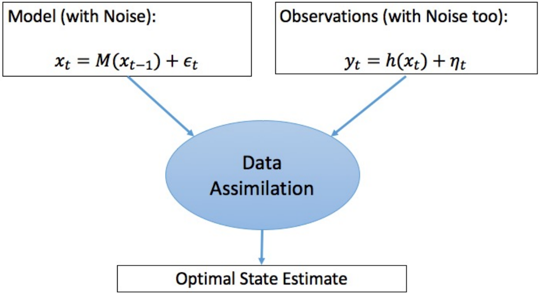
# Data Assimilation with Machine Learning Observation Operator

## an Application on Melt Pond Problem

Siyang Jing

Department of Mathematics, University of North Carolina at Chapel Hill
March 28th, 2019

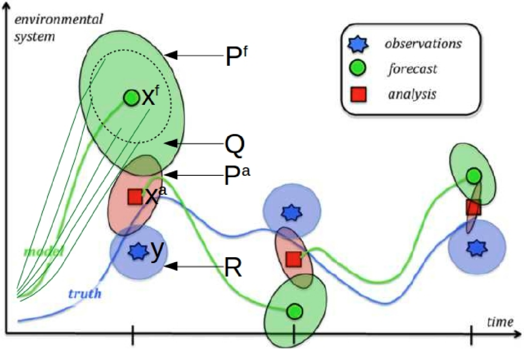
# Introduction to Data Assimilation



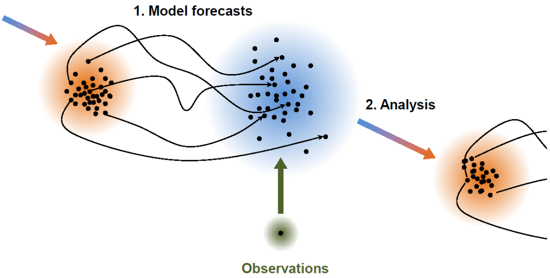

- Mathematical model is never a perfect description of the real world, e.g. dynamical systems from physical laws.
- Measurements, or observations, always come with error or uncertainty.
- Data assimilation is a way to combine these two types of information.

# Introduction to Data Assimilation

1. Prediction step:
   $\mathbf{x}_{t+1}^{f} = \mathcal{M}(\mathbf{x}_t^a)$
   $\mathbf{x}^f$: the forecast
2. Update step:
   $\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}^f))$
   $\mathbf{K}$: Kalman gain matrix
   $\mathbf{x}^a$: the analysis
3. Repeat the steps.



# Ensemble Kalman Filter



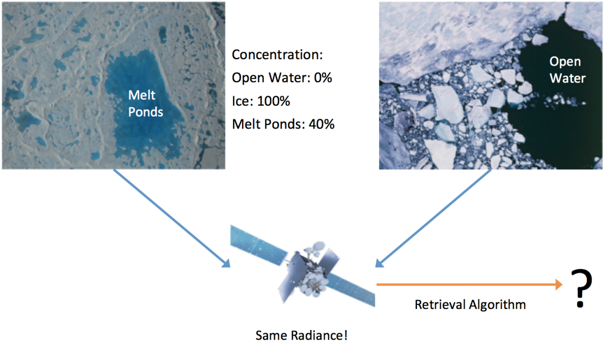

1. Ensemble members from last step
2. Evolve under model
3. Covariance matrix **P** from ensemble
4. Update ensemble using Kalman filter with **K** based on covariance **P**
5. New ensemble for the next step

# Sea Ice Concentration

- The polar regions are typically cloudy restricting continuous sea ice observations in the visible spectrum.
- Microwaves pass right through the clouds and microwave radiometry can provide continuous daily observations of sea ice concentration.
- NASA has an algorithm to retrieve ice concentration from satellite radiance.

# Problem with Melt Ponds






- Melt ponds have the same microwave signature as open water and obscure the signature of the ice below them. Therefore both the situations look roughly the same to the satellite.
- The retrieval algorithm will likely give incorrect ice concentration.
- Assimilating on wrong information makes the state estimation worse.
- e.g. Models predicts 90%, but observation is 60%. More error added!

# Assimilate with Radiance

- Change the observation space
  - Before: states $\rightarrow$ satellite radiance $\rightarrow$ (wrong) concentration
  - Proposed: states $\rightarrow$ satellite radiance
- Challenge: How to get $\mathcal{H}$: states $\rightarrow$ satellite radiance
  - Mathematical atmospheric model requires heavy computation. Not efficient and sometimes not accurate.
  - With sufficient historic data of sea ice optical images and corresponding satellite radiances, we can use machine learning to construct the observation operator.
- To test out our idea and to investigate the potential challenges in this approach, we start with a test bed model and synthetic data.

# Test Bed Model

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

- Instead of directly modeling ice concentration and other variables, we model the energy $E$ and maximum attainable albedo $\alpha_m$.
- Therefore, the state vector is

$$\mathbf{x} = [E \quad \alpha_m]$$

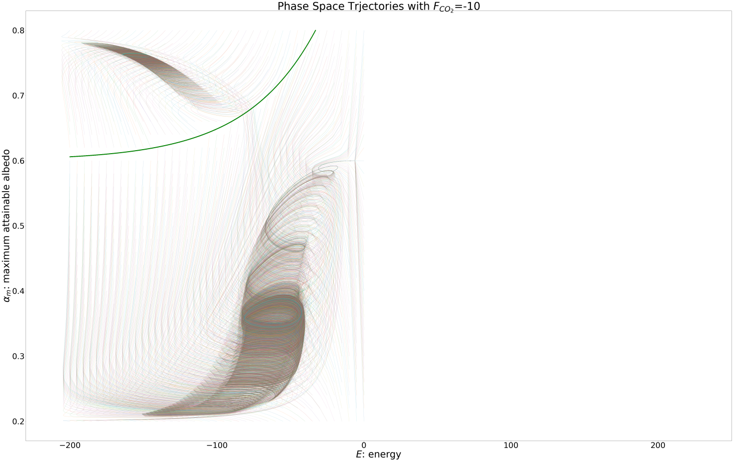
- The albedo can be calculated as

$$\alpha(E, \alpha_m) = \frac{\alpha_{ml} + \alpha_m}{2} + \frac{\alpha_{ml} - \alpha_m}{2} \tanh\left(\frac{E}{L_i h_c}\right)$$

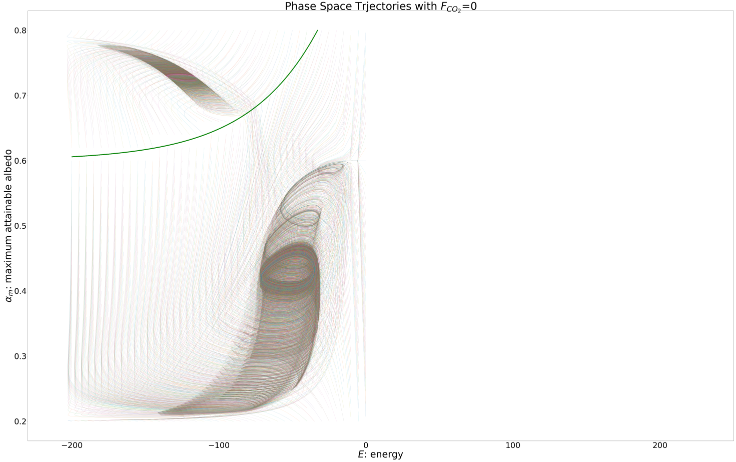
# Test Bed Model: Dynamical System

- When the temperature is high, which is corresponding to $\alpha(E, \alpha_m) > 0.6$
$\frac{dE}{dt} = [1 - \alpha(E, \alpha_m)]F_s(t) - F_0(t) + F_{co_2} - F_T(t)\frac{E}{c_{ml}H_{ml}} + F_B$
$\frac{d\alpha_m}{dt} = \frac{E^2}{K^2}\alpha_m\left(1 - \frac{\alpha_m}{0.8}\right) + \frac{K^2}{1+E^2}\alpha_m\left(1 - \frac{\alpha_m}{0.6}\right)$
- When the temperature is low, which is corresponding to $\alpha(E, \alpha_m) < 0.6$
$\frac{dE}{dt} = [1 - \alpha(E, \alpha_m)]F_s(t) - F_0(t) + F_{co_2} - F_T(t)\frac{E}{c_{ml}H_{ml}} + F_B$
$\frac{d\alpha_m}{dt} = \frac{K^2}{1+E^2}\alpha_m\left(1 - \frac{\alpha_m}{0.6}\right) + \frac{E^2}{K^2}\alpha_m\left(1 - \frac{\alpha_m}{0.2}\right)$
- One of the parameters of particular interest $F_{co_2}$: a combination of all exterior factors that contribute to the heat in the system, mainly the amount of $CO_2$.

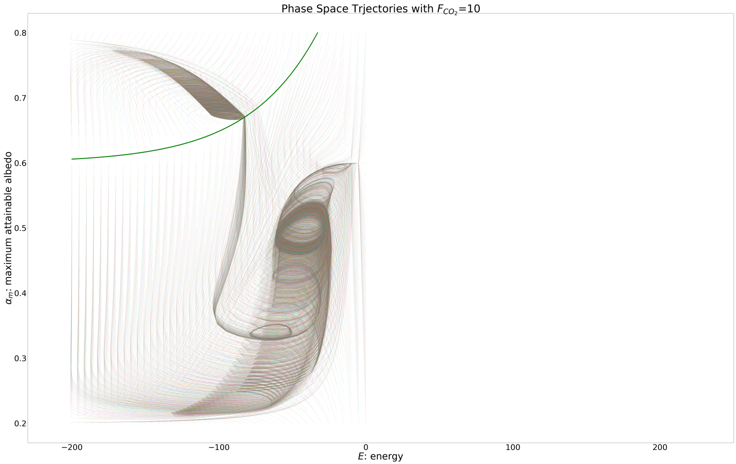
# Test Bed Model: Dynamical System



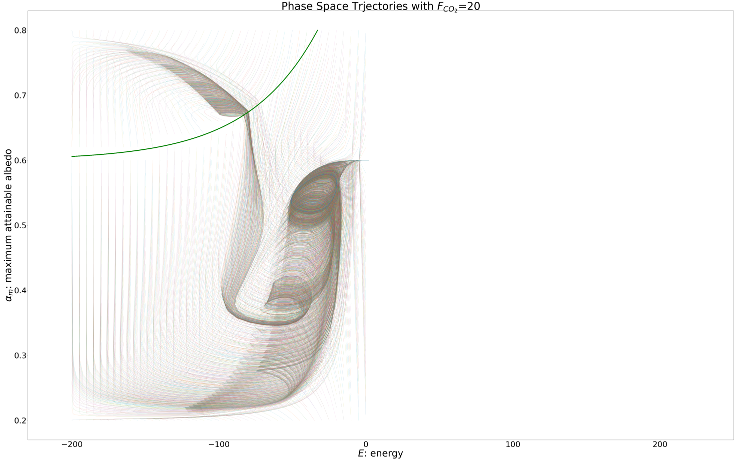
# Test Bed Model: Dynamical System



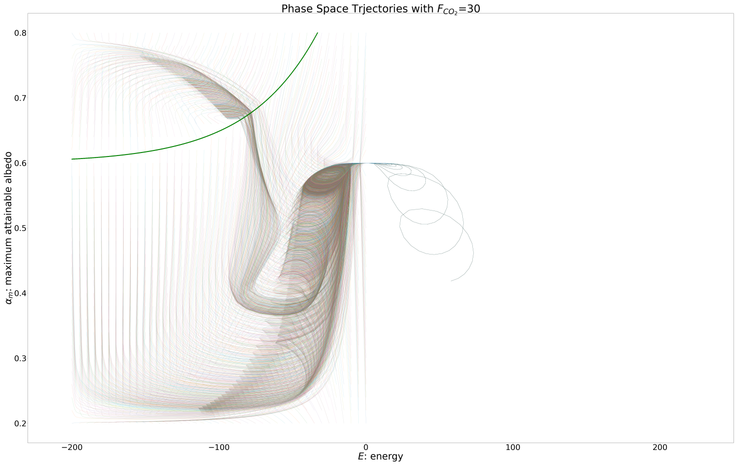
# Test Bed Model: Dynamical System



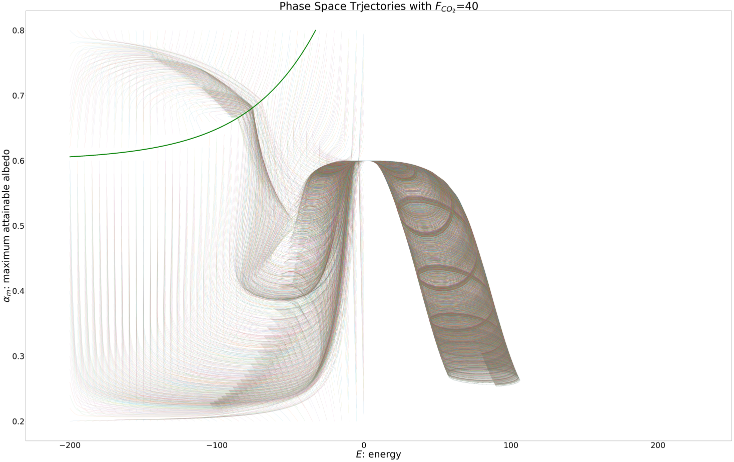
# Test Bed Model: Dynamical System



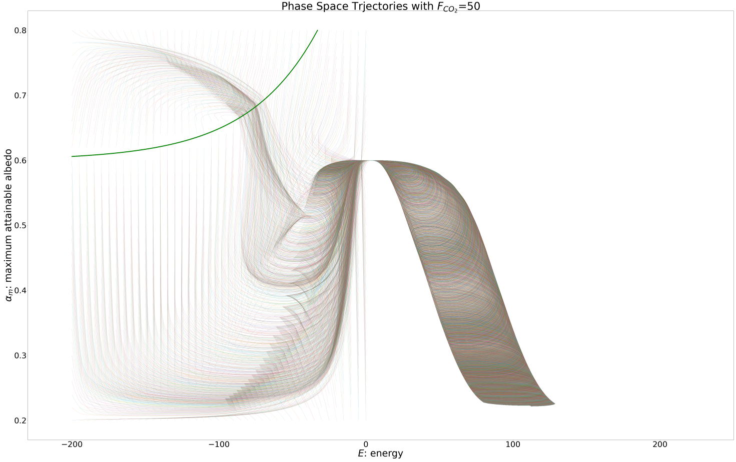
# Test Bed Model: Dynamical System



# Test Bed Model: Dynamical System

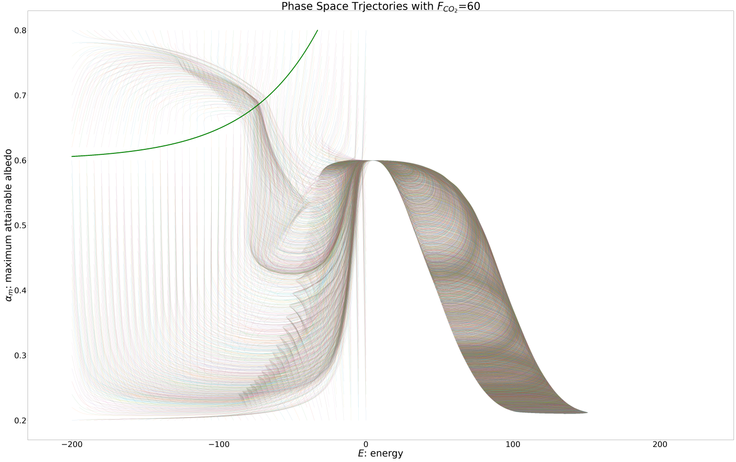


# Test Bed Model: Dynamical System

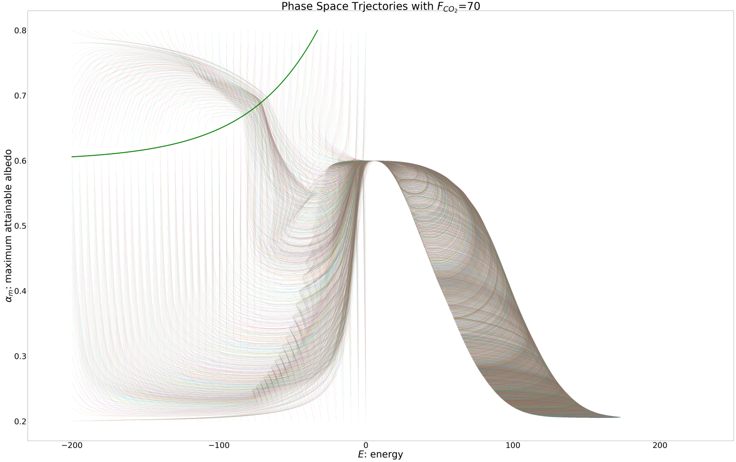


# Test Bed Model: Dynamical System

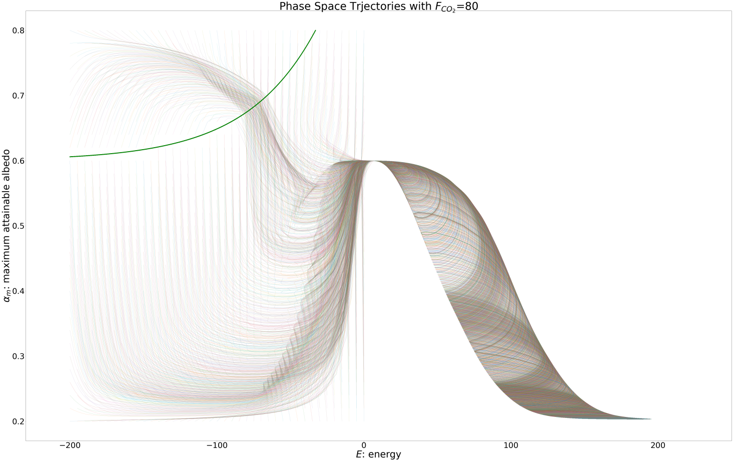
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL



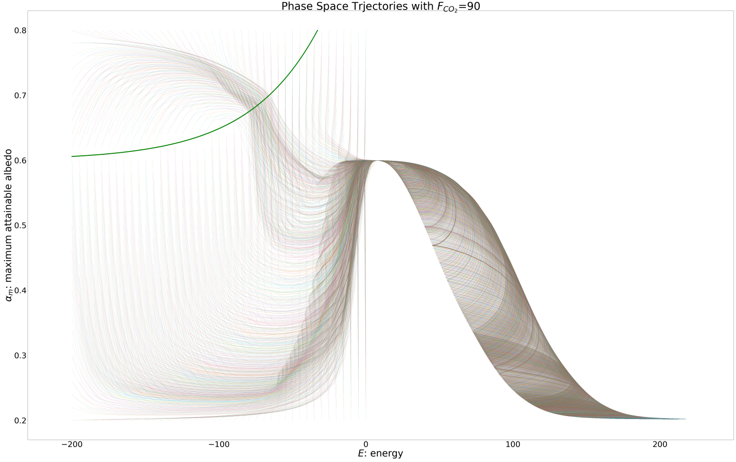
# Test Bed Model: Dynamical System



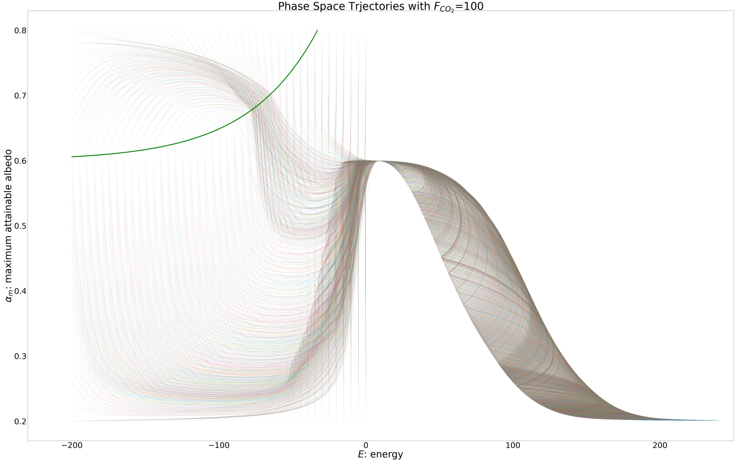
# Test Bed Model: Dynamical System



# Test Bed Model: Dynamical System



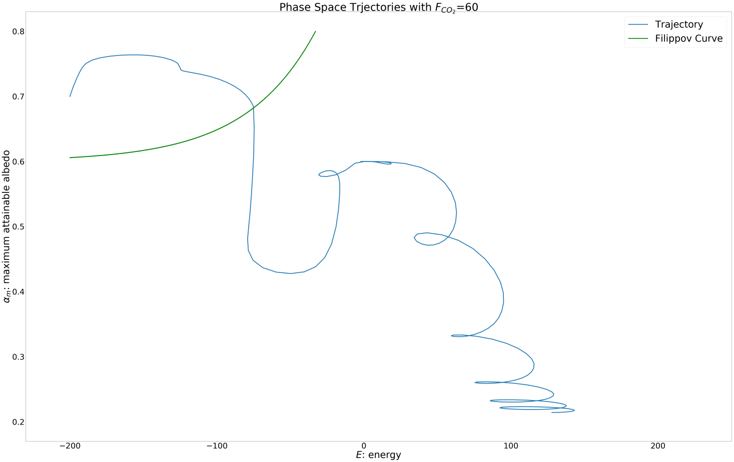
# Test Bed Model: Dynamical System



# Test Bed Model: Dynamical System

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

The trajectory that we use in data assimilation experiment.



# Test Bed Model: Observations

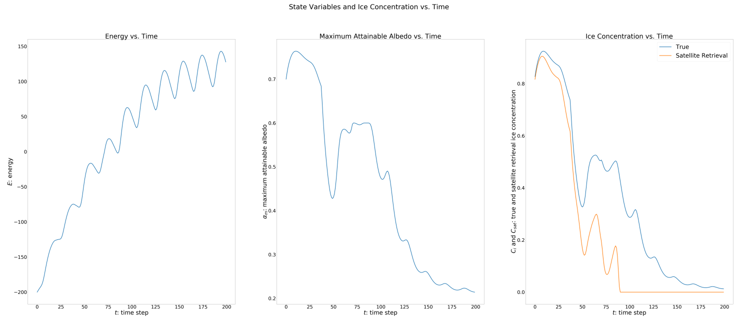


| | |
|---|---|
| Ice concentration | $C_i = 1 - \left(\frac{0.8 - \frac{1}{2}(\alpha_m + \alpha(E, \alpha_m))}{0.6}\right)$ |
| Pond concentration | $C_p = 1 - \frac{\alpha(E, \alpha_m)}{\alpha_m}$ |
| Satellite Radiances | $\lvert E\alpha_m \rvert$<br>$\alpha_m - \alpha(E, \alpha_m)$<br>$\alpha(E, \alpha_m)\lvert E \rvert$<br>$(0.5 + 0.4\tanh(\frac{50-E}{10}))(E + 273.15)$<br>$C_i C_p$ |
| Satellite retrieval concentration | $C_{sat} = \max(0, C_i - C_p)$ |

$C_{sat}$ is supposed to give $C_i$, but instead subtracts $C_p$ from it. If we still assume $C_{sat}$ gives the correct value of $C_i$ and use it in data assimilation, error will be incorporated in the analysis.

# Test Bed Model: Observations

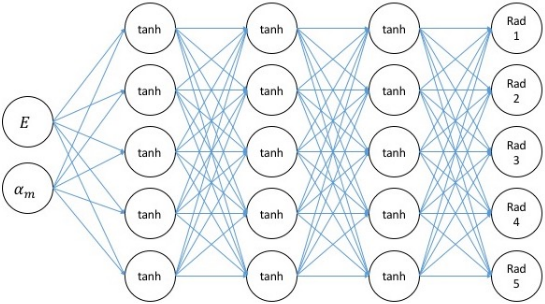
The following figure shows the error in satellite retrieved sea ice concentration.
Note that the error is especially large when energy is larger than $0$, corresponding to the situation when temperature rises and melt ponds starts to form.




This motivates us to construct an observation operator taking the state

# Machine Learning

- Since $\mathcal{H}$ is highly nonlinear, we decide to use neural network as the machine learning algorithm.
- A single layer of neural network: $\boldsymbol{x}_{out} = \tanh(\boldsymbol{W}\boldsymbol{x}_{in} + \boldsymbol{b})$
- We compose 4 layers to get the following architecture[1]:



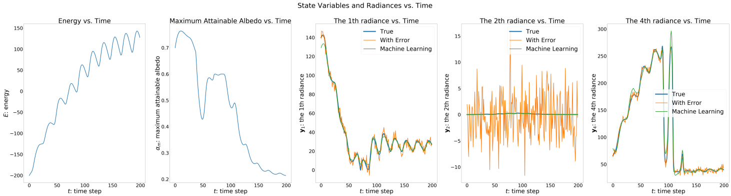

[1] Biases and techniques such as batch-norm and drop-out are not shown

# Data Generation

- $\alpha_m$ is only meaningful in $[0.2, 0.8]$. Energy lower than $-200$ would be unrealistic and uninteresting.
  - $\mathcal{H}_{\text{grid}}$: A benchmark model trained on $120 \times 180$ uniform grid data points in the state space $[0.2, 0.8] \times [-200, 250]$.
  - $\mathcal{H}_{\text{ML}}$: The model that mimics what we would get with the real world data, trained on trajectories in the phase space.
- Real data comes with error, so we need to add error to the generated observations. For each of the 5 radiances, we add Gaussian error with standard deviation proportional to the mean absolute value of that radiance, i.e. $\tilde{y}_{ij} = y_{ij} + \mathcal{N}(0, \lambda \bar{y}_j)$, where $\bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} |y_{ij}|$.
- We use different values of $\lambda$ from $0\%$ to $60\%$ to investigate the robustness of machine learning algorithm to error in training data.

# Evaluation

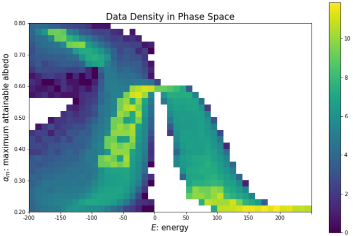
The following figure compares the machine learning predictions with the truth and noisy data with $\lambda = 40\%$.

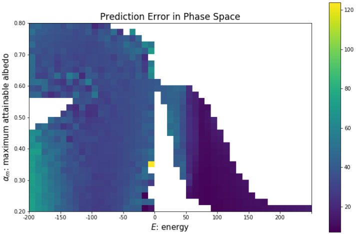



- It seems that machine learning is able to average out the noise and discover the real patter. For most values, machine learning prediction is very close to the true value.
- In fact, through experiment, we found out that, with appropriate techniques and sufficient data, machine learning is not severely affected by the amount of noise in the training data. The model trained with $\lambda = 60\%$ is only slightly worse than the model trained with $\lambda = 0\%$.

# Effect of Data Density

- For the equilibrium states, where sufficient training data exists, our model tends to perform quite well.
- For the transient state, where training data is extremely sparse, the prediction does not make much sense.

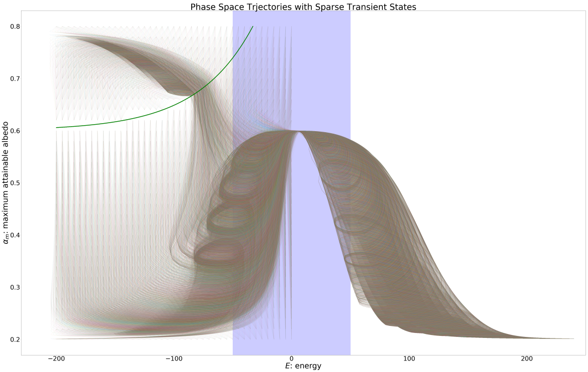



We address this issue in data assimilation with appropriate inflation.

# Training with Sparse Data

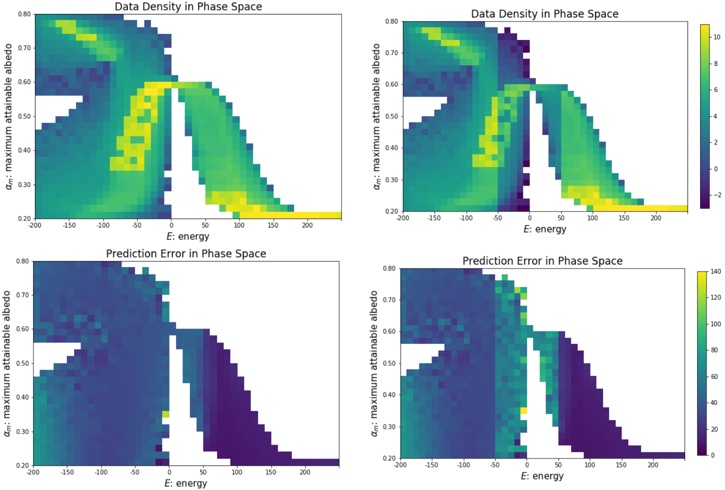
What if we don't have sufficient data in certain regions of interest?




We randomly drop $95\%$ / $70\%$ of the data in the shaded region to investigate the robustness of machine learning to sparse data, and the

# Training with Sparse Data: Evaluation

With only $5\%$ of the data, our neural network starts to make mistakes.



# Experiment

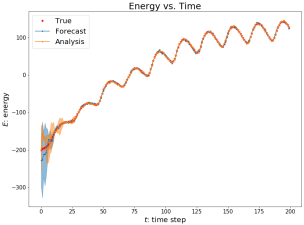


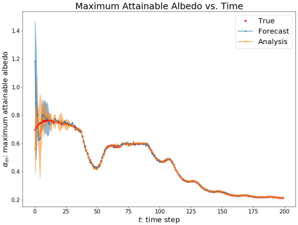
Experiment setup:

- Initial condition: $\mathbf{x}_0 = [0.7, -200]$, $F_{CO_2} = 60$
- Initial ensemble mean: $\mathbf{x}_0 + \mathcal{N}(0, \begin{bmatrix} 0.12 & 0 \\ 0 & 40 \end{bmatrix})$
- Initial ensemble spread: $\begin{bmatrix} 0.12 & 0 \\ 0 & 40 \end{bmatrix}$
- Size of ensemble: 100
- Time step: $0.05 \times 200$
- Observation error: $20\%$ of the mean absolute value

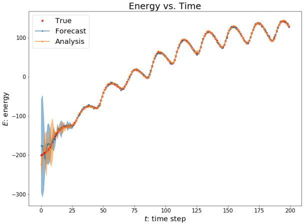
# Results: Baseline Model

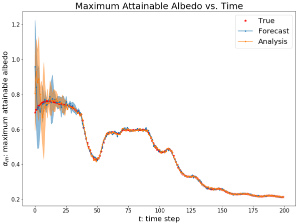
Machine learning model:




True radiance operator:

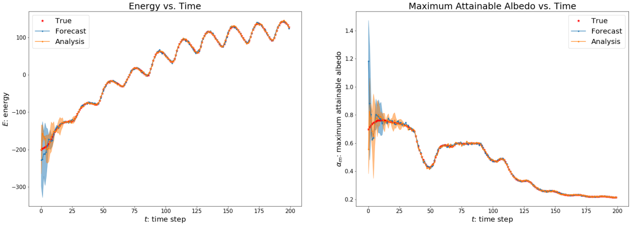


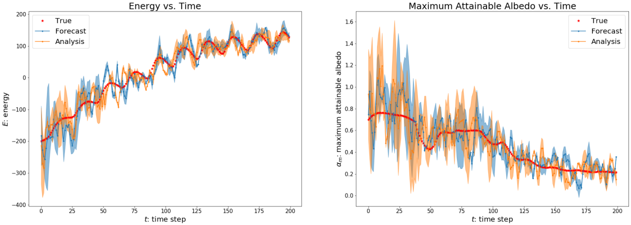
# Results: Baseline Model

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Machine learning model:

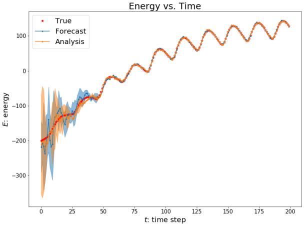


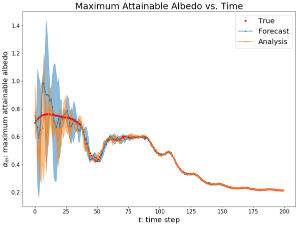

Retrieval algorithm:



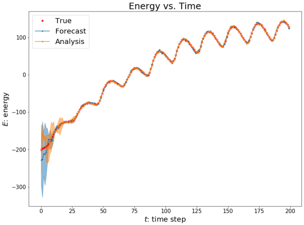
# Results: Realistic Model

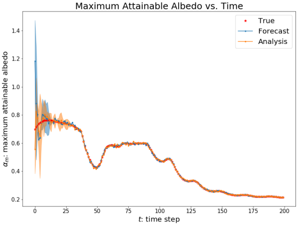
Machine learning model with trajectory training data:



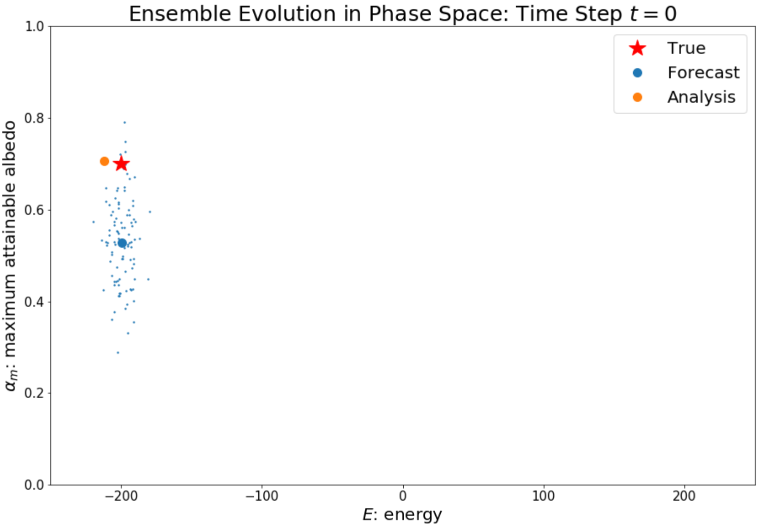




Machine learning model with grid training data:

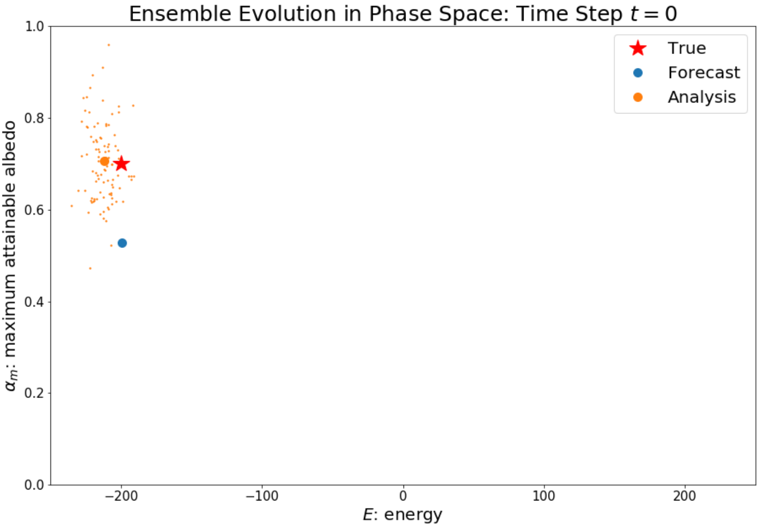


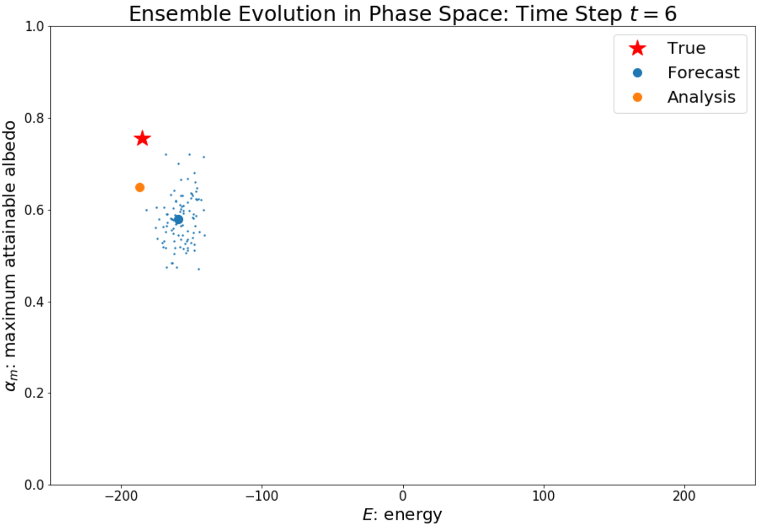
# Results: Realistic Model



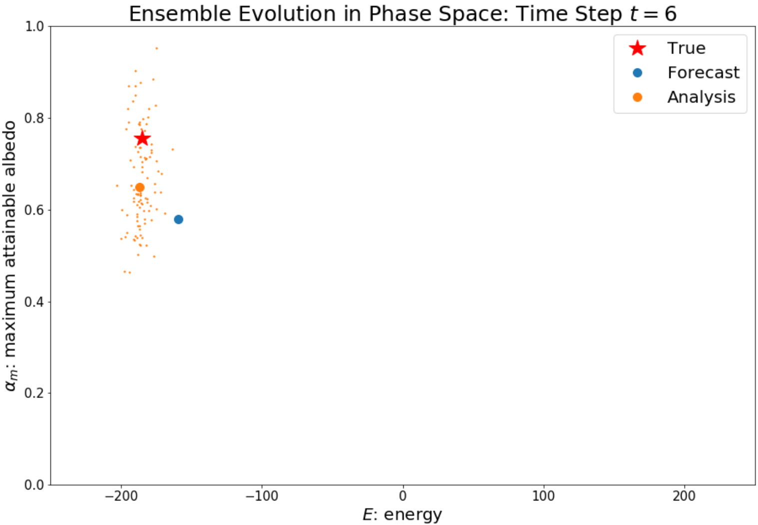
# Results: Realistic Model



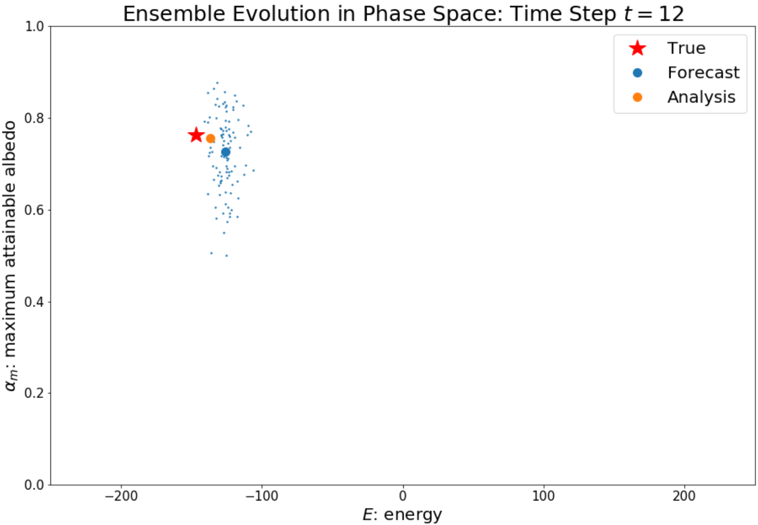
# Results: Realistic Model



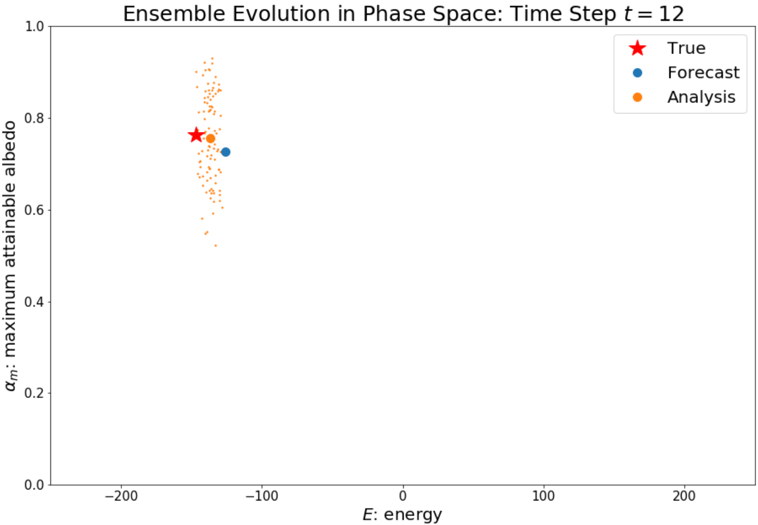
# Results: Realistic Model



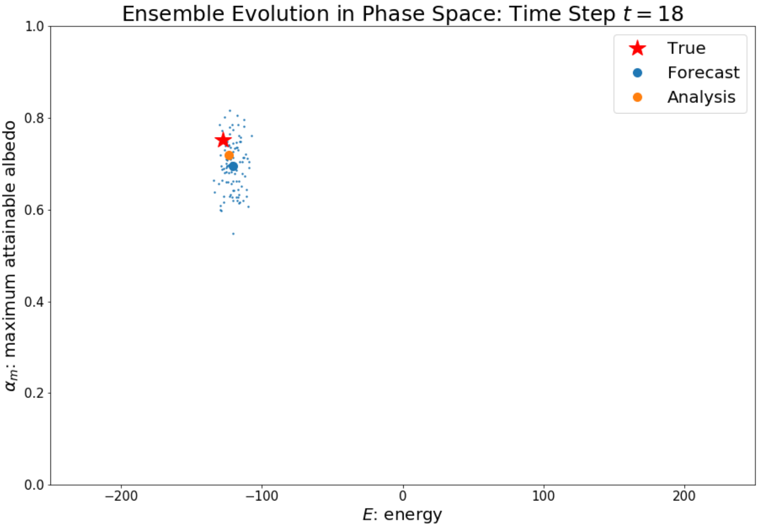
# Results: Realistic Model



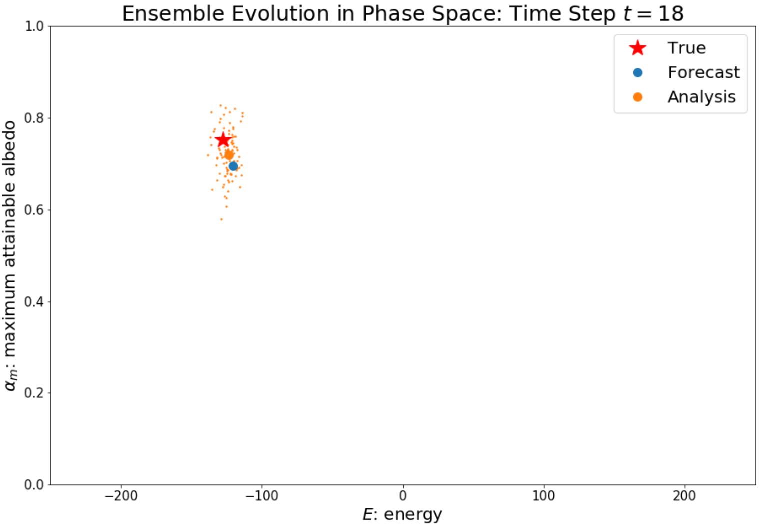
# Results: Realistic Model



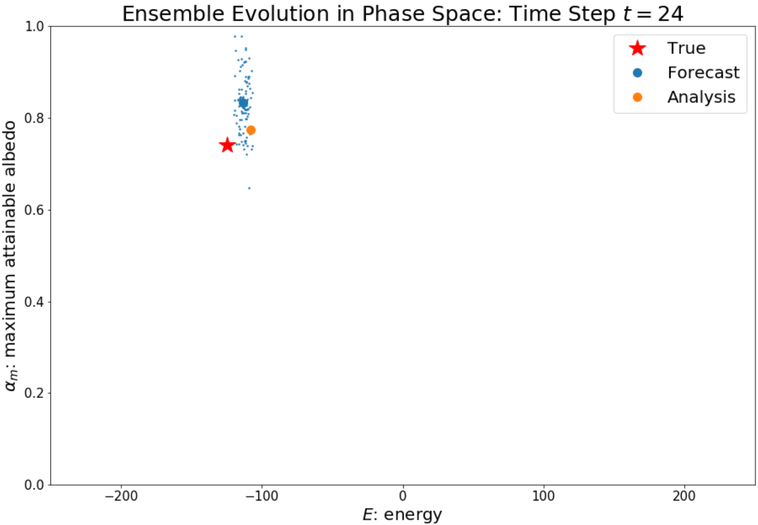
# Results: Realistic Model



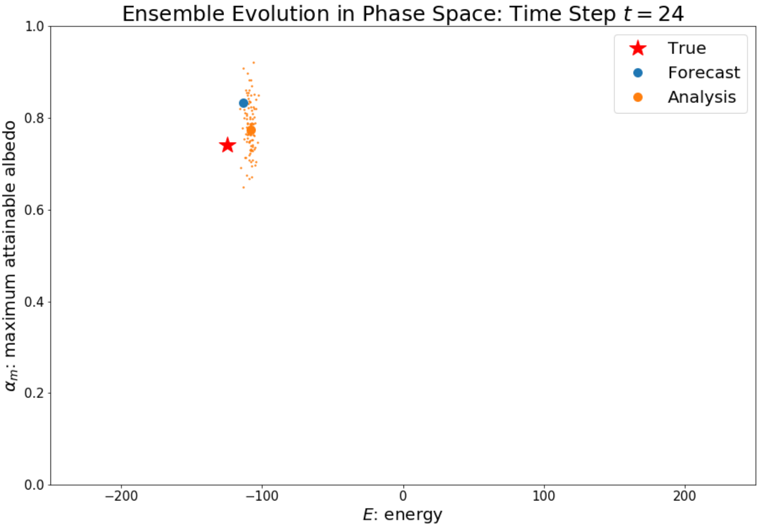
# Results: Realistic Model



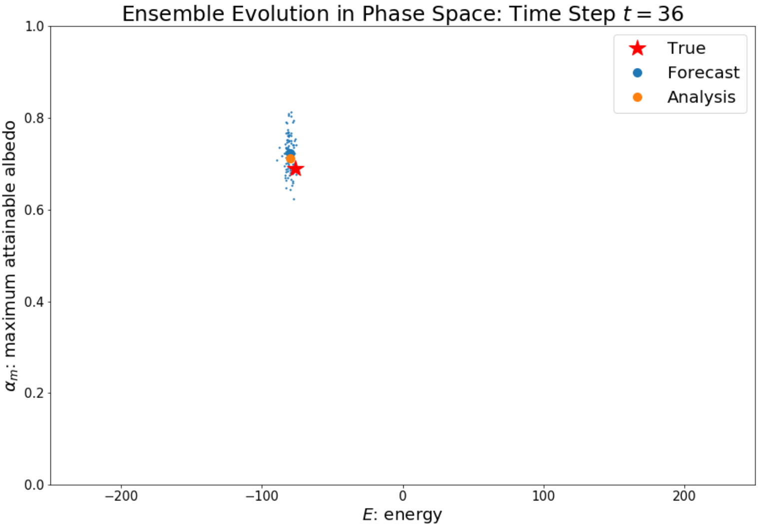
# Results: Realistic Model



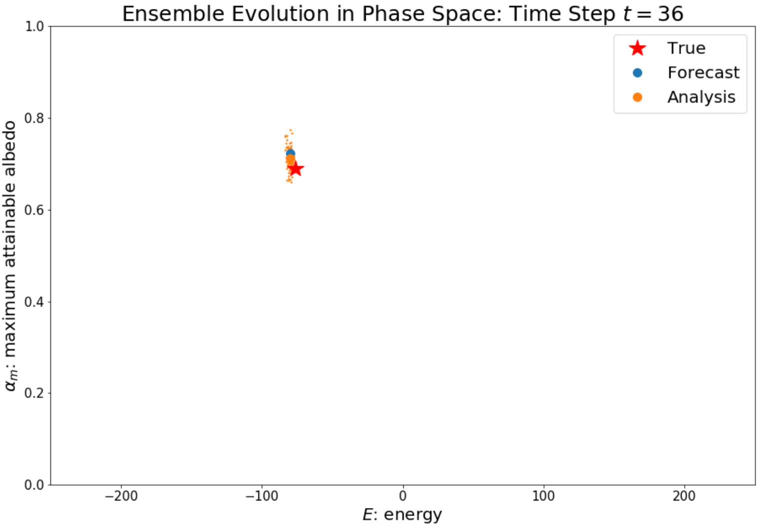
# Results: Realistic Model



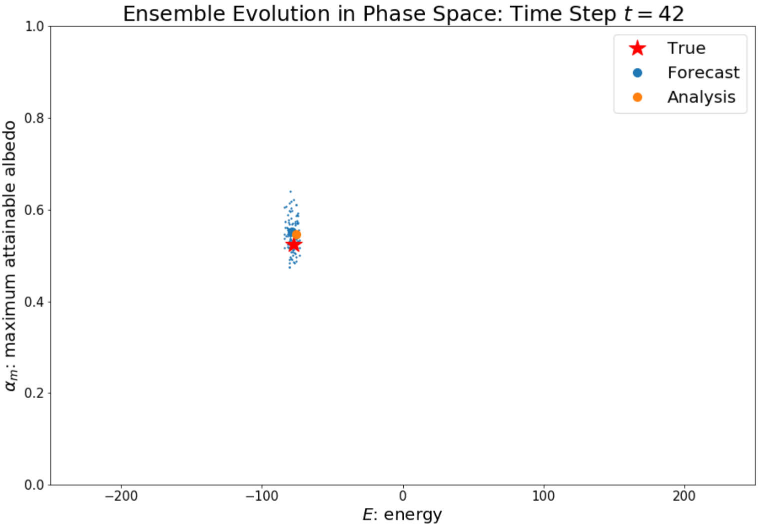
# Results: Realistic Model



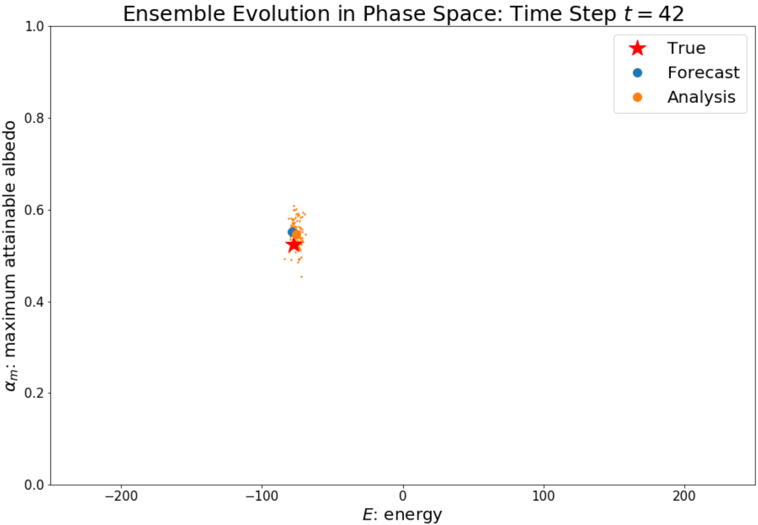
# Results: Realistic Model



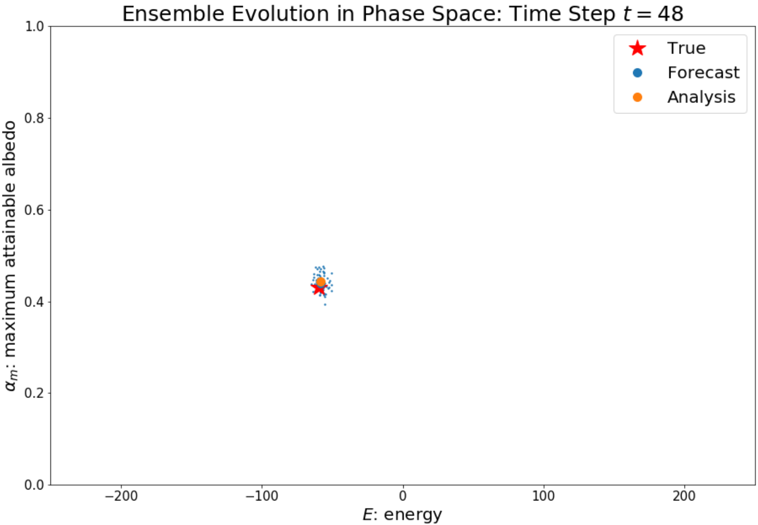
# Results: Realistic Model



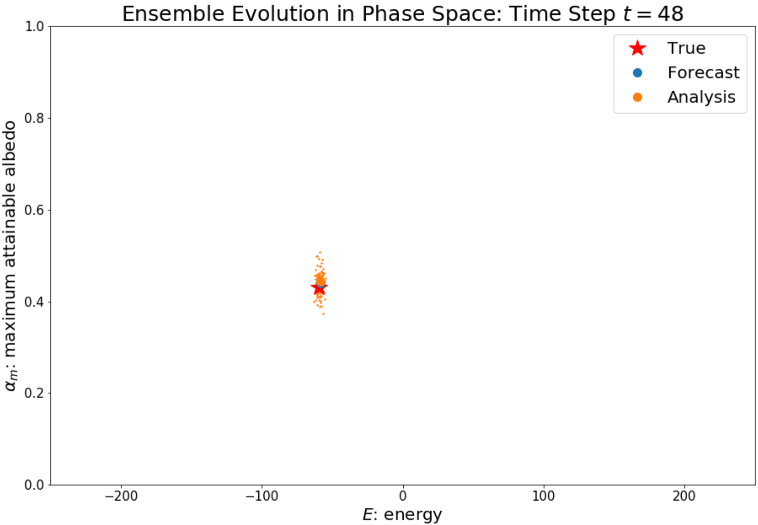
# Results: Realistic Model



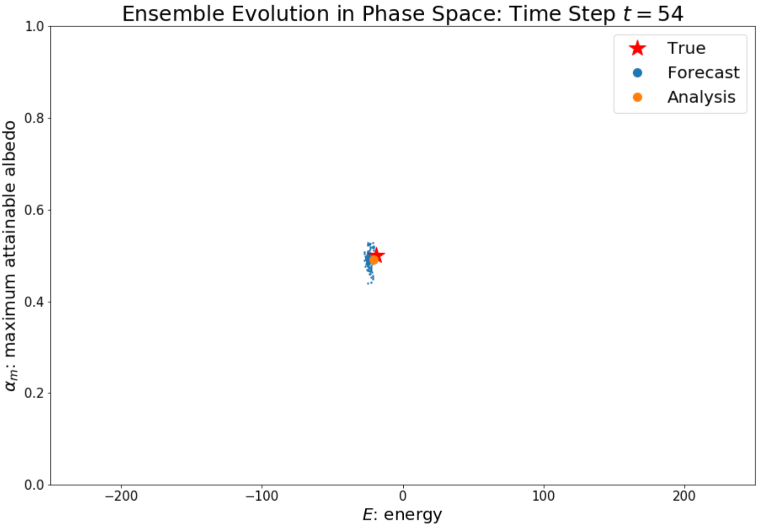
# Results: Realistic Model



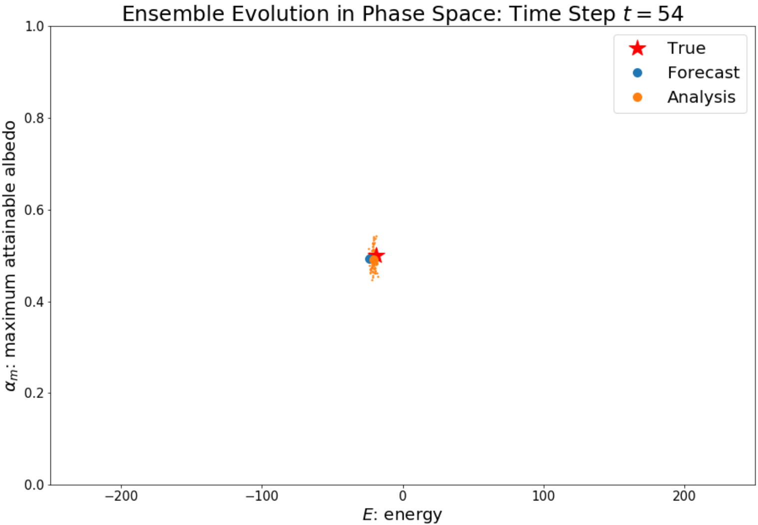
# Results: Realistic Model



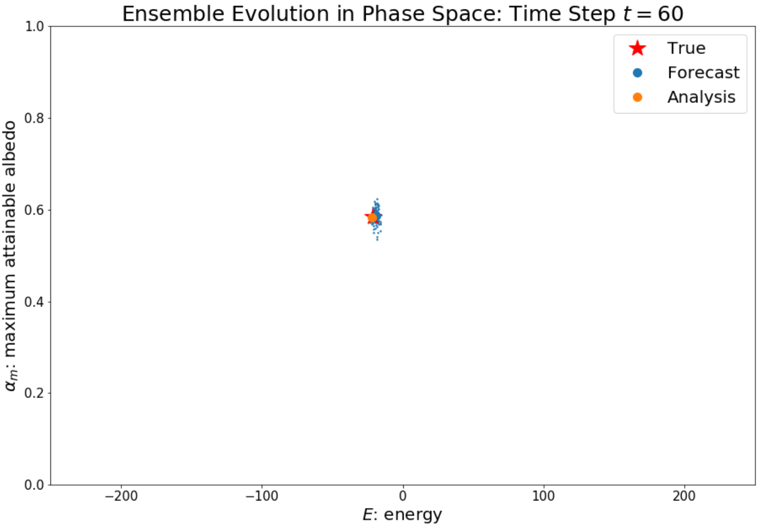
# Results: Realistic Model



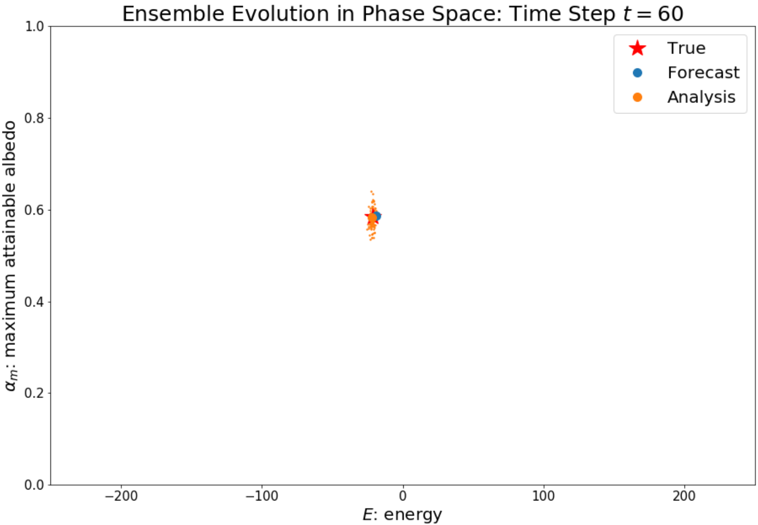
# Results: Realistic Model



# Results: Realistic Model



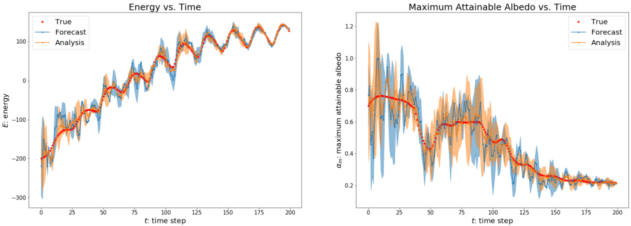
# Results: Realistic Model



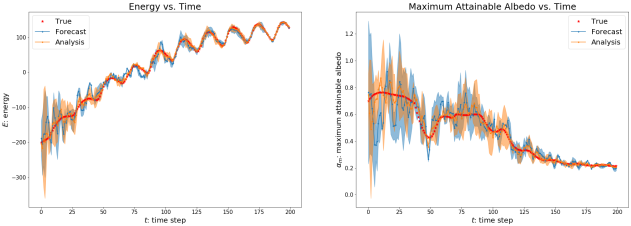
# Results: Sparse Data



Keep $5\%$ data in $E \in [-50, 50]$ for training:

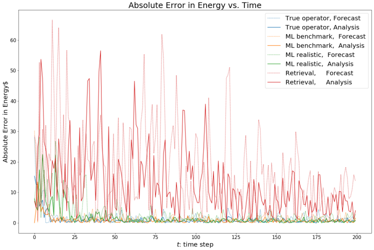


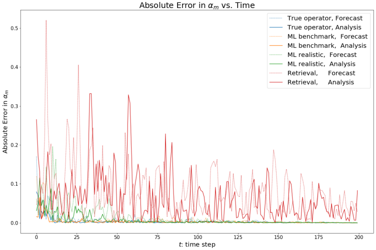

Keep $30\%$ data in $E \in [-50, 50]$ for training:



# Review and Comparison

- The figure below compares the performance of different observation operators in terms of forecast and analysis absolute error at each time step.
- Obviously, data assimilation with the retrieved concentration performs far worse than the others.

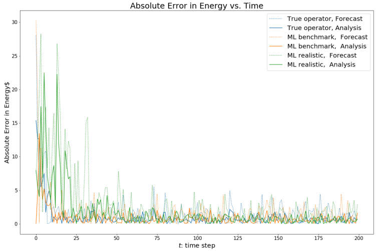


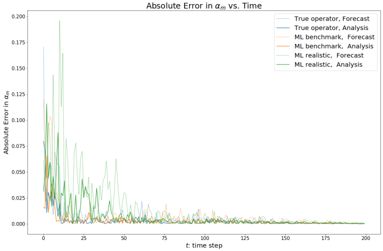



# Review and Comparison



- The behavior in $E$ and $\alpha_m$ are consistent.
- Around the first 30 time steps, $\mathcal{H}_{\text{ML}}$ does not perform very well. This is expected since, initially, the system is in transient states where the machine learning algorithm does not have sufficient data to learn the true observation operator.
- After a while, the error of $\mathcal{H}_{\text{ML}}$ decreases to roughly the same as $\mathcal{H}$ and $\mathcal{H}_{\text{grid}}$.



# Discussion

The experiments demonstrate that

1. Assimilating directly on satellite radiance with machine learning observation operator works well and far outperforms assimilating on inaccurately retrieved sea ice concentration. With sufficient amount of data, it is able to achieve roughly the same performance as the true observation operator.
2. Data density is an important issue. In transient states, without sufficient data, the machine learning observation operator could make large error. However, the performance is still better than assimilating with wrong concentration.
3. We can use the performance on test set to estimate the error covariance matrix of machine learning observation operator and apply covariance inflation accordingly at corresponding locations in the phase space.

# Thank you!

THE UNIVERSITY
*of* NORTH CAROLINA
*at* CHAPEL HILL

- This study is conducted in collaboration with Dr. Christian Sampson under the supervision of Prof. Christopher Jones.
- All images are from internet.

THE UNIVERSITY
*of* NORTH CAROLINA
*at* CHAPEL HILL