Ensemble-Smoothing under the Influence of Nonlinearity

Lars Nerger[1], Svenja Schulte[2], and Angelika Bunse-Gerstner[2]

1) Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
2) Center for Industrial Mathematics (ZeTeM), University of Bremen, Bremen, Germany
Contact: Lars.Nerger@awi.de

Introduction

Ensemble-smoothing can be used as a cost-efficient addition to ensemble square root Kalman filters to improve a reanalysis in data assimilation. To correct a past state estimate, the smoothing method utilizes the cross-covariances between the present filtered state ensemble and a past ensemble at the time instance where the smoothing should be performed. Using the cross-covariances relies on the assumption that the dynamics of the system under consideration are linear. For nonlinear models, it can be expected that the smoothing is suboptimal. In this study, twin assimilation experiments are used to assess the influence of nonlinearity on the performance of the smoother.

Filter and Smoother

For ensemble square-root filters (e.g. ETKF, ESTKF, SEIK) one can write:

Filter analysis step at time t_k $\quad \mathbf{X}^a_{k|k} = \mathbf{X}^f_{k|k-1}\mathbf{G}_k$

Notation:

State vector $\mathbf{x}^f \in \mathbb{R}^n$; Ensemble of N members $\mathbf{X}^f = \left[\mathbf{x}^{f(1)}, \ldots, \mathbf{x}^{f(N)}\right]$
$i|j$ denotes state at time i conditioned on observations up to time j.

Single smoother step $\quad \mathbf{X}^a_{k-1|k} = \mathbf{X}^a_{k-1|k-1}\mathbf{G}_k$

The matrix $\mathbf{G}_k$ from the filter analysis at time t_k is used.

Smoothing over multiple times $\quad \mathbf{X}^a_{i|k} = \mathbf{X}^a_{i|i}\prod_{j=i+1}^{k}\mathbf{G}_j$
$(t_{i+1}, t_{i+2}, \ldots, t_k)$

Experiments

Lorenz96 Model

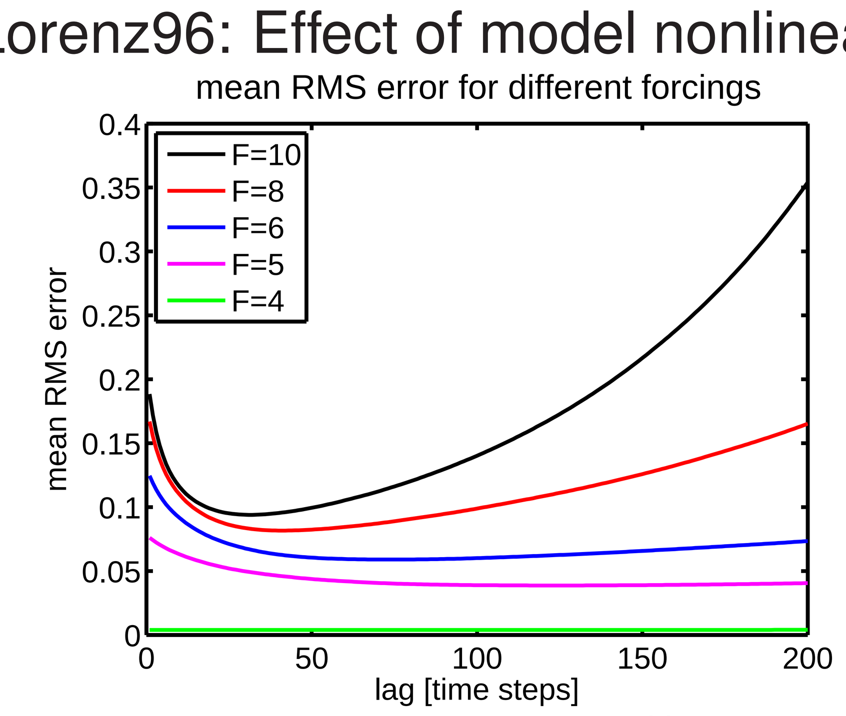
A forcing parameter F controls the nonlinearity of the Lorenz96 model [1]. Increasing F results in stronger nonlinearity. Assimilation experiments are performed over 20000 time steps with an ensemble of 34 members. Used is the Error Subspace Transform Kalman Filter (ESTKF)[2] implemented in PDAF [3, 4].

Global Ocean Model FESOM

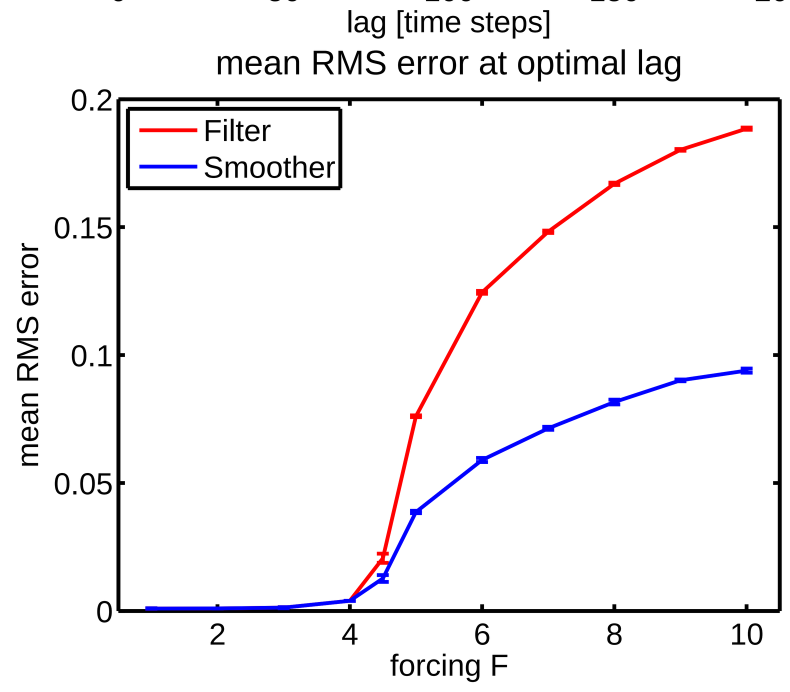
FESOM [5] is used in a global configuration. The state vector dimension is about 10 million. Synthetic observations of the sea surface height are assimilated each 10th day over one year. The ESTKF with regulated observation localization [6] is used with an ensemble of 16 members.

Assimilation with Lorenz96 Model

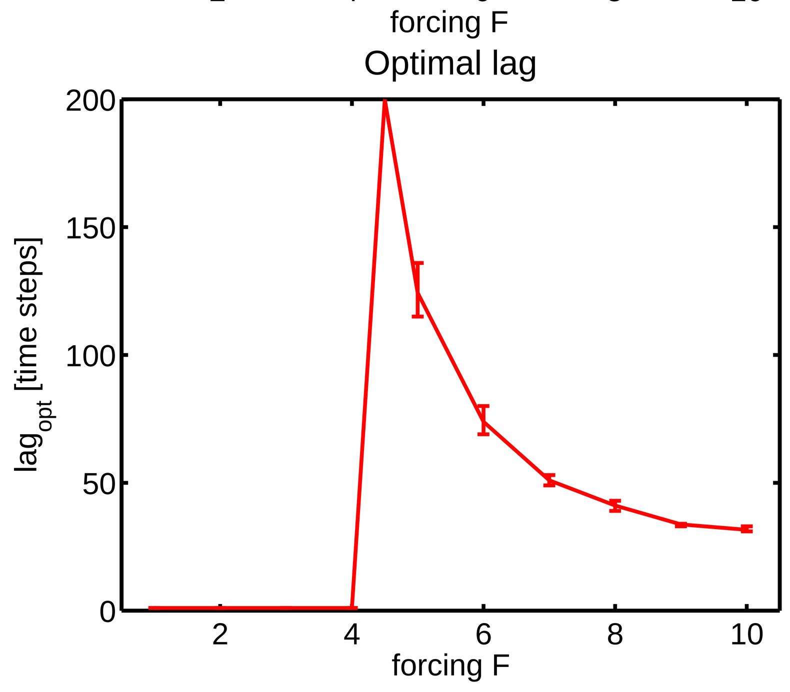
Lorenz96: Effect of model nonlinearity



Left: The model nonlinearity is controlled by the forcing F. Mean RMS errors are computed from time step 2000 to the end of each experiment. The mean RMS error increases with F. The smoother reduces the RMS error. There is an optimal lag at which the RMS error is minimal.

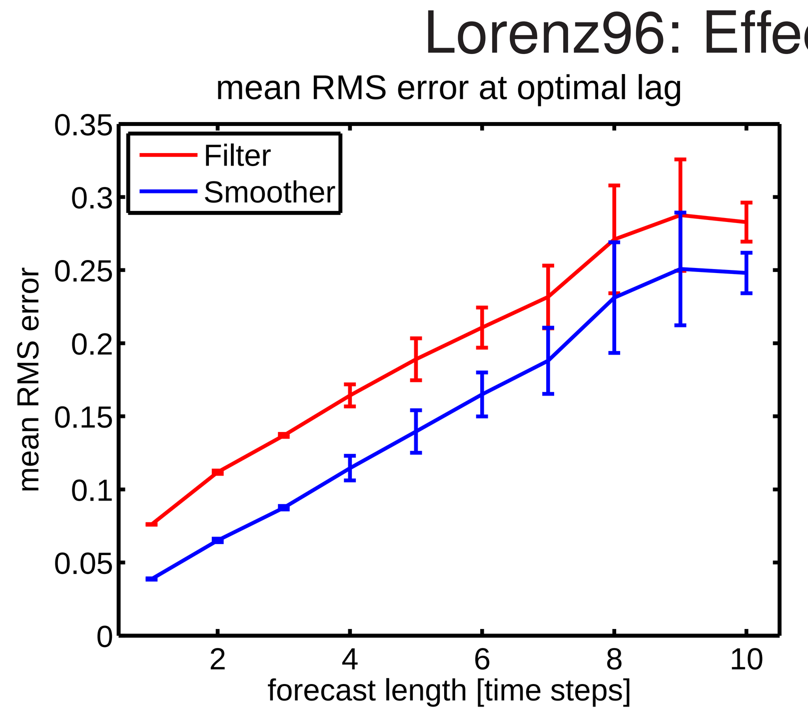


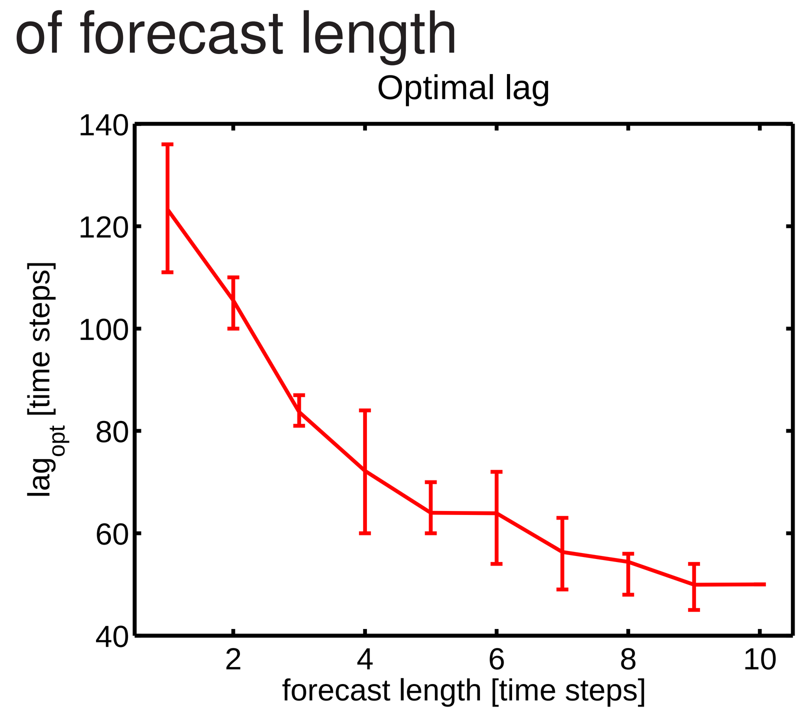
At the optimal lag, the RMS error grows strongly when F exceeds 4. At this value, the model dynamics become nonperiodic. The RMS error of the smoother is about 50% of the filter RMS error for all $F > 4$.



The RMS computation excludes the initial transient phase of the assimilation. After the transient phase, the smoother deteriorates the state estimates for the periodic cases with $F \leq 4$. Thus the optimal lag is 0. For $F > 4$ the optimal lag decreases approximately proportional to F^{-1}. The optimal lag is about 5 times the model error doubling time for $F = 10$.

Lorenz96: Effect of forecast length





Above: For fixed $F = 5$, the length of the forecast phase is increased. The RMS errors of the filter and smoother grow with the forecast length. The difference of both errors remains nearly constant. Thus, the nonlinearity of the model determines the impact of the smoother.

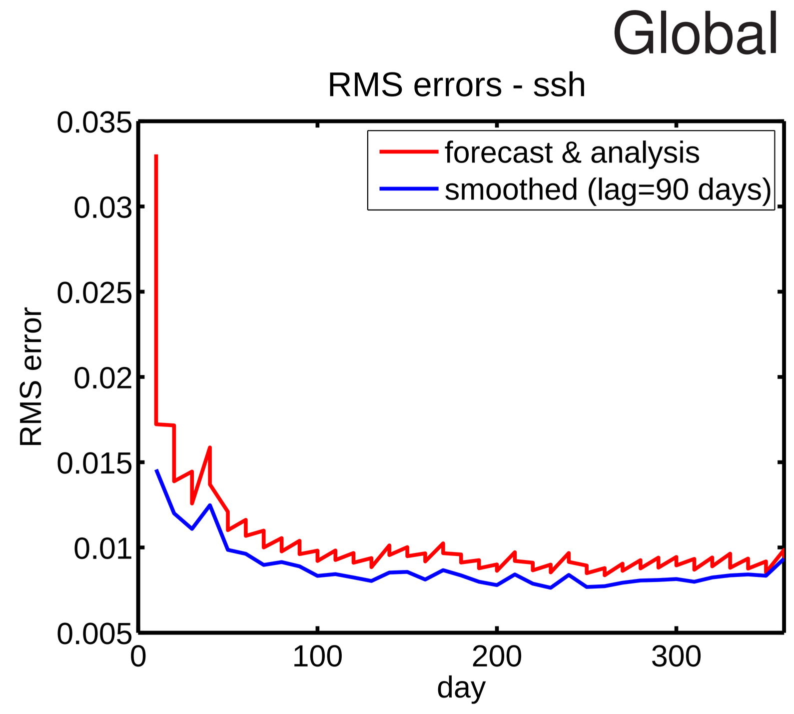
There is an optimal lag at which the error of the smoother estimate reaches its minimum. The optimal lag shrinks for growing forecast lengths from 123 to 50 time steps. The error bars show that the state estimate depends on the initial ensemble.

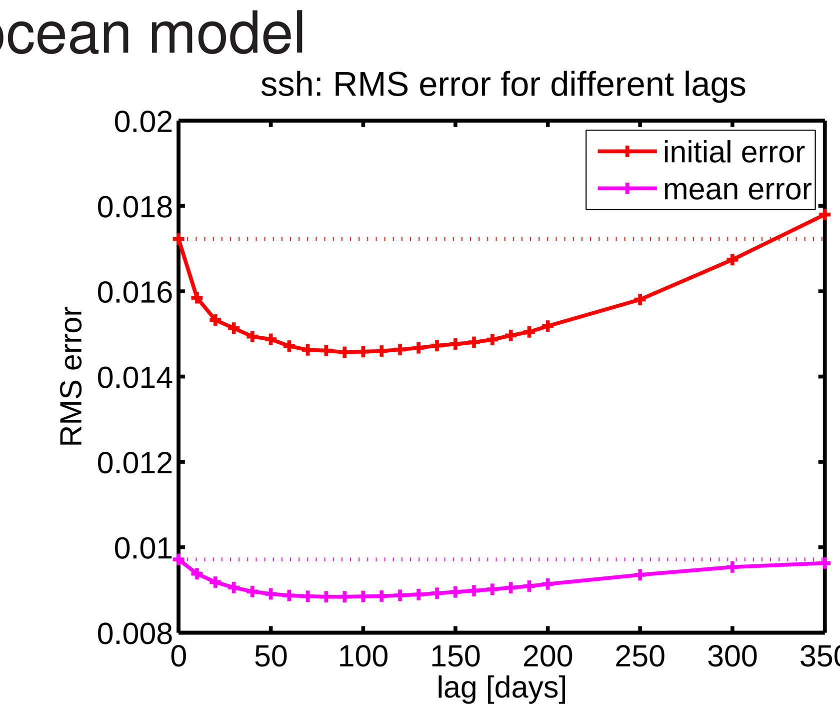
Assimilation with Global Ocean Model

Below: The assimilation of synthetic sea surface height (ssh) data significantly reduces the RMS errors of the forecasts and filter analyses. The smoother has an additional positive impact. It also improves the unobserved fields, but to a smaller extent.

The dependence of the smoother on the lag is similar to that with the Lorenz96 model. Small lags up to 50 days have a strong positive impact on the state estimate. The optimal lag for this configuration is at about 90 days.

Global ocean model





Conclusion

- Ensemble smoothing is suboptimal for nonlinear models.
- The nonlinearity determines the size of the smoother impact.
- There is an optimal smoother lag at which the positive impact of the smoother is maximal.
- The optimal lag is a multiple of the model's error-doubling time.

References

[1] E.N. Lorenz. (1996). Predictability - a problem partly solved. *Proceedings Seminar on Predictability*, ECMWF, Reading, UK, 1–18.

[2] L. Nerger et al. (2012) A Unification of Ensemble Square Root Kalman Filters. Mon. Wea. Rev. 140, 2335–2345

[3] Parallel Data Assimilation Framework (PDAF) – an open source framework for ensemble data assimilation. *http://pdaf.awi.de*

[4] L. Nerger and W. Hiller (2012) Software for Ensemble-based Data Assimilation Systems. Computers and Geosciences. In press. doi:10.1016/j.cageo.2012.03.026

[5] S. Danilov et al. (2004) A finite-element ocean model: Principles and evaluation. Ocean Modeling, 6, 125–150

[6] L. Nerger et al. (2012) A regulated localization scheme for ensemble-based kalman filters. Q. J. Roy. Meteor. Soc., 138, 802–812