

รายงานผลการศึกษา

Use Case :

การพยากรณ์ความเข้มข้นของ PM2.5 รายชั่วโมงและรายวันด้วยเทคนิค LSTM และการพยากรณ์ความเข้มข้นของ PM2.5 รายวันด้วยตัวแบบ VAR

โครงการออกแบบและการสร้างมาตรฐานกลางข้อมูลรัฐ
พร้อมการประยุกต์ใช้ข้อมูลขนาดใหญ่
กรณีศึกษามลพิษทางอากาศด้านฝุ่นละออง

ผู้วิจัย

อาจารย์ ดร.กรรณิการ์ หิรัญกสิ

อาจารย์ ดร.ปัญญา นันท อ้นพงษ์

คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

ตามที่มหาวิทยาลัยศิลปากรได้รับทุนอุดหนุนการวิจัยจากสำนักงานการวิจัยแห่งชาติ (วช.) ใน “โครงการออกแบบและการสร้างมาตรฐานกลางข้อมูลรัฐพร้อมการประยุกต์ใช้ข้อมูลขนาดใหญ่ กรณีศึกษามลพิษทางอากาศด้านฝุ่นละออง” ทางผู้วิจัยของโครงการดังกล่าว ได้มีการรวบรวมข้อมูลต่าง ๆ ที่เกี่ยวกับคุณภาพอากาศ โดยเฉพาะด้านฝุ่นละออง เช่น ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดไม่เกิน 10 ไมครอน (PM10) ก๊าซโอโซน (O3) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซไนโตรเจนไดออกไซด์ (NO2) และก๊าซซัลเฟอร์ไดออกไซด์ (SO2) เป็นต้น รวมถึงข้อมูลอุตุนิยมวิทยา ได้แก่ ความเร็วลม (wind speed : ws) ทิศทางลม (wind direction: wd) ค่าความชื้นสัมพัทธ์ (relative humidity : Rh) อุณหภูมิ (temperature : temp) และความกดอากาศ (Air pressure) เป็นต้น ที่รวบรวมจากสถานีตรวจวัดและรวบรวมไว้จากเว็บไซต์ต่าง ๆ อีกทั้งได้ทำการจัดการข้อมูลดังกล่าวเพื่อพร้อมใช้และเผยแพร่บนระบบ open data เพื่อให้ผู้ที่สนใจนำข้อมูลไปใช้ประโยชน์ต่อไป สามารถเข้าถึงได้จากแหล่งเดียว

ในการนี้เพื่อให้เห็นถึงประโยชน์การนำข้อมูลด้านฝุ่นละอองไปใช้ประโยชน์ทางหนึ่ง (Use Case) ผู้วิจัยจึงได้นำข้อมูลด้านฝุ่นละอองและข้อมูลคุณภาพอากาศของกรุงเทพมหานคร มาวิเคราะห์เพื่อใช้ประโยชน์ในการพยากรณ์ความเข้มข้นของ PM2.5 เนื่องจากกรุงเทพมหานครเป็นจังหวัดหนึ่งที่มีปัญหาเกี่ยวกับมลพิษทางและฝุ่นละออง อย่างไรก็ตาม เนื่องจากข้อจำกัดบางอย่าง ทำให้สถานีตรวจวัดทางอากาศของกรุงเทพมหานครบางแห่ง ไม่สามารถเก็บรวบรวมข้อมูลของคุณภาพอากาศบางอย่างได้ และช่วงระยะเวลาในการเก็บรวบรวมข้อมูลของสถานีมีความแตกต่างกัน เช่น สถานีตรวจวัดอากาศในสวนสาธารณะของกรุงเทพมหานคร จำนวน 20 สถานี มีข้อมูลเฉพาะความเข้มข้นของ PM2.5 เริ่มมีข้อมูลตั้งแต่ปี 2564 - มิ.ย. 2566 สถานีตรวจวัดทางอากาศในเขตของกรุงเทพมหานคร จำนวน 23 เขต เก็บรวบรวมข้อมูลความเข้มข้นของ PM2.5 ความเข้มข้นของ PM10 และข้อมูลทางอุตุนิยมวิทยา ตั้งแต่เดือนเมษายน 2561- มิ.ย. 2566 และสถานีตรวจวัดทางอากาศในเขตของกรุงเทพมหานครอีก 23 เขต ส่วนสถานีตรวจวัดคุณภาพอากาศแบบตู้คอนเทนเนอร์ จำนวน 4 สถานี ได้แก่ สถานีดินแดง, สถานีบริเวณสำนักงานเขตพระโขนง, สถานีบริเวณสำนักงานเขตราชบุรีบูรณะ ได้มีการเก็บรวบรวมข้อมูลของความเข้มข้นของ PM10, NOx, CO ตั้งแต่ปี 2549- มิ.ย.2566 สำหรับ PM2.5 เพิ่งติดตั้งเครื่องตรวจวัดเมื่อกุมภาพันธ์ 2563 และข้อมูลทางอุตุนิยมวิทยา (ws, wd, Rh, temp) มีการติดตั้งเครื่องตรวจวัดกรกฎาคม 2565 ส่วนสถานีตรวจวัดทางอากาศบริเวณสำนักงานเขตราชเทวี ได้ทำการตรวจวัดค่าความเข้มข้นของ PM10, NO2, CO, O3 และข้อมูลทางอุตุนิยมวิทยา เป็นรายชั่วโมง (ws, wd, Rh, temp) ตั้งแต่ปี 2553 - มิ.ย. 2566 และเริ่มตรวจวัดค่าความเข้มข้น PM2.5 เมื่อ 1 มกราคม 2557

ดังนั้นในการวิจัยนี้ ผู้วิจัยจึงได้นำข้อมูลด้านฝุ่นละอองของกรุงเทพมหานครมาใช้ประโยชน์เพื่อเป็นกรณีศึกษา 3 กรณี คือ 1) การพยากรณ์ความเข้มข้นของ PM2.5 รายชั่วโมง ใน n ชั่วโมงถัดไป จากค่าความเข้มข้นของ PM2.5 จำนวน 48 ชั่วโมงก่อนหน้า 2) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป โดยใช้ข้อมูล

ในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า ด้วยเทคนิค LSTM และ 3) การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายวัน โดยใช้ตัวแบบพยากรณ์ที่สามารถอธิบายความสัมพันธ์ของความเข้มข้นมลพิษต่าง ๆ สำหรับพยากรณ์ค่าความเข้มข้นของ PM2.5 ได้แก่ ค่าความเข้มข้นของ PM2.5 ย้อนหลังตามคาบเวลา และค่าความเข้มข้น PM10, CO, NO₂, O₃, CO₂ โดยการวิเคราะห์ข้อมูลทั้ง 2 กรณีศึกษาอาศัยวิธีการทางสถิติและวิทยาการข้อมูลเพื่อให้ได้ตัวแบบที่มีความเหมาะสมและแม่นยำในการพยากรณ์ สามารถนำไปใช้ประโยชน์ในการเฝ้าระวังและเตือนภัยต่อไป

1.2 วัตถุประสงค์ของการศึกษา

การศึกษานี้มีวัตถุประสงค์ ดังนี้

1. เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน n ชั่วโมงถัดไป โดยใช้ค่าความเข้มข้นของ PM2.5 รายชั่วโมงในอดีตจำนวน 48 ชั่วโมงก่อนหน้า ด้วยเทคนิคโครงข่ายประสาทเทียมแบบ Long-Short Term Memory (LSTM))

2. เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า ด้วยเทคนิค LSTM

3. เพื่อพยากรณ์ PM2.5 รายวัน โดยใช้ข้อมูลอนุกรมเวลาหลายชุด (multivariate time series) ได้แก่ ค่าความเข้มข้นของมลพิษทางอากาศ PM2.5, PM10, CO, NO₂, O₃ และ CO₂ รายวันในอดีต ด้วยตัวแบบเวกเตอร์การถดถอยในตัว (Vector autoregressive (VAR) model)

1.3 ขอบเขตการศึกษา

1.3.1 การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน n ชั่วโมงถัดไป และใน n วันถัดไป ด้วยเทคนิคโครงข่ายประสาทเทียมแบบ LSTM

ข้อมูลที่ใช้สำหรับการพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง คือ ค่าความเข้มข้นของ PM2.5 รายชั่วโมง จากสถานีเขตดินแดง ในช่วงวันที่ 1 ตุลาคม 2565 – 31 ตุลาคม 2565

1.3.2 การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายวัน ด้วยตัวแบบเวกเตอร์การถดถอยในตัว (VAR)

ข้อมูลที่ใช้สำหรับสร้างตัวแบบสำหรับการพยากรณ์ PM2.5 รายวัน ด้วยวิธี VAR คือ ข้อมูลคุณภาพอากาศของกรุงเทพมหานครจากกองจัดการคุณภาพอากาศและเสียง สำนักสิ่งแวดล้อม กรุงเทพมหานคร จำนวน 4 สถานีตรวจวัดคุณภาพอากาศแบบตู้คอนเทนเนอร์ ได้แก่ สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง สถานีบริเวณสำนักงานเขตราชบุรีบูรณะ และสถานีบริเวณสำนักงานเขตราชเทวี ประกอบด้วยตัวแปรต่าง ๆ ได้แก่ ค่าความเข้มข้นของมลพิษทางอากาศรายชั่วโมง 5 ตัวแปร คือ PM2.5, PM10, NO₂, CO และ O₃ โดยระยะเวลาของข้อมูลคุณภาพอากาศ เริ่มตั้งแต่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566

1.4 ประโยชน์ที่คาดว่าจะได้รับ

นำตัวแบบสำหรับการพยากรณ์ความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 2.5 จากการศึกษาไปใช้ในการพยากรณ์ความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 2.5 และมลพิษอากาศอื่นๆ ล่วงหน้าในระยะสั้นเป็นรายชั่วโมงหรือรายวันเพื่อนำไปใช้ในการจัดการและควบคุมปัญหามลพิษทางอากาศ

บทที่ 2

แนวคิดและทฤษฎีที่เกี่ยวข้องกับการศึกษา

การศึกษาปัจจัยที่ส่งผลต่อมลพิษอากาศของกรุงเทพมหานคร จำนวน 4 สถานี มีแนวคิดและทฤษฎีที่เกี่ยวข้องกับการศึกษา ดังต่อไปนี้

- 2.1 ความเข้มข้นของมลพิษอากาศและดัชนีคุณภาพอากาศ
- 2.2 แบบจำลองของการเรียนรู้ของเครื่องสำหรับข้อมูลอนุกรมเวลา
- 2.3 การทดสอบยูนิตรุตสำหรับการตรวจสอบคุณสมบัติคงที่ของอนุกรมเวลา
- 2.4 ตัวแบบเวกเตอร์การถดถอยในตัว (Vector autoregressive model)

2.1 ความเข้มข้นของมลพิษอากาศและดัชนีคุณภาพอากาศ

มลพิษทางอากาศ (Air Pollution) หมายถึง ภาวะของอากาศที่มีสารเจือปนอยู่ในปริมาณที่มากพอ และเป็นระยะเวลาานพอที่จะทำให้เกิดผลเสียต่อสุขภาพ อนามัยของมนุษย์ สัตว์ พืช และวัสดุต่างๆ สารดังกล่าวอาจเป็นธาตุหรือสารประกอบ ที่เกิดขึ้นเองตามธรรมชาติหรือเกิดจากการกระทำของมนุษย์ หรืออาจอยู่ในรูปของก๊าซ หยตของเหลว หรืออนุภาคของแข็งก็ได้ สารมลพิษทางอากาศหลักที่สำคัญคือ ฝุ่นละออง (SPM) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซซัลเฟอร์ไดออกไซด์ (SO₂) ก๊าซออกไซด์ของไนโตรเจน (NO_x) ก๊าซโอโซน (O₃)

- ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM_{2.5}) เป็นฝุ่นที่มีเส้นผ่านศูนย์กลางไม่เกิน 2.5 ไมครอน เกิดจากการเผาไหม้ทั้งจากยานพาหนะ การเผาวัสดุการเกษตร ไฟป่า และกระบวนการอุตสาหกรรม สามารถเข้าไปถึงถุงลมในปอดได้ เป็นผลทำให้เกิดโรคในระบบทางเดินหายใจ และโรคปอดต่างๆ หากได้รับในปริมาณมากหรือเป็นเวลานานจะสะสมในเนื้อเยื่อปอด ทำให้การทำงานของปอดเสื่อมประสิทธิภาพลง ทำให้หลอดลมอักเสบ มีอาการหอบหืด

- ฝุ่นละอองขนาดไม่เกิน 10 ไมครอน (PM₁₀) เป็นฝุ่นที่มีขนาดเส้นผ่านศูนย์กลางไม่เกิน 10 ไมครอน เกิดจากการเผาไหม้เชื้อเพลิง การเผาในที่โล่ง กระบวนการอุตสาหกรรม การบด การม่ หรือการทำให้เป็นผงจากการก่อสร้าง ส่งผลกระทบต่อสุขภาพเนื่องจากเมื่อหายใจเข้าไปสามารถเข้าไปสะสมในระบบทางเดินหายใจ

- ก๊าซโอโซน (O₃) เป็นก๊าซที่ไม่มีสีหรือมีสีฟ้าอ่อน มีกลิ่นฉุน ละลายน้ำได้เล็กน้อย เกิดขึ้นได้ทั้งในระดับบรรยากาศชั้นที่สูงจากผิวโลก และระดับชั้นบรรยากาศผิวโลกที่ใกล้พื้นดิน ก๊าซโอโซนที่เป็นสารมลพิษทางอากาศคือก๊าซโอโซนในชั้นบรรยากาศผิวโลก เกิดจากปฏิกิริยาระหว่างก๊าซออกไซด์ของไนโตรเจน และสารประกอบอินทรีย์ระเหยง่าย โดยมีแสงแดดเป็นตัวเร่งปฏิกิริยา มีผลกระทบต่อสุขภาพ โดยก่อให้เกิดการระคายเคืองตาและระคายเคืองต่อระบบทางเดินหายใจและเยื่อต่างๆ ความสามารถในการทำงานของปอดลดลง เหนื่อยเร็ว โดยเฉพาะในเด็ก คนชรา และคนที่เป็โรคปอดเรื้อรัง

- ก๊าซคาร์บอนมอนอกไซด์ (CO) เป็นก๊าซที่ไม่มีสี กลิ่น และรส เกิดจากการเผาไหม้ที่ไม่สมบูรณ์ของเชื้อเพลิงที่มีคาร์บอนเป็นองค์ประกอบ ก๊าซนี้สามารถสะสมอยู่ในร่างกายได้โดยจะไปรวมตัวกับฮีโมโกลบินในเม็ดเลือดแดงได้ดีกว่าออกซิเจนประมาณ 200-250 เท่า เมื่อหายใจเข้าไปทำให้ก๊าซชนิดนี้จะไปแย่งจับกับฮีโมโกลบินในเลือด เกิดเป็นคาร์บอกซีฮีโมโกลบิน (CoHb) ทำให้การลำเลียงออกซิเจนไปสู่เซลล์ต่างๆ ของร่างกายลดน้อยลง ส่งผลให้ร่างกายเกิดอาการอ่อนเพลีย และหัวใจทำงานหนักขึ้น

- ก๊าซไนโตรเจนไดออกไซด์ (NO₂) เป็นก๊าซที่ไม่มีสีและกลิ่น ละลายน้ำได้เล็กน้อย มีอยู่ทั่วไปในธรรมชาติ หรือเกิดจากการกระทำของมนุษย์ เช่น การเผาไหม้เชื้อเพลิงต่างๆ อุตสาหกรรมบางชนิด เป็นต้น ก๊าซนี้มีผลต่อระบบการมองเห็นและผู้ที่มีการหอบหืดหรือ โรคเกี่ยวกับทางเดินหายใจ

- ก๊าซซัลเฟอร์ไดออกไซด์ (SO₂) เป็นก๊าซที่ไม่มีสี หรืออาจมีสีเหลืองอ่อนๆ มีรสและกลิ่นที่ระดับความเข้มข้นสูง เกิดจากธรรมชาติและการเผาไหม้เชื้อเพลิงที่มีกำมะถัน (ซัลเฟอร์) เป็นส่วนประกอบ สามารถละลายน้ำได้ดี สามารถรวมตัวกับสารมลพิษอื่นแล้วก่อตัวเป็นอนุภาคฝุ่นขนาดเล็กได้ ก๊าซนี้มีผลกระทบโดยตรงต่อสุขภาพ ทำให้เกิดการระคายเคืองต่อเยื่อตา ผิวหนัง และระบบทางเดินหายใจ หากได้รับเป็นเวลานาน ๆ จะทำให้เป็นโรคหลอดลมอักเสบเรื้อรังได้

ดัชนีคุณภาพอากาศ หมายความว่า ดัชนีที่ใช้เป็นตัวแทนของคุณภาพอากาศ 1 ค่า ใช้เป็นตัวแทนค่าความเข้มข้นของสารมลพิษทางอากาศ 6 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM_{2.5}) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM₁₀) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซโอโซน (O₃) ก๊าซไนโตรเจนไดออกไซด์ (NO₂) และก๊าซซัลเฟอร์ไดออกไซด์ (SO₂)

ตารางที่ 2.1 เกณฑ์ของดัชนีคุณภาพอากาศของประเทศไทย

AQI	ความหมาย	สีที่ใช้	คำอธิบาย
0 - 25	คุณภาพอากาศดีมาก	ฟ้า	คุณภาพอากาศดีมาก เหมาะสำหรับกิจกรรมกลางแจ้งและการท่องเที่ยว
26 - 50	คุณภาพอากาศดี	เขียว	คุณภาพอากาศดี สามารถทำกิจกรรมกลางแจ้งและการท่องเที่ยวได้ตามปกติ
51 - 100	ปานกลาง	เหลือง	<u>ประชาชนทั่วไป</u> : สามารถทำกิจกรรมกลางแจ้งได้ตามปกติ <u>ผู้ที่ต้องดูแลสุขภาพเป็นพิเศษ</u> : หากมีอาการเบื้องต้น เช่น ไอ หายใจลำบาก ระคายเคืองตา ควรลดระยะเวลาการทำกิจกรรมกลางแจ้ง
101 - 200	เริ่มมีผลกระทบต่อสุขภาพ	ส้ม	<u>ประชาชนทั่วไป</u> : ควรเฝ้าระวังสุขภาพ ถ้ามีอาการเบื้องต้น เช่น ไอ หายใจลำบาก ระคายเคืองตา ควรลดระยะเวลาการทำกิจกรรมกลางแจ้ง หรือใช้อุปกรณ์ป้องกันตนเองหากมีความจำเป็น <u>ผู้ที่ต้องดูแลสุขภาพเป็นพิเศษ</u> : ควรลดระยะเวลาการทำกิจกรรม

AQI	ความหมาย	สีที่ใช้	คำอธิบาย
			กลางแจ้ง หรือใช้อุปกรณ์ป้องกันตนเองหากมีความจำเป็น ถ้ามีอาการทางสุขภาพ เช่น ไอ หายใจลำบาก ตาอักเสบ แน่นหน้าอก ปวดศีรษะ หัวใจเต้นไม่เป็นปกติ คลื่นไส้ อ่อนเพลีย ควรปรึกษาแพทย์
201 ขึ้นไป	คุณภาพอากาศดีมาก	แดง	ทุกคนควรหลีกเลี่ยงกิจกรรมกลางแจ้งหลีกเลี่ยงพื้นที่ที่มีมลพิษทางอากาศสูง หรือใช้อุปกรณ์ป้องกันตนเองหากมีความจำเป็น หากมีอาการทางสุขภาพควรปรึกษาแพทย์

การคำนวณดัชนีย่อยคุณภาพอากาศ

การคำนวณดัชนีย่อยคุณภาพอากาศจากค่าความเข้มข้นของสารมลพิษ จากข้อมูลผลการตรวจวัดคุณภาพอากาศ โดยแต่ละระดับของค่าความเข้มข้นของสารมลพิษทางอากาศเทียบเท่ากับค่าดัชนีคุณภาพอากาศที่ระดับต่าง ๆ ดังตารางที่ 2 และมีสูตรการคำนวณ ดังนี้

$$I = \frac{I_j - I_i}{X_j - X_i} (X - X_i) + I_i \quad (2.1)$$

ในสมการนี้

I = ค่าดัชนีย่อยคุณภาพอากาศ

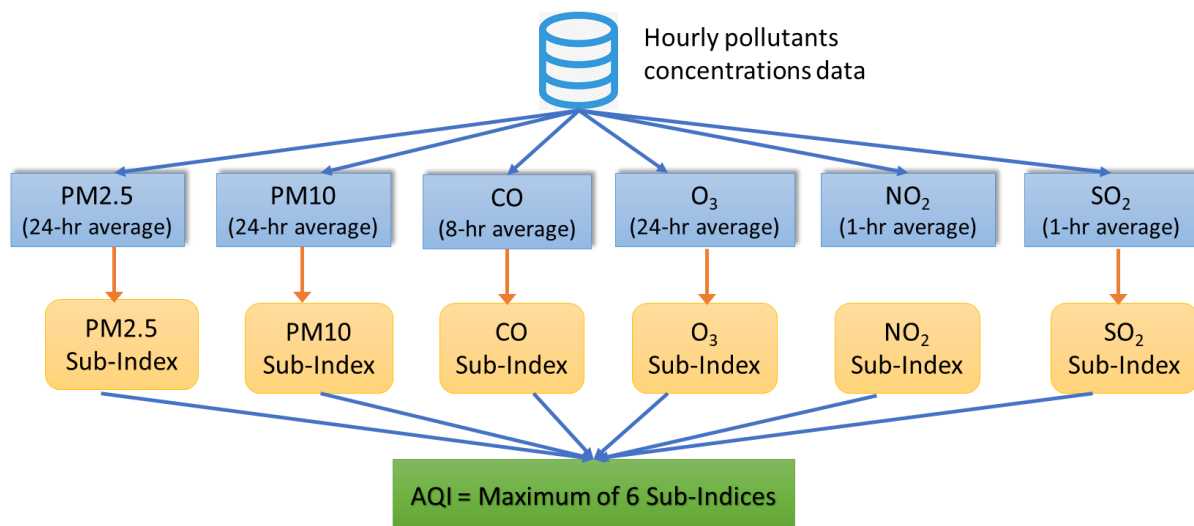
X = ความเข้มข้นของสารมลพิษทางอากาศจากการตรวจวัด

X_i, X_j = ค่าต่ำสุด, สูงสุด ของช่วงความเข้มข้นสารมลพิษที่มี X

I_i, I_j = ค่าต่ำสุด, สูงสุด ของช่วงดัชนีคุณภาพอากาศแถมเดียวกันกับช่วงความเข้มข้น X

ตารางที่ 2.2 ตารางเทียบค่าความเข้มข้นของสารมลพิษทางอากาศสำหรับคำนวณดัชนีคุณภาพอากาศ

ดัชนีคุณภาพอากาศ (I)	PM2.5 ($\mu\text{g}/\text{m}^3$)	PM10 ($\mu\text{g}/\text{m}^3$)	CO (ppm)	O ₃ (ppb)	NO ₂ (ppb)	SO ₂ (ppb)
0 - 25	0 - 15.0	0 - 50	0 - 4.4	0 - 35	0 - 60	0 - 100
26 - 50	15.1 - 25.0	51 - 80	4.5 - 6.4	36 - 50	61 - 106	101 - 200
51 - 100	25.1 - 37.5	81 - 120	6.5 - 9.0	51 - 70	107 - 170	201 - 300
101 - 200	37.6 - 75.0	121 - 180	9.1 - 30.0	71 - 120	171 - 340	301 - 400
201 ขึ้นไป	75.1 ขึ้นไป	181 ขึ้นไป	30.1 ขึ้นไป	121 ขึ้นไป	341 ขึ้นไป	401 ขึ้นไป

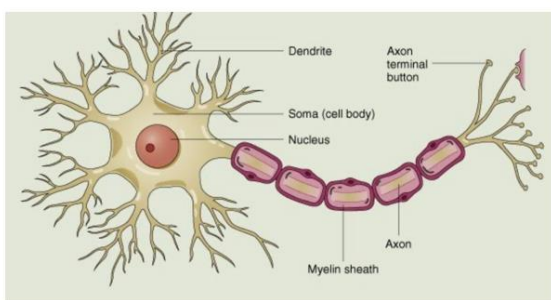


รูปที่ 2.1 ขั้นตอนการคำนวณดัชนีคุณภาพอากาศรายวัน

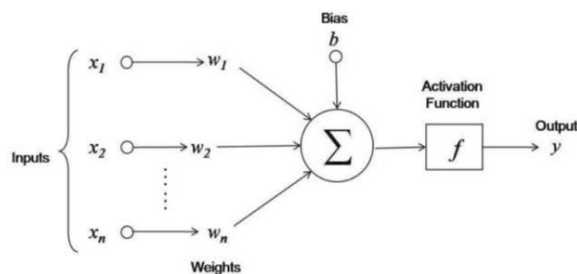
2.2 แบบจำลองของการเรียนรู้ของเครื่องสำหรับข้อมูลอนุกรมเวลา

2.2.1 โครงข่ายประสาทเทียม (Artificial Neural Network)

โครงข่ายประสาทเทียมเป็นโมเดลที่เป็นที่นิยมอย่างมากในปัจจุบัน เนื่องจากความสามารถในการทำนายที่แม่นยำในข้อมูลหลากหลายรูปแบบ เหมาะกับการนำไปใช้กับข้อมูลจำนวนมาก การออกแบบโมเดลนี้จะแบ่งออกเป็น node เล็ก ๆ โดย node แต่ละตัวถูกออกแบบมาโดยมีแรงบันดาลใจมาจากเซลล์ประสาท ดังรูปที่ 8 โดยแต่ละ node มีหลักการทำงานคือการรับข้อมูลจำนวนมากเข้าสู่ช่องทาง Input หรือเทียบได้กับ Axon ของเซลล์ประสาท ข้อมูลแต่ละตัวจะถูกคำนวณโดยการคูณน้ำหนักของ Axon แต่ละเส้น ก่อนที่กระแสสัญญาณจากทุกเส้นจะมารวมกัน และนำผลลัพธ์นี้เข้าสู่ Activation Function เพื่อคำนวณผลลัพธ์ของ node



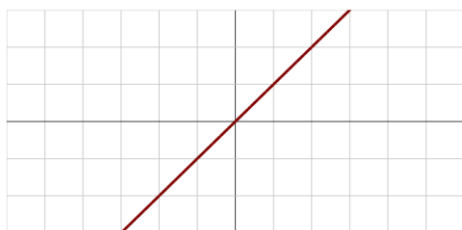
เซลล์ประสาท



Neural Node

รูปที่ 2.2 โหนดของโครงข่ายประสาทเทียม (Artificial Neural Node) ที่มีแรงบันดาลใจมาจากเซลล์ประสาท

การคำนวณ Activation Function สามารถทำได้โดยการเลือกใช้ Activation Function ให้เหมาะกับงาน โดย Activation Function มีให้เลือกหลากหลาย แต่ที่นิยมใช้ และเกี่ยวข้องกับงานนี้มีด้วยกัน 4 ฟังก์ชัน ได้แก่ Identity, Rectifier (ReLU), Sigmoid และ Hyperbolic Tangent (tanh) โดยรายละเอียดของแต่ละฟังก์ชัน เป็นไปตามรูปที่ 2.3



(a) Activation Function: Identity

$$\bullet f(x) = x$$

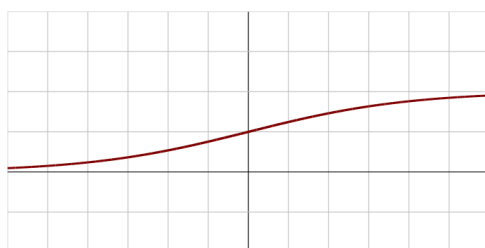
- ช่วงของคำตอบ (-inf, inf)
- คำนวณแล้วให้ผลเป็น signature โดยตรง



(b) Activation Function: Rectifier (ReLU)

$$\bullet (x)^+ = \max(0, x)$$

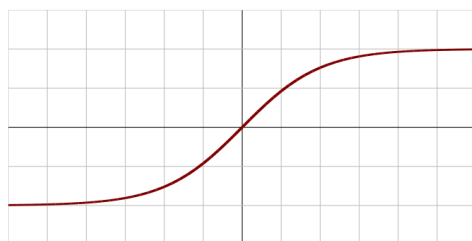
- ช่วงของคำตอบ [0, inf)
- ในทางปฏิบัติ ถูกใช้บ่อยที่สุด
- มักใช้ใน hidden layer (จะกล่าวต่อไป)



(c) Activation Function: Sigmoid

$$\bullet \sigma(x) = \frac{1}{1+e^{-x}}$$

- ช่วงของคำตอบ (0, 1)
- คำนวณแล้วให้ผลเป็นความน่าจะเป็นของคลาส 0 และ 1
- มักใช้ในงาน Classification



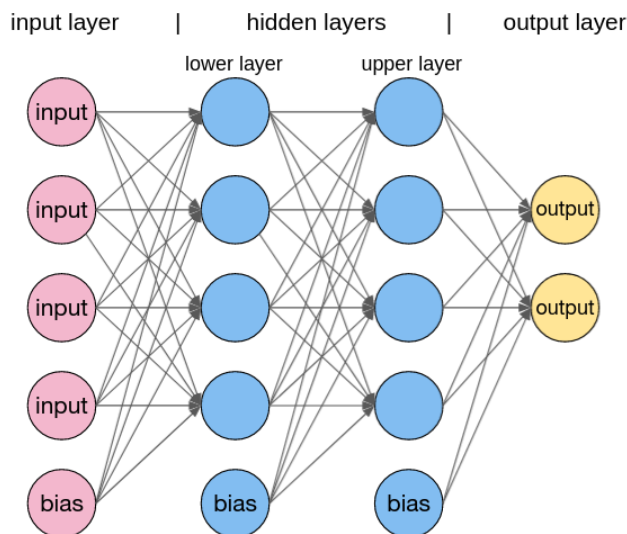
(d) Activation Function: Hyperbolic Tangent (tanh)

$$\bullet \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- ช่วงของคำตอบ (-1, 1)
- คำนวณแล้วให้ผลคล้ายความน่าจะเป็นของคลาส 0 และ 1

รูปที่ 2.3 กราฟของ Activation function ที่ใช้ในโครงข่ายประสาทเทียม

เมื่อเรานำแต่ละ neural node มาเชื่อมต่อกัน ก็จะได้โครงข่ายประสาทเทียมขนาดใหญ่ โดยเราสามารถแบ่งการเชื่อมต่อออกเป็นชั้น ๆ โดยในชั้นที่ทำหน้าที่รับข้อมูลนำเข้า เรียกว่า input layer, ชั้นที่นำข้อมูลส่งออก เรียกว่า output layer และชั้นที่อยู่ระหว่างทั้งสองชั้นดังกล่าว เรียกว่า hidden layer ดังรูปที่ 2.4



รูปที่ 2.4 การเชื่อมต่อของ neural node จำนวนมาก ก่อให้เกิดเป็น neural network ขนาดใหญ่

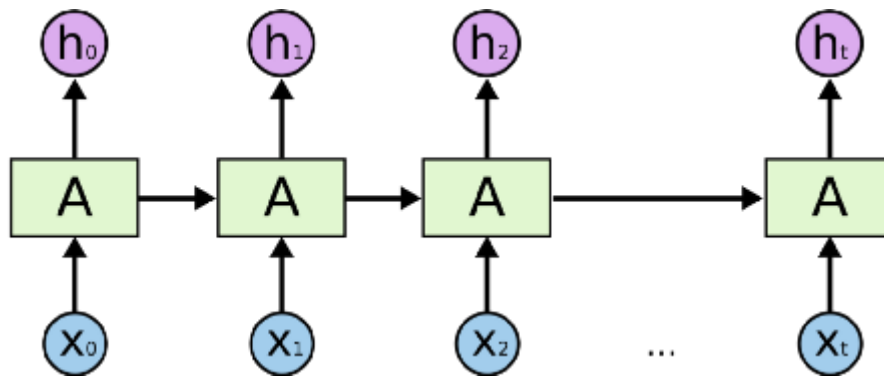
เนื่องจากปกติแล้ว เราจะมอง Artificial Neural Network เป็นฟังก์ชันที่รับ Input ประมวลผลออกมาเป็น Output โดยโมเดลจะมอง Input แต่ละตัวแยกกันชัดเจน ไม่ขึ้นต่อกัน Input ตัวอย่างเช่น หากเราใส่ข้อมูล PM2.5 ตั้งแต่เวลา 0.00 ถึง 6.00 เพื่อทำนาย PM2.5 ในเวลา 7.00 โมเดล Artificial Neural Network จะพิจารณาความเกี่ยวข้องกันระหว่างข้อมูลต่อไปนี้

- ระดับ PM2.5 เวลา 0.00 เทียบกับ 7.00
- ระดับ PM2.5 เวลา 1.00 เทียบกับ 7.00
- ระดับ PM2.5 เวลา 2.00 เทียบกับ 7.00
- ...
- ระดับ PM2.5 เวลา 6.00 เทียบกับ 7.00

โดยโมเดลจะไม่สนใจความสัมพันธ์ระหว่างพีเจอรกับพีเจอร (ระดับ PM2.5 ในเวลาที่ต่างกัน) ทำให้หากระดับ PM2.5 ค่อย ๆ เพิ่มขึ้นในแต่ละช่วงเวลา Neural Network จะไม่สนใจความสัมพันธ์เหล่านี้ แม้ว่าความสัมพันธ์ของข้อมูลดังกล่าวจะมีความสัมพันธ์กับระดับ PM2.5 ในชั่วโมงถัดไปเป็นอย่างมาก และจุดนี้เองคือจุดอ่อนของ Neural Network แบบดั้งเดิม ที่ยังทำงานกับข้อมูลชนิด Time series ได้ไม่เต็มประสิทธิภาพนัก

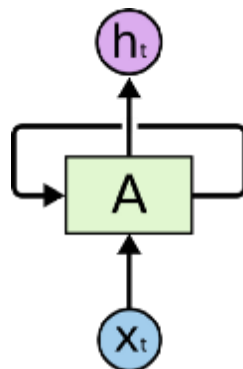
2.2.2 Recurrent Neural Network (RNN)

ใน ANN แบบดั้งเดิม อินพุตและเอาต์พุตทั้งหมดเป็นอิสระจากกัน แต่ในหลายกรณี การพิจารณาข้อมูลตามลำดับ จะทำให้เราสามารถคาดเดาสิ่งที่กำลังจะเกิดขึ้นต่อไปได้แม่นยำยิ่งขึ้น Recurrent Neural Network (RNN) คือ Artificial Neural Network แบบหนึ่งที่ย่อออกมาแก้ปัญหาสำหรับงานที่ข้อมูลมีลำดับ Sequence โดยใช้หลักการ Feed สถานะภายในของโมเดล กลับมาเป็น Input ใหม่ คู่กับ Input ปกติ เรียกว่า Hidden State, Internal State, Memory ช่วยให้โมเดลรู้จัก Pattern ของลำดับ Input Sequence ได้ การเชื่อมโยงของ RNN จะมีการเชื่อมโยงระหว่างขาออกของ neural node ในอดีต เข้ากับขาเข้าของ neural node ในปัจจุบัน และปัจจุบันไปอนาคตด้วย ทำให้ผลลัพธ์จากอดีตมีผลต่อการตัดสินใจในครั้งถัดไปด้วย โดยการเชื่อมต่อดังกล่าวสามารถแสดงเป็นแผนภาพได้ดังรูปที่ 2.5



รูปที่ 2.5 การเชื่อมต่อของ neural node ของโมเดล Recurrent Neural Network

จากรูปที่ 2.5 จะพบว่า neural node รับค่าเข้าสู่ระบบและให้ผลลัพธ์ทันที โดยที่จะส่งผลลัพธ์ส่งให้ node ถัดไปด้วย และ node ถัดไปจะทำงานซ้ำ ๆ เช่นเดียวกับ node ก่อนหน้า ทำให้เราสามารถย่อรูปของ RNN ได้ดังรูปที่ 2.6

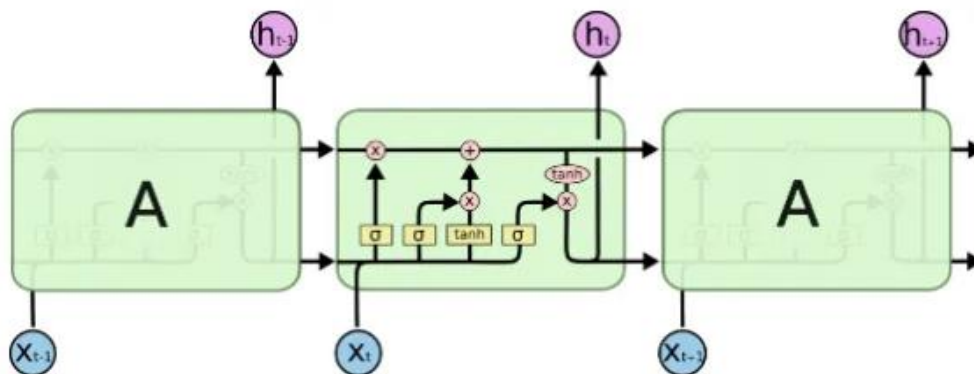


รูปที่ 2.6 การเชื่อมต่อของ neural node ของโมเดล Recurrent Neural Network หลังทำการลดรูป

ปัญหาหลักของ RNN คือไม่สามารถจัดการกับข้อมูลระยะยาวได้ เช่น หากข้อมูลจากอินพุตในช่วงแรกส่งผลกระทบต่อเวลาที่เรากำลังจะทำการทำนาย ข้อมูลที่ส่งผ่านไปในแต่ละ loop จะค่อย ๆ สูญหายเหมือนความทรงจำที่เลือนราง (ยิ่งเวลาผ่านไปนาน ยิ่งแสดงลักษณะเฉพาะไม่ชัดเจน) เราต้องการพัฒนาให้ RNN สามารถจัดเก็บและพิจารณาข้อมูลในระยะสั้นได้ (แบบที่ทำได้คืออยู่แล้ว) แต่เพิ่มความสามารถในการเรียกคืนข้อมูลระยะยาวได้เมื่อตรวจพบความเกี่ยวข้องกัน จึงเกิดเป็นโมเดล Long-Short Term Memory ในที่สุด

2.2.3 Long-Short Term Memory (LSTM)

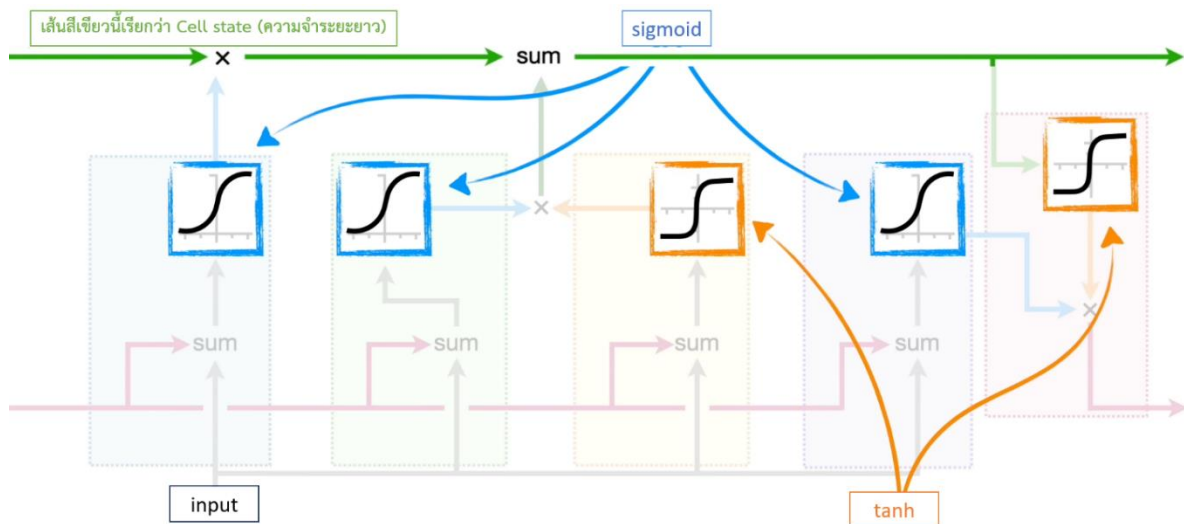
Long-Short Term Memory (LSTM) ถูกพัฒนาขึ้นเพื่อจัดการปัญหาของ RNN ที่ไม่สามารถจัดการกับข้อมูลระยะยาวได้ โดยการส่งผ่านค่าทั้งในระยะสั้น และระยะยาว ผ่าน neural node ไปเรื่อย ๆ ลักษณะโครงสร้างของโมเดล LSTM จะเป็นกล่องหลายกล่องต่อกัน แต่ละกล่องจะรับข้อมูลนำเข้าในแต่ละช่วงเวลา กล่องละ 1 ค่า จากนั้นจะทำการคำนวณผลลัพธ์ของ node นั้นโดยใช้พารามิเตอร์ 3 ตัว ได้แก่ ข้อมูลนำเข้าของช่วงเวลานั้น ข้อมูลจากหน่วยความจำระยะยาว และข้อมูลผลลัพธ์จาก node ก่อนหน้า (ระยะสั้น) ดังรูปที่ 2.7 ทำให้ผลลัพธ์ที่ได้จากโมเดลมีความสัมพันธ์กันกับลักษณะความสัมพันธ์ของข้อมูลทั้งในระยะยาว และระยะสั้นด้วย



The repeating module in an LSTM contains four interacting layers.

รูปที่ 2.7 โครงสร้างโดยรวมของเซลล์ LSTM ที่เชื่อมต่อกัน

ถ้าเราพิจารณาโครงสร้างเฉพาะเซลล์ LSTM เซลล์เดียวโดยแสดงรายละเอียดให้ชัดเจนยิ่งขึ้น จะสามารถแสดงได้ดังรูปที่ 2.8



รูปที่ 2.8 โครงสร้างโดยละเอียดภายในเซลล์ LSTM

จากรูปที่ 2.8 จะพบว่า เราสามารถแบ่งส่วนของเซลล์ LSTM ออกตามระบบ Activation Function ได้ 5 ส่วน โดยประกอบด้วย Sigmoid Function จำนวน 3 ตำแหน่ง และ tanh function จำนวน 2 ตำแหน่ง เราสามารถเขียนกรวยย่อย ๆ เพื่ออธิบายการทำงานของแต่ละฟังก์ชันได้ดังรูปที่ 16

จากกรอบแต่ละกรอบในรูปที่ 16 เมื่อนำมาประกอบกับกรอบอื่น ๆ จึงจะมีความสามารถในการทำหน้าที่เฉพาะตัว หากจับกลุ่มกรอบเหล่านี้อีกครั้ง เราจะสามารถนำมาจัดกลุ่มตามการทำงานได้เป็น 3 state ได้แก่

- การปรับอัตราส่วนของการจดจำ (Forget Gate) หรือกรอบซ้ายสุด เป็นกรอบที่ใช้คำนวณสัดส่วนของข้อมูลระยะยาวที่จะถูกปล่อยผ่าน Cell state ไปยังส่วนถัดไป โดยสัดส่วนที่ได้จะมาจากการคำนวณระหว่าง input และข้อมูลผลลัพธ์จากเซลล์ก่อนหน้า ตามทฤษฎีแล้ว ถ้าข้อมูลระยะยาวก่อนหน้านี้มีความเกี่ยวข้องกับผลลัพธ์น้อย สัดส่วนที่ได้นี้มีค่าน้อย

- อพเตทการจดจำระยะยาว (Input Gate) ประกอบด้วยกรอบที่ 2 และ 3 ซึ่งเป็นการประกอบกัันระหว่าง activation function 2 ตัว ผลลัพธ์ที่ได้จาก activation function ทั้ง 2 ตัวนี้จะถูกนำมารวมกัน และส่งไปอพเตทค่าที่กำลังจะผ่าน cell state ในส่วนนี้จะใช้ข้อมูลที่ได้จากข้อมูล input และผลลัพธ์จากเซลล์ก่อนหน้าในการพิจารณา

- อพเตทการจดจำระยะสั้น (Output Gate) ประกอบด้วยกรอบที่ 4 และ 5 (2 กรอบสุดท้าย) ซึ่งยังคงเป็นการประกอบกัันระหว่าง activation function ทั้งสองตัวเช่นเดียวกับ input gate แต่ในส่วนนี้ sigmoid function จะถูกคำนวณจากข้อมูล input ร่วมกับข้อมูลผลลัพธ์ที่ส่งผ่านมาจากเซลล์ LSTM ก่อนหน้า และ tanh function จะถูกคำนวณจากข้อมูล cell state ผลลัพธ์จากทั้งสองส่วนจะถูกนำมาคูณกัน ก่อนที่จะใช้เป็นคำตอบ และทำสำเนาไปยังเซลล์ถัดไป

2.3 การทดสอบยูนิตรูทสำหรับการตรวจสอบคุณสมบัติคงที่ของอนุกรมเวลา (unit root test for stationary time series)

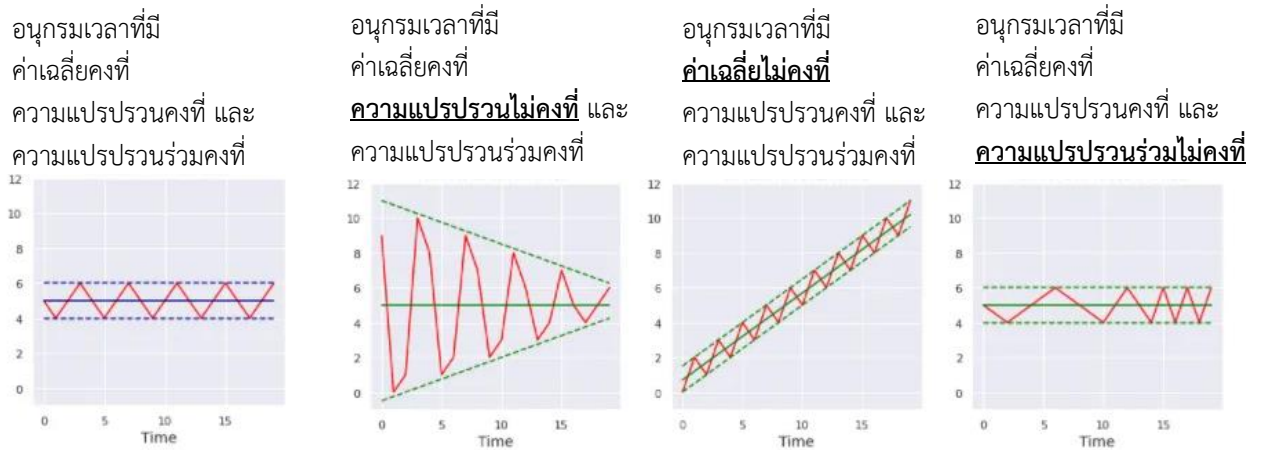
ข้อมูลอนุกรมเวลา มีลักษณะเป็นตัวแปรสุ่ม (Random Variables) และเมื่อนำมาเรียงกันตามลำดับเวลาที่เกิดขึ้น เรียกว่า กระบวนการสโตแคสติก (stochastic process) ซึ่งกระบวนการสโตแคสติก $\{Y_t; t = 1, 2, \dots, n\}$ จะเป็นกระบวนการคงที่ (stationary process) โดยแท้ ก็ต่อเมื่อการแจกแจงร่วมของตัวแปรสุ่ม ณ เวลาจุดต่าง ๆ กันในช่วง ๆ หนึ่งที่มีค่าวัด $y_{1+h}, y_{2+h}, \dots, y_{n+h}$ เป็นอิสระกับ h สำหรับทุก ๆ n กระบวนการนี้จะเป็นกระบวนการคงที่ในนัยที่กว้างขึ้น ถ้าค่าความคาดหวัง (expected value) ของ Y_t คงที่ ค่าความแปรปรวน (Variance) ของ Y_t คงที่ และค่าความแปรปรวนร่วม (Covariance) ของ Y_t และ Y_{t+h} จะไม่ขึ้นอยู่กับเวลา t แต่จะขึ้นอยู่กับช่วงเวลาที่ช้ากว่ากัน h สำหรับทุกค่า t นั่นคือ (Maddala and Kim, 1988)

$$E(Y_t) = E(Y_{t+h}) = \mu$$

$$Var(Y_t) = E(Y_t - \mu)^2 = E(Y_{t+h} - \mu)^2 = \sigma^2$$

$$E\{(Y_t - \mu)(Y_{t+h} - \mu)\} = \gamma_h$$

อนุกรมเวลาที่มีคุณสมบัติข้อใดข้อหนึ่งใน 3 ข้อนี้ แสดงว่า มีคุณสมบัติของความไม่คงที่ (Nonstationary) เรียกว่า อนุกรมเวลาแบบไม่คงที่ หรือกล่าวอีกนัยหนึ่ง คือ อนุกรมเวลานั้นมียูนิตรูท (nit root) ดังแสดงในรูปที่ 2.9



รูปที่ 2.9 อนุกรมเวลาที่มีคุณสมบัติคงที่และไม่คงที่

อนุกรมเวลา AR(1) หรือ ตัวแบบการถดถอยในตัวอันดับที่ 1 (autoregressive model at lag 1) ดังสมการ (2.2)

$$y_t = \phi y_{t-1} + e_t \tag{2.2}$$

ถ้ารากของสมการลักษณะผกผัน (inverse characteristic function) $1 - \phi z = 0$ ซึ่งก็คือ $z = 1/\phi$ มีค่าเท่ากับ 1 แล้ว จะเรียกว่า y_t มียูนิตรูท

จากนิยามข้างต้นจะได้ว่า อนุกรมเวลา AR(1) $y_t = \alpha + \phi y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim WN(0, \sigma^2)$ มียูนิตรูท ถ้า $\phi = 1$ กระบวนการยูนิตรูทของอนุกรมเวลานี้ไม่คงที่ (non-stationary) และเรียกว่าตัวแบบเดินสุ่ม (Random walk model) โดยที่ สัมประสิทธิ์ α เรียกว่า drift และถ้า y_t มียูนิตรูท เราจะได้ว่า Δy_t เป็นอนุกรมเวลาคงที่ และ y_t เป็น Difference stationary process

การทดสอบว่าอนุกรมเวลาชุดใด ๆ เป็นอนุกรมเวลามียูนิทรูทหรือไม่ เราจะใช้การทดสอบยูนิทรูทแบบถดถอยในตัวเป็นสมมติฐานว่าง และเรียกสั้น ๆ ว่า การทดสอบยูนิทรูท (unit root test) ซึ่งการทดสอบยูนิทรูทมีหลายวิธี วิธีหนึ่งที่เป็นที่นิยม คือ การทดสอบดิกกี-ฟลูเลอร์ (Dickey-fuller test หรือ DF test) เรียกว่า การทดสอบ DF โดยดิกกีและฟลูเลอร์ (Dickey and Fuller, 1979) ได้เสนอวิธีการทดสอบนี้บนพื้นฐานของความคลาดเคลื่อนสุ่ม e_t ในสมการ (2.2) มีการแจกแจงแบบเดียวกันและอิสระ (independently and identically distributed) และต่อมาซาอิดและดิกกี (Said and Dickey, 1984) ได้เสนอการทดสอบดิกกี-ฟลูเลอร์เสริม (augmented dickey-fuller test) เรียกว่า การทดสอบ ADF รายละเอียดของวิธีการทดสอบทั้งสองวิธี มีดังต่อไปนี้

การทดสอบยูนิทรูทด้วยวิธีการทดสอบ DF (unit root test with DF test)

การทดสอบยูนิทรูทด้วยวิธีการทดสอบ DF เป็นการทดสอบอนุกรมเวลาชุดใด ๆ ชุดหนึ่งที่อยู่ในรูปแบบของตัวแบบการถดถอยในตัวที่ความล่าช้า 1 คาบเวลา (autoregressive model at lag 1) หรือ AR(1) ดังสมการ (2.3)

$$y_t = \phi y_{t-1} + e_t \quad (2.3)$$

หรือ เขียนอยู่ในอีกรูปแบบหนึ่ง คือ ผลต่างของอนุกรมเวลา Δy_t โดยนำ y_{t-1} ลบทั้งสองข้างของสมการ (2.3) จะได้ว่า

$$\Delta y_t = (\phi - 1)y_{t-1} = \gamma y_{t-1} + e_t \quad (2.4)$$

เมื่อ $\gamma = \phi - 1$ และ e_t เป็นตัวแปรสุ่มที่มีการแจกแจงปกติที่ค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับ σ^2 เขียนแทนด้วย $e_t \sim i.i.d. N(0, \sigma^2)$

จากสมการ (2.4) การทดสอบยูนิทรูท มีสมมติฐานของการทดสอบ (Hypothesis) คือ

$$H_0: \gamma = 0 \text{ (unit root or non-stationary)}$$

$$H_1: \gamma \neq 0$$

มีตัวสถิติทดสอบ คือ

$$t_{\gamma=0} = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (2.5)$$

โดยที่ $\hat{\gamma}$ คือ ตัวประมาณกำลังสองน้อยที่สุด (least square estimate) และ $SE(\hat{\gamma})$ คือ ค่าคลาดเคลื่อนมาตรฐาน (standard error of estimate) ของตัวประมาณ $\hat{\gamma}$

ฟิลลิป (Phillips, 1987) ได้แสดงให้เห็นว่า ภายใต้สมมติฐานว่างของการทดสอบยูนิทรูท $H_0: \gamma = 0$

$$t_{\gamma=0} \rightarrow \frac{\int_0^1 W(r) dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}} \quad (2.6)$$

โดยที่ $W(r)$ คือ การเคลื่อนที่บราวเนียน (Brownian motion) จากกระบวนการไวเนอร์ (Wiener process) ที่นิยามอยู่บนช่วง (0,1)

เราจะได้การแจกแจงลู่เข้า (asymptotic distribution) ของตัวสถิติทดสอบ $t_{\phi=1}$ กำหนดโดยสมการ (2.6) เรียกว่า การแจกแจงทีของดิกกี-ฟลูเลอร์ (Dickey -fuller t-distribution) และกระบวนการทดสอบสมมติฐานว่าง $H_0: \gamma = 0$ ด้วยสถิติทดสอบ $t_{\gamma=0}$ เรียกว่า การทดสอบดิกกี-ฟลูเลอร์ (Dickey - fuller (DF) test)

การแจกแจงลู่เข้า (asymptotic distribution) ของตัวสถิติทดสอบ $t_{\gamma=0}$ ภายใต้ $\phi = 1$ ไม่ได้มีการแจกแจงที่แบบมาตรฐาน และค่าวิกฤตไม่สามารถหาได้จากตารางค่าสถิติที่เป็นมาตรฐาน ฟลูเลอร์ (1979) ได้เสนอค่าวิกฤตสำหรับสถิติทดสอบของการทดสอบ DF ดังแสดงในตารางที่ 2.1 เราจะปฏิเสธสมมติฐานว่าง

$H_0: \gamma = 0$ ถ้าค่าสถิติทดสอบ $t_{\gamma=0}$ มีค่ามากกว่าบริเวณวิกฤต ถ้าผลการทดสอบ คือ ไม่สามารถปฏิเสธสมมติฐาน H_0 ได้ จะได้ว่า เป็นอนุกรมเวลาแบบคงที่

ตารางที่ 2.3 แสดงค่าวิกฤตสำหรับการทดสอบดิคกีและฟูลเลอร์ ของข้อมูลอนุกรมเวลาที่อยู่ในรูปของการถดถอยสำหรับตัวแบบ AR(1) 3 กรณี คือ

1) Random walk without a drift : $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$ สมมติฐานของการทดสอบยูนิทรูท คือ

(I) $H_0: \gamma = 0$ vs $H_1: \gamma \neq 0$

2) Random walk with a drift : $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$ สมมติฐานของการทดสอบยูนิทรูท คือ

(I) $H_0: \gamma = 0$ vs $H_1: \gamma \neq 0$

(II) $H_0: \gamma = 0$ and $a_0 = 0$ vs $H_1: \gamma \neq 0$ or $a_0 \neq 0$

3) Random walk with a drift and time trend : $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$

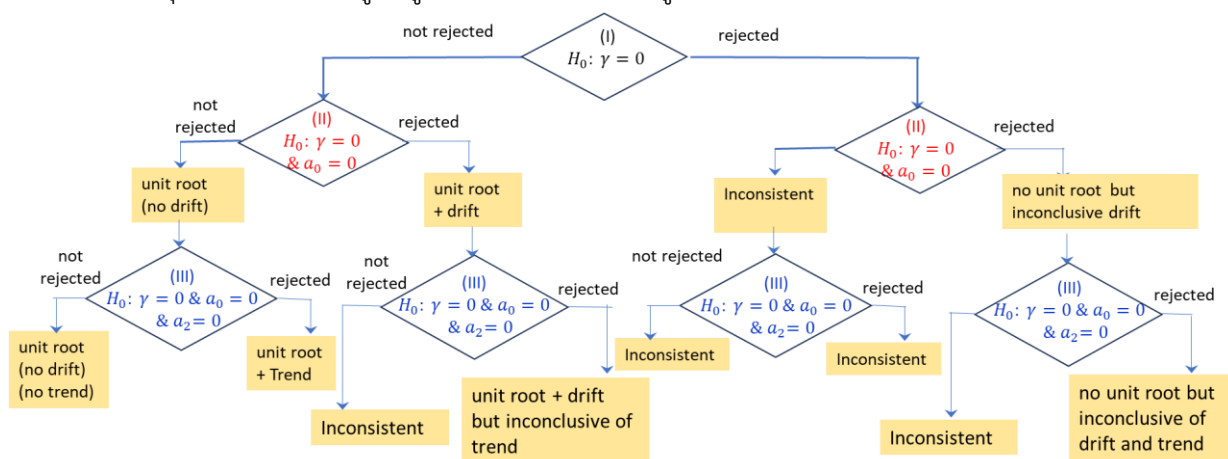
สมมติฐานของการทดสอบยูนิทรูท คือ

(I) $H_0: \gamma = 0$ vs $H_1: \gamma \neq 0$

(II) $H_0: \gamma = 0$ and $a_0 = 0$ vs $H_1: \gamma \neq 0$ or $a_0 \neq 0$

(III) $H_0: \gamma = 0$ & $a_0 = 0$ & $a_2 = 0$ vs $H_1: \gamma \neq 0$ or $a_0 \neq 0$ or $a_2 = 0$

การสรุปผลการทดสอบยูนิทรูททั้ง 3 กรณี แสดงดังรูปที่ 2.10



รูป 2.10 การสรุปผลการทดสอบยูนิทรูท

ตารางที่ 2.3 ค่าวิกฤตสำหรับการทดสอบดิคกีและฟูลเลอร์ (Maddala and Kim, 1988)

Sample size	K-test		T-test		F-test (a)		F-test (b)	
	1%	5%	1%	5%	1%	5%	1%	5%
Random walk without a drift : $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$								
25	-11.9	-7.7	-2.62	-1.95				
100	-13.3	7.9	-2.60	-1.95				
250	-13.6	-8.0	-2.58	-1.95				
500	-13.7	-8.0	-2.58	-1.95				
∞	-13.8	-8.1	-2.58	-1.95				

Random walk with a drift : $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$								
25	-17.2	-12.5	-3.75	-3.00	7.88	5.18		
50	-18.9	-13.3	-3.58	-2.93	7.06	4.86		
100	-19.8	-13.7	-3.51	-2.89	6.70	4.71		
250	-20.3	-14.0	-3.46	-2.88	6.52	4.63		
500	-20.5	-14.0	-3.44	-2.87	6.47	4.61		
∞	-20.7	-14.1	-3.43	-2.86	6.43	4.59		
Random walk with a drift and a time trend : $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$								
25	-22.5	-17.9	-4.38	-3.60	10.61	7.24	8.21	5.68
50	-25.7	-19.8	-4.15	-3.50	9.31	6.73	7.02	5.13
100	-27.4	-20.7	-4.04	-3.45	8.73	6.49	6.50	4.88
250	-28.4	-21.3	-3.99	-3.43	8.43	6.34	6.22	4.75
500	-28.9	-21.5	-3.98	-3.42	8.34	6.30	6.15	4.71
∞	-29.5	-21.8	-3.96	-3.41	8.27	6.25	6.09	4.68

Note: F-test (a) สำหรับ $H_0: \gamma = 0, a_0 = 0$ ในสมการ (2.14) และ $H_0: \gamma = 0, a_0 = 0$ ในสมการ (2.19)

F test (b) is for $H_0: \gamma = 0, a_0 = 0, a_2 = 0$ ในสมการ (สมการ (2.19))

แหล่งที่มา : Fuller (1976, p.371) for the K-test and p.373 for the t-test.

Dickey and fuller (1981, p.1063) for the F-test.

2.4 ตัวแบบเวกเตอร์การถดถอยในตัว (Vector autoregressive model) หรือตัวแบบ VAR

ตัวแบบเวกเตอร์การถดถอยในตัว เป็นตัวแบบที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ที่ไม่สามารถทราบถึงรูปแบบความสัมพันธ์ของตัวแปรต่าง ๆ มาก่อนล่วงหน้า หากใช้ข้อมูลในอดีตมาศึกษาความสัมพันธ์กัน ซึ่งตัวแปรภายในตัวใดตัวหนึ่งอาจจะขึ้นอยู่กับค่าในอดีตของตัวมันเองและตัวแปรภายในตัวอื่น ๆ และค่าปัจจุบันของตัวแปรภายนอก เนื่องจากไม่สามารถทราบรูปแบบความสัมพันธ์ของตัวแปรมาก่อนล่วงหน้า จึงทำให้ตัวแบบนี้ได้รับความนิยมในด้านการศึกษาความสัมพันธ์และผลกระทบของอนุกรมเวลาหลายชุดที่มีผลกระทบซึ่งกันและกันระหว่างการเปลี่ยนแปลงอย่างใดอย่างหนึ่ง

รูปแบบมาตรฐานของตัวแบบ VAR จะใช้วิเคราะห์ข้อมูลอนุกรมเวลาหลายตัวแปรและพิจารณาตัวแปรตามหรือตัวแปรภายใน (Dependent or endogenous Variables) หลายตัวพร้อมกัน ซึ่งตัวแปรภายในจะถูกกำหนดด้วยตัวอดีตของตัวมันเอง (Lagged variable) และจะถูกอธิบายด้วยตัวแปรล่าช้าของตัวแปรภายในตัวอื่น (Lagged of other variables)

กำหนดให้ $Y_{1t}, Y_{2t}, \dots, Y_{mt}$ คือ ข้อมูลอนุกรมเวลา m ชุด ณ เวลา t ใด ๆ โดยที่ $t = 1, 2, \dots, T$

เมื่อ T เป็นจำนวนคาบเวลา และกำหนดให้เวกเตอร์ $Y_t = [y_{1t}, y_{2t}, \dots, y_{mt}]'$ เป็นเวกเตอร์ของตัวแปรตามจำนวน m ตัว ณ เวลา t ใด ๆ

ตัวแบบเวกเตอร์การถดถอยในตัวเองลำดับที่ p เขียนแทนด้วย VAR(p) คือ

$$Y_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t \quad (2.7)$$

โดยที่

Y_{t-1}, \dots, Y_{t-p} คือ ค่าล่าช้า (lagged values) หรือค่าในอดีต (past values) ของเวกเตอร์ Y_t ย้อนหลัง 1 คาบเวลา ไปเรื่อย ๆ จนถึงย้อนหลัง p คาบเวลา ตามลำดับ

$A_0 = [a_{01}, a_{02}, \dots, a_{0m}]'$ คือ เวกเตอร์ค่าคงที่ (constants) ขนาด $m \times 1$

A_1, A_2, \dots, A_p คือ เมทริกซ์สัมประสิทธิ์การถดถอยของเวกเตอร์ Y_{t-1}, \dots, Y_{t-p} ขนาด $m \times m$

$\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{mt}]'$ คือ เวกเตอร์ของค่าความคลาดเคลื่อนสุ่มหรือตัวรบกวนขาว (white noise) ของ $Y_{1t}, Y_{2t}, \dots, Y_{mt}$ ตามลำดับ

เวกเตอร์ตัวรบกวน ε_t มีค่าเฉลี่ยเท่ากับเวกเตอร์ศูนย์ และไม่มีสหสัมพันธ์ในตัวเอง นั่นคือ

$$E(\varepsilon_t) = 0 \text{ และ } E(\varepsilon'_t \varepsilon_s) = 0, t \neq s$$

สมการ (3.1) เขียนอยู่ในรูปเมทริกซ์ได้ดังนี้

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{bmatrix} = A_0 + A_1 \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{m,t-1} \end{bmatrix} + A_2 \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \\ \vdots \\ y_{m,t-2} \end{bmatrix} + \dots + A_p \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{m,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{bmatrix}, t = 1, \dots, T \quad (2.8)$$

$$\text{โดยที่ } A_i = \begin{bmatrix} a_{11}^i & a_{12}^i & \dots & a_{1m}^i \\ a_{21}^i & a_{22}^i & \dots & a_{2m}^i \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^i & a_{m2}^i & \dots & a_{mm}^i \end{bmatrix}, i = 1, 2, \dots, p \quad (2.9)$$

การประมาณค่าพารามิเตอร์ในตัวแบบ VAR

เมื่อพิจารณาค่าสมการ จะพบว่า ตัวแปรสุ่มคลาดเคลื่อนในแต่ละสมการจะไม่มีความสัมพันธ์กันเอง ดังนั้น การใช้วิธีกำลังสองน้อยที่สุด (OLS) ในการประมาณค่าพารามิเตอร์ของสมการทั้งสองจะมีความแปรปรวนของตัวประมาณค่าต่ำที่สุด เราจึงสามารถหาค่าเฉลี่ยและความแปรปรวนของตัวแบบ VAR(1) ดังสมการต่อไปนี้

$$Y_t = A_0 + A_1 Y_{t-1} + \varepsilon_t$$

$$\text{ค่าเฉลี่ย : } E(Y_t) = \mu = (I - A_1)^{-1} A_0$$

$$\text{ความแปรปรวน : } \text{Var}(Y_t) = \Sigma + A_1 \Sigma A_1' + A_1^2 \Sigma (A_1^2)' + A_1^3 \Sigma (A_1^3)' + \dots$$

โดยที่ $A_1^j \rightarrow 0$ เมื่อ $j \rightarrow \infty$ นั้นหมายถึง ความแปรปรวนของอนุกรมเวลาทุกตัวที่อยู่ในเวกเตอร์ Y_t สามารถหาค่าได้

เมื่อเราทราบแล้วว่า อนุกรมเวลา n ชุดที่นำมาอยู่ในเวกเตอร์ Y_t ของตัวแบบ VAR จะต้องมีความเกี่ยวข้องต่อกันและกันได้ โดยมีจุดมุ่งหมายหลักของการวิเคราะห์ตัวแบบ VAR คือต้องการหาความสัมพันธ์ที่มีซึ่งกันและกันของอนุกรมเวลาใน Y_t ดังนั้น การเลือกลำดับ (p) ที่จะนำมาใช้ในตัวแบบ VAR ควรมีค่าที่

เหมาะสม โดยไม่ใช่ค่าที่ทำให้ต้องประมาณค่าพารามิเตอร์มากเกินไปจนความจำเป็น และไม่ใช่ค่าที่น้อยจนไม่สามารถแสดงถึงความสัมพันธ์ซึ่งกันและกันของอนุกรมเวลาในตัวแบบได้

การวิเคราะห์ตัวแบบ VAR มีข้อตกลงเบื้องต้น คือ ข้อมูลอนุกรมเวลา $y_{1t}, y_{2t}, \dots, y_{mt}$ จะต้องคงที่ (stationary) ในเชิงของความแปรปรวนและความแปรปรวนร่วม (variance-covariance stationary) ถ้าข้อมูลอนุกรมเวลาที่น่ามาวิเคราะห์ไม่เป็นไปตามเงื่อนไขดังกล่าว เราจะไม่สามารถมั่นใจได้ว่าการวิเคราะห์เวกเตอร์การถดถอยในตัวจะให้ผลลัพธ์ที่ถูกต้อง ซึ่งการตรวจสอบข้อมูลอนุกรมเวลา คงที่หรือไม่สามารถทำได้โดยหลายวิธี ดังนี้

1. พล็อตอนุกรมเวลาด้วยแผนภาพและตรวจสอบแนวโน้มและฤดูกาล (Visually Plotting) By
2. แบ่งอนุกรมเวลา (Splitting time series) ออกเป็นพาร์ติชันต่าง ๆ และเปรียบเทียบการอนุमानทางสถิติ

3. ทำการทดสอบ Augmented Dickey-Fuller test เพื่อตรวจสอบความคงที่

การศึกษานี้จะใช้วิธีการทดสอบ Augmented Dickey-Fuller test เพื่อตรวจสอบความคงที่ของอนุกรมเวลาก่อนที่จะทำการหาตัวแบบ VAR ตามรายละเอียดที่กล่าวไว้ในหัวข้อ 2.2

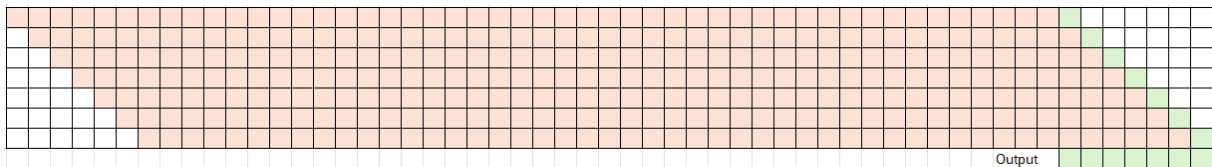
บทที่ 3

ขั้นตอนการดำเนินงาน

การศึกษานี้มีวัตถุประสงค์เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน n ชั่วโมงถัดไป และ ใน n วันถัดไป โดยใช้ค่าความเข้มข้นของ PM2.5 รายชั่วโมงในอดีต ด้วยเทคนิค LSTM และเพื่อพยากรณ์ PM2.5 รายวัน โดยใช้ข้อมูลอนุกรมเวลาหลายชุดด้วยตัวแบบ VAR ซึ่งขั้นตอนการดำเนินงานเพื่อตอบวัตถุประสงค์ทั้ง 2 ข้อนี้ มีรายละเอียดดังนี้

3.1 ขั้นตอนการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n ชั่วโมงถัดไป โดยใช้ข้อมูลในอดีตจำนวน 48 ชั่วโมงก่อนหน้า ด้วยเทคนิค LSTM

การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง จากข้อมูล n ชั่วโมงก่อนหน้า เป็นการนำข้อมูลค่าความเข้มข้นของ PM2.5 ก่อนหน้าช่วงเวลาที่จะทำนายมาใช้ในการสร้างตัวแบบ ดังรูปที่ 3.1 วิธีการนี้สามารถพยากรณ์ค่าความเข้มข้นของ PM2.5 ในช่วงเวลาที่ใกล้เคียงกับข้อมูลที่มีอยู่ได้อย่างแม่นยำ อย่างไรก็ตาม ระยะห่างระหว่างช่วงเวลาที่ต้องการพยากรณ์กับข้อมูลที่มีอยู่ อาจเป็นอุปสรรคหลักที่ต้องพบเจอ โดยเฉพาะอย่างยิ่ง เมื่อเราไม่มีข้อมูลสภาพอากาศอื่น ๆ มาประกอบด้วย ทำให้ความแม่นยำมีจำกัดและสามารถเชื่อถือได้ในระยะเวลาสั้น ๆ เท่านั้น



รูปที่ 3.1 แนวคิดการพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมงจากข้อมูล 48 ชั่วโมงก่อนหน้า

จากรูปที่ 3.1 การใช้ข้อมูลค่าความเข้มข้นของ PM2.5 รายชั่วโมง จำนวน 48 ชั่วโมงก่อนหน้า เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 ของ 1 ชั่วโมงถัดไป ซึ่งก็คือ เมื่อ $n=1$ กรอบสี่เหลี่ยมแสดงข้อมูลที่ใช้เป็นข้อมูลนำเข้าของโมเดล และกรอบสี่เหลี่ยมแสดงข้อมูลที่เป็นข้อมูลผลลัพธ์จากข้อมูลนำเข้าในบรรทัดเดียวกัน เมื่อนำผลลัพธ์ทุกบรรทัดมาประกอบกัน จะได้ผลการทำนายเป็นรายชั่วโมงต่อเนื่องกันไป

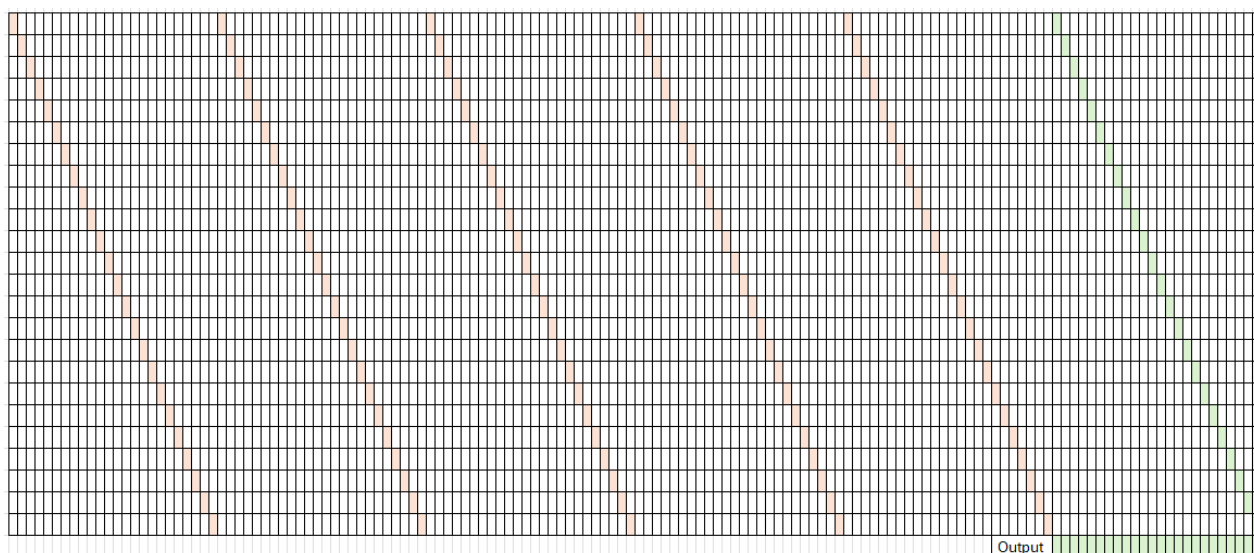
การหาตัวแบบสำหรับการพยากรณ์ตามรูปที่ 3.1 ผู้วิจัยใช้ข้อมูลค่าความเข้มข้นของ PM2.5 รายชั่วโมง จากสถานีตรวจวัดคุณภาพอากาศเขตดินแดง ในช่วงวันที่ 1 ตุลาคม 2565 – 31 ตุลาคม 2565 มาใช้เป็นชุดข้อมูลฝึกฝนเพื่อหาตัวแบบ โดยมีขั้นตอนดังนี้

1. ใช้ข้อมูลรายชั่วโมงจำนวน 48 ชั่วโมง (48 จุดข้อมูล) มาใช้ในการฝึกฝนตัวแบบ LSTM
2. ใช้ตัวแบบ LSTM พยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n ชั่วโมงถัดไป (ในรอบแรก กำหนดค่า $n=1$)

3. เมื่อได้ค่าพยากรณ์ของค่าความเข้มข้นของ PM2.5 ในชั่วโมงถัดไปแล้ว จะทำการขยับ window ของข้อมูลไปทางขวาอีก 1 รายชั่วโมง เพื่อฝึกฝนตัวแบบและค่าพยากรณ์ของค่าความเข้มข้นของ PM2.5 ในชั่วโมงถัดไป
4. ทำซ้ำ 1-3 ซ้ำ ๆ ไปจนถึงวันสุดท้ายของเดือน บันทึกผลของการพยากรณ์
5. เปลี่ยนค่า n เป็นค่าอื่น ในการทดลองนี้กำหนดให้ $n = 1, 2, 3, 4, 5, 6$ ทำซ้ำข้อ 1-4

3.2 ขั้นตอนการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า ด้วยเทคนิค LSTM

วิธีการการพยากรณ์ค่าความเข้มข้นของ PM2.5 โดยใช้ข้อมูลเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า เพื่อการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป เป็นการนำข้อมูลค่าความเข้มข้นของ PM2.5 ก่อนหน้าช่วงเวลาที่จะทำนายมาปรับปรุงโดยอาศัยรูปแบบของความผันแปร (การเพิ่มขึ้นและลดลง) ของ PM2.5 ในช่วงเวลาต่าง ๆ ของแต่ละวัน เช่น เราอาจสังเกตได้ว่าในตอนเช้าจะมีการเพิ่มขึ้นของความเข้มข้นของ PM2.5 มากกว่าช่วงเวลาอื่นของวัน และเหตุการณ์นี้มีแนวโน้มเกิดขึ้นซ้ำ ๆ ทุกวัน เราจึงทำการกระจายช่วงของข้อมูลออก แทนที่จะเป็นการใช้ทุกชั่วโมงของแต่ละวันดังวิธีการที่กล่าวไว้ในหัวข้อ 3.1 เราจะใช้เวลาทุก 24 ชั่วโมงแทน ดังรูปที่ 3.2



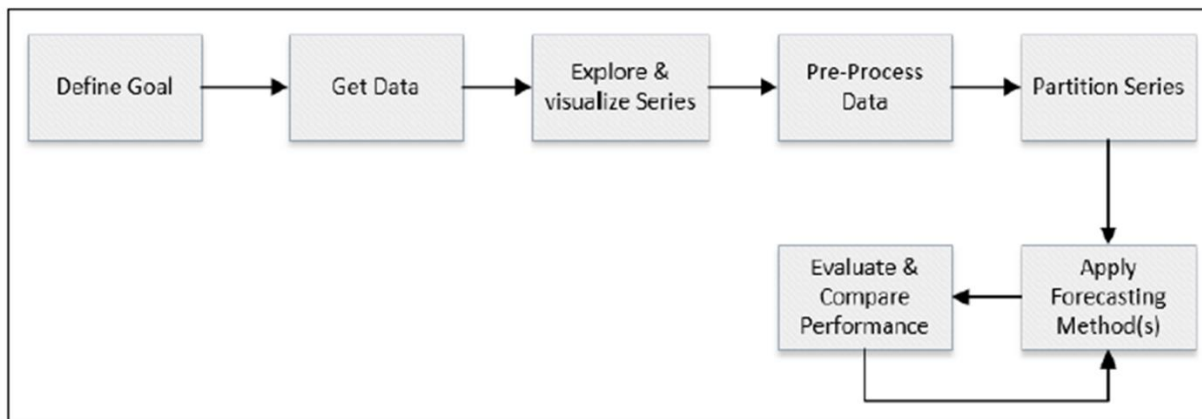
รูปที่ 21 วิธีการทำนายระดับฝุ่นละออง PM2.5 โดยใช้ประวัติข้อมูลเป็นรอบ 24 ชั่วโมง จำนวน 6 วันก่อนหน้า (เพื่อการแสดงตัวอย่าง ใช้ 6 วันเนื่องจากพื้นที่จำกัด ในการทดลองจริงใช้ 25 วัน) เพื่อทำนายระดับฝุ่น PM2.5 ใน 1 วันถัดไป กรอบสี่เหลี่ยมแสดงข้อมูลที่ใช้เป็นข้อมูลนำเข้าของโมเดล และกรอบสี่เหลี่ยมแสดงข้อมูลที่เป็นข้อมูลผลลัพธ์จากข้อมูลนำเข้าในบรรทัดเดียวกัน เมื่อนำผลลัพธ์ทุกบรรทัดมาประกอบกัน จะได้ผลการทำนายตลอด 24 ชั่วโมงของวันถัดไป

ผู้วิจัยใช้ข้อมูลค่าความเข้มข้นของ PM2.5 รายชั่วโมง จากสถานีตรวจวัดคุณภาพอากาศเขตดินแดง ในช่วงวันที่ 1 ตุลาคม 2565 – 31 ตุลาคม 2565 ตุลาคม 2565 มาใช้ในการฝึกฝนตัวแบบ โดยมีขั้นตอนดังนี้

1. ใช้ข้อมูลจำนวน 25 วัน (120 ชั่วโมง หรือ 120 จุดข้อมูล)
 2. ทำการจัดเรียงข้อมูลใหม่ โดยการเรียงข้อมูลในเวลาเดียวกันของแต่ละวันเป็นลำดับของข้อมูลที่ใช้ฝึกฝน และใช้ข้อมูลในเวลาเดียวกัน ของ n วันถัดไปเป็นข้อมูลทดสอบ สำหรับตัวแบบ LSTM เช่น ใช้ข้อมูลเวลา 7.00 น. ของวันที่ 1-25 ตุลาคม 2565 ในการฝึกฝน เพื่อคาดหวังให้ตัวแบบพยากรณ์ค่าความเข้มข้นของ PM2.5 ในวันที่ 26 ตุลาคม 2565 ในเวลา 7.00 น. จากนั้น ใช้ข้อมูลเวลา 8.00 น. ของวันที่ 1-25 ตุลาคม 2565 ในการฝึกฝน เพื่อคาดหวังให้ตัวแบบพยากรณ์ค่าความเข้มข้นของ PM2.5 ในวันที่ 26 ตุลาคม 2565 ในเวลา 8.00 น. ต่อเนื่องไปตามลำดับ เป็นต้น
 3. ฝึกฝนตัวแบบโดยใช้ข้อมูลทุกชุดที่เตรียมไว้
 4. ใช้ข้อมูล 5 วัน ในแต่ละเวลาเป็นข้อมูลนำเข้าของตัวแบบ LSTM เพื่อตัวแบบพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน $n \times 24$ ชั่วโมงถัดไป (ในรอบแรก กำหนดค่า $n=1$)
 5. เมื่อได้ค่าพยากรณ์ของค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงถัดไปแล้ว ทำการการขยับ window ของข้อมูลไปทางขวาอีก 1 ข้อมูล เพื่อฝึกฝนและพยากรณ์ค่าความเข้มข้นของ PM 2.5 ใน 24 ชั่วโมงถัดไป
 6. ทำข้อ 1-5 ซ้ำ ๆ ไปจนถึงวันสุดท้ายของเดือน บันทึกผลของการพยากรณ์
 7. เปลี่ยนค่า n เป็นค่าอื่น ในการศึกษาที่กำหนดให้ $n = 1, 2, 3, 4, 5$ ทำซ้ำข้อ 1-6 กล่าวคือเปลี่ยนวันที่จะทำนายจาก 26 ตุลาคม 2565 เป็น 27, 28, 29, 30 ตุลาคม 2565 แทน
- วิธีการนี้ทำให้ตัวแบบสามารถตรวจจับการเปลี่ยนแปลงของของค่าความเข้มข้นของ PM2.5 ในแต่ละวันในช่วงเวลาเดียวกันได้ ทำให้ประสิทธิภาพของตัวแบบ LSTM ที่สามารถพยากรณ์ค่าความเข้มข้นของ PM2.5 ล่วงหน้าในช่วงเวลาระยะยาวกว่า อย่างไรก็ตาม ระยะเวลาที่ต้องการพยากรณ์ล่วงหน้ายังคงมีจำกัด เนื่องจากความผันแปรของค่าความเข้มข้นของ PM2.5 ในแต่ละวัน ได้รับอิทธิพลของหลายปัจจัยแต่ไม่สามารถศึกษาการพยากรณ์ด้วยปัจจัยอื่นๆ ได้ เพราะความจำกัดของข้อมูล หากสามารถเก็บรวบรวมข้อมูลอื่นๆ เพิ่มเติม เช่น มลพิษอากาศอื่นๆ นอกเหนือจาก PM2.5 หรือ ข้อมูลอุตุนิยมวิทยา อาจจะใช้ตัวแบบอื่นในการพยากรณ์ค่าความเข้มข้นของ PM2.5

3.3 ขั้นตอนการพยากรณ์ PM2.5 รายวัน โดยใช้ข้อมูลอนุกรมเวลาหลายชุด ด้วยตัวแบบ VAR

การพยากรณ์ PM2.5 รายวัน โดยใช้ข้อมูลอนุกรมเวลาหลายชุด ด้วยตัวแบบ VAR เพื่อศึกษาความสัมพันธ์ของมลพิษทางอากาศต่างๆ มีขั้นตอนการดำเนินงานดังรูปที่ 3.3



รูปที่ 3.3 ขั้นตอนการดำเนินงานของการศึกษาปัจจัยที่ส่งผลต่อมลพิษทางอากาศทั้ง 5 ชนิด

จากรูปที่ 3.3 ขั้นตอนการดำเนินงานของการศึกษาปัจจัยที่ส่งผลต่อมลพิษทางอากาศ มีดังนี้

(1) การกำหนดเป้าหมายของการศึกษา (Define Goal) ซึ่งการศึกษานี้ มีเป้าหมายเพื่อศึกษาความสัมพันธ์มลพิษทางอากาศทั้ง 5 ชนิด ได้แก่ คือ PM2.5, PM10, NO₂ และ CO ซึ่งมีความสัมพันธ์และอิทธิพลซึ่งกันและกัน โดยใช้เทคนิคการวิเคราะห์ทางสถิติ ที่เรียกว่า ตัวแบบเวกเตอร์การถดถอยในตัว (Vector autoregressive model)

(2) การได้มาซึ่งข้อมูล (Get data) ในการศึกษาที่ใช้ข้อมูลคุณภาพอากาศของกรุงเทพมหานคร จำนวน 4 เขตที่มีสถานีตรวจวัดคุณภาพอากาศแบบตัวคอนเทนเนอร์ จำนวน 4 สถานี ซึ่งจะกล่าวถึงรายละเอียดในหัวข้อ 3.3.1-3.3.3

(3) การสำรวจและแสดงภาพของข้อมูลอนุกรมเวลา (Explore and Visualize series) เนื่องจากข้อมูลอนุกรมเวลา ที่นำมาใช้ในการศึกษา อาจยังมีข้อผิดพลาด เช่น ค่าของข้อมูลไม่ถูกต้องหรือมีข้อมูลสูญหาย จึงจำเป็นต้องตรวจสอบก่อนที่นำมาใช้วิเคราะห์ข้อมูล

(4) กระบวนการจัดเตรียมข้อมูลก่อนการวิเคราะห์ (Pre-process data) โดยจัดการข้อมูลที่มีค่าผิดพลาดและข้อมูลที่มีค่าสูญหาย ให้พร้อมใช้งาน

(5) การแบ่งข้อมูลอนุกรมเวลาออกเป็นส่วนๆ (Partition series) เป็นการแบ่งข้อมูลสำหรับการหาตัวแบบในการพยากรณ์ค่าความเข้มข้นของมลพิษทางอากาศ (training data set) และข้อมูลที่ใช้สำหรับประเมินประสิทธิภาพของการพยากรณ์ (test data set)

(6) การประยุกต์ใช้วิธีการทางอนุกรมเวลา เพื่อหาตัวแบบที่เหมาะสมสำหรับการพยากรณ์ (Apply Forecasting method) ค่าความเข้มข้นของมลพิษทางอากาศ

(7) การประเมินและเปรียบเทียบประสิทธิภาพ (evaluate and compare performance) ของการพยากรณ์โดยใช้ตัวแบบที่ได้จากขั้นตอนที่ (6)

3.3.1 การเตรียมข้อมูลและการจัดการข้อมูล

ข้อมูลคุณภาพอากาศของกรุงเทพมหานคร (จากกองจัดการคุณภาพอากาศและเสียง สำนักสิ่งแวดล้อม) จำนวน 4 เขตที่มีสถานีตรวจวัดคุณภาพอากาศแบบตู้คอนเทนเนอร์ จำนวน 4 สถานี ดังนี้

สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนงและสถานีบริเวณสำนักงานเขตราชวัตรบูรณะ มีการตรวจวัดค่ามลพิษอากาศรายชั่วโมง คือ PM₁₀, NO, NO₂, NO_x และ CO ตั้งแต่ ปี 2549 ส่วนค่า PM_{2.5} เริ่มมีการติดตั้งเมื่อกุมภาพันธ์ 2563 และเก็บข้อมูลการตรวจอากาศรายชั่วโมง ได้แก่ ความเร็วลม (ws) ทิศทางลม (wind direction: wd) ค่าความชื้นสัมพัทธ์ (relative humidity : Rh) อุณหภูมิอากาศ (temperature : temp) และความกดอากาศ (air pressure) ซึ่งเริ่มมีการติดตั้งในเดือนกรกฎาคม 2565 โดยข้อมูลทั้งหมดเก็บรวบรวมถึงเดือนมิถุนายน 2566

สถานีบริเวณสำนักงานเขตราชเทวี มีการตรวจวัดค่า PM_{2.5}, PM₁₀, NO, NO₂, NO_x, CO, O₃ และเก็บข้อมูลการตรวจอากาศรายวัน ได้แก่ ความเร็วลม (ws) ทิศทางลม (wind direction: wd) ค่าความชื้นสัมพัทธ์ (relative humidity : Rh) อุณหภูมิ (temperature : temp) และความกดอากาศ (Air pressure) เริ่มตั้งแต่ปี 2553 โดยข้อมูลเก็บทั้งหมดรวบรวมถึงเดือนมิถุนายน 2566 เป็นรายชั่วโมง

ข้อมูลคุณภาพอากาศของแต่ละสถานี เป็นไฟล์นามสกุล .xlsx มีรายละเอียดดังนี้

1) ข้อมูลของสถานีดินแดง มีจำนวนทั้งหมด 4 ไฟล์ คือ ไฟล์ข้อมูลปี 49-55 (7 sheets), ไฟล์ข้อมูลปี 56-64 (9 sheets) ไฟล์ข้อมูลปี 65 (12 sheets เดือน ม.ค.-ธ.ค.) และไฟล์ข้อมูลปี 66 (6 sheets เดือน ม.ค.-มิ.ย.)

2) ข้อมูลของสถานีบริเวณสำนักงานเขตพระโขนง มีจำนวนทั้งหมด 4 ไฟล์ คือ ไฟล์ข้อมูลปี 49-55 (7 sheets), ไฟล์ข้อมูลปี 56-64 (9 sheets) ไฟล์ข้อมูลปี 65 (12 sheets เดือน ม.ค.-ธ.ค.) และไฟล์ข้อมูลปี 66 (6 sheets เดือน ม.ค.-มิ.ย.)

3) ข้อมูลของสถานีบริเวณสำนักงานเขตราชวัตรบูรณะ มีจำนวนทั้งหมด 4 ไฟล์ คือ ไฟล์ข้อมูลปี 49-59 (11 sheets), ไฟล์ข้อมูลปี 60-64 (5 sheets) ไฟล์ข้อมูลปี 65 (12 sheets เดือน ม.ค.-ธ.ค.) และไฟล์ข้อมูลปี 66 (6 sheets เดือน ม.ค.-มิ.ย.)

4) สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนงและสถานีบริเวณสำนักงานเขตราชวัตรบูรณะมีจำนวนทั้งหมด 4 ไฟล์ คือ ไฟล์ข้อมูลปี 49-55 (7 sheets), ไฟล์ข้อมูลปี 56-64 (9 sheets) ไฟล์ข้อมูลปี 65 (12 sheets เดือน ม.ค.-ธ.ค.) และไฟล์ข้อมูลปี 66 (6 sheets เดือน ม.ค.-มิ.ย.)

ตารางที่ 3.1 แสดงวันที่เริ่มและวันสุดท้ายที่บันทึกข้อมูลและข้อมูลคุณภาพอากาศของในแต่ละสถานี

ตัวแปร	สถานีดินแดง		สถานีบริเวณสำนักงาน เขตพระโขนง		สถานีบริเวณสำนักงาน เขตราชวัชรปุณณะ		สถานีบริเวณสำนักงาน เขตราชเทวี	
	วันที่เริ่ม บันทึก ข้อมูล	วันสุดท้าย ที่บันทึก ข้อมูล	วันที่เริ่ม บันทึก ข้อมูล	วันสุดท้าย ที่บันทึก ข้อมูล	วันที่เริ่ม บันทึก ข้อมูล	วันสุดท้าย ที่บันทึก ข้อมูล	วันที่เริ่ม บันทึก ข้อมูล	วันสุดท้าย ที่บันทึก ข้อมูล
	มลพิษทางอากาศ							
PM2.5	07/01/63 16:00 น.	30/06/66 23.00 น.	20/01/63 10:00 น.	30/06/66 23.00 น.	09/01/63 14.00 น.	30/06/66 23.00 น.	01/01/57 01:00 น.	30/06/66 23.00 น.
PM10								
NO	01/01/49	30/06/66	01/01/49	30/06/66	01/01/49	30/06/66	01/01/53	30/06/66
NO2	17:00 น.	23.00 น.	00:00 น.	23.00 น.	17:00 น.	23.00 น.	00:00 น.	23.00 น.
NOx								
CO								
O3	ไม่มีข้อมูล		ไม่มีข้อมูล		ไม่มีข้อมูล		01/01/53 00:00 น.	30/06/66 23.00 น.
ข้อมูลอุตุนิยมวิทยา								
WS								
WD	25/07/65	30/06/66	22/07/65	30/06/66	22/07/65	30/06/66	01/01/53	30/06/66
Temp	13:00	23.00 น.	23:00 น.	23.00 น.	14:00 น.	23.00 น.	00:00 น.	23.00 น.
RH								
AP								

3.3.2 การตรวจสอบข้อมูลเพื่อเตรียมข้อมูลให้พร้อมใช้งาน

การตรวจสอบข้อมูลเพื่อเตรียมข้อมูลให้พร้อมใช้งาน มีขั้นตอนการดำเนินงานดังนี้

1) นำเข้าข้อมูลคุณภาพอากาศของแต่ละสถานี ซึ่งเป็นไฟล์นามสกุล .xlsx ในโปรแกรม Rstudio และจัดการข้อมูล โดยนำข้อมูลทั้ง 4 สถานีมา merge รวมกัน โดยปรับรูปแบบ (Format) ของข้อมูลในแต่ละไฟล์ให้เป็นรูปแบบเดียวกัน โดยที่สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง และ สถานีบริเวณสำนักงานเขตราชวัชรปุณณะ นำเข้าข้อมูลตั้งแต่ 1 มกราคม 2563 – มิถุนายน 2566 ส่วนข้อมูลของสถานีบริเวณสำนักงานเขตราชเทวีนำเข้าข้อมูลตั้งแต่ 1 มกราคม 2560 – มิถุนายน 2566

2) ทำการตรวจสอบ รูปแบบของ Date & Time ให้ถูกต้องและวันที่/เวลาของแต่ละปี ครอบคลุมไม่มีข้อมูลของค่าสูญหาย (Missing value) ของวันที่และเวลา

3) แสดงค่าสถิติสรุปข้อมูลคุณภาพอากาศรายชั่วโมงของทั้ง 4 สถานี คือ สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนงและสถานีบริเวณสำนักงานเขตราชกรีฑุรณะ และสถานีบริเวณสำนักงานเขตราชเทวี เพื่อตรวจสอบความถูกต้องของข้อมูลเบื้องต้น ซึ่งแสดงดังตารางที่ 3.2 ถึง 3.5

ตารางที่ 3.2 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีดินแดง (ก่อนการจัดการข้อมูล) ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566

ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	จำนวนข้อมูลสูญหาย
ข้อมูลความเข้มข้นของมลพิษอากาศ ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566							
PM2.5 ($\mu\text{g}/\text{m}^3$)	-9999.00	13.68	20.25	23.93	31.23	952.89	585
PM10 ($\mu\text{g}/\text{m}^3$)	0.00	45.0	59.51	65.59	79.66	949.34	425
NO (ppb)	-0.0013	8.6946	20.7458	33.668	46.305	332.00	1908
NO ₂ (ppb)	-0.3759	11.6974	19.8000	25.2558	32.0083	270.615	1963
NO _x (ppb)	-0.0005	21.6129	40.0974	53.7248	72.1640	360.9173	1963
CO (ppm)	-0.0425	0.3879	0.6229	0.6917	0.918	4.7051	1936
ข้อมูลอุตุนิยมวิทยาตั้งแต่วันที่ 25 กรกฎาคม 2565 ถึง 30 มิถุนายน 2566							
WS (m/s)	0.035	2.002	2.532	2.510	3.031	24.220	129
WD (°)	0.003	32.756	209.877	171.161	278.116	359.946	129
Temp (°C)	19.73	27.00	29.10	29.15	31.17	39.31	129
RH (%)	22.66	53.60	64.10	64.25	75.63	94.35	129
AP (mmbars)	964.6	1007.0	1009.0	1009.2	1011.0	1020.7	129

จากตารางที่ 3.2 ข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีดินแดง พบว่าความเข้มข้นของ PM2.5, PM10, NO, NO₂, NO_x และ CO มีค่าต่ำสุดเป็นค่าติดลบ และมีจำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายเท่ากับ 585, 425, 1908, 1963, 1963, 1936 ค่า ตามลำดับ ส่วนข้อมูลอุตุนิยมวิทยารายชั่วโมงตั้งแต่วันที่ 25 กรกฎาคม 2565 ถึง 30 มิถุนายน 2566 ของทั้ง 5 ตัวแปร มีค่าสูญหายเท่ากับ 129 ค่า

ตารางที่ 3.3 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีบริเวณสำนักงานเขตพระโขนง (ก่อนการจัดการข้อมูล)

ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	จำนวนข้อมูลสูญหาย
ข้อมูลความเข้มข้นของมลพิษอากาศ ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566							
PM _{2.5} (µg/m ³)	0.86	11.82	19.48	24.13	32.21	178.76	667
PM ₁₀ (µg/m ³)	0.66	29.63	46.99	54.14	72.00	1004.90	787
NO (ppb)	-6.515	4.490	17.000	45.242	69.000	369.410	1703
NO ₂ (ppb)	-0.04	9.35	16.63	23.84	35.89	123.9	1703
NO _x (ppb)	-8.732	14.520	35.200	69.091	109.730	434.070	1703
CO (ppm)	0.0000	0.4600	0.6800	0.9362	1.3100	4.8200	1446
ข้อมูลอุตุนิยมวิทยาตั้งแต่วันที่ 25 กรกฎาคม 2565 ถึง 30 มิถุนายน 2566							
WS (m/s)	-39.860	1.3000	1.6000	0.4246	2.0000	11.3200	507
WD (°)	-269.0	169.0	195.0	169.7	212.0	351.0	507
Temp (°C)	19.80	26.90	28.96	29.01	31.00	37.50	327
RH (%)	26.0	56.0	67.0	66.4	78.0	96.0	327
AP (mmbar)	941	1007	1009	1010	1011	1021	330

จากตารางที่ 3.3 ข้อมูลคุณภาพอากาศรายชั่วโมงของสถานีบริเวณสำนักงานเขตพระโขนง พบว่าความเข้มข้นของ NO, NO₂ และ NO_x มีค่าต่ำสุดเป็นค่าติดลบ จำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายของ PM_{2.5}, PM₁₀, NO, NO₂, NO_x และ CO เท่ากับ 667, 787, 1703, 1703, 1703 และ 1446 ค่าตามลำดับ ส่วนข้อมูลอุตุนิยมวิทยารายชั่วโมง พบว่า ความเร็วลม (WS) และทิศทางลม มีค่าต่ำสุดเป็นค่าติดลบ เช่นเดียวกัน และมีจำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายเท่ากับ 507, 507, 327, 327 และ 330 ค่า ตามลำดับ

ตารางที่ 3.4 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีบริเวณสำนักงานเขตราชวัชรปุระ (ก่อนการจัดการข้อมูล)

ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	จำนวนข้อมูลสูญหาย
ข้อมูลความเข้มข้นของมลพิษอากาศ ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566							
PM _{2.5} (µg/m ³)	-9999.000	12.923	20.000	2.427	30.152	158.120	307
PM ₁₀ (µg/m ³)	3.00	41.77	56.63	62.16	76.20	469.13	343
NO (ppb)	-0.0766	7.0000	20.1478	27.3618	38.6892	304.7000	1584
NO ₂ (ppb)	0.00	12.50	18.20	21.10	26.08	157.54	1583
NO _x (ppb)	0.00	22.28	40.33	48.46	64.60	389.10	1584
CO (ppm)	0.1644	0.3711	0.5400	0.6267	0.7966	11.2980	1696
ข้อมูลอุตุนิยมวิทยาตั้งแต่วันที่ 25 กรกฎาคม 2565 ถึง 30 มิถุนายน 2566							
WS (m/s)	0.01694	1.57621	2.18542	2.18254	2.80383	5.34000	1
WD (°)	0.0	69.0	218.0	184.1	184.1	360.0	1
Temp (°C)	19.80	26.66	28.83	28.87	30.99	38.10	0
RH (%)	26.42	58.20	68.29	68.39	79.55	79.55	0
AP (mmbar)	998	1007	1009	1010	1011	1021	0

จากตารางที่ 3.4 ข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีบริเวณสำนักงานเขตราชวัชรปุระ พบว่า ความเข้มข้นของ PM_{2.5} และ NO , NO₂, NO_x และ CO มีค่าต่ำสุดเป็นค่าติดลบ ส่วนความเข้มข้นของ NO₂ และ NO_x มีค่าต่ำสุดเป็นศูนย์ และมีจำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายของ PM_{2.5} , PM₁₀, NO, NO₂, NO_x และ CO เท่ากับ 507, 507, 327, 327 และ 330 ค่าตามลำดับ ส่วนข้อมูลอุตุนิยมวิทยารายชั่วโมง พบว่า ความเร็วลม (WS) และทิศทางลม มีจำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายเท่ากับ 1 ค่าเท่านั้น

ตารางที่ 3.5 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีบริเวณสำนักงานเขตราชเทวี (ก่อนการจัดการข้อมูล)

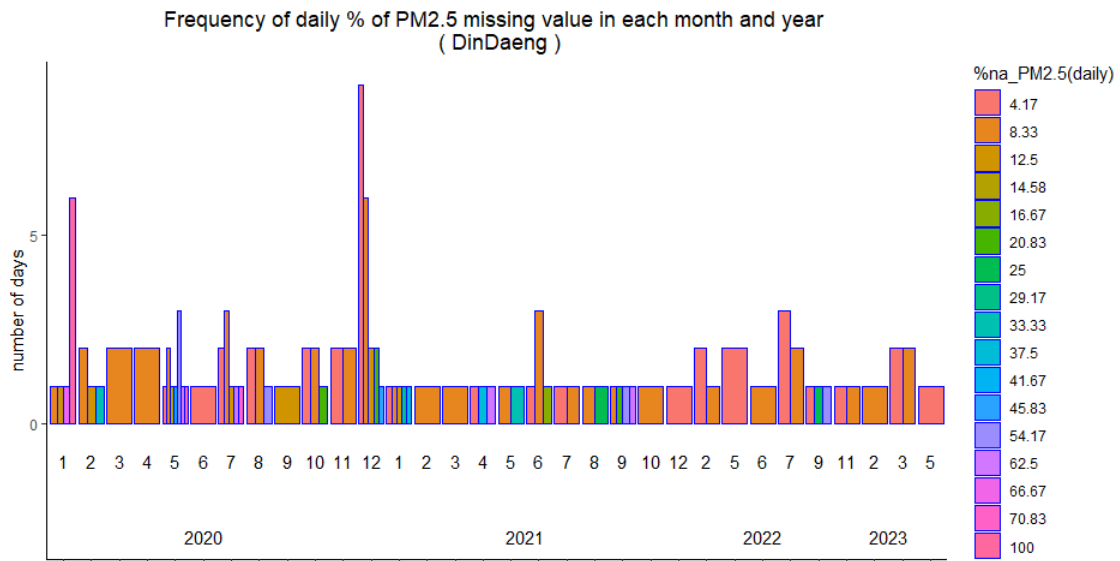
ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	จำนวนข้อมูลสูญหาย
ข้อมูลความเข้มข้นของมลพิษอากาศ ตั้งแต่วันที่ 1 มกราคม 2560 ถึง 30 มิถุนายน 2566							
PM _{2.5} (µg/m ³)	-23.00	14.00	22.00	25.91	33.00	473.00	851
PM ₁₀ (µg/m ³)	-9999.0	28.0	40.0	43.7	56.0	982.0	803
NO (ppb)	12.00	3.00	8.00	11.73	16.00	326.00	2764
NO ₂ (ppb)	0.00	17.00	23.00	24.87	31.00	185.00	2734
NO _x (ppb)	0.00	23.00	33.00	35.58	45.00	406.00	2764
CO (ppm)	-0.08	0.49	0.68	1.42	0.94	115.00	3465
O ₃ (ppm)	-10.00	4.00	11.00	15.77	22.00	268.00	2560
ข้อมูลอุตุนิยมวิทยาตั้งแต่วันที่ 1 มกราคม 2560 ถึง 30 มิถุนายน 2566							
WS (m/s)	-9999	0.20	0.30	-43.59	0.40	15.30	214
WD (°)	-9999	25.0	219.0	154.4	270.0	360.0	214
Temp (°C)	18.90	27.00	29.30	29.37	31.5	49.50	214
RH (%)	20.00	52.00	62.00	62.55	73.00	90.00	214
AP (mmbar)	992	1000	1002	1002	1004	1013	214

จากตารางที่ 3.5 ข้อมูลคุณภาพอากาศและข้อมูลอุตุนิยมวิทยารายชั่วโมงของสถานีบริเวณสำนักงานเขตราชเทวี พบว่า ความเข้มข้นของ PM_{2.5}, PM₁₀, CO และ O₃ มีค่าต่ำสุดเป็นค่าติดลบ ส่วนความเข้มข้นของ NO₂ และ NO_x มีค่าต่ำสุดเป็นศูนย์ และมีจำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายของ PM_{2.5}, PM₁₀, NO, NO₂, NO_x และ CO เท่ากับ 851, 803, 2764, 2734, 2764, 3465 และ 2560 ค่าตามลำดับ ส่วนข้อมูลอุตุนิยมวิทยารายชั่วโมง พบว่า ความเร็วลม (WS) และทิศทางลม มีค่าต่ำสุดเป็นค่าติดลบ ซึ่งมีค่าเท่ากับ -9999 มีทุกตัวแปรของข้อมูลอุตุนิยมวิทยา มีจำนวนค่าสังเกตรายชั่วโมงที่เป็นค่าสูญหายเท่ากับ 214 ค่า

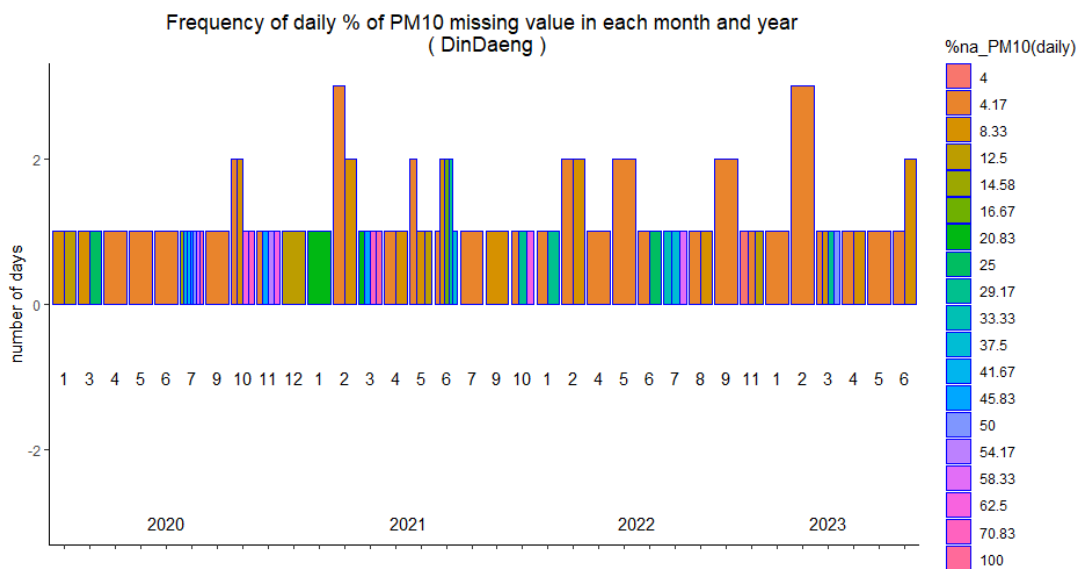
จากผลการสำรวจข้อมูลข้างต้น จะเห็นว่า ค่าสังเกตที่มีค่าต่ำสุดติดลบ ซึ่งเป็นค่าวัดที่ไม่ถูกต้อง และยังมีค่าสูญหายจำนวนหนึ่ง ดังนั้นจะต้องมีการจัดการข้อมูลก่อนที่จะนำข้อมูลไปใช้ต่อไป โดยแทนค่าสังเกตของตัวแปรต่าง ๆ ที่มีค่าติดลบ ด้วยค่า NA

4) ตรวจสอบจำนวนค่าสูญหาย (Missing value) ของตัวแปรต่าง ๆ ได้แก่ PM_{2.5}, PM₁₀, NO, NO₂, NO_x, CO, O₃ รายชั่วโมงของแต่ละวัน โดยคำนวณร้อยละของข้อมูลรายชั่วโมงของแต่ละวันและแจก

แจกแจงความถี่ (จำนวนวัน) ของค่าสัดส่วนข้อมูลที่มีค่าสูญหายในแต่ละเดือน และนำเสนอเป็นแผนภูมิแท่ง ดังรูปที่ 3.4 ถึง 3.5

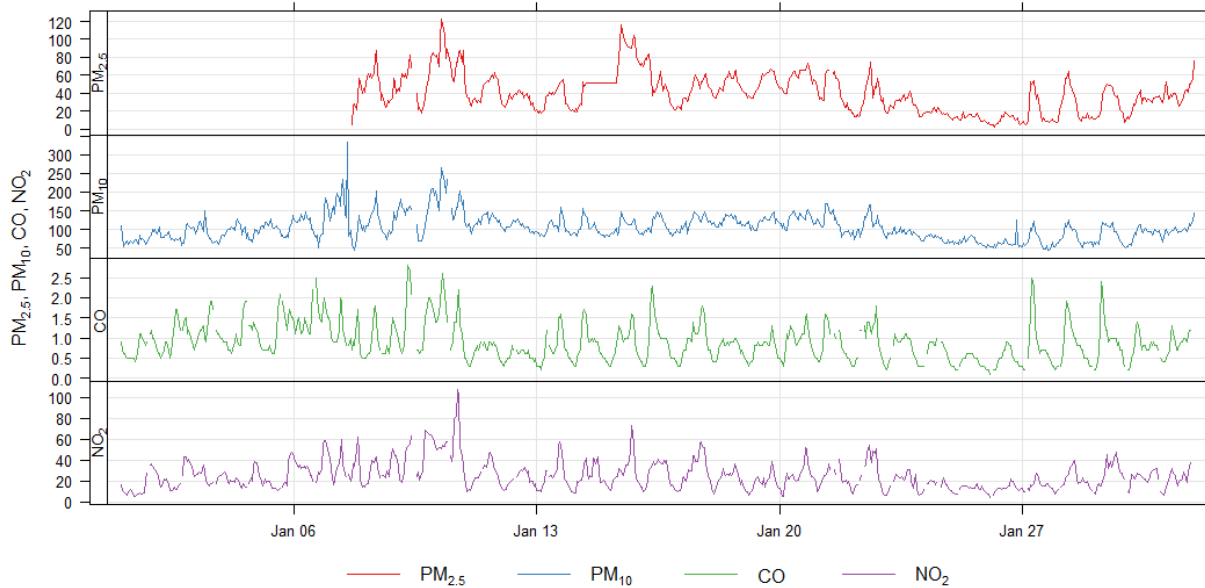


รูปที่ 3.4 การแจกแจงความถี่ของสัดส่วนข้อมูลสูญหายของ PM2.5 ในแต่ละเดือนของสถานีดินแดง

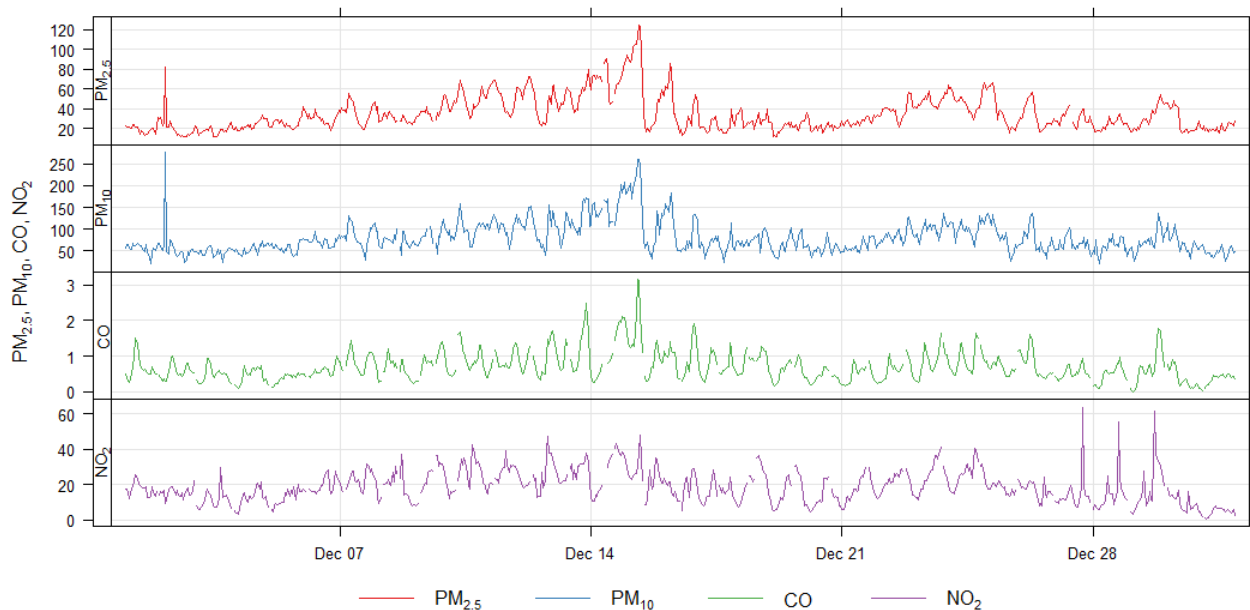


รูปที่ 3.5 การแจกแจงความถี่ของสัดส่วนข้อมูลสูญหายของ PM₁₀ ในแต่ละเดือนของสถานีดินแดง

รูปที่ 3.6 และ 3.7 แสดงกราฟเส้นค่าความเข้มข้นของ PM_{2.5} PM₁₀ CO และ NO₂ รายชั่วโมงในเดือนมกราคม ปี 2012 และเดือนธันวาคม เมื่อข้อมูลมีค่าสูญหายของค่าสูญหาย ตามลำดับ ซึ่งจะเห็นว่า ค่าสังเกตมีค่าสูญหาย ในหลายช่วงเวลา



รูปที่ 3.6 กราฟแสดงค่าความเข้มข้นของ PM_{2.5} PM₁₀ CO และ NO₂ รายชั่วโมงในเดือนมกราคม ปี 2012 เมื่อข้อมูลมีค่าสูญหาย



รูปที่ 3.7 กราฟแสดงค่าความเข้มข้นของ PM_{2.5} PM₁₀ CO และ NO₂ รายชั่วโมงในเดือนธันวาคม ปี 2012

3.3.3 การจัดการกับข้อมูลที่มีค่าสูญหาย

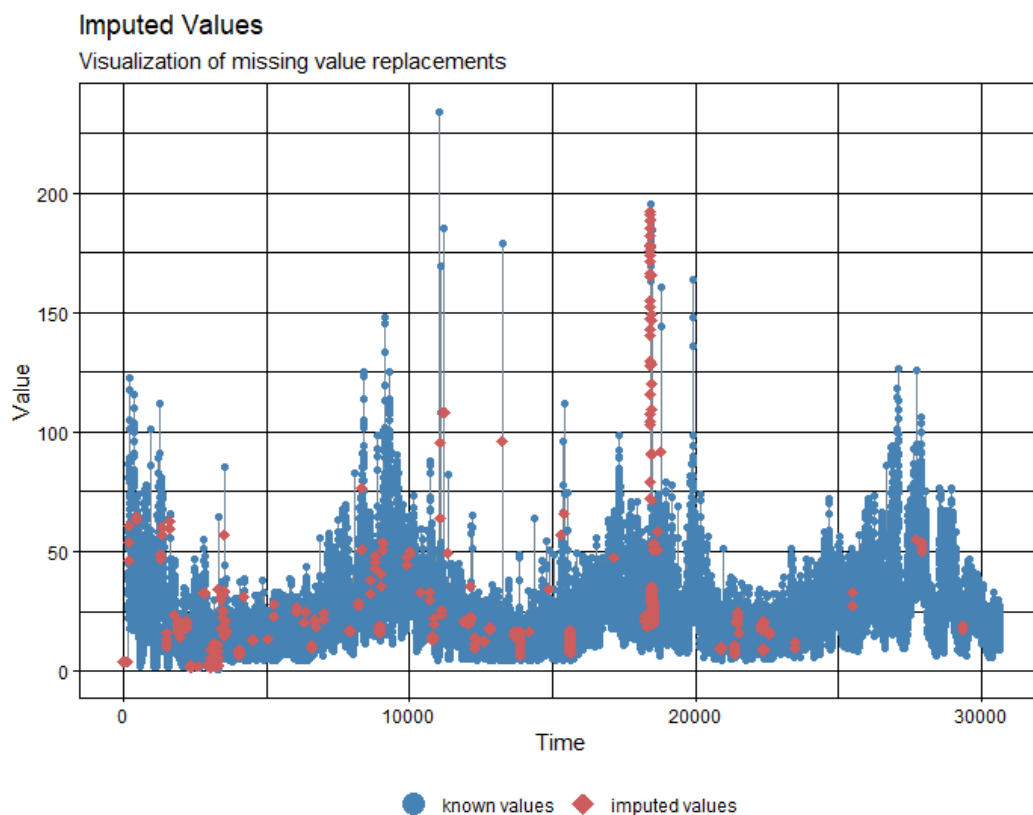
จากการสรุปข้อมูลที่มีค่าสูญหายข้างต้น ผู้วิจัยจึงได้ประมาณค่าสูญหาย โดยใช้วิธีการประมาณค่าสูญหาย ด้วยวิธี วิธีการประมาณค่าในช่วงเชิงเส้น (linear interpolation)

สมการของวิธีการประมาณค่าในช่วงเชิงเส้น สามารถเขียนอยู่ในรูป

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x - x_0) \quad (3.1)$$

เมื่อ x เป็นตัวแปรอธิบาย x_1 และ x_0 เป็นค่าสังเกตของตัวแปรอธิบาย x ที่ทราบค่า ซึ่งในข้อมูลอนุกรมเวลา หรือ เวลา (date and time) และ $f(x)$ คือ ค่าของตัวแปรตาม ซึ่งเป็นค่าความเข้มข้นของมลพิษอากาศแต่ละชนิด หรือค่าตัวแปรคุณนัยมวิทยา ณ เวลาที่ x โดยที่ $f(x_0)$ และ $f(x_1)$ คือ ค่าตัวแปรตาม ณ เวลาที่ x_0 และ x_1 ตามลำดับ (Junninen et al., 2004)

ตัวอย่างการประมาณค่าสูญหายของความเข้มข้น PM2.5 แสดงได้ดังรูปที่ 3.8



รูปที่ 3.8 การประมาณค่าสูญหายของความเข้มข้น PM2.5 จากสถานีดินแดง

หลังจากประมาณค่าสูญหาย จะได้ข้อมูลพร้อมสำหรับในการวิเคราะห์เพื่อศึกษาปัจจัยที่ส่งผลต่อ PM2.5 ในขั้นตอนต่อไป

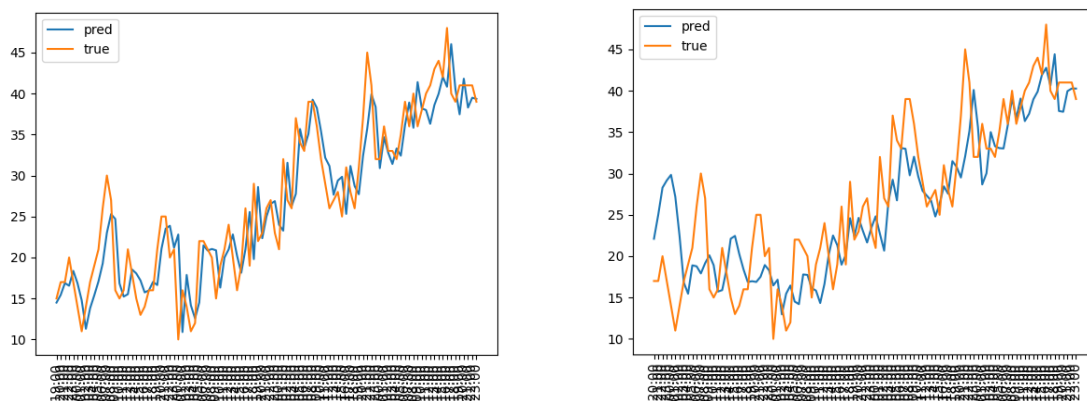
บทที่ 4

ผลการศึกษา

ผลการศึกษาปัจจัยที่ส่งผลต่อมลพิษทางอากาศจากข้อมูลคุณภาพอากาศของสถานีตรวจวัดคุณภาพอากาศทั้ง 4 สถานี คือ สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง สถานีบริเวณสำนักงานเขตราชวัชรบูรณะ และสถานีบริเวณสำนักงานเขตราชเทวี มีรายละเอียดดังนี้

4.1 การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n ชั่วโมงถัดไป โดยใช้ข้อมูลจำนวน 48 ชั่วโมงก่อนหน้า ด้วยเทคนิค LSTM

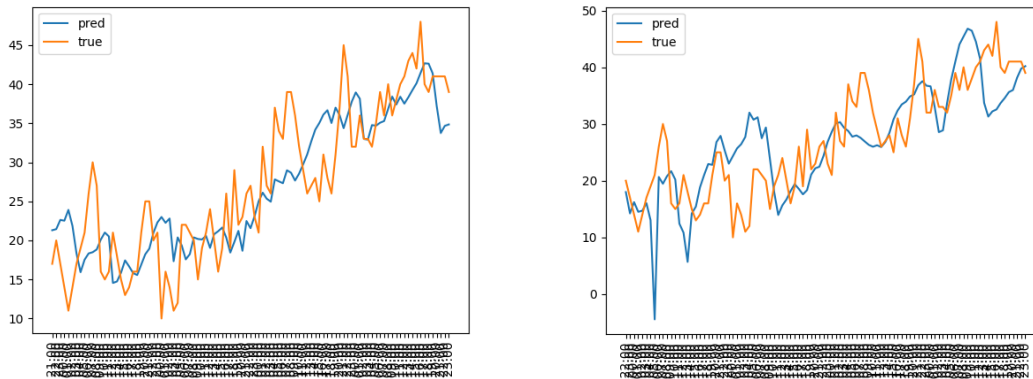
การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใช้ข้อมูลจำนวน 48 ชั่วโมงก่อนหน้า ด้วยเทคนิค LSTM ด้วยเทคนิค LSTM ผู้วิจัยใช้ข้อมูลค่าความเข้มข้นของ PM2.5 รายชั่วโมง จากสถานีตรวจวัดคุณภาพอากาศเขตดินแดง ในช่วงวันที่ 1 ตุลาคม 2565 – 31 ตุลาคม 2565 มาใช้เป็นชุดข้อมูลฝึกฝนเพื่อหาตัวแบบ ด้วยเทคนิค LSTM ได้ผลรูปที่ 4.1 -



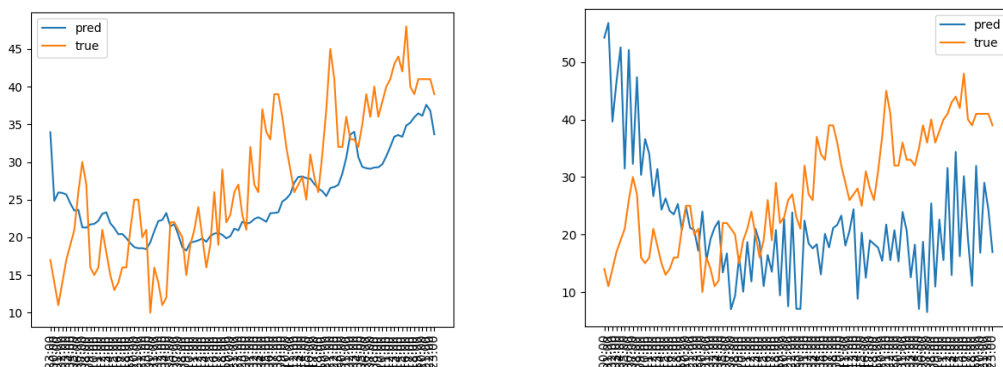
รูปที่ 4.1 (ซ้าย) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 1 ชั่วโมงถัดไป (ขวา) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 2 ชั่วโมงถัดไป โดยใช้ข้อมูล 48 ชั่วโมงก่อนหน้า

จากรูปที่ 4.1 แกนนอนแสดงช่วงเวลาในการทำงาน แกนตั้งแสดงค่าความเข้มข้นของ PM2.5 เส้นกราฟสีส้มแสดงปริมาณฝุ่นจริงและเส้นกราฟสีฟ้าแสดงค่าพยากรณ์ของ PM2.5 รายชั่วโมงที่ได้จากตัวแบบ LSTM จากข้อมูลรายชั่วโมง จำนวน 48 ชั่วโมงก่อนหน้า ในแต่ละกราฟหากเส้นกราฟสีฟ้าและเส้นกราฟสีส้มแสดงแนวโน้มใกล้เคียงกันมากกว่าจะแสดงให้เห็นถึงความแม่นยำในการทำงานที่สูงกว่า ในรูปที่ 4.1 (ซ้าย) แสดงผลลัพธ์ที่ได้จากการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 1 ชั่วโมงถัดไป ด้วยค่าความเข้มข้นของ PM2.5 รายชั่วโมงในอดีตจำนวน 48 ชั่วโมงก่อนหน้า จากกราฟพบว่าเส้นกราฟมีแนวโน้มเพิ่มขึ้นและ

ลดลงเหมือนกันในทุกช่วงเวลาจนแทบจะเท่ากันทั้งสองเส้น แสดงให้เห็นประสิทธิภาพในการทำนายที่แม่นยำมาก ส่วนรูปที่ 4.2 (ขวา) แสดงผลการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 2 ชั่วโมงถัดไป ด้วยค่าความเข้มข้นของ PM2.5 รายชั่วโมงในอดีตจำนวน 5 ชั่วโมงก่อนหน้า กราฟยังคงมีแนวโน้มเพิ่มขึ้นและลดลงในทำนองเดียวกัน แต่เห็นได้ชัดเจนว่าระดับของเส้นทั้งสองมีการแตกต่างกันมากกว่ารูปที่ 4.1 (ซ้าย)



รูปที่ 4.2 (ซ้าย) การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 3 ชั่วโมงถัดไป และ (ขวา) การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 4 ชั่วโมงถัดไป โดยใช้ข้อมูล 48 ชั่วโมงก่อนหน้า



รูปที่ 4.3 (ซ้าย) การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 5 ชั่วโมงถัดไป และ (ขวา) การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 6 ชั่วโมงถัดไป โดยใช้ข้อมูล 48 ชั่วโมงก่อนหน้า

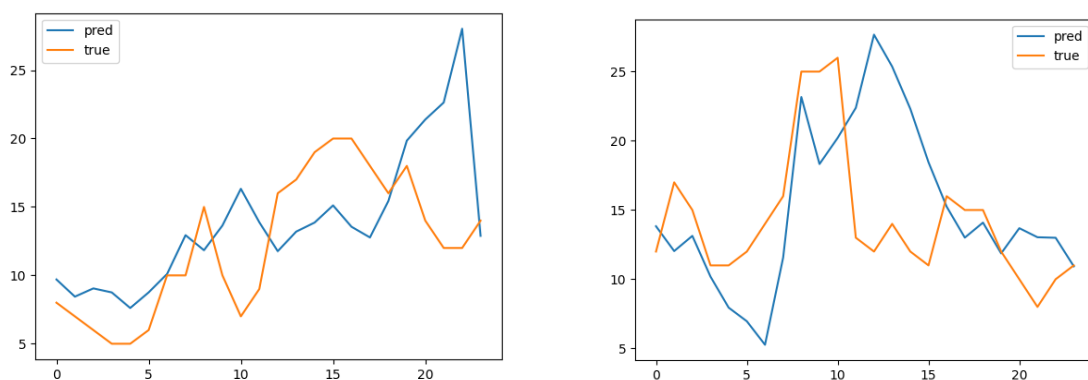
หากขยายช่วงเวลาการพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 3 หรือ 4 ชั่วโมงถัดไป ดังรูปที่ 4.2 กราฟยังคงมีแนวโน้มเพิ่มขึ้นและลดลงในทำนองเดียวกัน แต่เห็นได้ชัดเจนว่าระดับของเส้นทั้งสองมีการแตกต่างกันมากกว่าขึ้น และถ้าขยายช่วงเวลาการพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 5 หรือ 6 ชั่วโมงถัดไป ดังรูปที่ 4.3 (ซ้าย) และ (ขวา) ตามลำดับ จะพบว่า แนวโน้มของกราฟทั้งสองเส้นไม่สอดคล้องกันเลย ทำให้เราทราบถึงขีดจำกัดของเทคนิค LSTM ว่า สามารถใช้พยากรณ์ค่าความเข้มข้นของ

PM2.5 เป็นรายชั่วโมงได้ล่วงหน้าประมาณ 1 หรือ 2 ชั่วโมงถัดไป ได้แม่นยำกว่าการพยากรณ์ล่วงหน้าเป็นรายชั่วโมง 5 หรือ 6 ชั่วโมงถัดไป

จากผลการศึกษาการพยากรณ์ สรุปได้ว่าเทคนิค LSTM สามารถพยากรณ์ค่าความเข้มข้นของ PM2.5 ในชั่วโมงถัดไปได้อย่างแม่นยำ แต่ต้องใช้ข้อมูลหลายชั่วโมงก่อนหน้านี้ในการฝึกฝนตัวแบบ และถ้าพยากรณ์รายชั่วโมงล่วงหน้าที่ย่างจากชุดข้อมูลในอดีตมากขึ้น ความแม่นยำของค่าพยากรณ์จะลดลง จึงไม่เหมาะที่จะใช้พยากรณ์ล่วงหน้ามากกว่า 4 ชั่วโมง อย่างไรก็ตาม พยากรณ์ค่าความเข้มข้นของ PM2.5 เพียง 1 หรือ 2 ชั่วโมงล่วงหน้าในทางปฏิบัติอาจไม่สามารถนำไปใช้ในการประโยชน์ได้มากเท่าไรนัก หากต้องการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงล่วงหน้าหรือรายวัน จะต้องใช้ข้อมูลในอดีตจำนวนมากขึ้น ดังรายละเอียดในหัวข้อถัดไป

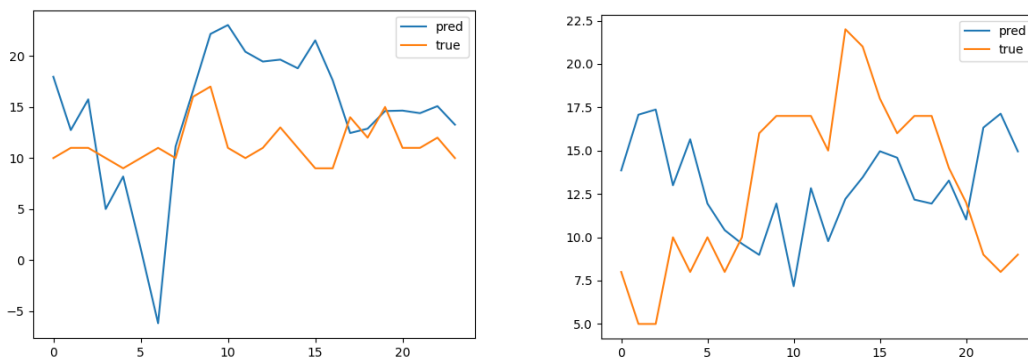
4.2 การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป โดยใช้ประวัติข้อมูลเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้านี้ ด้วยเทคนิค LSTM

การพยากรณ์ค่าความเข้มข้นของ PM2.5 โดยใช้ข้อมูลเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้านี้ เพื่อการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป เป็นการนำข้อมูลระดับฝุ่นละออง PM2.5 ก่อนหน้าช่วงเวลาที่จะพยากรณ์มาปรับปรุงโดยอาศัยรูปแบบของความผันแปรของของ PM2.5 ในช่วงเวลาต่าง ๆ ของแต่ละวัน ผู้วิจัยใช้ข้อมูลค่าความเข้มข้นของ PM2.5 รายชั่วโมง จากสถานีตรวจวัดคุณภาพอากาศเขตดินแดง ในช่วงวันที่ 1 ตุลาคม 2565 – 31 ตุลาคม 2565 ตุลาคม 2565 มาใช้ในการฝึกฝนตัวแบบ LSTM ผลการศึกษา แสดงดังรูปที่ 4.4

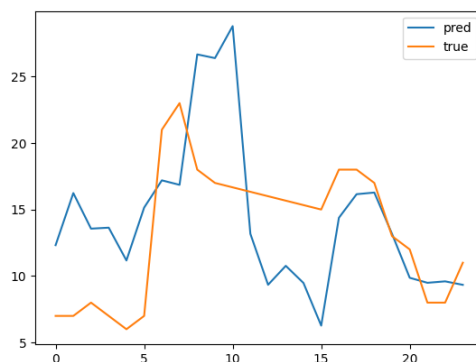


รูปที่ 4.4 (ซ้าย) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงถัดไป จำนวน 1 วัน และ (ขวา) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงถัดไป จำนวน 2 วัน โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วัน

จากรูปที่ 4.4 แขนงอนแสดงช่วงเวลาในการพยากรณ์ แขนงตั้งแสดงค่าความเข้มข้นของ PM2.5 เส้นกราฟสีส้มแสดงค่าสังเกตของความเข้มข้นของ PM2.5 และเส้นกราฟสีฟ้าแสดงค่าพยากรณ์ของความเข้มข้นของ PM2.5 ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 1 หรือ 2 วัน โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วัน จากรูป 4.4 (ซ้าย) จะพบว่า พยากรณ์ของความเข้มข้นของ PM2.5 ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 1 วัน หรือ $n=1$ กราฟของค่าพยากรณ์จะไม่ซ้อนทับกับกราฟของค่าจริงมากนัก ส่วนรูป 4.4 (ขวา) จะพบว่า พยากรณ์ของความเข้มข้นของ PM2.5 ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 2 วัน หรือ $n=2$ กราฟของค่าพยากรณ์จะมีค่าแตกต่างจากกราฟของค่าจริงมากกว่า



รูปที่ 4.4 (ซ้าย) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงถัดไป จำนวน 3 วัน และ (ขวา) การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงถัดไป จำนวน 4 วัน



รูปที่ 4.5 การพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงถัดไป จำนวน 5 วัน

จากรูปที่ 4.5 เส้นกราฟสีส้มแสดงค่าสังเกตของความเข้มข้นของ PM_{2.5} และเส้นกราฟสีฟ้าแสดงค่าพยากรณ์ของความเข้มข้นของ PM_{2.5} ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 3 หรือ 4 วัน โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วัน จากรูป 4.4 (ซ้าย) จะพบว่า พยากรณ์ของความเข้มข้นของ PM_{2.5} ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 3 วัน หรือ $n=3$ กราฟของค่าพยากรณ์บางช่วงเวลามีค่าแตกต่างจากค่าจริงพอสมควร ส่วนรูป 4.5 (ขวา) จะพบว่า พยากรณ์ของความเข้มข้นของ PM_{2.5} ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 4 วัน หรือ $n=4$ กราฟของค่าพยากรณ์จะมีค่าแตกต่างจากกราฟของค่าจริงค่อนข้างมาก และเห็นความแตกต่างระหว่างกราฟของค่าพยากรณ์และกราฟของค่าจริงได้มากขึ้นในรูปที่ 4.6 เมื่อพยากรณ์ของความเข้มข้นของ PM_{2.5} ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 5 วัน

ผลลัพธ์ข้างต้นจะเห็นว่า แนวโน้มที่ได้จากกราฟสามารถอธิบายการเพิ่มขึ้นและลดลงของปริมาณฝุ่นได้ในระยะเวลาที่จำกัด แม้จะมีช่วงเวลาที่น่าเชื่อถือมากกว่าการพยากรณ์ล่วงหน้าเป็นรายชั่วโมงจำนวนไม่กี่ชั่วโมง แต่เมื่อเพิ่มจำนวนวันที่ต้องการการพยากรณ์ล่วงหน้า ผลลัพธ์จากการพยากรณ์มีความแม่นยำลดลงเช่นเดียวกัน ดังนั้นการเรียนรู้จากรูปแบบของการเพิ่มขึ้นและลดลงของค่าความเข้มข้นของ PM_{2.5} ในอดีตเพื่อค่าพยากรณ์ความเข้มข้นของ PM_{2.5} ในช่วงรอบ 24 ชั่วโมงถัดไป จำเป็นต้องใช้ข้อมูลล่วงหน้าหลายวันในการฝึกฝนตัวแบบ

4.3 การพยากรณ์ค่าความเข้มข้นของ PM_{2.5} รายวัน ด้วยข้อมูลอนุกรมเวลาหลายชุด ด้วยตัวแบบ VAR

ในหัวข้อนี้จะศึกษาการพยากรณ์ค่าความเข้มข้นของ PM_{2.5} รายวัน ด้วยข้อมูลอนุกรมเวลาหลายชุด ประกอบด้วย ข้อมูลมลพิษอากาศ ได้แก่ ค่า PM_{2.5}, PM₁₀, NO₂, CO และ O₃ โดยใช้ข้อมูลคุณภาพอากาศของกรุงเทพมหานคร จำนวน 4 เขตที่มีสถานีตรวจวัดคุณภาพอากาศแบบตู้คอนเทนเนอร์ จำนวน 4 สถานี คือ สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง สถานีบริเวณสำนักงานเขตราชวัชรปุณณะ และสถานีบริเวณสำนักงานเขตราชเทวี ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566 จำนวน 911 วัน ผลการศึกษามีรายละเอียดดังต่อไปนี้

4.3.1 การสำรวจข้อมูลเบื้องต้นและการแสดงข้อมูลด้วยภาพของข้อมูล (Exploratory data analysis and data visualization)

ข้อมูลคุณภาพอากาศรายวันของสถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง สถานีบริเวณสำนักงานเขตราชวัชรปุณณะ และสถานีบริเวณสำนักงานเขตราชเทวี ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566 สรุปค่าสถิติพรรณนาได้ดังแสดงในตารางที่ 4.6- 4.10

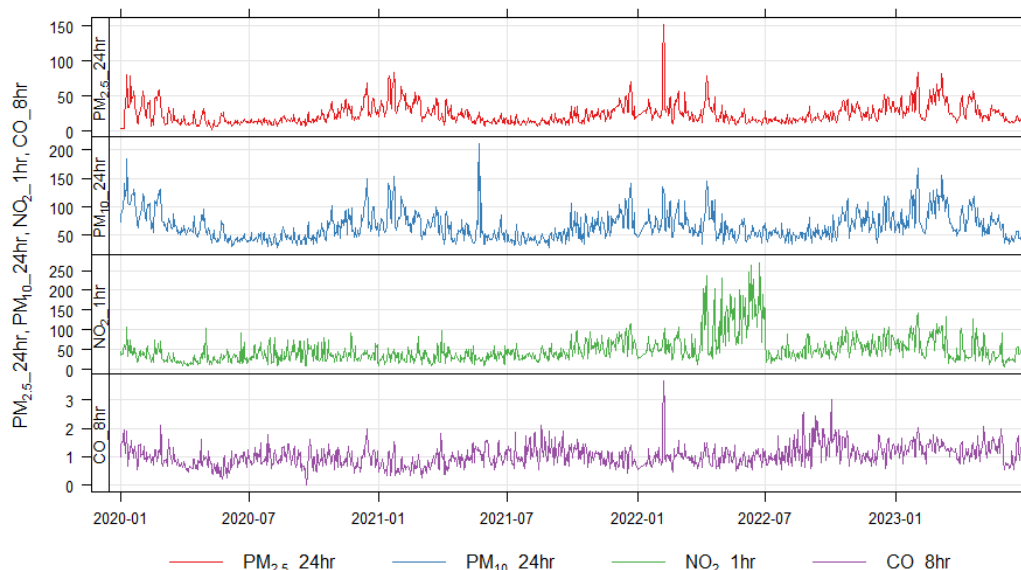
ตารางที่ 4.1 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศรายวันของสถานีดินแดง

ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	ค่าเบี่ยงเบนมาตรฐาน
PM _{2.5} (µg/m ³)	3.208	14.626	20.249	24.528	30.558	151.224	14.16
PM ₁₀ (µg/m ³)	28.04	47.75	59.63	65.76	77.35	210.91	24.11
NO ₂ (ppb)	7.00	27.76	40.08	49.45	61.12	270.62	34.46
CO (ppm)	0.0000	0.7861	1.0026	1.0412	1.2342	3.6703	0.36

จากตารางที่ 4.1 ข้อมูลคุณภาพอากาศรายวันของสถานีดินแดง พบว่า

- 1) ความเข้มข้นของ PM_{2.5} มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 3.208 และ 151.224 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 24.528 และ 14.16 µg/m³ ตามลำดับ
- 2) ความเข้มข้นของ PM₁₀ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 28.04 และ 210.91 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 65.76 และ 24.11 µg/m³ ตามลำดับ
- 3) ความเข้มข้นของ NO₂ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 7.00 และ 270.62 ppb ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 49.45 และ 34.46 ppb ตามลำดับ
- 4) ความเข้มข้นของ CO มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 0.0000 และ 3.6703 ppm ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.0412 และ 0.36 ppm ตามลำดับ
- 5)

Daily pollutant concentrations (DinDaeng)



รูปที่ 4.6 ความเข้มข้นของ PM_{2.5}, PM_{2.5}, NO₂ และ CO รายวันจากสถานีดินแดง

ตารางที่ 4.2 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศรายวันของสถานีบริเวณสำนักงานเขตพระโขนง

ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	ค่าเบี่ยงเบนมาตรฐาน
PM _{2.5} (µg/m ³)	2.912	13.144	20.143	24.398	32.193	95.887	14.59
PM ₁₀ (µg/m ³)	11.89	34.08	49.29	54.06	54.06	170.50	25.64
NO ₂ (ppb)	3.841	19.820	34.000	39.595	56.228	123.920	22.92
CO (ppm)	0.1325	0.6450	1.2431	1.2741	1.8175	3.1888	0.65

จากตารางที่ 4.2 ข้อมูลคุณภาพอากาศรายวันของสถานีบริเวณสำนักงานเขตพระโขนง พบว่า

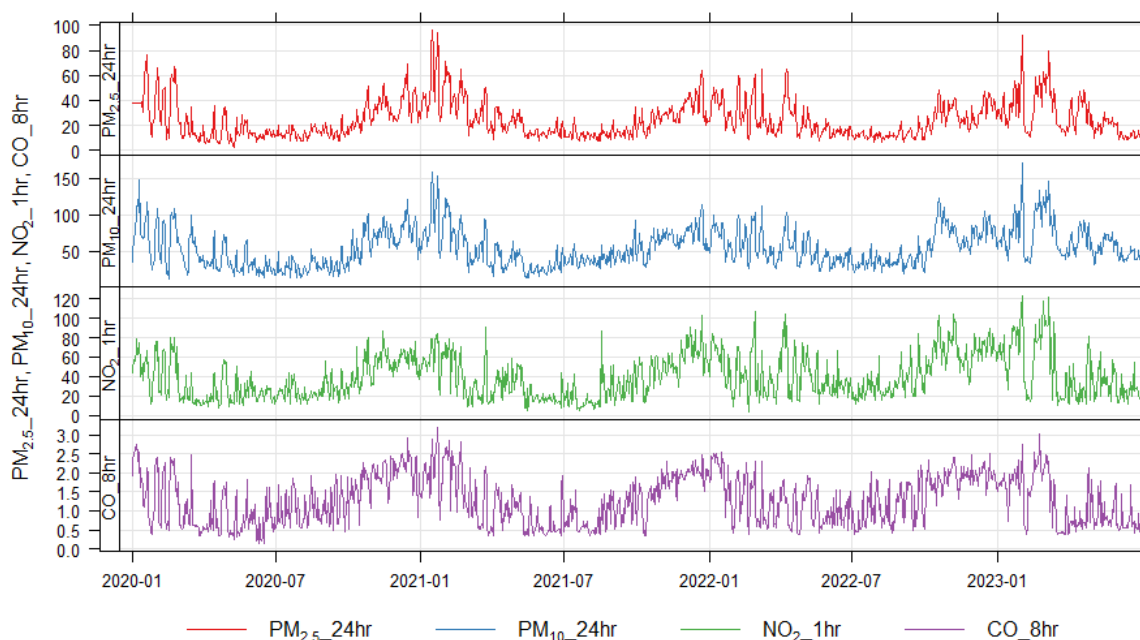
1) ความเข้มข้นของ PM_{2.5} มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 2.912 และ 95.887 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 24.398 และ 14.59 µg/m³ ตามลำดับ

2) ความเข้มข้นของ PM₁₀ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 11.89 และ 170.50 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 54.06 และ 25.64 µg/m³ ตามลำดับ

3) ความเข้มข้นของ NO₂ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 3.841 และ 123.920 ppb ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 39.595 และ 22.92 ppb ตามลำดับ

4) ความเข้มข้นของ CO มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 0.1325 และ 3.1888 ppm ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.2741 และ 0.65 ppm ตามลำดับ

Daily pollutant concentrations (PhraKhanong)



รูปที่ 4.7 ความเข้มข้นของ PM_{2.5}, PM_{2.5}, NO₂ และ CO รายวันจากสถานีบริเวณสำนักงานเขตพระโขนง

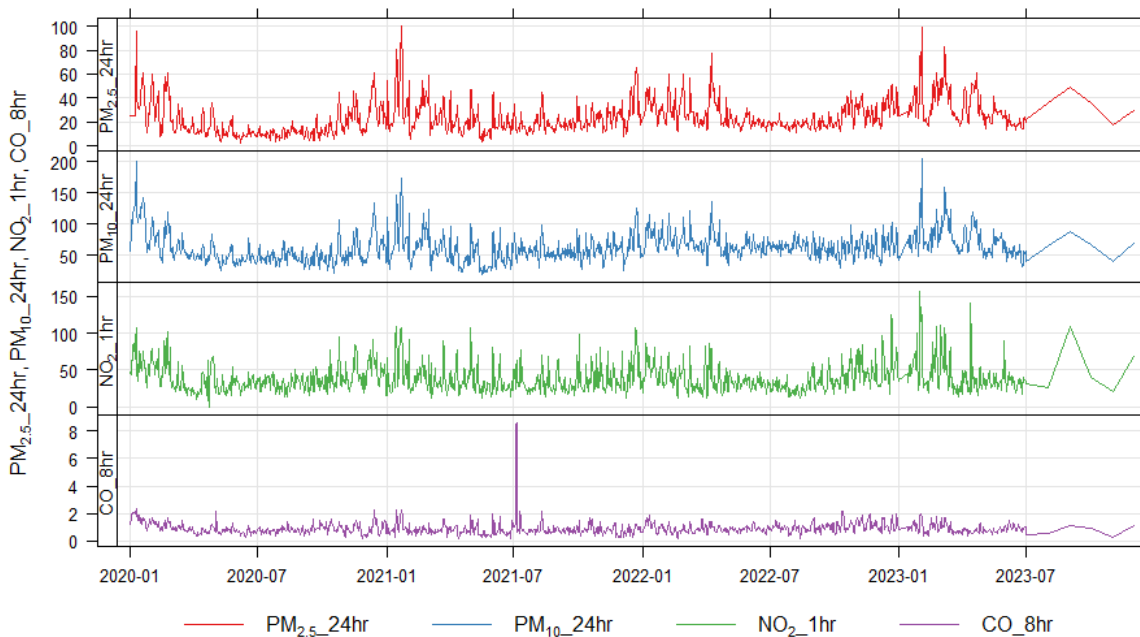
ตารางที่ 4.3 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศรายวันของสถานีบริเวณสำนักงานเขตราชบุรีบูรณะ

ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	ค่าเบี่ยงเบนมาตรฐาน
PM _{2.5} (µg/m ³)	2.708	14.444	20.250	23.612	29.406	99.833	13.33
PM ₁₀ (µg/m ³)	18.58	47.17	58.98	62.31	72.44	205.07	22.35
NO ₂ (ppb)	0.10	25.58	33.59	38.73	46.27	157.54	19.32
CO (ppm)	0.150	0.6569	0.8275	0.9112	1.0716	8.5878	0.47

จากตารางที่ 4.3 ข้อมูลคุณภาพอากาศรายวันของสถานีบริเวณสำนักงานเขตราชบุรีบูรณะ พบว่า

- 1) ความเข้มข้นของ PM_{2.5} มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 2.708 และ 99.833 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 23.612 และ 13.33 µg/m³ ตามลำดับ
- 2) ความเข้มข้นของ PM₁₀ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 18.58 และ 205.07 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 62.31 และ 22.35 µg/m³ ตามลำดับ
- 3) ความเข้มข้นของ NO₂ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 0.10 และ 157.54 ppb ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 38.73 และ 19.32 ppb ตามลำดับ
- 4) ความเข้มข้นของ CO มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 0.150 และ 8.588 ppm ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.912 และ 0.47 ppm ตามลำดับ

Daily pollutant concentrations (Ratburana)



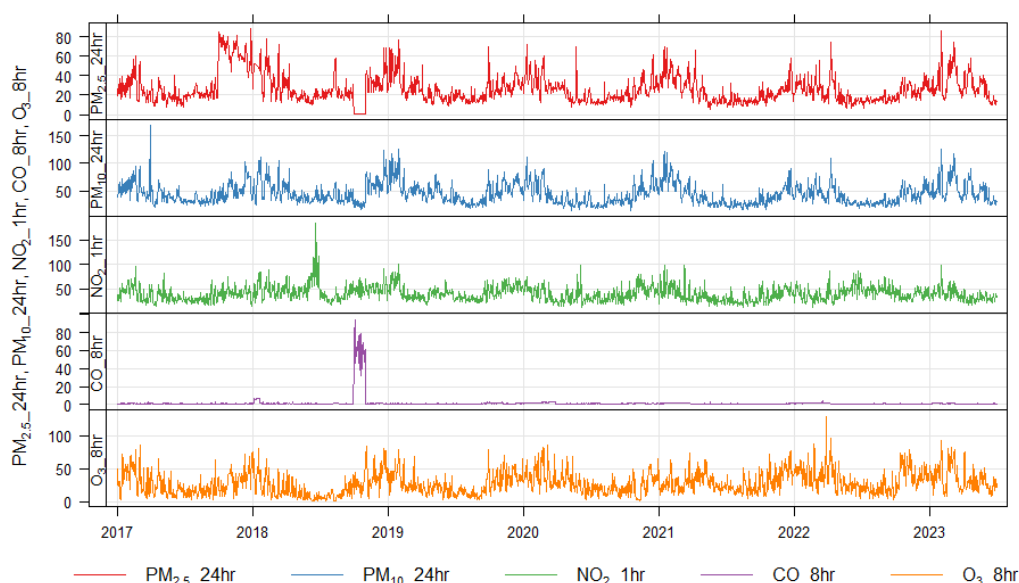
รูปที่ 4.8 ความเข้มข้นของ PM_{2.5}, PM₁₀, NO₂ และ CO รายวันจากสถานีบริเวณสำนักงานเขตราชบุรีบูรณะ

ตารางที่ 4.4 ค่าสถิติพรรณนาของข้อมูลคุณภาพอากาศรายวันของสถานีบริเวณสำนักงานเขตราชเทวี

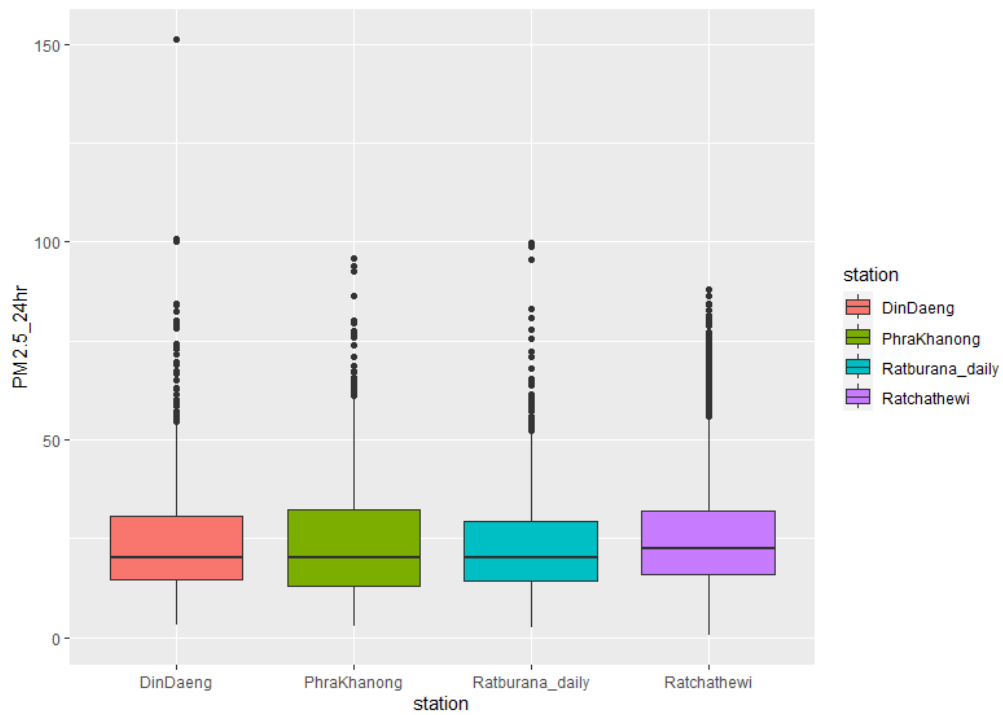
ตัวแปร	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	ค่าเบี่ยงเบนมาตรฐาน
PM _{2.5} (µg/m ³)	0.5975	16.0417	22.5417	26.1274	31.9167	87.8333	14.45
PM ₁₀ (µg/m ³)	13.46	29.62	39.79	44.16	54.71	168.69	18.89
NO ₂ (ppb)	12.00	29.00	37.00	40.02	48.00	185.00	15.06
CO (ppm)	0.1487	0.6525	0.8287	1.7481	1.1038	94.7500	7.01
O ₃ (ppm)	1.212	15.250	24.250	28.074	38.125	128.938	17.12

จากตารางที่ 4.4 ข้อมูลคุณภาพอากาศรายวันของสถานีบริเวณสำนักงานเขตราชเทวี พบว่า

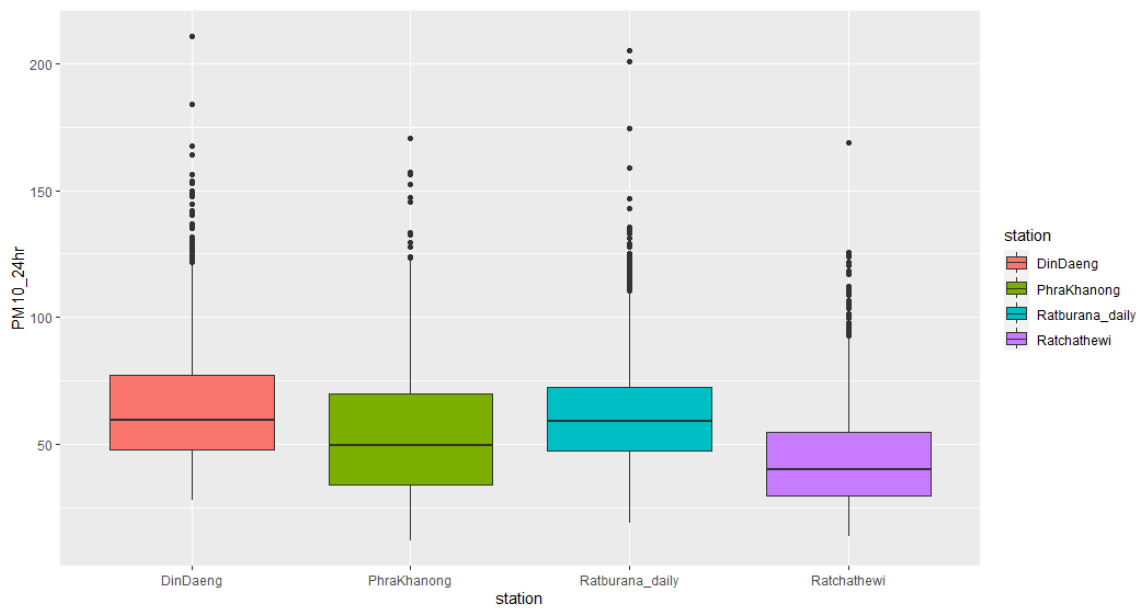
- 1) ความเข้มข้นของ PM_{2.5} มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 0.5975 และ 87.833 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 26.1274 และ 14.45 µg/m³ ตามลำดับ
- 2) ความเข้มข้นของ PM₁₀ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 13.46 และ 168.69 µg/m³ ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 44.16 และ 18.89 µg/m³ ตามลำดับ
- 3) ความเข้มข้นของ NO₂ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 12.00 และ 185.85 ppb ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 40.02 และ 15.06 ppb ตามลำดับ
- 4) ความเข้มข้นของ CO มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 0.1487 และ 94.7500 ppm ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.7481 และ 7.01 ppm ตามลำดับ
- 5) ความเข้มข้นของ O₃ มีค่าต่ำสุดและค่าสูงสุดเท่ากับ 1.212 และ 128.938 ppm ตามลำดับ และมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 28.074 และ 17.12 ppm ตามลำดับ



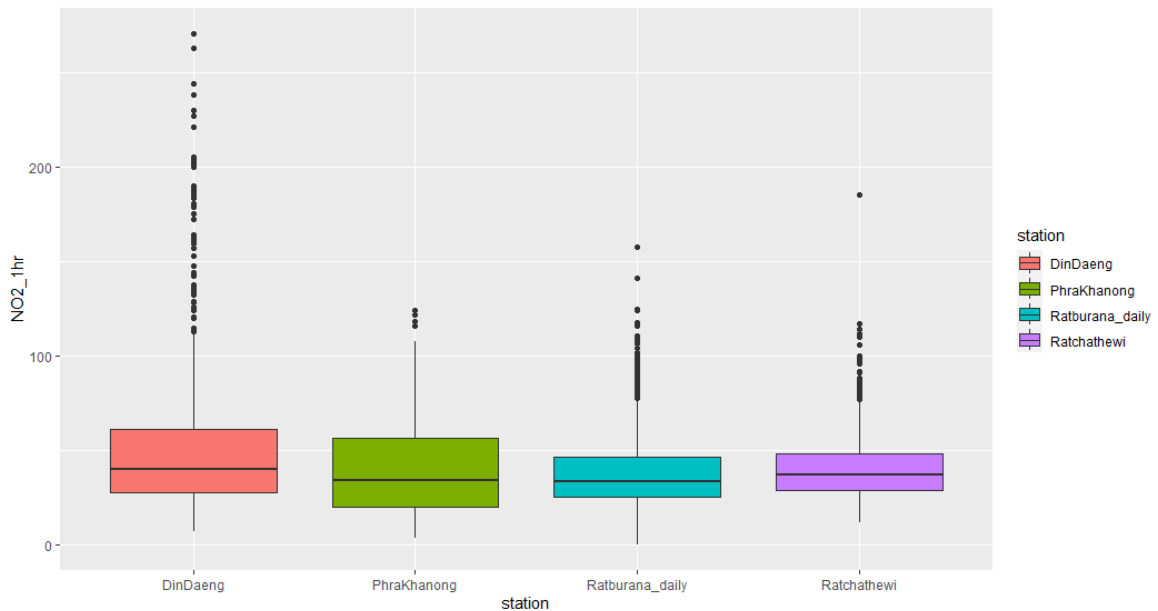
รูปที่ 4.10 ความเข้มข้นของ PM_{2.5}, PM₁₀, NO₂ และ CO รายวันจากสถานีบริเวณสำนักงานเขตราชเทวี



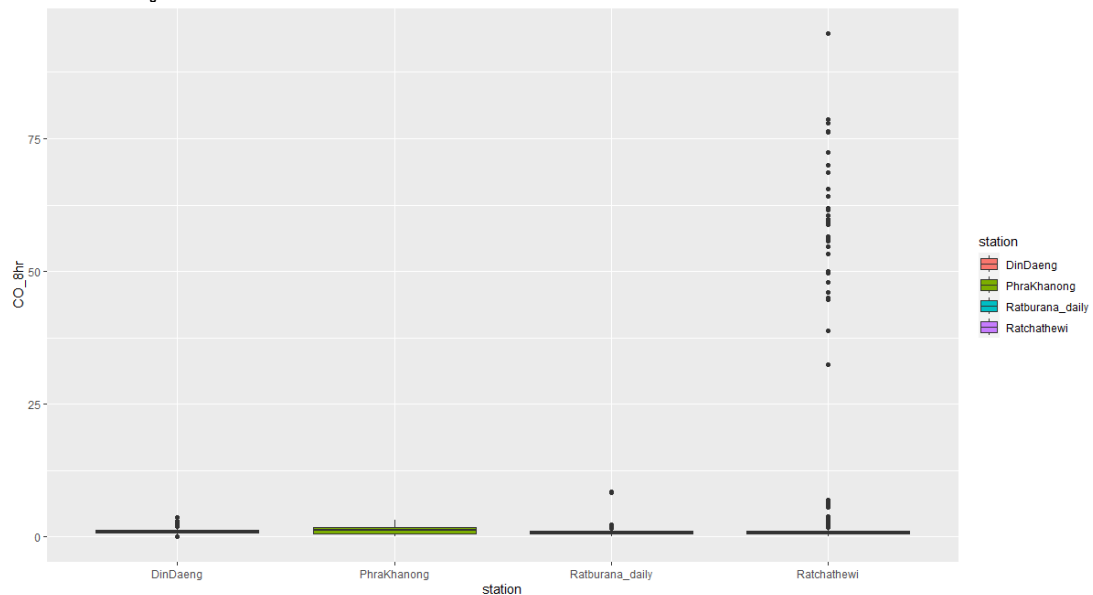
รูปที่ 4.11 แผนภาพกล่องของความเข้มข้นของ PM_{2.5} รายวันของ 4 สถานี



รูปที่ 4.12 แผนภาพกล่องของความเข้มข้นของ PM₁₀ รายวันของ 4 สถานี



รูปที่ 4.13 แผนภาพกล่องของความเข้มข้นของ PM₁₀ รายวันของ 4 สถานี



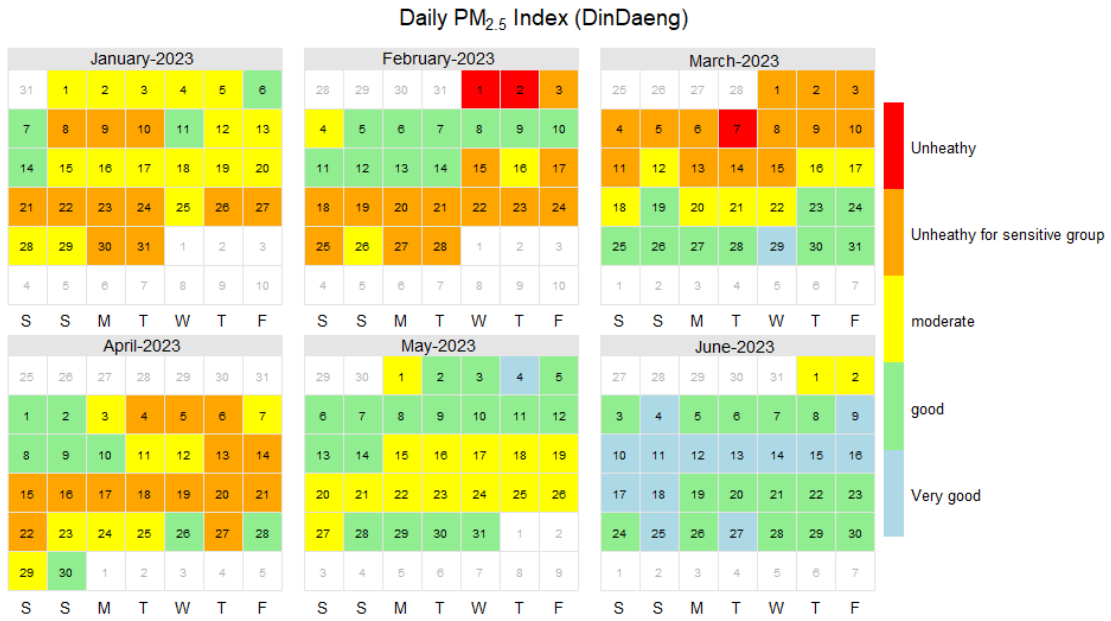
รูปที่ 4.14 แผนภาพกล่องของความเข้มข้นของ PM₁₀ รายวันของ 4 สถานี

4.3.2 ดัชนีคุณภาพอากาศรายวันของมลพิษอากาศ

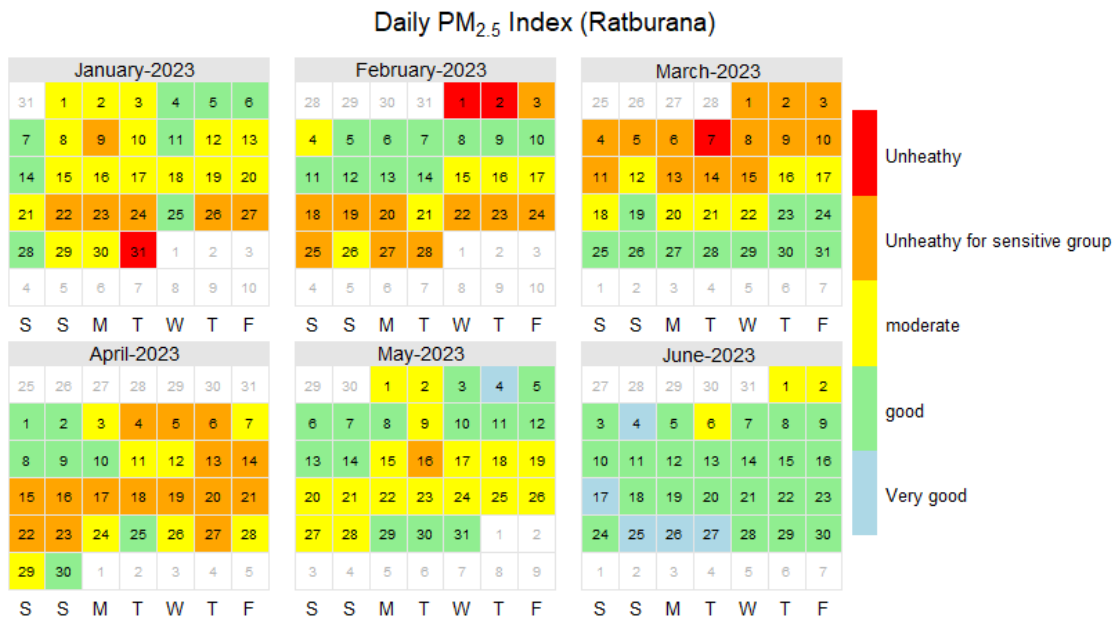
จากข้อมูลความเข้มข้นของมลพิษอากาศ 5 ชนิด ของสถานีต่าง ๆ สามารถนำมาคำนวณดัชนีคุณภาพอากาศย่อยของแต่ละมลพิษอากาศตามที่กล่าวไว้ในหัวข้อ 2.1 ซึ่งนำมาแสดงแผนภาพของดัชนีคุณภาพอากาศจาก 4 สถานี ตั้งแต่วันที่ 1 มกราคม 2563 ถึง 30 มิถุนายน 2566 โดยใช้แผนภาพปฏิทิน (calendarplot) ดังรูปที่ 4.15 – 4.18 และแผนภาพกล่อง (Boxplot) ของดัชนีคุณภาพอากาศรายวันจาก 4 สถานี ตั้งแต่วันที่ 1 มกราคม 2566 ถึง 30 มิถุนายน 2566 ดังรูปที่ 4.29 – 4.22

ตารางที่ 4.5 ค่าสถิติพรรณนาของดัชนีคุณภาพอากาศจากสถานีทั้ง 4 สถานี

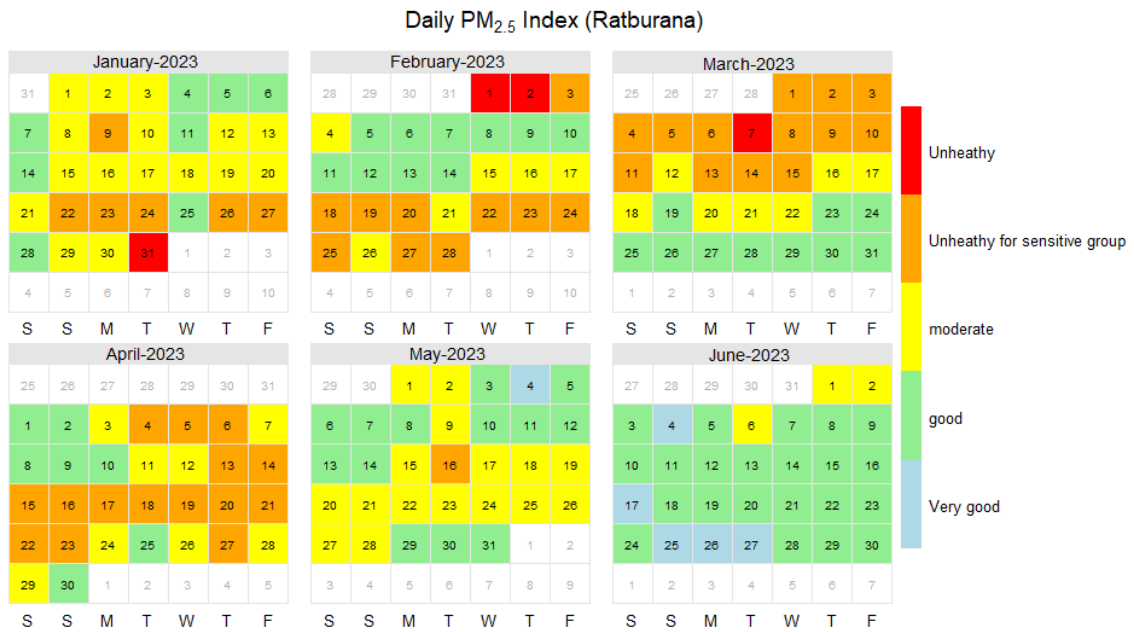
ดัชนีคุณภาพอากาศ (I)	ค่าต่ำสุด	ค่าควอร์ไทล์ที่ 1	ค่ามัธยฐาน	ค่าเฉลี่ย	ค่าควอร์ไทล์ที่ 3	ค่าสูงสุด	ค่าเบี่ยงเบนมาตรฐาน
สถานีดินแดง							
PM2.5	5.00	24.00	38.00	54.55	72.75	277.00	
PM10	14.00	24.00	33.00	41.13	48.00	231.00	
NO ₂	3.00	12.00	17.00	22.31	26.00	159.00	
CO	0.00	4.00	6.00	5.92	7.00	21.00	
สถานีบริเวณสำนักงานเขตพระโขนง							
PM2.5	5.00	22.00	38.00	55.12	79.00	222.00	
PM10	6.00	17.00	25.00	32.42	42.00	184.00	
NO ₂	2.00	8.00	14.00	16.98	23.00	64.00	
CO	1.00	4.00	7.00	7.24	10.00	18.00	
สถานีบริเวณสำนักงานเขตราชบุรีบูรณะ							
PM2.5	5.00	24.00	39.00	52.30	68.00	226.00	
PM10	9.00	24.00	33.00	37.85	44.00	225.00	
NO ₂	0.00	11.00	14.00	16.49	19.00	90.00	
CO	1.00	4.00	5.00	5.25	6.00	92.00	
สถานีบริเวณสำนักงานเขตราชเทวี							
PM2.5	1.00	28.00	44.00	59.13	78.00	214.00	
PM10	7.00	15.00	20.00	24.18	29.00	181.00	
NO ₂	5.00	12.00	15.00	16.86	20.00	109.00	
CO	1.00	4.00	5.00	8.54	6.00	266.00	
O ₃	1.00	11.00	17.00	25.10	30.00	209.00	



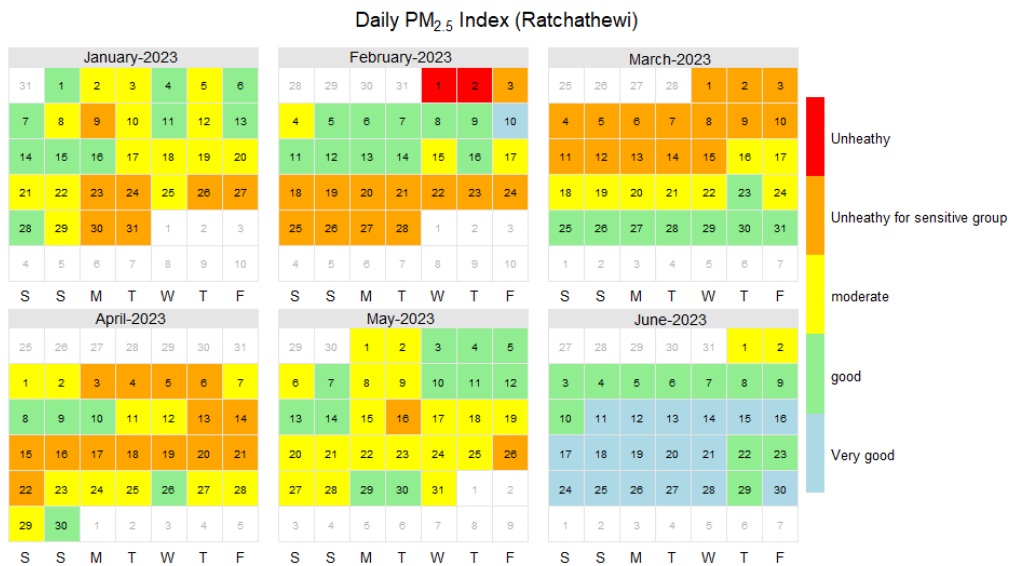
รูปที่ 4.15 แผนภาพปฏิทินของความเข้มข้นของ PM2.5 รายวันของสถานีดินแดง



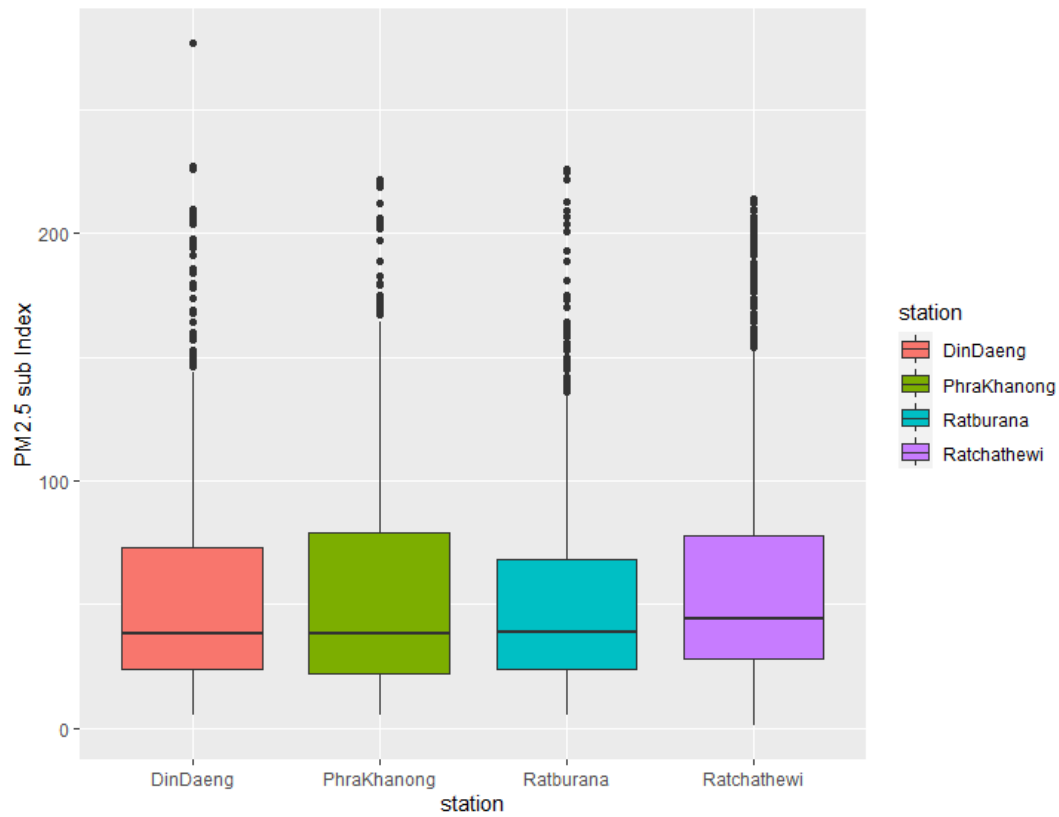
รูปที่ 4.16 แผนภาพปฏิทินของความเข้มข้นของ PM2.5 รายวันของสถานีบริเวณสำนักงานเขตพระโขนง



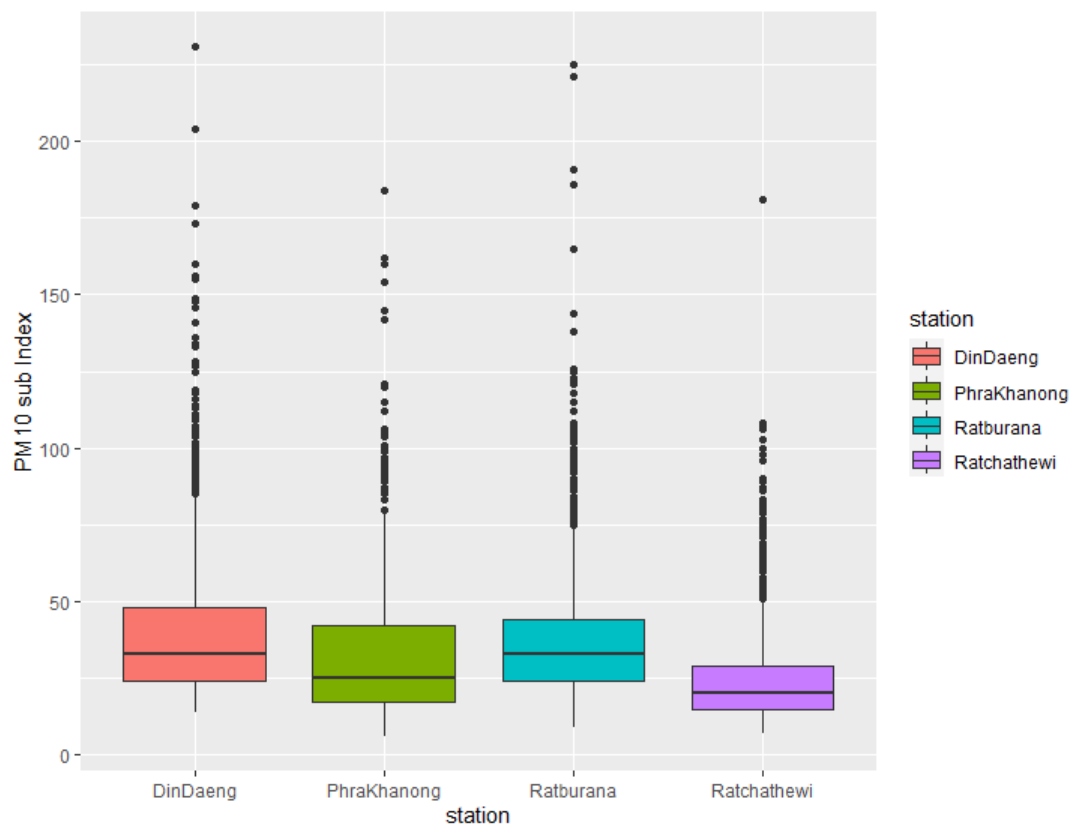
รูปที่ 4.17 แผนภาพปฏิทินของความเข้มข้นของ PM_{2.5} รายวันของสถานีบริเวณสำนักงานเขตราชบุรีบูรณะ



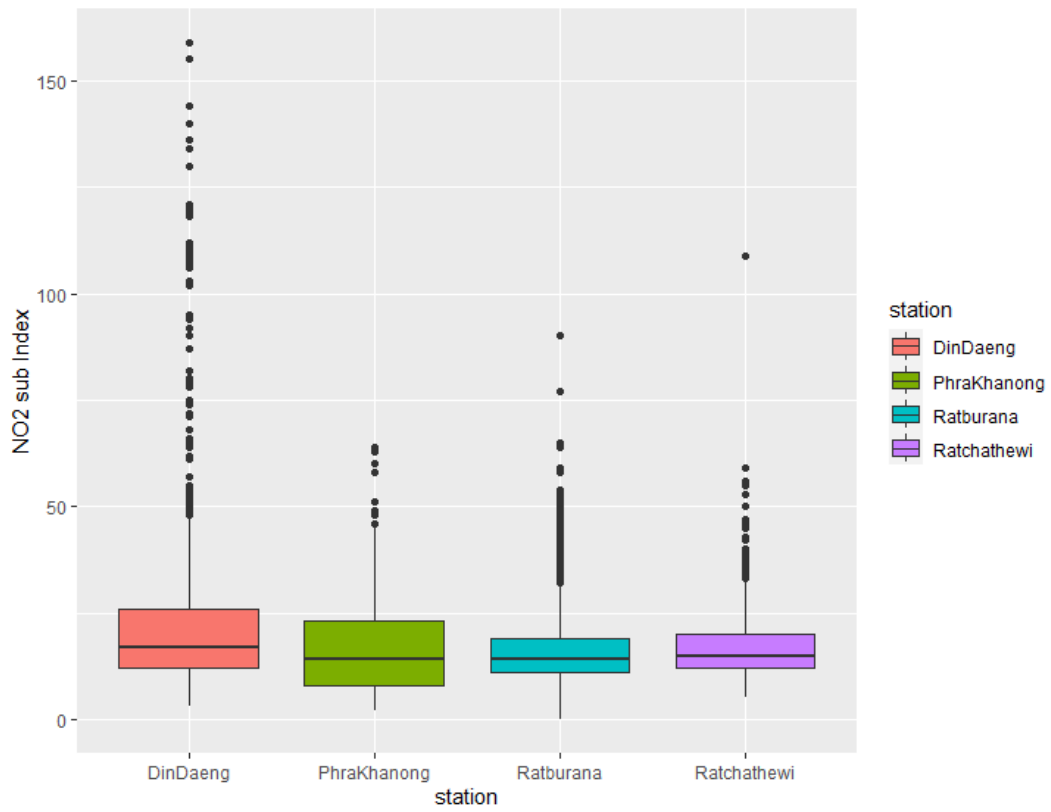
รูปที่ 4.18 แผนภาพปฏิทินของความเข้มข้นของ PM_{2.5} รายวันของสถานีบริเวณสำนักงานเขตราชเทวี



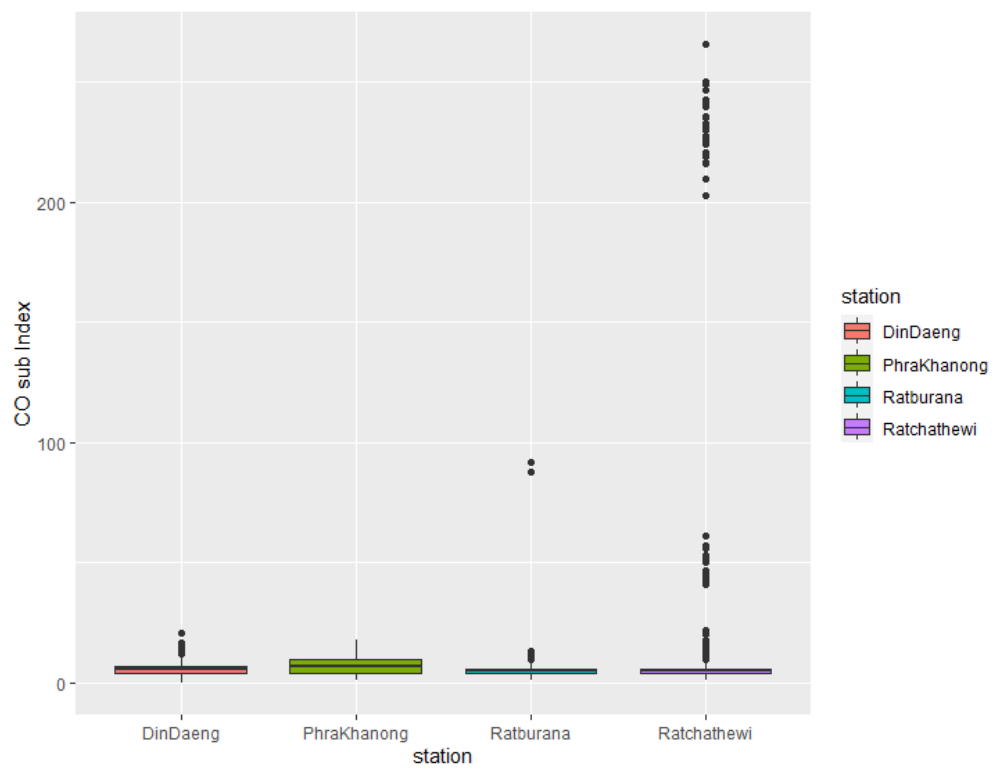
รูปที่ 4.19 แผนภาพกล่องของดัชนีคุณภาพอากาศของ PM_{2.5} รายวันของ 4 สถานี



รูปที่ 4.20 แผนภาพกล่องของดัชนีคุณภาพอากาศของ PM₁₀ รายวันของ 4 สถานี



รูปที่ 4.21 แผนภาพกล่องของดัชนีคุณภาพอากาศของ NO2 รายวันของ 4 สถานี



รูปที่ 4.22 แผนภาพกล่องของดัชนีคุณภาพอากาศของ CO รายวันของ 4 สถานี

4.3.3 การวิเคราะห์ปัจจัยที่ส่งผลต่อความเข้มข้นของมลพิษอากาศด้วยตัวแบบ VAR

การวิเคราะห์ปัจจัยที่ส่งผลต่อความเข้มข้นของมลพิษอากาศ 5 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซโอโซน (O3) และก๊าซไนโตรเจนไดออกไซด์ (NO2) ของข้อมูลคุณภาพอากาศจากสถานีทั้ง 4 สถานี ใช้วิธีการวิเคราะห์ที่เรียกว่า ตัวแบบเวกเตอร์การถดถอยในตัว (Vector autoregressive model) หรือตัวแบบ VAR โดยผู้วิจัยแบ่งชุดอนุกรมเวลารายวันที่จะทำการหาตัวแบบออกเป็นช่วงระยะเวลา 1 ปี คือ ตั้งแต่ 1 มกราคม 2565 ถึง 31 ธันวาคม 2566

ก่อนที่จะทำการวิเคราะห์ตัวแบบ VAR จะต้องตรวจสอบความคงที่ของอนุกรมเวลาด้วยการทดสอบยูนิตรูท ซึ่งผลการทดสอบยูนิตรูทสำหรับความเข้มข้นมลพิษทางอากาศ 5 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซโอโซน (O3) และก๊าซไนโตรเจนไดออกไซด์ (NO2) ของข้อมูลคุณภาพอากาศจากสถานีทั้ง 4 สถานี ด้วยการทดสอบ ADF test แล้วจึงหาตัวแบบในการพยากรณ์ความเข้มข้นของมลพิษแต่ละตัวจากข้อมูล จากนั้นนำสมการที่ได้ไปหาค่าพยากรณ์ 15 วันล่วงหน้า คือ วันที่ 1 มกราคม 2566 ถึง 15 มกราคม 2566 รายละเอียดของผลการวิเคราะห์ข้อมูลมีดังนี้

4.3.3.1 ผลการวิเคราะห์ตัวแบบเวกเตอร์การถดถอยในตัวสำหรับข้อมูลคุณภาพอากาศจากสถานีดินแดง

ผลการทดสอบยูนิตรูทสำหรับความเข้มข้นมลพิษทางอากาศ 5 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซไนโตรเจนไดออกไซด์ (NO2) และก๊าซคาร์บอนมอนอกไซด์ (CO) และก๊าซไนโตรเจนไดออกไซด์ (NO2) แสดงดังตารางที่ 4.6

ตารางที่ 4.6 การทดสอบยูนิตรูทของมลพิษทางอากาศจากข้อมูลคุณภาพอากาศจากสถานีดินแดง

ตัวแปร	Random walk	Random walk with a drift		Random walk with a drift and a time trend		
	(I) $H_0: \gamma = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(III) $H_0: \gamma = 0$ & $a_0 = 0$ & $a_2 = 0$
	t-test	t-test	F-test	t-test	F-test	F-test
PM2.5	-3.7716	-8.6968	37.8169	-8.7204	25.354	38.0308
PM10	-2.0378	-7.8327	30.6777	-7.8223	20.3979	30.5945
NO ₂	-3.332	-6.6781	22.2997	-6.7482	15.1856	22.7773
CO	-2.5127	-9.2694	42.9635	-9.817	32.1306	48.1933
ค่าวิกฤต ที่ $\alpha=0.05$	-1.95	-2.87	4.61	-3.42	4.71	6.30

จากตารางที่ 4.6 การทดสอบกระบวนการ Random walk ของอนุกรมเวลา สำหรับสมมติฐาน $H_0: \gamma = 0$ พบว่า ค่าสถิติทดสอบทีของความเข้มข้นของ PM2.5, PM10, NO₂, และ CO มีค่าเท่ากับ -3.7716, -2.0378, -3.332 และ -2.5127 ตามลำดับ ซึ่งทุกค่ามีค่าน้อยกว่าค่าวิกฤติ (ตกอยู่ในบริเวณวิกฤต) จึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 4 ชุดไม่มียูนิรูท และเมื่อทดสอบการทดสอบกระบวนการ Random walk with drift พบว่า สำหรับสมมติฐาน $H_0: \gamma = 0$ ค่าสถิติทดสอบทีของความเข้มข้นของ PM2.5, PM10, NO₂, CO และ O₃ สรุปผลได้อย่างเดียวกัน และการทดสอบสมมติฐาน (II) $H_0: \gamma = 0$ & $a_0 = 0$ พบว่า ค่าสถิติทดสอบ F ของความเข้มข้นของ PM2.5, PM10, NO₂ และ CO มีค่า 37.8169, 30.6777, 22.2997 และ 42.9635 ตามลำดับ ซึ่งตกอยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 4 ชุดไม่มียูนิรูทและมี drift

จากผลการทดสอบยูนิรูท สรุปได้ว่า อนุกรมเวลาของความเข้มข้นของมลพิษอากาศ ทั้ง 5 ชุดไม่มียูนิรูทและมี drift เราจึงสามารถใช้ตัวแบบ VAR ในการวิเคราะห์เพื่อหาสมการอธิบายความสัมพันธ์และพยากรณ์ความเข้มข้นของมลพิษอากาศ ทั้ง 4 ได้ ซึ่งผลการวิเคราะห์ แสดงดังตารางที่ 4.7

ตารางที่ 4.7 ผลการวิเคราะห์ตัวแบบ VAR ของข้อมูลคุณภาพอากาศจากสถานีดินแดง

ตัวแปรตาม : PM2.5 ($R^2 = 0.919$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	1.10992	0.05398	20.562	<2e-16	***
CO_8hr.l1	-3.14788	1.19214	-2.641	0.008655	**
PM2.5_24hr.l2	-0.50867	0.07165	-7.099	7.24e-12	***
PM2.5_24hr.l3	0.18816	0.05281	3.563	0.000418	***
const	8.6979	1.47183	5.91	8.26e-09	***
sd3	4.58248	1.57405	2.911	0.003835	**
ตัวแปรตาม : PM10 ($R^2 = 0.9625$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	0.54402	0.13124	4.145	4.28e-05	***
PM10_24hr.l1	0.54891	0.07126	7.702	1.45e-13	***
CO_8hr.l1	-7.09913	1.98301	-3.58	0.000393	***
PM2.5_24hr.l2	-0.38373	0.0878	-4.37	1.65e-05	***
PM2.5_24hr.l4	0.21144	0.06018	3.513	0.000502	***
const	28.08106	3.66629	7.659	1.93e-13	***
sd3	5.2883	2.58129	2.049	0.041252	*
ตัวแปรตาม : NO2 ($R^2 = 0.8725$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	0.54494	0.24905	2.188	0.02934	*
PM10_24hr.l1	-0.38236	0.1764	-2.168	0.03088	*
NO2_1hr.l1	0.54575	0.04822	11.319	<2e-16	***
CO_8hr.l2	-17.6815	4.41566	-4.004	7.63e-05	***
NO2_1hr.l3	0.1537	0.05212	2.949	0.00341	**
PM10_24hr.l4	-0.2404	0.10192	-2.359	0.0189	*
NO2_1hr.l4	0.12831	0.0579	2.216	0.02735	*
const	59.25217	10.14579	5.84	1.21e-08	***
ตัวแปรตาม : CO ($R^2 = 0.9274$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	0.011592	0.002148	5.398	1.26e-07	***
CO_8hr.l1	0.397171	0.050856	7.81	7.04e-14	***
PM2.5_24hr.l2	-0.00925	0.002159	-4.288	2.35e-05	***
NO2_1hr.l2	-0.00103	0.000378	-2.725	0.00675	**
PM2.5_24hr.l4	-0.00359	0.001709	-2.1	0.03642	*
CO_8hr.l4	0.22388	0.050598	4.425	1.30e-05	***
const	0.530384	0.0775	6.844	3.57e-11	***

*significant level at 0.05, **significant level at 0.01, ***significant level at 0.001

จากตารางที่ 4.7 พบว่า ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM2.5 อย่างมีนัยสำคัญทางสถิติ มี 5 ตัวแปร คือ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน, 2 วัน และ 3 วัน ความเข้มข้นของ CO ล่าช้า 1 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนมีนาคม ($sd_3 = 1$) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM2.5_t = 8.679 + 1.101 PM2.5_{t-1} - 3.148 CO_{t-1} - 0.509 PM2.5_{t-2} \\ + 0.0188 PM2.5_{t-3} + 4.582sd_3$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM10 อย่างมีนัยสำคัญทางสถิติ มี 6 ตัวแปร คือ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน, 2 วัน และ 4 วัน ความเข้มข้นของ PM10 ล่าช้า 1 วัน ความเข้มข้นของ CO ล่าช้า 1 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนมีนาคม ($sd_3 = 1$) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM10_t = 28.081 + 0.544 PM2.5_{t-1} + 0.549 PM10_{t-1} - 7.099 CO_{t-1} - 0.384 PM2.5_{t-2} \\ + 1.101 PM2.5_{t-4} + 5.288sd_3$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ NO2 อย่างมีนัยสำคัญทางสถิติ มี 7 ตัวแปร คือ ความเข้มข้นของ NO2 ล่าช้า 1 วัน และความเข้มข้นของ PM2.5 ล่าช้า 2 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$NO2_t = 59.252 + 0.545 PM2.5_{t-1} - 0.382 PM10_{t-1} + 0.546 NO2_{t-1} \\ - 17.68 CO_{t-2} + 0.154 NO_{t-3} - 0.240 PM2.5_{t-4} + 0.12831 NO2_{t-4}$$

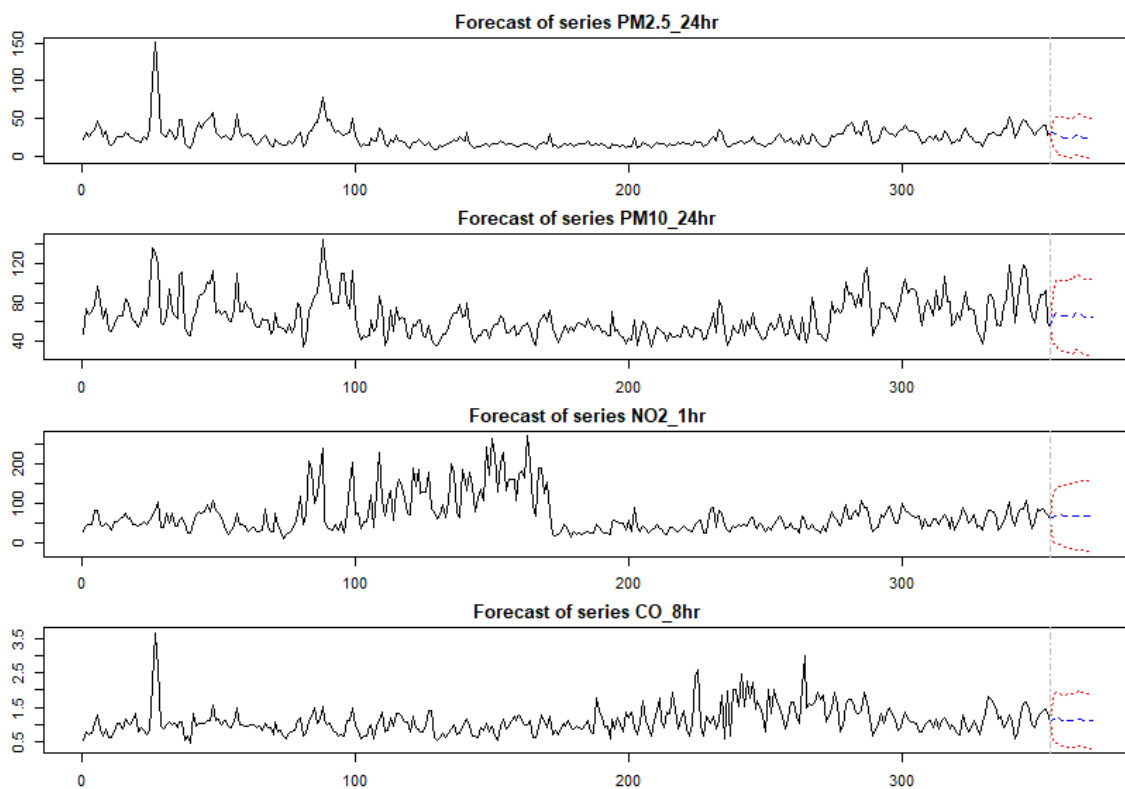
ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ CO อย่างมีนัยสำคัญทางสถิติ มี 6 ตัวแปร คือ ความเข้มข้นของ CO ล่าช้า 1 วัน ความเข้มข้นของ PM2.5 ล่าช้า 2 วัน และความเข้มข้นของ NO2 ล่าช้า 2 วันซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$CO_t = 0.530 + 0.116 PM2.5_{t-1} + 0.397 CO_{t-1} - 0.0009 PM2.5_{t-2} \\ - 0.0013 NO2_{t-2} + -0.003 PM2.5_{t-4} + 0.224 CO_{t-4}$$

จากสมการทั้ง 4 สมการข้างต้น สามารถคำนวณค่าพยากรณ์ของความเข้มข้นแต่ละชนิด ดังแสดงตารางที่ 4.8 และรูปที่ 4.23

ตารางที่ 4.8 ค่าพยากรณ์ของ PM2.5, PM10, NO2, CO และ O3 ของสถานีดินแดง ระหว่างวันที่ 1-15 มกราคม 2566

ลำดับ	date	PM2.5	PM10	NO2	CO
1	1/1/2023	31.20733	64.16754	66.08028	1.132593
2	1/2/2023	29.65525	69.27397	67.40273	1.178416
3	1/3/2023	27.20332	67.39912	69.83056	1.185976
4	1/4/2023	25.56341	65.8801	70.53004	1.110627
5	1/5/2023	24.93562	65.98484	68.92745	1.085799
6	1/6/2023	24.68979	66.17817	68.31822	1.098946
7	1/7/2023	24.38633	65.77969	69.08659	1.119268
8	1/8/2023	23.99247	64.99919	69.46907	1.11574
9	1/9/2023	28.257	69.65356	69.07782	1.108485
10	1/10/2023	28.57388	69.39073	69.46441	1.162116
11	1/11/2023	26.51344	67.33754	70.33001	1.153665
12	1/12/2023	24.89434	64.94473	69.69295	1.123717
13	1/13/2023	24.29925	64.6554	68.41762	1.094303
14	1/14/2023	24.16723	65.06996	68.2833	1.102231
15	1/15/2023	23.9938	64.9621	69.00637	1.116171



รูปที่ 4.23 กราฟแสดงการพยากรณ์ล่วงหน้า 15 วันของมลพิษอากาศ 4 ชนิดจากสถานีดินแดง

4.3.2 ผลการวิเคราะห์ตัวแบบเวกเตอร์การถดถอยในตัวสำหรับข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตพระโขนง

ผลการทดสอบยูนีรูทสำหรับความเข้มข้นมลพิษทางอากาศ 4 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซไนโตรเจนไดออกไซด์ (NO₂) และก๊าซคาร์บอนมอนอกไซด์ (CO) และก๊าซไนโตรเจนไดออกไซด์ (NO₂) แสดงดังตารางที่ 4.8

ตารางที่ 4.9 การทดสอบยูนีรูทของมลพิษทางอากาศจากข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตพระโขนง

ตัวแปร	Random walk	Random walk with a drift		Random walk with a drift and a time trend		
	(I) $H_0: \gamma = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(III) $H_0: \gamma = 0$ & $a_0 = 0$ & $a_2 = 0$
	t-test	t-test	F-test	t-test	F-test	F-test
PM2.5	-2.5775	-5.9628	17.7784	-5.9896	11.965	17.9466
PM10	-1.8871	-5.349	14.3061	-5.3797	9.6473	14.4705
NO ₂	-2.4293	-5.9495	17.6981	-6.2191	12.9008	19.3511
CO	-2.2554	-6.1209	18.7331	-6.3908	13.629	20.4432
ค่าวิกฤต ที่ $\alpha=0.05$	-1.95	-2.87	4.61	-3.42	4.71	6.30

จากตารางที่ 4.9 การทดสอบกระบวนการ Random walk ของอนุกรมเวลา สำหรับสมมติฐาน $H_0: \gamma = 0$ พบว่า ค่าสถิติทดสอบที่ของความเข้มข้นของ PM2.5 PM10 NO₂ และ CO มีค่าเท่ากับ -2.5775, -1.8871, -2.4293 และ -2.2554 ซึ่งมีค่าน้อยกว่าค่าวิกฤต (ตกอยู่ในบริเวณวิกฤต) จึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 4 ชุดไม่มียูนีรูท และเมื่อทดสอบการทดสอบกระบวนการ Random walk with drift พบว่า สำหรับสมมติฐาน $H_0: \gamma = 0$ ค่าสถิติทดสอบที่ของความเข้มข้นของ PM2.5, PM10 NO₂ CO และ O₃ สรุปผลได้อย่างเดียวกัน และการทดสอบสมมติฐาน (II) $H_0: \gamma = 0$ & $a_0 = 0$ พบว่า ค่าสถิติทดสอบ F ของความเข้มข้นของ PM2.5 PM10 NO₂ และ CO มีค่า 17.7784, 14.3061, 17.6981 และ 18.7331 ซึ่งตกอยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 4 ชุดไม่มียูนีรูทและมี drift

จากผลการทดสอบยูนีรูท สรุปได้ว่า อนุกรมเวลาของความเข้มข้นของมลพิษทางอากาศทั้ง 4 ชุดไม่มียูนีรูทและมี drift เราจึงสามารถใช้ตัวแบบ VAR ในการวิเคราะห์เพื่อหาสมการอธิบายความสัมพันธ์และพยากรณ์ความเข้มข้นของมลพิษทางอากาศ ทั้ง 4 ได้ ซึ่งผลการวิเคราะห์ แสดงดังตารางที่ 4.10

ตารางที่ 4.10 ผลการวิเคราะห์ตัวแบบ VAR ของข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตพระโขนง

ตัวแปรตาม : PM2.5 ($R^2 = 0.9425$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	0.7418	0.04383	16.923	<2e-16	***
NO2_1hr.l1	0.1946	0.03043	6.395	5.04e-10	***
CO_8hr.l1	-2.00908	0.93116	-2.158	0.031627	*
NO2_1hr.l2	-0.0836	0.02515	-3.324	0.000981	***
const	3.63487	0.85419	4.255	2.67e-05	***
sd2	2.72657	1.20242	2.268	0.023954	*
ตัวแปรตาม : PM10 ($R^2 = 0.9616$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM10_24hr.l1	0.72439	0.05031	14.399	<2e-16	***
NO2_1hr.l1	0.2533	0.05475	4.626	5.21e-06	***
CO_8hr.l1	-4.38573	1.74693	-2.511	0.0125	*
const	9.48385	1.67392	5.666	3.01e-08	***
ตัวแปรตาม : NO2 ($R^2 = 0.926$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM10_24hr.l1	0.19162	0.07104	2.697	0.00732	**
NO2_1hr.l1	0.63677	0.06118	10.408	<2e-16	***
PM2.5_24hr.l2	-0.25529	0.09622	-2.653	0.00833	**
CO_8hr.l2	4.1111	1.75348	2.345	0.0196	*
const	6.30416	2.06	3.06	0.00238	**
sd11	7.17351	2.64769	2.709	0.00707	**
ตัวแปรตาม : CO ($R^2 = 0.9409$)					
Variables	Estimate	Std. Error	t	p-value	significant
NO2_1hr.l1	0.008499	0.001647	5.159	4.13e-07	***
CO_8hr.l1	0.490669	0.058104	8.445	7.97e-16	***
PM10_24hr.l2	0.004885	0.002239	2.182	0.02976	*
NO2_1hr.l2	-0.00557	0.001857	-2.998	0.00291	**
PM10_24hr.l3	-0.00588	0.002011	-2.923	0.00369	**
CO_8hr.l3	0.262133	0.05603	4.678	4.12e-06	***
const	0.254436	0.056331	4.517	8.56e-06	***
sd11	0.147802	0.06931	2.132	0.03366	*

*significant level at 0.05, **significant level at 0.01, ***significant level at 0.001

จากตารางที่ 4.10 พบว่า ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM2.5 อย่างมีนัยสำคัญทางสถิติ มี 5 ตัวแปร คือ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน ความเข้มข้นของ NO2 ล่าช้า 1 วัน และ 2

วัน ความเข้มข้นของ CO ล่าช้า 1 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนกุมภาพันธ์ ($sd_2 = 1$ สำหรับเดือนกุมภาพันธ์ และ $sd_2 = 0$ เดือนอื่นๆ) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM2.5_t = 3.635 + 0.742 PM2.5_{t-1} + 0.195 NO2_{t-1} - 2.01 CO_{t-1} - 0.084 NO2_{t-2} + 2.727sd_2$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM10 อย่างมีนัยสำคัญทางสถิติ มี 3 ตัวแปร คือ ความเข้มข้นของ PM10 ล่าช้า 1 วัน ความเข้มข้นของ NO2 ล่าช้า 1 วัน และความเข้มข้นของ CO ล่าช้า 1 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM10_t = 9.484 + 0.724 PM10_{t-1} + 0.253 NO2_{t-1} - 4.386 CO_{t-1}$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ NO2 อย่างมีนัยสำคัญทางสถิติ มี 5 ตัวแปร คือ ความเข้มข้นของ PM10 ล่าช้า 1 วัน ความเข้มข้นของ NO2 ล่าช้า 1 วัน และความเข้มข้นของ PM2.5 ล่าช้า 2 วัน ความเข้มข้นของ CO ล่าช้า 2 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนพฤศจิกายน ($sd_{11} = 1$ สำหรับเดือนพฤศจิกายน และ $sd_{11} = 0$ เดือนอื่นๆ) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$NO2_t = 6.304 + 0.19162 PM10_{t-1} + 0.637 NO2_{t-1} - 0.255 PM2.5_{t-2} + 4.111 CO_{t-2} + 7.174sd_{11}$$

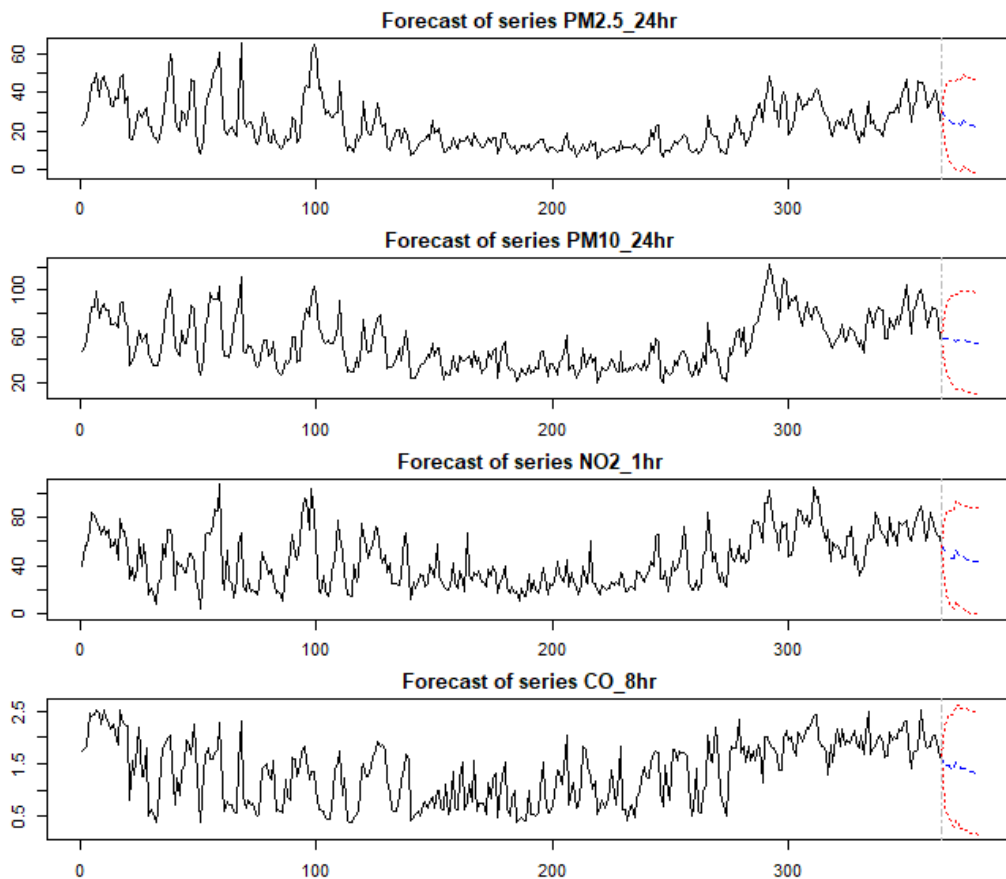
ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ CO อย่างมีนัยสำคัญทางสถิติ มี 7 ตัวแปร คือ ความเข้มข้นของ NO2 ล่าช้า 1 วันและ 2 วัน ความเข้มข้นของ CO ล่าช้า 1 วัน และ 3 วัน ความเข้มข้นของ PM10 ล่าช้า 3 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนพฤศจิกายน ($sd_{11} = 1$ สำหรับเดือนพฤศจิกายน และ $sd_{11} = 0$) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$CO_t = 0.254 + 0.0085 NO2_{t-1} + 0.491 CO_{t-1} + 0.0048 PM10_{t-1} - 0.0059 NO2_{t-2} - 0.0059 PM10_{t-3} + 0.262 CO_{t-3} + 0.148sd_{11}$$

จากสมการทั้ง 4 สมการข้างต้น สามารถคำนวณค่าพยากรณ์ของความเข้มข้นแต่ละชนิด ดังแสดงตารางที่ 4.11 และรูปที่ 4.24

ตารางที่ 4.11 ค่าพยากรณ์ของ PM2.5 PM10 NO2 CO และ O3 ของสถานีบริเวณสำนักงานเขตพระโขนง ระหว่างวันที่ 1-15 มกราคม 2566

ลำดับ	date	PM2.5	PM10	NO2	CO
1	1/1/2023	28.24841	58.30242	52.7577	1.46984
2	1/2/2023	26.97504	58.63469	49.48091	1.468131
3	1/3/2023	25.68635	58.05288	47.28101	1.468055
4	1/4/2023	24.57641	57.07453	46.08674	1.417698
5	1/5/2023	23.80574	56.28416	45.46747	1.389846
6	1/6/2023	23.26935	55.67692	52.17154	1.52399
7	1/7/2023	23.95833	56.34684	49.23287	1.491127
8	1/8/2023	23.4031	56.23189	48.17839	1.407072
9	1/9/2023	25.92715	56.25017	47.17391	1.415238
10	1/10/2023	24.9492	55.97317	46.33398	1.403465
11	1/11/2023	24.16794	55.61139	45.13525	1.374875
12	1/12/2023	23.48279	55.17108	44.50388	1.356016
13	1/13/2023	22.98979	54.77491	44.09938	1.344847
14	1/14/2023	22.62059	54.43445	43.86328	1.331926
15	1/15/2023	22.36055	54.18469	43.72763	1.321542



รูปที่ 4.24 กราฟแสดงการพยากรณ์ล่วงหน้า 15 วันของมลพิษอากาศ 4 ชนิดจากสถานีพระโขนง

4.3.3.3 ผลการวิเคราะห์ตัวแบบเวกเตอร์การถดถอยในตัวสำหรับข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตราษฎร์บูรณะ

ผลการทดสอบยูนิรูทสำหรับความเข้มข้นมลพิษทางอากาศ 4 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซไนโตรเจนไดออกไซด์ (NO₂) และก๊าซคาร์บอนมอนอกไซด์ (CO) และก๊าซไนโตรเจนไดออกไซด์ (NO₂) แสดงดังตารางที่ 4.12

ตารางที่ 4.12 การทดสอบยูนิรูทของมลพิษทางอากาศจากข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตราษฎร์บูรณะ

ตัวแปร	Random walk	Random walk with a drift		Random walk with a drift and a time trend		
	(I) $H_0: \gamma = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(III) $H_0: \gamma = 0$ & $a_0 = 0$ & $a_2 = 0$
	t-test	t-test	F-test	t-test	F-test	F-test
PM2.5	-2.5775	-7.4493	27.7483	-7.5147	18.8265	28.2373
PM10	-1.8871	-9.8067	48.0856	-10.5394	37.0531	55.5796
NO ₂	-2.4293	-9.0667	41.1027	-9.1379	27.8339	41.7509
CO	-2.2554	-10.9796	60.2753	-11.5674	44.6061	66.9091
ค่าวิกฤต ที่ $\alpha=0.05$	-1.95	-2.87	4.61	-3.42	4.71	6.30

จากตารางที่ 4.12 การทดสอบกระบวนการ Random walk ของอนุกรมเวลา สำหรับสมมติฐาน $H_0: \gamma = 0$ พบว่า ค่าสถิติทดสอบทีของความเข้มข้นของ PM2.5, PM10, NO₂, และ CO มีค่าเท่ากับ -2.5775, -1.8871, -2.4293 และ -2.2554 ซึ่งมีค่าน้อยกว่าค่าวิกฤต (ตกอยู่ในบริเวณวิกฤต) ยกเว้น PM10 จึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 3 ชุดไม่มียูนิรูท ยกเว้น PM10 และเมื่อทดสอบการทดสอบกระบวนการ Random walk with drift พบว่า สำหรับสมมติฐาน $H_0: \gamma = 0$ ค่าสถิติทดสอบทีของความเข้มข้นของ PM2.5, PM10, NO₂, CO และ O₃ สรุปผลได้ว่า อนุกรมเวลาทั้ง 4 ชุดไม่มียูนิรูท และการทดสอบสมมติฐาน (II) $H_0: \gamma = 0$ & $a_0 = 0$ พบว่า ค่าสถิติทดสอบ F ของความเข้มข้นของ PM2.5, PM10, NO₂ และ CO มีค่า 27.7483, 48.0856, 41.1027 และ 60.2753 ซึ่งตกอยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 4 ชุดไม่มียูนิรูทและมี drift

จากผลการทดสอบยูนิรูท สรุปได้ว่า อนุกรมเวลาของความเข้มข้นของมลพิษอากาศทั้ง 4 ชุดไม่มียูนิรูทและมี drift เราจึงสามารถใช้ตัวแบบ VAR ในการวิเคราะห์เพื่อหาสมการอธิบายความสัมพันธ์และพยากรณ์ความเข้มข้นของมลพิษอากาศ ทั้ง 4 ได้ ซึ่งผลการวิเคราะห์ แสดงดังตารางที่ 4.13

ตารางที่ 4.13 ผลการวิเคราะห์ตัวแบบ VAR ของข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตราชบุรีบูรณะ

ตัวแปรตาม : PM2.5 ($R^2 = 0.9483$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	1.00103	0.06407	15.623	<2e-16	***
PM10_24hr.l1	-0.20558	0.03432	-5.989	5.13e-09	***
NO2_1hr.l1	0.12059	0.02735	4.41	1.37e-05	***
NO2_1hr.l2	-0.12954	0.02458	-5.271	2.36e-07	***
const	13.8185	1.53886	8.98	<2e-16	***
ตัวแปรตาม : PM10 ($R^2 = 0.9656$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	0.42896	0.12886	3.329	0.000962	***
PM10_24hr.l1	0.33601	0.06903	4.868	1.70e-06	***
NO2_1hr.l1	0.18046	0.055	3.281	0.001135	**
NO2_1hr.l2	-0.23915	0.04943	-4.838	1.95e-06	***
const	34.93438	3.09487	11.288	<2e-16	***
ตัวแปรตาม : NO2 ($R^2 = 0.8959$)					
Variables	Estimate	Std. Error	t	p-value	significant
PM2.5_24hr.l1	0.69684	0.13516	5.156	4.19e-07	***
PM10_24hr.l1	-0.31923	0.07729	-4.131	4.51e-05	***
NO2_1hr.l1	0.49066	0.05869	8.36	1.40e-15	***
PM10_24hr.l2	-0.21866	0.05614	-3.895	0.000117	***
const	36.92616	3.75187	9.842	<2e-16	***
ตัวแปรตาม : CO ($R^2 = 0.9462$)					
Variables	Estimate	Std. Error	t	p-value	significant
NO2_1hr.l1	0.006663	0.000995	6.699	8.15e-11	***
CO_8hr.l1	0.536096	0.049949	10.733	<2e-16	***
PM10_24hr.l2	-0.00425	0.000974	-4.362	1.69e-05	***
NO2_1hr.l2	-0.00253	0.001085	-2.334	0.0201	*
const	0.575131	0.061245	9.391	<2e-16	***

*significant level at 0.05, **significant level at 0.01, ***significant level at 0.001

จากตารางที่ 4.13 พบว่า ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM2.5 อย่างมีนัยสำคัญทางสถิติ มี 4 ตัวแปร คือ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน ความเข้มข้นของ PM10 ล่าช้า 1 วัน ความเข้มข้นของ NO2 ล่าช้า 1 วัน และ 2 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM2.5_t = 13.819 + 1.001 PM2.5_{t-1} - 0.201 PM10_{t-1} - 0.120 NO2_{t-1} - 0.129 NO2_{t-2}$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM10 อย่างมีนัยสำคัญทางสถิติ มี 4 ตัวแปร คือ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน ความเข้มข้นของ PM10 ล่าช้า 1 วัน ความเข้มข้นของ NO2 ล่าช้า 1 วัน และ 2 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM10_t = 34.934 + 0.429 PM2.5_{t-1} + 0.336 PM10_{t-1} + 0.180 NO2_{t-1} - 0.239 NO2_{t-2}$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ NO2 อย่างมีนัยสำคัญทางสถิติ มี 4 ตัวแปร คือ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน ความเข้มข้นของ PM10 ล่าช้า 1 วัน และ 2 วัน ความเข้มข้นของ NO2 ล่าช้า 1 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$NO2_t = 36.926 + 0.697 PM2.5_{t-1} - 0.319 PM10_{t-1} + 0.491 NO2_{t-1} - 0.219 PM10_{t-2}$$

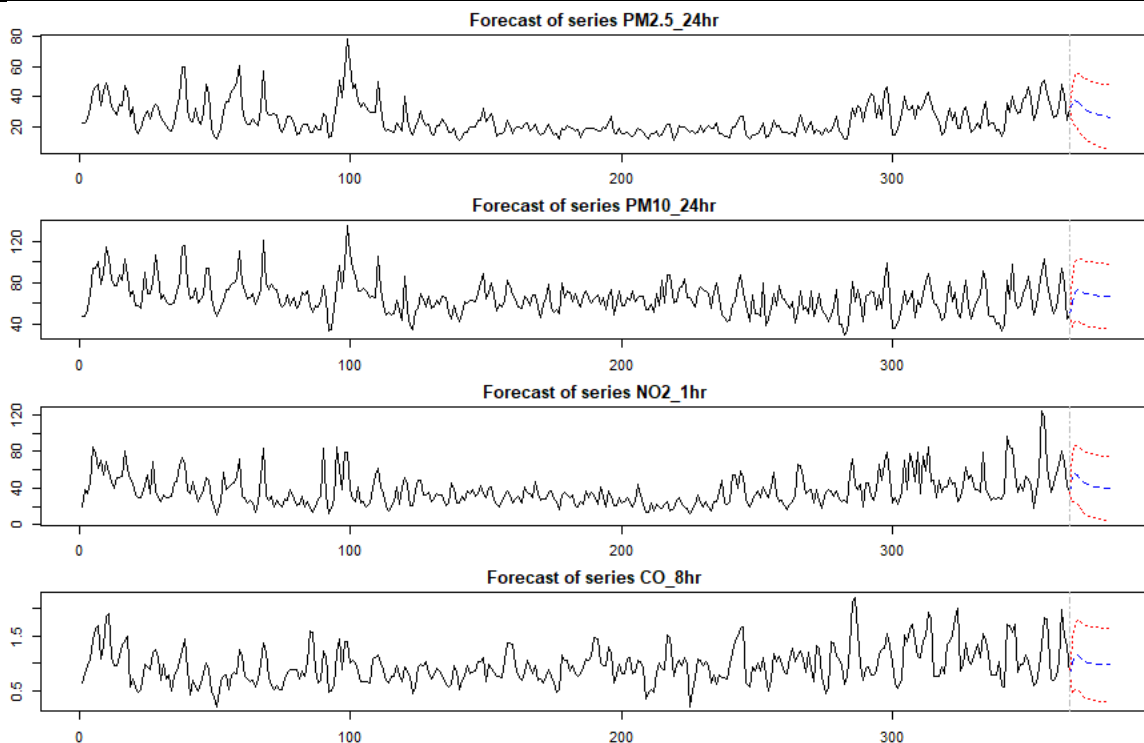
ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ CO อย่างมีนัยสำคัญทางสถิติ มี 4 ตัวแปร คือ ความเข้มข้นของ NO2 ล่าช้า 1 วันและ 2 วัน ความเข้มข้นของ CO ล่าช้า 1 วัน ความเข้มข้นของ PM10 ล่าช้า 2 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$CO_t = 0.575 + 0.0067 NO2_{t-1} + 0.536 CO_{t-1} - 0.00425 PM10_{t-2} - 0.00253 NO2_{t-2}$$

จากสมการทั้ง 4 สมการข้างต้น สามารถคำนวณค่าพยากรณ์ของความเข้มข้นแต่ละชนิด ดังแสดงตารางที่ 4.14 และรูปที่ 4.25

ตารางที่ 4.14 ค่าพยากรณ์ของ PM2.5 PM10 NO2 CO และ O3 ของสถานีบริเวณสำนักงานเขตราชบุรี
 บรูณะ ระหว่างวันที่ 1-15 มกราคม 2566

ลำดับ	date	PM2.5	PM10	NO2	CO
1	1/1/2023	34.73982	61.23845	51.2003	0.937083
2	1/2/2023	37.60833	71.21449	56.05025	1.122274
3	1/3/2023	36.95218	72.86609	54.51092	1.160471
4	1/4/2023	35.14195	71.70192	50.58983	1.116039
5	1/5/2023	33.29575	70.19475	47.41496	1.062977
6	1/6/2023	31.88255	69.26114	45.30635	1.028251
7	1/7/2023	30.81681	68.71997	43.91455	1.010027
8	1/8/2023	29.96654	68.33407	42.86589	1.000289
9	1/9/2023	29.24856	67.98327	42.00036	0.993905
10	1/10/2023	28.63342	67.652	41.27173	0.98901
11	1/11/2023	28.11	67.35231	40.66802	0.985213
12	1/12/2023	27.66924	67.09239	40.17516	0.982407
13	1/13/2023	27.30022	66.87142	39.77469	0.980421
14	1/14/2023	26.99181	66.68447	39.44843	0.97904
15	1/15/2023	26.73404	66.52625	39.18142	0.978078



รูปที่ 4.25 กราฟแสดงการพยากรณ์ล่วงหน้า 15 วันของมลพิษอากาศ 4 ชนิดจากสถานีราชบุรีบรูณะ

4.3.3.4 ผลการวิเคราะห์ตัวแบบเวกเตอร์การถดถอยในตัวสำหรับข้อมูลคุณภาพอากาศเขต ราชเทวี

ผลการทดสอบยูนิรูทสำหรับความเข้มข้นมลพิษทางอากาศ 5 ชนิด ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซโอโซน (O3) และก๊าซไนโตรเจนไดออกไซด์ (NO2) ของข้อมูลคุณภาพอากาศจากสถานีทั้ง 4 สถานี ด้วยการทดสอบ ADF test แสดงดังตารางที่ 4.15

ตารางที่ 4.15 การทดสอบยูนิรูทของมลพิษทางอากาศจากข้อมูลคุณภาพอากาศสถานีบริเวณสำนักงานเขตราชเทวี

ตัวแปร	Random walk	Random walk with a drift		Random walk with a drift and a time trend		
	(I) $H_0: \gamma = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(I) $H_0: \gamma = 0$	(II) $H_0: \gamma = 0$ & $a_0 = 0$	(III) $H_0: \gamma = 0$ & $a_0 = 0$ & $a_2 = 0$
	t-test	t-test	F-test	t-test	F-test	F-test
PM2.5	-2.2035	-5.919	17.5178	-5.9777	11.9113	17.8668
PM10	-1.8502	-5.8313	17.0028	-5.8802	11.5269	17.2896
NO ₂	-1.8747	-7.4233	27.5526	-7.4426	18.4922	27.7382
CO	-1.8489	-4.4547	9.9243	-5.7773	11.1276	16.6892
O ₃	-3.058	-6.9097	23.8717	-6.9358	16.0547	24.082
ค่าวิกฤต ที่ $\alpha=0.05$	-1.95	-2.87	4.61	-3.42	4.71	6.30

จากตารางที่ 4.15 การทดสอบกระบวนการ Random walk ของอนุกรมเวลา สำหรับสมมติฐาน $H_0: \gamma = 0$ พบว่า ค่าสถิติทดสอบที่ของความเข้มข้นของ PM2.5, PM10, NO2, CO และ O3 มีค่าเท่ากับ -2.2035, -1.8502, -1.8747, -1.8489 และ -3.058 ซึ่งอยู่นอกบริเวณวิกฤต (ยกเว้น PM2.5 และ O3) แต่เมื่อทดสอบการทดสอบกระบวนการ Random walk with drift พบว่า สำหรับสมมติฐาน $H_0: \gamma = 0$ ค่าสถิติทดสอบที่ของความเข้มข้นของ PM2.5, PM10, NO2, CO และ O3 มีค่าเท่ากับ -5.919, -5.8313, -7.4233, -4.4547 และ -6.9097 ซึ่งตกอยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธสมมติฐานว่าง แสดงว่า อนุกรมเวลาทั้ง 5 ชุดไม่มียูนิรูท และการทดสอบสมมติฐาน (II) $H_0: \gamma = 0$ & $a_0 = 0$ พบว่า ค่าสถิติทดสอบ F ของความเข้มข้นของ PM2.5, PM10, NO2, CO และ O3 มีค่า 17.5178, 17.0028, 27.5526, 9.9243 และ 23.8717 ซึ่งตกอยู่ในบริเวณวิกฤต ดังนั้นจึงปฏิเสธสมมติฐานว่างแสดงว่า อนุกรมเวลาทั้ง 5 ชุดไม่มียูนิรูทและมี drift

จากผลการทดสอบยูนิรูท สรุปได้ว่า อนุกรมเวลาของความเข้มข้นของมลพิษอากาศ ทั้ง 5 ชุดไม่มียูนิรูทและมี drift เราจึงสามารถใช้ตัวแบบ VAR ในการวิเคราะห์เพื่อพยากรณ์ความเข้มข้นของมลพิษอากาศได้ ซึ่งผลการวิเคราะห์ แสดงดังตารางที่ 4.16

ตารางที่ 4.16 ผลการวิเคราะห์ตัวแบบ VAR ของข้อมูลคุณภาพอากาศจากสถานีบริเวณสำนักงานเขตราชเทวี

ตัวแปรตาม : PM2.5 ($R^2 = 0.9534$)					
Variables	Estimate	Std. Error	t	p-value	signifcant
PM2.5_24hr.l1	0.77375	0.05534	13.983	<2e-16	***
O3_8hr.l1	0.08189	0.0198	4.137	4.4e-05	***
PM2.5_24hr.l2	-0.1766	0.06815	-2.591	0.00996	**
NO2_1hr.l2	-0.05383	0.02265	-2.377	0.01798	*
PM2.5_24hr.l3	0.13039	0.053	2.46	0.01437	*
const	5.53331	1.3702	4.038	6.6e-05	***
sd2	2.84622	1.00607	2.829	0.00493	**
ตัวแปรตาม : PM10 ($R^2 = 0.9675$)					
Variables	Estimate	Std. Error	t	p-value	signifcant
PM10_24hr.l1	0.72425	0.03997	18.118	<2e-16	***
O3_8hr.l1	0.15493	0.03608	4.294	2.27e-05	***
NO2_1hr.l2	-0.11957	0.03161	-3.783	0.000182	***
O3_8hr.l2	-0.11182	0.04241	-2.637	0.008734	**
O3_8hr.l3	0.07773	0.03386	2.295	0.022291	*
const	11.88938	1.89989	6.258	1.12e-09	***
ตัวแปรตาม : NO2 ($R^2 = 0.9488$)					
Variables	Estimate	Std. Error	t	p-value	signifcant
NO2_1hr.l1	0.47436	0.05207	9.11	<2e-16	***
PM10_24hr.l2	-0.19552	0.03713	-5.267	2.4e-07	***
NO2_1hr.l2	0.13539	0.05065	2.673	0.00786	**
const	24.10911	2.69063	8.96	<2e-16	***
ตัวแปรตาม : CO ($R^2 = 0.9099$)					
Variables	Estimate	Std. Error	t	p-value	signifcant
CO_8hr.l1	0.524508	0.05089	10.307	<2e-16	***
PM2.5_24hr.l2	-0.00357	0.001218	-2.93	0.00361	**
CO_8hr.l2	0.166951	0.058168	2.87	0.00435	**
NO2_1hr.l3	-0.00512	0.000878	-5.835	1.21e-08	***
CO_8hr.l3	0.202465	0.050004	4.049	6.32e-05	***
const	0.384217	0.053952	7.121	5.97e-12	***
sd4	0.118137	0.039789	2.969	0.00319	**
ตัวแปรตาม : O3 ($R^2 =$)					
Variables	Estimate	Std. Error	t	p-value	signifcant
PM10_24hr.l1	0.35042	0.08473	4.136	4.42e-05	***
O3_8hr.l1	0.65927	0.04606	14.315	<2e-16	***
PM10_24hr.l2	-0.40074	0.1055	-3.798	0.000171	***
NO2_1hr.l2	-0.17061	0.05025	-3.395	0.000763	***
PM10_24hr.l3	0.17566	0.08114	2.165	0.031049	*
const	13.57266	3.15658	4.3	2.21e-05	***
sd2	7.07091	2.20926	3.201	0.001495	**

จากตารางที่ 4.16 พบว่า ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM2.5 อย่างมีนัยสำคัญทางสถิติ มี 6 ตัวแปร คือ ความเข้มข้น PM2.5 ล่าช้า 1 วัน, 2 วัน และ 3 วัน ความเข้มข้นของ O3 ล่าช้า 1

วัน และ ความเข้มข้นของ NO₂ ล่าช้า 2 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนกุมภาพันธ์ ($sd_2 = 1$ สำหรับเดือนกุมภาพันธ์ และ $sd_2 = 0$ เดือนอื่นๆ) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM2.5_t = 5.533 + 0.77375 PM2.5_{t-1} + 0.08189 O3_{t-1} - 0.1766 PM2.5_{t-2} \\ - 0.054 NO2_{t-2} + 0.130 PM2.5_{t-3} + 2.846sd_2$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ PM₁₀ อย่างมีนัยสำคัญทางสถิติ มี 6 ตัวแปร คือ ความเข้มข้นของ PM₁₀ ล่าช้า 1 วัน ความเข้มข้นของ O₃ ล่าช้า 1 วัน, 2 วัน และ 3 วัน ความเข้มข้นของ NO₂ ล่าช้า 2 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$PM10_t = 11.889 + 0.724PM10_{t-1} + 0.155 O3_{t-1} - 0.120 NO2_{t-2} - 0.112 O3_{t-2} \\ + 0.078 O3_{t-3}$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ NO₂ อย่างมีนัยสำคัญทางสถิติ มี 3 ตัวแปร คือ ความเข้มข้นของ NO₂ ล่าช้า 1 วัน และ 2 วัน และความเข้มข้นของ PM₁₀ ล่าช้า 2 วัน ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$NO2_t = 24.109 + 0.474 NO2_{t-1} - 0.196 PM10_{t-2} + 0.136 NO2_{t-2}$$

ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ CO อย่างมีนัยสำคัญทางสถิติ มี 6 ตัวแปร คือ ความเข้มข้นของ CO ล่าช้า 1 วัน, 2 และ 3 วัน ความเข้มข้นของ PM_{2.5} ล่าช้า 2 วัน และความเข้มข้นของ NO₂ ล่าช้า 2 วัน และ ตัวแปรหุ่นฤดูกาลสำหรับเดือนเมษายน ($sd_4 = 1$ สำหรับเดือนเมษายน และ $sd_4 = 0$ เดือนอื่นๆ) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$CO_t = 0.384 + 0.525 CO_{t-1} - 0.0036 PM2.5_{t-2} + 0.167 CO_{t-2} - 0.0051 NO2_{t-3} \\ + 0.202 CO_{t-3} + 0.118sd_4$$

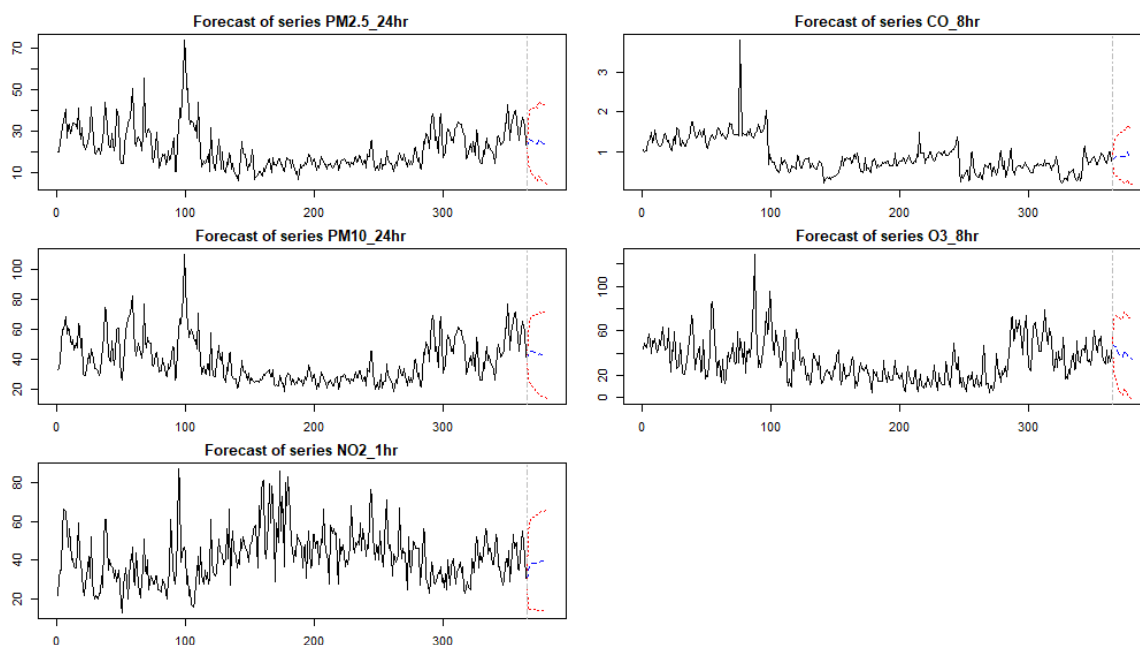
ปัจจัยที่มีผลต่อการพยากรณ์ค่าความเข้มข้นของ O₃ อย่างมีนัยสำคัญทางสถิติ มี 6 ตัวแปร คือ ความเข้มข้นของ PM₁₀ ล่าช้า 1 วัน, 2 วันและ 3 วัน ความเข้มข้นของ O₃ ล่าช้า 1 วัน และ ความเข้มข้นของ NO₂ ล่าช้า 2 วัน และตัวแปรหุ่นฤดูกาลสำหรับเดือนกุมภาพันธ์ ($sd_2 = 1$ สำหรับเดือนกุมภาพันธ์ และ $sd_2 = 0$ เดือนอื่นๆ) ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$O3_t = 13.573 + 0.350 PM10_{t-1} + 0.659 O3_{t-1} - 0.401 PM10_{t-2} - 0.171 NO2_{t-2} \\ + 0.176 PM10_{t-3} + 7.07sd_2$$

จากสมการทั้ง 4 สมการข้างต้น สามารถคำนวณค่าพยากรณ์ของความเข้มข้นแต่ละชนิด ดังแสดงตารางที่ 4.17 และรูปที่ 4.26

ตารางที่ 4.17 ค่าพยากรณ์ของ PM2.5 PM10 NO2 CO และ O3 ของสถานีบริเวณสำนักงานเขตราชเทวี ระหว่างวันที่ 1-15 มกราคม 2566

ลำดับ	date	PM2.5	PM10	NO2	CO	O3
1	1/1/2023	25.02838	43.04458	34.17029	0.822682	46.99132
2	1/2/2023	26.0552	44.25968	36.54264	0.856029	45.38264
3	1/3/2023	25.83668	45.415	37.65361	0.881072	42.37397
4	1/4/2023	25.45273	45.55485	38.26423	0.877947	40.42382
5	1/5/2023	25.10861	45.43254	38.47841	0.875866	38.74808
6	1/6/2023	24.71156	44.99552	38.63534	0.875002	37.6432
7	1/7/2023	24.31303	44.518	38.76269	0.871668	36.79868
8	1/8/2023	23.95231	44.01585	38.9298	0.869674	36.20147
9	1/9/2023	26.48226	43.55295	39.11967	0.868515	42.79555
10	1/10/2023	26.13636	44.22044	39.33054	0.867533	39.99853
11	1/11/2023	25.13563	43.46407	39.54679	0.974674	38.45335
12	1/12/2023	24.61438	43.47701	39.54741	0.912597	36.78484
13	1/13/2023	24.19441	43.15738	39.72486	0.900216	36.07285
14	1/14/2023	23.77268	42.88197	39.80659	0.905803	35.3533
15	1/15/2023	23.38408	42.49972	39.93188	0.895593	34.88249



รูปที่ 4.26 กราฟแสดงการพยากรณ์ล่วงหน้า 15 วันของมลพิษอากาศ 4 ชนิดจากสถานีราชเทวี

บทที่ 5 สรุปผลการศึกษา

การศึกษานี้ เป็นการศึกษานำข้อมูลด้านฝุ่นละอองไปใช้ประโยชน์ โดยแบ่งออกเป็น 3 กรณี คือ 1) เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน n ชั่วโมงถัดไป โดยใช้ค่าความเข้มข้นของฝุ่นละออง PM2.5 รายชั่วโมงจำนวน 48 ชั่วโมงก่อนหน้า ด้วยเทคนิค LSTM 2) เพื่อพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน n วันถัดไป โดยใช้ประวัติข้อมูลเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า ด้วยเทคนิค LSTM และ 3) เพื่อพยากรณ์ PM2.5 รายวัน โดยใช้ข้อมูลอนุกรมเวลาหลายชุด (multivariate time series) ได้แก่ ค่าความเข้มข้นของมลพิษทางอากาศ PM2.5, PM10, CO, NO₂, O₃ และ CO₂ รายวันในอดีต ด้วยตัวแบบเวกเตอร์การถดถอยในตัว (Vector autoregressive (VAR) model) สามารถสรุปผลได้ดังนี้

5.1 การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน n ชั่วโมงถัดไป โดยใช้ข้อมูลในอดีต 48 ชั่วโมงก่อนหน้า ด้วยเทคนิค LSTM

การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 1 หรือ 2 วัน ถัดไป ด้วยข้อมูลรายชั่วโมงในอดีต 48 ชั่วโมง มีความแม่นยำมากกว่า การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน 3, 4, 5 หรือ 4 วัน ถัดไป ดังนั้นเทคนิค LSTM เหมาะสมสำหรับการใช้พยากรณ์ค่าความเข้มข้นของ PM2.5 เป็นรายชั่วโมงได้ล่วงหน้าประมาณ 1 หรือ 2 ชั่วโมงถัดไป โดยต้องใช้ข้อมูลหลายชั่วโมงก่อนหน้าในการฝึกฝนตัวแบบ และถ้าพยากรณ์รายชั่วโมงล่วงหน้าห่างจากชุดข้อมูลในอดีตมากขึ้น ความแม่นยำของค่าพยากรณ์จะลดลง จึงไม่เหมาะที่จะใช้พยากรณ์ล่วงหน้ามากกว่า 4 ชั่วโมง อย่างไรก็ตาม การทำนายปริมาณฝุ่นเพียง 1 หรือ 2 ชั่วโมงล่วงหน้าในทางปฏิบัติอาจจะไม่สามารถนำไปใช้ในการประโยชน์ได้มากเท่าไรนัก หากต้องการพยากรณ์ค่าความเข้มข้นของ PM2.5 ใน 24 ชั่วโมงล่วงหน้า หรือ รายวัน จะต้องใช้ข้อมูลในอดีตจำนวนมากขึ้น

5.2 การพยากรณ์ค่าความเข้มข้นของ PM2.5 รายชั่วโมง ใน n วันถัดไป โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า ด้วยเทคนิค LSTM

การพยากรณ์ของความเข้มข้นของ PM2.5 ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 1 วัน โดยใช้ข้อมูลในอดีตเป็นรอบ 24 ชั่วโมง จำนวน 25 วันก่อนหน้า มีความแม่นยำมากกว่า การพยากรณ์ของความเข้มข้นของ PM2.5 ในรอบ 24 ชั่วโมงวันถัดไป จำนวน 2, 3, 4 หรือ 5 วัน ถัดไป และมีช่วงเวลาที่น่าเชื่อถือยาวนานกว่าการพยากรณ์ล่วงหน้าเป็นรายชั่วโมงจำนวนไม่กี่ชั่วโมง แต่เมื่อเพิ่มจำนวนวันที่ต้องการการพยากรณ์ล่วงหน้า ผลลัพธ์จากการพยากรณ์มีความแม่นยำลดลงเช่นเดียวกัน ดังนั้นการเรียนรู้จากรูปแบบของการเพิ่มขึ้นและลดลงของค่าความเข้มข้นของ PM 2.5 ในอดีตเพื่อค่าพยากรณ์ความเข้มข้นของ PM 2.5 ในช่วงรอบ 24 ชั่วโมง ถัดไป จำเป็นต้องใช้ข้อมูลล่วงหน้าหลายวันในการฝึกฝนตัวแบบ หากเป็นการทำนายในวันเดียวกัน ในทาง

ทฤษฎี ความแม่นยำมีแนวโน้มคงตัว แต่หากจำนวนวันที่พยากรณ์ล่วงหน้ามากขึ้น โดยมีระยะเวลาห่างจากชุดข้อมูลมากขึ้น ความแม่นยำของค่าพยากรณ์มีแนวโน้มลดลง ดังนั้นวิธีการนี้จึงเหมาะกับการทำนายแบบรายวันล่วงหน้า 1 วัน และสามารถใช้เป็นแนวทางในการสร้างระบบ Recommendation System ที่อาจพัฒนาในอนาคต

5.3 การพยากรณ์ PM2.5 รายวัน โดยใช้ข้อมูลอนุกรมเวลาหลายชุด ด้วยตัวแบบ VAR

การศึกษาปัจจัยที่ส่งผลต่อมลพิษทางอากาศ ได้แก่ ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) ฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน (PM10) ก๊าซโอโซน (O3) ก๊าซคาร์บอนมอนอกไซด์ (CO) ก๊าซไนโตรเจนไดออกไซด์ (NO2) ของสถานีที่ตรวจวัด 4 สถานี คือ สถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง สถานีบริเวณสำนักงานเขตราชบุรีบูรณะ และสถานีบริเวณสำนักงานเขตราชเทวี สามารถสรุปผล ดังตารางที่ 5.1

ตารางที่ 5.1 ปัจจัยที่ส่งผลต่อมลพิษทางอากาศของแต่ละสถานี

ตัวแปรตาม	สถานี	ตัวแปรทำนาย (Predictors)					
		PM2.5	PM10	NO2	CO	O3	Seasonal
PM2.5	ดินแดง	Lag1 Lag2 Lag3			Lag1		sd2
	พระโขนง	Lag1		Lag1 Lag2	Lag1		sd2
	ราชบุรีบูรณะ	Lag1	Lag1	Lag1 Lag2			
	ราชเทวี	Lag1 Lag2 Lag3		Lag1		Lag1	sd2
PM10	ดินแดง	Lag1 Lag2 Lag4	Lag1		Lag1		Sd3
	พระโขนง		Lag1	Lag1	Lag1		
	ราชบุรีบูรณะ	Lag1	Lag1	Lag1 Lag2			
	ราชเทวี		Lag1	Lag2		Lag1 Lag2	

ตัวแปรตาม	สถานี	ตัวแปรทำนาย (Predictors)					
		PM2.5	PM10	NO2	CO	O3	Seasonal
						Lag3	
NO2	ดินแดง	Lag1	Lag1 Lag4	Lag1 Lag3 Lag4	Lag2		
	พระโขนง	Lag2	Lag1	Lag1	Lag2		sd11
	ราษฎร์บูรณะ	Lag1	Lag1 Lag2	Lag1			
	ราชเทวี		Lag2	Lag1 Lag2			
CO	ดินแดง	Lag1 Lag4		Lag2	Lag1 Lag4		
	พระโขนง		Lag2 Lag3	Lag1 Lag2	Lag1 Lag3		sd11
	ราษฎร์บูรณะ		Lag1	Lag1 Lag2	Lag1		
	ราชเทวี	Lag2		Lag3	Lag1 Lag2 Lag3		sd4
O3	ราชเทวี		Lag1 Lag2 Lag3	Lag2		Lag1	sd2

จากตารางที่ 5.1 จะเห็นได้ว่า การพยากรณ์ความเข้มข้นของ PM2.5 ทั้ง 4 สถานี บางสถานีขึ้นอยู่กับค่าความเข้มข้นของ PM2.5 ล่าช้า 1 วัน หรือ บางสถานีขึ้นอยู่กับความเข้มข้นของ PM2.5 ล่าช้า 1 วัน 2 วัน และ 3 วัน และค่าความเข้มข้นของ NO2 1 วัน สำหรับ 3 สถานีจาก 4 สถานี โดยที่ ความเข้มข้นของ PM2.5 ล่าช้า 1 วัน มีความสัมพันธ์หรือส่งผลต่อค่าพยากรณ์ความเข้มข้นของ PM2.5 มากที่สุด นอกจากนี้ ยังพบว่า ความเข้มข้นของ PM2.5 จากสถานีดินแดง สถานีบริเวณสำนักงานเขตพระโขนง และสถานีบริเวณสำนักงานเขตพระโขนง ได้รับอิทธิพลของฤดูกาล คือ เดือนกุมภาพันธ์ จะมีค่าความเข้มข้นของ PM2.5 สูงกว่าเดือนอื่นๆ ดังนั้น จึงควรมีการเฝ้าระวังค่าความเข้มข้นของ PM2.5 โดยเฉพาะในเดือนกุมภาพันธ์

การประเมินประสิทธิภาพของตัวแบบการพยากรณ์ความเข้มข้นของมลพิษอากาศของ 4 สถานี โดยใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error: MAPE) แสดงในตารางที่ 5.2

ตารางที่ 5.2 ค่า MAPE ของสมการพยากรณ์ความเข้มข้นของมลพิษอากาศทั้ง 5 ชนิดของแต่ละ 4 สถานี

สถานี	ตัวแปรตาม				
	PM2.5	PM10	NO2	CO	O3
ดินแดง	0.178	0.141	0.224	0.192	-
พระโขนง	0.197	0.175	0.277	0.277	
ราษฎร์บูรณะ	0.305	0.315	0.438	0.600	
ราชเทวี	0.140	0.145	0.159	0.187	0.379

จากตารางที่ 5.2 พบว่า ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของสมการพยากรณ์ความเข้มข้นของ PM2.5 NO2 CO O3 ของสถานีบริเวณสำนักงานเขตราชเทวี มีค่าต่ำสุด เมื่อเทียบกับค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของสมการพยากรณ์ความเข้มข้นของมลพิษอากาศดังกล่าว ในสถานีอื่นๆ ส่วนค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของสมการพยากรณ์ความเข้มข้นของ PM10 ของสถานีบริเวณสำนักงานเขตราชเทวีมีค่าต่ำสุด รองลงมาจาก ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของสมการพยากรณ์ความเข้มข้นของ PM10 ของสถานีบริเวณสำนักงานเขตดินแดง และมีค่าต่างกันเล็กน้อย ซึ่งทั้งนี้อาจจะเป็นเพราะว่า สมการพยากรณ์ความเข้มข้นของมลพิษทางอากาศของสถานีบริเวณสำนักงานเขตราชเทวีได้จากตัวแบบ VAR ของสารมลพิษ 5 ชนิด ในขณะที่สมการพยากรณ์ความเข้มข้นของมลพิษทางอากาศของสถานีอื่นๆ ได้จากแบบ VAR ของสารมลพิษ 4 ชนิด จึงเป็นที่น่าสังเกตว่า การใช้ข้อมูลสารสนเทศของสารมลพิษทางที่มากขึ้น 1 ตัวแปร อาจส่งผลต่อความแม่นยำในการพยากรณ์ค่าความเข้มข้นของมลพิษทางอากาศแต่ละชนิดด้วยตัวแบบ VAR

บรรณานุกรม

- ประกาศกรมควบคุมมลพิษ เรื่อง ดัชนีคุณภาพอากาศของประเทศไทย. (2566, 16 พฤษภาคม). ราชกิจจานุเบกษา เล่ม 120.
- Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". **Journal of the American Statistical Association**, **74**, 427-431
- Maddala G. S. and In-Moo Kim. (1988). **Unit Roots, Cointegration, and Structural Change**.
- N.M. Noor et.al (2015). "Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set". **Materials Science Forum**, August 2014
- Said, S.E. and Dickey, D.A. (1984) Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. **Biometrika**, **71**, 599-607.
- Ugwuowo, I. F., & Asogwa, (2018). "A comparative analysis of vector autoregression model and neural networks". **International Journal of Mathematics and Statistics** **4(8)**, 1-13.