Covariance Localization in Strongly Coupled Data Assimilation

EMC Seminar @NCWCP

Takuma Yoshida and Eugenia Kalnay

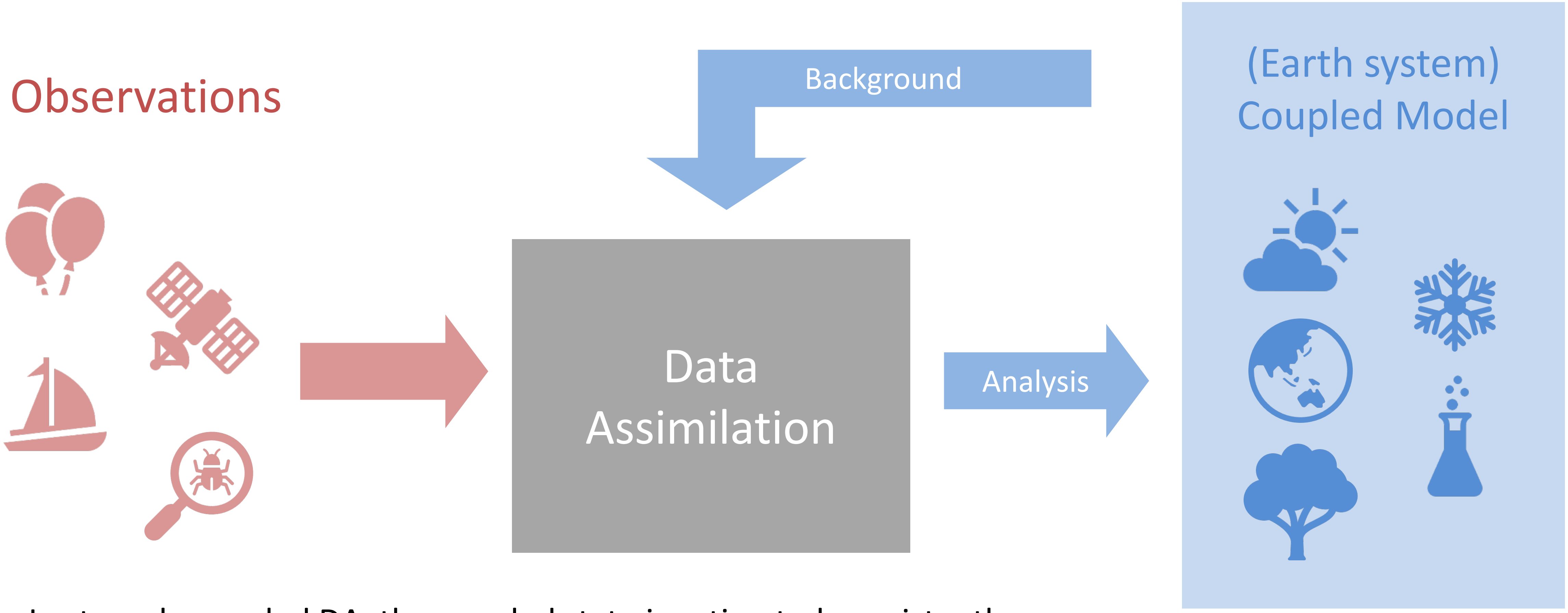
August 2, 2019

Department of Atmospheric and Oceanic Science,
University of Maryland

Special thanks to:
Prof. Stephen G. Penny
Prof. James A. Carton
Prof. Brian R. Hunt
Dr. Safa Motesharrei

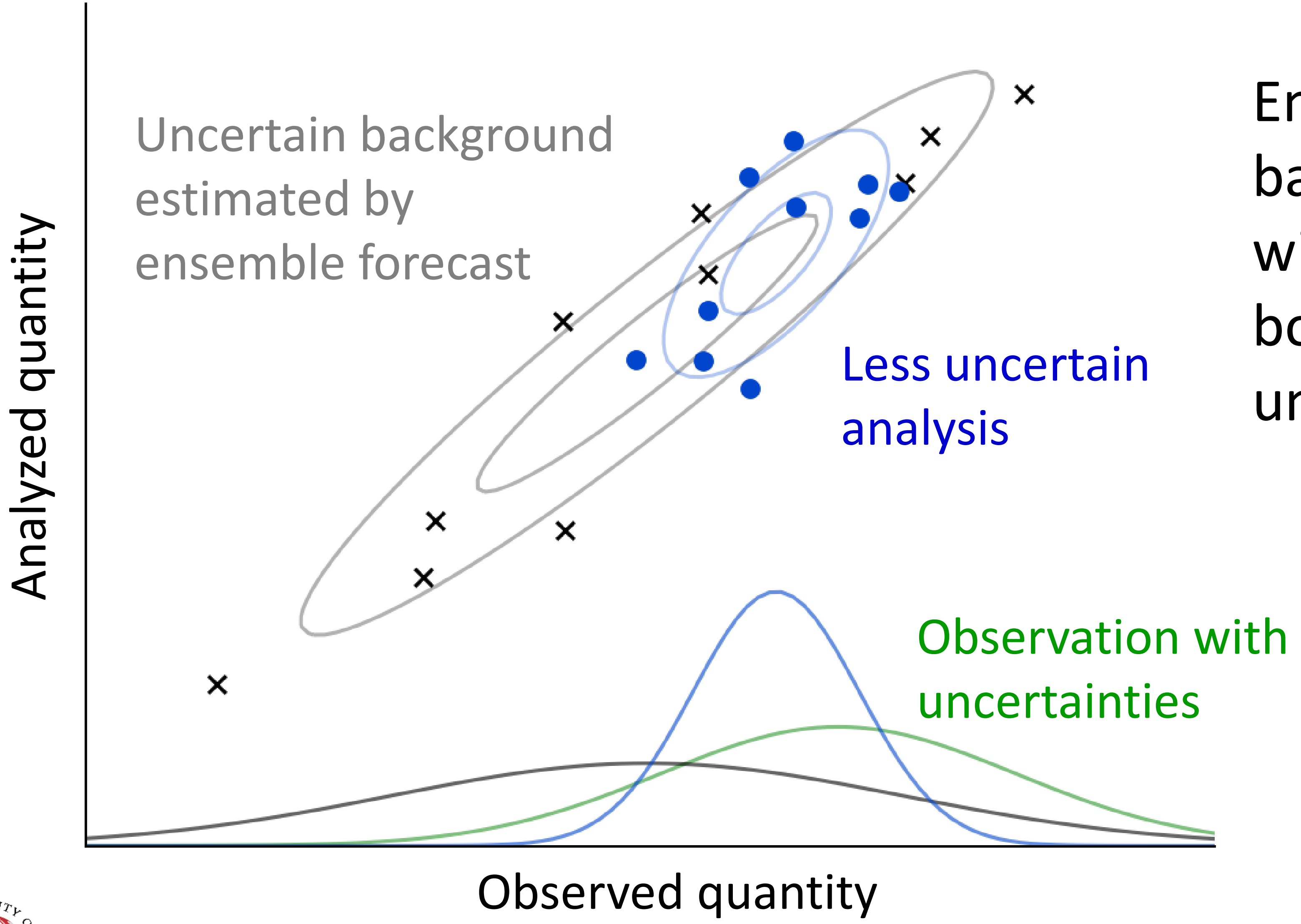
Self introduction

Strongly coupled DA – the ultimate goal



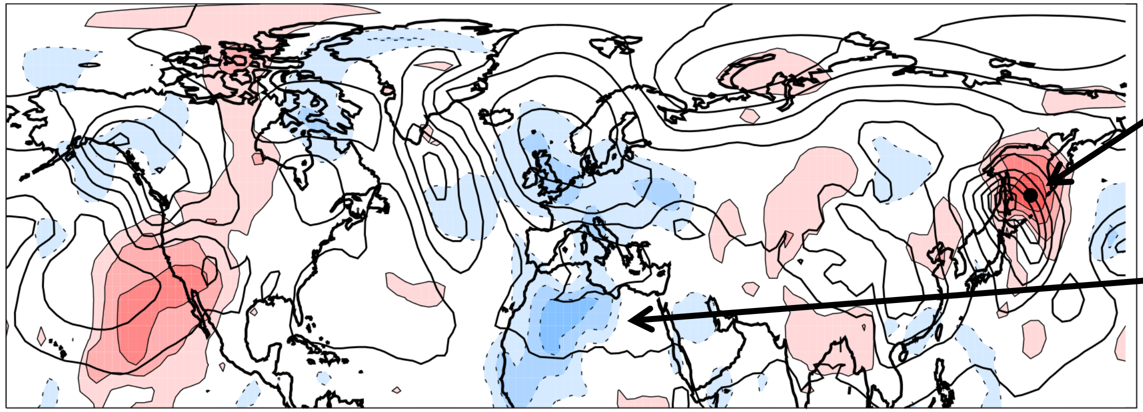
In strongly coupled DA, the coupled state is estimated consistently using observations from every subsystem of the earth.
All available constraints are brought together.

Data assimilation with ensemble Kalman filter



Ensemble forecast estimates background error covariance, with which we can constrain both observed and unobserved quantities

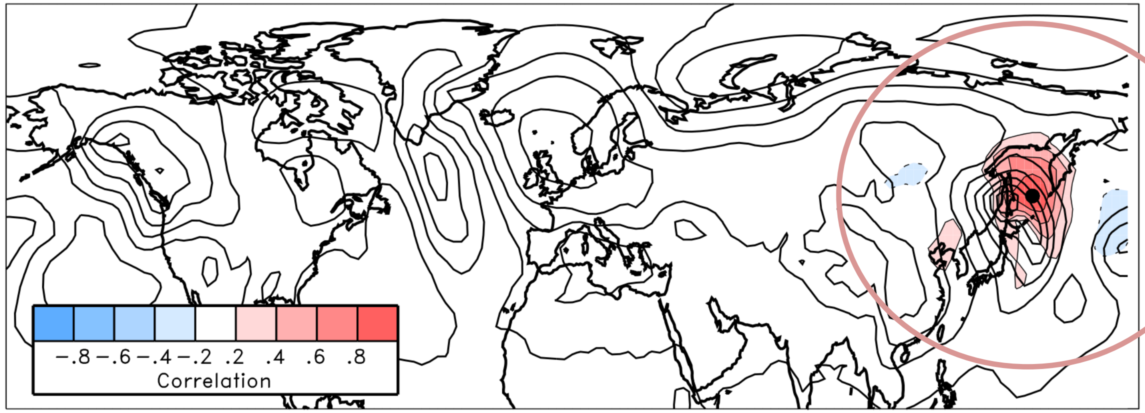
Covariance localization



An observation

The observation influences faraway analysis (shading), which is often detrimental

Localization

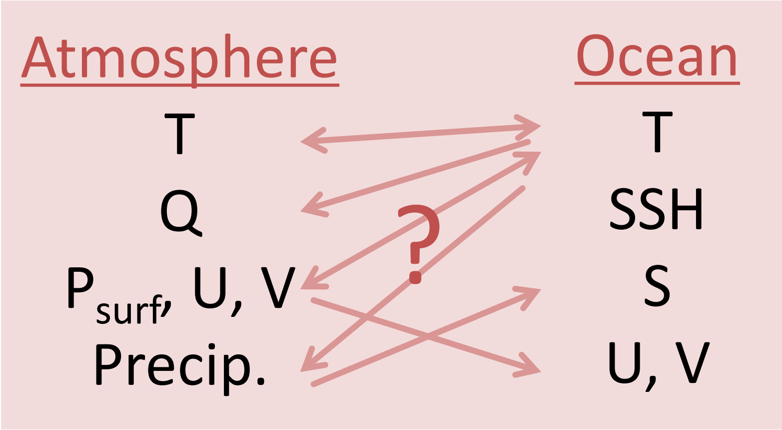


Influence of the observation is limited to its neighborhood, and analysis is often improved

Figure courtesy Hamill (2006)

Localization based on the type of variable

- Variable localization (Kang et al., 2011)
  - Localize the analysis depending on the observation/analysis variable types
  - By not assimilating CO_2 observations into some dynamical variables (and *vice versa*), analysis accuracy improves in experiments with dynamics-carbon coupled model
- How can we optimize localization for coupled Earth system models with growing complexity?
  - We may use physical intuitions as Kang et al.
  - We want a metric of relevance between each analysis variable and observation



Localization and strongly coupled DA

Localization

Assimilate select observations

(e.g., only within 2000 km)

↔

Strongly coupled DA

Assimilate more observations

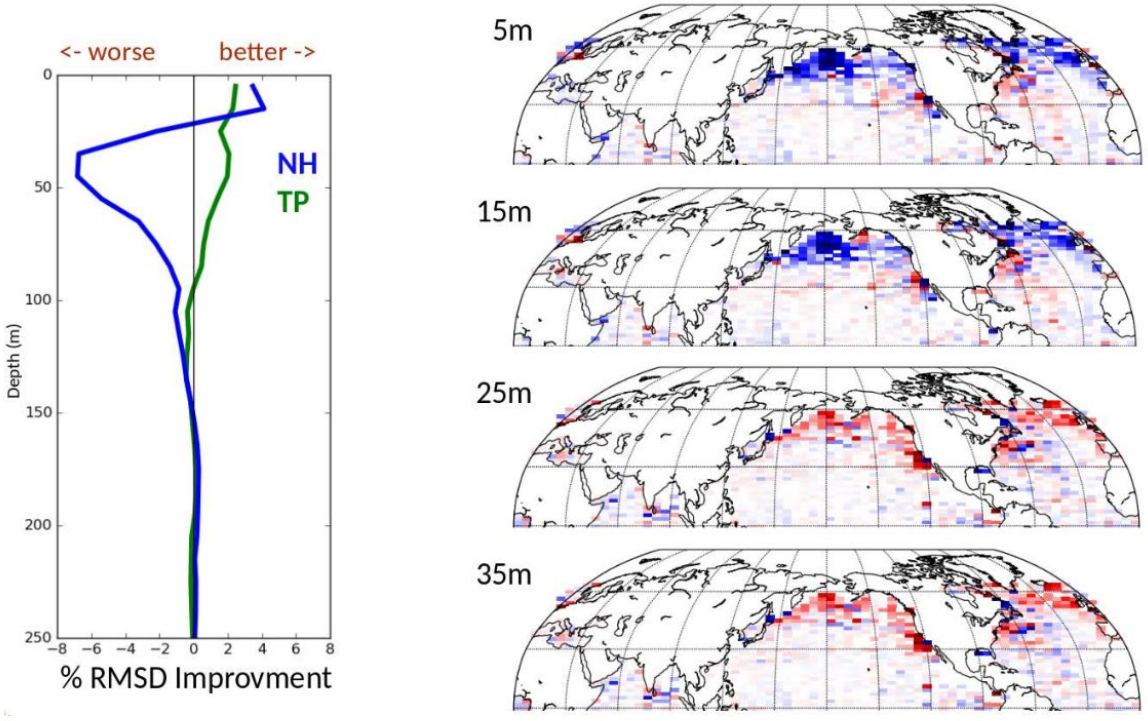
(e.g., atmospheric observations to ocean analysis)

Common problem

What kind of observations are most relevant to each variable's analysis?

- How to prioritize observations?
- When is strongly coupled DA beneficial?

Importance of localization for strongly coupled DA: example



Relative RMS of observation minus background (strongly vs weakly coupled) for ocean temperature

Travis Sluka (2018)

Using vertical and variable localization ... is vital when using a limited ensemble size.

Their suggestion: assimilating atmospheric observations only to the mixed layer

Outline

0. Introduction
1. Correlation-cutoff method and experiments with 9-variable coupled model
2. Localization modeling with neural networks – successful assimilation experiments with global atmosphere-ocean coupled model
3. Sudden and major change of dynamics found in coupled chaotic systems
4. Summary and future directions

Outline

0. Introduction
1. Correlation-cutoff method and experiments with 9-variable coupled model
2. Localization modeling with neural networks – successful assimilation experiments with global atmosphere-ocean coupled model
3. Sudden and major change of dynamics found in coupled chaotic systems
4. Summary and future directions

Yoshida, T., and E. Kalnay, (2018). Correlation-Cutoff Method for Covariance Localization in Strongly Coupled Data Assimilation. *Monthly Weather Review*, doi:10.1175/MWR-D-17-0365.1.

Mean squared error correlation for localization

Reduction of analysis error variance by single-observation assimilation

$$\frac{\sigma_{bi}^2 - \sigma_{ai}^2}{\sigma_{bi}^2} = \frac{\sigma_{y_b}^2}{\sigma_{y_b}^2 + \sigma_{y_o}^2} \operatorname{corr}^2(\delta x_{bi}, \delta y_b)$$

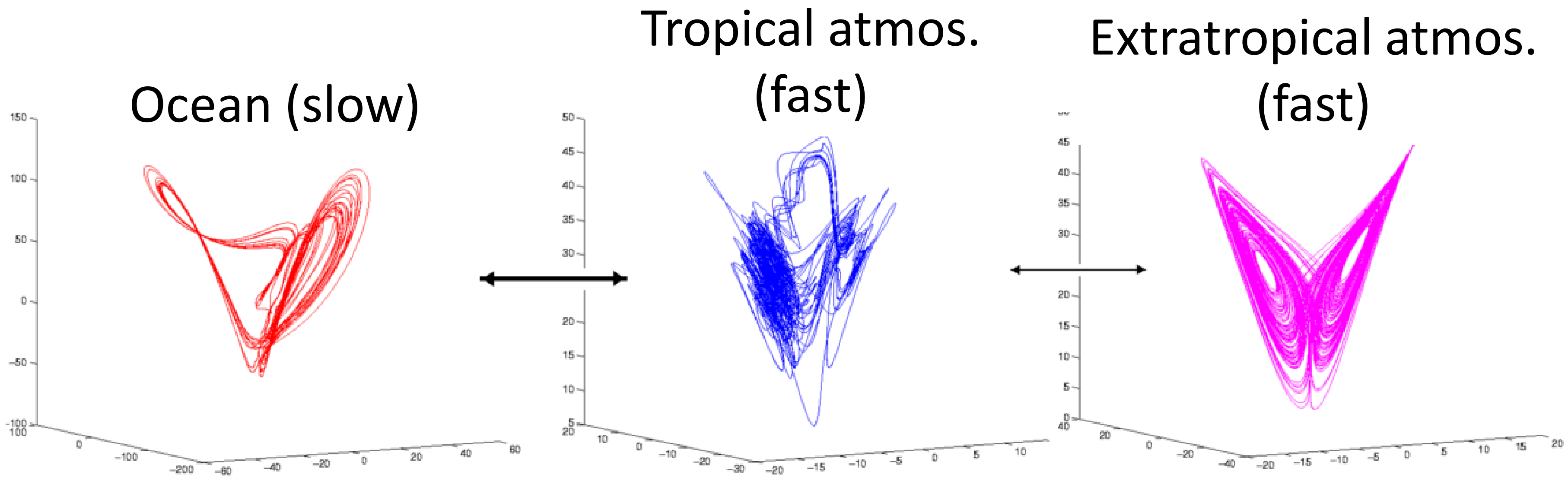
Relative accuracy of the observation to the background

Squared background error correlation between the analysis variable and observable

<u>Correlation-cutoff method</u>

Assimilate only observations whose background error is (on average) strongly correlated with the analysis variable's background error

Model: Peña and Kalnay (2004)



$$
\begin{aligned}
\dot{X} &= \tau\sigma(Y - X) - c(x_t + k_2) \\
\dot{Y} &= \tau r X - \tau Y - \tau S X Z + c(y_t + k_2) \\
\dot{Z} &= \tau S X Y - \tau b Z - c_z z_t
\end{aligned}
$$

$$
\begin{aligned}
\dot{x}_t &= \sigma(y_t - x_t) - c(SX + k_2) - c_e(Sx_e + k_1) \\
\dot{y}_t &= r x_t - y_t - x_t z_t + c(SY + k_2) + c_e(Sy_e + k_1) \\
\dot{z}_t &= x_t y_t - b z_t + c_z Z
\end{aligned}
$$

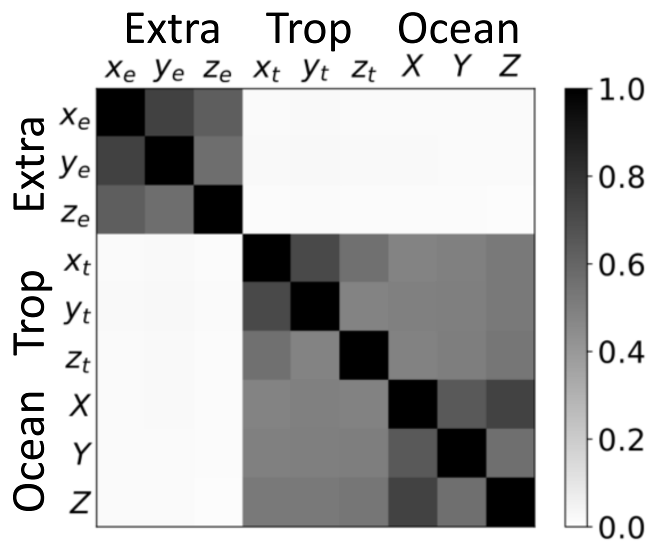
$$
\begin{aligned}
\dot{x}_e &= \sigma(y_e - x_e) - c_e(Sx_t + k_1) \\
\dot{y}_e &= r x_e - y_e - x_e z_e + c_e(Sy_t + k_1) \\
\dot{z}_e &= x_e y_e - b z_e
\end{aligned}
$$

- Three chaotic Lorenz models with different timescales are coupled with each other
- Shows chaotic coupled oscillations like ENSO in Ocean and Tropical Atmosphere

Figure: Peña and Kalnay (2004)

Obtain offline error statistics

1. Run an EnKF DA cycle (*offline cycle*)
2. Calculate background ensemble correlation for each pair of model variables and time
3. Calculate its temporal mean squared

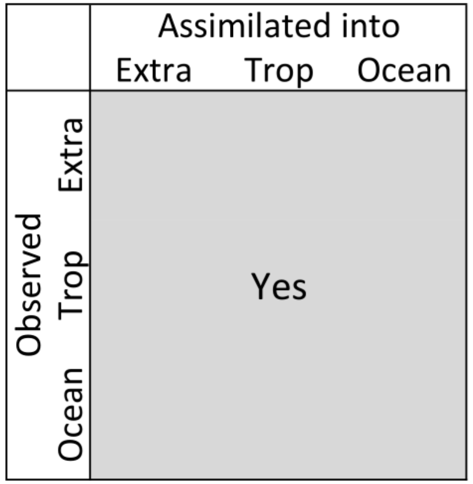


Strong background error correlation exists only within each subsystem and between “tropical atmosphere” and “ocean”

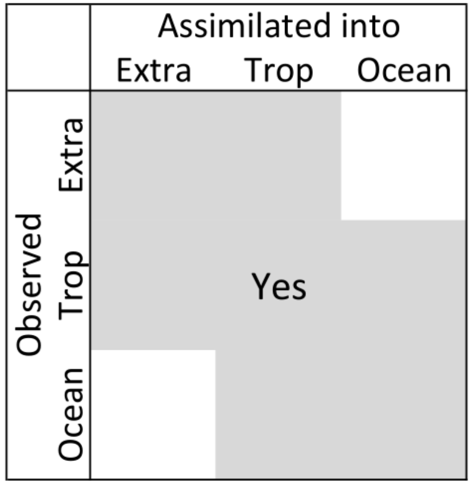
Different localizations for strongly coupled DA

Strongly coupled
(observations are mutually assimilated)

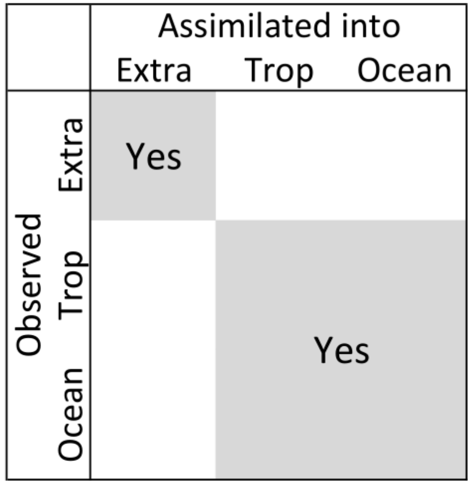
Standard strongly coupled



Adjacent

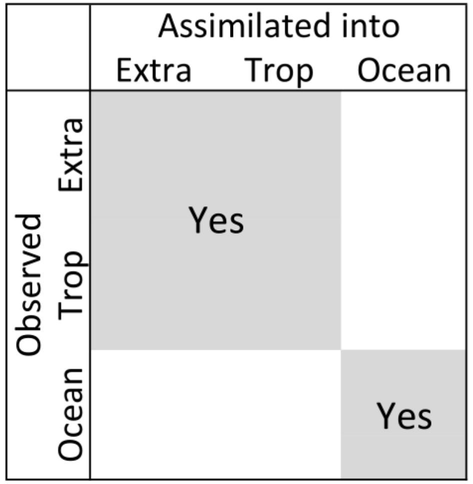


Correlation-cutoff

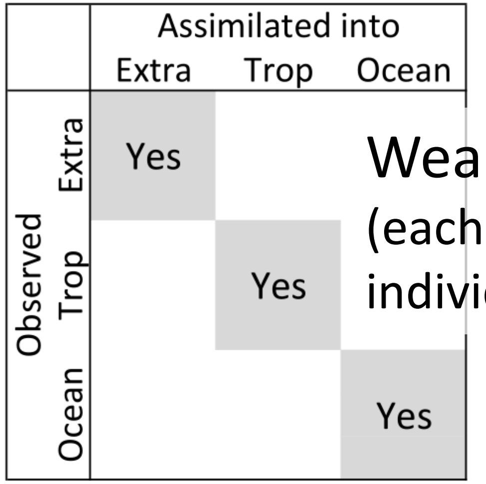


Guided

Atmos-coupling

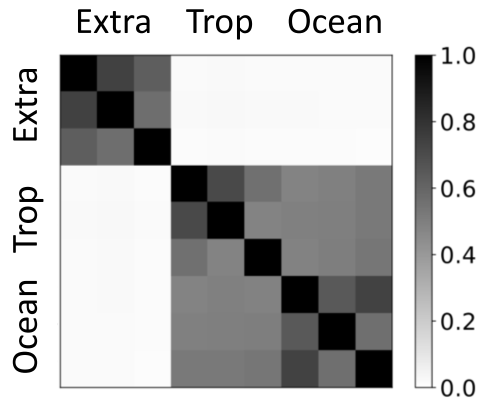


Weakly coupled



Weakly coupled
(each subsystem is individually analyzed)

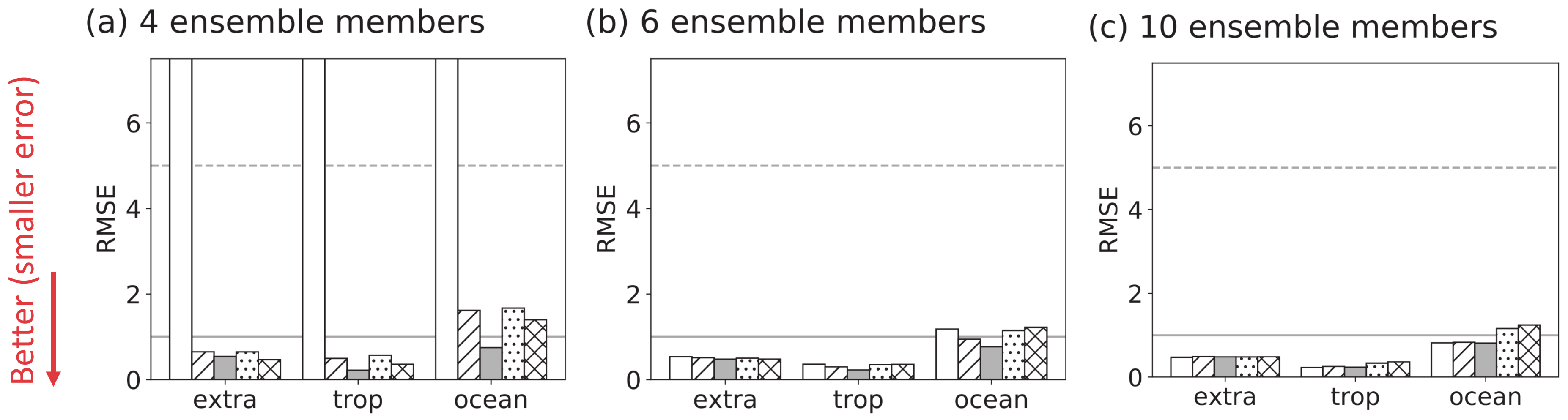
Mean squared background error correlations



Experimental settings (9-variable model)

Observations	3 of 9 variables (y_e, y_t, Y) are observed at the end of each assimilation window (every 8 timesteps; 0.08 nondimensional time units) Observation error of 1.0 for atmosphere, 5.0 for ocean
Assimilation algorithm	Local ensemble transform Kalman filter (LETKF) Every 8 timesteps
Ensemble size	4, 6, or 10
Covariance inflation	Adaptive multiplicative inflation
Localization	**5 different ways as explained**
Experiment length	75,000 timesteps (last 50,000 timesteps are verified)

Result: analysis RMS error



- **Localization with the correlation-cutoff method achieved smallest error in essentially all experimental settings and components**
- **Standard strongly coupled DA has larger error than weakly coupled DA when the ensemble size is insufficient**

Standard strongly coupled
Adjacent
Correlation-cutoff
Atmos-coupling
Weakly coupled
Obs error (atmos)
Obs error (ocean)

Conclusion (Correlation-cutoff method with a 9-variable model)

- The mean squared ensemble correlation helps to estimate the relevance between each analysis variable and observation
  - **We can selectively couple the analysis only where the background errors are well-correlated: correlation-cutoff method**
- The correlation-cutoff method is tested with a 9-variable coupled model
  - Achieved better analysis accuracy than weakly coupled DA and standard strongly coupled DA, especially for limited ensemble sizes

Outline

0. Introduction
1. Correlation-cutoff method and experiments with 9-variable coupled model
2. Localization modeling with neural networks – successful assimilation experiments with global atmosphere-ocean coupled model
3. Sudden and major change of dynamics found in coupled chaotic systems
4. Summary and future directions

Localization in general (observation space)

Attributes of analysis variable x_i

- Latitude/longitude/altitude
- Analysis variable type

Attributes of observable y_j

- Latitude/longitude/altitude
- Observation type

Some multivariate function

Localization weight

ρ_{ij} $(0 \leq \rho_{ij} \leq 1)$

Correlation-cutoff for a global coupled model

Attributes of analysis variable x_i

- Latitude/longitude/altitude
- Analysis variable type

Attributes of observable y_j

- Latitude/longitude/altitude
- Observation type

Localization function in observation space $g \circ f$

Nonlinear regression f

Expected squared background error correlation between (x_i, y_j)

Non-decreasing function g (*cutoff function*)

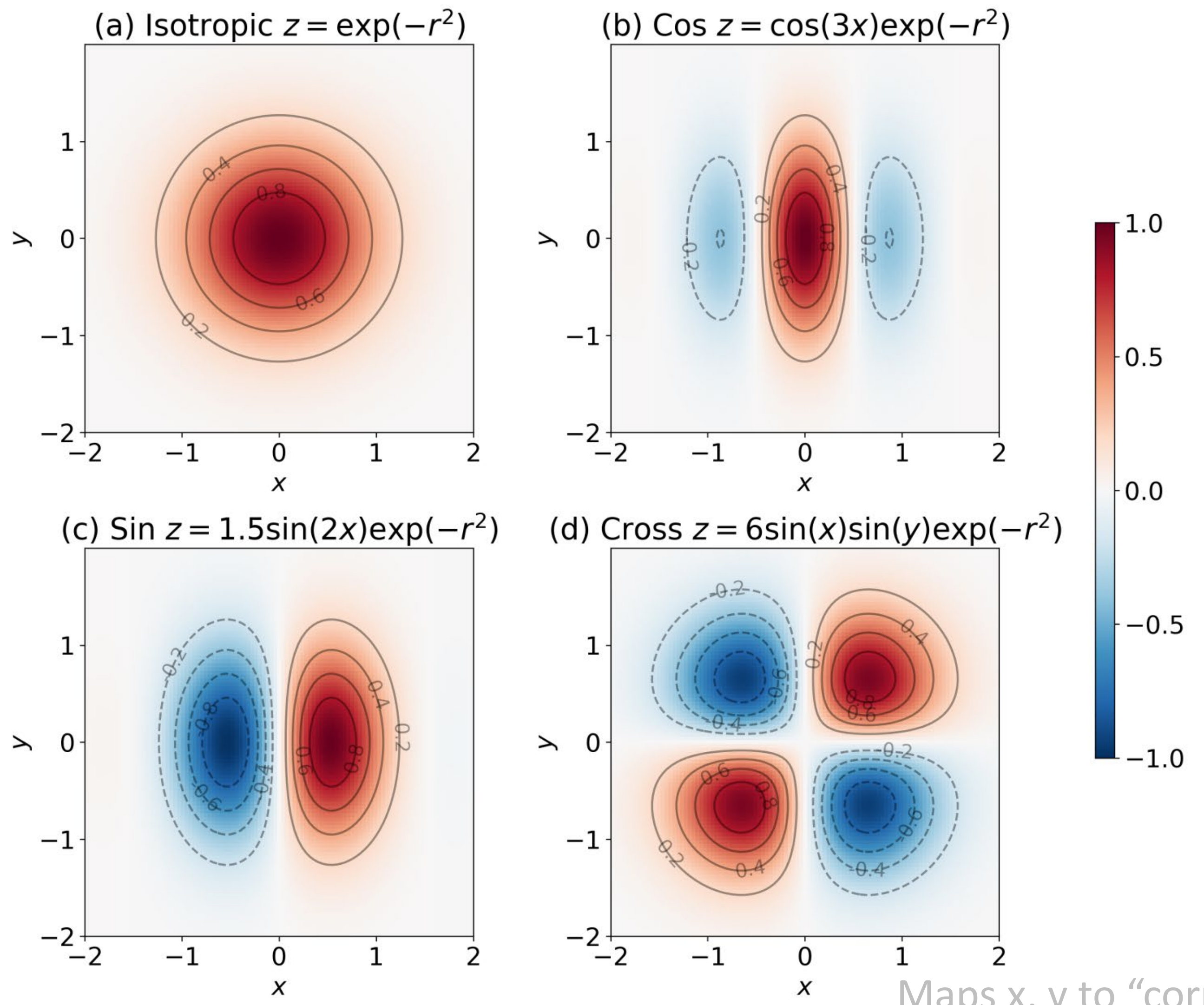
Localization weight ρ_{ij} $(0 \leq \rho_{ij} \leq 1)$

This can be done with neural networks and data from ensemble DA

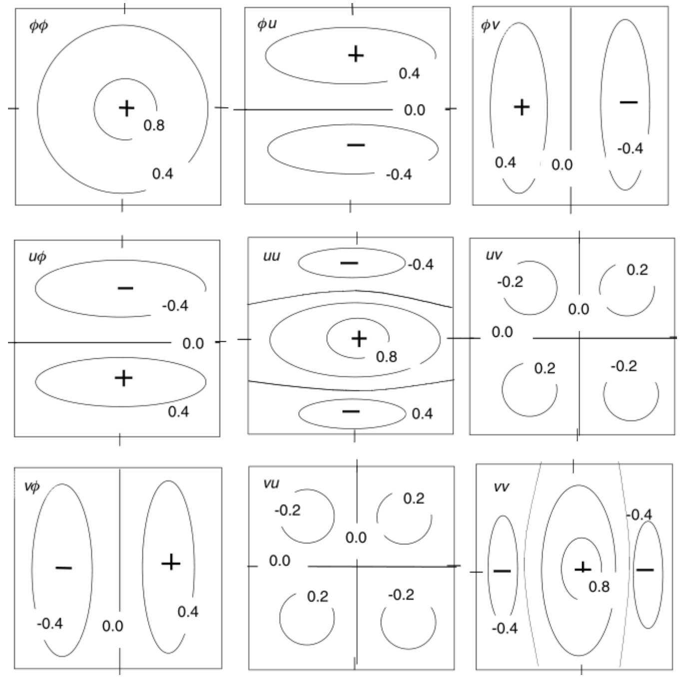
Univariate function to be specified

I will first show the capabilities of neural networks and then data assimilation experiments

Toy "correlation functions" for demonstration

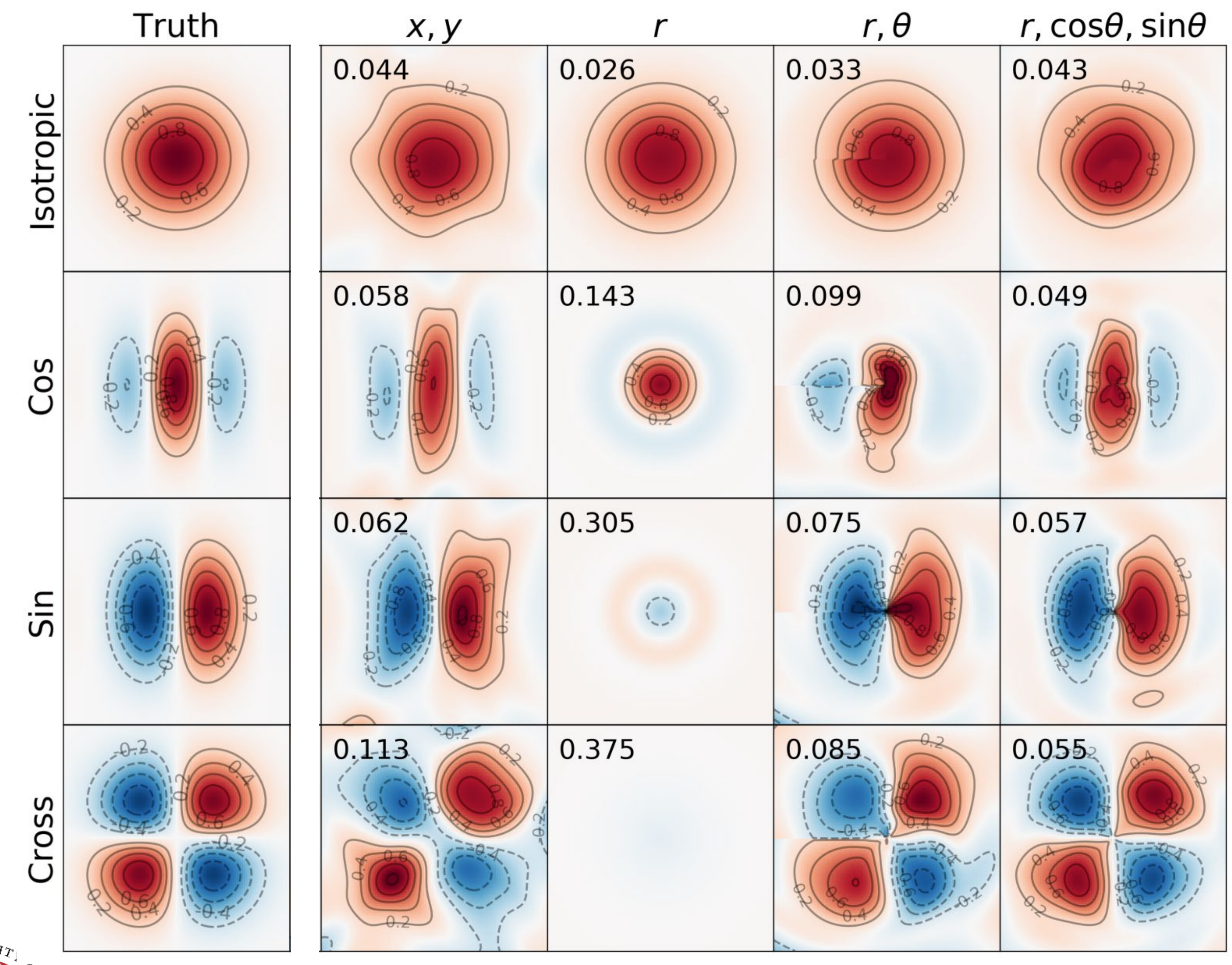


Multivariate error correlations under geostrophy



Maps x, y to "correlation"

Regression to “correlation functions”



Superimposed numbers are RMS regression error to validation dataset

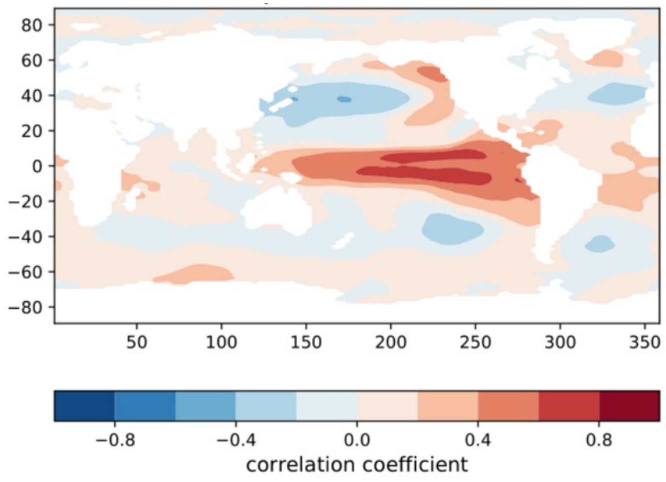
$$r = \sqrt{x^2 + y^2}$$
$$\theta = \arctan(y/x)$$

Minimal neural networks, with appropriate predictors, can reproduce spatially extended “correlation functions” from noisy training data

Fast Ocean Atmosphere Model (FOAM; Jacob 1997)

Atmospheric model (PCCM3)	
Horizontal resolution	R15 spectral (40 latitudes × 48 longitudes)
Vertical resolution	18 levels (hybrid σ-p)
Integration timestep	30 minutes 1 hour for radiation
Parameterized schemes	Convection Cloud Radiation Surface physics Vertical diffusion Gravity wave drag
Oceanic model (OM3)	
Horizontal resolution	128 latitudes × 128 longitudes (polar grid)
Vertical resolution	24 levels (z-coordinate)
Integration timestep	6 hours
Vertical mixing scheme	Bulk scheme (based on Richardson number)
Sea ice (CSIM 2.2.6)	
Horizontal resolution	128 latitudes × 128 longitudes (same as ocean)
Integration timestep	30 minutes
Modeled processes	Formation/melting Thermal conduction Snow on top Radiation
Land, Hydrology, and River Runoff	
Horizontal resolution	128 latitudes × 128 longitudes (same as ocean)
Integration timestep	30 minutes

- Low-resolution, affordable global atmosphere-ocean coupled model
- Simple processes of sea ice, land, and river are also modeled and coupled
- Despite its efficiency, it reproduces realistic climatology and natural variabilities including ENSO



El Niño naturally appearing as the first empirical orthogonal function of monthly global SST

Fitting error statistics of FOAM-LETKF

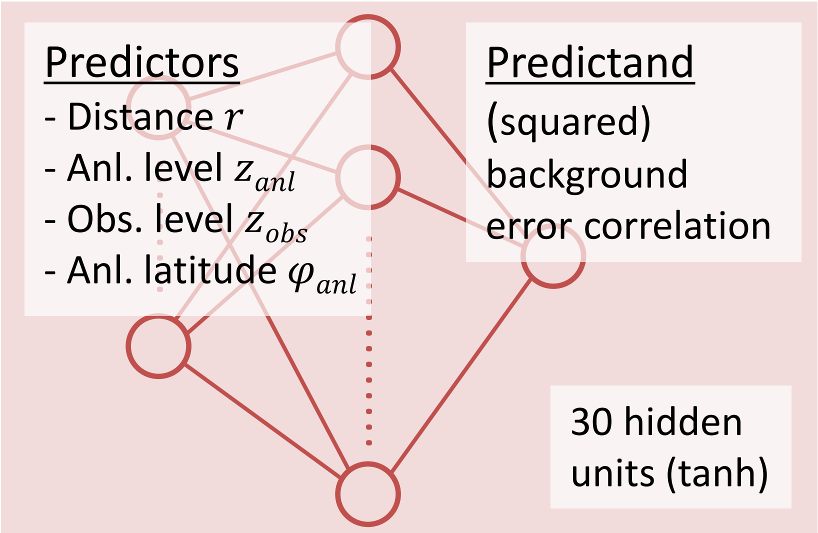
Training data

1. Run 1-year weakly coupled LETKF cycle (*offline cycle*)
2. For each pair of analysis variable and observation types, 8×10^6 training data are randomly sampled from the offline cycle

10 types of analysis variables/observations:
P_s, T, Q, U, V (atmosphere)
P_{top}, T, S, U, V (ocean)

Neural networks

For each analysis variable and observation types, a two-layer neural network is trained



Squared correlation is used for correlation-cutoff, but regression to raw error correlation is also shown for demonstration

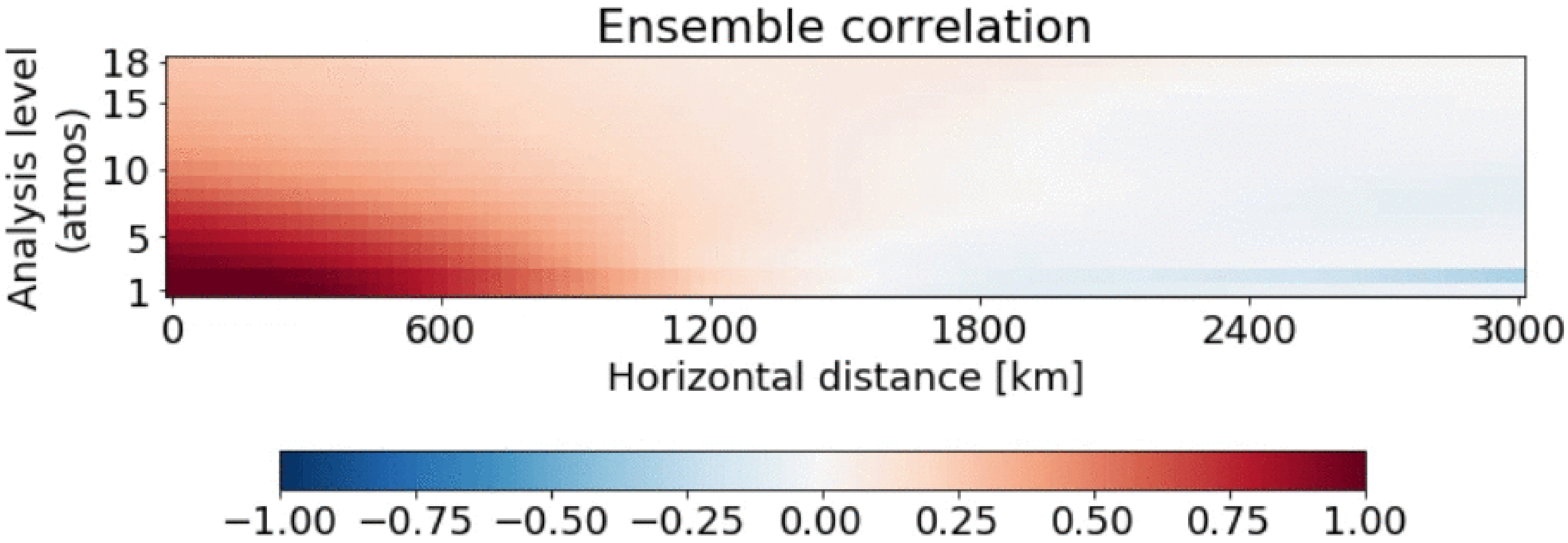
Obs elem atm_u

Anl elem atm_u

Obs level 1

Latitude -60

Squared correlation



Correlation-cutoff method for FOAM

Attributes of analysis variable x_i

- Analysis variable type
- Latitude φ_{anl}
- Altitude z_{anl}

Attributes of observable y_j

- Observation type
- Horizontal distance r (from x_i)
- Altitude z_{obs}

Localization function in observation space $g \circ f$

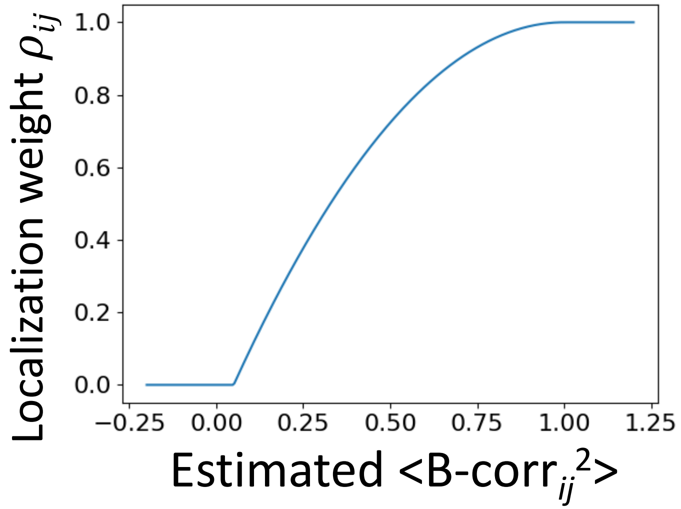
Nonlinear regression f

Expected squared background error correlation between (x_i, y_j)

Increasing function g (*cutoff function*)

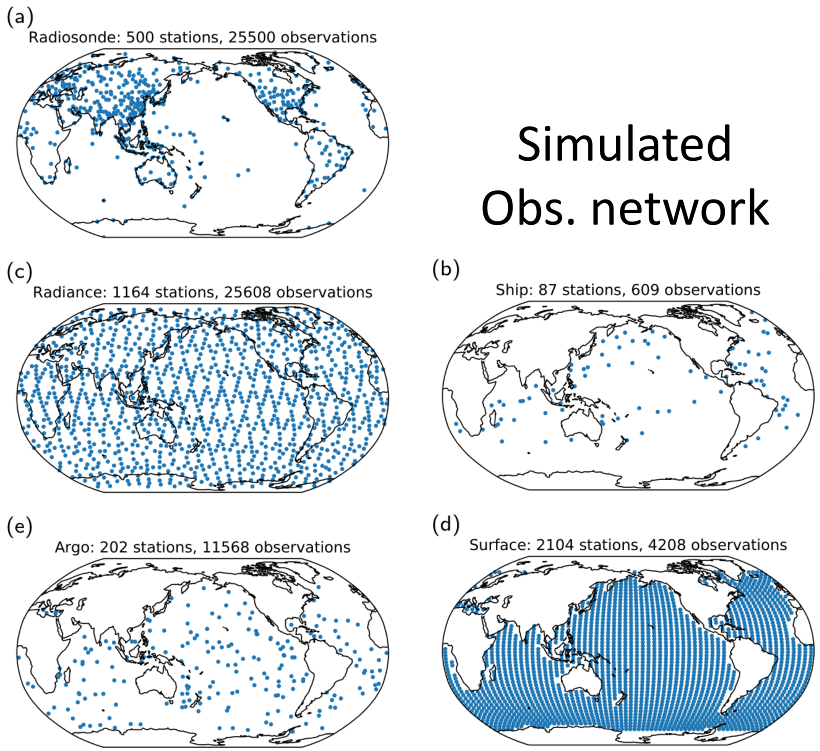
Localization weight ρ_{ij} $(0 \le \rho_{ij} \le 1)$

Neural networks discussed so far

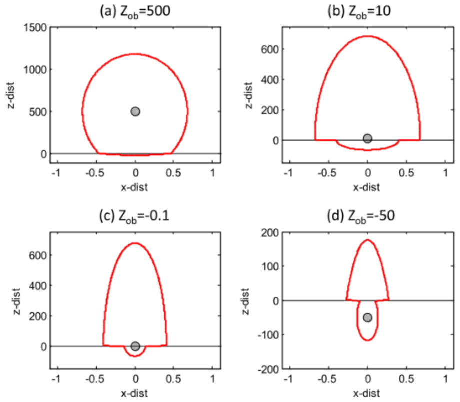


DA cycle experiments (OSSE)

	Cutoff	Control
Localization	Correlation-cutoff as in previous slide. No observations beyond 3000 km and 16 levels are evaluated	Horizontally 1000 km (atmos) and 400 km (ocean), vertically 3 model levels
Analysis	64-member strongly coupled LETKF with incremental analysis update. 24-hourly	
Covariance inflation	Relaxation to prior perturbations, 30% (atmos) and 90% (ocean)	
Analysis variables	Atmosphere: T, Q, U, V, P_s Ocean: T, S, U, V, P_{top}	
Period	One model year (from January 1) (first 30 days are omitted from evaluation)	



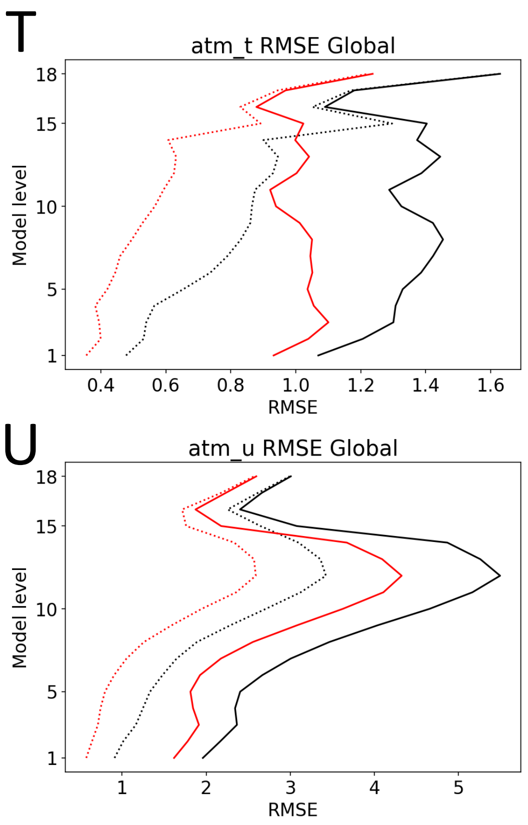
Simulated Obs. network



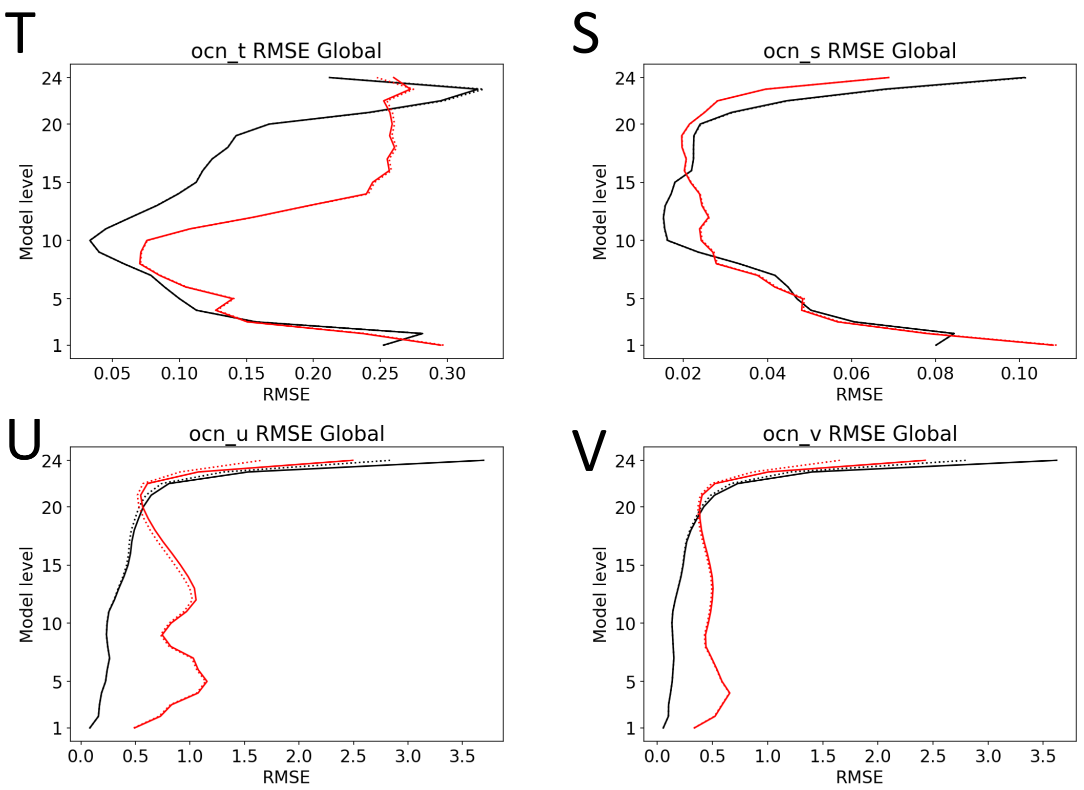
Cross-localization used for Control (Frolov et al., 2016)

Control vs Cutoff Vertical level - RMSE (time/space average)

Atmosphere: smaller error everywhere



Ocean: larger error



Better
(smaller error)

Black: Control
Red: Cutoff

Solid: Background (24h forecast) – Truth
Dotted: Analysis - Truth

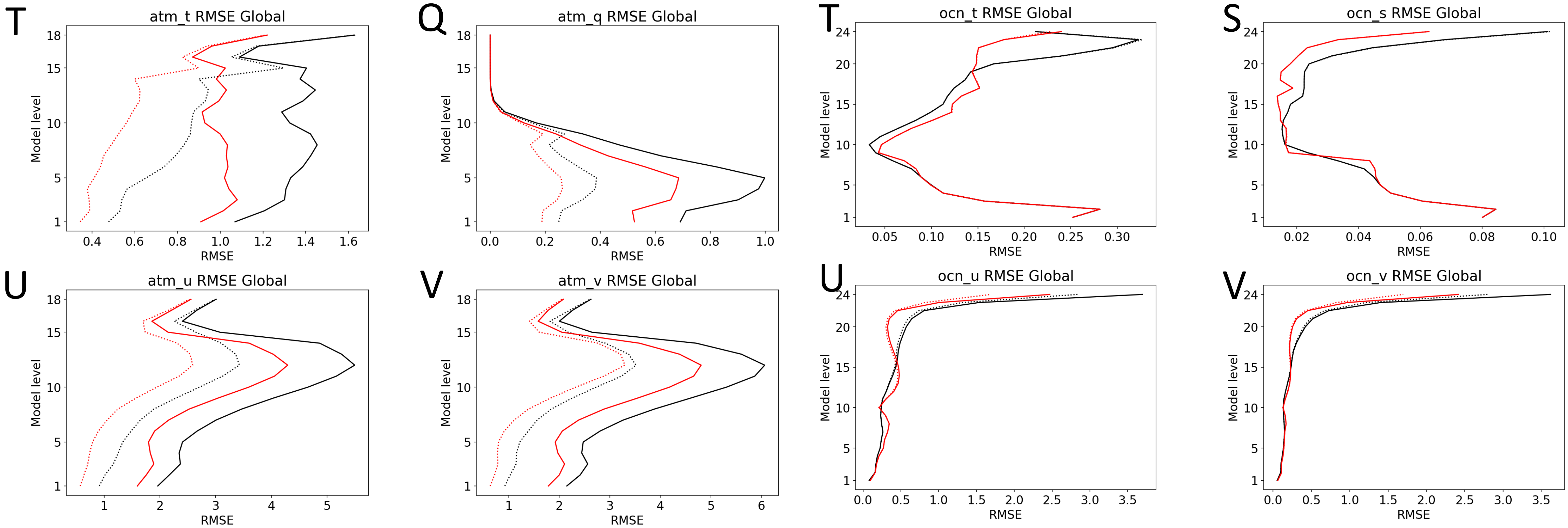
From Control vs Cutoff comparison

- Neural is consistently better than Control in the atmosphere
  - **If the training data is accurate, the correlation-cutoff method works as expected**
- Ocean, especially in deeper ocean, Cutoff is worse than Control
  - Since the 1-year offline experiment is not long enough for deep ocean to provide sufficiently independent samples, two tunings are made:
    - Ocean analysis below ~2300 m (unobserved depth) is turned off
    - Further, localization weights to ocean analysis variables are halved
  - **Cutoff-tuned** experiment

Vertical level - RMSE (time/space average)

Atmosphere: remains better

Ocean: now comparable to Control, improvements near surface



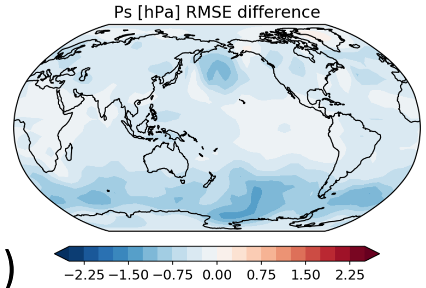
Better
(smaller error)

Black: Control
Red: Cutoff-tuned

Solid: Background (24h forecast) – Truth
Dotted: Analysis - Truth

Background (24h Forecast) errors

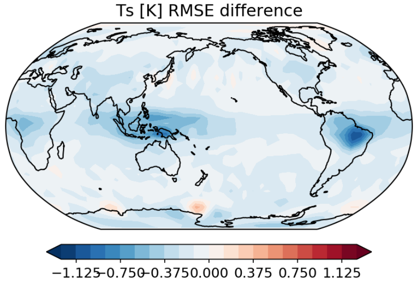
$\mathrm{RMSE}_{\mathrm{Cutoff\text{-}tuned}} - \mathrm{RMSE}_{\mathrm{Control}}$



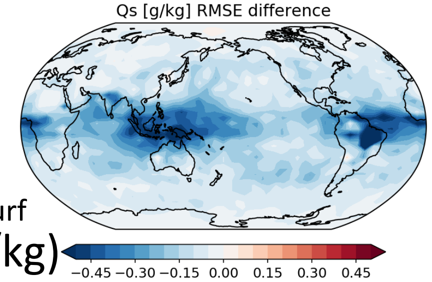
P_{surf} (hPa)

Atmosphere

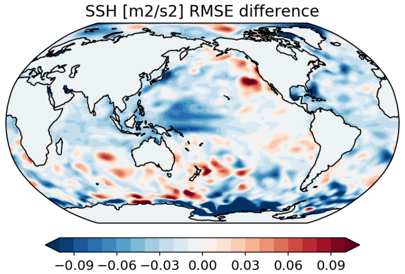
Ocean



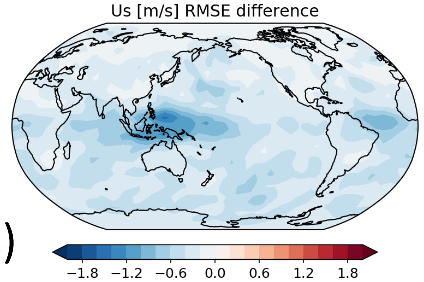
T_{surf} (K)



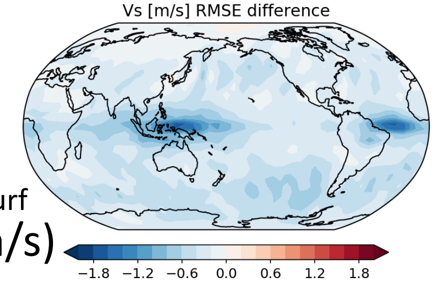
Q_{surf} (g/kg)



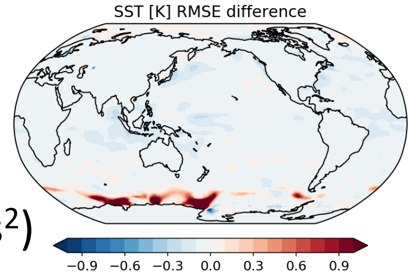
SST (K)



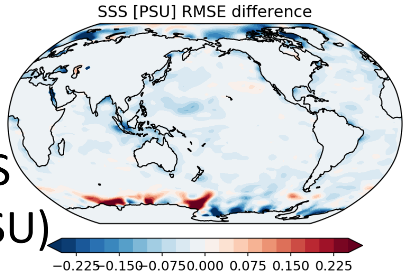
U_{surf} (m/s)



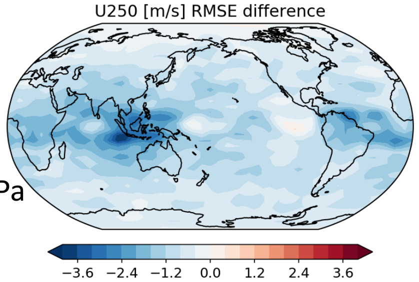
V_{surf} (m/s)



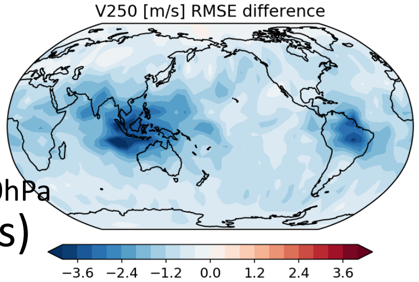
SSH (m^2/s^2)



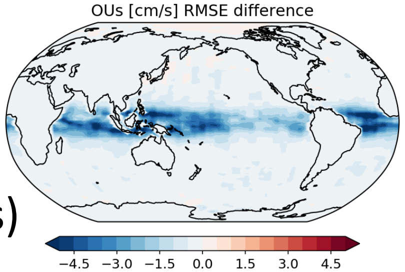
SSS (PSU)



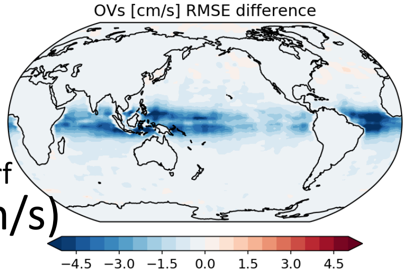
U_{250hPa} (m/s)



V_{250hPa} (m/s)



U_{surf} (cm/s)



V_{surf} (cm/s)

Conclusions (localization modeling with neural networks)

- Correlation-cutoff method works well in global atmosphere-ocean DA
  - We employed neural networks for a generic nonlinear regression method
  - Substantial improvement in the atmosphere for every level and variables. Largest improvement is in the tropics, where variable localization between mass and wind fields should be important
  - Internal ocean needed tuning possibly because its timescale is longer than the offline experiment
- Evaluation of neural networks takes ≤10% of analysis time
  - Indirect increase of computation inherent to variable localization exists
- We have also shown mathematical validity of the method

Future directions of correlation-cutoff method

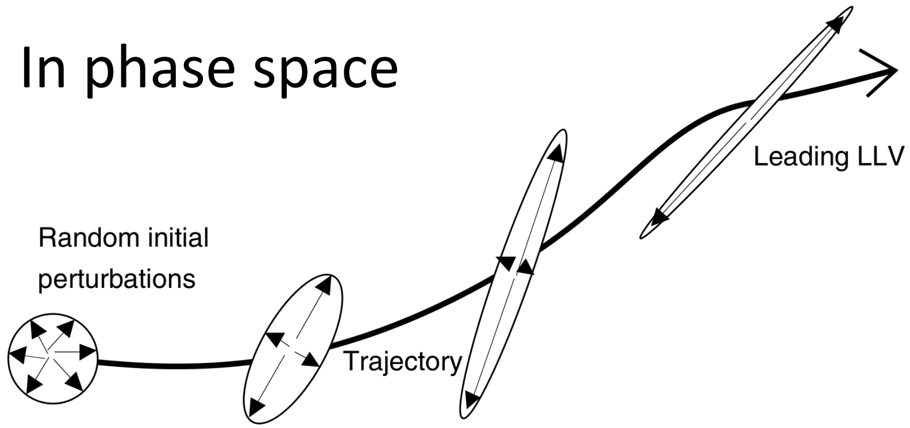
- Better understanding and representation of ocean
  - Combined with tuned distance-based localization
  - Longer DA cycle used for sampling the training data
  - Detection of “unreliable” statistics used in the training
- Thorough experiments with smaller models
  - Better cutoff function and its theoretical optimum
  - Dynamic balance of analysis
  - Iterative or online update of localization function
- Application to more realistic configurations and Earth system models

Outline

0. Introduction
1. Correlation-cutoff method and experiments with 9-variable coupled model
2. Localization modeling with neural networks – successful assimilation experiments with global atmosphere-ocean coupled model
3. Sudden and major change of dynamics found in coupled chaotic systems
4. Summary and future directions

Lyapunov exponents

In phase space



- Atmosphere and ocean are said to be *chaotic* when small error in the initial conditions will grow exponentially
- *Lyapunov exponents* are the long-term average growth rate (or decay rate) of errors to the first order
  - Chaotic dynamical systems have positive Lyapunov exponents
  - Can be estimated numerically

Figure courtesy Kalnay (2003)

ENSO-type model of Peña and Kalnay (2004)

(fast)
Tropical
atmosphere

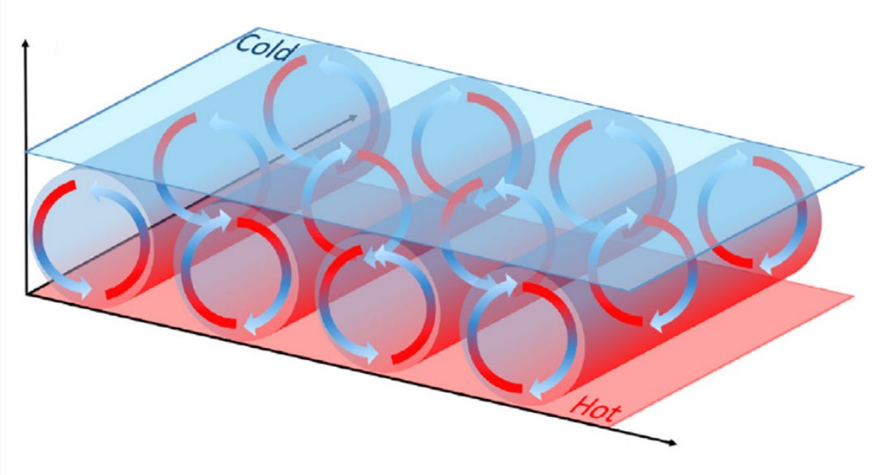
$$\dot{x}_t = \sigma(y_t - x_t) - \alpha c(SX + k_2)$$
$$\dot{y}_t = r x_t - y_t - x_t z_t + \alpha c(SY + k_2)$$
$$\dot{z}_t = x_t y_t - b z_t + \alpha c_z Z$$

(slow)
Ocean

$$\dot{X} = \tau\sigma(Y - X) - \alpha c(x_t + k_2)$$
$$\dot{Y} = \tau r X - \tau Y - \tau S X Z + \alpha c(y_t + k_2)$$
$$\dot{Z} = \tau S X Y - \tau b Z - \alpha c_z z_t.$$

- Convective motion (x_t, X) and thermal gradients (y_t, Y) and (z_t, Z) are coupled respectively
- $\alpha = 1$ in the original coupled model

Each Lorenz subsystem represents chaotic convective fluid



x_t, X: Rotational convective motion

y_t, Y: Horizontal T difference between ascending and descending currents

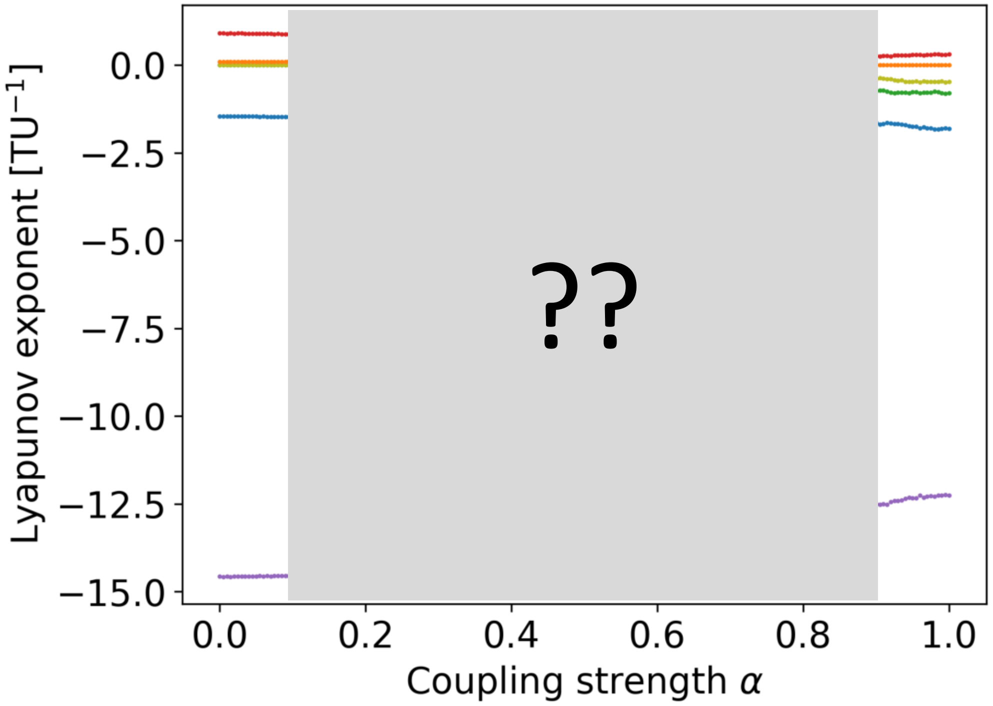
z_t, Z: Vertical T profile distortion

Figure courtesy Barna et al. (2017)

Uncoupled vs coupled models

In uncoupled limit, the 6 exponents originate from atmos/ocean

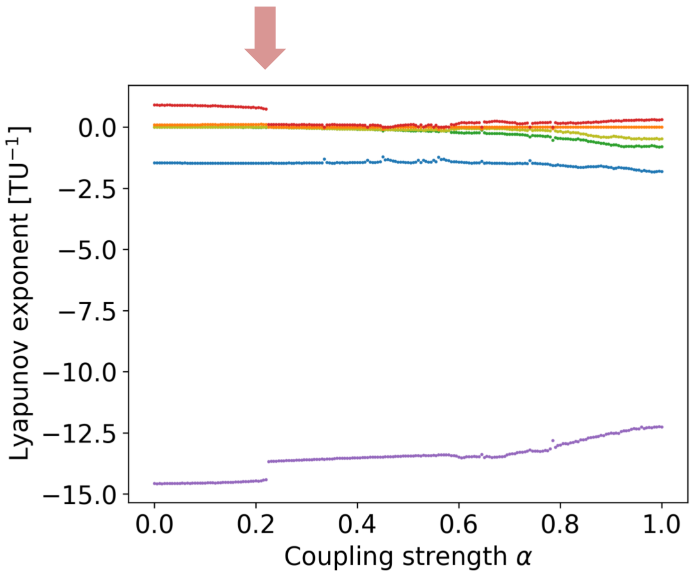
0.91 (atmos)
0.091 (ocean)
two 0's (atmos/ocean)
-1.5 (ocean)
-15 (atmos)

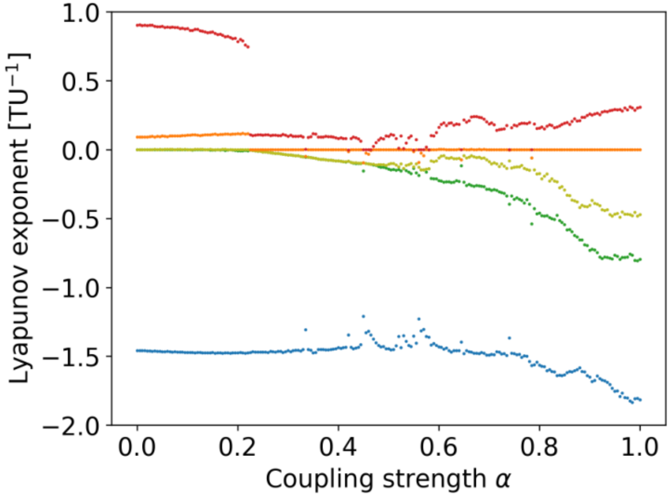


In coupled limit, the model's 6 exponents are

0.32
0
-0.47
-0.79
-1.8
-12

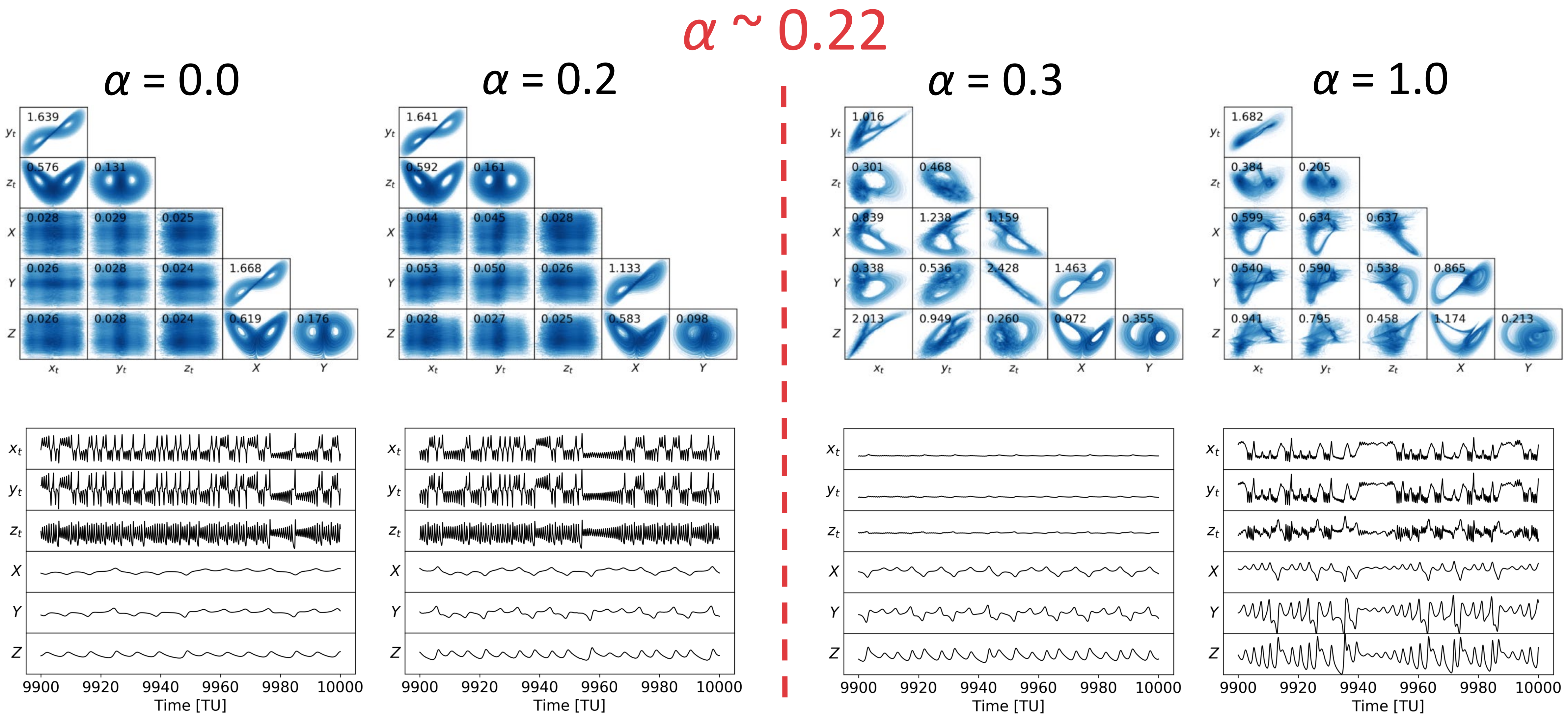
Series of models coupled incrementally





- Sudden change of Lyapunov spectrum at $\alpha \sim 0.22$
  - (From left to right,) largest positive exponent originating from the “Tropical atmosphere” suddenly disappears and another near-zero mode appears
  - With $\alpha < 0.22$, the model is qualitatively like the uncoupled model: continuous change of Lyapunov spectrum from $\alpha = 0$, two neutral exponents corresponding to temporal translation
  - With $\alpha > 0.22$, the model behaves as if an integrated chaotic system: with single neutral exponent corresponding to temporal translation

Attractors before/after the critical coupling strength



Bottom: The ordinate ranges shown are x_t, y_t ∈ [−25,25]; z_t ∈ [0,60]; X, Y ∈ [−100,100]; and Z ∈ [−70,130].

Discussion – change of dynamics in coupled systems

- A few parameter bifurcations exist, and we cannot trace Lyapunov exponents even if we continuously change the coupling strength
  - Similar to "*synchronization of chaos*" observed for coupled two or more similar dynamical systems
- This physically means that coupling can qualitatively stabilize some modes (and probably also destabilize in other systems)
  - Possibilities for regime changes, for example, the momentum coupling strength of mixed and boundary layers depends on vertical stratifications
  - Parameter estimation can be difficult due to discontinuous relationship between observed quantity and parameter
  - Severe misrepresentation of uncertainty with mis-specified model/parameter is possible

Outline

0. Introduction
1. Correlation-cutoff method and experiments with 9-variable coupled model
2. Localization modeling with neural networks – successful assimilation experiments with global atmosphere-ocean coupled model
3. Sudden and major change of dynamics found in coupled chaotic systems
4. Summary and future directions

Summary of thesis

1. Correlation-cutoff method – experiments with 9-variable coupled model
   - We have obtained the metric of "relevance" between analysis variables and observations
   - Our correlation-cutoff method improved the analysis of coupled 9-variable model
2. Localization modeling with neural networks – successful assimilation experiments with global atmosphere-ocean coupled model
   - Neural network enabled the implementation of the correlation-cutoff method to global atmosphere-ocean coupled model with realistic computation cost
   - Proof-of-concept experiments showed improved atmospheric analysis, especially in the tropics
   - We have tuned ocean analysis, and suggestions are made for further improvements
3. Sudden and major change of dynamics found in coupled chaotic systems
   - Dynamics of coupled model can discontinuously depend on the coupling strength
   - Implications for uncertainty estimate, regime change, and parameter estimation

Future directions of coupled DA

- Apply these methods to strongly coupled CFS-LETKF in collaboration with Eugenia Kalnay, Travis Sluka, and UMD-AOSC students
- Error growth of coupled system should be more thoroughly examined
  - E.g., resolution dependencies of error growth and stability of uncoupled/weakly coupled DA cycles
  - We have partially tackled this problem in another collaborative work (Penny et al. 2019)
- Data assimilation does not only improve state estimate but also enables
  - Detection and correction of model error (e.g., Bhargava et al., 2018)
  - Estimation of observation impact on forecasts (e.g., Chen and Kalnay, 2019)

⇒ Coupled DA will provide these by-products for coupled predictions

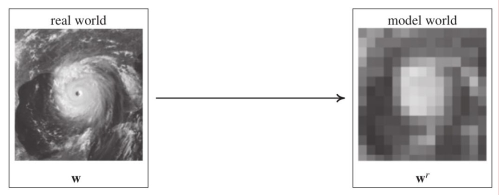
Acknowledgements

- Prof. Eugenia Kalnay for guiding me in every aspect of research
- Prof. Stephen G. Penny for sharing best knowledge on coupled DA
- Profs. James A. Carton, Brian Hunt, and Dr. Safa Motesharrei for valuable feedback
- Ms. Tammy Hendershot for making all the necessary arrangements
- Prof. Kayo Ide for organizing Weather and Chaos meetings and other events
- Drs. Yun Liu, Travis Sluka, Sreenivas Pentakota, Cheng Da, and Eviatar Bach for programs and data
- Drs. Takemasa Miyoshi and Daisuke Hotta for being my role models
- Drs. Tse-Chun Chen and Kriti Bhargava, Luyu Sun, and Chu-Chun Chang (and people mentioned above) for discussions
- Open source projects and Department of IT for infrastructure
- Japanese Government, Burgers program, and Ann Wylie Green fund for financial support
- Many people helped our personal lives in US
- My wife Rie for patience and encouragements, my son Itsuki for peace of minds

Backup slides

Localization and assimilation of averaged observations
(analogy to the superobbing)

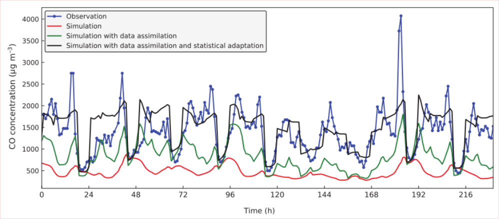
Spatial representation error



Countermeasure 1
Inflate **R** matrix so that it includes representation error as well as instrument error

Countermeasure 2
Spatial average of observations (superobs)

Temporal representation error



Countermeasure 1
Inflate **R** matrix so that it includes temporal representation error as well as instrument error (Equivalent to R-localization of LETKF)

Countermeasure 2
Temporal average of observations
(Huntley+ 2010, Tardif+ 2014/2015, Lu+ 2015ab)

Relative advantages

✓We can use the same analysis interval for fast/slow systems
✓Observation network can be nonstationary

✓ We can save analysis cost

Bottom line: both localization and averaged observations can handle the same problem.

Images from Janjic et al. (2018)

Computation cost is acceptable

- Sampling
  - 1E+9 pairs *total* for 100 pairs of variable types
  - Several hours with a single processor. Parallelizable
- Learning
  - 8E+6 samples × 3 epochs for *each* pair of observation and analysis variables
  - Tens of minutes with a single processor. Parallelizable

Except for IO, the training cost will be almost independent of model resolution.

- Evaluation of a neural network
  - O(100p) floating point computation for each analysis variable. This is less than LETKF's cost O($k^3 + pk^2$) (p: # of local observations, k: ensemble size)

Why neural network?

Method	Advantages	Disadvantages
Linear regression	Simple to implement Training is analytical	Linear
Lookup table	Nonlinear Simple to implement Training is analytical Fast to evaluate	Discontinuous Assumptions for boundaries Curse of dimensionality
Linear combination of nonlinear basis functions (e.g., polynomial fit)	Nonlinear Training is analytical	Assumptions for basis functions Curse of dimensionality
Neural network	Nonlinear Fewer assumptions Relatively tolerant of input dimensionality	Training requires iteration
Gaussian processes regression	Access to uncertainty Nonlinear Fewer assumptions	More expensive training

- We need less prior knowledge than other regression methods
  - With recent advancement of its methodology, it can be used as an almost end-to-end method
- Fast to evaluate once trained