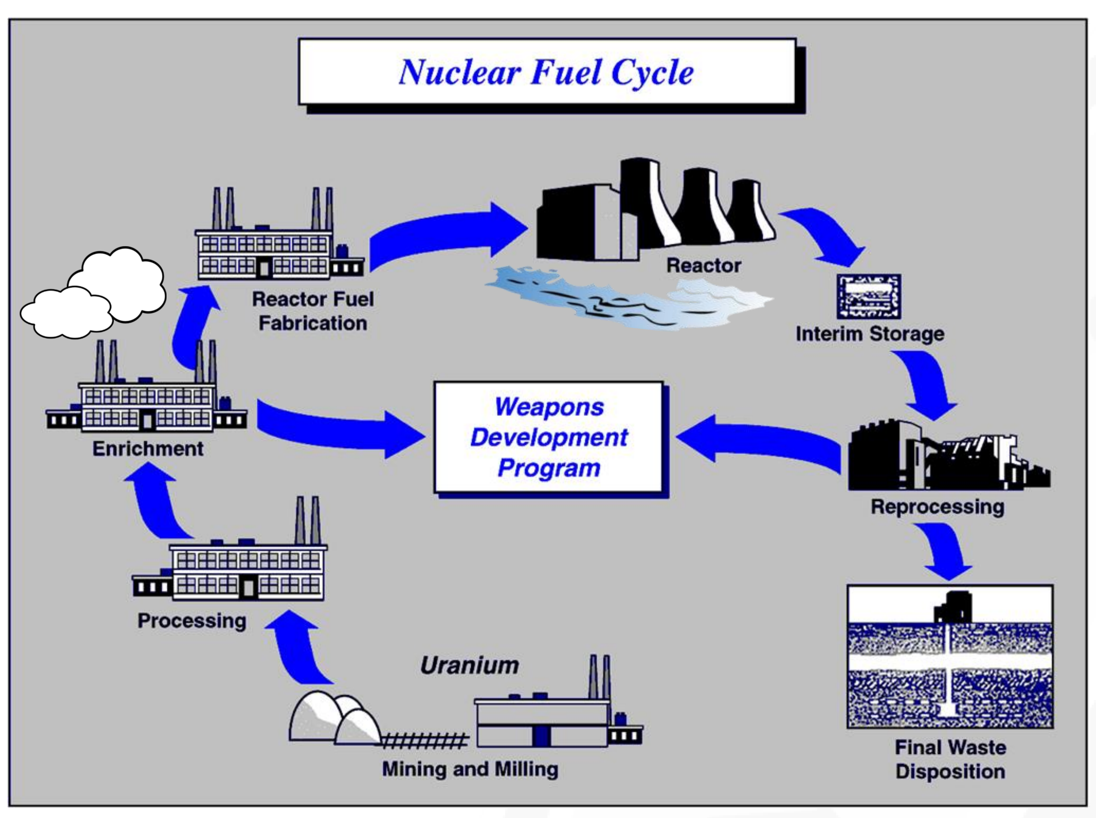
Penalized Ensemble Kalman Filters for High Dimensional Non-linear Systems

Elizabeth Hou[1], Earl Lawrence[2], Alfred O. Hero[1]
[1]University of Michigan, Ann Arbor and [2]Los Alamos National Laboratory
Alfred O. Hero, hero@eecs.umich.edu
Consortium for Verification Technology (CVT)

Motivation

- Nuclear fuel cycles release emissions into the surrounding environment
  - Heat, water vapor, CO2, SO2, Nox
- Declared and undeclared nuclear activities have different emission patterns
- Potential diversions could be:
  - Excessive emissions in the environment
  - Unusual chemicals detected by sensors
- Modeling the environment during declared activities makes it possible to detect unusual activity



System Model

- Environment is modeled as a spatial-temporal random field, governed by actual physical equations
- Sensors collect observations, these are noisy discrete data points generated by the environment

physical eqs. system noise

random variables

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \boldsymbol{\omega}_t$$

$$\mathbf{y}_t = h(\mathbf{x}_t) + \boldsymbol{\epsilon}_t$$

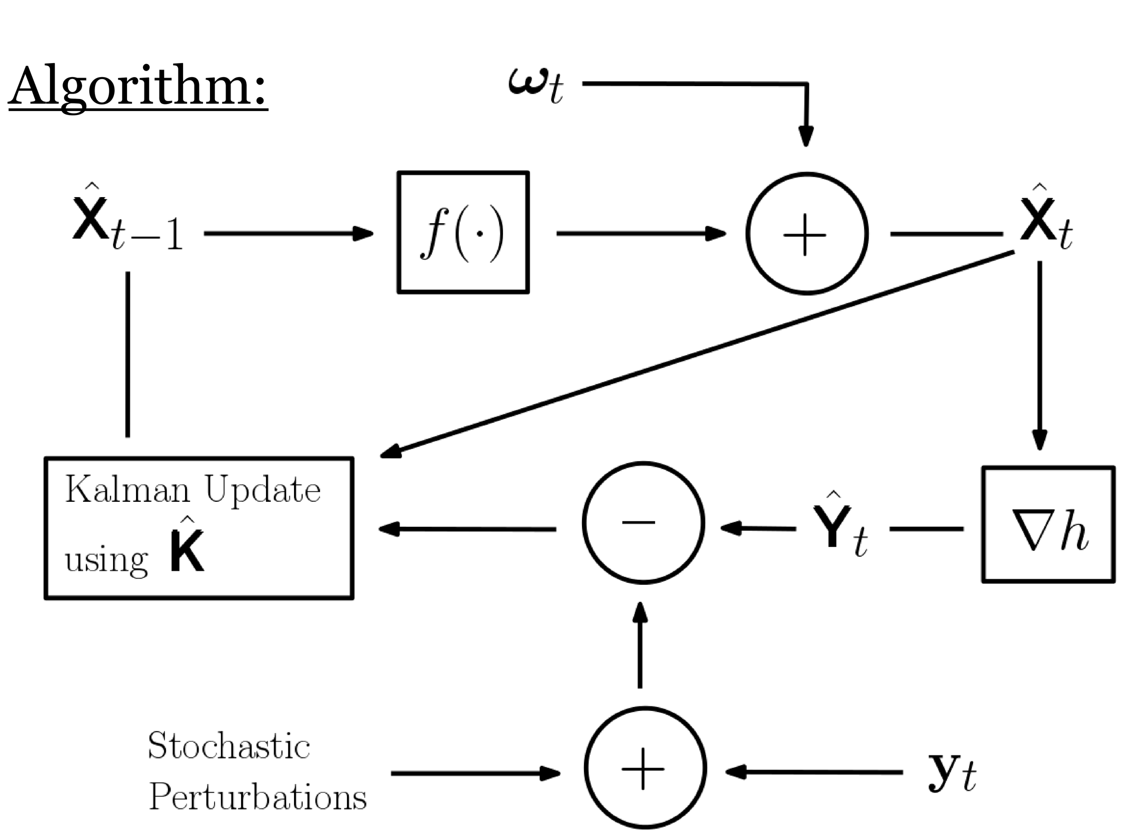
observations observation noise

distortion func.

Estimating the Environment

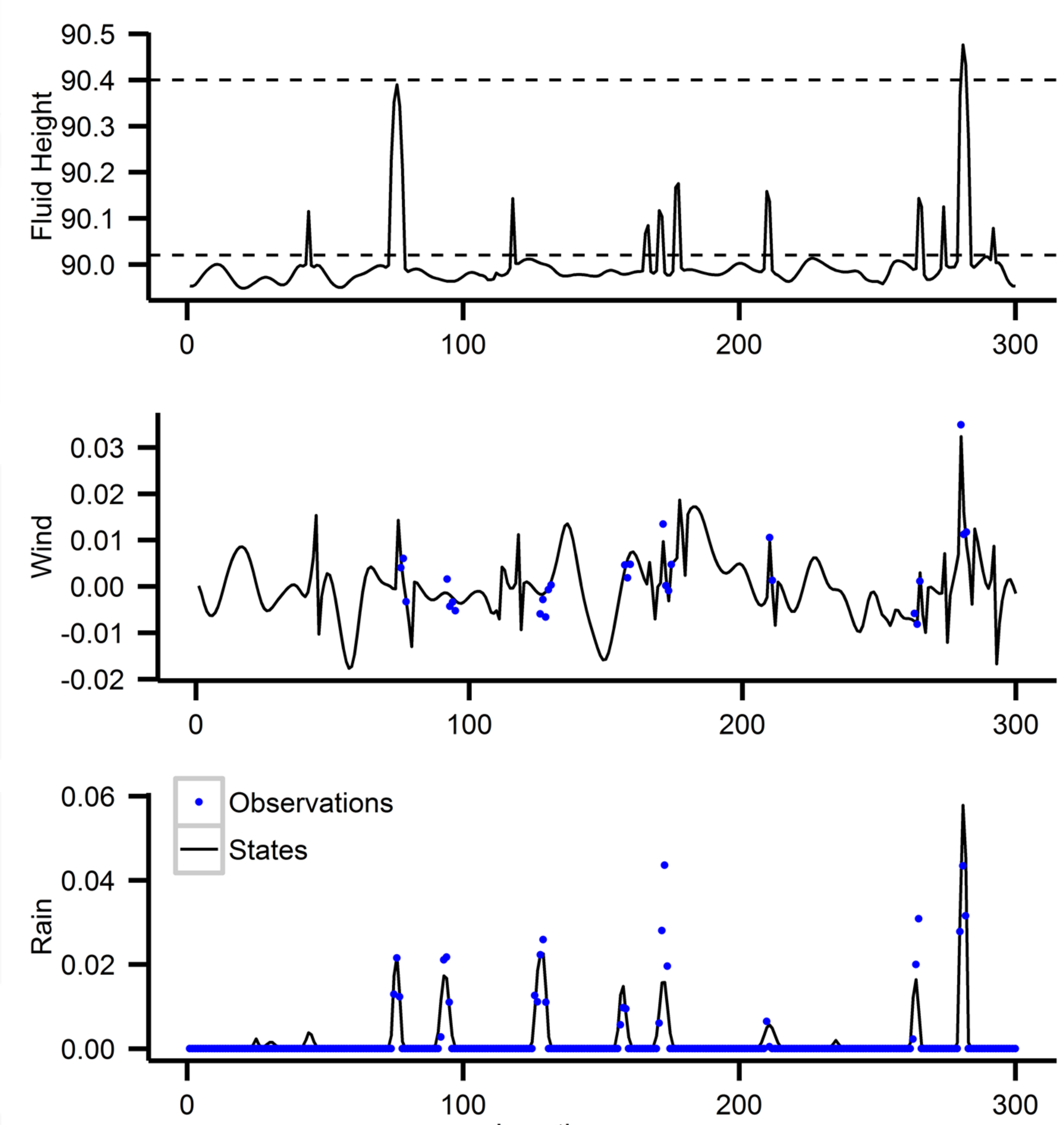
- When $f(\,)$ is very complicated, can't calculate $\nabla f(\,)$
  - So can't use (extended) Kalman filter
  - Must estimate the Kalman gain matrix $\mathbf{K}$ another way
- Ensemble Kalman filter [1] creates an ensemble of $\mathbf{x}_t$'s and uses this to calculate an estimator for $\mathbf{K}$
- Computationally expensive to create large ensemble
- So when $\mathbf{x}_t$ is large, the estimator $\hat{\mathbf{K}}$ is not very good
  - Many methods apply a taper matrix (using additional knowledge) to sample cov. in $\hat{\mathbf{K}}$
  - Our method (PEnKF) *learns* a better $\hat{\mathbf{K}}$
- The penalized ensemble Kalman filter [2] uses ℓ_1 penalty to promote sparsity in inverse sample cov.
  - Learns the interactions between variables
  - Is proven to require a smaller ensemble $\hat{\mathbf{X}}$ than standard EnKF

Algorithm:



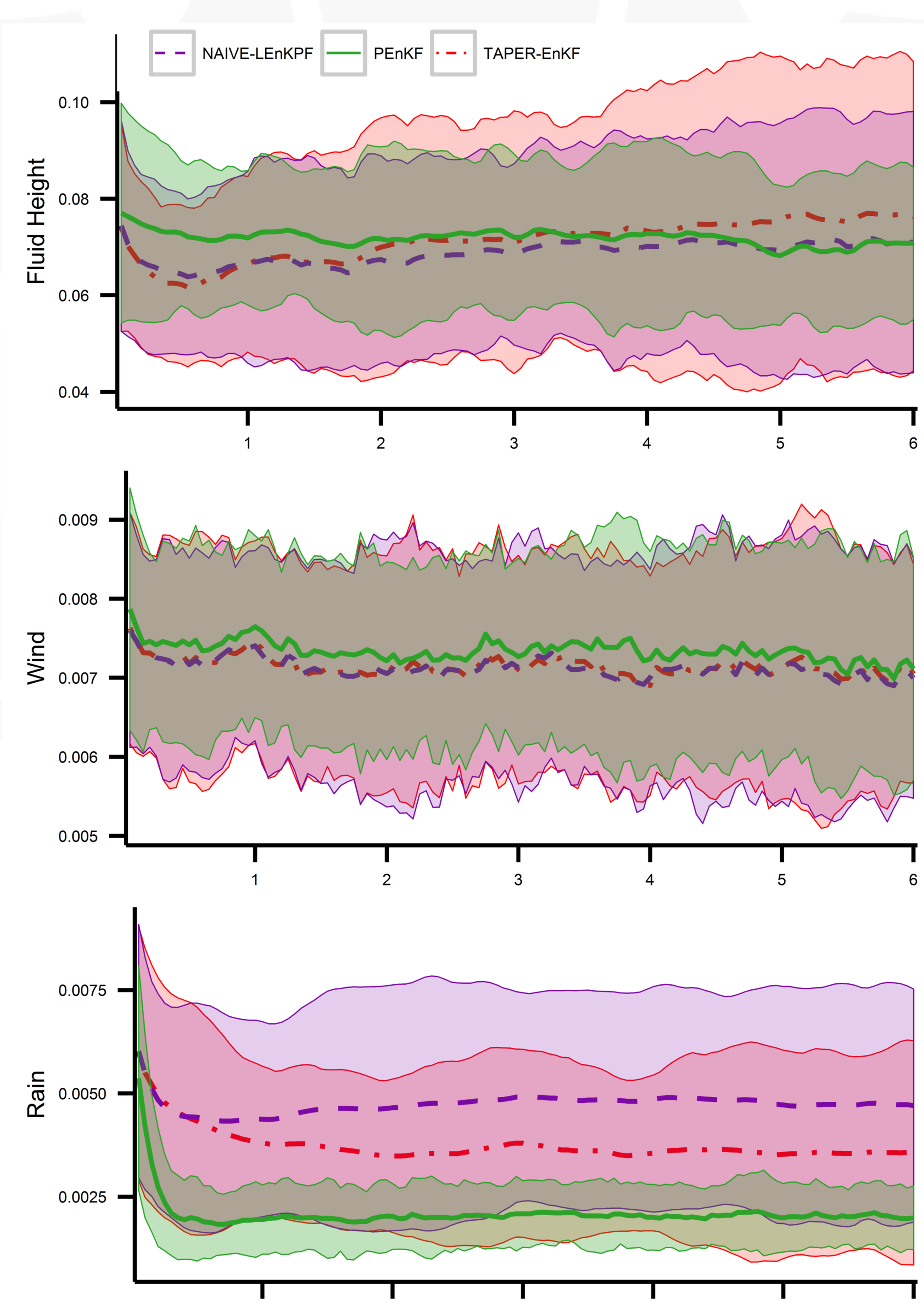
Cloud Convection System

- A system based on the modified shallow water equations of [3]
  - Models cloud convection with fluid dynamics
  - 3 Types of state variables: fluid height, rain content, wind speed
  - 300 locations for each type of state variable
- Observations every 5 seconds
  - Always observe rain content
  - Only observe wind speed where it is raining
  - Never observe fluid height



Root Mean Squared Error

- RMSE: Averaged over all 300 locations of each type
- Both NAIVE-LEnKPF [4] and TAPER-EnKF use *a prior* information about the true system
- PEnKF does equally well without this information
  - For rain, it does statistically better

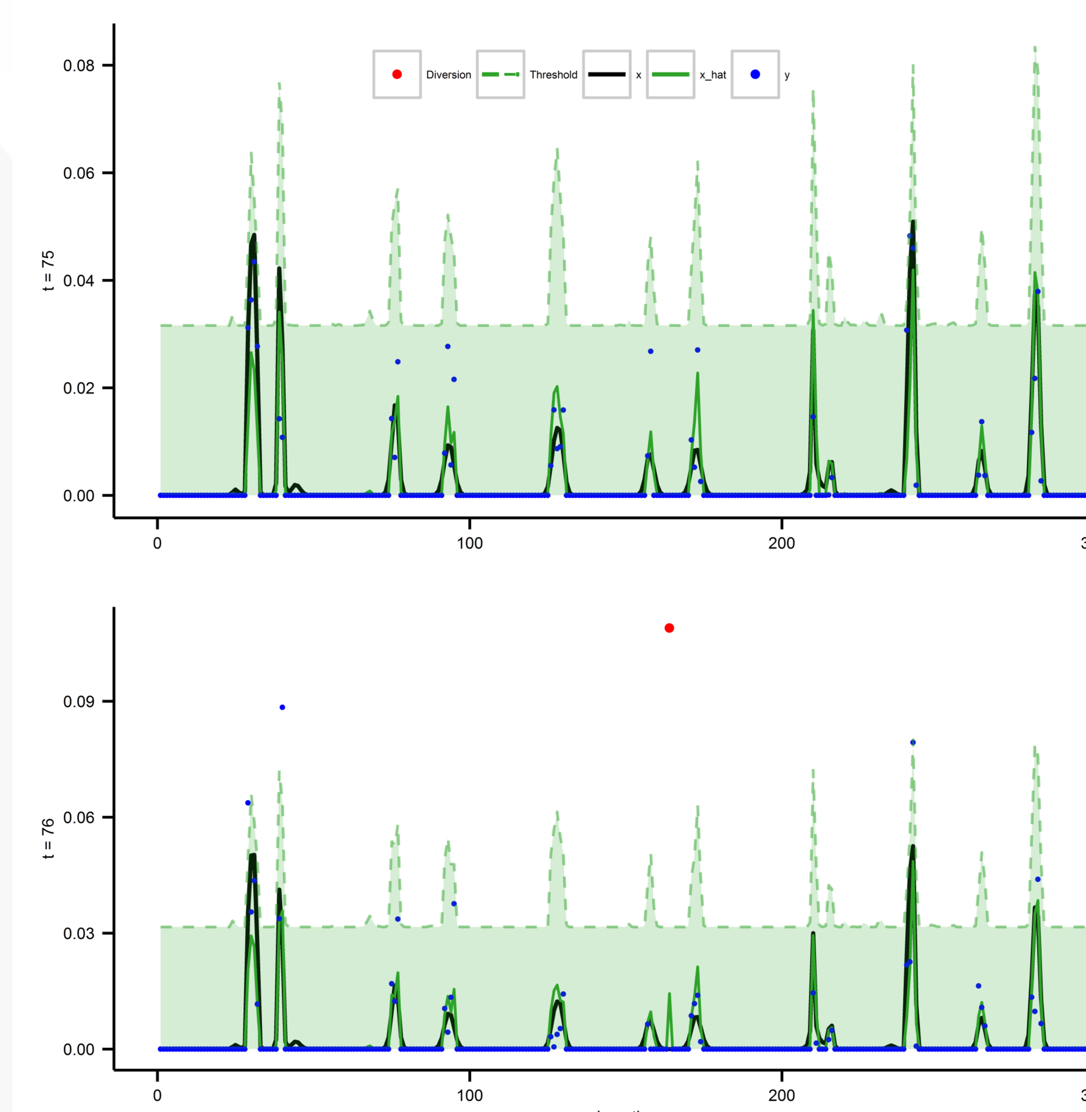


Diversion Detection

- Train the PEnKF for period of time without diversions (establish a baseline)
- Then for every new time point t,
  1) Calculate the threshold for false positive level α at every location i
     - $C^i(\alpha) = \frac{1}{n}\sum_{j=1}^{n} \hat{X}_t^{ij} + \Phi(\alpha)\sqrt{\tilde{P}_{ii}}$
  2) Test $y_t^i > C^i(\alpha)$
     - If true, declare an diversion at location i
     - Bypass the Kalman update at location i

Simulating Diversions

- Generate diversions as observations of rain content at locations where there shouldn't be (no clouds)
  - Can appear anytime after 3 hours (t = 60)
  - Last between 25 - 50 seconds
  - Have values between 0.1 - 0.15



- At false positive level α = 0.001. the threshold:
  - Correctly identifies the diversion at t = 75
  - Has 1 false positive (out of 299) at this t

Discussion

- Propose the PEnKF to estimate a large, complicated environment by incorporating data from many sources
- Propose a simple threshold to test for diversions and prevent them from corrupting the estimation
- Working on tighter test for more accurate detection with respect to a given false positive rate

References

[1] Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99 (C5), 10 143–10 162.
[2] Hou, E., E. Lawrence, and A. O. Hero, 2016: Penalized Ensemble Kalman Filters for High Dimensional Non-linear Systems. ArXiv e-prints, 1610.00195.
[3] Würsch, M., and G. C. Craig, 2014: A simple dynamical model of cumulus convection for data assimilation research. *Meteorologische Zeitschrift*, 23 (5), 483–490.
[4] Robert, S., and H. R. Künsch, 2016: Local Ensemble Kalman Particle Filters for efficient data assimilation. ArXiv e-prints, 1605.05476.

This work was partially supported by the Consortium for Verification Technology under Department of Energy National Nuclear Security Administration award number DE-NA0002534 and by the Laboratory Directed Research and Development program at Los Alamos National Laboratory under project 20150033DR, SHIELDS: Space Hazards Induced near Earth by Large Dynamic Storms - Understanding, Modeling, and Predicting.

National Nuclear Security Administration