

Correlation-based localization in iterative ensemble methods

Geir Evensen

Femke Vossepoel and Peter Jan van Leeuwen

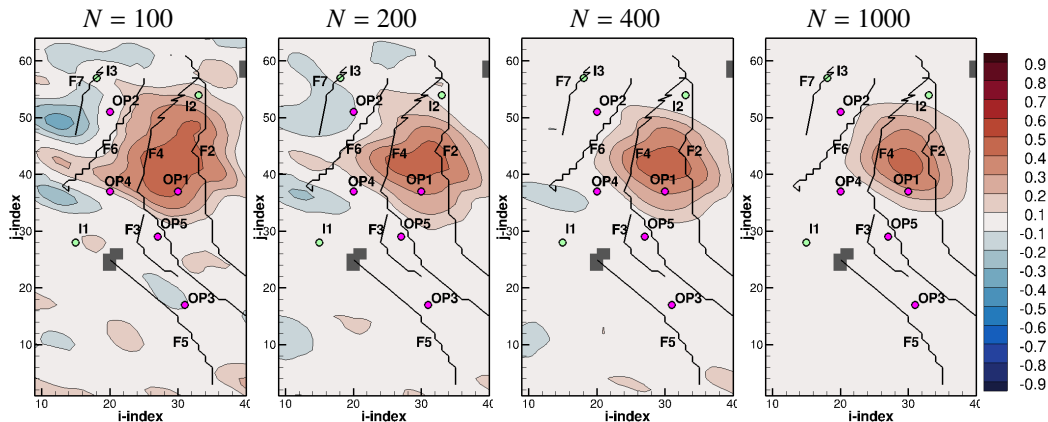


Petroleum parameter estimation

Problem definition:

- State variables are 3D fields of rock porosity and permeability.
- Measurements are the produced oil, gas, and water from different wells and time lapse seismic.
- Estimate reservoir properties to predict the produced fluids from existing and future wells.

Ensemble correlations: OP1-oilrate and porosity



Ensemble correlations

- We must use an unaffordably large ensemble size to “eliminate” spurious correlations.
- The nonlinear relation between model parameters and measurements reduces their correlation.
- Measurements are non-local and it is hard to use distance based localization.
- We must ensure a significant physical-correlation to spurious-noise ratio.

Global versus local analysis

- Global analysis

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T}. \quad (1)$$

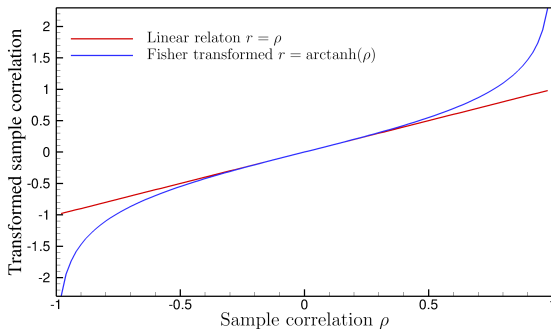
- Local analysis (each grid point has its own transition matrix based on selected measurements).

$$\mathbf{Z}_l^a = \mathbf{Z}_l^f \mathbf{T}_l. \quad (2)$$

- Many small inversions with a few selected measurements instead of one large using all.
- We are not estimating a huge low-rank covariance matrix.
- Local analysis is easy to parallelize.

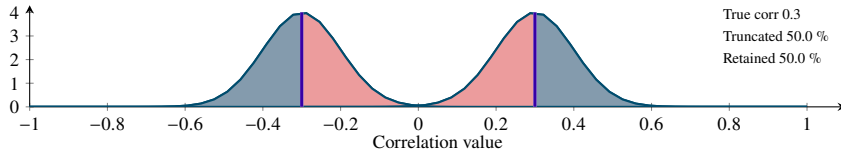
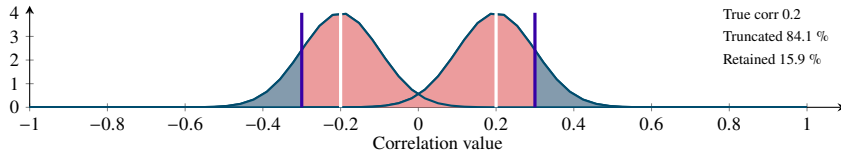
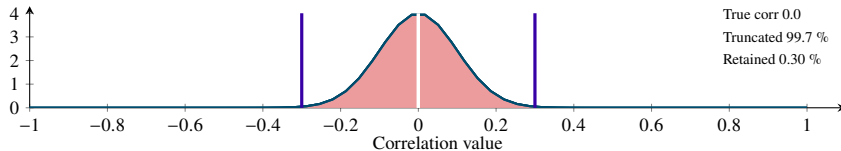
Properties of sample correlations

The Fisher transformed sample correlations are Gaussian distributed: $\mathcal{N}(r, 1/\sqrt{N-3})$



Small ensemble correlations are approximate samples of a Gaussian $\mathcal{N}(\rho, 1/\sqrt{N-3})$.

Correlation-based localization, $N = 100$



Introducing localization

Local analysis updates variables in a gridpoint using only “nearby” measurements.

- Distance-based truncation of “spatially” remote measurements.
- Correlation-based truncation of “weakly correlated” measurements.

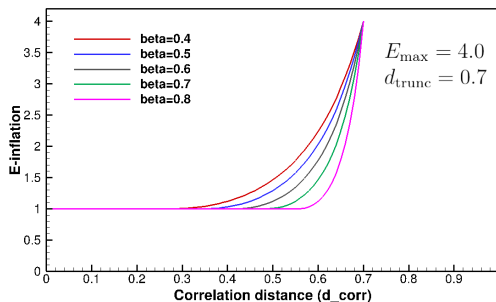
We define a **correlation distance**

$$d_{\text{corr}} = 1 - |\rho| \quad \in \quad [0 : 1]. \quad (3)$$

Truncating when the correlation distance

$$d_{\text{corr}} > d_{\text{trunc}} \approx 1 - 3/\sqrt{N}. \quad (4)$$

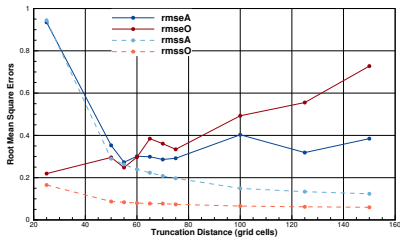
Additional inflation of remote measurement's errors



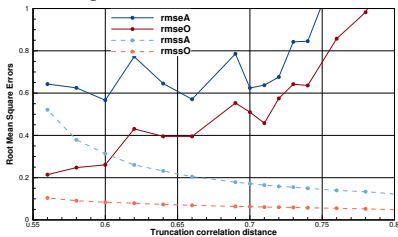
$$E_{\text{inflate}}(i, m) = \begin{cases} 1 & d_{\text{corr}}(i, m) \leq \beta d_{\text{trunc}} \\ \exp\left(\frac{d_{\text{corr}}(i, m) - \beta d_{\text{trunc}}}{b}\right)^2 & d_{\text{corr}}(i, m) > \beta d_{\text{trunc}} \end{cases}, \quad (5)$$

Results from the multiscale coupled Kuramoto-Sivashinsky model

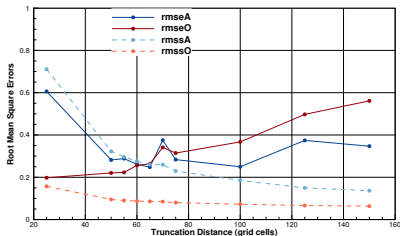
Distance localization: $E_{max} = 1.0$



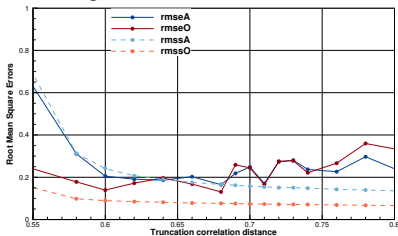
Adaptive localization: $E_{max} = 1.0$



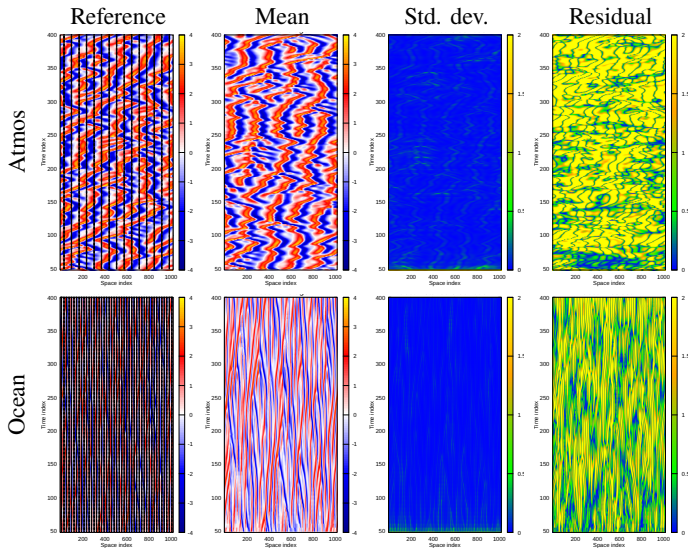
Distance localization: $E_{max} = 4.0$



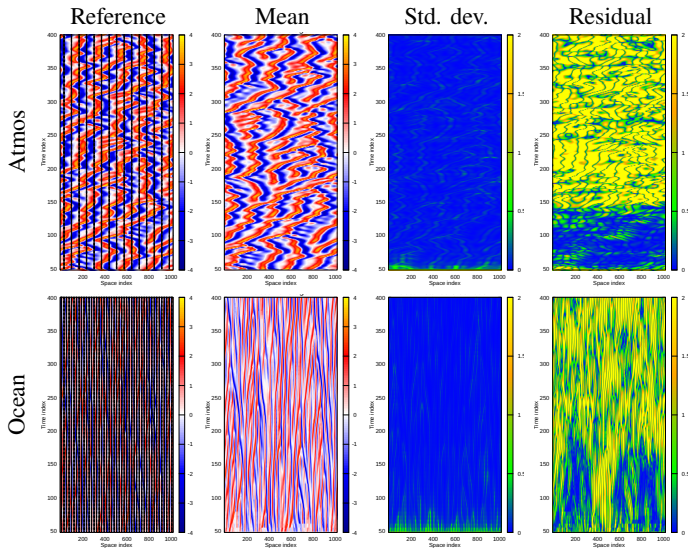
Adaptive localization: $E_{max} = 8.0$



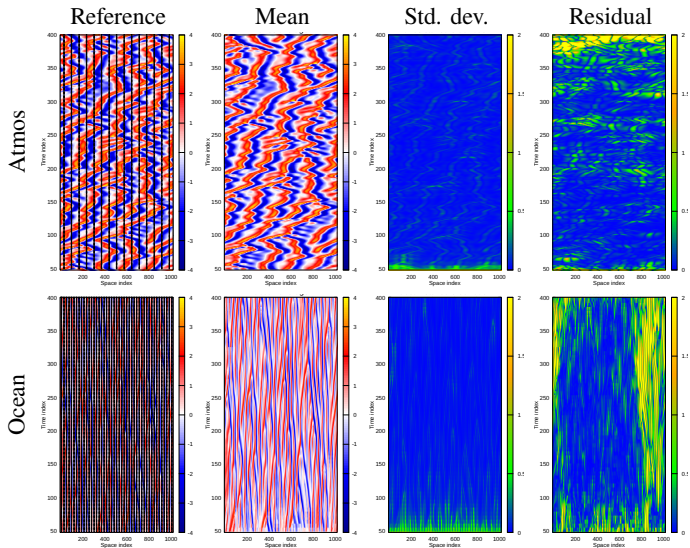
Global update, $N = 100$



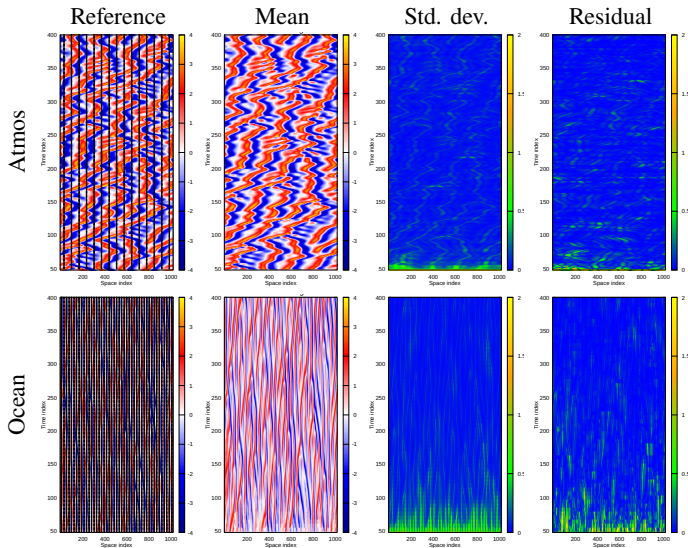
Global update, $N = 200$



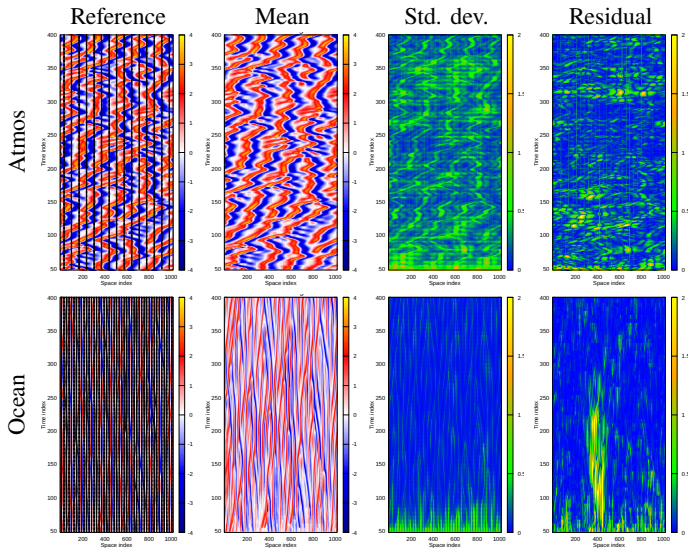
Global update, $N = 400$



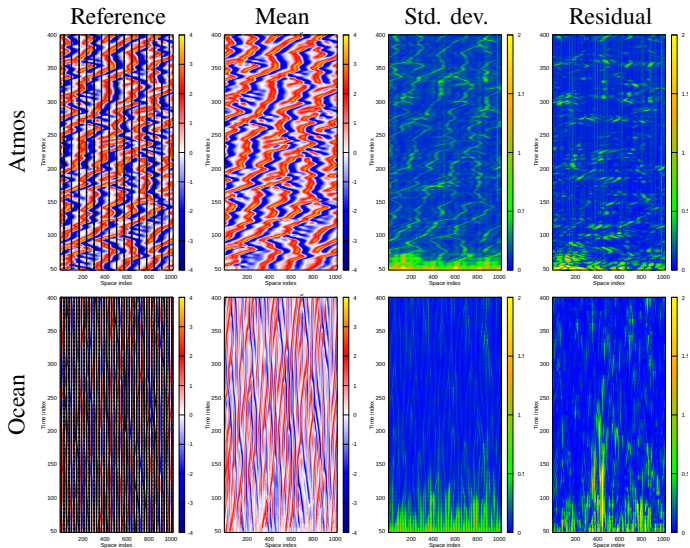
Global update, $N = 1000$



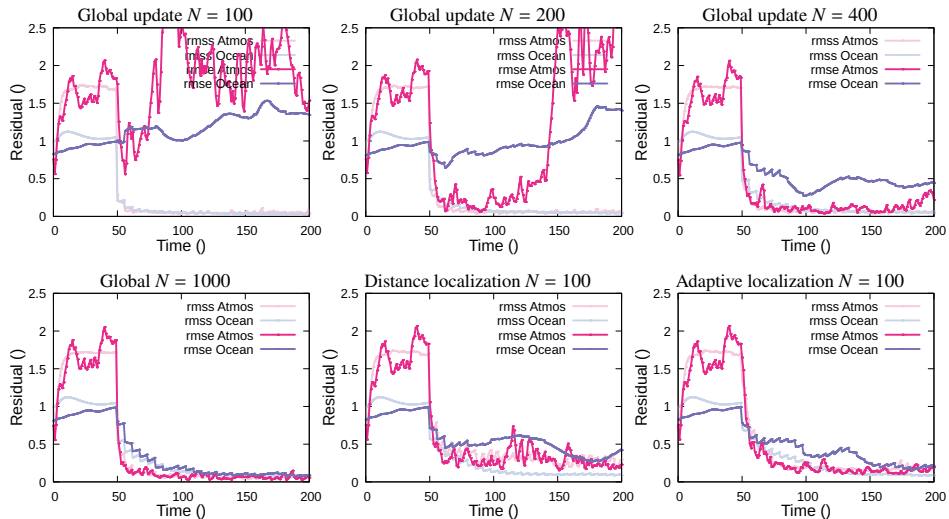
Distance localization, $N = 100$



Correlation localization, $N = 100$



Residuals



Conclusion

- Adaptive correlation-based localization is as good as the best distance-based localization.
- Be concerned about ensemble size and physical-correlation to spurious noise ratio.
- Computational cost of computing all correlations is nmN flops, but it is possible to simplify.
- See Femkes key-note presentation on Friday for more on localization in the KS model.



data-assimilation.no/workshops/EnKF-WS-2025



20th international EnKF workshop (2025)

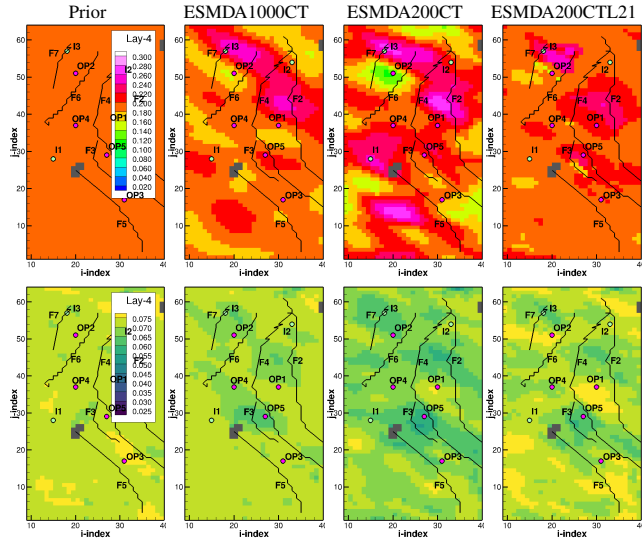
June 16 – 19, Ullensvang, Norway

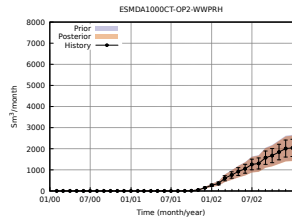
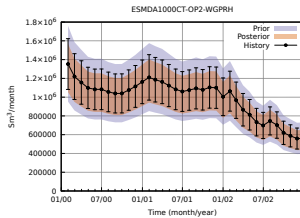
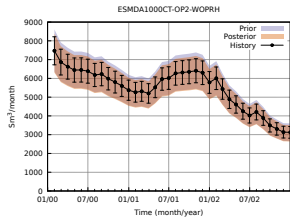
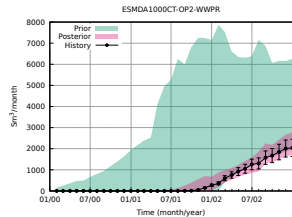
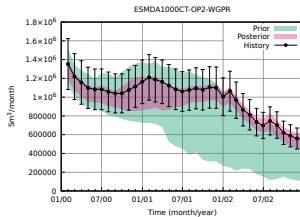
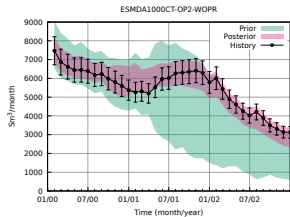
Abstracts and registration

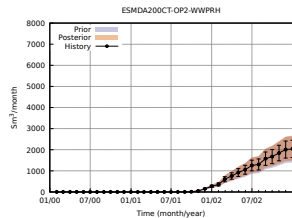
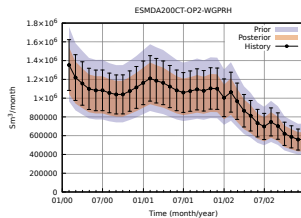
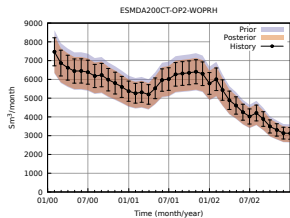
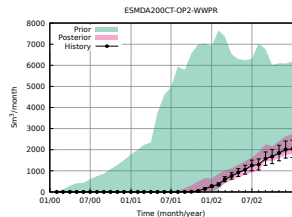
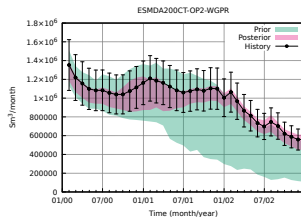
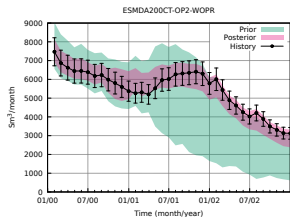
Ullensvang hotel



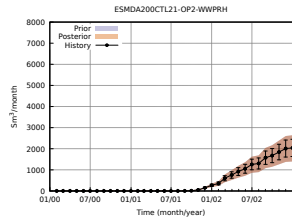
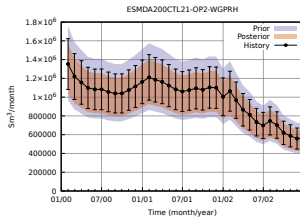
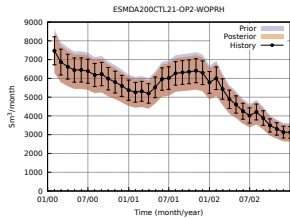
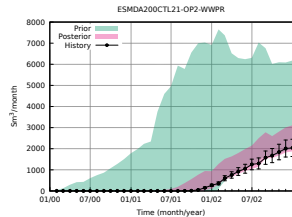
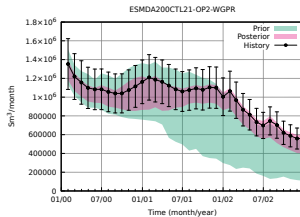
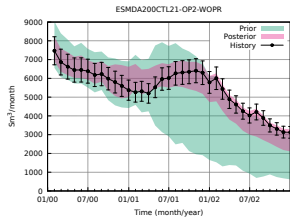
Reservoir application







ESMDA200CTL21



Bibliography

Evensen, G., F. C. Vossepoel, and P. J. Van Leeuwen. Iterative ensemble smoothers for data assimilation in coupled nonlinear multiscale models. *Mon. Weather Rev.*, 152, 2024. doi:[10.1175/MWR-D-23-0239.1](https://doi.org/10.1175/MWR-D-23-0239.1).

Coupled multiscale Kuramoto-Sivashinsky model

$$\frac{\partial A}{\partial t} = -A \frac{\partial A}{\partial x} - \frac{\partial^2 A}{\partial x^2} - \frac{1}{2} \frac{\partial^4 A}{\partial x^4} + 0.003(O - A), \quad (6)$$

$$\epsilon \frac{\partial O}{\partial t} = -O \frac{\partial O}{\partial x} - \frac{\partial^2 O}{\partial x^2} - \frac{\partial^4 O}{\partial x^4} + 0.003(A - O). \quad (7)$$

Spatial scale difference by using domain size of 32 for Atmos and 256 for Ocean.