# Probabilistic state estimation for coupled models

**Craig H. Bishop**[1]

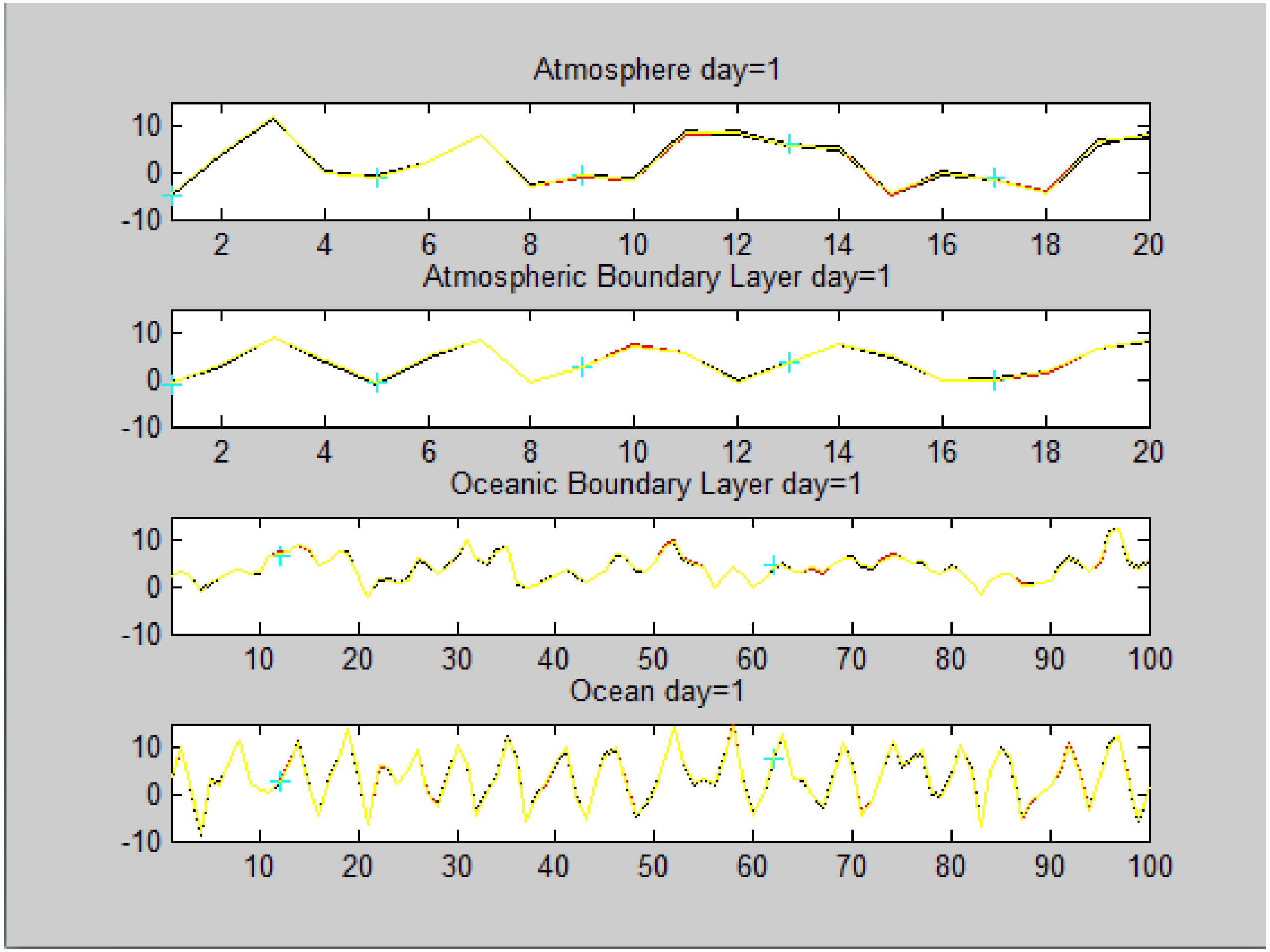
*[1]Marine Meteorology Division, Naval Research Laboratory, Monterey, USA*

# Outline

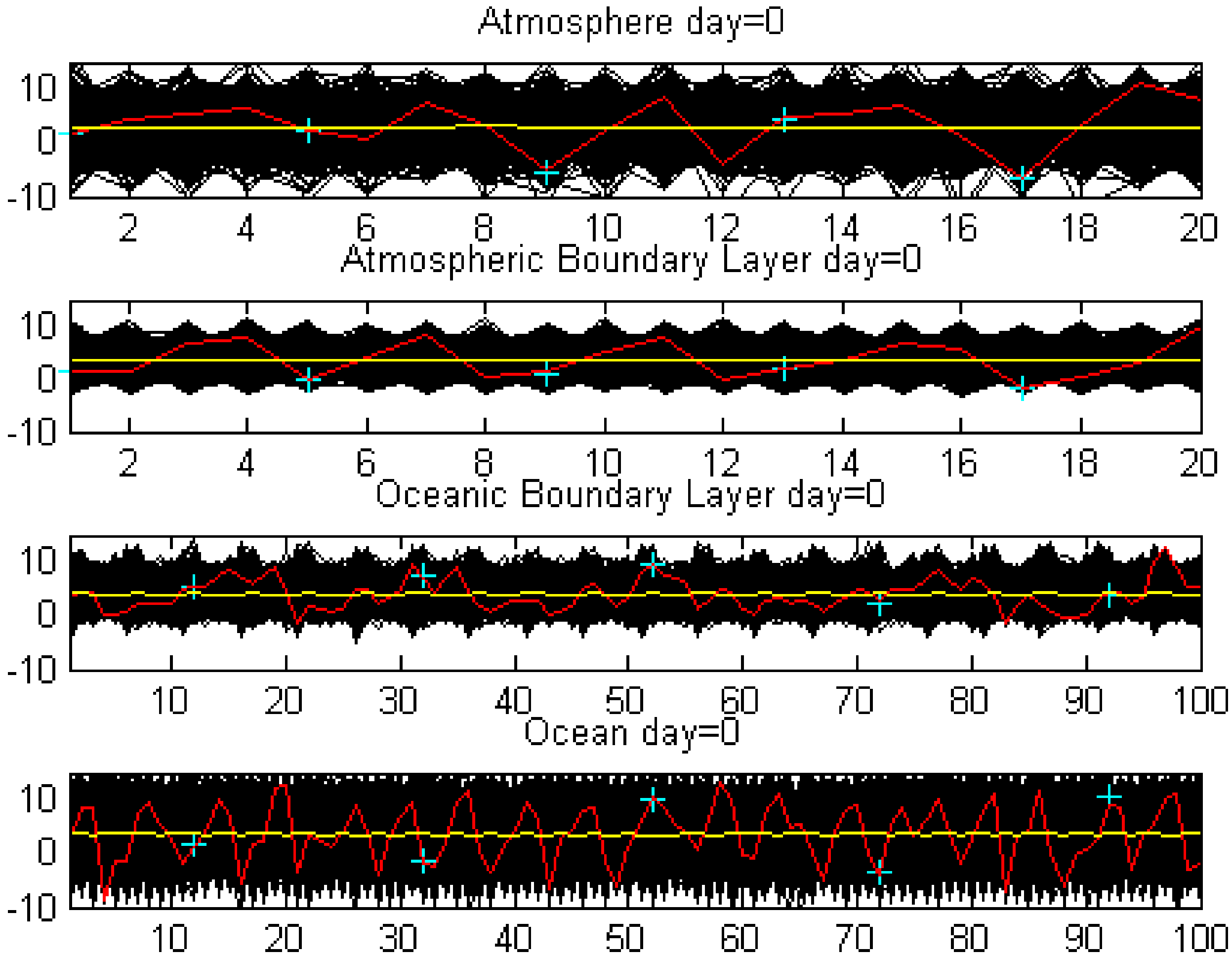
- Idealized coupled model of ocean-atmosphere system.
- Strongly coupled EnKF versus weakly coupled EnKF.
- On the extension of existing ocean (atmosphere) DA schemes to assimilate near interface obs (Frolov and Bishop, 2015).
- Local Ensemble Tangent Linear Models (LETLM) and their adjoints: an enabler for coupled 4DVAR and coupled model observation impact?

# Predictability of coupled model

## free coupled atmosphere-ocean ensemble



# Simple Coupled Data Assimilation Experiment



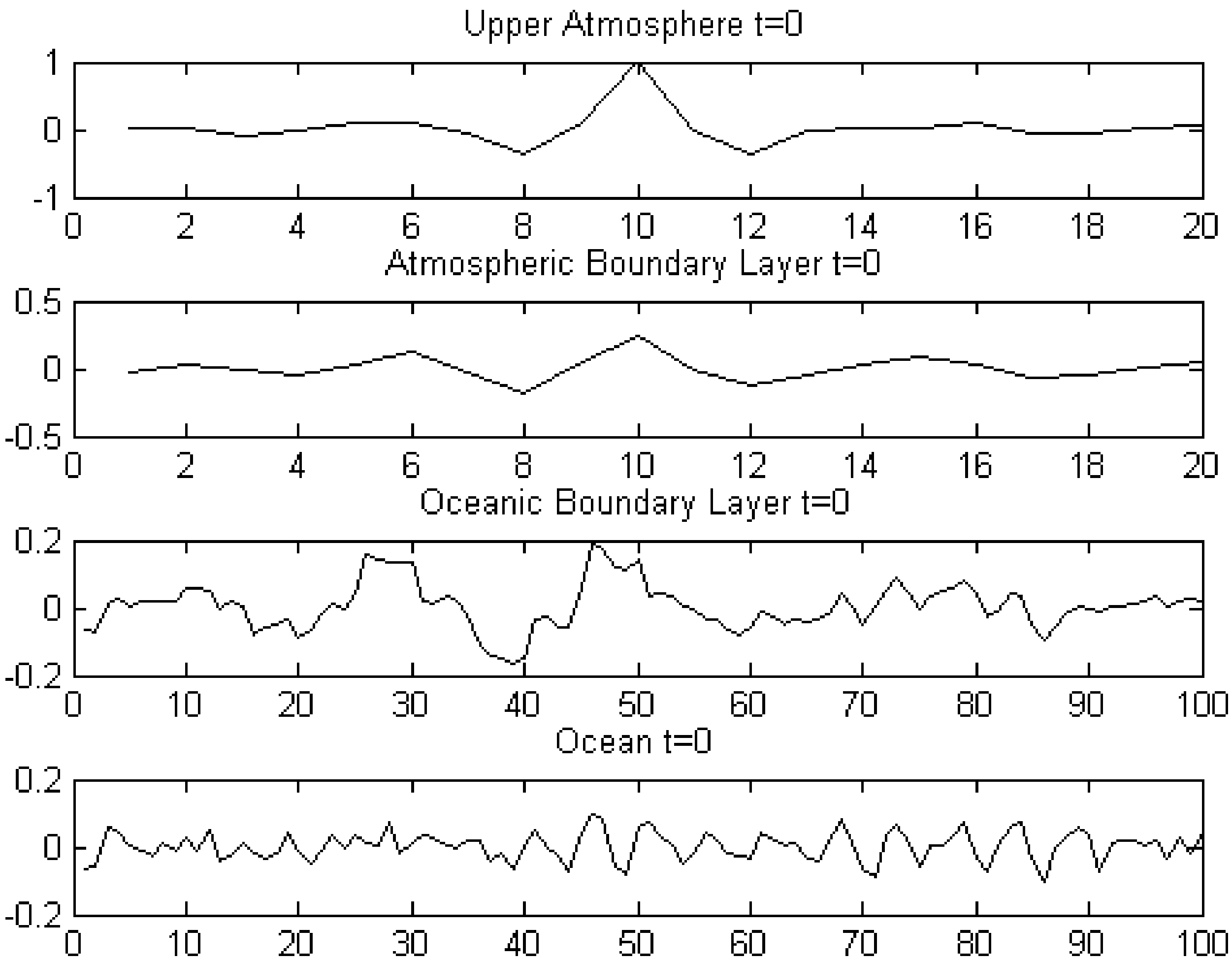

**Red line is true state**: Obtained from a random draw from the model climate

**Cyan +s are observations:** Truth plus observational noise

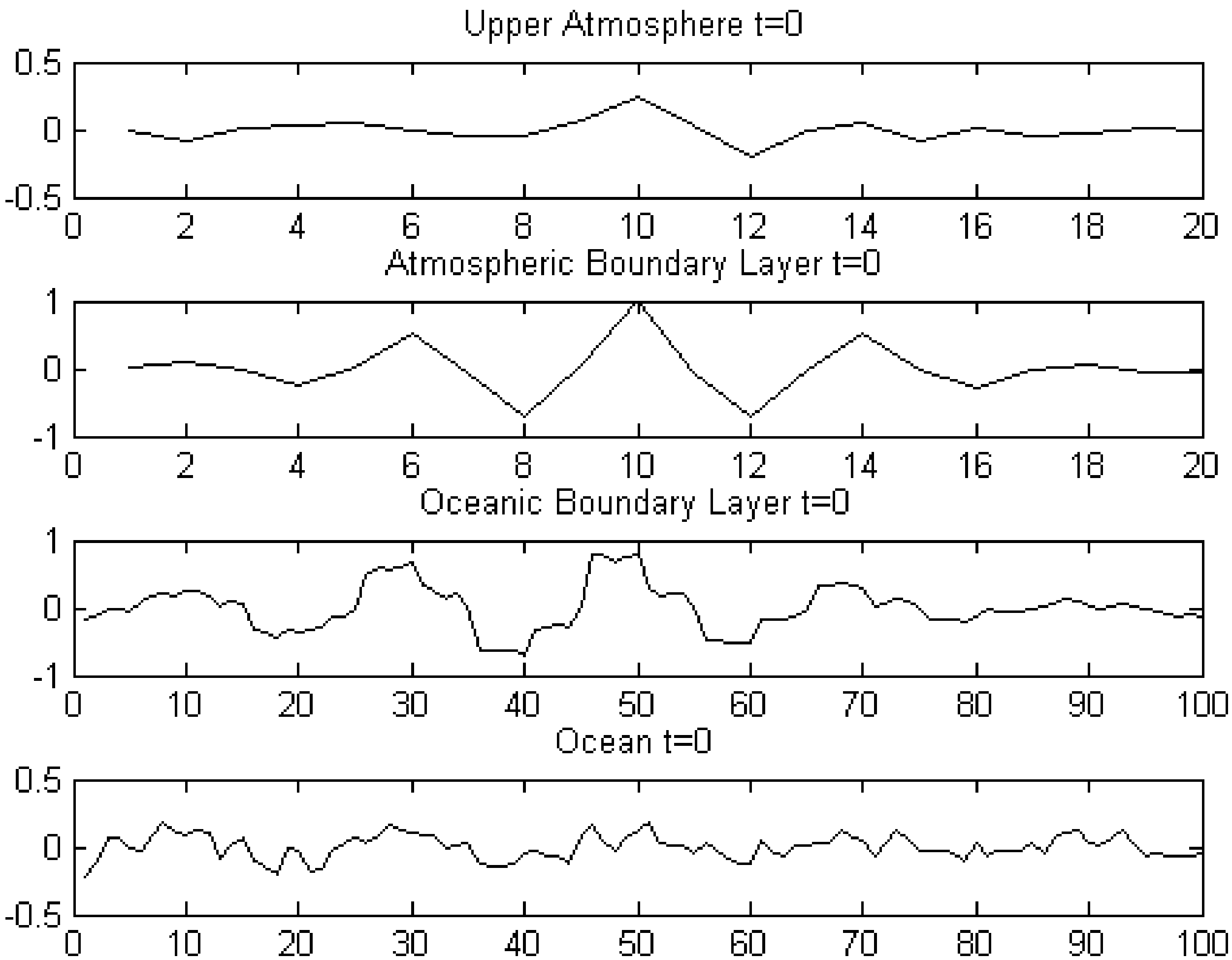
**Black lines are ensemble members**: Obtained from random draws from climatology

**Yellow line is ensemble mean**

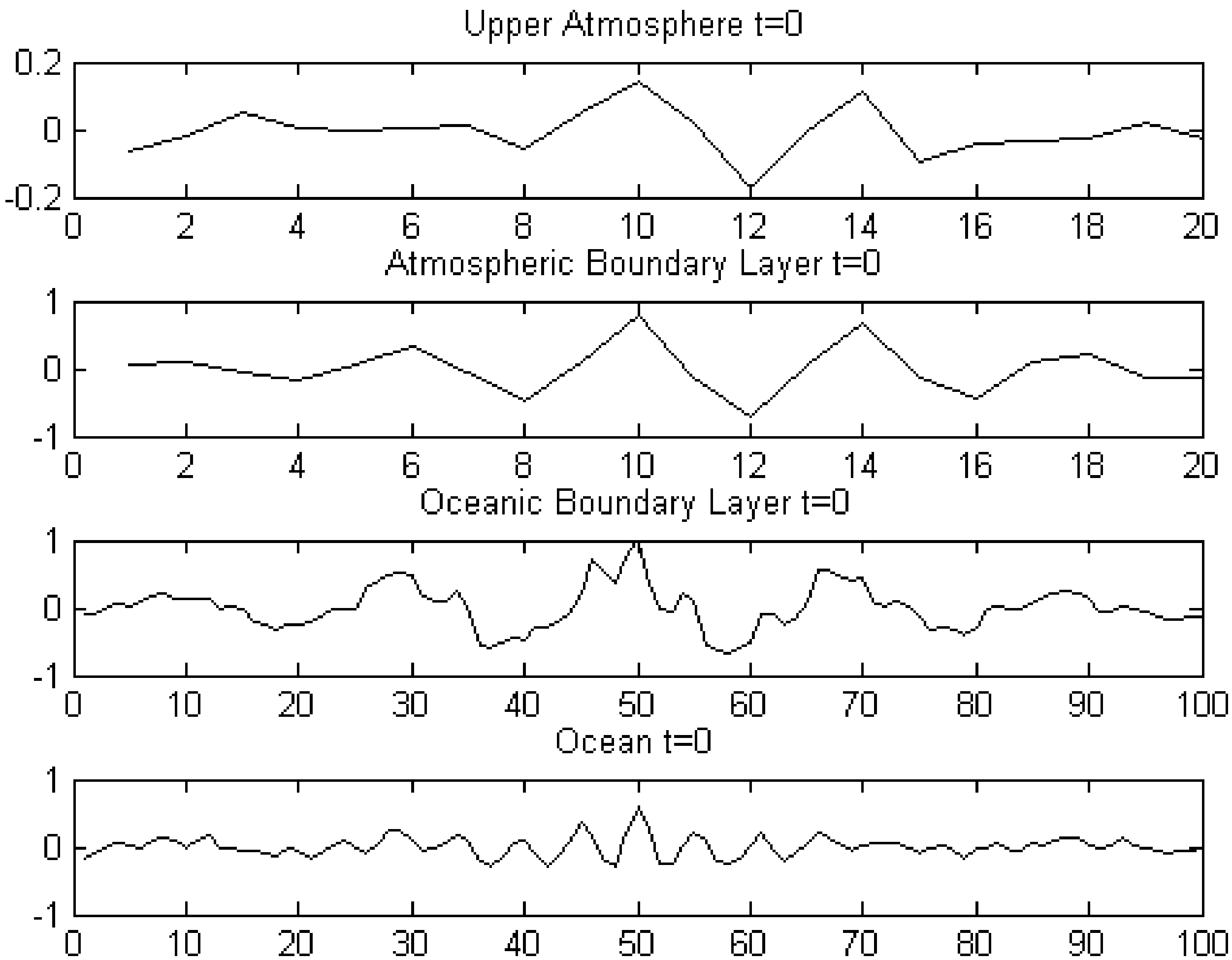
# Climatological correlation function for upper atmosphere



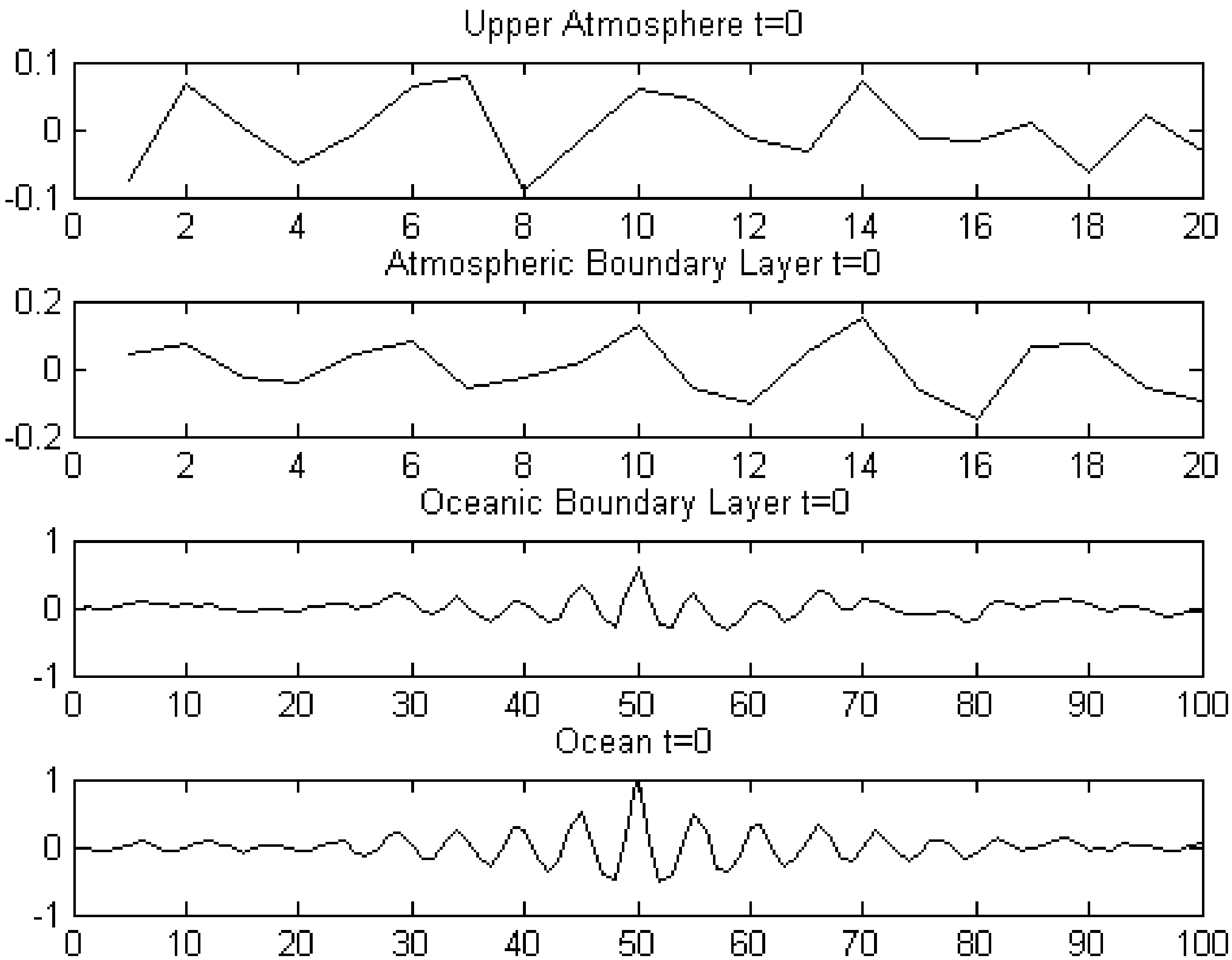
# Climatological correlation function for atmospheric BL



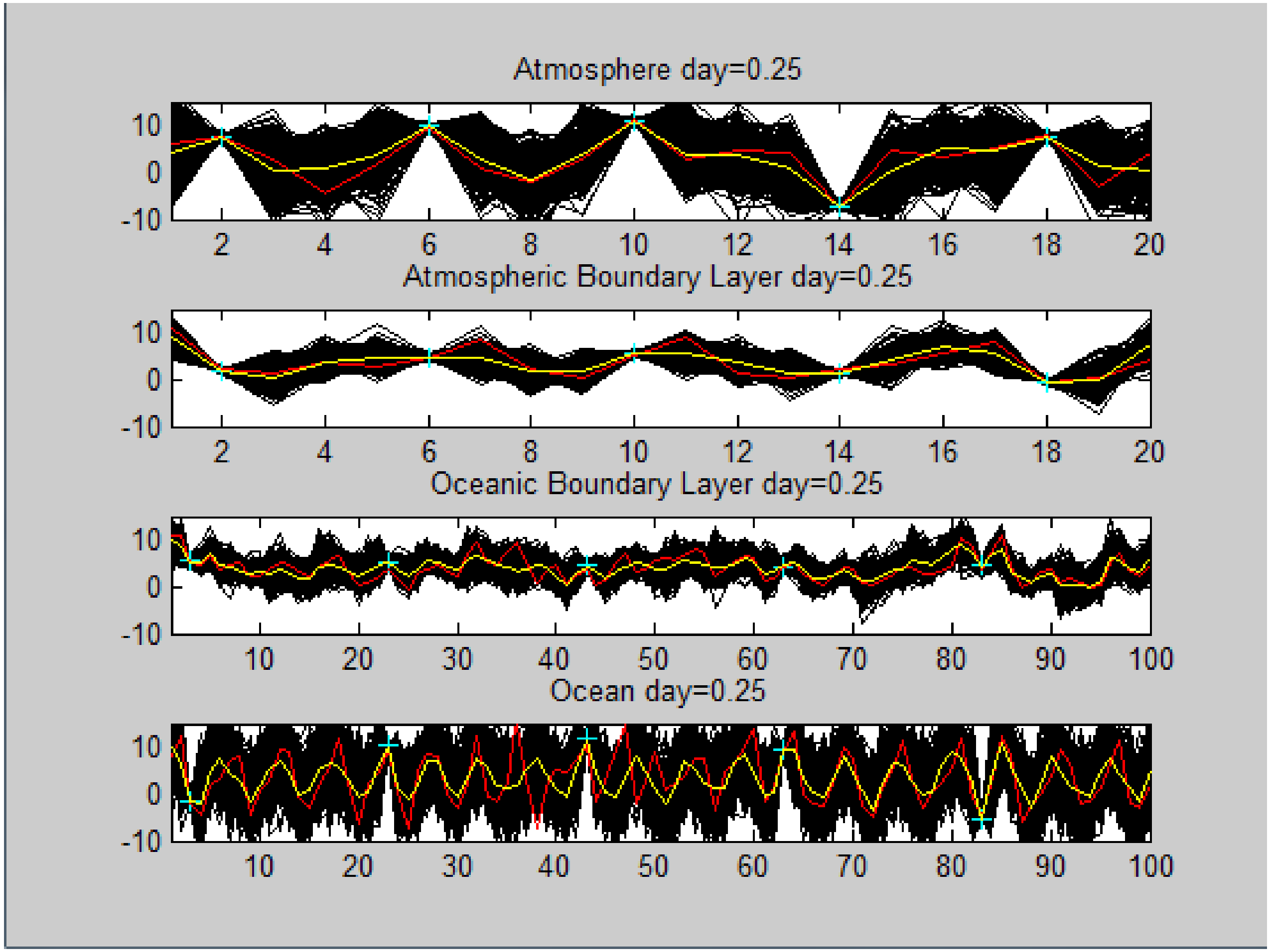
# Climatological correlation function for oceanic BL



# Climatological correlation function for ocean

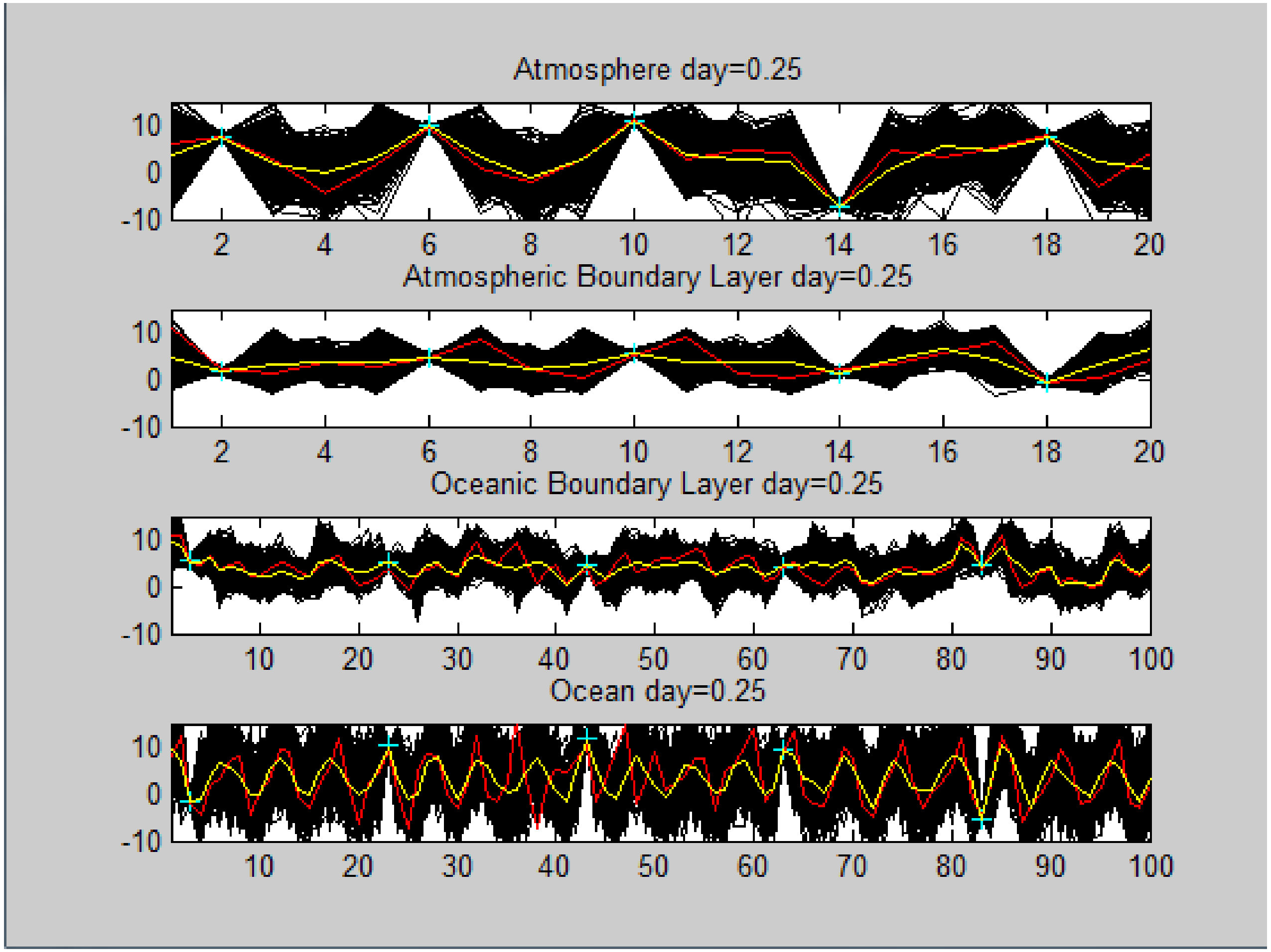


# *Strongly coupled* EnKF DA cycle




**Red line is true state**: Obtained from a random draw from the model climate
**Cyan +s are observations:** Truth plus observational noise
**Black lines are ensemble members**: Obtained from random draws from climatology
**Yellow line is ensemble mean**

# *Weakly coupled* EnKF DA cycle



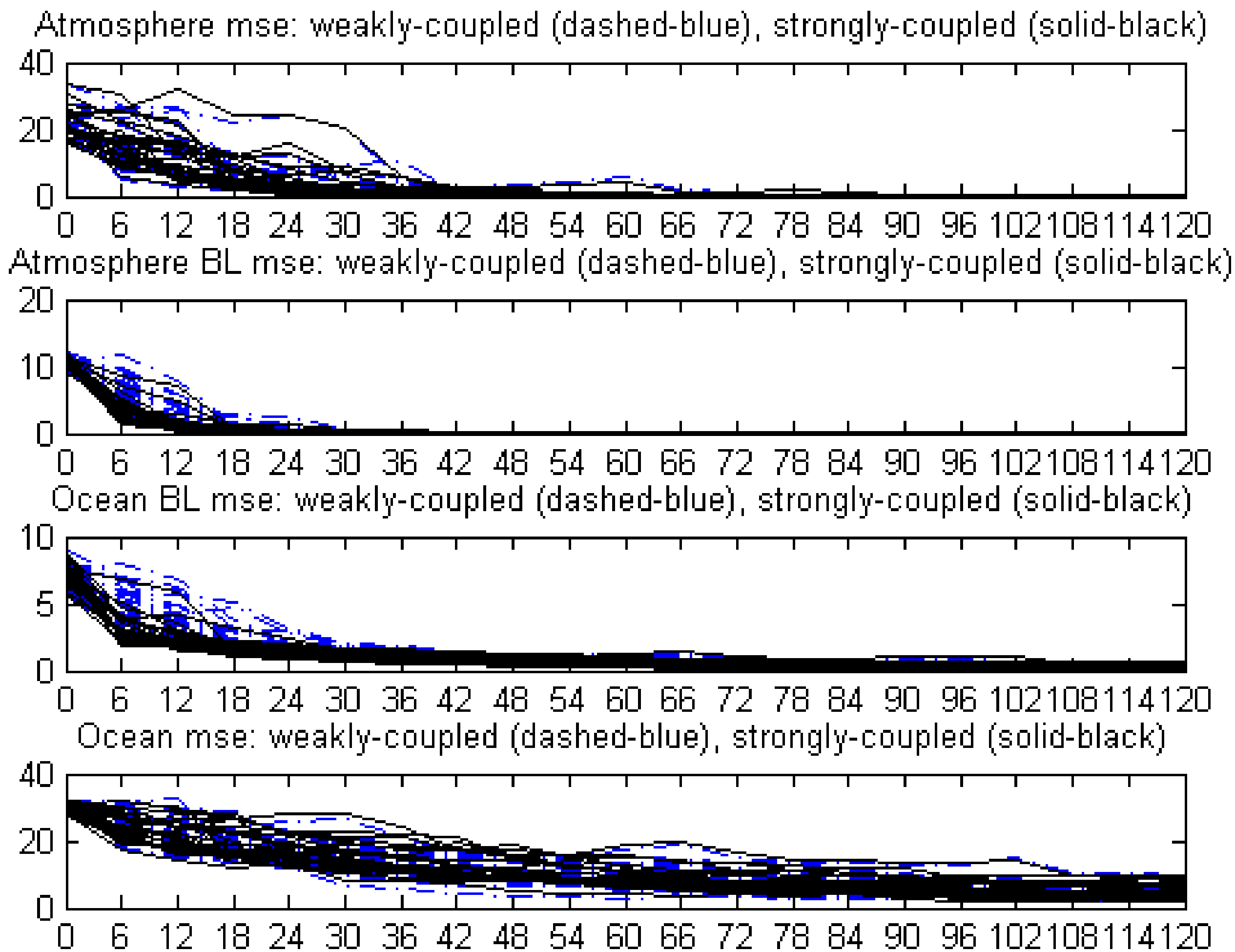

**Red line is true state**: Obtained from a random draw from the model climate

**Cyan +s are observations:** Truth plus observational noise

**Black lines are ensemble members**: Obtained from random draws from climatology

**Yellow line is ensemble mean**

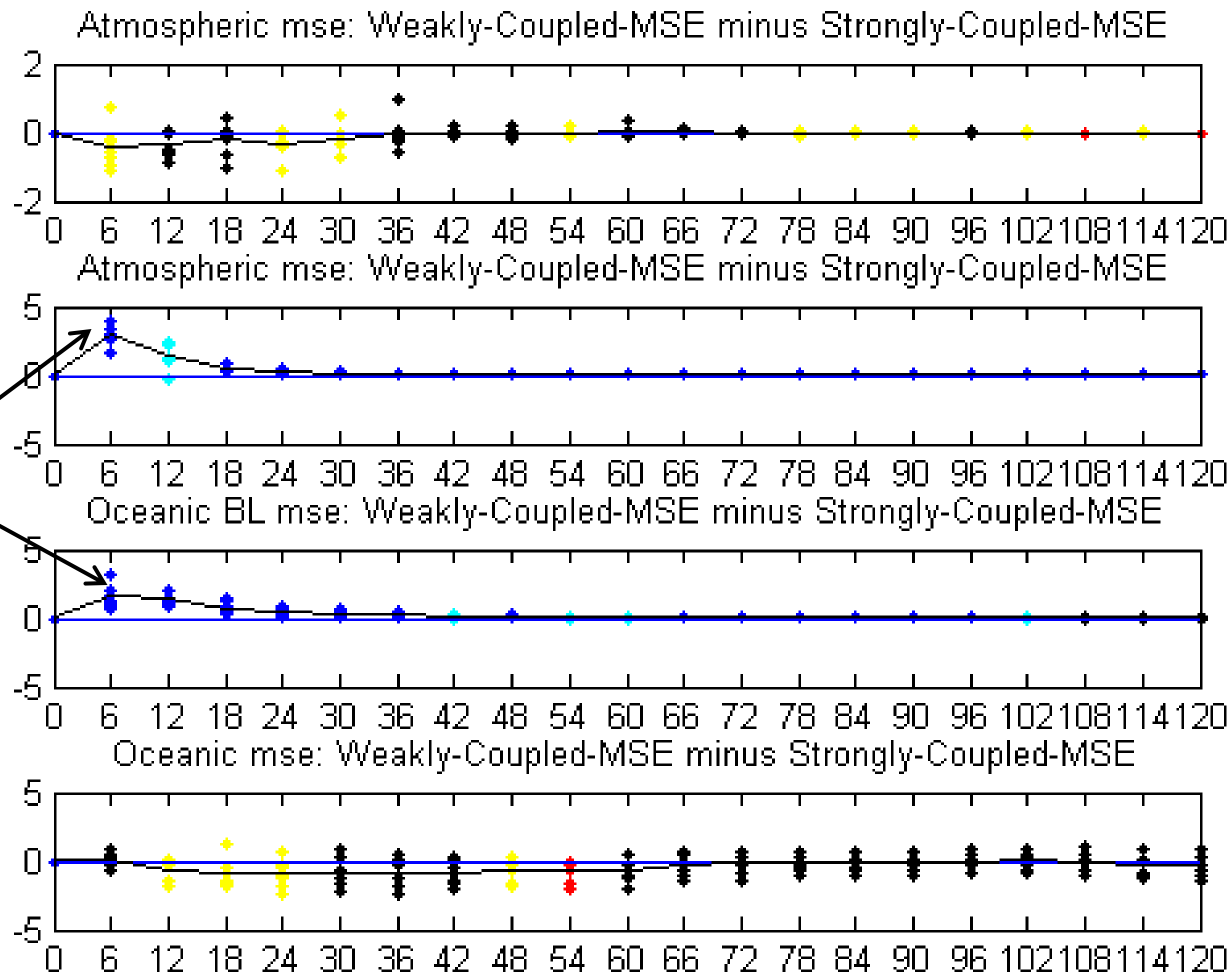
# Evolution of mean square error (mse) from 28 completely independent trials




y-axis is mse, x-axis is time (in hrs). DA was performed every 6 hrs. 7 independent DA experiments were performed. Blue lines pertain to weakly-coupled DA, black lines pertain to strongly-coupled DA.

**Strongly-coupled DA seems better than weakly coupled DA over the first (4-5 DA cycles) near the ocean-atmosphere interface**

# Significance of superiority of strongly coupled mse over first 20 cycles (5 days)

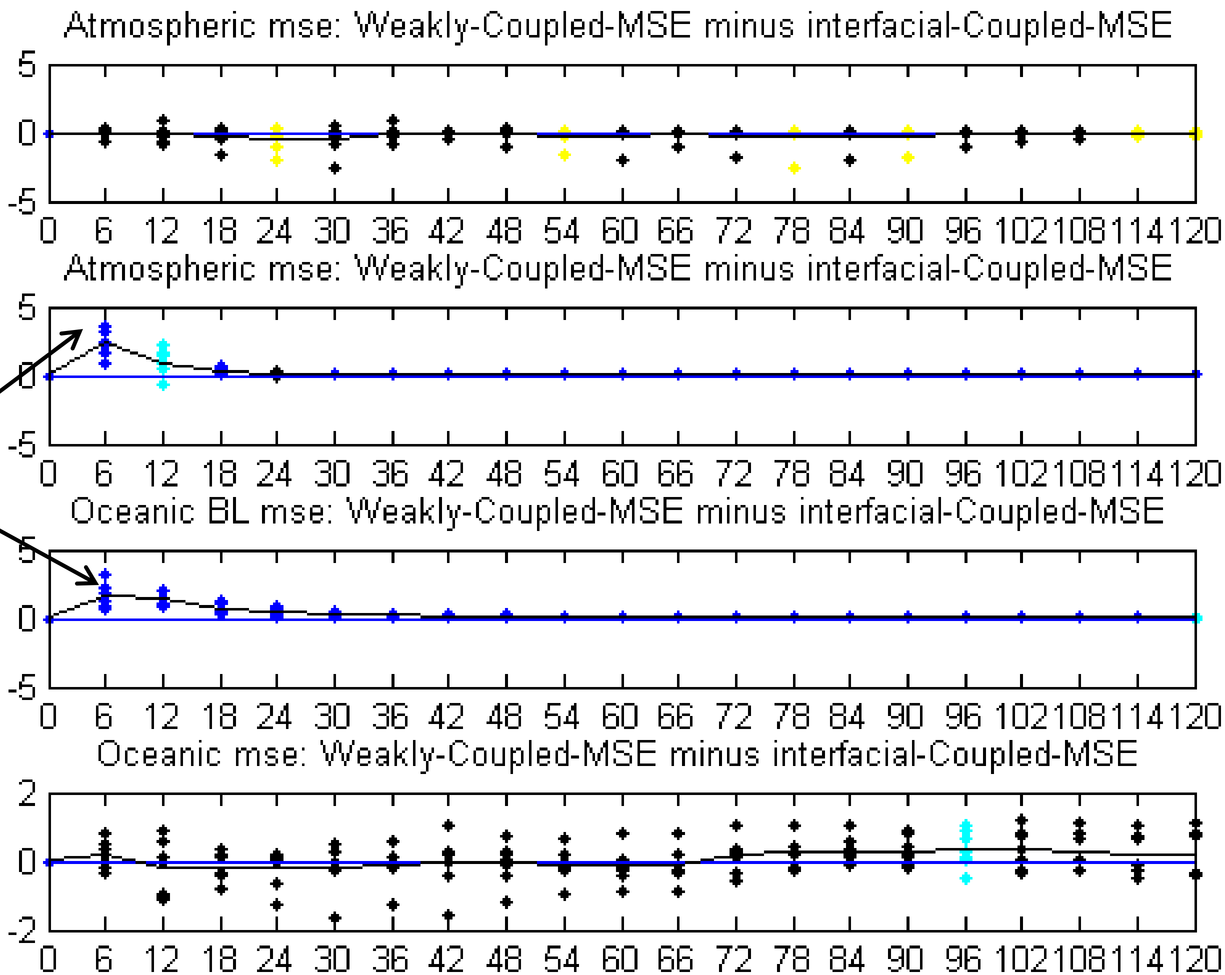



**20%-40% mse reduction due to coupled DA on first DA cycle**

Blue diamonds => 99% statistical confidence in superiority of strongly-coupled over weakly coupled

**Near ocean-atmosphere interface, strongly-coupled DA is significantly better than weakly coupled DA over first 5 days. Beyond that time, both methods give extremely small analysis error variances.**

# Significance of superiority of interfacial coupled mse over first 20 cycles (5 days)




**20%-40% mse reduction due to coupled DA on first DA cycle**

Frolov and Bishop (2015, MWR, in press)

**Near ocean-atmosphere interface, interfacial-coupled DA is significantly better than weakly coupled DA over first 5 days. <u>Interfacial DA does better away from the interface than fully coupled DA!</u>**

Black diamonds => statistical confidence, either way, is less than 95%

# The Coupled Model Tangent Linear (TL) Challenge

- Leading DA technique in meteorology is 4DVAR which relies on the Tangent Linear (TL) (or gradient) of the non-linear model and its Adjoint.
- No center has yet surmounted the barrier of building and maintaining TLs and adjoints of the entire coupled system.
  - A key challenge is that the sub-model components of coupled models may change on a time scale comparable to TL/adjoint development.
- Centers have created ensemble forecasts with the coupled model.
- If accurate TL/adjoints could be constructed from ensembles, it would be easier to make the TL/adjoints appropriately adjust to model changes.

# The Local Ensemble TLM: a possible enabler for strongly coupled 4DVAR

- Variables in a very local region around a model variable determine its change over a time step.
- Typically, only 27 – 350 variables will have any influence on the evolution of a single variable over a single time step.

$$
\begin{matrix}
T_{11} & T_{12} & T_{13} & T_{14} & T_{15} \\
T_{21} & T_{22} & T_{23} & T_{24} & T_{25} \\
T_{31} & T_{32} & \color{red}{T_{33}} & T_{34} & T_{35} \\
T_{41} & T_{42} & T_{43} & T_{44} & T_{45} \\
T_{51} & T_{52} & T_{53} & T_{54} & T_{55}
\end{matrix}
$$

In a 2D model, $T_{33}$*'s* evolution might only be affected by the variables within this local patch.

# The Local Ensemble TLM: a possible enabler for strongly coupled 4DVAR

- If an ensemble of $K$ perturbations are small enough to (automatically) satisfy the linearized governing equations then
- 

$$\left[\delta y_{m1},\delta y_{m2},\ldots,\delta y_{mK}\right]=\mathbf{M}_m\begin{bmatrix}\delta x_{11},\delta x_{12},\ldots,\delta x_{1nK}\\ \delta x_{21},\delta x_{22},\ldots,\delta x_{2K}\\ \vdots \\ \delta x_{n1},\delta x_{n2},\ldots,\delta x_{nK}\end{bmatrix};\text{ i.e. }\mathbf{y}_m=\mathbf{M}_m\mathbf{X}$$

where $\mathbf{M}_m$ is the local TL for the *m*th model grid point

$\delta x_{ij}$ is the jth ensemble pert of the *i*th variable at $t$

$\delta y_{mj}$ is the jth ensemble pert of the *m*th variable at $t+\delta t$

Hence, if ensemble size $K\geq n$, the TL is given by

$$\mathbf{M}_m=\mathbf{y}_m\mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T\right)^{-1}$$

This relation is PRECISELY correct for ensemble perts in the linear regime.

For large ensemble and perts in the non-linear regime, $\mathbf{M}_m$ is the BLUE of evolution.

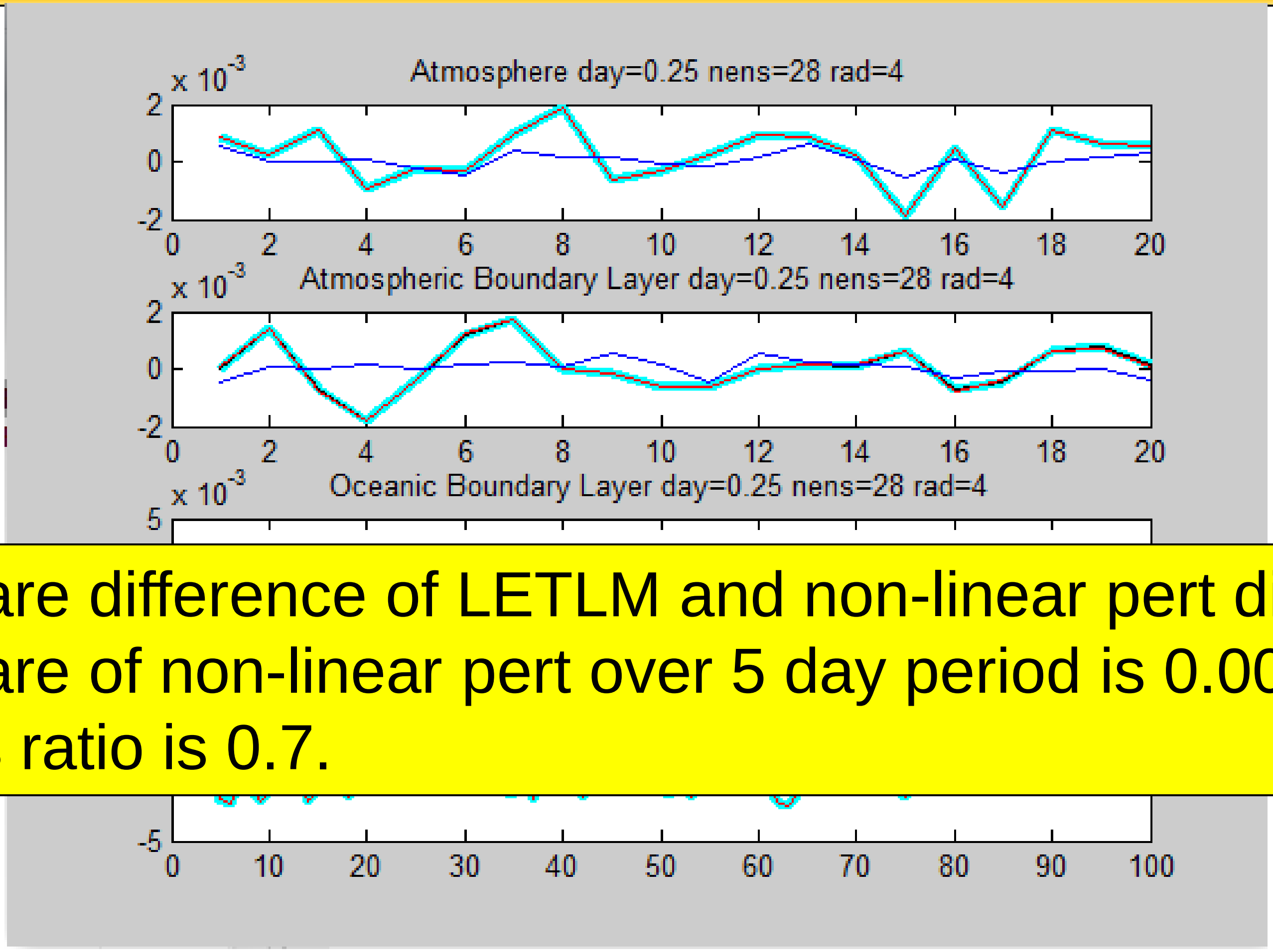
# Construction of Local Ensemble TLM (LETLM) for simple coupled model

- Simple coupled model (based on Model 1 of Lorenz, 2005) uses 2$^{nd}$ order Runge-Kutta time stepping. Vertical coupling via relaxation to neighbouring levels.
- Thus, patch size required 2-3 levels of 9 grid points
- Hence, needed $K$>27 ensemble members to precisely describe the linear dynamics – regardless of total number of variables in the model (240).

# Construction of Local Ensemble TLM (LETLM) for simple coupled model

- Simple coupled model (based on Model 1 of Lorenz, 2005) uses 2nd order Runge-Kutta time stepping. Vertical coupling via relaxation to neighbouring levels.
- Thus, patch size required 2-3 levels of 9 grid points.
- Hence, needed *K*>27 ensemble members to precisely describe the linear dynamics – regardless of total number of variables in the model (240).

# Tests of coupled LETLM with *K*=28 ensemble members




Mean square difference of LETLM and non-linear pert divided by mean square of non-linear pert over 5 day period is 0.0007. For ETLM, this ratio is 0.7.

**Black: Difference between 2 non-linear trajectories** (perfect TL predicts this

**Cyan line is global ETLM with *K*=480 members:** (Tracks black line perfectly)

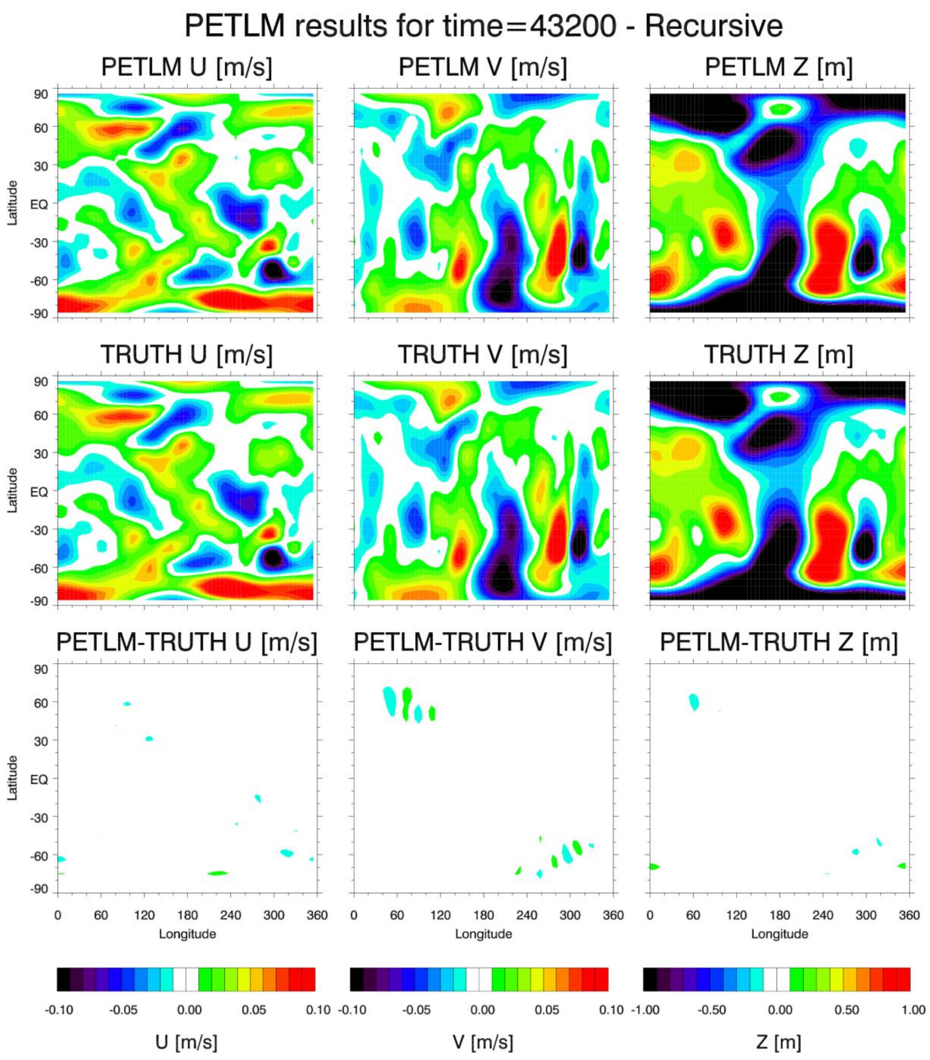
**Red line is LETLM**: (Often hidden by black line)

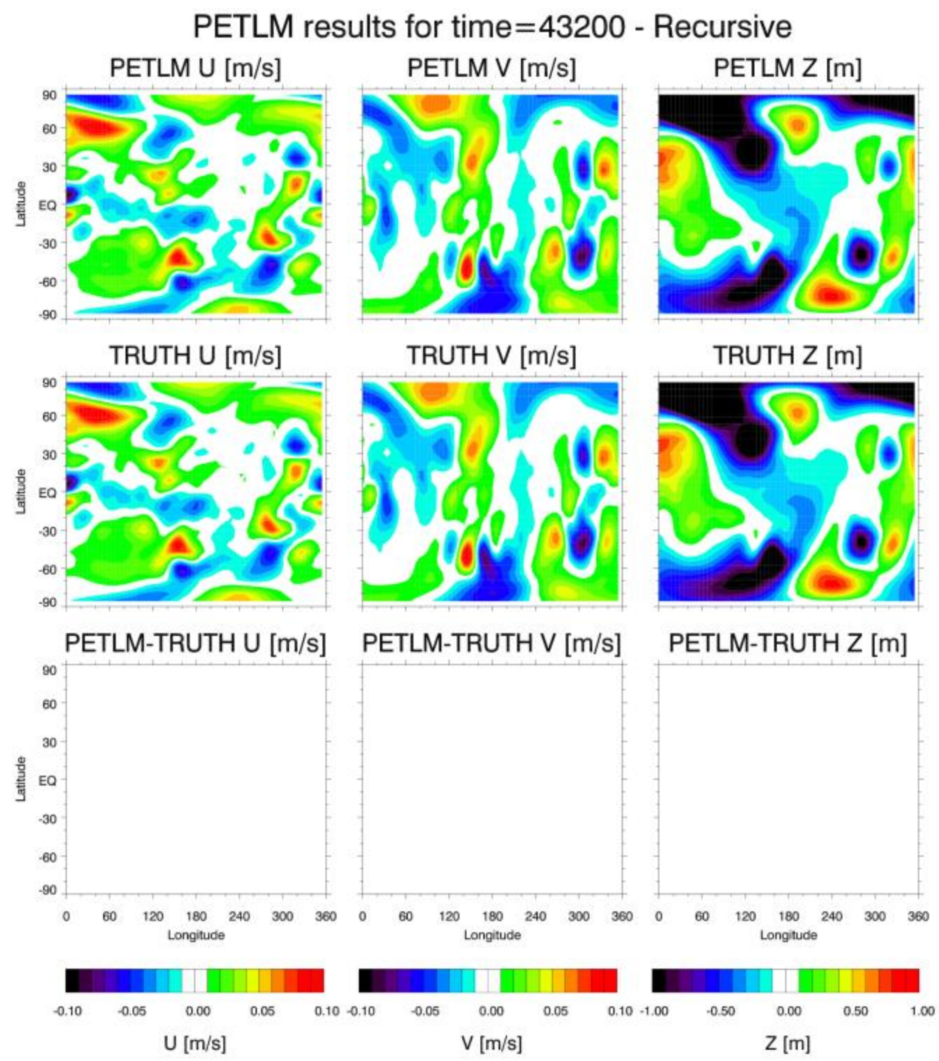
**Blue line is global ETLM with 28 members:** (Terrible performance)

# Tests of LETLM with a global spectral shallow water model ~*(T21,1500 DOF)*

- Number of ensemble members - 100
- Radius of the influence volume – 3000 km
- This gave 621 to 2601 variables per influence volume, depending on latitude.
- Length of integration – 12 hours
- Time step for non-linear model ~ 0.1 hour
- Time step for LETLM – 1 hour
- **LETLM will only be accurate if the number of dynamical Degrees Of Freedom (DOF) is no greater than ensemble size of 100.**
- **(Thanks to NRL DC's Doug Allen for putting this test together)**

# Tests of LETLM with a global spectral shallow water model ~*(T21,1500 DOF)*

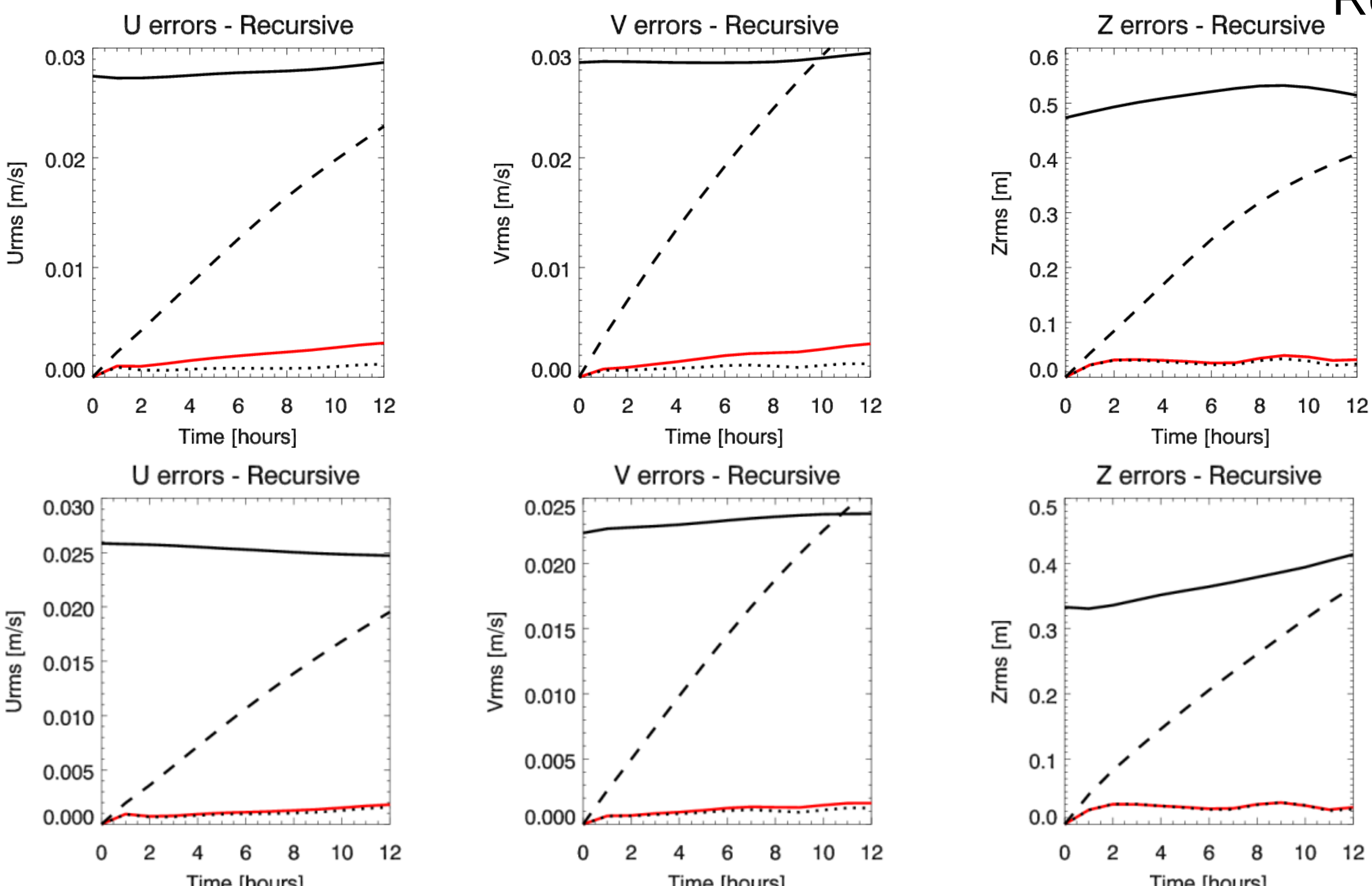



Multiple Ob (Pbclim) **12-hr forecasts** Multiple Ob (Pbens)

# Tests of LETLM with a global spectral shallow water model ~*(T21,1500 DOF)*

Solid: global RMS
Dashed: persistence
Dotted: TLM
Red: LETLM




Multiple obs (Pbclim)

Multiple obs (Pbens)

# Tests of LETLM with a global spectral shallow water model ~*(T21,1500 DOF)*

Solid: global RMS
Dashed: persistence
Dotted: TLM
Red: LETLM

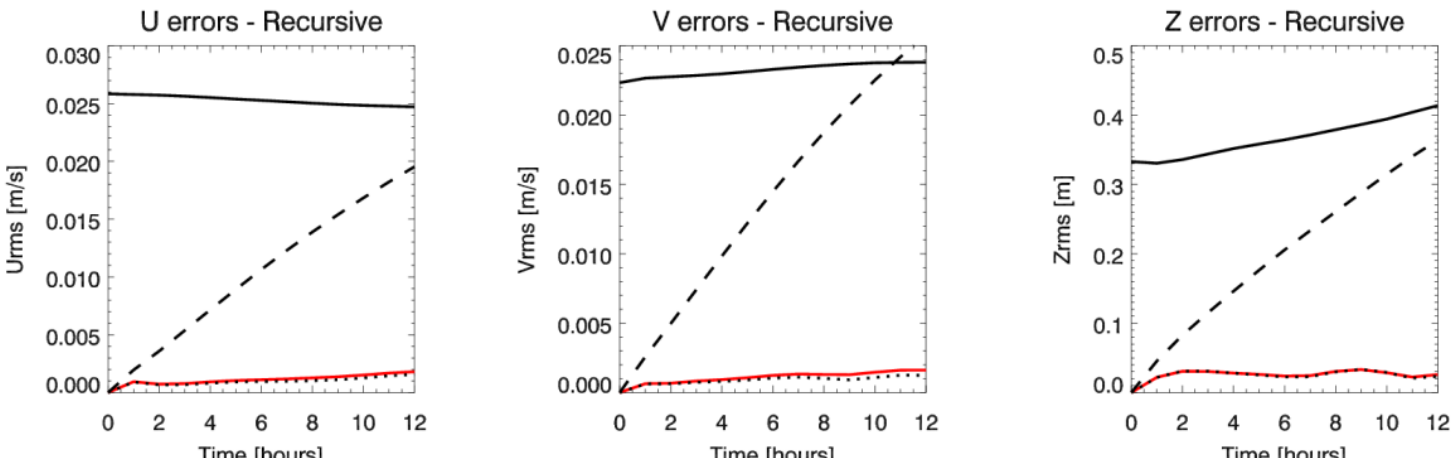
U errors - Recursive    V errors - Recursive    Z errors - Recursive

**These results demonstrate that accurate LETLMs are possible even when:**

**i. The LETLM time step is longer than the NLM time step**

**ii. Number of variables in influence volume exceeds the number of ensemble members**

**iii. Ensemble perturbation size is in the weakly non-linear regime**




Multiple obs (Pbens)

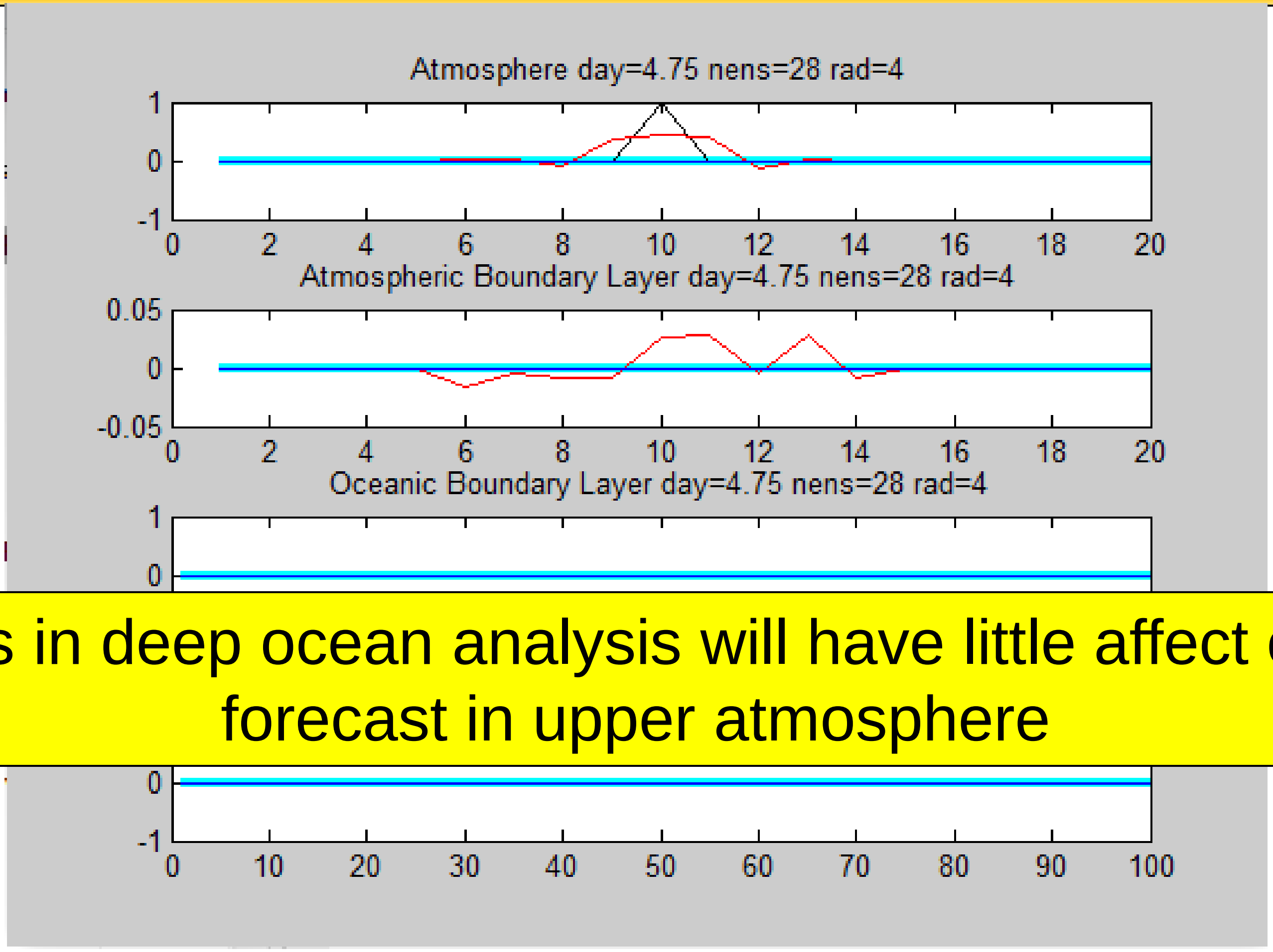
# Tests of adjoint of coupled LETLM with *K*=28 ensemble members

Checked that $\mathbf{x}^T(\mathbf{M}\mathbf{x}) = \left[\mathbf{x}^T(\mathbf{M}\mathbf{x})\right]^T = \mathbf{x}^T\mathbf{M}^T\mathbf{x}$ to machine precision over 5 day window.

Also checked that adjoint of LETLM with 28 members well-approximated transpose of ETLM with 480 members.

# Accurate LETLMs enable SVs and analysis and observation sensitivity

Target region in upper atmosphere



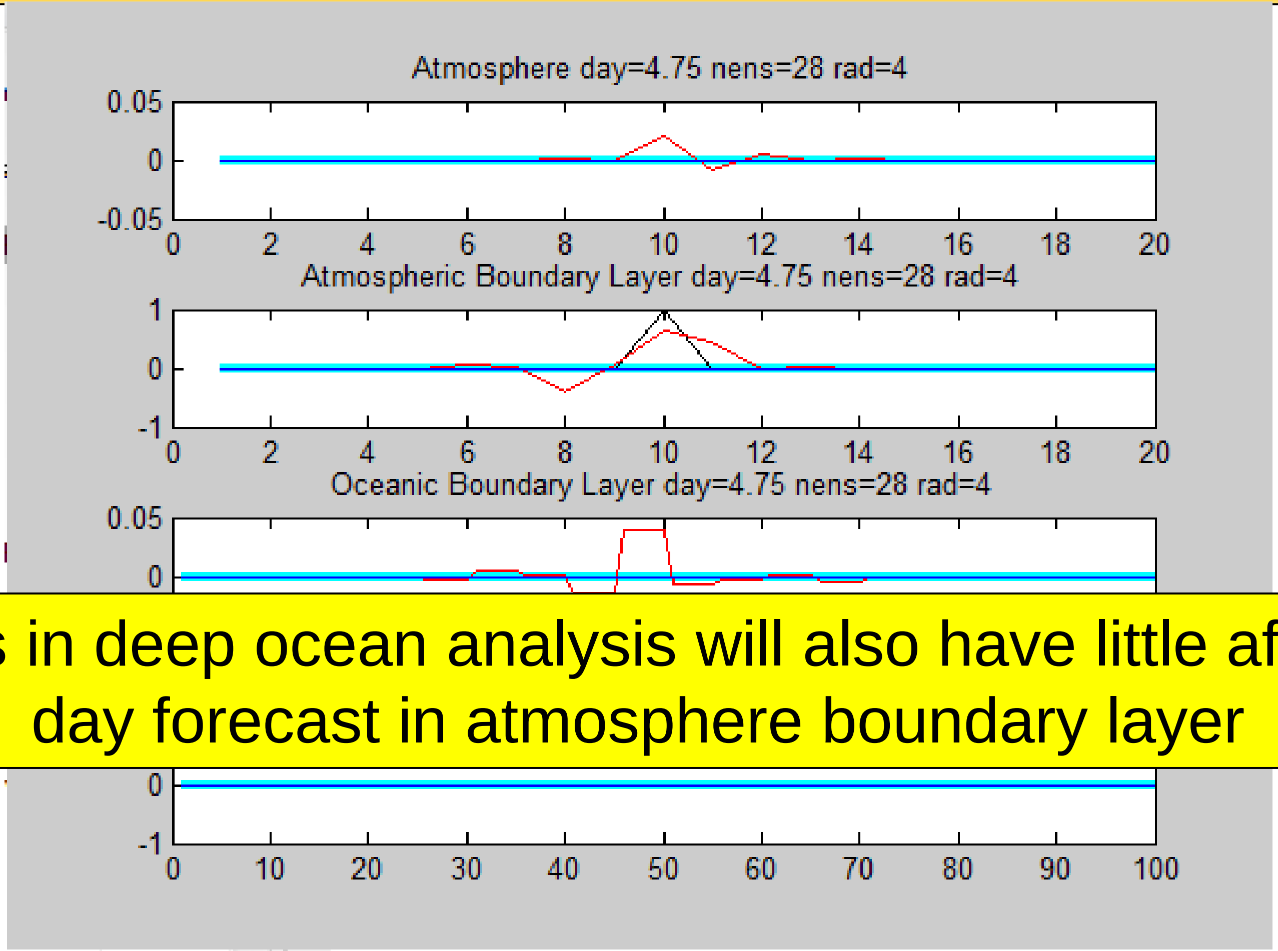

Changes in deep ocean analysis will have little affect on 5 day forecast in upper atmosphere

**Black: Location of target grid point (verification time is t=5 days)**

**Red line is d(target variable)/d(local variable)**: (i.e. good places to observe to improve forecast of target grid point at t=5 days).

# Accurate LETLMs enable SVs and analysis and observation sensitivity

Target region in atmospheric boundary layer



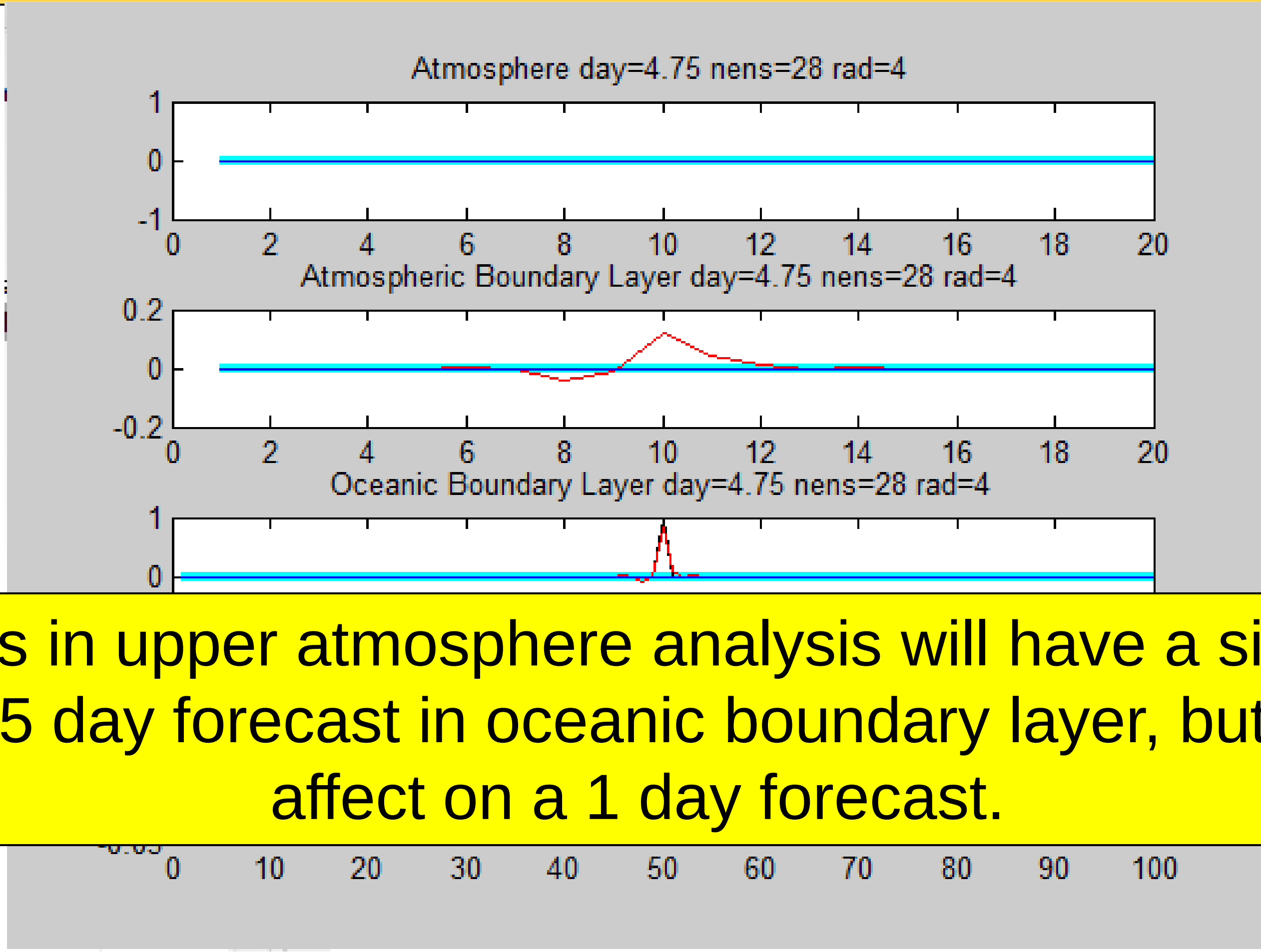

Changes in deep ocean analysis will also have little affect on 5 day forecast in atmosphere boundary layer

**Black: Location of target grid point (verification time is t=5 days)**

**Red line is d(target variable)/d(local variable)**: (i.e. good places to observe to improve forecast of target grid point at t=5 days).

# Accurate LETLMs enable SVs and analysis and observation sensitivity

Target region in oceanic boundary layer



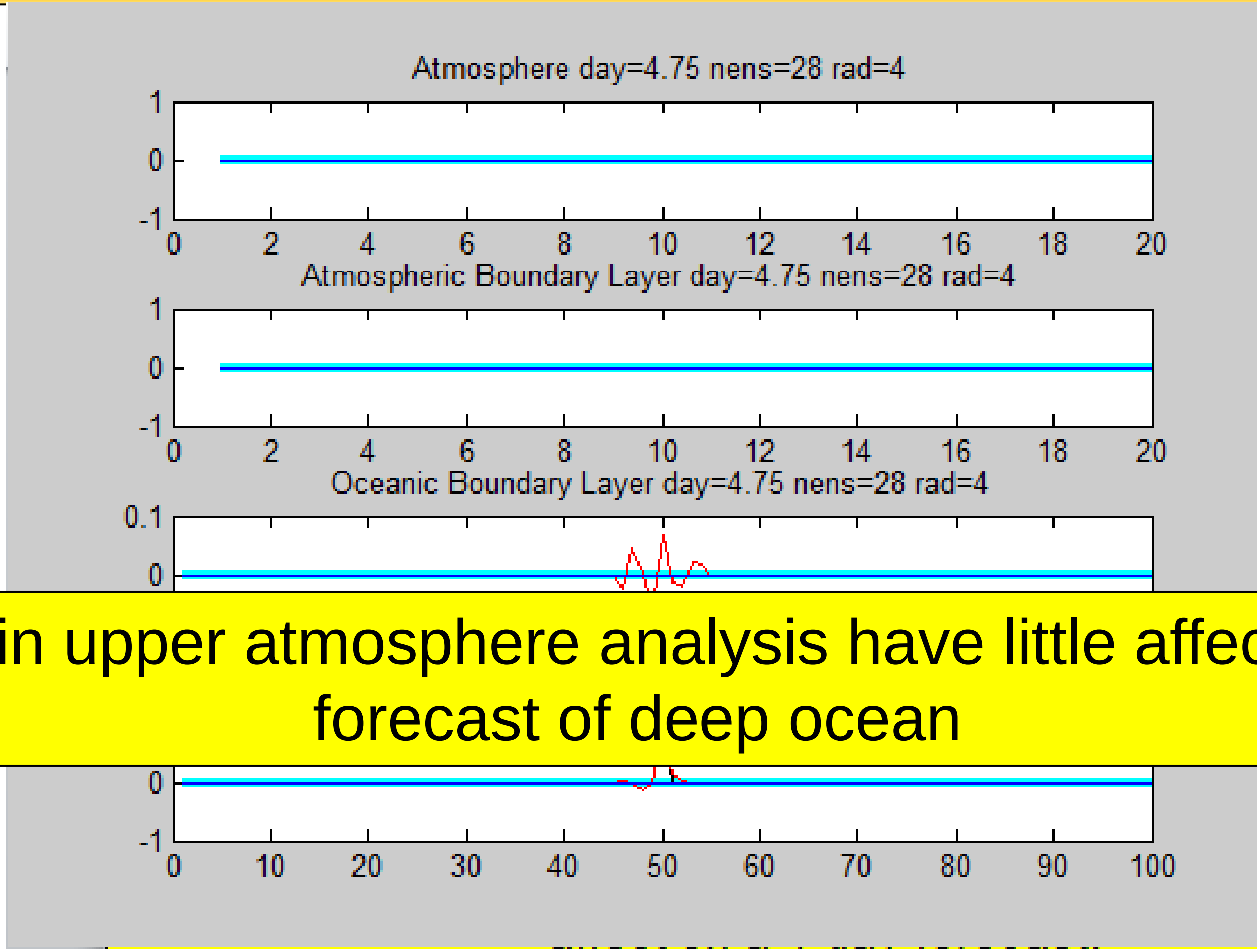

Changes in upper atmosphere analysis will have a significant affect on 5 day forecast in oceanic boundary layer, but very little affect on a 1 day forecast.

**Black: Location of target grid point (verification time is t=5 days)**

**Red line is d(target variable)/d(local variable)**: (i.e. good places to observe to improve forecast of target grid point at t=5 days).

# Accurate LETLMs enable SVs and analysis and observation sensitivity

Target region in deep ocean

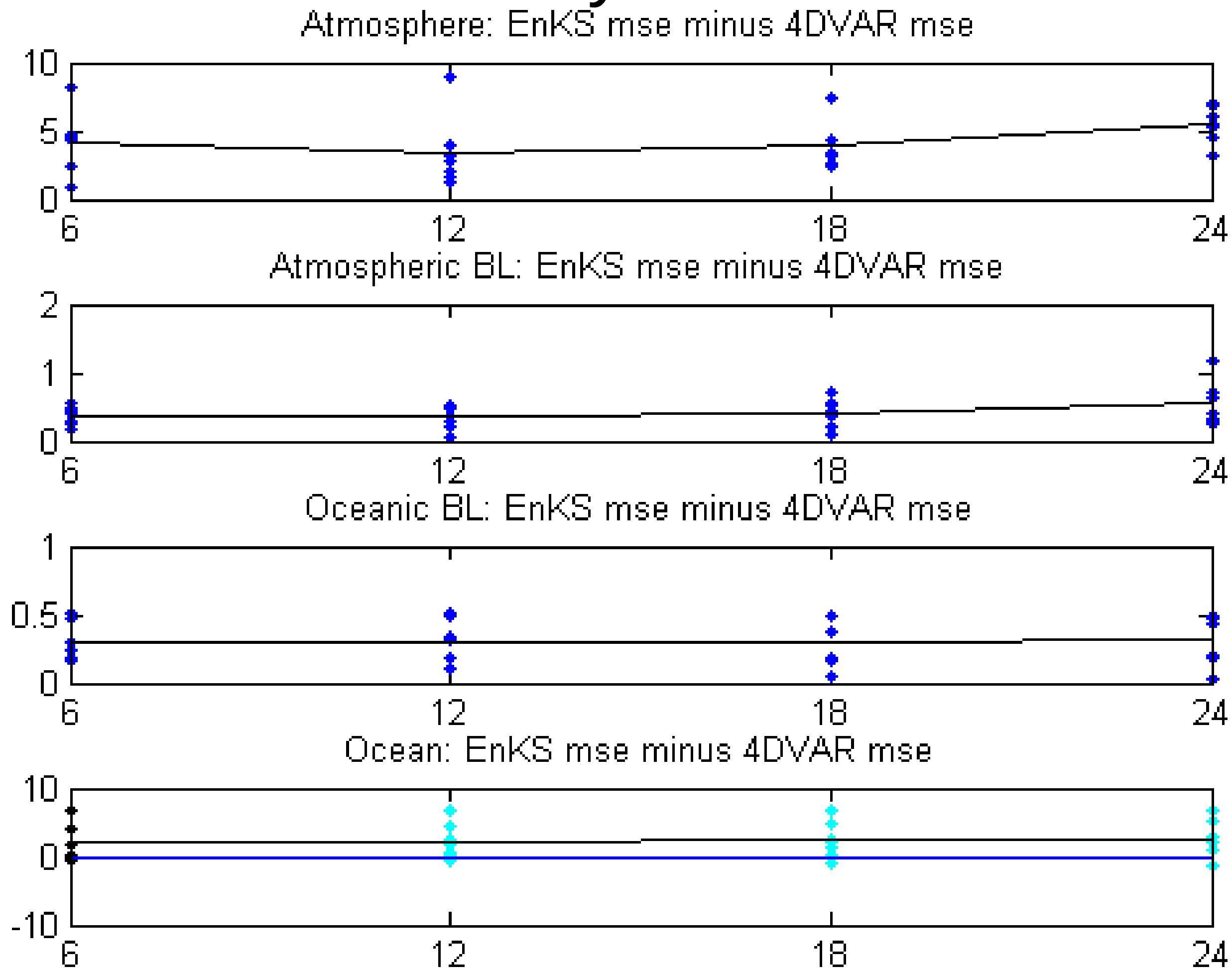



Changes in upper atmosphere analysis have little affect on 5 day forecast of deep ocean

**Black: Location of target grid point (verification time is t=5 days)**

**Red line is d(target variable)/d(local variable)**: (i.e. good places to observe to improve forecast of target grid point at t=5 days).

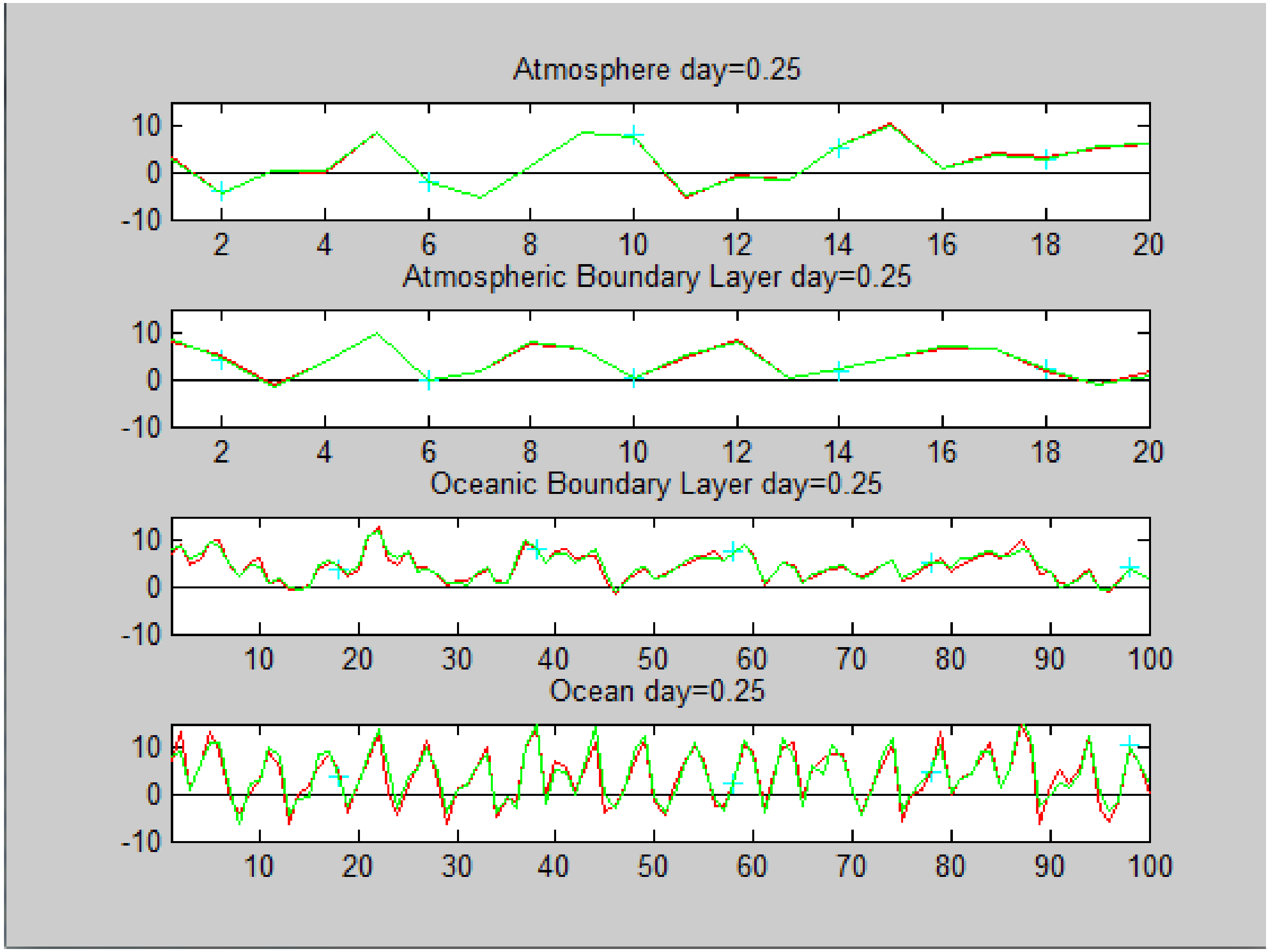
# Significance of superiority of 4DVAR over EnKS for 1 day window




Blue diamonds => 99% statistical confidence in superiority of strongly-coupled over weakly coupled

**4DVAR mode (4 outer loops) has profoundly lower mse than EnKS in first 24 hrs (a strongly non-linear regime) – differences much less at later times.**

# Long (5 day window) strong constraint 4DVAR (10 outer loops)




**Red line is true state**: Obtained from a random draw from the model climate

**Cyan +s are observations:** Truth plus observational noise

**Green line is posterior mode (4DVAR analysis after 10 outer loops)**

# Summary

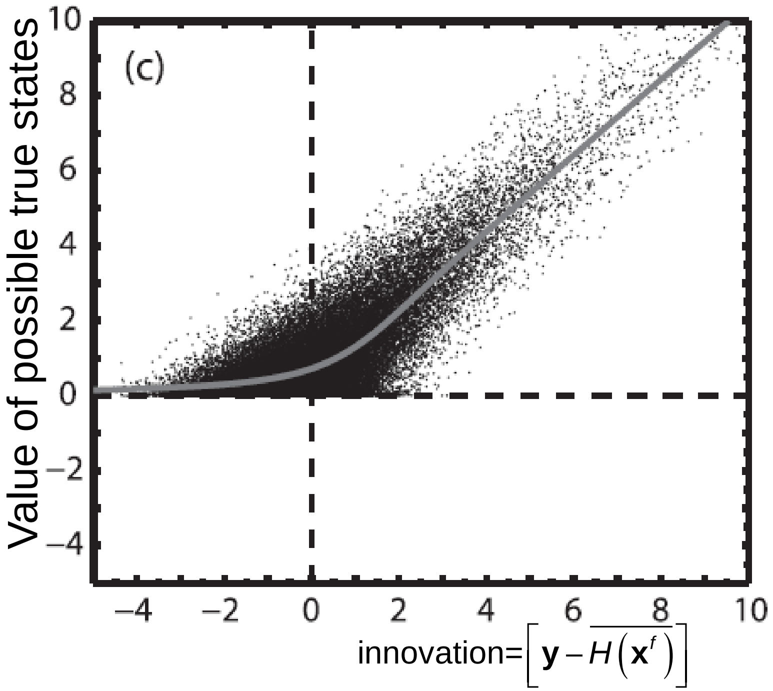
- An idealized coupled model with some qualitatively similar characteristics to the ocean-atmosphere system has been developed.
- Coupled EnKF outperformed uncoupled EnKF.
- Existing ocean (atmosphere) DA schemes can be extended to assimilate near interface obs (Frolov and Bishop, 2015). With simple model, interface solver EnKF performed at least as well as coupled EnKF.
- Theory for deriving accurate Local Ensemble Tangent Linear Models (LETLM) and their adjoints has been given.
- Implementation of LETLM in simple models enabled
    i. accurate prediction of differences between perturbed and non-perturbed non-linear trajectories
    ii. computation of gradient of forecast aspect with respect to analysis variables
    iii. 4DVAR

# Further Comment

- Depending on performance in large systems, LETLMs/Adjoints could replace existing TL/Adjoints OR just be used for those components of the coupled system for which no TLs exist.

# Data Assimilation 101

- 4DVAR/3DVAR with outer loop gives maxima (mode) of posterior pdf – ensemble is optional.
- EnKF/EnKS/EnOI and/or 4DVAR-without-outer-loop make ***linear*** estimates of posterior mean. EnKF/EnKS generate a posterior ensemble whose 2nd moment is independent of the value of the obs.




Hodyss (2011, MWR)

# Coupled DA Background

- Atmospheric DA currently uses DA methods that combine elements of 4DVAR and the EnKS. The DA window is 0.5 – 12 hrs
- Ocean DA schemes similar, but 3DVAR more common than 4DVAR and proxy ensembles more common than flow-dependent ensembles. DA window is 24-240+ hrs.
- Uncoupled DA: Only uncoupled models used in DA
- Weakly Coupled DA: Coupled model used for first guess but DA done separately in each fluid.
- Strongly Coupled DA: Atmospheric (oceanic) obs used to update ocean (atmosphere).
- What should be done to maximize the utility of coupled model data assimilation?