Programme of Research Forum 2018
Leveraging Big Data and Artificial Intelligence in Weather Observations and Forecasting

香港天文台
HONG KONG OBSERVATORY

From Big Data to Improvement of Daily to Seasonal Weather Forecasting

Dr. Ji CHEN
The Department of Civil Engineering
The University of the Hong Kong

18-Oct-2018

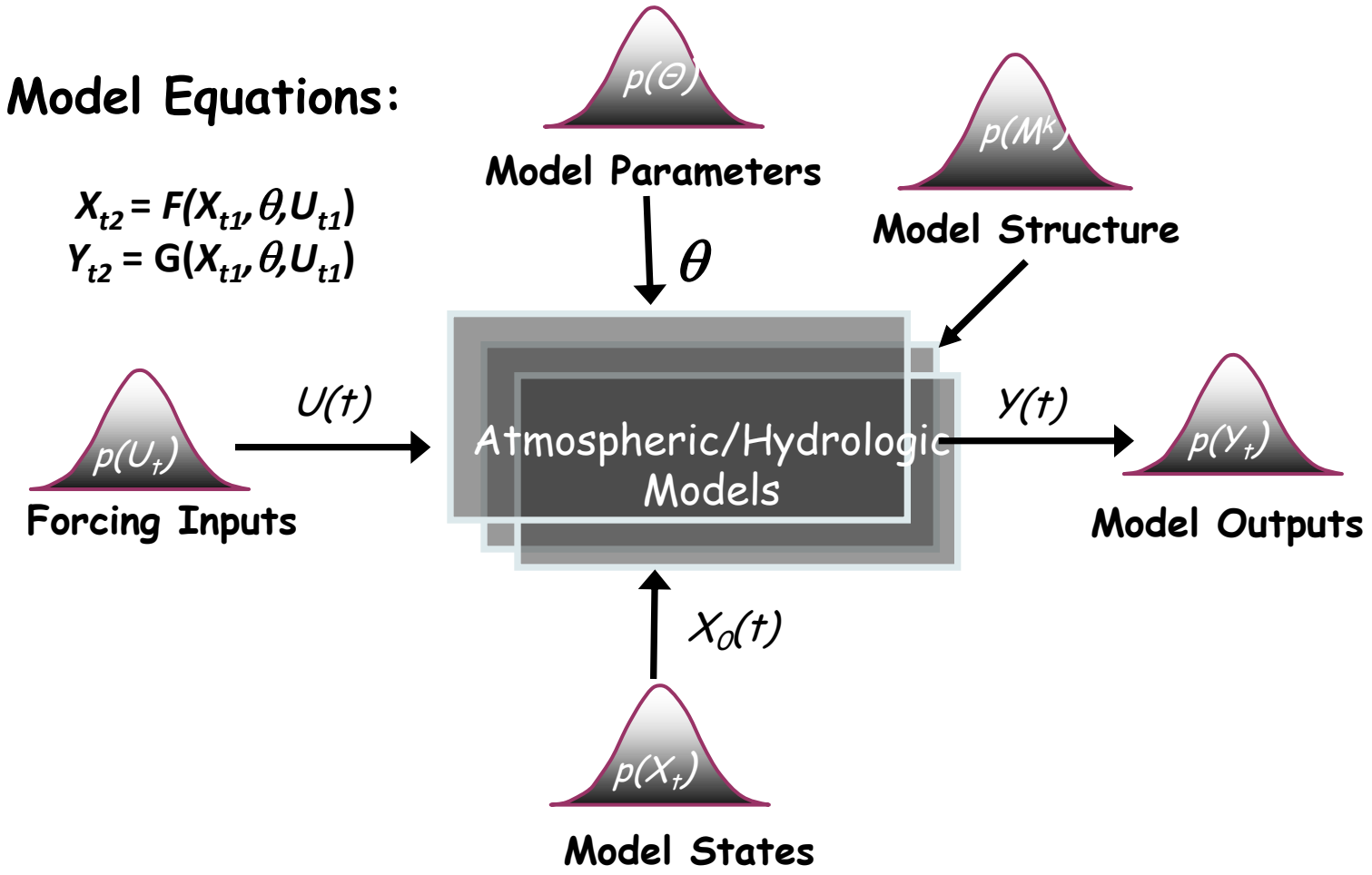
Introduction

- Weather forecasting has been a big data problem for a long time
  - Weather satellite and radar data
  - Global weather forecasting models
  - Regional weather forecasting models
  - Surface, upper air and aircraft data

- Quality control of big weather data
  - With Hugh quantities of data, quality can vary
  - Quality is particularly an issue with non-standard observations
  - Bad data must be removed

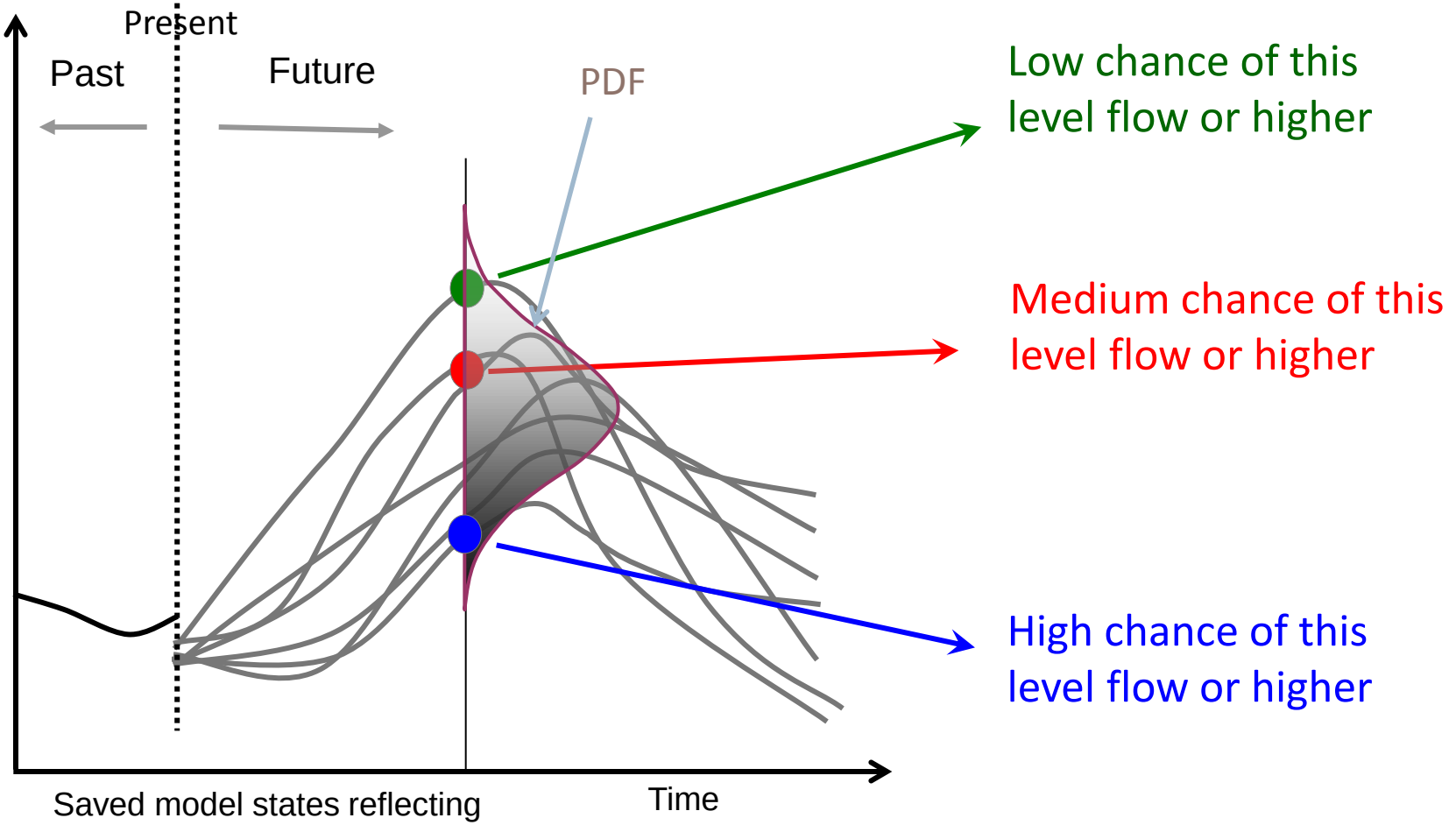
Picture from Internet

Uncertainties are prevalent in weather forecasting



Credit to Prof. DUAN, Qingyun

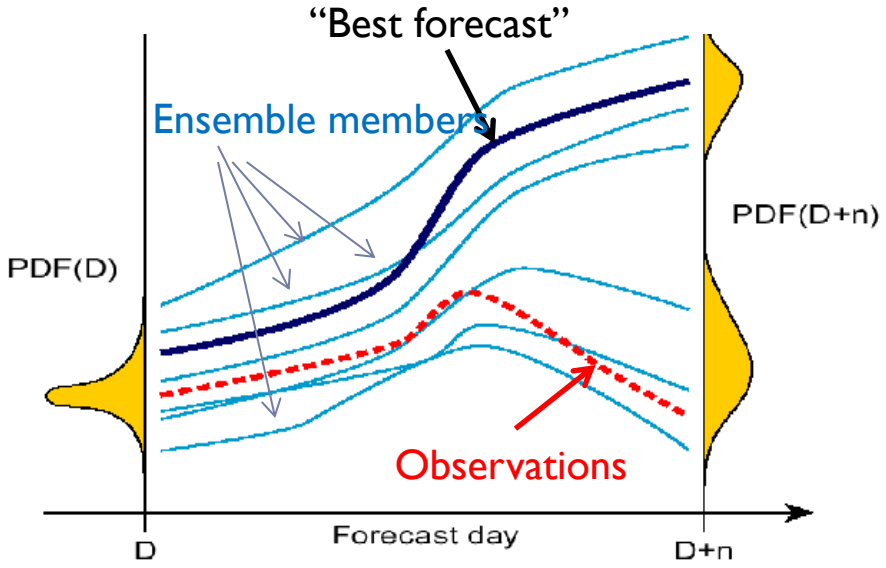
Ensemble Forecast : A set of forecasts of hydrologic events for pre-specified lead times, generated by perturbing different uncertain factors



Credit to Prof. DUAN, Qingyun

Advantages of Ensemble Forecasts

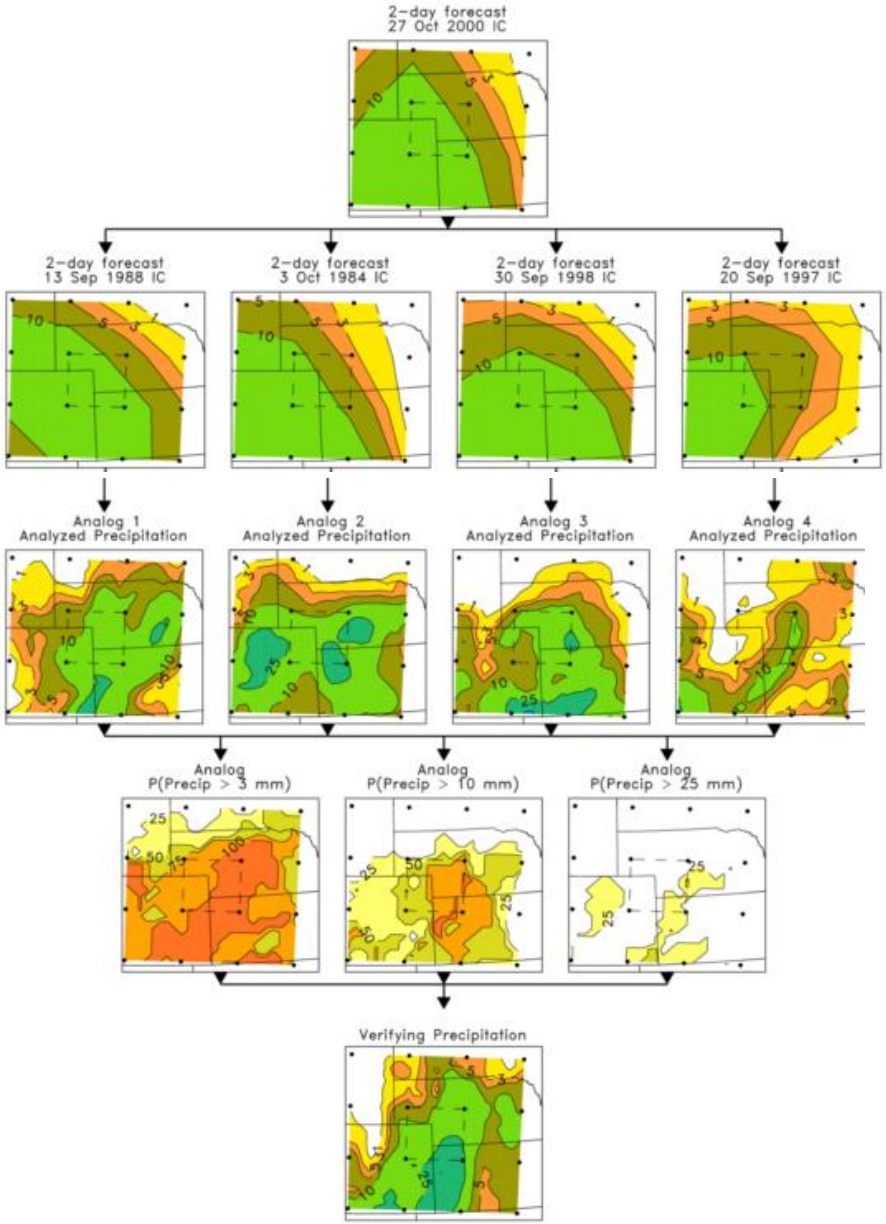
- To provide quantitative uncertainty information
  - Confidence information (for forecaster)
  - User-specified risk information (for user)
- To improve forecast accuracy
  - The average performance of ensemble predictions is better than any single prediction
- To extend forecast lead times
  - Meteorological predictions contain large uncertainties. Single valued predictions cannot express the uncertainty information. Therefore, they have shorter lead times



Credit to Prof. DUAN, Qingyun

Analog methods

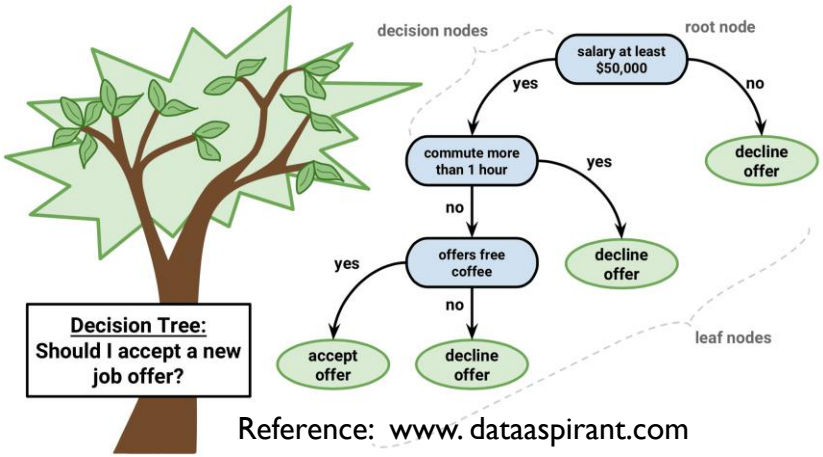
- Searches for “analogs”: the past forecast states that are similar to the current forecast states
- Calibrated probabilistic forecast is obtained from the frequency of the observed analogs
- Limitation: need sufficiently large reforecast dataset; assume stable climate. (Ref. Hamill, 2006)



Credit to Prof. DUAN, Qingyun

Multiple Regression (MR) and Classification and regression tree method (CART)

- **Multiple Regression (MR)** analysis is a linear statistical technique to find the best relationship between the dependent and explanatory variables.
- **CART** is a statistical data mining approach that divides the feature space into several sub-spaces and then fits regression model for each one.
- Advantages
  - Non-parametric method → use original data & selecting the impact variables multiple times
  - More suitable for nonlinear structure and large amount of data
  - Easy to understand and operate
- Construction of CART model
  - → recursive partitioning
  - → pruning
  - → fit regression model for each sub-space



Reference: www. dataaspirant.com

Credit to Dr. Demi SUN

Short-term forecast of Typhoon Variables

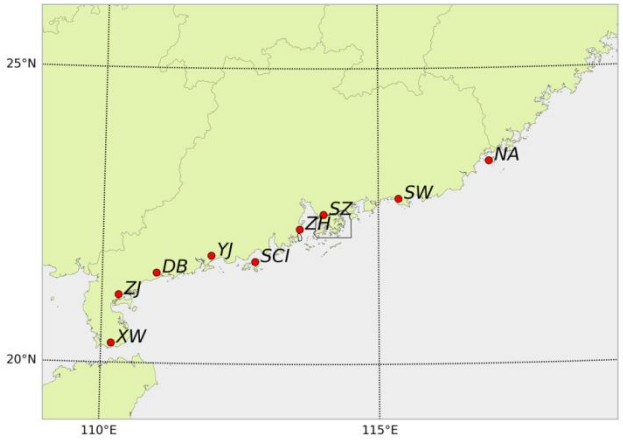
Li, Qinglan*; Lan, Hongping; Chan, Johnny C. L.; Cao, Chunyan; Li, Cheng; Wang, Xingbao. (2015) An Operational Statistical Scheme for Tropical Cyclone Induced **Rainfall Forecast**. ***Journal of Tropical Meteorology***, ISSN 1006-8775. June 2015, V21(2): 101-110

Li, Qinglan*, Xu, P., Wang, X., Lan, H., Cao, C., Li, G., Sun, L. and Zhang, L. An Operational Statistical Scheme for Tropical Cyclone Induced **Wind Gust Forecast**. ***Weather and Forecasting***, December 2016,V31(6):1817-1832

Li, Qinglan*, Zenglu Li, Yulong Peng, Xiaoxue Wang, Lei Li, Hongping Lan, Shengzhong Feng, Liqun Sun, Guangxin Li, Xiaolin Wei. Statistical Regression Scheme for **Intensity Prediction** of Tropical Cyclones in the Northwestern Pacific. ***Weather and Forecasting***, 2018,V33: 1299-1315.

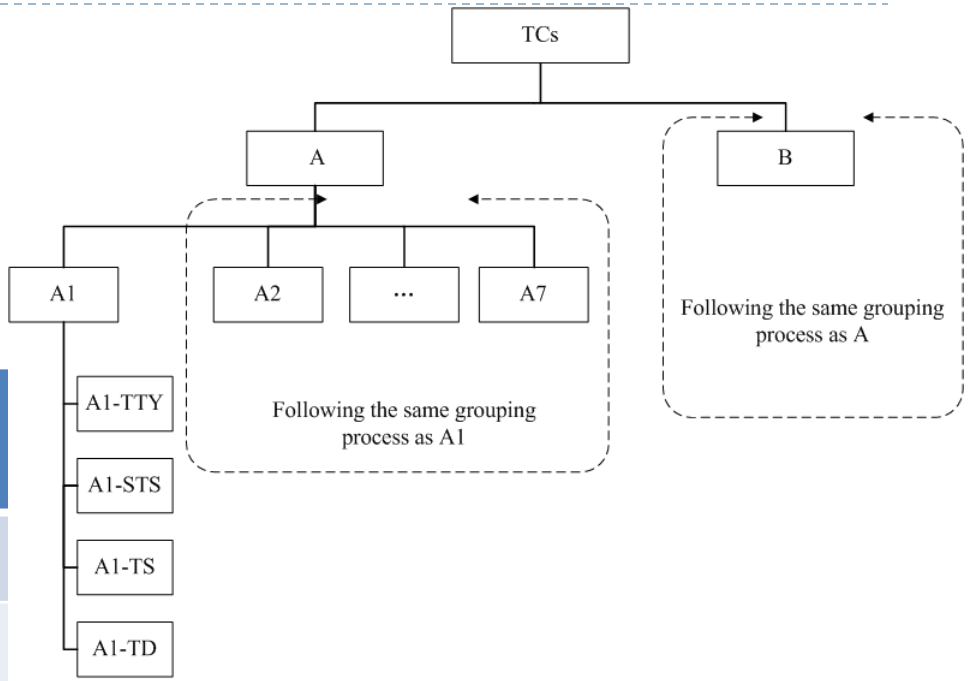
Dr. LI, Qinglan is a Professor and Ph.D. Supervisor in Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. Dr. Li obtained her PhD degree in The University of Hong Kong in 2010.

TC categorizing steps



Tropical Cyclone Intensity Scale according to CMA

Category	Abbreviation	Sustained Maximum Winds Near the Center of TCs
Super Typhoon	SuTY	>=51m/s
Severe Typhoon	STY	41.5 m/s ~ 50.9 m/s
Typhoon	TY	32.7 m/s ~ 41.4 m/s
Severe Tropical Storm	STS	24.5 m/s ~ 32.6 m/s
Tropical Storm	TS	17.2 m/s ~ 24.4 m/s
Tropical Depression	TD	10.8 m/s ~ 17.1 m/s

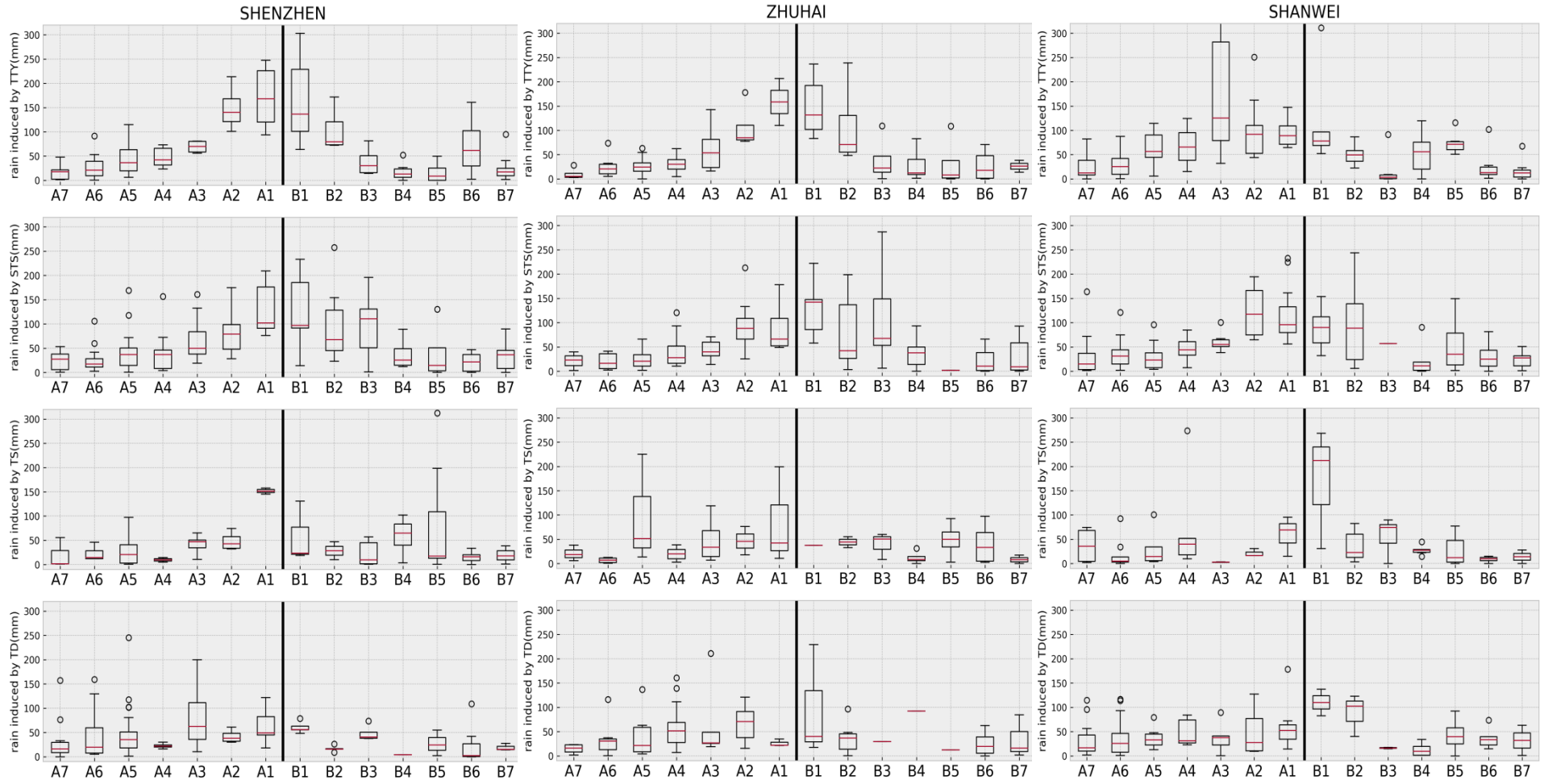


A：onshore wind

B：offshore wind

Credit to Dr. LI, Qinglan

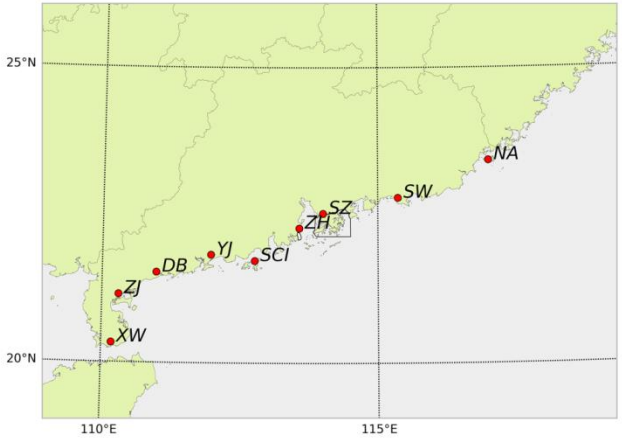
Boxplots for the station's rainfall

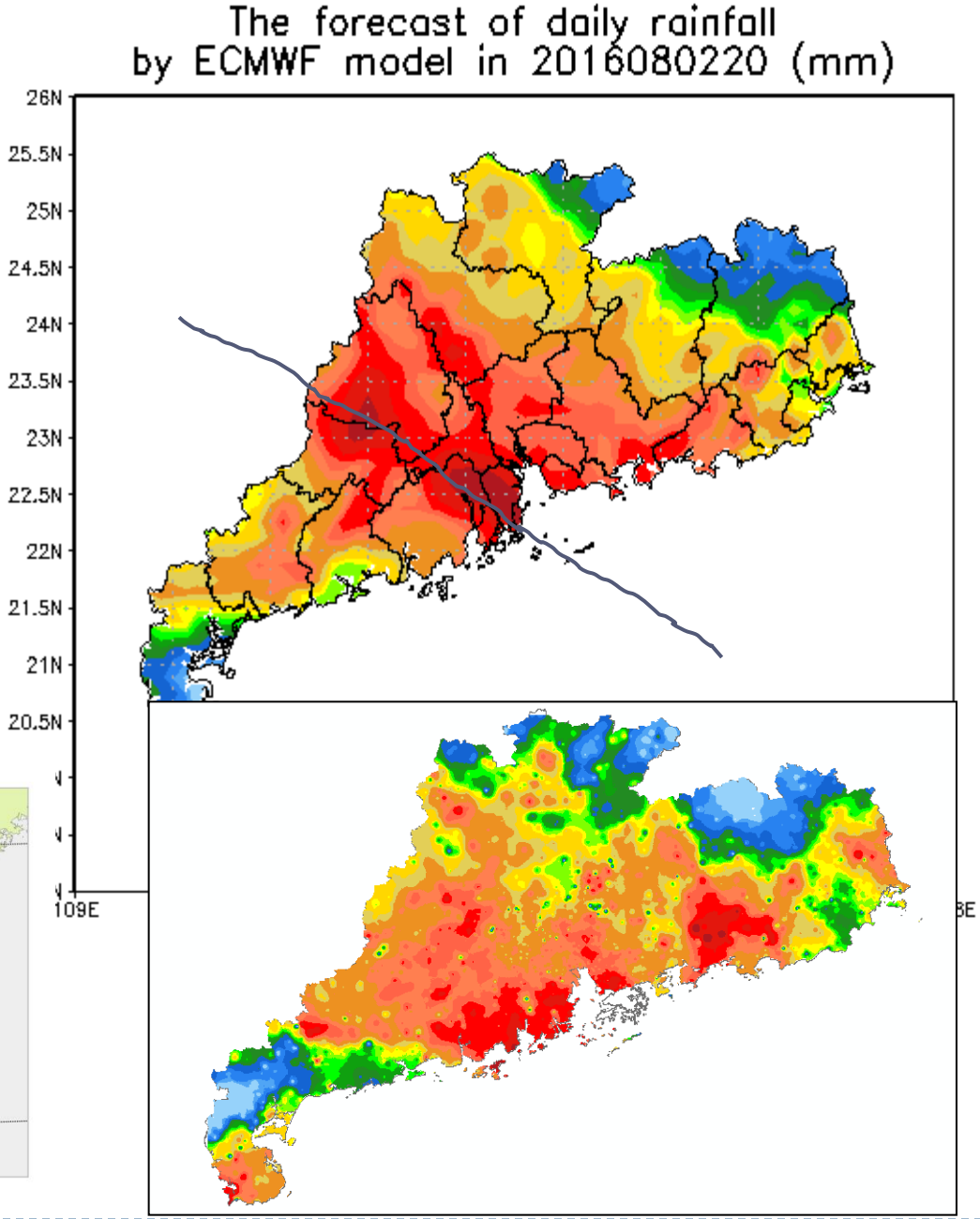


Maximum daily rainfall during landfalling period in Shenzhen (left) and Zhuhai (middle) and Shanwei (right)

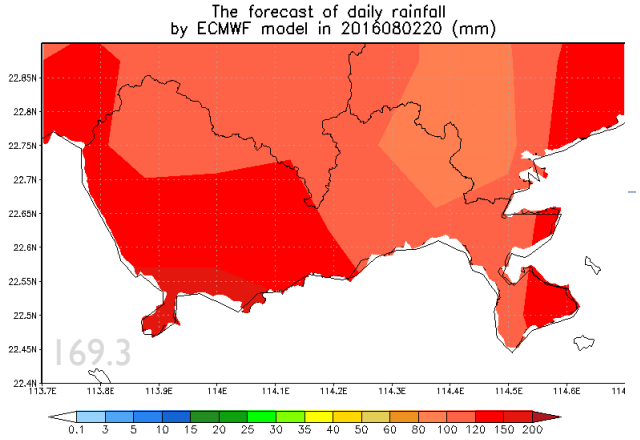
Credit to Dr. LI, Qinglan

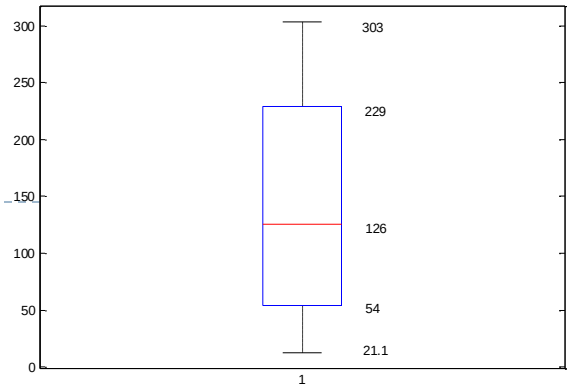
Typhoon Nida in 2016



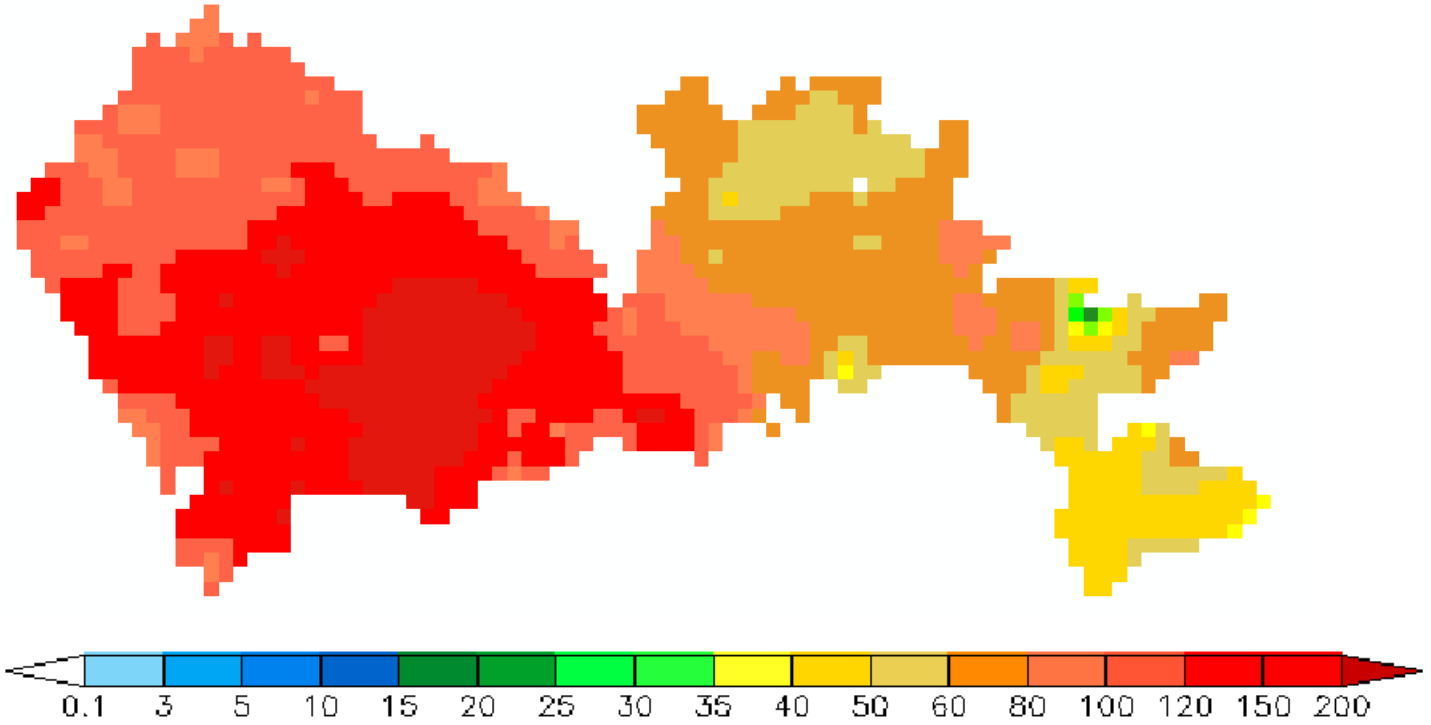


Credit to Dr. LI, Qinglan



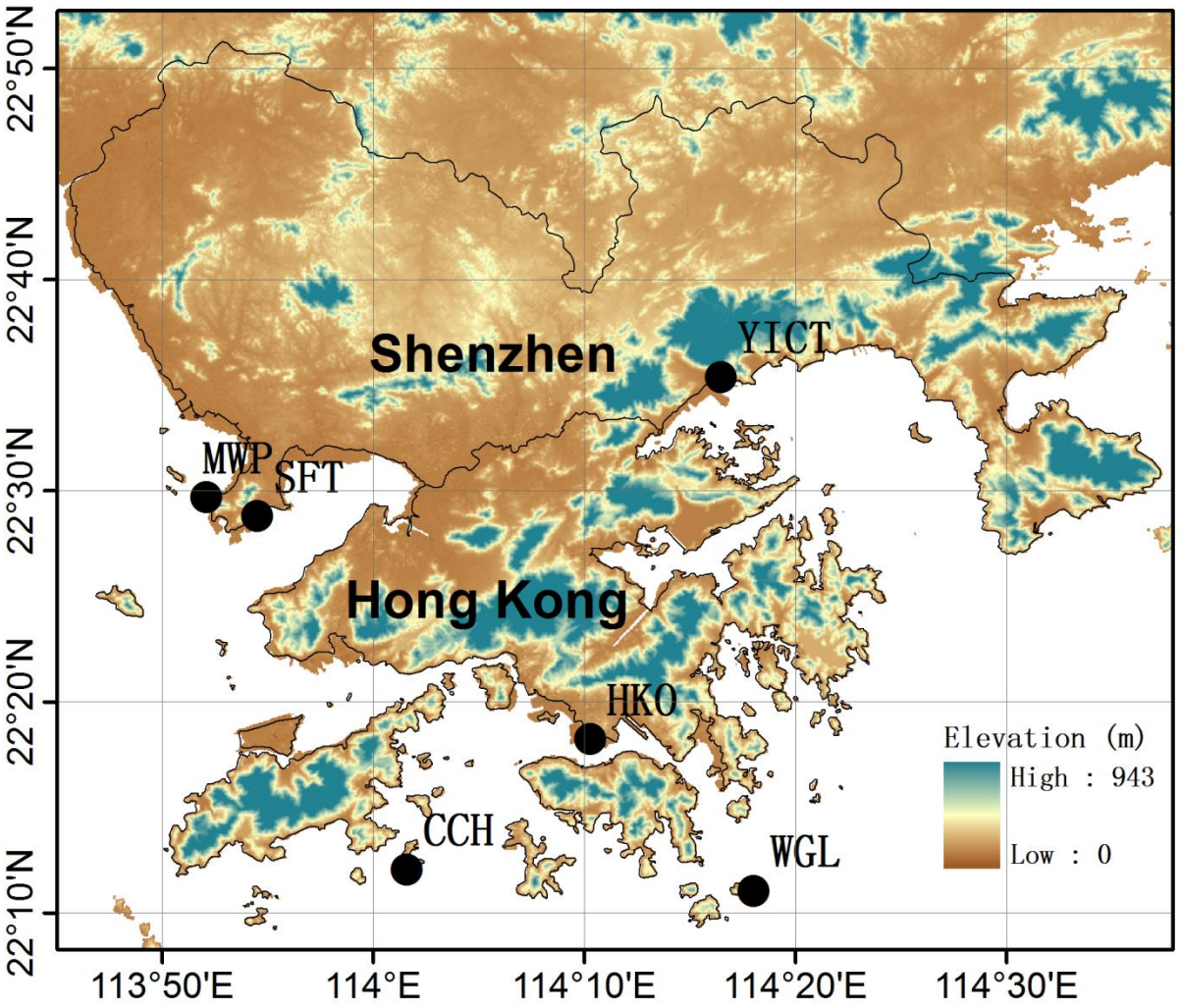


B1-TTY

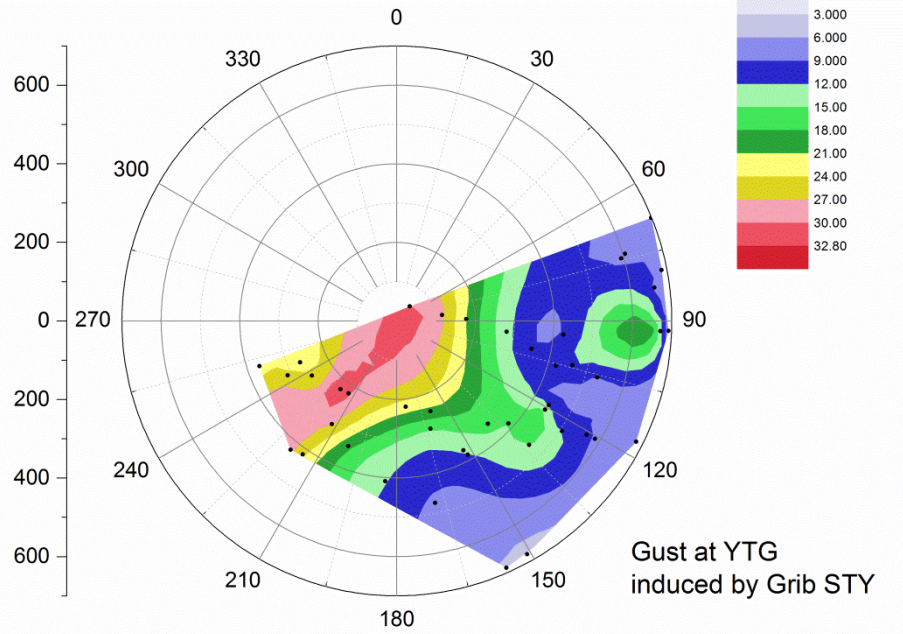


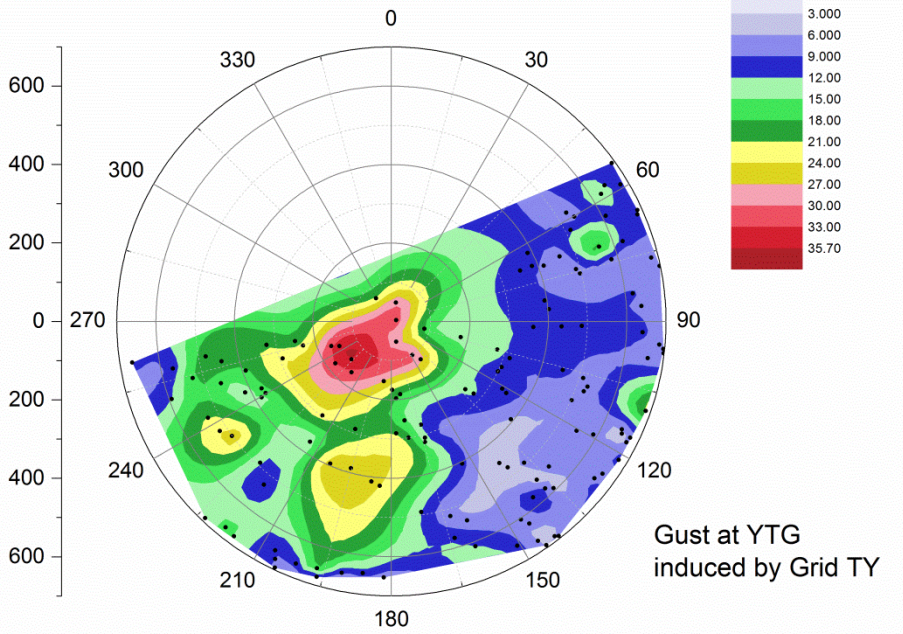
Credit to Dr. LI, Qinglan

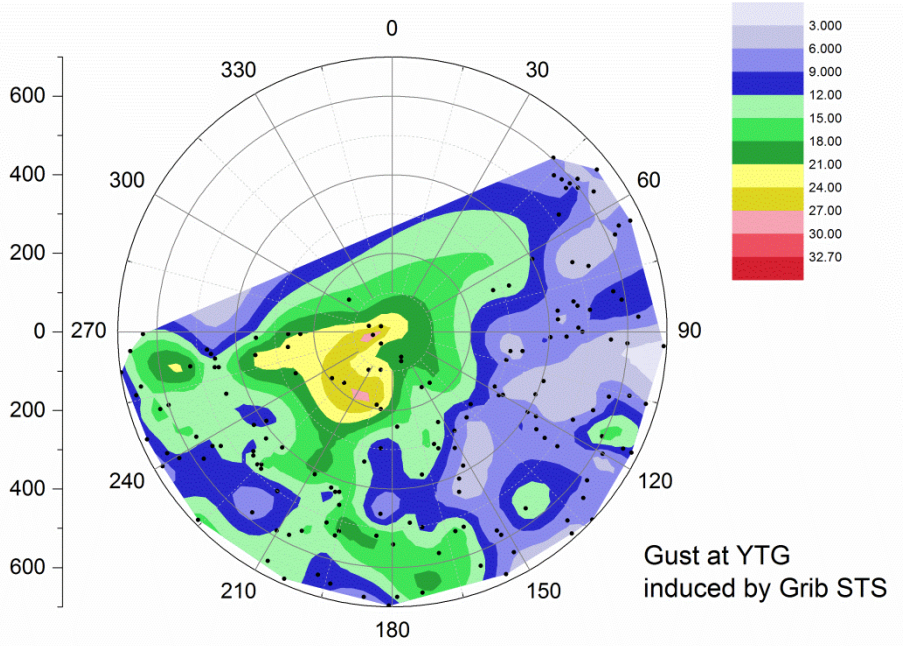
Short-term forecast of Gust

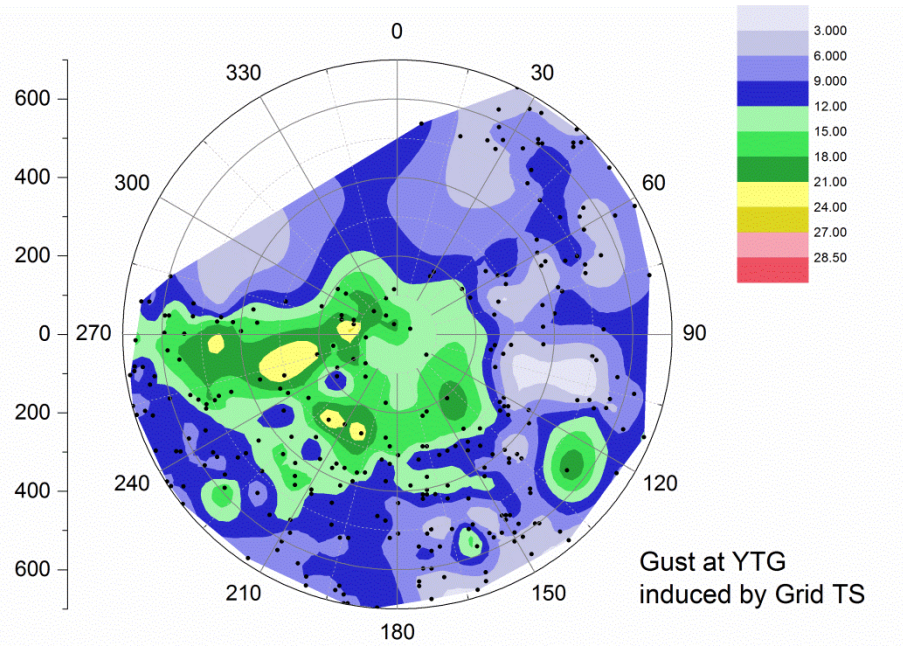


Credit to Dr. LI, Qinglan

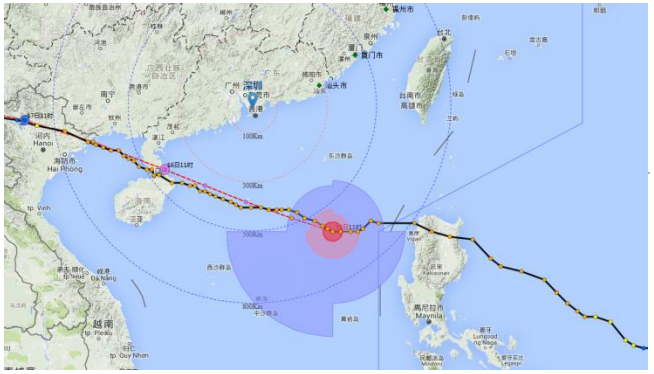

0
30
60
90
120
150
180
210
240
270
300
330
600
400
200
0
200
400
600
3.000
6.000
9.000
12.00
15.00
18.00
21.00
24.00
27.00
30.00
32.80
Gust at YTG
induced by Grib STY

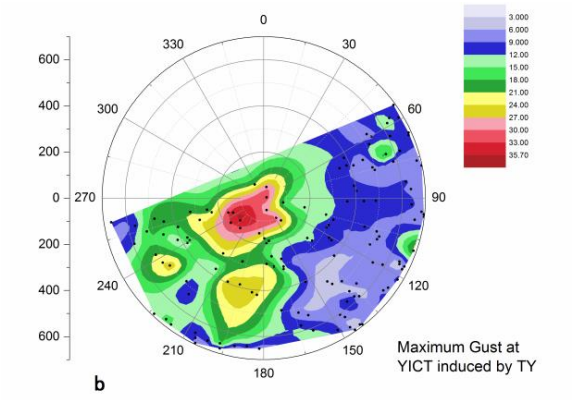

0
30
60
90
120
150
180
210
240
270
300
330
600
400
200
0
200
400
600
3.000
6.000
9.000
12.00
15.00
18.00
21.00
24.00
27.00
30.00
33.00
35.70
Gust at YTG
induced by Grid TY

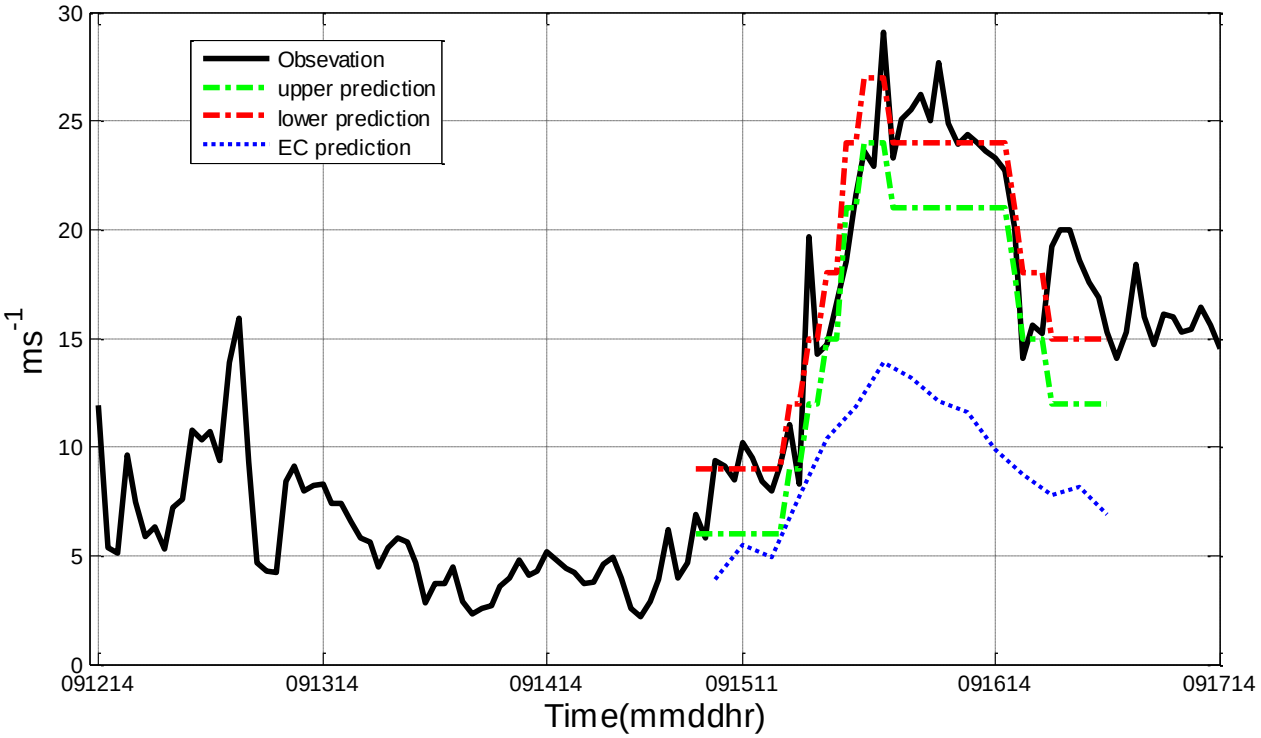

0
30
60
90
120
150
180
210
240
270
300
330
600
400
200
0
200
400
600
3.000
6.000
9.000
12.00
15.00
18.00
21.00
24.00
27.00
30.00
32.70
Gust at YTG
induced by Grib STS


0
30
60
90
120
150
180
210
240
270
300
330
600
400
200
0
200
400
600
3.000
6.000
9.000
12.00
15.00
18.00
21.00
24.00
27.00
28.50
Gust at YTG
induced by Grid TS

Typhoon Kalmaegi in 2014



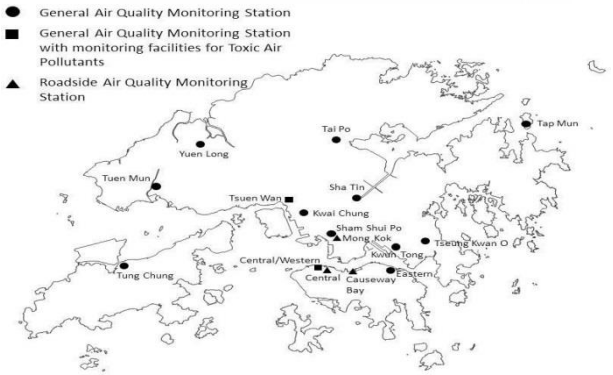




Credit to Dr. LI, Qinglan

Daily Forecast of Air Pollution Weather

- The Air Pollution Index (API) Reporting System is an important tool of risk communication. It informs the public of the local level of ambient air pollution, and the potential health risk it would impose, particularly on vulnerable groups.
- Hong Kong uses Air Quality Health Index (AQHI) since December 2013.
- The Environmental Protection Department (EPD) operates 13 general stations and 3 roadside stations for measuring concentrations of major air pollutants.
- The AQHI of the current hour is calculated from the sum of the percentage added health risk (%AR) of daily hospital admissions attributable to the 3-hour moving average concentrations of four criteria air pollutants.

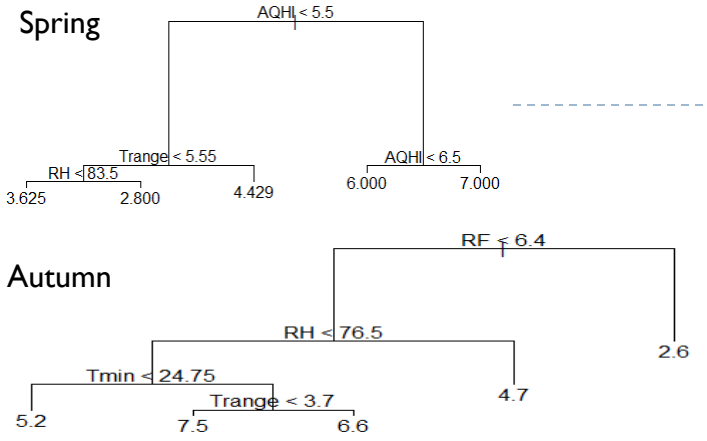


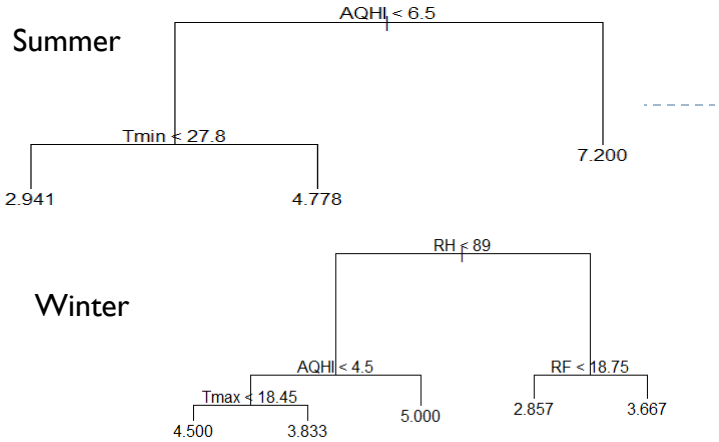
$$\%\mathrm{AR} = \%\mathrm{AR}(\mathrm{NO_2}) + \%\mathrm{AR}(\mathrm{SO_2}) + \%\mathrm{AR}(\mathrm{O_3}) + \max\big(\%AR(RSP), \%AR(FSP)\big)$$

$$\text{where } \%\mathrm{AR}(\mathrm{x}) = \big[\exp\big(\beta(\mathrm{x})\big) \times \mathrm{C}(\mathrm{x}) - 1\big] \times 100\%$$

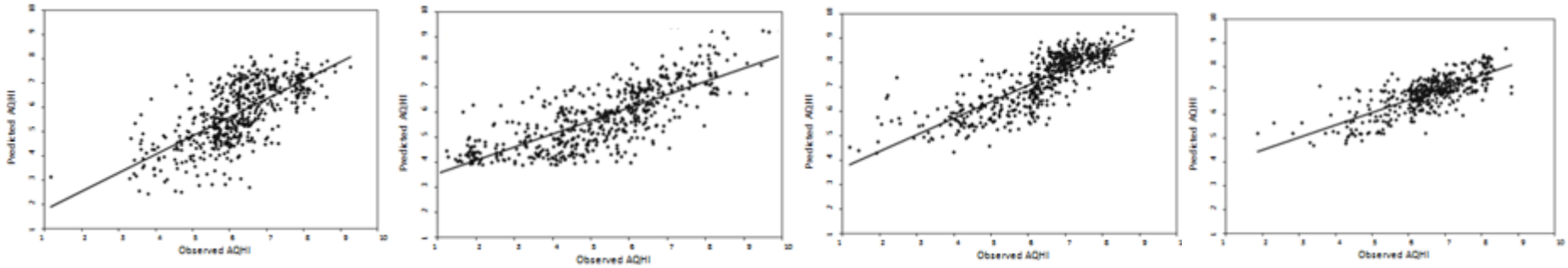
Credit to Dr. Demi SUN

➢ Regression trees for different seasons





➢ Observed and CART models hindcast AQHI in training period

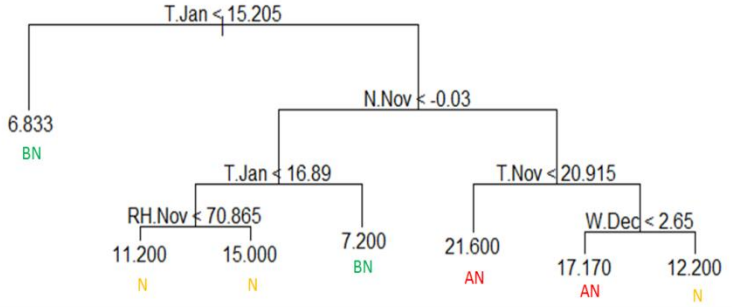


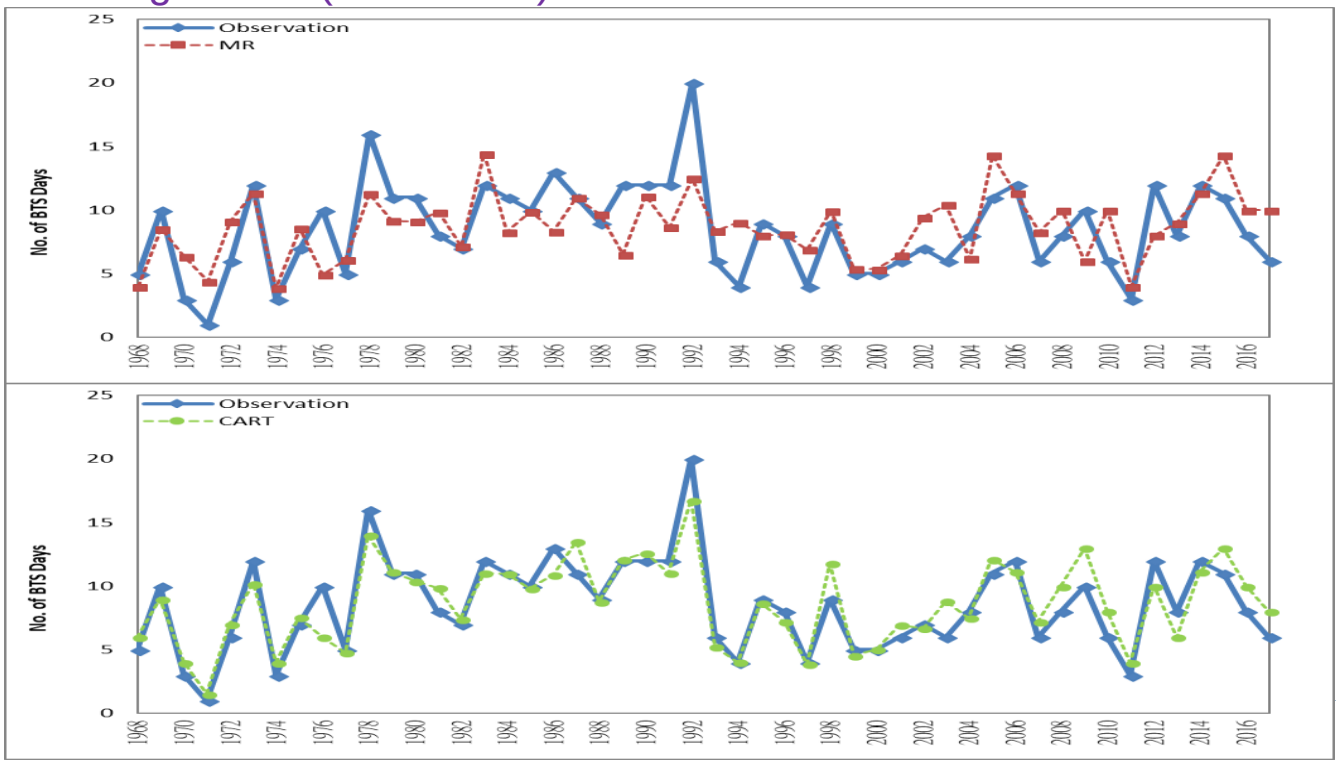
	Year	Spring	Summer	Autumn	Winter
Training Period	2014-2016	0.69	0.77	0.82	0.76
Validation Period	2017	0.71	0.66	0.55	0.46

Credit to Dr. Demi SUN

Seasonal Forecast of High-humidity Weather

- **Weather features :**
  - High humidity not caused by rainfall in spring in south China
  - Water vapor in the air condense to droplets on the surface of objectives (e.g. wall, floor, window…)
- **Definition:** $RH \geq 90\%\ and\ R = 0\ mm$
- **Peak season:** *February to April*
- Data are divided into Training Period (1968 – 2007) and Testing Period (2008-2017).





		Model	
		MR	CART
Training	r	0.65	0.95
	MSE	10.01	6.29
Testing	r	0.32	0.90
	MSE	11.00	8.72

Credit to Dr. Demi SUN

Seasonal Forecast of the Number of Tropical Cyclones

- Number of TCs affecting Hong Kong in a year:

  Training Period (1961 – 2000) and Testing Period (2001-2015).

- Average number of TCs affecting Hong Kong in a month (1961-2015)

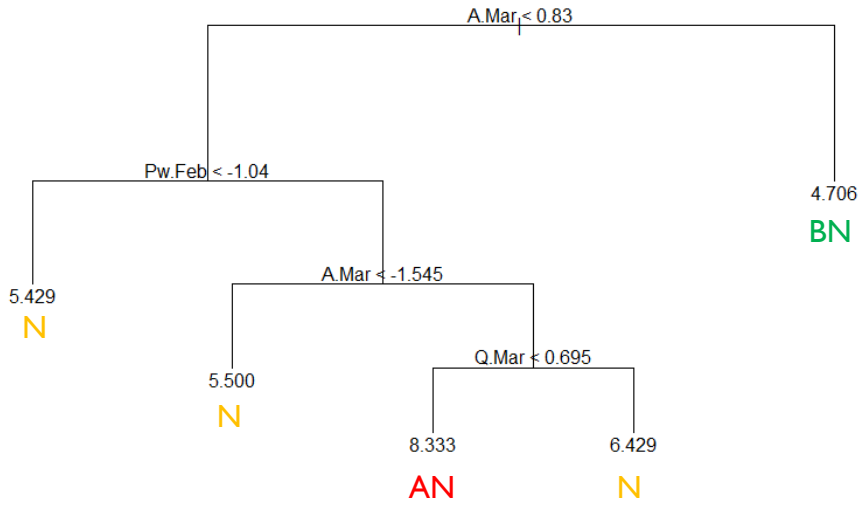
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Sum
0	0	0	0.1	0.2	**0.7**	**1.5**	**1.3**	**1.5**	**0.9**	0.1	0	6

- Category of number of TCs affecting Hong Kong in a year
  - "normal" status → the 1961-2015 long-term mean plus/minus half a standard deviation

Category	No. of tropical cyclones
Below normal (BN)	<=4
Near Normal (N)	5-7
Above normal (AN)	>=8

Credit to Dr. Demi SUN

- Pruned tree



Validation Period

Predicted / Observed	BN	N	AN
BN	**6/7**	1/7	0/7
N	0/7	**5/7**	2/7
AN	0/1	0/1	**1/1**

- Number of correct category forecasts by different methods

	Year	MR	CART
Training Period	1961-2000 (40)	27 (67.5%)	34 (85.0%)
Validation Period	2001-2015 (15)	10 (66.7%)	12 (80.0%)

Credit to Dr. Demi Sun

Seasonal Forecast of Summer and Annual Rainfall

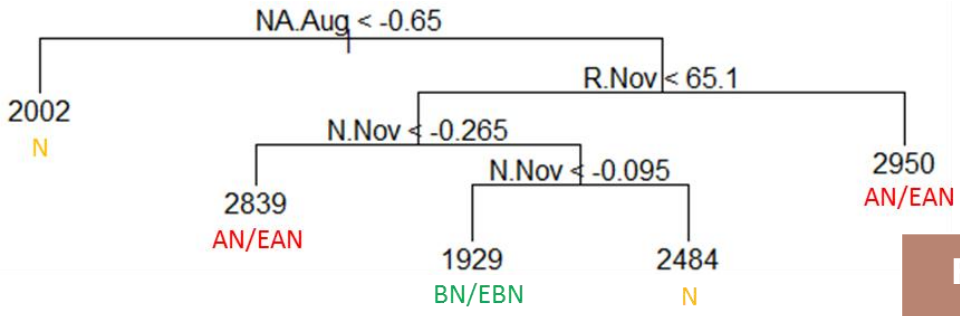
- **Objective:** to forecast the annual rainfall by the end of December and the summer rainfall by the end of May.
- Data are divided into Training Period (1961 – 2000) and Testing Period (2001-2015).
- Category of annual / summer rainfall in Hong Kong

Category	Annual Rainfall (mm)	Percentage	Summer Rainfall (mm)	Percentage
Extremely below normal (EBN)	< 1857	18%	< 799	8%
Below normal (BN)	1857 - 2107	16%	799 - 1037	33%
Near normal (N)	2017 - 2607	31%	1037 - 1513	38%
Above normal (AN)	2607 - 2856	18%	1513 - 1752	6%
Extremely above normal (EAN)	> 2856	16%	> 1752	15%

Credit to Dr. Demi SUN

Seasonal Forecast of Summer and Annual Rainfall

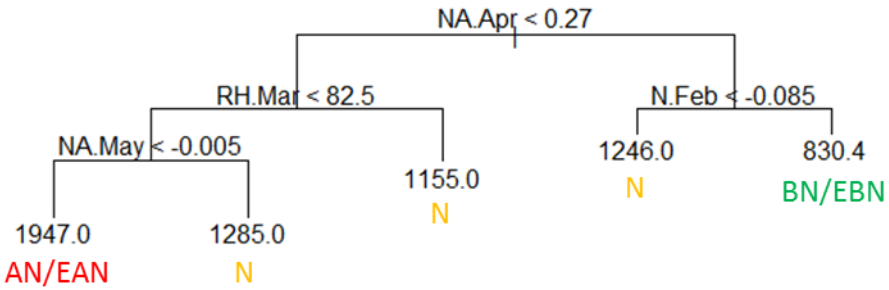
- Optimal tree for annual rainfall and its performance



Validation Period

Predicted / Observed	BN+EBN	N	AN+EAN
BN+EBN	**3/4**	1/4	0/4
N	0/3	**1/3**	2/3
AN+EAN	0/3	0/3	**3/3**

- Optimal tree for summer rainfall



Predicted / Observed	BN+EBN	N	AN+EAN
BN+EBN	**4/5**	1/5	0/5
N	0/3	**2/3**	1/3
AN+EAN	0/1	0/1	**1/1**

Credit to Dr. Demi SUN

THANK YOU !

References

- Camargo, S. J., and A. H. Sobel, 2005: Western North Pacific
- Chia, H. H., and C. F. Ropelewski, 2002: The interannual variability in the genesis location of tropical cyclones in the northwest Pacific. J. Climate
- Chan, J. C. L., 1985: Tropical cyclone activity in the northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. Mon. Wea. Rev. http://www.macaodaily.com
- Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. Bull. Amer. Meteor. Soc. Wallace, J. M., and Hobbs, P.V., 1977: Atmospheric Science: an Introduction survey, Academic Press.
- Hartmann, D. L., 1994: Global Physical Climatology, Academic Press.
- G.L. Hammer, N. Nicholls, C. Mitchell, 2000, Application of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems. Kluwer Academic Publishers.
- Zhang, S., S.-P. Xie, Q.-Y. Liu, Y.-Q. Yang, X.-G. Wang, and Z.-P. Ren, 2009: Seasonal variations of Yellow Sea fog: observations and mechanisms, Journal of Climate, Vol. 22
- Beiderwieden, E., Wolff, V., Hsia, Y.-J., and Klemm, O, 2008: It goes both ways: measurements of simultaneous evapotranspiration and fog droplet deposition at a montane cloud forest, Hydrol. Processes.
- JJ Boreux, J Guiot, 1992: A Fog Forecasting Method In A Deeply Embanked Valley, Atmospheric Environment, Part A, General Topics.
- Wang, B., and J. C. L. Chan, 2002: How strong ENSO events affect tropical storm activity over the western North Pacific. J. Climate
- Zhao, H., L. Wu, and R. Wang, 2014: Decadal variations of intense tropical cyclones over the western North Pacific during 1948–2010. Adv. Atmos. Sci.