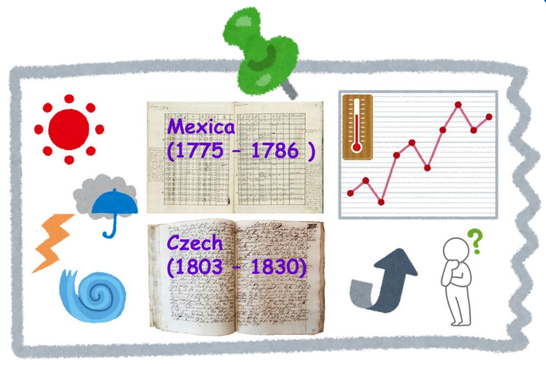
Impact of Gaussian transformation on cloud cover data assimilation for historical weather reconstruction

2024年5月24日　統計数理研究所　オープンハウス

王　小醒　　学際統計数理研究系　特任研究員

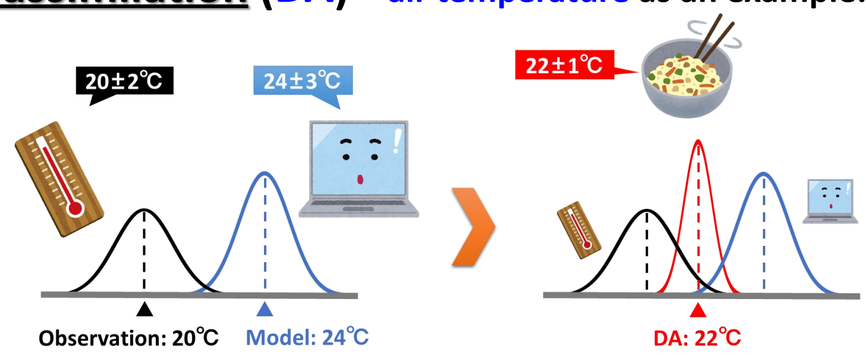
Key issue: How to get accurate weather data from descriptive information in ancient diaries?

- ✓ **Old diaries** record historical weather conditions.
- ✓ These records can be used to reconstruct weather **before instrumental measurements available**.
- ✓ **Data assimilation** is widely used because it can optimally combine observation with climate models.
- ✓ Cloud cover can be converted from the descriptive records but is difficult to assimilate because of its **non-Gaussian characteristic**.



- Method -

◆ Data assimilation (DA) – air temperature as an example:



Local ensemble transform Kalman filter (LETKF)

Kalman gain matrix　　observation operator

$$x_a = x_b + BH^T(R + HBH^T)^{-1}(y - Hx_b)$$

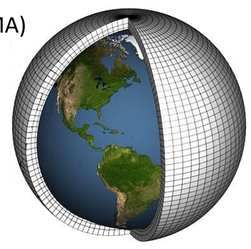
x_a: analysis state vector; x_b: model background state vector; R: observation error covariance matrix; B: background error covariance matrix; y: observation vector

R: {total cloud cover

B: { U, V, T, Q, cloud, precipitation, surface pressure, solar radiation, etc.

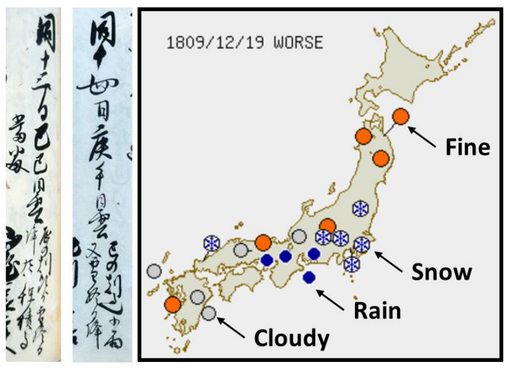
◆ Model – Global spectral model (GSM):

(Source: JMA)



- **Resolution:**
  6-hour; 192[lon]*94[lat]*28[level]
- **Variables:**
  3D: wind speed, air temperature, humidity
  2D: total cloud cover, solar radiation, surface pressure, precipitation

◆ Observation – Historical weather database (HWDB)



(Source: Hirosaki office diary)
Accessible from: http://tk2-202-10627.vs.sakura.ne.jp

Descriptive weather information:
"Fine", "cloudy", "rain", "snow", etc.

[Ichino et al., 2001]

Weather **category**:
1-sunny, **2**-cloudy, **3**-rainy

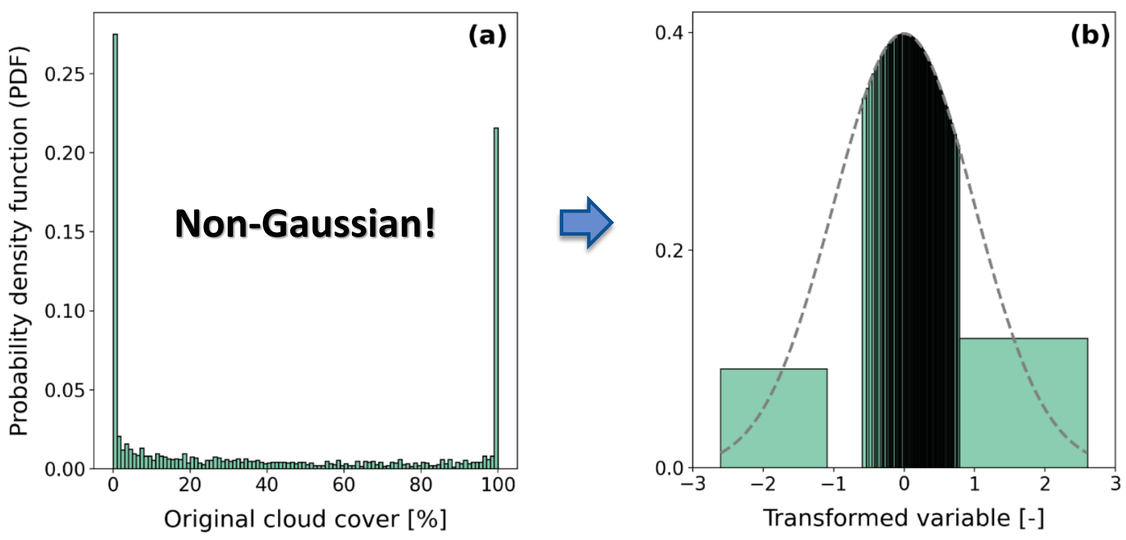
[Toride et al., 2017]

Total cloud cover: **0% - 100%**

- **Descriptive diary-based records** are firstly grouped into **three weather categories**, and then converted into the number of **total cloud cover values**.

◆ Gaussian transformation (GT)

$$\text{cloud}_{\text{trans}} = G^{-1}[\text{CDF}_{\text{orig}}] = \sqrt{2}\,\text{erf}^{-1}(2*\text{CDF}_{\text{orig}}-1)$$



- **GT** process is based on the **equivalent relationship of cumulative distribution functions** (CDFs) between the original total cloud cover and the transformed variable.

- Results -

◆ Performance evaluation in the idealized case

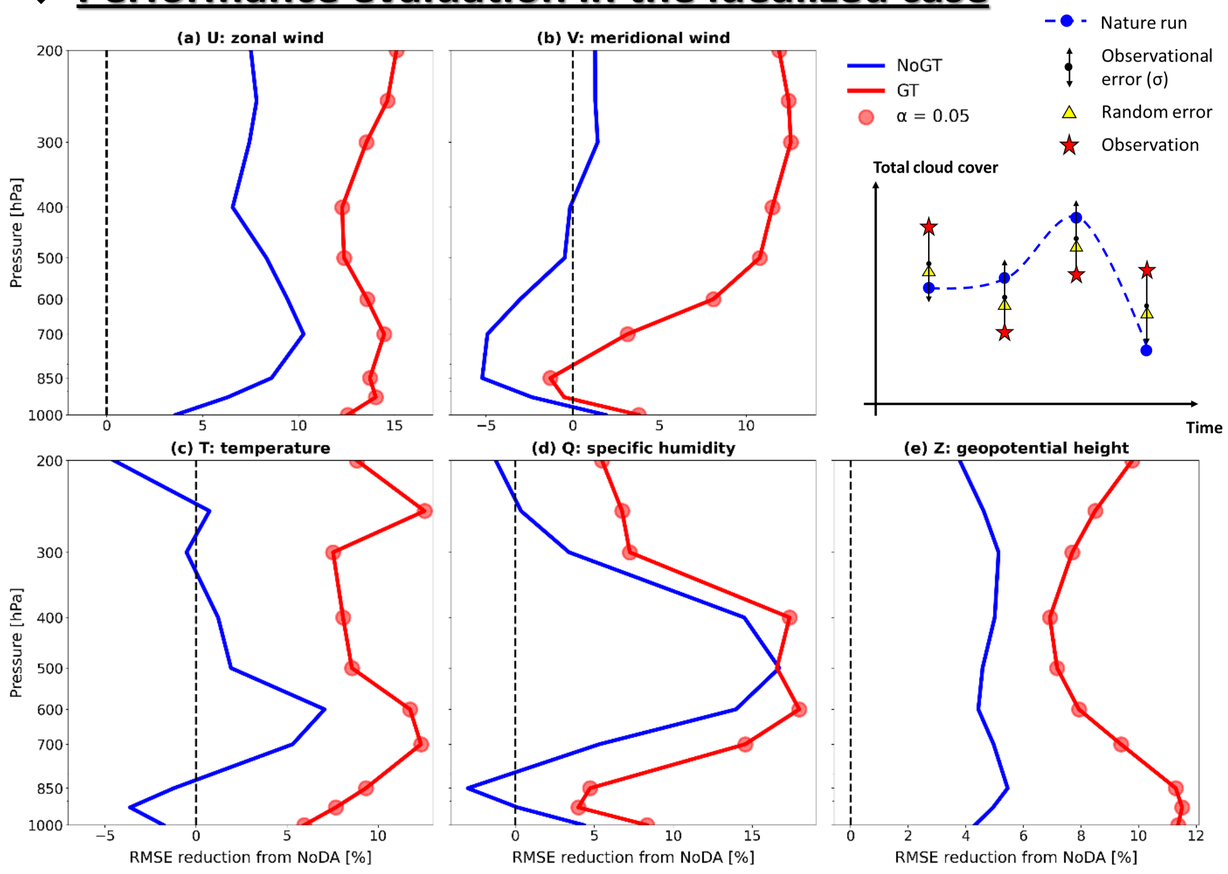


Figure1: 2-month root-mean-square error (RMSE) improvement profiles.

- Assimilating the Gaussian transformed cloud cover **improves RMSEs of both two- and three-dimensional variables**, showing the potential of GT to improve cloud cover assimilation.

◆ Practical reconstruction at the global scale

- This experiment assimilates cloud cover at sparsely distributed grid points **on a global scale**. Results indicate that the **impact of GT is confirmed over a wide region other than the observation sites**. In addition to Japan, diaries that record weather phenomena are distributed globally. This demonstrates the **potential of GT for global-scale historical weather reconstruction**.

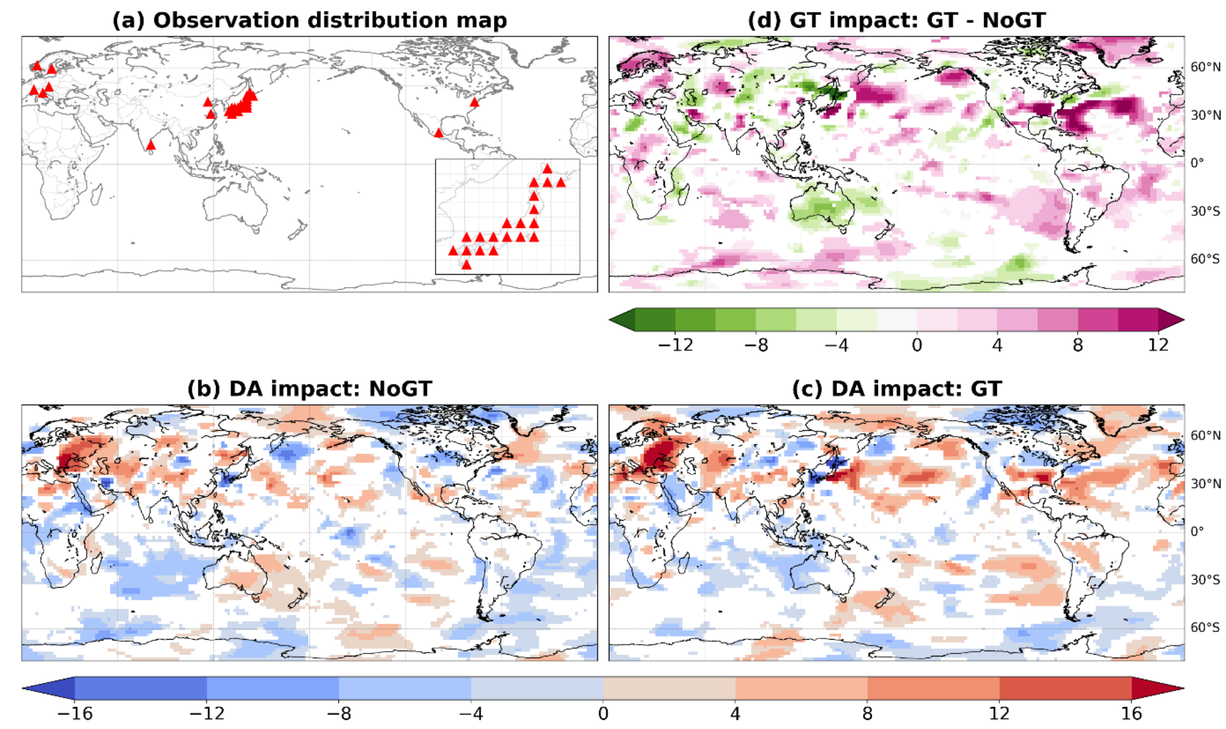


Figure 2: Horizontal distributions of DA impact and GT impact on 850 hPa temperature estimation. All colored grid points show statistical significance at the 1% level.

◆ Conclusion

- **Gaussian transformation (GT)** transforms the original cloud cover into a normal distribution shape, improving cloud cover assimilation accuracy. Results demonstrate the potential of GT in **high-resolution historical weather reconstruction** using old descriptive diaries.
- For more information, please also refer to https://doi.org/10.1175/MWR-D-22-0315.1.