Application of Machine Learning in Massive Rainfall Data Mining: Taking Shenzhen as an Example

Lei Li [a,b,c,f], Yuan-Yuan Liu [d], Wen-Hai Zhang [e]

[a] School of Atmospheric Sciences, Sun Yat-Sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai)
[b] Guangdong Provincial Observation and Research Station for Climate Environment and Air Quality Change in the Pearl River Estuary
[c] Key Laboratory of Tropical Atmosphere-Ocean System (Sun Yat-sen University), Ministry of Education
[d] China Institute of Water Resources and Hydropower Research
[e] Shenzhen Academy of Severe Storms Science
[f] Shenzhen National Climate Observatory

2023-09-13

Content

1. Background Information
2. Real-Time Monitoring Products
3. Data Mining based on Machine Learning
4. Outlook

Content

1. Background Information
2. Real-Time Monitoring Products
3. Data Mining based on Machine Learning
4. Outlook

The Characteristics of Shenzhen

- The first Special Economic Zone in China
- A young and vivid city
- An immigrant city
- Trade Center in China

The Figures of Shenzhen

- Economy: Gross Domestic Product 3.24 trillion CNY (0.44 trillion USD)
- Population: over 20 million
- Vehicle: 3.2 million (Densest in China)

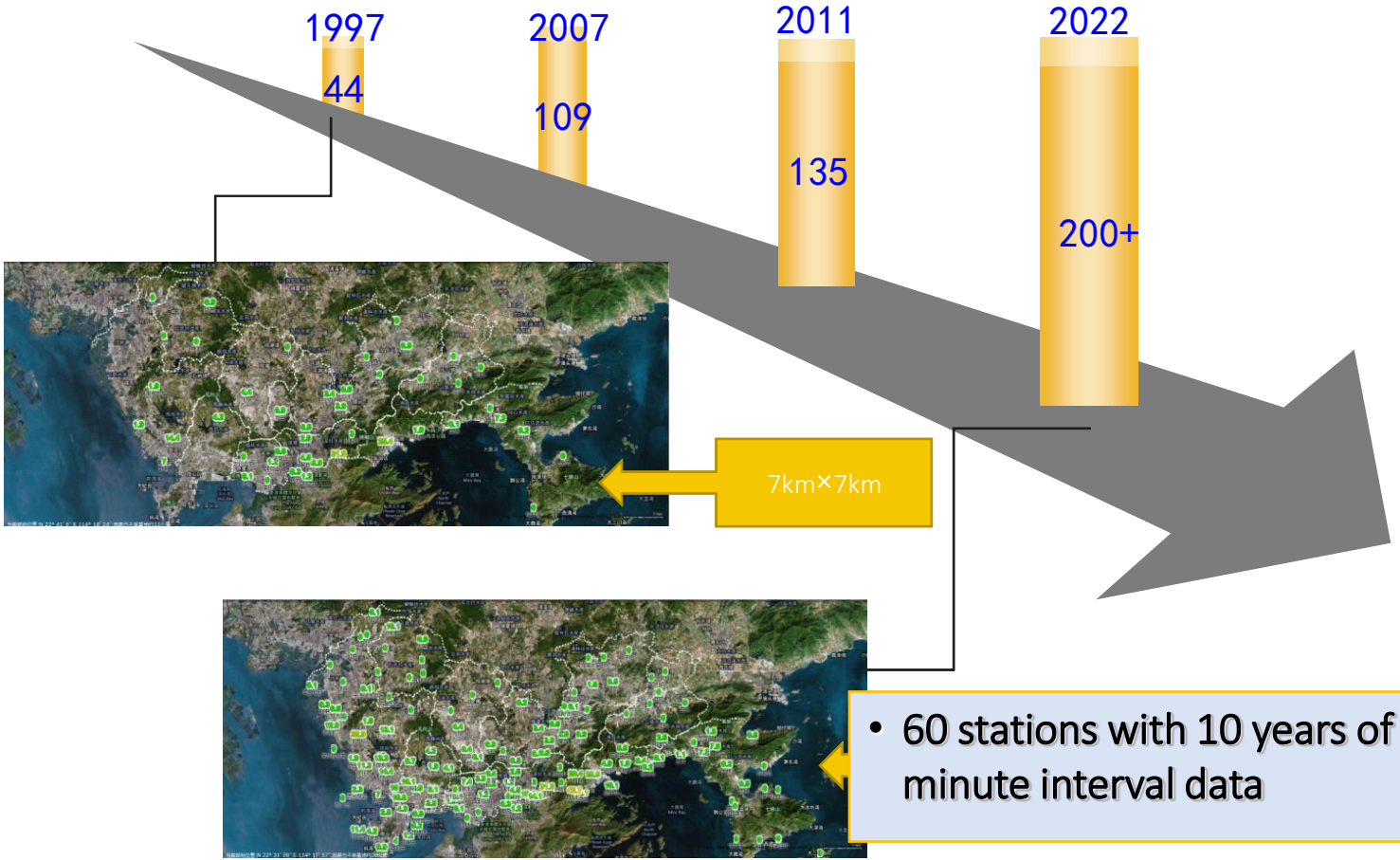
Typhoon, Monsoon Rainstorm, Severe Convection

Targets of Meteorological Observation

- Providing sufficient, accurate and real-time information for taking measure to protect our citizen from disastrous weathers

- Providing long-term, stable and continuous observed data to understand urban climate characteristics in Shenzhen

Auto Weather Stations Network in Shenzhen



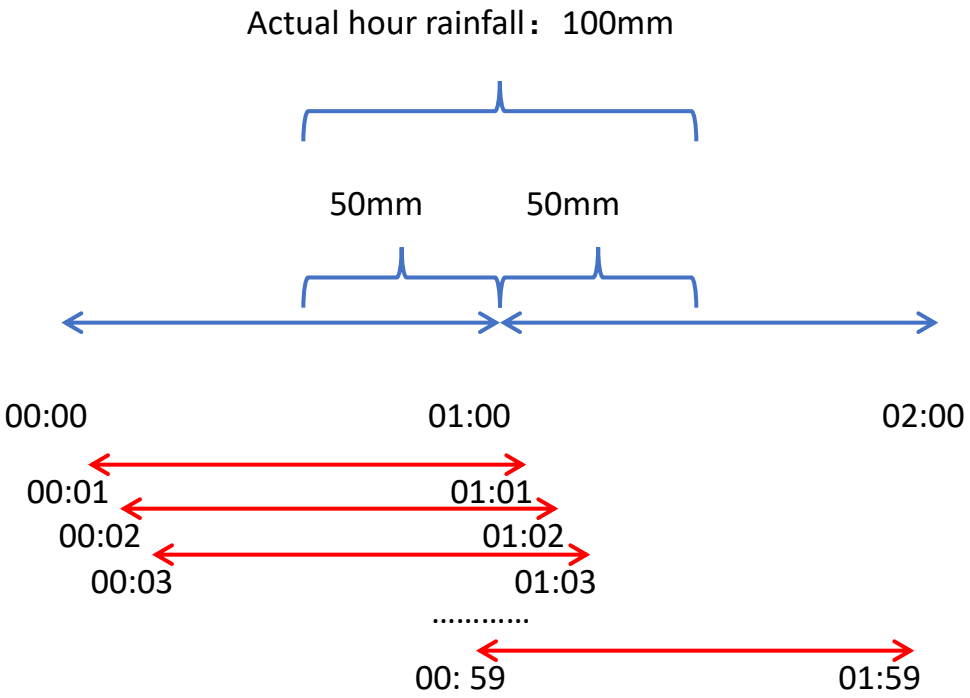
- 60 stations with 10 years of 1 minute interval data

How to make full use of our data resource?

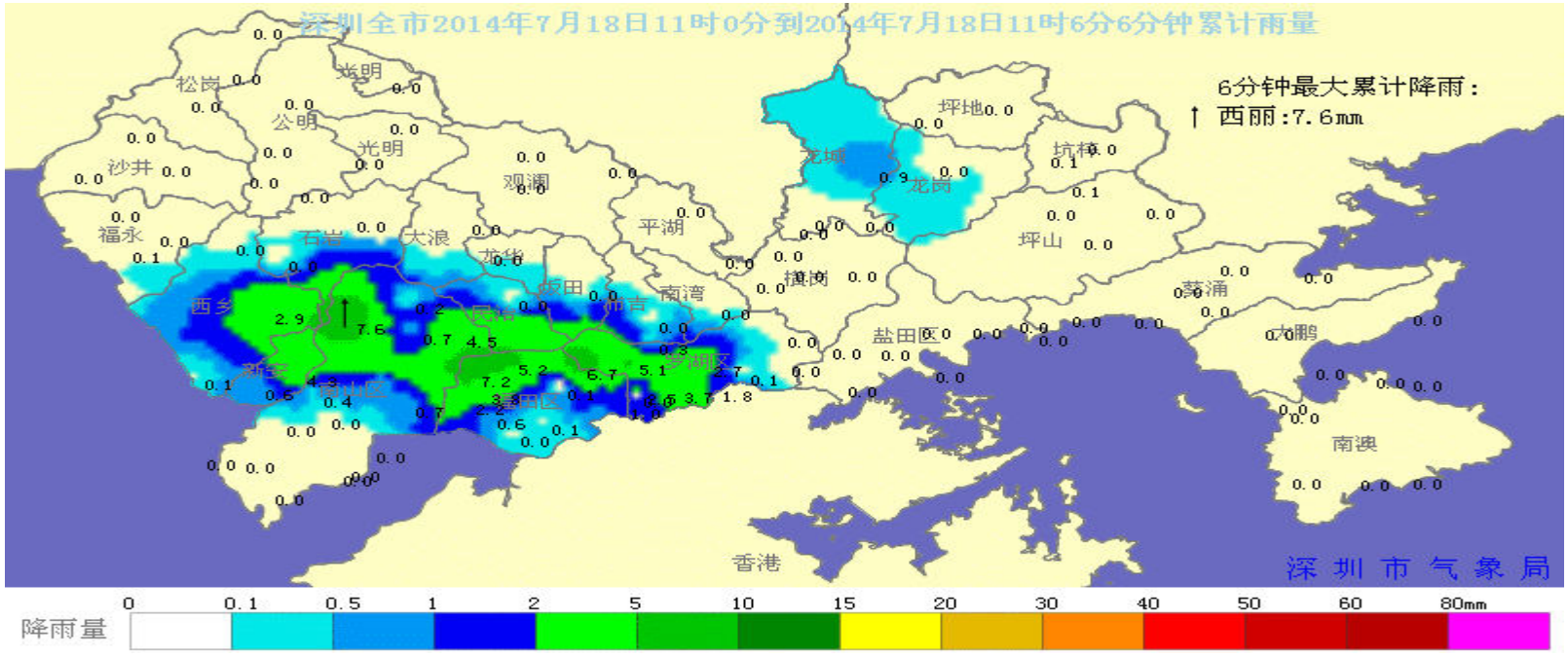
Content

1. Background Information
2. Real-Time Monitoring Products
3. Data Mining based on Machine Learning
4. Outlook

Products based on moving accumulated rainfall recorded by AWSs

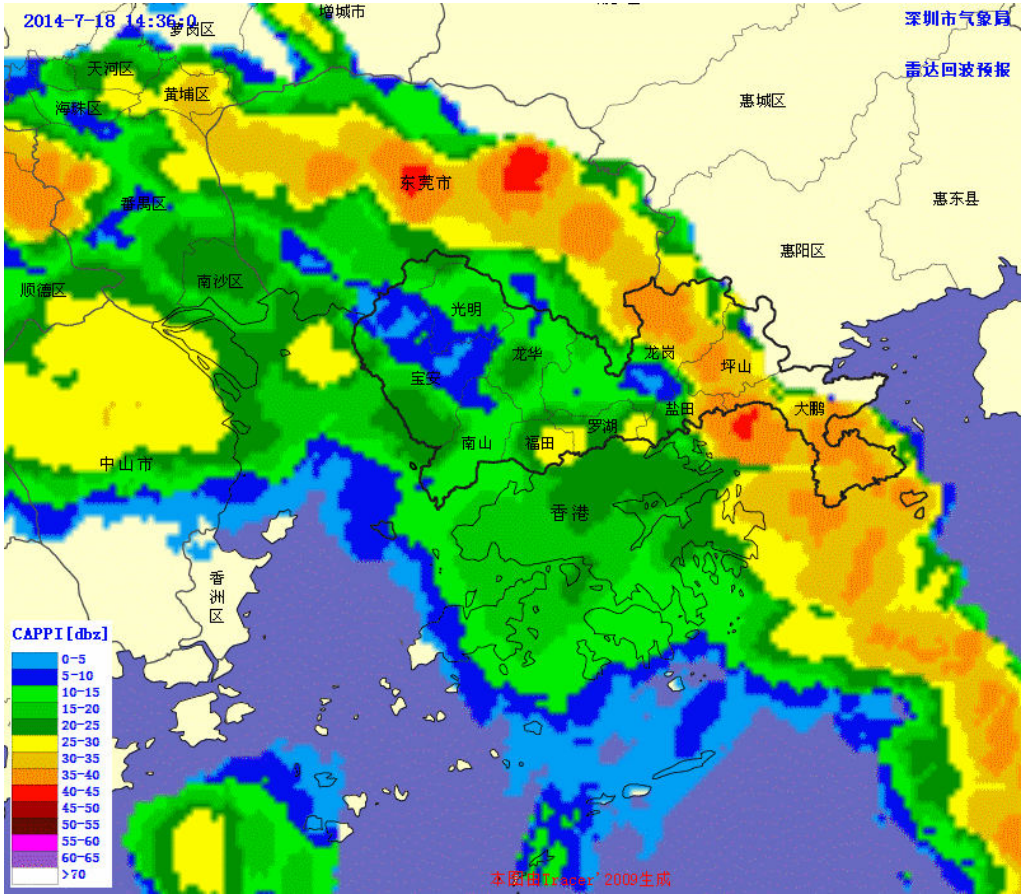


Movement of the rainfall belt



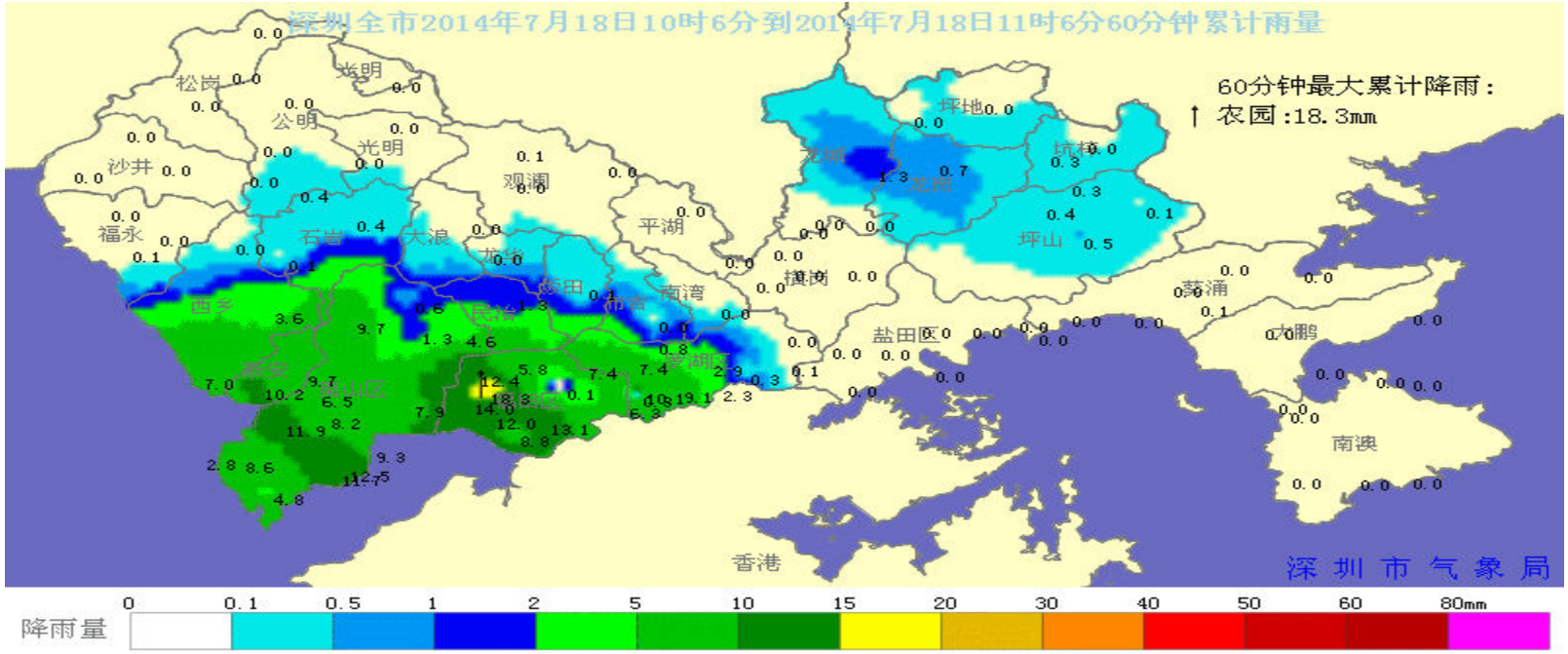
Typhoon Rammasun——6min moving accumulated rainfall

Movement of the rainfall



Typhoon Rammasun——radar image

Movement of the rainfall



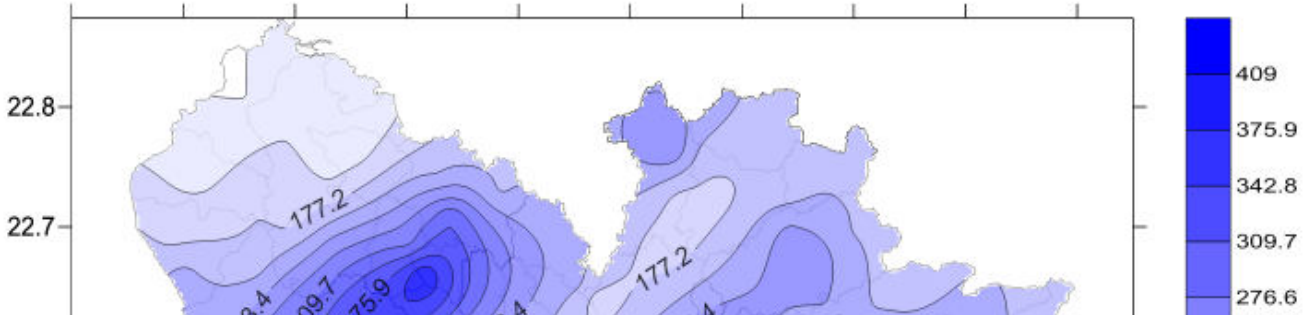
Typhoon Rammasun——60min moving accumulated rainfall

Content

1. Background Information
2. Real-Time Monitoring Products
3. Data Mining based on Machine Learning
4. Outlook

Early identification of the risk of extremely heavy rainstorm

2014.5.11 Heavy Rainstorm



排名	标识	极值(前100名)	出现区域	站号	站名	出现时间
1	日雨量极值	498	光明	G3727	塘家	2008-06-13
2	日雨量极值	472.3	宝安	G3733	黄麻布	2008-06-13
3	日雨量极值	443	龙华	G1155	龙华	2014-05-11
4	日雨量极值	432.8	南山	G3565	大学城	2014-05-11
5	日雨量极值	410	宝安	G3531	石岩水库	2008-06-13
6	日雨量极值	401.7	南山	G3766	大塘	2014-05-11
7	日雨量极值	387.9	大鹏	G3563	南澳	2016-05-21

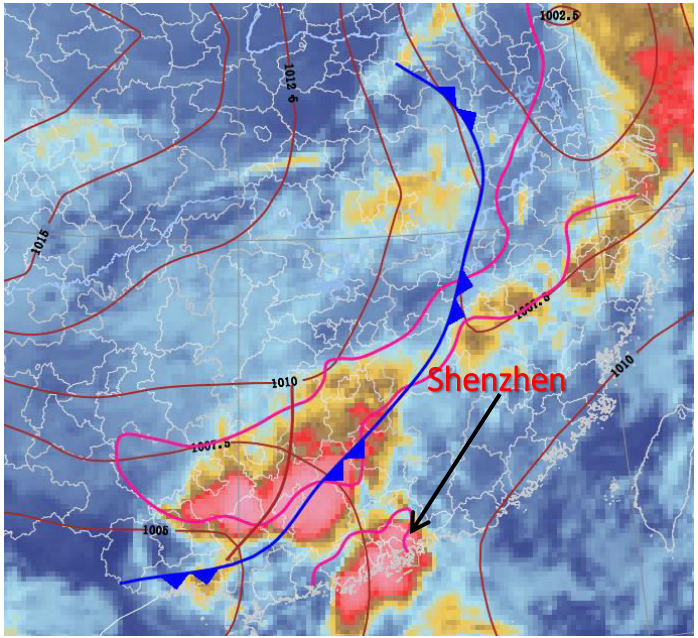
2014.5.11 Heavy Rainstorm

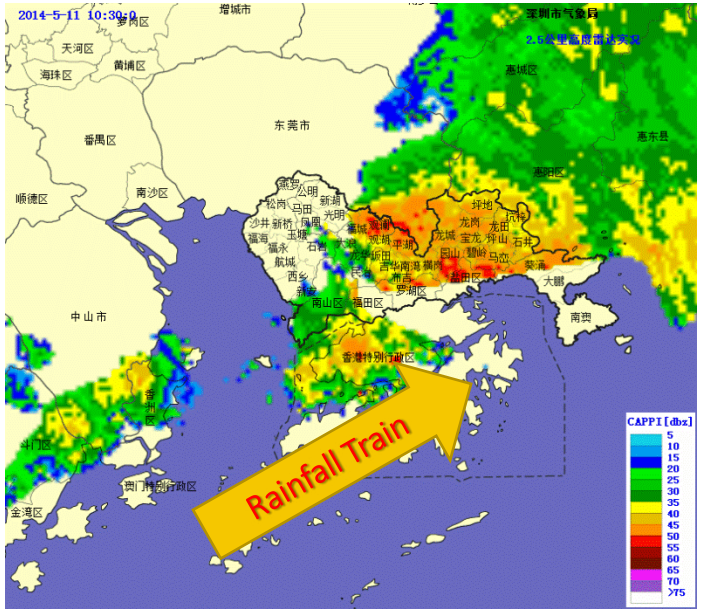






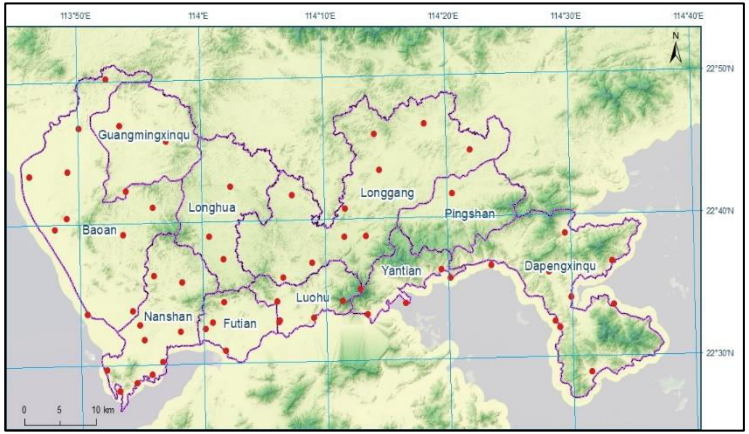

2014.5.11 Heavy Rainstorm

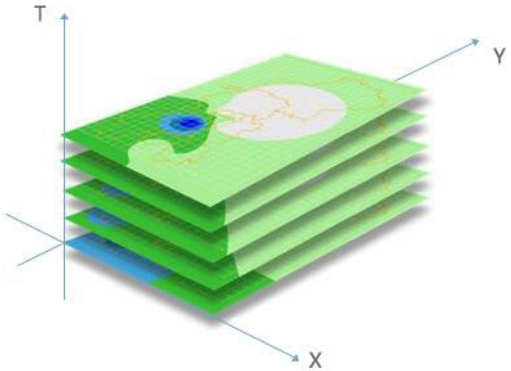




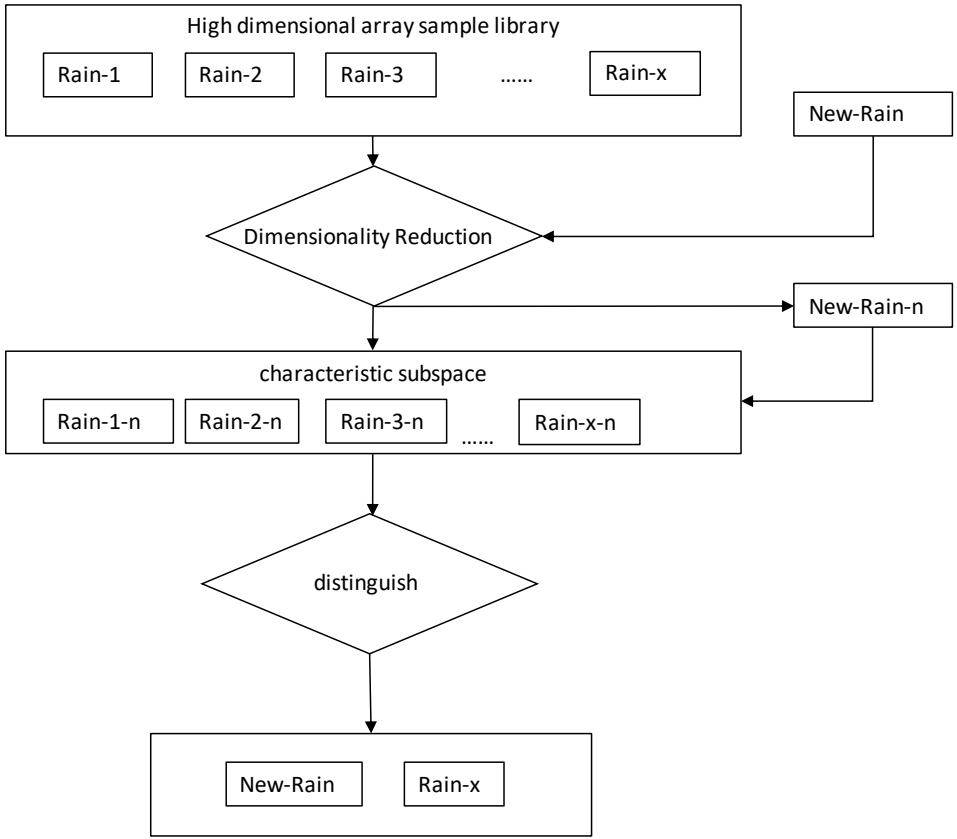
- When will the rainstorm stop?
- How much will the total precipitation be?

In-process Identification of Rainstorm based on Machine Learning

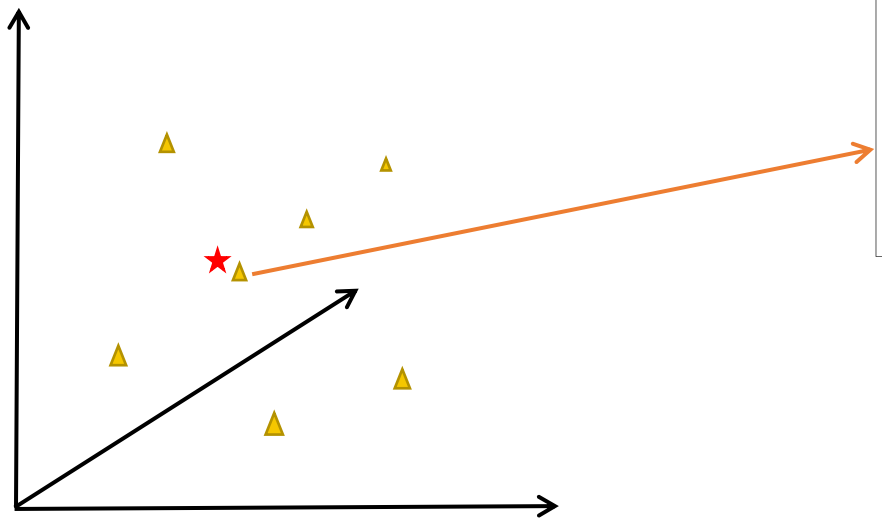




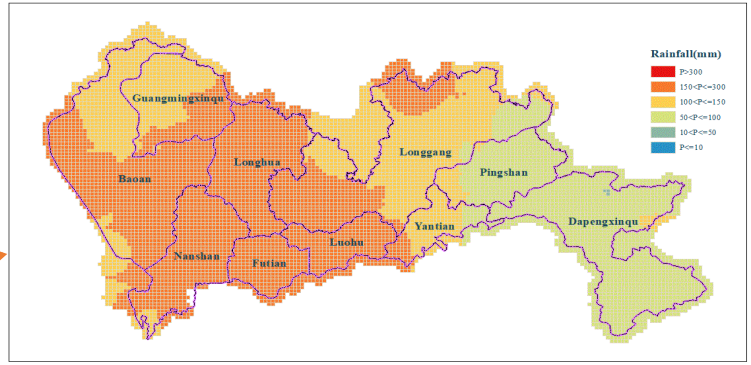
63 AWSs, 116 rainstorm events (2008-2017)



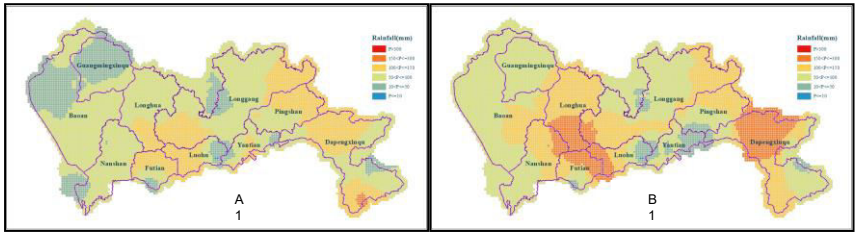
In-process Identification of Rainstorm based on Machine Learning

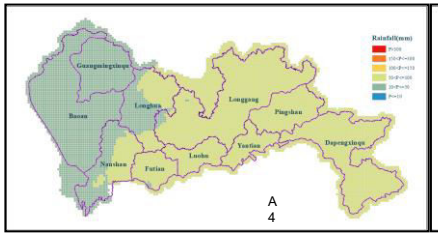


Duration, Intensity, Range

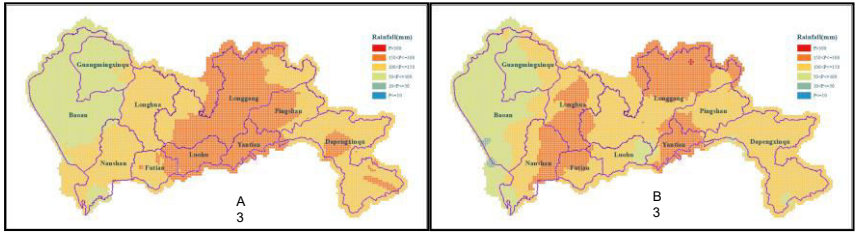


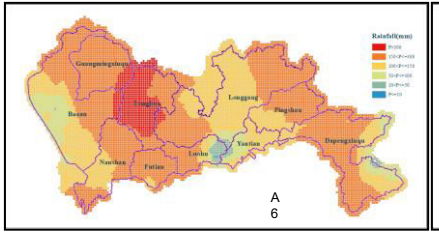
In-process Identification of Rainstorm based on Machine Learning

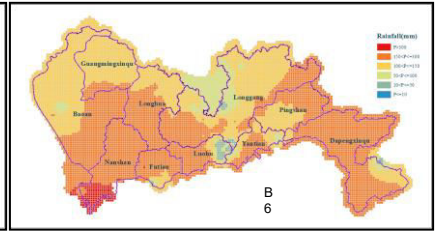




When an ongoing rainfall completes 1/3 of its whole duration, by identifying its “current” status, the historical rainstorm process closest to the ongoing process can be identified, and consequently the possible impact of the ongoing rainfall can be estimated.

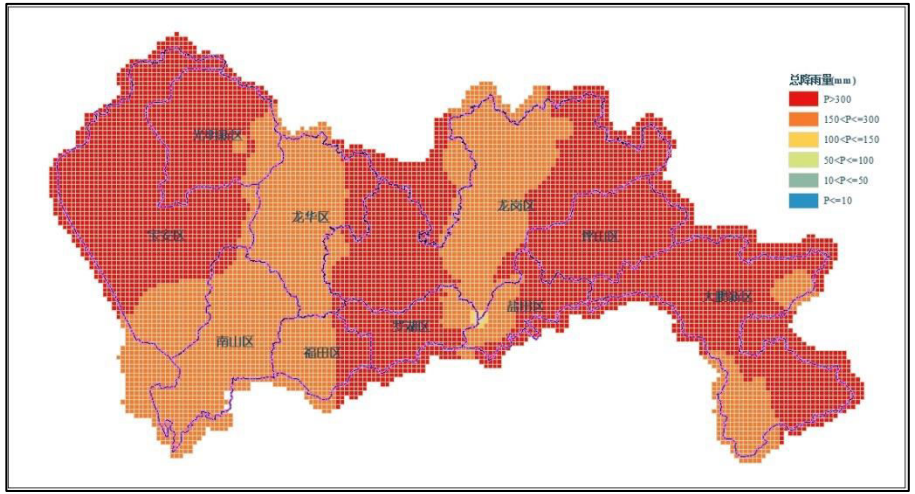




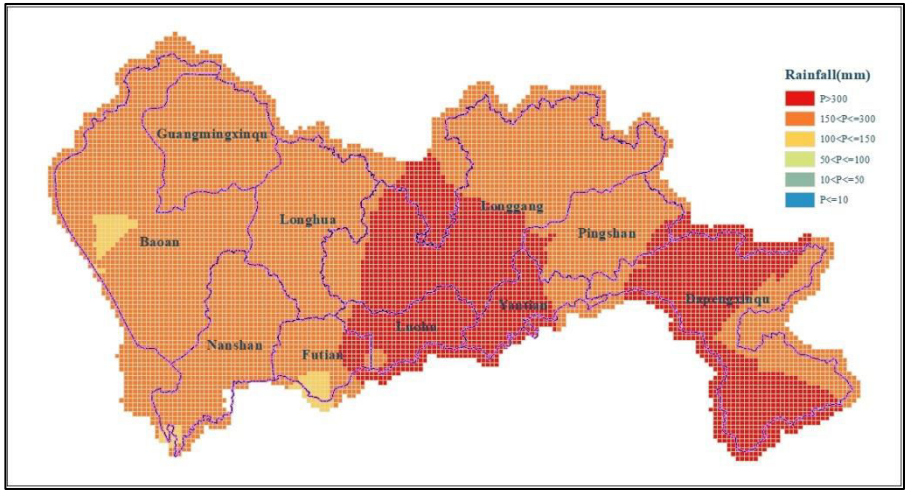


Liu Y Y, Li L*, Zhang W H, et al, 2019, Atmospheric Research

A new example

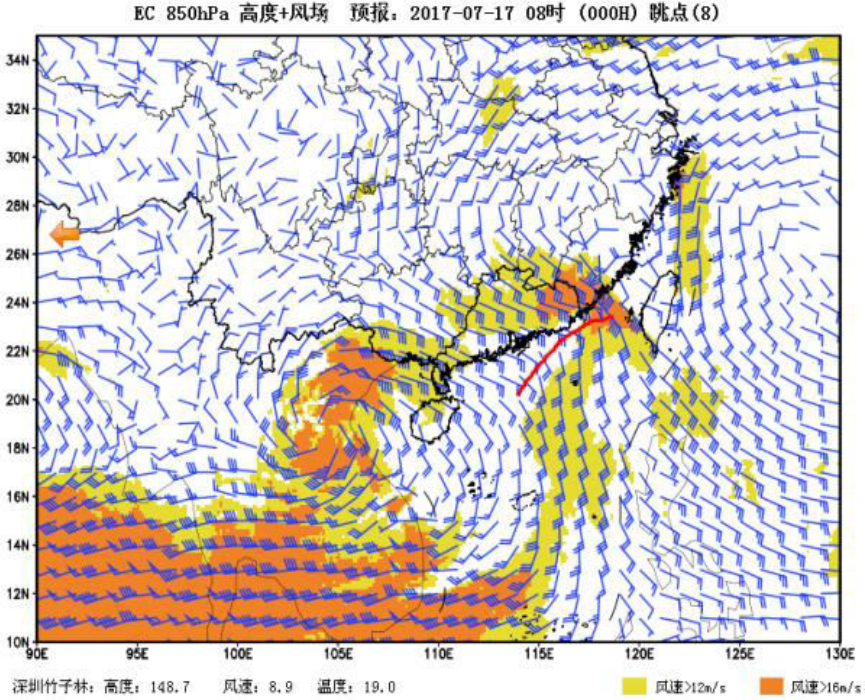


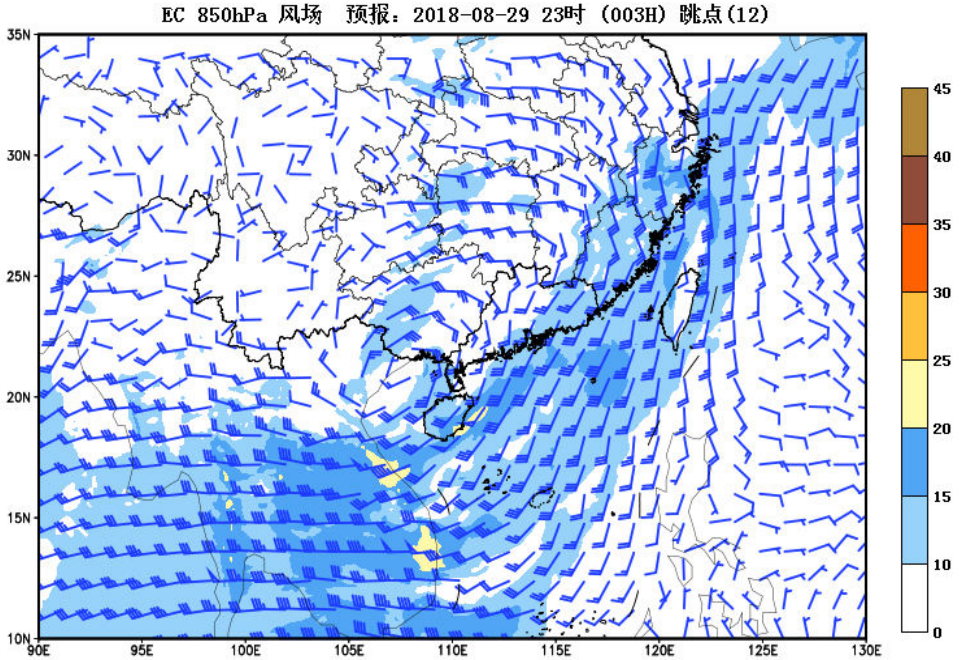
2018-8-27



2017-7-14

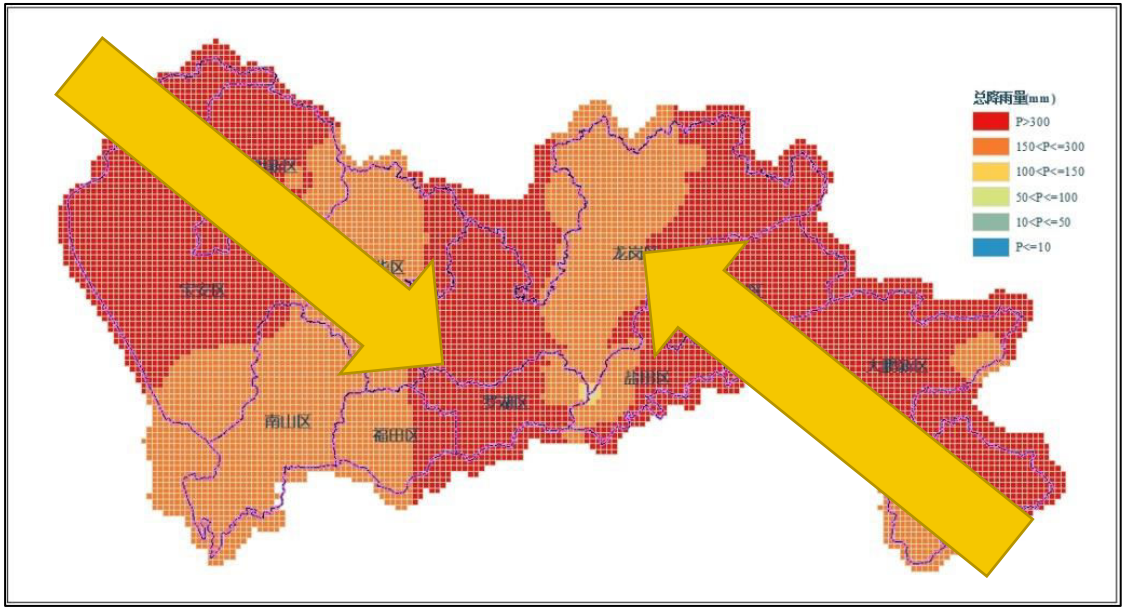
AI is more clever than me





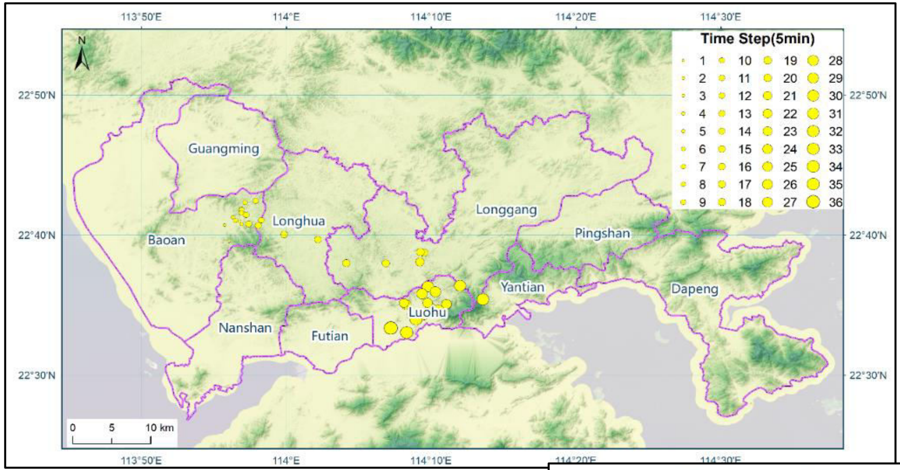
Dynamic spatial and temporal distribution characteristics of rainfall

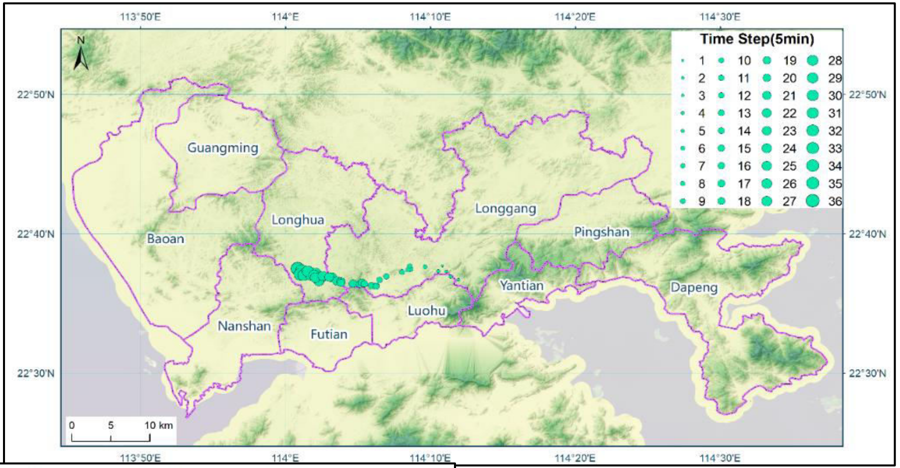
Why we do this ?

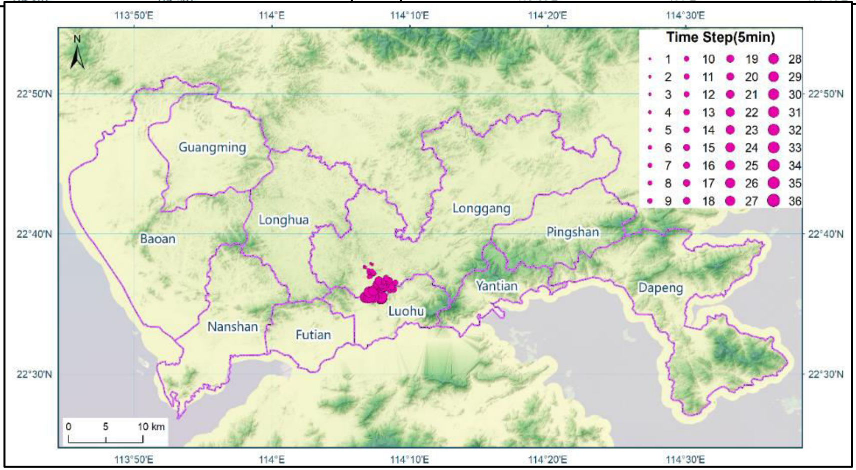


With the same rainfall distribution, but different rainfall "sequences", the hydrological response of the ground will be completely different, and disasters may also be different.

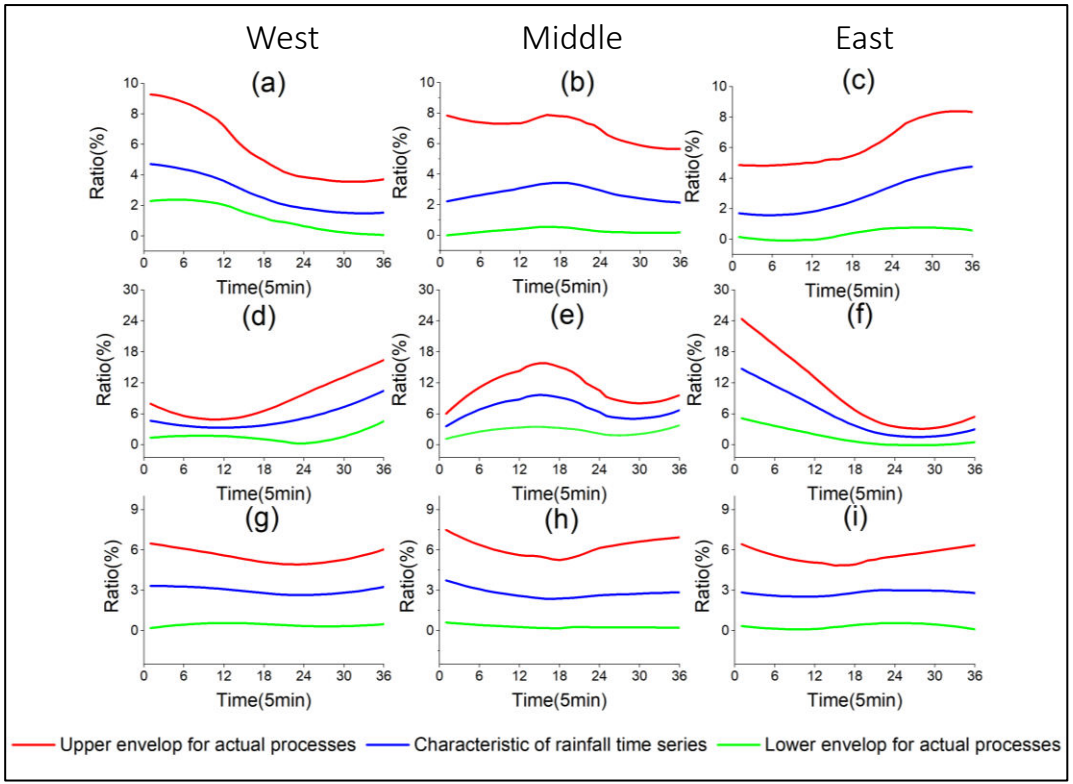
ML revealed movement trajectories of three types of rainfall





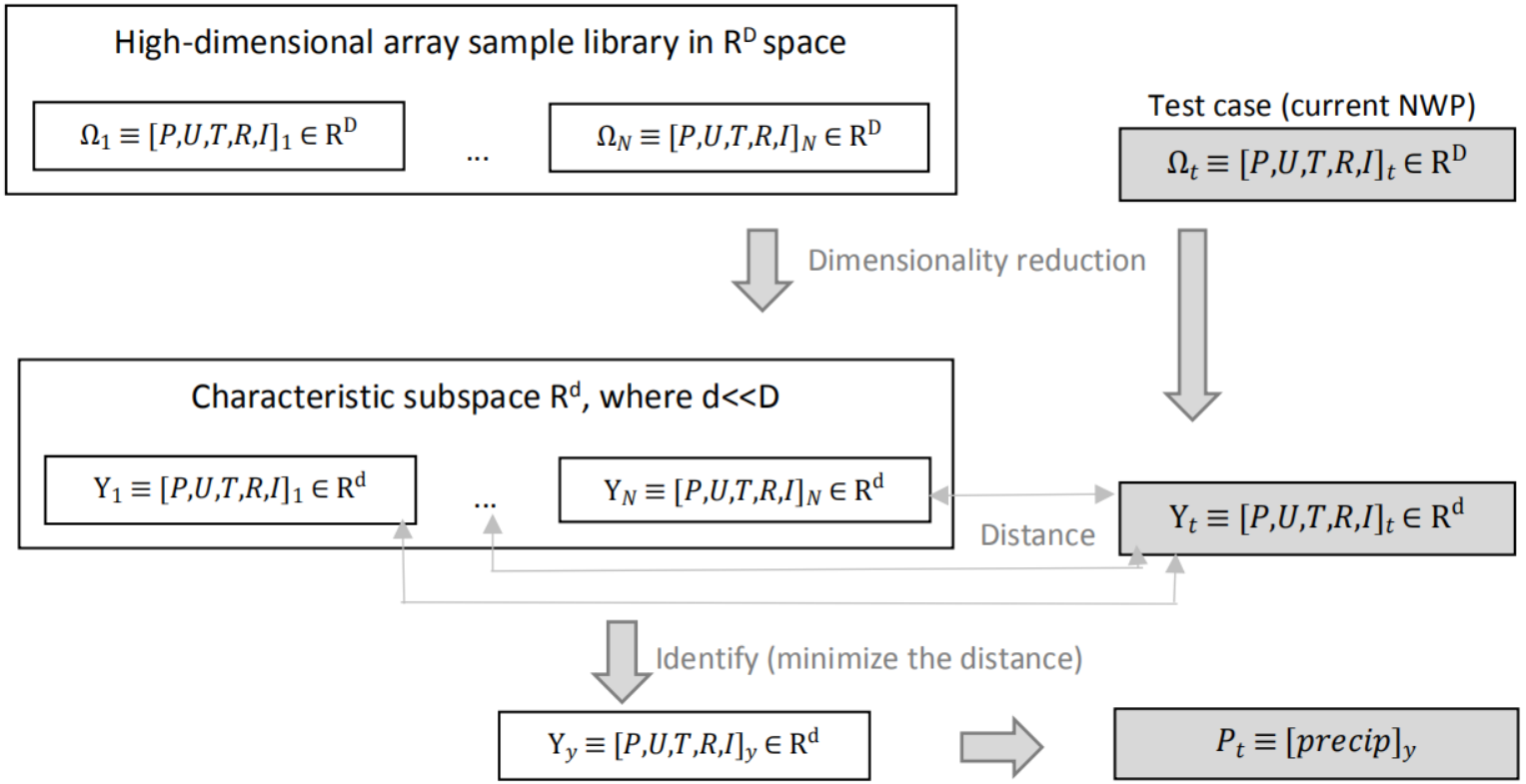


The temporal distribution in percentage of the rainfall during the whole process within different sub-regions.

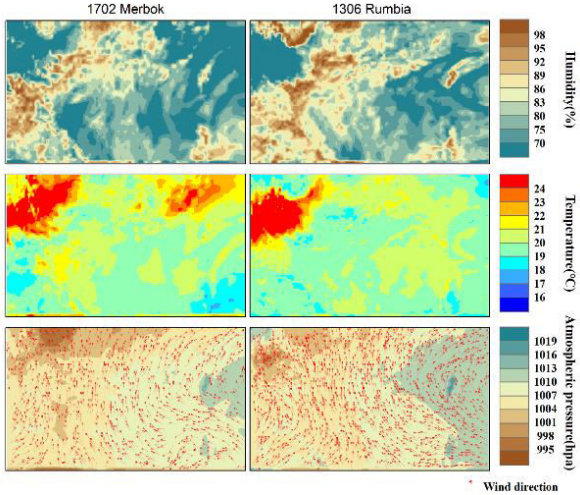


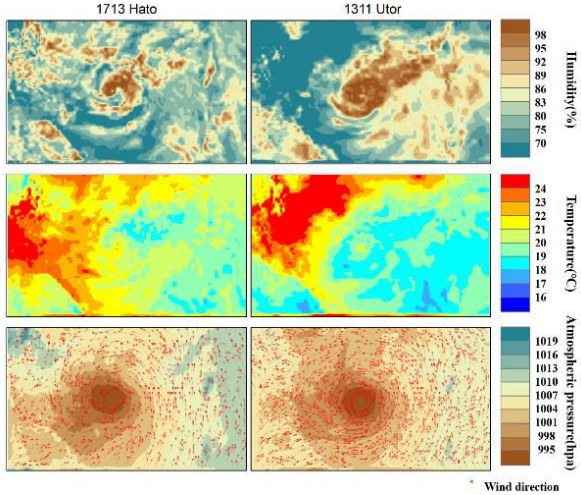
Find similar typhoons from historical numerical forecasting output data

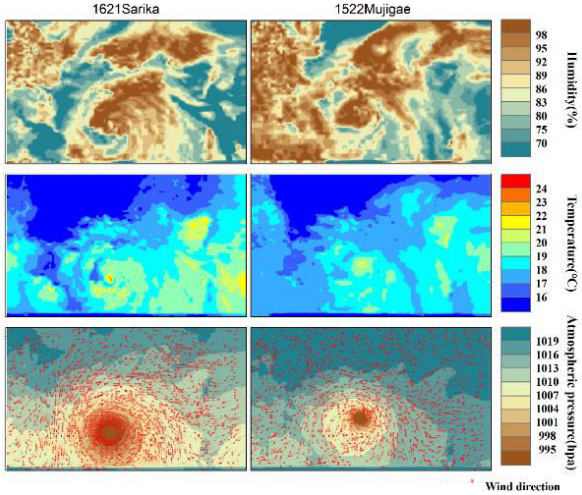
Method


High-dimensional array sample library in R^D space
$\Omega_1 \equiv [P,U,T,R,I]_1 \in R^D$
...
$\Omega_N \equiv [P,U,T,R,I]_N \in R^D$
Test case (current NWP)
$\Omega_t \equiv [P,U,T,R,I]_t \in R^D$
Dimensionality reduction
Characteristic subspace R^d, where d<<D
$Y_1 \equiv [P,U,T,R,I]_1 \in R^d$
...
$Y_N \equiv [P,U,T,R,I]_N \in R^d$
Distance
$Y_t \equiv [P,U,T,R,I]_t \in R^d$
Identify (minimize the distance)
$Y_y \equiv [P,U,T,R,I]_y \in R^d$
$P_t \equiv [precip]_y$

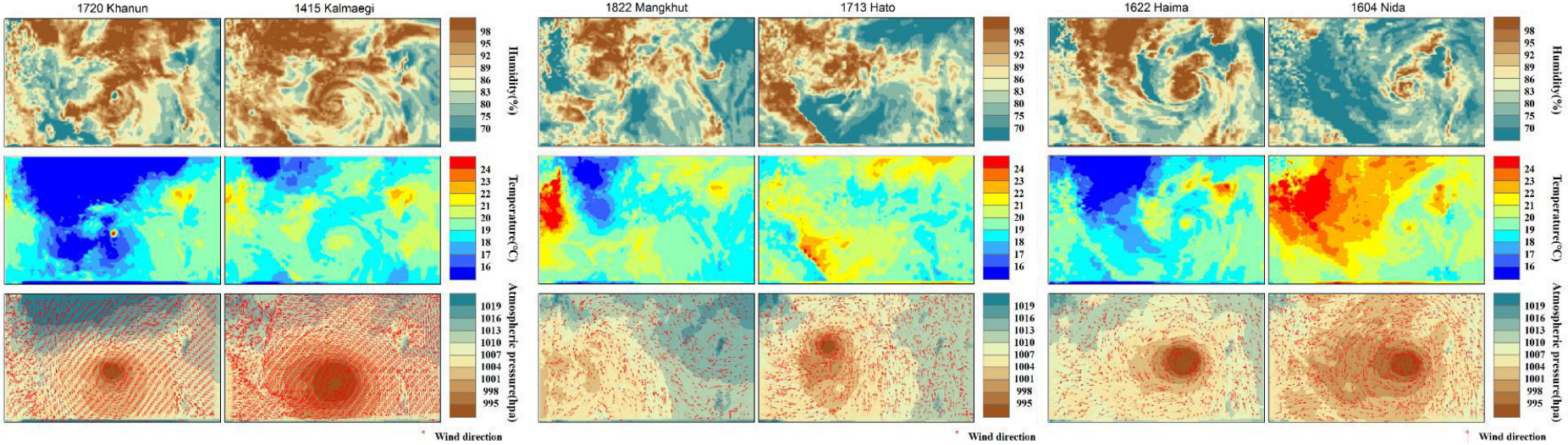
Validation


1702 Merbok
1306 Rumbia
Humidity(%)
98 95 92 89 86 83 80 75 70
Temperature(°C)
24 23 22 21 20 19 18 17 16
Atmospheric pressure(hpa)
1019 1016 1013 1010 1007 1004 1001 998 995
Wind direction


1713 Hato
1311 Utor
Humidity(%)
98 95 92 89 86 83 80 75 70
Temperature(°C)
24 23 22 21 20 19 18 17 16
Atmospheric pressure(hpa)
1019 1016 1013 1010 1007 1004 1001 998 995
Wind direction


1621Sarika
1522Mujigae
Humidity(%)
98 95 92 89 86 83 80 75 70
Temperature(°C)
24 23 22 21 20 19 18 17 16
Atmospheric pressure(hpa)
1019 1016 1013 1010 1007 1004 1001 998 995
Wind direction

Validation


1720 Khanun
1415 Kalmaegi
1822 Mangkhut
1713 Hato
1622 Haima
1604 Nida
Humidity(%)
98 95 92 89 86 83 80 75 70
Temperature(℃)
24 23 22 21 20 19 18 17 16
Atmospheric pressure(hpa)
1019 1016 1013 1010 1007 1004 1001 998 995
Wind direction

Validation

No.	Test TC cases (group A)	Most similar historical samples (group B)	Observed precipitation of samples in group A (mm/24h)	Predicted precipitation according to samples in group B (mm/24h)	Relative error by using ML method (%)	Predicted precipitation from IFS model (mm/24h)	Relative error by using IFS (%)	Predicted precipitation according to Li et al (2015a) (mm/24h)	Relative error by using Li's method (%)
1	1621 Sarika	1522 Mujigae	94.42	59.75*	36.72*	13.15	86.07	19.00	79.88
2	1622 Haima	1604 Nida	84.65	114.40*	35.14*	27.11	67.97	50.00	40.93
3	1702 Merbok	1306 Rumbia	97.79	41.60	57.46	0.93	99.05	70.00*	28.42*
4	1713 Hato	1311 Utor	50.14	56.04*	11.77*	1.75	96.51	62.00	23.65
5	1720 Khanun	1415 Kalmaegi	31.47	30.46*	3.21*	9.50	69.81	30.00	4.67
6	1822 Mangkhut	1713 Hato	30.26	50.14*	65.7*	6.71	77.83	63.00	108.20

Content

1. Background Information
2. Real-Time Monitoring Products
3. Data Mining based on Machine Learning
4. Outlook

What We Have Achieved

- An advanced monitoring network on precipitation

- Much deeper understanding on the rainfall characteristics in Shenzhen

- Many useful products in service

Challenges We Are Faced

- How to accurately predict the temporal distribution of the rainfall ? How to predict the start and end time? Which are very necessary for business arrangement.

- How to estimate the impact of Typhoon, Monsoon on Shenzhen in the future under the background of global warming?

- ……

Thanks!