

รายงานโครงการวิศวกรรมไฟฟ้า วิชา 2102499 ปีการศึกษา 2560

การพยากรณ์แบบเวียนเกิดสำหรับความเข้มแสงที่จุฬาลงกรณ์มหาวิทยาลัย

Recursive Estimation of Solar forecasting at Chulalongkorn University

นายโทนี่ ฟาง เลขประจำตัว 5730212721

อาจารย์ที่ปรึกษา ผศ.ดร. จิตโกมุท สังศิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

การพยากรณ์ความเข้มแสงโดยใช้อินพุทเวลาเป็นหนึ่งในวิธีการพยากรณ์ความเข้มแสง ซึ่งการพยากรณ์ความเข้มแสงมีประโยชน์คือผู้ใช้สามารถวางแผนเดินเครื่องกำเนิดไฟฟ้าและเป็นส่วนสำคัญในตลาดซื้อขายไฟฟ้า วัตถุประสงค์ของรายงานนี้คือการเปรียบเทียบการพยากรณ์ความเข้มแสงที่มีอัตราการชักตัวอย่างของข้อมูลที่ต่างกัน และการเปรียบเทียบการพยากรณ์ค่าความเข้มแสงที่ขึ้นต่างๆ ในรายงานนี้จะนำข้อมูลความเข้มแสงมาทำการลดอัตราการชักตัวอย่างและหาแบบจำลองที่เหมาะสมกับข้อมูลแต่ละชุด โดยการใช้วิธีภาวะน่าจะเป็นสูงสุดและใช้เกณฑ์ข้อสนเทศของเอนโทรปีและเกณฑ์ข้อสนเทศของเบย์ในการเลือกแบบจำลองที่เหมาะสมที่สุด จากนั้นจะนำแบบจำลองที่ได้มานั้นมาหาค่าพยากรณ์แล้วเปรียบเทียบความแม่นยำในการพยากรณ์ของแต่ละแบบจำลองว่ามีความแม่นยำอย่างไร และรายงานนี้ยังศึกษาค่าความผิดพลาดในการพยากรณ์ที่เวลาถัดไปบนแบบจำลองเดียวกัน ซึ่งรายงานนี้ให้ข้อสรุปว่าการพยากรณ์ความเข้มแสงที่เวลาไกลมักจะให้ค่าผิดพลาดที่มากกว่าเวลาใกล้

สารบัญ

1	บทนำ	5
2	วัตถุประสงค์	5
2.1	ขอบเขต	5
2.2	ผลลัพธ์ที่คาดหวัง	5
3	การหาแบบจำลองและการเลือกแบบจำลองที่เหมาะสมที่สุด	5
3.1	การหาคาบของสัญญาณ	7
3.2	การหาค่าแตกต่างของสัญญาณ	8
3.3	การหาสัมประสิทธิ์ของแบบจำลอง ARMA	8
3.4	การคำนวณค่าพยากรณ์	9
3.5	การเลือกอันดับของแบบจำลองที่เหมาะสมที่สุด	10
4	ผลลัพธ์ของโครงการ	11
4.1	การเตรียมข้อมูล	11
4.2	การหาคาบของสัญญาณ	11
4.3	การหาอันดับของการทำอนุพันธ์ตามฤดูกาลและการหาอันดับของการทำอนุพันธ์	12
4.4	การหาอันดับของแบบจำลอง ARMA ที่เหมาะสมที่สุด	13
5	การวิเคราะห์และวิจารณ์ผลการพยากรณ์	17
6	บทสรุป	20
7	ภาคผนวก	21
7.1	ฟังก์ชันอัตโนมัติสหสัมพันธ์ (Autocorrelation Function, ACF)	21
7.2	กราฟ ACF ที่ใช้ในการตัดสินใจหาค่า d และ D ของข้อมูลแต่ละชุด	22

สารบัญรูป

1	แผนผังการหาแบบจำลอง	6
2	ตัวอย่างข้อมูลความเข้มแสง	6
3	ค่าเฉลี่ยของข้อมูลความเข้มแสงที่เวลาต่างๆตลอดปี พ.ศ.2560	11
4	PSD ของข้อมูลแต่ละชุด	12
5	ACF และ PACF เมื่อ $d = 1$ และ $D = 1$ ของข้อมูลชุด 60 นาที	13
6	AIC และ BIC ของข้อมูลชุด 60 นาที	14
7	ค่าความผิดพลาดบนชุดข้อมูลตรวจสอบความสมเหตุสมผลของข้อมูลชุด 60 นาที	14
8	ACF และ PACF เมื่อ $d = 1$ และ $D = 1$ ของข้อมูลชุด 15 นาที	15
9	AIC และ BIC ของข้อมูลชุด 15 นาที	16
10	ค่าความผิดพลาดบนชุดข้อมูลตรวจสอบความสมเหตุสมผลของข้อมูลชุด 15 นาที	16
11	ACF และ PACF เมื่อ $d = 1$ และ $D = 1$ ของข้อมูลชุด 3 นาที	17
12	AIC และ BIC ของข้อมูลชุด 3 นาที	18
13	ผลการพยากรณ์ที่ขึ้นต่างๆของข้อมูลแต่ละชุด	19
14	ค่าความผิดพลาดที่ได้จากการพยากรณ์ค่าที่เวลาถัดไป 1 ชั่วโมงโดยเทียบกับข้อมูลที่วัดได้ของแบบจำลองต่างๆ	19
15	ค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 3 นาทีกับค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 15 นาทีเทียบกับข้อมูลจริง	20
16	ค่าความผิดพลาดของแบบจำลองของข้อมูลชุด 3 นาทีกับค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 15 นาทีเทียบกับข้อมูลจริง	20
17	ค่าพยากรณ์ของแบบจำลองทั้ง 3 เทียบกับข้อมูลจริง	20
18	ค่าความผิดพลาดของแบบจำลองของข้อมูลชุด 3 นาทีกับค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 15 นาทีเทียบกับข้อมูลจริง	20
19	ค่าความผิดพลาดที่ได้จากการพยากรณ์ค่าที่เวลาถัดไปของแบบจำลองต่างๆ	21
20	ตัวอย่างผลการพยากรณ์ที่ขึ้นต่างๆของข้อมูลแต่ละชุด	22
21	กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d และ D ของข้อมูลชุด 60 นาที	23
22	กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d และ D ของข้อมูลชุด 15 นาที	23
23	กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d และ D ของข้อมูลชุด 3 นาที	23

1 บทนำ

วิธีการพยากรณ์ความเข้มแสงนั้นมีได้หลายวิธี ซึ่งหนึ่งในวิธีการพยากรณ์นั้นคือการใช้อนุกรมเวลา (Time Series Model) ในการพยากรณ์ความเข้มแสง อนุกรมเวลาเป็นแบบจำลองรูปแบบหนึ่งที่ใช้ในการอธิบายลักษณะของสิ่งที่สนใจในอดีตและสามารถใช้ทำนายความเป็นไปอนาคตมากไปกว่านั้น อนุกรมเวลายังถูกใช้ในสาขาต่างๆ เช่น วิทยาศาสตร์และบัญชี เป็นต้น อนุกรมเวลายังถูกแบ่งออกเป็น 2 ประเภทนั้นคืออนุกรมเวลาที่มีสมบัตินิ่ง (stationary) และแบบจำลองอนุกรมเวลาที่ไม่สมบัตินิ่ง (non-stationary)

1. แบบจำลองอนุกรมเวลาที่มีสมบัตินิ่ง คือ แบบจำลองอนุกรมเวลาที่ใช้กับข้อมูลที่มีค่าเฉลี่ยเป็นค่าคงที่และมีฟังก์ชันอัตโนมัติสหสัมพันธ์ (autocorrelation function) ขึ้นกับระยะเวลาที่ห่างกัน เช่น แบบจำลอง AR, MA, ARMA, ARX, ARMAX
2. แบบจำลองอนุกรมเวลาที่ไม่สมบัตินิ่ง คือ แบบจำลองอนุกรมเวลาที่ใช้กับข้อมูลที่มีค่าเฉลี่ยไม่คงที่และมีฟังก์ชันอัตโนมัติสหสัมพันธ์ (autocorrelation function) ขึ้นกับระยะเวลาที่ห่างกัน เช่น แบบจำลอง ARIMA, ARIMAX

รายงานนี้จะใช้แบบจำลองอนุกรมเวลาที่ไม่สมบัตินิ่งในการอธิบายความเข้มแสงซึ่งอ้างอิงจาก [5] ที่ใช้แบบจำลอง SARIMA เมื่อมีการวัดของข้อมูลเข้ามา แบบจำลองจะถูกปรับเปลี่ยนโดยใช้ข้อมูลเดิมที่มีและข้อมูลใหม่ที่วัดเข้ามา และเนื่องด้วยการหาแบบจำลองอาจใช้เวลานาน ทฤษฎีเวียนเกิดจึงได้มีการนำมาใช้กับการหาแบบจำลอง เช่น recursive least-square (RLS) หรือ recursive prediction error method (RPEM)

ในปัจจุบันทฤษฎีเวียนเกิดได้มีการถูกนำมาใช้บ้างในการหาแบบจำลอง เช่น ในบทความวิจัยของ Peder Bacher, Henrik Madsen และ Henrik Aalborg Nielsen [7] ได้ใช้วิธีการ recursive least square กับแบบจำลอง AR, ARX และ Finite Impulse Response (FIR) ซึ่งนำมาใช้ในการพยากรณ์พลังงานแสงอาทิตย์ โดยบทความวิจัยนี้ได้แสดงให้เห็นว่าสัมประสิทธิ์ภายในแบบจำลองมีการเปลี่ยนแปลงเสมอเมื่อมีการนำข้อมูลใหม่เข้ามา

2 วัตถุประสงค์

วัตถุประสงค์ของโครงการนี้คือ

- เพื่อนำทฤษฎีเวียนเกิดมาใช้กับแบบจำลองความเข้มแสงที่จุฬาลงกรณ์มหาวิทยาลัย
- เพื่อนำไปใช้กับศูนย์พยากรณ์ความเข้มแสงที่จุฬาลงกรณ์มหาวิทยาลัย

2.1 ขอบเขต

- ใช้ทฤษฎีเวียนเกิดกับแบบจำลอง SARIMA
- ข้อมูลความเข้มแสงจะถูกวัดจากจุฬาลงกรณ์มหาวิทยาลัย
- ค่าที่พยากรณ์ได้จะแสดงผลผ่านเครื่องบริการที่จุฬาลงกรณ์มหาวิทยาลัย

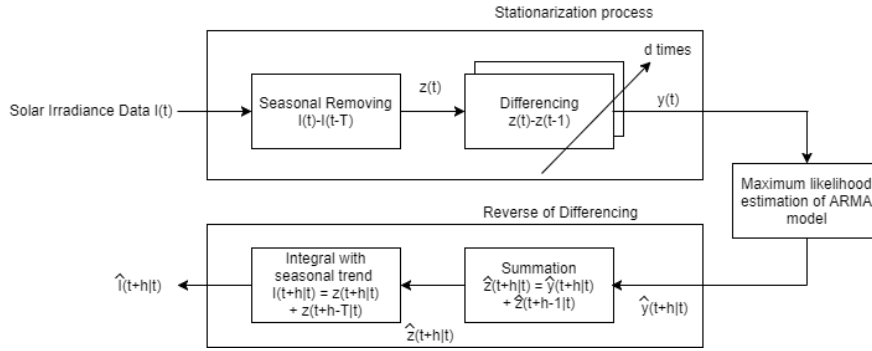
2.2 ผลลัพธ์ที่คาดหวัง

- สามารถนำทฤษฎีเวียนเกิดมาใช้กับแบบจำลองพยากรณ์ความเข้มแสงแบบทันทีบนเครื่องบริการของภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย

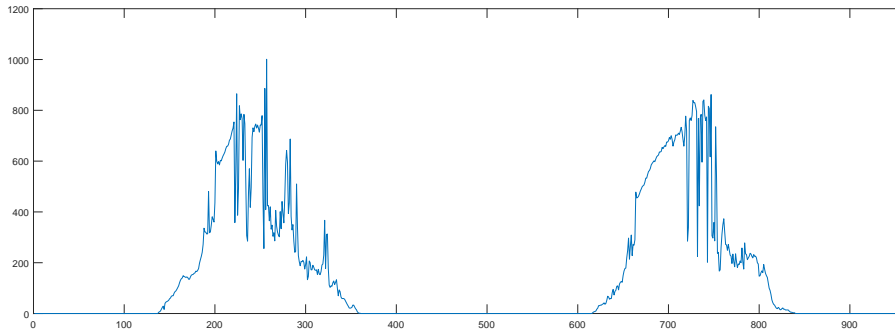
3 การหาแบบจำลองและการเลือกแบบจำลองที่เหมาะสมที่สุด

เนื้อหาส่วนนี้กล่าวถึงวิธีการหาแบบจำลองซึ่งแสดงดังรูป 1 เนื่องจากความเข้มแสงนั้นมีสมบัติเป็นคาบ หมายความว่าเราสามารถอธิบายส่วนที่เป็นคาบของความเข้มแสงด้วยอนุกรมฟูรีเยร์ หลังจากทำการลบส่วนที่เป็นคาบ จะพบว่าข้อมูลที่ได้จะมีลักษณะคล้ายสัญญาณสุ่ม แต่อาจจะมีส่วนที่ยังไม่เป็นคาบซึ่งยังไม่ถูกกำจัดโดยดูจากฟังก์ชันอัตโนมัติสหสัมพันธ์ (AutoCorrelation Function, ACF) ถ้าไม่มีการ cut-off ที่ lag ใด lag หนึ่ง ถือว่ายังไม่สมบัตินิ่ง (stationary) จึงได้ทำการหาค่าแตกต่างของสัญญาณ ซึ่งทำให้เรื่อยๆจนกว่าจะทำให้ข้อมูลมีสมบัตินิ่ง จากนั้นจึงนำข้อมูลที่ได้ไปหาสัมประสิทธิ์ภายในแบบจำลอง ARMA โดยใช้วิธีภาวะน่าจะเป็นสูงสุดในการหาสัมประสิทธิ์ภายในพหุนามถดถอยตัวเอง และพหุนามค่าเฉลี่ยเคลื่อนที่ จากนั้นจึงนำแบบจำลองที่ได้มาพยากรณ์ค่าความเข้มแสงที่เวลาถัดไป จากรูป 2 ข้อมูลความเข้มแสงสามารถวิเคราะห์ออกมาเป็นส่วนที่แสดงสมบัติเป็นคาบและส่วนที่เป็นการสุ่ม จึงสามารถใช้แบบจำลองประเภทเดียวกันในการประมาณค่าความเข้มแสงได้

จาก [5] ได้แสดงผลลัพธ์ว่าแบบจำลอง SARIMA(2, 2, 4)(0, 1, 1)₁₆ ให้ประสิทธิภาพในการพยากรณ์ความเข้มแสงได้ดีที่สุด ซึ่งในรายงานนี้จะใช้แบบจำลอง Seasonal AutoRegressive Integrated Moving Average หรือ SARIMA เพราะเป็นแบบจำลองในรูปแบบทั่วไปมากกว่า



รูป 1: แผนผังการหาแบบจำลอง



รูป 2: ตัวอย่างข้อมูลความเข้มแสง

ซึ่งแบบจำลอง SARIMA นิยามโดย

$$\tilde{A}(L)(1 - L^T)^D A(L)(1 - L)^d I(t) = \tilde{C}(L)C(L)e(t) \quad (1)$$

จาก (1) สามารถเขียนได้เป็น SARIMA(p, d, q)(P, D, Q) $_T$ โดยที่

$$\begin{aligned} A(L) &= 1 - (A_1L + A_2L^2 + \dots + A_pL^p) \\ C(L) &= 1 + (C_1L + C_2L^2 + \dots + C_qL^q) \\ \tilde{A}(L) &= 1 - (\tilde{A}_1L^T + \tilde{A}_2L^{2T} + \dots + \tilde{A}_PL^{PT}) \\ \tilde{C}(L) &= 1 + (\tilde{C}_1L^T + \tilde{C}_2L^{2T} + \dots + \tilde{C}_QL^{QT}) \end{aligned} \quad (2)$$

L คือตัวดำเนินการล่าหลัง (lag operator), T คือคาบของฤดูกาล (seasonal period), d คืออันดับของการทำปริพันธ์ (non-seasonal integrated order), D คืออันดับของการทำปริพันธ์ตามฤดูกาล (seasonal integrated order), $A(L)$ คือพหุนามถดถอยตัวเอง (AutoRegressive polynomial), $C(L)$ คือพหุนามค่าเฉลี่ยเคลื่อนที่ (Moving Average polynomial), $\tilde{A}(L)$ คือพหุนามถดถอยตัวเองที่มีเทรนด์ฤดูกาล (Seasonal AutoRegressive polynomial) และ $\tilde{C}(L)$ คือพหุนามค่าเฉลี่ยเคลื่อนที่ที่มีเทรนด์ฤดูกาล (Seasonal Moving Average polynomial)

จากรูป 2 จะพิจารณาว่าความเข้มแสงเกิดจากส่วนที่เป็นคาบบวกกับสัญญาณรบกวนซึ่งเขียนได้เป็นสมการ (3)

$$A(L)I(t) = s(t) + \alpha + C(L)e(t) \quad (3)$$

โดย $s(t) = s(t - T)$ คือส่วนที่เป็นคาบ α เป็นค่าคงที่และ $e(t)$ คือสัญญาณรบกวน จากรูป 1 กำหนดให้ $I(t)$ คือข้อมูลความเข้มแสง จากนั้น $I(t)$ ถูกลบด้วย $I(t - T)$ เพื่อกำจัดส่วนที่เป็นคาบ จะได้

$$\begin{aligned} A(L)I(t) - A(L)I(t - T) &= s(t) + \alpha + C(L)e(t) - s(t - T) - \alpha - C(L)e(t - T) \\ &= C(L)e(t) - C(L)e(t - T) \end{aligned}$$

$$A(L)(1 - L^T)I(t) = C(L)(1 - L^T)e(t) \quad (4)$$

$$A(L)(1 - L^T)I(t) = C(L)\tilde{C}(t)e(t) \quad (5)$$

แทนค่า $(1 - L^T)I(t) = z(t)$ และ $\tilde{C}(t)e(t) = \eta(t)$ จากนั้นทำการหาส่วนต่าง $z(t)$ เป็นจำนวน d ครั้ง จะได้

$$\begin{aligned} A(L)(1 - L)^d z(t) &= C(L)(1 - L)^d \eta(t) \\ A(L)(1 - L)^d z(t) &= C(L)v(t) \end{aligned} \quad (6)$$

แทนค่า $y(t) = (1 - L)^d z(t)$ จะได้

$$A(L)y(t) = C(L)v(t) \quad (7)$$

หลังจากใช้ ML ในการหาพารามิเตอร์ภายในแบบจำลองค่าเฉลี่ยเคลื่อนที่ถดถอยตัวเอง(AutoRegressive Moving Average, ARMA) เราจะได้แบบจำลอง ARMA ดังสมการ (8)

$$\begin{aligned} C(L)\hat{y}(t|t-1) &= (C(L) - A(L))y(t) \\ (C(L) - 1)\hat{y}(t|t-1) + \hat{y}(t|t-1) &= (C(L) - 1)y(t) - (A(L) - 1)y(t) \\ \hat{y}(t|t-1) &= (C(L) - 1)(y(t) - \hat{y}(t|t-1)) - (A(L) - 1)y(t) \\ \hat{y}(t|t-1) &= (C(L) - 1)e(t) - (A(L) - 1)y(t) \\ \hat{y}(t|t-1) &= (C_1L + C_2L^2 + \dots + C_qL^q)e(t) + \\ &\quad (A_1L + A_2L^2 + \dots + A_pL^p)y(t) \end{aligned} \quad (8)$$

จากนั้นจะได้แบบจำลอง ARIMA ซึ่งอ้างอิงจากรูป 1 ดังสมการ (9)

$$\begin{aligned} \hat{z}(t|t-1) &= \hat{y}(t|t-1) + \hat{z}(t-1|t-1) \\ \hat{z}(t|t-1) &= \hat{y}(t|t-1) + \hat{y}(t-1|t-1) + \hat{z}(t-2|t-1) \\ \hat{z}(t|t-1) &= \hat{y}(t|t-1) + \sum_{k=1}^n \hat{y}(t-k|t-1) \end{aligned} \quad (9)$$

และได้แบบจำลอง SARIMA ซึ่งอ้างอิงจากรูป 1 ดังสมการ (10)

$$\begin{aligned} \hat{I}(t|t-1) &= \hat{z}(t|t-1) + \hat{I}(t-T|t-1) \\ \hat{I}(t|t-1) &= \hat{z}(t|t-1) + \hat{z}(t-T|t-1) + \hat{I}(t-2T|t-1) \\ \hat{I}(t|t-1) &= \hat{z}(t|t-1) + \sum_{k=1}^n \hat{z}(t-kT|t-1) \end{aligned} \quad (10)$$

3.1 การหาคาบของสัญญาณ

เนื้อหาส่วนนี้กล่าวถึงการหาส่วนที่เป็นคาบของสัญญาณเพื่อนำมาลบออกจากข้อมูล โดยมีเป้าหมายคือหาคาบของสัญญาณและทำให้ข้อมูลชุดนั้นมีสมบัตินิ่ง จาก [3] ส่วนที่เป็นคาบสามารถเขียนให้อยู่ในรูปอนุกรมฟูริเยร์ได้ดังสมการ

$$s(t) = \sum_{i=1}^k (a_i \cos(\omega_i t) + b_i \sin(\omega_i t)) \quad (11)$$

โดย $t = 1, 2, \dots, N$, a_i เป็นสัมประสิทธิ์ของส่วนที่เป็นโคไซน์ในแต่ละความถี่ ω_i และ b_i เป็นสัมประสิทธิ์ของส่วนที่เป็นไซน์ในแต่ละความถี่ ω_i โดย ω_i จะหาจากการวิเคราะห์ข้อมูลในเชิงความถี่เพื่อหา power spectral density จากนั้นความถี่ที่มีพลังงานสูงจะถูกกำหนดให้เป็น ω_i โดยก่อนที่จะหา power spectral density ในรายงานนี้ใช้ขั้นตอนวิธีแปลงฟูริเยร์แบบเร็ว (Fast Fourier Transform, FFT) เพื่อเปลี่ยนข้อมูลให้อยู่ในเชิงความถี่ อ้างอิงจาก [4] Fast Fourier Transform เป็นหนึ่งในอัลกอริทึมที่ใช้ในการหาผลการแปลงฟูริเยร์เชิงวิฤต (Discrete Fourier Transform, DFT) ซึ่งแสดงใน (12)

$$S(k) = \sum_{t=0}^{N-1} s(t) e^{j2\pi kt/N} \quad (12)$$

โดย $\omega_k = \frac{2\pi k}{N}$ และ $k = 0, 1, \dots, N-1$. หลังจากใช้ FFT ก็จะมาหา $|S(k)|$ เพื่อหา ω_k ซึ่งจะมีเพียงความถี่ที่มีพลังงานสูงเท่านั้นที่จะถูกกำหนดให้เป็น ω_i ในสมการ (11).

3.2 การหาค่าแตกต่างของสัญญาณ

เนื้อหาส่วนนี้กล่าวถึงการหาค่าแตกต่างของข้อมูล (differencing) โดยมีเป้าหมายในการหาอันดับของการทำปริพันธ์ d จาก (5) ซึ่งถูกแทนด้วย

$$A(L)z(t) = C(L)e(t)$$

ข้อมูล $z(t)$ ถูกทำอนุปริพันธ์โดยการนำ $z(t-1)$ นำไปลบ จะได้

$$\Delta A(L)z(t) = A(L)z(t) - A(L)z(t-1) = A(L)(1-L)z(t)$$

และถ้าหาค่าแตกต่างของข้อมูล 2 ครั้งจะได้

$$\Delta^2 A(L)z(t) = \Delta A(L)z(t) - \Delta A(L)z(t-1) = A(L)(1-L)^2 z(t)$$

ถ้าหาค่าแตกต่างของข้อมูล d ครั้ง เราสามารถเขียนได้เป็นสมการ

$$\Delta^d A(L)z(t) = A(L)(1-L)^d z(t) \quad (13)$$

อ้างอิงจาก [5] ถ้า ACF ของข้อมูลชุดนี้มีการ cut-off ที่บาง lag หลังจากทำการหาค่าส่วนต่าง จำนวน d ครั้ง เราจะสรุปได้ว่าข้อมูลชุดนี้ ควรใช้แบบจำลองอนุกรมเวลาที่มีอันดับของการทำปริพันธ์ d และจะได้แบบจำลอง AutoRegressive Integrated Moving Average หรือ ARIMA ซึ่งมีสมการเป็น

$$A(L)(1-L)^d z(t) = C(L)v(t) \quad (14)$$

3.3 การหาสัมประสิทธิ์ของแบบจำลอง ARMA

เนื้อหาส่วนนี้จะเกี่ยวกับการหาแบบสัมประสิทธิ์ภายในแบบจำลอง ARMA ในรายงานนี้จะใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation, ML) ในการหาสัมประสิทธิ์ภายในพหุนามถดถอยตัวเอง และพหุนามค่าเฉลี่ยเคลื่อนที่ใน (2)

วิธีภาวะน่าจะเป็นสูงสุดเป็นหนึ่งในวิธีการหาค่าสัมประสิทธิ์ภายในพารามิเตอร์โดยการทำให้ค่าฟังก์ชันต้นทุนมีค่าสูงสุด โดยฟังก์ชันต้นทุนนิยามโดย

$$\mathcal{L}(y|\theta) = f(y(1), y(2), \dots, y(N)|\theta) \quad (15)$$

โดย $f(y(1), y(2), \dots, y(N)|\theta)$ ถูกนิยามด้วย

$$f(y(1), y(2), \dots, y(N)|\theta) = \prod_{t=1}^N \left(\frac{1}{\sigma\sqrt{2\pi}} \right) e^{-\frac{v(t)^2}{2\sigma^2}} \quad (16)$$

ใน (15), ฟังก์ชันต้นทุนถูกเรียกได้อีกอย่างคือฟังก์ชันภาวะน่าจะเป็นโดยที่

$$\theta = [A_1 \ A_2 \ \dots \ A_p \ C_1 \ C_2 \ \dots \ C_q \ \sigma^2]^T$$

$f(y|\theta)$ คือฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไข (conditional probability density function, conditional pdf) ของ $v(t)$ ใน (1) และ N เป็นจำนวนข้อมูล

จาก (7) สามารถเขียนได้อีกแบบเป็น

$$y(t) = A_1 y(t-1) + A_2 y(t-2) + \dots + A_p y(t-p) + v(t) + C_1 v(t-1) + C_2 v(t-2) + \dots + C_q v(t-q) \quad (17)$$

ดังนั้นเราสามารถหา $v(t)$ ในรูปของ $A_1, A_2, \dots, A_p, C_1, C_2, \dots, C_q$ จาก (17)

$$v(t) = y(t) - (A_1 y(t-1) + A_2 y(t-2) + \dots + A_p y(t-p)) - (C_1 v(t-1) + C_2 v(t-2) + \dots + C_q v(t-q)) \quad (18)$$

ถ้า $v(t)$ มีคุณสมบัติการแจกแจงแบบปกติ (normal distribution) โดยมีค่าเฉลี่ยเป็น 0 และมีค่าความแปรปรวนเป็น σ^2 ดังนั้น ฟังก์ชันลอการิทึมของฟังก์ชันภาวะน่าจะเป็น (log-likelihood function) จาก [2] สามารถเขียนได้เป็น

$$\mathcal{L}(\theta) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \sum_{t=1}^N \frac{v(t)^2}{2\sigma^2} \quad (19)$$

จาก [2], ถ้า $y(t)$ เป็นค่าจริงตั้งแต่ 1 ถึง p และ $v(t) = 0$ ตั้งแต่ $t = p, p-1, \dots, p-q+1$, จะได้ $y(t)$ ที่มีสมบัติการแจกแจงแบบปกติ ดังนั้น เราจะเริ่มที่ $t = p+1$ ในขณะเดียวกันฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขถูกเปลี่ยนเป็น (20)

$$\mathcal{L}(y|\theta) = f(y(p+1), y(p+2), \dots, y(N)|y(1), y(2), \dots, y(p), \theta) \quad (20)$$

ดังนั้นจะได้ฟังก์ชันความน่าจะเป็นคือ

$$\mathcal{L}(y|\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{N-p} e^{-\sum_{t=p+1}^N \frac{v(t)^2}{2\sigma^2}} \quad (21)$$

จากทฤษฎีของวิธีภาวะน่าจะเป็นสูงสุด เราสามารถหาค่าสัมประสิทธิ์ได้จากฟังก์ชันลอการิทึมของฟังก์ชันความน่าจะเป็นเช่นกัน ในที่สุด ค่าฟังก์ชันต้นทุนที่จะใช้ในการหาค่าสัมประสิทธิ์ถูกเปลี่ยนเป็น (22)

$$\log \mathcal{L}(y|\theta) = -\frac{N-p}{2} \log(2\pi) - \frac{N-p}{2} \log(\sigma^2) - \sum_{t=p+1}^N \frac{v(t)^2}{2\sigma^2} \quad (22)$$

โดย $\log \mathcal{L}(y|\theta)$ คือฟังก์ชันต้นทุนของปัญหานี้และพจน์สุดท้ายของ (22) สามารถเขียนอยู่ในรูป นอร์ม-2 ดังนั้นเราสามารถเขียนอยู่ในรูปเมทริกซ์ได้เป็น

$$\begin{bmatrix} y(p+1) \\ y(p+2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} y(p) & y(p-1) & \dots & y(1) & v(p) & v(p-1) & \dots & v(p-q+1) \\ y(p+1) & y(p) & \dots & y(2) & v(p+1) & v(p) & \dots & v(p-q+2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N) & y(N-1) & \dots & y(p) & v(N) & v(N-1) & \dots & v(p) \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \\ C_1 \\ C_2 \\ \vdots \\ C_q \end{bmatrix}$$

คำตอบเชิงปริมาณของ $\hat{\theta}_{ML}$ สามารถหาได้จากวิธีการหาค่าที่เหมาะสมที่สุดแบบมีเงื่อนไข เช่น interior-point เป็นต้น จาก [2] แบบจำลองอนุกรมเวลามีสมบัติเชิงเหตุ (causal) และสมบัตินกลับได้ (invertible) ดังนั้นเงื่อนไขดังกล่าวนั้นคือรากของพหุนามจะต้องอยู่ในวงกลมหนึ่งหน่วย ซึ่งเขียนได้เป็น

$$\begin{aligned} 1 - (A_1L + A_2L^2 + \dots + A_pL^p) &= 0 \\ 1 + (C_1L + C_2L^2 + \dots + C_qL^q) &= 0 \end{aligned}$$

3.4 การคำนวณค่าพยากรณ์

เนื้อหาส่วนนี้กล่าวถึงการคำนวณค่าพยากรณ์ โดยมีเป้าหมายในการหาค่าที่ได้จากการพยากรณ์โดยใช้ข้อมูลที่มีกับแบบจำลองที่หาสัมประสิทธิ์แล้ว จากสมการแบบจำลอง ARMA (7)

$$A(L)y(t) = C(L)v(t)$$

อ้างอิงจาก [6] และ [8] สมการที่ทำนายค่าที่เวลาถัดไปได้ใกล้เคียงที่สุดโดยพิสูจน์จากวิธี Prediction Error Method คือ

$$\hat{y}(t|t-1) = (1 - C^{-1}(L)A(L))y(t) \quad (23a)$$

$$e(t) = C^{-1}(L)A(L)y(t) \quad (23b)$$

เราจะได้สมการทำนายของแบบจำลอง ARMA(p, q)

$$\begin{aligned} C(L)\hat{y}(t|t-1) &= (C(L) - A(L))y(t) \\ (C(L) - 1)\hat{y}(t|t-1) + \hat{y}(t|t-1) &= (C(L) - 1)y(t) - (A(L) - 1)y(t) \\ \hat{y}(t|t-1) &= (C(L) - 1)(y(t) - \hat{y}(t|t-1)) - (A(L) - 1)y(t) \\ \hat{y}(t|t-1) &= (C(L) - 1)e(t) - (A(L) - 1)y(t) \\ \hat{y}(t|t-1) &= (C_1L + C_2L^2 + \dots + C_qL^q)e(t) + \\ &\quad (A_1L + A_2L^2 + \dots + A_pL^p)y(t) \end{aligned} \quad (24)$$

ดังนั้นเราสามารถคำนวณค่าที่เวลาถัดไป 1 ชั้นของแบบจำลอง ARMA(p, q)

$$\begin{aligned}\hat{y}(t+1|t) &= (C(L) - 1)e(t+1) + (1 - A(L))\hat{y}(t+1|t) \\ &= (C_1L + C_2L^2 + \dots + C_qL^q)e(t+1) + (A_1L + A_2L^2 + \dots + A_pL^p)\hat{y}(t+1|t) \\ &= C_1e(t) + C_2e(t-1) + C_3e(t-2) + \dots + C_qe(t-q+1) + \\ &\quad A_1\hat{y}(t|t) + A_2\hat{y}(t-1|t) + \dots + A_p\hat{y}(t-p+1|t)\end{aligned}\quad (25)$$

และสามารถคำนวณค่าที่เวลาถัดไป h ชั้นของแบบจำลอง ARMA(p, q)

$$\begin{aligned}\hat{y}(t+h|t) &= (C(L) - 1)e(t+h) + (1 - A(L))\hat{y}(t+h|t) \\ \hat{y}(t+h|t) &= C_1e(t+h-1) + C_2e(t+h-2) + C_3e(t+h-3) + \dots + C_qe(t+h-q) + \\ &\quad A_1\hat{y}(t+h-1|t) + A_2\hat{y}(t+h-2|t) + \dots + A_p\hat{y}(t+h-p|t)\end{aligned}\quad (26)$$

โดย

$$\hat{y}(t+h|t) = \begin{cases} \hat{y}(t+h|t) & t > 0 \\ y(t+h) & t \leq 0 \end{cases}\quad (27)$$

$$e(t+h|t) = \begin{cases} 0 & t > 0 \\ e(t+h) & t \leq 0 \end{cases}\quad (28)$$

ดังนั้นเราสามารถคำนวณค่าที่เวลาถัดไป h ชั้นของแบบจำลอง ARIMA($p, 1, q$)

$$\hat{z}(t+h|t) = \hat{y}(t+h|t) + \sum_{k=0}^n \hat{y}(t-k+h-1|t)\quad (29)$$

และก็สามารถคำนวณค่าที่เวลาถัดไป h ชั้นของแบบจำลอง SARIMA($p, 1, q$)(0, 1, 0) $_T$

$$\hat{I}(t+h|t) = \hat{z}(t+h|t) + \sum_{k=1}^n \hat{z}(t-kT+h|t)\quad (30)$$

โดยสูตรการคำนวณดังกล่าวนี้ได้มีการเปรียบเทียบกับคำสั่งต่างๆในซอฟต์แวร์ MATLAB ดังนี้

1. Infer

คำสั่ง **Infer** จะใช้แบบจำลองที่มีค่าสัมประสิทธิ์ ข้อมูลที่ต้องการหาและผลตอบสนองแรกเริ่ม(Presample response) เป็นตัวแปรขาเข้า ซึ่งจะให้ค่าที่พยากรณ์ถัดไป 1 ชั้นตามจำนวนข้อมูลที่ใส่เป็นตัวแปรขาเข้าเป็นตัวแปรขาออก

2. Forecast

คำสั่ง **Forecast** จะใช้แบบจำลองที่มีค่าสัมประสิทธิ์ จำนวนค่าที่จะพยากรณ์และผลตอบสนองแรกเริ่ม(Presample response) เป็นตัวแปรขาเข้า ซึ่งจะให้ค่าที่พยากรณ์ถัดไป h ชั้นตามจำนวนค่าที่จะพยากรณ์เป็นตัวแปรขาออก

3.5 การเลือกอันดับของแบบจำลองที่เหมาะสมที่สุด

เนื้อหาส่วนนี้กล่าวถึงการเลือกอันดับของแบบจำลองโดยใช้เกณฑ์ต่างๆเข้ามาใช้ในการเลือก หลังจากที่มีการหาคุณสมบัติบางอย่างของแบบจำลอง จึงต้องมีเกณฑ์ในการหาอันดับของแบบจำลอง SARIMA ที่เหมาะสมกับข้อมูล ในรายงานนี้ใช้เกณฑ์ข้อสนเทศของอาไคเกะ (Akaike Information Criterion, AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criterion, BIC) ทั้ง AIC และ BIC มีการได้อย่างเสียอย่างในเรื่องของความซับซ้อนของแบบจำลองและความสามารถในการปรับแบบจำลองเข้ากับข้อมูล ซึ่ง AIC และ BIC ถูกนิยามด้วย

$$AIC = -2\mathcal{L} + 2d\quad (31)$$

$$BIC = -2\mathcal{L} + d \log(N)\quad (32)$$

โดย \mathcal{L} คือฟังก์ชันลอการิทึมของฟังก์ชันภาวะน่าจะเป็น N คือจำนวนข้อมูล d คือจำนวนสัมประสิทธิ์ในแบบจำลอง

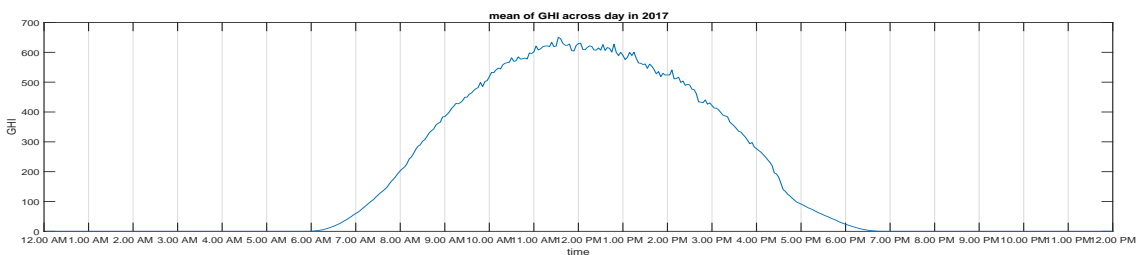
4 ผลลัพธ์ของโครงการงาน

เนื้อหาในส่วนนี้กล่าวถึงการเตรียมข้อมูลความเข้มแสงก่อนใช้หาแบบจำลอง ต่อมาเป็นผลการหาอันดับต่างๆของแบบจำลอง SARIMA โดยเริ่มจากการหาคาบของข้อมูลแล้วตามด้วยอันดับของการทำปริพันธ์ตามฤดูกาลและอันดับของการทำปริพันธ์ และสุดท้ายเป็นการหาอันดับของพหุนามต่างๆทั้ง 4 พหุนามที่ได้กล่าวไปเมื่อต้นรายงาน

4.1 การเตรียมข้อมูล

ในการทดลองนี้ได้ทำข้อมูลออกมาเป็นข้อมูลทั้งหมด 3 ชุดโดยมีวัตถุประสงค์ต่างกัน โดยข้อมูลที่นำมาใช้เป็นข้อมูลความเข้มแสงที่วัดได้จากเครื่องบริการ CU BEMS ซึ่งมีการบันทึกค่าทุกๆ 3 นาที โดยมีข้อมูลบางส่วนเกิดการสูญหายขึ้น ซึ่งวิธีการจัดการข้อมูลเหล่านี้เป็นไปตามขั้นตอนดังนี้

1. เริ่มจากการเติมวันที่ที่ไม่มีการบันทึกค่าโดยการนำอนุกรมเวลาที่ได้จากเครื่องบริการมาเทียบกับอนุกรมเวลาที่สร้างขึ้นเองเพื่อเปรียบเทียบหาเวลาที่หายไป
2. จากนั้นเป็นการตรวจสอบความถูกต้องของข้อมูลว่าเป็นไปตามลักษณะเฉพาะของความเข้มแสงหรือไม่ จาก [1] ได้กล่าวถึงปริมาณความเข้มแสงที่วัดได้จากอวกาศว่ามีค่าประมาณ 1367 วัตต์ต่อตารางเมตร ซึ่งหมายความว่าความเข้มแสงเมื่อผ่านชั้นบรรยากาศเข้ามายังโลกนั้นควรมีค่าต่ำกว่า 1367 วัตต์ต่อตารางเมตร ฉะนั้นค่าที่วัดได้มากกว่านั้นจะถือว่าค่าที่วัดได้นั้นต้องทำการเติมข้อมูลเข้ามาแทนที่ในเบื้องต้นได้กำหนดให้เป็นไม่มีการบันทึกค่าที่เวลานั้นๆ
3. จากนั้นจึงเติมข้อมูลซึ่งมีทั้งหมด 2 วิธีโดยขึ้นกับช่วงของข้อมูลที่ไม่มีการบันทึกค่า
 - (a) การประมาณค่าในช่วงเป็นแบบเชิงเส้น (linear interpolation) ใช้กับช่วงของข้อมูลที่หายเป็นระยะเวลาสั้นกว่าหรือเท่ากับ 30 นาที
 - (b) การหาค่าเฉลี่ยโดยการนำข้อมูลที่เวลาเดียวกันแต่คนละวันมาหาค่าเฉลี่ย ใช้กับช่วงของข้อมูลที่หายเป็นระยะเวลายาวกว่า 30 นาที
4. เสร็จแล้วจึงนำข้อมูลมาทำเป็นข้อมูล 3 ชุดที่มีค่าความถี่ต่างกัน ก็คือ 3, 15 และ 60 นาที โดยข้อมูลชุด 15 นาทีและข้อมูลชุด 60 นาทีนั้นเป็นการหาค่าเฉลี่ยตามช่วงเวลาที่ได้กล่าวไว้ก่อนหน้านี้
5. จากนั้นจึงได้มีการนำค่าที่วัดได้ในเวลากลางคืนออก ซึ่งในรายงานนี้ได้มีการนำค่าที่วัดได้ตั้งแต่ช่วงเวลา 0.00 น.-6.00 น. และ 19.00 น.-0.00 น. เพราะพบว่า ค่าเฉลี่ยของความเข้มแสงที่วัดได้ตลอดทั้งปีในช่วงเวลาดังกล่าวมีค่าเป็น 0 หรือใกล้เคียง 0 โดยกราฟที่แสดงค่าเฉลี่ยดังกล่าวได้แสดงในรูป 3

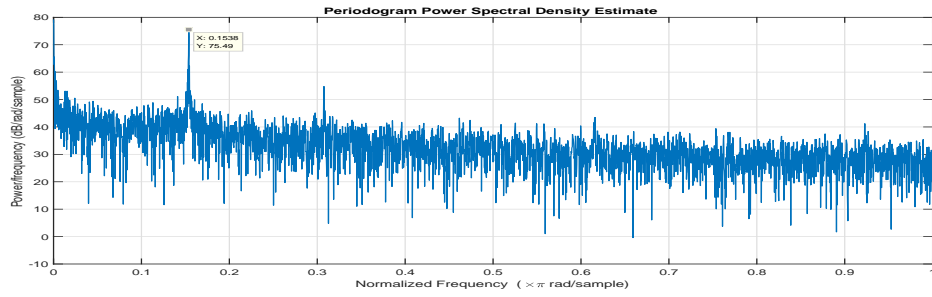


รูป 3: ค่าเฉลี่ยของข้อมูลความเข้มแสงที่เวลาต่างๆตลอดปี พ.ศ.2560

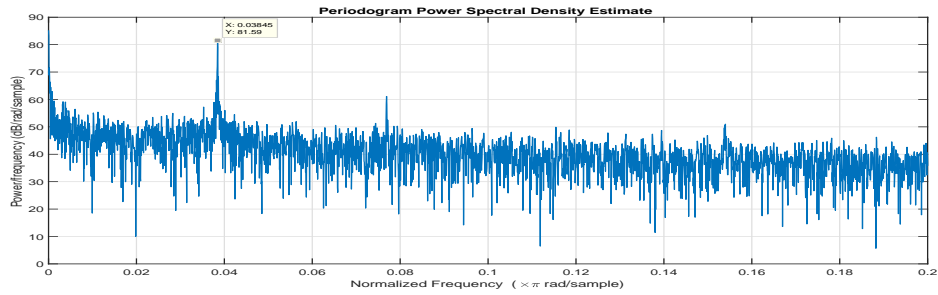
หลังจากที่มีการนำค่าที่วัดได้ในเวลากลางคืนของข้อมูลออกนั้น ข้อมูลในแต่ละชุดยังแบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลการสร้างแบบจำลองและชุดข้อมูลตรวจสอบความสมเหตุสมผล ในรายงานนี้กำหนดให้ชุดข้อมูลการสร้างแบบจำลองเป็นข้อมูลที่วัดในเดือนมกราคมถึงเดือนตุลาคม และกำหนดให้ชุดข้อมูลตรวจสอบความสมเหตุสมผลเป็นข้อมูลที่วัดในเดือนพฤศจิกายนและเดือนธันวาคม

4.2 การหาคาบของสัญญาณ

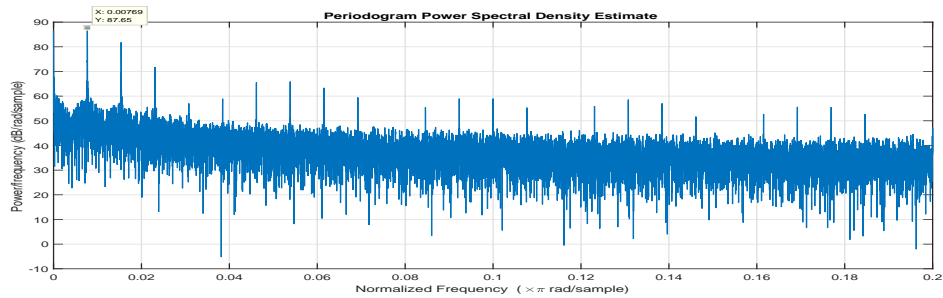
เป้าหมายของเนื้อหาในส่วนนี้คือการหาคาบ T ของแต่ละแบบจำลอง และทำให้ข้อมูลชุดนี้มีสมบัตินิ่ง ในรายงานนี้จะใช้คำสั่ง **periodogram** เพื่อหา PSD ของแต่ละข้อมูล ซึ่งผลที่ได้เป็นไปตามรูป 4 โดยข้อมูลชุด 60 นาทีจะมีคาบ $T = \frac{2\pi}{0.1538\pi} \approx 13$ ส่วนข้อมูลชุด 15 นาทีจะมีคาบ $T = \frac{2\pi}{0.03845\pi} \approx 52$ และข้อมูลชุด 3 นาทีจะมีคาบ $T = \frac{2\pi}{0.00769\pi} \approx 260$ ซึ่งค่า T ที่ได้ของข้อมูลแต่ละชุดคือจำนวนข้อมูลใน 1 วัน



(a) PSD ของข้อมูลชุด 60 นาที



(b) PSD ของข้อมูลชุด 15 นาที



(c) PSD ของข้อมูลชุด 3 นาที

รูป 4: PSD ของข้อมูลแต่ละชุด

4.3 การหาอันดับของการทำอนุพันธ์ตามฤดูกาลและการหาอันดับของการทำอนุพันธ์

เป้าหมายของการทำอนุพันธ์คือหาอันดับของการทำอนุพันธ์ตามฤดูกาล D และอันดับของการทำอนุพันธ์ d และทำให้ข้อมูลชุดนี้มีสมบัตินิ่ง หลังจากข้อมูลได้ถูกลบส่วนที่เป็นคาบออกไปแล้ว ข้อมูลชุดนี้อาจจะยังไม่มีสมบัตินิ่ง เพราะอาจจะยังเหลือส่วนที่ไม่เป็นคาบของข้อมูลโดยดูได้จาก ACF ซึ่งการหาอันดับที่เหมาะสมกับแบบจำลองเป็นไปตามตาราง 2 ทั้งนี้ รายงานนี้ให้ข้อมูลทั้ง 3 ชุดนั้นมีอันดับของการทำอนุพันธ์

ตาราง 1: ตารางวิเคราะห์ช่วงที่เหมาะสมต่อการหาอันดับของแบบจำลอง SARIMA

อันดับของสัมประสิทธิ์ที่ต้องการหา	ACF	PACF
AR(p)	Tails off ที่ lag p	หลัง lag p มีค่า ACF เป็น 0
MA(q)	หลัง lag q มีค่า ACF เป็น 0	Tails off ที่ lag q
SAR(P)	Tails off ที่ lag PT	หลัง lag PT มีค่า ACF เป็น 0
SMA(Q)	หลัง lag QT มีค่า ACF เป็น 0	Tails off ที่ lag QT
Seasonal Integrated Order(D)	มีการ cut-off ที่ lag ใด lag หนึ่งใน ACF โดยพิจารณาที่ lag $T, 2T, \dots$	ไม่ได้ใช้พิจารณา
Non-seasonal Integrated Order(d)	มีการ cut-off ที่ lag ใด lag หนึ่งใน ACF โดยพิจารณาที่ lag $1, 2, \dots$	ไม่ได้ใช้พิจารณา

ตามฤดูกาลและอันดับของการทำอนุพันธ์ที่เหมือนกันทั้ง 3 ชุดนั้นคือ $D = 1$ และ $d = 1$ ซึ่งกราฟที่แสดงการเลือกค่า d และ D อยู่ในรูป 21 22 และ 23 ซึ่งข้อมูลทั้ง 3 ชุดสามารถเลือกอันดับ D และ d เป็นได้ทั้ง 1 และ 2 แต่เนื่องด้วยความซับซ้อนของแบบจำลอง จึงเลือกอันดับ $D = 1$ และอันดับ $d = 1$

4.4 การหาอันดับของแบบจำลอง ARMA ที่เหมาะสมที่สุด

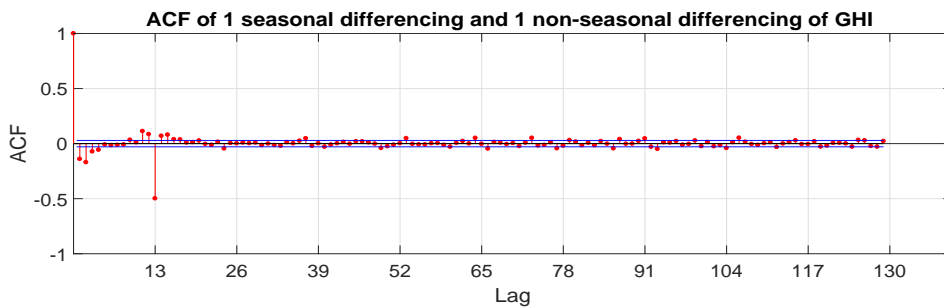
เป้าหมายของผลการทดลองส่วนนี้คือหาอันดับของแบบจำลองที่เหมาะสมที่สุดโดยหลังจากการทำ ML ก็จะหาค่า AIC กับ BIC ของแต่ละแบบจำลองว่ามีค่าเท่าไรเพื่อนำมาหาแบบจำลองที่เหมาะสมที่สุด ซึ่งจากกราฟ ACF และ PACF ของข้อมูลทั้ง 3 ชนิดได้ให้ช่วงของอันดับที่เหมาะสมกับการหาแบบจำลองทั้ง 3 ชนิดดังนี้ ซึ่งกราฟ ACF และ PACF ของข้อมูลแต่ละชุดได้แสดงดังหัวข้อย่อด้านล่าง

ตาราง 2: ช่วงของอันดับที่เหมาะสมกับแบบจำลองของข้อมูลชุดต่างๆ

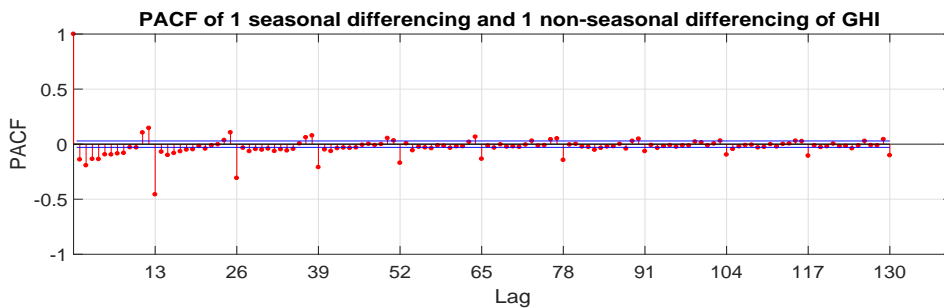
พหุนาม	ข้อมูลชุด 60 นาที	ข้อมูลชุด 15 นาที	ข้อมูลชุด 3 นาที
AR(p)	0-8	0-8	0-4
MA(q)	0-4	0-4	0-4
SAR(P)	0-4	0-6	0-2
SMA(Q)	0-1	0-1	0-1

1. ข้อมูลชุด 60 นาที

กราฟ ACF และ PACF ได้แสดงดังรูป 5 และนำช่วงของอันดับในตาราง 2 ไปหาค่า AIC และ BIC และนำ 5 แบบจำลองที่มีค่า AIC



(a) กราฟ ACF ที่ใช้ในการตัดสินใจหาช่วงของอันดับของแบบจำลอง



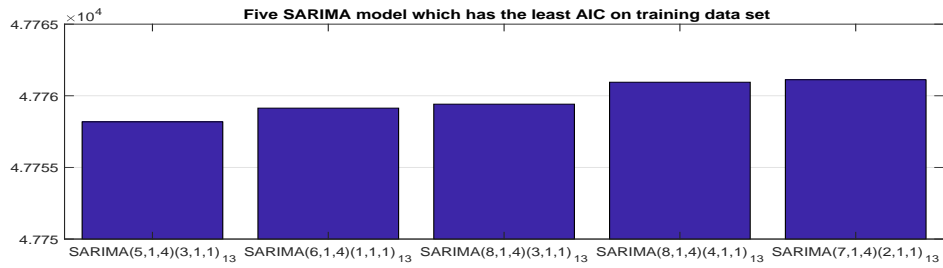
(b) กราฟ PACF ที่ใช้ในการตัดสินใจหาช่วงของอันดับของแบบจำลอง

รูป 5: ACF และ PACF เมื่อ $d = 1$ และ $D = 1$ ของข้อมูลชุด 60 นาที

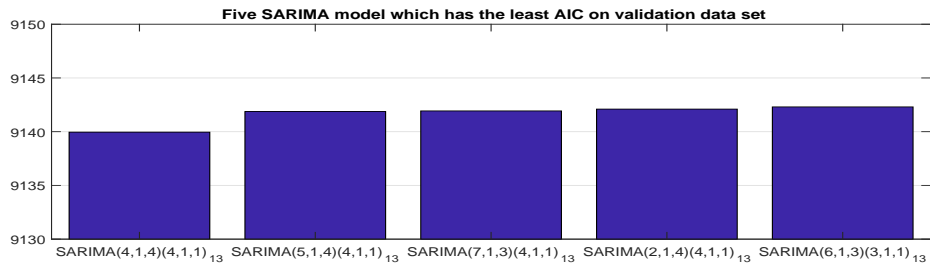
น้อยที่สุดและ 5 แบบจำลองที่มีค่า BIC น้อยที่สุด ซึ่งได้ทำทั้งข้อมูลการสร้างแบบจำลองและข้อมูลตรวจสอบความสมเหตุสมผล โดยทั้งหมดนี้ได้แสดงผลดังกราฟ 6

จากกราฟดังกล่าว พบว่าค่า AIC เลือกแบบจำลองที่มีอันดับสูง ในขณะที่ BIC มีเพียงแบบจำลอง SARIMA(1, 1, 2)(0, 1, 1)₁₃ ที่อยู่ในกราฟแสดงค่า BIC ที่น้อยที่สุดของทั้ง 2 ชุดข้อมูล และเมื่อนำแบบจำลองของข้อมูลชุดสร้างแบบจำลองมาหาค่าเฉลี่ยสัมบูรณ์ของค่าความผิดพลาด(MAE) และรากที่สองของค่าเฉลี่ยของค่าความผิดพลาดกำลังสอง(RMSE) จะได้ออกมาดังกราฟ 7 พบว่า แบบจำลอง SARIMA(1, 1, 2)(0, 1, 1)₁₃ มีทั้ง MAE และ RMSE น้อยที่สุด จึงสรุปว่า แบบจำลอง SARIMA(1, 1, 2)(0, 1, 1)₁₃ เหมาะสมกับข้อมูลชุดนี้ที่สุด โดย SARIMA(1, 1, 2)(0, 1, 1)₁₃ มีสมการเป็น

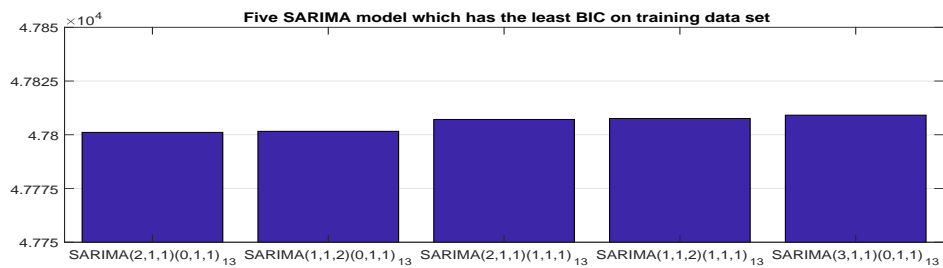
$$(1 - L^{13})(1 - 0.588L)(1 - L)I(t) = (1 - 0.933L^{13})(1 - 0.878L - 0.112L^2)e(t) \quad (33)$$



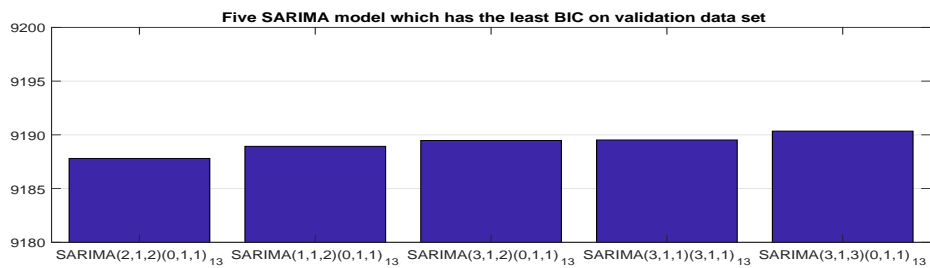
(a) AIC ของชุดข้อมูลการสร้างแบบจำลอง



(b) AIC ของชุดข้อมูลตรวจสอบความสมเหตุสมผล

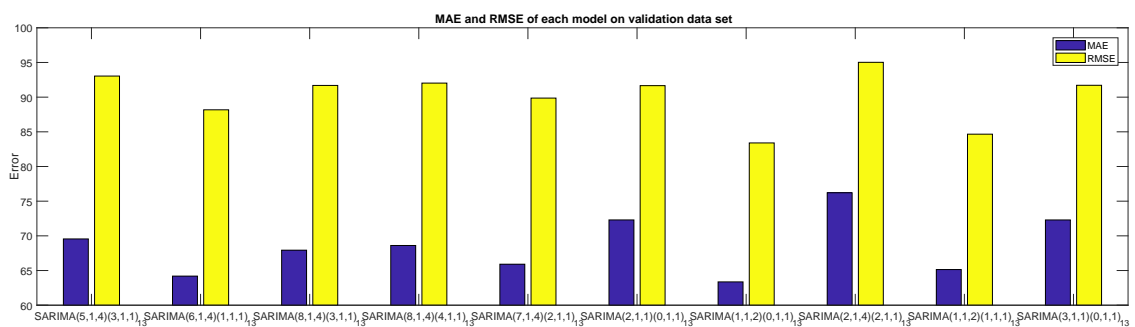


(c) BIC ของชุดข้อมูลการสร้างแบบจำลอง



(d) BIC ของชุดข้อมูลตรวจสอบความสมเหตุสมผล

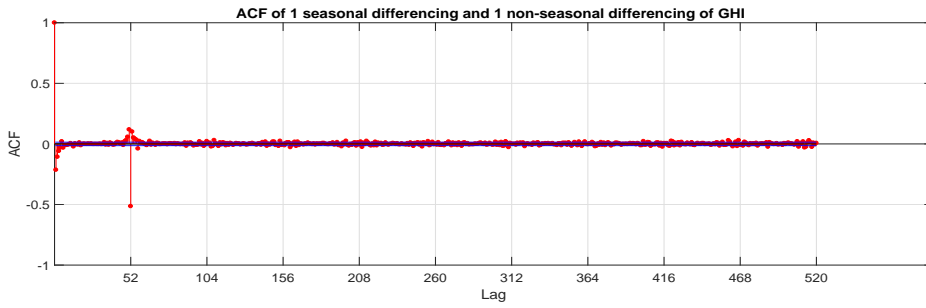
รูป 6: AIC และ BIC ของข้อมูลชุด 60 นาที



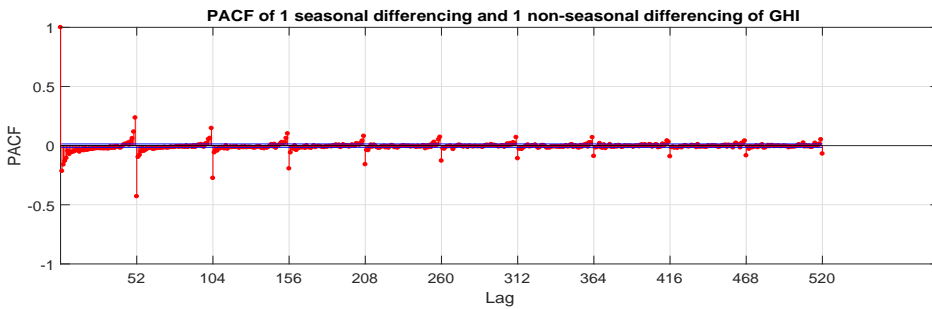
รูป 7: ค่าความผิดพลาดบนชุดข้อมูลตรวจสอบความสมเหตุสมผลของข้อมูลชุด 60 นาที

2. ข้อมูลชุด 15 นาที

กราฟ ACF และ PACF ได้แสดงดังรูป 8 และนำช่วงของอันดับในตาราง 2 ไปหาค่า AIC และ BIC และนำ 5 แบบจำลองที่มีค่า AIC



(a) กราฟ ACF ที่ใช้ในการตัดสินใจหาช่วงของอันดับของแบบจำลอง



(b) กราฟ PACF ที่ใช้ในการตัดสินใจหาช่วงของอันดับของแบบจำลอง

รูป 8: ACF และ PACF เมื่อ $d = 1$ และ $D = 1$ ของข้อมูลชุด 15 นาที

น้อยที่สุดและ 5 แบบจำลองที่มีค่า BIC น้อยที่สุด โดยทั้งหมดนี้ได้แสดงผลดังกราฟ 9

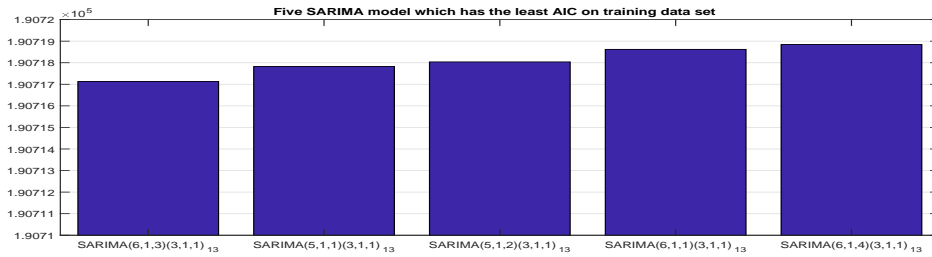
จากกราฟดังกล่าว พบว่าเมื่อนำแบบจำลองของข้อมูลชุดสร้างแบบจำลองมาหาค่าเฉลี่ยสัมบูรณ์ของค่าความผิดพลาด(MAE) และรากที่สองของค่าเฉลี่ยของค่าความผิดพลาดกำลังสอง(RMSE) จะได้ออกมาดังกราฟ 10 พบว่า แบบจำลอง SARIMA(1, 1, 4)(0, 1, 1)₅₂ มีค่า MAE น้อยที่สุด แต่แบบจำลอง SARIMA(2, 1, 2)(1, 1, 1)₅₂ มีค่า RMSE น้อยที่สุด เมื่อดูค่าลอการิทึมของฟังก์ชันภาวะน่าจะเป็นของ 2 แบบจำลองดังกล่าว จะพบว่า SARIMA(1, 1, 4)(0, 1, 1)₅₂ มีค่าลอการิทึมของฟังก์ชันภาวะน่าจะเป็นคือ -9.5359×10^4 ในขณะที่ SARIMA(2, 1, 2)(1, 1, 1)₅₂ มีค่าลอการิทึมของฟังก์ชันภาวะน่าจะเป็นคือ -9.5360×10^4 จึงสรุปว่า แบบจำลอง SARIMA(1, 1, 4)(0, 1, 1)₅₂ เหมาะสมกับข้อมูลชุดนี้ที่สุด โดย SARIMA(1, 1, 4)(0, 1, 1)₅₂ มีสมการเป็น

$$(1 - L^{52})(1 - 0.879L)(1 - L)I(t) = (1 - 0.951L^{52})(1 - 1.212L + 0.120L^2 + 0.060L^3 + 0.034L^4)e(t)$$

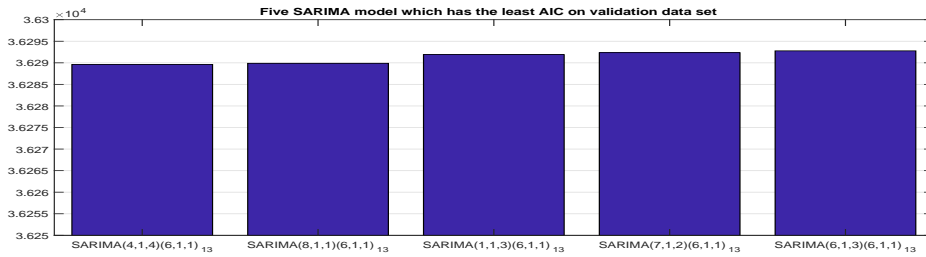
3. ข้อมูลชุด 3 นาที

ซึ่งกราฟ ACF และ PACF ได้แสดงดังรูป 11 และนำช่วงของอันดับในตาราง 2 ไปหาค่า AIC และ BIC และนำ 5 แบบจำลองที่มีค่า AIC น้อยที่สุดและ 5 แบบจำลองที่มีค่า BIC น้อยที่สุด โดยทั้งหมดนี้ได้แสดงผลดังกราฟ 12 ซึ่งพบว่าบนชุดข้อมูลการสร้างแบบจำลองนั้น แบบจำลอง SARIMA(3, 1, 4)(2, 1, 1)₂₆₀ มีทั้งค่า AIC และ BIC ที่น้อยที่สุด จึงสรุปว่าแบบจำลอง SARIMA(3, 1, 4)(2, 1, 1)₂₆₀ เหมาะสมกับข้อมูลชุดนี้ที่สุด โดย SARIMA(3, 1, 4)(2, 1, 1)₂₆₀ มีสมการเป็น

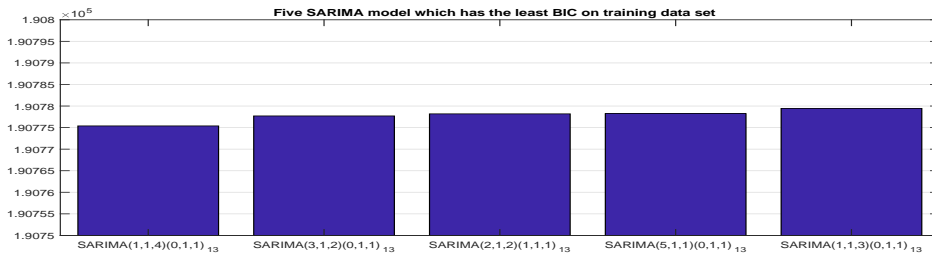
$$(1 + 0.016L^{260} - 0.016L^{520})(1 - L^{260})(1 - 0.533L - 0.719L^2 + 0.306L^3)(1 - L)I(t) = (1 - 0.947L^{260})(1 - 1.073L - 0.531L^2 + 0.682L^3 - 0.077L^4)e(t)$$



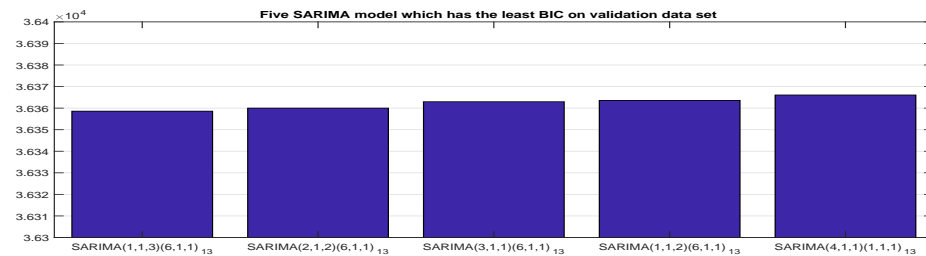
(a) AIC ของชุดข้อมูลการสร้างแบบจำลอง



(b) AIC ของชุดข้อมูลตรวจสอบความสมเหตุสมผล

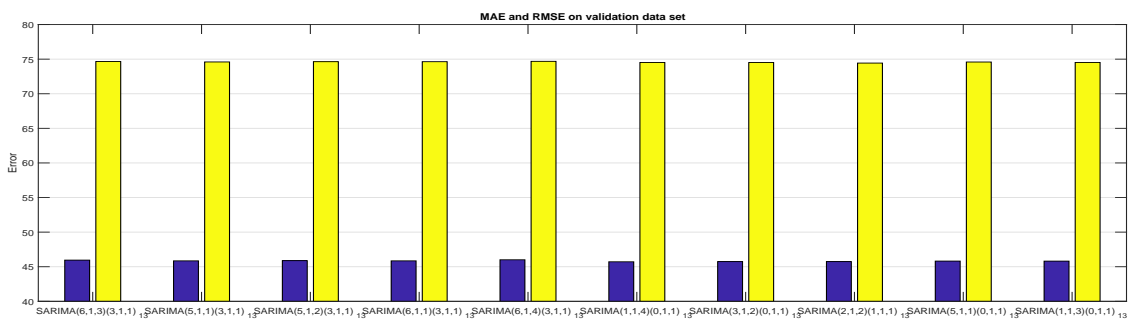


(c) BIC ของชุดข้อมูลการสร้างแบบจำลอง

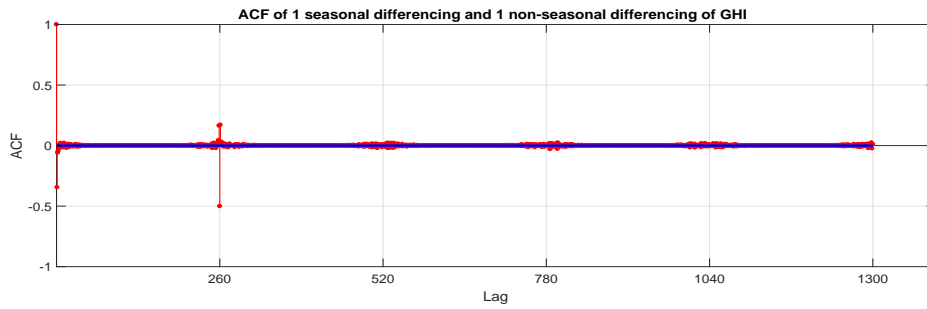


(d) BIC ของชุดข้อมูลตรวจสอบความสมเหตุสมผล

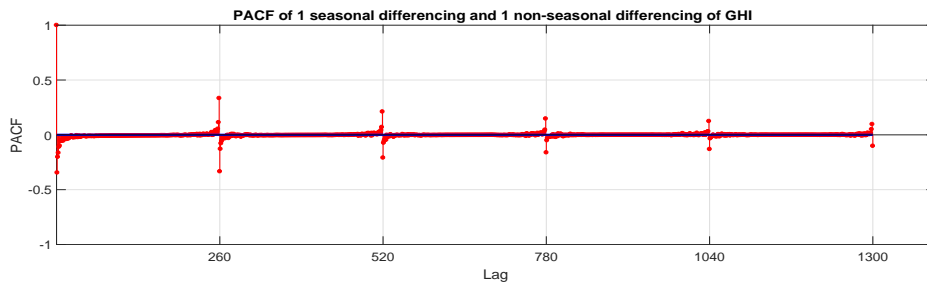
รูป 9: AIC และ BIC ของข้อมูลชุด 15 นาที



รูป 10: ค่าความผิดพลาดบนชุดข้อมูลตรวจสอบความสมเหตุสมผลของข้อมูลชุด 15 นาที



(a) กราฟ ACF ที่ใช้ในการตัดสินใจหาช่วงของอันดับของแบบจำลอง



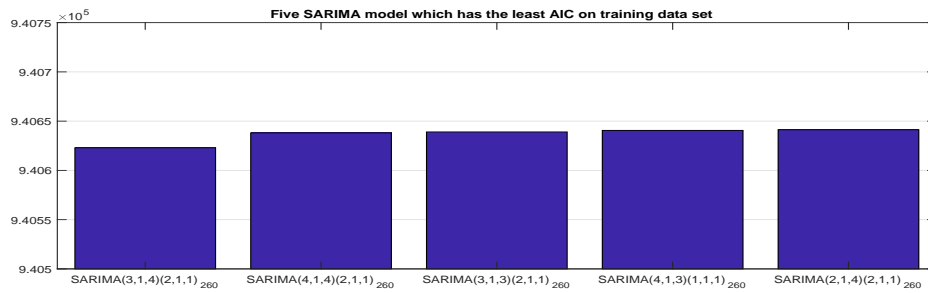
(b) กราฟ PACF ที่ใช้ในการตัดสินใจหาช่วงของอันดับของแบบจำลอง

รูป 11: ACF และ PACF เมื่อ $d = 1$ และ $D = 1$ ของข้อมูลชุด 3 นาที

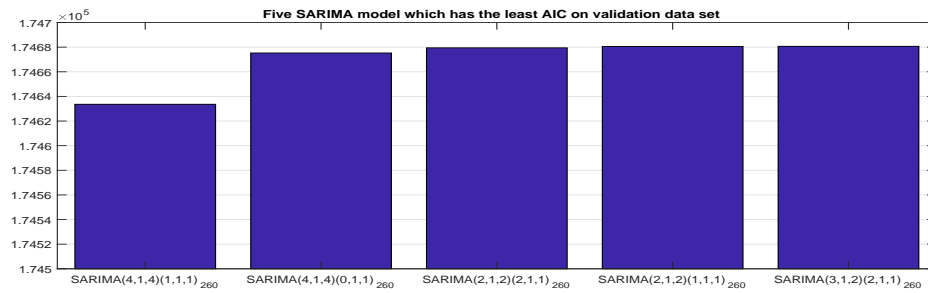
5 การวิเคราะห์และวิจารณ์ผลการพยากรณ์

เนื้อหาในส่วนนี้จะแบ่งออกเป็น 3 ส่วนคือการวิเคราะห์ผลการพยากรณ์ขั้นต่างๆบนแบบจำลองเดียวกัน ผลการพยากรณ์ค่าที่เวลาเดียวกันโดยใช้แบบจำลองที่มีอัตราการชักตัวอย่างต่างกัน และการวิเคราะห์ความแม่นยำในการพยากรณ์ค่าที่เวลาถัดไป

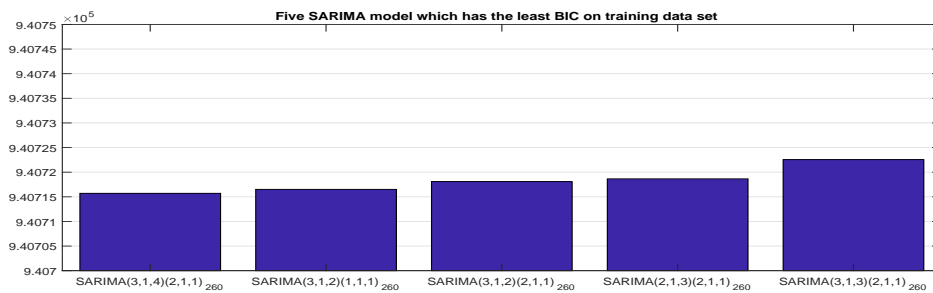
- การวิเคราะห์ความแม่นยำในการพยากรณ์ความเข้มแสงที่เวลาถัดไป 1 ชั้น
 ในหัวข้อนี้กล่าวถึงการวิเคราะห์การพยากรณ์ค่าที่เวลาถัดไป 1 ชั้นของแต่ละแบบจำลอง ซึ่งผลการพยากรณ์ค่าที่เวลาถัดไป 1 ชั้นเป็นไปตามกราฟ 20 และเมื่อวิเคราะห์ความผิดพลาดในการพยากรณ์ที่เวลาถัดไป 1 ชั้นจะพบว่าแบบจำลอง SARIMA(3, 1, 4)(2, 1, 1)₂₆₀ นั้นมีทั้ง MAE และ RMSE ที่น้อยที่สุดซึ่งแสดงในกราฟ 14 อาจเกิดจากการที่ข้อมูลเดิมมีอัตราการชักตัวอย่างเป็น 3 นาทีทำให้การพยากรณ์ค่าที่เวลาถัดไป 1 ชั้นมีค่าต่างจากข้อมูลจริงไม่มาก
- ผลการพยากรณ์ค่าที่เวลาเดียวกันโดยใช้แบบจำลองที่มีอัตราการชักตัวอย่างต่างกัน
 ในหัวข้อนี้กล่าวถึงการวิเคราะห์การพยากรณ์ค่าที่เวลาเดียวกันโดยใช้แบบจำลองที่มีอัตราการชักตัวอย่างต่างกัน ในรายงานนี้สามารถเปรียบเทียบเป็น 2 กรณีดังต่อไปนี้
 - การเปรียบเทียบแบบจำลองของชุดข้อมูล 3 นาทีกับแบบจำลองของชุดข้อมูล 15 นาที
 หัวข้อนี้เป็นการเปรียบเทียบค่าที่ได้จากการพยากรณ์ของแบบจำลองต่างๆ ซึ่งแบบจำลองของชุดข้อมูล 3 นาทีกับแบบจำลองของชุดข้อมูล 15 นาทีมีค่าที่เวลาเดียวกันทุกๆ 15 นาที ดังนั้นรายงานนี้จะพิจารณาค่าพยากรณ์ที่เวลา 6.00 น. 6.15 น. 6.30 น. 6.45 น. จนถึงเวลา 18.45 น. โดยผลของการพยากรณ์เป็นไปตามรูป 15 ซึ่งค่า MAE และ RMSE ได้แสดงผลดังรูป 16 ซึ่งสรุปได้ว่าการพยากรณ์โดยใช้แบบจำลองของข้อมูลชุด 3 นาทีแม่นยำกว่าการพยากรณ์โดยใช้อีกแบบจำลอง
 - การเปรียบเทียบแบบจำลองของชุดข้อมูล 3 นาที แบบจำลองของชุดข้อมูล 15 นาทีและแบบจำลองของชุดข้อมูล 60 นาที
 หัวข้อนี้เป็นการเปรียบเทียบค่าที่ได้จากการพยากรณ์ของแบบจำลองต่างๆ ซึ่งแบบจำลองของชุดข้อมูล 3 นาทีกับแบบจำลองของชุดข้อมูล 15 นาทีมีค่าที่เวลาเดียวกันทุกๆ 15 นาที ดังนั้นรายงานนี้จะพิจารณาค่าพยากรณ์ที่เวลา 6.00 น. 7.00 น. 8.00 น. จนถึงเวลา 18.00 น. โดยผลของการพยากรณ์เป็นไปตามรูป 17 ซึ่งค่า MAE และ RMSE ได้แสดงผลดังรูป 18 ซึ่งสรุปได้ว่าการพยากรณ์โดยใช้แบบจำลองของข้อมูลชุด 3 นาทีแม่นยำกว่าการพยากรณ์โดยใช้แบบจำลองอื่น
- การวิเคราะห์ผลการพยากรณ์ขั้นต่างๆบนแบบจำลองเดียวกัน
 ในหัวข้อนี้กล่าวถึงการวิเคราะห์ผลการพยากรณ์ขั้นต่างๆ โดยเทียบบนแบบจำลองเดียวกัน ในรายงานนี้จะเปรียบเทียบการพยากรณ์ค่าที่เวลาถัดไปเป็นจำนวน 10 ชั้น และหาค่าความผิดพลาดในการพยากรณ์ โดยมีสมมติฐานว่าค่าความผิดพลาดจะมีค่าเพิ่มขึ้นตามจำนวนชั้น และค่าที่พยากรณ์ในชั้นแรกจะแม่นยำที่สุด ซึ่งค่าความผิดพลาดที่ได้จากการพยากรณ์ค่าที่เวลาถัดไปเป็นจำนวน 10 ชั้นของแบบ



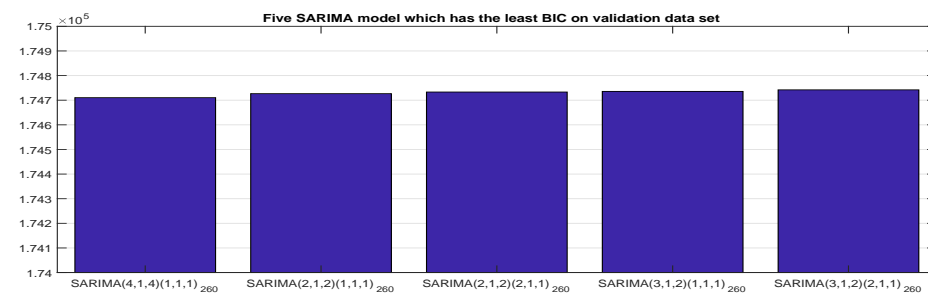
(a) AIC ของชุดข้อมูลการร้สร้างแบบจำลอง



(b) AIC ของชุดข้อมูลตรวจสอบความสมเหตุสมผล



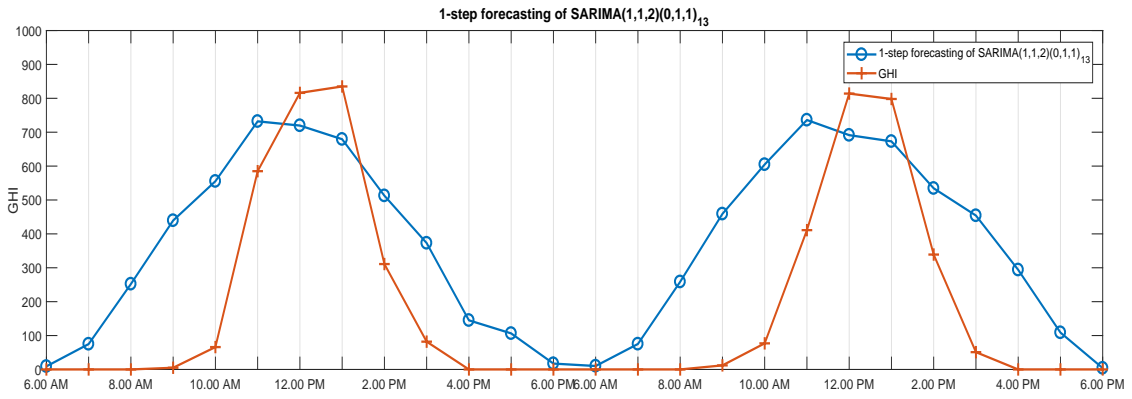
(c) BIC ของชุดข้อมูลการร้สร้างแบบจำลอง



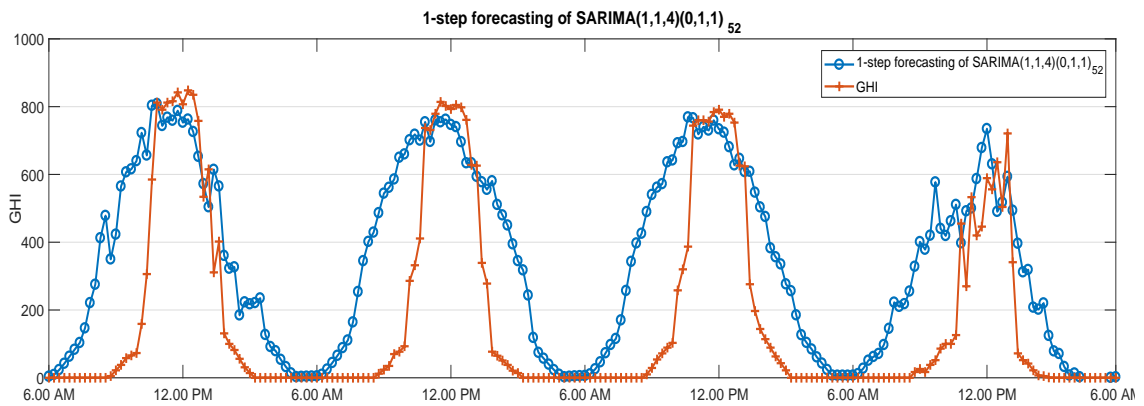
(d) BIC ของชุดข้อมูลตรวจสอบความสมเหตุสมผล

รูป 12: AIC และ BIC ของข้อมูลชุด 3 นาที

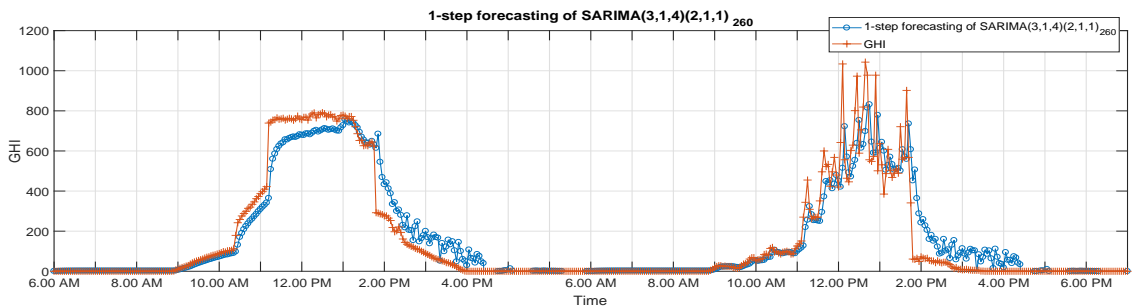
จำลองที่ได้เลือกมาได้แสดงผลในกราฟ 19 พบว่า สำหรับข้อมูลชุด 60 นาที ที่มีการพยากรณ์ค่าที่ขึ้น 9 และ 10 ดีกว่าการพยากรณ์ค่าที่ขึ้น 7 และ 8 อาจเกิดจากลักษณะเฉพาะของความเข้มแสงที่มีลักษณะเป็นคาบ ในขณะที่ข้อมูลชุด 3 และ 15 นาทีเป็นไปตามสมมติฐาน โดยตัวอย่างผลการพยากรณ์ในแต่ละชั้นของแต่ละแบบจำลองเป็นไปตามรูป 20



(a) ผลการพยากรณ์ความเข้มแสงที่เวลาลัดไป 1 ชั่วโมงของข้อมูลชุด 60 นาที

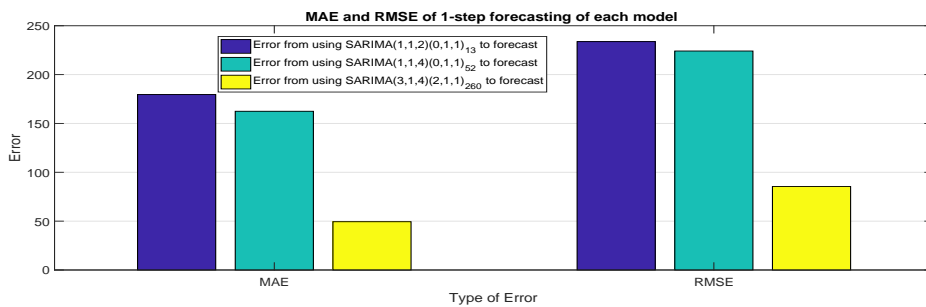


(b) ผลการพยากรณ์ความเข้มแสงที่เวลาลัดไป 1 ชั่วโมงของข้อมูลชุด 15 นาที

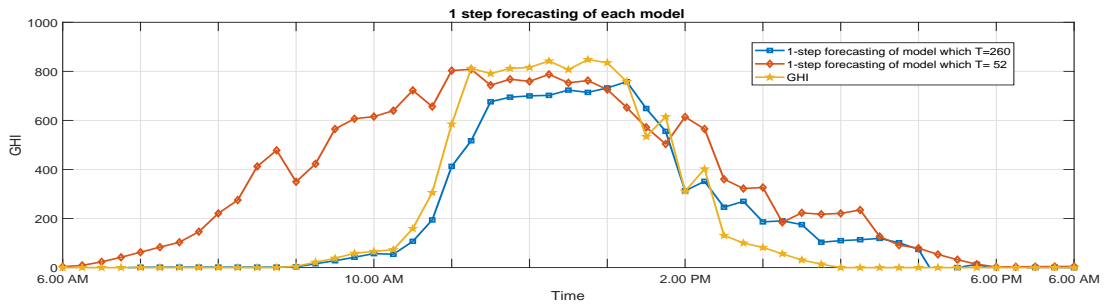


(c) ผลการพยากรณ์ความเข้มแสงที่เวลาลัดไป 1 ชั่วโมงของข้อมูลชุด 3 นาที

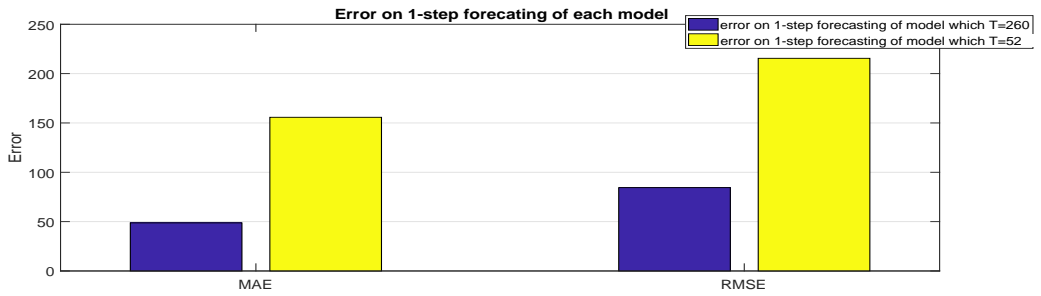
รูป 13: ผลการพยากรณ์ที่ขึ้นต่างๆของข้อมูลแต่ละชุด



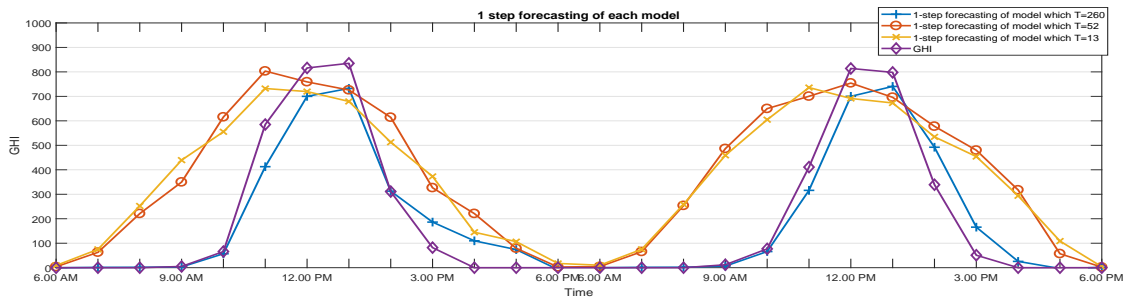
รูป 14: ค่าความผิดพลาดที่ได้จากการพยากรณ์ค่าที่เวลาลัดไป 1 ชั่วโมงโดยเทียบกับข้อมูลที่วัดได้ของแบบจำลองต่างๆ



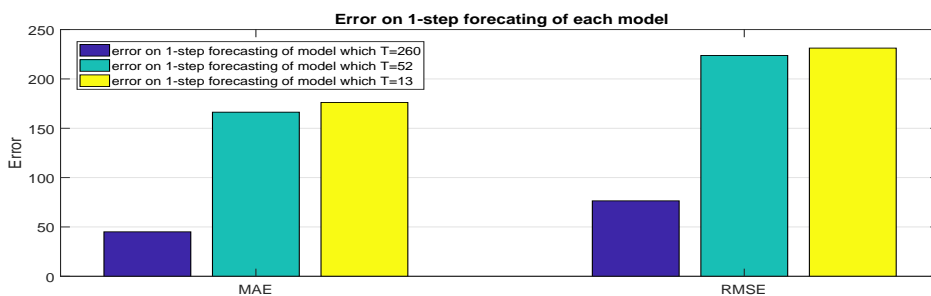
รูป 15: ค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 3 นาทีกับค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 15 นาทีเทียบกับข้อมูลจริง



รูป 16: ค่าความผิดพลาดของแบบจำลองของข้อมูลชุด 3 นาทีกับค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 15 นาทีเทียบกับข้อมูลจริง



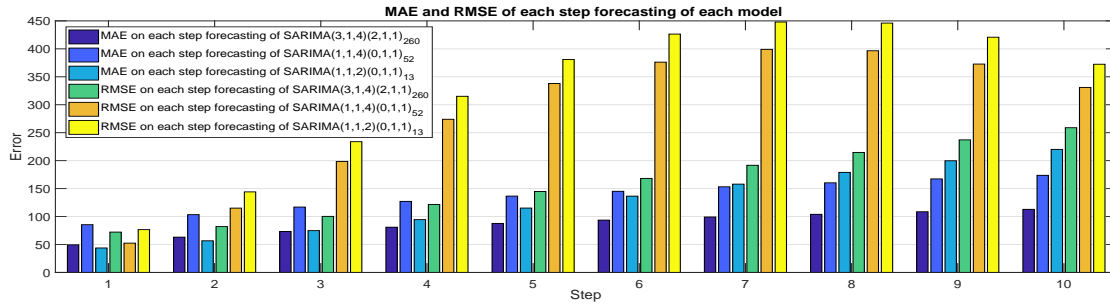
รูป 17: ค่าพยากรณ์ของแบบจำลองทั้ง 3 เทียบกับข้อมูลจริง



รูป 18: ค่าความผิดพลาดของแบบจำลองของข้อมูลชุด 3 นาทีกับค่าพยากรณ์ของแบบจำลองของข้อมูลชุด 15 นาทีเทียบกับข้อมูลจริง

6 บทสรุป

ในส่วนของการหาแบบจำลองของข้อมูลแต่ละชุดนั้น อันดับของแบบจำลองที่ได้ ในรายงานนี้จะมุ่งเน้นไปที่การพยากรณ์ค่าความเข้มแสงหลังจากที่มีการหาแบบจำลองแล้ว ซึ่งจะพบว่าการพยากรณ์ความเข้มแสงด้วยอนุกรมเวลาไม่เหมาะกับการพยากรณ์ในเวลาไกลๆ เพราะค่าความผิดพลาดจากการพยากรณ์โดยใช้แบบจำลองที่มีอัตราการซึกตัวอย่างที่น้อยนั้นมีค่าเยอะกว่าค่าความผิดพลาดจากการพยากรณ์โดยใช้แบบจำลองที่มีอัตราการซึกตัวอย่างที่น้อย ซึ่งดูได้จากกราฟทั้งหมดที่อยู่ในหัวข้อการวิเคราะห์และวิจารณ์ที่ผ่านมา



รูป 19: ค่าความผิดพลาดที่ได้จากการพยากรณ์ค่าที่เวลาถัดไปของแบบจำลองต่างๆ

เอกสารอ้างอิง

- [1] John A. Duffie and William A. Beckman. Solar engineering of thermal processes. 2013.
- [2] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [3] Nikolaos Kourentzes Juan R. Trapero and A. Martin. Short-term solar irradiation forecasting based on dynamic harmonic regression. *Energy*, 84:289–295, 2015.
- [4] W. Khovidhungij. *Signal, Systems and Control*. Chulalongkorn University Press, 2016.
- [5] V. Layanun and J. Songsiri. Solar irradiance forecasting for chulalongkorn university location using time series models. http://jitkomut.eng.chula.ac.th/group/vichaya_report.pdf, 2016.
- [6] L. Ljung. *System Identification: Theory for user*. Prentice-Hall, 1987.
- [7] Henrik Aalborg Nielsen Peder Bacher, Henrik Madsen. Online short-term solar power forecasting, 2009.
- [8] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, 1989.

7 ภาคผนวก

7.1 ฟังก์ชันอัตตสหสัมพันธ์ (Autocorrelation Function, ACF)

ฟังก์ชันอัตตสหสัมพันธ์นิยามโดย

$$ACF = \frac{R(\tau)}{R(0)} \quad (34)$$

โดยที่ฟังก์ชันความแปรปรวนของกลุ่มตัวอย่าง (sample autocovariance function) นิยามโดย

$$R(\tau) = \frac{1}{N} \sum_{t=\tau}^N y(t)y(t-\tau) \quad (35)$$

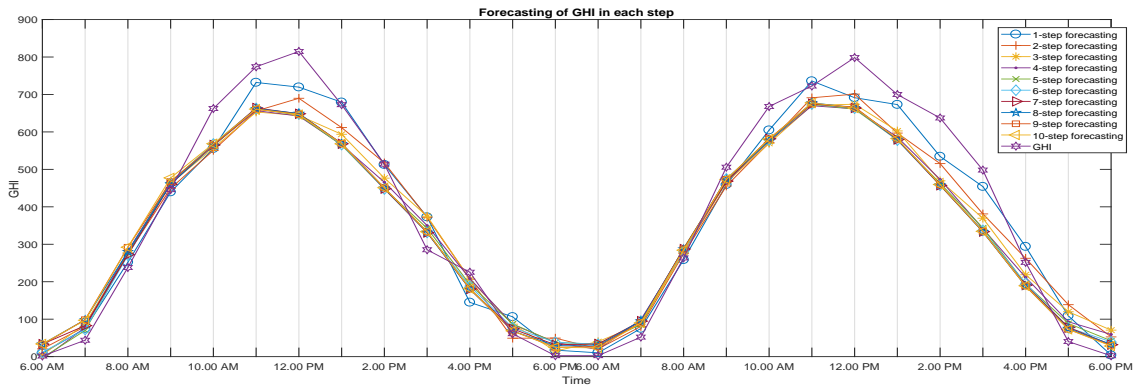
ในแบบจำลอง ARMA จะมี ACF ของ ARMA(p, q) ซึ่งแสดงในสมการ 36

$$R(\tau) - (A_1R(\tau-1) + A_2R(\tau-2) + \dots + A_pR(\tau-p)) = 0, 0 \geq \max(p, q+1) \quad (36)$$

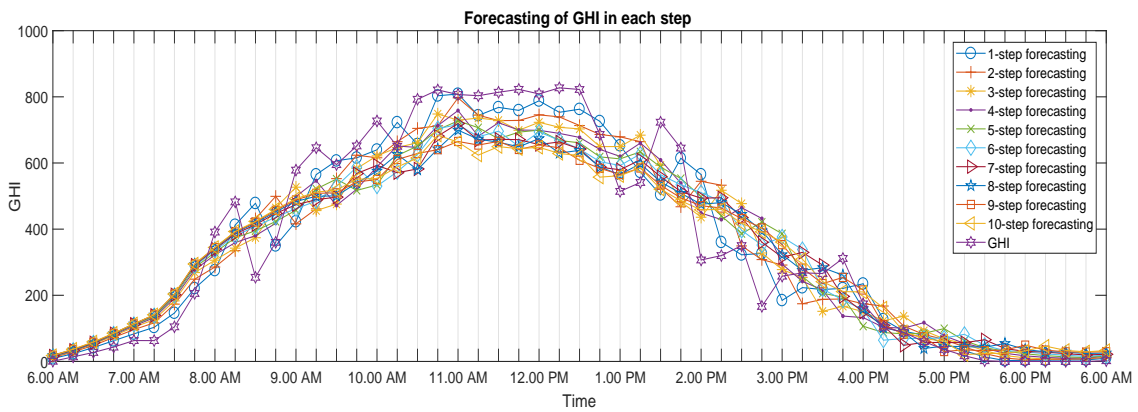
โดยมีเงื่อนไขเริ่มต้น

$$R(\tau) - \sum_{i=1}^p A_i R(\tau-i) = \sigma^2 \sum_{i=\tau}^q H(i-\tau), 0 \leq \tau \leq \max(p, q+1) \quad (37)$$

โดยที่ $H(z^{-1}) = \frac{y(z^{-1})}{v(z^{-1})} = \frac{C(z^{-1})}{A(z^{-1})}$



(a) ตัวอย่างผลการพยากรณ์ที่ขั้นต่างๆข้อมูลชุด 60 นาที

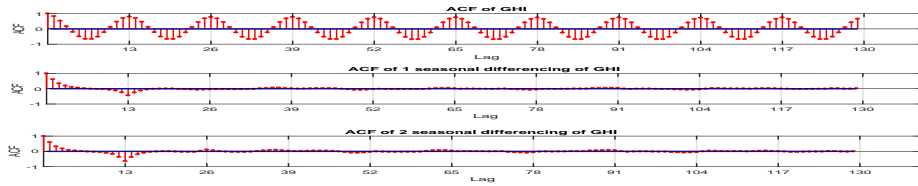


(b) ตัวอย่างผลการพยากรณ์ที่ขั้นต่างๆของข้อมูลชุด 15 นาที

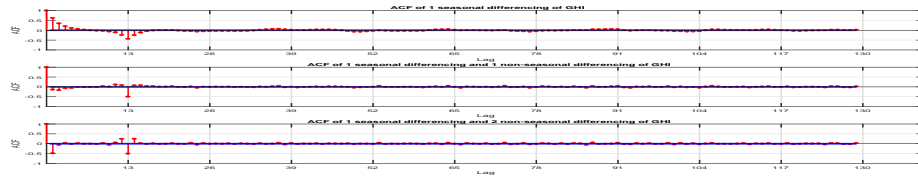
(c) ตัวอย่างผลการพยากรณ์ที่ขั้นต่างๆของข้อมูลชุด 3 นาที

รูป 20: ตัวอย่างผลการพยากรณ์ที่ขั้นต่างๆของข้อมูลแต่ละชุด

7.2 กราฟ ACF ที่ใช้ในการตัดสินใจหาค่า d และ D ของข้อมูลแต่ละชุด

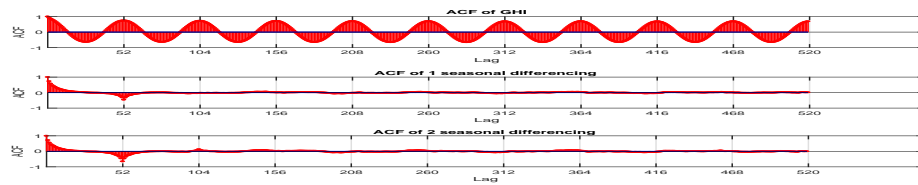


(a) กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ D

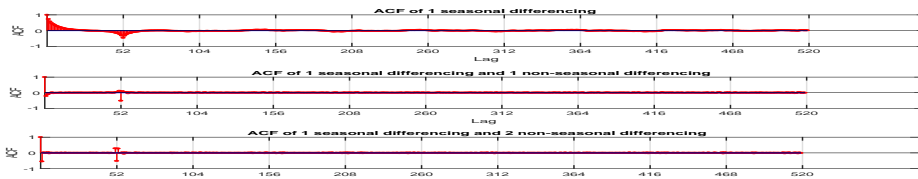


(b) กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d

รูป 21: กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d และ D ของข้อมูลชุด 60 นาที

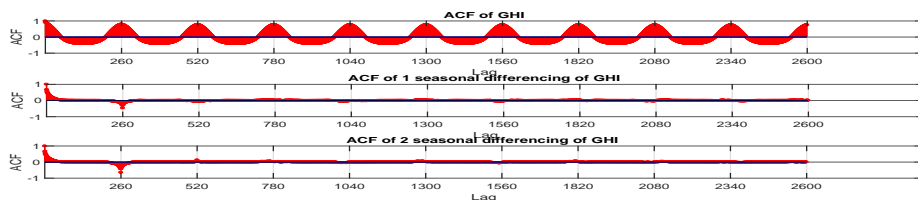


(a) กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ D

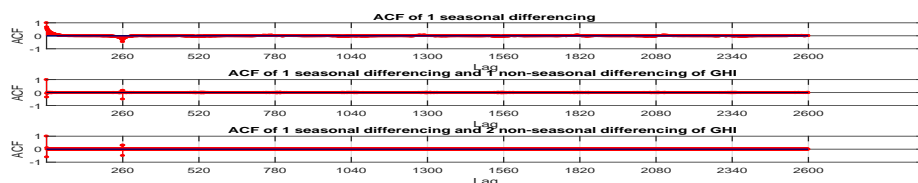


(b) กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d

รูป 22: กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d และ D ของข้อมูลชุด 15 นาที



(a) กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ D



(b) กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d

รูป 23: กราฟ ACF ที่ใช้ในการตัดสินใจหาอันดับของ d และ D ของข้อมูลชุด 3 นาที