Robust ensemble Kalman filtering under observation noise misspecification

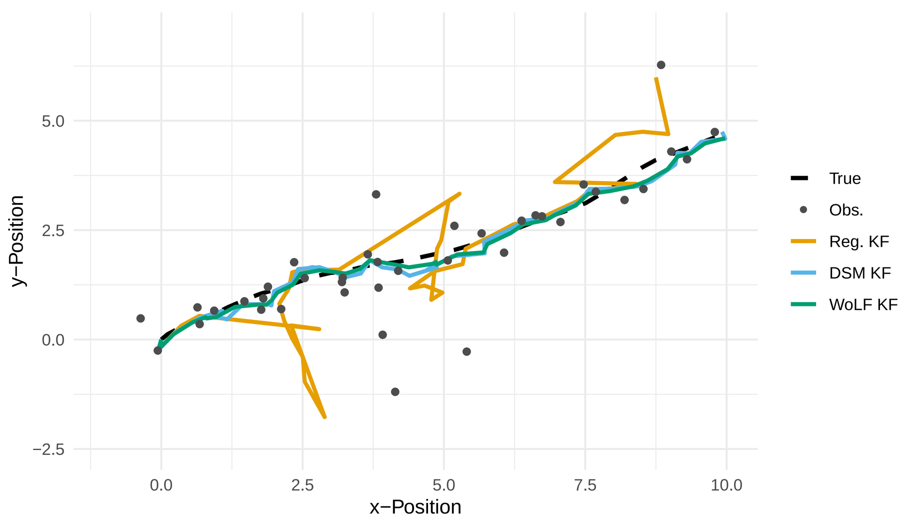
Contact: Hans Reimann
reimann@math.uni-heidelberg.de
Supervisor: Sebastian Reich

SFB 1294 Data Assimilation

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Summary: TL;DR

- **Context:** Bayesian filtering of dynamical systems
- **Problem:** Observation noise misspecification
- **Approach:** Generalized Bayesian inference with DSM
- **Theory:** Conjugacy, robustness, cov. stability and tuning in LGSS
- **Experiments:** EnKF in Lorenz 63 and LETKF in 40D Lorenz 96



Set up

For theory, assume linear, time discrete, time varying signal evolution equation and linear observation equation:

(S) $X_n = A_n X_{n-1} + C_n W_n$

(O) $Y_n = H_n X_n + \Gamma_n V_n$

- X_n in $\mathbb{R}^d$, Y_n in $\mathbb{R}^p$,
- W_n and V_n - independent standard Gaussian RVs,
- non-singular $Q_n = C_n C_n^T$ and $R_n = \Gamma_n \Gamma_n^T$, and
- $p(x_0) \sim n(x_0; m_0, P_0)$.

Ex.: Partially observed target tracking in two dimensions.

Problem

Bayesian learning is **information processing** (see [6]) via

$$p(x_n|y_{1:n}) = p(x_n|y_{1:(n-1)}) \times \exp\left(-\widehat{\mathrm{KL}}\left[p(y_n)\|p(y_n|x_n)\right]\right).$$

The KL divergence needs well-specified models:

"With a finite amount of samples [...] Bayes' posterior is highly sensitive to outliers in the data." [4]

For **Bayesian filtering**, this results in sensitivity of the analysis distribution in tail misspecification of the observation model.

Ex.: Assume Gaussian observation noise to employ KF; true DGP has heavy-tailed noise.

Idea: Utilize a generalized posterior [3] and replace KL divergence with diffusion score matching motivated by [1, 2].

Theory

Proposition 1 ***DSM analysis step***
The analysis step for the DSM Kalman filter is given via $p_{\mathrm{DSM}}(x_n|y_{1:n}) \sim n(x; m_n^a, P_n^a)$ with

$$N_n(y_n)^{-1} = 2k_n^2(y_n)R_n^{-1}$$
$$\tilde{y} = y_n - N_n(y_n)\nabla_{y_n} 2k_n^2(y_n)$$
$$\tilde{K}_n(y_n) = P_n^f H_n^T[N_n(y_n) + H_n P_n^f H_n^T]^{-1}$$
$$P_n^a = P_n^f - \tilde{K}_n(y_n) H_n P_n^f$$
$$m_n^a = m_n^f - \tilde{K}_n(y_n)[H_n m_n^f - \tilde{y}_n].$$

Algorithm 1 ***DSM Kalman filter***
1. *Forecast Step:*
   - *see regular Kalman filter*
2. *Analysis Step:*
   - *update according to Prop. 1*

Proposition 2 ***Global bias robustness***
$p_{\mathrm{DSM}}(x_n|y_{1:n})$ is globally bias robust for weight function $k : \mathcal{Y} \to \mathbb{R}$

$$k_n(y_n) = \left(1 + \frac{\|y_n - H_n m_n^f\|^2_{(H_n P_n^f H_n^T + R_n)^{-1}}}{q^2}\right)^{-\frac{1}{2}}$$

with tuning parameter $q > 0$.

Theorem 1 ***Asymptotic stability*** *via [5]*
Grant regular stability assumptions. Assume finite second moment for the DGP, then P_n^a is weakly stochastically bounded. Additionally assuming strict stationarity of the DGP, then P_n^a has a unique invariant measure and approaches it exponentially fast.

Proposition 3 ***Tuning in high dimensions***
Grant the LGSS is well-specified. For choosing the threshold parameter $q2 = d_Y$, the analysis precision of the DSM Kalman filter is an asymptotically unbiased w.r.t. the regular analysis precision for $d_Y \to \infty$.

Experiments

EnKF varieties in Lorenz 63 with x observed

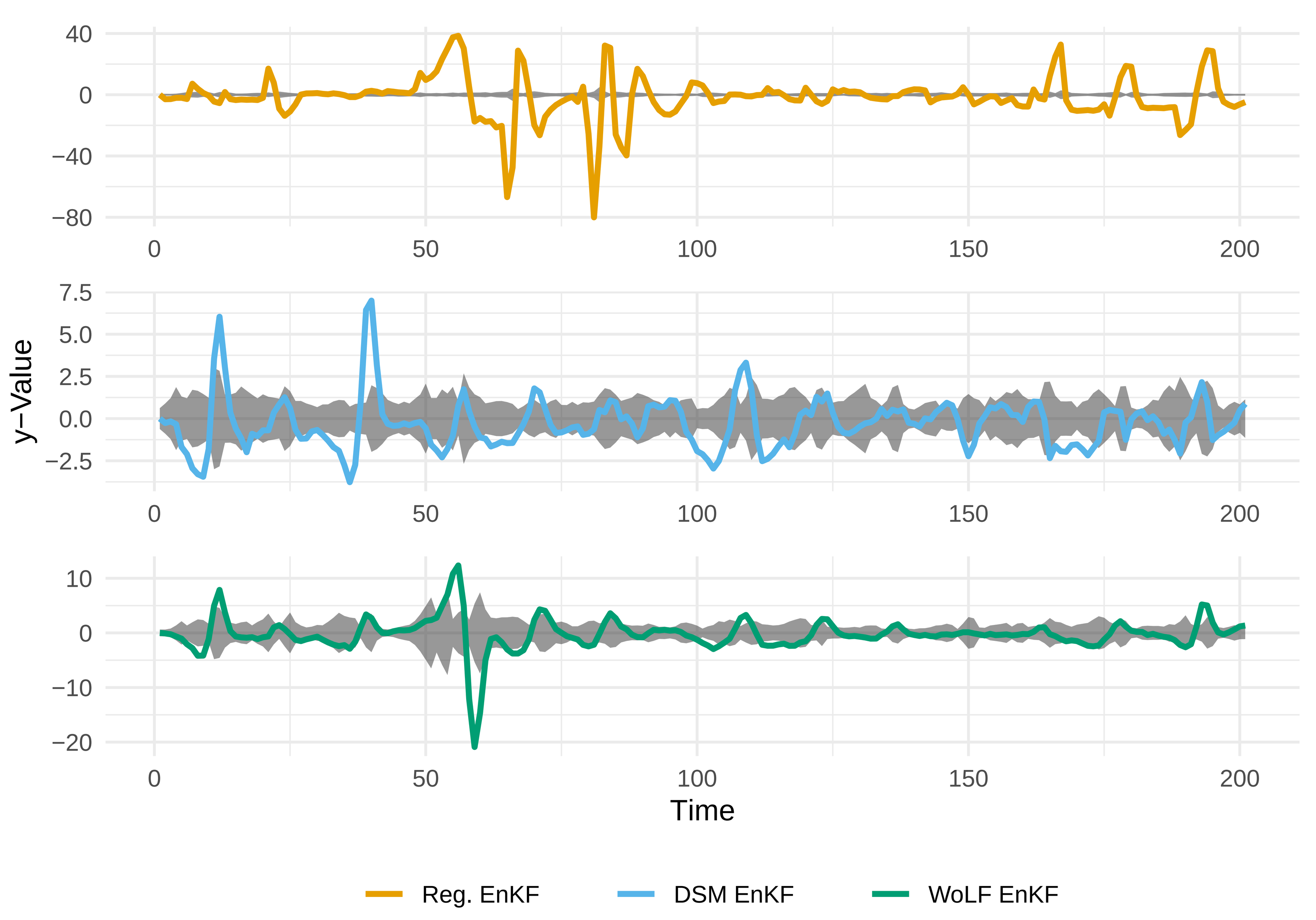


Figure 1: Centred trajectories of the unobserved y-component and their Gaussian approximation 95%-CIs in the contaminated model. .

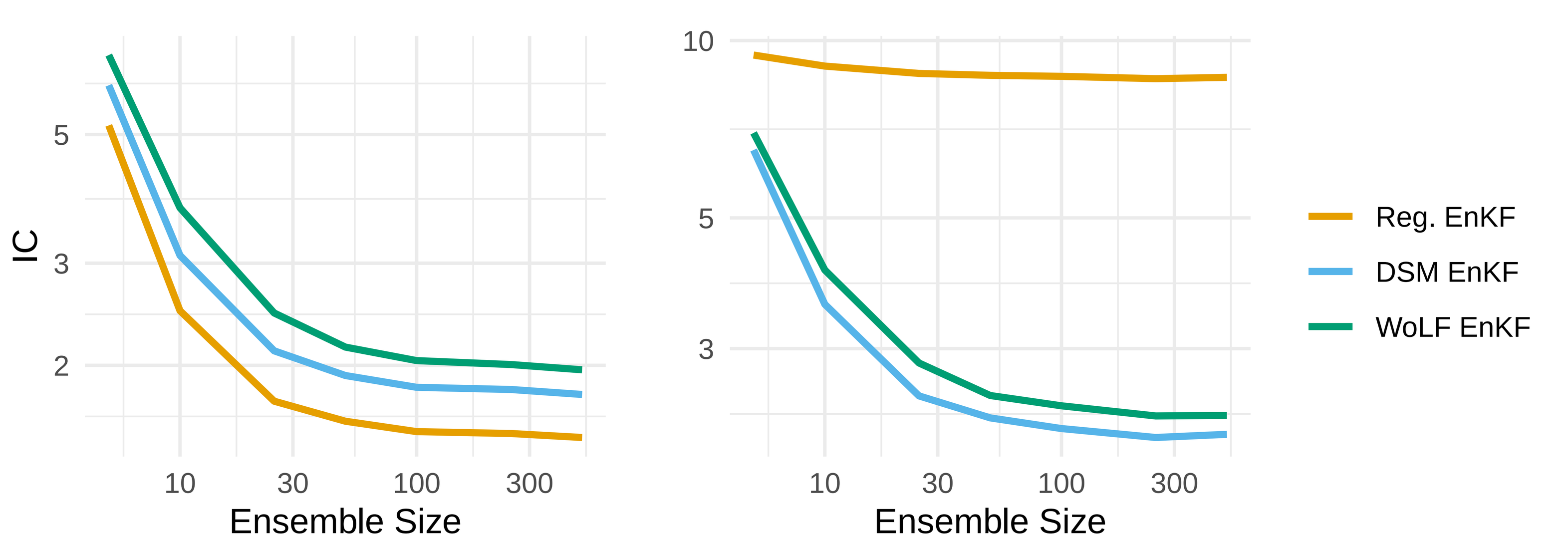


Figure 2: Averaged IC over M = 1000 repetitions for different ensemble sizes in the well-specified (left) and contaminated model (right).

LETKF varieties in Lorenz 96 with 40D

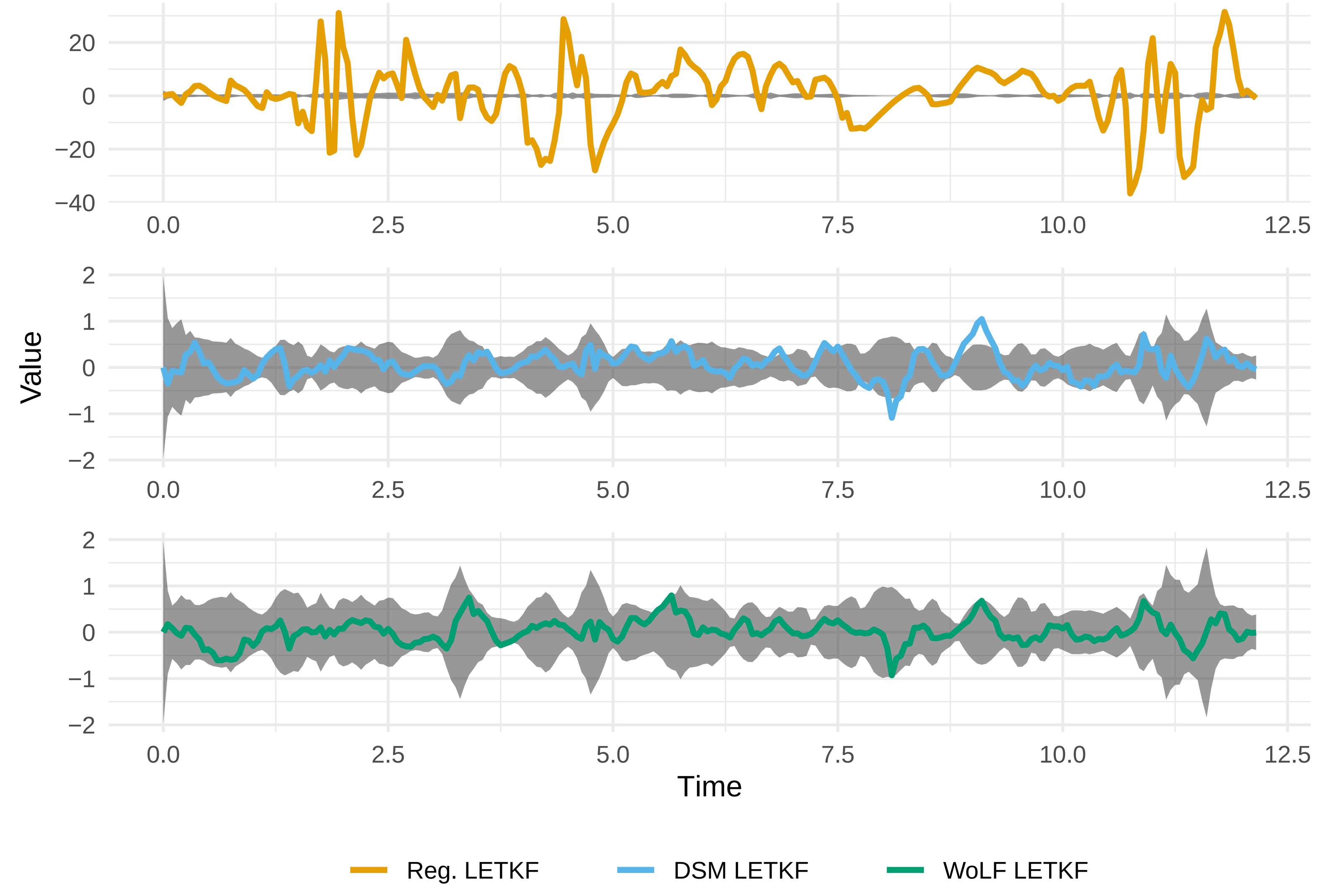


Figure 3: Centred trajectories of the first component and their Gaussian approximation 95%-CIs in the contaminated model.

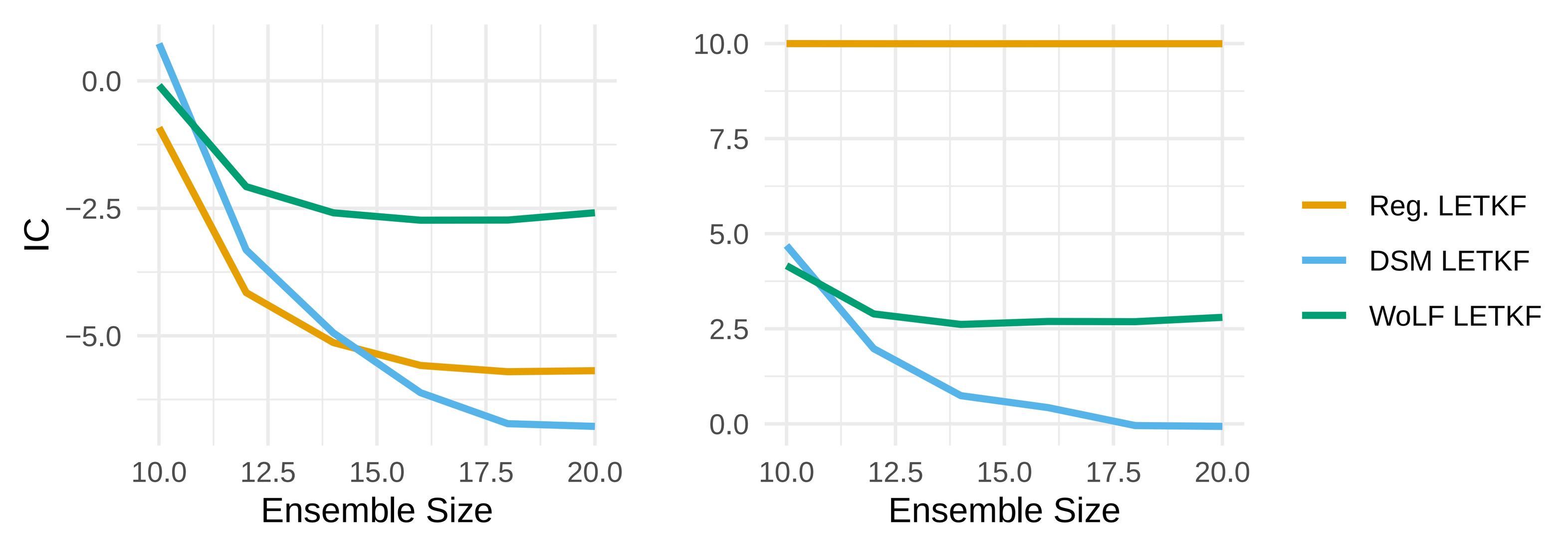


Figure 4: Averaged IC over M = 100 repetitions for different ensemble sizes in the well-specified (left) and contaminated model (right).

References

[1] M. Altamirano, F.-X. Briol, and J. Knoblauch. Robust and scalable bayesian online changepoint detection. *arXiv preprint arXiv:2302.04759*, 2023.

[2] M. Altamirano, F.-X. Briol, and J. Knoblauch. Robust and conjugate gaussian process regression. *arXiv preprint arXiv:2311.00463*, 2023.

[3] J. Jewson, J. Q. Smith, and C. Holmes. Principles of bayesian inference using general divergence criteria. *Entropy*, 20(6):442, 2018.

[4] L. Pacchiardi. Generalizing bayesian inference, Aug 2021. URL `http://www.lorenzopacchiardi.me/blog/2021/generalizedBayes/`.

[5] V. Solo. Stability of the kalman filter with stochastic time-varying parameters. In *Proceedings of 35th IEEE Conference on Decision and Control*, volume 1, pages 57–61. IEEE, 1996.

[6] A. Zellner. Optimal information processing and bayes's theorem. *The American Statistician*, 42(4):278–280, 1988.