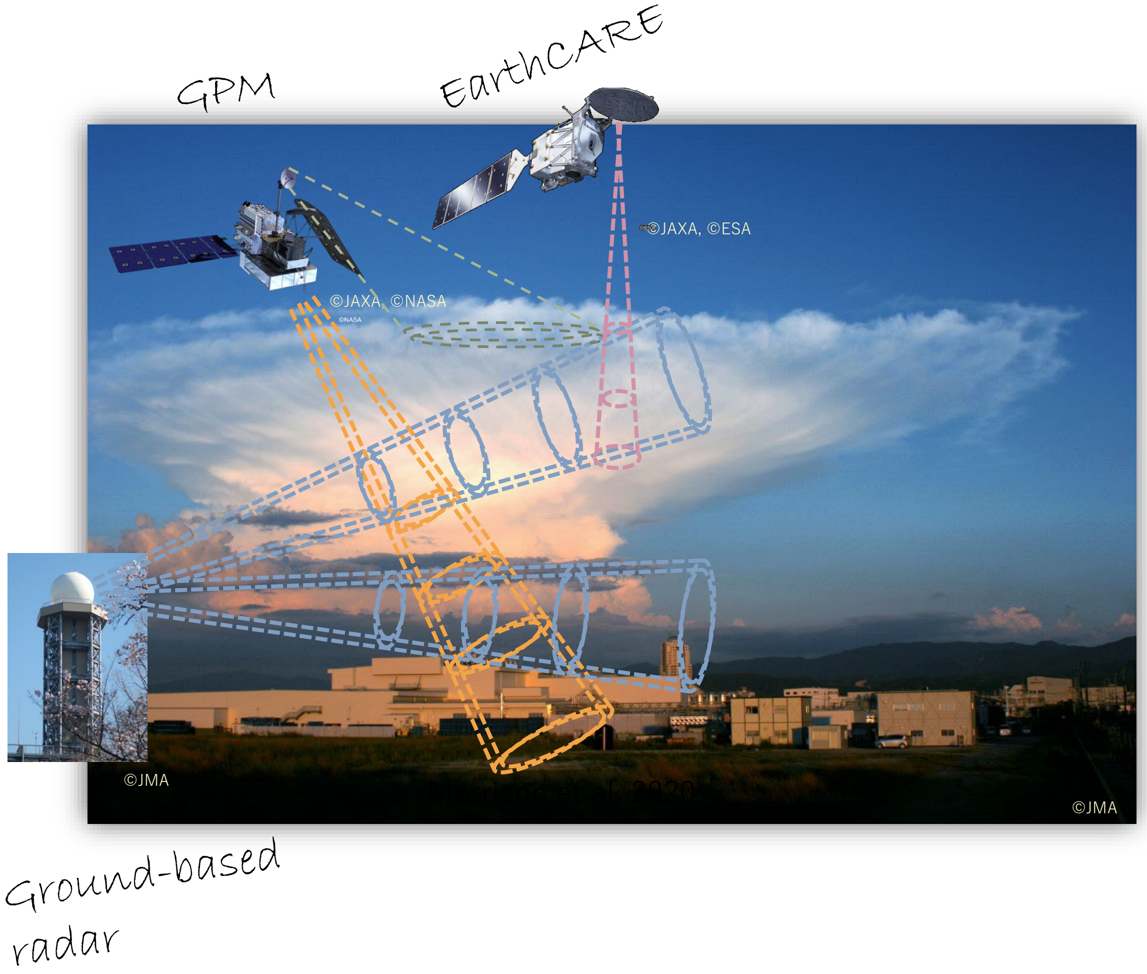
Generating background error covariances for hydrometeors with conditional generative adversarial networks

Yasutaka Ikuta

Meteorological Research Institute,
Japan Meteorological Agency

9th International Symposium on Data Assimilation, 20 Oct. 2023, Bologna, Italy

Introduction



- For improvement of radar assimilation, hydrometeors are used as control variables.
- However, there are some issues with creating a background error covariance matrix.
- Reasons for this include:
  - Hydrometeors are highly flow dependent.
  - Cloud microphysics has a strong non-linearity.
- In this study, we have created the flow-dependent background error covariance using deep learning.

Operational regional DA system at JMA

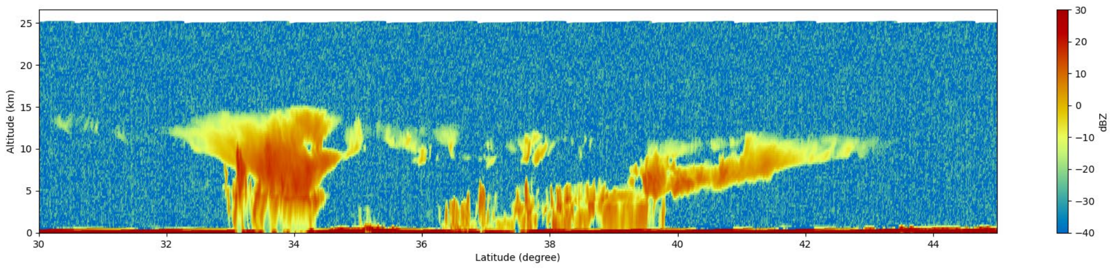
	Meso-scale Analysis	Local Analysis
Method	4DVar	Hybrid-3DVar
Horizontal grid spacing	DA: 5 km (outer), 15 km (inner) Forecast: 5 km	DA: 5 km Forecast: 2 km
Control variables	U: x-direction wind speed, V: y-direction wind speed, PT: potential temperature, Ps: surface pressure, Tg: soil temperature, μ: pseudo humidity, and Wg: soil volumetric water content	
Climatological background error	Surface type: sea / land Time: 00, 03, 06, 09, 12, 15, 18 21 UTC	
Ensemble background error	None	100 members (20 members with 5 different initial times)

Hydrometeors are not control variables

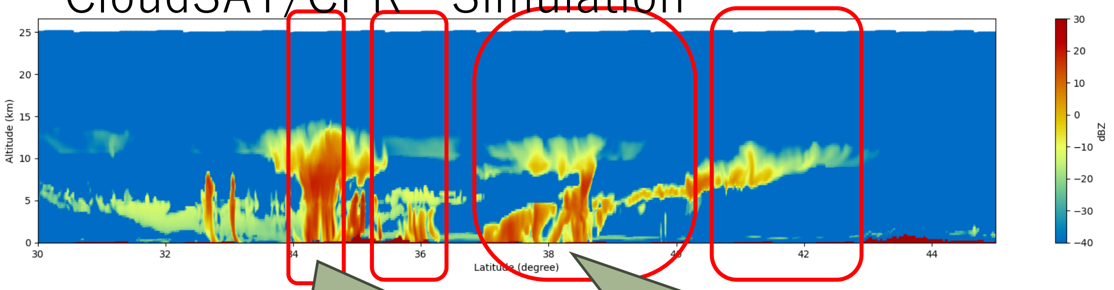
Ikuta et al. (2021), Yokota et al. (in revision)

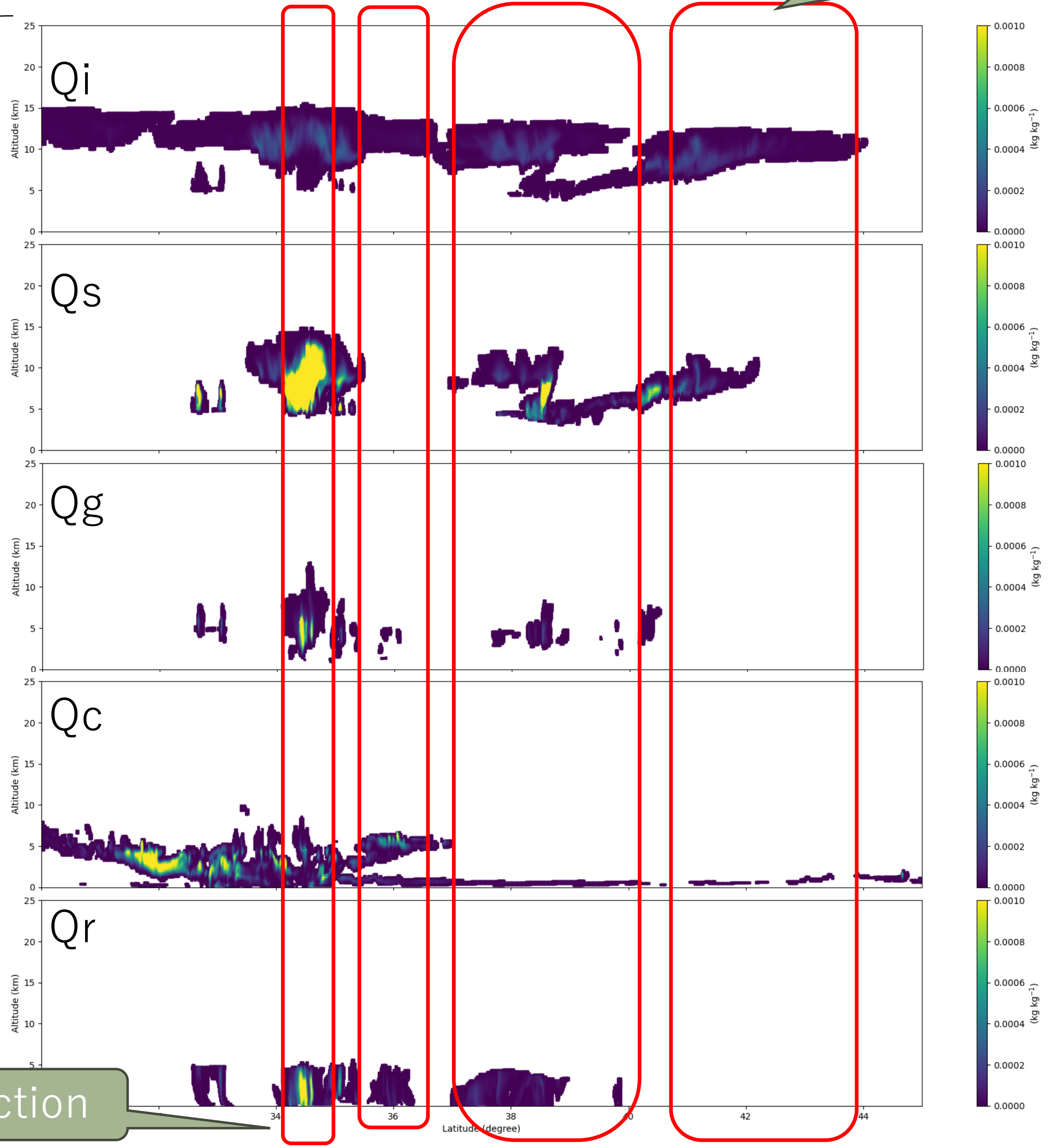
Vertical distribution of hydrometeors

CloudSAT/CPR　Observation



CloudSAT/CPR　Simulation

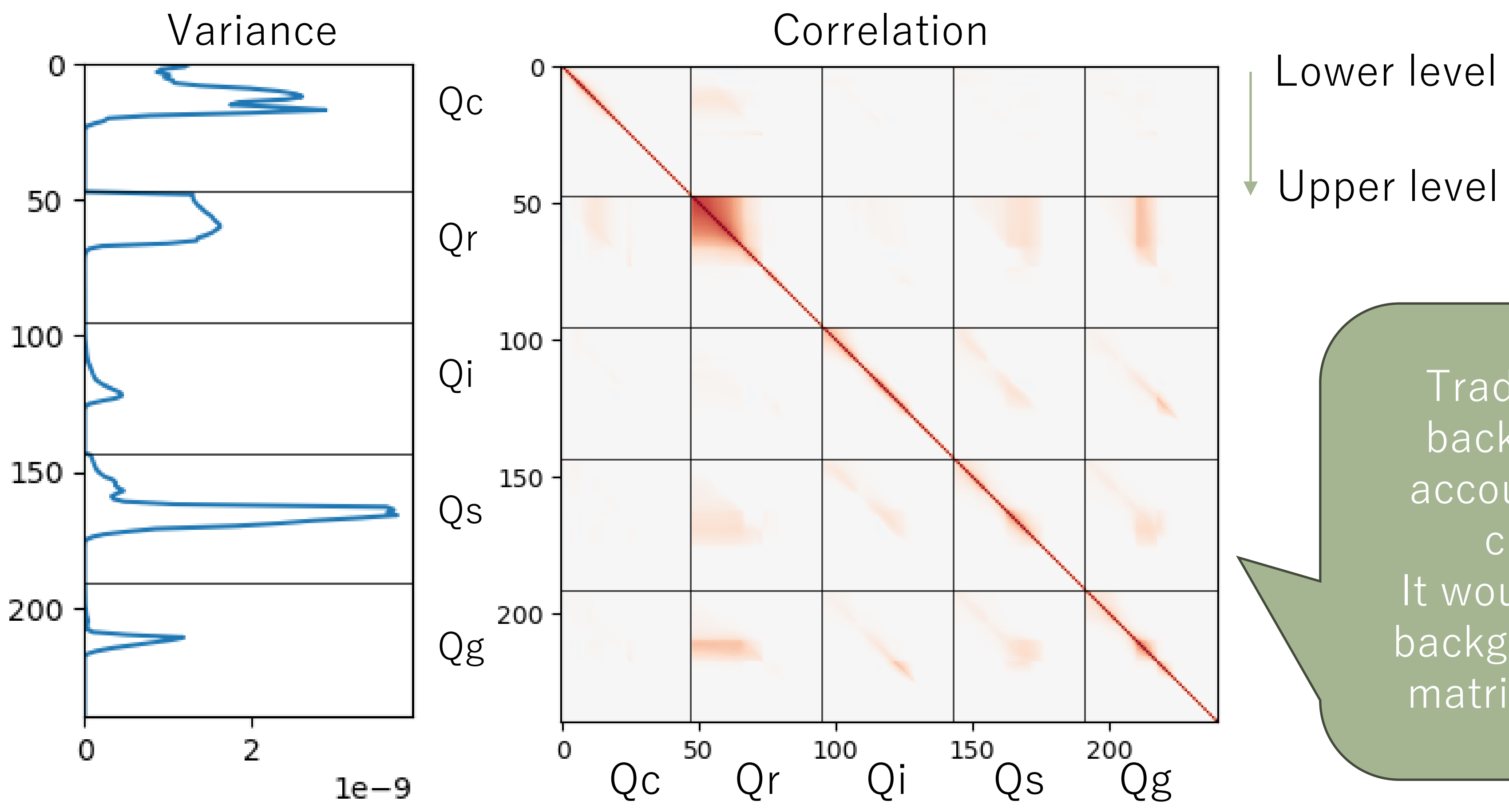




There are various patterns in the vertical distribution of hydrometeors.
Deep convection includes clouds, snow, graupel, and rain. On the other hand, there are grids with only clouds.

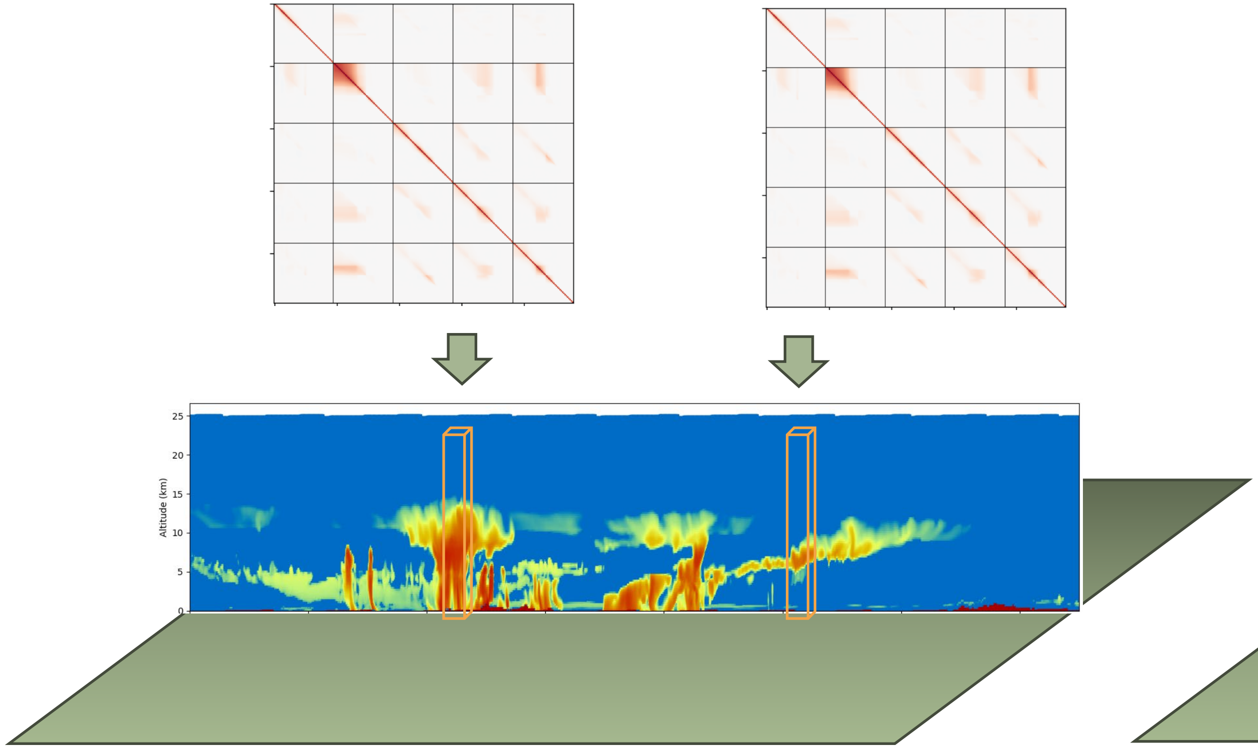
Climatological background error covariance of hydrometeors

Summer/winter average using 100-member Ensemble DA (EDA) around Japan

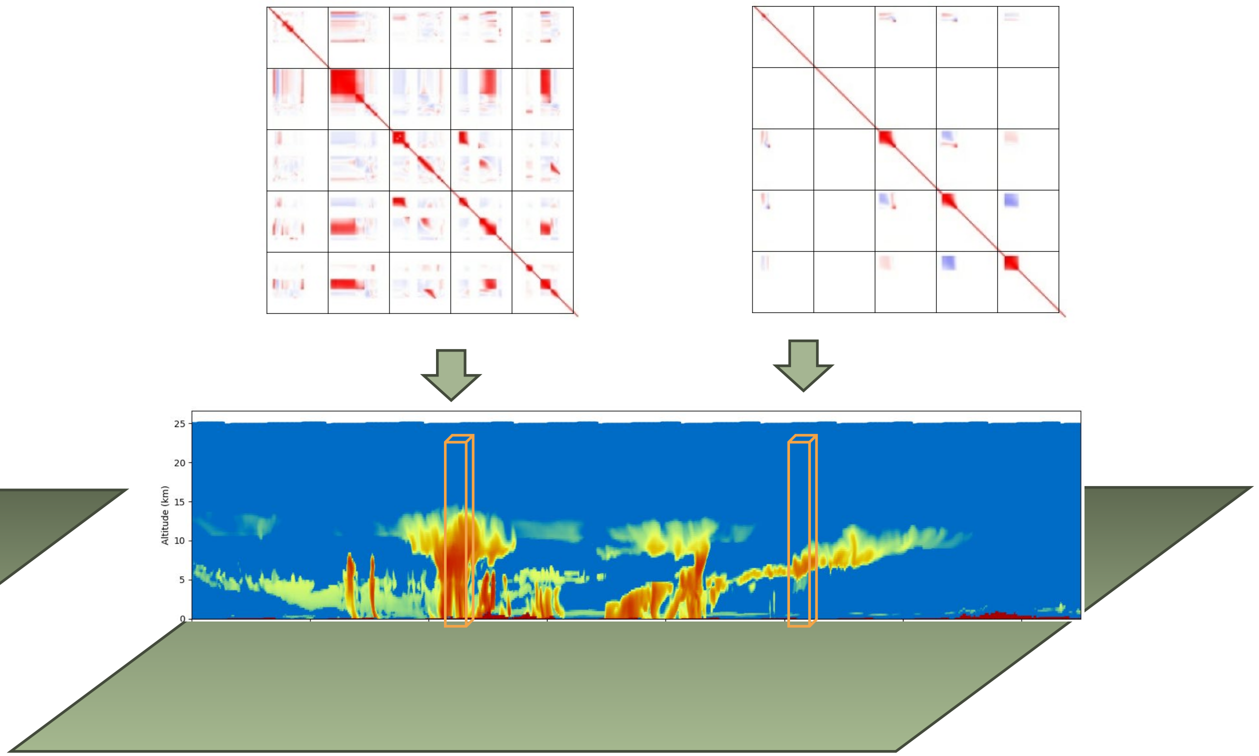


Traditional climatological background errors do not account for the diversity of cloud precipitation.
It would be useful to have a background error covariance matrix for various patterns.

Comparison with conventional or flow-dependent type



In conventional methods, the same BG is used everywhere.



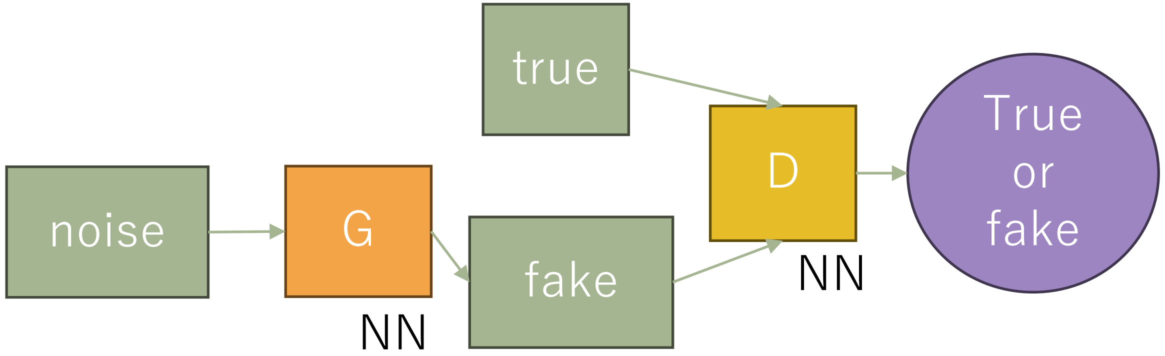
In the flow-dependent method, different BGs are used depending on the location.

New background error estimation

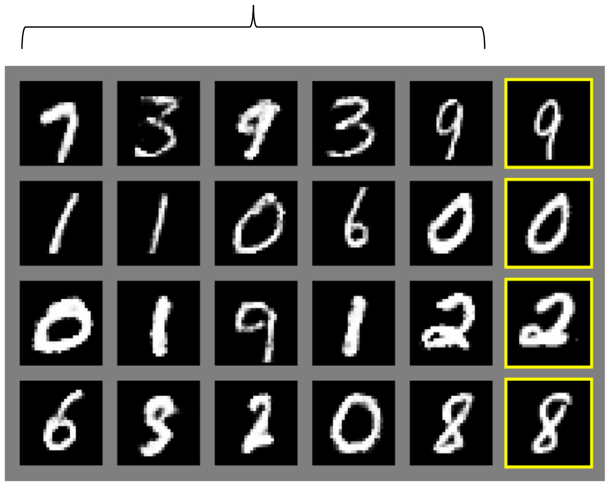
- Climatological background error
  - e.g. NMC method
  - Disadvantage: Cannot have the flow dependent.
- Ensemble background error
  - Created from EDA or ensemble forecast
  - Disadvantages: A huge number of members will be needed. Localization needs to be considered.
- The minimum number of members required is unknown due to the diversity of hydrometeors.
- If the background error of hydrometeors can be estimated using **deep learning**, the computational cost can be significantly reduced.

Method

- We estimate the background error covariance with generative DL.
- One of the popular classical genitive DL methods is Generative Adversarial Nets (**GAN**; Ian J. Goodfellow et al. 2014)



Generated images



Ian J. Goodfellow et al. (2014)

It is also called an "adversarial" generative network because it uses two neural networks, the Generator (G) and Discriminator (D), to compete against each other to learn data.

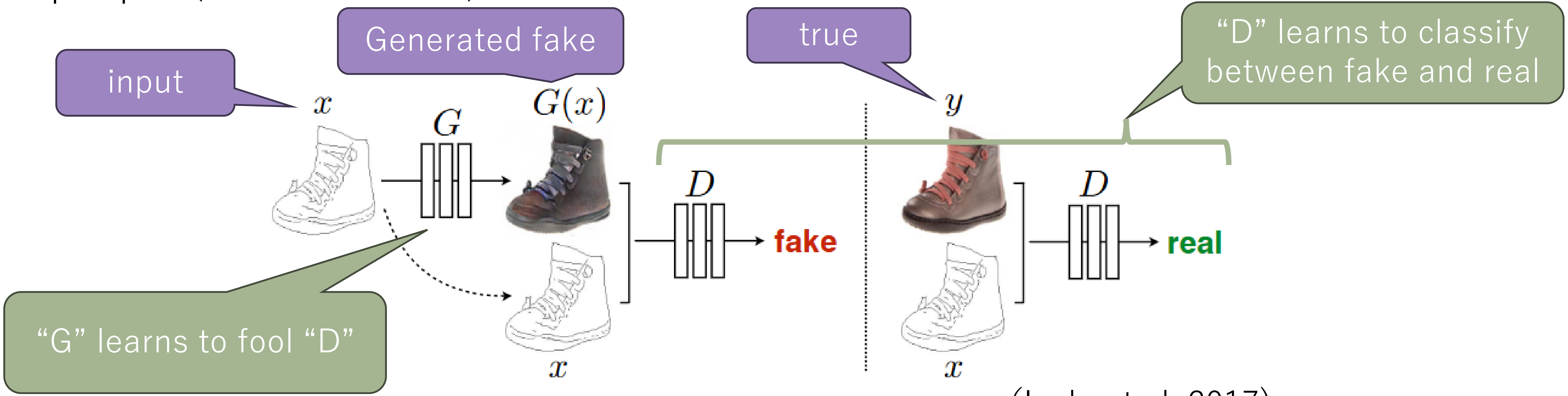
GAN can do the following:

- ✓ Generating non-existent data
- ✓ Conversion according to learned data characteristics
- ✓ Generate new data that includes features of the original data

However, GANs cannot control which images are generated.

Method

- Conditional Generative Adversarial Nets（**CGAN**; Mirza and Osindero 2014）
- pix2pix（Isola et al. 2017）



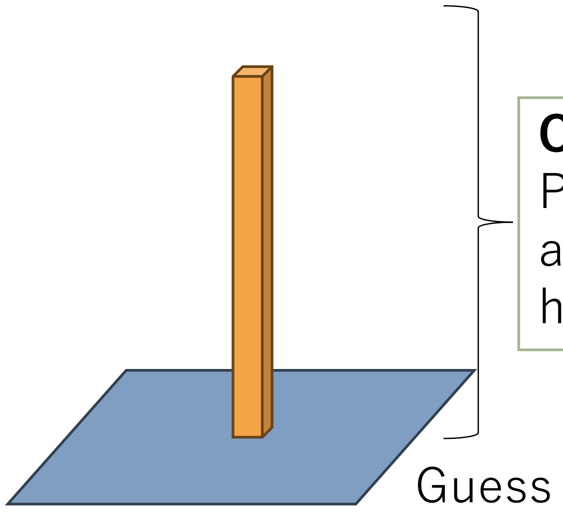
（Isola et al. 2017）

✓ CGAN is a GAN that trains conditioning by giving additional **condition information** to the Generator (G) and Discriminator (D).

✓ CGAN can **learn to accept** only the correct combination of real data and labels and **learn to reject** all other data. Unlike regular GAN, CGAN is used to generate images according to specified conditions.

Making of a tuple for training data

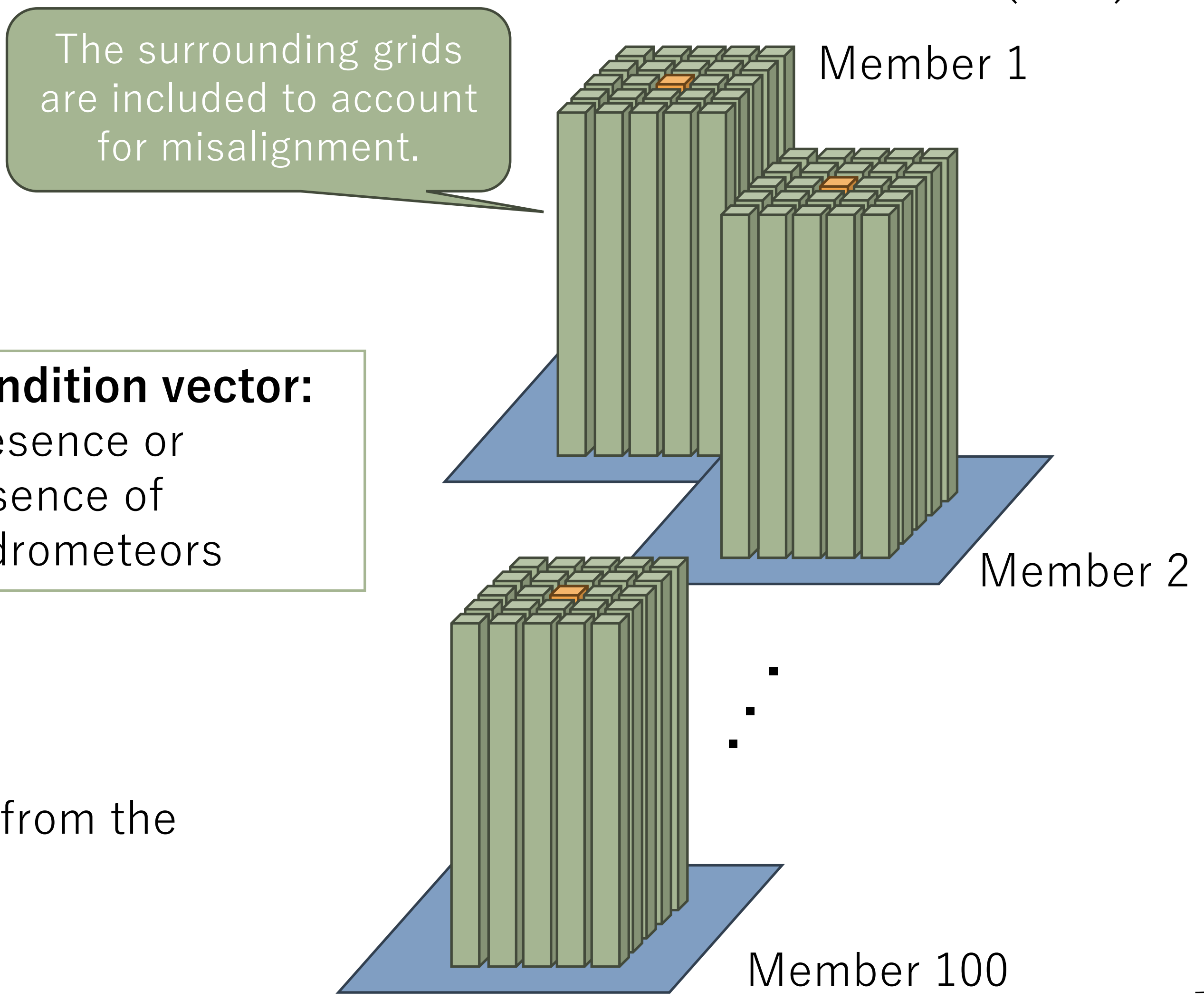
Deterministic DA



Condition vector:
Presence or absence of hydrometeors

Condition vector is created from the profile in guess,

Ensemble Data Assimilation (EDA)

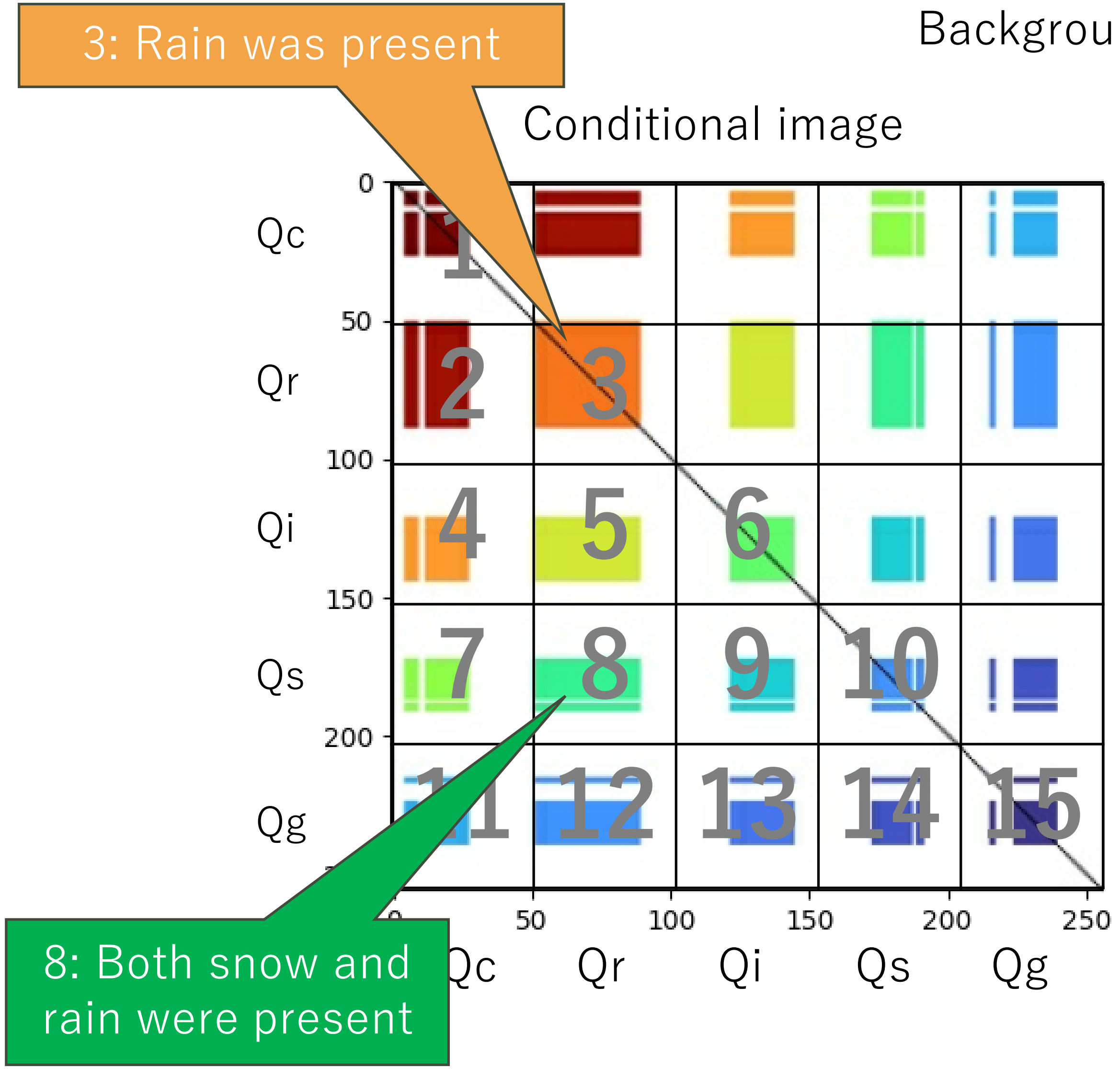


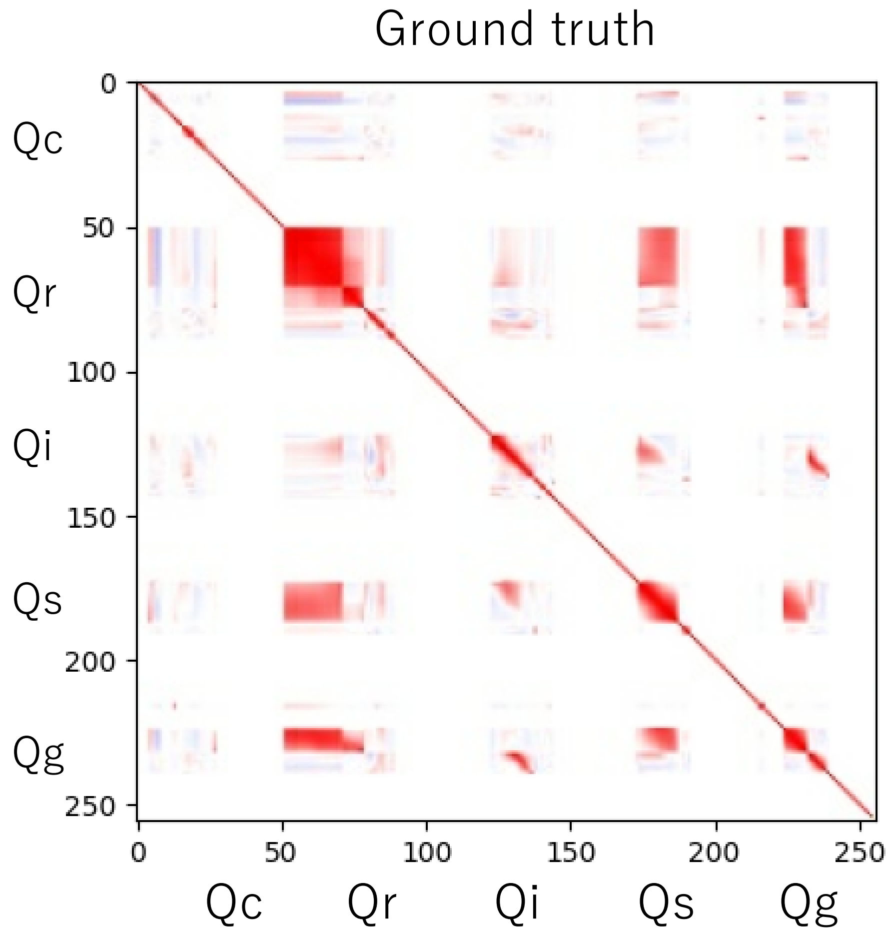
Ground truth:
Background error covariance of hydrometeors

$$\mathbf{B} = \langle (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T \rangle$$

Conditional image and ground truth image

Background error correlation

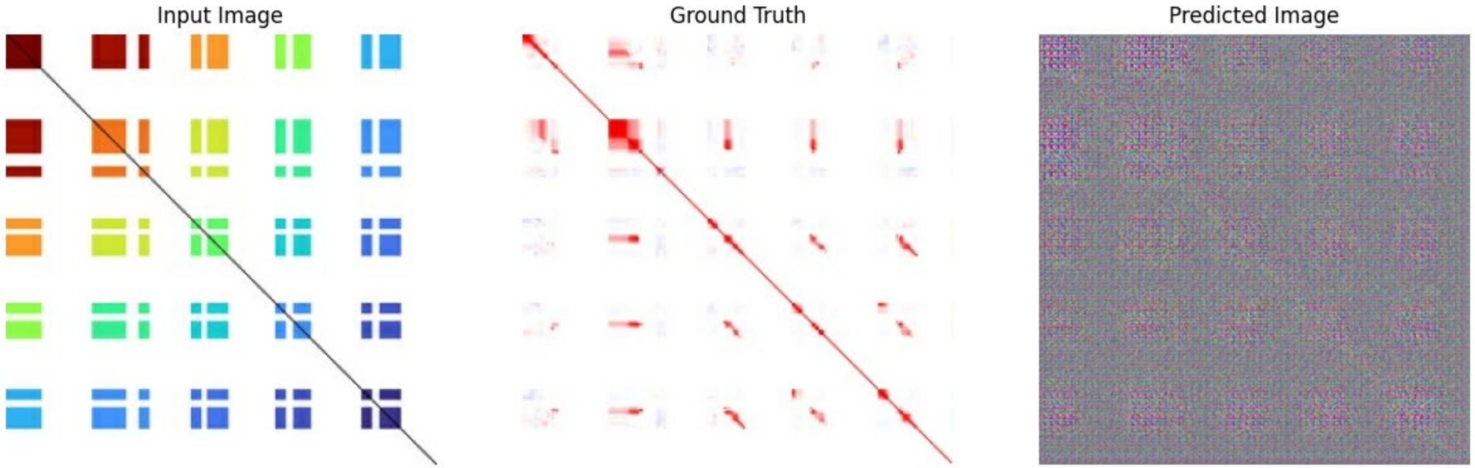




Result of learning

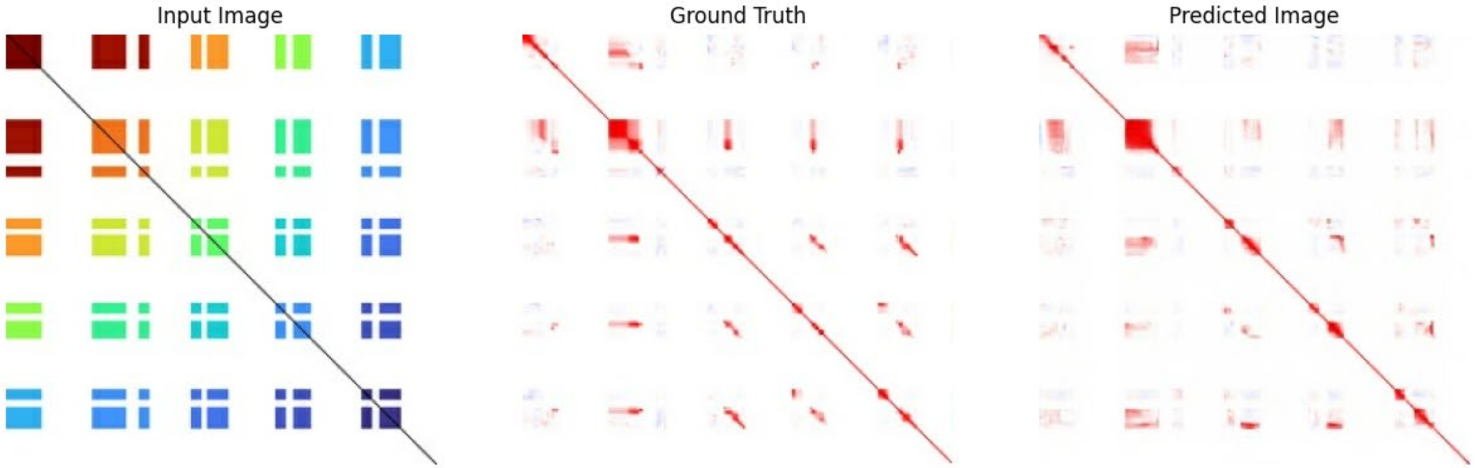
- TensorFlow
- Dataset
  - Training: 1,670
  - Validation: 50
  - Test: 50

First epoch



The initial image was white noise.

Epoch 40,000

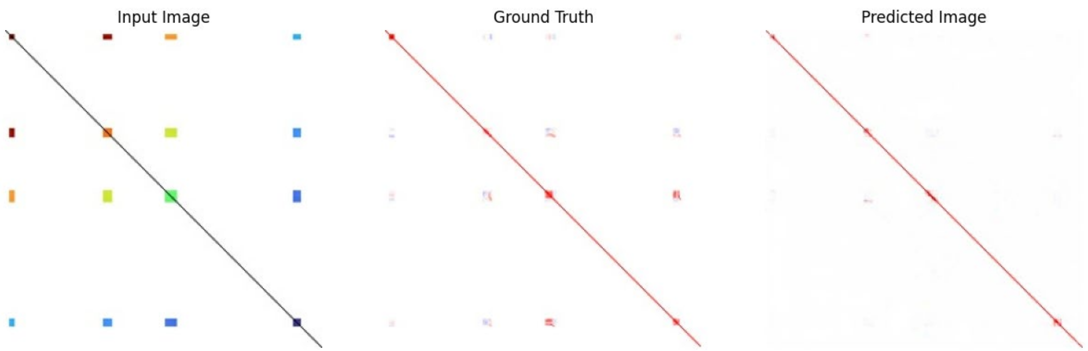


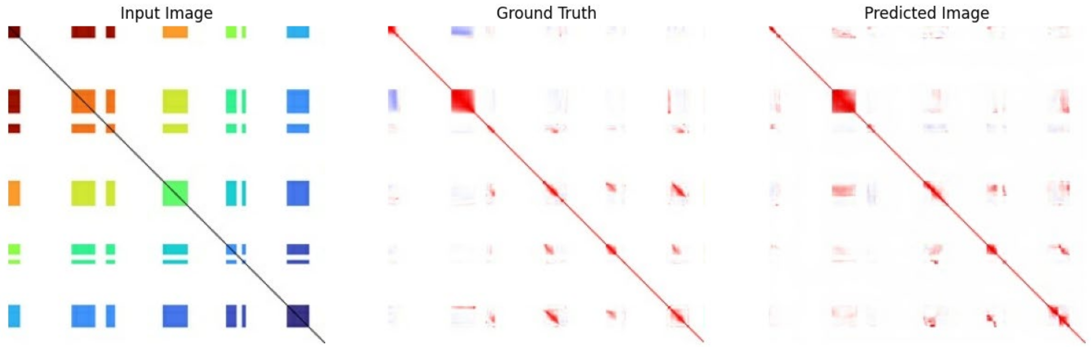
Successfully generated an image close to the ground truth

Test of learning result

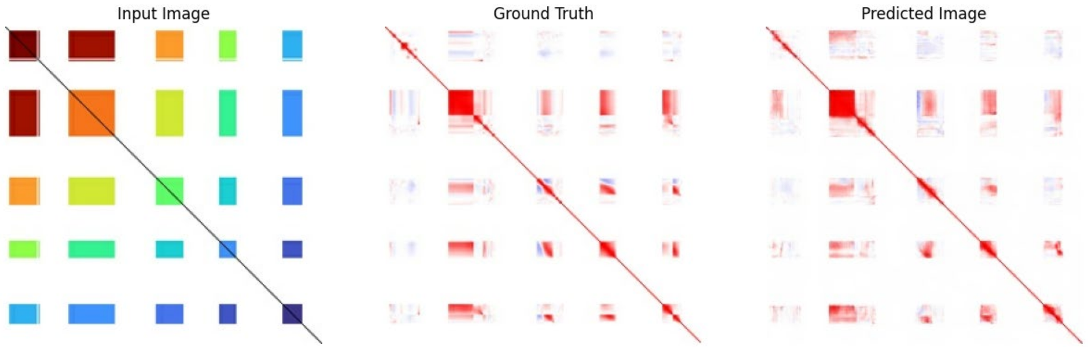
In all cases, the predicted structure was generally close to the true value.

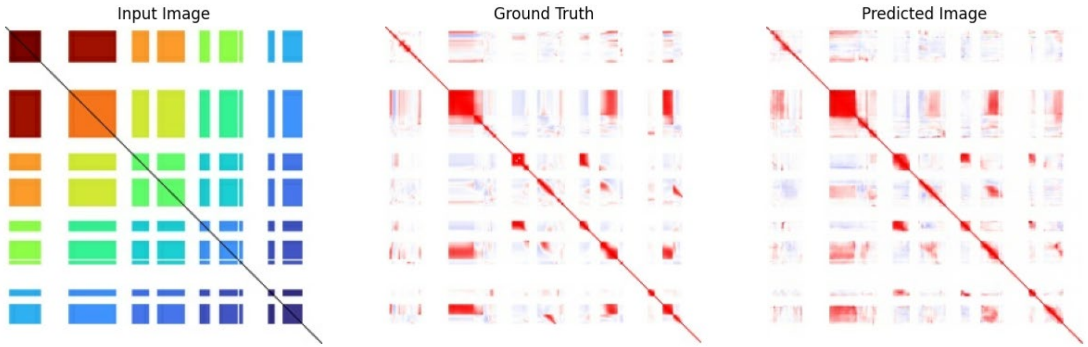
Not many hydrometeors are present.





Many hydrometeors are present.



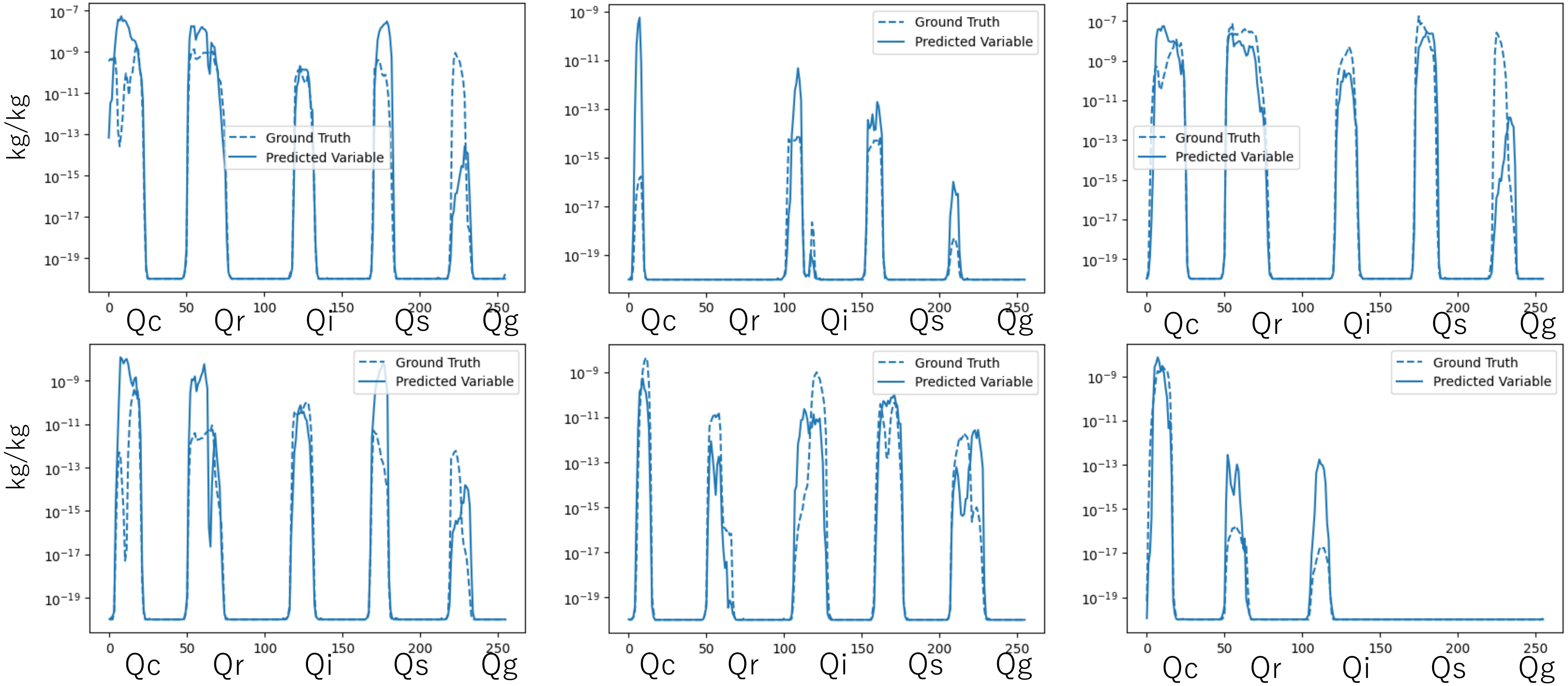


These results can be generated instantly from saved Checkpoints.

Test of learning result for variance

The structure was generated.

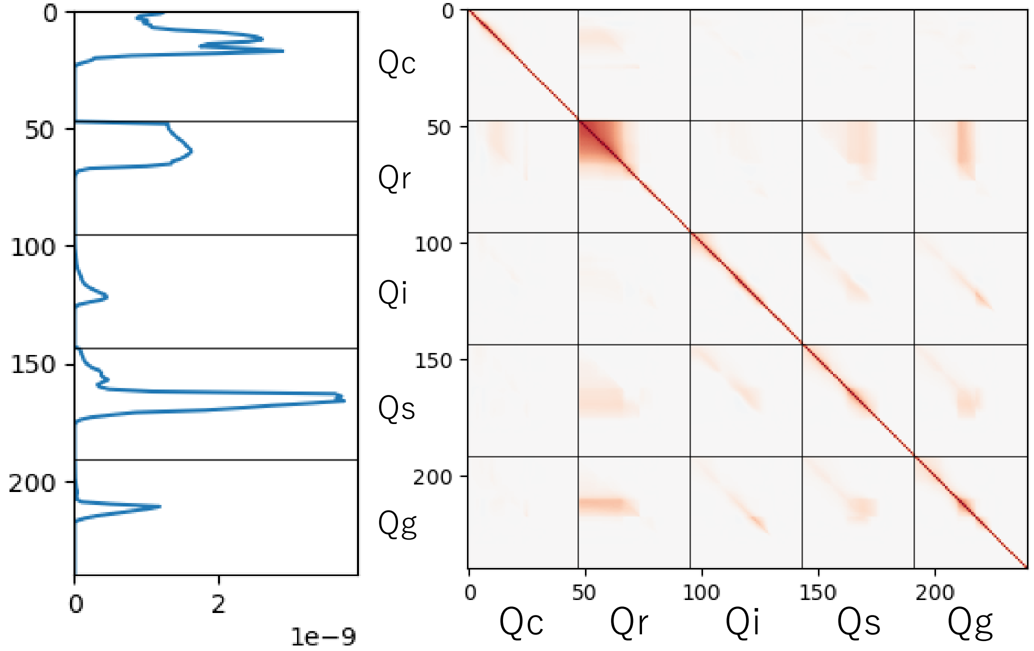
Variances are generated by CGAN using the same way of error correlation.



These results can be generated instantly from saved Checkpoints.

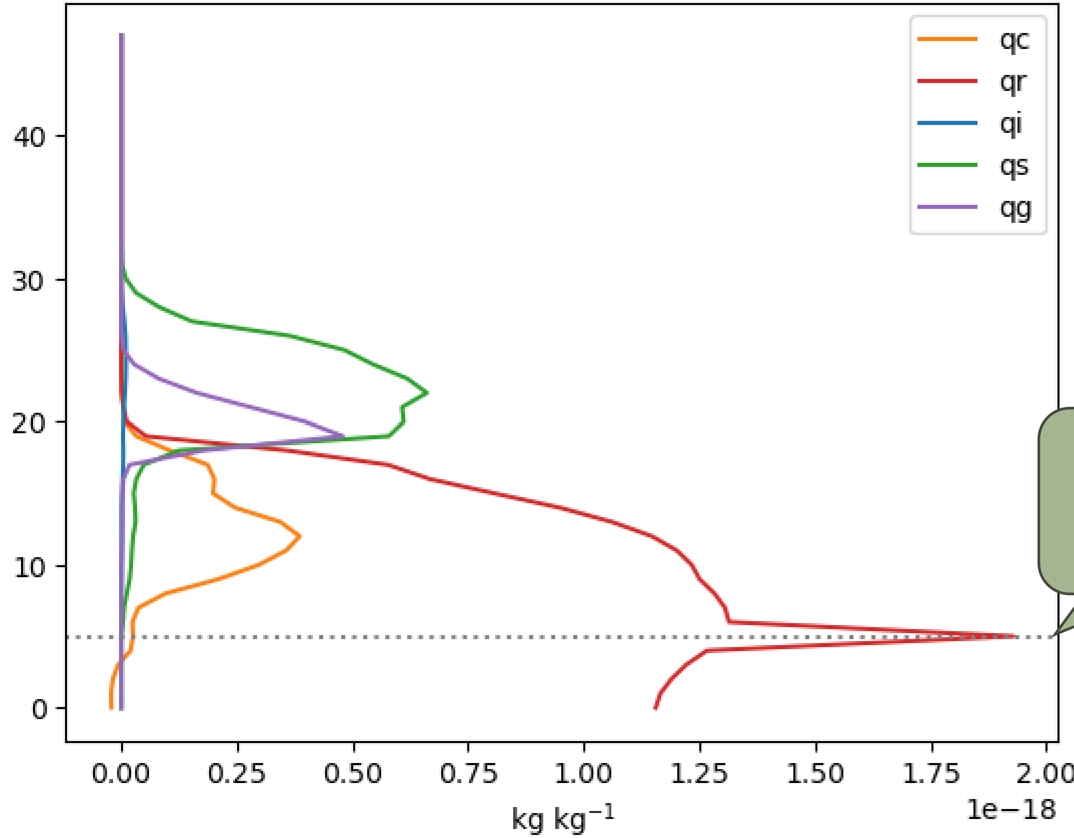
Vertical profile of analysis increment

Climatological BG error



In conventional methods, this background error is used anytime and anywhere.

<u>Idealized test of a single observation</u>

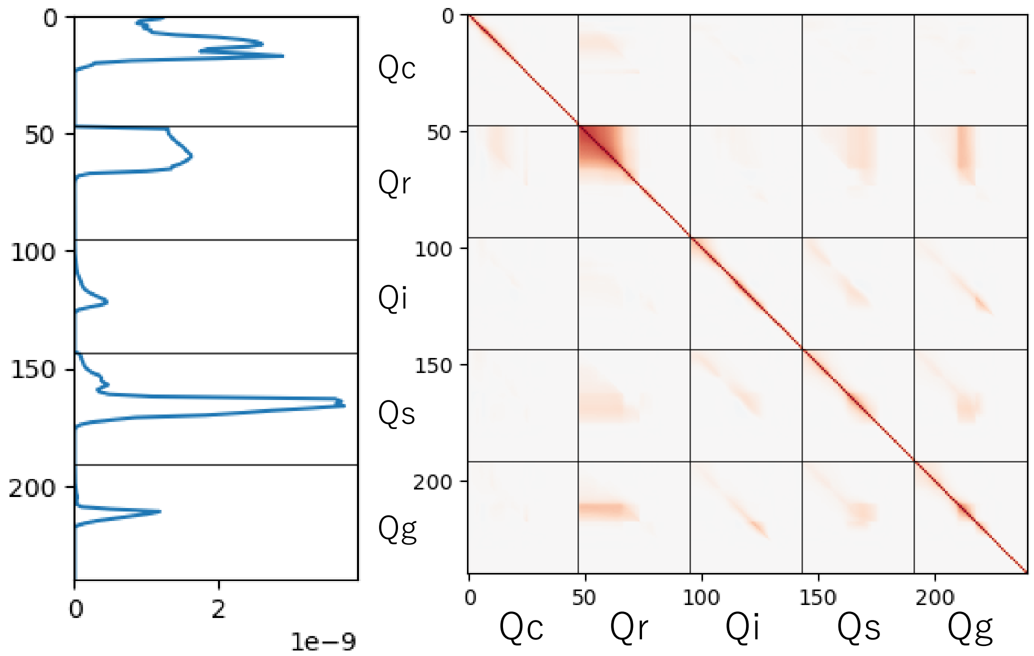


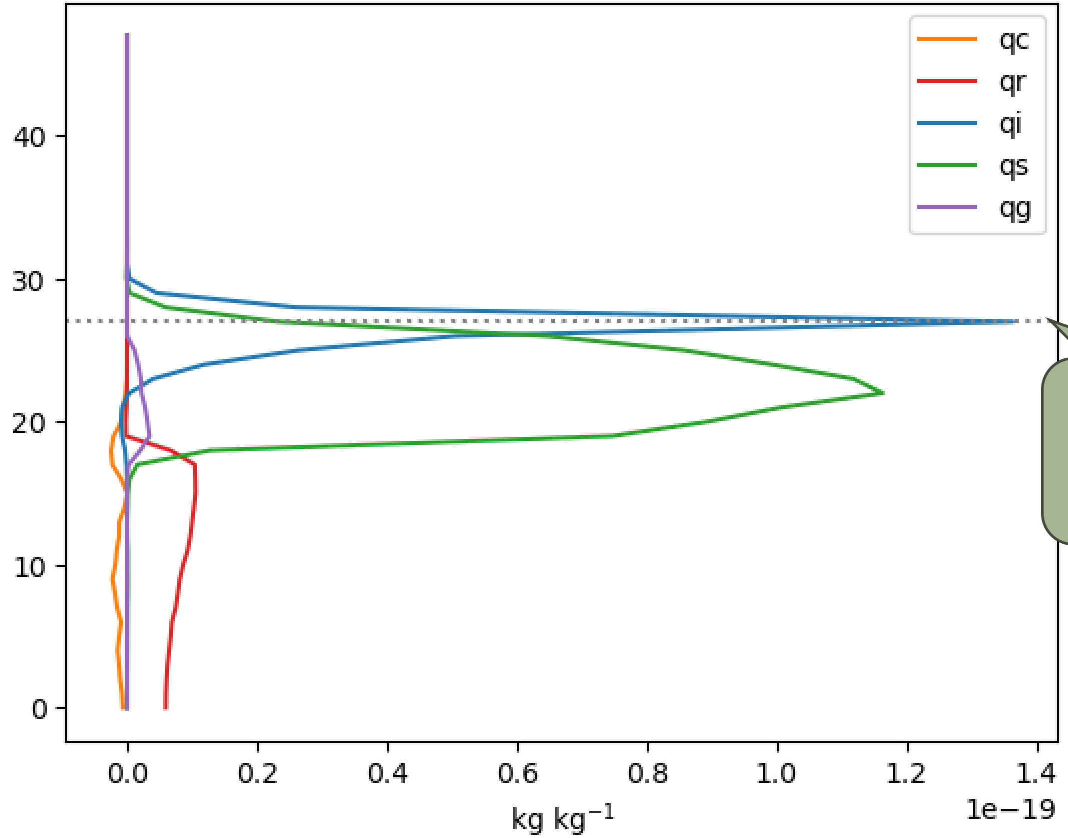
$$\frac{\delta Q_r^o}{\sigma_o} = 1$$

Vertical profile of analysis increment

Idealized test of a single observation

Climatological BG error



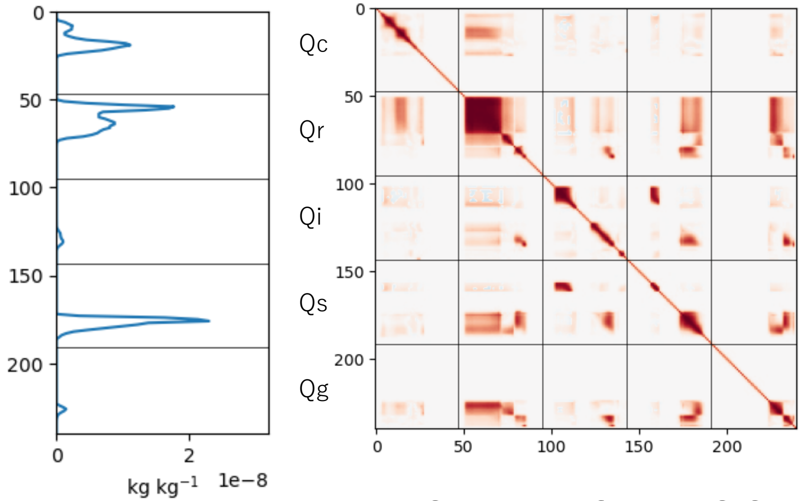


$$\frac{\delta Q_i^o}{\sigma_o} = 1$$

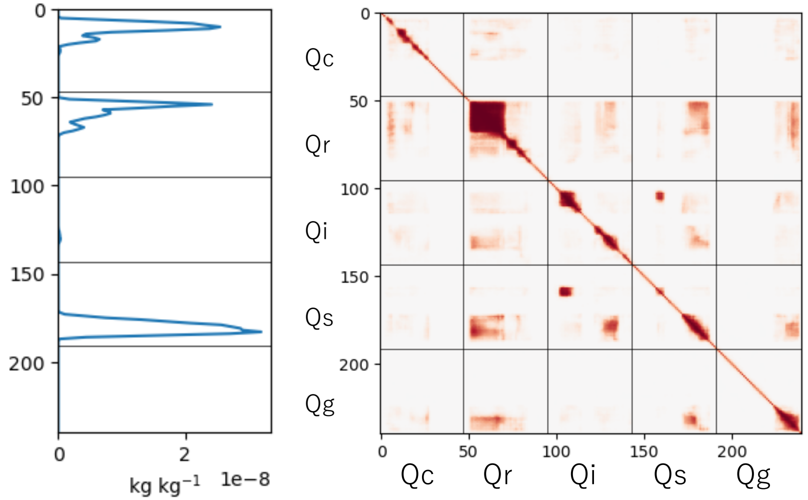
In conventional methods, this background error is used anytime and anywhere.

Vertical profile of analysis increment

BG Error from EDA

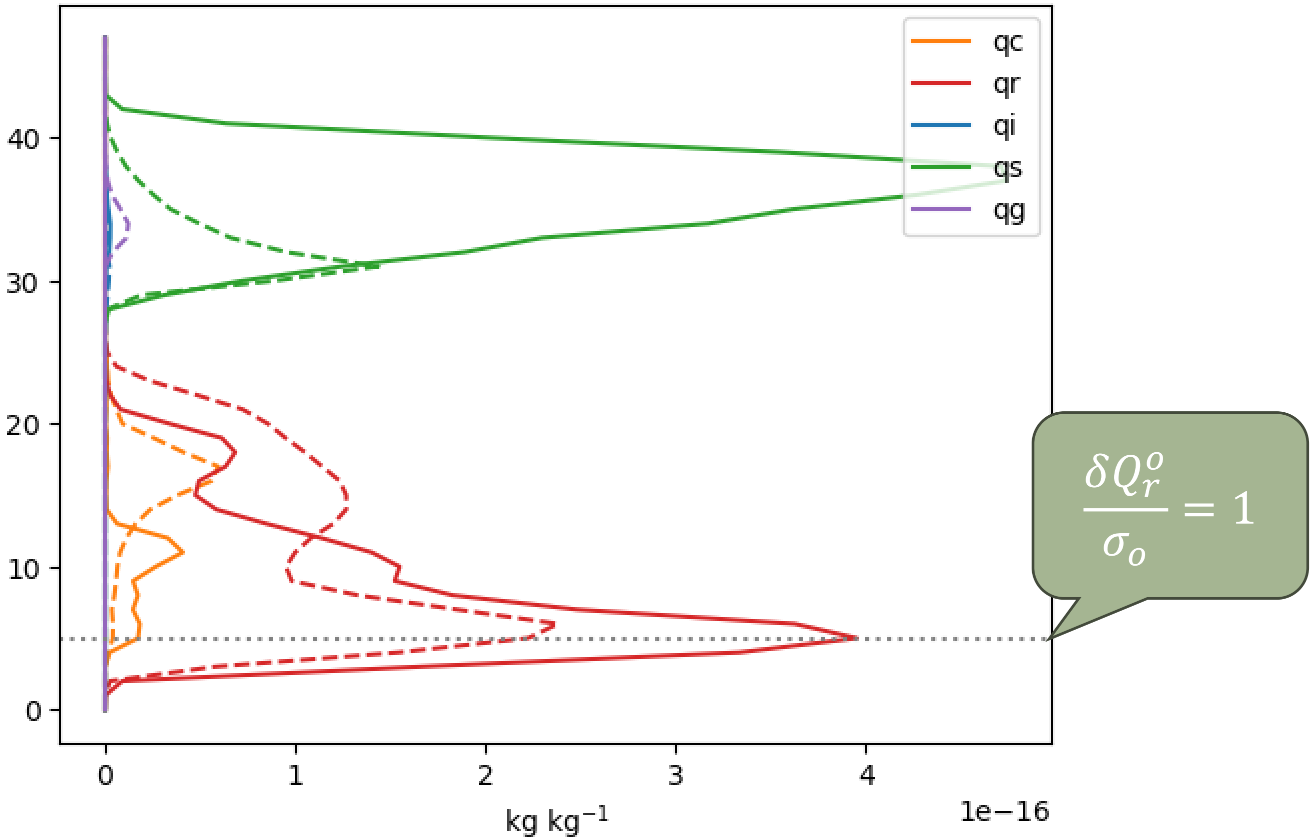


BG Error from CGAN



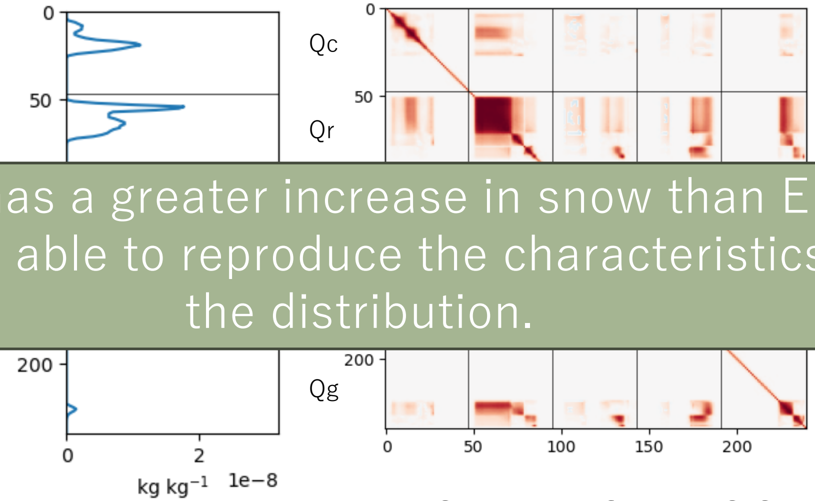
Idealized test of a single observation

Dashed: EDA
Solid: CGAN



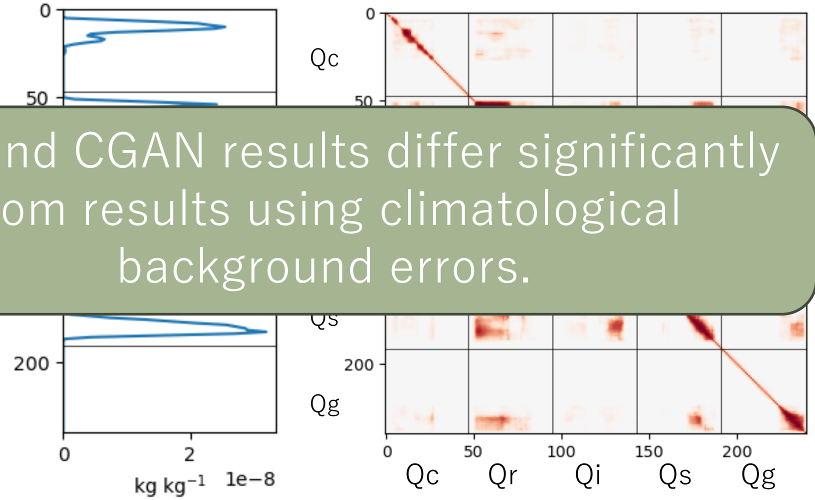
Vertical profile of analysis increment

BG Error from EDA



CGAN has a greater increase in snow than EDA, but it is able to reproduce the characteristics of the distribution.

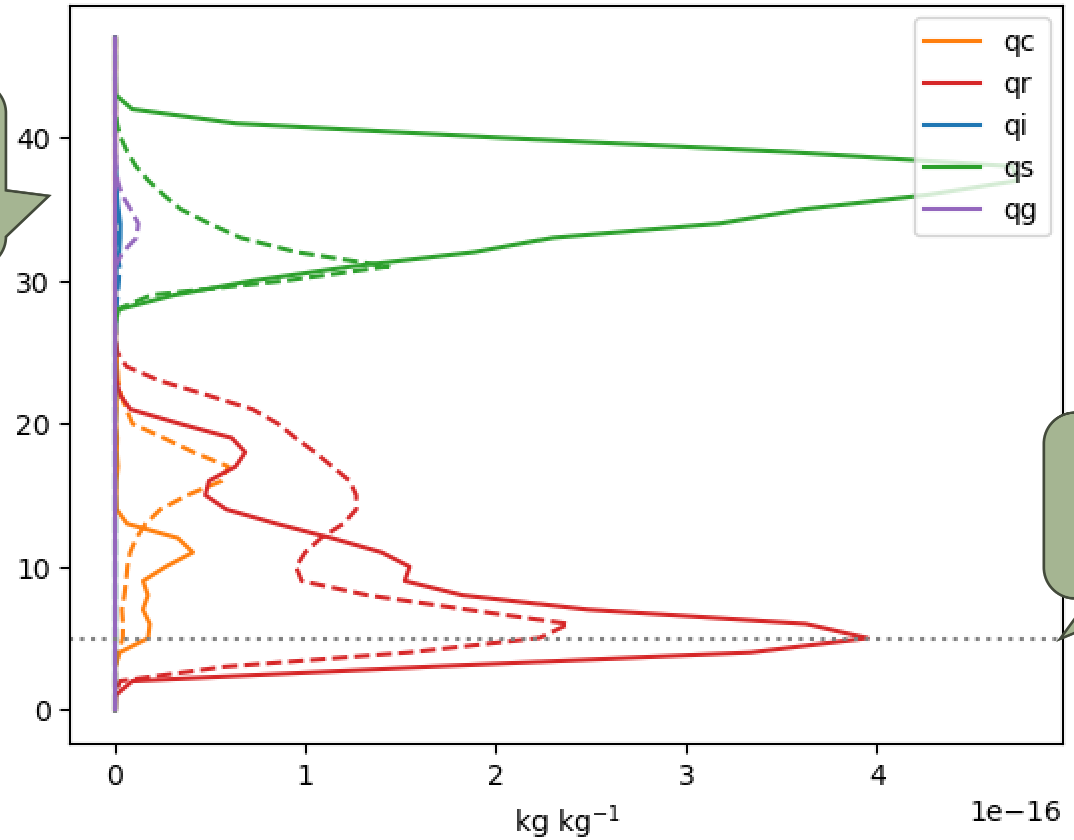
BG Error from CGAN



EDA and CGAN results differ significantly from results using climatological background errors.

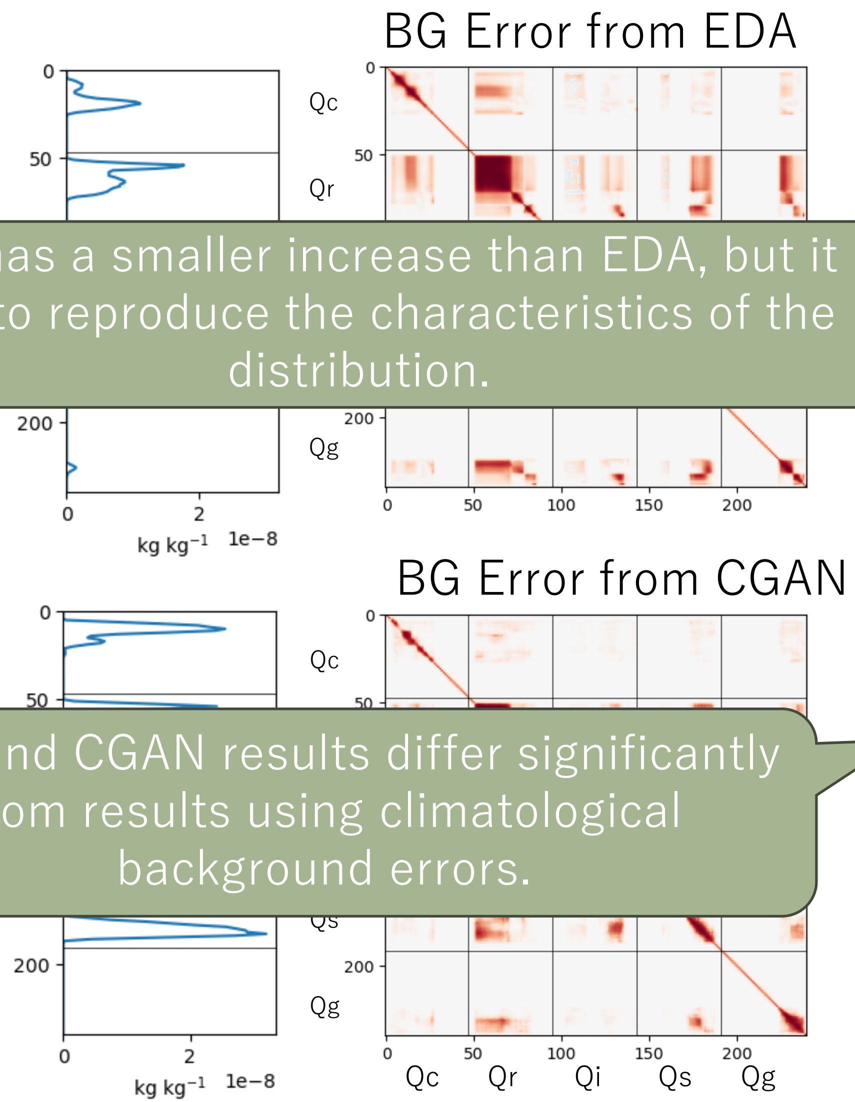
Idealized test of a single observation

Dashed: EDA
Solid: CGAN



$$\frac{\delta Q_r^o}{\sigma_o} = 1$$

Vertical profile of analysis increment

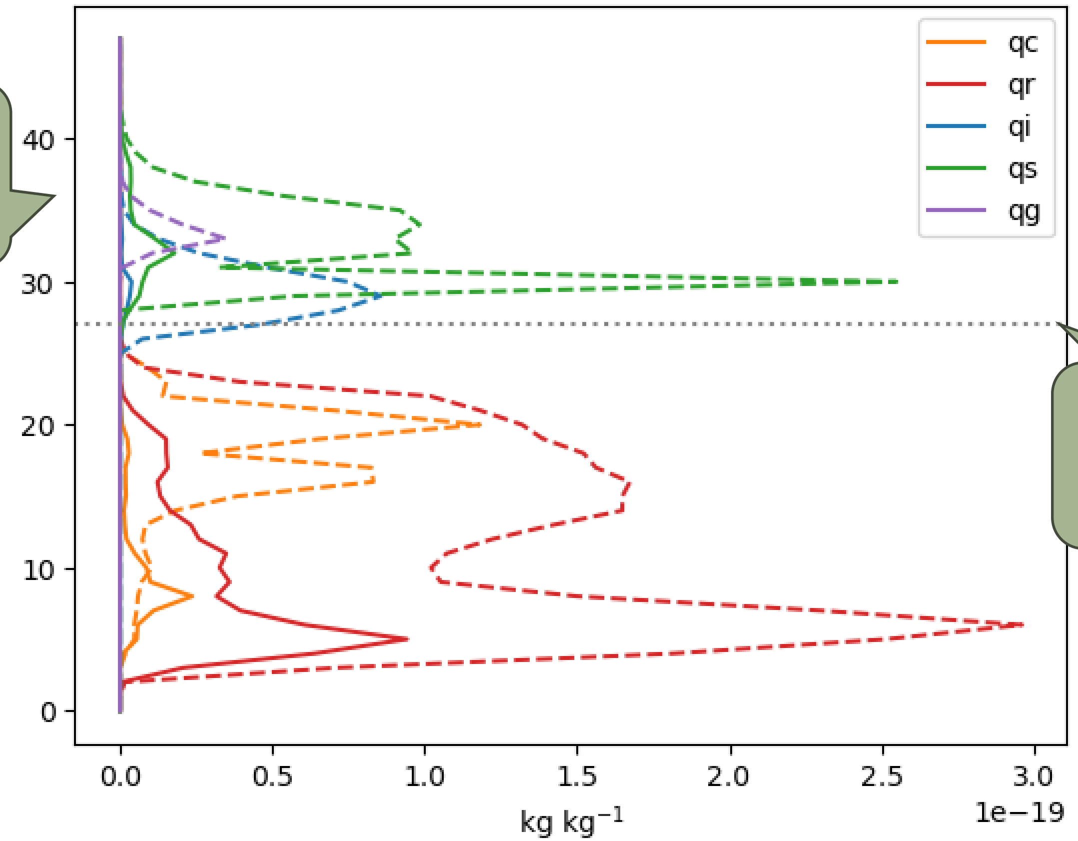


CGAN has a smaller increase than EDA, but it is able to reproduce the characteristics of the distribution.

EDA and CGAN results differ significantly from results using climatological background errors.

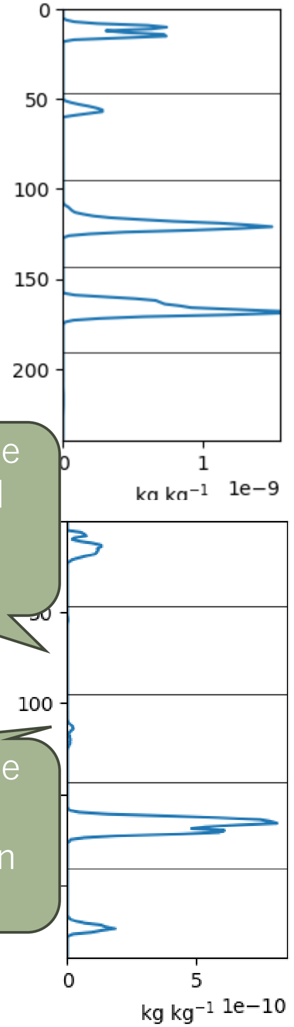
Idealized test of a single observation

Dashed: EDA
Solid: CGAN



$$\frac{\delta Q_i^o}{\sigma_o} = 1$$

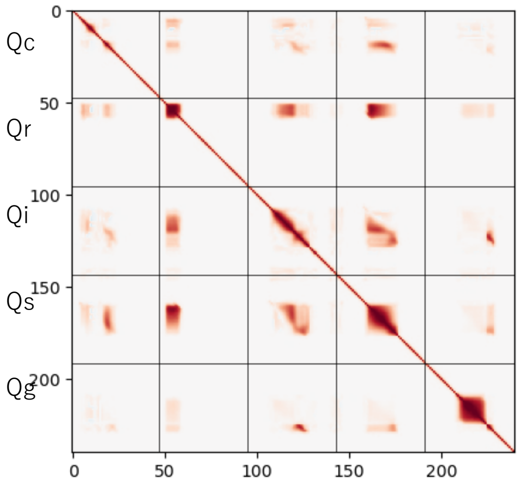
Vertical profile of analysis increment



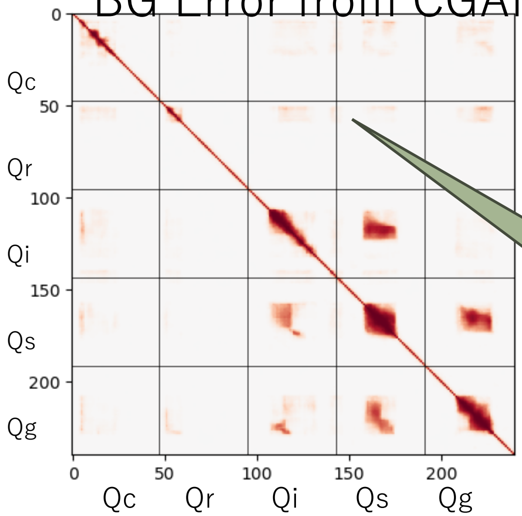
The variance of Qr could not be generated

The variance of Qi is smaller than EDA

BG Error from EDA



BG Error from CGAN

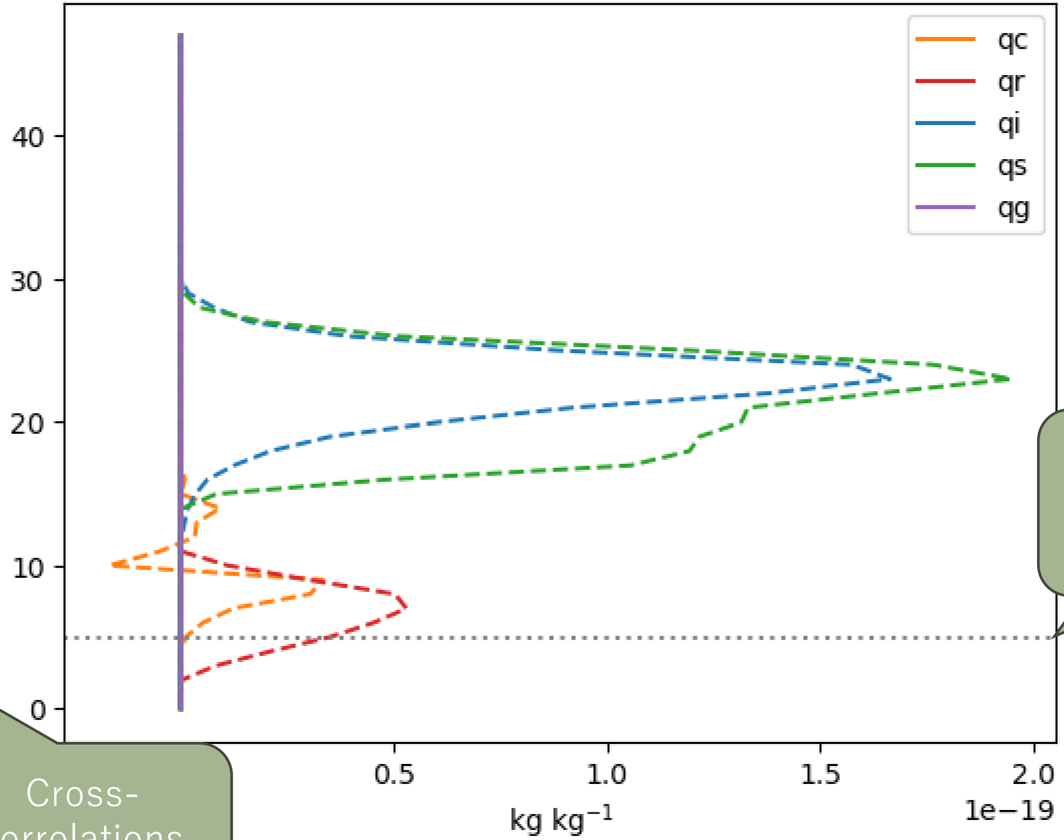


Cross-correlations are small

<u>Idealized test of a single observation</u>

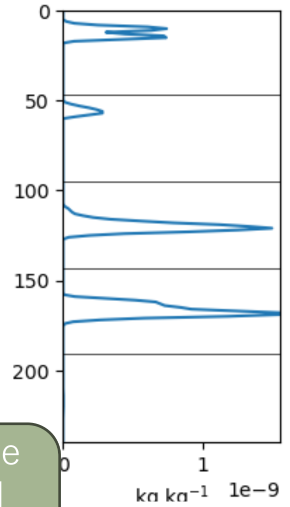
Failure Case

Dashed: EDA
Solid: CGAN

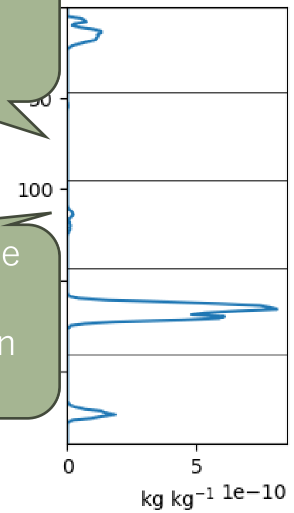


$$\frac{\delta Q_r^o}{\sigma_o} = 1$$

Vertical profile of analysis increment

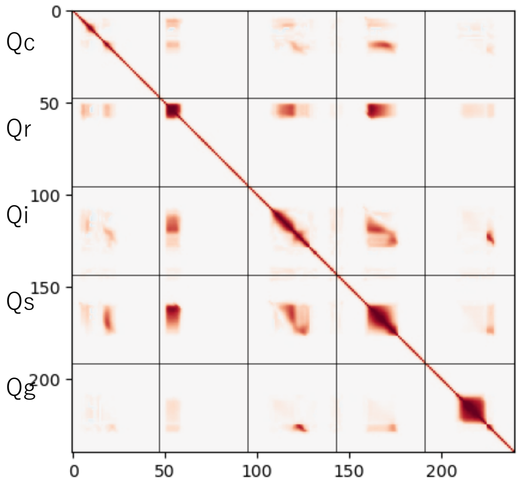


The variance of Qr could not be generated

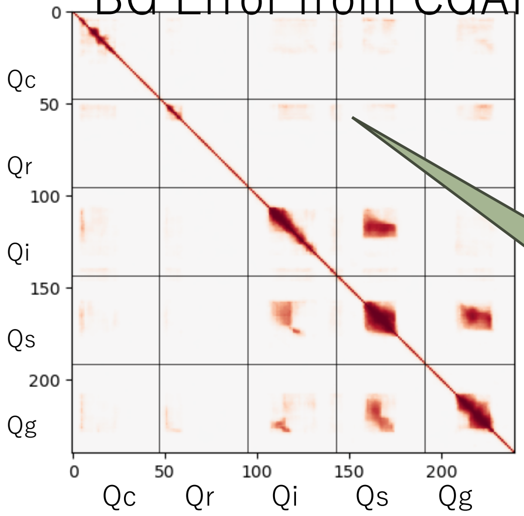


The variance of Qi is smaller than EDA

BG Error from EDA



BG Error from CGAN

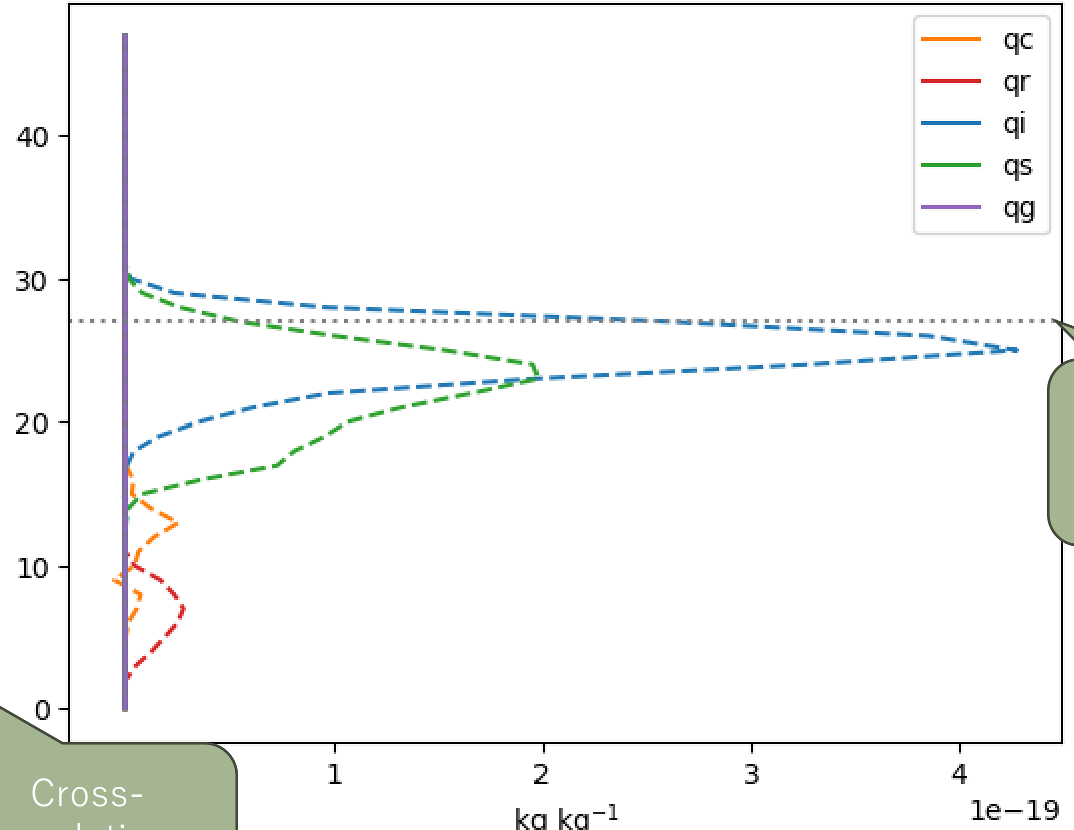


Cross-correlations are small

Idealized test of a single observation

Failure Case

Dashed: EDA
Solid: CGAN



$$\frac{\delta Q_i^o}{\sigma_o} = 1$$

Conclusion

- We tried to generate a background error correlation matrix using deep learning.
- The CGAN was used as the deep learning method.
- As a result of giving the presence or absence of hydrometeors as a condition vector, it was found that a matrix close to the ground truth could be generated.
- These results suggest obtaining flow-dependent hydrometeors background errors using deep learning without preparing ensemble predictions is possible.
- However, there are still issues that need to be resolved for practical application.
- Issues:
    - It is not a positive definite symmetric matrix. -> Adjust with post-processing
    - The generated variance has a large error. -> Enhance training of DL
    - Unexpected unbalanced and unphysical covariance matrix may be produced. -> Replace with climatological BG error

Future plans

- The background error correlation between vertical velocity, temperature, and hydrometeors is estimated using DL.