利用機器學習優化微物理參數法中的碰撞收集過程

洪語澤 陳正平 王理甲 蔡子衿 江琇瑛
Department of Atmospheric Sciences, National Taiwan University

Introduction

- Microphysics parameterizations play a critical role in high-resolution quantitative precipitation forecasts.
- However, the execution time of complex microphysics schemes is relatively long, posing a challenge in accelerating microphysics processes.

	CPU time (min)	
	Ideal case (1 CPU)	Real case (48 CPUs)
MORR	21 (28.0%)	87 (14.3%)
MY2	25 (33.3%)	--
WDM6	24 (32.0%)	124 (20.4%)
NTU-3M	75 (100%)	607 (100%)

To develop a machine learning model that can effectively improve and accelerate microphysics in the model.

Collision coalescence processes

$$\int \frac{dr_A^k}{dt} n(r)dr = \iint [r_C^k - r_A^k] K'(r_A, r_B, C_{air}) n_A(r_A) n_B(r_B)\, dr_A dr_B \quad K' = E_c \frac{\pi}{4}(r_A + r_B)^2 |V_A - V_B|$$

$$\int \frac{dr_B^k}{dt} n(r)dr = \iint [-r_B^k] K'(r_A, r_B, C_{air}) n_A(r_A) n_B(r_B)\, dr_A dr_B \quad E_c = E_{colli} \cdot E_{coal}$$

Why we focus on collision and coalescence processes?

1. Highly nonlinearity
2. Potential bias in current model
3. Can be computational expensive

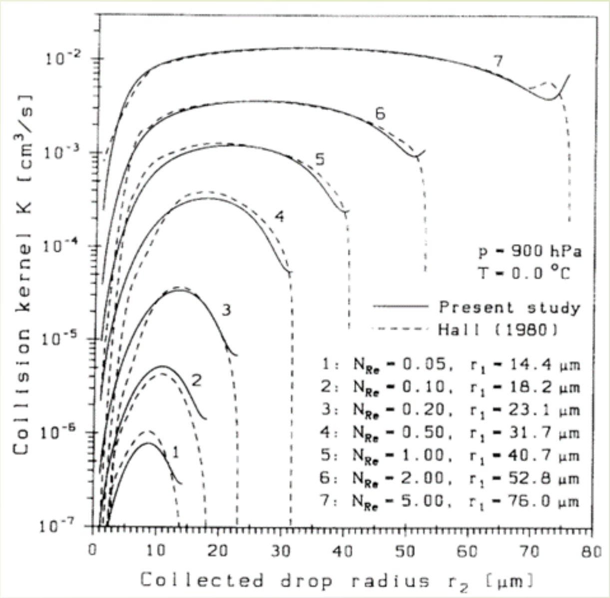
Methodology

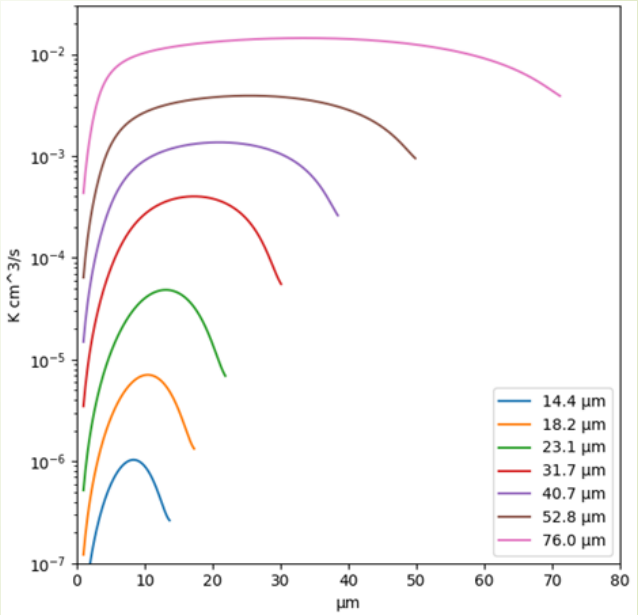
1. Building “theoretical model “
2. Producing training & validation data
3. Training machine learning model
4. Implement into WRF model
5. Evaluating the performance

Theoretical model

- Using C++ for speeding up
- Considering fall speed, collision efficiency, and using numerical integration, following Böhm's work
- Input variables (total ten variables)
  - Hydrometer distribution factors, aspect ratio, density
  - Air density, temperature
- Output variables
  - Change of hydrometer Moments

Böhm (1992c)

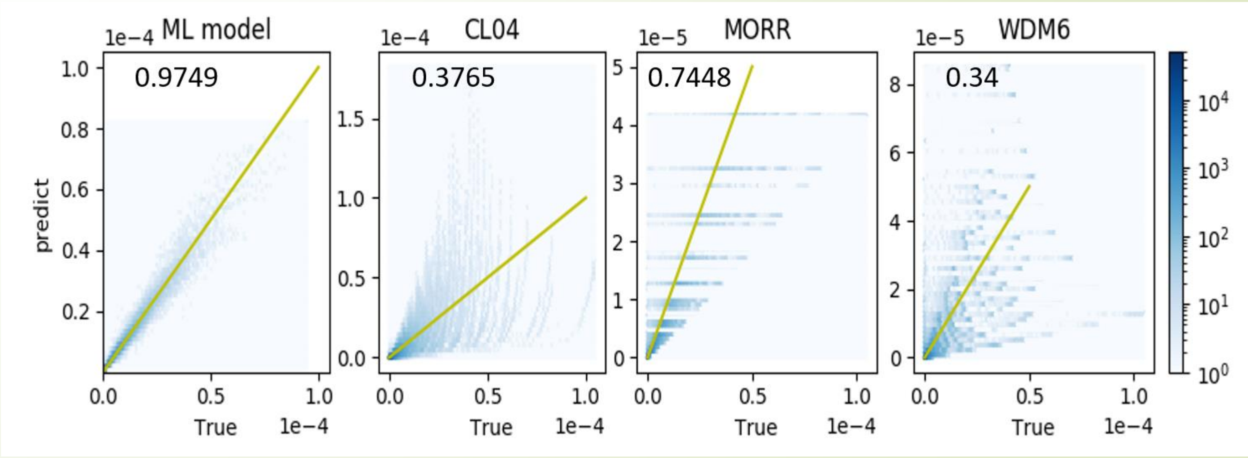




Machine learning Model

- **XGBoost** (eXtreme Gradient Boosting)
  - based on the gradient boosting framework
  - The tree-like model can be easily transform into FORTRAN code
- Training data and result
  - 1M dataset produced by theoretical model
  - The performance of ML model significantly better than other collision kernel in other microphysics scheme

```
double coaldM0(double * input) {
    double var0;
    if (input[3] >= 0.000038873906) {
        if (input[1] >= 0.0000058218866) {
            if (input[3] >= 0.0014235718) {
                if (input[3] >= 0.008511238) {
                    if (input[3] >= 0.018746888) {
                        if (input[9] >= 5.6202826) {
                            var0 = -0.9690507;
                        } else {
                            var0 = -0.7998701;
                        }
                    } else {
 ...
                    } else {
```

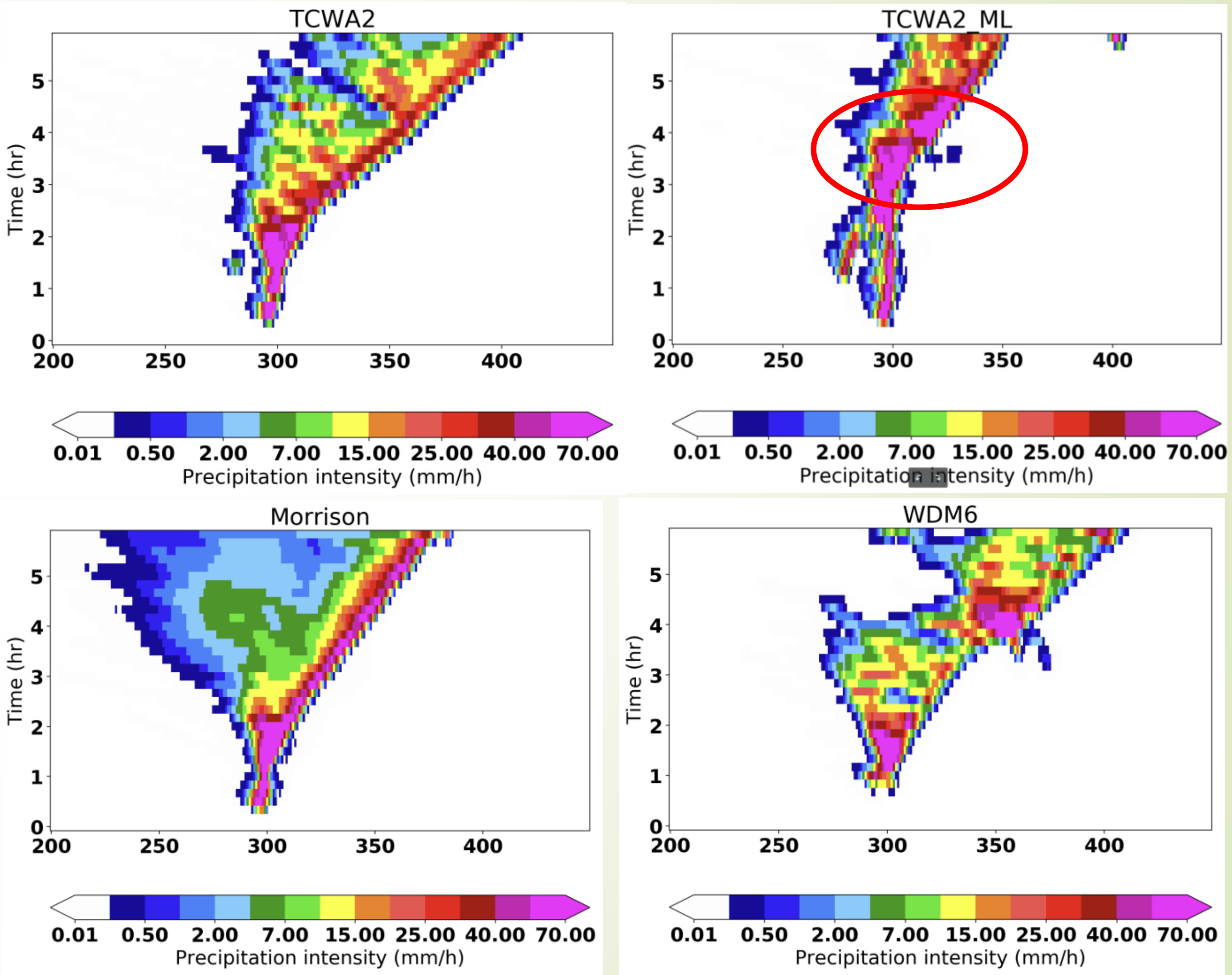


Evaluate the performance of ML model

- Implement into WRF 4.4.2
- Test with Idealized 2d squall line experiment
  - 2d squall line is common idealized case for testing MP scheme
  - Other physic schemes are disabled
- Experiment design:
  - **TCWA2:** CWA two moment MP scheme
  - **TCWA2_ML:** modified TCWA2 warm rain processes with machine learning model
  - **Morrison**
  - **WDM6**

Precipitation intensity

- **TCWA2_ML**
  - Higher convective rainfall
  - Much less stratiform precipitation
  - Squall line moves slower
- **Morrison**
  - Broadest stratiform precipitation

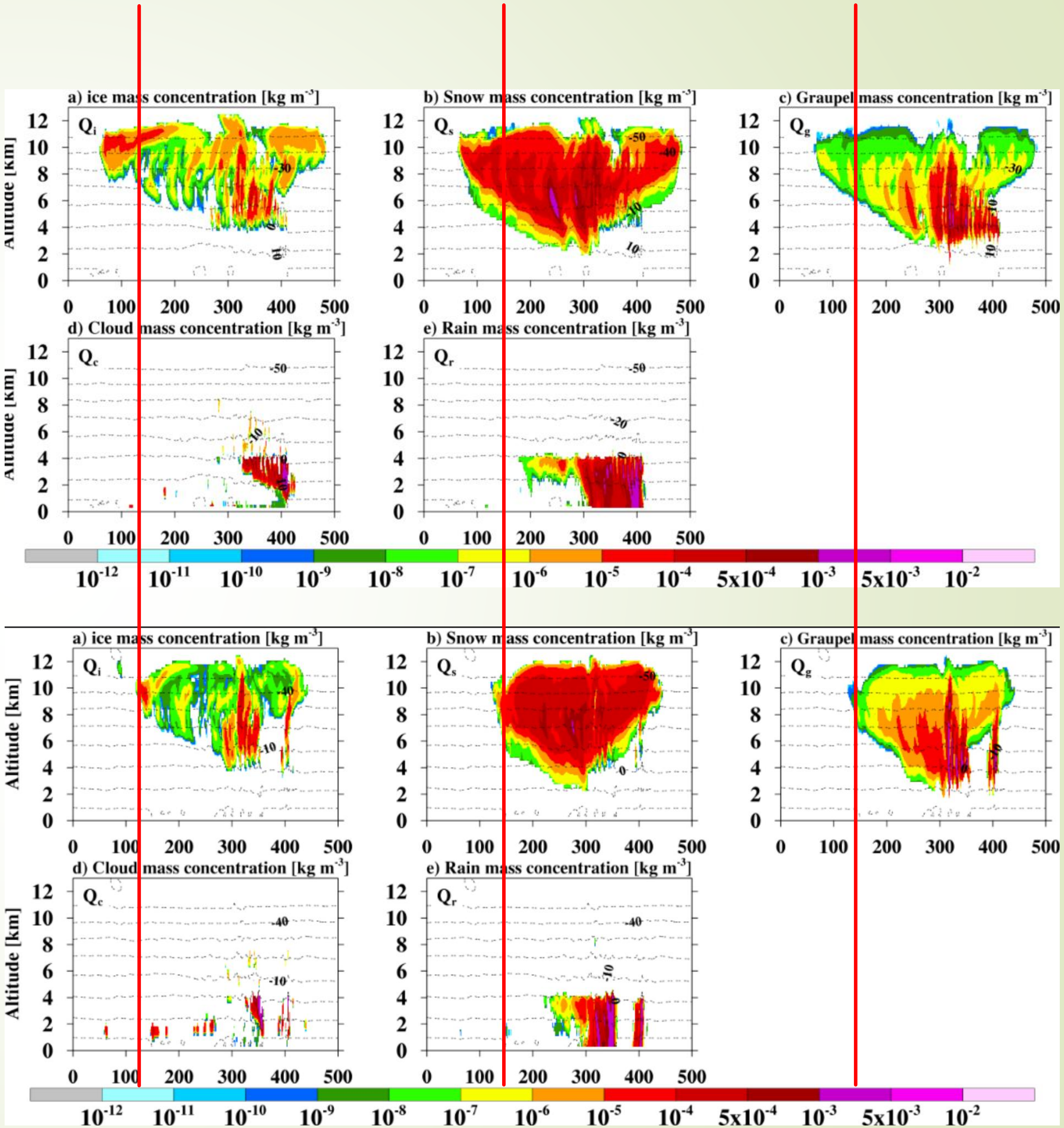


Hydrometer distribution at 6 hr

- Narrower development of stratiform
- May due to stronger warm rain processes

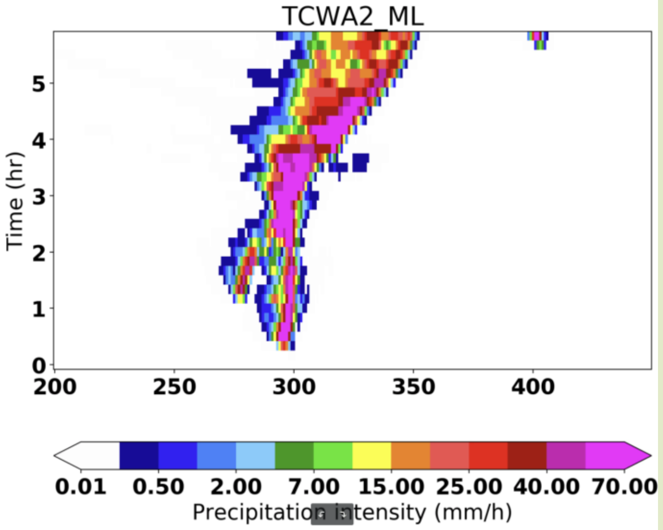
TCWA2

TCWA2_ml

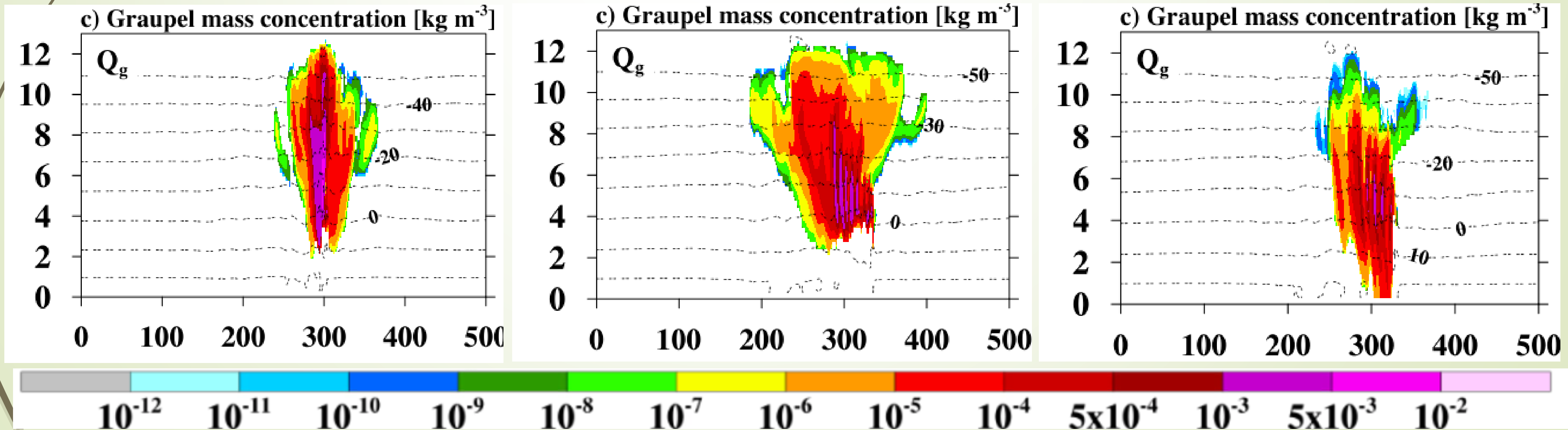


Graupel distribution at 3 hr

- Stronger warm rain processes in TCWA2_ML
- Higher rain drops increased rimming processes
- Higher graupel in TCWA2_ML



TCWA2_ML **TCWA2** **Morrison**



Computational efficiency

- 23% reduced computational time for TCWA2_ML. But still higher than Morrison and WDM6
- Note that Morrison and WDM6 is not complete 2-moment scheme
  - Morrison: cloud drop is one moment
  - WDM6: ice phase hydrometers are one moment

	TCWA2	TCWA2_ML	Morrison	WDM6
seconds	3461	2692	1182	963
%	100%	77%	34%	28%

Summary

1. We develop machine learning model using XGboost algorithm, and the training data is based on theoretical collision model
2. In squall line idealized case, machine learning model enhanced warm rain processes and reduced stratiform cloud development
3. 23% reduced computational time for TCWA2_ML.

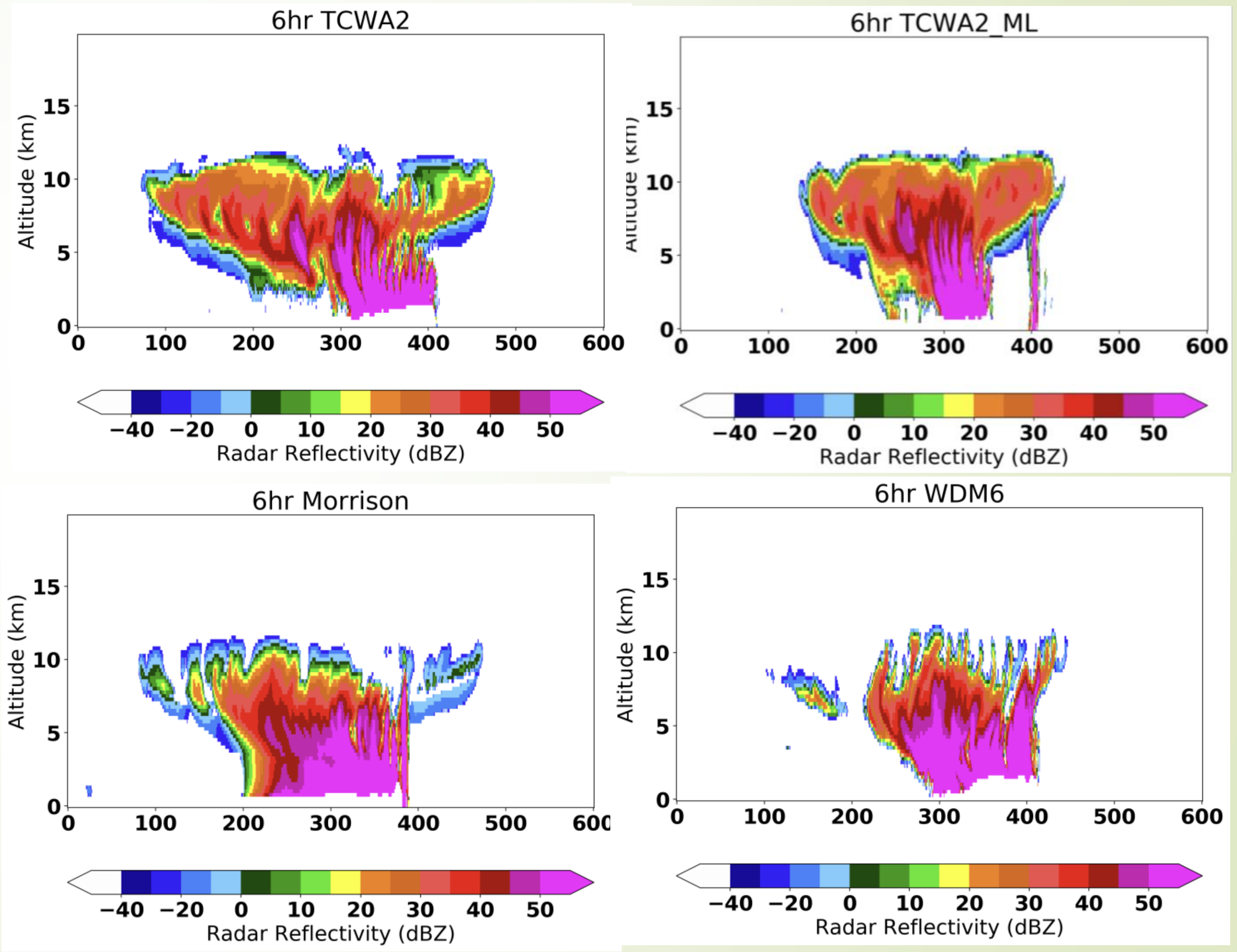
Future work

- Improve the theoretical model
  - Ice phase hydrometers density
  - coalescence efficiency
- Ice phase hydrometers collision coalescence processes
- Real cases simulation

Thank you for listening!

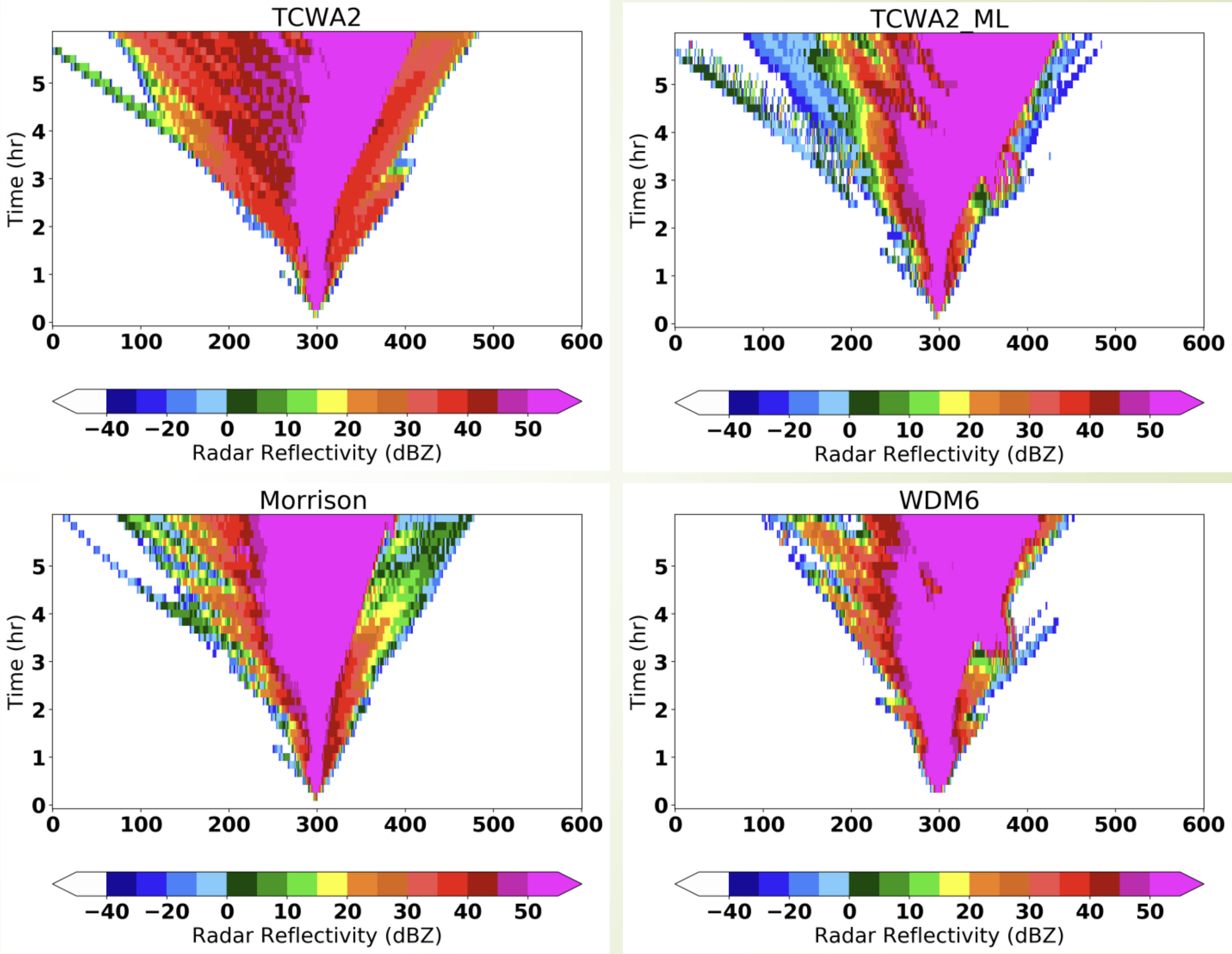
Radar reflectivity

- **TCWA2**相較**Morrison**有更廣泛的分佈，但雷達反射率在5公里以下和對流區域外較弱。顯示這些降水粒子不易到達地面，這主要是冰相粒子終端速度修正的結果
- 傳統參數法碰撞、合併效率高於理論值，而**TCWA2_ML**修正了碰撞合併效率後，降水粒子形成較慢，致使雷達回波因子普遍較弱。



最大回波地面投影

- **Morrison**和**TCWA2**的回波發展較趨向於層狀區，而**TCWA2_ML**與**WDM6**則只在對流區有較強的回波



發展碰撞收集機器學習參數法

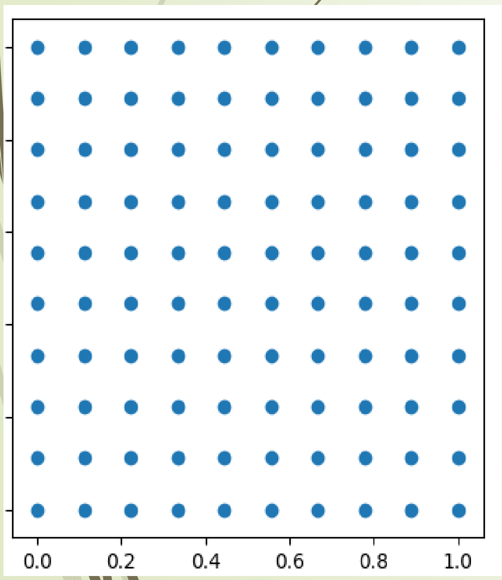
- 隨著AI技術發展，近年有些研究開始探討AI在微物理參數法上的應用，如
  - Geno and Alfonso, 2023
  - Takeshi and Wang, 2024

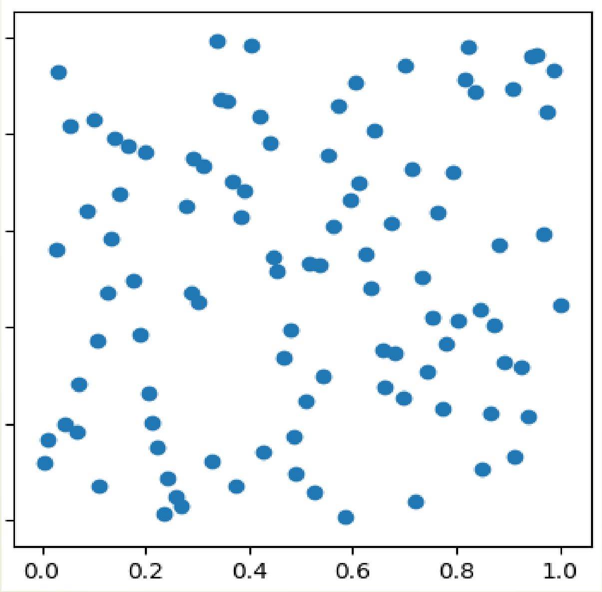
但在各物理過程尤其在碰撞合併核量的處理上，各有不足之處。

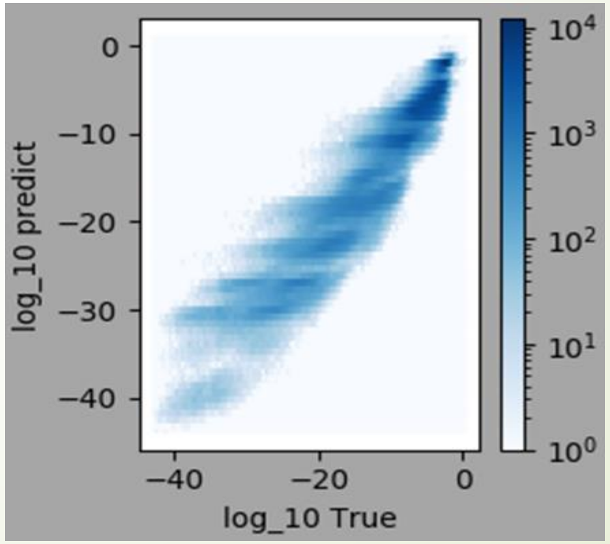
- 本計畫發展之機器學習模式考慮更完整的物理過程，如考慮空氣密度對合併係數的影響，冰相粒子合併過程等等。

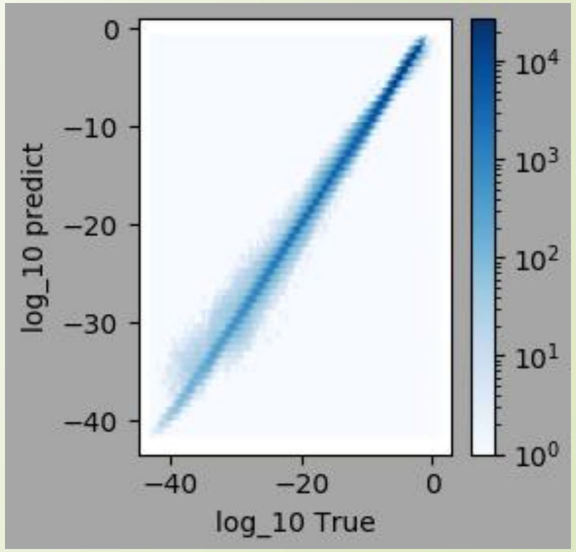
拉丁超立方抽樣

- 由於考慮的物理過程較為完整，使用輸入參數較多，用傳統的等間距取樣效果不佳。因此採用拉丁超立方法進行抽樣。
- 想在N個維度中取K個點，先在每個維度線性切成K個區間，在此維度裡的每個區間只會有一個點，但每個維度都有K個抽樣，可以解決蒙地卡羅抽樣造成抽樣點分佈不均的問題(Stein, 1987)



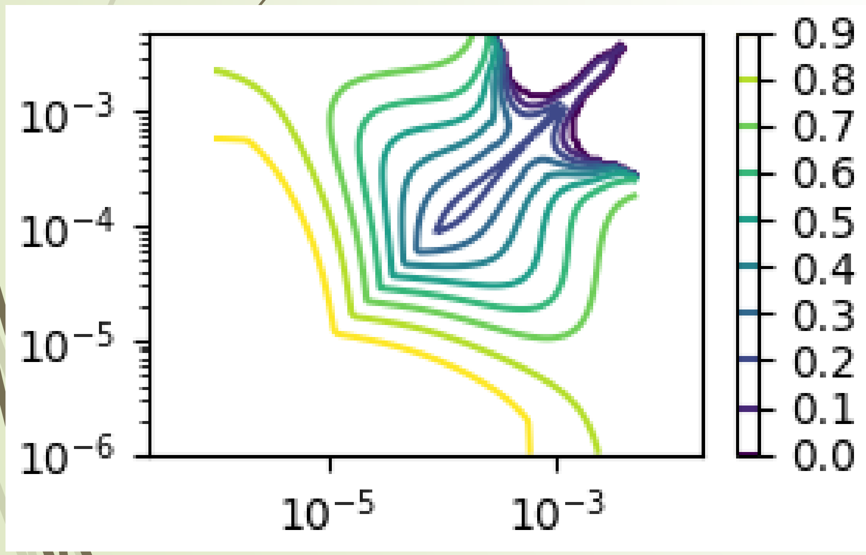






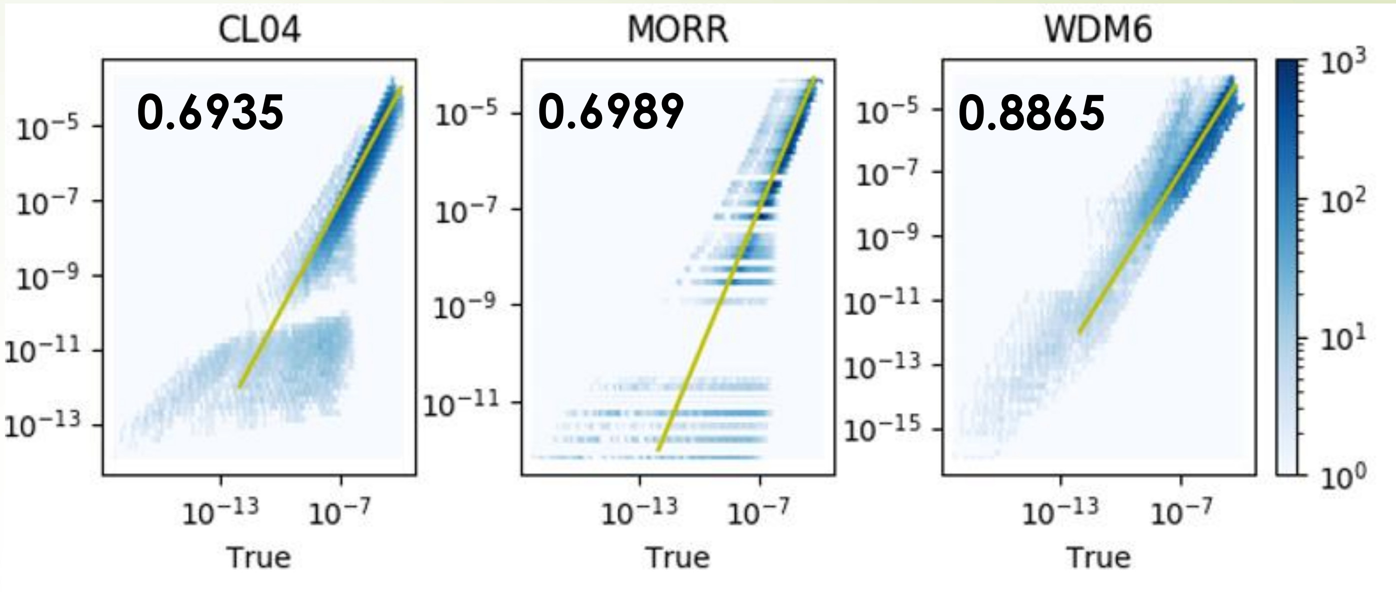
合併係數

- 水物碰撞之後的合併比例
- E_{coal}=1 in WDM6 & Morrison

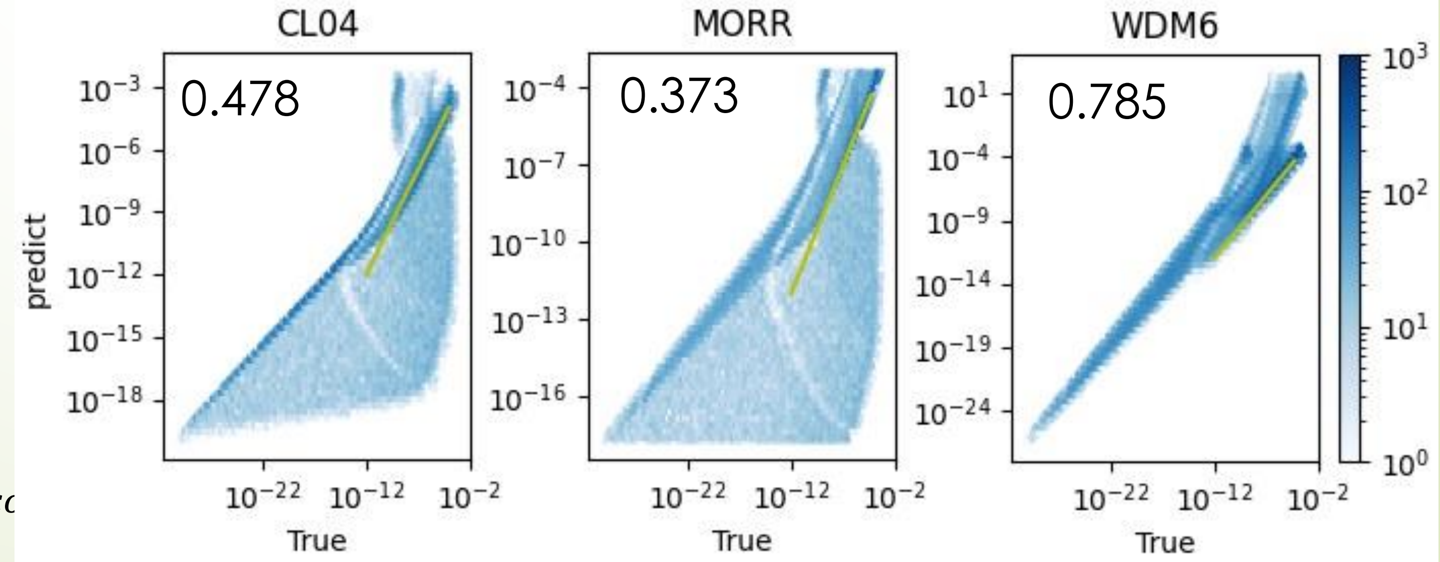


縱/橫座標為參與兩粒子之半徑，等值線顏色為E_{co}

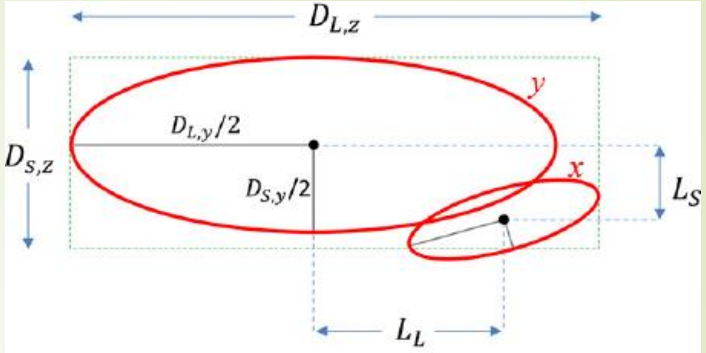
E_{coal}=1



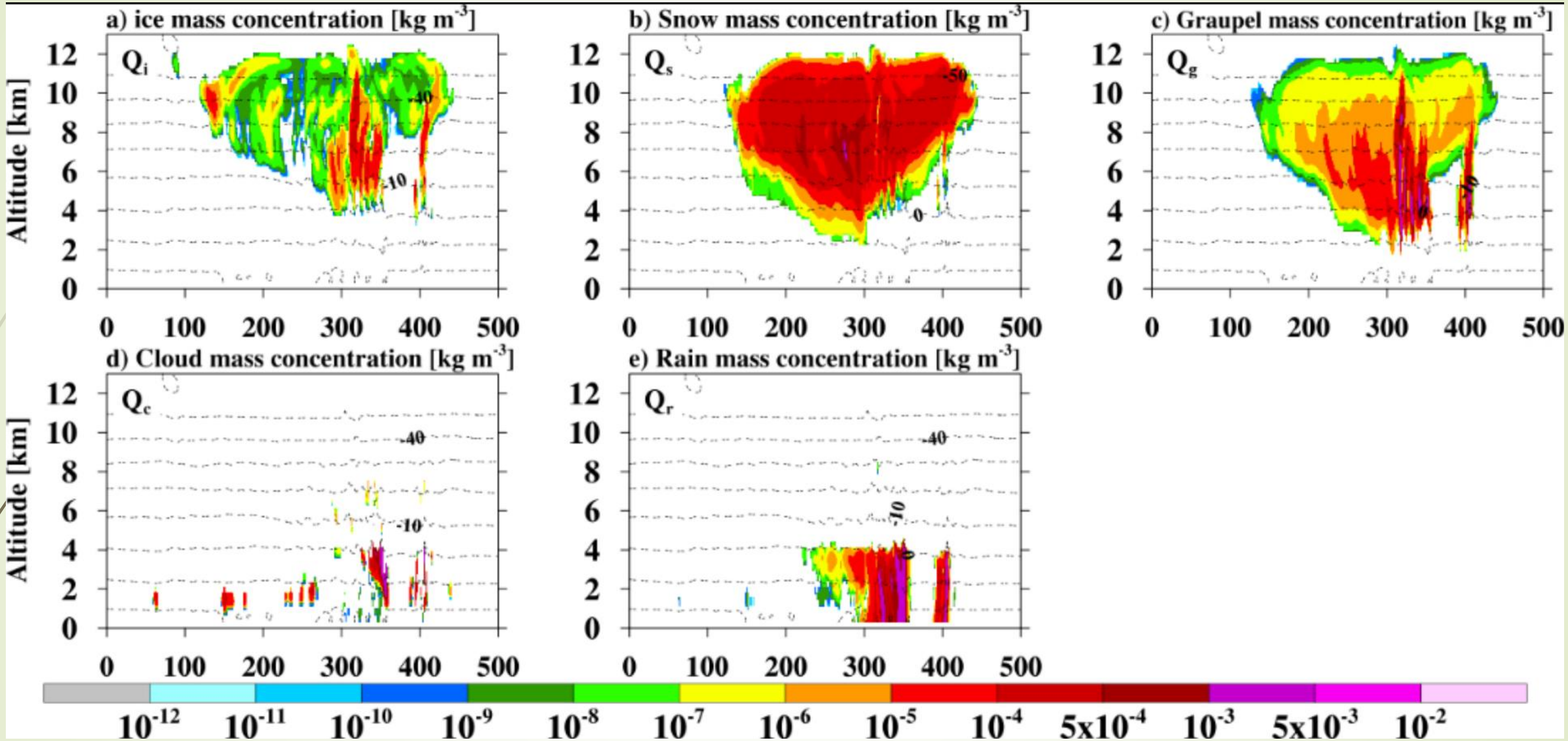
E_{coal}≠1



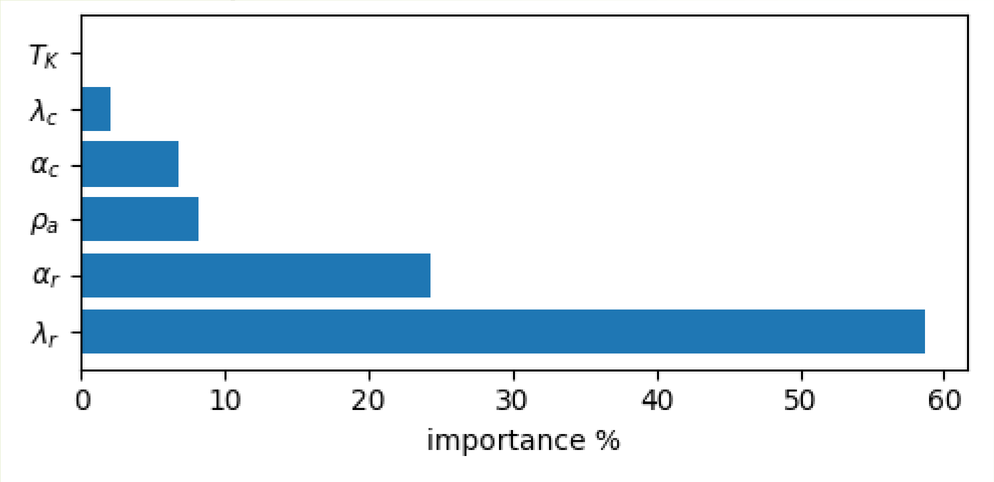
冰相水物合併係數



- 冰相粒子之間需要考慮合併後的形狀變化。這裡參考了Tsai and Chen (2020)對柱狀冰晶和盤狀冰晶之間合併後形狀變化的結果進行計算
- 冰相粒子和液相粒子的合併是假設液相粒子會結冰最後結合，所以$E_{coal} = 1$
- 修改了粒子的密度設定，另外設定了一個參數q來指定粒子為冰相或液相，當粒子為液相時，q必為1，且密度為997kg/m^3，粒子的縱橫比α使用Brandes et al. (2002)所提出的參數式計算；當粒子為冰相，其$0.05 < q \leq 1$，密度為q ·910kg/m^3，以及$0.05 < \alpha < 40$。



a) ice mass concentration [kg m-3]
b) Snow mass concentration [kg m-3]
c) Graupel mass concentration [kg m-3]
d) Cloud mass concentration [kg m-3]
e) Rain mass concentration [kg m-3]
Qi
Qs
Qg
Qc
Qr
Altitude [km]
0 2 4 6 8 10 12
0 100 200 300 400 500
-40
-10
-50
0
10-12 10-11 10-10 10-9 10-8 10-7 10-6 10-5 10-4 5x10-4 10-3 5x10-3 10-2



參數	α	$\lambda(1/m)$		$\rho_a(\mathrm{kg/m^3})$	**0.2-1.3**
雨滴	0-5	5e+3-5e+5		ΔT (K)	-20-+20
雲滴	0-20	1e+5-5e+6			