Examinations of nonlinearities and non-Gaussianities in the assimilation of all-sky microwave observations

Chih-Chi Hu[1] and Mingjing Tong[2]

[1] Princeton University, Cooperative Institute for Modeling the Earth System (CIMES)
[2] Geophysical Fluid Dynamics Laboratory (GFDL)

10th International Symposium on Data Assimilation (ISDA)
Oct 25, 2024

The careful specification of the observation error model plays an important role in the assimilation of satellite radiances

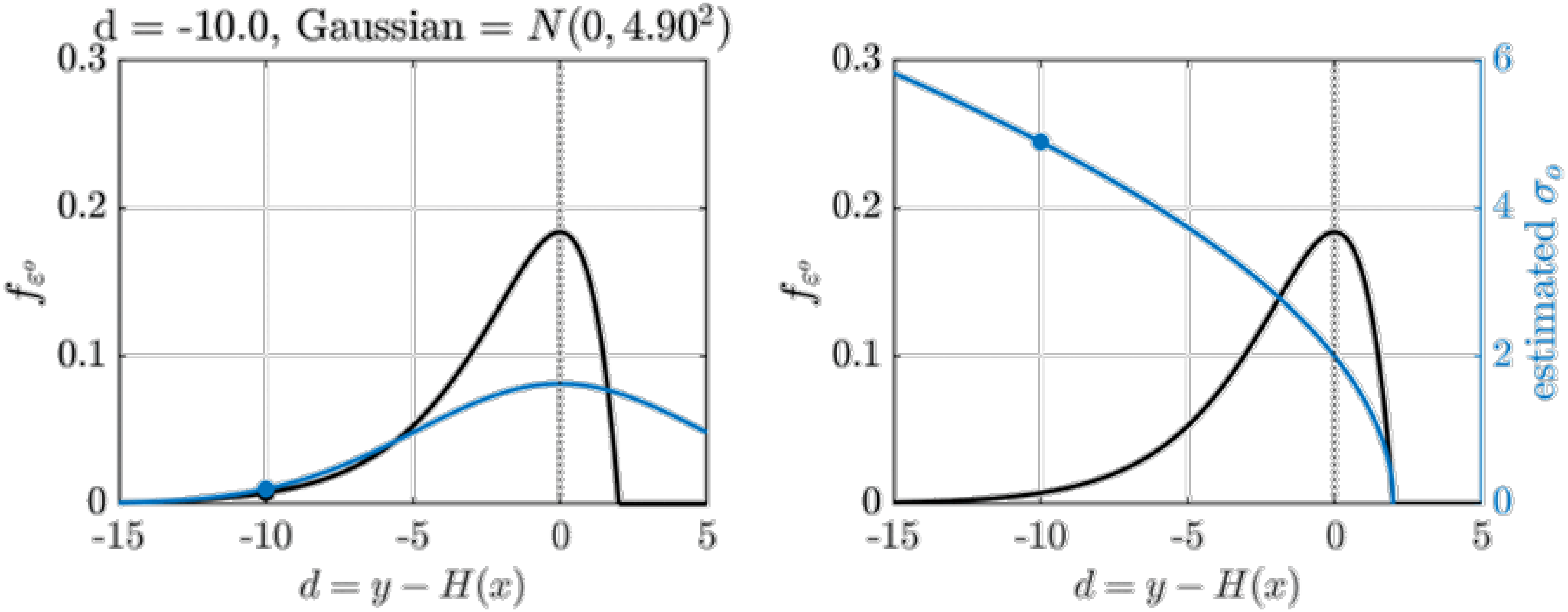
- **Mean**:
  - Variational bias correction (VarBC)

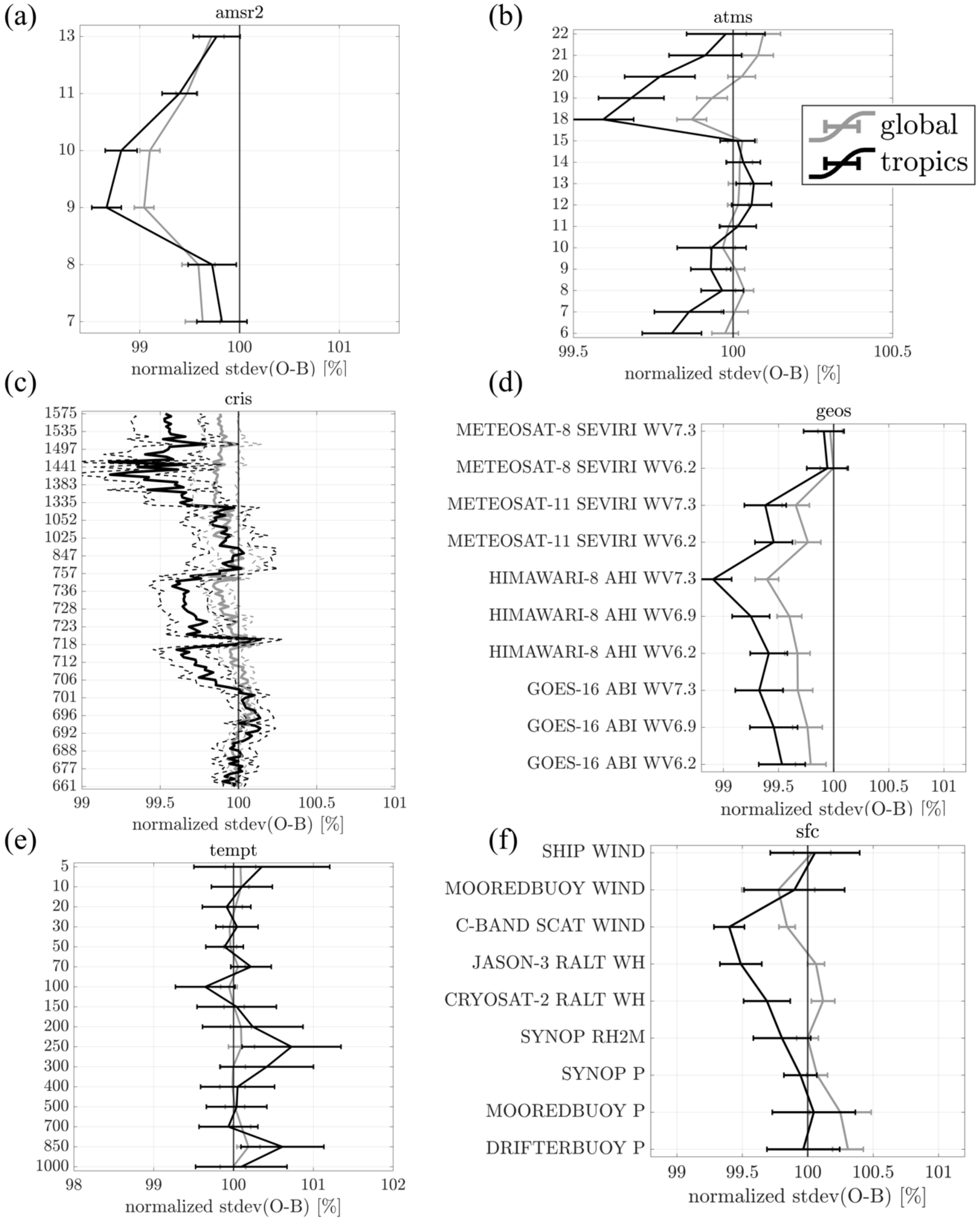
    (e.g., Derber and Wu 1998; Dee 2004; Zhu et al. 2014; Han and Bormann 2016)
- **Variance**:
  - State-dependent variance model *(e.g., Geer and Bauer 2011; Okamoto et al. 2014; Harnisch et al. 2016)*
  - Variational Quality Control (VarQC) *(Anderson and Jarvinen 1999)*
  - Adaptive Observation Error Inflation (AOEI) *(Minamide and Zhang 2017)*
- **Correlation**:
  - Spatial & inter-channel correlations

    (e.g, Bormann and Bauer 2010; Bormann et al. 2010; Weston et al. 2014; Stewart et al. 2014; Geer 2019; etc)

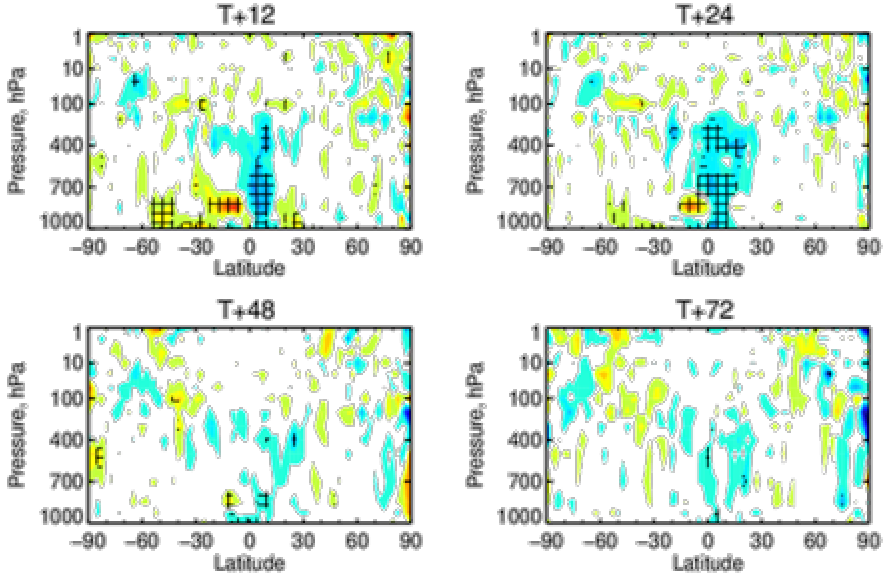
What about the non-Gaussianities in the observation error?

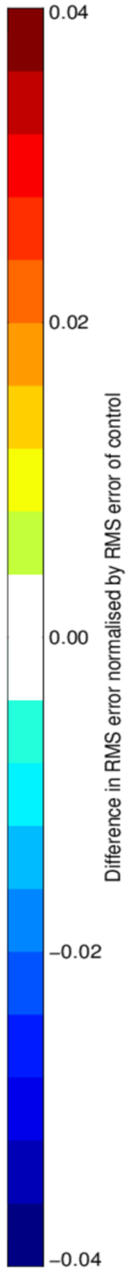
- A technique to estimate a non-parametric observation error PDF:
  - *Hu, C. C., Van Leeuwen, P. J., & Geer, A. J. (2024). A non-parametric way to estimate observation errors based on ensemble innovations. Quarterly Journal of the Royal Meteorological Society*
- To incorporate a non-parametric observation error into the "incremental-Var" method:
  - A "lookup table approach": evolving-Gaussian method *(Hu, Geer and van Leeuwen. To be submitted)*
  - **Key idea**: find a Gaussian for each value of $y - H(x)$, such that the local gradient of the Gaussian PDF is always the same as that of the non-Gaussian PDF



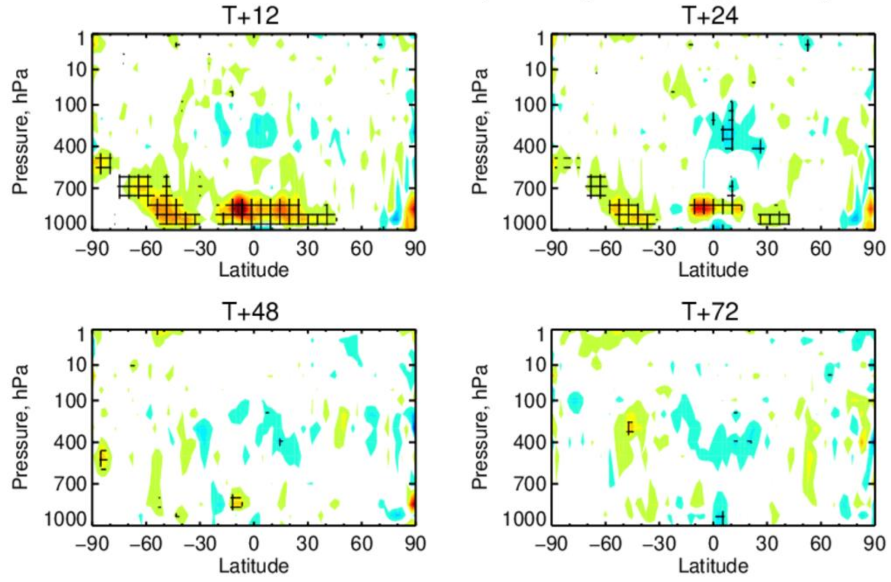


dRMS in R (only tropics)





dRMS in T (only tropics)



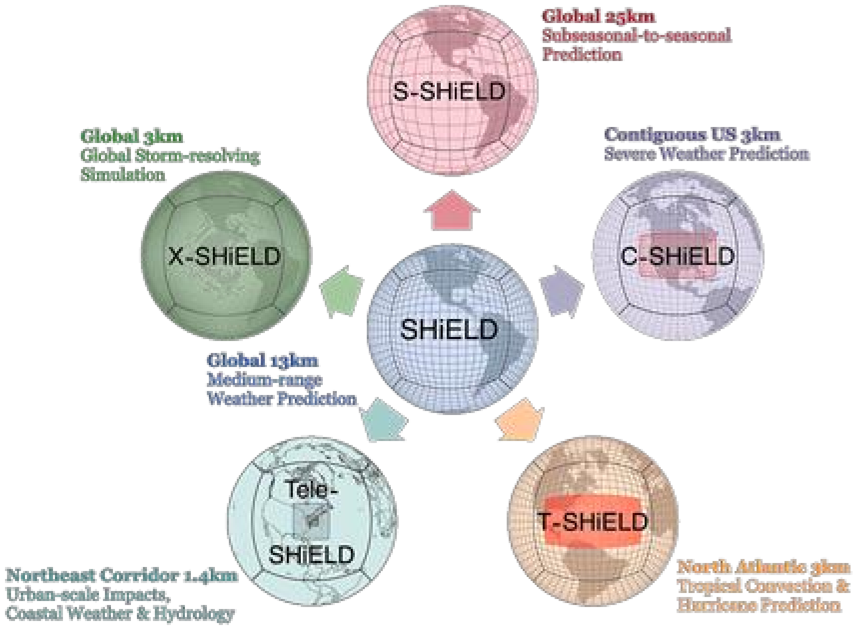
(Hu, Geer and van Leeuwen. To be submitted)

To examine the impact of non-Gaussian errors on a broader range of all-sky channels (AMSUA & ATMS)

- SHiELD (**S**ystem for **Hi**gh-resolution prediction on **E**arth-to-**L**ocal **D**omains)

https://www.gfdl.noaa.gov/shield/

- SHiELD-DA system:
  - GSI-based hybrid 4DEnVar, configured to closely match GFSv16.2
  - For fast testing:
    - C192 resolution (~ 50 km) global cycling DA
    - “ensemble replay” mode



- All-sky AMSUA and ATMS assimilation in GSI:
  - AMSUA: ch1-5 & ch15 / ATMS: ch1-6 & ch16-22 (allsky only over the ocean)
  - VarBC designed for all-sky channels *(Zhu et al. 2016; Zhu et al. 2019)*
  - Symmetric cloud error model *(Geer and Bauer 2011)* + “empirical inflation”

"Empirical inflation" downweighs most of the observations

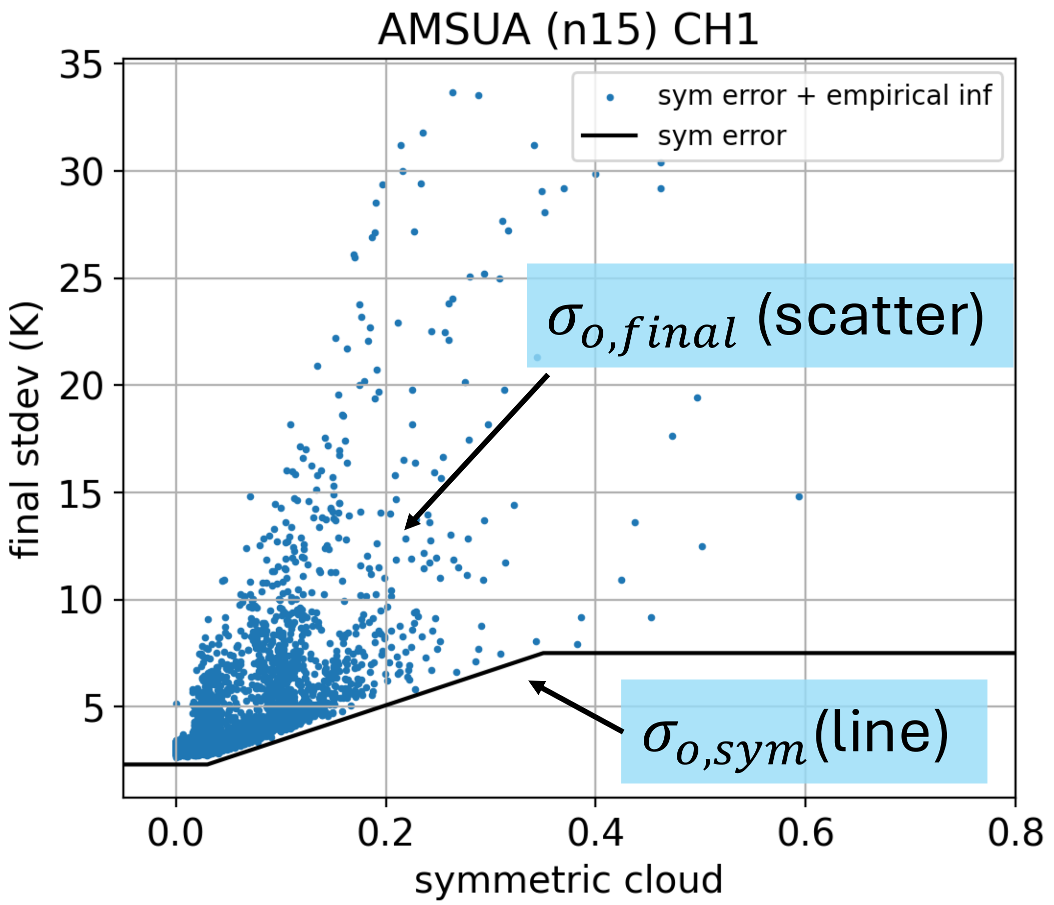
- In GSI, the observation error standard deviation is inflated through a situation-dependent inflation, which is empirical and requires careful tuning (hereafter, "empirical inflation")

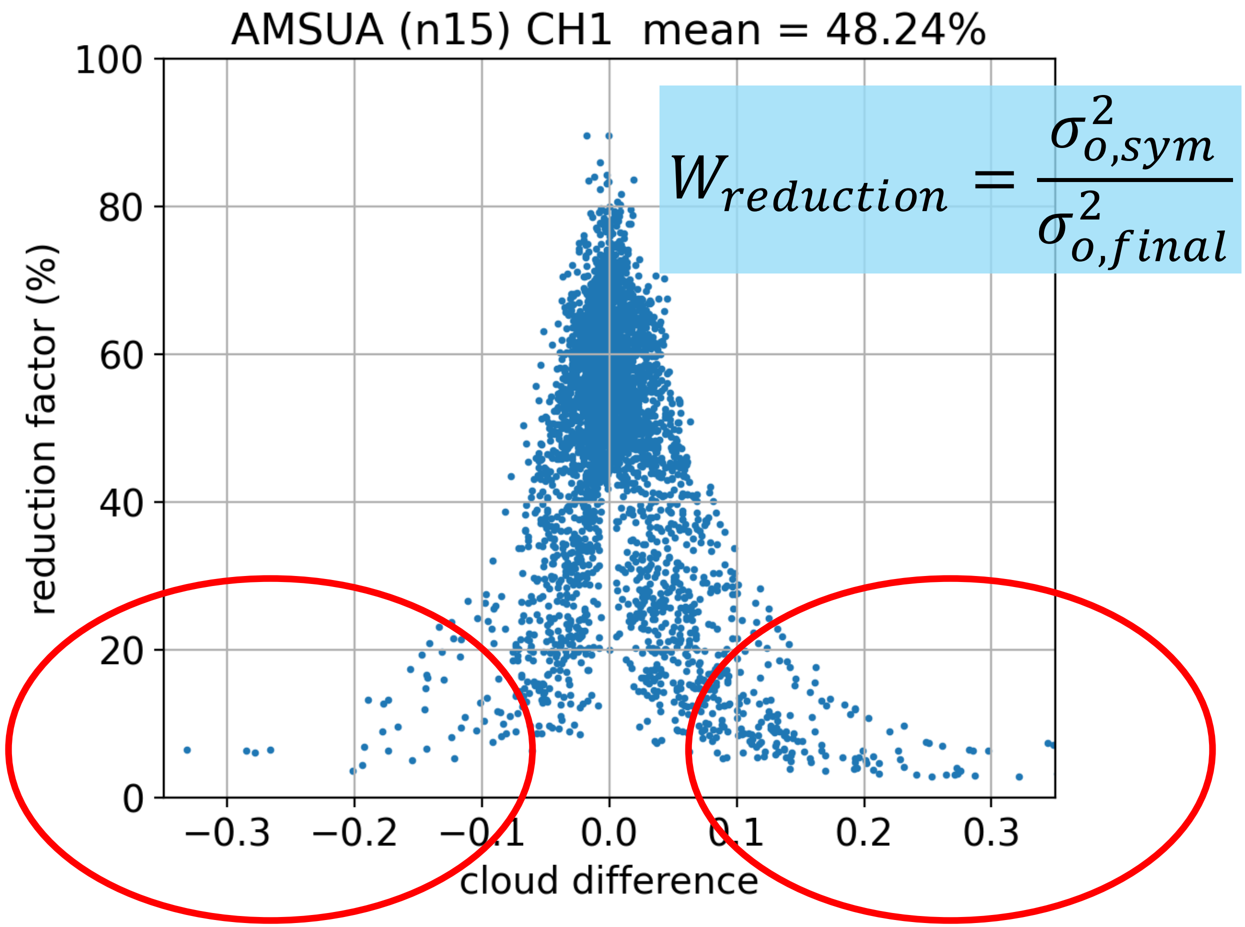
- $e = e_{sym} + \Delta e$
- $\Delta e = \Delta e_1 + \Delta e_2 + \Delta e_3 + \Delta e_4$
  - $\Delta e_1 = (1 - \delta)|\Delta T_B|$
    - $\delta = \begin{cases} 0, if\ (C_{obs} - C_{clear})(C_{model} - C_{clear}) < 0\ and\ |C_{obs} - C_{model}| \geq 0.005 \\ 1, otherwise \end{cases}$
  - $\Delta e_2 = \gamma_1 |C_{obs} - C_{model}|\ e_{sym}$
  - $\Delta e_3 = \gamma_2 \max(0, s - 9)\ e_{sym}$
  - $\Delta e_4 = \gamma_3 {w_s}^2\ e_{sym}$

where $\gamma_1 = 13,\ \gamma_2 = 1.5,\ \gamma_3 = 0.002$

“Empirical inflation” downweighs most of the observations

- In GSI, the observation error standard deviation is inflated through a situation-dependent inflation, which is empirical and requires careful tuning (hereafter, “empirical inflation”)
- The inflation amount is correlated to the cloud difference between model and observation





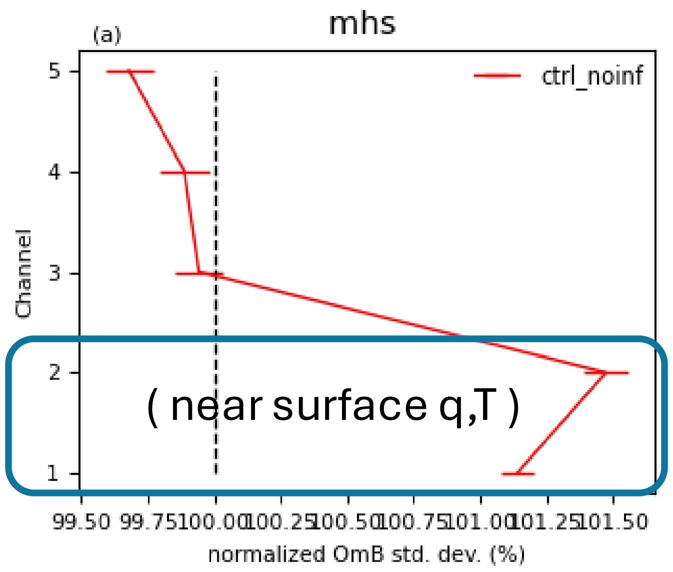
Empirical inflation .vs. non-Gaussian lookup table

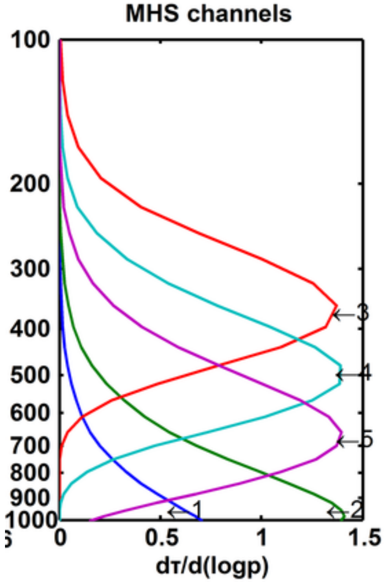
- The empirical inflation complicates the “lookup table approach” in the evolving-Gaussian method for the non-Gaussian observation errors.

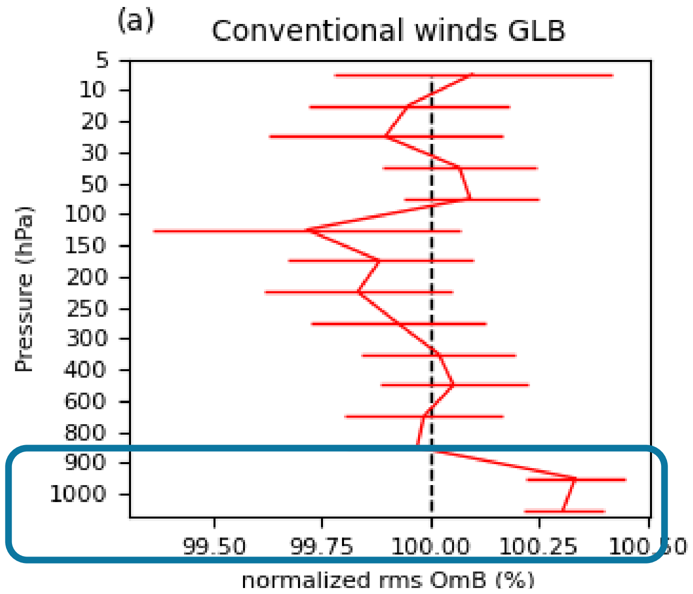
- Replace the empirical inflation by the “lookup table approach” :
  1) It’s not clean to compare Gaussian vs non-Gaussian error with empirical inflation
  2) The “cloud mismatch issue” has already been accounted for in the symmetric cloud error model. The empirical inflation seems to “double count” the issue.
  3) Empirical inflation is difficult to maintain

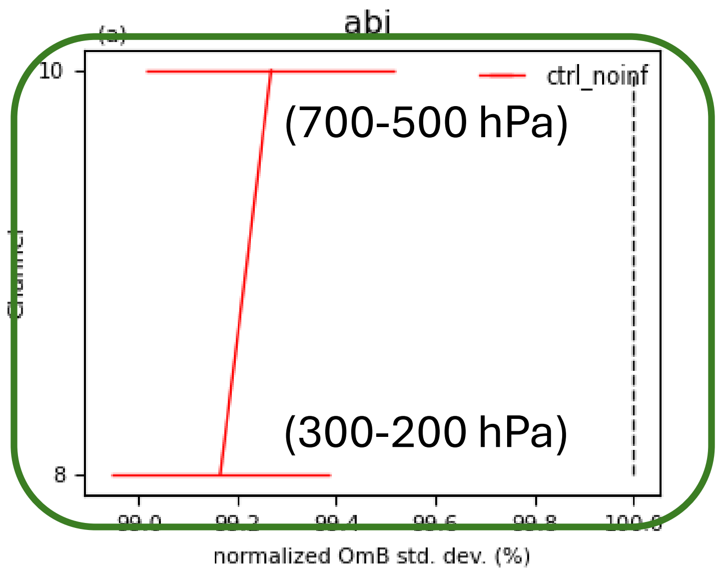
- However, turning off the empirical inflation has a dominant effect on the whole signal.

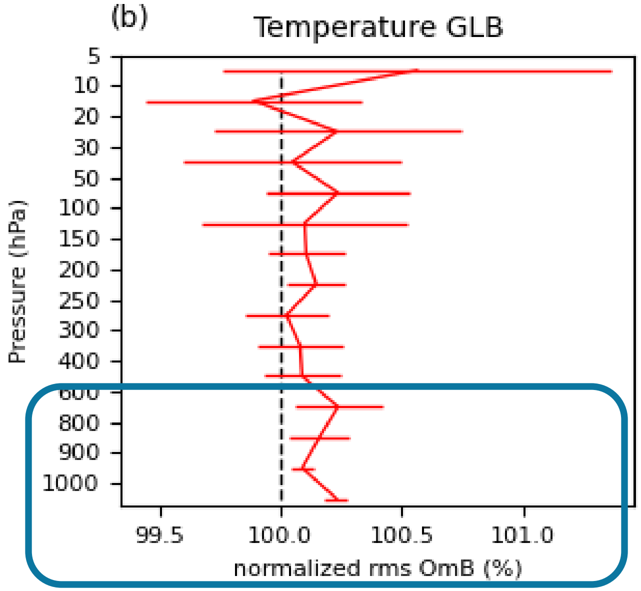
Mixed results from turning off the empirical inflation

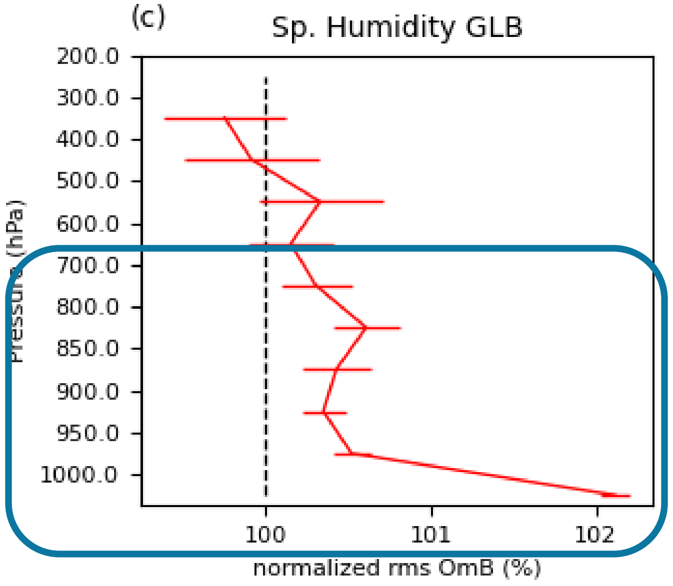




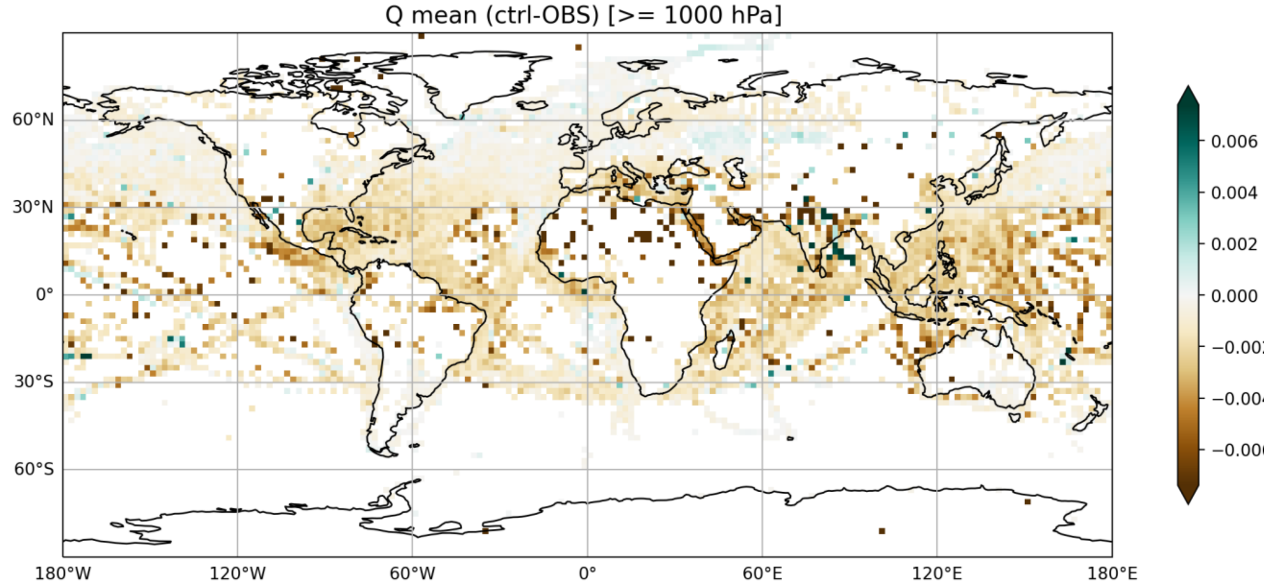




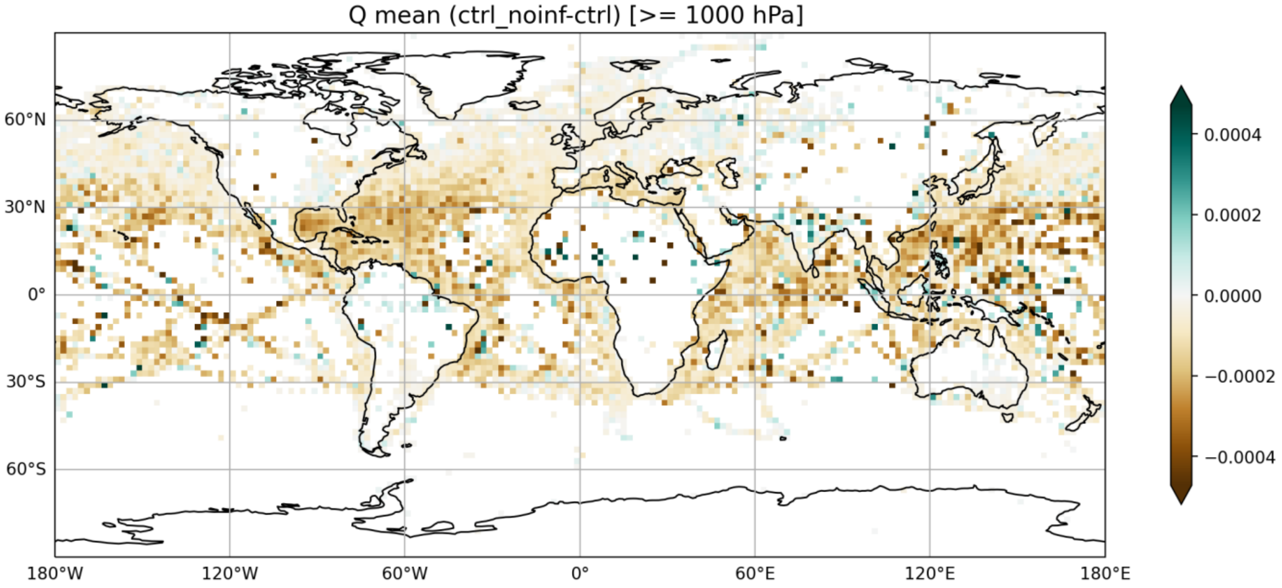




Turning off the empirical inflation worsens the near-surface dry biases

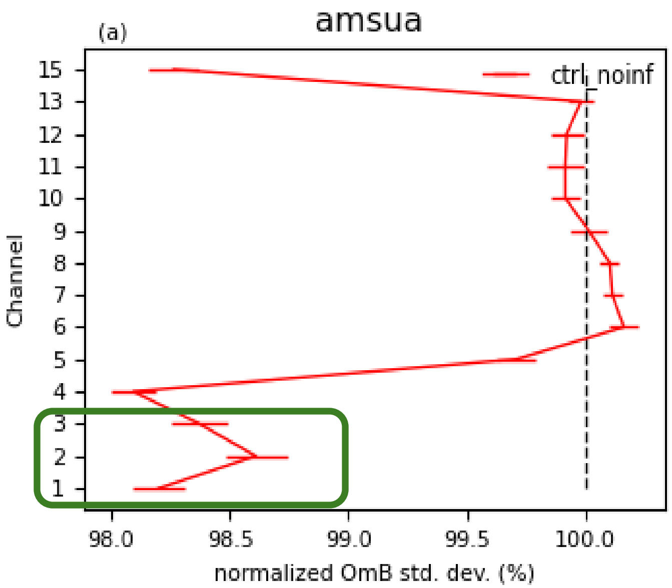


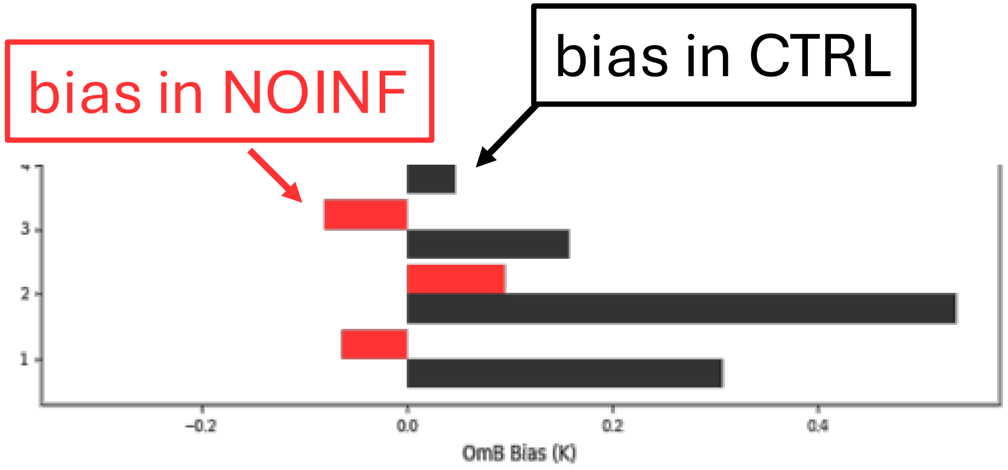
CTRL-OBS (humidity)

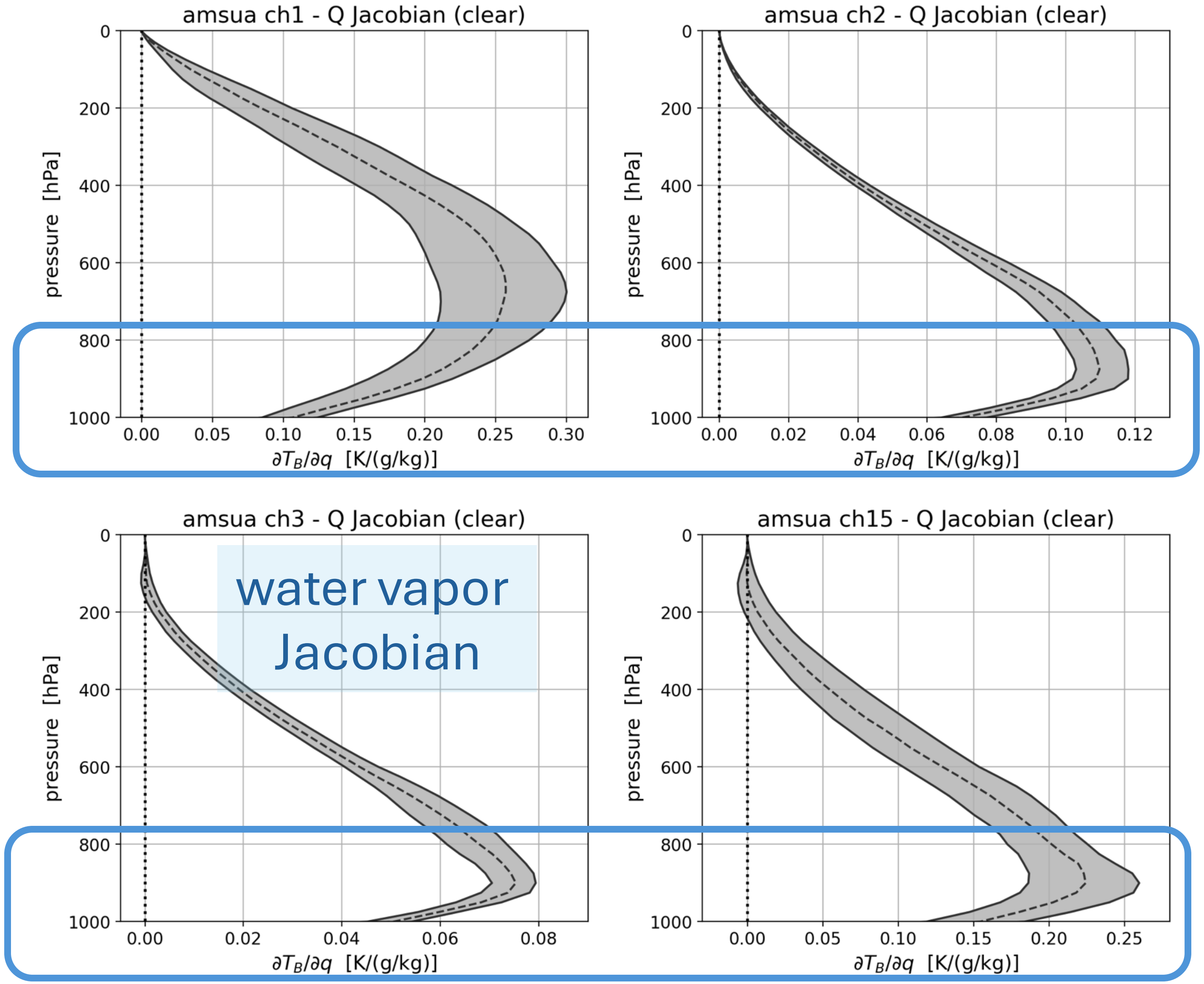


NOINF-CTRL (humidity)

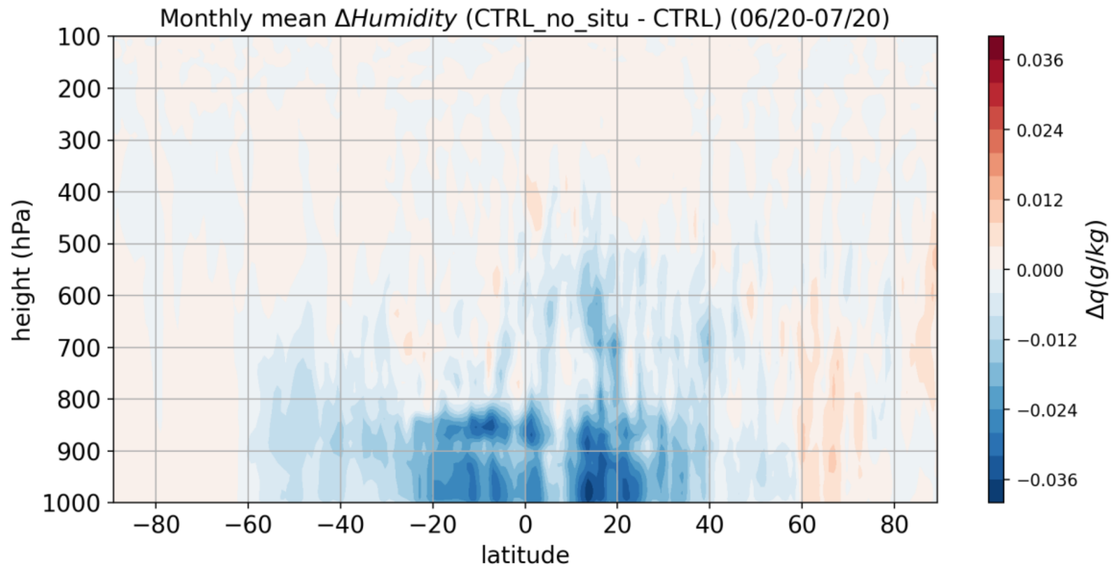
Improved observation space statistics for the window channels when empirical inflation is turned off

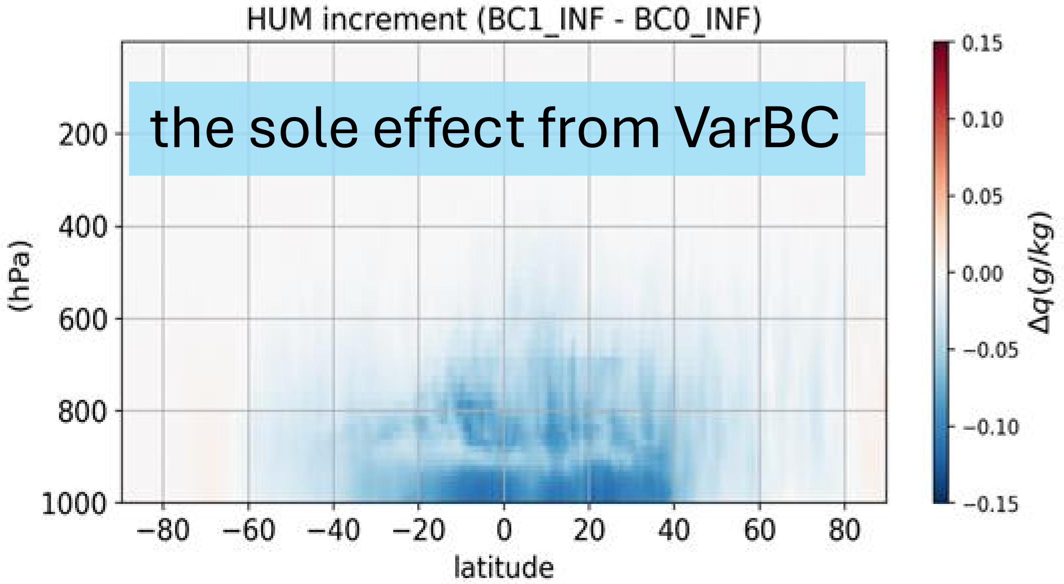




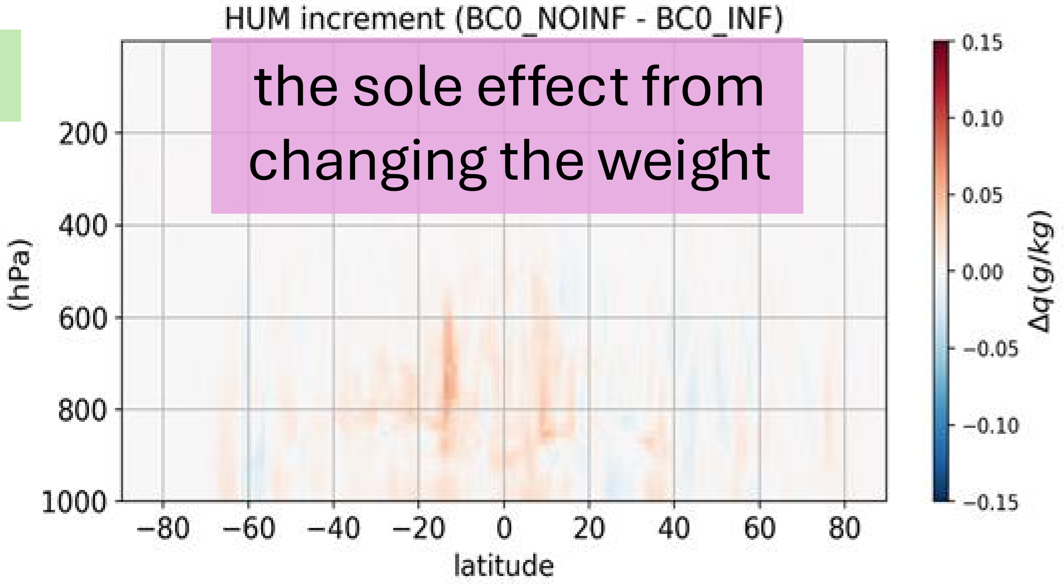


Turning off the empirical inflation also changes the VarBC



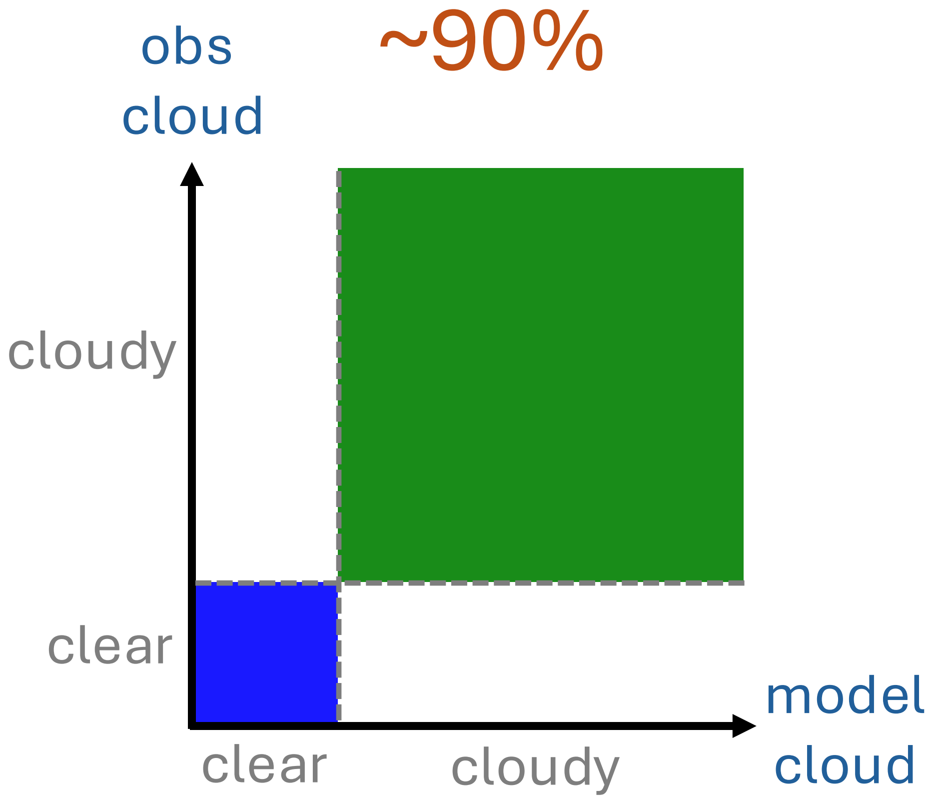


the “total effect” of turning off empirical inflation



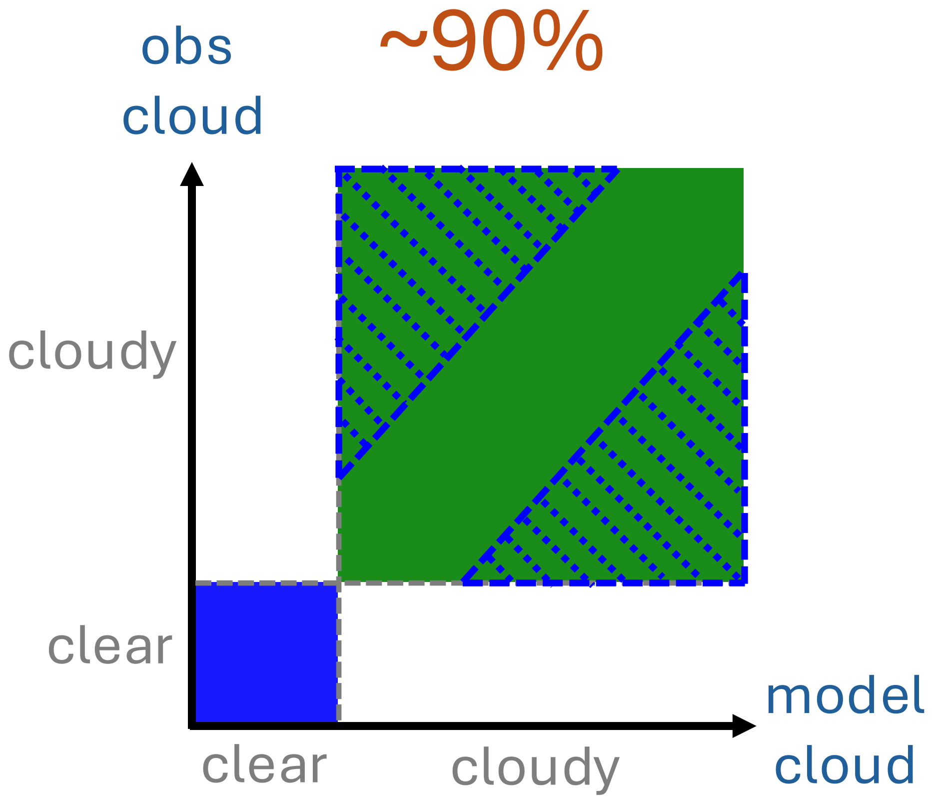
increase of the OBS weight (the stdev of OE)
⇒ change in VarBC
⇒ change in bias (the mean of OE)

Not all the assimilated observations go into VarBC in current GSI

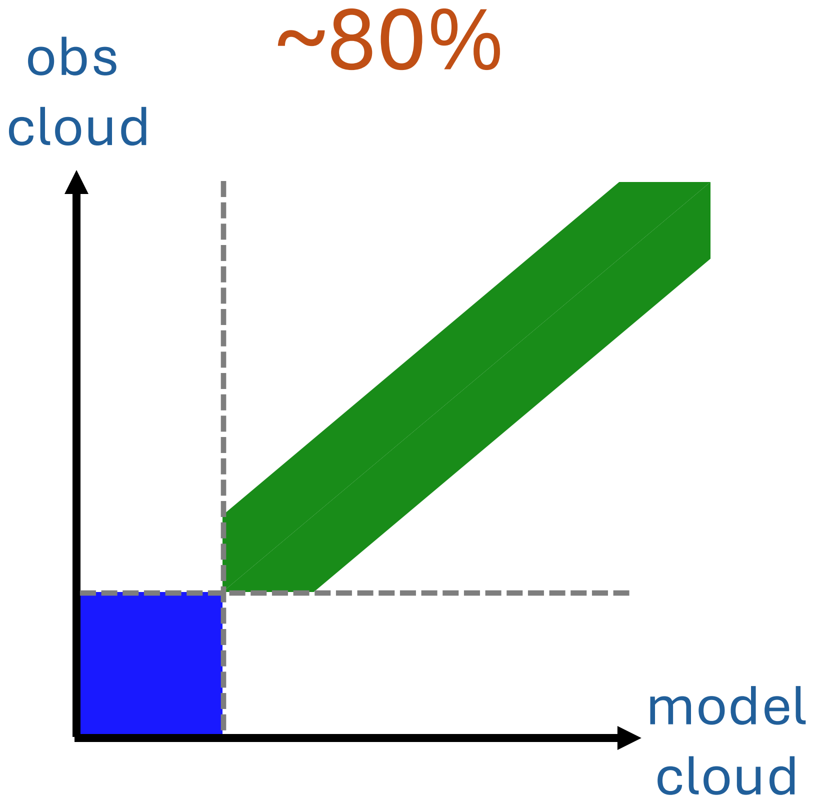


CTRL *(Zhu 2016; 2019)*

Not all the assimilated observations go into VarBC in current GSI



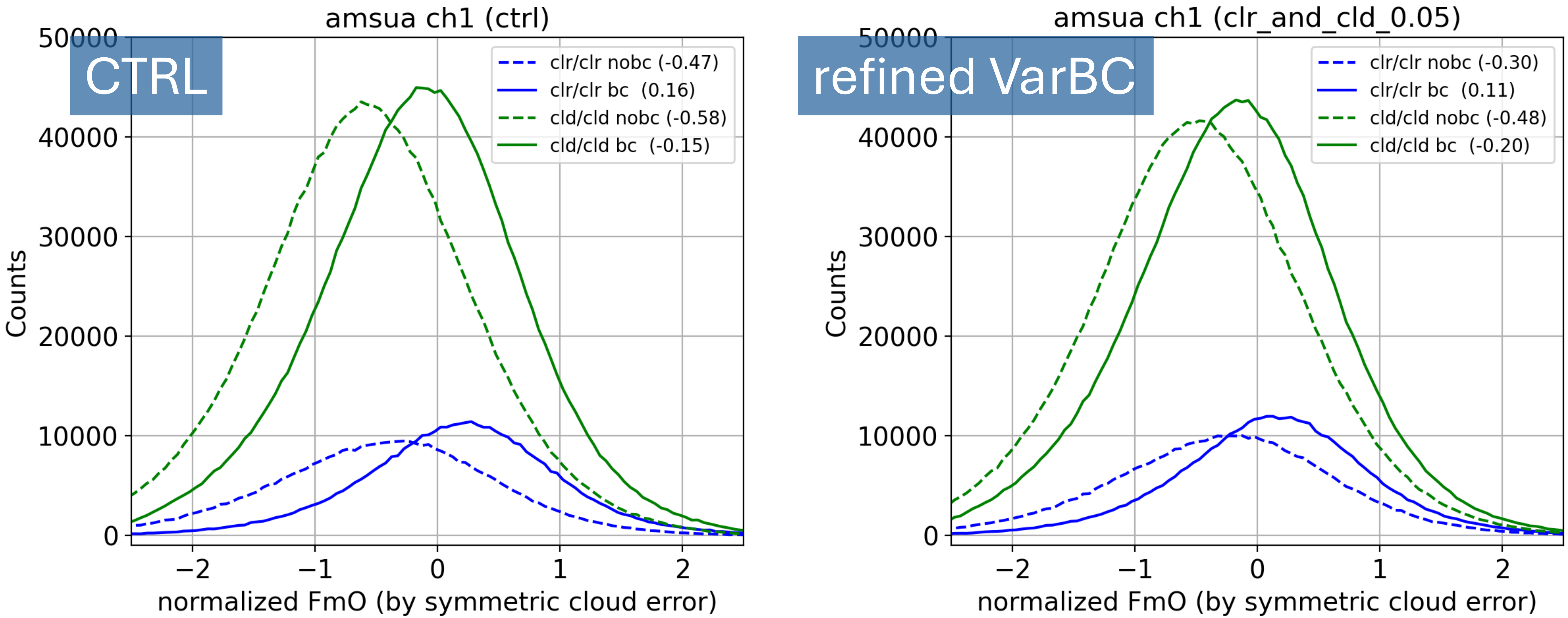
CTRL *(Zhu 2016; 2019)*



CLR_CLD_0.05

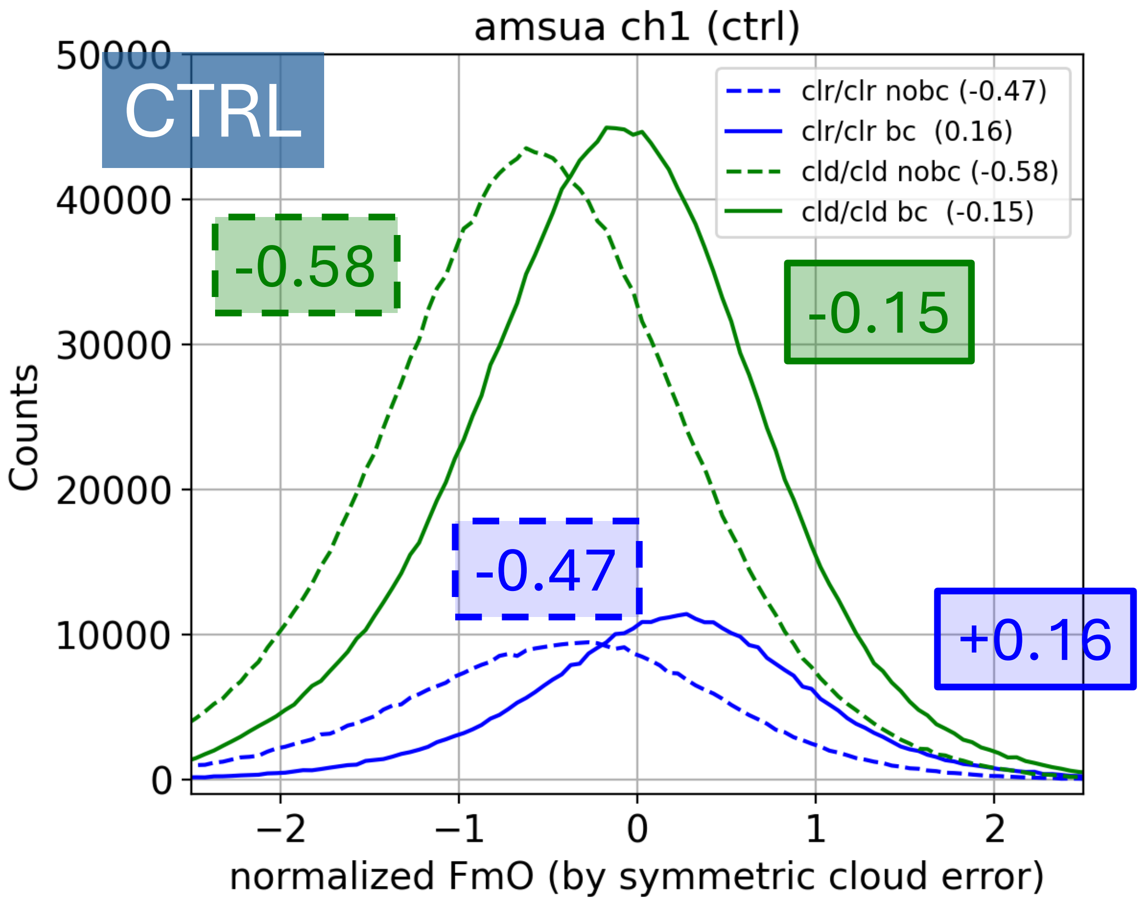
refined VarQC

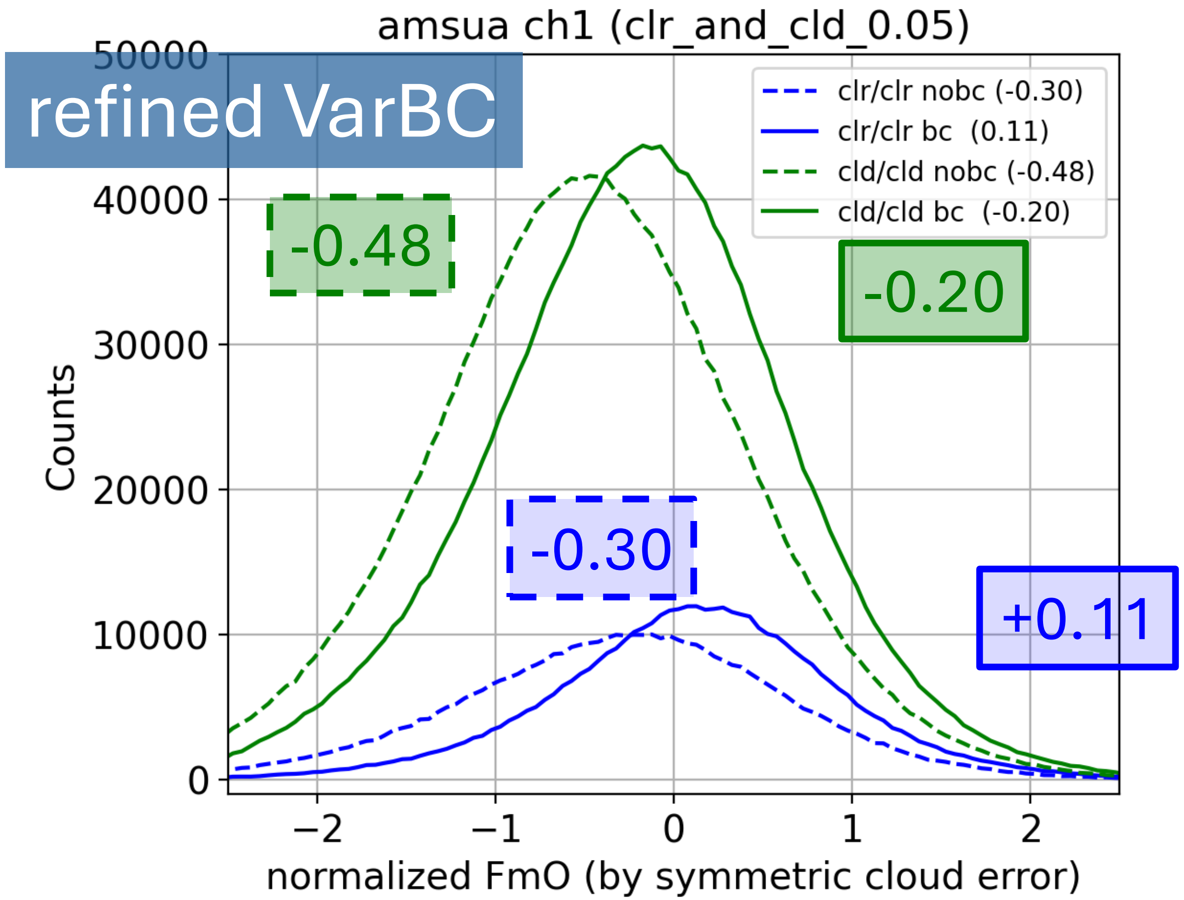
The sensitivity of VarBC to the empirical inflation (w/ empirical inflation)



colder model TB (dryer) warmer model TB (more humid)

The sensitivity of VarBC to the empirical inflation (w/ empirical inflation)

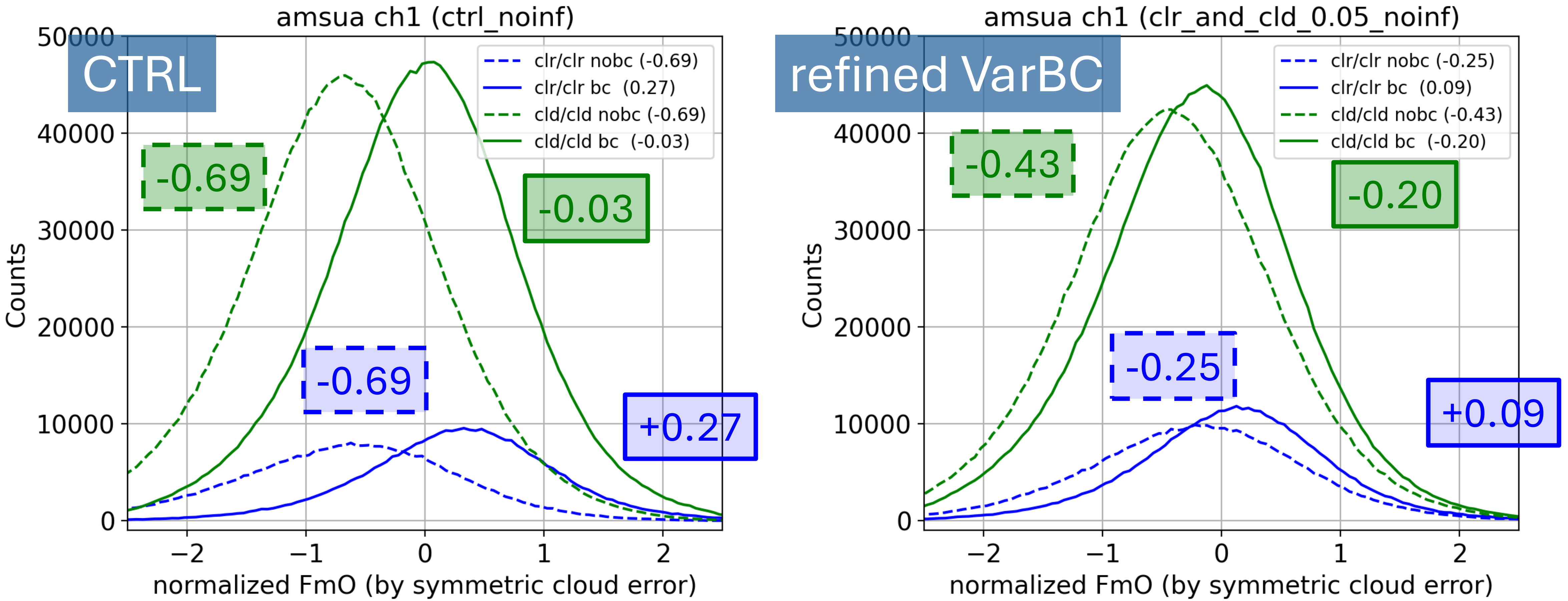




colder model TB (dryer) warmer model TB (more humid)

$\Delta = 0.43$ $\Delta = 0.63$ $\Delta = 0.28$ $\Delta = 0.41$

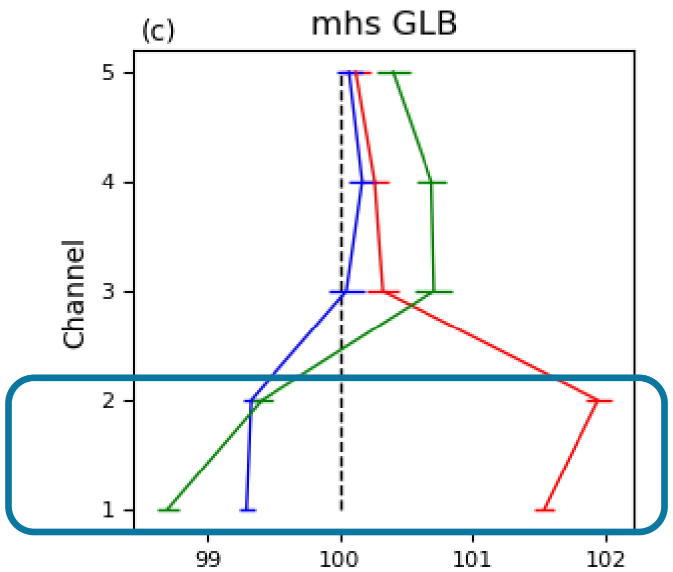
The sensitivity of VarBC to the empirical inflation (w/o empirical inflation)



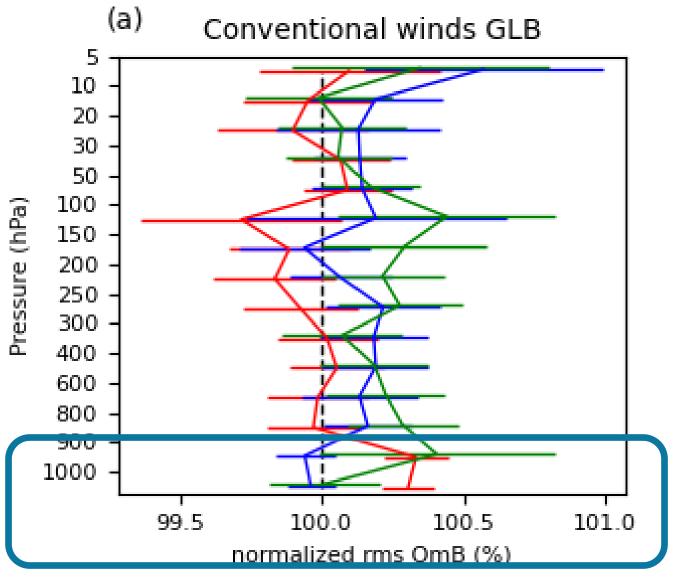
colder model TB (dryer) warmer model TB (more humid)

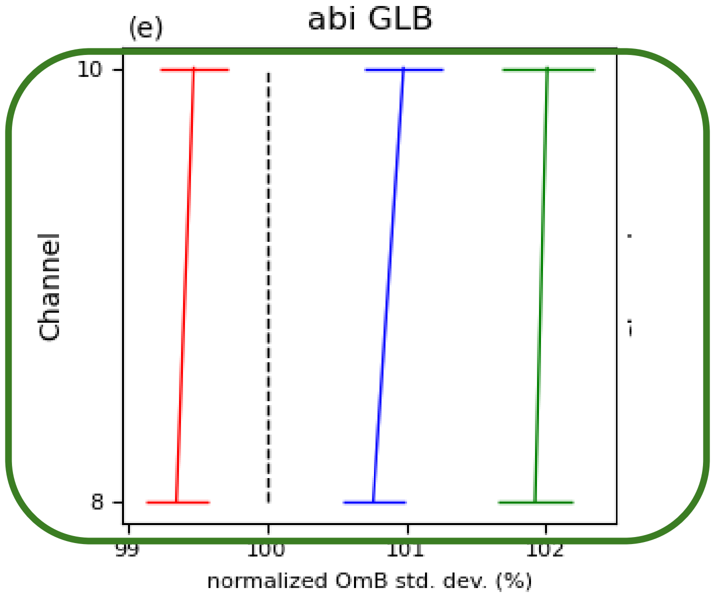
$\Delta = 0.66$ $\Delta = 0.96$ $\Delta = 0.23$ $\Delta = 0.34$

Improvement in low-level humidity, but mixed results for other variables



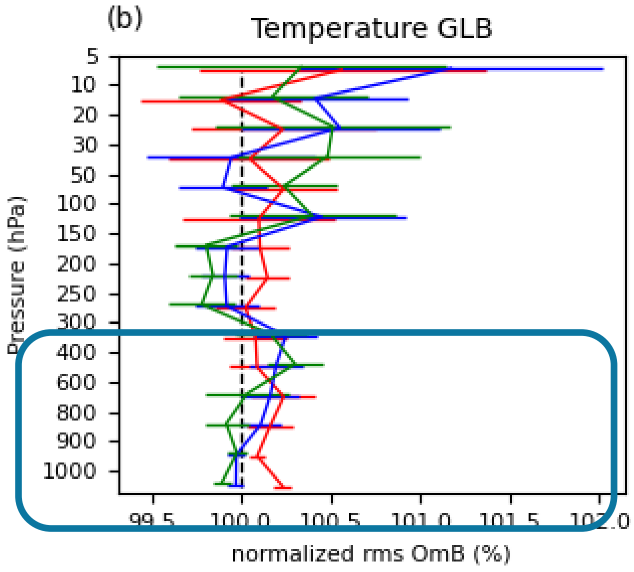
(near surface q,T)

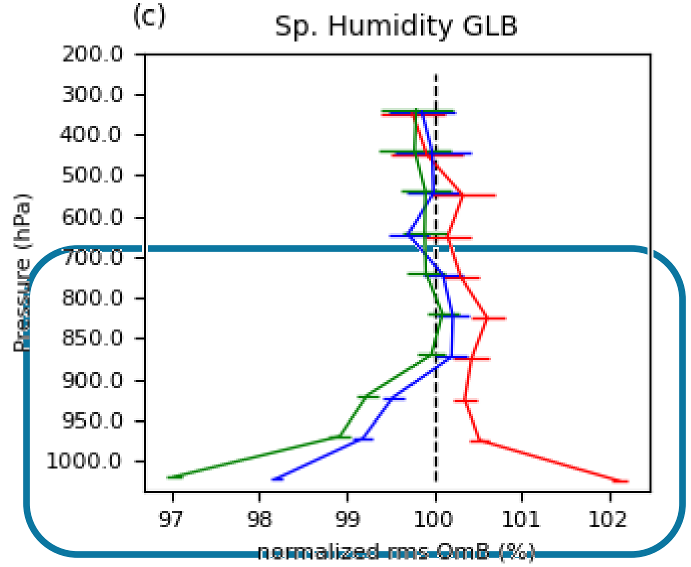




(700-500 hPa)

(300-200 hPa)





Conclusions

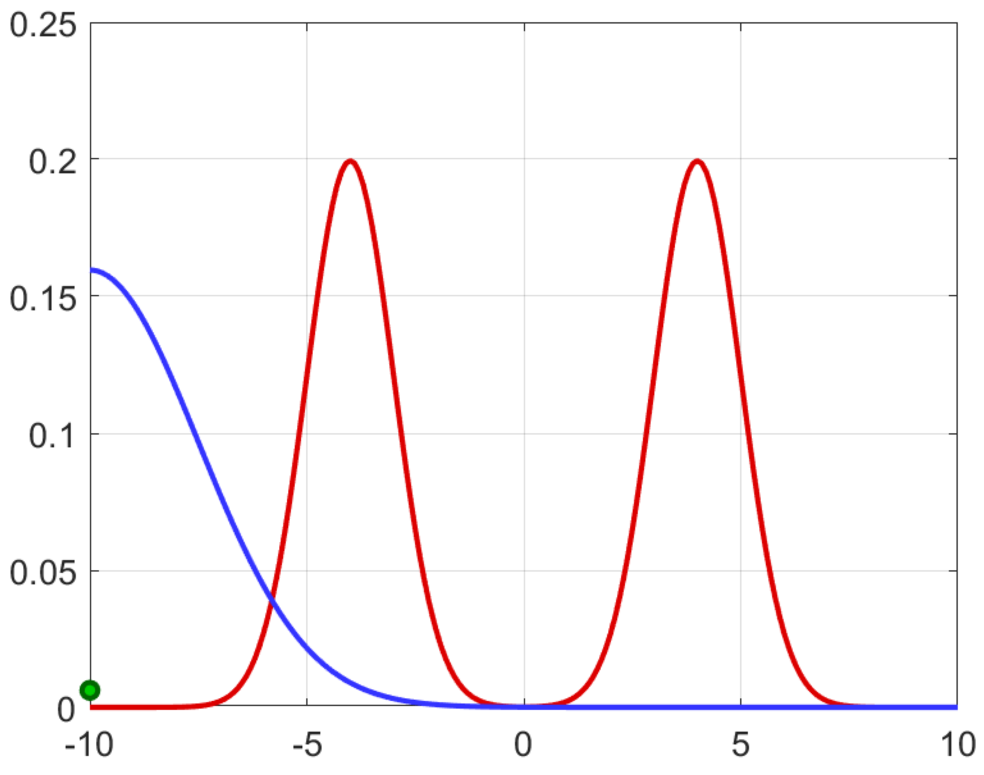
- In GSI, the empirical inflation significantly downweighs the all-sky AMSUA and ATMS observations, particularly for cloud mismatch observations.
- Turning off the empirical inflation changes the VarBC coefficients, whose effect dominates the effect of changing the standard deviation alone.
  - Mixed results, degradations in low-level dry biases
- The sensitivity of VarBC to empirical inflation is related to the data input to VarBC.
- A refined VarBC with a stricter control on cloudy data input shows much less sensitivity to the empirical inflation.
  - Improved low-level dry biases, while mixed results for other variables

Backup slides

DOEE can estimate the full distribution of the observation error using an ensemble

Hu, C. C., Van Leeuwen, P. J., & Geer, A. J. (2024). A non-parametric way to estimate observation errors based on ensemble innovations. Quarterly Journal of the Royal Meteorological Society

- **DOEE = Deconvolution-based Observation Error Estimation**



$$\begin{aligned} D &= Y - H(X) \\ &= [Y - H(X^{\text{truth}})] - [H(X) - H(X^{\text{truth}})] \\ &= \varepsilon^o - [H(X) - H(X^{\text{truth}})] \end{aligned}$$

$$f_D(d) = \int_{-\infty}^{\infty} f_{H(X^{truth})-H(X)}(d-\varepsilon_o) f_{\varepsilon_o}(\varepsilon_o)\, d\varepsilon_o$$

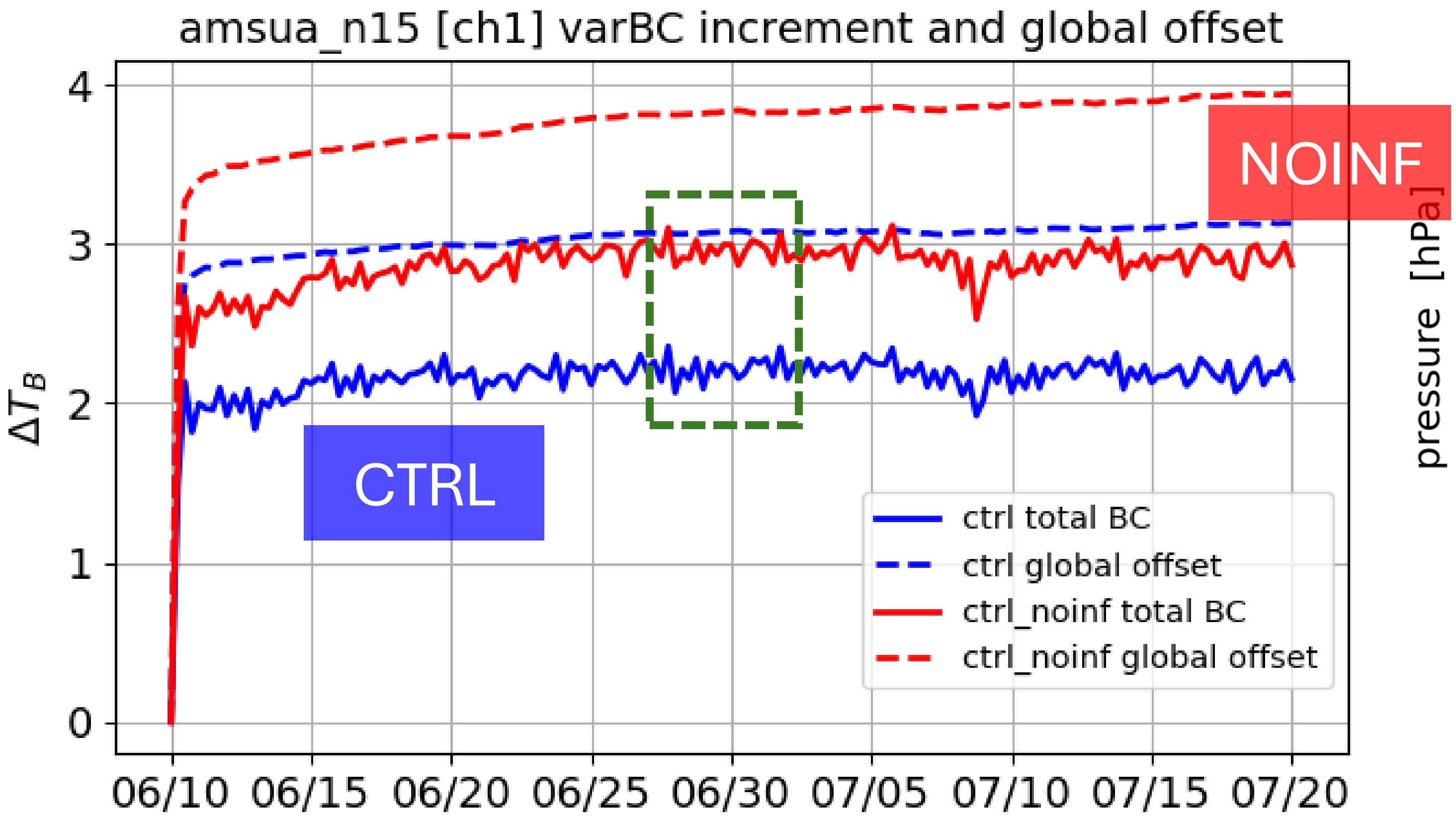
Smoothed pdf (input)

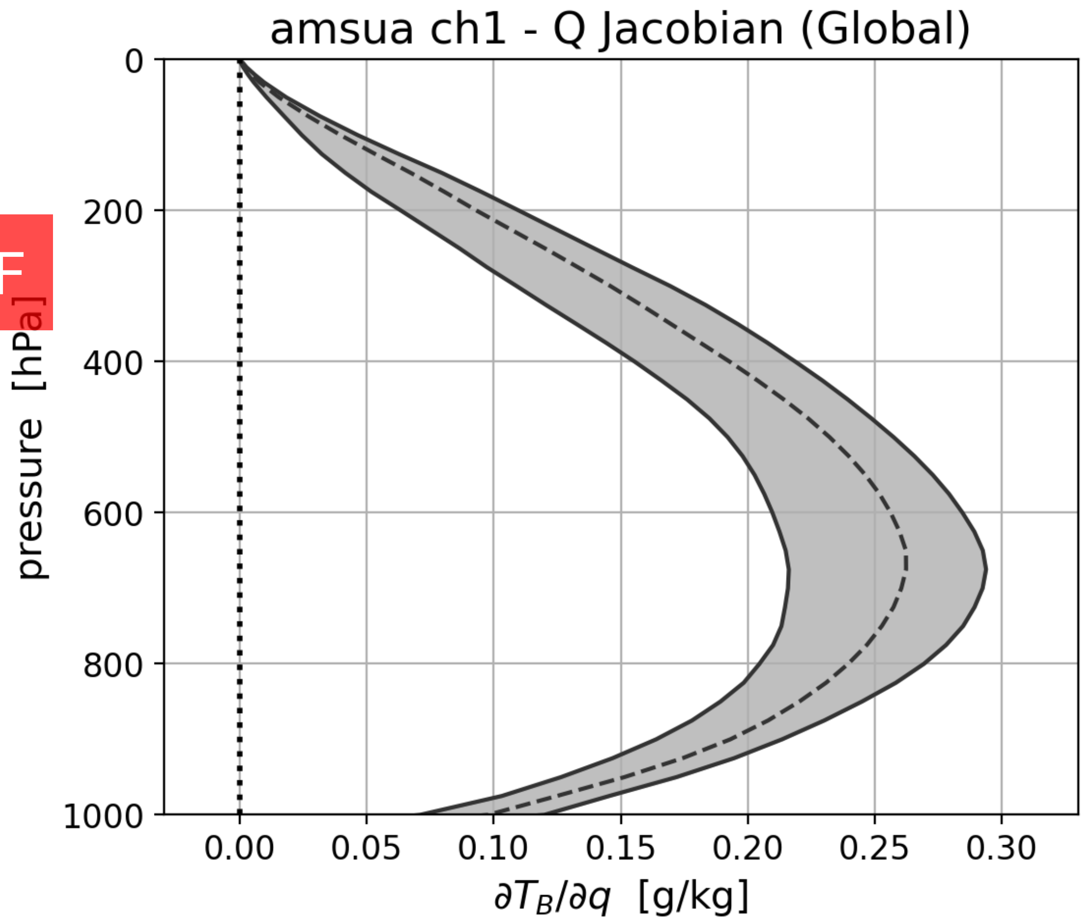
observation error pdf (output)

Assume each ensemble member and the truth are i.i.d. use the ensemble differences as samples (input)

Turning off the empirical inflation increases the VarBC increments for the AMSUA and ATMS window channels

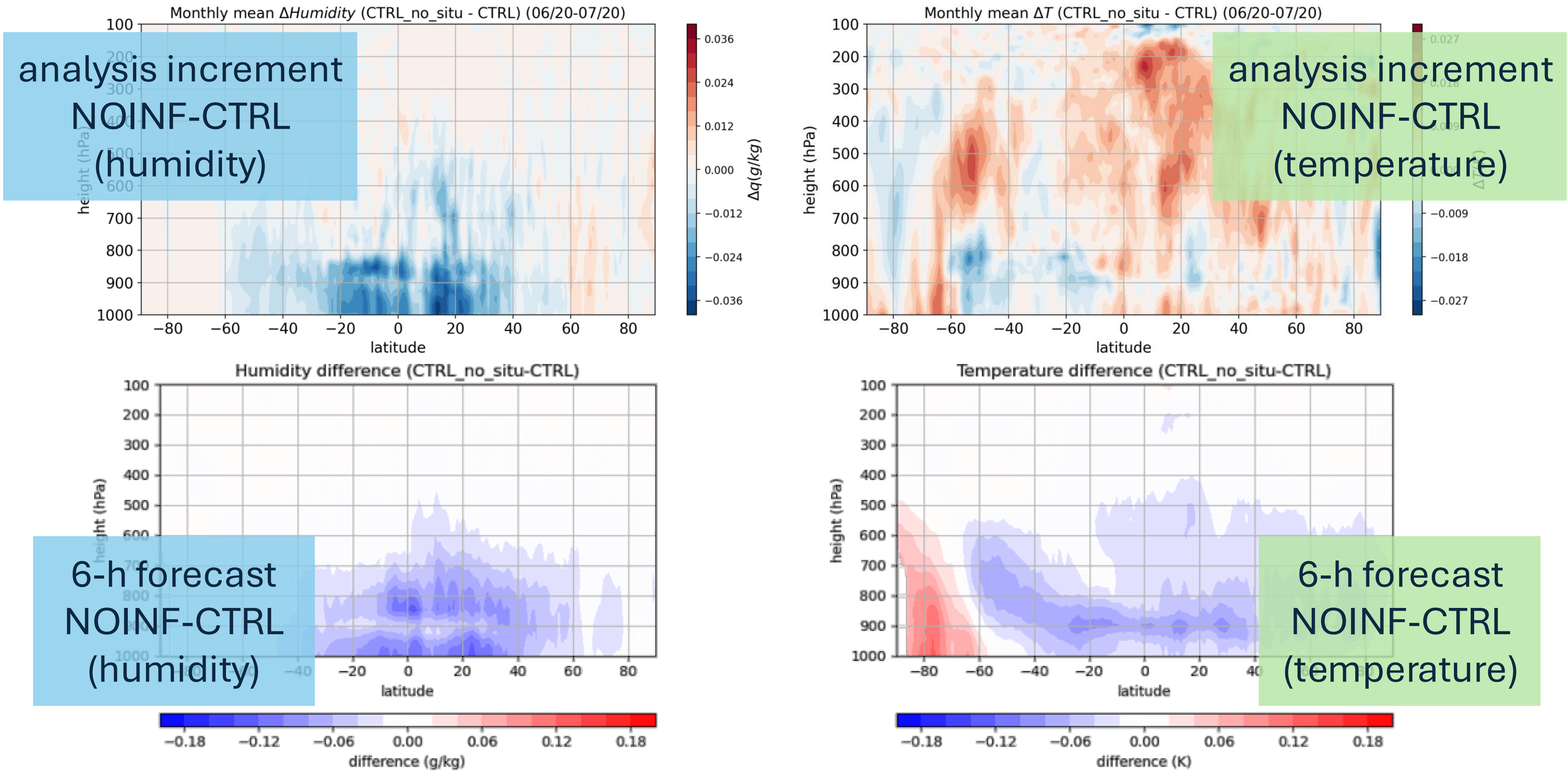
$$d = y - [H(x) + \Delta H_{VarBC}]$$

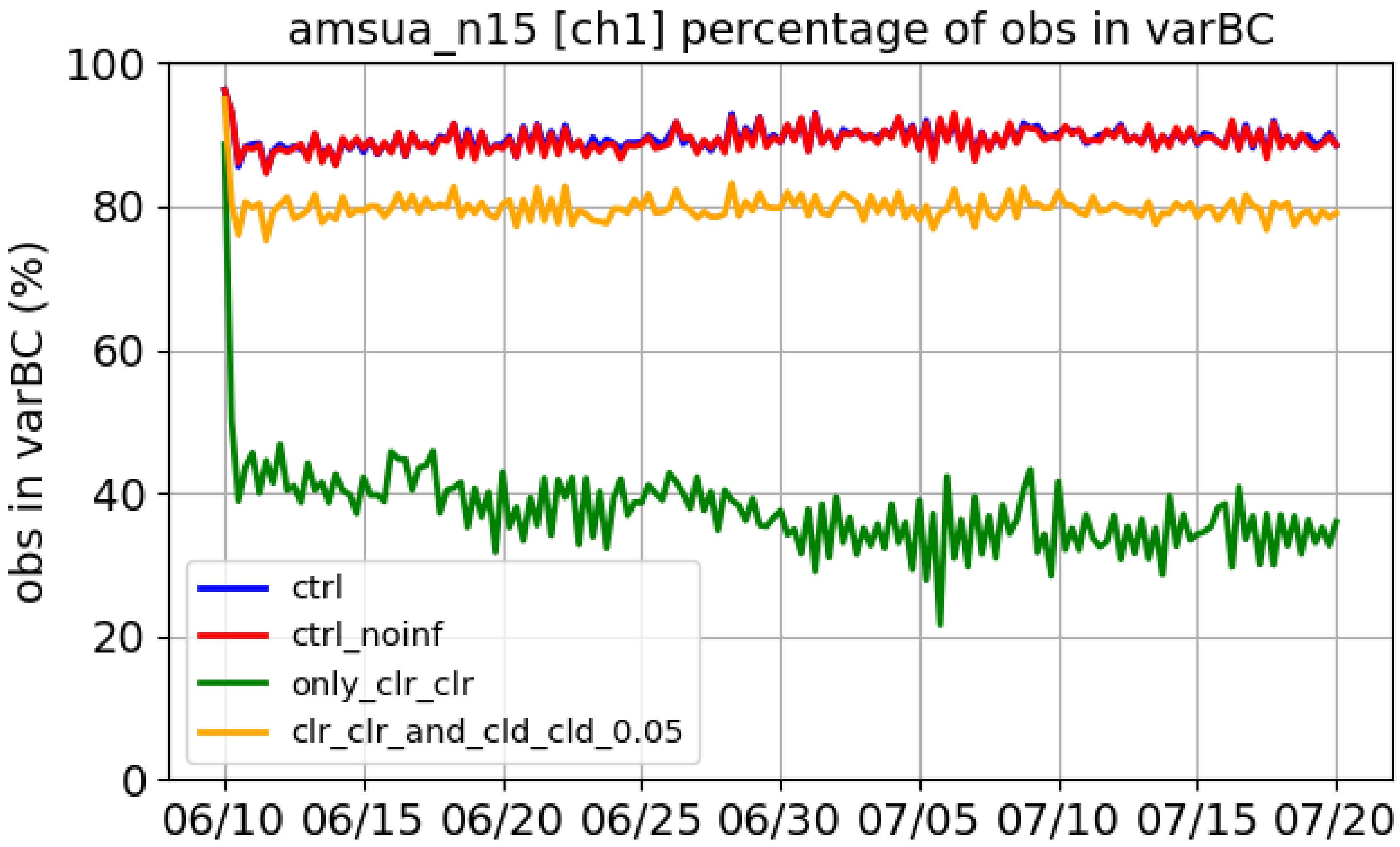




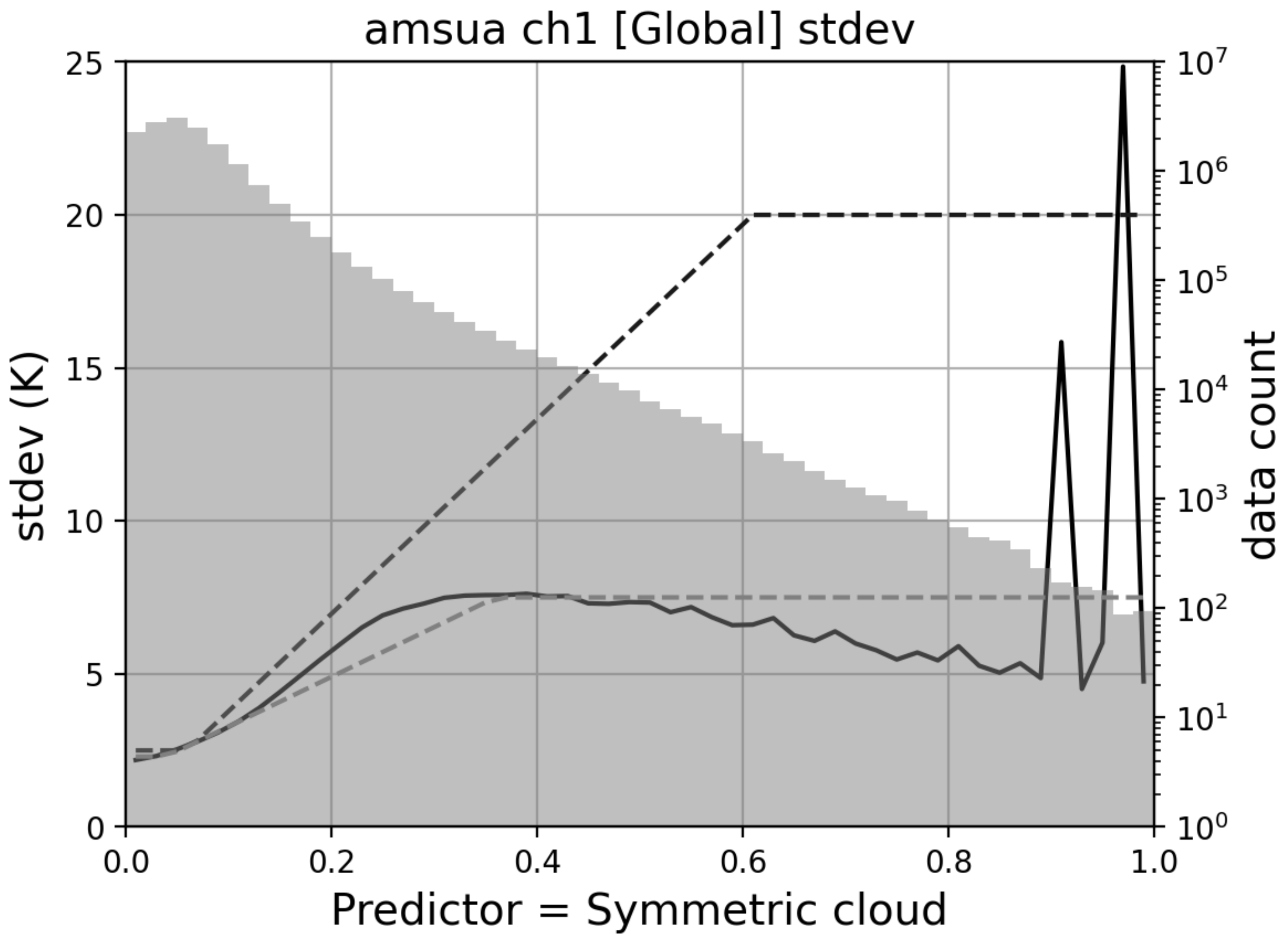
(Similar for channels: AMSUA ch1-3, 15, ATMS ch1-3, 16-17)

Direct impact from DA: introducing low-level dry biases





amsua_n15 [ch1] percentage of obs in varBC
obs in varBC (%)
100
80
60
40
20
0
06/10
06/15
06/20
06/25
06/30
07/05
07/10
07/15
07/20
ctrl
ctrl_noinf
only_clr_clr
clr_clr_and_cld_cld_0.05


amsua ch1 [Global] stdev
stdev (K)
data count
Predictor = Symmetric cloud
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7

evolving-Gaussian method examples