Model and Data Reduction Techniques for Data Assimilation

Colin Roberts

Colorado State University

Introduction

Introduction

Question: What is data assimilation?

Introduction

Question: What is data assimilation?

- "***Data assimilation*** is the technique whereby observational ***data*** are combined with output from a numerical ***model*** to produce an optimal estimate of the evolving state of the system." -Alan O'Neill

Motivation

Motivation

- Say we are given this satellite data taken over various times.
- We want to fill in the missing data given these measurements and a model.

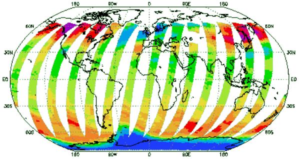


Figure: GOME, TM3-DAM data.

Motivation

- Say we are given this satellite data taken over various times.
- We want to fill in the missing data given these measurements and a model.

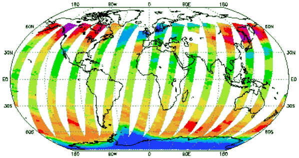


Figure: GOME, TM3-DAM data.

Motivation

- Not just an interpolation tool!
- Data assimilation is used to improve forecasts as well.

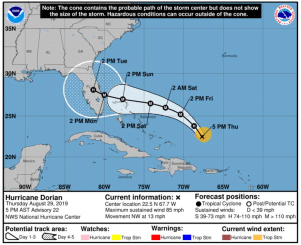


Figure: NOAA hurricane Dorian projection.

Goal

high dimensional nonlinear systems such as SWE

Related works

mention what others have tried ans why we're doing something different

Model and data

We take a temporally discretized system with $t = 0, 1, 2, \ldots, T$.

- State: $\boldsymbol{x}_t \in \mathbb{R}^M$.
- Model: $\boldsymbol{x}_{t+1} = \boldsymbol{F}_t(\boldsymbol{x}_t) + \boldsymbol{\sigma}_t$ with $\boldsymbol{\sigma}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma})$.
- Observation: $\boldsymbol{z}_t \in \mathbb{R}^D$.
- Observation operator: $\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t^{\text{truth}} + \boldsymbol{r}_t$ with $\boldsymbol{r} \sim \mathcal{N}(0, \boldsymbol{R})$ and $\boldsymbol{H}: \mathbb{R}^M \to \mathbb{R}^D$.

Particle filter

Algorithm:

- Start with L ***particles*** which are drawn around initial condition $\boldsymbol{x}_0$.
- Each particle ℓ starts with equal weighting $\boldsymbol{w}_0^\ell$
- Step each particle forward in time using $\boldsymbol{F}_0$ to get $\boldsymbol{x}_1$.
- Obtain the observation $\boldsymbol{z}_1$ and generate a prior $\mathrm{p}(\boldsymbol{z}_1|\boldsymbol{x}_1)$.
- Bayes' theorem generates the ***posterior*** $\mathrm{p}(\boldsymbol{x}_1|\boldsymbol{z}_1)$.
- ***Reweight*** by $\boldsymbol{w}_2^\ell \propto \boldsymbol{w}_1^\ell\, \mathrm{p}(\boldsymbol{z}_1|\boldsymbol{x}_1^\ell) \propto \exp[(\boldsymbol{z}_1 - \boldsymbol{H}\boldsymbol{x}_1)\boldsymbol{R}^{-1}(\boldsymbol{z}_1 - \boldsymbol{H}\boldsymbol{x}_1)]$ and require $\sum_{\ell=1}^{L} \boldsymbol{w}_1^\ell = 1$.
- Repeat for the next time steps.

Particle filter

Assimilation method: (Optimal Proposal) Particle Filter (OP-PF).

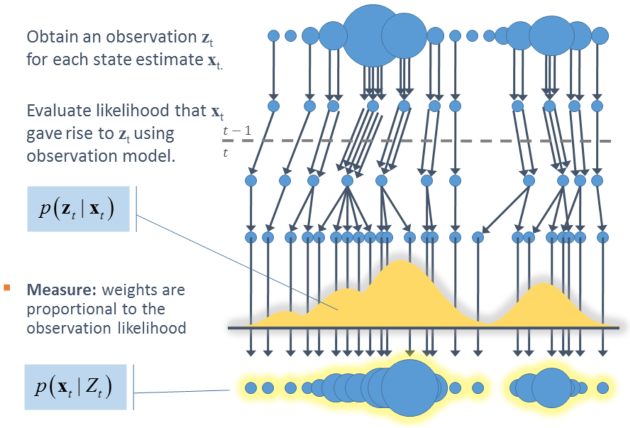
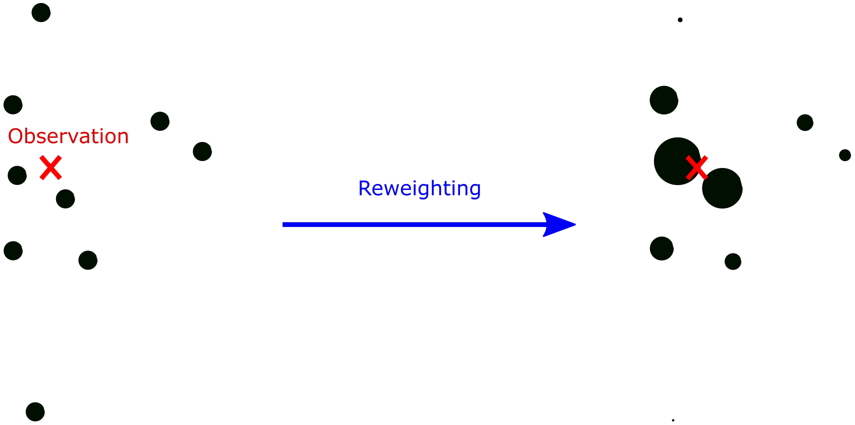


Figure: Sharath Srinivasan, towardsdatascience.com

Reweighting



Resampling

- At each time step, we also check the number of effective particles by $N_{\text{eff}} = \frac{1}{\sum_{\ell=1}^{L}(\mathbf{w}_t^\ell)^2}$.
- If N_{eff} drops below a threshold, we ***resample*** new particles with a Gaussian around the particle mean or latest observation.

Discussion

Why PF?

- PF is effective for nonlinear problems.
- PF captures non-Gaussian posteriors.

Other filtering methods can't accomplish both of these. **Why NOT PF?**

- Weight degeneration in high dimensional applications.
- Need $L \propto \exp(MD)$ particles to avoid weight degeneration.

Dimension reduction

Mitigating the curse of high dimensionality:

- Determine ***dynamically significant*** basis elements.
- These are basis elements capture large scale flow patterns and coherent structures over time.
- Project our problem onto these basis elements to reduce the dimensionality.

Reduction methods

- Assimilation in the Unstable Subspace (AUS)
  - Time dependent basis elements.
  - More computationally challenging.
- Proper Orthogonal Decomposition (POD)
  - Simple and cheap to implement.
- Dynamic Mode Decomposition (DMD)
  - More flexibility in choosing dynamically significant elements than POD.
  - Needs more knowledge of the dynamics we are capturing.

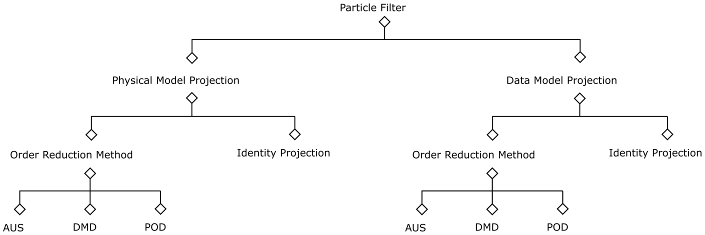
Intuition

1. Run the model and make observations up to time T.
2. Build snapshot matrix $\boldsymbol{X} = \begin{pmatrix} | & | & \cdots & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_T \\ | & | & \cdots & | \end{pmatrix}$.
3. Compute the singular value decomposition $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger$.
4. Choose the modes corresponding to the largest q singular values.

Projected data assimilation

- We can now determine $M^q < M$ and $D^q < M$ significant modes for the model and data respectively.
- We now perform DA on the dimensionally reduced problem by working solely with the coefficients of the dominant modes.
- We must balance keeping enough modes to represent the dynamics, but dropping enough to have effective PF.

Projection choices



What we're looking for

- Capture coherent structures.
- Low Root Mean Square Error (RMSE) from particle mean to truth.
- Low resampling rate.

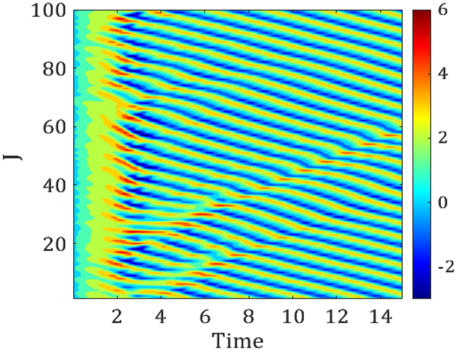
RMSE

Our method of measuring error is the RMSE

$$\text{RMSE} = \frac{(\boldsymbol{x}_t^{\text{truth}} - \mu_t)^\dagger \cdot (\boldsymbol{x}_t^{\text{truth}} - \mu_t)}{M}$$

Lorenz 96

- ODE describing meteorological quantities along lines of latitude.
- A staple comparison for other assimilation schemes.
- We ran with $M = 100$ and $F = 3.5$. Truth below.



Results

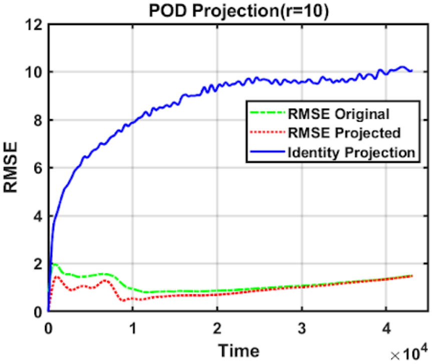


Figure: POD reduction to $M^q = 10$.

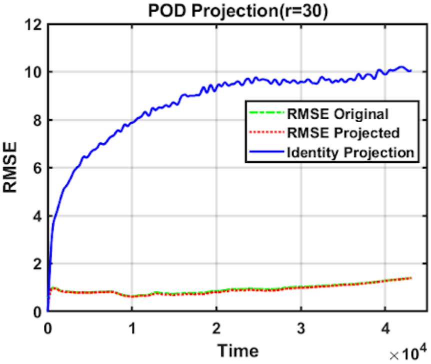


Figure: POD reduction to $M^q = 30$.

Shallow water equations

- PDE to solve for the height $h(x, y, t)$ of a column of water.
- Discretized x and y to get a grid of size $38,100$ (dimension of the problem).
- Truth was given by barotropic instability
  http://www.met.reading.ac.uk/~swrhgnrj/shallow_water_model/

For every run we have

- Total particles $L = 20$.
- Observation time is every 60 seconds
- Covariances: $\mathbf{\Sigma} = I$, $\boldsymbol{R} = 0.01I$.

Qualitative results

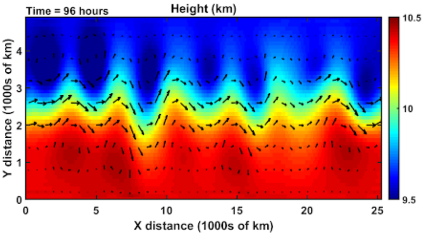


Figure: Truth run for shallow water equations.

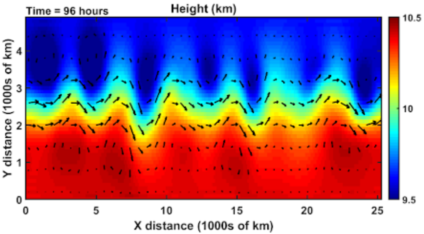


Figure: POD projection with $M^q = D^q = 100$ run for shallow water equations.

Results

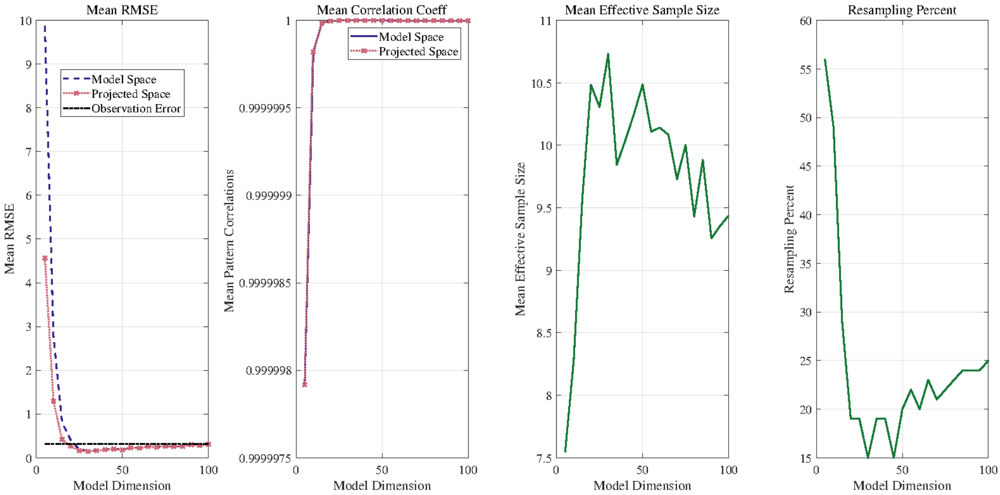


Figure: Fixed $D^q = 10$ and varied M^q. Observed every 1000th gridpoint every 60 seconds

Collaborators

- Aishah Albarakati, Clarkson University, U.S.;
- Rose Crocker, The University of Adelaide, Australia;
- Juniper Glass-Klaiber, Mount Holyoke College, U.S.;
- Sarah Iams, Harvard University, U.S.;
- Noah Marshall, University of British Colombia, Canada;
- Erik Van Vleck, University of Kansas, U.S.

Support and funding