Data assimilation as a control problem

Sebastian Reich

Universität Potsdam/ University of Reading

IPAM, May 8th 2018

Model:

$$\mathrm{d}x_t = f(x_t, \lambda)\,\mathrm{d}t + \sigma(x_t) \circ \mathrm{d}W_t, \qquad (x_0, \lambda) \sim \pi_0$$

Data/Observations:

(A) continuous-in-time

$$\mathrm{d}y_t = h(x_t, \lambda)\,\mathrm{d}t + R^{1/2}\mathrm{d}V_t$$

(B) discrete-in-time

$$y_{t_n} = h(x_{t_n}, \lambda) + R^{1/2}\,\Xi_{t_n}\,.$$

Goal: Approximate conditional PDF $\tilde{\pi}_t(x,\lambda) = \pi_t(x,\lambda|Y_t)$ where Y_t contains all the data up to time t and $\tilde{\pi}_0 = \pi_0$ at initial time.

Particle filters: Approximate conditional PDF $\tilde{\pi}_t$ by a set of particles

$$z_t^i = (x_t^i, \lambda_t^i), \qquad i = 1, \ldots, M,$$

with weights

$$w_t^i \geq 0, \qquad \text{s.t.} \sum_i w_t^i = 1\,.$$

E.g. sequential Monte Carlo.

Eulerian:

$$z_t^i = z_0^i \qquad \text{(particle locations are fixed)}$$

E.g., grid-based methods.

Lagrangian:

$$w_t^i = 1/M \qquad \text{(weights are fixed)}$$

subject of this talk.

Particle filters: Approximate conditional PDF $\tilde{\pi}_t$ by a set of particles

$$z_t^i = (x_t^i, \lambda_t^i), \qquad i = 1, \ldots, M,$$

with weights

$$w_t^i \geq 0, \qquad \text{s.t.} \sum_i w_t^i = 1.$$

E.g. sequential Monte Carlo.

Eulerian:

$$z_t^i = z_0^i \qquad \text{(particle locations are fixed)}$$

E.g., grid-based methods.

Lagrangian:

$$w_t^i = 1/M \qquad \text{(weights are fixed)}$$

subject of this talk.

A control approach to the continuous-in-time filtering problem:

Kalman gain factor:

$$K_t = P_{xh} R^{-1} \,,$$

P_{xh} denotes the covariance matrix between x_t and $h(x_t)$

Innovation:

$$\mathrm{d}I_t = \mathrm{d}y_t - \frac{1}{2}(h(x_t) + \bar{h}_t)\,\mathrm{d}t$$

or

$$\mathrm{d}I_t = \mathrm{d}y_t - h(x_t)\,\mathrm{d}t - R^{1/2}\mathrm{d}U_t$$

Ensemble Kalman-Bucy filter:

$$\mathrm{d}x_t = f(x_t, \lambda)\,\mathrm{d}t + \sigma(x_t) \circ \mathrm{d}W_t + K_t \circ \mathrm{d}I_t$$

Consider zero-drift & scalar SDE

$$\mathrm{d}x_t = \sigma(x_t) \circ \mathrm{d}W_t$$

with time-evolved expectation values:

$$\pi_t[f] = \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f]\,\mathrm{d}s, \qquad \mathcal{L}f := \frac{1}{2}\sigma\,\partial_x(\sigma\,\partial_x f)$$

Interacting particle representation:

$$\dot{x}_t = \frac{1}{2}\sigma(x_t) I_t(x_t), \qquad I_t := -\pi_t^{-1}\partial_x(\pi_t\,\sigma)$$

It holds that (Liouville plus integration by parts)

$$\hat{\pi}_t[f] := \pi_0[f] + \frac{1}{2}\int_0^t \pi_s[(\partial_x f)(\sigma I_s)]\,\mathrm{d}s$$

$$= \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f]\,\mathrm{d}s = \pi_t[f]\,.$$

Consider zero-drift & scalar SDE

$$\mathrm{d}x_t = \sigma(x_t) \circ \mathrm{d}W_t$$

with time-evolved expectation values:

$$\pi_t[f] = \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f]\,\mathrm{d}s, \qquad \mathcal{L}f := \frac{1}{2}\sigma\,\partial_x(\sigma\,\partial_x f)$$

Interacting particle representation:

$$\dot{x}_t = \frac{1}{2}\sigma(x_t) I_t(x_t), \qquad I_t := -\pi_t^{-1}\partial_x(\pi_t\,\sigma)$$

It holds that (Liouville plus integration by parts)

$$\hat{\pi}_t[f] := \pi_0[f] + \frac{1}{2}\int_0^t \pi_s[(\partial_x f)(\sigma I_s)]\,\mathrm{d}s$$
$$= \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f]\,\mathrm{d}s = \pi_t[f]\,.$$

Generalization of the ensemble Kalman-Bucy filter:

$$\mathrm{d}x_t = f(x_t, \lambda)\,\mathrm{d}t + \sigma(x_t) \circ \mathrm{d}W_t + K_t \circ \mathrm{d}I_t$$

Innovation $\mathrm{d}I_t$ as before, i.e.,

$$\mathrm{d}I_t = \mathrm{d}y_t - \frac{1}{2}(h(x_t) + \bar{h}_t)\,\mathrm{d}t$$

or

$$\mathrm{d}I_t = \mathrm{d}y_t - h(x_t)\,\mathrm{d}t - R^{1/2}\mathrm{d}U_t\,,$$

Kalman gain matrix $K_t = \nabla\psi_t$ with

$$-\nabla_x \cdot (\hat{\pi}_t \nabla_x \psi_t) = R^{-1}\hat{\pi}_t(h - \tilde{\pi}_t[h])\,.$$

It can be shown that $\tilde{\pi}_t[f] = \hat{\pi}_t[f]$.

The time-evolution of an ensemble of M particles x_t^i is given by

$$\mathrm{d}x_t^i = f(x_t^i, \lambda)\,\mathrm{d}t + \sigma(x_t^i) \circ \mathrm{d}W_t^i - \tilde{K}_t^i \circ \mathrm{d}I_t^i$$

Gain:

$$\tilde{K}_t^i := \sum_{j=1}^{M} x_t^j d_{ji} \approx \nabla_x \psi_t(x_t^i)$$

Innovation:

$$\mathrm{d}I_t^i = \mathrm{d}y_t - \frac{1}{2}(h(x_t^i) + \bar{h}_t)\,\mathrm{d}t$$

or

$$\mathrm{d}I_t^i = \mathrm{d}y_t - h(x_t^i)\,\mathrm{d}t - R^{1/2}\mathrm{d}U_t^i,$$

References

EnKBF:

- SR (BIT, 2011), Bergemann & SR (Meteorol. Zeitschrift, 2012)
- de Wiljes, Stannat & SR (ArXiv:1612.06065, 2016),
- Del Moral, Kurtzmann & Tugaut (ArXiv:1606.082566, 2016)

FPF:

- Yang, Mehta & Meyn (IEEE Trans. Autom. Contr., 2013)
- Taghvaei, de Wiljes, Mehta & SR (ArXiv:1702.07241),

Alternative control formulation:

- Crisan & Xiong (ESAIM, 2007), Crisan & Xiong (Stochastics, 2010)

Discrete–time observations:

$$y_{t_n} = h(Z_{t_n}) + R^{1/2}\,\Xi_{t_n}, \qquad n = 1,\ldots,N\,.$$

Likelihood function:

$$L(z_{[0,T]}) := \exp\left(-\frac{1}{2}\sum_n (y_{t_n} - h(z_{t_n}))^{\mathrm{T}} R^{-1} (y_{t_n} - h(z_{t_n}))\right).$$

Bayes:

$$\frac{\mathrm{d}\widehat{\mathbb{P}}_{[0,T]}}{\mathrm{d}\mathbb{P}_{[0,T]}}(Z^+_{[0,T]}) := \frac{L(Z^+_{[0,T]})}{\mathbb{P}_{[0,T]}[L]}\,.$$

The measure $\widehat{\mathbb{P}}_{[0,T]}$ solves the **filtering/smoothing problem** of SDE inference.

Discrete–time observations II

For simplicity: **single observation**, i.e.

$$N = 1, \quad R = I, \quad t_1 = T, \qquad L(z) = \frac{1}{2}\|y_T - h(z)\|^2 \, .$$

But keep recursive nature of sequential DA in mind!

Four **main players**:

- last analysis: π_0
- forecast based on last analysis: π_T
- new analysis at time $t = T$ (Bayes, filtering distribution): $\widehat{\pi}_T$
- smoothing distribution at $t = 0$: $\widehat{\pi}_0$

Standard sequential DA leads to a **discontinuous change** in distributions at observation time $t = T$ from π_T to $\widehat{\pi}_T$.

Forward SDE

$$dZ_t^+ = f_t(Z_t^+)dt + \sigma dW_t^+ ,$$

$X_0^+ \sim \pi_0$, $t \in [0, T]$, W_t^+ standard Brownian motion forward in time.

Generates **probability measure** $\mathbb{P}_{[0,T]}$ over $C([0,T], \mathbb{R}^{N_z})$ with **marginal densities** π_t, i.e. $Z_t \sim \pi_t$.

The same measure is generated by **backward SDE**

$$dZ_t^- = b_t(Z_t^-)dt + \sigma dW_t^- ,$$

W_t^- Brownian motion backward in time, $X_T^- \sim \pi_T$.

It holds that

$$b_t(z) = f_t(z) - \sigma^2 \nabla_z \log \pi_t(z) .$$

Forward SDE

$$dZ_t^+ = f_t(Z_t^+)dt + \sigma dW_t^+ ,$$

$X_0^+ \sim \pi_0$, $t \in [0, T]$, W_t^+ standard Brownian motion forward in time.

Generates **probability measure** $\mathbb{P}_{[0,T]}$ over $C([0,T], \mathbb{R}^{N_z})$ with **marginal densities** π_t, i.e. $Z_t \sim \pi_t$.

The same measure is generated by **backward SDE**

$$dZ_t^- = b_t(Z_t^-)dt + \sigma dW_t^- ,$$

W_t^- Brownian motion backward in time, $X_T^- \sim \pi_T$.

It holds that

$$b_t(z) = f_t(z) - \sigma^2 \nabla_z \log \pi_t(z) .$$

Fokker-Planck equation for marginals:

$$\begin{aligned}\partial_t \pi_t &= -\nabla_z \cdot (\pi_t f_t) + \frac{\sigma^2}{2} \Delta_z \pi_t \\ &= -\nabla_z \cdot (\pi_t b_t) - \frac{\sigma^2}{2} \Delta_z \pi_t \\ &= -\nabla_z \cdot (\pi_t u_t)\end{aligned}$$

with

$$u_t(z) = \frac{1}{2}(f_t(z) + b_t(z)) = f_t(z) - \frac{\sigma^2}{2} \nabla_z \log \pi_t(z) \,.$$

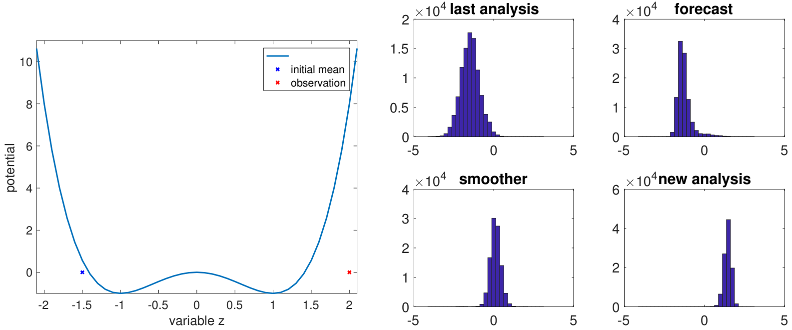
Replace forward SDE by **mean field equation**

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t = f_t(Z_t) - \frac{\sigma^2}{2} \nabla_z \log \pi_t(Z_t) \,, \qquad Z_0 \sim \pi_0 \,.$$

Remark. Generates path measure $\mathbb{Q}_{[0,T]}$ which is different from SDE measure $\mathbb{P}_{[0,T]}$; only marginals π_t agree!

Example

Scalar Brownian dynamics under a double well potential ($\sigma^2 = 0.5$):



The forecast and the new analysis at $T = 0.5$ are nearly singular with respect to each other.

The relation between the last analysis (π_0) and the smoother ($\widehat{\pi}_0$) is somewhat better. Exploited in optimal proposal density/auxiliary particle filters.

Discrete–time observations III

Forward-backward smoother iteration:

- **Forward**:

$$\mathrm{d}Z_t^+ = f(Z_t^+)\mathrm{d}t + \sigma W_t^+ ,$$

$Z_0^+ \sim \pi_0$. **Yields** π_t.

- **Backward**:

$$\mathrm{d}\widehat{Z}_t^- = f(\widehat{Z}_t^-)\mathrm{d}t - \sigma^2 \nabla_z \log \pi_t(\widehat{Z}_t^-)\mathrm{d}t + \sigma W_t^- ,$$

with $\widehat{Z}_T^- \sim \widehat{\pi}_T$ and

$$\widehat{\pi}_T(z) \propto L(z)\pi_T(z) .$$

Yields $\widehat{\pi}_t$.

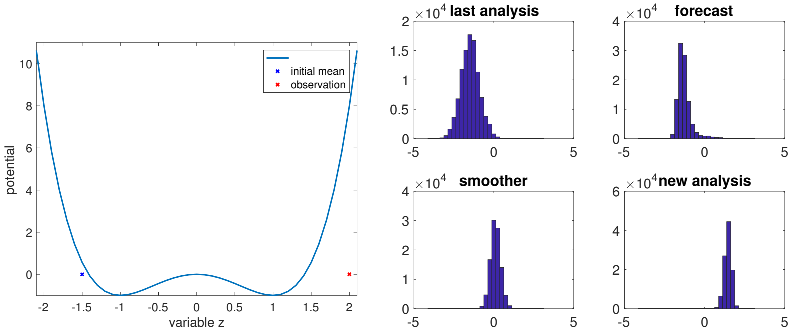
Smoother:

$$\mathrm{d}\widehat{Z}_t^+ = f(\widehat{Z}_t^+)\mathrm{d}t + \sigma^2 \nabla_z \log \frac{\widehat{\pi}_t}{\pi_t}(\widehat{Z}_t^+)\mathrm{d}t + \sigma W_t^+$$

$\widehat{Z}_0^+ \sim \widehat{\pi}_0$, $\widehat{Z}_T^+ \sim \widehat{\pi}_T$.

Example

Scalar Brownian dynamics under a double well potential ($\sigma^2 = 0.5$):



The forward smoother SDE links the smoother measure $\widehat{\pi}_0$ with $\widehat{\pi}_T$.

Still requires transforming π_0 into $\widehat{\pi}_0$ (but now at $t = 0$).

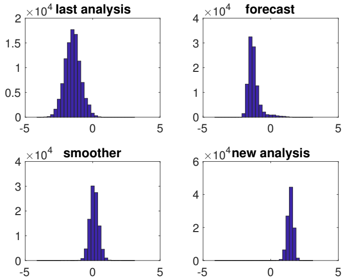
A different perspective on sequential DA:

Schrödinger problem. Find the measure $\widetilde{\mathbb{P}}_{[0,T]}$ which minimises the Kullback-Leibler divergence

$$\widetilde{\mathbb{P}}_{[0,T]} = \arg \inf_{\mathbb{Q} \ll \mathbb{P}} \mathrm{KL}(\mathbb{Q}_{[0,T]} || \mathbb{P}_{[0,T]})$$

subject to the constraints

$$\widetilde{\pi}_0 = q_0 = \pi_0, \qquad \widetilde{\pi}_T = q_T = \widehat{\pi}_T \, .$$



The measure $\widetilde{\mathbb{P}}_{[0,T]}$ is generated by a **controlled SDE**

$$\mathrm{d}\widetilde{Z}_t^+ = f(\widetilde{Z}_t^+)\mathrm{d}t + u_t(\widetilde{Z}_t^+)\mathrm{d}t + \sigma \mathrm{d}W_t^+ \, .$$

Schrödinger problem of DA II

Find an initial distribution ϕ_0^+ and its evolution ϕ_t^+ under the forward SDE

$$\mathrm{d}Z_t^+ = f(Z_t^+)\,\mathrm{d}t + \sigma \mathrm{d}W_t^+$$

such that the associated backward SDE

$$\mathrm{d}Z_t^- = \left(f(Z_t^-) - \sigma^2 \nabla_z \log \phi_t^+(Z_t^-)\right)\mathrm{d}t + \sigma \mathrm{d}W_t^-$$

with final condition $\phi_T^- := \widehat{\pi}_T$ leads to marginals ϕ_t^- such that

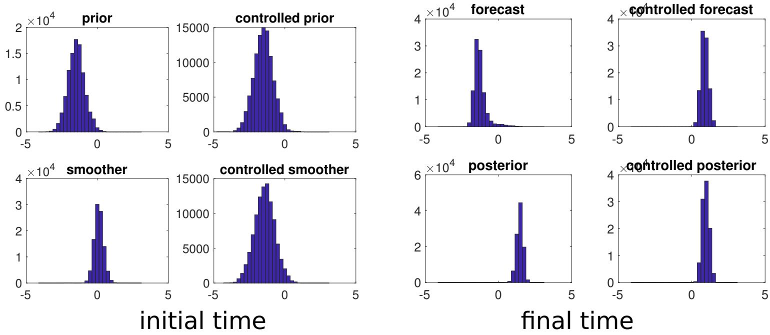
$$\phi_0^- = \pi_0\,.$$

Then the control

$$u_t = \sigma^2 \nabla_z \log \frac{\phi_t^-}{\phi_t^+}$$

solves the Schrödinger problem.

Double well potential, all densities are approximated as Gaussian (**linear, time-dependent control term**), ten iterations:



Remarks

- **Smoothing**:

$$\phi_0^+ = \pi_0 \quad \& \quad \phi_T^- = \widehat{\pi}_T \quad \Rightarrow \quad \widehat{\pi}_t = \phi_t^-$$

  Schrödinger:

$$\phi_0^- = \pi_0 \quad \& \quad \phi_T^- = \widehat{\pi}_T \quad \Rightarrow \quad \phi_t^+ \; / \; u_t$$

- Link to **Sinkhorn** and **Robbins & Monro** iterations: If

$$\pi_0(z) = \frac{1}{M}\sum_{i=1}^{M}\delta(z - z_0^i)$$

  then

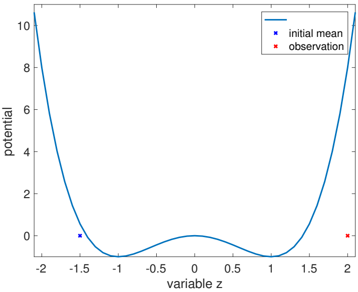
$$\phi_0^+(z) = \sum_{i=1}^{M}\alpha_i\delta(z - z_0^i), \qquad \sum_{i=1}^{M}\alpha_i = 1$$

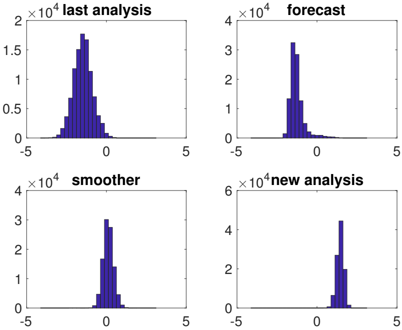
  leading to a Sinkhorn fixed point iteration in the weights $\{\alpha_i\}$ which involves taking expectation with respect to $\widehat{\pi}_T$.

References

- Schrödinger, E., Über die Umkehrung der Naturgesetze, Sitzungsberichte der Preuss. Phys. Math. Klasse, 1932.
- Yongxin Chen et al., On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint, J. Optim. Theory and Appl., 2016.
- Peyre, G., Cuturi, M., Computational Optimal Transport, arXiv:1803.00567, 2018.
- Fearnhead, P., Künsch, H.R., Particle filters and data assimilation, Annual Review of Statistics and Its Application, 2018.
- Guarniero et al., The iterated auxiliary particle filter, J. American Statist. Assoc., 2017.
- van Leeuwen, P.-J., Nonlinear data assimilation, Frontiers in Applied Dynamical Systems, Vol. 2, 2015.
- SR, Data assimilation, Acta Numerica, 2019.

We cannot, in general, implement the Schrödinger approach to sequential DA exactly.





Available realisations $Z_T^i \sim \pi_T$ with **importance weights**

$$w^i \propto \frac{\widehat{\pi}_T}{\pi_T}(Z_T^i)\,.$$

Instead of **resampling**, find **coupling/transformation**

$$\widehat{Z}_T = \nabla_z \psi(Z_T)\,,$$

$Z_T \sim \pi_T$ and $\widehat{Z} \sim \widehat{\pi}_T$.

More abstractly,

$$\widehat{Z}_T(a) = \int Z_T(a')\delta(a' - \nabla_a \psi(a))\mathrm{d}a'\,,$$

where A is some random reference variable. For example, $A = Z_T$.

Replace the integral by a sum and formally write

$$\widehat{Z}_T^j = \sum_{i=1}^{M} Z_T^i d_{ij}$$

with $a' \to i$, $Z_T(a') \to Z_T^i$, $a \to j$, $\widehat{Z}_T(a) \to \widehat{Z}_T^j$, $\delta(a' - \nabla_a \psi(a)) \to d_{ij}$. Need

$$\sum_{i=1}^{M} d_{ij} = 1, \qquad \frac{1}{M}\sum_j d_{ij} = w^i \,.$$

Select an "optimal" transformation through maximising correlation

$$V(D) = \frac{1}{M}\sum_{ij} d_{ij} Z_T^i \cdot Z_T^j = \frac{1}{M}\sum_j \widehat{Z}_T^j \cdot Z_T^j \,.$$

In addition, either $d_{ij} \geq 0$ (**Ensemble Transform Particle Filter**) or

$$\frac{1}{M-1}\sum_{i=1}^{M}(\widehat{Z}_T^i - \widehat{\bar{z}}_T)(\widehat{Z}_T^i - \widehat{\bar{z}}_T)^{\mathrm{T}} = \sum_{i=1}^{M} w^i (Z_T^i - \widehat{\bar{z}}_T)(Z_T^i - \widehat{\bar{z}}_T)^{\mathrm{T}}$$

(**Nonlinear Ensemble Transform Filter**).

Numerical example I

Lorenz-63 model, first component observed infrequently ($\Delta t = 0.12$) and with large measurement noise ($R = 8$):

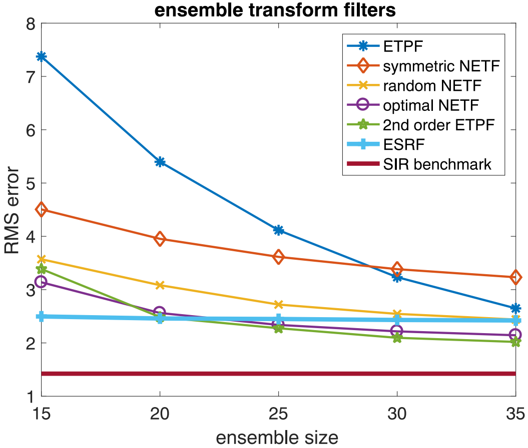


Figure: RMSEs for various second-order accurate LETFs compared to the ETPF, the ESRF, and the SIR PF as a function of the sample size, M.

Numerical example II

Hybrid filter: $\mathbf{P} := \mathbf{P}_{\mathrm{ESRF}}(\alpha)\;\mathbf{P}_{\mathrm{ETPF}}(1-\alpha)$.

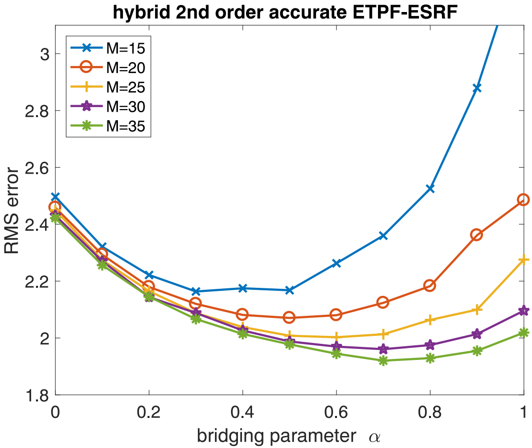


Figure: RMSEs for hybrid ESRF ($\alpha = 0$) and 2nd-order corrected NETF/ETPF ($\alpha = 1$) as a function of the sample size, M.

References

- Evensen, G., Data assimilation – the ensemble Kalman filter, Springer, 2009.
- SR, A nonparametric ensemble transform method for Bayesian inference, SIAM J. Sci. Comput., 2013.
- SR, Cotter, J., Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015.
- Tödter J., Ahrens, B., A second–order exact ensemble square root filter for nonlinear data assimilation, Month. Weather Rev., 2015.
- Chustagulprom, N. et al., A hybrid ensemble transform particle filter for nonlinear and spatially extended systems, SIAM/ASA J. UQ, 2016.
- Acevedo, W., et al., Second–order accurate ensemble transform particle filters, SIAM J. Sci. Comput., 2017.
- Fearnhead, P., Künsch, H.R., Particle filters and data assimilation, Annual Review of Statistics and Its Application, 2018.

Summary

- Continuous–in–time DA naturally leads to controlled interacting particle systems.
- Schrödinger problem provides an "optimal" mathematical framework for sequential DA with discrete–in–time observations.
- Numerical implementation nontrivial; good drift corrections can be derived using Gaussian approximations or kernel methods.
- Coupling arguments are central to derivation of interacting particle systems.
- Relevant to rare event simulations, optimal control problems and derivative–free optimization.

Collaborators

- Walter Acevedo
- Kay Bergemann
- Yuan Cheng
- Nawinda Chustagulprom
- Colin Cotter
- Jana de Wiljes
- Prashant Mehta
- Wilhelm Stannat
- Amari Taghvaei