การประมาณค่าข้อมูลน้ำฝนรายวันที่ขาดหายไปด้วยวิธีควอนไทล์
Gap Filling Daily Rainfall Data using Quantile Method

ศรีสุนี วุฒิวงศ์โยธิน* กุลสตรี ศรีจุมปา และคมกฤษณ์ โสภา

ภาควิชาวิศวกรรมโยธา คณะวิศวกรรมศาสตร์ มหาวิทยาลัยบูรพา จ.ชลบุรี
**Corresponding author; E-mail address: srisunee.wu@eng.buu.ac.th*

บทคัดย่อ

การประมาณค่าข้อมูลน้ำฝนรายวันที่ขาดหายไปมีความสำคัญเพื่อเติมชุดข้อมูลให้มีความสมบูรณ์ก่อนนำไปใช้วิเคราะห์ผลในการศึกษาด้านต่างๆ ที่เกี่ยวข้องกับน้ำ วิธีที่นิยมใช้ในการเติมค่าข้อมูลน้ำฝนรายวันที่ขาดหายไป ได้แก่ วิธีค่าเฉลี่ยคณิตศาสตร์ วิธีระยะทางผกผัน (Inverse Distance Weighting method, IDW) วิธีการเหล่านี้มักมีข้อจำกัดหลัก ได้แก่ มักให้ค่าปริมาณฝนรายวันที่ต่ำกว่าความเป็นจริง จำนวนวันฝนตกที่มากเกินไป ดังนั้นในงานวิจัยนี้ทำการศึกษาวิธีการเติมค่าสูญหายข้อมูลฝนด้วยวิธีการทางสถิติ คือ วิธีควอนไทล์ (Quantile method, QT) โดยการสร้างกราฟการแจกแจงความถี่แบบเบอร์นูลี่-แกมมา (Bernoulli-Gamma Distribution) จากข้อมูลฝนรายวันของสถานีเป้าหมายที่จะเติมค่าและสถานีอ้างอิงที่คัดเลือก โดยใช้สถานีตรวจวัดฝนในพื้นที่ลุ่มน้ำปิงตอนบน จำนวน 6 สถานี ทำการทดสอบการเติมค่าสูญหายที่เปอร์เซ็นต์สูญหาย 20% และ 40% ผลการศึกษา พบว่า วิธี QT ให้ค่าทางสถิติ ได้แก่ ค่าฝนรายวันมากที่สุด ฝนรายวันเฉลี่ย ความแปรปรวน และค่าเปอร์เซ็นต์ไทล์ที่ 99 ดีกว่าวิธี IDW แต่ให้ค่า RMSE และ MAE มากกว่าวิธี IDW เล็กน้อย เนื่องมาจาก QT มีค่าความแปรปรวนมากกว่า

คำสำคัญ: การเติมค่า ข้อมูลที่ขาดหายไป ฝนรายวัน วิธีควอนไทล์ วิธีระยะทางผกผัน

Abstract

Filling the gap of daily rainfall data is an important step to obtain a complete data set before further study that related to water resources. Generally, methods that use to fill the gap are arithmetic mean and inverse distant weighting (IDW). Some limitations of these traditional methods are underestimation of daily rainfall and overestimation of the number of rainy days, for example. Thus, this research attempts to study a gap filling daily rainfall data base upon statistical approach, quantile method (QT). By this method, a mixed probability distribution of Bernoulli-Gamma function is used to derive the target and source stations daily rainfall distributions of six selected rain stations located in upper Ping River basin. This study has tested such method to fill the missing daily rainfall at 20% and 40% percent missing. The study result reveals that QT generally yields a statistical value such as mean, maximum, variance, and 99% percentile better than IDW, but QT yields greater RMSE and MAE because QT produce higher in variance.

Keywords: Gap Filling, Missing Data, Daily Rainfall, Quantile Method, Inverse Distant Weighting Method

1. คำนำ

ข้อมูลฝนเป็นข้อมูลพื้นฐาน ที่เกี่ยวข้องกับการดำรงชีวิตของมนุษย์โดยตรงในกิจกรรมที่เกี่ยวข้องกับน้ำ ดังนั้น หากมีข้อมูลฝนที่เป็นชุดข้อมูลที่สมบูรณ์ จะทำให้การบริหารจัดการน้ำมีความแม่นยำ ลดความเสี่ยง เพิ่มความปลอดภัย ต่อชีวิตและทรัพย์สิน และมีประโยชน์ต่อการศึกษา เช่น การศึกษาการพัฒนาแหล่งน้ำ การศึกษาเพื่อหาแนวทางในการป้องกันและบรรเทาภัยพิบัติที่เกี่ยวกับน้ำ การออกแบบโครงสร้างชลศาสตร์ เช่น เขื่อน สะพาน ให้ได้ขนาดที่เหมาะสม ปลอดภัย เป็นต้น ในทางปฏิบัติการตรวจวัดข้อมูลด้วยเครื่องมือหรือวิธีต่างๆ เราไม่สามารถที่จะป้องกันการสูญหายของข้อมูลที่จะเกิดขึ้นได้ [1] การมีข้อมูลฝนขาดหายไป (missing data) อาจมีสาเหตุเนื่องมาจากเครื่องตรวจวัดน้ำฝนชำรุด ปัญหาจากเจ้าหน้าที่ไม่ได้ทำการบันทึกข้อมูล สมุดบันทึกหรือข้อมูลที่บันทึกสูญหาย จากเหตุสุดวิสัยภัยธรรมชาติ เช่น เกิดจากพายุฝนตกหนัก ทำให้เครื่องมือทำงานผิดปกติบางช่วงเวลา [2] ดังนั้นการประมาณค่าข้อมูลฝนรายวันที่ขาดหายไป จึงเป็นขั้นตอนสำคัญและจำเป็นเพื่อให้ได้ข้อมูลที่มีความสมบูรณ์และถูกต้องก่อนนำไปใช้ในการศึกษาและวิเคราะห์ในการศึกษาด้านทรัพยากรน้ำ การบริหารจัดการน้ำ สิ่งแวดล้อมและการเกษตร

ทั้งนี้การศึกษาวิธีการประมาณค่าข้อมูลฝนรายวันที่ขาดหายไป มีการศึกษามาหลายทศวรรษ เช่น วิธีสัดส่วนปกติ พัฒนาโดย Paulhus และ Kohler ปี ค.ศ. 1952 วิธีรีเกรชชั่น พัฒนาโดย Makhuva และคณะเมื่อปี ค.ศ.1997 วิธีทางสโตแคสติก พัฒนาโดย Zuccini และคณะ ปี ค.ศ.1992 วิธีระยะทางผกผัน (Inverse Distance Weighting Method, IDW) เริ่มคิดค้นโดย Shepard ในปี ค.ศ. 1968 มีการพัฒนาเรื่อยมาและประยุกต์ใช้มากมายในด้านวิทยาศาสตร์สาขาต่างๆ รวมถึงด้านอุทกวิทยา นอกจากนี้ยังมีวิธีการอื่นๆ อีกมากมาย ทั้งวิธีการแบบง่ายและค่อนข้างซับซ้อน วิธีทางสถิติ หรือการปรับปรุงวิธีดั้งเดิม (เช่น วิธีระยะทางผกผัน) แต่ก็ยังไม่มีข้อสรุปในการศึกษาว่าวิธีการใดดีที่สุด

สำหรับประเทศไทย วิธีการประมาณค่าข้อมูลฝนรายวันที่ขาดหายไปนิยมใช้วิธีอย่างง่าย ได้แก่ วิธีค่าเฉลี่ยคณิตศาสตร์ วิธีสัดส่วนปกติ และวิธีระยะทางผกผัน ซึ่งมักใช้ข้อมูลจากสถานีข้างเคียงจำนวนตั้งแต่ 2-4 มาใช้ในการประมาณค่า แต่เนื่องจากวิธีการดังกล่าวมักเป็นวิธีที่ได้จากการเฉลี่ยค่าจากสถานีข้างเคียง จึงมีข้อจำกัดหลัก ได้แก่ ให้ค่าฝนเฉลี่ยที่น้อยกว่าค่าจริง และให้ค่าจำนวนวันฝนตกที่มากกว่าความจริง เป็นต้น ซึ่งสาเหตุเกิดจากการเฉลี่ยค่าจากสถานีข้างเคียงนั้นเอง ทำให้ไม่สามารถประมาณค่าที่ใกล้เคียงโดยเฉพาะอย่างยิ่งกรณีเหตุการณ์ฝนสุดขีด (extreme storm) ได้ [3] จากข้อจำกัดดังกล่าวในกรณีที่ต้องการข้อมูลฝนรายวันสำหรับใช้ในการศึกษาวิเคราะห์เหตุการณ์วิกฤติต่างๆ เช่น การวิเคราะห์เหตุการณ์ฝนสูงสุด การ

คาดการณ์การเกิดภัยพิบัติ การเกิดน้ำป่าไหลหลากและดินโคลนถล่ม จำเป็นต้องใช้ข้อมูลที่มีความถูกต้อง ซึ่งในการประมาณค่าข้อมูลฝนรายวันที่ขาดหายไปด้วยวิธีการที่มีแนวโน้มที่ให้ค่าเฉลี่ยฝนรายวันน้อยกว่าความเป็นจริง อาจส่งผลให้ผลการศึกษาดังกล่าวมีความคลาดเคลื่อนได้

การวิจัยครั้งนี้จึงเสนอวิธีในการประมาณค่าข้อมูลฝนรายวันที่ขาดหายไปด้วยวิธีทางสถิติ ได้แก่ วิธีควอนไทล์ (Quantile Method, QT) เพื่อหาวิธีที่สามารถแก้ปัญหาข้อจำกัดกรณีให้ค่าเฉลี่ยฝนรายวันที่น้อยกว่าความจริง และทำการเปรียบเทียบกับวิธีระยะทางผกผัน (Inverse Distance Weighting Method, IDW) ซึ่งเป็นวิธีที่นิยมใช้มากที่สุดวิธีหนึ่ง โดยทำการทดลองกรณีมีค่าข้อมูลน้ำฝนรายวันขาดหายไปที่เปอร์เซ็นต์การสูญหาย 20% และ 40%

2. ข้อมูลที่ใช้ในการศึกษา

ข้อมูลที่ใช้ศึกษา ได้แก่ ข้อมูลฝนรายวันจากสถานีตรวจวัดภาคพื้นดินที่ตั้งอยู่ในลุ่มน้ำปิงตอนบน โดยคัดเลือกมาศึกษาเพื่อทดสอบวิธีการประมาณค่าฝนรายวันที่ขาดหายไปจำนวน 6 สถานี จากกลุ่มพื้นที่ตอนบน ตอนกลาง และตอนล่างของลุ่มน้ำ พื้นที่ละ 2 สถานี โดยคัดเลือกจากสถานีที่มีจำนวนข้อมูลสูญหายน้อยที่สุด (หรือมีจำนวนข้อมูลมากที่สุด) ช่วงเวลาข้อมูลที่ศึกษา ตั้งแต่ ม.ค. ปี ค.ศ. 1953 ถึง ธ.ค. ปี ค.ศ 2017 (65ปี) รายละเอียดข้อมูล 6 สถานีเป้าหมายที่ใช้ในการทดสอบแสดงดังตารางที่ 1

ตารางที่ 1 สถานีเป้าหมาย 6 สถานีที่คัดเลือกเพื่อใช้ในการเติมค่าสูญหาย

ลำดับ	รหัสสถานี	ชื่อสถานี	พิกัดแบบ UTM		ฝนรายปีเฉลี่ย (มม.)
			X	Y	
1	327009	สะเมิง จ.เชียงใหม่	471,837	2,084,142	1,160.31
2	327501	เมือง จ.เชียงใหม่	497,085	2,083,226	1,148.28
3	327003	จอมทอง จ.เชียงใหม่	465,843	2,036,358	907.62
4	329003	ป่าซาง จ.ลำพูน	493,735	2,048,255	954.94
5	327008	อมก๋อย จ.เชียงใหม่	431,844	1,9678,54	875.58
6	329006	โรงเรียนบ้านก้อ จัดสรร อ.ลี้ จ.ลำพูน	475,804	1,952,232	931.4

3. วิธีและขั้นตอนในการศึกษา

3.1 ขั้นตอนในการศึกษา

จากสถานีที่คัดเลือกจำนวน 6 สถานี ต่อไปนี้เรียกว่า “สถานีเป้าหมาย” (Target station) และสถานีข้างเคียงที่เลือก เรียกว่า “สถานีอ้างอิง” (Source station) ซึ่งขั้นตอนในการศึกษาการเติมค่าสรุปดังนี้

1) ตัดทิ้งวันที่มีข้อมูลขาดหายไปจริงๆ ทั้งหมดของสถานีเป้าหมายทั้ง 6 สถานี คงเหลือเฉพาะจำนวนวันที่มีข้อมูลตรวจวัดฝนรายวัน ในช่วง ม.ค. ปี ค.ศ. 1953 ถึง ธ.ค. ปี ค.ศ 2017 (65ปี) ซึ่งแต่ละสถานีมีจำนวนไม่เท่ากัน

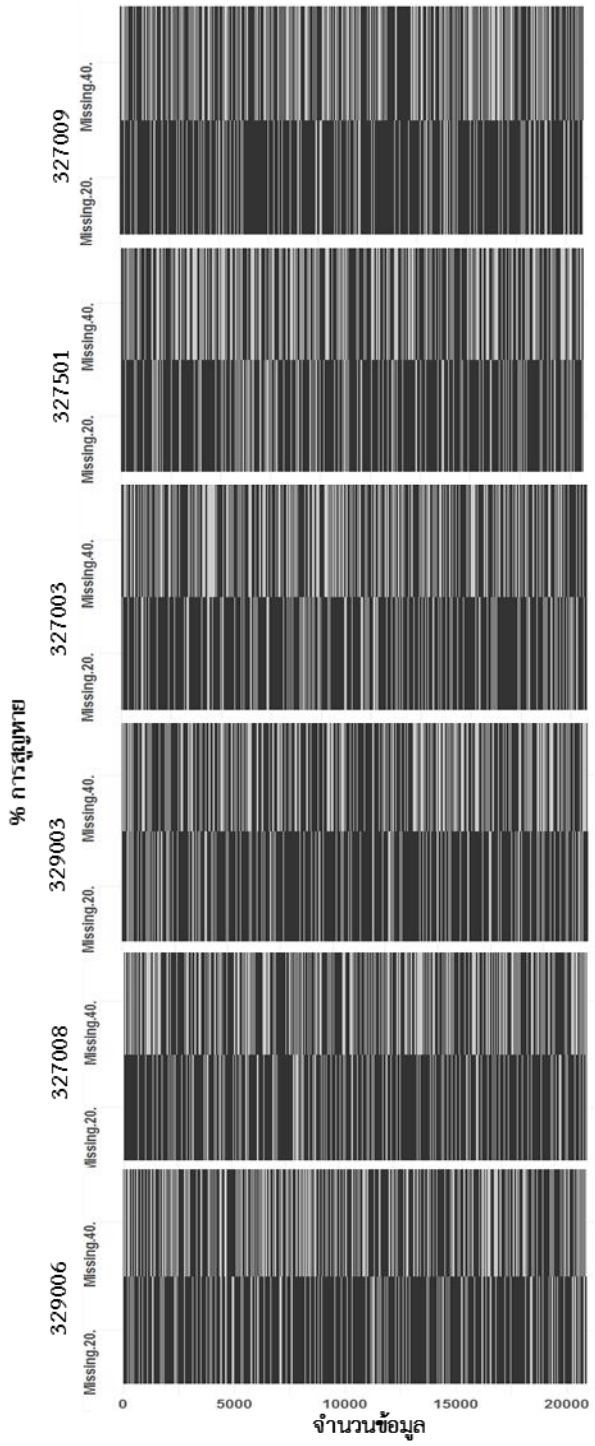
2) สร้างค่าข้อมูลฝนรายวันให้สูญหายแบบสุ่ม (Missing at Random, MAR) ให้กับสถานีเป้าหมายทั้ง 6 สถานี โดยทำให้มีเปอร์เซ็นต์การสูญหาย 20% และ 40% เพื่อใช้ทดสอบในการศึกษานี้ จำนวนข้อมูลและค่าสูญหายที่เปอร์เซ็นต์ต่างๆ ของแต่ละสถานี แสดงดังตารางที่ 2 และลักษณการสูญหายแบบสุ่มข้อมูลของแต่ละสถานีแสดงดังรูปที่ 1

3) ประมาณค่าข้อมูลฝนรายวันที่ขาดหายไปและเติมค่าฝนรายวันที่ขาดหายไปด้วยวิธีระยะทางผกผัน (IDW) และวิธีควอนไทล์ (QT)

4) ประเมินประสิทธิภาพ (Performance and Accuracy Assessment) และเปรียบเทียบระหว่างวิธี IDW และวิธี QT

ตารางที่ 2 จำนวนข้อมูลค่าสูญหายที่เปอร์เซ็นต์การสูญหายต่างๆ

สถานีเป้าหมาย	จำนวนข้อมูลทั้งหมด	จำนวนข้อมูลที่แต่ละเปอร์เซ็นต์การสูญหาย			
		20%		40%	
		ค่าสูญหาย	ข้อมูลที่มีค่า	ค่าสูญหาย	ข้อมูลที่มีค่า
พื้นที่ตอนบน					
1) 327009	23,406	**4,681**	18,725	**9,361**	14,045
2) 327501	23,186	**4,637**	18,549	**9,274**	13,912
พื้นที่ตอนกลาง					
3) 327003	23,314	**4,662**	18,652	**9,325**	13,989
4) 329003	22,135	**4,426**	17,709	**8,853**	13,282
พื้นที่ตอนล่าง					
5) 327008	21,513	**4,302**	17,211	**8,604**	12,909
6) 329006	20,881	**4,176**	16,705	**8,352**	12,529



รูปที่ 1 ลักษณะการสูญหายของค่าข้อมูลฝนรายวันแบบสุ่มของสถานีเป้าหมาย
หมายเหตุ: แถบสีดำ = ข้อมูลที่มีค่า และแถบสีเทา = ค่าสูญหาย

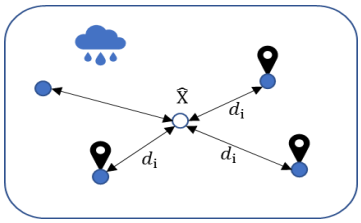
รายละเอียดวิธีการประมาณค่าและเติมค่าข้อมูลฝนรายวันที่ขาดหายไป และการประเมินประสิทธิภาพ กล่าวในหัวข้อดังต่อไปนี้

3.2 วิธีระยะทางผกผัน (Inverse distance Weighting Method, IDW)

วิธีระยะทางผกผัน (IDW) เป็นวิธีการประมาณค่าในช่วงเชิงพื้นที่ (spatial interpolation method) โดยใช้ค่าจากสถานีที่ทราบค่าข้อมูลที่อยู่ใกล้เคียงมาประมาณค่า ดังสมการที่ (1) และรูปที่ 2

$$\hat{X}=\frac{\Sigma_{i=1}^{m}\left(\frac{x_i}{d_i^2}\right)}{\Sigma_{i=1}^{m}\left(\frac{1}{d_i^2}\right)} \quad (1)$$

เมื่อ $\hat{X}$ = ปริมาณฝนที่ประมาณค่าได้
d_i = ระยะระหว่างสถานีเป้าหมายกับสถานีข้างเคียง (กม.)
x_i = ปริมาณน้ำฝนจากสถานีตรวจวัดฝนข้างเคียงที่ทราบค่า (มม.)



รูปที่ 2 วิธีระยะทางผกผัน

วิธีระยะทางผกผัน (IDW) ให้ความสำคัญกับสถานีที่อยู่ใกล้มากกว่าสถานีที่อยู่ห่างไกลออกไปด้วยการใช้ค่าถ่วงน้ำหนัก ซึ่งได้แก่ ส่วนกลับของระยะทางระหว่างสถานีที่เป้าหมายและสถานีข้างเคียงที่พิจารณายกกำลัง 2 ถ้าระยะทางเพิ่มขึ้นค่าถ่วงน้ำหนักจะลดลงอย่างรวดเร็ว ในการประมาณค่าข้อมูลฝนรายวันที่ขาดหายไปในการศึกษานี้กำหนดเลือกใช้ค่าสถานีอ้างอิง 3 สถานี มาคำนวณเพื่อประมาณค่าข้อมูลน้ำฝนให้กับสถานีเป้าหมาย

3.3 วิธีควอนไทล์ (Quantile Method, QM)

3.3.1 การแจกแจงความถี่แบบผสมเบอร์นูลี่-แกมม่า

วิธีควอนไทล์ เป็นวิธีทางสถิติที่ใช้ฟังก์ชันการแจกแจงความถี่ (Probability Distribution Function, PDF) หรือฟังก์ชันการแจกแจงความถี่แบบสะสม (Cumulative Distribution Function, CDF) สร้างจากข้อมูลฝนรายวันที่มีค่า โดยสันนิษฐานการแจกแจงข้อมูลน้ำฝนรายวันที่ใช้ในการศึกษานี้ใช้การแจกแจงความถี่แบบผสม ได้แก่ การแจกแจงเบอร์นูลี่-แกมม่า (Bernoulli-Gamma Distribution) โดยที่การแจกแจงแบบเบอร์นูลี่ ใช้สำหรับคำนวณหาความน่าจะเป็นของวันฝนไม่ตก ($1-\pi$) และความน่าจะเป็นของวันฝนตก (π) ส่วนการแจกแจงแบบแกมม่าใช้สำหรับวิเคราะห์รูปแบบการแจกแจงความถี่ของฝนรายวันซึ่งเป็นที่นิยมใช้ในด้านอุทกวิทยาและสำหรับข้อมูลฝน [3-5]

ฟังก์ชันการแจกแจงความถี่ (PDF) เบอร์นูลี่-แกมม่า แสดงดังนี้

$$g(x)=\begin{cases}\pi\gamma(x) & ; \text{ถ้า } x>0.1\\ 1-\pi & ; \text{ถ้า } x\leq 0.1\end{cases} \quad (2)$$

โดยที่ π คือ ความน่าจะเป็นของวันฝนตก
$1-\pi$ คือ ความน่าจะเป็นของวันฝนไม่ตก
$\gamma(x)$ คือ การแจกแจงความถี่แบบแกมม่า

ซึ่ง $$\gamma(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{\left(-\frac{x}{\beta}\right)} \quad (3)$$

และ $$\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}dx \quad \text{เมื่อ } \alpha>0 \quad (4)$$

หรือ $\Gamma(\alpha)=(\alpha-1)!$

เมื่อ α คือ พารามิเตอร์แรงรูปร่าง (shape parameter)
และ β คือ พารามิเตอร์มาตราส่วน (scale parameter)

สมการการแจกแจงความถี่แบบสะสม (Cumulative Distribution Function, CDF) แบบเบอร์นูลี่-แกมม่า คือ

$$G(x)=\begin{cases}1-\pi+\pi*\Gamma(x) & ; \text{ถ้า } x>0.1\\ 1-\pi & ; \text{ถ้า } x\le 0.1\end{cases} \quad (5)$$

เมื่อ $\Gamma(x)$ คือ ฟังก์ชันแกมม่าของฝนรายวัน

x คือ ปริมาณฝนรายวัน กำหนดให้ฝนตกมากกว่า 0.1 มม. เป็นวันที่มีฝนตก

ฟังก์ชันควอนไทล์ หรือส่วนกลับของ CDF (Inverse CDF) เขียนเป็นสมการดังนี้

$$G^{-1}(x)=\begin{cases}\Gamma^{-1}\left(\frac{p-1+\pi}{\pi}\right) & ; \text{ถ้า } \pi>1-p\\ 0 & ; \text{ถ้า } p\le 1-p\end{cases} \quad (6)$$

เมื่อ $\Gamma^{-1}(p)$ คือ ส่วนกลับของ *CDF* แบบแกมม่า

p คือ ค่าความน่าจะเป็น

3.3.2 ขั้นตอนการเติมค่าด้วยวิธีควอนไทล์

1) กำหนดสถานีที่ต้องการเติมค่าข้อมูลเป็นสถานีเป้าหมาย (Target Station, TS) และคัดเลือกสถานีอ้างอิง (Source Station, SS) จากสถานีที่มีค่าความสัมพันธ์แบบเพียร์สัน(r) ที่ดีที่สุดระหว่างข้อมูลฝนรายของ TS กับ SS ทั้งนี้ถ้าในวันที่สถานีเป้าหมายมีค่าสูญหาย และสถานีอ้างอิงลำดับแรก SS1 มีค่าสูญหายเช่นกัน จะทำการเลือกอ้างอิงค่าข้อมูลฝนจากสถานีอ้างอิงที่มีค่า r ในลำดับถัดไป สถานีอ้างอิงและค่าความสัมพันธ์เพียร์สันที่ใช้สำหรับแต่ละสถานีเป้าหมายในการศึกษานี้แสดงดังตารางที่ 3

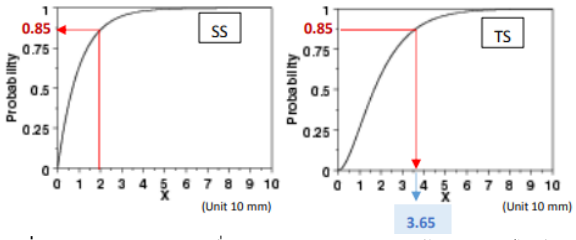
2) คำนวณหาค่าพารามิเตอร์ของฟังค์ชันการแจกแจงแบบเบอร์นูลี่-แกมม่า และสร้างกราฟแจกแจงความถี่แบบสะสมจากชุดข้อมูลที่มีค่าของสถานีเป้าหมาย (CDF-TS) และสร้างกราฟแจกแจงความถี่แบบสะสมของสถานีอ้างอิง (CDF-SS)

ตารางที่ 3 สถานีอ้างอิงที่คัดเลือกสำหรับวิธีควอนไทล์

สถานีเป้าหมาย	(รหัส) สถานีอ้างอิงที่คัดเลือก				
	SS1	SS2	SS3	SS4	SS5
327009	327021	070391	327020	070550	327501
r	0.82	0.81	0.8	0.78	0.77
327501	070391	327021	327020	327004	-
r	0.94	0.92	0.91	0.88	-
327003	327027	329005	329003	327010	-
r	0.85	0.85	0.8	0.76	-
329003	327024	329001	329002	329005	327003
r	0.85	0.82	0.8	0.8	0.8
327008	327022	329006	376001	-	-
r	0.78	0.78	0.7	-	-
329006	329007	376009	327022	329004	327008
r	0.81	0.81	0.8	0.8	0.78

3) ประมาณค่าฝนรายวันที่ขาดหายไปของสถานีเป้าหมาย โดยการอ่านค่าฝนรายวันจากสถานีอ้างอิง นำค่าฝนรายวันของสถานีอ้างอิงไปอ่านค่าความน่าจะเป็นจากกราฟ CDF-SS และนำค่าความน่าจะเป็นหรือโอกาสในการเกิดฝนตกของสถานีอ้างอิง มาอ่านค่าปริมาณฝนย้อนกลับจากกราฟ CDF-TS ก็จะได้ค่าฝนสำหรับเติมในวันที่ข้อมูลขาดหายของสถานีเป้าหมาย ดังแสดงดังรูปที่ 3

ตัวอย่าง สมมุติให้สถานีเป้าหมายมีค่าสูญหายในวันที่ 30 เม.ย. 2000 ทำการอ่านค่าฝนจากสถานีอ้างอิง (SS) ในวันเดียวกันได้ 20 มม. เมื่อนำไปอ่านค่าความน่าจะเป็นจากกราฟ CDF-SS จะได้ 0.85 (85%) จากนั้นนำค่าความน่าจะเป็น 0.85 ไปอ่านค่าปริมาณน้ำฝนจากกราฟ CDF-TS จะได้ค่าน้ำฝนรายวัน 36.5 มม.และนำค่านี้ไปเติมแทนค่าสูญหายให้กับสถานีเป้าหมายในวันที่ 30 เม.ย. ค.ศ. 2000



รูปที่ 3 ตัวอย่างการอ่านค่าเพื่อเติมค่าสูญหายฝนรายวันด้วยวิธีควอนไทล์

3.4 การประเมินประสิทธิภาพ (Performance Assessment)

การประเมินประสิทธิภาพกระทำโดย

1) การเปรียบเทียบค่าสถิติพื้นฐานระหว่างข้อมูลที่ประมาณค่าได้กับข้อมูลตรวจวัดจริง ได้แก่ ค่าฝนรายวันสูงสุด ค่าเฉลี่ยฝนรายวัน ความแปรปรวน และค่าฝนที่เปอร์เซ็นต์ไทล์ที่ 99

2) การวัดค่าความคลาดเคลื่อนกระทำโดยวิธีค่ารากที่สองความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error, RMSE) และค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) ดังสมการ

$$RMSE=\sqrt{\sum_{i=1}^{n}\frac{(x_{obs,i}-\hat{x}_{cal,i})^2}{n}} \quad (7)$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|x_{obs,i}-\hat{x}_{cal,i}\right| \quad (8)$$

เมื่อ $x_{obs,i}$ = ค่าฝนรายวันจากการตรวจวัดจริง

$\hat{x}_{cal,i}$ = ค่าฝนรายวันที่ได้จากการคำนวณเพื่อเติมค่า

n = จำนวนข้อมูลทั้งหมดที่ประมาณค่า

ถ้าค่าความคลาดเคลื่อน RMSE และ MAE น้อยแสดงว่าค่าที่คำนวณและเติมมีความใกล้เคียงค่าจริงมาก

4. ผลการทดสอบการเติมค่าข้อมูลฝนรายวันที่ขาดหายไป

4.1 การประเมินค่าประสิทธิภาพ: ค่าสถิติพื้นฐาน

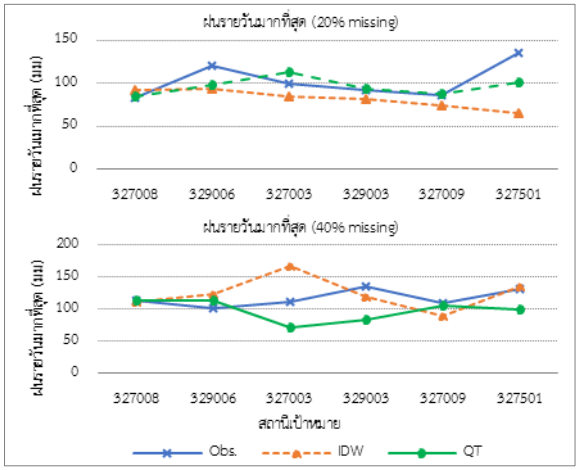
1) ฝนรายวันมากที่สุด: ได้จากการคำนวณหาค่าฝนรายวันที่มากที่สุดจากข้อมูลที่ทำการประมาณค่าทั้งหมด เปรียบเทียบกับค่าตรวจวัดจริง ดังตารางที่ 4 และรูปที่ 4 จากตารางจะเห็นได้ว่า วิธีควอนไทล์ (QT) ให้ค่าฝนรายวันที่มากที่สุดโดยรวมใกล้เคียงกว่าค่าตรวจวัดจริงคิดเป็นประมาณ 83.3% จากทั้งหมด 12 กรณี (6 สถานี x 2 เปอร์เซ็นต์การสูญหาย) และจากกราฟรูปที่ 4 สังเกตได้ว่าวิธี QT ให้ค่าฝนรายวันมากที่สุดมีแนวโน้มใกล้เคียงจำนวนมากกว่าวิธี IDW โดยเฉพาะอย่างยิ่งที่เปอร์เซ็นต์การมีค่าสูญหาย 20%

ตารางที่ 4 เปรียบเทียบฝนรายวันมากที่สุด

สถานีเป้าหมาย	เปอร์เซ็นต์การสูญหาย					
	20%			40%		
	Obs.	IDW	QT	Obs.	IDW	QT
327008	**83.4**	92.4	84.8	**110.9**	113.23	113.33

329006	**120.2**	93.2	98.5	**123.0**	112.79	102.38
327003	**99.0**	84.4	113.5	**168.5**	71.02	112.08
329003	**92.0**	81.8	92.9	**119.6**	82.65	136.49
327009	**86.3**	73.5	88.3	**89.0**	104.52	109.43
327501	**135**	65.3	100.6	**135.0**	98.63	130.85

หมายเหตุตาราง: Obs = ค่าตรวจวัดจริง IDW = ค่าที่คำนวณได้จากวิธีระยะทางผกผัน QT = ค่าที่คำนวณได้จากวิธีควอนไทล์ และค่าที่ขีดเส้นใต้ = ค่าที่ใกล้เคียงค่าตรวจวัดมากที่สุด

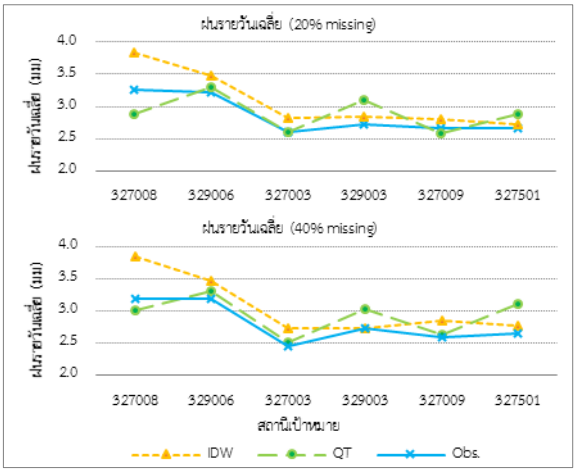


รูปที่ 4 เปรียบเทียบฝนรายวันมากที่สุดระหว่างค่าตรวจวัดจริงและวิธีประมาณค่าข้อมูลฝนรายวันที่ขาดหายไป

2) ฝนรายวันเฉลี่ย: จากข้อมูลที่ทำการคำนวณและเติมค่า เปรียบเทียบกับค่าตรวจวัดจริง เมื่อนำมาคำนวณหาค่าเฉลี่ย (ฝนรายวันเฉลี่ย) ได้ผลดังตารางที่ 5 และรูปที่ 5 จากตารางเห็นได้ว่า วิธีควอนไทล์ให้ค่าฝนรายวันโดยรวมใกล้เคียงกว่า (ค่าที่ขีดเส้นใต้) ค่าตรวจวัดจริงคิดเป็นความใกล้เคียงประมาณ 58.3 % จากทั้งหมด 12 กรณี และจากกราฟรูปที่ 5 สังเกตได้ว่าวิธี QT (สีเขียว) มีแนวโน้มใกล้เคียงกับค่าตรวจวัดจริง (สีฟ้า) มากกว่าวิธี IDW (สีส้ม)

ตารางที่ 5 เปรียบเทียบฝนรายวันเฉลี่ย

สถานีเป้าหมาย	เปอร์เซ็นต์การสูญหาย					
	20%			40%		
	Obs.	IDW	QT	Obs.	IDW	QT
327008	**3.3**	3.8	2.9	**3.2**	3.8	3.0
329006	**3.2**	3.5	3.3	**3.2**	3.5	3.3
327003	**2.6**	2.8	2.6	**2.4**	2.7	2.5
329003	**2.7**	2.8	3.1	**2.7**	2.7	3.0
327009	**2.7**	2.80	2.6	**2.6**	2.8	2.6
327501	**2.7**	2.7	2.9	**2.6**	2.8	3.1

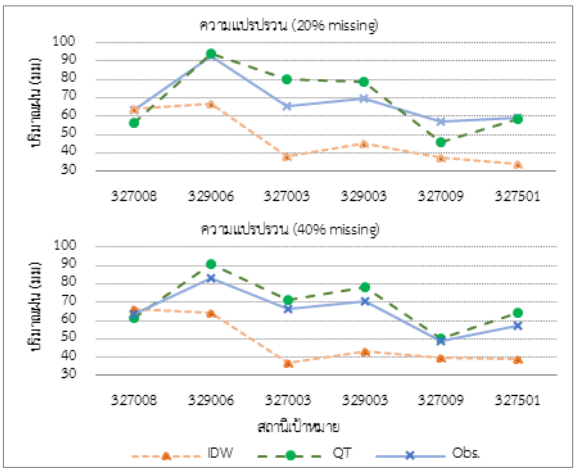


รูปที่ 5 เปรียบเทียบฝนรายวันเฉลี่ยระหว่างค่าตรวจวัดจริงและวิธีประมาณค่าข้อมูลฝนรายวันที่ขาดหายไป

3) ค่าความแปรปรวน: จากข้อมูลที่ทำการคำนวณและเติมค่ากับค่าตรวจวัดจริง นำมาคำนวณค่าความแปรปรวน ได้ผลดังตารางที่ 6 และรูปที่ 6 ซึ่งจะเห็นได้ว่า วิธี QT ให้ค่าความแปรปรวนฝนรายวันมากกว่าวิธี IDW แต่มีค่าใกล้เคียงกับค่าความแปรปรวนของค่าตรวจวัดจริงมากกว่าอย่างชัดเจน คิดเป็น 91.7 % ซึ่งสามารถเห็นได้ชัดดังในกราฟรูปที่ 6

ตารางที่ 6 เปรียบเทียบค่าความแปรปรวนฝนรายวัน

สถานีเป้าหมาย	เปอร์เซ็นต์การสูญหาย					
	20%			40%		
	Obs.	IDW	QT	Obs.	IDW	QT
327008	**63.3**	63.7	56.6	**63.9**	66.3	61.6
329006	**92.7**	66.7	94.2	**82.8**	64.5	91.0
327003	**65.3**	37.8	80.2	**66.5**	36.9	71.4
329003	**69.7**	44.8	78.9	**70.5**	43.4	78.4
327009	**56.7**	37.7	45.9	**48.9**	39.5	50.6
327501	**59.4**	34.2	58.5	**57.6**	38.7	64.0



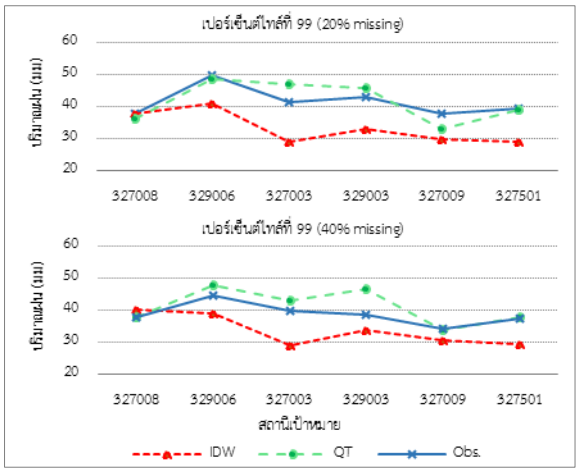
รูปที่ 6 เปรียบเทียบค่าความแปรปรวนฝนรายวันระหว่างค่าตรวจวัดจริงและวิธีประมาณค่าข้อมูลฝนรายวันที่ขาดหายไป

4) ค่าเปอร์เซ็นต์ไทล์ที่ 99: จากข้อมูลที่ทำการคำนวณและเติมค่า กับค่าตรวจวัดจริง นำมาหาค่าความเปอร์เซ็นต์ไทล์ที่ 99 ดังตารางที่ 7 และรูปที่ 7 ซึ่งค่าเปอร์เซ็นต์ไทล์ฝนรายวันที่ 99 หมายถึง โอกาสที่ฝนจะตกหนักหรือตกมากกว่า

ค่าดังกล่าวมีเพียง 1 % ค่าเปอร์เซ็นต์ไทล์ที่ 99 สามารถใช้แสดงถึงปริมาณฝนสุดขีด (extreme rainfall) ได้ จากการคำนวณพบว่า วิธี QT ให้ค่าเปอร์เซ็นต์ไทล์ที่ 99 ใกล้เคียงกว่าค่าตรวจวัดจริงมากกว่าวิธี IDW อย่างชัดเจน คิดเป็นความใกล้เคียงประมาณ 83.3 % และจากกราฟรูปที่ 9 เห็นได้ว่าวิธี QT (เส้นสีเขียว) มีแนวโน้มให้ค่าเปอร์เซ็นต์ไทล์ที่ 99 ใกล้เคียงกับค่าตรวจวัดจริงอย่างเห็นได้ชัด ขณะที่วิธี IDW มักให้ค่าต่ำกว่าตรงกับข้อจำกัดของวิธี IDW ที่มักให้ค่าฝนกรณีฝนสุดขีดต่ำกว่าค่าจริง

ตารางที่ 7 เปรียบเทียบค่าเปอร์เซ็นต์ไทล์ที่ 99 ของฝนรายวัน

สถานีเป้าหมาย	เปอร์เซ็นต์การสูญหาย					
	20%			40%		
	Obs.	IDW	QT	Obs.	IDW	QT
327008	**37.8**	37.5	36.2	**37.7**	40.0	37.7
329006	**49.6**	40.7	48.6	**44.5**	38.9	47.9
327003	**41.2**	29.0	47.0	**40.0**	29.1	43.2
329003	**42.8**	32.7	45.6	**38.8**	34.0	46.7
327009	**37.5**	29.8	32.8	**34.3**	30.5	33.8
327501	**39.4**	28.7	38.7	**37.4**	29.6	37.8



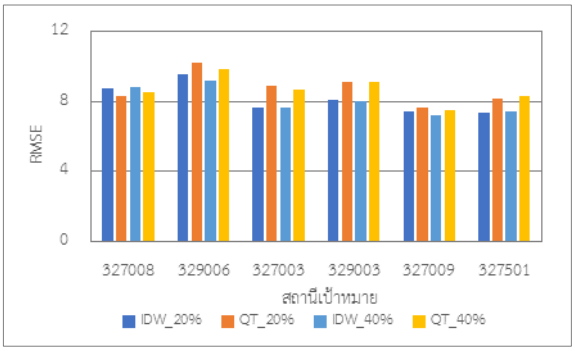
รูปที่ 7 เปรียบเทียบฝนรายวันเฉลี่ยระหว่างค่าตรวจวัดจริงและวิธีประมาณค่าข้อมูลฝนรายวันที่ขาดหายไป

4.2 การประเมินค่าประสิทธิภาพ: ค่าความคลาดเคลื่อน

ค่าความคลาดเคลื่อนระหว่างค่าจริงและค่าที่คำนวณได้ประเมินจากวิธีค่ารากที่สองความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) และค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์ (Mean Absolute Error, MAE) โดยถ้าค่า RMSE และ MAE มีค่าน้อย หมายถึง ให้ค่าความคลาดเคลื่อนน้อย ผลการคำนวณค่า RMSE แสดงดังตารางที่ 8 และรูปที่ 8 และค่า MAE แสดงดังตารางที่ 9 และรูป 8 จากผลการคำนวณพบว่าการเติมค่าทั้งวิธี IDW และ QT ให้ค่าความคลาดเคลื่อนไม่มากนัก โดยมีค่า RMSE อยู่ระหว่าง 7.2-10.2 โดยที่วิธี IDW มีค่า RMSE เฉลี่ยเท่ากับ 8.1 และวิธี QT มีค่า RMSE เฉลี่ยเท่ากับ 8.7 ส่วนค่า MAE ของทั้งสองวิธีมีค่าอยู่ระหว่าง 2.1-3.8 โดยที่วิธี IDW มีค่า MAE เฉลี่ยเท่ากับ 2.95 และวิธี QT มีค่า MAE เฉลี่ยเท่ากับ 3.15 ดังนั้นเมื่อเปรียบเทียบระหว่างวิธี พบว่า วิธี IDW ให้ค่าความคลาดเคลื่อน RMSE และ MAE น้อยกว่าวิธีควอนไทล์ ทั้งนี้เนื่องจาก IDW ให้ค่าฝนรายวันที่ประมาณค่าได้มีความแปรปรวนต่ำกว่าวิธี QT ขณะที่วิธี QT ให้ค่าความแปรปรวนสูงแต่สอดคล้องกับค่าตรวจวัดจริงมากกว่า ส่งผลให้การคำนวณค่าความคลาดเคลื่อน RMSE และ MAE ของวิธี QT มีค่ามากกว่า อย่างไรก็ตามค่าความคลาดเคลื่อนดังกล่าวของทั้งสองวิธีแตกต่างกันไม่มากนัก (แตกต่างน้อยกว่า 10%)

ตารางที่ 8 ค่ารากที่สองความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) ระหว่างวิธี IDW และ QT

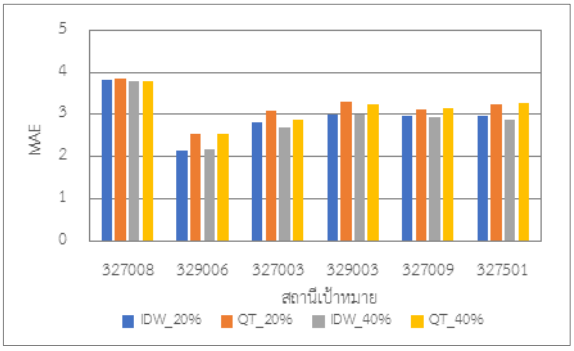
สถานีเป้าหมาย	เปอร์เซ็นต์การสูญหาย			
	20%		40%	
	IDW	QT	IDW	QT
327008	8.7	8.3	8.8	8.5
329006	9.5	10.2	9.2	9.9
327003	7.7	8.9	7.6	8.7
329003	8.1	9.1	8.0	9.1
327009	7.4	7.6	7.2	7.5
327501	7.4	8.2	7.4	8.3
เฉลี่ย	**8.1**	**8.7**	**8.1**	**8.7**



รูปที่ 8 ค่ารากที่สองความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE)

ตารางที่ 9 ค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) ระหว่างวิธี IDW และ QT

สถานีเป้าหมาย	เปอร์เซ็นต์การสูญหาย			
	20%		40%	
	IDW	QT	IDW	QT
327008	3.8	3.8	3.8	3.8
329006	2.1	2.5	2.2	2.5
327003	2.8	3.1	2.7	2.9
329003	3.0	3.3	3.0	3.2
327009	3.0	3.1	2.9	3.1
327501	3.0	3.2	2.9	3.3
เฉลี่ย	**3.0**	**3.2**	**2.9**	**3.1**



รูปที่ 9 ค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์

5. สรุปและวิจารณ์ผล

งานวิจัยศึกษานี้มีวัตถุประสงค์เพื่อศึกษาและพัฒนาวิธีการเติมค่าข้อมูลฝนรายวันที่ขาดหายไปด้วยวิธีควอนไทล์ (QT) ซึ่งใช้การแจกแจงความถี่แบบเบอร์นูลี่-แกมม่า เปรียบเทียบกับวิธี IDW ซึ่งเป็นวิธีที่นิยมใช้ทั่วไปมากที่สุดวิธีหนึ่ง ผลการศึกษาทั้งค่าทางสถิติพื้นฐานต่างๆ ได้แก่ ค่าฝนรายวันสูงสุด ฝนรายวันเฉลี่ย และความแปรปรวน การเติมค่าสูญหายด้วยวิธี QT ให้ค่าทางสถิติพื้นฐานโดยรวมที่เปอร์เซ็นต์การสูญหาย 20% และ 40% ดีกว่าวิธี IDW และปริมาณฝนที่ค่าเปอร์เซ็นต์ไทล์ 99 วิธี QT ให้ค่าใกล้เคียงกับค่าตรวจวัดมากกว่าวิธี IDW ค่อนข้างชัดเจน บ่งชี้ได้ว่าวิธี QT สามารถประมาณค่ากรณีเหตุการณ์ฝนสุดขีดหรือฝนตกหนักได้ดีกว่าวิธี IDW ซึ่งมักมีแนวโน้มให้ค่าฝนต่ำกว่าฝนตกหนักจริง

การประเมินประสิทธิภาพด้วยค่ารากที่สองความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) และค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) วิธี IDW ให้ค่าความคลาดเคลื่อนน้อยกว่าวิธี QT ทั้งนี้เนื่องจากวิธี IDW ให้ค่าความแปรปรวนต่ำกว่าค่าตรวจวัดจริง ซึ่งเกิดจากการเฉลี่ยค่าจากสถานีอ้างอิงที่อยู่ใกล้เคียงที่สุด 3 สถานีทำให้มีความแปรปรวนของค่าฝนน้อยกว่า ขณะที่วิธี QT มีค่าความแปรปรวนมากกว่าเนื่องจากอ้างอิงค่าฝนจากสถานีข้างเคียงที่มีค่า r ที่มากสุดเพียง 1 สถานีในการหาโอกาสการเกิดฝนตก ซึ่งวิธี QT ให้ค่าความแปรปรวนมากกว่าแต่สอดคล้องกับข้อมูลตรวจวัดจริงที่มีความแปรปรวนมากเช่นกัน ส่งผลให้ได้ค่าคำนวณ RMSE และ MAE มากกว่า อย่างไรก็ตามความคลาดเคลื่อนดังกล่าวของทั้งสองวิธีการเติมค่าสูญหายมีค่าแตกต่างกันไม่มากนัก

ผลการศึกษาโดยรวมสรุปได้ว่าวิธี QT สามารถใช้ประมาณค่าและเติมค่าสูญหายข้อมูลฝนรายวันได้โดยรวมดีกว่าวิธี IDW โดยเฉพาะอย่างยิ่งกรณีเหตุการณ์ฝนสุดขีด (extreme rainfall) ดังนั้นประโยชน์ที่ได้จากการศึกษานี้ คือ วิธีการเติมค่าสูญหายข้อมูลที่ให้ค่าใกล้เคียงกว่าวิธีดั้งเดิม ส่งผลให้การศึกษาวิเคราะห์งานด้านต่างๆ ที่เกี่ยวข้องสามารถให้ผลที่ถูกต้องและแม่นยำมากขึ้นโดยเฉพาะอย่างยิ่งกรณีเหตุการณ์ฝนตกหนัก

กิตติกรรมประกาศ

ขอขอบคุณนายณัฐพงศ์ บุญประเสริฐ นางสาวสุธินี วงศ์แสง นิสิตที่ดูแลภายใต้การจัดทำโครงงานทางวิศวกรรมศาสตร์ระดับปริญญาวิศวกรรมศาสตรบัณฑิต ที่มีส่วนช่วยในการจัดเตรียมข้อมูล การริเริ่มเพื่อทดลองแนวคิดและการพัฒนาวิธีควอนไทล์ส่งผลให้งานวิจัยนี้สำเร็จลุล่วงได้

เอกสารอ้างอิง

[1] J. Kim and J. H. Ryu, "A Heuristic Gap Filling Method for Daily Precipitation Series," *Water Resources Management,* journal article vol. 30, no. 7, pp. 2275-2294, May 01 2016, doi: 10.1007/s11269-016-1284-z.

[2] C. Simolo, M. Brunetti, M. Maugeri, and T. Nanni, "Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach," *International Journal of Climatology,* vol. 30, no. 10, pp. 1564-1576, 2010, doi: 10.1002/joc.1992.

[3] M. M. Hasan and B. F. W. Crokea, "Filling gaps in daily rainfall data : a statistical approach," in *20th International Congress on Modelling and Simulation*, Adelaide, Australia, 1–6 December 2013 2013.

[4] D. A. Mooley, "Gamma Distribution Probability Model for Asian Summer Monsoon Monthly Rainfall," *Monthly Weather Review,* vol. 101, no. 2, pp. 160-176, 1973, doi: 10.1175/1520-0493(1973)101<0160:Gdpmfa>2.3.Co;2.

[5] H. Aksoy, "Use of gamma distribution in hydrological analysis," *Turkish Journal of Engineering and Environmental Sciences,* vol. 24, no. 6, pp. 419-428, 2000.