Study of AIoT Weather Forecast System Technology in Thailand

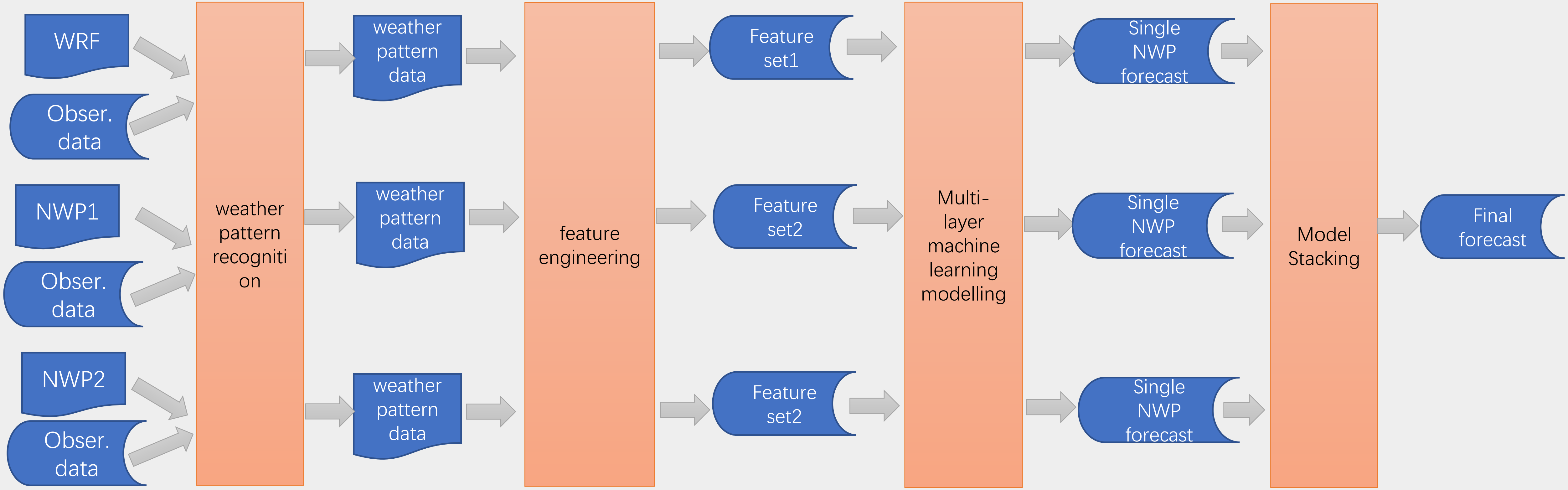
Machine Learning

Dr. Miao Lin

Agenda / Monthly Report Template

- Preparation / Misc
  - Overall architecture / plan
  - Observation data
  - Machine learning workshop
- Model Development
  - Temperature / humidity model : %
  - Rainfall model: %
  - Wind model: %
- Monthly Evaluation
  - Latest month evaluation
  - Performance comparison (comparison with other NWPs: IBM, EC)
  - Explanations, issues studied etc.
- To-dos

Overall Architecture



Solution overview: Temperature / Humidity Forecast

- Input data
  - NWP Data
    - Global forecast: GFS, EC
    - Our own forecast: WRF
    - Weather attr. from surface level
  - Observation data
    - Thailand weather station data
- Target data
  - Use station data as the learning target
- Weather data clustering and preprocessing
  - Weather pattern clustering
- Feature set
  - Common features
  - Features at surface level
- Single layer modelling
  - First layer
    - Modeling temperature directly

Solution overview: Rainfall Forecast

- Input data
  - NWP Data
    - Global forecast: GFS, EC
    - Our own forecast: WRF
    - Weather attr. from various level
  - Observation data
    - GPM: global rainfall data
    - Thailand weather station data
    - Radar image data

- Target data
  - Use GPM as the learning target
  - Combine GPM and station data

- Weather data clustering and preprocessing
  - GPM smoothing
  - Map station data and GPM data

- Feature set
  - Common features
  - Features at various pressure level
  - Features between different level

- Two layers of modelling
  - First layer
    - Modeling rainfall probability
  - Second layer
    - Modeling rainfall volume

Solution plan – Main principles

- **One model per weather attribute (at least)**
  - Rainfall, temperature, humidity, wind: 4 different models

- **Pipeline first, accuracy later**
  - Use less challenging attribute to build the machine learning pipeline
  - Difficulty order
    - rainfall > wind > temperature = humidity
  - Deploying order
    - temperature & humidity -> rainfall -> wind

- **Improvement strategy**
  - **Model retrain / tuning**
    - Use latest / most relevant observation
    - Tune input weather attributes
    - Design model per area per season
  - **New model design / development**

Technical Stack for machine learning components

- Machine learning techniques
  - Linear modelling
  - Tree-based modelling
  - Time series modelling
  - Deep learning models
    - CNN, LSTM
    - ConvLSTM, TrajGRU
    - 3D Conv in CNN
    - GAN
  - Model ensemble
  - Feature selection / reduction

- Other techniques
  - Data query: *sql, mongoDB*, etc
  - Working environment: *Linux, Bash*
  - Code review / management: *git*
  - Python environment: *miniconda*
  - Regular forecast job: *crontab*

Solution Plan – development timeline

p1 Preparation: observation data, weather pattern analysis, prepare first workshop

p2 Build ***temperature*** forecast model pipeline

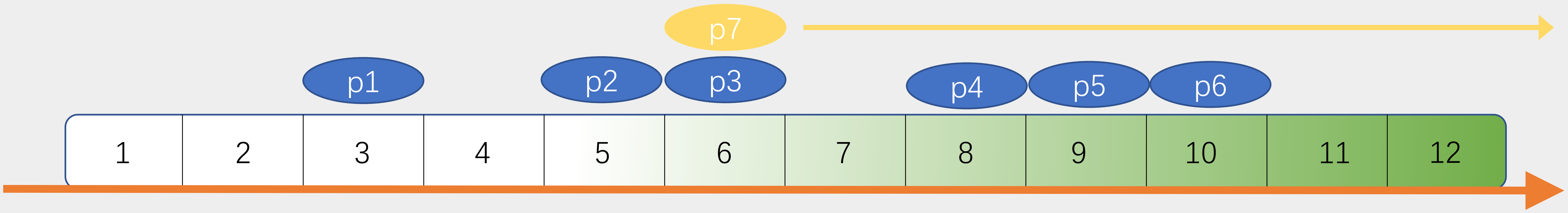
p3 Build ***humidity*** forecast model pipeline

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

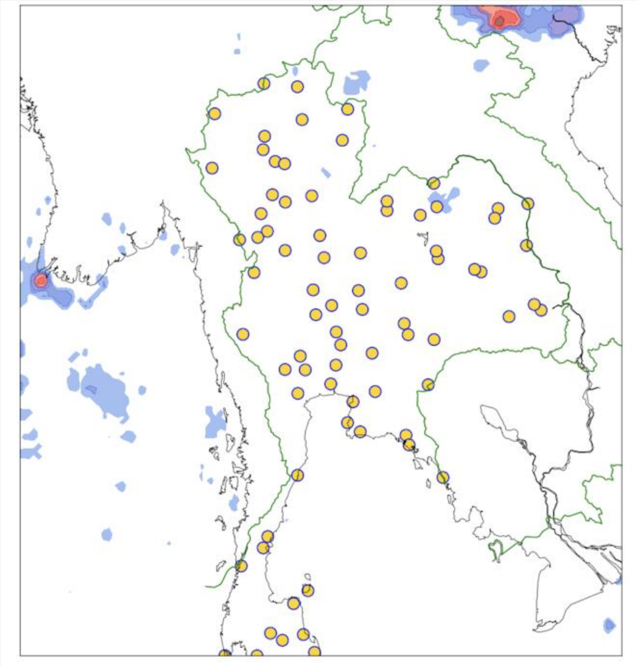


Observations: general requirements

Weather attributes	Learning target	Periods	Frequency	Properties
Rainfall	Station	At least 1 year	Hourly / 3 hourly	• Small coverage: a few Grid points • Most accurate
	Radar data or image	At least 9 months in rainy season (July, Aug, Sep 2017~2019)	5 minutes	• Large coverage • Whole domain • Less accurate
	GPM		30 minutes	• From AIoT platform • Good coverage gridded product • Less accurate
Temperature / humidity / Wind	Station	At least 1 year	Hourly	
Upper air sounding		At least 1 year	6 or 12 hourly	• Accurate upper air measurement • Less frequent • Poor coverage

Observations analysis (1)

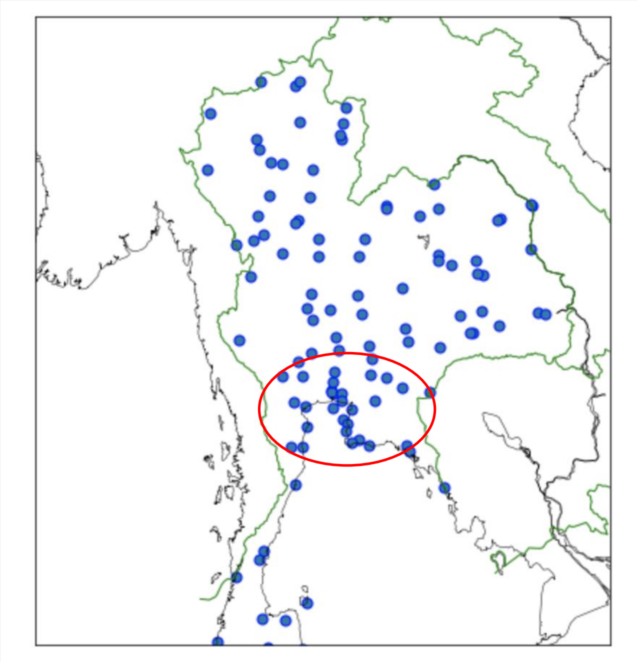
data	No. of Stations	Periods	Freq.	Usage	Issues
tmd_weather3Hours	126	1 month	3 hour	• ML model learning target • evaluation	Short periods, better more than 6 months



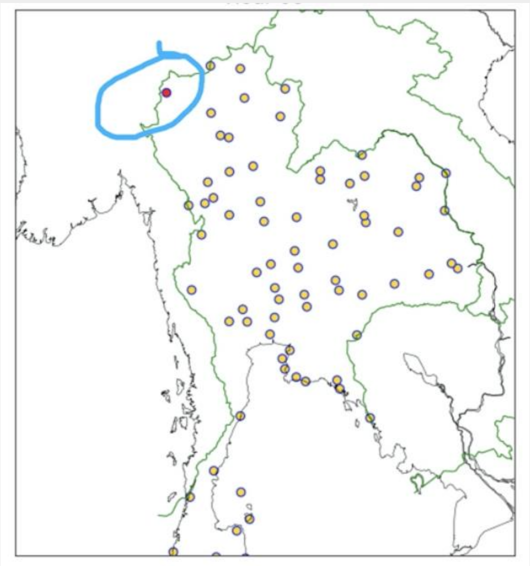
Obs. from website

Obs. from TMD

TMD has more stations!



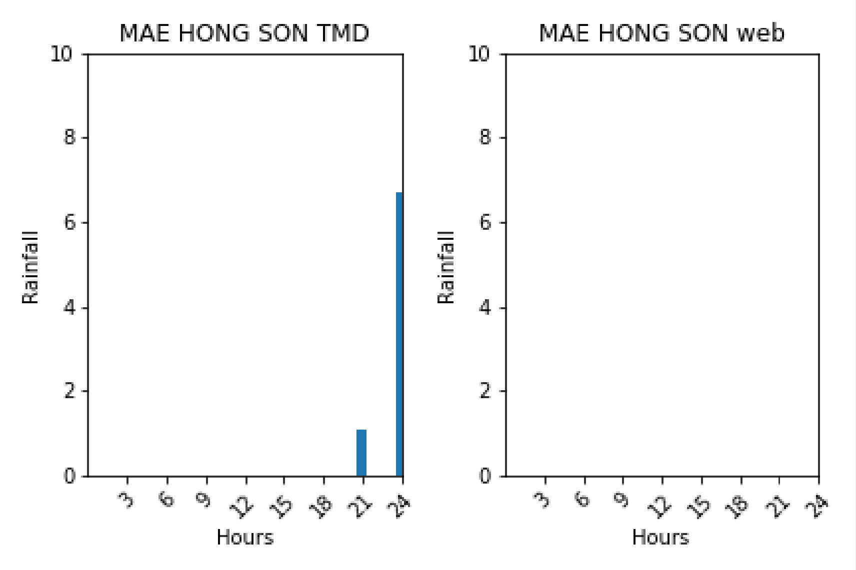
Observations analysis (2)



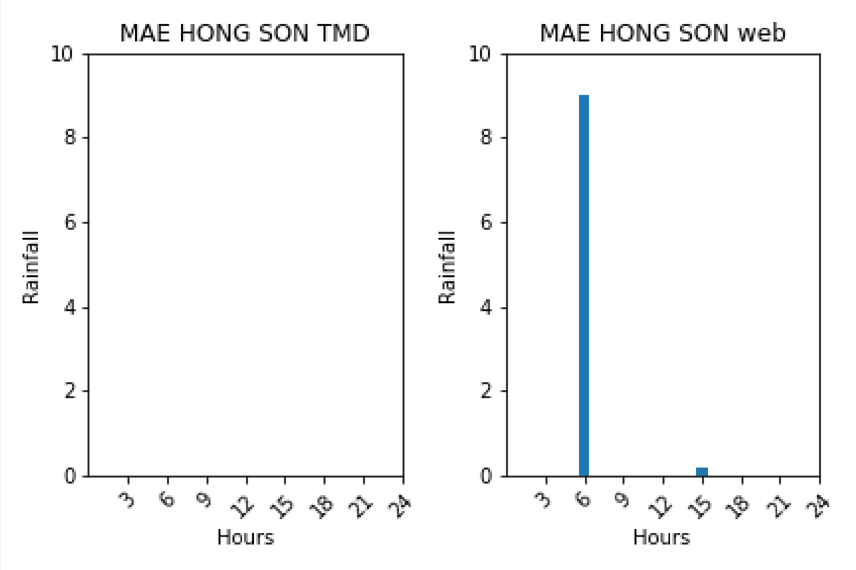
Rainfall comparison between TMD data and weather station data from website

- http://www.aws-observation.tmd.go.th/web/reports/weather_minute.asp
- Use Mae Hong Son as an example (in blue circle)
- Some inconsistency for raining cases
- ***Suggestion***: get more data for long term comparison

2019-12-28 TMD vs web



2019-12-29 TMD vs web



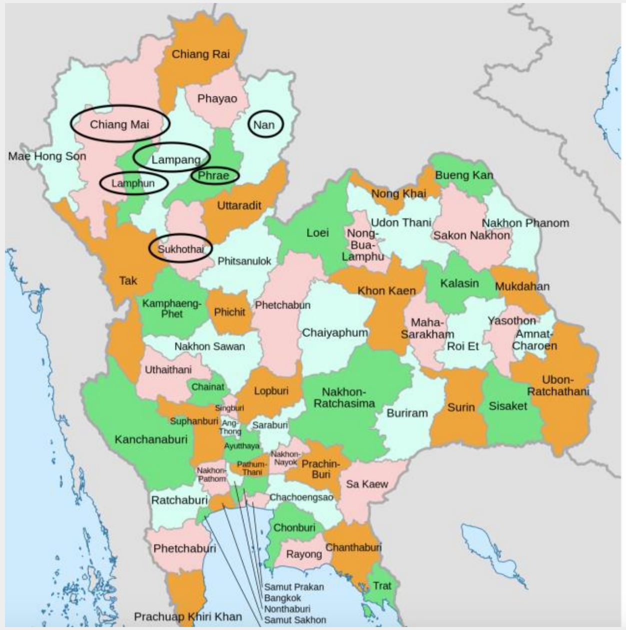
Observations analysis (3)

From TMD

data	No. of Stations	Periods	Freq.	Usage
rain_tmd_2017-2019	12	3 years: 2017-2019	• Daily • One record per day	• Evaluation only

Issues:

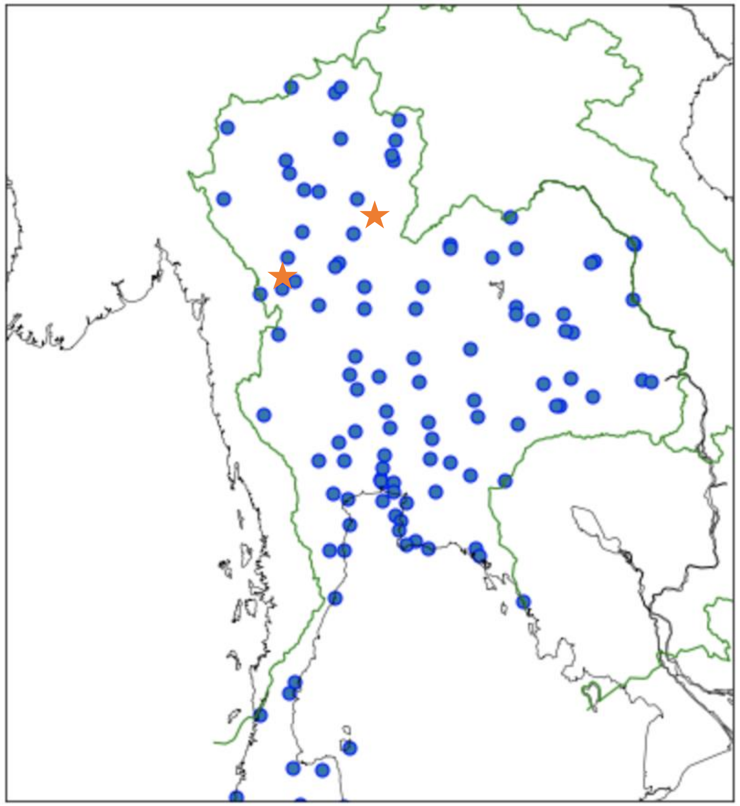
- Low frequency, cannot use in modelling, better hourly / 3 hourly
- No lat/lon for each station



Observations analysis (4)

Observation data for two dams

	Sirikit dam catchment	Bhumibol dam catchment
Location	• Lat: 17.76 • Lon: 100.56	• Lat: 17.24 • Lon: 98.97
File TMD 2017-2019	Cannot verify	Cannot verify
File TMD 3 hours	• No nearby record • Nearest one is UTTARADIT (17.61, 100.1) • ~70km	• Station: 48377 (17.24, 99.002), ~4km • No raining from 2019-12-09 to 2020-01-09
File Web	• No nearby record • Nearest one same as above	• Station 6: same • No raining from 2019-12-09 to 2020-01-09
Remarks	• Need to have a nearby weather station ~10km • With raining period records • From July to Sep yearly	• With raining period records • From July to Sep yearly



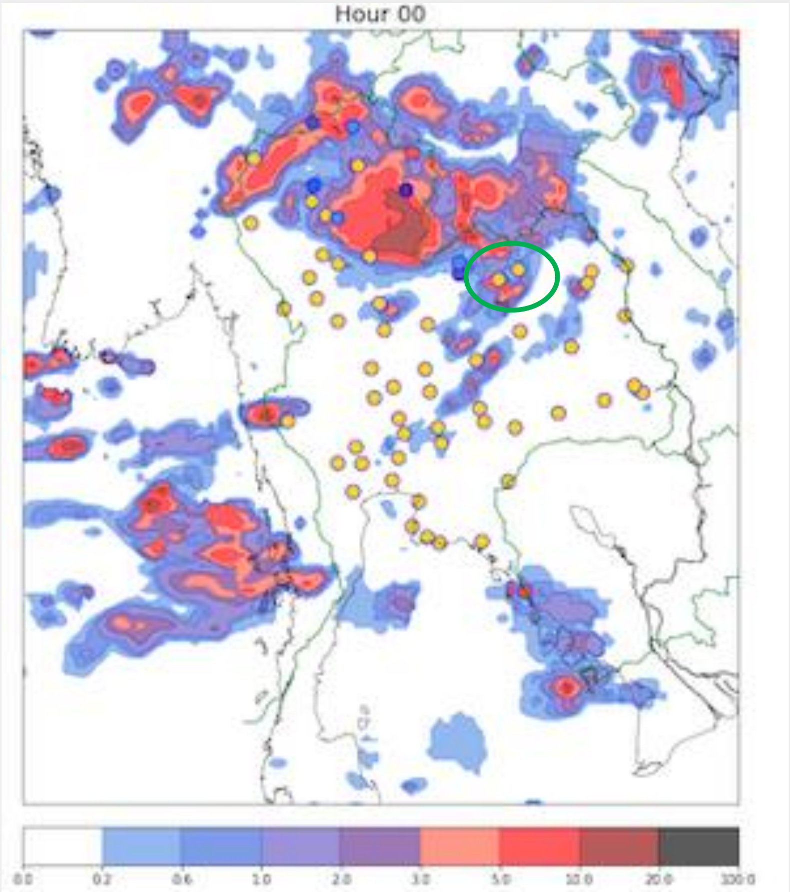
GPM calibration

GPM
- Good coverage
- Whole domain
- Less accurate

Weather station
- Small coverage
- A few grid points
- More accurate

New observation data?

Background: GPM data
Station (circles):
no rain raining



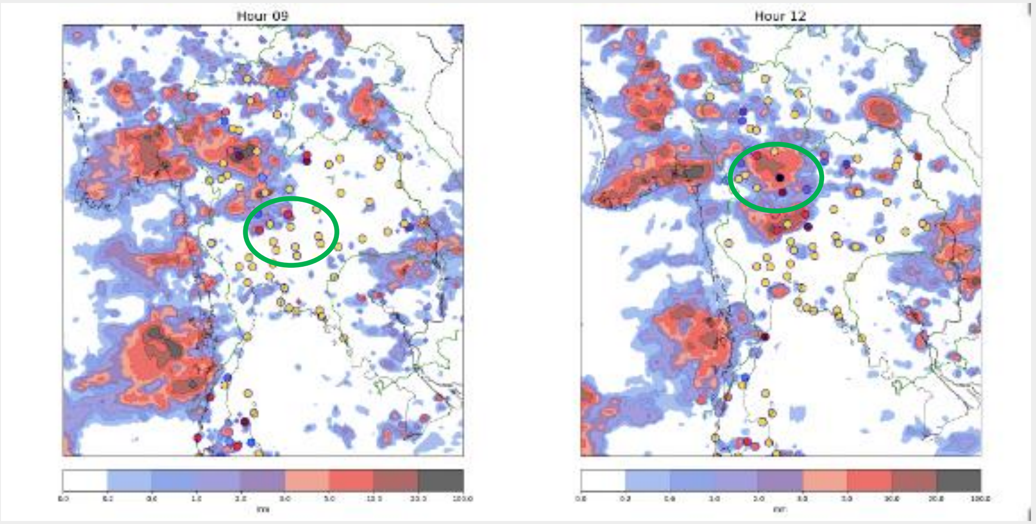
GPM calibration: methodology

- Correct on hour interval basis can offset geographic error
  - 3, 6, 8 and 24 hours summation
- Match distributions
  - Filter small rainfall – cutoff point
- Region to point
  - Add more nearby rainfall information to the points, e.g., 5*5, 7*7, 9*9
- Linear regression
  - Principal component analysis
- Neural network model
  - Learn region feature to point

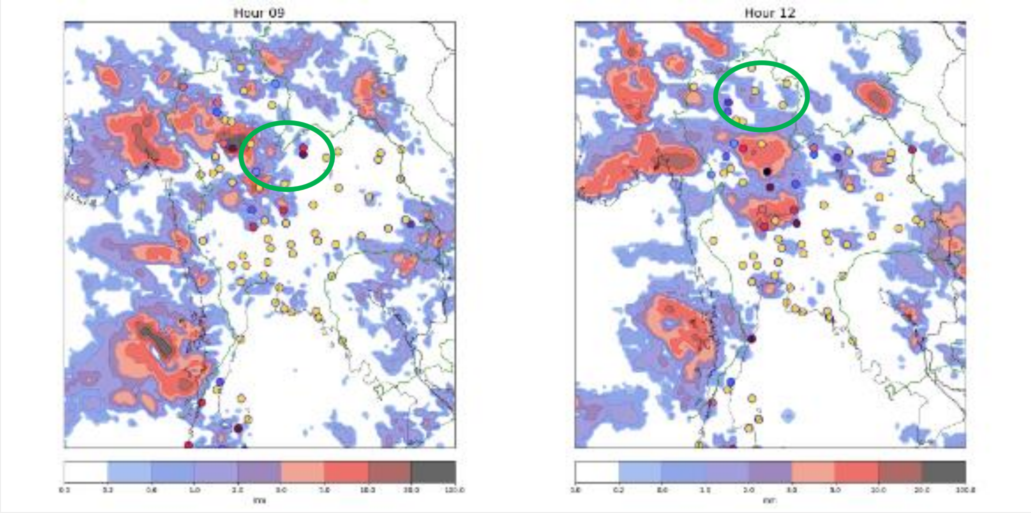
GPM calibration: results

Hour aggregation	GPM loss vs station	ML loss vs station
3	7.48	4.86
6	12.94	7.59
8	16.03	8.93
24	32.58	16.76

Before calibration



After calibration



Machine Learning workshop

Topic 1: use machine learning in weather forecast overview

- Weather forecast overview
- Why / How to use machine learning in weather forecasts?
- Performance overview of using machine learning in weather forecast
- Customer uses cases
- Q&A

Topic 2: case study of weather forecast using machine learning models

- Case study 1
  - rainfall forecast using multi-layer XGBoost
  - 25 minutes
- Case study 2
  - Wind speed forecast using deep learning models
  - 25 minutes
- Q&A

Thank You

Appendix

Project Schedule

			Month																	
ID	AIoT Weather Forecast System Project Schedule	Duration	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
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																		

Communication Protocol

Communication Methods: LINE, Email, Zoom Meeting

- Downscaling – Sun Xiangming, xiangming.sun@envision-digital.com
- Machine Learning – Lin Miao, miao.lin@envision-digital.com
- General / Commercial matters – Henry Tay, henry.tay@envision-digital.com

  – Tony Song, guiting.song@envision-digital.com