Are linear model operators still needed in four-dimensional data assimilation?

Jon Poterjoy

Department of Meteorology
The Pennsylvania State University

Friday 9th May, 2014

Introduction

There are currently two competing 4-D data assimilation methods for operational NWP.

E4DVar: 4DVar with a mix of ensemble and static background error covariance.

4DEnVar: same as above, but with linearized model operations replaced by 4-D ensemble.

4DVar

Let the probability distribution of the model state at $t = 0$ be given by $\mathbf{x}_0 \sim N(\mathbf{x}_0^b, \mathbf{B})$.

4DVar minimizes a cost function to find the $\mathbf{x}_0^a$ that fits $\mathbf{x}_0^b$ and the set of observations, $\mathbf{y}_t$ for $t = 0, 1, 2, ..., T$.

The model state is integrated forward in time using M_t and tranformed to observation space using H_t.

Incremental 4DVar

Incremental 4DVar uses linear operators to transform increments at $t = 0$ into observation space at $t = \tau$:

$$\begin{aligned} H_\tau[M_\tau(\mathbf{x}_0)] &= H_\tau[M_\tau(\mathbf{x}_0^b + \delta\mathbf{x}_0)], \\ &\approx H_\tau[M_\tau(\mathbf{x}_0^b)] + \mathbf{H}_\tau \mathbf{M}_\tau \delta\mathbf{x}_0, \end{aligned}$$

where $\mathbf{H}_\tau$ and $\mathbf{M}_\tau$ are linearized about $\mathbf{x}_t^b$ between $t = 0$ and $t = \tau$.

Solution for single observation

For a set of observations at $t = \tau$, the minimum of the 4DVar cost function is given by

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \overline{\mathbf{x}_0'[H_\tau(\mathbf{x}_\tau)']^T}[\overline{H_\tau(\mathbf{x}_\tau)'[H_\tau(\mathbf{x}_\tau)']^T} + \mathbf{R}_\tau]^{-1}[\mathbf{y}_\tau - H_\tau(\mathbf{x}_\tau)].$$

$\overline{\mathbf{x}_0'[H_\tau(\mathbf{x}_\tau)']^T} \leftarrow$ Covariance between $\mathbf{x}_0$ and $\mathbf{x}_\tau$ in observation space.

$\overline{H_\tau(\mathbf{x}_\tau)'[H_\tau(\mathbf{x}_\tau)']^T} \leftarrow$ Variance of $\mathbf{x}_\tau$ in observation space.

E4DVar and 4DEnVar

For <u>ensemble **B**</u> with <u>no localization</u> and <u>no hybrid covariance</u>, E4DVar is the same as 4DEnVar, but with additional linear approximations for H_τ and M_τ.

E4DVar:

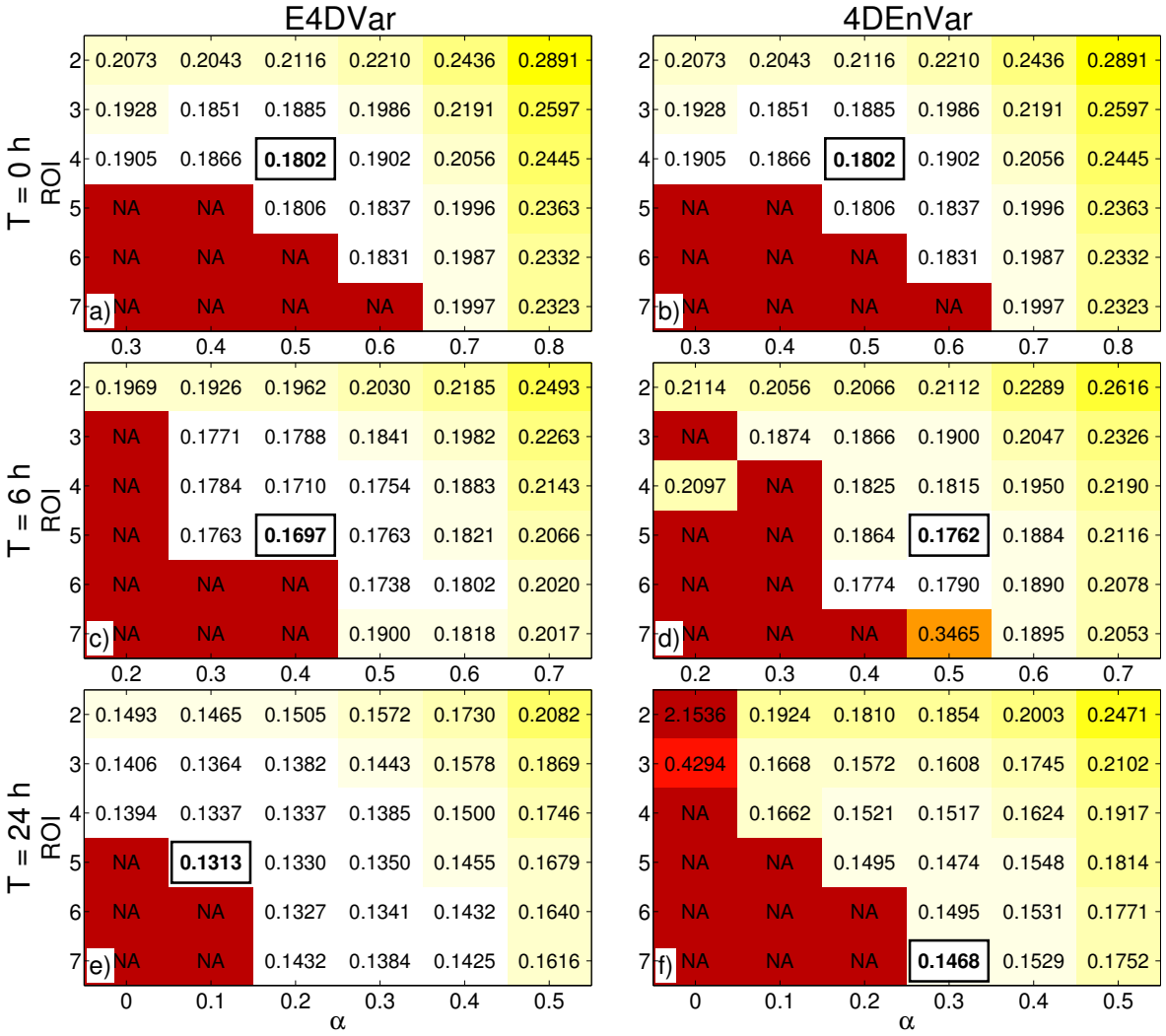
$$\begin{aligned}\overline{\mathbf{x}_0'[H_\tau(\mathbf{x}_\tau)']^T} &\approx [\mathbf{H}_\tau\mathbf{M}_\tau\mathbf{B}]^T \\ &= \frac{1}{N_e-1}\sum_{n=1}^{N_e}\mathbf{x}_{0,n}'[\mathbf{H}_\tau\mathbf{M}_\tau\mathbf{x}_{0,n}']^T,\end{aligned}$$

4DEnVar:

$$\overline{\mathbf{x}_0'[H_\tau(\mathbf{x}_\tau)']^T} \approx \frac{1}{N_e-1}\sum_{n=1}^{N_e}\mathbf{x}_{0,n}'[H_\tau(\mathbf{x}_{\tau,n})']^T$$

Cycling DA experiments with Lorenz-96

Results from cycling data assimilation experiments in <u>full ensemble mode</u> with <u>localization</u> show that E4DVar persistently outperforms 4DEnVar as assimilation window length T increases.



4-D covariance in Lorenz-96 model

Using the Lorenz-96 model, E4DVar and 4DEnVar methods are applied for estimating $\mathbf{P}_{0,\tau}$, where

$$\mathbf{P}_{0,\tau} = \overline{\mathbf{x}_0' \mathbf{x}_\tau'^T}.$$

In each case, 10 samples are drawn from $N(\mathbf{x}_0^b, \mathbf{B}^c)$ to estimate the error statistics at $t = 0$.

$\mathbf{x}_0^b$ is a model state after a spin-up period, and $\mathbf{B}^c$ is a climatological covariance, estimated over 10 000 data assimilation cycles.

Localization

E4DVar: $\mathbf{P}_{0,\tau}$ is estimated after localizing the covariance at the beginning of the assimilation window

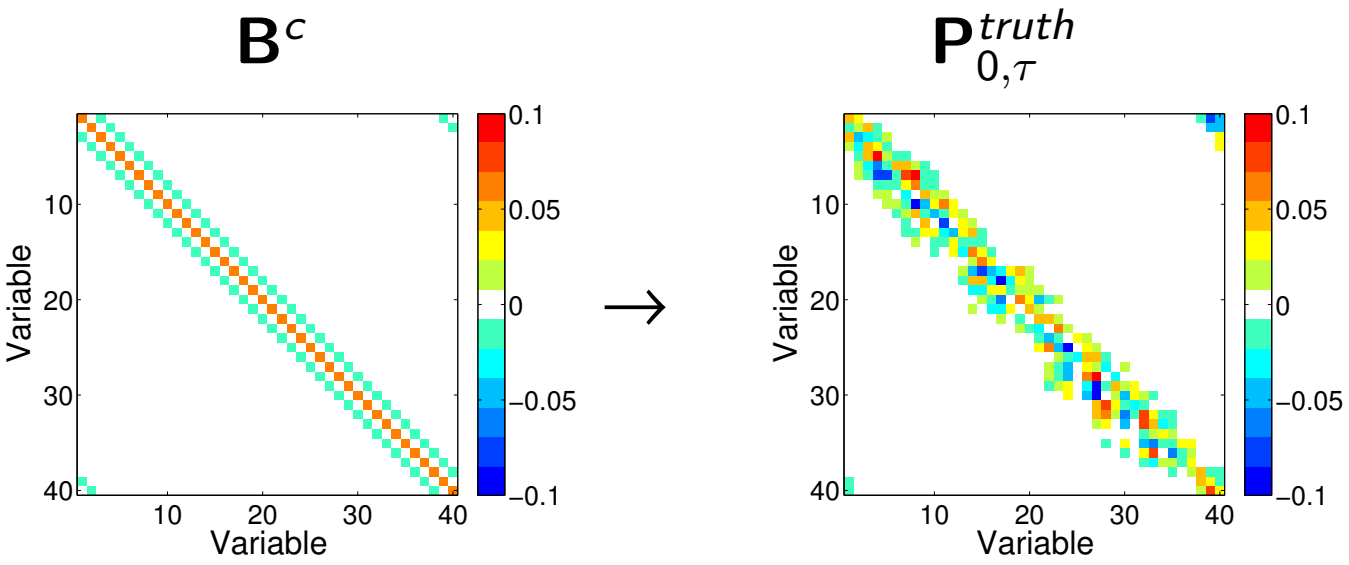
$$\mathbf{B} \quad \rightarrow \quad \mathbf{B} \circ \rho.$$

4DEnVar: $\mathbf{P}_{0,\tau}$ is localized after being estimated from unlocalized $\mathbf{B}$

$$\mathbf{P}_{0,\tau} \quad \rightarrow \quad \mathbf{P}_{0,\tau} \circ \rho.$$

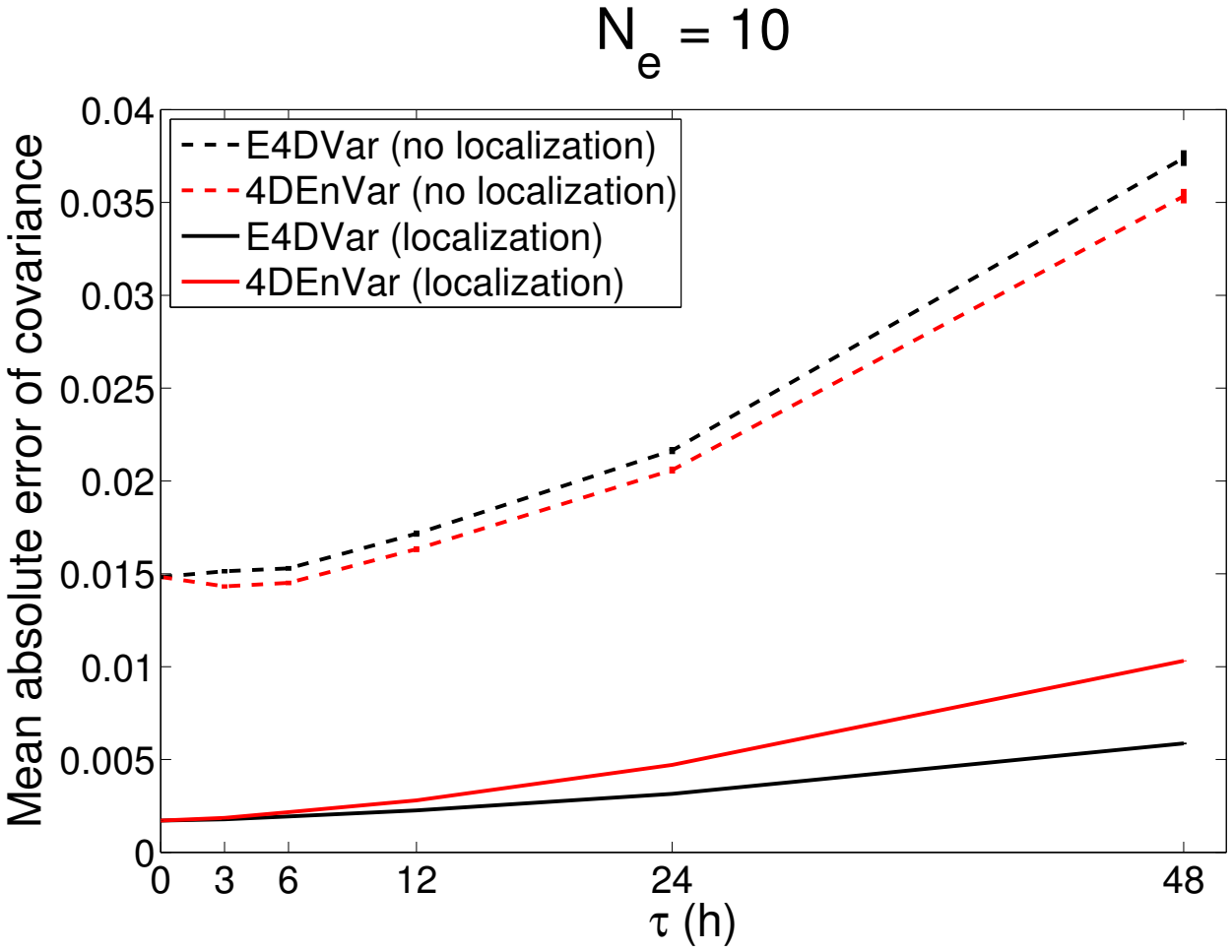
True covariance

The true $\mathbf{P}_{0,\tau}$ is estimated from a 10^5-member ensemble forecast, initialized with samples drawn from $N(\mathbf{x}_0^b, \mathbf{B}^c)$.



True $\mathbf{P}_{0,\tau}$ for $\tau = 24$ h.

Results from 100 trials



E4DVar outperforms 4DEnVar because of better localization!

Results from 100 trials

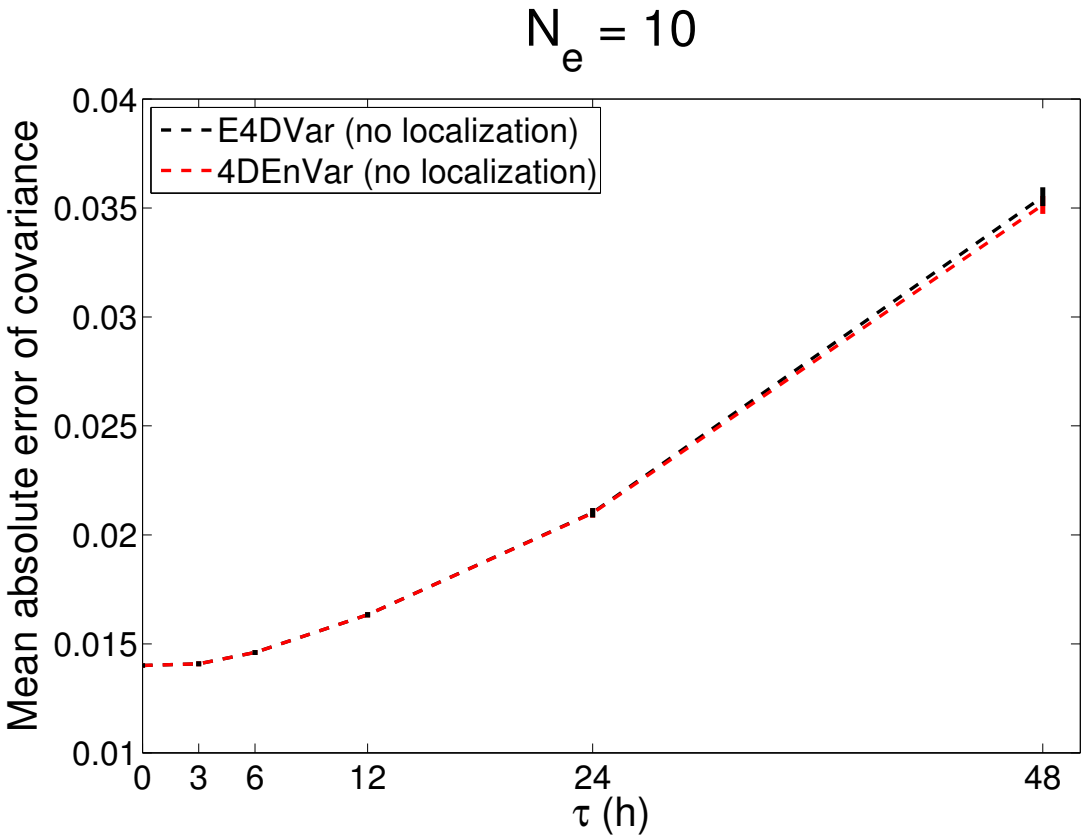
Without localization, 4DEnVar will always provide a more accurate $\mathbf{P}_{0,\tau}$ than E4DVar because it makes fewer approximations.

With localization, E4DVar is more accurate because sampling errors at $t = 0$ are removed before propagating to $t = \tau$.

Adaptive 4-D, flow-following localization in 4DEnVar is an active area of research (e.g., Bishop and Hodyss, MWR 2011).

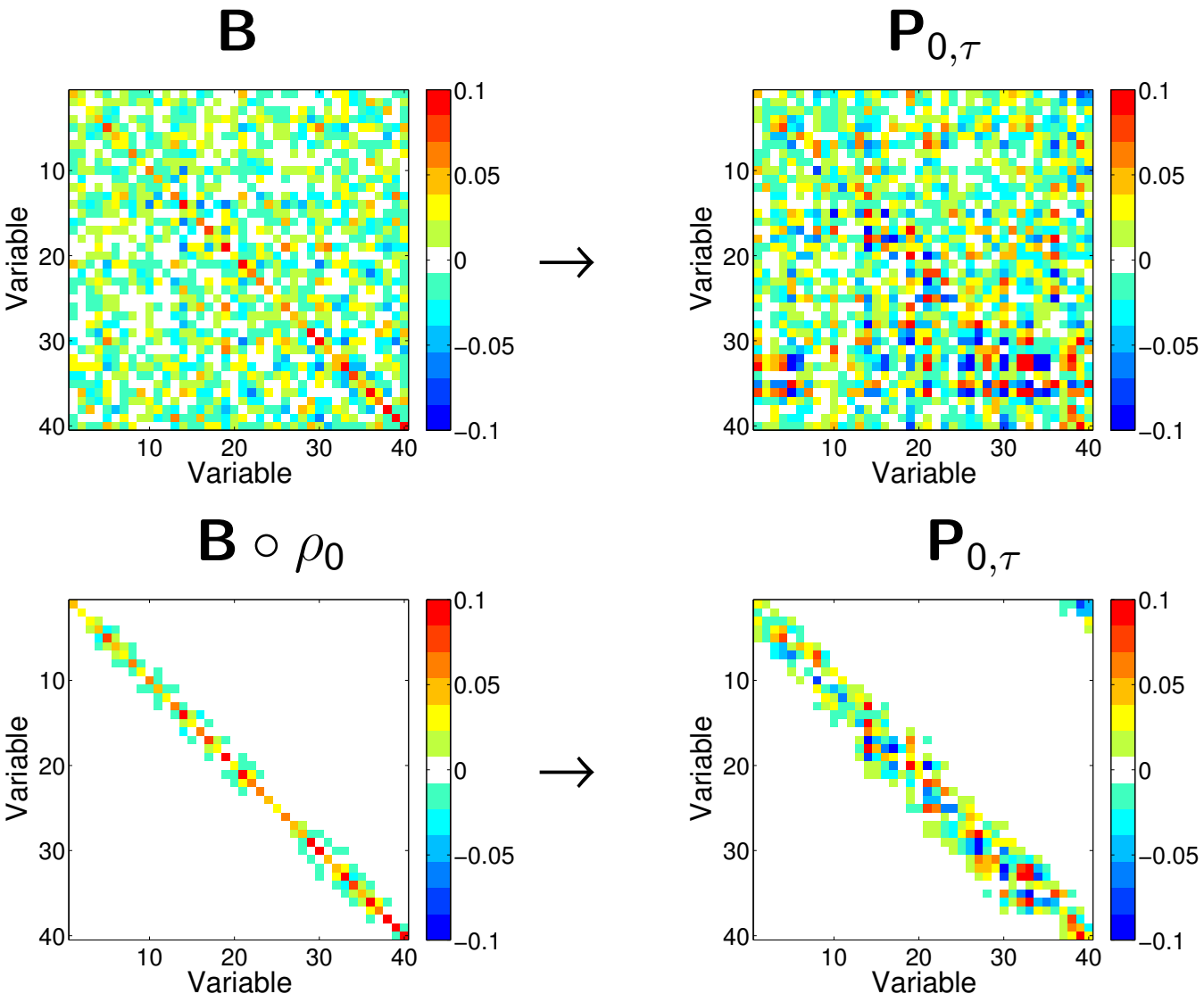
Equivalent time-dependent localization

E4DVar becomes a better proxy for 4DEnVar, when $\mathbf{M}_\tau$ uses the perfect base state $\mathbf{x}_0^b$, rather than a sample mean.



The E4DVar method can estimate the ρ required at $t = \tau$ that gives the same result as localizing $\mathbf{B}$.

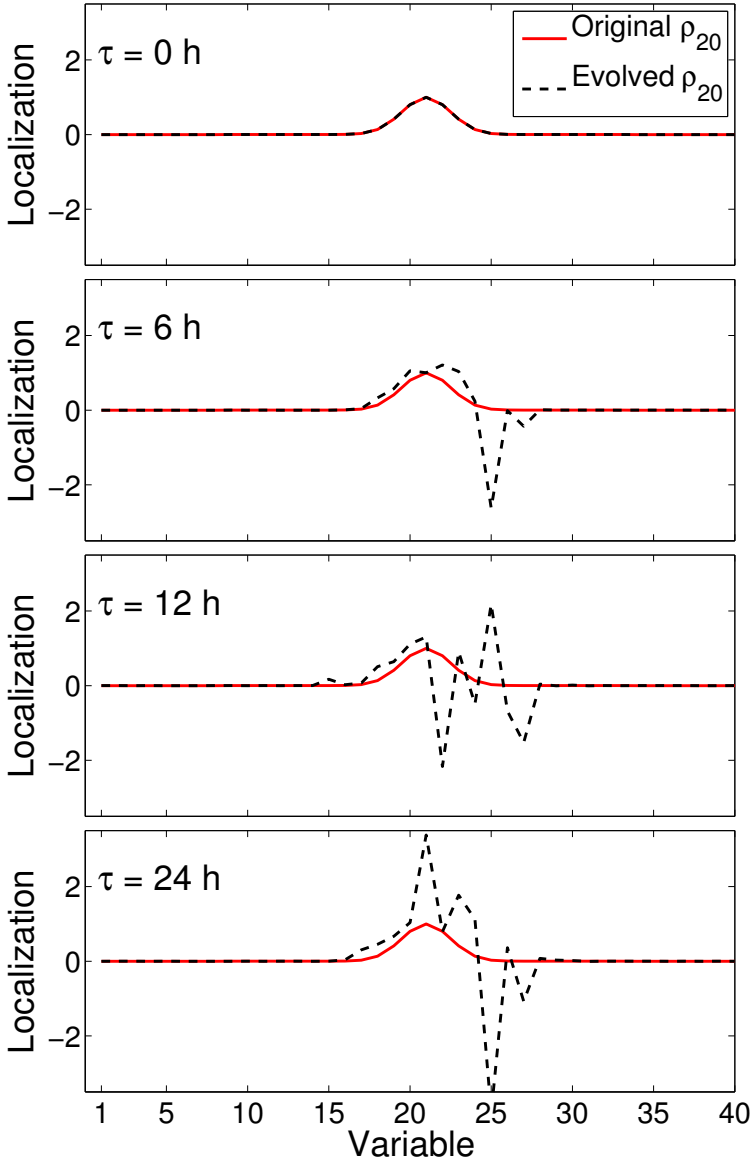
Covariance with and without localization



Time-dependent ρ for variable 20

The function required to localize the covariance between variable 20 and the remaining state variables in $\mathbf{P}_{0,\tau}$ is plotted (dashed black) for $\tau = 0, 6, 12$ and 24 h.

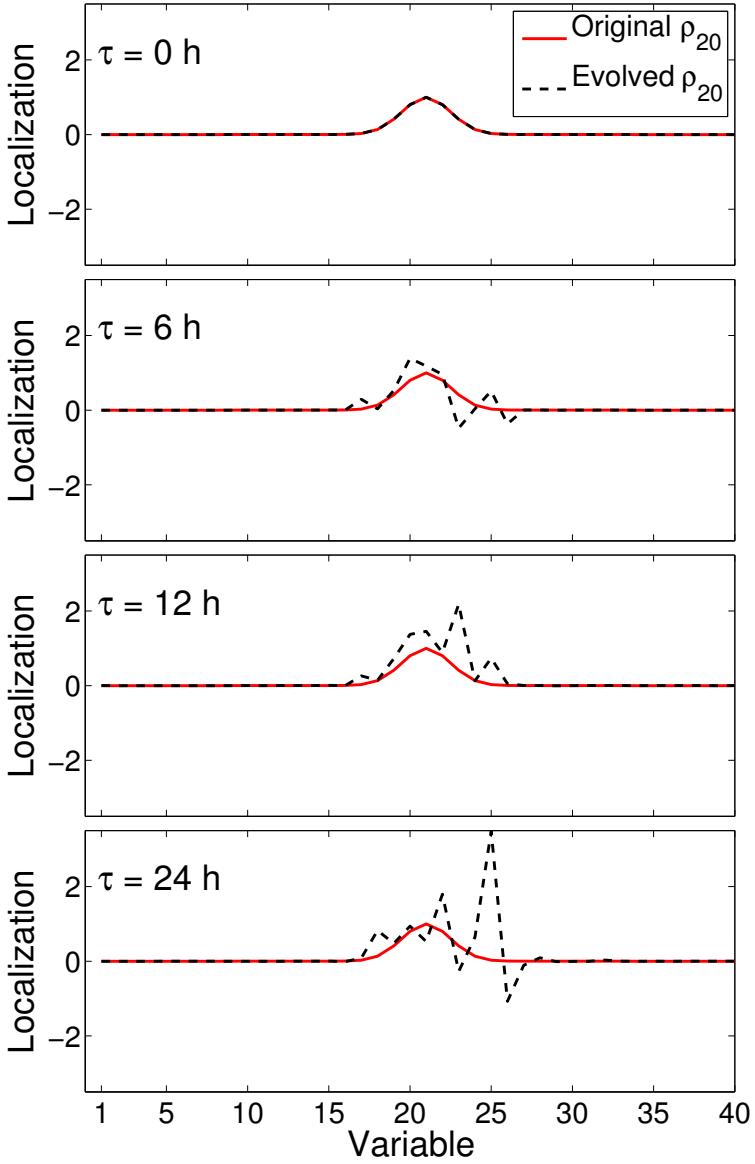
The initial correlation function is plotted in red.



Time-dependent ρ for variable 20

The experiment is repeated using 10 new random samples.

The time-dependent localization associated with the ρ used at $t = 0$ depends on the additive corrections made to **B** during localization to remove spurious correlations—this relationship is not easily parameterized.



Summary

In full ensemble mode with no static covariance component, E4DVar consistently outperforms 4DEnVar for the Lorenz (1996) system.

Experiments using ensembles and a tangent linear model to estimate time-dependent covariance show that E4DVar largely from its localization strategy.

An example using the tangent linear model to estimate the 4-D localization needed to achieve the same result as **B**-localization in E4DVar demonstrate major challenges for 4DEnVar.