Assimilation of cloud profiling radar using machine-learning-generated background errors and local ensemble tangent linear models

Yasutaka Ikuta

Meteorological Research Institute / Japan Meteorological Agency

JAXA's EORA3 supported this research

International Data Assimilation Symposium, October 24, 2024, Kobe

Application of ML for flow-dependent DA

4DVar
- ✓ BG error
  - ✓ Climatological and EDA
- ✓ NL using a physics-based model
- ✓ TL/AD and LETLM/AD

4DVar with ML This presentation
- ✓ BG error
  - ✓ Climatological and ML
- ✓ NL using ML?
- ✓ TL/AD using ML?

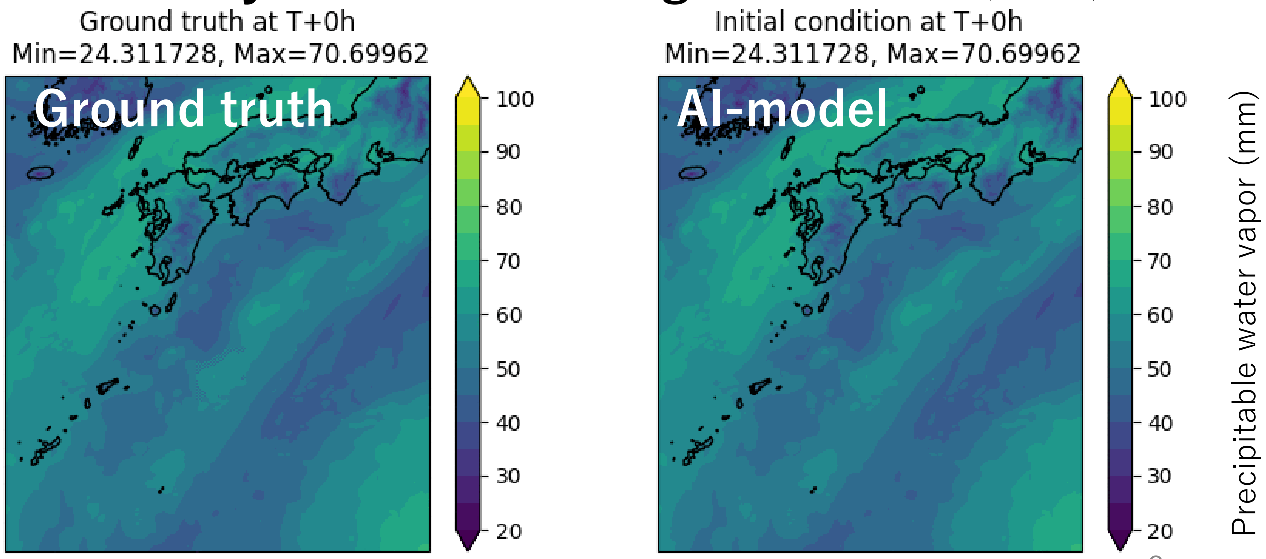
LETLM (e.g. Payne 2021)

$$\delta\mathbf{x}(t+1) = \mathbf{M}\delta\mathbf{x}(t)$$

$$\mathbf{M} = \delta\mathbf{x}(t)\delta\mathbf{X}^T(\delta\mathbf{X}\delta\mathbf{X}^T)^+$$

LETLM can easily perform TL for complex physical schemes.

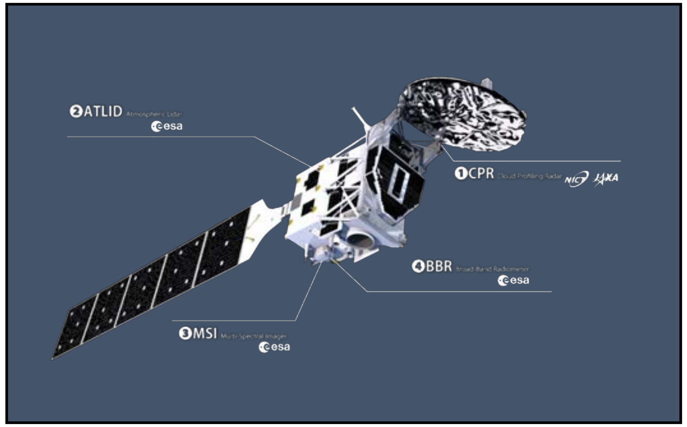
Preliminary result of GNN regional model (5km)



Init: 0000 UTC 24 Aug. 2024

Assimilation of Cloud Profiling Radar (CPR)

Launched on 29 May 2024

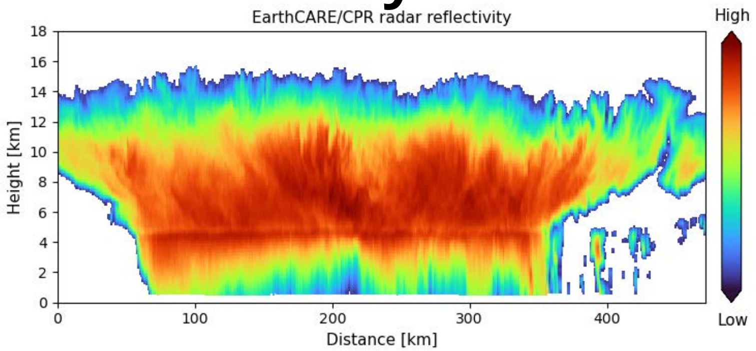


<u>EarthCARE/CPR</u>

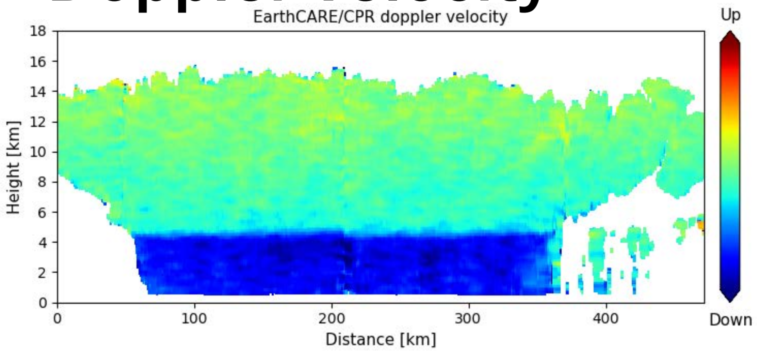
- ✓ CPR is developed by JAXA and NICT
- ✓ W-band radar
- ✓ Reflectivity and Doppler velocity
  - ✓ Information on hydrometeors

However, **CloudSAT/CPR's data** was used since EarthCARE/CPR's data has not yet been made public.

Reflectivity



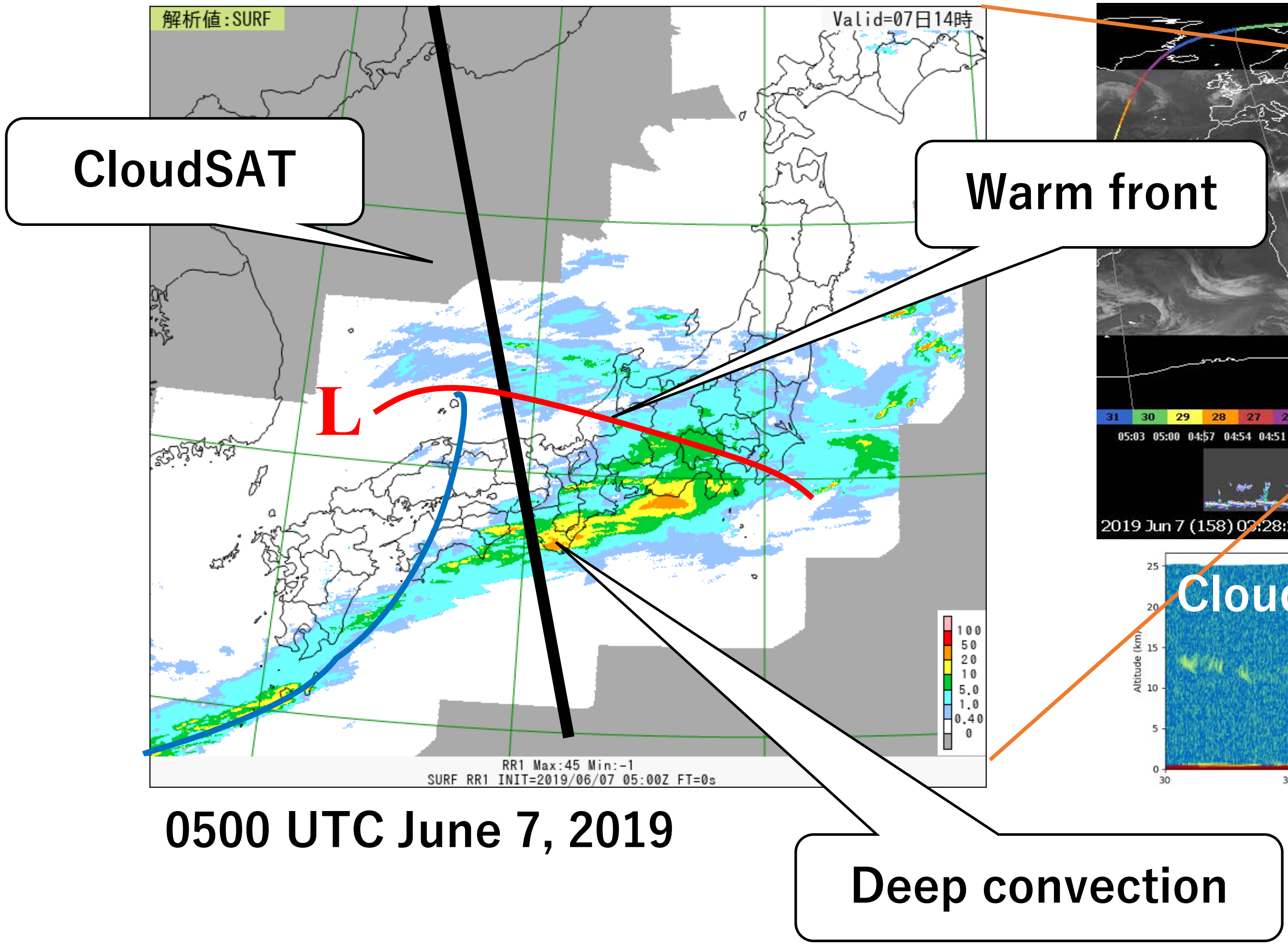
Doppler velocity



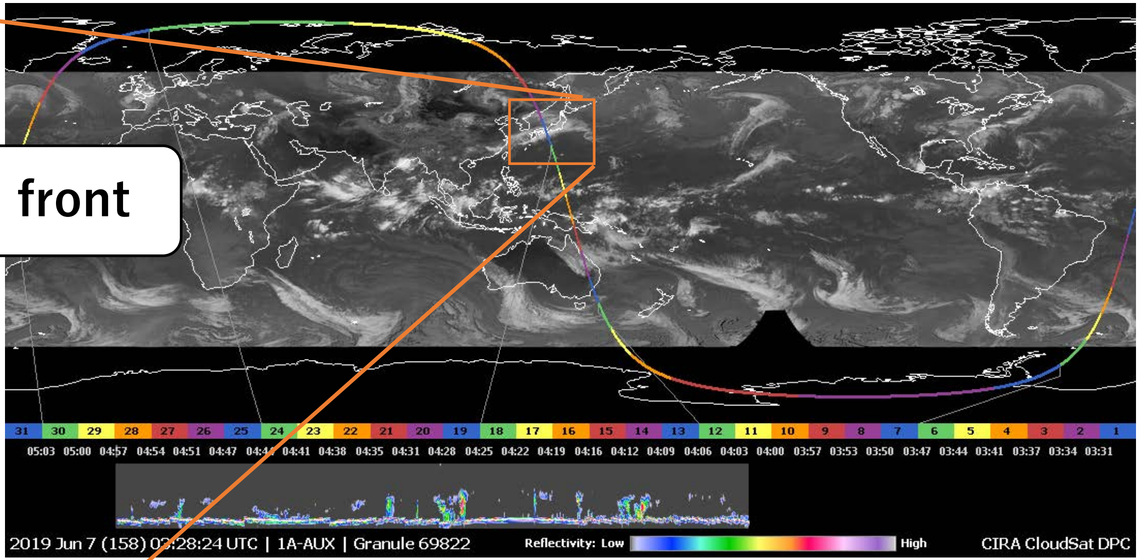
https://www.satnavi.jaxa.jp/ja/news/2024/10/04/9923/index.html

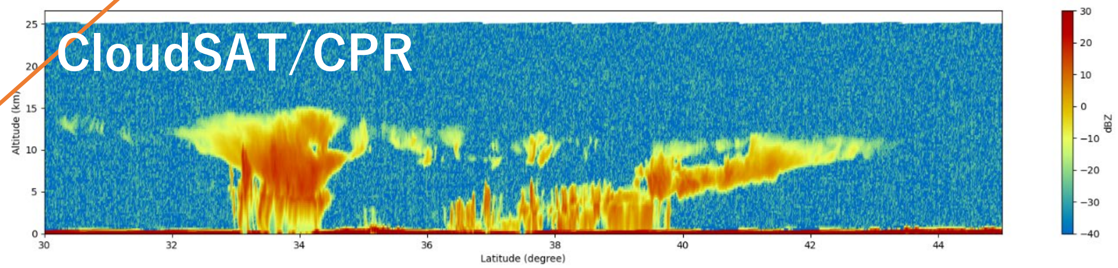
Case study of the precipitation system

Ground-based radar observation



0500 UTC June 7, 2019





Data assimilation system in this study

Method	Hybrid-4DVar
Horizontal grid spacing	Outer model: 5 km Inner model: 15 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, Wg: soil volumetric water content **Qi: mixing ratio of cloud ice** **Qs: mixing ratio of snow** **Qg: mixing ratio of graupel** **Qc: mixing ratio of cloud water** **Qr: mixing ratio of rain**
Local ensemble tangent-linear model	Cloud microphysics scheme
Climatological background error	NMC method
Ensemble background error	100-member EDA (15km)

In this study. we focus on **BG error of hydrometeors.**

- ✓ This system is based on JMA's regional data assimilation system (Ikuta et al. 2020).
- ✓ Related developments were introduced by Kawabata et al. (P1.02) on Monday.

LETLM for cloud microphysics scheme

Tendency of microphysics scheme

$$x(t+1) = x(t) + \left(\frac{dx(t)}{dt}\right)_{mp} \Delta t$$

$$\left(\frac{dx(t)}{dt}\right)_{mp} = \mathbf{M}_{mp}\,\mathbf{K}_{diag}\,x(t)$$

Output **LETLM** **Input**

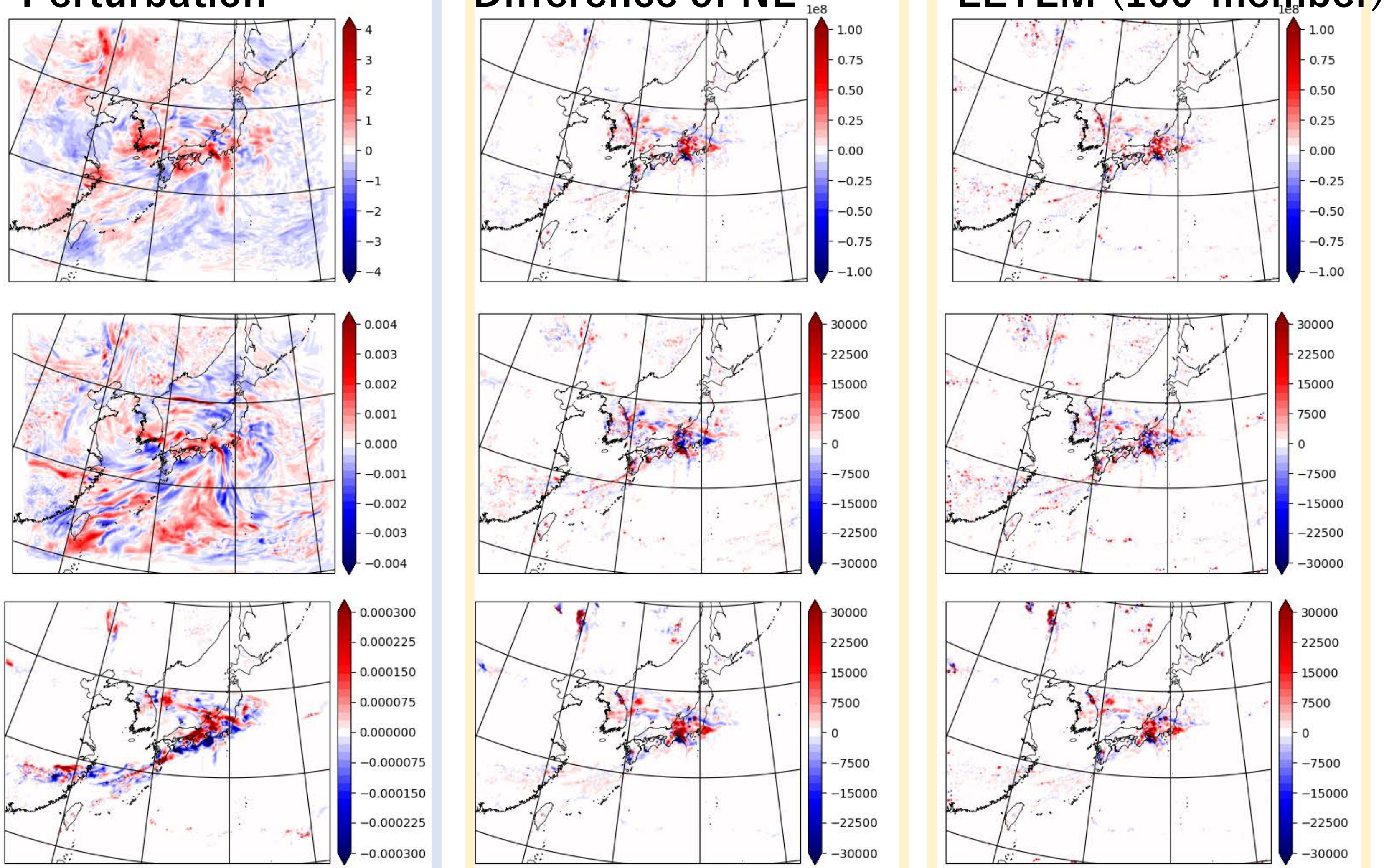
$$x(t) = \begin{pmatrix} \delta\left(\frac{\rho}{J}\right) \\ \delta\left(\frac{\rho\theta_m}{J}\right) \\ \delta\left(\frac{\rho Q_v}{J}\right) \\ \delta\left(\frac{\rho Q_i}{J}\right) \\ \delta\left(\frac{\rho Q_s}{J}\right) \\ \delta\left(\frac{\rho Q_g}{J}\right) \\ \delta\left(\frac{\rho Q_c}{J}\right) \\ \delta\left(\frac{\rho Q_r}{J}\right) \end{pmatrix} \quad \frac{dx(t)}{dt} = \begin{pmatrix} \delta\frac{d}{dt}\left(\frac{\rho}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho\theta_m}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho Q_v}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho Q_i}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho Q_s}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho Q_g}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho Q_c}{J}\right) \\ \delta\frac{d}{dt}\left(\frac{\rho Q_r}{J}\right) \end{pmatrix} \quad \mathbf{K}_{diag}x(t) = \begin{pmatrix} \delta p \\ \delta\theta \\ \delta Q_v \\ \delta Q_i \\ \delta Q_s \\ \delta Q_g \\ \delta Q_c \\ \delta Q_r \end{pmatrix}$$

Prognostic variable **Tendency** **Diagnostic variable**

LETLM for cloud microphysics scheme

Perturbations of tendency are qualitatively similar

Input Perturbation	Output Difference of NL	Output LETLM (100-member)



Potential temperature $\delta\theta$ — $\delta \frac{d}{dt}\left(\frac{\rho\theta_m}{J}\right)$

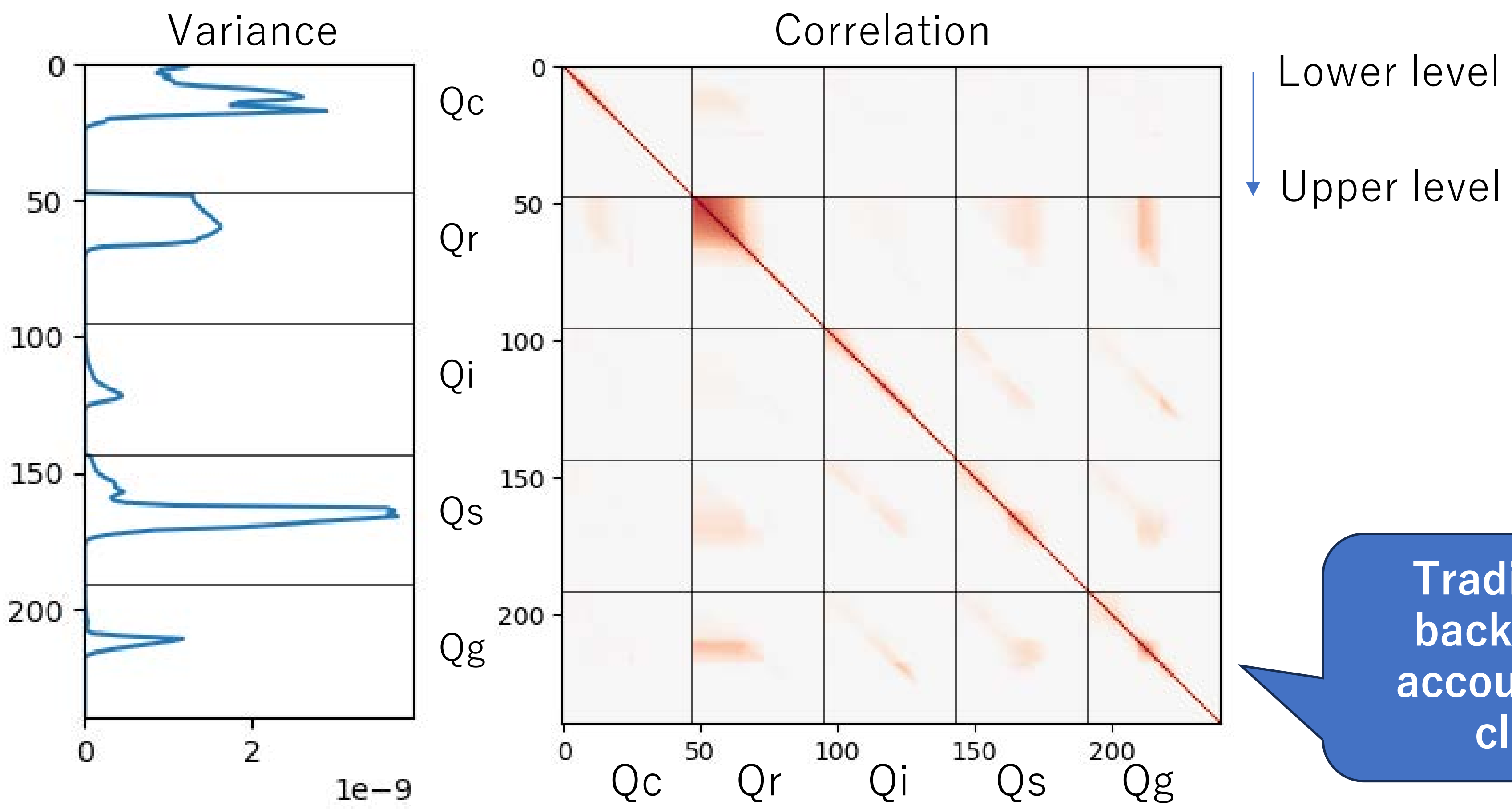
Specific humidity δQ_v — $\delta \frac{d}{dt}\left(\frac{\rho Q_v}{J}\right)$

Mixing ratio of rain δQ_r — $\delta \frac{d}{dt}\left(\frac{\rho Q_r}{J}\right)$

at 10-level

Climatological background error covariance of hydrometeors

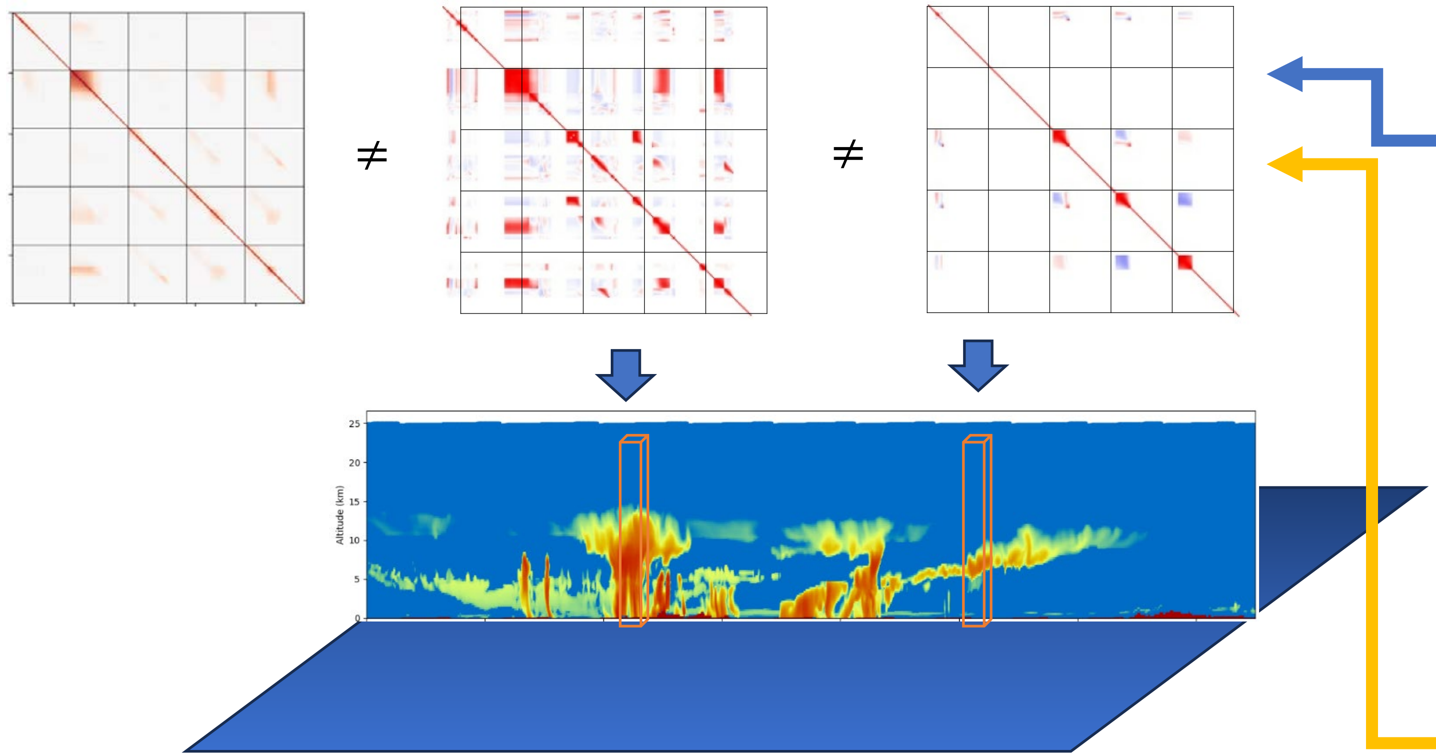
Summer/winter average using 100-member Ensemble DA (EDA) around Japan



Traditional climatological background errors do not account for the diversity of cloud precipitation.

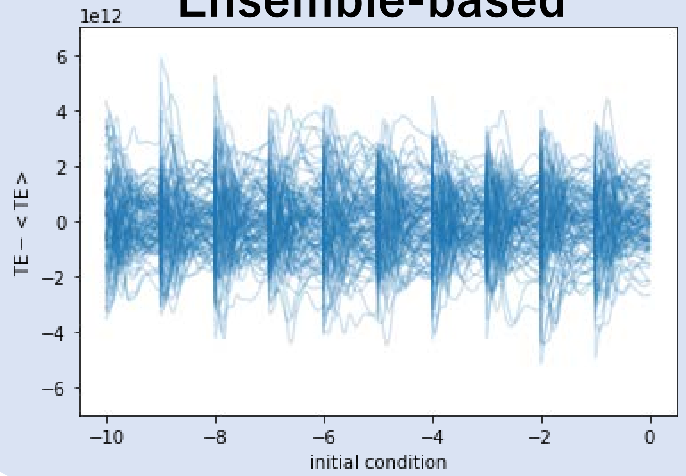
Flow-dependent background covariance

Climatological | **Situation dependent**



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

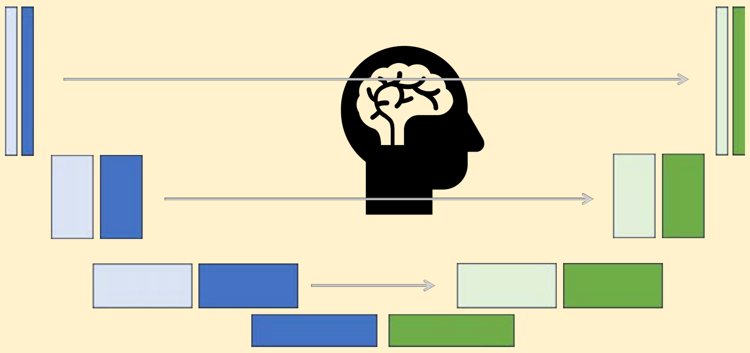
Ensemble-based



CNTL

or

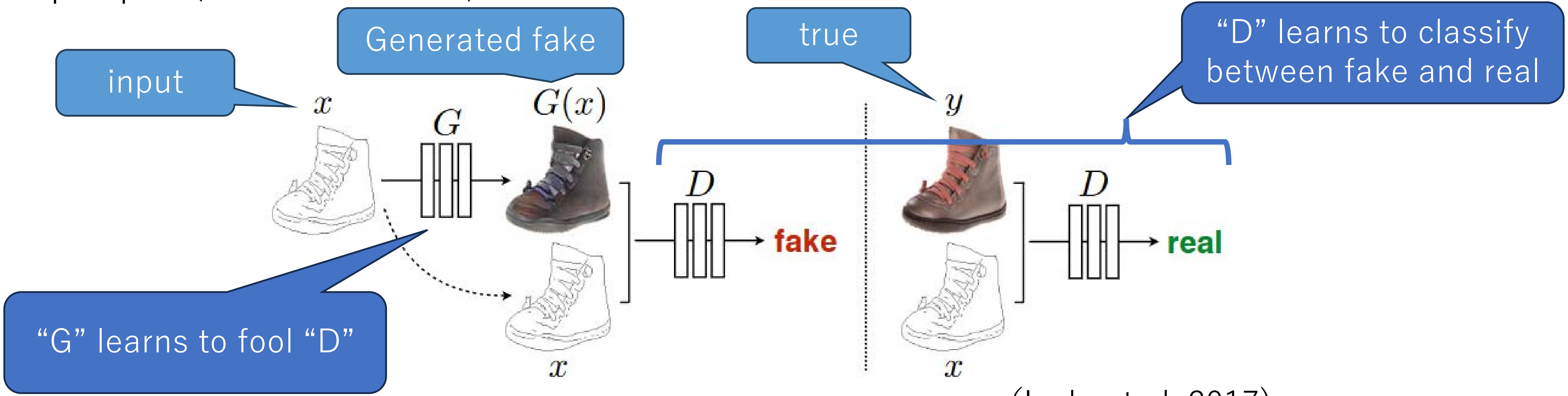
ML-based



TEST

ML-based BG error estimating method

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

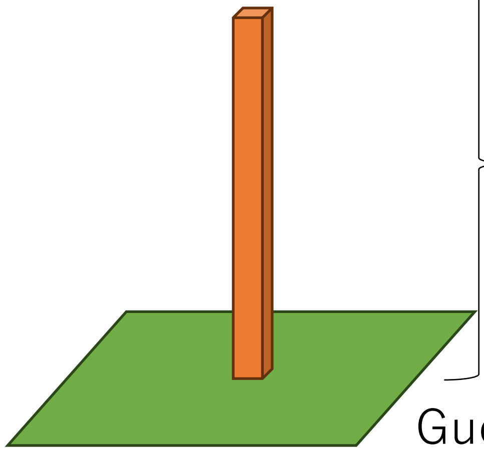


(Isola et al. 2017)

✓ CGAN trains the Generator (G) and Discriminator (D) by giving them additional **condition information**.
✓ CGAN can **learn to accept** only the correct combination of real data and labels and **learn to reject** all other data. CGAN generates images according to specified conditions.

Making of a tuple for training data

Deterministic DA

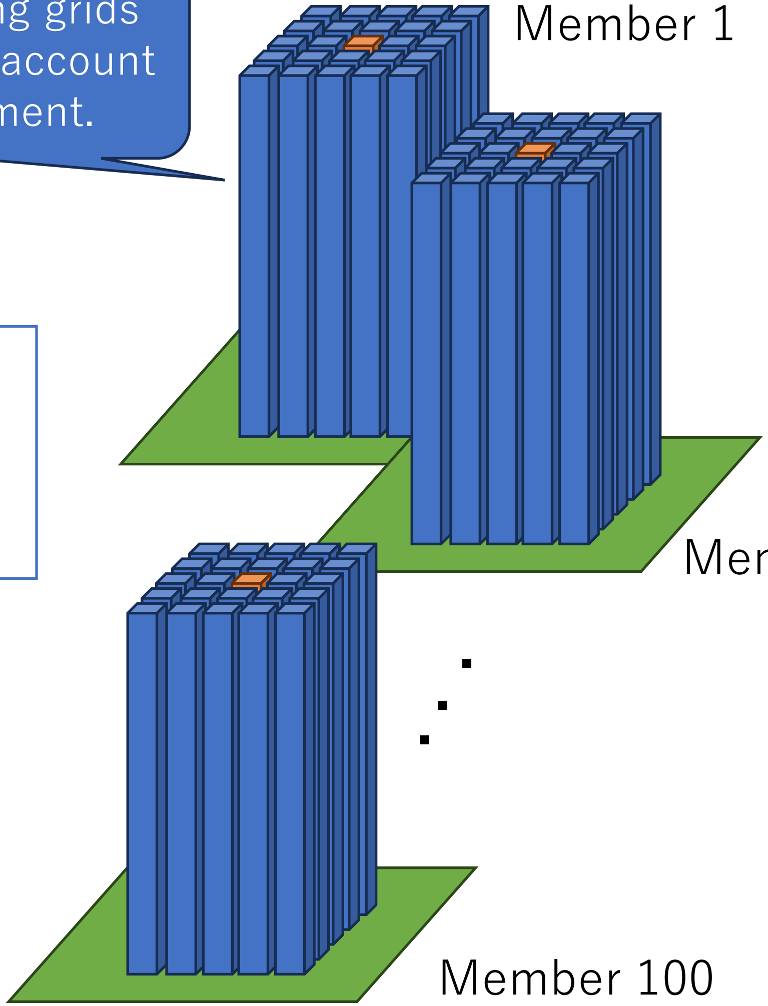


Condition vector: Presence or absence of hydrometeors

Condition vector is created from the profile in guess,

Ensemble Data Assimilation (EDA)

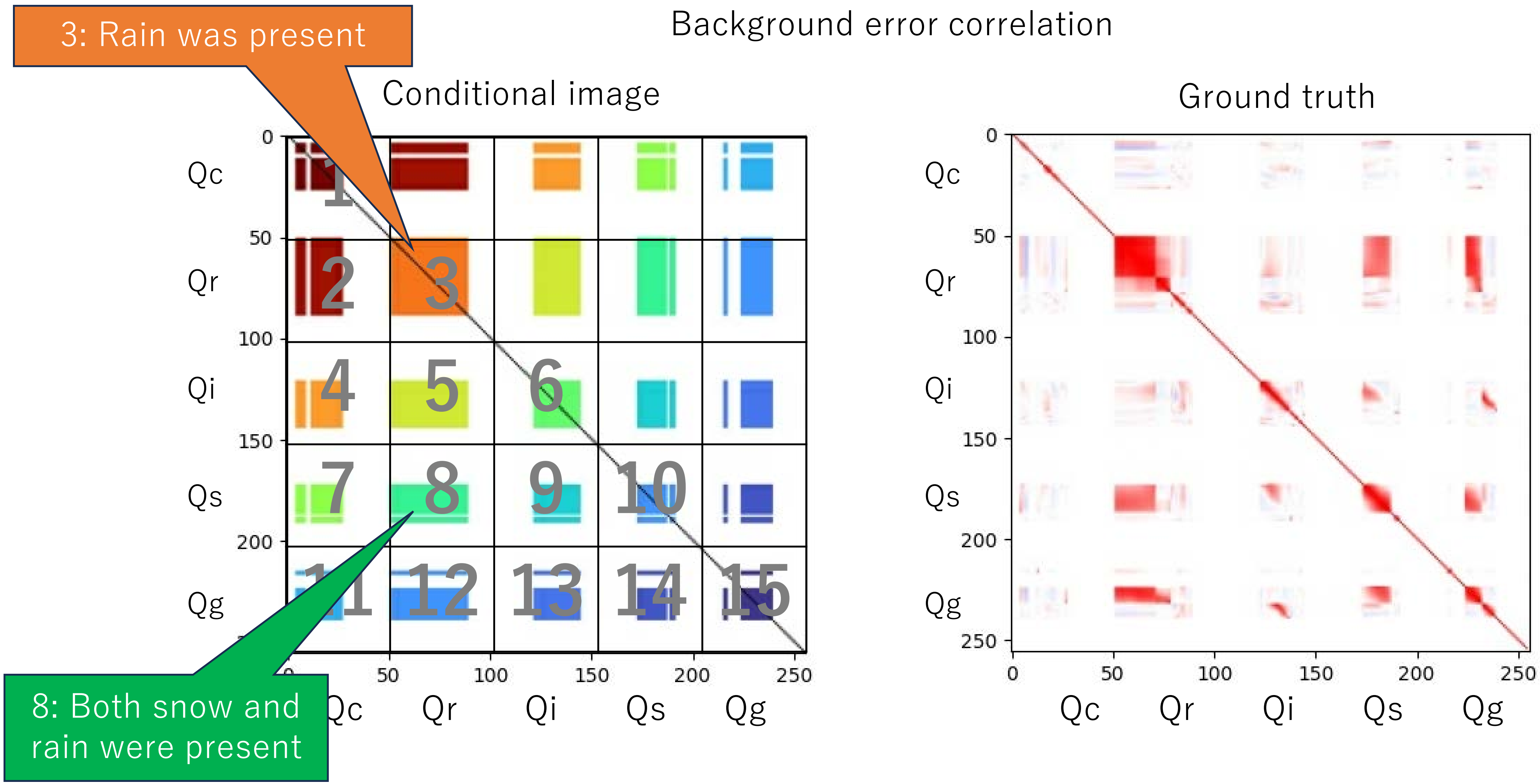
The surrounding grids are included to account for misalignment.



Ground truth: Background error covariance of hydrometeors

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

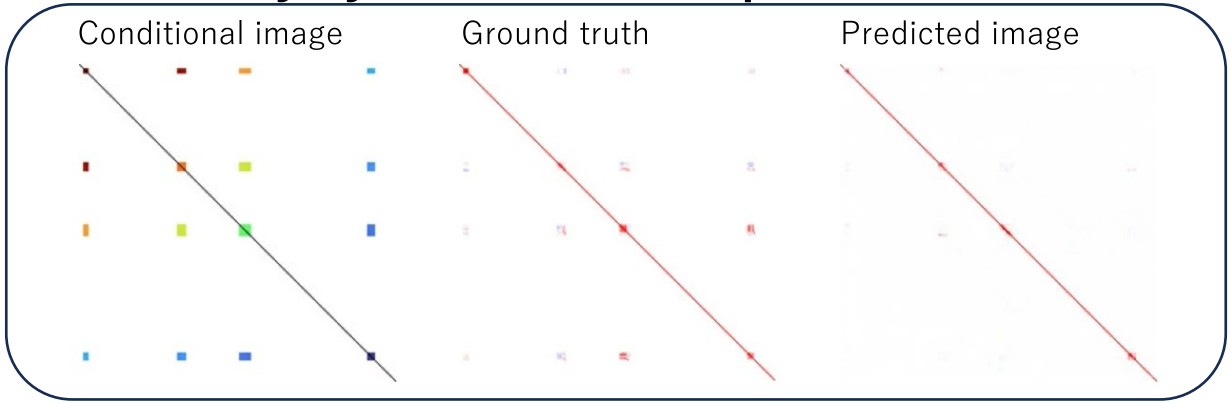
Conditional image and ground truth image

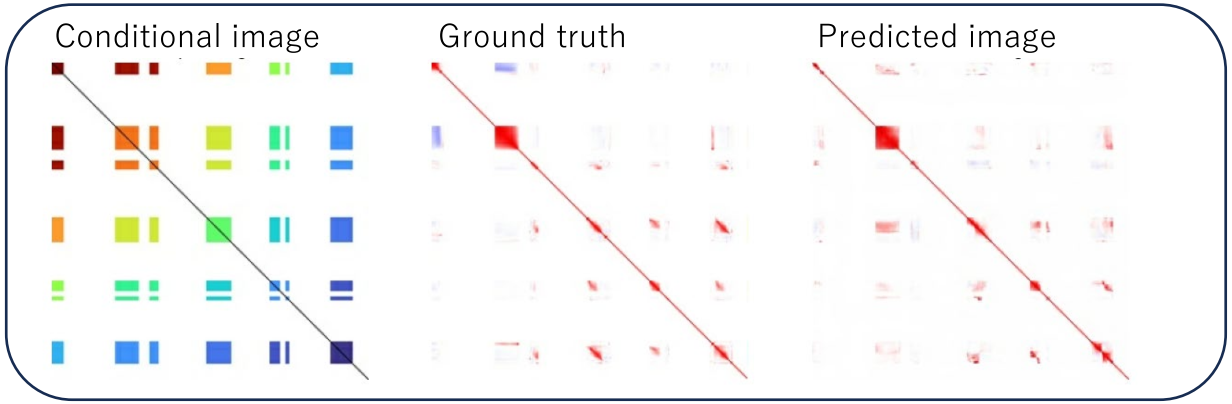


Learning results

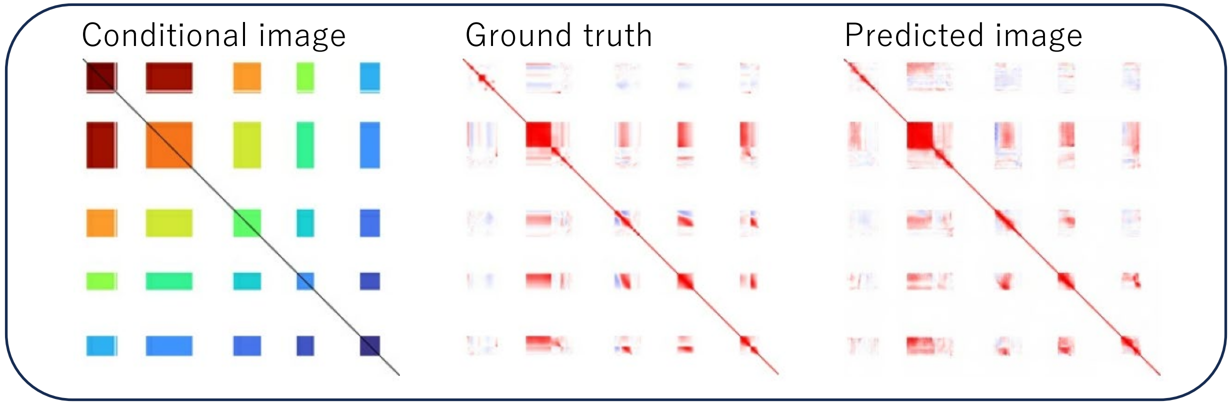
In all cases, the predicted structure was generally close to the ground truth.

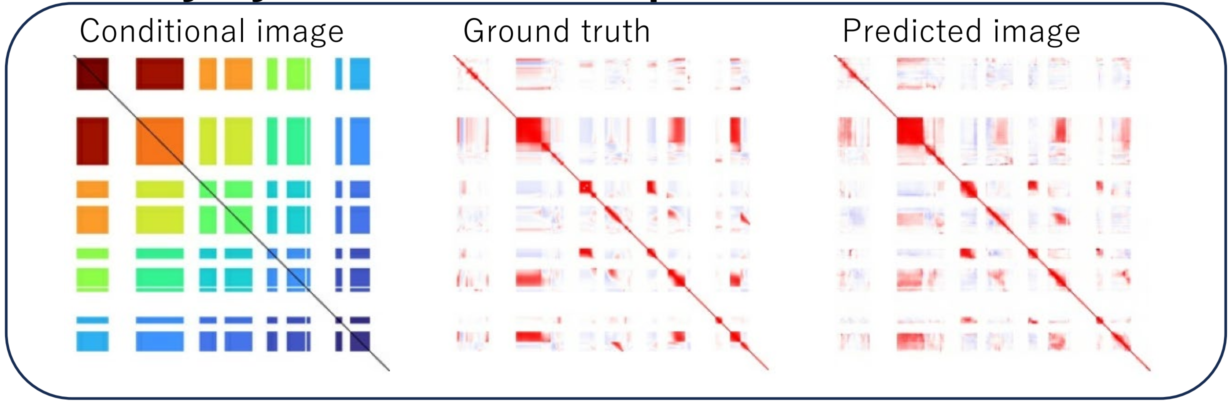
Not many hydrometeors are present.





Many hydrometeors are present.



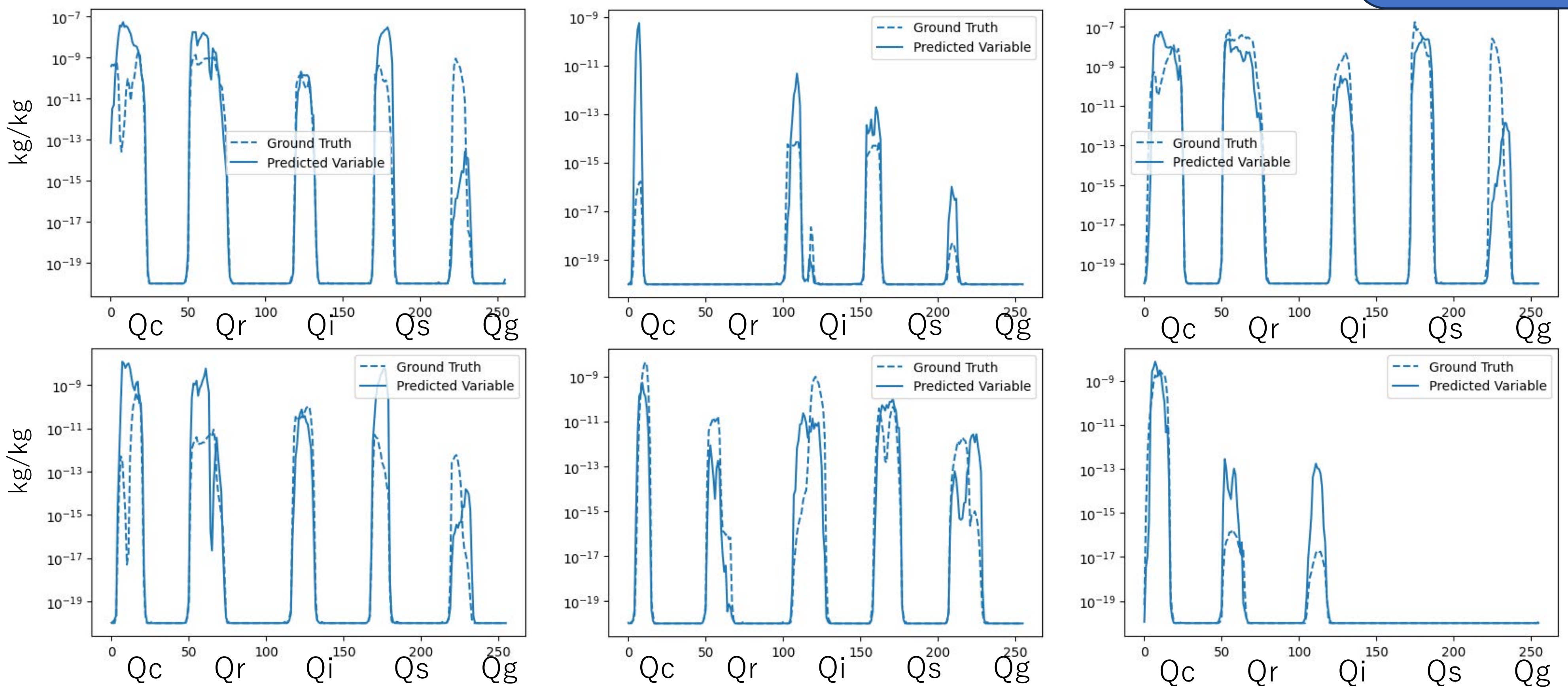


These results can be generated instantly from saved Checkpoints.

Test of learning result for variance

The structure of variance was generated. But magnitudes are different.

Variances are generated by CGAN using the same way of error correlatio



These results can be generated instantly from saved Checkpoints.

Re-scaling to match the climatological BG error variance

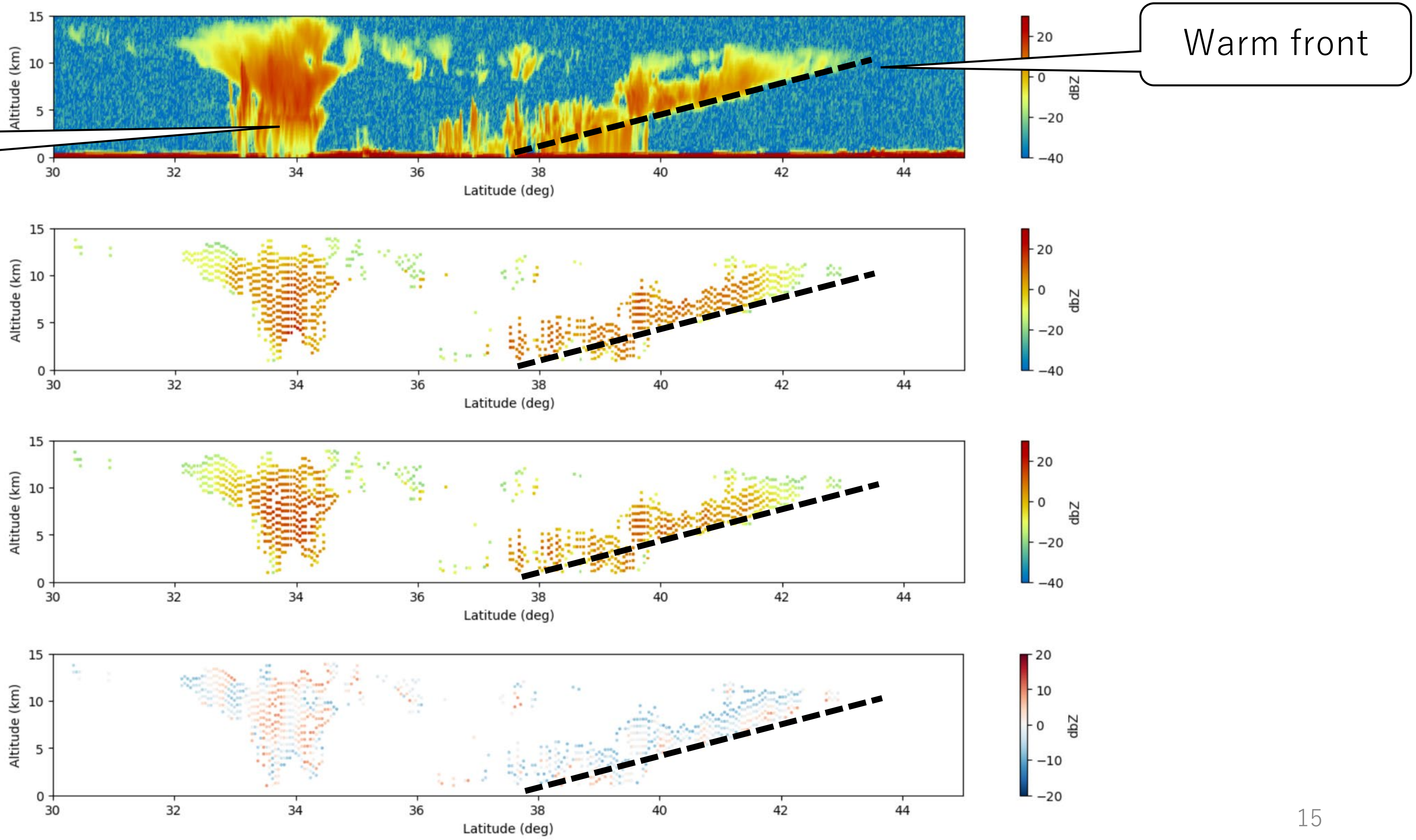
Observed reflectivity and the first guess

Observation

Thinned observation after QC

First-guess at observation points

Observation minus the first-guess



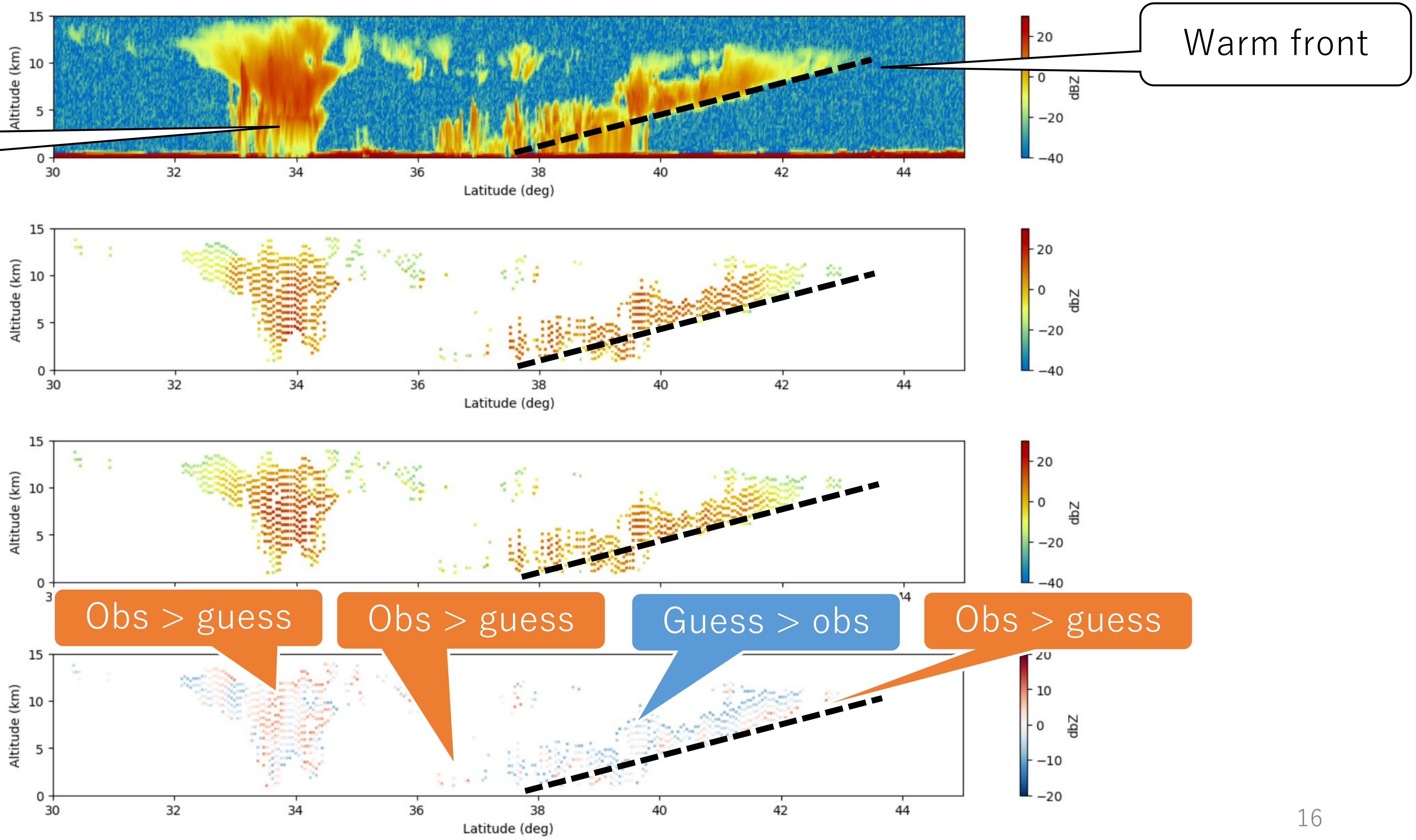
Observed reflectivity and the first guess

Observation

Thinned observation after QC

First-guess at observation points

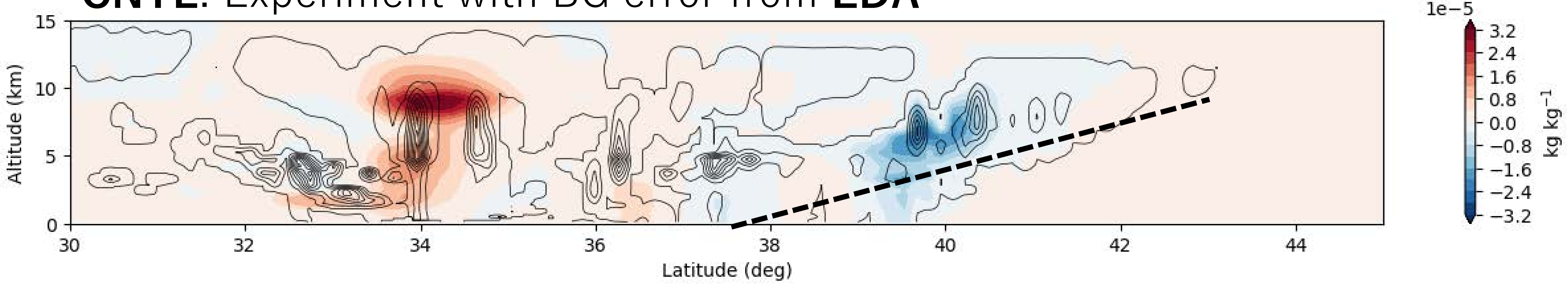
Observation minus the first-guess



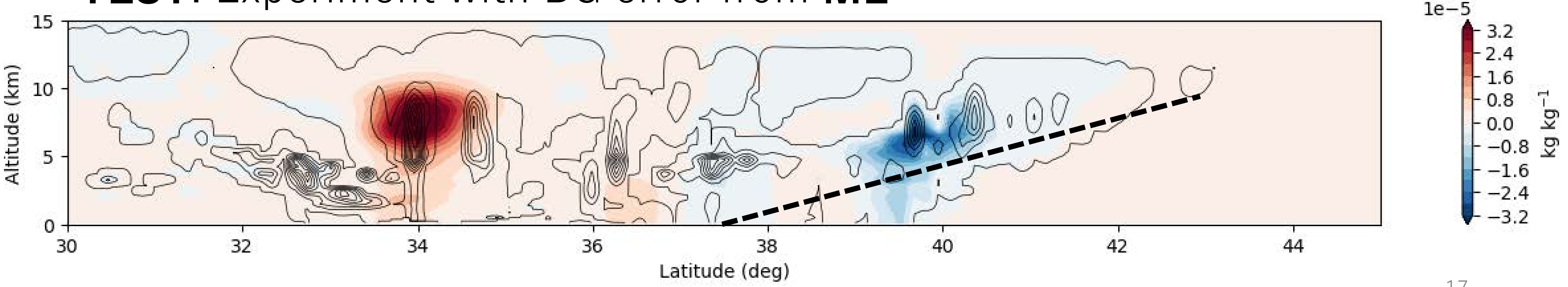
Analysis increments

$$\delta Q = \delta Q_c + \delta Q_r + \delta Q_i + \delta Q_s + \delta Q_g$$

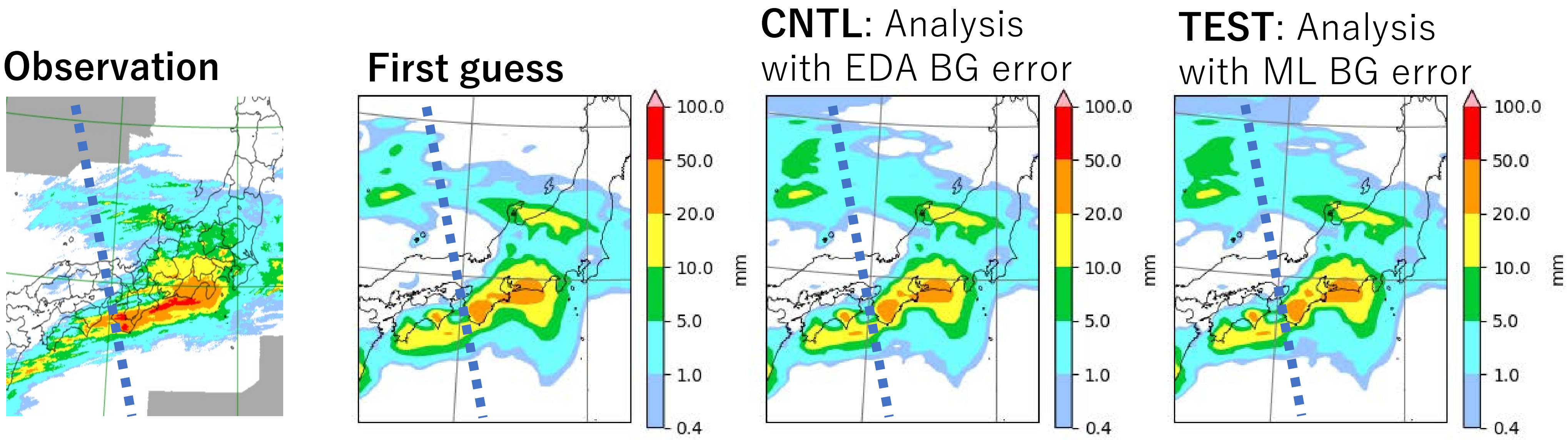
CNTL: Experiment with BG error from **EDA**

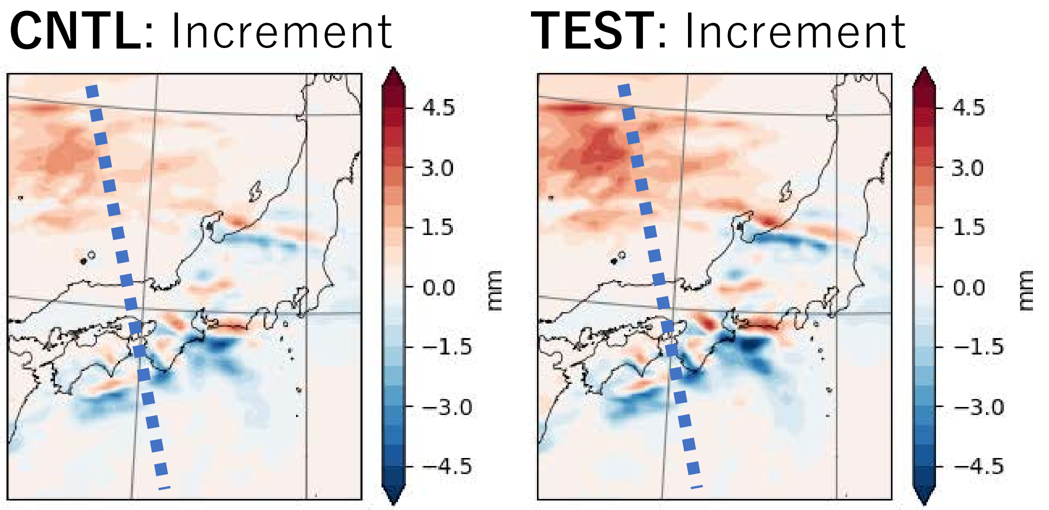


TEST: Experiment with BG error from **ML**



Impact on precipitation in assimilation window





- ✓ CPR assimilation increased precipitation.
- ✓ ML's BG error performance is similar to EDA's BG error

Conclusion

- We tried to generate a background error correlation matrix using CGAN.
    - As a result of giving the presence or absence of hydrometeors as a condition vector, it was found that a covariance matrix close to the ground truth could be generated.
    - These results suggest obtaining flow-dependent hydrometeors background errors using ML without preparing ensemble predictions is possible.
- The analysis using ML's BG error was able to achieve precipitation simulations similar to the analysis using EDA's BG error.
- However, there are still issues that need to be resolved for practical application.
    - The generated variance has a large error.

Future plans

- Utilize EarthCARE's reflectivity and Doppler velocity
- The background error correlation between vertical velocity, temperature, and hydrometeors is estimated using ML.
- Replace physics-based model with ML-model.