

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)



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...



SEVERE Geomagnetic Storm ALERT

Updated
2024-05-10
2:00 p.m. EDT

WHAT: Geomagnetic Storm has strengthened and reached G4 conditions

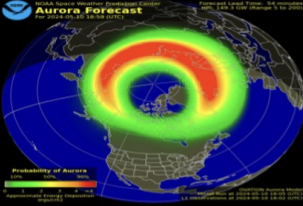
G4

KEY MESSAGES

What is a severe geomagnetic storm?
A major disturbance in Earth's magnetic field; often varying intensity between lower levels and severe storm conditions over the course of the event.

What you should do?
The Public should stay properly informed of storm progression by visiting our webpage.

Possible Technology Effects
Infrastructure operators have been notified to take action to mitigate any possible impacts. Possible increased and more frequent voltage control problems - normally mitigable. Increased possibility of anomalies or effects to satellite operations. More frequent and longer periods of GPS degradation possible.



NOAA Space Weather Prediction Center
Aurora Forecast
For 2024-05-10 18:00 (UTC)
Forecast Lead Time: 54 minutes
Min: 1.4h, 3.0h (Range 5 to 200h)

Probability of Aurora
10% 20% 50% 80% 100%

1-4 5-8 9-12 13-16 17-20 21-24 25-28 29-32 33-36 37-40 41-44 45-48 49-52 53-56 57-60 61-64 65-68 69-72 73-76 77-80 81-84 85-88 89-92 93-96 97-100 101-104 105-108 109-112 113-116 117-120 121-124 125-128 129-132 133-136 137-140 141-144 145-148 149-152 153-156 157-160 161-164 165-168 169-172 173-176 177-180 181-184 185-188 189-192 193-196 197-200 201-204 205-208 209-212 213-216 217-220 221-224 225-228 229-232 233-236 237-240 241-244 245-248 249-252 253-256 257-260 261-264 265-268 269-272 273-276 277-280 281-284 285-288 289-292 293-296 297-300 301-304 305-308 309-312 313-316 317-320 321-324 325-328 329-332 333-336 337-340 341-344 345-348 349-352 353-356 357-360 361-364 365-368 369-372 373-376 377-380 381-384 385-388 389-392 393-396 397-400 401-404 405-408 409-412 413-416 417-420 421-424 425-428 429-432 433-436 437-440 441-444 445-448 449-452 453-456 457-460 461-464 465-468 469-472 473-476 477-480 481-484 485-488 489-492 493-496 497-500 501-504 505-508 509-512 513-516 517-520 521-524 525-528 529-532 533-536 537-540 541-544 545-548 549-552 553-556 557-560 561-564 565-568 569-572 573-576 577-580 581-584 585-588 589-592 593-596 597-600 601-604 605-608 609-612 613-616 617-620 621-624 625-628 629-632 633-636 637-640 641-644 645-648 649-652 653-656 657-660 661-664 665-668 669-672 673-676 677-680 681-684 685-688 689-692 693-696 697-700 701-704 705-708 709-712 713-716 717-720 721-724 725-728 729-732 733-736 737-740 741-744 745-748 749-752 753-756 757-760 761-764 765-768 769-772 773-776 777-780 781-784 785-788 789-792 793-796 797-800 801-804 805-808 809-812 813-816 817-820 821-824 825-828 829-832 833-836 837-840 841-844 845-848 849-852 853-856 857-860 861-864 865-868 869-872 873-876 877-880 881-884 885-888 889-892 893-896 897-900 901-904 905-908 909-912 913-916 917-920 921-924 925-928 929-932 933-936 937-940 941-944 945-948 949-952 953-956 957-960 961-964 965-968 969-972 973-976 977-980 981-984 985-988 989-992 993-996 997-1000

NOAA Space Weather Prediction Center
U.S. Department of Commerce

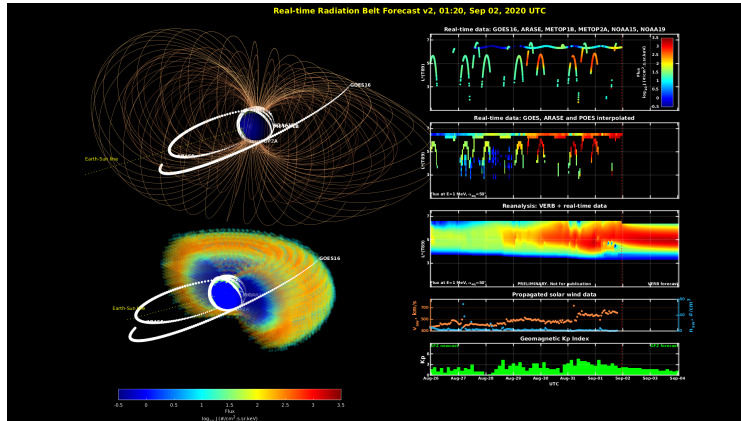


National Oceanic and
Atmospheric Administration
U.S. Department of Commerce

Safeguarding Society with Actionable Space Weather Information

Space Weather Prediction Center;
Boulder, CO

Space weather: estimating the effects of solar storms



A. M. Castillo Tibocho, 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

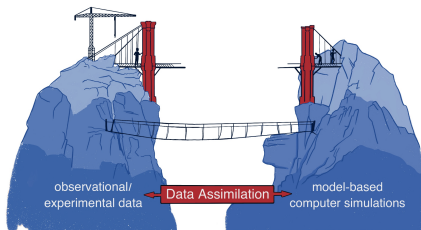
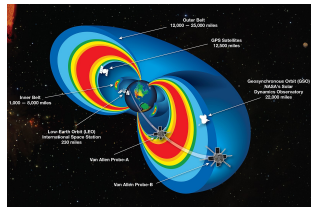
Model

$$\partial_t f_\alpha + v_\alpha \cdot \nabla f_\alpha + Z_\alpha e \left(E + \frac{v_\alpha}{c} \times B \right) \cdot \partial_v f_\alpha = 0$$

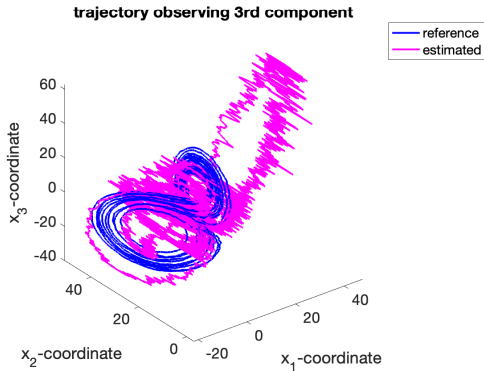
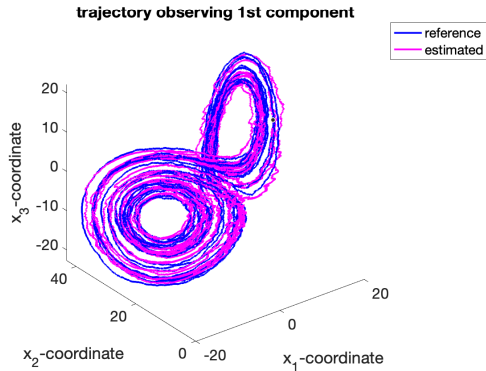
$$\nabla \cdot E - 4\pi e \int \sum Z_\alpha f_\alpha dv = 0$$

Vlasov-Poisson equation $\nabla \times E = 0$ 1

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:

$$(Model) \quad dX_t = f(X_t, \lambda)dt + \sqrt{2}C dW_t$$

$$(Observations) \quad dY_t = h(X_t)dt + R^{1/2}dV_t$$

Filtering Problem: Estimate filtering distribution

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

Kushner-Stratonovich equation:

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

where

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

Linear Model: Kalman- Bucy Filter

$$d\bar{x}_t = A\bar{x}_t dt + bdt - P_t H^T R^{-1} (H\bar{x}_t dt - dY_t)$$

Ensemble Kalman Bucy Filter (EnKBF)

Deterministic EnKBF:

$$dX_t^i = f(X_t^i)dt + CdW_t^i - \frac{1}{2}P_t^M H^T R^{-1} \left(h(X_t^i)dt + \bar{h}_t^M dt - 2dY_t \right)$$

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

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

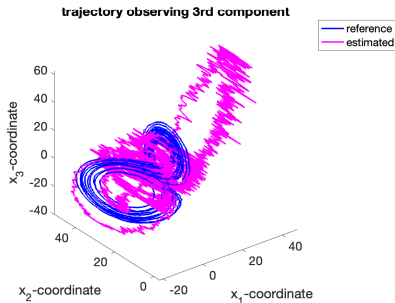
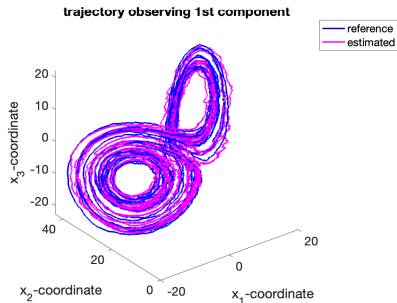
Interacting particle representation of the model error:

$$dX_t^i = f(X_t^i)dt + CC^T(P_t^M)^{-1}(X_t^i - \bar{x}_t^M)dt \\ - \frac{1}{2}P_t^M H^T R^{-1} \left(h(X_t^i)dt + \bar{h}_t^M dt - 2dY_t \right)$$

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



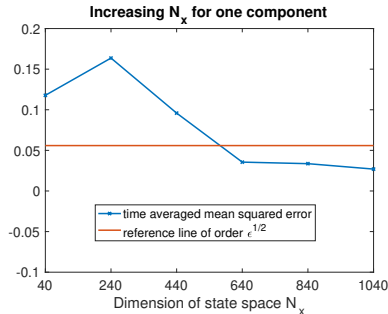
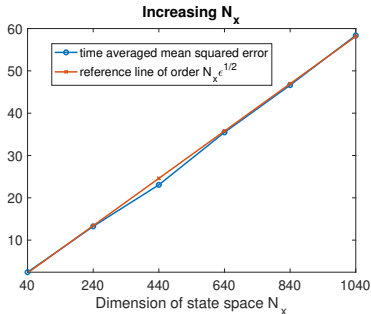
Proposition: Under the assumption of *short range interactions* and $dX_t = f(X_t)dt + \sqrt{2}dW_t$ and $dY_t = X_t dt + R^{1/2}dB_t$ with $R = \Theta(\epsilon)$ we obtain bounds:

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

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:

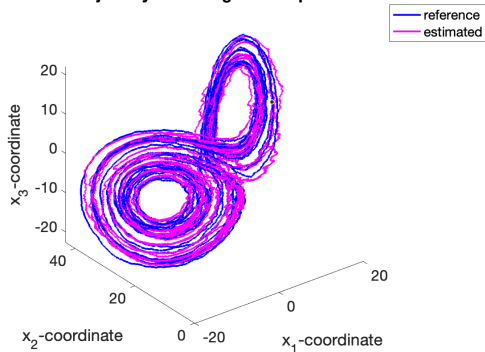
1. $\mathbb{E}[e_i^2(t)] \leq C\sqrt{\epsilon} \quad \forall i$ (**Expected error for fixed t**)
- 2.

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

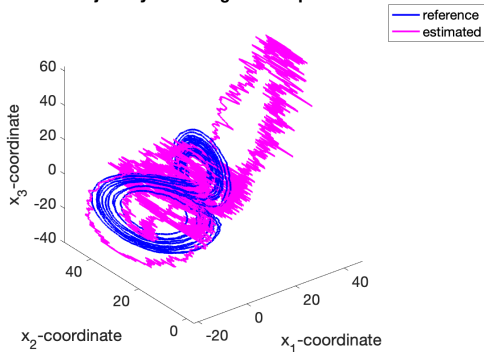
Let's be a bit more realistic

Partially observed setting

trajectory observing 1st component



trajectory observing 3rd component

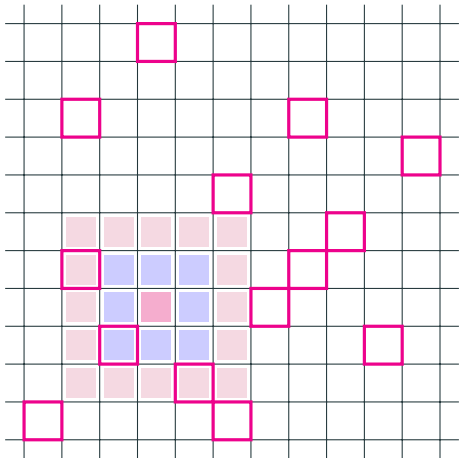


Ansatz to combat challenge:

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

Fixed partial observations

Neighboring observations



Need bound for:

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

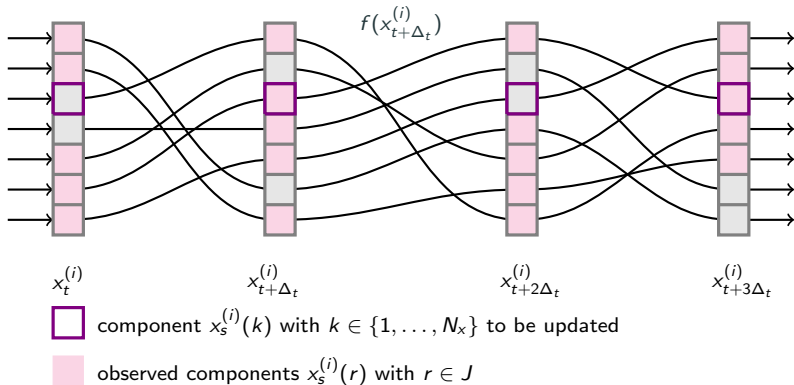
Assume: have an observation at **location** j . Considering component k ($k \neq j$) yields:

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

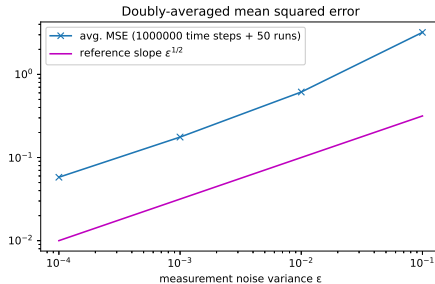
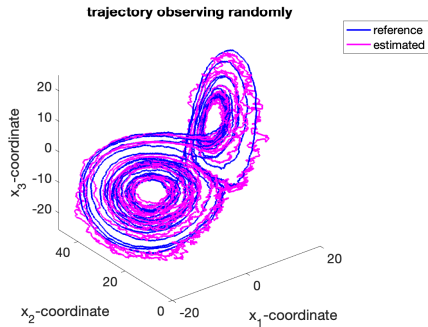
$$\mathbb{E}[e_k(t)^2] \leq \begin{cases} c_j \sqrt{\epsilon}, & c_{j,k} = \Theta(1), \\ C, & c_{j,k} = \Theta(\epsilon). \end{cases} \quad (2)$$

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

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

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

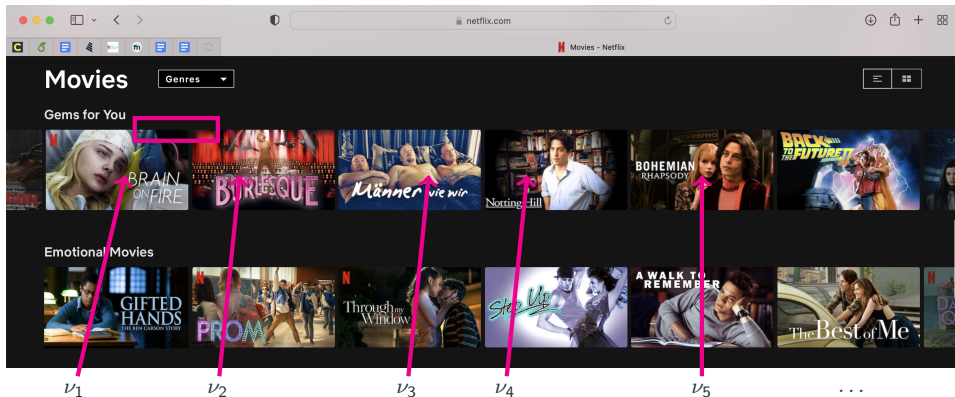


a_3



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

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

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

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

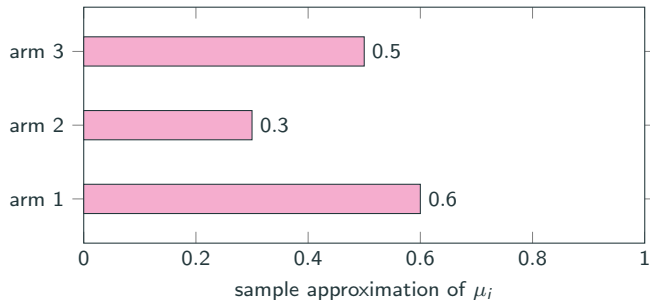
Bernoulli example

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 $a_1, \dots, a_T$ is the random variable

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

Goal: find a sequential sampling strategy

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

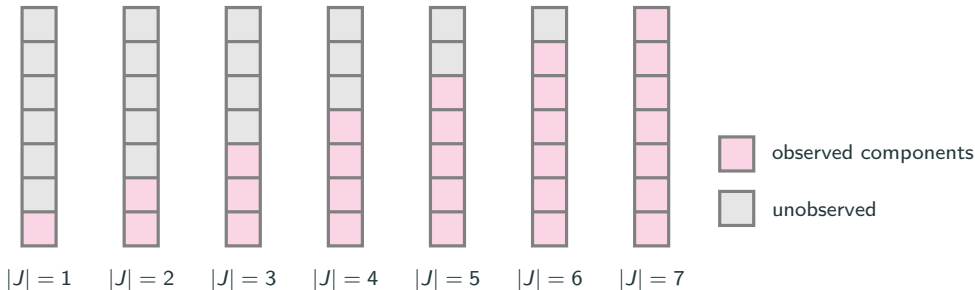
that minimises the regret

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 $|J|$.

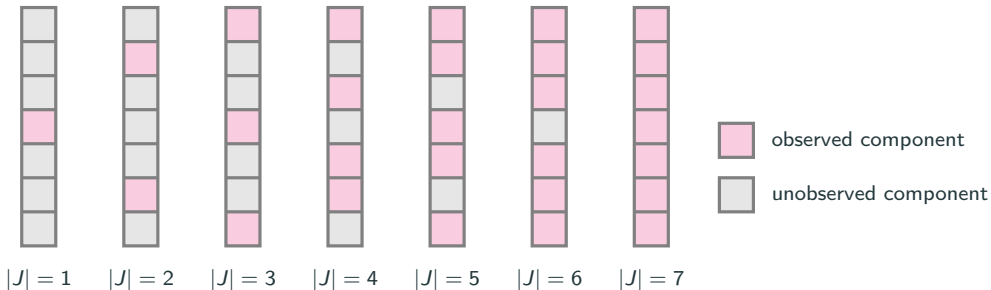


Algorithmic Ansatz: Empirically estimate the probability of "success" for the investigated sizes $|J|$ in terms of the coverage.

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:

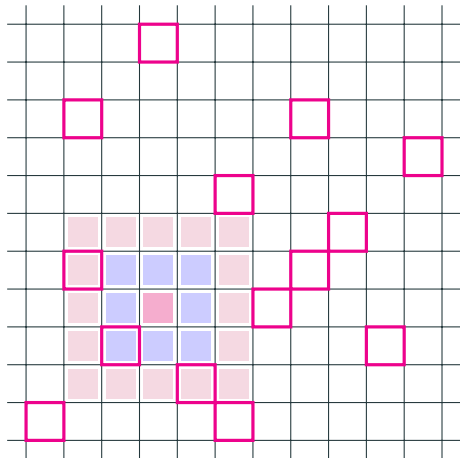
- Minimize regret in determining the optimal size $|J|$ for a given fixed frequency.
- Alternatively, minimize regret in finding the optimal frequency λ for a given fixed size $|J|$.



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 $\triangleright$ fraction of state indices "covered"

```

covered  $\leftarrow$  0
for  $k \leftarrow 0$  to  $N_x - 1$  do
  for all  $j \in J$  do
     $d \leftarrow \min(|j - k|, N - |j - k|)$ 
    if  $d \leq l$  then
       $c_{kj} \leftarrow \frac{(P_t[k, j])^2}{(\text{diag}(P_t)[k])^2 + 10^{-12}}$ 
      if  $c_{kj} \geq \tau_{\text{corr}}$  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

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

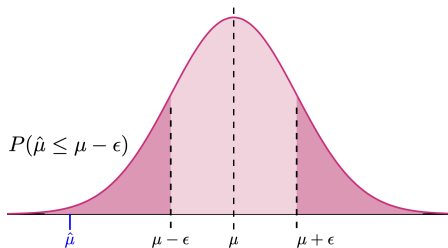
Confidence Bounds: For $X_i - \mu$ be independent and σ -subgaussian for all i :

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

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 $1 - \delta$



Algorithm 2 UCB(δ)

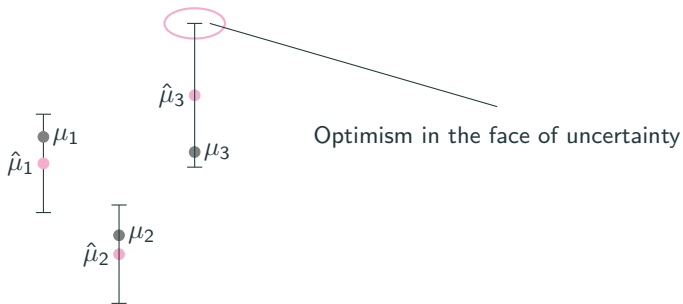
Initialization: Play each machine once;

for $t = 1, 2, 3, \dots$ **do**

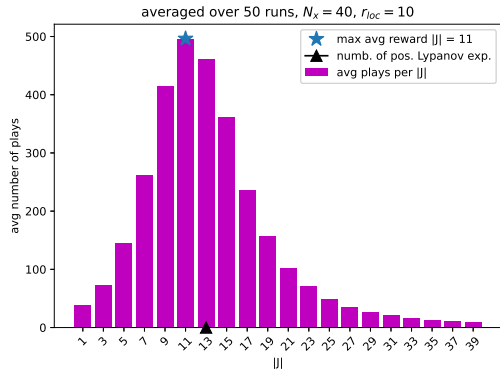
 Perform action $a_{t+1} = \arg \max_{a \in \mathcal{A}} \hat{\mu}_a(t) + \sqrt{\frac{2 \log(1/\delta)}{N_t(a)}}$

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

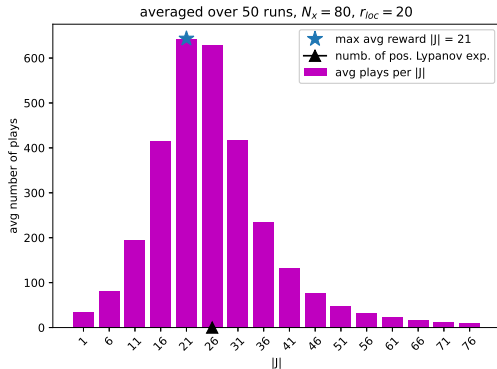
end for = 0



Sequential learning for Lorenz 96 model



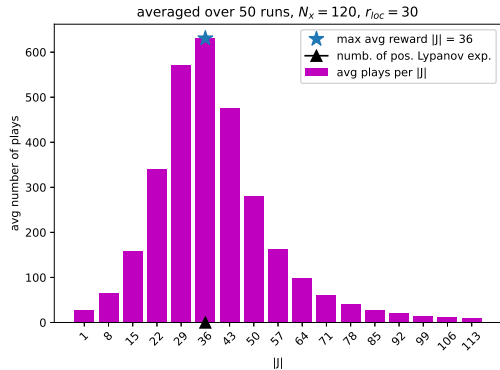
(a) $N_x = 40$



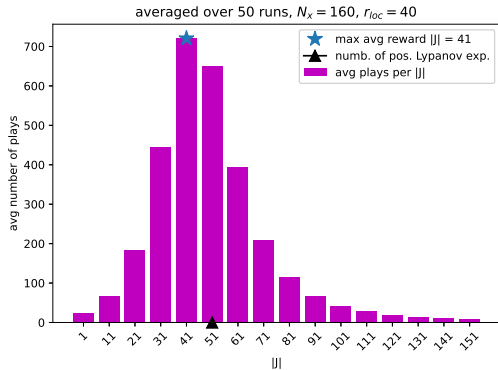
(b) $N_x = 80$

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.

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 (I-EnKBF)

Covariance adjustment

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

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

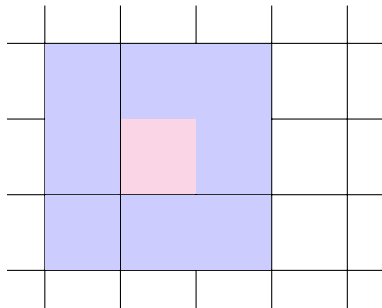
where ρ is an appropriate
Gaspari Cohn function.

Localised EnKBF

$$\begin{aligned} dX_t^i &= f(X_t^i)dt + \sigma^2 P_t^\dagger (X_t^i - \bar{X}_t)dt \\ &\quad - \frac{1}{2} P_t^L H^T (R R^T)^{-1} (H X_t^i dt + H \bar{X}_t dt - 2dY_t) \end{aligned}$$

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

Spatial grid



Assumptions

- **Short range interaction:** There is a sequence of Lipschitz constants $\mathcal{F}_k$, such that for any $X = [x_1, \dots, x_{N_x}]$ and $X' = [x'_1, \dots, x'_{N_x}]$, 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 I-EnKBF driven by the current mask, i.e.,

$$dX_t^i = f(X_t^i)dt + \sigma^2 P_t^\dagger (X_t^i - \bar{X}_t)dt - \frac{1}{2} P_t^L H_{J_k}^\top (RR^\top)^{-1} (H_{J_k} X_t^i dt + \textcolor{red}{H}_{J(t)} \bar{X}_t dt - 2dY_t)$$

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

