Localizing High-Dimensional Covariance Estimates with Hierarchical Rank Structure

Robin Armstrong[1], Samuel Otto[2], and Anil Damle[3]

American Geophysical Union 2025 Annual Meeting

NG21A: Advances in Data Assimilation, Data Fusion, Machine Learning, Predictability, and Uncertainty Quantification in the Geosciences I Oral

December 16th, 2025, New Orleans, Louisiana

[1] Cornell University, Center for Applied Mathematics
[2] Cornell University, Sibley School of Mechanical and Aerospace Engineering
[3] Cornell University, Department of Computer Science

My Collaborators

Anil Damle
Cornell University
Dept. of Computer Science

Samuel Otto
Cornell University
Sibley School of Mechanical & Aerospace Engineering

Covariance Matrix Estimation

- Data assimilation (DA) represents forecast uncertainty using a prior probability distribution P_0.
- Ensemble DA represents P_0 by an ensemble:

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \sim P_0.$$

- The **prior covariance matrix** allows information to spread from observed variables onto unobserved ones:

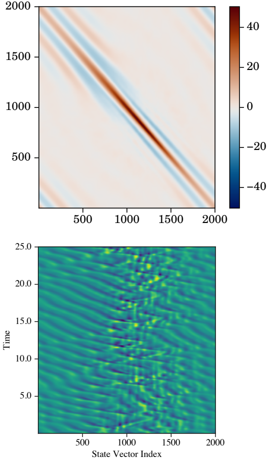
$$\mathbf{C} := \mathrm{Cov}[P_0].$$

- This matrix must be estimated from the ensemble:

$$\mathbf{C} \approx \widehat{\mathbf{C}} := \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}},$$

where $\overline{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i$.

- This approximation is very inaccurate when $m \ll n$.



Correlation-Based Localization

- We address undersampling with **localization** [Hamill et al., 2001, Vishny et al., 2024]:

$$\mathbf{C}_{i,j} = \ell_{i,j}\widehat{\mathbf{C}}_{i,j},$$

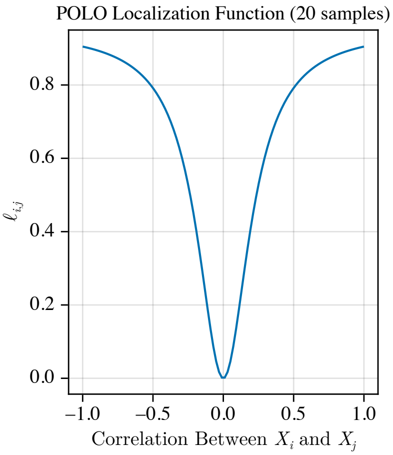
...where $\widehat{\mathbf{C}}$ = sample covariance of the small ensemble.

- **Prior Optimal LOcalization (POLO)** [Vishny et al., 2024] is the optimal localization function for multivariate Gaussian samples:

$$\mathbf{C}_{i,j} = \ell_{i,j}\widehat{\mathbf{C}}_{i,j}, \qquad \ell_{i,j} = \frac{(m-1)\rho_{i,j}^2}{1+m\rho_{i,j}^2},$$

where m = ensemble size, $\rho_{i,j}$ = **true correlation** between state variables i and j.

- In practice we must estimate $\rho_{i,j}$ from the samples.



Data Sparsity

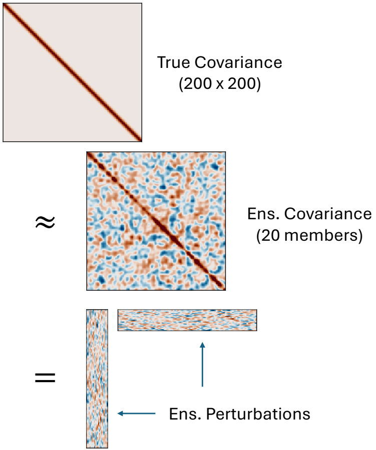
- No matter how we estimate the covariance, we need the estimate to be **data sparse**; representable in $\ll n^2$ floating point numbers.
- **Unlocalized covariance of size-*m* ensemble is low-rank**; we pay $\mathcal{O}(mn)$ to store the samples.
- **Correlation-based localization destroys low-rank strucrure**; we need a different representation.
- Recompressing to low-rank form will not work.

Eckart-Young Theorem [Eckart and Young, 1936]

If $\mathbf{C}$ is a covariance matrix and $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ are its eigenvalues, then

$$\|\widetilde{\mathbf{C}}_k - \mathbf{C}\|_{\mathrm{F}} \geq \mathcal{E}^{(k)}_{\mathrm{min}} := \sqrt{\lambda^2_{k+1} + \lambda^2_{k+2} + \ldots}$$

for any rank-k approximation $\widetilde{\mathbf{C}}_k$.



Hierarchical Rank Structure

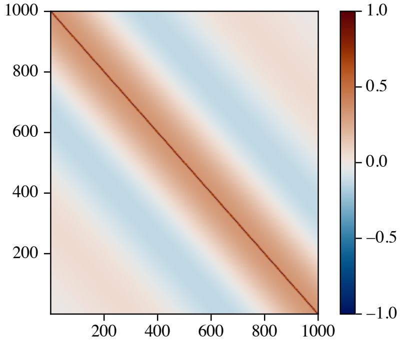
Today I will show you...

...a way to improve the efficiency of correlation-based localization using **hierarchical rank structure**.

- Informally: *correlations vary more smoothly at long distances than at short distances.*
- More formally: *cross-covariances between well-separated domains are low-rank.*

Definition

A **rank-k $\mathcal{H}$-matrix** is a data structure for representing an $n \times n$ hierarchically rank-structured matrix in $\mathcal{O}(nk \log n)$ floating point numbers while supporting fast linear algebra operations (e.g., matvecs, linear system solves).



Hierarchical Rank Structure

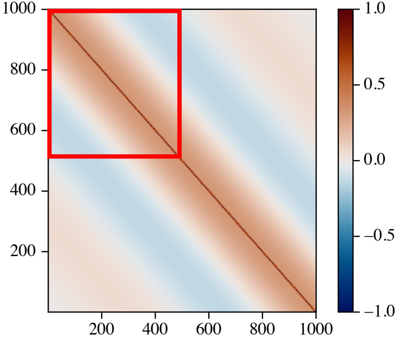
Today I will show you...

...a way to improve the efficiency of correlation-based localization using **hierarchical rank structure**.

- Informally: *correlations vary more smoothly at long distances than at short distances.*
- More formally: *cross-covariances between well-separated domains are low-rank.*

Definition

A **rank-k $\mathcal{H}$-matrix** is a data structure for representing an $n \times n$ hierarchically rank-structured matrix in $\mathcal{O}(nk \log n)$ floating point numbers while supporting fast linear algebra operations (e.g., matvecs, linear system solves).



Hierarchical Rank Structure

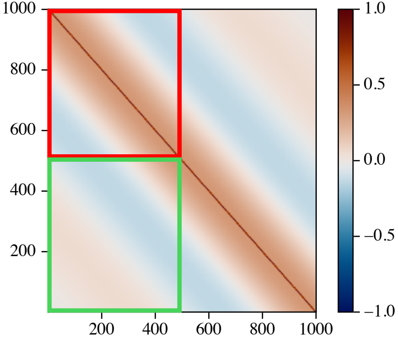
Today I will show you...

...a way to improve the efficiency of correlation-based localization using **hierarchical rank structure**.

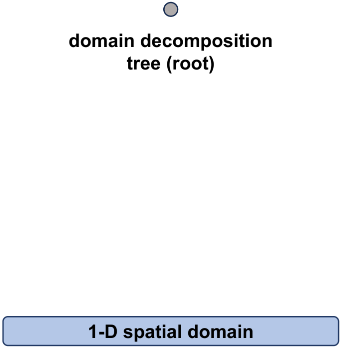
- Informally: *correlations vary more smoothly at long distances than at short distances*.
- More formally: *cross-covariances between well-separated domains are low-rank.*

Definition

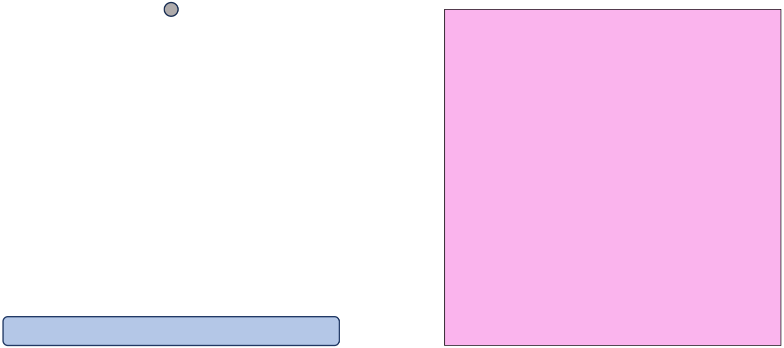
A **rank-k $\mathcal{H}$-matrix** is a data structure for representing an $n \times n$ hierarchically rank-structured matrix in $\mathcal{O}(nk \log n)$ floating point numbers while supporting fast linear algebra operations (e.g., matvecs, linear system solves).



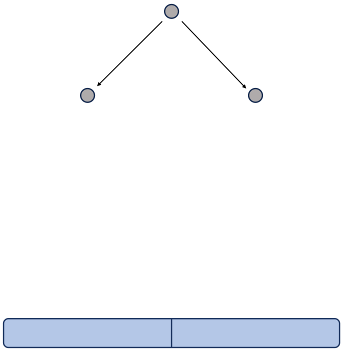
Compression Using Hierarchical Rank Structure

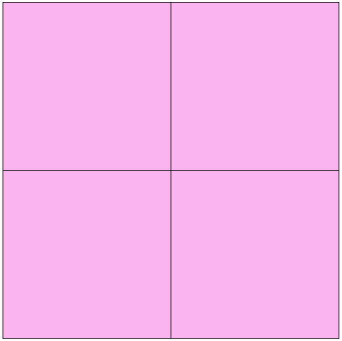


Compression Using Hierarchical Rank Structure

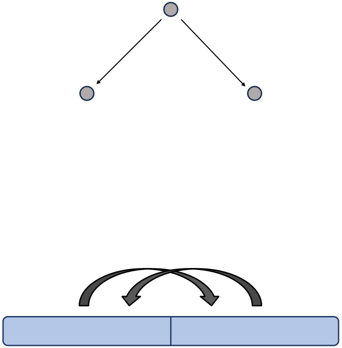


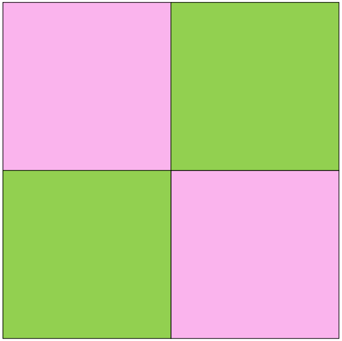
Compression Using Hierarchical Rank Structure



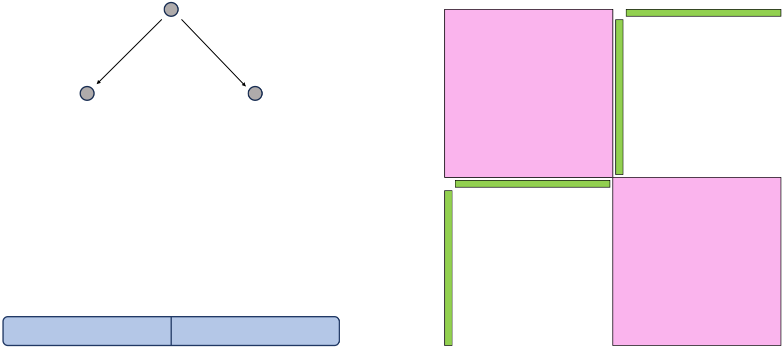


Compression Using Hierarchical Rank Structure

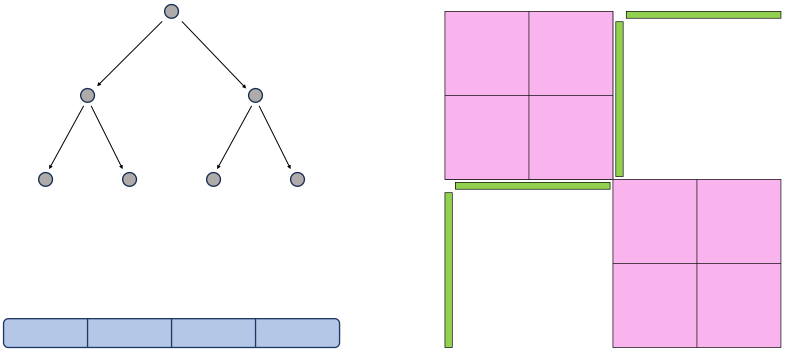




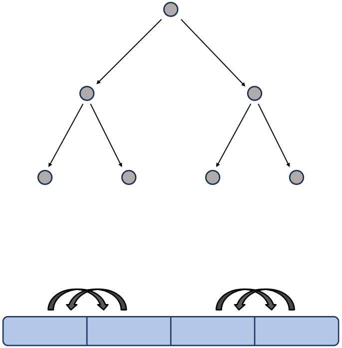
Compression Using Hierarchical Rank Structure

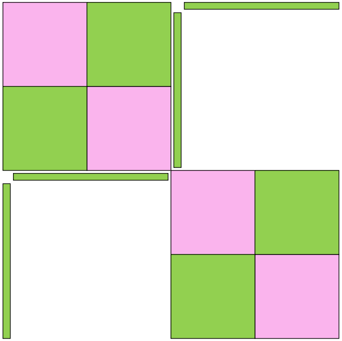


Compression Using Hierarchical Rank Structure

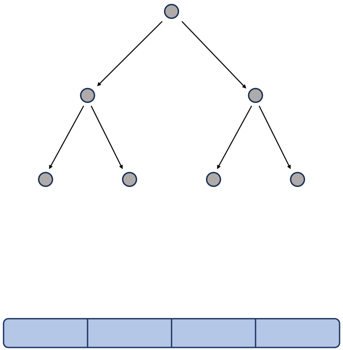


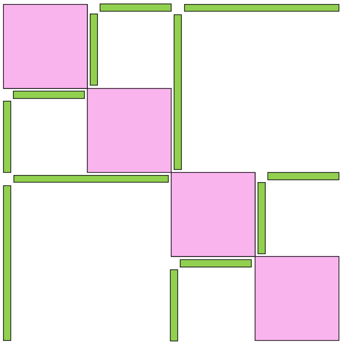
Compression Using Hierarchical Rank Structure



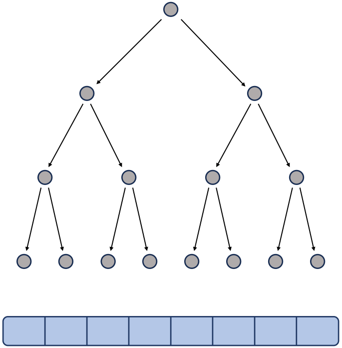


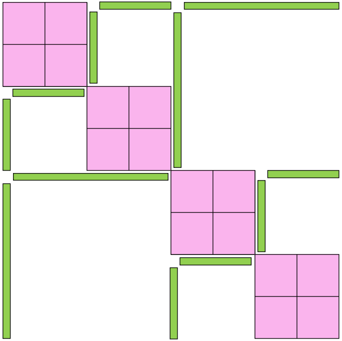
Compression Using Hierarchical Rank Structure



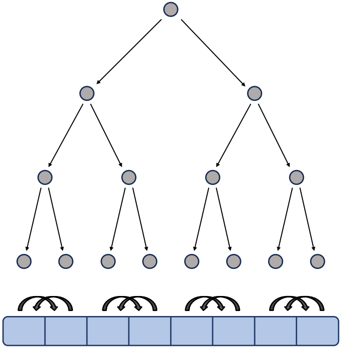


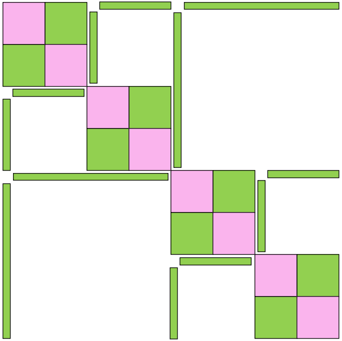
Compression Using Hierarchical Rank Structure



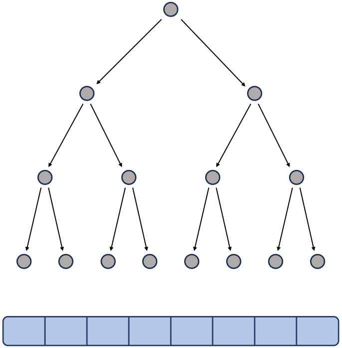


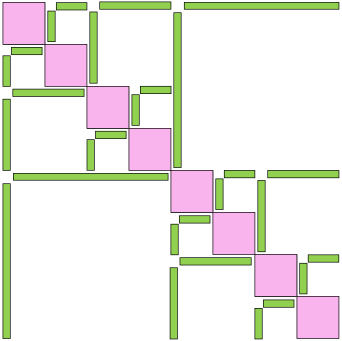
Compression Using Hierarchical Rank Structure





Compression Using Hierarchical Rank Structure





Estimating Cross Covariances

- Let $\mathcal{X}$, $\mathcal{Y}$ be well-separated subdomains of space.
- **The goal:** Localize the cross-covariance matrix

$$\widehat{\mathbf{C}}_{\mathcal{X},\mathcal{Y}} = \mathbf{Z}_{\mathcal{X}}\mathbf{Z}_{\mathcal{Y}}^{\mathrm{T}},$$

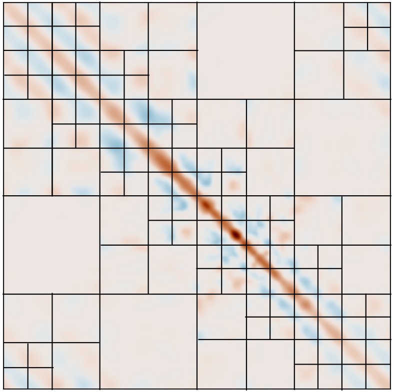
where $\mathbf{Z}_{\mathcal{X}}$ (resp. $\mathbf{Z}_{\mathcal{Y}}$) = perturbations on $\mathcal{X}$ (resp. $\mathcal{Y}$). We need the end result to be in low-rank form.

- "Baseline" localized covariance (not low-rank):

$$\mathbf{C}_{\mathcal{X},\mathcal{Y}} = (\mathbf{Z}_{\mathcal{X}}\mathbf{Z}_{\mathcal{Y}}^{\mathrm{T}}) * \ell_{\mathrm{POLO}}((\mathbf{P}_{\mathcal{X}}\mathbf{Z}_{\mathcal{X}})(\mathbf{P}_{\mathcal{Y}}\mathbf{Z}_{\mathcal{Y}})^{\mathrm{T}}, m),$$

where $\mathbf{P}_{\mathcal{X}}$, $\mathbf{P}_{\mathcal{Y}}$ are smoothing transformations, and $\ell_{\mathrm{POLO}}(\cdot, m)$ = POLO localizer for m members.

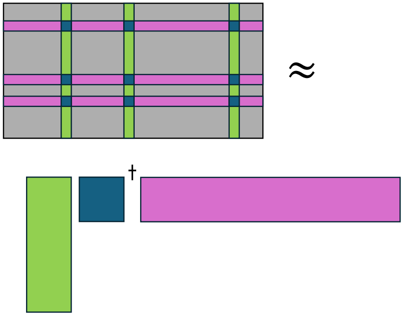
- POLO function provides adaptivity to ensemble size, smoothing transformations provide more robustness against sampling noise.



Example: Localized covariance.

Skeletal Approximations

- We need our cross-covariance estimates to be **low-rank** in order for the overall estimate to be computationally efficient.
- We use a **generalized Nÿstrom approximation** [Murray et al., 2023]:
  1. Select a small number of *skeleton rows*.
  2. Select a small number of *skeleton columns*.
  3. Approximate the remaining rows/columns in terms of the skeleton rows/columns.
- We only ever form the skeleton rows/columns; we never form the entire cross-covariance block.
- **How to choose skeleton rows/columns?** Main ingredients:
  1. Gauss-Legendre quadrature, and
  2. column-pivoted QR factorization.



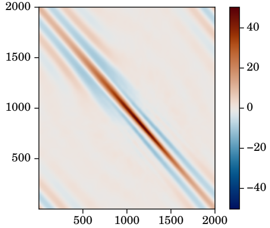
Test Case 1: "Storm Track" Dynamics

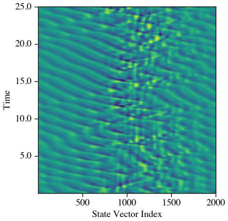
- A modification of "model II" from [Lorenz, 2005]. Like the Lorenz '96 model, but:
  - admits waves much larger than the grid spacing, and
  - has a "stable" region of strong damping and a "chaotic" region of weak damping.

  Based off a system from [Bishop et al., 2017].
- **Domain:** 2000 grid points in 1D with periodic boundary conditions.
- **Partition:** recursive bisection until domain has at most 10 grid points.
- **Admissibility criterion:**

$$\min\{\ell(\mathcal{X}), \ell(\mathcal{Y})\} \leq d(\mathcal{X}, \mathcal{Y}),$$

where $\ell(\cdot) =$ domain length, and $d(\mathcal{X}, \mathcal{Y}) = \inf_{x\in\mathcal{X}, y\in\mathcal{Y}} |x - y|$.





Test Case 2: 2D Quasigeostrophic Turbulence

- **Quasigeostrophic flow** approximates the motion of a rotating fluid where Coriolis and pressure-gradient forces are nearly in balance [Daley, 1991].
- **Domain:** 128×128 grid on a 2D square with periodic boundary conditions.
- **Partition:** bisecting rectangles until longest side spans at most 10 gridpoints.
- **Admissibility criterion:**

$$\min\{\ell(\mathcal{X}), \ell(\mathcal{Y})\} \leq d(\mathcal{X}, \mathcal{Y}),$$

  where $\ell(\cdot) =$ max sidelength of rectangle, and $d(\mathcal{X}, \mathcal{Y}) = \inf_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\mathbf{x} - \mathbf{y}\|_2$.
- Simulated with code from `https://github.com/jswhit/sqgturb`.

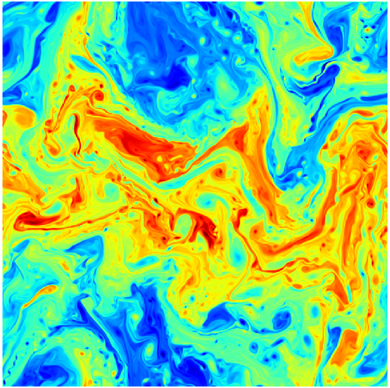
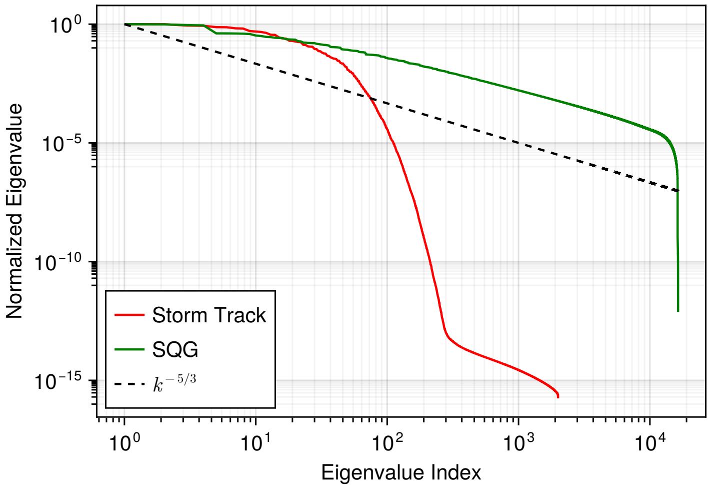
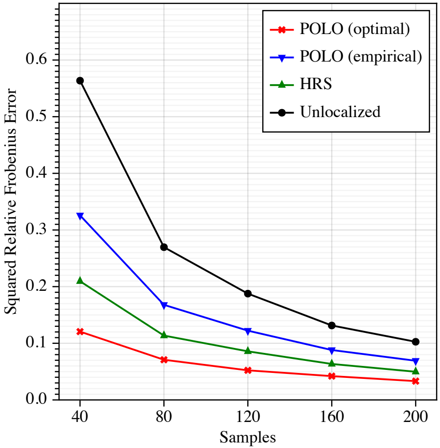


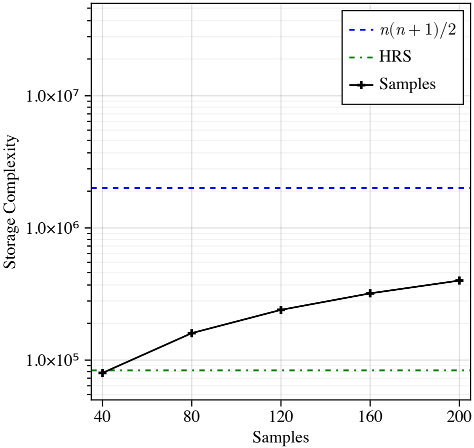
Figure: from `https://github.com/jswhit/sqgturb`.

Problem Difficulty

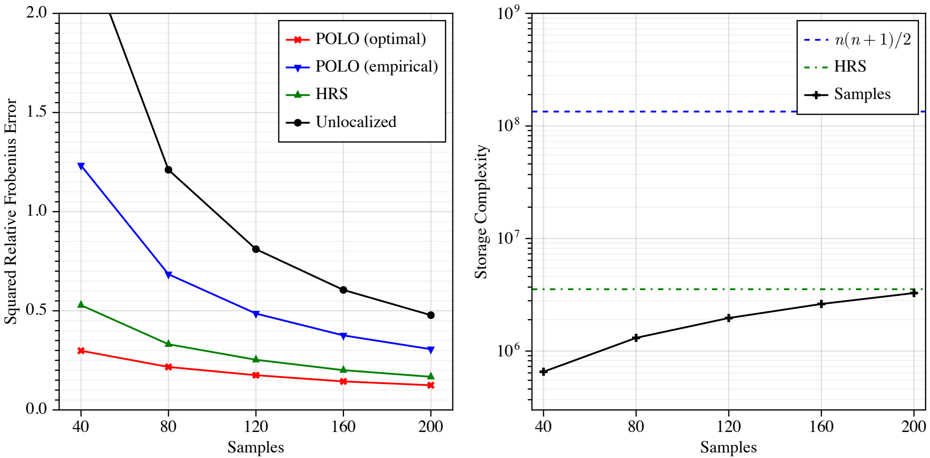


Results: “Storm Track” Dynamics





Results: Quasigeostrophic Turbulence



Test Case 3: Data Assimilation

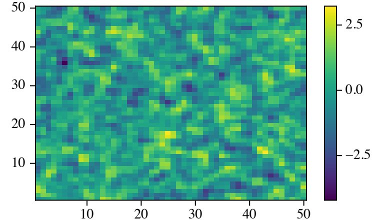
- **Model**: a 2D Gaussian process on a 50×50 grid.
- **Observing system**: a 5×5 grid of "sensors" that observe a weighted average over a small nearby region.
- **Error measure 1**: relative analysis variance error.

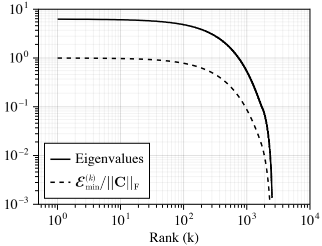
$$E_1 = \frac{1}{n} \sum_{i=1,\, j=1}^{n} \frac{|v_{ij} - \widehat{v}_{ij}|}{v_{ij}},$$

  where v_{ij} (resp. $\widehat{v}_{ij}$) = true (resp. ensemble) analysis variance at gridpoint (i, j).
- **Error measure 2**: relative Kalman gain accuracy.

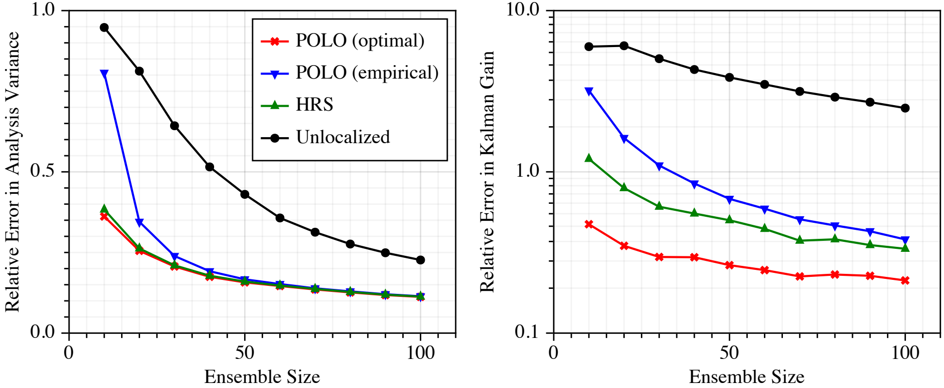
$$E_2 = \|\mathbf{K}\|_2^{-1} \|\widehat{\mathbf{K}} - \mathbf{K}\|_2,$$

  where $\mathbf{K}$ (resp. $\widehat{\mathbf{K}}$) = true (resp. localized ensemble) Kalman gain matrix.





Results: Data Assimilation



Conclusions

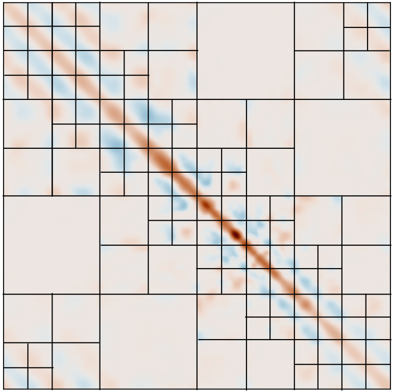
Summary

High-dimensional covariance estimation from a limited number of samples is a challenging problem arising in DA. **Localization** is critical for dealing with the effects of undersampling. **Hierarchical rank structure** provides an effective framework for localization.

Future Directions

- Using a different hierarchical matrix format: recursive skeletonization [Minden et al., 2017].
- Enforcing positive definiteness (related to the above).
- Testing on model reduction and cycled DA problems.

- **Thank you!**



References I

Bishop, C. H., Whitaker, J. S., and Lei, L. (2017).
Gain form of the ensemble transform Kalman filter and its relevance to satellite data assimilation with model space ensemble covariance localization.
Monthly Weather Review, 145(11):4575 – 4592.

Daley, R. (1991).
Atmospheric Data Analysis.
Cambridge Atmosphere and Space Science Series. Cambridge University Press.

Eckart, C. and Young, G. (1936).
The approximation of one matrix by another of lower rank.
Psychometrika, 1(3):211–218.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001).
Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter.
Monthly Weather Review, 129(11).

Lorenz, E. N. (2005).
Designing chaotic models.
Journal of the Atmospheric Sciences, 62(5):1574 – 1587.

References II

Minden, V., Ho, K. L., Damle, A., and Ying, L. (2017).
A recursive skeletonization factorization based on strong admissibility.
Multiscale Modeling & Simulation, 15(2):768–796.

Murray, R., Demmel, J., Mahoney, M. W., Erichson, N. B., Melnichenko, M., Malik, O. A., Grigori, L., Luszczek, P., Dereziński, M., Lopes, M. E., Liang, T., Luo, H., and Dongarra, J. (2023).
Randomized numerical linear algebra : A perspective on the field with an eye to software.

Vishny, D., Morzfeld, M., Gwirtz, K., Bach, E., Dunbar, O. R. A., and Hodyss, D. (2024).
High-dimensional covariance estimation from a small number of samples.
Journal of Advances in Modeling Earth Systems, 16(9):e2024MS004417.
e2024MS004417 2024MS004417.