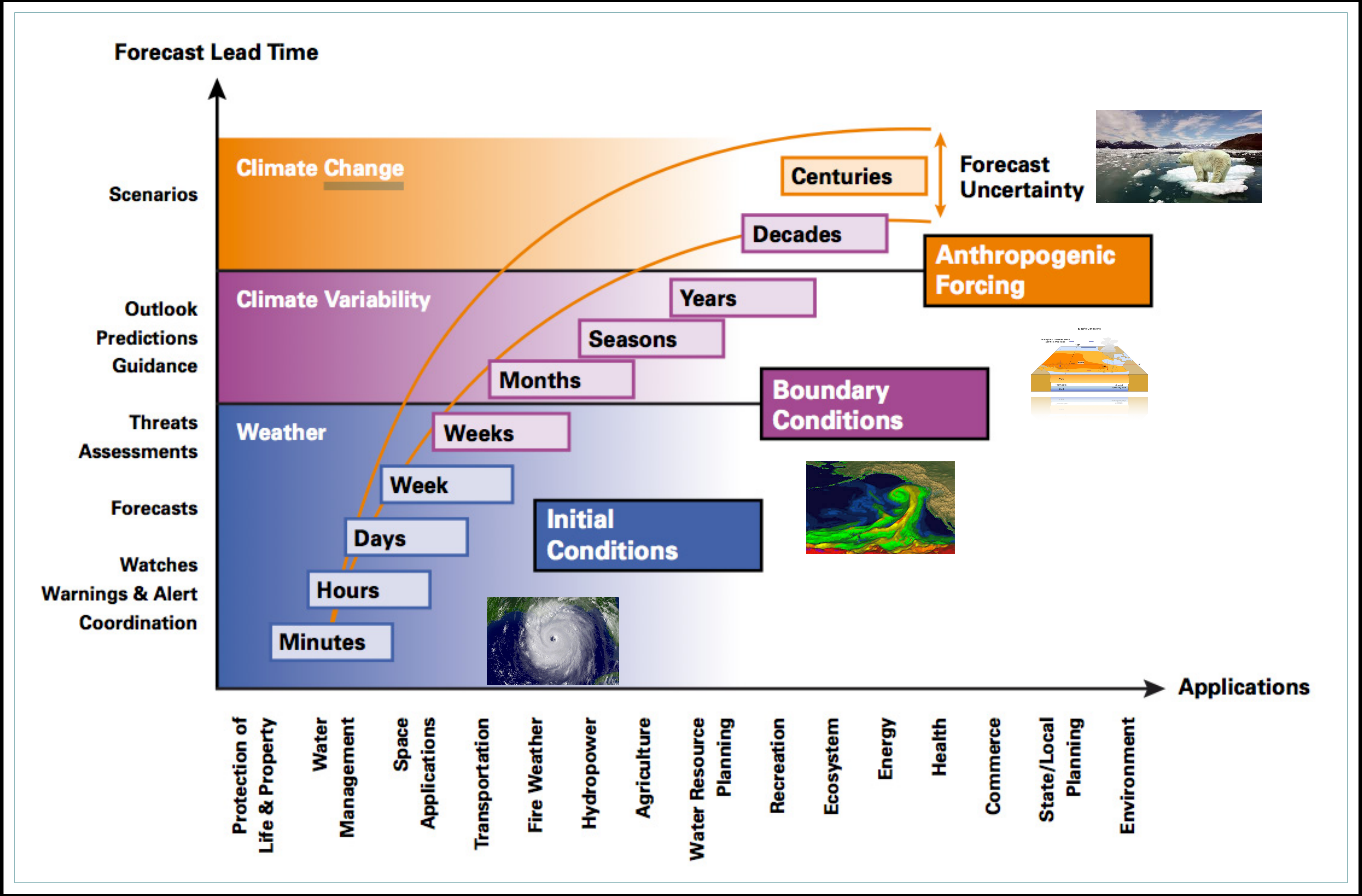
Machine Learning for Data Assimilation and Model Error Representation

Aneesh Subramanian

Hurricane Beryl over the Caribbean (taken by NASA astronaut Matthew Dominick) source: nasa.gov/ISS

Motivation

- Improve probabilistic weather & climate forecasts
- Better representation of uncertainty in forecasts



source: https://www.wmo.int/

Learning Goals

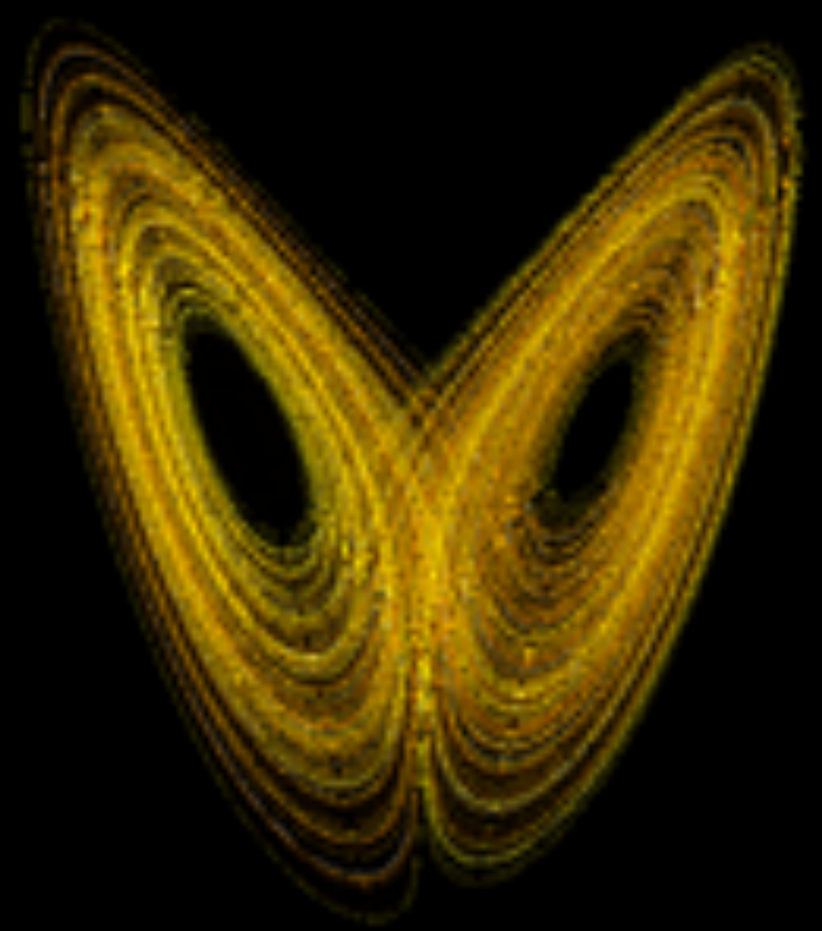
- Learn why representing initial condition uncertainty is necessary in ensemble forecasts
- Learn about the two main sources for uncertainty and how they impact ensemble forecasts
- Learn how machine learning can augment traditional data assimilation
- Learn why stochastic parametrization is necessary to represent model uncertainty
- Learn how machine learning can be used for model uncertainty representation

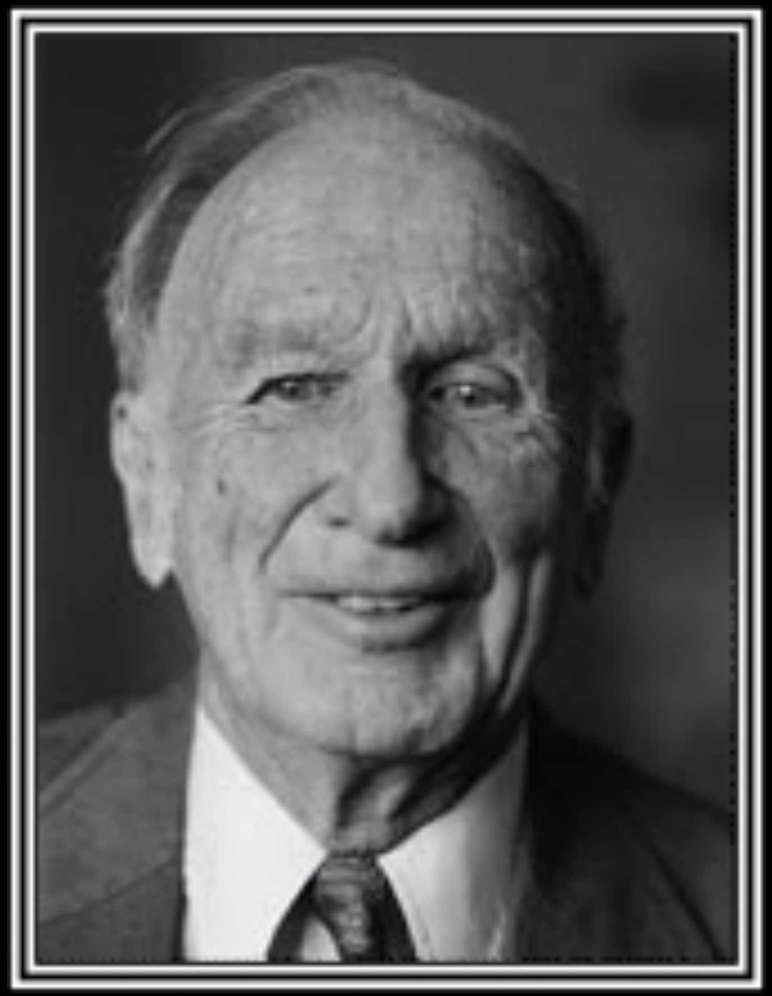

DURING THIS TRIAL, WE INTEND TO PROVE THAT THIS BUTTERFLY, BY NEGLIGENTLY FLAPPING HIS WINGS, SET IN MOTION A CHAIN OF EVENTS, RESULTING IN A TORNADO, MAKING HIM FINANCIALLY RESPONSIBLE...
WHAMOND 9-27

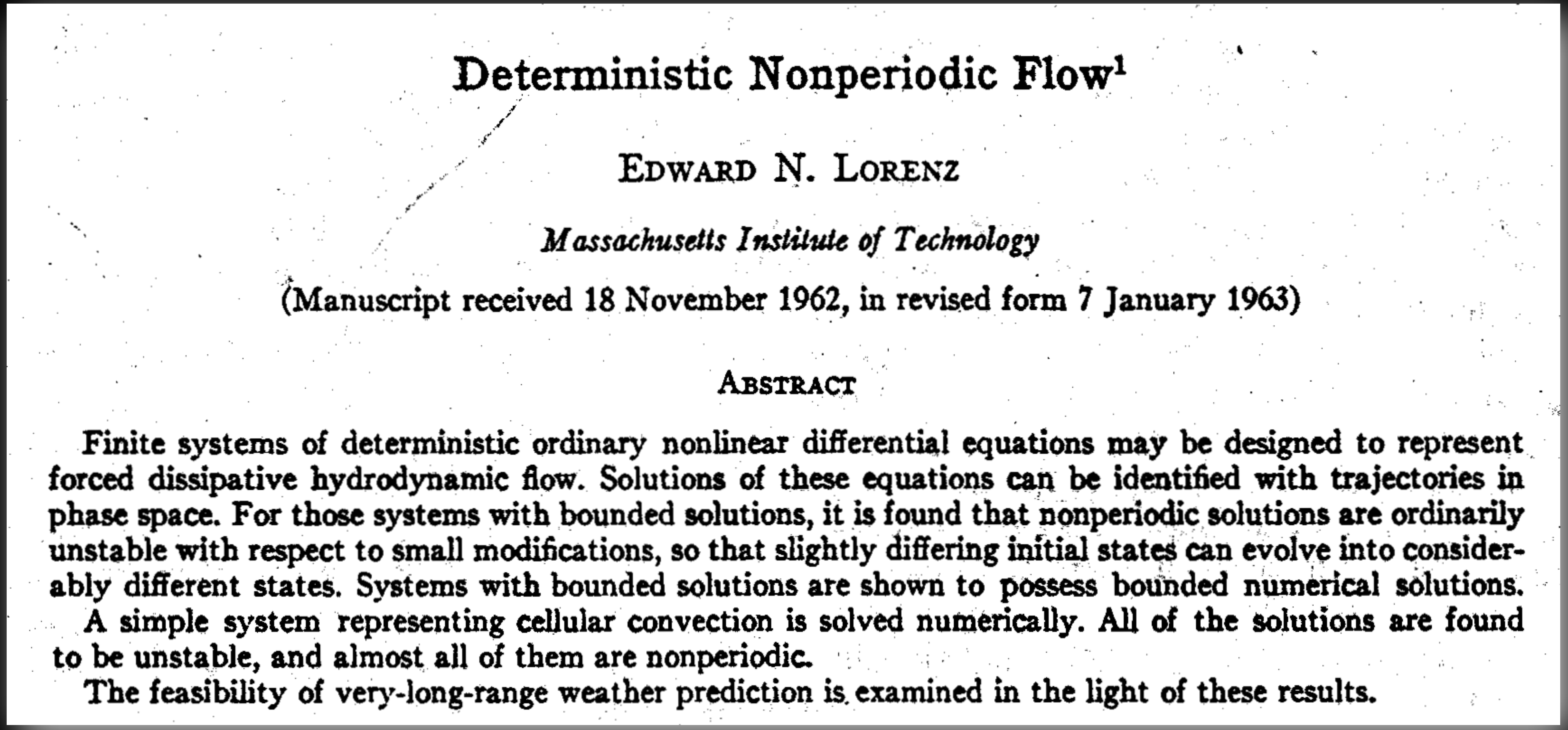
Predictability in a deterministic nonperiodic flow

"Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?"

-(Lorenz 1972)







Deterministic Nonperiodic Flow[1]

EDWARD N. LORENZ

Massachusetts Institute of Technology

(Manuscript received 18 November 1962, in revised form 7 January 1963)

ABSTRACT

Finite systems of deterministic ordinary nonlinear differential equations may be designed to represent forced dissipative hydrodynamic flow. Solutions of these equations can be identified with trajectories in phase space. For those systems with bounded solutions, it is found that nonperiodic solutions are ordinarily unstable with respect to small modifications, so that slightly differing initial states can evolve into considerably different states. Systems with bounded solutions are shown to possess bounded numerical solutions.

A simple system representing cellular convection is solved numerically. All of the solutions are found to be unstable, and almost all of them are nonperiodic.

The feasibility of very-long-range weather prediction is examined in the light of these results.

Sensitive dependence to initial conditions

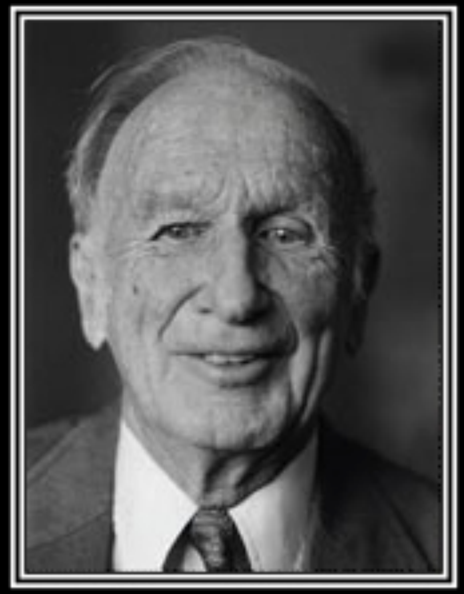
"Finite time for error in representation of small scales to affect accuracy of simulation of large scales, no matter how small in scale and hence amplitude this model error is"

-(Lorenz 1969)

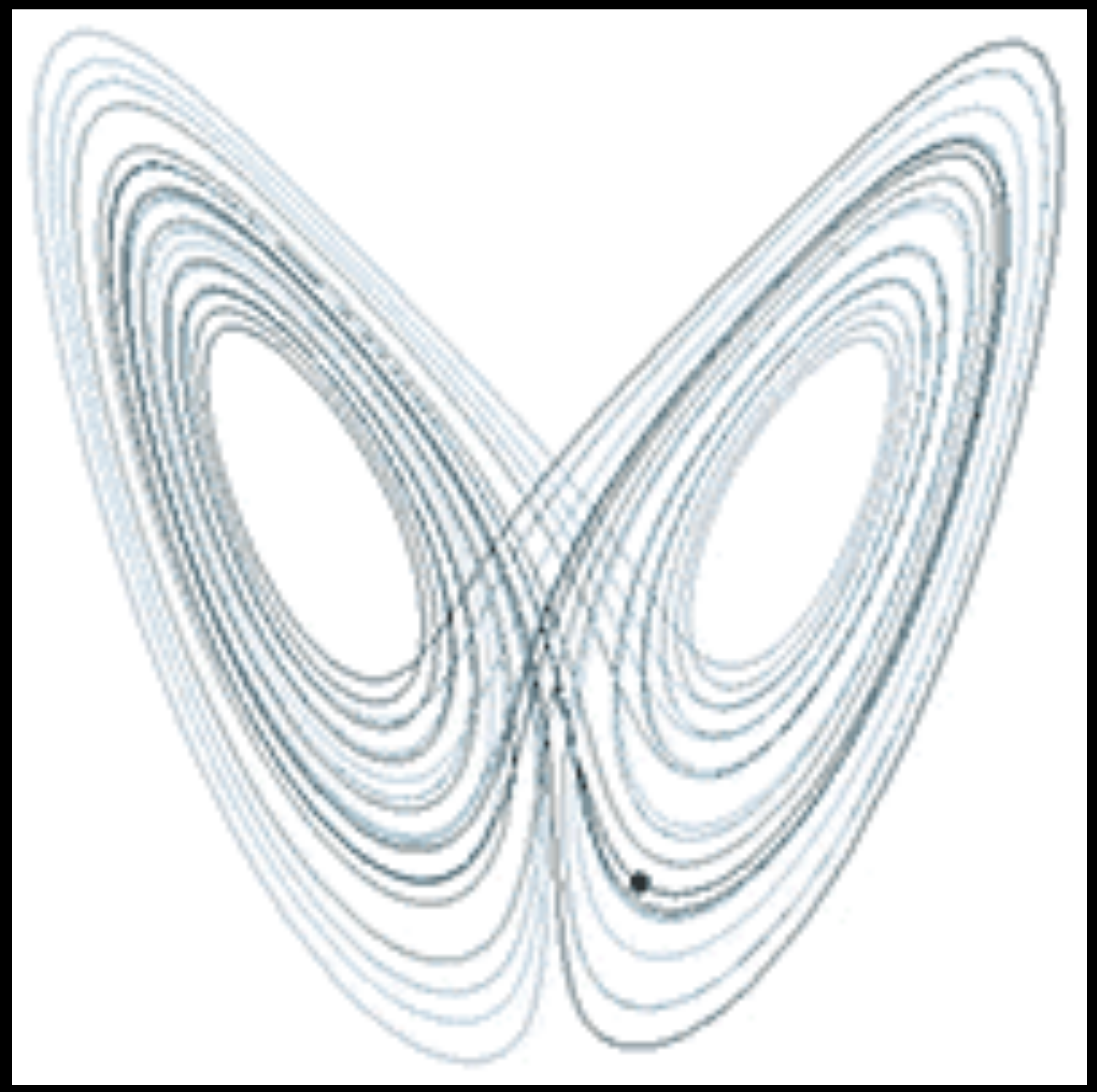
$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = rx - y - xz$$

$$\frac{dz}{dt} = xy - bz$$

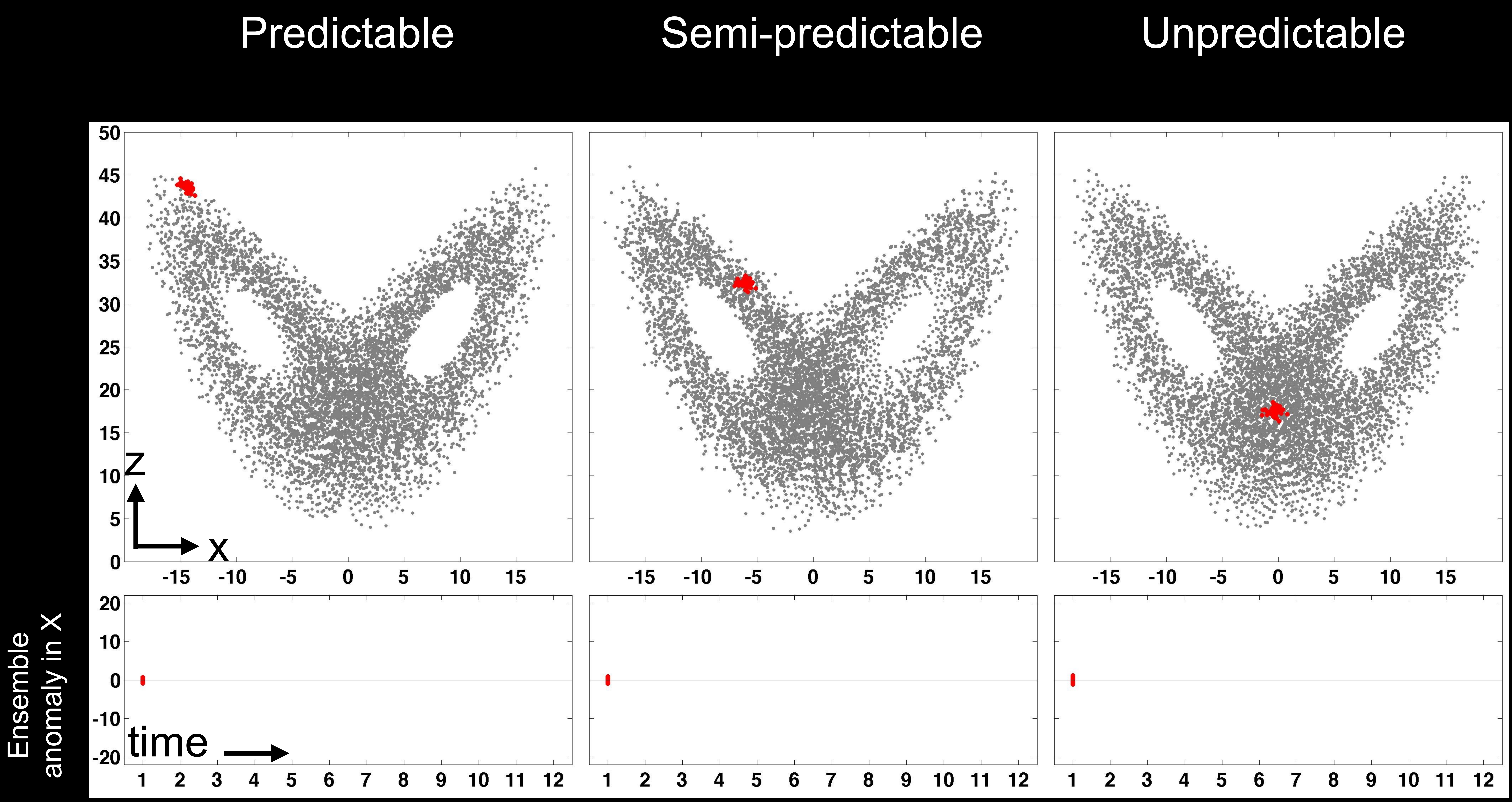


r = 28, σ = 10, and b = 8/3

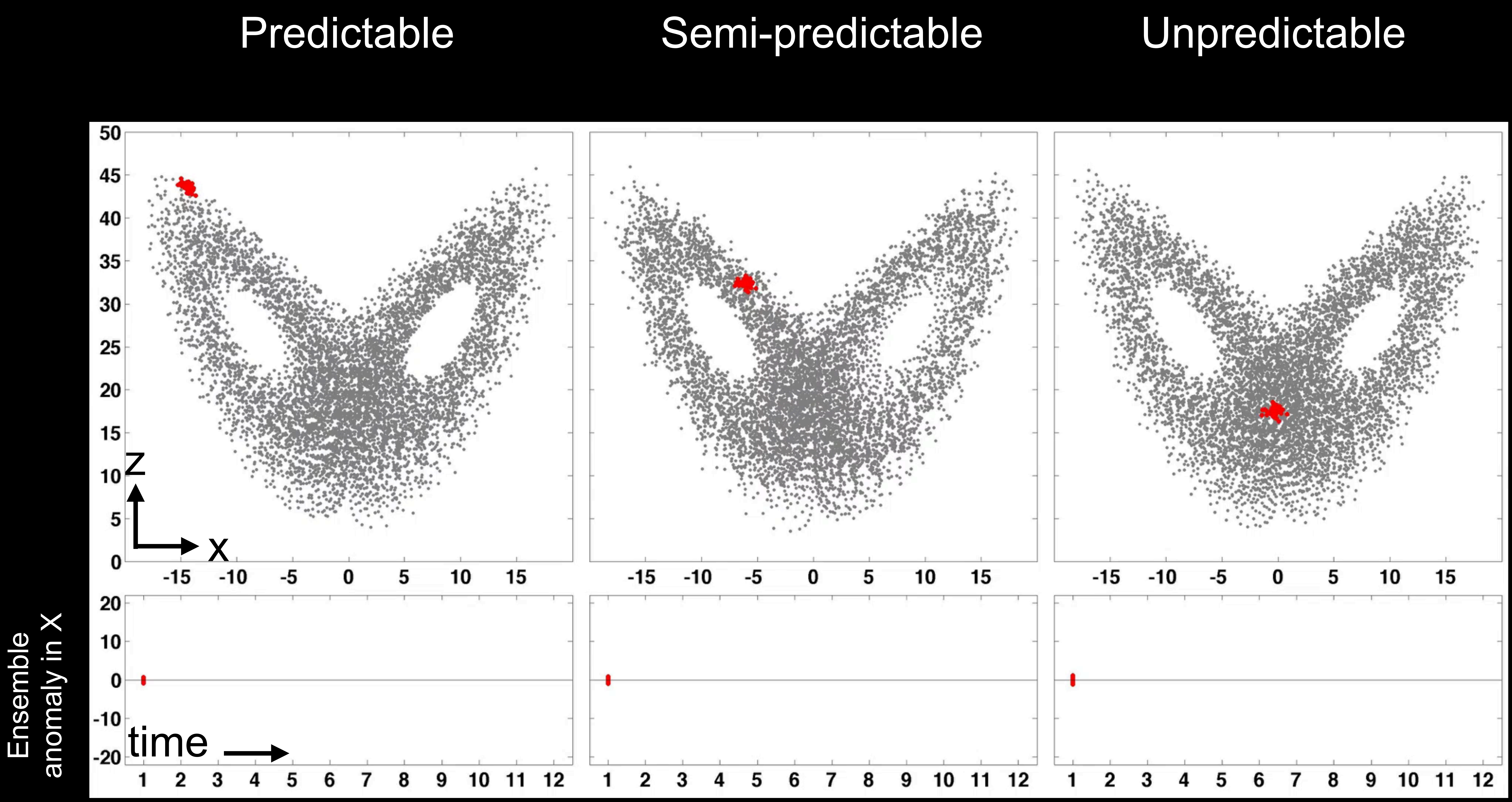


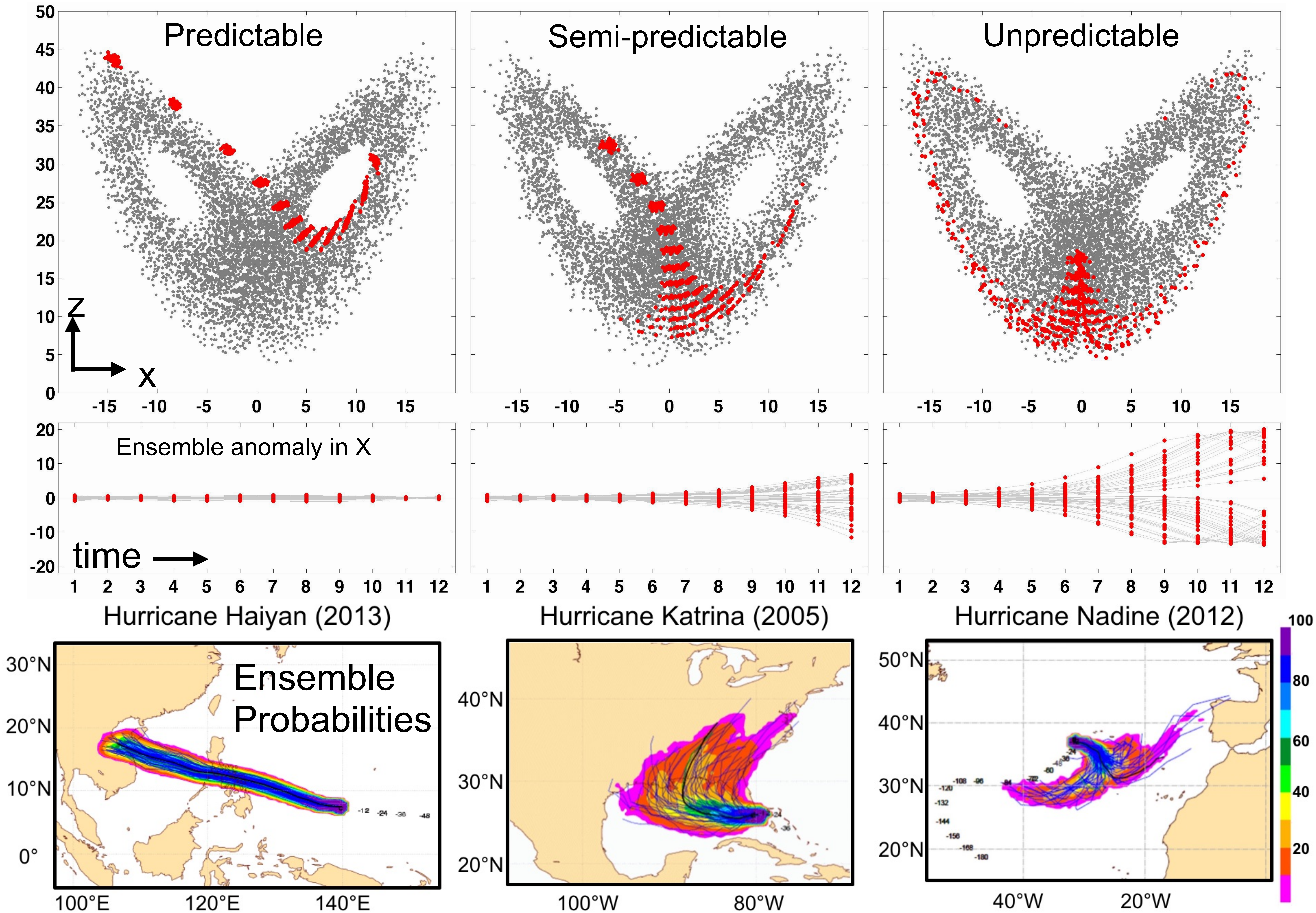
source: wikipedia

Ensemble Forecast with Initial Uncertainty



Ensemble Forecast with Initial Uncertainty




Predictable
Semi-predictable
Unpredictable
Z
X
Ensemble anomaly in X
time
Hurricane Haiyan (2013)
Hurricane Katrina (2005)
Hurricane Nadine (2012)
Ensemble Probabilities
30°N
20°N
10°N
0°
100°E
120°E
140°E
40°N
30°N
20°N
100°W
80°W
50°N
40°N
30°N
20°N
40°W
20°W
100
80
60
40
20

DISCUSSION QUESTION

Which of the following statements is true regarding initial condition error and model error in ensemble forecasting?

A) Initial condition error primarily affects long-term forecasts, while model error primarily affects short-term forecasts.

B) Initial condition error primarily affects short-term forecasts, while model error primarily affects long-term forecasts.

C) Both initial condition error and model error equally affect short-term and long-term forecasts.

D) Initial condition error only affects oceanic forecasts, while model error only affects atmospheric forecasts.

DISCUSSION QUESTION

Which of the following statements is true regarding initial condition error and model error in ensemble forecasting?

A) Initial condition error primarily affects long-term forecasts, while model error primarily affects short-term forecasts.

B) Initial condition error primarily affects short-term forecasts, while model error primarily affects long-term forecasts.

C) Both initial condition error and model error equally affect short-term and long-term forecasts.

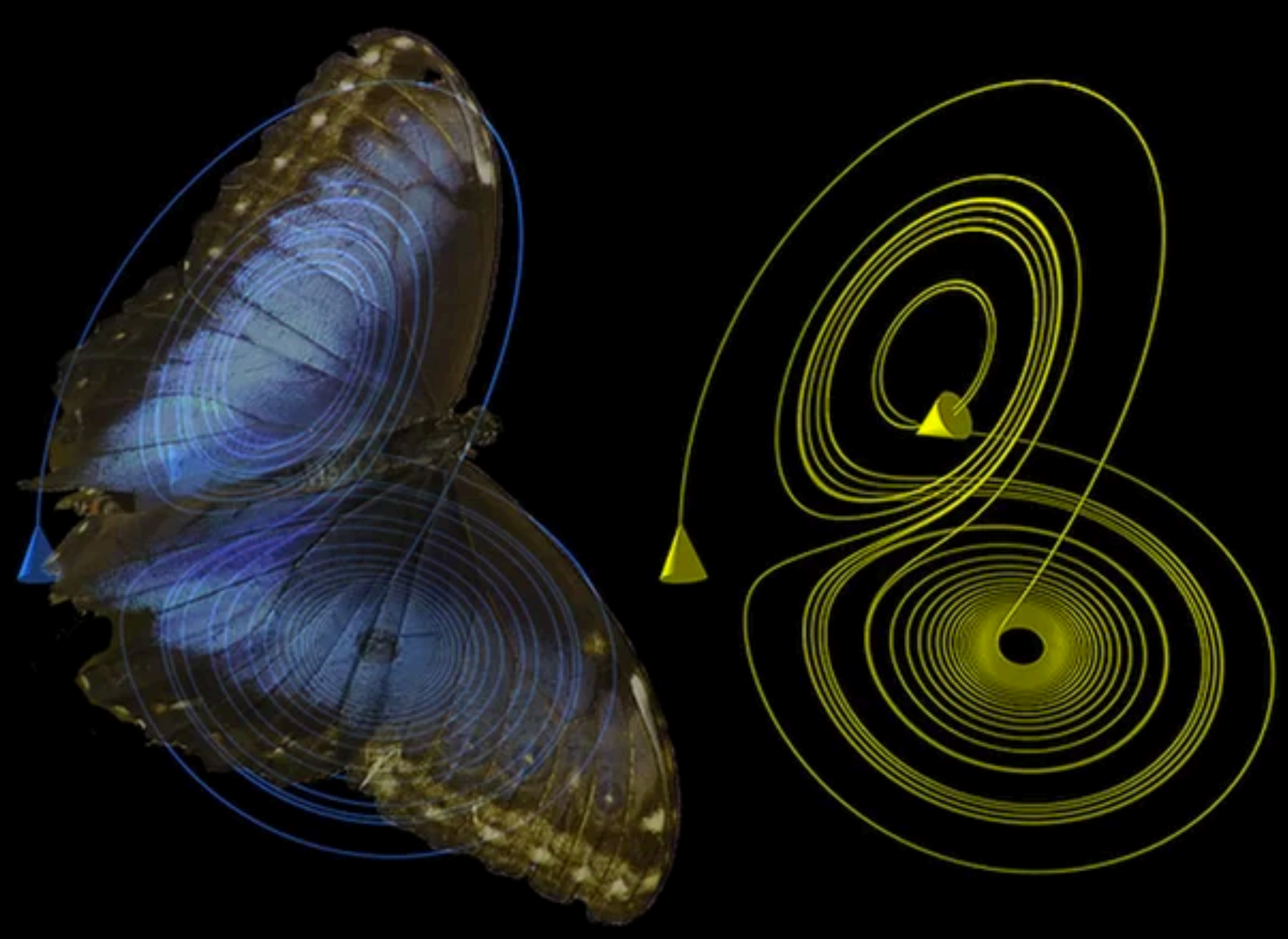
D) Initial condition error only affects oceanic forecasts, while model error only affects atmospheric forecasts.

Sources for uncertainty in earth system forecasts

Two main sources for uncertainty:

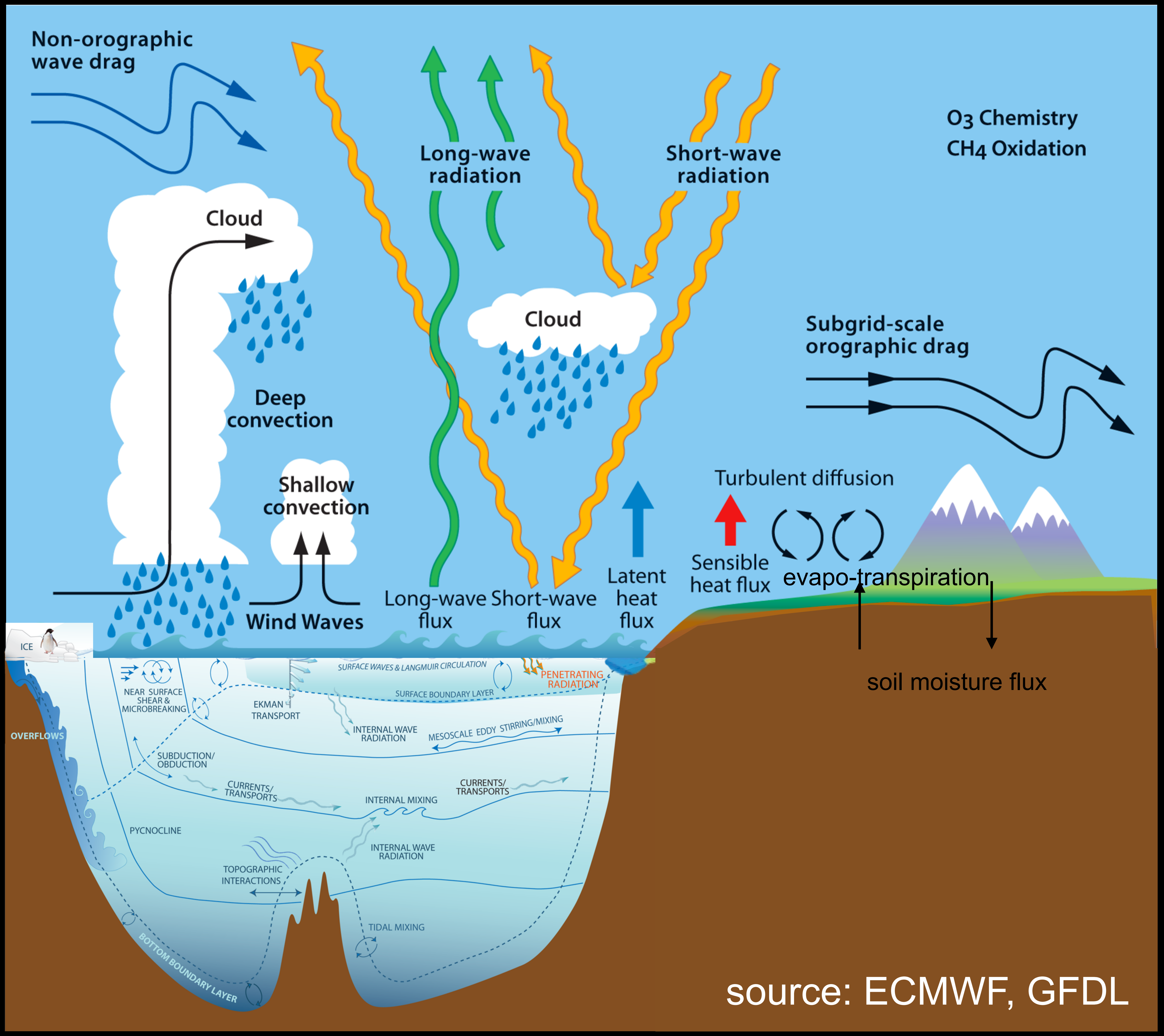
1. Initial condition uncertainty
2. Model error from sub-grid scale parameterization (from represented sub-grid scale processes and unrepresented physical processes)

Initial condition uncertainty



source: https://www.livescience.com/

Model uncertainty

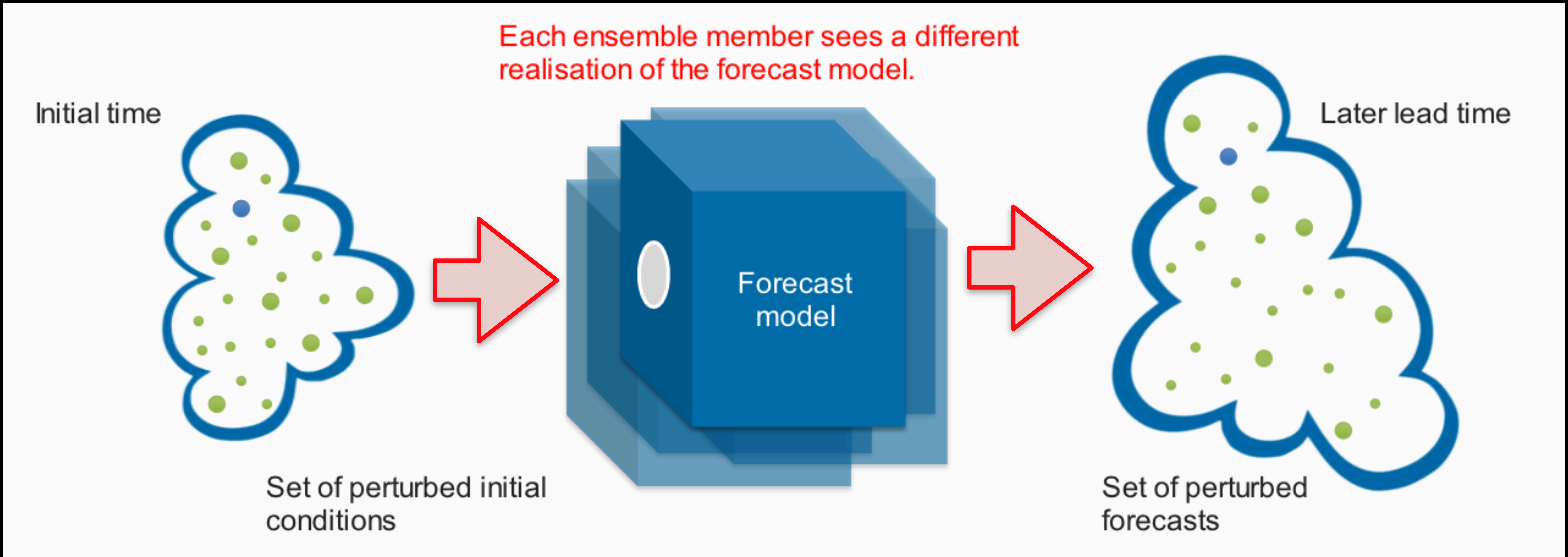


source: ECMWF, GFDL

Kalman filter equations

$$\tilde{\boldsymbol{x}}(t+\Delta t,-) = \boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$
$$\boldsymbol{P}(t+\Delta t,-) = \boldsymbol{A}\boldsymbol{P}(t)\boldsymbol{A}^T$$
$$\tilde{\boldsymbol{x}}(t+\Delta t) = \tilde{\boldsymbol{x}}(t+\Delta t,-) + \boldsymbol{K}(t+\Delta t)\left(\boldsymbol{y}(t+\Delta t) - \boldsymbol{E}\tilde{\boldsymbol{x}}(t+\Delta t,-)\right)$$
$$\boldsymbol{K}(t+\Delta t) = \boldsymbol{P}(t+\Delta t,-)\boldsymbol{E}^T\left(\boldsymbol{E}\boldsymbol{P}(t+\Delta t,-)\boldsymbol{E}^T + \boldsymbol{R}(t+\Delta t)\right)^{-1}$$
$$\boldsymbol{P}(t+\Delta t) = \boldsymbol{P}(t,-) - \boldsymbol{K}(t+\Delta t)\boldsymbol{E}\boldsymbol{P}(t,-)$$

Stochastic perturbations in an ensemble forecast model



Stochastic (random) perturbations added to the model tendencies.
This should act to increase the spread in the ensemble but not always!

What qualifies as a reliable ensemble forecast?

What qualifies as a reliable ensemble forecast?

Provides a realistic estimate of the forecast uncertainty

DISCUSSION QUESTION

In ensemble forecasting, what are the main implications of an underdispersive ensemble?

A) Overconfident predictions, underestimating variability and risk of extremes.

B) Accurate range, conservative predictions.

C) Equally reliable as overdispersive.

D) Broader range, less certain predictions.

DISCUSSION QUESTION

In ensemble forecasting, what are the main implications of an underdispersive ensemble?

A) Overconfident predictions, underestimating variability and risk of extremes.

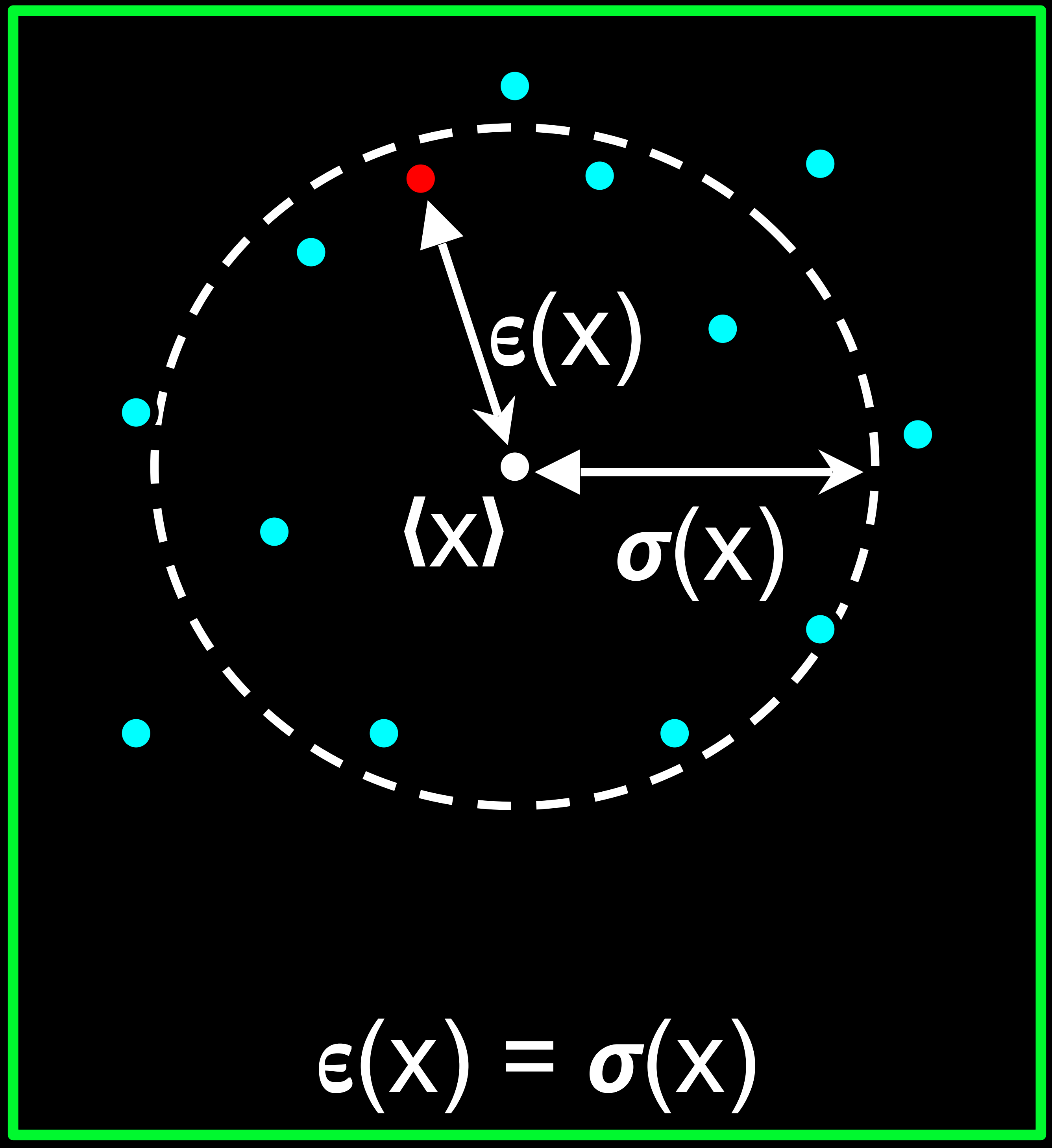
B) Accurate range, conservative predictions.

C) Equally reliable as overdispersive.

D) Broader range, less certain predictions.

Ensemble forecasting for uncertainty prediction

In a reliable ensemble, ensemble spread is a predictor of the ensemble error



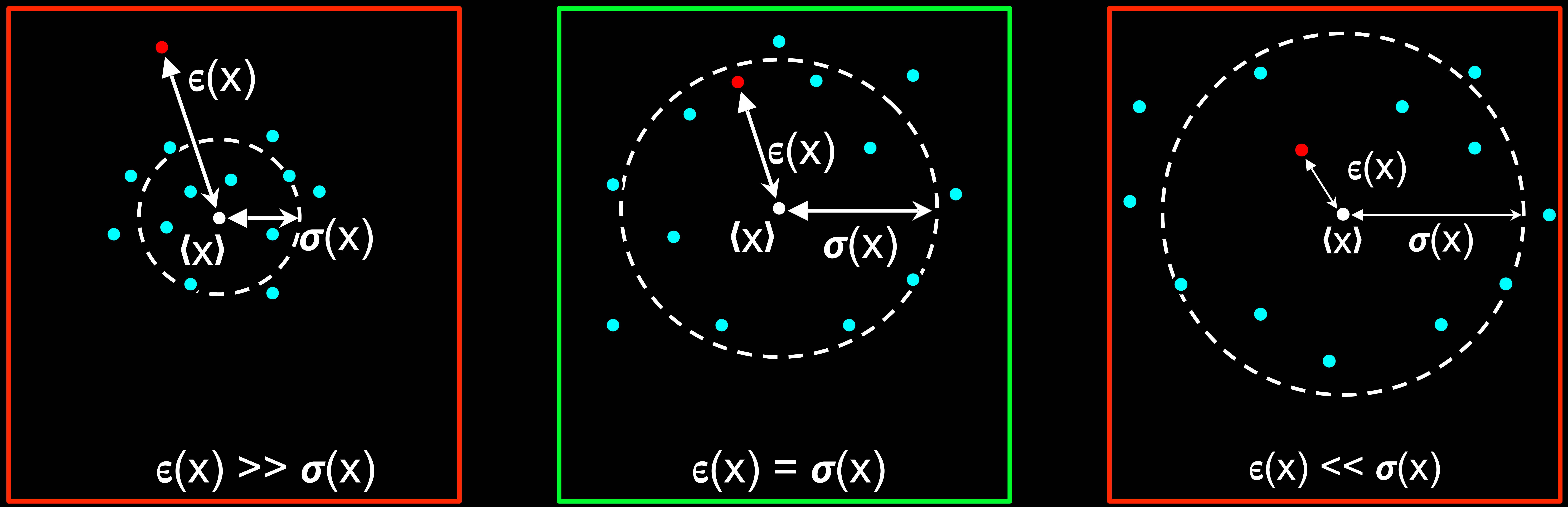
reliable ensemble

Averaged over many ensemble forecasts

- truth / observations
- ensemble member
- ensemble mean

Ensemble forecasting for uncertainty prediction

In a reliable ensemble, ensemble spread is a predictor of the ensemble error



under-dispersive ensemble | reliable ensemble | over-dispersive ensemble

Averaged over many ensemble forecasts

- truth / observations
- ensemble member
- ensemble mean

Ensemble Forecasting

Main two sources for uncertainty in model forecasts:

1. Initial condition uncertainty
2. Model error from sub-grid scale parameterization

The Ensemble Forecasting system (ECMWF) simulates the effect of :

- Initial condition uncertainty using Singular Vector + Ensemble Data Assimilation perturbations for the initial conditions
- Model uncertainties (2 stochastic schemes, SPPT and SKEB)

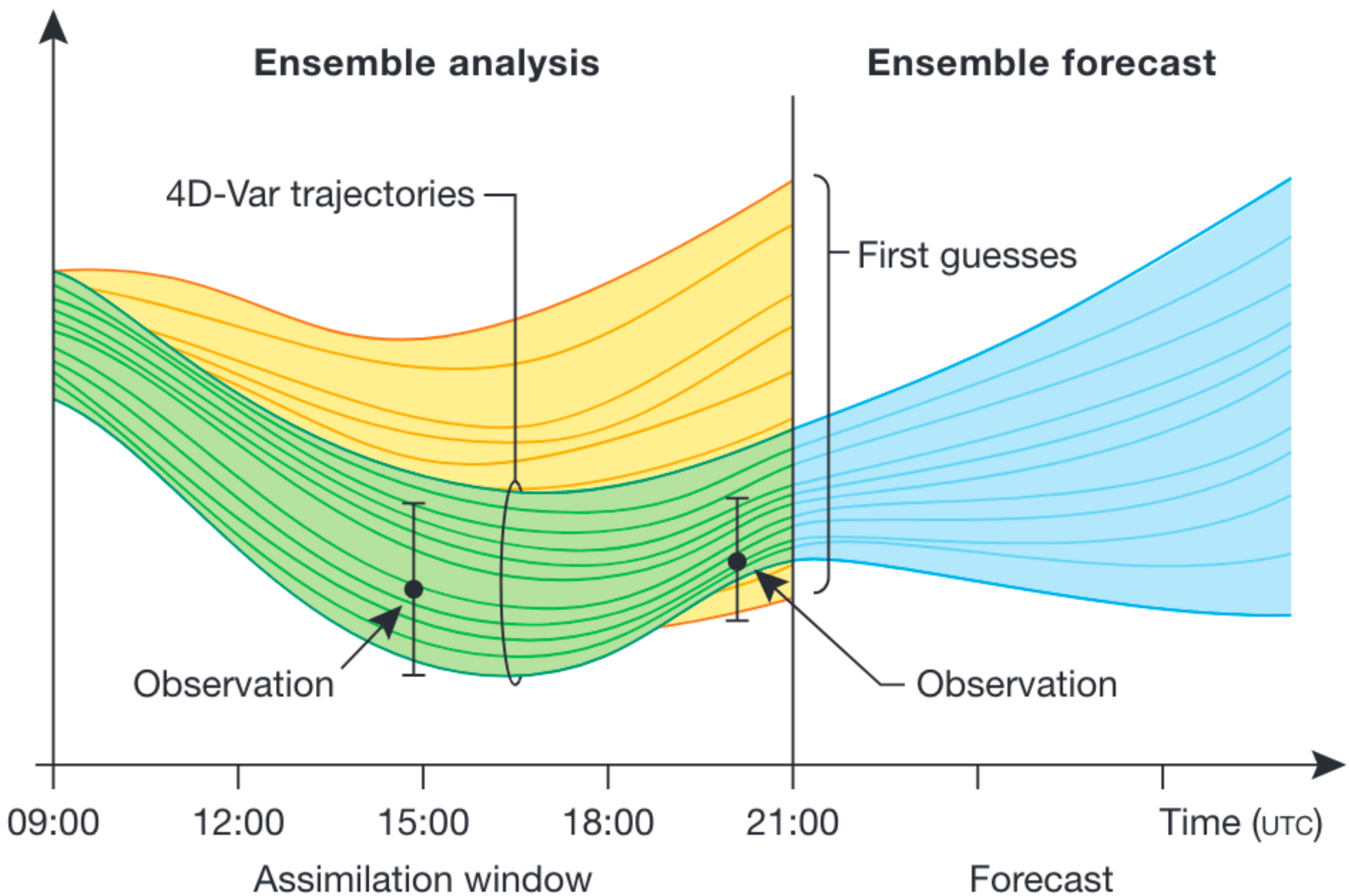
Stochastic parameterization has been successfully demonstrated in numerical weather prediction (e.g., Buizza et al., 1999; Shutts and Palmer, 2007; Palmer et al., 2009) and monthly to seasonal prediction (Weisheimer et al., 2011).

Berner et al. (2016) : A very good review of stochastic parameterization for weather and climate models

Kalman filter equations

$$\tilde{\boldsymbol{x}}(t+\Delta t,-) = \boldsymbol{A}\tilde{\boldsymbol{x}}(t)$$
$$\boldsymbol{P}(t+\Delta t,-) = \boldsymbol{A}\boldsymbol{P}(t)\boldsymbol{A}^T$$
$$\tilde{\boldsymbol{x}}(t+\Delta t) = \tilde{\boldsymbol{x}}(t+\Delta t,-) + \boldsymbol{K}(t+\Delta t)\left(\boldsymbol{y}(t+\Delta t) - \boldsymbol{E}\tilde{\boldsymbol{x}}(t+\Delta t,-)\right)$$
$$\boldsymbol{K}(t+\Delta t) = \boldsymbol{P}(t+\Delta t,-)\boldsymbol{E}^T\left(\boldsymbol{E}\boldsymbol{P}(t+\Delta t,-)\boldsymbol{E}^T + \boldsymbol{R}(t+\Delta t)\right)^{-1}$$
$$\boldsymbol{P}(t+\Delta t) = \boldsymbol{P}(t,-) - \boldsymbol{K}(t+\Delta t)\boldsymbol{E}\boldsymbol{P}(t,-)$$

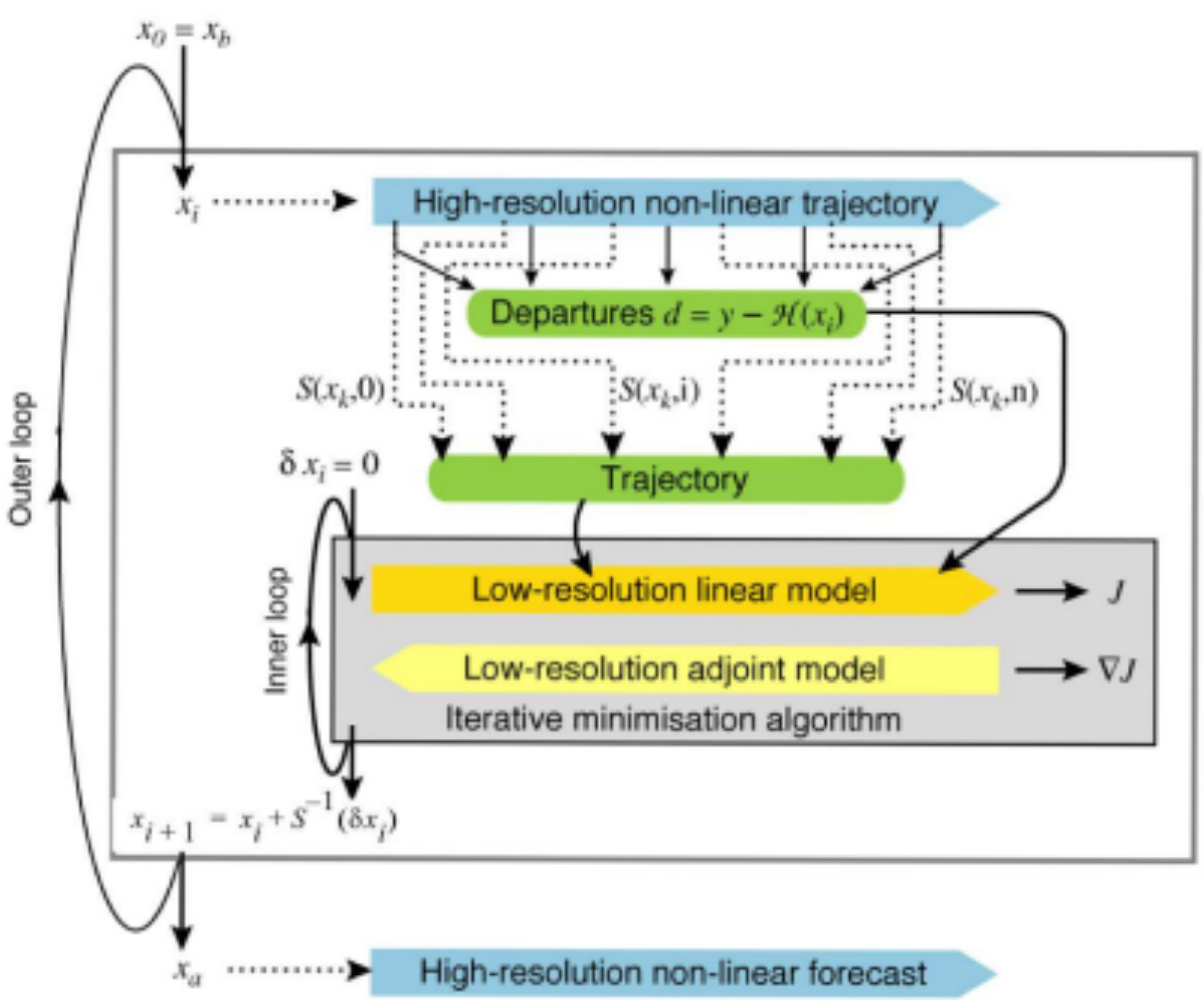
DA in Operational Forecast Models



Traditional DA Methods are Computationally Expensive

Bottlenecks may include:

- The need for multiple model runs
- Model space to observation space mapping
- Inversion of very large matrices
- Computing gradients in very large dimensional spaces

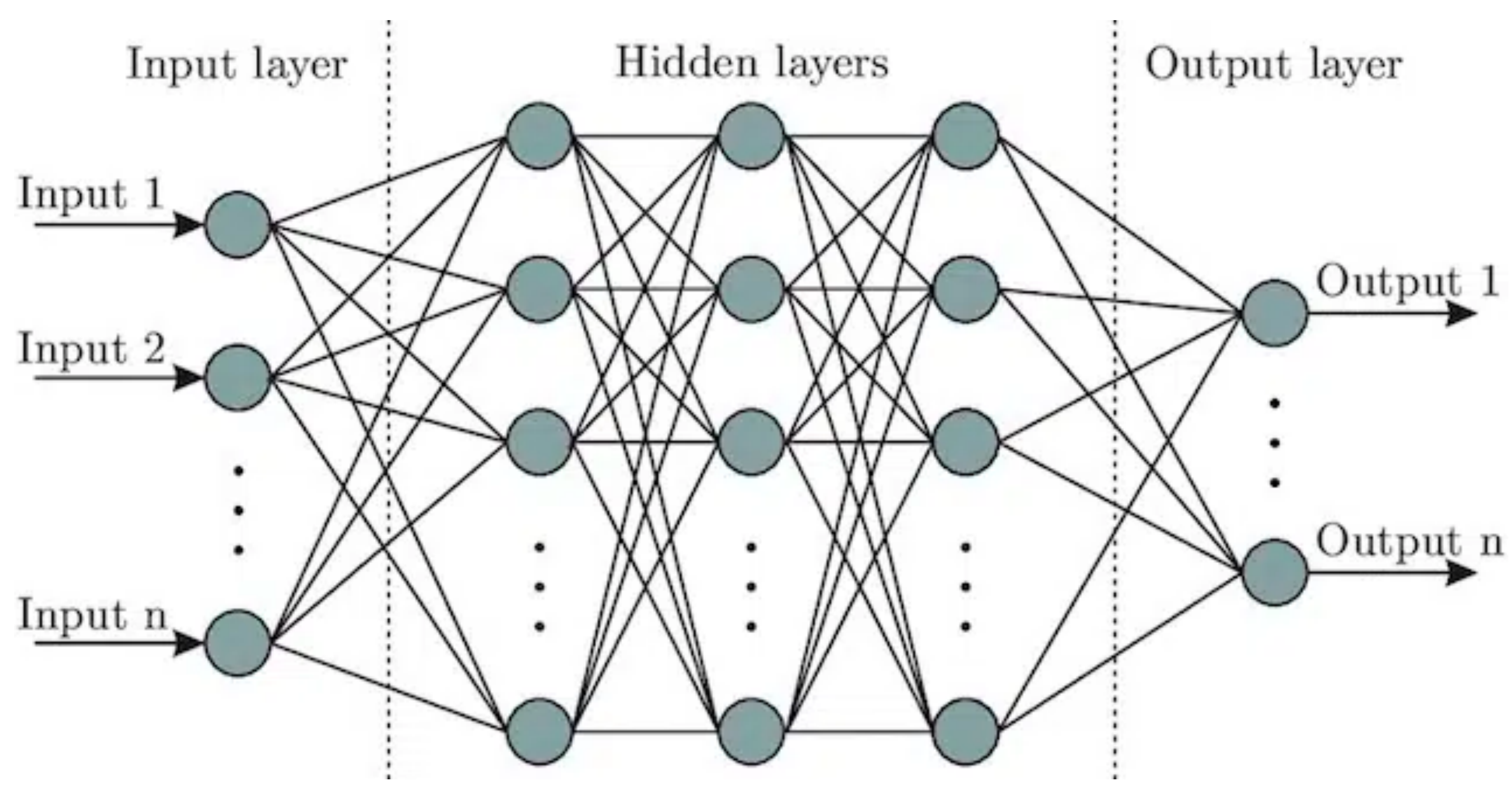


ECMWF

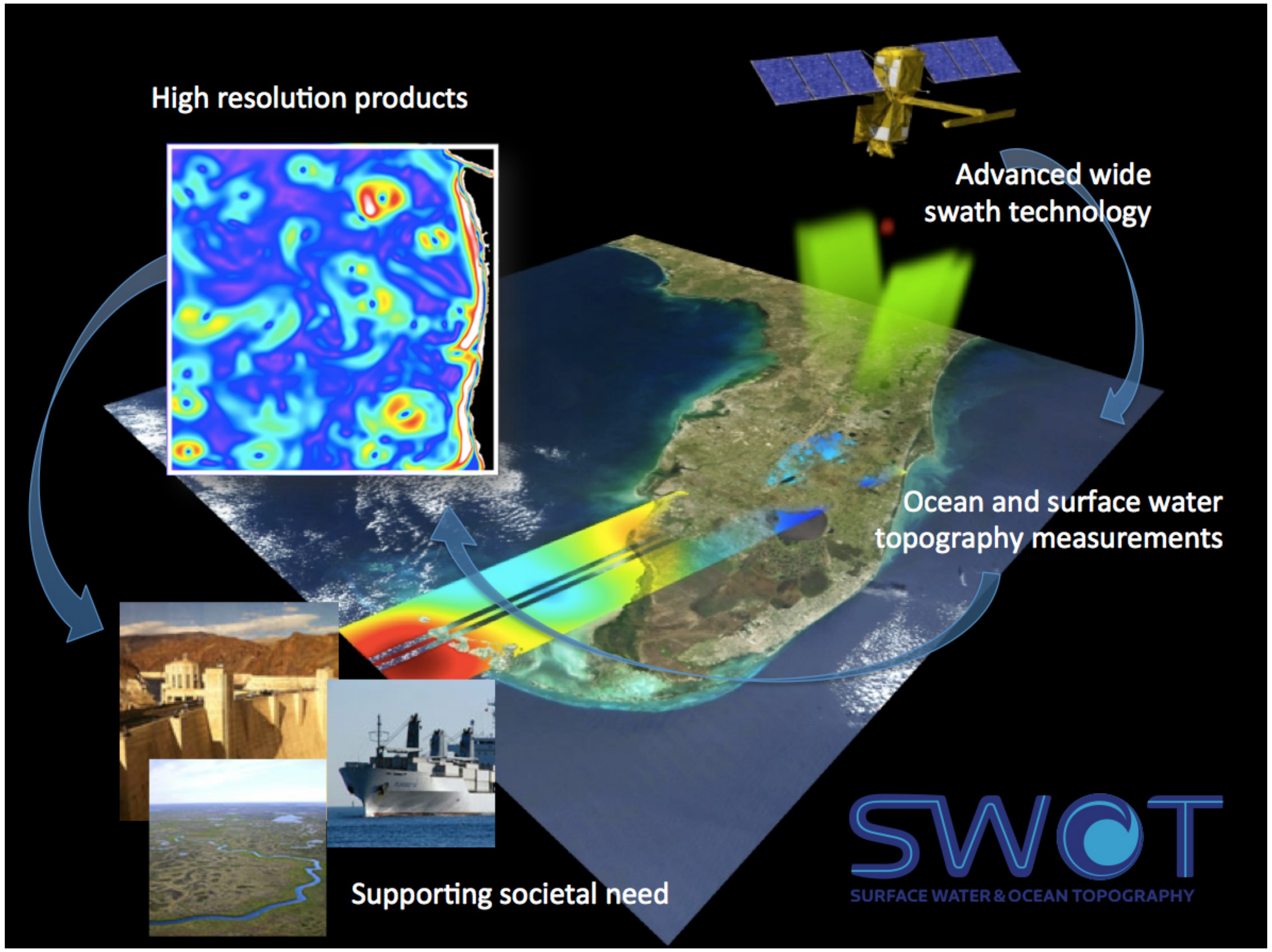
High Resolution Observations are Underutilized

- Large quantities of observations are not assimilated
- Incorporating these data into forecast models has the potential to significantly improve skill

Machine learning may be able to help



https://searchengineland.com/machine-learning-search-terms-concepts-algorithms-383913



NASA JPL

Machine Learning Has Been Combined with DA

- Model surrogates (Bocquet 2023; Brajard et al., 2020)
- Observation operator (Liang et al., 2023; Jing et al., 2019)
- Variable resolution models (Barthélémy et al., 2022)
- Bias correction (Chapman et al., 2019; 2022)

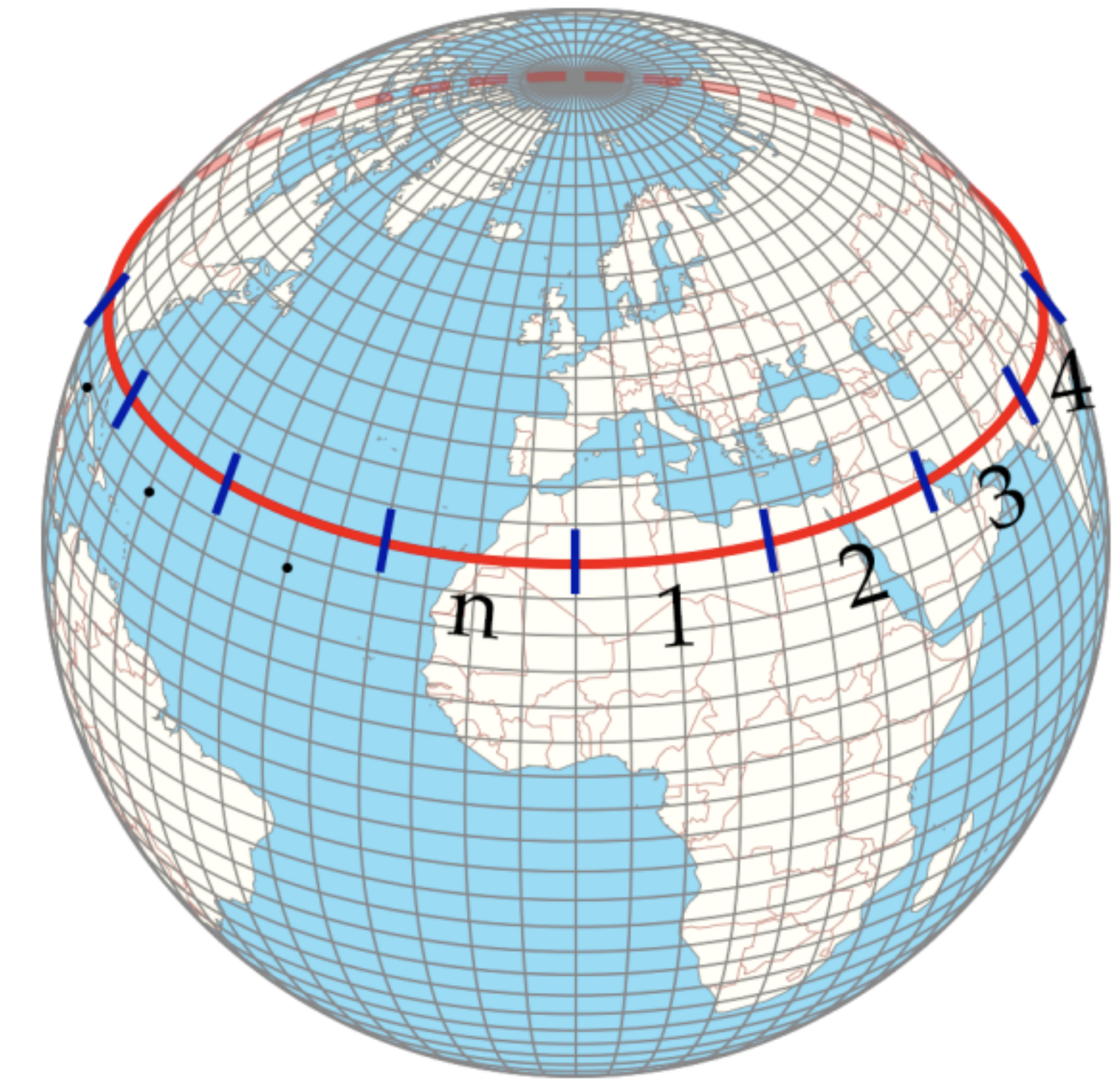
Limited work attempting to perform assimilation directly with ML

Research Questions

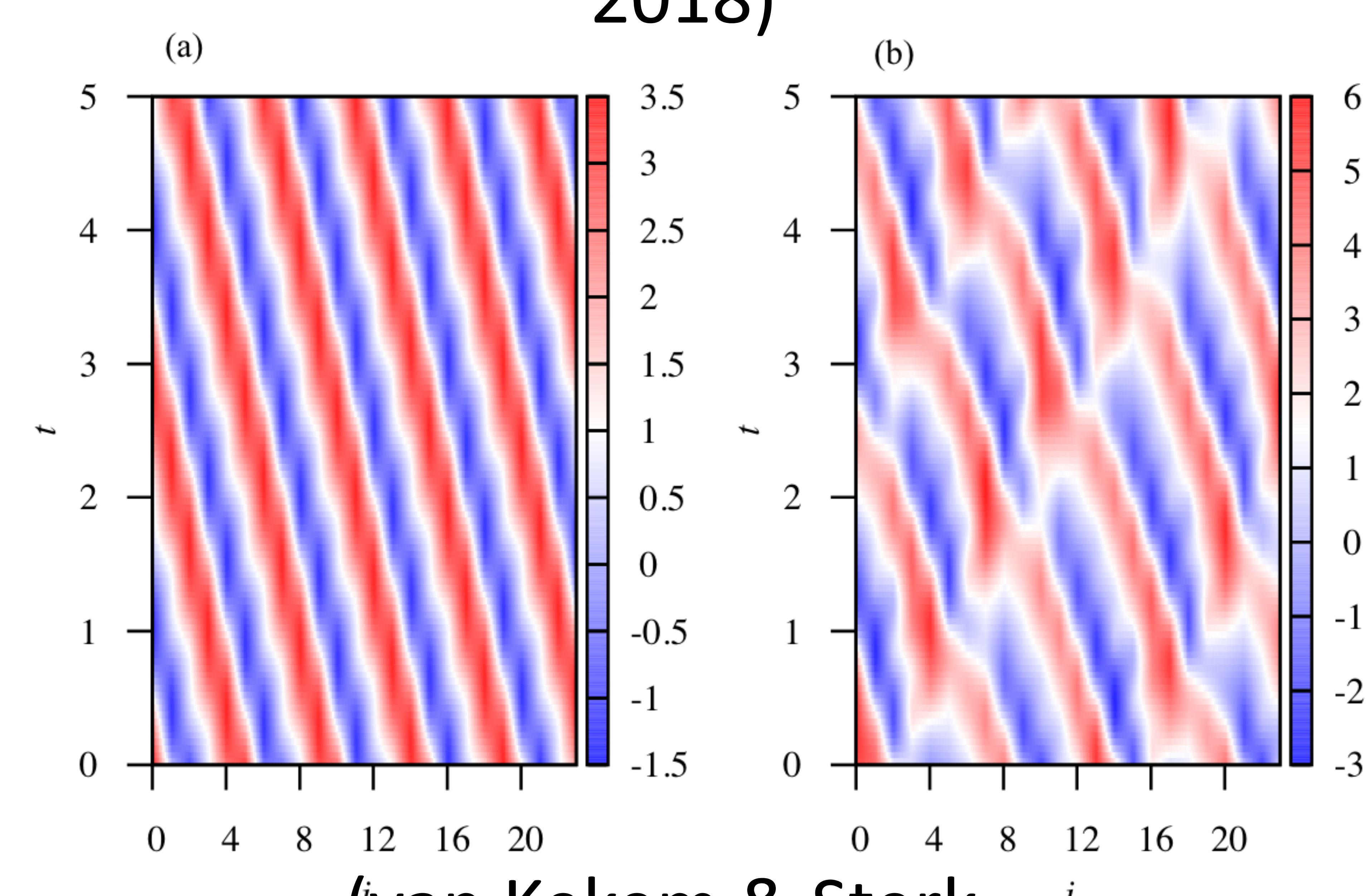
1. Can a simple CNN be trained to emulate a traditional DA method successfully?
2. Can an augmented method, using the trained CNN only for high resolution observations, outperform a traditional method in which high resolution observations are ignored?

Lorenz-96 as a Test System

- Set of N discrete differential equations
- Analogous to a single state variable zonally
- Chaotic for some choices of forcing parameter
- Used for testing DA methods



(van Kekem, 2018)

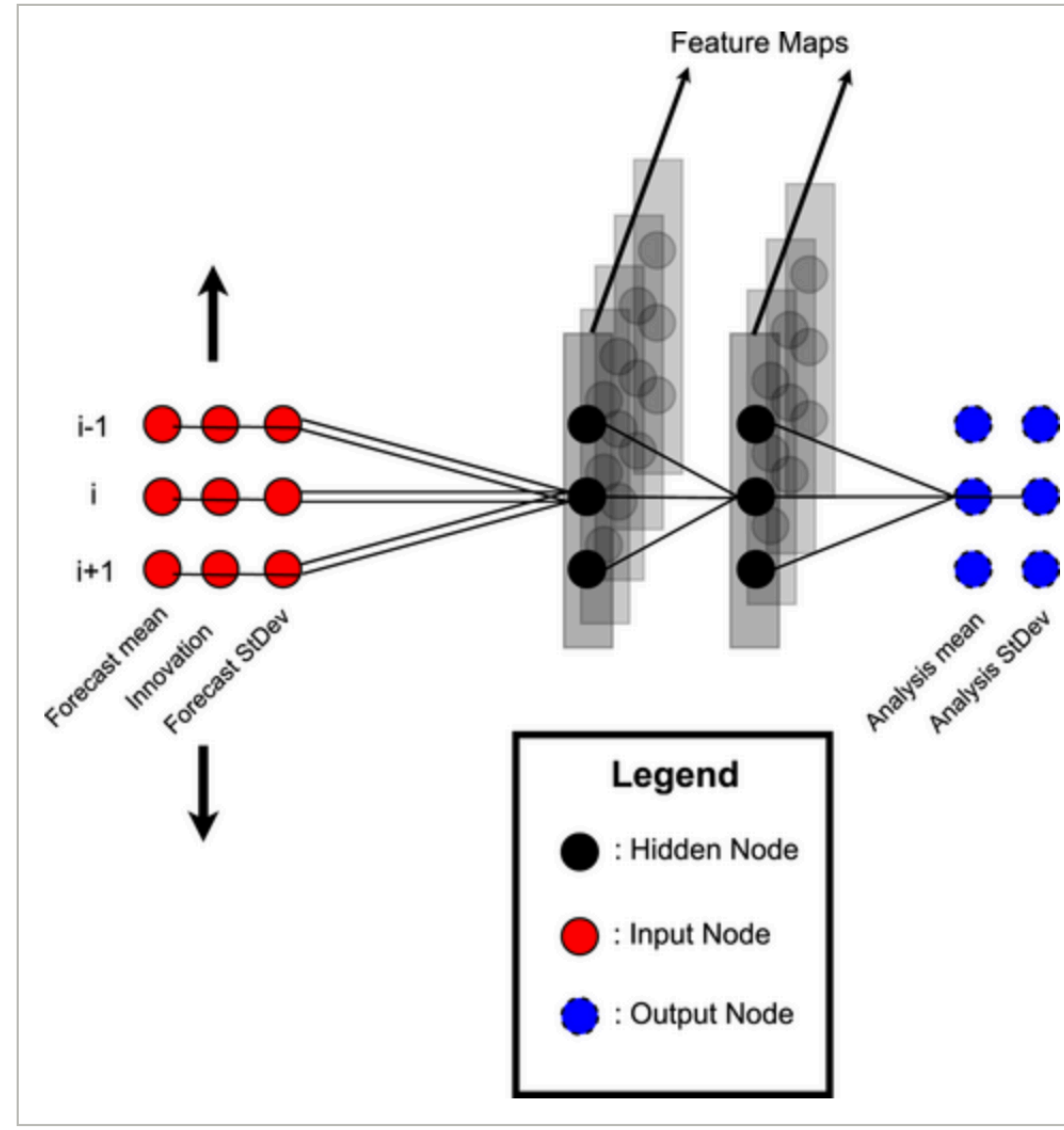


(van Kekem & Sterk, 2018)

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F$$

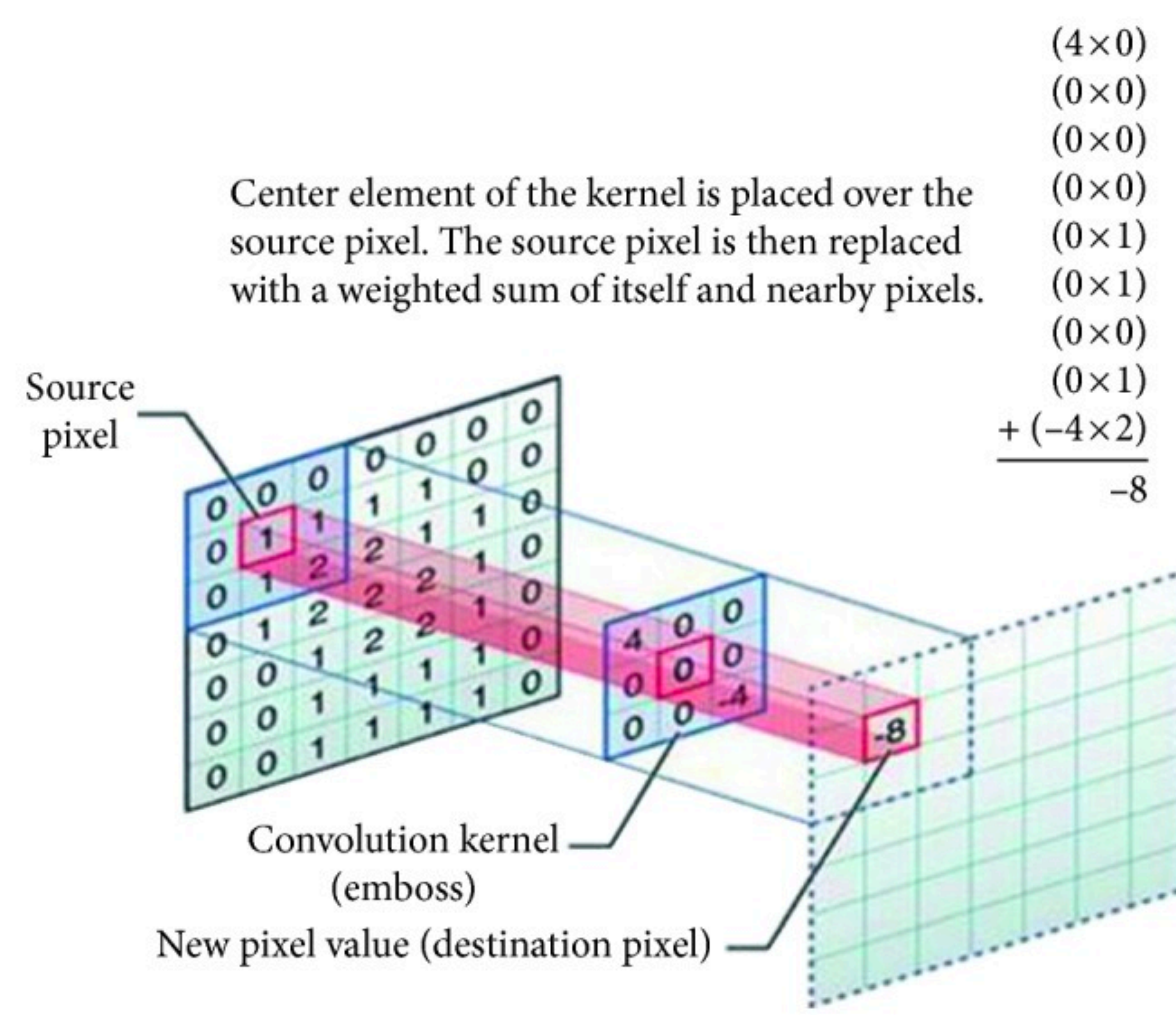
Neural Network Assimilation

- Input:
  - Forecast mean
  - Forecast standard deviation
  - Innovation
- Output:
  - Analysis mean
  - Analysis standard deviation
- Train to replicate EnKF analysis ensemble statistics



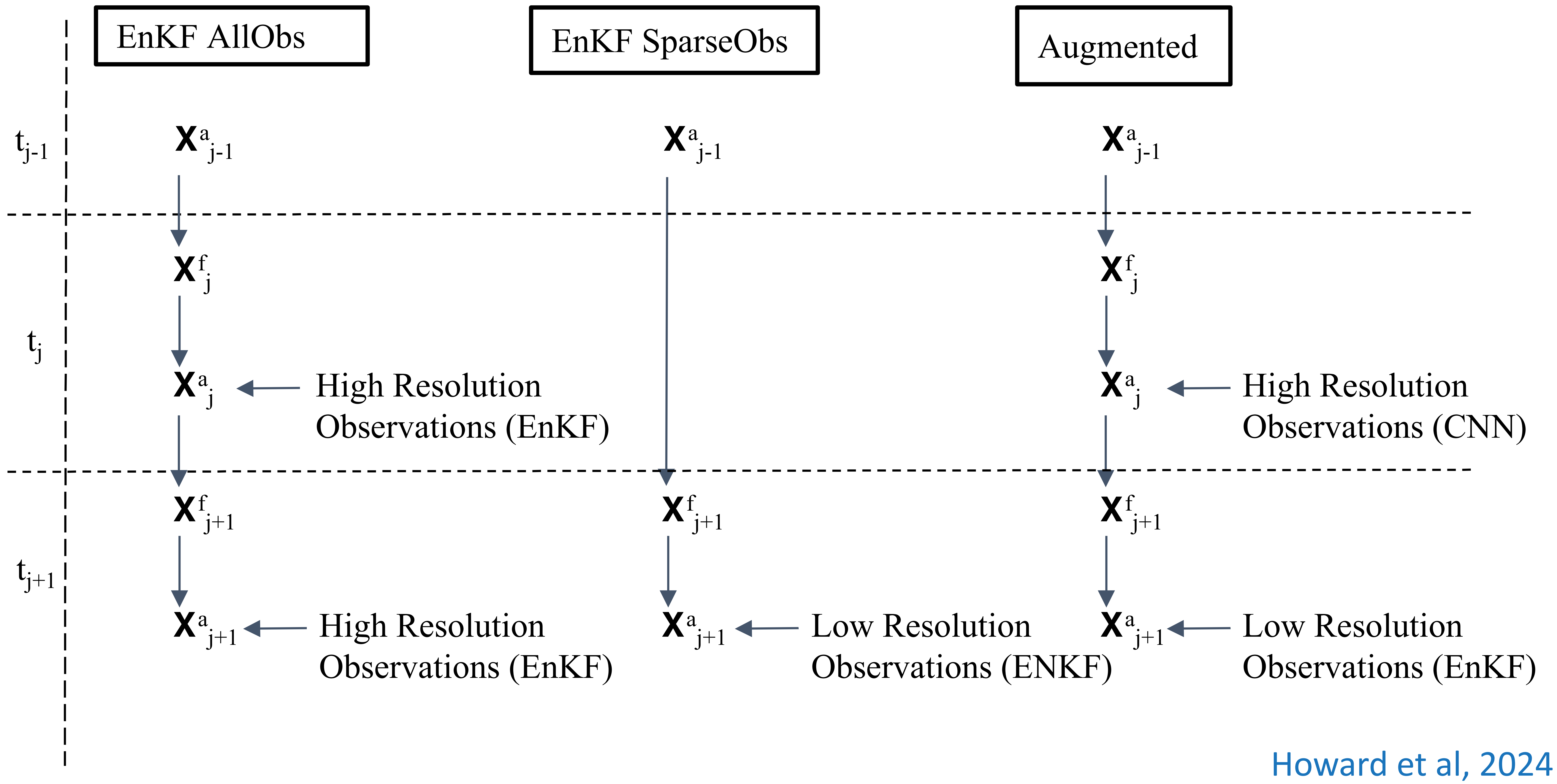
Howard et al, 2024

Input



Output

Machine Learning Augmented Method



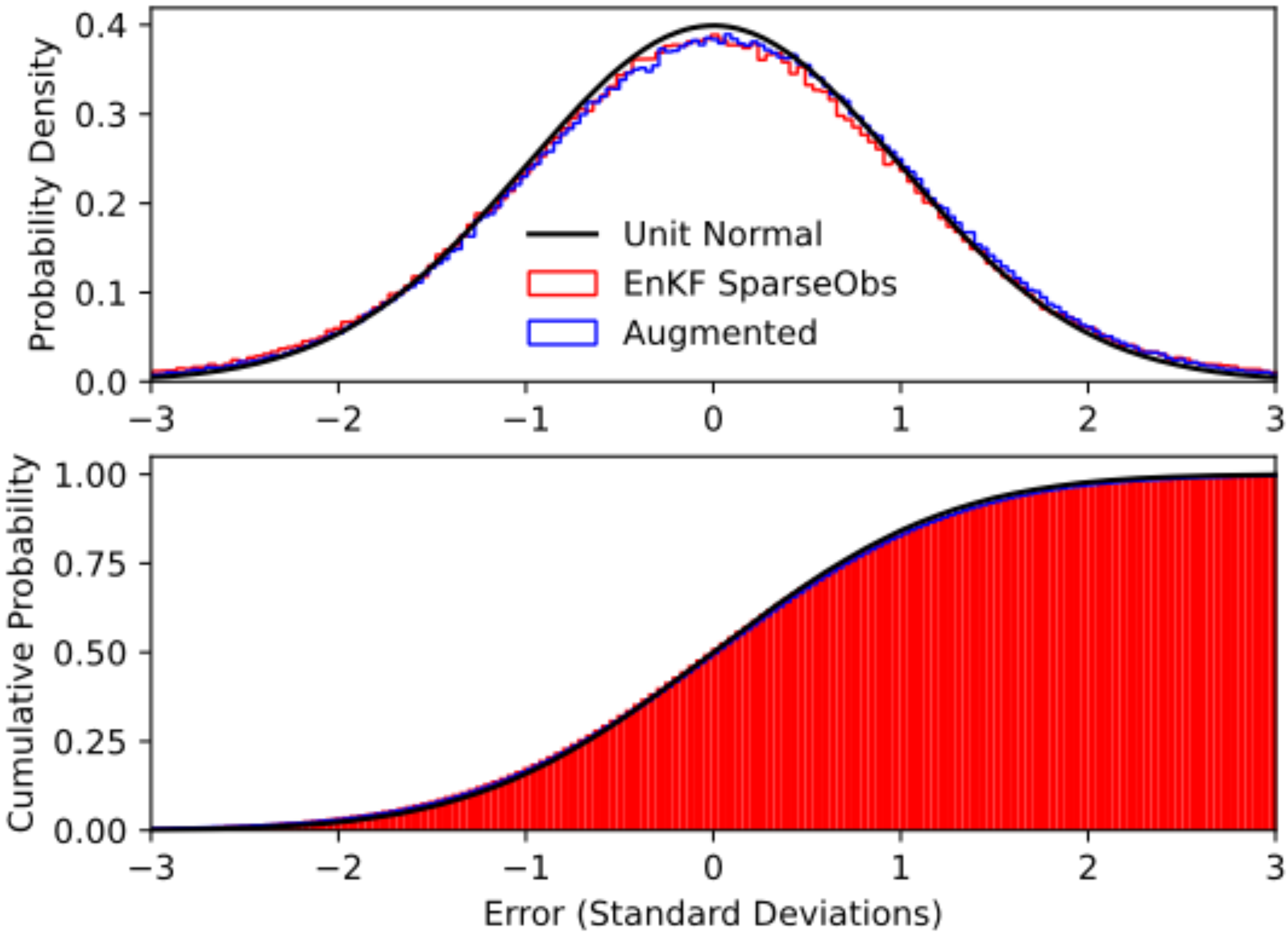
Howard et al, 2024

- EnKF AllObs
  - Assimilate all synthetic observations
- EnKF SparseObs
  - Assimilate spatially and temporally thinned observations with the EnKF
- Augmented
  - Assimilated spatially and temporally thinned observations with EnKF
  - Assimilate temporally thinned, spatially dense observations with trained neural network

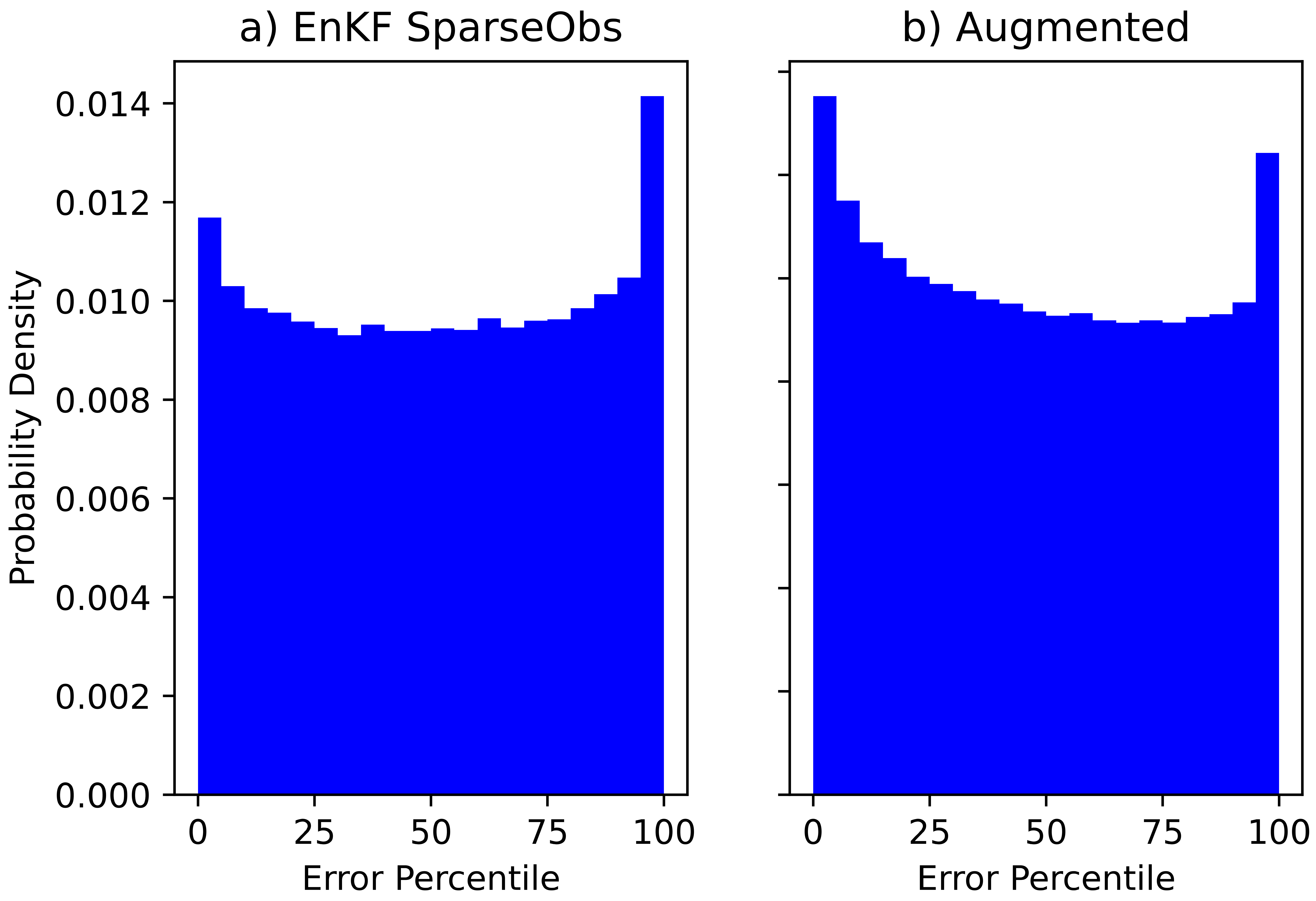
Augmented Method Produces Improved RMSE

Howard et al, 2024

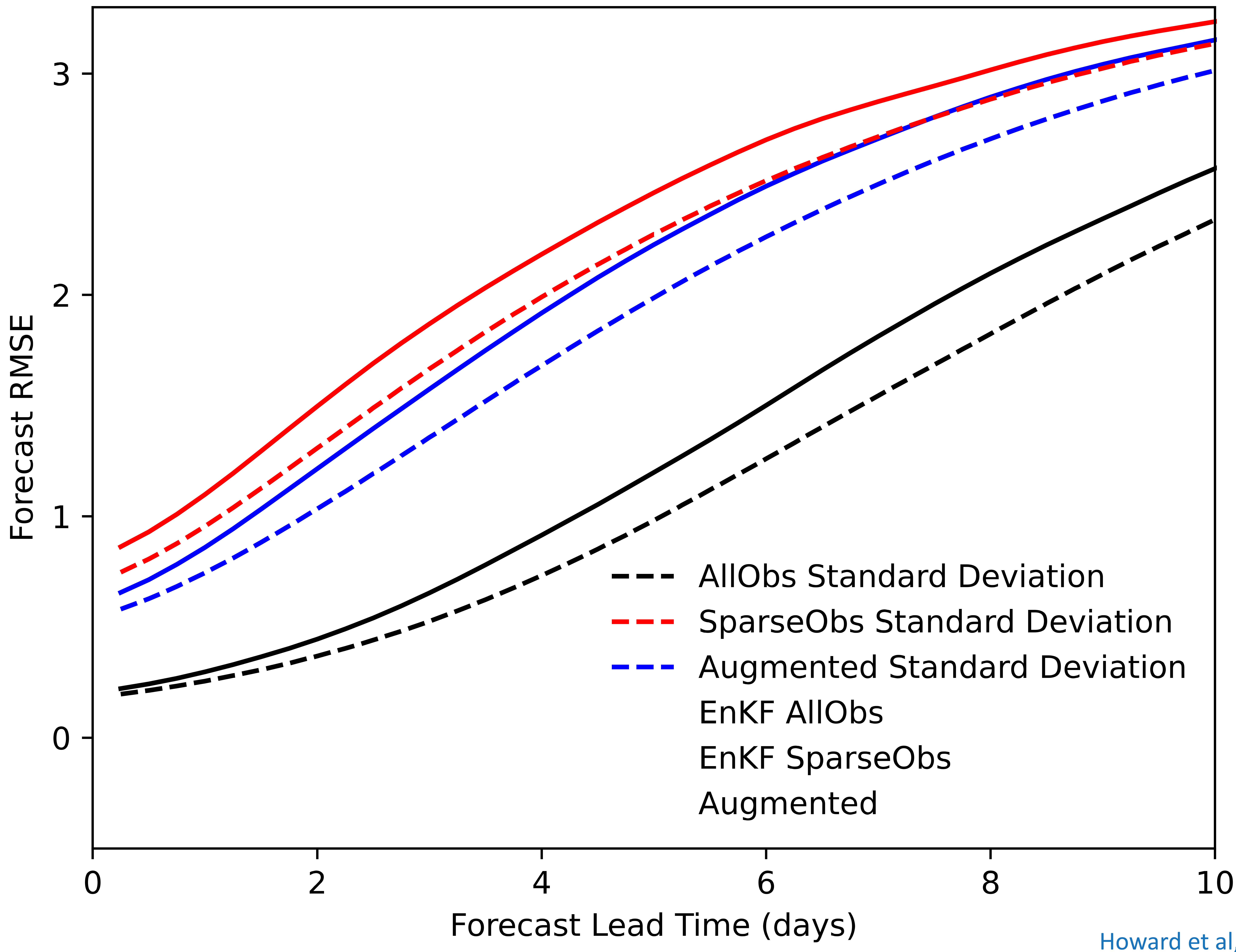
Error Distribution



Reliability of Uncertainty Estimation



Forecasts Initialized with Augmented Method are More Accurate



Howard et al, 2024

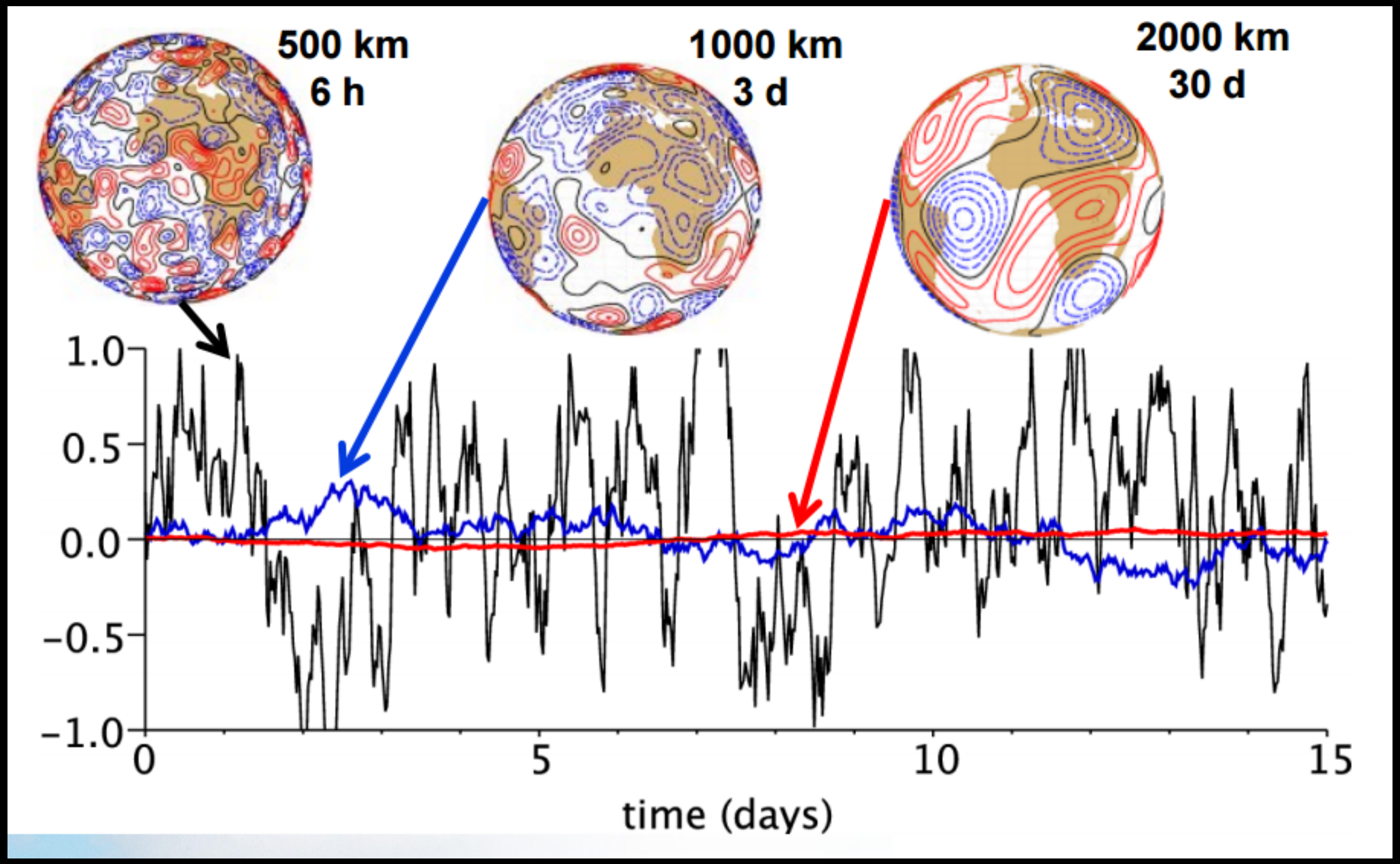
TAKE HOME MESSAGE

Using a trained neural network to assimilate high-resolution data improves forecast accuracy in a synthetic system

Howard, L. J., Subramanian, A., & Hoteit, I. (2024). A machine learning augmented data assimilation method for high-resolution observations. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS003774. https://doi.org/10.1029/2023MS003774

Stochastic parameterization to represent model error

Stochastically Perturbed Parameterization Tendencies (SPPT)



- The net parameterized physics tendency:

$$\boldsymbol{X} = X_U, X_V, X_T, X_Q$$

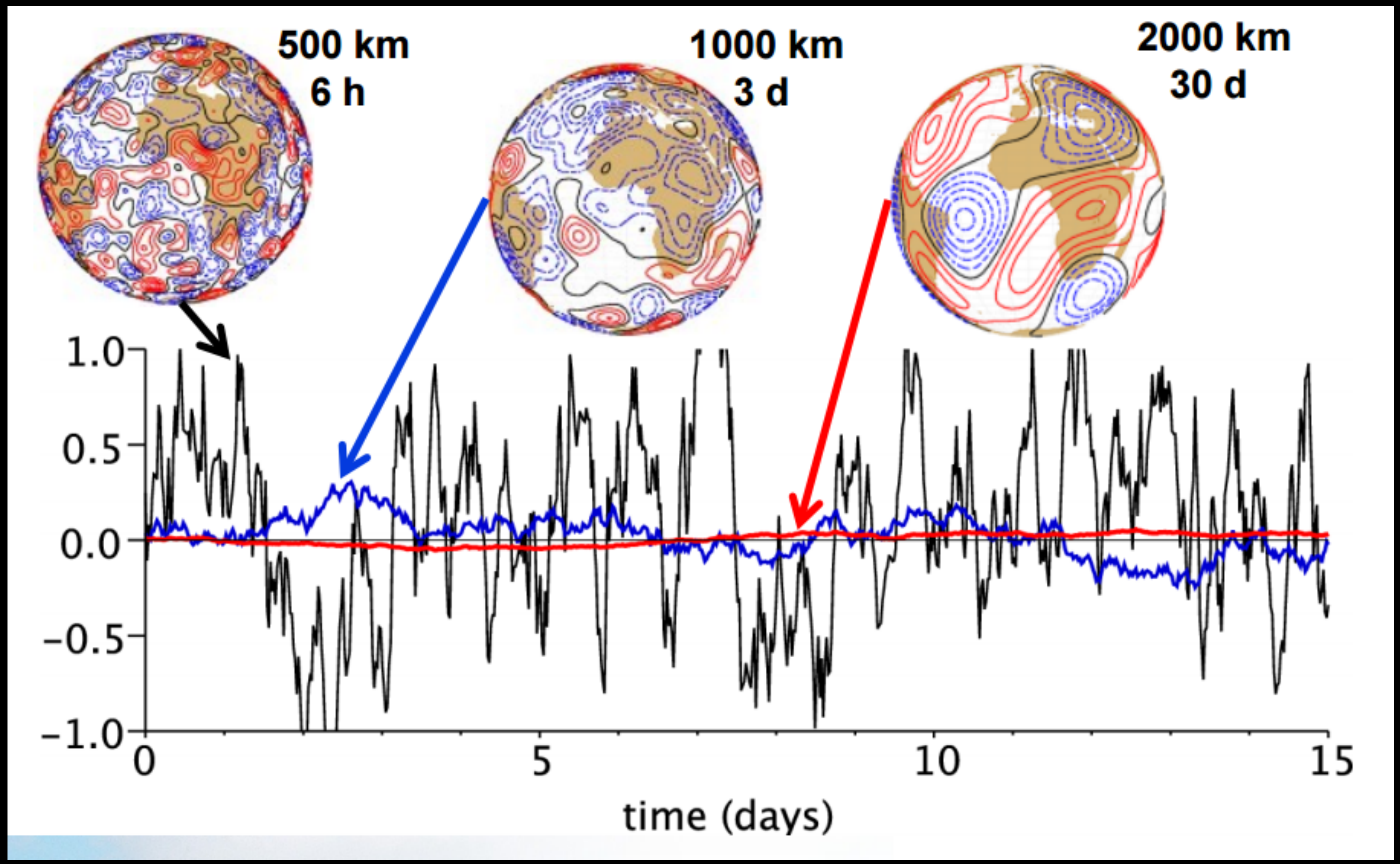
coming from : (radiation schemes, gravity wave drag, vertical mixing, convection, cloud physics)

- Perturbed with multiplicative noise $\boldsymbol{X}' = (1 + \mu r)\boldsymbol{X}$

Stochastic Kinetic Energy Backscatter (SKEB)

Stochastic parameterization to represent model error

Stochastically Perturbed Parameterization Tendencies (SPPT)



- The net parameterized physics tendency:

  $\boldsymbol{X} = X_U, X_V, X_T, X_Q$

coming from : (radiation schemes, gravity wave drag, vertical mixing, convection, cloud physics)

- Perturbed with multiplicative noise $\boldsymbol{X}' = (1 + \mu r)\boldsymbol{X}$

Stochastic Kinetic Energy Backscatter (SKEB)

- Simulates a missing and uncertain process
- Parameterizes upscale transfer of energy from sub-grid scales

Shutts and Palmer 2004, Shutts 2005, Berner et al. 2009

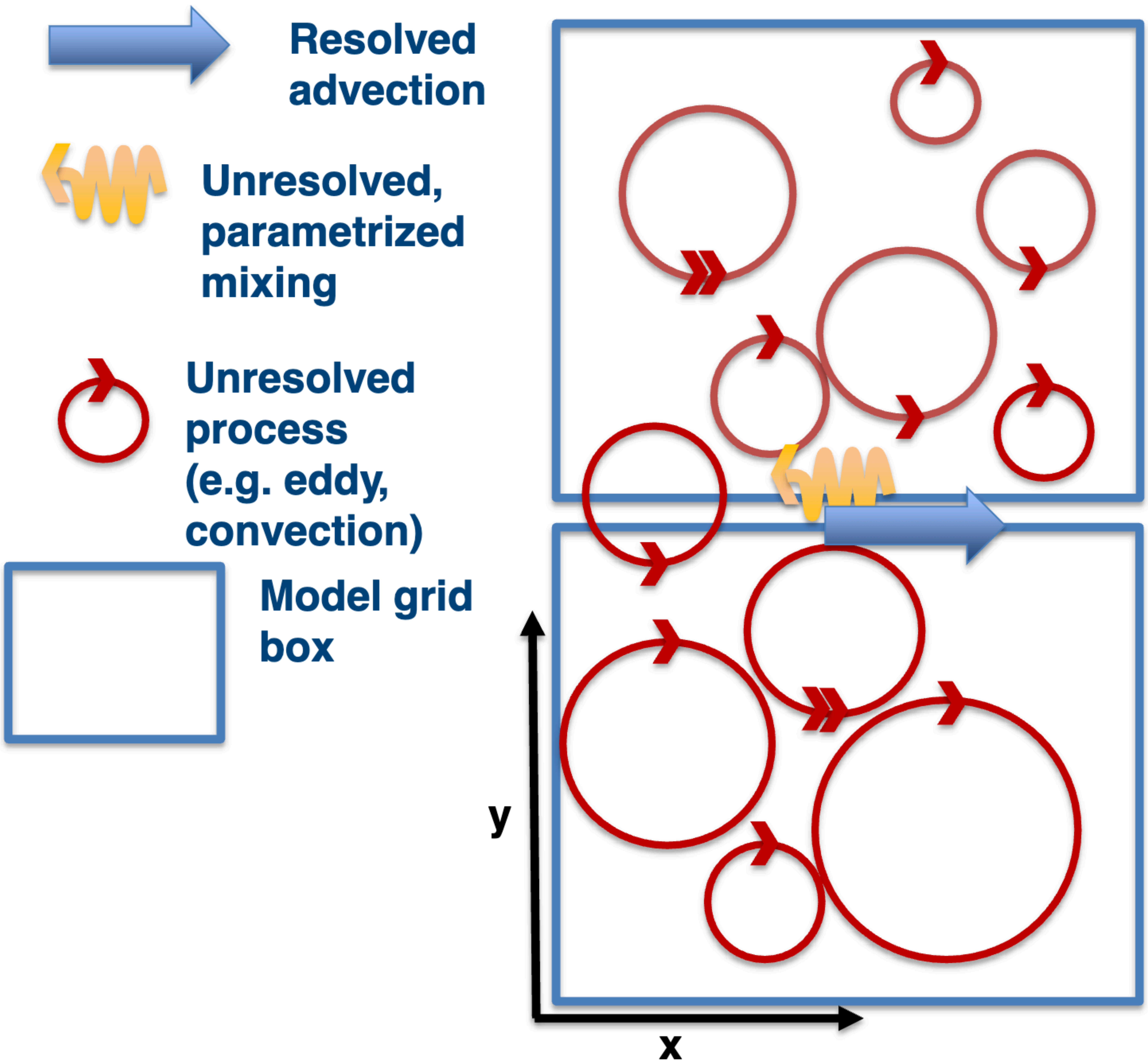
$$F_\phi = \left(\frac{b_R D_{tot}}{B_{tot}}\right)^{1/2} F^*$$

F^* is the 3D random pattern

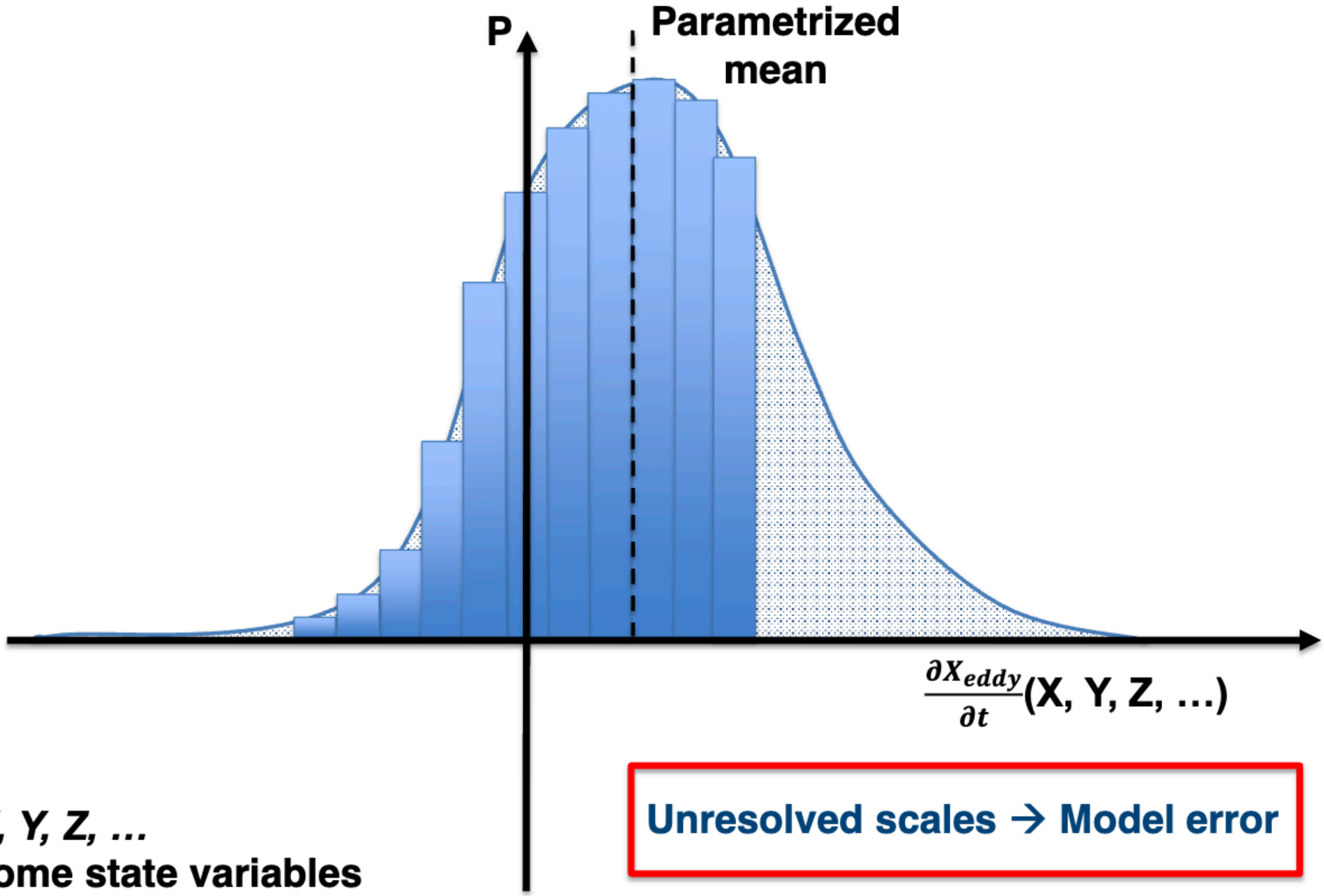
B_{tot} is the mean KE input by convective updrafts

D_{tot} is the dissipation rate

Subgrid-scale Ocean Processes



Subgrid-scale Ocean Processes



Unresolved scales → Model error

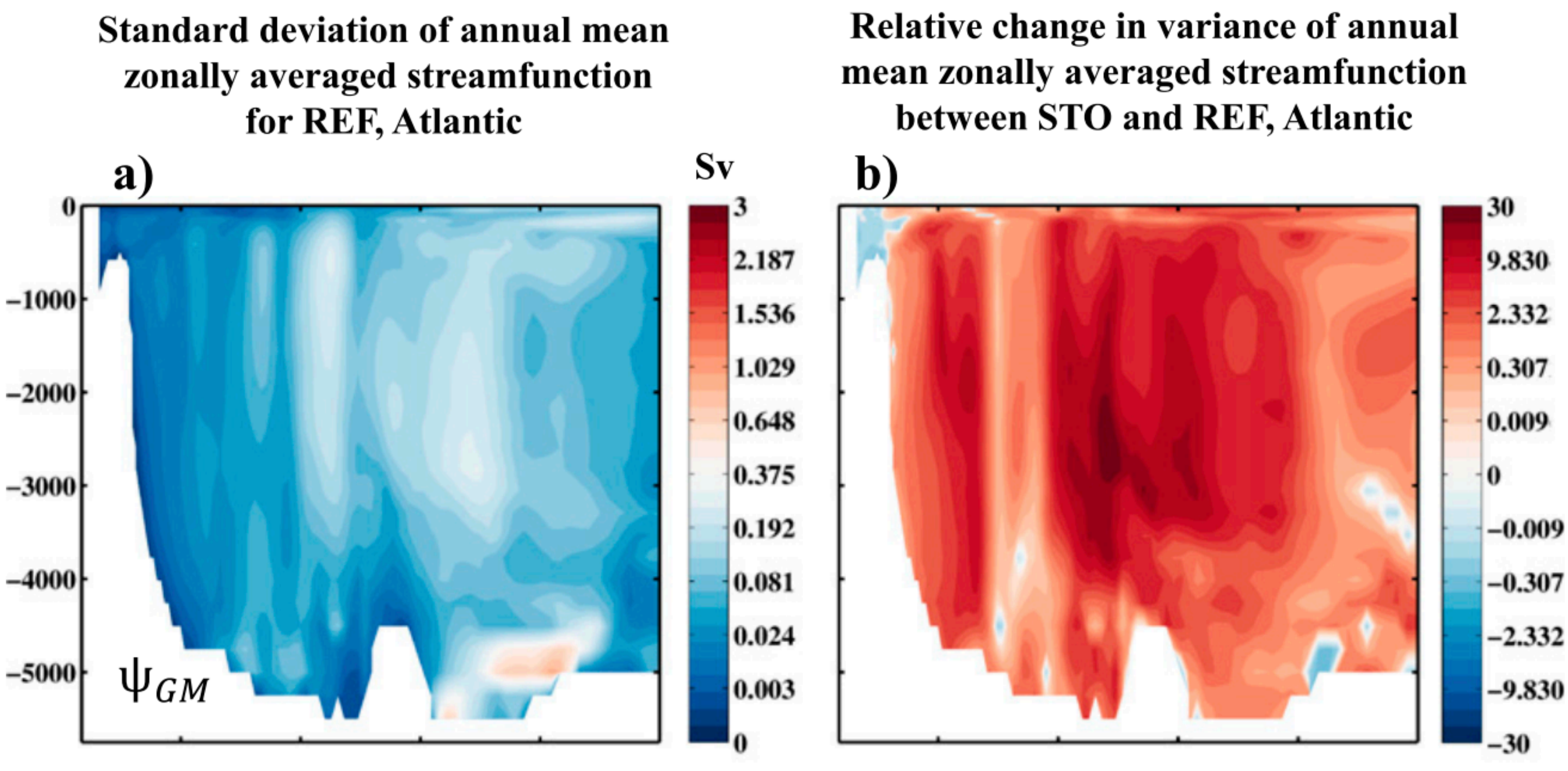
X, Y, Z, …
some state variables

Stochastic Subgrid-Scale Ocean Mixing: Impacts on Low-Frequency Variability

STEPHAN JURICKE, TIM N. PALMER, AND LAURE ZANNA
Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, United Kingdom

Stochastic perturbations :

$$P^{\mathrm{sto}}(i,j) = [1+\xi(i,j)]P^{\mathrm{ref}}(i,j),$$



Stochastic perturbations to GM scheme strengthens inter annual variability of AMOC

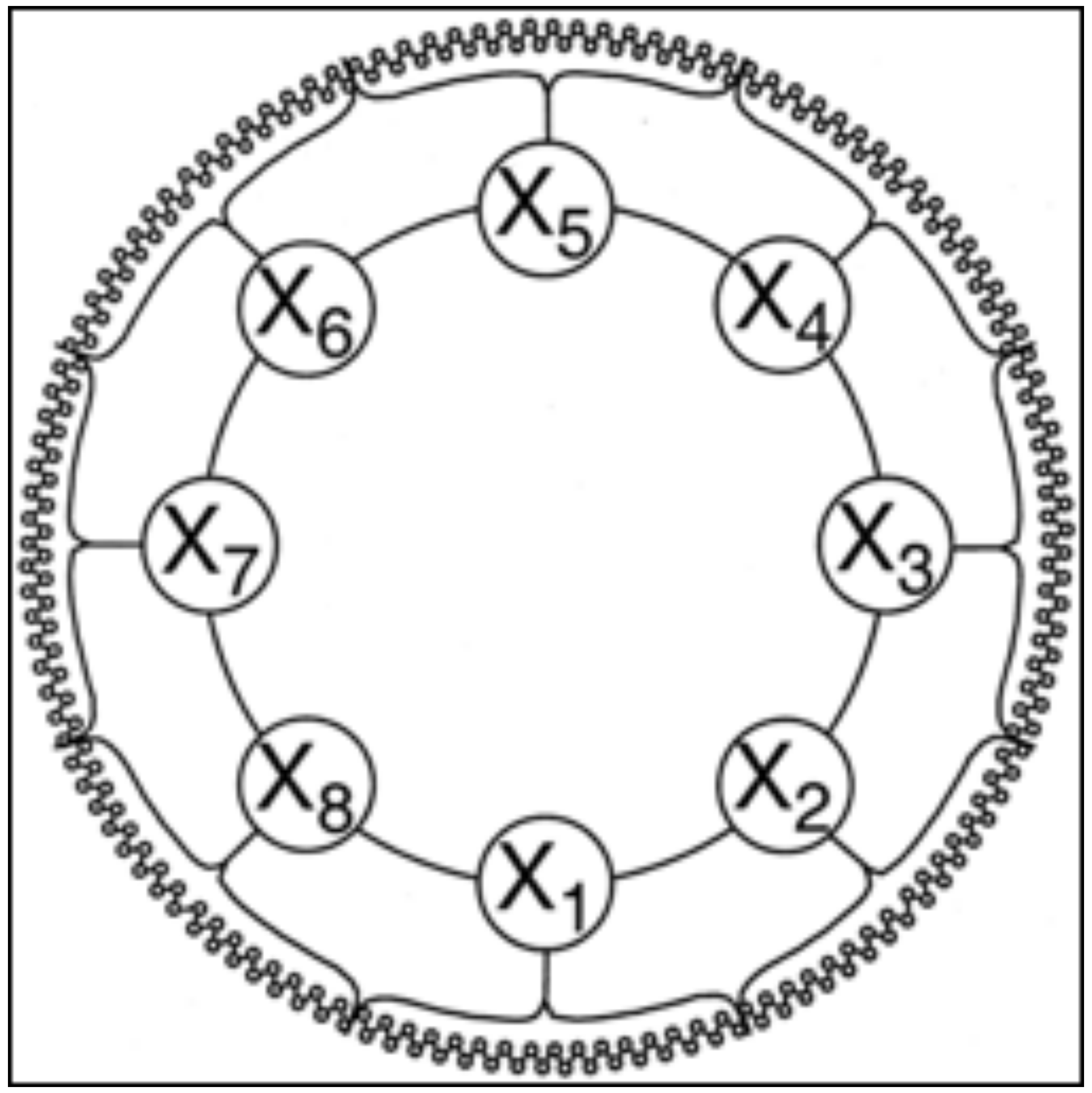
Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz ’96 Model

Lorenz ‘96 Model

Subgrid U(X)

$$\frac{dX_k}{dt} = -X_{k-1}\left(X_{k-2} - X_{k+1}\right) - X_k + F - \frac{hc}{b}\sum_{j=J(k-1)+1}^{kJ} Y_j$$

$$\frac{dY_j}{dt} = -cbY_{j+1}\left(Y_{j+2} - Y_{j-1}\right) - cY_j + \frac{hc}{b} X_{\text{int}[(j-1)/J]+1}$$



Parameter Settings:

No. X variables, K=8	Forcing term, F = 20
No. Y variables / X, J = 32	Spatial scale ratio, b = 10
Coupling constant, h=1	, Timescale ratio, c = 10

1. Integrate both equations, RK4, dt = 0.001 => “true atmosphere”
2. Forecast problem: assume Y variables unresolved. Integrate X equation RK2, dt = 0.005. Must parametrise U(X)

Wilks, 2005
Arnold et al, 2013

Gagne, D. J., H. Christensen, A. C. Subramanian, A. Monahan (2019): Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model, JAMES, 12, e2019MS001896. https://doi.org/10.1029/2019MS001896.

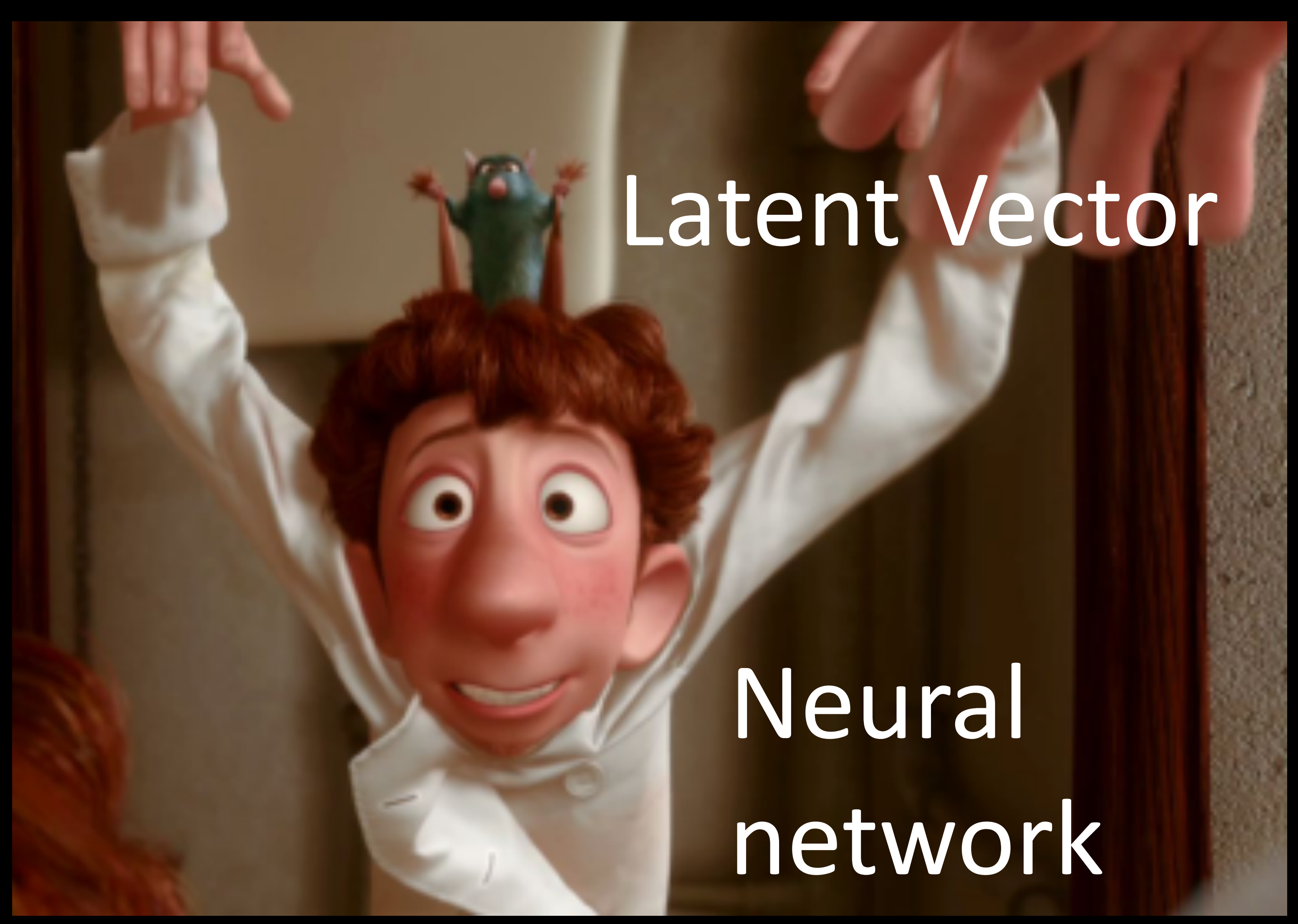
Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model

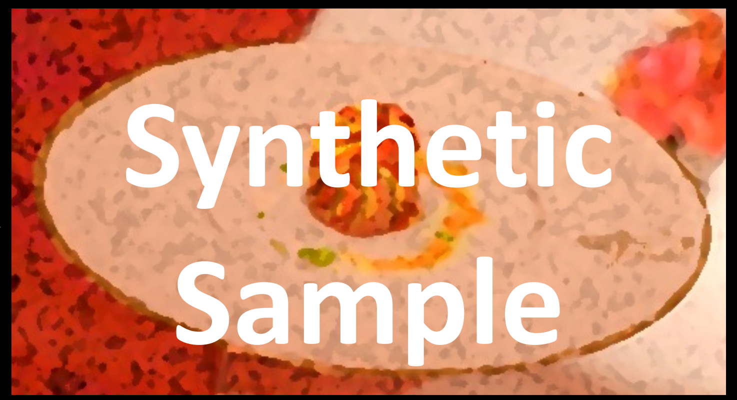
David John Gagne II, Hannah Christensen, Aneesh Subramanian, Adam Monahan, 2020

Use discriminative model to train a generative model
Originally proposed by Goodfellow et al. (2014)

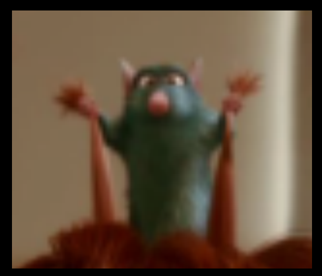
Generator: Creates synthetic samples drawn from training data based on latent vector.

Critic: Determines which samples are real or synthetic. Adaptive loss function.





Feedback



X_{t-1}, U_{t-1}, $z_{AR1\ N(0,1)}$

U_t

https://upload.wikimedia.org/wikipedia/en/e/e1/Ratatouille-remy-control-linguini.png
http://www.imdb.com/character/ch0009859/mediaviewer/rm988253440

GAN Configurations

Type	Inputs
XU	X_{t-1} and U_{t-1}
X	X_{t-1}

-

Type	Noise SD
lrg	1
med	0.1
sml	0.01
tny	0.001

-

Type	Noise Time Correlation
w	0
r	Estimated from offline deterministic U_t residual autocorrelation

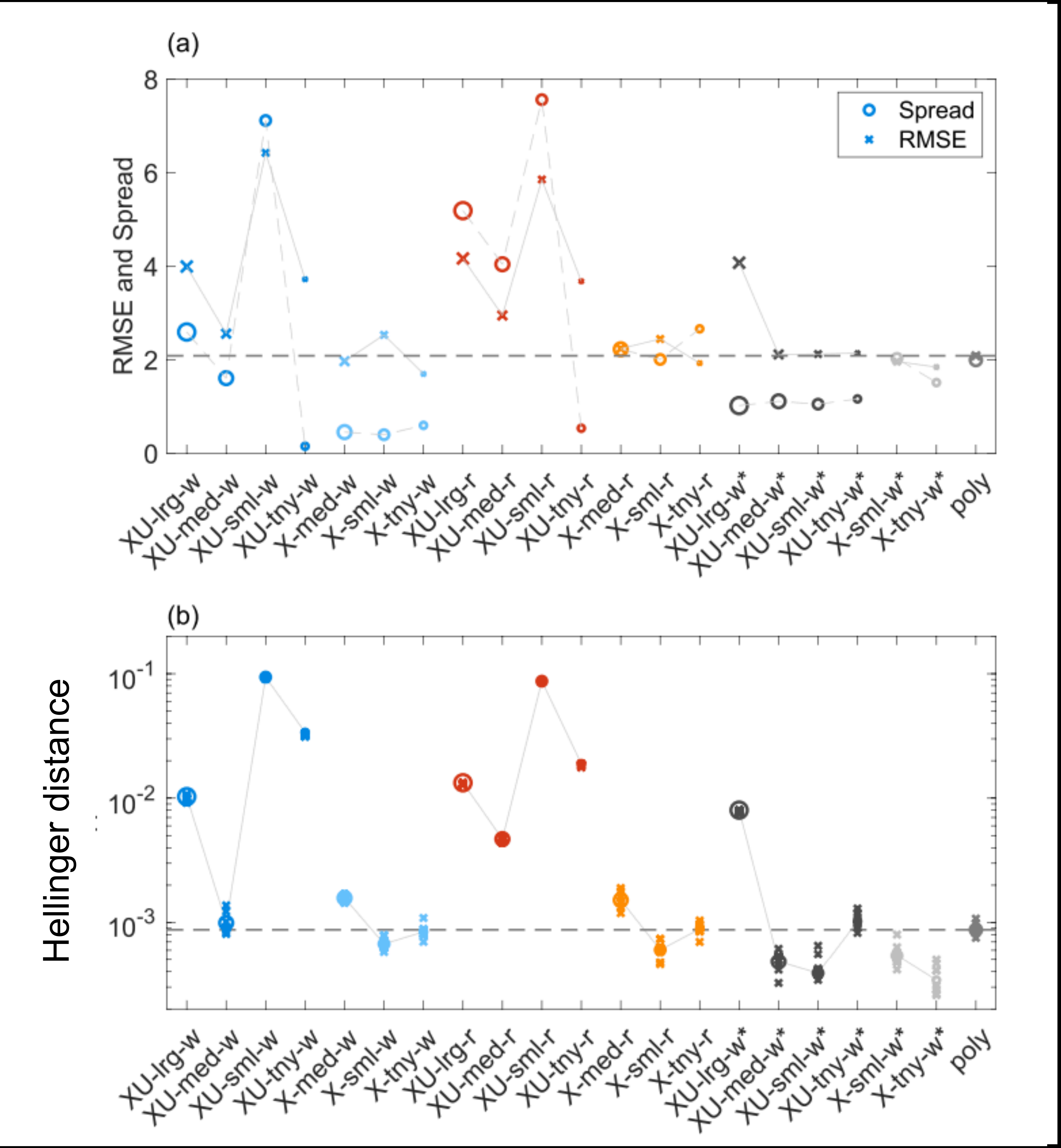
Example: XU-med-w

Weather Forecast and Climate Run Summary Statistics

Weather forecasts were performed for 750 initial conditions with 50 ensemble members for each forecast.

GAN parameterized models with X-only input have lower RMSE than poly but are overdispersive

Climate analysis: 20,000 MTU integration (~270 "years"), discard first 2,000 MTU (30 "years") as spinup.
PDF measures of parameterized model climate compared to truth run shows some of the stochastic GAN parameterizations helps improve climate representation compared to polynomial fit parameterizations (analogous to current day climate model parameterizations)



Smaller Hellinger distance: forecast pdf closer to 'true' pdf

Summary

- Mean bias is reduced in stochastic prediction models
- Extreme weather events are better represented
- High frequency perturbations have nonlinear rectifications improving low frequency variability in models
- Machine learning approaches can be used to improve both weather and climate forecasts and uncertainty in these forecasts


"I believe that the ultimate climate models..will be stochastic, ie random numbers will appear somewhere in the time derivatives"
- Lorenz (1975).
Thank You

Ensemble Kalman Filter

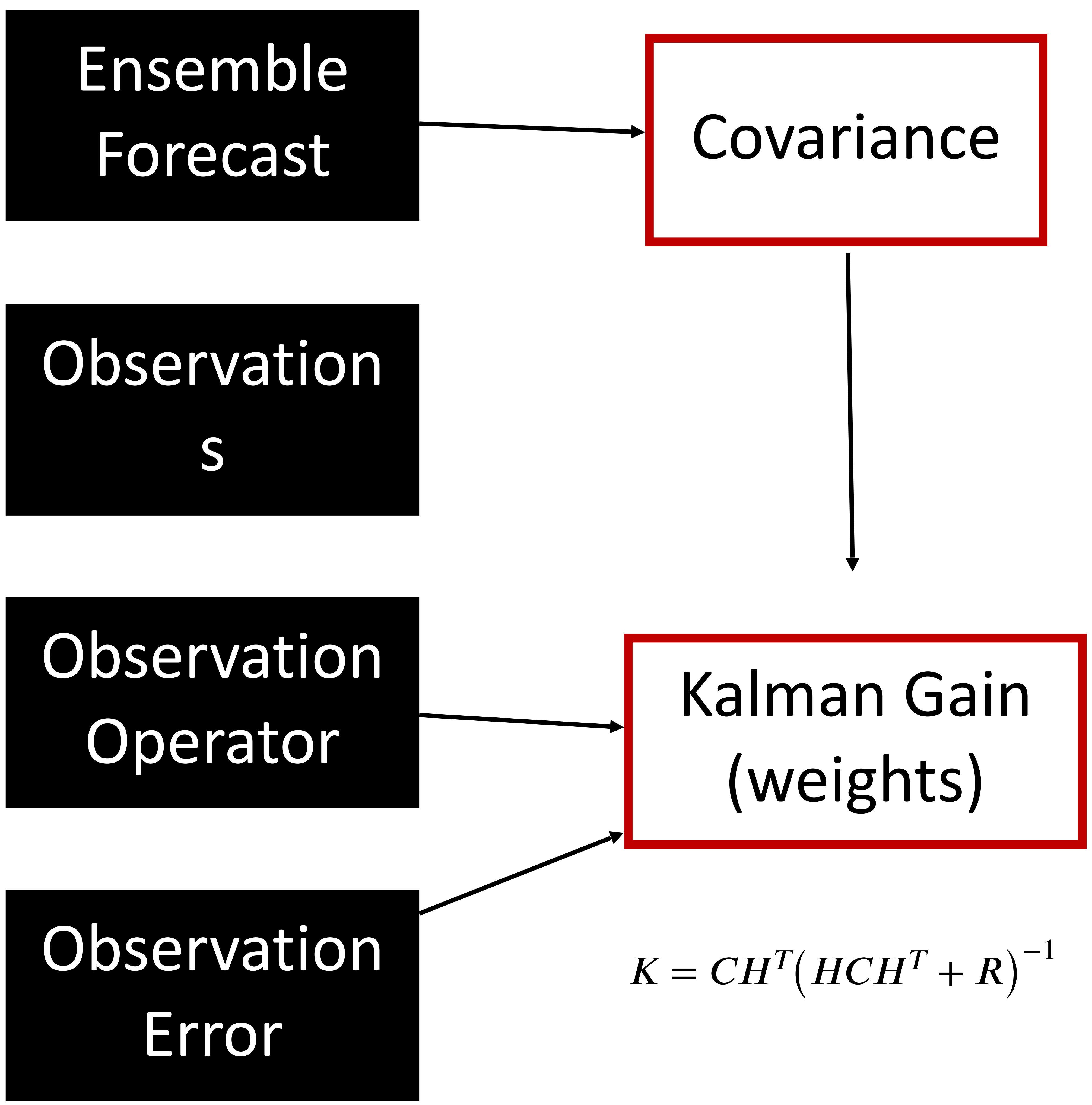
Ensemble Forecast

Observations

Observation Operator

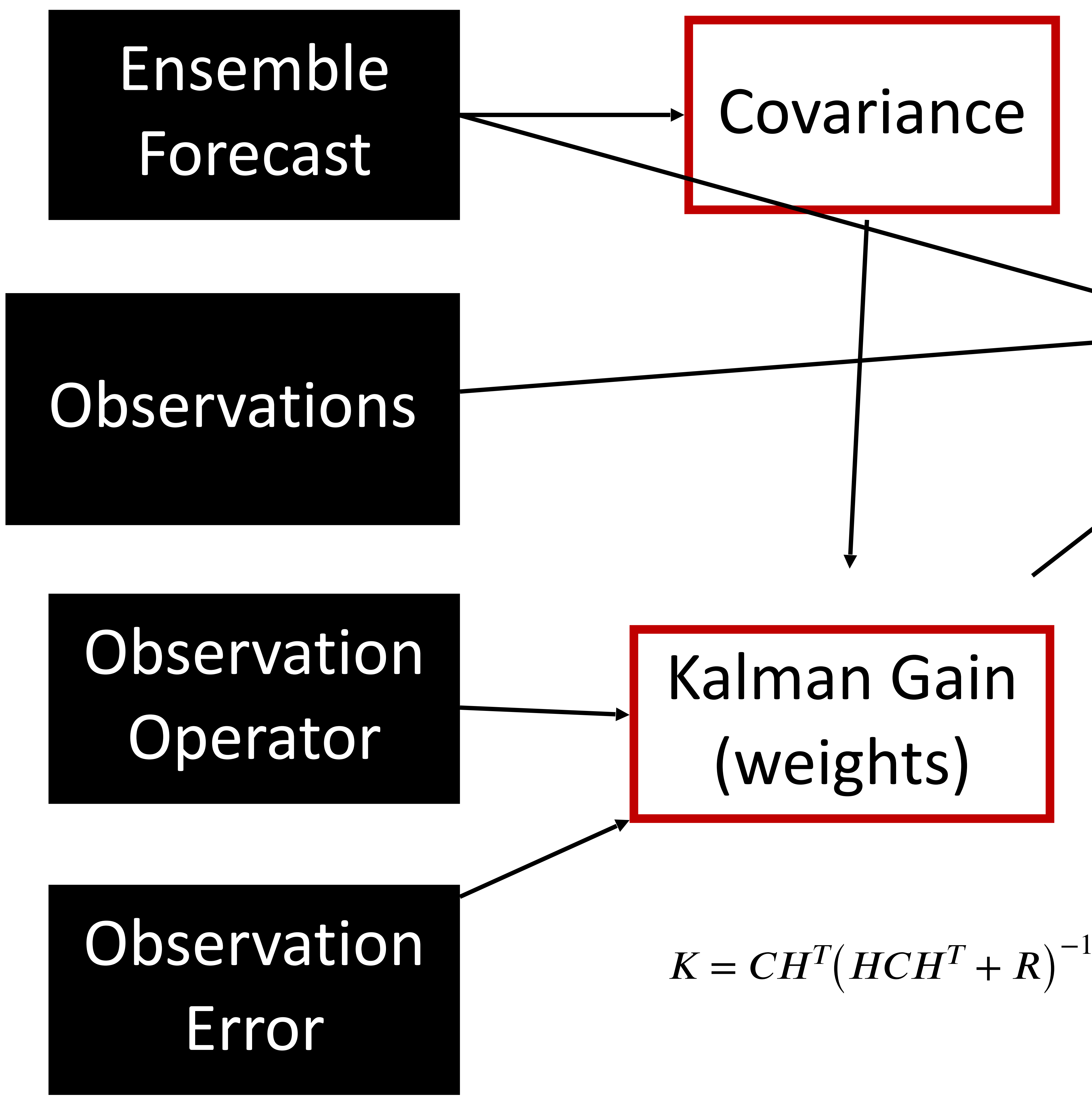
Observation Error

Ensemble Kalman Filter



$$K = CH^T\left(HCH^T + R\right)^{-1}$$

Ensemble Kalman Filter



$$X^a = X^f + K(Y - HX^f)$$

Generating Training Data

Generate “true” trajectory by integrating Lorenz system

↓

Generate synthetic observations by adding normally distributed noise

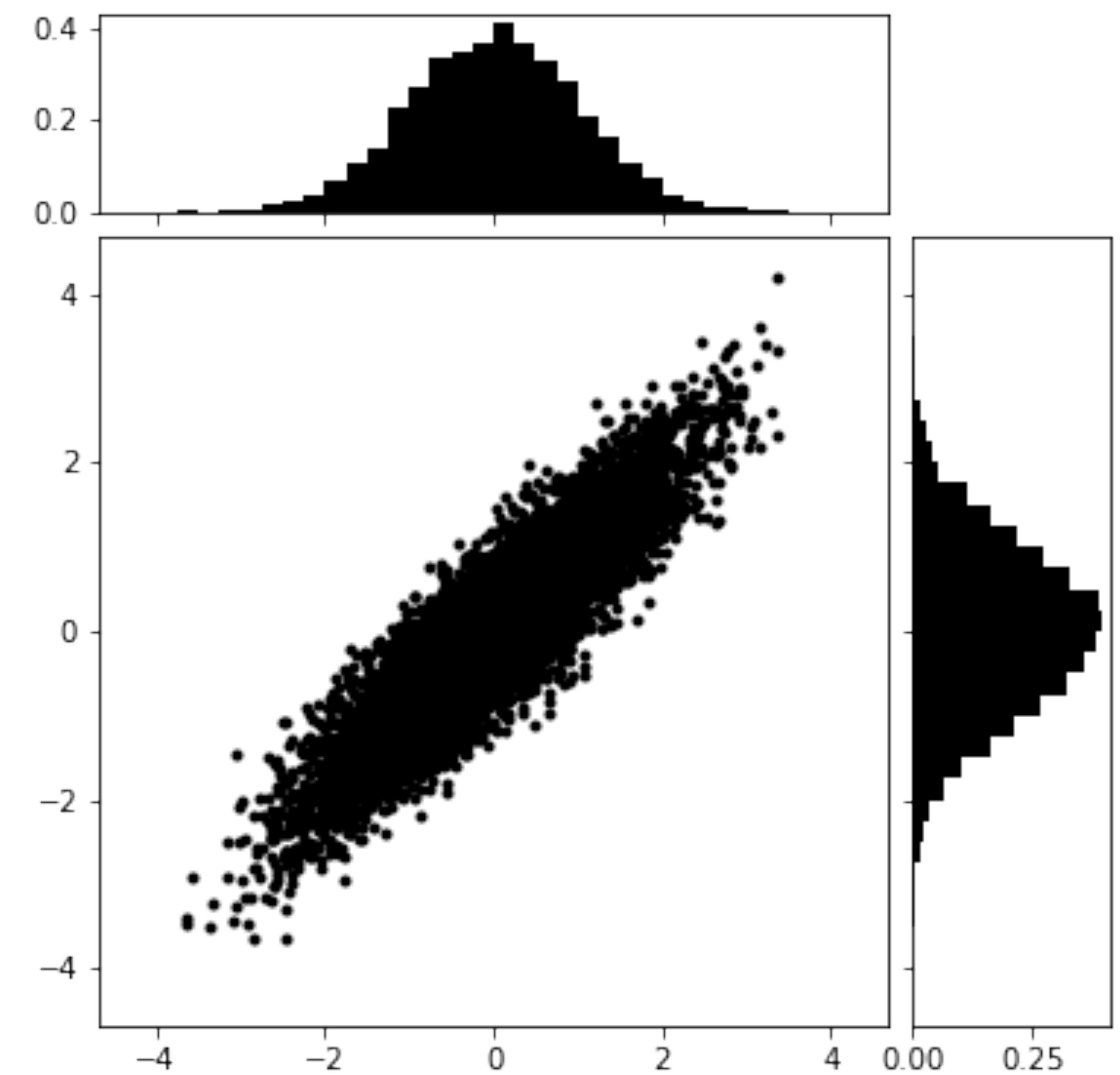
↓

Assimilate synthetic observations with EnKF

Parameter	**Value**
Assimilation Time Step	0.05 Model Units/6 Hours
Max Time	2000 Model Units
Localization Distance	5 grid points
Inflation Factor	1 (no inflation)
Observation Error Standard Deviation	30% of climatological standard deviation
Ensemble Size	100

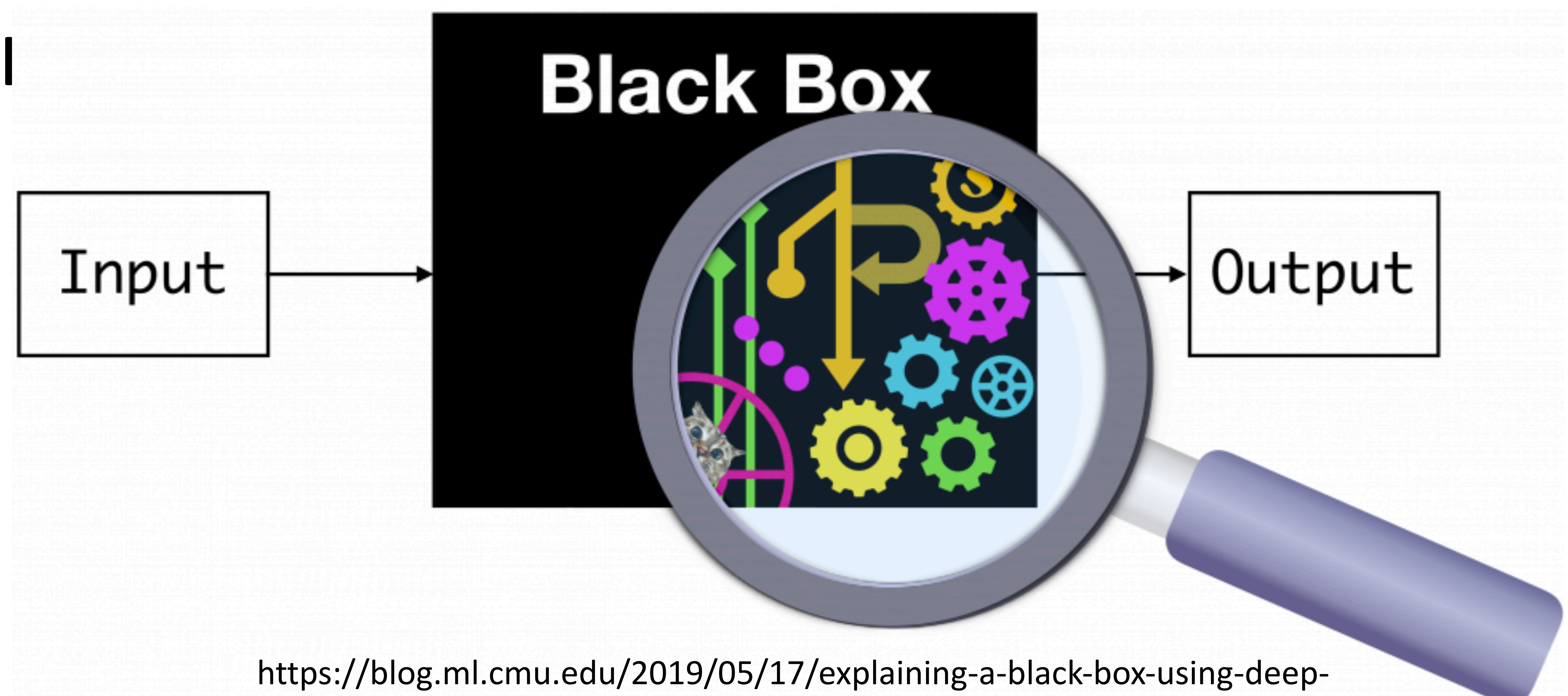
Regeneration of Ensemble

- Adjust forecast perturbations to match standard deviations
- Weighted average of forecast and adjusted perturbations (tuned)



SHAP Values Quantify impact of Input Features

- “Explainable AI”
- SHAP values approximate the contribution to the output (prediction) magnitude from each input va



https://blog.ml.cmu.edu/2019/05/17/explaining-a-black-box-using-deep-variational-information-bottleneck-approach/

SHAP Values for Linear Regression Model

- Linear regression model:

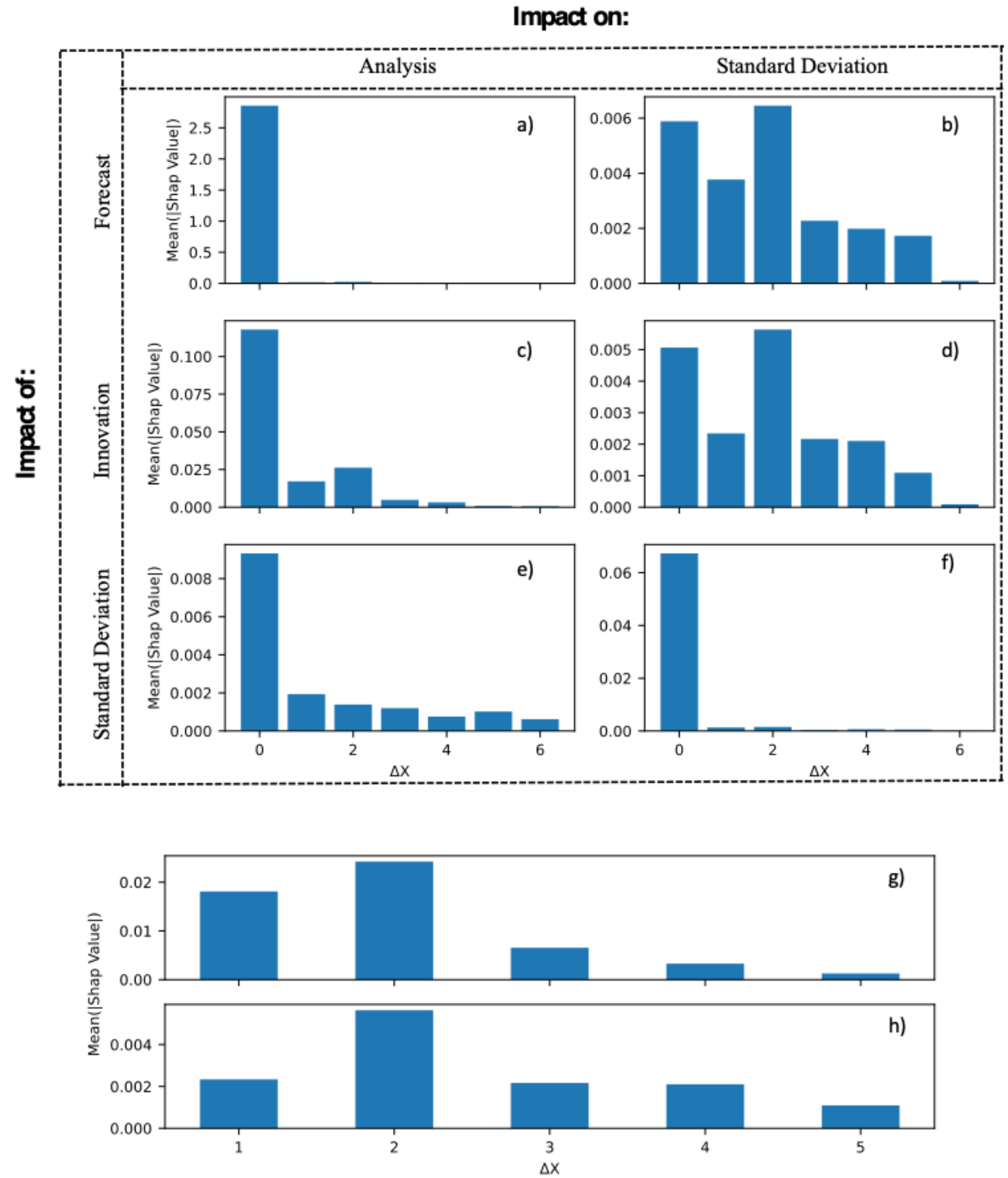
$$y = \sum_i c_i x_i$$

- SHAP values approximate the contribution to the output (prediction) magnitude from each input variable

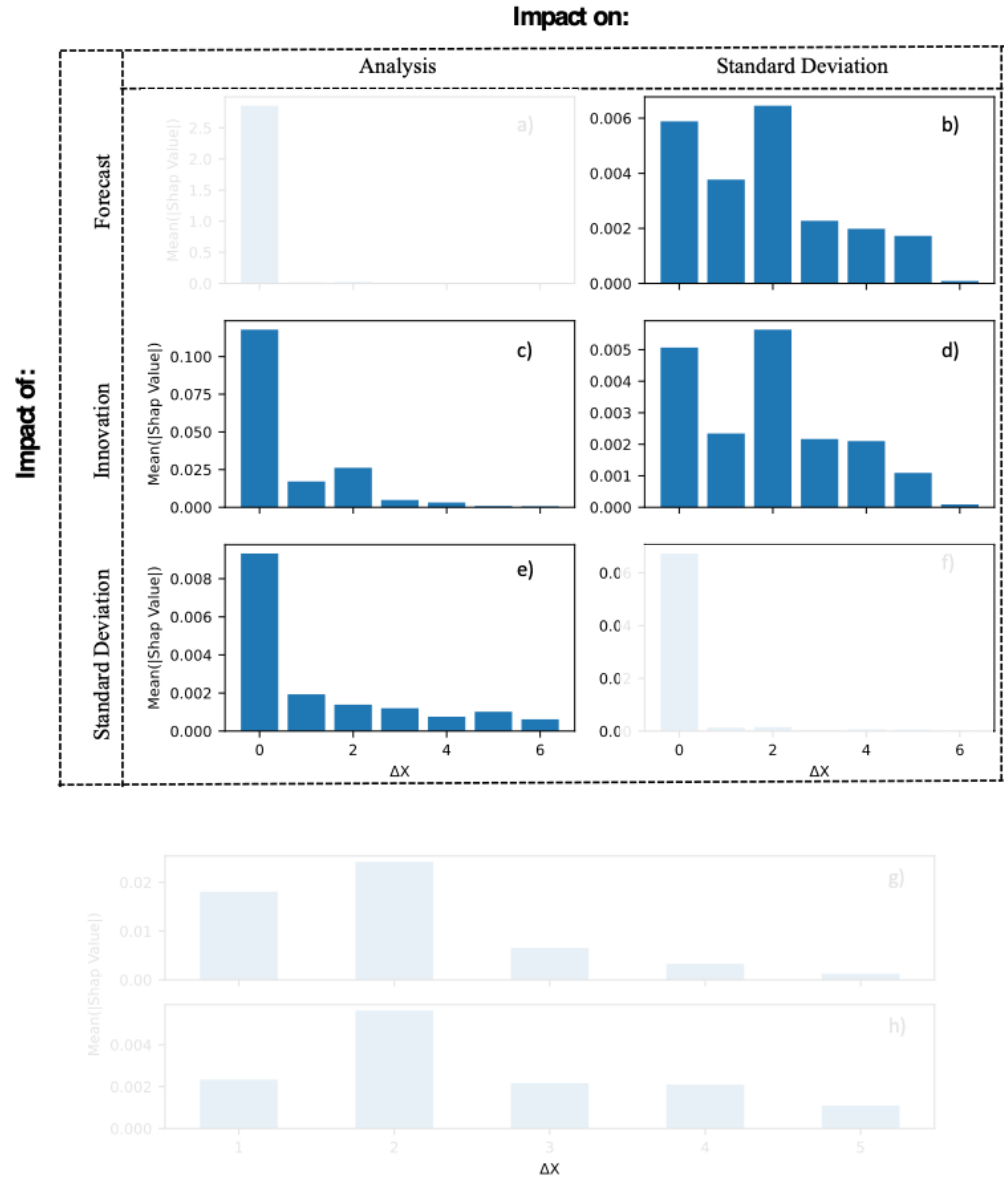
SHAP values for input variable i: $\phi_i = c_i(x_i - E[x_i])$

Model prediction as a function of SHAP values: $y - E[y] = \sum(\phi_i)(x_i - E[x_i])$

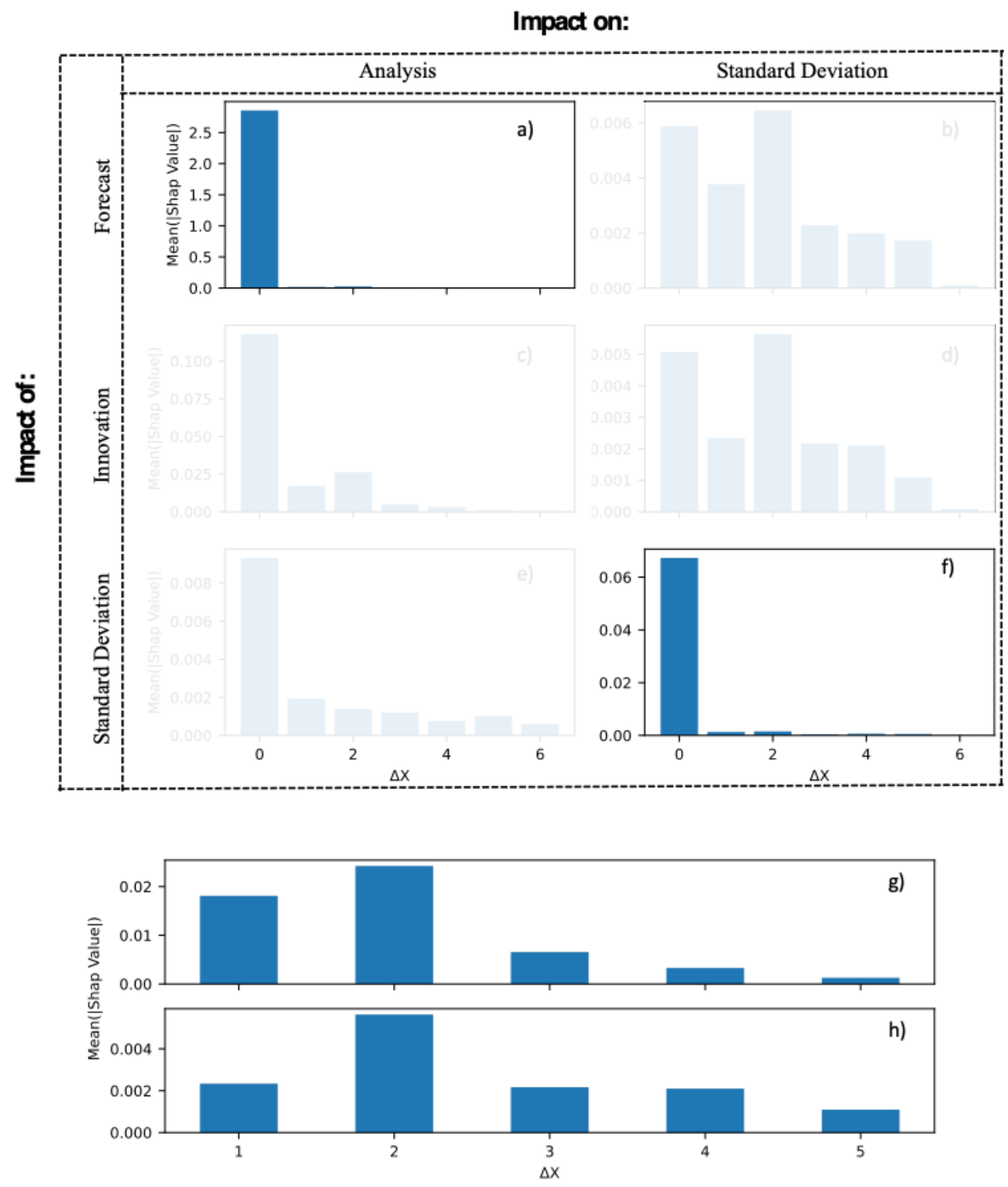
- Expect:
  - Monotonic decrease
  - Value at 2 greater than at 1 or 3



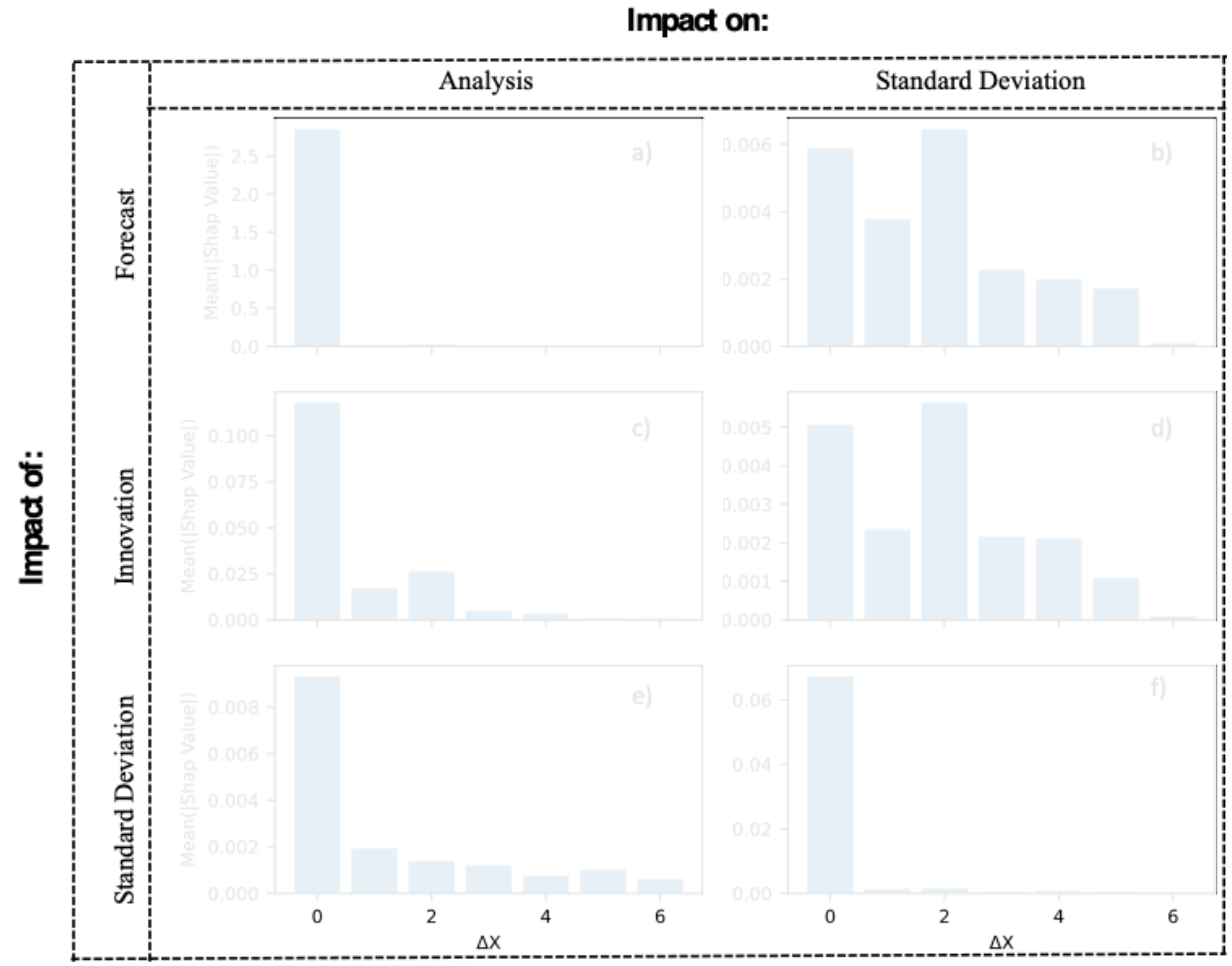
- Expect:
  - Monotonic decrease
  - Value at 2 greater than at 1 or 3

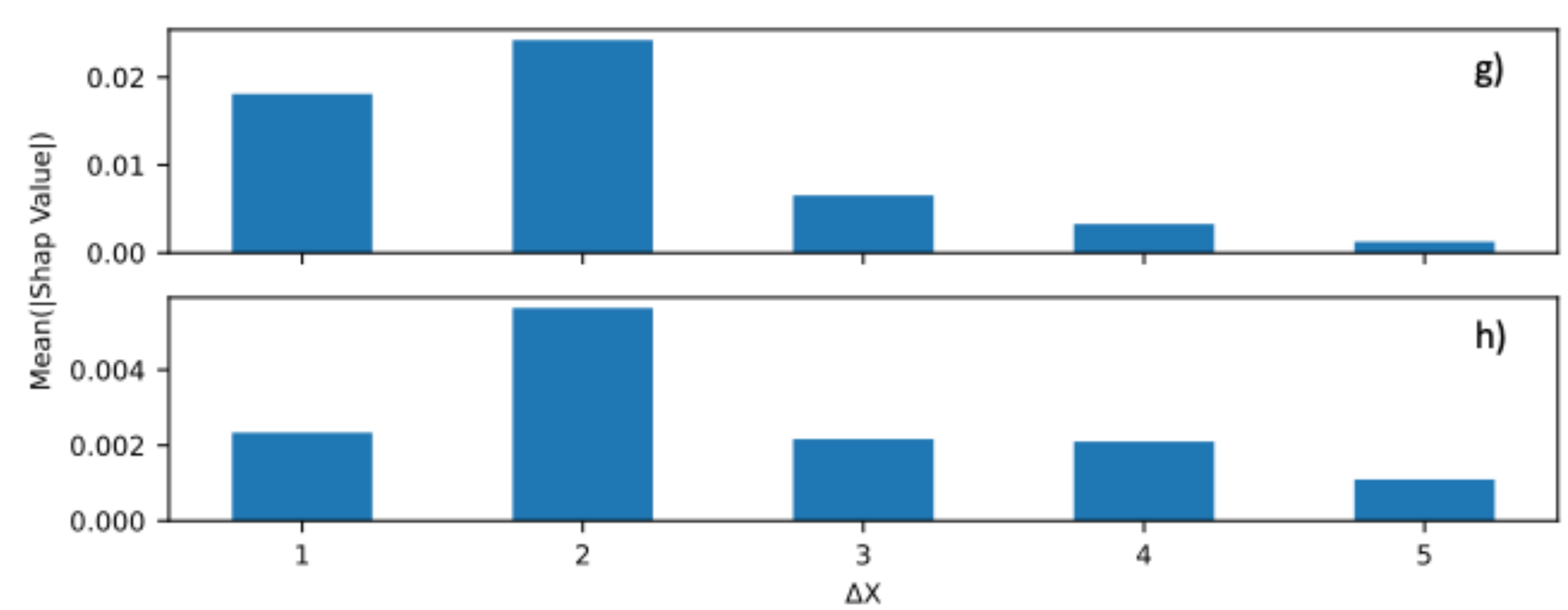


- Expect:
  - Monotonic decrease
  - Value at 2 greater than at 1 or 3



- Expect:
  - Monotonic decrease
  - Value at 2 greater than at 1 or 3





Conclusions

- A simple neural network can be trained to emulate a traditional DA method
- Using a trained network on high-resolution data improves accuracy in an augmented method
- Explainable AI methods suggest that the network learns the correlation structure of the underlying system