บทที่ 5

การพัฒนาแบบจำลองในการพยากรณ์

ดังที่ได้กล่าวมาแล้วว่าอนุกรมเวลาโดยทั่วไปประกอบด้วยส่วนที่เป็นดีเทอร์มินีสติก และส่วนที่เป็นสโตแคสติก ในส่วนที่เป็นดีเทอร์มินีสติกอาจจะประกอบไปด้วย ส่วนแนวโน้ม (Trend) ส่วนฤดูกาล (Seasonal) และส่วนวงจร (Cycle) ในส่วนที่เป็นสโตแคสติกก็อาจจะประกอบด้วยกระบวนการทางสโตแคสติกที่เป็นอิสระ (การสุ่มอย่างเดียว) หรือกระบวนการทางสโตแคสติกที่แปรตามการสุ่ม สามารถเขียนอยู่ในรูปสมการได้ว่า

$$Z_{forecasting} = Z_{trend} + Z_{seasonal} + Z_{cycle} + Z_{stochastic} + \varepsilon \qquad (5\text{-}1)$$

ที่ $Z_{forecasting}$ เป็นปริมาณฝนที่คาดหมายที่ประกอบไปด้วยส่วนประกอบต่างๆ ของอนุกรม

Z_{trend} เป็นปริมาณฝนที่เกิดจากส่วนแนวโน้ม (Trend)

$Z_{seasonal}$ เป็นปริมาณฝนที่เกิดจากส่วนฤดูกาล (Seasonal)

Z_{cycle} เป็นปริมาณฝนที่เกิดจากส่วนวงจร (Ceycle)

$Z_{stochastic}$ เป็นปริมาณฝนที่เกิดจากส่วนสโตแคสติก (Stochastic)

E เป็นปริมาณฝนที่เกิดจากการสุ่มที่เหมาะสม

จากการตรวจสอบข้อมูลเบื้องต้น และการวิเคราะห์ทางสโตแคสติกของอนุกรมชุดต่างๆ ของแต่ละสถานีนั้น ทำให้ทราบถึงพฤติกรรมของอนุกรมชุดต่างๆ และในอนุกรมชุดต่างๆ มีในบางอนุกรมมีส่วนของแนวโน้มอยู่ และได้ทำการย้ายส่วนแนวโน้มออกก่อนการวิเคราะห์ทางสโตแคสติก แบบจำลองที่เหมาะสมในส่วนของสโตแคสติก คือ แบบจำลองการถดถอยในตัวเอง (Autoregressive Model) AR

อนุกรมเวลาอุทกวิทยาเป็นจำนวนมากที่พฤติกรรมความสัมพันธ์โดยลำดับเวลา มีค่านัยสำคัญ (Significance) นั้นคือ ค่าของตัวแปรสุ่มที่คาบเวลาอื่น ๆ ความสัมพันธ์ของตัวแปรสุ่ม X ที่คาบเวลาหนึ่ง ด้วยค่า k คาบเวลาก่อน ๆ ได้กำหนดโดย ρ_k และเรียก k เป็นอันดับความสัมพันธ์โดยลำดับ หรือช่วงเวลาถัดไป k (Multilag) ดังนั้นจะได้ว่าเป็นแบบจำลอง

หลายช่วงเวลาถัดไป (Multilag Model) หรือ แบบจำลองอันดับที่สูง ๆ หรือ แบบจำลองการถดถอยในตัวเองอันดับสูง ๆ สามารถเขียนได้ตาม

$$X_t = b_0 + b_1X_{t-1} + b_2X_{t-2} + \ldots + b_kX_{t-k} + \varepsilon_t$$

ค่าของ X_{t-k} อาจจะแทนค่าข้อมูลจริง หรือ Natural Logarithms ในกรณีของแบบจำลองปกติ องค์ประกอบสุ่มกลายเป็น

$$\varepsilon_t = \sigma_x w_t \sqrt{1-R^2}$$

ที่ σ^2_x เป็นความแปรปรวนของ X

R^2 เป็นสัมประสิทธิ์หลายมิติ (Multilag) ของค่าที่หาได้ระหว่าง X_t และ $X_{t-1}, X_{t-2}, \ldots, X_{t-k}$

w_t เป็นค่าการสังเกตุสุ่มจาก N(0,1) และ

b เป็นค่าสัมประสิทธิ์การถดถอยหลายมิติ (Multilag)

ผลจากการวิเคราะห์ทางสโตแคสติกของอนุกรมชุดต่าง ๆ ของแต่ละสถานี ปรากฏว่าสำหรับอนุกรมฝนรายปี สถานีกรมอุตุนิยมวิทยา (41013) มีค่าอันดับของแบบจำลอง AR เป็น 3 หรือ AR(3) สถานีสนามบินดอนเมือง (41063) มีค่าอันดับของแบบจำลอง AR เป็น 2 หรือ AR(2) สถานีเขตมีนบุรี (41052) มีค่าอันดับของแบบจำลอง AR เป็น 9 หรือ AR(9) และสถานีเขตหนองแขม (41202) มีค่าอันดับแบบจำลอง AR เป็น 4 หรือ AR(4) ซึ่งสามารถเขียนสมการในการพัฒนาแบบจำลองการพยากรณ์ฝนรายปีได้ว่า

สถานีกรมอุตุนิยมวิทยา (41013)

$$X_t = b_0 + b_3X_{t-3}$$

สถานีสนามบินดอนเมือง (41063)

$$X_t = b_0 + b_2X_{t-2}$$

สถานีเขตมีนบุรี (41052)

$$X_t = b_0 + b_9 X_{t-9}$$

สถานีเขตหนองแขม (41202)

$$X_t = b_0 + b_4 X_{t-4}$$

ในการศึกษาครั้งนี้ได้พิจารณาค่าสัมประสิทธิ์ความสัมพันธ์ของเวลาสำหรับฝนรายปีในเทอมที่มีค่านัยสำคัญ (Significance) ของ lag time k มาพิจารณาส่วนเทอมอื่นๆ ที่ไม่ Significance ไม่นำมาพิจารณา หรือถือว่ามีค่าเป็น 0 เพื่อดูพฤติกรรมของส่วนประกอบต่างๆ ของแบบจำลองที่ใช้ในการศึกษาครั้งนี้

5.1 การพัฒนาแบบจำลองการพยากรณ์ปริมาณฝนรายปี

จากสมการ (5-1) การพยากรณ์ปริมาณฝนของอนุกรมชุดหนึ่งๆ นั้น สำหรับฝนรายปีส่วนแนวโน้ม ถ้ามีก็นำมาพิจารณาด้วย ส่วนฤดูกาล และส่วนวงจร สำหรับฝนรายปีไม่นำมาพิจารณา เนื่องจากข้อมูลที่สนใจใช้ช่วงเวลาเป็นปี ดังนั้นสามารถเขียนสมการสำหรับแบบจำลองการพยากรณ์ปริมาณฝนรายปีได้ว่า

$$Z_i = \phi_k Z_{i-k} \qquad (5\text{-}2)$$

ที่ $Z_i = \dfrac{X_i - \mu_x}{\sigma_x}$

$\phi_k = \rho_x(k)$ เป็นค่าความสัมพันธ์ช่วงเวลาโดยลำดับ k โดยพิจารณาค่าที่ significance หาได้จากสมการ (2-30)

จากสมการ (5-2) จะได้ว่า

$$X_i = \mu_x + \rho_x(k)\,(X_{i-k} - \mu_x) \qquad (5\text{-}3)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ที่ X_i เป็นค่าของปริมาณฝนที่เวลา i

μ_x เป็นค่าเฉลี่ยของ X

$\rho_x(k)$ เป็นค่าความสัมพันธ์ในตัวเองที่ช่วงเวลาถัดไป k

แบบจำลองนี้กำหนดว่าค่าของ X ในคาบเวลาหนึ่งขึ้นอยู่กับค่าของ X ในคาบเวลาก่อนหน้านี้ที่ปี i-k

วิธีการสำหรับการหาค่าของ X_i คือทำการประมาณค่า μ_x, σ_x และ $\rho_x(k)$ โดย $\bar{X}$, S_x และ $r_x(k)$ ตามลำดับ ดังแสดงในตารางที่ 5-1,2,3,4

สมการ (5-5) ทำการหาค่าของ X ต่าง ๆ ด้วยค่าเฉลี่ย (μ_x) และความสัมพันธ์โดยลำดับอันดับ k ($\rho_x(k)$) ความสัมพันธ์โดยลำดับเป็นแบบธรรมดาในทางอุทกวิทยา และขึ้นอยู่กับการใช้เป็นแบบจำลองที่สำคัญด้วย โดยค่าเฉลี่ย และความสัมพันธ์โดยลำดับ เปลี่ยนแปลงไปตามข้อมูลที่เพิ่มขึ้นทุกๆ ปี หรือจะต้องหาค่าเหล่านี้ใหม่ทุกครั้ง ก่อนที่จะทำการคาดหมายปริมาณฝนในปีต่อไป สังเกตุได้ว่าถ้า $\rho_x(k)= 0$ สมการ (5-5) ค่า X_i จะเท่ากับค่าเฉลี่ย (μ_x) แต่ถ้า $\rho_x(k) = 1$ สมการ (5-5) หาได้แน่นอนโดยสมบูรณ์ว่า X_i กำหนดได้โดย X_{i-k} หรือ $X_i = X_{i-k}$ นั่นเอง จากการคาดหมายปริมาณฝนรายเดือน พบว่าสถานีกรมอุตุนิยมวิทยา(41013) สามารถอธิบายค่า R^2 ได้ว่ามีค่าเท่ากับ 0.6161 สถานีสนามบินดอนเมือง(41063) สามารถอธิบายค่า R^2 ได้ว่ามีค่าเท่ากับ 0.95 สถานีเขตมีนบุรี(41052) สามารถอธิบายค่า R^2 ได้ว่ามีค่าเท่ากับ 0.9167 และสถานีเขตหนองแขม (41202) สามารถอธิบายค่า R^2 ได้ว่ามีค่าเท่ากับ 0.6358

5.2 การพัฒนาแบบจำลองการพยากรณ์ฝนรายเดือน

ในการพัฒนาแบบจำลองการพยากรณ์ปริมาณฝนรายเดือนหรือแบบจำลองการถดถอยในตัวเองของฤดูกาล ช่วงถัดไปหนึ่งหน่วย รวมกับความสัมพันธ์ช่วงเวลาในอนุกรมรายปีของแต่ละเดือนที่พิจารณาสามารถเขียนสมการได้เป็น

$$Z_{i,j} = a_1 Z_{i-1,j-1} + b_k Z_{i-k,j} \qquad (5\text{-}6)$$

ที่ $Z_{i,j} = \dfrac{X_{i,j} - \mu_{x,j}}{\sigma_{x,j}}$

$a_1 = \rho_{x,j-1}(1)$ เป็นค่าความสัมพันธ์ในตัวเองอันดับหนึ่ง สำหรับฝนรายเดือน j-1 โดยพิจารณาค่าที่ Significance หาได้จากสมการ (2-30)

$b_k = \rho_{x,j}(k)$ เป็นค่าความสัมพันธ์ช่วงเวลาโดยลำดับ k สำหรับฝนรายปีในเดือน j โดยพิจารณาค่าที่ Significance หาได้จากสมการ (2-30)

$X_{i-k,j}$ เป็นค่าปริมาณฝนที่เดือน j ปีที่มีความสัมพันธ์ช่วงเวลา k เป็นปี i-k ที่พิจารณา

จากสมการ (5-6) จะได้ว่า

$$X_{i,j} = \mu_{x,j} + \rho_{x,j}(1)\frac{\sigma_{x,j}}{\sigma_{x,j-1}}(X_{i-1,j-1} - \mu_{x,j-1}) + \rho_{x,j}(k)(X_{i-k,j} - \mu_{x,j}) \qquad (5\text{-}7)$$

โดยที่ ตัวห้อย i คือปีที่พิจารณา และ j คือ เดือนที่พิจารณาในปีนั้น และ j มีค่าจาก 1 ถึง 12 หรือ j = 1, 2, ..., m ค่า $\mu_{x,j}$ เป็นค่าเฉลี่ยของ X ในเดือนที่ j $\mu_{x,j}$ ประมาณได้โดย $\bar{X}_j$ ที่

$$\bar{X}_j = \sum_{i=1}^{n} \frac{X_{i,j}}{n} \qquad (5\text{-}8)$$

ด้วย n เท่ากับจำนวนปีของข้อมูล และ

$X_{i,j}$ ค่าข้อมูลเดือนที่ j ของปีที่ i ในทำนองเดียวกัน

$\sigma^2_{x,j}$ เป็นค่าที่ประมาณได้โดย $S^2_{x,j}$

$\rho_{x,j}(1)$ ค่าที่ประมาณได้โดย $r_{x,j}(1)$ ซึ่งเป็นค่าความสัมพันธ์โดยลำดับอันดับหนึ่งระหว่างเดือนที่ต่อเนื่องกันเช่น ถ้าปริมาณฝนที่พิจารณาคือ $\rho_{x,4}(1)$ ก็จะเป็นความสัมพันธ์โดยลำดับอันดับหนึ่งระหว่าง เดือน 4 และ 5 สามารถประมาณได้โดย

$$r_{x,j}(1) = \frac{\sum_{i=1}^{n} X_{i,j}X_{i,j+1} - nX_jX_{j+1}}{(n-1)\, S_{x,j}\, S_{x,j+1}} \qquad (5\text{-}9)$$

ที่

$$S^2_{x,j} = \frac{\sum_{i=1}^{n} X^2_{i,j} - nX^2_j}{n-1} \qquad (5\text{-}10)$$

ในสมการ (5-9) มีบางตัวที่มีปัญหาที่ j = m ในกรณีนี้ j+1 ควรจะถือตาม 1 โดยที่เดือนแรกต่อไปเป็นเดือนที่ m (เดือนธันวาคมต่อไปเป็นเดือนมกราคม)

ในการใช้พารามิเตอร์ประชากรประมาณได้จากค่าสถิติที่ตรงกัน มีปัญหาอย่างหนึ่งคือ ค่า $X_{i,j+1}$ จริง ๆ แล้วเท่ากับ $X_{i+1,1}$ ที่ j = m ดังนั้น $X_{i,13}$ หรือ เดือนที่ 13 ในปีที่ i จริง ๆ แล้ว คือ $X_{i+1,1}$ หรือค่าเดือนแรกในปีต่อไป

5.3 การปรับปรุงค่าพยากรณ์ฝนรายเดือน

โดยปกติแล้วปริมาณฝนรายปีจะมาจากการรวมปริมาณฝนรายเดือนของในปีนั้นๆ สามารถเขียนอยู่ในรูปสมการได้ว่า

$$X_i = f(x_{i,j}) + \varepsilon_{i,j} = \sum_{j=1}^{12} X_{i,j} + \varepsilon_{i,j} \qquad (5\text{-}11)$$

โดยที่ X_i ปริมาณฝนรวมรายปีรวมจากฝนรายเดือนที่ปี i

$f(x_{i,j})$ เป็นฟังค์ชั่นของปริมาณฝนแต่ละเดือน

$\varepsilon_{i,j}$ ค่าความผิดพลาดในเดือน j ปีที่ i

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 5-1 จะเห็นว่า ค่า X_i ที่ได้จากการพยากรณ์ฝนรายเดือน รวมกันนั้น ไม่เท่ากันกับปริมาณฝนที่เป็นจริง หรือเขียนอยู่ในรูปสมการได้ว่า

$$\varepsilon_{i,j} + \sum_{i=1}^{12} x_{i,j} \neq X_i$$

ดังจะพบได้ว่าเปอร์เซนต์ความคลาดเคลื่อนแต่ละสถานีคือ

สถานีกรมอุตุนิยมวิทยา(41013) มีความคลาดเคลื่อน 34.2 เปอร์เซนต์

สถานีสนามบินดอนเมือง(41063) มีความคลาดเคลื่อน -7.07 เปอร์เซนต์

สถานีเขตมีนบุรี(41052) มีความคลาดเคลื่อน 2.25 เปอร์เซนต์

สถานีเขตหนองแขม(41202) มีความคลาดเคลื่อน 1.99 เปอร์เซนต์

ในการปรับแก้ค่าพยากรณ์สามารถทำได้โดยการเฉลี่ยค่าความผิดพลาดให้กับทุกๆ เดือน โดยใช้ Weight ของปริมาณฝนแต่ละเดือนเป็นเกณฑ์จากค่าความผิดพลาดรวม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5-1 ผลการพัฒนาแบบจำลอง(สถานีกรมอุตุนิยมวิทยา)

STATION		41013 Bangkok Metropolis												
YEAR	PARAMETER	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1991	lag(k)	5	13	9	3	5	6	3	8	8	3	9	2	3
	$X_{i-k,j}$	0.0	103.6	176.0	139.3	504.5	90.5	177.8	574.5	453.8	380.3	44.3	0.0	2097.3
	$X_{i-1,j}$	1.4	1.2	50.7	15.5	220.2	59.3	159.9	172.5	258.6	401.0	22.6	0.0	1362.9
	$\mu_{x,j}$	8.9	27.9	27.5	64.0	204.0	151.0	156.1	201.7	346.6	252.1	50.1	8.6	1499.9
	$\rho_{x,j}(k)$	-0.345	0.201	0.524	-0.347	0.207	-0.420	-0.374	0.376	0.285	0.314	0.215	0.231	-0.316
	$\sigma_{x,j}$	14.1	35.9	35.3	55.8	130.3	81.8	63.9	100.8	110.1	116.8	44.1	19.9	289.2
	$\rho_{x,j}(1)$	0.448	0.502	0.487	0.583	0.515	0.482	0.453	0.497	0.181	0.664	0.266	0.254	
	$X_{i,j}$	10.4	34.6	92.1	55.7	200.2	181.6	113.5	344.6	361.3	275.5	86.2	3.3	1311.1

Note: lag(k) = year X = Monthly Rainfall (mm) i = 1991 j = 1,2,...,12

ตารางที่ 5-2 ผลการพัฒนาแบบจำลอง(สถานีสนามบินดอนเมือง)

STATION		41063 Don Muang Airport												
YEAR	PARAMETER	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
	lag(k)	19	17	19	1	15	12	14	15	13	3	19	2	2
1991	$X_{i-k,j}$	0.0	2.1	84.9	37.6	163.2	78.4	192.2	262.7	289.9	341.7	3.1	0.0	1442.1
	$X_{i-1,j}$	1.1	0.0	24.7	37.6	186.9	87.2	86.3	101.1	247.8	474.3	47.6	0.0	1294.6
	$\mu_{x,j}$	8.0	18.3	26.6	62.9	157.4	140.7	146.9	197.2	286.9	205.3	36.7	13.7	1300.5
	$\rho_{x,j}(k)$	0.156	0.516	0.615	-0.305	0.351	0.262	0.296	0.451	0.246	0.351	0.396	0.318	0.371
	$\sigma_{x,j}$	18.0	26.0	29.7	51.0	81.1	62.6	60.8	96.1	115.6	126.9	30.9	28.8	305.6
	$\rho_{x,j}(1)$	0.479	0.544	0.584	0.695	0.608	0.564	0.537	0.530	0.352	0.791	0.290	0.269	
	$X_{i,j}$	4.4	5.2	51.1	68.7	131.5	138.2	131.0	175.3	226.4	238.1	75.2	12.3	1353.0

Note: lag(k) = year X = Monthly Rainfall (mm) i = 1991 j = 1,2,...,12

ตารางที่ 5-3 ผลการพัฒนาแบบจำลอง(สถานีเขตมีนบุรี)

STATION		41052 A. Min Buri												
YEAR	PARAMETER	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
	lag(k)	19	13	11	14	17	12	6	15	4	17	15	17	9
1990	$X_{i-k,j}$	0.0	129.4	20.9	87.0	232.3	76.6	202.7	245.6	235.9	96.8	4.1	4.9	1421.4
	$X_{i-1,j}$	0.9	0.0	13.4	35.5	183.1	125.8	170.1	247.6	296.3	239.0	12.5	0.0	1324.2
	$\mu_{x,j}$	11.4	17.9	14.7	80.9	190.0	155.5	166.6	196.4	302.6	191.4	41.4	9.0	1377.8
	$\rho_{x,j}(k)$	0.518	0.829	0.379	0.635	0.509	0.573	0.375	0.377	-0.335	0.503	0.402	0.290	-0.525
	$\sigma_{x,j}$	26.7	37.1	18.5	61.5	87.6	82.5	48.9	101.3	121.5	89.3	43.0	24.5	284.1
	$\rho_{x,j}(1)$	0.752	0.758	0.819	0.842	0.789	0.770	0.764	0.600	0.648	0.865	0.646	0.642	
	$X_{i,j}$	-0.8	99.4	10.3	81.2	157.1	105.2	166.6	220.5	361.8	140.8	46.2	-2.8	1354.9

Note: lag(k) = year X = Monthly Rainfall (mm) i = 1990 j = 1,2,...,12

ตารางที่ 5-4 ผลการพัฒนาแบบจำลอง(สถานีเขตหนองแขม)

STATION		41202 A. Nong Khaem												
YEAR	PARAMETER	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
	lag(k)	17	12	4	2	11	12	13	17	4	5	4	5	4
1991	$X_{i-k,j}$	3.5	26.7	5.0	10.0	34.4	55.5	161.0	149.9	279.2	0.0	328.4	9.2	1349.6
	$X_{i-1,j}$	0.0	0.0	52.6	10.6	153.1	81.1	158.0	42.2	397.8	302.7	87.7	0.0	1285.8
	$\mu_{x,j}$	12.5	18.1	19.2	54.4	149.5	129.9	130.9	162.9	271.8	224.2	65.6	9.1	1248.1
	$\rho_{x,j}(k)$	0.733	0.448	0.253	0.373	0.516	0.291	0.416	0.295	-0.437	-0.321	0.429	0.362	-0.246
	$\sigma_{x,j}$	23.8	32.1	20.9	54.5	95.2	67.1	65.9	90.0	93.3	149.8	68.9	16.1	297.9
	$\rho_{x,j}(1)$	0.564	0.577	0.569	0.584	0.529	0.527	0.506	0.531	-0.089	0.452	0.334	0.285	
	$X_{i,j}$	2.1	12.4	8.8	87.4	45.4	109.6	118.2	177.8	202.1	278.2	194.7	10.9	1223.1

Note: lag(k) = year X = Monthly Rainfall (mm) i = 1991 j = 1,2,...,12