Optimal Transportation and its use in data assimilation and sequential Bayesian inference

Walter Acevedo, Jana de Wiljes, and Sebastian Reich

Universität Potsdam/ University of Reading

CMO-BIRS, Oaxaca, May 3rd 2017

Given:

- M samples $\mathbf{z}_i^{\mathrm{f}}$ from a RV Z^{f} with PDF $\pi^{\mathrm{f}}(\mathbf{z})$ (prior)
- normalized importance weights $w_i \propto \pi(\mathbf{y}_{\mathrm{obs}}|\mathbf{z}_i^{\mathrm{f}})$ (likelihood)

Desired:

- M samples $\mathbf{z}_i^{\mathrm{a}}$ from a RV Z^{a} with PDF (posterior)

$$\pi^{\mathrm{a}}(\mathbf{z}) \propto \pi(\mathbf{y}_{\mathrm{obs}}|\mathbf{z})\, \pi^{\mathrm{f}}(\mathbf{z}).$$

- typically achieved by sampling from a discrete RV

$$\widehat{Z}^{\mathrm{a}}(\omega) \in \{\mathbf{z}_i^{\mathrm{f}}\}_{i=1,\ldots,M}$$

with $\mathbb{P}[\widehat{Z}^{\mathrm{a}}(\omega) = \mathbf{z}_i^{\mathrm{f}}] = w_i$ (resampling with replacement).

Q: How to make this work for high-dimensional problems and relatively small sample sizes M.

Problem setting

Given:

- M samples $\mathbf{z}_i^{\mathrm{f}}$ from a RV Z^{f} with PDF $\pi^{\mathrm{f}}(\mathbf{z})$ (prior)
- normalized importance weights $w_i \propto \pi(\mathbf{y}_{\mathrm{obs}}|\mathbf{z}_i^{\mathrm{f}})$ (likelihood)

Desired:

- M samples $\mathbf{z}_i^{\mathrm{a}}$ from a RV Z^{a} with PDF (posterior)

$$\pi^{\mathrm{a}}(\mathbf{z}) \propto \pi(\mathbf{y}_{\mathrm{obs}}|\mathbf{z})\,\pi^{\mathrm{f}}(\mathbf{z}).$$

- typically achieved by sampling from a discrete RV

$$\widehat{Z}^{\mathrm{a}}(\omega) \in \{\mathbf{z}_i^{\mathrm{f}}\}_{i=1,\ldots,M}$$

with $\mathbb{P}[\widehat{Z}^{\mathrm{a}}(\omega) = \mathbf{z}_i^{\mathrm{f}}] = w_i$ (resampling with replacement).

Q: How to make this work for high-dimensional problems and relatively small sample sizes M.

Problem setting

Given:

- M samples $\mathbf{z}_i^{\mathrm{f}}$ from a RV Z^{f} with PDF $\pi^{\mathrm{f}}(\mathbf{z})$ (prior)
- normalized importance weights $w_i \propto \pi(\mathbf{y}_{\mathrm{obs}}|\mathbf{z}_i^{\mathrm{f}})$ (likelihood)

Desired:

- M samples $\mathbf{z}_i^{\mathrm{a}}$ from a RV Z^{a} with PDF (posterior)

$$\pi^{\mathrm{a}}(\mathbf{z}) \propto \pi(\mathbf{y}_{\mathrm{obs}}|\mathbf{z})\,\pi^{\mathrm{f}}(\mathbf{z}).$$

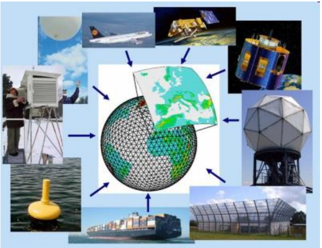
- typically achieved by sampling from a discrete RV

$$\widehat{Z}^{\mathrm{a}}(\omega) \in \{\mathbf{z}_i^{\mathrm{f}}\}_{i=1,\ldots,M}$$

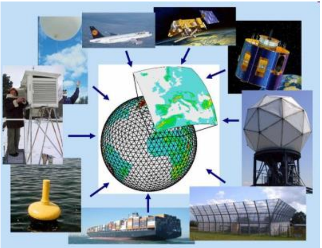
with $\mathbb{P}[\widehat{Z}^{\mathrm{a}}(\omega) = \mathbf{z}_i^{\mathrm{f}}] = w_i$ (resampling with replacement).

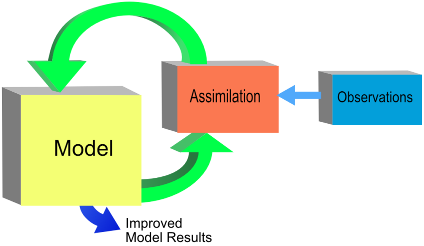
Q: How to make this work for high-dimensional problems and relatively small sample sizes M.

Numerical Weather Prediction



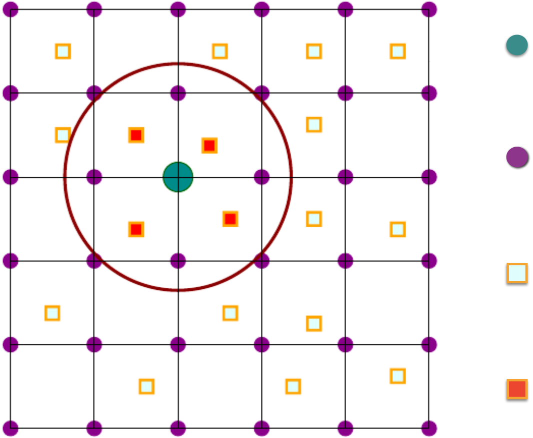
- **Model**: highly nonlinear discretized partial differential equations
- **Data**: heterogeneous mix of ground-, airborne-, satellite-based and radar data
- 24/7 **data assimilation** service for optimal weather prediction
- non-traditional **particle filters** (PF) with $M = \mathcal{O}(10^2)$ particles for models with dimension of state space $N = \mathcal{O}(10^7)$ being used operationally





- **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
- non-traditional **particle filters** (PF) with $M = \mathcal{O}(10^2)$ particles for models with dimension of state space $N = \mathcal{O}(10^7)$ being used operationally

Key idea: Localization



Classic PF: Resampling with replacement

Resampling interpreted as discrete Markov chain

$$\mathbf{P} \in \mathbb{R}^{M\times M}$$

s.t. $p_{ij} \geq 0$ and

$$\sum_i p_{ij} = 1, \qquad \frac{1}{M}\sum_j p_{ij} = w_i.$$

and

$$\mathbf{z}_j^{\mathrm{a}} = \mathbf{z}_i^{\mathrm{f}} \quad \text{with probability} \quad p_{ij}\,.$$

Example. Monomial resampling

$$\mathbf{P}^0 := \mathbf{w} \otimes \mathbf{1} = \begin{pmatrix} w_1 & w_1 & \cdots & w_1 \\ w_2 & w_2 & \cdots & w_2 \\ \vdots & \vdots & \ddots & \vdots \\ w_M & w_M & \cdots & w_M \end{pmatrix}.$$

Classic PF: Resampling with replacement

Resampling interpreted as discrete Markov chain

$$\mathbf{P} \in \mathbb{R}^{M \times M}$$

s.t. $p_{ij} \geq 0$ and

$$\sum_i p_{ij} = 1, \qquad \frac{1}{M} \sum_j p_{ij} = w_i.$$

and

$$\mathbf{z}_j^{\mathrm{a}} = \mathbf{z}_i^{\mathrm{f}} \quad \text{with probability} \quad p_{ij}\,.$$

Example. Monomial resampling

$$\mathbf{P}^0 := \mathbf{w} \otimes \mathbf{1} = \begin{pmatrix} w_1 & w_1 & \cdots & w_1 \\ w_2 & w_2 & \cdots & w_2 \\ \vdots & \vdots & \ddots & \vdots \\ w_M & w_M & \cdots & w_M \end{pmatrix}.$$

Define resampling Markov chain $\mathbf{P}^\lambda$:

$$\mathbf{P}^\lambda = \operatorname{argmin} \sum_{ij} p_{ij} \left\{ \|\mathbf{z}_i^{\mathrm{f}} - \mathbf{z}_j^{\mathrm{f}}\|^2 + \frac{1}{\lambda} \ln \frac{p_{ij}}{p_{ij}^0} \right\}$$

for given $\lambda > 0$ subject to

$$p_{ij} \geq 0, \qquad \sum_i p_{ij} = 1, \qquad \frac{1}{M} \sum_j p_{ij} = w_i.$$

Remark.

- $\lambda \to 0$: $\mathbf{P}^0 = \mathbf{w} \otimes \mathbf{1}$ (monomial resampling).
- $\lambda \to \infty$: $\mathbf{P}^\infty$ solves the optimal coupling/transport problem.
- Effective iterative solvers are available [Cuturi, 2013].

Sinkhorn approximation I

Define resampling Markov chain $\mathbf{P}^\lambda$:

$$\mathbf{P}^\lambda = \text{argmin} \sum_{ij} p_{ij} \left\{ \|\mathbf{z}_i^{\mathrm{f}} - \mathbf{z}_j^{\mathrm{f}}\|^2 + \frac{1}{\lambda} \ln \frac{p_{ij}}{p_{ij}^0} \right\}$$

for given $\lambda > 0$ subject to

$$p_{ij} \geq 0, \qquad \sum_i p_{ij} = 1, \qquad \frac{1}{M} \sum_j p_{ij} = w_i.$$

Remark.

- $\lambda \to 0$: $\mathbf{P}^0 = \mathbf{w} \otimes \mathbf{1}$ (monomial resampling).
- $\lambda \to \infty$: $\mathbf{P}^\infty$ solves the optimal coupling/transport problem.
- Effective iterative solvers are available [Cuturi, 2013].

Prior and posterior means:

$$\bar{\mathbf{z}}^{\mathrm{f}} = \frac{1}{M}\sum_i \mathbf{z}_i^{\mathrm{f}}, \qquad \bar{\mathbf{z}}^{\mathrm{a}} = \sum_i w_i \mathbf{z}_i^{\mathrm{f}}$$

Mean value for each column of the resampling Markov chain:

$$\bar{\mathbf{z}}_j^{\mathrm{a}} = \sum_i \mathbf{z}_i^{\mathrm{f}} p_{ij}$$

Reformulated Sinkhorn cost:

$$J(\mathbf{P}) = -2\sum_j (\bar{\mathbf{z}}_j^{\mathrm{a}} - \bar{\mathbf{z}}^{\mathrm{a}})\cdot(\mathbf{z}_j^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}) + \frac{1}{\lambda}\sum_{ij} p_{ij} \ln \frac{p_{ij}}{p_{ij}^0} + \text{constant}$$

Remark. Monomial resampling:

$$-2\sum_j (\bar{\mathbf{z}}_j^{\mathrm{a}} - \bar{\mathbf{z}}^{\mathrm{a}})\cdot(\mathbf{z}_j^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}) = 0.$$

Sinkhorn approximation II

Prior and posterior means:

$$\overline{\mathbf{z}}^{\mathrm{f}} = \frac{1}{M}\sum_i \mathbf{z}_i^{\mathrm{f}}, \qquad \overline{\mathbf{z}}^{\mathrm{a}} = \sum_i w_i \mathbf{z}_i^{\mathrm{f}}$$

Mean value for each column of the resampling Markov chain:

$$\overline{\mathbf{z}}_j^{\mathrm{a}} = \sum_i \mathbf{z}_i^{\mathrm{f}} p_{ij}$$

Reformulated Sinkhorn cost:

$$J(\mathbf{P}) = -2\sum_j (\overline{\mathbf{z}}_j^{\mathrm{a}} - \overline{\mathbf{z}}^{\mathrm{a}}) \cdot (\mathbf{z}_j^{\mathrm{f}} - \overline{\mathbf{z}}^{\mathrm{f}}) + \frac{1}{\lambda}\sum_{ij} p_{ij} \ln \frac{p_{ij}}{p_{ij}^0} + \text{constant}$$

Remark. Monomial resampling:

$$-2\sum_j (\overline{\mathbf{z}}_j^{\mathrm{a}} - \overline{\mathbf{z}}^{\mathrm{a}}) \cdot (\mathbf{z}_j^{\mathrm{f}} - \overline{\mathbf{z}}^{\mathrm{f}}) = 0.$$

Non-classic particle filters (PFs) [Reich and Cotter, 2015]:

$$\mathbf{z}_j^{\mathrm{a}} = \sum_i \mathbf{z}_i^{\mathrm{f}} d_{ij}$$

with transformation matrix $\mathbf{D} = \{d_{ij}\}$ subject to

$$\sum_{i=1}^{M} d_{ij} = 1 \qquad \frac{1}{M}\sum_{j=1}^{M} d_{ij} = \widehat{w}_i.$$

Remark.

- Non-classic PFs been pioneered in the EnKF community.
- Mostly developed for Gaussian likelihood functions and Gaussian approximations to the prior distribution.
- Entries d_{ij} can take negative values.

Non-classic particle filters (PFs) [Reich and Cotter, 2015]:

$$\mathbf{z}_j^{\mathrm{a}} = \sum_i \mathbf{z}_i^{\mathrm{f}} d_{ij}$$

with transformation matrix $\mathbf{D} = \{d_{ij}\}$ subject to

$$\sum_{i=1}^{M} d_{ij} = 1 \qquad \frac{1}{M}\sum_{j=1}^{M} d_{ij} = \widehat{w}_i.$$

Remark.

- Non-classic PFs been pioneered in the EnKF community.
- Mostly developed for Gaussian likelihood functions and Gaussian approximations to the prior distribution.
- Entries d_{ij} can take negative values.

Definition. A LETF is called first-order if

$$\frac{1}{M}\sum_i \mathbf{z}_i^a = \bar{\mathbf{z}}^a$$

i.e.

$$\frac{1}{M}\sum_j d_{ij} = w_i \quad \Longleftrightarrow \quad \mathbf{D1} = M\mathbf{w}$$

Result.

$$\mathbf{D} = \mathbf{P}^\lambda$$

is first-order for any $\lambda \geq 0$.

First-order accuracy

Definition. A LETF is called first-order if

$$\frac{1}{M}\sum_i \mathbf{z}_i^a = \overline{\mathbf{z}}^a$$

i.e.

$$\frac{1}{M}\sum_j d_{ij} = w_i \quad \Longleftrightarrow \quad \mathbf{D1} = M\mathbf{w}$$

Result.

$$\mathbf{D} = \mathbf{P}^\lambda$$

is first-order for any $\lambda \geq 0$.

Second-order accuracy

Definition A first-order LETF is called second-order if

$$\mathbf{D} = \mathbf{w} \otimes \mathbf{1} + \mathbf{S} = \mathbf{P}^0 + \mathbf{S}$$

with $\mathbf{S}$ such that $\mathbf{S1} = \mathbf{0}$ and

$$\mathbf{S}\mathbf{S}^{\mathrm{T}} = M(\mathbf{W} - \mathbf{w} \otimes \mathbf{w})$$

where $\mathbf{W} = \mathrm{diag}\,(\mathbf{w})$.

Remark.

- The posterior samples reproduce the covariance matrix defined through the importance weights.
- $\mathbf{D} = \mathbf{P}^{\lambda}$ is **not** second-order accurate for any λ. In fact, the posterior samples underestimate the covariance.
- But the first-order $\mathbf{D} = \mathbf{P}^{\infty}$ leads to $\widehat{\pi}^{a} \to \pi^{a}$ as $M \to \infty$ (ETPF, [Reich, 2013]) with

$$\widehat{\pi}^{a}(\mathbf{z}) = \frac{1}{M} \sum_{i} \delta(\mathbf{z}_i^{a} - \mathbf{z}).$$

Second-order accuracy

Definition A first-order LETF is called second-order if

$$\mathbf{D} = \mathbf{w} \otimes \mathbf{1} + \mathbf{S} = \mathbf{P}^0 + \mathbf{S}$$

with $\mathbf{S}$ such that $\mathbf{S1} = \mathbf{0}$ and

$$\mathbf{S}\mathbf{S}^{\mathrm{T}} = M(\mathbf{W} - \mathbf{w} \otimes \mathbf{w})$$

where $\mathbf{W} = \mathrm{diag}\,(\mathbf{w})$.

Remark.

- The posterior samples reproduce the covariance matrix defined through the importance weights.
- $\mathbf{D} = \mathbf{P}^\lambda$ is **not** second-order accurate for any λ. In fact, the posterior samples underestimate the covariance.
- But the first-order $\mathbf{D} = \mathbf{P}^\infty$ leads to $\widehat{\pi}^{\mathrm{a}} \to \pi^{\mathrm{a}}$ as $M \to \infty$ (ETPF, [Reich, 2013]) with

$$\widehat{\pi}^{\mathrm{a}}(\mathbf{z}) = \frac{1}{M} \sum_i \delta(\mathbf{z}_i^{\mathrm{a}} - \mathbf{z}).$$

Result. Any second-order accurate LETF is of the form

$$\mathbf{D} = \mathbf{P}^0 + \mathbf{S}\mathbf{Q}, \qquad \mathbf{S} := \sqrt{M}(\mathbf{W} - \mathbf{w} \otimes \mathbf{w})^{1/2},$$

with $\mathbf{Q}$ being an orthogonal matrix s.t. $\mathbf{Q}\mathbf{1} = \mathbf{1}$.

Remark.

- Second-order accurate LETFs have been proposed by [Xiong et al., 2006] and [Tödter and Ahrens, 2015] corresponding to $\mathbf{Q} = \mathbf{I}$ or $\mathbf{Q}$ randomly chosen.
- $\mathbf{D}$ satisfies

$$\sum_i d_{ij} = 1, \qquad \frac{1}{M}\sum_j d_{ij} = w_i$$

  but entries can take negative values contrary to the resampling Markov chains $\mathbf{P}^\lambda$.

Result. Any second-order accurate LETF is of the form

$$\mathbf{D} = \mathbf{P}^0 + \mathbf{S}\mathbf{Q}, \qquad \mathbf{S} := \sqrt{M}(\mathbf{W} - \mathbf{w} \otimes \mathbf{w})^{1/2},$$

with $\mathbf{Q}$ being an orthogonal matrix s.t. $\mathbf{Q}\mathbf{1} = \mathbf{1}$.

Remark.

- Second-order accurate LETFs have been proposed by [Xiong et al., 2006] and [Tödter and Ahrens, 2015] corresponding to $\mathbf{Q} = \mathbf{I}$ or $\mathbf{Q}$ randomly chosen.
- $\mathbf{D}$ satisfies

$$\sum_i d_{ij} = 1, \qquad \frac{1}{M}\sum_j d_{ij} = w_i$$

  but entries can take negative values contrary to the resampling Markov chains $\mathbf{P}^\lambda$.

Find $\mathbf{Q}$ in

$$\mathbf{P} = \mathbf{P}^0 + \mathbf{S}\mathbf{Q}, \qquad \mathbf{S} := \sqrt{M}(\mathbf{W} - \mathbf{w} \otimes \mathbf{w})^{1/2},$$

such that the Sinkhorn cost ($\lambda = \infty$)

$$J(\mathbf{Q}) = -2 \sum_j (\mathbf{z}_j^{\mathrm{a}} - \bar{\mathbf{z}}^{\mathrm{a}}) \cdot (\mathbf{z}_j^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}})$$

is minimized.

Proposition [de Wiljes et al., 2016]

The optimal $\mathbf{Q}$ is given by

$$\mathbf{Q} = \mathbf{U}\mathbf{V}^{\mathrm{T}}$$

with orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ obtained from the SVD of

$$\mathbf{S}(\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}}\mathbf{Z}^{\mathrm{f}} = \mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}.$$

with

$$\mathbf{Z}^{\mathrm{f}} = \left(\mathbf{z}_1^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}, \mathbf{z}_2^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}, \ldots, \mathbf{z}_M^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}\right) \in \mathbb{R}^{N_z \times M}.$$

Find **Q** in

$$\mathbf{P} = \mathbf{P}^0 + \mathbf{S}\mathbf{Q}, \qquad \mathbf{S} := \sqrt{M}(\mathbf{W} - \mathbf{w} \otimes \mathbf{w})^{1/2},$$

such that the Sinkhorn cost ($\lambda = \infty$)

$$J(\mathbf{Q}) = -2 \sum_j (\mathbf{z}_j^{\mathrm{a}} - \bar{\mathbf{z}}^{\mathrm{a}}) \cdot (\mathbf{z}_j^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}})$$

is minimized.

Proposition [de Wiljes et al., 2016]

The optimal **Q** is given by

$$\mathbf{Q} = \mathbf{U}\mathbf{V}^{\mathrm{T}}$$

with orthogonal matrices **U** and **V** obtained from the SVD of

$$\mathbf{S}(\mathbf{Z}^{\mathrm{f}})^{\mathrm{T}}\mathbf{Z}^{\mathrm{f}} = \mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}.$$

with

$$\mathbf{Z}^{\mathrm{f}} = \left(\mathbf{z}_1^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}, \mathbf{z}_2^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}, \ldots, \mathbf{z}_M^{\mathrm{f}} - \bar{\mathbf{z}}^{\mathrm{f}}\right) \in \mathbb{R}^{N_z \times M}.$$

Alternatively: Second-order accurate LETF through a correction to the Sinkhorn approximation [de Wiljes et al., 2016]:

$$\mathbf{D} = \mathbf{P}^{\lambda} + \mathbf{C} = \mathbf{P}^{0} + \mathbf{B} + \mathbf{C}$$

with

$$\mathbf{B} = \mathbf{P}^{\lambda} - \mathbf{P}^{0},$$

and symmetric $\mathbf{C}$ subject to $\mathbf{C1} = \mathbf{0}$.

Requires solution of a **continuous-time algebraic Riccati equation** in $\mathbf{C}$:

$$M(\mathbf{W} - \mathbf{w} \otimes \mathbf{w}) - \mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{C}\mathbf{C} + \mathbf{B}\mathbf{C} + \mathbf{C}\mathbf{B}^{\mathrm{T}}$$

Remark. Fully “observable”, hence solutions exist.

Second-order corrected LETF

Alternatively: Second-order accurate LETF through a correction to the Sinkhorn approximation [de Wiljes et al., 2016]:

$$\mathbf{D} = \mathbf{P}^{\lambda} + \mathbf{C} = \mathbf{P}^{0} + \mathbf{B} + \mathbf{C}$$

with

$$\mathbf{B} = \mathbf{P}^{\lambda} - \mathbf{P}^{0},$$

and symmetric $\mathbf{C}$ subject to $\mathbf{C}\mathbf{1} = \mathbf{0}$.

Requires solution of a **continuous-time algebraic Riccati equation** in $\mathbf{C}$:

$$M(\mathbf{W} - \mathbf{w} \otimes \mathbf{w}) - \mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{C}\mathbf{C} + \mathbf{B}\mathbf{C} + \mathbf{C}\mathbf{B}^{\mathrm{T}}$$

Remark. Fully "observable", hence solutions exist.

Numerical example I

Gaussian prior, non-Gaussian likelihood:

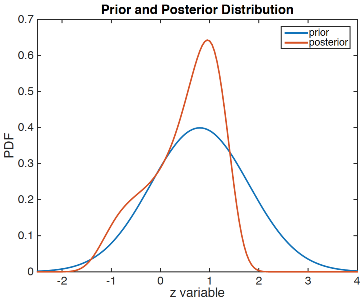


Figure: Prior and posterior distribution for the single Bayesian inference step

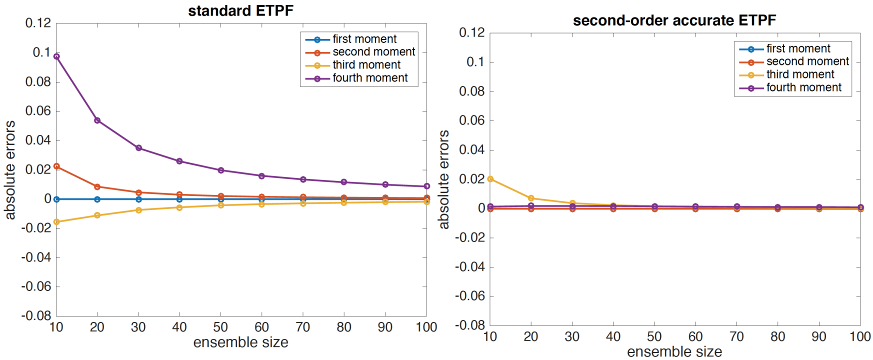


Figure: Absolute errors in the first four moments of the posterior distribution as obtained from the standard Sinkhorn LETF ($\lambda = \infty$) (left panel) and the second-order corrected Sinkhorn LETF (right panel).

Numerical example II

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

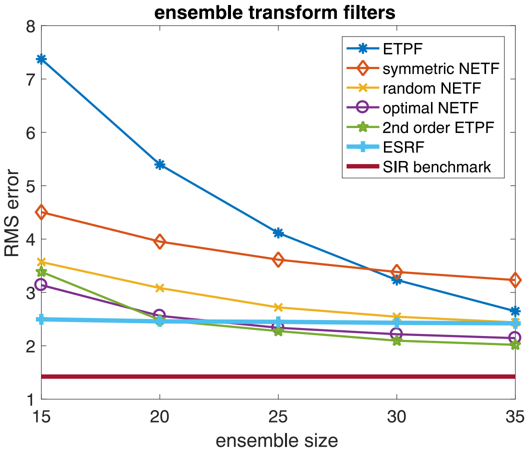


Figure: RMSEs for various second-order accurate LETFs compared to the ETPF, the ESRF, and the SIR PF as a function of the sample size, M.

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

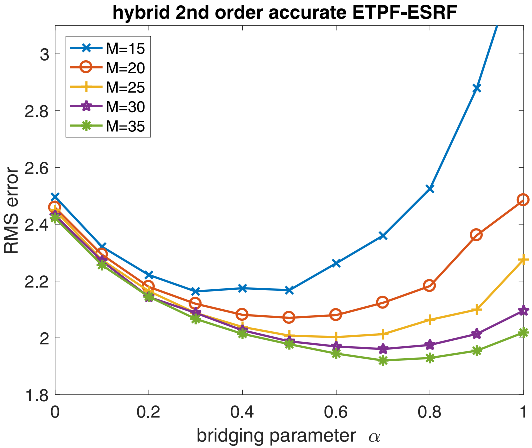


Figure: RMSEs for hybrid ESRF ($\alpha = 0$) and 2nd-order corrected LETF/ETPF ($\alpha = 1$) as a function of the sample size, M.

Lorenz-96 model, discretized nonlinear advection equation, 40 grid points, every second observed.
Hybrid filter $\mathbf{P} := \mathbf{P}_{\mathrm{LETKF}}(\alpha)\, \mathbf{P}_{\mathrm{ETPF}}(1-\alpha)$ + localization.

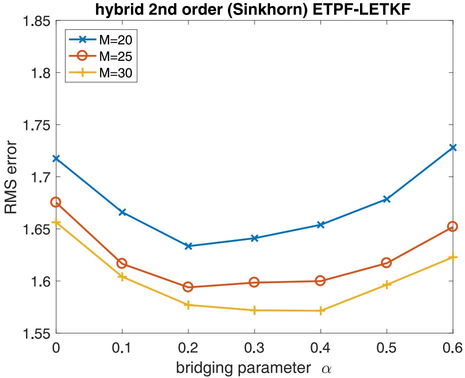


Figure: RMSE for hybrid LETKF ($\alpha = 0$) and 2nd-order corrected LETF/ETPF ($\alpha = 1$).

Conclusions

- The resampling step of a SIR particle filter can be replaced by a deterministic transformation step – **variance reduction, increase in bias**.
- There is a **systematic family of options**: ETPF, NETF, Sinkhorn + 2nd order correction, ... all with pros and cons; **currently being implemented into DWD DA test system**
- All these methods allow for **localization** and **hybridization** with an EnKF [Chustagulprom et al., 2016] and, hence, application to spatially extended systems.
- All these methods can be applied to non-Gaussian likelihoods and combined with **optimal proposal steps** of all flavors.
- Approach is applicable to any problem which requires **coupling of samples** from different distributions (e.g. multi-level MC, pseudo-marginal MCMC, approximation of the Barycenters in the Wasserstein space etc.)

Conclusions

- The resampling step of a SIR particle filter can be replaced by a deterministic transformation step – **variance reduction, increase in bias**.
- There is a **systematic family of options**: ETPF, NETF, Sinkhorn + 2nd order correction, ... all with pros and cons; **currently being implemented into DWD DA test system**
- All these methods allow for **localization** and **hybridization** with an EnKF [Chustagulprom et al., 2016] and, hence, application to spatially extended systems.
- All these methods can be applied to non-Gaussian likelihoods and combined with **optimal proposal steps** of all flavors.
- Approach is applicable to any problem which requires **coupling of samples** from different distributions (e.g. multi-level MC, pseudo-marginal MCMC, approximation of the Barycenters in the Wasserstein space etc.)

Conclusions

- The resampling step of a SIR particle filter can be replaced by a deterministic transformation step – **variance reduction, increase in bias**.
- There is a **systematic family of options**: ETPF, NETF, Sinkhorn + 2nd order correction, ... all with pros and cons; **currently being implemented into DWD DA test system**
- All these methods allow for **localization** and **hybridization** with an EnKF [Chustagulprom et al., 2016] and, hence, application to spatially extended systems.
- All these methods can be applied to non-Gaussian likelihoods and combined with **optimal proposal steps** of all flavors.
- Approach is applicable to any problem which requires **coupling of samples** from different distributions (e.g. multi-level MC, pseudo-marginal MCMC, approximation of the Barycenters in the Wasserstein space etc.)

Conclusions

- The resampling step of a SIR particle filter can be replaced by a deterministic transformation step – **variance reduction, increase in bias**.
- There is a **systematic family of options**: ETPF, NETF, Sinkhorn + 2nd order correction, ... all with pros and cons; **currently being implemented into DWD DA test system**
- All these methods allow for **localization** and **hybridization** with an EnKF [Chustagulprom et al., 2016] and, hence, application to spatially extended systems.
- All these methods can be applied to non-Gaussian likelihoods and combined with **optimal proposal steps** of all flavors.
- Approach is applicable to any problem which requires **coupling of samples** from different distributions (e.g. multi-level MC, pseudo-marginal MCMC, approximation of the Barycenters in the Wasserstein space etc.)

Conclusions

- The resampling step of a SIR particle filter can be replaced by a deterministic transformation step – **variance reduction, increase in bias**.
- There is a **systematic family of options**: ETPF, NETF, Sinkhorn + 2nd order correction, ... all with pros and cons; **currently being implemented into DWD DA test system**
- All these methods allow for **localization** and **hybridization** with an EnKF [Chustagulprom et al., 2016] and, hence, application to spatially extended systems.
- All these methods can be applied to non-Gaussian likelihoods and combined with **optimal proposal steps** of all flavors.
- Approach is applicable to any problem which requires **coupling of samples** from different distributions (e.g. multi-level MC, pseudo-marginal MCMC, approximation of the Barycenters in the Wasserstein space etc.)

References I

Chustagulprom, N., Reich, S., and Reinhardt, M. (2016).
A hybrid ensemble transform filter for nonlinear and spatially extended dynamical systems.
SIAM/ASA J Uncertainty Quantification, 4:592–608.

Cuturi, M. (2013).
Sinkhorn distances: Lightspeed computation of optimal transport.
In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc.

de Wiljes, J., Acevedo, W., and Reich, S. (2016).
A second-order accurate ensemble transform particle filter.
Technical Report ArXiv:1608.08179, University of Potsdam.

Reich, S. (2013).
A non-parametric ensemble transform method for Bayesian inference.
SIAM J Sci Comput, 35:A2013–A2014.

References II

Reich, S. and Cotter, C. (2015).
Probabilistic Forecasting and Bayesian Data Assimilation.
Cambridge University Press.

Tödter, J. and Ahrens, B. (2015).
A second-order exact ensemble square root filter for nonlinear data assimilation.
Mon. Wea. Rev., 143:1347–1367.

Xiong, X., Navon, I., and Uzungoglu, B. (2006).
A note on the particle filter with posterior Gaussian resampling.
Tellus, 85A:456–460.