าของอนุกรมเวลาตัวอย่าง และทะเลสาบ Superior มีค่าสัมประสิทธิ์ความเบ้คล้ายคลึงกัน แต่ทะเลสาบ Erie มีค่าสัมประสิทธิ์ของอนุกรมเวลาสังเคราะห์ต่ำกว่าค่าสัมประสิทธิ์ของอนุกรมเวลาตัวอย่าง จึงควรปรับแบบจำลองเพื่อแก้ปัญหานี้ต่อไป

ตารางที่ 7.10 คุณสมบัติทางสถิติของอนุกรมเวลารายเดือนตัวอย่างและอนุกรมเวลารายเดือนที่สังเคราะห์ 20 ชุด ของ NBS ของทะเลสาบ Erie

	<i>Mean (\bar{y})</i>	<i>St . Deviation (s)</i>	<i>Skewness Coeff. (g)</i>
Historic Sample	16.797	40.122	0.648
Generated Sample			
1	16.898	41.875	0.546
2	16.417	40.398	0.652
3	17.268	42.357	0.647
4	17.157	41.401	0.414
5	16.558	39.237	0.721
6	17.090	40.467	0.473
7	16.557	39.598	0.490
8	16.994	39.979	0.541
9	17.019	40.175	0.604
10	17.060	42.238	0.670
11	16.842	40.100	0.628
12	16.932	40.849	0.558
13	16.994	41.152	0.464
14	16.692	39.771	0.469
15	16.818	39.750	0.553
16	16.924	39.048	0.511
17	16.305	38.920	0.614
18	16.690	39.460	0.614
19	17.027	40.701	0.465
20	16.731	39.721	0.631
Mean of 20 values	16.849	40.360	0.563

ตารางที่ 7.11 คุณสมบัติทางสถิติของอนุกรมเวลารายเดือนตัวอย่างและอนุกรมเวลารายเดือนที่สังเคราะห์ 20 ชุด ของ NBS ของทะเลสาบ Superior

	Mean (\bar{y})	St . Deviation (s)	Skewness Coeff. (g)
Historic Sample	71.932	82.160	0.489
Generated Sample			
1	71.861	82.534	0.523
2	71.429	80.368	0.465
3	71.912	83.005	0.482
4	72.828	84.521	0.602
5	72.181	84.933	0.466
6	71.546	79.440	0.483
7	72.481	82.042	0.416
8	71.800	81.416	0.496
9	71.511	80.457	0.419
10	72.357	82.478	0.436
11	71.560	81.160	0.451
12	71.922	83.157	0.484
13	72.029	83.011	0.544
14	72.406	83.493	0.403
15	72.519	85.262	0.649
16	72.343	84.889	0.599
17	71.711	78.496	0.431
18	71.973	83.051	0.589
19	72.212	82.151	0.482
20	71.243	80.263	0.507
Mean of 20 values	71.991	82.206	0.496

ตัวอย่าง Lag Zero Cross Correlation Matrix (M_0) ของอนุกรมเวลาที่สังเคราะห์ 1 ชุด คือ

$$M_0 = \begin{bmatrix} 1.000 & 0.622 & 0.272 & 0.516 \\ 0.622 & 1.000 & 0.232 & 0.505 \\ 0.272 & 0.232 & 1.000 & 0.415 \\ 0.516 & 0.505 & 0.415 & 1.000 \end{bmatrix}$$

ซึ่งมีคุณสมบัติใกล้เคียงกับ M_0 ของอนุกรมเวลาตัวอย่างในตารางที่ 7.8

ขั้นที่ 5 Reliability ของแบบจำลอง

คำนวณค่าพารามิเตอร์จากอนุกรมเวลาสังเคราะห์แต่ละชุด (อนุกรมเวลาที่สังเคราะห์ขึ้นทั้งหมด 20 ชุด) และวิเคราะห์หาการแจกแจงความน่าจะเป็นของค่าพารามิเตอร์ ซึ่งสามารถนำมาคำนวณหาช่วงความเชื่อมั่นของค่าพารามิเตอร์ต่อไป

7.13 ข้อจำกัดของการจำลองแบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic

- (1) แบบจำลองแบบหลายตัวแปรกับอนุกรมเวลา Periodic ต้องรักษา Dependence Structure ทั้งทางด้านเวลาและพื้นที่ จึงทำให้มีพารามิเตอร์มากและซับซ้อน ทางเลือกที่นิยมใช้คือการจำลองด้วยแบบจำลองแบบแยกส่วน (Disaggregation Model) ซึ่งจะได้กล่าวถึงในบทที่ 8
- (2) แบบจำลอง AR เป็นแบบจำลองแบบ Short Term Memory และแบบจำลองหลายตัวแปรโดยทั่วไปนิยมใช้ AR(1) และ AR(2) เท่านั้น เนื่องจากความยุ่งยากในการประมาณพารามิเตอร์สำหรับแบบจำลองหลายตัวแปรที่มีลำดับสูงๆ
- (3) แบบจำลอง ARMA มีข้อดีคือสามารถรักษาทั้ง Time และ Space Dependence แต่การใช้แบบจำลองหลายตัวแปรแบบ ARMA ยังมีปัญหาในการประมาณค่าพารามิเตอร์ ดังนั้นปกติจึงนิยมใช้ทางเลือกในการจำลองโดยรักษา Time Dependence ก่อน แล้วจึงหาวิธีรักษา Space Dependence ที่หลัง แต่มีข้อจำกัดรักษา Space Dependence ในทางปฏิบัติซึ่งใช้เฉพาะ M_0 เท่านั้น

7.14 การประยุกต์แบบหลายตัวแปรกับอนุกรมเวลาแบบ Periodic

แบบจำลองหลายตัวแปรแบบ Periodic มักใช้ในการสังเคราะห์ข้อมูลรายเดือนของระบบแหล่งน้ำที่มีหลายอ่างเก็บน้ำ อาจใช้ในการสังเคราะห์ปริมาณน้ำท่ารายเดือน หรือฝนรายเดือนแล้วแปลงเป็นน้ำท่าโดยใช้แบบจำลอง ฝน-น้ำท่า ก็ได้ หรืออาจใช้ในการสังเคราะห์ทั้งฝนและน้ำท่าพร้อมๆ กันก็ได้ เช่นกรณีโครงการชลประทาน Majes-Siguas ใช้ Multivariate AR(1) ในการสังเคราะห์ปริมาณน้ำท่ารายเดือน กรณีที่แบบจำลองหลายตัวแปรแบบ Periodic ไม่สามารถจำลองคุณสมบัติรายปีได้ อาจต้องใช้แบบจำลองแยกส่วน (Disaggregation Model) ในบทที่ 8 แทน

7.15 เอกสารอ้างอิง

Salas, J., J.W.Delleur, V.Yevjevich and W.L.Lane. 1980. Applied Modelling of Hydrologic Time Series. Water Resources Publications, USA. 484p.

บทที่ 8

การจำลองอนุกรมเวลาด้วยแบบจำลองแบบแยกส่วน

DISAGGREGATION MODELING

8.1 คำนำ

แบบจำลองแยกส่วนคือแบบจำลองที่พัฒนาขึ้นมาเพื่อแก้ปัญหา Parsimony ของพารามิเตอร์ของแบบจำลองหลายตัวแปร (Multivariate) กรณีนำไปใช้ในการสังเคราะห์อนุกรมเวลาแบบ Periodic ซึ่งมีจำนวนสถานีมาก แบบจำลองแยกส่วนยังช่วยให้สามารถรักษาคูสมบัติทางสถิติมากกว่า 1 ระดับ เช่นสามารถรักษาคูสมบัติทางสถิติของอนุกรมเวลารายปี และรายฤดูกาล แบบจำลองแยกส่วนพัฒนาจากสมมติฐานว่าตัวแปรมีสหสัมพันธ์เชิงเส้น

Harms และ Campbel (1967) เป็นคนแรกที่เสนอแนวคิดในการทำแบบจำลองแยกส่วน โดยพัฒนาต่อจากแบบจำลองของ Thomas และ Fiering(1962) แบบจำลองแยกส่วนที่ถูกพัฒนาและเป็นที่ยอมรับกันโดยทั่วไปเป็นครั้งแรกคือแบบจำลองของ Valencia และ Schaake(1973) ซึ่งถือเป็นจุดเริ่มต้นที่มีการใช้แบบจำลองแยกส่วนอย่างจริงจังในการสังเคราะห์อนุกรมเวลาทางอุทกวิทยา และ Lane(1979) เป็นคนแรกที่พยายามพัฒนาคอมพิวเตอร์โปรแกรมเพื่อช่วยในการสร้างแบบจำลองทางสโตแคสติกแบบแยกส่วน

แบบจำลองแยกส่วนสามารถแบ่งออกเป็น 2 ลักษณะคือ (1) การแยกส่วนตามเวลา (Temporal Disaggregation) เพื่อแยกอนุกรมเวลารายปีเป็นรายฤดูกาล (Annual to Seasonal Disaggregation) และ (2) การแยกส่วนตามพื้นที่ (Spatial Disaggregation) เพื่อแยกอนุกรมเวลารายปีของสถานีหลักเป็นอนุกรมเวลารายปีของสถานีย่อย (Key to Sub-station Disaggregation) ในบทนี้จะกล่าวถึงแบบจำลองแยกส่วน 3 แบบ คือแบบจำลองแบบพื้นฐาน (Basic Model) แบบจำลองแบบขยาย (Extended Model) และแบบจำลองแบบย่อ (Condensed Model)

8.2 แบบจำลองแยกส่วนทั่วไป (General Disaggregation Models)

แบบจำลองแยกส่วนทั่วไปจะอยู่ในรูปของแบบจำลองเชิงเส้น หรือ Linear Dependence Model ดังนี้

$$Y = AX + B\varepsilon \quad [8.1]$$

เมื่อ Y =Column Vector ของค่าปัจจุบันของอนุกรมเวลาซึ่งสัมพันธ์กับค่าปัจจุบันของอนุกรมเวลา X หรือ Dependent Series

X =Column Vector ของค่าปัจจุบันของอนุกรมเวลาหลักหรืออนุกรมเวลาแบบอิสระตามเวลา (Key หรือ Time Independent Series)

ถ้าค่าเฉลี่ยของ X และ Y ไม่เท่ากับศูนย์ แบบจำลองในสมการ [8.1] จะต้องมีพารามิเตอร์เมทริกซ์ C เพิ่มดังสมการ

$$Y = AX + B\underline{\varepsilon} + C \quad [8.2]$$

เมื่อ C = พารามิเตอร์เมทริกซ์ ซึ่งเป็น Column Vector ขนาดเท่ากับเมทริกซ์ Y

8.3 แบบจำลองแยกส่วนตามเวลาแบบ Single Site (Single Site Temporal Disaggregation Model)

การจำลองแยกส่วนตามเวลาแบบ Single Site คือแบบจำลองซึ่งแยกอนุกรมเวลารายปีของ 1 สถานี เป็นอนุกรมเวลารายฤดูกาลเช่นรายเดือน มีรูปแบบที่สำคัญ 3 แบบคือ

- (1) แบบจำลองพื้นฐาน (Basic Model) (Valencia and Schaake, 1973)

$$Y = AX + B\underline{\varepsilon} \quad [8.3]$$

- (2) แบบจำลองแบบขยาย (Extended Model) (Mejia and Rousselle, 1976)

$$Y = AX + B\underline{\varepsilon} + CZ \quad [8.4]$$

เมื่อ Y = Column Vector ของค่ารายเดือนของปีปัจจุบัน ซึ่งมีขนาด (12×1)

X = ค่าปัจจุบันของอนุกรมเวลารายปี (1×1)

Z = Column Vector ของค่ารายเดือนจำนวน w' เดือนของปีก่อนหน้านั้น ซึ่งมีขนาด $(w' \times 1)$

$\underline{\varepsilon}$ = Column Vector ของค่าปัจจุบันของ Completely Random Series หรือ Stochastic Term
ซึ่งเมทริกซ์มีขนาด (12×1)

A = พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด (12×1)

B = พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด (12×12)

C = พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(12 \times w')$

กรณีที่จำนวนสถานีเท่ากับ n และ w' เท่ากับ 12 จะสามารถเขียนแบบจำลองแยกส่วนแบบขยายเต็มรูปได้ ดังรูปที่ 8.2

$$\begin{bmatrix} y_{v,1}^{(1)} \\ y_{v,2}^{(1)} \\ \vdots \\ y_{v,12}^{(1)} \\ y_{v,1}^{(2)} \\ y_{v,2}^{(2)} \\ \vdots \\ y_{v,12}^{(2)} \\ \vdots \\ y_{v,1}^{(m)} \\ y_{v,2}^{(m)} \\ \vdots \\ y_{v,12}^{(m)} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{12n,1} & a_{12n,2} & \dots & a_{12n,n} \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1,12n} \\ b_{21} & b_{22} & \dots & b_{2,12n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ b_{12n,1} & b_{12n,2} & \dots & b_{12n,12n} \end{bmatrix} \begin{bmatrix} \varepsilon_1^{(1)} \\ \varepsilon_2^{(1)} \\ \vdots \\ \varepsilon_{12}^{(1)} \\ \varepsilon_1^{(2)} \\ \varepsilon_2^{(2)} \\ \vdots \\ \varepsilon_{12}^{(2)} \\ \vdots \\ \varepsilon_1^{(m)} \\ \varepsilon_2^{(m)} \\ \vdots \\ \varepsilon_{12}^{(m)} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1,12n} \\ c_{21} & c_{22} & \dots & c_{2,12n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_{12n,1} & c_{12n,2} & \dots & c_{12n,12n} \end{bmatrix} \begin{bmatrix} y_{v-1,1}^{(1)} \\ y_{v-1,2}^{(1)} \\ \vdots \\ y_{v-1,12}^{(1)} \\ y_{v-1,1}^{(2)} \\ y_{v-1,2}^{(2)} \\ \vdots \\ y_{v-1,12}^{(2)} \\ \vdots \\ y_{v-1,1}^{(m)} \\ y_{v-1,2}^{(m)} \\ \vdots \\ y_{v-1,12}^{(m)} \end{bmatrix}$$

รูปที่ 8.2 Full Form of Extended Disaggregation Model: Y=AX+Bε+CZ

(3) แบบจำลองแบบย่อ (Condensed Model) (Lane, 1979) ใช้จำลองที่ละ 1 Period เช่น 1 เดือน

$$Y_{\tau} = A_{\tau}X + B_{\tau}\underline{\varepsilon} + C_{\tau}Y_{\tau-1} \tag{8.5}$$

เมื่อ τ = เดือนปัจจุบัน เมื่อ $\tau = 1$ ถึง 12 ($w = 12$)

Y_{τ} = คำนี้าทำของเดือนปัจจุบัน τ

X = คำนี้าทำของปีปัจจุบัน

$Y_{\tau-1}$ = คำนี้าทำของเดือนก่อนหน้านั้น $\tau-1$

$\underline{\varepsilon}$ = Column Vector ของค่าปัจจุบันของ Completely Random Series หรือ Stochastic Term

$A_{\tau}, B_{\tau}, C_{\tau}$ = ค่าพารามิเตอร์ของเดือน τ

8.4 แบบจำลองแยกส่วนตามเวลาแบบ Multisite (Multisite Temporal Disaggregation Model)

แบบจำลองแยกส่วนตามเวลาแบบ Multisite คือแบบจำลองซึ่งแยกส่วนอนุกรมเวลารายปีเป็นรายฤดูกาล สำหรับกรณีที่มีอนุกรมเวลามากกว่า 1 ชุด หรือเป็นแบบ Multisite แทนที่จะเป็นแบบ Single Site ตามที่กล่าวถึงในหัวข้อ 8.2 รูปแบบของแบบจำลองจะคล้ายกับแบบ Single Site เช่น แบบพื้นฐาน (Basic) แบบขยาย (Extended) และแบบย่อ (Condensed) แต่เมทริกซ์มีขนาดใหญ่กว่าตามขนาดของ Multisite

ในหัวข้อนี้จะกล่าวถึง แบบจำลองแยกส่วนแบบขยายสำหรับกรณี Multisite Temporal

$$Y = AX + B\underline{\varepsilon} + CZ \tag{8.6}$$

เมื่อ Y =Column Vector ของค่าปัจจุบันของอนุกรมเวลารายเดือน ซึ่งมีขนาด ($nw \times 1$)

X =Column Vector ของค่าปัจจุบันของอนุกรมเวลารายปี ซึ่งมีขนาด $(n \times 1)$

Z = Column Vector ของค่าปีก่อนหน้าของอนุกรมเวลารายเดือน ซึ่งมีขนาด $(nw' \times 1)$

$\underline{\epsilon}$ = Column Vector ของค่าปัจจุบันของ Completely Random Series หรือ Stochastic Term
ซึ่งมีขนาด $(nw \times 1)$

A =พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(nw \times n)$

B =พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(nw \times nw)$

C =พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(nw \times nw')$

n =จำนวน Sites ที่พิจารณา

w =จำนวนฤดูกาลใน 1 ปี

8.5 แบบจำลองแยกส่วนตามพื้นที่ (Spatial Disaggregation Model)

แบบจำลองแยกส่วนตามพื้นที่ คือแบบจำลองที่ใช้แยกอนุกรมเวลารายปีของสถานีหลัก (Key Stations) เป็นอนุกรมเวลารายปีของสถานีย่อย (Substation) ซึ่งจำนวนสถานีหลักอาจมากกว่า 1 สถานีก็ได้ และสถานีย่อยคือสถานีที่มีสหสัมพันธ์เชิงเส้นกับสถานีหลัก

รูปแบบของแบบจำลองแยกส่วนตามพื้นที่ จะมีรูปแบบคล้ายกับแบบจำลองแยกส่วนตามเวลา เช่น แบบพื้นฐาน (Basic) แบบขยาย (Extended) และแบบย่อ (Condensed)

Lane(1979) เสนอแนะแบบจำลองแยกส่วนตามพื้นที่ โดยใช้แบบจำลองแบบขยายของ Mejia and Roussele ดังนี้

$$Y = AX + B\underline{\epsilon} + CZ \quad [8.7]$$

เมื่อ Y = Column Vector ของค่ารายปีของสถานีย่อยของปีปัจจุบัน ซึ่งมีขนาด $(n \times 1)$

X = Column Vector ของค่ารายปีของสถานีหลักของปีปัจจุบัน ซึ่งมีขนาด $(m \times 1)$

Z = Column Vector ของค่ารายปีของสถานีย่อยของปีก่อนหน้า ซึ่งมีขนาด $(n \times 1)$

$\underline{\epsilon}$ = Column Vector ของค่าปัจจุบันของ Completely Random Series หรือ Stochastic Term
ซึ่งมีขนาด $(n \times 1)$

A = พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(n \times m)$

B = พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(n \times n)$

C = พารามิเตอร์เมทริกซ์ ซึ่งมีขนาด $(n \times n)$

n = จำนวน Sites ที่พิจารณา

8.6 แบบจำลองแยกส่วนตามพื้นที่และตามเวลา (Spatial and Temporal Disaggregation Model)

บางกรณีอาจต้องทำการจำลองโดยการจำลองแยกส่วนตามพื้นที่ที่ผสมกับการจำลองแยกส่วนตามเวลา เช่น ถ้าต้องการสังเคราะห์อนุกรมเวลารายเดือนของน้ำท่าที่ไหลเข้าระบบอ่างเก็บน้ำ ซึ่งกรณีนี้อาจทำได้โดยใช้แบบจำลองหลายตัวแปรกับอนุกรมเวลารายเดือนโดยตรง แต่จำนวนพารามิเตอร์จะมาก และอาจมีปัญหา Parsimony ของพารามิเตอร์ได้ ทางเลือกคือการใช้แบบจำลองแยกส่วนตามพื้นที่และตามเวลา

เช่น ถ้าต้องการสังเคราะห์อนุกรมเวลารายเดือนของน้ำท่าที่ไหลเข้าระบบอ่างเก็บน้ำ ซึ่งมีจำนวน $n+1$ อ่างเก็บน้ำ กำหนดให้ 1 อ่างเป็นสถานีหลัก และอีก n อ่างเก็บน้ำเป็นสถานีย่อย การจำลองจะสามารถทำได้หลายแบบ ดังนี้

แบบที่ 1-Multivariate Periodic Time Series Model

แบบที่ 2-AR/ARMA + Spatial + Temporal Disaggregation Model

ขั้นที่ 1-จำลองเพื่อสังเคราะห์ปริมาณน้ำท่ารายปีของสถานีหลัก (Key Station) โดยใช้แบบจำลอง AR หรือ ARMA

ขั้นที่ 2-จำลองแบบแยกส่วนตามพื้นที่ เพื่อสังเคราะห์ปริมาณน้ำท่ารายปีของ n สถานีย่อย จากปริมาณน้ำท่ารายปีของสถานีหลัก

ขั้นที่ 3-จำลองแบบแยกส่วนตามเวลา เพื่อสังเคราะห์ปริมาณน้ำท่ารายเดือนของ $n+1$ อ่างเก็บน้ำ จากปริมาณน้ำท่ารายปีของ $n+1$ อ่างเก็บน้ำ

แบบที่ 3-Multivariate Annual Time Series Model + Temporal Disaggregation Model

ขั้นที่ 1-ใช้ Multivariate Annual Time Series เพื่อสังเคราะห์ปริมาณน้ำท่ารายปีของ $n+1$ อ่างเก็บน้ำ

ขั้นที่ 2-จำลองแบบแยกส่วนตามเวลา เพื่อสังเคราะห์ปริมาณน้ำท่ารายเดือนของ $n+1$ อ่างเก็บน้ำ จากปริมาณน้ำท่ารายปีของ $n+1$ อ่างเก็บน้ำ

8.7 การประมาณค่าพารามิเตอร์ของแบบจำลอง

8.7.1 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาพื้นฐาน (Basic Model)

แบบจำลองแยกส่วนแบบพื้นฐาน (Basic Model)

$$Y = AX + B\varepsilon$$

$(nw \times 1) = (nw \times n)(n \times 1) + (nw \times nw)(nw \times 1) \dots \dots \dots$ ขนาดเมทริกซ์

$$E(Y) = 0 \text{ และ } E(X) = 0$$

สมการสำหรับคำนวณหาพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาพื้นฐาน คือ

$$\hat{A} = S_{YX}S_{XX}^{-1} \quad [8.8]$$

$$\hat{B}\hat{B}^T = S_{YY} - S_{YX}S_{XX}^{-1}S_{XY} = S_{YY} - \hat{A}S_{XY} \quad [8.9]$$

เมื่อ $S_{XX} = E(XX^T) = \text{Variance-Covariance Matrix of } X$

$$S_{YY} = E(YY^T) = \text{Variance-Covariance Matrix of } Y$$

$$S_{YX} = E(YX^T) = \text{Covariance Matrix of } Y \text{ and } X$$

$$S_{XY} = E(XY^T) = \text{Covariance Matrix of } X \text{ and } Y$$

การ Derive หาพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาพื้นฐาน (Basic Model)

$$Y = AX + B\underline{\varepsilon}$$

$$E(YX^T) = E(AXX^T + B\underline{\varepsilon}X^T) = AE(XX^T)$$

$$\text{ให้ } E(YX^T) = S_{YX} = \sum_{v=1}^N \frac{Y_v X_v^T}{N-1}$$

$$E(XX^T) = S_{XX} = \sum_{v=1}^N \frac{X_v X_v^T}{N-1}$$

$$S_{YX} = \hat{A}S_{XX}$$

$$\hat{A} = S_{YX}S_{XX}^{-1}$$

$$E(YY^T) = E(AX + B\underline{\varepsilon})Y^T$$

$$= E(AXY^T) + E(B\underline{\varepsilon})(AX + B\underline{\varepsilon})^T$$

$$= AE(XY^T) + E(B\underline{\varepsilon}X^T A^T + B\underline{\varepsilon}\underline{\varepsilon}^T B^T)$$

$$= AE(XY^T) + E(B\underline{\varepsilon}\underline{\varepsilon}^T B^T)$$

$$S_{YY} = \hat{A}S_{XY} + \hat{B}\hat{B}^T$$

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A}S_{XY}$$

8.7.2 ขั้นตอนการคำนวณหาค่าพารามิเตอร์ (Calculation Procedure)

ในหัวข้อนี้จะแสดงขั้นตอนการคำนวณหาค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาพื้นฐาน โดยกำหนดให้

$i = 1, \dots, n$ เมื่อจำนวนสถานี $n = 2$

$\tau = 1, \dots, \omega$ เมื่อจำนวนฤดูกาล (No. of Periods) ในหนึ่งปี $\omega = 3$

$v = 1, \dots, N$ เมื่อ N คือจำนวนปีของอนุกรมเวลา

$x(i, v) =$ ปริมาณน้ำท่ารายปีในปีที่ v ของสถานี i

$y(i, v, \tau) =$ ปริมาณน้ำท่ารายเดือนในเดือน τ ปีที่ v ของสถานี i

$S_{xx} = E(XX^T) = \text{Variance-Covariance Matrix of } X$

$$S_{XX} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} x(1, v) \\ x(2, v) \end{bmatrix} [x(1, v) \quad x(2, v)]$$

$$S_{XX} = \frac{1}{N} \begin{bmatrix} \sum x^2(1, v) & \sum x(1, v)x(2, v) \\ \sum x(2, v)x(1, v) & \sum x^2(2, v) \end{bmatrix}$$

[8.10]

	i=1			i=2			i=1	i=2
	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=1$	$\tau=2$	$\tau=3$		
v	$y(1, v, 1)$	$y(1, v, 2)$	$y(1, v, 3)$	$y(2, v, 1)$	$y(2, v, 2)$	$y(2, v, 3)$	$x(1, v)$	$x(2, v)$
1	x	x	x	y	y	y	$\sum_{\tau} y(1,1,\tau)$	$\sum_{\tau} y(2,1,\tau)$
2	x	x	x	y	y	y	$\sum_{\tau} y(1,2,\tau)$	$\sum_{\tau} y(2,2,\tau)$
.
N	x	x	x	y	y	y	$\sum_{\tau} y(1,N,\tau)$	$\sum_{\tau} y(2,N,\tau)$
Mean	0	0	0	0	0	0	0	0
Var	1	1	1	1	1	1	Var(x(1))	Var(x(2))

ในการทำงานเดียวกันจะสามารถคำนวณหา S_{YY} , S_{XY} หรือ S_{YX} ได้ดังนี้

$$S_{YY} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y(1, v, 1) \\ y(1, v, 2) \\ y(1, v, 3) \\ y(2, v, 1) \\ y(2, v, 2) \\ y(2, v, 3) \end{bmatrix} [y(1, v, 1), y(1, v, 2), y(1, v, 3), y(2, v, 1), y(2, v, 2), y(2, v, 3)] \quad [8.11]$$

$$S_{YX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y(1, v, 1) \\ y(1, v, 2) \\ y(1, v, 3) \\ y(2, v, 1) \\ y(2, v, 2) \\ y(2, v, 3) \end{bmatrix} [x(1, v), x(2, v)] \quad [8.12]$$

$$S_{XY} = S_{YX}^T \quad [8.13]$$

ค่า S_{XX} , S_{YY} และ S_{XY} คือค่า Variance-Covariance ของ X และ Y ซึ่งสามารถคำนวณได้โดยใช้ฟังก์ชันใน Excel ซึ่งทำได้ ดังนี้

จากเมนูบาร์เลือก DATA+DATA ANALYSIS + COVARIANCE แต่ต้องจำไว้ว่าค่า Variance-Covariance ที่หาได้โดยวิธีนี้ คือ Biased Variance-Covariance ต้องปรับแก้โดยการคูณค่าที่คำนวณได้โดย $(N/N-1)$ หรือใช้ Function $\{=SUMPRODUCT(A,B)/(N-1)\}$ ดังตัวอย่างในตารางที่ 8.1

ตารางที่ 8.1 ตัวอย่างการคำนวณ Variance-Covariance Matrix ด้วยโมดูล COVAIANCE ใน DATA ANALYSIS และฟังก์ชัน SUMPRODUCT

v	$y(1, v, 1)$	$y(1, v, 2)$	$y(1, v, 3)$	$y(2, v, 1)$	$y(2, v, 2)$	$y(2, v, 3)$	$x(1, v)$	$x(2, v)$	
1	-13.3	-11.7	0.8	-4.7	1.4	5.4	-24.2	2.1	
2	-11.3	-7.7	2.8	-1.7	-2.6	-4.6	-16.2	-8.9	
3	-10.3	-5.7	3.8	1.3	-0.6	-3.6	-12.2	-2.9	
4	-16.3	-2.7	-4.2	20.3	0.4	-2.6	-23.2	18.1	
5	-14.3	-13.7	-3.2	1.3	1.4	-1.6	-31.2	1.1	
6	-13.3	-11.7	-2.2	12.3	-2.6	-0.6	-27.2	9.1	
7	69.7	6.3	-1.2	-6.7	-1.6	0.4	74.8	-7.9	
8	-16.3	10.3	-0.2	-5.7	-0.6	1.4	-6.2	-4.9	
9	36.7	28.3	0.8	-7.7	2.4	2.4	65.8	-2.9	
10	-11.3	8.3	2.8	-8.7	2.4	3.4	-0.2	-2.9	
Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Use function SUMPRODUCT(A,B)/(N-1)									
S_{yy}	850.68	212.54	-2.29	-111.46	-1.76	16.13	1060.93	-97.08	S_{yx}
	212.54	175.12	7.18	-60.77	9.87	15.76	394.84	-35.14	
	-2.29	7.18	7.29	-15.71	0.20	0.64	12.18	-14.87	
	-111.46	-60.77	-15.71	89.34	-5.58	-15.91	-187.93	67.86	
	-1.76	9.87	0.20	-5.58	3.60	3.60	8.31	1.62	
16.13	15.76	0.64	-15.91	3.60	10.27	32.53	-2.04		
S_{xy}	1060.93	394.84	12.18	-187.93	8.31	32.53	1467.96	-147.09	S_{xx}
	-97.08	-35.14	-14.87	67.86	1.62	-2.04	-147.09	67.43	
Use COVARIANCE ใน DATA ANALYSIS จะได้ Biased Covariance ต้องปรับแก้ด้วย (N/N-1)									
	$y(1, v, 1)$	$y(1, v, 2)$	$y(1, v, 3)$	$y(2, v, 1)$	$y(2, v, 2)$	$y(2, v, 3)$	$x(1, v)$	$x(2, v)$	
$y(1, v, 1)$	765.61								
$y(1, v, 2)$	191.29	157.61							
$y(1, v, 3)$	-2.06	6.46	6.56						
$y(2, v, 1)$	-100.31	-54.69	-14.14	80.41					
$y(2, v, 2)$	-1.58	8.88	0.18	-5.02	3.24				
$y(2, v, 3)$	14.52	14.18	0.58	-14.32	3.24	9.24			
$x(1, v)$	954.84	355.36	10.96	-169.14	7.48	29.28	1321.16		
$x(2, v)$	-87.37	-31.63	-13.38	61.07	1.46	-1.84	-132.38	60.69	
S_{yy}	850.68								
	212.54	175.12							
	-2.29	7.18	7.29						
	-111.46	-60.77	-15.71	89.34					
	-1.76	9.87	0.20	-5.58	3.60				
16.13	15.76	0.64	-15.91	3.60	10.27				
S_{xy}	1060.93	394.84	12.18	-187.93	8.31	32.53	1467.96		
	-97.08	-35.14	-14.87	67.86	1.62	-2.04	-147.09	67.43	

ขนาดของ Variance-Covariance Matrices ของ X และ Y และพารามิเตอร์เมทริกซ์คือ

$$\begin{aligned}
 S_{XX} &= (n \times n) \\
 S_{YY} &= (nw \times nw) \\
 S_{YX} &= (nw \times n) \\
 S_{XY} &= (n \times nw) \\
 S_{XX}^{-1} &= (n \times n) \\
 \hat{A} &= S_{YX}S_{XX}^{-1} \dots (nw \times n) (n \times n) = (nw \times n) \\
 \hat{A}S_{YX} &= (nw \times n) (n \times nw) = (nw \times nw) \\
 \hat{B}\hat{B}^T &= S_{YY} - \hat{A}S_{XY} \dots (nw \times nw) - (nw \times nw) = (nw \times nw)
 \end{aligned}$$

แบบจำลองแยกส่วนพื้นฐานแบบเต็มรูป

$$Y = AX + B\underline{\varepsilon}$$

$$\begin{bmatrix} y(1, v, 1) \\ y(1, v, 2) \\ y(1, v, 3) \\ y(2, v, 1) \\ y(2, v, 2) \\ y(2, v, 3) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{bmatrix} \begin{bmatrix} x(1, v) \\ x(2, v) \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{bmatrix} \begin{bmatrix} \varepsilon(1, v, 1) \\ \varepsilon(1, v, 2) \\ \varepsilon(1, v, 3) \\ \varepsilon(2, v, 1) \\ \varepsilon(2, v, 2) \\ \varepsilon(2, v, 3) \end{bmatrix}$$

8.7.3 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบขยาย (Extended Model)

แบบจำลองแยกส่วนแบบขยาย (Extended Model) (Mejia and Rousselle, 1976)

$$Y = AX + B\underline{\varepsilon} + CZ$$

$(nw \times 1) = (nw \times n)(n \times 1) + (nw \times n)(n \times 1) + (nw \times n)(n \times 1) \dots$ ขนาดเมทริกซ์

$$E(Y) = 0$$

$$E(X) = 0$$

$$E(Z) = 0$$

สมการสำหรับคำนวณหาพารามิเตอร์ของแบบจำลองแยกส่วนแบบขยาย คือ

$$\hat{A} = (S_{YX} - S_{YZ}S_{ZZ}^{-1}S_{ZX})(S_{XX} - S_{XZ}S_{ZZ}^{-1}S_{ZX})^{-1} \quad [8.14]$$

$$\hat{C} = (S_{YZ} - \hat{A}S_{XZ})S_{ZZ}^{-1} \quad [8.15]$$

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A}S_{XX}\hat{A}^T - \hat{A}S_{XZ}\hat{C}^T - \hat{C}S_{ZX}\hat{A}^T - \hat{C}S_{ZZ}\hat{C}^T \quad [8.16]$$

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A}S_{XY} - \hat{C}S_{ZY} \quad [8.17]$$

การ Derive ค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบขยาย (Extended Model)

$$Y = AX + B\underline{\varepsilon} + CZ$$

$$E(XY^T) = E \left[X(AX + B\underline{\varepsilon} + CZ)^T \right] = E(XX^T A^T + X\underline{\varepsilon}^T B^T + XZ^T C^T)$$

$$S_{XY} = S_{XX}\hat{A}^T + S_{XZ}\hat{C}^T$$

$$E(ZY^T) = E \left[Z(AX + B\underline{\varepsilon} + CZ)^T \right] = E(ZX^T A^T + Z\underline{\varepsilon}^T B^T + ZZ^T C^T)$$

$$S_{ZY} = S_{ZX}\hat{A}^T + S_{ZZ}\hat{C}^T$$

$$E(YZ^T) = E(AX + B\underline{\varepsilon} + CZ)Z^T$$

$$E(YZ_T) = E(AXZ^T + B\underline{\varepsilon}Z^T + CZZ^T)$$

$$S_{YZ} = \hat{A}S_{XZ} + \hat{C}S_{ZZ}$$

$$\hat{C} = (S_{YZ} - \hat{A}S_{XZ})S_{ZZ}^{-1}$$

$$E(YX^T) = E(AXX^T + B\underline{\varepsilon}Z^T + CZX^T)$$

$$\begin{aligned} S_{YX} &= \hat{A}S_{XX} + \hat{C}S_{ZX} = \hat{A}S_{XX} + (S_{YZ} - \hat{A}S_{XZ})S_{ZZ}^{-1}S_{ZX} \\ &= \hat{A}S_{XX} + S_{YZ}S_{ZZ}^{-1}S_{ZX} - \hat{A}S_{XZ}S_{ZZ}^{-1}S_{ZX} \end{aligned}$$

$$\hat{A} = (S_{YX} - S_{YZ}S_{ZZ}^{-1}S_{ZX})(S_{XX} - S_{XZ}S_{ZZ}^{-1}S_{ZX})^{-1}$$

$$\begin{aligned} E(YY^T) &= E(AX + B\underline{\varepsilon} + CZ)Y^T = E(AXY^T + B\underline{\varepsilon}Y^T + CZY^T) \\ &= E \left(AXY^T + B\underline{\varepsilon}(AX + B\underline{\varepsilon} + CZ)^T + CZY^T \right) \end{aligned}$$

$$S_{YY} = \hat{A}S_{XY} + \hat{B}\hat{B}^T + \hat{C}S_{ZY}$$

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A}S_{XY} - \hat{C}S_{ZY}$$

8.7.4 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบย่อ (Condensed Model)

(Lane, 1979)

$$Y_t = A_t X + B_t \underline{\varepsilon} + C_t Y_{t-1}$$

$$(n \times 1) = (n \times n)(n \times 1) + (n \times n)(n \times 1) + (n \times n)(n \times 1) \dots \dots \dots \text{ขนาดเมทริกซ์}$$

$$\begin{aligned} \hat{A}_t &= [S_{YX}(t, t) - S_{YY}(t, t-1)S_{YY}^{-1}(t-1, t-1)S_{YX}(t-1, t)] \\ &\quad [S_{XX}(t, t) - S_{XY}(t, t-1)S_{YY}^{-1}(t-1, t-1)S_{YX}(t-1, t)]^{-1} \end{aligned} \quad [8.18]$$

$$\hat{C}_t = [S_{YY}(t, t-1) - \hat{A}_t S_{XY}(t, t-1)]S_{YY}^{-1}(t-1, t-1) \quad [8.19]$$

$$\hat{B}_t \hat{B}_t^T = S_{YY}(t, t) - \hat{A}_t S_{XY}(t, t) - \hat{C}_t S_{YY}(t-1, t) \quad [8.20]$$

การ Derive หาพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบย่อ (Condensed Model)

$$Y_\tau = A_\tau X + B_\tau \underline{\varepsilon} + C_\tau Y_{\tau-1}$$

$$E(Y_\tau Y_\tau^T) = E(A_\tau X Y_\tau^T + B_\tau \underline{\varepsilon} Y_\tau^T + C_\tau Y_{\tau-1} Y_\tau^T)$$

$$S_{YY}(\tau, \tau - 1) = \hat{A}_\tau S_{XY}(\tau, \tau - 1) + \hat{C}_\tau S_{YY}(\tau - 1, \tau - 1)$$

$$\hat{C}_\tau = (S_{YY}(\tau, \tau - 1) - \hat{A}_\tau S_{XY}(\tau, \tau - 1)) S_{YY}^{-1}(\tau - 1, \tau - 1)$$

$$E(Y_\tau X^T) = E(A_\tau X X^T + B_\tau \underline{\varepsilon} X^T + C_\tau Y_{\tau-1} X^T)$$

$$\begin{aligned} S_{YX}(\tau, \tau) &= \hat{A}_\tau S_{XX}(\tau, \tau) + \hat{C}_\tau S_{YX}(\tau - 1, \tau) \\ &= \hat{A}_\tau S_{XX}(\tau, \tau) + S_{YY}(\tau, \tau - 1) S_{YY}^{-1}(\tau - 1, \tau - 1) S_{YX}(\tau - 1, \tau) \\ &\quad - \hat{A}_\tau S_{XY}(\tau, \tau - 1) S_{YY}^{-1}(\tau - 1, \tau - 1) S_{YX}(\tau - 1, \tau) \end{aligned}$$

$$\begin{aligned} \hat{A}_\tau &= [S_{YX}(\tau, \tau) - S_{YY}(\tau, \tau - 1) S_{YY}^{-1}(\tau - 1, \tau - 1) S_{YX}(\tau - 1, \tau)] \\ &\quad [S_{XX}(\tau, \tau) - S_{XY}(\tau, \tau - 1) S_{YY}^{-1}(\tau - 1, \tau - 1) S_{YX}(\tau - 1, \tau)]^{-1} \end{aligned}$$

$$E(Y_\tau Y_\tau^T) = E(A_\tau X + B_\tau \underline{\varepsilon} + C_\tau Y_{\tau-1}) Y_\tau^T$$

$$\begin{aligned} S_{YY}(\tau, \tau) &= \hat{A}_\tau S_{XY}(\tau, \tau) + \hat{B}_\tau E(\underline{\varepsilon} X^T A^T + \underline{\varepsilon} \underline{\varepsilon}^T B_\tau^T + \underline{\varepsilon} Y_{\tau-1}^T C_\tau^T) + \hat{C}_\tau S_{YY}(\tau - 1, \tau) \\ &= \hat{A}_\tau S_{XY}(\tau, \tau) + \hat{B}_\tau \hat{B}_\tau^T + \hat{C}_\tau S_{YY}(\tau - 1, \tau) \end{aligned}$$

$$\hat{B}_\tau \hat{B}_\tau^T = S_{YY}(\tau, \tau) - \hat{A}_\tau S_{XY}(\tau, \tau) - \hat{C}_\tau S_{YY}(\tau - 1, \tau)$$

8.7.5 การประมาณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามพื้นที่แบบขยายกับกรณี Multisite

$$Y = AX + B\underline{\varepsilon} + CZ$$

กรณีที่มีสถานีหลัก(Key Station) m สถานี และสถานีย่อย n สถานี

$$(n \times 1) = (n \times m)(m \times 1) + (n \times n)(n \times 1) + \dots \dots \dots \text{ขนาดเมทริกซ์}$$

พารามิเตอร์ของแบบจำลองนี้จะคล้ายกับแบบจำลองแยกส่วนตามเวลาแบบขยายคือ

$$\hat{A} = [S_{YX} - S_{YZ} S_{YY}^{-1} S_{XZ}^T] [S_{XX} - S_{XZ} S_{YY}^{-1} S_{XZ}^T]^{-1}$$

$$\hat{C} = [S_{YZ} - \hat{A} S_{XZ}] S_{YY}^{-1}$$

$$\hat{B} \hat{B}^T = S_{YY} - \hat{A} S_{XY} - \hat{C} S_{YZ}^T$$

เนื่องจาก Z คือ Y ของปีก่อนหน้านั้น หรือ $Z = Y(1)$ จะสามารถเขียนสมการพารามิเตอร์ใหม่ได้ ดังนี้

$$\hat{A} = [S_{YX} - S_{YY}(1) S_{YY}^{-1}(1) S_{XY}^T(1)] [S_{XX} - S_{XY}(1) S_{YY}^{-1}(1) S_{XY}^T(1)]^{-1} \quad [8.21]$$

$$\hat{C} = [S_{YY}(1) - \hat{A} S_{XY}(1)] S_{YY}^{-1} \quad [8.22]$$

$$\hat{B} \hat{B}^T = S_{YY} - \hat{A} S_{XY} - \hat{C} S_{YY}^T(1) \quad [8.23]$$

8.7.6 วิธีการคำนวณค่า Variance-Covariance Matrixes สำหรับแบบจำลองแยกส่วนตามเวลาแบบย่อ (Condensed Model)

การคำนวณค่าพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาแบบย่อในสมการ 8.18-8.20 จะเกี่ยวข้องกับ การประมาณค่า Variance-Covariance Matrices ของ $S_{YY}(\tau, \tau)$, $S_{YX}(\tau, \tau)$, $S_{XX}(\tau, \tau)$, $S_{YY}(\tau, \tau-1)$, $S_{YX}(\tau-1, \tau)$ เมื่อ τ คือ ฤดูกาล เช่น เดือน

ตามหลักโมเมนต์

$$S_{YY}(\tau-1, \tau-1) = S_{YY}(\tau, \tau) \text{ ของเดือนก่อนหน้านั้น และ}$$

$$S_{XY}(\tau, \tau-1) = S_{YX}^T(\tau-1, \tau)$$

กรณี $n = 2$

$$S_{YY}(\tau, \tau) = \frac{1}{N-1} \sum_{V=1}^N \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v,\tau}^{(1)} & y_{v,\tau}^{(2)} \end{bmatrix} \quad [8.24]$$

$$S_{YX}(\tau, \tau) = \frac{1}{N-1} \sum_{V=1}^N \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad [8.25]$$

$$S_{XX}(\tau, \tau) = \frac{1}{N-1} \sum_{V=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad [8.26]$$

$$S_{YY}(\tau, \tau-1) = \frac{1}{N-1} \sum_{V=1}^N \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v,\tau-1}^{(1)} & y_{v,\tau-1}^{(2)} \end{bmatrix} \quad [8.27]$$

$$S_{YX}(\tau-1, \tau) = \frac{1}{N-1} \sum_{V=1}^N \begin{bmatrix} y_{v,\tau-1}^{(1)} \\ y_{v,\tau-1}^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad [8.28]$$

$S_{XX}(\tau, \tau)$ มีค่าเหมือนกันทุกเดือน เนื่องจาก X คือค่าของอนุกรมเวลารายปี

8.7.7 วิธีการคำนวณค่า Variance-Covariance Matrixes สำหรับแบบจำลองแยกส่วนตามพื้นที่ (Spatial)

สำหรับแบบจำลองแยกส่วนตามพื้นที่ X, Y, Z คืออนุกรมเวลารายปี ถ้ากำหนดให้ มีสถานีหลัก (X) เท่ากับ $m=2$ สถานี และสถานีย่อย (Y) เท่ากับ $n = 4$ สถานี จะสามารถคำนวณหาค่า Variance-Covariance Matrices S_{YY} , S_{XX} , S_{YX} , $S_{YY}(1)$ และ $S_{XY}(1)$ ได้ดังนี้

$$S_{YY} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_v^{(1)} \\ y_v^{(2)} \\ y_v^{(3)} \\ y_v^{(4)} \end{bmatrix} \begin{bmatrix} y_v^{(1)} & y_v^{(2)} & y_v^{(3)} & y_v^{(4)} \end{bmatrix} \quad [8.29]$$

$$S_{XX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad [8.30]$$

$$S_{YX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_v^{(1)} \\ y_v^{(2)} \\ y_v^{(3)} \\ y_v^{(4)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad [8.31]$$

$$S_{YY}(1) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_v^{(1)} \\ y_v^{(2)} \\ y_v^{(3)} \\ y_v^{(4)} \end{bmatrix} \begin{bmatrix} y_{v-1}^{(1)} & y_{v-1}^{(2)} & y_{v-1}^{(3)} & y_{v-1}^{(4)} \end{bmatrix} \quad [8.32]$$

$$S_{XY}(1) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} y_{v-1}^{(1)} & y_{v-1}^{(2)} & y_{v-1}^{(3)} & y_{v-1}^{(4)} \end{bmatrix} \quad [8.33]$$

8.8 การทดสอบความเหมาะสมของแบบจำลอง (Goodness of Fit Test)

วิธีการทดสอบความเหมาะสมของแบบจำลองแยกส่วน จะเหมือนกับวิธีการทดสอบความเหมาะสมของแบบจำลอง AR, ARMA, ARIMA และ Multivariate ตามที่กล่าวมาแล้ว ซึ่งประกอบด้วย

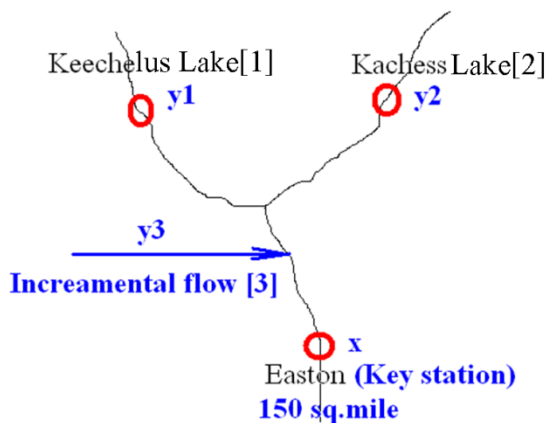
- (1) การทดสอบสมมติฐานความเป็นอิสระและการแจกแจงแบบปกติของ Residual Series
- (2) การเปรียบเทียบ Model Correlogram, Generated Correlogram กับ Historical Correlogram
- (3) การทดสอบ Parsimony ของพารามิเตอร์

8.9 ตัวอย่างการจำลองแบบแยกส่วน

กลุ่มน้ำ Yakima ดังรูปที่ 8.3 มีอนุกรมเวลาน้ำท่ารายเดือน 4 สถานีคือ (1) Keechelus Lake (2) Kachess Lake (3) Incremental Flow และ (4) Easton Diversion Dam ดังแสดงในตารางที่ 8.2 กำหนดให้ Easton เป็น

สถานีหลัก (Key Station) และ Keechelus, Kachess, Incremental Flow เป็นสถานีย่อย (Sub-stations) ต้องการ
สร้างแบบจำลองเพื่อสังเคราะห์อนุกรมเวลารายเดือนของ 4 สถานีในกลุ่มน้ำ Yakima ดังนี้

- (1) สังเคราะห์อนุกรมเวลารายปีของสถานีหลัก Easton Diversion Dam
- (2) สร้างแบบจำลองแยกส่วนตามพื้นที่ (Spatial Disaggregation Model) เพื่อสังเคราะห์อนุกรมเวลารายปีของสถานีย่อย (Keechelus Lake, Kachess Lake และ Incremental Flow) จากอนุกรมเวลาของสถานีหลัก
- (3) สร้างแบบจำลองแยกส่วนตามเวลา (Multisite Temporal Disaggregation Model) เพื่อสังเคราะห์อนุกรมเวลารายเดือนของสถานีย่อยจากอนุกรมเวลารายปี



รูปที่ 8.3 สถานีหลักและสถานีย่อยในกลุ่มน้ำ Yakima

ตารางที่ 8.2 อนุกรมเวลาตัวอย่างของสถานีหลักและสถานีย่อยของกลุ่มน้ำ Yakima ระหว่างปี 1926-1940

		Year															
		1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	
Annual data	Keechelus	Y(1)	183.1	234.4	251.2	156.2	160.4	176.6	278.5	345.7	321.6	248.8	219.7	201.1	215.9	213.6	186.1
	Kachees	Y(2)	158.1	220.3	233.6	134.7	134.8	152.0	240.2	303.7	304.5	223.5	207.3	174.3	192.3	183.2	153.5
	Difference	Y(3)	126.1	184.6	227.1	131.7	132.1	108.5	188.1	264.6	275.5	233.2	207.1	142.3	190.5	170.3	110.6
	Easton	X	467.3	639.3	711.9	422.6	427.3	437.1	706.8	914.0	901.6	705.5	634.1	517.7	598.7	567.1	450.2
May Data	Keechelus	Y5(1)	24.7	7.9	21.5	11.3	18.2	21.9	32.2	8.4	47.7	14.5	15.1	12.0	9.7	14.1	22.8
	Kachees	Y5(2)	23.3	7.1	20.9	11.9	19.1	21.8	34.0	7.4	46.9	15.7	15.9	11.3	10.3	15.0	20.2
	Difference	Y5(3)	22.2	12.0	22.4	11.5	20.5	13.6	28.9	10.4	39.3	17.8	47.6	11.8	13.3	14.8	15.4
	Easton	X5	70.2	27.0	64.8	34.7	57.8	57.3	95.1	26.2	133.9	48.0	16.6	35.1	33.3	43.9	58.4
June Data	Keechelus	Y6(1)	30.3	19.1	21.0	15.3	36.9	26.1	32.4	18.1	48.2	15.0	34.9	22.4	34.7	29.2	3.4
	Kachees	Y6(2)	29.3	22.4	22.1	16.0	34.5	24.3	34.4	20.0	47.6	16.8	38.1	21.0	33.0	30.1	30.4
	Difference	Y6(3)	19.8	25.4	22.5	15.4	36.7	24.6	35.9	19.5	29.2	19.7	42.7	21.4	28.9	26.7	20.0
	Easton	X6	79.3	66.9	65.6	46.7	108.1	75.0	102.7	57.6	125.0	51.5	115.0	64.8	96.6	86.0	83.7

ขั้นที่ 1 การสังเคราะห์อนุกรมเวลาน้ำท่ารายปีของสถานีหลัก (Easton Diversion Dam) โดย
แบบจำลอง AR(1)

$$y_t = 0.568y_{t-1} + 132\varepsilon_t$$

เมื่อ y_t = Annual Flow – Mean Annual Flow

ขั้นที่ 2 การสังเคราะห์อนุกรมเวลาน้ำท่ารายปีของสถานีย่อยจากอนุกรมเวลาน้ำท่ารายปีของสถานีหลัก โดยใช้แบบจำลองแยกส่วนตามพื้นที่แบบขยาย (Extended Spatial Disaggregation Model) ในสมการ [8.7]

$$Y = AX + B\underline{\varepsilon} + CZ$$

เมื่อ

Y=Column Matrix ของอนุกรมเวลาน้ำท่ารายปีปัจจุบันของ 3 สถานีย่อย

X= Column Matrix ของอนุกรมเวลาน้ำท่ารายปีปัจจุบันของสถานีหลัก

Z= Column Matrix ของอนุกรมเวลาน้ำท่ารายปีของปีก่อนหน้านั้นของ 3 สถานีย่อย

คำนวณหาค่าพารามิเตอร์เมทริกซ์ \hat{A} , \hat{B} , \hat{C} จากสมการ [8.14], [8.15] และ [8.17]

$$\hat{A} = (S_{YX} - S_{YZ}S_{ZZ}^{-1}S_{ZX})(S_{XX} - S_{XZ}S_{ZZ}^{-1}S_{ZX})^{-1}$$

$$\hat{C} = (S_{YZ} - \hat{A}S_{XZ})S_{ZZ}^{-1}$$

$$\hat{B}\hat{B}^T = S_{YY} - \hat{A}S_{XY} - \hat{C}S_{ZY}$$

ผลการคำนวณ Variance-Covariance Matrices และค่าพารามิเตอร์ คือ

$$S_{YY} = \begin{bmatrix} 3101 & 3002 & 2701 \\ 3002 & 2977 & 2762 \\ 2701 & 2762 & 2877 \end{bmatrix}$$

$$S_{YY}(1) = \begin{bmatrix} 1844 & 1609 & 861.5 \\ 1967 & 1760 & 1038 \\ 2168 & 2037 & 1416 \end{bmatrix}$$

$$S_{XX} = [25880]$$

$$S_{YX} = \begin{bmatrix} 8803 \\ 8741 \\ 8340 \end{bmatrix}$$

$$S_{YX}(1) = [5980 \quad 5407 \quad 3316]$$

$$\hat{A} = \begin{bmatrix} 0.3532 \\ 0.3372 \\ 0.3096 \end{bmatrix}; \text{column sum of } \hat{A} = 1.00$$

$$\hat{C} = \begin{bmatrix} 0.5112 & -0.6510 & 0.03759 \\ 0.1707 & -0.1734 & -0.02162 \\ -0.6819 & 0.8244 & -0.01597 \end{bmatrix}; \text{column sum of } \hat{C} = [0 \quad 0 \quad 0]$$

$$\hat{B}\hat{B}^T = \begin{bmatrix} 64.29 & 15.50 & -79.79 \\ 15.50 & 21.10 & -36.60 \\ -79.79 & -36.60 & 116.4 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 8.018 & 0 & 0 \\ 1.933 & 4.167 & 0 \\ -9.951 & -4.167 & 0 \end{bmatrix}; \text{column sum of } \hat{B} = [0 \quad 0 \quad 0]$$

ตัวอย่างการคำนวณ Variance-Covariance Matrices ด้วยฟังก์ชัน SUMPRODUCT ในตารางที่ 8.3

ตารางที่ 8.3 Variance-Covariance Matrices สำหรับแบบจำลองแยกส่วนตามพื้นที่แบบขยาย

Year (t)	Annual data(t)				Annual Data(t-1)			
	Keechelus Y1(t)	Kachees Y2(t)	Difference Y3(t)	Easton X(t)	Keechelus Y1(t-1)	Kachees Y2(t-1)	Difference Y3(t-1)	Easton X(t-1)
1926	-43.1	-43.6	-52.7	-139.4	-40.1	-48.2	-68.2	-156.5
1927	8.2	18.6	5.8	32.6	-43.1	-43.6	-52.7	-139.4
1928	25.0	31.9	48.3	105.2	8.2	18.6	5.8	32.6
1929	-70.0	-67.0	-47.1	-184.1	25.0	31.9	48.3	105.2
1930	-65.8	-66.9	-46.7	-179.4	-70.0	-67.0	-47.1	-184.1
1931	-49.6	-49.7	-70.3	-169.6	-65.8	-66.9	-46.7	-179.4
1932	52.3	38.5	9.3	100.1	-49.6	-49.7	-70.3	-169.6
1933	119.5	102.0	85.8	307.3	52.3	38.5	9.3	100.1
1934	95.4	102.8	96.7	294.9	119.5	102.0	85.8	307.3
1935	22.6	31.5	44.7	98.8	95.4	102.8	96.7	294.9
1936	-6.5	5.6	28.3	27.4	22.6	31.5	44.7	98.8
1937	-25.1	-27.4	-36.5	-89.0	-6.5	5.6	28.3	27.4
1938	-10.3	-9.4	11.7	-8.0	-25.1	-27.4	-36.5	-89.0
1939	-12.6	-18.5	-8.5	-39.6	-10.3	-9.4	11.7	-8.0
1940	-40.1	-48.2	-68.2	-156.5	-12.6	-18.5	-8.5	-39.6
Mean	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Y1(t)	Y2(t)	Y3(t)	X(t)	Y1(t-1)	Y2(t-1)	Y3(t-1)	X(t-1)
Y1(t)	3101	3002	2701	8803	1844	1609	861	4315
Y2(t)	3002	2977	2762	8741	1967	1760	1038	4765
Y3(t)	2701	2762	2877	8340	2168	2037	1416	5621
X(t)	8803	8741	8340	25884	5980	5407	3316	14702

ขั้นที่ 3 การสังเคราะห์อนุกรมเวลาน้ำท่ารายเดือนจากอนุกรมเวลาน้ำท่ารายปีของสถานีย่อย โดยใช้แบบจำลองแยกส่วนตามเวลาแบบย่อ (Condensed Form Multisite Temporal Disaggregation Model)

$$Y_t = A_t X + B_t \varepsilon + C_t Y_{t-1}$$

คำนวณหาค่าพารามิเตอร์เมทริกซ์ \hat{A}_t , \hat{B}_t , \hat{C}_t จากสมการ [8.18], [8.19] และ [8.20]

$$\hat{A}_t = [S_{YX}(t, t) - S_{YY}(t, t-1)S_{YY}^{-1}(t-1, t-1)S_{YX}(t-1, t)]$$

$$[S_{XX}(t, t) - S_{XY}(t, t-1)S_{YY}^{-1}(t-1, t-1)S_{YX}(t-1, t)]^{-1}$$

$$\hat{C}_t = [S_{YY}(t, t-1) - \hat{A}_t S_{XY}(t, t-1)]S_{YY}^{-1}(t-1, t-1)$$

$$\hat{B}_t \hat{B}_t^T = S_{YY}(t, t) - \hat{A}_t S_{XY}(t, t) - \hat{C}_t S_{YY}(t-1, t)$$

ผลการคำนวณ Variance-Covariance Matrices และค่าพารามิเตอร์เดือนมิถุนายน(เดือนที่ 6) คือ

$$S_{XY}(6,5) = \begin{bmatrix} 174.3 & 171.9 & 174.0 \\ 175.0 & 173.7 & 182.2 \\ 111.3 & 119.1 & 163.7 \end{bmatrix}$$

$$S_{YY}(6,5) = \begin{bmatrix} 67.49 & 67.13 & 48.38 \\ 62.46 & 63.01 & 47.61 \\ 18.92 & 23.44 & 21.12 \end{bmatrix}$$

$$S_{YY}(5,5) = \begin{bmatrix} 110.7 & 109.3 & 76.57 \\ 109.3 & 109.4 & 47.61 \\ 76.57 & 77.23 & 60.76 \end{bmatrix}$$

$$S_{YX}(6,6) = \begin{bmatrix} 37.56 & 38.15 & 33.47 \\ 91.86 & 96.61 & 95.29 \\ 28.21 & 37.76 & 63.73 \end{bmatrix}$$

$$S_{XX}(6,6) = \begin{bmatrix} 3101 & 3002 & 2701 \\ 3002 & 2977 & 2762 \\ 2701 & 2762 & 2877 \end{bmatrix}$$

$$S_{YY}(6,6) = \begin{bmatrix} 87.56 & 79.81 & 45.08 \\ 79.81 & 76.27 & 47.87 \\ 45.08 & 47.87 & 58.20 \end{bmatrix}$$

$$\hat{A} = \begin{bmatrix} 0.22130 & -0.4229 & 0.17680 \\ 0.12260 & -0.2549 & 0.12860 \\ -0.05735 & 0.7423 & -0.02786 \end{bmatrix}$$

$$\hat{C} = \begin{bmatrix} 1.4160 & -0.8238 & 0.2171 \\ 0.5396 & -0.1149 & 0.3167 \\ -3.133 & 2.9610 & 0.5494 \end{bmatrix}$$

$$\hat{B}\hat{B}^T = \begin{bmatrix} 38.70 & 36.61 & 31.47 \\ 36.61 & 35.84 & 31.64 \\ 31.47 & 31.64 & 37.07 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 6.221 & 0 & 0 \\ 5.884 & 1.103 & 0 \\ 5.059 & 1.697 & 2.934 \end{bmatrix}$$

ตัวอย่างการคำนวณ Variance-Covariance Matrices ด้วยฟังก์ชัน SUMPRODUCT ในตารางที่ 8.4

ตารางที่ 8.4 Variance-Covariance Matrices สำหรับแบบจำลองแยกส่วนตามเวลาแบบย่อ

Year (t)	Annual data				May Data				June Data			
	Y1(t)	Y2(t)	Y3(t)	X(t)	Y1(t,5)	Y2(t,5)	Y3(t,5)	X(t,5)	Y1(t,6)	Y2(t,6)	Y3(t,6)	X(t,6)
1926	-43.1	-43.6	-52.7	-139.4	5.9	4.6	4.2	14.6	2.5	1.3	-6.1	-2.4
1927	8.2	18.6	5.8	32.6	-10.9	-11.6	-6.0	-28.6	-8.7	-5.6	-0.5	-14.8
1928	25.0	31.9	48.3	105.2	2.7	2.2	4.4	9.2	-6.8	-5.9	-3.4	-16.1
1929	-70.0	-67.0	-47.1	-184.1	-7.5	-6.8	-6.5	-20.9	-12.5	-12.0	-10.5	-35.0
1930	-65.8	-66.9	-46.7	-179.4	-0.6	0.4	2.5	2.2	9.1	6.5	10.8	26.4
1931	-49.6	-49.7	-70.3	-169.6	3.1	3.1	-4.4	1.7	-1.7	-3.7	-1.3	-6.7
1932	52.3	38.5	9.3	100.1	13.4	15.3	10.9	39.5	4.6	6.4	10.0	21.0
1933	119.5	102.0	85.8	307.3	-10.4	-11.3	-7.6	-29.4	-9.7	-8.0	-6.4	-24.1
1934	95.4	102.8	96.7	294.9	28.9	28.2	21.3	78.3	20.4	19.6	3.3	43.3
1935	22.6	31.5	44.7	98.8	-4.3	-3.0	-0.2	-7.6	-12.8	-11.2	-6.2	-30.2
1936	-6.5	5.6	28.3	27.4	-3.7	-2.8	-1.4	-8.0	7.1	10.1	16.8	34.0
1937	-25.1	-27.4	-36.5	-89.0	-6.8	-7.4	-6.2	-20.5	-5.4	-7.0	-4.5	-16.9
1938	-10.3	-9.4	11.7	-8.0	-9.1	-8.4	-4.7	-22.3	6.9	5.0	3.0	14.9
1939	-12.6	-18.5	-8.5	-39.6	-4.7	-3.7	-3.2	-11.7	1.4	2.1	0.8	4.3
1940	-40.1	-48.2	-68.2	-156.5	4.0	1.5	-2.6	2.8	5.6	2.4	-5.9	2.0
Mean	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Use SUMPRODUCT(A,B)/(N-1)												
	Y1(t)	Y2(t)	Y3(t)	X(t)	Y1(t,5)	Y2(t,5)	Y3(t,5)	X(t,5)	Y1(t,6)	Y2(t,6)	Y3(t,6)	X(t,6)
Y1(t)	3101	3002	2701	8803	174.3	171.9	174.0	520.2	37.56	91.26	28.21	157.6
Y2(t)	3002	2977	2762	8741	175.0	173.7	182.2	531.0	38.15	95.95	37.76	172.5
Y3(t)	2701	2762	2877	8340	111.3	119.1	163.7	394.1	33.47	94.42	63.73	192.5
X(t)	8803	8741	8340	25884	460.6	464.7	519.9	1445.2	109.2	281.6	129.7	522.6
Y1(t,5)	174.3	175.0	111.3	460.6	110.67	109.34	76.57	296.6	67.5	62.5	18.9	148.9
Y2(t,5)	171.9	173.7	119.1	464.7	109.34	109.41	77.23	296.0	67.1	63.0	23.4	153.6
Y3(t,5)	174.0	182.2	163.7	519.9	76.57	77.23	60.76	214.6	48.4	47.6	21.1	117.1
X(t,5)	520.2	531.0	394.1	1445.2	296.6	296.0	214.6	807.1	183.0	173.2	63.5	419.6
Y1(t,6)	37.6	38.2	33.5	109.2	67.49	67.13	48.38	183.0	87.56	79.87	45.08	212.5
Y2(t,6)	91.3	96.0	94.4	281.6	62.53	63.05	47.62	173.2	79.87	76.32	47.78	203.9
Y3(t,6)	28.2	37.8	63.7	129.7	18.92	23.44	21.12	63.5	45.08	47.78	58.20	151.1
X(t,6)	157.6	172.5	192.5	522.6	148.9	153.6	117.1	419.6	212.5	203.9	151.1	567.6

8.10 ประโยชน์และข้อจำกัดของการจำลองแยกส่วน (Usefulness and Limitations of Disaggregation Modeling)

แบบจำลองแบบแยกส่วนมีข้อดีและประโยชน์หลายประการ อาทิเช่น

- (1) การสังเคราะห์อนุกรมเวลารายฤดูกาลจากอนุกรมเวลารายปี (Annual to Seasonal Disaggregation) ซึ่งเป็นไปตามธรรมชาติของกระบวนการ ที่ผลรวมของข้อมูลรายฤดูกาลจะเท่ากับรายปี และค่าสถิติรายปีและรายเดือนมีความสำคัญต่อการศึกษาทางอุทกวิทยา
- (2) การสังเคราะห์อนุกรมเวลาของสถานีย่อยจากอนุกรมเวลาของสถานีหลัก (Key to Sub-station Disaggregation) ซึ่งสอดคล้องกับความเป็นจริงทางด้านอุทกวิทยาในกรณีที่ผลรวมของอนุกรมเวลาของสถานีย่อยเท่ากับอนุกรมเวลาของสถานีหลักหรือมีความสัมพันธ์เชิงเส้นกับอนุกรมเวลาของสถานีหลัก
- (3) กรณีที่อนุกรมเวลาที่มีความยาวไม่เท่ากัน ซึ่งปกติอนุกรมเวลาของสถานีหลักจะยาวกว่า จึงมีประโยชน์ในการต่อขยายข้อมูลของสถานีย่อย

- (4) การจำลองแบบแยกส่วนจะช่วยลดจำนวนพารามิเตอร์ จึงมีผลดีต่อการทดสอบ Parsimony ของพารามิเตอร์

ข้อเสียและข้อจำกัดของแบบจำลองแบบแยกส่วน คือ

- (1) การใช้แบบจำลองแบบแยกส่วนไม่เหมาะกับอนุกรมเวลาที่มีช่วงเวลาที่สั้น เช่นรายวัน ซึ่งทำให้ต้องแยกเหตุการณ์ออกจากกัน เช่นพายุฝนก่อให้เกิดฝนตกต่อเนื่อง 3 วัน การวิเคราะห์รายวันจะต้องแยกฝนที่เกิดจากพายุออกเป็น 3 เหตุการณ์
- (2) การจำลองแบบแยกส่วน วิเคราะห์โดยใช้เมทริกซ์ ซึ่งต้องเขียนโปรแกรมคอมพิวเตอร์ช่วยในการคำนวณที่ซับซ้อน
- (3) การประมาณค่าพารามิเตอร์โดยวิธีโมเมนต์ถึงแม้จะทำได้ง่าย แต่ค่าที่ได้มัก Biased และต้องมีการปรับค่าให้เป็น Unbiased ซึ่งวิธีการปรับค่าให้เป็น Unbiased สำหรับแบบจำลองแบบแยกส่วนยังไม่ค่อยได้มีการพัฒนา
- (4) กรณีที่จำนวนสถานีมาก อาจมีปัญหาด้าน Parsimony ของพารามิเตอร์ได้

เกณฑ์การใช้แบบจำลองแบบแยกส่วน

- (1) จำนวนพารามิเตอร์ต้องเท่ากับจำนวน โมเมนต์ที่ใช้ประมาณค่าพารามิเตอร์
- (2) จำนวนค่าสังเกต Y ต้องมากกว่าจำนวน โมเมนต์ และมากกว่าจำนวนพารามิเตอร์
- (3) โครงสร้างของแบบจำลองต้องสอดคล้องกับ โมเมนต์ที่ต้องการรักษาไว้ (Preserved)
- (4) โมเมนต์ที่แบบจำลองต้องการรักษาไว้ต้องสอดคล้องกับ โมเมนต์ของอนุกรมเวลาตัวอย่างที่นำไปใช้ประมาณค่าพารามิเตอร์
- (5) จำนวน Stochastic Term ต้องเท่ากับจำนวนเทอมของเมทริกซ์ Y

ตารางที่ 8.5 แสดงการเปรียบเทียบจำนวนพารามิเตอร์พารามิเตอร์ของแบบจำลองแยกส่วนตามเวลาที่ทั้ง 3 แบบ ซึ่งพบว่าจะมีปัญหาในการสังเคราะห์อนุกรมเวลารายฤดูกาลจากอนุกรมเวลารายปี ถ้าจำนวนสถานีมาก เช่น $n=20$ การสังเคราะห์อนุกรมเวลารายเดือนต้องใช้พารามิเตอร์หลายหมื่นตัว

ตารางที่ 8.6 แสดงปัญหาเกี่ยวกับ Parsimony ถ้ามีอัตราส่วนระหว่างจำนวนข้อมูลตัวอย่างต่อจำนวนพารามิเตอร์ (R) น้อยกว่า 10

ตารางที่ 8.5 จำนวนพารามิเตอร์ของแบบจำลองแยกส่วนตามเวลา

จำนวนสถานี(n)	จำนวนฤดูกาล(w)	Basic Model	Extended Model	Condensed Model
1	12	156	168	36
1	24	600	624	72
1	52	2,756	2,808	156
2	12	624	672	144
2	24	2,400	2,496	288
3	12	1,404	1,512	324
20	12	62,400	67,200	14,400
n	w	$n^2w+n^2w^2$	$2n^2w+n^2w^2$	$3n^2w$
n	12	$156n^2$	$168n^2$	$36n^2$
1	w	$w+w^2$	$2w+w^2$	$3w$
2	w	$4w+4w^2$	$8w+4w^2$	$12w$
n Small	w Large	$\sim w^2$	$\sim w^2$	$\sim w^2$
n Large	w Small	$\sim n^2$	$\sim n^2$	$\sim n^2$

ตารางที่ 8.6 แนวทางการพิจารณา Parsimony ของพารามิเตอร์

อัตราส่วนจำนวนข้อมูลตัวอย่างต่อจำนวนพารามิเตอร์ (R)	ข้อพิจารณา
$R < 1$	Impossible
$1 < R < 3$	Foolish
$3 \leq R < 5$	Poor
$5 \leq R < 10$	Fair
$10 \leq R < 20$	Good
$20 \leq R$	Very Good

8.11 ตัวอย่างการประยุกต์ใช้แบบจำลองแยกส่วน (Practical Applications of Disaggregation Models)

Valencia และ Schaake(1973) ใช้แบบจำลองแยกส่วนใน 3 โครงการคือ

- (1) การสังเคราะห์หอนุกรมเวลาฝนใน Puerto Rico
- (2) การสังเคราะห์หอนุกรมเวลาน้ำท่าของแม่น้ำโคโลราโด และ
- (3) การจำลองความต้องการน้ำของระบบประปาเมืองบอสตัน

Curry และ Bras(1978) ใช้แบบจำลองแยกส่วนในการสังเคราะห์น้ำท่ารายเดือนของแม่น้ำไนล์ โดยใช้เทคนิค Broken Line ในการสังเคราะห์หอนุกรมเวลารายปี แล้วจึงใช้แบบจำลองแยกส่วนระดับที่ 1 เพื่อแยกหอนุกรมเวลารายปีเป็นรายครึ่งปี ระดับที่ 2 เพื่อแยกหอนุกรมเวลารายครึ่งปีเป็นราย 4 เดือน ระดับที่ 3 เพื่อแยกหอนุกรมเวลาราย 4 เดือนเป็นรายเดือน

8.12 เอกสารอ้างอิง

Salas,J., J.W.Delleur, V.Yevjevich and W.L.Lane. 1980. Applied Modeling of Hydrologic Time Series, Water Resources Publications, USA. 484p.

8.13 ภาคผนวก

APPENDIX A8.1 ESTIMATION OF COVARIANCES

The main object of this appendix is to illustrate the estimation procedure for the covariance matrices used in the disaggregation models presented in this chapter.

Two approaches may be taken to estimate covariance matrices. One is to estimate the elements of the matrix individually using standard estimation techniques. The second is to use a matrix approach which estimates all of the covariances at one time. This second approach will be used here. The estimation of covariances will be illustrated for the three temporal disaggregation schemes and for one spatial disaggregation scheme presented in this chapter. A simple example will be used to clarify the procedure.

For the temporal disaggregation models, the following notation will be used. Consider the periodic (seasonal) data in question to be denoted as $y_{v,\tau}^{(i)}$, $i = 1, \dots, n$; $v = 1, \dots, N$; and $\tau = 1, \dots, w$; with i referring to one of the individual sites or time series, v being the year, and τ being the interval (season) of the year. The limits n , N and w are respectively the total number of times series, the total number of years of data, and the total number of intervals (season) in the year. The annual data is denoted as $x_v^{(i)}$ with i and v taking on the same meaning as before. Further consider that the means have been subtracted from the original data so that the data now have zero means. Note that this is done on a season by season and site by site basis. There are therefore n annual means and w seasonal means involved. For ease of presentation, the estimation procedure for the temporal models will be illustrated for the case of $n = 2$ and $w = 3$.

COVARIANCES ESTIMATES FOR THE BASIC TEMPORAL DISAGGREGATION MODEL

The estimation of parameter matrices A and B as given by Eqs. (8.14) and (8.15) require the sample covariance matrices S_{XX} , S_{YX} and S_{YY} . The population value corresponding to S_{XX} may be expressed as

$$E(XX^T) = E \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (A8.1)$$

where $E(\cdot)$ is an expected value operator. The sample estimate S_{XX} , is obtained from

$$S_{XX} = \frac{1}{N} \sum_{v=1}^N (X_v X_v^T) \quad (\text{A8.2})$$

or

$$S_{XX} = \frac{1}{N} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.3})$$

Equation (A8.3) is adjusted slightly to correct for the bias introduced in the covariance because of the need for estimating the means. The adjustment amounts to reducing the denominator of Eqs. (A8.2) and (A8.3) by one, resulting in

$$S_{XX} = \frac{1}{N-1} \sum_{v=1}^N (X_v X_v^T) \quad (\text{A8.4})$$

and

$$S_{XX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.5})$$

These "unbiased" forms are used in this chapter. The reader should note that many authorities use the biased estimation equations. The alternative approach, estimating the individual elements of S_{XX} separately, may be written as

$$s_{XX}^{(i,j)} = \frac{1}{N-1} \sum_{v=1}^N x_v^{(i)} x_v^{(j)} \quad (\text{A8.6})$$

where $s_{XX}^{(i,j)}$ is the element of S_{XX} from the i -th row and j -th column.

The covariance matrix S_{YX} , made up of the covariances between the seasonal and annual series, is given by

$$S_{YX} = \frac{1}{N-1} \sum_{v=1}^N (Y_v X_v^T) \quad (\text{A8.7})$$

or

$$S_{YX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v,1}^{(1)} \\ y_{v,2}^{(1)} \\ y_{v,3}^{(1)} \\ y_{v,1}^{(2)} \\ y_{v,2}^{(2)} \\ y_{v,3}^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.8})$$

Again, the individual elements can be equivalently estimated as

$$s_{YX}(i,j) = \frac{1}{N-1} \sum_{v=1}^N y_{v,\tau}^{(k)} x_v^{(j)} \quad (\text{A8.9})$$

where

$$i = \omega k + \tau \quad (\text{A8.10})$$

Having illustrated the individual element estimation technique for both S_{XX} and S_{YX} , only the matrix approach will be covered for the remaining covariance matrix estimates. Individual element estimation equations are fairly easy to determine and that task is left to those interested readers.

The covariances among the seasonal series, S_{YY} is estimated by

$$S_{YY} = \frac{1}{N-1} \sum_{v=1}^N (Y_v Y_v^T) \quad (\text{A8.11})$$

or

$$S_{YY} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v,1}^{(1)} \\ y_{v,2}^{(1)} \\ y_{v,3}^{(1)} \\ y_{v,1}^{(2)} \\ y_{v,2}^{(2)} \\ y_{v,3}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v,1}^{(1)} & y_{v,2}^{(1)} & y_{v,3}^{(1)} & y_{v,1}^{(2)} & y_{v,2}^{(2)} & y_{v,3}^{(2)} \end{bmatrix} \quad (\text{A8.12})$$

COVARIANCE ESTIMATES FOR THE EXTENDED TEMPORAL DISAGGREGATION MODEL

The parameter estimates for this model are given by Eqs. (8.16), (8.17), and (8.18). These equations require sample estimates for S_{XX} , S_{YX} , S_{YY} , S_{ZZ} , S_{ZX} , and S_{YZ} . The first three of these are the same as those given for the basic temporal model. To determine the covariances involving Z , consider the case where Z contains two seasons of the previous years data. For this example it has dimensions of $(2n \times 1)$ or (4×1) and has the form

$$Z_v = \begin{bmatrix} y_{v-1,2}^{(1)} \\ y_{v-1,3}^{(1)} \\ \text{-----} \\ y_{v-1,2}^{(2)} \\ y_{v-1,3}^{(2)} \end{bmatrix} \quad (\text{A8.13})$$

Before presenting the estimation equations, it is appropriate to point out that there are two basic types of moment estimators, open series and circular series estimators. For the preceding equations, there are no differences between the two approaches. For the circular series approach, the series is taken to repeat itself so that values of $y_{v,\tau}^{(i)}$ and $x_v^{(i)}$ for $v = 0$ or $v = n + 1$ have meaning and may be used in calculations. In this approach $y_{0,\tau}^{(i)}$ equals $y_{N,\tau}^{(i)}$, $x_0^{(i)}$ equals $x_N^{(i)}$, $y_{N+1,\tau}^{(i)}$ equals $y_{1,\tau}^{(i)}$, and $x_{N+1}^{(i)}$ equals $x_1^{(i)}$. For the open series approach the assumption of continual repetition of the sample series is not made. In actuality the value of the estimates are only changed a negligible amount. The estimation procedure is made slightly more complicated for the open series approach. For this approach, the summations no longer always go from one to N as they do for the circular series approach; they go only over all pairs of values for which both elements of the pair are defined. In addition the mean values (to subtract from the series) are usually calculated only over the values used in the covariance calculation. For this chapter, only the circular series approach is used.

Thus the covariance matrix estimate, S_{ZZ} , may be calculated as

$$S_{ZZ} = \frac{1}{N-1} \sum_{v=1}^N (Z_v Z_v^T) \quad (A8.14)$$

or

$$S_{ZZ} = \frac{1}{N-1} \sum_{N=1}^N \begin{bmatrix} y_{v-1,2}^{(1)} \\ y_{v-1,3}^{(1)} \\ y_{v-1,2}^{(2)} \\ y_{v-1,3}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v-1,2}^{(1)} & y_{v-1,3}^{(1)} & y_{v-1,2}^{(2)} & y_{v-1,3}^{(2)} \end{bmatrix} \quad (A8.15)$$

If all of the seasons are included in Z , then S_{ZZ} would be identical to S_{YY} .

The covariance matrix estimate, S_{ZX} , may be calculated as

$$S_{ZX} = \frac{1}{N-1} \sum_{v=1}^N (Z_v X_v^T) \quad (A8.16)$$

or

$$S_{ZX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v-1,2}^{(1)} \\ y_{v-1,3}^{(1)} \\ y_{v-1,2}^{(2)} \\ y_{v-1,3}^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (A8.17)$$

Likewise, S_{ZY} may be calculated as

$$S_{ZY} = \frac{1}{N-1} \sum_{v=1}^N (Z_v Y_v^T) \quad (A8.18)$$

or

$$S_{ZY} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v-1,2}^{(1)} \\ y_{v-1,3}^{(1)} \\ y_{v-1,2}^{(2)} \\ y_{v-1,3}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v,1}^{(1)} & y_{v,2}^{(1)} & y_{v,3}^{(1)} & y_{v,1}^{(2)} & y_{v,2}^{(2)} & y_{v,3}^{(2)} \end{bmatrix} \quad (A8.19)$$

COVARIANCE ESTIMATES FOR THE CONDENSED TEMPORAL DISAGGREGATION MODEL

The parameter estimates for this model are given by Eqs. (8.19), (8.20) and (8.21). The required moment estimates are $S_{YY}(\tau, \tau)$, $S_{YX}(\tau, \tau)$, $S_{XX}(\tau, \tau)$, $S_{YY}(\tau, \tau-1)$, $S_{YX}(\tau-1, \tau)$ using the notation described in Sec. 8.3.4 and where τ is the season. Moment $S_{YY}(\tau-1, \tau-1)$ of Eqs. (8.19) and (8.20) is simply $S_{YY}(\tau, \tau)$ of the previous season and moment $S_{XY}(\tau, \tau-1)$ is simply the transpose of $S_{YX}(\tau-1, \tau)$ so these moment estimates will not be illustrated separately. For this model there are a separate set of moments and parameters for each season. The required moments may be estimated as

$$S_{YY}(\tau, \tau) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v,\tau}^{(1)} & y_{v,\tau}^{(2)} \end{bmatrix} \quad (\text{A8.20})$$

$$S_{YX}(\tau, \tau) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.21})$$

$$S_{XX}(\tau, \tau) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.22})$$

$$S_{YY}(\tau, \tau-1) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v,\tau}^{(1)} \\ y_{v,\tau}^{(2)} \end{bmatrix} \begin{bmatrix} y_{v,\tau-1}^{(1)} & y_{v,\tau-1}^{(2)} \end{bmatrix} \quad (\text{A8.23})$$

and

$$S_{YX}(\tau-1, \tau) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_{v,\tau-1}^{(1)} \\ y_{v,\tau-1}^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.24})$$

$S_{XX}(\tau, \tau)$ has the same value for all seasons of the year.

COVARIANCE ESTIMATES FOR THE SPATIAL DISAGGREGATION MODEL

For this model, both Y and X are annual time series, the Y column matrix composed of annual values for the substations and the X column matrix composed of annual values for the key stations. For illustrative purposes, consider the case of four substations and two key stations. The required moment estimates, S_{YY} , S_{XX} , S_{YX} , $S_{YY}(1)$ and $S_{XY}(1)$, may be estimated as

$$S_{YY} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_v^{(1)} \\ y_v^{(2)} \\ y_v^{(3)} \\ y_v^{(4)} \end{bmatrix} \begin{bmatrix} y_v^{(1)} & y_v^{(2)} & y_v^{(3)} & y_v^{(4)} \end{bmatrix} \quad (\text{A8.25})$$

$$S_{XX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.26})$$

$$S_{YX} = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_v^{(1)} \\ y_v^{(2)} \\ y_v^{(3)} \\ y_v^{(4)} \end{bmatrix} \begin{bmatrix} x_v^{(1)} & x_v^{(2)} \end{bmatrix} \quad (\text{A8.27})$$

$$S_{YY}(1) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} y_v^{(1)} \\ y_v^{(2)} \\ y_v^{(3)} \\ y_v^{(4)} \end{bmatrix} \begin{bmatrix} y_{v-1}^{(1)} & y_{v-1}^{(2)} & y_{v-1}^{(3)} & y_{v-1}^{(4)} \end{bmatrix} \quad (\text{A8.28})$$

and

$$S_{XY}(1) = \frac{1}{N-1} \sum_{v=1}^N \begin{bmatrix} x_v^{(1)} \\ x_v^{(2)} \end{bmatrix} \begin{bmatrix} y_{v-1}^{(1)} & y_{v-1}^{(2)} & y_{v-1}^{(3)} & y_{v-1}^{(4)} \end{bmatrix} \quad (\text{A8.29})$$

บทที่ 9

ข้อพิจารณาในการประยุกต์ใช้แบบจำลองสโตแคสติก CONSIDERATIONS IN MODEL APPLICATION

วัตถุประสงค์ของบทนี้ คือการนำเสนอข้อพิจารณาในการนำแบบจำลองสโตแคสติกที่กล่าวถึงในหนังสือเล่มนี้ไปประยุกต์ใช้ในงานทางวิศวกรรมชลประทาน ซึ่งจะมีผลต่อความสำเร็จหรือความล้มเหลวในการนำแบบจำลองไปใช้ในงานทางวิศวกรรมชลประทาน ซึ่งอนุกรมหลักที่กล่าวถึงคือ อนุกรมเวลาน้ำท่าในหนังสือเล่มนี้ แต่อย่างไรก็ตามแนวคิดและวิธีการที่กล่าวถึงสามารถนำไปประยุกต์ใช้กับอนุกรมเวลาอื่นๆ ได้

9.1 การเตรียมข้อมูลอนุกรมเวลา (Pretreatment of Historical Data)

การเตรียมข้อมูลอนุกรมเวลา หมายถึง กรรมวิธีในการปรับอนุกรมเวลาตัวอย่าง (Historical Data) ให้มีสภาพเป็น Homogeneous เพื่อสะท้อนคุณสมบัติของกลุ่มน้ำตามธรรมชาติและไม่ได้เปลี่ยนแปลงไปตามเวลา (Stationary) รวมถึงการปรับแก้ความผิดพลาดในการเก็บและบันทึกข้อมูล (System Error) การเติมข้อมูลที่ขาดหายไป (Missing Data) หรือการต่อขยายข้อมูล (Extension) การเตรียมห้องข้อมูลถือเป็นเรื่องจำเป็นอันดับแรก ก่อนที่จะเริ่มทำการวิเคราะห์ และสร้างแบบจำลองสโตแคสติก

9.1.1 การแปลความหมายข้อมูล (Data Compilation)

อนุกรมเวลาทางอุทกวิทยาโดยทั่วไป มีข้อผิดพลาดทั้งในรูปของการเก็บและบันทึกข้อมูล (System Error) และความผิดพลาดตามธรรมชาติ (Random Error) เช่น มีการเปลี่ยนจุดตรวจวัดน้ำ หรือการเปรียบเทียบ Stage-Discharge Relationship ซึ่งข้อมูลอนุกรมเวลาที่มีการพิมพ์มักจะมีการปรับแก้ความผิดพลาดต่างๆ ไว้แล้ว แต่อย่างไรก็ตามผู้ใช้ควรมีการพิจารณาปรับแก้ด้วยตัวเองถ้าจำเป็น

การเปรียบเทียบจุดตรวจวัดน้ำเพียงเล็กน้อยสามารถปรับแก้ได้ง่าย โดยใช้อัตราส่วนพื้นที่ หรือ Proportion Correction แต่ถ้ามีการย้ายจุดตรวจวัดจากจุดเดิมมากจนมีผลต่อข้อมูล อาจต้องใช้วิธีการอื่นๆ ประกอบ กรณีที่ Rating Curve ของเดือนหนึ่งมีการเปลี่ยนแปลงอย่างมากจากเดือนที่แล้ว อาจต้องมีการปรับค่าที่สะท้อนถึงการเปลี่ยนแปลงแบบเฉียบพลันของ Rating Curve ผู้ที่ใช้ข้อมูลที่มีการตีพิมพ์เผยแพร่ควรต้องอ่านคำอธิบายที่มาของข้อมูลอย่างละเอียดว่ามีการปรับแก้ข้อมูลมาแล้วหรือไม่ และปรับอย่างไร ข้อมูลของช่วงแรกๆ ของแม่น้ำสายหลักมักมีการเติมข้อมูลโดยใช้ข้อมูลจากสถานีใกล้เคียง นอกจากนั้นข้อผิดพลาดเกี่ยวกับข้อมูลที่อาจเกิดขึ้น ได้แก่ การใส่ตำแหน่งทศนิยมผิดที่

วิธีการตรวจสอบความผิดพลาดของข้อมูลแบบง่ายๆ คือการพล็อตกราฟ ซึ่งสามารถทำได้ 4 แบบ คือการพล็อตอนุกรมเวลา การพล็อต Simple Mass Curve การพล็อตข้อมูล 2 สถานีเปรียบเทียบและ Double Mass Curve

(1) กราฟอนุกรมเวลา ซึ่งจะทำให้เห็นข้อผิดพลาดของข้อมูลได้ง่าย ยกตัวอย่าง อนุกรมเวลาของน้ำท่าทำ Hoover Dam ในรูปที่ 9.1 แสดงให้เห็นว่าอนุกรมเวลาน้ำท่ามีการเปลี่ยนแปลงอย่างมากหลังปี 1935 หลังจากเขื่อน Hoover เริ่มเก็บกักน้ำ และหลังปี 1963 เมื่อเขื่อน Glen Canyon เริ่มการเก็บกักน้ำ

(2) กราฟ Simple Curve Curve คือ กราฟที่แสดงค่าสะสม (Running Sum) กับเวลาการเปลี่ยนแปลงของความลาดชันของเส้นแบบผิดปกติ อาจบอกถึงข้อผิดพลาดในข้อมูลได้

(3) การพล็อตข้อมูลของ 2 สถานีใกล้เคียงเปรียบเทียบ จะทำให้เห็นถึงความสัมพันธ์ระหว่าง 2 สถานี ตลอดจนการเปลี่ยนแปลงความสัมพันธ์ของข้อมูลระหว่าง 2 สถานีนั้น Multivariate ที่กล่าวถึงในหนังสือเล่มนี้จะอยู่บนพื้นฐานของความสัมพันธ์เชิงเส้น

(4) Double Mass Curve นิยมใช้ในการตรวจสอบความคงเส้นคงวา (Consistency) ของข้อมูล

9.1.2 การเติมข้อมูลที่ขาดหายและการต่อขยายข้อมูล (Data Fill-in and Extension)

กรณีใช้ออนุกรมเวลา 1 ชุด ข้อมูลที่ขาดหายไปไม่มีผลมากในการวิเคราะห์ และอาจถือว่าข้อมูลที่เหลืออยู่เป็นข้อมูลที่ต่อเนื่องได้ อย่างไรก็ตามการเติมข้อมูลที่ขาดหายไปโดยใช้ข้อมูลจากสถานีใกล้เคียงจะมีผลดีกว่า โดยเฉพาะข้อมูลที่ขาดหายไปเป็นข้อมูลช่วงสำคัญ เช่น แฉ่งมาก (Extremely Dry) หรือน้ำมาก (Extremely Wet) ถ้าขาดข้อมูลช่วงดังกล่าวถือว่าขาดข้อมูลที่สำคัญ และจะทำให้เกิด Bias ในการสร้างแบบจำลอง การเติมข้อมูลที่ขาดหายไป คือการ Interpolate ข้อมูลขณะที่การต่อขยายข้อมูล คือการ Extrapolate ข้อมูล ซึ่งจะได้ข้อพิจารณาในการต่อขยายข้อมูล เช่นเดียวกับการเติมข้อมูลที่ขาดหายไป

การเติมข้อมูลที่ขาดหายไป หรือการต่อขยายข้อมูล โดยใช้ค่าเฉลี่ย ถือเป็นข้อผิดพลาดที่สำคัญ เพราะจะทำให้ความแปรปรวน (Variance) ของข้อมูลลดน้อยลง การเติมข้อมูลที่ขาดหายไปหรือการต่อขยายข้อมูลโดยใช้สมการ Regression และใช้ข้อมูลจากสถานีใกล้เคียง ควรพิจารณาเพิ่ม Stochastic หรือ Random Error Term เข้าไปด้วย

ในกรณี Multivariate จำเป็นที่อนุกรมเวลาทุกชุดต้องมีข้อมูลที่สมบูรณ์ และมีช่วงเวลาเดียวกัน

9.1.3 การปรับข้อมูลให้เป็นไปตามสภาพธรรมชาติ (Reduction of Data to Natural Conditions)

หลักการพื้นฐานทางสโตแคสติก คือ ต้องสร้างแบบจำลองของกระบวนการตามธรรมชาติ ซึ่งไม่เปลี่ยนแปลงไปตามเวลา ดังนั้นจึงต้องมีการปรับอนุกรมเวลาตัวอย่างให้เป็นอนุกรมเวลาแบบ Unregulated-Undepleted เพื่อเป็นตัวแทนของ Virgin Basin

อาจต้องมีการปรับค่าหลายครั้ง ทั้งการปรับแบบตรงไปตรงมาและการปรับแบบซับซ้อนที่ต้องการความรู้ความชำนาญ เป็นพิเศษการปรับค่าเนื่องจากความจุ การระเหย และการเก็บกักน้ำในคลังของ

อ่างเก็บน้ำ หรือการปรับค่าเนื่องจากการผันน้ำ และ Return Flow ถือเป็นารปรับแบบตรงไปตรงมา แต่ปกติแล้วอาจมีปัญหาข้อมูลสำหรับการปรับไม่ครบถ้วน เป็นข้อมูลเกี่ยวกับการใช้น้ำของพืช

นอกจากนี้อาจมีการเปลี่ยนแปลงในคุณสมบัติของกลุ่มน้ำในการผลิตน้ำท่า อันเนื่องจากการตัดไม้ทำลายป่าในวงกว้าง การเพาะปลูกแบบเลื่อนลอย ไฟฟ้า หรือการพัฒนาในเขตกลุ่มน้ำ (Urbanization) ซึ่งมีผลกระทบต่อกระบวนการผัน-น้ำท่า การปรับแก้จำเป็นต้องมีการตรวจสอบกระบวนการเกิดผัน-น้ำท่าของกลุ่มน้ำอย่างละเอียด

9.2 การเลือกแบบจำลอง (Model Selection)

ในการเลือกแบบจำลอง มีคำถามที่น่าสนใจแต่ตอบยาก คือถ้าเราใช้แบบจำลองที่มีความซับซ้อนมากขึ้น ผลลัพธ์จะเปลี่ยนไปหรือไม่ ถ้าคำตอบคือไม่เปลี่ยนแปลง ก็แสดงว่าไม่จำเป็นแบบจำลองที่ซับซ้อนมากกว่า ที่สำคัญคือการตรวจสอบข้อมูลที่สังเคราะห์ใหม่เปรียบเทียบกับค่าสถิติของอนุกรมเวลาตัวอย่าง

9.2.1 สิ่งสำคัญเกี่ยวกับสถิติของอนุกรมเวลาตัวอย่าง

ในการจำลองทางสโตแคสติก คือการรักษาคุณสมบัติทางสถิติของอนุกรมเวลาตัวอย่างไว้อย่างน้อยข้อมูลที่สังเคราะห์ใหม่ต้องรักษาค่าเฉลี่ยและค่าความแปรปรวนของอนุกรมเวลาตัวอย่างได้ และจะดีกว่าถ้าสามารถรักษาคุณสมบัติทางสถิติของโมเมนต์ที่มีลำดับสูงกว่านั้นได้ เป็นค่าสถิติที่แสดงรูปร่างของการแจกแจง (Shape of Distribution) ตามหลักการแล้วกระบวนการแปลงอนุกรมเวลาให้มีการแจกแจงปกติ (Normalization) จะช่วยรักษาคุณสมบัติของโมเมนต์ที่มีลำดับสูงได้อย่างมีประสิทธิภาพอยู่แล้ว การศึกษาส่วนใหญ่พบว่า Annual Lag-one Serial Correlation มีความสำคัญต่อผลลัพธ์มาก นอกจากนี้คุณสมบัติทางสถิติอื่นๆ ที่ควรต้องรักษาไว้สำหรับแบบจำลองรายปี คือ Lag-two Serial Correlation และดรชนีที่แสดงถึง Long-Term Persistence แบบจำลองรายเดือนหรือแบบจำลองแบบ Periodic ปกติต้องรักษา 2-3 log Periodic Correlation ไว้ แต่อาจไม่จำเป็นต้องรักษาคุณสมบัติทางสถิติรายปี แบบจำลอง Disaggregation มีข้อดีคือสามารถรักษาคุณสมบัติได้มากกว่า 1 ระดับ

9.2.2 การรักษาคุณสมบัติทางสถิติของอนุกรมเวลาตัวอย่าง (Preservation of Historical Statistics)

สิ่งสำคัญที่กล่าวถึงในการจำลองทางสโตแคสติก คือเรื่องของ Parsimony ของพารามิเตอร์ ซึ่งเกี่ยวข้องกับการเลือกพารามิเตอร์ที่มีความสำคัญ โดยไม่จำเป็นต้องมีจำนวนพารามิเตอร์มากเกินไป จนจำนวนข้อมูลที่มีไม่สามารถประมาณค่าพารามิเตอร์ที่แม่นยำได้ การที่ไม่สามารถรักษาคุณสมบัติของ

Parsimony ได้ จะทำให้แบบจำลองไม่สามารถสังเคราะห์อนุกรมเวลาที่มีความแปรปรวน (Variability) ตามที่ต้องการได้

ค่าสถิติบางตัวของอนุกรมเวลา บางตัวอาจมีค่าไม่ต่างจากศูนย์ ซึ่งแสดงว่าไม่จำเป็นต้องรักษาคุณสมบัติในอนุกรมเวลาที่สังเคราะห์ใหม่

9.2.3 วัตถุประสงค์ในการสังเคราะห์ข้อมูล (Propose for Generation)

การเลือกเทคนิคในการจำลองทางสโตแคสติก ไม่เพียงเกี่ยวข้องกับคุณสมบัติทางสถิติของอนุกรมเวลาเท่านั้น แต่ยังเกี่ยวข้องกับวัตถุประสงค์ในการนำข้อมูลสังเคราะห์ไปใช้งานด้วย โดยทั่วไปควรเริ่มจากแบบจำลองที่ง่าย ๆ เช่น $AR(p)$ ก่อน ไม่ควรเลือกใช้แบบจำลองที่ซับซ้อนจนเกินความจำเป็น และพึงจำไว้ว่าเทคนิคแต่ละเทคนิคมีความเหมาะสมกับการประยุกต์ใช้งานแต่ละประเภท

9.2.4 ความอ่อนไหวหรือความไวของผลลัพธ์ (Sensitivity of Results)

การเลือกรูปแบบของแบบจำลองจะมีผลต่อการสังเคราะห์ข้อมูลมากน้อยเท่าใด อาจประเมินได้ยาก ดังนั้นจึงต้องพิจารณาถึงความอ่อนไหว หรือความไวของผลลัพธ์ ถ้าแบบจำลองนั้นให้ผลลัพธ์ที่เปลี่ยนแปลงมากขณะที่ ค่าสถิติมีการเปลี่ยนแปลงน้อย แสดงว่าแบบจำลองนั้นอ่อนไหวต่อค่าสถิติตัวนั้น และโดยทั่วไปข้อมูลที่มีไม่เพียงพอต่อการประมาณค่าสถิติที่แม่นยำได้ ดังนั้นจึงจำเป็นต้องเลือกแบบจำลองที่เหมาะสมกับข้อมูลที่มี เลือกค่าสถิติที่ต้องการรักษาไว้ตามลักษณะข้อมูลที่มี

9.2.5 การวิเคราะห์ทั้งลุ่มน้ำ (Reginald Analysis)

ประเทศที่กำลังพัฒนามักประสบปัญหาข้อมูลนั้นไม่เพียงพอที่จะประมาณค่าพารามิเตอร์ที่แม่นยำที่สามารถให้ผลลัพธ์ที่น่าเชื่อถือหรือมั่นคงได้ (Reliable or Stable) และบางกรณีไม่มีข้อมูลสำหรับการประมาณพารามิเตอร์เลย กรณีนี้มี 2 ทางเลือก คือ (1) หาข้อมูลที่ขาดหายไปและ (2) ใช้หลักการวิเคราะห์พารามิเตอร์ของกลุ่มน้ำ หลักของการวิเคราะห์ทั้งลุ่มน้ำ คือสมมติว่าพารามิเตอร์มีค่าคงที่หรือเปลี่ยนแปลงไปตามพื้นที่หรือลุ่มน้ำ ตามสมมติฐานนี้จะคำนวณค่าพารามิเตอร์จากสถานที่ที่มีข้อมูล แล้วทำการ Interpolate เพื่อสร้างเส้น Contour ของพารามิเตอร์ วิธีนี้สามารถใช้กับน้ำฝน อุณหภูมิ แต่ไม่สามารถใช้ได้กับอัตราการไหลเนื่องจากอัตราการไหลเปลี่ยนแปลงไปตามขนาดพื้นที่ลุ่มน้ำ รูปร่างลุ่มน้ำ และรูปแบบการวางตัว (Orientation) ของลุ่มน้ำ

9.3 การประยุกต์ใช้แบบจำลอง (Model Application)

อนุกรมเวลาที่สังเคราะห์ขึ้นตามหลักการของสโตแคสติก สามารถใช้ในการศึกษาแทนอนุกรมเวลาตัวอย่าง (Historical Data) และซึ่งอาจให้คำตอบที่แตกต่างจากการวิเคราะห์โดยใช้อนุกรมเวลาตัวอย่าง อย่างไรก็ตามผลลัพธ์จะออกมาเป็นเช่นไร ขึ้นอยู่กับ Time Dependence และ/หรือ Space Dependence ของอนุกรมเวลา ในหัวข้อนี้จะกล่าวถึง แนวทางการนำแบบจำลองสโตแคสติกไปประยุกต์ใช้งานในด้านต่างๆ

9.3.1 การหาขนาดอ่างเก็บน้ำ (Reservoir Sizing Studies)

ตัวอย่างการนำแบบจำลองสโตแคสติกไปประยุกต์ใช้งานที่สำคัญคือ การหาขนาดอ่างเก็บน้ำขนาดใหญ่ ซึ่งมีความสามารถเก็บกักน้ำไว้ใช้งานได้ยาวนานกว่า 1 ปี หรือที่เรียกว่า Annual Carryover หลักการในการคำนวณหาขนาดอ่างเก็บน้ำจะสัมพันธ์ yield หรือพุดง่ายๆ ว่า ความจุ และ yield (ผลผลิต) นั้นมีความสัมพันธ์กัน และยังสัมพันธ์กับตัวแปรหรือปัจจัยอื่นๆ เช่น ความต้องการใช้น้ำ (Potential Demand) ของพื้นที่บริการ ธรณีวิทยาและภูมิศาสตร์ของทำเลที่ตั้งอ่างเก็บน้ำ ค่าก่อสร้าง ซึ่งปัจจัยเหล่านี้จะมีผลต่อขนาดอ่างเก็บน้ำที่เหมาะสมในเชิงเศรษฐศาสตร์

วิธีการหาขนาดอ่างเก็บน้ำนี้ โดยปกติจะใช้ข้อมูลตัวอย่าง และปกติจะเลือกใช้เฉพาะข้อมูลในช่วงวิกฤติ (Critical Period) ช่วงวิกฤติในที่นี้จะหมายถึง ช่วงเวลาที่ปริมาณน้ำที่ไหลเข้าอ่างและความต้องการใช้น้ำมีความแตกต่างกันมากที่สุด แต่การใช้ข้อมูลช่วงวิกฤติ หรือการใช้อนุกรมเวลาตัวอย่าง จะทำให้ไม่ทราบค่าความเสี่ยงที่ระบบไม่สามารถทำงานได้ (Risk of Failure)

กระบวนการหาขนาดอ่างเก็บน้ำจะเกี่ยวข้องกับ Economic Optimization เป็นสำคัญ ซึ่งจำเป็นต้องมีข้อมูลเชิงตัวเลขที่น่าเชื่อถือได้ และมีความถูกต้องแม่นยำสูง และข้อมูลปริมาณการไหลของน้ำเข้าอ่างเป็นหนึ่งในข้อมูลที่ขาดต่อการแจกแจงที่ถูกต้องแม่นยำ ดังนั้นการใช้แบบจำลองสโตแคสติกสังเคราะห์ข้อมูลเพื่อการหาขนาดอ่างเก็บน้ำจึงเป็นทางเลือกที่สำคัญอันหนึ่ง

9.3.2 การศึกษาการปฏิบัติการอ่างเก็บน้ำ (Reservoir Operation Studies)

การศึกษากิจการปฏิบัติการอ่างเก็บน้ำ คือวิธีการในการหาแนวทางในการจัดการอ่างเก็บน้ำ หรือการประเมินผลของนโยบายหรือแผนการปฏิบัติการอ่างเก็บน้ำที่ใช้งานอยู่ การสังเคราะห์ข้อมูลอนุกรมเวลาจะทำให้สามารถประมาณความน่าจะเป็นที่การปฏิบัติการอ่างเก็บน้ำจะล้มเหลว (Failure) ดังนั้น ในการหากลยุทธ์การปฏิบัติการอ่างเก็บน้ำที่ดีที่สุด (Optimal Operation Strategy) จึงต้องผนวกเทคนิคทาง สโตคาสติกเข้าไป การ Optimization Scheme

9.3.3 การศึกษาลุ่มน้ำ (Basin-Wide Studies)

การศึกษาลุ่มน้ำ คือการหาวิธีการปฏิบัติการของระบบอ่างเก็บน้ำ หรือการประเมินผลการดำเนินงานของโครงการแหล่งน้ำในลุ่มน้ำ หลักการบริหารจัดการน้ำแบบบูรณาการที่เรียกว่า Total Water Management หรือ Integrated Water Resource Management หรือ Comprehensive Basin Planning ถือเป็นหลักการที่เป็นที่ยอมรับกันโดยทั่วไปในการพัฒนาโครงการแหล่งน้ำ หรืออ่างเก็บน้ำใหม่ในลุ่มน้ำ ตามหลักการนี้ จำเป็นต้องมีการประเมินเปรียบเทียบผลประโยชน์และค่าลงทุน กรณี With and Without Projects และการประเมินจะไม่จำกัดเฉพาะผลประโยชน์และค่าลงทุนของโครงการนั้นเท่านั้น แต่ต้องประเมินผลประโยชน์และค่าลงทุนที่เกิดกับทั้งลุ่มน้ำด้วย เนื่องจากโครงการใหม่จะมีผลกระทบต่อทั้งลุ่มน้ำ เช่น ถ้ามีการพัฒนาโครงการชลประทานบริเวณตอนบนของลุ่มน้ำ จะทำให้มีปริมาณสำหรับการผลิตกระแสไฟฟ้าพลังงานและใช้น้ำน้อยลงในพื้นที่ท้ายน้ำ แต่ในทางกลับกันอ่างเก็บน้ำที่สร้างในพื้นที่ตอนบนของลุ่มน้ำจะช่วยทำให้ปริมาณน้ำที่พื้นที่ลุ่มน้ำตอนล่างได้รับมีความแน่นอนและมั่นคงมากขึ้น ซึ่งถือเป็นผลดีของการสร้างอ่างเก็บน้ำ ในการศึกษาจำเป็นต้องใช้ข้อมูลอนุกรมเวลาที่สังเคราะห์ขึ้นใหม่ เพื่อจะได้สามารถวิเคราะห์ค่าความเสี่ยงต่างๆ ได้

9.4 ข้อจำกัดของแบบจำลอง (Model Limitations)

แบบจำลองสโตแคสติกที่มีการพัฒนาขึ้นใหม่จะแก้ไขข้อบกพร่องในแบบจำลองเก่า แต่ส่วนใหญ่ก็ยังมีข้อจำกัดในการใช้งาน ซึ่งข้อจำกัดบางข้อก็ยังเป็นที่ยกเถียงกันอยู่ หรือไม่มีความสำคัญมากนัก และปัญหาเรื่อง Parsimony ของพารามิเตอร์ ยังเป็นอุปสรรคต่อการพัฒนาแบบจำลองในปัจจุบัน

9.4.1 Short-and Long-Term Persistence

ความจำเป็นในการรักษาคุณสมบัติเกี่ยวกับ Long -Term Persistence ของอนุกรมเวลา ยังเป็นที่ถกเถียงกันอยู่ เช่น จะใช้ค่าอะไรเพื่อแสดง Long -Term persistence ของอนุกรมเวลา จะจัดการอย่างไรต่อ Bias ในค่าสถิติ และต้องใช้ข้อมูลยาวนานเท่าใดเพื่อการประมาณค่าสถิติดังกล่าว ดังนั้นในหนังสือเล่มนี้จึงไม่ได้เน้นการพัฒนาแบบจำลองที่รักษาคุณสมบัติของ Long -Term Persistence

ทั้งแบบจำลอง AR และ ARMA สามารถรักษาคุณสมบัติของ Short-Term Persistence ได้เป็นอย่างดี ถึงแม้ ARMA สามารถรักษาคุณสมบัติของ Long-Term Persistence ได้ แต่ก็สูญเสียความสามารถในการรักษาคุณสมบัติของ Short-Term Persistence บางส่วน ทำให้ไม่สามารถรักษาคุณสมบัติของ Short-Term Persistence ที่มี Lag มากกว่า 1 ได้ดีเท่าแบบจำลอง AR และโดยทั่วไป Correlation ที่มี Lag มากๆ มักไม่เหมาะกับการสังเคราะห์ข้อมูลใหม่ (Generation of Data)

9.4.2 ข้อจำกัดของแบบจำลองแบบ (Periodic)

ตามทฤษฎีแล้ว แบบจำลองแบบ Periodic ซึ่งสามารถรักษาคุณสมบัติทางสถิติรายปีได้อย่างอัตโนมัติเป็นสิ่งที่ต้องการ แต่สิ่งสำคัญคือการวิเคราะห์เพื่อระบุกระบวนการตามธรรมชาติที่แท้จริงของอนุกรมเวลา และการฟิตแบบจำลองแบบ Periodic ให้เข้ากับอนุกรมเวลาอย่างถูกต้องแม่นยำ ซึ่งถ้าทำได้แบบจำลอง Periodic ก็จะสามารถรักษาคุณสมบัติทางสถิติรายปีได้ ในทางตรงกันข้าม ถ้าไม่สามารถระบุกระบวนการตามธรรมชาติ หรือการฟิตแบบจำลองไม่แม่นยำก็จะทำให้เกิดปัญหาในการรักษาคุณสมบัติทางสถิติรายปีได้ ซึ่งถ้าเป็นเช่นนั้นจะถือเป็นข้อจำกัดที่สำคัญของแบบจำลองแบบ Periodic

การประมาณค่าพารามิเตอร์ของแบบจำลอง Periodic จะยุ่งยากกว่าการประมาณค่าพารามิเตอร์ของแบบจำลองรายปี (Annual Model) เนื่องจากแบบจำลอง Periodic มีพารามิเตอร์มากและมักมีปัญหาเกี่ยวกับ Parsimony ของพารามิเตอร์ วิธีการ Smooth ค่าสถิติ เช่น Fourier Series Estimate จะช่วยแก้ปัญหาเกี่ยวกับ Parsimony ของพารามิเตอร์

9.4.3 ข้อจำกัดของแบบจำลอง Disaggregation

แบบจำลอง Disaggregation มีข้อดีคือการเพิ่มความคล่องตัว (Versatility) เข้าไปในกระบวนการจำลอง แต่มีข้อเสียที่สำคัญคือ มีพารามิเตอร์มากเกินไป เมื่อเปรียบเทียบกับข้อมูลตัวอย่างที่มี จึงต้องระวังเรื่อง Parsimony

ยิ่งกว่านั้นถ้ามีการแปลงข้อมูล (Transform Data) อาจทำให้ไม่สามารถรักษาคุณสมบัติของแบบจำลองได้ เช่น ถ้าใช้แบบจำลอง Disaggregation ในการสังเคราะห์ข้อมูลรายเดือน โดยมีอนุกรมเวลารายปีเป็นอนุกรมเวลาหลัก ผลรวมของอนุกรมเวลารายเดือนที่สังเคราะห์ขึ้นมาอาจรวมแล้วไม่เท่ากับอนุกรมเวลารายปี ถ้ามีการแปลงข้อมูล และโดยทั่วไปอนุกรมเวลาทางอุทกวิทยาไม่ได้มีการแจกแจงแบบปกติ แต่แบบจำลองสโตแคสติกต้องการอนุกรมเวลาที่มีการแจกแจงปกติ และจำเป็นต้องมีการแปลงอนุกรมเวลาตัวอย่าง (x_t) ให้เป็นอนุกรมเวลาที่มีการแจกแจงแบบปกติ (y_t) ซึ่งต้องทำด้วยความระมัดระวัง

9.4.4 สรุป

ตำราเล่มนี้กล่าวถึงขั้นตอนที่ใช้ในทางปฏิบัติในการจำลองอนุกรมเวลาทางอุทกวิทยา การสร้างแบบจำลองที่กล่าวถึง จะครอบคลุมเฉพาะแบบจำลองที่เป็นที่ยอมรับและใช้งานกัน โดยทั่วไป

เทคนิคการจำลองทางสโตแคสติก สามารถแบ่งออกได้เป็น 3 ขั้นตอน คือ

ขั้นที่ 1 การเตรียมข้อมูล ซึ่งได้แก่การเติมข้อมูลที่ขาดหายไป หรือการต่อขยายข้อมูล และการตรวจสอบการเปลี่ยนแปลงของอนุกรมเวลา

ขั้นที่ 2 การสร้างแบบจำลองทางคณิตศาสตร์ ซึ่งตำราเล่มนี้ได้กล่าวถึงการจำลองทางคณิตศาสตร์ไว้อย่างละเอียด

ขั้นที่ 3 การประยุกต์ใช้งาน ซึ่งรวมถึงเทคนิคในการสร้างแบบจำลองและการนำเสนอผลลัพธ์

9.5 เอกสารอ้างอิง

Salas, J, J.W.Delleur, V.Yeyjavich and W.L Lane. 1980. **Applied Modeling of Hydrologic Time Series.** Water Resources Publications. USA. pp.484

Chapter 10

Markov Chain Modelling

10.1 Definition

Let X = a discrete random variable which can be described by a probability distribution function, $f(x)$, where $\sum f(x) = 1.0$.

$\{X_0, X_1, X_2, \dots, X_n\}$ = a sequence (set) of $N+1$ random variables

A stochastic process (in time domain) is a description of the relation between the random variables in $\{X_0, X_1, X_2, \dots, X_n\}$.

10.2 States of the system and their probabilities

Let $\{a_1, a_2, \dots, a_i, a_j, \dots, a_s\}$ be the states of system or all possible outcomes of the random variable. In this case, there are s states. Assuming that X is a random variable representing the weather conditions on any day. The states of weather conditions in any day can be defined as $\{a_1 = \text{sunny day}, a_2 = \text{rainy day}\}$ or s equals to 2.

At any time t , assume that $t=0$ is the initial time period (period=0) for Markov chain, the probability that the random variable (X_0 or $X_{t=0}$) falls into any state which can be defined as follows:

$$\left. \begin{aligned} P(X_0 = a_1) &= p_1^{(0)} \\ P(X_0 = a_2) &= p_2^{(0)} \\ &\cdot \\ P(X_0 = a_i) &= p_i^{(0)} \\ P(X_0 = a_j) &= p_j^{(0)} \\ P(X_0 = a_s) &= p_s^{(0)} \end{aligned} \right\} \quad (10.1)$$

$$p_1^{(0)} + p_2^{(0)} + \dots + p_s^{(0)} = 1.00 \quad (10.2)$$

$$\text{Let } \underline{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_i^{(0)}, p_j^{(0)}, \dots, p_s^{(0)}) \quad (10.3)$$

= An initial probability vector or a probability vector for initial time period, $t=0$.

$$\underline{p}^{(n)} = (p_1^{(n)}, p_2^{(n)}, \dots, p_i^{(n)}, p_j^{(n)}, \dots, p_s^{(n)}) \quad (10.4)$$

= a probability vector for time period n .

10.3 Conditional probability and transition matrix

Let “ a_i ” be the state of system at time period t which changes to the state “ a_j ” in time period $t+1$ as shown in Figure 1. The conditional probability of state change from “ a_i ” to “ a_j ” is defined as shown below.

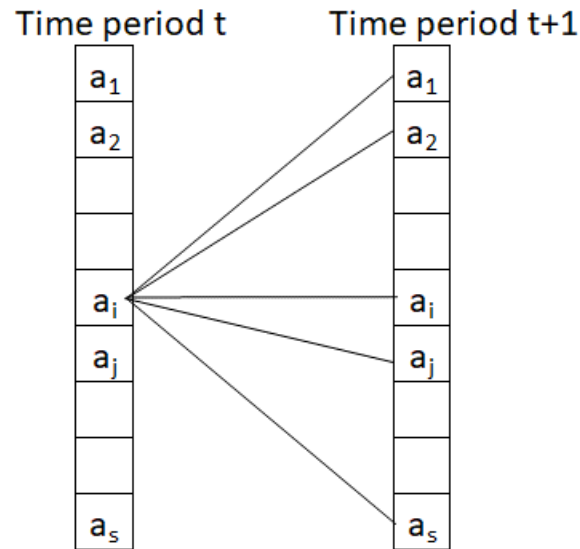


Figure 10.1 Transformation of state of system from one time period to the next time period

One step transition

$$P(X_{t+1} = a_j | X_t = a_i) = p_{ij}^{(1)} \quad (10.5)$$

n steps transition

$$P(X_{t+n} = a_j | X_t = a_i) = p_{ij}^{(n)} \quad (10.6)$$

$$P(X_{t+n} = a_j | X_t = a_i) = P(X_n = a_j | X_0 = a_i) = p_{ij}^{(n)} \quad (10.7)$$

Where $p_{ij}^{(1)}$ and $p_{ij}^{(n)}$ are one step and n-step conditional probability respectively. Equation 10.7 is valid only if a steady state condition exists. Under the steady state condition, the conditional probability does not change with time

Transition matrix is the matrix of conditional probabilities for changing the state of system from period t to period $t+n$. n is the number of steps in transition.

For one time step ($t=1$), the transition matrix is given below

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdot & p_{1i} & p_{1j} & \cdot & p_{1s} \\ p_{21} & p_{22} & \cdot & p_{2i} & p_{2j} & \cdot & p_{2s} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{i1} & p_{i2} & \cdot & p_{ii} & p_{ij} & \cdot & p_{is} \\ p_{j1} & p_{j2} & \cdot & p_{ji} & p_{jj} & \cdot & p_{js} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{s1} & p_{s2} & \cdot & p_{si} & p_{sj} & \cdot & p_{ss} \end{bmatrix} \tag{10.8}$$

Where P = one-step transition matrix

p_{ij} = 1-step conditional probability of changing from state i to state j.

Assume a steady state condition, the transition matrix for multi-step can be derived below.

$$P^2 = P \cdot P \tag{10.9}$$

$$P^3 = P^2 \cdot P \tag{10.10}$$

$$P^n = P^{n-1} \cdot P \tag{10.11}$$

Where P^n = n steps transition matrix

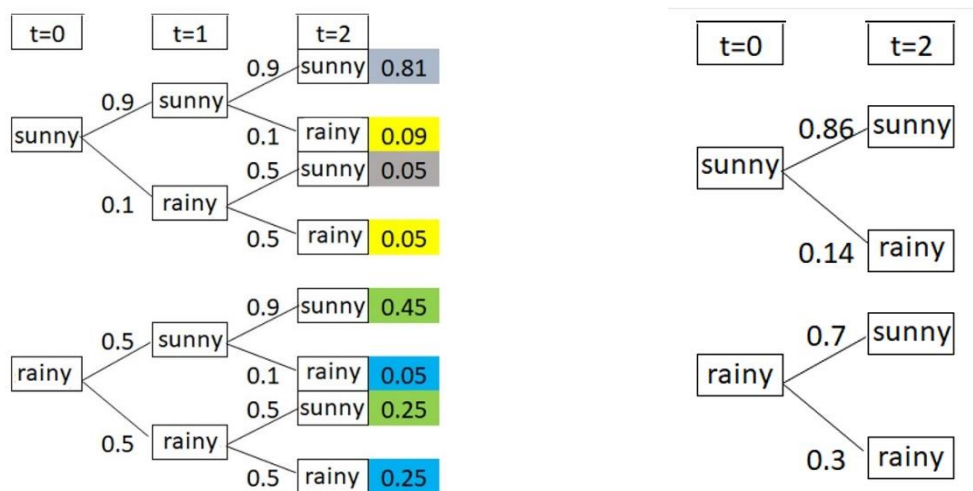
Example 10.1 If the 1-step transition matrix of weather conditions is shown below:

$$P = \begin{bmatrix} P(a_1|a_1) & P(a_2|a_1) \\ P(a_1|a_2) & P(a_2|a_2) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$$

The 2-step transition matrix is calculated below.

$$P^2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.86 & 0.14 \\ 0.70 & 0.30 \end{bmatrix}$$

Figure 10.2 shows the derivation of n-step transition matrix from 1-step transition matrix under steady state condition using Tree diagram.



Transition from t=0 to t=1 and to t=2

Transition from t=0 to t=2

Figure 10.2 Transforming the transition matrix by Tree diagram.

10.4 Markov chain model

Markov chain model is a discrete stochastic process in which the probability distribution of the state in the current stage or time period(t) is conditionally dependent of the state in the past time period(t-1, t-2,). This characteristic is called the **Markov Property**.

$$\underline{p}^{(n)} = \underline{p}^{(n-1)}P = \underline{p}^{(0)}P^n \quad (10.12)$$

$$\underline{p}^{(1)} = \underline{p}^{(0)}P \quad (10.13)$$

$$\underline{p}^{(2)} = \underline{p}^{(1)}P = \underline{p}^{(0)}P^2 \quad (10.14)$$

$$\underline{p}^{(3)} = \underline{p}^{(2)}P = \underline{p}^{(0)}P^3 \quad (10.15)$$

Where

$\underline{p}^{(0)}$, $\underline{p}^{(1)}$, $\underline{p}^{(2)}$, $\underline{p}^{(3)}$ and $\underline{p}^{(n)}$ = probability vector for time period 0, 1, 2, 3 and n respectively.

P , P^2 , P^3 and P^n = transition matrix for time step 1, 2, 3 and n respectively.

10.4.1 First order Markov chain model

First order Markov chain is the method for modelling a stochastic process which the outcome at time t (x_t) depends on the outcome at time t-1 (x_{t-1}), but independent of x_{t-2} , x_{t-3} , ..., x_0 . According to this assumption, the conditional probability can be written as,

$$P(X_t=a_j | X_{t-1}=a_i, X_{t-2}=a_k, X_{t-3}=a_l, \dots, X_0=a_q) = P(X_t=a_j | X_{t-1}=a_i) \quad (10.16)$$

Equation 10.16 above is one step conditional probability.

Let $\{a_i, a_j, \dots, a_s\}$ = the states of system

$$p_{ij}^{(1)} = P(X_t=a_j | X_{t-1}=a_i) \quad (10.17)$$

If the Markov chain is a homogeneous process or a steady state,

$P(X_t=a_j | X_{t-1}=a_i) = p_{ij}^{(1)}$ imply that $p_{ij}^{(1)}$ is independent of time period(t).

Let $P = 2$ states-one step steady state transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad (10.18)$$

$$\sum_{j=1}^s p_{ij} = 1.00 \quad (10.19)$$

where s = No. of states

The transition probability matrix P can be determined from historical data.

Let n_{ij} = no. of times the state of system changed from state i at time t to state j at time t+1. The sample estimate conditional probability can be calculated by equation below.

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^s n_{ij}} \quad ; i \text{ and } j=1,2, \dots, s \quad (10.20)$$

where

\hat{p}_{ij} = sample estimate conditional probability

Example 10.2 Estimate the conditional probability from the records of daily weather conditions are shown in Table 10.1.

Table 10.1 Observed daily weather conditions

Day	Weather condition	Day	Weather condition	Day	Weather condition
1	rainy	11	sunny	21	rainy
2	sunny	12	rainy	22	rainy
3	sunny	13	rainy	23	sunny
4	sunny	14	rainy	24	sunny
5	rainy	15	sunny	25	rainy
6	rainy	16	sunny	26	sunny
7	sunny	17	rainy	27	sunny
8	sunny	18	sunny	28	sunny
9	sunny	19	sunny	29	rainy
10	sunny	20	sunny	30	sunny

Counting the no. of change from state i to state j is shown in Table 10.2.

Table 10.2 No. of times the state of system change from i to j

i	n_{ij}		$\sum_{j=1}^s n_{ij}$
	j		
	sunny	rainy	total
sunny	12	6	18
rainy	7	4	11

From Table 10.2, the sample estimate conditional probability is calculated as shown in Table 10.3.

Table 10.3 Sample estimate conditional probability

i	\hat{p}_{ij}		
	j		
	sunny	rainy	total
sunny	0.667	0.333	1.0
rainy	0.636	0.364	1

Steady state Markov Chain

$$\text{Let } \underline{p}^{(n)} = (p_1^{(n)}, p_2^{(n)}, p_3^{(n)}, \dots, p_j^{(n)}, \dots, p_s^{(n)}) \quad (10.21)$$

$p_j^{(n)}$ = probability that the chain is in state j and time period n .

$\underline{p}^{(0)}$ = initial probability vector

$$\underline{p}^{(1)} = \underline{p}^{(0)}P$$

$$\underline{p}^{(2)} = \underline{p}^{(1)}P = \underline{p}^{(0)}P^2$$

$$\underline{p}^{(n)} = \underline{p}^{(n-1)}P = \underline{p}^{(0)}P^n \quad (10.22)$$

$$\underline{p}^{(t+n)} = \underline{p}^{(t)}P^n = \underline{p}^{(0)}P^{t+n} \quad (10.23)$$

Where P^n = n power of P

For large n , $\underline{p}^{(n)}$ is less dependent on $\underline{p}^{(0)}$ showing that the chain is in steady state.

Thus

$$\underline{p}^{(n)} = \underline{p}^{(t+n)} \quad (10.24)$$

$$\text{and } p_{ij}^{(n)} = p_j^{(n)} \text{ or } p_j^{(n)} \text{ is independent of initial state } (i). \quad (10.25)$$

Thus under steady state condition and n is large

$$\underline{p}^{(n)} = \underline{p}^{(t+n)} = \underline{p} \quad (10.26)$$

$$\underline{p} = \underline{p}P \quad (10.27)$$

10.4.2 n^{th} order Markov chain model

State j in time period t depends on the state of the system of n previous time periods. The conditional probability for n time period can be written below.

$$P(X_t=a_j | X_{t-1}=a_i, X_{t-2}=a_k, X_{t-3}=a_l, \dots, X_0=a_q) = P(X_t=a_j | X_{t-1}=a_i, X_{t-2}=a_k, X_{t-3}=a_l, \dots, X_{t-n}=a_p) \quad (10.28)$$

10.5 A Simple Weather Markov Chain Model

Let assume 2 states Markov chain for the weather conditions prediction, i.e.

State	Weather conditions
0	Sunny day
1	Rainy day

The conditional probabilities of weather conditions based on a given the weather on the preceding day can be represented by a transitional matrix as follow.

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} P(\text{sunny}|\text{sunny}) & P(\text{rainy}|\text{sunny}) \\ P(\text{sunny}|\text{rainy}) & P(\text{rainy}|\text{rainy}) \end{bmatrix}$$

The graphical representation of the above transition matrix is shown in Figure 10.3.

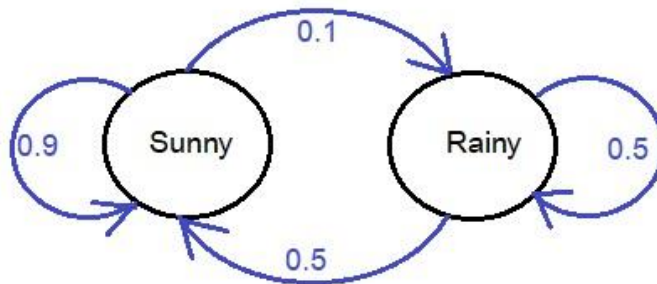


Figure 10.3 Graphical representation of transition probability matrix

The matrix P represents the weather model in which a sunny day is 90% likely to be followed by another sunny day, and a rainy day is 50% likely to be followed by another rainy day. The columns can be labelled "sunny" and "rainy" respectively, and the rows can be labelled in the same order.

p_{ij} is the probability that, if a given day is in state i , it will be followed by a day of state j .

Notice that the rows of P sum to 1.00.

10.5.1 Predicting the weather condition

If the weather on day 0 is known to be sunny. This is represented by a vector in which the "sunny" entry is 100%, and the "rainy" entry is 0% or

$$\underline{p}^{(0)} = (1 \ 0)$$

Where $\underline{p}^{(0)}$ is the probability vector representing the current weather condition.

Thus, the weather on day 1 can be predicted by:

$$\underline{p}^{(1)} = \underline{p}^{(0)}P = (1 \ 0) \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = (0.9 \ 0.1)$$

Thus, there is a 90% chance that day 1 will be sunny.

The weather on day 2 can be predicted in the same way:

$$\underline{p}^{(2)} = \underline{p}^{(1)}P = (0.9 \ 0.1) \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = (0.86 \ 0.14)$$

or

$$\underline{p}^{(2)} = \underline{p}^{(0)}P^2 = (1 \ 0) \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}^2 = (1 \ 0) \begin{bmatrix} 0.86 & 0.14 \\ 0.70 & 0.30 \end{bmatrix} = (0.86 \ 0.14)$$

General rules for predicting the weather on day n are

$$\underline{p}^{(n)} = \underline{p}^{(n-1)}P = \underline{p}^{(0)}P^n$$

10.5.2 Steady state of the weather condition

In this example, the predictions for the weather on more distant days are increasingly inaccurate and tend towards a [steady state vector](#). This vector represents the probabilities of sunny and rainy weather on any day which is independent of the initial weather condition.

The steady state vector is defined as:

$$\underline{p} = \lim_{n \rightarrow \infty} \underline{p}^n \quad (10.29)$$

where $\underline{p} = (p_1, p_2)$ = steady state vector or the probability vector of weather condition on any day.

p_1 = the probability to be sunny day

p_2 = the probability to be rainy day

But only converges to a strictly positive vector if P is a [regular](#) transition matrix (that is, there is at least one P^n with all non-zero entries).

Since the \underline{p} is independent from initial conditions, it must be unchanged when transformed by P . This makes \underline{p} an [eigenvector](#) (with [eigenvalue](#) = 1), and the vector \underline{p} can be derived from the transition matrix P . The example on weather condition prediction, let

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\underline{p}P = \underline{p}$$

$$\underline{p}P = \underline{p}I$$

$$\underline{p}(P - I) = 0$$

$$\underline{p} \left(\begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = 0$$

$$(p_1 \ p_2) \begin{bmatrix} -0.1 & 0.1 \\ 0.5 & -0.5 \end{bmatrix} = 0$$

$$0.1p_1 + 0.5p_2 = 0$$

$$0.1p_1 - 0.5p_2 = 0$$

and since p_1 and p_2 are the probability vector we know that

$$p_1 + p_2 = 1$$

Solving this pair of simultaneous equations to obtain the steady state vector, \underline{p}

$$0.1(1-p_2)-0.5p_2 = 0$$

$$p_2 = 0.1/0.6 = 0.167$$

$$p_1 = 1 - 0.167 = 0.833$$

$$(p_1 \ p_2) = (0.833, 0.167)$$

In conclusion, in the long term, 83.3% is sunny day and 16.7% is rainy day.

Prove that the \underline{p} is independent of the initial conditions, it must be unchanged when transformed by P ($\underline{p}P = \underline{p}$) as shown by Equation (10.27) or transformed by n time steps ($\underline{p} = \lim_{n \rightarrow \infty} \underline{p}^n$) as shown in Equation (10.29).

Prove 1: $\underline{p}P = \underline{p}$

$$\underline{p}P = [0.833333 \quad 0.166667] \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = [0.833333 \quad 0.166667]$$

Prove 2: $\underline{p} = \lim_{n \rightarrow \infty} \underline{p}^n = \underline{p}^{(0)}P^n$

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.86 & 0.14 \\ 0.70 & 0.30 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.8376 & 0.1642 \\ 0.8120 & 0.1880 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} 0.8334426 & 0.1666674 \\ 0.8327872 & 0.1672128 \end{bmatrix}$$

$$P^{16} = \begin{bmatrix} 0.8333334 & 0.1666666 \\ 0.8333330 & 0.1666667 \end{bmatrix}$$

$$P^{32} = \begin{bmatrix} 0.8333333 & 0.1666667 \\ 0.8333333 & 0.1666667 \end{bmatrix}$$

$$\underline{p} = \underline{p}^{(0)}P^{32} = [1 \quad 0] \begin{bmatrix} 0.8333333 & 0.1666667 \\ 0.8333333 & 0.1666667 \end{bmatrix}$$

$$\underline{p} = [0.8333333 \quad 0.1666667]$$

10.6 The Cola Example

Example 10.3: Suppose the entire cola industry produces only two colas (cola 1 and cola 2). Given that

- a person last purchased cola 1, there is a 90% chance that her next purchase will be cola 1.
- a person last purchased cola 2, there is an 80% chance that her next purchase will be cola 2.

Answer the following questions

Question 1: If a person is currently a cola 2 purchaser, what is the probability that she will purchase cola 1 two purchases from now?

Question 2: If a person is currently a cola 1 purchaser, what is the probability that she will purchase cola 1 three purchases from now?

Question 3: Suppose 60% of all people now drink cola 1 and 40% now drink cola 2. Three purchases from now, what fraction of all purchasers will be drinking cola 1?

Solution

Two-state Markov chain due to two types of Cola available in the market ($s=2$).

State 1 = person last purchased cola 1

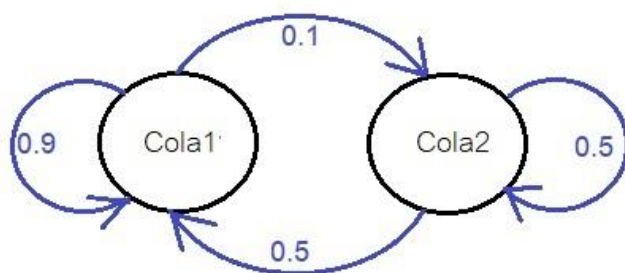
State 2 = person last purchased cola 2

Question 1: If a person is currently a cola 2 purchaser, what is the probability that she will purchase cola 1 two purchases from now?

Determine $P(X_2=\text{cola 1} | X_0=\text{Cola 2}) = p_{21}^{(2)}$

Let X_t = the type of cola purchased at any time period (t) with the transition matrix

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$



Given $X_0 = \text{cola 2}$, thus $\mathbf{p}^{(0)} = (0 \ 1)$

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)}\mathbf{P} = (0 \ 1) \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = (0.2 \ 0.8)$$

$$\mathbf{p}^{(2)} = \mathbf{p}^{(1)}\mathbf{P} = (0.2 \ 0.8) \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = (0.34 \ 0.66)$$

$$P(X_2 = \text{cola 1} | X_0 = \text{cola 2}) = 0.34 \quad \text{Ans}$$

Alternative 1

$$\mathbf{p}^{(2)} = \mathbf{p}^{(0)}\mathbf{P}^2$$

$$\mathbf{P}^2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

$$\mathbf{p}^{(2)} = \mathbf{p}^{(0)}\mathbf{P}^2 = [0 \ 1] \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix} = [0.34 \ 0.66]$$

Alternative 2

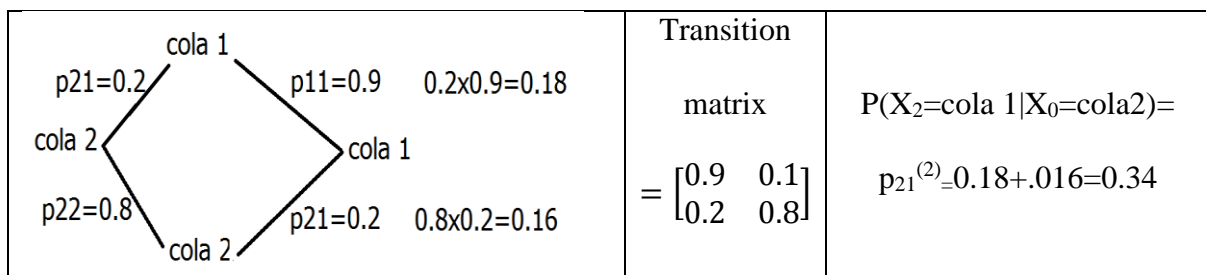


Figure 10.4 Solve Question 1 by Tree diagram

Question 2: If a person is currently a cola 1 purchaser, what is the probability that she will purchase cola 1 three purchases from now?

Determine $P(X_3 = \text{cola 1} | X_0 = \text{cola 1}) = p_{11}^{(3)}$

Given $X_0 = \text{cola 1}$, thus $\mathbf{p}^{(0)} = (1 \ 0)$

$$\mathbf{p}^{(3)} = \mathbf{p}^{(0)}\mathbf{P}^3$$

$$\mathbf{P}^3 = \mathbf{P}^2\mathbf{P} = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix}$$

$$\mathbf{p}^{(3)} = \mathbf{p}^{(0)}\mathbf{P}^3 = [1 \ 0] \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix} = [0.781 \ 0.219]$$

$$P(X_3=\text{cola 1}|X_0=\text{cola 1})= p_{11}^{(3)} = 0.781$$

Alternative: Solve Question 2 by Tree diagram

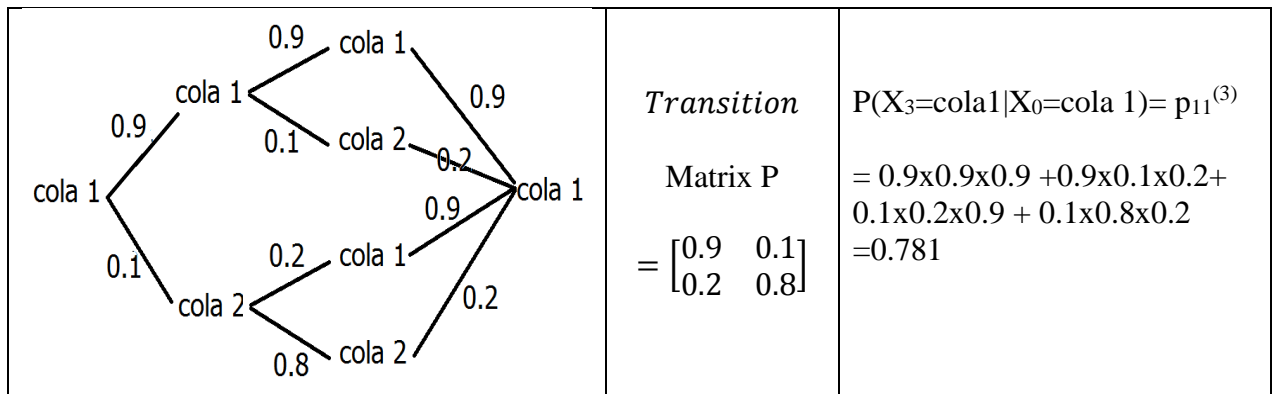


Figure 10.5 Solve Question 2 by Tree diagram

Question 3: Suppose 60% of all people now drink cola 1 and 40% now drink cola 2. Three purchase from now, what fraction of all purchasers will be drinking cola 1?

Determine $P(X_3=\text{cola 1} | \underline{p}^{(0)}) = p_{1_}^{(3)}$

Given $\underline{p}^{(0)} = (0.6 \ 0.4)$

$$P^3 = \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix}$$

$$\underline{p}^{(3)} = \underline{p}^{(0)} P^3$$

$$\underline{p}^{(3)} = \underline{p}^{(0)} P^3 = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix} = \begin{bmatrix} 0.6438 & 0.3562 \end{bmatrix}$$

$$P(X_3=\text{cola 1} | \underline{p}^{(0)}) = p_{1_}^{(3)} = 0.6438$$

Alternative: Solve Question 3 by Tree diagram

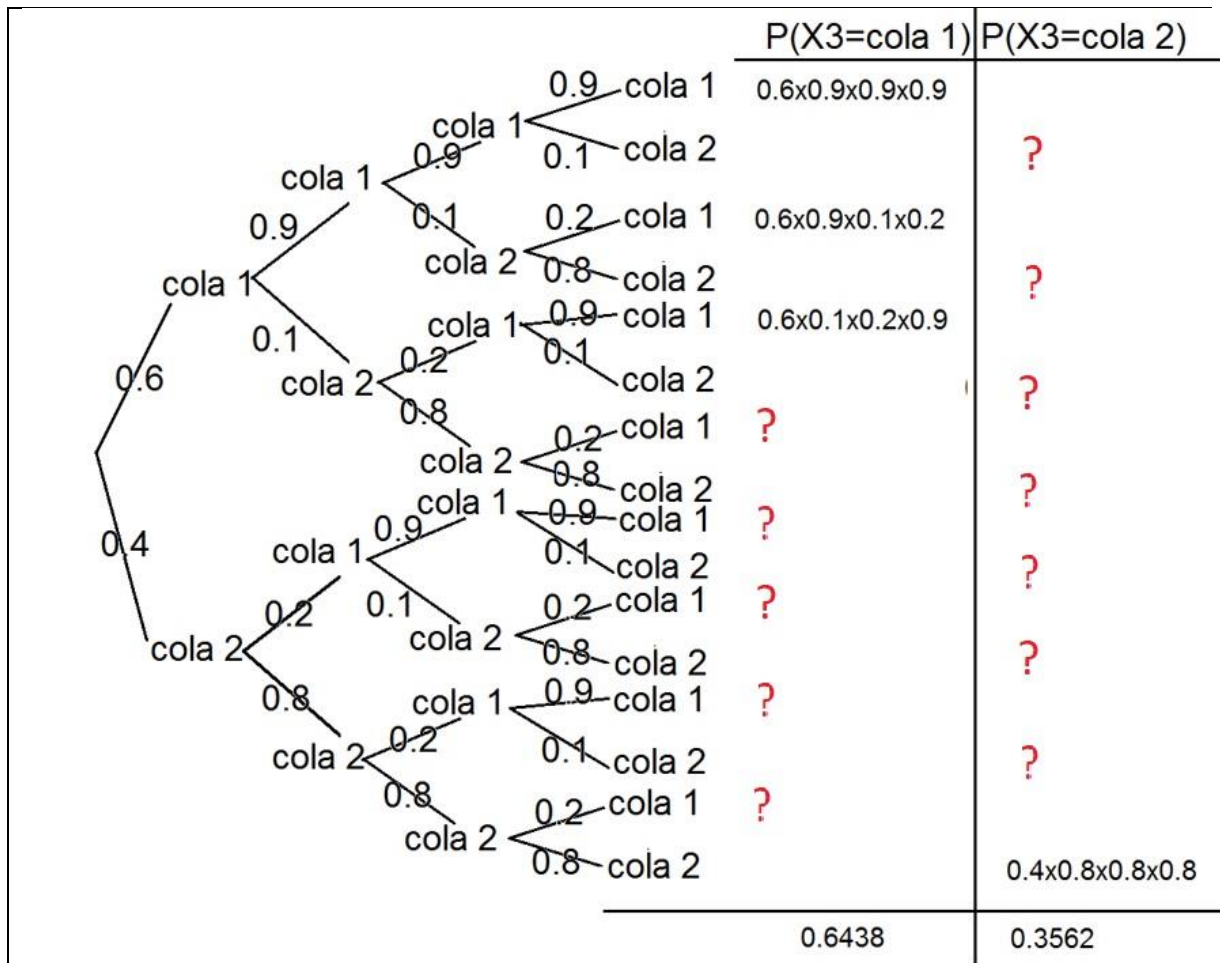
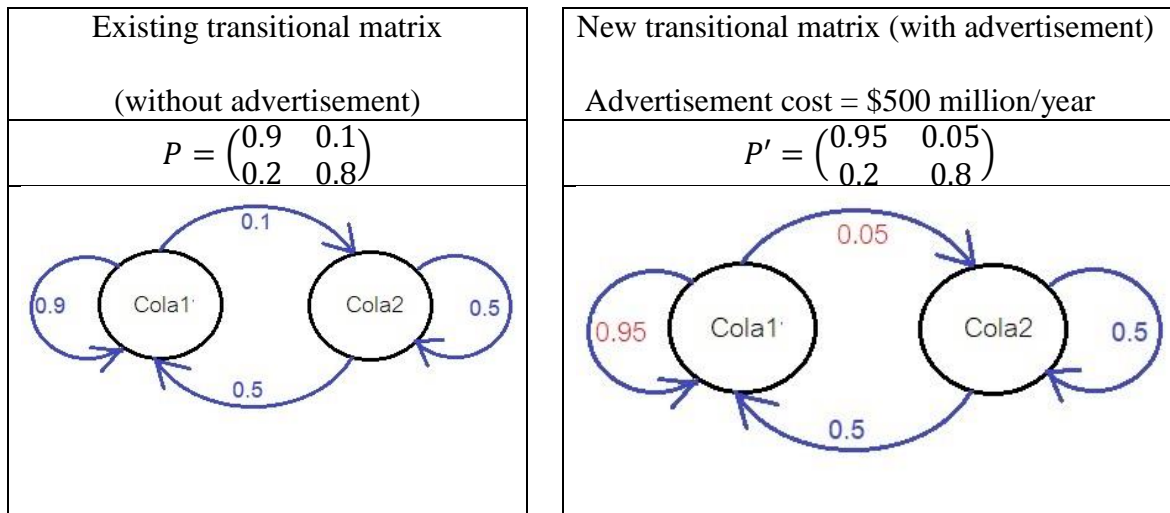


Figure 10.6 Solve Question 3 by Tree diagram

Example 10.4 From Example 10.3, suppose that each customer makes one purchase of cola during any week (52 weeks=1 year). Suppose there are 100 million cola customers. One selling unit of cola costs the company \$1 to produce and is sold for \$2. If the cost of advertisement is \$500 million per year, the advertising firm guarantees to decrease from 10% to 5% the fraction of cola 1 customers who switch to cola 2 after a purchase. Should the company makes cola 1 hire the advertising firm?

Solution

The advertising firm guarantees “to decrease from 10% to 5% the fraction of cola 1 customers who switch to cola 2 after a purchase”. The transition matrix with and without advertisement can be defined as follow:



At steady state: $\underline{p} = \underline{p} P$

Let the steady state probability vector, $\underline{p} = (p_1 \ p_2)$

Without advertisement case

$$(p_1 \ p_2) = (p_1 \ p_2) \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = (0.9p_1 + 0.2p_2 \quad 0.1p_1 + 0.8p_2)$$

$$p_1 = 0.9p_1 + 0.2p_2$$

$$p_2 = 0.1p_1 + 0.8p_2$$

Then $0.1p_1 = 0.2p_2$

$$p_1 = 2p_2$$

Since $p_1 + p_2 = 1$

Thus $p_2 = 1/3$ and $p_1 = 2/3$

Steady state vector $\underline{p} = \left(\frac{2}{3}, \frac{1}{3}\right)$

Cola sell profit = \$1/unit

Total Cola sell profit = 52 weeks/year x 100,000,000 customers x \$1/unit

$$= \$5,200,000,000 \text{ /year}$$

Total Cola 1 sell profit = $(2/3) \times 5,200,000,000 = \$3,466,666,667 \text{ /year}$

Total Cola 2 sell profit = $(1/3) \times 5,200,000,000 = \$1,733,333,333 \text{ /year}$

With advertisement case

$$(p_1 \ p_2) = (p_1 \ p_2) \begin{bmatrix} 0.95 & 0.05 \\ 0.20 & 0.80 \end{bmatrix} = (0.95p_1 + 0.20p_2 \quad 0.05p_1 + 0.80p_2)$$

$$p_1 = 0.95p_1 + 0.20p_2$$

$$p_2 = 0.05p_1 + 0.80p_2$$

Then $0.05p_1 = 0.2p_2$

$$p_1 = 4p_2$$

Since $p_1 + p_2 = 1$

Thus $p_2 = 1/5$ and $p_1 = 4/5$

Steady state vector $\underline{p} = \left(\frac{4}{5}, \frac{1}{5}\right)$

$$\begin{aligned} \text{Total Cola 1 sell profit} &= (4/5) \times \$5,200,000,000 - 500,000,000 \\ &= \$4,160,000,000 - 500,000,000 = \$3,660,000,000 \end{aligned}$$

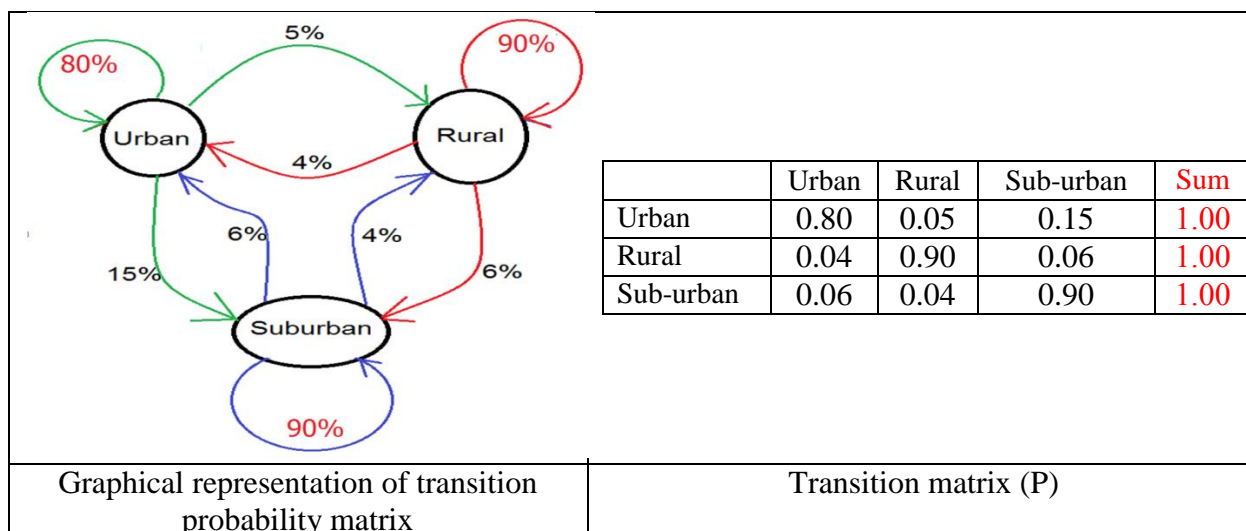
$$\begin{aligned} \text{Total Cola 2 sell profit} &= (1/5) \times \$5,200,000,000 \\ &= \$1,040,000,000 \end{aligned}$$

$$\begin{aligned} \text{Cola 1 sell profit increase} &= \text{Profit with advertisement} - \text{Profit without advertisement} \\ &= \$3,660,000,000 / \text{year} - \$3,466,666,667 / \text{year} \\ &= \$193,333,333 / \text{year} \end{aligned}$$

Therefore Cola 1 company should hire the advertisement agency since the company will gain net profit by \$193,333,333 annually.

10.7 Classification of State of Markov Chain**10.7.1 State Classification - Population distribution example**

Each American family is classified as living in an urban, rural and suburban location. During a given year, 15% of all urban families move to a suburban and 5% move to a rural location; 6% of all suburban families move to an urban location, and 4% move to rural location; finally 4% of all rural families move to an urban location and 6% move to a suburban location.



- (a) If a family now lives in an urban location, what is the probability that it will live in an urban area two years from now? or A rural area ? or A suburban area ?
- (b) Suppose that at present, 40% of all families live in an urban area, 35% live in a sub-urban, 25% live in a rural area. Two years from now, what % of American families will live in an urban area?
- (c) What problems might occur if this model were used to predict the future population distribution of the United States?

Solution

$$(a) \quad \underline{p}^{(2)} = \underline{p}^{(0)}P^2$$

$$\underline{p}^{(0)} = (1 \ 0 \ 0)$$

$$P = \begin{bmatrix} 0.80 & 0.05 & 0.15 \\ 0.04 & 0.90 & 0.06 \\ 0.06 & 0.04 & 0.90 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.80 & 0.05 & 0.15 \\ 0.04 & 0.90 & 0.06 \\ 0.06 & 0.04 & 0.90 \end{bmatrix} \begin{bmatrix} 0.80 & 0.05 & 0.15 \\ 0.04 & 0.90 & 0.06 \\ 0.06 & 0.04 & 0.90 \end{bmatrix} = \begin{bmatrix} 0.6510 & 0.0910 & 0.2580 \\ 0.0716 & 0.8144 & 0.1140 \\ 0.1036 & 0.0750 & 0.8241 \end{bmatrix}$$

$$\underline{p}^{(2)} = \underline{p}^{(0)}P^2 = [1 \ 0 \ 0] \begin{bmatrix} 0.6510 & 0.0910 & 0.2580 \\ 0.0716 & 0.8144 & 0.1140 \\ 0.1036 & 0.0750 & 0.8241 \end{bmatrix} = [0.6510 \ 0.0910 \ 0.2580]$$

$$P(X_2=\text{Urban}|X_0=\text{Urban})=0.6510$$

$$P(X_2=\text{Rural}|X_0=\text{Urban})=0.0910$$

$$P(X_2=\text{Sub-urban}|X_0=\text{Urban})=0.2580$$

$$(b) \quad \underline{p}^{(0)} = (0.40, 0.25, 0.35)$$

$$\mathbf{p}^{(2)} = \mathbf{p}^{(0)} P^2$$

$$\mathbf{p}^{(2)} = \mathbf{p}^{(0)} P^2 = [0.40 \quad 0.25 \quad 0.35] \begin{bmatrix} 0.6510 & 0.0910 & 0.2580 \\ 0.0716 & 0.8144 & 0.1140 \\ 0.1036 & 0.0750 & 0.8241 \end{bmatrix} = \\ [0.3510 \quad 0.2590 \quad 0.3900]$$

(c) Determine the steady state vector

$$P = \begin{bmatrix} 0.80 & 0.05 & 0.15 \\ 0.04 & 0.90 & 0.06 \\ 0.06 & 0.04 & 0.90 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.6510 & 0.0910 & 0.2580 \\ 0.0716 & 0.8144 & 0.1140 \\ 0.1036 & 0.0750 & 0.8241 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.457045 & 0.152701 & 0.390253 \\ 0.116733 & 0.678313 & 0.204954 \\ 0.157911 & 0.132113 & 0.709977 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} 0.288341 & 0.224928 & 0.486731 \\ 0.164898 & 0.505011 & 0.330091 \\ 0.199707 & 0.207524 & 0.592769 \end{bmatrix}$$

$$P^{16} = \begin{bmatrix} 0.217434 & 0.279455 & 0.503111 \\ 0.196744 & 0.360628 & 0.442628 \\ 0.210184 & 0.272735 & 0.517081 \end{bmatrix}$$

$$P^{32} = \begin{bmatrix} 0.208005 & 0.298758 & 0.493237 \\ 0.206764 & 0.305754 & 0.487482 \\ 0.208042 & 0.298119 & 0.493838 \end{bmatrix}$$

$$P^{64} = \begin{bmatrix} 0.207653 & 0.300533 & 0.491814 \\ 0.207644 & 0.300586 & 0.491771 \\ 0.207653 & 0.300528 & 0.491818 \end{bmatrix}$$

$$P^{128} = \begin{bmatrix} 0.20765 & 0.300546 & 0.491803 \\ 0.20765 & 0.300546 & 0.491803 \\ 0.20765 & 0.300546 & 0.491803 \end{bmatrix}$$

It takes too long time to reach the steady state. Therefore the future population distribution of USA may not be predicted by Markov chain.

10.7.2 State Classification - Coin tossing example

At time 0, I have THB 200. At time 1, 2, 3, ..., I play a game (coin tossing) in which I bet THB 100 with probability $p(=0.5)$ that I win the game, and with probability $(1-p)$ that I lose the game. My game is to increase my money to THB 400, and as soon as I do, the game is over. The game is also over if my money is reduced to THB 0.

If we define X_t to be the state of my money after t game is played, then X_0, X_1, \dots, X_t may be viewed as a discrete time stochastic process. Note that $X_0=200$ is a known constant, but X_1 and later X_t 's are random. Given that there is a probability $p=0.5$ that $X_1=300$ and a probability $1-p=0.5$ that $X_1=100$. Note that if $X_t=400$, then the game is over or X_{t+1} and all later X_t 's will also equal to THB 400. Similarly, if $X_t=0$, then the game is over or X_{t+1} and all later X_t 's will also equal to THB 0.

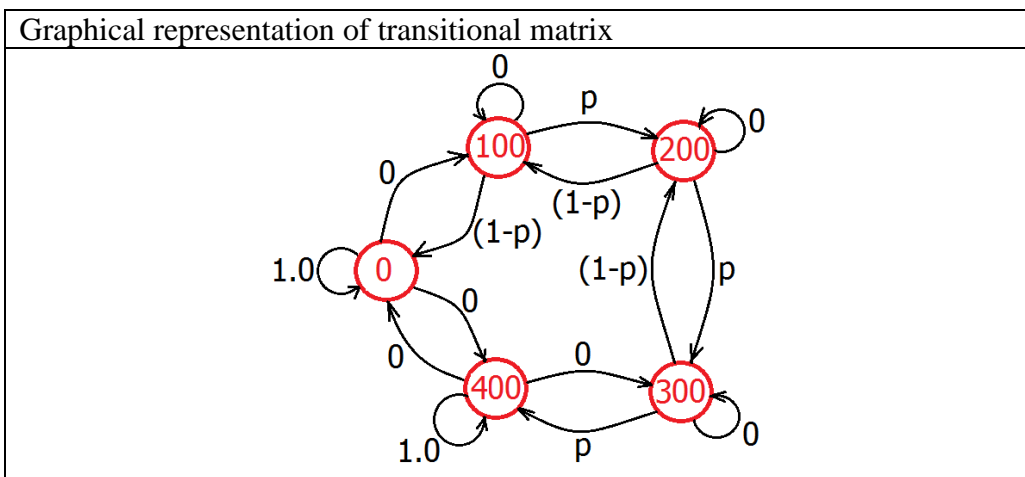
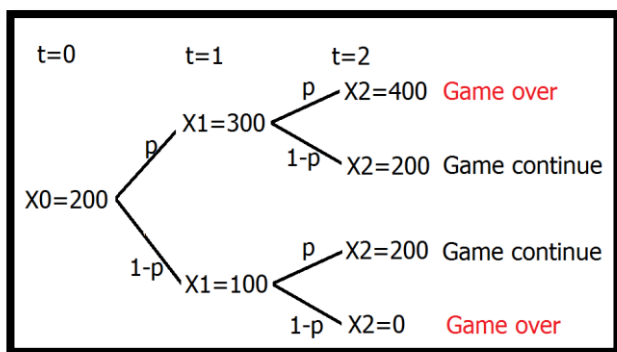
What is the transition matrix?

Solution

$X_t = \{0, 100, 200, 300, 400\}$ = State of Markov Chain (=5 states)

No. of states = 5

Tree diagram



Transition matrix

	0 THB	100 THB	200 THB	300 THB	400 THB
0 THB	1	0	0	0	0
100 THB	$1-p=0.5$	0	$p=0.5$	0	0
200 THB	0	$1-p=0.5$	0	$p=0.5$	0
300 THB	0	0	$1-p=0.5$	0	$p=0.5$
400 THB	0	0	0	0	1

Questions

(1) What is the probability vector after playing the game twice?

Solution

$$\underline{p}^{(2)} = \underline{p}^{(0)} P^2$$

$$\underline{p}^{(0)} = [0 \quad 0 \quad 1 \quad 0 \quad 0]$$

$$P^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.25 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\underline{p}^{(2)} = [0 \quad 0 \quad 1 \quad 0 \quad 0] \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.25 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= [0.25 \quad 0 \quad 0.5 \quad 0 \quad 0.25]$$

From Tree diagram

	$a=0$	$a=100$	$a=200$	$a=300$	$a=400$
$P(X_2=a/X_0=200)$	$(1-p)^2=0.25$	0	$2p(1-p)=0.5$	0	$p^2=0.25$

(2) From the coin tossing problem above, after playing the game twice, what is the probability that I will have THB 300?

$$\text{Solution } P(X_2=300/X_0=200) = 0$$

(3) After playing the game twice, what is the probability that I will have THB 200?

$$\text{Solution } P(X_2=200/X_0=200) = 2p(1-p) = 2*0.5*0.5 = 0.5$$

(3) After playing the game three times, what is the probability that I will have THB 200?

$$\text{Solution } P(X_3=200/X_0=200) = 0$$

(4) What is the probability vector after playing the game three times?

Solution

From Tree diagram

	$a=0$	$a=100$	$a=200$	$a=300$	$a=400$
$P(X_3=a/X_0=200)$	$(1-p)^2$	$2p(1-p)^2$	0	$2p^2(1-p)$	p^2
$p_{a,200}^{(3)}$	0.25	0.25	0	0.25	0.25

Markov model

$$\underline{p}^{(3)} = \underline{p}^{(0)} P^3$$

$$P^3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.25 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.625 & 0 & 0.25 & 0 & 0.125 \\ 0.25 & 0.25 & 0 & 0.25 & 0.25 \\ 0.125 & 0 & 0.25 & 0 & 0.625 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\underline{p}^{(3)} = [0 \quad 0 \quad 1 \quad 0 \quad 0] \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.625 & 0 & 0.25 & 0 & 0.125 \\ 0.25 & 0.25 & 0 & 0.25 & 0.25 \\ 0.125 & 0 & 0.25 & 0 & 0.625 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= [0.25 \quad 0.25 \quad 0 \quad 0.25 \quad 0.25]$$

10.8 Steady State Probabilities and Mean First Passage Times

Mean first passage times is the expected number of transitions from the current state “i” to reach the final state “j”. Mean first passage times can be calculated by the following equation.

$$m_{ij} = p_{ij} + \sum_{k \neq j} p_{ik} (1 + m_{kj}) \quad (10.30)$$

$$p_{ij} + \sum_{k \neq j} p_{ik} = \sum_{j=1}^s p_{ij} = 1 \quad (\text{for any } i \text{ and } j=1 \text{ to } s) \quad (10.31)$$

$$m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj} \quad (10.32)$$

$$m_{ii} = \frac{1}{p_i} \quad (10.33)$$

Where

m_{ij} = the mean first passage times from a state "i" to a state "j"

m_{kj} = the mean first passage times from a state "k" to a state "j"

m_{ii} = the mean first passage times from a state "i" and return to a state "i"

p_{ij} = transition probability from a state "i" to a state "j"

p_{ik} = transition probability from a state "i" to a state "k"

p_i = steady state probability for a state "i"

Example 10.5 From the example 10.3 (Cola industry problem),

$$\text{Transition probability matrix} = P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

$$\text{Steady state vector} = \underline{p} = (p_1, p_2) = \left(\frac{2}{3}, \frac{1}{3}\right)$$

Solution

$$m_{11} = \frac{1}{p_1} = \frac{1}{2/3} = 1.5$$

$$m_{22} = \frac{1}{p_2} = \frac{1}{1/3} = 3$$

$$m_{12} = 1 + p_{11}m_{12} = 1 + 0.9m_{12}$$

$$0.1m_{12} = 1$$

$$m_{12} = \frac{1}{0.1} = 10$$

$$m_{21} = 1 + p_{22}m_{21} = 1 + 0.8m_{21}$$

$$0.2m_{21} = 1$$

$$m_{21} = \frac{1}{0.2} = 5$$

10.9 Types of States in a Markov Chain

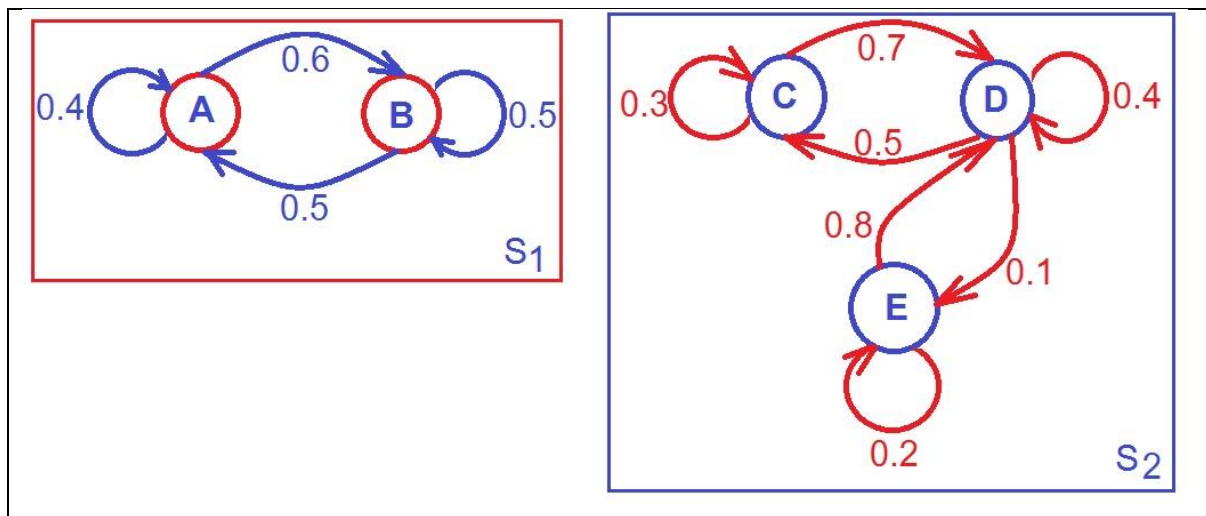
Mathematicians classify the types of states of a Markov Chain according to their properties and relationship among the states in the transition probability matrix.

Let (A, B, C, D, E) be 5 states in Markov Chain with the transition probability matrix P as follows.

$$P = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{pmatrix}$$

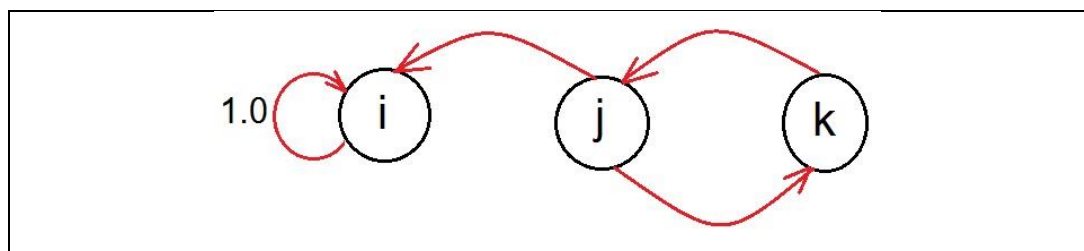
$$P = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$$

The graphical representation of P is shown below.

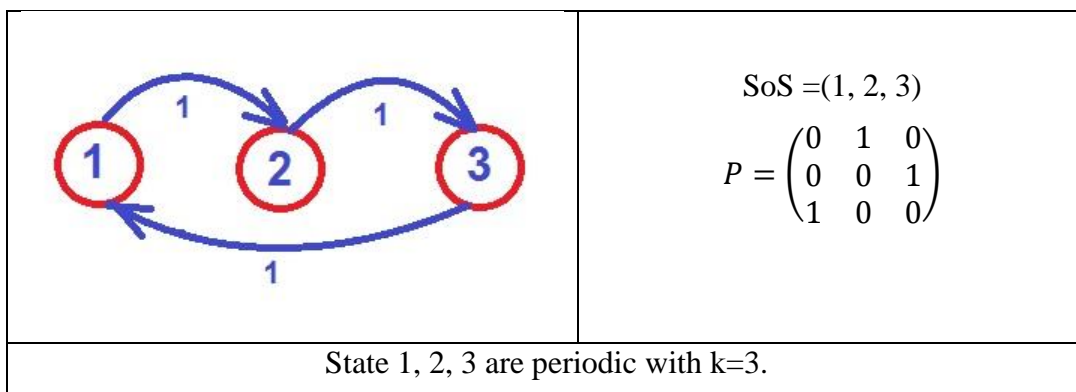


Definition of Terms

1. State E is Reachable from State C and a vice versa. Thus States C and E are said to be Communicate States and so are States A and B.
2. Set of states (S_1) in a Markov Chain is a Closed Set, since any state in S_1 is not reachable from any state outside S_1 . Similarly S_2 is a Closed Set.
3. State i is an Absorbing State if $p_{11} = 1$. This indicates that once entering the absorbing state, leaving is impossible. Therefore the absorbing state is a closed set containing only one state, see Figure below.



4. From the figure above, State k is a Transient State if there exists State i that is reachable from State k, but State k is not reachable from State i.
5. Recurrent State is the state that is not transient state. States A, B, C, D, E are all recurrent states.
6. Recurrent state is periodic if all paths leaving from State i and transition back to State i within a multiple of k transitions ($k > 1$) as shown in the Figure below.



If a recurrent state is not periodic, it is called Aperiodic.

7. If all states in a Markov Chain are recurrent, aperiodic and communicate with each other, the chain is said to be Ergodic.

$P_1 = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/4 & 3/4 \end{pmatrix}$	$P_1 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 2/3 & 1/3 \\ 0 & 0 & 1/4 & 3/4 \end{pmatrix}$
Ergodic	Non-ergodic

Example 10.6 The state college admissions office has modeled the path of a student through state college as a Markov Chain. There are 6 states of the system namely Freshman(F), Sophomore(So), Junior(J), Senior(Sen), Quit(Q) and Graduate(G). The transition probability matrix is given below:

	F.	So.	J.	Sen.	Q.	G.
F.	0.10	0.80	0	0	0.10	0
So.	0	0.10	0.85	0	0.05	0
J.	0	0	0.15	0.80	0.05	0
Sen.	0	0	0	0.10	0.05	0.85
Q.	0	0	0	0	1	0
G.	0	0	0	0	0	1

Each student’s state is observed at the beginning of each fall semester. For example, if a student is a junior at the beginning of the current fall semester, there is an 80% chance that

he will be a senior at the beginning of the next fall semester, a 15% chance that he will be a junior, and a 5% chance that he will have quit. (We assume that once a student quits, he never reenrolls.) You are asked to answer the following questions.

- If a student enters State College as Freshman, how many years can he/she expect to spend as a student at the State College?
- What is the probability that a freshman graduates?

Solution

Let partition the given transition matrix (P) into 4 sub-matrices as shown below.

$$P = \begin{array}{|c|c|c|} \hline & Q & R \\ \hline & O & I \\ \hline \end{array}$$

$Q = \begin{pmatrix} 0.1 & 0.8 & 0 & 0 \\ 0 & 0.1 & 0.85 & 0 \\ 0 & 0 & 0.15 & 0.8 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}$	$R = \begin{pmatrix} 0.1 & 0 \\ 0.05 & 0 \\ 0.05 & 0 \\ 0.05 & 0.85 \end{pmatrix}$
$O = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$I - Q = \begin{pmatrix} 0.9 & -0.8 & 0 & 0 \\ 0 & 0.9 & -0.85 & 0 \\ 0 & 0 & 0.85 & -0.20 \\ 0 & 0 & 0 & 0.9 \end{pmatrix}$$

$$(I - Q)^{-1} = \begin{pmatrix} 1.1111 & 0.9877 & 0.9877 & 0.8779 \\ 0 & 1.1111 & 1.1111 & 0.9877 \\ 0 & 0 & 1.1765 & 1.0458 \\ 0 & 0 & 0 & 0.1111 \end{pmatrix}$$

From the inverse matrix (I-Q) above, we can concluded that if a student enters the State College as a freshman:

the expected time as Freshman = 1.1111 years

the expected time as Sophomore = 0.9877 years

the expected time as Junior = 0.9877 years

the expected time as senior = 0.8779 years

Thus, the expected time as a student = 3.9643 years

$$(I - Q)^{-1}R = \begin{pmatrix} 1.1111 & 0.9877 & 0.9877 & 0.8779 \\ 0 & 1.1111 & 1.1111 & 0.9877 \\ 0 & 0 & 1.1765 & 1.0458 \\ 0 & 0 & 0 & 0.1111 \end{pmatrix} \begin{pmatrix} 0.1 & 0 \\ 0.05 & 0 \\ 0.05 & 0 \\ 0.05 & 0.85 \end{pmatrix}$$

$$(I - Q)^{-1}R = \begin{pmatrix} & Q & G \\ F & 0.2538 & 0.7462 \\ So & 0.1605 & 0.8395 \\ J & 0.1111 & 0.8889 \\ Sen & 0.0556 & 0.9444 \end{pmatrix}$$

The probability that a freshman graduates = 0.7462

10.10 References

https://www.stat.berkeley.edu/~aldous/150/takis_exercises.pdf (One Hundred, Solved, Exercises for the subject: Stochastic Processes by Takis Konstantopoulos)

<https://nms.kcl.ac.uk/ton.coolen/allnotes/MarkovChains.pdf> (Markov Chains Compact Lecture Notes and Exercises September 2009 ACC Coolen Department of Mathematics King's College London)

<https://www.stat.auckland.ac.nz/~fewster/325/notes/ch8.pdf> (Chapter 8: Markov Chains)

https://vknight.org/OR_Methods/Markov_Chains/Markov_Chains_Exercise_Sheet-Solutions.pdf
(Markov Chains Exercise Sheet – Solutions Last updated: October 17, 2012)

10.11 Exercises

(1) A Company has two machines. During any day, each machine that is working at the beginning of the day has a $1/3$ chance of breaking down. If a machine breaks down during the day, it is sent to a repair facility and will be working two days after it breaks down. (Thus, if a machine breaks down during day $t-1$, it will be working at the beginning of day $t+1$.)

Let the state of the system be the number of machines working at the beginning of the day, formulate a transition probability matrix for this situation.

(2) Consider an inventory system in which the sequence of events during each period is as follows: (1) We observe the inventory level (call it i) at the beginning of the period. (2) If $i \leq 1$, $4-i$ units are ordered. If $i \geq 2$, 0 units are ordered. Delivery of all ordered units is immediate. (3) With probability $1/3$, 0 units are demanded during the period; with probability $1/3$, 1 unit is demanded during the period; and with probability $1/3$, 2 units are demanded during the period. (4) We observe the inventory level at the beginning of the next period.

Let a period of state be the period at beginning inventory level. Determine the transition matrix that could be used to model this inventory system as a Markov chain and determine the steady state vector.

(3) An urn contains two unpainted balls at present. We choose a ball at random and flip a coin. If the chosen ball is unpainted and the coin comes up heads, we paint the chosen ball red; if the chosen ball is unpainted and the coin comes up tails, we paint the chosen unpainted ball black. If the ball has already been painted, then (whether head or tail has been tossed) we change the color of the ball (from red to black or from black to red). To model this situation as a stochastic process, we define time t to be the time after the coin has been flipped for the t^{th} time and the chosen ball has been painted. The state at any time may be described by the vector $[u(\text{unpainted}), r(\text{red}), b(\text{black})]$, where u is the number of unpainted balls in the urn, r is the number of red balls in the urn, and b is the number of black balls in the urn. We are given that $X_0=[2,0,0]$. After the first coin toss, one ball will have been painted either red or black, and the state will be either $[1, 1, 0]$ or $[1, 0, 1]$. Hence, we can be sure that $X_t=[1, 1, 0]$ or $X_t=[1, 0, 1]$. Clearly, there must be some sort of relation between the X_t 's. For example, if $X_t=[0, 2, 0]$, we can be sure that X_{t+1} will be $[0, 1, 1]$. Please determine the transition matrix of this process.

(4) From Problem (3), after the two balls are painted, what is the probability that the state is $[0,2,0]$?

(5) Consider two stocks. Stock 1 always sells for THB 10 or THB 20. If stock 1 is selling for THB 10 today, there is a 0.80 chance that it will sell for THB 10 tomorrow. If it is selling for THB 20 today, there is a 0.90 chance that it will sell for THB 20 tomorrow. Stock 2 always sells for THB 10 or THB 25. If stock 2 sells today for THB 10, there is a 0.90 chance that it will sell tomorrow for THB 10. If it sells today for THB 25, there is a 0.85 chance that it will sell tomorrow for THB 25. On the average, which stock will sell for a higher price? Find and interpret all mean first passage times.

(6) At the beginning of each year, my car is in good, fair or broken-down condition. A good car will be good at the beginning of next year with the probability 0.85; fair with probability

0.10; or broken-down with probability 0.05. A fair car will be fair at the beginning of the next year with probability 0.70 or broken-down with probability 0.30. It costs THB 600,000 to purchase a good car; a fair car can be traded in for THB 200,000; and a broken-down car has no trade-in value and must immediately be replaced by a good car. It costs THB 100,000 per year to operate a good car and THB 150,000 to operate a fair car. Should I replace my car as soon as it becomes a fair car, or should I drive my car until it breaks down? Assume that the cost of operating a car depends on the type of car on hand at the beginning of the year (after a new car, if any, arrives).

(7) Three balls are divided between two containers. During each period a ball is randomly chosen and switched to other container.

(a) Find (in steady state) the probability that a container will contain 0, 1, 2 or 3 balls.

(b) if container 1 contains no balls, on the average how many periods will go by before it again contains no balls?

(8) *Work-Force Planning.* An engineering firm employs three types of engineers: junior engineers, senior engineers and partners. During a given year, there is a 0.15 probability that a junior engineer will be promoted to senior engineer and a 0.05 probability the he or she will leave the firm. Also there is a 0.20 probability that a senior engineer will be promoted to partner and a 0.10 probability that he or she will leave the firm. There is a 0.05 probability that a partner will leave the firm. The firm never demotes an engineer.

There are many interesting questions the firm might want to answer. For example, what is the probability that a new hired junior engineer will leave the firm before becoming a partner? On the average, how long does a newly hired junior engineer stay with the firm?

(9) *Payoff Insurance Company* charges a customer according to his or her accident history. A customer who has had no accident during the last two years is charged a THB 3,000 annual premium. Any customer who has had an accident during each of the last two years is charged a THB 12,000 annual premium. A customer who has had an accident during only one of the last two years is charged an annual premium of THB 9,000. A customer who has an accident during the last year has a 10% chance of having an accident during the current year. If a customer has not had an accident during the last year, there is only a 3% chance that he or she will have an accident during the current year. During a given year, what is the average premium paid by a Payoff customer? (Hint: In case of difficulty, try a four-state Markov Chain.

(10) The state college admissions office has modeled the path of a student through state college as a Markov Chain. There are 6 states of the system namely Fresherman(F), Sophomore(So), Junior(J), Senior(Sen), Quit(Q) and Graduate(G). The transition probability matrix is given below:

	F.	So.	J.	Sen.	Q.	G.
F.	0.10	0.80	0	0	0.10	0
So.	0	0.10	0.85	0	0.05	0
J.	0	0	0.15	0.80	0.05	0
Sen.	0	0	0	0.10	0.05	0.85

<i>Q.</i>	0	0	0	0	1	0
<i>G.</i>	0	0	0	0	0	1

Each student's state is observed at the beginning of each fall semester. For example, if a student is a junior at the beginning of the current fall semester, there is an 80% chance that he will be a senior at the beginning of the next fall semester, a 15% chance that he will be a junior, and a 5% chance that he will have quit. (We assume that once a student quits, he never reenrolls.) You are asked to answer the following questions.

- (a) What is the probability that a freshman graduates?
- (b) What is the probability that a freshman quits?
- (c) What is the probability that a Sophomore graduates?
- (d) What is the probability that a Sophomore quits?
- (e) What is the probability that a Junior graduates?
- (f) What is the probability that a Junior quits?
- (g) What is the probability that a Senior graduates?
- (h) What is the probability that a Senior quits?

(11) A newspaper has obtained the following information about its subscribers: During the first year as subscribers, 20% of all subscribers cancel their subscriptions. Of those who have subscribed for one year, 10% cancel during the second year. Of those who have been subscribing for more than 2 years, 4% will cancel during any given year. On the average, how long does a subscriber subscribe to the newspaper?

(12) A forest consists of two types of trees: those that are 0-1.5 meters and those that are taller than 1.5 meters. Each year, 40% of all 0-1.5 meters tall trees die, 10% are sold for THB 700 each, 30% stay between 0-1.5 meters, and 20% grow to more than 1.5 meters. Each year, 50% of all trees taller than 1.5 meters are sold for THB 1,650, 20% are sold for THB 1,050, and 30% remain in the forest.

(a) What is the probability that a 0-1.5 meters tall tree will die before being sold?

(b) If a tree (less than 1.5 meters) is planted, what is the expected revenue earned from that tree?

(13) An important machine is known to never last more than four months. During its first month of operation it fails 10% of the time. If the machine completes its first month, then it fails during its second month 20% of the time. If the machine completes its second month, then it fails during its third month 50% of the time. If the machine completes its third month, then it is sure to fail by the end of the fourth month. At the beginning of each month we must decide whether or not to replace our machine with a new machine. It costs THB 50,000 to purchase a new machine, but if a machine fails during a month, we incur a cost of THB 100,000 (due to factory downtime) and must replace the machine (at the beginning of the next month) with a new machine. Three maintenance policies are under consideration:

Policy 1: Plan to replace a machine at the beginning of its fourth month of operation.

Policy 2: Plan to replace a machine at the beginning of its third month of operation.

Policy 3: Plan to replace a machine at the beginning of its second month of operation.

Which policy will give the lowest average monthly cost?

Appendix A Random Variable and Probability Distribution Functions

A.1 Random Variable

Random variable is the variable that one cannot predict its value confidently. The outcome of the random variable can only be described by the probability distribution function.

There are two types of random variables.

- (1) Discrete random variable
- (2) Continuous random variable

Let X = Random variable with $f(x)$ as probability distribution function
 $f(x)$ = Probability mass function (pmf) for discrete variable or probability density function (pdf) for continuous variable.
 $F(x)$ = Cumulative probability distribution function (cdf)

$$F(x) = \sum_{min}^x f(x) \quad \text{for discrete random variable} \quad (A1.1)$$

$$F(x) = \int_{min}^x f(x) dx \quad \text{for continuous random variable} \quad (A1.2)$$

A.2 Probability Distribution Functions of Random Variable

The probability distribution functions of the random variable are shown in Figure A.1 and A.2.

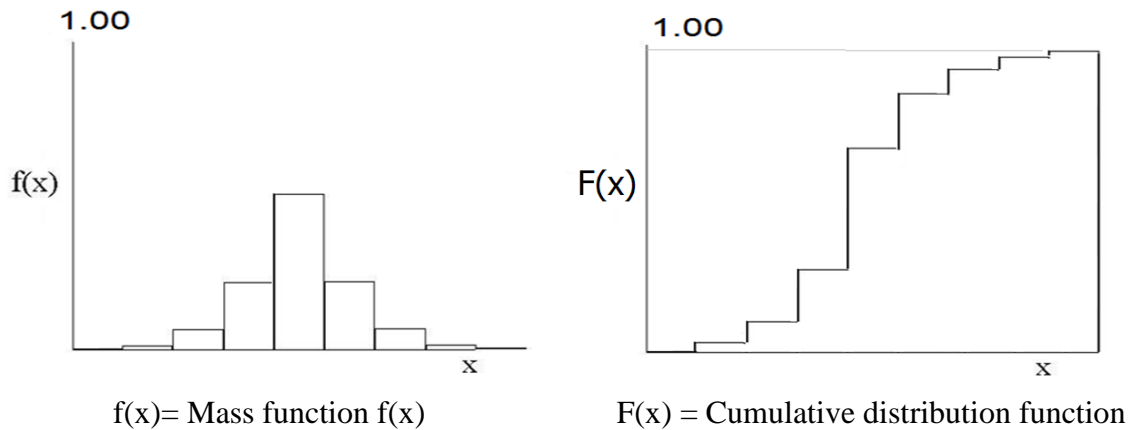


Figure A.1 Discrete probability distribution function of discrete random variable

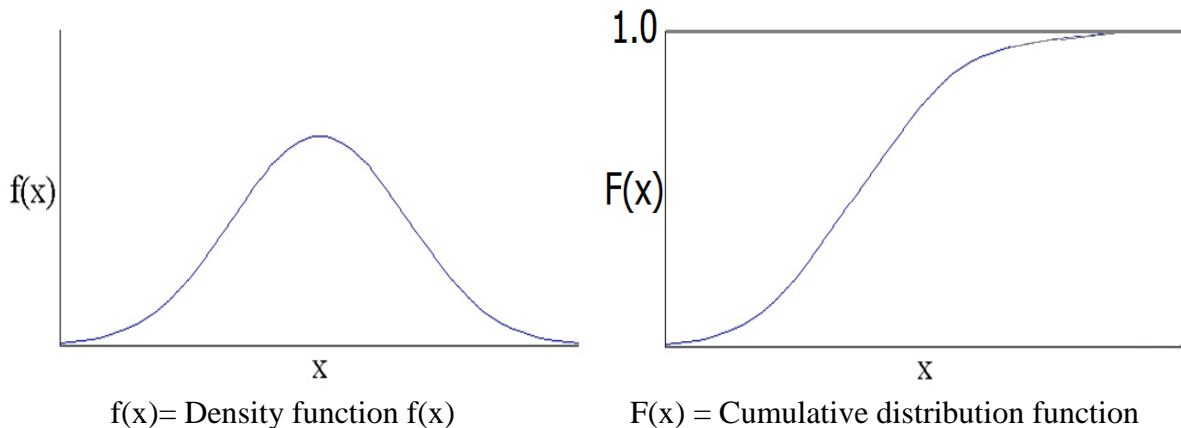


Figure A.2 Continuous probability distribution function of continuous random variable

Table A.1 Discrete Probability Distribution Functions and their properties

No.	Distribution	Range of X	f(x)	μ	σ^2
1	Uniform	(x_1, x_2, \dots, x_k)	$u(x; k) = \frac{1}{k}$	$\frac{1}{k} \sum_{i=1}^k x_i$	$\frac{1}{k} \sum_{i=1}^k x_i^2 - \mu^2$
2	Binomial	0, 1, 2, ..., n	$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$	np	npq
3	Multinomial	$x_1 = 0, 1, 2, \dots, n$ $x_2 = 0, 1, 2, \dots, n$ $x_k = 0, 1, 2, \dots, n$	$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n)$ $= \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$ $\sum_{i=1}^k x_i + x_2 + \dots + x_k = n$ $p_1 + p_2 + \dots + p_k = 1.0$	np_1 np_2 . np_k	$np_1 q_1$ $np_2 q_2$. $np_k q_k$
4	Bivariate Hypergeometric	0, 1, 2, ..., Min(k, n)	$h(x; N, n, k)$ $= \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$	$n \frac{k}{N}$	$\left(\frac{N-n}{N-1}\right) n \frac{k}{N} \left(1 - \frac{k}{N}\right)$
5	Multivariate Hypergeometric	$x_1 = 0, 1, 2, \dots, \text{Min}(a_1, n)$ $x_2 = 0, 1, 2, \dots, \text{Min}(a_2, n)$. $x_k = 0, 1, 2, \dots, \text{Min}(a_k, n)$	$f(x_1, x_2, \dots, x_k; a_1, a_2, \dots, a_k, N, n)$ $= \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_k}{x_k}}{\binom{N}{n}}$ $\sum_{i=1}^k x_i + x_2 + \dots + x_k = n$ $a_1 + a_2 + \dots + a_k = N$	$n \frac{a_1}{N}$ $n \frac{a_2}{N}$. $n \frac{a_k}{N}$	$\left(\frac{N-n}{N-1}\right) n \frac{a_1}{N} \left(1 - \frac{a_1}{N}\right)$ $\left(\frac{N-n}{N-1}\right) n \frac{a_2}{N} \left(1 - \frac{a_2}{N}\right)$. $\left(\frac{N-n}{N-1}\right) n \frac{a_k}{N} \left(1 - \frac{a_k}{N}\right)$
6	Poisson	0, 1, 2,	$p(x; \mu) = \frac{e^{-\mu} \mu^x}{x!}$	μ	μ
7	Negative Binomial	k, k+1,	$b^*(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
8	Geometric	1, 2,	$g(x; p) = p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

The various types of probability distribution function and their properties are presented in Table A.1 and A.2.

$$\mu = E(X) = \sum x f(x) \tag{A.3}$$

$$\sigma^2 = E(X - \mu)^2 = E(X^2) - \mu^2 \quad (\text{A.4})$$

$$E(X^2) = \sum x^2 f(x) \quad (\text{A.5})$$

Table A.2 Continuous Probability Distribution Functions and their properties

No.	Distribution	Range of X	f(x)	μ	σ^2	Remark
1	Uniform	$a \leq x < b$	$\frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	
2	Normal	$-\infty \leq x < \infty$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	
3	Standard Normal	$-\infty \leq z < \infty$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$	$\mu_z=0$	$\sigma_z^2=1$	
4	Gamma	$0 \leq x < \infty$	$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$ ($\alpha > 0, \beta > 0$)	$\alpha\beta$	$\alpha\beta^2$	
5	Exponential	$0 \leq x < \infty$	$\frac{1}{\beta} e^{-\frac{x}{\beta}}$ ($\beta > 0$)	β	β^2	Gamma with $\alpha=1$
6	Chi Square (χ^2)	$0 \leq x < \infty$	$\frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}$ ($\nu = n-1$)	ν	2ν	Gamma with $\alpha = \frac{\nu}{2}$, $\beta = 2$
7	Weibull	$x > 0$	$\alpha\beta x^{\beta-1} e^{-\alpha x^\beta}$ ($\alpha > 0, \beta > 0$)	$\alpha^{-\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right)$	$\alpha^{-\frac{2}{\beta}} \Gamma\left(1 + \frac{2}{\beta}\right) - \left[\alpha^{-\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right)\right]^2$	If $\beta=1$; Weibull=Exponential If $\beta > 1$; Weibull \approx Normal

$$\mu = E(X) = \int x \cdot f(x) dx \quad (\text{A.4})$$

$$\sigma^2 = E(X - \mu)^2 = E(X^2) - \mu^2 \quad (\text{A.5})$$

$$E(X^2) = \int x^2 f(x) \quad (\text{A.6})$$

The Microsoft Excel is a good tool for learning about the probability distribution functions. Table A.3 shows common probability distribution functions.

Table A.3 Probability distribution functions in Microsoft Excel

ฟังก์ชัน	คำอธิบาย	
BETADIST	ส่งกลับฟังก์ชันการแจกแจงแบบเบต้าสะสม (Beta Cumulative Distribution Function)	C
BETAINV	ส่งกลับค่าผกผันของฟังก์ชันการแจกแจงสะสมสำหรับการแจกแจงแบบเบต้าที่ระบุ	C
BINOMDIST	ส่งกลับความน่าจะเป็นของการแจกแจงแบบทวินามสำหรับแต่ละชุด	D
CHIDIST	ส่งกลับค่าความน่าจะเป็นด้านเดียวของการแจกแจงแบบไค-สแควร์ (Chi-squared Distribution)	C
CHIINV	ส่งกลับค่าผกผันของความน่าจะเป็นด้านเดียวของการแจกแจงแบบไค-สแควร์	C
EXPONDIST	ส่งกลับค่าการแจกแจงแบบเอ็กซ์โพเนนเชียล (Exponential Distribution)	C
FDIST	ส่งกลับค่า F ของการแจกแจงความน่าจะเป็น	C
FINV	ส่งกลับค่าผกผันของค่า F ของการแจกแจงความน่าจะเป็น	C
GAMMADIST	ส่งกลับค่าการแจกแจงแบบแกมมา (Gamma Distribution)	C
GAMMAINV	ส่งกลับค่าผกผันของการแจกแจงแบบแกมมาสะสม (Gamma Cumulative Distribution)	C
HYPGEOMDIST	ส่งกลับค่าการแจกแจงแบบไฮเพอร์จีโอเมตริก (Hypergeometric Distribution)	D
LOGINV	ส่งกลับค่าผกผันของการแจกแจงแบบ Lognormal	C
LOGNORMDIST	ส่งกลับค่าการแจกแจงสะสมแบบ Lognormal	C
NEGBINOMDIST	ส่งกลับค่าการแจกแจงแบบทวินามที่เป็นลบ	D
NORMDIST	ส่งกลับค่าการแจกแจงแบบปกติสะสม (Normal Cumulative Distribution)	C
NORMINV	ส่งกลับค่าผกผันของการแจกแจงแบบปกติสะสม	C
NORMSDIST	ส่งกลับค่าการแจกแจงแบบปกติมาตรฐานสะสม (Standard Normal Cumulative Distribution)	C
NORMSINV	ส่งกลับค่าผกผันของการแจกแจงแบบปกติมาตรฐานสะสม	C
POISSON	ส่งกลับค่าการแจกแจงปัวซอง (Poisson Distribution)	D
WEIBULL	ส่งกลับค่าการแจกแจงแบบ Weibull	C

Remark: C = continuous, D = discrete

A.3 Properties of Random Variables and Related Theory

A.3.1 คุณสมบัติสำคัญที่ควรทราบเกี่ยวกับตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)

- (1) $f(x) \geq 0$ for all x
- (2) $f(x)$ คือ Probability Mass Function (pmf) หรือเรียกสั้น ๆ ว่า Probability Function
- (3) $\sum_x f(x) = 1$
- (4) $P(X=x) = f(x)$
- (5) $F(x) = P(X \leq x) = \sum_{t \leq x} f(t) =$ Cumulative Probability Distribution Function
หรือเรียกสั้น ๆ ว่า Cumulative Distribution Function (cdf) ดังแสดงในรูปที่ 1
- $F(x) = P(X \leq x) = \sum_{t \leq x} f(t) =$ Non-exceedence Probability
- (6) $P(X > x) = 1 - P(X \leq x) =$ Exceedence Probability
- (7) $P(a < X \leq b) = \sum_{x=a}^{x=b} f(x) = F(b) - F(a)$

A.3.2 คุณสมบัติสำคัญที่ควรทราบเกี่ยวกับตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

- (1) $f(x) \geq 0$ for all x (Probability Density Function, pdf)
- (2) $\int_{-\infty}^{\infty} f(x) dx = 1$
- (3) $F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx =$ Cumulative Probability Distribution Function (cdf)
 $=$ Non-exceedence probability
- (4) $P(a < X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$

A.3.3 การแจกแจงความน่าจะเป็นร่วม (Joint Probability Distribution)

ถ้าผลการทดลองขึ้นอยู่กับตัวแปรสุ่มสองตัว (X, Y)

$f(x, y) =$ Joint Probability Distribution ของ X และ Y ซึ่งก็คือ $P(X=x, Y=y)$ หรือ $P(X \cap Y)$

(1) Discrete Joint Probability Distributions

คุณสมบัติที่สำคัญที่ควรทราบคือ

$$(1.1) f(x, y) \geq 0 \text{ สำหรับทุกค่าของ } x, y$$

$$(1.2) \sum_x \sum_y f(x, y) = 1$$

$$(1.3) P(X = x, Y = y) = f(x, y)$$

$$(1.4) P[(X, Y) \subset A] = \sum_A \sum f(x, y)$$

(2) Continuous Joint Probability Distributions

คุณสมบัติสำคัญที่ควรทราบคือ

$$(2.1) f(x, y) \geq 0 \text{ สำหรับทุกค่าของ } x, y$$

$$(2.2) f(x, y) = \text{Joint Probability Density Function}$$

$$(2.3) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

$$(2.4) P[(X, Y) \subset A] = \iint_A f(x, y) dx dy$$

(3) Marginal Probability Distributions

Marginal Probability Distributions คือการแจกแจงความน่าจะเป็นของตัวแปรสุ่มตัวใดตัวหนึ่ง เช่น X หรือ Y ซึ่งคำนวณจาก Joint Probability Distributions คุณสมบัติที่สำคัญของ Marginal distribution function แสดงอยู่ในตารางที่ A.4

ตารางที่ A.4 คุณสมบัติของ Marginal Probability Distributions ที่ควรทราบ

Marginal Distribution of X, g(x)	Marginal Distribution of Y, h(y)
(1) $g(x) = \sum_y f(x, y) \dots$ (Discrete) $= \int_y f(x, y) dy \dots$ (Continuous)	$h(y) = \sum_x f(x, y) \dots$ (Discrete) $= \int_x f(x, y) dx \dots$ (Continuous)
(2) $\sum_x \sum_y f(x, y) = \sum_x g(x) = 1$ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_{-\infty}^{\infty} g(x) dx = 1$	$\sum_x \sum_y f(x, y) = \sum_y h(y) = 1$ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_{-\infty}^{\infty} h(y) dy = 1$
(3) $f(x, y) = g(x) h(y)$ ถ้า X และ Y (Independent)	
(4) $f(x, y) = g(x) f(y x)$ ถ้า Y ขึ้นกับ X (Conditional) $= h(y) f(x y)$ ถ้า X ขึ้นกับ Y (Conditional)	

A.3.4 การแจกแจงแบบมีเงื่อนไข (Conditional Probability Distributions)

กรณีที่ตัวแปรสุ่ม X และ Y ไม่เป็นอิสระต่อกัน การแจกแจงความน่าจะเป็นของตัวแปรสุ่มตัวหนึ่ง เช่น Y จะขึ้นอยู่กับค่าของตัวแปรสุ่ม X ว่ามีค่าเท่ากับเท่าใด เช่น

$$P(Y = y | X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{P(X \cap Y)}{P(X = x)} \quad (A.7)$$

คุณสมบัติที่สำคัญของ Conditional Probability Distributions คือ

$$f(y|x) = \frac{f(x, y)}{g(x)} \text{ เมื่อ } g(x) > 0 \quad (A.8)$$

$$f(x|y) = \frac{f(x, y)}{h(y)} \text{ เมื่อ } h(y) > 0 \quad (A.9)$$

A.3.5 การเป็นอิสระของตัวแปรสุ่ม (Independence)

ถ้า Random variable มีลักษณะเป็นตัวแปรสุ่มแบบอิสระ (Independent variable)

Conditional probability distribution = Marginal probability distribution

ดังนั้น ถ้า X เป็นอิสระ (Independent) จาก Y

$$f(x|y) = g(x) \quad (A.10)$$

และ ถ้า Y เป็นอิสระจาก X

$$f(y|x) = h(y) \quad (A.11)$$

จากสมการที่ A.8 และ A.9 ถ้า X และ Y เป็นอิสระต่อกัน

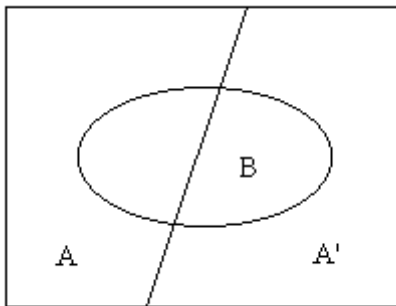
$$f(x, y) = g(x) \cdot h(y) \quad (\text{A.10})$$

A.4 ทฤษฎีของ Bayes (Bayes' Theorem)

ใช้หา Conditional Probability ของเหตุการณ์

กำหนดให้ทราบค่า $P(A)$, $P(B|A)$, $P(B|A')$ ดังรูปที่ A.3

ถ้า A, A' เป็น Mutually Exclusive Event หรือเหตุการณ์ที่ไม่เกิดร่วมกัน และ $P(A)$ หรือ $P(A')$ ไม่เท่ากับ 0 จงหา $P(A|B)$, $P(A'|B)$



รูปที่ A.3 Ven Diagram แสดงเหตุการณ์ A, A' และ B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(B \cap A) + P(B \cap A')} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')} \quad (\text{A.11})$$

$$P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(A' \cap B)}{P(B \cap A) + P(B \cap A')} = \frac{P(B|A')P(A')}{P(B|A')P(A') + P(B|A)P(A)} \quad (\text{A.12})$$

A.5 การคาดคะเนทางคณิตศาสตร์ (Mathematical Expectation)

การคาดคะเนทางคณิตศาสตร์คือวิธีการคำนวณหาค่าพารามิเตอร์ของตัวแปรสุ่มเช่น Mean, Variance และ Covariance โดยวิธีการคำนวณหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มหรือฟังก์ชันของตัวแปรสุ่ม

กฎเกี่ยวกับ Expected Value

$$\begin{aligned} (1) E(c) &= c \\ (2) E(c \cdot u(X)) &= c E(u(X)) \\ (3) E\left(\sum_{i=1}^k u_i(X)\right) &= \sum_{i=1}^k E(u_i(X)) \\ (4) E((aX+b)^n) &= \sum_{i=0}^n \binom{n}{i} a^{n-i} b^i E(X^{n-i}) \\ E((aX+b)^2) &= a^2 E(X^2) + 2ab E(X) + b^2 \end{aligned}$$

A.6 การหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มหรือฟังก์ชันของตัวแปรสุ่ม

ให้ X เป็นตัวแปรสุ่ม มีฟังก์ชันการแจกแจงความน่าจะเป็นเท่ากับ $f(x)$

ถ้า X เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$E(X) = \sum_x x \cdot f(x) = \mu \quad (\text{A.13})$$

= Population Mean ของ X

= μ_x = Moment ที่ 1 รอบจุดศูนย์

= เป็นพารามิเตอร์ที่แสดงจุดศูนย์กลางของการแจกแจง

ให้ \bar{X} คือ Sample mean ของ X หรือ \bar{X} คือ Estimator ของ μ ซึ่งคำนวณหาได้จากสมการ

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{A.14})$$

ถ้า X เป็นตัวแปรสุ่มต่อเนื่อง

$$E(X) = \int_x x \cdot f(x) dx = \mu \quad (\text{A.15})$$

ถ้า Z = u(X) = ฟังก์ชันใด ๆ ของ X

$$E(Z) = E(u(X)) = \mu_z = \mu_{u(x)}$$

$$= \sum_x u(x) f(x) \quad \dots\dots\dots (\text{Discrete}) \quad (\text{A.16})$$

$$= \int_x u(x) \cdot f(x) dx \quad \dots\dots\dots (\text{Continuous}) \quad (\text{A.17})$$

A.7 การหาค่าเฉลี่ย (Expected Value) ของตัวแปรสุ่มที่เป็นฟังก์ชันของตัวแปรสุ่มร่วมกัน 2 ตัว

ให้ X, Y เป็นตัวแปรสุ่ม มีฟังก์ชันความน่าจะเป็นร่วมกันเท่ากับ f(x,y)

ให้ Z = u(X,Y) = ฟังก์ชันใด ๆ ของ X, Y

$$E(Z) = E(u(X,Y)) = \mu_z = \mu_{u(x,y)} \quad (\text{A.18})$$

(1) กรณีตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)

$$E(Z) = \sum_y \sum_x u(x,y) \cdot f(x,y) \quad (\text{A.19})$$

$$E(X) = \mu_x = \sum_y \sum_x x \cdot f(x,y)$$

$$= \sum_x x \cdot g(x) \quad (\text{A.20})$$

$$E(Y) = \mu_y = \sum_y \sum_x y \cdot f(x,y)$$

$$= \sum_y y \cdot h(y) \quad (\text{A.21})$$

ถ้า Z = XY

$$E(Z) = E(XY)$$

$$= \sum_y \sum_x xy \cdot f(x,y) \quad (\text{A.22})$$

(2) กรณีตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

$$E(Z) = \int_y \int_x u(x,y) \cdot f(x,y) dx dy \quad (\text{A.23})$$

$$\begin{aligned}
 E(X) &= \int_y \int_x x f(x,y) dx dy \\
 &= \int_x x \cdot g(x) dx
 \end{aligned} \tag{A.24}$$

$$\begin{aligned}
 E(Y) &= \int_y \int_x y f(x,y) dx dy \\
 &= \int_y y \cdot h(y) dy
 \end{aligned} \tag{A.25}$$

ถ้า $Z = XY$

$$\begin{aligned}
 E(Z) &= E(XY) \\
 &= \int_y \int_x xy \cdot f(x,y) dx dy
 \end{aligned} \tag{A.26}$$

ถ้า X, Y เป็นอิสระต่อกัน (Independent)

$$f(x,y) = g(x) \cdot h(y) \tag{A.27}$$

$$\begin{aligned}
 E(Z) = E(XY) &= \int_y \int_x xy \cdot g(x)h(y)dx dy = \int_y (\int_x x \cdot g(x)dx)y \cdot h(y)dy \\
 &= \int_y E(X) y \cdot h(y)dy = E(X) \int_y y \cdot h(y)dy \\
 E(XY) &= E(X) \cdot E(Y)
 \end{aligned} \tag{A.28}$$

A.8 การหาความแปรปรวน (Variance) ของตัวแปรสุ่มและฟังก์ชันของตัวแปรสุ่ม

ความแปรปรวน (Variance) คือพารามิเตอร์ของตัวแปรสุ่ม (X) หรือของฟังก์ชันการแจกแจงความน่าจะเป็น (f(x)) ซึ่งแสดงรูปร่างของการแจกแจงความน่าจะเป็น สามารถคำนวณหาได้โดยใช้ Expected Value

$$\begin{aligned}
 \text{ให้ } \text{Var} (X) = \sigma^2 &= \text{ความแปรปรวนของ } X \text{ (Population Variance)} \\
 \text{Var} (X) = \sigma^2 &= E(X-\mu)^2 \tag{A.29} \\
 &= \sum_x (x-\mu)^2 \cdot f(x) \quad \dots\dots\dots \text{Discrete} \\
 &= \int_x (x-\mu)^2 \cdot f(x) dx \quad \dots\dots\dots \text{Continuous} \\
 &= \text{Moment ที่ 2 รอบ } \mu
 \end{aligned}$$

σ คือ ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของ X

ถ้า $Z = u(X) =$ ฟังก์ชันใด ๆ ของ X

$$\begin{aligned}
 \text{Var} (Z) &= \text{Var} (u(X)) = \sigma_{u(x)}^2 \\
 &= E(u(X)-\mu_{u(x)})^2
 \end{aligned} \tag{A.30}$$

ให้ S^2 คือ Sample variance ของ X หรือ S^2 คือ Unbiased Estimator ของ σ^2 ซึ่งคำนวณหาได้จาก
สมการ

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \tag{A.31}$$

A.9 กฎเกี่ยวกับ Variance

(1) $Z = aX+b$

$$\sigma_z^2 = \sigma_{ax+b}^2 = a^2 \sigma^2$$

(2) $Z = aX+bY$ เมื่อ X, Y คือตัวแปรสุ่ม 2 ตัว

$$\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab\sigma_{xy}$$

เมื่อ σ_{xy} = Covariance ของ X และ Y

(3) $Z = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ เมื่อ X_1, X_2, \dots, X_n คือตัวแปรสุ่ม

$$\sigma_z^2 = a_1^2 \sigma_{x1}^2 + a_2^2 \sigma_{x2}^2 + \dots + a_n^2 \sigma_{xn}^2$$

ถ้า X_1, X_2, \dots, X_n independent

Covariance จะมีค่าเป็น 0

A.10 ความแปรปรวนร่วมกันของตัวแปรสุ่ม (Covariance)

ให้ X และ Y เป็นตัวแปรสุ่มมีฟังก์ชันความน่าจะเป็นร่วมกันเท่ากับ $f(x,y)$ และมีค่าเฉลี่ยเท่ากับ μ_x และ μ_y ตามลำดับ

ให้ $Cov(X,Y) = \sigma_{xy}$ = ความแปรปรวนร่วมกันของตัวแปรสุ่ม X และ Y (Population Covariance)

$$\begin{aligned} Cov(X,Y) &= \sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] \\ &= \sum_y \sum_x (x-\mu_x)(y-\mu_y) f(x,y) \dots \dots \dots \text{Discrete} \end{aligned} \tag{A.32}$$

$$= \int_y \int_x (x-\mu_x)(y-\mu_y) f(x,y) dx dy \dots \dots \dots \text{Continuous} \tag{A.33}$$

หรือ $E[(X-\mu_x)(Y-\mu_y)] = E(XY) - \mu_x \mu_y$

ให้ S_{xy} คือ Sample covariance ของ X และ Y หรือ S_{xy} คือ Estimator ของ σ_{xy} ซึ่งคำนวณหาได้

จากสมการ

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{A.34}$$

ถ้า X, Y = independent

$$E(XY) = E(X) E(Y) = \mu_x \mu_y \tag{A.35}$$

$$\sigma_{xy} = 0 \tag{A.36}$$

A.11 สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

ให้ X และ Y เป็นตัวแปรสุ่มมีฟังก์ชันความน่าจะเป็นร่วมเท่ากับ $f(x,y)$ และมีค่าเฉลี่ยเท่ากับ μ_x และ μ_y ตามลำดับ

ให้ ρ = ค่าสัมประสิทธิ์สหสัมพันธ์ (Population correlation coefficient)

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} \quad (\text{A.37})$$

ให้ r คือ Sample correlation coefficient ของ X และ Y หรือ r คือ Estimator ของ ρ ซึ่งคำนวณหาได้จากสมการ

$$r = \frac{S_{xy}}{\sqrt{S_x^2 S_y^2}} \quad \text{หรือ}$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{A.38})$$

A.12 สัมประสิทธิ์ความเบ้ (Skewness Coefficient)

สัมประสิทธิ์ความเบ้ (Skewness Coefficient) คือพารามิเตอร์ของตัวแปรสุ่ม (X) ซึ่งแสดงความเบ้ (Skewness) ของการแจกแจงความน่าจะเป็น สามารถคำนวณหาได้จากโมเมนต์ที่ 3 รอบจุดศูนย์กลาง ดังนี้

ให้ γ = ค่าสัมประสิทธิ์ความเบ้ของ X (Population Skewness Coefficient)

$$\gamma = \frac{E(X - \mu)^3}{\sigma^3} \quad (\text{A.39})$$

ให้ g คือ Sample skewness coefficient ของ X หรือ g คือ Unbiased Estimator ของ γ ซึ่งคำนวณหาได้จากสมการ

$$g = \frac{N \cdot \sum_{i=1}^n (x_i - \bar{x})^3}{(N-1)(N-2)S^3} \quad (\text{A.40})$$

A.13 สัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Autocorrelation Coefficient)

ให้ x_t คืออนุกรมเวลาของ x เมื่อ $t = 1$ ถึง n หรือ (x_1, x_2, \dots, x_n)

ให้ c_k คือ Lag k autocovariance ของ x_t ซึ่งคำนวณหาได้จากสมการ

$$C_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{n-k-1} \quad \text{เมื่อ } 0 < k < n \quad \text{หรือ} \quad (A.41)$$

$$C_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x}_t)(x_{t+k} - \bar{x}_{t+k})}{n-k-1} \quad (A.42)$$

ให้ r_k คือ Lag k autocorrelation coefficient หรือ Correlogram ของ x_t ซึ่งคำนวณหาได้จากสมการ

$$r_k = \frac{C_k}{C_0} \quad \text{หรือ}$$

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x}_t)(x_{t+k} - \bar{x}_{t+k})}{\sqrt{\sum_{t=1}^{n-k} (x_t - \bar{x}_t)^2 \sum_{t=1}^{n-k} (x_{t+k} - \bar{x}_{t+k})^2}} \quad (A.43)$$

A.14 แบบฝึกหัด

(1) กำหนดให้อนุกรมเวลา X_t และ Y_t มีค่าดังตารางที่ A.5 จงคำนวณหา

1. ค่าเฉลี่ย (mean)
2. ค่าความแปรปรวน (variance)
3. ค่าความแปรปรวนร่วม (covariance)
4. ค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient)
5. ค่าสัมประสิทธิ์ความเบ้ (skewness coefficient)
6. ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (lag k autocorrelation function) เมื่อ $k=0$ to 10

พร้อมหาข้อสรุปเกี่ยวกับ อนุกรมเวลา X_t และ Y_t ดังต่อไปนี้

1. อนุกรมเวลาทั้งสองมีการแจกแจงแบบปกติหรือไม่
2. อนุกรมเวลา X_t และ Y_t เป็นอิสระต่อกัน
3. อนุกรมเวลา X_t และ Y_t มีลักษณะเป็นแบบ Stationary หรือไม่ (Stationary คืออนุกรมเวลาที่ค่าพารามิเตอร์ไม่เปลี่ยนแปลงไปตามเวลา)

ตารางที่ A.5 ข้อมูลสำหรับแบบฝึกหัดข้อ (1)

t	X_t	Y_t
1	9.545	14.24
2	11.12	13.44
3	10.68	12.18
4	8.215	15.73
5	7.199	14.63
6	12.63	12.79
7	11.58	13.93
8	8.891	19.85
9	10.67	16.13
10	9.639	10.52
11	10.31	16.76
12	12.73	18.37
13	5.996	15.38
14	12.76	14.39
15	10.92	15.02
16	10.07	14.68
17	9.248	14.17
18	8.773	9.482
19	8.612	22.05
20	6.906	14.04
21	9.31	14.91
22	7.975	14.78
23	8.696	16.14
24	10.21	11.5
25	13.09	16.38
26	10.03	14.04
27	7.519	17.72
28	9.797	17.38
29	10.49	18.34
30	5.84	12.21

(2) Construct graphs showing the probability mass function (pmf) or probability density function (pdf) of the following distribution functions.

(2.1) Discrete probability distribution

pmf	Given data
1. $u(x;k)$	Let X be Uniform random variable showing the numbers $\{1, 2, 3, 4, 5, 6\}$ obtained from a dice tossing experiment.
2. $b(x; n, p)$	Let X be Binomial random variable showing the no. of getting 6 from tossing a dice 10 times.
3. $f(x_1, x_2, x_3; p_1, p_2, p_3, n)$	A box contains 20 color balls; 8 red, 7 green and 5 blue. Let X_1, X_2, X_3 be Multinomial random variables showing the no. of red, green and blue balls drawn from the experiment of drawing one ball at a time for 3 times. The ball has to return to the box before next drawing.
4. $h(x; N, n, k)$	A box containing 20 color balls; 8 red, 7 green and 5 blue. Let X be Bivariate hypergeometric random variable showing

	the no. of red balls drawn from the experiment of drawing one ball at a time for 10 times with no return ball.																																										
5. $f(x_1, x_2, x_3; a_1, a_2, a_3, N, n)$	A box containing 20 color balls; 8 red, 7 green and 5 blue. Let X_1, X_2, X_3 be Multivariate hypergeometric random variables showing the no. of red, green and blue balls drawn from the experiment of drawing one ball at a time for 3 times with no return ball.																																										
6. $p(x; \mu)$	Let X be Poisson random variable representing the no. of rainy days in June. The historical data of June rainy days are given in Table A.6. <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="6">Table A.6 June rainy days(days)</th> </tr> <tr> <th>Year</th> <th>x(t)</th> <th>Year</th> <th>x(t)</th> <th>Year</th> <th>x(t)</th> </tr> </thead> <tbody> <tr> <td>2548</td> <td>2</td> <td>2553</td> <td>7</td> <td>2558</td> <td>0</td> </tr> <tr> <td>2549</td> <td>0</td> <td>2554</td> <td>0</td> <td>2559</td> <td>0</td> </tr> <tr> <td>2550</td> <td>1</td> <td>2555</td> <td>10</td> <td>2560</td> <td>8</td> </tr> <tr> <td>2551</td> <td>0</td> <td>2556</td> <td>2</td> <td>2561</td> <td>4</td> </tr> <tr> <td>2552</td> <td>3</td> <td>2557</td> <td>5</td> <td>2562</td> <td>3</td> </tr> </tbody> </table>	Table A.6 June rainy days(days)						Year	x(t)	Year	x(t)	Year	x(t)	2548	2	2553	7	2558	0	2549	0	2554	0	2559	0	2550	1	2555	10	2560	8	2551	0	2556	2	2561	4	2552	3	2557	5	2562	3
Table A.6 June rainy days(days)																																											
Year	x(t)	Year	x(t)	Year	x(t)																																						
2548	2	2553	7	2558	0																																						
2549	0	2554	0	2559	0																																						
2550	1	2555	10	2560	8																																						
2551	0	2556	2	2561	4																																						
2552	3	2557	5	2562	3																																						
7. $b^*(x; k, p)$	Let X be Negative binomial random variable showing the no. of dice tossing experiment to obtain 6 for 5 times.																																										
8. $g(x; p)$	Let X be geometric random variable showing the no. of dice tossing experiment to obtain 6 for the first time.																																										

(2.2) Continuous probability distribution

pdf	Given data																																										
1. $f(x) = \frac{1}{b-a}$	Let X be Uniform random variable showing the daily water demand of a man which varies between 50-650 liters/day.																																										
2. $f(x; \mu, \sigma)$	Let X be Normal random variable showing the annual rainfall. The historical records of annual rainfall are given in Table A.7. <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="6">Table A.7 Annual rainfall(mm)</th> </tr> <tr> <th>Year</th> <th>x(t)</th> <th>Year</th> <th>x(t)</th> <th>Year</th> <th>x(t)</th> </tr> </thead> <tbody> <tr> <td>2548</td> <td>800</td> <td>2553</td> <td>990</td> <td>2558</td> <td>1,700</td> </tr> <tr> <td>2549</td> <td>915</td> <td>2554</td> <td>1,050</td> <td>2559</td> <td>1,900</td> </tr> <tr> <td>2550</td> <td>700</td> <td>2555</td> <td>1,100</td> <td>2560</td> <td>800</td> </tr> <tr> <td>2551</td> <td>1,200</td> <td>2556</td> <td>1,200</td> <td>2561</td> <td>750</td> </tr> <tr> <td>2552</td> <td>950</td> <td>2557</td> <td>1,500</td> <td>2562</td> <td>930</td> </tr> </tbody> </table>	Table A.7 Annual rainfall(mm)						Year	x(t)	Year	x(t)	Year	x(t)	2548	800	2553	990	2558	1,700	2549	915	2554	1,050	2559	1,900	2550	700	2555	1,100	2560	800	2551	1,200	2556	1,200	2561	750	2552	950	2557	1,500	2562	930
Table A.7 Annual rainfall(mm)																																											
Year	x(t)	Year	x(t)	Year	x(t)																																						
2548	800	2553	990	2558	1,700																																						
2549	915	2554	1,050	2559	1,900																																						
2550	700	2555	1,100	2560	800																																						
2551	1,200	2556	1,200	2561	750																																						
2552	950	2557	1,500	2562	930																																						
3. $f(x; \alpha, \beta)$	Let X be Gamma random variable showing the annual maximum event rainfall. The historical records of annual maximum event rainfall are given in Table A.8. <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="6">Table A.8 Annual maximum event rainfall (mm)</th> </tr> <tr> <th>Year</th> <th>x(t)</th> <th>Year</th> <th>x(t)</th> <th>Year</th> <th>x(t)</th> </tr> </thead> <tbody> <tr> <td>2548</td> <td>380</td> <td>2553</td> <td>375</td> <td>2558</td> <td>515</td> </tr> <tr> <td>2549</td> <td>515</td> <td>2554</td> <td>450</td> <td>2559</td> <td>614</td> </tr> <tr> <td>2550</td> <td>500</td> <td>2555</td> <td>540</td> <td>2560</td> <td>371</td> </tr> <tr> <td>2551</td> <td>200</td> <td>2556</td> <td>380</td> <td>2561</td> <td>275</td> </tr> <tr> <td>2552</td> <td>275</td> <td>2557</td> <td>430</td> <td>2562</td> <td>575</td> </tr> </tbody> </table>	Table A.8 Annual maximum event rainfall (mm)						Year	x(t)	Year	x(t)	Year	x(t)	2548	380	2553	375	2558	515	2549	515	2554	450	2559	614	2550	500	2555	540	2560	371	2551	200	2556	380	2561	275	2552	275	2557	430	2562	575
Table A.8 Annual maximum event rainfall (mm)																																											
Year	x(t)	Year	x(t)	Year	x(t)																																						
2548	380	2553	375	2558	515																																						
2549	515	2554	450	2559	614																																						
2550	500	2555	540	2560	371																																						
2551	200	2556	380	2561	275																																						
2552	275	2557	430	2562	575																																						

4. $f(x; \beta)$	Let X be Exponential random variable showing the annual maximum event rainfall. The historical records of annual maximum event rainfall are given in Table A.8.
5. $f(x; \nu)$	Let X be Chi-square random variable showing the annual maximum event rainfall. The historical records of annual maximum event rainfall are given in Table A.8.
6. $f(x; \alpha, \beta)$	Let X be Weibull random variable showing the annual maximum event rainfall. The historical records of annual maximum event rainfall are given in Table A.8. Assume $\beta=2$.

(2.3) The demand and supply of a water resource system are randomly distributed with the statistics given in Table below. Determine the risk of water shortage if the probability distribution functions of both demand and supply are assumed to be (a) Uniform (b) Normal and (c) Gamma.

	Demand	Supply
Mean	2.5	5
Standard deviation	0.9	1.5

(2.4) The spillway of a dam was designed as an Ogee-type with the design discharging capacity of 2,000 m³/s. The spillway freeboard is 3 m. The spillway discharge formula is given below.

$$Q = CLH^{1.5}$$

Where

Q = spillway discharge in m³/s

C = discharge coefficients = 2.0

L = width of the spillway crest = 50m.

H = design head in m.

There are uncertainty of the discharge (Q) and the discharge coefficient (C) data in the design of spillway. It is estimated that the coefficient of variation (CV) of Q and C are 0.3 and 0.4 respectively. What is the risk of dam over-topping from the uncertainties of Q and C ? (Hint: Normal distribution is assumed.)

(2.5) Let $Y = Y' + \varepsilon$ where Y = observed variate, Y' = model variate

ε = model error. Prove that R^2 is always greater than NSE.

$$R^2 = \frac{\text{Cov}(Y, Y')^2}{\text{Var}(Y)\text{Var}(Y')} \quad (1)$$

$$NSE = 1 - \frac{\sum(y-y')^2}{\sum(y-\bar{y})^2} \quad (2)$$

If the model error term (ε) is independent with the mean equals to zero, $R^2 = NSE = \frac{\text{Var}(\varepsilon)}{\text{Var}(Y)}$.

(2.6) Determine the population mean (μ) and variance (σ^2) of the stream flow (z), soil moisture storage (SW) and Groundwater storage (GW) of the watershed model in Assignment No.1 of Chapter 1

(2.7) From the data given in Table A.9, determine the parameters of the ARMA(1,1) as shown in the textbook (Chapter 1) and the parameters of ARMA(2,2) as given in Assignment No.1 of Chapter 1.

t	x(t)	z(t)
0	1,304.40	500.00
1	876.00	428.01
2	996.00	483.53
3	1,183.00	562.44
4	1,591.00	714.31
5	1,561.00	753.37
6	1,717.00	836.82
7	519.00	516.09
8	1,228.00	687.85
9	1,618.00	814.64
10	906.00	632.27
11	1,987.00	938.84
12	1,966.00	983.21
13	1,011.00	735.99
14	681.00	611.11
15	1,355.00	772.81
16	1,244.00	747.55
17	1,535.00	834.62
18	628.00	579.79
19	1,275.00	733.30
20	849.00	611.84
21	970.00	627.65
22	779.00	561.22
23	1,972.00	900.36
24	1,257.00	742.42
25	1,607.00	849.78
26	1,981.00	984.86
27	1,672.00	932.91
28	1,520.00	901.37
29	1,652.00	943.09
30	996.00	755.92

Appendix B

Review Basic Matrix Operation

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (3 \times 3 \text{ square matrix})$$

$$A^T = A \quad (\text{Symmetric matrix}) \quad (\text{B.1})$$

$$(aA + bB)^T = aA^T + bB^T \quad (\text{B.2})$$

$$(AB)^T = B^T A^T \quad (\text{B.3})$$

$$AB = I \text{ then } B = A^{-1} \quad (\text{Inverse of matrix A}) \quad (\text{B.4})$$

$$b_{ij} = \frac{-1^{i+j} d_{ij}}{|A|} \quad (\text{B.5})$$

d_{ij} = determinant of sub-matrix formed by eliminating the i row and j column of matrix A

$|A|$ = determinant of matrix

$$|A| = \sum_j (-1)^{1+j} a_{1j} \cdot d_{1j} \quad ; j=1 \text{ to } 3 \quad (\text{B.6})$$

$$\begin{aligned} |A| &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} - b \begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} + c \begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix} \\ &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= aei + bfg + cdh - ceg - bdi - afh. \end{aligned}$$

$$AA^{-1} = A^{-1}A = I \quad (\text{B.7})$$

Power of matrix

$$A^0 = I$$

$$A^1 = A$$

$$A^2 = A.A$$

$$A^k = A.A.A \dots A \quad (\text{multiplying } k \text{ of Matrix } A) \quad (\text{B.8})$$

If $|A| \neq 0$ then $A = \text{non-singular matrix}$

Matrix A is positive definite if

$$|a_{11}| > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0 \quad (\text{B.9})$$

Matrix A is positive semi-definite if

$$|a_{11}| \geq 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \geq 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \geq 0 \quad (\text{B.10})$$

Appendix C

First Order Analysis of Uncertainty

C.1 นิยาม First Order Analysis of Uncertainty

First Order Analysis of Uncertainty (FAU) คือการประมาณค่าความผิดพลาดของตัวแปรตาม (Dependent Variable) ในแบบจำลอง ซึ่งเป็นฟังก์ชันของตัวแปรอิสระ (Independent Variables) ตั้งแต่ 1 ตัวขึ้นไป

$$w = f(x) \quad (C.1)$$

X = ตัวแปรอิสระ (Independent variable)

W = ตัวแปรตาม (Dependent variable)

f(.) = แบบจำลอง

การประมาณค่า w อาจมีความผิดพลาดได้จาก 2 สาเหตุ คือ

- 1) แบบจำลอง หรือฟังก์ชัน $f(.)$ ไม่ถูกต้อง หรือ
- 2) ค่า x ไม่ถูกต้อง (uncertainty)

สมมติว่า $f(.)$ ถูกต้อง

ถ้า $x = \bar{x}$

$$w = \bar{w} = f(\bar{x}) \quad (C.2)$$

แต่ถ้า x จริง มีค่าต่างจาก \bar{x} จะมีผลทำให้การประมาณค่า w ผิดพลาดได้

C.2 Taylor Series Expansion

Taylor Series Expansion ของ $f(x)$ เมื่อ x แปรผันรอบค่า \bar{x}

$$w = f(\bar{x}) + \frac{df}{dx}(x - \bar{x}) + \frac{1}{2!} \frac{d^2f}{dx^2}(x - \bar{x})^2 + \dots \quad (C.3)$$

ถ้าพิจารณาว่า Second Order Term หรือเทอมที่มี Order สูงขึ้น มีค่าน้อยมากจนตัดทิ้งได้

$$w - \bar{w} = \frac{df}{dx}(x - \bar{x}) = \text{Error} = \varepsilon \quad (C.4)$$

$$E(w - \bar{w})^2 = E\left(\frac{df}{dx}(x - \bar{x})\right)^2 = \sigma_w^2 = \sigma_\varepsilon^2$$

$$\sigma_w^2 = \left(\frac{df}{dx}\right)^2 E(x - \bar{x})^2 = \left(\frac{df}{dx}\right)^2 \sigma_x^2 \quad (\text{C.5})$$

ถ้า $w = f(x_1, x_2, \dots, x_n)$

$$\sigma_w^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_{x_n}^2 \quad (\text{C.6})$$

C.3 การวิเคราะห์ความไม่แน่นอนอันดับที่ 1 ของสมการ Manning

สมการ Manning

$$Q = \frac{1}{n} AR^{2/3} S^{1/2}$$

ถ้า y เป็นตัวแปรตาม และ Q เป็นตัวแปรอิสระที่มีความไม่แน่นอน

$$y = f(Q)$$

จากสมการที่ (C.5)

$$\sigma_y^2 = \left(\frac{\partial f}{\partial Q}\right)^2 \sigma_Q^2 \quad (\text{C.7})$$

$$Q = \frac{1}{n} AR^{2/3} S^{1/2}$$

$$\begin{aligned} \frac{dQ}{dy} &= \left(\frac{1}{n} S^{1/2}\right) \frac{dAR^{2/3}}{dy} = \left(\frac{1}{n} S^{1/2}\right) \left(\frac{2}{3} AR^{-1/3} \frac{dR}{dy} + R^{2/3} \frac{dA}{dy}\right) \\ &= \left(\frac{1}{n} AR^{2/3} S^{1/2}\right) \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right) = Q \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right) \end{aligned}$$

$$\frac{df}{dQ} = \frac{dy}{dQ} = \frac{1}{Q \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)} \quad (\text{C.8})$$

$$\sigma_y^2 = \left(\frac{1}{Q \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)}\right)^2 \sigma_Q^2$$

$$\sigma_y^2 = \frac{\sigma_Q^2}{Q^2 \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)^2} = \frac{CV_Q^2}{\left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)^2} \quad (\text{C.9})$$

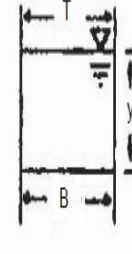
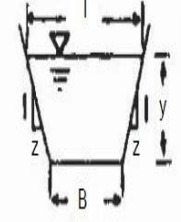
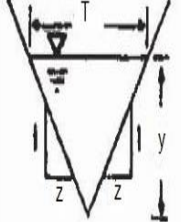
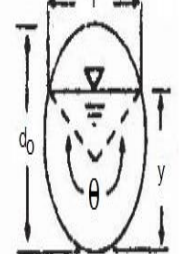
Section	Rectangle	Trapezoidal	Triangle	Circular
				
$\frac{2dR}{3Rdy} + \frac{1dA}{A dy}$	$\frac{5B + 6y}{3y(B + 2y)}$	$\frac{(B + 2zy)(5B + 6y\sqrt{1+z^2}) + 4zy^2\sqrt{1+z^2}}{3y(B + zy)(B + 2y\sqrt{1+z^2})}$	$\frac{8}{3y}$	$\frac{4(2\sin\theta + 3\theta - 5\theta\cos\theta)}{3d_0(\theta - \sin\theta)\sin(\theta/2)}$ where $\theta = 2\cos^{-1}\left(1 - \frac{2y}{d_0}\right)$

Figure C.1 Section Factor (K)

ถ้า y เป็นตัวแปรตาม และ Q, n, S เป็นตัวแปรอิสระที่มีความไม่แน่นอน

$$\sigma_y^2 = \frac{cV_Q^2 + cV_n^2 + \frac{1}{4}cV_S^2}{\left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)^2} \tag{C.10}$$

$$n = \frac{1}{Q} AR^{2/3} S^{1/2}$$

$$\begin{aligned} \frac{dn}{dy} &= \left(\frac{1}{Q} S^{1/2}\right) \frac{dAR^{2/3}}{dy} \\ &= \left(\frac{1}{Q} S^{1/2}\right) \left(\frac{2}{3} AR^{-1/3} \frac{dR}{dy} + R^{2/3} \frac{dA}{dy}\right) \\ &= \left(\frac{1}{Q} AR^{2/3} S^{1/2}\right) \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right) \\ &= n \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right) \end{aligned}$$

Let $K = \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)$ (C.11)

$$\frac{dn}{dy} = nK \tag{C.12}$$

$$\frac{df}{dn} = \frac{dy}{dn} = \frac{1}{n \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)}$$

$$S^{-1/2} = AR^{2/3}$$

$$\frac{dS^{-1/2}}{dy} = \left(\frac{1}{nQ}\right) \frac{dAR^{2/3}}{dy}$$

$$\frac{1}{2}S^{-3/2} \frac{dS}{dy} = \left(\frac{1}{nQ}\right) \left(\frac{2}{3}AR^{-1/3} \frac{dR}{dy} + R^{2/3} \frac{dA}{dy}\right)$$

$$\frac{dS}{dy} = (2S^{3/2}) \left(\frac{1}{nQ} AR^{2/3}\right) \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right) = 2S \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)$$

$$\frac{df}{dS} = \frac{dy}{dS} = \frac{1}{2S \left(\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy}\right)}$$

$$\frac{df}{dS} = \frac{1}{2SK}$$

(C.13)

C.4 ตัวอย่างการใช้ FAU วิเคราะห์ความเสี่ยง

ทางน้ำรูปสี่เหลี่ยมผืนผ้า กว้าง (B) = 50 เมตร และมีข้อมูลที่เกี่ยวข้องกับการไหลในทางน้ำคือ

ค่าความลาดชัน (S) = 1% $Q_{\text{design}} = 5,000$ cms $CV_Q = 30\%$ $n = 0.035$ และ $CV_n = 15\%$ จงหาค่าความแปรปรวนของความลึกในการไหลของน้ำในทางน้ำ (S_y^2) อันเป็นผลจากความไม่แน่นอนของอัตราการไหล (Q) และค่า n

ถ้าบ้านที่ปลูกริมทางน้ำมีพื้นสูงกว่า $y_{\text{design}} = 1$ เมตร จงหาความน่าจะเป็นที่น้ำจะท่วมบ้าน สมมติว่า Y มีการแจกแจงแบบปกติ

วิธีทำ

$$Q = \frac{1.49}{n} AR^{2/3} S^{1/2}$$

$$Q = 5,000, B = 50, n = 0.035, S = 1\%$$

By trial&error, $y = 9.6155$ ฟุต

$$\frac{2}{3R} \frac{dR}{dy} + \frac{1}{A} \frac{dA}{dy} = \frac{3B+6y}{3y(B+2y)} = \frac{3*50+6*9.6155}{3*9.6155(50+2*9.6155)} = 0.104$$

$$S_y = 3.2251$$

บ้านจะถูกน้ำท่วมถ้า $y > 9.6155 + 1 = 10.6155$ เมตร

$$P(y > 10.6155) = P[z > (10.6155 - 9.6155)/3.2251] = P(z > 0.3101) = 1 - 0.6217 = 0.3783$$

C.5 ความเสี่ยง (Risk)

ความเสี่ยงคือความน่าจะเป็นของเหตุการณ์ที่โหลด (Load) ที่กระทำต่อระบบมีค่าสูงกว่าความสามารถในการรับโหลด (Capacity) และมีผลทำให้ระบบล้มเหลว หรือ Risk of Failure

$$\text{Risk} = P(\text{Load} > \text{Capacity}) \quad (\text{C.14})$$

Let L = Load and c^* = Capacity

กรณีที่ Load มีความไม่แน่นอน แต่ Capacity = c^* (มีค่าคงที่)

$$\text{Risk} = P(L > c^*) = \int_{c^*}^{\text{upper}} f(\text{Load}) dL$$

กรณีที่ทั้ง Load และ Capacity ต่างก็มีความไม่แน่นอน จะสามารถหาค่าความเสี่ยงรวม (Composite Risk) ได้จากสมการ

$$\text{Risk} = P(L > C)$$

$$\text{Risk} = \int_{\text{lower}}^{\text{upper}} \int_C^{\text{upper}} [f(L)dL]f(C)dC \quad (\text{C.15})$$

จาก Figure C.2 ถ้า Load มีการแจกแจงแบบปกติ มี Mean= 3 และ Standard Deviation= 1 และ Capacity (c^*)= 5

$$\text{Risk} = P(L > 5) = 0.023$$

จาก Figure C.2 ถ้า Load และ Capacity มีการแจกแจงแบบปกติ $\mu_L = 3, \sigma_L = 1, \mu_C = 5, \sigma_C = 0.75$

$$\text{Risk} = P(L > C) = 0.052$$

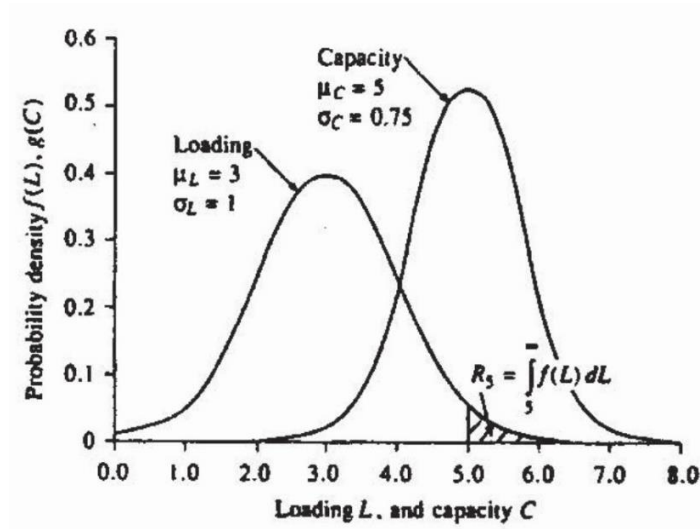


Figure C.2 Composite Risk