

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเพื่อหาการแจกแจงความเข้มข้นเฉลี่ยรายวันของ PM_{10} และการประยุกต์ใช้ โดยมีวิธีการดำเนินการวิจัยตามขั้นตอนดังนี้

รวบรวมข้อมูล

งานวิจัยครั้งนี้ได้รับความอนุเคราะห์จากสำนักจัดการคุณภาพอากาศและเสียง กรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม ข้อมูลความเข้มข้นของ PM_{10} เป็นข้อมูลความเข้มข้นของ PM_{10} รายชั่วโมง ตั้งแต่เดือนมกราคม พ.ศ. 2548 ถึง เดือนธันวาคม พ.ศ. 2553 จำนวน 5 สถานี ดังนี้

- สถานีหน้าพระลาน จังหวัดสระบุรี (S1)
- สถานีกระทรวงวิทย์ ฯ (B1)
- สถานีกรมการขนส่ง (B2)
- สถานีจุฬา (B3)
- สถานีดินแดง (B4)

คัดเลือกข้อมูล

ผู้วิจัยได้ทำการแปลงข้อมูลความเข้มข้นของ PM_{10} รายชั่วโมง เป็นข้อมูลความเข้มข้นของ PM_{10} เฉลี่ยรายวัน ด้วยวิธีการหาค่าเฉลี่ยของข้อมูลความเข้มข้นรายชั่วโมงของ PM_{10} ในวันนั้น ๆ ซึ่งหากวันใดมีจำนวนข้อมูลน้อยกว่า 16 ค่า (จาก 24 ค่า) ให้ตัดข้อมูลในวันนั้นออก เนื่องจากจำนวนข้อมูลน้อยเกินไปไม่สามารถใช้เป็นตัวแทนค่าเฉลี่ยได้ เช่น

ในวันที่ 20 เมษายน พ.ศ. 2548 พบว่า มีข้อมูลความเข้มข้นรายชั่วโมงของ PM_{10} จำนวน 24 ค่า (จาก 24 ค่า) จึงสามารถนำข้อมูลของวันนี้มาหาค่าเฉลี่ยได้ เท่ากับ 83.79 ไมโครกรัมต่อลูกบาศก์เมตร ดังตารางที่ 3

ในวันที่ 21 เมษายน พ.ศ. 2548 พบว่า ความเข้มข้นรายชั่วโมงของ PM_{10} มีจำนวนข้อมูล 12 ค่า จะเห็นว่ามีจำนวนข้อมูลน้อยกว่า 16 ค่า (จาก 24 ค่า) จึงไม่สามารถนำข้อมูลของวันนี้มาหาค่าเฉลี่ยได้ ดังนั้น จึงตัดข้อมูลวันที่ 21 เมษายน พ.ศ. 2548 ออก ดังตารางที่ 4

ตารางที่ 3 ข้อมูลความเข้มข้นรายชั่วโมงของ PM₁₀ ณ สถานีหน้าพระลาน จังหวัดสระบุรี
เมื่อวันที่ 20 เมษายน พ.ศ. 2548

เวลา	ความเข้มข้นรายชั่วโมงของ PM ₁₀	เวลา	ความเข้มข้นรายชั่วโมงของ PM ₁₀
01.00 น.	20	13.00 น.	47
02.00 น.	21	14.00 น.	53
03.00 น.	33	15.00 น.	29
04.00 น.	35	16.00 น.	66
05.00 น.	31	17.00 น.	68
06.00 น.	26	18.00 น.	268
07.00 น.	25	19.00 น.	278
08.00 น.	15	20.00 น.	351
09.00 น.	12	21.00 น.	278
10.00 น.	34	22.00 น.	142
11.00 น.	55	23.00 น.	63
12.00 น.	47	24.00 น.	14

ตารางที่ 4 ข้อมูลความเข้มข้นรายชั่วโมงของ PM₁₀ ณ สถานีหน้าพระลาน จังหวัดสระบุรี
เมื่อวันที่ 21 เมษายน พ.ศ. 2548

เวลา	ความเข้มข้นรายชั่วโมงของ PM ₁₀	เวลา	ความเข้มข้นรายชั่วโมงของ PM ₁₀
01.00 น.	-	13.00 น.	16
02.00 น.	-	14.00 น.	17
03.00 น.	-	15.00 น.	14
04.00 น.	-	16.00 น.	19
05.00 น.	-	17.00 น.	14
06.00 น.	-	18.00 น.	-
07.00 น.	-	19.00 น.	169
08.00 น.	-	20.00 น.	210
09.00 น.	-	21.00 น.	184

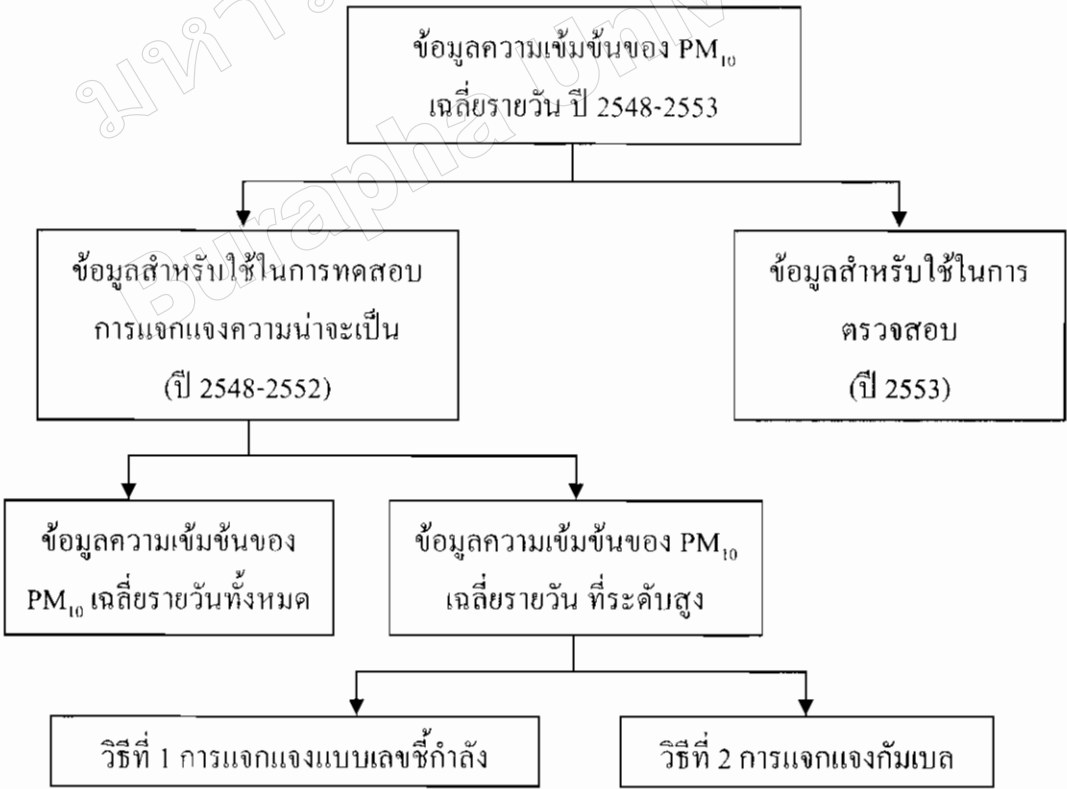
ตารางที่ 4 (ต่อ)

เวลา	ความเข้มข้นรายชั่วโมงของ PM_{10}	เวลา	ความเข้มข้นรายชั่วโมงของ PM_{10}
10.00 น.	-	22.00 น.	149
11.00 น.	-	23.00 น.	88
12.00 น.	23	24.00 น.	87

เมื่อได้ข้อมูลความเข้มข้นของ PM_{10} เฉลี่ยรายวัน แล้ว จึงแบ่งข้อมูล ออกเป็น 2 ส่วน ดังนี้

- 1. ข้อมูลในปี พ.ศ. 2548-2552 นำมาใช้เพื่อหาการแจกแจงความน่าจะเป็นทั้งในกรณีของความเข้มข้นของ PM_{10} เฉลี่ยรายวันทั้งหมด และความเข้มข้นของ PM_{10} เฉลี่ยรายวันที่ระดับสูง
- 2. ข้อมูลในปี พ.ศ. 2553 ใช้สำหรับการตรวจสอบความเหมาะสมของการแจกแจงความน่าจะเป็นของความเข้มข้นของ PM_{10} เฉลี่ยรายวันทั้งหมด และความเข้มข้นของ PM_{10} เฉลี่ยรายวันที่ระดับสูง ที่ได้จากการศึกษาในส่วนที่ 1

การคัดเลือกข้อมูลความเข้มข้นของ PM_{10} เฉลี่ยรายวัน สรุปได้ดังภาพที่ 8



ภาพที่ 8 แผนผังแสดงการคัดเลือกข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10}

วิเคราะห์ข้อมูลเบื้องต้น

การวิเคราะห์ข้อมูลเบื้องต้นเริ่มจากการสร้างกราฟการกระจายของข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ในแต่ละสถานี ตามลำดับวันที่ของการเก็บข้อมูล และนับจำนวนข้อมูลที่เกินมาตรฐาน ($120 \mu\text{m}$ ไมโครกรัมต่อลูกบาศก์เมตร) และคำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน ค่าสูงสุด และค่าต่ำสุด

ศึกษาการแจกแจงความเข้มข้นเฉลี่ยรายวันของ PM_{10}

ตอนที่ 1 ศึกษาการแจกแจงความเข้มข้นเฉลี่ยรายวันของ PM_{10} ทั้งหมด

การแจกแจงความน่าจะเป็นพื้นฐานที่เลือกมาศึกษาในงานวิจัยนี้ได้แก่ การแจกแจงล็อกนอร์มอล การแจกแจงไวบูลล์ การแจกแจงแกมมา และการแจกแจงเพียร์สัน โดยมีขั้นตอนดังนี้

1. ประเมินค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด ในการประมาณค่าพารามิเตอร์ของแต่ละการแจกแจงนั้น จะนำสมการความหนาแน่นของความน่าจะเป็นของแต่ละการแจกแจงมาใช้ในการประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด เพื่อหาตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimator: MLE) ซึ่งจะให้ค่าประมาณของพารามิเตอร์ที่มีความแปรปรวนต่ำเสมอ

การแจกแจงล็อกนอร์มอล

$$\ln \mu = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad (1)$$

$$(\ln \sigma)^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \ln \mu)^2 \quad (2)$$

การแจกแจงไวบูลล์

$$\beta = \left[\left(\sum_{i=1}^n \ln x_i^\beta \right) \left(\sum_{i=1}^n x_i^\beta \right)^{-1} - \frac{1}{n} \sum_{i=1}^n \ln x_i \right]^{-1} \quad (3)$$

$$\lambda = \left(\frac{1}{n} \sum_{i=1}^n x_i^\beta \right)^{1/\beta} \quad (4)$$

การแจกแจงแกมมา

$$\ln \theta + \psi(k) = \frac{\sum_{i=1}^n \ln x_i}{n} \quad (5)$$

$$\theta k = \frac{\sum_{i=1}^n x_i}{n} \quad (6)$$

จะได้

$$\ln \theta + \psi(k) = \frac{\sum_{i=1}^n \ln x_i}{n} - \ln \left(\frac{\sum_{i=1}^n x_i}{n} \right) \quad (7)$$

จากสมการ (7) จะใช้วิธีนิวตัน-ราฟสัน มาช่วยในการคำนวณหาค่าพารามิเตอร์ k จากนั้นจึงนำไปแทนในสมการ (7) เพื่อหาพารามิเตอร์ θ

การแจกแจงเพียร์สัน ชนิดที่ 5

$$\beta = \frac{\alpha + 1}{\alpha} \left[1 / \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right) \right] \quad (8)$$

$$\frac{\partial \ln[\Gamma(\alpha + 1)]}{\partial \alpha} = \ln(\alpha \beta) - \frac{1}{n} \sum_{i=1}^n \ln x_i \quad (9)$$

จากสมการ (9) จะใช้วิธีนิวตัน-ราฟสัน มาช่วยในการคำนวณหาค่าพารามิเตอร์ β จากนั้นจึงนำไปแทนในสมการ (9) เพื่อหาพารามิเตอร์ α

2. นำค่าประมาณของพารามิเตอร์ที่ได้แทนค่าลงในตัวแบบของการแจกแจงความน่าจะเป็น
3. นำค่าความน่าจะเป็นที่ได้จากข้อ 2 มาหาค่าประมาณของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ทั้งหมด ในแต่ละการแจกแจง โดยมีตำแหน่งเปอร์เซ็นต์ไทล์ตรงกับของข้อมูลจริง
4. ตรวจสอบความเหมาะสมของการแจกแจงของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ทั้งหมด ดังนี้

4.1 วิธีโคลโมโกรอฟ สเมอร์นอฟ

ตรวจสอบความเหมาะสมของการแจกแจงของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ทั้งหมด โดยพิจารณาจากการแจกแจงความน่าจะเป็นที่มีค่าสถิติทดสอบ D ค่าที่สุด และยอมรับสมมติฐานหลัก

4.2 วิธีแอนเดอร์สัน คาร์ริง

ตรวจสอบความเหมาะสมของการแจกแจงของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ทั้งหมด โดยพิจารณาจากการแจกแจงความน่าจะเป็นที่มีค่าสถิติทดสอบ A_N^2 ค่าที่สุด และยอมรับสมมติฐานหลัก

5. ตรวจสอบความสามารถในการพยากรณ์ความเข้มข้นเฉลี่ยรายวันของ PM_{10} ในปี 2553 ด้วยการหาค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย จากสมการ (10)

$$RMSE = \sqrt{\frac{1}{n} \left(\sum_{i=1}^n (x_i - \hat{x}_i)^2 \right)} \quad (10)$$

เมื่อ x_i คือ ความเข้มข้นเฉลี่ยรายวันของ PM_{10}
 $\hat{x}_i$ คือ ค่าพยากรณ์
 n คือ จำนวนข้อมูลในแต่ละกลุ่มตัวอย่าง

ตอนที่ 2 หาการแจกแจงของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่ระดับสูง

การแจกแจงความน่าจะเป็น โดยทั่วไป มีความเหมาะสมกับข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} แต่ไม่เหมาะสมข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่ระดับสูง เพื่อลดความผิดพลาดในการพยากรณ์ที่ระดับความเข้มข้นสูง จึงประยุกต์ใช้การแจกแจงเลขชี้กำลัง และการแจกแจง กัมเบล เพื่อพยากรณ์ความเข้มข้นของข้อมูลที่ระดับสูง นอกจากนี้ การแจกแจงเลขชี้กำลัง และการแจกแจงกัมเบลยังสามารถใช้ในการพยากรณ์รอบการเกิดซ้ำและความน่าจะเป็นที่ความเข้มข้นเฉลี่ยรายวันของ PM_{10} จะเกินมาตรฐานคุณภาพอากาศได้อีกด้วย

วิธีที่ 1 การแจกแจงเลขชี้กำลัง (Exponential Distribution)

การคัดเลือกข้อมูล

เนื่องจากกรมควบคุมมลพิษได้กำหนดมาตรฐานคุณภาพอากาศเกี่ยวกับฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมโครเมตร (PM_{10}) ในบรรยากาศต้องมีค่าไม่เกิน 120 ไมโครกรัมต่อลูกบาศก์เมตร ซึ่ง

เป็นระดับความเข้มข้นที่อาจก่อให้เกิดผลกระทบต่อสุขภาพของมนุษย์ได้ ซึ่งการคัดเลือกข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่ระดับสูง สำหรับการแจกแจงเลขชี้กำลัง จะกำหนดค่าเปอร์เซ็นต์ไทล์ที่ครอบคลุมค่ามาตรฐานคุณภาพอากาศของ PM_{10} ที่ 120 ไมโครกรัมต่อลูกบาศก์เมตร และคัดเลือกข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่มีค่าเกินเปอร์เซ็นต์ไทล์ที่กำหนดไว้ดังนี้

1. เรียงข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ในแต่ละสถานีจากความเข้มข้นน้อยไปหามาก

2. หาค่าเปอร์เซ็นต์ไทล์ที่ระดับความเข้มข้น 120 ไมโครกรัมต่อลูกบาศก์เมตร ดังนี้
สถานีหน้าพระลาน จังหวัดสระบุรี พบว่า ที่ระดับความเข้มข้น 120 ไมโครกรัมต่อลูกบาศก์เมตร ตรงกับเปอร์เซ็นต์ไทล์ที่ 70

สถานีกระทรวงวิทย์ฯ พบว่า ที่ระดับความเข้มข้น 120 ไมโครกรัมต่อลูกบาศก์เมตร ตรงกับเปอร์เซ็นต์ไทล์ที่ 88

สถานีกรมการขนส่ง พบว่า ที่ระดับความเข้มข้น 120 ไมโครกรัมต่อลูกบาศก์เมตร ตรงกับเปอร์เซ็นต์ไทล์ที่ 98

สถานีจุฬา พบว่า ที่ระดับความเข้มข้น 120 ไมโครกรัมต่อลูกบาศก์เมตร ตรงกับเปอร์เซ็นต์ไทล์ที่ 93

สถานีดินแดง พบว่า ที่ระดับความเข้มข้น 120 ไมโครกรัมต่อลูกบาศก์เมตร ตรงกับเปอร์เซ็นต์ไทล์ที่ 85

3. เพื่อให้ข้อมูลที่ได้อักรอบคลุมค่ามาตรฐานคุณภาพอากาศ (120 ไมโครกรัมต่อลูกบาศก์เมตร) จึงกำหนดค่าเปอร์เซ็นต์ไทล์ให้ต่ำจากที่คำนวณได้ในข้อ 2 อีก 10 เปอร์เซ็นต์ ดังนี้

สถานีหน้าพระลาน จังหวัดสระบุรี กำหนดให้ใช้ค่าเปอร์เซ็นต์ไทล์ที่ 60

สถานีกระทรวงวิทย์ฯ กำหนดให้ใช้ค่าเปอร์เซ็นต์ไทล์ที่ 78

สถานีกรมการขนส่ง กำหนดให้ใช้ค่าเปอร์เซ็นต์ไทล์ที่ 88

สถานีจุฬา กำหนดให้ใช้ค่าเปอร์เซ็นต์ไทล์ที่ 83

สถานีดินแดง กำหนดให้ใช้ค่าเปอร์เซ็นต์ไทล์ที่ 75

4. ข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่มีค่าสูงกว่าค่าเปอร์เซ็นต์ไทล์ที่กำหนดไว้ในแต่ละสถานี กำหนดให้เป็น x_i และนำไปใช้ในการวิเคราะห์ขั้นต่อไป โดยแต่ละสถานีมีจำนวนข้อมูล ดังนี้

สถานีหน้าพระลาน จังหวัดสระบุรี มีจำนวนข้อมูล 697 ค่า

สถานีกระทรวงวิทย์ฯ มีจำนวนข้อมูล 352 ค่า

สถานีกรมการขนส่ง	มีจำนวนข้อมูล 167 ค่า
สถานีจุฬา	มีจำนวนข้อมูล 280 ค่า
สถานีดินแดง	มีจำนวนข้อมูล 431 ค่า

5. ศึกษาการแจกแจงความน่าจะเป็นของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่ระดับสูง นำค่า x_n ที่ได้จากข้อ 4 มาหาการแจกแจงความน่าจะเป็นของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่ระดับสูง โดยมีขั้นตอน ดังนี้

สำหรับการแจกแจงเลขชี้กำลัง มีฟังก์ชันความหนาแน่นของความน่าจะเป็น คือ

$$f(x) = b_m \exp[-b_m(x_n - \phi)] \quad , \phi \leq x_n < \infty \quad (11)$$

และฟังก์ชันความน่าจะเป็นสะสม คือ

$$F(x) = 1 - \exp[-y_n] \quad (12)$$

เมื่อ

$$y_n = b_m(x_n - \phi)$$

โดย x_n

คือ ความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่มีค่าเกินกว่าเปอร์เซ็นต์ไทล์ที่

กำหนด

b_m คือ พารามิเตอร์รูปร่าง , $b_m > 0$

ϕ คือ พารามิเตอร์ตำแหน่ง , $\phi > 0$

ในการหาฟังก์ชันความน่าจะเป็นสะสม จะต้องทำการประมาณค่าพารามิเตอร์ของการแจกแจงแบบเลขชี้กำลัง คือ b_m และ ϕ ซึ่งหาได้ ดังนี้

$$b_m = \frac{\sigma_n}{S_n} \quad \text{และ} \quad \phi = \bar{x}_n - \frac{\bar{y}_n}{b_m}$$

เมื่อ

$$S_n^2 = (\overline{x_n^2} - \bar{x}_n^2)N_1 / (N_1 - 1) \quad \sigma_n^2 = (\overline{y_n^2} - \bar{y}_n^2)$$

$$\overline{x_n^2} = \frac{1}{N_1} \sum_{r=1}^n x_n^2(r)$$

$$\bar{x}_n = \frac{1}{N_1} \sum_{r=1}^n x_n(r)$$

$$\overline{y_n^2} = \frac{1}{N_1} \sum_{r=1}^n y_n^2(r) \quad \bar{y}_n = \frac{1}{N_1} \sum_{r=1}^n y_n(r)$$

โดยที่ $y_n(r)$ สามารถหาได้จากความสัมพันธ์ระหว่างตัวแปร y_n และ P_{rN_1} ดังนี้

$$y_n(r) = -\ln(1 - P_{rN_1})$$

เมื่อ N_1 คือ จำนวนข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่มีค่าเกินกว่า
เปอร์เซ็นต์ไทล์ที่กำหนด

P_{rN_1} คือ ค่าความน่าจะเป็นที่จะได้ลำดับที่ r จาก N_1 ค่า

$$P_{rN_1} = \frac{N_1 - r + 1}{N_1 + 1}$$

วิธีที่ 2 การแจกแจงกัมเบล

การคัดเลือกข้อมูล

การแจกแจงกัมเบล จะคัดเลือกข้อมูลโดยใช้ข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} สูงสุดในแต่ละเดือน (Lu & Fang, 2003) ดังนี้

1. นำข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} แบ่งออกตามปีที่เก็บข้อมูล ได้ 5 กลุ่ม แต่ละกลุ่มจะถูกแบ่งย่อยตามเดือน ได้ 12 กลุ่มย่อย ดังนั้นได้กลุ่มย่อยทั้งหมด 60 กลุ่มย่อย

2. นำข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} ในแต่ละกลุ่มย่อย มาเรียงลำดับจากมากไปน้อย ตั้งแต่ 1 ถึง n ดังนี้ $x_1 > x_2 > \dots > x_n$ และให้ค่าความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่มีค่าสูงสุดในแต่ละกลุ่มย่อยคือ x_m

3. ได้ข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} สูงสุดในแต่ละเดือน (x_m) จำนวน 60 ค่า เพื่อใช้ในการวิเคราะห์ขั้นต่อไป

4. ศึกษาการแจกแจงความน่าจะเป็นของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ที่ระดับสูง สำหรับการแจกแจงกัมเบลมีฟังก์ชันความหนาแน่นของความน่าจะเป็น คือ

$$f_{mn}(x_m) = \frac{1}{\sigma} \exp[-y_m - \exp(-y_m)] \quad (13)$$

และฟังก์ชันความน่าจะเป็นสะสม คือ

$$G_{mn}(x_m) = \exp[-\exp(-y_m)] \tag{14}$$

เมื่อ

$$y_m(x_m) = a_m(x_m - u_m)$$

โดย a_m คือ พารามิเตอร์เร่งขนาด , $a_m > 0$

u_m คือ พารามิเตอร์บ่งตำแหน่ง

ในการหาฟังก์ชันความน่าจะเป็นสะสม จะต้องทำการประมาณค่าพารามิเตอร์ของการแจกแจงเลขชี้กำลัง คือ a_m และ u_m ซึ่งหาได้ดังนี้

$$\hat{a}_m = \frac{\sigma_m}{S_m} \text{ และ } \hat{u}_m = \bar{x}_m - \frac{\bar{y}_m}{\hat{a}_m}$$

เมื่อ

$$S_m^2 = (\overline{x_m^2} - \bar{x}_m^2)N / (N - 1) \qquad \sigma_m^2 = (\overline{y_m^2} - \bar{y}_m^2)$$

$$\overline{x_m^2} = \frac{1}{N} \sum_{r=1}^n x_m^2(r) \qquad \bar{x}_m = \frac{1}{N} \sum_{r=1}^n x_m(r)$$

$$\overline{y_m^2} = \frac{1}{N} \sum_{r=1}^n y_m^2(r) \qquad \bar{y}_m = \frac{1}{N} \sum_{r=1}^n y_m(r)$$

โดยที่ $y_m(r)$ สามารถหาได้จากความสัมพันธ์ระหว่างตัวแปร y_n และ P_{rN} ดังนี้

$$y_m(r) = -\ln(-\ln P_{rN})$$

เมื่อ N คือ จำนวนข้อมูลความเข้มข้นเฉลี่ยรายวันของ PM_{10} สูงสุดในแต่ละเดือน
 P_{rN} คือ ค่าความน่าจะเป็นที่จะได้ลำดับที่ r จาก N ค่า

$$P_{rN} = \frac{N - r + 1}{N + 1}$$

ศึกษารอบการเกิดซ้ำ

สำหรับรอบการเกิดซ้ำ $R(x_c)$ เป็นระยะเวลาเฉลี่ยที่ความเข้มข้นเฉลี่ยรายวันของ PM_{10} จะเกินมาตรฐาน. x_c , สามารถคำนวณได้ ดังนี้
กรณีการแจกแจงแบบเลขชี้กำลัง

$$R(x_c) = \frac{1}{(1-f)(1-F_r(x_c))} \tag{15}$$

เมื่อ $F_r(x_c)$ คือ ความน่าจะเป็นสะสมที่ความเข้มข้นเฉลี่ยรายวันของ PM_{10} มีค่า 120 ไมโครกรัมต่อลูกบาศก์เมตร
 $R(x_c)$ คือ ระยะเวลาที่เหตุการณ์นั้นจะกลับมาเกิดซ้ำอีก
 x_c คือ ค่ามาตรฐานคุณภาพอากาศของ PM_{10} กำหนดเป็น 120 ไมโครกรัมต่อลูกบาศก์เมตร
 f คือ ค่าเปอร์เซ็นต์ไทล์ที่กำหนด
กรณีการแจกแจงกัมเบล

$$R(x_c) = \frac{1}{1-G_{mm}(x_c)} \tag{16}$$

เมื่อ $G_{mm}(x_c)$ คือ ความน่าจะเป็นสะสมของค่ามาตรฐานคุณภาพอากาศของ ความเข้มข้นเฉลี่ยรายวันของ PM_{10} (ที่ 120 ไมโครกรัมต่อลูกบาศก์เมตร)
 $R(x_c)$ คือ ระยะเวลาที่เหตุการณ์นั้นจะกลับมาเกิดซ้ำอีก
 x_c คือ ค่ามาตรฐานคุณภาพอากาศของ PM_{10} (120 ไมโครกรัมต่อลูกบาศก์เมตร)

การประมาณค่าการลดมลพิษจากแหล่งกำเนิด

การประมาณค่าการลดมลพิษจากแหล่งกำเนิด ใช้เป็นข้อมูลอ้างอิงในการวางมาตรการในการลดการปล่อย PM_{10} ออกสู่บรรยากาศในปีถัดไป เพื่อควบคุมไม่ให้ระดับความเข้มข้นของ PM_{10} มีค่าเกินมาตรฐานคุณภาพอากาศ อย่างไรก็ตาม การประมาณค่าการลดมลพิษจากแหล่งกำเนิด

จะมีประสิทธิภาพเมื่อ พื้นที่ที่ทำการศึกษายู่ภายใต้สภาวะที่กำหนดคือ การกระจายของมลพิษจากแหล่งกำเนิด และสภาพอากาศไม่มีการเปลี่ยนแปลง

เราสามารถประมาณค่าการลดมลพิษจากแหล่งกำเนิด เพื่อตอบสนองมาตรฐานคุณภาพอากาศ ได้ดังนี้

กรณีที่ 1 การแจกแจงเลขชี้กำลัง

$$ESR = \frac{x_c - \bar{x}}{x_c} \times 100 \pm S_{x_c} \quad (17)$$

- เมื่อ x_c คือ ค่ามาตรฐานคุณภาพอากาศของ PM_{10} (120 ไมโครกรัมต่อลูกบาศก์เมตร)
 $\bar{x}$ คือ ค่าเฉลี่ยของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ในแต่ละสถานี
 S_{x_c} คือ ค่าความคลาดเคลื่อนของการพยากรณ์

โดยค่าความคลาดเคลื่อนของการพยากรณ์สามารถหาได้ดังนี้

$$S_{x_c} = S_y \sqrt{1 + \frac{1}{n} + \frac{(x_c - \bar{x})^2}{(n-1)S_x^2}} \quad (18)$$

- เมื่อ x_c คือ ค่ามาตรฐานคุณภาพอากาศของ PM_{10} (120 ไมโครกรัมต่อลูกบาศก์เมตร)
 $\bar{x}$ คือ ค่าเฉลี่ยของความเข้มข้นเฉลี่ยรายวันในแต่ละสถานี
 S_y คือ ค่าความคลาดเคลื่อนของ y
 S_x คือ ค่าความคลาดเคลื่อนของ x

กรณีที่ 2 การแจกแจงกัมเบล

$$ESR = \frac{x_c - \bar{x}}{x_c} \times 100 \pm S_{x_c} \quad (19)$$

- เมื่อ x_c คือ ค่ามาตรฐานคุณภาพอากาศของ PM_{10} (120 ไมโครกรัมต่อลูกบาศก์เมตร)
 $\bar{x}$ คือ ค่าเฉลี่ยของความเข้มข้นเฉลี่ยรายวันของ PM_{10} ในแต่ละสถานี
 S_{x_c} คือ ค่าความคลาดเคลื่อนของการพยากรณ์

โดยค่าความคลาดเคลื่อนของการพยากรณ์สามารถหาได้ดังนี้

$$S_c = \left[\frac{1}{N} (1 + 1.1396K + 1.1K^2) \right]^{1/2} s \quad (20)$$

เมื่อ s คือ เบี่ยงเบนมาตรฐานของตัวอย่างขนาด N ($N = 60$)

$$K \text{ คือ ความถี่ของปัจจัย โดย } K = -\frac{\sqrt{6}}{\pi} \left\{ 0.5772 + \ln \left(\ln \left(\frac{R(x_i)}{R(x_i) - 1} \right) \right) \right\}$$

$R(x_i)$ คือ รอบการเกิดซ้ำ