บทที่ 3
แนวทางการศึกษาและทฤษฎีที่ใช้

การวิเคราะห์ความถี่น้ำหลากในครั้งนี้ ประกอบด้วย 2 ส่วน คือ การคัดเลือกฟังก์ชันการแจกแจงความน่าจะเป็น และการวิเคราะห์ความถี่น้ำหลากในลักษณะภูมิภาค ซึ่งทั้ง 2 ส่วนมีแนวทางการศึกษาและทฤษฎีที่ใช้ดังหัวข้อต่อไปนี้

3.1 ฟังก์ชันการแจกแจงความน่าจะเป็น

ฟังก์ชันการแจกแจงความน่าจะเป็น (*probability distribution functions*) เป็นฟังก์ชันทางสถิติที่แสดงถึงความน่าจะเป็นของการเกิดค่าของตัวแปรสุ่ม (random variate) ซึ่งในการวิเคราะห์ขนาดและความถี่น้ำหลากนั้น ฟังก์ชันการแจกแจงจะถูกสมมุติขึ้น และค่าพารามิเตอร์ของฟังก์ชันการแจกแจงที่สมมุติ จะประเมินจากกลุ่มตัวอย่างข้อมูลที่มีการจดบันทึกไว้ ซึ่งในการศึกษาครั้งนี้ ฟังก์ชันการแจกแจงความถี่ที่ใช้ศึกษาเปรียบเทียบ ประกอบด้วย Log-Normal 2 Parameter, Pearson Type III, Log Pearson Type III และ Gumbel เหตุผลในการเลือกฟังก์ชันความถี่ทั้ง 4 แบบ เนื่องจากเป็นวิธีที่นิยมใช้กันอย่างแพร่หลาย

การประมาณค่าพารามิเตอร์ของฟังก์ชันการแจกแจงที่ทำในการศึกษาครั้งนี้ ใช้การประมาณค่าโดยวิธีโมเมนต์ (Moments Method) และวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) ซึ่งรูปแบบของฟังก์ชันการแจกแจงดังกล่าว มีรายละเอียดดังนี้

3.1.1 <u>การแจกแจงแบบ Log-Normal 2 Parameter</u>

ฟังก์ชันการแจกแจงแบบ Log-Normal 2 Parameter คล้ายคลึงกับการแจกแจงแบบ Normal เพียงแต่ใช้ค่า Log ของตัวแปรแทนที่

ถ้าตัวแปรสุ่ม x มีการแจกแจงแบบลอกนอร์มอล ดังนั้น Y = ln x จะมีการแจกแจงแบบ Normal ซึ่งฟังก์ชันความหนาแน่นน่าจะเป็น ที่มีการแจกแจงแบบลอกนอร์มอล 2 พารามิเตอร์ เป็นดังนี้

$$F(x) = \left(\frac{1}{x\sigma_y\sqrt{2\pi}}\right) \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu_y}{\sigma_y}\right)^2\right] \quad \text{สำหรับ } x>0,\ y=\ln x \quad \text{.......................(3.1)}$$

โดย σ_y และ μ_y เป็นพารามิเตอร์ ซึ่ง σ_y คือ ค่าเบี่ยงเบนมาตรฐานของตัวแปร y และ μ_y คือ ค่าเฉลี่ยของตัวแปร y

การประเมินพารามิเตอร์โดยวิธีโมเมนต์ ซึ่งประมาณได้จาก

$$\bar{x} = \exp[\mu_y + (1/2)\sigma_y^2] \qquad \text{.......................................(3.2)}$$

$$S_x^2 = \exp[2[\mu_y + \sigma_y^2](\exp[\sigma_y^2] - 1) \qquad \text{.......................................(3.3)}$$

ค่าพารามิเตอร์จากวิธีจากวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) สามารถหาได้จากสมการ

$$\mu_y = \sum_{i=1}^{n} \ln x_i / n \qquad \text{.......................................(3.4)}$$

$$\sigma_y^2 = \sum_{i=1}^{n} (\ln x_i - \bar{x})^2 / n \qquad \text{.......................................(3.5)}$$

3.1.2 <u>การแจกแจงแบบ Pearson Type III</u>

รูปแบบของฟังก์ชันการแจกแจง มีรายละเอียดดังนี้

$$F(X) = \frac{1}{\alpha\Gamma(\beta)} [(\frac{\chi-\gamma}{\alpha})^{\beta-1} \exp(-(\frac{\chi-\gamma}{\alpha}))] \qquad \text{.......................................(3.6)}$$

เนื่องจากการแจกแจงนี้ คล้ายกับการแจกแจงแบบ Log Pearson Type III การอธิบายค่าพารามิเตอร์รวมทั้งวิธีประเมินพารามิเตอร์ จะแสดงไว้ในหัวข้อ 3.1.3

3.1.3 <u>การแจกแจงแบบ Log Pearson Type III</u>

การแจกแจงนี้ *U.S. Federal Water Resources Council* นำเสนอในปี *ค.ศ. 1967* สำหรับการวิเคราะห์ความถี่น้ำหลาก ซึ่งมีรูปสมการแบบเดียวกับการแจกแจงแบบ Pearson Type III เพียงแต่ใช้ค่า ln x แทนที่ในสมการ โดยกำหนดให้ Y= lnx การแจกแจงนี้เป็นส่วนหนึ่งของการแจกแจงแบบ Log-Normal 2 Parameter เพียงแต่มีค่าสัมประสิทธิ์ความเบ้ (coefficient of skewness , Cs) ในขณะที่การแจกแจงแบบ Log-Normal 2 Parameter มีค่าสัมประสิทธิ์ความเบ้ของอนุกรม ln x เท่ากับ ศูนย์ (Hann,1977) โดยมีรูปสมการดังนี้

$$F(X) = \frac{1}{\alpha\Gamma(\beta)} [(\frac{\ln x-\gamma}{\alpha})^{\beta-1} \exp(-(\frac{\ln x-\gamma}{\alpha}))] \qquad \text{.......................................(3.7)}$$

โดย α คือ scale parameter β คือ shape parameter γ คือ location parameter และ $\Gamma(\beta)$ คือ gamma function ซึ่งค่าพารามิเตอร์ของฟังก์ชันทั้งสองนั้น มีความสัมพันธ์กับค่าเฉลี่ย ความแปรปรวน และค่าสัมประสิทธิ์ความเบ้ โดยประเมินได้ดังนี้

การประเมินพารามิเตอร์โดยวิธีโมเมนต์

$$\bar{Y} = \gamma + S_Y(\beta)^{1/2} \qquad \text{...(3.8)}$$

$$S_y^{\,2} = \beta(\alpha)^2 \qquad \text{.................................(3.9)}$$

$$\beta = 2/(Cs)^2 \qquad \text{................................(3.10)}$$

$$Cs = \hat{\gamma}_1\sqrt{\frac{n(n-1)}{n-2}[1+\frac{8.5}{n}]} \qquad \text{...................................(3.11)}$$

$$\hat{\gamma}_1 = \frac{n}{(n-1).(n-2)s}\sum_{i=1}^{n}(X_i-\bar{X})^3 \qquad \text{................................(3.12)}$$

สำหรับสมการที่ (11) Bobee และ Robitaille (1976) เสนอแนะในการปรับแก้ค่าสัมประสิทธิ์ความเบ้ (coefficient of skewness) เนื่องจากค่าที่ประเมินได้จากข้อมูลโดยตรง ($\hat{\gamma}_1$) มีแนวโน้ม bias

การประเมินพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method)

การแจกแจงแบบ Pearson Type III

$$\frac{1}{\alpha^2}\sum_{i=}^{n}(x_i-\gamma)-\frac{n\beta}{\alpha}=0 \qquad \text{...(3.13)}$$

$$-n\frac{\Gamma'(\beta)}{\Gamma(\beta)}+\sum_{i=}^{n}\ln(x_i-\gamma)-n\ln\alpha=0 \qquad \text{...(3.14)}$$

$$\frac{n}{\alpha}-(\beta-1)\sum_{i=}^{n}(\frac{1}{(x_i-\gamma})=0 \qquad \text{...(3.15)}$$

$$\varphi(\beta)=\ln(\beta+2)-\frac{1}{2(\beta+2)}-\frac{1}{12(\beta+2)^2}-\frac{1}{120(\beta+2)^4}-\frac{1}{252(\beta+2)^6}-\frac{1}{(\beta+1)}-\frac{1}{\beta} \quad \text{...(3.16)}$$

ค่า $\varphi(\beta)$ ในสมการ (3.16) เท่ากับ $\frac{\Gamma'(\beta)}{\Gamma(\beta)}$

การแจกแจงแบบ Log Pearson Type III

$$\sum_{i=}^{n} \ln(x_i - \alpha) = n\alpha\beta \qquad \text{...(3.17)}$$

$$n\varphi(\beta) = \sum_{i=1}^{n} \ln\left[(\ln x_i - \gamma)/\alpha \right] \qquad \text{...(3.18)}$$

$$n = \alpha(\beta - 1)\sum_{i=1}^{n} \frac{1}{\ln[(\ln x_i - \gamma)} \qquad \text{...(3.19)}$$

3.1.4 <u>การแจกแจงแบบ Gumbel</u>

การแจกแจงนี้นำเสนอโดย Gumbel และเป็นวิธีที่ใช้กันอย่างแพร่หลายในประเทศไทย ในการวิเคราะห์ความถี่น้ำหลาก (Sabur, 1982) ซึ่งมีรูปสมการดังนี้

$$F(x) = \alpha \exp\{-\alpha (x-\beta) - \exp[-\alpha (x-\beta)]\} \qquad \text{...(3.20)}$$

โดย α คือค่า concentration parameter และ β คือ ค่า measure of central tendency ซึ่งสามารถหาค่าได้จากสมการดังนี้

การประเมินโดยวิธีโมเมนต์

$$\alpha = 1.2825/ S_x \qquad \text{...(3.21)}$$

$$\beta = \bar{x} - 0.450 S_x \qquad \text{...(3.22)}$$

การประมาณพารามิเตอร์จากวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method)

ซึ่งค่าของ α, β จะหาได้โดยวิธีการคำนวณซ้ำ (Iterative procedure) จากสมการ

$$\sum_{i=1}^{n} x_i \exp(-\alpha x_i) - (x - 1/\alpha)\sum_{i=1}^{n} \exp(-\alpha x_i) = 0 \qquad \text{....................................(3.23)}$$

$$\beta = 1/\alpha \ln(n/\sum_{i=1}^{n} \exp(-\alpha x_i)) \qquad \text{...(3.24)}$$

3.2 สมการความถี่โดยทั่วไป

Chow (1964) เสนอหลักการในการประเมินค่า จากเหตุการณ์ใด ๆ ซึ่งสามารถที่จะหาค่าได้ในรูปของสมการดังนี้

$$x = \bar{X} + \Delta x \qquad \text{...(3.25)}$$

โดยที่ x แทนค่าของเหตุการณ์ใด ๆ และ Δx เป็นส่วนที่เบี่ยงเบนไปจากค่าเฉลี่ย $\bar{X}$ และตั้งสมมุติฐานว่าค่า Δx มีค่าเท่ากับผลคูณของค่าเบี่ยงเบนมาตรฐาน S กับค่าปัจจัยความถี่ (K_T) ดังนั้นสมการความถี่ทั่วไป (*general frequency equation*) สำหรับการคำนวณค่าเหตุการณ์ที่รอบปีการเกิดต่าง ๆ เมื่อแทนค่ารอบปีและค่าเหตุการณ์ของรอบปีนั้นด้วยสัญลักษณ์ T และ X_T ตามลำดับคือ

$$X_T = \bar{X} + K_T S \qquad \text{...(3.26)}$$

และ สมการที่ (3.26) สามารถจัดให้อยู่ในรูปแบบได้ดังนี้

$$X_T = \bar{X} + (1 + C_V K_T) \qquad \text{...(3.27)}$$

เมื่อ X_T = ปริมาณการไหลที่รอบปีการเกิดซ้ำเฉลี่ย T ปี

S = ค่าเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างข้อมูล

$\bar{x}$ = ค่าเฉลี่ยของตัวอย่างข้อมูล

C_V = ค่าสัมประสิทธิ์ของความแปรปรวน (coefficient of variation)

K_T = ค่าปัจจัยความถี่ (frequency factor)

ค่าปัจจัยความถี่ (*frequency factor*, K_T) ในแต่ละการแจกแจงมีการจัดทำไว้ในรูปของตาราง และสามารถที่จะประเมินจากฟังก์ชันของการแจกแจงความน่าจะเป็นได้ดังนี้ (Kite, 1977)

<u>ฟังก์ชันการแจกแจงแบบ Log-normal 2 parameter</u>

$$K_T = \frac{\exp\ \{[\ln(1+c_v^{\,2})]^{1/2}\, t - [\ln(1+c_v^{\,2})]/2\} - 1}{c_v} \qquad \text{............(3.28)}$$

เมื่อ $C_v = \frac{S}{\bar{X}}$ = ค่าสัมประสิทธิ์ของความแปรปรวน

ค่า t จากสมการ (3.28) คือ ค่าปัจจัยความถี่ของการแจกแจงแบบปกติ เท่ากับ standard normal deviate และ Chow (1964) ได้แสดงตารางสำหรับหาค่า t จากค่าสัมประสิทธิ์ความเบ้ (C_s) หรือจากค่าค่าสัมประสิทธิ์ของความแปรปรวนไว้ด้วย แต่ค่า K_T ที่ประเมินจากการแจกแจงแบบนี้ ใช้ค่าสัมประสิทธิ์ความเบ้, $C_s = 0$

<u>การแจกแจงแบบ Pearson Type III</u>

$$K_T \approx t+(t^2-1)t+\frac{C_s}{6}+\frac{1}{3}(t^3-6t)(\frac{C_s^{\,2}}{6})-(t^2-1)(\frac{C_s^{\,3}}{6})+t(\frac{C_s^{\,4}}{6})+\frac{1}{3}(\frac{C_s^{\,5}}{6}) \quad \text{..............(3.29)}$$

เมื่อ C_s = สัมประสิทธิ์ความเบ้ของตัวอย่างข้อมูล

t = ค่า standard normal deviate

<u>การแจกแจงแบบ Log Pearson Type III</u>

ค่า K_T ของการแจกแจงนี้ ประเมินจากสมการเดียวกันกับการแจกแจงแบบ Pearson Type III สมการ (3.29) เพียงแต่ค่า C_s ประเมินจากข้อมูลที่จัดในรูป logarithm

<u>ฟังก์ชันการแจกแจงแบบ Gumbel</u>

$$K_T = \frac{Y_T-\bar{Y}_n}{S_n} \quad \text{..........(3.30)}$$

โดยที่ $Y_T = -\ln[-\ln(T-1)/T)]$ ....(3.31)

$$\bar{Y}_n = \frac{1}{2}\sum_{i=1}^{n} Y_m \quad \text{..........(3.32)}$$

$$S_n = [\frac{1}{n}\sum_{i=1}^{n}(Y_m-\bar{Y}_n)^2]^{1/2} \quad \text{..........(3.33)}$$

เมื่อ n = จำนวนข้อมูล

m = ลำดับข้อมูลซึ่งได้จากการจัดเรียง

T = รอบปีการเกิดซ้ำ ในการศึกษาประเมินจากสมการของ Weibull (Benson, 1962)

$= (n+1)/m$

Y_n = ค่าเฉลี่ยของอนุกรมข้อมูล x

ซึ่งจากสมการ (3.30) นี้ ถ้าค่า $n \rightarrow \infty$ และแทนค่า α, β จะได้ค่า K_T ดังนี้

$$K_T = -\{0.45+0.7797 \ln(-\ln[1-1/T])\} \qquad \text{...(3.34)}$$

3.2 การทดสอบความเหมาะสมของฟังก์ชันการแจกแจงความน่าจะเป็น

การทดสอบความเหมาะสม (Test of Goodness of Fit) ของการแจกแจงความน่าจะเป็นแบบต่าง ๆ มีวัตถุประสงค์เพื่อหาฟังก์ชันการแจกแจงที่สามารถปรับเข้ากับกลุ่มตัวอย่างข้อมูลทางอุทกวิทยาได้ดีที่สุด ซึ่ง Kite (1977) กล่าวถึงผลการวิจัยเกี่ยวกับการเลือกฟังก์ชันการแจกแจง ซึ่งแสดงให้เห็นว่าจะไม่มีฟังก์ชันการแจกแจงแบบใดแบบหนึ่งที่จะมีความเหมาะสมกับปรากฏการณ์ทางอุทกวิทยาได้ทั้งหมด ในทุกพื้นที่ และทุกชุดข้อมูล ดังนั้น จึงต้องมีการทดสอบความเหมาะสมของฟังก์ชันการแจกแจงความน่าจะเป็น ก่อนเลือกใช้ในการวิเคราะห์

ในการศึกษานี้ เป็นการทดสอบเพื่อพิจารณาว่ากลุ่มตัวอย่างของข้อมูลปริมาณน้ำสูงสุดรายปีที่มีการจดบันทึกไว้ในพื้นที่ลุ่มน้ำภาคเหนือและภาคตะวันออกเฉียงเหนือ มีลักษณะการแจกแจงแบบใดแบบหนึ่งตามที่กำหนดหรือไม่ ซึ่งวิธีการที่นิยมใช้ในการทดสอบมี 3 วิธี คือ การทดสอบแบบ Chi-Square การทดสอบแบบ Least Square และการทดสอบแบบ Kolmogorov-Smirnov ซึ่งทั้ง Chi-Square และวิธี Kolmogorov-Smirnov ทดสอบความเชื่อมั่นที่ระดับ 95% (Yevjevich, 1972)

3.3.1 <u>วิธี Chi-Square</u>

เป็นการทดสอบความเหมาะสม โดยพิจารณาจากการเปรียบเทียบการแจกแจงความถี่ที่ได้จากการสังเกต และความถี่ของเหตุการณ์ที่ได้จากการคาดหมายไว้ (expected) ตามลักษณะการแจกแจงที่ทดสอบ ตัวทดสอบสถิติที่ใช้ ประเมินจากสมการดังต่อไปนี้

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad \text{...(3.35)}$$

โดย χ^2 = ค่าสถิติ Chi-Square

o_i = ค่าความถี่ที่ได้จากการสังเกตในชั้นที่ i

E_i = ค่าความถี่ที่คาดว่าจะได้ในชั้นที่ i ตามลักษณะการแจกแจงที่ทดสอบ

k = จำนวนชั้น (class)

ในการกำหนดจำนวนและระยะของช่วงชั้น (number and length of class intervals) มีผู้เสนอแนะและแสดงความคิดเห็นไว้หลายแบบ โดย Markovic (1965) แสดงความคิดเห็นเกี่ยวกับการมีจำนวนชั้น (classes) มากเกินไปนั้น อาจทำให้ในบางชั้นไม่มีค่าของเหตุการณ์อยู่เลย หรืออาจมีน้อยเกินไปจนทำให้ผลคำนวณที่ได้ผิดปกติไป (irregular) แต่ถ้ามีจำนวนชั้นน้อยเกินไป อาจทำให้ผลการทดสอบออกมาไม่เด่นชัด ซึ่งที่ผ่านมายังไม่มีการกำหนดจำนวนชั้นให้เป็นที่ยอมรับกันโดยทั่วไป นักสถิติส่วนมากจะกำหนดจำนวนชั้นไม่ควรน้อยกว่า 10 ชั้น และไม่มากกว่า 20 ชั้น (โดยไม่ได้มีทฤษฎีพิสูจน์ยืนยันหลักการนี้) ในทางปฏิบัตินั้นโดยทั่วไปกำหนดเหตุการณ์ที่คาดหมายในแต่ละช่วงชั้นไม่ควรน้อยกว่า 5 ค่า และมีจำนวนช่วงชั้นไม่น้อยกว่า 5 ช่วงชั้น [Yevjevich (1977), จุฬาลงกรณ์มหาวิทยาลัย คณาจารย์วิชาคณิตศาสตร์ (2523)]

สำหรับการกำหนดระยะช่วงชั้นกำหนดได้ 2 แบบ คือ

ก. แต่ละช่วงชั้นมีการเพิ่มค่าของเหตุการณ์เท่า

ข. แต่ละช่วงชั้นมีการเพิ่มค่าความน่าจะเป็นเท่ากัน ถ้าแบ่งค่าความน่าจะเป็นที่เหตุการณ์เกิดขึ้นออกเป็นช่วงเท่า ๆ กัน จำนวนความถี่ของข้อมูลที่คาดหมาย $E_j = n/k$ เมื่อ n คือจำนวนข้อมูล

ซึ่งในการศึกษานี้ ได้กำหนดช่วงชั้นแบบเพิ่มค่าความน่าจะเป็นเท่ากัน โดยมีจำนวนชั้นต่ำสุดเท่ากับ 5 และเกณฑ์พิจารณาความเหมาะสม โดยการเปรียบเทียบค่า χ^2 กับค่าวิกฤติของ Chi-Square (χ_o^2) ที่ได้จากตารางสถิติที่ระดับความเชื่อมั่น 95 % และชั้นแห่งความเป็นอิสระ (degree of freedom) เท่ากับ k-p-1, p คือจำนวนพารามิเตอร์ของฟังก์ชันการแจกแจงที่ทดสอบ

เกณฑ์การทดสอบ คือ ถ้า $\chi^2 < \chi_o^2$ หมายถึงการยอมรับสมมุติฐานว่ากลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะตามการแจกแจงที่ทดสอบ ที่ระดับความเชื่อมั่น 95 % , $\chi^2 \geq \chi_o^2$ เป็นการไม่ยอมรับสมมุติฐานว่ากลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะตามการแจกแจงที่ทดสอบ ที่ระดับความเชื่อมั่น 95 % ซึ่งค่า χ^2 จะมีค่าน้อย เมื่อความถี่ที่ได้จากการสังเกตและความถี่ที่คาดว่าจะได้ตามทฤษฎี มีค่าใกล้เคียงกัน และ ค่า χ^2 จะมากเมื่อความถี่ต่างกัน และการทดสอบแบบ Chi-Square นี้ ยังใช้ในการเปรียบเทียบความเหมาะสมในแต่ละการแจกแจง เมื่อผลการทดสอบสามารถยอมรับความเหมาะสมของฟังก์ชันการแจกแจงมากกว่า 1 แบบ และใช้เกณฑ์ตัดสินว่าฟังก์ชันที่สามารถปรับเข้ากับข้อมูลได้ดีที่สุด จะมีค่า χ^2 น้อยที่สุด

3.3.2 วิธี Kolmogorov-Smir nov

การทดสอบแบบ Kolmogorov-Smimov เป็นการทดสอบที่ใช้เกณฑ์ผลต่างสูงสุดของความน่าจะเป็นสะสมของค่าที่ได้จากการ plotting position (empirical frequency distribution) กับค่าที่ประเมินได้จากฟังก์ชันการแจกแจง ซึ่งมีขนาดของข้อมูลเท่ากัน ผลต่างสูงสุดของค่าดังกล่าว ประเมินจากสมการดังต่อไปนี้

$$D_n = \max_{x} \left| F(x) - S_n(x) \right| , \forall x \qquad \text{.......................................(3.36)}$$

โดย D_n = ค่าทดสอบสถิติ Kolmogorov-Smirnov

$F(x)$ = ความน่าจะเป็นสะสมของ x ตามลักษณะของฟังก์ชันการแจกแจงที่ทดสอบ

$S_n(x)$ = ความน่าจะเป็นสะสมของ x ที่ประมาณจากข้อมูล โดย $S_n(x)$ ได้จากการ plotting position โดยวิธี Weibull คือ $\frac{m}{n+1}$ m คือลำดับของข้อมูลที่ได้จากการจัดเรียง และ n คือจำนวนข้อมูล

ค่า D_n โดยทั่วไปสามารถหาได้ง่ายและสะดวก โดยการวาดกราฟการแจกแจงความถี่ของข้อมูล (Empirical Frequency Distribution) เทียบกับกราฟที่ได้จากฟังก์ชันการแจกแจง และประเมินค่า D_n จากกราฟ

เกณฑ์การทดสอบ พิจารณาจากการนำค่า D_n ที่ประเมินได้เปรียบเทียบกับค่า D_c ซึ่งเป็นค่าวิกฤตของ Kolmogorov-Smirnov ซึ่งค่า D_n กำหนดจากความสัมพันธ์ของขนาดกลุ่มตัวอย่างข้อมูล (n) และระดับนัยสำคัญ (α) ในการศึกษา กำหนดค่าระดับนัยสำคัญเท่ากับ 5 % ซึ่งเป็นเกณฑ์ที่ใช้โดยทั่วไป

หลักการพิจารณา คือ

1. ถ้า $D_n < D_c$ หมายถึงการยอมรับสมมุติฐานว่ากลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะตามการแจกแจงที่ทดสอบ ที่ระดับนัยสำคัญ α

2. $D_n \geq D_c$ หมายถึงการไม่ยอมรับสมมุติฐานว่ากลุ่มตัวอย่างสุ่มมาจากประชากรที่มีลักษณะตามการแจกแจงที่ทดสอบ ที่ระดับนัยสำคัญ α

3. ในการเปรียบเทียบแต่ละการแจกแจง การแจกแจงที่สามารถปรับเข้ากับข้อมูลได้ดีที่สุด จะมีค่าสถิติ D_n น้อยที่สุด

3.3.4 วิธี Least Square

การทดสอบแบบ Least Square เป็นอีกทางเลือกหนึ่งของการทดสอบความเหมาะสม นอกเหนือจากการทดสอบแบบ Chi-square และ Kolmogorov-Smirnov โดยอาศัยหลักการของผลรวมของผลต่างกำลังสองระหว่างขนาดของเหตุการณ์ที่ได้จากการฟังก์ชันการแจกแจง กับค่าที่ได้จากการสังเกตทุกค่าที่มีความน่าจะเป็นเท่ากัน โดยอาศัยหลักการกำลังสองน้อยที่สุด (Least Square) ซึ่งสามารถหาค่าได้จากสมการดังนี้

$$SE = \left[\sum_{i=1}^{n} \frac{(X_i - Y_i)^2}{n-p} \right]^{1/2} \qquad \text{..(3.37)}$$

โดย SE = ค่าความคลาดเคลื่อนมาตรฐาน (standard error)

X_i = ข้อมูลปริมาณน้ำหลากสูงสุดรายปีที่จดบันทึกไว้ และมีการจัดเรียงข้อมูลในลำดับ i

Y_i = ปริมาณน้ำหลากที่ประเมินจากฟังก์ชั่นการแจกแจงและสอดคล้องกับลำดับ i

p = จำนวนพารามิเตอร์ของฟังก์ชั่นการแจกแจงที่ทดสอบ

n = จำนวนข้อมูล

การทดสอบวิธีนี้มีข้อจำกัด คือ ผลที่ได้จากการคำนวณขึ้นอยู่กับวิธีของ plotting position ที่เลือกใช้ในการประมาณค่าความน่าจะเป็นของค่าเหตุการณ์ทุก ๆ ค่าของชุดข้อมูลที่บันทึกไว้ และการทดสอบวิธีนี้จะใช้เฉพาะการเปรียบเทียบความเหมาะสมในแต่ละการแจกแจงเท่านั้น เกณฑ์ที่ใช้ในการทดสอบ คือ ฟังก์ชันการแจกแจงใดที่สามารถปรับเข้ากับข้อมูลได้ดีที่สุด จะมีค่าความแตกต่าง (SE) น้อยสุด

3.4 การวิเคราะห์ความถี่น้ำหลากในลักษณะภูมิภาค

การวิเคราะห์ความถี่น้ำหลากในลักษณะภูมิภาค (Regional Flood Frequency Analysis) เป็นวิธีการที่มีประโยชน์และมีความจำเป็นอย่างมากต่อการวิเคราะห์ขนาดและความถี่ของน้ำหลากในบริเวณที่ไม่มีการเก็บข้อมูล หรือที่ที่มีการเก็บข้อมูลแต่สถิติความยาวข้อมูลสั้นเกินไป (น้อยกว่า 10 ปี) ซึ่งการวิเคราะห์ความถี่น้ำหลากในลักษณะภูมิภาค จะพิจารณาจากสถานีวัดน้ำอื่น ๆ ที่มีสถิติความยาวข้อมูลเพียงพอ โดยสถานีนั้นมีความคล้ายกันทางสภาพภูมิอากาศและอุตุอุทกวิทยา และหารูปแบบความสัมพันธ์ เพื่อนำไปประยุกต์ใช้กับบริเวณ หรือสถานีที่พิจารณาซึ่งมีข้อจำกัดทางด้านข้อมูล

3.4.1 แนวทางการศึกษา

จากการศึกษาที่ผ่านมาโดยทั่วไป พบว่า การเกิดสภาพน้ำหลากในพื้นที่ลุ่มน้ำต่าง ๆ ขึ้นอยู่กับปัจจัยต่าง ๆ ที่เข้ามาเกี่ยวข้อง เช่น ปริมาณฝนที่ตกในพื้นที่ สภาพภูมิประเทศ และลักษณะการใช้ประโยชน์ที่ดิน ดังนั้น ในการศึกษานี้ จึงได้นำรูปแบบความสัมพันธ์แบบสมการสหสัมพันธ์เชิงซ้อน (multiple regression) ใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างปริมาณน้ำหลากในรอบปีการเกิดต่าง ๆ กับปัจจัยต่าง ๆ ในพื้นที่ลุ่มน้ำ ซึ่งในการศึกษา แบ่งการวิเคราะห์ความถี่ในลักษณะภูมิภาคออกเป็น 2 พื้นที่ คือ พื้นที่ลุ่มน้ำภาคเหนือ และภาคตะวันออกเฉียงเหนือ โดยพิจารณาจากสถานีวัดน้ำท่าในภูมิภาค ที่มีขนาดพื้นที่ลุ่มน้ำตั้งแต่ 100 – 3,900 ตารางกิโลเมตร โดยในแต่ละพื้นที่พิจารณาความสัมพันธ์ของปริมาณน้ำหลากในรอบปีการเกิด 2, 5, 10, 20, 50 และ 100 ปี กับข้อมูลลักษณะทางกายภาพของลุ่มน้ำ พื้นที่ป่าไม้ในลุ่มน้ำ และแปรความสัมพันธ์ระหว่างปริมาณฝนสูงสุดรายวันกับปริมาณน้ำหลาก ออกเป็น 3 กรณี คือ

กรณีที่ 1 พิจารณาปริมาณน้ำหลากในรอบปีการเกิด ต่าง ๆ สัมพันธ์กับปริมาณฝนในเทอมเดี่ยว

$$Q_T = f(R_T, A, S, L, Lc, F, H) \quad \text{...(1)}$$

โดยที่
- A = พื้นที่ลุ่มน้ำ (ตร.กม.)
- R_T = ปริมาณฝนสูงสุดรายวันในรอบปีการเกิด T ปี (มม.)
- S = ความลาดชันเฉลี่ยของลำน้ำ (%)
- L = ความยาวของลำน้ำหลัก (กม.)
- Lc = ความยาวของลำน้ำหลักจากจุดใกล้ศูนย์ถ่วงถึงจุดออกของลำน้ำ (กม.)
- F = พื้นที่ป่าไม้ปกคลุมในพื้นที่ลุ่มน้ำ (%)
- H = ความสูงของพื้นที่ลุ่มน้ำ (ม.)

กรณีที่ 2 พิจารณาปริมาณน้ำหลากในรอบปีการเกิดต่าง ๆ สัมพันธ์กับปริมาตรฝน

$$Q_T = f((R_T{*}A), A, S, L, Lc, F, H) \quad \text{...(2)}$$

กรณีที่ 3 พิจารณาอัตราส่วนของปริมาณน้ำหลากในรอบปีการเกิดต่าง ๆ กับขนาดพื้นที่ลุ่มน้ำสัมพันธ์กับปริมาณฝนในเทอมเดี่ยว

$$Q_T/A = f(R_T, A, S, L, Lc, F, H) \quad \text{...(3)}$$

โดยค่า Q_T และ R_T ของแต่ละสถานี ได้จากการวิเคราะห์ขนาดและความถี่ด้วยวิธีการแจกแจงที่สามารถปรับเข้ากับข้อมูลได้ดีที่สุด จาก 4 ฟังก์ชันการแจกแจงที่นำมาทดสอบ และในการศึกษาหาความสัมพันธ์ระหว่างปริมาณน้ำหลากกับองค์ประกอบต่าง ๆ ของลุ่มน้ำนี้ ทำการตรวจสอบสัมประสิทธิ์สหสัมพันธ์ของสมการ ที่ระดับนัยสำคัญ 0.05 หรือระดับนัยสำคัญ 5 % และเปรียบเทียบผลการศึกษาระหว่างภาคเหนือและภาคตะวันออกเฉียงเหนือ ในเชิงความแตกต่างของลักษณะพื้นที่

3.4.2 รูปแบบสมการสหสัมพันธ์เชิงซ้อน

สมการสหสัมพันธ์เชิงซ้อน (multiple regression) ตามแนวทางการศึกษา มีรูปแบบของสมการทั่วไปดังนี้

3.4.2.1 *สมการทั่วไปของสมการสหสัมพันธ์เชิงซ้อน*

สมการสหสัมพันธ์เชิงซ้อน ใช้เพื่อตรวจสอบความสัมพันธ์ระหว่างปริมาณน้ำหลากในรอบปีการเกิดต่าง ๆ ซึ่งเป็นตัวแปรตาม (dependent variable) และปัจจัยต่าง ๆ ในพื้นที่ลุ่มน้ำซึ่งเป็นกลุ่มตัวแปรอิสระ (group of independent variables) สมการทั่วไปของสหสัมพันธ์เชิงซ้อนประกอบด้วย ตัวแปรตาม 1 ตัว และตัวแปรอิสระ ตั้งแต่ 2 ตัวขึ้นไป โดยมีรูปสมการดังต่อไปนี้

$$Q_T = bo\, X_1^{b1} X_2^{b2} \ldots\ldots X_n^{bn} \qquad (3.38)$$

โดย Q_T = ปริมาณน้ำหลากในรอบปีการเกิดต่าง ๆ (ตัวแปรตาม)

X_1 ถึง X_n = ข้อมูลลักษณะทางกายภาพของลุ่มน้ำ, พื้นที่ป่าไม้ในพื้นที่ลุ่มน้ำ และปริมาณฝนสูงสุดรายวันที่สอดคล้องกับรอบปีการเกิด (กลุ่มตัวแปรอิสระ)

bo = ค่าคงที่ของสมการสหสัมพันธ์เชิงซ้อน

b1 ถึง bn = ค่าสัมประสิทธิ์ของสมการสหสัมพันธ์เชิงซ้อนสำหรับตัวแปรอิสระต่าง ๆ

จากสมการ (3.38) แปลงให้อยู่ในรูปของสมการสหสัมพันธ์เชิงซ้อนแบบเส้นตรง ได้ดังนี้

$$\ln(Q_T) = Y = bo + b1 \ln(X_1) + b2 \ln(X_2) + \ldots\ldots + bn \ln(X_n) \qquad (3.39)$$

ในการประเมินสมการสหสัมพันธ์เชิงซ้อนแบบเส้นตรงของสมการ (3.39) ใช้วิธีกำลังสองน้อยสุด (Least Square Method) โดยทำให้ผลรวมกำลังสองของส่วนเบี่ยงเบน (ผลต่างของค่าสังเกต และค่าที่ประเมินได้จากสมการสหสัมพันธ์เชิงซ้อน) มีค่าน้อยสุด

เกณฑ์ที่ใช้ตรวจสอบสมการสหสัมพันธ์ที่ประเมินได้ พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน (multiple correlation coeffcient, R) และค่าความคลาดเคลื่อนมาตรฐาน (standard errors of estimate, SEE) โดยประเมินจากสมการดังต่อไปนี้

1. **ค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน** (multiple correlation coeffcient, R)

$$R = Sreg / S$$

$$= \sqrt{\left(\sum_{i=1}^{n}(\hat{Y}-\bar{Y})^2 / \sum_{i=1}^{n}(Y-\bar{Y})^2\right.} \quad \text{.....................................(3.41)}$$

โดยที่

Sreg = ส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามที่อธิบายด้วยสมการสหสัมพันธ์เชิงซ้อน

S = ส่วนเบี่ยงเบนมาตรฐานทั้งหมด เท่ากับ ผลรวมของค่าส่วนเบี่ยงเบนมาตรฐานของ ตัวแปรตามที่อธิบายได้ด้วยสมการสหสัมพันธ์ กับส่วนเบี่ยงเบนมาตรฐานของตัวแปรตาม ซึ่งเกิดจากแหล่งอื่นที่อธิบายไม่ได้

$\hat{Y}$ = ค่าปริมาณน้ำหลาก (ตัวแปรตาม) ที่ประเมินจากสมการสหสัมพันธ์

$\bar{Y}$ = ค่าเฉลี่ยของปริมาณน้ำหลาก (ตัวแปรตาม) ที่นำมาวิเคราะห์

Y = ค่าปริมาณน้ำหลากที่นำมาวิเคราะห์

ขนาดของค่า R มีค่าอยู่ระหว่าง 0.0 ถึง ± 1.0 ถ้าค่าเข้าใกล้ +1.0 แสดงว่ามีความสัมพันธ์กันโดยตรงในขั้นที่ดี และถ้าค่าเข้าใกล้ 0.0 แสดงว่า มีความสัมพันธ์กันน้อย และถ้าค่าเข้าใกล้ –1.0 แสดงว่ามีความสัมพันธ์ผกผันในขั้นที่ดี

2. ค่าความคลาดเคลื่อนมาตรฐาน (standard errors of estimate,SEE)

$$SEE = \sqrt{\left(\sum_{i=1}^{n}(Y-\hat{Y})^2 /(n-k-1)\right.} \quad \text{.....................................(3.42)}$$

โดยที่

SEE = ส่วนเบี่ยงเบนมาตรฐานของตัวแปรตาม ซึ่งเกิดจากแหล่งอื่นที่อธิบายไม่ได้ด้วยสมการสหสัมพันธ์เชิงซ้อน

k = จำนวนตัวแปรอิสระ

n = จำนวนข้อมูล

โดยที่ ค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน และความคลาดเคลื่อนมาตรฐานของสมการสหสัมพันธ์เชิงซ้อนที่ประเมินได้ในแตะละภูมิภาคนั้น จะทำการเปรียบเทียบกับสมการความสัมพันธ์ระหว่างปริมาณน้ำหลากในรอบปีการเกิดต่าง ๆ กับตัวแปรพื้นที่ลุ่มน้ำอย่างเพียงเดียว ซึ่งเป็นรูปแบบความสัมพันธ์ที่ใช้กันทั่วไปในประเทศไทย