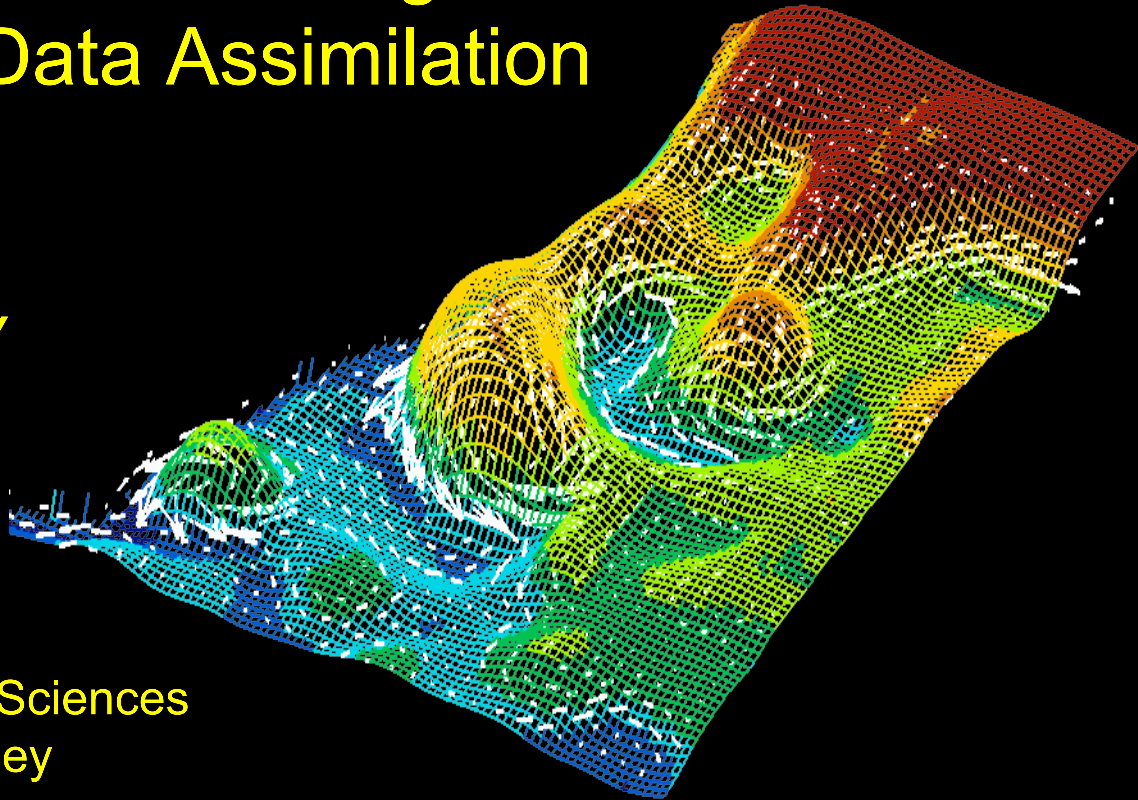
Predictability of Mesoscale Variability in the East Australian Current given Strong Constraint Data Assimilation



John Wilkin
Javier Zavala-Garay
and Hernan Arango

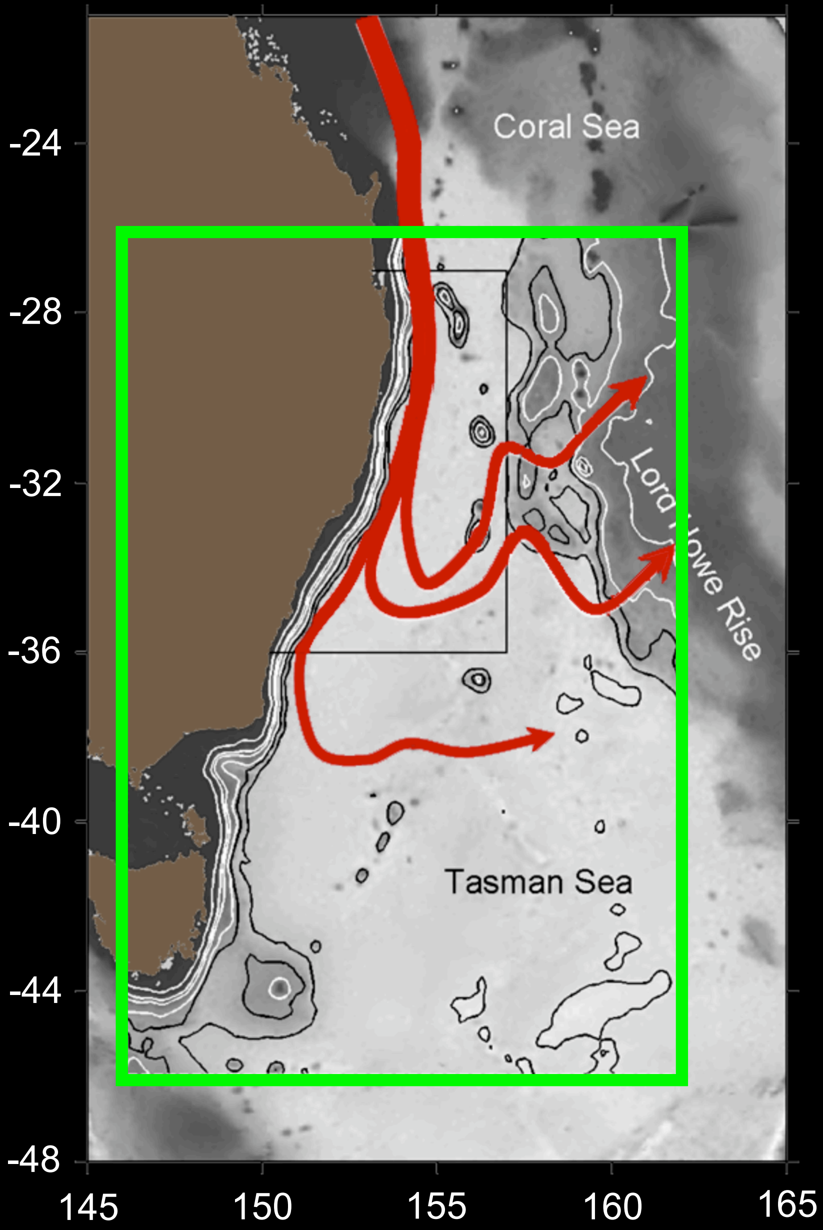
Institute of Marine and Coastal Sciences
Rutgers, University of New Jersey
New Brunswick, NJ, USA

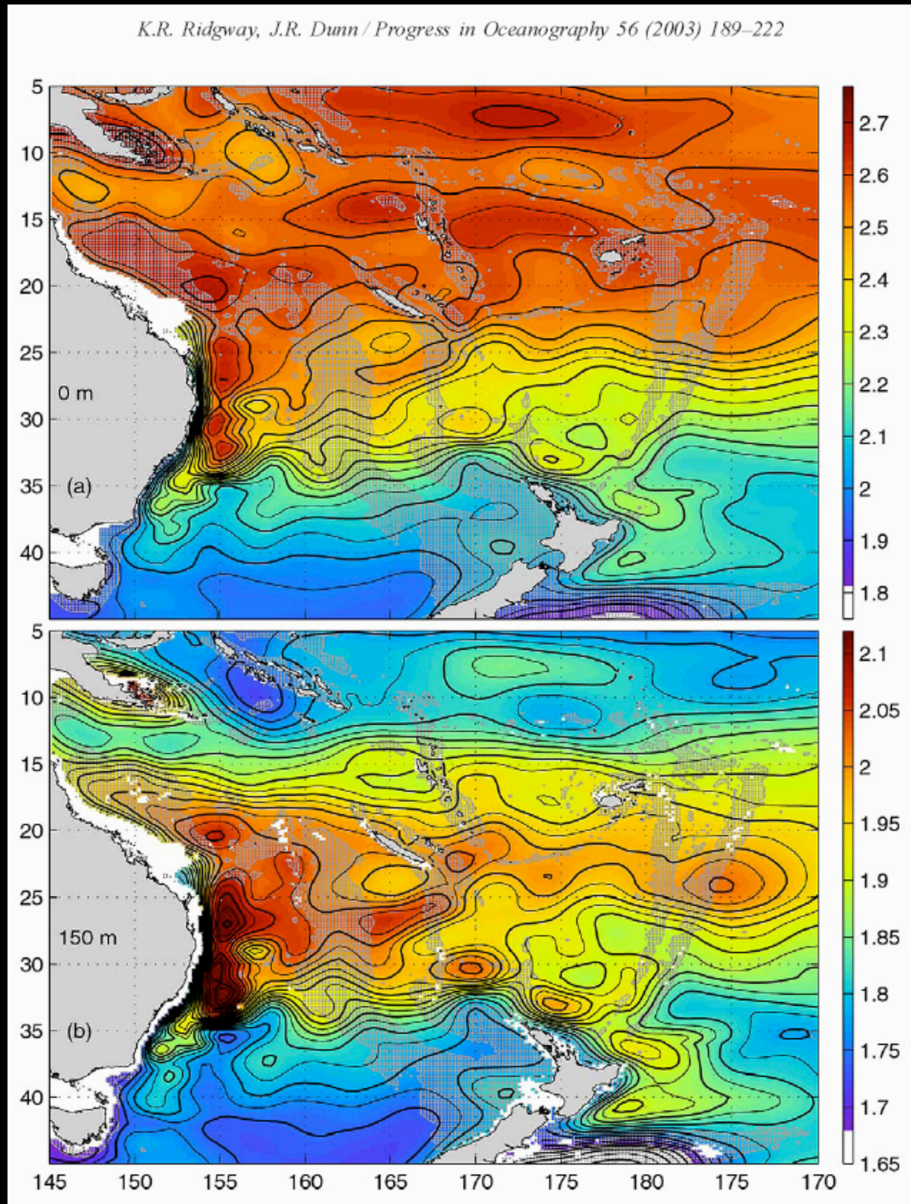
RUTGERS MARINE & COASTAL SCIENCES

jwilkin@rutgers.edu
http://marine.rutgers.edu

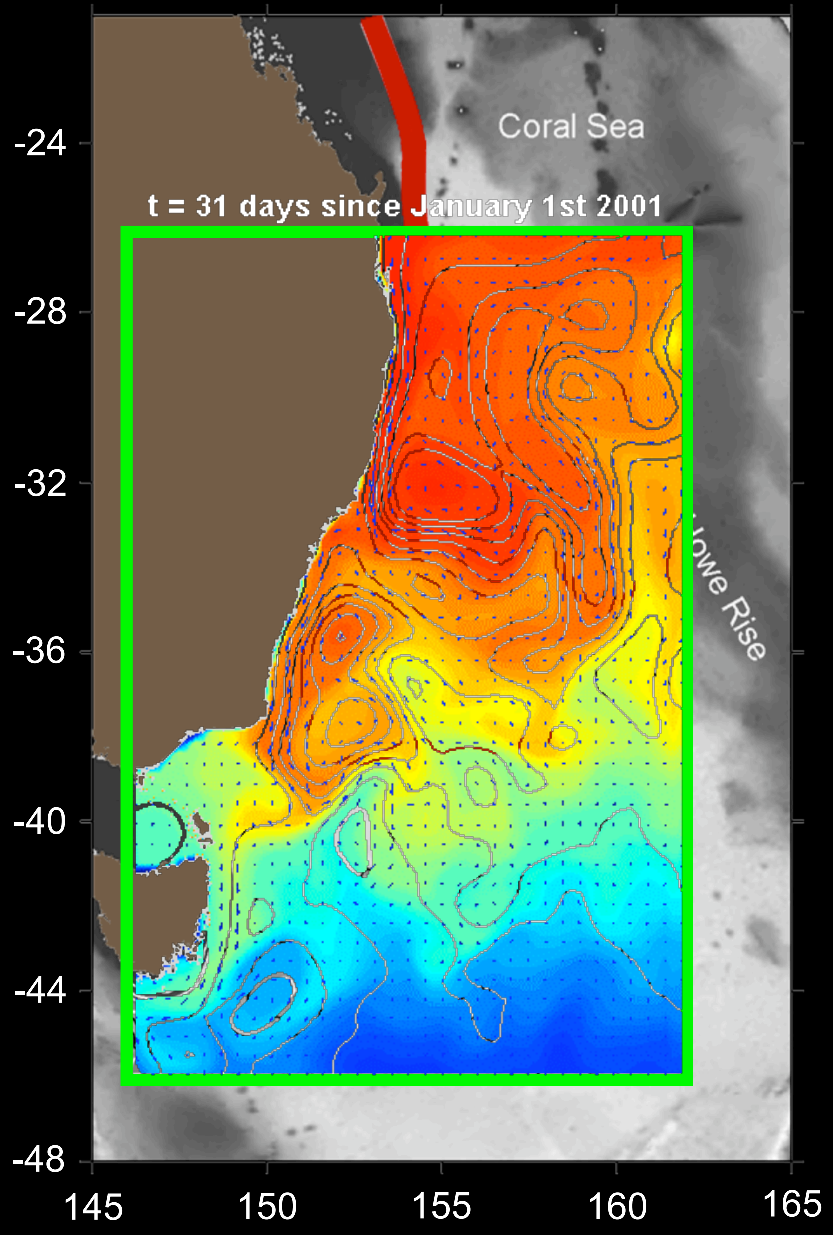
Ocean Surface Topography Mission/Jason-2
Surveying Earth's Oceans

East Australia Current ROMS Model for IS4DVAR





East Australia Current ROMS Model for IS4DVAR



ROMS* Model Configuration

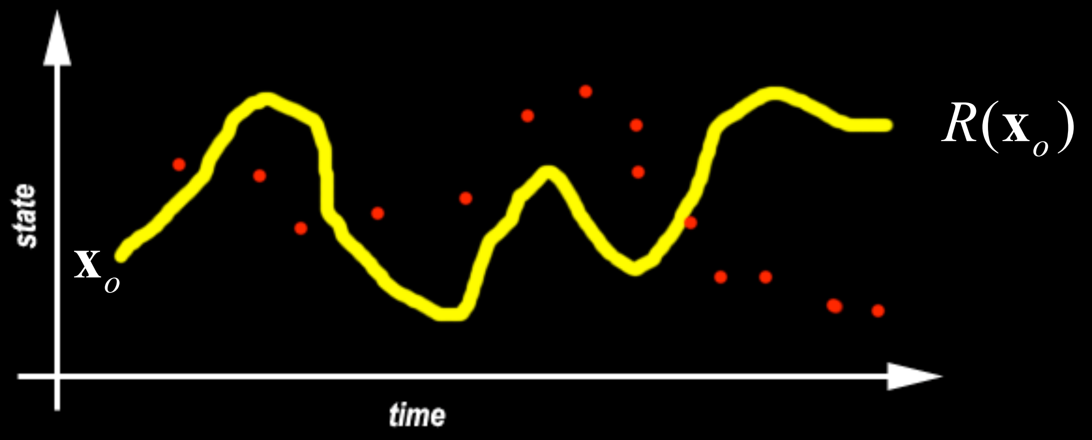
Resolution	0.25 x 0.25 degrees
Grid	64 x 80 x 30
Δx, Δy	~ 25 km
Δt	1080 sec
Bathymetry	16 to 4895 m
Open boundaries	Global NCOM (2001 and 2002)
Forcing	Global NOGAPS, daily
De-correlation scale	100 km, 150 m
N outer, N inner loops	10, 3

* http://myroms.org

1/8º resolution version simulates complex EOF “eddy” and “wave” modes of satellite SST and SSH in EAC separation:

Wilkin, J., and W. Zhang, 2006, Modes of mesoscale sea surface height and temperature variability in the East Australian Current J. Geophys. Res. 112, C01013, doi:10.1029/2006JC003590

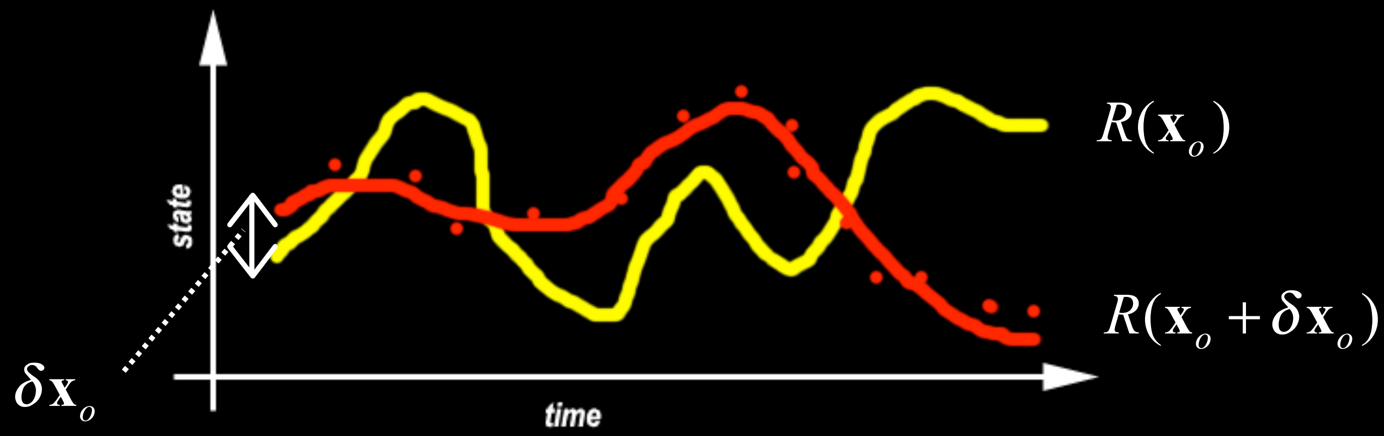
IS4DVAR*



- Given a first guess (the forward trajectory)…
- and given the available data…

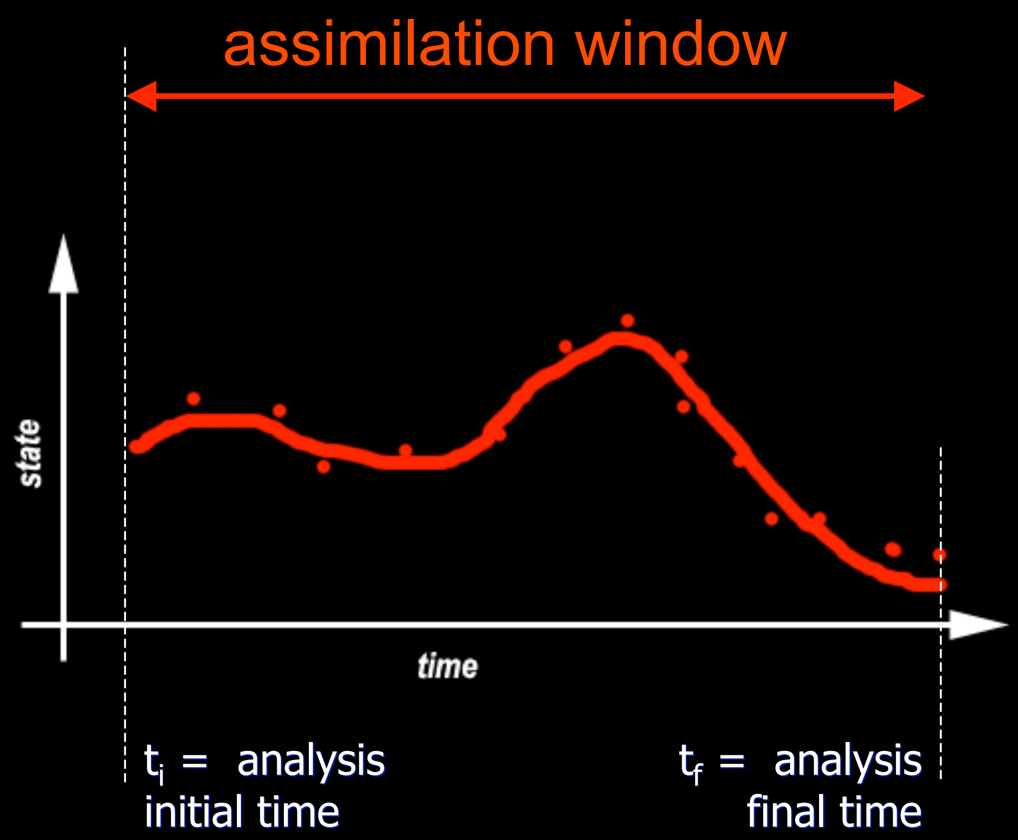
**Incremental Strong Constraint 4-Dimensional Variational data assimilation*

IS4DVAR



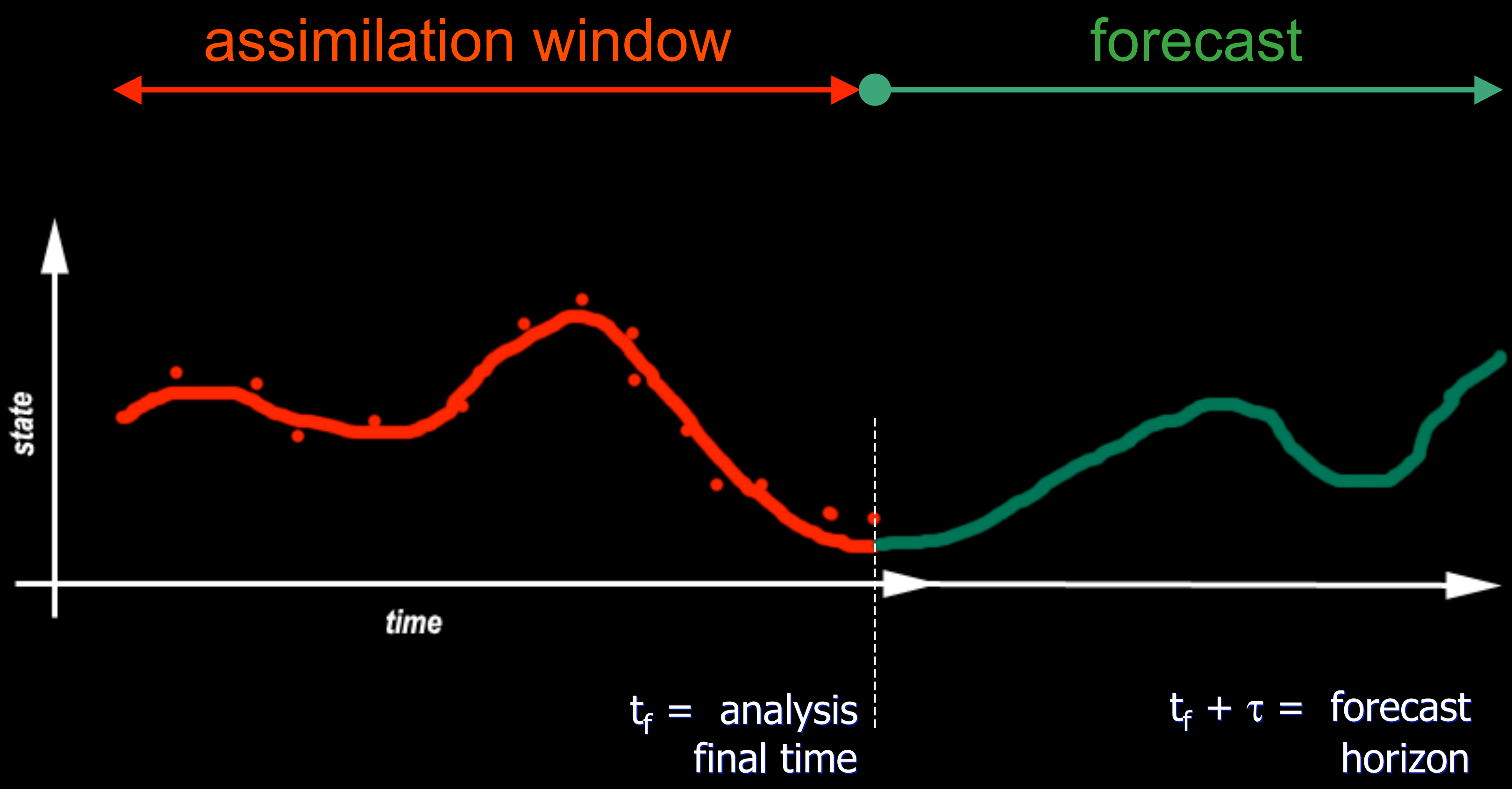
- Given a first guess (the forward trajectory)…
- and given the available data…
- what change (or increment) to the initial conditions (IC) produces a new forward trajectory that better fits the observations?

The best fit becomes the *analysis*

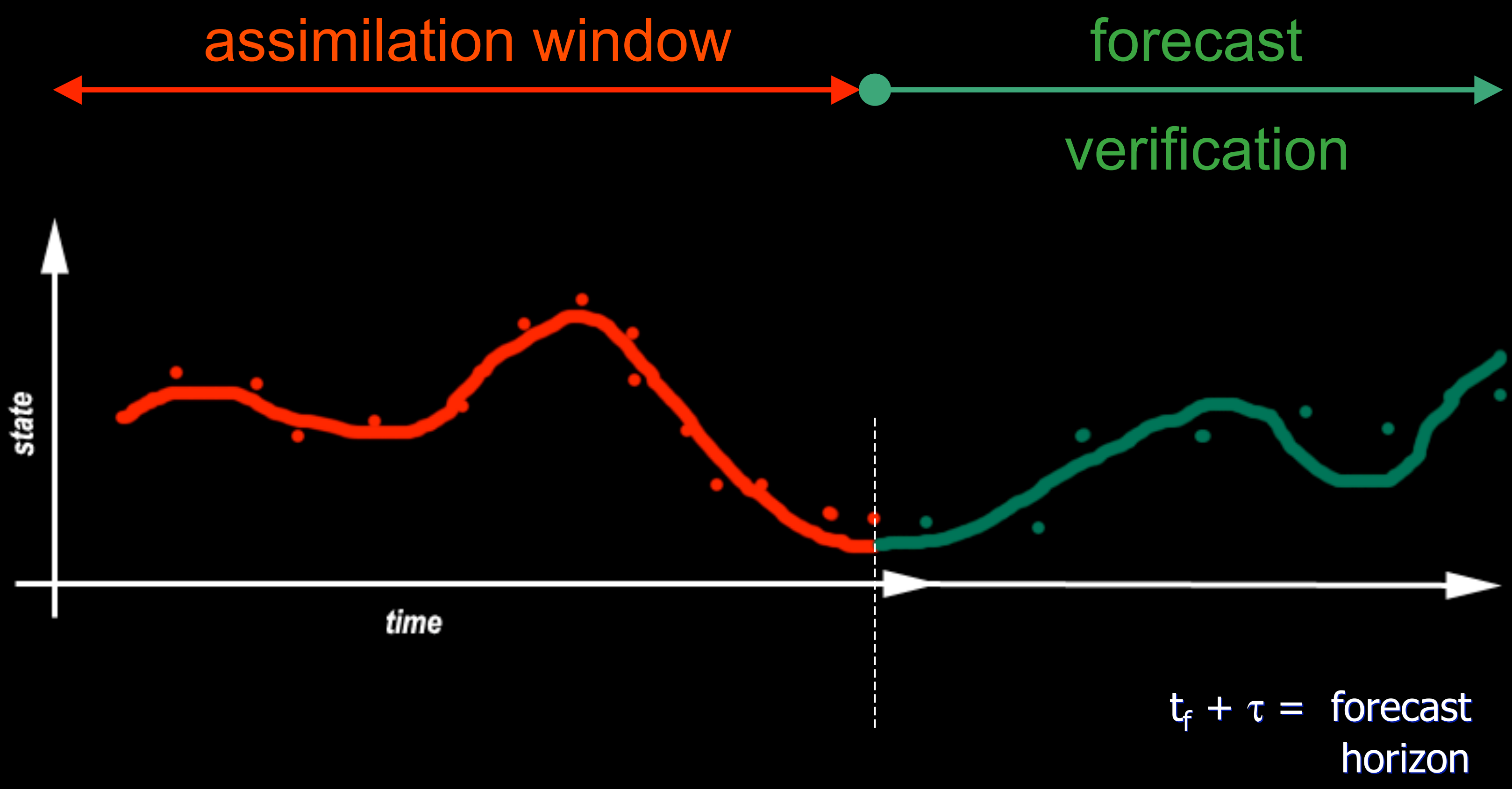


The *strong constraint* requires the trajectory satisfies the *physics* in ROMS.
The Adjoint enforces the consistency among state variables.

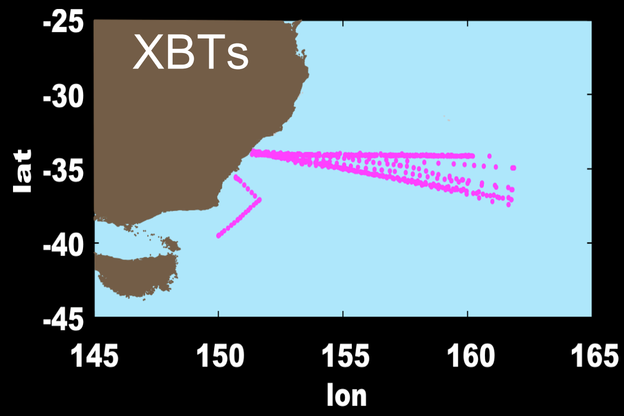
The final analysis state becomes the initial conditions for the *forecast* window



Forecast verification is with respect to data not yet assimilated



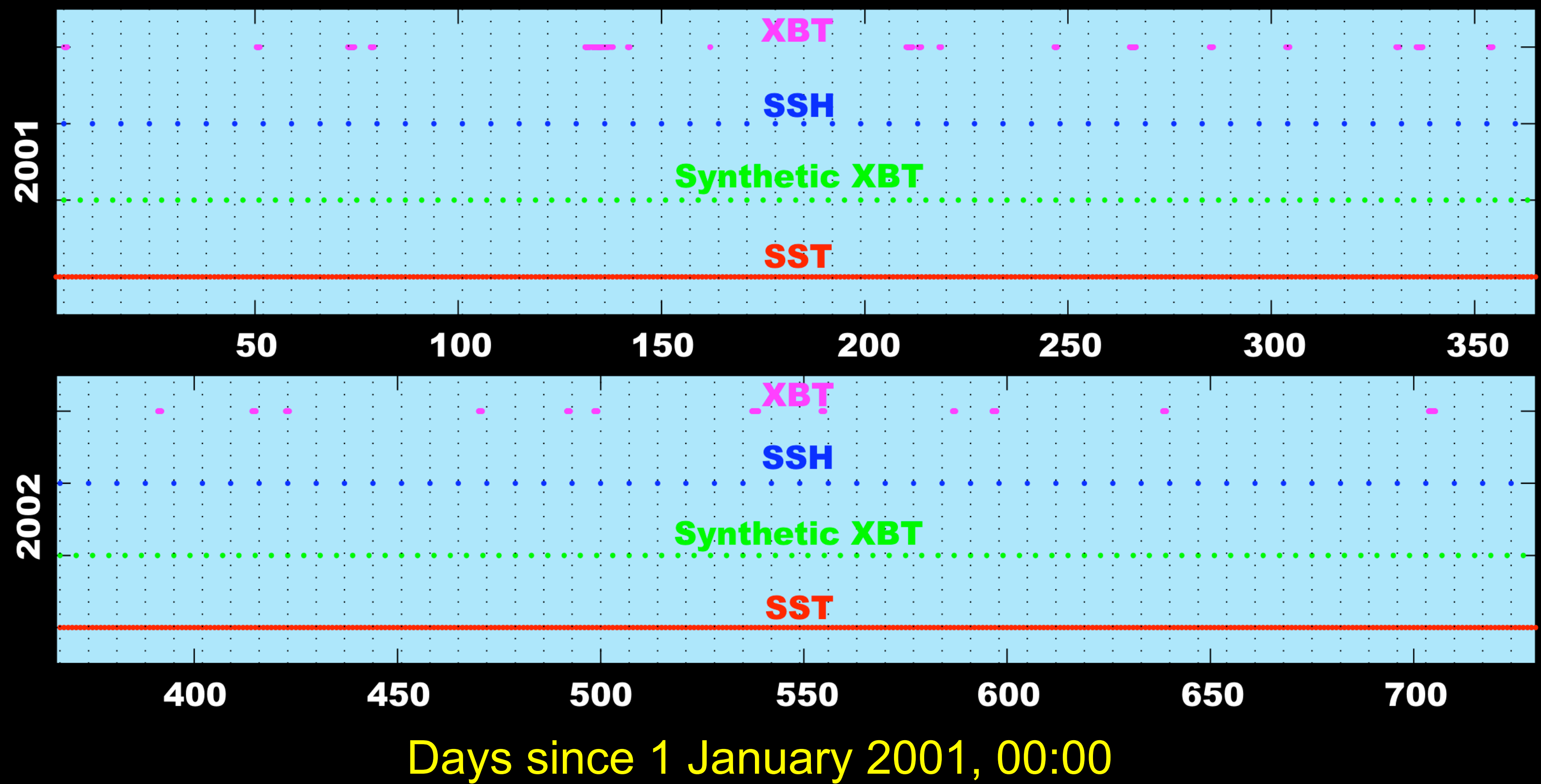
4DVar Observations and Experiments



<u>7-Day IS4DVAR Experiments</u>

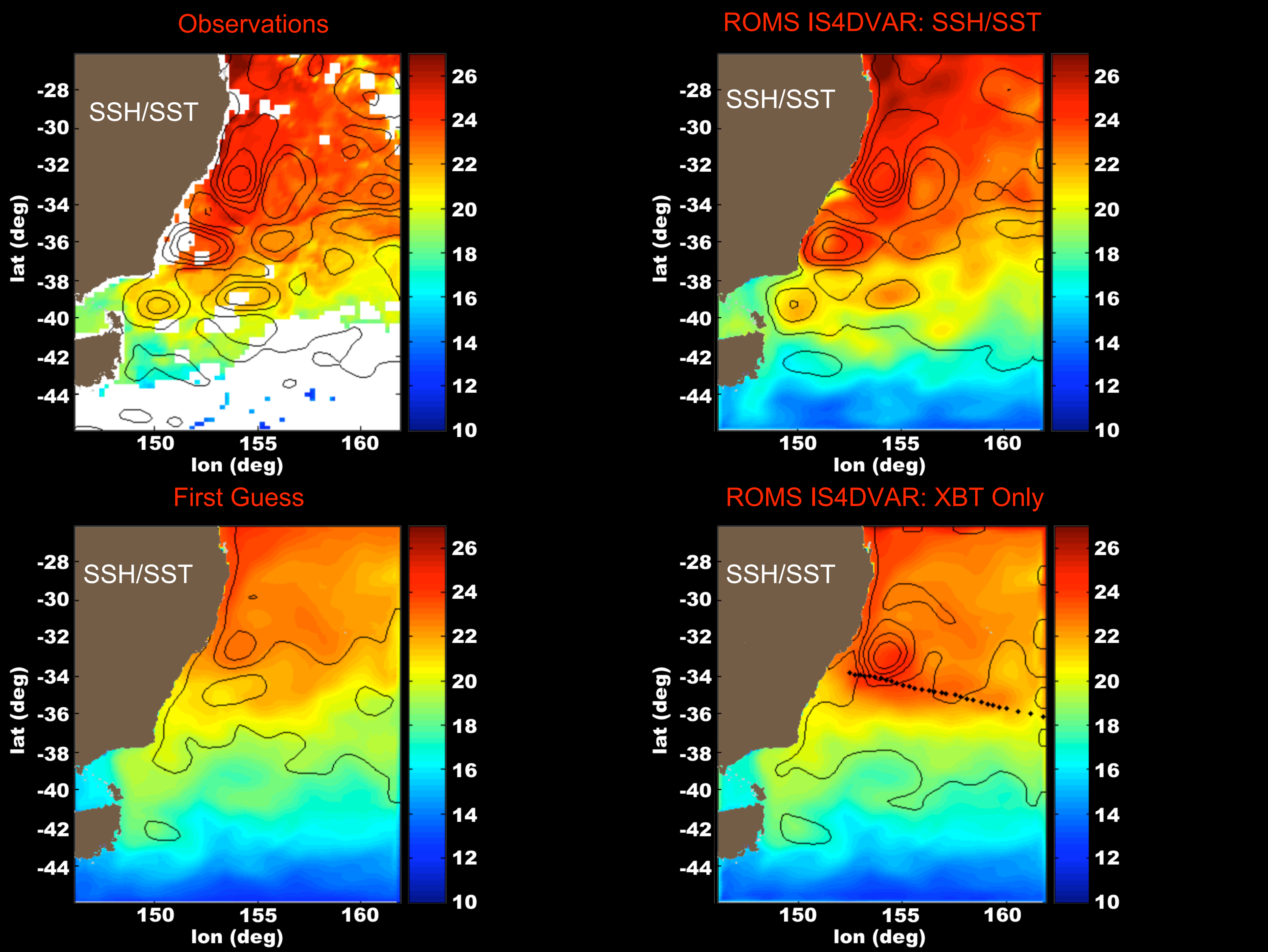
E1: SSH, SST
E2: SSH, SST, XBT

SSH 7-Day Averaged AVISO
SST Daily CSIRO HRPT



EAC IS4DVAR

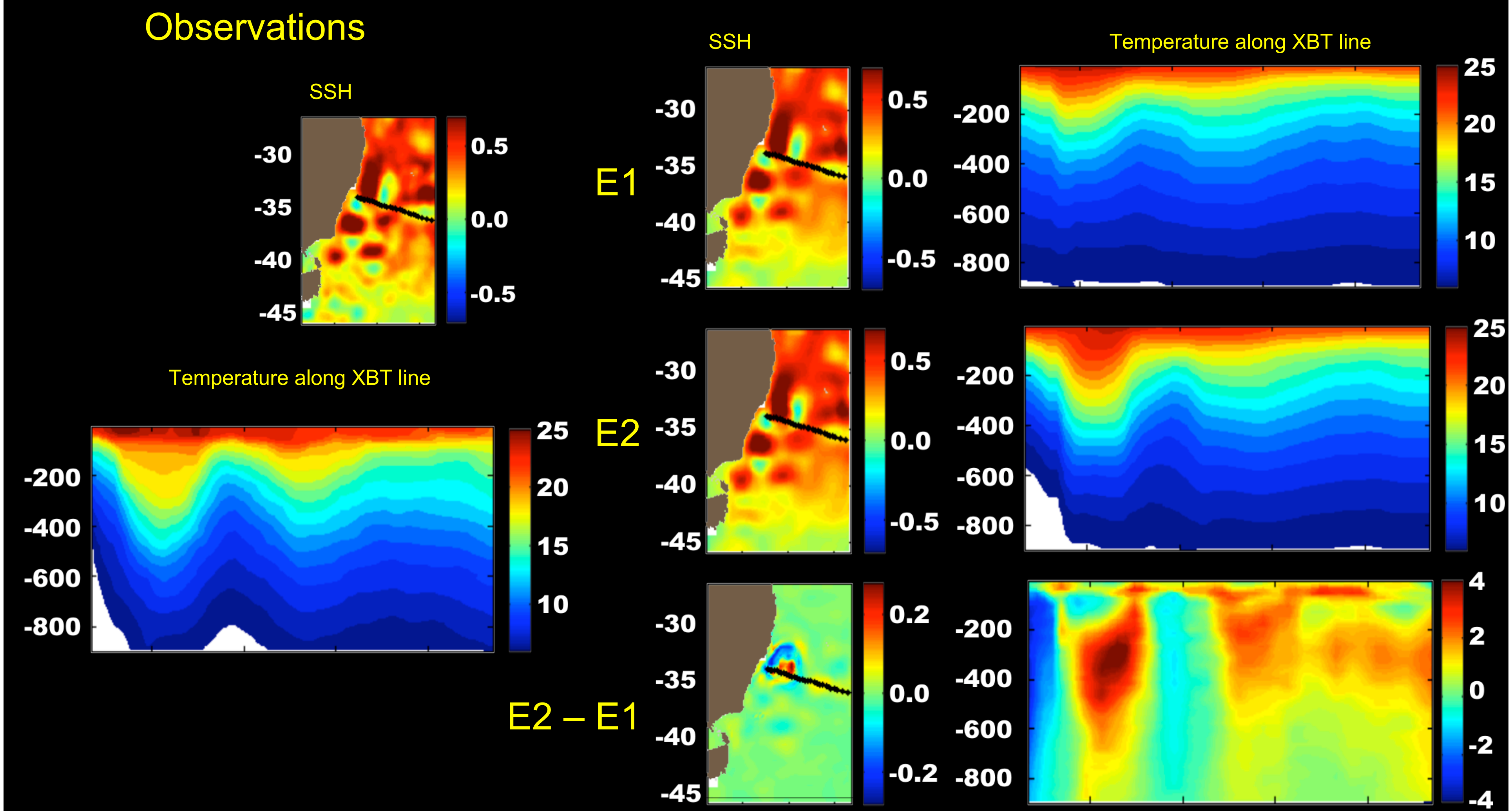
Assimilating surface vs. sub-surface observations



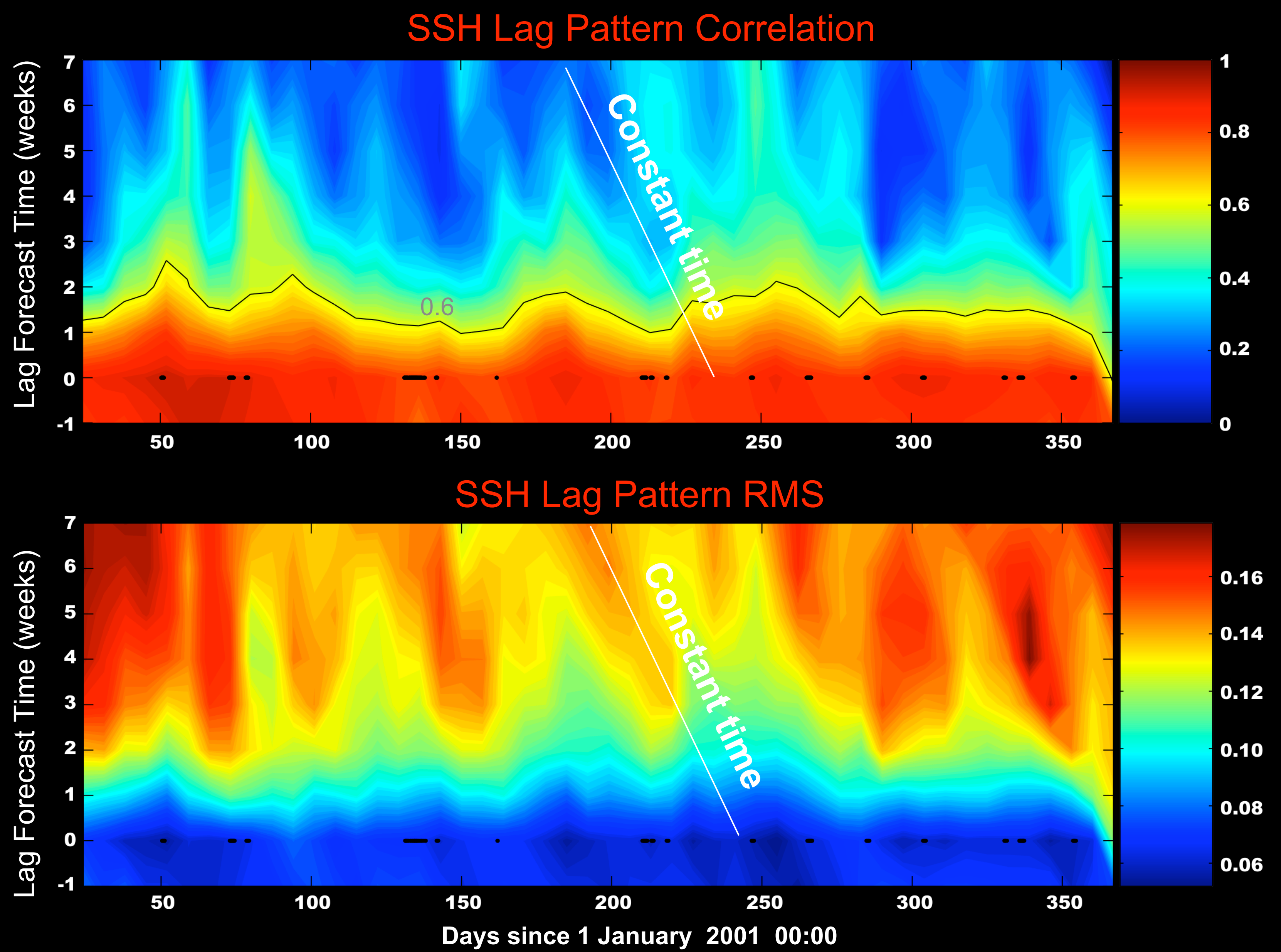
EAC IS4DVAR

7-Day 4DVar Assimilation cycle

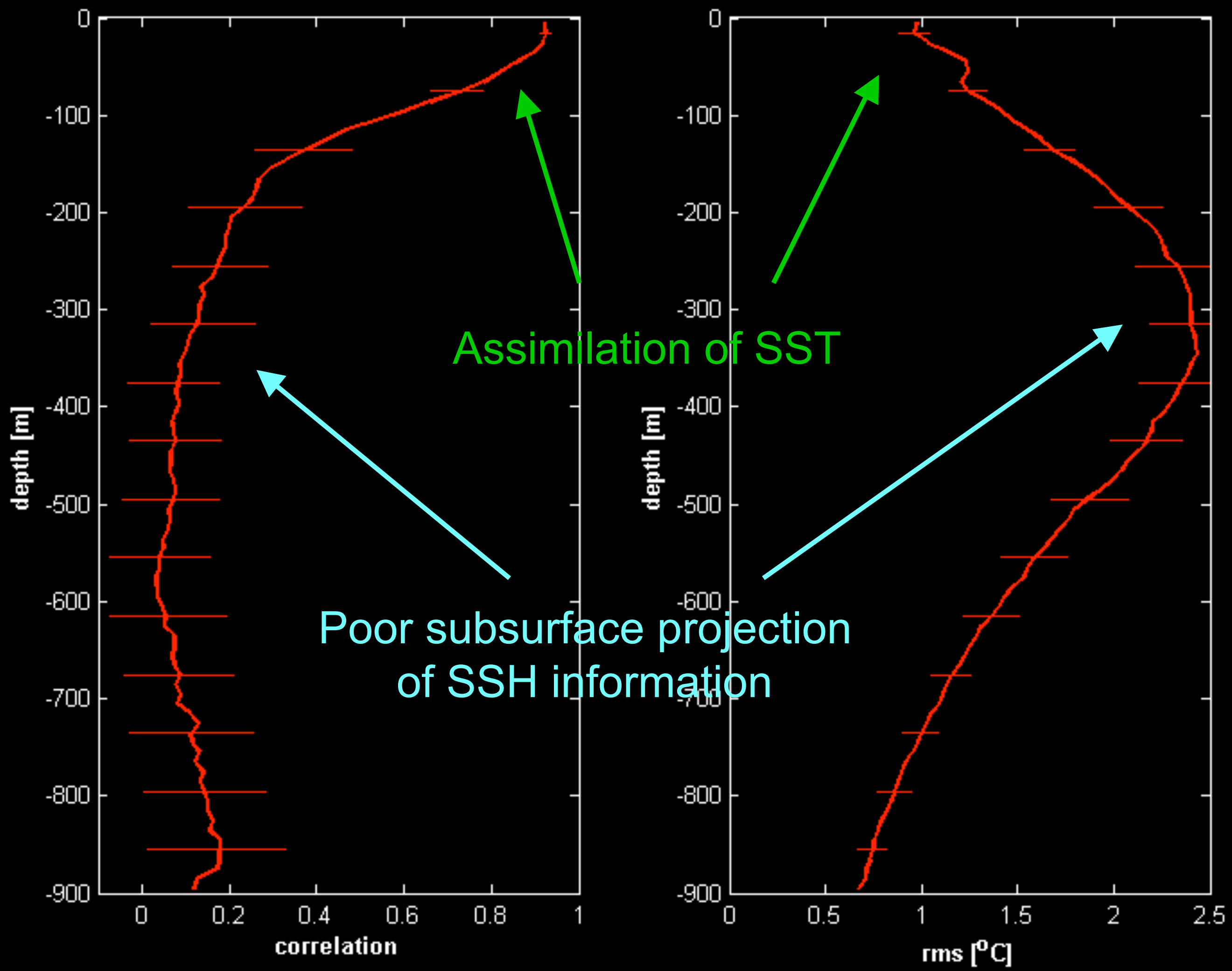
E1: SSH, SST Observations
E2: SSH, SST, XBT Observations



Forecast SSH correlation and RMS error: Experiment E2



Comparison between ROMS temperature *analysis* (*fit*) and *withheld observations* (all available XBTs); the XBT data were not assimilated – they are used here only to evaluate the quality of the reanalysis.



E1: SSH + SST

The subsurface projection of surface only satellite data is less skillful than we would like.

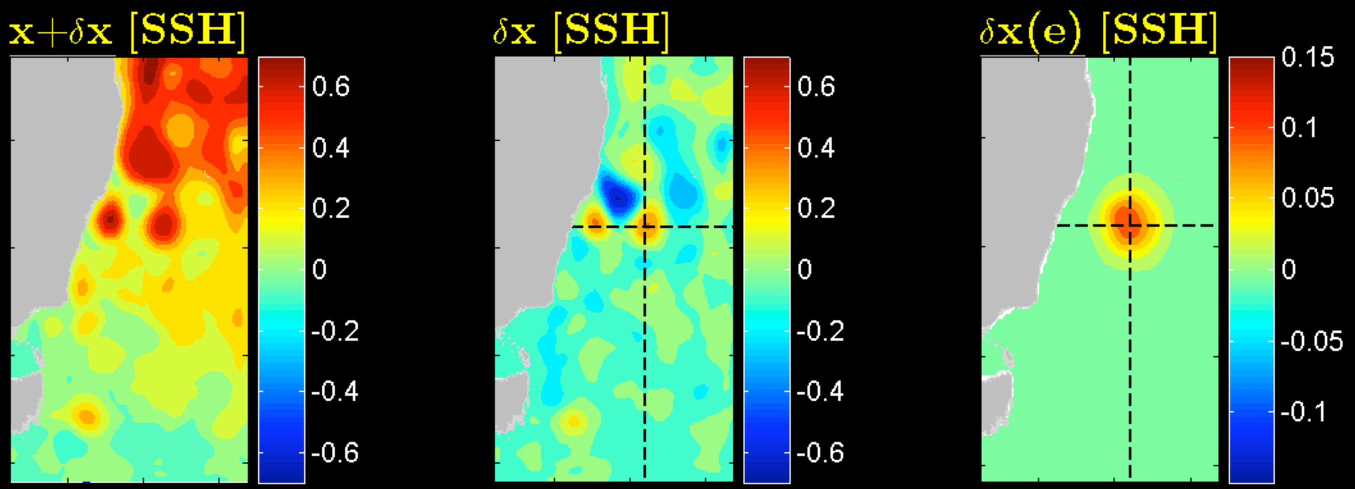
These errors also adversely affect the forecast.

Transferring information from one state variable to another, and projecting surface to subsurface

Three ways:

1) The Adjoint Model
2) Empirical statistical correlations to generate synthetic data
   - *Here, T(z) from SSH and SST*
3) Modeling of the background covariance matrix
   - *e.g. via the hydrostatic/geostrophic relation*

(1) Adjoint



For a single observation (e.g. SSH at one grid point) the increment is given by:

$$\delta\,\mathbf{x} = \mathbf{c}\;\mathbf{B}\;\mathbf{M}^T\,\mathbf{e}$$

Basic IS4DVAR procedure:

J = *model-data misfit*

$$J(x) = \underbrace{\tfrac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)}_{J_b} + \underbrace{\sum_{i=1}^{N_{obs}} \tfrac{1}{2}(\mathbf{H}_i\mathbf{x}_i-\mathbf{y}_i)^T \mathbf{O}^{-1}(\mathbf{H}_i\mathbf{x}_i-\mathbf{y}_i)}_{J_o}$$

The "best" simulation minimizes *L over an interval: t' = [0, τ]*

At extrema of $\int_0^\tau L\,dt'$ we require:

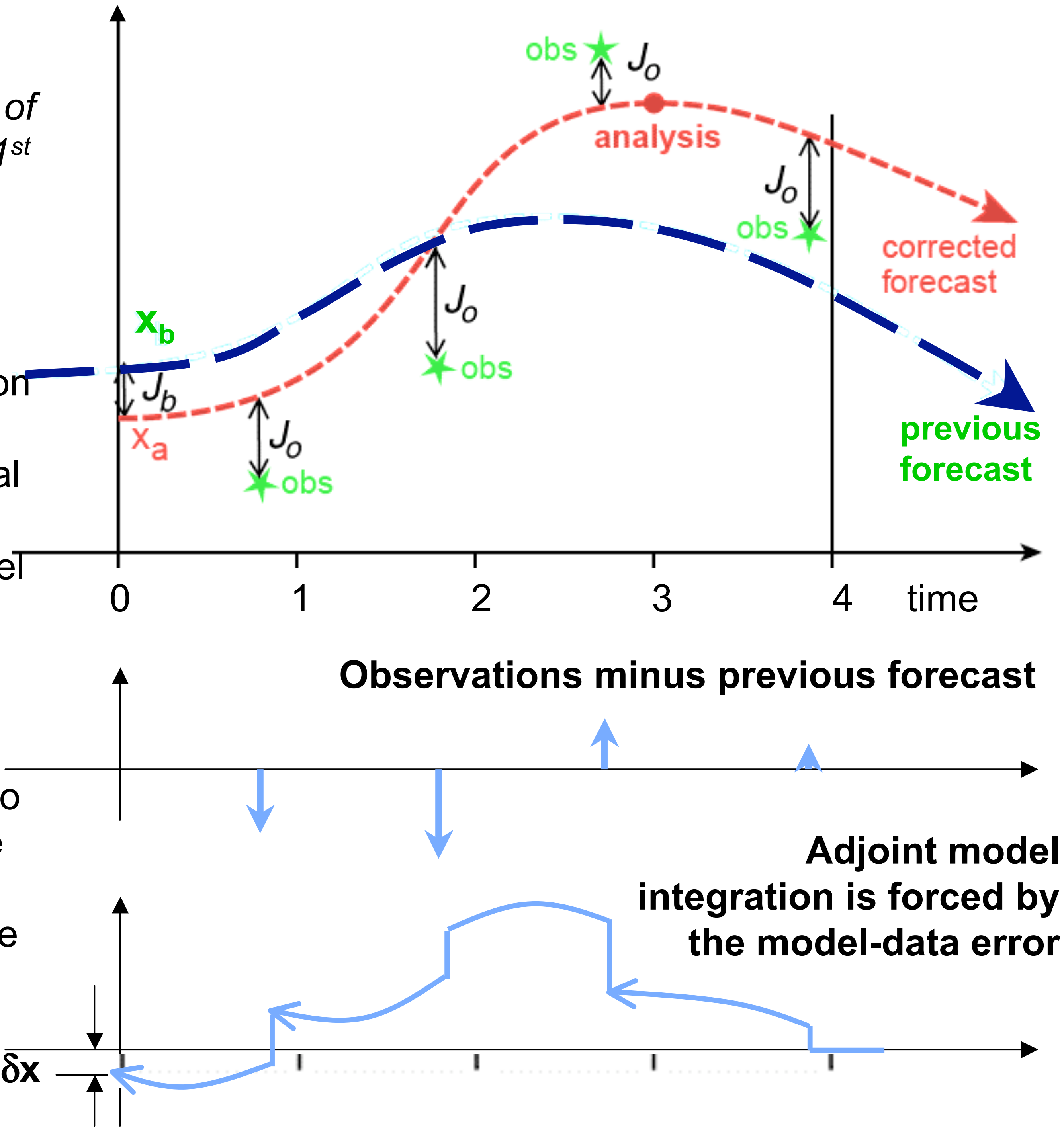
$$L = J(\mathbf{x}) + \sum_{i=1}^{N} \boldsymbol{\lambda}_i^T \left(\frac{d\mathbf{x}_i}{dt} - \mathbf{N}(\mathbf{x}_i) - \mathbf{F}_i \right)$$

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{\lambda}_i} = 0 \quad \Rightarrow \dfrac{d\mathbf{x}_i}{dt} - \mathbf{N}(\mathbf{x}_i) - \mathbf{F}_i = 0 & \textit{NLROMS} \\ \dfrac{\partial L}{\partial \mathbf{x}_i} = 0 \quad \Rightarrow -\dfrac{d\boldsymbol{\lambda}_i}{dt} - \left(\dfrac{\partial \mathbf{N}}{\partial \mathbf{x}}\right)^T \boldsymbol{\lambda}_i = \delta_{im} \mathbf{H}^T \mathbf{O}^{-1} (\mathbf{H}\mathbf{x}_m - \mathbf{y}_m) & \textit{ADROMS} \\ \dfrac{\partial L}{\partial \mathbf{x}(0)} = 0 \Rightarrow \mathbf{B}^{-1}(\mathbf{x}(0) - \mathbf{x}_b) = \boldsymbol{\lambda}(0) & \textit{coupling of NL \& AD} \\ \dfrac{\partial L}{\partial \mathbf{x}(\tau)} = 0 \Rightarrow \boldsymbol{\lambda}(\tau) = 0 & \textit{i.c. of ADROMS} \end{cases}$$

x_b = *model state (background) at end of previous cycle, and 1st guess for the next forecast*

In 4DVAR assimilation the adjoint gives the sensitivity of the initial conditions to mis-match between model and data

A descent algorithm uses this sensitivity to iteratively update the initial conditions, x_a, (analysis) to minimize $J_b + J_o$



Basic IS4DVAR procedure:

$$J(x) = \underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)}_{J_b} + \underbrace{\sum_{i=1}^{N}\frac{1}{2}(\mathbf{H}_i\mathbf{x}_i-\mathbf{y}_i)^T \mathbf{O}^{-1}(\mathbf{H}_i\mathbf{x}_i-\mathbf{y}_i)}_{J_o}$$

= model-data misfit

Outer-loop (10)

(1) Choose an $\mathbf{x}(0) = \mathbf{x}_b(0)$

(2) Integrate NLROMS $t \in [0,\tau]$ and save $\mathbf{x}(t)$

Inner-loop (3)

(a) Choose a $\delta\mathbf{x}(0)$

(b) Integrate TLROMS $t \in [0,\tau]$ and compute ***J***

(c) Integrate ADROMS $t \in [\tau,0]$ to yield $\dfrac{\partial J_o}{\partial \delta\mathbf{x}(0)} = \lambda(0)$

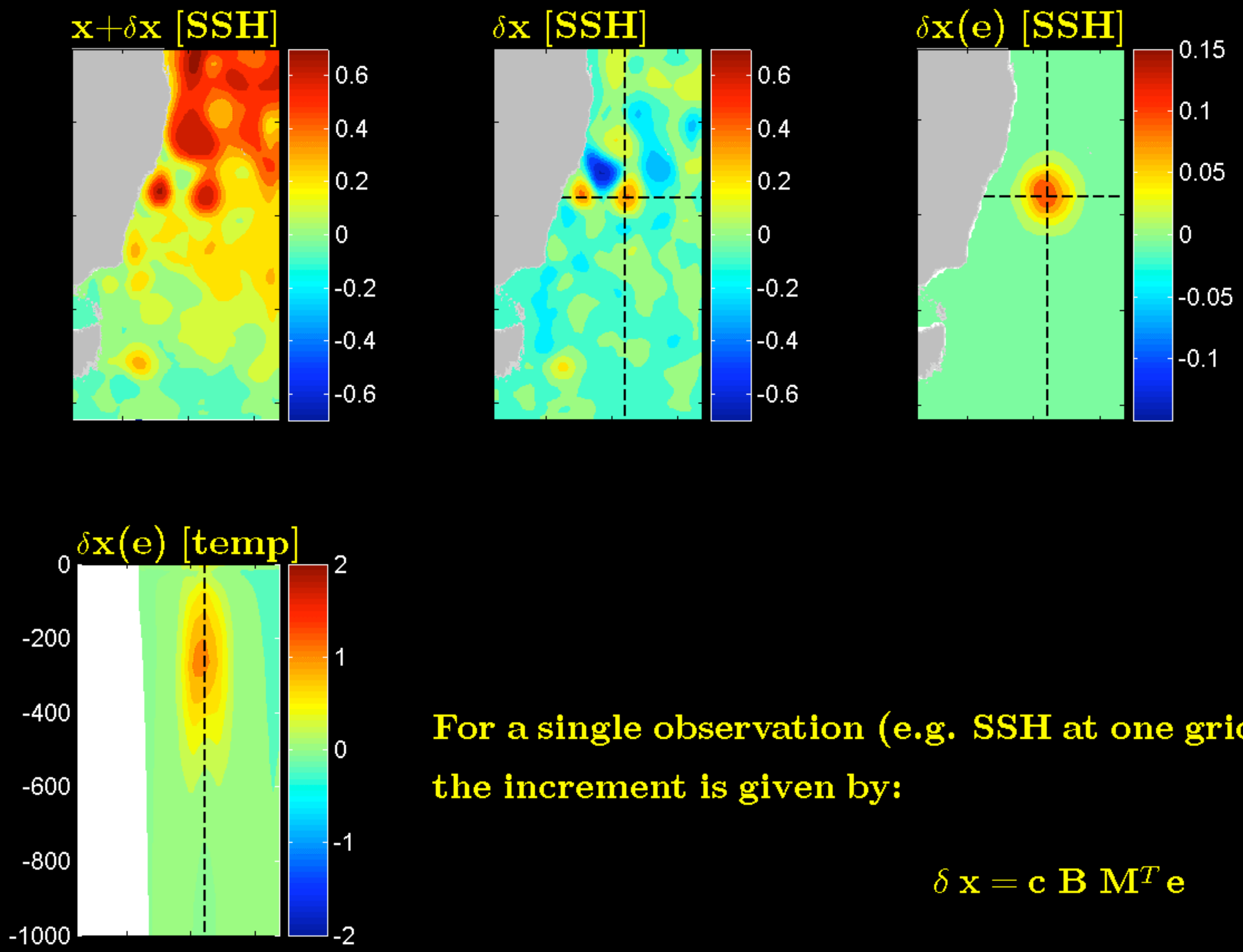
(d) Compute $\dfrac{\partial J}{\partial \delta\mathbf{x}(0)} = \mathbf{B}^{-1}\delta\mathbf{x}(0) - \lambda(0)$

(e) Use a descent algorithm to determine a "down gradient" correction to $\delta\mathbf{x}(0)$ that will yield a smaller value of ***J***

(f) Back to (b) until converged

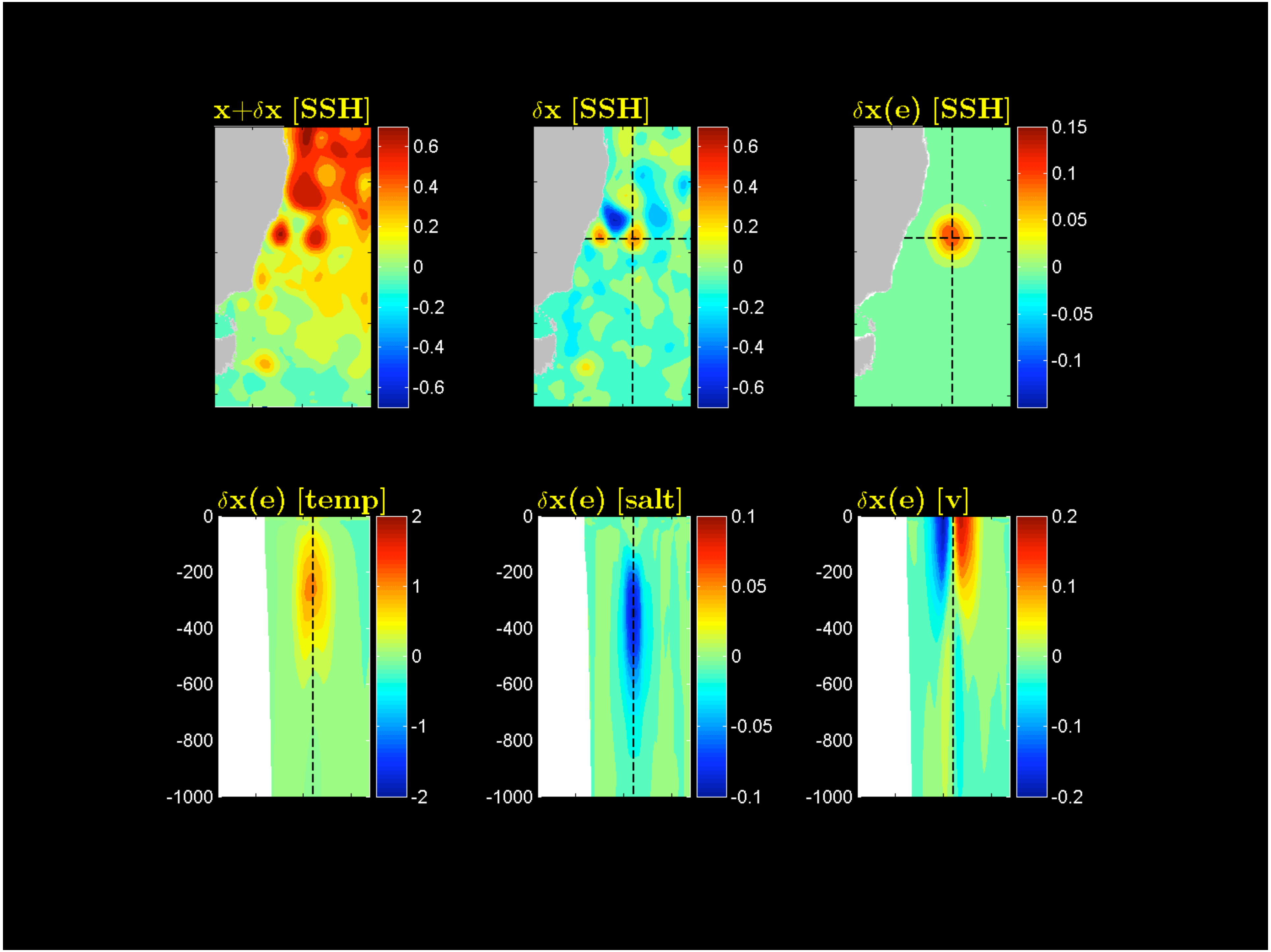
(3) Compute new $\mathbf{x}(0) = \mathbf{x}(0) + \delta\mathbf{x}(0)$ and back to (2) until converged

NLROMS = Non-linear forward model; TLROMS = Tangent linear; ADROMS = Adjoint



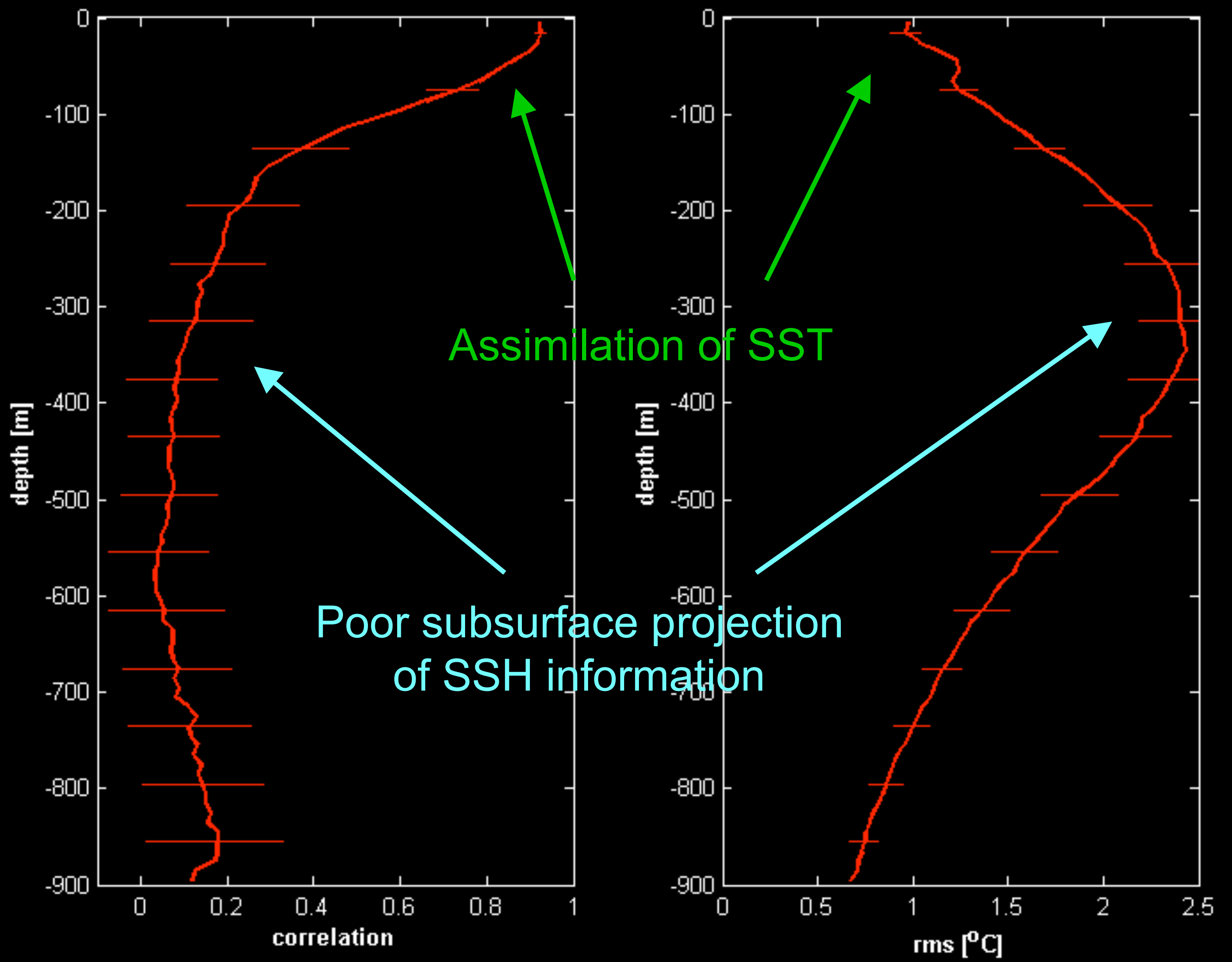
For a single observation (e.g. SSH at one grid point) the increment is given by:

$$\delta\,\mathbf{x} = \mathbf{c}\ \mathbf{B}\ \mathbf{M}^{T}\,\mathbf{e}$$



x+δx [SSH]
δx [SSH]
δx(e) [SSH]
δx(e) [temp]
δx(e) [salt]
δx(e) [v]

Comparison between ROMS temperature *analysis* (*fit*) and *withheld observations* (all available XBTs); the XBT data were not assimilated – they are used here only to evaluate the quality of the reanalysis.



E1: SSH + SST

The subsurface projection of surface only satellite data is less skillful than we would like.

These errors also adversely affect the forecast.

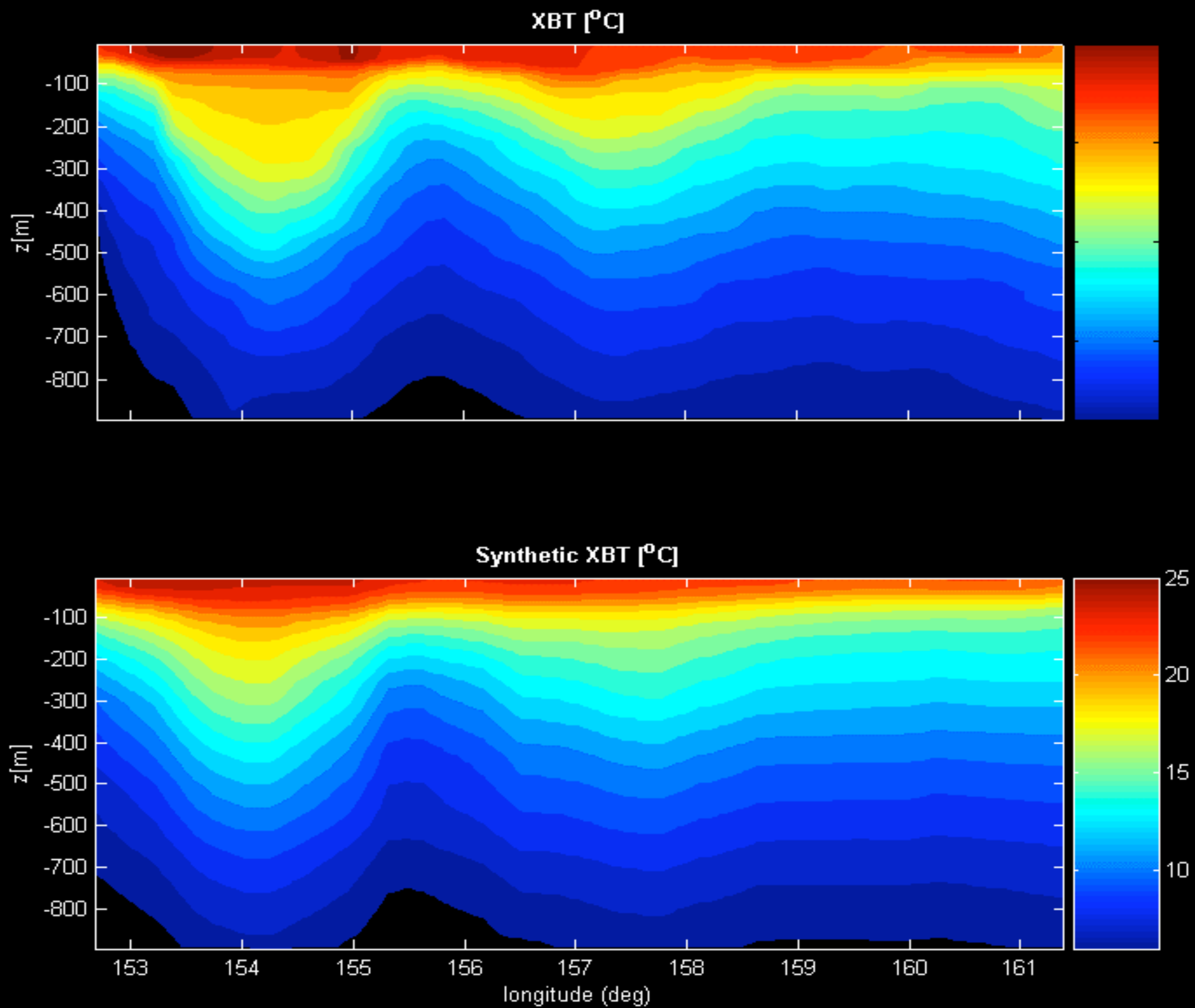
(2) Synthetic XBT/CTD

A statistical subsurface projection using regression of SSH and SST on EOFs of historical [*dynhgt*, *T(z), S(z)*] observed profiles

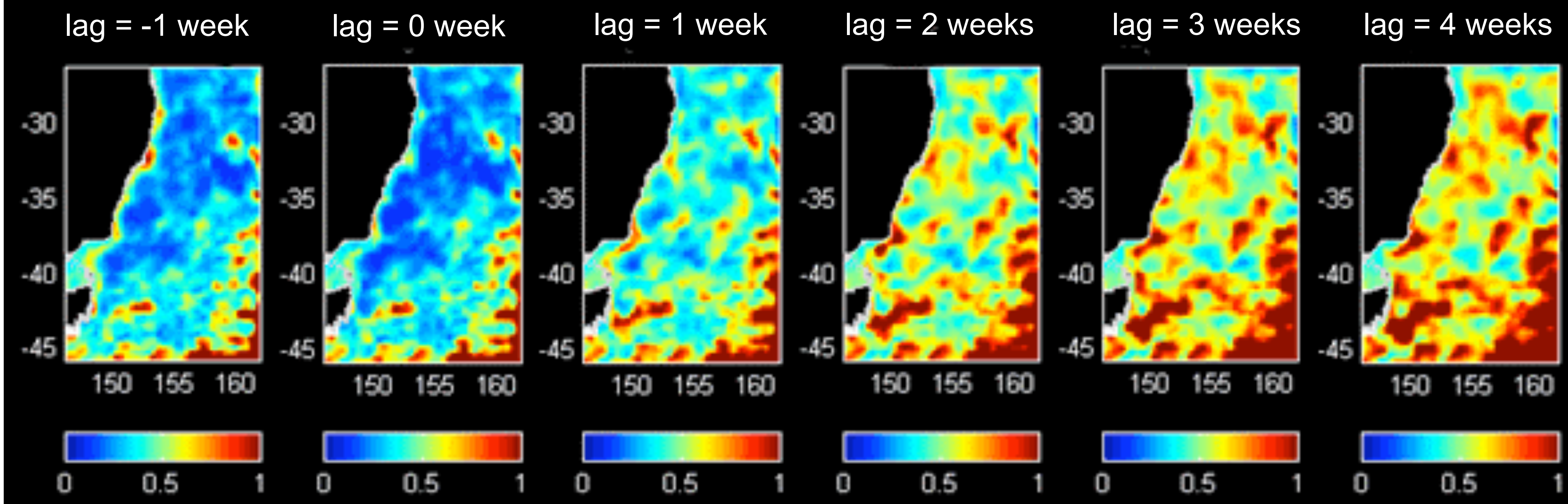
Example:

VOS XBT transect

Syn-XBT analysis



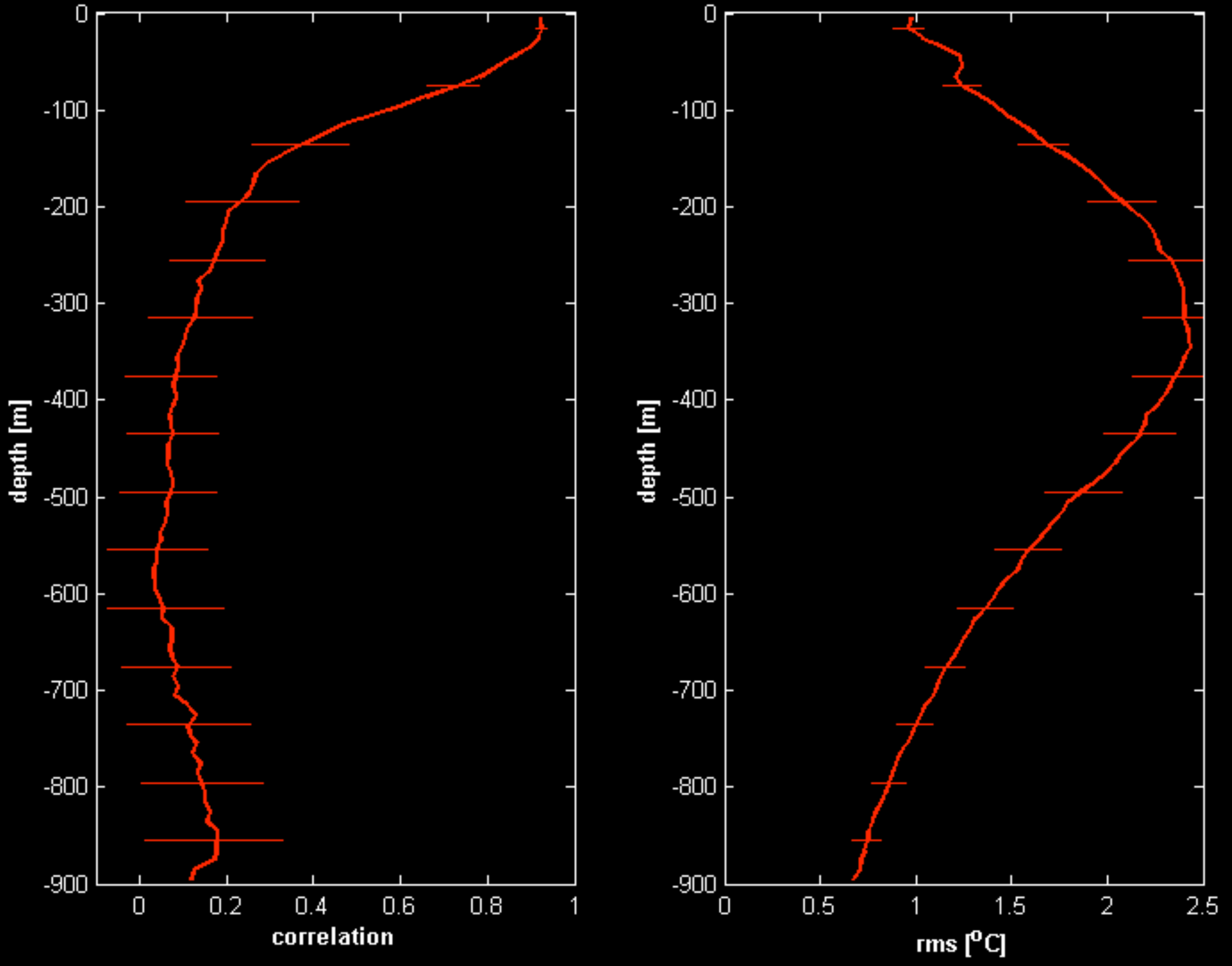
RMS error normalized by the expected variance in SSH



Forecast RMS error:

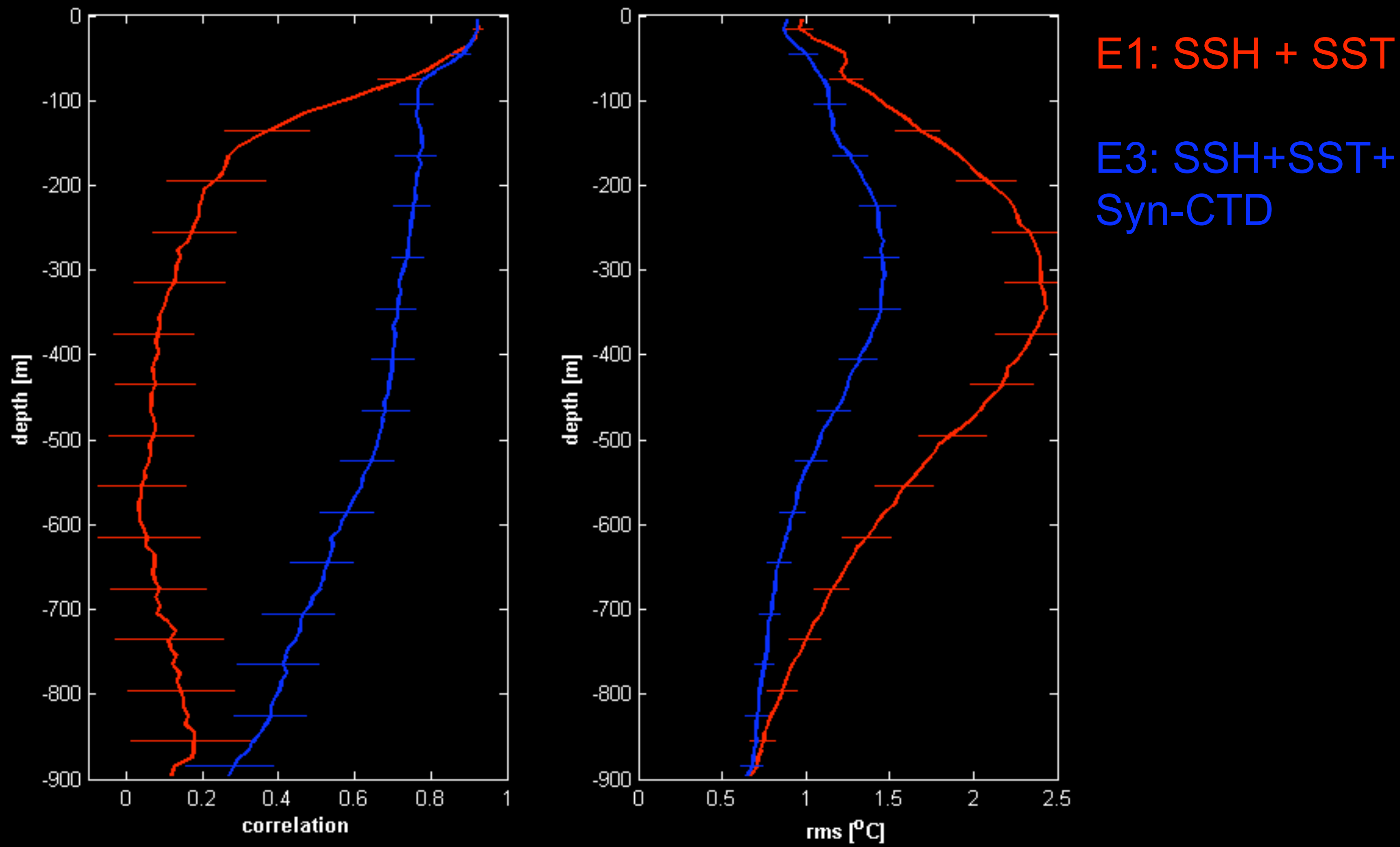
- typically < 0.5 out to 2 weeks forecast
- grows fastest at the open boundaries

Comparison between ROMS temperature *analysis* (*fit*) and *withheld observations* (all available XBTs); the XBT data were not assimilated – they are used here only to evaluate the quality of the reanalysis.



E1: SSH + SST

Comparison between ROMS temperature *analysis* (*fit*) and *withheld observations* (all available XBTs); the XBT data were not assimilated – they are used here only to evaluate the quality of the reanalysis.



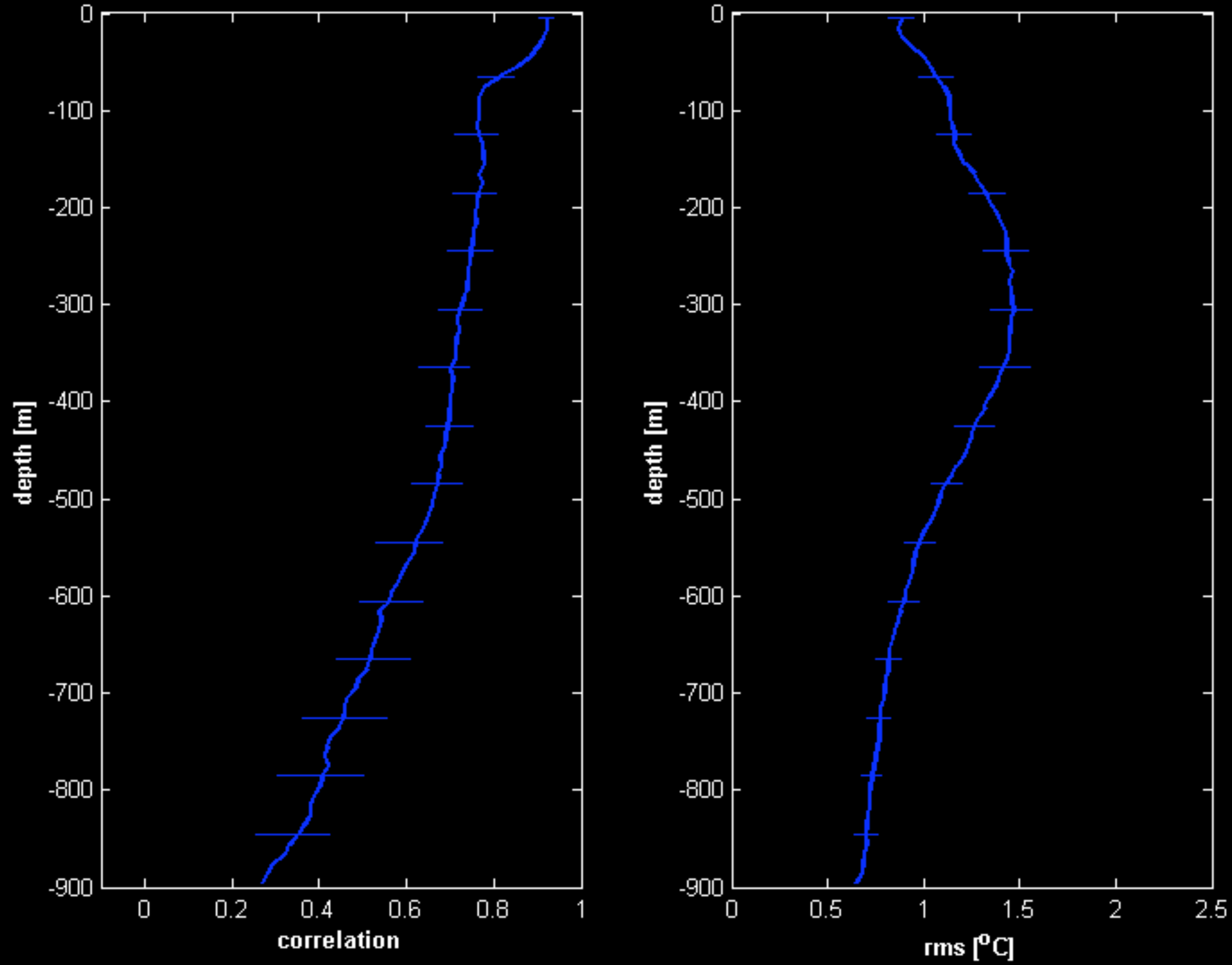
Comparison between ROMS subsurface temperature *predictions* and all *XBT observations* in 2001-2002

E3: SSH+SST+ Syn-CTD

0 lag – analysis skill

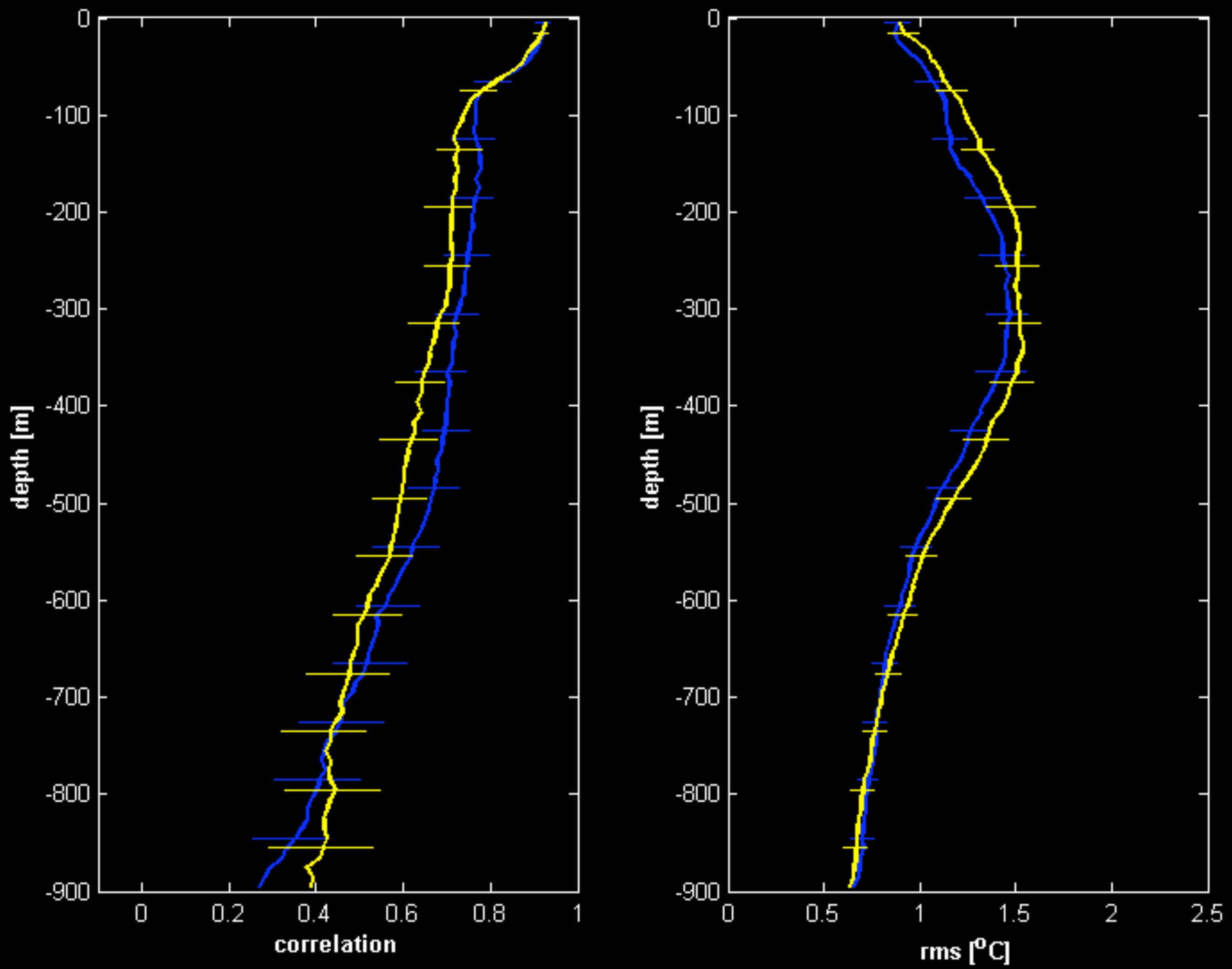
correlation

RMS error (^{o}C)



Comparison between ROMS subsurface temperature predictions and all XBT observations in 2001-2002

correlation RMS error (°C)



E3: SSH+SST+ Syn-CTD

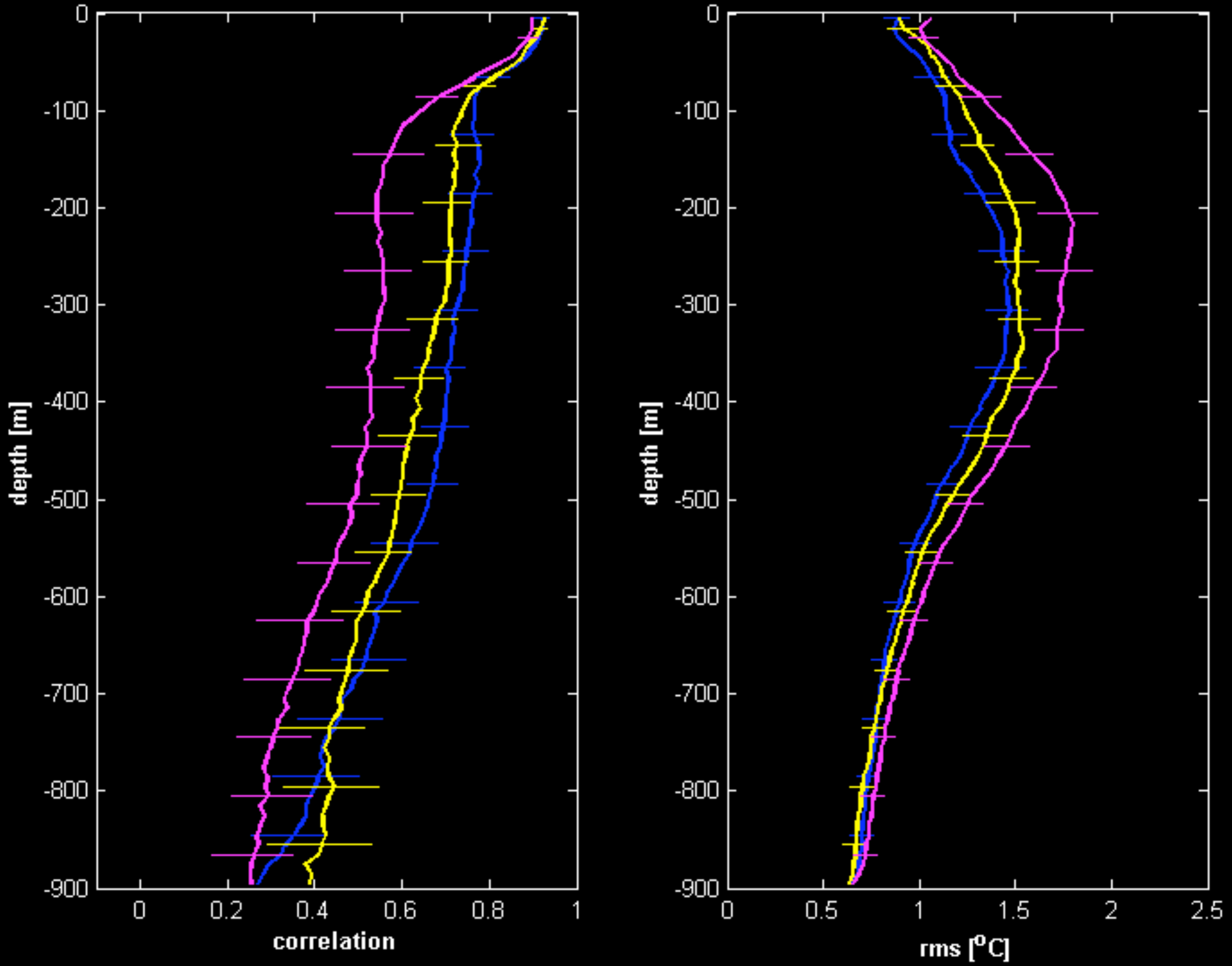
0 lag – analysis skill

1 week lag – little loss of skill

Comparison between ROMS subsurface temperature *predictions* and all *XBT observations* in 2001-2002

E3: SSH+SST+ Syn-CTD

correlation RMS error (°C)



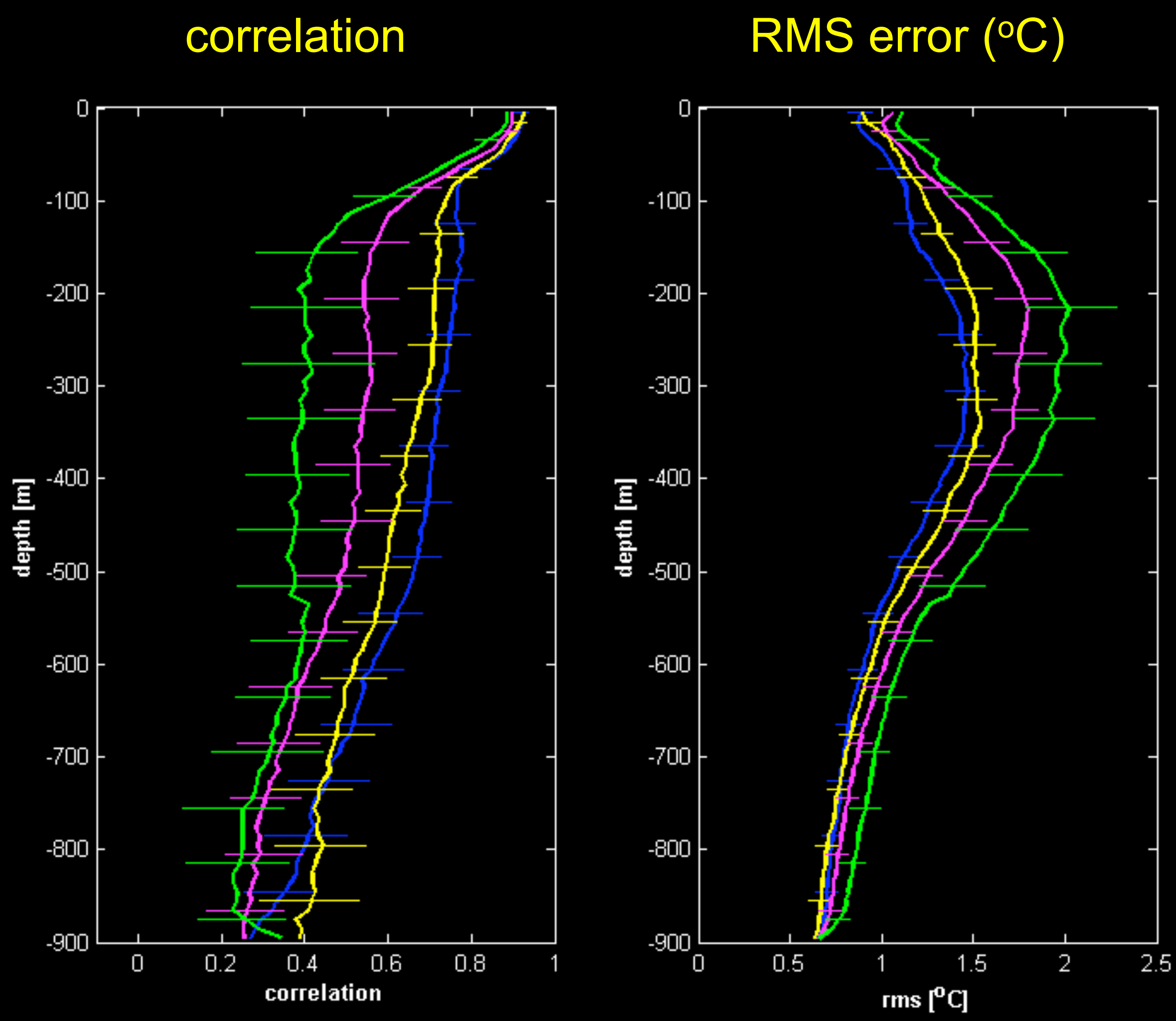
0 lag – analysis skill

1 week lag – little loss of skill

2 week lag – forecast begins to deteriorate

Comparison between ROMS subsurface temperature *predictions* and all *XBT observations* in 2001-2002

E3: SSH+SST+ Syn-CTD



0 lag – analysis skill

1 week lag – little loss of skill

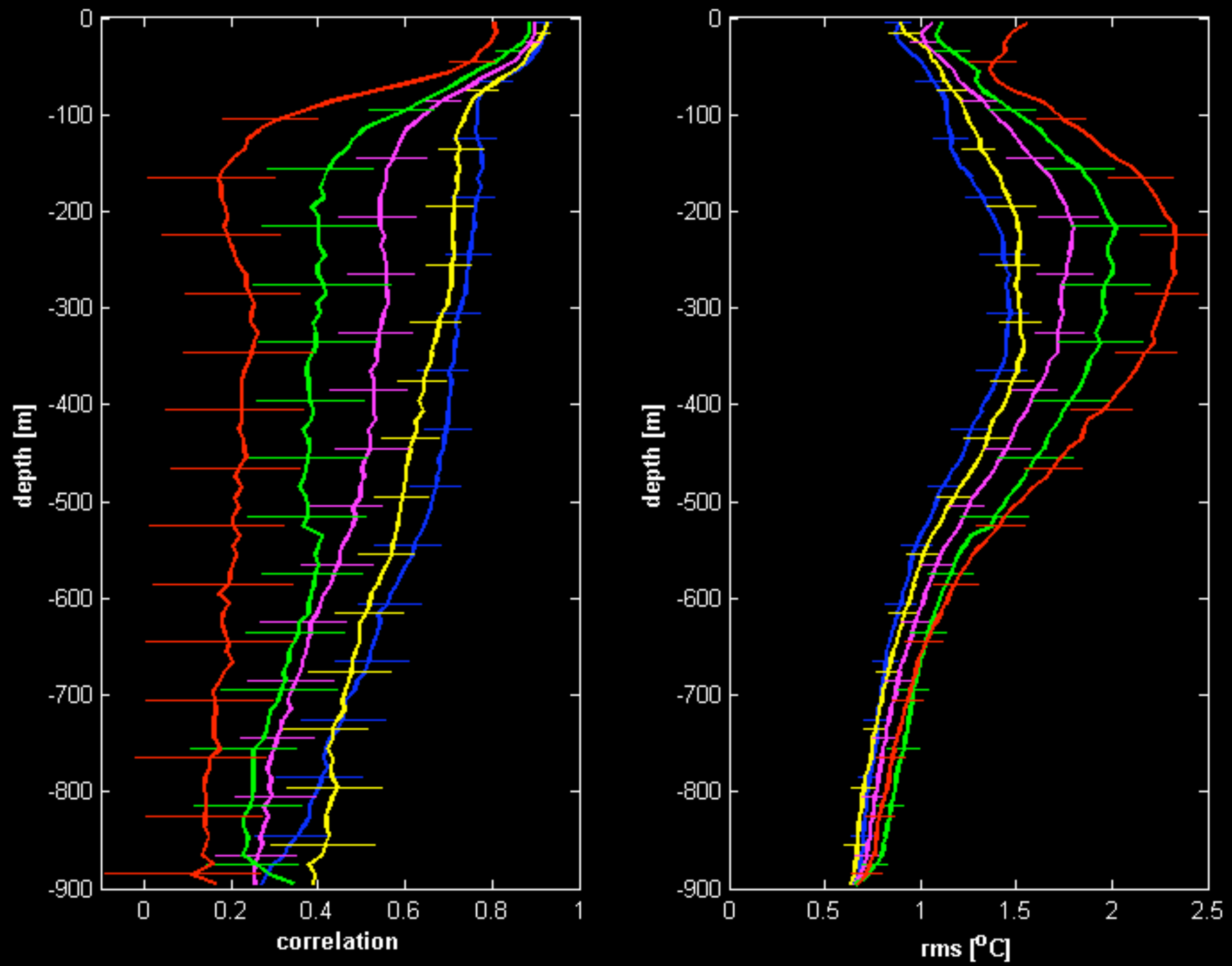
2 week lag – forecast begins to deteriorate

3 week lag – forecast still better than …

Comparison between ROMS subsurface temperature *predictions* and all *XBT observations* in 2001-2002

correlation

RMS error (^oC)



E3: SSH+SST+ Syn-CTD

0 lag – analysis skill

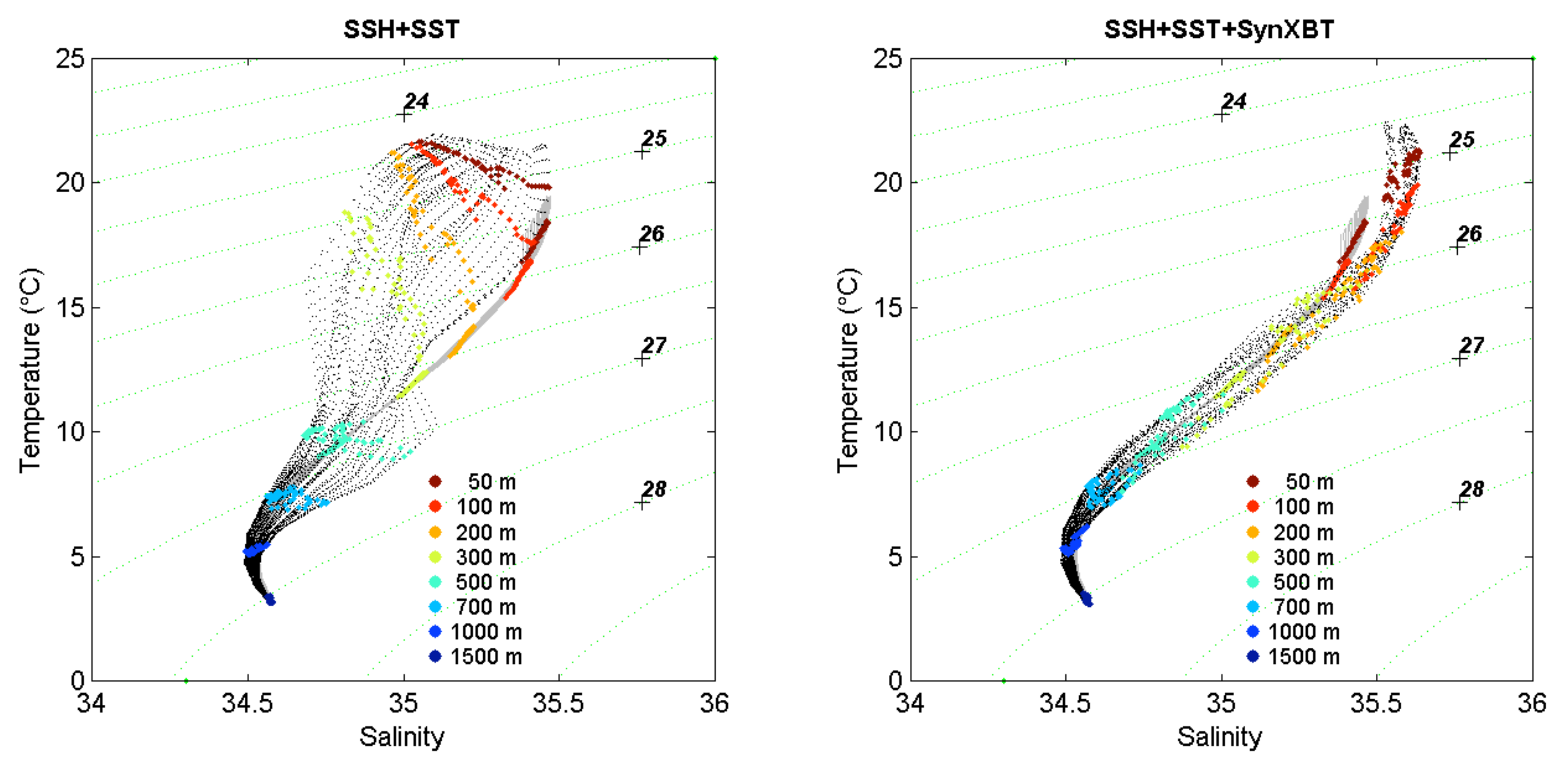
1 week lag – little loss of skill

2 week lag – forecast begins to deteriorate

3 week lag – forecast still better than …

no assimilation

E1 E3*



E1: SSH, SST

***E3: SSH, SST, Syn-CTD**

Syn-CTD 4-day CSIRO subsurface projection of satellite obs to T(z), S(z)

Forecast uncertainty:
Ensemble predictions using Singular Vectors of the forecast

The eigenvectors of … $R^T(t,0)\ W\ R(0,t)$

ADROMS

TLROMS

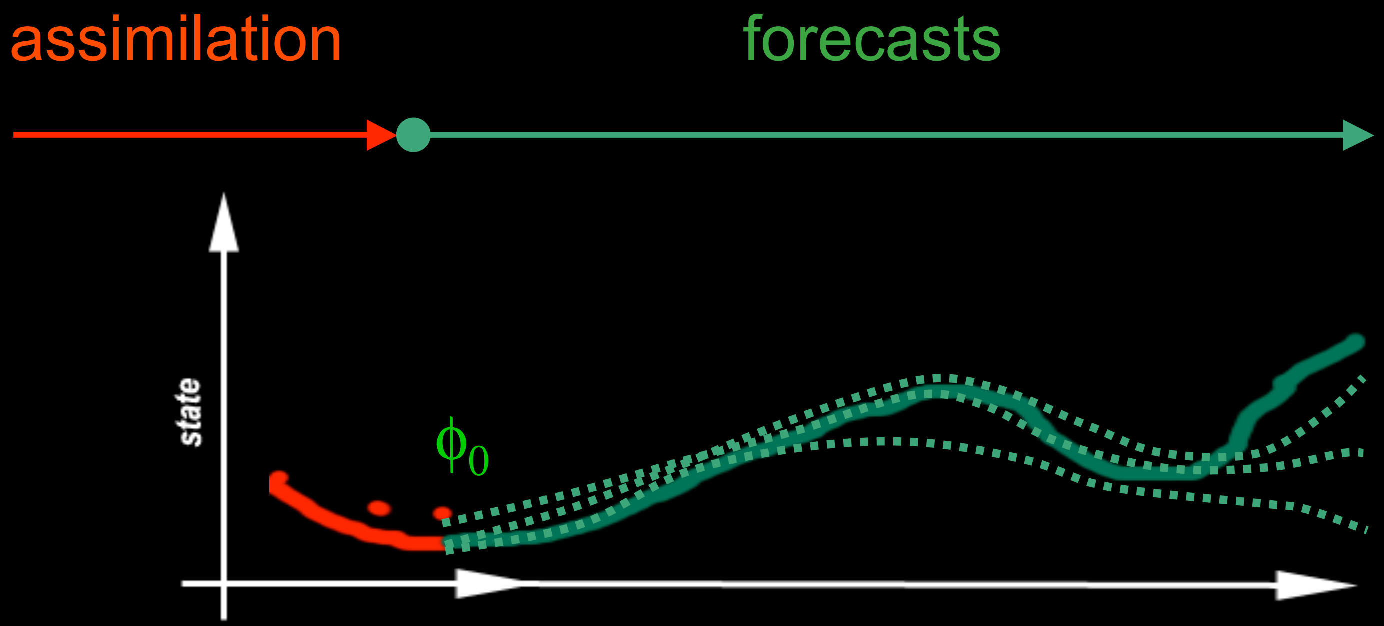
weights to define norm

…having the largest eigenvalues, are the fastest growing perturbations of the Tangent Linear model.

They correspond to the right *Singular Vectors* of *R(0,t) (the ROMS Tangent Linear propagator)*

These describe perturbations to the initial conditions that lead to the greatest uncertainty in the forecast

Forecast uncertainty:
Ensemble predictions using Singular Vectors of the forecast



SV of forecast $R(\phi_0)$

Perturb ϕ_0 for ensemble of IC:
r scales $max|\delta\phi|$ to 3 cm

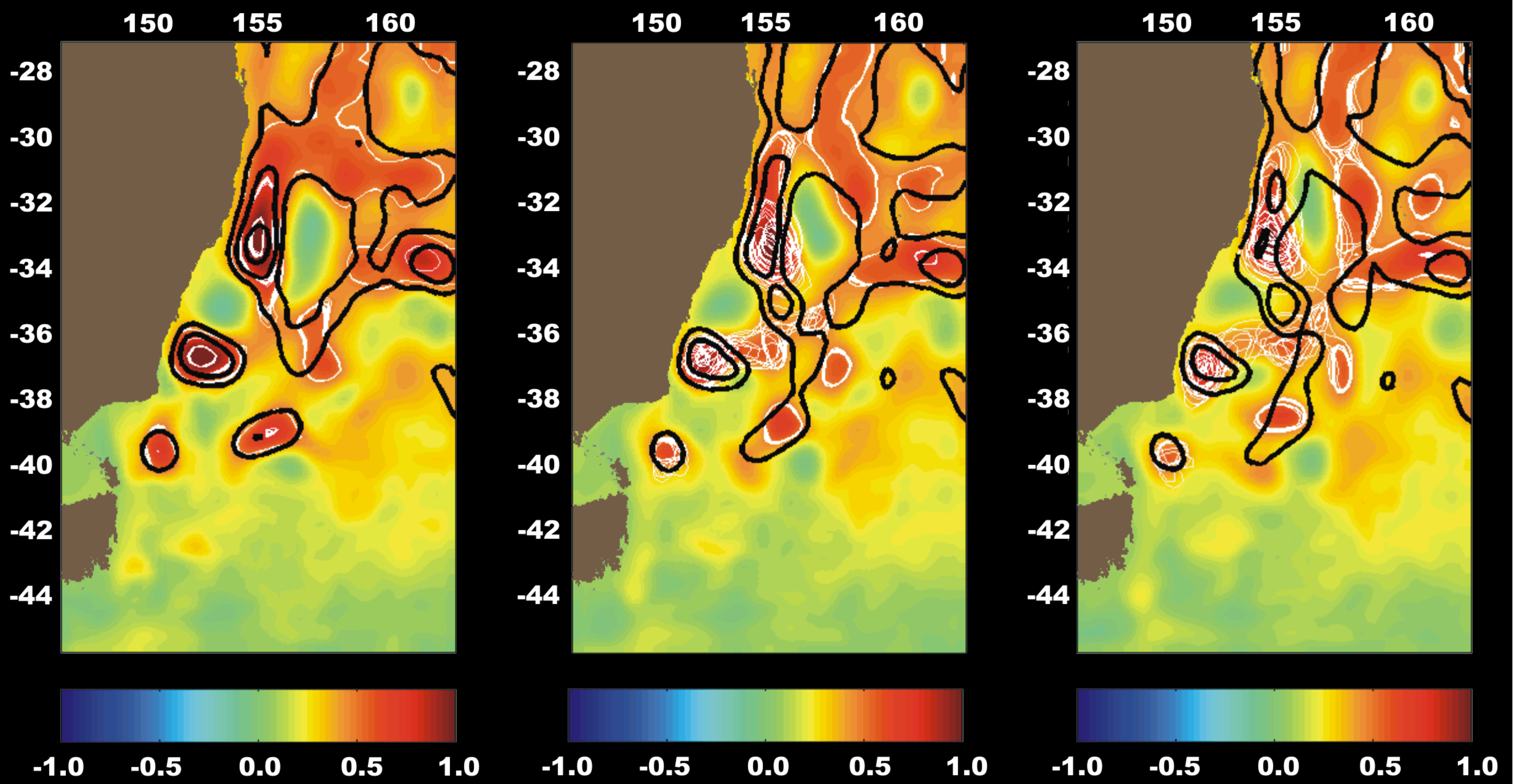
$$\delta\phi = r\sum_{i=1}^{10} a_i SV_i \quad , a_i \ are \ N(0,1)$$

Forecast uncertainty:
Ensemble predictions using Singular Vectors of the forecast

- The optimal perturbations when we include XBTs are more realistic: they tend to be concentrated at the surface, where most of the instability takes place.

White contours: Ensemble set
Color: Ensemble mean
Black contour: Observed SSH

Ensemble Prediction: E1



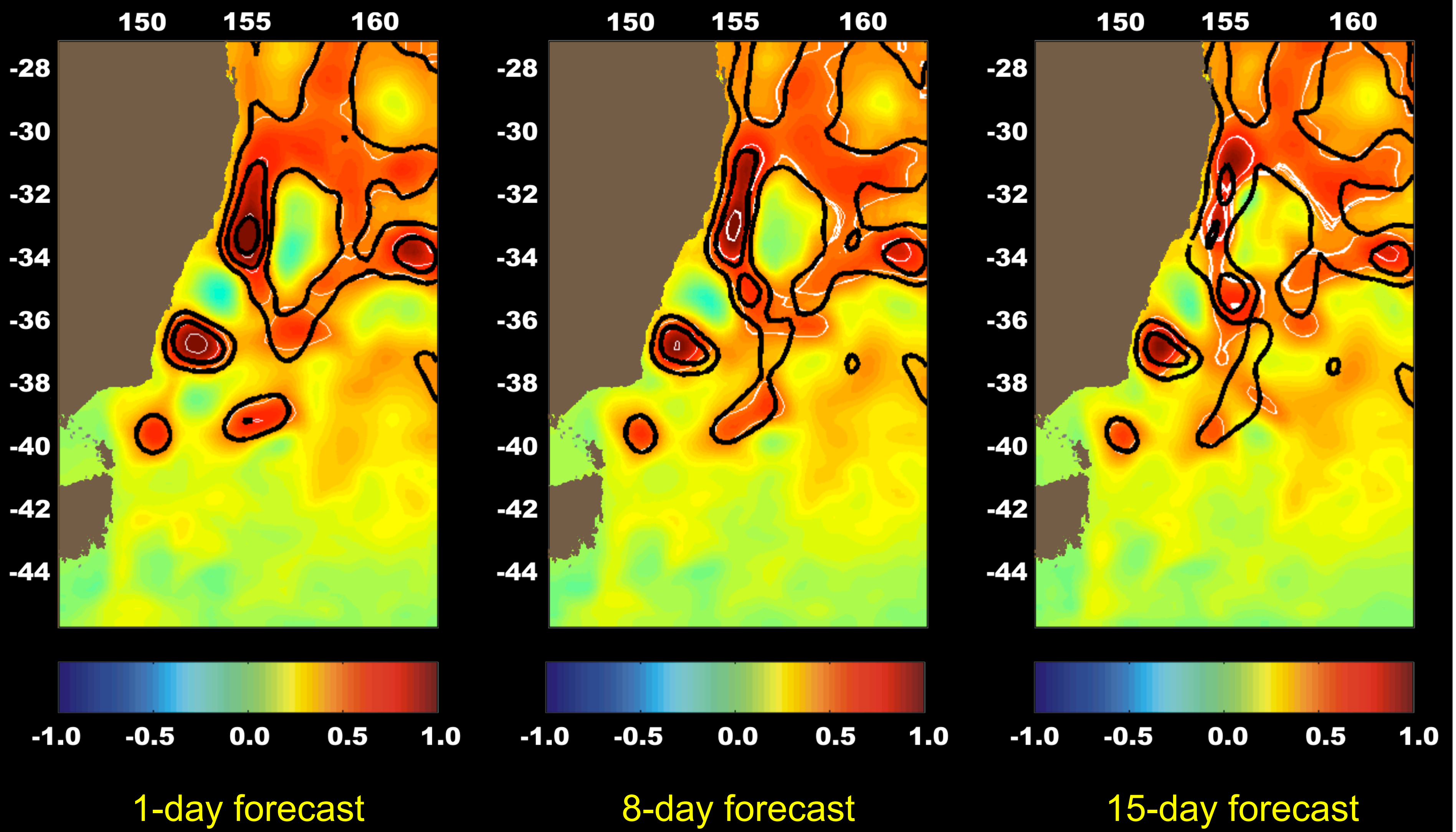
1-day forecast

8-day forecast

15-day forecast

Ensemble Prediction: E2

White contours: Ensemble set
Color: Ensemble mean
Black contour: Observed SSH



Forecast uncertainty:
Ensemble predictions using Singular Vectors of the forecast

- The optimal perturbations when we include XBTs are more realistic: they tend to be concentrated at the surface, where most of the instability takes place.
- When used in an ensemble prediction system the spread of E2 is smaller and verifies better with observations than that of E1.

- Subsurface XBT data significantly improves the forecast
- We have a further source of subsurface *information* based on surface observations: *synthetic-CTD*
  - a statistically-based proxy deduced from historical EAC data

Conclusions

- Skillful ocean state predictions up to 2+ weeks
- Assimilation of SST and SSH constrains surface well
- Subsurface information required (adjoint not enough)
  - improves estimate of the subsurface
  - makes forecasts more stable to uncertainty in IC
- Synthetic-CTD subsurface projection adds significant analysis and forecast skill
  - syn-CTD is a linear empirical relationship, suggesting a simple dynamical relationship links surface to subsurface variability
  - could be built in to the background error covariance (Weaver et al 2006, "…*balance operator for variational ocean data assimilation …*", QJRMS)
- Singular Vectors demonstrate ensemble predictions and uncertainty estimation
- Computational effort: 1 week analysis + forecast takes 4 hours on 8-processors (AMD Opteron-250) (1/4^o resolution)

Future Work

- Include balance terms in the IS4DVAR
- Improve surface forcing and open boundary conditions
  - better external analysis - BLUELINK
  - include boundary data in control variables
  - improve surface forcing via weak constraint data-assimilation (WS4DVAR)
- Use along-track SSH data instead of gridded multi-satellite analysis
- Explore sensitivity to length of assimilation window