MAR513-Lecture 6: Data Assimilation

Why do we need the data assimilation?

- Models are not perfect. Most of ocean models even do not resolve the realistic world in both time and space scales.
- Data are limited in space and time, and also at most times they are not accurate enough.
- Increase needs for the ocean forecast, particularly in coastal oceans.

How could we achieve our goal? Does it achievable? What critical points do we need consider?

- The model should be capable of reproducing right physics and simulating the fields of currents and water property at a certain accuracy. The data assimilation should be used to improve the accuracy rather than to add additional dynamics into the system. It is particularly true for the purpose of the forecast application.

- Data is usually insufficient, the data assimilation does not always work as one expects.

- Advanced data assimilation methods usually requires a huge computational power.

- Any data assimilation method should be tested for its reliability and capability of improving the simulation using the so-called "twin experiments".

Data Assimilation Methods

1. Nudging-directly merge model-predicted values to observation given a priori statistical assumption about the model noise and errors in the observation data

2. Optimal interpolation (OI)-uses the error covariance of the observations and model predictions to find their most likely linear combination. OI requires a priori statistical assumptions of the model noise and observational errors

3. Adjoint (variational) methods-based on control theory, in which a cost function, defined by the difference between model-derived and measured quantities, is minimized in a least-square sense under the constraint that the governing equations of the model remain satisfied

4. Kalman filters (RRKF, EnKF, EnSKF, EnTKF, SEIK)-the most sophisticated statistical approaches through the Kalman gain.

Nudging method

$\alpha(x,y,z,t)$ is a variable selected to be assimilated;
$F(\alpha, x,y,z,t)$ is the sum of all the terms in the governing equation of $\alpha(x,y,z,t)$;

$$\frac{\partial \alpha(x,y,z,t)}{\partial t} = F\left(\alpha,x,y,z,t\right) + G_\alpha \frac{\sum_{i=1}^{N} W_i^2(x,y,z,t)\gamma_i\left(\alpha_o - \alpha_m\right)_i}{\sum_{i=1}^{N} W_i\left(x,y,z,t\right)}$$

where α_o is the observed value; α_m is the model predicted value; N is the number of observational points within the search area; γ_i is the data quality factor at the ith observational point with a range from 0 to 1; G_a is the nudging factor that keeps the nudging term to be scaled by the slowly physical adjustment process. G_a must satisfy the numerical stability criterion given by

$$G_a < 1/\Delta t$$

$W_i\ (x,y,z,t)$ is a product of horizontal, vertical, temporal and directional weight functions given as

$$W_i(x,y,z,t) = w_{xy} \bullet w_\sigma \bullet w_t \bullet w_\theta$$

Weight functions

$$w_{xy} = \begin{cases} \dfrac{R^2 - r_o^2}{R^2 + r_o^2}, & 0 \le r_o \le R \\ 0, & r_o > R \end{cases}$$

R is the search radius
r_o is the distance from the location where the data exist

$$w_\sigma = \begin{cases} 1 - \dfrac{|\sigma_{obs} - \sigma|}{R_\sigma}, & |\sigma_{obs} - \sigma| \le R_\sigma \\ 0, & |\sigma_{obs} - \sigma| > R_\sigma \end{cases}$$

R_o is the vertical search range

$$w_t = \begin{cases} 1, & |t - t_o| < T_w/2 \\ \dfrac{T_w - |t - t_o|}{T_w/2}, & T_w/2 \le |t - t_o| \le T_w \\ 0, & |t - t_o| > T_w \end{cases}$$

T_w is the half assimilation window

$$w_\theta = \frac{\big||\Delta\theta| - 0.5\pi\big| + c_1\pi}{(0.5 + c_1)\pi}$$

$\Delta\theta$ is the directional difference between the local isobath and the computational point with a c_1 constant ranging from 0.05 to 0.5.

The OI method

Let X_f, X_a and X_o be the model forecast, assimilated (analysis) and observed values of a model variable X, respectively, and assume that they satisfy a linear relationship given as

$$X_a = X_f + \sum_{k=1}^{M} a_k (X_{o,k} - X_{f,k})$$

where M is the total data points involved in the optimal interpolation for X at a node point.

Defining that the true value of X is X_T at the assimilated node and $X_{T,k}$ at the kth observed point, $e_a = X_a - X_T$; $e_f = X_f - X_T$; $e_{o,k} = X_{o,k} - X_{T,k}$; and $e_{f,k} = X_{f,k} - X_{T,k}$, the analysis error e_a is equal to

$$e_a = e_f + \sum_{k=1}^{M} a_k (e_{o,k} - e_{f,k})$$

The analysis error covariance $P_a = e_a^2$, which is given as

$$P_a = [e_f + \sum_{k=1}^{M} a_k(e_{o,k} - e_{f,k})][e_f + \sum_{k=1}^{M} a_k(e_{o,k} - e_{f,k})$$

$$= e_f^2 + 2\sum_{k=1}^{M} a_k(e_f e_{o,k} - e_f e_{f,k}) + [\sum_{k=1}^{M} a_k(e_{o,k} - e_{f,k})]^2$$

In the least square fitting method, the error in e_a must be a minimum when the first differentiation condition of $\partial P_a / \partial a_k = 0$ is satisfied, i.e.,

$$\begin{cases} (e_{o,1} - e_{f,1})\sum_{k=1}^{M} a_k(e_{o,k} - e_{f,k}) = (e_f e_{f,1} - e_f e_{o,1}) \\ (e_{o,1} - e_{f,2})\sum_{k=1}^{M} a_k(e_{o,k} - e_{f,k}) = (e_f e_{f,2} - e_f e_{o,2}) \\ \vdots \\ \vdots \\ (e_{o,M} - e_{f,M})\sum_{k=1}^{M} a_k(e_{o,k} - e_{f,k}) = (e_f e_{f,M} - e_f e_{o,M}) \end{cases}$$

Assuming that $e_{f,k}$ is not correlated with $e_{o,k}$, then we have

$$\begin{cases} \sum_{k=1}^{M} a_k (e_{o,k} e_{o,1} + e_{f,k} e_{f,1}) = e_f e_{f,1} \\ \sum_{k=1}^{M} a_k (e_{o,k} e_{o,2} + e_{f,k} e_{f,2}) = e_f e_{f,2} \\ \vdots \\ \vdots \\ \sum_{k=1}^{M} a_k (e_{o,k} e_{o,M} + e_{f,k} e_{f,M}) = e_f e_{f,M} \end{cases}$$

or

$$\begin{cases} (e_{o,1}^2 + e_{f,1}^2) a_1 + (e_{o,2} e_{0,1} + e_{f,2} e_{f,1}) a_2 \cdots + (e_{o,M} e_{o,1} + e_{f,M} e_{f,1}) a_M = e_f e_{f,1} \\ (e_{o,1} e_{o,2} + e_{f,1} e_{f,2}) a_1 + (e_{o,2}^2 + e_{f,2}^2) a_2 \cdots + (e_{o,M} e_{o,2} + e_{f,M} e_{f,2}) a_M = e_f e_{f,2} \\ \vdots \\ \vdots \\ (e_{o,1} e_{o,2} + e_{f,1} e_{f,2}) a_1 + (e_{o,2} e_{o,M} + e_{f,2} e_{f,M}) a_2 \cdots + (e_{o,M}^2 + e_{f,M}^2) a_M = e_f e_{f,M} \end{cases}$$

This can be written in matrix form as

$$\hat{P} \cdot \hat{a} = \hat{f}$$

where

$$\hat{P} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1M} \\ P_{21} & P_{22} & \cdots & P_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ P_{M1} & P_{M2} & \cdots & P_{MM} \end{pmatrix}; \; \hat{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix}; \; \hat{f} = \begin{pmatrix} e_f e_{f,1} \\ e_f e_{f,2} \\ \vdots \\ e_f e_{f,M} \end{pmatrix}$$

and

$$P_{i,k} = e_{o,i} e_{o,k} + e_{f,i} e_{f,k}, \; i = 1, 2, \ldots.\mathrm{M}; \; k = 1,2,\ldots..\mathrm{M}.$$

When the observational and forecast error covariance values are known or specified, parameter a_k can be determined by using a state-of-the-art linear algebraic equation solver to solve the matrix.

In real applications, for simplification, we can assume that the observational errors are zero and the forecast error covariance satisfies a normal distribution given by

$$P_{ik} = e^{-(\frac{r_{ik}}{d})^2}$$

where r_{ik} is the horizontal distance between i and k points and d is the correlation radius. With this approach, the OI scheme should be very similar to the nudging data assimilation scheme.

Note:

The nudging and OI data assimilation methods are practical approaches for the purpose of model application to the real-time simulation and assimilation. However, they lack rigorous scientific support and are not generally useful for sensitivity studies of model parameters.

Adjoint Assimilation Methods

The adjoint data assimilation is conducted using a variational method. The governing equations of the ocean circulation model can be written in the form of vectors as

$$\frac{\partial \mathbf{x}}{\partial t} = F(\mathbf{x},\mathbf{c})$$

where

$\mathbf{x}$ is a matrix array consisting of dependent variables such as u, v, w, ζ, T, and s;

F is a nonlinear operator including the advective, Coriolis, pressure and diffusive terms for the momentum equations and advective and diffusive terms for temperature and salinity equations;

$\mathbf{c}$ is a matrix array containing the model parameters such as drag coefficient, light attenuation lengths, and boundary and initial conditions.

The cost function(to measure the "distance" (error) between the observations and model prediction) is defined as

$$\boxed{J(\mathbf{x},\mathbf{c}) = \int_{T_1}^{T_2}\int_{\Omega}[\frac{K}{2}(\mathbf{x}-\mathbf{x}_o)^2 + \frac{K_c}{2}(\mathbf{c}-\mathbf{c}_o)^2]\mathrm{d}\Omega\,\mathbf{dt}}$$

where $\boldsymbol{x}_o$ and $\boldsymbol{c}_o$ are the observed vectors for dependent variables and model parameters, respectively, and $\boldsymbol{K}$ and $\boldsymbol{K}_c$ are validity coefficients, Ω is the numerical computational domain; t represents time and (T_1, T_2) is the time integration window.

The Lagrange function is defined as

$$L(\mathbf{x},\lambda,\mathbf{c}) = \int_{T_1}^{T_2}\int_{\Omega}\{\frac{K}{\mathbf{2}}(\mathbf{x}-\mathbf{x}_o)^2 + \frac{K_c}{2}(\mathbf{c}-\mathbf{c}_o)^2 + \lambda[\frac{\partial\mathbf{x}}{\partial t} - F(\mathbf{x},\mathbf{c})]\}\,\mathbf{d}\Omega\,\mathbf{dt}$$

[note, $(\mathbf{x}\text{-}\mathbf{x}_o)^2 = (\mathbf{x}\text{-}\mathbf{x}_o)(\mathbf{x}\text{-}\mathbf{x}_o)^T$; λ is a matrix array called the Lagrange multipliers for X (also denoted as adjoint variables)]

The variational method used in the adjoint data assimilation model attempts to find the solution of $\mathbf{x}$ with a minimum value of $\boldsymbol{L}$ with respect of $\boldsymbol{\lambda}$, $\mathbf{x}$, and $\boldsymbol{c}$ in term of least square fitting, *i.e.*,

$$\delta L(\lambda,\mathbf{x},\mathbf{c}) = \frac{\partial L}{\partial\lambda}\partial\lambda + \frac{\partial L}{\partial\mathbf{x}}\delta x + \frac{\partial L}{\partial\mathbf{c}}\delta\mathbf{c} = 0$$

This condition is equivalent to solving the Lagrange-Euler equations that satisfy the constraints in the form of

$$\frac{\partial L}{\partial\lambda} = 0 \quad \Rightarrow \quad \text{Forward equation}$$

$$\frac{\partial L}{\partial\mathbf{x}} = 0 \quad \Rightarrow \quad \text{Adjoint equation}$$

$$\frac{\partial L}{\partial\mathrm{c}} = 0 \quad \Rightarrow \quad \text{Parmater control equation}$$

$$L(\mathbf{x},\lambda,\mathbf{c}) = \int_{T_1}^{T_2}\int_{\Omega}\{\frac{K}{\mathbf{2}}(\mathbf{x}-\mathbf{x}_o)^2 + \frac{K_c}{2}(\mathbf{c}-\mathbf{c}_o)^2 + \lambda[\frac{\partial \mathbf{x}}{\partial t} - F(\mathbf{x},\mathbf{c})]\}\mathbf{d}\Omega\mathbf{dt}$$

$$\frac{\partial L(\mathbf{x},\lambda,\mathbf{c})}{\partial L} = \int_{T_1}^{T_2}\int_{\Omega}\{\frac{\partial \mathbf{x}}{\partial t} - F(\mathbf{x},\mathbf{c},t)\}\mathbf{d}\Omega\mathbf{dt} = 0 \quad \Rightarrow \boxed{\frac{\partial \mathbf{x}}{\partial t} - F(\mathbf{x},\mathbf{c},t) = 0}$$

Forward model

$$\frac{\partial L(\mathbf{x},\lambda,\mathbf{c})}{\partial \mathbf{x}} = \int_{T_1}^{T_2}\int_{\Omega}\{K(\mathbf{x}-\mathbf{x}_o) + \frac{\partial \lambda}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial t} - \lambda\frac{\partial F(\mathbf{x},\mathbf{c},t)}{\partial \mathbf{x}}\}\mathbf{d}\Omega\mathbf{dt} = 0$$

$$\Rightarrow \boxed{\frac{\partial \lambda}{\partial t} = -K(\mathbf{x}-\mathbf{x}_o) + \lambda\frac{\partial F(\mathbf{x},\mathbf{c},t)}{\partial \mathbf{x}}}$$

Adjoint model

$$\frac{\partial L(\mathbf{x},\lambda,\mathbf{c})}{\partial \mathbf{c}} = \int_{T_1}^{T_2}\int_{\Omega}\{K_c(\mathbf{c}-\mathbf{c}_o) - \lambda\frac{\partial F(\mathbf{x},\mathbf{c},t)}{\partial \mathbf{c}}\}\mathbf{d}\Omega\mathbf{dt} = 0$$

$$\Rightarrow \boxed{\mathbf{c} = \mathbf{c}_o + \frac{\lambda}{K_c}\frac{\partial F(\mathbf{x},\mathbf{c},t)}{\partial \mathbf{c}}}$$

Parameter control equation

Adjoint Data Assimilation Approach