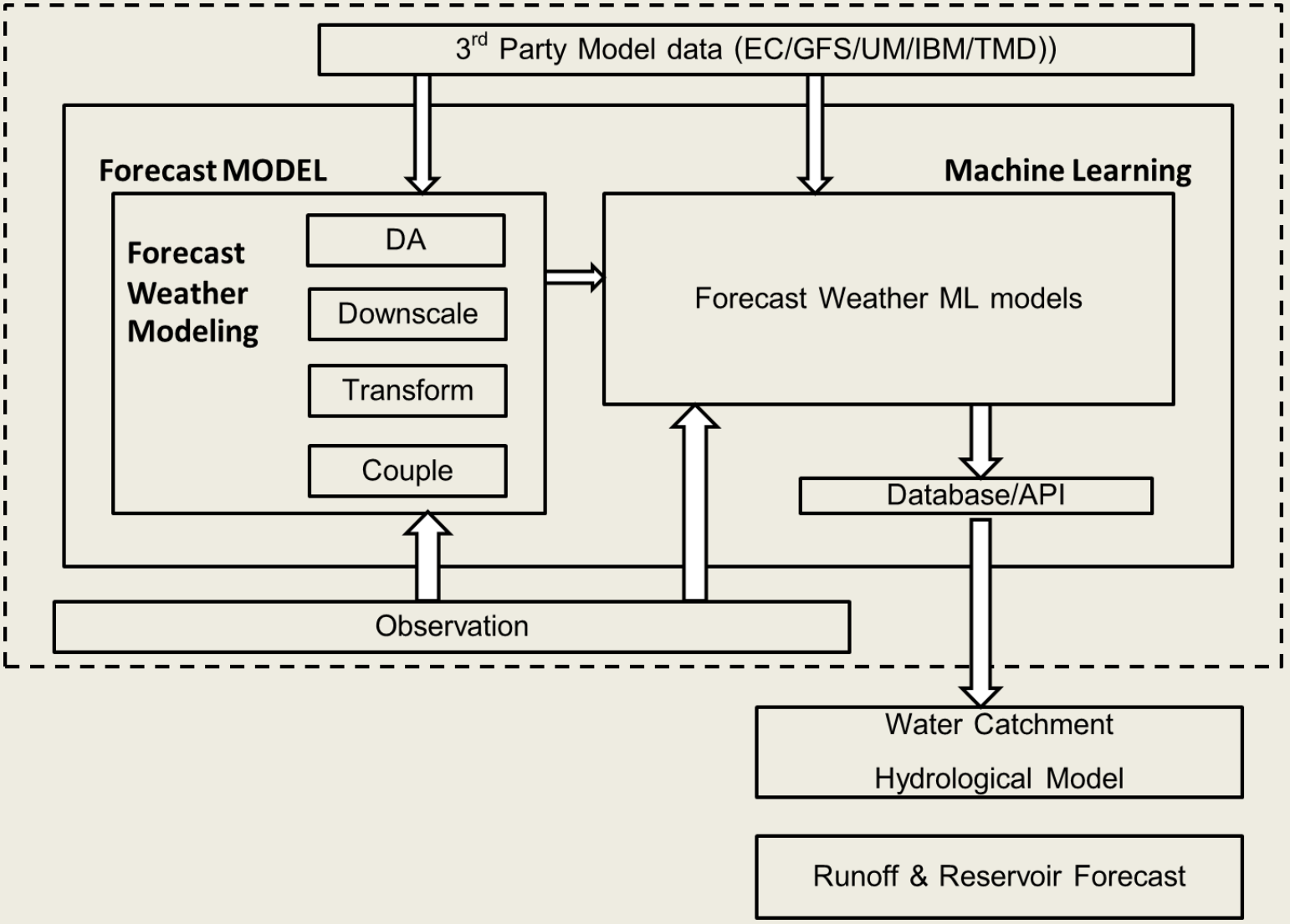
ระบบวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อการวางแผนงานการบริหารจัดการน้ำ และการอบรม

Big data analysis system For water management planning and Training

โดย หน่วยปฏิบัติการวิจัยระบบการจัดการแหล่งน้ำ

แนวคิดการทำงาน

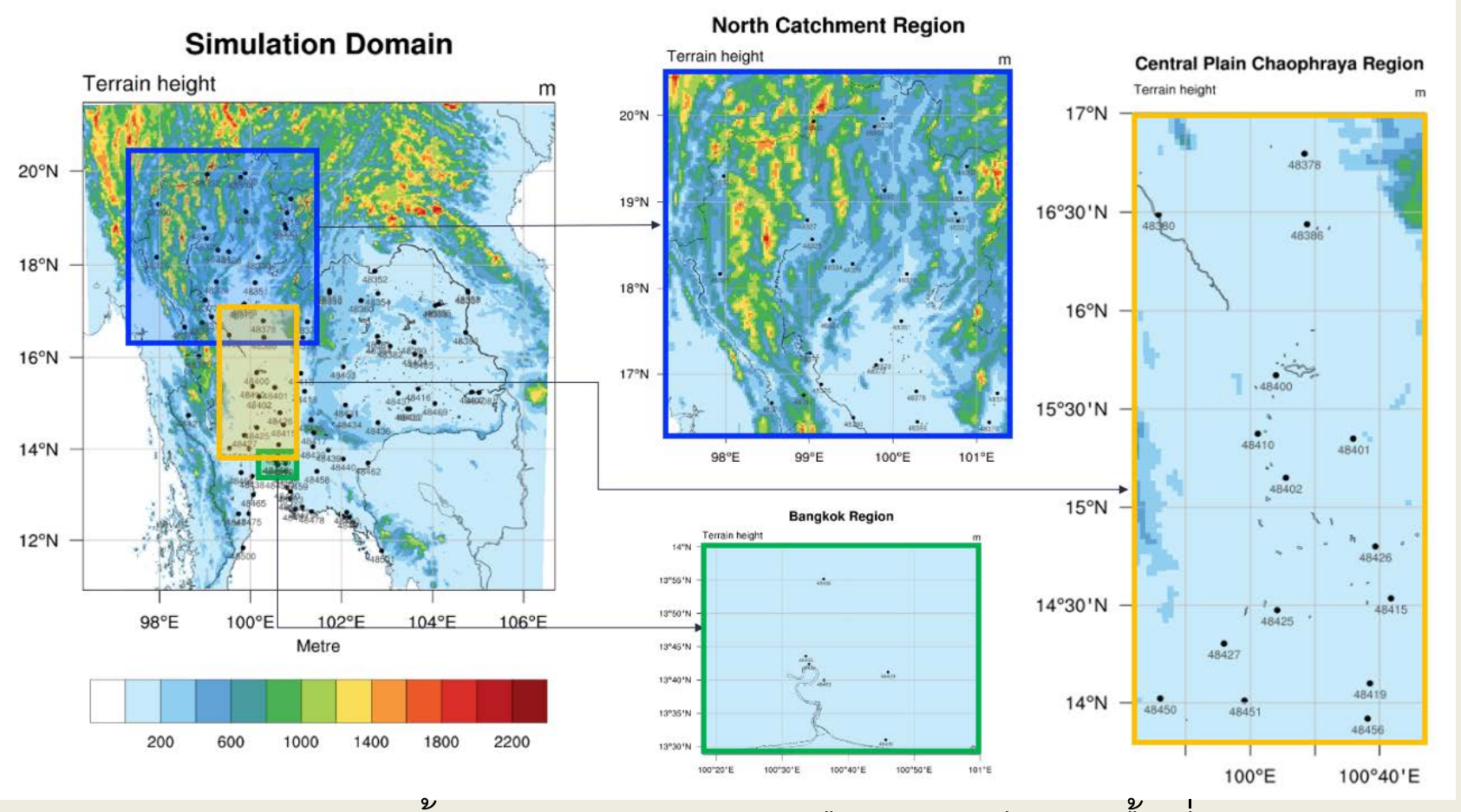


ผลการดำเนินงาน

Technology Solution

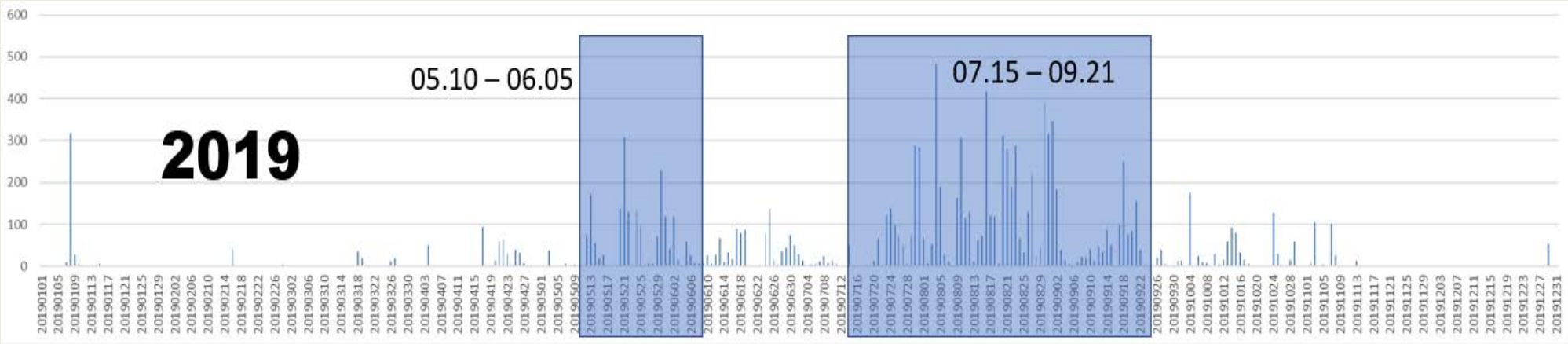
- **Sub-selection of global NWP model**
  ECMWF's Integrated Forecasting System cycle 45r1
- **Optimising regional NWP model**
  Dry season and rain season
- **Utilising observations as much as possible**
  Envision AIoT platform, Thailand side
- **Application of tailored machine learning algorithm**
  Dr Lin Miao's talk in the afternoon
- **High-Performance Computer (HPC)**
  HPC has been used since July 2019 for Thailand domain

ผลการดำเนินงาน



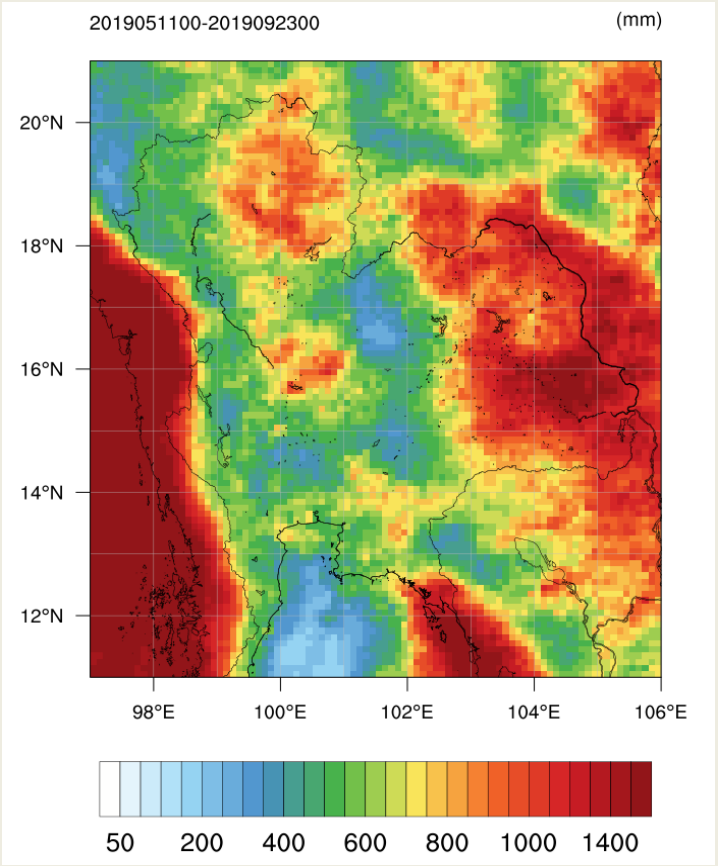
โดเมนการวิเคราะห์สำหรับทั้งสามภูมิภาค: ภาคเหนือ (กล่องสีฟ้า), พื้นที่ธรรมดาภาคกลาง (กล่องสีน้ำตาล) และพื้นที่กรุงเทพฯ (กล่องสีเขียว) ที่มีสถานีอากาศพื้นผิวติดฉลาก

ผลการดำเนินงาน

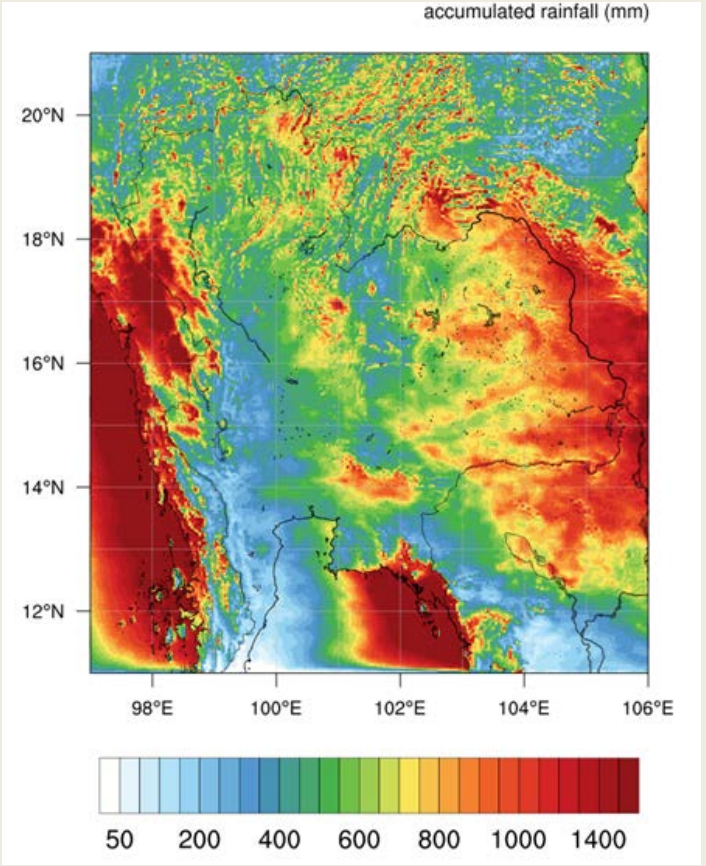


Sub-select ระยะเวลา 3 เดือน สำหรับการเตรียมชุดข้อมูลสำหรับการเรียนรู้ด้วย ML algorithm

ผลการดำเนินงาน



a



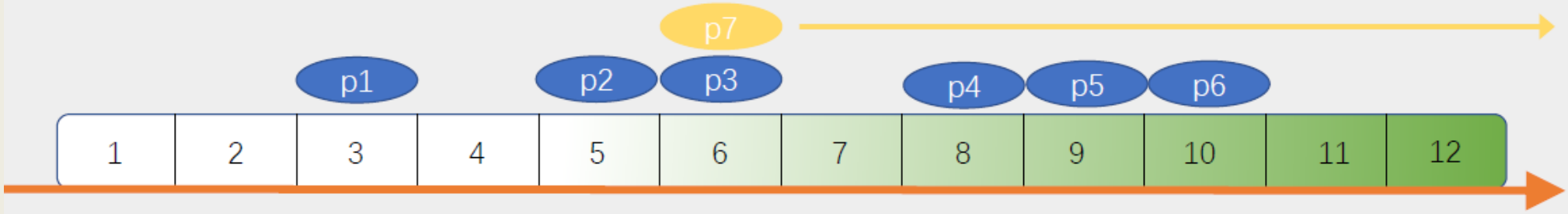
b

ช่วงเวลา 3 เดือน สะสมปริมาณน้ำฝนของ (a) การสังเกตการณ์ดาวเทียมและ (b) การพยากรณ์อากาศ

ผลการดำเนินงาน

Machine Learning Development Timeline

- p1 Preparation: observation data, weather pattern analysis, prepare first workshop
- p2 Build ***temperature*** forecast model pipeline
- p3 Build ***humidity*** forecast model pipeline
- p4 Build ***rainfall*** forecast model pipeline and evaluate/tune ***temperature/humidity*** model
- p5 Build ***wind*** forecast model pipeline and evaluate/tune ***rainfall*** model
- p6 Evaluate/tune ***wind*** model
- p7 Optimize models for a particular area / region



ผลการดำเนินงาน

Observation Requirement

Weather attributes	Learning target	Periods	Frequency	Properties
Rainfall	Station ✓	At least 1 year	Hourly / 3 hourly	Small coverage, accurate
	Radar image	At least 6 months, L3/L2/ Image format	5 minutes / 15 minutes	Large coverage, whole domain, less accurate
	GPM			Self-collected, less accurate than Radar
Temperature	Station ✓	At least 1 year	Hourly	
Humidity				
Wind				

ผลการดำเนินงาน

ข้อจำกัดและโซลูชั่นที่มีศักยภาพ

โครงการได้ยกระดับการสร้างแบบจำลอง NWP ที่ทันสมัยและเทคนิค ML เพื่อดูการคาดการณ์ระยะสั้นของปริมาณน้ำฝนอุณหภูมิความชื้นและลมในความละเอียดสูงมาก แม้ว่าโครงการจะมีระยะเวลาอันสั้น แต่งานในอนาคตบางอย่างยังคงเป็นไปได้ที่จะปรับปรุงการคาดการณ์ ข้อจำกัดอย่างหนึ่งของโครงการปัจจุบันคือปัญหาความไม่แน่นอนยังไม่ได้รับการแก้ไขอย่างเต็มที่ ในการพยากรณ์ของการจำลองความละเอียดสูงมากโดยเฉพาะอย่างยิ่งการพยากรณ์ปริมาณน้ำฝน ถึงแม้ว่าอัลกอริทึม ML จะช่วยได้บ้างโดยการลบความ Bias ออกจากทั้งความเข้มและตำแหน่ง แต่ฐานของการคาดการณ์ก็คือหลังจากการพยากรณ์ที่กำหนดขึ้นเพียงครั้งเดียว วิธีการของ Ensemble เป็นวิธีการแก้ปัญหาที่มีประสิทธิภาพในการจัดการกับความไม่แน่นอนได้ดียิ่งขึ้นโดยดูที่การคาดการณ์หลายครั้งและทำให้ความไม่แน่นอนลดลงมากที่สุด

ผลการดำเนินงาน

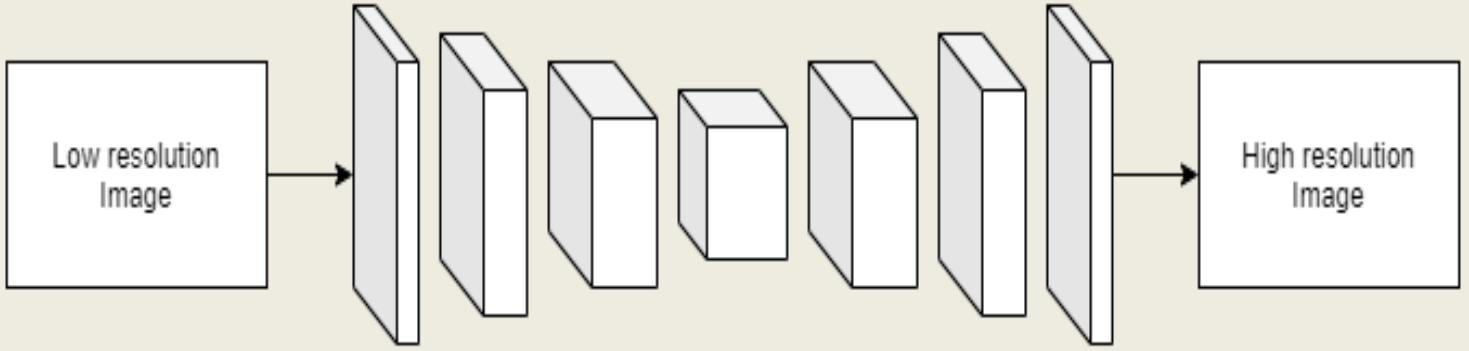
ข้อจำกัดและโซลูชั่นที่มีศักยภาพ (ต่อ)

ข้อจำกัดอีกประการหนึ่งคือคุณภาพของเงื่อนไขเริ่มต้นที่เกี่ยวข้องกับการประยุกต์ใช้เทคนิคการรวบรวมข้อมูล ณ เวลาศูนย์ ในโครงการปัจจุบันการรวบรวมข้อมูลเพื่อปรับปรุงคุณภาพของเงื่อนไขเริ่มต้นนั้นดำเนินการเฉพาะในระดับ Global ซึ่งเราเชื่อว่าการสำรวจในพื้นที่จำนวนมากไม่ได้ถูกใช้งาน เช่น ปริมาณน้ำฝนจากเรดาร์ และความเร็วเชิงมุม ซึ่งเราต้องการข้อมูลความละเอียดสูงในระดับภูมิภาค เพื่อปรับปรุงสภาพเริ่มต้นซึ่งจะช่วยปรับปรุงการพยากรณ์

ผลการดำเนินงาน

การพัฒนา Machine Learning ของนักวิจัย

ในปัจจุบันเทคโนโลยีด้านการทำ Image Processing มีความก้าวหน้าเป็นอย่างมาก โดยเฉพาะการใช้ Convolution Neural Network มาวิเคราห์ภาพถ่าย เช่น การตรวจจับวัตถุภายในภาพ การจำแนกประเภทภาพ เพิ่มความละเอียดรูปภาพ เป็นต้น

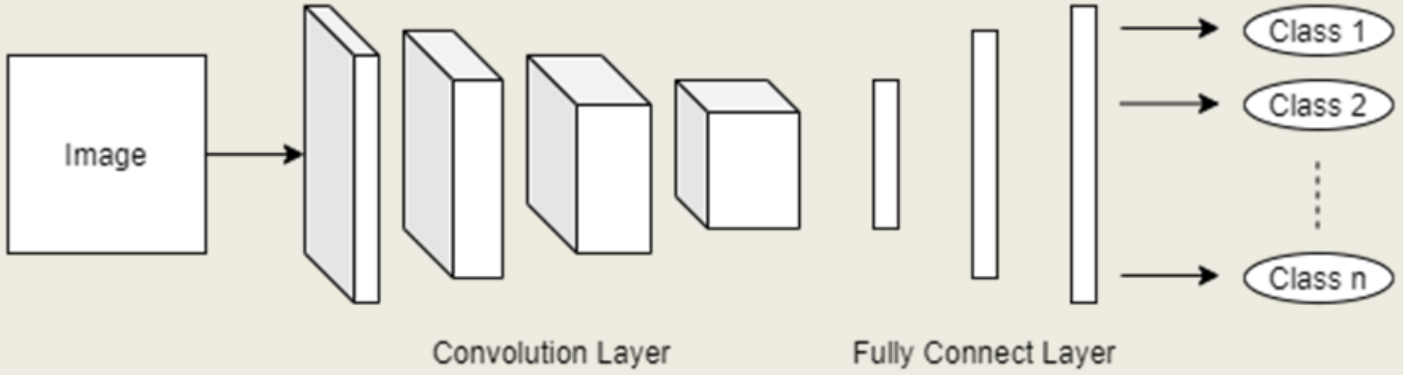


ตัวอย่างการใช้ Convolution Neural Network เพื่อเพิ่มความละเอียดรูปภาพ

ผลการดำเนินงาน

การพัฒนา Machine Learning ของนักวิจัย (ต่อ)

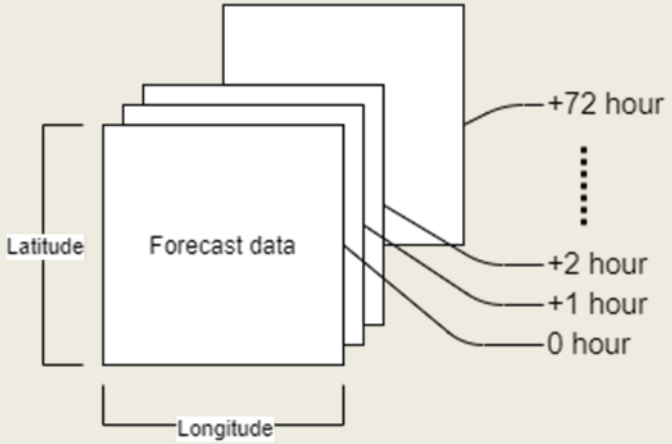
ในการทดลองนี้ เพื่อเพิ่มความแม่นยำของการพยากรณ์ปริมาณฝน เราจะประยุกต์ใช้ Convolution Neural Network เพื่อนำข้อมูลการพยากรณ์อากาศ (Low resolution) มาคำนวนประมาณฝนที่ตกบนพื้นที่ (High resolution) โดยจะแบ่งขั้นตอนการทำงานเป็น 4 ขั้นตอน ดังนี้



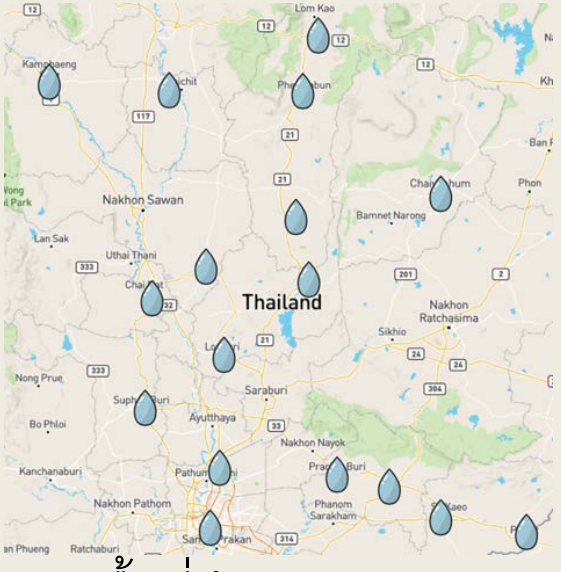
รูปตัวอย่างการใช้ Convolution Neural Network เพื่อการจำแนกประเภทภาพ

ผลการดำเนินงาน

1. **เตรียมข้อมูล** Input ในขั้นตอนนี้จะเป็นการเตรียม Input ที่จะใช้ในการทำนายปริมาณฝนบนพื้นที่ โดยจะนำข้อมูลพยากรณ์อากาศจาก HII ที่เป็น grid ประมาณ 3x3 กม.
   ทุก 1 ชัวโมง ล่วงหน้า 72 ชั่วโมง มาทำการ Preprocess ให้อยู่ในรูปแบบ 3D array คล้ายรูปภาพ โดยพื้นที่ศึกษาจะอยู่บริเวณภาคกลางที่ latitude 13.56102 ถึง 16.990646 และ longitude 99.2199 ถึง 102.77574



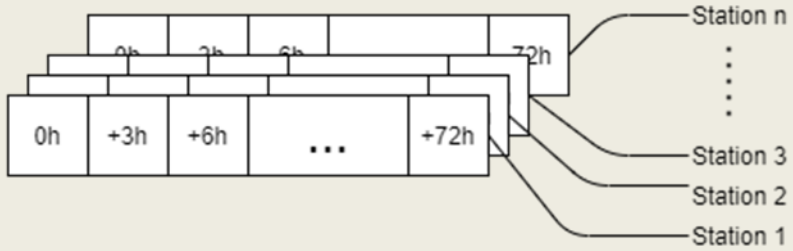
รูปตัวอย่างการ Preprocess ข้อมูลปริมาณฝน ณ. เวลาต่างๆให้อยู่ในรูปแบบ Array 3 มิติ



รูปแสดงพื้นที่ศึกษาและตำแหน่งสถานีอุตุนิยมวิทยา

ผลการดำเนินงาน

2. **เตรียมข้อมูลปริมาณฝนที่จะใช้เป็น Target สำหรับ Neural Network** ในขั้นตอนนี้จะนำข้อมูลปริมาณฝนจาก TMD ย้อนหลัง 72 ชั่วโมง ให้อยู่ในรูปแบบ 2D array และกำหนดเดือนของการพยากรณ์อากาศโดยแปลงเป็นลักษณะ One hot encoding



รูปตัวอย่างการจัดกลุ่มข้อมูลผลการตรวจวัดลักษณะอากาศ

เดือน	1	2	3	4	5	6	7	8	9	10	11	12
มกราคม	1	0	0	0	0	0	0	0	0	0	0	0
กุมภาพันธ์	0	1	0	0	0	0	0	0	0	0	0	0
มีนาคม	0	0	1	0	0	0	0	0	0	0	0	0
เมษายน	0	0	0	1	0	0	0	0	0	0	0	0
พฤษภาคม	0	0	0	0	1	0	0	0	0	0	0	0
มิถุนายน	0	0	0	0	0	1	0	0	0	0	0	0
กรกฎาคม	0	0	0	0	0	0	1	0	0	0	0	0
สิงหาคม	0	0	0	0	0	0	0	1	0	0	0	0
กันยายน	0	0	0	0	0	0	0	0	1	0	0	0
ตุลาคม	0	0	0	0	0	0	0	0	0	1	0	0
พฤศจิกายน	0	0	0	0	0	0	0	0	0	0	1	0
ธันวาคม	0	0	0	0	0	0	0	0	0	0	0	1

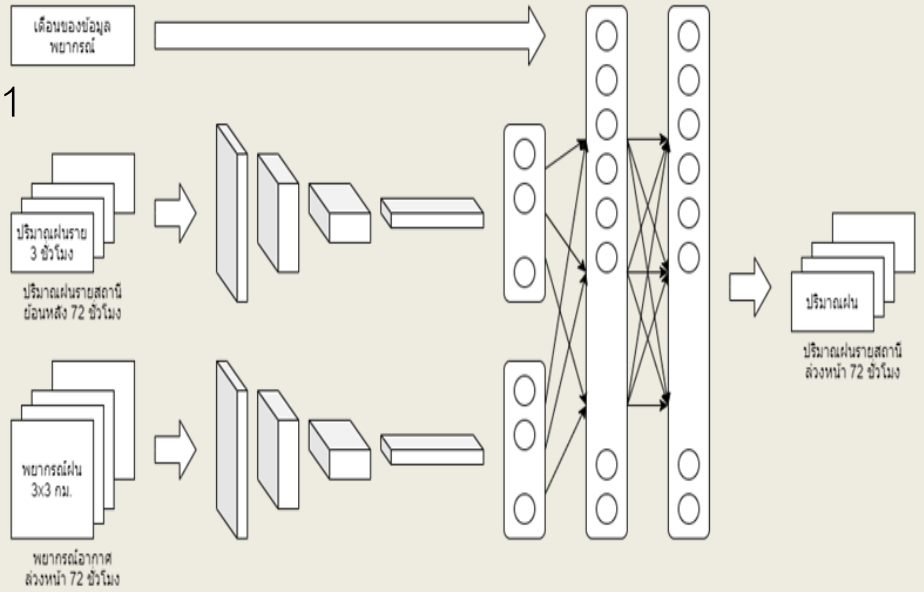
รูปตัวอย่างการแปลงข้อมูลให้อยู่ในลักษณะ One hot encoding

ผลการดำเนินงาน

3. **สร้าง Neural Network สำหรับประมาณปริมาณฝนจุดตามตำแหน่งสถานีอุตุนิยมวิทยา** ในขั้นตอนนี้จะทำการสร้าง Neural Network เพื่อคำนวนปริมาณปริมาณฝนจุดที่ตกบริเวณสถานีอุตุนิยมวิทยา โดยใช้ข้อมูล

1) ข้อมูลพยากรณ์ปริมาณฝนปีความละเอียด 3x3 กม.ทุก 1 ชั่วโมง ล่วงหน้า 72 ชั่วโมง
2) ข้อมูลปริมาณฝนจากสถานีอุตุนิยมวิทยาย้อนหลัง 72 ชั่วโมง
3) ข้อมูลระบุเดือนของพยากรณ์อากาศ

เป็น input สำหรับ Neural Network และใช้ข้อมูลผลการตรวจวัดลักษณะอากาศราย 3 ชั่วโมงจากกรมอุตุนิยมวิทยาที่ช่วงเวลาตรงกับข้อมูลพยากรณ์ปริมาณฝนเป็น Target



รูปอธิบายการทำงานของ Model เพื่อคำนวนหาปริมาณฝนที่ตกบริเวณสถานีอุตุนิยมวิทยา

ผลการดำเนินงาน

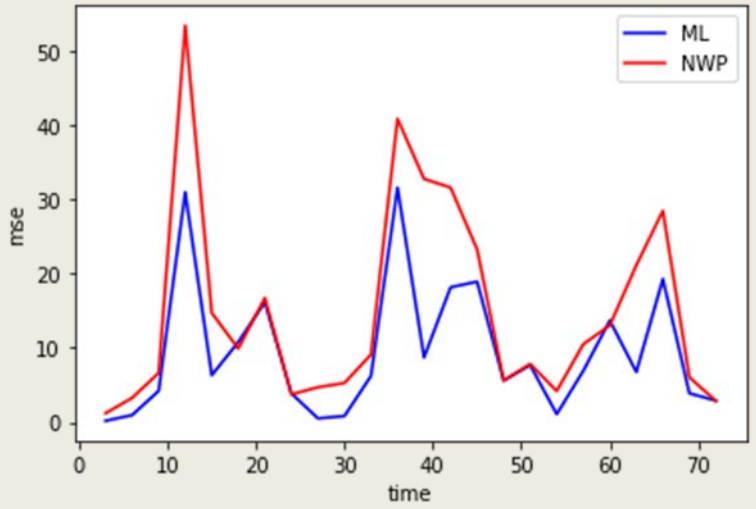
4. **การสอน Neural Network** การสอน Neural Network จะเริ่มจากการแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่

   1) Training set ใช้ข้อมูลวันที่ 18 ถึง 25 ของทุกเดือน
   2) Validation set ใช้ข้อมูลวันที่ 4 ถึง 5 ของทุกเดือน
   3) Test setใช้ข้อมูลวันที่ 11 ถึง 12 ของทุกเดือน

   โดยในการสอนจะใช้ Training set ในการสอน และในระหว่างการสอนจะมีการวัดผลด้วย Validation set เพื่อตรวจสอบว่า model overfit กับชุดข้อมูลหรือไม่ หาก error บนชุดข้อมูล Validation set ไม่ลดลงต่อเนื่อง 50 รอบให้ถือว่า model overfit กับชุดข้อมูลแล้ว ให้ทำการหยุดสอน model และใช้ model version ที่ให้ error บนชุดข้อมูล Validation set น้อยที่สุดเป็น model สำหรับวัดผลต่อไป

ผลการดำเนินงาน

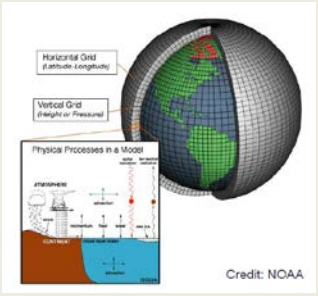
5. **การวัดผล** การวัดผลจะทำการวัดผลบน Test set โดยคำนวนค่า mean square error ของ output จาก model กับปริมาณฝนที่วัดได้จริงจากสถานีอุตุนิยมวิทยา เปรียบเทียบกับปริมาณฝนที่คำนวนจาก grid ที่ใกล้สถานีที่สุดกับปริมาณฝนที่วัดได้จริง

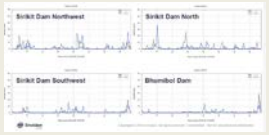


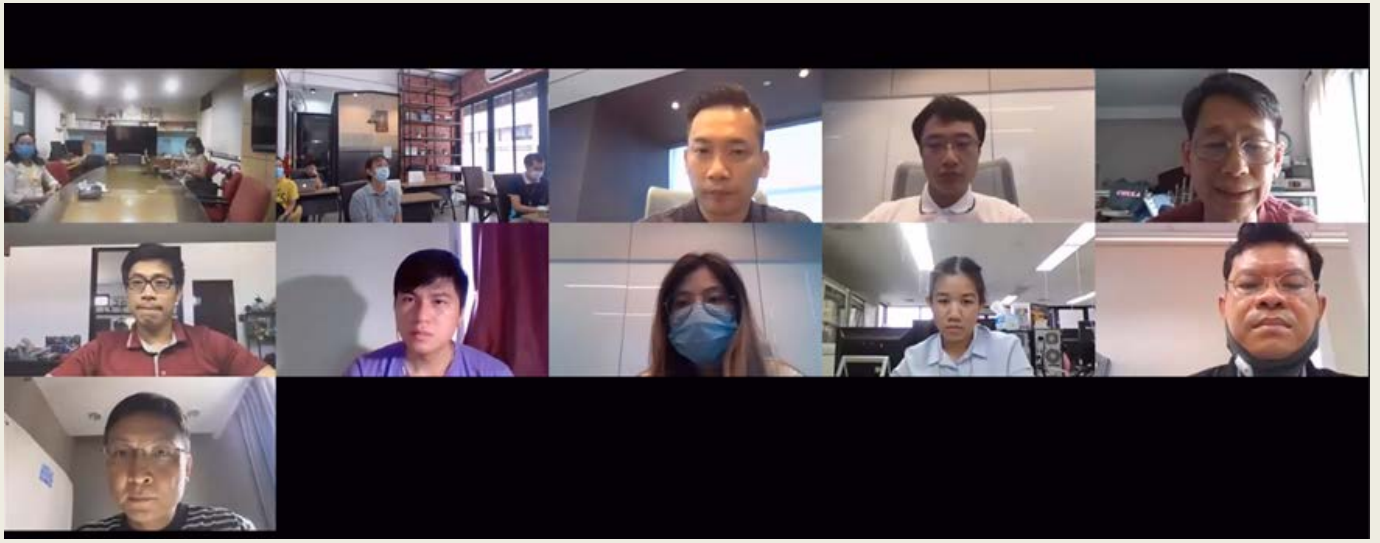
รูปแสดง mean square error ระหว่าง output ของ machine learning กับ nwp

ผลการดำเนินงาน

การอบรม

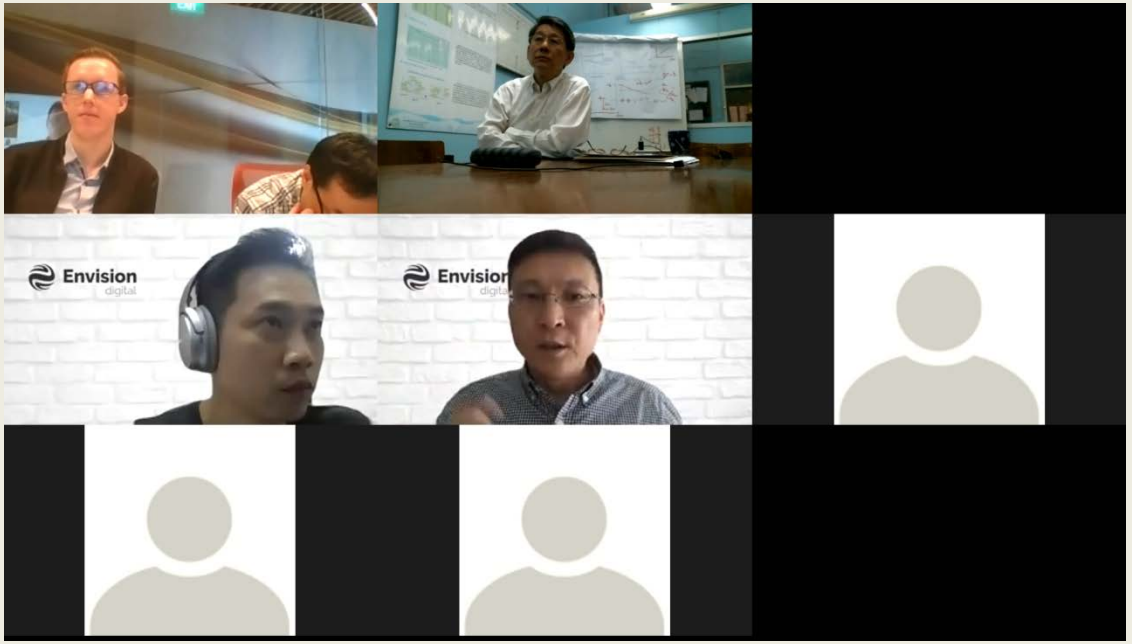






ผลการดำเนินงานปัจจุบันของโครงการมีความเกี่ยวเนื่องโดยตรงกับโครงการ “การพัฒนาระบบวิเคราะห์ข้อมูลฝนขนาดใหญ่ เพื่อการวางแผนงานการบริหารจัดการน้ำ” ซึ่งปัจจุบันโครงการดังกล่าวยังดำเนินการล่าช้ากว่ากำหนดประมาณ 3 เดือน ทำให้ทางโครงการยังไม่สามารถจัดการอบรมได้ตามที่กำหนดแผนไว้ ทั้งนี้ทางโครงการอาจขอปรับเปลี่ยนแนวคิดการอบรมแบบกลุ่มเป็นการอบรมรายบุคคลเพื่อช่วยในการพัฒนาหลักสูตรและบุคลากรที่ได้จะสามารถนำความรู้ที่ได้ไปขยายผลทดแทนการอบรมแบบเดิมที่ติดข้อจำกัดจากปัญหาสถานการณ์โควิด-19

ผลการดำเนินงาน



ID	AIoT Weather Forecast System Project Schedule	Duration
1	Proposed Overall Schedule	360 days
2	Kick Off Meeting	1day
3	Dynamic downscaling the NWP data to 1km by 1km Resolution and compare and share the results monthly starting from 2nd month.	180 days
4	Applying machine learning algorithms to the calibrated NWP data for further improve the accuracy. The result will be compare and share monthly. An User Interface will also be developed to visualize the weather forecast data.	180 days
5	Deliver operational forecast data for 6 months.	180 days
6	Final Report with Recommendations	1 day
7	Final Meeting and Future Roadmap	1 day

Month 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Figure 7 Project Schedule Timeline

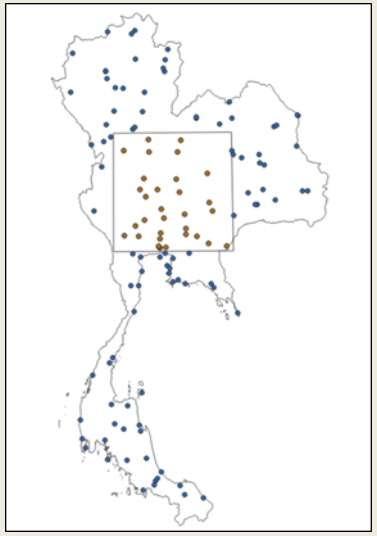
The results will be shared in a report every quarterly and at the end of the project, a final report with recommendations will be compiled and submitted to the screening committee on the 15th month.

A final meeting will be held at 18th month to discuss the final conclusions and future roadmap recommendations,

ทางโครงการได้หารือกับทีมผู้เชียวชาญประเทศสิงคโปรเพื่อหาทางปรับแผนการดำเนินงาน เนื่องจากปัญหา COVID-19 และพยายามเร่งรัดให้เกิดการอบรมในช่วงเดือนตุลาคม โดยมีท่านประธาน SIP และคณะกรรมการตรวจรับร่วมให้ความคิดเห็น

ผลการดำเนินงาน

การดำเนินงานในส่วนทีมวิจัยกับกรมอุตุนิยมวิทยา



1. การเลือกพื้นที่ ละติจูดที่ 13.56102-16.990646 ลองติจูด 99.2199-102.77574 และสถานี ในการศึกษา
2. โมเดลพยากรณ์ที่ใช้พัฒนาควรเป็น การพยากรณ์ 3 วัน ขนาด 6x6 กม. ประมวลผลทุก 1 ชัวโมง
3. การประเมินความแม่นยำใช้มาตรฐานของ WMO
4. ประเมินผลการพยากรณ์ และเพิ่มข้อมูล เรดาร์ กับ ข้อมูลดาวเทียม
5. จัดการอบรมการทำงานและหารือถึงการผลิตข้อมูลให้ผู้ใช้ (ปลายเดือนกันยายน)

ผลการดำเนินงาน

การดำเนินงานในส่วนที่มวิจัยต่างประเทศ (ตุลาคม)

ส่วนที่ 1: ภาพรวมของ Machine Learning Models สำหรับ การพยากรณ์ฝน (1 ชัวโมง)

- สรุปการทบทวนงานวิจัยที่เกี่ยวข้องกับ Machine Learning Models สำหรับการเพิ่มประสิทธิภาพการพยากรร์ฝนด้วย NWP
- หารือเกี่ยวกับข้อดีข้อเสีบของบางกรรมวิธีของ ML

ส่วนที่ 2ก: การออกแบบ Machine Learning Model สำหรับการพยากรณ์ฝน (2 ชัวโมง)

- การออกแบบและสร้าง ML Model ทำอย่างไร
- การฝึก ML Model และอะไรคือเกณฑ์ประเมิน ทำอย่างไร
- กรณีศึกษาของขอบเขตประเทศไทย – หน้าแล้ง และ หน้าฝน
- การปรับปุงเพิ่มเติม Model ทำอย่างไร
- เปิดให้หารือ

ผลการดำเนินงาน

การดำเนินงานในส่วนที่มวิจัยต่างประเทศ (ตุลาคม)

ส่วนที่ 2ข: การออกแบบ Deep Learning Models สำหรับการพยากรณ์ฝน (2 ชัวโมง)

- การออกแบบและสร้าง ML Model ทำอย่างไร
- การฝึกและปรับแต่ง ML Model และอะไรคือเกณฑ์ประเมิน ทำอย่างไร
- กรณีศึกษาของขอบเขตประเทศไทย – หน้าแล้ง และ หน้าฝน
- การปรับปุงเพิ่มเติม Model ทำอย่างไร
- เปิดให้หารือ

ผลที่ได้รับ (ชุดที่ 1-1)

เดือนที่	กิจกรรม (activities)	ผลที่คาดว่าจะได้รับ (outputs)
3 เดือนที่ 1	1. พัฒนาชุดข้อมูลเพื่อพัฒนาระบบการวิเคราะห์	1. ระบบประมวลและการวิเคราะห์
3 เดือนที่ 2	1. จัด module การวิเคราะห์ ในเรื่องที่ 1 การประมาณฝนตก 2. พัฒนาการลดทอนขนาดโมเดล NWP จากของทีมวิจัยและจากกรมอุตุนิยมวิทยาเพื่อเปรียบเทียบและแบ่งปันข้อมูล	1. การประมาณฝนตก(พื้นที และความเข้มของฝน) 2. เปรียบเทียบและแบ่งปันข้อมูลการพยากรณ์

ผลที่ได้รับ (ชุดที่ 1-2)

เดือนที่	กิจกรรม (activities)	ผลที่คาดว่าจะได้รับ (outputs)
3 เดือนที่ 3	1. เชื่อมโยงในเรื่องที่ 2ปรับปรุงข้อมูลฝนทำนายให้ถูกต้องมากขึ้นโดย Machine learning Algorithms	1. การประมาณฝนตก(พื้นที และความเข้มของฝน) 2. เปรียบเทียบและแบ่งปันข้อมูลการพยากรณ์ 3. Machine Learning Algorithms ที่เพิ่มความถูกต้องใน การพยากรณ์

ผลที่ได้รับ (ชุดที่ 1-3)

เดือนที่	กิจกรรม (activities)	ผลที่คาดว่าจะได้รับ (outputs)
3 เดือนที่ 4	1. เชื่อมโยงในเรื่องที่ 3 การประมาณน้ำท่าไหลเข้าเขื่อน (เพื่อบรรลุเป้าหมายการเพิ่มเก็บกักน้ำ15 %) 2. อบรมถ่ายทอดความรู้ผ่าน การประชุมเชิงปฏิบัติการ 4 ครั้ง	1. เชื่อมโยงกับโครงการการปล่อยน้ำเขื่อน 2. ถ่ายทอดความรู้สู่ผู้เกี่ยวข้อง

หมายเหต : จากการเปลี่ยนแปลงของเขตงานของ ประธาน SIP ตัวหนังสือ แดง ทางทีมวิจัยกำลังหารือในการปรับกิจกรรมเพื่อความเหมาะสม

ผลที่ได้รับ (ชุดที่ 2-1)

เดือนที่	กิจกรรม (activities)	ผลที่คาดว่าจะได้รับ (outputs)
3 เดือนที่ 1	1. พัฒนาหลัดสูตร การอบรมครั้งที่ 1 เรื่องการพยากรณ์และการจัดเตรียมข้อมูล	1. หลักสูตรเรื่องการพยากรณ์และการจัดเตรียมข้อมูล
3 เดือนที่ 2	1. จัดอบรมครั้งที่ 1 เรื่องการพยากรณ์และการจัดเตรียมข้อมูล 2. พัฒนาหลักสูตร การจัดอบรมครั้งที่ 2 เรื่องเทคนิค Downscaling ข้อมูลฝน	1. การอบรมเรื่องการพยากรณ์และการจัดเตรียมข้อมูล 2. หลักสูตรเรื่องเทคนิค Downscaling ข้อมูลฝน

ผลที่ได้รับ (ชุดที่ 2-2)

เดือนที่	กิจกรรม (activities)	ผลที่คาดว่าจะได้รับ (outputs)
3 เดือนที่ 3	1. จัดอบรมครั้งที่ 2 เรื่องเทคนิค Downscaling 2. พัฒนาหลักสูตร การอบรมครั้งที่ 3 เรื่องการใช้ Machine Learning ในการพยากรณ์ฝน	1. การอบรมเรื่องเทคนิค Downscaling 2. หลักสูตรเรื่องการใช้ Machine Learning ในการพยากรณ์ฝน
3 เดือนที่ 4	1. จัดอบรมครั้งที่ 3 เรื่องการใช้ Machine Learning ในการพยากรณ์ฝน	1. เชื่อมโยงกับโครงการการปล่อยน้ำเขื่อน 2. ถ่ายทอดความรู้สู่ผู้เกี่ยวข้อง

หมายเหตุ : จากการเปลี่ยนแปลงของเขตงานของ ประธาน SIP ตัวหนังสือ แดง ทางทีมวิจัยกำลังหารือในการปรับกิจกรรมเพื่อความเหมาะสม

ขอบคุณครับ