Transitioning to strongly coupled data assimilation for Earth system initialization

Prof. Stephen G Penny
University of Maryland College Park

EnKF Workshop, Norway - 3 June 2019

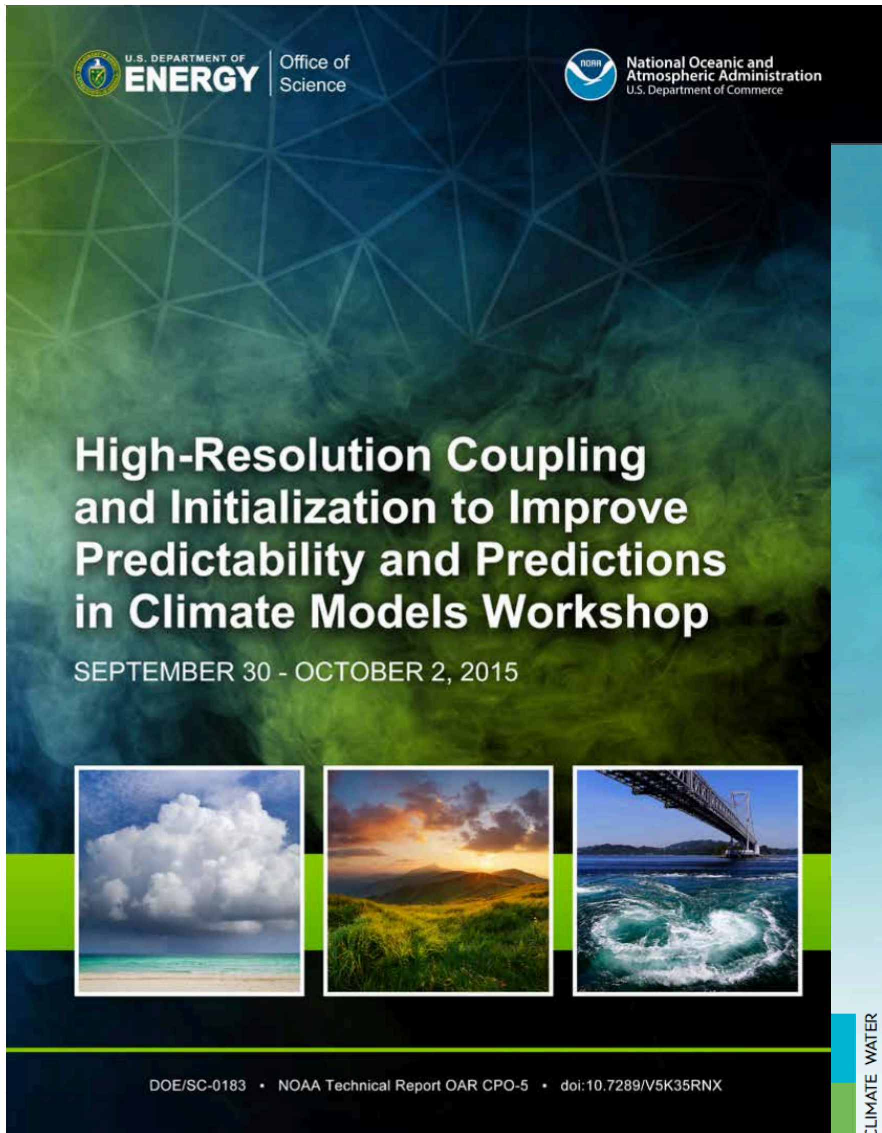
Overview

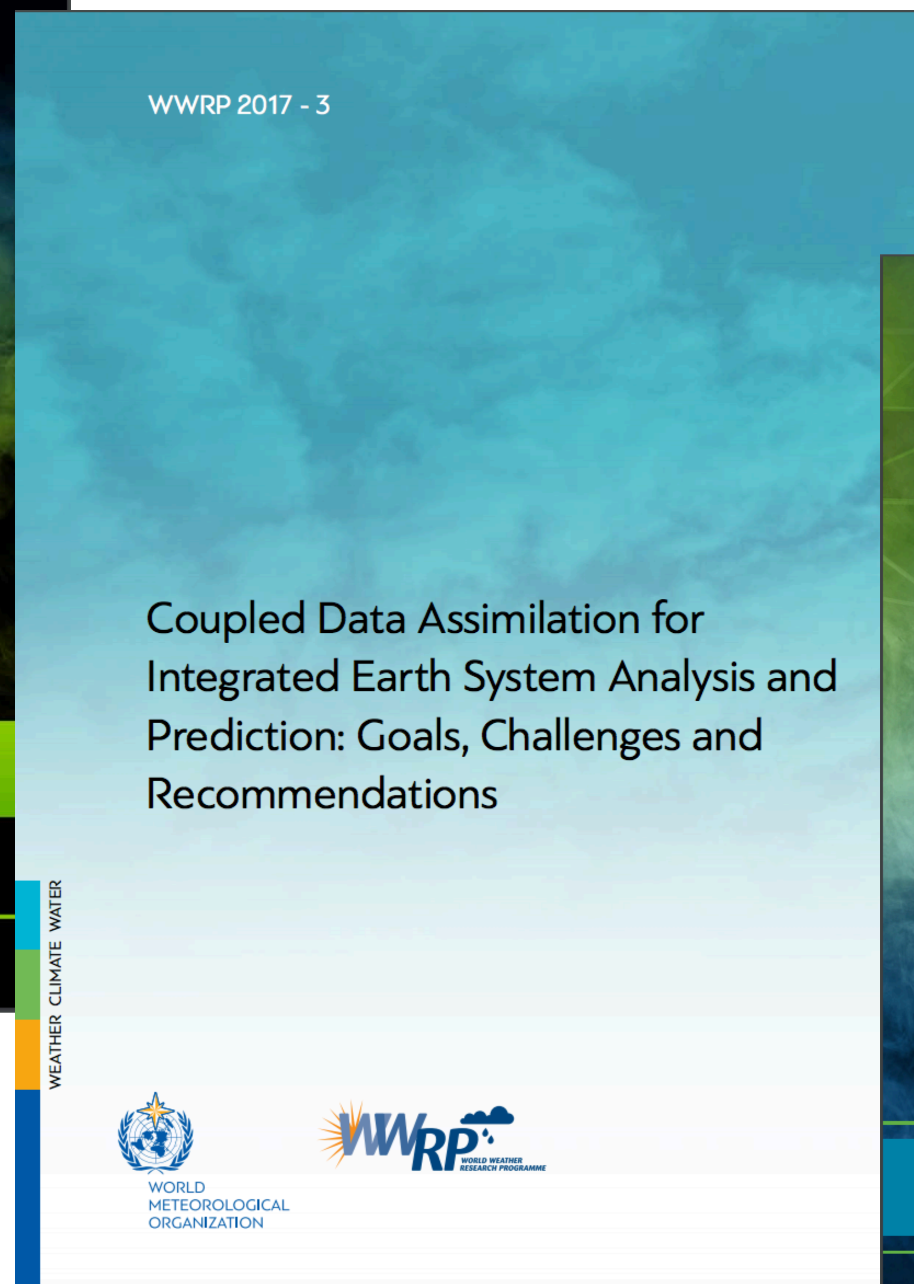
- Brief background
- Motivation for Coupled Data Assimilation (CDA)
- Prior results using Strongly Coupled Data Assimilation (SCDA)
- Our results using SCDA with a simple coupled QG model
- Extending to more realistic systems

Overview

- Brief background
- Brief Introduction to DA from my perspective
- Motivation for Coupled Data Assimilation (CDA)
- Prior results using Strongly Coupled Data Assimilation (SCDA)
- Our results using SCDA with a simple coupled QG model
- Extending to more realistic systems

Workshops on Earth system model initialization





Overview

- Brief Bio/background
- Motivation for Coupled Data Assimilation (CDA)
- Prior results using Strongly Coupled Data Assimilation (SCDA)
- Our results using SCDA with a simple coupled QG model
- Extending to more realistic systems

Motivation for CDA

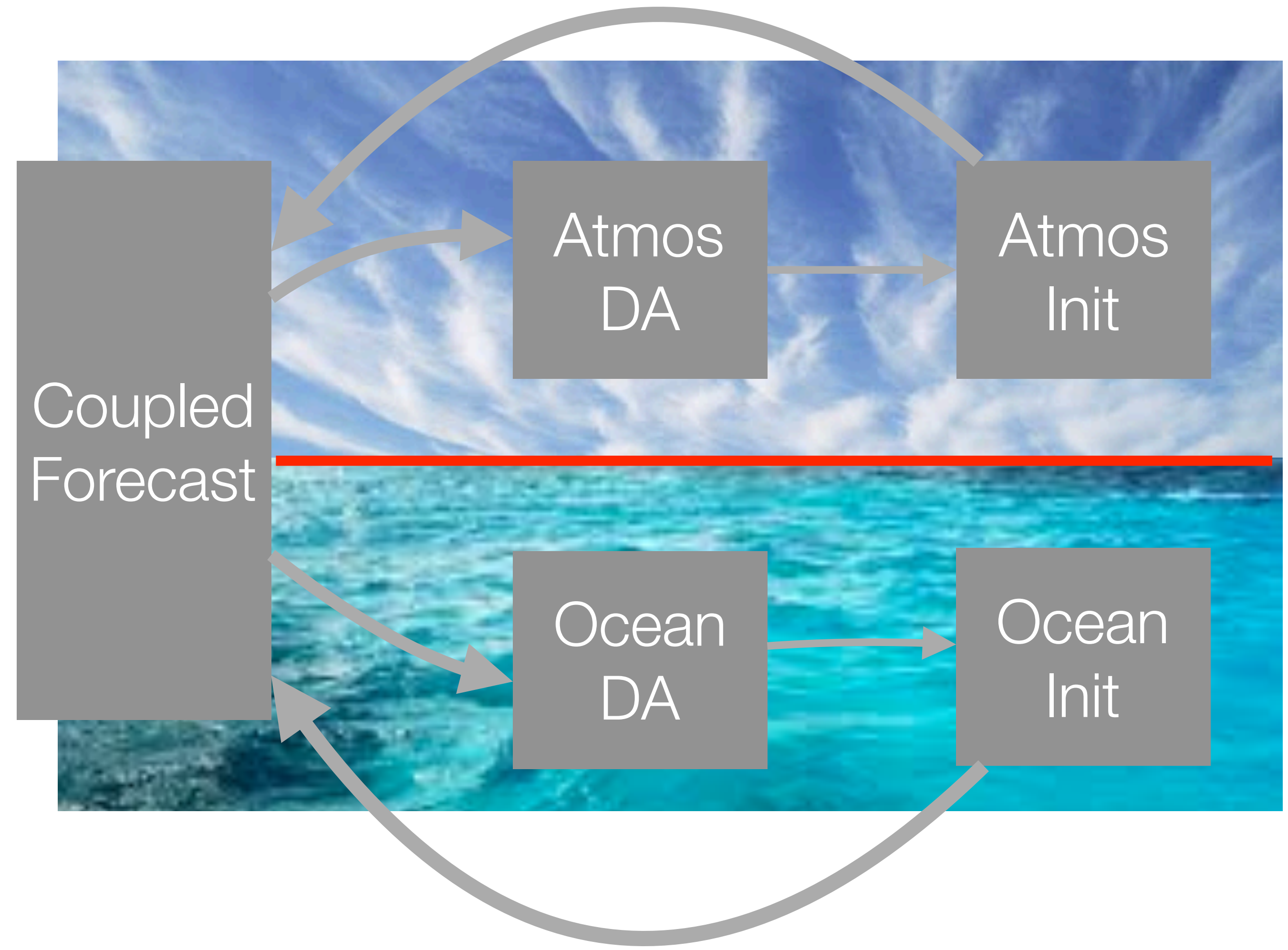
- Coupled data assimilation (CDA) is characterized by the use of a coupled forecast model, but more generally focuses on the assimilation of information from multiple spatiotemporal scales, often derived from different components of the Earth system.

- Weakly coupled DA (WCDA) allows information to be transferred between scales via the forward model integration

- Strongly coupled DA (SCDA) attempt to transfer information instantaneously at the analysis time, and also in the model

Aside - definitions

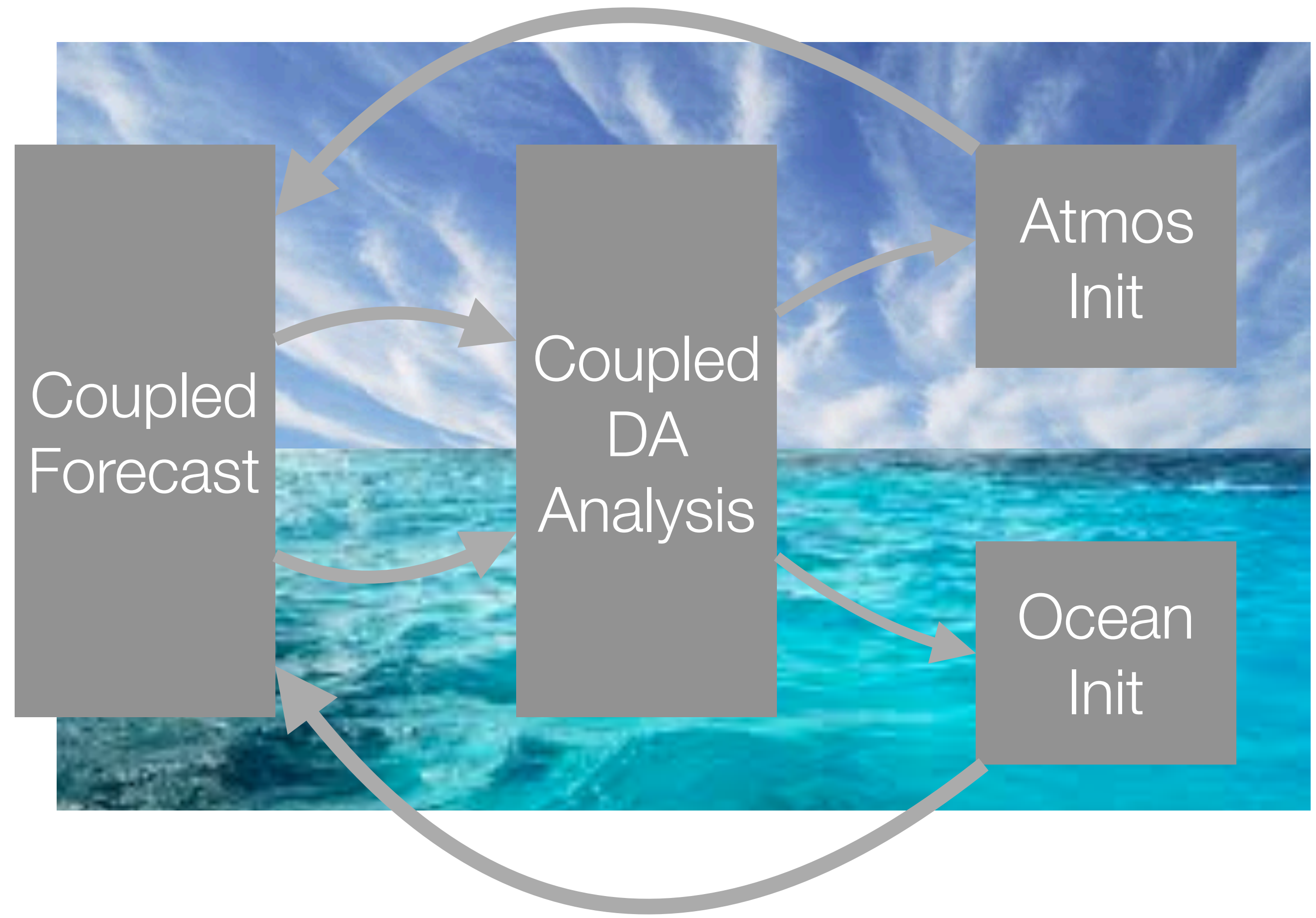
- Weakly coupled data assimilation (WCDA) means -
- Strongly coupled data assimilation (SCDA) means -
- At this point, when I discuss ‘Coupled Data Assimilation’ (CDA), I implicitly refer to SCDA.



Weakly Coupled Data Assimilation

Aside - definitions

- Weakly coupled data assimilation (WCDA) means -
- Strongly coupled data assimilation (SCDA) means -
- At this point, when I discuss 'Coupled Data Assimilation' (CDA), I implicitly refer to SCDA.



Strongly Coupled Data Assimilation

Overview

- Brief Bio/background
- Motivation for Coupled Data Assimilation (CDA)
- Prior results using Strongly Coupled Data Assimilation (SCDA)
- Our results using SCDA with a simple coupled QG model
- Extending to more realistic systems

A review of SCDA applied to Simple models

- Han et al. (2013): *Lorenz atmosphere and a pycnocline ocean model*
    - “Results show that it **requires a large ensemble size to improve the assimilation quality** by applying coupling error covariance in an ensemble coupled data assimilation system… It is also found that **a fast-varying medium has more difficulty being improved using observations in slow-varying media by applying coupling error covariance** because the linear regression from the observational increment in slow-varying media has difficulty representing the high-frequency information of the fast-varying medium.”

A review of SCDA applied to Simple models

- Liu et al. (2013): *Lorenz atmosphere and Jin ocean model*
  - **SCDA that assimilates observations in both the atmosphere and ocean and that employs the coupled covariance matrix outperforms the WCDA alternative**.
  - Assimilation of synoptic atmospheric variability was critical for the improvement of both the atmospheric state and the oceanic state through coupled covariance, especially in the midlatitude system
  - **The assimilation of synoptic atmospheric observation alone improved the coupled state almost as much as assimilating additional oceanic observations**, while the assimilation of oceanic observations had little impact on the atmosphere.

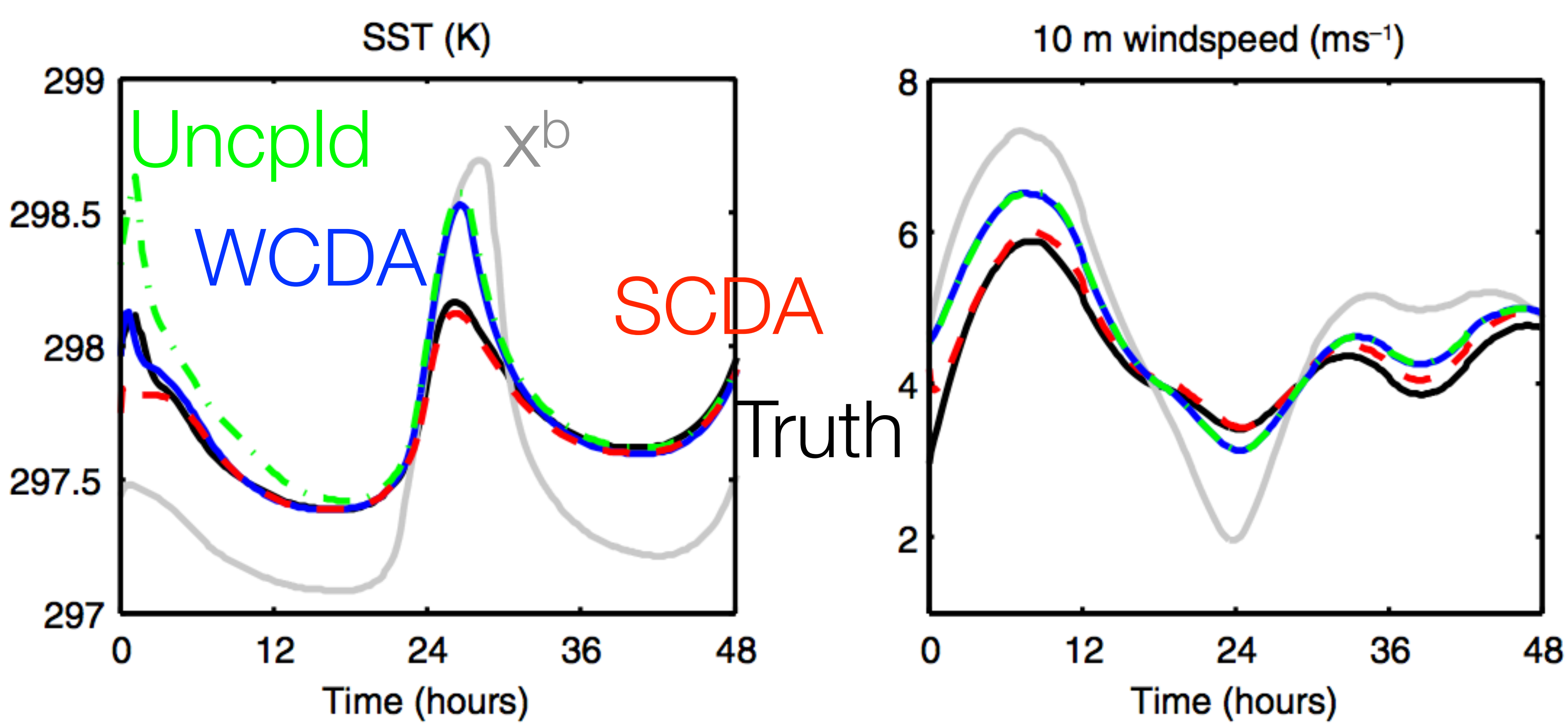
A review of SCDA applied to Simple models

- Tardif et al. (2014): *Lorenz (1984) atmosphere and Stommel 3-box ocean model*
  - Forcing the idealized ocean model with atmospheric analyses is inefficient at recovering the slowly evolving MOC
  - Daily assimilation rapidly leads to accurate MOC analyses, provided a comprehensive set of oceanic observations is available for assimilation
  - **In the absence of sufficient observations in the ocean, the assimilation of time-averaged atmospheric observations proves to be more effective for MOC initialization** than either forcing the ocean or assimilating sparse ocean observations.

A review of SCDA applied to Simple models

- Smith et al. (2015): *idealized single-column atmos/ocean model*
  - Incremental 4D-Var - “When compared to uncoupled initialisation, **coupled assimilation is able to produce more balanced initial analysis fields, thus reducing initialisation shock and its impact on the subsequent forecast**.”

Forecasts:



A review of SCDA applied to Simple models

- Smith et al. (2017): *idealized single-column atmos/ocean model*
  - "**consider cross correlations rather than cross covariances** because different components of the coupled state vector have very different levels of variability; standardizing prevents variables with large error variances from dominating the structure of the covariance matrix”
  - “Within the boundary region there is **notable variation in the strength and structure of the error cross correlations** between **summer and winter**, and between **day and night**. "
  - “atmosphere–ocean forecast error cross correlations are very state and model dependent…the static **B** formulation assumed in traditional 4D-Var may not be sufficient”

A review of SCDA applied to Simple models

- Smith et al. (2018): *idealized single-column atmos/ocean model*
  - “compare methods for improving the rank and conditioning of multivariate sample error covariance matrices for [CDA].”
  - “The first method, **reconditioning, alters the matrix eigenvalues directly**; this preserves the correlation structures but does not remove sampling noise."
  - “The second method, **model state-space localization via the Schur product, effectively removes sample noise but can dampen small cross-correlation signals**.”

A review of SCDA applied to Intermediate Complexity models

- Lu et al. (2015): *FOAM Low resolution Earth system GCM*
  - The use of time-averaged surface temperature observations was necessary for SCDA to outperform WCDA, otherwise SCDA performed worse than WCDA in the midlatitudes
  - Results may have been influenced by the small ensemble size (16), coarse model grid (7.5° x 4.5° atmosphere and 2.8° x 1.4° ocean), and use of monthly SST data

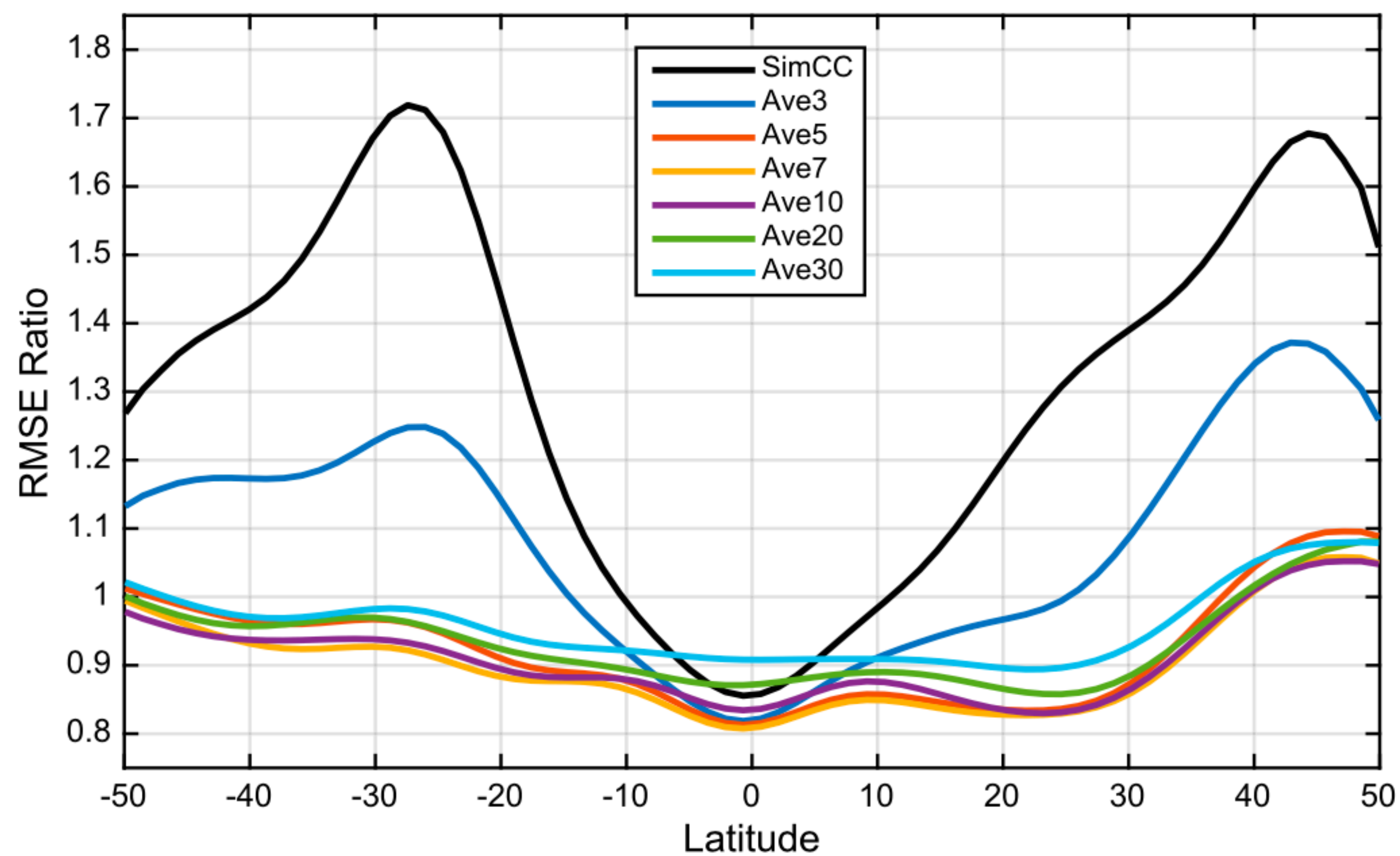


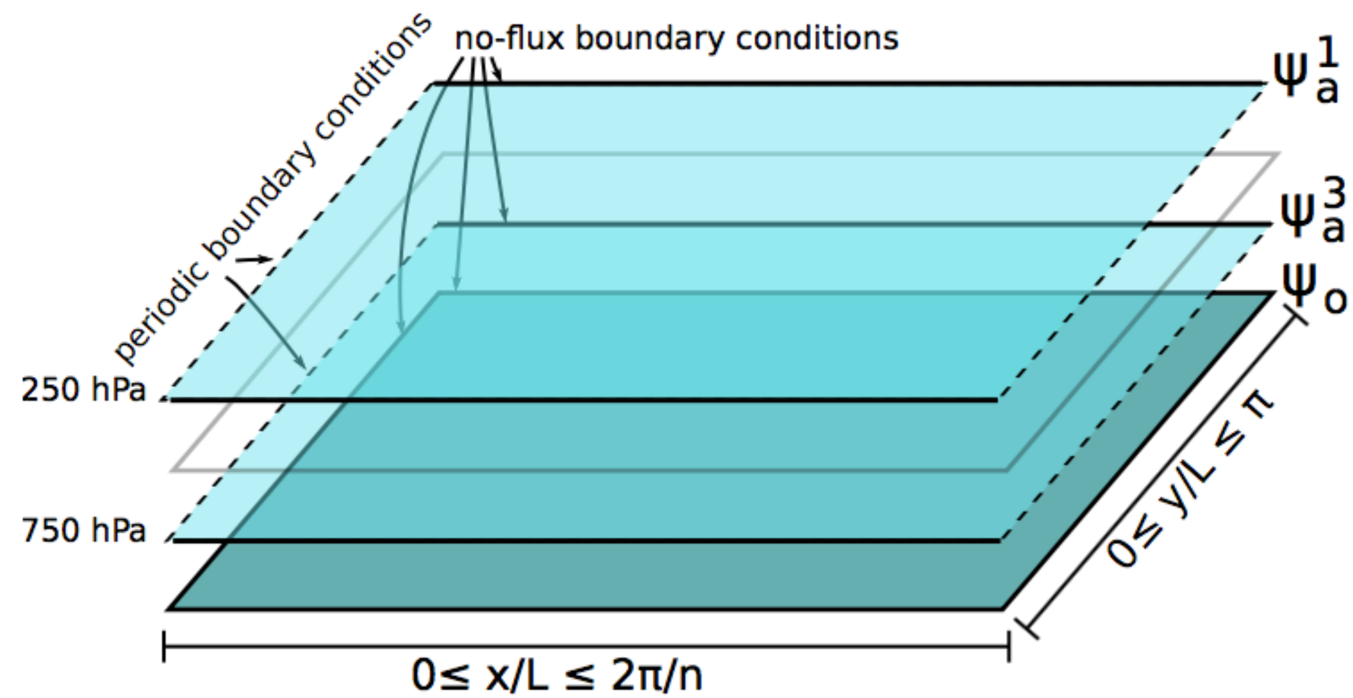
FIG. 5. Zonal-mean RMSE of monthly SST from the SimCC experiment and the LACC experiments with different averaging lengths, normalized by the WCDA experiment.

Overview

- Brief Bio/background
- Motivation for Coupled Data Assimilation (CDA)
- Prior results using Strongly Coupled Data Assimilation (SCDA)
- Our results using SCDA with a simple coupled QG model
- Extending to more realistic systems

Modular Arbitrary Order Ocean Atmosphere Model (MAOOAM)

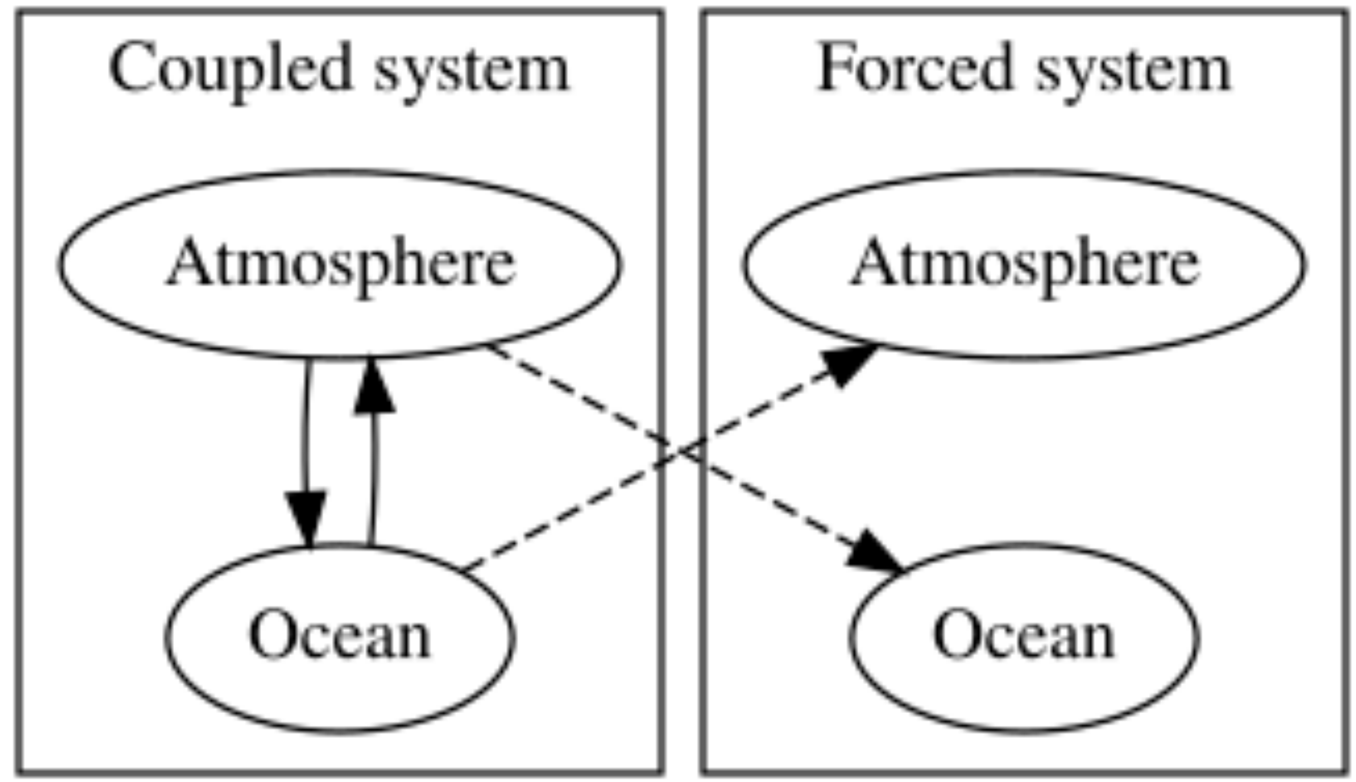
- Truncated QG model
- 2-layer atmosphere (fast component), 1-layer ocean (slow component)
- Coupled dynamics and thermodynamics
- Tangent Linear Model (TLM) available for investigation of Lyapunov exponents and experimentation with 4D-Var



De Cruz et al. (2016)
Vannitsem and Lucarini (2016)

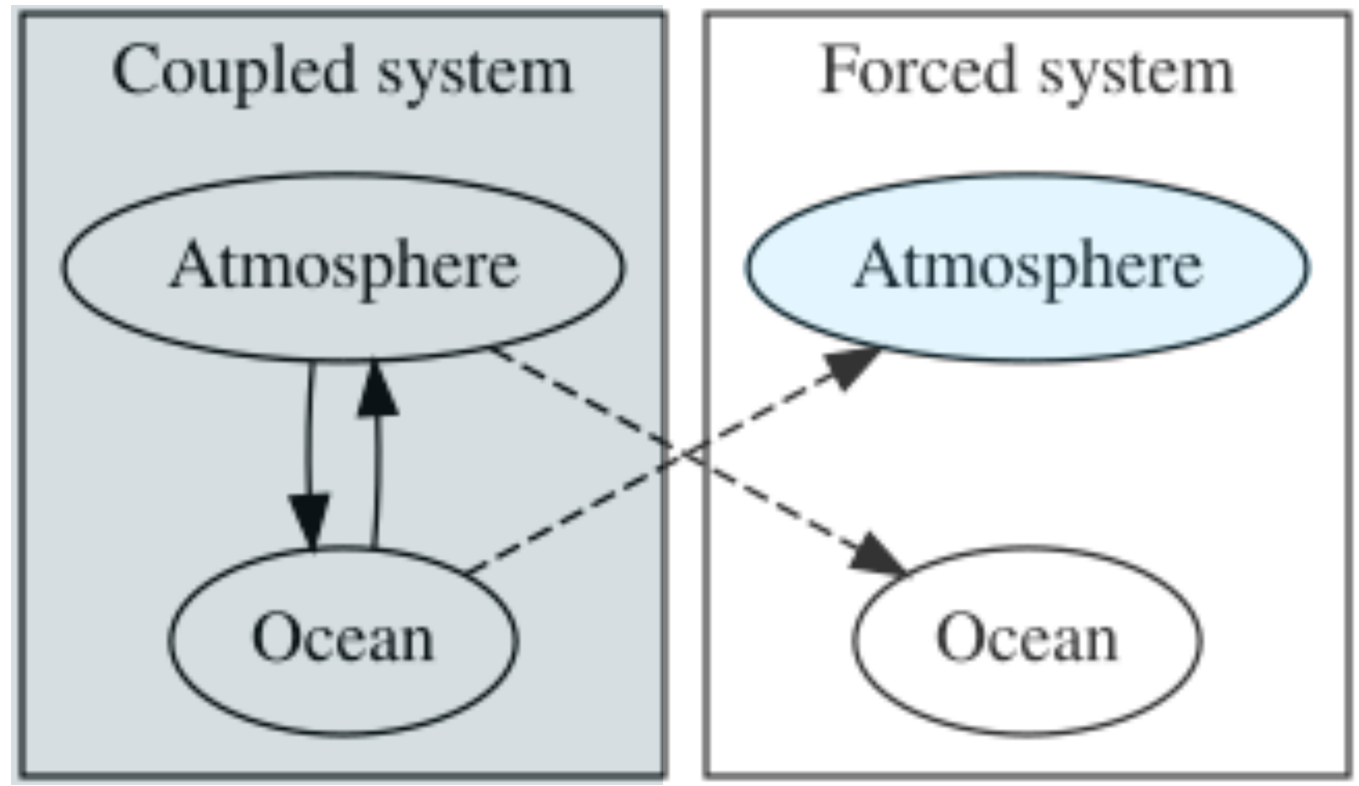
Examining the forced and coupled systems

- We examine:
  - Atmosphere forced by the coupled ocean state
  - Ocean forced by the coupled atmospheric state
  - Fully coupled modeling system



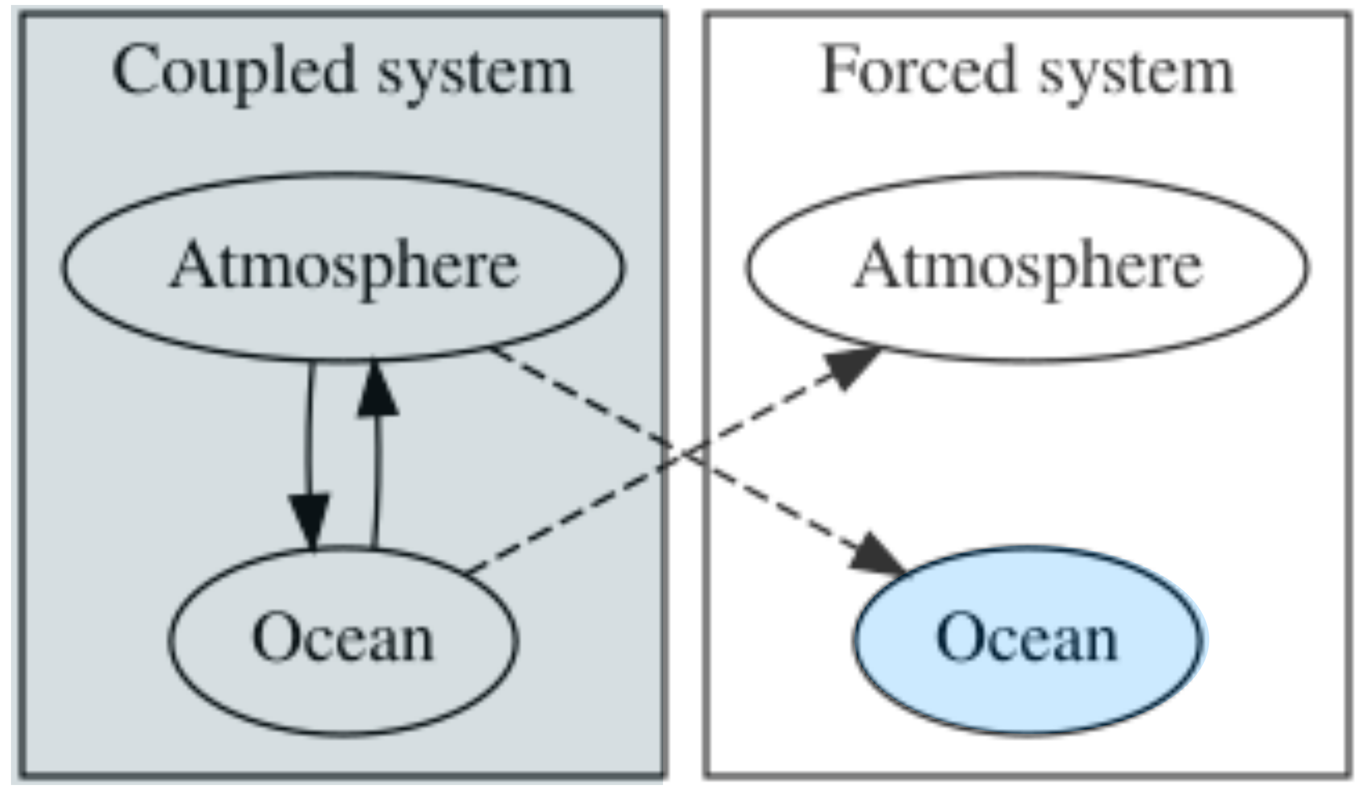
Examining the forced and coupled systems

- We examine:
  - Atmosphere forced by the coupled ocean state
  - Ocean forced by the coupled atmospheric state
  - Fully coupled modeling system



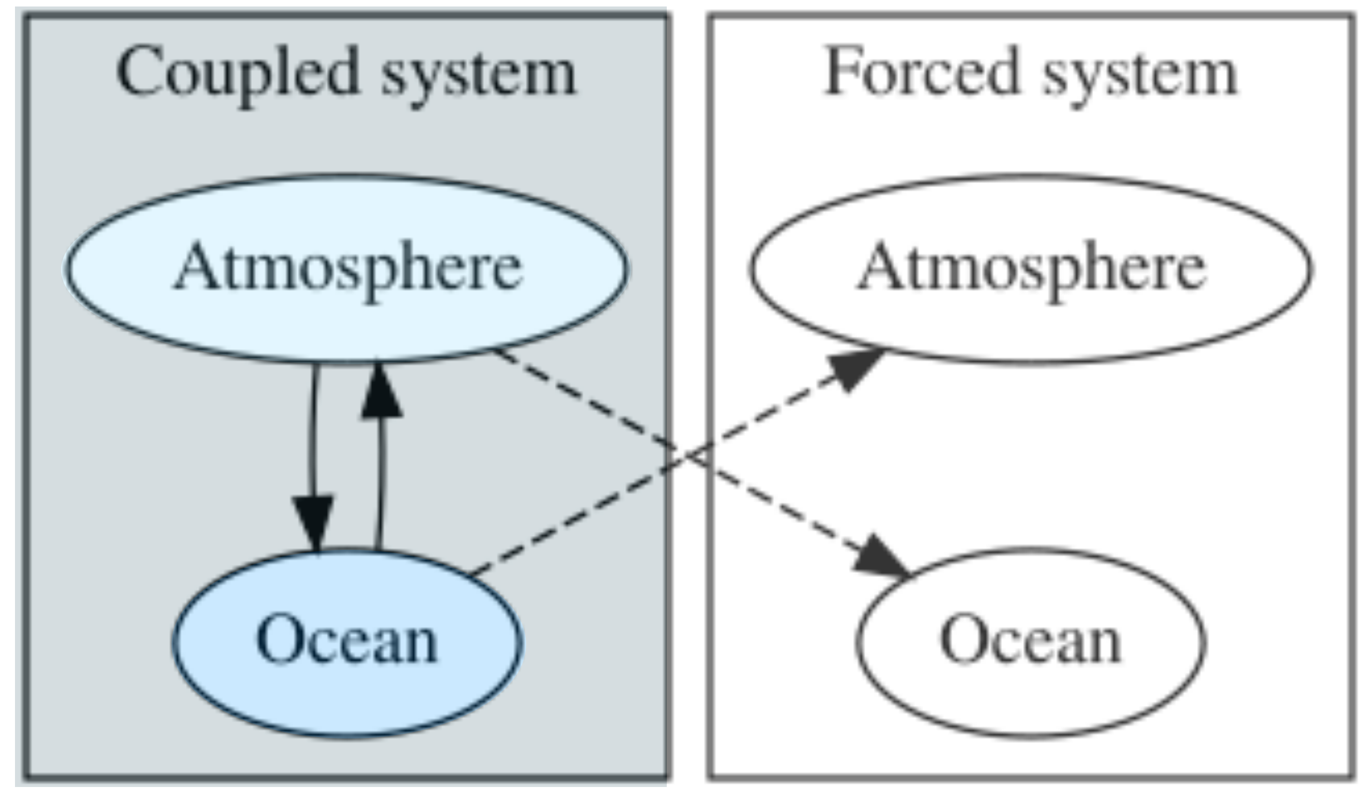
Examining the forced and coupled systems

- We examine:
  - Atmosphere forced by the coupled ocean state
  - Ocean forced by the coupled atmospheric state
  - Fully coupled modeling system



Examining the forced and coupled systems

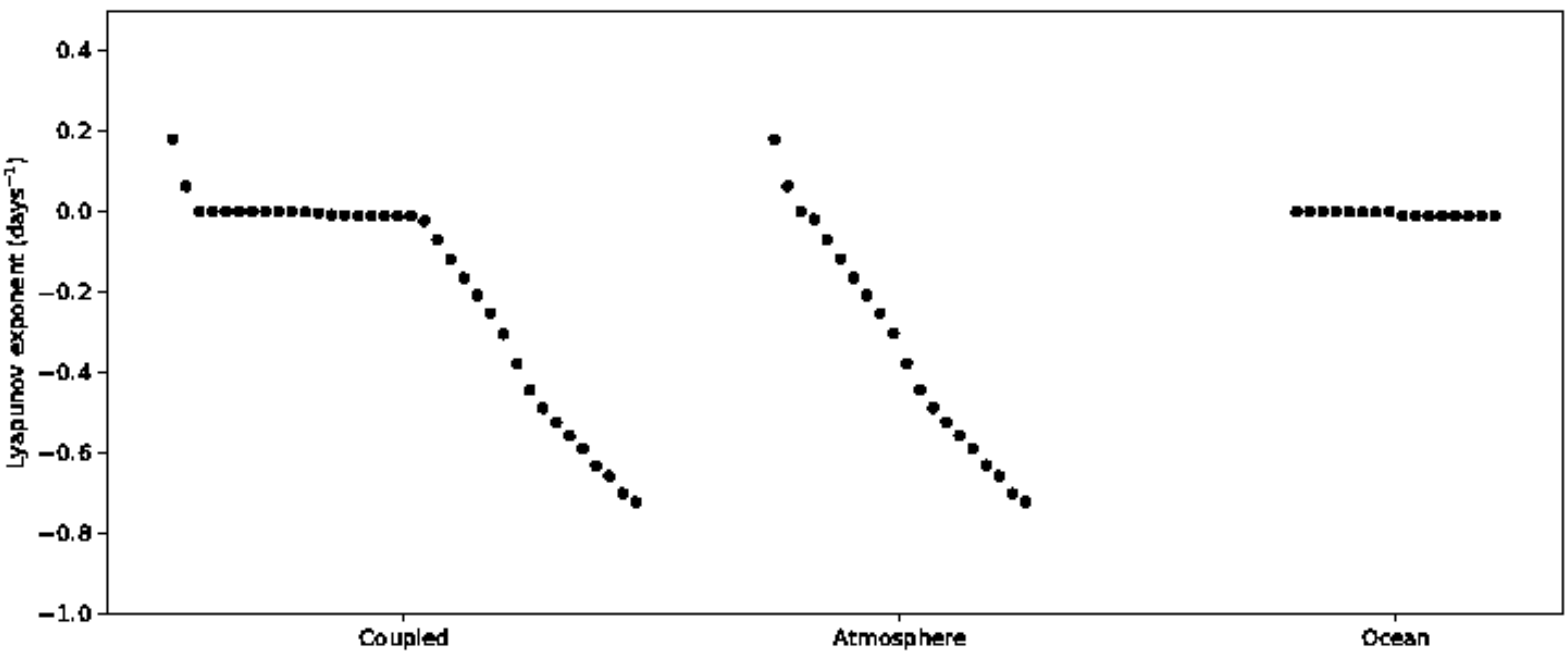
- We examine:
  - Atmosphere forced by the coupled ocean state
  - Ocean forced by the coupled atmospheric state
  - Fully coupled modeling system



*The attempt is to emulate the typical transition process in an operational center like NCEP

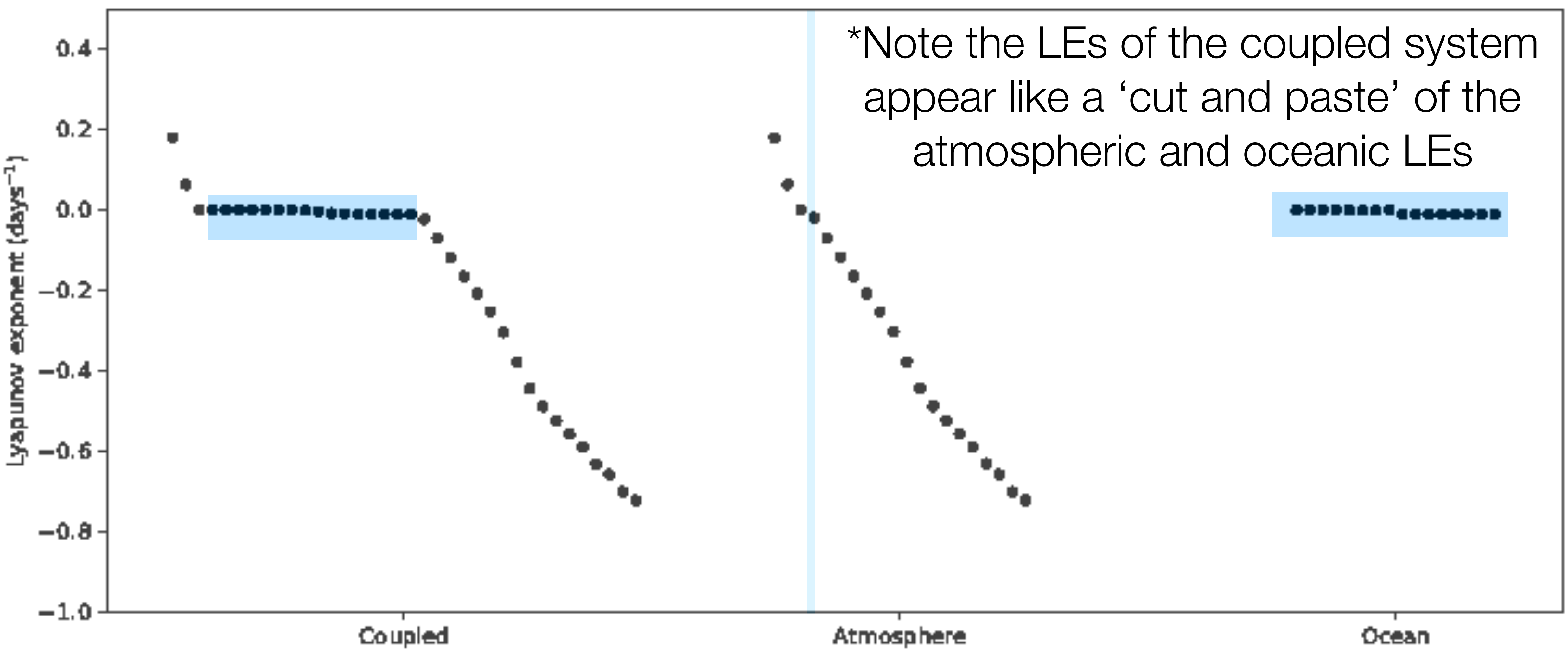
Lyapunov spectrum of coupled system, forced atmosphere, and forced ocean

- The discrepancy in scales can be characterized by the ratio the magnitudes of Lyapunov Exponents (LEs)



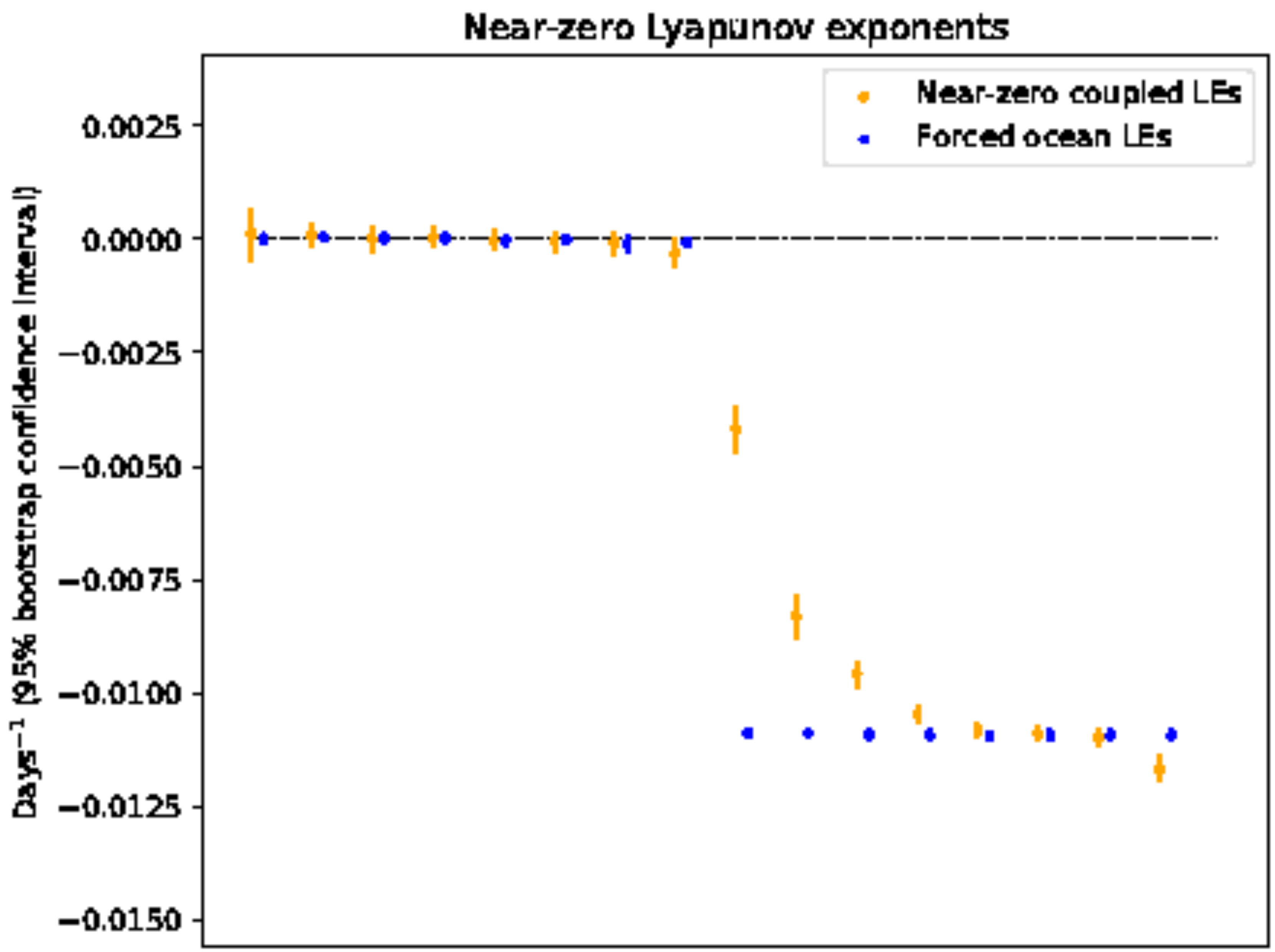
Lyapunov spectrum of coupled system, forced atmosphere, and forced ocean

- The discrepancy in scales can be characterized by the ratio the magnitudes of Lyapunov Exponents (LEs)



Comparing forced ocean LEs with corresponding coupled LEs

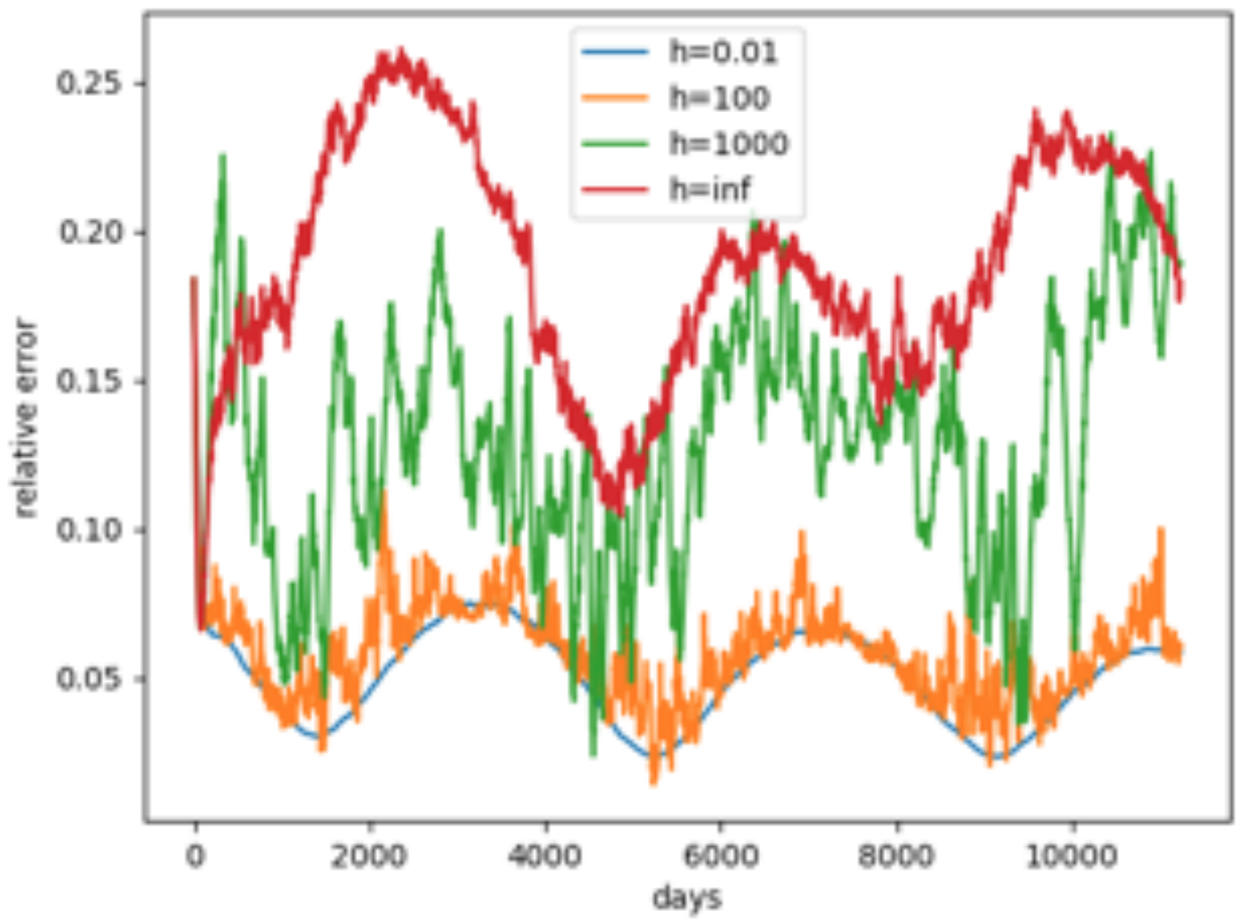
- What appears as a 'jump' in the forced ocean Lyapunov spectrum becomes a smooth transition in the coupled system



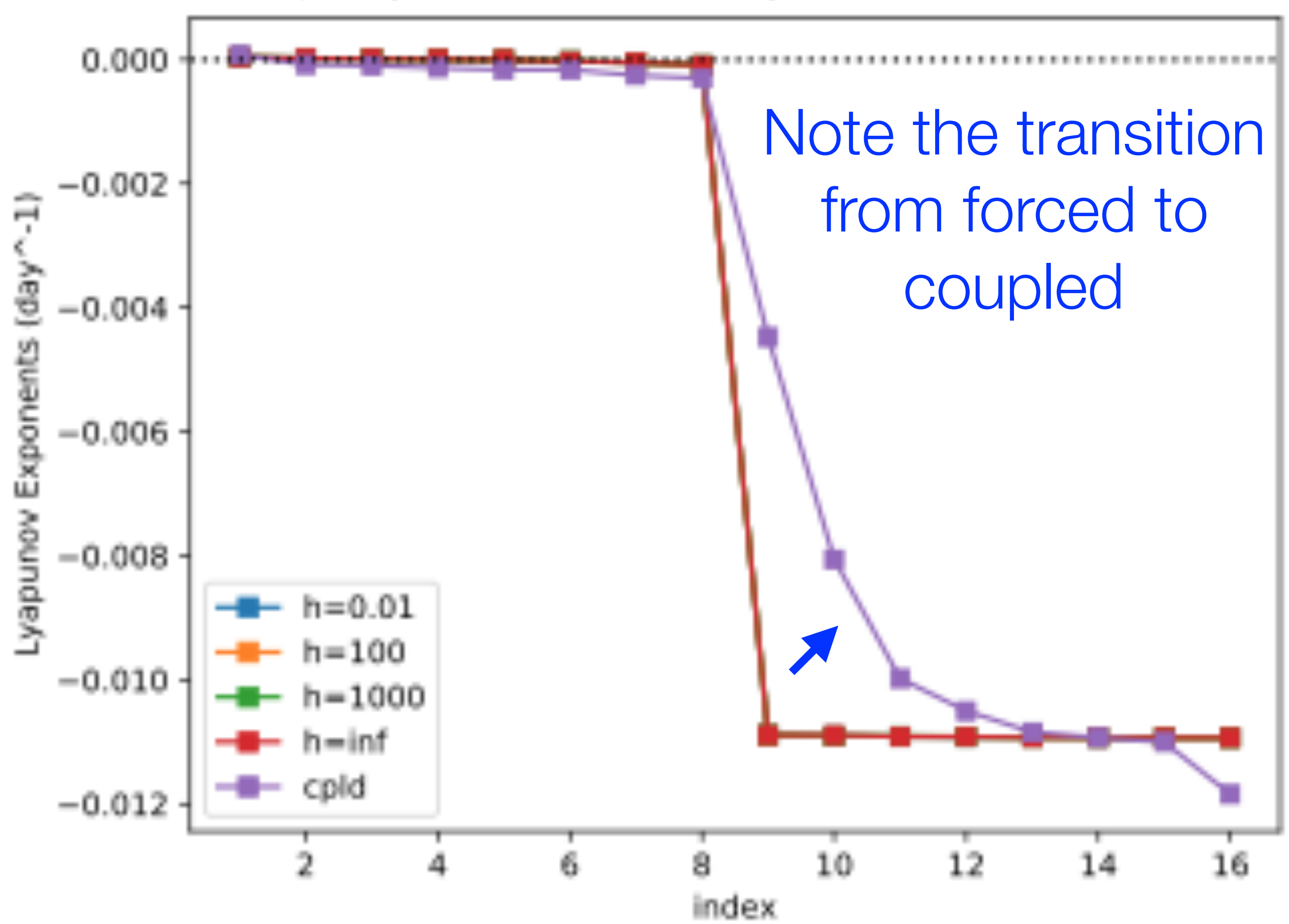
Lyapunov stability of the forced system

- Even the forced atmosphere and forced ocean (shown below) do not synchronize when provided with accurate forcing.

Error over time:



Lyapunov exponents:

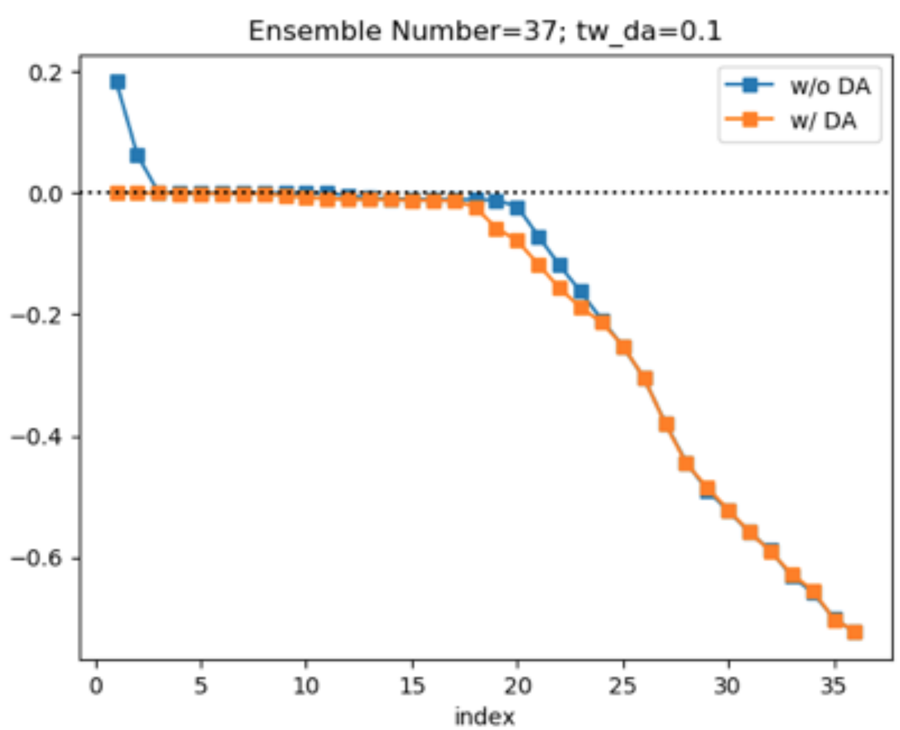


Data assimilation stabilizes growing errors

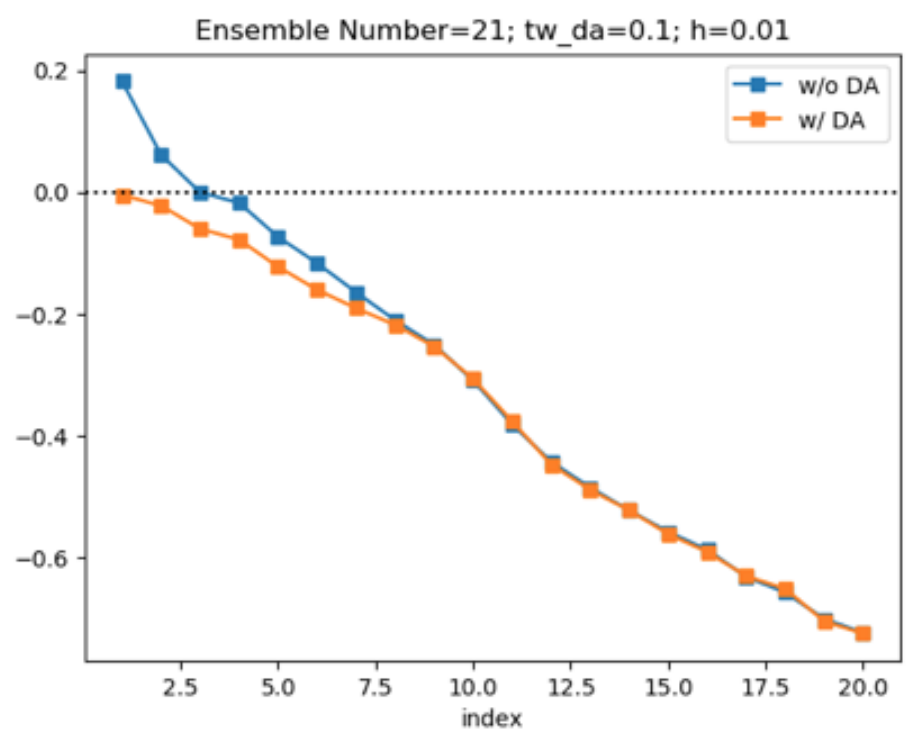
- Data assimilation provides a forcing towards the ‘true’ state that constrains growing errors

- The drives the (conditional) Lyapunov exponents negative, indicating stability

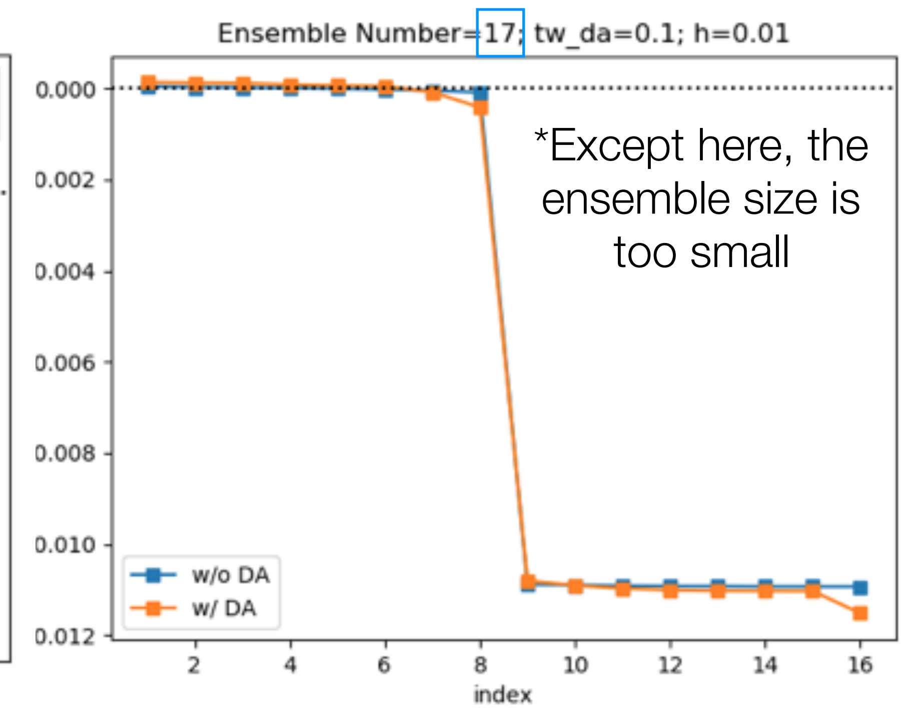
Coupled System



Forced Atmosphere

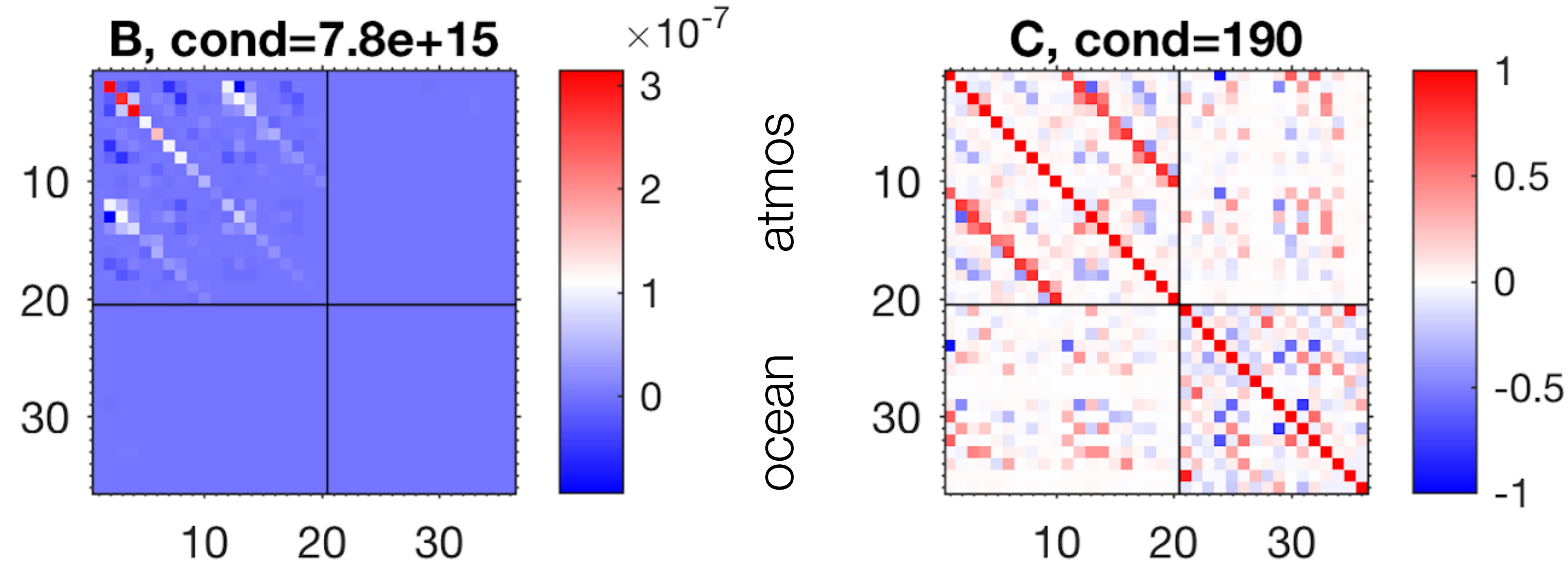


Forced Ocean



Variational CDA

- Building the climatological error covariance matrix **B**

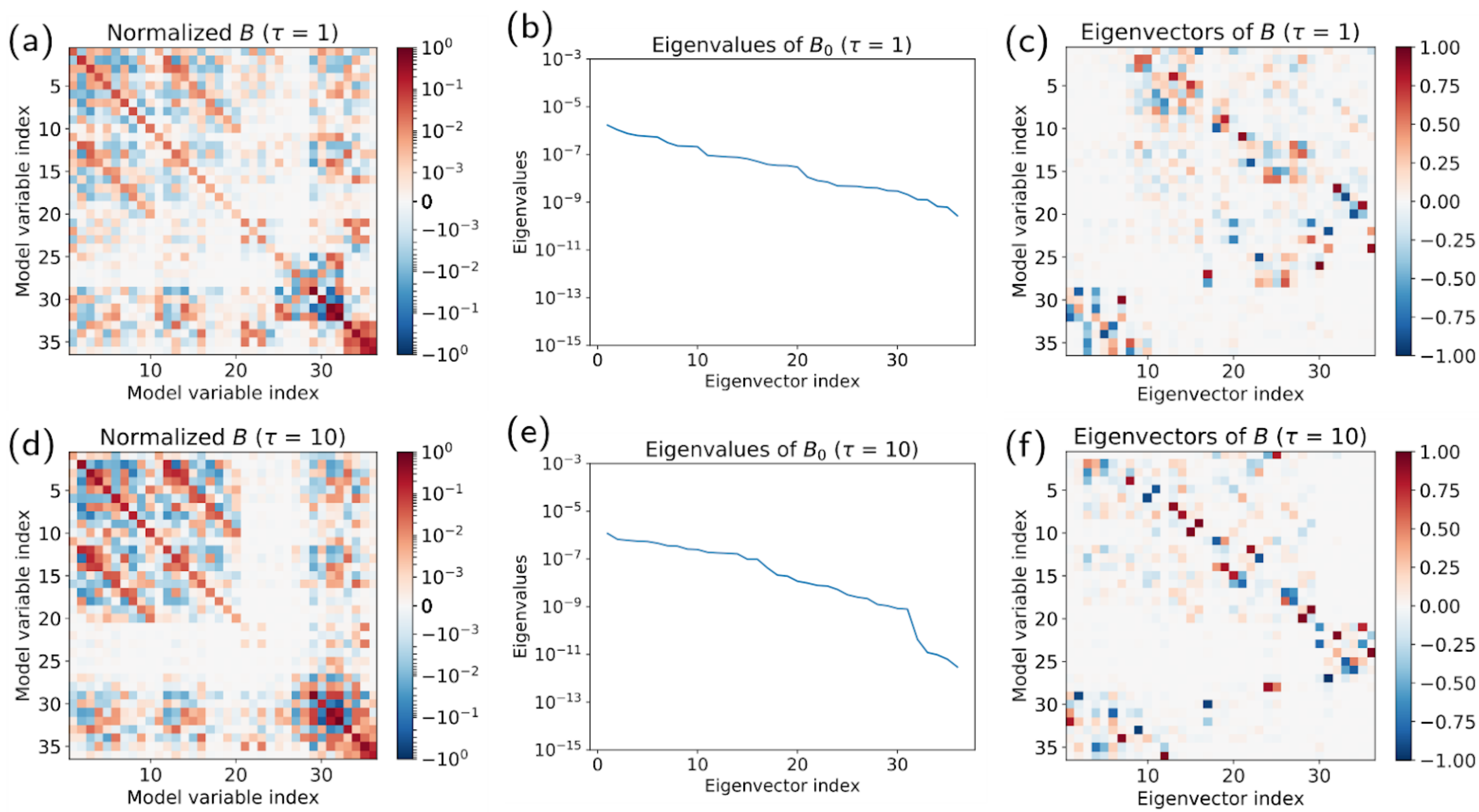


- Due to the highly disparate scales, the **B** matrix is ill-conditioned (i.e. ratio of largest to smallest eigenvalue >>1)

- Either transforming to the correlation matrix (e.g. Smith et al. 2018) or using the control variable transform can mitigate this issue

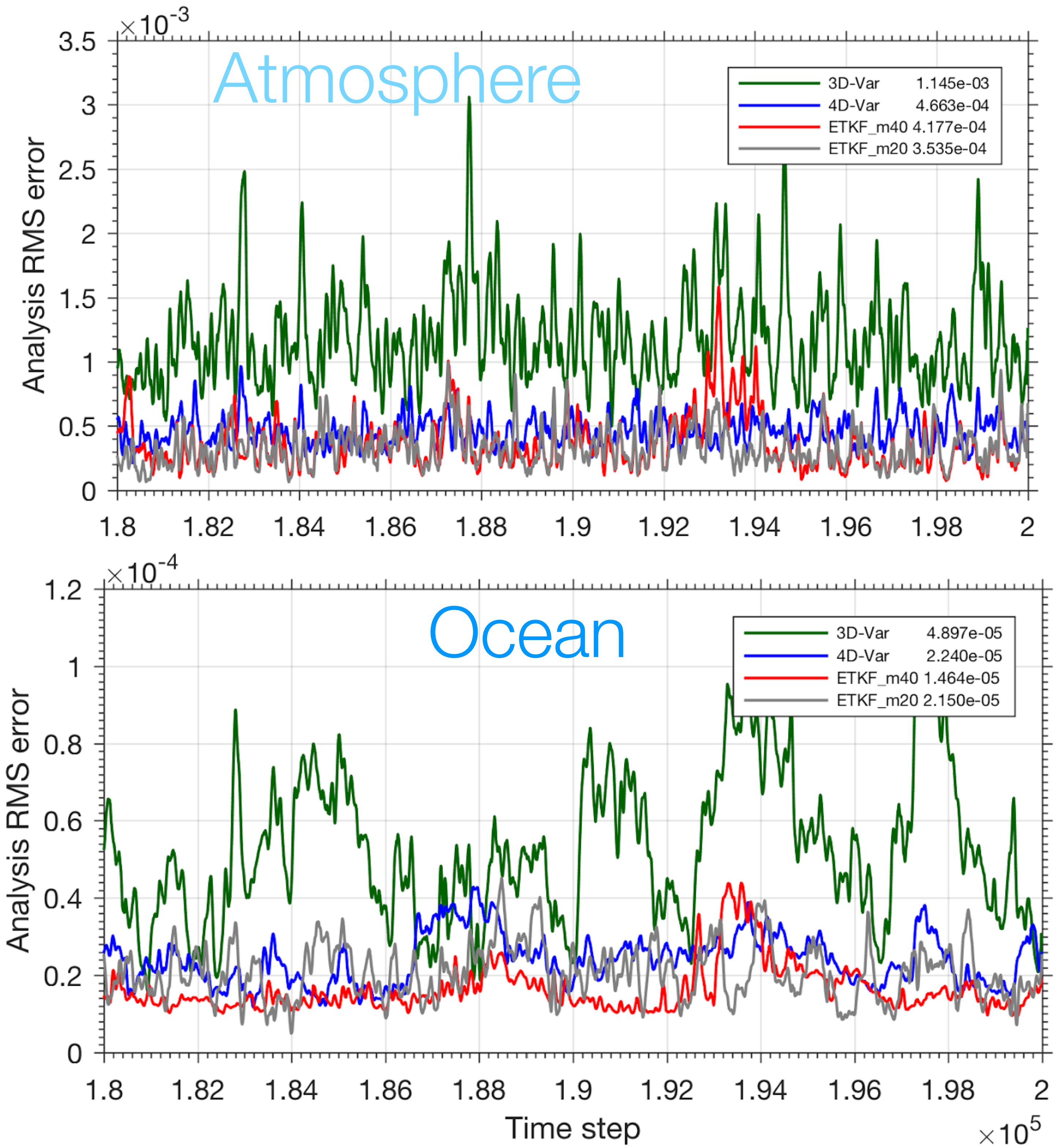
Climatological forecast error covariance **B** at various lead times

- The structure of **B** changes depending on the lead time of the forecast
- This may indicate that building **B** matrices for different timescales may be beneficial



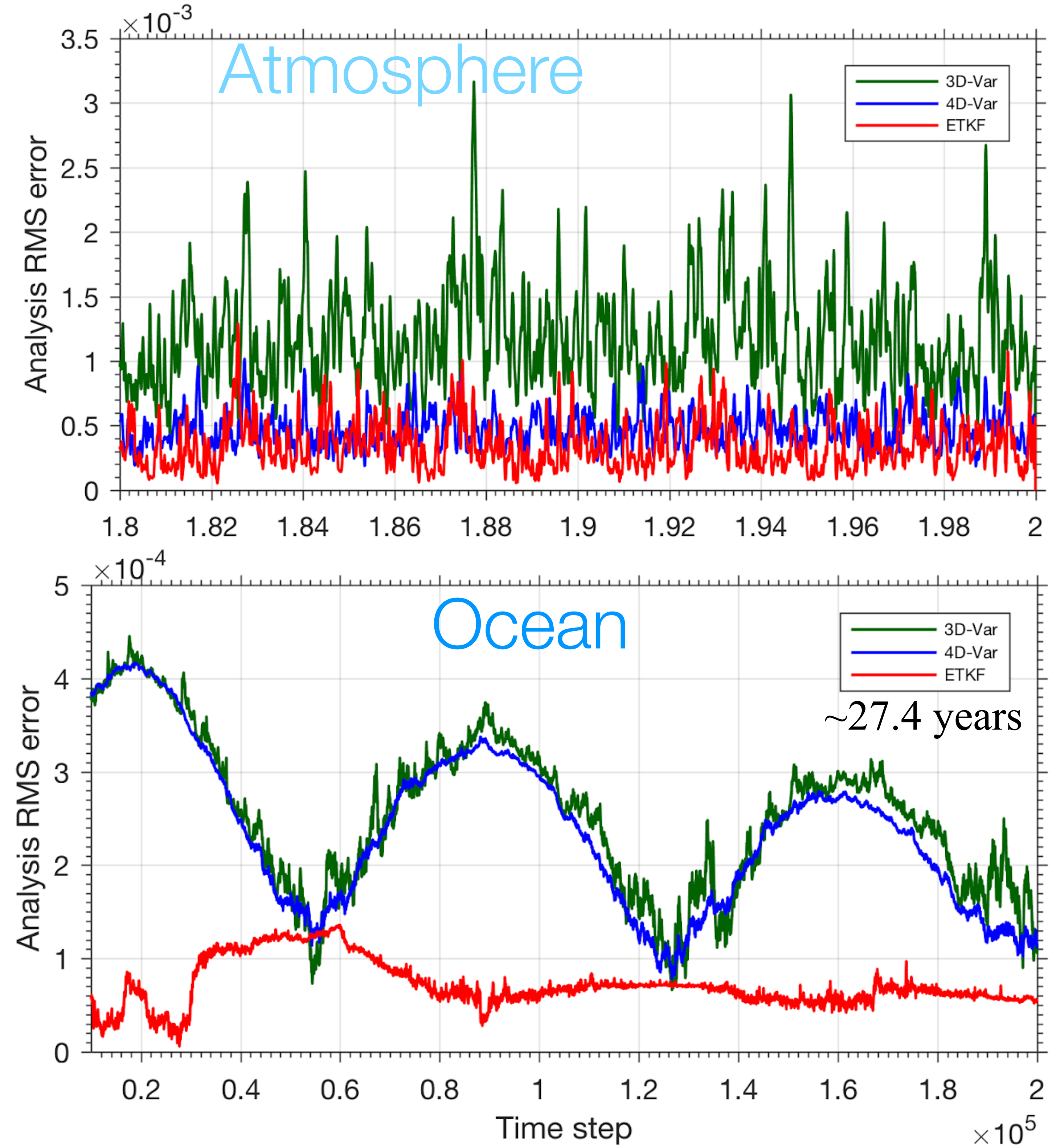
Assimilating observations in the entire coupled domain using 3D-Var, 4D-Var, and the ETKF

- The 3D-Var generally produces the lowest accuracy analysis.
- The accuracy of 4D-Var and the ETKF (*k*=40 or *k*=20) are comparable.



Assimilating **only atmospheric observations** using 3D-Var, 4D-Var, and the ETKF (*k*=40)

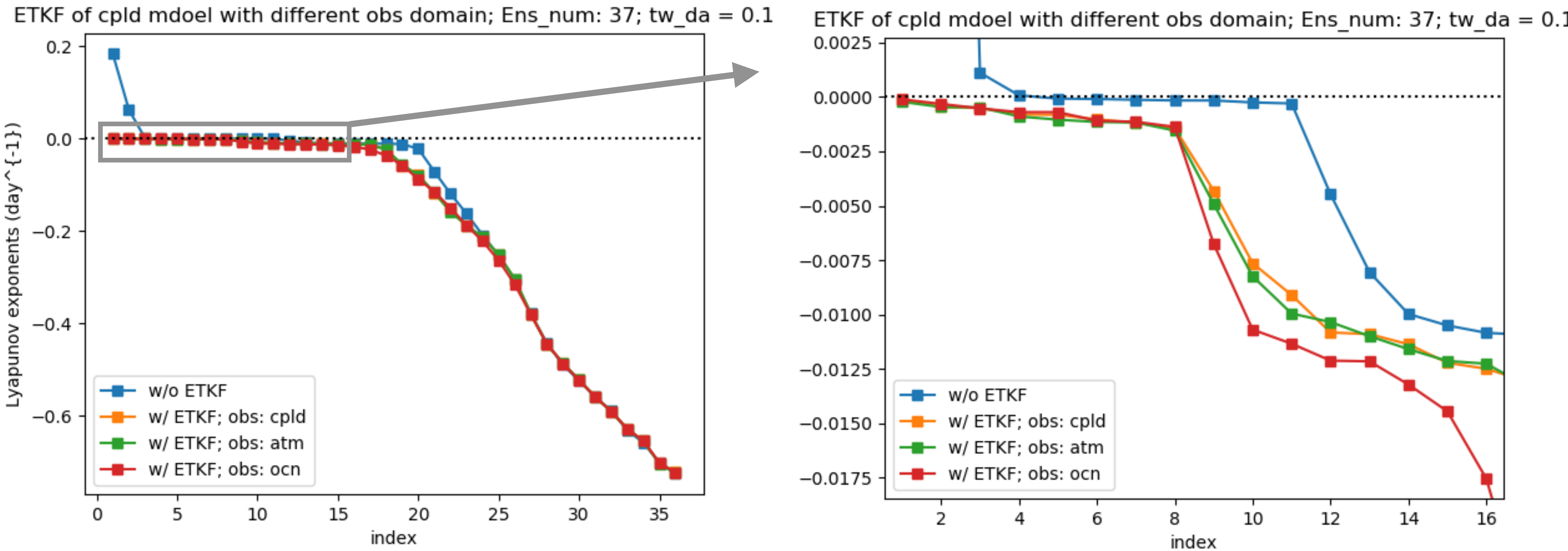
- The accuracy of the 4D-Var and ETKF are compatible in the atmosphere
- The accuracy is degraded in the ocean compared to observing the full domain
- There is a decadal oscillation in the error of the variational solution in the ocean, likely due to the static climatological error covariance matrix



More explorations using the ETKF

- We examine a number of questions using the ETKF as our exploratory DA tool
- For example we compare:
  - Observing coupled state versus only atmosphere or ocean
  - Observing the model native spectral space or transformed physical grid space
  - Using fixed or mobile observing network
  - Varying ensemble sizes and analysis cycle windows
  - Examining various forecast lead times

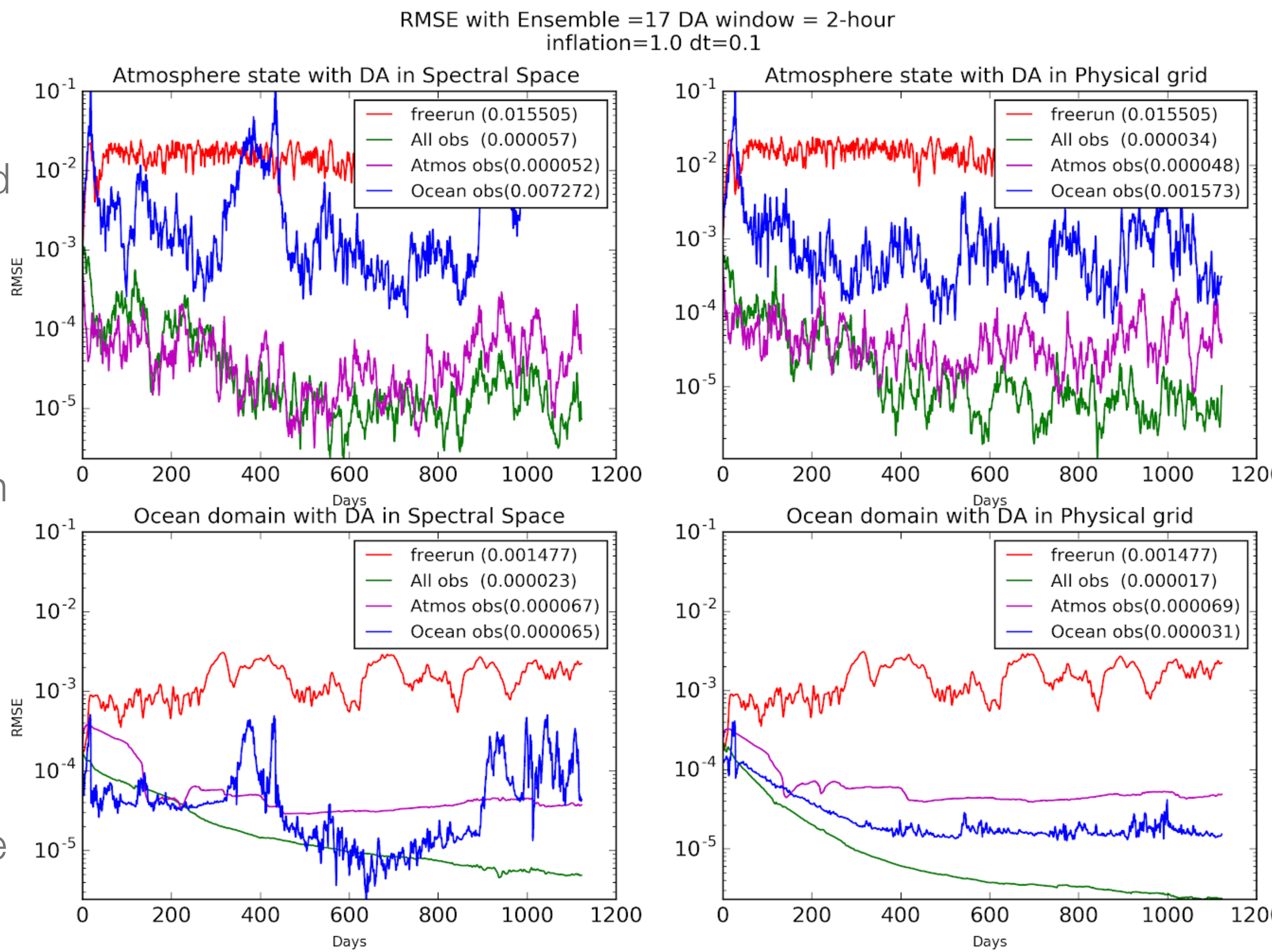
Stability of ETKF when **observing** atmos / ocean / coupled systems



Here, there is a large ensemble size (k=37) and a short assimilation window (tau=0.1)

Comparing ETKF with observations in atmos/ocean and in *model spectral* or *transformed physical grid*

- Best accuracy achieved when observing the entire coupled system and applying CDA
- Applying CDA with only atmospheric observations is still relatively accurate in both domains.
- Assimilating only ocean observations degrades atmospheric state estimate (as may be expected)

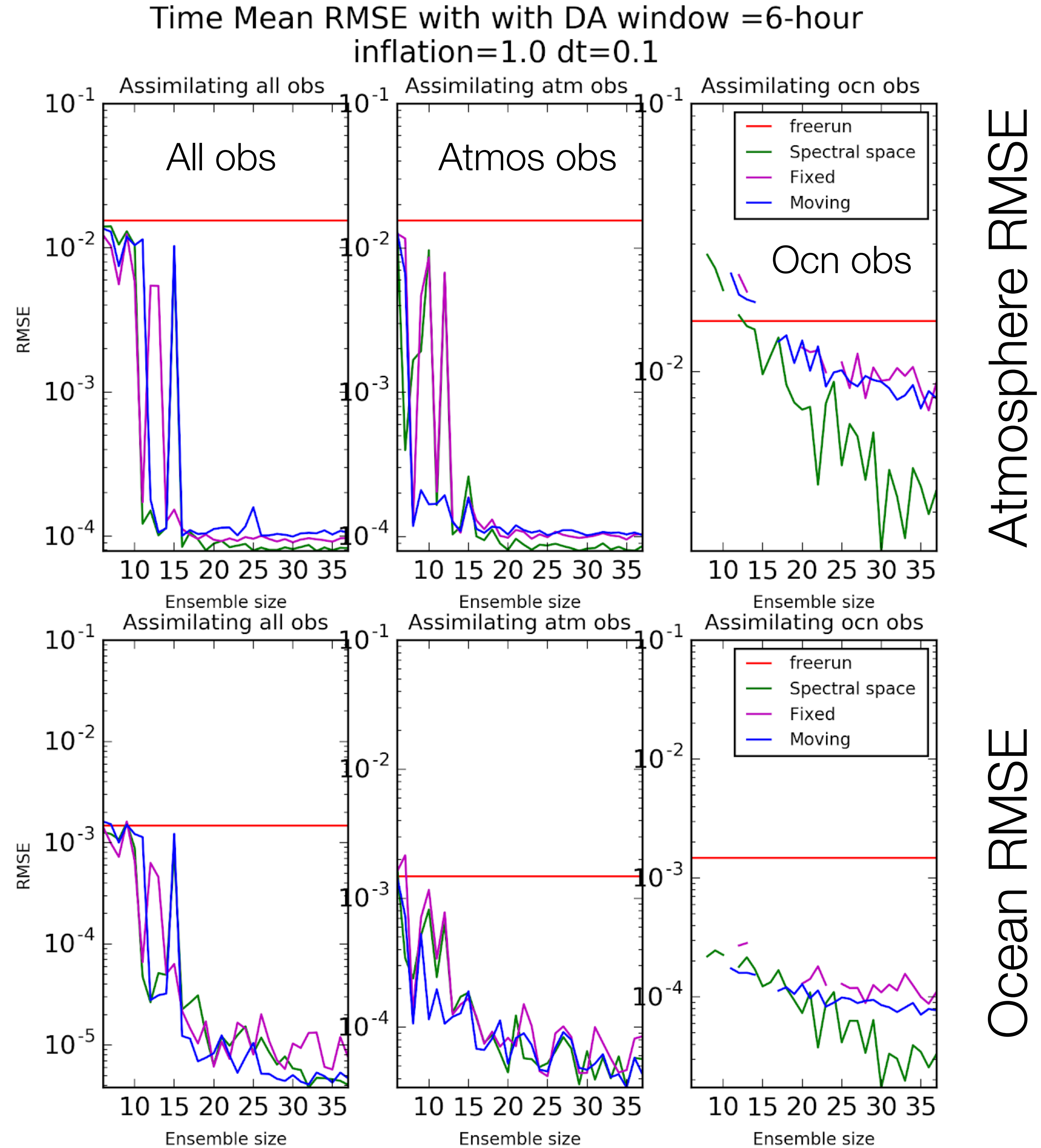


Examining stability while **varying ensemble size and observing networks**

- There is a more gradual transition to stability as ensemble size is increased (versus uncoupled system)

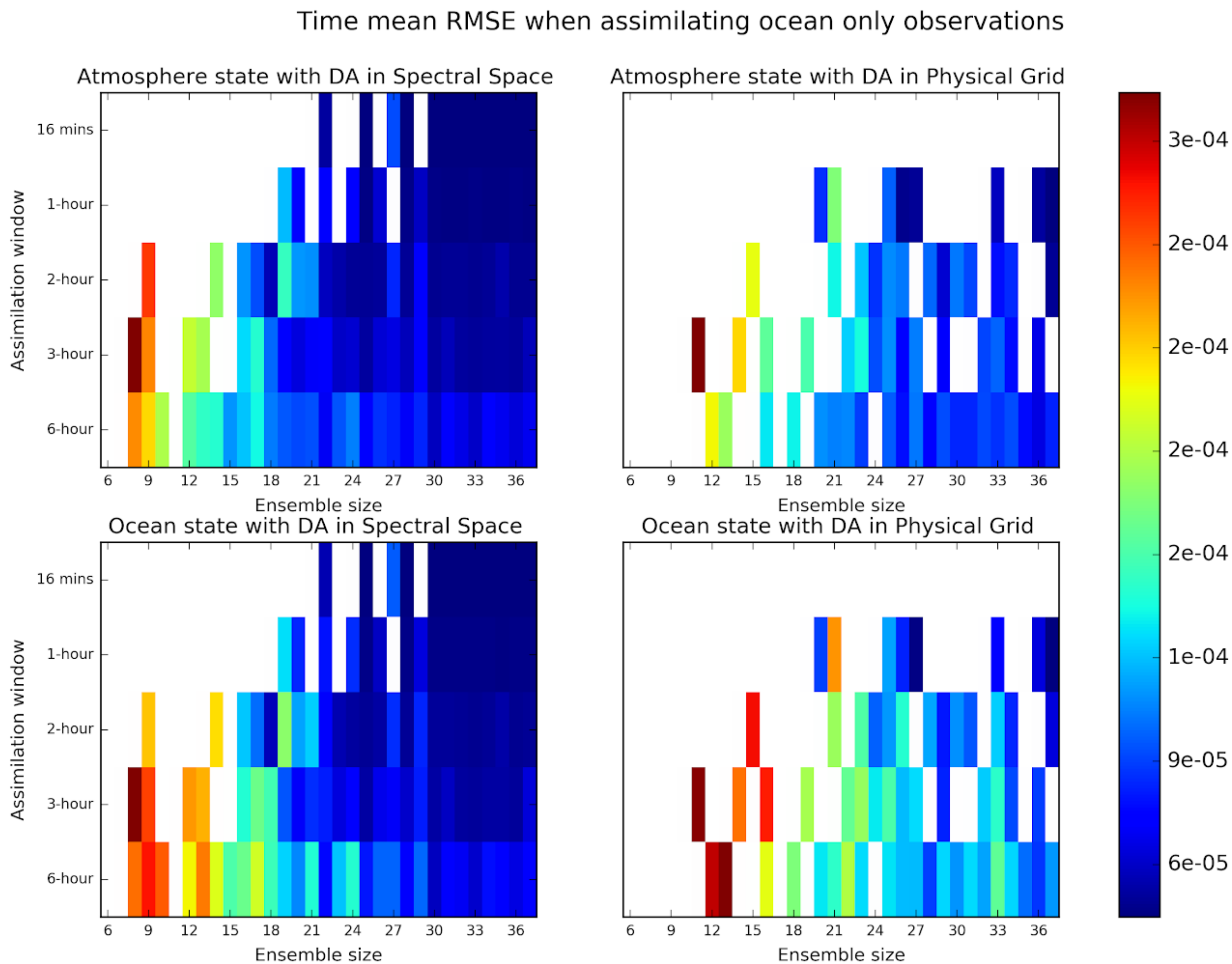
- Best accuracy occurs when assimilating all observations (atmos/ocean)

- With sufficient ensemble size, ocean observations alone can constrain the coupled system, at reduced accuracy.



CDA with **only ocean observations**

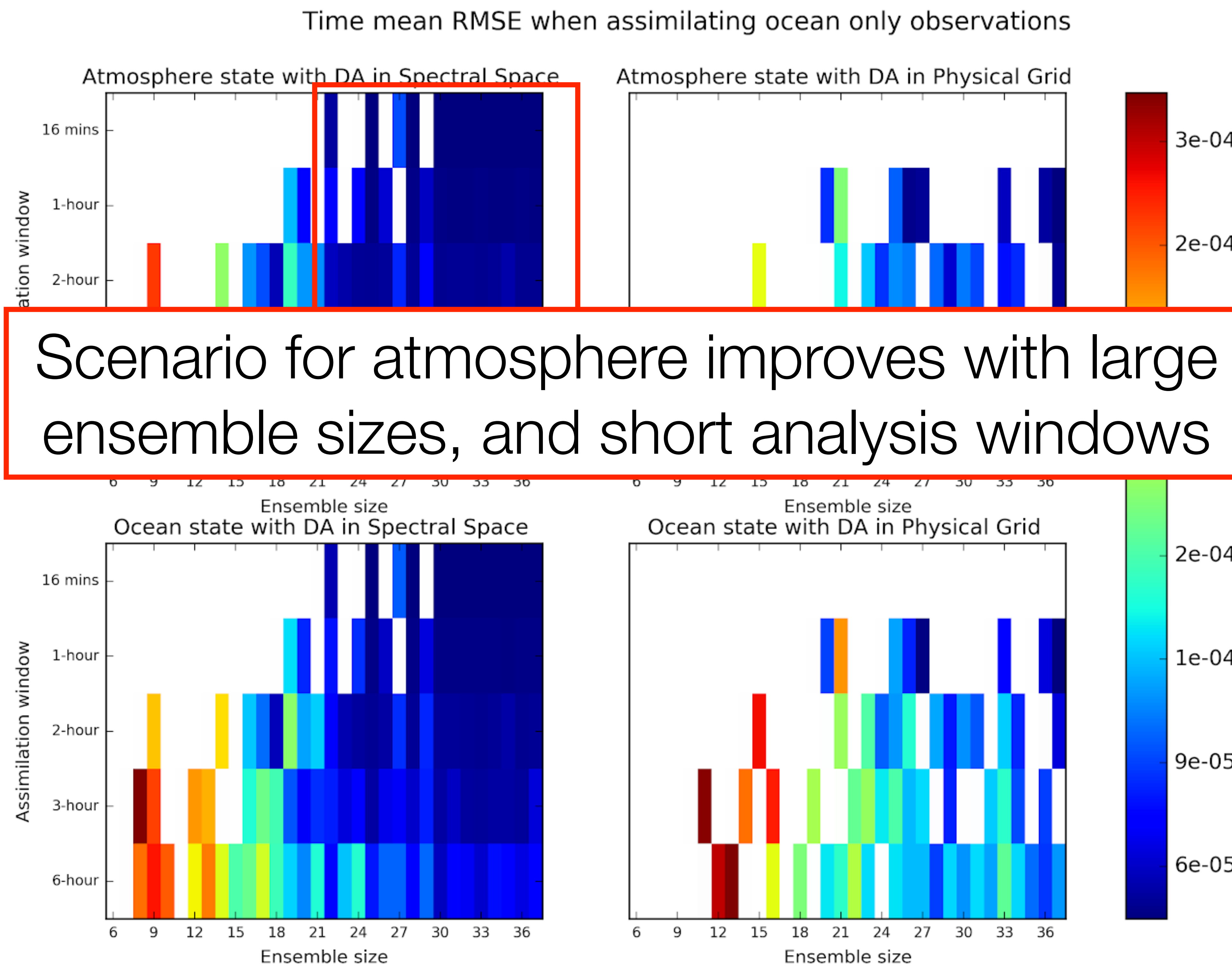
- Assimilation errors are smallest when using large ensembles and small analysis cycle windows
- Observing the native model spectral space is more stable. Observing the transformed physical grid space leads to model instabilities that may indicate many more observations are needed



CDA with **only ocean observations**

- Assimilation errors are smallest when using large ensembles and small analysis cycle windows

- Observing the native model spectral space is more stable. Observing the transformed physical grid space leads to model instabilities that may indicate many more observations are needed

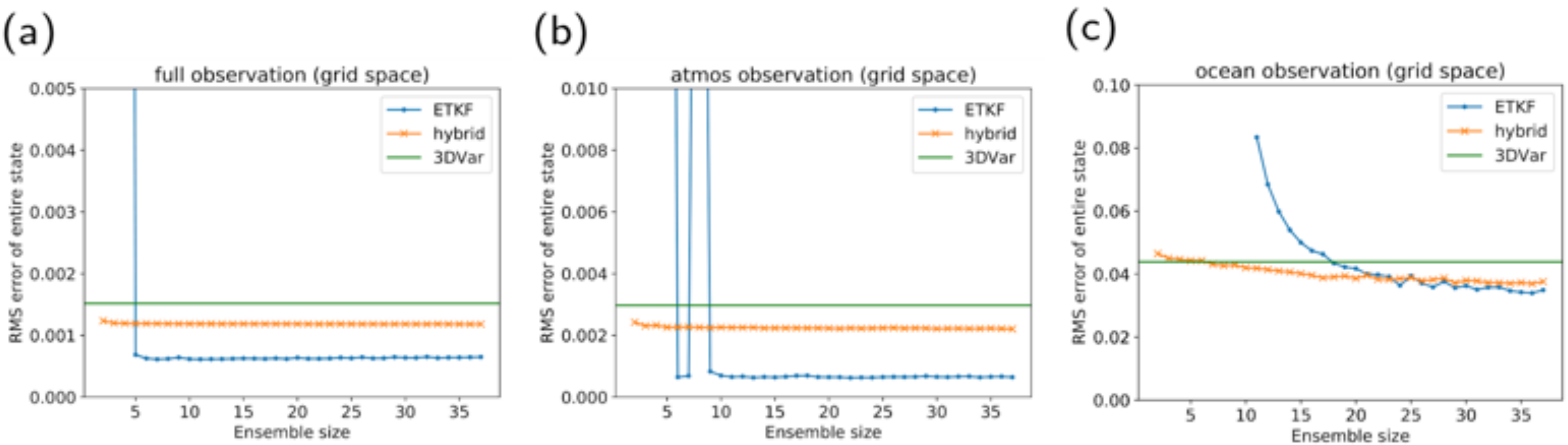


Scenario for atmosphere improves with large ensemble sizes, and short analysis windows

Hybrid-Gain CDA

- Similarly to the forced systems, the Hybrid-Gain CDA is effective when observing only atmospheric observations at stabilizing the filter at small ensemble sizes, when the ETKF otherwise diverges

- Unlike the forced system, the gaining of stability when observing only the ocean is very gradual with increasing ensemble size. The Hybrid-Gain CDA provides stability at low ensemble sizes and comparable results with large ensemble sizes.



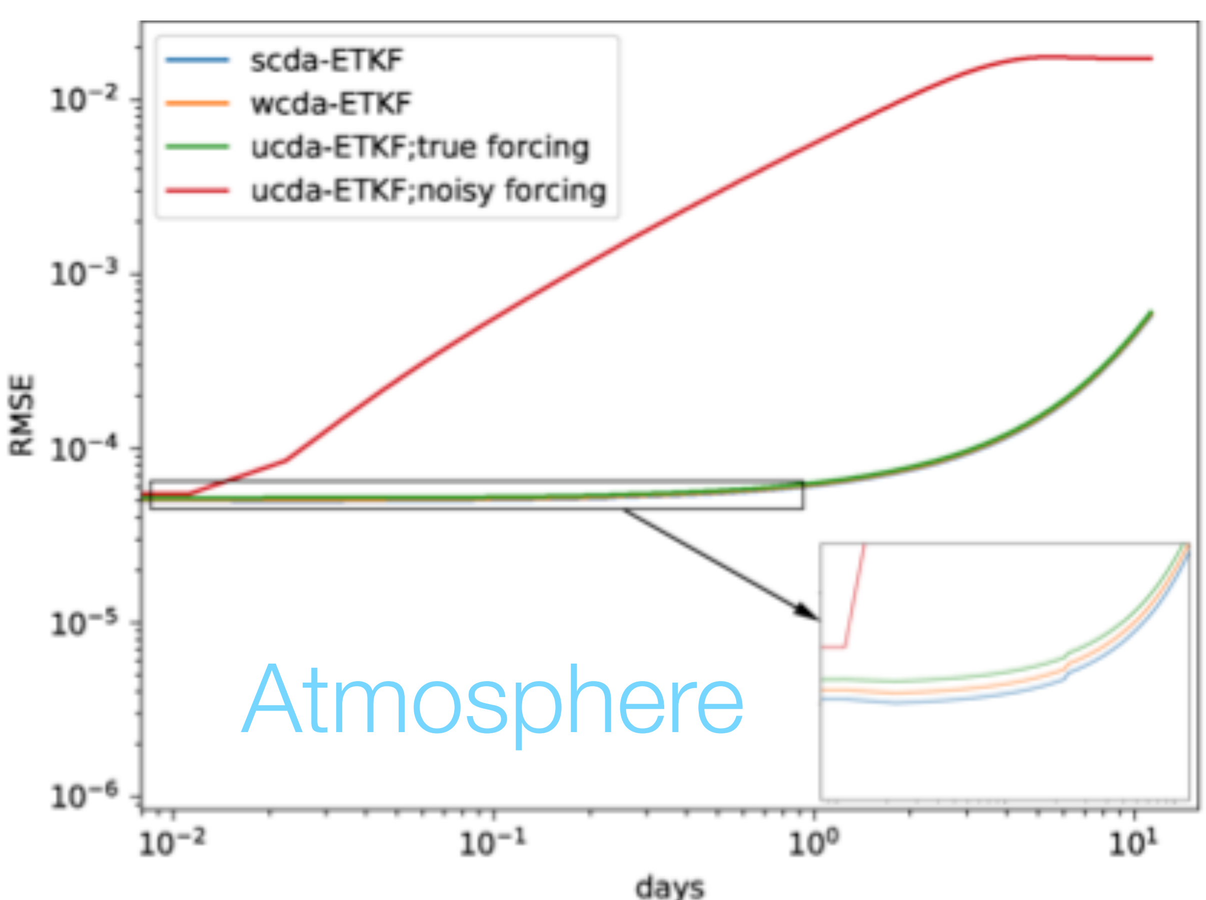
(here, ETKF uses relaxation to prior)

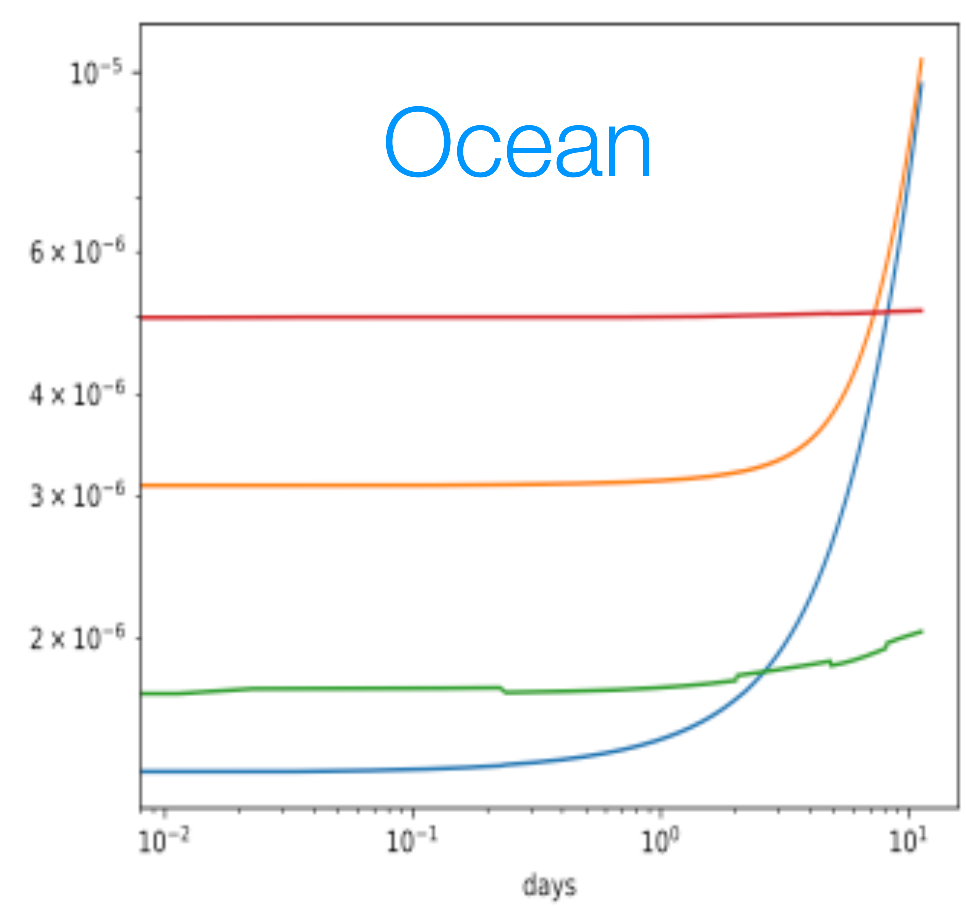
Forecast accuracy at various lead times

- Forecast accuracy in MAOOAM initialized with ETKF is similar in atmosphere for SCDA, WCDA, and uncoupled perfect forcing case. Diverges for noisy forcing case.
- Forecast accuracy in the ocean is most accurate with SCDA for the first 48 hours versus the perfect forcing case, and out to about 1 week versus the WCDA cases.

*RMSE of forecasts with lead times ranging from 0 to 10 days initialized from the analyses produced from 36,000 DA cycles

**noisy forcing uses white noise with magnitude 10% of climatological variability



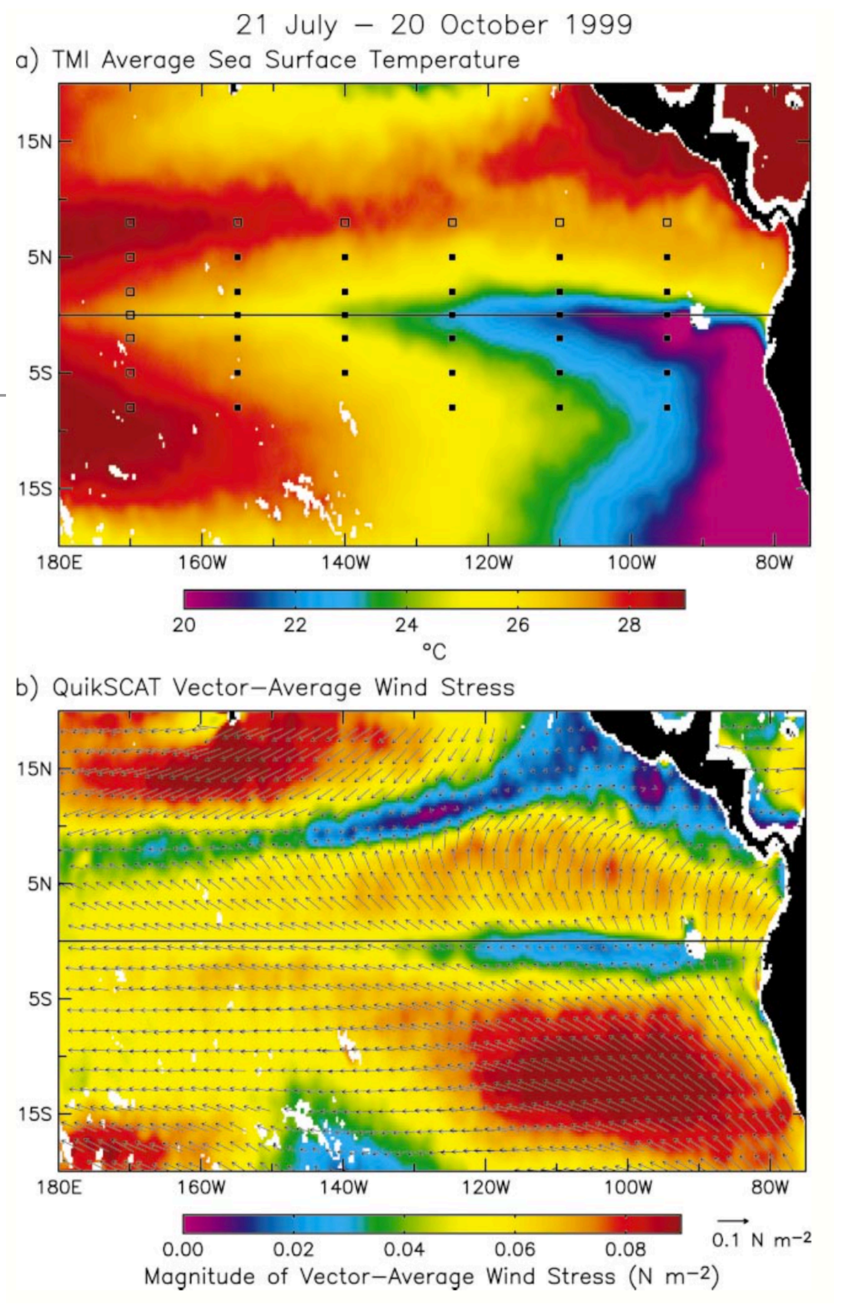


Overview

- Brief Bio/background
- Motivation for Coupled Data Assimilation (CDA)
- Prior results using Strongly Coupled Data Assimilation (SCDA)
- Our results using SCDA with a simple coupled QG model
- Extending to more realistic systems

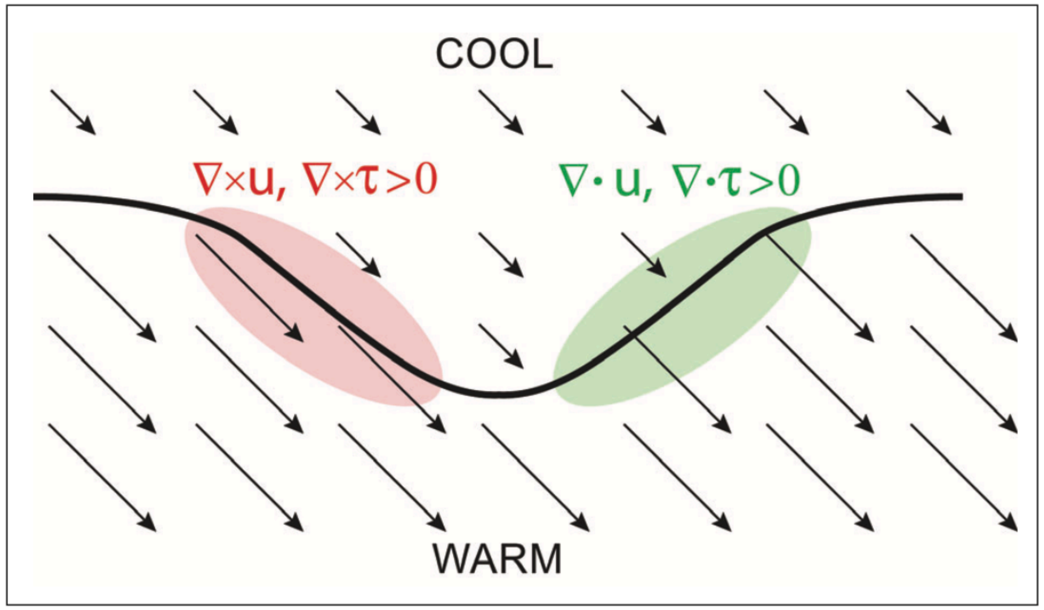
SST and Surface Wind Interaction

- Stability of atmospheric boundary layer is affected by SST
- Wind stress divergence correlates with cold to warm SST, and wind stress convergence with warm to cold SST, strongest with winds aligning with SST gradient



Chelton et al. (2001)

- Due to sensitivity in lateral variations, the wind stress curl is strongest where winds align with isotherms.



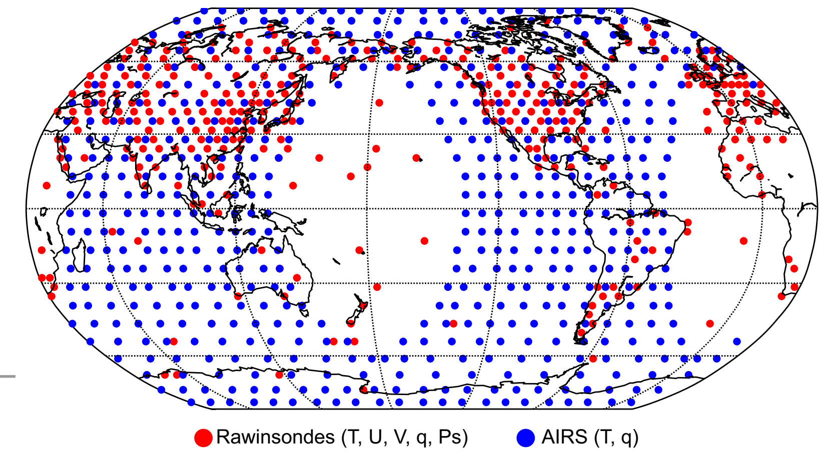
Chelton and Xie (2010)

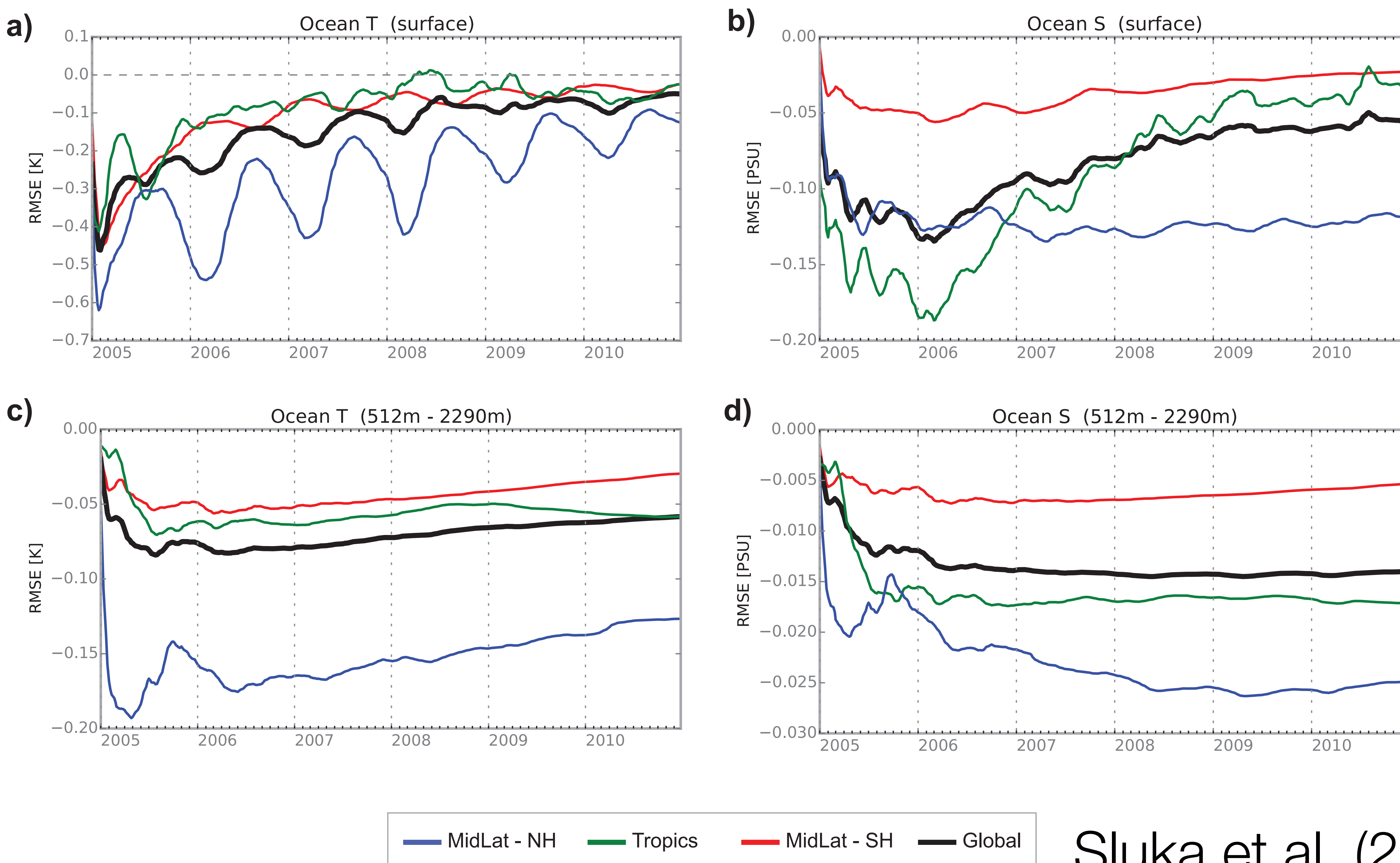
Applying SCDA to an Intermediate Complexity model

- Sluka at al. (2016): *Coupled SPEEDY/NEMO model*
    - Assimilate atmospheric observations to update the ocean directly via SCDA and compared to WCDA
    - T30 atmosphere with 2º ocean telescoping to 0.25º in tropics, using LETKF with an ensemble size of 40 members updated at a 6-hour analysis cycle
    - Shows large reduction in errors using SCDA vs WCDA

Sluka, Penny, Kalnay, Miyoshi (2016)

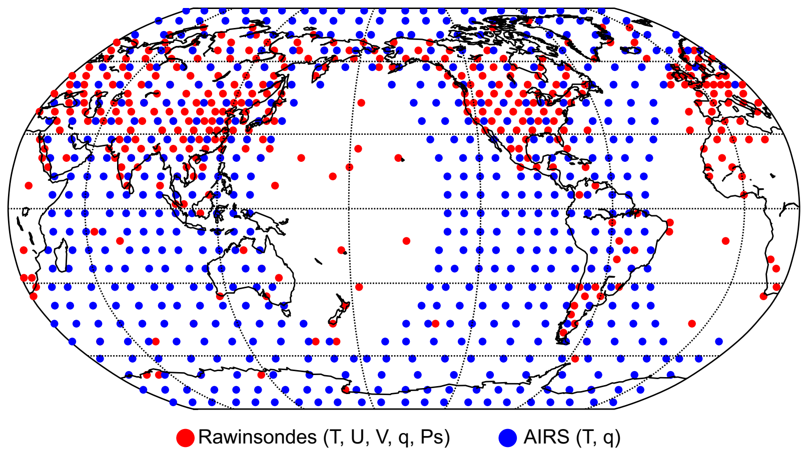
Reduction in analysis error using SCDA versus WCDA baseline

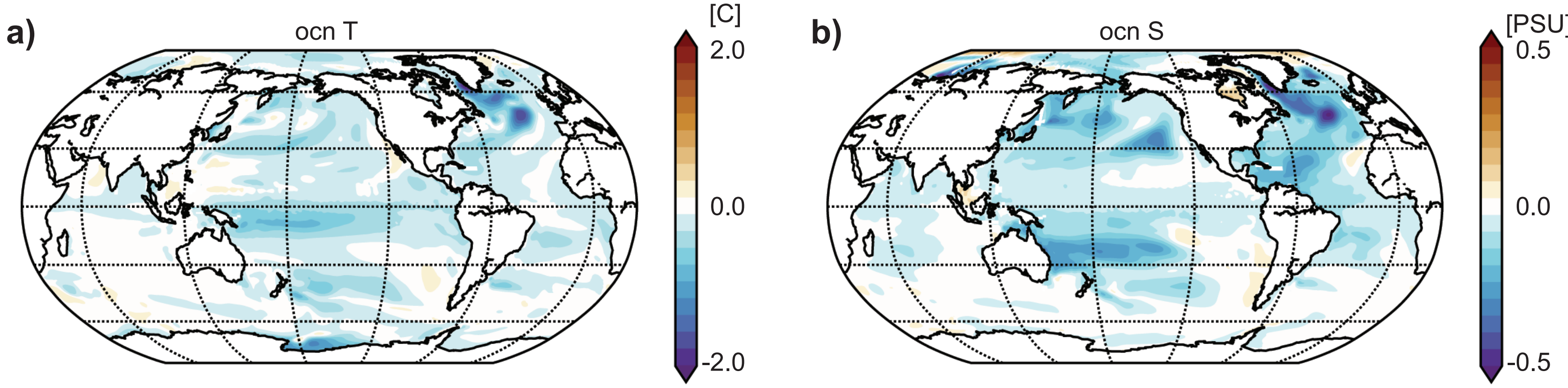




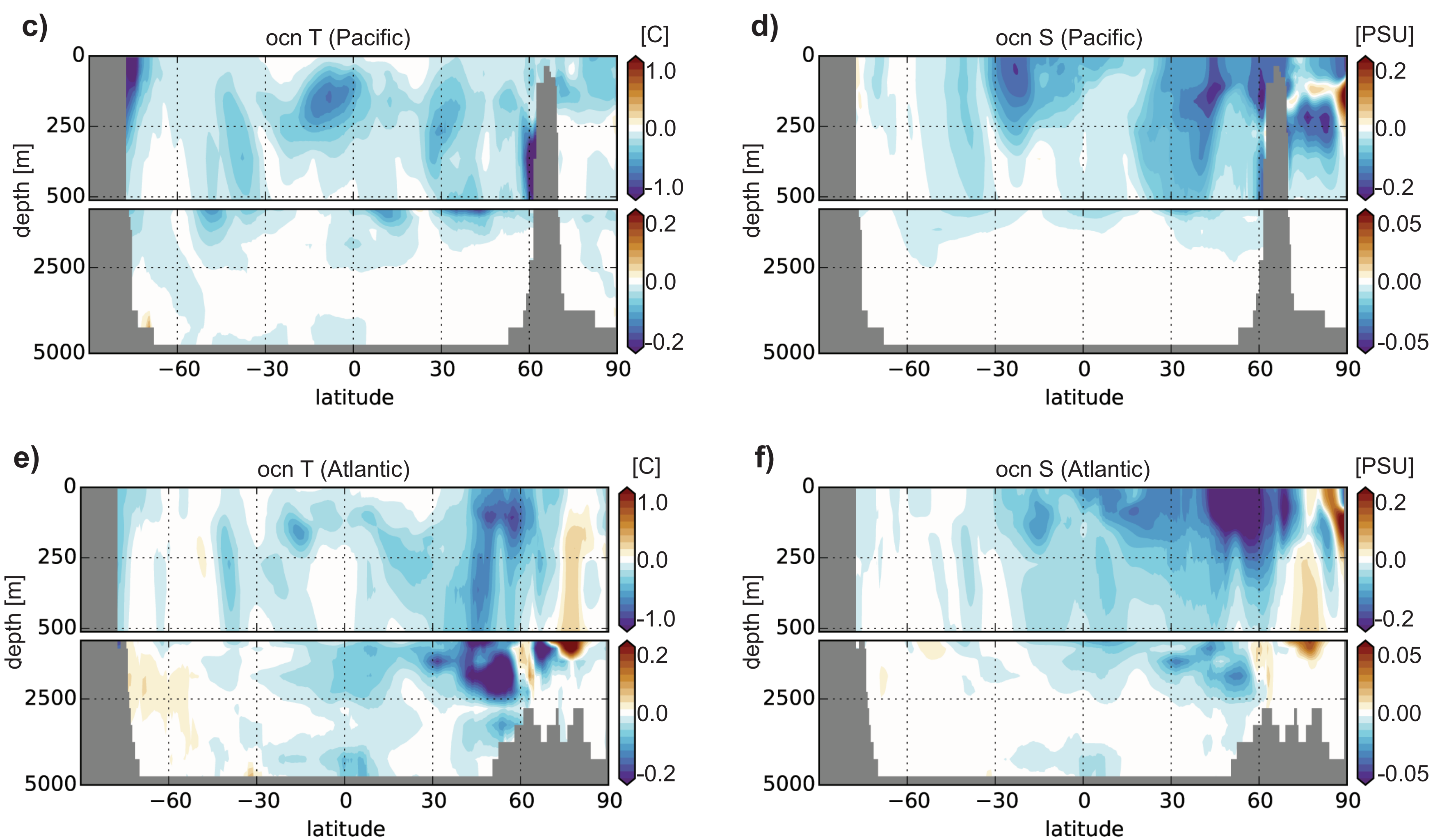
Sluka et al. (2016)

Reduction in analysis error using SCDA versus WCDA baseline





Surface T and S RMSE reduction

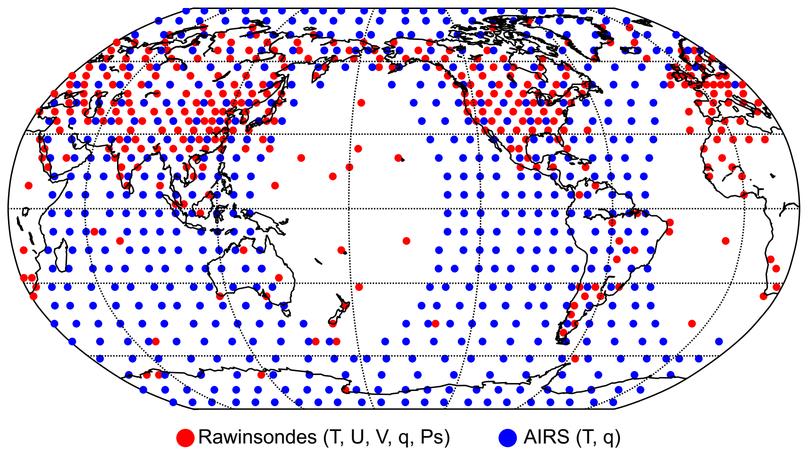


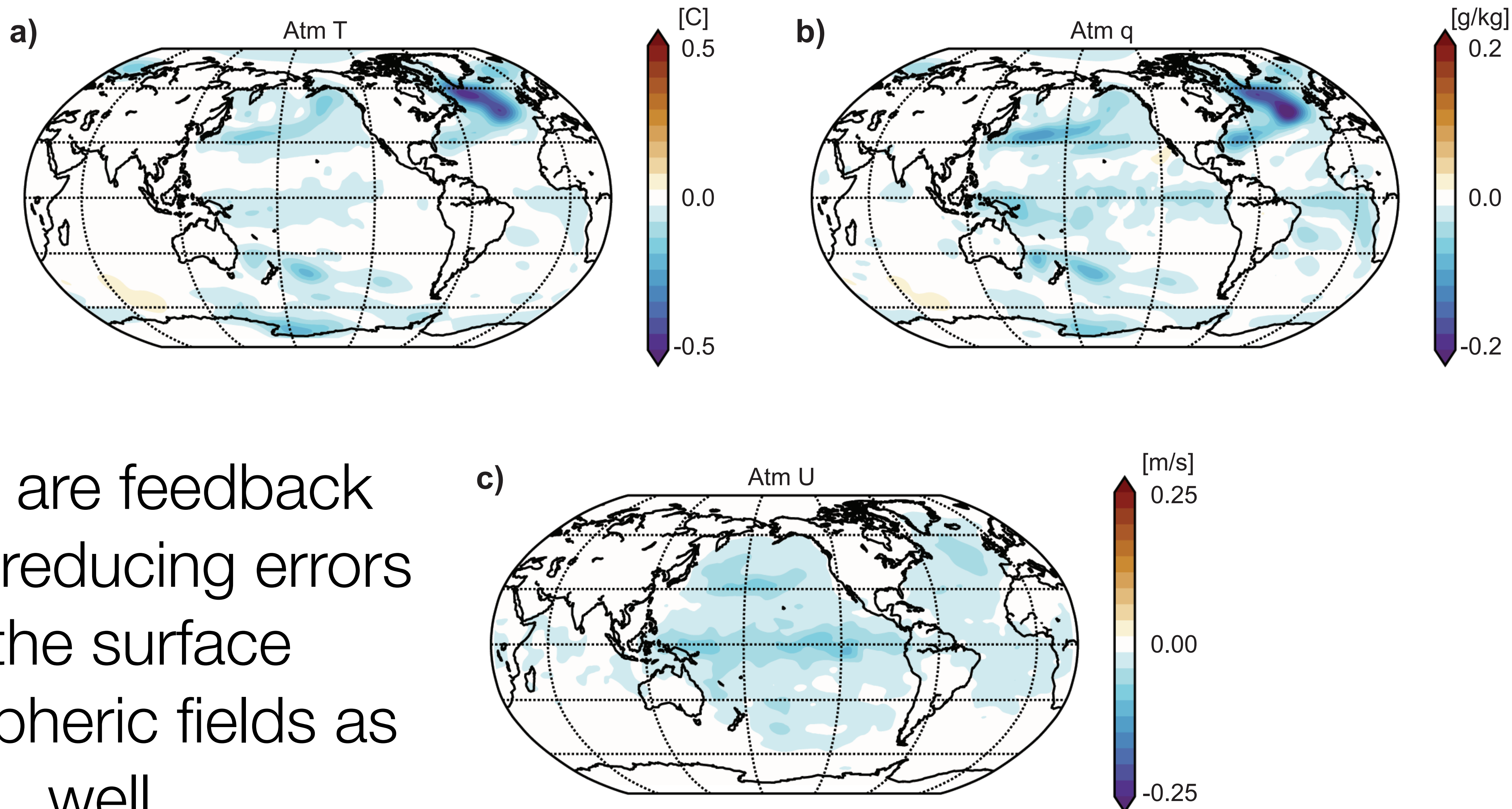
Zonal average RMSE reduction in the Pacific

Zonal average RMSE reduction in the Atlantic

Sluka et al. (2016)

Reduction in analysis error using SCDA versus WCDA baseline



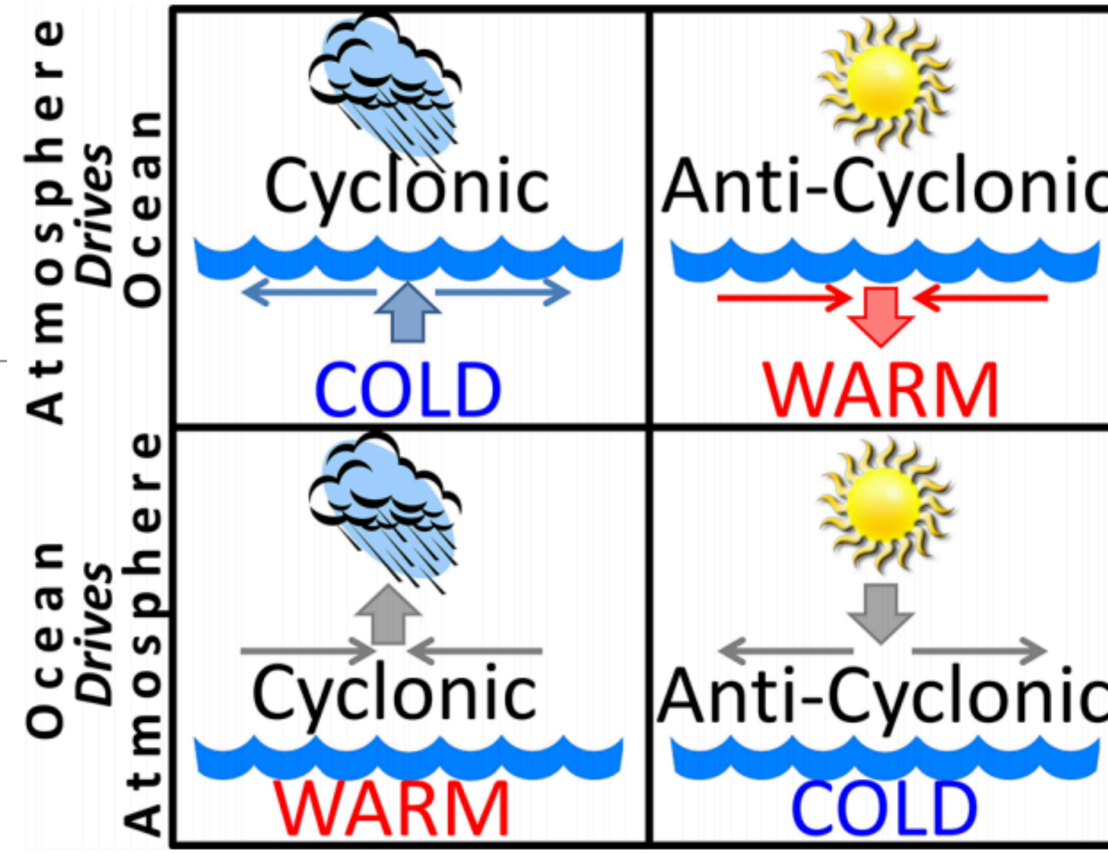


There are feedback effects reducing errors in the surface atmospheric fields as well.

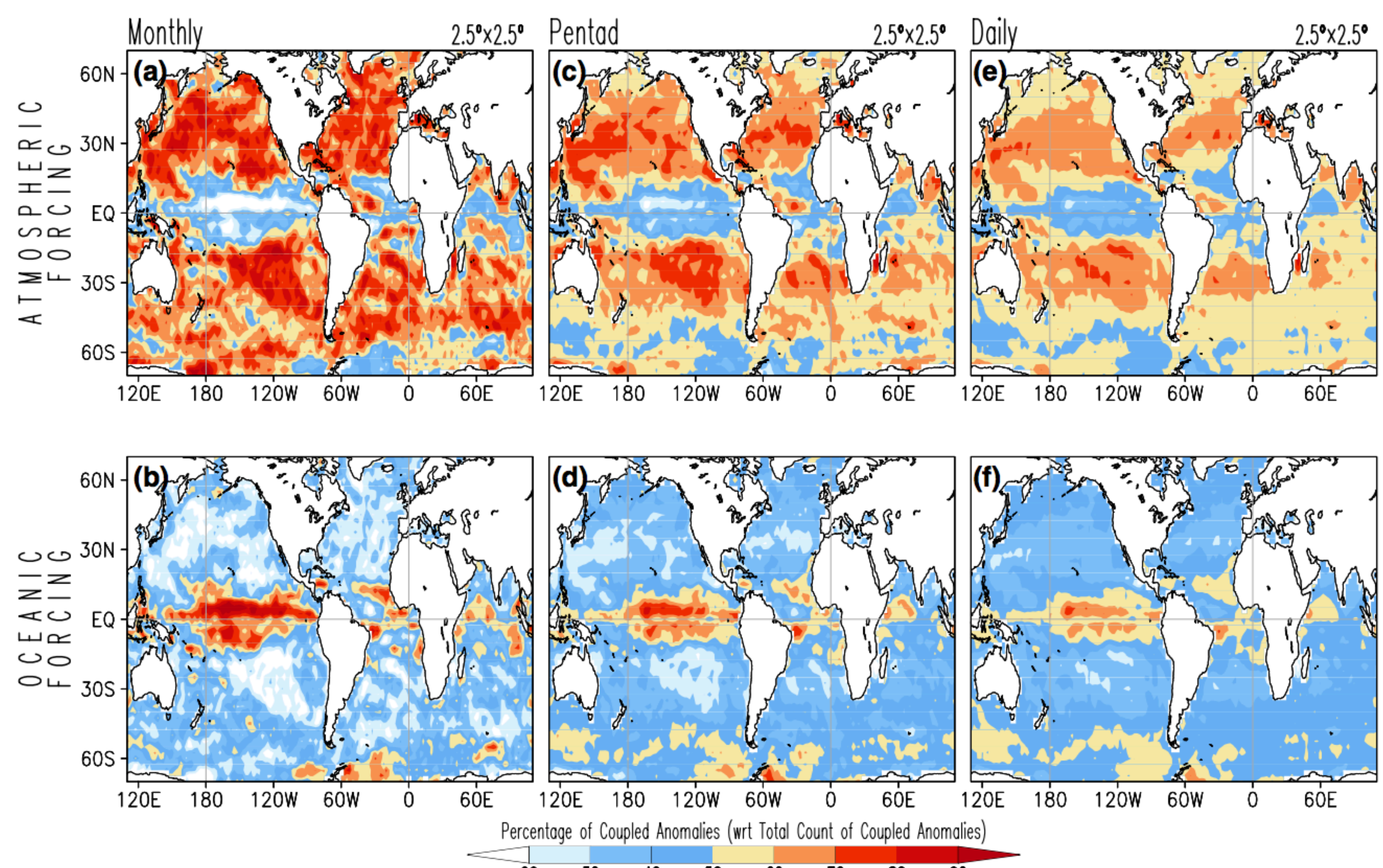
Coupled Anomalies

- Relationship between slowly varying SST anomalies and low-level (850 mb) atmospheric vorticity anomalies.
- Examination of CMIP5 model output and NOAA reanalysis products show coupled anomalies driven by atmos in the midlatitudes and by the ocean in the tropics.
- Coupled anomalies exist in Atmospheric reanalyses due to assimilation of observations

Premise of attribution:

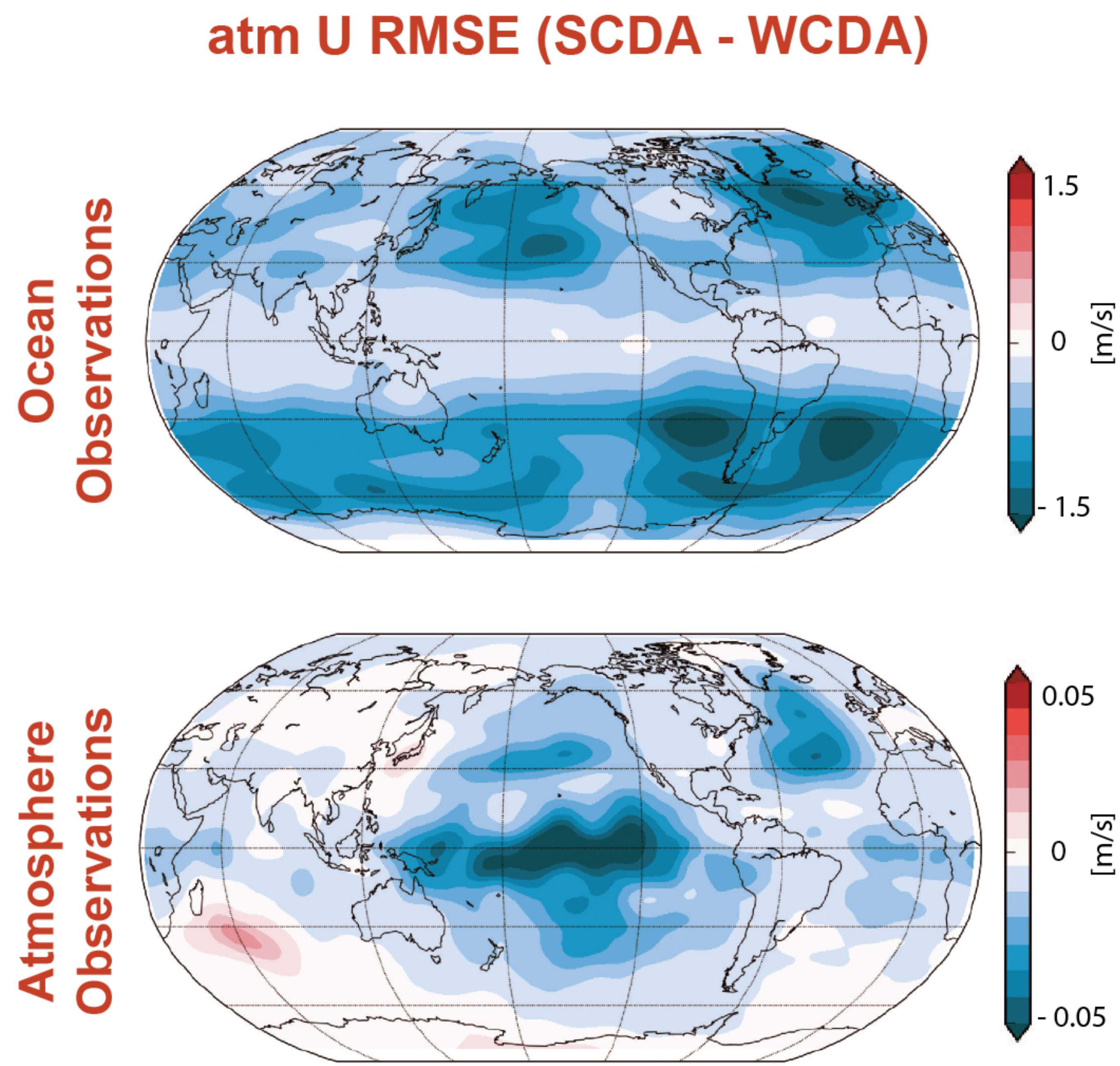


Ruiz−Barradas et al. (2017)



Geographically Dependent benefits of SCDA

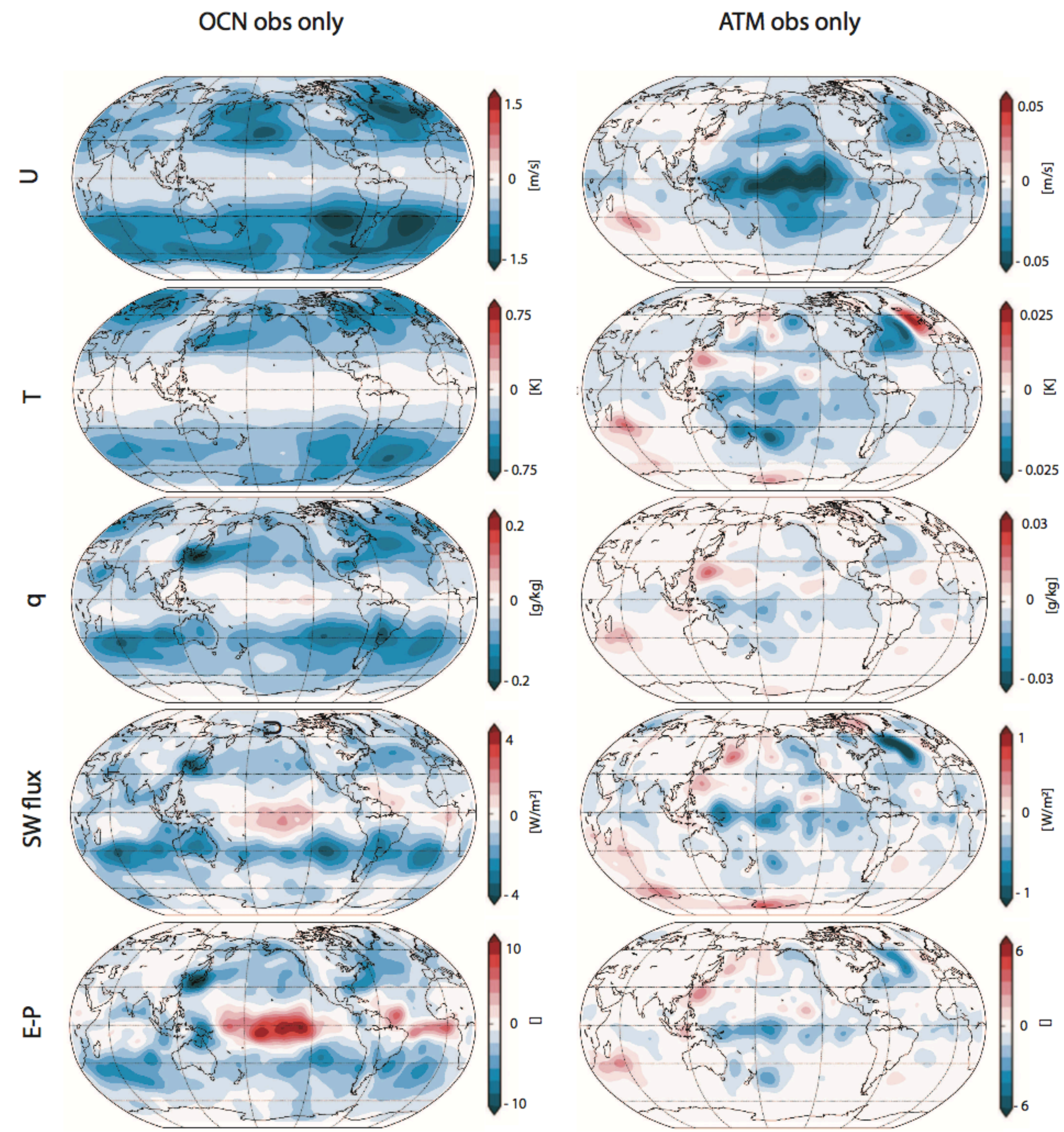
- Additional work with the SPEEDY/NEMO coupled model *(Sluka, 2018 Ph.D. Thesis)* indicates similar patterns of improvement due to SCDA
- For example: observations of the 'downstream' system improve 'upstream' state



Coupled Data Assimilation

- Additional experiments show that using SCDA to assimilate observations across domains tends to improve the coupled model state when observations are assimilated from the ‘downstream’ component to correct the ‘upstream’ state (w.r.t. information flow).

SCDA - WCDA errors (blue is improved)



Sluka Ph.D. Thesis

Estimating Vertical Error Correlations

Using real data, vertical localization appears necessary, e.g. in the Northern Atlantic/ Pacific (below), but the exact error correlations are model-dependent (right) - meaning there are lingering coupled modeling errors that need to be addressed.

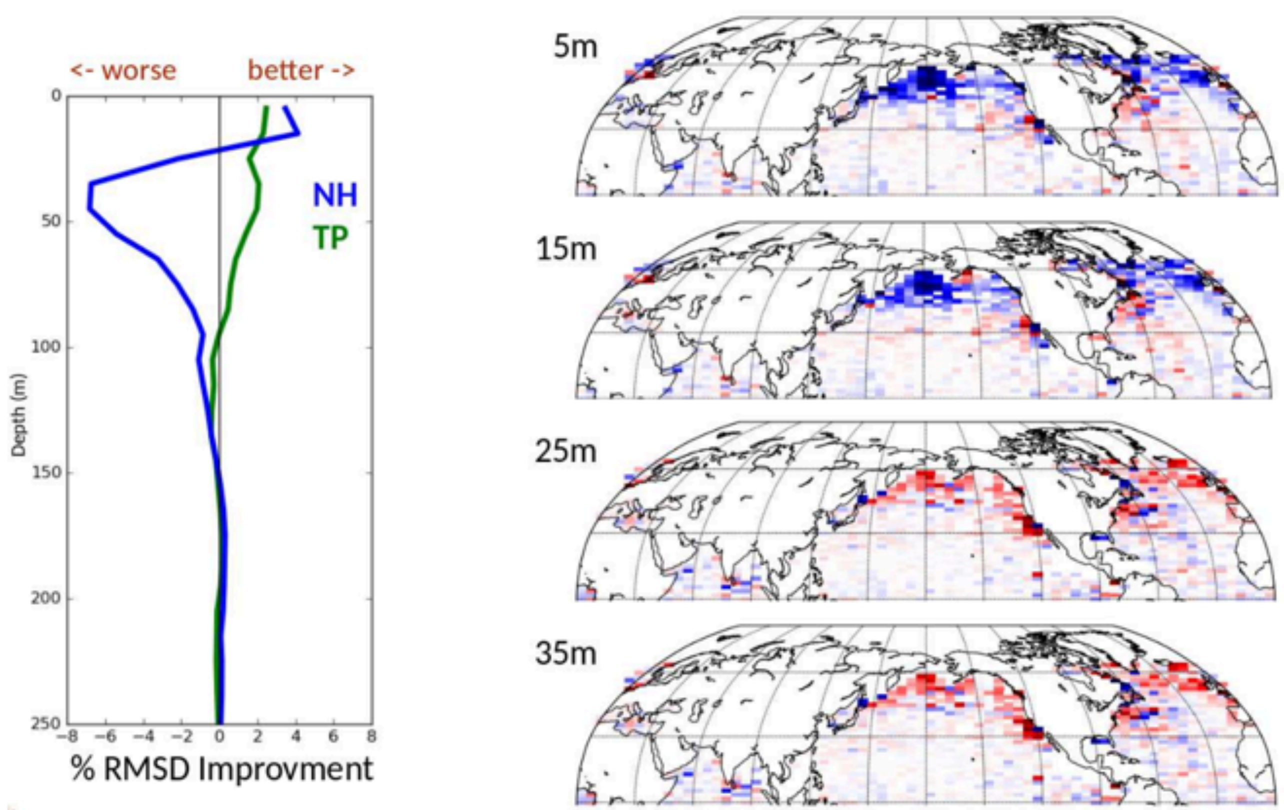
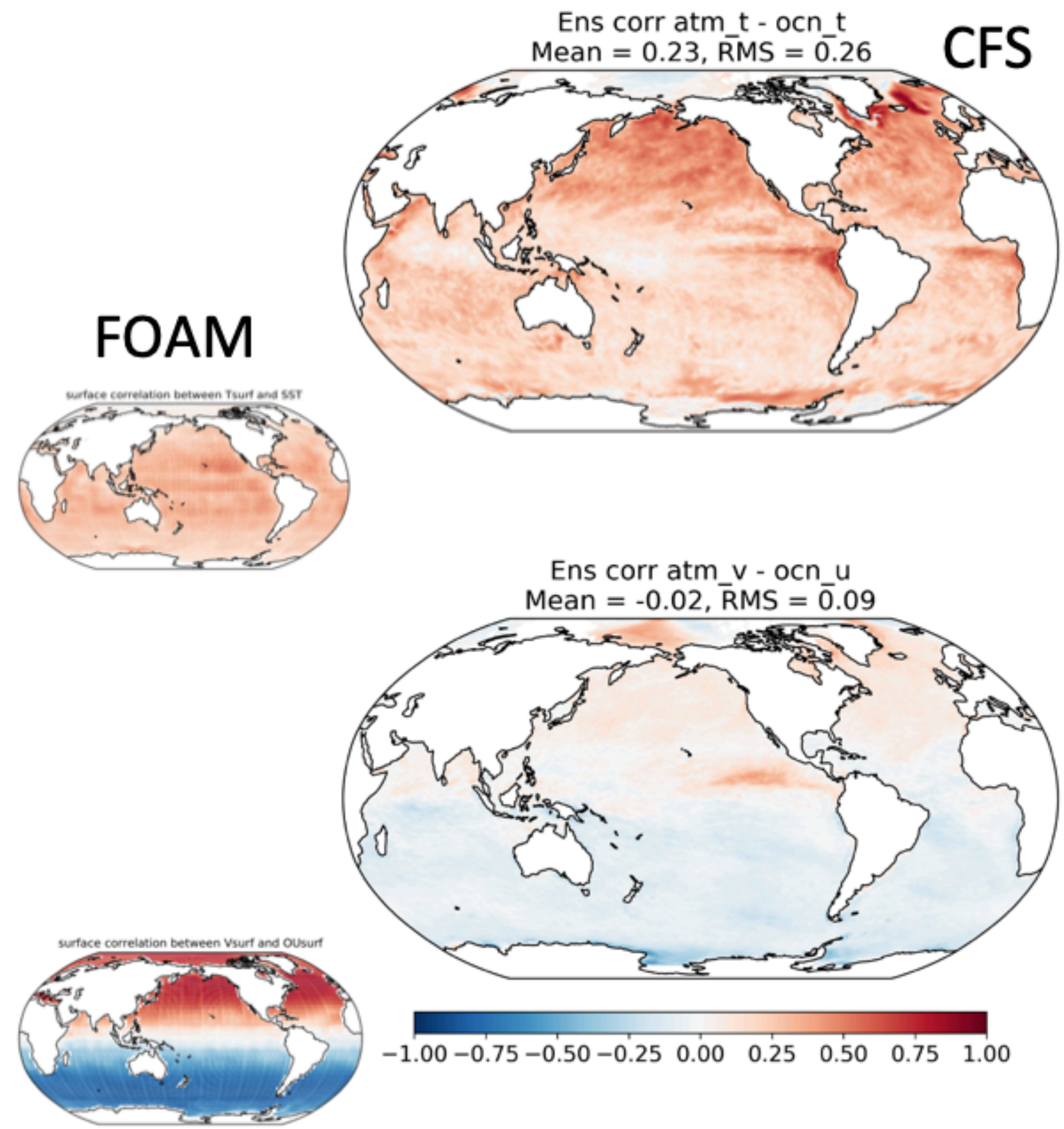


Figure 3.13: STRONG - WEAK change in observation minus forecast (O-F) RMSD for ocean temperature. Averaged over the tropics (TP) and Northern Hemisphere (NH) at various depths (left) and shown spatially (right). For the spatial plot blue is an RMSD improvement, red is a degradation.

Sluka (2018)



Courtesy: Takuma Yoshida

Conclusion

- SCDA produces superior coupled state estimates and forecasts in idealized scenarios (vs. uncoupled or WCDA)
- With appropriate configuration, 1-way strong coupling can also constrain an unobserved component of the coupled system
- Additional complications arise as model complexity increases, so increased study of CDA is needed with more realistic Earth system models.
- Applying SCDA to coupled models using real observational data will likely require improvements to the modeling at the interface.

* The work applying CDA to the MAOOAM coupled QG model will be available online in the *Journal of Advances in Modeling Earth Systems (JAMES) in the near future - Penny et al. (2019)*.