A comparison of sequential data assimilation schemes for ocean prediction with HYCOM

Twin Experiments

A. Srinivasan, University of Miami, Miami, FL

E. P. Chassignet, COAPS, Florida State University, Tallahassee, FL

O. M. Smedstad, QinetiQ North America, Stennis Space Center, MS, USA

T. M. Chin, University of Miami, Miami, FL

F. Counillon, Nansen Center, Bergen, Norway

J. M. Brankart and P. Brasseur LEGI, Grenoble, France

W. C. Thacker, AOML, NOAA, Miami, FL

J. A. Cummings, NRL Monterey, CA

Assimilation Schemes for HYCOM

- Multivariate Optimal Interpolation (MVOI)
  J. A. Cummings; Lorenc 1981, Daley 1991, Cummings 2005
- Ensemble Optimal Interpolation (EnOI)
  F. Counillon; Evensen 2003, Oke 2002
- Ensemble Reduced Order Information Filter (EnROIF)
  T. M. Chin; Chin etal 1999,01
- Fixed basis variant of the SEEK filter (SEEK)
  J. M. Brankart and P. Brasseur; Pham et al. 1998

Objective: To assess the performance of these schemes in identically configured twin experiments assimilating synthetic altimeter and surface temperature obs.

Domain & Model Configuration: 1/12° Gulf of Mexico nested in the 1/12° North Atlantic

Data Update Equation

Common linear formula for updating the model-forecast $\mathbf{x}^f$ to obtain data-analysis $\mathbf{x}^a$:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{1}$$

$\mathbf{y}$ is the data to be assimilated
$\mathbf{H}$ is the observation operator
$\mathbf{K}$ is an optimization parameter often called the *gain matrix*

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{2}$$

$\mathbf{P}^f$ is the forecast error covariance
$\mathbf{R}$ is the observation error covariance

Main difference among the four methods is in the

numerical representation of $\mathbf{P}^f$

MVOI

- $\mathbf{P}^f$ is separated into several variances and correlations

$$\mathbf{C}_h = (1 + s_h)exp(-s_h) \quad (3)$$
$$\mathbf{C}_v = (1 + s_v)exp(-s_v) \quad (4)$$
$$\mathbf{C}_f = (1 + s_f)exp(-s_f) \quad (5)$$
$$\mathbf{C}_b = \mathbf{C}_f\mathbf{C}_h\mathbf{C}_v \quad (6)$$

where $\mathbf{C}_h$ and $\mathbf{C}_v$ are the horizontal and vertical correlations and s_h and s_v are scaled distances.

- analysis on z levels (42 zlevels between 0-2500 m)
- state variables: U, V, T, S and intf. Pressure
- dynamical method (Cooper-Haines 1996) for vertical projection
- simultaneous 3D analysis
- obs errors uncorrelated

EnOI

In EnOI, the forecast covariance matrix is essentially the *sample covariance* of an ensemble of model states

$$\mathbf{P}^f_{\text{EnOI}} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\mathbf{x}^f_m - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}^f_m - \overline{\mathbf{x}}^f\right)^T \qquad (7)$$

where $\mathbf{x}^f_m$ is the m^{th} sample of the forecast ensemble, $\overline{\mathbf{x}}^f$ is the ensemble mean, and M is the number of samples. The analysis update is computed as:

$$\mathbf{x}^a = \mathbf{x}^f + \alpha\mathbf{P}^f\mathbf{H}^T(\alpha\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \qquad (8)$$

where $\alpha \in (0,1]$ is a parameter introduced to allow different weights for the ensemble and measurements.

State Varaibles: UT, VT, UB, VB, PB, DP, T, S, SSH, SST

Vertical Projection: static correlations; analysis in observation space

EnROIF

EnROIF uses a *Markov random field* (MRF) to model the forecast error process as

$$e(i,j) \;=\; \sum_{(\triangle i, \triangle j)\,\in\,\mathcal{N}} \gamma(i,j,\triangle i,\triangle j)\, e(i-\triangle i, j-\triangle j) \;+\; \delta(i,j) \qquad (9)$$

where $\mathcal{N}$ specifies a set of local grid locations, γ is the regression coefficient (small matrix), and $\delta(i,j)$ is a white noise with unit variance. Computation for the state analysis $\mathbf{x}^a$ in ROIF is performed as

$$\mathbf{L}^a(\mathbf{x}^a - \mathbf{x}^f) \;=\; \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \qquad (10)$$

where the sparsely banded analysis information matrix $\mathbf{L}^a = \mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is numerically inverted.

State Varaibles: DP, SSH, SST

Vertical Projection: static correlations; 2D vertically decoupled analysis

Fixed basis SEEK

In SEEK filter, $\mathbf{P}^f$, is assumed to be of low rank, M and is represented by dominant modes of *empirical orthogonal functions* (EOFs) as

$$\mathbf{P}^f_{\text{SEEK}} = \frac{1}{M-1}\sum_{m=1}^{M}\mathbf{S}^f_m\mathbf{S}^{fT}_m \tag{11}$$

where $\mathbf{S}^f_m$, $m = 1, \ldots, M$, are the M most dominant EOF modes. In

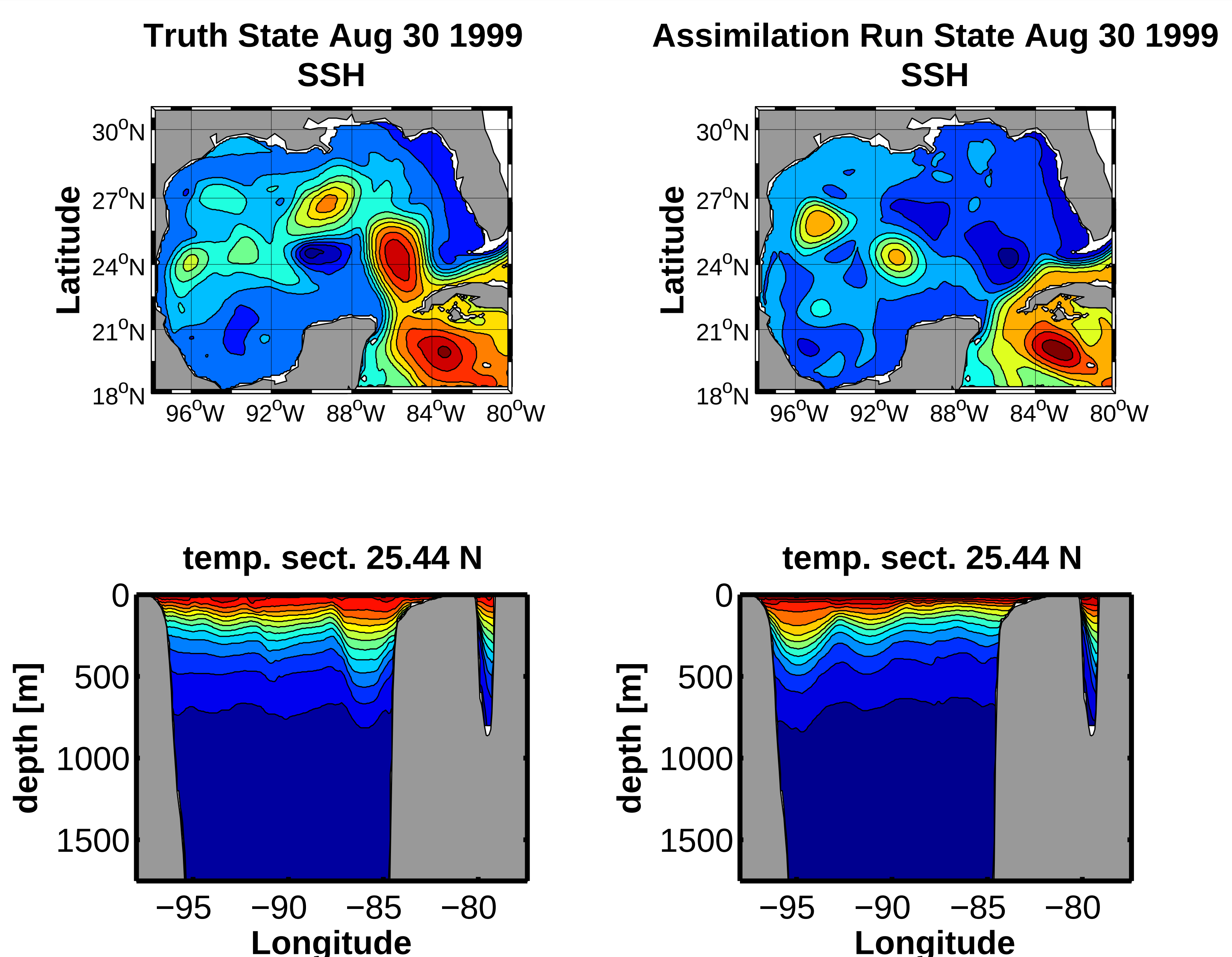
the SEEK analysis the Kalman Gain is rewritten using the Sherman-Morrrison-Woodberry matrix identity as

$$\mathbf{K} = \mathbf{S}^f[\mathbf{I} + \left(\mathbf{H}\mathbf{S}^f\right)\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{S}^f\right)^T]^{-1}\left(\mathbf{H}\mathbf{S}^f\right)\mathbf{R}^{-1} \tag{12}$$

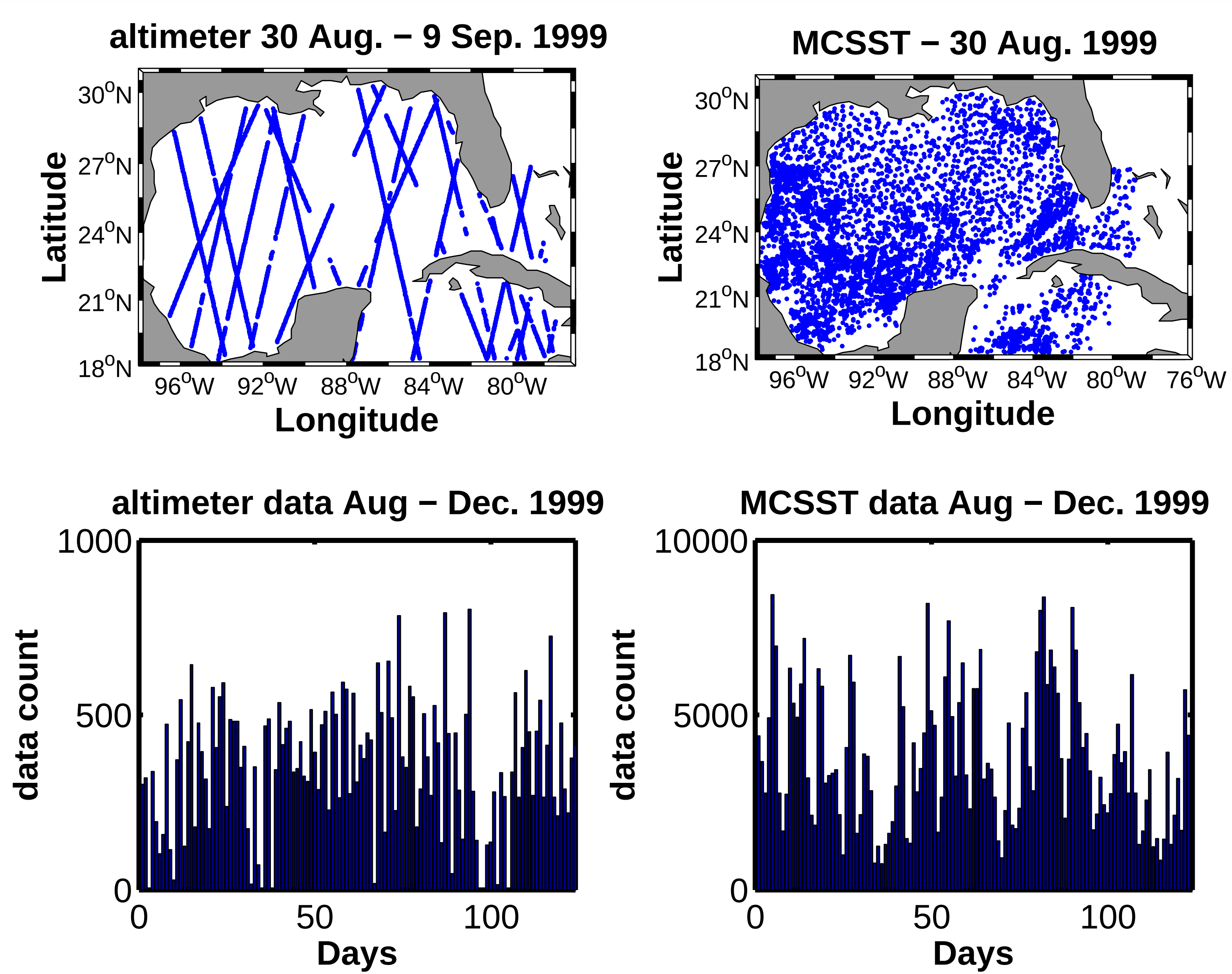
State Varaibles: UT, VT, UB, VB, PB, DP, T, S, SSH, SST

Vertical Projection: static correlations; analysis in reduced space

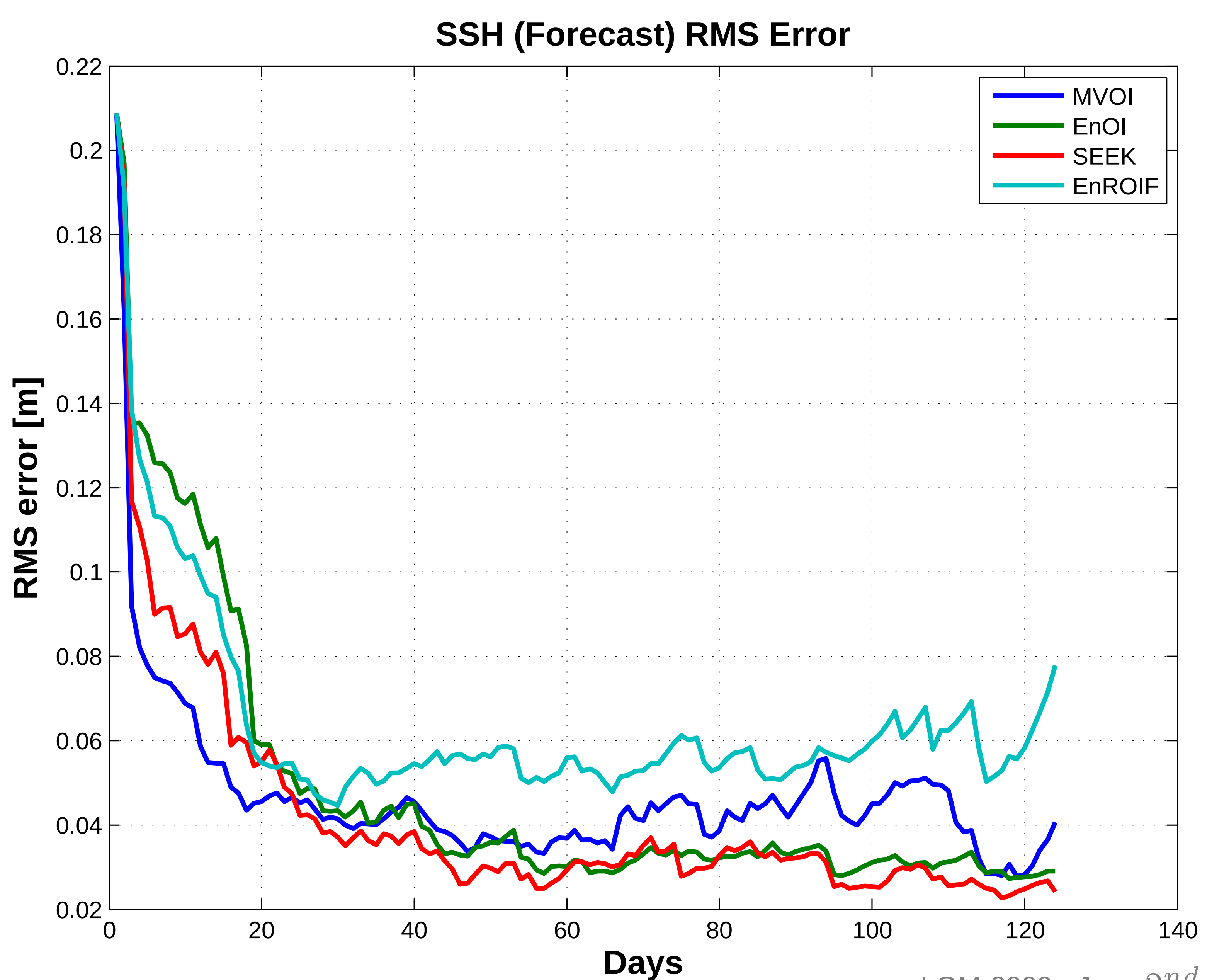
Twin Experiments: Initial States



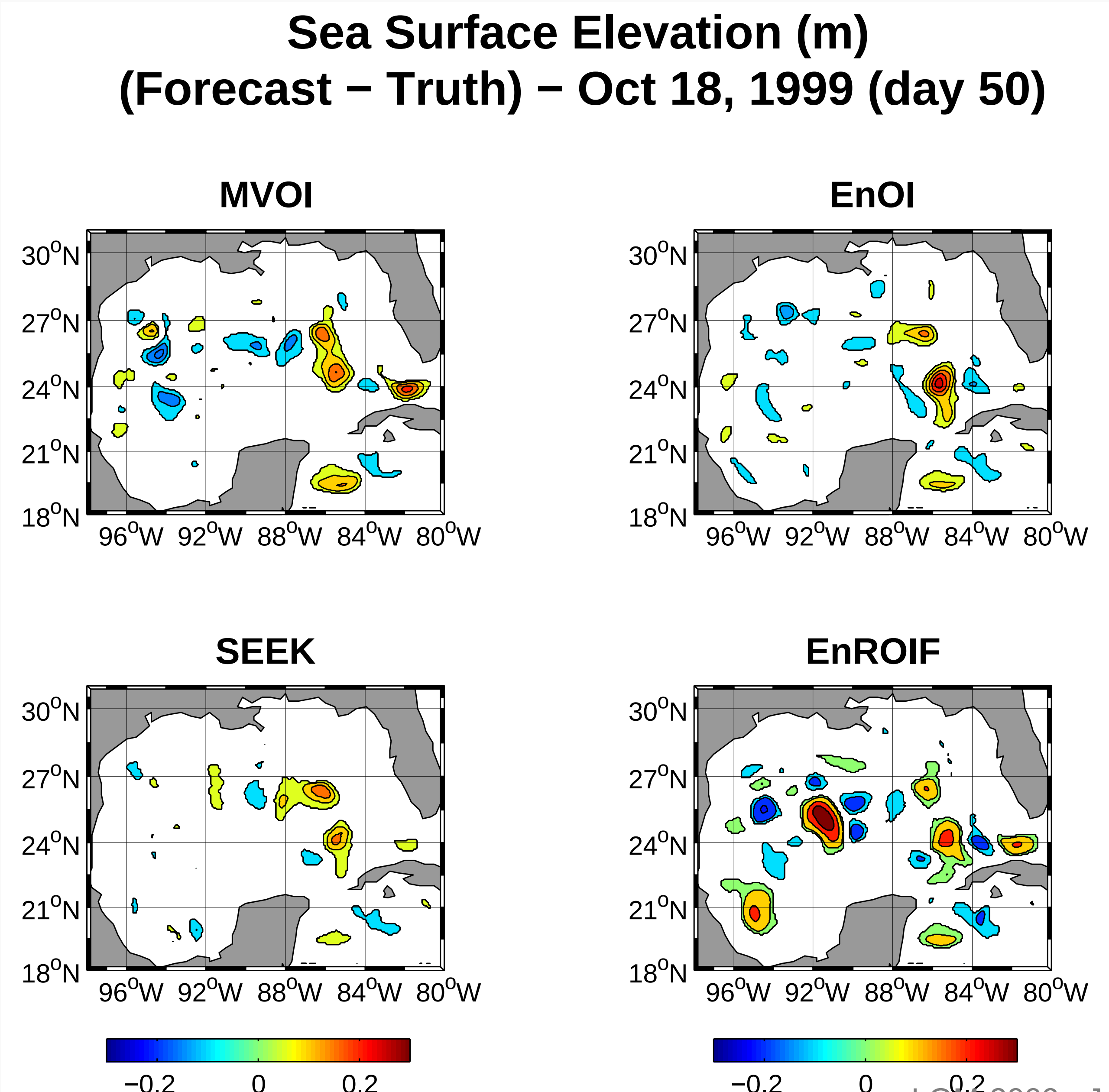
Twin Experiments: Data



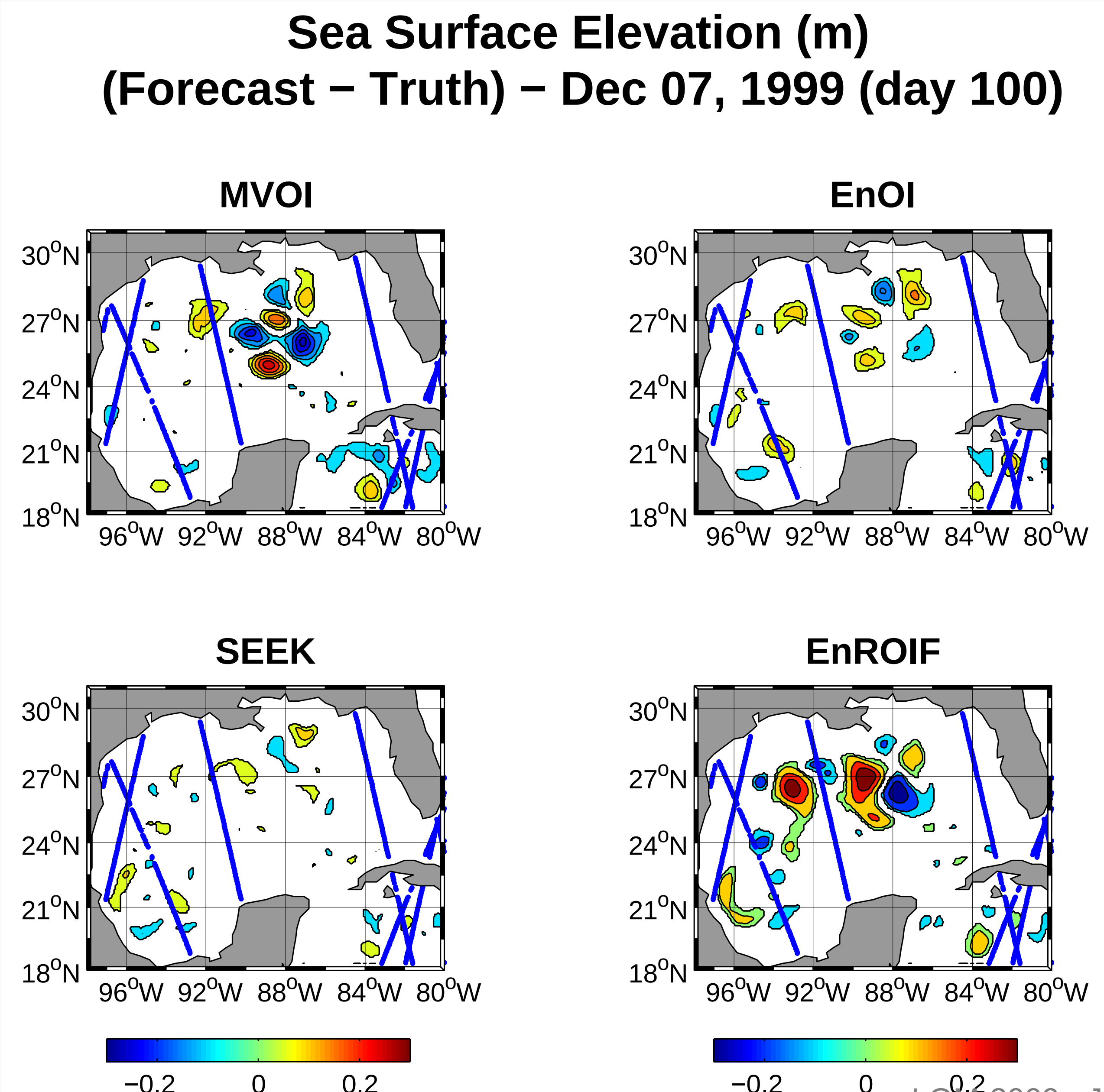
Twin Experiments: SSH



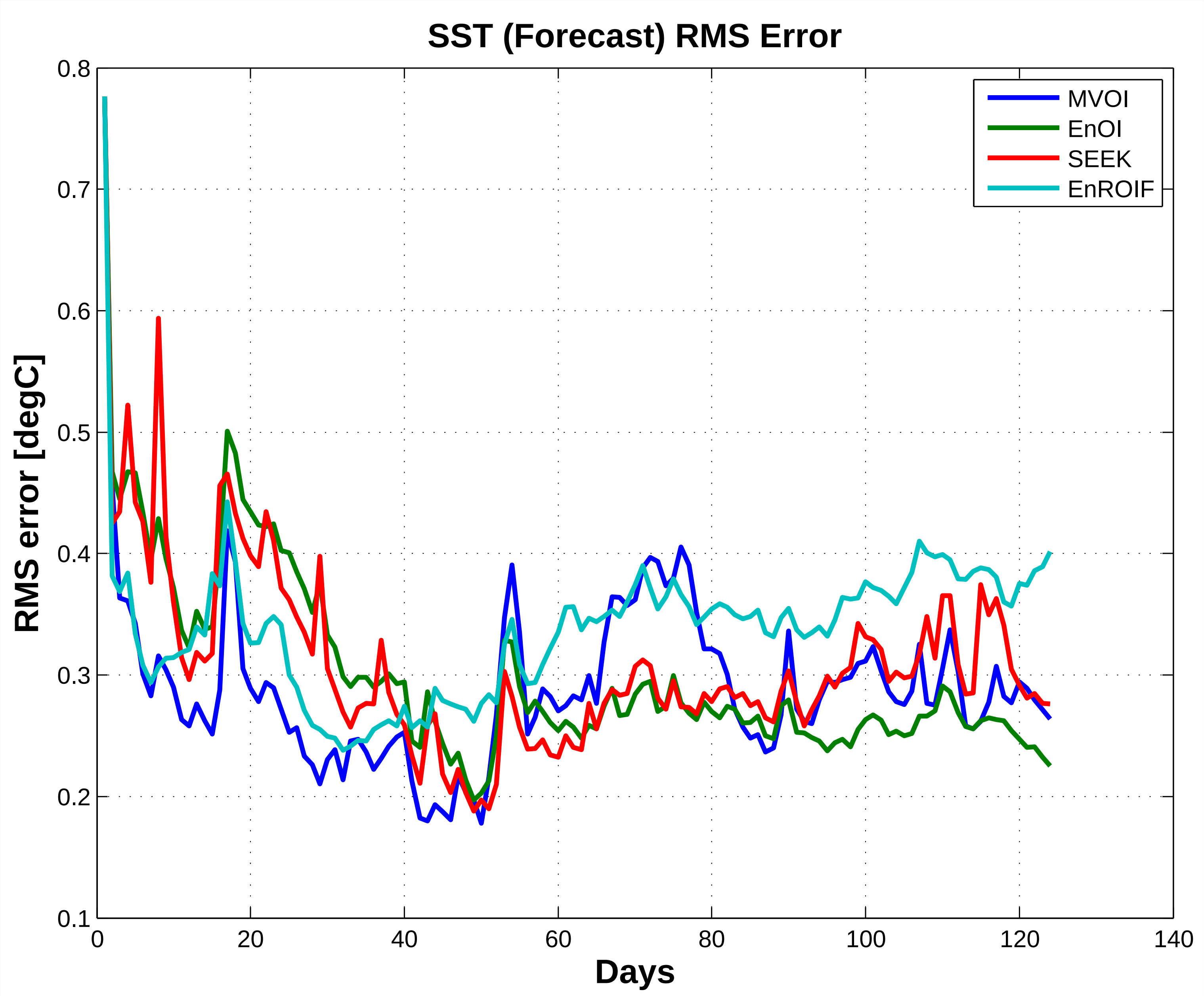
Twin Experiments: SSH



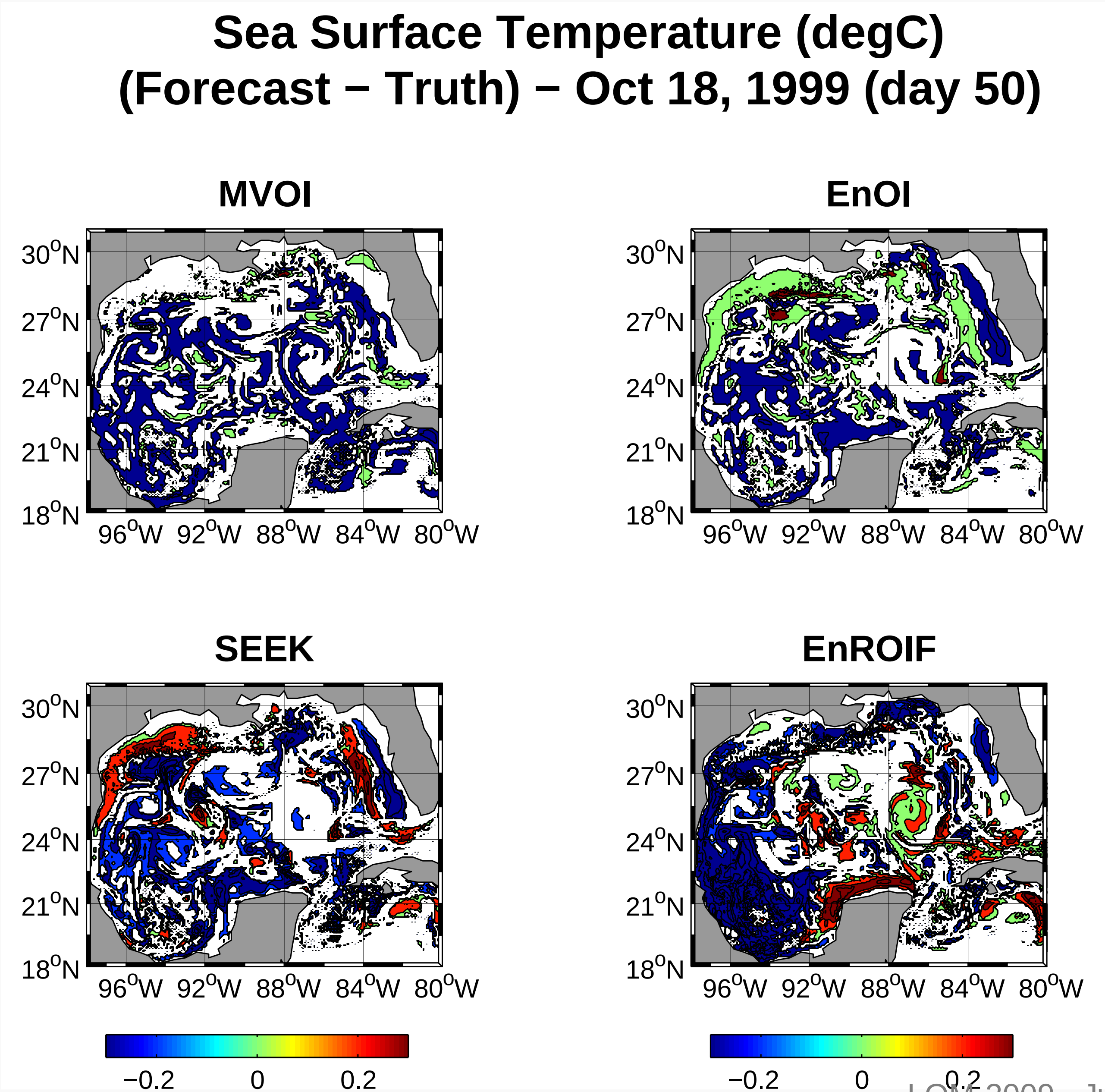
Twin Experiments: SSH



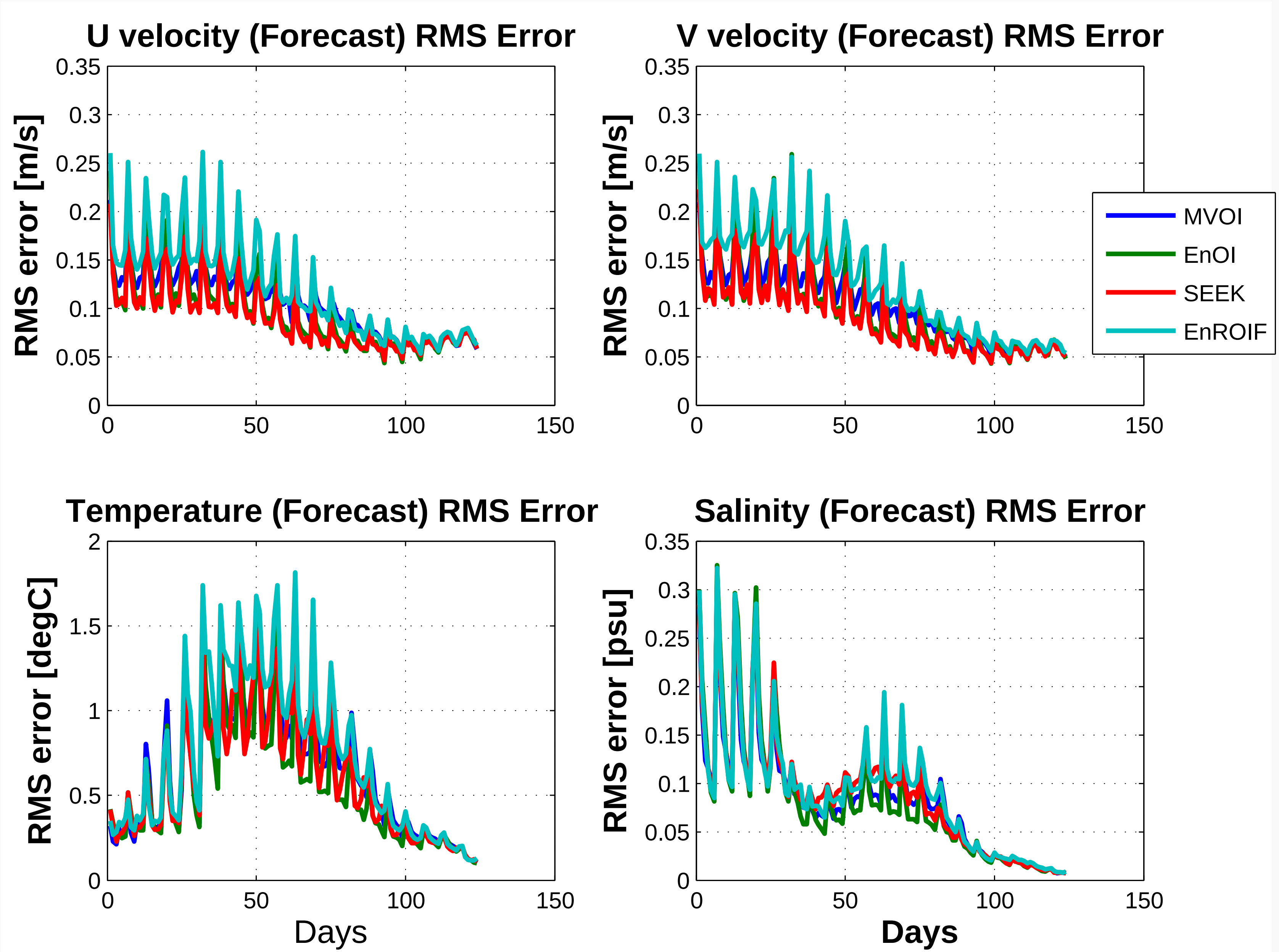
Twin Experiments: SST



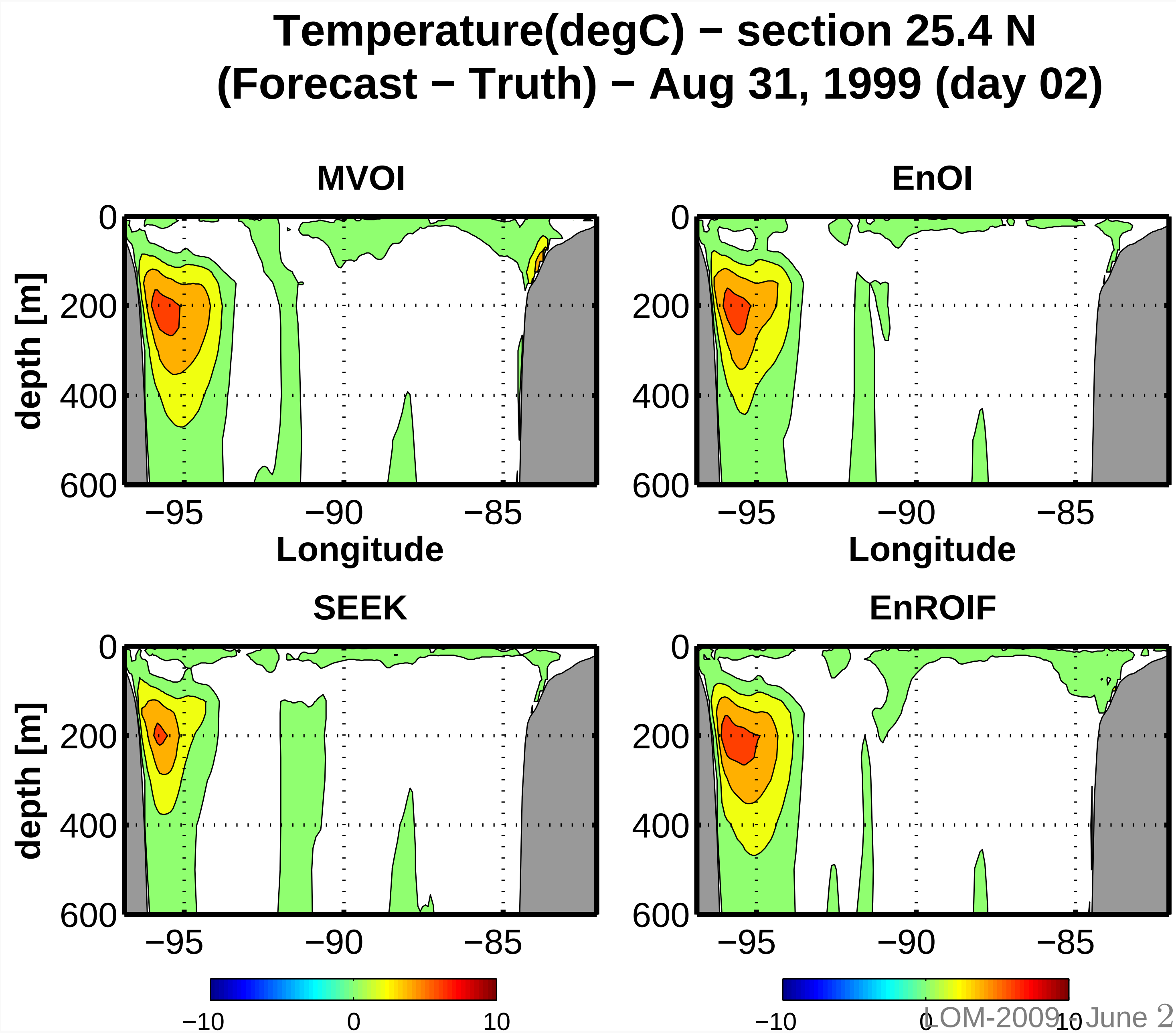
Twin Experiments: SST



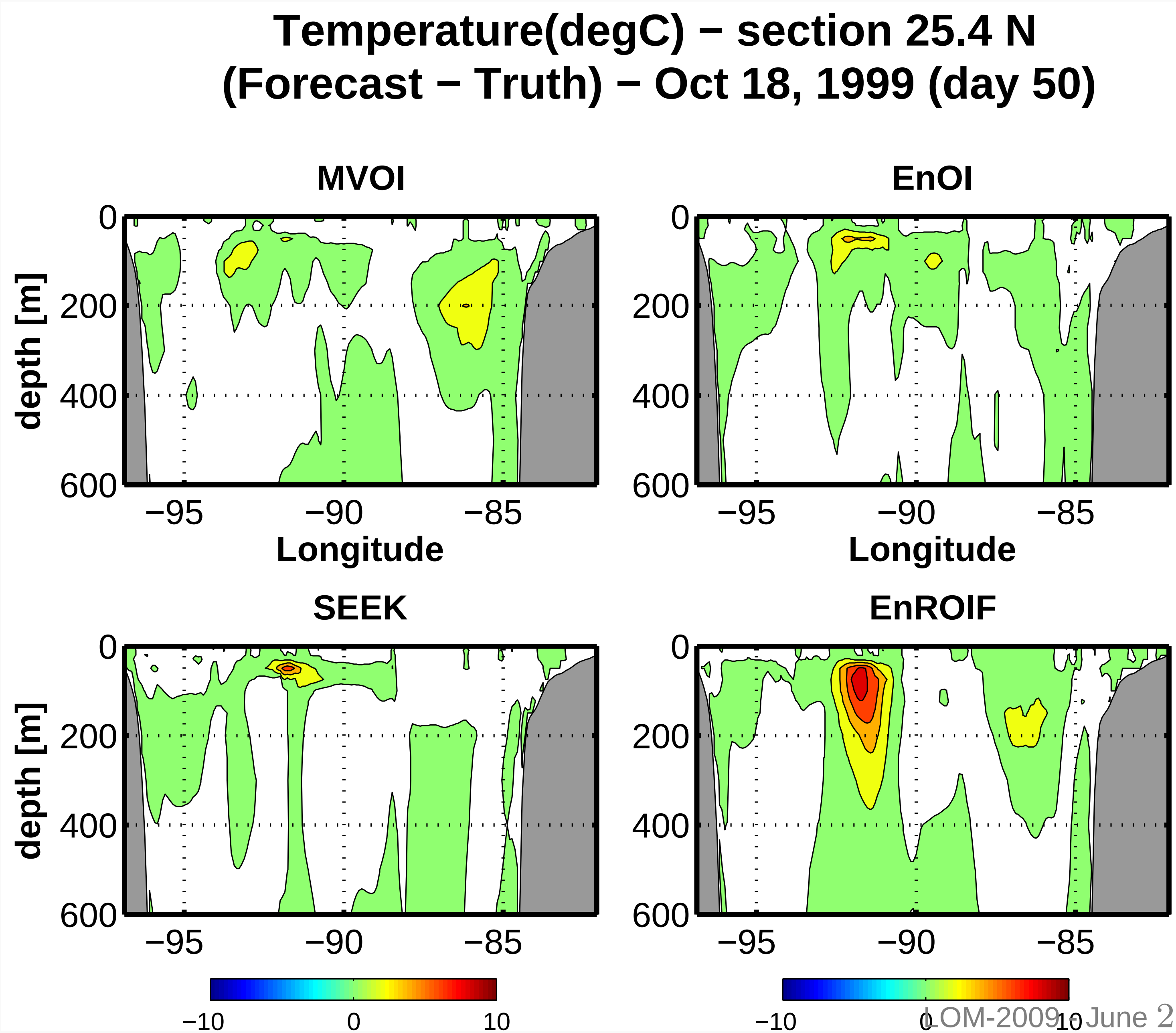
Twin Experiments: Unobserved Variables [u,v, t,s]



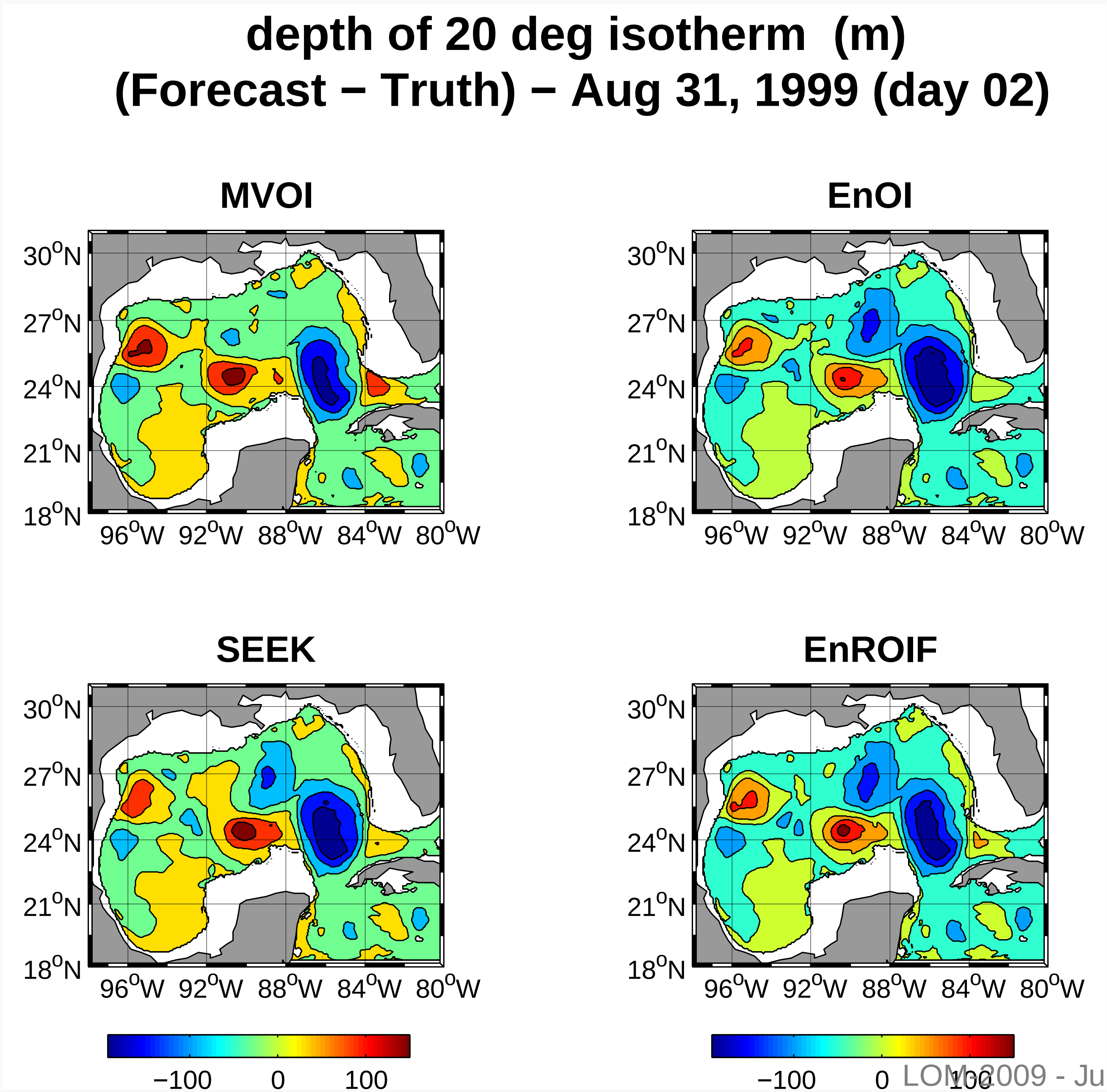
Twin Experiments:



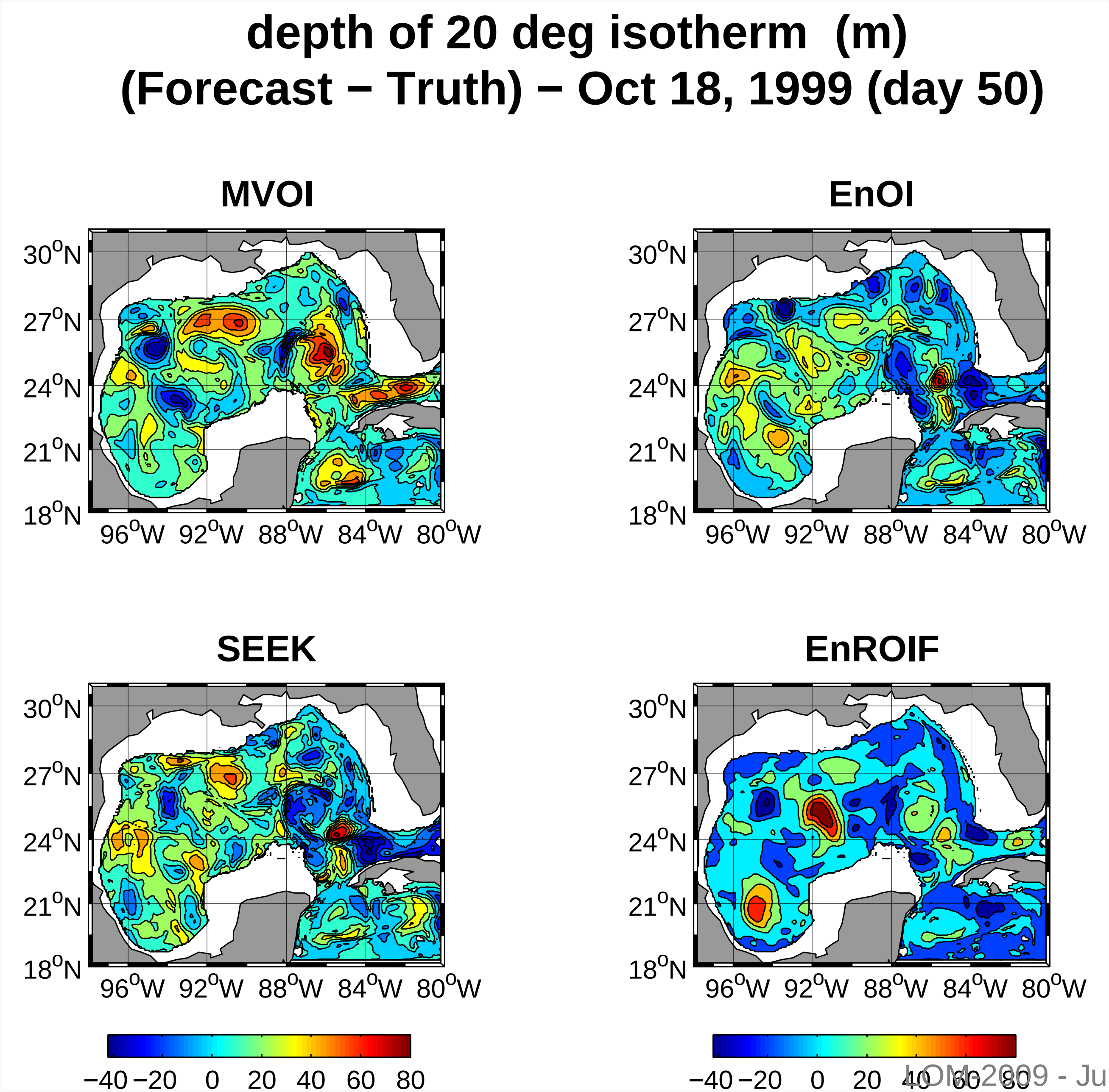
Twin Experiments:



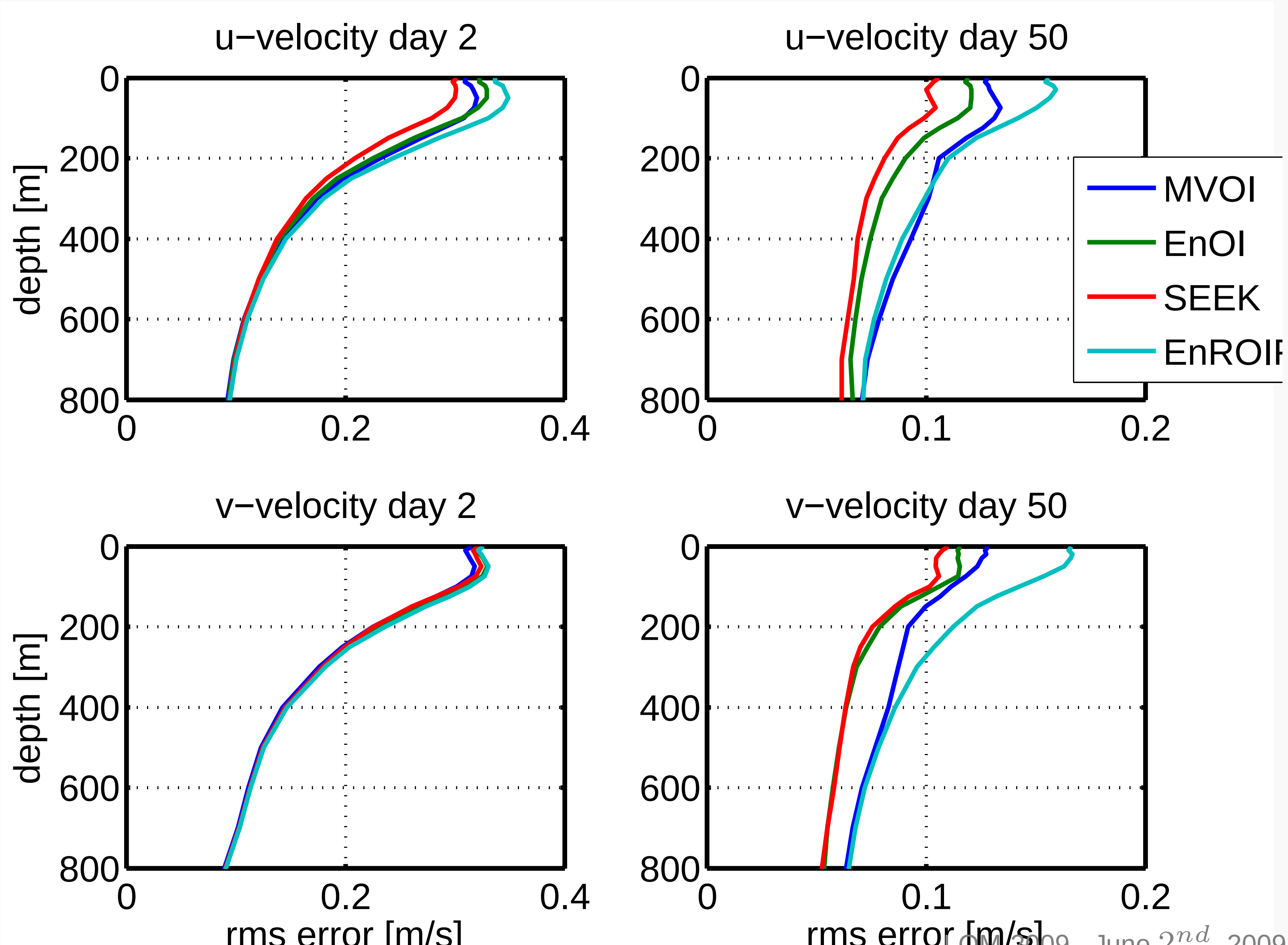
Twin Experiments:



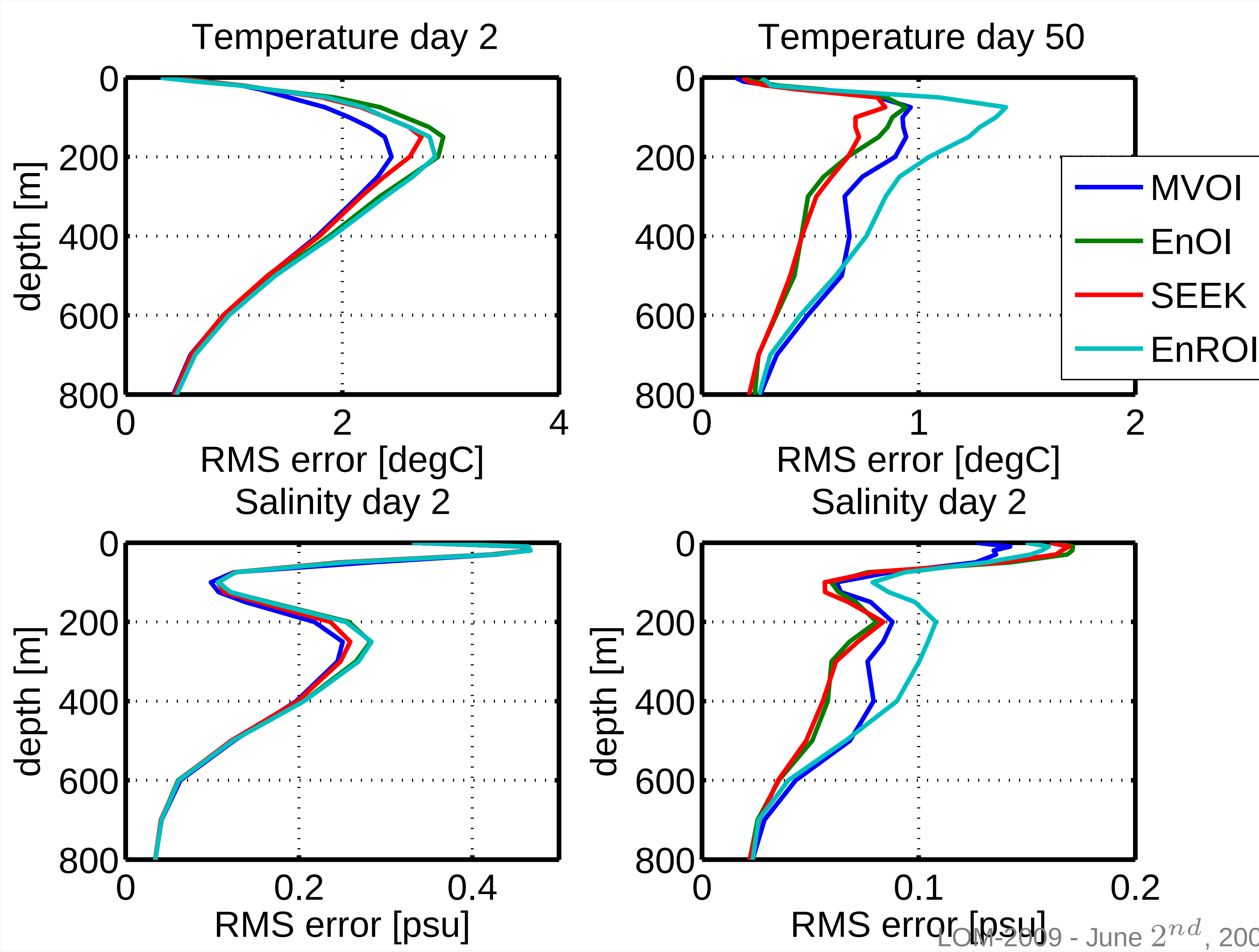
Twin Experiments:



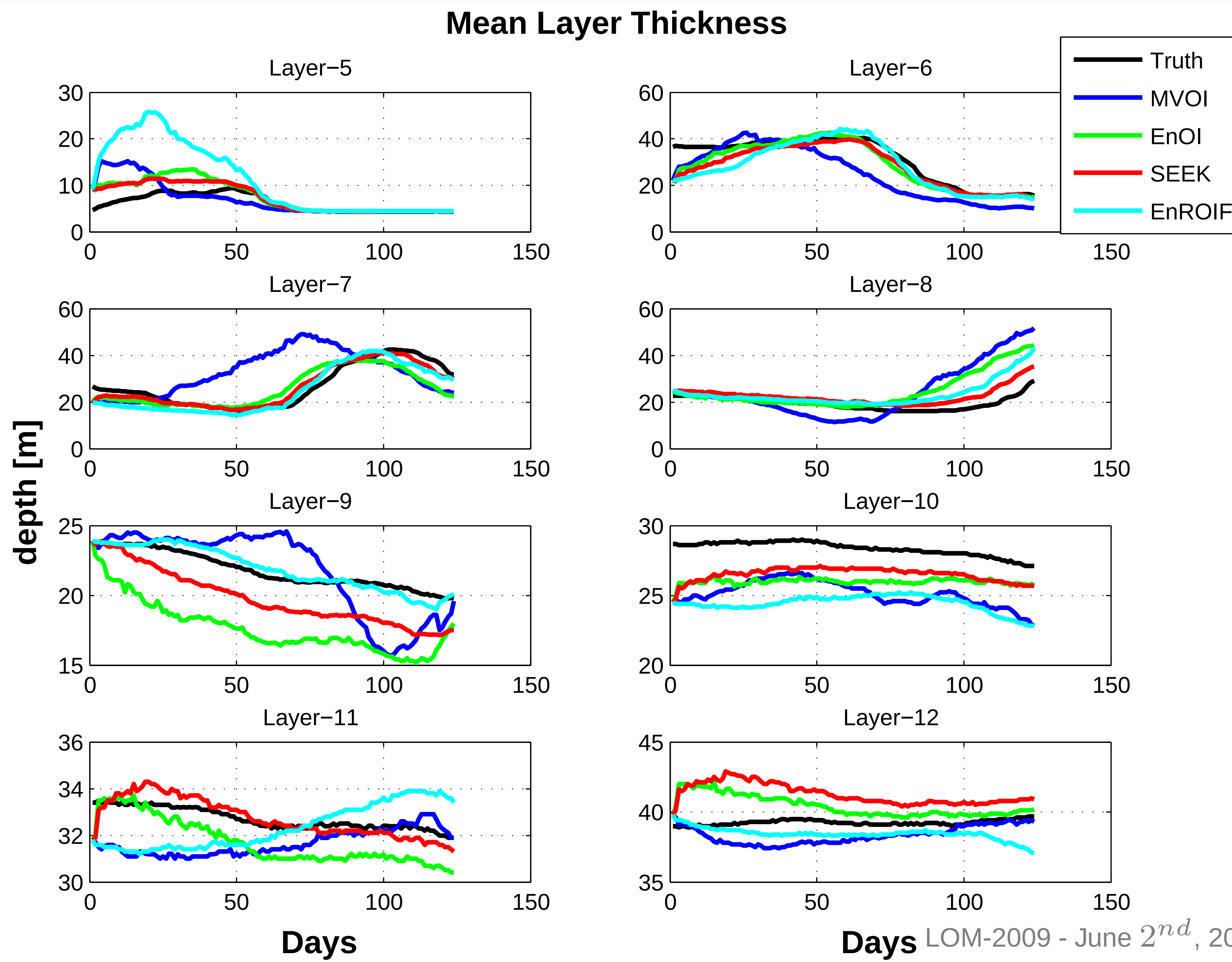
Twin Experiments: 3D U&V



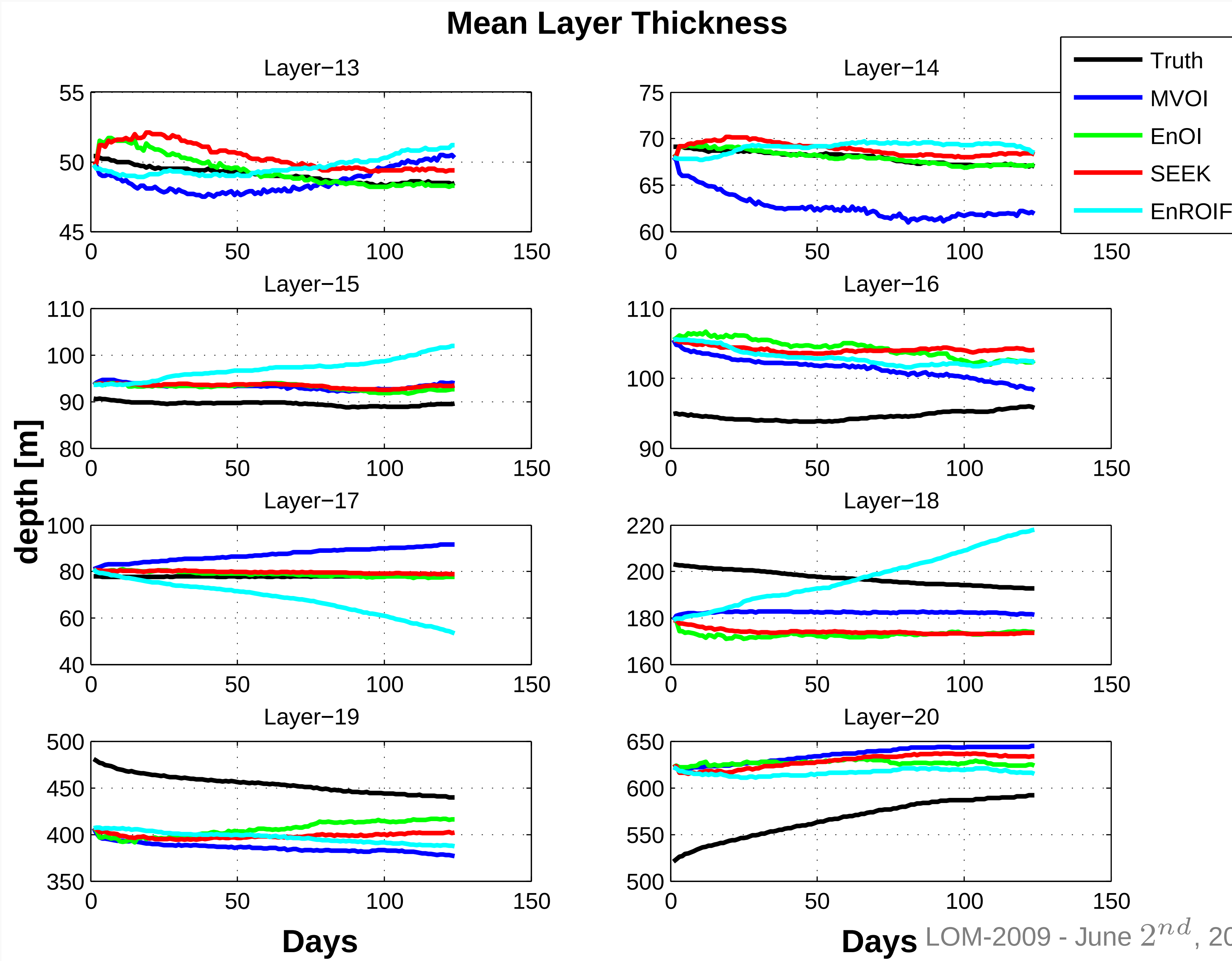
Twin Experiments: 3D T&S



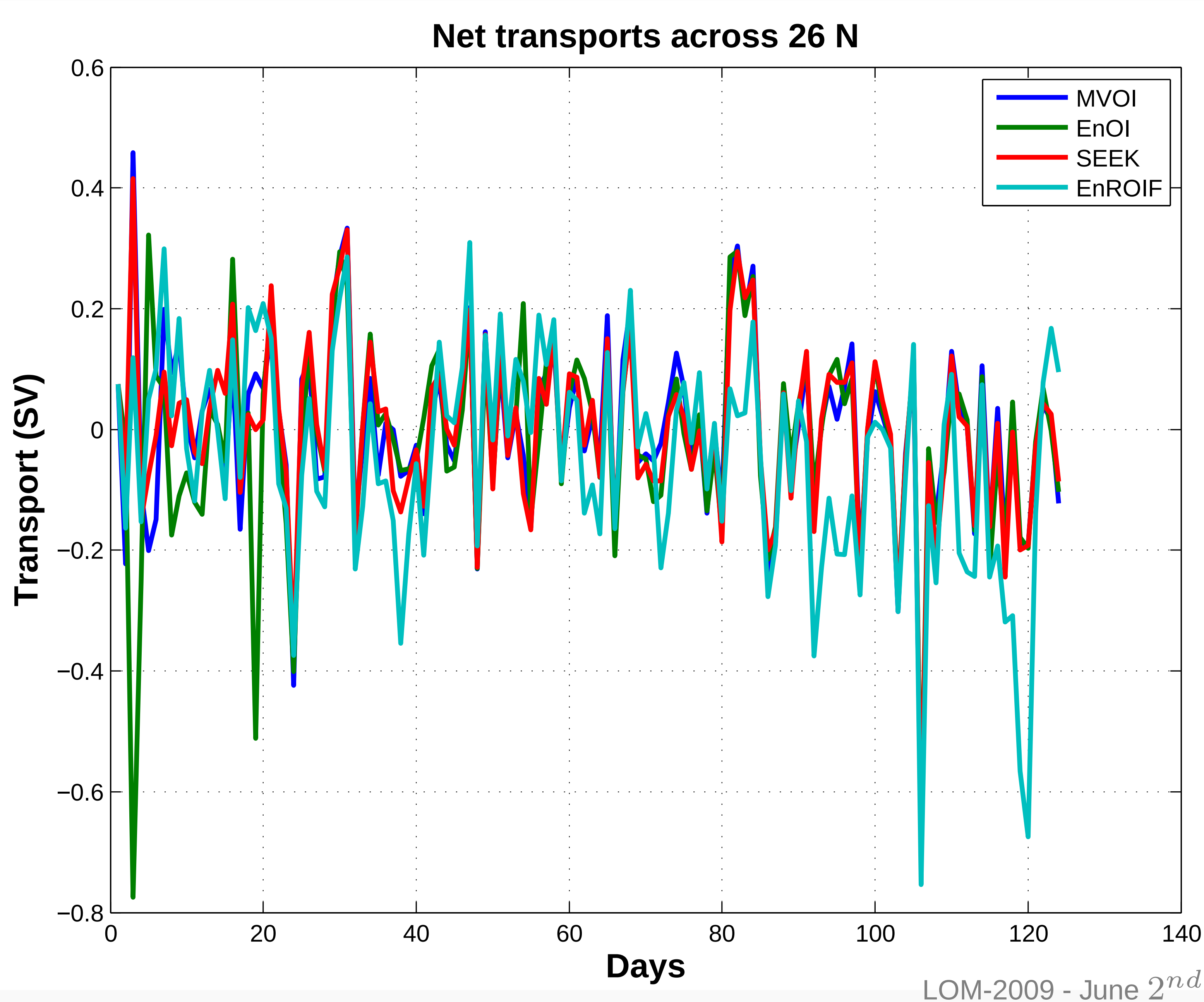
Twin Experiments: Layer Thickness



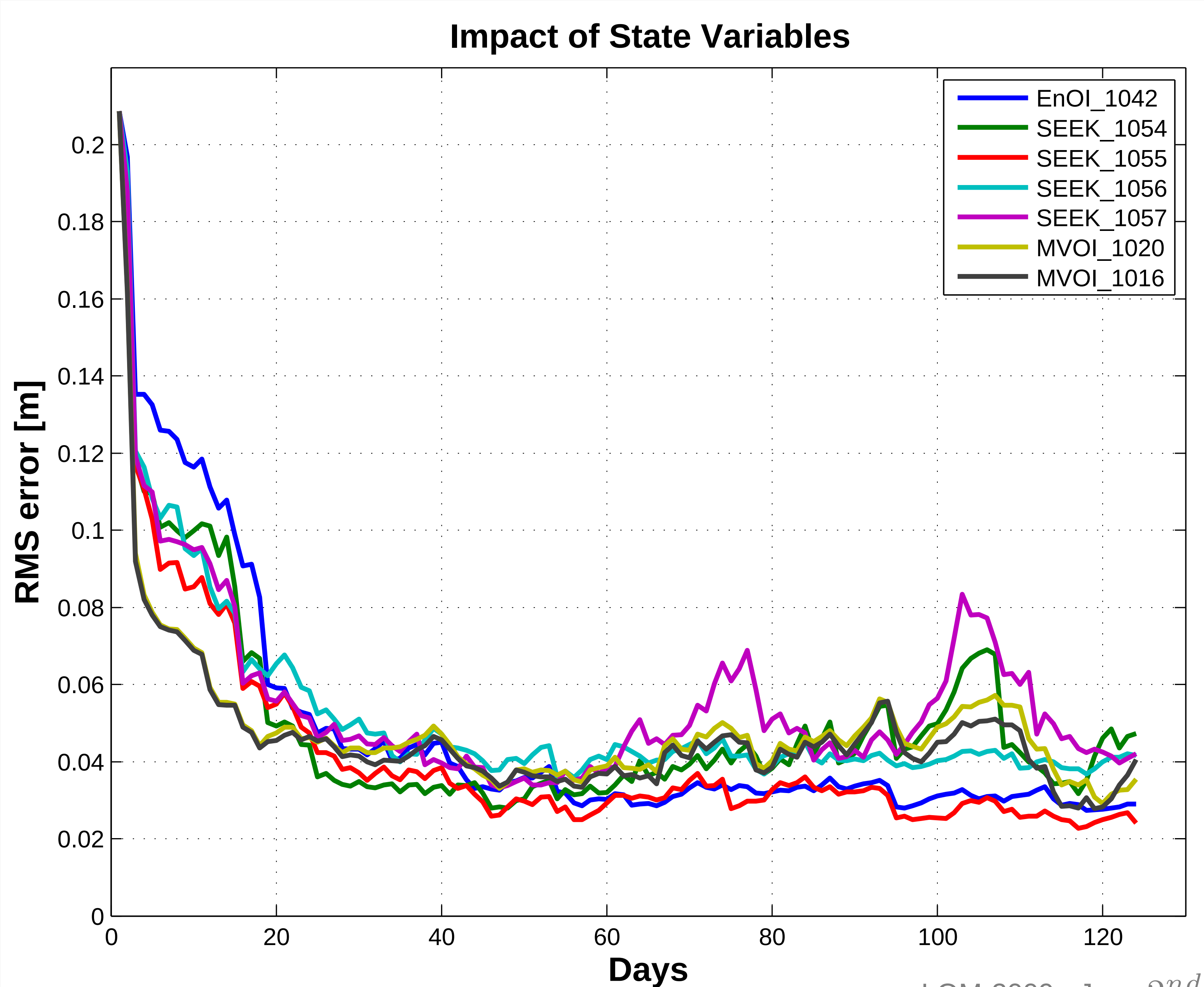
Twin Experiments: Layer Thickness



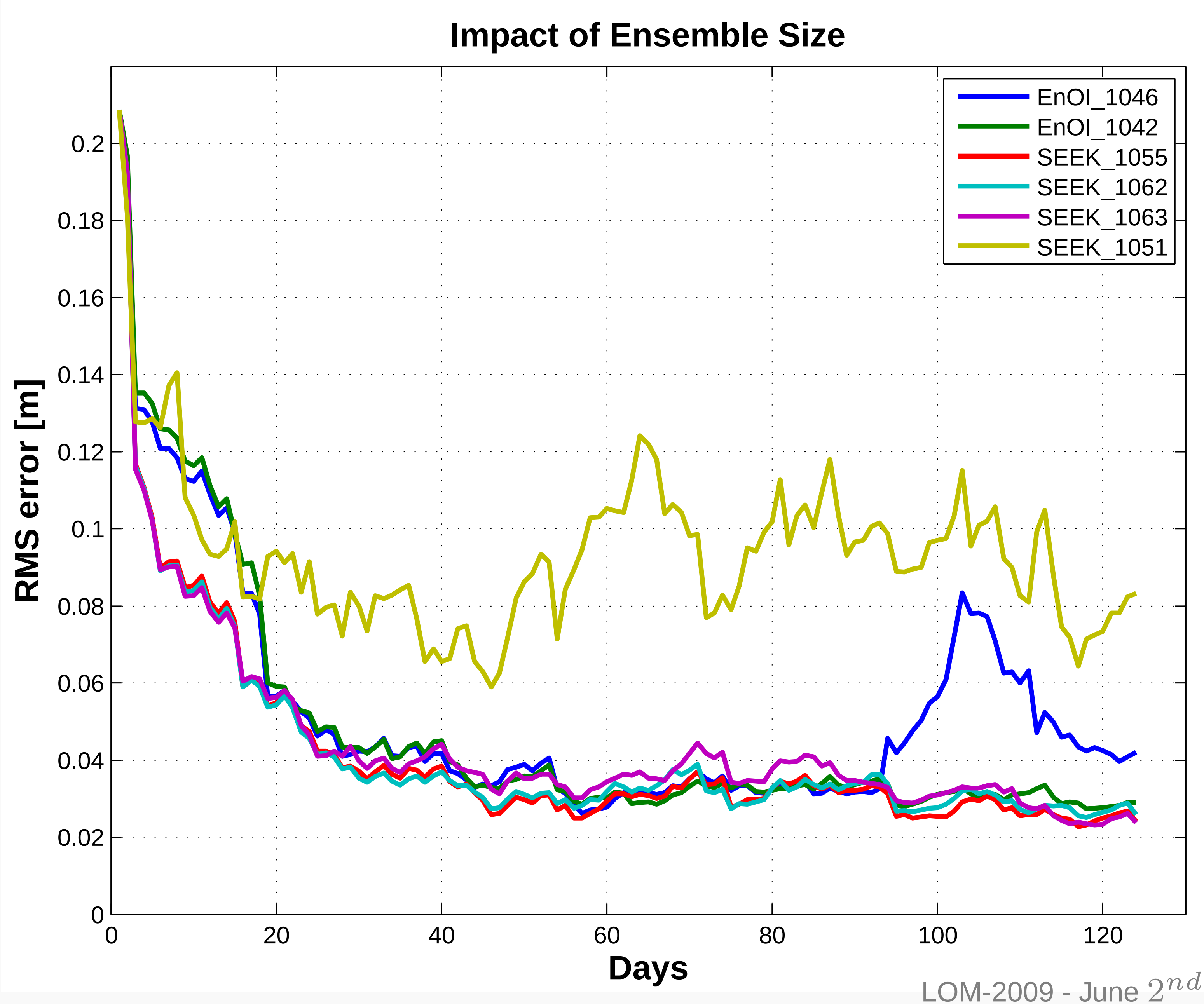
Twin Experiments: Net Transports



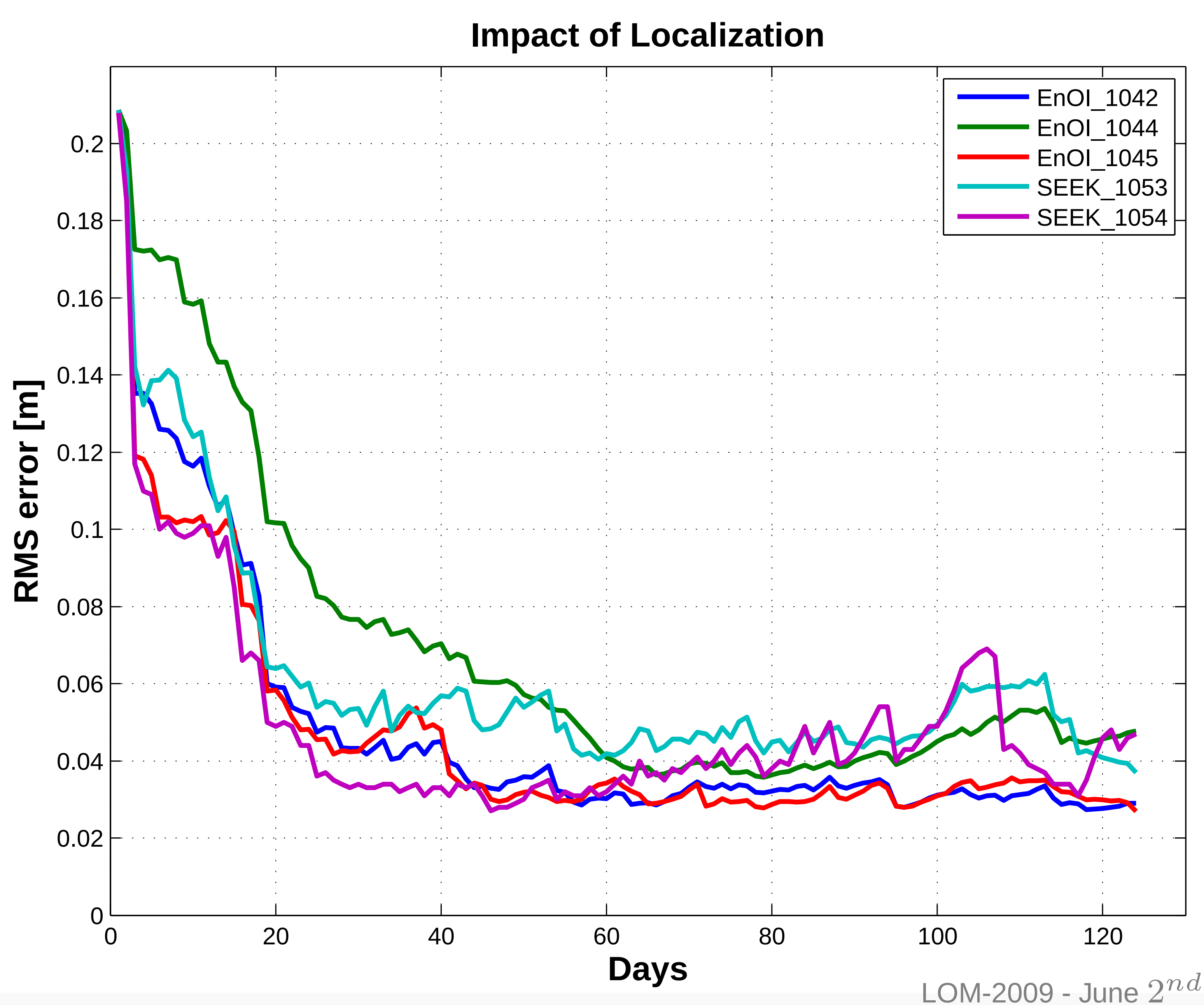
Twin Experiments: State Variables



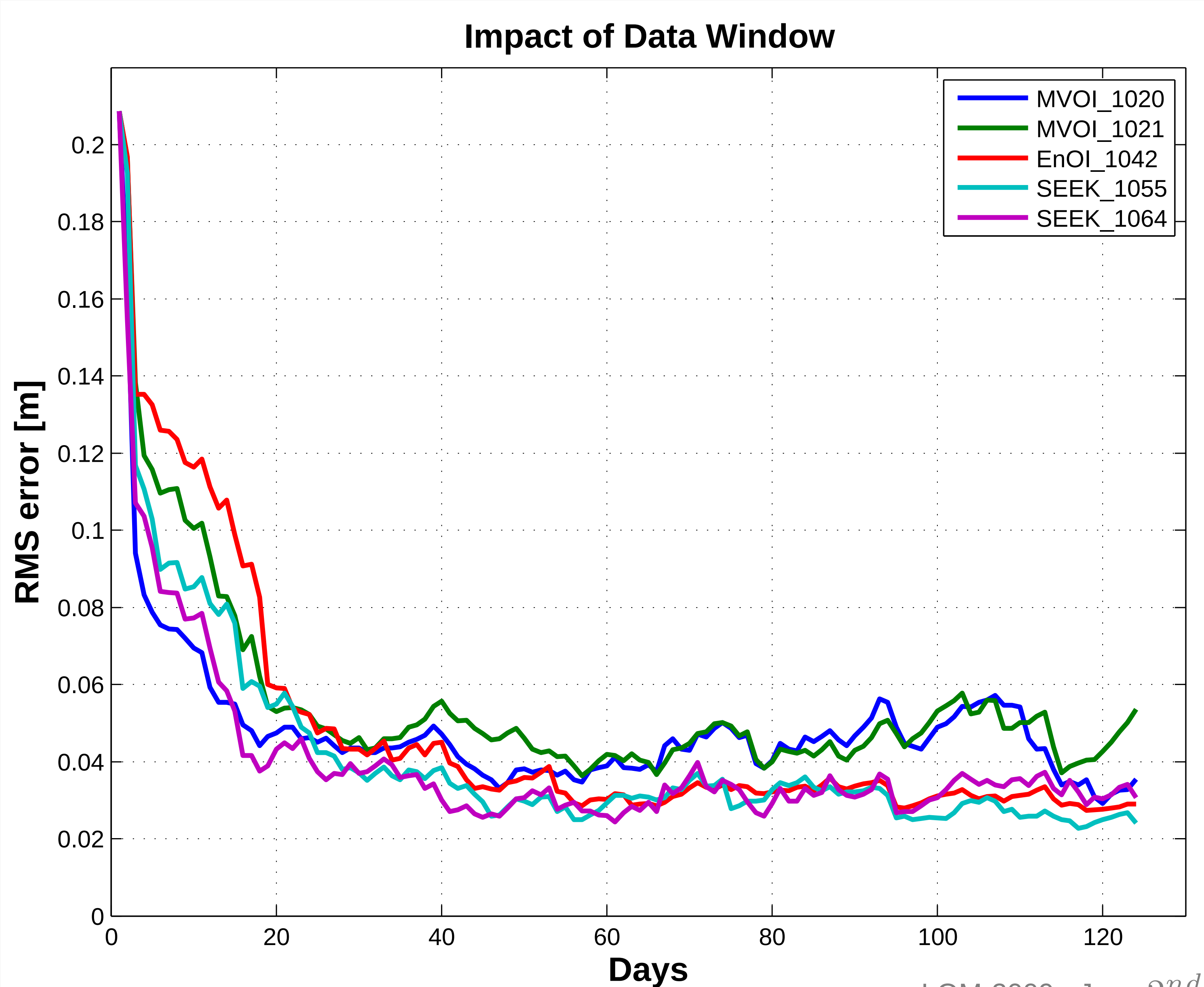
Twin Experiments: Ensemble Size and Covariance Rank



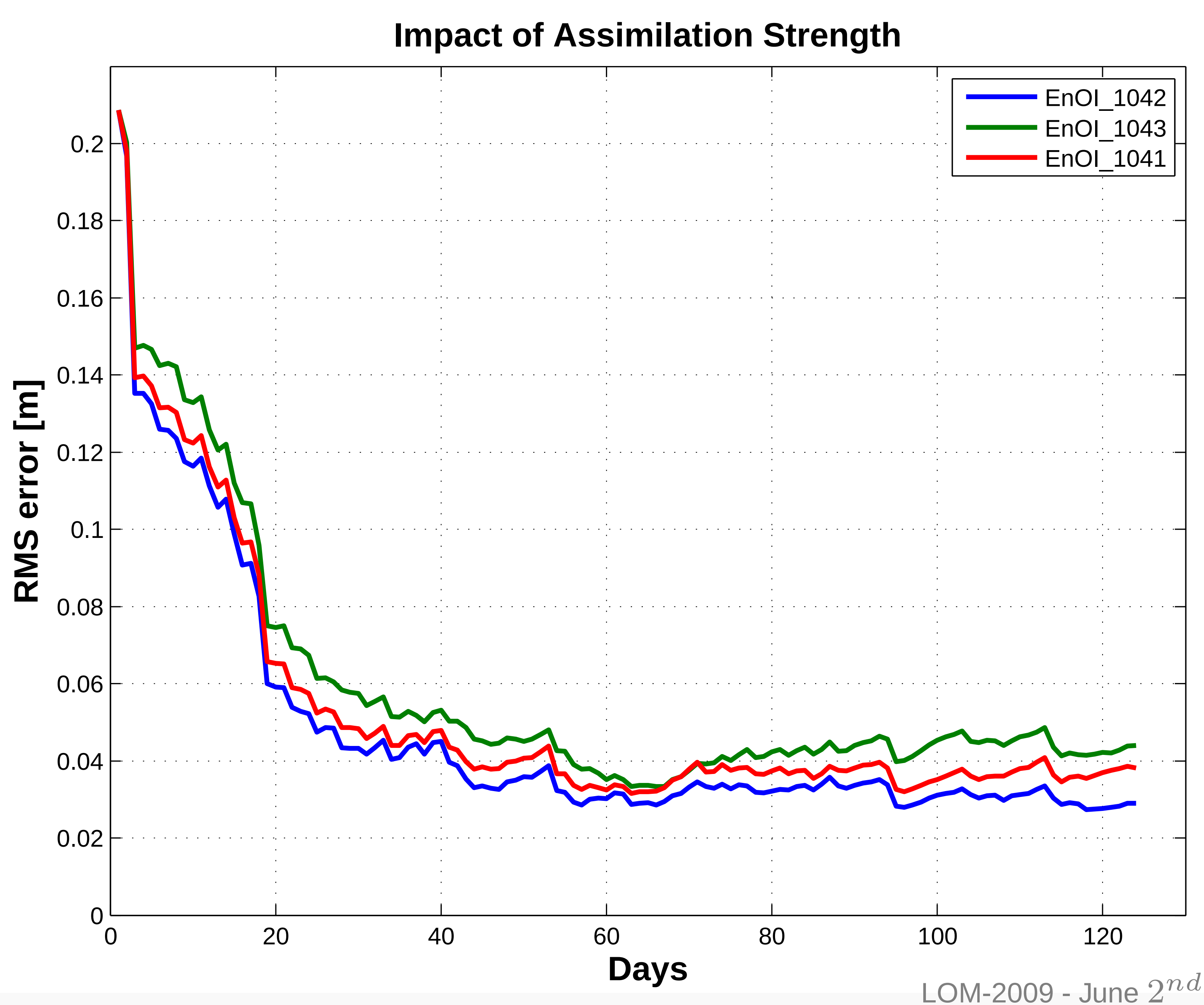
Twin Experiments: Localization



Twin Experiments: Data Window



Twin Experiments: Assimilation Strength



Computational Costs

Scheme	Scalability	Memory Req.	Typical Wall Clock/update
MVOI	$O(m^3/V)$	–	22 min
EnOI	$O(m^3)$	nN	31 min
EnROIF	–	sN	8 min
SEEK	$O(r^3)$	nN	14 min

n ==> no of members in the ensemble
N ==> size of the forecast model
s ==> size of the MRF neighborhood
r ==> covariance rank (== no of members in our experiments)
v == no of analysis volumes in MVOI
all experiments were done on 8 core/2.6 GHz Intel/12GB RAM)

Summary

- All schemes are able to reduce errors in both observed and unobserved variables.
- 3D multivariate (MVOI, EnOI, SEEK) are clearly better than decoupled 2D scheme used currently in EnROIF
- Best performance is obtained when all state variables are used in the estimation space
- Subsurface correction based on correlations(EnOI and SEEK) are better balanced and seem more effective than the dynamical method(MVOI)
- The nearly identical performance of EnOI and SEEK is both expected and reassuring
- Several parameters (ensemble size, covariance rank, radius of localization etc.) have to be explored for a given problem to find a good solution that is also cost effective.