A smooth and scalable localised ensemble transport particle filter

MATT GRAHAM (m.m.graham@nus.edu.sg)

ALEX THIERY (a.h.thiery@nus.edu.sg)

Motivation

Filtering in spatially-extended dynamical systems is a challenging task with important applications such as *numerical weather prediction* (NWP). The *ensemble Kalman filter* (EnKF) is a popular inference method in geophysical models, however in systems with non-linear dynamics or observation operators or non-Gaussian noise inference quality can be poor. *Particle filters* (PFs) provide consistent inferences in more general models, however the required ensemble size for accurate inference scales exponentially with dimension. *Localisation* approaches which exploit low dependence between state variables at spatially distant points have been key to scaling EnKF methods to large spatially-extended models. Localising PF methods is however challenging as the resampling step can introduce artificial discontinuities in the system state when applied locally.

Model definition and notation

Generative model:

$\mathbf{x}_1 = F_1(\mathbf{u}_1)$, $\quad \mathbf{u}_1 \sim \mu_1$,

$\mathbf{x}_t = F_t(\mathbf{x}_{t-1}; \mathbf{u}_t)$, $\quad \mathbf{u}_t \sim \mu_t \quad \forall t \in 2{:}T$,

$\mathbf{y}_t = G_t(\mathbf{x}_t; \mathbf{v}_t)$, $\quad \mathbf{v}_t \sim \nu_t \quad \forall t \in 1{:}T$.

System state $\mathbf{x}_t \in \mathcal{X} \quad \forall t \in 1{:}T$,
observation $\mathbf{y}_t \in \mathcal{Y} \quad \forall t \in 1{:}T$,
state noise $\mathbf{u}_t \in \mathcal{U} \quad \forall t \in 1{:}T$,
observation noise $\mathbf{v}_t \in \mathcal{V} \quad \forall t \in 1{:}T$.

Observation distribution $\quad g_t(\boldsymbol{y} \,|\, \boldsymbol{x}_t)\,\eta(\mathrm{d}\boldsymbol{y}) = \mathbb{P}(\mathbf{y}_t \in \mathrm{d}\boldsymbol{y} \,|\, \mathbf{x}_t = \boldsymbol{x}_t)$

Filtering distribution $\quad \pi_{t|t}(\mathrm{d}\boldsymbol{x}) = \mathbb{P}(\mathbf{x}_t \in \mathrm{d}\boldsymbol{x} \,|\, \mathbf{y}_1 = \boldsymbol{y}_1, \dots \mathbf{y}_t = \boldsymbol{y}_t)$

Predictive distribution $\quad \pi_{t+1|t}(\mathrm{d}\boldsymbol{x}) = \mathbb{P}(\mathbf{x}_{t+1} \in \mathrm{d}\boldsymbol{x} \,|\, \mathbf{y}_1 = \boldsymbol{y}_1, \dots \mathbf{y}_t = \boldsymbol{y}_t)$

Filtering problem

Infer filtering distributions $\{\pi_{t|t}\}_{t=1}^{T}$ given observations $\{\boldsymbol{y}_t\}_{t=1}^{T}$.

Filtering distributions can be computed recursively by iterating two updates:

$$\cdots \longrightarrow \pi_{t|t} \xrightarrow{\text{prediction}} \pi_{t+1|t} \xrightarrow{\text{assimilation}} \pi_{t+1|t+1} \longrightarrow \cdots$$

Prediction update: $$\pi_{t+1|t}(\mathrm{d}\boldsymbol{x}) = \int_{\mathcal{U}} \int_{\mathcal{X}} \delta_{F_{t+1}(\boldsymbol{x}'; \boldsymbol{u})}(\mathrm{d}\boldsymbol{x})\, \pi_{t|t}(\mathrm{d}\boldsymbol{x}')\, \mu_{t+1}(\mathrm{d}\boldsymbol{u}).$$

Assimilation update: $$\pi_{t+1|t+1}(\mathrm{d}\boldsymbol{x}) = \frac{g_{t+1}(\boldsymbol{y}_{t+1} \,|\, \boldsymbol{x})}{\int_{\mathcal{X}} g_{t+1}(\boldsymbol{y}_{t+1} \,|\, \boldsymbol{x}')\, \pi_{t+1|t}(\mathrm{d}\boldsymbol{x}')} \pi_{t+1|t}(\mathrm{d}\boldsymbol{x}).$$

Ensemble filters

Particle ensemble $\{\boldsymbol{x}_{t|t}^{n}\}_{n=1}^{N}$ approximates $\pi_{t|t}$ at each time index t

$$\pi_{t|t}(\mathrm{d}\boldsymbol{x}) \approx \frac{1}{N} \sum_{n=1}^{N} \delta_{\boldsymbol{x}_{t|t}^{n}}(\mathrm{d}\boldsymbol{x}).$$

Prediction update: $\boldsymbol{u}_{t+1}^{n} \sim \mu_{t+1}$, $\boldsymbol{x}_{t+1|t}^{n} = F_{t+1}(\boldsymbol{x}_{t|t}^{n}; \boldsymbol{u}_{t+1}^{n}) \quad \forall n \in 1{:}N$.

Ensemble Kalman filter [1,2]

Simulate observations: $\boldsymbol{v}_{t+1}^{n} \sim \nu_{t+1}$, $\boldsymbol{y}_{t+1|t}^{n} = G_{t+1}(\boldsymbol{x}_{t+1|t}^{n}; \boldsymbol{v}_{t+1}^{n}) \quad \forall n \in 1{:}N$.

Assimilation update: $\quad \boldsymbol{x}_{t+1|t+1}^{n} = \boldsymbol{x}_{t+1|t}^{n} + \boldsymbol{K}_{t+1}(\boldsymbol{y}_{t+1} - \boldsymbol{y}_{t+1|t}^{n}) \quad \forall n \in 1{:}N$

$\boldsymbol{K}_{t+1}$ is a Monte Carlo estimate of Kalman gain using $\{\boldsymbol{x}_{t+1|t}^{n}, \boldsymbol{y}_{t+1|t}^{n}\}_{n=1}^{N}$.

Consistent as $N \to \infty$ only when F_t, G_t linear and μ_t, ν_t Gaussian $\forall t \in 1{:}T$.

Particle filter [3,4]

Importance weights: $\tilde{w}_{t+1}^{n} = g_{t+1}(\boldsymbol{y}_{t+1} \,|\, \boldsymbol{x}_{t+1|t}^{n})$, $\quad w_{t+1}^{n} = \frac{\tilde{w}_{t+1}^{n}}{\sum_{m=1}^{N} \tilde{w}_{t+1}^{m}} \quad \forall n \in 1{:}N$.

Assimilation update: $\quad \boldsymbol{x}_{t+1|t+1}^{n} = \sum_{m=1}^{N} \mathsf{R}_{t+1}^{n,m} \boldsymbol{x}_{t+1|t}^{m} \quad \forall n \in 1{:}N$,

with matrix $\mathbf{R}_{t+1} \in \{0,1\}^{N \times N}$ satisfying $\mathbf{R}_{t+1}\mathbf{1} = \mathbf{1}$, $\quad \mathbb{E}\left[\mathbf{R}_{t+1}^{\mathsf{T}}\mathbf{1}\right] = N\boldsymbol{w}_{t+1}$.

Consistent as $N \to \infty$ for arbitrary $\{F_t, G_t, \mu_t, \nu_t\}$, however ensemble size N required for fixed error scales exponentially with model dimension [5].

Ensemble transform particle filter [6]

Assimilation update: $\quad \boldsymbol{x}_{t+1|t+1}^{n} = \sum_{m=1}^{N} T_{t+1}^{n,m} \boldsymbol{x}_{t+1|t}^{m} \quad \forall n \in 1{:}N$,

with matrix $\boldsymbol{T}_{t+1}$ the solution to the discrete *optimal transport* (OT) problem

$$\boldsymbol{T}_{t+1} = \operatorname*{argmin}_{\boldsymbol{T} \in \mathcal{P}_N(\boldsymbol{w}_{t+1})} \boldsymbol{T} \cdot \boldsymbol{C}_{t+1}, \quad \mathcal{P}_N(\boldsymbol{w}) = \left\{ \boldsymbol{T} \in \mathbb{R}^{N \times N} : \boldsymbol{T}\mathbf{1} = \mathbf{1},\ \boldsymbol{T}^{\mathsf{T}}\mathbf{1} = N\boldsymbol{w} \right\}$$

Typically $C_{t+1}^{m,n} = |\boldsymbol{x}_{t+1|t}^{m} - \boldsymbol{x}_{t+1|t}^{n}|_2^2 \quad \forall m \in 1{:}N,\ n \in 1{:}N$.

Consistent as $N \to \infty$. Still requires exponential scaling of N with dimension.

Spatially extended systems

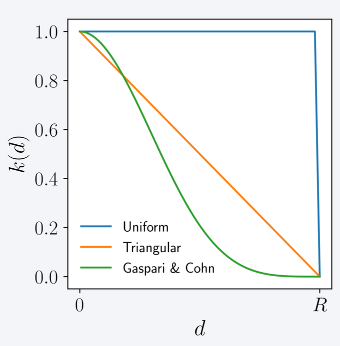
Let $(\mathcal{S}, d)$ be a compact metric space representing the spatial domain and $\tilde{\mathcal{X}}$ the field the state variables take values in for each $s \in \mathcal{S} \Rightarrow \mathcal{X} = \tilde{\mathcal{X}}^{\mathcal{S}}$.

Assume system is independently observed at a finite set of locations $\{s_d \in \mathcal{S}\}_{d=1}^{D}$ so that the log observation density decomposes as

$$\log g_t(\boldsymbol{y}_t \,|\, \boldsymbol{x}) = \sum_{d=1}^{D} \ell_{t,d}(\boldsymbol{y}_t(s_d) \,|\, \boldsymbol{x}).$$

In practical implementations $\mathcal{S}$ will be discretised in to a mesh. The state space $\mathcal{X}$ is then a vector space with a finite but typically high dimension – e.g. operational NWP systems have meshes with order of 10^8 elements.

Localisation [7,8]



For systems with large spatial domains, typically state variables at distant points will have low dependence - *decay of correlations* property. In EnKF methods this property is exploited to perform *localised* assimilation updates. A *localisation kernel* $k : \mathbb{R}_+ \to [0,1]$ weights the effect of the observations at each s_d on the state variables at a point s by $k \circ d(s, s_d)$. The kernel is chosen such that $k(0) = 1$ and $k(r) = 0 \ \ \forall r > R$ for a *localisation radius* R.

Localised particle filters [9,10]

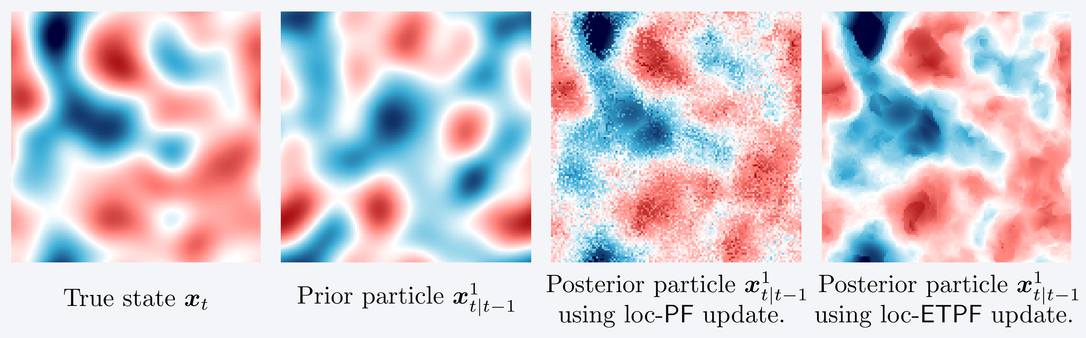
Localised log weights: $\quad \log \tilde{w}_t^{n}(s) = \sum_{d=1}^{D} \ell_{t,d}(\boldsymbol{y}_t(s_d) \,|\, \boldsymbol{x}_{t|t-1}^{n})\, k \circ d(s, s_d).$

How to implement assimilation update with localised weights?

Loc-PF: Generate resampling matrices $\mathbf{R}_t(s)$ independently for each $s \in \mathcal{S}$ leads to noise in state fields which are typically spatially smooth.

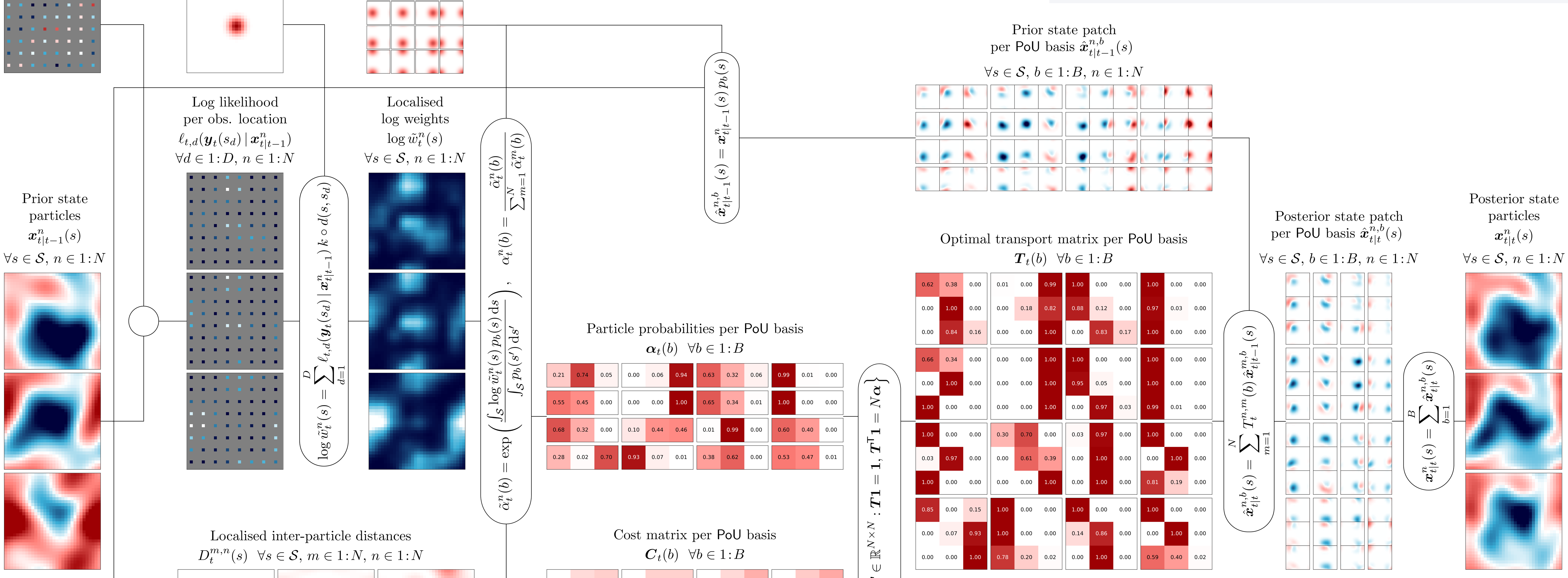
Loc-ETPF [11]: Computing OT matrix $\boldsymbol{T}_t(s)$ for each $s \in \mathcal{S}$ decreases discontinuities however still non-smooth and computationally demanding.

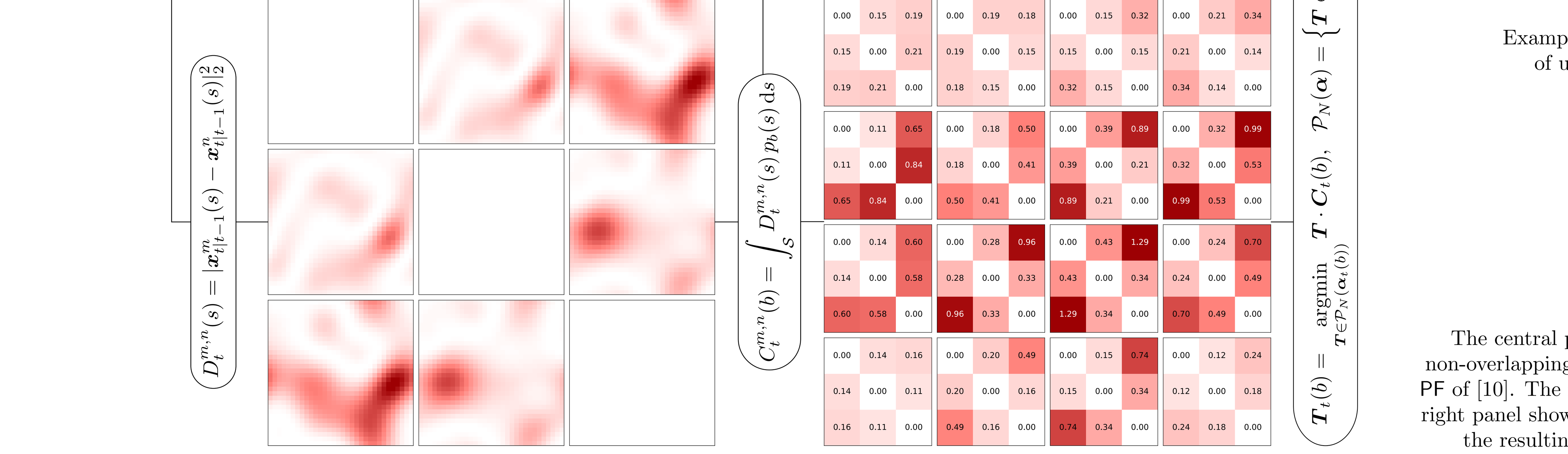
Example of applying localised PF assimilation updates to two-dimensional spatial Gaussian process model:



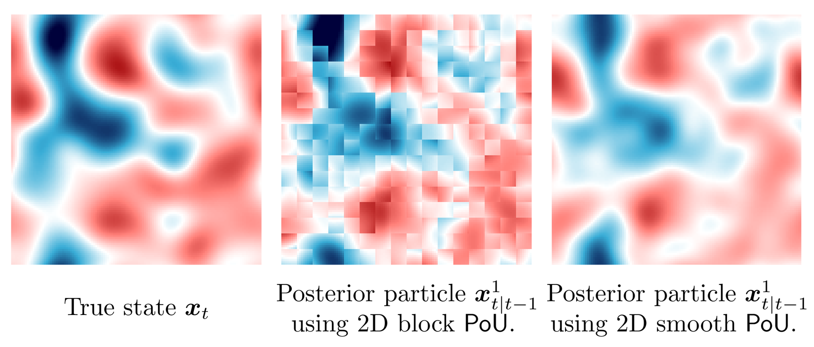
Proposed method

We propose an alternative localised ETPF method which is able to maintain spatial smoothness in the updated particles while also significantly reducing the computational cost by requiring few OT problems to be solved.





Examples proposed localised assimilation updates using two different partitions of unity applied to spatial Gaussian process model example from above:



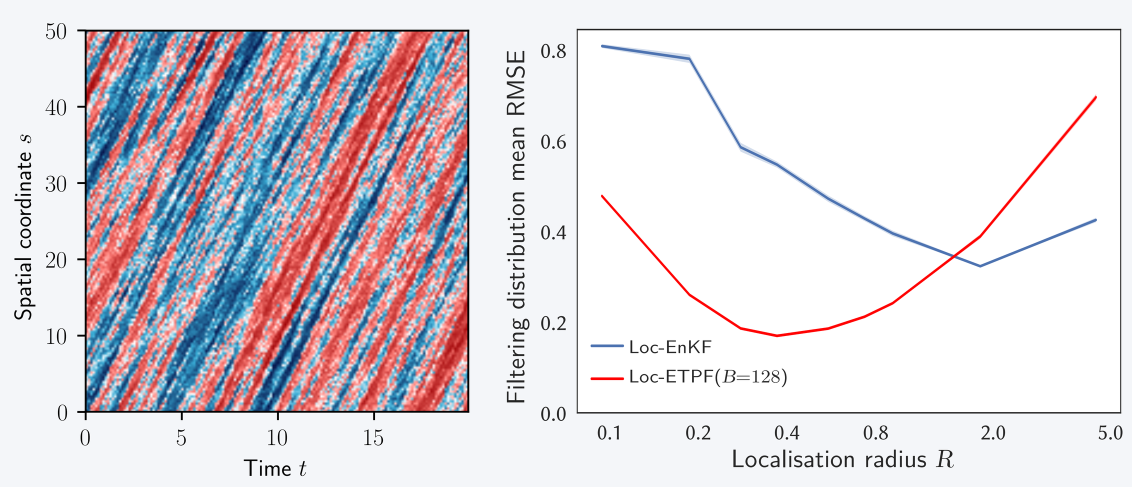
The central panel shows a particle computed using a piecewise-constant PoU consisting of non-overlapping rectangular blocks, with the resulting algorithm a ETPF variant of the blocked PF of [10]. The resulting state particles show discontinuities reflecting the non-smooth PoU. The right panel shows a particle computed using a PoU formed of smooth 2D 'bump' functions, with the resulting particles maintaining the smoothness of the inputted prior state particles.

1D transformed stochastic turbulence model

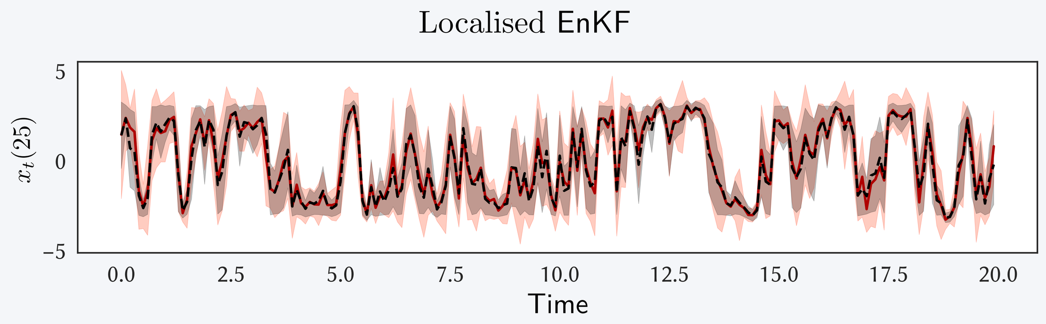
As a test model we considered a bijective non-linear transformation of a linear-Gaussian model. Specifically for a smooth bijection $\boldsymbol{\phi} : \mathcal{X} \to \mathcal{X}$ we simulate a model with state update and observation operators

$$F_t(\boldsymbol{x}; \boldsymbol{u}) = \boldsymbol{\phi}\left(\boldsymbol{A}\boldsymbol{\phi}^{-1}(\boldsymbol{x}) + \boldsymbol{B}\boldsymbol{u}\right) \quad \text{and} \quad G_t(\boldsymbol{x}; \boldsymbol{v}) = \boldsymbol{H}\boldsymbol{\phi}^{-1}(\boldsymbol{x}) + \boldsymbol{v},$$

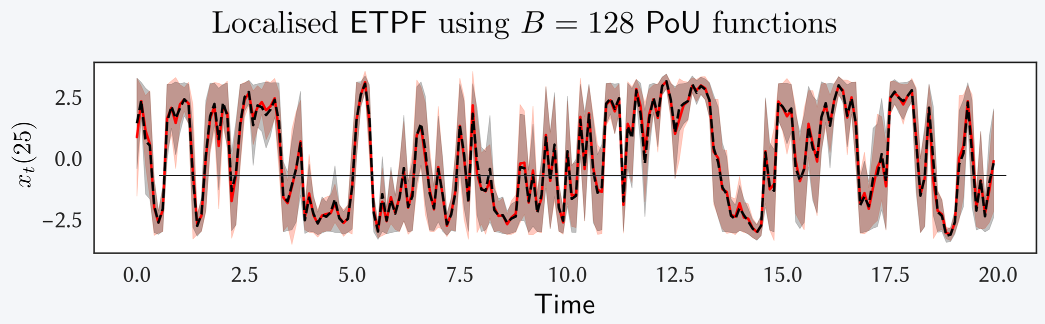
and Gaussian state and observation noise distributions μ_t and ν_t. We performed exact inference in the underlying linear-Gaussian model using a Kalman filter and pushed samples from the Gaussian filtering distributions through $\boldsymbol{\phi}$ to get exact samples from the non-Gaussian filtering distributions. For the base linear-Gaussian model we used a 1D stochastic partial differential equation model of turbulence on a periodic domain proposed in [12].

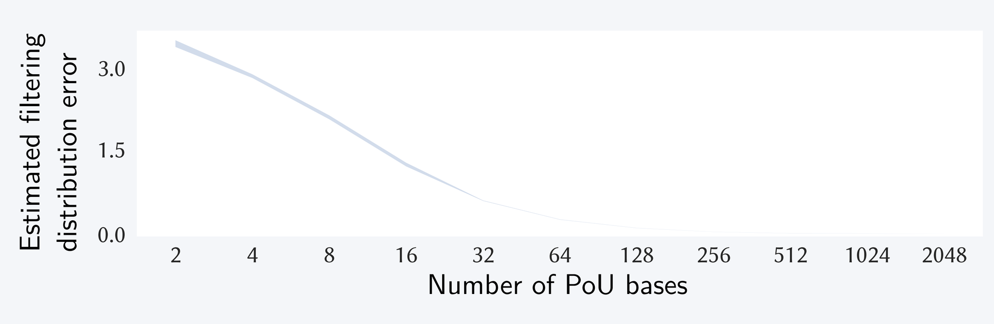


Localised EnKF



Localised ETPF using $B = 128$ PoU functions





References

1. Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Mon Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* **99** 10143–10162.
2. Burgers, G., van Leeuwen, P. J. and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review* **126** 1719–1724.
3. Gordon, N. J., Salmond, D. J. and Smith, A. F. (1993). Novel approach to nonlinear/non- Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)* **140** 107–113. IET.
4. Del Moral, P. (1996). Non-linear filtering: interacting particle resolution. *Markov processes and relate fields* **2** 555–581.
5. Snyder, C., Bengtsson, T., Bickel, P. and Anderson, J. (2008). Obstacles to high- dimensional particle filtering. *Monthly Weather Review* **136** 4629–4640.
6. Reich, S. (2013). A nonparametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing* **35** A2013–A2024.
7. Hamill, T. M., Whitaker, J. S. and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review* **129** 2776–2790.
8. Hunt, B. R., Kostelich, E. J. and Szunyogh, I. (2007). Efficient data assimilation for spatiotemp chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena* **230** 112–126.
9. Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review* 4089–4114.
10. Rebeschini, P., Van Handel, R. (2015). Can local particle filters beat the curse of dimensionalit *The Annals of Applied Probability* **25** 2809–2866.
11. Cheng, Y. and Reich, S. (2015). Assimilating data into scientific models: An optimal coupling spective. In *Nonlinear Data Assimilation* 75–118. Springer.
12. Majda, A.J. and Harlim, J., (2012). *Filtering complex turbulent systems.* CUP.