Data-Driven Data Assimilation

Yicun Zhen, Pierre Tandeo

IMT Atlantique
Lab-STICC
UBL
Brest, France

SpatialTempMeteo Workshop
Rennes, France
Nov. 28-30, 2018

Outline

A Brief Review of Analog Forecasting Method

AnDA = Analog Forecast + Data Assimilation

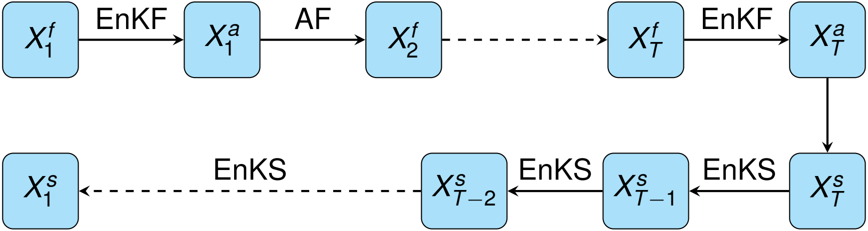
Assumption

Having a huge amount of data (analogs (A_i) + successors (S_i)) so that every corner of the attractor of the system is covered.

- Given the state estimate X_t
- Find the k (=50, for instance) analogs that are the closest to X_t : $A_1, ..., A_k$
- Linearly regress $S_1, ..., S_k$ on $A_1, ..., A_k$: S = LA+b
- Apply the linear relation on X_t: $X_{t+1} \leftarrow LX_t + b$
- Covariance inflation: $X_{t+1} \leftarrow X_{t+1} + \mathcal{N}(0, \alpha(k)C_{t+1})$ where $C_{t+1} = cov(S - (LA + b))$

Analog Data Assimilation (AnDA) for Reanalysis

- Ensemble Kalman filter(EnKF) for data assimilation;
- Analog forecast (AF) for state forecast;
- Ensemble Kalman smoother (EnKS) for calculating the state reanalysis.

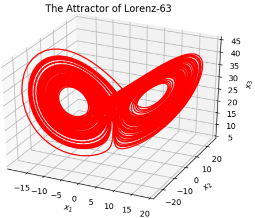


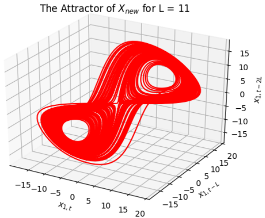
Time-Delayed Analog Forecast

Motivation

Taken's Theorem (1981): under certain conditions, a strange attractor can be resconstructed using lagged partial observations.

Construct time-delayed analogs $A_t^{new} = (A_t, A_{t-L}, ..., A_{t-kL})$ and similarly for successors and state estimates.





Experimental Design

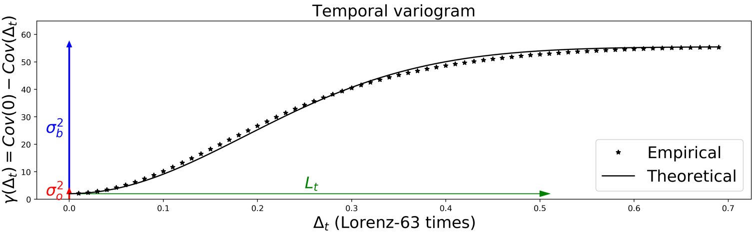
Lorenz 63

- dt $= 0.01$, $\mathrm{dt}_{obs} = 0.08$, $y_t^o = x_{1,t} + \xi_t$, $\mathrm{R}_{obs} = 2.0$
- Use time-delayed analogs $X_t^{analog} = (x_{1,t}, x_{1,t-7}, x_{1,t-14})$
- 50 ensemble members

Lorenz 96

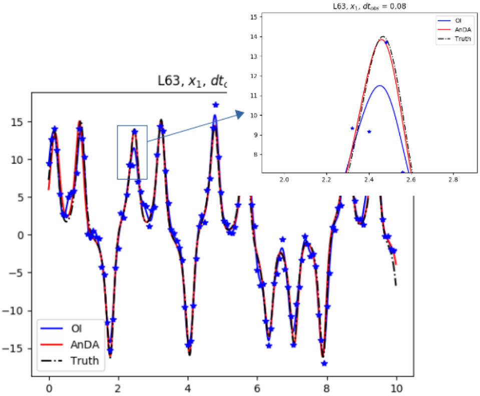
- dt = 0.05 , $\mathrm{dt}_{obs} = 0.20$, $\mathrm{R}_{obs} = 1.0$
- $X_t = (x_{1,t}, x_{3,t}, \ldots, x_{39,t})$, $y_t^o = (x_{1,t}, x_{5,t}, \ldots, x_{37,t}) + \xi_t$
- $x_{i,t}^{analog} = (x_{i-2,t}, x_{i,t}, x_{i+2,t}, x_{i-2,t-1}, x_{i,t-1}, x_{i+2,t-1})$
- 50 ensemble members

The variogram of L63

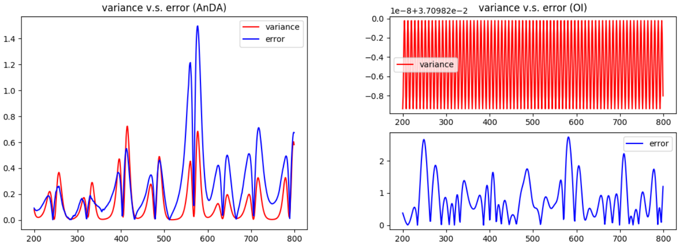


Numerical results of AnDA on L63 and L96

	OI	AnDA	$\sqrt{R_{obs}}$
L63	1.04	0.68	1.414
L96	2.0	0.95	1.0



Numerical results of AnDA on L63



The variance and error of AnDA/OI estimates.

A case when the observations are sparse in time

Lorenz 63, dt = 0.01, $dt_{obs} = 0.50$

	OI	AnDA	$\sqrt{R_{obs}}$
RMSE	4.78	6.57	1.414

Why does AnDA not work?

- Analog forecast (with locally linear regression) well captures the dynamics near the attractor of the system, but not the dynamics off the attractor;
- If the initial value is far away from the attractor, the locally linear operator might enlarge the error;
- Ensemble Kalman smoother also becomes unstable when the observation is sparse.

An Analog-Based Interpolation Method

- Find the k analogs that best match the observed trajectory;
- Use the mean of these k analogs as the interpolated result;
- Localization needs to be incorporated.

$$\{A_{1,t},..,A_{k,t}\} = \underset{A_{.,t}}{\text{argmin}} \sum_{\Delta t \in [-\Delta T, \Delta T]} w(\Delta t) \| h_{t+\Delta t}(A_{t+\Delta t}) - y^{o}_{t+\Delta t} \|^2$$

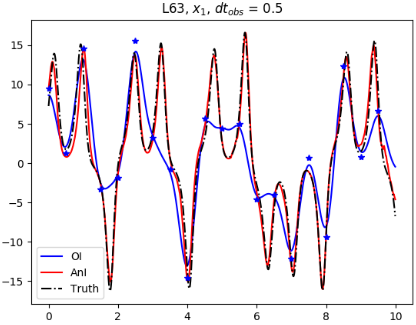
$$w(\Delta t) = \exp(-(\Delta t)^2 / L^2)$$

$$\widehat{A}_t = \frac{1}{k} \sum_i A_{i,t}$$

Mathematically, $\widehat{A}_t = \mathbb{E}(x_t | y_0, y_1, ..., y_T)$

Numerical Results on L63

	AnDA	OI	AnI	$\sqrt{R_{obs}}$
RMSE	6.57	4.78	1.49	1.414

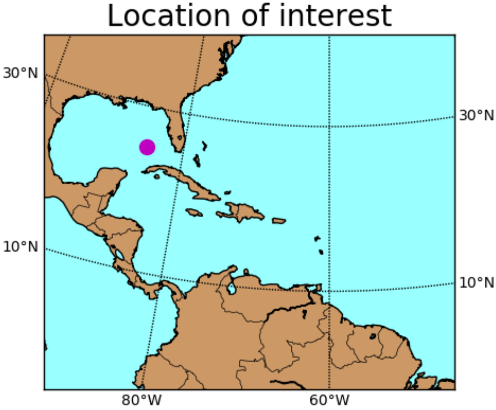


Numerical Results on Simulated SSH and Observations

- Region of interest: 10° X 10°, (40 X 36 grid);
- Data downloaded from OCCIPUT: 50 ensemble members, 20 years.
- Simulated obs at every grid-point, with $R_{obs} = 10\%$ of the climatological covariance of SSH;
- dt_{obs} = 15 days;
- The first 50 principal components of SSH are used to construct the analogs;
- AnI is implemented for each node separately.
- OI is implemented on the original SSH data.

Numerical Results on Simulated SSH and Obs

	AnI	OI
RMSE	0.033	0.0368



Summary and Future Work

- Analog-based methods outperform OI in Lorenz 63 and Lorenz 96.
- Analog interpolation slightly outperforms OI with simulated SSH data.
- Explore the potential of analog-based methods handling real satellite data