Sequential Learning Methods for High-Dimensional Data Assimilation

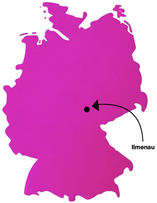
Jana de Wiljes

jana.de-wiljes@tu-ilmenau.de

www.dewiljes-lab.com

21st international EnKF workshop (2026)

SFB 1294 Data Assimilation



TECHNISCHE UNIVERSITÄT ILMENAU

UrbanAIR

The Origin Story...

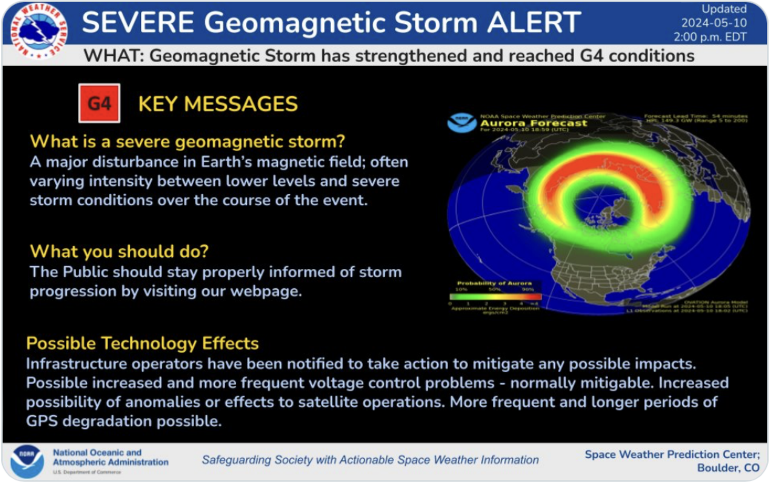
Polar lights in surprising places

Friday 10 of May 2024 in Switzerland, Southern California, and Tuscany

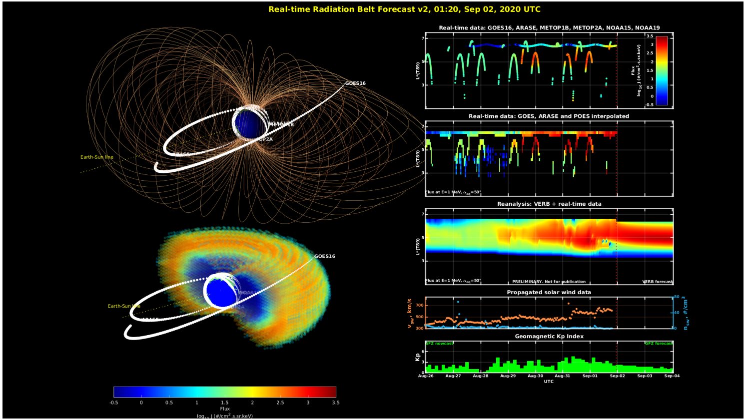
The other side of the coin...

NOAA Space Weather Prediction Center @NWSSWPC · May 10

G4 conditions have been observed...



Space weather: estimating the effects of solar storms



A. M. Castillo Tibocha, J. de Wiljes, Y. Shprits, and N. A. Aseev (2021). *Reconstructing the dynamics of the outer electron radiation belt by means of ensemble Kalman filtering with the VERB-3D Code*. Space Weather 19(10), e2020SW002672.

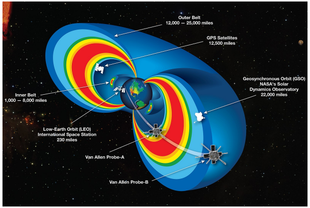
How many observations are necessary ?

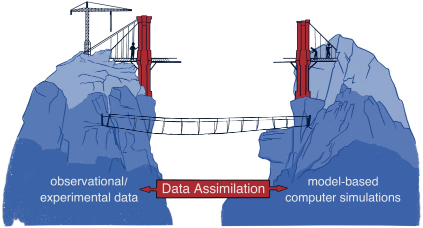
Goal: Estimating the phase space density $f_\alpha(t, x, v)$ of charged particles α inside of Earth's magnetosphere

Model

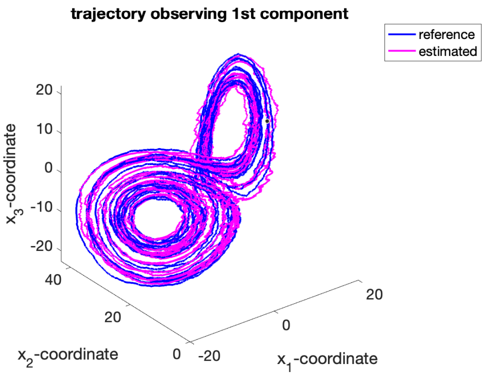
$$
\begin{aligned}
\partial_t f_\alpha + v_\alpha \cdot \nabla f_\alpha + Z_\alpha e\,(E + \frac{v_\alpha}{c} \times B) \cdot \partial_v f_\alpha &= 0 \\
\nabla \cdot E - 4\pi e \int \sum Z_\alpha f_\alpha dv &= 0 \\
\text{Vlasov-Poisson equation} \qquad \nabla \times E &= 0 \qquad 1
\end{aligned}
$$

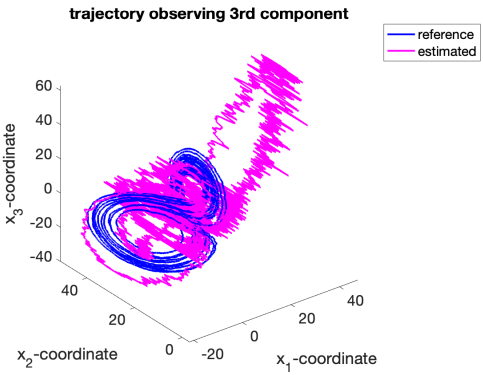
Data





Where should I observe? What frequency?





L63:

- observing first or second component leads to good results
- observing only third component leads to blow-up

Continuous filtering problem

Mathematical framework:

$$(\textit{Model}) \qquad \mathrm{d}X_t = f(X_t, \lambda)\mathrm{d}t + \sqrt{2}C\mathrm{d}W_t$$

$$(\textit{Observations}) \qquad \mathrm{d}Y_t = h(X_t)\mathrm{d}t + R^{1/2}\mathrm{d}V_t$$

Filtering Problem: Estimate filtering distribution

$$\pi_t(A) = \mathbb{P}[X_t \in A | Y_{0:t}]$$

Kushner-Stratonovich equation:

$$\begin{aligned} \mathrm{d}\pi_t[g] &= \pi_t[f \cdot \nabla g]\mathrm{d}t + \pi_t[\nabla \cdot D\nabla g]\mathrm{d}t \\ &\quad + \left(\pi_t[gh] - \pi_t[g]\pi_t[h]\right)^{\mathrm{T}} R^{-1} \left(\mathrm{d}Y_t - \pi_t[h]\mathrm{d}t\right) \end{aligned}$$

where

$$\pi_t[g] := \int_{\mathbb{R}^{N_x}} g(x)\pi_t(x)dx$$

Linear Model: Kalman- Bucy Filter

$$\mathrm{d}\bar{x}_t = A\bar{x}_t\mathrm{d}t + b\mathrm{d}t - P_t H^{\mathrm{T}} R^{-1}(H\bar{x}_t\mathrm{d}t - \mathrm{d}Y_t)$$

Ensemble Kalman Bucy Filter (EnKBF)

Deterministic EnKBF:

$$\mathrm{d}X_t^i = f(X_t^i)\mathrm{d}t + C\mathrm{d}W_t^i - \frac{1}{2}P_t^M H^{\mathrm{T}} R^{-1}\left(h(X_t^i)\mathrm{d}t + \bar{h}_t^M \mathrm{d}t - 2\mathrm{d}Y_t\right)$$

Equations of motion closed through the empirical mean $\bar{x}_t^M$ and empirical covariance matrix P_t^M

$$\bar{h}_t^M = \frac{1}{M}\sum_{i=1}^{M} h(X_t^i), \quad P_t^M = \frac{1}{M-1}\sum_{i=1}^{M}(X_t^i - \bar{x}_t^M)(X_t^i - \bar{x}_t^M)^{\mathrm{T}}$$

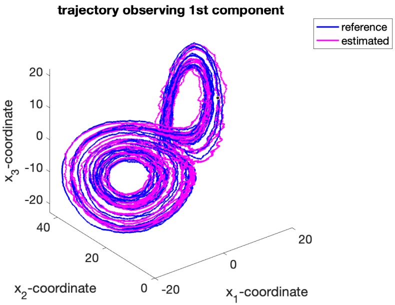
Interacting particle representation of the model error:

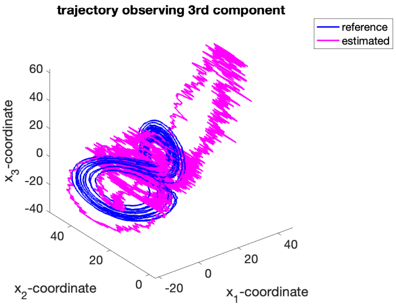
$$\begin{aligned}\mathrm{d}X_t^i = f(X_t^i)\mathrm{d}t &+ CC^\top (P_t^M)^{-1}(X_t^i - \bar{x}_t^M)\mathrm{d}t \\ &- \frac{1}{2}P_t^M H^{\mathrm{T}} R^{-1}\left(h(X_t^i)\mathrm{d}t + \bar{h}_t^M \mathrm{d}t - 2\mathrm{d}Y_t\right)\end{aligned}$$

J. de Wiljes, S. Reich, and W. Stannat (2018). *Long-time stability and accuracy of the ensemble Kalman-Bucy filter for fully observed processes and small measurement noise.* SIAM Journal on Applied Dynamical Systems, 17(2), 1152–1181

A. Taghvaei, J. de Wiljes, P. G. Mehta, S. Reich (2018). *Kalman Filter and its Modern Extensions for the Continuous-time Nonlinear Filtering Problem* Journal of Dynamic Systems, Measurements, and Control, 140(3): 030904

Upper and lower bounds on covariance





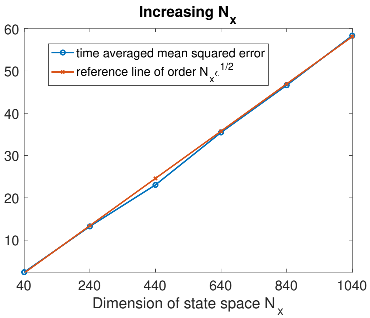
Proposition: Under the assumption of *short range interactions* and $dX_t = f(X_t)dt + \sqrt{2}dW_t$ and $dY_t = X_t dt + R^{1/2} dB_t$ with $R = \Theta(\epsilon)$ we obtain bounds:

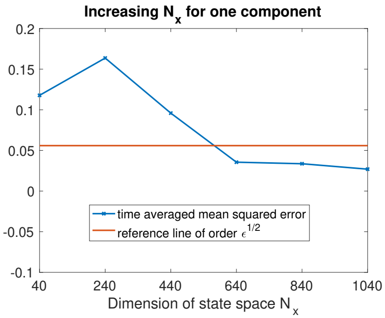
$$\Theta(\sqrt{\epsilon}) = \lambda_{\min}(\epsilon) \leq \|P_t\|_{\min} \leq \lambda_{\max}(\epsilon) = \Theta(\sqrt{\epsilon}) \text{ for } t > t_* = \Theta(\sqrt{\epsilon})$$

J. de Wiljes, S. Reich, and W. Stannat (2018). *Long-time stability and accuracy of the ensemble Kalman-Bucy filter for fully observed processes and small measurement noise.* SIAM Journal on Applied Dynamical Systems, 17(2), 1152–1181

J. de Wiljes and X. T. Tong (2020). *Analysis of a localised nonlinear Ensemble Kalman Bucy Filter with complete and accurate observations.* Nonlinearity, 33(9), 4752–4782

Analysis for component-wise error



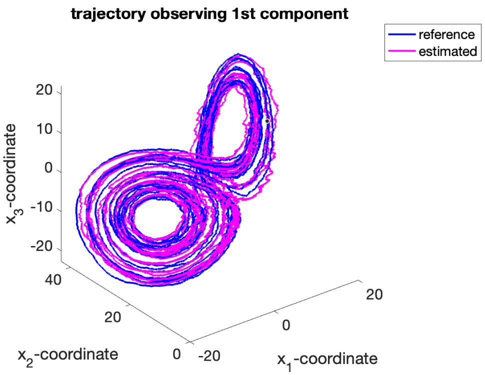


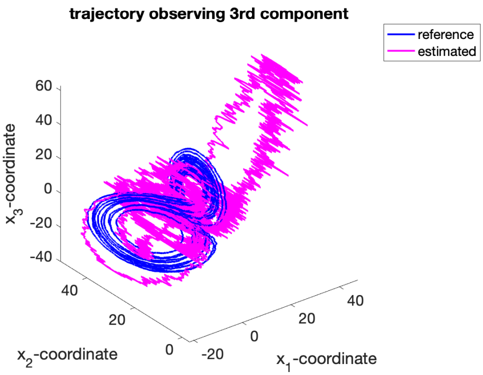
Theorem: Under the previous assumptions the following holds:

1. $\mathbb{E}[e_i^2(t)] \leq C\sqrt{\epsilon} \quad \forall i$ **(Expected error for fixed** t**)**
2. $$\mathbb{E}_{t_0}\left[\sup_{t_0 \leq t \leq T} e_i^2(t)\right] \leq \max_i\{e_i^2(t_0)\} + C\sqrt{\epsilon}\log(T/\sqrt{\epsilon}) \quad \forall T > t_0 \text{ (\textbf{Uniform Error in} } [t_0, T])$$

Let's be a bit more realistic

Partially observed setting

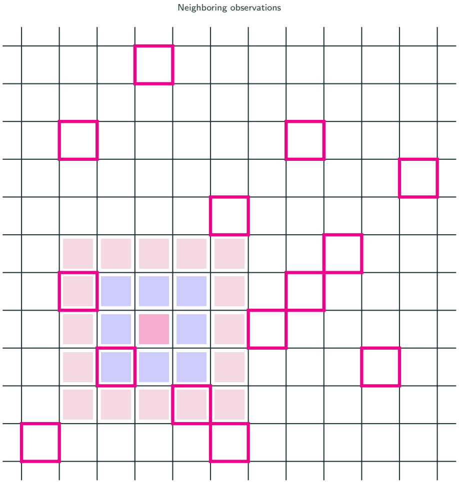




Ansatz to combat challenge:

- Use properties of the underlying system (ODE or PDE) to determine crucial components
- randomise

Fixed partial observations



Need bound for:

$$\frac{d}{dt}P_t = (F_t + F_t^T) + (P_t^\dagger P_t + P_t P_t^\dagger) - \frac{1}{2\epsilon}(P_t^L \Omega P_t + P_t \Omega P_t^L)$$

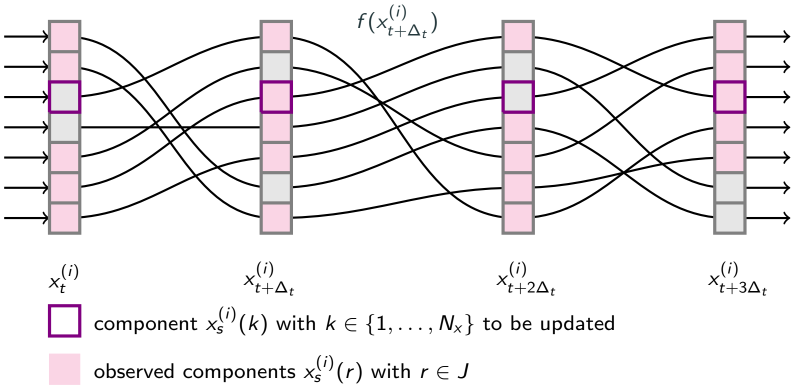
Assume: have an observation at location j. Considering component k ($k \neq j$) yields:

$$[P_t^L \Omega P_t]_{k,k} = [P_t]_{k,j}^2 \phi_{k,j} = c_{j,k} \phi_{k,j} [P_t]_{k,k}^2 \tag{1}$$

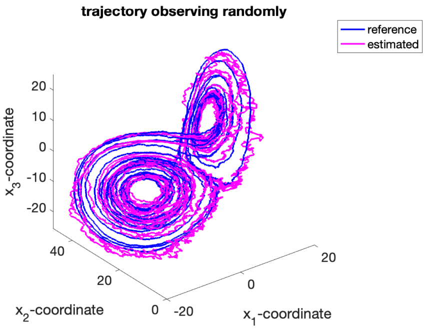
$$\mathbb{E}[e_k(t)^2] \leq \begin{cases} C_j \sqrt{\varepsilon}, & c_{j,k} = \Theta(1), \\ C, & c_{j,k} = \Theta(\varepsilon). \end{cases} \tag{2}$$

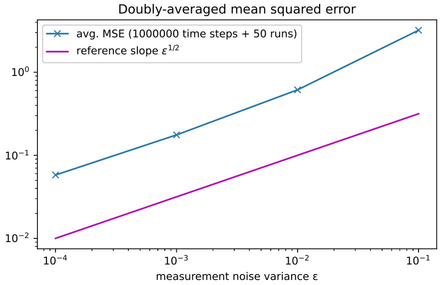
Randomised partial observations every time step

Uniformly random draws of the set $J_t \sim \mathcal{J}$ with $J_t \subset \{1, \dots, N_x\}$ with and fixed $N_J := |J|$ fixed



Expected error over random observations





Theorem: For $|J_\tau|$ *large enough*, i.e., for all k the probability that there is a $j \in J$ with $c_{j,k} \leq \kappa$ is large, yields

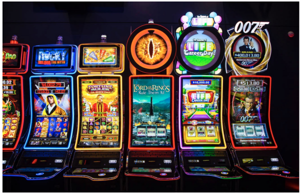
$$\mathbb{E}_{J\sim\mathcal{J}}\left[\mathbb{E}[e_i^2(t)]\right] \leq C_\tau^J\sqrt{\epsilon} \quad \forall i \text{ } \textbf{(Expected error for fixed t)}$$

N. Abedini , J. de Wiljes, and S. Dubinkina (2025). *Filtering with Randomised Observations: Sequential Learning of Relevant Subspace Properties and Accuracy Analysis* arXiv:2509.04867

Intro to Sequential Learning

Multi-armed bandits

Choose from K options
to receive a
high reward and
to reduce loss after T rounds



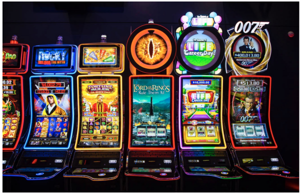
Examples:

- Which advertising campaign generates the largest revenue?

a_1 a_2 a_3 a_4

Multi-armed bandits

Choose from K options
to receive a
high reward and
to reduce loss after T rounds



Examples:

- Which vaccine should enter next stage of clinical trials?

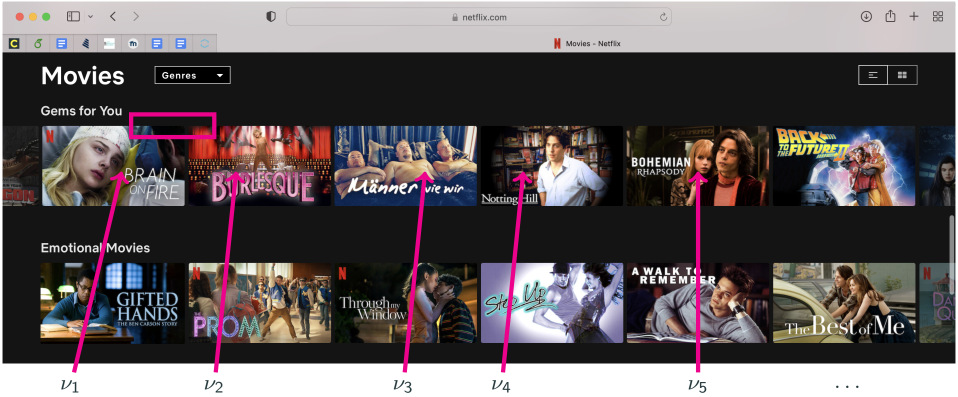
a_1

a_2

a_3

a_4

Example recommender system



For the t-th visit of the app/website

- recommend a movie a_t
- observe $R_t \sim \nu_{a_t}$ (e.g., a rating, number of clicks or of times watched)

Multi-armed bandits

Stochastic K-Armed Bandit

A stochastic K-Armed Bandit is a collection of distributions

$$\nu = (\nu_a : a \in \mathcal{A})$$

where $\mathcal{A}$ is a set of actions (arms) and $|\mathcal{A}| = K$ and

$$\mu_a(\nu) = \int_{-\infty}^{\infty} x \nu_a(x) dx \tag{3}$$

Procedure: in each round $t \in \{1, \ldots, T\}$

1. learner chooses an action $A_t = a$
2. receives reward $R_t \sim \nu_a$ (independent from the past)

Bernoulli example

Bernoulli setting:

- $\{(\mathcal{B}(\mu_i))_i : \mu_i \in [0,1]\}$
- Reward when choosing **arm a** at **time t**

$$R_t = \begin{cases} 1 & \text{with probability } \mu_a \\ 0 & \text{with probability } 1-\mu_a \end{cases}$$

Example $|\mathcal{A}| = 3$:

Stochastic Bandit Problem

Let the the largest mean of all the arms be denoted

$$\mu^*(\nu) = \max_{a \in \mathcal{A}} \mu_a(\nu)$$

Regret

The T-period regret of the sequence of random actions $a_1, \ldots, a_T$ is the random variable

$$\mathcal{R}(\nu, T) = T\mu^*(\nu) - \mathbb{E}[\sum_{t=1}^{T} R_t] \tag{4}$$

Goal: find a sequential sampling strategy

$$A_{t+1} = \pi_t(A_1, R_1, \ldots, A_t, R_t) \tag{5}$$

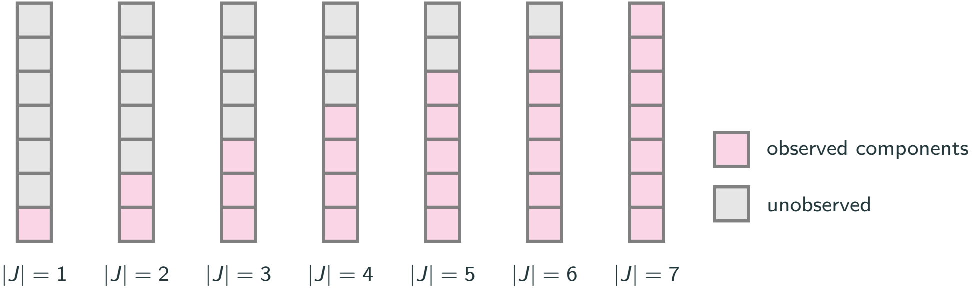
that minimises the regret

Use Sequential Learning for DA

Sequential learning in the context of DA

Goal: Sequentially develop a strategy π that depends on previous choices and rewards to achieve the following objectives:

- Minimize regret in determining the optimal size $|J|$ for a given fixed frequency.
- Alternatively, minimize regret in finding the optimal frequency λ for a given fixed size $|J|$.

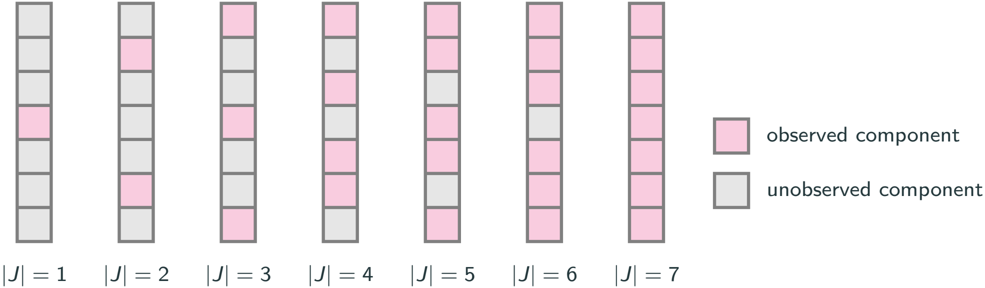


Algorithmic Ansatz: Empirically estimate the probability of "success" for the investigated sizes $|J|$ in terms of the coverage.

Sequential learning in the context of DA

Goal: Sequentially develop a strategy π that depends on previous choices and rewards to achieve the following objectives:

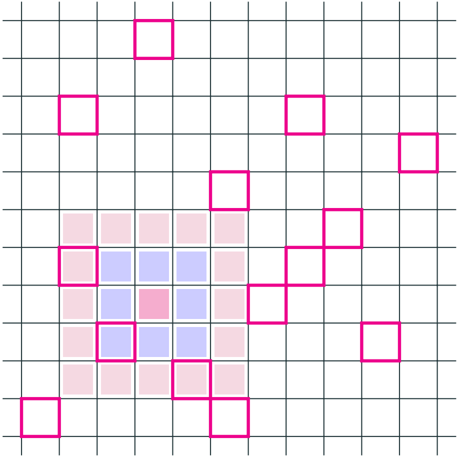
- Minimize regret in determining the optimal size $|J|$ for a given fixed frequency.
- Alternatively, minimize regret in finding the optimal frequency λ for a given fixed size $|J|$.



Algorithmic Ansatz: Empirically estimate the probability of "success" for the investigated sizes $|J|$ in terms of the coverage.

Computationally efficient approximative coverage estimation

Coverage check for each grid cell



spatially correlated w.r.t. cell

cell being checked

observed cells

Algorithm 1 Compute approximate coverage ratio κ

Require: index set J, localisation radius l, forecast-error covariance $P_t \in \mathbb{R}^{N_x \times N_x}$, correlation threshold τ_{corr}

Ensure: κ ▷ fraction of state indices "covered"

    covered $\leftarrow 0$

    for $k \leftarrow 0$ **to** $N_x - 1$ **do**

        for all $j \in J$ **do**

            $d \leftarrow \min\big(|j-k|,\ N - |j-k|\big)$

            if $d \leq l$ **then**

$$c_{kj} \leftarrow \frac{(P_t[k,j])^2}{(\mathrm{diag}(P_t)[k])^2 + 10^{-12}}$$

                if $c_{kj} \geq \tau_{\mathrm{corr}}$ **then**

                    covered $\leftarrow$ *covered* $+ 1$

                    break

                end if

            end if

        end for

    end for

    $\kappa \leftarrow$ *covered* $/ N_x = 0$

Choice of reward structure

Ansatz: the estimated reward is

$$\mathcal{R}_{a_\tau} = \underbrace{\beta\,\kappa}_{coverage} - \underbrace{\alpha \frac{N_J}{N_x}}_{\text{penalty term preventing too large } N_J} - \underbrace{\gamma \cdot \frac{1}{N_x} \sum_{i=1}^{n} P_t[i,i]}_{\text{spread}} \quad \text{with } a_\tau = N_J$$

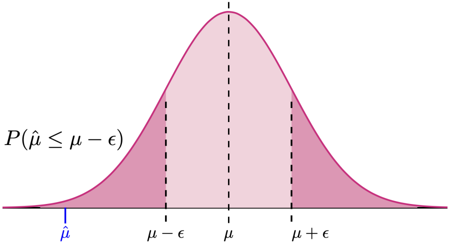
Confidence Bounds: For $X_i - \mu$ be independent and σ-subgaussian for all i:

$$\mathbb{P}(\hat{\mu} \geq \mu + \epsilon) \leq \underbrace{\exp\left(-\frac{n\epsilon^2}{2\sigma^2}\right)}_{\delta}$$

for any $\epsilon \geq 0$, i.e.,

$$\hat{\mu} - \underbrace{\sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}}}_{\epsilon} \leq \mu \leq \hat{\mu} + \underbrace{\sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}}}_{\epsilon}$$

with probability at least $1 - \delta$



UCB

Algorithm 2 UCB(δ)

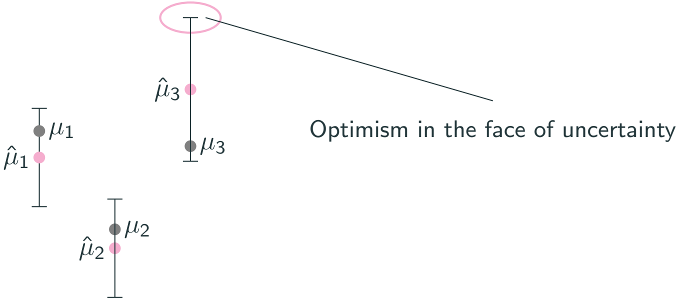
Initialization: Play each machine once;

for $t = 1, 2, 3....$ **do**

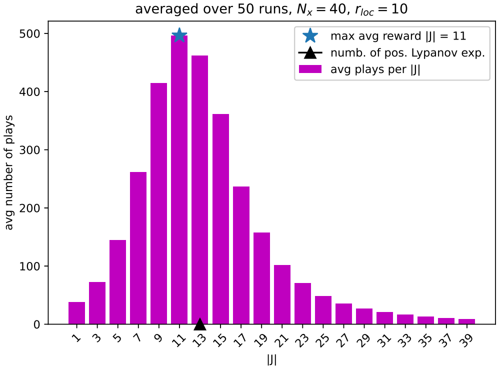
Perform action $a_{t+1} = \arg\max_{a \in \mathcal{A}} \hat{\mu}_a(t) + \sqrt{\frac{2\log(1/\delta)}{N_t(a)}}$

Update $\hat{\mu}_{a+1}(t+1)$ and $N_{t+1}(a+1)$

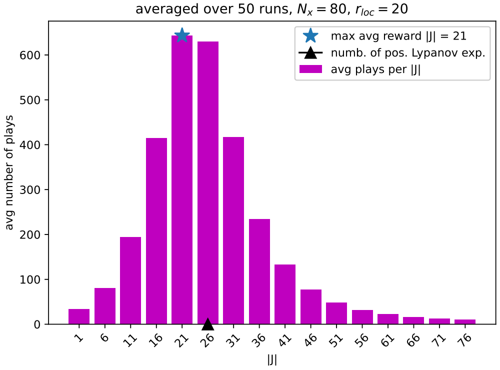
end for$=0$



Sequential learning for Lorenz 96 model



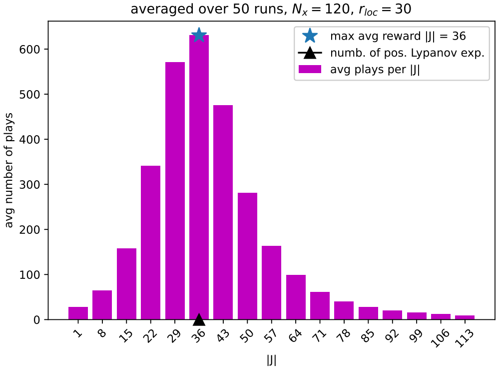
(a) $N_x = 40$



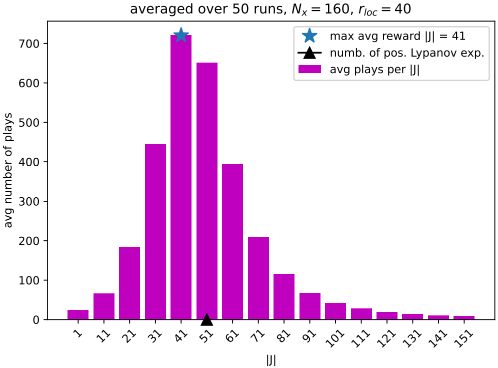
(b) $N_x = 80$

Figure 1: Plots display number of times different number of $\|J\|$ being used in the filter for dimensions $N_x \in \{40, 80\}$, the corresponding number of positive Lyapunov exponents is displayed too.

Sequential learning for Lorenz 96 model



(a) $N_x = 120$



(b) $N_x = 160$

Figure 2: Plots display number of times different number of $\|J\|$ being used in the filter for dimensions $N_x \in \{120, 160\}$, the corresponding number of positive Lyapunov exponents is displayed too.

Outlook and challenges

Next steps:

- different rewards choices
- find an appropriate sequential learning algorithm that takes dependencies into account
- find associated regret bounds
- choose a Bayesian sequential learning ansatz

Challenges:

- highly susceptible to hyper parameter
- in practice the regret has to be estimated - not enough samples computational complexity
- spatial and temporal dependencies
- learn areas of the state space with N_J

Thank you!

Jana de Wiljes
TU Ilmenau
jana.de-wiljes@tu-ilmenau.de
www.dewiljeslab.de

This research has been partially funded by the Deutsche Forschungsgemeinschaft (DFG) Project-ID 318763901-SFB1294.

Localised Ensemble Kalman Bucy Filter (l-EnKBF)

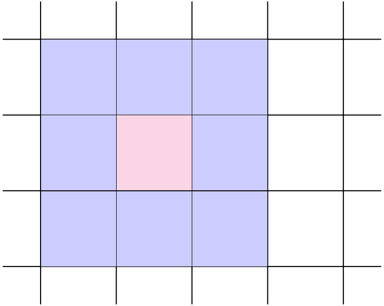
Covariance adjustment

$$[P_t^L]_{i,j} := [P_t]_{i,j}\phi_{i,j}.$$

$$\phi_{i,j} = \rho(\frac{d(i,j)}{l})$$

where ρ is an appropriate Gaspari Cohn function.

Spatial grid



Localised EnKBF

$$dX_t^i = f(X_t^i)dt + \sigma^2 P_t^\dagger (X_t^i - \overline{X}_t)dt$$
$$- \frac{1}{2} P_t^L H^T (RR^T)^{-1}(HX_t^i dt + H\overline{X}_t dt - 2dY_t)$$

where $[P_t^\dagger]_{i,i} = [P_t]_{i,i}^{-1}, \quad [P_t^\dagger]_{i,j} = 0, \quad \forall i,j = 1,\dots,n,\ i \neq j$

Assumptions

- **Short range interaction:** There is a sequence of Lipschitz constants $\mathcal{F}_k$, such that for any $X = [x_1, \ldots, x_{N_x}]$ and $X' = [x'_1, \ldots, x'_{N_x}]$, the following holds

$$|f_i(X) - f_i(X')| \leq \sum_{j=1}^{N_x} \mathcal{F}_{d(i,j)} |x_j - x'_j|.$$

- **Diagonally dominant:** there is a $q < 1$ such that

$$\sum_{j \neq i} \phi_{i,j} \leq q.$$

  Moreover, the interaction between components can be dominated by a constant $C_{\mathcal{F}}$ multiple of the matrix structure ϕ:

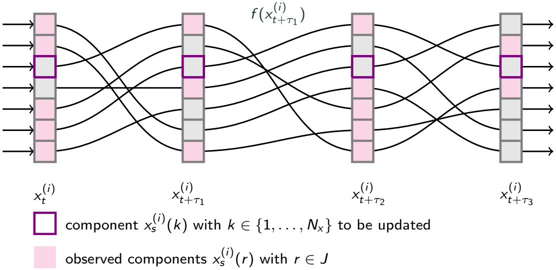
$$\mathcal{F}_{d(i,j)} \leq C_{\mathcal{F}} \phi_{i,j} \quad \forall i, j.$$

Randomised partial observations

Let $\{\tau_k\}_{k\geq 1}$ be the jump times of a Poisson process with constant intensity $\lambda > 0$ and the corresponding uniformly random draws of sets $J_k := J_{\tau_k} \subset \{1, \dots, N_x\}$. Each particle obeys the continuous-time l-EnKBF driven by the current mask, i.e.,

$$dX_t^i = f\left(X_t^i\right)dt + \sigma^2 P_t^{\dagger}\left(X_t^i - \bar{X}_t\right)dt - \tfrac{1}{2} P_t^L H_{J_k}^{\top}(RR^{\top})^{-1}\left(H_{J_k} X_t^i dt + H_{J(t)} \bar{X}_t dt - 2dY_t\right)$$

and $J(t) = J_k \quad$ for $t \in [\tau_k, \tau_{k+1})$



N. Abedini , J. de Wiljes, and S. Dubinkina (2025). *Filtering with Randomised Observations: Sequential Learning of Relevant Subspace Properties and Accuracy Analysis*

arXiv:2509.04867