



Verification Tools for the Seven-Day Weather Forecast with Weather Companies Worldwide in Taiwan

Chun-Han Huang, Chung-Kung Lin, Ping-Yu Lin, Treng-Shi Huang, Kuo-Chen Lu

Weather Forecast Center, CWB

2020/10/15

Outline



- Motivation
- Data Usage
- Temperature
- Rain
- Summary

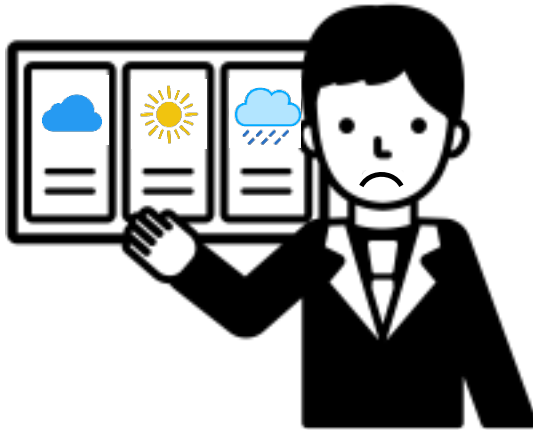
Motivation

APP / Web

- People who get weather forecast from not only CWB now

Insufficient

- Good enough guidance for Forecaster?



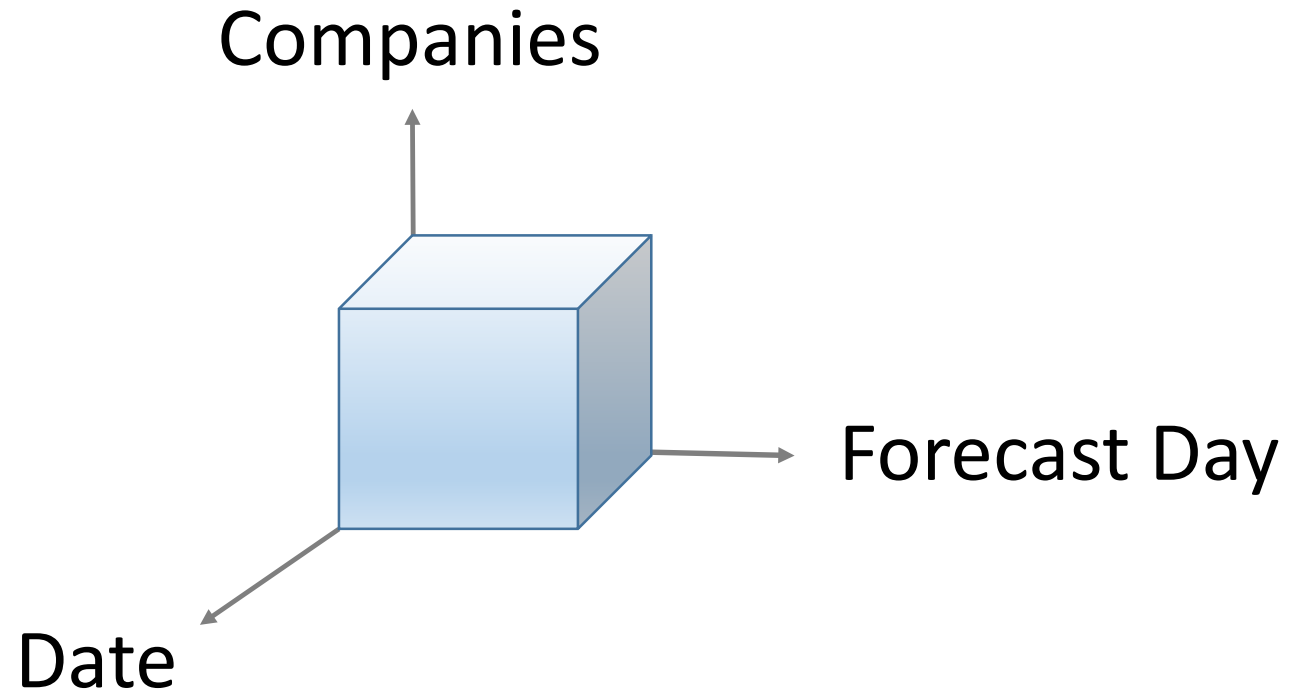
Advantage

- Local model? Global model?
- Temperature? Precipitation?

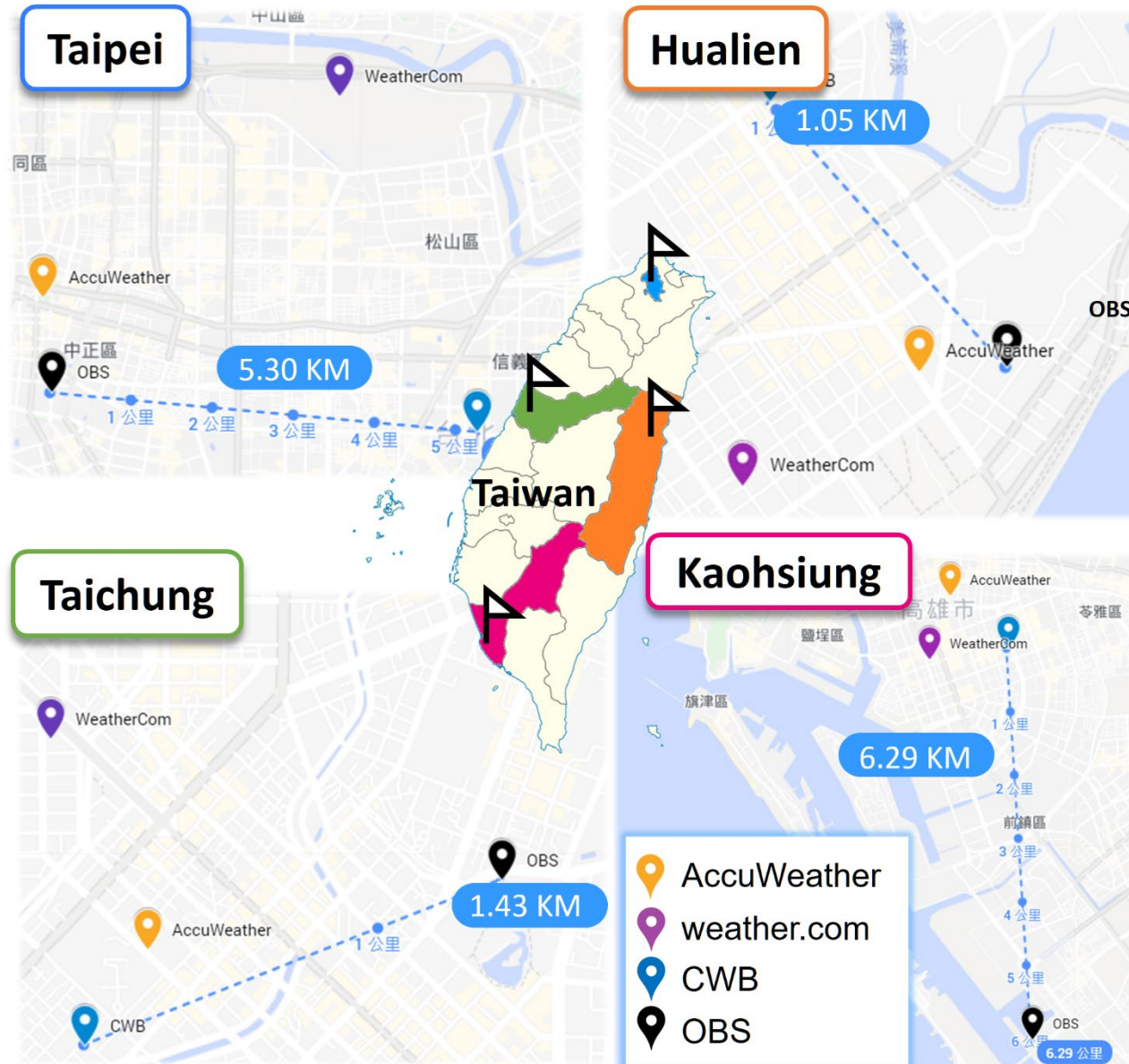
Strategies

- Build up a check and verification operation

Data



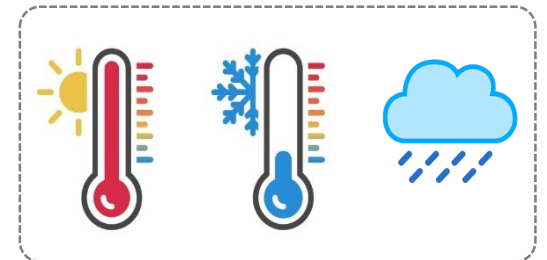
Geolocation of Weather Company in Four Areas



Weather Companies



Verification Items



Data Formation



Temperature High

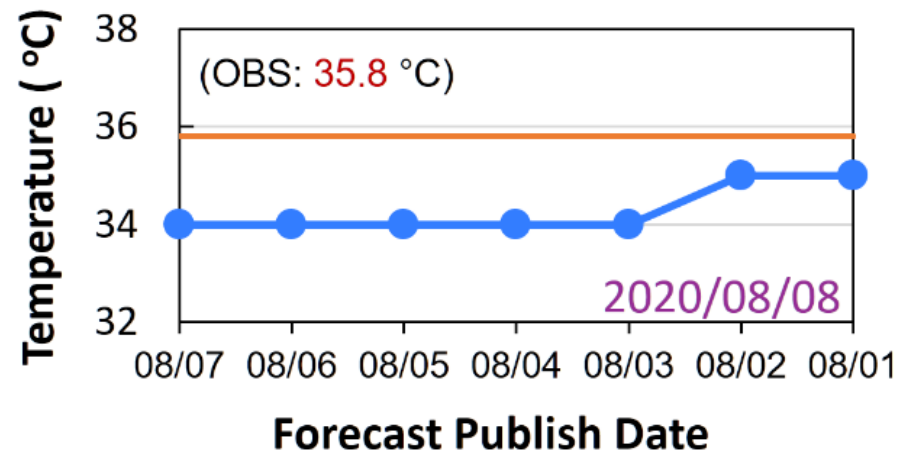


Rain

Forecast Date

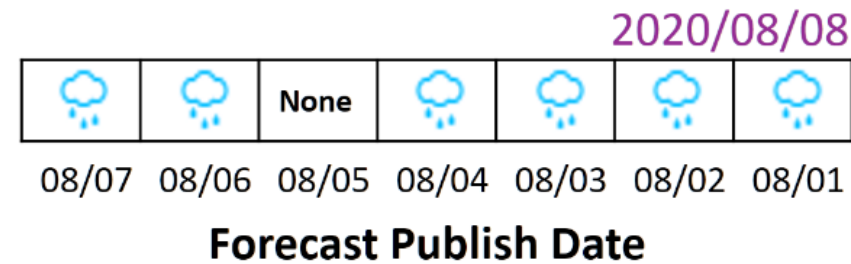
	2	3	4	5	6	7	8	9	10	11	12	13	14
1	35	33	34	35	35	35	35						
2		32	33	34	35	35	35	34					
3			33	34	34	34	34	34	34				
4				34	35	35	34	33	33	33			
5					35	34	34	33	31	32	33		
6						34	34	34	33	32	33	34	
7							34	34	33	32	32	34	34

(°C)



Forecast Date

	2	3	4	5	6	7	8	9	10	11	12	13	14
1	☁	☁	☁	None	None	☁	☁						
2		☁	☁	☁	None	None	☁	☁					
3			☁	☁	☁	☁	None	☁	☁				
4				☁	☁	☁	☁	☁	☁	☁			
5					☁	☁	☁	☁	☁	☁	☁		
6						☁	☁	☁	☁	☁	☁	☁	
7							☁	☁	☁	☁	☁	None	None



Verification

Verification



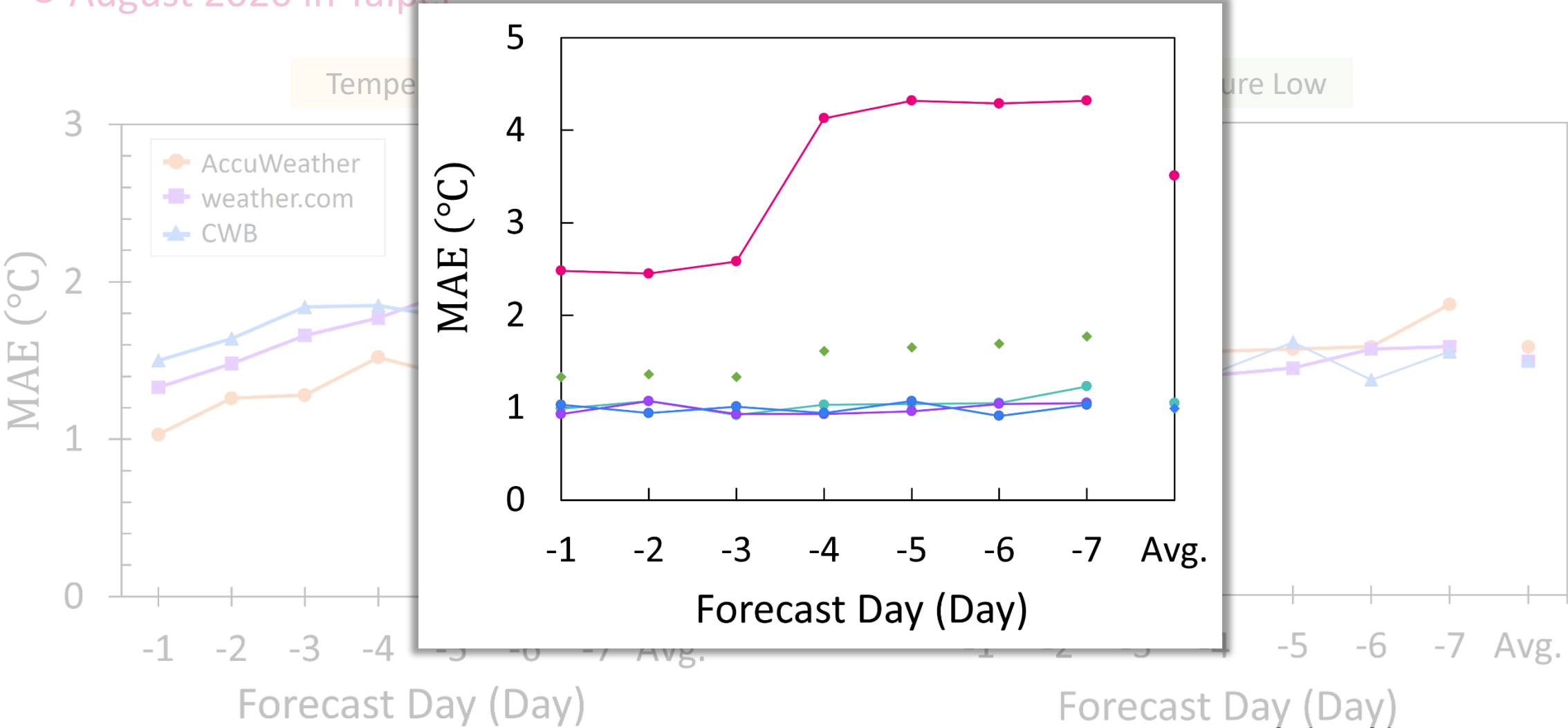
Temperature

High & Low



The Companies' Monthly MAE Verification of the Temperature

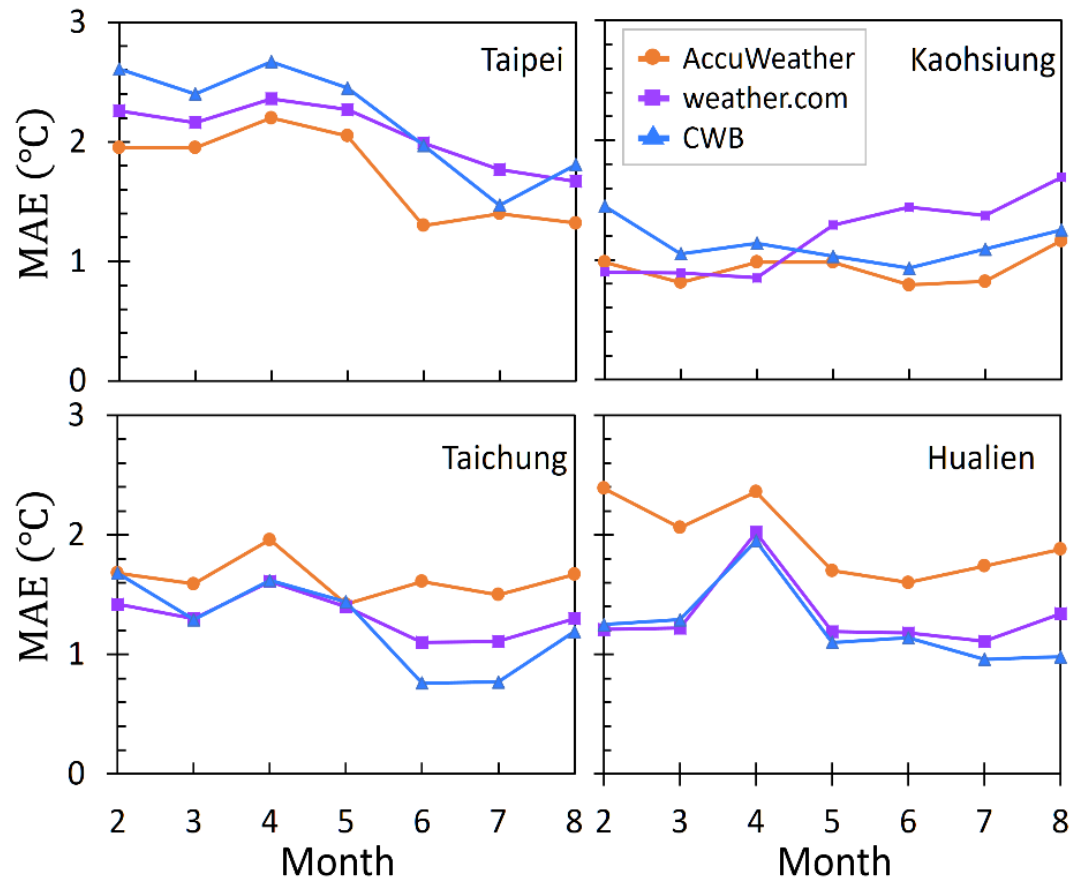
● August 2020 in Taipei



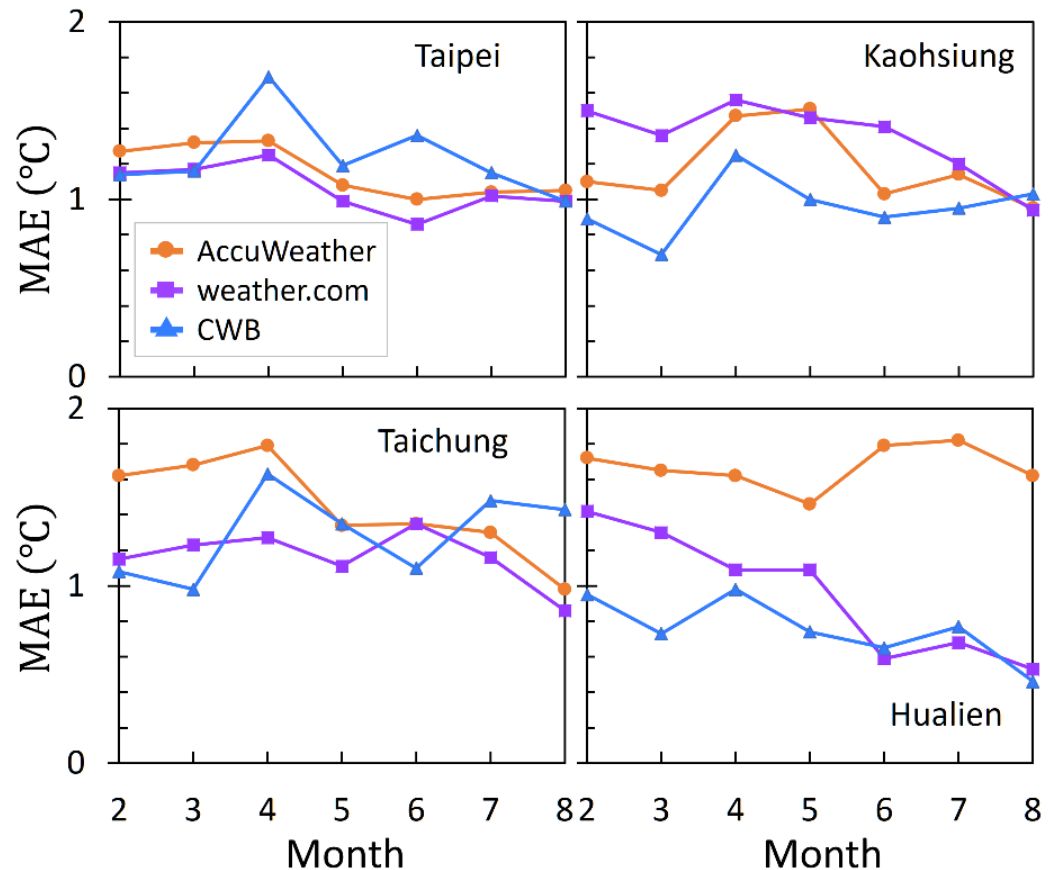
The Companies' Monthly MAE Verification of the Temperature

● August 2020 in Taipei

Temperature High

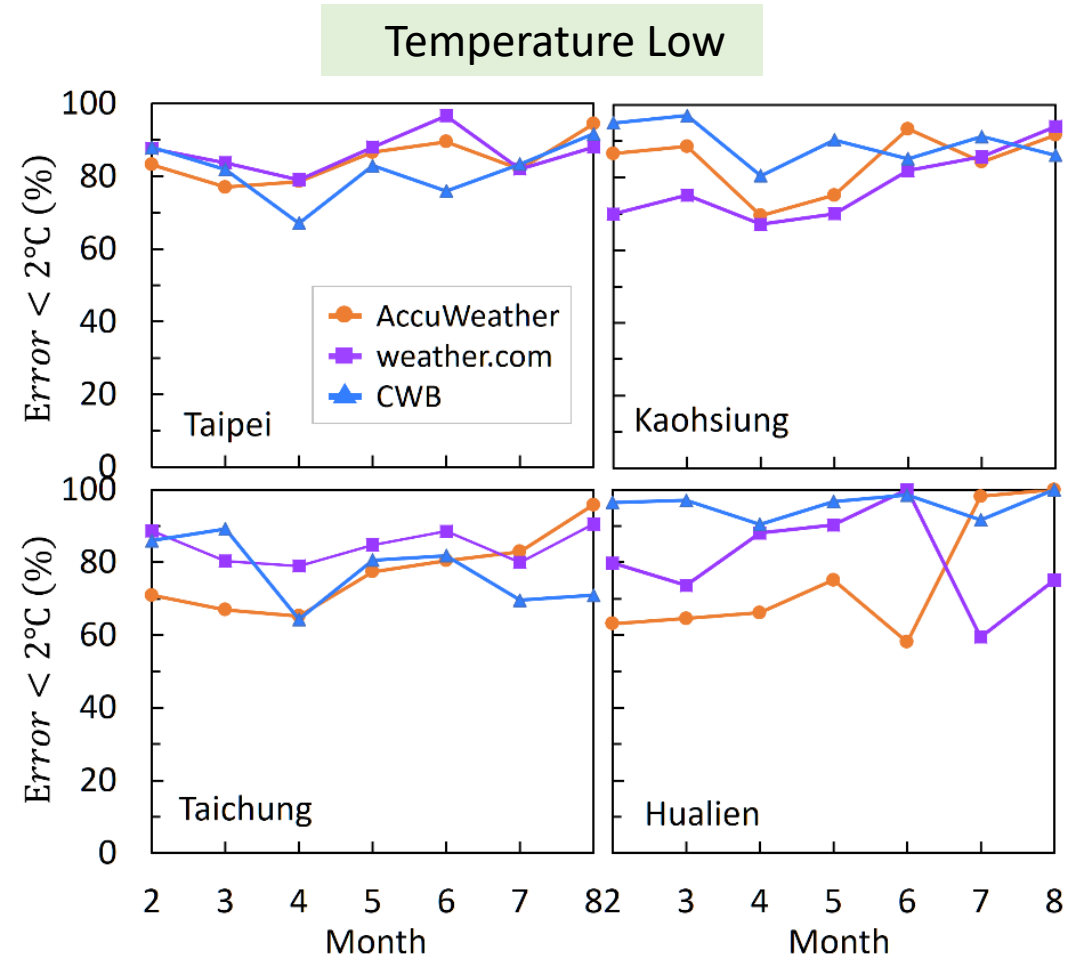
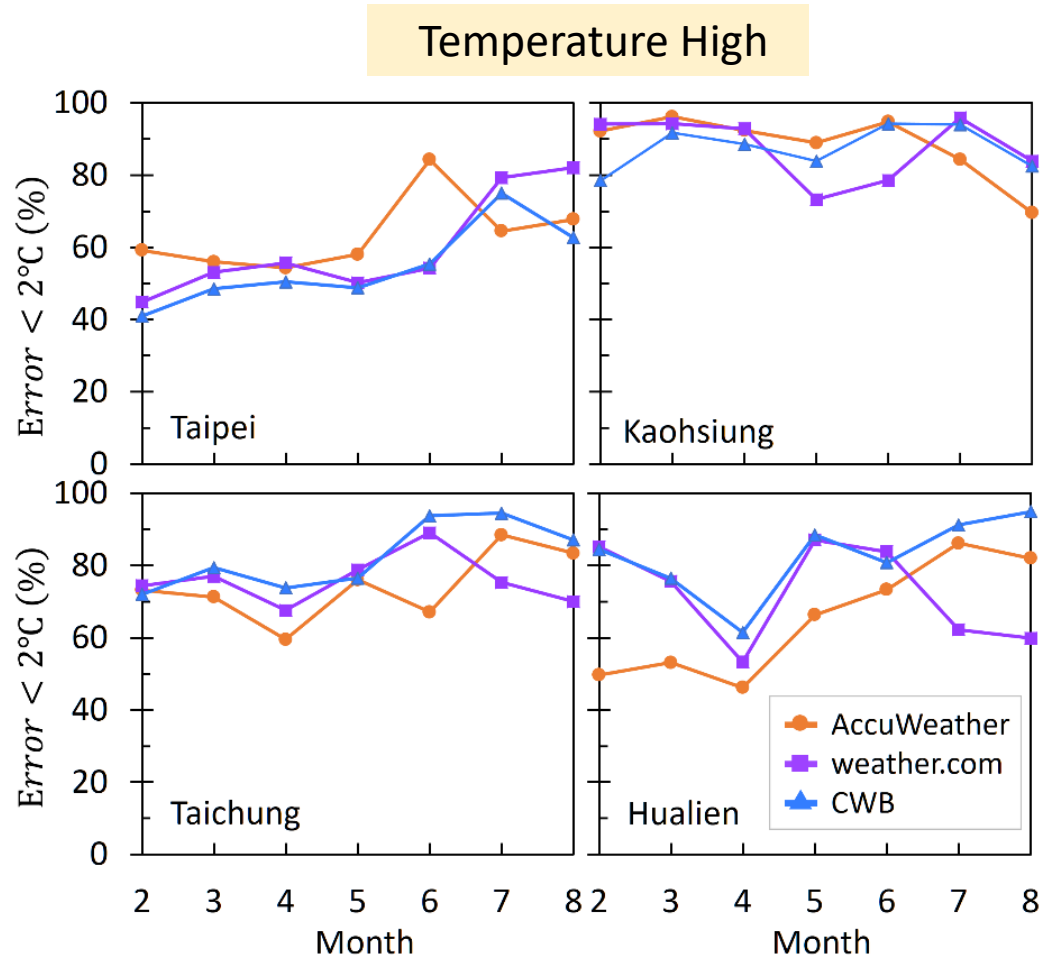


Temperature Low



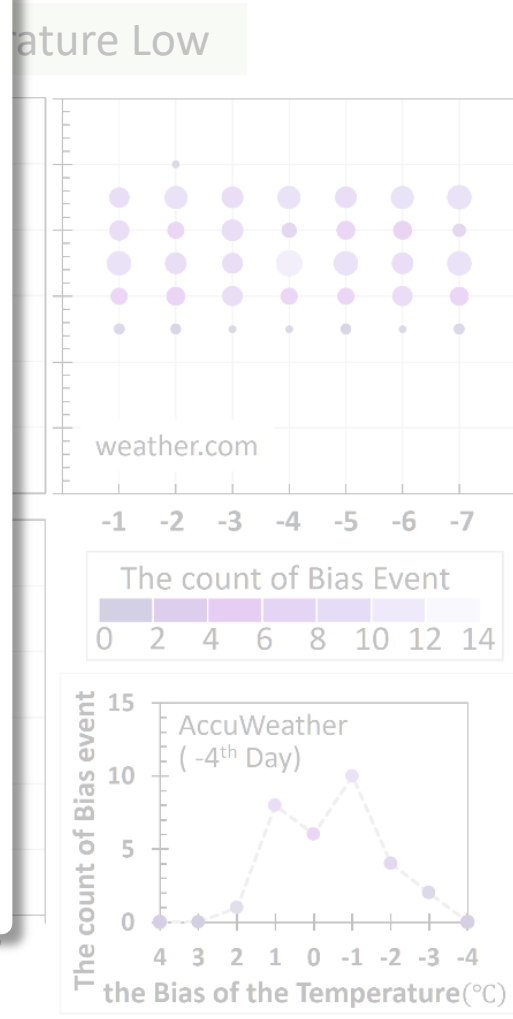
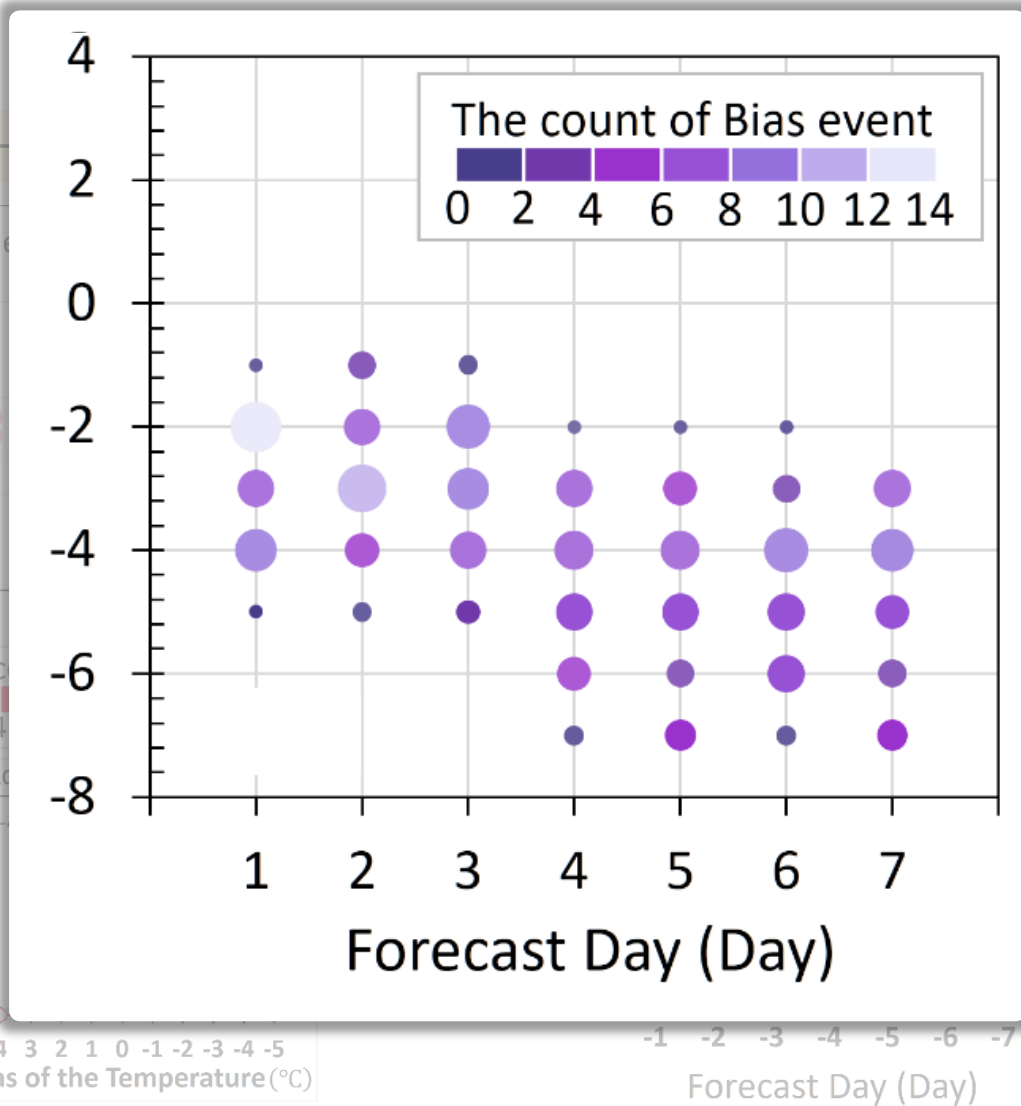
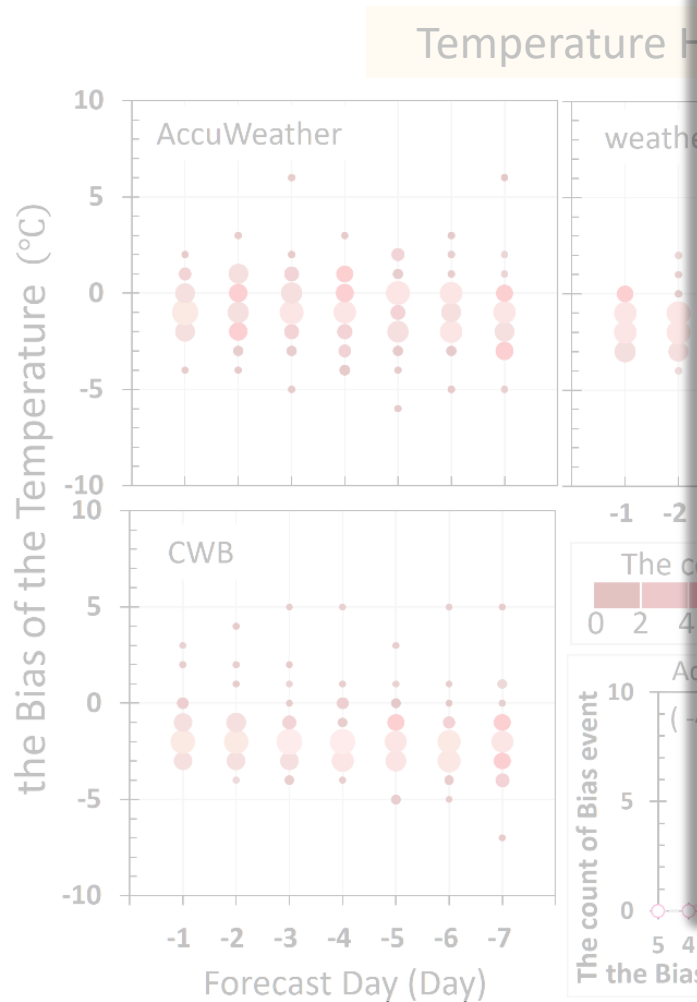
The Companies' Monthly Bias Verification of the Temperature

● August 2020 in Taipei



The Companies' Monthly Bias Analysis of the Temperature

● August 2020 in Taipei





Rain



The Companies' Monthly Verification of the Rain Weather

● August 2020 in Taipei

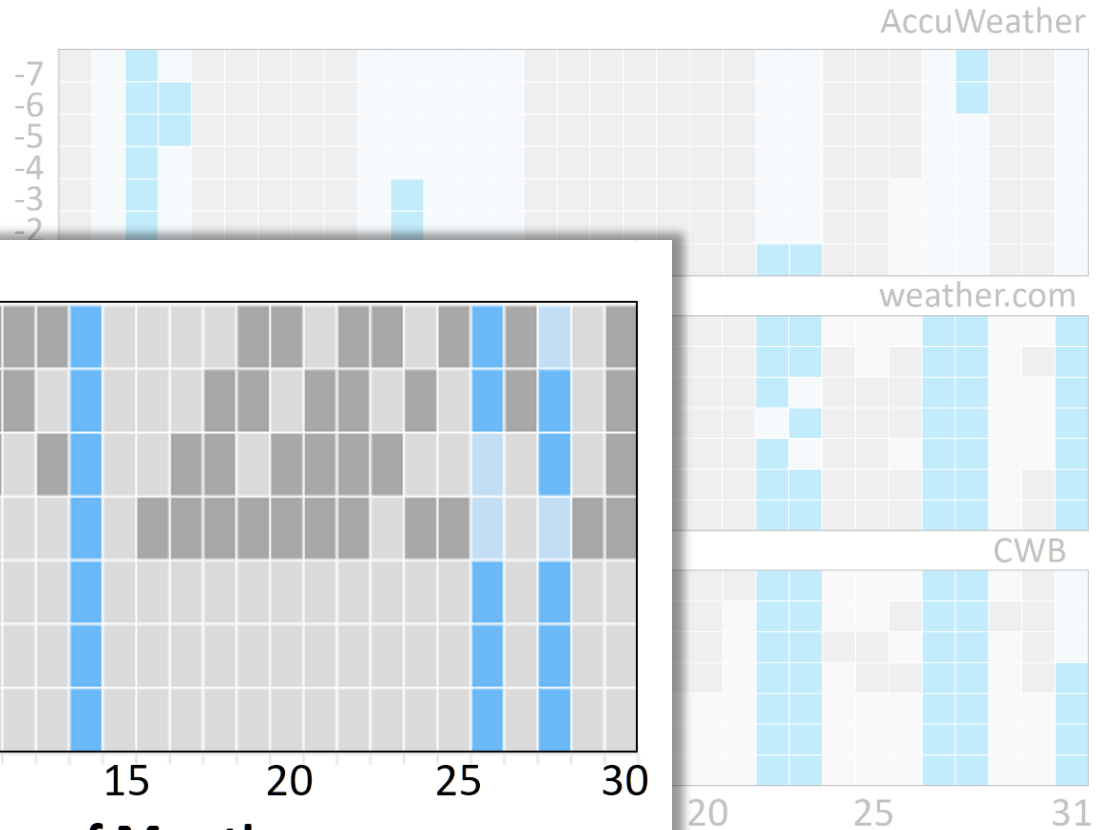
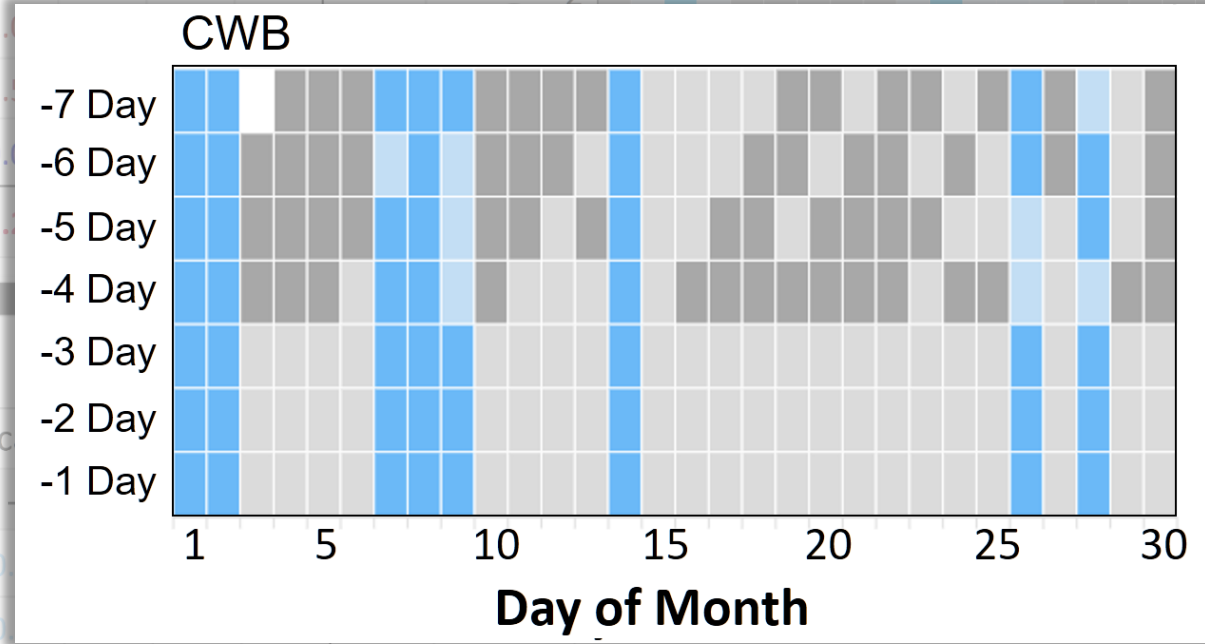
• Bias Score

Company	Forecast Day							Avg.
	-1	-2	-3	-4	-5	-6	-7	
weather.com	1.54	1.62	1.69	2.0				
CWB	1.23	1.38	1.54	1.1				
AccuWeather	0.38	0.23	0.23	0.9				
Avg per Day	1.05	1.07	1.15	1.1				



• CSI

Company	Forecast Day							Avg.
	-1	-2	-3	-4	-5	-6	-7	
weather.com	0.71	0.72	0.57	0.9				
CWB	0.42	0.42	0.42	0.9				
AccuWeather	0.29	0.14	0.14	0.08	0.15	0.23	0.15	0.17
Avg per Day	0.47	0.43	0.38	0.39	0.42	0.40	0.37	

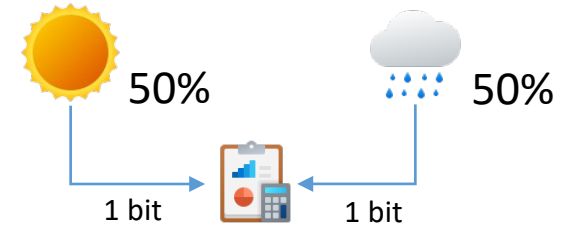


Forecast \ Obs.		Contingency table	
		Rain	No Rain
Forecast	Rain	● Hits	● False Alarm
	No Rain	● Missed	● Negative Correct

Entropy in Information Theory

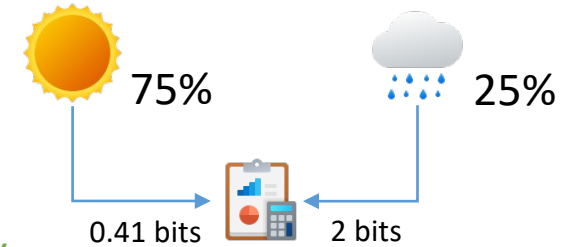
1

bit = uncertainty divided by 2 (states)
useful message = 1 bit = $\log_2 2$



2

bits = uncertainty divided by 4 (states)
useful message {
= 2 bits = $\log_2 \left(\frac{1}{0.25}\right)$ for rainy
= 0.41 bits = $\log_2 \left(\frac{1}{0.75}\right)$ for sunny



→ on average = $25\% \times \log_2 \left(\frac{1}{0.25}\right) + 75\% \times \log_2 \left(\frac{1}{0.75}\right) = 0.81$ bits

3

• Shannon Entropy : $H(P) = -P \log_2(P)$, where P as probability

$$H(\mathbf{P}) = - \sum_i p_i \log_2(p_i)$$

Entropy in Information Theory

4

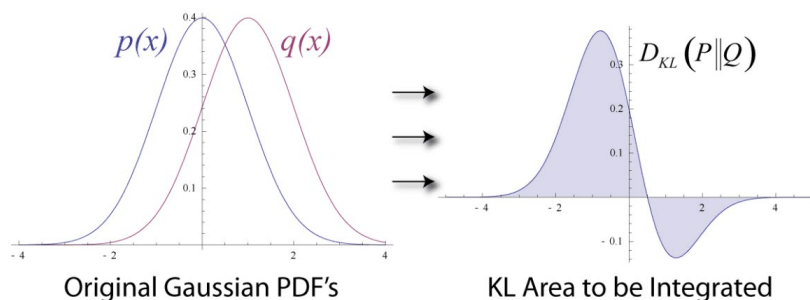
- Shannon Entropy : $H(P) = -P \log_2(P)$
- Cross-Entropy : $H(P, Q) = -P \log_2(Q) \geq$ Shannon Entropy

the cost from identifying event of probability P (Ground Truth) and probability Q (Prediction)

5

- Kullback-Leibler divergence : $D_{KL}(P \parallel Q) =$ Cross-Entropy – Shannon Entropy
(without symmetry)

$$= H(P, Q) - H(P) = \sum_x^N P(x) * \log_2 \left(\frac{P(x)}{Q(x)} \right)$$



distance from probability Q to probability P

6

- Jensen-Shannon divergence: $\frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(M \parallel P) , M = \frac{P + Q}{2}$
(with symmetry)

Peter Dayan, 1995: Neural Computation 7, pp. 889-904

Kullback, S.; Leibler, R. A. , 1951: "On Information and Sufficiency", . Ann. Math. Statist. 22, no. 1, 79--86.

Jianhua Lin, 1991:"Divergence Measures Based on the Shannon Entropy",IEEE Transacions of information theory, vol. 37, no. 1

https://en.wikipedia.org/wiki/Kullback-Leibler_divergence

Mean-Fluctuation-Continuation divergence, **MFC**-divergence

7

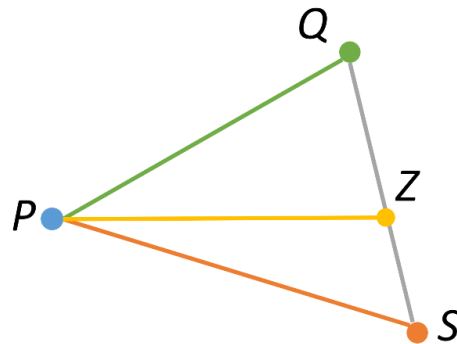
A new way to find the Entropy with two reference probability

Mean-Fluctuation-Continuation divergence (MFC-divergence):

$$D_{MFC}(P \parallel Z) = \frac{1}{\sqrt{2}} \sqrt{D_{KL}(P \parallel Q)^2 + D_{KL}(P \parallel S)^2 - 2D_{KL}(Q \parallel Z)^2}, \text{ with symmetry}$$

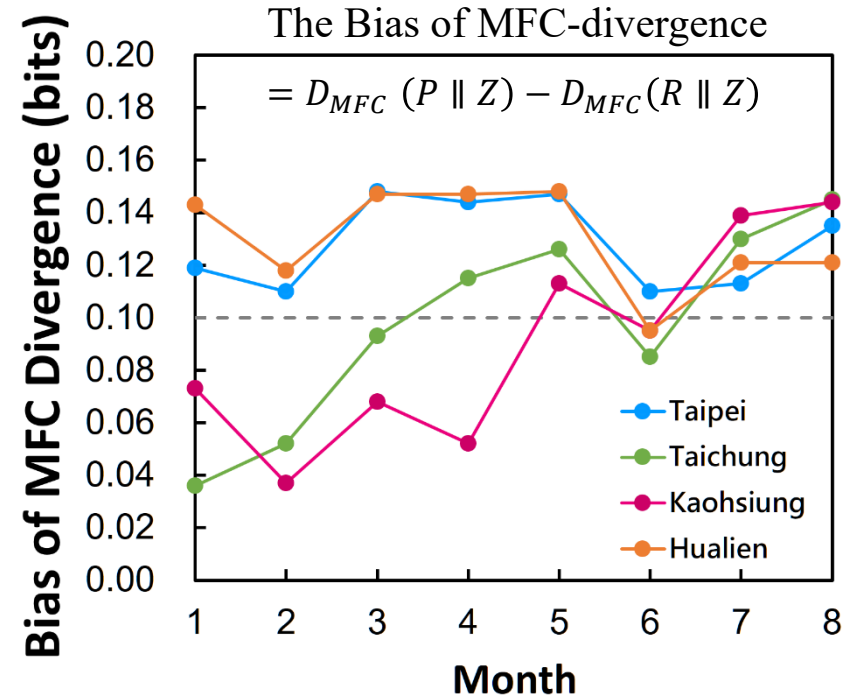
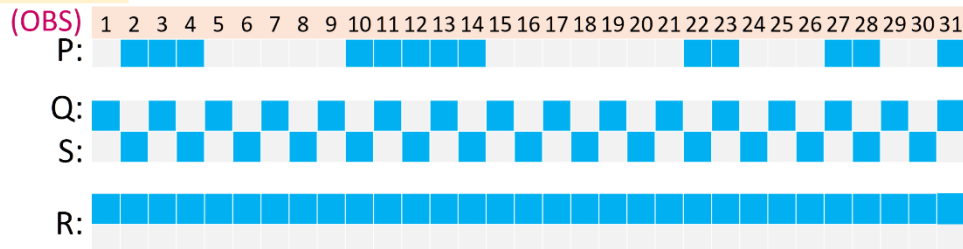
$$\text{, where } D_{KL}(Q \parallel Z) = \frac{D_{KL}(Q \parallel S) * D_{KL}(S \parallel Q)}{2}$$

, distance from probability Q and probability S to probability P



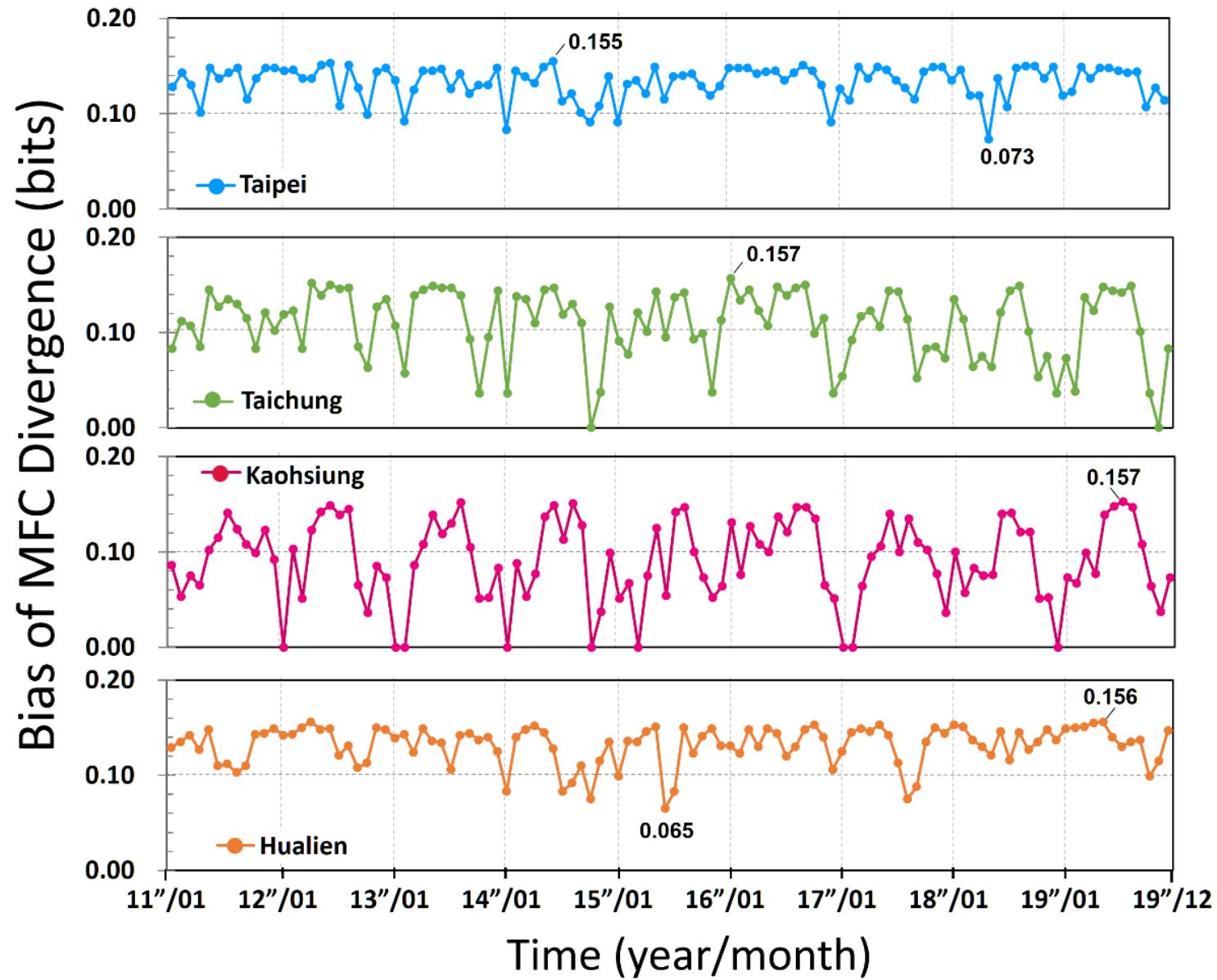
$$0 \leq D_{KL}(P \parallel Z) \leq \max[D_{KL}(P \parallel Q), D_{KL}(P \parallel S)]$$

Taipei Date: 2020/08, Entropy = 0.135 (bits) ■ Rain, ■ No rain



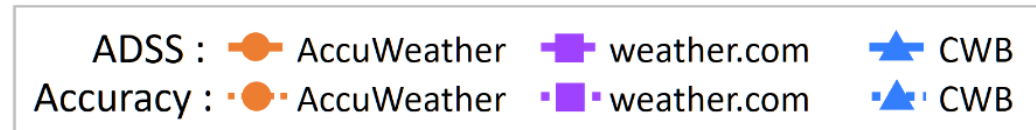
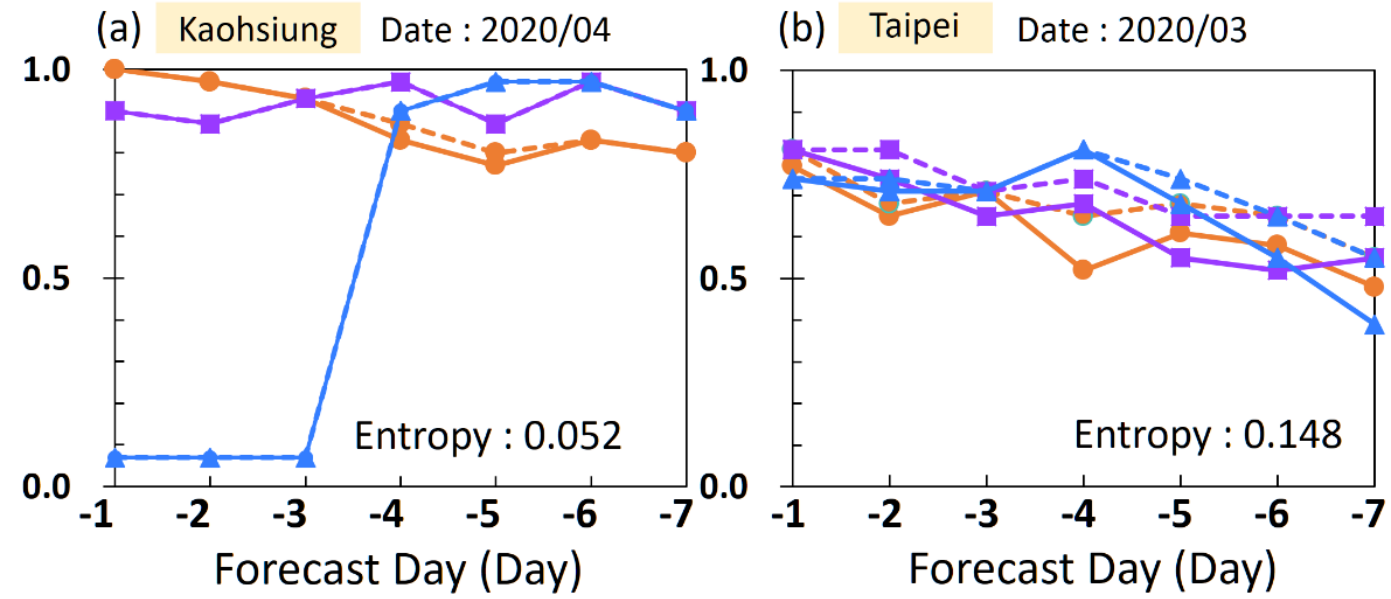
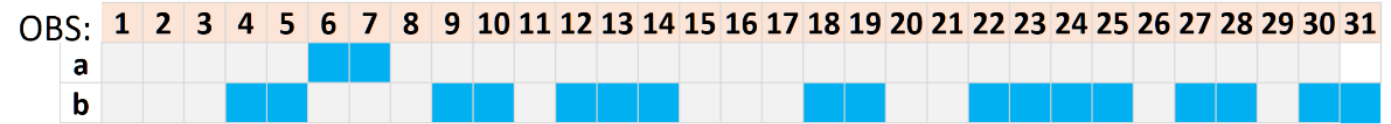
Here the notation R is a probability function in which all values are filled with zero or one (inverse to the zero).

The Entropy of the Rain Weather Observation in 2011-2019



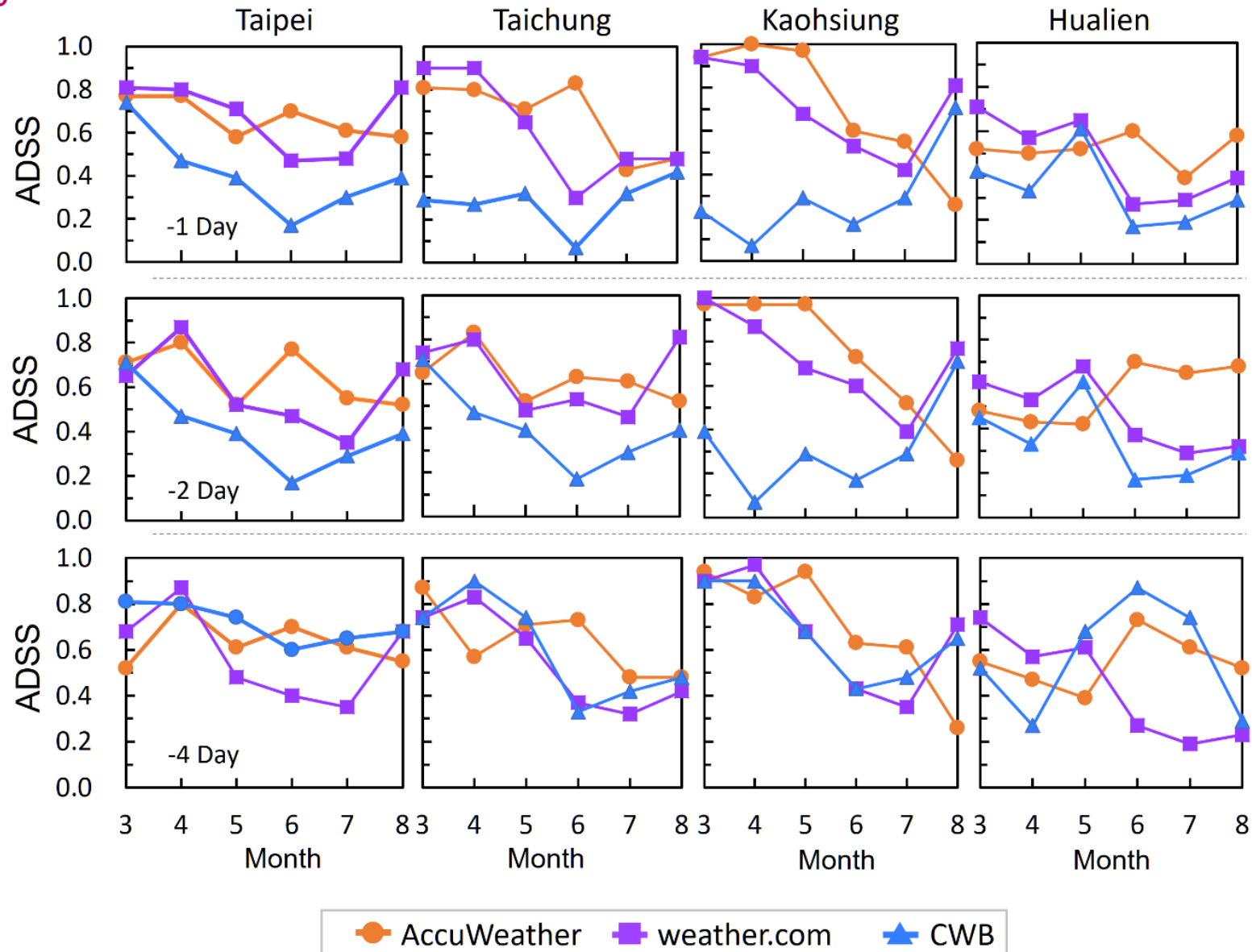
The Accuracy of Data Stability Score, ADSS

Contingency Table		
(Event Count)	Obs. Rain	Obs. No Rain
Forecast Rain	A Hits	B False Alarms
Forecast No Rain	C Misses	D Negative Corrects
Accuracy	= $\frac{A + D}{A + B + C + D}$	
ADSS	= $\frac{A + D - E}{A + B + C + D} \leq \text{Accuracy}$	
E : The amount of discontinuous event of hits and negative corrects.		

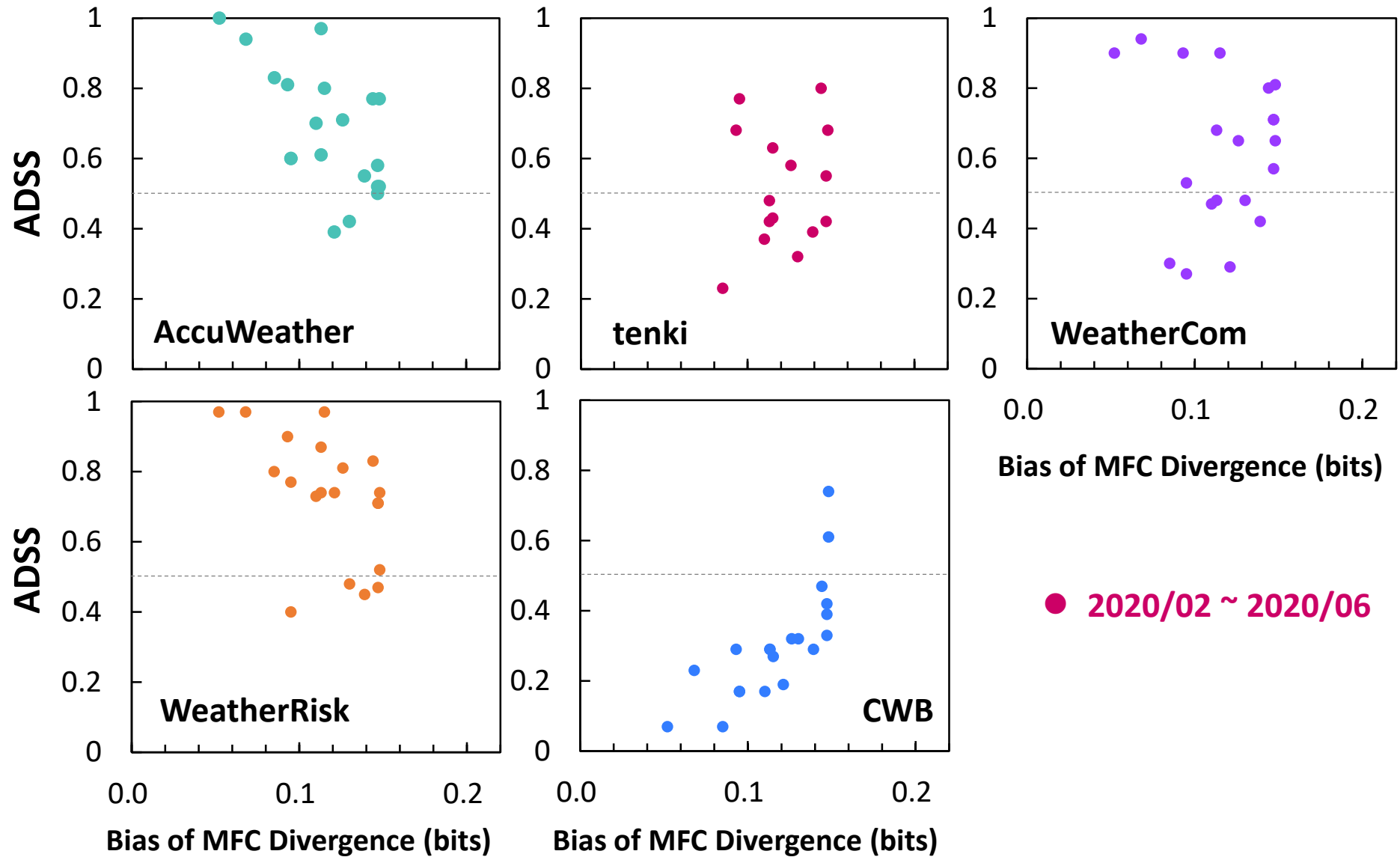


The ADSS of the Companies in Four Cities

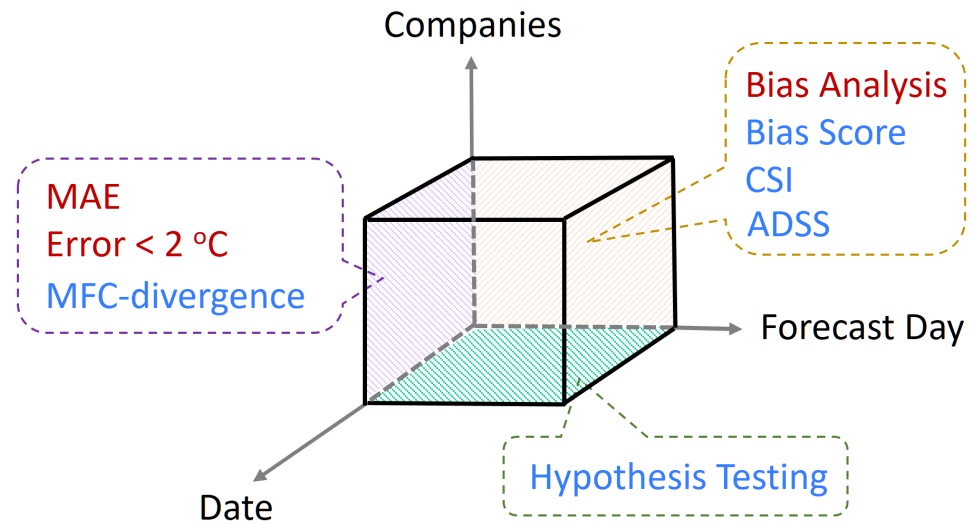
● 2020



The Bias of MFC-divergence of the Companies in Four Cities



Summary



The dimension of the seven-day verification in temperature and rain, and its related diagnostic tools

- There is a simple and clearly way to show **the weather forecast from multiple companies in web style**.
- It shows not only all forecast data, but the **trend of the bias** with different score index
- By analyzing the verification data, it may be a way to improve our 7-day forecasts product

All developed tools will be opened on the open-source platform, such as Git.



Thanks for Your Attention and Listening
