Research collaborators for large ensemble simulations

Empirical optimal vertical localization derived from large ensembles

トビアス　ネッカー **Dr. Tobias Necker** - Universität Wien, Vienna, Austria   tobias.necker@univie.ac.at

Co-authors:

P. Griewank — Universität Wien, Vienna, Austria

M. Weissmann — Universität Wien, Vienna, Austria

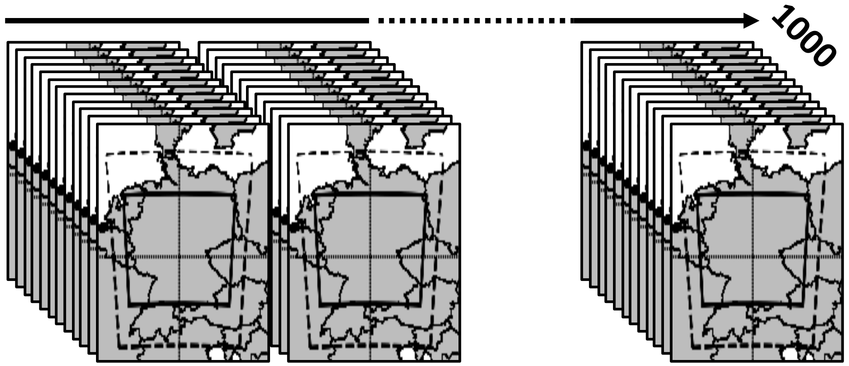
T. Honda — Hokkaido University, Sapporo, Japan

T. Miyoshi — RIKEN Center for Computational Science, Kobe, Japan

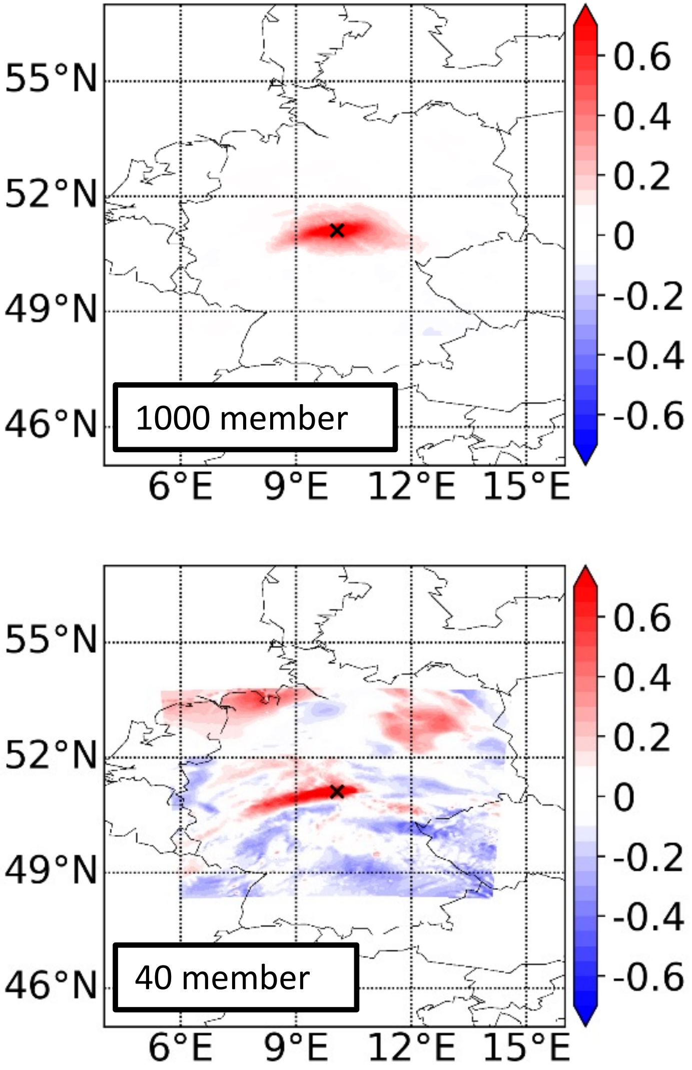
ISDA Bologna 2023

Outline

1. Empirical Optimal Localization (EOL) method
2. Results
   a. Optimal model space covariance localization
   b. Optimal gain localization for satellite observations
3. Take-home messages



Motivation



Example for spurious correlations on convective scales (Correlation of T2m to T2m)

Data assimilation heavily relies on accurate error covariance estimates

$$K = P^b H^T \left(H P^b H^T + R \right)^{-1}$$

Spread and weight observation information

Sampling errors in error correlations pose severe issue for data assimilation requiring localization

$$cov(\Delta x^b, \Delta y^b) = r(\Delta x^b, \Delta y^b)\, \sigma(\Delta x^b)\, \sigma(\Delta y^b)$$

Different reasons why we need localization?

1. **Sampling errors**
   (e.g., spurious correlations)
2. **Rank deficiency**
   (e.g., regularization / matrix inversions)
3. **Algorithm efficiency**
   (e.g., degrees of freedom in LETKF)

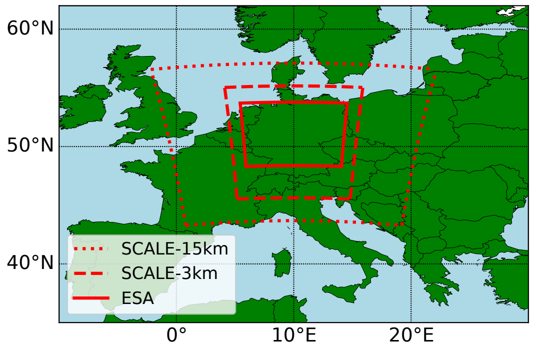
Motivation: Better understanding of error covariances and localization

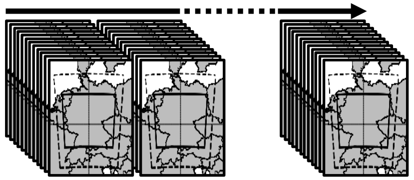
Our approach for fundamental research:

- Apply large ensemble to study sampling errors and improve localization
- 1000-member ensemble **(Necker et al. 2020)** provides background ensemble forecasts over Germany with 3km horizontal resolution and 30 vertical levels

Experimental setting:

- We assume 1000-member correlation as truth (r^{1000})
- Subsampling strategy: 25 40-member subsample correlations (r^{40})
- First guess: 3h lead time background convective-scale forecasts





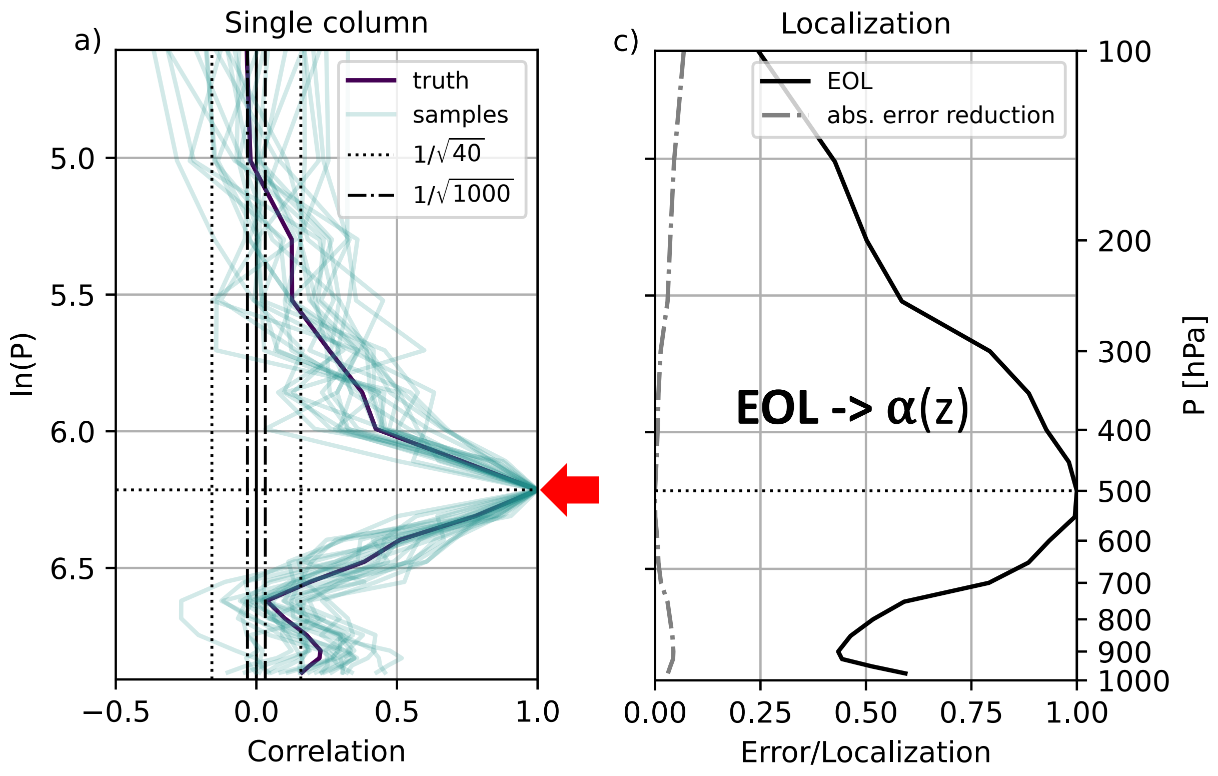
Main research questions:

a) How can we improve state-of-the-art vertical localization approaches?

b) How to achieve positive-definiteness of derived covariance and localization matrices?

c) How should we localize vertical error correlations of non-local visible and infrared satellite observations?

Empirical Optimal Localization (EOL) – Example for single vertical column



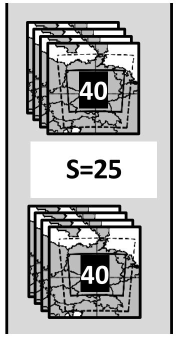
Goal: Derive an optimal localization from 1000-member ensemble using subsampling strategy

Strategy: Minimize the MSD between sample correlations and the 1000-m correlation truth using cost function J

$$J(\alpha) = \sqrt{\sum_{s=1}^{S} \left(\alpha r_s^{40} - r^{1000}\right)^2}$$

Solving for α yields simple formula for optimal localization:

$$\rightarrow \quad \boxed{\alpha_{single} = \frac{\sum_{s=1}^{S} r_s^{40} r^{1000}}{\sum_{s=1}^{S} (r_s^{40})^2}}$$



α : Optimal localization factor

r^{40} : 40-member ensemble correlations

r^{1000} : 1000-member ensemble correlation

S : Number of 40-member subsamples

Single column example:

- Temperature to temperature sample correlations
- Reference level 500hPa to all other vertical levels

How does the Empirical Optimal Localization (EOL) method work?

Properties of the empirical optimal localization (EOL) method (see Necker et al. 2023)

- Requires large ensemble as truth
- Subsampling allows deriving optimal localization
- Optimizes the covariance (not necessarily the analysis)
- Approach can easily be adapted to estimate optimal localization directly for Kalman gain K

EOL can be computed for single correlations or correlations grouped in batches of interest:

Option 1:

$$\alpha_{single} = \frac{\sum_{s=1}^{S} r_s^{40} r^{1000}}{\sum_{s=1}^{S} (r_s^{40})^2}$$

Optimal localization for single correlations

- Ideal to study situation-dependence or variability

Option 2:

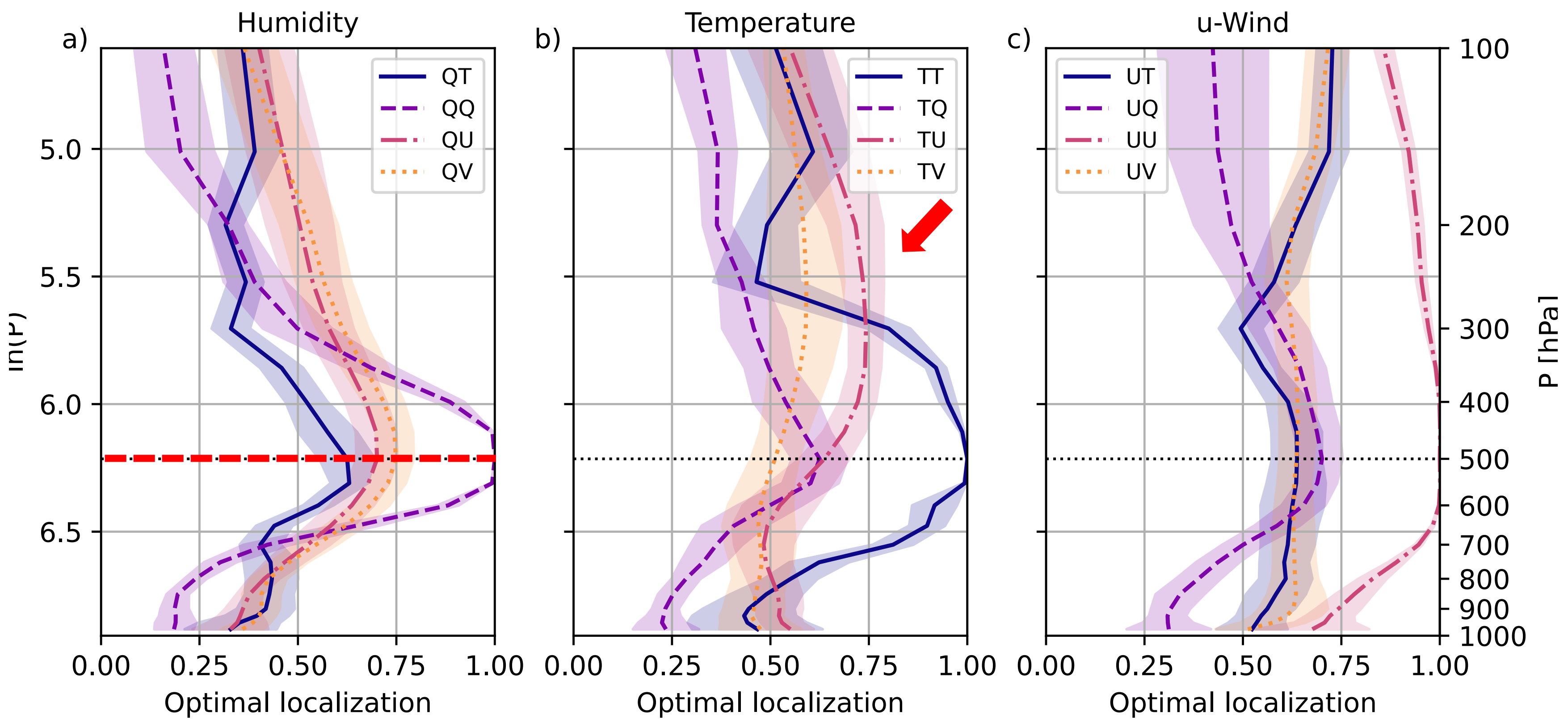
$$\alpha_{batch} = \frac{\sum_{s=1}^{S} \sum_{k=1}^{K} r_{s,k}^{40} r_k^{1000}}{\sum_{s=1}^{S} \sum_{k=1}^{K} (r_{s,k}^{40})^2}$$

Optimal localization for grouped correlations in batches defined by specific criteria

- Allows deriving: Domain-uniform, variable-, observation- or situation-dependent localization
- K: Number of combined correlations (e.g., variables, temproally, vertically, horizontally, …)

How should an optimal domain-uniform vertical localization look like?

T: Temperature / Q: Spec. humidity / U: Zonal wind / V: Meridional wind



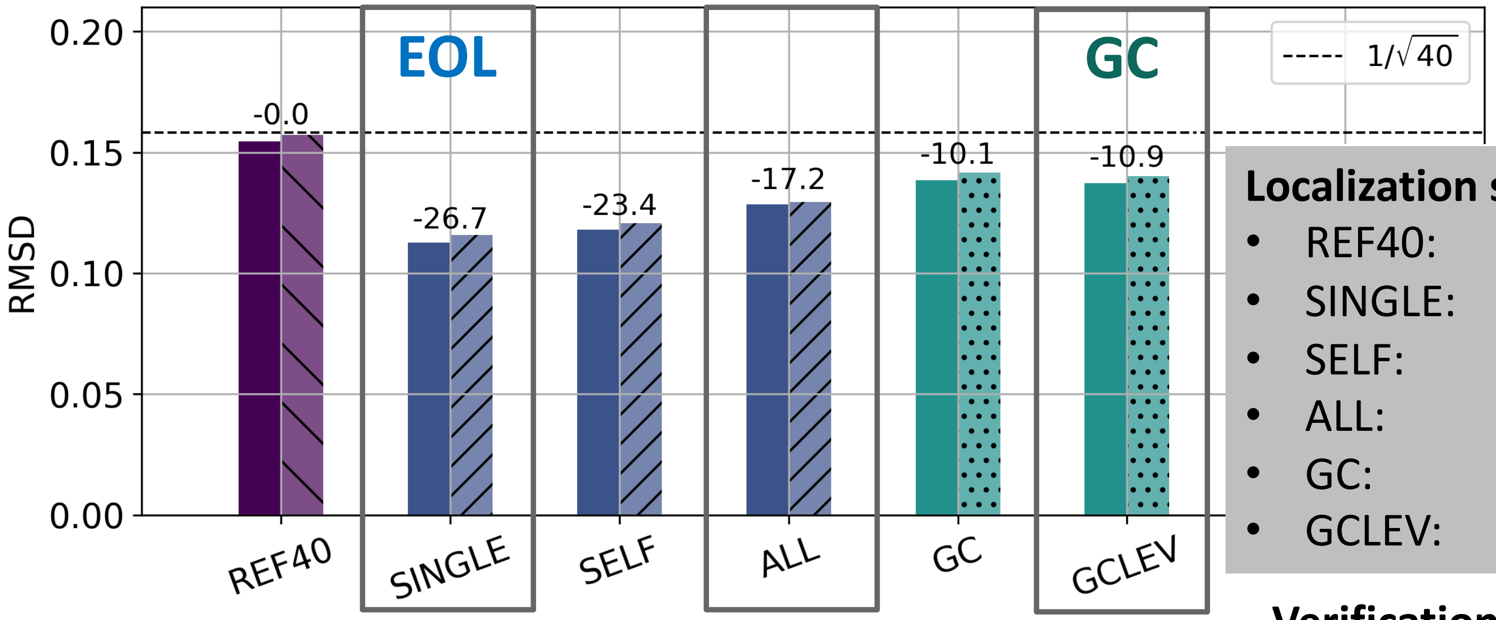
EOL for 500hPa

- Single variable pairs
- **Domain-uniform / non-adaptive**
- Shading: day-to-day variability
- **Reference level 500hPa**

Conclusions

- Substantially different localization scales and shapes for humidity, temperature, and wind
- Self- & cross-correlations behave systematically different
- Maximum of localization function not necessarily at reference level

Estimated error reduction (%) for different localization approaches



Configuration: Domain-uniform localization verified for two independent forecasts

Localization settings

- REF40: Reference without localization
- SINGLE: Variable-dependent EOL
- SELF: Self/cross correlation-dependent EOL
- ALL: One EOL function for all situations
- GC: Optimally tuned Gaspari-Cohn (GC)
- GCLEV: Height-dependent optimally tuned GC

Verification: Against 1000-member correlation truth

Conclusions

- GC vs ALL: Localization functions (shapes) can be improved compared to GC
- GC vs SINGLE: Variable or correlation dependent EOL outperform common methods by factor two

→ Applying better localization functions and treating variable pairs differently bears substantial room for improvements

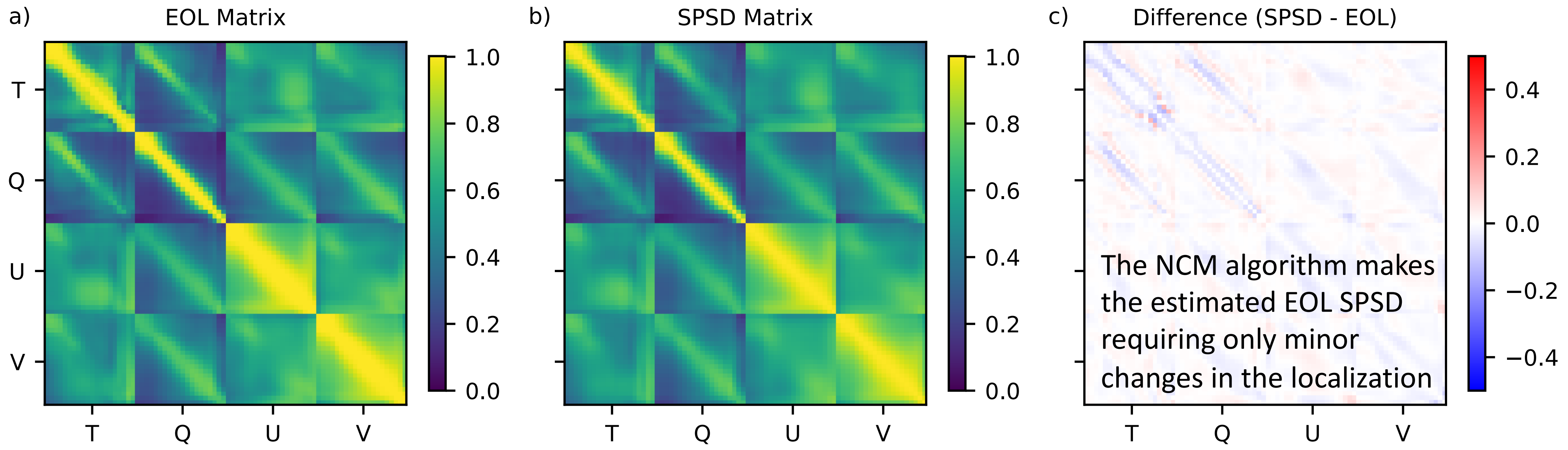
Does the EOL yield a symmetric positive semi-definite (SPSD) localization?

Localized cov-matrix: Needs to be symmetric positive semi-definite (PSD) to ensure a unique optimal analysis

Problem: Localization methods do not necessarily yield a positive semi-definite covariance matrix

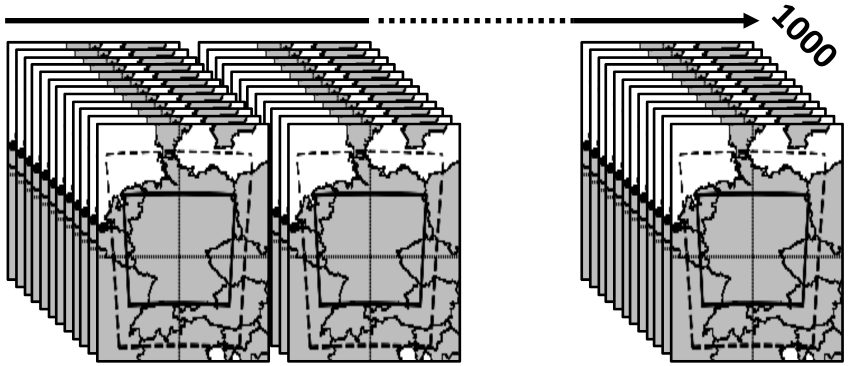
<u>We tested different methods for achieving PSD of localization:</u>

1. **Fitting suitable localization functions such as GC:** GC ensures PSD by function design
2. **Regularization** or **eigen-decomposition:** Requires matrix manipulation to mitigate inflation of variances
3. **Nearest Correlation Matrix (NCM) algorithm** (Higham 2002): great results -> Potentially many use-cases in DA/ML

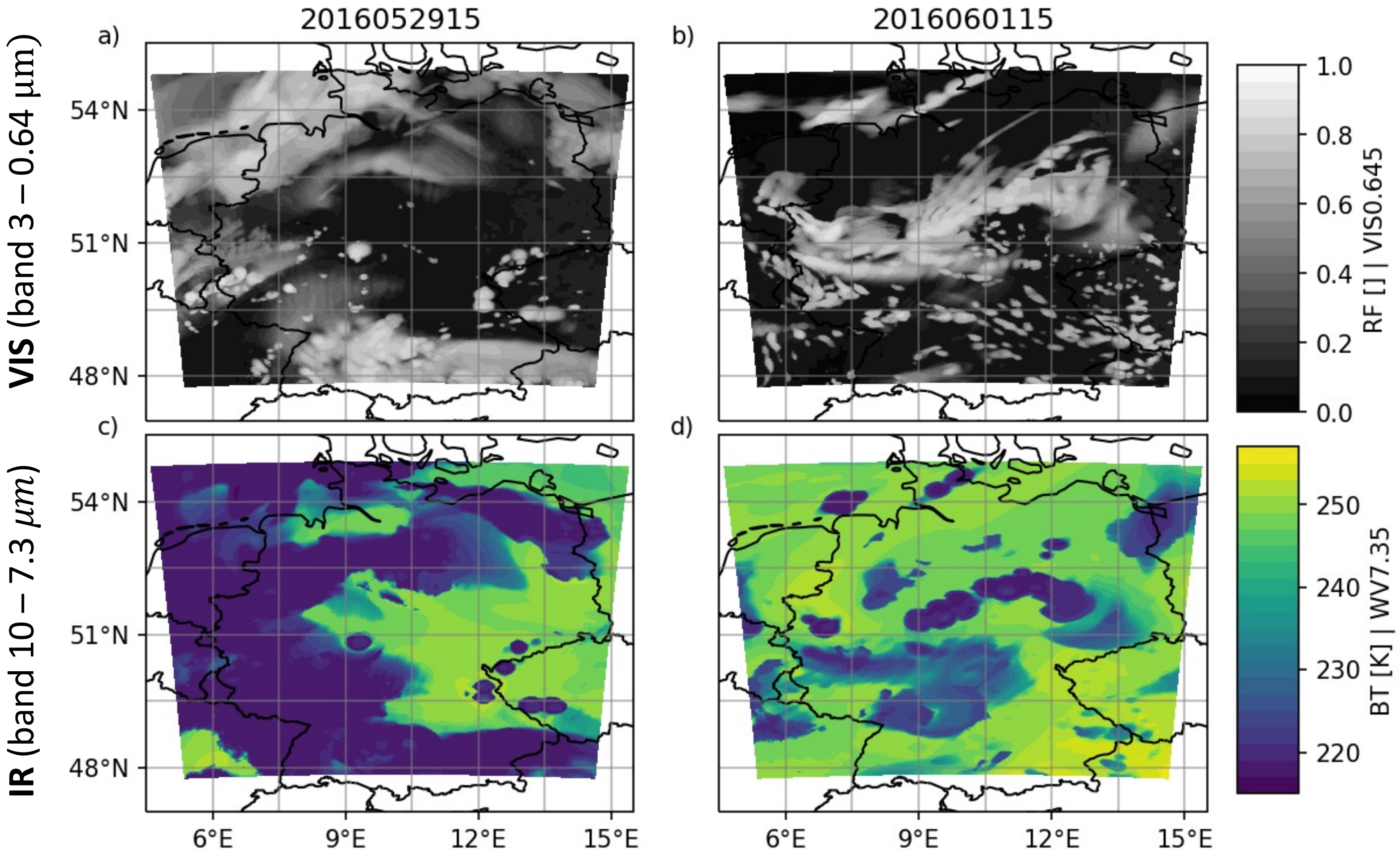


Outline

1. Empirical Optimal Localization (EOL) method
2. Results
   a. Optimal model space covariance localization
   b. Optimal gain localization for satellite observations
3. Take-home messages



1000-member ensemble: Synthetic satellite images based on RTTOV



Setup:

- 2-days in May/June 2016, 15UTC
- 1000-member correlation as truth: r^{1000}
- Localization for 40-member ensemble assuming EAKF or LETKF DA

Today: Two channels of Himawari-8

- **IR** (band 10 – 7.35 μm)
- **VIS** (band 3 – 0.64 μm)

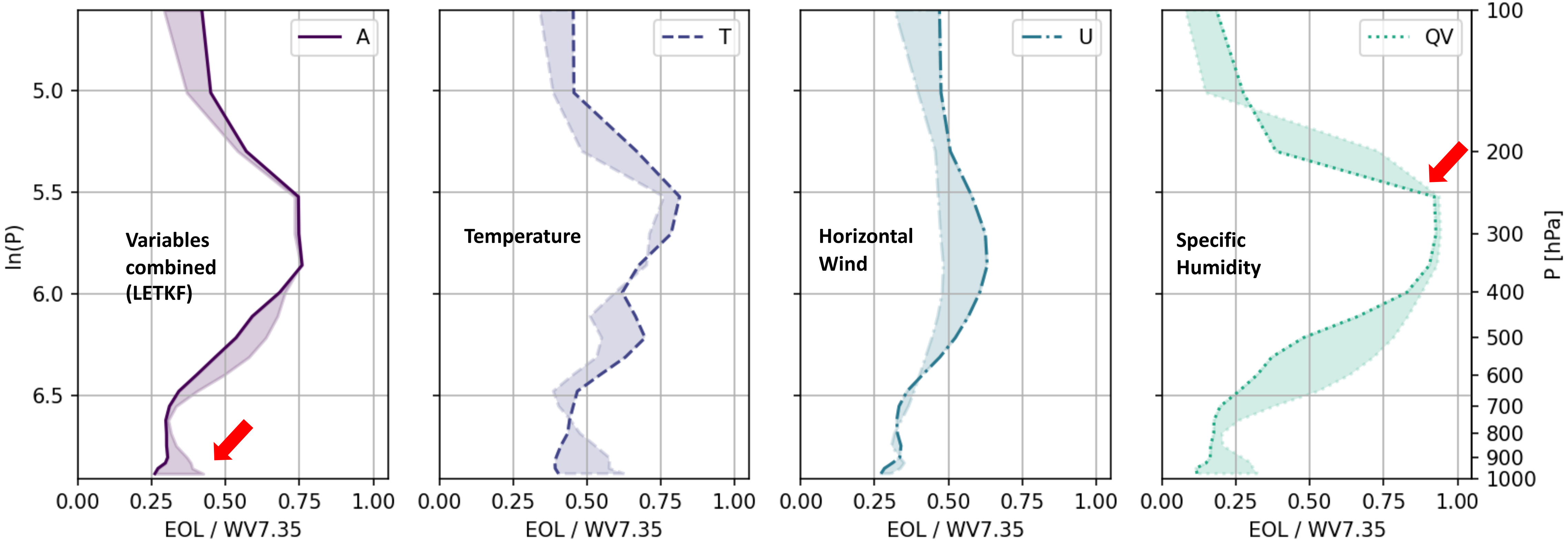
$$\Delta x = \alpha \, r \, \Delta y$$

Goal: *Localization α of error correlation r from satellite model equivalent y to state variable x*

Thanks to T. Honda for computing the synthetic satellite images

Domain-uniform EOL / Infrared 7.35µm channel

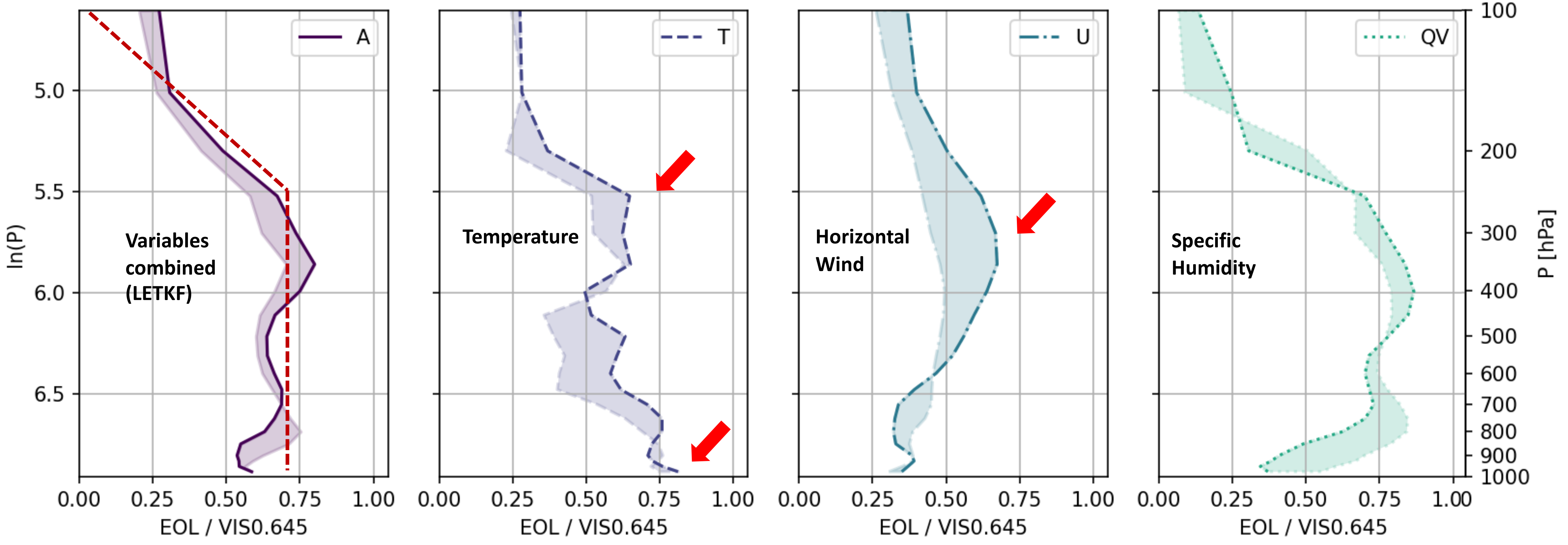
→ **Domain-uniform EOL mainly peaks between 250-400 hPa; Humidity revelas asymmetric localization shape**



IR (band 10 – 7.35 μm) | Solid = Case 1 | Dashed = Case 2 | Shading = Variability

Domain-uniform EOL / Visible 0.64 µm channel

→ **Localization exhibits multiple peaks; VIS 0.6µm strongly correlated with T in lower troposphere / surface levels**

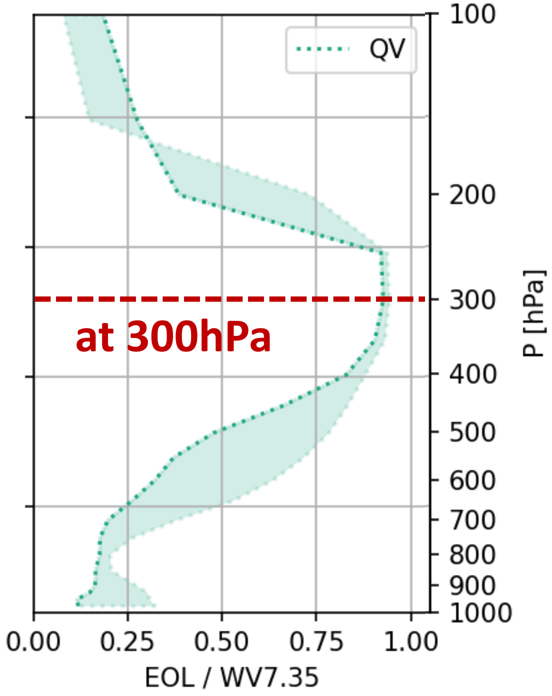


VIS (band 3 – 0.64 µm) | Solid = Case 1 | Dashed = Case 2 | Shading = Variability

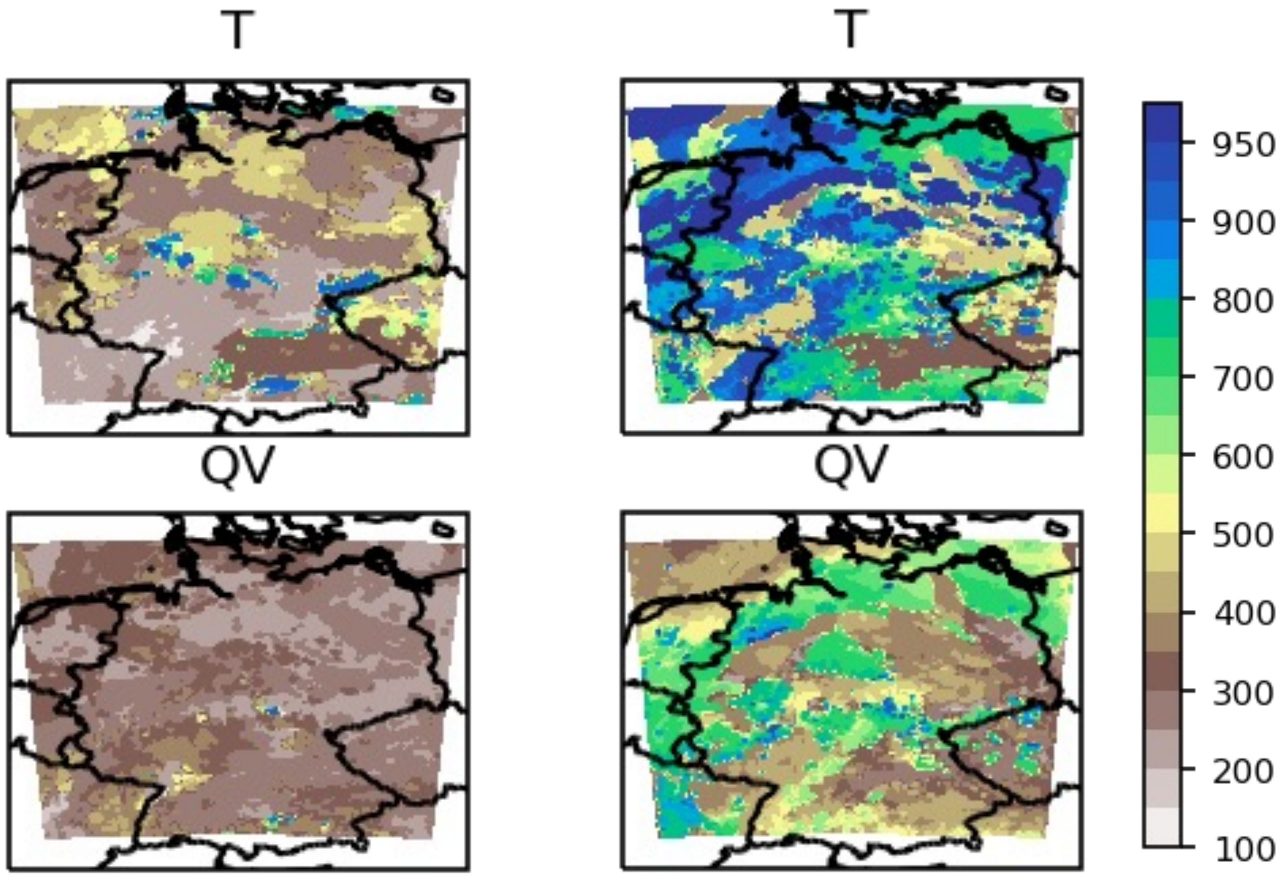
EOL for diagnostics: How situation dependent should localization be?

→ **Visible channels shows higher situation-dependence and would benefit more from adaptive localization**

Diagnostic: Derive level (hPa) of maximum column EOL

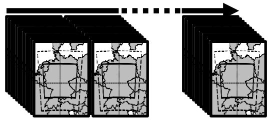


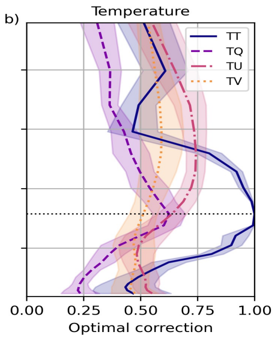
≈ good height estimate of GC center

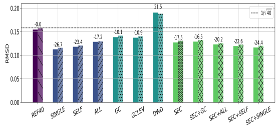


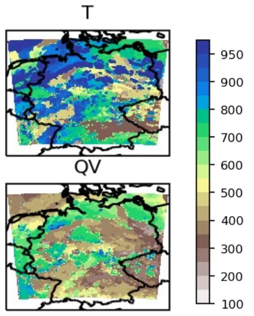
IR (band 10 – 7.35 μm) **VIS** (band 3 – 0.64 μm)

Conclusions – Take away massages









Empirical Optimal Localization (EOL) method allows deriving optimal localization from large ensembles using subsampling

$$\alpha_{single} = \frac{\sum_{s=1}^{S} r_s^{40} r^{1000}}{\sum_{s=1}^{S} (r_s^{40})^2}$$

Vertical model space covariance localization:

- Positive-definiteness: NCM algorithm (Higham, 2002) interesting tool for DA or ML
- Localization shapes, scales and approaches: Variable- and height-dependence
  --> *substantial room for improvements over current GC-based localizations*

Vertical observation space localization of satellite observations:

- Localization should be channel, variable, and cloud situation dependent
- Visible channels benefit more from situation-dependent localization than WV channels

Reference: Necker et al. 2023:
Guidance on how to improve vertical covariance localization based on a 1000-member ensemble
Nonlin. Processes Geophys., 30, 13–29 https://doi.org/10.5194/npg-30-13-2023

1000-member ensemble and localization references

Anderson, J. L. 2012: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. Mon. Wea. Rev., 140, 2359– 2371. https://doi.org/10.1175/MWR-D-11-00013.1

Anderson, J. L., 2016: Reducing Correlation Sampling Error in Ensemble Kalman Filter Data. Assimilation. Mon. Wea. Rev., 144, 913–925. https://doi.org/10.1175/MWR-D-15-0052.1

Destouches, M, Montmerle, T, Michel, Y, Ménétrier, B. 2021: Estimating optimal localization for sampled background-error covariances of hydrometeor variables. Q. J. R. Meteorol. Soc., 147, 74– 93. https://doi.org/10.1002/qj.3906

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. ., 125, 723–757. https://doi.org/10.1002/qj.49712555417

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev. ., 126, 796–811, https://doi.org/10.1175/1520-0493(1998)126,0796:DAUAEK.2.0.CO;2

Lei, L. and Anderson, J. L. 2014: Comparisons of Empirical Localization Techniques for Serial Ensemble Kalman Filters in a Simple Atmospheric General Circulation Model, Monthly Weather Review, 142, 739 – 754, https://doi.org/10.1175/MWR-D-13-00152.1

Nicholas J. Higham, **2002:** Computing the nearest correlation matrix—a problem from finance, IMA Journal of Numerical Analysis, Volume 22, Issue 3, Pages 329–343, https://doi.org/10.1093/imanum/22.3.329

Necker et al. 2020a: Necker, T, Geiss, S, Weissmann, M, Ruiz, J, Miyoshi, T, Lien, G-Y. (2020): A convective-scale 1,000-member ensemble simulation and potential applications. Q J R Meteorol Soc.; 146: 1423– 1442. https://doi.org/10.1002/qj.3744

Necker et al. 2020b: Necker, T., M. Weissmann, Y. Ruckstuhl, J. Anderson, and T. Miyoshi, (2020): Sampling Error Correction Evaluated Using a Convective-Scale 1000-Member Ensemble. Mon. Wea. Rev., 148, 1229–1249, https://doi.org/10.1175/MWR-D-19-0154.1

observations. Q J R Meteorol Soc., 149(754), 1546–1571. Available from: https://doi.org/10.1002/qj.4464

Necker et al. 2023: Necker, T., Hinger, D., Griewank, P. J., Miyoshi, T., and Weissmann, M. (2023): Guidance on how to improve vertical covariance localization based on a 1000-member ensemble, Nonlin. Processes Geophys., 30, 13–29, https://doi.org/10.5194/npg-30-13-2023

Necker et al. 2024: Necker, T., Honda, T., Griewank, P. J., Miyoshi, T., and Weissmann, M.: Situation-dependence of all-sky satellite correlations and localization in ensemble data assimilation. Planed submission to Q J R Meteorol Soc., (in preparation)