

การทำนายค่าความต้องการออกซิเจนทางชีวเคมีสำหรับน้ำในแม่น้ำเจ้าพระยา
โดยกระบวนการอัลฟาทรिमอาร์มา



บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์และวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2558

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

BIOCHEMICAL OXYGEN DEMAND PREDICTION FOR CHAOPHRAYA RIVER WATER
BY ALPHA -TRIMMED ARIMA PROCESS

Miss Chadaphim Photphanloet



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Applied Mathematics and
Computational Science

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2015

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การทำนายค่าความต้องการออกซิเจนทางชีวเคมีสำหรับ
น้ำในแม่น้ำเจ้าพระยาโดยกระบวนการอัลฟาทรिमอาร์มา

โดย

นางสาวชฎาพิมพ์ พตพันธ์เลิศ

สาขาวิชา

คณิตศาสตร์ประยุกต์และวิทยาการคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... คณบดีคณะวิทยาศาสตร์

(รองศาสตราจารย์ ดร. พลกฤษณ์ แสงวณิช)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(อาจารย์ ดร. กิติพร พลายมาศ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์)

..... กรรมการ

(อาจารย์ ดร. จิราพรรณ สุนทรโชติ)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ภัทรสินี ภัทรโกศล)

..... กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร. สุรัตนา สังข์หนู)

ชฎาพิมพ์ พตพันธ์เลิศ : การทำนายค่าความต้องการออกซิเจนทางชีวเคมีสำหรับน้ำในแม่น้ำเจ้าพระยาโดยกระบวนการอัลฟาทริมอาร์มา (BIOCHEMICAL OXYGEN DEMAND PREDICTION FOR CHAOPHRAYA RIVER WATERBY ALPHA -TRIMMED ARIMA PROCESS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. รัชลิดา ลิปิกรณ์, 65 หน้า.

น้ำเป็นหนึ่งในปัจจัยที่สำคัญที่สุดสำหรับการดำรงชีวิตอย่างยั่งยืนของมนุษยชาติ นอกจากนี้แหล่งน้ำที่เพียงพอแล้ว คุณภาพของน้ำก็มีความสำคัญ กล่าวได้ว่า คุณภาพน้ำต้องมีความปลอดภัยสำหรับมนุษย์และเป็นไปตามมาตรฐานคุณภาพน้ำ คุณภาพน้ำสามารถตรวจวัดได้โดยใช้คุณสมบัติทางกายภาพ เคมีและชีวภาพ วัตถุประสงค์ของการศึกษาครั้งนี้คือการประยุกต์การวิเคราะห์อนุกรมเวลาในการสร้างแบบจำลองและทำนายค่าความต้องการออกซิเจนทางชีวเคมี (BOD) ของน้ำที่ 11 สถานีตรวจวัดตามแนวแม่น้ำเจ้าพระยาในอนาคต ข้อมูล BOD ที่นำมาใช้ในการศึกษาถูกเก็บรวบรวมโดยกรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อมตั้งแต่ปี 1996 –2013 อย่างไรก็ตาม ข้อมูล BOD ที่ได้รับไม่สมบูรณ์และไม่เหมาะสมสำหรับตัวแบบ ARIMA ดังนั้นการศึกษานี้แนะนำเสนอกระบวนการใหม่ในการปรับข้อมูล BOD และเลือกรูปแบบ ARIMA ที่เหมาะสมที่สุดสำหรับชุดข้อมูลที่ไม่สมบูรณ์ดังกล่าว กระบวนการนี้เรียกว่า กระบวนการอัลฟาทริมอาร์มา กระบวนการนี้เริ่มต้นจากการตรวจสอบการกระจายของข้อมูล การแต่งขอบข้อมูลที่ผิดปกติและสุดท้ายเลือกตัวแบบ ARIMA ที่เหมาะสมที่สุด กระบวนการที่นำเสนอสามารถใช้ทั้งข้อมูลอนุกรมเวลาที่มีความแปรผันตามฤดูกาลและไม่มีความแปรผันตามฤดูกาล ผลการทำนายโดยรวมแสดงให้เห็นว่า การนำเสนอกระบวนการอัลฟาทริมอาร์มา สามารถเพิ่มความแม่นยำของการทำนายค่า BOD

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาควิชา	คณิตศาสตร์และวิทยาการ	ลายมือชื่อนิสิต
	คอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาหลัก
สาขาวิชา	คณิตศาสตร์ประยุกต์และวิทยาการ	
	คณนา	
ปีการศึกษา	2558	

5771950923 : MAJOR APPLIED MATHEMATICS AND COMPUTATIONAL SCIENCE

KEYWORDS: TIME SERIES ANALYSIS, BIOCHEMICAL OXYGEN DEMAND, ARIMA MODEL

CHADAPHIM PHOTPHANLOET: BIOCHEMICAL OXYGEN DEMAND PREDICTION
FOR CHAOPHRAYA RIVER WATERBY ALPHA -TRIMMED ARIMA PROCESS.

ADVISOR: ASSOC. PROF. RAJALIDA LIPIKORN, Ph.D., 65 pp.

Water is one of the most important factor for sustainable livelihood of mankind. In addition to having adequate water resources, the quality of water is also important. That is to say that water quality must be safe for mankind and meet the water quality standard. The water quality can be measured based on its physical, chemical, and biological properties. The objective of this study is to apply time series analysis to model and predict Biochemical Oxygen Demand (BOD) of water at 11 monitoring stations along the Chaophraya River in the future. The BOD data that were used in our study were collected by the Pollution Control Department, Ministry of Natural Resources and Environment during 1996 – 2013. However, the obtained BOD data are not perfect and are not suitable for ARIMA model. This study thus proposes a new process to adjust BOD data and to select the most suitable ARIMA models for such imperfect sets of data. This process is called alpha-trimmed ARIMA process. The process starts from checking distribution of data, trimming outlier data, and finally selecting the most suitable ARIMA model. The proposed process can be applied to both seasonal and nonseasonal time series data. The overall prediction results show that the proposed alpha-trimmed ARIMA process can improve the accuracy of BOD prediction.

Department: Mathematics and Student's Signature

Computer Science Advisor's Signature

Field of Study: Applied Mathematics and
Computational Science

Academic Year: 2015

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาให้ความรู้ คำแนะนำ และคำปรึกษาต่าง ๆ ที่ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี

ขอกราบขอบพระคุณ อาจารย์ ดร.กิติพร พลายมาศประธานกรรมการ ผู้ช่วยศาสตราจารย์ ดร.ภัทรสินี ภัทรโกศล กรรมการ อาจารย์ ดร.จิราพรรณ สุนทรโชติ กรรมการ และ ผู้ช่วยศาสตราจารย์ ดร.สุรัตนา สังข์หนูนุ กรรมการภายนอกมหาวิทยาลัย ที่ให้คำปรึกษา คำแนะนำและแก้ไขข้อบกพร่องต่าง ๆ ในงานวิจัยนี้ ซึ่งทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น

ขอกราบขอบพระคุณบิดาและมารดา ตลอดจนพี่น้องในครอบครัวและเพื่อน ๆ ทุกคนที่คอยเป็นกำลังใจและช่วยเหลือผู้วิจัยมาโดยตลอด

ขอขอบคุณโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่สนับสนุนทุนการศึกษามาโดยตลอด

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ	ญ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ประโยชน์ที่ได้รับ	2
1.4 ลำดับขั้นตอนในการเสนอผลการวิจัย	2
1.5 ขอบเขตของการวิจัย.....	3
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	4
2.1 แนวความคิดพื้นฐานเกี่ยวกับการพยากรณ์.....	4
2.2 อนุกรมเวลาและสหสัมพันธ์	5
2.3 การทดสอบแนวโน้มและฤดูกาล.....	7
2.4 เทคนิคการทำให้เรียบ	9
2.5 การวิเคราะห์อนุกรมเวลาที่เป็นสเตชันนารีแบบบอกซ์เจนคินส์	14
2.6 การวิเคราะห์อนุกรมเวลาที่ไม่เป็นสเตชันนารีแบบบอกซ์เจนคินส์	22
2.7 การทดสอบการแจกแจงแบบปกติ.....	25
บทที่ 3 การทำนายค่า BOD สำหรับน้ำในแม่น้ำเจ้าพระยาโดยกระบวนการอัลฟาทริมา.....	28
3.1 พื้นที่และการเก็บรวบรวมข้อมูล.....	28

3.2 การทำนายค่า BOD โดยใช้กระบวนการอัลฟาทริมา.....	31
บทที่ 4 ผลการวิเคราะห์ข้อมูลวิจัย	38
4.1 ข้อมูลชุดที่ 1	38
4.2 ข้อมูลชุดที่ 2	45
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	60
รายการอ้างอิง	63
ประวัติผู้เขียนวิทยานิพนธ์	65



สารบัญตาราง

ตารางที่ 1 ลักษณะของ ρ_K และ ρ_{KK} สำหรับตัวแบบ ARMA(p,q)19

ตารางที่ 2 ที่ตั้งของสถานีของจุดตรวจวัดคุณภาพน้ำ แม่น้ำเจ้าพระยา.....30

ตารางที่ 3 ตัวอย่างชุดข้อมูล BOD34

ตารางที่ 4 การแต่งขอบข้อมูลครั้งที่ 1 สำหรับชุดข้อมูล BOD.....36

ตารางที่ 5 การแต่งขอบข้อมูลครั้งที่ 2 สำหรับชุดข้อมูล BOD36

ตารางที่ 6 การแต่งขอบข้อมูลครั้งที่ 3 สำหรับชุดข้อมูล BOD.....37

ตารางที่ 7 ค่า α สำหรับชุดข้อมูล37

ตารางที่ 8 ตัวแบบของกระบวนการอัลฟาทรिमอาริม่าของข้อมูลชุดที่ 1 สำหรับแต่ละสถานี.....38

ตารางที่ 9 ค่า α ของข้อมูลชุดที่ 1 สำหรับแต่ละสถานี.....39

ตารางที่ 10 เปรียบเทียบค่า RMSE ระหว่างกระบวนการอัลฟาทรिमอาริม่ากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 1.....39

ตารางที่ 11 เปรียบเทียบเปอร์เซ็นต์ของความคลาดเคลื่อนสัมพัทธ์ ระหว่างกระบวนการอัลฟาทรिमอาริม่ากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 1.....40

ตารางที่ 12 ตัวแบบของกระบวนการอัลฟาทรिमอาริม่าของข้อมูลชุดที่ 2 สำหรับแต่ละสถานี.....45

ตารางที่ 13 ค่า α ของข้อมูลชุดที่ 2 สำหรับแต่ละสถานี46

ตารางที่ 14 เปรียบเทียบค่า RMSE ระหว่างกระบวนการอัลฟาทรिमอาริม่ากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 2.....47

ตารางที่ 15 เปรียบเทียบเปอร์เซ็นต์ของความคลาดเคลื่อนสัมพัทธ์ ระหว่างกระบวนการอัลฟาทรिमอาริม่ากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 2.....48

สารบัญภาพ

ภาพที่ 1	สถานีจุดตรวจวัดคุณภาพน้ำทั้ง 18 สถานี.....	29
ภาพที่ 2	กระบวนการอัลฟาทริมอาร์มีมา	31
ภาพที่ 3	อิทธิพลแก๊รมของตัวอย่างชุดข้อมูล BOD.....	35
ภาพที่ 4	กราฟแสดงค่า BOD ของสถานีที่ 3 สำหรับข้อมูลชุดที่ 1	41
ภาพที่ 5	กราฟแสดงค่า BOD ของสถานีที่ 17 สำหรับข้อมูลชุดที่ 1	42
ภาพที่ 6	กราฟแสดงค่า BOD ของสถานีที่ 18 สำหรับข้อมูลชุดที่ 1	43
ภาพที่ 7	กราฟแสดงค่า BOD ของสถานีที่ 20 สำหรับข้อมูลชุดที่ 1	44
ภาพที่ 8	กราฟแสดงค่า BOD ของสถานีที่ 1 สำหรับข้อมูลชุดที่ 2	49
ภาพที่ 9	กราฟแสดงค่า BOD ของสถานีที่ 3 สำหรับข้อมูลชุดที่ 2	50
ภาพที่ 10	กราฟแสดงค่า BOD ของสถานีที่ 6 สำหรับข้อมูลชุดที่ 2	51
ภาพที่ 11	กราฟแสดงค่า BOD ของสถานีที่ 8 สำหรับข้อมูลชุดที่ 2	52
ภาพที่ 12	กราฟแสดงค่า BOD ของสถานีที่ 10 สำหรับข้อมูลชุดที่ 2	53
ภาพที่ 13	กราฟแสดงค่า BOD ของสถานีที่ 12 สำหรับข้อมูลชุดที่ 2	54
ภาพที่ 14	กราฟแสดงค่า BOD ของสถานีที่ 15 สำหรับข้อมูลชุดที่ 2	55
ภาพที่ 15	กราฟแสดงค่า BOD ของสถานีที่ 16.1 สำหรับข้อมูลชุดที่ 2.....	56
ภาพที่ 16	กราฟแสดงค่า BOD ของสถานีที่ 17 สำหรับข้อมูลชุดที่ 2.....	57
ภาพที่ 17	กราฟแสดงค่า BOD ของสถานีที่ 18 สำหรับข้อมูลชุดที่ 2.....	58
ภาพที่ 18	กราฟแสดงค่า BOD ของสถานีที่ 20 สำหรับข้อมูลชุดที่ 2.....	59

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

น้ำเป็นทรัพยากรธรรมชาติที่มีความสำคัญที่สุด ต่อการดำรงชีวิตของมนุษย์ดังพระราชดำรัสของพระบาทสมเด็จพระเจ้าอยู่หัวภูมิพลอดุลยเดช พระราชทานไว้ว่า

“...หลักสำคัญว่า ต้องมีน้ำบริโภค น้ำใช้ น้ำเพื่อการเพาะปลูก เพราะว่าชีวิต อยู่ที่นั่น ถ้ามีน้ำคนอยู่ได้ ถ้าไม่มีน้ำคนอยู่ไม่ได้ ไม่มีไฟฟ้าคนอยู่ได้ แต่ถ้ามีไฟฟ้าไม่มีน้ำคนอยู่ไม่ได้...”

(พระราชดำรัสของพระบาทสมเด็จพระเจ้าอยู่หัวภูมิพลอดุลยเดช พระราชทานไว้เมื่อวันที่ 17 มีนาคม 2529 ณ พระตำหนักจิตรลดารโหฐาน)

นอกจากจะต้องมีแหล่งน้ำและปริมาณน้ำที่เพียงพอแล้ว สิ่งสำคัญที่ต้องพิจารณา คือ คุณภาพของน้ำ ซึ่งคุณภาพน้ำหมายถึง ความเหมาะสมของน้ำเพื่อใช้ในกิจกรรมเฉพาะของมนุษย์ ถ้า น้ำในแหล่งน้ำมีคุณภาพไม่ดี แม้จะมีปริมาณมากเพียงใดก็ไร้ประโยชน์ ในอดีตที่ผ่านมาส่วนใหญ่ น้ำจะมีคุณภาพดี แต่ปัจจุบันสภาพแวดล้อมเปลี่ยนแปลงไปมาก เนื่องจากการพัฒนาด้านอุตสาหกรรม ด้านเกษตรกรรม รวมทั้งการเพิ่มขึ้นของประชากร ทำให้มีการปล่อยของเสียและสารพิษสู่สิ่งแวดล้อม มากขึ้น ส่งผลต่อคุณภาพน้ำในแหล่งน้ำนั้น ๆ และคาดว่าในอนาคตอันใกล้ ทรัพยากรน้ำจะเป็น ทรัพยากรที่มีข้อจำกัด เนื่องจากปัญหาแหล่งน้ำเสื่อมโทรม และมลพิษทางน้ำ จึงจำเป็นที่จะต้องให้ ความสำคัญต่อการอนุรักษ์ทรัพยากรน้ำ เพื่อให้ได้น้ำที่มีปริมาณพอเหมาะ เพียงพอแก่ความต้องการ ให้ได้น้ำที่มีคุณภาพดี ใสสะอาด ไม่ขุ่นข้น และตามลำห้วยลำธารก็ให้น้ำไหลอย่างสม่ำเสมอตลอดไป รวมทั้งเพื่อเพิ่มระดับน้ำใต้ดินและลดการสูญเสียจากน้ำไหลบ่า

กรมควบคุมมลพิษ โดยสำนักจัดการคุณภาพน้ำ กระทรวงทรัพยากรธรรมชาติและ สิ่งแวดล้อม มีการติดตามตรวจวัดคุณภาพน้ำของแม่น้ำเจ้าพระยาทุก ๆ 3 เดือน โดยมีสถานีจุด ตรวจวัดคุณภาพน้ำทั้งหมด 18 สถานี ซึ่งผลการตรวจวัดคุณภาพน้ำในแต่ละปี ได้รายงานผ่านเว็บไซต์ ของกรมควบคุมมลพิษ

ดังนั้นผู้วิจัยจึงสนใจการวิเคราะห์อนุกรมเวลาของข้อมูลคุณภาพน้ำ สำหรับแม่น้ำเจ้าพระยา โดยพิจารณาจากค่าความต้องการออกซิเจนทางชีวเคมี (Biochemical Oxygen Demand: BOD) ซึ่งเป็นปริมาณออกซิเจนที่จุลินทรีย์ต้องการใช้ในการย่อยสลายสารอินทรีย์ในน้ำ โดยการประยุกต์ใช้การวิเคราะห์อนุกรมแบบบ็อกซ์-เจนคินส์ (Box-Jenkins) เพื่อสร้างกระบวนการใหม่ที่สามารถทำนายค่า BOD ได้ และเปรียบเทียบผลการทำนายค่า BOD สำหรับกระบวนการที่สร้างขึ้นใหม่กับเทคนิคการทำให้เรียบ (Smoothing method) และการพยากรณ์อนุกรมเวลาด้วยวิธีของบ็อกซ์-เจนคินส์ ซึ่งกระบวนการใหม่ที่สร้างขึ้นนั้นมีแม่นยำกว่าเทคนิคการทำให้เรียบและการพยากรณ์อนุกรมเวลาด้วยวิธีของบ็อกซ์-เจนคินส์ และงานวิจัยนี้จะเป็นประโยชน์ต่อภาครัฐและประชาชนในการวางแผนเพื่อหาแนวทางแก้ปัญหาในอนาคต

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาพฤติกรรมของค่า BOD ที่สถานีจุดตรวจวัดคุณภาพน้ำของแม่น้ำเจ้าพระยา ณ เวลาต่าง ๆ และใช้ตัวแบบเชิงคณิตศาสตร์ทำนายค่า BOD ในอนาคตจากอนุกรมเวลาในอดีต

1.3 ประโยชน์ที่ได้รับ

1. สามารถทำนายค่า BOD ที่สถานีจุดตรวจวัดคุณภาพน้ำ ณ เวลาต่าง ๆ ของแม่น้ำเจ้าพระยาในอนาคต โดยใช้ค่า BOD จากอนุกรมเวลาในอดีตได้อย่างถูกต้องแม่นยำ
2. สามารถสร้างกระบวนการอัลฟาทรिमอริมาที่เหมาะสมที่สุด สำหรับชุดข้อมูลคุณภาพน้ำที่ทำการศึกษา
3. กระบวนการที่สร้างขึ้นสามารถประยุกต์ใช้กับข้อมูลคุณภาพน้ำที่มีลักษณะใกล้เคียงกับข้อมูลของแม่น้ำเจ้าพระยาได้

1.4 ลำดับขั้นตอนในการเสนอผลการวิจัย

- ในบทที่ 2 จะกล่าวถึงความรู้พื้นฐานสำหรับการสร้างตัวแบบและงานวิจัยที่เกี่ยวข้อง
- ในบทที่ 3 จะกล่าวถึงรายละเอียด ขั้นตอนวิธี ของกระบวนการที่สร้างขึ้นมาใหม่
- ในบทที่ 4 จะแสดงผลการศึกษา
- ในบทสุดท้าย จะกล่าวถึงข้อสรุป ปัญหาและงานวิจัยที่น่าสนใจในอนาคต

1.5 ขอบเขตของการวิจัย

1. ข้อมูล BOD ของแม่น้ำเจ้าพระยาในปี พ.ศ. 2539-2556 ถูกเก็บรวบรวมข้อมูลโดยกรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม จากสถานีจุดตรวจวัดคุณภาพน้ำ 18 สถานี

2. การวิจัยในครั้งนี้ศึกษาค่า BOD จากอนุกรมเวลาในอดีตโดยศึกษาจากชุดข้อมูล 2 ชุด คือ ชุดที่ 1 ข้อมูล BOD ถูกเก็บรวบรวมในเดือนมิถุนายนและธันวาคมจำนวน 33 ค่าและชุดที่ 2 ข้อมูล BOD ถูกเก็บรวบรวมในเดือนกุมภาพันธ์และสิงหาคมจำนวน 34 ค่า นั่นคือ จำนวนฤดูกาลต่อปี (L) เท่ากับ 2

3. การวิจัยในครั้งนี้วิเคราะห์คุณภาพน้ำจากค่า BOD เท่านั้น โดยไม่สนใจปัจจัยอื่น ๆ และทำนายค่า BOD ในอนาคตจากอนุกรมเวลาในอดีต ในแต่ละชุดข้อมูล

4. การวิจัยในครั้งนี้เปรียบเทียบประสิทธิภาพของเทคนิคการทำให้เรียบและการพยากรณ์อนุกรมเวลาด้วยวิธีของบ็อกซ์-เจนกินส์กับกระบวนการที่สร้างขึ้นใหม่เท่านั้น



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยนี้ผู้วิจัยเริ่มจากการวิเคราะห์หอนุกรมเวลาของค่า BOD โดยอาศัยเทคนิคการพยากรณ์เชิงปริมาณ (Qualitative forecasting technique) คือเทคนิคการทำให้เรียบและการวิเคราะห์หอนุกรมแบบบอกซ์เจนคินส์ เพื่อสร้างรูปแบบหรือสมการพยากรณ์ สำหรับการพยากรณ์ข้อมูลหรือเหตุการณ์ในอนาคต

2.1 แนวความคิดพื้นฐานเกี่ยวกับการพยากรณ์

2.1.1 ความหมายและความสำคัญของการพยากรณ์

การพยากรณ์ หมายถึงการคาดคะเนหรือการทำนายการเกิดเหตุการณ์ต่าง ๆ ในอนาคต โดยอาศัยข้อมูล ประสบการณ์ ความรู้ความสามารถของผู้พยากรณ์ที่เกิดขึ้นในอดีต มาศึกษารูปแบบของการเกิดเหตุการณ์ในอนาคต การพยากรณ์มีบทบาทสำคัญอย่างมากในการวางแผนและการตัดสินใจเกี่ยวกับการดำเนินงานของบุคคลทุกอาชีพ และขององค์กรต่าง ๆ เช่น การค้า การส่งออก การวางแผนเกี่ยวกับธุรกิจ การเกษตร การสาธารณสุข เป็นต้น

2.1.2 การตรวจสอบความแม่นยำของการพยากรณ์

หลังจากการเลือกวิธีการพยากรณ์แล้ว ขั้นตอนถัดไปก็คือ การนำข้อมูลมาวิเคราะห์เพื่อจะพยากรณ์ต่อไป สำหรับการตรวจสอบความแม่นยำของการพยากรณ์เชิงคุณภาพนั้นจะไม่มีหลักเกณฑ์ที่แน่ชัด หรือหลักเกณฑ์ที่แน่นอนเหมือนกับการตรวจสอบความแม่นยำของการพยากรณ์เชิงปริมาณ เพราะการพยากรณ์เชิงคุณภาพต้องอาศัยประสบการณ์ความรู้ความชำนาญของผู้พยากรณ์เป็นหลัก ส่วนการพยากรณ์เชิงปริมาณมีการนำหลักสถิติไปใช้กับข้อมูลที่เก็บรวบรวมมาได้

สำหรับค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) เป็นการตรวจสอบความแม่นยำของการพยากรณ์เชิงปริมาณ โดยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เป็นการวัดความแม่นยำจากค่าเฉลี่ยผลรวมกำลังสองของค่าความคลาดเคลื่อนของการพยากรณ์ (Sum Square Error: SSE) โดยจะให้ความสำคัญกับค่าพยากรณ์ที่แตกต่างไปจากค่าจริงมาก ดังนั้นในงานวิจัยนี้จะใช้ RMSE (Root Mean Square Error) แทน SSE และ MSE เพราะการใช้ RMSE เป็นการปรับค่าความคลาดเคลื่อนให้มีหน่วยวัดเดียวกับค่า BOD ที่วัดได้

โดยที่

$$SSE = \sum_{i=1}^n \left(Y_i - \hat{Y}_i \right)^2$$

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}$$

และ

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

เมื่อ Y_i แทนข้อมูลสังเกตที่ i สำหรับ $i=1, 2, 3, \dots, n$

\hat{Y}_i แทนค่าพยากรณ์ที่ i สำหรับ $i=1, 2, 3, \dots, n$

n แทนจำนวนข้อมูลทั้งหมด

2.2 อนุกรมเวลาและสหสัมพันธ์

2.2.1 อนุกรมเวลาและส่วนประกอบของอนุกรมเวลา

อนุกรมเวลาเป็นค่าข้อมูลหรือค่าสังเกตที่เก็บรวบรวมตามลำดับเวลาที่เกิดขึ้นอย่างต่อเนื่อง โดยช่วงเวลาที่เกิดขึ้นรวบรวมข้อมูลอาจจะเป็นรายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปีก็ได้ เช่น ยอดขายสินค้าแผนกซูเปอร์มาร์เก็ต ของห้างสรรพสินค้าแห่งหนึ่ง ยอดขายอาจจะเป็นรายวัน รายเดือน รายปีก็ได้ หรืออาจจะเป็นราคาหุ้นที่เปลี่ยนแปลงไปในแต่ละวัน การวิเคราะห์อนุกรมเวลา มีวัตถุประสงค์หลักคือ เพื่อศึกษาแบบแผนการเคลื่อนไหวของอนุกรมเวลาและสร้างรูปแบบอนุกรม แทนลักษณะการเคลื่อนไหวดังกล่าว

ก่อนการวิเคราะห์อนุกรมเวลาเพื่อการพยากรณ์ค่าในอนาคต ต้องทราบก่อนว่าอนุกรมเวลา มีการเคลื่อนไหวเนื่องจากส่วนประกอบหลักใดบ้าง เพื่อกำหนดรูปแบบอนุกรมเวลาจากส่วนประกอบหลักที่เป็นไปได้ ได้แก่

1. แนวโน้ม หมายถึงการเคลื่อนไหวของอนุกรมเวลาในระยะยาวว่าจะมีแนวโน้มเพิ่มขึ้นหรือลดลง และลักษณะของแนวโน้มมีหลายลักษณะ เช่น แนวโน้มเส้นตรง แนวโน้มเอ็กซ์โพเนนเชียล แนวโน้มกำลังสอง แนวโน้มกำลัง เป็นต้น

2. ความแปรผันตามฤดูกาล หมายถึงการเคลื่อนไหวของข้อมูลที่เกิดขึ้นเนื่องจากอิทธิพลของฤดูกาล ซึ่งจะเกิดขึ้นซ้ำ ๆ ในแต่ละช่วง โดยลักษณะการเคลื่อนไหวในแต่ละช่วงไม่แตกต่างกัน

3. ความแปรผันตามวัฏจักร หมายถึงการเคลื่อนไหวของอนุกรมเวลาที่เก็บรวบรวมเป็นระยะเวลาต่อเนื่องกันหลายปีอาจจะแสดงอิทธิพลของวัฏจักร โดยหนึ่งวัฏจักรจะคลุมระยะเวลาหลายปี

4. เหตุการณ์ที่ผิดปกติ หมายถึงการเคลื่อนไหวของข้อมูลที่ไม่ใช่รูปแบบที่แน่นอน ลักษณะของข้อมูลที่เกิดขึ้นส่วนใหญ่จะเป็นลักษณะของเหตุการณ์ที่ไม่ได้คาดการณ์ไว้ล่วงหน้า

2.2.2 สัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Coefficient auto correlation : r_K)

สัมประสิทธิ์สหสัมพันธ์ในตัวเอง จะเป็นค่าวัดความสัมพันธ์ระหว่างข้อมูลอนุกรมเวลาในชุดเดียวกันโดยมีช่วงเวลาห่างกันเท่ากับ K (lag K) เช่น การหาความสัมพันธ์ระหว่างข้อมูลอนุกรมเวลา $Y_t, Y_{t-1}, Y_{t-2}, \dots$ กับ $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$ จะเห็นว่าข้อมูลอนุกรมเวลาดังกล่าวเป็นข้อมูลชุดเดียวกัน แต่มีช่วงเวลาห่างกันเท่ากับ 1 (lag 1) สัมประสิทธิ์สหสัมพันธ์ในตัวเองมีสูตรการคำนวณดังนี้

$$r_K = \frac{\sum_{t=1}^{n-K} (Y_t - \bar{Y})(Y_{t+K} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

โดยที่ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

K คือ จำนวนช่วงเวลาที่ข้อมูลอยู่ห่างกัน สำหรับ $K = 1, 2, 3, \dots$

n คือ จำนวนข้อมูลทั้งหมด

\bar{Y} คือ ค่าเฉลี่ยของข้อมูลทั้งหมด

และความแปรปรวนของ r_K ($\sigma_{r_K}^2$) มีสูตรการคำนวณคือ

$$\sigma_{r_K}^2 = \begin{cases} \frac{1}{n} & , K = 1 \\ \frac{1}{n} \left(1 + 2 \sum_{j=1}^{K-1} r_j^2 \right) & , K = 2, 3, 4, \dots \end{cases}$$

คุณสมบัติของ r_K มีดังนี้

1. r_K มีค่าอยู่ระหว่าง -1 ถึง 1 เมื่อขนาดของ r_K ที่วัดด้วย $|r_K|$ มีค่าเข้าใกล้ 0 แสดงว่ามีความสัมพันธ์กันต่ำ ถ้า r_K มีค่าน้อยกว่า 0 แสดงว่ามีความสัมพันธ์กันในทิศทางที่ตรงกันข้าม และถ้า r_K มีค่ามากกว่า 0 แสดงว่ามีความสัมพันธ์กันในทิศทางไปทางเดียวกัน

2. $r_K = r_{-K}$ คือการวัดความสัมพันธ์ระหว่าง Y_t กับ Y_{t-K} หรือระหว่าง Y_t กับ Y_{t-K} ซึ่งต่างก็เป็นค่าสังเกตที่อยู่ห่างกัน K ค่าไม่แตกต่างกัน

r_K เป็นค่าวัดลักษณะของตัวอย่างที่สุ่ม และเป็นค่าประมาณของ ρ_K ซึ่ง ρ_K จะเป็นค่าวัดลักษณะของประชากร ค่า r_K มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น $\sigma_{r_K}^2$ หรือ $r_K \sim N(0, \sigma_{r_K}^2)$ สามารถทดสอบสมมติฐานเกี่ยวกับ ρ_K ได้โดยกำหนดสมมติฐานหลักและสมมติฐานทางเลือกดังนี้

$$H_0 : \rho_K = 0$$

$$H_1 : \rho_K \neq 0$$

ตัวสถิติที่ใช้ทดสอบคือ $Z = \frac{r_K}{S_{r_K}} = \frac{r_K}{\sqrt{\frac{1}{n}}} = \sqrt{n} r_K$ จะยอมรับ H_0 เมื่อ $-Z_{\frac{\lambda}{2}} \leq Z \leq Z_{\frac{\lambda}{2}}$ ที่ระดับ

นัยสำคัญเท่ากับ λ ในทำนองเดียวกันอาจจะใช้ค่า r_K เป็นสถิติที่ใช้ทดสอบก็ได้ ซึ่งจะยอมรับ H_0 เมื่อ $-Z_{\frac{\lambda}{2}} \cdot \frac{1}{\sqrt{n}} \leq r_K \leq Z_{\frac{\lambda}{2}} \cdot \frac{1}{\sqrt{n}}$ โดยทั่วไปการทดสอบสมมติฐานนิยมใช้ระดับนัยสำคัญที่ 0.05 ดังนั้น การยอมรับ H_0 ก็ต่อเมื่อ ($Z_{0.025} = 1.96$)

$$-1.96 \cdot \frac{1}{\sqrt{n}} \leq r_K \leq 1.96 \cdot \frac{1}{\sqrt{n}}$$

2.2.3 สัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน (Partial coefficient auto correlation :

r_{KK})

สัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนจะเป็นค่าวัดความสัมพันธ์ระหว่างค่าสังเกตที่อยู่ห่างกัน K ช่วงเวลา เมื่อกำหนดให้ค่าสังเกต ณ เวลาอื่นคงที่ เช่น การหาค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนระหว่าง $Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}$ กับ $Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, Y_{t-5}$ ซึ่งมีช่วงเวลาห่างกันอยู่ 1 โดยกำหนดให้ $Y_{t+1}, Y_{t+2}, Y_{t+3}, \dots$ และ $Y_{t-6}, Y_{t-7}, Y_{t-8}, \dots$ มีค่าคงที่ สำหรับสูตรที่ใช้คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนเป็นดังนี้

$$r_{KK} = \begin{cases} r_1 & , K = 1 \\ \frac{r_K - \sum_{j=1}^{K-1} r_{K-1,j} r_{K-j}}{1 - \sum_{j=1}^{K-1} r_{K-1,j} r_j} & , K = 2, 3, 4, \dots \end{cases}$$

เมื่อ $r_{K,j}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนที่ช่วงเวลาห่างกัน K และ j

โดยที่ $r_{K,j} = r_{K-1,j} - r_{KK} \cdot r_{K-1,K-j}$ สำหรับ $j=1, 2, 3, \dots, K-1$

และ $\sigma_{r_{KK}}^2 = \frac{1}{n}$ สำหรับ $j=1, 2, 3, \dots, K$

2.3 การทดสอบแนวโน้มและฤดูกาล

ในการพิจารณาการเคลื่อนไหวของอนุกรมเวลา เริ่มจากการพล็อตกราฟของอนุกรมเวลา เพื่อวิเคราะห์อนุกรมเวลาว่า มีการเคลื่อนไหวเนื่องจากส่วนประกอบหลักใดบ้าง แต่ในบางครั้งไม่

สามารถสรุปจากพล็อตกราฟของอนุกรมเวลา จึงต้องใช้การทดสอบสมมติฐานเพื่อหารูปแบบอนุกรมเวลาที่เหมาะสม การทดสอบสมมติฐานมีทั้งแบบไม่อิงพารามิเตอร์และการทดสอบแบบอิงพารามิเตอร์ การทดสอบแบบไม่อิงพารามิเตอร์ใช้เมื่อข้อมูลที่นำมาวิเคราะห์ไม่มีการแจกแจงตามข้อสมมติของการทดสอบแบบอิงพารามิเตอร์ การทดสอบแบบไม่อิงพารามิเตอร์ให้อำนาจของการทดสอบที่ต่ำกว่าอำนาจของการทดสอบแบบอิงพารามิเตอร์ เพราะ ไม่พิจารณาขนาดของค่าสังเกตแต่พิจารณาเฉพาะลำดับและทิศทางของค่าสังเกตเท่านั้น [1, 2] โดยผู้วิจัยจะนำการทดสอบดังกล่าวมาใช้ในการทดสอบอนุกรมเวลาของค่า BOD ด้วย

การทดสอบแนวโน้มแบบไม่อิงพารามิเตอร์มีหลายการทดสอบ เช่น การทดสอบแบบวิ่ง (runs test) การทดสอบแบบจุดกลับ (turning point test) การทดสอบแบบเครื่องหมาย (sign test) การทดสอบสหสัมพันธ์ของ Daniels และการทดสอบสหสัมพันธ์ของ Kendall ส่วนการทดสอบแนวโน้มแบบอิงพารามิเตอร์ ได้แก่ การทดสอบสัดส่วนของ Von Neumann การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง และการวิเคราะห์การถดถอย โดยแต่ละการทดสอบมีตัวทดสอบสถิติที่ต่างกันสำหรับอนุกรมเวลาขนาดใหญ่และขนาดเล็ก

การทดสอบฤดูกาลแบบไม่อิงพารามิเตอร์ ได้แก่ การทดสอบความแตกต่างระหว่างกลุ่มของ Kruskal และ Wallis ซึ่งเป็นการทดสอบที่ใช้ในการวิเคราะห์ทางสถิติเพื่อทดสอบว่า ไม่มีความแตกต่างระหว่างกลุ่มหรือทดสอบว่าแต่ละตัวอย่างมาจากประชากรเดียวกัน โดยนำการทดสอบความแตกต่างระหว่างกลุ่มของ Kruskal และ Wallis มาใช้เพื่อทดสอบกับอนุกรมเวลาที่ไม่มีการเคลื่อนไหวจากฤดูกาล โดยมีฤดูกาลเป็นกลุ่ม ส่วนการทดสอบฤดูกาลแบบอิงพารามิเตอร์ ได้แก่ การทดสอบ $H_0 : \rho_L = 0$ และการวิเคราะห์การถดถอย

ในงานวิจัยนี้ใช้การทดสอบสัดส่วนของ Von Neumann สำหรับการทดสอบแนวโน้มแบบอิงพารามิเตอร์ และการทดสอบ $H_0 : \rho_L = 0$ สำหรับการทดสอบฤดูกาลแบบอิงพารามิเตอร์ โดยมีวิธีการคำนวณดังนี้

2.3.1 การทดสอบสัดส่วนของ Von Neumann สำหรับการทดสอบแนวโน้มแบบอิงพารามิเตอร์

การทดสอบสัดส่วนของ Von Neumann เป็นการทดสอบแนวโน้มที่ตัวทดสอบสถิติได้จากการพิจารณาขนาดของผลต่างระหว่างค่าข้อมูลและค่าข้อมูลก่อนหน้า (Y_t และ Y_{t-1})

การทดสอบสัดส่วนของ Von Neumann มีขั้นตอนดังนี้

1. กำหนด H_0 : อนุกรมเวลาไม่มีแนวโน้มและความคลาดเคลื่อนเป็นอิสระกัน
 H_1 : อนุกรมเวลามีแนวโน้มและความคลาดเคลื่อนเป็นอิสระกัน

2. ใช้ตัวทดสอบสถิติ

$$M = \frac{\sum_{t=2}^n (Y_t - Y_{t-1})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

ที่ระดับนัยสำคัญ λ มีช่วงวิกฤต $CR: M < M_{p,n}$ หรือ $M > M_{1-p,n}$ และ $p = \frac{\lambda}{2}$ ซึ่งค่า $M_{p,n}$ เป็นค่าวิกฤตจากตารางสถิติ ส่วนค่า $M_{1-p,n}$ คำนวณได้จาก $M_{1-p,n} = 4 - M_{p,n}$

2.3.2 การทดสอบ $H_0: \rho_L = 0$ สำหรับการทดสอบฤดูกาลแบบอิงพารามิเตอร์

อนุกรมเวลาที่มีการเคลื่อนไหวเนื่องจากฤดูกาลจะมีค่า r_K สูง สำหรับ K เท่ากับผลคูณของ $\frac{L}{2}$ หรือ $K = \frac{L}{2}, L, \frac{3L}{2}, 2L, \dots$ เมื่อ L เป็นจำนวนฤดูกาลต่อปี การทดสอบว่าอนุกรมเวลาที่มีการเคลื่อนไหว เนื่องจากฤดูกาลจึงพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อมูลในอนุกรมเวลาที่อยู่ห่างกัน L ช่วงเวลา หรือ r_L นั่นคือเมื่ออนุกรมเวลาที่มีการเคลื่อนไหวเนื่องจากฤดูกาล ข้อมูลในฤดูกาลเดียวกันแต่ต่างปีกันมีสหสัมพันธ์ในทางเดียวกัน การทดสอบ $H_0: \rho_L = 0$ กับ $H_1: \rho_L > 0$ ใช้ตัวทดสอบสถิติ r_L ที่ระดับนัยสำคัญ λ มีช่วงวิกฤตเป็น $CR: r_L \geq \frac{Z_\lambda}{\sqrt{n}}$ ซึ่ง Z_λ เป็นค่าวิกฤตจากตารางสถิติ

2.4 เทคนิคการทำให้เรียบ

การทำให้เรียบ หมายถึงการใช้ข้อมูลหรือค่าสังเกตในอดีตส่วนหนึ่งหรือทั้งหมดในการสร้างสมการพยากรณ์ โดยน้ำหนักที่ให้กับค่าสังเกตแต่ละค่าแตกต่างกัน วิธีการให้น้ำหนักแก่ค่าสังเกตหรือข้อมูลในอดีตนั้นผู้คิดค้นไว้หลายวิธีด้วยกันขึ้นอยู่กับลักษณะข้อมูลหรือค่าสังเกตว่ามีส่วนประกอบของแนวโน้มหรือฤดูกาลหรือไม่ เหตุผลประการสำคัญที่มีการใช้เทคนิคการทำให้เรียบ เนื่องจากในข้อมูลอนุกรมเวลามักมีความแปรผันเนื่องจากเหตุการณ์ที่ผิดปกติรวมอยู่ด้วย ซึ่งความแปรผันเนื่องจากเหตุการณ์ที่ผิดปกติจะทำให้ไม่สามารถเห็นส่วนประกอบอื่น ๆ การกำจัดอิทธิพลของความแปรผันเนื่องจากเหตุการณ์ที่ผิดปกติจะทำให้สามารถเห็นส่วนประกอบอื่น ๆ ของอนุกรมเวลาได้และสามารถจะพยากรณ์ค่าอนุกรมเวลาในอนาคตได้

สำหรับเทคนิคการทำให้เรียบมีด้วยกันอยู่หลายวิธี ขึ้นอยู่กับลักษณะของข้อมูลอนุกรมเวลา ถ้าอนุกรมเวลาไม่มีแนวโน้มและฤดูกาลจะใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย (simple moving average method) วิธีการเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก (weighted moving average method) และวิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่าย (single exponential smoothing method) ถ้าอนุกรมเวลามีแนวโน้มแต่ไม่มีความแปรผันตามฤดูกาล จะใช้วิธีการเฉลี่ยเคลื่อนที่สองครั้ง

(double moving average method) วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสองครั้ง (double exponential smoothing method) และวิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสามครั้ง (triple exponential smoothing method) ถ้าอนุกรมเวลาไม่มีแนวโน้มแต่มีความแปรผันตามฤดูกาล จะใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลที่มีความแปรผันตามฤดูกาล (seasonal simple exponential smoothing method) ถ้าอนุกรมเวลา มีทั้งแนวโน้มและความแปรผันตามฤดูกาล จะใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลด้วยวิธีของวินเตอร์ (triple exponential smoothing : Winters's method) [3-6]

จากการวิเคราะห์ลักษณะการเคลื่อนไหวของอนุกรมเวลา BOD พบว่าชุดข้อมูล BOD มี 3 ลักษณะคืออนุกรมเวลาไม่มีแนวโน้มและฤดูกาล อนุกรมเวลาที่มีแนวโน้มแต่ไม่มีความแปรผันตามฤดูกาลและอนุกรมเวลาไม่มีแนวโน้มแต่มีความแปรผันตามฤดูกาล ดังนั้นในงานวิจัยนี้จะนำเสนอเทคนิคการทำให้เรียบที่ทำให้ชุดข้อมูล BOD มีค่า RMSE น้อยที่สุด นั่นคือ การเฉลี่ยเคลื่อนที่อย่างง่ายสำหรับอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสองครั้งสำหรับอนุกรมเวลาที่มีแนวโน้มแต่ไม่มีความแปรผันตามฤดูกาล และวิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลที่มีความแปรผันตามฤดูกาลสำหรับอนุกรมเวลาไม่มีแนวโน้มแต่มีความแปรผันตามฤดูกาล

2.4.1 การเฉลี่ยเคลื่อนที่อย่างง่าย

การเฉลี่ยเคลื่อนที่อย่างง่าย เป็นวิธีการพยากรณ์ที่เหมาะสมกับข้อมูลที่มีลักษณะการเคลื่อนไหวอยู่ในระดับคงที่หรือค่อนข้างราบเรียบตามแนวนอน และเหมาะกับการพยากรณ์ระยะสั้น การพยากรณ์ตามวิธีนี้จะมีการนำค่าสังเกตหรือข้อมูลล่าสุดจำนวนหนึ่ง เช่น 3 ค่าหรือ 5 ค่า มาหาค่าเฉลี่ย ซึ่งการเฉลี่ยนี้เป็นการเฉลี่ยที่ให้น้ำหนักกับค่าสังเกตแต่ละค่าเท่ากัน สำหรับจำนวนค่าสังเกตที่จะนำมาหาค่าเฉลี่ยนั้นถ้าจำนวนข้อมูลมีการเคลื่อนไหวมากควรใช้จำนวนค่าสังเกตมาก และในทางกลับกันถ้าข้อมูลค่อนข้างเรียบควรใช้จำนวนค่าสังเกตน้อยลงตามลำดับ

สำหรับตัวแบบของข้อมูลที่มีลักษณะการเคลื่อนไหวอยู่ในระดับคงที่จะมีตัวแบบเป็น

$$Y_t = \beta_0 + \varepsilon_t$$

เมื่อ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

β_0 คือ ค่าพารามิเตอร์ของตัวแบบ

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t

สำหรับสูตรที่ใช้ในการพยากรณ์มีดังนี้

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-k+1}}{k}$$

โดยที่ \hat{Y}_{t+1} คือ ค่าพยากรณ์ ณ เวลา $t+1$

Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

k คือ จำนวนข้อมูลที่นำมาหาค่าเฉลี่ย

2.4.2 การทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสองครั้ง

การทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสองครั้ง จะมีชื่อเรียกอีกสองแบบคือ Brown's one-parameter linear model และ linear exponential smoothing วิธีการพยากรณ์แบบนี้เหมาะกับข้อมูลที่มีการเคลื่อนไหวเป็นแนวโน้มแบบเส้นตรงและไม่มีความแปรผันตามฤดูกาล นอกจากนี้ยังเหมาะสมกับการพยากรณ์ในระยะสั้นจนถึงการพยากรณ์ในระยะปานกลาง ข้อมูลที่ใช้ในการคำนวณหาค่าพยากรณ์ควรมีอย่างน้อย 5 ค่า ขึ้นไป

สำหรับข้อมูลที่มีลักษณะเป็นแนวโน้มแบบเส้นตรงจะมีตัวแบบเป็น

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

เมื่อ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

β_0, β_1 คือ ค่าพารามิเตอร์ของตัวแบบ

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t

ค่าของ β_0 และ β_1 ในตัวแบบนั้นเป็นค่าพารามิเตอร์ซึ่งไม่ทราบค่า จึงต้องประมาณด้วยค่า a และ b สำหรับวิธีการพยากรณ์แบบนี้ค่าของ a และ b จะเปลี่ยนแปลงไปตามเวลา (t) และมีสูตรสำหรับคำนวณหาค่าพยากรณ์ดังนี้

$$\hat{Y}_{t+m} = a_t + b_t m$$

โดยที่ \hat{Y}_{t+m} คือ ค่าพยากรณ์ ณ เวลา $t+m$

a_t, b_t คือ ค่าประมาณค่าพารามิเตอร์ β_0 และ β_1 ณ เวลา t

m คือ จำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า

เมื่อ $a_t = 2E_{1,t} - E_{2,t}$

$$b_t = \left(\frac{g}{1-g} \right) (E_{1,t} - E_{2,t})$$

$$E_{1,t} = gY_t + (1-g)E_{1,t-1}$$

$$E_{2,t} = gE_{1,t} + (1-g)E_{2,t-1}$$

ซึ่ง $E_{1,t}$ คือ ค่าที่ได้จากการทำให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่ายในครั้งที่ 1 ณ เวลา t

$E_{2,t}$ คือ ค่าที่ได้จากการทำให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่ายในครั้งที่ 2 ณ เวลา t

g คือ ค่าคงที่การทำให้เรียบ

การสร้างสมการพยากรณ์ตามวิธีการข้างต้นจะเกี่ยวข้องกับค่าที่สำคัญ 2 ค่า คือ

1. ค่าเริ่มต้นของ $E_{1,t}$ และ $E_{2,t}$ ซึ่งมีวิธีการกำหนดไว้ 2 แบบคือ
- ก. กำหนดให้ $E_{1,1} = E_{2,1} = Y_1$ แล้วหา $E_{1,t}$ กับ $E_{2,t}$ เมื่อ $t = 2, 3, 4, \dots$
- ข. กำหนดให้ $E_{1,0} = \hat{\beta}_0 - \frac{(1-g)}{g} \hat{\beta}_1$ และ $E_{2,0} = \hat{\beta}_0 - \frac{2(1-g)}{g} \hat{\beta}_1$

$$\text{ซึ่ง } \hat{\beta}_0 = \bar{Y} - \frac{(n+1)}{2} \hat{\beta}_1 \text{ และ } \hat{\beta}_1 = \frac{12 \left\{ \sum_{t=1}^n \left(t - \frac{n+1}{2} \right) Y_t \right\}}{n^3 - n}$$

เมื่อได้ค่า $E_{1,0}$ และ $E_{2,0}$ แล้วจึงหา $E_{1,t}$ และ $E_{2,t}$ เมื่อ $t = 2, 3, 4, \dots$ ต่อไป

2. ค่าคงที่การทำให้เรียบ (g) ในทางปฏิบัตินิยมกำหนดค่า g ให้น้อยกว่าค่า g ที่ใช้ในวิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่ายหรือผู้พยากรณ์เป็นผู้กำหนดเองก็ได้ หรืออาจจะใช้โปรแกรมสำเร็จรูป ค้นหาค่า g โดยใช้หลักเกณฑ์ที่ทำให้ SEE หรือ MSE หรือ RMSE มีค่าต่ำสุด

2.4.3 การทำให้เรียบสำหรับอนุกรมเวลาที่มีความแปรผันตามฤดูกาล

ในหัวข้อที่ผ่านมาได้กล่าวถึงเทคนิคการทำให้เรียบสำหรับข้อมูลที่ไม่มีแนวโน้มและข้อมูลที่มีแนวโน้ม สำหรับข้อมูลที่มีความแปรผันตามฤดูกาลเพียงอย่างเดียวก็มีวิธีการทำให้เรียบเหมือนกัน เพื่อที่จะพยากรณ์ข้อมูลในอนาคตต่อไป สำหรับตัวแบบอนุกรมเวลาที่มีส่วนประกอบ ความแปรผันตามฤดูกาลเพียงอย่างเดียวได้แก่

$$Y_t = \beta_0 + S_t + \varepsilon_t \quad \text{สำหรับตัวแบบการบวก}$$

$$Y_t = \beta_0 S_t \varepsilon_t \quad \text{สำหรับตัวแบบการคูณ}$$

เมื่อ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

β_0 คือ ค่าพารามิเตอร์ของตัวแบบ

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t

S_t คือ ค่าความแปรผันตามฤดูกาล ณ เวลา t

จากตัวแบบดังกล่าวจะประมาณค่า β_0 ด้วยค่าการทำให้เรียบ (E_t) และประมาณ S_t ด้วยค่าวัดอิทธิพลของฤดูกาลหรือค่าดัชนีฤดูกาล (\hat{S}_t) ซึ่งการประมาณค่าดังกล่าวจะมีด้วยกันอยู่ 2 วิธีดังนี้

2.4.3.1 การทำให้เรียบแบบเอกซ์โพเนนเชียล จะขึ้นอยู่กับตัวแบบของอนุกรมเวลาว่าเป็นตัวแบบการบวกหรือการคูณ ซึ่งจะมีสูตรการคำนวณที่แตกต่างกันดังนี้

1. เมื่อเป็นตัวแบบการบวก

$$E_t = \begin{cases} \frac{Y_1 + Y_2 + \dots + Y_L}{L} & , t = L \\ g \left(Y_t - \hat{S}_{t-L} \right) + (1-g) E_{t-1} & , t > L \end{cases}$$

ค่า E_t ไม่มีค่าเกิดขึ้นเมื่อ $t < L$

$$\hat{S}_t = \begin{cases} Y_t - E_L & , t \leq L \\ \delta(Y_t - E_t) + (1 - \delta)\hat{S}_{t-L} & , t > L \end{cases}$$

สมการการพยากรณ์ คือ $\hat{Y}_{t+m} = E_t + \hat{S}_{t+m}$

เมื่อ E_t คือ ค่าการทำให้เรียบของข้อมูลซึ่งเป็นค่าประมาณของ β_0

\hat{S}_t คือ ค่าดัชนีฤดูกาลซึ่งเป็นค่าประมาณของ S_t

g คือ ค่าคงที่การทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์ และ $0 \leq g \leq 1$

δ คือ ค่าคงที่การทำให้เรียบระหว่างค่าฤดูกาลจริงกับค่าประมาณฤดูกาล และ $0 \leq \delta \leq 1$

L คือ จำนวนฤดูกาล โดย $L=4$ เมื่อข้อมูลเป็นรายไตรมาส $L=7$ เมื่อข้อมูลเป็นรายวัน และ $L=12$ เมื่อข้อมูลเป็นรายเดือน

\hat{Y}_{t+m} คือ ค่าพยากรณ์ ณ เวลา $t+m$

m คือ จำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า

\hat{S}_{t+m} คือ ค่าดัชนีฤดูกาล ณ เวลา $t+m$ ซึ่ง $\hat{S}_{t+m} = \hat{S}_{t-L+1}$ เมื่อช่วงเวลาที่ $t+m$ ตกอยู่ในฤดูกาลที่ i สำหรับ $i=1,2,3,\dots,L$

2. เมื่อเป็นตัวแบบการคูณ

$$E_t = \begin{cases} \frac{Y_1 + Y_2 + Y_3 + \dots + Y_L}{L} & , t = L \\ \frac{gY_t}{\hat{S}_{t-L}} + (1-g)E_{t-1} & , t > L \end{cases}$$

ค่า E_t ไม่มีค่าเกิดขึ้นเมื่อ $t < L$

$$\hat{S}_t = \begin{cases} \frac{Y_t}{E_L} & , t \leq L \\ \frac{\delta Y_t}{E_t} + (1 - \delta)\hat{S}_{t-L} & , t > L \end{cases}$$

สมการการพยากรณ์ คือ $\hat{Y}_{t+m} = E_t + \hat{S}_{t+m}$ ส่วนความหมายของค่าต่าง ๆ จะเหมือนกับตัวแบบการบวก

2.4.3.2 การทำให้เรียบด้วยค่าความคลาดเคลื่อน

1. เมื่อเป็นตัวแบบการบวก

$$E_t = \begin{cases} \frac{Y_1 + Y_2 + Y_3 + \dots + Y_L}{L} & , t = L, L+1 \\ E_{t-1} + ge_{t-1} & , t > L+1 \end{cases}$$

$$\hat{S}_t = \begin{cases} Y_t - E_L & , t \leq L \\ \hat{S}_{t-L} + \delta(1-g)e_t & , t > L \end{cases}$$

$$\hat{Y}_t = E_t + \hat{S}_{t-L} \quad \text{เมื่อ } t \leq n$$

ค่า \hat{Y}_t ไม่มีค่าเกิดขึ้นเมื่อ $t \leq L$

สมการการพยากรณ์ คือ $\hat{Y}_t = E_{n+1} + \hat{S}_{n+m}$ เมื่อ $t = n+m$, $m = 1, 2, 3, \dots$ และ n คือ จำนวนข้อมูลทั้งหมด

2. เมื่อเป็นตัวแทนการคูณ

$$E_t = \begin{cases} \frac{Y_1 + Y_2 + Y_3 + \dots + Y_L}{L} & , t = L, L+1 \\ E_{t-1} + g e_{t-1} \hat{S}_{t-L} & , t > L+1 \end{cases}$$

ค่า E_t ไม่มีค่าเกิดขึ้นเมื่อ $t < L$

$$\hat{S}_t = \begin{cases} \frac{Y_t}{E_L} & , t \leq L \\ \hat{S}_{t-L} + \frac{\delta(1-g)e_t}{E_t} & , t > L \end{cases}$$

$$\hat{Y}_t = E_t \cdot \hat{S}_{t-L} \quad \text{เมื่อ } t \leq n$$

ค่า \hat{Y}_t ไม่มีค่าเกิดขึ้นเมื่อ $t < L$

สมการการพยากรณ์ คือ $\hat{Y}_t = E_{n+1} \cdot \hat{S}_{n+m}$ เมื่อ $t = n+m$ และ $m = 1, 2, 3, \dots$

2.5 การวิเคราะห์อนุกรมเวลาที่เป็นสเตชันนารีแบบบอซเจนคินส์

2.5.1 แนวความคิดพื้นฐานเกี่ยวกับอนุกรมเวลาแบบบอซเจนคินส์

การพยากรณ์อนุกรมเวลาด้วยวิธีของบอซเจนคินส์ เป็นวิธีการพยากรณ์ที่ยุ่งยากซับซ้อน และต้องใช้ข้อมูลและเวลาในการคำนวณค่อนข้างมาก แต่อย่างไรก็ตามวิธีนี้ก็มิใช่ข้อดีหลายประการ เช่น สามารถใช้ได้กับข้อมูลที่มีการเคลื่อนไหวทุกประเภทและเป็นวิธีการที่มีความแม่นยำของการพยากรณ์ค่อนข้างสูงมากเมื่อเปรียบเทียบกับวิธีการพยากรณ์แบบอื่นๆ สำหรับจำนวนข้อมูลที่ใช้ในการวิเคราะห์ด้วยวิธีของบอซเจนคินส์ ควรมีอย่างน้อย 30 รายการ หรือ 30 ค่าขึ้นไป [7-9]

การพยากรณ์อนุกรมเวลาด้วยวิธีนี้ผู้ที่คิดค้นคือ George E.P. Box และ Gwilym M. Jenkins โดยได้เขียนไว้ในหนังสือชื่อ Time Series Analysis: Forecasting and Control และพิมพ์จำหน่ายครั้งแรกในปี ค.ศ.1970 การกำหนดตัวแบบให้กับอนุกรมเวลาตามวิธีของบอซเจนคินส์ จะกำหนดตัวแบบอยู่ในกลุ่มของตัวแบบ ARMA(p,q) (Auto Regressive and Moving Average order p and q

model) หรือ ARIMA(p,d,q) (Auto Regressive Integrated Moving Average of order p, d and q) ตัวแบบ ARMA(p,q) จะเป็นตัวแบบที่ใช้กับข้อมูลอนุกรมเวลาที่มีลักษณะเป็นสเตชันนารี (stationary) ส่วนตัวแบบ ARIMA(p,d,q) จะเป็นตัวแบบที่ใช้กับข้อมูลอนุกรมเวลาที่มีลักษณะไม่เป็นสเตชันนารี (nonstationary) ดังนั้นตามวิธีของ บอกซ์-เจนกินส์ [10-12] สามารถจะแบ่งข้อมูลอนุกรมเวลาออกเป็น 2 แบบดังนี้

1. อนุกรมเวลาที่เป็นสเตชันนารี (stationary time series) เป็นอนุกรมเวลาที่ค่าสังเกต (Y_t) มีคุณสมบัติทางสถิติคือ ค่าเฉลี่ย ($E(Y_t)$) ค่าความแปรปรวน ($V(Y_t)$) และฟังก์ชันความน่าจะเป็นของค่าสังเกต ณ เวลาต่าง ๆ กันคงที่ กล่าวคือ ไม่เปลี่ยนแปลงตามเวลาที่เปลี่ยนไป เช่น อนุกรมเวลา $Y_1, Y_2, Y_3, \dots, Y_n$ เป็นอนุกรมเวลาที่เป็นสเตชันนารี แสดงว่า ความสัมพันธ์เชิงสถิติระหว่าง $Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_{t+T-1}$ ที่จุดเริ่มต้น t จะไม่ต่างจากความสัมพันธ์เชิงสถิติระหว่าง $Y_{t+j}, Y_{t+j-1}, Y_{t+j-2}, \dots, Y_{t+j+T-1}$ ที่จุดเริ่มต้น $t+j$ สำหรับ $j=1,2,3,\dots, n$ การพิจารณาว่าอนุกรมเวลาชุดใดชุดหนึ่งเป็นสเตชันนารีหรือไม่ พิจารณาได้จาก

ก. ค่าเฉลี่ย ($E(Y_t)$) คงที่สำหรับทุก ๆ ค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็น ส่วน ๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากนัก จะสรุปได้ว่าค่าเฉลี่ยคงที่

ข. ค่าความแปรปรวน ($V(Y_t)$) คงที่สำหรับทุก ๆ ค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็น ส่วน ๆ แล้วหาค่าส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลาแต่ละส่วน ถ้าค่าส่วนเบี่ยงเบนมาตรฐานในแต่ละส่วนย่อยไม่แตกต่างกันมากนัก จะสรุปได้ว่าค่าความแปรปรวนคงที่

ค. การพล็อตกราฟของอนุกรมเวลาแล้วดูการเคลื่อนไหวของอนุกรมเวลา ถ้าการเคลื่อนไหวของอนุกรมเวลามีแนวโน้ม และ/หรือฤดูกาล แสดงว่าอนุกรมเวลาชุดนั้นไม่สเตชันนารี

ง. พิจารณาคอเรโลแกรม (correlogram) ของ r_K ซึ่งเป็นกราฟที่เกิดจากการพล็อตค่าสัมประสิทธิ์ที่สหสัมพันธ์ในตัวเองที่ lag K (coefficient autocorrelation at lag K : r_K) กับค่า K ถ้าอนุกรมเวลาเป็น สเตชันนารีค่า r_K มีลักษณะลดลง (die down) อย่างรวดเร็ว เมื่อ K มีค่าเพิ่มขึ้น แต่ถ้าค่า r_K มีลักษณะลดลงค่อนข้างช้า เมื่อ K มีค่าเพิ่มขึ้น และ/หรือค่า r_K มีค่าค่อนข้างสูงที่ $K = L, 2L, 3L, \dots$ เมื่อ L เป็นจำนวนฤดูกาล แสดงว่าอนุกรมเวลาชุดนั้นไม่สเตชันนารี

สำหรับการกำหนดตัวแบบให้กับอนุกรมเวลาที่เป็นสเตชันนารี จะกำหนดตัวแบบในรูป ARMA(p,q) ซึ่งตัวแบบ ARMA(p,q) จะประกอบด้วยตัวแบบ AR(p) (Auto Regressive order p) และตัวแบบ MA(q) (Moving Average order q)

2. อนุกรมเวลาที่ไม่เป็นสเตชันนารี (nonstationary time series) เป็นอนุกรมเวลาที่ค่าสังเกต (Y_t) มีคุณสมบัติทางสถิติไม่คงที่ คือเปลี่ยนแปลงไปตามเวลาที่เปลี่ยนไป อนุกรมเวลาที่ไม่เป็นสเตชันนารี จะใช้ตัวแบบ ARMA(p,q) ไม่ได้ต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่

ที่มีคุณสมบัติเป็นสเตชันนารีเสียก่อนแล้วจึงจะใช้ตัวแบบ ARMA(p,q) ได้ การแปลงอนุกรมเวลาที่ไม่เป็นสเตชันนารีให้เป็นอนุกรมเวลาที่เป็นสเตชันนารีสามารถกระทำด้วยวิธีการดังต่อไปนี้

ก. การหาผลต่าง (regular differencing) ถ้าอนุกรมเวลาเดิม คือ Y_t เป็นอนุกรมเวลาที่มีแนวโน้ม เราแปลงให้เป็นอนุกรมเวลาที่ไม่มีความโน้ม คือ Z_t โดยที่ $Z_t = \nabla^d Y_t$ เมื่อ d เป็นลำดับของการหาผลต่าง เช่น

$$\begin{aligned} d = 1, Z_t &= \nabla^1 Y_t = Y_t - Y_{t-1} \\ d = 2, Z_t &= \nabla^2 Y_t = \nabla(Y_t - Y_{t-1}) \\ &= \nabla Y_t - \nabla Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \end{aligned}$$

ข. การหาผลต่างฤดูกาล (seasonal differencing) ถ้าอนุกรมเวลาเดิม คือ Y_t เป็นอนุกรมเวลาที่มีความแปรผันตามฤดูกาลเข้ามาเกี่ยวข้อง สามารถแปลงให้เป็นอนุกรมเวลาที่ไม่มีความแปรผันตามฤดูกาลคือ Z_t โดยที่ $Z_t = \nabla_L^D Y_t$ เมื่อ D เป็นลำดับของการหาผลต่างฤดูกาล และ L เป็นจำนวนฤดูกาลต่อปี เช่น ถ้าอนุกรมเวลาเดิมเป็นอนุกรมเวลารายไตรมาส ($L = 4$)

$$\begin{aligned} \text{เช่น } D = 1, Z_t &= \nabla_4 Y_t = Y_t - Y_{t-4} \\ D = 2, Z_t &= \nabla_4^2 Y_t = \nabla_4(Y_t - Y_{t-4}) \\ &= \nabla_4 Y_t - \nabla_4 Y_{t-4} \\ &= (Y_t - Y_{t-4}) - (Y_{t-4} - Y_{t-8}) \\ &= Y_t - 2Y_{t-4} + Y_{t-8} \end{aligned}$$

การหาผลต่างแบบนี้จะทำกี่ครั้งก็ได้จนกว่าจะได้อนุกรมเวลาที่เป็นสเตชันนารี

ค. การหาผลต่างและผลต่างฤดูกาล ถ้าอนุกรมเวลาเดิม คือ Y_t เป็นอนุกรมเวลาที่มีทั้งแนวโน้มและความแปรผันตามฤดูกาล การแปลงอนุกรมเวลาให้เป็นสเตชันนารีทำได้โดย การหาผลต่างและการหาผลต่างฤดูกาลควบคู่กันไป ค่า d และ D จะมีค่าเป็นเท่าไรนั้นขึ้นอยู่กับว่าอนุกรมเวลาที่แปลงเป็นอนุกรมเวลาสเตชันนารีหรือยัง เช่น

ถ้าอนุกรมเวลาเดิมเป็นอนุกรมเวลาที่มีทั้งแนวโน้มและเป็นอนุกรมเวลารายไตรมาส ($L = 4$) เมื่อ $d = 1$ และ $D = 1$ จะได้

$$\begin{aligned} Z_t &= \nabla^d \nabla_L^D Y_t = \nabla^1 \nabla_4^1 Y_t = \nabla(Y_t - Y_{t-4}) \\ &= \nabla Y_t - \nabla Y_{t-4} \\ &= (Y_t - Y_{t-1}) - (Y_{t-4} - Y_{t-5}) \\ &= Y_t - Y_{t-1} + Y_{t-4} - Y_{t-5} \end{aligned}$$

ถ้าอนุกรมเวลาเดิมเป็นอนุกรมเวลาที่มีแนวโน้มและความแปรผันตามฤดูกาลแบบรายเดือน ($L = 12$) เมื่อ $d = 1$ และ $D = 1$ จะได้

$$\begin{aligned}
Z_t &= \nabla^d \nabla_L^D Y_t = \nabla^1 \nabla_{12}^1 Y_t = \nabla(Y_t - Y_{t-12}) \\
&= \nabla Y_t - \nabla Y_{t-12} \\
&= (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) \\
&= Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}
\end{aligned}$$

ง. ถ้าอนุกรมเวลาเดิม (Y_t) มีความแปรปรวน ($V(Y_t)$) ไม่คงที่ จะแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาที่เป็นสเตชันนารีโดยใช้ลอการิทึมหรือรากที่สองหรือยกกำลังสอง เช่น $Z_t = \log(Y_t)$ หรือ $Z_t = \sqrt{Y_t}$ เป็นต้น

เมื่อมีการแปลงอนุกรมเวลาเดิมที่ไม่เป็นสเตชันนารีให้เป็นอนุกรมเวลาใหม่ที่เป็นสเตชันนารี ตัวแบบสำหรับอนุกรมเวลาใหม่คือ ARMA(p,q) ส่วนตัวแบบสำหรับอนุกรมเวลาเดิมคือ ARIMA(p,d,q) ตัวอย่างเช่น จากอนุกรมเวลาเดิม คือ Y_t แปลงเป็นอนุกรมเวลาใหม่คือ Z_t ซึ่ง $Z_t = \nabla^d Y_t$ และเมื่ออนุกรมเวลา Z_t เป็นสเตชันนารีแล้ว จะหาตัวแบบ ARMA(p,q) ให้กับอนุกรมเวลา Z_t ได้เป็น $Z_t \sim \text{ARMA}(p,q)$ ส่วนตัวแบบ ARIMA(p,d,q) ให้กับอนุกรมเวลา Y_t ได้เป็น $Y_t \sim \text{ARIMA}(p,d,q)$ สำหรับตัวแบบ ARIMA นี้จะมี p เป็นอันดับของ AR q เป็นอันดับของ MA และ d เป็นจำนวนครั้งที่หาผลต่างเพื่อให้อนุกรมเวลา Z_t เป็นสเตชันนารี ตัวอย่างของตัวแบบ ARIMA(p,d,q) ของอนุกรมเวลา Y_t เช่น

เมื่อ $d=0$, $\text{ARIMA}(p,0,q) = \text{ARMA}(p,q)$

เมื่อ $p=0$, $\text{ARIMA}(0,d,q) = \text{IMA}(d,q)$

เมื่อ $q=0$, $\text{ARIMA}(p,d,0) = \text{ARI}(p,d)$

2.5.2 ตัวแบบ ARMA(p,q)

ตัวแบบ ARMA(p,q) มี p เป็นอันดับของ AR และ q เป็นอันดับของ MA จำนวนพารามิเตอร์ในตัวแบบจะเท่ากับ $p+q+1$ ตัวแบบที่กำหนดให้กับอนุกรมเวลามักจะเป็นตัวแบบที่มีจำนวนพารามิเตอร์น้อย สำหรับวิธีของบอกซ์เจนคินส์ได้ตัวแบบ ARMA(p,q) เมื่ออนุกรมเวลามีคุณสมบัติเป็นสเตชันนารี มีดังนี้

1. ตัวแบบการถดถอยในตัวเองอันดับ p (autoregressive model of order p : AR(p)) มีตัวแบบทั่วไปดังนี้

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

โดยที่ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

δ คือ ค่าคงที่

ϕ_i คือ ค่าพารามิเตอร์ของการถดถอยในตัวเองที่ $i, i = 1, 2, \dots, p$

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t ซึ่ง ε_t มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ σ_ε^2 และเป็นอิสระต่อกัน ($\varepsilon_t \sim IN(0, \sigma_\varepsilon^2)$)

เนื่องจาก

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} = \delta + \varepsilon_t$$

สามารถจะเขียนในรูปแบบของ backward operator ได้เป็น

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \delta + \varepsilon_t$$

ตัว B เรียกว่า ตัวถอยหลัง (backward shift หรือ backshift) โดยที่ $B^p Y_t = Y_{t-p}$

2. ตัวแบบค่าเฉลี่ยเคลื่อนที่อันดับ q (moving average model of order q : MA(q)) มีตัวแบบทั่วไปดังนี้

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

โดยที่ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

μ คือ ค่าคงที่

θ_i คือ ค่าพารามิเตอร์ของค่าเฉลี่ยเคลื่อนที่ที่ i , $i = 1, 2, \dots, q$

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t ซึ่ง ε_t มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0

ความแปรปรวนเท่ากับ σ_ε^2 และเป็นอิสระต่อกัน ($\varepsilon_t \sim IN(0, \sigma_\varepsilon^2)$)

จากตัวแบบค่าเฉลี่ยเคลื่อนที่ที่สามารถจะเขียนในรูปแบบของ backward operator ได้เป็น

$$Y_t = \mu + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

เมื่อ $B^q \varepsilon_t = \varepsilon_{t-q}$

3. ตัวแบบผสมการถดถอยในตัวเองและค่าเฉลี่ยเคลื่อนที่อันดับที่ p และ q (mixed autoregressive and moving average model of order p and q : ARMA(p, q)) มีตัวแบบทั่วไปดังนี้

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

ตัวแบบลักษณะนี้จะมีเทอมของ AR อยู่ p เทอม และเทอมของ MA อยู่ q เทอม

2.5.3 การกำหนดตัวแบบอนุกรมเวลา

ในการกำหนดตัวแบบ ARMA(p, q) ที่เหมาะสมให้กับข้อมูลอนุกรมเวลาพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนที่ได้จากข้อมูล (r_K และ r_{KK}) ว่าสอดคล้องกับค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง และสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน (ρ_K และ ρ_{KK}) ของตัวแบบใด ซึ่งค่า ρ_K และ ρ_{KK} ของแต่ละตัวแบบ ARMA(p, q) จะมีลักษณะที่แตกต่างกันออกไป ส่วนค่า r_K และ r_{KK} จะเป็นค่าประมาณของ ρ_K และ ρ_{KK} ตามลำดับ

ตารางที่ 1 ลักษณะของ ρ_K และ ρ_{KK} สำหรับตัวแบบ ARMA(p,q)

ตัวแบบ	ρ_K	ρ_{KK}
ARMA(0,0)	$\rho_K = 0$ สำหรับ $K = 1, 2, \dots$	$\rho_{KK} = 0$ สำหรับ $K = 1, 2, \dots$
AR(1)	ρ_K มีค่าเข้าใกล้ 0 เมื่อ K มีค่ามากขึ้น	$\rho_{KK} \begin{cases} \neq 0, & K = 1 \\ = 0, & K \geq 2 \end{cases}$
AR(2)	ρ_K มีค่าเข้าใกล้ 0 เมื่อ K มีค่ามากขึ้น	$\rho_{KK} \begin{cases} \neq 0, & K = 1, 2 \\ = 0, & K \geq 3 \end{cases}$
MA(1)	$\rho_K \begin{cases} \neq 0, & K = 1 \\ = 0, & K \geq 2 \end{cases}$	ρ_{KK} มีค่าเข้าใกล้ 0 เมื่อ K มีค่ามากขึ้น
MA(2)	$\rho_K \begin{cases} \neq 0, & K = 1, 2 \\ = 0, & K \geq 3 \end{cases}$	ρ_{KK} มีค่าเข้าใกล้ 0 เมื่อ K มีค่ามากขึ้น
ARMA(1,1)	ρ_K มีค่าเข้าใกล้ 0 เมื่อ K มีค่ามากขึ้น	ρ_{KK} มีค่าเข้าใกล้ 0 เมื่อ K มีค่ามากขึ้น

ในปี 1978 Schwarz [13] ได้คิดค้น ค่า BAYESIAN INFORMATION CRITERION (BIC) ซึ่งเป็นเครื่องมือวัดในทางสถิติ ที่สามารถใช้ในการกำหนดตัวแบบอนุกรมเวลาได้ และในปี 1998 Fraley และ Raftery [14] ได้กล่าวว่า ค่าที่ต่ำที่สุดของ BIC จะใช้ในการเลือกตัวแบบที่ดีที่สุดพอดี โดยสามารถคำนวณค่า BIC ได้จากสมการต่อไปนี้

$$BIC = \ln(\sigma_e^2) + \frac{v}{n} \ln(n)$$

โดยที่ σ_e^2 คือความแปรปรวนของค่าความคลาดเคลื่อน

n คือจำนวนข้อมูล

v คือจำนวนพารามิเตอร์ในตัวแบบ

2.5.4 การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ เป็นขั้นตอนหลังจากการกำหนดตัวแบบ ARMA(p,q) ให้กับอนุกรมเวลาแล้ว การประมาณค่าพารามิเตอร์ตามวิธีของบอกซ์-เจนคินส์ จะมีอยู่ 2 แบบด้วยกันคือ การประมาณค่าเบื้องต้นและการประมาณด้วยวิธีกำลังสองน้อยที่สุด

1. การประมาณค่าเบื้องต้น

การประมาณค่าเบื้องต้นจะใช้สมการที่แสดงความสัมพันธ์ระหว่าง ρ_K และพารามิเตอร์ในตัวแบบ จำนวนสมการที่ใช้จะเท่ากับอันดับที่ปรากฏในตัวแบบ โดยจะประมาณค่า ρ_K ด้วย r_K ในสมการความสัมพันธ์ดังกล่าว แล้วแก้สมการหาค่าประมาณของพารามิเตอร์ (ค่า ϕ_i หรือ θ_i) ส่วนค่าประมาณของ δ หรือ μ จะได้จากสมการที่แสดงความสัมพันธ์ระหว่างค่าเฉลี่ยของอนุกรมเวลา

($E(Y) = \mu_Y$) กับพารามิเตอร์ในตัวแบบ ตัวอย่างเช่น ตัวแบบ AR(1) จะหา $\hat{\phi}_1$ ได้จากการแก้สมการ $\rho_1 = \phi_1$ และ $\hat{\delta}$ ได้จากการแก้สมการ $\mu_Y = \frac{\delta}{1 - \phi_1}$ เมื่อแทน ρ_1 และ μ_Y ด้วย r_1 และ \bar{Y} ใน

สมการความสัมพันธ์ทั้งสองตามลำดับ จะได้ค่าประมาณ $\hat{\phi}_1$ และ $\hat{\delta}$ หรือถ้าเป็นตัวแบบ MA(1) จะ

หาค่า $\hat{\theta}_1$ ได้จากการแก้สมการ $\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$ และ $\hat{\mu}$ ได้จาก $\mu_Y = \mu$ เมื่อแทน ρ_1 และ μ_Y ด้วย

r_1 และ \bar{Y} ในสมการแสดงความสัมพันธ์ทั้งสองตามลำดับ ซึ่งก็จะได้ค่าประมาณ $\hat{\theta}_1$ และ $\hat{\mu}$ ตามต้องการ

2. การประมาณด้วยวิธีกำลังสองน้อยที่สุด

การประมาณด้วยวิธีกำลังสองน้อยที่สุด จะเป็นการประมาณค่าพารามิเตอร์ในตัวแบบที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (SSE) มีค่าต่ำสุด วิธีการนี้จะทำได้โดยการแก้สมการปกติ เพราะว่าการคำนวณมีความยุ่งยากและเสียเวลามาก จึงต้องอาศัยเทคนิคการวิเคราะห์เชิงตัวเลข และคอมพิวเตอร์ช่วยในการคำนวณโดยมีการกำหนดค่าเริ่มต้น (initial value) ก่อน แล้วใช้

คอมพิวเตอร์ (ในที่นี้ใช้โปรแกรม SPSS for Windows และ Minitab) ทำการคำนวณซ้ำๆ กัน (iteration) หลายรอบจนกว่าค่าประมาณที่ได้และค่า SEE เริ่มมีค่าคงที่

2.5.5 การตรวจสอบความเหมาะสมของตัวแบบ

เมื่อมีการกำหนดตัวแบบและประมาณค่าพารามิเตอร์แล้ว ต้องมีการตรวจสอบว่าตัวแบบที่เลือกไว้มีความเหมาะสมกับอนุกรมเวลาหรือไม่ การตรวจสอบมีหลายวิธีและวิธีการตรวจสอบส่วนใหญ่จะอาศัยค่าความคลาดเคลื่อนของการพยากรณ์ (e_t) ซึ่งค่าดังกล่าวเป็นค่าประมาณของ ε_t ในตัวแบบ โดยที่ $e_t = Y_t - \hat{Y}_t$ สำหรับวิธีการตรวจสอบที่จะกล่าวถึงในที่นี้มีทั้งหมด 2 วิธีดังนี้

1. พิจารณา $\rho_K(e_t) = 0$ สำหรับ $K = 1, 2, 3, \dots, m$ หรือไม่ นั่นคือเป็นการพิจารณาว่าค่าความคลาดเคลื่อนของการพยากรณ์ที่อยู่ห่างกัน $1, 2, 3, \dots, m$ ช่วงเวลาเป็นอิสระกันหรือไม่

การทดสอบดังกล่าวสามารถจะกำหนดสมมติฐานหลักและสมมติฐานทางเลือกได้ดังนี้

$$H_0 : \rho_1(e_t) = \rho_2(e_t) = \dots = \rho_m(e_t) = 0 \quad \text{สำหรับ } K = 1, 2, 3, \dots, m$$

$$H_1 : \rho_K(e_t) \text{ อย่างน้อยหนึ่งค่าที่ไม่เท่ากับ } 0 \quad \text{สำหรับ } K = 1, 2, 3, \dots, m$$

สำหรับตัวสถิติที่ใช้ในการทดสอบคือ ตัวสถิติของ Box และ Pierce หรือ ตัวสถิติของ Box และ Liung โดยที่ตัวสถิติของ Box และ Pierce คือ

$$Q = n \sum_K^m r_K^2(e_t)$$

ส่วนตัวสถิติของ Box และ Liung คือ

$$Q_m = n(n+2) \sum_{K=1}^m \left\{ \frac{r_K^2(e_t)}{n-K} \right\}$$

เมื่อ n คือจำนวนอนุกรมเวลาของ e_t ทั้งหมด

m คือช่วงเวลาที่ห่างกันมากที่สุดของ e_t ที่นำมาพิจารณา

จะปฏิเสธ H_0 ก็ต่อเมื่อตัวสถิติ Q หรือ $Q_m \geq \chi_{\lambda, m-a}^2$ เมื่อ a คือจำนวนพารามิเตอร์ในตัวแบบทั้งหมดรวมค่าคงที่ด้วย (μ, δ) แสดงว่ามี $\rho_K(e_t)$ อย่างน้อยหนึ่งค่าที่ไม่เท่ากับ 0 สำหรับ $K = 1, 2, 3, \dots, m$ หรือค่าความคลาดเคลื่อนไม่เป็นอิสระต่อกัน หรือตัวแบบที่เรากำหนดให้กับอนุกรมเวลายังไม่เหมาะสม

2. การทดสอบว่าแต่ละพารามิเตอร์ในตัวแบบมีค่าเท่ากับ 0 หรือไม่

โดยกำหนดสมมติฐานหลักและสมมติฐานทางเลือกดังนี้

$$H_0 : \theta = 0$$

$$H_1 : \theta \neq 0$$

เมื่อ θ เป็นพารามิเตอร์ สำหรับตัวสถิติที่ใช้ทดสอบคือ T โดยที่

$$T = \frac{\hat{\theta}}{S_{\hat{\theta}}}$$

เมื่อ $\hat{\theta}$ คือค่าประมาณของพารามิเตอร์ θ

$S_{\hat{\theta}}$ คือค่าความคลาดเคลื่อนมาตรฐานของ $\hat{\theta}$

จะยอมรับ H_0 เมื่อ $-T_{\frac{\lambda}{2}, d.f.} \leq T \leq T_{\frac{\lambda}{2}, d.f.}$ และจะปฏิเสธ H_0 เมื่อ $T \leq -T_{\frac{\lambda}{2}, d.f.}$ หรือ

$T > T_{\frac{\lambda}{2}, d.f.}$ โดยระดับนัยสำคัญเท่ากับ λ และ $d.f. = n - 1$

2.5.6 การพยากรณ์

เมื่อได้มีการตรวจสอบแล้วว่าตัวแบบที่กำหนดให้กับอนุกรมเวลาเหมาะสมแล้ว ขั้นตอนต่อไปจะมีการนำเอาสมการที่สร้างขึ้นจากตัวแบบไปใช้ในการพยากรณ์ การพยากรณ์ส่วนใหญ่จะเป็นการพยากรณ์แบบล่วงหน้า และการพยากรณ์แบบล่วงหน้าจะมี 2 แบบคือการพยากรณ์แบบจุด และการพยากรณ์แบบช่วง สำหรับสมการพยากรณ์แบบล่วงหน้า m ช่วงเวลาที่ได้จากตัวแบบ ARMA(p,q)

ต่อไปจะนำเสนอการวิเคราะห์ห้อนุกรมเวลาที่ไม่เป็นสเตชันนารี ซึ่งอนุกรมเวลาที่ไม่เป็นสเตชันนารีสามารถทำให้เป็นอนุกรมเวลาที่เป็นสเตชันนารีได้ ดังที่กล่าวไว้ข้างต้น

2.6 การวิเคราะห์ห้อนุกรมเวลาที่ไม่เป็นสเตชันนารีแบบบอกซ์-เจนคินส์

สำหรับตัวแบบที่ใช้กับอนุกรมเวลาที่ไม่เป็นสเตชันนารี มีดังนี้

2.6.1 ตัวแบบสำหรับอนุกรมเวลาที่มีแนวโน้ม

สำหรับตัวแบบที่ใช้กับอนุกรมเวลาที่มีแนวโน้มคือ ตัวแบบ ARIMA(p,d,q) (Auto Regressive Integrated Moving Average of order p, d and q) ตัวแบบนี้จะมี p เป็นอันดับของ AR q เป็นอันดับของ MA และ d เป็นจำนวนครั้งที่หาผลต่างเพื่อให้อนุกรมเวลา z_t เป็นสเตชันนารี ตัวแบบ ARIMA(p,d,q) ถ้าไม่มีเทอมของตัวแบบการถดถอยในตัวเอง (AR) จะเรียกว่า IMA(d,q) (Integrated Moving Average of order d and q) และถ้าไม่มีเทอมของตัวแบบค่าเฉลี่ยเคลื่อนที่ (MA) จะเรียกว่า ARI(p,d) (Auto Regressive Integrated of order p and d) ดังนั้นตัวแบบ ARIMA(p,d,q) สามารถจะเขียนในรูปของ backward operator ได้ดังนี้

$$(1-B)^d (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p) y_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (1)$$

เมื่อ B เรียกว่า backshift โดยที่ $B^p y_t = y_{t-p}$ และ $B^q \varepsilon_t = \varepsilon_{t-q}$

θ_0 คือค่าคงที่

ϕ_i คือค่าพารามิเตอร์ของ AR ตัวที่ i สำหรับ $i = 1, 2, 3, \dots, p$

θ_j คือค่าพารามิเตอร์ของ MA ตัวที่ j สำหรับ $j = 1, 2, 3, \dots, q$

2.6.2 ตัวแบบสำหรับอนุกรมเวลาที่มีความแปรผันตามฤดูกาล

อนุกรมเวลาหลายชุดหรือหลายประเภท เช่น อนุกรมเวลาเกี่ยวกับราคาผลผลิตทางการเกษตร จำนวนนักท่องเที่ยว ปริมาณน้ำฝน อนุกรมเวลาเหล่านี้จะพบว่ามีลักษณะซ้ำกันเป็นช่วง (periodic) ซึ่งจะซ้ำกันหรือคล้ายกันทุก L หน่วยเวลา ลักษณะที่คล้ายกันเป็นช่วงเป็นการเปลี่ยนแปลงตามฤดูกาล ดังนั้น ถ้าให้ L จำนวนฤดูกาลต่อปี อนุกรมเวลา y_t จะคล้ายกับ $y_{t \pm iL}; i = 1, 2, 3, \dots$ และถ้าข้อมูลเป็นรายเดือน $L = 12$ หรือ ถ้าข้อมูลเป็นรายไตรมาส $L = 4$ เป็นต้น

การสร้างตัวแบบถดถอยในตัวเองรวมการเฉลี่ยเคลื่อนที่ (ARIMA) ของอนุกรมเวลาที่มีความแปรผันตามฤดูกาล จะสร้างได้เหมือนกับตัวแบบการถดถอยในตัวเองรวมการเฉลี่ยเคลื่อนที่ของอนุกรมเวลาที่ไม่มีความแปรผันตามฤดูกาล คือ การกำหนดตัวแบบ การประมาณค่า การตรวจสอบตัวแบบ และการพยากรณ์ แต่สำหรับ ข้อมูลที่มีความแปรผันตามฤดูกาลจะหาผลต่างของค่าสังเกตด้วยช่วงห่าง L หน่วยเวลา ดังนั้นการหาสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนจะห่างกัน L หน่วยเวลา

2.6.2.1 ตัวแบบการถดถอยในตัวเองที่มีฤดูกาล (Seasonal Auto Regressive of order P: SAR(P)_L) สำหรับข้อมูลที่มีฤดูกาลคงที่ (เป็นสเตชันนารี) และมีตัวแบบการถดถอยในตัวเอง จะมีตัวแบบที่เขียนในรูปของ backward operator ได้ดังนี้

$$(1 - \phi_1 B^L - \phi_2 B^{2L} - \dots - \phi_P B^{PL}) y_t = \theta_0 + \varepsilon_t$$

2.6.2.2 ตัวแบบเฉลี่ยเคลื่อนที่ที่มีฤดูกาล (Seasonal Moving Average of order Q: SMA(Q)_L)

สำหรับข้อมูลที่มีฤดูกาลคงที่ (เป็นสเตชันนารี) และมีตัวแบบเฉลี่ยเคลื่อนที่ สามารถเขียนในรูปของ backward operator ได้ดังนี้

$$y_t = \theta_0 + (1 - \theta_1 B^L - \theta_2 B^{2L} - \dots - \theta_Q B^{QL}) \varepsilon_t$$

2.6.2.3 ตัวแบบการถดถอยในตัวเองและการเฉลี่ยเคลื่อนที่ที่มีฤดูกาล (Seasonal Mixed and Moving Average of order P, Q: SARMA(P,Q)_L)

สำหรับข้อมูลที่มีฤดูกาลคงที่ (เป็นสเตชันนารี) และมีตัวแบบการถดถอยในตัวเองและการเฉลี่ยเคลื่อนที่ จะเขียนในรูปของ backward operator ได้ดังนี้

$$(1 - \phi_1 B^L - \phi_2 B^{2L} - \dots - \phi_P B^{PL}) y_t = \theta_0 + (1 - \theta_1 B^L - \theta_2 B^{2L} - \dots - \theta_Q B^{QL}) \varepsilon_t$$

2.6.2.4 ตัวแบบการถดถอยในตัวเองรวมการเฉลี่ยเคลื่อนที่ที่มีฤดูกาล (Seasonal Auto Regressive Integrated Moving Average of order P, D, Q: SARIMA(P,D,Q)_L)

ตัวแบบ SARIMA(P,D,Q)_L จะเป็นตัวแบบที่ใช้กับข้อมูลที่มีฤดูกาลไม่คงที่ (Nonstationary) โดยจะต้องมีการหาผลต่างของข้อมูลด้วย ช่วงห่างกัน L หน่วยเวลา เพื่อให้ข้อมูลเป็นสเตชันนารี ตัว

แบบ SARIMA (P,D,Q)_L เป็นตัวแบบที่มี P เป็นอันดับของ SAR (Seasonal Auto Regressive) Q เป็นอันดับของ AMA (Seasonal Moving Average) และ D เป็นจำนวนครั้งที่หาผลต่างฤดูกาล ดังที่กล่าวไว้ในหัวข้อ 2.5.1 (2. ข.) เพื่อให้ได้อนุกรมเวลา (y_t) ที่ไม่เป็นสเตชันนารี เนื่องจากฤดูกาลเป็นอนุกรมชุดใหม่ (z_t) ที่เป็น สเตชันนารี ส่วนค่า L เป็นจำนวนฤดูกาลต่อปีหรือเป็นความยาวของฤดูกาล ตัวแบบ SARIMA(P,D,Q)_L ถ้าไม่มีเทอมของ AR (P=0) จะเรียกว่า SIMA(D,Q) และถ้าไม่มีเทอมของ MA (Q=0) จะเรียกว่า SARI(P,D) แต่ถ้า D=0 (มีฤดูกาลแต่เป็นสเตชันนารี) จะเรียกว่า SARMA(P,Q)_L และถ้า P=0, D=0 สามารถเขียนในรูปของ backward operator ได้ดังนี้

$$(1 - B^L)^D (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{PL} B^{PL}) y_t = \theta_0 + (1 - \theta_1 B^L - \theta_2 B^{2L} - \dots - \theta_{QL} B^{QL}) \varepsilon_t$$

2.6.3 ตัวแบบสำหรับอนุกรมเวลาที่มีแนวโน้มและความแปรผันตามฤดูกาล

เมื่อข้อมูลที่เก็บรวบรวมมาได้เป็นอนุกรมเวลาที่มีแนวโน้มรวมทั้งมีความแปรผันตามฤดูกาลผสมกันอยู่ ตัวแบบที่นิยมใช้กับอนุกรมเวลาดังกล่าวจะเป็นตัวแบบในลักษณะตัวแบบการคูณของอนุกรมเวลาที่มีฤดูกาลและไม่มีฤดูกาล หลักการสร้างตัวแบบดังกล่าวจะประกอบด้วยตัวแบบ 2 ส่วนคือ ตัวแบบ ARIMA(p,d,q) เป็นตัวแบบที่ใช้กับอนุกรมเวลาที่มีแนวโน้ม และ SARIMA(P,D,Q) เป็นตัวแบบที่ใช้กับอนุกรมเวลาที่มีความแปรผันตามฤดูกาล ตัวแบบนี้สามารถจะเขียนในรูปของ backward operator ได้ดังนี้

$$(1 - B)^d (1 - B^L)^P (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (1 - \phi_L B^L - \phi_{2L} B^{2L} - \dots - \phi_{PL} B^{PL}) y_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) (1 - \theta_L B^L - \theta_{2L} B^{2L} - \dots - \theta_{QL} B^{QL}) \varepsilon_t \quad (2)$$

สัญลักษณ์ที่ใช้กับตัวแบบนี้คือ ARIMA(p,d,q)(P,D,Q)_L หรือ SARIMA(p,d,q) (P,D,Q)_L ในการทำงานวิจัยนี้จะใช้ ARIMA(p,d,q) (P,D,Q)_L

2.6.4 ขั้นตอนของบ็อกซ์-เจนกินส์สำหรับอนุกรมเวลาที่ไม่เป็นสเตชันนารี

การใช้วิธีบ็อกซ์-เจนกินส์สำหรับอนุกรมเวลาที่ไม่เป็นสเตชันนารีไม่แตกต่างจากการใช้กับอนุกรมเวลาที่เป็นสเตชันนารี จะเพิ่มเพียงขั้นตอนของการแปลงอนุกรมเวลาเดิมที่ไม่เป็นสเตชันนารีให้เป็นอนุกรมเวลาใหม่ที่เป็นสเตชันนารี กล่าวคือ ถ้าอนุกรมเวลามีแนวโน้มให้หาผลต่างของอนุกรมเวลาจนได้อนุกรมเวลาใหม่ที่เป็นสเตชันนารี ถ้าอนุกรมเวลามีอิทธิพลของฤดูกาล ให้หาผลต่างฤดูกาลของอนุกรมเวลาจนได้อนุกรมเวลาใหม่ที่เป็นสเตชันนารี ถ้าอนุกรมเวลามีแนวโน้มและอิทธิพลของฤดูกาลให้หาทั้งผลต่างและผลต่างของฤดูกาลของอนุกรมเวลาจนได้อนุกรมเวลาใหม่ที่เป็นสเตชันนารี และถ้าอนุกรมเวลามีค่าความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิมเป็นอนุกรมเวลาใหม่ โดยอาจจะใช้ลอการิทึม ($z_t = \ln y_t$) หรือลอการิทึม ($z_t = \sqrt{y_t}$) ขึ้นอยู่กับชุดข้อมูล จนกว่าอนุกรมเวลาใหม่จะมีค่าความแปรปรวนคงที่ เมื่ออนุกรมเวลาที่แปลงแล้วเป็นสเตชันนารี จะทำตามขั้นตอนของบ็อกซ์-เจนกินส์ดังนี้

1. กำหนดตัวแบบ ARIMA(p,d,q) หรือ SARIMA(P,D,Q)_L หรือ ARIMA(p,d,q)(P,D,Q)_L ให้กับอนุกรมเวลาเดิม (y_t) โดยพิจารณาจากคอโรลแกรมของ r_K และ r_{KK} ของอนุกรมเวลาใหม่ (z_t) ถ้าตัวแบบ ARMA(p,q) ให้พิจารณาค่า r_K และ r_{KK} เมื่อ $K=1,2,3, \dots$ ถ้าตัวแบบ SARMA(P,Q)_L ให้พิจารณาค่า r_K และ r_{KK} เมื่อ $K=L, 2L, 3L, \dots$ นอกจากนี้การกำหนดตัวแบบให้กับอนุกรมเวลาเดิม (y_t) อาจจะหาค่า d หรือ D ก่อนเพื่อให้อนุกรมเวลาเป็นสเตชันนารีแล้วทดลองหาตัวแบบหลาย ๆ ตัวแบบแล้วเลือกตัวแบบตามหลักการซึ่งเป็นหลักมาตรฐานคือ

ก. เลือกตัวแบบที่มีค่า SBC และ SSE น้อยที่สุด

ข. เลือกตัวแบบที่มีจำนวนพารามิเตอร์น้อยที่สุด

ค. จะไม่เลือกตัวแบบที่มีการยอมรับสมมติฐานที่ว่าค่าพารามิเตอร์มีค่าเท่ากับ 0 ($\phi_1, \phi_2, \dots, \phi_p, \phi_L, \phi_{2L}, \dots, \phi_{pL}, \theta_1, \theta_2, \dots, \theta_q$ และ $\theta_L, \theta_{2L}, \dots, \theta_{qL}$)

2. ประมาณค่าพารามิเตอร์ ในที่นี้จะใช้วิธีกำลังสองน้อยที่สุด โดยใช้โปรแกรม SPSS for Windows และ Minitab ช่วยคำนวณหา

3. ตรวจสอบความเหมาะสมของตัวแบบ เมื่อมีการกำหนดตัวแบบและการประมาณค่าพารามิเตอร์แล้วต้องมีการตรวจสอบว่าตัวแบบที่เลือกไว้มีความเหมาะสมกับอนุกรมเวลาหรือไม่

4. พยากรณ์ ซึ่งการพยากรณ์จะมีทั้งการพยากรณ์แบบจุดและการพยากรณ์แบบช่วง โดยส่วนใหญ่จะใช้โปรแกรมในการคำนวณหา

2.7 การทดสอบการแจกแจงแบบปกติ

การทดสอบว่าข้อมูลมีการแจกแจงแบบปกติ หรือแจกแจงแบบไม่ปกตินั้น สามารถทดสอบสถิติด้วย Kolmogorov-Smirnov Test เมื่อใช้ข้อมูลมากกว่า 50 กรณี หรือ Shapiro-Wilk test เมื่อใช้ข้อมูลน้อยกว่า 50 กรณี ซึ่งในงานวิจัยนี้เราใช้การทดสอบสถิติ Shapiro-Wilk test โดยกำหนดสมมติฐานหลักและสมมติฐานทางเลือกดังนี้ [15-17]

H_0 : ชุดข้อมูลมีการแจกแจงแบบปกติ

H_1 : ชุดข้อมูลไม่มีการแจกแจงแบบปกติ

คำนวณหาตัวสถิติทดสอบโดยเรียงลำดับค่าสังเกต n ค่าจากน้อยไปมาก

$$Y_1 < Y_2 < \dots < Y_n$$

ตัวสถิติทดสอบ คือ

$$W = \frac{\left[\sum_{i=1}^k a_{n-i+1} (Y_{n-i+1} - Y_i) \right]^2}{\sum_{i=1}^n (Y_i - \bar{Y})}$$

โดยที่ a_{n-i+1} คือ สัมประสิทธิ์จากการเปิดตารางของ Shapiro-Wilk test ในการทดสอบการแจกแจงปกติ

Y_i คือ ข้อมูลหรือค่าสังเกต ณ เวลา i สำหรับ $i = 1, 2, \dots, k$

$$\text{และ } k = \begin{cases} \frac{n}{2} & , n \text{ เป็นจำนวนคู่} \\ \frac{n-1}{2} & , n \text{ เป็นจำนวนคี่} \end{cases}$$

จะยอมรับ H_0 เมื่อ $p\text{-value} = P(W < w)$ มีค่ามากกว่าค่าระดับนัยสำคัญ ($\lambda = 0.05$) โดยค่า $p\text{-value} = P(W < w)$ ได้จากตาราง Percentage point of the W test เมื่อ n มีค่าตั้งแต่ 3 ถึง 50

ในปัจจุบันมีหลายงานวิจัยที่ใช้ตัวแบบ ARIMA สำหรับการวิเคราะห์อนุกรมเวลาเพื่อทำนายค่าในอนาคต เช่น ในปี ค.ศ. 2015 Chen, Guan, Yun, Li และ Recknagel ได้ศึกษาความเข้มข้นของคลอโรฟิลล์เอ ซึ่งมีผลต่อการแพร่พันธุ์ของสาหร่ายเซลล์เดียว โดยสาหร่ายเซลล์เดียวจะจับตัวกันเป็นแพ ส่งผลกระทบต่อการเพาะเลี้ยงสัตว์น้ำ และทำให้ประชาชนไม่มีน้ำดื่ม น้ำใช้ ทางผู้วิจัยจึงสนใจที่จะทำนายความเข้มข้นของคลอโรฟิลล์เอ โดยใช้ตัวแบบ ARIMA และใช้ตัวแบบการถดถอยเชิงเส้นหลายตัวแปร โดยมีอุณหภูมิ น้ำ ความโปร่งใสในน้ำ ความเข้มข้นของไนโตรเจนอนินทรีย์รวมและความเข้มข้นของฟอสเฟต เป็นปัจจัยในการเกิดความเข้มข้นของคลอโรฟิลล์เอ ผลการวิจัยพบว่า ตัวแบบ ARIMA สามารถทำนายความเข้มข้นของคลอโรฟิลล์เอได้ดีกว่าตัวแบบการถดถอยเชิงเส้นหลายตัวแปร โดย RMSE ของตัวแบบ ARIMA และตัวแบบการถดถอยเชิงเส้นหลายตัวแปร คือ 28.98 ug/l and 40.20 ug/l ตามลำดับ [18] ในปีเดียวกัน Arya และ Zhang ได้มีการประยุกต์ใช้ตัวแบบ ARIMA ในการวิเคราะห์อนุกรมเวลา เพื่อทำนายอุณหภูมิและปริมาณออกซิเจนที่ละลายในน้ำ จาก 4 สถานีของแม่น้ำ Stillaguamish ทางตอนเหนือของ Snohomish และทิศตะวันออกเฉียงใต้ของ southeastern Skagit Counties ใน western Washington State ผลการวิจัยพบว่าตัวแบบของอนุกรมเวลาสามารถทำนายค่าในอนาคตได้ และเป็นตัวแบบที่มีความเหมาะสมและถูกต้องสูง [19] และในปีเดียวกัน Mirzavand และ Ghazavi ได้ใช้ตัวแบบ ARIMA ในการวิเคราะห์อนุกรมเวลาเพื่อทำนาย ระดับน้ำตาลในประเทศอิหร่านจำนวน 36 กลุ่มที่ถูกแบ่งออกเป็น 5 กลุ่ม ผลการวิจัยพบว่า สามารถทำนายระดับน้ำตาลแต่ละกลุ่มได้ ซึ่งให้ความถูกต้องที่สูง นั่นคือ RMSE เท่ากับ 0.89, 0.89, 0.95, 0.95 และ 0.75 ตามลำดับ [20] จากงานวิจัยล่าสุดที่กล่าวมาและงานวิจัยอีกเป็นจำนวนมากในอดีต แสดงให้เห็นว่า ARIMA สามารถใช้ในการทำนายค่าพารามิเตอร์ทางน้ำได้เป็น

อย่างดี ดังนั้นงานวิจัยในครั้งนี้ ผู้วิจัยจะศึกษาพฤติกรรมของค่า BOD และประยุกต์ใช้การวิเคราะห์อนุกรมเวลาแบบบ็อกซ์-เจนกินส์ สำหรับการสร้างตัวแบบใหม่เพื่อทำนายค่า BOD ในอนาคต



บทที่ 3

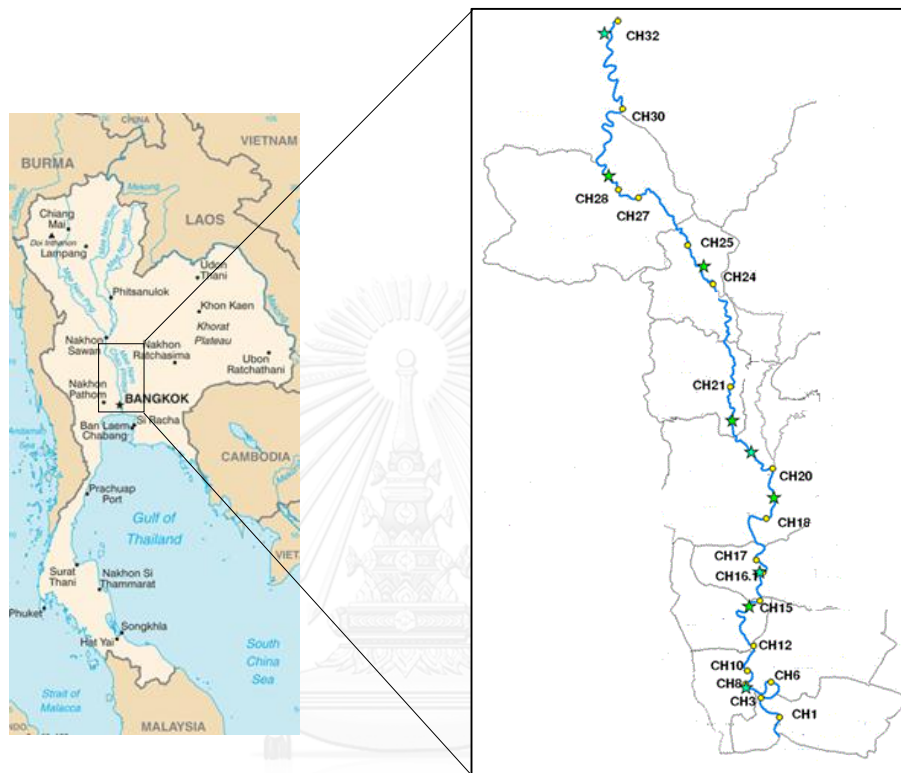
การทำนายค่า BOD สำหรับน้ำในแม่น้ำเจ้าพระยาโดยกระบวนการอัลฟาทริมา

3.1 พื้นที่และการเก็บรวบรวมข้อมูล

ข้อมูล BOD ของแม่น้ำเจ้าพระยาถูกเก็บรวบรวมทั้งหมดจาก 18 สถานีจุดตรวจวัดคุณภาพน้ำ แม่น้ำเจ้าพระยาที่สถานีที่ 1 (CH1) ถึงสถานีที่ 32 (CH32) ดังภาพที่ 1 และตารางที่ 2 โดยเจ้าหน้าที่ กรมควบคุมมลพิษ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อมเป็นผู้ตรวจวัดคุณภาพเพื่อนำมาวิเคราะห์คุณภาพน้ำ ซึ่งข้อมูล BOD ถูกเก็บรวบรวมในช่วงปี พ.ศ. 2539 – 2556 โดยในแต่ละปี มีความถี่ในการเก็บรวบรวมข้อมูลไม่เท่ากัน (ปีละ 3-6 ครั้ง) ทางผู้ทำวิจัยจึงจัดชุดข้อมูลใหม่เพื่อให้ข้อมูล BOD เป็นอนุกรมเวลาโดยแบ่งข้อมูลออกเป็น 2 ชุด คือ

ชุดที่ 1 ข้อมูล BOD ถูกเก็บรวบรวมในเดือนมิถุนายน ($t = 1, 3, 5, \dots, 33$) และธันวาคม ($t = 2, 4, 6, \dots, 32$) จำนวน 33 ค่า

ชุดที่ 2 ข้อมูล BOD ถูกเก็บรวบรวมในเดือนกุมภาพันธ์ ($t = 1, 3, 5, \dots, 33$) และสิงหาคม ($t = 2, 4, 6, \dots, 32, 34$) จำนวน 34 ค่า



ภาพที่ 1 สถานีจุดตรวจวัดคุณภาพน้ำทั้ง 18 สถานี

ที่มา : http://www.pcd.go.th/info_serv/images_water/chaopraya50_34B.gif

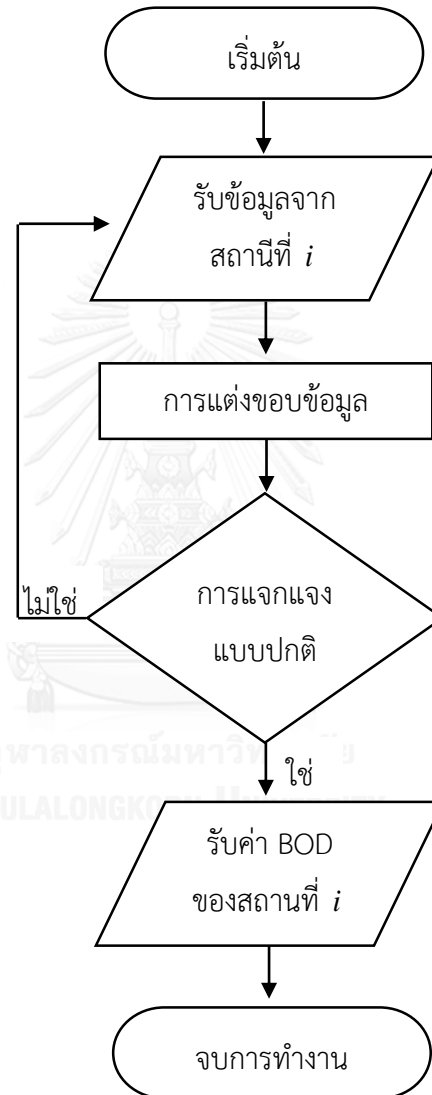
http://renuga158.blogspot.com/2013/07/blog-post_1424.html

ตารางที่ 2 ที่ตั้งของสถานีของจุดตรวจวัดคุณภาพน้ำ แม่น้ำเจ้าพระยา

สถานี	ที่ตั้งสถานีของจุดตรวจวัดคุณภาพน้ำ แม่น้ำเจ้าพระยา	ละติจูด	ลองจิจูด
1	พระสมุทรเจดีย์ อ.เมือง จ.สมุทรปราการ	672495 N	1503718 E
3	หน้าที่ว่าการ อ.พระประแดง จ.สมุทรปราการ	666474 N	1510119 E
6	ท่าเรือกรุงเทพ(สะพานปลา) เขตยานนาวา กทม.	669779 N	1515411 E
8	สะพานกรุงเทพฯ เขตดาวคะนอง กทม.	661571 N	1514712 E
10	สะพานพุทธยอดฟ้า เขตสัมพันธวงศ์ กทม.	662225 N	1519063E
12	สะพานพระรามหก อ.บางกรวย จ.นนทบุรี	664167 N	1527303 E
15	สะพานนนทบุรี อ.ปากเกร็ด จ.นนทบุรี	666174 N	1542211 E
16.1	จุดสูบน้ำดิบเพื่อการประปาสำแล อ.เมือง จ.ปทุมธานี	667292 N	1551717 E
17	บริเวณ อ.สามโคก จ.ปทุมธานี	665076 N	1555811 E
18	โรงงานกระดาษบางปะอิน อ.บางปะอิน จ.พระนครศรีอยุธยา	668291 N	1569511 E
20	ป้อมเพชร ต.สำเภากลุ่ม อ.พระนครศรีอยุธยา จ.พระนครศรีอยุธยา	670298 N	1586207 E
21	สะพานข้ามแม่น้ำเจ้าพระยา อ.เมือง จ.อ่างทอง	656788 N	1613207 E
24	สะพานข้ามแม่น้ำเจ้าพระยา อ.เมือง จ.สิงห์บุรี	651186 N	1647307 E
25	บริเวณใต้ตลาด อ.อินทรบุรี จ.สิงห์บุรี	643068 N	1660098 E
27	เขื่อนเจ้าพระยา อ.เมือง จ.ชัยนาท	627201 N	1675915 E
28	ศาลากลาง อ.เมือง จ.ชัยนาท	620791 N	1678621 E
30	สะพานสมเด็จพระวันรัตน์ ต.ท่าน้ำอ้อย อ.พยุหะคีรี จ.นครสวรรค์	622189 N	1705401 E
31	วัดมณีวงศ์ ต.ย่านมัทรี อ.พยุหะคีรี จ.นครสวรรค์	619792 N	1719201 E
32	สะพานเดชาติวงศ์ อ.เมือง จ.นครสวรรค์	620681 N	1734498 E

3.2 การทำนายค่า BOD โดยใช้กระบวนการอัลฟาทรिमอาร์ริมา

ในงานวิจัยนี้ ผู้ทำวิจัยนำเสนอกระบวนการอัลฟาทรिमอาร์ริมาที่ประกอบด้วย 3 ขั้นตอนหลัก นั่นคือ การแต่งขอบข้อมูล (data trimming) การพิจารณาการแจกแจงแบบปกติของข้อมูล และการเลือกตัวแบบและทำนายค่า BOD ดังภาพที่ 2



ภาพที่ 2 กระบวนการอัลฟาทรिमอาร์ริมา

3.2.1 การแต่งขอบข้อมูล

ในขั้นตอนของการแต่งขอบข้อมูล จะพิจารณาจากค่าความเบ้ (Skewness) ของค่า BOD ในแต่ละสถานี โดยค่าความเบ้มีสูตรคำนวณดังนี้

$$\text{Skewness} = \frac{\sum_{t=1}^n \left[\frac{(Y_t - \bar{Y})}{S} \right]^3}{n}$$

โดยที่ Y_t คือ ค่า BOD ณ เวลาที่ t

\bar{Y} คือ ค่าเฉลี่ย BOD ของสถานีที่ i

S คือ ส่วนเบี่ยงเบนมาตรฐานของค่า BOD

n คือ จำนวนข้อมูลทั้งหมด

ถ้า Skewness = 0 แสดงว่า ข้อมูลมีลักษณะสมมาตรเป็นรูปประฆังคว่ำ

ถ้า Skewness > 0 แสดงว่า ข้อมูลมีลักษณะเบ้ขวา ค่า BOD ที่มากที่สุดจะถูกแต่งขอบ (trim) แล้วแทนค่าที่ถูกแต่งขอบด้วยค่าทางสถิติ 3 ค่าที่แตกต่างกัน คือ ค่าเฉลี่ยของ BOD ค่ามัธยฐานของ BOD และค่า BOD ณ เวลาที่ $t - 2$

ถ้า Skewness < 0 แสดงว่า ข้อมูลมีลักษณะเบ้ซ้าย ค่า BOD ที่น้อยที่สุดจะถูกแต่งขอบ แล้วแทนค่าที่ถูกแต่งขอบด้วยค่าทางสถิติ 3 ค่าที่แตกต่างกันดังที่กล่าวข้างต้น นั่นคือหลังจากการแต่งขอบ จะได้ข้อมูลชุดใหม่ 3 ชุด คือ ชุดข้อมูลที่แทนด้วยค่าเฉลี่ยของ BOD ชุดข้อมูลที่แทนด้วยค่ามัธยฐานของ BOD และชุดข้อมูลที่แทนด้วยค่า BOD ณ เวลาที่ $t - 2$

3.2.2 การพิจารณาการแจกแจงแบบปกติของข้อมูล (Normal distribution)

หลังจากผ่านขั้นตอนการแต่งขอบ จะทดสอบการแจกแจงแบบปกติของชุดข้อมูลด้วยตัวทดสอบสถิติ Shapiro-Wilk test เพราะจำนวนข้อมูล BOD ของแต่ละสถานีมีจำนวนน้อยกว่า 50 ค่า [15-17] ซึ่งผู้ทำวิจัยสามารถใช้โปรแกรม SPSS for Windows และ Minitab ช่วยในการวิเคราะห์

สำหรับงานวิจัยนี้ การแต่งขอบข้อมูลผู้ทำวิจัยจะแต่งขอบข้อมูลไม่เกิน 3 ครั้ง เพราะถ้าทำการแต่งขอบข้อมูลมากเกินไป จะทำให้ตัวแบบที่สร้างขึ้นเป็นชุดข้อมูลของค่าประมาณ ไม่ใช่ค่าจริง ซึ่งอาจทำให้การทำนายค่า BOD มีความแม่นยำน้อย ดังนั้น อัลฟา (α) คือจำนวนครั้งที่ทำการแต่งขอบข้อมูลจนกระทั่งข้อมูลมีการแจกแจงแบบปกติ

3.2.3 การเลือกตัวแบบและทำนายค่า BOD

ในขั้นตอนนี้จะนำข้อมูลที่มีการแจกแจงแบบปกติมาทำนายค่า BOD โดยการวิเคราะห์อนุกรมเวลาแบบบอซซ์-เจนคินส์ซึ่งเป็นการสร้างตัวแบบ 3 ตัวแบบที่แตกต่างกัน (ชุดข้อมูลที่ถูกแทนด้วยค่าเฉลี่ยของ BOD ชุดข้อมูลที่ถูกแทนด้วยค่ามัธยฐานของ BOD และชุดข้อมูลที่ถูกแทนด้วยค่า BOD ณ เวลาที่ $t - 2$) สำหรับข้อมูล BOD ในแต่ละสถานี ดังนั้นจะมีตัวแบบ 3 ตัวแบบที่ดีที่สุดและ

สามารถทำนายค่า BOD ในอนาคตได้ 3 ค่า สำหรับขั้นตอนในการสร้างตัวแบบโดยการวิเคราะห์อนุกรมเวลาแบบบอซเจนคินส์ เพื่อทำนายค่า BOD ในอนาคตของงานวิจัยนี้มีขั้นตอนดังนี้

1. พิจารณาความเป็นสเตชันนารี เพื่อกำหนดค่า d โดยการพิจารณาจากลักษณะของกราฟ r_k และ r_{kk} ของข้อมูล BOD
2. กำหนดตัวแบบให้มีความเหมาะสม (เลือกค่า p และ q) นั่นคือผู้ทำวิจัยนำค่า BIC มาช่วยในการกำหนดตัวแบบเพื่อลดความยุ่งยากของกระบวนการบอซเจนคินส์ โดยการพิจารณาค่า BIC ที่น้อยที่สุด [13, 14] ซึ่งสามารถใช้โปรแกรม SPSS for Windows ช่วยในการคำนวณได้
3. ประเมินค่าพารามิเตอร์ต่าง ๆ ในสมการ (1) หรือ (2) โดยวิธีกำลังสองน้อยที่สุด ซึ่งสามารถใช้โปรแกรม SPSS for Windows และ Minitab ช่วยในการคำนวณได้
4. ทำนายค่า BOD ในอนาคต จากการสร้างตัวแบบทั้ง 3 ตัวแบบสำหรับชุดข้อมูล BOD ในแต่ละสถานี
5. ทำการเลือกตัวแบบเพียงตัวแบบเดียว ที่ทำให้ผลการทำนายค่า BOD มีความแม่นยำมากที่สุด โดยการพิจารณาค่า α ที่น้อยที่สุด หากตัวแบบที่สร้างขึ้นมามีค่า α ที่เท่ากัน ขั้นตอนวิธีจะเลือกตัวแบบที่มี RMSE ที่น้อยที่สุด นั่นคือขั้นตอนวิธีจะเลือกตัวแบบที่ดีที่สุดเพียงตัวแบบเดียวสำหรับแต่ละสถานีเพื่อทำนายค่า BOD ในอนาคต

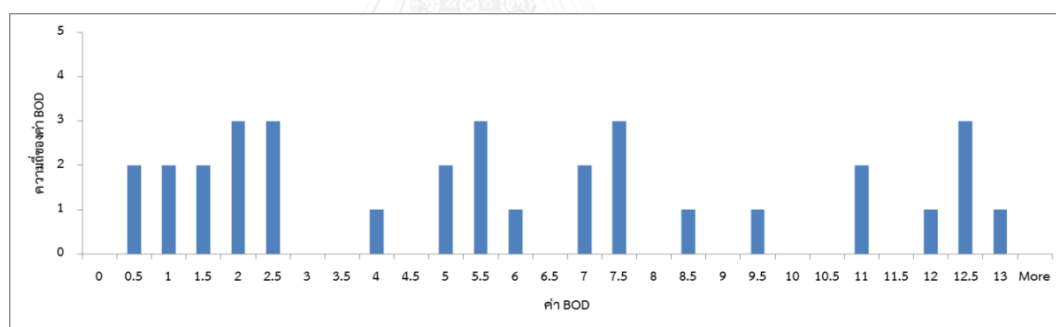
ต่อไปจะเป็นการนำเสนอตัวอย่าง การใช้กระบวนการอัลฟาทริมาอริมา สำหรับตัวอย่างชุดข้อมูล BOD ที่ได้จากการสุ่มค่า BOD ในช่วง 0-13 mg/l โดยใช้โปรแกรม Microsoft Excel จำนวน 33 ค่า ดังตารางที่ 3 โดยชุดข้อมูล BOD นี้มีค่าเฉลี่ยและค่ามัธยฐาน เท่ากับ 5.70 และ 5.17 ตามลำดับ

ตารางที่ 3 ตัวอย่างชุดข้อมูล BOD

เวลา (t)	BOD (mg/l)
1	6.73
2	2.50
3	5.13
4	1.22
5	5.17
6	7.49
7	1.57
8	8.34
9	12.09
10	7.44
11	1.53
12	0.17
13	11.63
14	10.54
15	7.36
16	12.90
17	4.72
18	9.36
19	2.38
20	0.78
21	1.98
22	5.95
23	2.47

24	0.79
25	3.85
26	7.00
27	12.07
28	4.96
29	5.34
30	10.92
31	12.13
32	0.14
33	1.32

จากข้อมูล BOD ในตารางที่ 3 สามารถเขียนเป็นฮิสโทแกรมได้ โดยให้ความกว้างของอันตรภาคชั้นเท่ากับ 0.5 ดังภาพที่ 1



ภาพที่ 3 ฮิสโทแกรมของตัวอย่างชุดข้อมูล BOD

จากฮิสโทแกรมของตัวอย่างชุดข้อมูล BOD ในภาพที่ 3 นำมาวิเคราะห์ค่าความเบ้ของชุดข้อมูล BOD ณ $t = 1, \dots, 33$ นั้นคือสามารถคำนวณค่าความเบ้ได้ 0.327 ซึ่งค่าความเบ้มากกว่า 0 แสดงว่าข้อมูลชุดนี้มีลักษณะเบ้ขวา ดังนั้นจะทำการแต่งขอบข้อมูลครั้งที่ 1 โดยการแต่งขอบค่า BOD ที่มากที่สุดจำนวน 1 ค่า (BOD=12.90 mg/l ณ $t = 16$) แล้วแทนค่าที่ถูกแต่งขอบด้วยค่าทางสถิติทั้ง 3 ค่า ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน และค่า BOD ณ $t - 2$ (สำหรับ $t = 16$) ที่คำนวณจาก $t = 1, \dots, 15, 17, \dots, 33$ ซึ่งผลที่ได้คือ 5.47, 5.15 และ 10.54 ($t = 16$) ตามลำดับ ดังตารางที่ 4

ในการแต่งขอบข้อมูลครั้งที่ 1 จะได้ชุดข้อมูลใหม่ 3 ชุด คือชุดข้อมูลที่ค่า BOD ณ $t = 16$ ถูกแทนด้วยค่าเฉลี่ยของ BOD ชุดข้อมูลที่แทนด้วยค่ามัธยฐานของ BOD และชุดข้อมูลที่แทนด้วยค่า BOD ณ $t - 2$ ต่อไปจะพิจารณาการแจกแจงแบบปกติของชุดข้อมูลทั้ง 3 ชุด โดยตัวทดสอบสถิติ Shapiro-Wilk test พบว่าชุดข้อมูลทั้ง 3 ชุด ไม่มีการแจกแจงแบบปกติ

ตารางที่ 4 การแต่งขอบข้อมูลครั้งที่ 1 สำหรับชุดข้อมูล BOD

เวลา (t)	BOD (mg/l)	แทนค่า BOD ด้วยค่าทางสถิติ		
		ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t - 2$
16	12.90	5.47	5.15	10.54

จากการแต่งขอบข้อมูลครั้งที่ 1 ชุดข้อมูลทั้ง 3 ชุด ยังไม่มีการแจกแจงแบบปกติ จึงต้องพิจารณาค่าความเบ้ของข้อมูล BOD ที่คำนวณจาก $t = 1, \dots, 15, 17, \dots, 33$ นั่นคือคำนวณค่าความเบ้ได้ 0.348 ซึ่งค่าความเบ้มากกว่า 0 แสดงว่าข้อมูลมีลักษณะเบ้ขวา ดังนั้นจึงทำการแต่งขอบข้อมูลครั้งที่ 2 โดยการแต่งขอบค่า BOD ที่มากที่สุดเพิ่มอีก 1 ค่า (BOD=12.13mg/l ณ $t = 31$) แล้วแทนค่าที่ถูกแต่งขอบด้วยค่าทางสถิติทั้ง 3 ค่า ได้แก่ ค่าเฉลี่ย ค่ามัธยฐานและค่า BOD ณ $t-2$ (สำหรับ $t = 16, 31$) ที่คำนวณจาก $t = 1, \dots, 15, 17, \dots, 30, 32, 33$ ซึ่งผลที่ได้คือ 5.26, 5.15 และ 10.54 ($t = 16$), 5.34 ($t = 31$) ตามลำดับ ดังตารางที่ 5 จากนั้นพิจารณาการแจกแจงแบบปกติของชุดข้อมูลทั้ง 3 ชุด โดยตัวทดสอบสถิติ Shapiro-Wilk test พบว่าชุดข้อมูลที่แทนด้วยค่าเฉลี่ยมีการแจกแจงแบบปกติ และชุดข้อมูลที่แทนด้วยค่ามัธยฐานและค่า BOD ณ $t - 2$ ข้อมูลไม่มีการแจกแจงแบบปกติ

ตารางที่ 5 การแต่งขอบข้อมูลครั้งที่ 2 สำหรับชุดข้อมูล BOD

เวลา (t)	BOD (mg/l)	แทนค่า BOD ด้วยค่าทางสถิติ		
		ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t - 2$
16	12.90	5.26	5.13	10.54
31	12.13	5.26	5.13	5.34

จากการแต่งขอบข้อมูลครั้งที่ 2 ชุดข้อมูลที่แทนด้วยค่ามัธยฐานและค่า BOD ณ $t-2$ ข้อมูลไม่มีการแจกแจงแบบปกติ จึงต้องพิจารณาค่าความเบ้ของข้อมูล BOD ที่คำนวณจาก $t=1, \dots, 15, 17, \dots, 30, 32, 33$ นั่นคือคำนวณค่าความเบ้ได้ 0.387 ซึ่งค่าความเบ้มากกว่า 0 แสดงว่าข้อมูลมีลักษณะเบ้ขวา ดังนั้นจึงทำการแต่งขอบข้อมูลครั้งที่ 3 โดยการแต่งขอบค่า BOD ที่มากที่สุดเพิ่มอีก 1 ค่า (BOD=12.09 mg/l ณ $t=9$) แล้วแทนค่าที่ถูกแต่งขอบด้วยค่าทางสถิติทั้ง 2 ค่า ได้แก่ ค่ามัธยฐานและค่า BOD ณ $t-2$ (สำหรับ $t=9, 16, 31$) ที่คำนวณจาก $t=1, \dots, 8, 10, \dots, 15, 17, \dots, 30, 32, 33$ คือ 5.05 และ 1.57 ($t=9$), 10.54 ($t=16$), 5.34 ($t=31$) ตามลำดับ ดังตารางที่ 6 จากนั้นจะพิจารณาการแจกแจงแบบปกติของชุดข้อมูลทั้ง 2 ชุด โดยตัวทดสอบสถิติ Shapiro-Wilk test พบว่าชุดข้อมูลที่แทนด้วยค่ามัธยฐานมีการแจกแจงแบบปกติ และชุดข้อมูลที่แทนด้วยค่า BOD ณ $t-2$ ไม่มีการแจกแจงแบบปกติ

ตารางที่ 6 การแต่งขอบข้อมูลครั้งที่ 3 สำหรับชุดข้อมูล BOD

เวลา (t)	BOD (mg/l)	แทนค่า BOD ด้วยค่าทางสถิติ	
		ค่ามัธยฐาน	ณ $t-2$
9	12.09	5.05	1.57
16	12.90	5.05	10.54
31	12.13	5.05	5.34

ดังที่กล่าวไว้ข้างต้นว่า ในงานวิจัยนี้ค่าสูงสุดของ α เป็น 3 ดังนั้นในขั้นตอนของการสร้างตัวแบบ จะมีชุดข้อมูล 2 ชุด นั่นคือชุดข้อมูลที่แทนด้วยค่าเฉลี่ยโดยการแต่งขอบครั้งที่ 2 และชุดข้อมูลที่แทนด้วยค่ามัธยฐานโดยการแต่งขอบครั้งที่ 3 ดังตารางที่ 7 ซึ่งจะเห็นได้ว่าชุดข้อมูลที่แทนด้วยค่าเฉลี่ยมีค่า α น้อยที่สุด จึงนำข้อมูลชุดนี้มาหาตัวแบบ เพื่อทำนายค่า BOD ในอนาคต ดังที่กล่าวไว้ในหัวข้อ 3.2.3

ตารางที่ 7 ค่า α สำหรับชุดข้อมูล

	แทนค่า BOD ด้วยค่าทางสถิติ		
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t-2$
ชุดข้อมูล	$\alpha = 2$	$\alpha = 3$	-

บทที่ 4

ผลการวิเคราะห์ข้อมูลวิจัย

ในบทนี้ผู้ทำวิจัยแสดงผลการทดลองของการทำนายค่า BOD สำหรับน้ำในแม่น้ำเจ้าพระยาของข้อมูล 11 สถานี คือ สถานีที่ 1, 3, 6, 8, 10, 12, 15, 16.1, 17, 18 และ 20 เนื่องจากข้อมูล BOD ที่สถานีปลายทางน้ำมีปัญหาด้านสิ่งแวดล้อมมากกว่าสถานีต้นน้ำ จึงเลือกพิจารณาจากข้อมูลสถานีที่ 1 ถึงสถานีที่ 20 ทางผู้ทำวิจัยได้วิเคราะห์ข้อมูลทั้ง 11 สถานี โดยผู้วิจัยเริ่มจากการวิเคราะห์อนุกรมเวลาแบบบอซซ์เจนคินส์และเทคนิคการทำให้เรียบ พบว่า เมื่อวิเคราะห์อนุกรมเวลาแบบบอซซ์เจนคินส์ของชุดข้อมูลทั้ง 2 ชุด สำหรับแต่ละสถานี ได้ตัวแบบ ARIMA(0,0,0) ซึ่งค่าพยากรณ์อนุกรมของค่า BOD ณ เวลาต่าง ๆ และการทำนายค่า BOD ในอนาคต คือ ค่าเฉลี่ยของชุดข้อมูลนั้น ๆ ทางผู้วิจัยจึงได้สร้างกระบวนการใหม่คือ กระบวนการอัลฟาทรिमอาร์ริมา ซึ่งในบทนี้ได้วิเคราะห์ข้อมูลโดยใช้กระบวนการอัลฟาทรिमอาร์ริมา และทำการเปรียบเทียบผลการทำนายกับผลที่ได้จากการใช้เทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) ส่วนสถานีที่ไม่ได้นำเสนอสามารถใช้กระบวนการอัลฟาทรिमอาร์ริมาได้เช่นเดียวกัน

4.1 ข้อมูลชุดที่ 1

จากการวิเคราะห์ข้อมูล BOD ของข้อมูลชุดที่ 1 จำนวน 32 ค่า ที่เก็บรวบรวมในเดือนมิถุนายนและธันวาคมทั้ง 11 สถานี พบว่า เมื่อผู้ทำวิจัยใช้กระบวนการอัลฟาทรिमอาร์ริมา มีเพียง 4 สถานี คือ สถานีที่ 3, 17, 18 และ 20 ที่สามารถหาตัวแบบของกระบวนการอัลฟาทรिमอาร์ริมาได้ ดังตารางที่ 8 โดยแต่ละตัวแบบของแต่ละสถานีมีค่า α ที่แตกต่างกัน ดังตารางที่ 9 และอีก 7 สถานีไม่สามารถหาตัวแบบจากกระบวนการดังกล่าวได้ เนื่องจากทำการแต่งขอบข้อมูล 3 ครั้ง แล้วข้อมูลก็ยังไม่มีการแจกแจงแบบปกติ

ตารางที่ 8 ตัวแบบของกระบวนการอัลฟาทรिमอาร์ริมาของข้อมูลชุดที่ 1 สำหรับแต่ละสถานี

สถานีที่	ตัวแบบของกระบวนการอัลฟาทรिमอาร์ริมา		
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t-2$
3	ARIMA(0,0,0)	-	ARIMA(0,0,0)(0,1,1) ₂
17	ARIMA(1,0,3)	ARIMA(1,0,3)	ARIMA(1,0,3)
18	ARIMA(1,0,1)	ARIMA(1,0,1)	ARIMA(1,0,1)
20	ARIMA(1,0,3)	ARIMA(1,0,3)	ARIMA(1,0,3)

ตารางที่ 9 ค่า α ของข้อมูลชุดที่ 1 สำหรับแต่ละสถานี

สถานีที่	กระบวนการอัลฟาทรिमอริมา		
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t-2$
3	$\alpha = 3$	-	$\alpha = 3$
17	$\alpha = 3$	$\alpha = 3$	$\alpha = 3$
18	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$
20	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$

จากตารางที่ 9 จะเห็นได้ว่าค่า α ของสถานีที่ 3 เมื่อแทนค่าด้วยค่ามัธยฐานไม่มีค่าของ α เพราะผู้ทำวิจัยทำการแต่งขอบข้อมูล 3 ครั้ง แล้วข้อมูลก็ยังไม่มีการแจกแจงแบบปกติ นอกนั้นค่า α ในแต่ละสถานีมีค่าเท่ากัน สำหรับตัวแบบทั้ง 3 แบบ ดังนั้นผู้ทำวิจัยจะพิจารณาค่า RMSE ที่น้อยที่สุดของกระบวนการอัลฟาทรिमอริมา เพื่อเลือกตัวแบบที่ดีที่สุด พบว่า สถานีที่ 3, 18 และ 20 เลือกตัวแบบที่แทนข้อมูลด้วยค่า BOD ณ $t-2$ และสถานีที่ 17 เลือกตัวแบบที่แทนข้อมูลด้วยค่าเฉลี่ย ดังตารางที่ 10

จากนั้นเปรียบเทียบค่า RMSE โดยนำชุดข้อมูล BOD ที่สามารถใช้กระบวนการอัลฟาทรिमอริมาได้ (ชุดข้อมูล BOD ของสถานีที่ 3, 17, 18 และ 20) มาใช้เทคนิคการทำให้เรียบ พบว่า อนุกรมเวลาทั้ง 4 สถานีใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล และเมื่อวิเคราะห์อนุกรมเวลาแบบบ็อกซ์เจนคินส์ทั้ง 4 สถานี ได้ตัวแบบ ARIMA(0,0,0) ดังตารางที่ 10

ตารางที่ 10 เปรียบเทียบค่า RMSE ระหว่างกระบวนการอัลฟาทรिमอริมากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 1

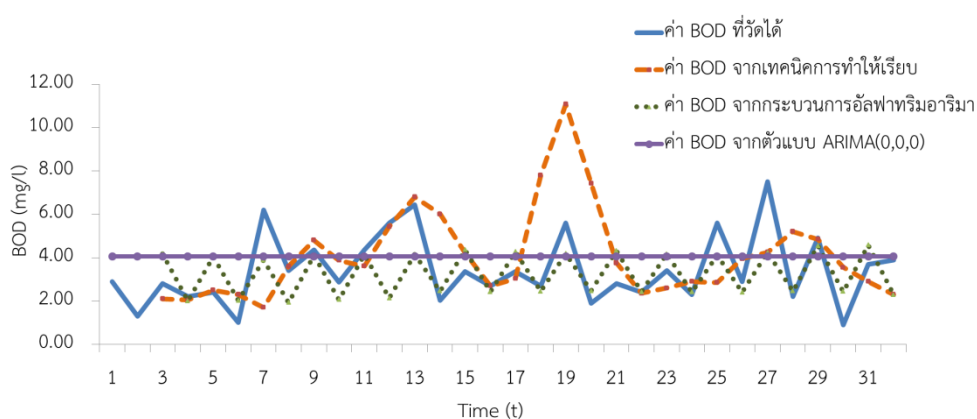
สถานีที่	ค่า RMSE				
	กระบวนการอัลฟาทรिमอริมา			เทคนิคการทำให้เรียบ	ตัวแบบ ARIMA(0,0,0)
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t-2$		
3	1.5856	-	1.3705	2.9065	2.5769
17	0.5054	0.5141	0.5690	1.2142	0.9644
18	0.7243	0.7248	0.7234	0.8059	0.8143
20	0.5354	0.5335	0.5325	0.9626	0.6597

เมื่อได้ตัวแบบที่ดีที่สุดแล้ว ผู้ทำวิจัยจึงทำนายค่า BOD ในอนาคต ณ $t = 33$ เพื่อคำนวณเปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ของกระบวนการอัลฟาทรिमอาร์ิม่าเปรียบเทียบกับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) ดังตารางที่ 11 พบว่า เปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ของกระบวนการอัลฟาทรिमอาร์ิม่าดีกว่าเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) โดยกระบวนการอัลฟาทรिमอาร์ิม่าสามารถลดเปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ได้มากกว่า 50 % สำหรับสถานีที่ 3 สามารถลดได้ 50 % สำหรับสถานีที่ 20 และสถานีที่ 17, 18 เมื่อใช้กระบวนการอัลฟาทรिमอาร์ิม่ายังคงดีกว่าเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0)

ตารางที่ 11 เปรียบเทียบเปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ ระหว่างกระบวนการอัลฟาทรिमอาร์ิม่ากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 1

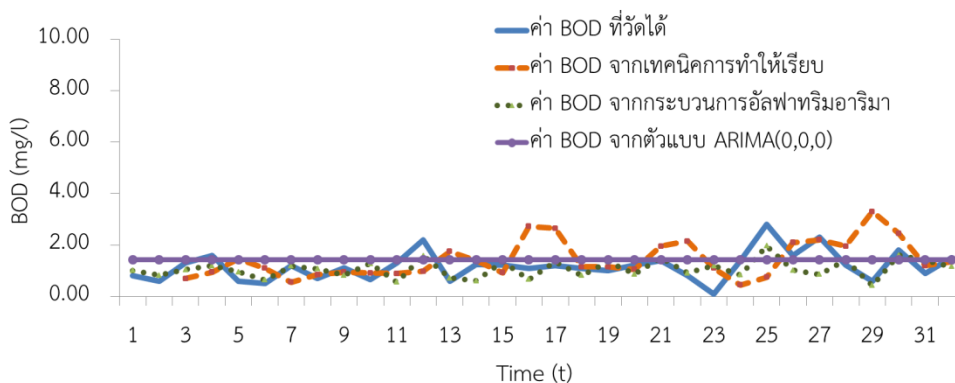
สถานีที่	กระบวนการอัลฟาทรिमอาร์ิม่า	เทคนิคการทำให้เรียบ	ตัวแบบ ARIMA(0,0,0)
3	3.2580 %	30.9091 %	26.3636 %
17	46.8591 %	65.7143 %	59.4286 %
18	60.7373 %	61.3636 %	66.1364 %
20	13.8246 %	26.2857 %	52.5714 %

ภาพที่ 4-7 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ระหว่างค่าจริง ค่าที่ได้จากการ ใช้เทคนิคการทำให้เรียบ ตัวแบบ ARIMA(0,0,0) และค่าที่ได้จากกระบวนการอัลฟาทรिमอาร์ิม่าสำหรับ $t = 1, 2, 3, \dots, 32$ ของสถานีที่ 3, 17, 18 และ 20 ตามลำดับ



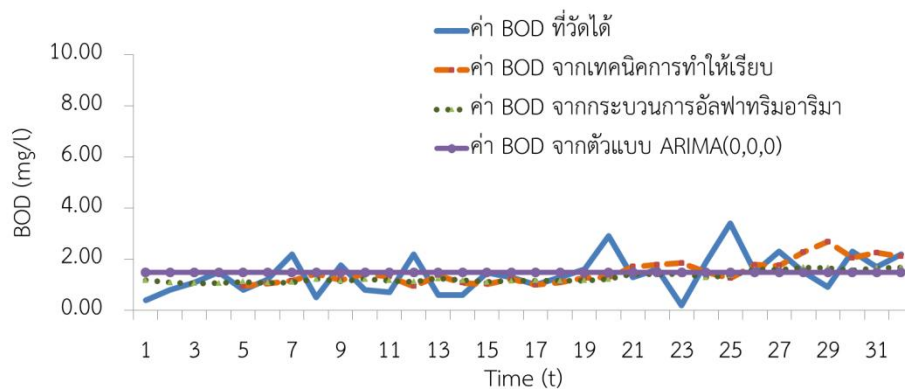
ภาพที่ 4 กราฟแสดงค่า BOD ของสถานีที่ 3 สำหรับข้อมูลชุดที่ 1

จากภาพที่ 4 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 3 ระหว่างค่าจริง และเทคนิคทั้ง 3 เทคนิค จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล และค่า BOD จากตัวแบบ ARIMA(0,0,0) มีการเคลื่อนไหวที่ไม่ใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ (RMSE=2.9065, 2.5769 ตามลำดับ) ส่วนกราฟของค่า BOD จากกระบวนการอัลฟาทริมอาร์ิม่าที่แทนข้อมูลด้วยค่า BOD ณ $t-2$ (ARIMA(0,0,0)(0,1,1)₂) มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ ส่งผลให้ค่า RMSE ของกระบวนการอัลฟาทริมอาร์ิม่ามีค่าน้อยที่สุด (RMSE=1.3705)



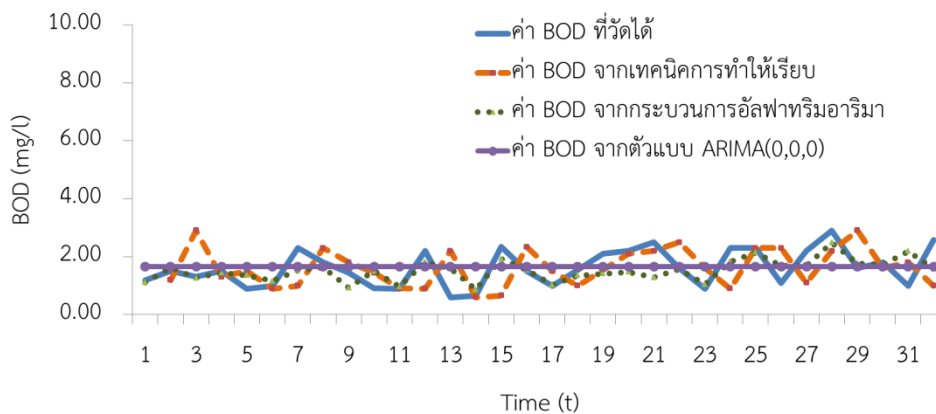
ภาพที่ 5 กราฟแสดงค่า BOD ของสถานีที่ 17 สำหรับข้อมูลชุดที่ 1

จากภาพที่ 5 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 17 ระหว่างค่าจริง และเทคนิคทั้ง 3 เทคนิค จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล และค่า BOD จากตัวแบบ ARIMA(0,0,0) มีการเคลื่อนไหวที่ไม่ใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ (RMSE=1.2142, 0.9644 ตามลำดับ) ส่วนกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอริมาที่แทนข้อมูลด้วยค่าเฉลี่ย (ARIMA(1,0,3)) มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ ส่งผลให้ค่า RMSE ของกระบวนการอัลฟาทรिमอริมามีค่าน้อยที่สุด (RMSE=0.5054)



ภาพที่ 6 กราฟแสดงค่า BOD ของสถานีที่ 18 สำหรับข้อมูลชุดที่ 1

จากภาพที่ 6 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 18 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มี แนวโน้มและฤดูกาล (RMSE=0.8059) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=0.8143) และค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมาที่แทนข้อมูลด้วยค่า BOD ณ $t - 2$ (ARIMA(1,0,1)) ซึ่งมีค่า RMSE=0.7234 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมา มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



ภาพที่ 7 กราฟแสดงค่า BOD ของสถานีที่ 20 สำหรับข้อมูลชุดที่ 1

จากภาพที่ 7 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 20 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มี แนวโน้มและฤดูกาล (RMSE=0.9626) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=0.6597) และค่า BOD จากกระบวนการอัลฟาทริมอาร์ริมาที่แทนข้อมูลด้วยค่า BOD ณ $t-2$ (ARIMA(1,0,3)) ซึ่งมีค่า RMSE=0.5325 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทริมอาร์ริมา มีการ เคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด

4.2 ข้อมูลชุดที่ 2

สำหรับการวิเคราะห์ข้อมูล BOD ของข้อมูลชุดที่ 2 จำนวน 33 ค่า ที่เก็บรวบรวมในเดือน กุมภาพันธ์และสิงหาคม พบว่า ทั้ง 11 สถานีสามารถหาตัวแบบจากกระบวนการอัลฟาทรिमอาร์มาได้ ดังตารางที่ 12 โดยแต่ละตัวแบบ ของแต่ละสถานีมีค่า α ที่แตกต่างกันดังตารางที่ 13

ตารางที่ 12 ตัวแบบของกระบวนการอัลฟาทรिमอาร์มาของข้อมูลชุดที่ 2 สำหรับแต่ละสถานี

สถานีที่	ตัวแบบของกระบวนการอัลฟาทรिमอาร์มา		
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t - 2$
1	ARIMA(1,0,1)(1,1,1) ₂	ARIMA(1,0,1)(1,1,1) ₂	ARIMA(1,0,1)(0,1,0) ₂
3	ARIMA(0,0,0)	ARIMA(0,0,0)	ARIMA(0,0,0)
6	ARIMA(0,0,1)	ARIMA(0,0,1)	ARIMA(1,0,1)
8	ARIMA(1,0,1)(0,1,0) ₂	ARIMA(1,0,1)(0,1,0) ₂	ARIMA(1,0,1)(0,1,0) ₂
10	ARIMA(1,0,1)(0,1,0) ₂	ARIMA(1,0,1)(0,1,0) ₂	ARIMA(1,0,1)(0,1,0) ₂
12	ARIMA(0,0,1)(0,1,0) ₂	ARIMA(1,0,1)(0,1,0) ₂	ARIMA(1,0,1)(0,1,0) ₂
15	ARIMA(1,0,2)	ARIMA(1,0,2)	ARIMA(1,0,2)
16.1	ARIMA(2,0,3)	-	ARIMA(2,0,3)
17	ARIMA(2,0,2)	-	-
18	ARIMA(1,0,3)	ARIMA(1,0,1)	ARIMA(1,0,3)
20	ARIMA(1,0,1)	ARIMA(1,0,1)	ARIMA(1,0,1)

ตารางที่ 13 ค่า α ของข้อมูลชุดที่ 2 สำหรับแต่ละสถานี

สถานีที่	กระบวนการอัลฟาทรिमอริมา		
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t-2$
1	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$
3	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$
6	$\alpha = 1$	$\alpha = 1$	$\alpha = 2$
8	$\alpha = 2$	$\alpha = 3$	$\alpha = 2$
10	$\alpha = 3$	$\alpha = 3$	$\alpha = 3$
12	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$
15	$\alpha = 3$	$\alpha = 3$	$\alpha = 3$
16.1	$\alpha = 1$	-	$\alpha = 1$
17	$\alpha = 1$	-	-
18	$\alpha = 3$	$\alpha = 3$	$\alpha = 3$
20	$\alpha = 1$	$\alpha = 1$	$\alpha = 1$

จากตารางที่ 13 จะเห็นได้ว่าค่า α ของสถานีที่ 6 และ 8 มีค่าไม่เท่ากัน ผู้ทำวิจัยจะเลือก α ที่น้อยที่สุดของสถานีที่ 6 และ 8 คือ $\alpha = 1$ และ $\alpha = 2$ ตามลำดับ ซึ่งมีค่า α 2 ค่าที่เท่ากัน ผู้ทำวิจัยจะพิจารณาค่า RMSE ที่น้อยที่สุดของกระบวนการอัลฟาทรिमอริมา เพื่อเลือกตัวแบบที่ดีที่สุดในการทำนายค่า BOD ในอนาคต นั่นคือ ผู้ทำวิจัยเลือกตัวแบบที่ถูกแทนข้อมูลด้วยค่ามัธยฐาน และ ณ $t-2$ สำหรับสถานีที่ 6 และ 8 ตามลำดับดังตารางที่ 14 และเมื่อผู้ทำวิจัยพิจารณาสถานีที่เหลือ พบว่าในแต่ละสถานีมีค่า α เท่ากัน ดังนั้นผู้ทำวิจัยจะพิจารณาค่า RMSE ที่น้อยที่สุดของกระบวนการอัลฟาทรिमอริมา นั่นคือ สถานีที่ 3, 15, 17, 18 และ 20 เลือกตัวแบบที่แทนข้อมูลด้วยค่าเฉลี่ย สถานีที่ 1 และ 10 เลือกตัวแบบที่แทนข้อมูลด้วยค่ามัธยฐาน และสถานีที่ 12 และ 16.1 เลือกตัวแบบที่แทนข้อมูลด้วย ณ $t-2$ ดังตารางที่ 14

จากนั้นเปรียบเทียบค่า RMSE โดยนำชุดข้อมูล BOD ที่สามารถใช้กระบวนการอัลฟาทรिमอริมาได้ (ชุดข้อมูล BOD ของทั้ง 11 สถานี) มาใช้เทคนิคการทำให้เรียบและการวิเคราะห์อนุกรมเวลาแบบบอซซ์เจนคินส์ พบว่า เมื่อใช้เทคนิคการทำให้เรียบมี 8 สถานีใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล คือสถานีที่ 1, 3, 10, 15, 16.1, 17, 18, 20 มี 2 สถานีใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลที่มีความผันแปรตามฤดูกาล สำหรับตัวแบบการคูณ เนื่องจากอนุกรมเวลาไม่มีแนวโน้มมีความแปรผันตามฤดูกาลและตัวแบบการคูณมีค่า RMSE น้อยที่สุด คือสถานีที่ 8, 12 และมีเพียง 1 สถานีที่ใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสองครั้ง เนื่องจากอนุกรมเวลามีแนวโน้มแต่ไม่มีความแปรผันตามฤดูกาล คือสถานีที่ 6 และเมื่อวิเคราะห์อนุกรมเวลาแบบบอซซ์เจนคินส์ทั้ง 11 สถานีได้ตัวแบบ ARIMA(0,0,0) ดังตารางที่ 14

ตารางที่ 14 เปรียบเทียบค่า RMSE ระหว่างกระบวนการอัลฟาทรिमอริมากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 2

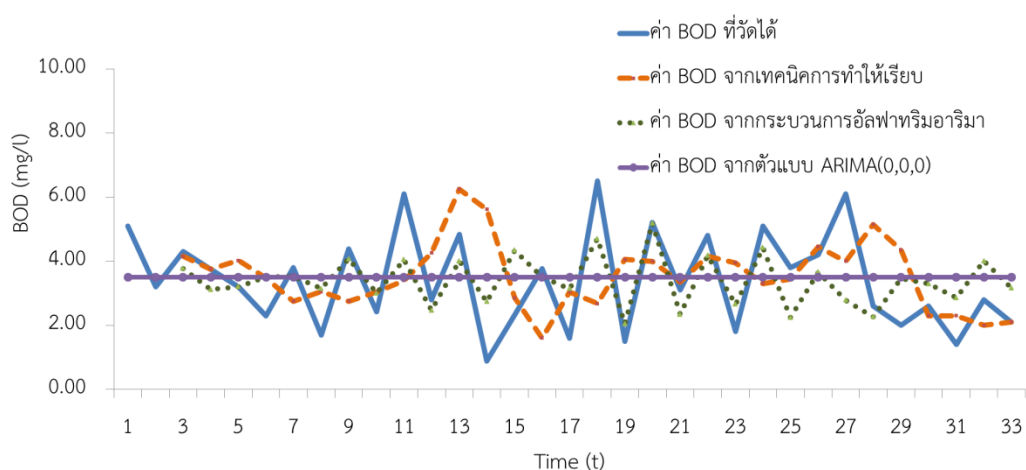
สถานีที่	ค่า RMSE				
	กระบวนการอัลฟาทรिमอริมา			เทคนิคการทำให้เรียบ	ตัวแบบ ARIMA(0,0,0)
	ค่าเฉลี่ย	ค่ามัธยฐาน	ณ $t-2$		
1	1.2351	1.2265	1.3666	2.0348	1.5593
3	1.5280	1.5288	1.5785	2.2341	1.6691
6	1.9415	1.9231	1.7639	2.6394	2.5809
8	2.0330	1.7711	1.9182	2.2145	2.4688
10	1.6323	1.6313	1.6520	2.5524	2.4688
12	1.1836	1.0914	1.0690	1.4871	1.4929
15	0.4019	0.4036	0.4244	1.0746	0.7624
16.1	0.3609	-	0.3601	0.6098	0.4350
17	0.2838	-	-	0.5039	0.3737
18	0.4078	0.4151	0.4142	1.3215	0.9190
20	0.4899	0.4901	0.4907	0.7816	0.5377

เมื่อได้ตัวแบบที่ดีที่สุดแล้ว ผู้ทำวิจัยจึงทำนายค่า BOD ในอนาคต ณ $t = 34$ เพื่อคำนวณเปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ของกระบวนการอัลฟาทรिमอารีมา เปรียบเทียบกับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) ดังตารางที่ 15 พบว่า สำหรับสถานีที่ 1, 6, 8 และ 12 เมื่อใช้กระบวนการอัลฟาทรिमอารีมาสามารถลดเปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ได้มากกว่า 50 % สำหรับสถานีที่ 3, 10, 15, 16.1, 17, 18 และ 20 เมื่อใช้กระบวนการอัลฟาทรिमอารีมายังคงดีกว่าเทคนิคการทำให้เรียบ และสำหรับสถานี 3, 10, 15, 16.1, 17, 18 และ 20 ค่าเปอร์เซ็นต์ของค่าความคลาดเคลื่อนสัมพัทธ์ของตัวแบบ ARIMA(0,0,0) มีค่ามากกว่ากระบวนการอัลฟาทรिमอารีมาและเทคนิคการทำให้เรียบ

ตารางที่ 15 เปรียบเทียบเปอร์เซ็นต์ของความคลาดเคลื่อนสัมพัทธ์ ระหว่างกระบวนการอัลฟาทรिमอารีมากับเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) สำหรับข้อมูลชุดที่ 2

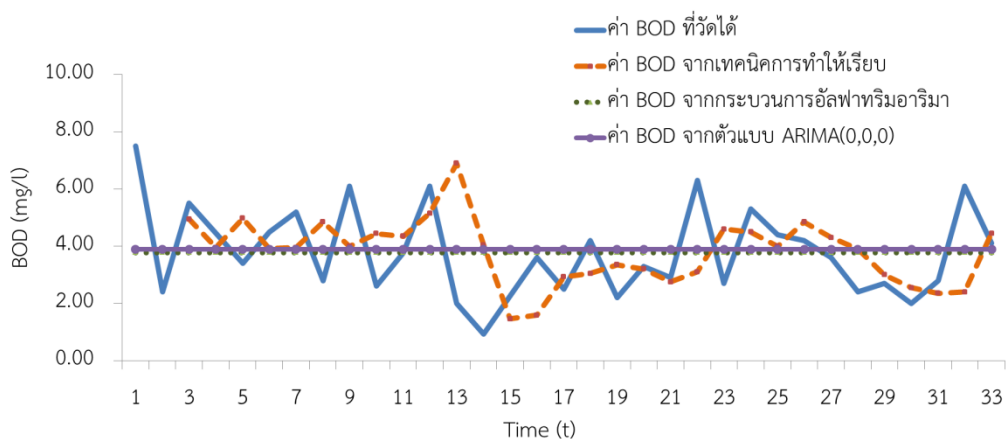
สถานีที่	กระบวนการอัลฟาทรिमอารีมา	เทคนิคการทำให้เรียบ	ตัวแบบ ARIMA(0,0,0)
1	6.30 %	38.75 %	12.50 %
3	10.0 %	21.43 %	7.14 %
6	18.97 %	39.46 %	27.84 %
8	6.12 %	13.67 %	28.00 %
10	28.44 %	44.74 %	8.16 %
12	2.46 %	11.15 %	5.77 %
15	46.04 %	54.55 %	28.64 %
16.1	48.18 %	70.37 %	58.52 %
17	68.89 %	70.0 %	64.67 %
18	46.55 %	50.0 %	38.33 %
20	31.25 %	40.53 %	30.53 %

ภาพที่ 8-20 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ระหว่างค่าจริง ค่าที่ได้จากการใช้เทคนิคการทำให้เรียบ ตัวแบบ ARIMA(0,0,0) และค่าที่ได้จากกระบวนการอัลฟาทรिमอาร์ิมา สำหรับ $t = 1, 2, 3, \dots, 33$ ของสถานีที่ 1, 3, 6, 8, 10, 12, 15, 16.1, 17, 18 และ 20 ตามลำดับ



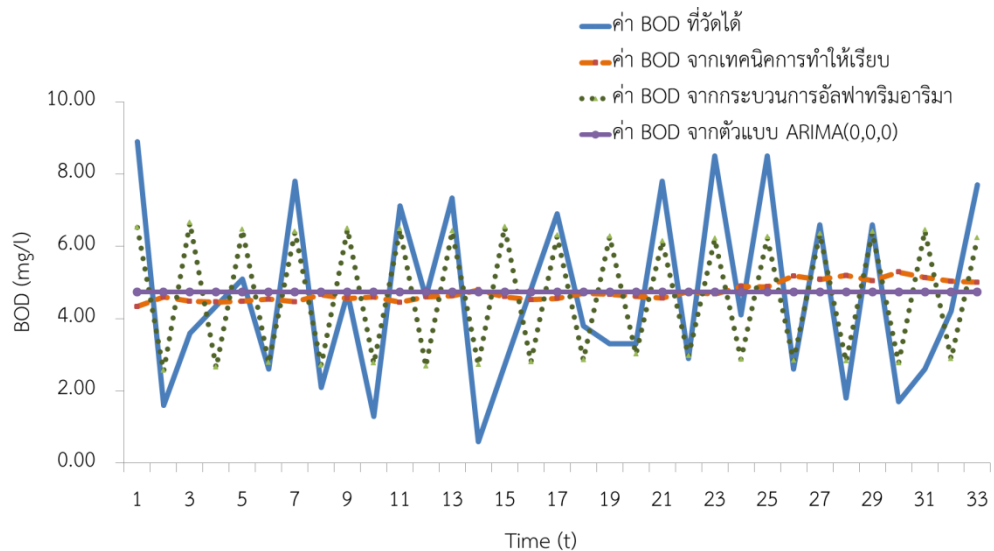
ภาพที่ 8 กราฟแสดงค่า BOD ของสถานีที่ 1 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 8 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 1 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล (RMSE=2.0348) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=1.5593) และค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมาที่แทนข้อมูลด้วยค่ามัธยฐาน (ARIMA(1,0,1)(1,1,1)₂) ซึ่งมีค่า RMSE=1.2265 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมา มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



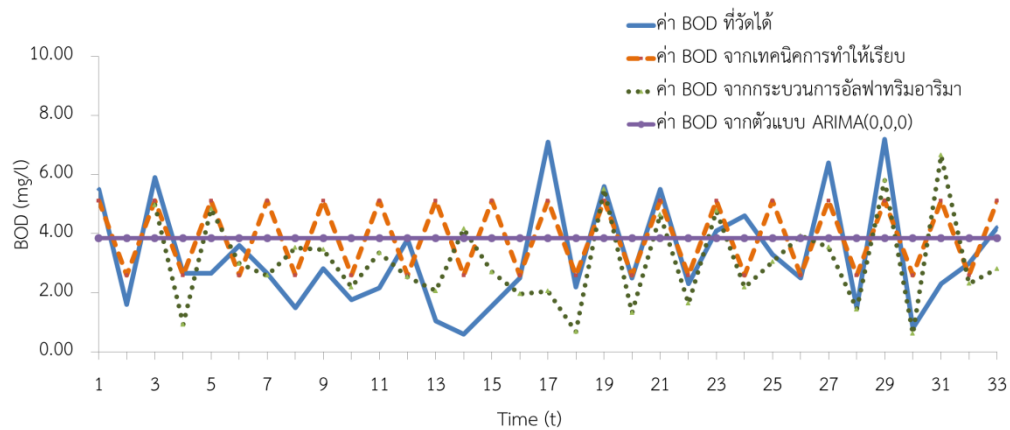
ภาพที่ 9 กราฟแสดงค่า BOD ของสถานีที่ 3 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 9 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 3 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้ม และฤดูกาล (RMSE=2.2341) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=1.6691) และค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมาที่แทนข้อมูลด้วยค่าเฉลี่ย (ARIMA(0,0,0)) ซึ่งมีค่า RMSE=1.5280 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมา มีการเคลื่อนไหวเดียวกับ ตัวแบบ ARIMA(0,0,0) แต่ค่า RMSE ของกระบวนการอัลฟาทรिमอาร์ิมา มีค่าน้อยกว่า



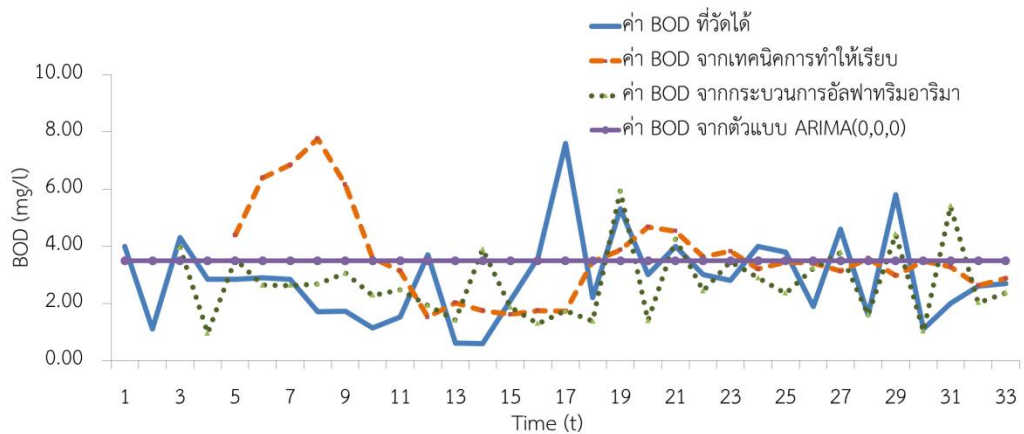
ภาพที่ 10 กราฟแสดงค่า BOD ของสถานีที่ 6 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 10 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 6 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลซ้ำสองครั้ง เนื่องจากอนุกรมเวลามีแนวโน้มแต่ไม่มีความแปรผันตามฤดูกาล ($RMSE=2.6394$) ค่า BOD จากตัวแบบ $ARIMA(0,0,0)$ ($RMSE=2.5809$) และค่า BOD จากกระบวนการอัลฟาทรिमอริมาที่แทนข้อมูลด้วยค่ามัธยฐาน ($ARIMA(0,0,1)$) ซึ่งมีค่า $RMSE=1.9231$ จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอริมามีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า $RMSE$ น้อยที่สุด



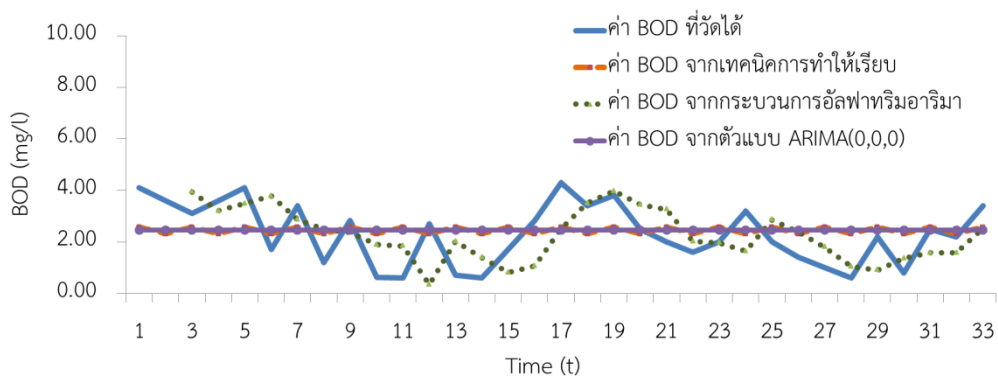
ภาพที่ 11 กราฟแสดงค่า BOD ของสถานีที่ 8 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 11 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 8 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลที่มีความผันแปรตามฤดูกาล เนื่องจากอนุกรมเวลาไม่มีแนวโน้มมีความแปรผันตามฤดูกาล (RMSE=2.2145) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=2.4688) และค่า BOD จากกระบวนการอัลฟาทริมอาร์ิม่าที่แทนข้อมูลด้วยค่า BOD ณ $t-2$ (ARIMA(1,0,1)(0,1,0)₂) ซึ่งมีค่า RMSE=1.9182 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทริมอาร์ิม่ามีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



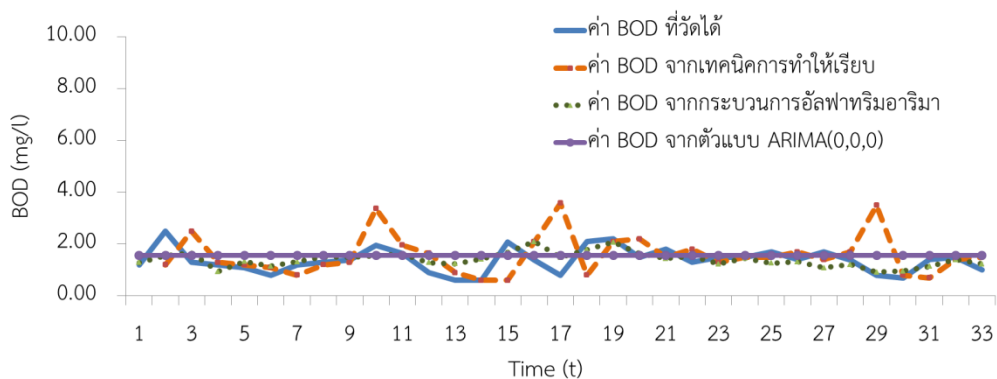
ภาพที่ 12 กราฟแสดงค่า BOD ของสถานีที่ 10 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 12 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 10 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มี แนวโน้มและฤดูกาล (RMSE=2.5524) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=2.4688) และค่า BOD จากกระบวนการอัลฟาทริมอาร์ิมาที่แทนข้อมูลด้วยค่ามัธยฐาน (ARIMA(1,0,1)(0,1,0)₂) ซึ่งมีค่า RMSE=1.6313 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทริมอาร์ิมา มีการ เคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



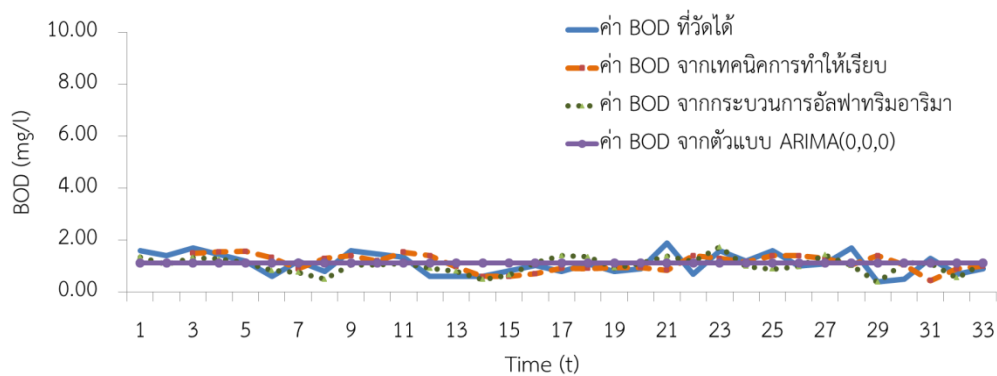
ภาพที่ 13 กราฟแสดงค่า BOD ของสถานีที่ 12 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 13 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 12 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลที่มีความผันแปรตามฤดูกาล เนื่องจากอนุกรมเวลาไม่มีแนวโน้มมีความแปรผันตามฤดูกาล (RMSE=1.4871) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=1.4929) และค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมาที่แทนข้อมูลด้วยค่า BOD ณ $t-2$ (ARIMA(1,0,1)(0,1,0)₂) ซึ่งมีค่า RMSE=1.0690 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมามีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



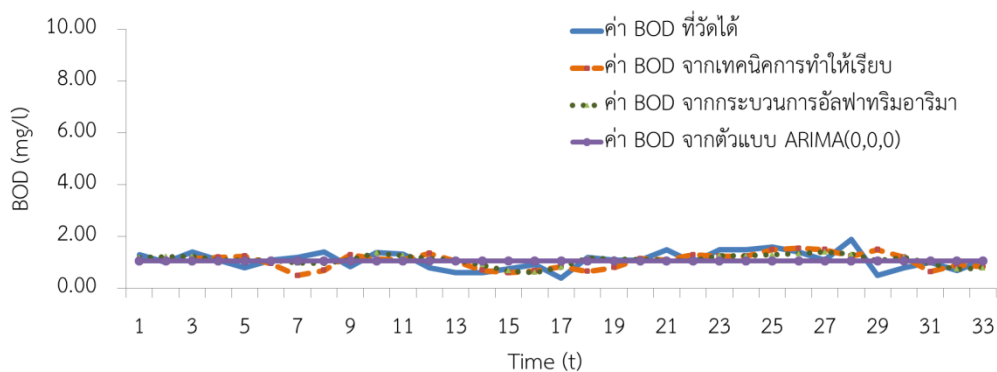
ภาพที่ 14 กราฟแสดงค่า BOD ของสถานีที่ 15 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 14 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 15 ระหว่างค่าจริง และเทคนิคทั้ง 3 เทคนิค จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล และค่า BOD จากตัวแบบ ARIMA(0,0,0) มีการเคลื่อนไหวที่ไม่ใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ (RMSE=1.0746, 0.7624 ตามลำดับ) ส่วนกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิมาที่แทนข้อมูลด้วยค่าเฉลี่ย (ARIMA(1,0,2)) มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ ส่งผลให้ค่า RMSE ของกระบวนการอัลฟาทรिमอาร์ิมา มีค่าน้อยที่สุด (RMSE=0.4019)



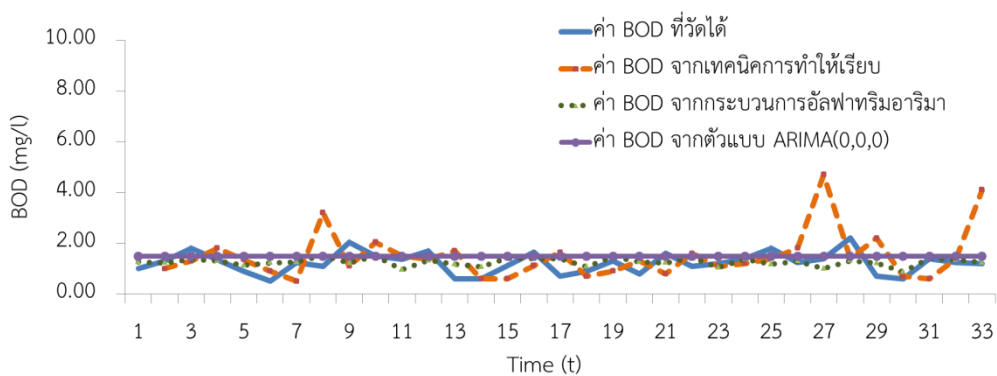
ภาพที่ 15 กราฟแสดงค่า BOD ของสถานีที่ 16.1 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 15 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 16.1 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล (RMSE=0.6098) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=0.4350) และค่า BOD จากกระบวนการอัลฟาทรिमอริมาที่แทนข้อมูลด้วยค่า BOD ณ $t-2$ (ARIMA(2,0,3)) ซึ่งมีค่า RMSE=0.3601 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอริมา มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



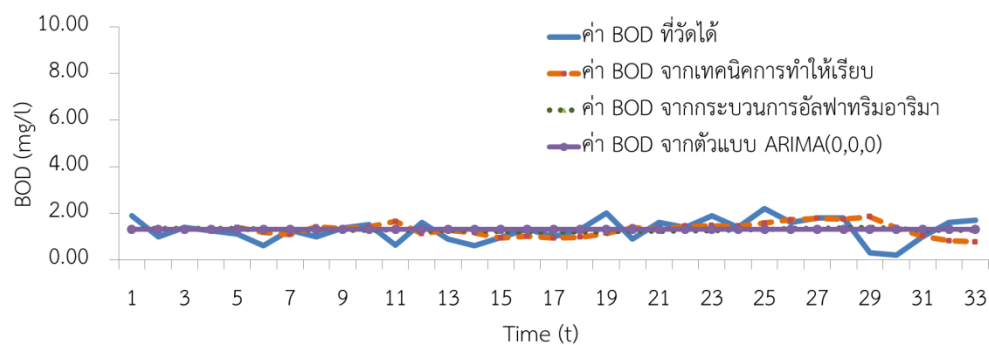
ภาพที่ 16 กราฟแสดงค่า BOD ของสถานีที่ 17 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 16 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 17 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มี แนวโน้มและฤดูกาล (RMSE=0.5039) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=0.3737) และค่า BOD จากกระบวนการอัลฟาทริมอาร์ิมาที่แทนข้อมูลด้วยค่าเฉลี่ย (ARIMA(2,0,2)) ซึ่งมีค่า RMSE=0.2838 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทริมอาร์ิมา มีการ เคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้และมีค่า RMSE น้อยที่สุด



ภาพที่ 17 กราฟแสดงค่า BOD ของสถานีที่ 18 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 17 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 18 ระหว่างค่าจริง และเทคนิคทั้ง 3 เทคนิค จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มีแนวโน้มและฤดูกาล และค่า BOD จากตัวแบบ ARIMA(0,0,0) มีการเคลื่อนไหวที่ไม่ใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ (RMSE=1.3215, 0.9190 ตามลำดับ) ส่วนกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอริมาที่แทนข้อมูลด้วยค่าเฉลี่ย (ARIMA(1,0,3)) มีการเคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD ที่วัดได้ ส่งผลให้ค่า RMSE ของกระบวนการอัลฟาทรिमอริมามีค่าน้อยที่สุด (RMSE=0.4078)



ภาพที่ 18 กราฟแสดงค่า BOD ของสถานีที่ 20 สำหรับข้อมูลชุดที่ 2

จากภาพที่ 18 เป็นกราฟแสดงการเปรียบเทียบของค่า BOD ของสถานีที่ 20 ระหว่างค่าจริง ค่า BOD จากเทคนิคการทำให้เรียบที่ใช้วิธีการเฉลี่ยเคลื่อนที่อย่างง่าย เนื่องจากอนุกรมเวลาไม่มี แนวโน้มและฤดูกาล (RMSE=0.7816) ค่า BOD จากตัวแบบ ARIMA(0,0,0) (RMSE=0.5377) และค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิม่าที่แทนข้อมูลด้วยค่าเฉลี่ย (ARIMA(1,0,1)) ซึ่งมีค่า RMSE=0.4899 จะเห็นได้ว่า ลักษณะกราฟของค่า BOD จากกระบวนการอัลฟาทรिमอาร์ิม่ามีการ เคลื่อนไหวใกล้เคียงกับกราฟของค่า BOD จากตัวแบบ ARIMA(0,0,0) และมีค่า RMSE น้อยที่สุด

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 การอภิปรายผลสำหรับข้อมูลชุดที่ 1

จากการวิเคราะห์ข้อมูลของข้อมูลชุดที่ 1 โดยใช้กระบวนการอัลฟาทรिमอาร์มา มีเพียง 4 สถานีที่สามารถหาตัวแบบได้นั้น และอีก 7 สถานี ไม่สามารถหาตัวแบบได้ เนื่องจากผู้ทำวิจัยทำการแต่งขอบข้อมูล 3 ครั้ง แล้วข้อมูลก็ยังไม่มีการแจกแจงแบบปกติ ซึ่งข้อมูลที่ไม่มีการแจกแจงแบบปกติ หรือการกระจายของข้อมูลไม่มีการกระจายแบบโค้งปกติ นั้น เกิดจากช่วงเวลาต่าง ๆ ของกิจกรรมของมนุษย์ การปล่อยน้ำของโรงงานอุตสาหกรรมต่าง ๆ ที่ตั้งอยู่บริเวณสถานีตรวจวัดคุณภาพน้ำซึ่งส่งผลกระทบต่อค่า BOD นั่นคือ ค่า BOD อาจจะมีค่าที่สูงหรือต่ำกว่าปกติได้ และจากตารางที่ 7 พบว่าตัวแบบที่ใช้ในการทำนายค่า BOD ในอนาคต สำหรับสถานีต่าง ๆ มีดังนี้

สำหรับสถานีที่ 3 เมื่อแทนค่า BOD ด้วยค่าเฉลี่ย ได้ตัวแบบ ARIMA(0,0,0) นั่นคือชุดข้อมูล BOD เป็นอนุกรมเวลาที่เป็นสเตชันนารี ($d=0$) และ $p=q=0$ และเมื่อแทนค่า BOD ด้วยค่า BOD ณ $t-2$ ได้ตัวแบบ ARIMA(0,0,0)(0,1,1)₂ นั่นคือชุดข้อมูล BOD เป็นอนุกรมเวลาที่เป็นสเตชันนารี ($d=0$) ที่มีความแปรผันตามฤดูกาล ($D=1$) และ $p=q=P=0$ โดยที่ $L=2$ และเมื่อแทนค่า BOD ด้วยค่ามัธยฐาน พบว่าไม่มีตัวแบบ เพราะไม่มีค่าของ α เนื่องจากผู้ทำวิจัยทำการแต่งขอบข้อมูล 3 ครั้ง แล้วข้อมูลก็ยังไม่มีการแจกแจงแบบปกติ

สำหรับสถานีที่ 17 และ 20 พบว่า ตัวแบบของทั้ง 2 สถานี เมื่อแทนค่า BOD ด้วยค่าทางสถิติทั้ง 3 ค่า มีตัวแบบเดียวกันคือ ARIMA(1,0,3) นั่นคือชุดข้อมูล BOD ของทั้ง 2 สถานีเป็นอนุกรมเวลาที่เป็นสเตชันนารี ($d=0$) และ $p=1, q=3$

สำหรับสถานีที่ 18 เมื่อแทนค่า BOD ด้วยค่าทางสถิติทั้ง 3 ได้ตัวแบบ ARIMA(1,0,1) นั่นคือชุดข้อมูล BOD เป็นอนุกรมเวลาที่เป็นสเตชันนารี ($d=0$) และ $p=q=1$

จากทั้ง 4 สถานีจะพบว่า สถานีที่ 17, 18 และ 20 ไม่ว่าจะแทนค่าทางสถิติด้วยค่าใด ก็ไม่มีผลต่อชุดข้อมูล BOD เพราะมีตัวแบบเดียวกันแต่มีผลในการเลือกตัวแบบที่ดีที่สุด เพียงตัวแบบเดียว ส่วนสถานีที่ 3 การแทนค่าทางสถิติที่ต่างกัน มีผลต่อการสร้างตัวแบบและการเลือกตัวแบบที่ดีที่สุด

เมื่อเปรียบเทียบค่า RMAE ของทั้ง 3 ตัวแบบ พบว่า กระบวนการอัลฟาทรिमอาร์มามีค่าน้อยที่สุดสำหรับแต่ละสถานี และเมื่อพิจารณาค่าเปอร์เซ็นต์ของความคลาดเคลื่อนสัมพัทธ์ของทั้ง 3 ตัวแบบ พบว่า กระบวนการอัลฟาทรिमอาร์มามีค่าน้อยที่สุด นั่นคือ การใช้กระบวนการอัลฟาทรिमอาร์มามีความแม่นยำมากกว่า เทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0)

5.2 การอภิปรายผลสำหรับข้อมูลชุดที่ 2

จากการวิเคราะห์ข้อมูลของข้อมูลชุดที่ 2 โดยใช้กระบวนการอัลฟาพริมาอริมา ทั้ง 11 สถานี สามารถหาตัวแบบได้นั้น จากตารางที่ 12 พบว่าตัวแบบที่ใช้ในการทำนายค่า BOD ในอนาคต สำหรับ สถานีต่าง ๆ มีดังนี้

สำหรับสถานีที่ 1 และ 6 เมื่อแทนค่า BOD ด้วยค่าเฉลี่ยและค่ามัธยฐานได้ตัวแบบเดียวกัน คือ ตัวแบบ ARIMA(1,0,1)(1,1,1)₂ และ ARIMA(0,0,1) ตามลำดับ นั่นคือไม่ว่าจะแทนค่าทางสถิติด้วย ค่าเฉลี่ยหรือค่ามัธยฐาน ก็ไม่มีผลต่อชุดข้อมูล BOD เพราะมีตัวแบบเดียวกันแต่มีผลในการเลือกตัวแบบที่ดีที่สุด เพียงตัวแบบเดียวและเมื่อแทนค่า BOD ด้วยค่า BOD ณ $t-2$ สำหรับสถานีที่ 1 และ 6 ได้ตัวแบบ ARIMA(1,0,1)(0,1,0)₂ และ ARIMA(1,0,1) ตามลำดับ

สำหรับสถานีที่ 3, 8, 10, 15 และ 20 พบว่าตัวแบบของทั้ง 5 สถานี เมื่อแทนค่า BOD ด้วย ค่าทางสถิติทั้ง 3 ค่า ในแต่ละสถานีได้ตัวแบบเดียวกันคือ ตัวแบบ ARIMA(0,0,0), ARIMA(1,0,1)(0,1,0)₂, ARIMA(1,0,1)(0,1,0)₂, ARIMA(1,0,2) และ ARIMA(1,0,1) ตามลำดับ นั่นคือ ทั้ง 5 สถานี ไม่ว่าจะแทนค่าทางสถิติด้วยค่าใด ก็ไม่มีผลต่อชุดข้อมูล BOD เพราะมีตัวแบบเดียวกันแต่ มีผลในการเลือกตัวแบบที่ดีที่สุด เพียงตัวแบบเดียว

สำหรับสถานีที่ 12 เมื่อแทนค่า BOD ด้วยค่ามัธยฐานและค่า BOD ณ $t-2$ ได้ตัวแบบ เดียวกันคือ ตัวแบบ ARIMA(1,0,1)(0,1,0)₂ นั่นคือการแทนค่าชุดข้อมูล BOD ด้วยค่ามัธยฐานและค่า BOD ณ $t-2$ ไม่มีผลต่อชุดข้อมูล BOD เพราะมีตัวแบบเดียวกันแต่มีผลในการเลือกตัวแบบที่ดีที่สุด เพียงตัวแบบเดียว และเมื่อแทนค่า BOD ด้วยค่าเฉลี่ย ได้ตัวแบบ ARIMA(0,0,1)(0,1,0)₂

สำหรับสถานีที่ 16.1 และ 18 เมื่อแทนค่า BOD ด้วยค่าเฉลี่ยและค่า BOD ณ $t-2$ ได้ตัวแบบเดียวกันคือ ตัวแบบ ARIMA(2,0,3) และ ARIMA(1,0,3) ตามลำดับ นั่นคือการแทนค่าชุดข้อมูล BOD ด้วยค่าเฉลี่ยและค่า BOD ณ $t-2$ ไม่มีผลต่อชุดข้อมูล BOD เพราะมีตัวแบบเดียวกันแต่มีผล ในการเลือกตัวแบบที่ดีที่สุด เพียงตัวแบบเดียว ส่วนการแทนค่า BOD ด้วยค่ามัธยฐานสำหรับสถานีที่ 16.1 นั้นพบว่าไม่มีตัวแบบ เพราะไม่มีค่าของ α เนื่องจากผู้ทำวิจัยทำการแต่งขอบข้อมูล 3 ครั้ง แล้ว ข้อมูลก็ยังไม่มีการแจกแจงแบบปกติ และสถานีที่ 18 เมื่อแทนค่า BOD ด้วยค่ามัธยฐานได้ตัวแบบ ARIMA(1,0,1)

สำหรับสถานีที่ 17 เมื่อแทนค่า BOD ด้วยค่าเฉลี่ย ได้ตัวแบบ ARIMA(2,0,2) และเมื่อแทนค่า BOD ด้วยค่ามัธยฐานและค่า BOD ณ $t-2$ พบว่าไม่มีตัวแบบ เพราะไม่มีค่าของ α เนื่องจาก ผู้ทำวิจัยทำการแต่งขอบข้อมูล 3 ครั้ง แล้วข้อมูลก็ยังไม่มีการแจกแจงแบบปกติ นั่นคือการแทนค่า ทางสถิติที่ต่างกัน มีผลต่อการสร้างตัวแบบแต่ไม่มีผลต่อการเลือกตัวแบบที่ดีที่สุด เพราะมีเพียงตัวแบบเดียวเท่านั้น

เมื่อเปรียบเทียบค่า RMAE ของทั้ง 3 ตัวแบบ พบว่า ในแต่ละสถานี กระบวนการอัลฟาทรिम อาริมามีค่าน้อยที่สุด และเมื่อพิจารณาค่าเปอร์เซ็นต์ของความคลาดเคลื่อนสัมพัทธ์ของทั้ง 3 ตัวแบบ พบว่า มีเพียง 5 สถานี ที่เมื่อใช้ กระบวนการอัลฟาทรिम อาริมามีค่าน้อยกว่าอีก 2 ตัวแบบนี้คือ การใช้กระบวนการอัลฟาทรिम อาริมามีความแม่นยำมากกว่า ส่วนอีก 6 สถานี พบว่า ตัวแบบ ARIMA(0,0,0) มีค่าน้อยที่สุด นั่นคือ การทำนายค่า BOD ในอนาคตมีค่าใกล้เคียงกับค่าเฉลี่ยของอนุกรมเวลาในอดีต

สำหรับเทคนิคการทำให้เรียบและการวิเคราะห์อนุกรมเวลาแบบบ็อกซ์เจนคินส์ เป็นตัวแบบที่ใช้ในการวิเคราะห์อนุกรมเวลา เพื่อการพยากรณ์ ซึ่งเป็นเครื่องมือสำคัญในการวางแผนและการตัดสินใจสำหรับเหตุการณ์ในอนาคตได้ ในวิทยานิพนธ์นี้ ผู้ทำวิจัยได้นำ การวิเคราะห์ค่าความเบ้ การแต่งขอบข้อมูลและการแจกแจงแบบปกติมาประยุกต์ใช้กับการวิเคราะห์อนุกรมเวลาแบบบ็อกซ์เจนคินส์ สำหรับชุดข้อมูล BOD ในแต่ละสถานีทั้ง 2 ชุดพบว่า การทำนายค่า BOD ของกระบวนการอัลฟาทรिम อาริมามีการใช้งานได้จริง และให้ผลลัพธ์ในทิศทางที่ดี นั่นคือ เมื่อเปรียบเทียบประสิทธิภาพในการทำนายค่า BOD ของกระบวนการอัลฟาทรिम อาริมามี โดยรวมสามารถทำนายได้ดีกว่าเทคนิคการทำให้เรียบและตัวแบบ ARIMA(0,0,0) ซึ่งงานวิจัยในครั้งนี้เป็นเพียงการเริ่มต้นภายใต้แนวความคิดที่ว่า ข้อมูล BOD ในแต่ละสถานี มีการแจกแจงแบบปกติ แต่น้ำในแม่น้ำเจ้าพระยามีทิศทางการไหลของน้ำและบางสถานีของจุดตรวจวัดคุณภาพมีบ้านเรือน โรงงาน อุตสาหกรรมต่าง ๆ ตั้งอยู่ใกล้ ๆ ซึ่งมีผลต่อค่า BOD ณ เวลาต่าง ๆ ทำให้มีค่า BOD บางส่วนมีค่าที่สูงหรือต่ำกว่าปกติ สำหรับในบางสถานี

การวิเคราะห์อนุกรมเวลาแบบบ็อกซ์เจนคินส์เป็นตัวแบบที่ใช้กับข้อมูลที่มีความสัมพันธ์เป็นเส้นตรง (Linear) เท่านั้น โดยงานที่เกี่ยวข้องกับการทำนายคุณภาพน้ำ ข้อมูลจะมีลักษณะผสมระหว่างความสัมพันธ์ที่เป็นเส้นตรงและไม่เป็นเส้นตรง (Nonlinear) ซึ่งในงานวิจัยนำเครือข่ายประสาทเทียม (Artificial Neural Network) มาใช้เพื่อสร้างตัวแบบสำหรับข้อมูลที่ไม่เป็นเส้นตรง จากเหตุผลดังกล่าวนี้ ในอนาคตทางผู้ศึกษามีความสนใจที่จะสร้างตัวแบบที่ผสมระหว่างการวิเคราะห์อนุกรมเวลาแบบบ็อกซ์เจนคินส์และ ANN สำหรับการทำนายค่า BOD ในอนาคต เพื่อเพิ่มประสิทธิภาพในการทำนายให้ได้ความแม่นยำที่สูงมากขึ้น

รายการอ้างอิง

- [1] M. Milenković, L. Švadlenka, V. Melichar, N. Bojović, and Z. Avramović, "SARIMA modelling approach for railway passenger flow forecasting," *Transport*, pp. 1-8, 2016.
- [2] J. L. Vile, J. W. Gillard, P. R. Harper, and V. A. Knight, "Time-dependent stochastic methods for managing and scheduling Emergency Medical Services," *Operations Research for Health Care*, vol. 8, pp. 42-52, 2016.
- [3] B. Billah, M. L. King, R. D. Snyder, and A. B. Koehler, "Exponential smoothing model selection for forecasting," *International Journal of Forecasting*, vol. 22, pp. 239-247, 2006.
- [4] E. S. Gardner, "Exponential smoothing: The state of the art—Part II," *International Journal of Forecasting*, vol. 22, pp. 637-666, 2006.
- [5] S. Makridakis and M. Hibon, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *International Journal of Forecasting*, vol. 7, pp. 317-330, 1991.
- [6] A. B. K. Rob J. Hyndman, Ralph D. Snyder, Simone Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting* vol. 18, pp. 439–454, 2002.
- [7] S. E. A. a. F. Ahiakpor, "ARIMA (autoregressive integrated moving average) approach to predicting inflation in Ghana," *Journal of Economics and International Finance*, vol. 3, pp. 328-336, 2011.
- [8] C.-F. C. A. R. T. BAILLIE, "Small Sample Bias in Conditional Sumoof-Squares Estimators of Fractionally Integrated ARMA Models," *Empirical Economics* vol. 18, pp. 791-806, 1993.
- [9] C. S. H. JEFFREY L. CALLEN, CLARENCE. Y. KWAN, PATRICKC .Y. YIP, "An Empirical Investigation of the Random Character of Annual Earnings," *Journal of Accounting, Auditing & Finance* vol. 8, pp. 151-162, 1993.
- [10] B. A. Enock Mintah Ampaw, Seth Opoku Larbi, Samuel Lartey, "Time Series Modelling of Rainfall in New Juaben Municipality of the Eastern Region of

- Ghana," *International Journal of Business and Social Science* vol. 8, pp. 116-129, 2013.
- [11] B. A. Jan G. de Gooijer, Ann Gould and Lecily Robinson, "Methods for Determining the Order of an Autoregressive-Moving Average Process: A Survey," *International Statistical Review / Revue Internationale de Statistique*, vol. 53, pp. 301-329, 1985.
- [12] M. Khorasani, M. Ehteshami, H. Ghadimi, and M. Salari, "Simulation and analysis of temporal changes of groundwater depth using time series modeling," *Modeling Earth Systems and Environment*, vol. 2, 2016.
- [13] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, pp. 461-464, 1978.
- [14] C. F. A. A. E. RAFTERY, "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," *The Computer Journal* vol. 41, pp. 578-588, 1998.
- [15] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: a guide for non-statisticians," *Int J Endocrinol Metab*, vol. 10, pp. 486-9, Spring 2012.
- [16] R. Lee, M. Qian, and Y. Shao, "On Rotational Robustness of Shapiro-Wilk Type Tests for Multivariate Normality," *Open Journal of Statistics*, vol. 04, pp. 964-969, 2014.
- [17] J. T. Zofia Hanusz, "<Normality Tests for Statistical Analysis 2.pdf>," *Colloquium Biometricum*, vol. 44, pp. 35-41, 2014.
- [18] Q. Chen, T. Guan, L. Yun, R. Li, and F. Recknagel, "Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: Feasibilities and potentials," *Harmful Algae*, vol. 43, pp. 58-65, 2015.
- [19] F. K. A. a. L. Zhang, "Time series analysis of water quality parameters at Stillaguamish River using order series method," *Stochastic Environmental Research and Risk Assessment*, vol. 29, pp. 227-239, 2015.
- [20] M. Mirzavand and R. Ghazavi, "A Stochastic Modelling Technique for Groundwater Level Forecasting in an Arid Environment Using Time Series Methods," *Water Resources Management*, vol. 29, pp. 1315-1328, 2015.





ประวัติผู้เขียนวิทยานิพนธ์

ชื่อ ชฎาพิมพ์ พตพันธ์เลิศ
ประวัติการศึกษา ปริญญาตรี วิทยาศาสตร์บัณฑิต
มหาวิทยาลัยสงขลานครินทร์ พ.ศ. 2556

ผลงานตีพิมพ์

Photphanloet C., Treeratanajaru W., Cooharojananone N. and Lipikorn R.,
"Biochemical Oxygen Demand Prediction for Chaophraya River Using
Alpha-Trimmed ARIMA Model," 13th International Joint Conference on
Computer Science and Software Engineering 2016 (2016).

