Overview of Data Assimilation in Ecosystem Modeling

Yvette H. Spitz

Oregon State University, College of Oceanic and Atmospheric Sciences
Corvallis, OR 97331
yvette@coas.oregonstate.edu
Issues with ecosystem modeling -> data assimilation

• Need to define the number of variables to represent a given ecosystem (3 to 32 variables) and the model pathways

• Need to define the parameterization of the various links in the ecosystem

BUT

• Most of the model parameters are little known and some of them not measurable. Large uncertainties in the pathways
Data assimilation and ecosystem modeling (0D to 1D)

Parameter estimation (a few examples)
- Ocean Station P: Matear (1995), Prunet et al. (1996)
- Mesocosm experiments: Vallino (2000)

Parameter estimation and model configuration
- Central Equatorial Pacific: Friedrichs (2001)
Forcing and flux estimates

- River discharge and advection/diffusion: Spitz et al. (2005)

Model comparisons and validations - JGOFS Testbed

- Friedrichs et al. (2007)
parameter estimation
and
model configuration
Data assimilation technique

Variational adjoint technique

First Guess Parameters

Data

Model

Adjoint Model

Cost Function

Gradient

Descent direction line search

New parameters

Stop if grad J < ε

Optimization Package

(N1QN3, Gilbert and Lemarechal)
Nitrogen based Ecosystem model

Fasham et al. (1990)
(29 parameters)

\[ \frac{\partial P}{\partial t} = J(Q_1 + Q_2)P - \gamma_1 J(Q_1 + Q_2)P - \mu_1 P - G_1 - P(m+h^+)/MLD \]
Two U.S. JGOFS long term time series

Available data to constrain the system

Primary productivity = $\left(1 - \gamma_1\right) J (Q_1+Q_2) P$

Bacteria productivity = $U_1 + U_2$

PON = $P + Z + B + DET$
We now have 50 parameters

Spitz et al, 2001
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Original model</th>
<th>Orig. mod. + ass.</th>
<th>Modif. mod. + ass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll</td>
<td>0.34 (-0.33)</td>
<td>0.09 (-0.23)</td>
<td>0.04 (0.86)</td>
</tr>
<tr>
<td>Nitrate</td>
<td>0.28 (0.87)</td>
<td>0.12 (0.84)</td>
<td>0.05 (0.98)</td>
</tr>
<tr>
<td>PON</td>
<td>0.60 (0.48)</td>
<td>0.11 (0.36)</td>
<td>0.12 (0.76)</td>
</tr>
<tr>
<td>Bacteria</td>
<td>0.16 (0.01)</td>
<td>0.08 (-0.37)</td>
<td>0.02 (0.56)</td>
</tr>
</tbody>
</table>

Month of year: J M M J S N
Annual fluxes for the upper mixed-layer

Original model

New model with data assimilation
Estimation of model parameters and configuration of model pathways using the upper mixed-layer observations at HOT

Without data assimilation
Ecosystem model for HOT

DON/DOC → Phytoplankton
Phytoplankton → Chlorophyll-a
Chlorophyll-a → Mesozoo.
Mesozoo. → N2 fixation
N2 fixation → DON/DOC

Ammonium → Bacteria
Bacteria → Nano/Microzoo.
Nano/Microzoo. → Detritus
Detritus → Nitrate
Nitrate → DON/DOC
Estimation of model parameters and configuration of model pathways using the upper mixed-layer observations at HOT

Without data assimilation

With data assimilation and nitrogen fixation
Dinitrogen fixation at HOT

- HOT observations (Karl et al., 1997)
- Model estimates

• Using observed trichome abundances: $21.90 \pm 10.95$ mmol N m$^{-2}$ yr$^{-1}$
• From the model results: $25.81 \pm 15.32$ mmol N m$^{-2}$ yr$^{-1}$
# 16 Parameters that differ by more than 20%

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>BATS</th>
<th>HOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_p$</td>
<td>phytoplankton maximum growth rate</td>
<td>3.6700</td>
<td>1.5900</td>
</tr>
<tr>
<td>$k_1$</td>
<td>half-saturation for phyto NO3 uptake</td>
<td>0.5970</td>
<td>0.4200</td>
</tr>
<tr>
<td>$k_2$</td>
<td>half-saturation for phyto NH4 uptake</td>
<td>0.0339</td>
<td>0.0532</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>phytoplankton specific mortality rate</td>
<td>0.0722</td>
<td>0.1100</td>
</tr>
<tr>
<td>$\beta_b$</td>
<td>zoo assimilation efficiency of bact</td>
<td>0.8700</td>
<td>0.6900</td>
</tr>
<tr>
<td>$\beta_d$</td>
<td>zoo assimilation efficiency of det</td>
<td>0.4500</td>
<td>0.6400</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>zooplankton specific excretion rate</td>
<td>0.1250</td>
<td>0.0790</td>
</tr>
<tr>
<td>ggen</td>
<td>bact gross growth eff. of nitro DOC</td>
<td>0.6280</td>
<td>0.2230</td>
</tr>
<tr>
<td>$V_{b_amino}$</td>
<td>bacteria max. growth amino DOM</td>
<td>0.2530</td>
<td>0.8590</td>
</tr>
<tr>
<td>$k_{don}$</td>
<td></td>
<td>1.3390</td>
<td>0.9970</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>large zoo. assimilation eff. of zoo</td>
<td>0.4399</td>
<td>0.8910</td>
</tr>
<tr>
<td>cnp</td>
<td>C:N for phyto</td>
<td>5.1600</td>
<td>6.1800</td>
</tr>
<tr>
<td>cnz</td>
<td>C:N for zoo</td>
<td>5.0200</td>
<td>3.7800</td>
</tr>
<tr>
<td>cnb</td>
<td>C:N for bact</td>
<td>5.3800</td>
<td>6.6400</td>
</tr>
<tr>
<td>$V$</td>
<td>sinking of detritus</td>
<td>1.3498</td>
<td>0.6424</td>
</tr>
<tr>
<td>No</td>
<td>Minimum NO3 at 100 m</td>
<td>0.3951</td>
<td>0.2443</td>
</tr>
</tbody>
</table>
Ecosystem model coupled to a 3D circulation model

Model 7: Spitz et al., 2001
parameter estimation
and
Physical forcing
Project AMORE (ULb, MUMM, VUB, OSU)
Advanced MOdelling and Research on Eutrophication

The most visible symptom of eutrophication: Phaeocystis foam on Middlekerke Beach (photo by J.P. Mommaerts)

http://www.ulb.ac.be/assoc/esa/AMORE
Observations at Station 330 between 1988 and 1999
MIRO ECOSYSTEM STRUCTURE

126 parameters to estimate

Lancelot et al., 2005
Twin experiments, modeled observations

The model is run for

1) the western channel (WCH)

2) the French coastal zone with the Seine discharge (FCZ) and input from WCH

3) the Belgian coastal zone with the Scheldt discharge (BCZ) and input from FCZ

Each region is characterized by its own area, depth, water temperature and light conditions
Model results and observations

First guess
(parameters from lab. exp. and guesses)

+ Observations

After twin experiments and changing the most sensitive parameters + modification of some pathways
Impact of the three-dimensional flow field on the ecosystem response

In a 3D coupled physical-biological model that will allow to represent the spatial and temporal behavior of the ecosystem, we have

$$B(t) = B(t-1) - \text{advection} + \text{diffusion} + \text{biological (source+sink)}$$

We will consider that the advection and diffusion terms are unknown. They will be estimated via data assimilation.
Assimilation of the 3D surface model observations and estimation of the forcing

---

---

---

---

---

---

---

---
Recovery with the 0D model of the physical contribution to the concentrations when assimilating the 3D model observations

- Physical contribution from the 3D model simulation
- Recovery with the 0D model and data assimilation

Units: concentration per hour
Recovery with the 0D model of the physical contribution to the concentrations when assimilating the 3D model observations

Physical contribution from the 3D model simulation

Recovery with the 0D model and data assimilation

Units: concentration per hour
Estimation of model parameters and advection/diffusion

Estimated parameters agree well with new lab. experiments

Advection/diffusion terms remain to be verified with a 3D simulation using the new set of parameters
Coupling to 3D circulation model of the North Sea

Lacroix et al. (2007)
Model Intercomparison

U.S. JGOFS Testbed

Marjorie Friedrichs

Larry Anderson
Fei Chai
Scott Doney
Jeff Dusenberry
Raleigh Hood
Dennis McGillicuddy
Yvette Spitz

Rob Armstrong
Jim Christian
John Dunne
Masahiko Fujii
John Klinck
Markus Schartau
Jerry Wiggert

Friedrichs et al, 2007 (JGR)
Experiment 1: Individual assimilation

Experiment 2: Simultaneous assimilation  
Which types of models reproduce mean PP and chl?

Experiment 3: Cross validation  
Which types of models are most portable?
Initial model-data comparison (pre-assimilation)
EqPac + Arabian Sea

Model data misfit

= chlorophyll
= export flux
= primary production
= nitrate
= zooplankton

Model Number (increasing complexity)

LST 1 2 3 4 5 6 7 8 9 10 11 12
Cost function comparison Expt 1 & 2

Expt 1
- Indiv EP
- Simultaneous
- Indiv AS

Expt 2

Model Number (increasing complexity)
Production vs Chlorophyll: Expt. 2

- Models with multiple P size classes are better able to reproduce chl
- Nearly all models underestimate production
- No relationship between number of P (or Z) compartments, and how well production is reproduced
Production vs Chlorophyll: Expt. 2

- Only models with iron are capable of reproducing observed PP in EqPac, and these models do much better in AS as well.
Cost function comparison Expt 1 & 3

![Cost function comparison graph]

- **Expt 1**: individ EP, individ AS

Model Number (increasing complexity)

Cost function
- Models with more P and Z size classes are not necessarily more portable than models with single P and Z size classes.
Conclusions of first part

- The variational adjoint method is a powerful tool to estimate model parameters and missing ecosystem pathways
  - For each system, we found a set of parameters that is representative of the species
  - We found that the pathways are different for the two oligotrophic environments
  - We were able to estimate the local forcing for the station 330.

- Success at estimating parameters and reducing the cost function, but what does it mean?

- Caution should be taken when using this technique: definition of the cost function, the cost function minimum should be a global minimum, availability of the observations.

- In all the examples, 0D or 1D. But if we look at basin scale problem we cannot anymore use the variational adjoint method.
Data assimilation and ecosystem modeling (3D)

**parameter estimation**
- Losa et al. (2004)

**state estimation (Kalman filter type)**
- Carmillet et al. (2001)
- Eknes and Evensen (2002)
- Natvik and Evensen (2003)

**error of representation**
- Richman et al. (2005)
3D coupled ecosystem circulation model and error estimates
Atmospheric forcing

• Wind stress and non solar radiation affect directly the circulation and indirectly the ecosystem

• Solar radiation affects directly the ocean circulation and ecosystem (i.e. photosynthesis)

Circulation model

• Mixing scheme, grid resolution etc

Ecosystem model

• Parameters and pathways
Wind Stress (dyne cm\(^{-2}\))

Mean Difference between NCEP/DOE and NCEP/NCAR (1992-2001)

Downward Short Wave Radiation (W m\(^{-2}\))

Mean Difference between NCEP/DOE and NCEP/NCAR (1992-2001)

The mean varies between 100 and 220 W m\(^{-2}\)
Correlation between NCEP/DOE and NCEP/NCAR (1992-2001)

Surface temperature

Surface Chla
Correlation between model chlorophyll-a simulations

**Correlation between**

*NCEP/DOE* and *NCEP/NCAR*

**Correlation between**

*BATS* and *HOT* parameters
Large number of state variables prohibits solving the full system

→ **Reduced State Space Kalman Filter** *(Richman et al, 2005)*

1) Compute the **multivariate empirical orthogonal functions** (EOF’s) of our 23 year time series of deviations from the seasonal cycle,

2) A statistical test is performed in order to estimate the number of significant degrees of freedom. *(Preisendorfer (1988)) (35 modes accounting for 59% of the total variance)*

3) Recast the Kalman filter problem in terms of a Reduced State Space of approximately 35 EOFs instead of $10^5$ discrete points

4) We estimate the **multivariate model error covariance** $P^f$ by performing linear regressions to fit the EOF's of the SST model data misfits with the temperature components of the model multivariate EOF's.

5) Using the estimated model covariance, we calculate the Kalman gain and the update the model to combine with the observations.
Model and AVHRR-Seasonal Anomaly of SST First EOF

- **First Spatial EOF**
  - Model SST
    - 7%
  - AVHRR SST
    - 4%
  - AVHRR-Model
    - 3%

- **First Temporal Amplitude**
  - 1980 to 2000
Correlation before and after assimilation of AVHHR SST
**Representation error**

The Kalman filter blending of the model and the observations made a modest improvement of the model outputs.

*Why was not there a bigger impact?*

The model cannot represent all of the variability observed in the data.

Using the Reduced State Space, we can estimate this error of representation.

The difference between the model data misfit and the EOF representation of this misfit (error of representation) gives us information on where improvement is needed.
EOFs of the model temperature error of representation.

Numbers in upper right corners of panels are percent of total variance.
Conclusions of second part

- Remotely sensed observations are the only data that allow us tackle the problem of model error of representation on spatial and temporal scales needed to address climate impact.

- Data assimilation, such as Kalman filter (Reduced State Space) can help us to estimate the error of representation of the circulation model. This technique will be applied to estimate these errors in the ecosystem model and the coupled circulation/ecosystem model.

- Observations on longer times and of new kinds will help improve model and reduce the errors. Climatic change assessment will also be improved.
Conclusions

• Long term time series (ORION, IOOS etc) and remote sensed observations are needed to estimate model parameters, state variables and to calibrate the model results, BUT what are we really missing from space and mooring, gliders?