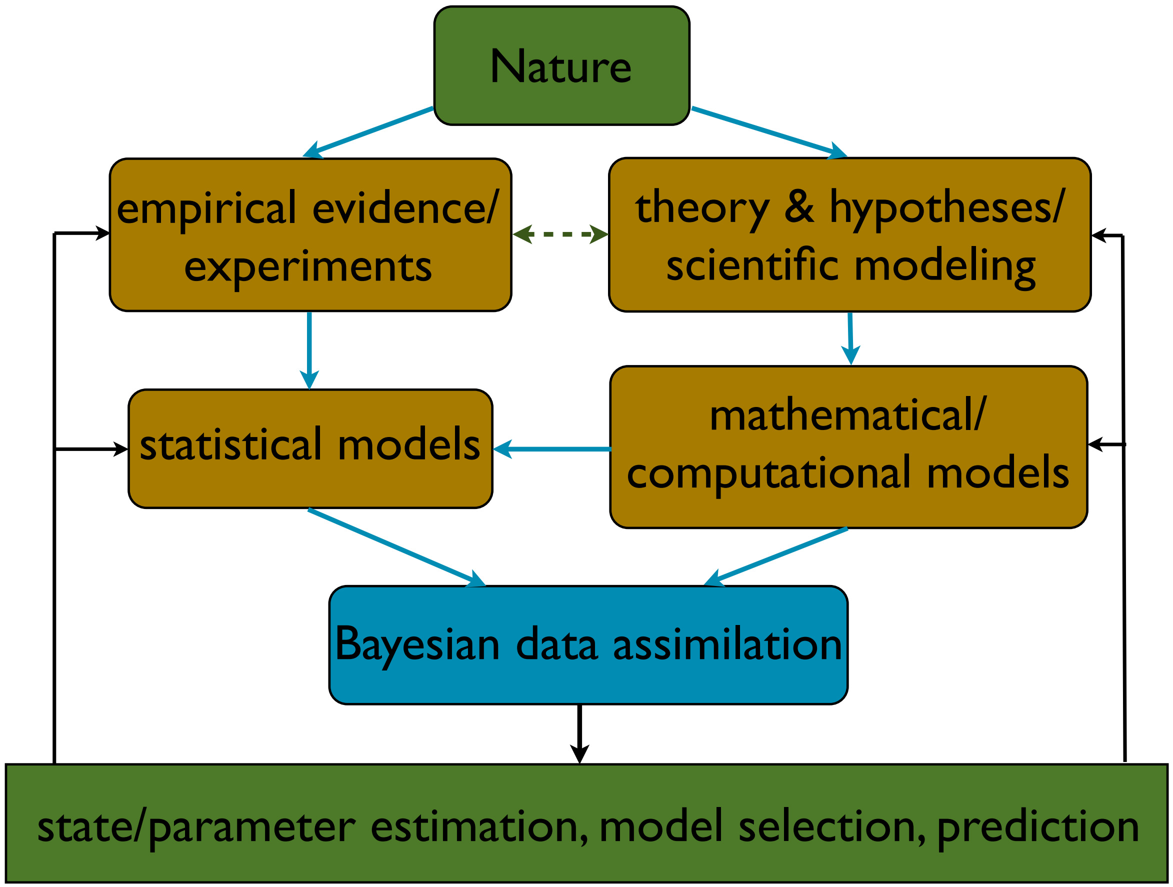
Ensemble transform filters for geophysical data assimilation

Sebastian Reich in collaboration with Georg Gottwald (University of Sydney), Kay Bergemann, Seoleun Shin (Uni Potsdam), Eugenia Kalnay, Kayo Ide, Javier Amezcua (U of Maryland)

1. Data assimilation (DA)



2. A simplified mathematical problem statement

1) Mathematical model

$$\dot{\mathbf{x}} = f(\mathbf{x})$$

2) Random initial conditions $\mathbf{x}(0)$, i.e. $\mathbf{x}(0) \sim \rho_0$ for some given probability density function (PDF) $\rho_0(\mathbf{x})$.

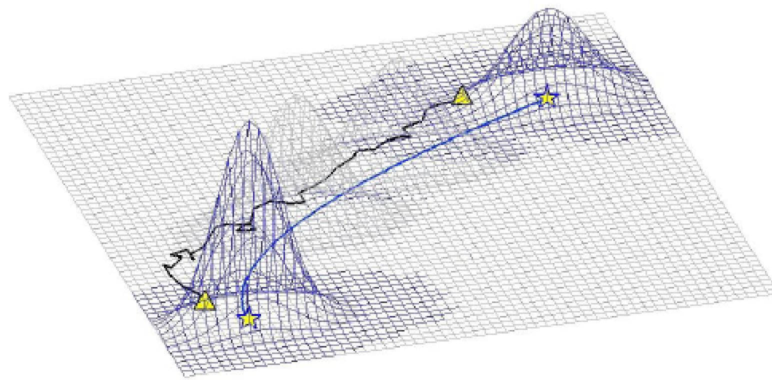
3) Observations/measurements

$$\mathbf{y}(t_j) = \mathbf{H}\mathbf{x}(t_j) + \eta_j, \qquad \mathbf{y} \in \mathbb{R}^k, \quad \mathbf{H} \in \mathbb{R}^{k \times n},$$

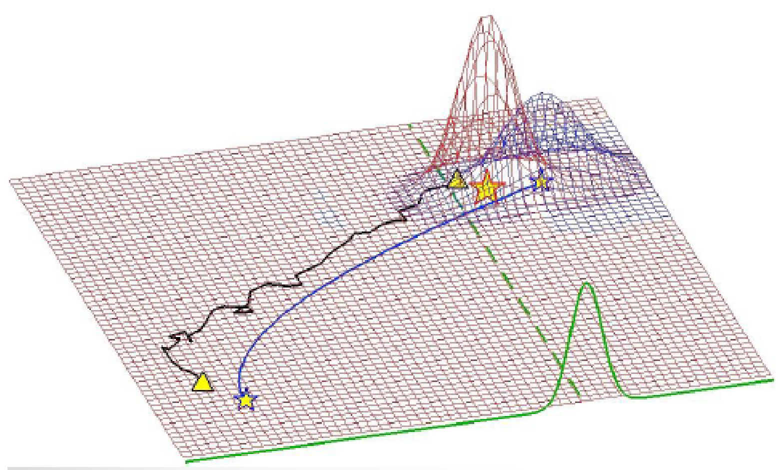
$k < n$, at discrete times $t_j > 0$, $j = 1, \ldots, J$, subject to random measurement errors:

$$\eta_j \sim \mathsf{N}(\mathbf{0}, \mathbf{R}).$$

Liouville (forecast): $\mathbf{x} = (x_1, x_2)$, $\pi_{\text{post}}(\mathbf{x}, t_j) \to \pi_{\text{prior}}(\mathbf{x}, t_{j+1})$



Bayes (assimilation): $y = x_1 + \eta$, $\pi_{\text{prior}}(\mathbf{x}, t_{j+1}) \to \pi_{\text{post}}(\mathbf{x}, t_{j+1})$



(figures courtesy Chris Jones)

3. Ensemble propagation and particle filters

Particle/Monte-Carlo methods. Given a collection

$$\mathbf{X}(t) = [\mathbf{x}_1(t) \,|\, \mathbf{x}_2(t) \,|\, \cdots \,|\, \mathbf{x}_m(t)] \in \mathbb{R}^{n \times m}$$

of m independent solutions of

$$\dot{\mathbf{x}} = f(\mathbf{x})$$

with initial conditions $\mathbf{x}_i(0) \sim \rho_0(\mathbf{x})$.

The empirical measure

$$\rho_{\text{em}}(\mathbf{x}, t) = \sum_{i=1}^{m} \alpha_i \delta(\mathbf{x} - \mathbf{x}_i(t))$$

is an exact (weak) solution to Liouville's equation ($\delta(\cdot)$ Dirac's delta function, $\alpha_i > 0$ (constant) weights).

Data assimilation for particle approximation at t_j.

Bayes' formula:

$$\rho(\mathbf{x}, t_j^+) = \sum_{i=1}^{m} \alpha_i \, \delta(\mathbf{x} - \mathbf{x}_i(t_j)).$$

with new weights

$$\alpha_i \Rightarrow \pi(\mathbf{y}(t_j)|\mathbf{x}_i(t_j)) \times \alpha_i.$$

Problem: Particles are still sampled from the prior and not from the posterior PDF. The mismatch is compensated for by non-uniform weight factors (importance sampling). Near mutual singularity of measures leads to degeneracy of weights.

4. Assimilation as a continuous deformation of probability I

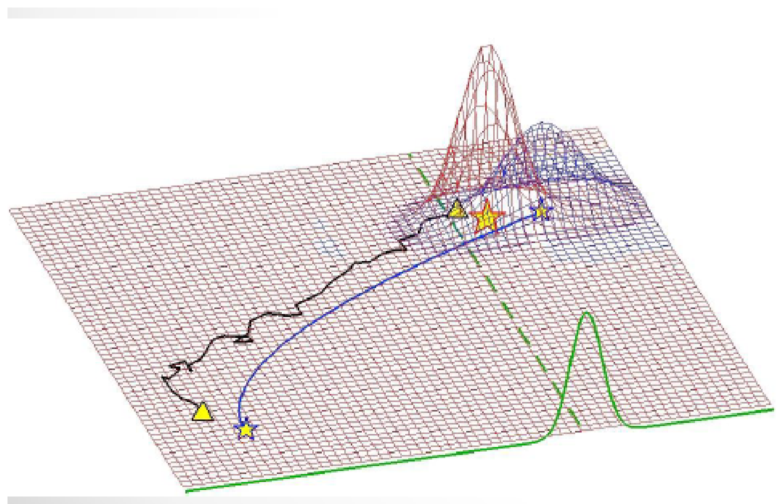
Dynamics and assimilation are not easily compatible in a numerical sense.

We have to shift focus somehow. Here is our basic line of attack:

We view the change of measure induced by Bayes' theorem as an (optimal) transportation problem.

See Crisan and Xiong (2010) for a related approach in the context of continuous time filtering.

Bayes (assimilation): $\pi_{\text{prior}}(\mathbf{x}, t_{j+1}) \to \pi_{\text{post}}(\mathbf{x}, t_{j+1})$



Task: Shuffle blue prior into red posterior: **x**' = T(**x**) s.t.

$$\pi_{\text{post}}(T(\mathbf{x}))|DT(\mathbf{x})| = \pi_{\text{pr}}(\mathbf{x}).$$

Negative log-likelihood:

$$L = (\mathbf{H}\mathbf{x} - \mathbf{y}(t_j))^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}(t_j))/2.$$

Bayes' theorem:

$$\pi_{\mathsf{post}}(\mathbf{x}) = \pi(\mathbf{x}|\mathbf{y}(t_j)) \;\propto\; \pi_{\mathsf{pr}}(\mathbf{x}) e^{-L} = \pi_{\mathsf{pr}}(\mathbf{x}) \prod_{n=1}^{N} e^{-L/N}$$

Limit $N \to \infty$, Taylor expansion in $\Delta s = 1/N$:

$$\frac{\partial \rho}{\partial s} = -\rho \left(L - \mathbb{E}_\rho[L]\right),$$

$s \in [0, 1]$, $\rho(\mathbf{x}, 0) = \pi_{\mathsf{pr}}(\mathbf{x})$.

"Kushner-Stratonovich" equation for intermittent data assimilation (Reich, 2011):

$$\frac{\partial \rho}{\partial t} = -\nabla_{\mathbf{x}}(\rho f) - \sum_{j \geq 1} \delta_\epsilon(t - t_j)\rho(L - \bar{L})$$

with $\delta_\epsilon(\cdot)$ is a (compact) mollifier to the Dirac delta function with support $[t_j - \epsilon, t_j + \epsilon]$ and

$$\bar{L} = \mathbb{E}_\rho[L].$$

If $\Psi(x)$ is a smooth observable, then

$$\frac{d}{dt}\bar{\Psi} = \overline{\nabla_x \Psi \cdot f} - \sum_{j \geq 1} \delta_\epsilon(t - t_j)\overline{(\Psi - \bar{\Psi})(L - \bar{L})}$$

One can in particular set $\rho = \rho_{\text{em}}$, $\Psi = x$ or $\Psi = (x - \bar{x})(x - \bar{x})^T$.

4. Assimilation as a continuous deformation of probability II

For demonstration purposes, consider a single degree of freedom:

We got to transport particles distributed according to the heap on the left (prior) such that they form the heap on the right (posterior). Direct approach: Monge-Ampere, nonlinear elliptic PDF.

Goal: Find vector field $g(\mathbf{x};\rho) \in L^2(\mathrm{d}\rho, \mathbb{R}^n)$ such that

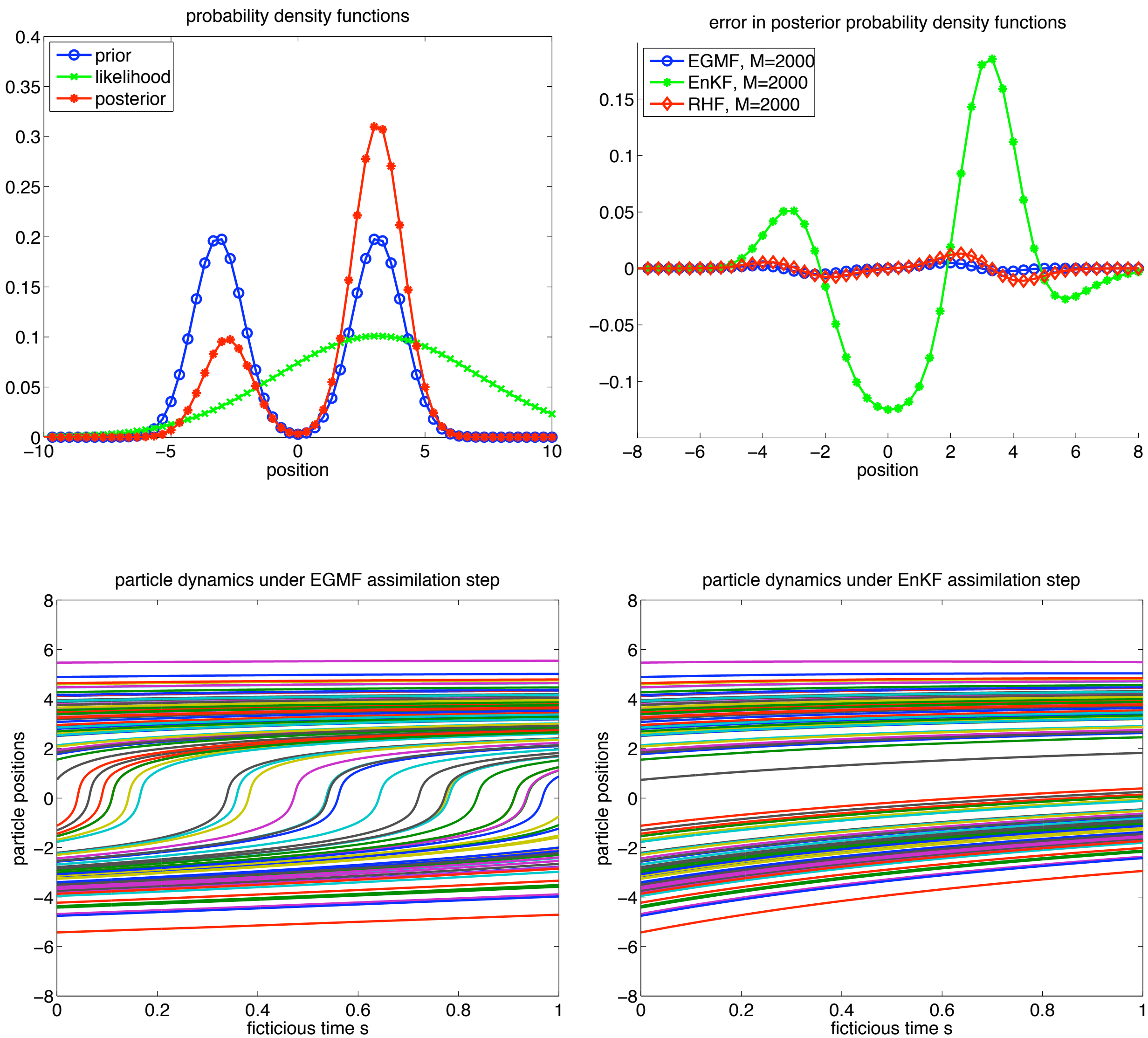
$$\nabla_{\mathbf{x}} \cdot (\rho g) = \rho \{L - \bar{L}\}.$$

The vector field g is not unique. Find minimizer with respect to the norm (see Otto, 2001, for an application in gradient flow dynamics):

$$\left\| \frac{\partial \rho}{\partial s} \right\|_{\rho}^{2} = \inf_{v \in L^2(\mathrm{d}\rho, \mathbb{R}^n)} \left\{ \int \mathrm{d}\rho \, v^T \mathbf{M} v : \frac{\partial \rho}{\partial s} + \nabla_{\mathbf{x}} \cdot (\rho v) = 0 \right\}.$$

Gradient flow:

$$g(\mathbf{x};\rho) = \mathbf{M}^{-1} \nabla_{\mathbf{x}} \psi(\mathbf{x};\rho).$$


probability density functions
prior
likelihood
posterior
position
error in posterior probability density functions
EGMF, M=2000
EnKF, M=2000
RHF, M=2000
position
particle dynamics under EGMF assimilation step
particle positions
ficticious time s
particle dynamics under EnKF assimilation step
particle positions
ficticious time s

Formulation of a complete dynamics-assimilation step (Reich, 2010):

$$\dot{\mathbf{x}} = f(\mathbf{x}) + \delta_\epsilon(t - t_j)\mathbf{M}^{-1}\nabla_{\mathbf{x}}\psi(\mathbf{x};\rho), \tag{1}$$

$$\rho_t = -\nabla_{\mathbf{x}} \cdot (\rho\dot{\mathbf{x}}) \tag{2}$$

closed by the elliptic PDE

$$\nabla_{\mathbf{x}} \cdot (\rho\mathbf{M}^{-1}\nabla_{\mathbf{x}}\psi) = \rho\,\{L - \bar{L}\}. \tag{3}$$

for $t \in [t_j - \epsilon, t_j + \epsilon]$.

This system constitutes an example of a Vlasov-McKean system (infinite-dimensional interacting particle system).

Algorithmic summary of data assimilation step:

1) Use the ensemble of solutions $\mathbf{x}_i(t)$, $i = 1, \ldots, m$, to define a statistical model $\tilde{\rho}(\mathbf{x}; \{\mathbf{x}_i(t)\})$.

2) Solve

$$\nabla_{\mathbf{x}} \cdot (\tilde{\rho} \mathbf{M}^{-1} \nabla_{\mathbf{x}} \psi) = \tilde{\rho} \left\{ L - \mathbb{E}_{\tilde{\rho}}[L] \right\} \tag{4}$$

numerically or by quadrature for the potential ψ.

3) Propagate particles under vector field $g = \mathbf{M}^{-1} \nabla \psi$ (gradient flow) .

Choices for $\tilde{\rho}$ include

(a) Gaussian PDF parametrized by the ensemble mean and covariance matrix (ensemble Kalman filter (EnKF))

(b) Gaussian mixture models/ Gaussian kernel density estimators. Resulting elliptic PDE can be solved analytically!

(c) Empirical PDF and weak form of McKean-Vlasov; e.g. find vector field such that ensemble mean and covariance matrix are consistently propagated (different from EnKF!)

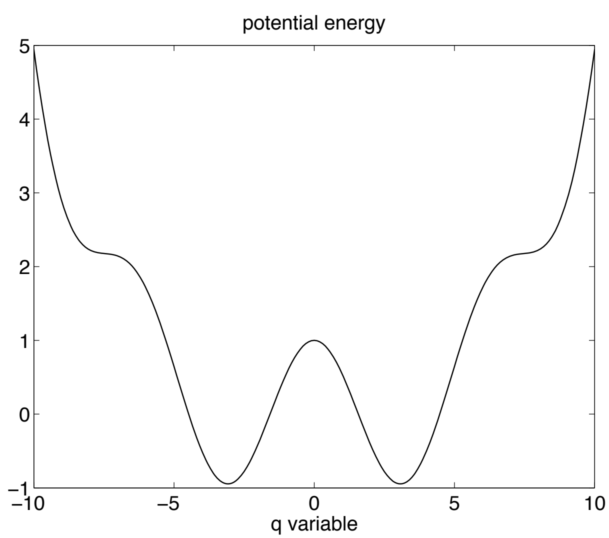
The second and third approach give rise to new filters (Reich 2011/2012).

6. Numerics I: Langevin dynamics under bistable potential

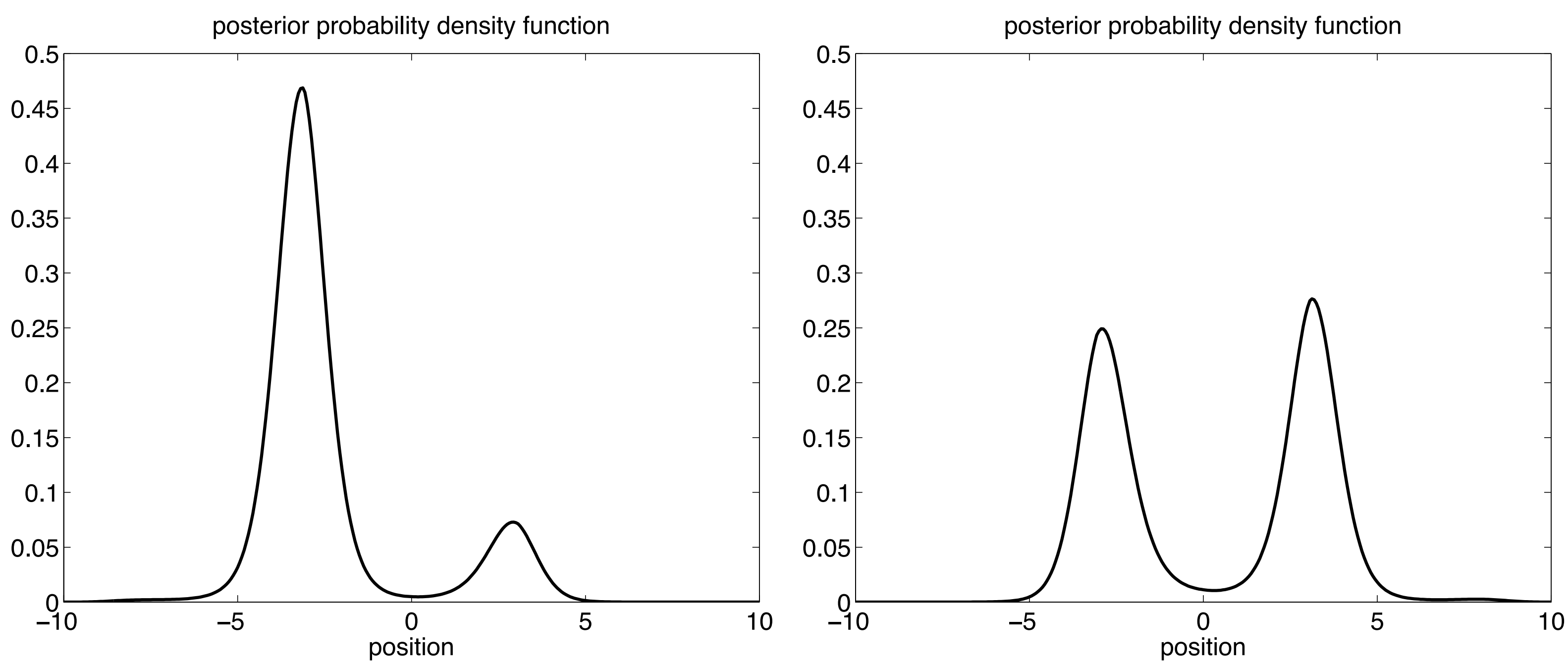
Second-order Langevin dynamics under a one-dimensional double-well potential $V(q)$:

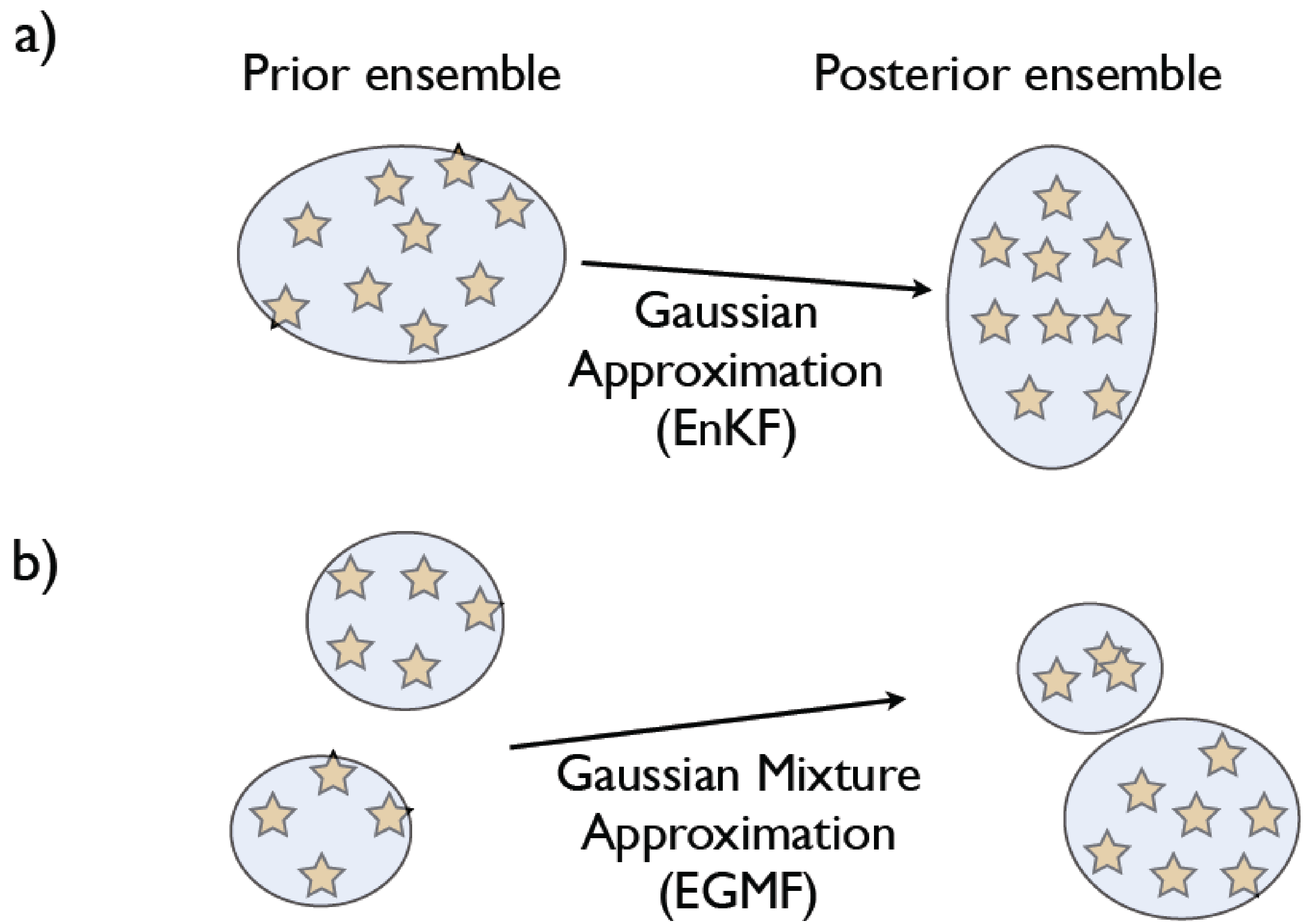
$$\mathrm{d}v = -V'(q)\mathrm{d}t - \gamma v \mathrm{d}t + \sqrt{2\sigma}\mathrm{d}w(t), \qquad \mathrm{d}q = v\mathrm{d}t,$$

$w(t)$ denotes standard Brownian motion, potential $V(q)$ given by

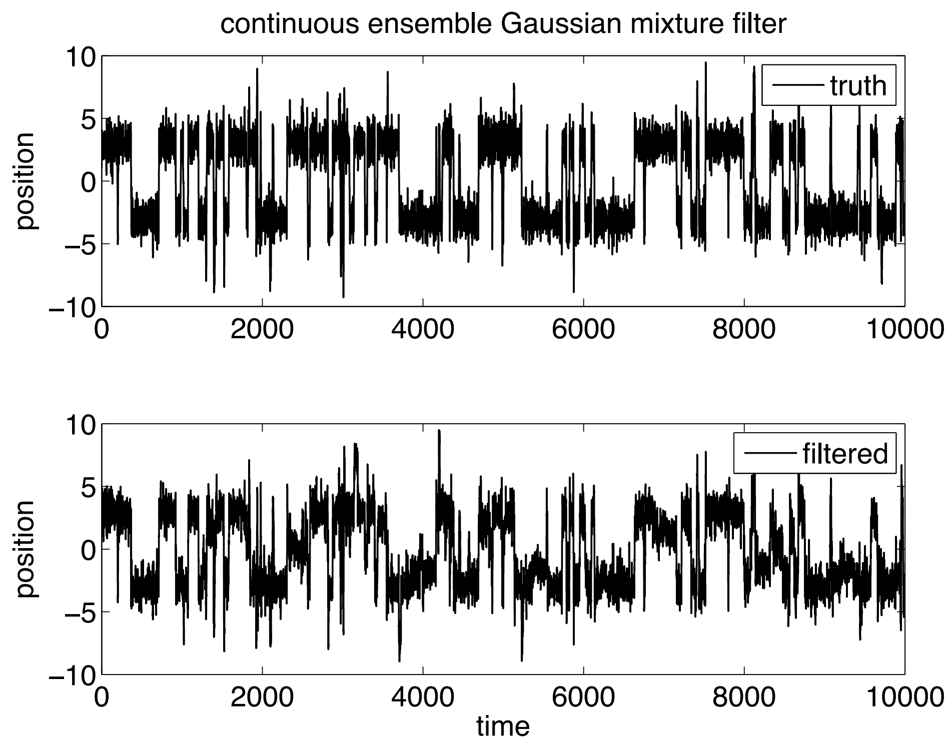


Assimilate velocities v only (!); posterior PDFs in positions q:




a)
Prior ensemble
Posterior ensemble
Gaussian
Approximation
(EnKF)
b)
Gaussian Mixture
Approximation
(EGMF)

Results from EGMF ($M = 50$ ensemble members):



RMSE $= 1.9148$ for EGMF compared to 2.331 for the EnKF. A bi-Gaussian is used in 25% of the assimilation steps.

7. First and second-order filter consistency

Bayes' theorem implies

$$\bar{\Psi} = \frac{\mathbb{E}_{\pi_{\mathrm{pr}}}[\Psi e^{-L}]}{\mathbb{E}_{\pi_{\mathrm{pr}}}[e^{-L}]}$$

for the posterior expectation value of an observable Ψ.

Set $\pi_{\mathrm{pr}} = \rho_{\mathrm{em}}$ (ensemble), $\Psi = x$ (ensemble mean) or $\Psi = (x - \bar{x})(x - \bar{x})^T$ (ensemble covariance matrix).

The ensemble Kalman filter (EnKF) is not consistent with respect to the Bayesian posterior mean and covariance matrix.

Example for classic moment closure problem:

$$\frac{d}{ds}\bar{\Psi} = -\overline{(\Psi - \bar{\Psi})(L - \bar{L})}.$$

Simple example:

$x \in \mathbb{R}$, obs $y = x + \eta$, prior with mean $\bar{x} = 0$ and variance P:

$$\Psi = x^2 \to \bar{\Psi} = P$$

Then

$$\begin{aligned} \frac{d}{ds} P &= -\frac{1}{2} \overline{(x^2 - P)(x - y)^2} \\ &= \frac{1}{2} \left\{ \overline{x^4} - P^2 - 2\overline{x^3} y \right\} \end{aligned}$$

For Gaussians: $\overline{x^4} = 3P^2$, $\overline{x^3} = 0$ (EnKF closure).

Appropriate corrections proposed by Lei & Bickel (2011) (NLEAF) and Reich (2011) (MEnKF). Requires computing the square root of covariance matrix.

Basic idea:

x_i^p are the proposal ensemble members with mean $\bar{x}^p$ and covariance matrix $\bar{P}^p$. These can be obtained from a standard EnKF.

The Bayesian posterior mean and covariance are denoted by $\bar{x}$ and $\bar{P}$, respectively.

The analysed ensemble members are now given by

$$x_i^a = \bar{x} + \bar{P}^{1/2}(\bar{P}^p)^{-1/2}(x_i^p - \bar{x}^p).$$

Lorenz-63 model:

Mean RMSE over 2000 assimilation cycles with $t_j = 0.05\,j$ for the Lorenz-63 model using five different filter algorithms and three different ensemble sizes. FD (filter divergence) is being used whenever the mean RMSE exceeds the observation error $R = 4$.

	EnKF	MEnKF1	MEnKF2	NLEAF1	NLEAF2
$M = 10$	0.4405	1.2045	FD	FD	FD
$M = 40$	0.3004	0.3140	0.2510	0.3772	FD
$M = 400$	0.3272	0.3262	0.2375	0.3037	0.2336

Continuous & consistent update:

$$\frac{d}{ds}\bar{x} = -\overline{L\Delta x} \quad \text{(ensemble mean)},$$

and

$$\frac{d}{ds}\Delta x = -\frac{1}{2}(L - \bar{L})\Delta x - \frac{1}{2}\overline{L\Delta x} \quad \text{(ensemble deviations)}$$

with $\Delta x = x - \bar{x}$.

Ensemble filter equations for assimilation at t_j:

$$\frac{d}{dt}x_i = f(x_i) - \frac{1}{2}\delta_\epsilon(t - t_j)\left\{(L_i - \bar{L})\Delta x_i + \overline{L\Delta x}\right\}.$$

EnKF closure seems preferable. Work in progress ...

8. Summary

Ensemble deformation techniques lead to a dynamical systems interpretation of data assimilation and provide an alternative to particle filters with (stochastic) resampling (sequential Monte Carlo methods).

Future work:

1) analyse filter algorithms from dynamical systems point of view,
2) minimum variance estimate (EnKF) vs. MAP estimate (4DVar)
3) smoothing and parameter estimation, model selection
4) data assimilation for multi-scale problems,
5) assimilation of information aside from “observations”,
6) links to inverse problems (e.g. L_1 regularization)

Publications:

1) Bergemann, Gottwald, Reich: Ensemble propagation and continuous matrix factorization algorithms, QJ, 2009.
2) Bergemann, Reich: A localization technique for ensemble Kalman filters, QJ, 2010.
3) Bergemann, Reich: A mollified ensemble Kalman filter, QJ, 2010.
4) Reich: A dynamical systems framework for intermittent data assimilation, BIT, 2011.
5) Reich: A Gaussian-mixture ensemble transform filter, QJ, 2011.
6) Gottwald, Mitchell, Reich: Controlling overestimation of error covariance in ensemble Kalman filters with sparse observations: A variance limiting Kalman filter, MWR, 2011.
7) Bergemann, Reich: An ensemble Kalman-Bucy filter for continuous data assimilation, submitted.
8) Amezcua, Kalnay, Ide, Reich: Using the ensemble Kalman-Bucy filter in an ensemble framework, submitted.
9) Reich, Shin: On the consistency of ensemble transform filter formulations, submitted.