



รายงาน

การวิเคราะห์ข้อมูลภัยแล้งในบริเวณลุ่มแม่น้ำยม

วิชา CS385 : Human-Centered Data Science

จัดทำโดย

- | | | |
|---------------|--------------|------------|
| 1. ชนากานต์ | ศิริกษา | 6109610144 |
| 2. ฐิตารีย์ | ลาเวชกุล | 6109610193 |
| 3. ปัณณรัตน์ | วัฒน์มณีสิริ | 6109680154 |
| 4. ปรีดารัตน์ | วัณภูงกุร | 6109680360 |

โครงการนี้เป็นส่วนหนึ่งของวิชา คพ.385 วิทยาการข้อมูลที่มีมนุษย์เป็นศูนย์กลาง

สาขาวิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2563

บทคัดย่อ

ประเทศไทยประสบกับปัญหาอุทกภัยและภัยแล้งมาโดยตลอด ทั้งนี้มาจากสภาพแวดล้อมทางธรรมชาติจากการกระทำของมนุษย์ และจากการบริหารจัดการด้านการลดผลกระทบจากภัยพิบัติทางธรรมชาติที่ขาดประสิทธิภาพ ส่งผลทำให้เกิดความเดือดร้อนของประชาชนความเสียหายทางเศรษฐกิจ และความไม่มั่นคงในการดำรงชีวิตอยู่อย่างปลอดภัย ดังจะเห็นได้จากภัยแล้งที่เกิดขึ้นในปี พ.ศ.2558 ได้สร้างความเสียหายอย่างรุนแรงในพื้นที่ภาคเหนือของประเทศไทย โดยเฉพาะบริเวณเขตลุ่มแม่น้ำยม

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาสถานการณ์ภัยพิบัติอันเนื่องมาจากน้ำ(ภัยแล้ง) ในลุ่มน้ำยม แนวโน้มของปริมาณฝนรายปี และบริโภคไดของลุ่มน้ำยม มีความผันผวนของปริมาณฝนมากที่สุด เพื่อมากำหนดแนวโน้มการเกิดฝนรายปีของพื้นที่ในลุ่มน้ำยม เพื่อแก้ปัญหาภัยแล้งในลุ่มแม่น้ำยมได้ดียิ่งขึ้น

ดังนั้นทางผู้จัดทำจึงได้ออกแบบโมเดลและวิเคราะห์ข้อมูล โดยใช้ข้อมูลปัจจัยทางสภาพอากาศ ข้อมูลพายุ และข้อมูลป่าไม้ ซึ่งพื้นที่ที่นำมาวิเคราะห์จะแยกพื้นที่ออกเป็น 3 ส่วนคือ ตอนบน ตอนกลาง และตอนล่าง ซึ่งแบ่งตามระดับความสูงของน้ำทะเล เพื่อช่วยให้สามารถนำมาเปรียบเทียบกัน เพื่อใช้ในการวิเคราะห์ข้อมูลให้สามารถเห็นภาพได้จัดเจนมากขึ้น โดยงานวิจัยนี้สร้างขึ้นเพื่อแก้ปัญหาภัยแล้งในลุ่มแม่น้ำยม และปรับปรุงระบบจัดการน้ำให้ดียิ่งขึ้น เพื่อให้ชาวบ้าน เกษตรกรมีน้ำใช้ทำการเกษตร อุปโภค บริโภค อย่างพอดีตลอดทั้งปี

สารบัญ

บทคัดย่อ	2
สารบัญ	3
สารบัญภาพ	5
ภาพรวม	10
ความเป็นมาและความสำคัญของโครงการ	10
วัตถุประสงค์	11
ขอบเขตการศึกษา	11
หลักการและเหตุผลของการศึกษา	11
คำถ้ามที่อยู่ในขอบเขตของการศึกษา	12
บทสรุปการศึกษา	12
แหล่งของเครื่องมือและโปรแกรมที่ใช้ในการศึกษา	12
การคัดเลือกข้อมูล(DataSelection)	14
กระบวนการคัดเลือกข้อมูล	14
ขอบเขต และการกำหนดความต้องการของข้อมูล (Data Requirements)	14
วิธีการสะสมข้อมูล (Data Collection)	19
การเตรียมข้อมูล(DataPreparation)เพื่อทำการวิเคราะห์	20
วิธีที่ใช้ในการวิเคราะห์ข้อมูล(Methods) และ scripts ที่ใช้ในการวิเคราะห์	35
การเลือกโมเดล (Model Selection)	40
การเลือกวิธีสร้างโมเดล และใช้ข้อมูล	45
ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล	52
บทสรุป	72

สรุปผลลัพธ์ที่และตอบคำถามในขอบเขตการศึกษา	72
เงื่อนไข ข้อจำกัดในการวิเคราะห์ที่ควรทราบ	74
คำแนะนำในการต่อยอดการศึกษา	75
บรรณานุกรม	76

สารบัญภาพ

ภาพที่ 1 ภาพแสดงสภาพภูมิประเทศและลำน้ำสาขาในลุ่มแม่น้ำยม	15
ภาพที่ 2 ภาพแสดงปริมาณฝนในแต่ละบริเวณของลุ่มแม่น้ำยม	16
ภาพที่ 3 ภาพแสดงการดึงข้อมูลจาก Google drive เพื่อเลือกสถานีในการวิเคราะห์ข้อมูล	20
ภาพที่ 4.1 ภาพแสดงการ Combining data ระหว่าง ข้อมูลปัจจัยสภาพอากาศ และข้อมูลป่าไม้	21
ภาพที่ 4.2 ภาพแสดงการ Combining data ระหว่าง ข้อมูลปัจจัยสภาพอากาศ และข้อมูลป่าไม้	22
ภาพที่ 5 ภาพแสดงข้อมูลในตารางที่ทำการรวมแล้วของสถานีทต.เขียงม่วน	22
ภาพที่ 6 ภาพแสดงข้อมูลในตารางที่ทำการรวมแล้วของสถานีออบต.วังชี้น	23
ภาพที่ 7 ภาพแสดงข้อมูลในตารางที่ทำการรวมแล้วของสถานีโรงฆ่าสัตว์ ทม.สวรรคโลก	23
ภาพที่ 8 ภาพแสดงข้อมูลในตารางที่ทำการลบคอลัมน์ที่ไม่จำเป็นออกเรียบร้อยแล้วทั้ง 3 สถานี	24
ภาพที่ 9 ภาพแสดงข้อมูลในตารางที่ทำการเช็คเพื่อหา Missing Value	25
ภาพที่ 10 ภาพแสดงข้อมูลในตารางที่ทำการลบข้อมูลปี 2010-2011 เรียบร้อยแล้ว	26
ภาพที่ 11 ภาพแสดงการเปลี่ยนข้อมูล -999.0 เป็น NaN	26
ภาพที่ 12 ภาพแสดงการคำนวณหา Missing Value ที่เหลือของข้อมูล	27
ภาพที่ 13 ภาพแสดงการเปรียบเทียบค่าสถิติก่อนและหลังเติม Missing Value ของข้อมูล	28
ภาพที่ 14 ภาพแสดงวิธีการแปลงข้อมูล เพื่อให้ข้อมูลเป็นรายวัน	29
ภาพที่ 15 ภาพแสดงการแปลงข้อมูล เพื่อให้ข้อมูลเป็นรายวัน	30
ภาพที่ 16 ภาพแสดงการกำหนดฟังก์ชันในการวิเคราะห์ปริมาณน้ำฝนในแต่ละปี	31
ภาพที่ 17 ภาพแสดงการกำหนดฟังก์ชันที่ใช้การวิเคราะห์ปัจจัยต่างๆ	32
ภาพที่ 18 ภาพแสดงการกำหนดฟังก์ชัน Plot Graph	32

ภาพที่ 19 ภาพแสดงการนำฟังก์ชันที่กำหนดมาวิเคราะห์ข้อมูล	33
ภาพที่ 20 ภาพแสดงตัวอย่างของข้อมูลหลังผ่านการวิเคราะห์	33
ภาพที่ 21 ภาพแสดงตัวอย่างการนำข้อมูลมาแสดงผล	33
ภาพที่ 22 ภาพแสดงการ Plot Graph ข้อมูลที่ต้องการวิเคราะห์ของสถานีต่อนบน คือ ข้อมูลปริมาณน้ำฝน สะสม	34
ภาพที่ 23 ภาพแสดงอนุกรมเวลาของสถานีทต. เชียงม่วน	35
ภาพที่ 24 ภาพแสดงอนุกรมเวลาของสถานีอุบต. วังชื่น	37
ภาพที่ 25 ภาพแสดงอนุกรมเวลาของสถานีโรงฆ่าสัตว์ ทม.สวรรคโลก	38
ภาพที่ 26 ภาพแสดงฟังก์ชันการหาความน่าจะเป็น Stationary	39
ภาพที่ 27 ภาพแสดงความน่าจะเป็นที่จะเป็น Stationary สถานีทต. เชียงม่วน	39
ภาพที่ 28 ภาพแสดงความน่าจะเป็นที่จะเป็น Stationary สถานีอุบต. วังชื่น	39
ภาพที่ 29 ภาพแสดงความน่าจะเป็นที่จะเป็น Stationary สถานีโรงฆ่าสัตว์ ทต. สวรรคโลก	40
ภาพที่ 30 ภาพแสดงผลลัพธ์การหาพารามิเตอร์จาก auto_arima	41
ภาพที่ 31 ภาพแสดงการเปรียบเทียบปริมาณฝนสะสมจากการทำนายด้วย SARIMA	41
ภาพที่ 32 ภาพแสดงประสิทธิภาพของโมเดล SARIMA	42
ภาพที่ 33 ภาพแสดงการเปรียบเทียบปริมาณฝนสะสมจากการทำนายด้วย Prophet	42
ภาพที่ 34 ภาพแสดงประสิทธิภาพของโมเดล Prophet	42

ภาพที่ 35 ภาพแสดงโครงสร้างของ LSTM	43
ภาพที่ 36 ภาพแสดงการเปรียบเทียบปริมาณฟันสะสมจากการทำนายด้วย LSTM	43
ภาพที่ 37 ภาพแสดงประสิทธิภาพของโมเดล LSTM	44
ภาพที่ 38 ภาพแสดงการเปรียบเทียบการทำนายปริมาณฟันสะสมจาก 3 โมเดล	44
ภาพที่ 39 ภาพแสดงการเปรียบเทียบ RMSE จาก 3 โมเดล	44
ภาพที่ 40 ภาพแสดง Correalation ระหว่างฟีเจอร์	45
ภาพที่ 41 ภาพแสดงเปรียบเทียบการทำนายปริมาณฟันสะสมด้วย Multivariate values ของโมเดล Prophet	46
ภาพที่ 42 ภาพแสดงค่า RMSE การทำนายปริมาณฟันสะสมด้วย Multivariate values ของโมเดล Prophet	46
ภาพที่ 43 ภาพแสดง Correalation ระหว่างฟีเจอร์หลังการทำ Feature Engineering	47
ภาพที่ 44 ภาพแสดงเปรียบเทียบการทำนายปริมาณฟันสะสมหลังทำ Feature Engineering ของโมเดล Prophet	48
ภาพที่ 45 ภาพแสดงค่า RMSE การทำนายปริมาณฟันสะสมหลังทำ Feature Engineering ของโมเดล Prophet	48
ภาพที่ 46 ภาพแสดงการเปรียบเทียบการทำนายปริมาณฟันสะสมระหว่าง Multivariate valures และหลังทำ Feature Engineering ของโมเดล Prophet	49
ภาพที่ 47 ภาพแสดงค่า RMSE การทำนายปริมาณฟันสะสมโมเดล Prophet ทั้ง 3 ครั้ง	49
ภาพที่ 48 ภาพแสดงฟังก์ชันสร้างโมเดลทำนาย Prophet	50
ภาพที่ 49 ภาพแสดงฟังก์ชัน Data pipeline	51
ภาพที่ 50 ภาพแสดงตัวอย่างการใช้งานฟังก์ชัน Data pipeline	51
ภาพที่ 51 ภาพแสดงปริมาณน้ำฝนสะสม 10 ปีล่ามน้ำยมตอนบน	52

ภาพที่ 52 ภาพแสดงปริมาณฟันสะสมย้อนหลัง 30 ปี	53
ภาพที่ 53 ภาพแสดงอุณหภูมิ 10 ปีล่ามนำ้มต่อนบน	54
ภาพที่ 54 ภาพแสดงอุณหภูมิเฉลี่ยย้อนหลัง 30 ปี	55
ภาพที่ 55 ภาพแสดงปริมาณน้ำฝนสะสม 10 ปีล่ามนำ้มต่อนกลาง	56
ภาพที่ 56 ภาพแสดงอุณหภูมิ 10 ปีล่ามนำ้มต่อนกลาง	57
ภาพที่ 57 ภาพแสดงปริมาณน้ำฝนสะสม 10 ปีล่ามนำ้มต่อนล่าง	58
ภาพที่ 58 ภาพแสดงอุณหภูมิ 10 ปีล่ามนำ้มต่อนล่าง	59
ภาพที่ 59 ภาพแสดงปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนบนจากการทำงานปี ค.ศ.2012-2021	62
ภาพที่ 60 ภาพแสดงแนวโน้มของปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนบนจากการทำงานปี ค.ศ.2012-2021	63
ภาพที่ 61 ภาพแสดงช่วงเวลาปริมาณฟันสะสมที่เกิดในพื้นที่ลุ่มน้ำนำ้มต่อนบน	63
ภาพที่ 62 ภาพแสดงปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนบนจากการทำงานปี ค.ศ.2012-2030	64
ภาพที่ 63 ภาพแสดงแนวโน้มปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนบนจากการทำงานปี ค.ศ.2012-2030	64
ภาพที่ 64 ภาพแสดงปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนกลางจากการทำงานปี ค.ศ.2012-2021	65
ภาพที่ 65 ภาพแสดงแนวโน้มปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนกลางจากการทำงานปี ค.ศ.2012-2021	66
ภาพที่ 66 ภาพแสดงช่วงเวลาปริมาณฟันสะสมที่เกิดในลุ่มน้ำนำ้มต่อนกลาง	66
ภาพที่ 67 ภาพแสดงปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนกลางจากการทำงานปี ค.ศ.2012-2030	67
ภาพที่ 68 ภาพแสดงแนวโน้มปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนกลางจากการทำงานปี ค.ศ.2012-2030	67
ภาพที่ 69 ภาพแสดงปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนล่างจากการทำงานปี ค.ศ.2012-2021	68
ภาพที่ 70 ภาพแสดงแนวโน้มปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนล่างจากการทำงานปี ค.ศ.2012-2021	69
ภาพที่ 71 ภาพแสดงช่วงเวลาปริมาณฟันสะสมที่เกิดขึ้นในลุ่มน้ำนำ้มต่อนล่าง	69
ภาพที่ 72 ภาพแสดงปริมาณฟันสะสมลุ่มน้ำนำ้มต่อนล่างจากการทำงานปี ค.ศ.2012-2030	70

ภาพที่ 73 ภาพแสดงแนวโน้มปริมาณผนังสมคุ์มัยมต้อนร่างจากการทำนายปี ค.ศ.2012-2030

71

1. ภาพรวม

1.1. ความเป็นมาและความสำคัญของโครงงาน

น้ำเป็นสารประกอบที่พบมากถึง 3 ใน 4 ส่วนของพื้นโลก โดยส่วนใหญ่อยู่ในสภาพน้ำเดิมในทะเลและมหาสมุทรประมาณ 97 เปอร์เซ็นต์ เป็นน้ำแข็งตามขั้วโลกประมาณ 2 เปอร์เซ็นต์ และเป็นน้ำจืดตามแม่น้ำลำคลองต่างๆ ประมาณ 1 เปอร์เซ็นต์ ถ้าโลกเราปราศจากน้ำสิ่งมีชีวิตต่างๆ บนโลกก็จะไม่สามารถดำรงชีวิตอยู่ได้ โดยเฉพาะมนุษย์ซึ่งจำเป็นต้องใช้น้ำในการอุปโภคบริโภค และประโยชน์ของน้ำสำหรับมนุษย์นั้น ยังมีอีกมากมายหลายอย่าง ได้แก่ การใช้เพื่อชำระล้างร่างกายและเครื่องนุ่งห่ม ใช้ในการประกอบอาหาร การเกษตร การล้างทำความสะอาดสระน้ำและสาธารณะอื่น ๆ การพักผ่อนหย่อนใจ การอุตสาหกรรม การขับเคลื่อนสิ่งสกปรก การผลิตกระแสไฟฟ้าพลังน้ำ การคมนาคม และการป้องกันอัคคีภัย จึงนับว่าน้ำเป็นทรัพยากรที่มีประโยชน์มากหมายต่อมนุษย์อย่างแท้จริง

โดยแหล่งน้ำของประเทศไทยในปัจจุบันของเรามีอยู่ทั่วทุกที่ แต่ก็ยังมีปัญหาการขาดแคลนน้ำในบางพื้นที่อยู่ อย่างเช่น ภัยแล้งในคุ่น้ำมูล ภัยแล้งในคุ่น้ำเจ้าพระยา ภัยแล้งในคุ่น้ำยม ฯลฯ ซึ่งปัญหาการขาดแคลนน้ำและภัยแล้งส่วนใหญ่ มีสาเหตุมาจากการไม่ตัดตามฤดูกาลและเกิดสภาวะฝนทึ่งช่วงที่ติดต่อ กันยาวนานในพื้นที่คุ่น้ำ ประกอบกับระบบนิเวศของคุ่น้ำที่เปลี่ยนแปลงไป ผลกระทบจากการขยายตัวของชุมชน กิจกรรมทางเศรษฐกิจต่างๆ รวมทั้งการขยายพื้นที่การเกษตรทั้งในเขตชลประทานและนอกเขตชลประทาน ทำให้มีความต้องการน้ำจีดเพิ่มมากขึ้น ซึ่งทำให้เกิดการขาดแคลนน้ำในคุ่น้ำ ขณะที่ศักยภาพในการพัฒนาแหล่งน้ำเก็บกักน้ำต้นทุนของพื้นที่คุ่น้ำมีจำนวนจำกัด อ่างเก็บน้ำขนาดใหญ่ อ่างเก็บน้ำขนาดกลาง และขนาดเล็ก ฝ่ายตามลำน้ำ หน่องบึงธรรมชาติ และสร้างเก็บน้ำที่มีอยู่ยังไม่สามารถเก็บกักน้ำไว้ใช้ได้อย่างเพียงพอต่อความต้องการเพื่อการเกษตร การอุปโภคบริโภคและภาคอุตสาหกรรม แหล่งเก็บกักน้ำและแหล่งน้ำธรรมชาติที่มีอยู่เกิดการตื้นเขินไม่สามารถเก็บกักน้ำได้เต็มประสิทธิภาพ ดังนั้นผู้จัดทำจึงได้ทำการวิเคราะห์ข้อมูลภัยแล้ง โดยเลือกวิเคราะห์ในบริเวณคุ่น้ำ แม่น้ำ ยอม เพื่อให้ทราบถึงแนวโน้มการเกิดฝนและปัญหภัยแล้งอย่างแท้จริงว่าต้นเหตุของปัญหาน้ำเกิดขึ้นจากอะไร และทำการคิดแนวทางการแก้ปัญหาอย่างเหมาะสมและเป็นไปได้มากที่สุด

1.2. วัตถุประสงค์

- 1.2.1. เพื่อนำไปศึกษาและวิเคราะห์แนวโน้มปริมาณน้ำฝนในบริเวณลุ่มแม่น้ำยมที่เกิดขึ้นในระยะ 9 ปีที่ผ่านมา (ค.ศ.2012-2020) และทำการสร้างแบบจำลองเพื่อคาดการณ์ปริมาณน้ำฝนที่จะเกิดขึ้นในปี ค.ศ.2020 โดยใช้ข้อมูลปัจจัยสภาพอากาศร่วมกับปัจจัยที่ทำให้เกิดฝนอื่นๆ
- 1.2.2. เพื่อนำไปศึกษาและวิเคราะห์ความผันผวนของปริมาณน้ำฝนในลุ่มแม่น้ำยมในระยะ 9 ปีที่ผ่านมา(ค.ศ.2012-2020)

1.3. ขอบเขตการศึกษา

การวิเคราะห์ภัยแล้งในลุ่มน้ำยมจะวิเคราะห์จากข้อมูลปริมาณฝนสะสมจากสถานีทด. เชียงม่วน, สถานีอุบต. วังชี้น, และสถานีโรงฆ่าสัตว์ ทต. สوارคโลก ร่วมกับข้อมูลปัจจัยสภาพอากาศ พื้นที่ป่าไม้ในจังหวัดที่อยู่ในบริเวณลุ่มน้ำยม และข้อมูลพายุที่มีผลกระทบต่อประเทศไทย โดยใช้ข้อมูลระหว่าง ปี ค.ศ.2012-2020 และศึกษาปริมาณน้ำฝนของพื้นที่บริเวณลุ่มน้ำยมตั้งแต่ปี ค.ศ. 2012-2030

1.4. หลักการและเหตุผลของการศึกษา

เหตุผลที่เราเลือกศึกษาภัยแล้งในลุ่มแม่น้ำยม เนื่องจากตั้งแต่ปี ค.ศ.2009 ก็ได้เห็นข่าวภัยแล้งในลุ่มแม่น้ำยมมาตลอดจนถึงปัจจุบัน ซึ่งช่วงปี ค.ศ.2017 จ.อุดรธานี ที่อยู่ในพื้นที่ของลุ่มแม่น้ำยมที่เกิดสถานการณ์ภัยแล้งรุนแรงที่สุดในรอบ 60 ปีและช่วงที่ผ่านมาหลายปีก็ได้มีโครงการที่สร้างขึ้นมาเพื่อแก้ปัญหาต่างๆแต่ก็ไม่สามารถแก้ปัญหาได้ตรงจุดจริงๆ ซึ่งภัยแล้งที่เกิดขึ้นนั้นทำให้เกิดผลกระทบต่างๆมากมาย ไม่ว่าจะเป็นชาวบ้านไม่มีน้ำประปาใช้ ข้าวในนาเสียหาย ฯลฯ ทำให้ทางผู้จัดทำได้มีข้อสงสัยขึ้นว่าทำไม่เสียงบโครงการไปทุกปีเพื่อแก้ปัญหานี้แต่กลับไม่มีอะไรที่เปลี่ยนแปลงไปเลย และยังแย่ลงอีกด้วย ทางผู้จัดทำจึงทำการศึกษา เพื่อวิเคราะห์ต้นเหตุของภัยแล้งที่เกิดขึ้นในลุ่มแม่น้ำยม ให้สามารถนำข้อมูลที่วิเคราะห์ได้ไปทำการแก้ปัญหาให้ตรงจุดยิ่งขึ้น โดยใช้หลักการแก้ปัญหาที่ต้นเหตุควบคู่ไปกับแก้ปัญหาที่ปลายเหตุ เนื่องจากการแก้ปัญหาที่ต้นเหตุนั้น ในทางปฏิบัติอาจจะทำได้ยากและใช้เวลานานแต่ก็ใช่ว่าจะทำไม่ได้ และจะใช้การแก้ปัญหาที่ปลายเหตุไปด้วยเพื่อช่วยให้ปัญหาระเทลงระหว่างที่ใช้เวลาแก้ต้นเหตุ และทางผู้จัดทำได้ใช้ข้อมูลปัจจัยทางสภาพอากาศ ข้อมูลพายุ และข้อมูลป่าไม้ มาใช้ประกอบการวิเคราะห์ เนื่องจากปัจจัยที่ทำให้เกิดฝนหลักๆคือ น้ำ แสงอาทิตย์ พื้นที่ป่าไม้ และปัจจัยอื่นๆเช่น

การเกิดพายุในแต่ละพื้นที่ของประเทศไทย ทำให้ผู้จัดทำนำข้อมูลเหล่านี้มาใช้ประกอบการวิเคราะห์ในงานนี้

1.5. คำาณทืออยูในขอบเขตของการศึกษา

1.5.1. แนวโน้มของปริมาณฝนรายปีมีการเปลี่ยนแปลงอย่างไร

1.5.2. บริเวณใดของลุ่มน้ำยามมีความผันผวนของปริมาณฝนมากที่สุด

1.6. บทสรุปการศึกษา

จากการศึกษาพบว่าปริมาณน้ำฝนสะสมในระยะเวลา 9 ปีย้อนหลัง(ค.ศ.2012-2020) แนวโน้มของลุ่มน้ำยามตอนบนและตอนกลางลดลงอย่างต่อเนื่อง แต่ลุ่มน้ำยามตอนล่างมีแนวโน้มที่ปริมาณน้ำฝนสะสมต่อปีสูงขึ้น ซึ่งเมื่อเทียบกับค่าเฉลี่ยของน้ำฝน 30 ปีย้อนหลังนั้นพบว่าปริมาณน้ำฝนสะสมทั้งลุ่มน้ำยามตอนบน ตอนกลาง และตอนล่างนั้นอยู่ในช่วงที่ต่ำกว่าค่าเฉลี่ยของปริมาณน้ำฝนย้อนหลัง 30 ปีของลุ่มน้ำยาม และมีแนวโน้มของปริมาณฝนสะสมลดลงอย่างต่อเนื่องในอีก 10 ปีข้างหน้า(ค.ศ. 2021-2030) ซึ่งทำให้ความผันผวนของปริมาณฝนในช่วงฤดูฝนที่อาจทำให้เกิดภาวะฝนทึ่งช่วงในอนาคต และจากปัจจัยของพายุหมุนเขตร้อน โดยเฉพาะลุ่มน้ำยามตอนล่างที่มีลักษณะภูมิประเทศเป็นที่ราบลุ่มไม่มีแนวเทือกเขาภักดีหรือชัลรอบพายุทำให้มีความผันผวนของปริมาณน้ำฝนมากกว่าบริเวณอื่น มีความรุนแรงมากขึ้นจน ดังนั้นการที่มีปริมาณน้ำฝนลดน้อยลงในแต่ละปี อีกทั้งอุณหภูมิที่เพิ่มสูงสีบเนื่องมาจากการศึกษาอุณหภูมิเฉลี่ยต่อปีของพื้นที่ลุ่มน้ำยามตอนบน ตอนกลาง และตอนล่างเทียบกับอุณหภูมิเฉลี่ย 30 ปีย้อนหลังจากการรวมอุตุนิยมวิทยาพบว่าในบริเวณลุ่มน้ำยามนั้นมีอุณหภูมิเฉลี่ยต่อปีสูงกว่าอุณหภูมิเฉลี่ย 30 ปีย้อนหลังทั้งสิ้น จึงสามารถสรุปได้ว่า ทั้งสองปัจจัยนี้เป็นสาเหตุที่ทำให้ลุ่มน้ำยามประสบภัยแล้งอย่างรุนแรง และต่อเนื่องยาวนานมากขึ้น

1.7. แหล่งของเครื่องมือและโปรแกรมที่ใช้ในการศึกษา

- เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลและสร้างโมเดลการศึกษารังนี้ใช้ Google Colaboratory อยู่ในไฟล์ชื่อ ForecastingModel.ipynb โดยอัปโหลดไว้ที่ Github : https://github.com/punnarat-w/Yomriver_analysis.git

หมายเหตุ สามารถทดลองใช้งานโมเดลได้โดยการ Clone link ดังกล่าวหรือคลิกที่ชื่อไฟล์ใน link และ Open in Colab เพื่อลองใช้งานโมเดล ซึ่งเพื่อให้การศึกษารังนี้ Reproducibility มากที่สุดทางผู้จัดทำได้มีการลำดับขั้นตอนพร้อมอธิบายกระบวนการ

และเหตุผลในการศึกษาแต่ละขั้นตอน รวมถึงสร้าง Data Pipeline พร้อมวิธีการใช้งานอย่างละเอียดเพื่อให้ผู้อ่านสามารถลองปรับเปลี่ยนสถานีที่ใช้ในการวิเคราะห์และคาดการณ์ข้อมูลได้ตามความเหมาะสม

- แหล่งรวมข้อมูลที่ใช้ในการศึกษา

https://drive.google.com/drive/folders/1ezFS9OX_Mmj8V-rSclCaBi8Mqthv2hA-?usp=sharing

2. การคัดเลือกข้อมูล(DataSelection)

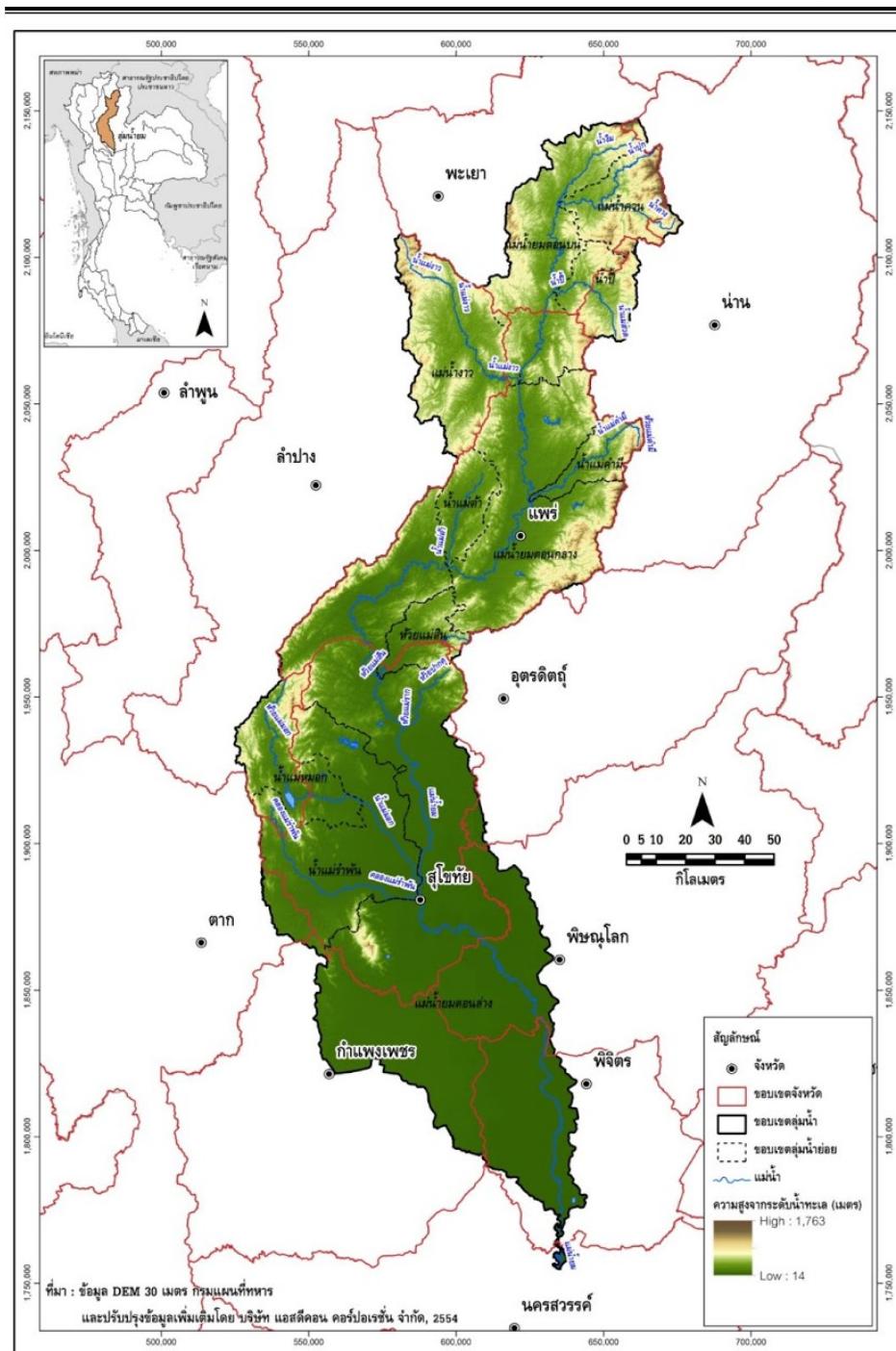
2.1. กระบวนการคัดเลือกข้อมูล

ในการเลือกข้อมูล ทางผู้จัดทำได้ทำการแบ่งลุ่มน้ำยมออกเป็น ลุ่มน้ำยมตอนบน ตอนกลาง และตอนล่าง โดยแบ่งตามระดับความสูงของน้ำทะเล เพื่อให้ข้อมูลนั้นครอบคลุมและเกิดความเอียงของข้อมูลให้น้อยที่สุด ส่วนของการเลือกสถานีในแต่ละตอนนั้นได้ทำการเลือกสถานีที่ติดกับแม่น้ำยมเส้นหลักมากที่สุดครอบคลุมพื้นที่ในเส้นทางของแม่น้ำยม เพื่อให้สามารถวิเคราะห์ปัญหาได้ครอบคลุมพื้นที่บริเวณใกล้เคียงแม่น้ำมากที่สุดเป็นหลัก และข้อมูลปัจจัยสภาพอากาศของแต่ละสถานีจำเป็นต้องมีข้อมูล 9 ปี ข้อมูลไม่เว้นช่วงหายเป็นเวลานาน เพื่อนำมาใช้ในการวิเคราะห์และนำรายได้อย่างแม่นยำมากยิ่งขึ้น

2.2. ขอบเขต และการกำหนดความต้องการของข้อมูล (Data Requirements)

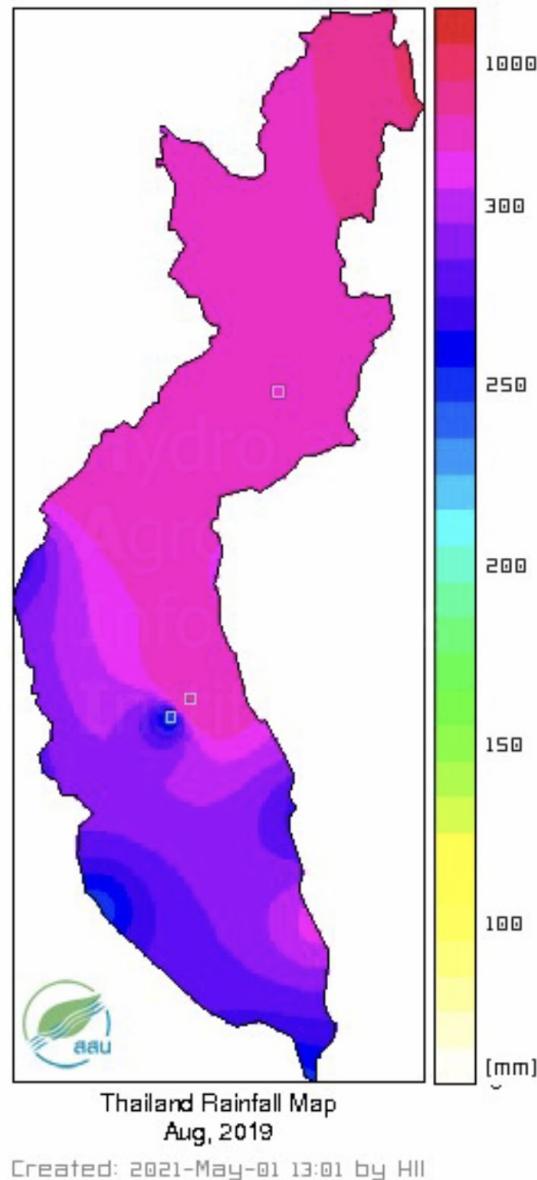
เนื่องจากการศึกษาในครั้งนี้ไม่สามารถศึกษาจากทุกพื้นที่ในบริเวณลุ่มน้ำยมได้ด้วยเหตุผลในเรื่องของการขอข้อมูล และการเก็บข้อมูลจากบางพื้นที่ไม่สามารถเก็บได้อย่างทั่วถึงและสม่ำเสมอ ทางผู้จัดทำจึงต้องทำการเลือกสถานีที่เก็บข้อมูลที่ต้องใช้ในการวิเคราะห์จากบางสถานีจากพื้นที่บางส่วนที่สามารถเป็นตัวแทนอิบยาลีภาระรวมของบริเวณลุ่มน้ำยมได้ โดยทำการศึกษารูปแบบสภาพภูมิประเทศ การกระจายตัวของฝนในบริเวณพื้นที่ของลุ่มน้ำยมว่ามีลักษณะอย่างไร ในแต่ละพื้นที่มีรูปแบบที่เหมือนหรือต่างกันอย่างไร เพื่อเลือกบริเวณพื้นที่หลักมาวิเคราะห์ในการศึกษาครั้งนี้

การเลือกข้อมูลศึกษาในครั้งนี้จะแบ่งพื้นที่เป็น 3 ส่วน คือ ตอนบน ตอนกลาง และตอนล่าง โดยแบ่งตามระดับความสูงจากน้ำทะเล ซึ่งพื้นที่ลุ่มน้ำยมตอนบนจะมีความสูงจากระดับน้ำทะเลมากกว่าบริเวณพื้นที่อื่นในลุ่มน้ำยม และจากสภาพภูมิประเทศคลุ่มน้ำยมตอนบนเป็นพื้นที่ที่เป็นเทือกเขาส่วนใหญ่ พื้นที่ลุ่มน้ำยมตอนกลางจะเป็นพื้นที่ที่เป็นภูเขาในระดับที่ต่ำลงมากจากลุ่มน้ำยมตอนบน และลุ่มน้ำยมตอนล่างพื้นที่ส่วนใหญ่จะเป็นที่ราบลุ่มดังภาพ



ภาพที่ 1 ภาพแสดงสภาพภูมิประเมศและลำน้ำสาขาในลุ่มแม่น้ำยม

ซึ่งปัจจัยความสูงของพื้นที่และพื้นที่ป่าไม้ที่ส่งผลต่อฝนที่ตกในพื้นที่นั้นๆ และจากการศึกษาแผนที่แสดงปริมาณน้ำฝนสะสมจากสถานีสังเคราะห์ (องค์การมหาชน) พบว่าปริมาณน้ำฝนมีความแตกต่างกันตามพื้นที่ซึ่งสอดคล้องกับสภาพภูมิประเทศ ความสูงจากระดับน้ำทะเลและดังภาพปริมาณน้ำฝนสะสมในปี ค.ศ. 2019



ภาพที่ 2 ภาพแสดงปริมาณฝนในแต่ละบริเวณของลุ่มน้ำแม่น้ำยม

ด้วยเหตุนี้จึงทำให้การวิเคราะห์ในครั้งนี้จะแบ่งเป็นลุ่มน้ำยมตอนบน ตอนกลาง และตอนล่าง

- ตอนบน ครอบคลุมพื้นที่ อำเภอสองแคว จังหวัด่น่าน อำเภอเชียงม่วน อำเภออดอกคำใต้ อำเภอปง จังหวัดพะ夷า อำเภอสอง จังหวัดแพร่ อำเภอจาว จังหวัดลำปาง
- ตอนกลาง ครอบคลุมพื้นที่อำเภอเด่นชัย อำเภอเมืองแพร่ อำเภอร้องกวาง อำเภอสอง อำเภอสอง อำเภอสูงเม่น อำเภอหนองม่วงไข่ จังหวัดแพร่
- ตอนล่าง ครอบคลุมพื้นที่อำเภอบึงสามัคคี อำเภอทรายทองวัฒนา อำเภอไทรราม อำเภอพرانกระต่าย อำเภอเมืองกำแพงเพชร อำเภอalanกระปือ จังหวัดกำแพงเพชร อำเภอชุมแสง จังหวัดนครสวรรค์ อำเภอบึงนาราง อำเภอโพทะเล อำเภอโพธิ์ประทับช้าง อำเภอเมืองพิจิตร อำเภอวชิรบารมี อำเภอสะพานทิน อำเภอสามจ่าม จังหวัดพิจิตร อำเภอบางกระทุ่ม อำเภอบางระกำ อำเภอพรหมพิราม อำเภอเมืองพิษณุโลก จังหวัดพิษณุโลก อำเภอสอง อำเภอวังชิ้น จังหวัดแพร่ อำเภออง្ញາග្រลาศ อำเภอคีรีมาศ อำเภอบ้านด่านลานหอย อำเภอเมืองสุโขทัย อำเภอศรีนคร อำเภอศรีสัชนาลัย อำเภอศรีสำโรง อำเภอสารคโลก จังหวัดสุโขทัย อำเภอพิชัย จังหวัดอุตรดิตถ์

ข้อมูลที่ใช้ในการวิเคราะห์

โดยข้อมูลที่ใช้ในการวิเคราะห์ในครั้งนี้จะเป็นข้อมูลในระยะเวลา 9 ปีอยู่ในช่วงปี ค.ศ.2012-2020 เนื่องจากในระยะเวลา 9 ปีมานี้เกิดเหตุการณ์ภัยพิบัติ การเปลี่ยนแปลงทางธรรมชาติต่างๆ ออาทิเช่น ภาวะโลกร้อน อุทกภัยปี ค.ศ.2012 เป็นต้น ซึ่งข้อมูลในช่วงเวลา 9 ปีจะสามารถให้รายละเอียดที่นำไปสู่การวิเคราะห์ภัยแล้งของลุ่มแม่น้ำยมที่ใกล้เคียงกับ สถานการณ์ปัจจุบัน และคาดการณ์สถานการณ์อนาคตได้อย่างมีประสิทธิภาพ

1. ข้อมูลเกี่ยวกับปัจจัยสภาพอากาศ

- ปริมาณน้ำฝนสะสมต่อวัน (หน่วย มิลลิเมตร) โดยความถี่ในการวัดเป็นรายวัน ในช่วงปี ค.ศ.2012-2020
- อุณหภูมิรายวัน(หน่วย องศาเซลเซียส) โดยความถี่ในการวัดเป็นรายวันในช่วงปี ค.ศ.2012-2020

- ความกดอากาศ (หน่วย hPa) โดยความถี่ในการวัดเป็นรายวันในช่วงปี ค.ศ.2012-2020
- ความเร็วลมรายวัน (หน่วย นิอต) โดยความถี่ในการวัดเป็นรายวันในช่วงปี ค.ศ.2012-2020

จากสถานีดังต่อไปนี้

- สถานีทต.เชียงม่วน ต.เชียงม่วน อ.เชียงม่วน จ.พะเยา จากสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน)
- สถานีอุบต.วังชิ้น อ.วังชิ้น จ.แพร่ จากสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน)
- สถานีโรงฆ่าสัตว์ ทม.สوارคโคโลกา อ.สوارคโคโลกา จ.สุโขทัย จากสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน)

2. ข้อมูลพื้นที่ป่าไม้

- พื้นที่ป่าไม้ (หน่วย ไร่) โดยความถี่ในการวัดเป็นรายปีในช่วงปี ค.ศ.2012-2020 จากศูนย์สารสนเทศกรมป่าไม้

จากพื้นที่ในจังหวัดต่อไปนี้

- จังหวัด พะเยา ในลุ่มแม่น้ำயม
 - จังหวัด แพร่ ในลุ่มแม่น้ำယม
 - จังหวัด สุโขทัย ในลุ่มแม่น้ำယม
- สาเหตุที่เลือกช่วงความถี่ในการวัดเป็นรายปีเนื่องจาก การเปลี่ยนแปลงของปริมาณพื้นที่ป่าไม้ในรายวันและรายเดือนมีช่วงของการเปลี่ยนแปลงที่ค่อนข้างต่ำ

3. ข้อมูลพาย

- จำนวนครั้งที่เกิดพายุหมุนเขตร้อนรวมในแต่ละเดือนบริเวณลุ่มแม่น้ำယม (หน่วย ครั้ง) โดยความถี่ในการวัดเป็นรายเดือนในช่วงปี ค.ศ.2012-2020 (รูปแบบไฟล์ .xlsx หรือ .csv) จากกรมอุตุนิยมวิทยา

2.3. วิธีการสะสมข้อมูล (Data Collection)

หลังจากทำการเลือกข้อมูลที่ต้องใช้ในการวิเคราะห์เรียบร้อยแล้ว ทางผู้จัดทำได้ทำการติดต่อขอข้อมูลเกี่ยวกับปัจจัยสภาพอากาศกับทางสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) ตามรายละเอียดที่ได้กล่าวในข้างต้น ส่วนข้อมูลพื้นที่ป่าไม้ทางผู้จัดทำได้ขอรุ่มมาจากเว็บไซต์ ศูนย์สารสนเทศ สำนักแผนงานและสารสนเทศ กรมป่าไม้ และทำการบันทึกข้อมูลในรูปแบบของไฟล์ .csv ซึ่งได้ทำการเลือกพื้นที่ป่าไม้ที่อยู่ในบริเวณลุ่มแม่น้ำยมตอนบน ตอนกลาง และตอนล่าง

3. การเตรียมข้อมูล(DataPreparation)เพื่อทำการวิเคราะห์

ในขั้นตอนนี้ผู้จัดทำได้ทำการใช้ Google Colab ซึ่งเต็มคือ Google Colaboratory เป็นบริการ Software as a Service (SaaS) โฮสต์โปรแกรม Jupyter Notebook บน Cloud จาก Google ในการเขียนโค้ดเพื่อวิเคราะห์ข้อมูลภัยแล้งในบริเวณลุ่มน้ำยม โดยภาษาที่ใช้คือ Python Version 3.6.9 และใช้ Google Drive ซึ่งเป็นโปรแกรมการจัดการเอกสารออนไลน์ของทาง Google หรือพื้นที่เก็บข้อมูลระบบคลาวด์ เพื่อใช้ในการเก็บข้อมูล Raw Data ที่ได้รับมาจาก สถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน)

3.1. Retrieving Data

ในขั้นตอนนี้ทางผู้จัดทำได้ทำการดึงข้อมูลดิบ(Raw Data) ซึ่งเป็นข้อมูลสถานีจัดเก็บน้ำ มาแสดงเพื่อทำการวิเคราะห์และเลือกสถานีน้ำที่ต้องการนำมาใช้ในการวิเคราะห์ข้อมูลภัยแล้งในบริเวณลุ่มน้ำยม ดังภาพ

```
# importing pandas library
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# download description file
!gdown --id 1GJdM5L1SH1Cs95xFef1f3HT7kLT4wd-f

Downloading...
From: https://drive.google.com/uc?id=1GJdM5L1SH1Cs95xFef1f3HT7kLT4wd-f
To: /content/st.csv
100% 20.8k/20.8k [00:00<00:00, 7.97MB/s]

def getData(filename):
    data = pd.read_csv(filename)
    # data['Datetime'] = pd.to_datetime(data['Datetime'])
    # data = data.set_index('Datetime')
    return data

description = getData('st.csv')
description
```

	code	name	lat	long	tambon	amphoe	province	basin	region	station_type
0	BDLH	ทศ.ล้านหลวง	17.008562	99.582995	ล้านหลวง	บ้านด่านล้านหลวง	สุไหทัย	ภาคเหนือ	R	NaN
1	SAMN	ทศ.สามงาม	16.512774	100.208275	สามงาม	สามงาม	พิจิตร	ภาคเหนือ	R	NaN
2	SNGT	อบต.บ้านกลาง	18.482018	100.177530	บ้านกลาง	สอง	แพร่	ภาคเหนือ	R	NaN
3	KPPT	อบจ.กำแพงเพชร	16.410858	99.521990	อ่างทอง	เมืองกำแพงเพชร	กำแพงเพชร	ภาคเหนือ	R	NaN
4	PKTI	ทศ.พวนกระดาย	16.653760	99.575782	พวนกระดาย	พวนกระดาย	กำแพงเพชร	ภาคเหนือ	R	NaN
...
109	VLGE13	ชุมแสงสังคราม	16.858560	100.059642	ชุมแสงสังคราม	บางระกำ	พิษณุโลก	ภาคเหนือ	W	NaN
110	VLGE15	ชุมชนหนองเนินไก	16.434798	99.367315	นาบ่อค่า	เมืองกำแพงเพชร	กำแพงเพชร	ภาคเหนือ	R	NaN
111	VLGE12	นครป่าหมาก	16.623374	100.369210	นครป่าหมาก	บางกระฐุม	พิษณุโลก	ภาคเหนือ	W	NaN
112	TBW014	สะพานข้ามคลองเกเรียงไกร	15.843182	100.211475	บางเตียน	ชุมแสง	นครสวรรค์	ภาคเหนือ	W	NaN
113	TBR013	อบต.ท่าเสา	15.990014	100.221606	ท่าเสา	โพทะเล	พิจิตร	ภาคเหนือ	R	NaN

114 rows × 10 columns

ภาพที่ 3 ภาพแสดงการดึงข้อมูลจาก google drive เพื่อเลือกสถานีในการวิเคราะห์ข้อมูล

3.2. Download data

หลังจากเลือกสถานีที่ต้องการนำมาวิเคราะห์ได้แล้ว ทางผู้จัดทำได้ทำการดึงข้อมูลปัจจัยสภาพอากาศ และข้อมูลป่าไม้ มาใช้ประกอบการวิเคราะห์ข้อมูล โดยใช้การดึงข้อมูลแบบเดียว กับการดึงข้อมูลสถานานน้ำ

3.3. Data Preparation

เมื่อดึงข้อมูลมาเรียบร้อยแล้ว จึงทำการรวมข้อมูลระหว่าง 2 ข้อมูล คือ ข้อมูลปัจจัยสภาพอากาศ และข้อมูลป่าไม้ เข้าด้วยกัน

3.3.1. Combining data

เป็นขั้นตอนการรวมข้อมูลโดยใช้การ Merging data sets ซึ่งการรวมข้อมูลเป็นตารางเดียวกันนั้น ได้ใช้ความสัมพันธ์ของช่วงเวลา 1 ปีในการ mapping ซึ่งเราได้เพิ่มคอลัมน์ที่ชื่อว่า Key เพื่อเข้มความสัมพันธ์ระหว่าง 2 ตารางเข้าด้วยกัน โดย Key ที่ชื่อว่า 11 หมายถึง เป็นข้อมูลของปี 2011 Key ที่ชื่อว่า 12 หมายถึง เป็นข้อมูลของปี 2012 เรื่อยๆไปจนถึง Key ที่ชื่อว่า 20 หมายถึง เป็นข้อมูลของปี 2020 ซึ่งจะสร้างขึ้นทั้งในข้อมูลปัจจัยสภาพอากาศและข้อมูลป่าไม้ โดยข้อมูลที่ทำการรวมนั้น จะแบ่งออกเป็นทั้งหมด 3 ตาราง คือ ตอบบน ตอบกลาง และตอบล่าง เพื่อใช้ในการวิเคราะห์ตามที่ได้บอกจุดประสงค์และเหตุผลไว้ข้างต้น ดังภาพ

```
def insertkey_forest(df):
    # insert column key in nf_df for merge data
    nf_10 = df[df['Year'] == '2010']
    nf_10.insert(0, "Key", "10")
    nf_11 = df[df['Year'] == '2011']
    nf_11.insert(0, "Key", "11")
    nf_12 = df[df['Year'] == '2012']
    nf_12.insert(0, "Key", "12")
    nf_13 = df[df['Year'] == '2013']
    nf_13.insert(0, "Key", "13")
    nf_14 = df[df['Year'] == '2014']
    nf_14.insert(0, "Key", "14")
    nf_15 = df[df['Year'] == '2015']
    nf_15.insert(0, "Key", "15")
    nf_16 = df[df['Year'] == '2016']
    nf_16.insert(0, "Key", "16")
    nf_17 = df[df['Year'] == '2017']
    nf_17.insert(0, "Key", "17")
    nf_18 = df[df['Year'] == '2018']
    nf_18.insert(0, "Key", "18")
    nf_19 = df[df['Year'] == '2019']
    nf_19.insert(0, "Key", "19")
    nf_20 = df[df['Year'] == '2020']
    nf_20.insert(0, "Key", "20")
    df = pd.concat([nf_10, nf_11, nf_12, nf_13, nf_14, nf_15, nf_16, nf_17, nf_18, nf_19, nf_20], sort=False)
    return df
```

ภาพที่ 4.1 ภาพแสดงการ Combining data ระหว่าง ข้อมูลปัจจัยสภาพอากาศ และข้อมูลป่าไม้

```

def prepare_forestdata(forest_df):
    nf_df = forest_df[forest_df['Province'] == 'ต่อไปนี้']
    mf_df = forest_df[forest_df['Province'] == 'ตอนกลาง']
    sf_df = forest_df[forest_df['Province'] == 'ตอนล่าง']
    nf_df = nf_df.astype(str)
    mf_df = mf_df.astype(str)
    sf_df = sf_df.astype(str)
    nf_df = insertkey_forest(nf_df)
    mf_df = insertkey_forest(mf_df)
    sf_df = insertkey_forest(sf_df)
    return nf_df, mf_df, sf_df

def combine_with_forest(north_df, nf_df, mid_df, mf_df, south_df, sf_df):
    data_north = pd.merge(north_df, nf_df, on="Key", how="outer")
    data_mid = pd.merge(mid_df, mf_df, on="Key", how="outer")
    data_south = pd.merge(south_df, sf_df, on="Key", how="outer")
    return data_north, data_mid, data_south

```

ภาพที่ 4.2 ภาพแสดงการ Combining data ระหว่าง ข้อมูลปัจจัยสภาพอากาศ และข้อมูลป่าไม้

จากภาพ หลังจากข้อมูลได้ทำการรวมเป็นตารางเดียวกันแล้ว จะได้ตารางที่มีข้อมูล
ปัจจัยสภาพอากาศ และข้อมูลป่าไม้ ของทั้ง 3 สถานีดังนี้

สถานีทท. เชียงใหม่

	Key	code	date	time	rain1h	temp_out	press	humid	Year	Province	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
0	10	CGMN	2010-01-01	00:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
1	10	CGMN	2010-01-01	01:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
2	10	CGMN	2010-01-01	02:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
3	10	CGMN	2010-01-01	03:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
4	10	CGMN	2010-01-01	04:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
...
96427	20	CGMN	2020-12-31	19:00	0.0	24.8	981.8	61.4	NaN	NaN	NaN	NaN	NaN	NaN
96428	20	CGMN	2020-12-31	20:00	0.0	23.8	982.7	64.0	NaN	NaN	NaN	NaN	NaN	NaN
96429	20	CGMN	2020-12-31	21:00	0.0	23.2	983.6	63.5	NaN	NaN	NaN	NaN	NaN	NaN
96430	20	CGMN	2020-12-31	22:00	0.0	22.7	983.9	64.9	NaN	NaN	NaN	NaN	NaN	NaN
96431	20	CGMN	2020-12-31	23:00	0.0	22.3	984.0	65.9	NaN	NaN	NaN	NaN	NaN	NaN

96432 rows × 14 columns

ภาพที่ 5 ภาพแสดงข้อมูลในตารางที่ทำการรวมแล้วของสถานีทท.เชียงใหม่

สถานีอุบต.วังชิ้น

	Key	code	date	time	rain1h	temp_out	press	humid	Year	Province	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
0	10	WCHN	2010-01-01	00:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
1	10	WCHN	2010-01-01	01:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
2	10	WCHN	2010-01-01	02:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
3	10	WCHN	2010-01-01	03:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
4	10	WCHN	2010-01-01	04:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
...
96427	20	WCHN	2020-12-31	19:00	0.0	22.8	998.5	81.9	NaN	NaN	NaN	NaN	NaN	NaN
96428	20	WCHN	2020-12-31	20:00	0.0	21.3	999.6	86.4	NaN	NaN	NaN	NaN	NaN	NaN
96429	20	WCHN	2020-12-31	21:00	0.0	19.8	1000.3	89.5	NaN	NaN	NaN	NaN	NaN	NaN
96430	20	WCHN	2020-12-31	22:00	0.0	18.6	1000.8	93.0	NaN	NaN	NaN	NaN	NaN	NaN
96431	20	WCHN	2020-12-31	23:00	0.0	17.5	1001.3	93.6	NaN	NaN	NaN	NaN	NaN	NaN

96432 rows × 14 columns

ภาพที่ 6 ภาพแสดงข้อมูลในตารางที่ทำการรวมแล้วของสถานีอุบต.วังชิ้น

สถานีโรงไฟฟ้าสัตหี ทม.สวรรค์โลก

	Key	code	date	time	rain1h	temp_out	press	humid	Year	Province	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
0	10	SWKL	2010-01-01	00:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
1	10	SWKL	2010-01-01	01:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
2	10	SWKL	2010-01-01	02:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
3	10	SWKL	2010-01-01	03:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
4	10	SWKL	2010-01-01	04:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN	NaN	NaN
...
96427	20	SWKL	2020-12-31	19:00	0.0	22.8	1008.1	61.3	NaN	NaN	NaN	NaN	NaN	NaN
96428	20	SWKL	2020-12-31	20:00	0.0	22.7	1009.2	61.8	NaN	NaN	NaN	NaN	NaN	NaN
96429	20	SWKL	2020-12-31	21:00	0.0	21.1	1010.1	66.0	NaN	NaN	NaN	NaN	NaN	NaN
96430	20	SWKL	2020-12-31	22:00	0.0	19.8	1010.7	70.2	NaN	NaN	NaN	NaN	NaN	NaN
96431	20	SWKL	2020-12-31	23:00	0.0	19.3	1010.8	70.6	NaN	NaN	NaN	NaN	NaN	NaN

96432 rows × 14 columns

ภาพที่ 7 ภาพแสดงข้อมูลในตารางที่ทำการรวมแล้วของสถานีโรงไฟฟ้าสัตหี ทม.สวรรค์โลก

3.3.2. Data Cleansing

เมื่อทำการรวมตารางเสร็จเรียบร้อยแล้ว จึงจะเริ่มขั้นตอนการ Cleaning เพื่อทำให้ข้อมูลเกิด Error น้อยที่สุด โดยจะทำการวิเคราะห์ข้อมูลในตารางก่อนเพื่อเข้ามีคอลัมน์ใหม่ บ้างที่เราไม่จำเป็นต้องใช้ในการวิเคราะห์ครั้งนี้ เมื่อเจอแล้วจะทำการลบคอลัมน์นั้นทิ้งไปเพื่อให้ง่ายต่อการวิเคราะห์ข้อมูล

3.3.2.1. Unnecessary data

ซึ่งการลบคอลัมน์นั้น จากการวิเคราะห์แล้วพบว่าคอลัมน์ 'Province' และ 'Year' ออกจากตารางของทั้ง 3 สถานี เนื่องจากไม่จำเป็นต้องใช้ในการวิเคราะห์ข้อมูลแล้ว ดังภาพ

```
def drop_column(df, column):
    df = df.drop(columns = column, axis = 1)
    return df
```

```
north_df = drop_column(north_df, ['Province', 'Year'])
north_df.head()
```

	Key	code	date	time	rain1h	temp_out	press	humid	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
0	10	CGMN	2010-01-01	00:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
1	10	CGMN	2010-01-01	01:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
2	10	CGMN	2010-01-01	02:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
3	10	CGMN	2010-01-01	03:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
4	10	CGMN	2010-01-01	04:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN

```
mid_df = drop_column(mid_df, ['Province', 'Year'])
mid_df.head()
```

	Key	code	date	time	rain1h	temp_out	press	humid	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
0	10	WCHN	2010-01-01	00:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
1	10	WCHN	2010-01-01	01:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
2	10	WCHN	2010-01-01	02:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
3	10	WCHN	2010-01-01	03:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
4	10	WCHN	2010-01-01	04:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN

```
south_df = drop_column(south_df, ['Province', 'Year'])
south_df.head()
```

	Key	code	date	time	rain1h	temp_out	press	humid	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
0	10	SWKL	2010-01-01	00:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
1	10	SWKL	2010-01-01	01:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
2	10	SWKL	2010-01-01	02:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
3	10	SWKL	2010-01-01	03:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
4	10	SWKL	2010-01-01	04:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN

ภาพที่ 8 ภาพแสดงข้อมูลในตารางที่ทำการลบคอลัมน์ที่ไม่จำเป็นออกเรียบร้อยแล้วทั้ง 3 สถานี

3.3.2.2. Error and Outliers

เมื่อทำการลบคอลัมน์ที่ไม่จำเป็นเรียบร้อยแล้ว จึงทำการเช็คข้อมูลอีกรอบเพื่อหา Missing Value ของทั้ง 3 สถานี ยกตัวอย่างดังภาพ

- สถานีต่อนบน: สถานีทต.เชียงม่วน

```
#สถานีต่อนบน north_df
north_df.describe()
```

	rain1h	temp_out	press	humid
count	96432.000000	96432.000000	96432.000000	96432.000000
mean	-352.427800	-349.46950	274.772087	-397.266665
std	477.432796	494.92824	947.046163	532.817777
min	-999.000000	-999.000000	-999.000000	-999.000000
25%	-999.000000	-999.000000	-999.000000	-999.000000
50%	0.000000	23.00000	975.800000	43.000000
75%	0.000000	28.90000	980.000000	83.000000
max	58.800000	44.60000	995.000000	100.000000

```
north_df[ 'code' ].unique()
array(['CGMN'], dtype=object)
```

```
north_df[ 'date' ].unique()
```

```
array(['2010-01-01', '2010-01-02', '2010-01-03', ..., '2020-12-29',
       '2020-12-30', '2020-12-31'], dtype=object)
```

```
north_df[ 'time' ].unique()
```

```
array(['00:00', '01:00', '02:00', '03:00', '04:00', '05:00', '06:00',
       '07:00', '08:00', '09:00', '10:00', '11:00', '12:00', '13:00',
       '14:00', '15:00', '16:00', '17:00', '18:00', '19:00', '20:00',
       '21:00', '22:00', '23:00'], dtype=object)
```

ภาพที่ 9 ภาพแสดงข้อมูลในตารางที่ทำการเช็คเพื่อหา Missing Value

จากการจะแสดงให้เห็นว่ามีค่าที่ผิดปกติคือค่า -999.000000 ทำให้ทางผู้จัดทำทำการวิเคราะห์ข้อมูลในตารางอีกรอบเพื่อหาสาเหตุค่าเหล่านี้ จึงพบว่า ในช่วงปี 2010-2020 ข้อมูลปัจจัยสภาพอากาศเริ่มเก็บข้อมูลตั้งแต่ปี 2012-2020 และข้อมูลป่าไม้เริ่มเก็บข้อมูลตั้งแต่ปี 2013-2019 ทำให้มีข้อมูลที่เป็น -999.000000 อยู่เป็นจำนวนมากมาก ผู้จัดทำจึงทำการลบข้อมูลตั้งแต่ปี 2010-2011 ของข้อมูลปัจจัยสภาพอากาศและข้อมูลป่าไม้ และทำการวิเคราะห์ข้อมูลตั้งแต่ปี 2012-2020 แทน ยกตัวอย่างดังภาพ

```
def drop_range_unavailable(df):
    idx = df['rain1h'].ne(-999).idxmax()
    df = df.drop(range(0, idx))
    return df
```

```
north_df = drop_range_unavailable(north_df)
north_df.head()
```

	Key	code	date	time	rain1h	temp_out	press	humid	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
15278	11	CGMN	2011-09-29	14:00	0.0	34.8	974.0	73.0	NaN	NaN	NaN	NaN
15279	11	CGMN	2011-09-29	15:00	0.0	33.8	974.0	70.0	NaN	NaN	NaN	NaN
15280	11	CGMN	2011-09-29	16:00	-999.0	-999.0	-999.0	-999.0	NaN	NaN	NaN	NaN
15281	11	CGMN	2011-09-29	17:00	0.0	32.8	975.0	72.0	NaN	NaN	NaN	NaN
15282	11	CGMN	2011-09-29	18:00	0.0	28.9	977.0	85.0	NaN	NaN	NaN	NaN

ภาพที่ 10 ภาพแสดงข้อมูลในตารางที่ทำการลบข้อมูลปี 2010-2011 เรียบร้อยแล้ว

จากการ หลังจากทำการลบข้อมูลปี 2010-2011 เรียบร้อยแล้วจะพบว่า แม้ ข้อมูลจะเริ่มมีการบันทึกในช่วงปี 2011 แต่ปริมาณของข้อมูลที่เก็บนั้นมีเพียงไม่กี่เดือน เท่านั้นเมื่อเทียบกับปีอื่นทำให้ในการวิเคราะห์เปรียบเทียบระหว่างปีในภายหลังอาจเกิด ปัญหาได้จึงทำการลบข้อมูลเริ่มต้นที่มีจำนวนเดือนในปี 2011 ออกและเริ่มทำการ วิเคราะห์ข้อมูลที่ปี 2012 เมื่อlobข้อมูลเสร็จสิ้นพบว่ามีข้อมูลที่เป็น -999.0 อยู่เป็น จำนวนมาก จึงทำการแทนที่ข้อมูล -999.0 เป็น NaN โดยใช้การ replace ดังภาพ

```
def replacewithnan(df):
    df = df.replace({-999: np.nan})
    return df
north_df = replacewithnan(north_df)
mid_df = replacewithnan(mid_df)
south_df = replacewithnan(south_df)
display(north_df)
display(mid_df)
display(south_df)
```

	Key	code	date	time	rain1h	temp_out	press	humid	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area
15278	11	CGMN	2011-09-29	14:00	0.0	34.8	974.0	73.0	NaN	NaN	NaN	NaN
15279	11	CGMN	2011-09-29	15:00	0.0	33.8	974.0	70.0	NaN	NaN	NaN	NaN
15280	11	CGMN	2011-09-29	16:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15281	11	CGMN	2011-09-29	17:00	0.0	32.8	975.0	72.0	NaN	NaN	NaN	NaN
15282	11	CGMN	2011-09-29	18:00	0.0	28.9	977.0	85.0	NaN	NaN	NaN	NaN
...
96427	20	CGMN	2020-12-31	19:00	0.0	24.8	981.8	61.4	NaN	NaN	NaN	NaN
96428	20	CGMN	2020-12-31	20:00	0.0	23.8	982.7	64.0	NaN	NaN	NaN	NaN
96429	20	CGMN	2020-12-31	21:00	0.0	23.2	983.6	63.5	NaN	NaN	NaN	NaN
96430	20	CGMN	2020-12-31	22:00	0.0	22.7	983.9	64.9	NaN	NaN	NaN	NaN
96431	20	CGMN	2020-12-31	23:00	0.0	22.3	984.0	65.9	NaN	NaN	NaN	NaN

81154 rows × 12 columns

ภาพที่ 11 ภาพแสดงการเปลี่ยนข้อมูล -999.0 เป็น NaN

หลังจากทำการแทนที่ข้อมูล -999.0 เป็น NaN เรียบร้อยแล้ว ขั้นตอนต่อไปจึงทำการหา Missing Value ที่ยังเหลืออยู่ในข้อมูล ซึ่งจะใช้ .isna() ซึ่งเป็นไลบรารีของ Pandas ในการหา และใช้ .sum() ในการคำนวณผลรวมของ Missing Value ทั้งหมด ในแต่ละคอลัมน์ของข้อมูลแต่ละสถานี ยกตัวอย่างดังภาพ

```
#North
print("สถานีทต. เชียงม่วน")
display(north_df.isna().sum())
#Mid
print("\n\nสถานีอุบต. วังชื่น")
display(mid_df.isna().sum())
#South
print("\n\nสถานีโรงฆ่าสัตว์ ทม. สวรรค์โลก")
display(south_df.isna().sum())
```

สถานีทต. เชียงม่วน

Key	0
code	0
date	0
time	0
rain1h	18748
temp_out	20142
press	19051
humid	27080
Province_Area	19810
Forest_Area	19810
Percent_of_Province_Area	19810
Non_Forest_Area	19810
dtype:	int64

ภาพที่ 12 ภาพแสดงการคำนวณหา Missing Value ที่เหลือของข้อมูล

เมื่อคำนวณหา Missing Value เส้นจะเรียบร้อยแล้ว จะพบว่ามี Missing Value อยู่เป็นจำนวนมาก จึงต้องทำการหาวิธีเติมข้อมูลที่หายไป เนื่องจากไม่สามารถตัดทิ้งได้ โดยเลือกวิธีเติมข้อมูลที่จะทดสอบกับตัวอย่างข้อมูลมา 3 วิธีด้วยกันได้แก่ เติมด้วยค่าศูนย์, เติมด้วยค่าฐานนิยม และเติมด้วยการ Random ข้อมูลในกลุ่มข้อมูล หรือคลัมน์เดียวกันมาใส่แทนเพื่อกำจัด Missing Value ซึ่งหาวิธีที่สามารถคงรักษาการกระจายตัวค่าเฉลี่ย ค่าทางสถิติของข้อมูลได้ใกล้เคียงกับข้อมูลเดิมมากที่สุด โดยไม่ได้นำการเติมข้อมูลด้วยค่าเฉลี่ยมาทดสอบเพื่อป้องกันการเอนเอียงของข้อมูล

Before fill missing values					After fill missing values with zero				
	rain1h	temp_out	press	humid		rain1h	temp_out	press	humid
count	62406.000000	61012.000000	62103.000000	54074.000000	count	81154.000000	81154.000000	81154.000000	81154.000000
mean	0.106663	27.609932	978.881743	74.091486	mean	0.082022	20.757291	749.088066	49.368152
std	1.060647	6.512036	4.300071	21.530116	std	0.931183	13.195658	414.912005	39.107868
min	0.000000	6.400000	964.100000	17.500000	min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	23.500000	976.000000	57.000000	25%	0.000000	9.800000	970.000000	0.000000
50%	0.000000	27.000000	979.000000	79.000000	50%	0.000000	25.000000	977.000000	57.000000
75%	0.000000	32.300000	982.000000	94.000000	75%	0.000000	30.400000	981.000000	88.000000
max	58.800000	44.600000	995.000000	100.000000	max	58.800000	44.600000	995.000000	100.000000

After fill missing values with random					After fill missing values with mode				
	rain1h	temp_out	press	humid		rain1h	temp_out	press	humid
count	81154.000000	81154.000000	81154.000000	81154.000000	count	81154.000000	81154.000000	81154.000000	81154.000000
mean	0.106965	27.611207	978.886415	74.144407	mean	0.082022	26.962161	978.674753	82.736809
std	1.070122	6.512672	4.296931	21.500652	std	0.931183	5.757814	3.780150	21.403567
min	0.000000	6.400000	964.100000	17.500000	min	0.000000	6.400000	964.100000	17.500000
25%	0.000000	23.500000	976.000000	57.000000	25%	0.000000	24.900000	977.000000	68.000000
50%	0.000000	27.000000	979.000000	79.000000	50%	0.000000	25.000000	978.000000	94.000000
75%	0.000000	32.300000	982.000000	94.000000	75%	0.000000	30.400000	981.000000	100.000000
max	58.800000	44.600000	995.000000	100.000000	max	58.800000	44.600000	995.000000	100.000000

ภาพที่ 13 ภาพแสดงการเปรียบเทียบค่าสถิติก่อนและหลังเติม Missing Value ของข้อมูล

จากการสรุปได้ว่าการเติมข้อมูลด้วยค่า Random สามารถให้ค่าทางสถิติของข้อมูลที่ใกล้เคียงกับข้อมูลต้นฉบับมากที่สุดจึงเลือกใช้วิธีนี้ในการเติมข้อมูล หลังจากการทำ Random เพื่อกำจัด Missing Value และ ข้อมูลในตารางที่เป็น NaN ทั้งหมด จะถูกเปลี่ยนเป็นค่าที่ใกล้เคียงกับข้อมูลเดิม และเพื่อเช็คให้แน่ใจว่าข้อมูลในตารางนั้นไม่มี

Missing Value และ จึงทำการใช้ .isna() ซึ่งเป็นไลบรารีของ Pandas ในการหา และใช้ .sum() ในการคำนวณผลรวมของ Missing Value ทั้งหมดในแต่ละคอลัมน์ของข้อมูล แต่ละสถานีอีกรังส์ เพื่อให้แน่ใจว่าข้อมูลในตารางไม่มี Missing Value จริงๆ และให้สามารถนำข้อมูลไปวิเคราะห์ได้สะดวกยิ่งขึ้น เป็นอันเสร็จขั้นตอนการเตรียมข้อมูล (Data Preparation)

3.3.3. Data Transformation

เมื่อทำการเตรียมข้อมูลเรียบร้อยแล้วจะได้ข้อมูลจำนวนฝนตก อุณหภูมิ ความกดอากาศ ความเร็วลม และพื้นที่ป่าไม้เป็นรายชั่วโมง ซึ่งทางผู้จัดทำต้องการ ข้อมูลที่กล่าวไปเป็นรายวัน เพื่อใช้ในการวิเคราะห์ได้ง่ายยิ่งขึ้น โดยทางผู้จัดทำได้ทำการแปลงข้อมูล(Data Transformation) ดังนี้

```
def transformtoOneday(df):
    df_new = df.groupby('date')[['temp_out', 'press', 'humid']].mean()
    df_new['rain1d'] = df.groupby('date')[['rain1h']].sum()
    df_temp = df.groupby('date')[['Key', 'code', 'Province_Area', 'Forest_Area', 'Percent_of_Province_Area', 'Non_Forest_Area']].first()
    dff = pd.concat([df_temp, df_new], axis=1)
    dff['Date'] = df[['date']].unique()
    first_column = dff.pop('Date')
    dff.insert(0, 'Date', first_column)
    dff.reset_index(drop=True, inplace=True)
    return dff

north_df = transformtoOneday(north_df)
mid_df = transformtoOneday(mid_df)
south_df = transformtoOneday(south_df)
```

ภาพที่ 14 ภาพแสดงวิธีการแปลงข้อมูล เพื่อให้ข้อมูลเป็นรายวัน

จากภาพ เมื่อทำการแปลงข้อมูลโดยการนำอุณหภูมิ(temp_out) ความกดอากาศ(press) และความเร็วลม(humid)เฉลี่ยกันต่อวัน โดยใช้ .mean() ใน การหาค่าเฉลี่ย และรวมปริมาณฝนตก(rain1h)ในแต่ละชั่วโมงต่อวัน โดยใช้ .sum() ในการหาผลรวมปริมาณน้ำฝน ในส่วนของพื้นที่ต่างๆ อาทิเช่น พื้นที่ป่าไม้(Forest_Area) ในแต่ละวันมีการเท่ากัน จึงทำการนำค่าแรกมาใช้ในการรวม แต่ละวัน โดยใช้ .first() จากนั้นทำการรวมเมื่อทำการรวมตารางเข้าด้วยกันแล้ว จะได้ข้อมูลของลุ่มแม่น้ำยมของทั้ง 3 สถานีในรูปแบบของรายวันดังภาพที่ 15 เพื่อสามารถนำไปวิเคราะห์ตามวัตถุประสงค์ได้สะดวกยิ่งขึ้น เป็นอันเสร็จสิ้น การแปลงข้อมูล(Data Transformation)

```
#overview data
print("สถานีทด.เชียงม่วน")
display(north_df.head())
print("\nสถานีอุบล.วังชัน")
display(mid_df.head())
print("\nสถานีโรงฆ่าลัต์ ทม.สวรรคโลก")
display(south_df.head())
```

สถานีทด.เชียงม่วน

	Date	Key	code	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area	temp_out	press	humid	rain1d
0	2011-09-29	11	CGMN	19,254,452.30	12,181,835.00	61.32	116.05	29.800000	977.200000	78.660000	0.0
1	2011-09-30	11	CGMN	19,254,452.30	12,177,345.86	61.32	116.03	29.641667	976.708333	71.083333	0.0
2	2011-10-01	11	CGMN	19,254,452.30	12,115,729.16	60.97	117.08	27.079167	977.750000	74.566667	60.2
3	2011-10-02	11	CGMN	19,254,452.30	12,177,345.86	61.32	116.03	28.129167	979.433333	76.333333	7.8
4	2011-10-03	11	CGMN	19,254,452.30	12,177,345.86	61.32	116.03	27.808333	979.458333	79.087500	8.0

<

สถานีอุบล.วังชัน

	Date	Key	code	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area	temp_out	press	humid	rain1d
0	2011-05-03	11	WCHN	4,051,912.64	2,515,083.57	62.07	37.93	24.700000	994.580000	85.000000	0.2
1	2011-05-04	11	WCHN	4,051,912.64	2,600,029.96	64.17	35.83	28.679167	994.408333	79.891667	2.4
2	2011-05-05	11	WCHN	4,051,912.64	2,610,863.30	64.44	35.56	28.754167	995.191667	75.112500	0.6
3	2011-05-06	11	WCHN	4,051,912.64	2,600,029.96	64.17	35.83	28.891667	995.008333	73.387500	0.0
4	2011-05-07	11	WCHN	4,051,912.64	2,623,909.73	64.76	35.24	27.216667	994.637500	80.679167	0.0

<

สถานีโรงฆ่าลัต์ ทม.สวรรคโลก

	Date	Key	code	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area	temp_out	press	humid	rain1d
0	2011-04-11	11	SWKL	84,183,434.55	41,275,108.50	42.42	748.56	30.425000	1001.900000	62.875000	0.0
1	2011-04-12	11	SWKL	84,183,434.55	41,652,816.09	43.38	736.0	30.370833	1001.112500	67.683333	0.2
2	2011-04-13	11	SWKL	84,183,434.55	41,275,108.50	42.42	748.56	29.108333	1000.966667	61.720833	0.2
3	2011-04-14	11	SWKL	84,183,434.55	41,661,864.73	43.46	734.97	29.308333	1002.354167	72.741667	0.0
4	2011-04-15	11	SWKL	84,183,434.55	41,728,212.55	43.51	734.36	30.675000	1002.987500	72.229167	0.0

ภาพที่ 15 ภาพแสดงการแปลงข้อมูล เพื่อให้ข้อมูลเป็นรายวัน

3.3.4. Data Exploration

หลังจากข้อมูลที่ใช้ในการวิเคราะห์ผ่านกระบวนการเตรียมแล้ว ขั้นตอนต่อไปคือ การสำรวจข้อมูล โดยมีหลักการคือการเตรียมข้อมูลอีกรึ่งเพื่อนำข้อมูลมาวิเคราะห์และแสดงผลในรูปแบบต่างๆ (Visualization) เช่น แผนภูมิ หรือกราฟ

3.3.4.1. การวิเคราะห์ปริมาณน้ำฝนสะสมในลุ่มแม่น้ำยมแต่ละปีเพื่อนำไปศึกษาและวิเคราะห์แนวโน้มปริมาณน้ำฝนที่เกิดขึ้นในระยะ 9 ปีที่ผ่านมา (ค.ศ. 2012-2020)

- กำหนดพังก์ชันในการวิเคราะห์ปริมาณน้ำฝนในแต่ละปี

```
# function for calculate annual rainfall
def transformtoPerYear(df) :
    df['year'] = pd.to_datetime(df['Date']).apply(lambda x: '{year}'.format(year=x.year))
    rain = df.copy()
    temp = df.copy()
    df_new = rain.groupby('year')['rainid'].sum()
    df_new_temp = temp.groupby('year')[['temp_out']].mean()
    df_temp = df.groupby('year')[['Key', 'code', 'Province_Area', 'Forest_Area', 'Percent_of_Province_Area', 'Non_Forest_Area']].first()
    dff = pd.merge(df_temp.reset_index(), df_new.reset_index(), on="year")
    dff = pd.merge(dff.reset_index(), df_new_temp.reset_index(), on="year")
    dff = dff.rename(columns = {'rainid': 'rainy'}, inplace = False)
    return dff
```

ภาพที่ 16 ภาพแสดงการกำหนดพังก์ชันในการวิเคราะห์ปริมาณน้ำฝนในแต่ละปี

จากการ หลังจากข้อมูลได้ถูกจัดเตรียมแล้ว ผู้จัดทำได้กำหนดพังก์ชันที่ใช้ในการวิเคราะห์ข้อมูล โดยเปลี่ยนข้อมูลปริมาณน้ำฝนสะสมในแต่ละวันให้เป็นสะสมต่อปี (ค.ศ. 2012-2020)

3.3.4.2. การวิเคราะห์ปริมาณปัจจัยอื่นที่ส่งผลต่อปริมาณน้ำฝนในลุ่มแม่น้ำยมแต่ละปีเพื่อนำไปศึกษาและวิเคราะห์ความผันผวนของปริมาณน้ำฝนในลุ่มแม่น้ำยมในระยะ 9 ปีที่ผ่านมา (ค.ศ. 2012-2020)

- กำหนดพังก์ชันที่ใช้การวิเคราะห์ปัจจัยต่างๆ โดยปัจจัยต่างๆที่กำหนดไว้ คือ ปัจจัยด้านสภาพอากาศ เช่น อุณหภูมิ ความกดอากาศ ความเร็วลมเฉลี่ย , พื้นที่ป่าไม้ และ พื้นที่ชุมชน

```
# function for calculate average pressure
def transformPressoPerYear(df) :
    df['year'] = pd.to_datetime(df['Date']).apply(lambda x: '{year}'.format(year=x.year))
    rain = df.copy()
    press = df.copy()
    df_new = rain.groupby('year')['rainid'].sum()
    df_new_press = press.groupby('year')[['press']].mean()
    df_press = df.groupby('year')[['Key','code','Province_Area','Forest_Area','Percent_of_Province_Area','Non_Forest_Area']].first()
    dff = pd.merge(df_press.reset_index(), df_new.reset_index(), on="year")
    dff = pd.merge(dff.reset_index(), df_new_press.reset_index(), on="year")
    dff = dff.rename(columns = {'rainid': 'rainiy'}, inplace = False)
    return dff

# function for calculate average humidity
def transformHumidtoPerYear(df) :
    df['year'] = pd.to_datetime(df['Date']).apply(lambda x: '{year}'.format(year=x.year))
    rain = df.copy()
    hum = df.copy()
    df_new = rain.groupby('year')['rainid'].sum()
    df_new_hum = hum.groupby('year')[['humid']].mean()
    df_hum = df.groupby('year')[['Key','code','Province_Area','Forest_Area','Percent_of_Province_Area','Non_Forest_Area']].first()
    dff = pd.merge(df_hum.reset_index(), df_new.reset_index(), on="year")
    dff = pd.merge(dff.reset_index(), df_new_hum.reset_index(), on="year")
    dff = dff.rename(columns = {'rainid': 'rainiy'}, inplace = False)
    return dff
```

ภาพที่ 17 ภาพแสดงการกำหนดฟังก์ชันที่ใช้การวิเคราะห์ปัจจัยต่างๆ

จากภาพ จะเป็นการวิเคราะห์ปัจจัยที่กล่าวมาข้างต้นโดยปรับเปลี่ยนจากค่าเฉลี่ยต่อวันเป็นค่าเฉลี่ยต่อปี เพื่อใช้ในการวิเคราะห์ควบคู่ไปกับปริมาณน้ำฝนสะสมต่อปี

3.3.4.3. การกำหนดฟังก์ชัน Plot Graph

```
#function for plotting graph
from matplotlib import pyplot as plt
def plot(x, y, title, x_label, y_label):
    plt.figure(figsize=(10,10))

    # plotting the points
    plt.plot(x.values, y.values, color='green', linestyle='dashed', linewidth = 3,
              marker='o', markerfacecolor='blue', markersize=12)

    # naming the x axis
    plt.xlabel(x_label)
    # naming the y axis
    plt.ylabel(y_label)

    # giving a title to my graph
    plt.title(title)

    # function to show the plot
    plt.show()
```

ภาพที่ 18 ภาพแสดงการกำหนดฟังก์ชัน Plot Graph

จากภาพ จะเป็นการกำหนดฟังก์ชันเพื่อออกแบบรูปแบบกราฟเพื่อให้สามารถเข้าใจการวิเคราะห์ข้อมูลได้ง่ายขึ้น โดยใช้ไลบรารี Matplotlib

3.3.4.4. การนำฟังก์ชันที่กำหนดมาวิเคราะห์ข้อมูลและ Plot Graph

โดยตัวอย่างการนำข้อมูลมาใช้ในการสำรวจและวิเคราะห์จะยกตัวอย่าง

เป็นการใช้งานฟังก์ชัน ร่วมกับข้อมูลของสถานี ทต. เชียงม่วน

```
#ส่วนนี้ดอนบัน
north_df_year = transformtoPerYear(north_df)
north_df_pyear = transformPressstoPerYear(north_df)
north_df_hyear = transformHumidtoPerYear(north_df)
north_yp = pd.merge(north_df_year,north_df_pyear,how='left',on =['index','rain1y','year','Key','code','Province_Area','Forest_Area','Percent_of_Province_Area','Non_Forest_Area'])
north = pd.merge(north_yp,north_df_hyear,how='left',on =['index','rain1y','year','Key','code','Province_Area','Forest_Area','Percent_of_Province_Area','Non_Forest_Area'])
display(north)
```

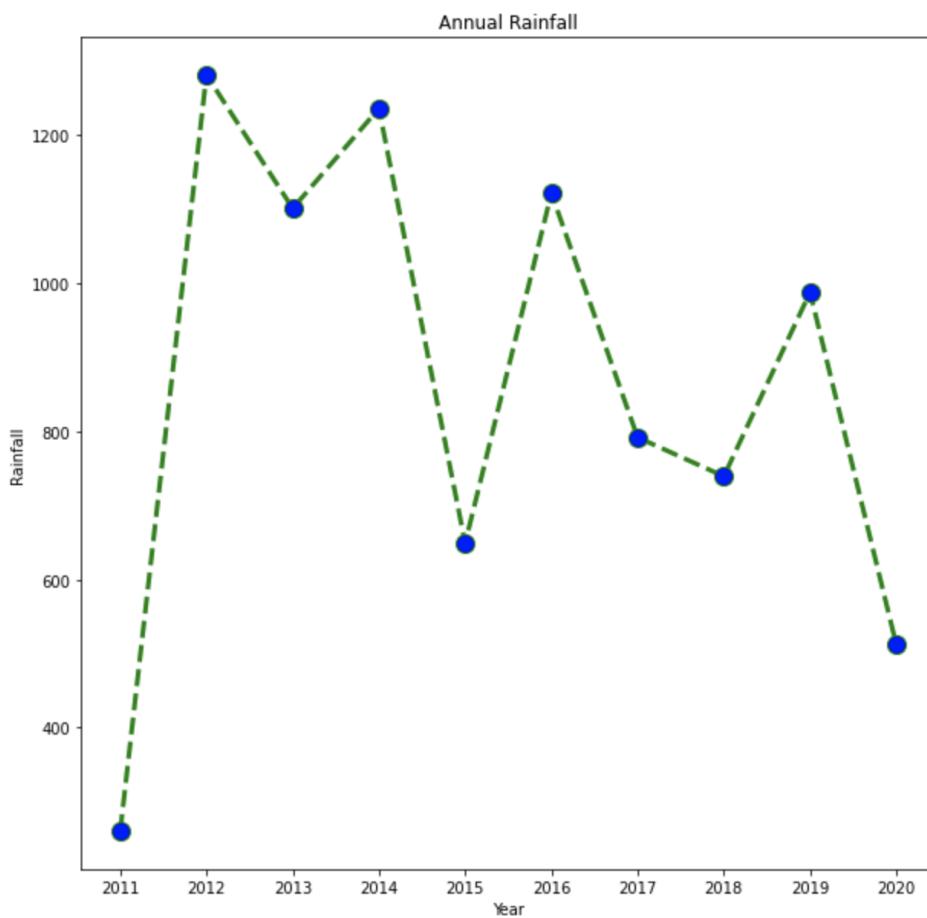
ภาพที่ 19 ภาพแสดงการนำฟังก์ชันที่กำหนดมาวิเคราะห์ข้อมูล

	index	year	Key	code	Province_Area	Forest_Area	Percent_of_Province_Area	Non_Forest_Area	rain1y	temp_out
0	0	2011	11	CGMN	19,254,452.30	12,181,835.00			61.32	116.05 259.6 27.662589
1	1	2012	12	CGMN	19,254,452.30	12,115,729.16			60.97	117.08 1281.8 27.935337
2	2	2013	13	CGMN	19,254,452.30	12,181,835.00			61.32	116.05 1101.4 28.002386
3	3	2014	14	CGMN	19,254,452.30	12,197,151.71			61.41	115.77 1236.4 26.572820
4	4	2015	15	CGMN	19,254,452.30	12,177,345.86			61.32	116.03 648.8 26.696986
5	5	2016	16	CGMN	19,254,452.30	12,177,949.25			61.3	116.11 1121.8 27.980863
6	6	2017	17	CGMN	19,254,452.30	12,115,729.16			60.97	117.08 791.8 27.974292
7	7	2018	18	CGMN	19,254,452.30	12,128,810.16			61.05	116.86 739.2 27.886073
8	8	2019	19	CGMN	19,254,452.30	12,103,104.21			60.92	117.24 988.0 28.011530
9	9	2020	20	CGMN	19,254,452.30	12,177,345.86			61.32	116.03 511.8 27.429440

ภาพที่ 20 ภาพแสดงตัวอย่างของข้อมูลหลังผ่านการวิเคราะห์

```
# plot graph north zone
plot(north_df_year['year'], north_df_year['rain1y'], "Annual Rainfall", "Year", "Rainfall")
print("\n")
```

ภาพที่ 21 ภาพแสดงตัวอย่างการนำข้อมูลมาแสดงผล



ภาพที่ 22 ภาพแสดงการ Plot Graph ข้อมูลที่ต้องการวิเคราะห์ของสถานีตอนบน คือ ข้อมูลปริมาณน้ำฝนสะสม

4. วิธีที่ใช้ในการวิเคราะห์ข้อมูล(Methods) และ scripts ที่ใช้ในการวิเคราะห์

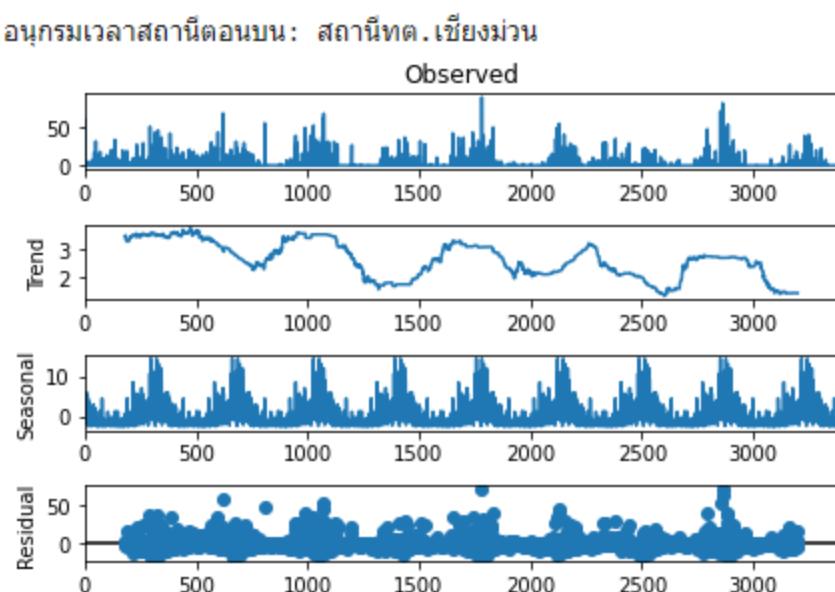
4.1. ตรวจสอบลักษณะของข้อมูล

เนื่องจากอัลกอริทึมที่ใช้สำหรับการทำนายปริมาณฝนสะสมในอนาคตต้อง บางอัลกอริทึมมีข้อจำกัดของข้อมูลที่จะใช้ในการสอนโมเดล เช่น ARIMA อัลกอริทึมที่เป็นที่นิยมในการทำนายข้อมูลในอนาคตแต่มีข้อจำกัดที่อัลกอริทึมนี้ไม่เหมาะสมกับข้อมูลที่เป็น Non-Stationary หรือข้อมูลที่มีลักษณะความแปรปรวนไม่คงที่ขึ้นอยู่กับระยะเวลา เป็นต้น ทำให้ต้องมีการตรวจสอบลักษณะของข้อมูลก่อน

4.1.1. การสำรวจอนุกรมเวลาของข้อมูล (Time Series)

เป็นการสำรวจข้อมูลปริมาณฝนสะสมตามความผันแปรของระยะเวลาว่ามีแนวโน้ม หรือการผันแปรของข้อมูลแบบเดิมๆ ในระยะเวลาที่ผ่านมา หรือข้อมูลปริมาณฝนนั้นไม่มีความผันแปรตามฤดูกาล ซึ่งการสำรวจนี้ทำให้เห็นถึงฤดูกาล แนวโน้มของข้อมูล รวมถึงถูกใช้ในการตัดสินใจเลือกโมเดลที่ใช้ได้ในภายหลัง

4.1.1.1. อนุกรมเวลาสถานีตอนบน: สถานีทศ. เชียงม่วน

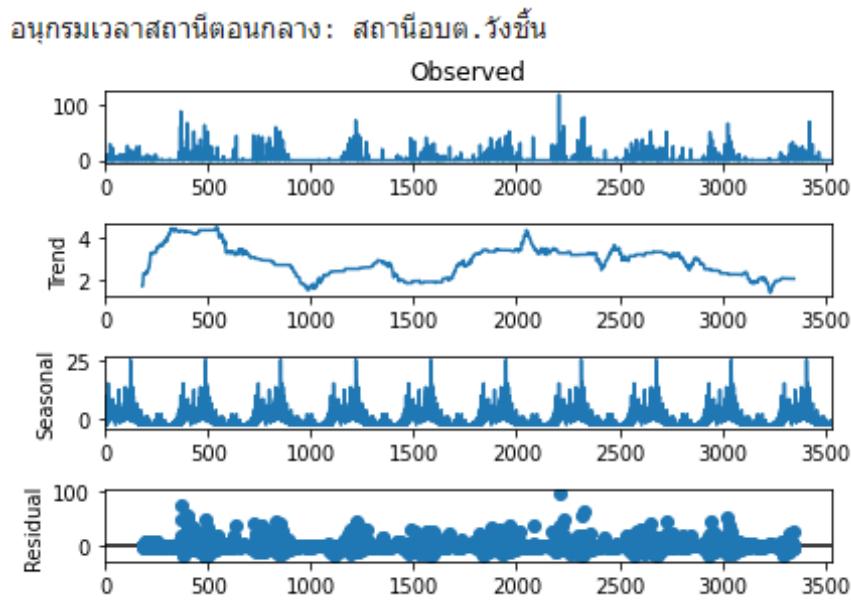


ภาพที่ 23 ภาพแสดงอนุกรมเวลาของสถานีทศ. เชียงม่วน

จากภาพ Observed หรือกราฟด้านบนสุดหมายถึงข้อมูลปริมาณฝนสะสมของสถานีปกติ และเมื่อนำกราฟนั้นมาแยกกองค์ประกอบของอนุกรมเวลาจะได้เป็นกราฟต่างๆ ดังนี้

- Trend แนวโน้มของข้อมูลจะสังเกตได้ว่าแนวโน้มของปริมาณผนังมีเพิ่มขึ้นและลดลง แต่การลดลงจะลดค่อนข้างเยอะกว่าการขึ้น จากการที่แนวเส้นของกราฟลดลงไปทางขวาทำให้ข้อมูลปริมาณน้ำฝนสะสมมีแนวโน้มลดลงตามเวลา
- Seasonal หรือค่าการแปรผันตามฤดูกาลในรอบ 1 ปีจะสังเกตได้ว่าเส้นของกราฟเป็นลักษณะแบบเดียวกันซ้ำ ๆ แสดงว่าปริมาณผนังสะสมของพื้นที่ตอนบนของลุ่มน้ำน้ำยามีความผันแปรแบบเดิมๆตามฤดูกาล
- Residual คือข้อมูลปริมาณน้ำฝนจาก Observed หลังจากแยก Trend และ Seasonal ออก ซึ่งจะเป็นการผันแปรตามปัจจัยต่างๆ ที่เกิดขึ้นในระยะเวลาหนึ่น แต่ไม่ได้เกิดอย่างสม่ำเสมอ

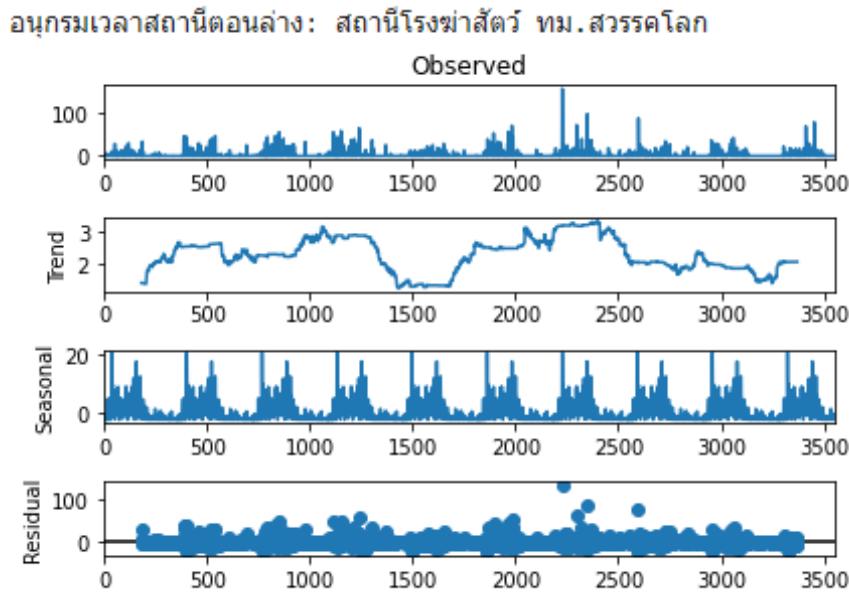
4.1.1.2. อันุกรมเวลาสถานีตอนคลาง: สถานีออบต.วังชี้น



ภาพที่ 24 ภาพแสดงอันุกรมเวลาของสถานีออบต. วังชี้น

- Trend แนวโน้มของปริมาณฝนสะสมของพื้นที่ในลุ่มแม่น้ำยมตอนคลางมีแนวโน้มที่ไม่คงที่ในช่วงปีต้นๆ และมีแนวโน้มลดลงในช่วงปีท้ายๆ
- Seasonal จะสังเกตได้ว่ากราฟมีลักษณะรูปแบบซ้ำๆ ในระยะเวลาที่ผ่านมา แสดงว่าข้อมูลปริมาณฝนสะสมในพื้นที่ลุ่มแม่น้ำยมตอนคลางมีความแปรผันตามฤดูกาล

4.1.1.3. อนุกรมเวลาสถานีต่อน้ำ: สถานีโรงฆ่าสัตว์ ทม.สวรรคโลก



ภาพที่ 25 ภาพแสดงอนุกรมเวลาของสถานีโรงฆ่าสัตว์ ทม.สวรรคโลก

- Trend แนวโน้มของข้อมูลมีลักษณะเพิ่มขึ้นสูงทีละเล็กน้อย และมีช่วงประมาณปีกลาง ในระยะเวลา 10 ปีที่ข้อมูลลดลงอย่างมาก แล้วแนวโน้มค่อยกลับเพิ่มไปในลักษณะเดิม แต่ในช่วงปีหลังมาเนี้ี้ ข้อมูลปริมาณฝนสะสมกลับค่อยๆ ลดลง
- Seasonal ลักษณะของกราฟเหมือนกับกราฟของสถานีในพื้นที่ตอนบน และตอนกลาง ของลุ่มน้ำแม่น้ำยม กล่าวคือพื้นที่ตอนใต้บริเวณลุ่มน้ำแม่น้ำยมมีความแปรผันตามฤดูกาลเช่นเดียวกัน

จากการสำรวจอนุกรมเวลาของข้อมูลปริมาณน้ำฝนสะสมในบริเวณพื้นที่ลุ่มน้ำแม่น้ำยมตอนบน ตอนกลางและตอนล่างมี Trend Seasonal ความผันแปรตามเวลาหรือคือเป็นข้อมูล Non-Stationary และเพื่อประสิทธิภาพของการตรวจสอบก็มีการใช้ adfuller จาก statsmodels.tsa.stattools มาคำนวณค่าทางสถิติหากความผันแปรจะเป็นที่ข้อมูลจะเป็น Stationary หรือไม่

```
#Check Stationarity Data
def ad_test(dataset):
    dfstest = adfuller(dataset, autolag = 'AIC')
    print("1. ADF : ",dfstest[0])
    print("2. P-Value : ", dfstest[1])
    print("3. Num Of Lags : ", dfstest[2])
    print("4. Num Of Observations Used For ADF Regression:", dfstest[3])
    print("5. Critical Values :")
    for key, val in dfstest[4].items():
        print("\t{}: {} - The data is {} stationary with {}% confidence".format(key, val
        , "not" if val<dfstest[0] else "", 100-int(key[:-1])))
```

ภาพที่ 26 ภาพแสดงฟังก์ชันการทดสอบความน่าจะเป็น Stationary

ผลลัพธ์ที่ได้เป็นดังนี้

- สถานีทต. เชียงใหม่

```
1. ADF : -6.917608944197283
2. P-Value : 1.168515472678681e-09
3. Num Of Lags : 26
4. Num Of Observations Used For ADF Regression: 3355
5. Critical Values :
    1%: -3.4323006141089776 - The data is stationary with 99% confidence
    5%: -2.86240186752734 - The data is stationary with 95% confidence
    10%: -2.567228789048532 - The data is stationary with 90% confidence
```

ภาพที่ 27 ภาพแสดงความน่าจะเป็นที่จะเป็น Stationary สถานีทต. เชียงใหม่

- สถานีอุบต. วังชิ้น

```
1. ADF : -5.9552670506424334
2. P-Value : 2.100542165234618e-07
3. Num Of Lags : 30
4. Num Of Observations Used For ADF Regression: 3500
5. Critical Values :
    1%: -3.43221974356695 - The data is stationary with 99% confidence
    5%: -2.8623661465665307 - The data is stationary with 95% confidence
    10%: -2.5672097721632654 - The data is stationary with 90% confidence
```

ภาพที่ 28 ภาพแสดงความน่าจะเป็นที่จะเป็น Stationary สถานีอุบต. วังชิ้น

- สถานีโรงฆ่าสัตว์ ทต.สวรรค์โลก

```

1. ADF : -9.004785003047061
2. P-Value : 6.354784581422205e-15
3. Num Of Lags : 17
4. Num Of Observations Used For ADF Regression: 3535
5. Critical Values :
 1%: -3.432201217785406 - The data is stationary with 99% confidence
 5%: -2.862357963491113 - The data is stationary with 95% confidence
 10%: -2.567205415735552 - The data is stationary with 90% confidence

```

ภาพที่ 29 ภาพแสดงความน่าจะเป็นที่จะเป็น Stationary สถานีโรงฆ่าสัตว์ ทต. สวรรค์โลก

จากการคำนวณข้อมูลจากพื้นที่ทั้ง 3 พบร้าข้อมูลเป็น Stationary ผิดกับสมมติฐานที่ตรวจสอบจากอนุกรมเวลา ก่อนหน้านี้เพียงอย่างเดียว แต่เนื่องด้วยข้อมูลมี Trend และ Seasonal จึงสรุปได้ว่าข้อมูลปริมาณฝนสะสมนั้นเป็น Stationary ที่มีแนวโน้มหรือ Trend ของข้อมูลอยู่

4.2. การเลือกโมเดล (Model Selection)

ในการศึกษาครั้งนี้ได้เลือกอัลกอริทึมเพื่อใช้สำหรับการทำนายอนาคตมา 3 อัลกอริทึม ด้วยกัน ได้แก่ SARIMA (อัลกอริทึม ARIMA สำหรับข้อมูล Stationary ที่มี Trend), Facebook Prophet และ LSTM (Long Short-Term Memory) ซึ่งเป็นอัลกอริทึมที่มีความนิยมใช้ทำนายอนาคตอย่างแพร่หลาย โดยการทดสอบสร้างโมเดลเพื่อหาโมเดลที่เหมาะสมกับข้อมูลมากที่สุด จะใช้การทดสอบแบบ Univariate value หรือก็คือใช้เพียงข้อมูลวันที่และปริมาณฝนสะสม เท่านั้น เพื่อดูความเหมาะสมกับข้อมูลในเบื้องต้นเท่านั้น หากใช้ข้อมูลหลายอย่างจะต้องตั้งค่า โมเดลมากขึ้น เตรียมข้อมูลใหม่มากขึ้น เพราะแต่ละอัลกอริทึมที่เลือกทดสอบใช้ข้อมูลในลักษณะ ที่ต่างกันซึ่งใช้เวลานาน และข้อมูลที่ใช้ในการทดสอบจะเลือกข้อมูลจากสถานีทต. เชียงม่วน ของ พื้นที่ลุ่มแม่น้ำยมตอนบนเป็นตัวแทนในการทดสอบ แบ่งเป็นชุดข้อมูลสำหรับสร้างโมเดล 80% (ประจำปี 2012-2018) ชุดข้อมูลสำหรับทดสอบโมเดล 20% (ประจำปี 2019-2020) เพราะ จากลักษณะของข้อมูล และค่าความน่าจะเป็นที่คำนวณหา Stationary ของข้อมูลจาก 3 สถานี ใกล้เคียงกัน สามารถเลือก 1 สถานีเพื่อเป็นตัวแทนในการทดสอบได้ การวัดผลของการทดสอบ โมเดลวัดผลด้วยค่า RMSE (Root-mean square error) เพื่อหาค่าความคาดเคลื่อนจากข้อมูล จริง

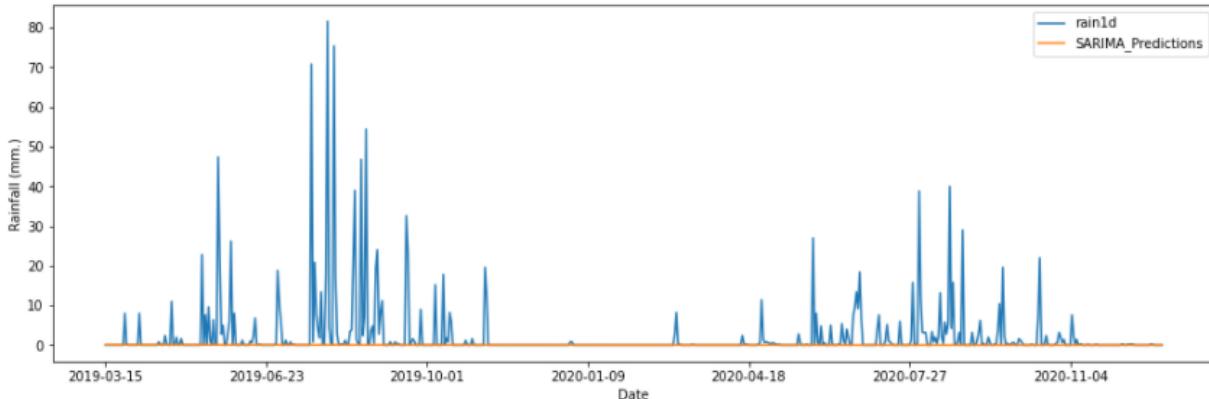
4.2.1. SARIMA

- ใช้ auto_arima ช่วยหาพารามิเตอร์ และโมเดลในตระกูล ARIMA ที่เหมาะสมกับข้อมูลปริมาณฝนสะสม จากผลลัพธ์ได้ โมเดล SARIMAX และชุดพารามิเตอร์ (5,0,1) สำหรับการตั้งค่าโมเดล

Dep. Variable:	y	No. Observations:	3288
Model:	SARIMAX(5, 0, 1)	Log Likelihood	-11006.425
Date:	Thu, 10 Jun 2021	AIC	22028.850
Time:	04:52:45	BIC	22077.634
Sample:	0	HQIC	22046.315
	- 3288		
Covariance Type:	opg		

ภาพที่ 30 ภาพแสดงผลลัพธ์การหาพารามิเตอร์จาก auto_arima

- ผลการทำนายข้อมูลปริมาณฝนสะสมกับข้อมูลชุดทดสอบเป็นดังนี้ หมายเหตุ rain1d คือปริมาณฝนสะสมจริงของข้อมูลชุดทดสอบ แกน y แทนปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุดข้อมูลทดสอบ



ภาพที่ 31 ภาพแสดงการเปรียบเทียบปริมาณฝนสะสมจากการทำนายด้วย SARIMA

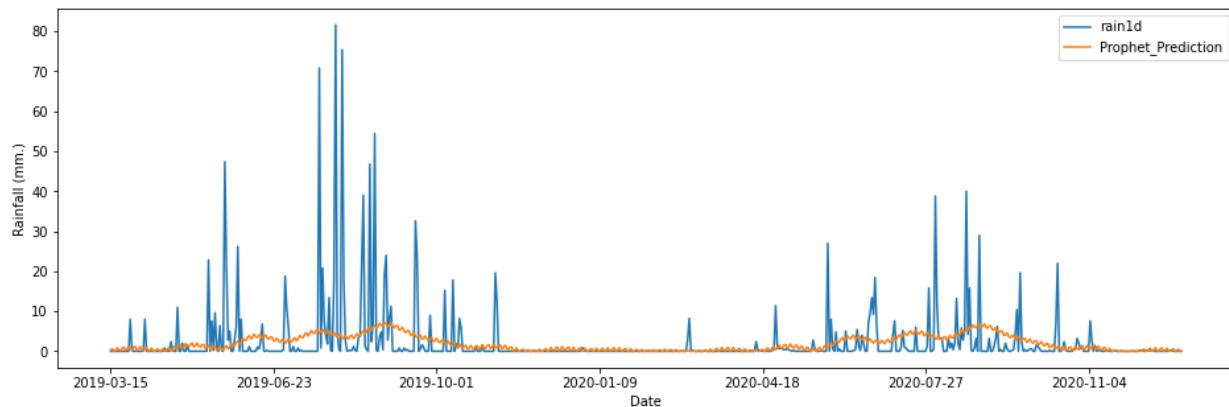
- สรุปการทดสอบประสิทธิภาพโมเดล SARIMA

SARIMA Prediction
RMSE Error: 8.097730242005639

ภาพที่ 32 ภาพแสดงประสิทธิภาพของโมเดล SARIMA

4.2.2. Facebook Prophet

- ผลการทำนายข้อมูลปริมาณฝนสะสมกับข้อมูลชุดทดสอบเป็นดังนี้
หมายเหตุ rain1d คือปริมาณฝนสะสมจริงของข้อมูลชุดทดสอบ แกน y และ
ปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุดข้อมูล
ทดสอบ



ภาพที่ 33 ภาพแสดงการเปรียบเทียบปริมาณฝนสะสมจากการทำนายด้วย Prophet

- สรุปการทดสอบประสิทธิภาพโมเดล Prophet

Prophet Prediction
RMSE Error: 7.54119914621526

ภาพที่ 34 ภาพแสดงประสิทธิภาพของโมเดล Prophet

4.2.3. LSTM (Long Short-Term Memory)

- การสร้างโมเดลทดสอบ LSTM มีโครงสร้างโมเดลทดสอบแบบ Sequential ประกอบไปด้วย 2 เลเยอร์ ในレイอิร์แรกมีนิวรอนจำนวน 10 โหนดดังนี้

```
generator = TimeseriesGenerator(scaled_train_data, scaled_train_data, length=n_input, batch_size=32)

lstm_model = Sequential()
lstm_model.add(LSTM(10, activation='relu', input_shape=(n_input, n_features)))
lstm_model.add(Dense(1))

lstm_model.compile(optimizer='adam', loss='mse')

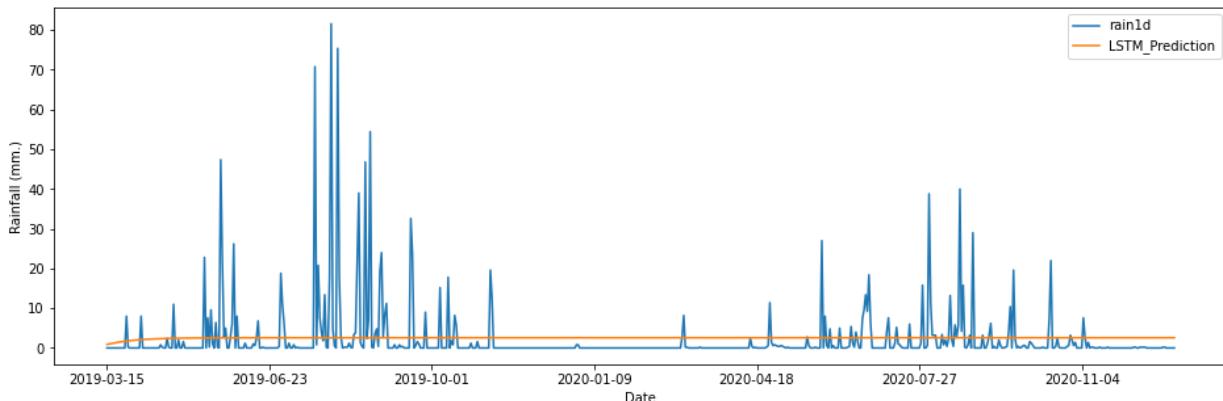
lstm_model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 10)	480
dense (Dense)	(None, 1)	11
<hr/>		
Total params: 491		
Trainable params: 491		
Non-trainable params: 0		

ภาพที่ 35 ภาพแสดงโครงสร้างของ LSTM

- ผลการทำนายข้อมูลปริมาณฝนสะสมกับข้อมูลชุดทดสอบเป็นดังนี้
หมายเหตุ rain1d คือปริมาณฝนสะสมจริงของข้อมูลชุดทดสอบ แกน y และ
ปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุดข้อมูล
ทดสอบ



ภาพที่ 36 ภาพแสดงการเปรียบเทียบปริมาณฝนสะสมจากการทำนายด้วย LSTM

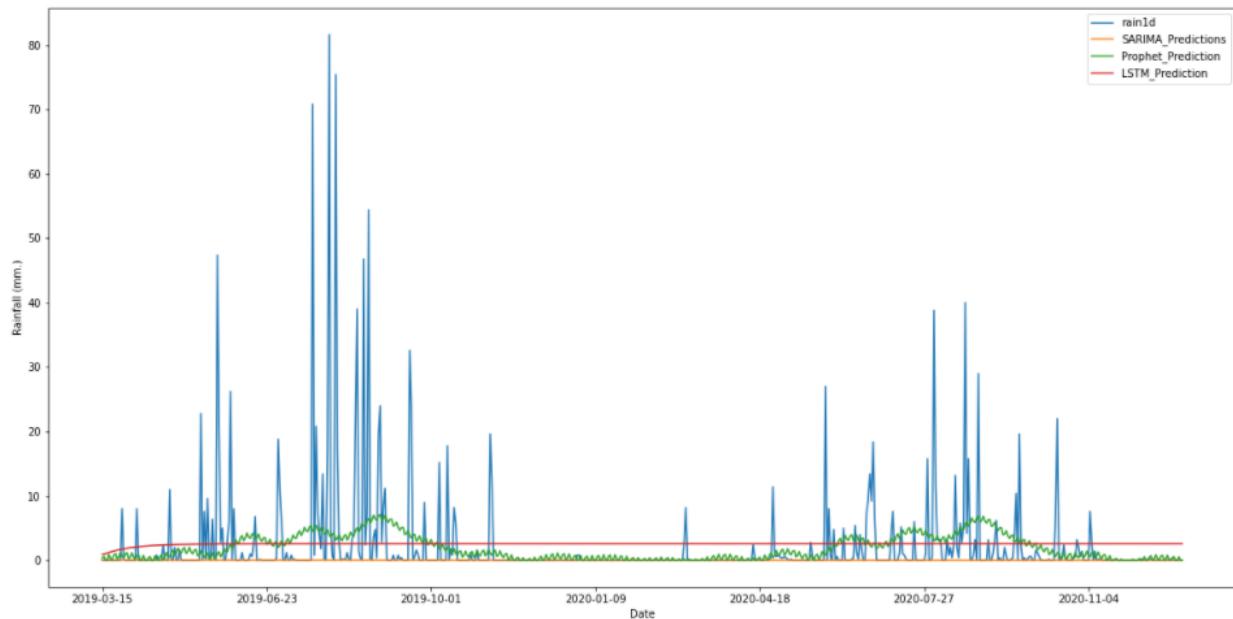
- สรุปการทดสอบประสิทธิภาพโมเดล LSTM

LSTM Prediction
RMSE Error: 7.778129078979698

ภาพที่ 37 ภาพแสดงประสิทธิภาพของโมเดล LSTM

4.2.4. สรุปผลการทดสอบเลือกโมเดล

- ผลการทำนายปริมาณฝนสะสมจากทั้ง 3 โมเดลเป็นดังนี้
หมายเหตุ rain1d คือปริมาณฝนสะสมจริงของข้อมูลชุดทดสอบ แกน y แทน
ปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุดข้อมูล
ทดสอบ



ภาพที่ 38 ภาพแสดงการเปรียบเทียบการทำนายปริมาณฝนสะสมจาก 3 โมเดล

- ค่า RMSE ของทั้ง 3 โมเดลจากการทดสอบเป็นดังนี้

SARIMA RMSE: 8.097730242005639
Prophet RMSE: 7.54119914621526
LSTM RMSE: 7.778129078979698

ภาพที่ 39 ภาพแสดงการเปรียบเทียบ RMSE จาก 3 โมเดล

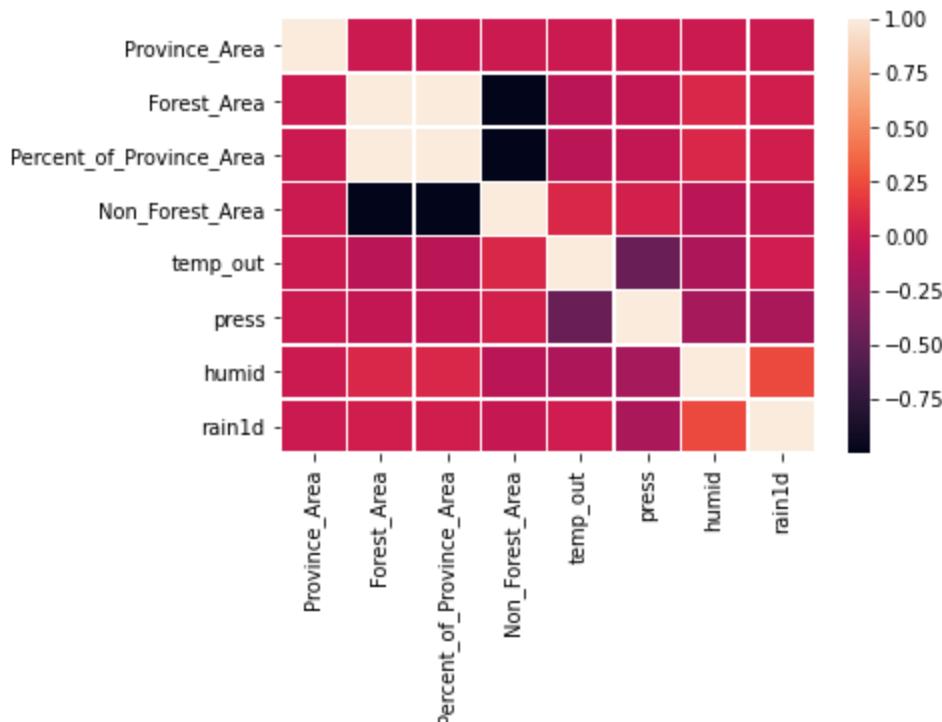
จากการทดสอบจะสังเกตได้ว่าค่า RMSE ของโมเดล Prophet ให้ค่าต่ำที่สุด ซึ่งหมายความว่าการทำนายด้วยโมเดล Prophet จะให้ค่าคลาดเคลื่อนจากความเป็นจริงน้อยที่สุด (+, - 7.541) ส่งผลให้โมเดล Prophet เป็นโมเดลหลักที่ใช้ในการศึกษาครั้งนี้

4.3. การเลือกวิธีสร้างโมเดล และใช้ข้อมูล

หลังจากเลือกโมเดลที่จะใช้ในการศึกษาได้แล้วต่อมาจะเป็นการเลือกวิธีการสร้างโมเดลโดยเลือกจากรูปแบบของข้อมูลที่สอนโมเดล โดยจะทดสอบประสิทธิภาพการทำนายของโมเดลด้วยข้อมูลสถานีทต. เชิงม้วนของพื้นที่ตอนบนของลุ่มน้ำนัยม และวัดผลด้วยค่า RMSE

4.3.1. Multivariate values

Multivariate values หรือก็คือใช้ข้อมูลปัจจัยสภาพอากาศอื่นๆ ได้แก่ อุณหภูมิเฉลี่ยต่อวัน ความกดอากาศเฉลี่ยต่อวัน ความเร็วลมเฉลี่ยต่อวัน และพื้นที่ป่า มาประกอบการทำนายโดยมีการหาค่า Correlation ระหว่างฟีเจอร์เพื่อหาความสำคัญของแต่ละฟีเจอร์ เพื่อตัดฟีเจอร์ที่ไม่จำเป็นต่อการทำนายได้ดังนี้

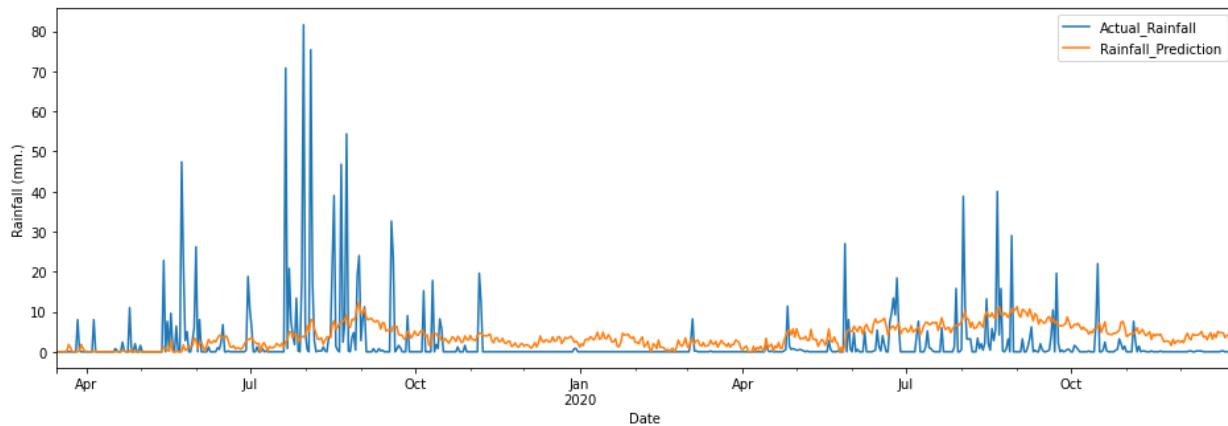


ภาพที่ 40 ภาพแสดง Correlation ระหว่างฟีเจอร์

ค่าที่ได้จะอยู่ระหว่างค่า -1 ถึง 1 โดย -1 คือฟีเจอร์นั้นแปรผันต่อกัน ส่วนค่า 1 คือระหว่าง 2 นั้นแปรผันตรงตามกัน ส่วน 0 คือไม่มีความเกี่ยวข้องกัน หากภาพจะเห็นได้ว่าค่า ความสัมพันธ์ระหว่าง rain1d หรือปริมาณฝนสะสมมีความสัมพันธ์กับฟีเจอร์อื่นๆ ค่อนข้างน้อย ใกล้เคียงกัน (เข้าใกล้ 0) มีเพียง humid หรือความเร็วลมที่สูงกว่าฟีเจอร์อื่นเพียงเล็กน้อย นอกจากนี้มีการตัดฟีเจอร์ค่าเหมือนกันทุกข้อมูลอย่าง Province_Area และข้อมูลซ้ำซ้อนอย่าง Percent_of_Province_Area, Non_Forest_Area ออก

ผลลัพธ์การทำนายของ Prophet ด้วย Multivariate values เป็นดังนี้

หมายเหตุ Actual_Rainfall คือปริมาณฝนสะสมจริงของข้อมูลทดสอบ
แกน y แทนปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุด
ข้อมูลทดสอบ



ภาพที่ 41 ภาพแสดงเปรียบเทียบการทำนายปริมาณฝนสะสมด้วย Multivariate values ของโมเดล Prophet

โดยค่า RMSE ของการทำนายโดยใช้ Multivariate values เป็นดังนี้

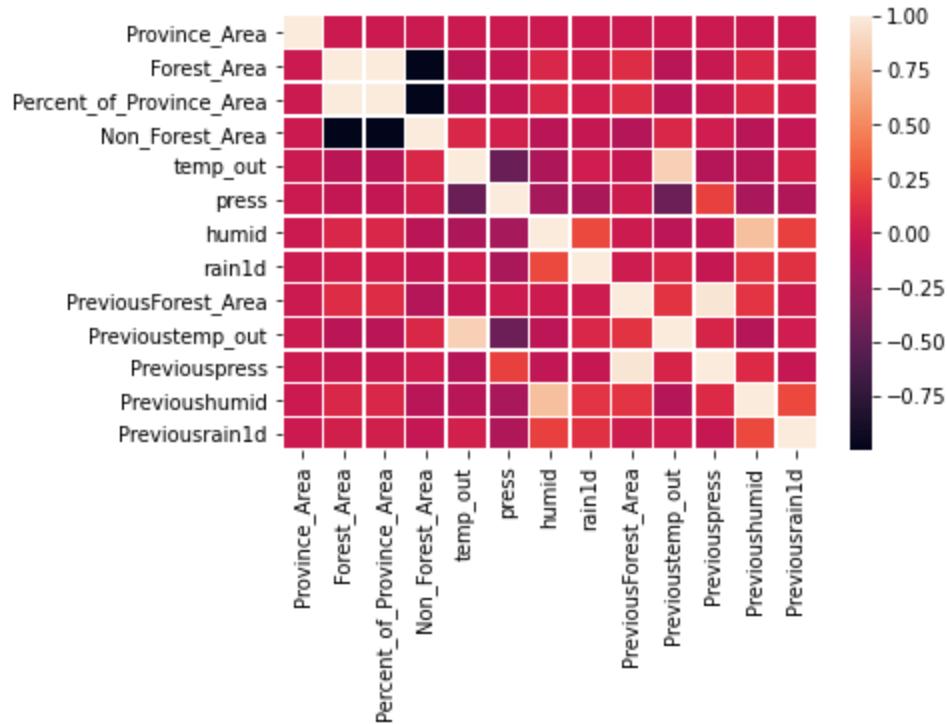
Prophet Prediction
RMSE Error: 7.800005077387621

ภาพที่ 42 ภาพแสดงค่า RMSE การทำนายปริมาณฝนสะสมด้วย Multivariate values ของโมเดล Prophet

ซึ่งจากค่า RMSE พบร้าได้ค่ามากกว่าการสอนโมเดลด้วย Univariate value จึงนำไปสู่
การทดสอบโมเดลด้วยวิธี Feature engineering ใน 4.3.2.

4.3.2. Feature engineering

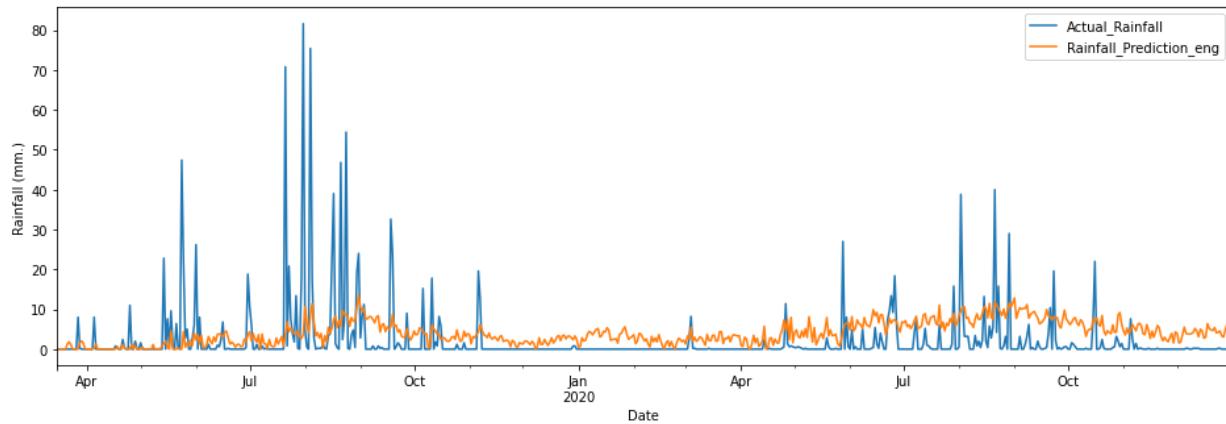
การเพิ่มจำนวนฟีเจอร์ข้อมูลด้วยข้อมูลของวันก่อนหน้า อย่างเช่น ข้อมูลปริมาณฝน สะสมวันที่ 29 กันยายน ค.ศ. 2012 จะมีฟีเจอร์ข้อมูลปริมาณฝนสะสมของวันก่อนหน้า ซึ่งก็คือ วันที่ 28 กันยายน ค.ศ. 2012 ร่วมด้วย โดยมีการหาความสัมพันธ์ Correlation เช่นเดียวกับการทำ Multivariate values ดังนี้



ภาพที่ 43 ภาพแสดง Correalation ระหว่างฟีเจอร์หลังการทำ Feature Engineering

จากภาพจะเห็นได้ว่า ความสัมพันธ์ระหว่าง rain1d หรือปริมาณฝนสะสมมีความสัมพันธ์กับฟีเจอร์อื่นๆ ที่เพิ่มขึ้นมาค่อนข้างน้อย(เข้าใกล้ 0)ใกล้เคียงกับฟีเจอร์อื่นๆ จาก Multivariate values ก่อนหน้านี้ จึงไม่มีการตัดฟีเจอร์ออกนอกจาก Province_Area , Percent_of_Province_Area และ Non_Forest_Area ออก

ผลลัพธ์การทำนายของ Prophet หลังทำ Feature Engineering เป็นดังนี้
 หมายเหตุ Actual_Rainfall คือปริมาณฝนสะสมจริงของข้อมูลชุดทดสอบ
 แกน y แทนปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุด
 ข้อมูลทดสอบ



ภาพที่ 44 ภาพแสดงเปรียบเทียบการทำนายปริมาณฝนสะสมหลังทำ Feature Engineering ของโมเดล Prophet

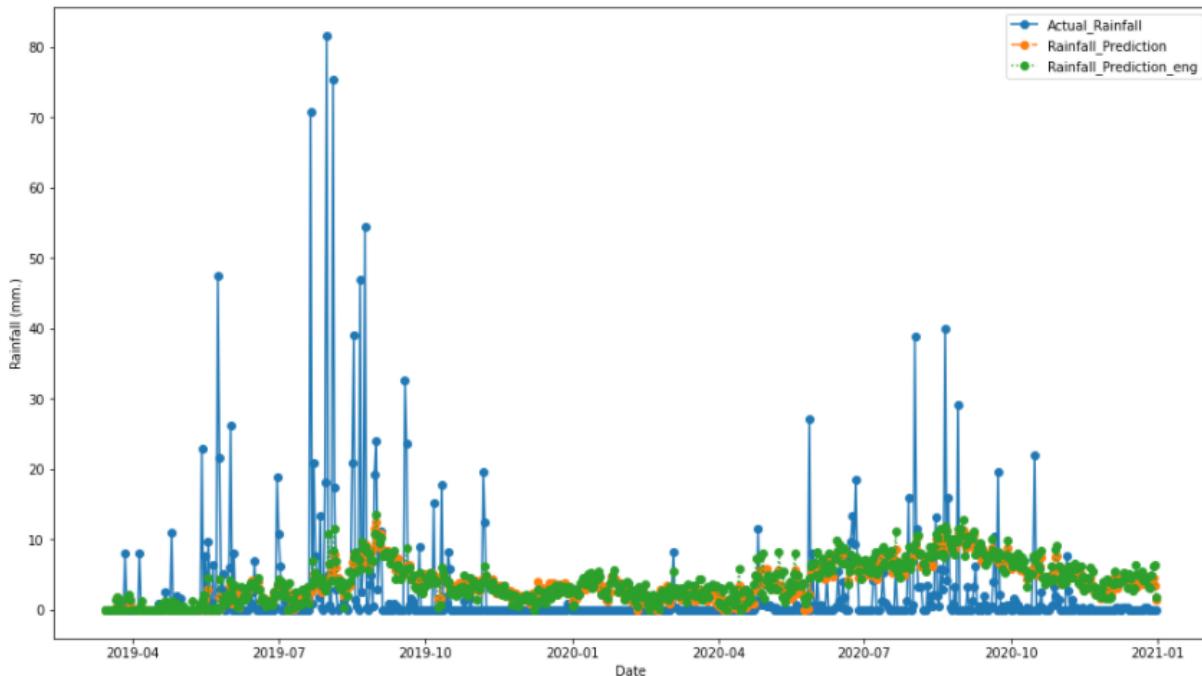
โดยค่า RMSE ของการทำนายหลังทำ Feature Engineering เป็นดังนี้

Prophet Prediction after feature engineering
RMSE Error: 7.872461648526516

ภาพที่ 45 ภาพแสดงค่า RMSE การทำนายปริมาณฝนสะสมหลังทำ Feature Engineering ของโมเดล Prophet

4.3.3. สรุปผลการเลือกวิธีสร้างโมเดล และการใช้ข้อมูล

จากการทดสอบสามารถแสดงภาพเปรียบเทียบการทำนายได้ดังนี้
 หมายเหตุ Actual_Rainfall คือปริมาณฝนสะสมจริงของข้อมูลชุดทดสอบ
 แกน y แทนปริมาณฝนสะสม (หน่วยมิลลิเมตร) แกน x แทน วันเดือนปีของชุด
 ข้อมูลทดสอบ



ภาพที่ 46 ภาพแสดงการเปรียบเทียบการทำนายปริมาณฝนสะสมกว่าง Multivariate values และหลังทำ Feature Engineering ของโมเดล Prophet
 เมื่อเทียบดูค่า RMSE จากการทดสอบของโมเดล Prophet ทั้ง 3 ครั้งได้แก่
 Univariate value, Multivariate values และ Multivariate valuesหลังทำ Feature
 Engineering ได้ดังนี้

```

Prophet Prediction with univariate values
RMSE Error: 7.54119914621526
Prophet Prediction with multivariate values
RMSE Error: 7.800005077387621
Prophet Prediction with multivariate values after feature engineering
RMSE Error: 7.872461648526516
  
```

ภาพที่ 47 ภาพแสดงค่า RMSE การทำนายปริมาณฝนสะสมโมเดล Prophet ทั้ง 3 ครั้ง

สรุปได้ว่าการใช้โมเดล Prophet ทำนายปริมาณน้ำฝนสะสมด้วย Univariate value ให้ประสิทธิภาพการทำนายได้ใกล้เคียงกันค่าจริงมากที่สุด

4.4. Scripts ที่ใช้ในการวิเคราะห์

4.4.1. Facebook Prophet

ส่วนของโปรแกรมที่ใช้สร้างโมเดลเป็นดังนี้ โดยมีพารามิเตอร์ 3 อย่างที่ใช้ในการสร้างได้แก่ df คือข้อมูลจากสถานีที่ต้องการวิเคราะห์, period แทนช่วงเวลาที่ต้องการให้ทำนาย เช่น ทำนาย 10 วัน หรือ 10 ปี ให้ใส่ period = 10, freq แทนความถี่ของช่วงที่ทำนาย ‘D’ แทนช่วงความถี่เป็นวัน ‘M’ แทนช่วงความถี่เป็นเดือน และ ‘Y’ แทนช่วงความถี่เป็นปี

```
def forecast_model(df, period, freq):
    #Preprocessing
    df_new = df[['Date', 'rain1d']]
    df_new = df_new.rename(columns={'Date':'ds', 'rain1d':'y'})
    #Modeling
    m = Prophet()
    m.fit(df_new)
    future = m.make_future_dataframe(periods=period, freq= freq)
    prophet_pred = m.predict(future)
    prophet_pred.loc[(prophet_pred['yhat'] < 0.0), 'yhat']= 0.0
    return prophet_pred, m
```

ภาพที่ 48 ภาพแสดงฟังก์ชันสร้างโมเดลทำนาย Prophet

เนื่องจากเป็น Univariate value จึงใช้ข้อมูลแค่วัน (Date) และปริมาณฝนสะสม (rain1d) เริ่มจากการเปลี่ยนชื่อคอลัมน์เพราะ Prophet จำกัดชื่อคอลัมน์แค่ 2 ชื่อได้แก่ ds แทนคอลัมน์ของวัน และ y แทนคอลัมน์ของข้อมูลที่จะทำนายหรือที่คือปริมาณฝนสะสม หลังจากนั้นนำข้อมูลสอนโมเดล และทำการทำนายตามช่วงเวลาที่ระบุในพารามิเตอร์ และมีการปรับค่าถ้าค่าติดลบให้เท่ากับ 0 เนื่องจากค่าที่ทำนายออกมามีค่าบางส่วนที่ติดลบ ซึ่งในความเป็นจริงปริมาณฝนสะสมไม่มีทางติดลบได้ น้อยที่สุดคือมีค่าเท่ากับ 0 และทำการส่งโมเดลและค่าที่ทำนายคืน

4.4.2. Data Pipeline

ฟังก์ชันของโปรแกรมที่สามารถใช้งานเพื่อทำนายข้อมูลในอนาคตได้เป็นดังนี้ โดยมีการเรียบเรียงกระบวนการทำงานจัดการข้อมูล สร้างโมเดลและแสดงผลภายในฟังก์ชัน มีพารามิเตอร์หลัก 4 อย่างด้วยกันคือ df คือข้อมูลจากสถานีที่ต้องการวิเคราะห์, forest_df แทนข้อมูลพื้นที่ป่าไม้, province แทนพื้นที่ตอนบน ตอนกลาง หรือตอนล่างที่สถานี df ตั้งอยู่, period แทนช่วงเวลาที่ต้องการให้ทำนาย, freq แทนความถี่ของช่วงที่ทำนาย

```
def forecasting_pipeline(df, forest_df, province, period, freq):

    #Combining data
    df = insertkey(df)
    f_df = prepare_forestdata(forest_df, Province = province)
    df = combine_with_forest(df, f_df)

    #Cleansing data
    df = drop_column(df, ['Province','Year'])
    df = drop_range_unavailable(df)
    df = replacewithnan(df)
    df = fill_missing_value(df)

    #Transform data
    df = transformtoOneday(df)

    #Modeling
    forecast, model = forecast_model(df, period, freq)

    #Visualization
    model.plot(forecast)
    plt.ylabel('Rainfall (mm.)')
    plt.xlabel('Date')
    print('ปริมาณฝนสะสมลุ่มแม่น้ำยมจากการทำนาย')
    plt.show()
    plt.close()

    model.plot_components(forecast)
    print('องค์ประกอบของข้อมูลปริมาณฝนสะสมลุ่มแม่น้ำยมจากการทำนาย')
    plt.show()
    plt.close()
```

ภาพที่ 49 ภาพแสดงฟังก์ชัน Data pipeline

```
forecast_demo = forecasting_pipeline(df, forest_df, province='ตอนบน', period=365, freq='D')
```

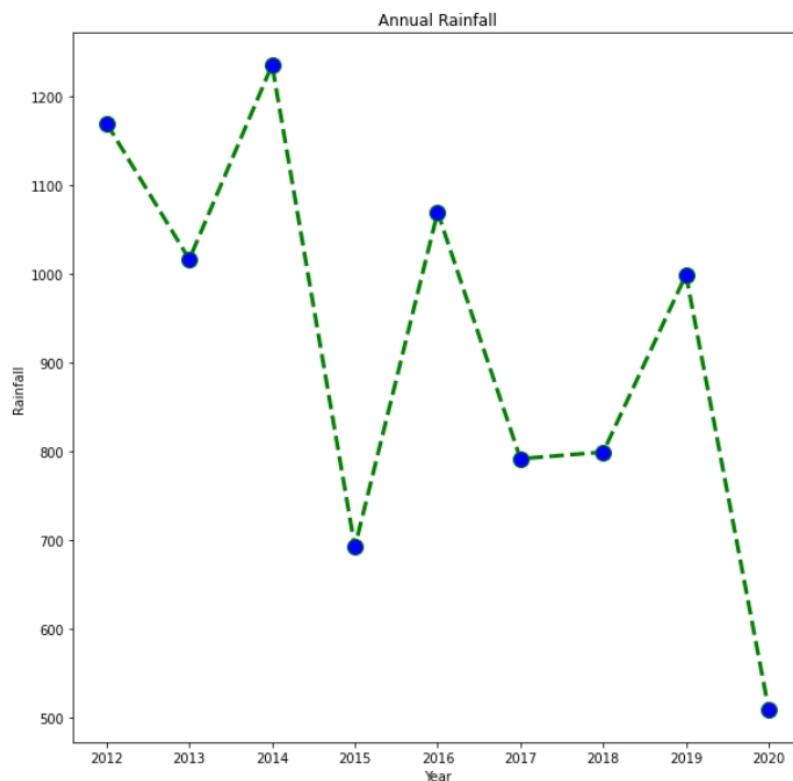
ภาพที่ 50 ภาพแสดงตัวอย่างการใช้งานฟังก์ชัน Data pipeline

5. ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล

จากการสำรวจข้อมูลและวิเคราะห์เกี่ยวกับปริมาณน้ำฝน ปัจจัยสภาพอากาศ พื้นที่ป่าไม้ และปัจจัยอื่นๆ ในระยะเวลา 9 ปี (ค.ศ. 2012 - ค.ศ. 2020)

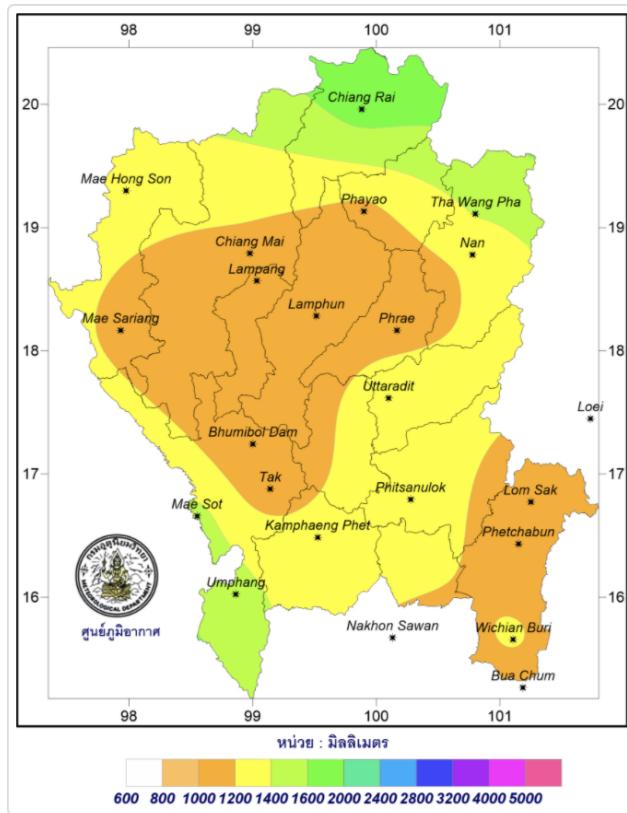
5.1. วิเคราะห์ปริมาณน้ำฝนสะสมรายปีในระยะเวลา 9 ปี (ค.ศ. 2012 - ค.ศ. 2020) จากสถานีตอนบນ

จากการวิเคราะห์ข้อมูลและนำเสนอในรูปแบบของกราฟ พบร่วมปริมาณน้ำฝนสะสมต่อปียังเพิ่มขึ้นและลดลงไม่คงที่ แต่ก็มีแนวโน้มที่จะลดลง ซึ่งในปีที่ปริมาณน้ำฝนสะสมรายปีสูงที่สุดคือปีค.ศ. 2014 ที่ 1,236 มิลลิเมตร หลังจากนั้นก็ลดระดับลงอย่างเห็นได้ชัดในปีค.ศ. 2015 หลังจากนั้นจึงเพิ่มระดับขึ้น และรักษาระดับอยู่ 700 - 1,000 มิลลิเมตร ก่อนจะลดลงอีกรั้งต่อมาที่สุดในรอบ 9 ปี ในปีค.ศ. 2020 ที่ 508.4 มิลลิเมตร



ภาพที่ 51 ภาพแสดงปริมาณน้ำฝนสะสม 10 ปีล่ามต่อนบบ

เมื่อนำข้อมูลปริมาณน้ำฝนสะสมเฉลี่ยรายปีย้อนหลัง 30 ปี (ค.ศ. 1581 - ค.ศ. 2010) จากข้อมูลของศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยามาศึกษาเพิ่มเติม

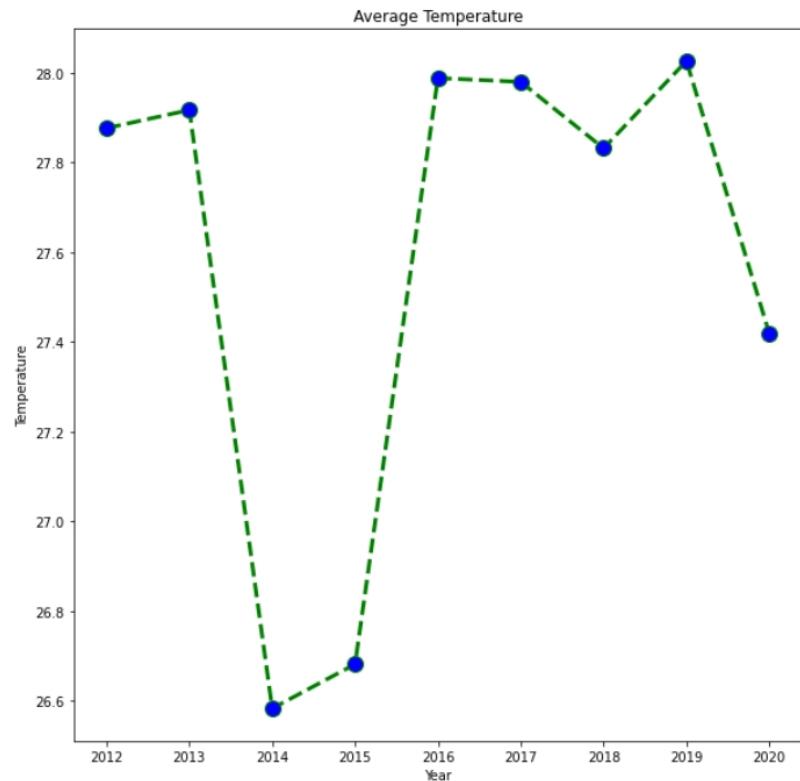


ภาพที่ 52 ภาพแสดงปริมาณฝนสะสมย้อนหลัง 30 ปี

พบว่าปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลัง 30 ปี ของบริเวณกลุ่มแม่น้ำயมสถานีต่อนบนนั้น มีค่าปริมาณอยู่ระหว่าง 1,000 - 1,200 มิลลิเมตร เมื่อเปรียบเทียบกับปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลังในช่วงปี ค.ศ.2012 - ค.ศ. 2020 จะพบว่าในระยะเวลา 9 ปีที่ผ่านมาดังนี้ มี 4 ปีที่ปริมาณน้ำฝนเฉลี่ยสะสมอยู่ในช่วงระดับเดียวหรือมากกว่าปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลัง 30 ปี คือ ค.ศ.2012 , ค.ศ.2013 , ค.ศ.2014 , ค.ศ.2016

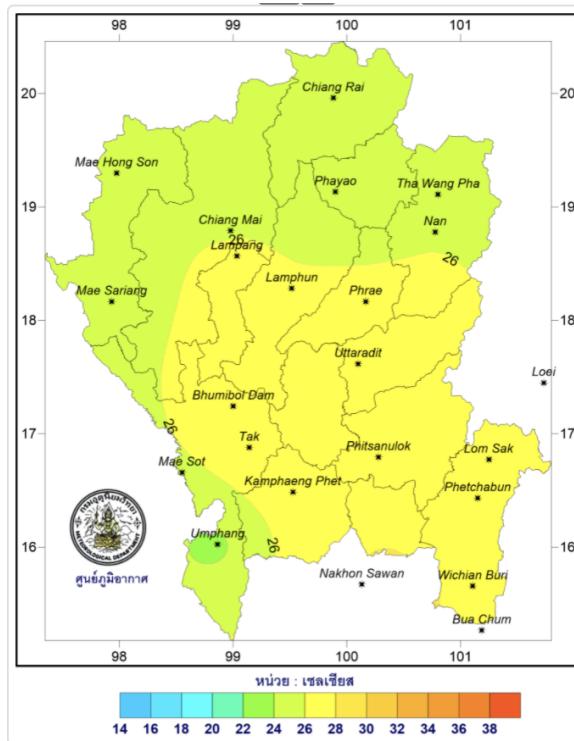
5.2. วิเคราะห์อุณหภูมิเฉลี่ยรายปีในระยะเวลา y 9 ปี (ค.ศ.2012 - ค.ศ. 2020) จากสถานีต่อนบน

จากการวิเคราะห์ข้อมูลและนำเสนอในรูปแบบของกราฟ จะพบว่าในช่วงปี ค.ศ.2012 - ค.ศ. 2020 อุณหภูมิเฉลี่ยของลุ่มน้ำຍมต่อนบนนั้นมีค่าอยู่ระหว่าง 26 - 28 เซลเซียส ถึงจะลดระดับลงมาที่ 26 เซลเซียสในปี ค.ศ.2014 และ ค.ศ.2015 แต่ก็มีแนวโน้มที่จะคงระดับอยู่ในช่วง 27.8 - 28 เซลเซียส



ภาพที่ 53 ภาพแสดงอุณหภูมิ 10 ปีล่าม้ำยมต่อนบน

ซึ่งเมื่อนำข้อมูลอุณหภูมิเฉลี่ยรายปีย้อนหลัง 30 ปี (ค.ศ.1581 - ค.ศ.2010) จากข้อมูลของศูนย์
ภูมิอากาศ จากรบอุตุนิยมวิทยามาศึกษาเพิ่มเติม

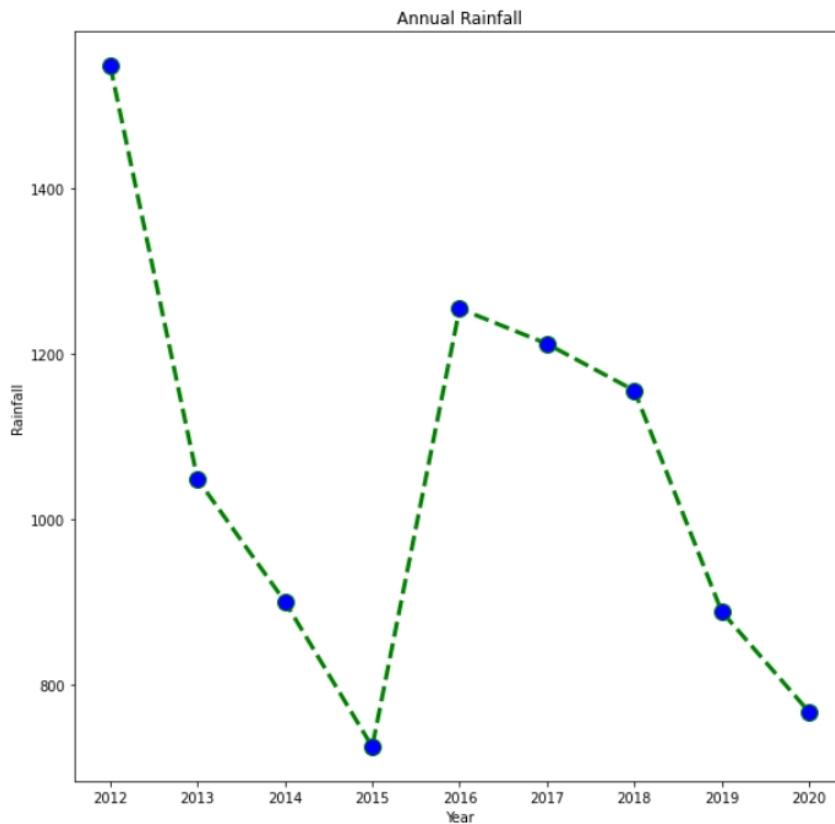


ภาพที่ 54 ภาพแสดงอุณหภูมิเฉลี่ยย้อนหลัง 30 ปี

พบว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปีของบริเวณลุ่มน้ำแม่น้ำยมสถานีต่อนบนนั้น มีค่าประมาณอยู่ระหว่าง 24 - 26 เซลเซียส เมื่อเปรียบเทียบกับอุณหภูมิเฉลี่ยย้อนหลังในช่วง ค.ศ.2012 - ค.ศ. 2020 จะพบว่าในระยะเวลา 9 ปีที่ผ่านมาด้านนี้ ทุกปีมีอุณหภูมิเฉลี่ยสูงกว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปีในทุกปีประมาณ 2 - 3 เซลเซียส

5.3. วิเคราะห์ปริมาณน้ำฝนสะสมรายปีในระยะเวลา 9 ปี (ค.ศ.2012 - ค.ศ. 2020) จากสถานีต่อนกลาง

จากการวิเคราะห์ข้อมูล พบว่าในปี ค.ศ.2012 มีปริมาณฝนสะสมอยู่ที่ 1,549 มิลลิเมตร เป็นปริมาณสะสมที่มากที่สุดในรอบ 9 ปี ก็จะลดลงเรื่อยๆ และมีค่าต่ำที่สุดในปีค.ศ.2015 อยู่ที่ 725.6 มิลลิเมตร แต่ในปี ค.ศ.2016 ปริมาณน้ำฝนสะสมก็กลับเพิ่มขึ้นเกือบ 2 เท่าของก่อนหน้า อยู่ที่ 1,255.2 มิลลิเมตร หลังจากนั้นก็ลดระดับลงอีกรอบตั้งแต่ปีค.ศ.2017 เป็นต้นมา



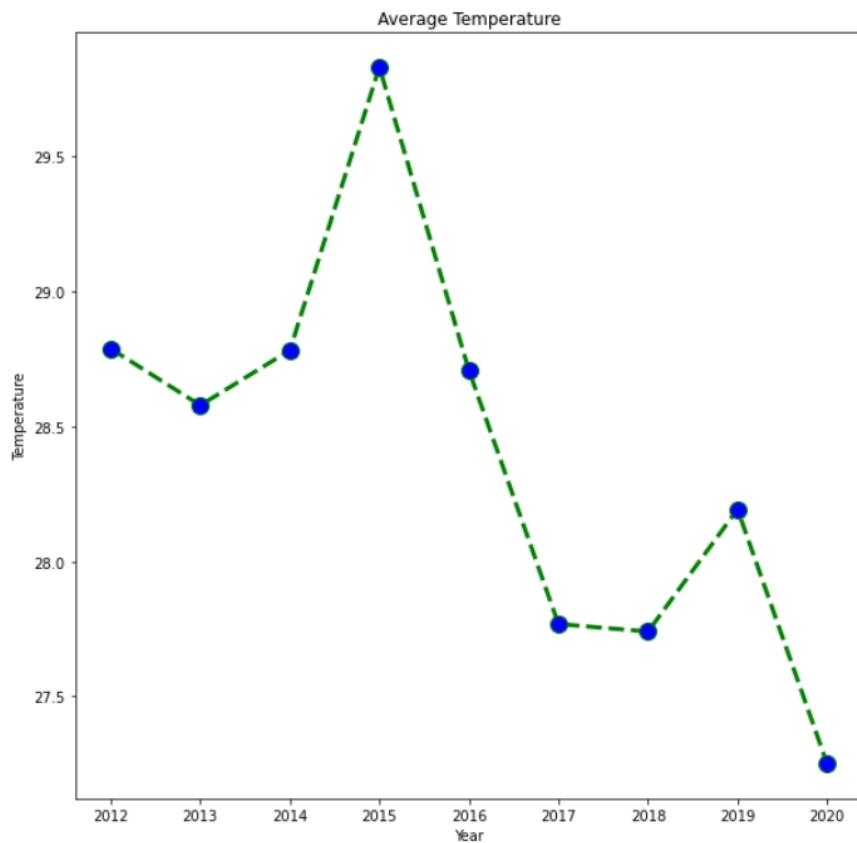
ภาพที่ 55 ภาพแสดงปริมาณน้ำฝนสะสม 10 ปีล่ามั่นหมายต่อนกลาง

เมื่อนำข้อมูลปริมาณน้ำฝนสะสมเฉลี่ยรายปีย้อนหลัง 30 ปี (ค.ศ.1581 - ค.ศ.2010)

จากข้อมูลของศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา(รูปภาพที่ 52)มาศึกษาเพิ่มเติม พบว่าปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลัง 30 ปีของบริเวณลุ่มแม่น้ำมายสถานีตอนกลางนั้น มีค่าประมาณอยู่ระหว่าง 1,000 - 1,200 มิลลิเมตร เมื่อเปรียบเทียบกับปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลังในช่วงปี ค.ศ.2012 - ค.ศ. 2020 จะพบว่าในระยะเวลา 9 ปีที่ผ่านมาตนั้น มี 5 ปีที่ปริมาณน้ำฝนเฉลี่ยสะสมอยู่ในช่วงระดับเดียวหรือมากกว่าปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลัง 30 ปี คือ ค.ศ.2012 , ค.ศ.2013 , ค.ศ.2016 , ค.ศ.2017 , ค.ศ.2018

5.4. วิเคราะห์อุณหภูมิเฉลี่ยรายปีในระยะเวลา 9 ปี (ค.ศ.2012 - ค.ศ. 2020) จากสถานีต่อนกลาง

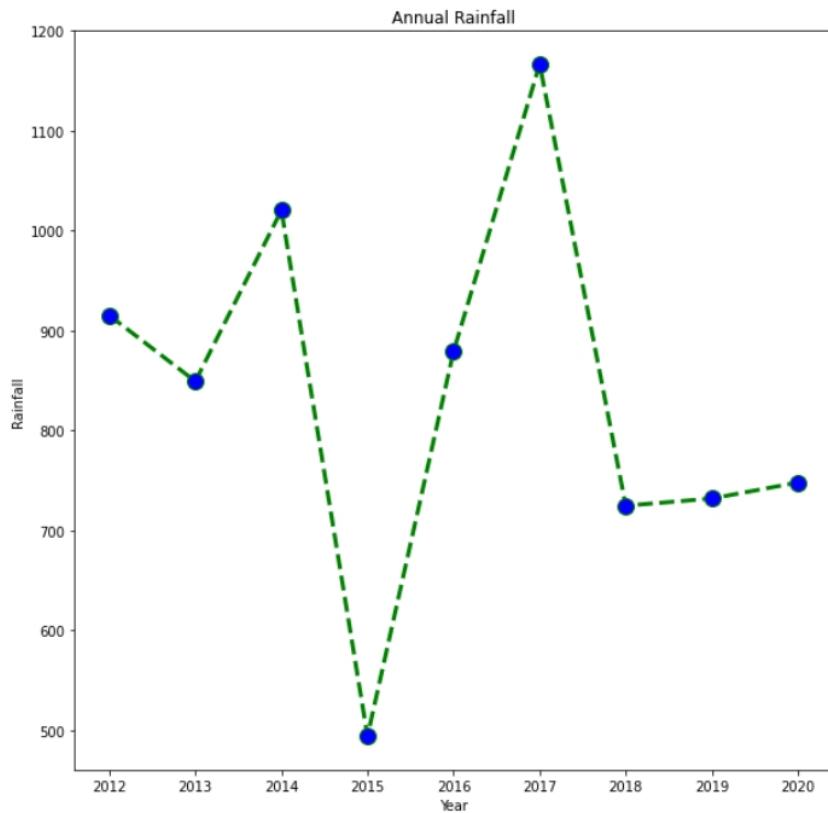
จากการวิเคราะห์ข้อมูลและนำเสนอในรูปแบบของกราฟ จะพบว่าในปี ค.ศ.2012 - ค.ศ. 2020 อุณหภูมิเฉลี่ยของลุ่มน้ำแม่น้ำยมตอนกลางนั้นมีค่าอยู่ระหว่าง 27 - 29 เซลเซียส ถึงปีค.ศ. 2015จะมีค่าสูงสุดแต่ก็ยังมีแนวโน้มที่จะลดลง



ภาพที่ 56 ภาพแสดงอุณหภูมิ 9 ปีลุ่มน้ำยมตอนกลาง
ซึ่งเมื่อนำข้อมูลอุณหภูมิเฉลี่ยรายปีย้อนหลัง 30 ปี (ค.ศ.1581 - ค.ศ.2010) จากข้อมูลของศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา(รูปภาพที่ 54)มาศึกษาเพิ่มเติมพบว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปีของบริเวณลุ่มน้ำยมสถานีต่อนกลางนั้น มีค่าประมาณอยู่ระหว่าง 26 - 28 เซลเซียส เมื่อเปรียบเทียบกับอุณหภูมิเฉลี่ยย้อนหลังในช่วง ค.ศ.2012 - ค.ศ. 2020 จะพบว่าในระยะเวลา 9 ปีที่ผ่านมา น้ำ นอกจากราชปี ค.ศ.2017 , ค.ศ.2018 , ค.ศ.2020 ปีอื่นๆจะมีอุณหภูมิเฉลี่ยสูงกว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปีในทุกปีประมาณ 1 - 2 เซลเซียส

5.5. วิเคราะห์ปริมาณน้ำฝนสะสมรายปีในระยะเวลา 9 ปี (ค.ศ.2012 - ค.ศ. 2020) จากสถานีตอนล่าง

จากการวิเคราะห์ข้อมูล พบร่วมปริมาณน้ำฝนสะสมในช่วงระยะเวลา 9 ปีมีปริมาณที่เพิ่มขึ้นลดลงสลับกันไป ก่อนจะผันผวนอย่างเห็นได้ชัดในช่วงปีค.ศ. 2014 - ค.ศ.2018 ซึ่งในปี ค.ศ.2015 มีปริมาณน้ำฝนสะสมอยู่ที่ 494.4 มิลลิเมตร เป็นปริมาณน้ำฝนสะสมที่น้อยที่สุดในรอบ 9 ปีและเพิ่มสูงขึ้นเรื่อยๆ จนกระทั่งเพิ่มสูงที่สุดในรอบ 9 ปีในปี ค.ศ.2017 อุ่นที่ 1,166.4 มิลลิเมตร ก่อนจะลดระดับลงมากในปีค.ศ. 2018 และค่อยๆเพิ่มขึ้นเล็กน้อย

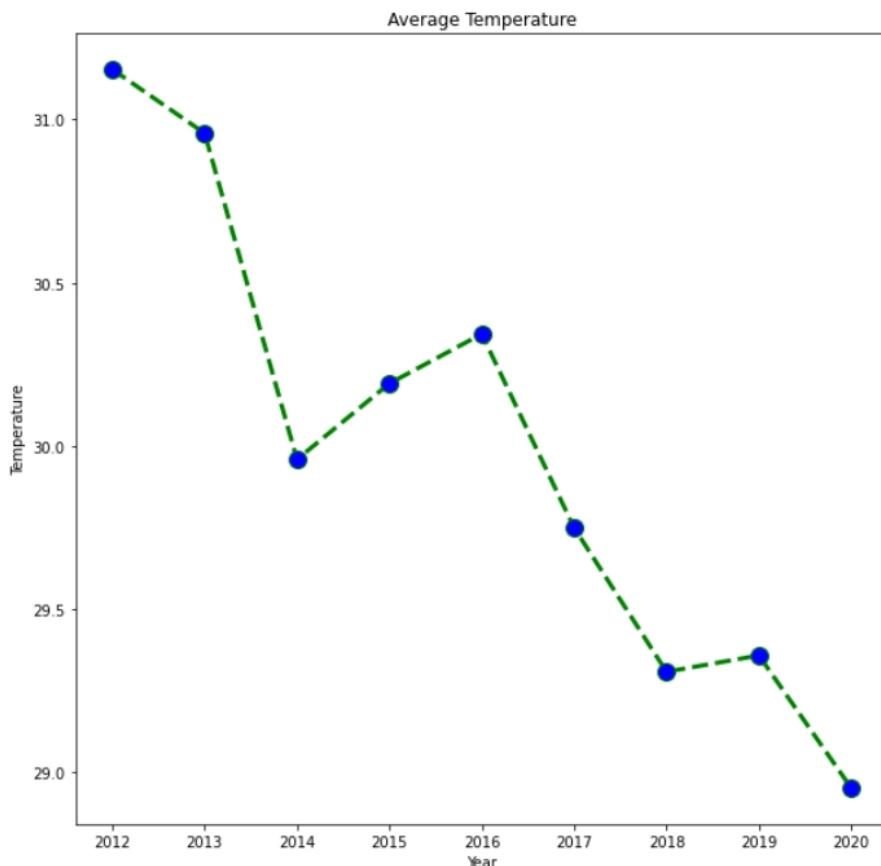


ภาพที่ 57 ภาพแสดงปริมาณน้ำฝนสะสม 10 ปีล่าม้วยตอนล่าง

เมื่อนำข้อมูลปริมาณน้ำฝนเฉลี่ยรายปีย้อนหลัง 30 ปี (ค.ศ.1581 - ค.ศ.2010) จากข้อมูลของศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา(รูปภาพที่ 52)มาศึกษาเพิ่มเติม พบร่วมปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลัง 30 ปีของบริเวณลุ่มน้ำแม่น้ำยมสถานีตอนล่างนั้น มีค่าประมาณอยู่ระหว่าง 1,200 - 1,400 มิลลิเมตร เมื่อเปรียบเทียบกับปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลังในช่วงปี ค.ศ.2012 - ค.ศ. 2020 จะพบว่าในระยะเวลา 9 ปีที่ผ่านมาดังนั้น ไม่มีปีใดที่ปริมาณน้ำฝนเฉลี่ยสะสมอยู่ในช่วงระดับเดียวหรือมากกว่าปริมาณน้ำฝนสะสมเฉลี่ยย้อนหลัง 30 ปี

5.6. วิเคราะห์อุณหภูมิเฉลี่ยรายปีในระยะเวลา 9 ปี (ค.ศ.2012 - ค.ศ. 2020) จากสถานีต่อนล่าง

จากการวิเคราะห์ข้อมูลและนำเสนอในรูปแบบของกราฟ จะพบว่าในปีค.ศ.2012 - ค.ศ. 2020 อุณหภูมิเฉลี่ยของลุ่มน้ำน้ำยมต่อนล่างนั้นมีค่าอยู่ระหว่าง 28 - 31 เซลเซียส และมีแนวโน้มลดลงอย่างเห็นได้ชัด



ภาพที่ 58 ภาพแสดงอุณหภูมิ 10 ปีลุ่มน้ำยมต่อนล่าง

ซึ่งเมื่อนำข้อมูลอุณหภูมิเฉลี่ยรายปีย้อนหลัง 30 ปี (ค.ศ.1581 - ค.ศ.2010) จากข้อมูลของศูนย์ภูมิอากาศกรมอุตุนิยมวิทยา(รูปภาพที่ 54)มาศึกษาเพิ่มเติมพบว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปีของบริเวณลุ่มน้ำน้ำยมสถานีต่อนล่างนั้น มีค่าประมาณอยู่ระหว่าง 26 - 28 เซลเซียส เมื่อเปรียบเทียบกับอุณหภูมิเฉลี่ยย้อนหลังในช่วง ค.ศ .2012 - ค.ศ. 2020 จะพบว่าในระยะเวลา 9 ปีที่ผ่านมานั้น ทุกปีมีอุณหภูมิเฉลี่ยสูงกว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปีในทุกปีประมาณ 2 - 3 เซลเซียส

5.7. วิเคราะห์และเปรียบเทียบปริมาณผนังสมสมของ 3 บริเวณ

เมื่อดูกราฟของทั้ง 3 บริเวณและวิเคราะห์แนวโน้มพบว่า แนวโน้มของสถานีต่อนบนและตอนกลาง มีแนวโน้มที่จะลดลงโดยเฉพาะสถานีตอนกลางที่กราฟจะโค้งลงตั้งแต่ปีค.ศ.2016 เป็นต้นมา แต่สถานีตอนบนปริมาณน้ำฝนสะสมยังเพิ่มและลดลงลับกันไปแต่ถ้าดูแนวโน้มโดยรวมก็ยังแสดงให้เห็นว่า ปริมาณน้ำฝนสะสมลดลงเรื่อยๆ ในทางกลับกันสถานีตอนล่าง ที่ถึงแม้จะมีปริมาณน้ำฝนสะสมเพิ่มขึ้นลดลงลับกัน แต่ก็มีแนวโน้มที่จะเพิ่มขึ้นในปีค.ศ.2018 เป็นต้นมา

ดังนั้นจากการวิเคราะห์กราฟทั้ง 3 จะพบจุดที่น่าสนใจคือ

5.7.1. ปริมาณน้ำฝนสะสมของบริเวณลุ่มแม่น้ำยมตอนบนและตอนกลางในปี ค.ศ.2014 ที่มีค่าสูงที่สุดในรอบ 9 ปี จากการสืบค้นข้อมูลเพิ่มเติมจากบันทึกเหตุการณ์น้ำท่วม สถาบันสารสนเทศทรัพยากรน้ำ พบว่าในปี ค.ศ.2014 ประเทศไทยได้รับอิทธิพลจากพายุคัลแมกีและมรสุมที่เข้ามาทางประเทศไทยตอนบน ส่งผลให้บริเวณภาคเหนือตอนบน ภาคตะวันออกเฉียงเหนือ รวมถึงลุ่มแม่น้ำตอนกลางและบน ได้รับผลกระทบฝนตกชุก ทำให้สถานีตอนบนและตอนกลางในปีค.ศ. 2014 มีปริมาณสูงสุด

5.7.2. ปริมาณน้ำฝนสะสมของบริเวณลุ่มแม่น้ำยมทั้ง 3 พื้นที่ในปีค.ศ. 2015 ที่ทั้งสถานีตอนบน ตอนกลาง และตอนล่างมีปริมาณน้ำฝนสะสมต่ำลงมากอย่างเห็นได้ชัด จากการสืบค้นข้อมูลเพิ่มเติมจากเว็บไซต์ศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา พบว่าในปีค.ศ. 2015 ประเทศไทยประสบภัยแล้งอย่างหนักเนื่องจากปราภูภารณ์เคลื่อนไหวระดับรุนแรง ทำให้ฝนไม่ตกต่อเนื่องนานหลายเดือนและทำให้ฝนไม่ตกตามฤดูกาล ส่งผลให้หลายพื้นที่ในประเทศประสบภัยแล้ง รวมถึงบริเวณลุ่มแม่น้ำยมด้วย จึงทำให้ปริมาณน้ำฝนสะสมของปีค.ศ.2015 มีค่าต่ำลงใน 3 พื้นที่

5.7.3. ปริมาณน้ำฝนสะสมของบริเวณลุ่มแม่น้ำยมตอนล่างในปีค.ศ.2017 ซึ่งจะพบว่ามีปริมาณน้ำฝนสะสมมากที่สุดในช่วง 9 ปี ในขณะที่ปริมาณผนังสะสมของบริเวณลุ่มแม่น้ำยมในปีค.ศ.2017 ของบริเวณตอนบนและตอนกลางไม่ได้มีปริมาณที่สูงที่สุด จากข้อมูลของศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา พบว่าในปีค.ศ.2017 นั้น ประเทศไทยได้รับผลกระทบจากมรสุมตะวันตกเฉียงใต้ตั้งแต่เดือนมกราคม และพายุหมุนจากทะเลจีนใต้ในเดือน พฤษภาคม นอกจากนี้ยังได้รับผลกระทบจากพายุดีเปรสชันจากตะวันตกเฉียงใต้อย่างต่อเนื่องถึงสิ้นปี ซึ่งส่งผลให้ภาคตะวันตก ภาคกลาง ภาคเหนือตอนล่าง รวมถึงลุ่มแม่น้ำ

ym ผนตกเกือบทตลอดปี ปริมาณน้ำฝนสะสมในปีค.ศ. 2017 ในสถานีต่อนล่างจึงมีค่าสูงแตกต่างจากสถานีต่อนบนและตอนกลางที่ได้รับผลกระทบน้อยกว่า

- 5.7.4.** ปริมาณน้ำฝนสะสมของบริเวณลุ่มแม่น้ำยมทั้ง 3 พื้นที่ในปีค.ศ. 2020 จากราฟจะสังเกตได้ว่าในปีค.ศ. 2020 บริเวณลุ่มแม่น้ำยมมีค่าน้อยมาก โดยเฉพาะลุ่มแม่น้ำยมตอนบนที่ปีค.ศ. 2020 เป็นปีที่ปริมาณน้ำฝนน้อยที่สุดในรอบ 9 ปี จากการศึกษาข้อมูลของบันทึกภัยแล้ง ปีค.ศ. 2020 จักคลังข้อมูลน้ำแห้งชาติ พบร้า ประเทศไทยได้รับอิทธิพลจากปรากฏการณ์โอลนีโว ที่ส่งผลให้ฝนตกน้อยลงต่อเนื่อง ส่งผลให้ปริมาณน้ำฝนสะสมของปีค.ศ. 2020 มีค่าต่ำมากทั้ง 3 พื้นที่

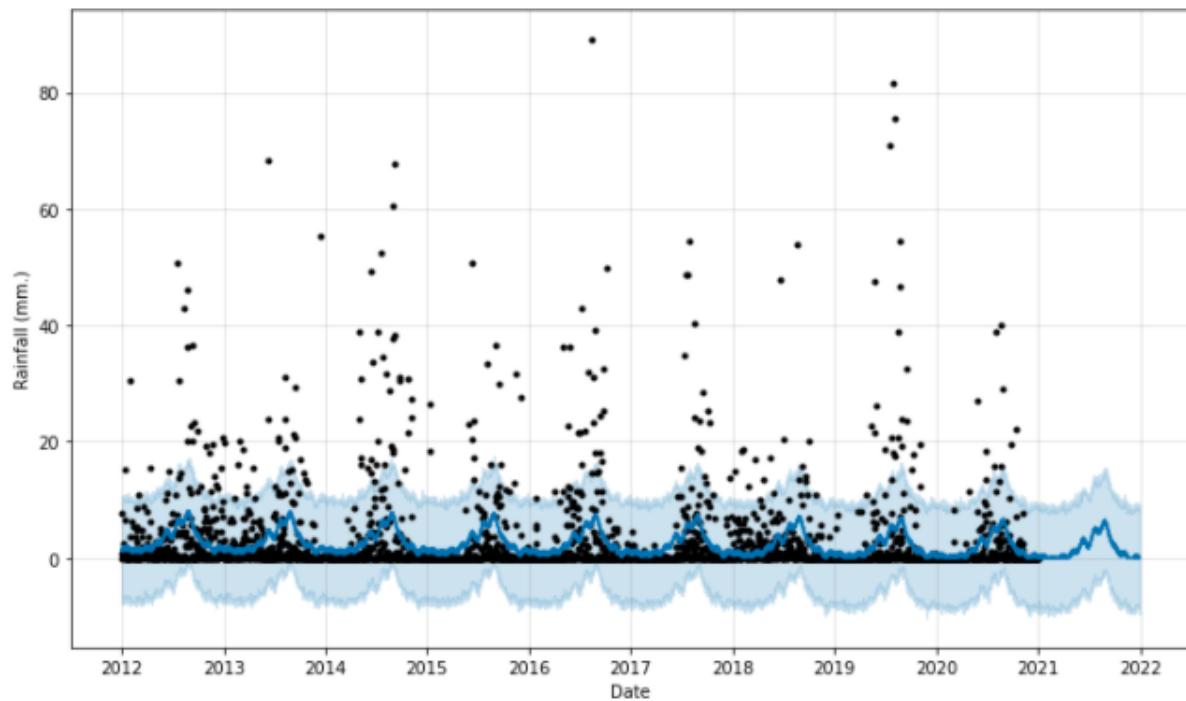
5.8. วิเคราะห์แนวโน้มปริมาณฝนสะสมในบริเวณลุ่มแม่น้ำยมในอนาคต

ในการวิเคราะห์หาแนวโน้มปริมาณฝนสะสมในอนาคตของแต่ละพื้นที่จะใช้ช่วงความถี่ของข้อมูลเป็นวัน เพื่อให้เห็นความแปรผันตามฤดูกาลและวิเคราะห์หาช่วงเวลาการเกิดฝนในระหว่างปี

5.8.1. พื้นที่ลุ่มแม่น้ำยมตอนบน (สถานี ทต. เชียงม่วน)

จากการคาดการณ์แนวโน้มปริมาณฝนสะสมในปีค.ศ. 2021 พบร้าปริมาณฝนสะสมจะมีความแปรผันตามฤดูกาลเช่นเดิมเหมือนปีที่ผ่านมา แต่แนวโน้มของปริมาณฝนจะลดลงตามการคาดการณ์ เอียงของเส้นกราฟสีฟ้าที่ลากเข้าใกล้ 0 ดังแผนภาพข้างล่าง โดยจุดสีดำคือค่าปริมาณฝนสะสมจริงในปี ค.ศ. 2012-2020 ส่วนเส้นสีฟ้าคือปริมาณฝนที่โมเดลคาดการณ์ และแบบสีฟ้าอ่อนคือช่วงความคลาดเคลื่อนไม่แน่นอนของข้อมูลที่อาจจะเกิดขึ้นในบริเวณแถบสีฟ้าอ่อน

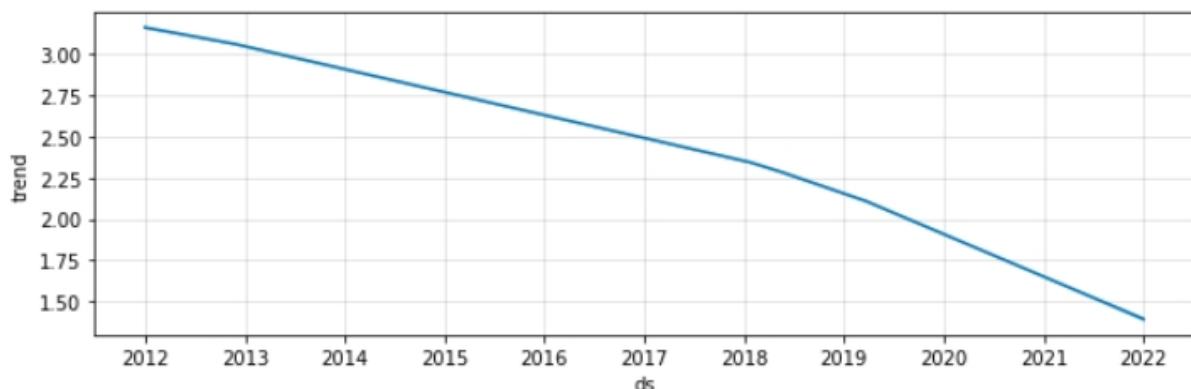
ปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการท่านายปี ค.ศ. 2012-2021



ภาพที่ 59 ภาพแสดงปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการท่านายปี ค.ศ. 2012-2021

ซึ่งเมื่อลองวิเคราะห์แยกองค์ประกอบของปริมาณเวลาของปริมาณน้ำฝนสะสมตั้งแต่ปีค.ศ. 2012-2021 เพื่อหาแนวโน้มของปริมาณน้ำฝนดังภาพด้านล่างพบว่า แนวโน้มของปริมาณฝนในปีที่ผ่านมาลดลงอย่างเห็นได้ชัด และในปีค.ศ. 2021 มีแนวโน้มลดลงอย่างต่อเนื่อง แสดงว่าในช่วงเวลา 9 ปีที่ผ่านมาจนถึงปัจจุบันปริมาณของฝนที่ลดลงส่งผลให้พื้นที่บริเวณลุ่มน้ำยมต่อนบนพบเจอกับปัญหาภัยแล้งที่รุนแรงขึ้นกว่าเดิมในทุกๆ ปี

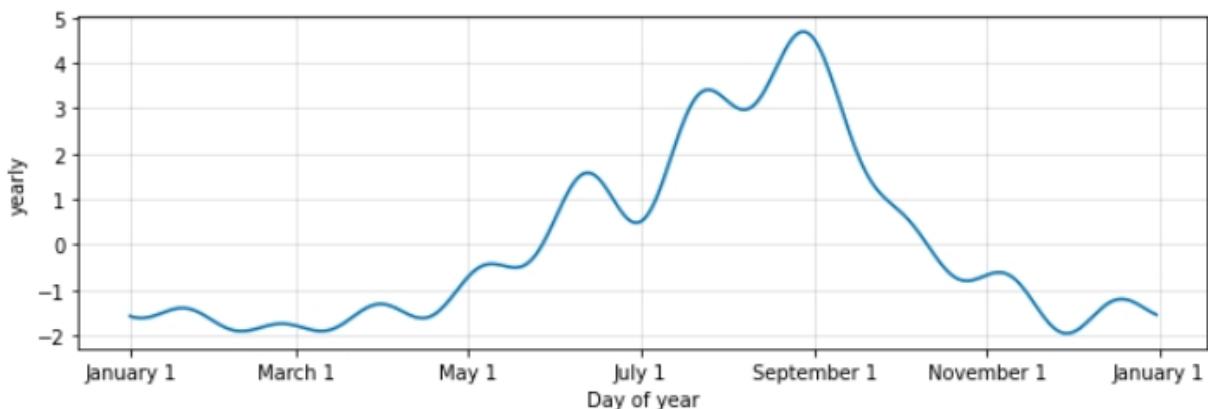
แนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการท่านายปี ค.ศ. 2012-2021



ภาพที่ 60 ภาพแสดงแนวโน้มของปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการท่านายปี ค.ศ. 2012-2021

นอกจากนี้พบว่าลักษณะการเกิดฝนหรือช่วงฤดูของปริมาณฝนในรอบปีของ

พื้นที่บริเวณลุ่มน้ำยมต่อนบนจะมีฝนในช่วงประมาณเดือนมิถุนายน-เดือนตุลาคมซึ่งอยู่ในช่วงฤดูฝน โดยเดือนที่มีฝนตกมากที่สุดจะเป็นเดือนกันยายน ส่วนช่วงที่มีปริมาณฝนน้อยมากที่สุดจะเป็นช่วงเดือนมกราคม-เดือนมีนาคม และช่วงที่มีความผันผวนมากที่สุดจะเป็นช่วงประมาณท้ายเดือนมิถุนายน-ต้นเดือนกรกฎาคม ดังภาพด้านล่าง

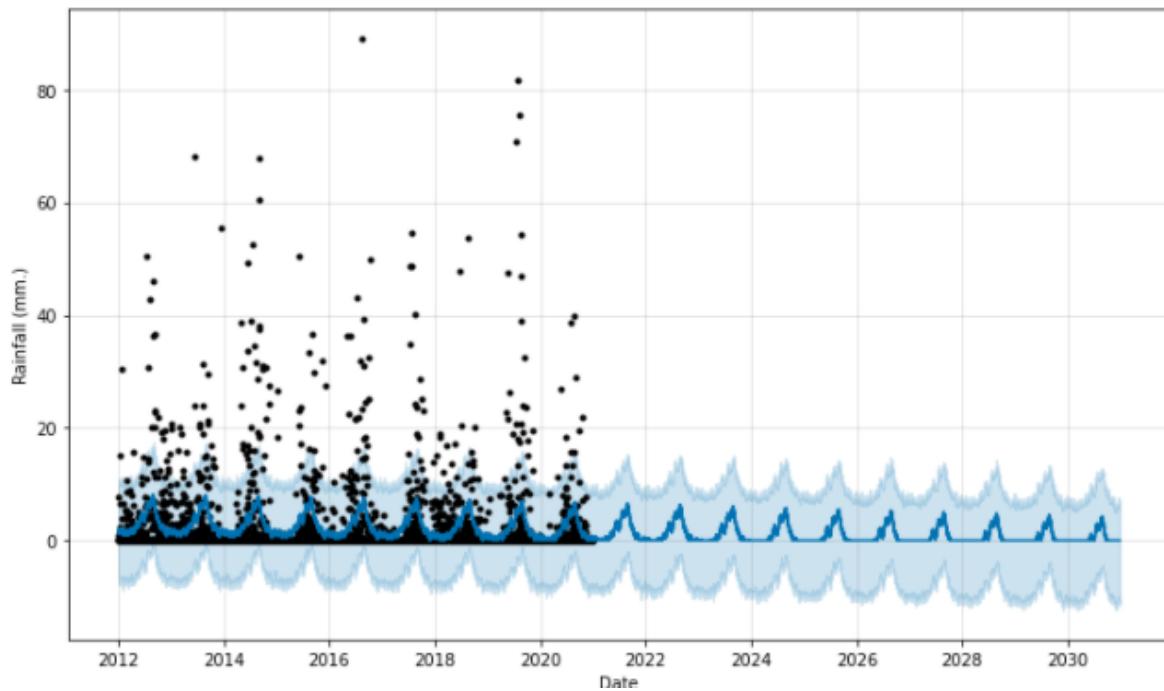


ภาพที่ 61 ภาพแสดงช่วงเวลาปริมาณฝนสะสมที่เกิดในพื้นที่ลุ่มน้ำยมต่อนบน

เมื่อพิจารณาคาดการณ์แนวโน้มปริมาณฝนสะสมของพื้นที่ลุ่มน้ำยมต่อนบน

ในอีก 10 ปีข้างหน้า (ค.ศ. 2021-2030) ดังภาพด้านไป

ปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการทํางานปี ค.ศ. 2012-2030

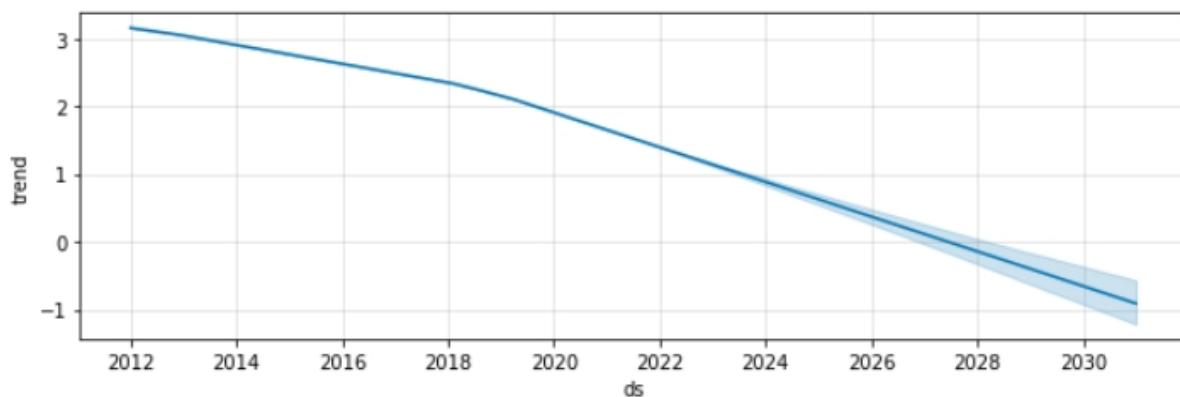


ภาพที่ 62 ภาพแสดงปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการทํางานปี ค.ศ. 2012-2030

จะสังเกตได้ว่าลักษณะของปริมาณฝนตามฤดูกาลของข้อมูลมีลักษณะที่คล้าย

กันในทุกๆ ปี แต่ระยะห่างระหว่างยอดความสูงของปริมาณฝนสะสมสูงสุด กับระดับ
ของปริมาณฝนสะสมต่ำสุดในแต่ละปี ระยะห่างเริ่มลดลงต่อเนื่อง และเมื่อแยกองค์
ประกอบตามอนุกรมเวลาเพื่อหาแนวโน้มเป็นดังต่อไปนี้

แนวโน้มของปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการทํางานปี ค.ศ. 2012-2030



ภาพที่ 63 ภาพแสดงแนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อนบนจากการทํางานปี ค.ศ. 2012-2030

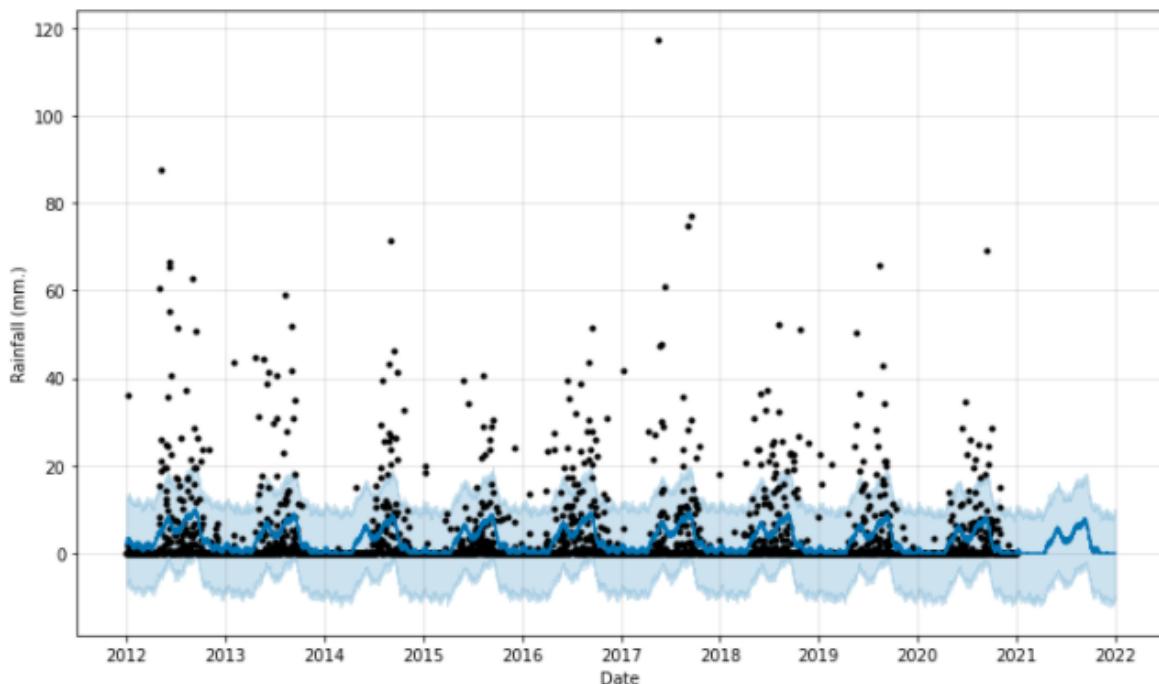
จากภาพแสดงว่าแนวโน้มของปริมาณฝนในอีก 10 ปีต่อมาจะคงลดลงอย่างต่อ
เนื่อง แม้จะอาจจะเกิดความคาดเคลื่อนได้ตามพื้นที่สีฟ้าอ่อนช่วงที่สามารถคาดเคลื่อน

ได้ก็ยังทำให้แนวโน้มลดต่ำลงกว่าแนวโน้มก่อนปีที่ผ่านมาซึ่งการที่ปริมาณฝนลดลงอย่างต่อเนื่องอาจส่งผลให้พื้นที่ลุ่มน้ำมีน้ำท่วมต่อนบนพบร่องกับความแห้งแล้งจากปริมาณฝนที่ลดลงมากขึ้นเรื่อยๆ ในอนาคต

5.8.2. พื้นที่ลุ่มน้ำมีน้ำท่วมต่อนกลาง (สถานี อบต. วังชี้น)

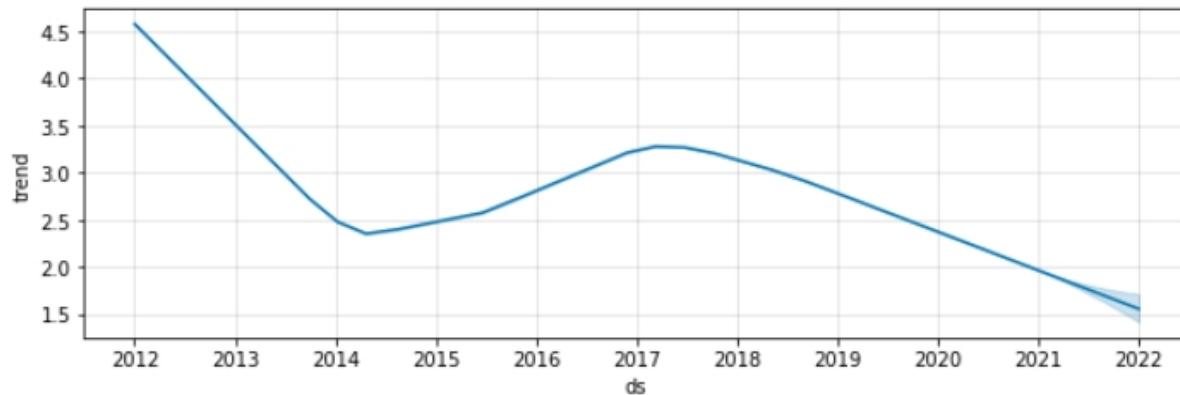
จากการคาดการณ์ปริมาณฝนสะสมบนภาคต้องพื้นที่ลุ่มน้ำมีน้ำท่วมต่อนกลาง ในปี ค.ศ. 2021 พบว่าปริมาณฝนสะสมมีลักษณะรูปแบบตามฤดูกาลในลักษณะเดียวกับปีที่ผ่านมา จึงมีการแยกกองค์ประกอบอนุกรมเวลาจากภาพด้านล่างเพื่อหาแนวโน้มของปริมาณฝนที่ชัดเจนมากยิ่งขึ้น

ปริมาณฝนสะสมลุ่มน้ำมีน้ำท่วมต่อนกลางจากการทํางานปี ค.ศ. 2012-2021



ภาพที่ 64 ภาพแสดงปริมาณฝนสะสมลุ่มน้ำมีน้ำท่วมต่อนกลางจากการทํางานปี ค.ศ. 2012-2021 เมื่อแยกกองค์ประกอบแล้วจะได้แนวโน้มของข้อมูลปริมาณฝนตั้งภาพด้านล่าง จะสังเกตได้ว่าปริมาณฝนสะสมของพื้นที่ลุ่มน้ำมีน้ำท่วมต่อนกลางมีความผันผวนมากกว่าพื้นที่ลุ่มน้ำมีน้ำท่วมต่อนบนที่มีแนวโน้มลดลงอย่างเดียว เพียงแต่เมื่อผ่านปีค.ศ. 2017 เป็นต้นมาจนถึงปีค.ศ. 2021 แนวโน้มของปริมาณฝนสะสมลดลงอย่างต่อเนื่อง

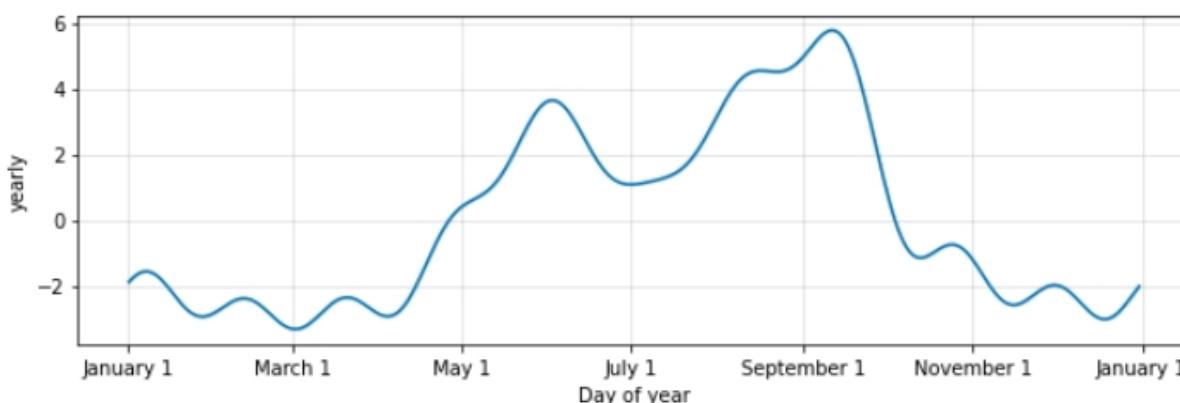
แนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อนกลางจากการทำงานปี ค.ศ. 2012-2021



ภาพที่ 65 ภาพแสดงแนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อนกลางจากการทำงานปี ค.ศ.2012-2021

นอกจากนี้มีการวิเคราะห์ช่วงเวลาการเกิดฝนของพื้นที่ลุ่มน้ำยมต่อนกลางได้ดัง

ภาพข้างล่าง



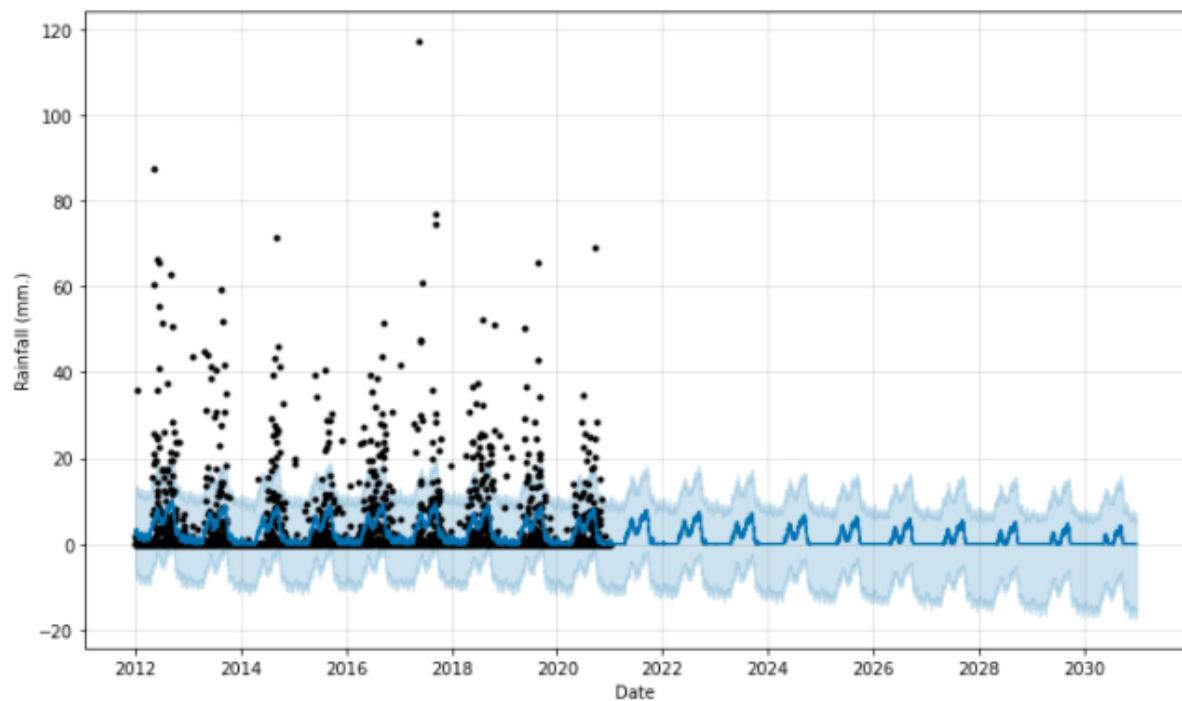
ภาพที่ 66 ภาพแสดงช่วงเวลาปริมาณฝนสะสมที่เกิดในลุ่มน้ำยมต่อนกลาง

สังเกตได้ว่าปริมาณการเกิดฝนอยู่ในช่วงประมาณเดือนพฤษภาคม-เดือน

ตุลาคม โดยเดือนกันยายนมีปริมาณฝนสะสมสูงสุดในรอบปี เช่นเดียวกับพื้นที่ลุ่มน้ำยม ตอนบน แต่ในช่วงประมาณท้ายเดือนมิถุนายน-ต้นเดือนกรกฎาคมมีความผันผวนของปริมาณน้ำฝนที่กินระยะเวลาค่อนข้างนานกว่าช่วงอื่นๆ ทั้งที่เป็นช่วงที่การเกิดฝนกว่าช่วงอื่นๆ ในปี ส่วนช่วงที่เกิดฝนน้อยหรือไม่มีฝนเลยจะอยู่ในช่วงหลังเดือนตุลาคมถึงเดือนเมษายนในปีถัดไป

เมื่อคาดการณ์ปริมาณฝนสะสมที่จะเกิดขึ้นในพื้นที่ลุ่มน้ำยมในระยะเวลา 10 ปี ถัดมา (ค.ศ. 2021-2030) จะได้ผลลัพธ์ดังนี้

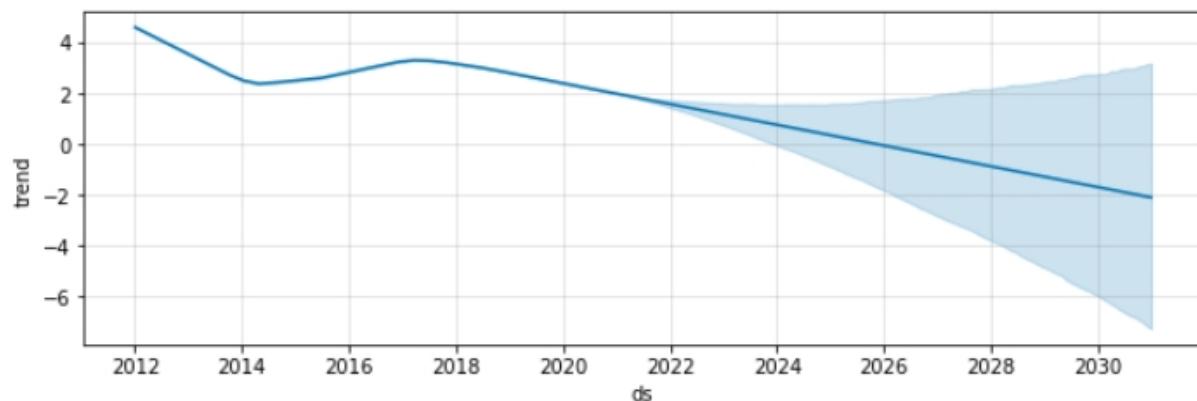
ปริมาณฝนสะสมลุ่มน้ำยมต่อนกลางจากการทำนายปี ค.ศ. 2012-2030



ภาพที่ 67 ภาพแสดงปริมาณฝนสะสมลุ่มน้ำยมต่อนกลางจากการทำนายปี ค.ศ.2012-2030

พบว่าในระยะเวลา 10 ปีต่อมาจากการคาดการณ์ลักษณะการเกิดฝนของพื้นที่ลุ่มน้ำยมต่อนกลางยังคงเกิดในลักษณะเดิม แต่ระยะห่างระหว่างยอดที่มีปริมาณฝนสูงสุดกับระดับปริมาณฝนที่ต่ำสุดมีระยะห่างน้อยลงเรื่อยๆ จนรวมกับว่าช่วงที่มีความผันผวนมากที่สุดในเดือนกรกฎาคมปริมาณฝนสะสมจะลดต่ำลงเท่ากับ 0 ซึ่งอยู่ในช่วงที่มีการเกิดฝนมากที่สุด ตามแนวโน้มดังต่อไปนี้

แนวโน้มของปริมาณฝนสะสมลุ่มน้ำยมต่อนกลางจากการทำนายปี ค.ศ. 2012-2030



ภาพที่ 68 ภาพแสดงแนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อนกลางจากการทำนายปี ค.ศ.2012-2030

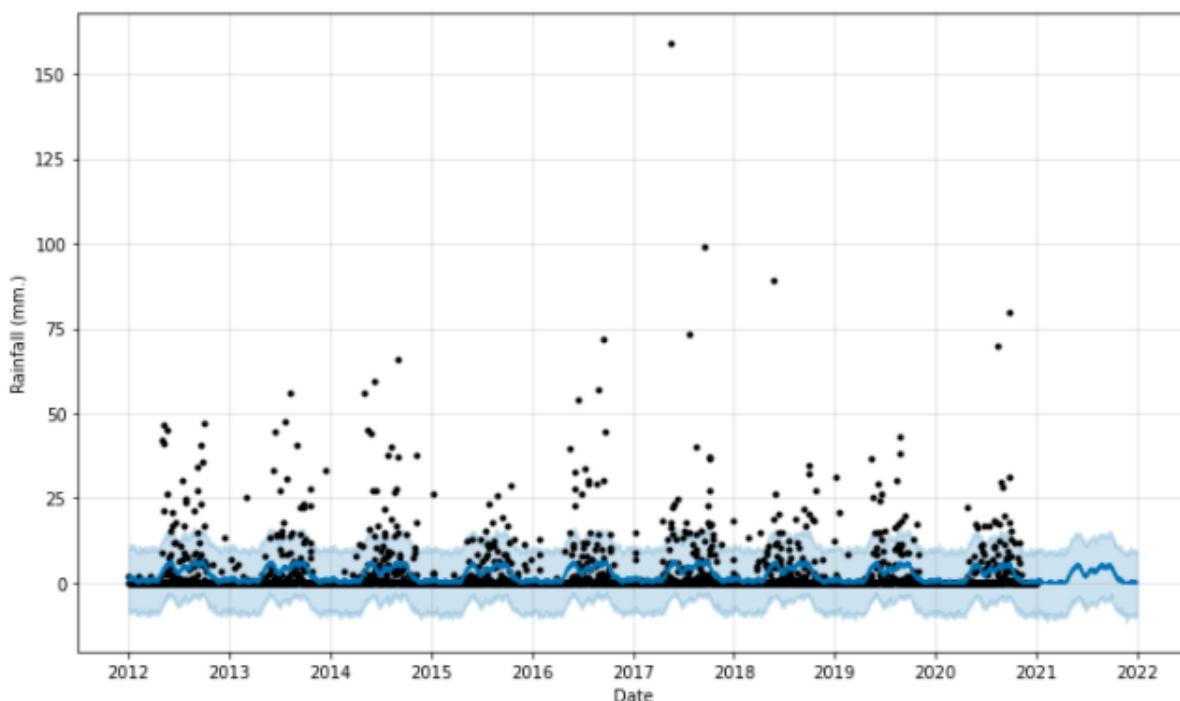
แสดงได้ว่าในอนาคตจากปริมาณการเกิดฝนจะลดลงอย่างต่อเนื่องแต่อย่างไรก็ดีช่วงความคลาดเคลื่อนที่สามารถเกิดได้ในอนาคตก็อาจพาแนวโน้มของปริมาณฝนสูงขึ้นหรือต่ำลงกว่าที่แผนภาพแสดง แต่จากผลของโมเดลที่เลือกทำนายว่าแนวโน้มปริมาณฝนในอนาคตลดลงนั้นทำให้ความผันผวนของปริมาณน้ำฝนในช่วงเดือนกรกฎาคมจะรุนแรงมากขึ้นกว่าเดิม ซึ่งอาจเกิดเป็นฝนทึ่งช่วงได้ในเดือนกรกฎาคม ทำให้พื้นที่ลุ่มน้ำแม่น้ำยมต่อนกลางจะพบเจอกับภัยแล้งได้มากขึ้นในเดือนนี้ นอกเหนือจากภัยแล้งที่พบเจอในช่วงเดือนตุลาคม-เดือนเมษายนปีถัดไป

5.8.3. พื้นที่ลุ่มน้ำแม่น้ำยมต่อนล่าง (สถานี โรงไฟฟ้าสัตหีว ทต. สารคดโลก)

จากการคาดการณ์ปริมาณฝนสะสมของพื้นที่ลุ่มน้ำแม่น้ำยมต่อนล่างในปีค.ศ. 2021

ได้ผลลัพธ์ดังนี้

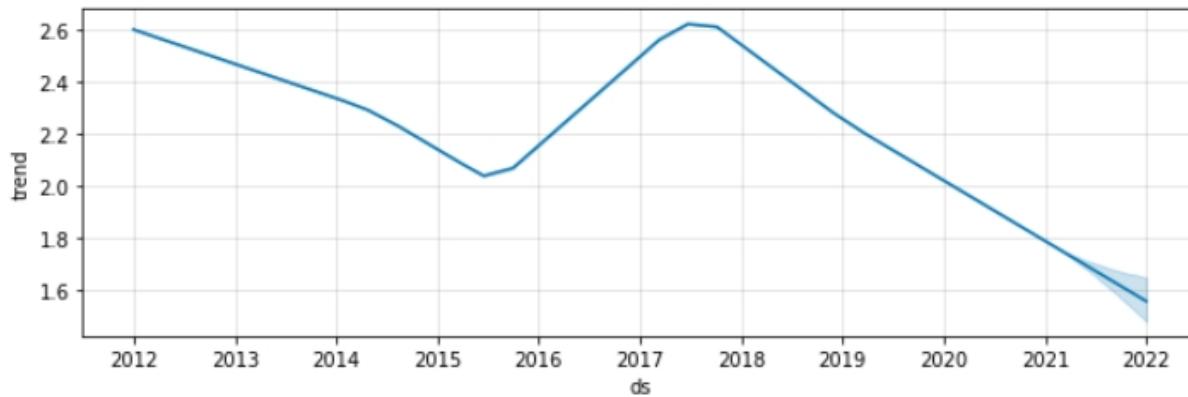
ปริมาณฝนสะสมลุ่มน้ำแม่น้ำยมต่อนล่างจากการทำนายปี ค.ศ. 2012-2021



ภาพที่ 69 ภาพแสดงปริมาณฝนสะสมลุ่มน้ำแม่น้ำยมต่อนล่างจากการทำนายปี ค.ศ. 2012-2021

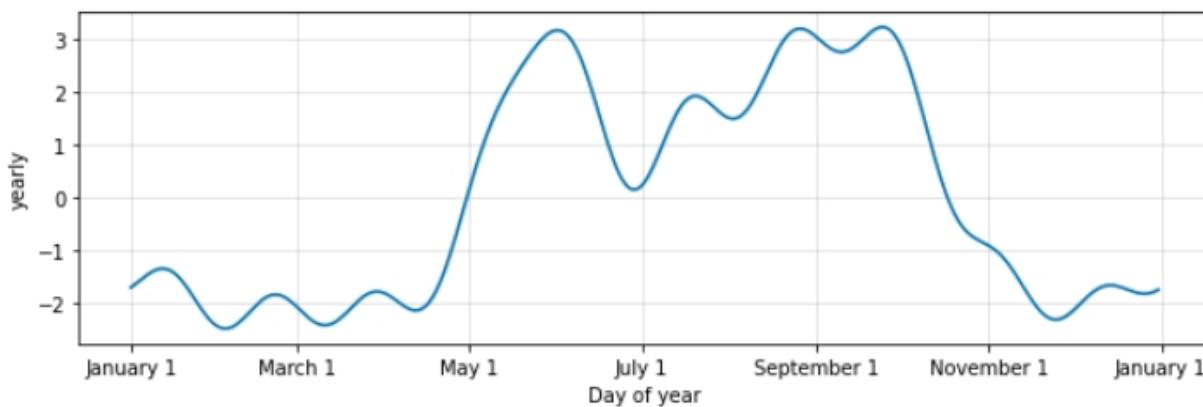
ลักษณะของปริมาณฝนสะสมในปีค.ศ. 2021 มีลักษณะเดียวกับปีที่ผ่านมา เมื่อเทียบกับพื้นที่ 2 ส่วนก่อนหน้านี้เพียงแต่ความสูงของกราฟ หรือก็คือปริมาณของน้ำฝนสะสมโดยภาพรวมแล้วน้อยกว่าพื้นที่ทั้ง 2 ส่วน แต่แนวโน้มของปริมาณฝนสามารถดูได้จากแผนภาพต่อไปนี้

แนวโน้มของปริมาณฝนสะสมลุ่มน้ำยมต่อน้ำจากการทำนายปี ค.ศ. 2012-2021



ภาพที่ 70 ภาพแสดงแนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อน้ำจากการทำนายปี ค.ศ. 2012-2021

แนวโน้มของปริมาณฝนสะสมจากปีที่ผ่านมา ก็มีแนวโน้มสูงขึ้นและลดลง แต่หลังจากปีค.ศ. 2018-2021 แนวโน้มของปริมาณฝนลดลงอย่างต่อเนื่อง และเมื่อตัดช่วงเวลาการเกิดฝนในแต่ละรอบปีได้ผลลัพธ์ดังนี้



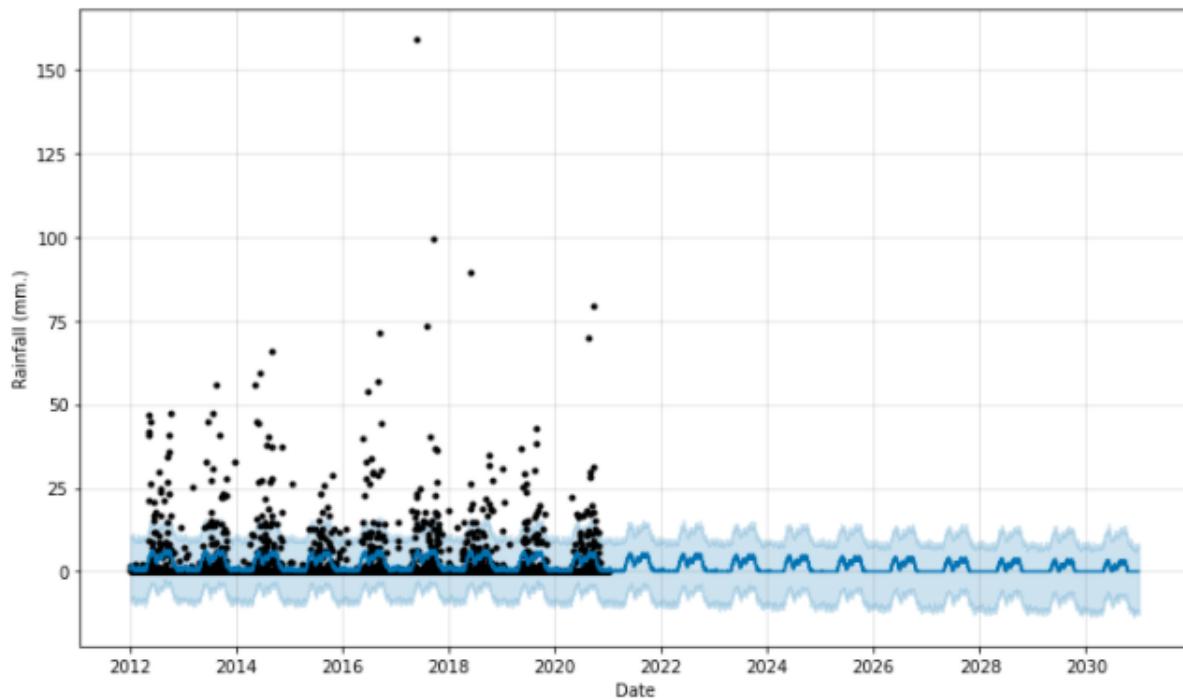
ภาพที่ 71 ภาพแสดงช่วงเวลาปริมาณฝนสะสมที่เกิดขึ้นในลุ่มน้ำยมต่อน้ำล่าง

ลักษณะของช่วงเวลาการเกิดฝนคล้ายกับพื้นที่ลุ่มน้ำยมต่อน้ำกลาง เนื่องจากช่วงเวลาการเกิดฝนในพื้นที่ลุ่มน้ำยมต่อน้ำล่างนี้ มีปริมาณฝนสะสมสูงในช่วงเดือนพฤษภาคม-เดือนตุลาคม และช่วงที่มีความผันผวนของปริมาณน้ำฝนจะเกิดในช่วงประมาณห้ามเดือนมิถุนายน-ต้นเดือนกรกฎาคม เช่นเดียวกัน แต่ช่วงที่เกิดปริมาณฝนสูงสุดจะมี 2 ช่วง ด้วยกันคือช่วงประมาณเดือนมิถุนายน และช่วงเดือนตุลาคม และช่วงที่ฝนตกน้อยหรือไม่ตกเลยจะเป็นในช่วงหลังเดือนตุลาคมจนถึงเดือนเมษายนในปีถัดไป

จากการคาดการณ์แนวโน้มปริมาณฝนสะสมที่จะเกิดขึ้นใน 10 ปีต่อมา (ค.ศ.

2021-2030) ได้ผลลัพธ์ดังนี้

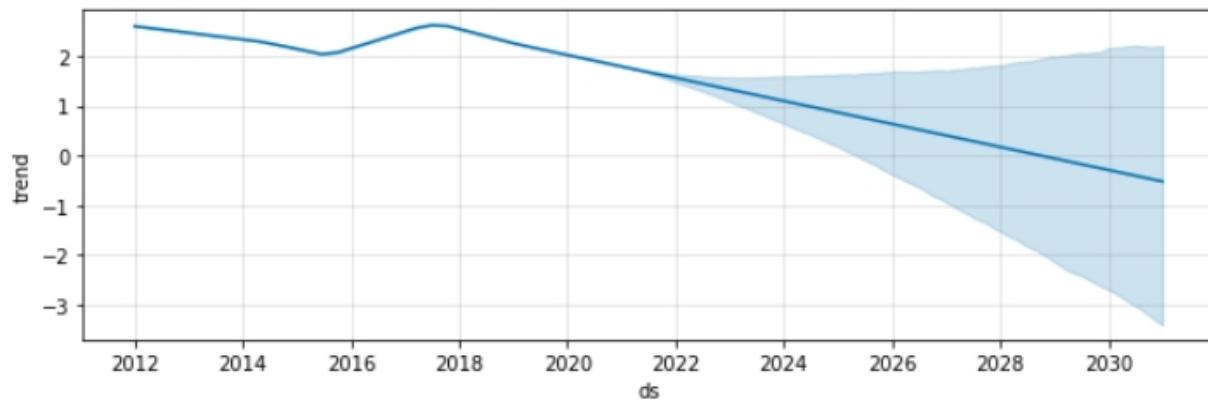
ปริมาณฝนสะสมลุ่มน้ำยมต่อนล่างจากการทำนายปี ค.ศ. 2012-2030



ภาพที่ 72 ภาพแสดงปริมาณฝนสะสมลุ่มน้ำยมต่อนล่างจากการทำนายปี ค.ศ.2012-2030

สังเกตได้ว่าแนวโน้มของปริมาณฝนสะสมมีลักษณะเหมือนกับพื้นที่ 2 ส่วนที่ผ่านมาคือลักษณะของปริมาณฝนในแต่ละปีเกิดในรูปแบบเดิม แต่ระยะห่างระหว่างจุดสูงสุดและต่ำสุดของปริมาณฝนสะสมน้อยลงเรื่อยๆ และเข่นเดียวกับพื้นที่ลุ่มน้ำยมต่อนกลางช่วงเดือนกรกฎาคมที่เป็นช่วงที่มีความผันผวนจะรุนแรงขึ้น โดยลดต่ำลงจนเข้าใกล้ 0 ซึ่งอาจทำให้เกิดฝนทึ่งช่วงในช่วงฤดูฝนได้ ซึ่งก่อให้เกิดภัยแล้งในเดือนกรกฎาคมได้ในอนาคต นอกจากภัยแล้งในช่วงเดือนตุลาคม-เมษายนปีถัดไป ตามปกติ

แนวโน้มของปริมาณฝนสะสมลุ่มน้ำยมต่อน้ำจากการทำนายปี ค.ศ. 2012-2030



ภาพที่ 73 ภาพแสดงแนวโน้มปริมาณฝนสะสมลุ่มน้ำยมต่อน้ำจากการทำนายปี ค.ศ.2012--2030

จากภาพแสดงแนวโน้มของปริมาณฝนสะสมในระยะเวลา 10 ปีต่อมาแสดงให้เห็นชัดเจนขึ้นว่าปริมาณฝนสะสมในอนาคตนั้นมีแนวโน้มลดลงอย่างต่อเนื่องตามการคาดการณ์ของโมเดล แต่จากแผนภาพด้านบนสังเกตได้ว่าพื้นที่สีฟ้าอ่อนหรือกึ่กือซึ่งของความคาดเคลื่อนของค่าปริมาณน้ำฝนในอนาคตนั้นกินพื้นที่เป็นบริเวณกว้างซึ่งอาจจะลดต่ำลงอย่างรวดเร็วหรือพาแนวโน้มของปริมาณน้ำฝนกลับขึ้นสูงได้อย่างไม่แน่นอนในอนาคต

6. บทสรุป

6.1. สรุปผลลัพธ์ที่และตอบคำถามในขอบเขตการศึกษา

จากการวิเคราะห์ปริมาณผนนฟันสะสมรายปีในระยะเวลา 9 ปีย้อนหลังพบว่าแนวโน้มของปริมาณผนฟันสะสมในลุ่มแม่น้ำยมต่อนบนและตอนกลางมีแนวโน้มที่จะลดลงต่อเนื่อง แต่บริเวณลุ่มแม่น้ำยมตอนล่างมีแนวโน้มที่จะเพิ่มขึ้นเพียงเล็กน้อย แต่จากข้อมูลของศูนย์ประมวลวิเคราะห์สถานการณ์น้ำ กรมชลประทาน รายงานว่าปริมาณน้ำผนฟันสะสมเฉลี่ยย้อนหลัง 30 ปีของลุ่มแม่น้ำยมมีปริมาณอยู่ในช่วง 1,000 - 1,200 มิลลิเมตร เมื่อเปรียบเทียบกับปริมาณผนฟันสะสมย้อนหลัง 9 ปีของลุ่มแม่น้ำยมแล้วพบว่า ในแต่ละปีบริเวณลุ่มแม่น้ำยมจะมีปริมาณน้ำผนฟันสะสมน้อยกว่าค่าเฉลี่ยย้อนหลัง 30 ปีประมาณร้อยละ 50 ของปี โดยเฉพาะบริเวณลุ่มแม่น้ำยมตอนล่าง ไม่มีปีใดเลยที่ปริมาณน้ำผนฟันสะสมต่ำไปมากกว่าค่าเฉลี่ยย้อนหลัง 30 ปี

นอกจากนี้เมื่อนำข้อมูลอุณหภูมิเฉลี่ยรายปีในระยะเวลา 9 ปีเปรียบเทียบกับอุณหภูมิเฉลี่ยย้อนหลัง 30 ปี จากรายงานอุตุนิยมวิทยาพบว่าข้อมูลอุณหภูมิเฉลี่ยรายปีบริเวณลุ่มแม่น้ำยมทั้งตอนบน ตอนกลาง และตอนล่าง มีค่ามากกว่าอุณหภูมิเฉลี่ยย้อนหลัง 30 ปี ทั้งสิ้น และเนื่องจากพื้นที่ป่าไม้ในบริเวณลุ่มน้ำยมในระยะเวลา 9 ปีที่ผ่านมา มีจำนวนลดลง และเมื่อเทียบระหว่างพื้นที่ป่าไม้ของบริเวณลุ่มน้ำยมตอนบน ตอนกลาง และตอนล่าง นั้นพบว่าทางตอนล่างมีพื้นที่พักอาศัยมากกว่าพื้นที่ป่าไม้ เป็นอัตราส่วน 56:44 (พื้นที่พักอาศัย:พื้นที่ป่าไม้) ซึ่งเป็นอีกเหตุผลหนึ่งที่ทำให้พื้นที่ตอนล่างมีแนวโน้มประสบภัยธรรมชาติ เช่น ขาดแคลนน้ำได้มากกว่าพื้นที่ส่วนอื่น ทั้งปริมาณผนฟันสะสมที่ยังคงต่ำกว่าค่าเฉลี่ย และพื้นที่พักอาศัยมากกว่าพื้นที่อื่นๆ ความต้องการบริโภคน้ำจึงมากขึ้นส่วนทางกับปริมาณน้ำที่มี

จากปัจจัยของปริมาณผนฟันสะสม อุณหภูมิ และพื้นที่ป่าไม้ตามที่กล่าวมานี้ทำให้สามารถวิเคราะห์ได้ว่าการที่พื้นที่ป่าไม้ลดลง อุณหภูมิเฉลี่ยรายปีสูงขึ้น ประกอบกับปริมาณน้ำผนฟันสะสมที่น้อยกว่าค่าเฉลี่ยอย่างต่อเนื่องหลายปีนับล้านเป็นปัจจัยที่ทำให้เกิดภัยแล้งในบริเวณลุ่มแม่น้ำยม

การคาดการณ์ปริมาณผนฟันสะสมของโมเดลจากพื้นที่ลุ่มน้ำยมทั้ง 3 พื้นที่ในช่วงเวลา 10 ปีต่อมา (ค.ศ. 2021-2030) พบว่าปริมาณผนฟันสะสมทั้ง 3 พื้นที่ในลุ่มแม่น้ำยมมีแนวโน้มลดลงอย่างต่อเนื่องในทุกๆ ปีในช่วงเวลา 10 ปีต่อมาแม้ว่าในระยะเวลา 9 ปีก่อนหน้านี้(ค.ศ.2012-2020)ที่ปริมาณผนฟันสะสมรายปีของพื้นที่ลุ่มน้ำยมตอนล่างเพิ่มสูงขึ้น โดยพื้นที่ลุ่มน้ำยมตอนบนจะมีแนวโน้มปริมาณผนฟันสะสมลดลงเร็วกว่าพื้นที่ในตอนกลางและตอนล่าง แต่พื้นที่ทั้ง 3 ยังคงมีรูปแบบลักษณะปริมาณผนฟันสะสมตามฤดูกาลในลักษณะ

เดิมในทุกๆ ปี เพียงความต่างระหว่างปริมาณฝนสูงสุดและปริมาณฝนต่ำสุดในรอบปีนั้น หรือคือปริมาณฝนสะสมเป็น 0 มิลลิเมตรต่างกันน้อยลงอย่างต่อเนื่อง จากการที่แนวโน้มของปริมาณฝนสะสมลดลงอย่างต่อเนื่องในทุกๆ ปีนี้นำไปสู่ปัญหาภัยแล้งของลุ่มน้ำยมที่รุนแรงขึ้น อย่างต่อเนื่อง

ในส่วนของการวิเคราะห์ความผันผวนของแต่ละพื้นที่บริเวณลุ่มน้ำยม จากการนำกราฟปริมาณน้ำฝนสะสมในระยะเวลา 9 ปี ของลุ่มน้ำยมตอนบน ตอนกลาง ตอนล่าง มาเปรียบเทียบพบว่า บริเวณลุ่มน้ำยมตอนล่าง มีปริมาณน้ำฝนที่แตกต่างจากตอนบนและตอนกลาง เช่น ในระยะเวลา 9 ปี ในลุ่มน้ำยมตอนบนและตอนกลาง ปีที่ปริมาณน้ำฝนสะสมมากที่สุดคือปี ค.ศ. 2014 แต่ในลุ่มน้ำยมตอนล่าง ปีที่ปริมาณน้ำฝนสะสมมากที่สุดคือปี ค.ศ. 2017 นอกจากนี้ บริเวณลุ่มน้ำยมนั้นจะได้รับผลกระทบจากพายุซึ่งเดือนกรกฎาคม ถึงปลายเดือนตุลาคม คือพายุได้ผ่านในมหาสมุทรแปซิฟิกพัดผ่านเข้ามาทางภาคตะวันออกเฉียงเหนือตอนบน ทำให้มีผลกระทบต่อพื้นที่ลุ่มน้ำยมตอนบนและตอนกลาง แต่เนื่องจากภูมิประเทศของตอนบนและตอนกลางมีที่อุ่นเขารายล้อมอยู่จึงทำให้สามารถชะลอน้ำลง และลดความรุนแรงของพายุได้ ส่วนในลุ่มน้ำยมตอนล่างนั้นจะได้รับผลกระทบจากพายุหมุนเขตร้อนในทะเลจีนใต้ที่พัดผ่านเข้ามาทางภาคตะวันออกเฉียงเหนือตอนล่าง ผ่านภาคกลางและภาคตะวันออก ซึ่งลักษณะภูมิประเทศของภาคกลางและภาคตะวันออกไม่มีพื้นที่ที่เป็นที่อุ่นเขารายล้อมอยู่จึงทำให้ขาดปัจจัยสำคัญในการชะลอน้ำและลดความรุนแรงของพายุ ส่งผลให้บริเวณลุ่มน้ำยมตอนล่างจะได้รับผลกระทบจากพายุมากที่สุดทำให้มีปริมาณฝนสะสมที่สูงขึ้นผิดปกติในช่วงเวลาหนึ่ง ดังนั้นหากที่กล่าวมาข้างต้น จึงสามารถวิเคราะห์ได้ว่าบริเวณที่ปริมาณน้ำฝนผันผวนมากที่สุดคือ บริเวณลุ่มน้ำยมตอนล่าง

รูปแบบปริมาณฝนสะสมที่เกิดในช่วงเวลาในแต่ละปีจากการศึกษาปริมาณฝนสะสมจากพื้นที่ลุ่มน้ำยมทั้ง 3 พื้นที่ในช่วงเวลา 10 ปีต่อมา(ค.ศ. 2021-2030) พบว่าช่วงเวลาที่มีฝนสะสมสูงขึ้นจะอยู่ในช่วงระหว่างเดือนพฤษภาคม-เดือนตุลาคม โดยทั้ง 3 พื้นที่มีช่วงของความผันผวนของปริมาณฝนสะสมกล่าวคือในช่วงที่มีปริมาณฝนสะสมสูงในฤดูฝนนั้นจะมีช่วงเวลาหนึ่งที่ปริมาณฝนกลับลดต่ำลงผิดปกติและค่อยเพิ่มขึ้นสูงซึ่งช่วงเวลานั้นคือช่วงประมาณท้ายเดือนมิถุนายน-ต้นเดือนกรกฎาคมเช่นเดียวกันกับพื้นที่ทั้ง 3 ส่วน และจากการที่แนวโน้มของปริมาณฝนลดลงอย่างต่อเนื่องในระยะเวลา 10 ปีต่อมา (ค.ศ. 2021-2030)

ทำให้ช่วงเวลาที่เกิดความผันผวนนั้นลดต่ำลงมากขึ้นตามแนวโน้ม จนเข้าใกล้หรือปริมาณฝนสะสมในช่วงเวลานั้นเท่ากับ 0 มิลลิเมตร หรือภาวะของฝนทึ่งช่วง ซึ่งจะทำให้พื้นที่ในลุ่มน้ำยมทั้ง 3 พื้นที่นั้นพบเจอกับภัยแล้งเพิ่มมากขึ้นในเดือนกรกฎาคมนokหนึ่งจากช่วงเวลาปกติที่มักเกิดภัยแล้ง

แต่อย่างไรก็ติดผลของการคาดการณ์ด้วยโมเดลในการศึกษาครั้งนี้นั้นยังมีช่วงของค่าที่ไม่แน่นอนของปริมาณน้ำฝนในอีก 10 ปีข้างหน้าจึงไม่สามารถยืนยันได้ว่าปริมาณน้ำฝนจะลดลงจริงตามผลการคาดการณ์ เพียงเป็นการวิเคราะห์เพื่อหาแนวโน้มในอนาคตที่อาจจะเกิดขึ้นได้

6.2. เงื่อนไข ข้อจำกัดในการวิเคราะห์ที่ควรทราบ

- ข้อมูลจากลุ่มน้ำยมแบ่งเป็น 3 ตอน คือ ตอนบน ตอนกลาง และตอนล่าง โดยแบ่งตามความสูงของระดับน้ำทะเล ซึ่งในแต่ละตอนได้มีการเลือกข้อมูลจาก 1 สถานี ที่ติดกับแม่น้ำสายหลักของลุ่มน้ำยมมากที่สุด ระยะเวลาของข้อมูลทั้ง 3 สถานี มีดังนี้
 - ลุ่มน้ำยมตอนบน สถานีทต.เชียงม่วน มีข้อมูลตั้งแต่วันที่ 29 กันยายน พ.ศ. 2554 ถึง 31 ธันวาคม พ.ศ. 2563 ระยะเวลา 10 ปี
 - ลุ่มน้ำยมตอนกลาง สถานีออบต.วังชีน มีข้อมูลตั้งแต่วันที่ 3 พฤษภาคม พ.ศ. 2554 ถึง 31 ธันวาคม พ.ศ. 2563 ระยะเวลา 10 ปี
 - ลุ่มน้ำยมตอนล่าง สถานีโรงฆ่าสัตว์ ทม.สวรรคโลก มีข้อมูลตั้งแต่วันที่ 11 เมษายน พ.ศ. 2554 ถึง 31 ธันวาคม พ.ศ. 2563 ระยะเวลา 10 ปี
- การทดสอบเลือกโมเดลหลักที่ใช้ในการศึกษา เลือกจากโมเดลที่มีค่า RMSE น้อยที่สุด โดยโมเดลทั้ง 3 ที่ใช้ทดสอบใช้การตั้งค่าเริ่มต้นของอัลกอริทึม ทดสอบกับ Univariate value หรือก็คือข้อมูลปริมาณฝนสะสมอย่างเดียวเท่านั้น และใช้ข้อมูลจากสถานีทต. เชียงม่วนของพื้นที่ตอนบนเป็นตัวแทนในการทดสอบเลือกโมเดล
- การวิเคราะห์ภัยแล้งในบริเวณลุ่มน้ำยมนี้เป็นการวิเคราะห์จากข้อมูลปริมาณฝนสะสมของพื้นที่เป็นหลัก

6.3. คำแนะนำในการต่อยอดการศึกษา

- วิเคราะห์ปริมาณน้ำฝนของถุ่มน้ำยมหากาความสัมพันธ์กับดัชนีเคลื่อนไหว ลานีญ่า
- นำข้อมูลจากทุกสถานีของถุ่มแม่น้ำในการวิเคราะห์
- วิเคราะห์โครงการพัฒนาของรัฐบาลถึงข้อดีและข้อเสีย และวิธีป้องกันที่มีประสิทธิภาพ
- เปรียบเทียบข้อมูลระหว่างสถานีของสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) กับ กรมอุตุนิยมวิทยา
- ให้คำแนะนำในการรับมือกับภาวะฝนทึ่งช่วงและช่วงที่เกิดความผันผวนของปริมาณฝน รวมถึงระบุค่าของปริมาณฝนที่อาจจะเกิดขึ้น
- ปรับปรุงโมเดลโดยลดช่วงของค่าความไม่แน่นอน หรือพัฒนาโมเดลด้วยอัลกอริทึมอื่น

บรรณานุกรม

- สำนักขลประทานที่ 4 จังหวัดกำแพงเพชร.(ไม่ระบุ).**ภัยแล้ง (Drought).**จาก https://www.ldd.go.th/WEB_UNCCD/new_Nov_54/page1.htm .สืบค้นเมื่อวันที่ 8 มิถุนายน 2564
- ศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา.(ไม่ระบุ) **สถิติภูมิอากาศคลาบ 30 ปี พ.ศ. 2524-2553.** จาก <http://climate.tmd.go.th/statistic/stat30y> สืบค้นเมื่อวันที่ 7 มิถุนายน 2564
- Krish Hariharan.(2563).**Time Series Forecasting — ARIMA vs Prophet.**จาก <https://medium.com/analytics-vidhya/time-series-forecasting-arima-vs-prophet-5015928e402a>. สืบค้นเมื่อวันที่ 1 มิถุนายน 2564
- Logistic Cafe.(2563).**การวิเคราะห์อนุกรมเวลา (Time Series Analysis) คืออะไร ?**จาก <https://www.logisticafe.com/2009/08/time-series-analysis/>. สืบค้นเมื่อวันที่ 7 มิถุนายน 2564
- Nachiketa Hebbar.(2563).**Time Series Forecasting With ARIMA Model in Python for Temperature Prediction.**จาก <https://medium.com/swlh/temperature-forecasting-with-arima-model-in-python-427b2d3bcb53>. สืบค้นเมื่อวันที่ 1 มิถุนายน 2564
- Perth N.(ไม่ระบุ).**ข้อมูล Time Series คืออะไร และทำไม่คนเรียน Data Science ควรสนใจ.**จาก https://blog.datath.com/time-series-forecasting/#bth_thi_8_ARIMA_Models. สืบค้นเมื่อวันที่ 30 พฤษภาคม 2564
- P0011.(2562).**หลักการทำงานของ ARIMA ในการทำนายค่าของ Time Series.**จาก <https://algoaddict.com/blog/54520/arima>. สืบค้นเมื่อวันที่ 30 พฤษภาคม 2564