

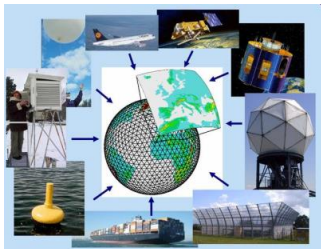


Data assimilation as an optimal control problem and applications to UQ

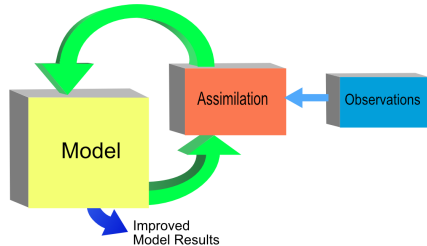
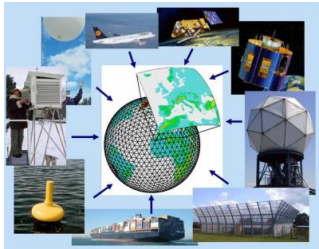
Walter Acevedo, Angwenyi David, Jana de Wiljes & Sebastian Reich

Universität Potsdam/ University of Reading

IPAM, November 13th 2017



- ▶ **Model:** highly nonlinear discretized partial differential equations
- ▶ **Data:** heterogeneous mix of ground-, airborne-, satellite-based and radar data
- ▶ 24/7 data assimilation service for optimal weather prediction



- ▶ **Model:** highly nonlinear discretized partial differential equations
- ▶ **Data:** heterogeneous mix of ground-, airborne-, satellite-based and radar data
- ▶ **24/7 data assimilation** service for optimal weather prediction

Model:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t, \quad (x_0, \lambda) \sim \pi_0$$

Data/Observations:

(A) continuous-in-time

$$dy_t = h(x_t, \lambda) dt + R^{1/2} dV_t$$

(B) discrete-in-time

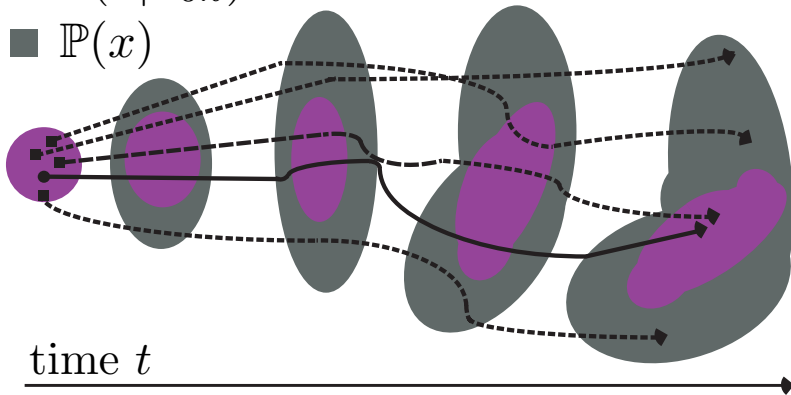
$$y_{t_n} = h(x_{t_n}, \lambda) + R^{1/2} \Xi_{t_n}.$$

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.

Time evolved marginal distributions:

■ $\mathbb{P}(x|Y_{0:t})$

■ $\mathbb{P}(x)$



Applications

- ▶ assimilation of observations into computer model (e.g. weather forecasting),
- ▶ assimilation of synthetic data from a high resolution model into a model of lower resolution (e.g. parameter estimation),
- ▶ rare event simulation (data comes from possible rare event scenarios),
- ▶ solution of inverse problems, minimization (model dynamics is trivial, e.g., $dx_t = 0$).

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

$$z_t^i = (x_t^i, \lambda_t^i), \quad i = 1, \dots, M,$$

with weights

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

E.g. sequential Monte Carlo.

Eulerian:

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

E.g., grid-based methods.

Lagrangian:

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

subject of this talk.

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

$$z_t^i = (x_t^i, \lambda_t^i), \quad i = 1, \dots, M,$$

with weights

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

E.g. sequential Monte Carlo.

Eulerian:

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

E.g., grid-based methods.

Lagrangian:

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

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:

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t$$

Ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

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:

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t$$

Ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

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:

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t$$

Ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

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:

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t$$

Ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

- (i) Draw samples $x_0^i, i = 1, \dots, M$ from the initial distribution π_0 .
- (ii) Solve the **interacting particle system**

$$dx_t^i = f(x_t^i, \lambda)dt + \sigma(x_t^i) \circ dW_t^i + K_t \circ dl_t^i,$$

with **innovation**

$$dl_t^i = dy_t - dy_t^i, \quad dy_t^i := h(x_t^i)dt + R^{-1/2}dU_t^i,$$

and **gain factor**

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

$$P_{xh} := \frac{1}{M-1} \sum_{i=1}^M (x_t^i - \bar{x}_t)(h(x_t^i) - \bar{h}_t)^\top,$$

$i = 1, \dots, M$.

Consider zero-drift & scalar SDE

$$dx_t = \sigma(x_t) \circ dW_t$$

with time-evolved expectation values:

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

Interacting particle representation:

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

It holds that (Liouville plus integration by parts)

$$\begin{aligned} \hat{\pi}_t[f] &:= \pi_0[f] + \frac{1}{2} \int_0^t \pi_s[(\partial_x f)(\sigma l_s)] ds \\ &= \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f] ds = \pi_t[f]. \end{aligned}$$

Consider zero-drift & scalar SDE

$$dx_t = \sigma(x_t) \circ dW_t$$

with time-evolved expectation values:

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

Interacting particle representation:

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

It holds that (Liouville plus integration by parts)

$$\begin{aligned} \hat{\pi}_t[f] &:= \pi_0[f] + \frac{1}{2} \int_0^t \pi_s[(\partial_x f)(\sigma l_s)] ds \\ &= \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f] ds = \pi_t[f]. \end{aligned}$$

Consider zero-drift & scalar SDE

$$dx_t = \sigma(x_t) \circ dW_t$$

with time-evolved expectation values:

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

Interacting particle representation:

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

It holds that (Liouville plus integration by parts)

$$\begin{aligned} \hat{\pi}_t[f] &:= \pi_0[f] + \frac{1}{2} \int_0^t \pi_s[(\partial_x f)(\sigma l_s)] ds \\ &= \pi_0[f] + \int_0^t \pi_s[\mathcal{L}f] ds = \pi_t[f]. \end{aligned}$$

Generalization of the ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

Innovation dl_t as before, i.e.,

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t,$$

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 $\tilde{\pi}_t[f] = \hat{\pi}_t[f]$.

Generalization of the ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

Innovation dl_t as before, i.e.,

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t,$$

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 $\tilde{\pi}_t[f] = \hat{\pi}_t[f]$.

Generalization of the ensemble Kalman-Bucy filter:

$$dx_t = f(x_t, \lambda) dt + \sigma(x_t) \circ dW_t + K_t \circ dl_t$$

Innovation dl_t as before, i.e.,

$$dl_t = dy_t - \frac{1}{2}(h(x_t) - \bar{h}_t) dt$$

or

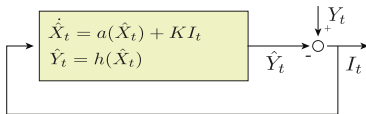
$$dl_t = dy_t - h(x_t) dt - R^{1/2} dU_t,$$

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

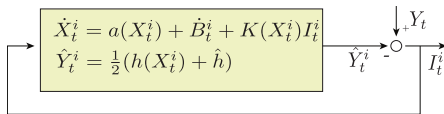
$$-\nabla_x \cdot (\hat{\pi}_t \nabla_x \psi_t) = R^{-1} \hat{\pi}_t (h - \tilde{\pi}_t[h]).$$

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

Structural form of filter formulations:



(a): Kalman Filter



(b): Feedback Particle Filter

A) **Diffusion maps**: Define elliptic operator

$$\mathcal{L}\pi\psi := -\pi^{-1}\nabla_x \cdot (\pi \nabla_x \psi)$$

and introduce the approximation

$$\frac{e^{\epsilon\mathcal{L}\pi} - \text{Id}}{\epsilon} \psi \approx h - \pi[h]$$

The exponential $e^{\epsilon\mathcal{L}\pi}$ can be approximated using diffusion maps.

B) **Optimal transport**: Find optimal coupling T_ϵ between π and

$$\pi_\epsilon = \pi(1 + \epsilon(h - \pi[h])).$$

Then

$$\nabla_x \psi(x) \approx \frac{T_\epsilon(x) - x}{\epsilon}.$$

A) **Diffusion maps**: Define elliptic operator

$$\mathcal{L}\pi\psi := -\pi^{-1}\nabla_x \cdot (\pi \nabla_x \psi)$$

and introduce the approximation

$$\frac{e^{\epsilon\mathcal{L}\pi} - \text{Id}}{\epsilon} \psi \approx h - \pi[h]$$

The exponential $e^{\epsilon\mathcal{L}\pi}$ can be approximated using diffusion maps.

B) **Optimal transport**: Find optimal coupling T_ϵ between π and

$$\pi_\epsilon = \pi(1 + \epsilon(h - \pi[h])).$$

Then

$$\nabla_x \psi(x) \approx \frac{T_\epsilon(x) - x}{\epsilon}.$$

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

$$dx_t^i = f(x_t^i, \lambda) dt + \sigma(x_t^i) \circ dW_t^i - \tilde{K}_t^i \circ dl_t^i$$

Gain:

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

Innovation:

$$dl_t^i = dy_t - \frac{1}{2}(h(x_t^i) - \bar{h}_t) dt$$

or

$$dl_t^i = dy_t - h(x_t^i) dt - R^{1/2} dU_t^i,$$

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), David, de Wiljes & SR (ArXiv:1709.09199)

Alternative control formulation:

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

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), David, de Wiljes & SR (ArXiv:1709.09199)

Alternative control formulation:

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

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), David, de Wiljes & SR (ArXiv:1709.09199)

Alternative control formulation:

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

Bayes:

$$\pi(x|y) \propto \pi(y|x) \pi(x)$$

Task: Given a set of samples $x_0^i \sim \pi(x)$, produce an estimator for expectation values with respect to the posterior distribution.

Application: Learn models

$$y_{\text{out}} = \Psi_x(y_{\text{in}}),$$

parametrized by x , which gives rise to the likelihood $\pi(y|x)$.

Standard methodologies:

- ▶ importances sampling
- ▶ MCMC
- ▶ homotopy methods: **learning models by making them interact**

Bayes:

$$\pi(x|y) \propto \pi(y|x) \pi(x)$$

Task: Given a set of samples $x_0^i \sim \pi(x)$, produce an estimator for expectation values with respect to the posterior distribution.

Application: Learn models

$$y_{\text{out}} = \Psi_x(y_{\text{in}}),$$

parametrized by x , which gives rise to the likelihood $\pi(y|x)$.

Standard methodologies:

- ▶ importances sampling
- ▶ MCMC
- ▶ homotopy methods: **learning models by making them interact**

Homotopy: Define family of distributions

$$\pi_\alpha(x) \propto e^{-\alpha L} \pi(x),$$

where $L(x) := -\ln \pi(y|x)$ and $\alpha \geq 0$. The posterior is obtained for $\alpha = 1$.

The function L can also stand for a tilting potential in rare event simulations/ importance sampling.

Dynamic formulation:

► **Bayes**

$$\frac{\partial}{\partial \alpha} \pi_\alpha = -\pi_\alpha (L - \pi_\alpha[L]), \quad \pi[e^{-L}] = 1 + \int_0^1 \pi_\alpha[L] dt,$$

► **Liouville**

$$\frac{\partial}{\partial \alpha} \pi_\alpha = -\nabla_x \cdot (\pi_\alpha \nabla_x \psi_\alpha)$$

and parameter dynamics given by

$$\dot{x} = \nabla_x \psi_\alpha(x).$$

Homotopy: Define family of distributions

$$\pi_\alpha(x) \propto e^{-\alpha L} \pi(x),$$

where $L(x) := -\ln \pi(y|x)$ and $\alpha \geq 0$. The posterior is obtained for $\alpha = 1$.

The function L can also stand for a tilting potential in rare event simulations/ importance sampling.

Dynamic formulation:

► **Bayes**

$$\frac{\partial}{\partial \alpha} \pi_\alpha = -\pi_\alpha (L - \pi_\alpha[L]), \quad \pi[e^{-L}] = 1 + \int_0^1 \pi_\alpha[L] dt,$$

► **Liouville**

$$\frac{\partial}{\partial \alpha} \pi_\alpha = -\nabla_x \cdot (\pi_\alpha \nabla_x \psi_\alpha)$$

and parameter dynamics given by

$$\dot{x} = \nabla_x \psi_\alpha(x).$$

Particle flow filter:

$$\dot{x}_\alpha^i = \nabla_x \psi_\alpha(x_\alpha^i), \quad x_0^i \sim \pi(x),$$

$\alpha \in [0, 1]$, with potential ψ_α solving

$$\nabla_x \cdot (\pi_\alpha \nabla_x \psi_\alpha) = \pi_\alpha (L - \pi_\alpha(L)).$$

Numerical implementation:

- ▶ diffusion maps
- ▶ optimal transportation
- ▶ (iterative) ensemble Kalman filter

Particle flow filter:

$$\dot{x}_\alpha^i = \nabla_x \psi_\alpha(x_\alpha^i), \quad x_0^i \sim \pi(x),$$

$\alpha \in [0, 1]$, with potential ψ_α solving

$$\nabla_x \cdot (\pi_\alpha \nabla_x \psi_\alpha) = \pi_\alpha (L - \pi_\alpha(L)).$$

Numerical implementation:

- ▶ diffusion maps
- ▶ optimal transportation
- ▶ (iterative) ensemble Kalman filter

Extensions:

- ▶ **Randomized particle flow:**

$$\begin{aligned} dx_\alpha &= \epsilon \nabla_x \ln \pi_\alpha(x) dt + \sqrt{2\epsilon} dW_\alpha + \nabla_x \psi_\alpha(x) dt \\ &= \nabla_x \{ \psi_\alpha(x) + \epsilon \ln \pi_0(x) - \epsilon \alpha L(x) \} dt + \sqrt{2\epsilon} dW_\alpha, \end{aligned}$$

for $\alpha \geq 0$. Here $\epsilon \geq 0$ is a parameter.

It still holds that

$$\partial_\alpha \pi_\alpha = -\pi_\alpha (L - \pi_\alpha[L]).$$

- ▶ **Minimization:** Method for finding

$$x^* = \arg \min V(x)$$

if $\alpha \rightarrow \infty$, $\epsilon \propto \alpha^{-1}$, and

$$\pi_\alpha \propto e^{-\alpha V} \pi_0.$$

Extensions:

- ▶ **Randomized particle flow:**

$$\begin{aligned} dx_\alpha &= \epsilon \nabla_x \ln \pi_\alpha(x) dt + \sqrt{2\epsilon} dW_\alpha + \nabla_x \psi_\alpha(x) dt \\ &= \nabla_x \{ \psi_\alpha(x) + \epsilon \ln \pi_0(x) - \epsilon \alpha L(x) \} dt + \sqrt{2\epsilon} dW_\alpha, \end{aligned}$$

for $\alpha \geq 0$. Here $\epsilon \geq 0$ is a parameter.

It still holds that

$$\partial_\alpha \pi_\alpha = -\pi_\alpha (L - \pi_\alpha[L]).$$

- ▶ **Minimization:** Method for finding

$$x^* = \arg \min V(x)$$

if $\alpha \rightarrow \infty$, $\epsilon \propto \alpha^{-1}$, and

$$\pi_\alpha \propto e^{-\alpha V} \pi_0.$$

Particle flow:

- ▶ Moser (1965),
- ▶ Daum et al. (SPIE, 2010), Daum et al. (SPIE, 2017)
- ▶ SR (BIT, 2011)
- ▶ Heng, Doucet & Pokern (2015, ArXiv:1509.08787)

Extensions:

- ▶ iterative Kalman filters (Dean Oliver, Mark Bocquet, Marco Iglesias, SR ...)
- ▶ sequential Monte Carlo & rejuvenation (... , Beskos, Crisan & Jasra (Annals of Applied Prob, 2014), ...)
- ▶ Kalman filter as minimizer (Schilling & Stuart (2016, ArXiv:1602.02020))
- ▶ stochastic minimization techniques/ simulated annealing (...)

Particle flow:

- ▶ Moser (1965),
- ▶ Daum et al. (SPIE, 2010), Daum et al. (SPIE, 2017)
- ▶ SR (BIT, 2011)
- ▶ Heng, Doucet & Pokern (2015, ArXiv:1509.08787)

Extensions:

- ▶ iterative Kalman filters (Dean Oliver, Mark Bocquet, Marco Iglesias, SR ...)
- ▶ sequential Monte Carlo & rejuvenation (... , Beskos, Crisan & Jasra (Annals of Applied Prob, 2014), ...)
- ▶ Kalman filter as minimizer (Schilling & Stuart (2016, ArXiv:1602.02020))
- ▶ stochastic minimization techniques/ simulated annealing (...)

- ▶ ensemble Kalman filter has triggered renewed interest in Lagrangian filtering and data assimilation
- ▶ stability and accuracy of the resulting interacting particle systems is largely unknown
- ▶ many applications outside classic filtering context such as rare event simulations and whenever a change of measure arises
- ▶ exploitation of geometric properties? E.g. data assimilation for highly oscillatory Hamiltonian systems (Reinhardt, Hastermann, Klein, SR, ArXiv:1708.03570)
- ▶ ...

- ▶ ensemble Kalman filter has triggered renewed interest in Lagrangian filtering and data assimilation
- ▶ stability and accuracy of the resulting interacting particle systems is largely unknown
- ▶ many applications outside classic filtering context such as rare event simulations and whenever a change of measure arises
- ▶ exploitation of geometric properties? E.g. data assimilation for highly oscillatory Hamiltonian systems (Reinhardt, Hastermann, Klein, SR, ArXiv:1708.03570)
- ▶ ...