บทที่ 2
ทฤษฎีและหลักการที่ใช้ในการศึกษา

ในการสังเคราะห์ข้อมูลน้ำฝนรายวันจะประกอบไปด้วยขั้นตอน 2 ขั้นตอนคือ ขั้นตอนของการสังเคราะห์วันที่เกิดฝนตก (Precipitation Occurrence Process) และขั้นตอนของการสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตก (Precipitation Amounts Process)

โดยใช้วิธี Two-state, First-order Markov Chain ในขั้นตอนของการสังเคราะห์วันที่เกิดฝนตก ส่วนขั้นตอนของการสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตกจะทำการสังเคราะห์โดยใช้การแจกแจง 2 วิธี แล้วนำมาเปรียบเทียบกันโดยใช้ค่าทางสถิติเป็นตัววัดความเหมาะสมว่าวิธีการแจกแจงใดให้ข้อมูลฝนรายวันใกล้เคียงกับข้อมูลจริงมากที่สุด ซึ่งวิธีการแจกแจง 2 วิธีที่นำมาใช้นั้นได้แก่ วิธี Two-parameter Gamma Distribution และ วิธี Mixed Exponential Distribution

ในการสังเคราะห์ค่าปริมาณน้ำฝนรายวันที่กล่าวมาเบื้องต้นทั้งหมด มีทฤษฎีและหลักการดังต่อไปนี้

2.1 การสังเคราะห์วันที่เกิดฝนตก (Precipitation Occurrence Process)

การสังเคราะห์วันที่เกิดฝนตกในการศึกษานี้เลือกใช้แบบจำลองที่รู้จักกันโดยทั่วไป คือ Two-state, First-order Markov Chain

2.1.1 Two-state, First-order Markov Chain

Markov Chain เป็นขบวนการทางสโตแคสติกที่มีคุณสมบัติคือ ค่าที่เกิดขึ้นในวันที่ t (X_t) จะไม่ขึ้นกับค่า X_{t-2} , X_{t-3} , ... , X_0 ซึ่งเป็นค่าลำดับที่เกิดขึ้นในขบวนการก่อนที่จะดำเนินมาถึงค่าในลำดับที่ X_{t-1} แต่จะขึ้นกับค่าที่เกิดขึ้นในวันที่ t - 1 (X_{t-1}) เท่านั้น (Haan, 1977) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$\text{Prob}\,(X_t = a_j \mid X_{t-1} = a_i , X_{t-2} = a_k , X_{t-3} = a_l , \ldots, X_0 = a_q) =$$
$$\text{Prob}\,(X_t = a_j \mid X_{t-1} = a_i) \qquad (2.1)$$

ค่า Prob ($X_t = a_j \mid X_{t-1} = a_i$) คือ ค่าความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability) เป็นค่าความน่าจะเป็นซึ่งขบวนการที่เกิดขึ้นที่เวลา t จะอยู่ใน State ที่ j และที่เวลา t - 1 ขบวนการจะอยู่ใน State ที่ i

สมการที่ (2.1) แสดงให้เห็นว่าค่าความน่าจะเป็นเงื่อนไขนี้จะเป็นอิสระต่อ State ที่เกิดขึ้นก่อนหน้านั้นที่เวลาที่ t - 1

ค่า Prob ($X_t = a_j \mid X_{t-1} = a_i$) เรียกกันโดยทั่วไปว่า "One Step Transition Probability" ซึ่งคือค่าความน่าจะเป็นที่ขบวนการทำการ Transition จาก State ที่ a_i ไปยัง State ที่ a_j ภายในช่วงเวลาเดียวหรือขั้นตอนเดียว ค่า Prob ($X_t = a_j \mid X_{t-1} = a_i$) มักจะใช้ในรูปแบบของ $p_{ij}(t)$ ซึ่งแสดงถึงค่าความน่าจะเป็นของขั้นตอนจาก a_i ไปยัง a_j ในช่วงเวลาที่ t ถ้า $p_{ij}(t)$ เป็นอิสระต่อ t สามารถบอกว่าได้ว่า Markov Chain นี้เป็น Homogeneous

$$\text{Prob}(X_t = a_j \mid X_{t-1} = a_i) = p_{ij} \tag{2.2}$$

ถ้าขบวนการถูกแบ่งออกเป็น m States ขบวนการจะถูกกำหนดโดย m^2 Transition Probabilities และสามารถเขียนให้อยู่ในรูป Matrix ได้ดังนี้

$$P = \left[p_{ij}\right] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ p_{m1} & & & p_{mm} \end{bmatrix} \tag{2.3}$$

ในการนำ Two-state, First-order Markov Chain มาใช้ในการสังเคราะห์วันที่เกิดฝนตก (Todorovic and Woolhiser, 1975; Wilks, 1998) ให้ X_t แทนเหตุการณ์ 2 เหตุการณ์ในการเกิดฝน คือ เกิดฝนตก หรือไม่เกิดฝนตกในวันที่ t

$$\begin{aligned} X_t &= 0 \quad \text{เมื่อวันที่ t ไม่เกิดฝนตก} \\ &= 1 \quad \text{เมื่อวันที่ t เกิดฝนตก} \end{aligned} \tag{2.4}$$

ดังนั้นอนุกรมเวลาของปริมาณน้ำฝนคือ

$$Y_t = r_t X_t \tag{2.5}$$

เมื่อ r_t แทนค่า ปริมาณน้ำฝนที่ไม่เป็นศูนย์ ดังนั้น $Y_t = 0$ เมื่อ $X_t = 0$ และ $Y_t = r_t$ เมื่อ $X_t = 1$

Two-state, First-order Markov Chain สำหรับ X_t จะเป็นไปตามสมมติฐานที่ว่า ความน่าจะเป็นของวันฝนตก กำหนดโดยการเกิดฝนหรือไม่เกิดฝนของวันก่อนหน้านั้น ดังนั้นจะมีรูปแบบ Matrix of Transition Probabilities ดังนี้ (Clarke, 1998)

$$p = \begin{bmatrix} Prob(X_t = 0 \mid X_{t-1} = 0) & Prob(X_t = 1 \mid X_{t-1} = 0) \\ Prob(X_t = 0 \mid X_{t-1} = 1) & Prob(X_t = 1 \mid X_{t-1} = 1) \end{bmatrix} \quad (2.6)$$

โดยที่ $\text{Prob}\ (X_t = 0 \mid X_{t-1} = 0) = P_{00}$ (2.7a)

และ $\text{Prob}\ (X_t = 1 \mid X_{t-1} = 0) = P_{01}$ (2.7b)

และ $\text{Prob}\ (X_t = 0 \mid X_{t-1} = 1) = P_{10}$ (2.7c)

และ $\text{Prob}\ (X_t = 1 \mid X_{t-1} = 1) = P_{11}$ (2.7d)

Matrix of Transition Probabilities สามารถเขียนในเทอมของค่าคงที่ P_{00}, P_{01}, P_{10} และ P_{11} ได้ดังนี้

$$P = \text{State at Time i -1} \begin{array}{c c} & \text{State at Time i} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \end{array} \quad (2.8)$$

ค่าพารามิเตอร์ P_{01} และ P_{11} เป็น Conditional Probability ของวันฝนตกที่เกิดหลังวันฝนไม่ตก และ Conditional Probability ของวันฝนตกที่เกิดหลังวันฝนตกตามลำดับ และ $P_{00} + P_{01} = 1$ และ $P_{10} + P_{11} = 1$

กำหนดให้ Transition Probability ใน สมการที่ (2.7b) และ (2.7d) เป็น Critical Probability โดย

$$\begin{aligned} P_c &= P_{01} \quad \text{เมื่อ } X_{t-1} = 0, \\ &= P_{11} \quad \text{เมื่อ } X_{t-1} = 1 \end{aligned} \quad (2.9)$$

สังเคราะห์ค่า Uniform Random Number ซึ่งอยู่ในช่วง 0 ถึง 1 มาเปรียบเทียบกับ Condition Probability ในสมการที่ (2.9) โดยวันที่ฝนตกจะถูกสังเคราะห์เมื่อค่า Uniform Random Number ซึ่งอยู่ในช่วง 0 ถึง 1 นี้มีค่าน้อยกว่า Condition Probability ทำให้ได้ค่าต่อไปใน X_t series คือ

$$\begin{aligned} X_t &= 1 \quad \text{เมื่อ } U_t \leq P_c \\ &= 0 \quad \text{เมื่อ } U_t \geq P_c \end{aligned} \quad (2.10)$$

2.1.2 การตรวจสอบการสังเคราะห์วันที่เกิดฝนตกโดยใช้ค่าทางสถิติ

ในการตรวจสอบการสังเคราะห์วันที่เกิดฝนตกที่สังเคราะห์ได้โดยวิธี Two-state, First-order Markov Chain นี้ว่ามีความเหมาะสมอย่างไร สามารถทำได้โดยใช้ค่าทางสถิติคือ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์การแปรผันของจำนวนวันที่เกิดฝนตกของข้อมูลที่

สังเคราะห์ได้เปรียบเทียบกับข้อมูลจริง แล้วนำมาเป็นตัววัดความเหมาะสม รายละเอียดทฤษฎีของค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์การแปรผัน ดังแสดงได้ในหัวข้อ 2.4.1 ถึง 2.4.3 ตามลำดับ

วันที่เกิดฝนตกซึ่งสังเคราะห์ได้จาก Two-state, First-order Markov Chain ในหัวข้อนี้จะถูกนำไปสังเคราะห์ค่าปริมาณน้ำฝนรายวันโดยการแจกแจง 2 วิธี ได้แก่ Two-parameter Gamma Distribution และ Mixed Exponential Distribution ในหัวข้อ 2.2 และ 2.3 ตามลำดับ แล้วนำมาเปรียบเทียบกันว่าวิธีใดให้ความเหมาะสมทางสถิติกว่ากัน

2.2 การสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตก (Precipitation Amounts Process) โดย Two-parameter Gamma Distribution

การสังเคราะห์ปริมาณน้ำฝนรายวันโดยใช้ Two-parameter Gamma Distribution มีขั้นตอนการสังเคราะห์ 2 ขั้นตอนคือ ขั้นแรกทำการหาค่าพารามิเตอร์ δ และ λ โดยใช้ Method of Maximum Likelihood เมื่อได้ค่าพารามิเตอร์ทั้งสองแล้ว ขั้นต่อมาคือการนำค่าพารามิเตอร์ทั้งสองมาทำการสังเคราะห์ค่า Gamma Random Variable โดยในการศึกษาครั้งนี้ใช้วิธี Method of Whittaker (1973) ในการสังเคราะห์ ซึ่งขั้นตอนนี้ก็คือการสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตกที่สังเคราะห์โดย Two-state, First-order Markov Chain ในหัวข้อที่ 2.1 นั่นเอง

Two-parameter Gamma Distribution มี Probability Density Function คือ

$$f(x)=\frac{x^{\delta-1}\exp[-x/\lambda]}{\lambda^{\delta}\Gamma[\delta]} \qquad (2.11)$$

เมื่อ $x>0$

และ $f(x)=0$ เมื่อ $x<0$

ค่า δ คือ Shape Parameter และ ค่า λ คือ Scale Parameter โดยที่ $\delta>0$ และ $\lambda>0$ และมีคุณสมบัติทางสถิติที่สำคัญคือ $E(x)=\delta\lambda$ และ $Var(x)=\delta\lambda^2$

2.2.1 การประมาณค่าพารามิเตอร์ของ Two-parameter Gamma Distribution โดย Method of Maximum Likelihood

การหาค่าพารามิเตอร์ทั้งสองของ Two-parameter Gamma Distribution ซึ่งได้แก่ δ และ λ สามารถประมาณหาค่าได้โดยใช้ Method of Maximum Likelihood มีทฤษฎีดังต่อไปนี้

Method of Maximum Likelihood ซึ่งพัฒนาขึ้นโดย R.A. Fisher (1922) (Chow et al., 1988) มีหลักการที่ว่าค่าที่ดีที่สุดของค่าพารามิเตอร์ของการแจกแจงความน่าจะเป็น (Probability Distribution) จะหาได้โดยทำการหาค่าที่มากที่สุด (Maximize) ของ Likelihood หรือ Joint Probability Distribution ของข้อมูลที่ทำการเก็บบันทึกได้ และมี Likelihood Function คือ

$$L = \prod_{i=1}^{n} f(x_i) \tag{2.12}$$

เนื่องจาก Probability Density Function ส่วนใหญ่อยู่ในรูปแบบของ Exponential บางครั้งจึงจัดสมการให้อยู่ในรูปแบบที่ง่ายต่อการนำไปใช้ ดังแสดงในสมการที่ (2.13) ซึ่งเรียกว่า Log-likelihood Function

$$\ln L = \sum_{i=1}^{n} \ln\left[f(x_i)\right] \tag{2.13}$$

Log-likelihood Function ของ Two-parameter Gamma Distribution เพื่อใช้ในการหาค่าพารามิเตอร์ δ และ λ มีรูปแบบดังนี้

$$\ln L = \sum_{i=1}^{n} \ln\left[\frac{x_i^{\delta-1}\exp[-x_i/\lambda]}{\lambda^{\delta}\Gamma[\delta]}\right] \tag{2.14}$$

เขียนใหม่ได้คือ

$$\ln L = \sum_{i=1}^{n}\left[(\delta-1)\ln x_i - \frac{x_i}{\lambda} - \delta\ln\lambda - \ln\Gamma(\delta)\right] \tag{2.15}$$

ในการหาค่าพารามิเตอร์ δ และ λ สามารถทำได้โดยการแก้สมการที่ (2.15) โดยการทำ Partial Derivative เทียบกับพารามิเตอร์แต่ละตัว และให้เท่ากับศูนย์ (Clarke, 1984) จะได้ออกมา 2 สมการคือ

$$\sum_{i=1}^{n} \ln x_i - n\ln\lambda - n\frac{\partial}{\partial\delta}\left[\ln\Gamma(\delta)\right] = 0 \tag{2.16}$$

และ

$$\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i - \frac{n\delta}{\lambda} = 0 \tag{2.17}$$

จากสมการที่ (2.16) ทำการประมาณค่าเทอมที่เป็น Gamma Function ดังสมการที่ (2.19)

$$\psi(\delta) = \frac{\partial}{\partial\delta}\left[\ln\Gamma(\delta)\right] \tag{2.18}$$

ประมาณได้ดังนี้

$$\psi(\delta) = \ln\delta - \frac{1}{2\delta} - \frac{1}{12\delta^2} \tag{2.19}$$

จากสมการที่ (2.17) สามารถทำให้อยู่ในรูปแบบเพื่อนำไปประมาณค่าพารามิเตอร์ δ ได้ดังนี้

$$\lambda = \frac{1}{n\delta}\sum_{i=1}^{n} x_i \tag{2.20}$$

แทนที่ค่าประมาณในสมการที่ (2.19) และค่าพารามิเตอร์ λ ในสมการที่ (2.20) ลงในสมการที่ (2.16) เพื่อหาค่าพารามิเตอร์ δ ดังนี้

$$\frac{1}{2\delta} + \frac{1}{12\delta^2} = \ln\bar{x} - \frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{2.21}$$

สมการที่ (2.21) สามารถเขียนใหม่ได้ดังนี้

$$12A\delta^2 - 6\delta - 1 = 0 \tag{2.22}$$

เมื่อ

$$A = \ln\bar{x} - \frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{2.23}$$

จากสมการที่ (2.22) สามารถนำมาหาค่าพารามิเตอร์ δ ได้ และเนื่องจากสมการที่ (2.22) เป็นสมการ Quadratic ดังนั้นจะได้ว่า

$$\delta = \frac{1 + \sqrt{1 + \frac{4}{3}A}}{4A} \tag{2.24}$$

แทนค่าสมการที่ (2.24) ลงในสมการที่ (2.20) ก็จะสามารถหาค่าพารามิเตอร์ของ λ ได้ โดย

$$\lambda = \frac{\bar{x}}{\delta} \tag{2.25}$$

พบว่าค่าพารามิเตอร์ δ ที่ได้จากสมการที่ (2.24) นั้น มีค่าเบี่ยงเบนเล็กน้อยสำหรับการหาค่าพารามิเตอร์ δ ดังนั้นเพื่อให้ได้ค่าพารามิเตอร์ δ ที่แม่นยำมากยิ่งกว่า จึงได้มีการประมาณค่า δ_* สำหรับค่าพารามิเตอร์ δ (Bobée and DesGroseilliers, 1985) ดังสมการที่ (2.26)

$$\delta_* = \delta - \Delta\delta \tag{2.26}$$

เมื่อ δ คือค่าพารามิเตอร์ที่ได้จากสมการที่ (2.24) และค่า $\Delta\delta$ หาได้จากสมการที่ (2.27) ดังนี้

$$\Delta\delta = (0.04475)(0.26)^{\delta} \tag{2.27}$$

ดังนั้นค่าพารามิเตอร์ λ จึงประมาณได้โดยการหาค่าของพารามิเตอร์ λ_* ดังสมการที่ (2.28)

$$\lambda_* = \frac{\bar{x}}{\delta_*} \tag{2.28}$$

2.2.2 การสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตกสำหรับวิธี Two-parameter Gamma Distribution โดยใช้ Method of Whittaker (1973)

เมื่อได้ค่าพารามิเตอร์ δ และ λ จากหัวข้อที่ 2.2.1 แล้ว นำค่าพารามิเตอร์ทั้งสองมาทำการสังเคราะห์ค่า Gamma Random Variable โดยใช้ Method of Whittaker (1973) ซึ่งก็คือการสังเคราะห์ค่าปริมาณน้ำฝนรายวันสำหรับวิธี Two-parameter Gamma Distribution นั่นเอง วิธีการสังเคราะห์มีขั้นตอนดังต่อไปนี้

ให้ k เป็นส่วนที่เป็นจำนวนเต็ม (Integer Part) ของ Shape Parameter (δ) และ ให้ $p = \delta - k$ ดังนั้นจะเกิด 2 กรณี คือ

(i) $\delta < 1 : k = 0$ และ $p = \delta$

(ii) $\delta > 1 : k \geq 1$ และ $p = 0$ หรือ $p > 0$ ขึ้นกับว่า δ เป็นจำนวนเต็มหรือไม่

พิจารณากรณี (i)

ให้ $u = 1/p$ และ $v = 1/(1 - p)$

ขั้นที่ 1 สร้างค่า Uniform Random Number ในช่วง 0 ถึง 1 มาสองตัว เป็น U_1 และ U_2

$S_1 = U_1^u$ และ $S_2 = U_2^v$

ถ้า $S_1 + S_2 \leq 1$ ทำต่อที่ ขั้นที่ 2 และในทางกลับกัน ถ้า $S_1 + S_2 > 1$ ให้กลับมาทำที่ขั้นที่ 1 ใหม่

ขั้นที่ 2 สร้างค่า Uniform Random Number ในช่วง 0 ถึง 1 อีกตัวหนึ่ง เป็น U_3

และให้ $y = S_1 / (S_1 + S_2)$

ดังนั้น
$$r = x_1 = -y \ln U_3 \qquad (2.29)$$

พิจารณากรณี (ii)

ก. ถ้า $p = 0$ (เนื่องจาก δ = จำนวนเต็ม) ดังนั้น

$$r = z = -\ln\left(\prod_{i=1}^{k} r_i\right) \qquad (2.30)$$

โดย r_i แทนค่า Uniform Random Number ในช่วง 0 ถึง 1

ข. ถ้า $p > 0$ ทำตามขั้นตอนที่ 1 และ ขั้นที่ 2 จะได้ค่า x_1 ตามสมการที่ (2.29) และหาค่า z ตามสมการที่ (2.17) ดังนั้น

$$r = x_1 + z \qquad (2.31)$$

เนื่องจากค่า r ที่ได้จากสมการที่ (2.29), (2.30) และ (2.31) เป็นค่าปริมาณน้ำฝนรายวันสำหรับ One-parameter Gamma Distribution ดังนั้นทำให้เป็น Two-parameter Gamma Distribution ได้โดยการนำค่า Scale Parameter (λ) มาคูณกับค่า r

2.3 การสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตก (Precipitation Amounts Process) โดย Mixed Exponential Distribution

การสังเคราะห์ค่าปริมาณน้ำฝนรายวันโดยใช้ Mixed Exponential Distribution มีขั้นตอนการสังเคราะห์ 2 ขั้นตอนคือ ขั้นแรกทำการหาค่าพารามิเตอร์ α, β_1 และ β_2 โดยใช้วิธี Method of Maximum Likelihood แล้วนำ Log-likelihood Function ที่ได้มาเป็น Objective Function ของวิธี Nelder and Mead Algorithm ซึ่งเป็นวิธีการ Simplex ที่นำมาหาค่าพารามิเตอร์ที่เหมาะสม (Optimal Parameter Value) โดยไม่ใช้การหาค่าอนุพันธ์ของ Objective Function ซึ่งยุ่งยาก เมื่อได้ค่าพารามิเตอร์ทั้งสามแล้ว ขั้นต่อมาคือการนำค่าพารามิเตอร์ทั้งสามมาทำการสังเคราะห์ค่า Exponential Random Variable โดยในการศึกษาครั้งนี้ใช้วิธี Inverse Probability Distribution ในการสังเคราะห์ ซึ่งขั้นตอนนี้ก็คือการสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตกที่สังเคราะห์ได้โดย Two-state, First-order Markov Chain ในหัวข้อที่ 2.1 นั่นเอง

Mixed Exponential Distribution มี Probability Density Function คือ

$$f(x)=\frac{\alpha}{\beta_1}\exp\left[\frac{-x}{\beta_1}\right]+\frac{1-\alpha}{\beta_2}\exp\left[\frac{-x}{\beta_2}\right] \tag{2.32}$$

โดย x > 0, $0\le\alpha\le 1$ และ $\beta_1\ge\beta_2>0$ โดยที่ $\alpha(n)$ คือ Mixing Probability

2.3.1 การประมาณค่าพารามิเตอร์ของ Mixed Exponential Distribution โดย Method of Maximum Likelihood

การหาค่าพารามิเตอร์ทั้งสามของ Mixed Exponential Distribution ซึ่งได้แก่ α, β_1 และ β_2 สามารถประมาณหาค่าได้โดยใช้ Method of Maximum Likelihood ซึ่งทฤษฎีได้กล่าวไว้ข้างต้นในหัวข้อที่ 2.2.1 แล้ว

จากสมการที่ (2.13) จะได้ว่า Log-likelihood Function ของ Mixed Exponential Distribution เพื่อใช้ในการหาค่าพารามิเตอร์ α, β_1 และ β_2 มีรูปแบบดังนี้ (Woolhiser and Roldán, 1982)

$$\log L=\sum_{i=1}^{n}\left\{\ln\left[\frac{\alpha}{\beta_1}\exp\left(\frac{-x_i}{\beta_1}\right)+\frac{1-\alpha}{\beta_2}\exp\left(\frac{-x_i}{\beta_2}\right)\right]\right\} \tag{2.33}$$

เนื่องจาก Log-likelihood Function ในสมการที่ (2.33) ได้แสดงให้เห็นว่าพารามิเตอร์ทั้งสามของ Mixed Exponential Distribution อยู่ในรูปแบบ Implicit Form ดังนั้นจึงได้นำวิธีการทาง Simplex มาช่วยในการหาค่าพารามิเตอร์ α, β_1 และ β_2 แทนที่จะทำการหาค่าอนุพันธ์ซึ่งยุ่งยากแทน โดยวิธีทาง Simplex ที่เลือกมาใช้ในการศึกษาครั้งนี้คือ Nelder and Mead Algorithm มีทฤษฎีดังนี้

ในการหาค่าพารามิเตอร์ α, β_1 และ β_2 ของ Mixed Exponential Distribution โดยวิธี Nelder and Mead Algorithm ในขั้นตอนแรกจะต้องมีค่าพารามิเตอร์เริ่มต้นของทั้ง 3 พารามิเตอร์นี้ก่อน

ค่าพารามิเตอร์ที่จะนำไปใช้เป็นค่าประมาณเริ่มต้นประมาณได้โดย Method of Moments (Rider, 1961; Woolhiser and Roldán, 1982) ซึ่งคำนวณจากการให้ค่า α มีค่าอยู่ในช่วงตั้งแต่ 0.2 ถึง 0.8 โดยมีระยะห่างแต่ละช่วงเป็น 0.1 ในขณะที่ค่า β_1 มีค่าอยู่ในช่วงตั้งแต่ $0.2\bar{x}_j$ ถึง $0.8\bar{x}_j$ โดยมีระยะห่างแต่ละช่วงเป็น $0.1\bar{x}_j$ (โดยที่ j เท่ากับช่วงเวลาที่ทำการศึกษา) ดังนั้นสำหรับแต่ละคู่ของ α และ β_1 จะทำให้ค่า β_2 สามารถหาได้ดังสมการที่ (2.34)

$$\bar{x}_j = \alpha\beta_1 + (1-\alpha)\beta_2 \tag{2.34}$$

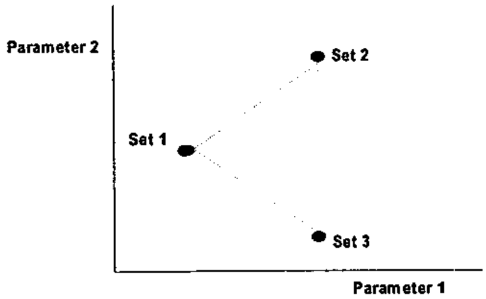
เมื่อได้ค่าพารามิเตอร์เริ่มต้นทั้งหมดของ α, β_1 และ β_2 แล้ว นำไปหาค่าพารามิเตอร์ทั้งสาม โดยการใช้พารามิเตอร์เริ่มต้นทีละชุดเป็นค่าพารามิเตอร์เริ่มต้น วิธีการ Nelder and Mead Algorithm จะให้ค่าพารามิเตอร์ที่ดีที่สุด ซึ่งให้ค่า Objective Function ที่ดีที่สุดออกมา

Nelder and Mead Algorithm (Nelder and Mead, 1965) เป็นวิธีการหาค่าพารามิเตอร์ที่เหมาะสม (Optimal Parameter Value) โดยไม่ใช้การหาค่าอนุพันธ์ของ Objective Function ซึ่งยุ่งยาก แต่จะใช้ขั้นตอนการหาค่าที่ง่ายและตรงไปตรงไปตรงมาไม่ซับซ้อนแทน ในการค้นหาค่าพารามิเตอร์ที่เหมาะสมนี้ ค่าพารามิเตอร์ที่ประมาณได้จะถูกเลือกโดยวิธีซึ่งได้จากการทำซ้ำไปซ้ำมาในขั้นแรกจะทำให้ได้ค่าประมาณที่ดี ทำให้ค่าประมาณที่ไม่ดีจะถูกปฏิเสธ และทำให้เกิดการสังเคราะห์ค่าประมาณที่ดีกว่าขึ้นมาจากค่าประมาณที่ดี

Nelder and Mead ทำการหาค่าพารามิเตอร์ที่เหมาะสมโดยการใช้ค่า Simplex ซึ่งเป็นเซตของค่าพารามิเตอร์ที่ต่างกัน สำหรับแบบจำลองที่มีค่าพารามิเตอร์ n ตัว จะทำให้ Simplex มี n+1 เซตของพารามิเตอร์ที่ต่างกัน เช่น ถ้าแบบจำลองมีค่าพารามิเตอร์ 2 ตัว จะเกิด Simplex ที่ประกอบไปด้วยเซตของค่าประมาณของพารามิเตอร์ทั้ง 2 ตัวอยู่ 3 เซต

ในทางเรขาคณิต แบบจำลองที่มีพารามิเตอร์ n ตัว สามารถทำให้เห็นได้เป็นรูปทรงที่มีมิติ ซึ่ง Simplex นี้จะเป็นเหมือนรูปทรงที่มี n มิติ และแต่ละเซตของพารามิเตอร์ที่เกิดขึ้น

ใน Simplex ซึ่งเท่ากับ n+1 เซตจะทำให้เกิดจุดยอดของรูปทรงที่มีมิตินี้ ในกรณีของแบบจำลองที่มีค่าพารามิเตอร์ 2 ตัว จะทำให้เกิด Simplex เป็นรูปทรง 3 เหลี่ยมที่มี 2 มิติ ดังแสดงในรูปที่ 2.1

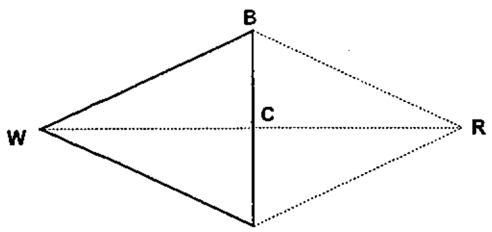


รูปที่ 2.1 Simplex เริ่มต้นสำหรับแบบจำลองที่มี 2 พารามิเตอร์

Nelder and Mead Algorithm เป็นวิธีการที่ทำให้เกิดการเปลี่ยนแปลง Simplex เพื่อหาค่าจุดยอดที่ทำให้เกิดค่าต่ำที่สุดของ Objective Function โดยมีวิธีการดังต่อไปนี้

2.3.1.1 การเปรียบเทียบ (Comparison): ในขั้นตอนการเปลี่ยนแปลงขั้นแรกนี้จะทำการหาจุดยอดของ Simplex ที่ทำให้เกิดค่าที่แย่ที่สุดซึ่งเท่ากับค่าที่มากที่สุดของ Objective Function และจุดยอดที่ให้ค่าที่ดีที่สุดซึ่งเท่ากับค่าที่น้อยที่สุดของ Objective Function จากรูปที่ 2.2 จะให้ชื่อว่า W และ B ตามลำดับ

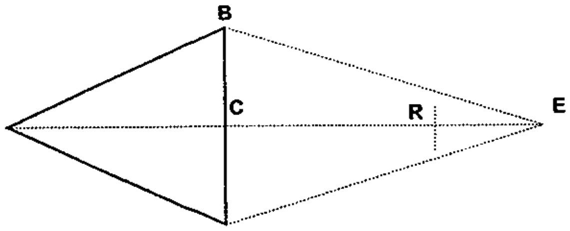
2.3.1.2 การสะท้อน (Reflection): ขั้นตอนต่อมาคือการหาค่าจุดศูนย์กลางของจุดยอดทั้งหมด ยกเว้นจุดยอด W และให้จุดศูนย์กลางนี้ชื่อว่า C ดังรูปที่ 2.2 วิธีการต่อมาจะกำหนดให้เกิดเส้นตรงลากมาจาก W ผ่านจุดศูนย์กลาง C และทำการสะท้อนระยะทาง WC ออกไปในแนวเส้นตรงเพื่อที่จะกำหนดจุดยอดใหม่คือจุด R ดังแสดงในรูปที่ 2.2



Xi (reflected) = Xi (centroid) + 1.0 [Xi (centroid) - Xi (worst)]

รูปที่ 2.2 การสะท้อนกลับของ Simplex

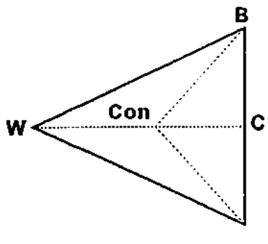
2.3.1.3 การขยายช่วง (Expansion): ถ้าเซตของค่าพารามิเตอร์ที่ได้จากจุดยอด R ให้ค่าที่ดีกว่า หรือมีค่าดีเท่ากับค่าจุดยอดที่ให้ค่าที่ดีที่สุดแล้ว ขั้นตอนต่อไปคือการขยาย Simplex ออกไปในทิศทางเดียวกัน ดังแสดงในรูปที่ 2.3 ซึ่งเป็นการกำหนดจุดยอดที่ทำการขยายออกไปและให้มีชื่อว่า E ดังรูปที่ 2.3 ถ้าจุดยอดที่ขยายออกไปนี้ให้ค่าที่ดีกว่าจุดยอดที่ให้ค่าที่ดีที่สุดแล้ว จุดยอดที่ให้ค่าที่แย่ที่สุดจะถูกแทนที่โดยจุดยอดที่ขยายออกไปนี้ (จุด E) แต่ถ้าจุดยอดที่ขยายออกไปนี้ (จุด E) ให้ค่าที่ไม่ดีเท่ากับจุดยอดที่ให้ค่าที่ดีที่สุดแล้ว จุดยอดที่ให้ค่าที่แย่ที่สุดจะถูกแทนที่โดยจุดยอดที่ถูกสะท้อนออกไป (จุด R)



X_i (expanded) = X_i + 2.0 [X_i (reflected) - X_i (centroid)]

รูปที่ 2.3 การขยายช่วงของ Simplex

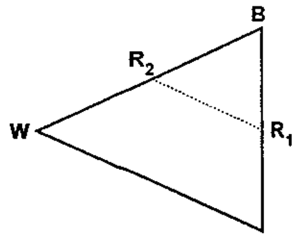
2.3.1.4 การหดช่วง (Contraction): ถ้าจุดยอดที่ถูกสะท้อนออกไป (จุด R) นี้ให้ค่าที่แย่กว่าค่าจุดยอดที่ให้ค่าที่ดีที่สุด แต่ให้ค่าดีกว่าจุดยอดอื่นๆ (ยกเว้นจุดยอดที่ให้ค่าที่แย่ที่สุด) Simplex จะหดช่วงโดยการแทนที่จุดยอดที่ให้ค่าที่แย่ที่สุดด้วยจุดยอดที่สะท้อนออกไป (จุด R) ถ้าจุดยอดที่สะท้อนออกไป (จุด R) ไม่ได้ให้ค่าที่ดีไปกว่าค่าอื่นๆ (ยกเว้นจุดยอดที่ให้ค่าที่แย่ที่สุด) จะทำการหดช่วงของ Simplex จากรูปที่ 2.4 จุดยอดที่ให้ค่าที่แย่ที่สุดจะถูกย้ายไปตามแนวเส้นเข้าไปยังจุดศูนย์กลาง ถ้าค่า Objective Function ของจุดยอดที่ถูกหดช่วงเข้าไปนี้ (จุด con) ให้ค่าที่ดีกว่าจุดยอดที่ให้ค่าที่แย่ที่สุดจะถูกแทนที่โดยจุดยอดที่ถูกหดช่วงเข้าไป (จุด con)



X_i (contracted) = X_i (centroid) - 0.5 [X_i (centroid) - X_i (worst)]

รูปที่ 2.4 การหดช่วงของ Simplex

2.3.1.5 การลดช่วง (Reduction): ถ้าจุดยอดที่ถูกหดช่วงเข้าไป (จุด con) นี้ ไม่ได้ทำให้ค่า Objective Function ดีขึ้น จะทำการลดช่วงของ Simplex โดยย้ายทุกจุดไปทางจุดยอดที่ให้ค่าที่ดีที่สุด จะได้จุดใหม่คือ R_1 และ R_2 ดังแสดงในรูปที่ 2.5



X_{ij} **(reduced) =** X_i **(best) + 0.5 [**X_{ij} **-** X_i **(best)]**

รูปที่ 2.5 การลดช่วงของ Simplex

วิธีการ Nelder and Mead จะสิ้นสุดเมื่อเป็นไปตามเกณฑ์ดังต่อไปนี้

ก. เป็นไปตามสมการ

$$\sqrt{\sum_{j=1, j \,//\, worst}^{n} \frac{(z_j - z_c)^2}{n-1}} < tolerance \qquad (2.35)$$

เมื่อ n = จำนวนของพารามิเตอร์

j = Index of a Vertex

c = Index of Centroid Vertex

z_j และ z_c = ค่าของ Objective Function ของจุดยอด j และ c ตามลำดับ

ข. จำนวนของการทำซ้ำเท่ากับ 50 เท่าของจำนวนของพารามิเตอร์

ค่าพารามิเตอร์ของ Objective Function จะมีค่าเท่ากับค่าของจุดยอดที่ให้ค่าที่ดีที่สุดเมื่อ Nelder and Mead สิ้นสุด และถือว่าค่านี้เป็นค่าพารามิเตอร์ที่มีความเหมาะสม

2.3.2 การสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตกสำหรับวิธี Mixed Exponential Distribution โดยใช้ Inverse Probability Distribution

เมื่อได้ค่าพารามิเตอร์ α, β_1 และ β_2 จากหัวข้อที่ 2.3.1 แล้ว นำค่าพารามิเตอร์ทั้งสามมาทำการสังเคราะห์ค่า Exponential Random Variable โดยวิธี Inverse Probability

Distribution ซึ่งก็คือการสังเคราะห์ค่าปริมาณน้ำฝนรายวันของวันที่เกิดฝนตกนั่นเอง (Wilks, 1988) วิธีการสังเคราะห์มีขั้นตอนดังต่อไปนี้

ก. ทำการสร้าง Uniform Random Number ที่อยู่ในช่วง 0 ถึง 1 (U_t)

ข. นำ U_t มาเปรียบเทียบกับค่า α

ค. ถ้าค่า $U_t < \alpha$ จะเลือกค่า β_1 สำหรับวันที่ t

ง. และในทางกลับกันถ้าค่า $U_t > \alpha$ จะเลือกค่า β_2 สำหรับวันที่ t

จ. ค่าปริมาณน้ำฝนรายวันที่ไม่เป็นศูนย์ของวันที่ t (r_t) จะหาได้จากสมการ

$$r_t = r_{\min} - \beta \ln v_t \tag{2.36}$$

โดย β เท่ากับ β_1 หรือ β_2

และ $v_t \sim U(0,1)$ คือค่า Uniform Random Number ที่อยู่ในช่วง 0 ถึง 1 อีกตัวหนึ่ง

และ $r_{\min}$ คือ ค่าต่ำสุดของปริมาณน้ำฝนรายวันที่วัดได้

2.4 คุณสมบัติทางสถิติที่ใช้ในการตรวจสอบความเหมาะสมของการสังเคราะห์ข้อมูลน้ำฝนรายวัน

คุณสมบัติทางสถิติที่นำมาใช้เป็นตัวตรวจสอบความเหมาะสมว่าการสังเคราะห์ข้อมูลน้ำฝนรายวันโดยใช้วิธีการแจกแจงแบบใดจะให้ค่าใกล้เคียงกับข้อมูลจริงมากที่สุด ได้แก่ ค่าเฉลี่ยเลขคณิต ค่าเบี่ยงเบนมาตรฐาน สัมประสิทธิ์การแปรผัน สัมประสิทธิ์ความเบ้ ค่าปริมาณน้ำฝนรายวันสูงสุดในรอบปีการเกิดซ้ำต่างๆ โดยทฤษฎีกัมเบล และค่า Akaike Information Criterion (AIC) โดยมีรายละเอียดดังต่อไปนี้

2.4.1 ค่าเฉลี่ยเลขคณิต (Arithmetic Mean)

ค่าเฉลี่ยเลขคณิตสำหรับค่าตัวแปรสุ่ม (Random Variable, x) หรือค่าเฉลี่ยเลขคณิตของประชากร (μ) คือ $E(x)$ หาได้จากการอินทิเกรตค่าตัวแปรสุ่ม (x) และ Probability Density Function, $f(x)$ ของตัวแปรสุ่มนั้นๆ บนช่วงที่เป็นไปได้ของค่าตัวแปรสุ่ม โดยมีสมการดังนี้

$$E(x) = \mu = \int_{-\infty}^{\infty} x f(x) dx \tag{2.37}$$

ค่า $E(x)$ คือโมเมนต์ลำดับที่หนึ่งรอบจุดศูนย์กลางของค่าตัวแปรสุ่ม ซึ่งเป็นค่าพารามิเตอร์ที่ใช้วัดจุดกึ่งกลาง หรือ การเข้าสู่ศูนย์กลางของการแจกแจงทางสถิติ

การประมาณค่าเฉลี่ยของข้อมูลตัวอย่างทางสถิติ หาได้จากสมการที่ (2.38)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.38}$$

เมื่อ $\bar{x}$ = ค่าเฉลี่ยเลขคณิต

x_i = ค่าของข้อมูลลำดับที่ i

n = จำนวนของข้อมูลทั้งหมด

2.4.2 ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation)

ค่าการกระจายของข้อมูลวัดได้จากค่าความแปรปรวน (Variance, σ^2) ซึ่งเป็นค่าโมเมนต์ลำดับที่สองรอบค่าเฉลี่ย โดยมีสมการดังนี้

$$E[(x-\mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \tag{2.39}$$

ค่าความแปรปรวนสำหรับข้อมูลตัวอย่างทางสถิติ หาได้จากสมการที่ (2.40)

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2.40}$$

ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation, s) ของข้อมูลตัวอย่างทางสถิติหาได้จากรากที่สองของความแปรปรวน ดังสมการที่ (2.41)

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.41}$$

เมื่อ s = ค่าเบี่ยงเบนมาตรฐาน

$\bar{x}$ = ค่าเฉลี่ยเลขคณิต

x_i = ค่าของข้อมูลลำดับที่ i

n = จำนวนของข้อมูลทั้งหมด

2.4.3 สัมประสิทธิ์การแปรผัน (Coefficient of Variation)

สัมประสิทธิ์การแปรผัน (Coefficient of Variation) หรือ สัมประสิทธิ์ส่วนเบี่ยงเบนมาตรฐาน (Coefficient of Standard Deviation) มีค่าดังสมการที่ (2.42)

$$C_V = \frac{s}{\bar{x}} \tag{2.42}$$

เมื่อ C_V = สัมประสิทธิ์การแปรผัน

s = ค่าเบี่ยงเบนมาตรฐาน

$\bar{x}$ = ค่าเฉลี่ยเลขคณิต

2.4.4 สัมประสิทธิ์ความเบ้ (Coefficient of Skewness)

ความสมมาตรของการแจกแจงรอบค่าเฉลี่ยจะหาได้จากค่าความเบ้ (Skewness) ซึ่งเป็นค่าโมเมนต์ที่สามรอบค่าเฉลี่ย

$$E\left[(x-\mu)^3\right] = \int_{-\infty}^{\infty}(x-\mu)^3 f(x)dx \tag{2.43}$$

ทำการหารค่าความเบ้ (Skewness) ในสมการที่ (2.43) ด้วย s^3 จะได้ค่าสัมประสิทธิ์ความเบ้ (Coefficient of Skewness) ของข้อมูลตัวอย่างทางสถิติ ดังแสดงในสมการ (2.44)

$$C_s = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)s^3} \tag{2.44}$$

เมื่อ C_s = สัมประสิทธิ์ความเบ้

s = ค่าเบี่ยงเบนมาตรฐาน

$\bar{x}$ = ค่าเฉลี่ยเลขคณิต

x_i = ค่าของข้อมูลลำดับที่ i

n = จำนวนของข้อมูลทั้งหมด

2.4.5 ค่าปริมาณน้ำฝนรายวันสูงสุดในรอบปีการเกิดซ้ำต่างๆ โดยทฤษฎีกัมเบล

ทฤษฎีกัมเบล (Gumbel Distribution) หรือ Extreme Value Type I (EVI) (Chow et al., 1988) มี Probability Density Function คือ

$$F(x) = \exp\left[-\exp\left(-\frac{x-u}{\alpha}\right)\right] \quad , -\infty \le x \le \infty \tag{2.45}$$

โดยที่

$$\alpha = \frac{\sqrt{6}s}{\pi} \tag{2.46}$$

และ

$$u = \bar{x} - 0.5772\alpha \tag{2.47}$$

ค่าพารามิเตอร์ u คือค่า Mode ของการแจกแจง (คือจุดที่ให้ค่าสูงสุดของ Probability Density Function) และจะได้ Reduced Variate, y ดังสมการที่ (2.48)

$$y = \frac{x-u}{\alpha} \tag{2.48}$$

แทนค่า Reduce Variate, y ในสมการที่ (2.48) ลงในสมการที่ (2.45) ทำให้ได้สมการที่ (2.49)

$$F(x) = \exp[-\exp(-y)] \tag{2.49}$$

จากสมการที่ (2.49) หาค่า y ได้ดังสมการที่ (2.50)

$$y = -\ln\left[\ln\left(\frac{1}{F(x)}\right)\right] \tag{2.50}$$

สมการที่ (2.51) แสดงความสัมพันธ์ระหว่างค่าความน่าจะเป็นของการเกิดเหตุการณ์ที่สามารถเกิดได้ทุกๆ เหตุการณ์ มีค่าเท่ากับส่วนกลับของรอบปีการเกิดซ้ำ (Return Period, T) เนื่องจาก

$$\begin{aligned}\frac{1}{T} &= P(X \geq x_T) \\ &= 1 - P(x < x_T) \\ &= 1 - F(x_T)\end{aligned}$$

ดังนั้น

$$F(x_T) = \frac{T-1}{T} \tag{2.51}$$

แทนค่าสมการที่ (2.51) ลงในสมการที่ (2.50) จะได้ว่า

$$y_T = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right] \tag{2.52}$$

จากความสัมพันธ์ของ x และ y ในสมการที่ (2.48) สามารถเขียนความสัมพันธ์ของ x_T และ y_T ได้ดังนี้

$$x_T = u + \alpha y_T \tag{2.53}$$

2.4.6 Akaike Information Criterion (AIC)

ทฤษฎีที่นำมาใช้ในการพิจารณาเพื่อช่วยในการตัดสินใจว่าการแจกแจงชนิดใดมีความเหมาะสมกว่ากันคือ Akaike Information Criterion (AIC) (Akaike, 1974; Salas et al., 1980; Woolhiser and Roldán, 1982) มีรูปแบบสมการคือ

$$AIC = -2\log L + 2k \tag{2.54}$$

เมื่อ $\log L$ คือ ค่าของ Log-likelihood Function และ k เท่ากับจำนวนของพารามิเตอร์ โดยการแจกแจงใดที่ให้ค่า AIC น้อยจะเป็นการแจกแจงที่เหมาะสมกว่า

ในการหาค่า AIC ของวิธี Two-parameter Gamma Distribution จะมีรูปแบบของ Log-likelihood Function ดังนี้

$$\log L_g = \sum_{i=1}^{26}\sum_{j=1}^{p(i)}\left[(\delta(i)-1)\log x_{ij} - \frac{x_{ij}}{\lambda(i)} - \delta(i)\log\lambda(i) - \log\Gamma(\delta(i))\right] \qquad (2.55)$$

เมื่อ $p(i)$ = จำนวนของวันที่เกิดฝนตกในช่วงเวลาที่ i

x_{ij} = ค่าปริมาณน้ำฝนของวันที่ j ในช่วงเวลาที่ i

และในการหาค่า AIC ของวิธี Mixed Exponential Distribution มีรูปแบบของ Log-likelihood Function คือ

$$\log L_m = \sum_{i=1}^{26}\sum_{j=1}^{p(i)}\left\{\log\left[\frac{\alpha(i)}{\beta_1(i)}\exp\left(\frac{-x_{ij}}{\beta_1(i)}\right) + \frac{1-\alpha(i)}{\beta_2(i)}\exp\left(\frac{-x_{ij}}{\beta_2(i)}\right)\right]\right\} \qquad (2.56)$$

เมื่อ $p(i)$ = จำนวนของวันที่เกิดฝนตกในช่วงเวลาที่ i

x_{ij} = ค่าปริมาณน้ำฝนของวันที่ j ในช่วงเวลาที่ i

เมื่อหาค่า $\log L_g$ และ $\log L_m$ ได้แล้ว จึงนำไปหาค่า AIC ดังสมการที่ (2.54) โดยที่จำนวนพารามิเตอร์ k ของวิธี Two-parameter Gamma Distribution มีค่าเท่ากับ 52 เนื่องจากวิธี Two-parameter Gamma Distribution มีพารามิเตอร์ 2 ตัว และในการศึกษาครั้งนี้ทำการแบ่งช่วงเวลาเป็น 26 ช่วงเวลา ดังนั้นจำนวนพารามิเตอร์ทั้งหมดที่ใช้จึงมี 52 ตัว ส่วนจำนวนพารามิเตอร์ k ของวิธี Mixed Exponential Distribution มีค่าเท่ากับ 78 เนื่องจากวิธี Mixed Exponential Distribution มีพารามิเตอร์ 3 ตัว และในการศึกษาครั้งนี้ทำการแบ่งช่วงเวลาเป็น 26 ช่วงเวลา ดังนั้นจำนวนพารามิเตอร์ทั้งหมดที่ใช้จึงมี 78 ตัว แล้วนำค่า AIC ที่ได้มาเปรียบเทียบกัน โดยถ้า AIC ของวิธีใดมีค่าน้อยกว่า จะสามารถบอกได้ว่าวิธีนั้นมีความเหมาะสมในการสังเคราะห์ปริมาณน้ำฝนรายวันกว่า