

การย้ายข้อมูล เพื่อเพิ่มความเชื่อถือได้ในการประเมินความถี่ของน้ำท่วม ตอน 1

TRANSFER OF INFORMATION TO IMPROVE ESTIMATES OF FLOOD FREQUENCIES, PART I

วราวุธ วุฒินิชย์ Ph. D. *

1. บทนำ

พารามิเตอร์ที่สำคัญที่สุดในการแจกแจงความเป็นไปได้ (Probability Distribution) ของการเกิดน้ำท่วมคือ ค่าเฉลี่ย (Mean) แวเรียนซ์ (Variance) และสัมประสิทธิ์ความบิดเบี้ยว (Skewness Coefficient) เนื่องจากไม่รู้ค่าที่แท้จริงของพารามิเตอร์เหล่านี้จึงจำเป็นต้องหาจากข้อมูลน้ำท่วม อย่างไรก็ตาม ความเชื่อถือได้ในการหาค่าพารามิเตอร์ขึ้นอยู่กับจำนวนข้อมูลน้ำท่วมที่มีอยู่ ถ้ามีข้อมูลน้อยการประเมินค่าพารามิเตอร์จะมีโอกาสผิดพลาดมากแต่ความผิดพลาดเหล่านี้จะลดลงได้โดยการย้ายข้อมูล (Transfer of information) น้ำท่วมจากสถานีวัดน้ำใกล้เคียงซึ่งมีข้อมูลยาวนานกว่ามาใช้ในการหาค่าพารามิเตอร์

ข้อมูลน้ำท่วมของสถานีที่มีข้อมูลยาวนานกว่าจะสามารถนำเอามาใช้เป็นพื้นฐานในการขยายข้อมูลน้ำท่วม (Extending flood records) ของสถานีที่มีข้อมูลสั้นกว่าที่วัดในช่วงเวลาเดียวกันได้ ข้อมูลที่ถูกขยาย (Extended flood records) นี้ จะสามารถนำไปใช้หาค่าพารามิเตอร์ใหม่ได้ อย่างไรก็ตาม การขยายข้อมูลอาจไม่ทำให้การหาค่าพารามิเตอร์มีความเชื่อถือได้มากขึ้นก็ได้ ทั้งนี้ขึ้นอยู่กับความสอดคล้องกันของข้อมูลที่น่ามาใช้

ในที่นี้จะกล่าวถึงเทคนิคการวิเคราะห์แบบรีเกรสชัน (Regression) ในการย้ายข้อมูลน้ำท่วมจากสถานีใกล้เคียงที่มีข้อมูลยาวนานกว่ามาใช้ที่สถานีที่มีข้อมูลสั้นกว่า เพื่อประโยชน์ดังต่อไปนี้

* ผู้ช่วยศาสตราจารย์ ภาควิชาวิศวกรรมชลประทาน คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตร-
ศาสตร์ กำแพงแสน

1. เพิ่มความถูกต้องในการหาค่าพารามิเตอร์ของข้อมูลน้ำท่วมที่กำหนดให้
2. ขยายข้อมูลน้ำท่วมที่กำหนดให้ ให้ยาวขึ้น
3. หาข้อมูลน้ำท่วมที่ขาดหายไป (Missing flood records)

ตอนที่ 1 นี้ จะกล่าวถึงการใช้รีเกรชันเส้นตรงแบบง่ายในการย้ายข้อมูล พร้อมทั้งยกตัวอย่างประกอบ และในตอนที่ 2 จะได้กล่าวถึงการใช้รีเกรชันเส้นตรงแบบหลายตัวแปรในการย้ายข้อมูล

2. รีเกรชันเส้นตรงแบบง่าย (Simple Linear Regression)

สมการรีเกรชันเส้นตรงแบบง่ายเป็นที่นิยมใช้กันมากในการย้ายข้อมูลอุทกวิทยาระหว่างสองสถานะ

ให้ X เป็นตัวแปรแรนด้อม (Random) ของสถานะที่มีข้อมูลยาวนานกว่า และ Y เป็นตัวแปรแรนด้อมของสถานะที่มีข้อมูลสั้นกว่า

ซึ่งไม่จำเป็นว่าตัวแปรแรนด้อมทั้งสองตัวจะต้องเป็นชนิดเดียวกัน เช่น ถ้า X เป็นปริมาณน้ำท่วม Y อาจเป็นปริมาณน้ำท่วมหรือน้ำฝน หรืออื่นๆ ก็ได้ สมมติให้ N_1 เป็นจำนวนข้อมูลชุดที่สั้นกว่า และ $(N_1 + N_2)$ เป็นจำนวนข้อมูลชุดที่ยาวกว่า ชุดข้อมูล X และ Y จะแทนด้วย

$$X_1, X_2, \dots, X_{N_1+1}, \dots, X_{N_1+N_2}$$

$$Y_1, Y_2, \dots, Y_{N_1}$$

สมมติว่าตัวแปรแรนด้อม X และ Y มีการแจกแจงความเป็นไปได้แบบปกติแบบมีตัวแปร 2 ตัว (Bivariate Normal Distribution) โดยมี $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$, และ r เป็นพารามิเตอร์ ซึ่ง

μ_X และ σ_X^2 คือค่าเฉลี่ยและแวกเรียนซ์ของป๊อปูเลชัน (Population) ของตัวแปรแรนด้อม X

μ_Y และ σ_Y^2 คือค่าเฉลี่ยและแวกเรียนซ์ของป๊อปูเลชันของตัวแปรแรนด้อม Y

r คือค่าสัมประสิทธิ์สหพันธ์ระหว่างตัวแปร (Cross-Correlation Coefficient)

ถ้าการแจกแจงของตัวแปร X และ Y ไม่เป็นแบบปกติตามที่สมมติไว้ จะต้องมีการแปลงข้อมูล (Transformation) ให้มีการแจกแจงแบบปกติ ซึ่งโดยทั่วไปแล้วการแปลงข้อมูลด้วยลอการิทึม (Logarithmic transformation) จะให้ค่าที่มีการแจกแจงเป็นแบบปกติโดยประมาณ

สมการรีเกรชันเส้นตรงแบบง่ายระหว่างตัวแปร X และ Y คือ

$$Y_t = a + bX_t \quad \text{----- (1)}$$

เมื่อ Y_t คือค่ารีเกรชันของค่า X_t ที่กำหนดให้

a และ b คือพารามิเตอร์ของป้อนป้อนของสมการ

อย่างไรก็ตามสมการที่ (1) ไม่ได้อธิบายแวกเรียนซ์ทั้งหมดของ Y อธิบายเพียงแค่ r^2 ของแวกเรียนซ์เท่านั้น ข้อบกพร่องอื่นนี้จะแก้ไขได้โดยการเพิ่มแวกเรียนซ์ $(1 - r^2) \sigma_Y^2$ ทางด้านขวามือของสมการที่ (1) สมการรีเกรชันเส้นตรงแบบง่ายจะสามารถเขียนได้ใหม่ดังนี้

$$Y_t = a + bX_t + \alpha(1-r^2)^{1/2} \sigma_Y \epsilon_t \quad \text{---- (2)}$$

เมื่อ ϵ_t คือตัวแปรแรนด้อมปกติ (Normal Random Variate) ซึ่งมีค่าเฉลี่ยเป็นศูนย์ และแวกเรียนซ์เท่ากับหนึ่ง ϵ_t เรียกสั้นๆ ว่านอยส์ สัมประสิทธิ์ $\theta = 0$ สำหรับสมการที่ไม่มีนอยส์ และ $\theta = 1$ สำหรับสมการที่มีนอยส์ และสัมประสิทธิ์ α จะใช้เพื่อลดการลำเอียง (Bias) ในการใช้แวกเรียนซ์ประเมิน (estimated variance) ของ Y (Matalas and Jacobs, 1964)

3. การประเมินค่าพารามิเตอร์

ค่าประเมินของพารามิเตอร์ a และ b ของสมการที่ (2) ที่หาโดยวิธีลีสแควร์ (Least Square) คือ

$$\hat{a} = \bar{Y}_1 - \hat{b} \bar{X}_1 \quad \text{----- (3)}$$

และ $\hat{b} = \hat{r} S_1(Y)/S_1(X)$ ----- (4)

เมื่อสัมประสิทธิ์สหสัมพันธ์ $\hat{r}$ หาได้จากสมการ

$$\hat{r} = \frac{\frac{1}{N_1} \sum_{t=1}^{N_1} (X_t - \bar{X}_1)(Y_t - \bar{Y}_1)}{S_1(X) \cdot S_1(Y)} \text{ ----- (5)}$$

ค่าเฉลี่ย $\bar{X}_1, \bar{Y}_1$ และความเบี่ยงเบนมาตรฐาน $S_1(X)$ และ $S_1(Y)$ คำนวณได้จากสมการ

$$\bar{X}_1 = \frac{1}{N_1} \sum_{t=1}^{N_1} X_t \text{ ----- (6)}$$

$$\bar{Y}_1 = \frac{1}{N_1} \sum_{t=1}^{N_1} Y_t \text{ ----- (7)}$$

$$S_1(X) = \left[\frac{1}{N_1-1} \sum_{t=1}^{N_1} (X_t - \bar{X}_1)^2 \right]^{1/2} \text{ ----- (8)}$$

$$\text{และ } S_1(Y) = \left[\frac{1}{N_1-1} \sum_{t=1}^{N_1} (Y_t - \bar{Y}_1)^2 \right]^{1/2} \text{ ----- (9)}$$

และหาค่าสัมประสิทธิ์ α ได้จากสมการ (Matalas and Jacobs, 1964)

$$\alpha^2 = \frac{N_2(N_1-4)(N_1-1)}{(N_2-1)(N_1-3)(N_1-2)} \text{ ----- (10)}$$

สมการที่ (2) จะสามารถเขียนใหม่ในรูปของพารามิเตอร์ประเมินได้ดังนี้

$$Y_t = \bar{Y}_1 + \hat{r} \frac{S_1(Y)}{S_1(X)} (X_t - \bar{X}_1) + \alpha \theta \sqrt{1 - \hat{r}^2} S_1(Y) \epsilon_t \text{ ----- (11)}$$

สมการที่ (11) สามารถใช้ขยายข้อมูล $Y_1, \dots, Y_{N_1}$ โดยใช้ข้อมูลที่ยาวกว่า $X_1, \dots, X_{N_1}, \dots, X_{N_1+N_2}$ หรือพูดง่าย ๆ ใช้สมการที่ (11) คำนวณค่า $\hat{Y}_{N_1+1}, \dots, \hat{Y}_{N_1+N_2}$ จากข้อมูล $X_{N_1+1}, \dots, X_{N_1+N_2}$ ซึ่งข้อมูลที่ขยายให้ยาวขึ้นแล้ว $Y_1, \dots, Y_{N_1}, \hat{Y}_{N_1+1}, \dots, \hat{Y}_{N_1+N_2}$ จะสามารถใช้คำนวณหาค่าเฉลี่ยและแปรปรวนใหม่ได้ดังนี้

$$\bar{Y} = \bar{Y}_1 + \frac{N_2}{(N_1+N_2)} \hat{b} (\bar{X}_2 - \bar{X}_1) \quad \text{-----} (12)$$

$$\begin{aligned} \text{และ } S^2(Y) = & \frac{1}{(N_1+N_2-1)} [(N_1-1)S_1^2(Y) + (N_2-1)\hat{b}^2 S_2^2(X) + \\ & \frac{N_1 N_2}{(N_1+N_2)} \hat{b}^2 (\bar{X}_2 - \bar{X}_1)^2 + (N_2-1) \theta^2 \alpha^2 (1-\hat{r}^2) \cdot \\ & S_1^2(Y)] \quad \text{-----} (13) \end{aligned}$$

$$\text{เมื่อ } \bar{X}_2 = \frac{1}{N_2} \sum_{t=N_1+1}^{N_1+N_2} X_t$$

$$\text{และ } S_2^2(X) = \frac{1}{(N_2-1)} \sum_{t=N_1+1}^{N_1+N_2} (X_t - \bar{X}_2)^2 \quad \times$$

4. เกณฑ์ในการตัดสินใจว่าค่าพารามิเตอร์ที่ประเมินมีความเชื่อถือได้มากขึ้นหรือไม่

ในการขยายข้อมูล Y ที่วัดได้จากสถานีวิจัยน้ำแห่งหนึ่งโดยใช้ข้อมูลที่ยาวกว่า X จากสถานีวิจัยน้ำอีกแห่งหนึ่งตามวิธีเกรซซัน อาจเป็นที่สงสัยว่าข้อมูล N_2 ค่าที่เพิ่มขึ้นจะช่วยให้การประเมินค่าพารามิเตอร์ของ Y ดีขึ้นหรือไม่ อาจเป็นไปได้ว่าข้อมูล N_2 ค่าที่เพิ่มขึ้น

ทำให้ค่า พารามิเตอร์ที่ประเมินได้มีความเชื่อถือได้น้อยลง จึงจำเป็นต้องมีกฎเกณฑ์ในการตัดสินว่า พารามิเตอร์ที่ประเมินจากข้อมูลที่ถูกขยายให้ยาวขึ้นมีความเชื่อถือได้มากขึ้นหรือไม่ ปกติจะใช้ แวเรียนซ์ของพารามิเตอร์เป็นตัววัดความเชื่อถือได้ในการประเมิน

ถ้าแวลเรียนซ์ของพารามิเตอร์ที่คำนวณจากข้อมูลที่ขยายให้ยาวขึ้น ($N_1 + N_2$) มีค่ามากกว่าแวลเรียนซ์ของข้อมูลเดิม (N_1) แสดงว่าข้อมูลที่ขยายให้ยาวขึ้นได้ความเชื่อถือได้ในการประเมินค่าพารามิเตอร์น้อยลง (Fiering, 1963) ในทางกลับกันถ้าแวลเรียนซ์จากข้อมูลที่ขยายให้ยาวขึ้นน้อยกว่าค่าแวลเรียนซ์จากข้อมูลเดิม การขยายข้อมูลโดยใช้เทคนิครีเกรชันจะทำให้การประเมินค่าพารามิเตอร์มีความเชื่อถือได้มากขึ้นและควรที่จะนำมาใช้

ในทางปฏิบัติมักใช้หลักข้อมูลสัมพัทธ์ (Relative Information) ในการวัดความเชื่อถือได้ของพารามิเตอร์ที่ประเมิน ข้อมูลสัมพัทธ์ I คืออัตราส่วนของแวลเรียนซ์ $\text{Var}(\hat{\gamma}_1)$ ของพารามิเตอร์ที่ประเมินจากข้อมูลเดิมซึ่งมี N_1 ค่าต่อแวลเรียนซ์ $\text{Var}(\hat{\gamma})$ ของข้อมูลที่ถูกต้องให้ยาวขึ้นซึ่งมี ($N_1 + N_2$) ค่า หรือ

$$I = \frac{\text{Var}(\hat{\gamma}_1)}{\text{Var}(\hat{\gamma})} \quad \text{-----} \quad (14)$$

ถ้า I มากกว่าหนึ่ง หมายความว่าแวลเรียนซ์ของค่าพารามิเตอร์ ที่ประเมินจากข้อมูลเดิมมากกว่าแวลเรียนซ์ของพารามิเตอร์ที่ประเมินจากข้อมูลที่ถูกต้องหรือหมายความว่าวิธีรีเกรชันช่วยลดแวลเรียนซ์ (Fiering, 1963) ทำให้การประเมินค่าพารามิเตอร์จากข้อมูลที่ขยายให้ยาวขึ้นให้ค่าที่เชื่อถือได้มากกว่าค่าที่ประเมินได้จากข้อมูลเดิม กรณีนี้แสดงว่าควรใช้วิธีรีเกรชันในการขยายข้อมูล

5. เกณฑ์ในการตัดสินว่าค่าเฉลี่ยที่ประเมินได้มีความเชื่อถือได้มากขึ้น

ถ้าใช้สมการที่ (11) ในการขยายข้อมูล และใช้สมการที่ (12) ในการหาค่าเฉลี่ยจากข้อมูลที่ขยายให้ยาวขึ้น Fiering (1963) และ Metalas and Jacobs (1964) เสนอแนะวิธีการตรวจสอบความเชื่อถือได้ โดยใช้ค่าแวลเรียนซ์ของ $\bar{Y}$ ซึ่งคำนวณได้จากสมการ

$$\text{Var}(\bar{Y}) = \frac{\sigma_Y^2}{N_1} \left[1 - \frac{N_2}{N_1 + N_2} \left(r^2 - \frac{(1-r^2)}{(N_1-3)} \right) \right] \quad (15)$$

ถ้า $\bar{Y}$ เป็นตัวแทนของ μ_Y ได้ดีกว่า $\bar{Y}_1$

$$\text{Var}(\bar{Y}) < [\text{Var}(\bar{Y}_1) = \frac{\sigma_Y^2}{N_1}]$$

$$\text{หรือ } r > \left(\frac{1}{N_1 - 2} \right)^{\frac{1}{2}} \quad (16)$$

ค่าทางด้านขวามือของสมการที่ (16) เรียกว่าสัมประสิทธิ์สหสัมพันธ์ต่ำสุดวิกฤต (Critical Minimum Correlation Coefficient) ซึ่งเป็นเกณฑ์ในการตัดสินใจว่าเฉลี่ยที่ประเมินได้มีความเชื่อถือได้มากขึ้นหรือไม่

6. เกณฑ์ที่ใช้ตัดสินว่าค่าแนวโน้มที่ประเมินได้มีความเชื่อถือได้มากขึ้น

วิธีรีเกรชันจะเพิ่มความเชื่อถือได้ในการประเมินแนวโน้มถ้า

$$\text{Var}[S^2(Y)] < \text{Var}[S_1^2(Y)]$$

เมื่อ $\text{Var}[S^2(Y)]$ คือแนวโน้มของแนวโน้มที่ประเมินจากข้อมูลที่ขยายให้ยาวขึ้นแล้ว

$\text{Var}[S_1^2(Y)]$ คือแนวโน้มของแนวโน้มที่ประเมินจากข้อมูลเดิม

ในการใช้สมการรีเกรชันที่ (11) ในการขยายข้อมูล Metalas and Jacobs (1964) แนะนำการคำนวณแนวโน้มของ $S^2(Y)$ ที่ประเมินได้จากข้อมูลที่ขยายให้ยาวขึ้นโดยใช้สมการ

$$\text{Var}[S^2(Y)] = \frac{2\sigma_Y^4}{(N_1-1)} + \frac{N_2\sigma_Y^4}{(N_1+N_2-1)^2} [Ar^2 + Br^2 + C] \quad (17)$$

เมื่อ A, B และ C เป็นฟังก์ชันของ N_1, N_2 และ θ

$$A = \frac{(N_1-6)(N_2+2)(N_1-8)}{(N_1-5)(N_1-3)} - \frac{2N_2(N_1-4)^2}{(N_1-3)^2} \theta^2 + \frac{4(N_1-4)(\theta^2-2)}{(N_1-3)} \quad (18)$$

$$B = \frac{6(N_2+2)(N_1-6)}{(N_1-5)(N_1-3)} + \frac{2(N_1^2-N_1-14)}{(N_1-3)} - \frac{2(N_1+N_2-1)(N_1-4)(1-\theta^2)}{(N_1-3)} + \frac{2N_2(N_1-4)(N_1-5)}{(N_1-3)^2} \theta^2 - \frac{2(N_1-4)(N_1+3)}{(N_1-3)} \theta^2 - \frac{2N_1N_2(N_1-4)^2 \theta^4}{(N_1-2)(N_1-3)^2} \quad (19)$$

และ

$$C = \frac{2(N_1+1)}{(N_1-3)} + \frac{3(N_2+2)}{(N_1-5)(N_1-3)} - \frac{(N_1+1)(2N_1+N_2-2)}{(N_1-1)} + \frac{2(N_1-4)(N_1+N_2-1)}{(N_1-3)(1-\theta^2)} + \frac{2N_2(N_1-4)}{(N_1-3)^2} \theta^2 + \frac{2(N_1+1)(N_1-4)}{(N_1-3)} \theta^2 + \frac{N_1N_2(N_1-4)^2}{(N_1-2)(N_1-3)^2} \theta^4 \quad (20)$$

$S^2(Y)$ ของข้อมูลที่ถูกขยายให้ยาวขึ้นจะเป็นตัวแทนของ σ_Y^2 ดีกว่า $S_1^2(Y)$ ของข้อมูลเดิม ถ้า

$$\text{Var} [S^2(Y)] < [\text{Var}(S_1^2(Y) = \frac{2\sigma_Y^4}{(N_1-1)}]$$

หรือ

$$|r| > \left[\frac{-B + \sqrt{B^2 - 4AC}}{2A} \right]^{1/2} \quad \text{-----} \quad (21)$$

ค่าทางด้านขวามือของสมการที่ (21) คือค่าสัมประสิทธิ์สหสัมพันธ์ต่ำสุดวิกฤต ตารางที่ 1 และ 2 แสดงค่าต่ำสุดวิกฤตของสัมประสิทธิ์สหสัมพันธ์สำหรับ N_1, N_2 และ $\theta = 0$ และ $\theta = 1$ ตามลำดับ ค่าในตารางแสดงให้เห็นว่า ปกติค่า r วิกฤตสำหรับกรณี $\theta = 0$ จะสูงกว่ากรณีที่ $\theta = 1$ และเมื่อ $\theta = 0$ ค่า r วิกฤตจะเพิ่มเมื่อ N_1 และ N_2 เพิ่ม เมื่อ $\theta = 1$ ผลจะตรงกันข้าม

7. ข้อควรสังเกต

เกณฑ์ในการวัดว่าค่าเฉลี่ยของแวลเรียนซ์ของพารามิเตอร์ที่ประเมินได้มีความเชื่อถือได้เพิ่มขึ้นหรือไม่ มีสมมติฐานว่าข้อมูลทั้งสิ้นและยาวมีการแจกแจงเป็นแบบปกติ ถ้าข้อมูลไม่ได้มีการแจกแจงเป็นแบบปกติ จะต้องแปลงข้อมูลเดิมให้อยู่ในรูปการแจกแจงแบบปกติเสียก่อนโดยใช้ลอกอะริทึม หรือสแควร์รูท

เนื่องจากไม่มีเกณฑ์ในการวัดความเชื่อถือได้ของพารามิเตอร์อื่นนอกจากค่าเฉลี่ยและแวลเรียนซ์ จึงต้องใช้วิธีซิมูเลชัน (Simulation) ในการหาเกณฑ์โดยประมาณในการวัดว่าขนาดน้ำท่วมที่มีความเป็นไปได้ 1% ($Q_{0.01}$) หรือ 2% ($Q_{0.02}$) มีความเชื่อถือได้มากขึ้นหรือไม่ หรือเพื่อหาเกณฑ์ในการลดค่าผิดพลาดมาตรฐาน (Standard error) ของการประเมินค่า $Q_{0.01}$ และ $Q_{0.02}$ ลง 10% ค่าสัมประสิทธิ์สหสัมพันธ์ต่ำสุดในตารางที่ 3 จะใช้วัดว่าการประเมินน้ำท่วมที่มีความเป็นไปได้ 1 หรือ 2% จะมีความเชื่อถือได้เพิ่มขึ้นอย่างน้อย 10% (U.S. Water Resources Council, 1976)

ตาราง 1 ค่าตัวสุกฤกษ์ของ r ในการเพิ่มความเชื่อถือได้ในการประเมินความเสี่ยง : $\theta = 0$ สมการที่ 2

N_1 N_2	6	8	10	12	14	16	18	20	25	30	35	40	45	50	55	60
6.	.81	.69	.63	.63	.63	.64	.65	.65	.66	.67	.68	.68	.68	.69	.69	.69
8.	.82	.72	.70	.70	.70	.71	.72	.72	.73	.73	.74	.74	.74	.75	.75	.75
10.	.82	.74	.73	.73	.73	.74	.75	.75	.76	.76	.77	.77	.77	.78	.78	.78
12.	.82	.75	.74	.75	.75	.76	.76	.77	.78	.78	.79	.79	.79	.80	.80	.80
14.	.82	.76	.75	.76	.76	.77	.77	.78	.79	.80	.80	.80	.81	.81	.81	.81
16.	.82	.77	.76	.77	.77	.78	.78	.79	.80	.81	.81	.82	.82	.82	.82	.82
18.	.83	.77	.77	.77	.78	.78	.79	.80	.81	.81	.82	.82	.83	.83	.83	.83
20.	.83	.77	.77	.78	.79	.80	.81	.82	.83	.83	.84	.84	.85	.85	.85	.85
25.	.83	.78	.78	.79	.80	.81	.82	.83	.84	.85	.85	.86	.86	.86	.86	.86
30.	.83	.78	.79	.80	.81	.82	.83	.84	.85	.85	.86	.86	.87	.87	.87	.87
35.	.83	.79	.79	.80	.81	.82	.83	.84	.85	.85	.86	.87	.87	.87	.88	.88
40.	.83	.79	.79	.80	.81	.82	.83	.84	.85	.85	.86	.87	.87	.87	.88	.88
45.	.83	.79	.79	.80	.81	.82	.83	.84	.85	.85	.86	.87	.87	.87	.88	.88
50.	.83	.79	.79	.80	.81	.82	.83	.84	.85	.85	.86	.87	.87	.87	.88	.88
55.	.83	.79	.80	.81	.82	.83	.84	.85	.85	.86	.87	.87	.87	.88	.88	.88
60.	.83	.79	.80	.81	.82	.83	.84	.85	.85	.86	.87	.87	.88	.88	.88	.88

ตาราง 2 ค่าตัวสุกฤกษ์ของ r ในการเพิ่มความเชื่อถือได้ในการประเมินความเสี่ยง : $\theta = 1$ สมการที่ 2

N_1 N_2	6	8	10	12	14	16	18	20	25	30	35	40	45	50	55	60
6.	.82	.71	.64	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
8.	.82	.71	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
10.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
12.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
14.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
16.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
18.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
20.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
25.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
30.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
35.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
40.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
45.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
50.	.82	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
55.	.83	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28
60.	.83	.72	.65	.59	.55	.52	.49	.47	.42	.39	.36	.34	.32	.30	.29	.28

ตาราง 3 ค่า r ค่าสุดซึ่งใช้เป็นเกณฑ์แสดงความเชื่อถือได้ของการประเมิน $Q_{0.01}$ และ $Q_{0.02}$ เพิ่มขึ้น 10% (U.S. Water Resources Council, 1976)

r	$(N_1 + N_2)/N_2$
0.9	2.0
0.85	3.0
0.80	4.0

8. ตัวอย่าง

ข้อมูลน้ำท่วมประจำเดือนของ Trinchera Creek และ Ute Creek แสดงไว้ในตารางที่ 4 ต้องการประเมินหาค่าเฉลี่ย และแนวร้อยละของน้ำท่วมประจำปีของ Trinchera Creek ที่มีความเชื่อถือได้มากขึ้น โดยการย้ายข้อมูลจาก Ute Creek ซึ่งมีข้อมูลยาวนานกว่า

ในตัวอย่างนี้จะแสดงวิธีการย้ายข้อมูลน้ำท่วมประจำเดือนจาก Ute Creek มาใช้ที่ Trinchera Creek เพื่อเพิ่มความเชื่อถือได้ในการประเมินค่าพารามิเตอร์ของน้ำท่วมประจำปีที่ Trinchera Creek

การคำนวณมีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 ให้ Y'_t คือปริมาณน้ำท่วมประจำเดือนของ Trinchera Creek ระหว่าง 1961-1970 ซึ่งมีจำนวนข้อมูล $\tilde{N}_1 = 12(10) = 120$ เดือน

X'_t คือปริมาณน้ำท่วมประจำเดือนของ Ute Creek ระหว่าง 1956-1975 ซึ่งมีจำนวนข้อมูล $\tilde{N}_1 + \tilde{N}_2 = 12(20) = 240$ เดือน หรือ $\tilde{N}_2 = 120$ เดือน

ข้อมูลปริมาณน้ำท่วมประจำเดือนระหว่าง 1961-1970 ของทั้งสองลำน้ำจะถูกนำมาใช้หาความสัมพันธ์ร่วม

คือความสัมพันธ์ร่วม

ขั้นที่ 2 ใช้ลอการิทึมแปลงข้อมูลน้ำท่วมประจำเดือนให้มีการแจกแจงเป็นแบบปกติดังนี้

$$Y_t = \log Y'_t \quad (\text{สำหรับ Trinchera Creek})$$

$$X_t = \log X'_t \quad (\text{สำหรับ Ute Creek})$$

ขั้นที่ 3 ใช้สมการที่ (5) หาสัมประสิทธิ์สหสัมพันธ์ $\hat{r}$ ของข้อมูล X_t และ Y_t ในช่วง 1961-1970, $\hat{r} = 0.856$

ขั้นที่ 4 สำหรับ $\tilde{N}_1 = 120$ ค่าวิกฤต r ในการเพิ่มความเชื่อถือได้ของการประเมินค่าเฉลี่ย หาได้จากสมการที่ (16)

$$r = \left(\frac{1}{118} \right)^{\frac{1}{2}} = 0.092$$

ค่า $\hat{r} = 0.856$ ซึ่งมากกว่าค่า r ค่าวิกฤต (0.092) แสดงว่าการย้ายข้อมูลจาก Ute Creek มาใช้ที่ Trinchera Creek ช่วยเพิ่มความเชื่อถือได้ในการประเมินค่าเฉลี่ย

จากตารางที่ 2 เมื่อ $\theta = 1$, $\tilde{N}_1 = 120$ และ $\tilde{N}_2 = 120$
 r จะน้อยกว่า 0.28 (เมื่อ $N_1 = 60$ และ $N_2 = 60$,
 $r = 0.28$)

เนื่องจาก $\hat{r} (= 0.856)$ มากกว่า r ค่าวิกฤต (0.28) แสดงว่าการย้ายข้อมูลด้วยวิธีเรขาคณิตตรงแบบง่ายช่วยเพิ่มความเชื่อถือได้ทั้งในการประเมินค่าเฉลี่ยและแนวโน้มของลอการิทึมของน้ำท่วมประจำเดือนของ Trinchera Creek

ขั้นที่ 5 ใช้สมการที่ (11) ขยายข้อมูลน้ำท่วมประจำเดือนของ Trinchera Creek ในช่วง (1956-1960) และ (1971-1975) โดยอาศัยข้อมูลของ Ute Creek ซึ่งมีขั้นตอนในการประเมินดังต่อไปนี้

- หาค่าเฉลี่ยและแนวโน้มของลอการิทึมของข้อมูลน้ำท่วมประจำเดือนของ Trinchera Creek ระหว่าง 1961-1970

$$\bar{Y}_1 = 1.253$$

$$S_1(Y) = 0.354$$

ตาราง 4 ข้อมูลน้ำท่วมประจำเดือน (Monthly Floods) ของ Trinchera Creek (1961-1970)
และ Ute Creek (1956-1975) ที่ยูทิลิตี้ Fort Garland รัฐโคโลราโด

	Trinchera Creek											
	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT
1961	15,000	12,000	9,600	8,500	7,000	13,000	43,000	102,000	104,000	35,000	39,000	26,000
1962	16,000	13,000	9,000	9,000	11,000	21,000	75,000	115,000	60,000	31,000	18,000	10,000
1963	11,000	8,200	8,800	5,500	7,000	18,000	16,000	16,000	13,000	12,000	8,200	10,000
1964	7,500	6,900	6,200	6,400	6,500	9,000	23,000	60,000	48,000	18,000	17,000	13,000
1965	10,000	9,500	7,600	0,400	8,600	8,300	35,000	81,000	95,000	58,000	36,000	21,000
1966	15,000	13,000	13,000	12,000	7,600	11,000	28,000	53,000	43,000	22,000	26,000	11,000
1967	8,300	9,100	8,700	7,000	7,500	9,500	11,000	20,000	23,000	15,000	20,000	13,000
1968	11,000	10,000	8,300	5,000	6,000	9,500	16,000	87,000	116,000	47,000	44,000	20,000
1969	14,000	13,000	8,600	8,600	8,400	10,000	30,000	87,000	84,000	44,000	41,000	31,000
1970	18,000	16,000	20,000	12,000	12,000	14,000	35,000	119,000	97,000	38,000	25,000	22,000

Ute Creek

	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT
1956	18,000	16,000	7,800	5,000	7,000	30,000	72,000	81,000	124,000	44,000	27,000	8,200
1957	11,000	10,000	8,200	4,500	5,000	6,000	24,000	67,000	91,000	27,000	49,000	9,600
1958	16,000	14,000	6,500	6,700	7,000	12,000	54,000	149,000	132,000	24,000	68,000	36,000
1959	2,400	3,200	500	1,600	3,000	5,000	15,000	74,000	152,000	189,000	101,000	41,000
1960	7,400	7,000	6,100	2,800	2,000	5,800	11,000	30,000	28,000	4,500	8,200	700
1961	11,000	9,000	11,000	5,500	5,500	11,000	55,000	102,000	86,000	34,000	49,000	52,000
1962	18,000	12,000	9,000	9,500	10,000	19,000	64,000	85,000	46,000	37,000	21,000	14,000
1963	8,800	6,200	5,000	4,800	6,400	22,000	17,000	14,000	9,600	22,000	28,000	27,000
1964	3,100	2,700	1,800	1,700	2,000	3,500	14,000	33,000	19,000	14,000	34,000	8,400
1965	7,700	4,000	3,500	4,000	3,500	9,600	49,000	71,000	152,000	108,000	124,000	76,000
1966	27,000	14,000	14,000	8,000	7,500	19,000	38,000	69,000	53,000	101,000	80,000	10,000
1967	7,800	7,800	8,800	5,000	5,500	8,800	16,000	33,000	70,000	55,000	63,000	107,000
1968	15,000	12,000	7,000	5,000	5,000	8,800	15,000	95,000	104,000	112,000	168,000	26,000
1969	14,000	11,000	5,600	6,400	6,000	12,000	51,000	71,000	95,000	76,000	141,000	19,000
1970	17,000	12,000	8,000	7,000	7,500	7,000	29,000	74,000	92,000	68,000	41,000	136,000
1971	22,000	16,000	15,000	11,000	11,000	13,000	15,000	22,000	30,000	33,000	12,000	20,000
1972	18,000	10,000	10,000	7,500	8,000	11,000	15,000	20,000	21,000	8,000	8,500	16,000
1973	21,000	10,000	6,400	5,500	6,200	7,800	31,000	198,000	172,000	96,000	46,000	15,000
1974	12,000	7,300	5,000	4,000	6,500	13,000	20,000	58,000	22,000	13,000	24,000	4,700
1975	7,500	7,300	4,700	4,000	4,500	7,500	42,000	117,000	111,000	89,000	25,000	28,000

- หาค่าเฉลี่ยและแนวเรียงนซ์ของลอกอะริทึมของข้อมูลน้ำท่วมประจำเดือนของ UTe Creek ระหว่าง 1961-1970

$$\bar{X}_1 = 1.274$$

$$S_1(X) = 0.492$$

$$\hat{r} = 0.856$$

$$\theta = 1$$

$$\alpha = \left[\frac{120(120-4)(120-1)}{(120-1)(120-3)(120-2)} \right]^{\frac{1}{2}} \\ = 1.004$$

แทนค่าสัมประสิทธิ์ต่างๆ ลงในสมการที่ 11 จะได้สมการรีเกรชันเส้นตรงแบบง่าย

$$Y_t = 1.253 + 0.856 \left(\frac{(0.354)}{(0.492)} \right) (X_t - 1.274) + \\ 1.004(1)(1-0.856^2)^{\frac{1}{2}}(0.354) \epsilon_t$$

$$Y_t = 0.468 + 0.616 X_t + 0.184 \epsilon_t$$

เมื่อ

$$Y_t = \text{ปริมาณน้ำท่วมประจำเดือนของ Trinchera Creek}$$

$$X_t = \text{ปริมาณน้ำท่วมประจำเดือนของ Ute Creek}$$

$$\epsilon_t = \text{ตัวแปรแรนด้อมซึ่งมีการแจกแจงแบบปกติ (Normal Variate) มีค่าเฉลี่ยเท่ากับศูนย์ และมีแนวเรียงนซ์เท่ากับหนึ่ง}$$

สมการรีเกรชันเส้นตรงแบบง่ายนี้สามารถขยายข้อมูลของ Trinchera Creek ระหว่าง (1956-1960) และ (1971-1975) ได้ โดยใช้ข้อมูลของ Ute Creek และเลขแรนด้อมปกติที่มีค่าเฉลี่ยเท่ากับ 0 และแนวเรียงนซ์เท่ากับ 1 ที่สร้างขึ้นมา (Generated Normal Random Numbers [0, 1])

ยกตัวอย่าง จากตารางที่ 4 ปริมาณน้ำท่วมในเดือนกรกฎาคม 1959 ของ Ute Creek คือ $X'_t = 189$ ลบ.ฟุต/วินาที หรือ $X_t = \log 189 = 2.276$ เลขแรมค่อมปกติ $[0, 1]$ ที่สร้างขึ้น $\epsilon_t = -0.348$

$$Y_t = 0.468 + 0.616(2.276) + 0.184(-0.348) \\ = 1.8062$$

$$Y'_t = \log^{-1}(1.8062) = 64.01 \text{ ลบ.ฟุต/วินาที}$$

∴ ปริมาณน้ำท่วมของ Trinchera Creek ในเดือนกรกฎาคม 1959 ที่คำนวณได้ = 64.01 ลบ.ฟุต/วินาที

ในทำนองเดียวกันจะสามารถคำนวณหาปริมาณน้ำท่วมของ Trinchera Creek ในเดือนต่างๆ ระหว่าง (1956-1960) และ (1971-1975) ได้ ดังแสดงในตารางที่ 5 และจากข้อมูลน้ำท่วมประจำปีจะสามารถคำนวณหาปริมาณน้ำท่วมประจำปีระหว่าง (1956-1960) และ (1971-1975) ได้ ดังแสดงไว้ในตารางที่ 6

ขั้นที่ 6 ค่าเฉลี่ยและแวนเรียนซ์ของ Trinchera Creek สามารถคำนวณใหม่ โดยใช้สมการที่ 12 และ 13 ได้ดังนี้

$$\bar{Y}' = 68.87 \\ S^2(Y') = 1061.27$$

สรุป

ค่า $\bar{Y}'$ และ $S^2(Y')$ ของ Trinchera Creek ที่คำนวณได้ในขั้นที่ 6 จะสามารถนำไปใช้ในการวิเคราะห์ความถี่ของน้ำท่วมที่ให้ความเชื่อถือได้มากกว่าการวิเคราะห์โดยใช้ค่าเฉลี่ยและค่าแวนเรียนซ์ของข้อมูลเดิม

ตาราง 5 ข้อมูลปริมาณน้ำท่าประจำปีของ Trinchera Creek พายัพเพิ่มขึ้นโดยใช้สมการที่ 11

	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT
1956	3.271	19.261	14.408	6.291	11.658	13.701	30.167	39.311	51.224	36.584	40.311	11.922
1957	8.371	8.142	8.258	8.725	6.772	4.189	37.445	32.679	26.628	20.859	63.570	13.973
1958	11.872	17.281	8.921	7.313	5.374	13.669	36.244	93.133	48.354	15.377	48.229	31.148
1959	3.325	8.370	3.936	5.850	4.259	14.081	17.606	56.711	49.259	64.009	33.000	26.681
1960	8.699	6.634	13.470	4.916	3.100	7.579	18.288	35.937	21.302	7.331	11.283	1.896
	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT
1971	20.225	16.611	17.003	14.257	24.307	13.820	12.945	18.781	31.321	21.976	13.634	33.812
1972	15.412	13.069	14.272	8.363	13.293	9.679	22.521	28.918	12.111	12.135	6.045	15.249
1973	16.122	7.793	11.258	9.636	29.377	9.802	16.158	78.987	93.835	52.487	63.905	11.394
1974	7.039	7.081	6.702	3.774	11.439	11.976	17.379	46.629	26.931	22.478	8.004	5.049
1975	12.284	9.920	10.528	9.084	10.193	9.697	18.022	39.261	76.248	24.470	23.092	27.036

ตาราง 6 ข้อมูลปริมาณน้ำท่าประจำปีของ Trinchera Creek พายัพเพิ่มขึ้นโดยใช้สมการที่ 11

Year	Annual Flood	Year	Annual Flood
1956	51.22	1971	33.81
1957	63.57	1972	28.92
1958	93.13	1973	93.84
1959	64.01	1974	46.63
1960	35.94	1975	76.25

เอกสารอ้างอิง

1. Fiering, M.B., 1963, "Use of Correlation to Improve Estimates of the Mean and Variance.", Professional Paper 434-C, U.S. Geological Survey.
2. Matalas, N.C., and Jacobs, B., 1964, "A Correlation Procedure for Augmenting Hydrologic Data", Professional Paper 434-E, U.S. Geological Survey.
3. Salas, J.D. 1980, "Transfer of Information to Improve Estimates of Flood Frequencies", Hydrology for Transportation Engineers, U.S. Department of Transportation.
4. U.S Water Resources Council, 1976, "Flood Flow Frequency", Bulletin No. 17 of the Hydrology Committee, Washington, D.C.