Data Assimilation, Adaptive Sampling and Adaptive Modeling for Coupled Physical-Biological-Ecosystem Ocean Dynamics and Predictions

Pierre F.J. Lermusiaux

Division of Engineering and Applied Sciences, Harvard University

Main collaborators on this topic

HU: **A. Robinson, P. Haley, W. Leslie, J. McCarthy, O. Logoutov and X. Liang**

AOSN2-MURI: **N. Leonard, J. Marsden, F. Lekien, J. Bellingham, Y. Chao, S. Majumdar**

Univ. Trieste/Venice: **G. Cossarini, C. Solidoro, R. Pastres**

http://www.deas.harvard.edu/~pierrel

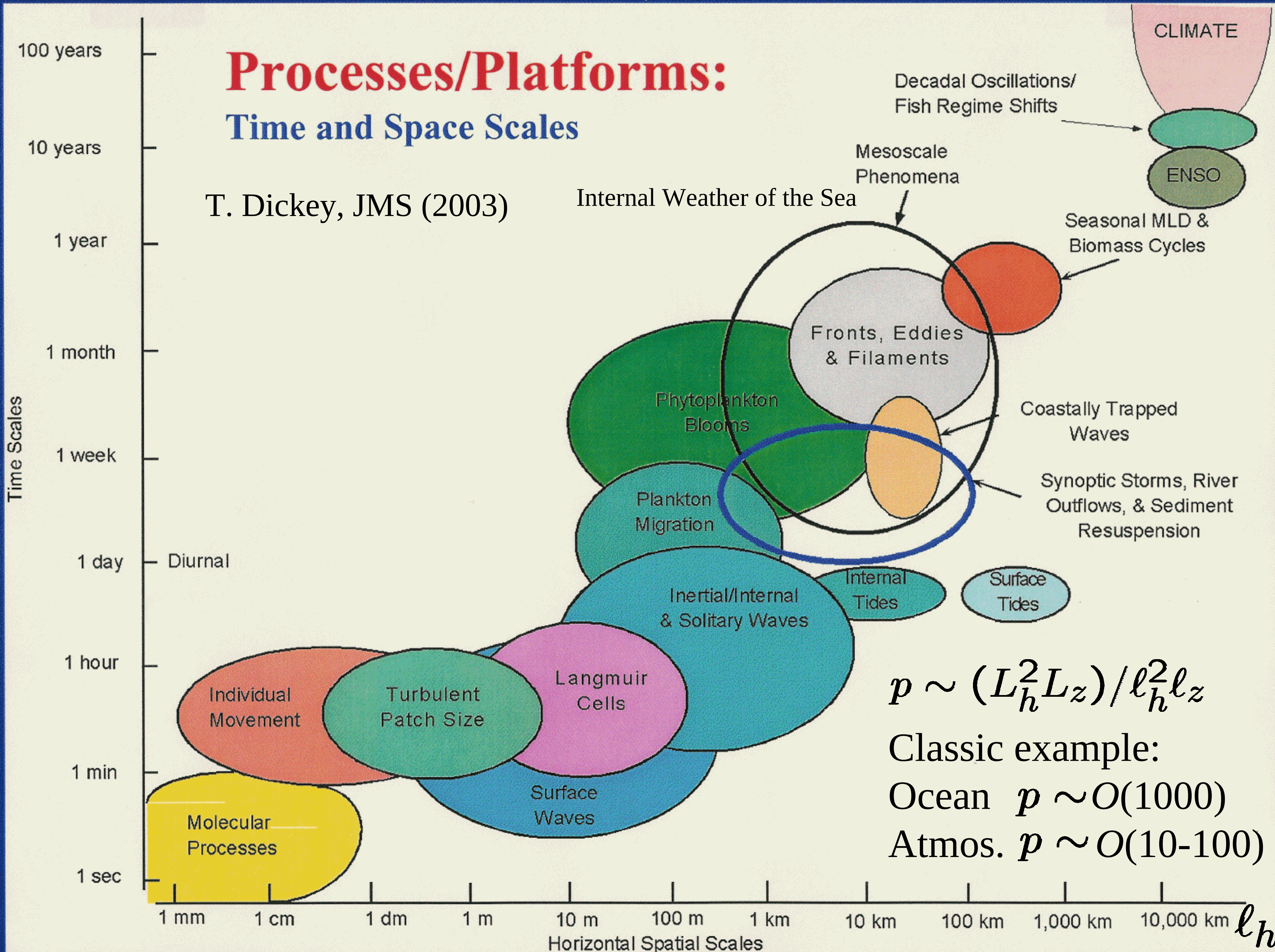
1. **Interdisciplinary Ocean Science and Data Assimilation**
2. **Different Methods and their Applications**
3. **Error Subspace Statistical Estimation (ESSE)**
   - **Assimilation of Water Quality data in a 3D Ecosystem Model of the Lagoon of Venice**
   - **Smoothing and Biogeochemical Dominant Dynamical Balances (Mass Bay/ Monterey Bay)**
   - **Error Forecasting, Adaptive Sampling and Adaptive Modeling in Monterey Bay**
4. **Conclusions**

Halifax- NPCDS Workshop, August 18, 2005

Processes/Platforms:

Time and Space Scales

T. Dickey, JMS (2003)



$$p \sim (L_h^2 L_z)/\ell_h^2 \ell_z$$

Classic example:

Ocean $p \sim O(1000)$

Atmos. $p \sim O(10\text{-}100)$

ℓ_h

OCEANIC FOOD WEB:
Multiple trophic relations

e.g. leading to adult herring (arrows show energy flow)

- **Interactions of Physical and Biological/Chemical Dynamical Processes, e.g.**
  - **Primary Productivity**
  - **The Biological Pump and its Role in the Changing Global Carbon Cycle**

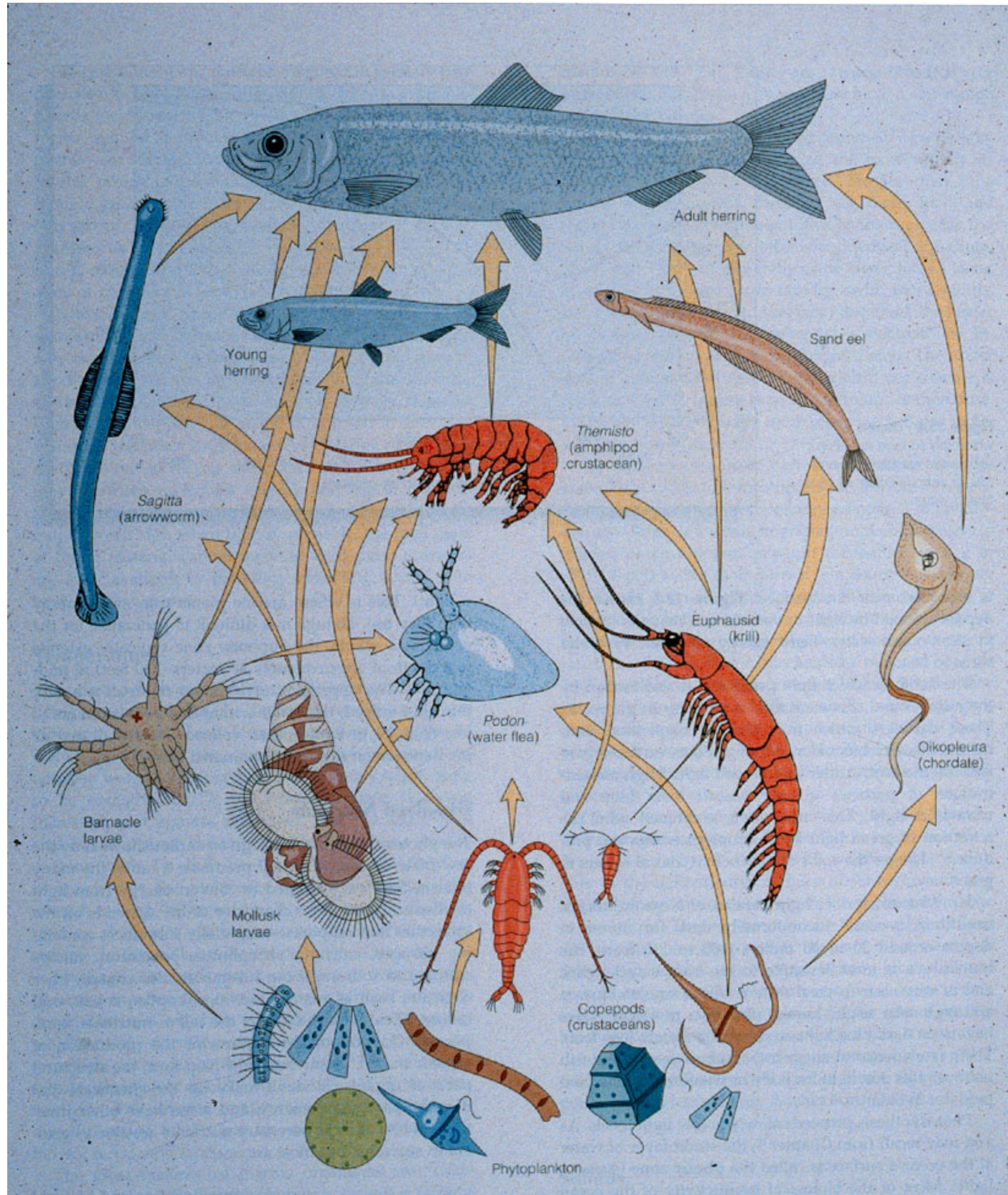


Figure 13.7 A food web, illustrating the major trophic relationships leading to an adult herring. The arrows show the direction of energy flow.

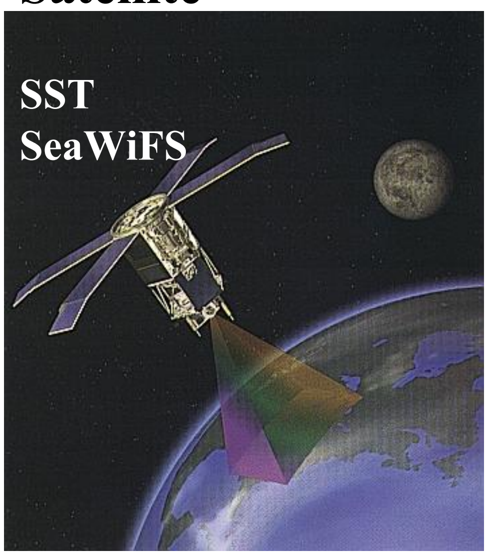
Physical and Multidisciplinary Observations

AUV

Aircraft

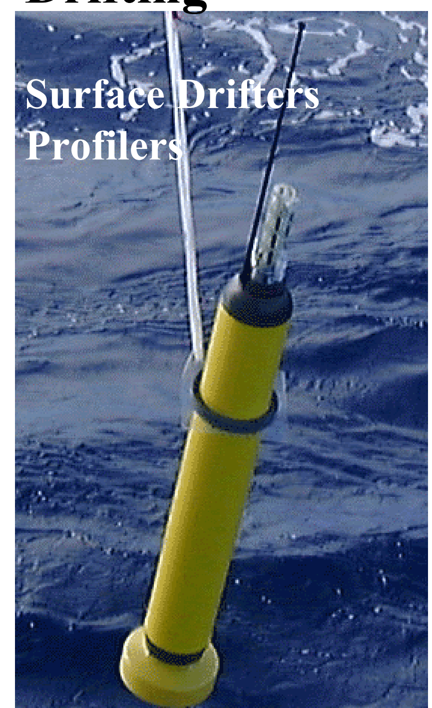
Ships

Satellite



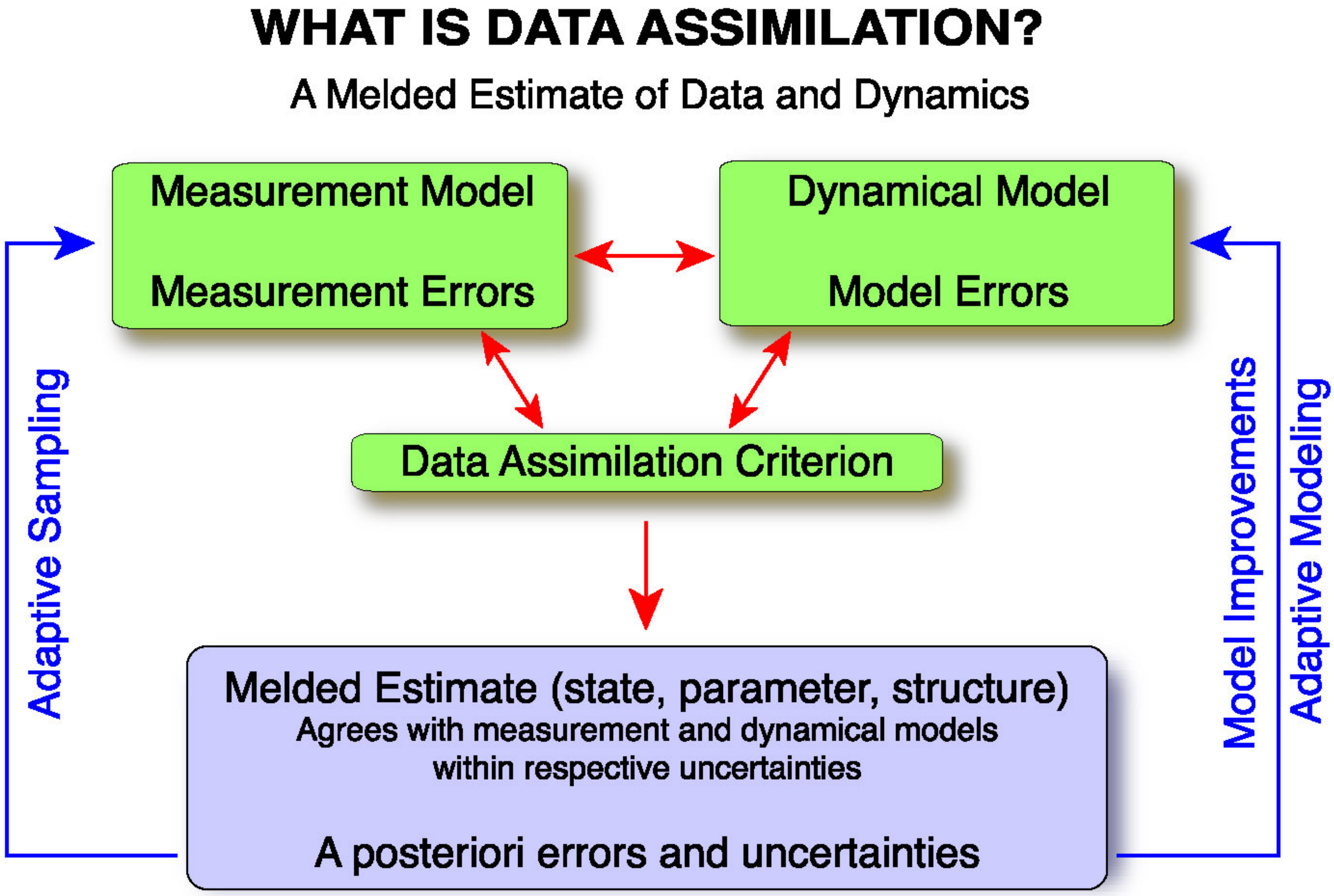
Moored/Fixed

Drifting



Interdisciplinary Ocean Science and Data Assimilation

- From observations and a priori conservation laws, fundamental ocean science formulates models, usually differential equations, which aim to explain the dynamics of the sea phenomena under study
- Estimation of four-dimensional fields and parameters in the ocean is challenging
  - Multiple interactive scales in space and time (e.g. ocean weather: 1-100km, 1-10days)
  - Large domains (e.g. 10-1000 km during 10-1000 days: 10^5-10^7 model points)
  - Limited ocean data (10^4-10^5 data points)
- Coupled physical-biogeochemical-ecosystem-optical-acoustical modeling and estimations initiated
- Substantial advances require interdisciplinary data assimilation:
  - Quantitative combinations of data and models, in accord with uncertainties
  - Model reductions, simplifications and understanding

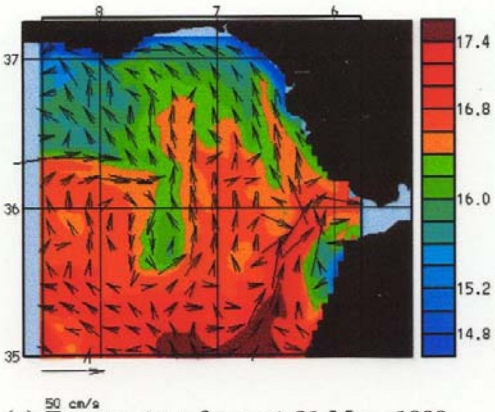


e.g. Robinson A.R., P.F.J. Lermusiaux and N.Q. Sloan, III (1998). *Data Assimilation*. The Sea, Vol. 10.

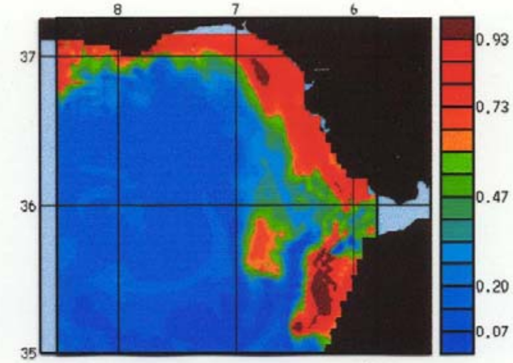
Robinson A.R. and P.F.J. Lermusiaux (2002). *DA for physical-biological interactions*. The Sea, Vol.12.

Gulf of Cadiz, 1998

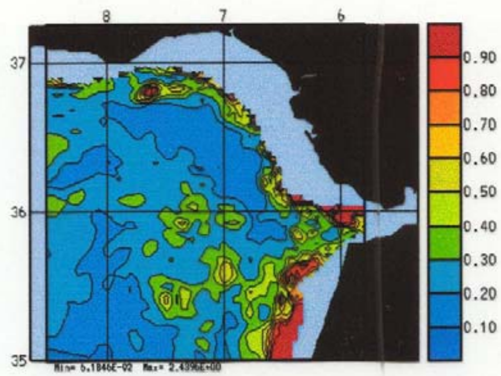
Real-Time HOPS/ESSE physical-ecosystem predictions



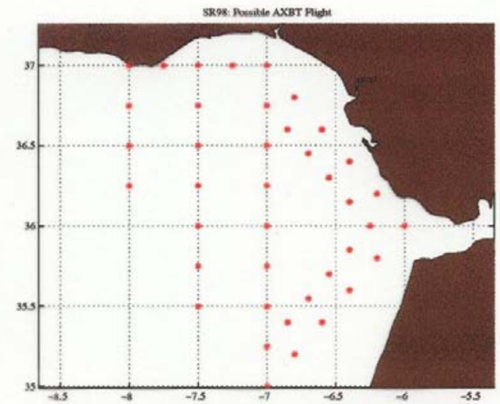
(a) Temperature forecast 21 Mar. 1998



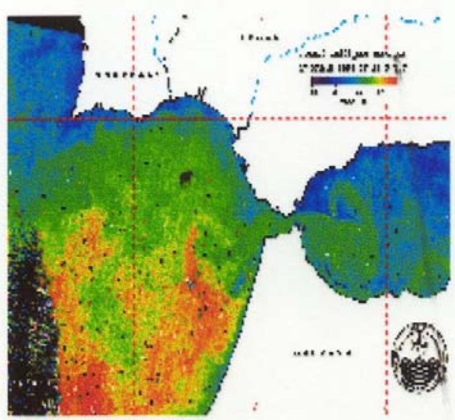
(b) Chlorophyll forecast 21 Mar. 1998



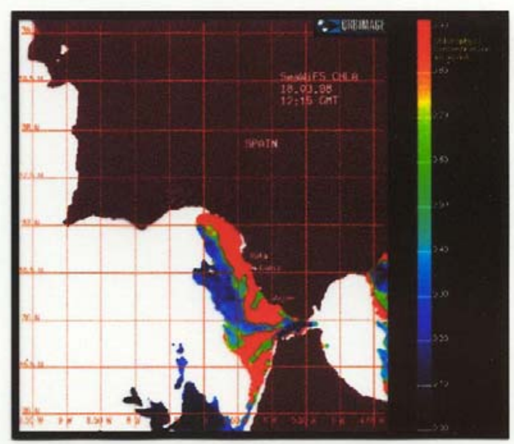
(c) Temperature forecast error (100m)



(d) Adaptively designed sampling



(e) SST from AVHRR 17 Mar. 1998



(f) Chlorophyll from SeaWiFS 18 Mar. 1998

Generic Data Assimilation Problem

Dynamical models:

$$d\phi_i + \mathbf{u} \cdot \nabla \phi_i dt - \nabla(K_i \nabla \phi_i) dt = B_i(\phi_1, \ldots, \phi_i, \ldots, \phi_n) dt + d\eta_i \qquad (i = 1, \cdots, n)$$

e.g. i= u, v, T, $\cdots$, ZOO, $\cdots$, p

Parameter equations:

$$dP_\ell = C_\ell(\phi_1, \ldots, \phi_i, \ldots, \phi_n) dt + d\zeta_\ell \qquad (\ell = 1, \cdots, p)$$

e.g. $P_\ell = \{ K_i, R_i, \cdots \}$

Measurement models:

$$y_j = \mathcal{H}_j(\phi_1, \cdots, \phi_i, \cdots, \phi_n) + \epsilon_j \qquad (j = 1, \cdots, m)$$

e.g. $y_j = \{ XBT_j, Fluo_j, SSH_j, CODAR_j \}$

Assimilation criterion:

$$\min_{\phi_i, P_\ell} \quad J(d\eta_i\ ,\ d\zeta_\ell\ ,\ \epsilon_j\ ,\ q_\eta\ ,\ q_\zeta\ ,\ q_\epsilon)$$

CLASSES OF DATA ASSIMILATION SCHEMES

	Error Evol.	Criterion
• **Estimation Theory (Filtering and Smoothing)**		
1. Direct Insertion, Blending, Nudging	- Linear	
2. Optimal interpolation	- Linear	LS
3. Kalman filter/smoother	- Linear	LS
4. Bayesian estimation (Fokker-Plank equations)	- Non-lin.	Non-LS
5. Ensemble/Monte-Carlo methods	- Non-lin.	LS/Non-LS
6. Error-subspace/Reduced-order methods: Square-root filters, e.g. SEEK	- (Non)-Lin.	LS
7. Error Subspace Statistical Estimation (ESSE): 5 and 6	-Non-lin.	LS/Non-LS
• **Control Theory/Calculus of Variations (Smoothing)**		
1. “Adjoint methods” (+ descent)	- Lin. adj.	LS
2. Generalized inverse (e.g. Representer method + descent)	- Lin. adj.	LS
• **Optimization Theory (Direct local/global smoothing)**		
1. Descent methods (Conjugate gradient, Quasi-Newton, etc)	- Lin	LS/Non-LS
2. Simulated annealing, Genetic algorithms	- Non-lin.	LS/Non-LS
• **Hybrid Schemes**		
• Combinations of the above		

Control Theory

(Calculus of Variation Approach, Variational Assimilation)

Variational Adjoint Method

$$\min_{\hat{\boldsymbol{\psi}}_k} J_N = \boldsymbol{\epsilon}_0^T \mathbf{P}_0^{-1} \boldsymbol{\epsilon}_0 + \sum_{k=1}^{N-1} \mathbf{v}_k^T \mathbf{R}_k^{-1} \mathbf{v}_k + \sum_{k=1}^{N} 2\boldsymbol{\lambda}_{k-1}^T \mathbf{w}_{k-1} \quad (19)$$

Dynamical model $\hat{\boldsymbol{\psi}}_k = \mathbf{A}_{k-1} \hat{\boldsymbol{\psi}}_{k-1} \quad k = 1, \ldots, N$ (20)

Initial condition $\hat{\boldsymbol{\psi}}_0 = \boldsymbol{\Psi}_0 + \mathbf{P}_0 \mathbf{A}_0^T \boldsymbol{\lambda}_0$ (21)

Adjoint model $\boldsymbol{\lambda}_{k-1} = \mathbf{A}_k^T \boldsymbol{\lambda}_k + \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{H}_k \hat{\boldsymbol{\psi}}_k)$ (22)

$k = 1, \ldots, N-1$

Initial condition $\boldsymbol{\lambda}_{N-1} = 0$ (23)

Model is strong constraint: variational adjoint method assumes no model error

Georges Bank
(NW Atlantic)

Estimate full biological source term (RHS) from data (Pseudicalanus spp.)

McGillicuddy et al (1998)

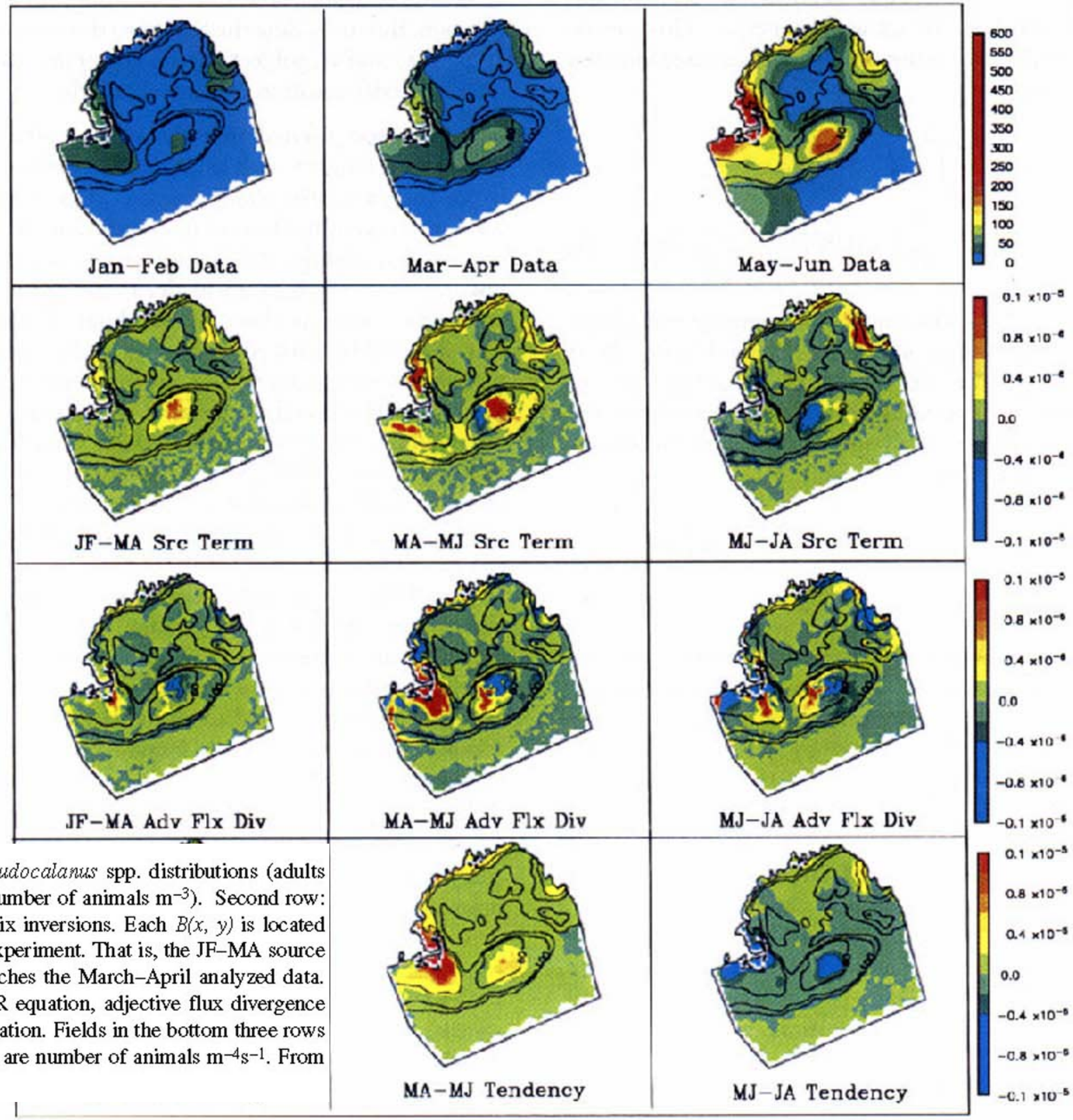


Figure 12.15 Top row: bimonthly climatological *Pseudocalanus* spp. distributions (adults only) objectively analyzed from the MARMAP data (number of animals m^{-3}). Second row: three source terms $B(x, y)$ resulting from three of the six inversions. Each $B(x, y)$ is located directly below the analyzed data used to initialize the experiment. That is, the JF–MA source term results in a forward model integration which matches the March–April analyzed data. Last two rows: two of the remaining terms in the ADR equation, adjective flux divergence and overall tendency, averaged over the period of integration. Fields in the bottom three rows have been normalized to the bottom depth, so the units are number of animals $m^{-4}s^{-1}$. From D. J. McGillicuddy, Jr. et al. (1998)

Direct Minimization Methods

(descent methods, simulated annealing, genetic algorithms, etc)

Comparisons of methods for the estimation of biogeochemical parameters

Vallino, JMR (2000)

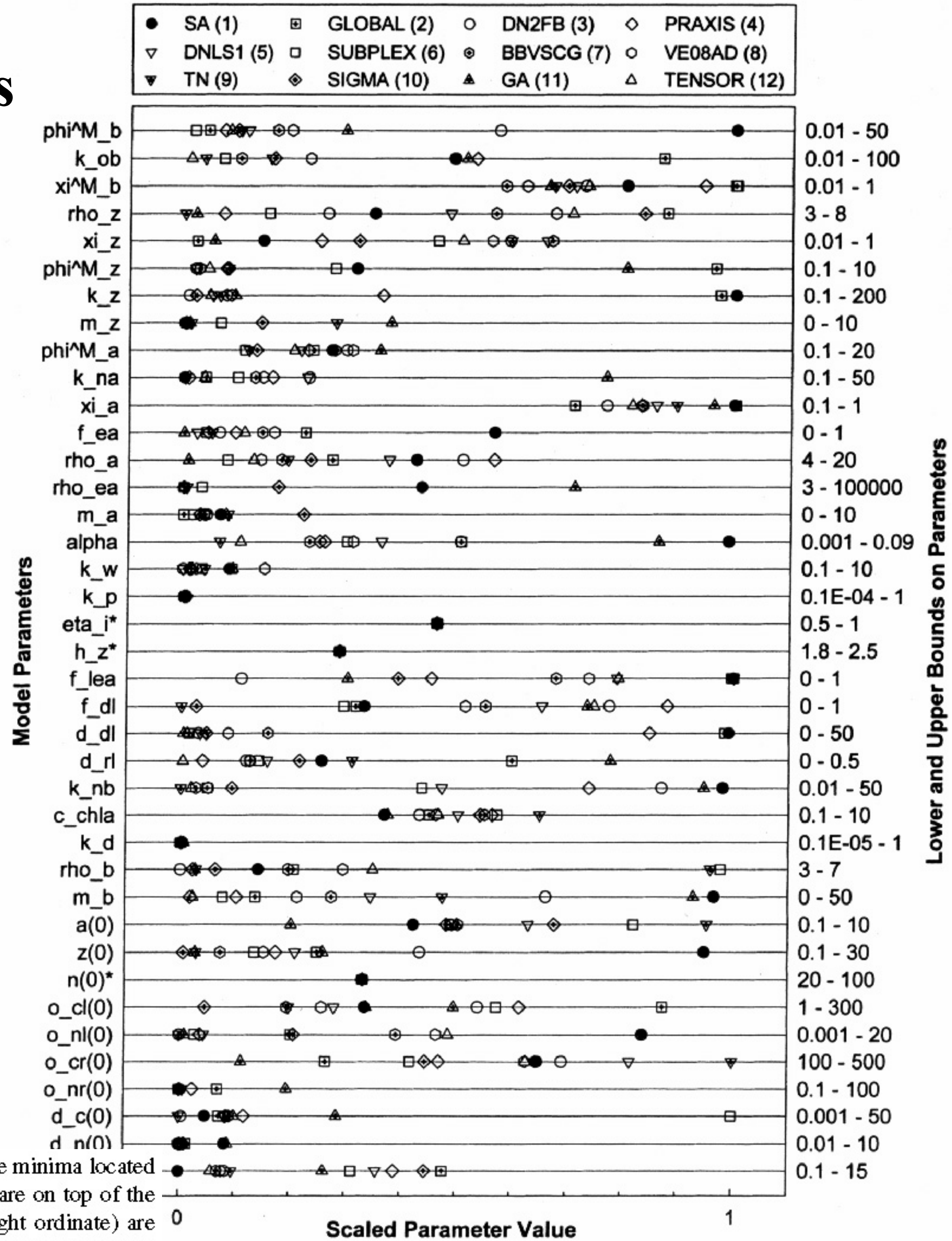


Figure 12.8 Optimized, scaled (0 to 1) parameter values associated with each of the minima located by the twelve optimization routines (abbreviated routine names and their symbols are on top of the figure). Model parameters (left ordinate) and their absolute parameter bounds (right ordinate) are described in Vallino (2000). Parameters marked with an asterisk (left ordinate) were held constant during the data assimilation. From Vallino (2000).

Data Assimilation via ESSE

Table 1. Filtering/Smoothing via ESSE: Continuous-Discrete Problem Statement

Dynamical Model:	$d\widehat{\mathbf{x}} = \mathcal{M}(\widehat{\mathbf{x}})\,dt + d\widehat{\boldsymbol{\eta}}\,,\;$ with $\widehat{\mathbf{x}}(\mathbf{r}_0, t_0) = \widehat{\mathbf{x}}_0 + \widehat{\mathbf{n}}(0)\,.$
Measurement Model:	$\mathbf{y}^o_k = \mathcal{H}(\mathbf{x}_k) + \widehat{\boldsymbol{\epsilon}}_k\,.$
Estimation Criterion:	
Estimate Error Subpace:	$\left\{ \text{Find } \mathbf{P}^p_k = \mathbf{E}_k \boldsymbol{\Pi}_k \mathbf{E}^T_k \text{ with rank}(\mathbf{E}_k) = p \;\middle|\; \min_{\boldsymbol{\Pi}_k, \mathbf{E}_k} \lVert \mathbf{P}_k - \mathbf{P}^p_k \rVert \right\}$
Estimate State by Min. Err. Var. in ES:	$\left\{ \text{Find } \widehat{\mathbf{x}}_k \;\middle|\; \min_{\widehat{\mathbf{x}}_k} J_k = \text{tr}\left[\mathbf{P}^p_k(+)\right] \text{ using } [\mathbf{y}^o_0, \ldots, \mathbf{y}^o_k / \mathbf{y}^o_N] \right\}$

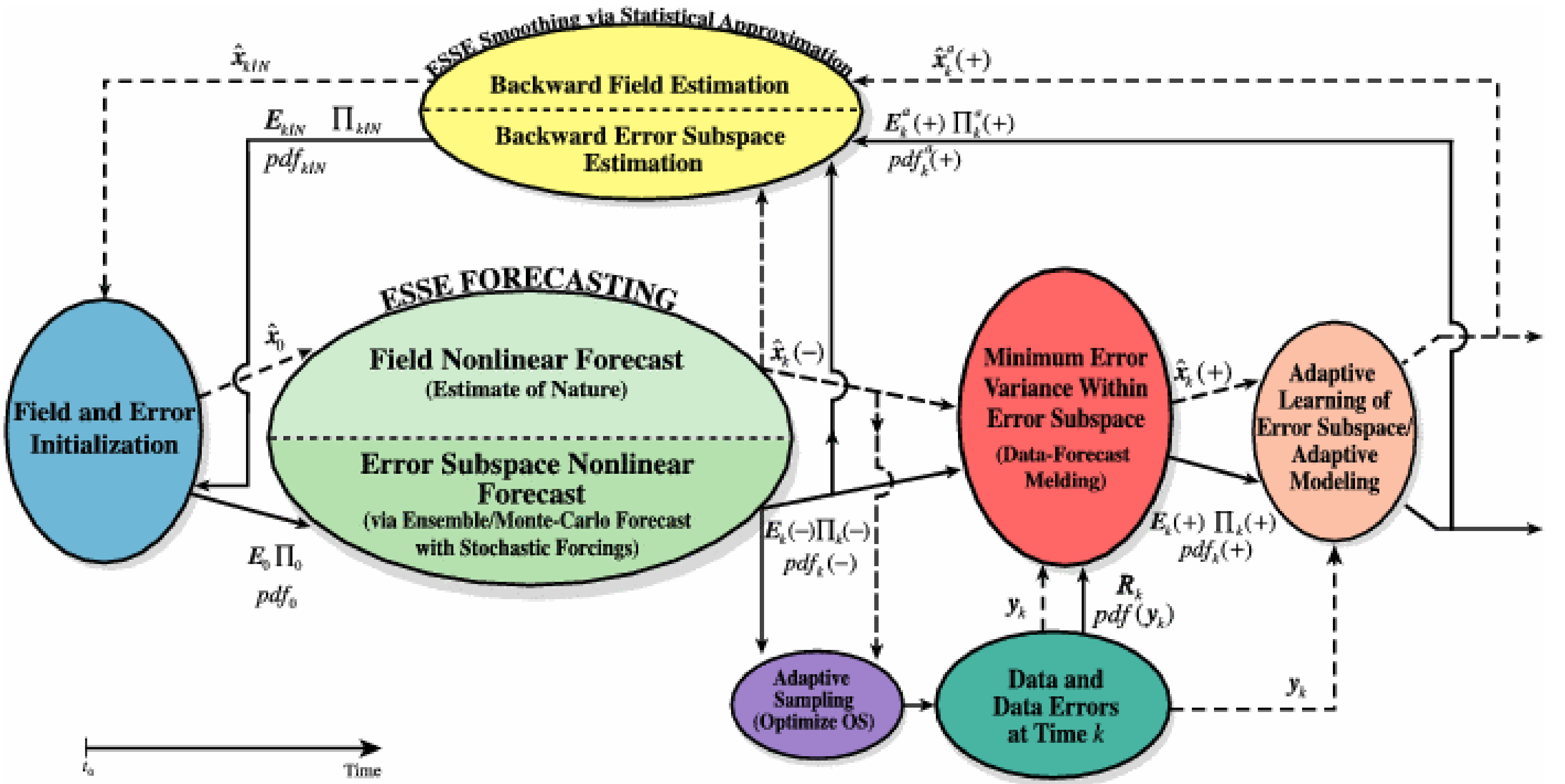
◦ **Optimal error space reduction and Min. Err. Var. combined:**

"Estimate the ocean evolution by minimizing the largest (most energetic) expected errors, in agreement with the full dynamical model and measurement model (data) constraints, and their respective uncertainties."

o **Linked to POD/Polynomial Chaos, but with time-varying error KL basis:**

$$\mathbf{x}(x, t, \theta) = \bar{\mathbf{x}}(x, t) + \sum_{i=1}^{M} \sqrt{\lambda_i}\;\; \phi^s_u(\mathbf{x}, t)\; \zeta_i(\theta)$$

Error Subspace Statistical Estimation (ESSE)



- Uncertainty forecasts (with dynamic error subspace, error learning)
- Ensemble-based (with nonlinear and stochastic primitive eq. model (HOPS)
- Multivariate, non-homogeneous and non-isotropic Data Assimilation (DA)
- Consistent DA and adaptive sampling schemes
- Software: not tied to any model, but specifics currently tailored to HOPS

Nonlinear Dynamical State and Error Subspace Ensemble Forecast for Sequential Statistical Estimation

- **Cond. Mean Estimates:**

  Central forecast: $\widehat{\mathbf{x}}^{\mathrm{cf}}_{k+1}(-) \;\Big|\; d\widehat{\mathbf{x}} = \mathcal{M}(\widehat{\mathbf{x}}, t)\, dt \,, \quad \text{with } \widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k(+) \,.$

  Ensemble mean: $\widehat{\mathbf{x}}^{\mathrm{em}}_{k+1}(-) \doteq \mathcal{E}^q\{\widehat{\mathbf{x}}^j_{k+1}(-)\} \quad .$

- **ES Initial Conditions:** $\widehat{\mathbf{x}}^j_k(+) = \widehat{\mathbf{x}}_k(+) + \mathbf{E}_k(+)\,\boldsymbol{\pi}^j_k(+) + n^j_k \,, \quad j = 1, ..., q \,,$

  with either,

  $$\boldsymbol{\pi}^j_k(+) = \boldsymbol{\Pi}^{\frac{1}{2}}_k(+)\,\mathbf{u}^j \,,$$

  $$\boldsymbol{\pi}^j_k(+) = \boldsymbol{\Pi}^{\frac{1}{2}}_k(+)\,\mathbf{u}^j \,, \text{ with dynamical/dataconstraints, or,}$$

  $$\boldsymbol{\pi}^j_k(+) = \boldsymbol{\Sigma}(+)\left(\mathbf{V}^T_+\right)^j \,, \text{ with (32) or (33a–c),}$$

  where $\mathbf{u}$ is of zero mean and covariance $\mathbf{I}^p$.

- **Ensemble Forecast:** $\widehat{\mathbf{x}}^j_{k+1}(-) \;\Big|\; d\widehat{\mathbf{x}}^j = \mathcal{M}(\widehat{\mathbf{x}}^j, t)\, dt + d\mathbf{w} \,, \text{ with } \widehat{\mathbf{x}}^j_k = \widehat{\mathbf{x}}^j_k(+) \,, j = 1, ..., q \,.$

  where $\mathcal{E}\{d\mathbf{w}(t) d\mathbf{w}^T(t)\} \doteq \mathbf{Q}(t)\, dt \doteq \mathbf{B}(t)\,\mathbf{B}^T(t)\, dt$ and $\mathbf{B}(t) \in \mathrm{R}^{n \times r}$.

- **ES Forecast:** $\mathbf{M}_{k+1}(-) = \left[\widehat{\mathbf{x}}^j_{k+1}(-) - \widehat{\mathbf{x}}_{k+1}(-)\right] \,, \quad j = 1, \ldots, q \,,$

  decomposed into, $\boldsymbol{\Pi}_{k+1}(-) \doteq \frac{1}{q}\boldsymbol{\Sigma}^2_{k+1}(-)$ and $\mathbf{E}_{k+1}(-)$ of rank $p \leq q$, where,

  $$\left\{\boldsymbol{\Sigma}_{k+1}(-)\,, \mathbf{E}_{k+1}(-) \;\Big|\; \mathrm{SVD}_p(\mathbf{M}_{k+1}(-)) = \mathbf{E}_{k+1}(-)\boldsymbol{\Sigma}_{k+1}(-)\mathbf{V}^T_{k+1}(-)\right\}$$

  and the operator $\mathrm{SVD}_p(\cdot)$ selects the rank-p SVD.

- **Convergence Crit.:**

  $$\rho = \frac{\sum_{i=1}^{k} \sigma_i(\boldsymbol{\Pi}^{\frac{1}{2}}\mathbf{E}^T\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Pi}}^{\frac{1}{2}})}{\sum_{i=1}^{\tilde{p}} \sigma_i(\widetilde{\boldsymbol{\Pi}})} \geq \alpha \,,$$

  where α is a chosen convergence limit ($1 - \epsilon \leq \alpha \leq 1$),
  $k = \min\,(\tilde{p}, p)$ and
  $\sigma_i(\cdot)$ selects the singular value number i.

STOCHASTIC FORCING MODEL: Sub-grid-scales

II. 3d Random Noise, Exponentially Decorrelated in Time and 2-Grid Point Decorrelated in Space

$$\mathbf{d}\boldsymbol{\psi}^t = \mathbf{f}^{PE}\left(\boldsymbol{\psi}^t, t\right)\, dt + \mathbf{B}^{fc}(t)\, \mathbf{d}\tilde{\mathbf{w}}^c \ . \tag{86a}$$

$$\mathbf{d}\tilde{\mathbf{w}}^c = -\boldsymbol{\beta}^c \tilde{\mathbf{w}}^c\, dt + \mathbf{dw}^c \ , \tag{86b}$$

where symbols denote the:

- **discrete-space PE state vector:** $\boldsymbol{\psi} = (\hat{\mathbf{u}}, \hat{\mathbf{v}}, \mathbf{T}, \mathbf{S}, \mathbf{p})^T \in \mathbb{R}^n$
- **coarse 3d white noise:** $\mathbf{w}_k^c$
- **Coarse 3d Gauss-Markov process:** $\tilde{\mathbf{w}}_k^c$,

  i.e. $\mathbf{dw}^c = (\mathbf{dw}_{\hat{u}}^c, \mathbf{dw}_{\hat{v}}^c, \mathbf{dw}_T^c, \mathbf{dw}_S^c, \mathbf{dw}_{\psi}^c)^T$
- **PE dynamical model operator:** $\mathbf{f}^{PE}(\cdot, t)$
- **linear extrapolation matrix, from coarse to fine state:** $\mathbf{B}^{fc}(t)$

Data-Forecast Melding:
Minimum Error Variance within Error Subspace

TRUNCATED Minimum Sample ES Variance, Linear Update (subscript k omitted)

Dynamical State Update: $\widehat{\mathbf{x}}_{(+)} = \widehat{\mathbf{x}}_{(-)} + \mathbf{K}^p \left(\mathbf{y}^{\mathrm{o}} - \mathcal{H}\big(\widehat{\mathbf{x}}_{(-)}\big)\right)$.

Sample ES Optimal Gain: $\mathbf{K}^p = \mathbf{E}_- \boldsymbol{\Pi}_{(-)} \mathbf{H}^{p^T} \left(\mathbf{H}^p \boldsymbol{\Pi}_{(-)} \mathbf{H}^{p^T} + \mathbf{R}\right)^{-1}$, where $\mathbf{H}^p \doteq \mathbf{H}\mathbf{E}_-$.

Sample ES Cov. Update: $\mathbf{L}\boldsymbol{\Pi}_{(+)}\mathbf{L}^T = \boldsymbol{\Pi}_{(-)} - \boldsymbol{\Pi}_{(-)} \mathbf{H}^{p^T} \left(\mathbf{H}^p \boldsymbol{\Pi}_{(-)} \mathbf{H}^{p^T} + \mathbf{R}\right)^{-1} \mathbf{H}^p \boldsymbol{\Pi}_{(-)}$.
$\mathbf{E}_+ = \mathbf{E}_- \mathbf{L}$.

ADAPTIVE LEARNING of the Error Subspace (subscript k omitted)

$$\widehat{\mathbf{n}}_{(+)} = \mathbf{K}_{\mathrm{trc}} \left(\mathbf{y}^{\mathrm{o}} - \mathcal{H}\big(\widehat{\mathbf{x}}_{(+)}\big)\right) ,$$

$$\mathbf{K}_{\mathrm{trc}} = \mathbf{E}_{\mathrm{trc}(-)} \, \boldsymbol{\Pi}_{\mathrm{trc}(-)} \, \mathbf{H}_{\mathrm{trc}}^T \left(\mathbf{H}_{\mathrm{trc}} \, \boldsymbol{\Pi}_{\mathrm{trc}(-)} \, \mathbf{H}_{\mathrm{trc}}^T + \mathbf{R}\right)^{-1}, \text{ where } \mathbf{H}_{\mathrm{trc}} \doteq \mathbf{H}\mathbf{E}_{\mathrm{trc}(-)} .$$

$$\mathbf{E}_+^{\mathrm{a}} \boldsymbol{\Sigma}^{\mathrm{a}}{}_{(+)} \mathbf{V}_+^{\mathrm{a}\,T} = \mathrm{SVD}_{p+1}\left(\left[\mathbf{E}_+ \boldsymbol{\Sigma}_{(+)} \, , \, \widehat{\mathbf{n}}_{(+)}\right]\right) ,$$

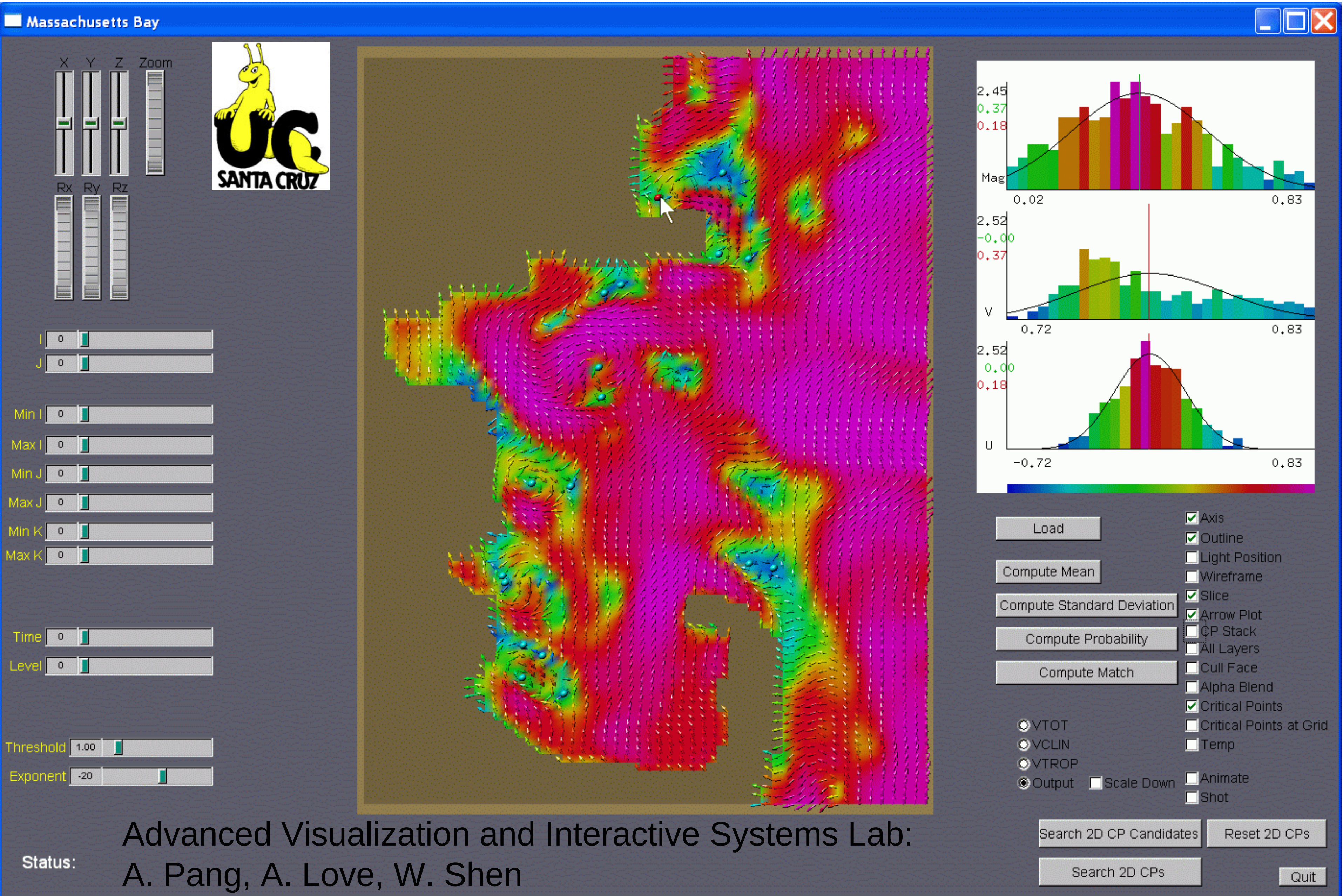
$$\boldsymbol{\Pi}^{\mathrm{a}}{}_{(+)} = \frac{1}{q+1} \boldsymbol{\Sigma}^{\mathrm{a}\,2}{}_{(+)} .$$

Ocean Regions and Experiments/Operations for which ESSE has been utilized in real-time

- Strait of Sicily (AIS96-RR96), Summer 1996
- Ionian Sea (RR97), Fall 1997
- Gulf of Cadiz (RR98), Spring 1998
- Massachusetts Bay (LOOPS), Fall 1998
- Georges Bank (AFMIS), Spring 2000
- Massachusetts Bay (ASCOT-01), Spring 2001
- Monterey Bay (AOSN-2), Summer 2003

For publications, email me or see http://www.deas.harvard.edu/~pierrel

Interactive Visualization and Targeting of pdfs



Advanced Visualization and Interactive Systems Lab:
A. Pang, A. Love, W. Shen

Assimilation of Water Quality data in a 3D Ecosystem Model of the Lagoon of Venice

Time/Space Scales considered

- Seasonal and lagoon-scale

Ecosystem model of the lagoon

- Trophic Diffusion Model (TDM)
- Released in 1987. Since then, evaluated against various regional data sets (model is here assumed to be a strong constraint)

Data

- MELa1: Monitoring of Environmental Lagoon program (MAV/CVN)
- Relatively rich data set on trophic state

Objectives

1. Sensitivity of the ecosystem to its initial conditions and boundary forcing: Time-scales of responses to driving forces/memory of the system
2. Ecosystem field estimation and seasonal dynamics based on ESSE data assimilation
3. Updated assessment of water quality

Cossarini, G., Lermusiaux, P. F. J. , Solidoro, C.

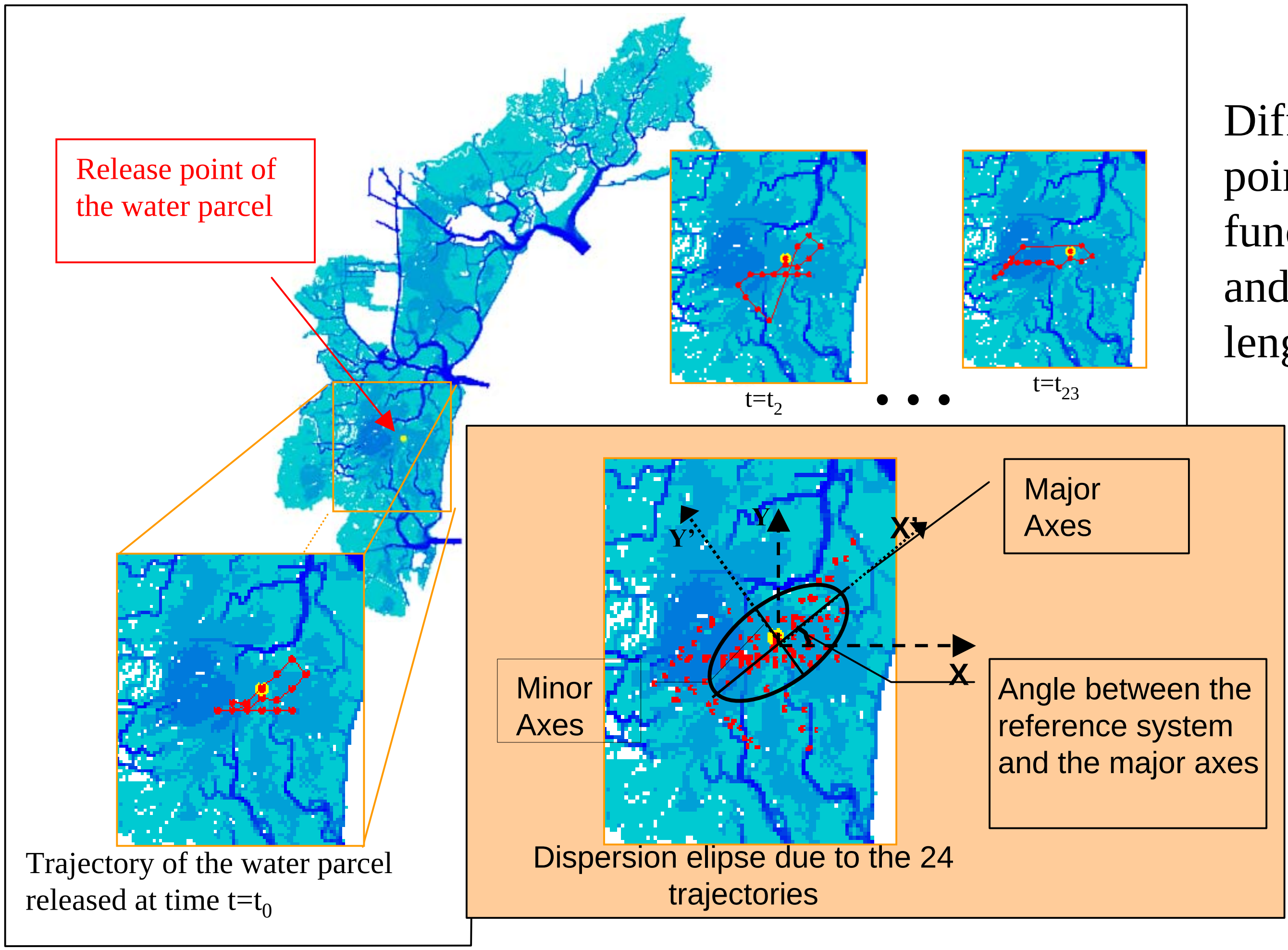
Physical Model: Transport formulation

1) time scale longer than tidal period

2) residual currents negligible

→ Anisotropic eddy diffusion

No advective term



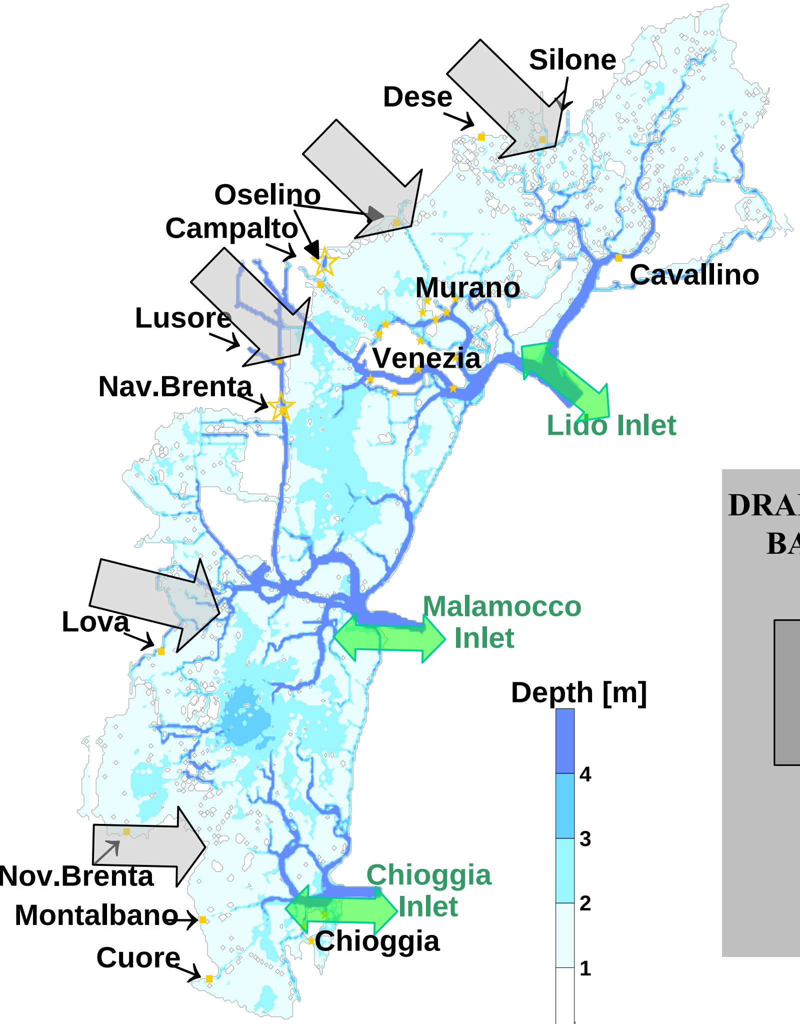
Diffusivity tensor of each point is computed as function of a mean velocity and a characteristic length (axes of ellipse)

each point has different mixing energy and principal mixing direction

TDM - Trophic Diffusive Model

domain and boundaries

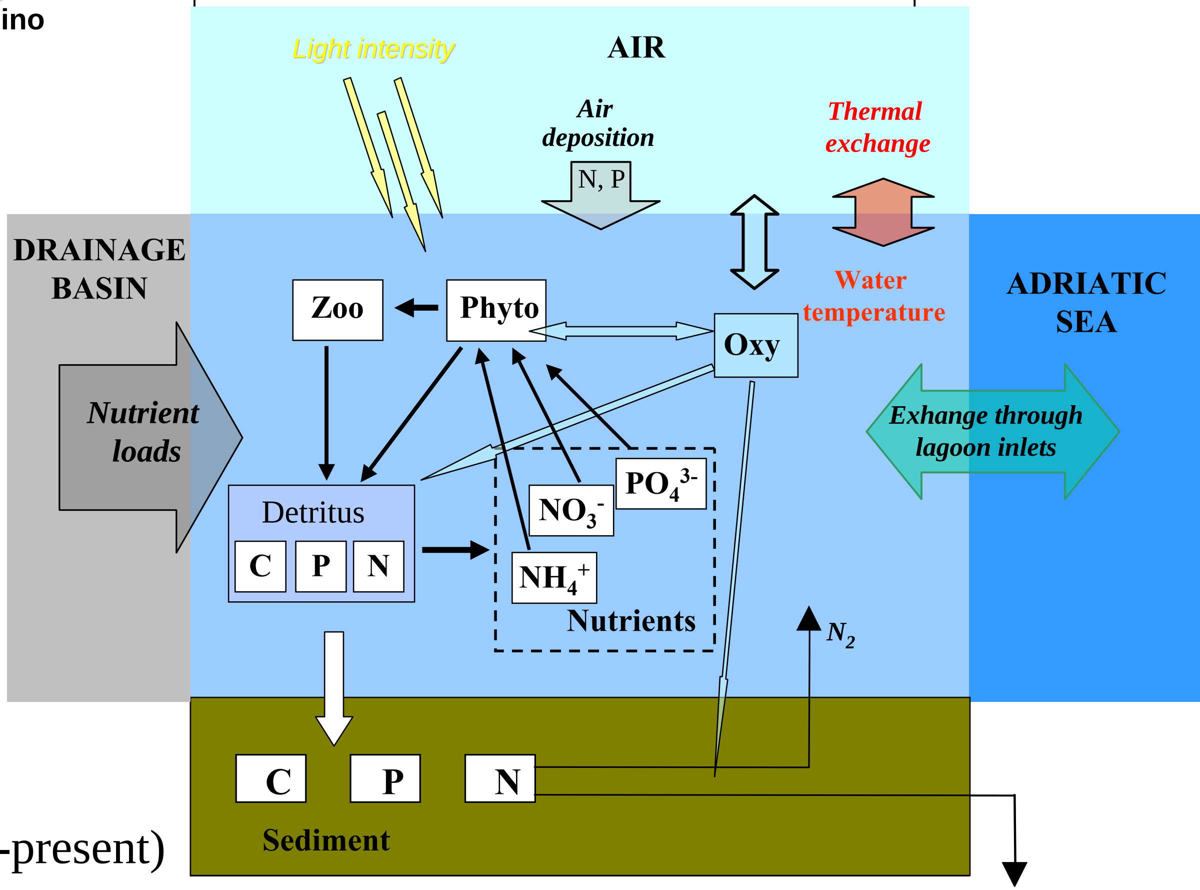
70000 cells of 100x100x1 m^3



3D coupled model

$$\frac{\partial \mathbf{C_i}}{\partial t} = w_{si} \frac{\partial \mathbf{C_i}}{\partial z} + \nabla \mathbf{K} \nabla \mathbf{C_i} + \mathbf{F}(C, T, I, ..)$$

Biological Model (12 s.v.)



Dejak, Pastres, Solidoro et al. (1987-present)

Data for the year 2001

MELa1-Monitoring Environmental Lagoon
(1th example of monitoring for the whole lagoon)

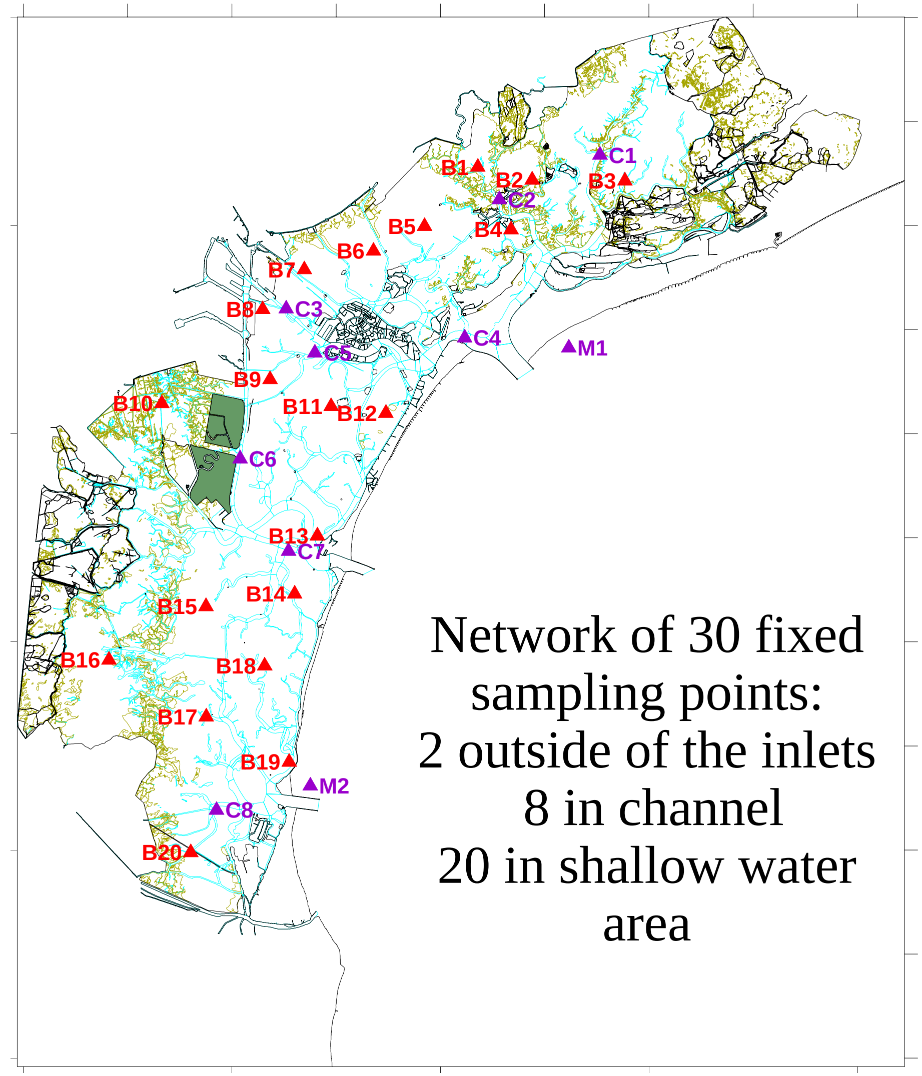
Temperatura
salinità
conducibilità
DO, pH, Eh
Torbidità
alc tot
sol sosp tot
chl a
N-NH4
N-NO3
N-NO2
TDN
P-PO4
TDP
TOC
POC
DOC
foepigm.

AS, Cu, Hg,Pb,
Zn,Cd, Cr, Ni

Monthly sampling campaigns from Sept 2000 to Aug 2003

12 monthly campaigns of 2001

Consorzio Venezia Nuova
Magistrato alle Acque
Venezia



Network of 30 fixed sampling points:
2 outside of the inlets
8 in channel
20 in shallow water area

Time-scales of driving forces/memory of the system
Sensitivity of the ecosystem to its initial conditions and boundary forcing

Field Initial Conditions (IC) affect simulations only for a limited time interval

Model evolution mainly driven by boundary conditions (nutrient loads from drainage basin, exchanges at the inlets): after some time system looses memory of IC

⇓

DA of in situ field data is effective only during this time interval (internal predictability limit)

This time interval has been estimated in 2 sets of experiments

experiments #1: Effects of re-initializations for each month of 2001 (simulation starting from monthly OA maps *vs.* reference simulation)

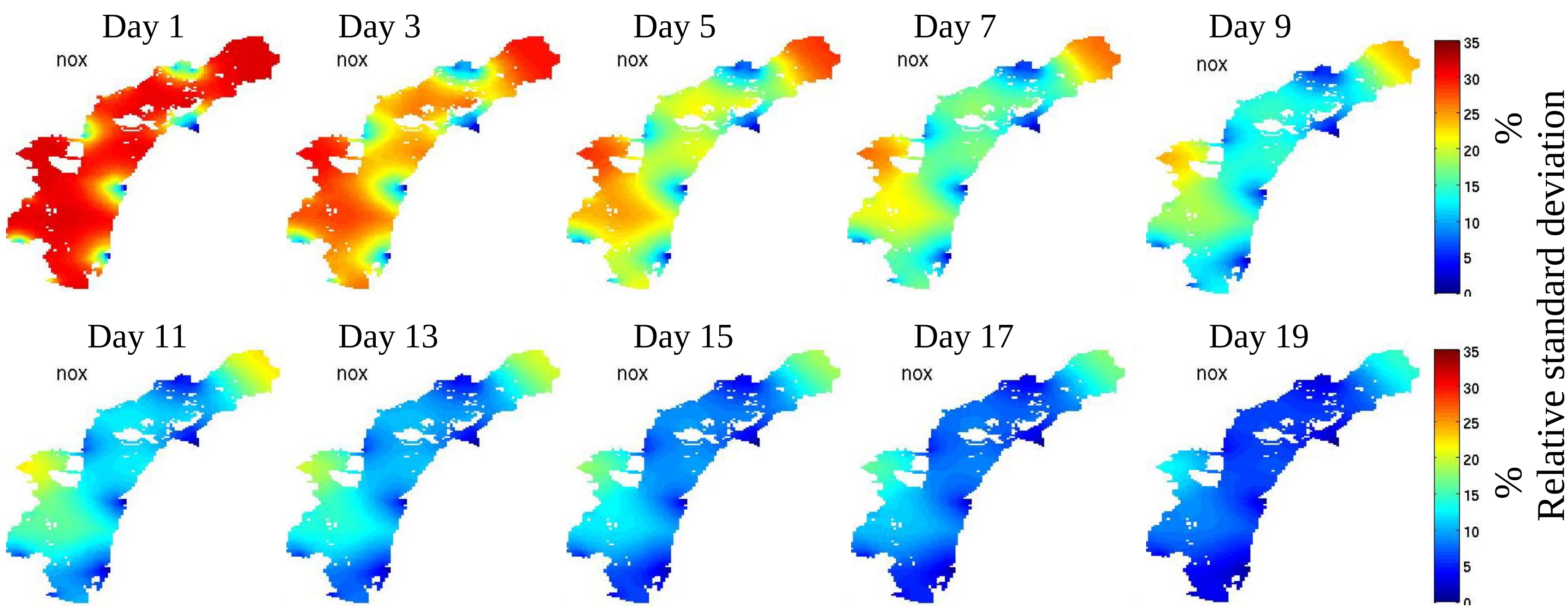
experiments #2 : Evolving an ensemble of simulations obtained by stochastic initialization at the DA time event (for each month of 2001).

Experiments #2

Analysis of ensemble of simulations obtained by stochastic initialization

$$IC^{j} = IC^{ref} \cdot w \cdot \sigma \qquad w = N(0,1) \qquad \sigma = 30\%$$

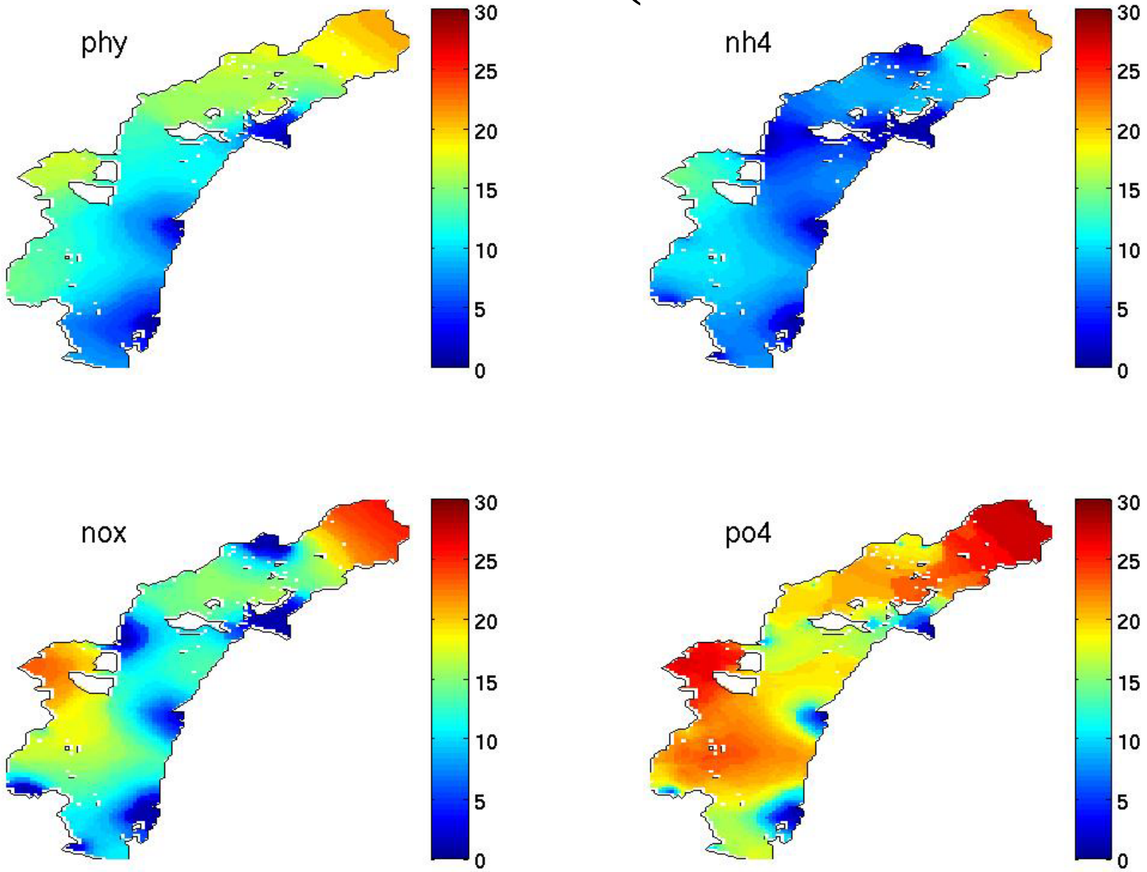
Sequence of maps of the standard deviation of the ensemble



Criteria: stop when standard deviation is halved

Experiments #2: Results

Day at which the std of the ensemble is halved respect the one at start time (mean of the 12 monthly runs)



	Spatial average [day]
Phy	12
NH_4^+	8.4
NO_x^-	13.6
PO_4^{3-}	19.9

For the whole system, the time interval, τ, is around 10-15 days … It shows a spatial variability

DA of in situ field data can give information about the system for a time interval of about two weeks after the assimilation events [frequency of the sampling in *MELa1* is around 30 days]

Objective 1: Results

Sets of experiments show that:

1. BCs are the main forcing of the system

2a. The system remembers IC for about 10-15 days

2b. This time interval depends on the specific variable or process under consideration and it shows a spatial variability (5-20 days)

Objective 2 :

Ecosystem field estimation and seasonal dynamics based on ESSE data assimilation

- Stochastic models of uncertainty/variability in BCs: generate ESSE ensemble
- Assimilate in situ field data every months for 2001
- Results shown: seasonal evolution of ecosystem ESSE eigenvectors and ecosystem fields

Ensemble of stochastic BCs: Nutrient loads

New estimates of fields (via ESSE)

Ensemble of nutrient loads *L* from the drainage basin at day *k*

$$L_k^j = L_k \cdot \left(1 + w \cdot \sigma_R \cdot f_R\right) + L_k \cdot w \cdot \sigma_L$$

j=1:q   q=100

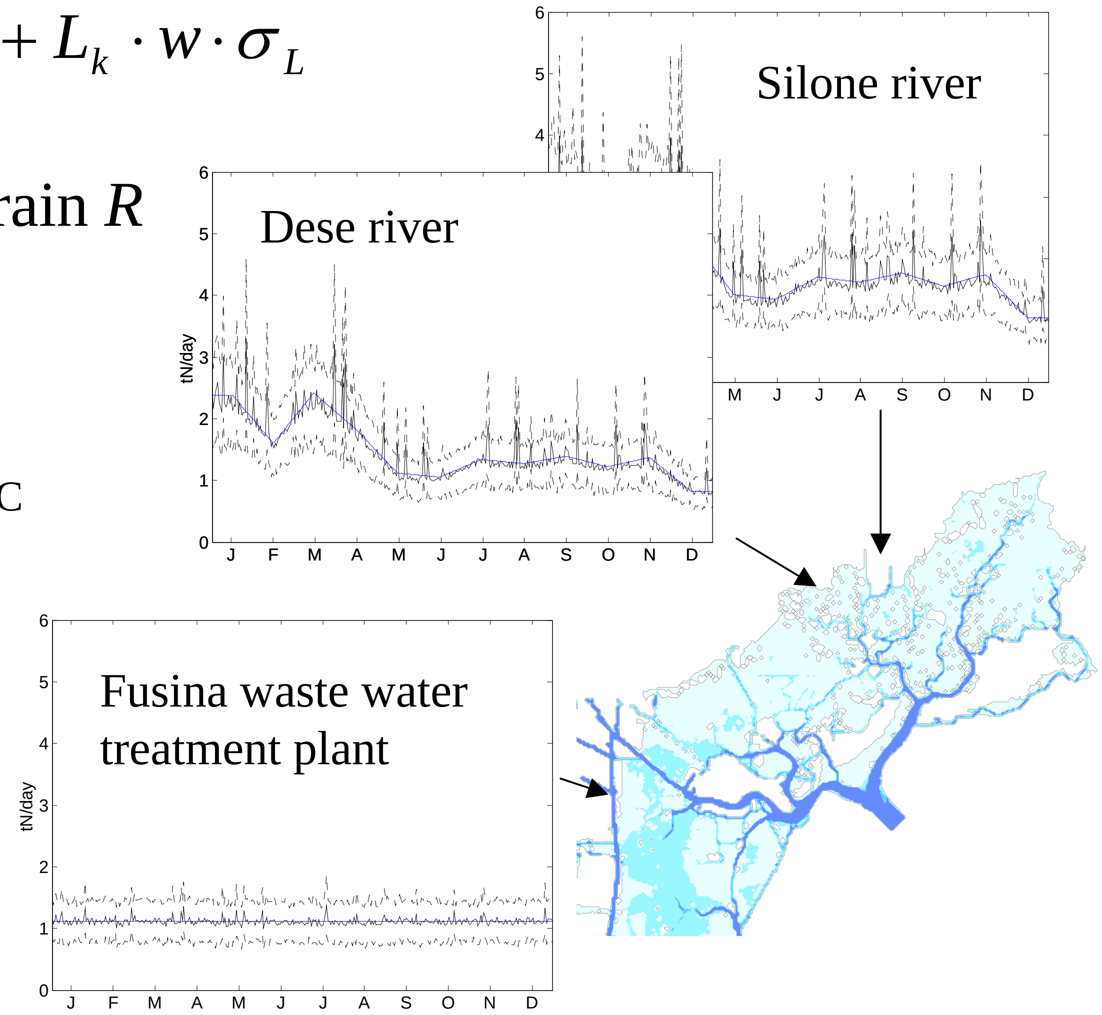
f_R is a function of daily rain *R*

$$f_R = \frac{R_{k-\tau_R} - \overline{R_{month}}}{R_{\max} - \overline{R_{month}}}$$

σ_R and σ_L are the noise covariances in BC

they vary for each point sources and are assumed constant in time

w is a random number of normal distribution: p(*w*)~N(0,1)



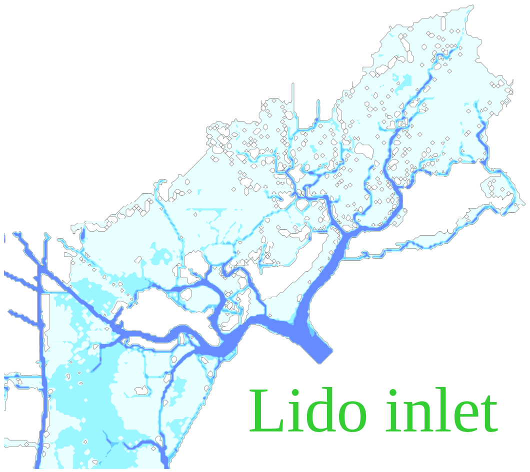
Ensemble of stochastic BCs: Inlets exchanges

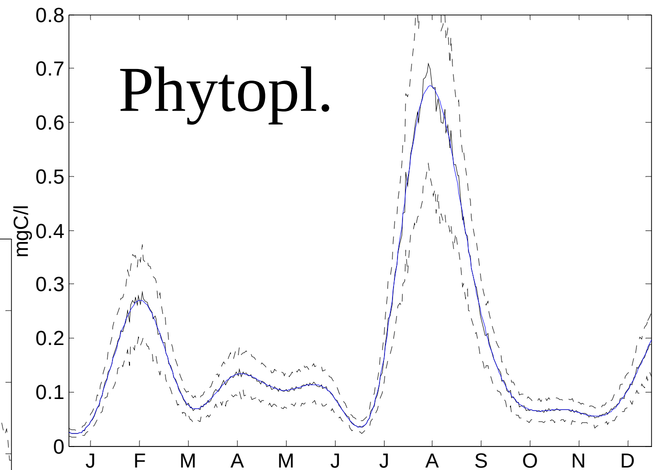
New estimates of fields (via ESSE)

Ensemble of the evolutions of the concentration of state variables at the inlets

$$C_k^j = C_k \cdot (1 + \sigma_c \cdot w)$$

j=1:q   q=100

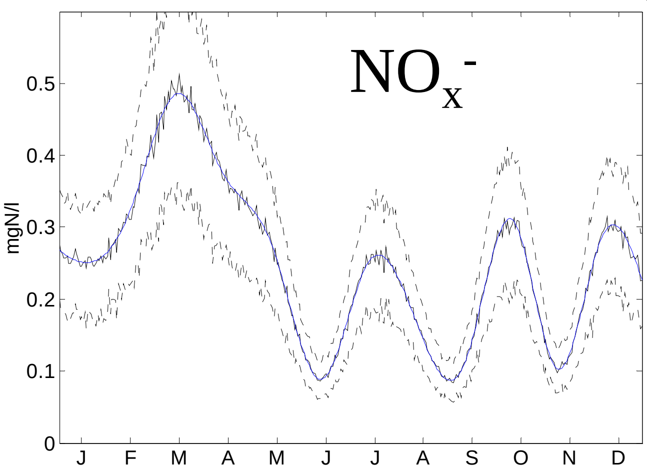




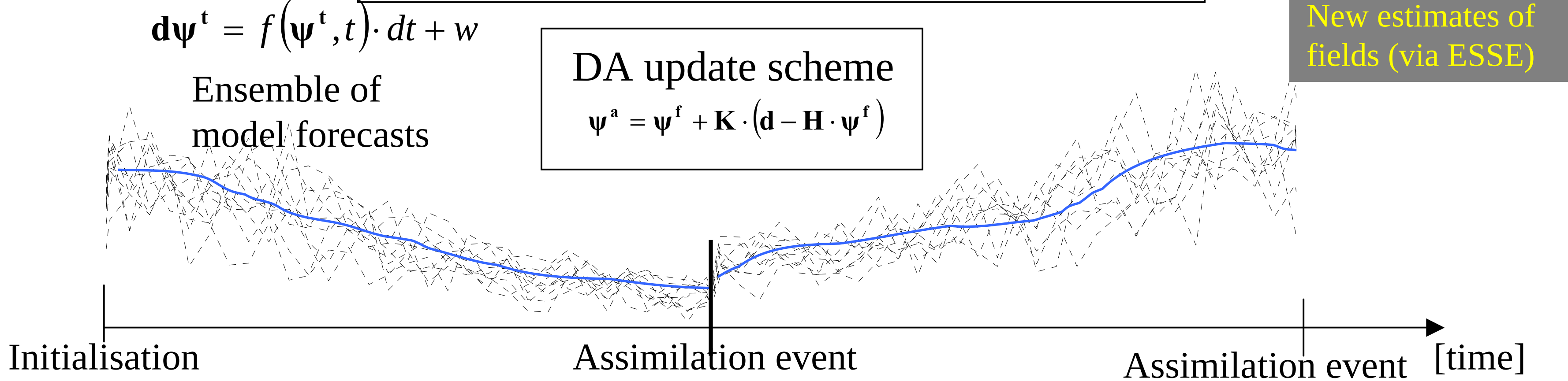
σ_C is the noise covariance in boundary

It is assumed constant in time

w is a random number of normal distribution: p(*w*)~N(0,1)



Ensemble of forecasts



M (error sample matrix)

Members of the ensemble (q)

Points of hte domain x # variables

$$
\mathbf{M}^* = \begin{bmatrix}
\Psi^1_{phy,1}-\Psi^t_{phy,1} & \Psi^2_{phy,1}-\Psi^t_{phy,1} & \ldots\Psi^j_{phy,1}-\Psi^t_{phy,1} & \ldots & \Psi^q_{phy,1}-\Psi^t_{phy,1} \\
\Psi^1_{phy,2}-\Psi^t_{phy,2} & \Psi^2_{phy,2}-\Psi^t_{phy,2} & & & \\
\ldots & \ldots & & & \\
\Psi^1_{phy,np}-\Psi^t_{phy,np} & & \Psi^j_{phy,np}-\Psi^t_{phy,np} & & \\
\Psi^1_{nh,1}-\Psi^t_{nh,1} & & \ldots & & \\
\ldots & \ldots & \Psi^j_{phy,i}-\Psi^t_{phy,i} & \ldots & \\
\Psi^1_{nh,np}-\Psi^t_{nh,np} & & .. & & \\
\ldots & & & &
\end{bmatrix}
$$

The mean of the ensemble of forecasts is chosen as the estimate of the true state of the system

ESSE scheme

$$\mathbf{P} = \frac{\mathbf{M}\cdot\mathbf{M}^{\mathbf{T}}}{q} = \mathbf{E}\cdot\mathbf{\Pi}\cdot\mathbf{E}^{\mathbf{T}}$$

Where $\mathbf{E}$ and $\mathbf{\Pi} = \frac{1}{q}\cdot\mathbf{\Sigma}^{\mathbf{2}}$ are respectively the matrix of eigenvectors and eigenvalues of the decomposition of **M** matrix

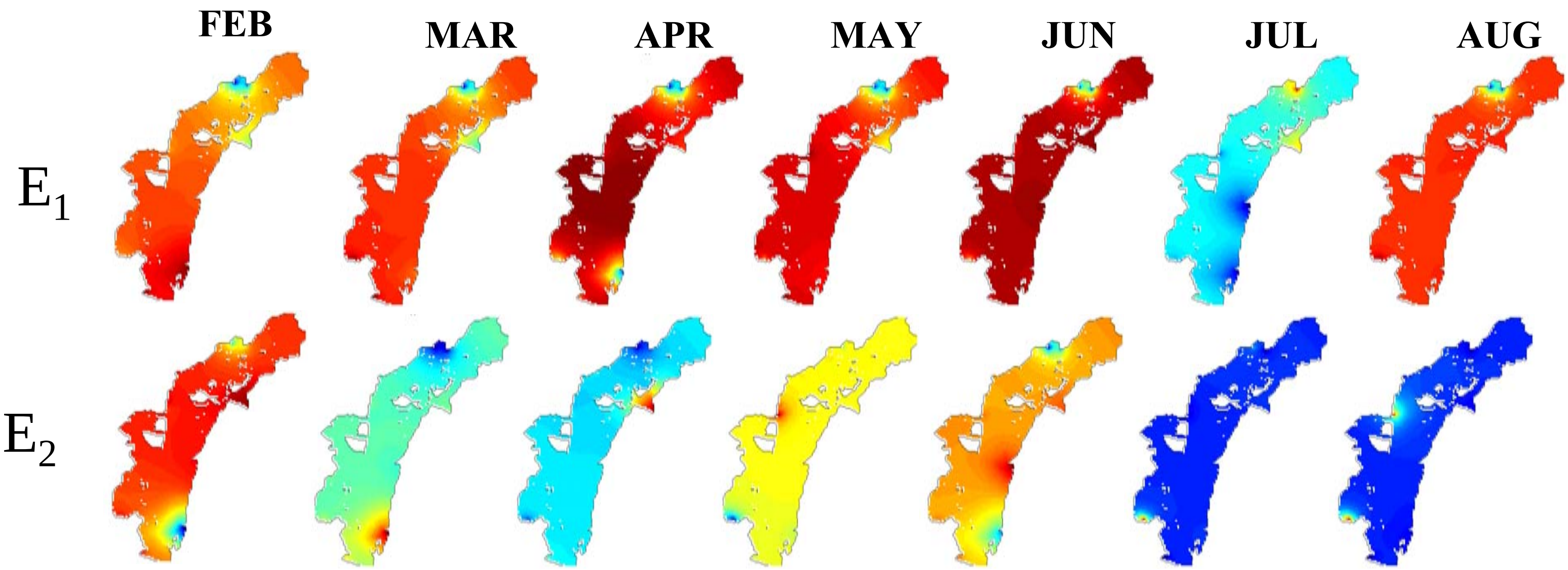
The *a posteriori* state is obtained using **E** and **Π**:

$$\boldsymbol{\psi}^{\mathbf{a}} = \boldsymbol{\psi}^{\mathbf{f}} + \mathbf{E}\cdot\mathbf{\Pi}\cdot\tilde{\mathbf{E}}^{\mathbf{T}}\cdot\tilde{\mathbf{H}}\cdot\left(\tilde{\mathbf{H}}\cdot\tilde{\mathbf{E}}\cdot\mathbf{\Pi}\cdot\tilde{\mathbf{E}}^{\mathbf{T}}\cdot\tilde{\mathbf{H}} + \tilde{\mathbf{R}}\right)^{-1}\cdot\left(\tilde{\mathbf{d}} - \tilde{\mathbf{H}}\cdot\tilde{\boldsymbol{\psi}}^{\mathbf{f}}\right)$$

1° result: first eigenvectors explain the major part of the variance of **M** (uncertainty of the system)

New estimates of fields (via ESSE)

Sequence of the first two eigenvectors of NO_x^-

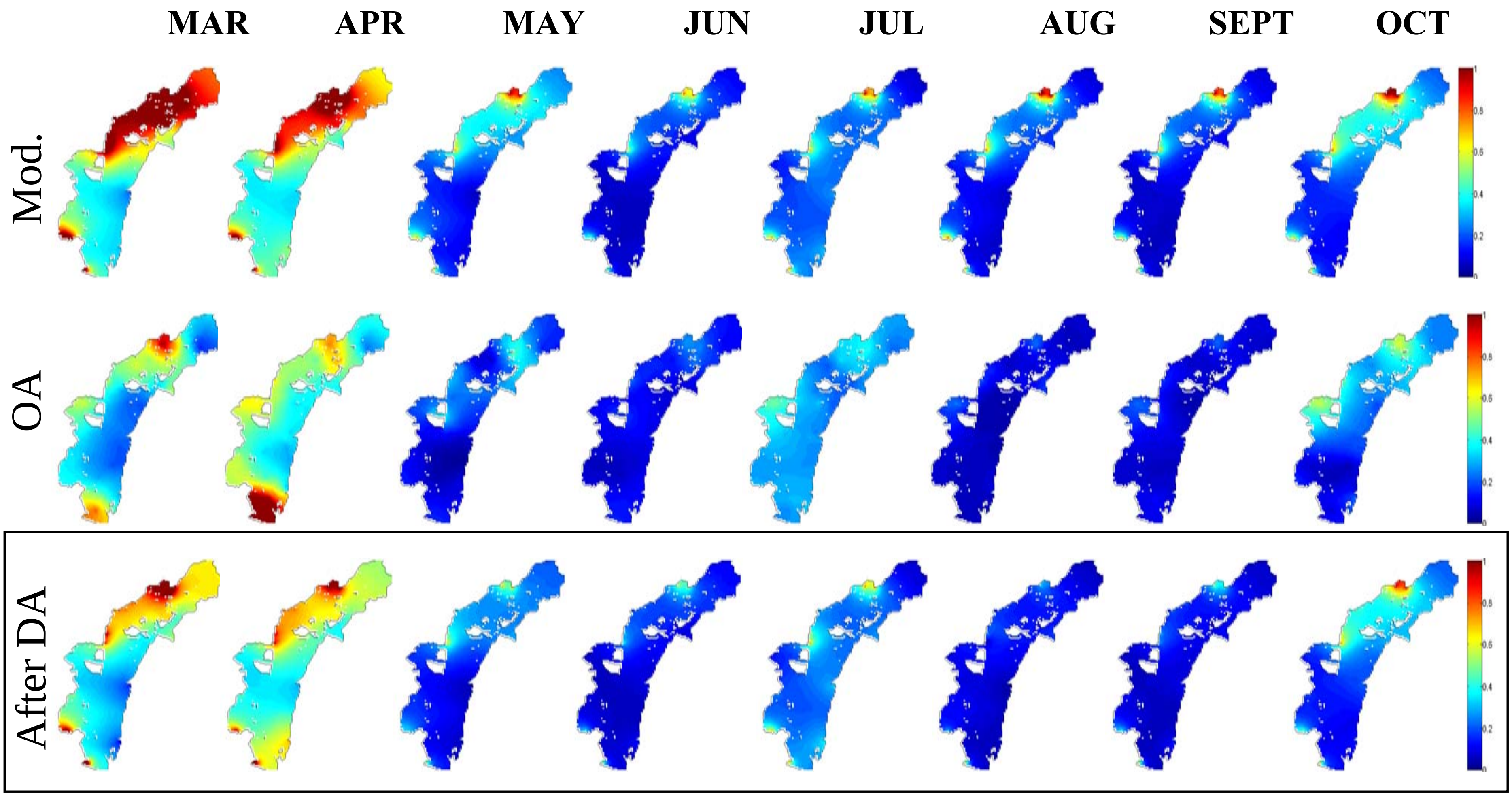


First eigenvectors indicate Dese and Silone as major sources of the variability during all the months

The second eigenvectors show the influence of the other sources. They vary from month to month (e.g. Chioggia inlet in Feb-Mar).

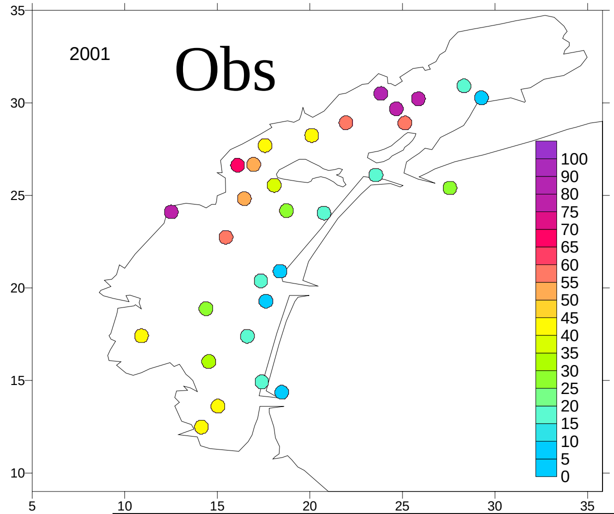
2° result: Evolution of new monthly estimate of NO_x^- maps

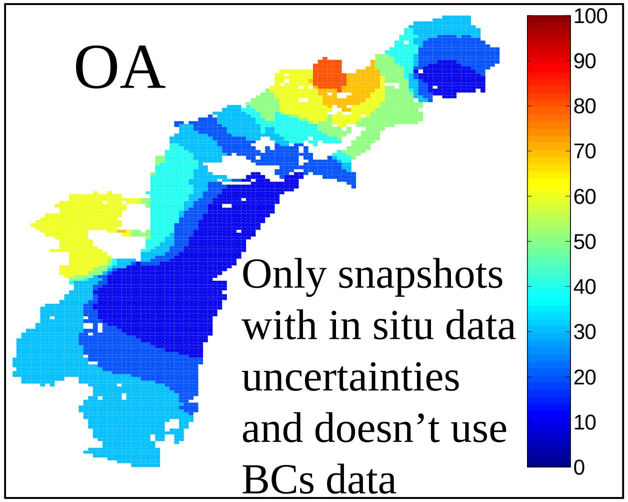
New estimates of fields (via ESSE)

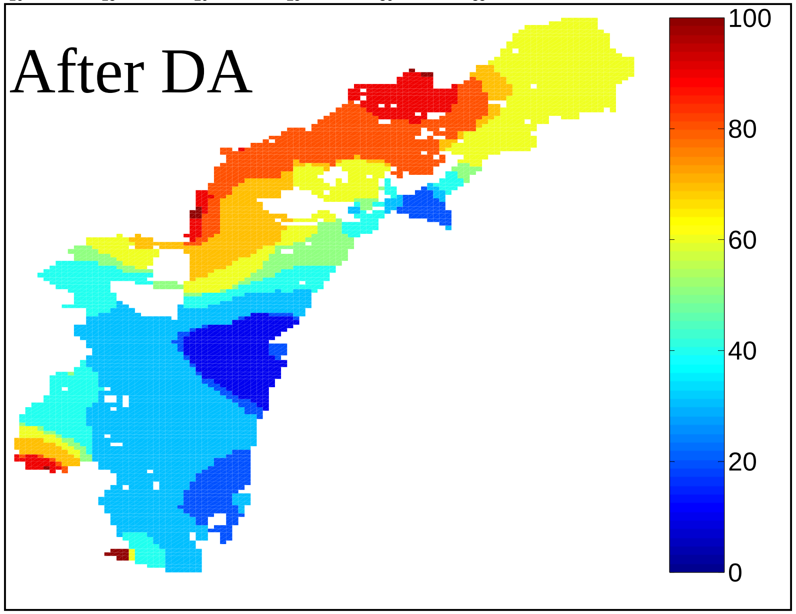


Special legislation (Ronchi/Costa Dlg) set Water Quality Targhet for nutrients. WQT of DIN (NH+NO) is 0.35 mg/l

Spatial distribution of the percentage of time above the Water Quality Target





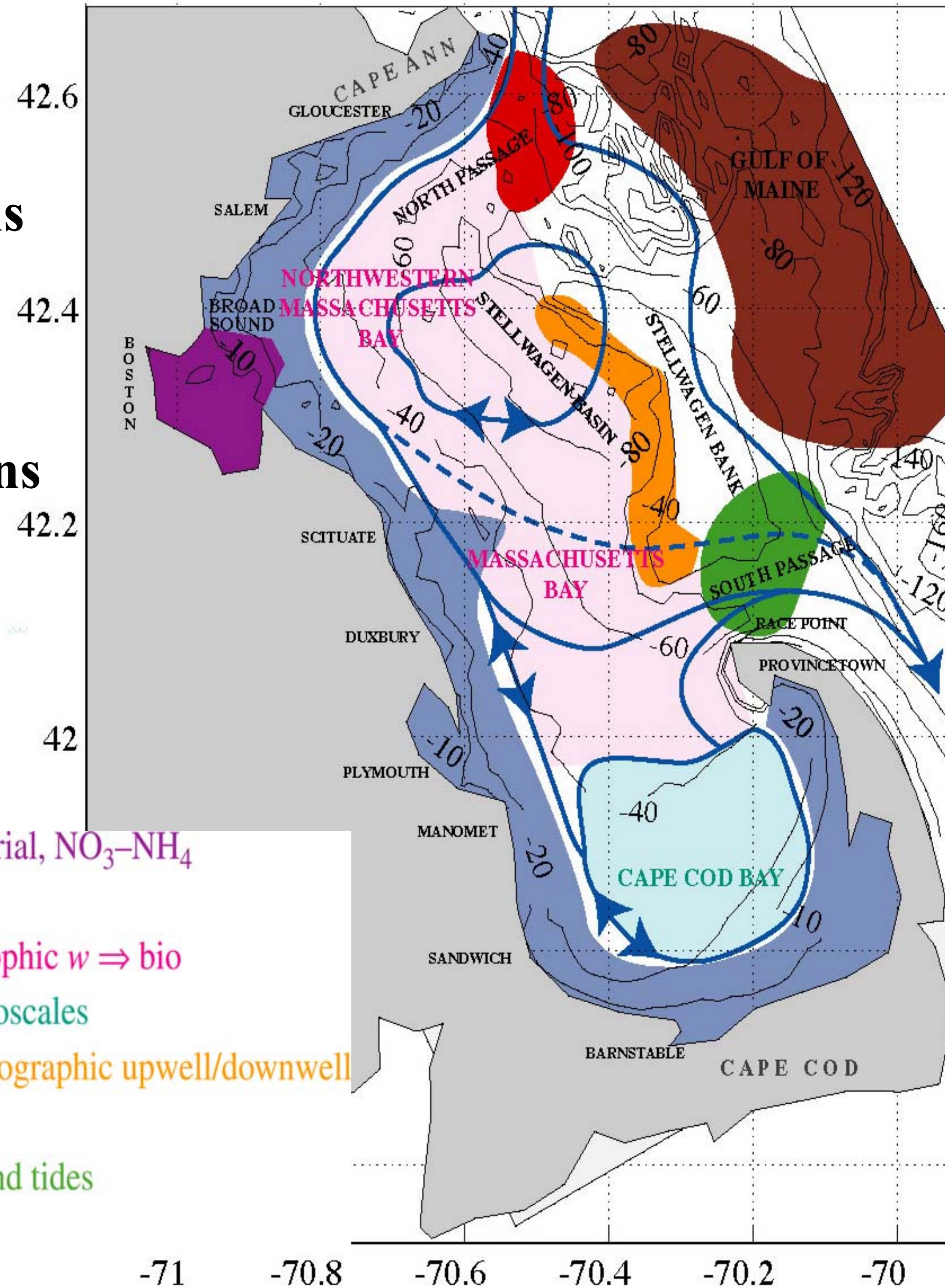


A Quest for Dominant Dynamical Balances

- **Ocean dynamics is complex, with multiple scales, processes and features**
- **Ultimate basic understanding can be relatively simple but hard to reach**
- **Modern approach:**
  - Combine data and dynamical models quantitatively for realistic studies
  - A road towards understanding and simplified dynamics
- **Many oceanic features can be described by limited number terms, said to be in approximate ``balance'': e.g. geostrophy, Ekman layer**
- **Focus here: explore dominant (dynamical) biogeochemical and biogeochemical-physical balances in coastal ecosystems**
- **Such balances are essential constraints for optimal sampling, biogeochemical initialization and selection of model parameters**

Massachusetts Bay

Horizontal Circulation Patterns for stratified conditions
(not present at all times)
and
Coupled bio-physical sub-regions in late summer
(Dominant dynamics for trophic enrichment and accumulation)



Boston Harbor: Charles River, sediments, toxic material, NO_3–NH_4

Along Coast: upwelling/downwelling ⇒ bio ↑/↓

Open Bay: submesoscale/mesoscale eddies. Ageostrophic *w* ⇒ bio

Cape Cod Bay: Horizontal bio advection and submesoscales

West of Stellwagen Bank: GOM meanders, tides, topographic upwell/downwell

Offshore: GOM meanders

Race Point: Multiple bio advections, accumulation, and tides

Cape Ann: Physical instabilities at GOM inflow

COUPLED PHYSICAL-BIOGEOCHEMICAL MODELS

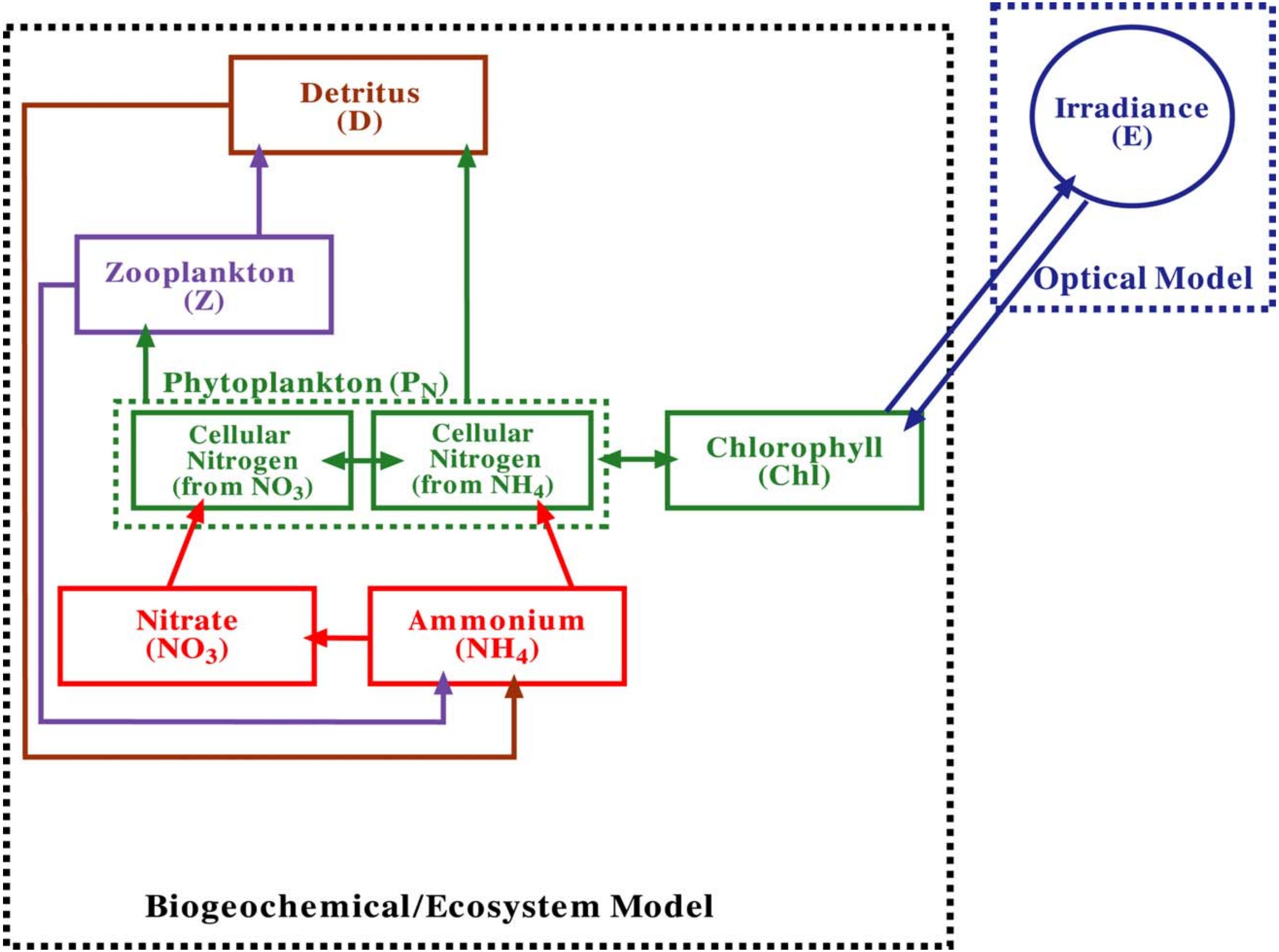
- **Physical model: Primitive-Equation (PDE, x, y, z, t: HOPS)**

Horiz. Mom.	$\frac{D\mathbf{u_h}}{Dt} + f\,\mathbf{e}_3 \wedge \mathbf{u}_h = -\frac{1}{\rho_0}\nabla_h\, p_w + \nabla_h \cdot (A_h \nabla_h \mathbf{u}_h) + \frac{\partial\; A_v\; \partial\mathbf{u}_h/\partial z}{\partial z}$	**(1-2)**
Vert. Mom.	$\rho g + \frac{\partial p_w}{\partial z} = 0$	**(3)**
Thermal en.	$\frac{DT}{Dt} = \nabla_h \cdot (K_h \nabla_h T) + \frac{\partial\; K_v\; \partial T/\partial z}{\partial z}$	**(4)**
Cons. of salt	$\frac{DS}{Dt} = \nabla_h \cdot (K_h \nabla_h S) + \frac{\partial\; K_v\; \partial S/\partial z}{\partial z}$	**(5)**
Cons. of mass	$\nabla \cdot \mathbf{u} = 0$	**(6)**
Eqn. of state	$\rho(\mathbf{r}, z, t) = \rho(T, S, p_w)$	**(7)**

- **Biogeochemical model: Generic ADR equation (PDE, x, y, z, t)**

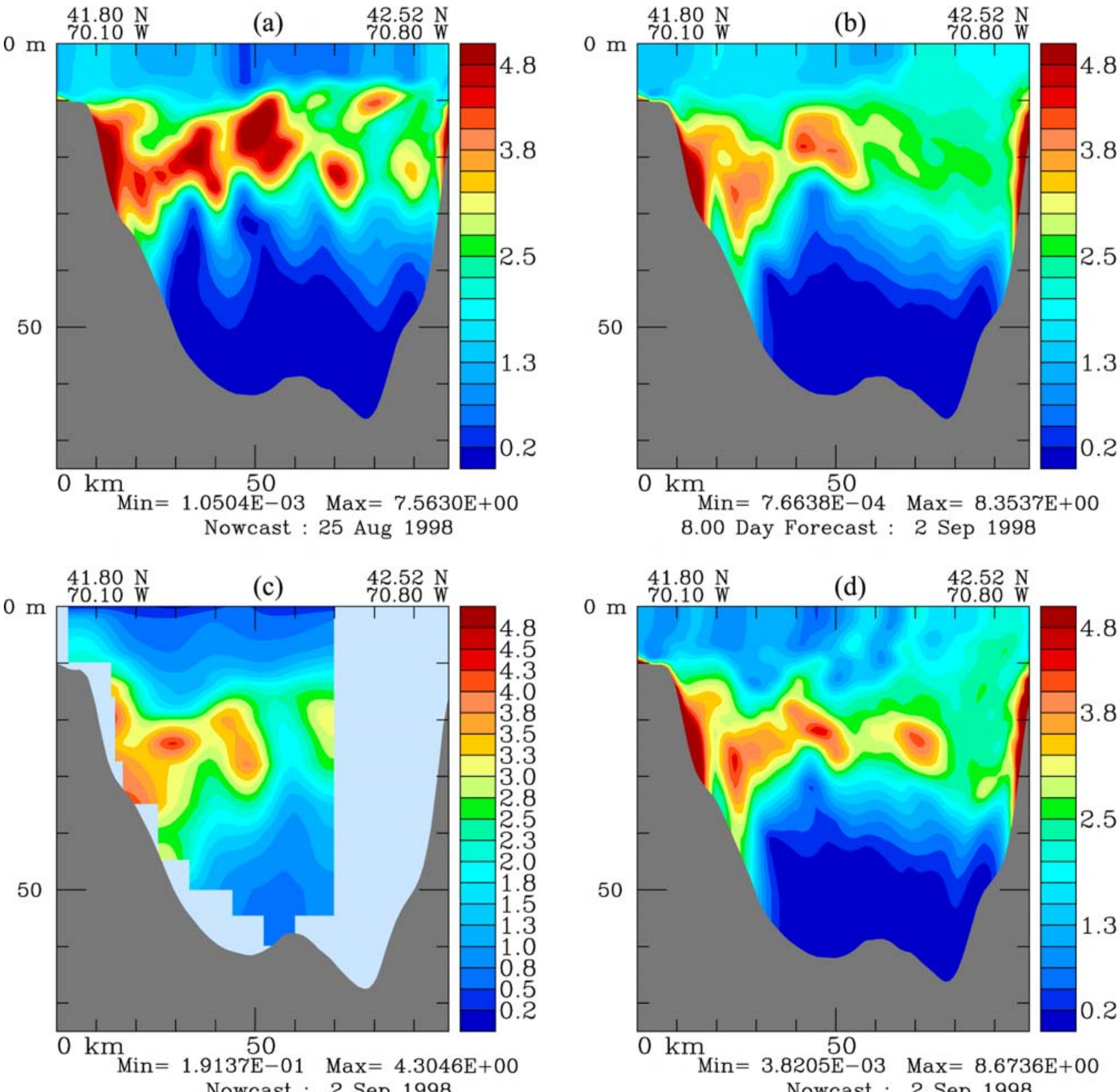
$$\frac{\partial \phi_i}{\partial t} + \mathbf{u} \cdot \nabla \phi_i - \nabla_h (A_i \nabla_h \phi_i) - \frac{\partial\; K_i\, \partial \phi_i/\partial z}{\partial z} = \mathcal{B}_i(\phi_1, \ldots, \phi_i, \ldots, \phi_7) \qquad (\mathbf{8-14})$$

$i = \mathrm{NO3},\ \mathrm{P}_{NO3},\ \mathrm{ZOO},\ \mathrm{NH4},\ \mathrm{DET},\ \mathrm{CHL},\ \mathrm{P}_{NH4}$



Schematic representation of ecosystem model (seven state variables/compartments)

Coupled Physical-Biogeochemical **Smoothing** via ESSE



Cross-sections in Chl-a fields, from south to north along main axis of Massachusetts Bay, with:

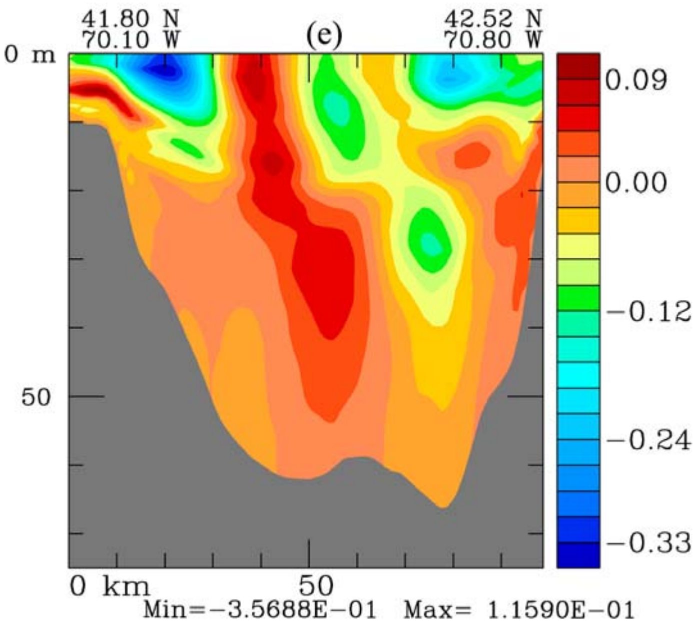
a) Nowcast on Aug. 25

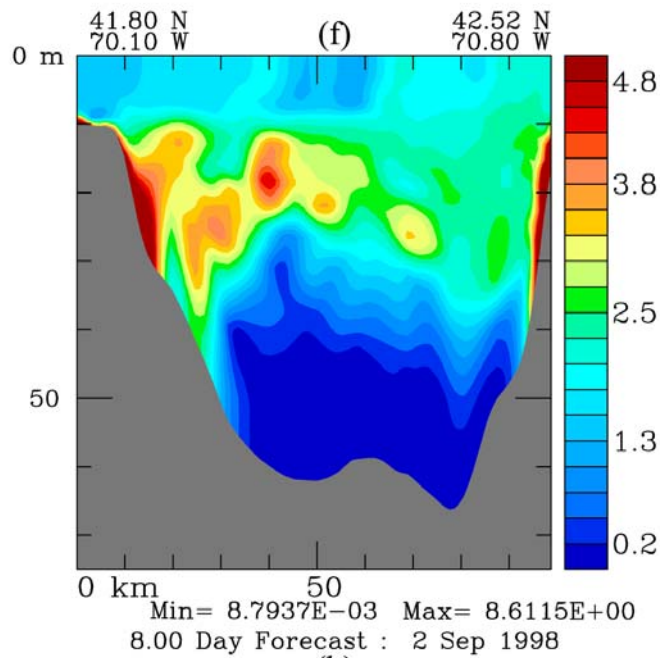
b) Forecast for Sep. 2

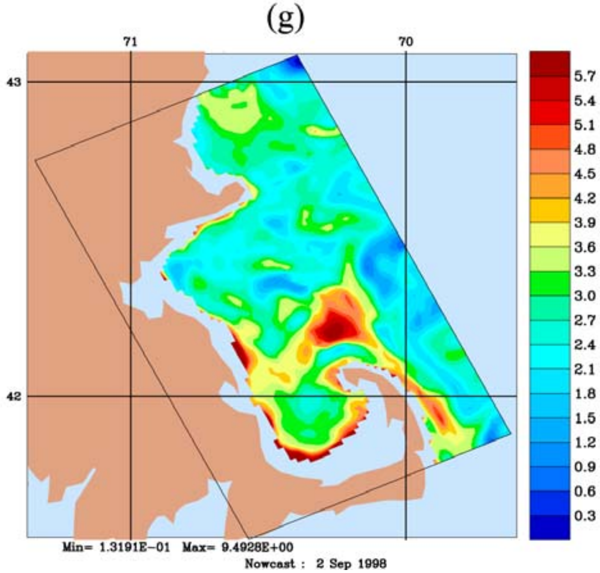
c) 2D objective analysis for Sep. 2 of Chl-a data collected on Sep. 2–3

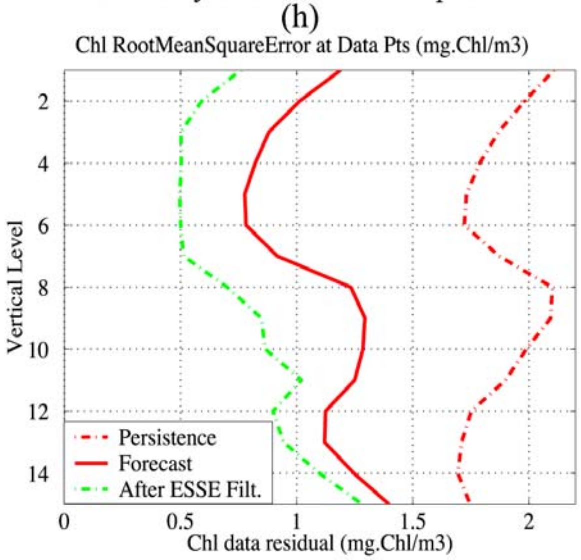
d) ESSE filtering estimate on Sep. 2

Coupled Physical-Biogeochemical DA via ESSE (continued)









e) Difference between ESSE smoothing estimate on Aug. 25 and nowcast on Aug. 25

f) Forecast for Sep. 2, starting from ESSE smoothing estimate on Aug. 25

(g): as d), but for Chl-a at 20 m depth

(h): RMS differences between Chl-a data on Sep. 2 and Chl-a estimates at data-points as a function of depth (specifically, "RMS-error" for persistence, dynamical forecast and ESSE filtering estimate).
Note: Internal predictability computed by ESSE to be ≥ 2 weeks

Dominant dynamical balances for initial biogeochemical fields/parameters

Circadian (daily) 0th order biological balance

Phyto eqn.:

$$\frac{1}{T}\int_0^T V^{NO_3} + V^{NH_4} dt \simeq gZ + n_3 P$$

$$V^{NH_4} = \frac{P_P}{\theta^C_{Chl}} \frac{NH_4}{k_{NH_4} + NH_4} P \ , \quad V^{NO_3} = \frac{P_P}{\theta^C_{Chl}} \frac{NO_3 \, e^{-\psi NH_4}}{k_{NO_3} + NO_3} P \ , \quad P_P = P_m (1 - e^{-\alpha E/P_m}) e^{-\beta E/P_m}$$

$$Optical \ \ model : E(x,y,z,t) = E(x,y,0,t) e^{-(k_w z + k_c \int_0^z \, Chl \, dz)}$$

Zoo eqn.: *Select non-zero root,*

$$0 \simeq (1 - \gamma_1 - \gamma_2) g - n_1 - n_2 Z \quad , \quad g = R_m (1 - e^{-\Lambda P})$$

Nitrate eqn.:

$$\frac{1}{T}\int_0^T V^{NO_3} dt \simeq k_N [NH_4]$$

Ammonium eqn.:

$$\frac{1}{T}\int_0^T V^{NH_4} dt \simeq -k_N [NH_4] + \gamma_1 g Z + (1 - \epsilon_1) n_1 Z + (1 - \epsilon_2) n_2 Z^2 + k_D D$$

Detritus eqn.: $\int_0^H dz$ *with either null* $D = 0$ *or well-mixed* $\frac{\partial D}{\partial z} = 0$ *surf./bot. BC*

$$\nu_D \frac{\partial D}{\partial z} + k_D D \simeq \gamma_2 g Z + \epsilon_1 n_1 Z + \epsilon_2 n_2 Z^2 + n_3 P + n_4 P^2$$

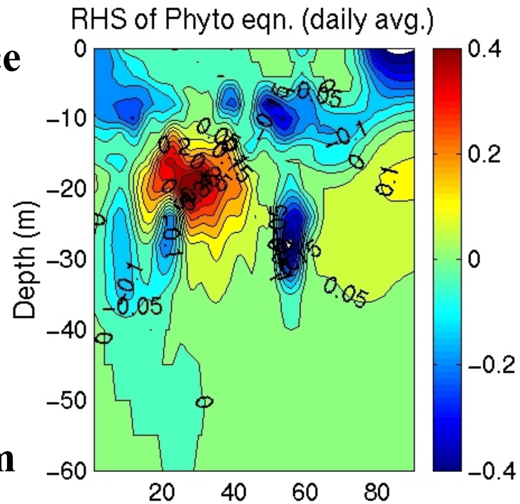
Chlorophyll: *OA-ed from Fluo. For P, assume fully-acclimated C:CHL*

Balance subject to observed variables and parameters constraints

Daily Average of RHS Terms: Residuals vs. Order of Magnitude

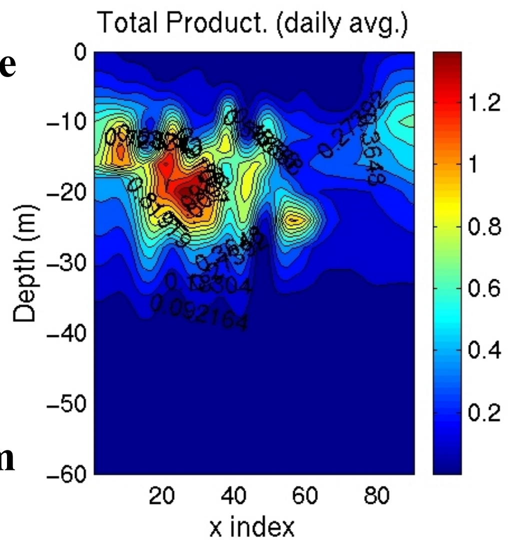
Mass. Bay: Aug 21, 1998

Surface



Bottom

Surface



Bottom

Cape Cod

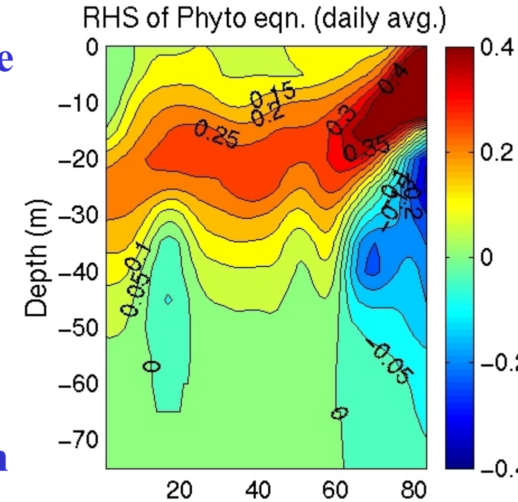
Cape Ann

Residuals

Orders of magnitude

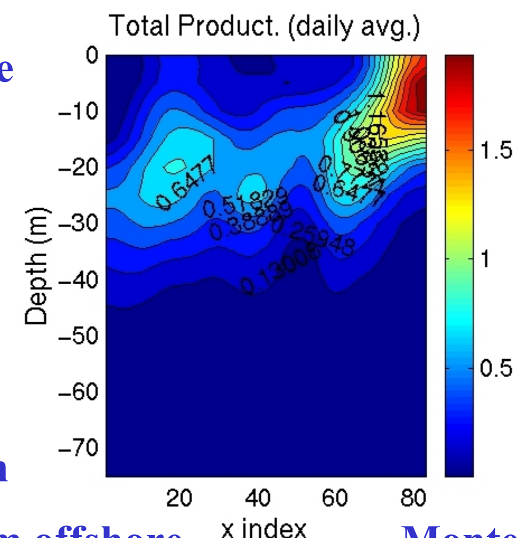
Monterey Bay: Aug 04, 2003

Surface



Bottom

Surface

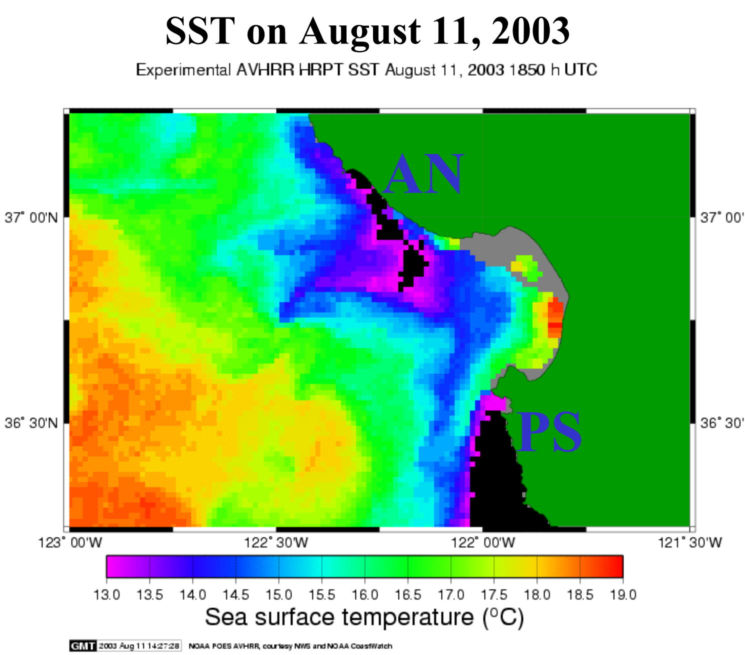


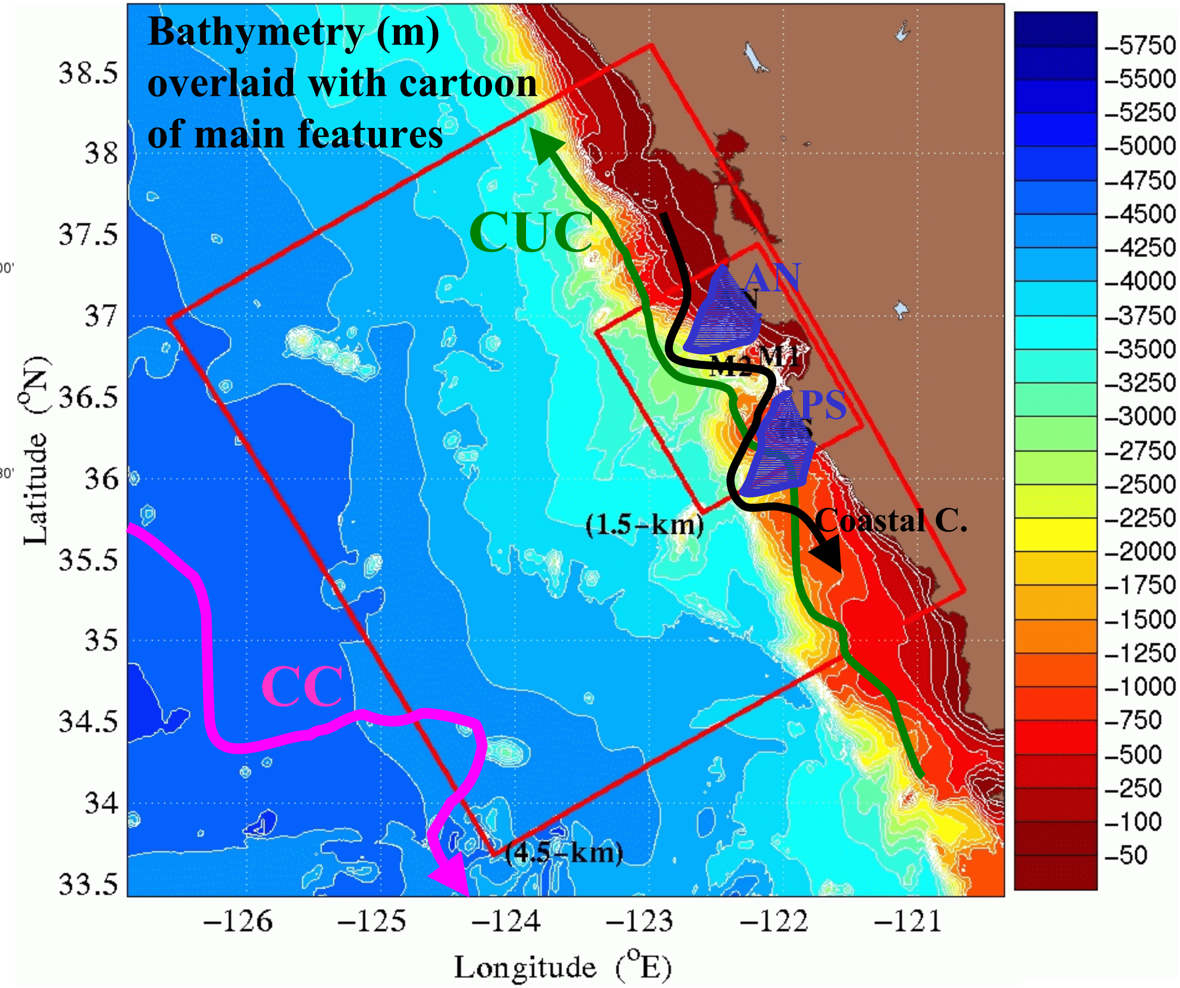
Bottom

100km offshore

Monterey

REGIONAL FEATURES of Monterey Bay and California Current System and Real-time Modeling Domains (AOSN2, 4 Aug. – 3 Sep., 2003)

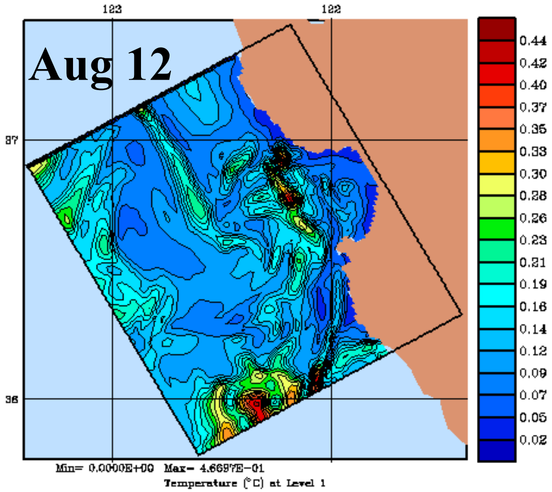




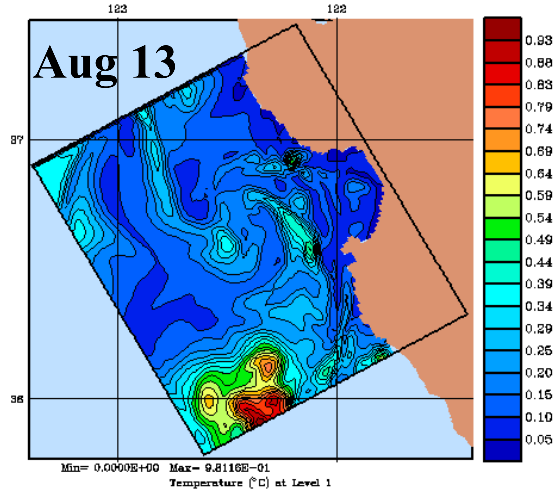
REGIONAL FEATURES

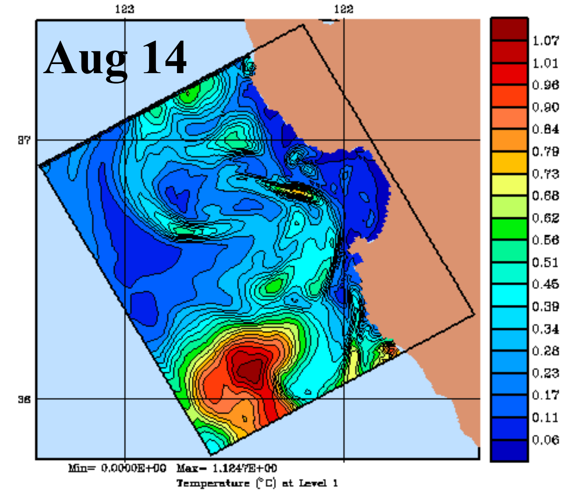
- **Upwelling centers at Pt AN/ Pt Sur**:…..……..Upwelled water advected equatorward and seaward
- **Coastal current, eddies, squirts, filam., etc:**….Upwelling-induced jets and high (sub)-mesoscale var. in CTZ
- **California Undercurrent (CUC)**:……………..Poleward flow/jet, 10-100km offshore, 50-300m depth
- **California Current (CC)**:……………………Broad southward flow, 100-1350km offshore, 0-500m depth

ESSE Surface Temperature Error Standard Deviation Forecasts

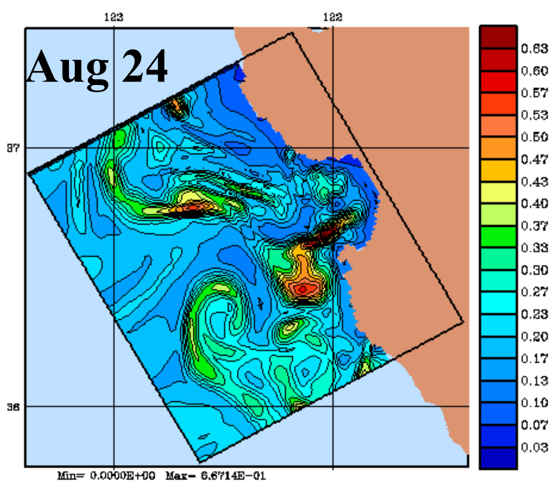


Start of Upwelling

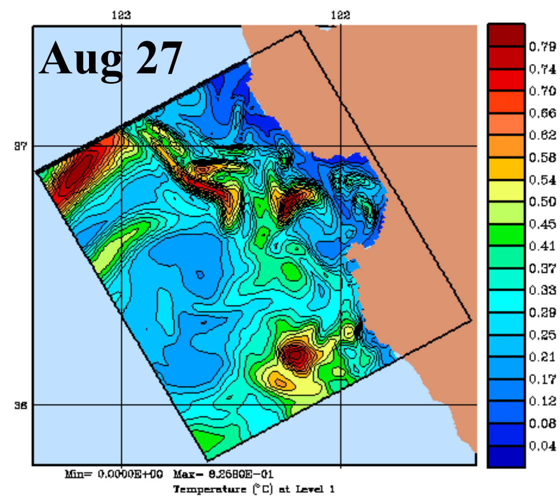


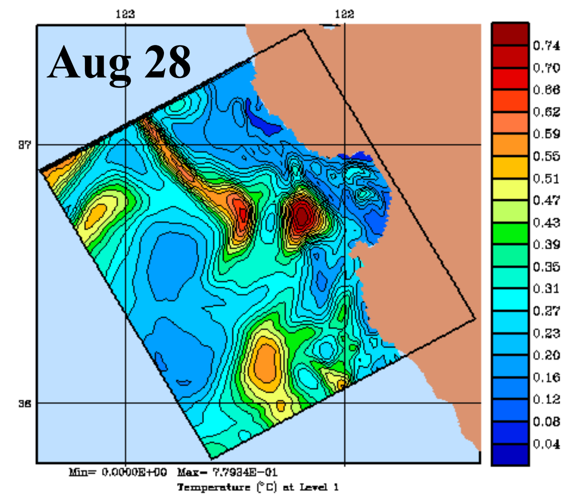


First Upwelling period



End of Relaxation





Second Upwelling period

- **Real-time consistent error forecasting, data assimilation and adaptive sampling (1 month)**
- **ESSE results described in details and posted on the Web daily (see AOSN2 page at HU)**

Oceanic Adaptive Sampling: Multiple Facets

Foci	- Optimal ocean science (Physics, Acoustics and/or Biology) - Demonstration of adaptive sampling value, etc.
Objective Fields	i. Maintain synoptic accuracy (e.g. upwelling, BL or CUC/CCS coverage) ii. Minimize uncertainties (e.g. uncertain ocean estimates), or iii. Maximize the sampling of expected events (e.g. start of upwelling/ relaxation, dynamics of upwelling filament, small scales/model errors) Multidisciplinary or not Local, regional or global, etc.
Time and Space Scales	i. Tactical scales (e.g. minutes-to-hours adaptation by each glider) ii. Strategic scales (e.g. hours-to-days adaptation for glider group/cluster) iii. Experiment scales
Assumptions	- Fixed or variable environment (w.r.t. asset speeds) - Objective field depends on the predicted data values or not - Operational, time and cost constraints, or not, etc.
Methods	Bayesian-based, Nonlinear programming, (Mixed)-integer programming, Simulated Annealing, Genetic algorithms, Neural networks, Fuzzy logics

For each of the 5 categories, there are multiple choices (only a few listed here)
Choices set the type of adaptive sampling research

a. Adaptive sampling via ESSE

- Objective: Minimize predicted trace of full error covariance (T,S,U,V error std Dev).
- Scales: Strategic/Experiment (not tactical yet). Day to week.
- Assumptions: Small number of pre-selected tracks/regions (based on quick look on error forecast and constrained by operation)
- Problem solved: e.g. Compute today, the tracks/regions to sample tomorrow, that will most reduce uncertainties the day after tomorrow.

- Objective field changes during computation and is affected by data to-be-collected

- Model errors ***Q*** can account for coverage term

Dynamics:	$dx = M(x)dt + d\eta$	$\eta \sim N(0, Q)$
Measurement:	$y = H(x) + \varepsilon$	$\varepsilon \sim N(0, R)$

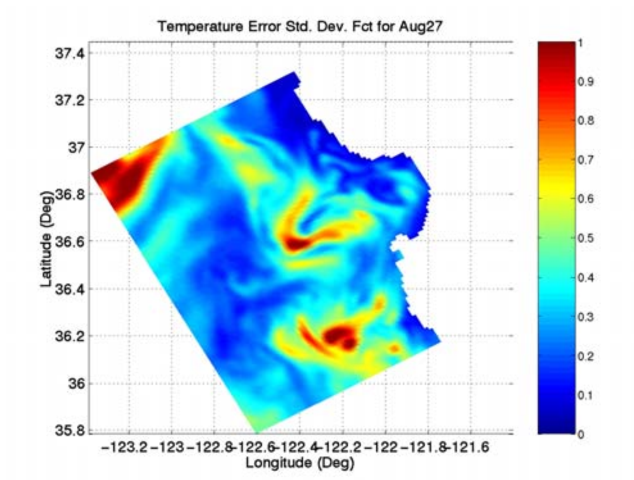
Non-lin. Err. Cov.:

$$dP/dt = \langle (x - \hat{x})(M(x) - M(\hat{x}))^T \rangle + \langle (M(x) - M(\hat{x})(x - \hat{x})^T \rangle + Q$$

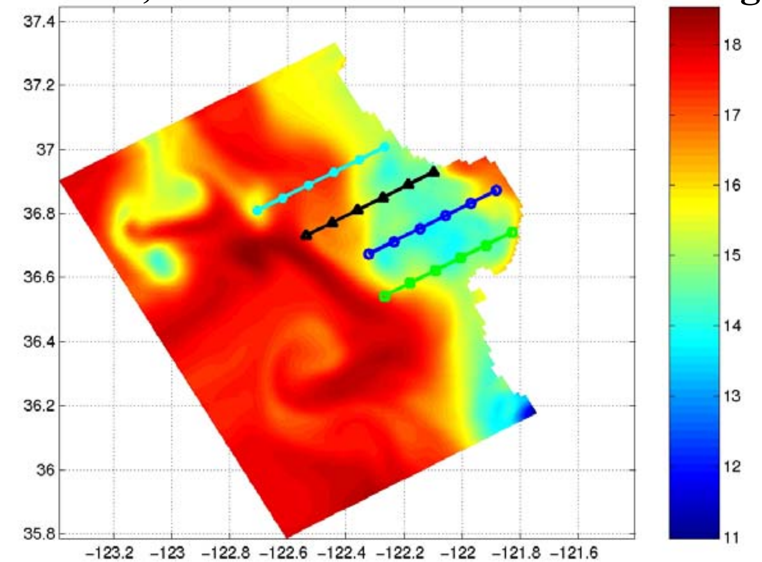
Metric or Cost function: e.g. Find future H_i and R_i such that

$$\underset{Hi,Ri}{Min} \quad tr(P(t_f)) \qquad or \qquad \underset{Hi,Ri}{Min} \int_{t_0}^{t_f} tr(P(t))\ dt$$

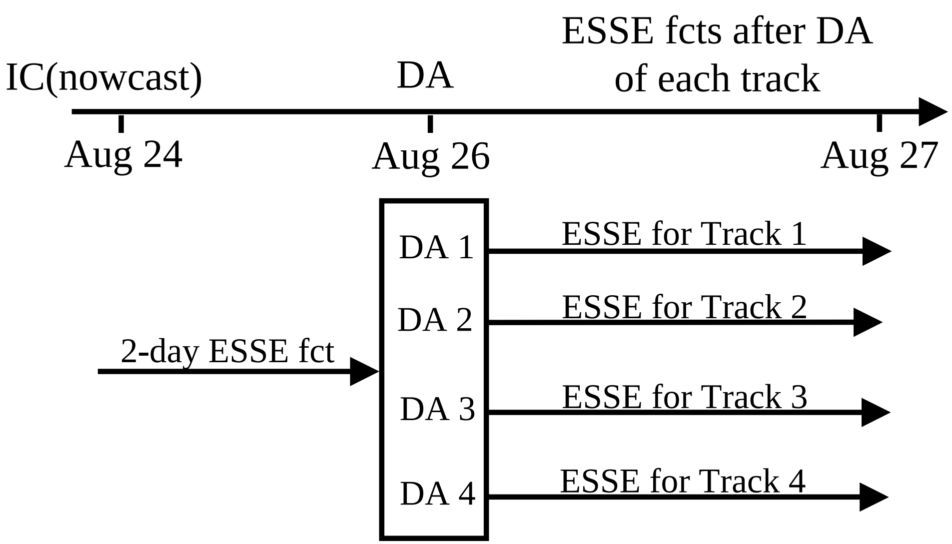
Which sampling on Aug 26 optimally reduces uncertainties on Aug 27?



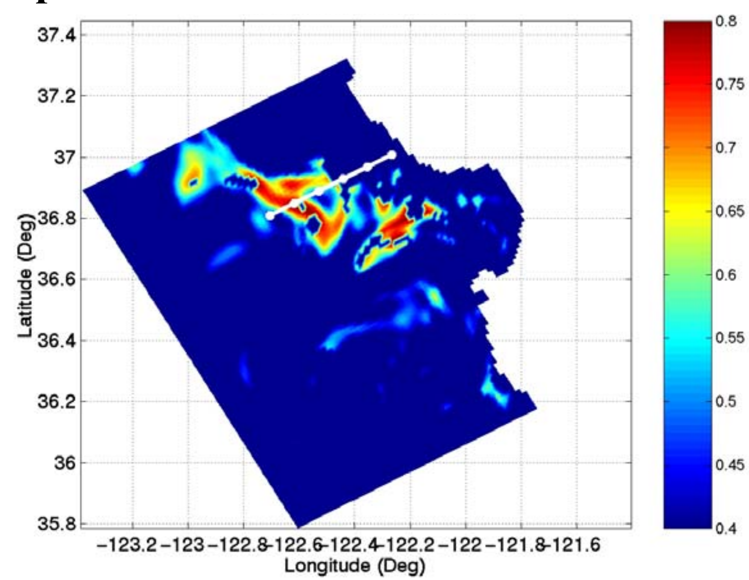
4 candidate tracks, overlaid on surface T fct for Aug 26



- **Based on nonlinear error covariance evolution**
- **For every choice of adaptive strategy, an ensemble is computed**



Best predicted relative error reduction: track 1



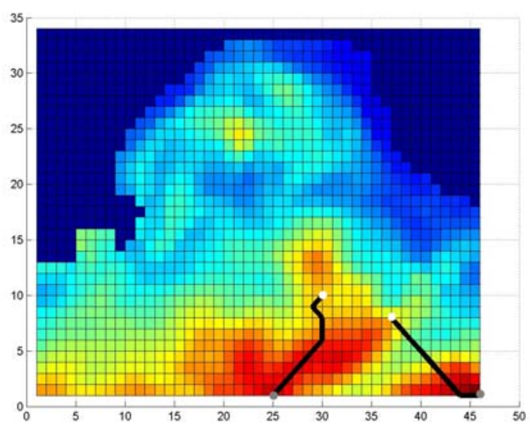
b. Optimal Paths Generation for a “fixed” objective field

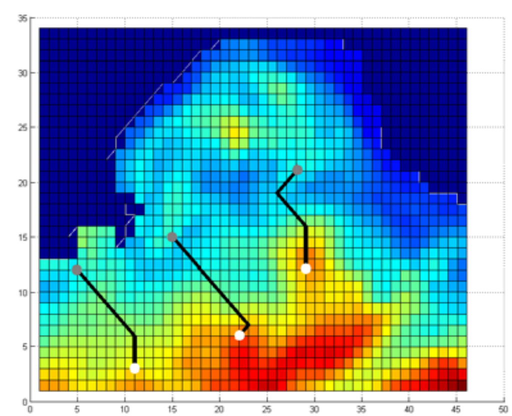
(Namik K. Yilmaz, P. Lermusiaux and N. Patrikalakis)

- Objective: Minimize ESSE error standard deviation of temperature field
- Scales: Strategic/Tactical
- Assumptions
  - Speed of platforms >> time-rate of change of environment
  - Objective field fixed during the computation of the path and is not affected by new data
- Problem solved: assuming the error is like that now and will remain so for the next few hours, where do I send my gliders/AUVs?
- Method: *Combinatorial optimization (Mixed-Integer Programming, using Xpress-MP code)*
  - Objective field (error stand. dev.) represented as a piecewise-linear: solved *exactly* by MIP
  - Possible paths defined on discrete grid: set of possible path is thus finite (but large)
  - Constraints imposed on vehicle displacements dx, dy, dz for meaningful path

Example:
Two and Three Vehicles, 2D objective field (3D examples also done)

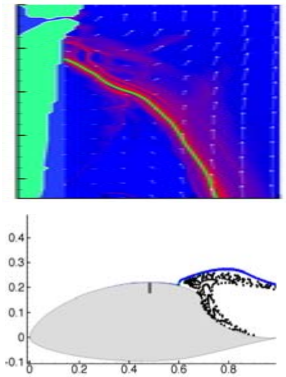
Grey dots: starting points
White dots: MIP optimal end points



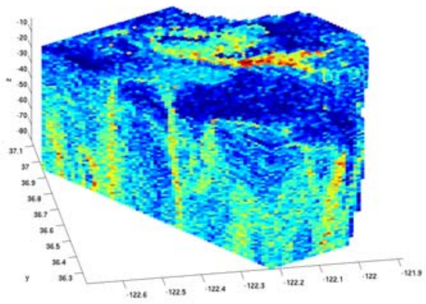


c. Lagrangian Coherent Structures (LCS):

Defined by extrema in "Direct Lyapunov Exponent" (DLE scalar field)



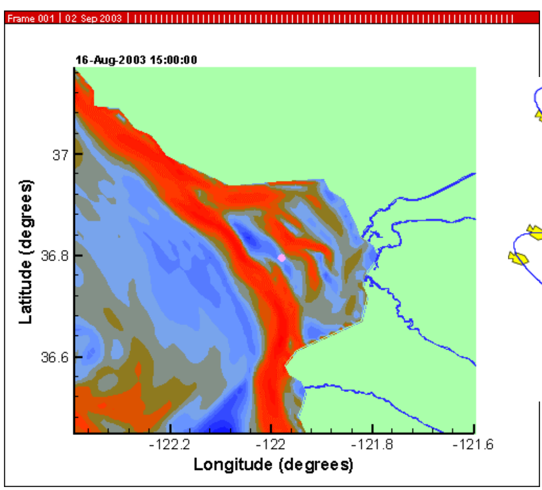
LCS are barriers in many flows



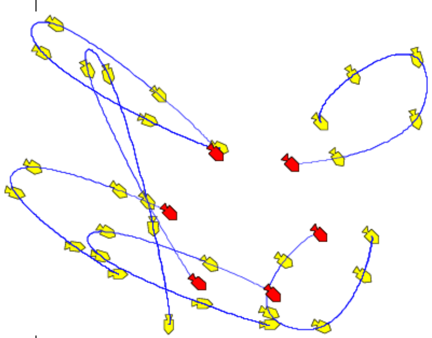
3D capabilities being developed

Objectives:

- Extend LCS capabilities to 3D
- Relate to biological and geophysical features (temperature fronts, plankton plumes)
- Use LCS for optimal path-planning of underwater gliders (single and groups—deployment and transit)
- Optimization tools for Princeton error metric



Optimal path lies on an LCS



Optimize group deployment

Approach:

- Compute LCS from Observational and Model Predicted oceanographic flow data
- New theory and flux estimates shows precisely how LCS act as barriers
- Use Mangen or GAIO for LCS computations and NTG or DMOC for optimization.

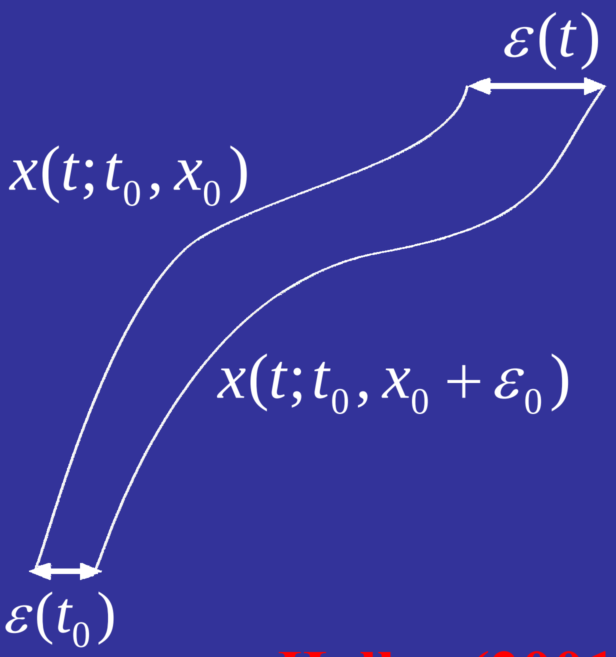
Principal Investigator:

Prof. Jerrold E. Marsden
Dept. of Control and Dynamical Systems
California Institute of Technology

Direct Lyapunov Exponent (DLE)

The finite-time Lyapunov exponent is the maximum exponential growth about a trajectory.

$$\sigma_T(t_0, \mathbf{x}_0) = \frac{1}{T} \ln \max_{\epsilon(t_0)} \left\{ \frac{\epsilon(t)}{\epsilon(t_0)} \right\}$$



Haller (2001)

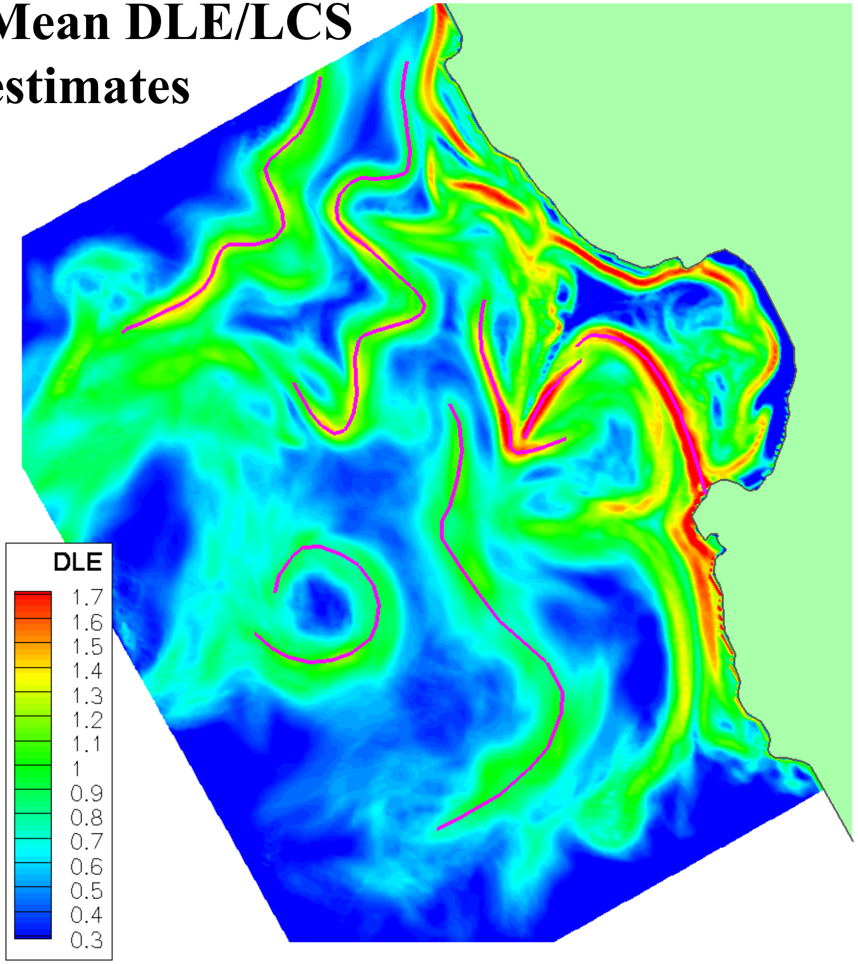
More precisely:

$$\sigma_T = \frac{1}{T} \ln \lambda_{\max} \left\{ \frac{\partial \mathbf{x}(\mathbf{t_0{+}T}, \mathbf{t_0}, \mathbf{x_0})}{\partial \mathbf{x}_0}^{\top} \frac{\partial \mathbf{x}(\mathbf{t_0{+}T}; \mathbf{t_0}, \mathbf{x_0})}{\partial \mathbf{x}_0} \right\}$$

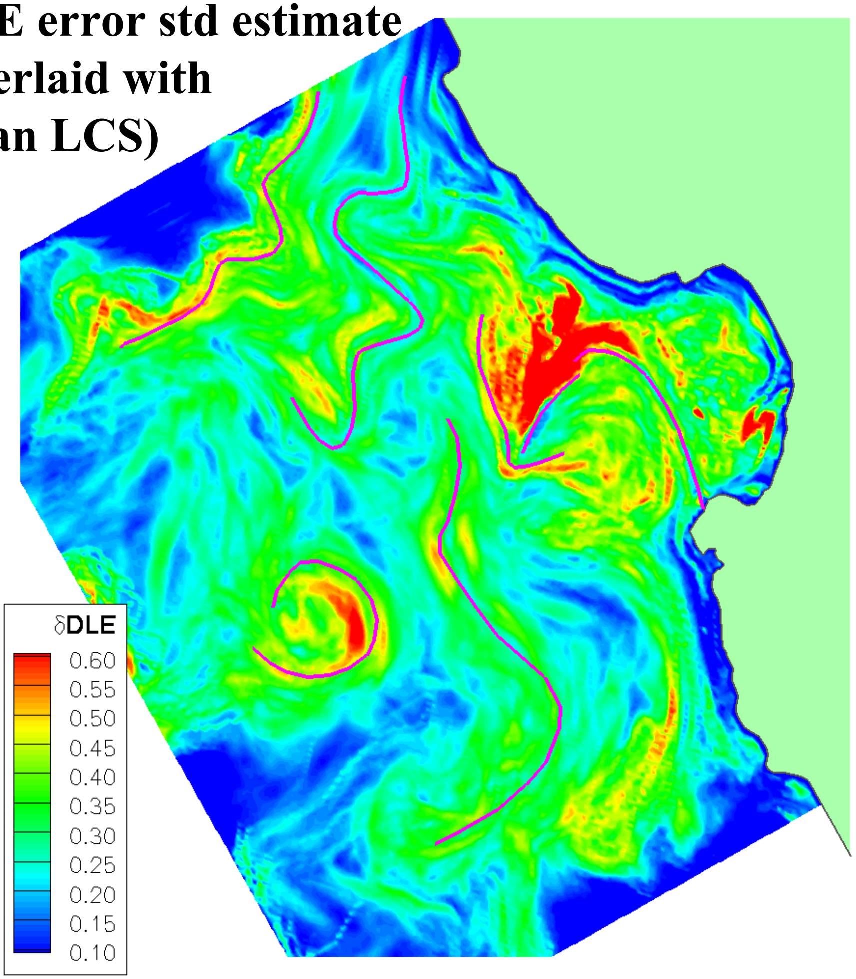
The DLE is also the normalized log of the maximum singular value of the finite time strain tensor dx/dx_0. Of course, $x \in R^2$ here!

Example: DLE/LCS and their Uncertainties for Aug 26-29, 2003 Upwelling Period

Mean DLE/LCS estimates



DLE error std estimate (overlaid with mean LCS)



See: Lermusiaux and Lekien, Aug. 2005, to appear.
for "Dynamical System Methods in Fluid Dynamics", Oberwolfach, Germany.

d. Different Objective Fields: Flux and/or Term-by-term Balances

- **Physical model: Primitive-Equation (PDE, x, y, z, t: HOPS)**

Horiz. Mom. $\frac{D\mathbf{u_h}}{Dt} + f\,\mathbf{e}_3 \wedge \mathbf{u}_h = -\frac{1}{\rho_0}\nabla_h p_w + \nabla_h \cdot (A_h \nabla_h \mathbf{u}_h) + \frac{\partial\, A_v\, \partial\mathbf{u}_h/\partial z}{\partial z}$

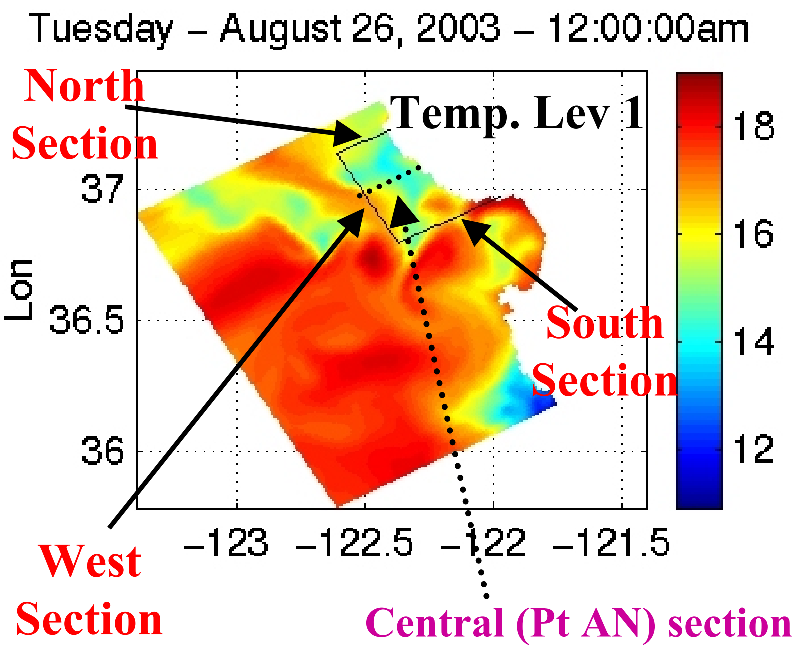
Vert. Mom. $\rho g + \frac{\partial p_w}{\partial z} = 0$

Thermal en. $\frac{DT}{Dt} = \nabla_h \cdot (K_h \nabla_h T) + \frac{\partial\, K_v\, \partial T/\partial z}{\partial z}$

Cons. of salt $\frac{DS}{Dt} = \nabla_h \cdot (K_h \nabla_h S) + \frac{\partial\, K_v\, \partial S/\partial z}{\partial z}$

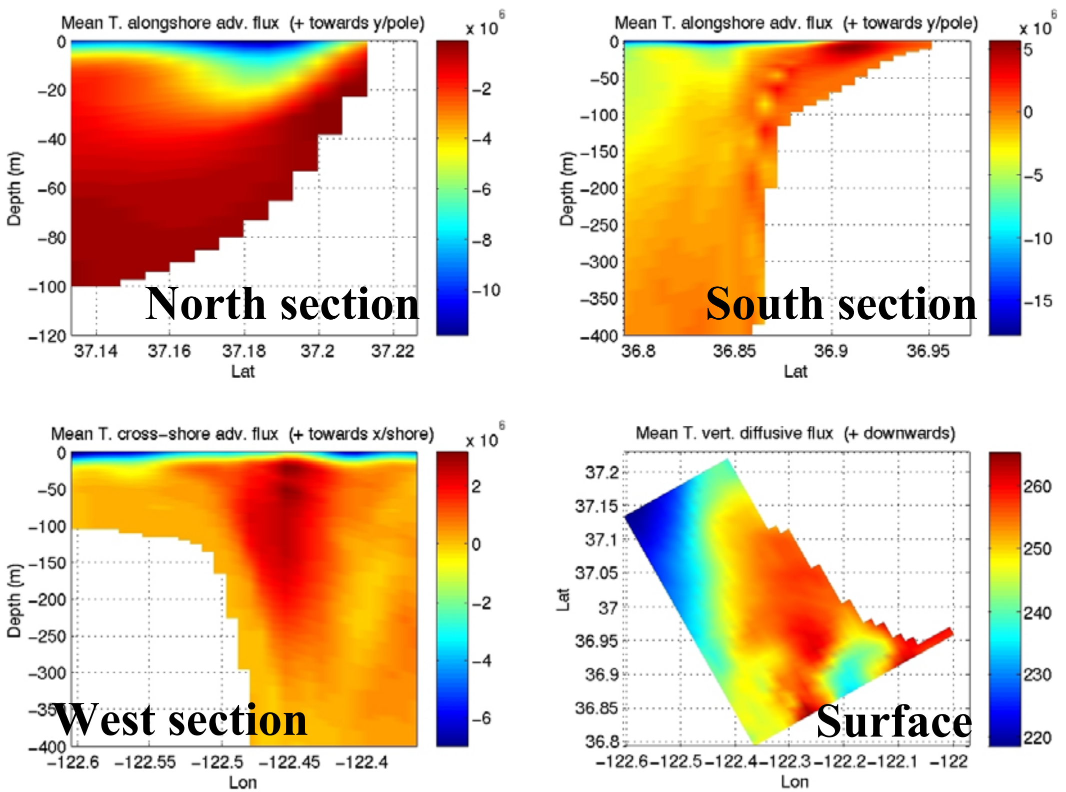
Cons. of mass $\nabla \cdot \mathbf{u} = 0$

Eqn. of state $\rho(\mathbf{r}, z, t) = \rho(T, S, p_w)$



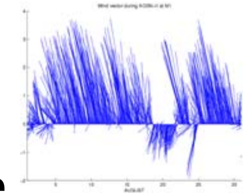
Heat Flux Balances: 4 fluxes normal to each side averaged over first upwelling period

Mean Fluxes (W/m2) over: August 6, 2003 – 10:30:00pm –> August 13, 2003 – 4:30:00am GMT

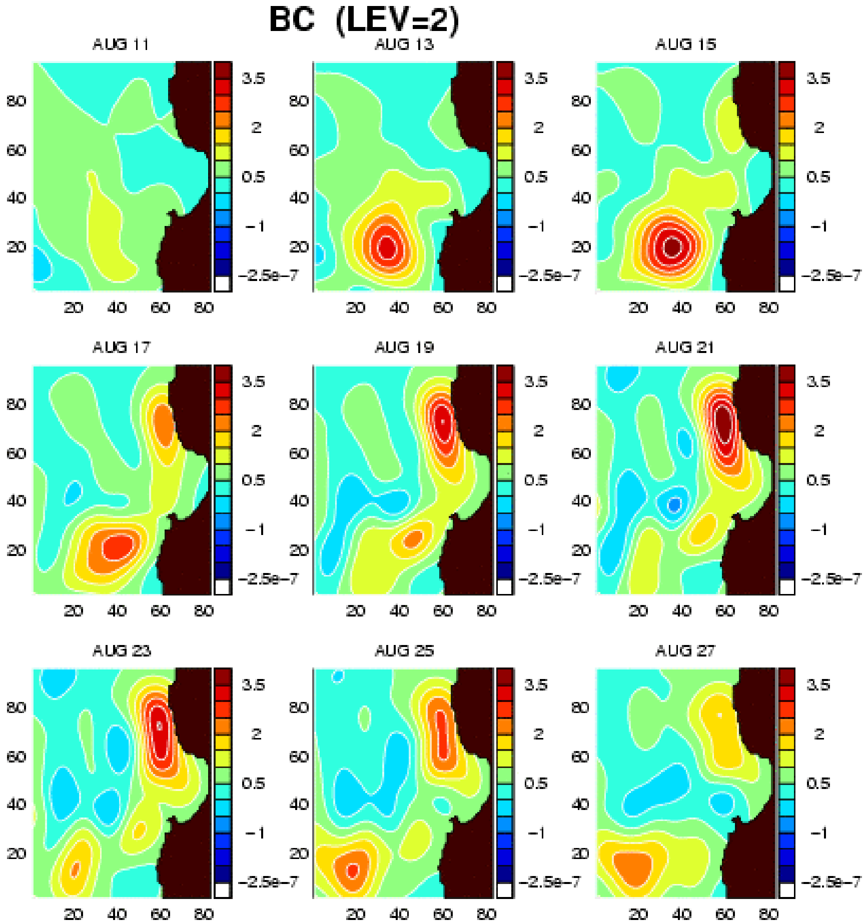


e. Objective Field: Multi-Scale Energy and Vorticity Analysis

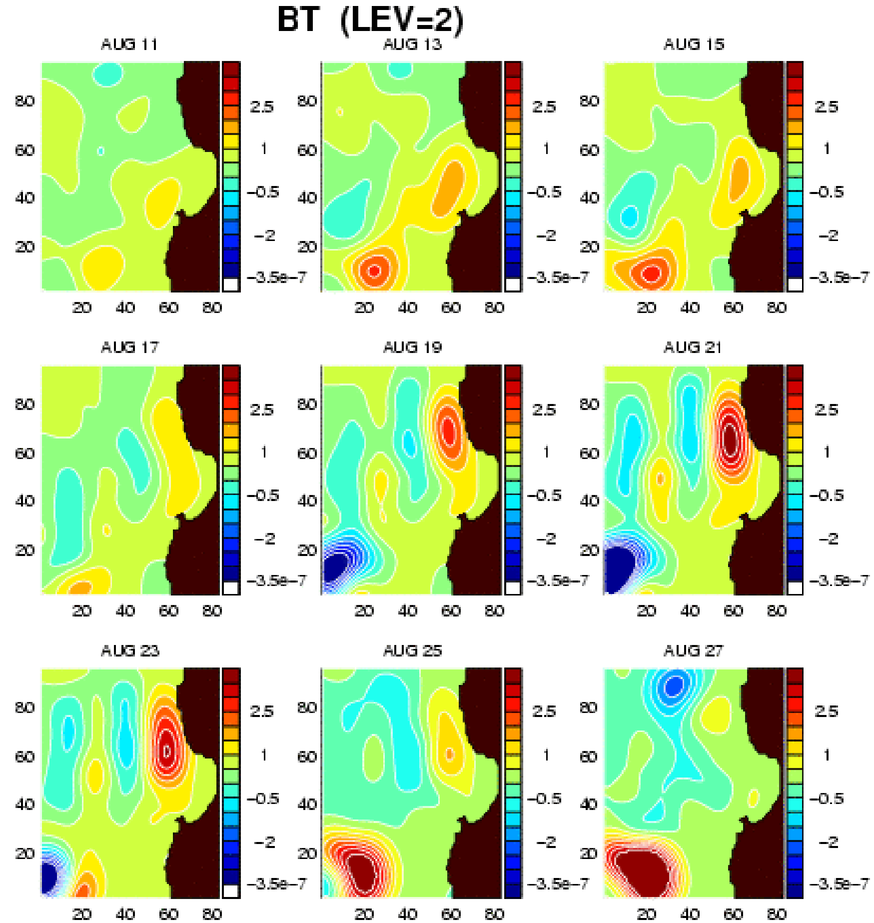
Two-scale window decomposition in space and time of energy eqns: 11-27 August 2003



Transfer of APE from large-scale to meso-scale



Transfer of KE from large-scale to meso-scale



- Center west of Pt. Sur: winds destabilize the ocean directly.
- Center near the Bay: winds enter the balance on the large-scale window and release energy to the meso-scale window during relaxation.

X. San Liang

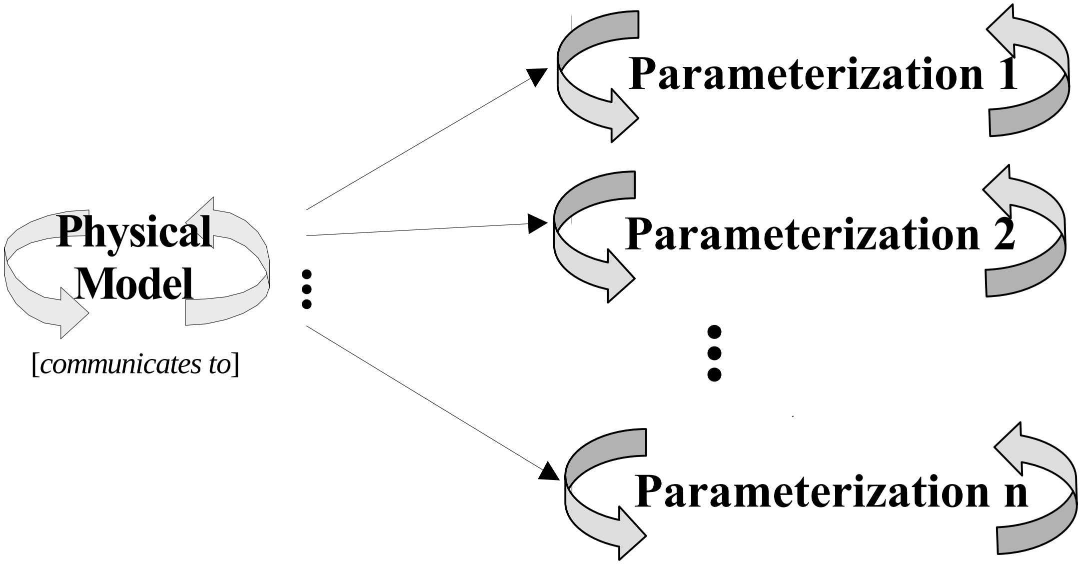
Oceanic Adaptive Modeling: Motivations and Concepts

- Physical and biogeochemical ocean dynamics can be intermittent and highly variable, and can involve interactions on multiple scales
- In general, oceanic fields and interactions that matter vary in time and space
- Model uncertainties can be (very) large, especially for biogeochemical processes

- For efficient forecasting, model structures and parameters should evolve and respond dynamically to new data injected into the executing prediction system
  - Correction of model biases
  - Comparison of competing models and better scientific understanding
  - Multi-model data assimilation
  - Automated evolution of model structures as a function of model-data misfits

- A model quantity (parameters, structures, state-variables) is said to be adaptive if its formulation, classically assumed constant, is made a function of data values
  - Physical regime transition (e.g., well-mixed to stratified) and evolving/unknown turbulent mixing parameterizations
  - Variations of biological assemblages with time and space (e.g., variable zooplankton dynamics, summer to fall phytoplankton populations, etc) and evolving biogeochemical rates and ratios

Towards Real-time Physical Adaptive Models



- Different Types of Adaptation:
  - Physical model with multiple parameterizations in parallel (hypothesis testing)
  - Physical model with a single adaptive parameterization (adaptive evolution)

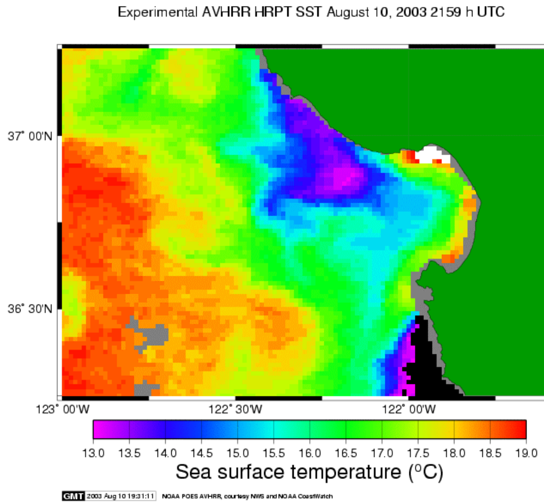
- Model selection based on quantitative dynamical/statistical study of data-model misfits

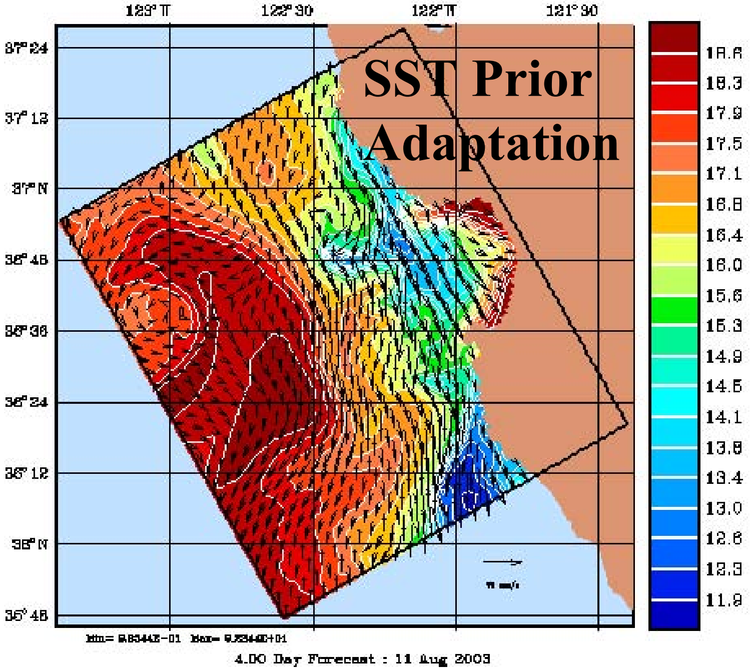
Quasi-Automated Real-time Physical Calibration during AOSN2

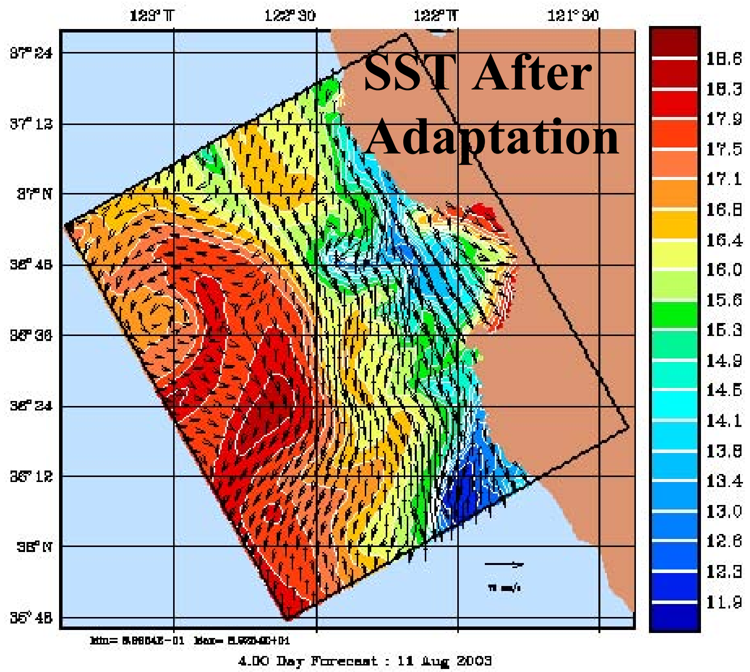
Prior to AOSN2, ocean models calibrated to historical conditions judged to be similar to these expected in August 2003.

Ten days in the experiment:

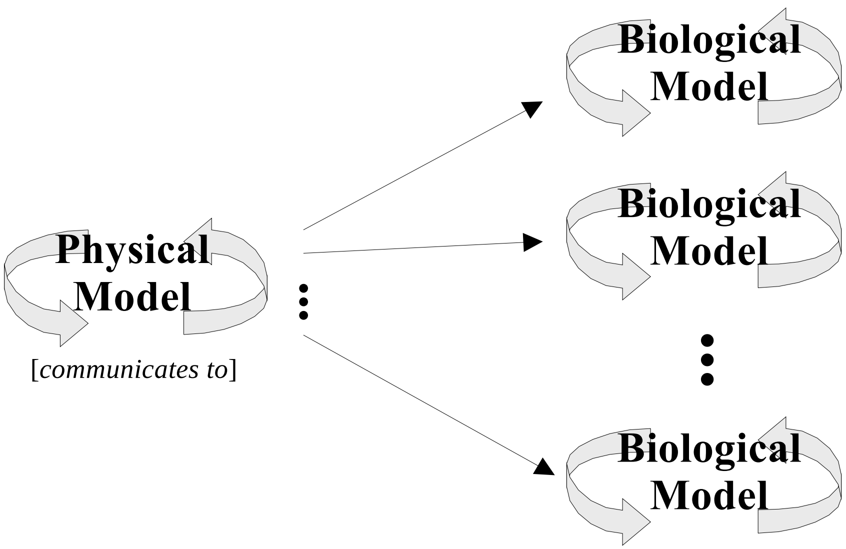
- Parameterization of the transfer of atmos. fluxes to upper layers (SBL mixing) adapted to new 2003 data
- 20 sets of parameter values and 2 mixing models tested
- Configuration with smallest RMSE/higher PCC improved upper-layer T and S fields, and currents

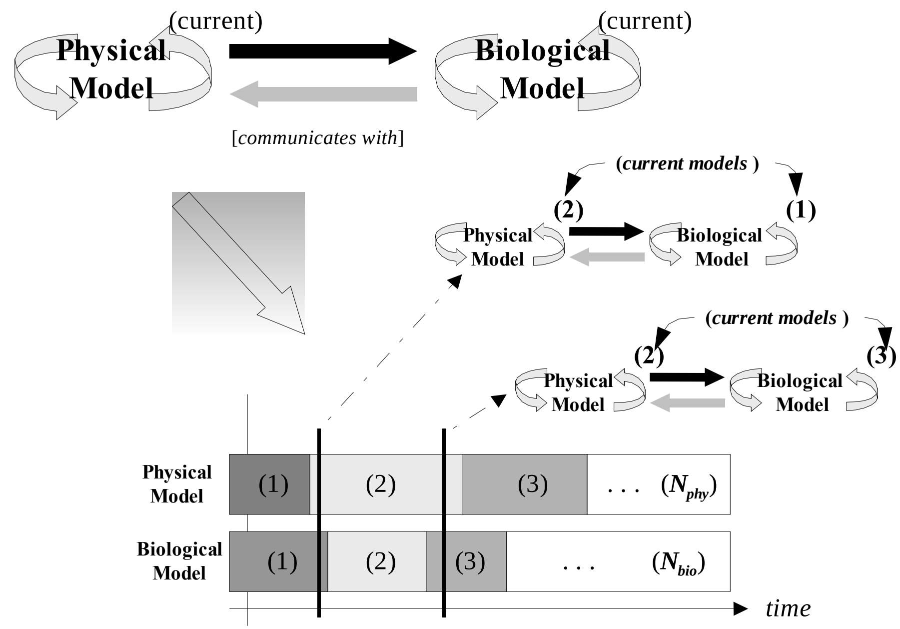






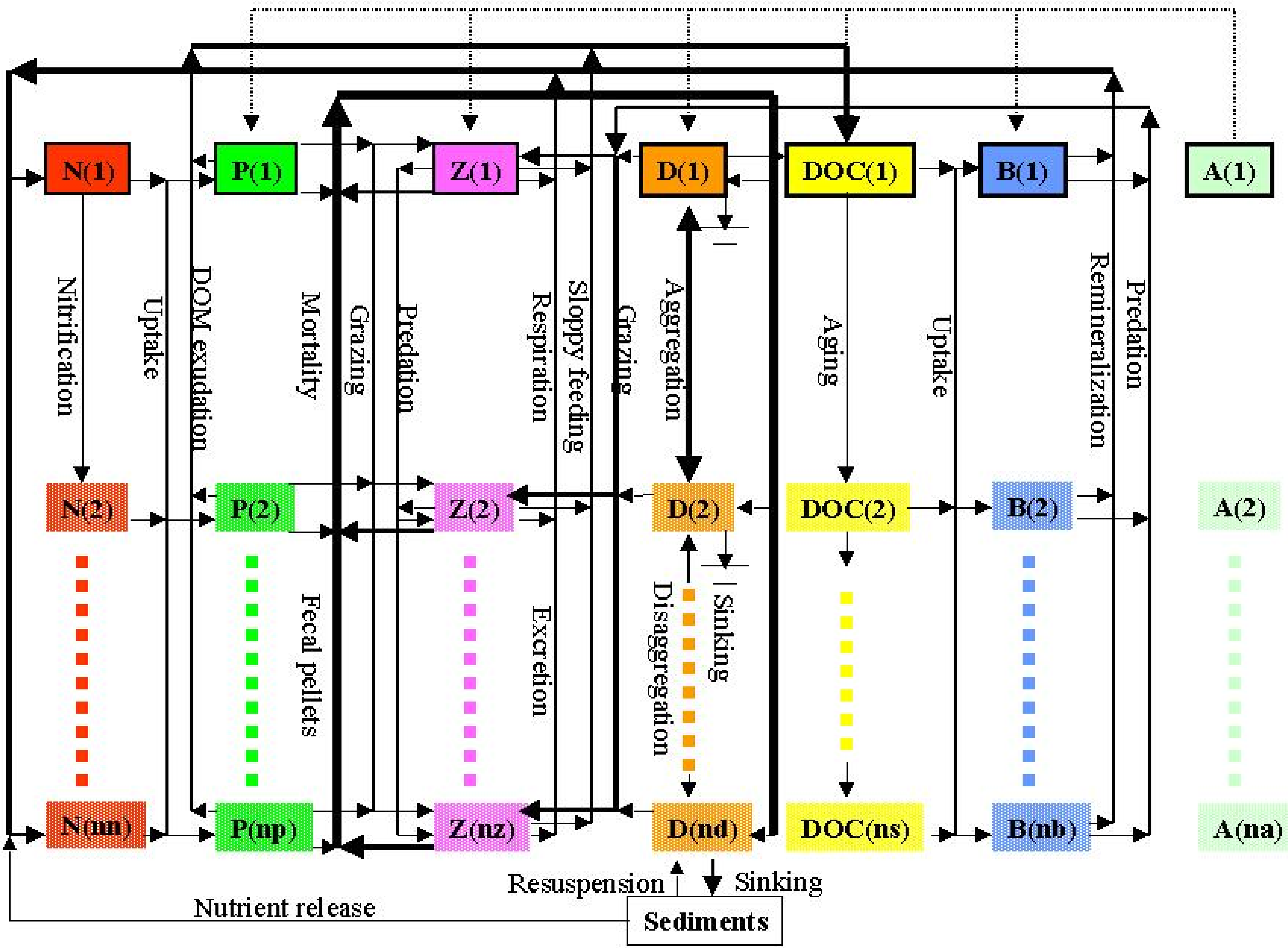
Towards Real-time Adaptive Coupled Models





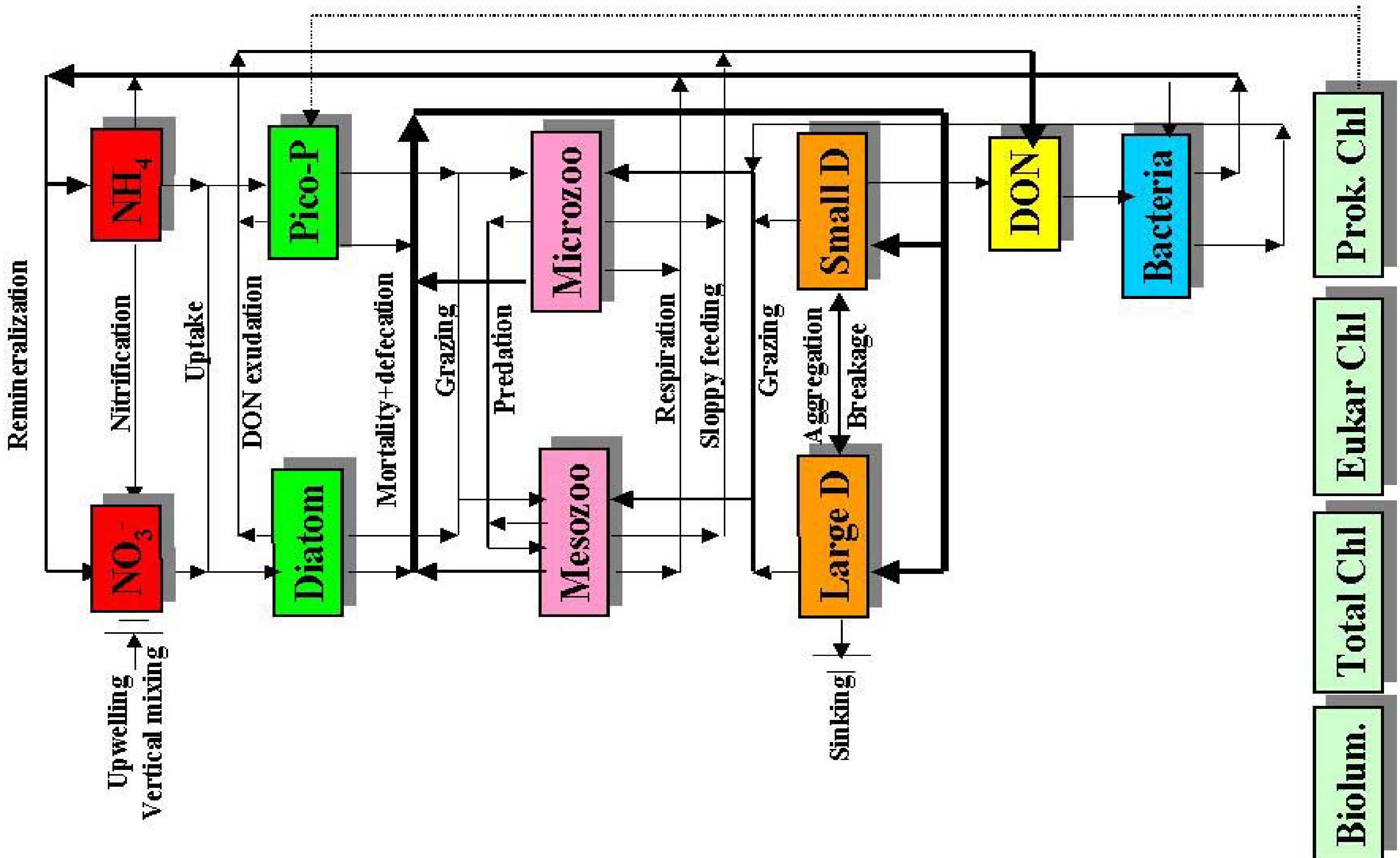
- Different Types of Adaptive Couplings:
  - Adaptive physical model drives multiple biological models (biology hypothesis testing)
  - Adaptive physical model and adaptive biological model proceed in parallel, with some independent adaptation
- Ongoing and Future Numerical Implementation
  - For performance and scientific reasons, both modes are being implemented using message passing for parallel execution
  - Mixed language programming (using C function pointers and wrappers for functional choices)

I. Generalized Adaptable Biological Model



(R.C. Tian, P.F.J. Lermusiaux, J.J. McCarthy and A.R. Robinson, HU, 2004)

A Priori Biological Model for Monterey Bay



Another configuration with PO_4 and $Si(OH)_4$

Example: Use P data to select parameterizations of Z grazing

Table 1. Parameterization of grazing on multiple types of prey with passive selection (g_{max}: maximum grazing rate; K: Half-saturation constant (but saturation constant in Eq. 1); P_0 threshold below which grazing is zero; p_i: preference coefficient; ? , *a*, ? : constant).

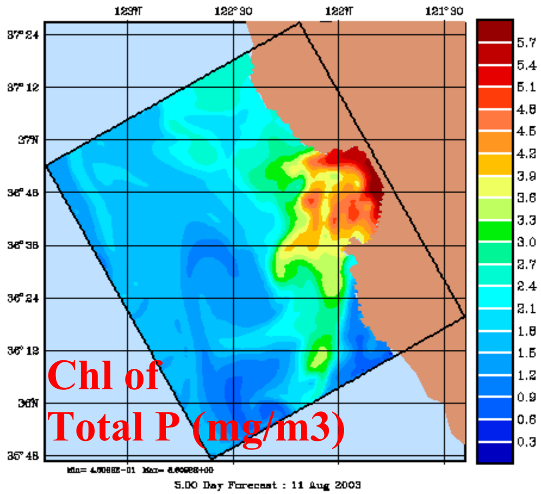
Function	References
(1) Rectilinear $g_i = \begin{cases} g_{max} \frac{p_i P_i}{K}, & for \quad R \le K \\ g_{max}, & for \quad R > K \end{cases}, \quad R = \sum_{i=1}^{n} p_i P_i$	Armstrong, 1994
(2) Ivlev function for each prey type: $g_i = g_{max}(1 - e^{-\alpha_i P_i})$	Leonard et al., 1999
(3) Ivlev function with interference between prey types: $g_i = g_{max}(1 - e^{-\alpha R})\frac{p_i P_i}{R}$, with $R = \sum_{i=1}^{n} p_i P_i$	Hofmann and Ambler, 1988
(4) Mechanistic disc function: $g_i = g_{max} \frac{a_i N_i}{1 + \sum_{j=1}^{n} a_j \tau_j N_j}$	Murdoch and Oaten, 1975; Holt, 1983; Gismervik and Anderson, 1997; Strom and Loukos, 1998
(5) Michaelis Menten Function: $g_i = g_{max} \frac{p_i P_i}{K + \sum_{j=1}^{n} p_j P_j}$	Murdoch, 1973; Real, 1977; Moloney and Field, 1991; Verity, 1991; Gismervik and Anderson, 1997; Strom and Loukos, 1998
(6) Threshold MM function: $g_i = g_{max}\left(\frac{R - P_0}{K + R - P_0}\right)\frac{p_i P_i}{R}$, with $R = \sum_{i=1}^{n} p_i P_i$	Evans, 1988; Lancelot et al., 2000
(7) Modified MM function: $g_i = g_{max} \frac{p_i P_i}{1 + \sum_{j=1}^{n} p_j P_j}$	Verity, 1991; Fasham et al. (1999) and Tian et al. (2001)

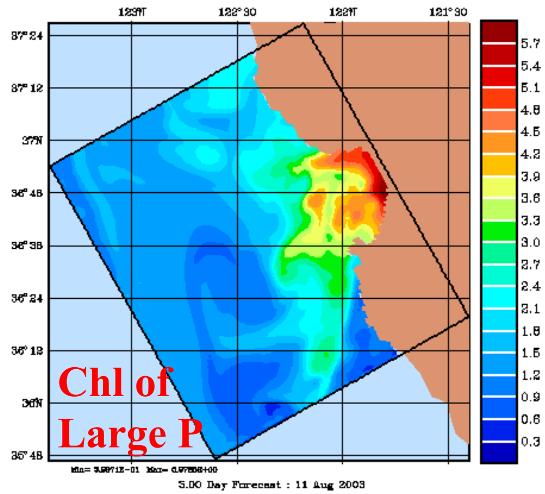
Table 2. Parameterization of grazing on multiple types of prey with active switching selection (g_{max}: maximum grazing rate; K: Half-saturation constant; P_0 threshold below which grazing is zero; p_i: preference coefficient; α, a, τ: constant).

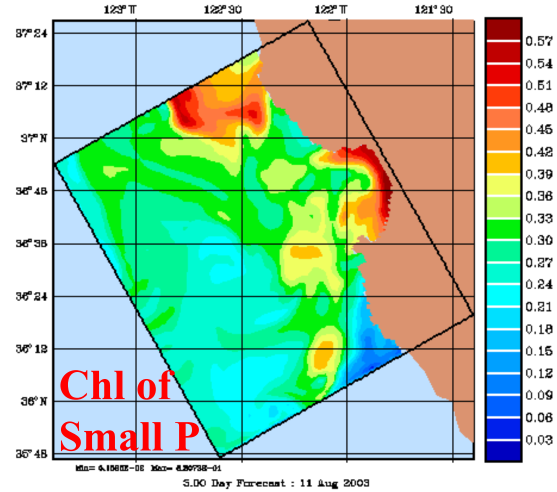
Function	References
(1) Switching MM predation: $g_i = g_{max} \frac{p_i P_i^2}{K \sum_{j=1}^{n} p_j P_j + \sum_{j=1}^{n} p_j P_j^2}$	Fasham et al., 1990; Strom and Loukos, 1998; Pitchford and Brindley, 1999; Spitz et al., 2001
(2) Mechanistic disc switching predation: $g_i = g_{max} \frac{b_i N_i^2}{(1 + c_i N_i)(1 + \sum_{j=1}^{n} \frac{b_j h_j N_j^2}{1 + c_j N_j^2})}$	Chesson, 1983
(3) Generalized switching function: $g_i = g_{max} a_i \frac{(p_i P_i)^m}{\sum_{i=1}^{n} (p_i P_i)^m}$	Tansky, 1978; Teramoto, 1979; Matsuda et al., 1986
(4) Generalized switching function: $g_i = g_{max} \frac{(p_i P_i)^m}{\left(\sum_{i=1}^{n} (p_i P_i)\right)^m}$	Vance, 1978
(5) Generalized switching MM function: $g_i = g_{max} \frac{(p_i P_i)^m}{1 + \sum_{i=1}^{n} (p_i P_i)^m}$	Gismervik and Andersen (1997)
(6) Generalized switching MM function: $g_i = g_{max} \frac{(p_i (P_i - P_{0i}))^m}{1 + \sum_{i=1}^{n} (p_i (P_i - P_{0i}))^m}$	This work

Towards automated quantitative model aggregation and simplification
(Older Simulations: done in Sep.-Oct. 2004)

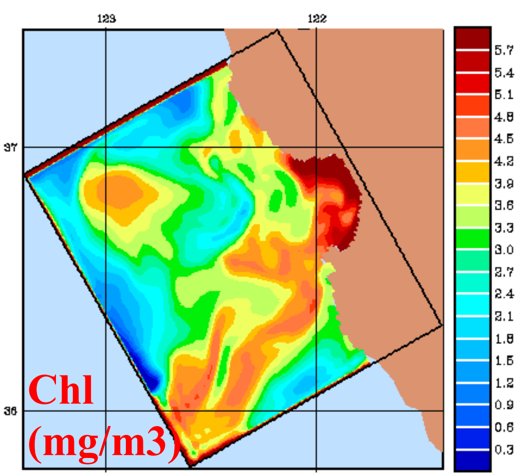
A priori configuration of generalized model on Aug 11 during an upwelling event

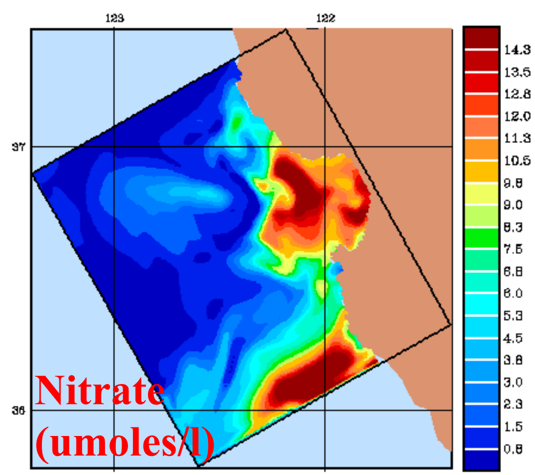


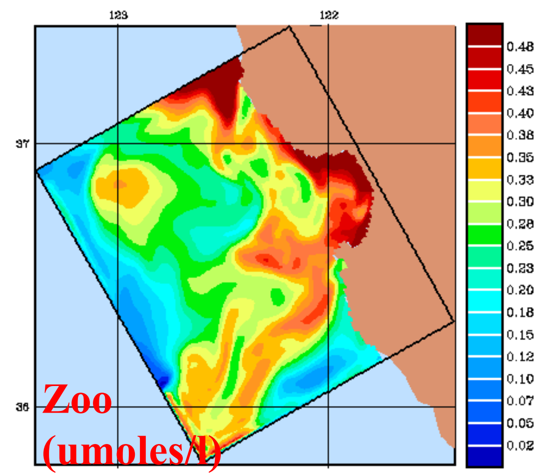




NPZ configuration of generalized model on Aug 11 during same upwelling event





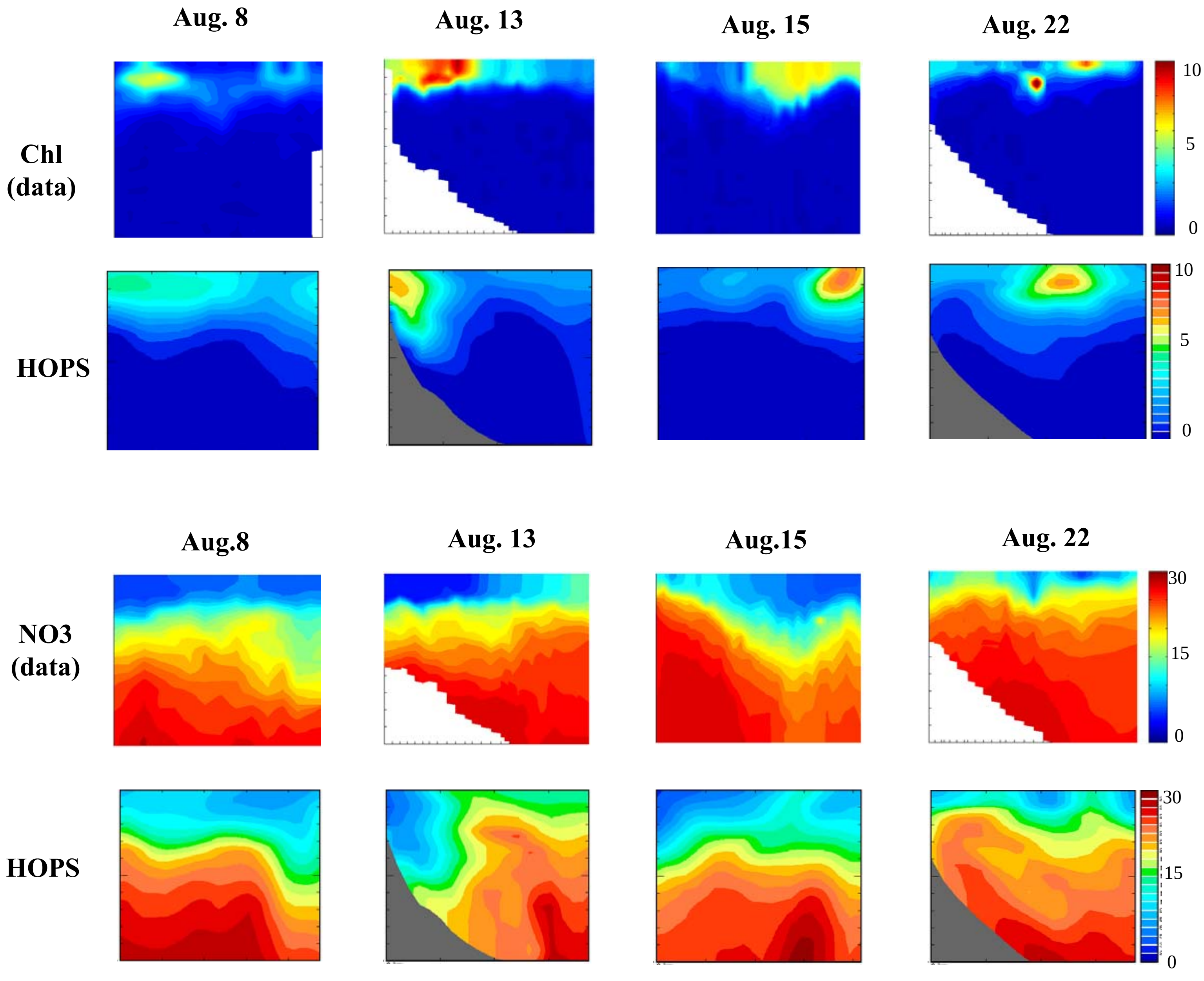


Dr. Rucheng Tian

Cross-Section in Chl μg/l and NO_3 (phys. DA, no bio. DA)

Observations (S. Haddock et al) vs. Simulations

Aug 06 - Aug 18: Upwelling
Aug 19 - Aug 23: Relaxation
Aug 27 - Aug 30: Upwelling



- **Several Chl hot-spots position and amplitudes, and nutricline tilts, captured but bio. model vertical resolution not sufficient**

PROCESSES

1. **Deeper nutricline and stronger blooms during upwelling**
2. **Much smaller scale hot-spots and shallower nutricline during relaxation due to ocean-driven (sub)-mesoscales**

Forecast Error Analyses and Optimal (Multi) Model Estimates

- Forecast Error Analyses: *Learn individual model forecast errors in an on-line fashion from model-data misfits based on Maximum-Likelihood*
- Model Fusion: *Combine models via Maximum-Likelihood based on the current estimates of their forecast errors*

3-steps strategy, using model-data misfits and error parameter estimation

1. Select appropriate/convenient forecast error parameterization
   - Approximate forecast error covariances and biases models as efficient parametric family:

$$\mathbf{B} \approx \tilde{\mathbf{B}}(\boldsymbol{\alpha}); \qquad \boldsymbol{\mu} \approx \tilde{\boldsymbol{\mu}}(\boldsymbol{\beta});$$

   - Limit number of free parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ *(for now: error length scale and variance)*

2. Adaptively determine forecast error parameters from model-data misfits based on the Maximum-Likelihood principle:

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} p(\boldsymbol{\mathcal{Y}}|\boldsymbol{\Theta})$$

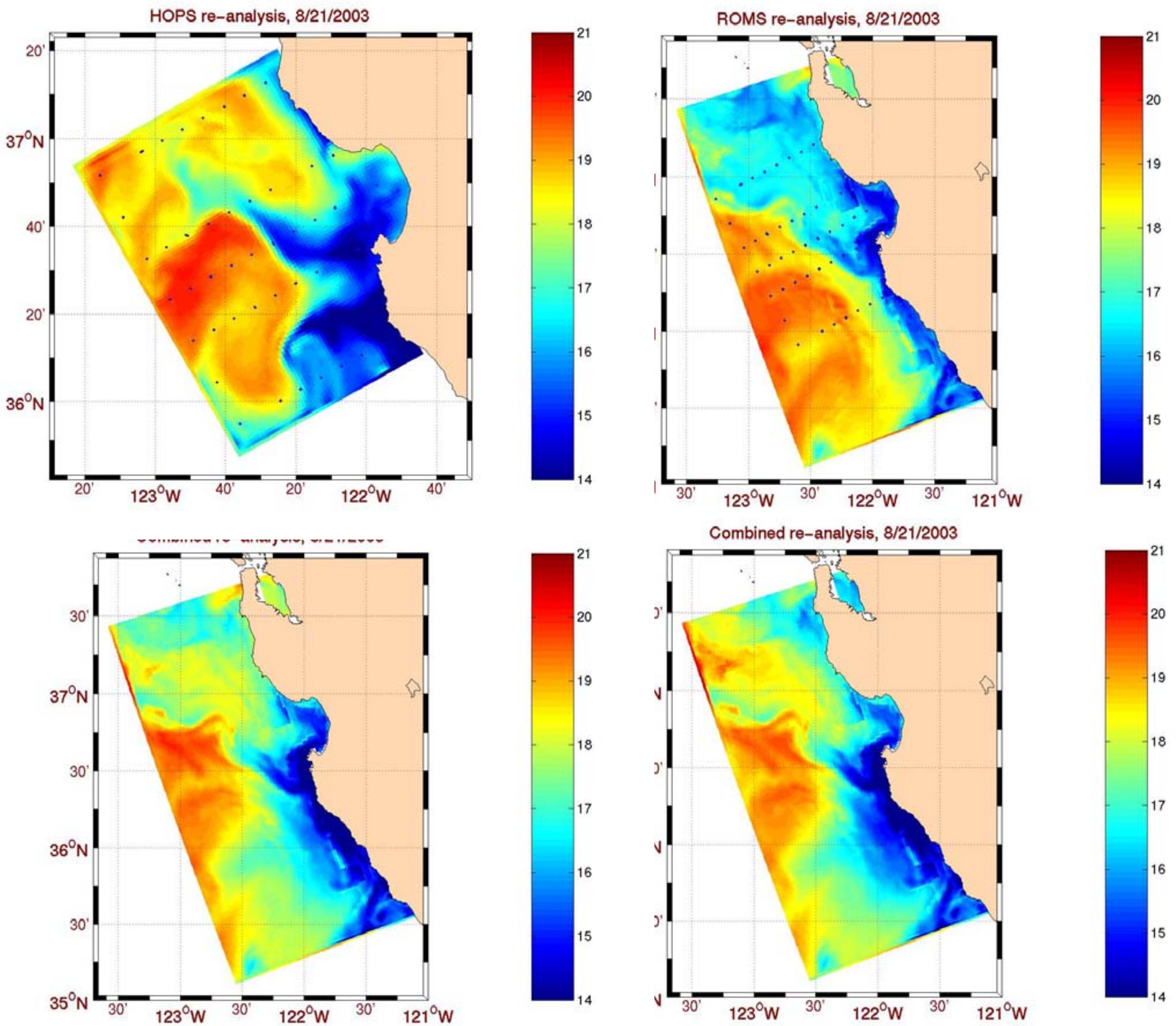
   $\boldsymbol{y}$ is the data, $\boldsymbol{\Theta}$ the vector of $\boldsymbol{\alpha}$'s and $\boldsymbol{\beta}$'s of each model

3. Combine model forecasts via Maximum-Likelihood based on the current estimates of error parameters

O. Logoutov

Forecast Error Analyses and Optimal (Multi) Model Estimates

Two-Model Forecasting Example



HOPS and ROMS SST forecast

Left – HOPS (re-analysis)

Right – ROMS (re-analysis)

Combined SST forecast

Left – with *a priori* error parameters

Right – with Maximum-Likelihood error parameters

CONCLUSIONS and FUTURE

- ESSE powerful nonlinear scheme for interdisciplinary estimation of oceanic state variables and error fields via multivariate physical-biogeochemical-ecosystem-acoustical data assimilation
- Main results shown:
  - Lagoon of Venice: Estimated time-scales/memory of ecosystem, seasonal field evolutions and new assessment of water quality
  - Mass Bay/ Monterey Bay: Smoothing and biogeochemical dominant dynamical balances
  - Monterey Bay: Error Forecasting (fields and LCS), Adaptive Sampling,Adaptive Modeling
- Entering a new era of fully interdisciplinary oceanic dynamical system science, combining models and data
- Multiple novel and challenging opportunities, for example:
  - Quantitative assimilation feedbacks, e.g. via Adaptive (Bayesian) estimation/learning
    - Adaptive modeling/system identification (optimal parameters, structures, state variables and multi-model combinations, model validation/verification)
    - Adaptive sampling (optimal data type, quantity and time-space locations)
    - Adaptive model reductions and simplifications
  - Theory and applications of environmental ocean science
    - Dominant dynamical balances for fundamental understanding, and for weak constraints