

**DATA ASSIMILATION TECHNIQUES IN MODELING
OCEAN PROCESSES**

By

R. Mahadevan, A. A. Fernandes and S.W.A. Naqvi

Technical Report No. NIO/TR-10/2001



**NATIONAL INSTITUTE OF OCEANOGRAPHY
DONA PAULA, GOA 403 004**

August 2001

CONTENTS

Acknowledgements

Abstract

1. Introduction

2. Variational Formalism

2.1 Advection Problem

3. Solution Methods

3.1. Penalty Algorithm

3.2 Optimal Control Method

4. Application of Data Assimilation Technique

5. One Dimensional Model of Upper Ocean

6. One Dimensional Model of Water Column Biogeochemistry

7. Summary

References

Appendix - A

Objective Analysis of Ocean Data

Appendix - B

Extreme Value Problems

Appendix - C

Descent Algorithms

ACKNOWLEDGEMENTS

The authors express their sincere thanks to Dr. E. Desa, Director, National Institute of Oceanography, Goa for supporting this study on data assimilation and for his constant encouragement. The authors would like to thank Dr. T. Pankajakshan for going through the manuscript, giving valuable suggestions and introducing us to the literature on mixed layer dynamics. We express our grateful thanks to Ms. Felcy Alphonso for the secretarial assistance and to Mr. Udaykumar Javali for drafting the figures, which helped in quickly bringing out this report.

DATA ASSIMILATION TECHNIQUES IN MODELING

OCEAN PROCESSES

Abstract:

Three main classes of procedures in data analysis and assimilation *viz.* Objective Analysis, Optimal Interpolation and variational method, used to process the observed data on atmospheric and ocean parameters are briefly reviewed. The variational method of data assimilation, its mathematical description and the classical variational approach, which leads to the Euler-Lagrange equations, are presented. Two methods of solving data assimilation problems, *viz.* the penalty algorithm and the optimal control theory method are discussed. Application of data assimilation technique to physical problems governed by linear and nonlinear ordinary differential equations are explained. The formulation of data assimilation technique in one-dimensional models of upper ocean and water column biogeochemistry is presented.

1. INTRODUCTION

In meteorology and oceanography, the procedures used to process the observed data on atmospheric and ocean parameters, distributed in space and time to generate homogeneous fields are usually called data analysis or assimilation. These homogeneous fields are prerequisites for various practical applications and theoretical studies. The fields produced by an analysis or assimilation must satisfy two basic requirements. On the one hand, they must be close to the observations and on the other hand they must satisfy the dynamic relationships, which govern the process observed.

Three main classes of procedures have been used in analysis or assimilation. The first class comprises of spatial interpolation, known as 'Objective Analysis' in meteorology (Cressman, 1959; Barnes, 1973). A brief description of objective analysis is presented in Appendix-A. The second class of procedures is known as 'Optimal

Interpolation'. The fields generated by these procedures are the linear combinations of observations, which minimize the mean quadratic error between predicted and observed values. Optimal interpolation techniques have been introduced mainly by Gandin (1965) in meteorology and are being widely used in large scale prediction models, see e.g. Lorenc (1981). The third class of analysis is based on variational method. They produce fields, which minimize a given measure of the distance between the analysis and the observations, while at the same time satisfy an explicit dynamic constraint. The dynamic constraints will normally be expressed by one or more differential equations. These methods have the advantage of providing better consistency between the analysis and dynamics as expressed by the constraints.

In this report, the variational method of data assimilation or fitting dynamics to data is briefly reviewed. Two methods of solving data assimilation problems, viz. penalty algorithm and optimal control theory method are discussed. The formulation of data assimilation technique in one-dimensional models of, upper ocean and water column biogeochemistry is presented.

The practical usefulness of variational methods for meteorological problems are pointed out very early by Sasaki (1955, 1958), but in spite of that these methods have not been fully utilized. Probably, the complex mathematical technicality of these methods and their high computational costs have limited their use. This technique received more attention in meteorological applications after 1985, e.g. Derber (1985), Lewis and Derber (1985), Le Dimet and Talagrand (1986), Hoffman (1986), Talagrand and Courtier (1987), Courtier and Talagrand (1987), and Lorenc (1988).

Applications of data assimilation techniques in oceanography did not receive much attention compared to that in meteorology. Closer to the idea of data assimilation or fitting dynamics to data in oceanography, is the dynamical method used to produce ocean circulation pattern from hydrographic observations. These early attempts were followed by more sophisticated models which used inverse method, variational method, direct minimization method and optimal control techniques, to introduced more observed data into the models. Some investigators have referred all the above methods as inverse methods.

Wunsch (1978) reviewed the inverse methods to fit data to dynamics and using an inverse method determined North Atlantic general circulation West of 50° W. Bennett (1982) described the theory of numerous inverse methods for applications in oceanography.

Bennett and McIntosh (1982) used a variational method in the investigation of tidal prediction. Provost(1983) and Provost and Salmon (1986) have used a variational technique to simulate hydrographic station data to estimate the three-dimensional field of geostrophic velocities. Malanotte-Rizzoli and Holland (1986, 1988) used a quasi-geostrophic general circulation model to investigate the effect of data insertion into a numerical model. They examined how data insertion in local limited regions affects the dynamics of the model and also how the dynamics spreads the inserted information into different regions. Smedstad and O'Brien (1991) followed a variational approach of strong constraints and applied it to equatorial Pacific Ocean to determine the solution of the reduced gravity model, which fits the observations.

Thacker and Long (1988) and Thacker (1988) have described a variational assimilation method in their papers. The method is illustrated using a simple model of the wind-driven equatorial ocean. It is shown how surface elevation and wind stress observations can be used to recover the model state. In Thacker (1988) the errors in fitting a model to inadequate data is discussed. Results show that for the simple wave model of Thacker and Long (1988) a reasonable fit can be obtained even if the number of observations is less than the number of degrees of freedom of the model. Basu *et al.* (1993) used the numerical scheme proposed by Thacker and Long (1988) to assimilate data into a one-dimensional nonlinear shallow water equations model with state variables representing potential and current.

Copeland and Bayne (1998) used the direct minimization method to study tidal flow in coastal areas accommodating large quantities of measured data within the model domain. LeDimet and Talagrand (1986) discussed different algorithms for analysis and assimilation of data into models. They applied the optimal control techniques to solve the one-dimensional version of shallow water equations.

The need for using assimilation techniques in Oceanography can be explained as follows:

(a) The conventional or direct fluid flow simulation models use numerical methods (FDM or FEM or spectral methods) to solve the equations governing the process. In these models, certain measured process parameters are used to define the initial and boundary conditions. For this reason these problems are generally referred to as initial / boundary value problems. In these problems, except for calibration and validation of model prediction, there is no way to use field data which have been collected with elaborate care, efforts and sophisticated equipment; that is we have only limited use for such data

This problem is becoming more serious with the availability of improved measuring techniques to collect large quantity of good quality data. The drifting and moored buoy systems and remote sensing techniques provide large quantity of good quality data describing the variability of physical parameters. In such a fast changing situation, it is not proper to continue with the conventional modeling techniques, which are not capable of using the large quantity of data, the availability of which is continuously increasing.

(b) In conventional methods, sometimes, the error resulting from model formulation arising from the simplification of dynamic equations and in specifying the model inputs exceed the errors in the measured data. In such cases, models which include a large quantity of measured data are expected to provide more realistic solutions.

(c) Assimilation methods also have the potential to produce more accurate solutions, since it allows both data and equations to control the evolution of the solution. Hence, adopting the data assimilation techniques in numerical modeling of ocean processes is considered to be a better modeling practice.

2. VARIATIONAL FORMALISM

In this formalism a given measure of the distance between the model and the observation is minimized. The primary field of interest in oceanography is the time dependent flow field and the equations governing the flow field can be written as,

$$G[Y] = 0$$

$$\text{and } G[Y] \equiv (\partial Y / \partial t) - F(Y), \quad (2.1)$$

where the components of the vector Y are the dependent oceanographic fields such as u , v , w , p etc., and t is the time. These equations are defined over some spatial and temporal domain ' σ ' with the boundary ' Γ '. F may be a linear or a nonlinear operator. Additional information to this problem will be given by ' b ', which can be considered as a control variable. For instance, b can consist of initial conditions or boundary conditions or a combination of both. When b is specified, a unique solution $Y(b)$ of Eqn.(2.1) can be estimated.

If the observations ' \underline{Y} ' on oceanographic fields at homogeneously or irregularly distributed set of points over σ are given, we are interested in obtaining a solution of Eqn.(2.1) which is **close** to the observation \underline{Y} . The **closeness** is defined as a function of Y and \underline{Y} and it is called the **cost function** ' J '. Though there are several ways of defining the cost function J in data assimilation problems, it is often defined as,

$$J[Y] = \{Y - \underline{Y}, Y - \underline{Y}\} \quad (2.2)$$

where $\{ \cdot, \cdot \}$ is the inner product defined in the spatial and temporal domain as,

$$\{Y - \underline{Y}, Y - \underline{Y}\} = \int (Y - \underline{Y})^2 d\sigma \quad (2.3)$$

where $d\sigma = dx \cdot dt$. The integral in Eqn.(2.3) must be replaced by an appropriate finite sum if \underline{Y} consists of discrete observations.

We seek a solution to Eqn.(2.1) which minimizes the function J . This problem is referred to as **Data Assimilation** or **fitting dynamics to the observed data**. This problem of model fitting should be particularly appropriate for highly asymptotic and non-homogeneous data sets, such as, data from fixed moorings, from drifting buoys, from satellites and from relatively unscheduled shipboard observations.

The mathematical problem stated above is a **constrained variational (minimization) problem** with the model Eqn.(2.1) representing the constraint on the cost

function ' J ' (Eqn.2.2). The general minimization and constrained minimization problems are discussed in Appendix-B.

A systematic approach for solving problems of this type is based on a generalization of the classical **Lagrange multiplier technique** used in a finite dimensional space. Under simple regularity condition (*i.e.* differentiability of Y to a sufficient order), it can be shown that the problem of determining the stationary point of Eqn.(2.2) under the constraint Eqn.(2.1) is equivalent to the problem of determining the stationary point of the Lagrangian $L[Y, \Lambda]$. The **Lagrangian ' L '** for the present problem is defined as a scalar function of Y and Λ in the form,

$$L[Y, \Lambda] = J[Y] + \{\Lambda, G[Y]\} \quad (2.4)$$

where Λ an unknown function in the domain σ , is called the Lagrange multiplier. In this formulation the constrained problem is replaced by an unconstrained problem with respect to the functions Y and Λ .

Then for the stationarity of L, we require the first order variation δL to be zero for arbitrary perturbations of δY and $\delta \Lambda$. This requirement gives the Euler - Lagrange equations of the problem, which in principle can be solved to estimate the desired field. However, except in special cases, directly solving this problem is not easy.

This classical variational analysis approach is also referred to as the **strong constraint formalism**

2.1 Strong constraint formalism of Advection Problem.

To illustrate this **strong constraint formalism**, let us consider the linear advection equation,

$$G[C] = 0,$$

$$\text{and } G[C] \equiv (\partial C / \partial t) + v(x) (\partial C / \partial x) \quad (2.5)$$

where C is an unknown scalar function and $v(x)$ is a known function of x. Observations on $\underline{C}(x, t)$ over some domain σ in the (x, t) plane are given. We seek a solution to Eqn.(2.5) (*i.e.* C(x,t)) which minimizes the function,

$$J[C] = \int (C - \underline{C})^2 dx \cdot dt \quad (2.6)$$

Eqns.(2.5) and (2.6) constitute a constrained variational problem.

Under simple regularity condition, the problem of determining the stationary points of Eqn.(2.6) under the constraint Eqn.(2.5) is equivalent to the problem of determining the stationary points of the Lagrangian $L[C, \Lambda]$ defined as,

$$L[C, \Lambda] = J[C] + \{\Lambda, G\} \quad (2.7)$$

where the Lagrange multiplier Λ is an unknown function of x and t . The second term in the above integral defines an inner product in the (x, t) space. In this formulation the constrained problem is replaced by an unconstrained problem with respect to the functions C and Λ .

Then the first-order variation δL resulting from perturbations δC and $\delta \Lambda$ is equal to,

$$\begin{aligned} \delta L = \int [2(C - \underline{C}) \delta C + \Lambda \{ [\partial (\delta C)/\partial t] + v [\partial (\delta C)/\partial x] \}] dx dt \\ + \int [(\partial C/\partial t) + v (\partial C/\partial x)] \delta \Lambda dx dt \end{aligned} \quad (2.8)$$

Integration by parts of the terms containing the derivatives of δC in the above equation leads to,

$$\begin{aligned} \delta L = \int [2(C - \underline{C}) - (\partial \Lambda/\partial t) - (\partial \Lambda v/\partial x)] \delta C dx dt \\ + \int [(\partial C/\partial t) + v (\partial C/\partial x)] \delta \Lambda dx dt + \int_{\Gamma} \Lambda V \cdot n \delta C d\Gamma \end{aligned} \quad (2.9)$$

where n is a unit vector normal to the boundary Γ of σ , directed outwards and V is the vector with components v and 1 in the x and t directions respectively.

For given C and Λ , L will be stationary if and only if δL is zero for any arbitrary perturbations δC and $\delta \Lambda$. This will be the case if and only if the three following equations are satisfied,

$$(\partial \Lambda/\partial t) + [\partial(\Lambda v)/\partial x] - 2(C - \underline{C}) = 0$$

$$(\partial C/\partial t) + v (\partial C/\partial x) = 0$$

$$\text{and} \quad \Lambda = 0 \quad \text{along} \quad \Gamma \quad (2.10a,b,c)$$

where Eqn.(2.10a) is the adjoint equation of the governing Eqn.(2.5). The Eqns.(2.10a, b) with the boundary condition Eqn.(2.10c) are the Euler-Lagrange equations of the problem. Except in particular cases, it is difficult to directly solve the above equations. Hence, different techniques are developed in which the original problem is transformed

by changing either the cost function J or the constraint or both into a sequence of unconstrained problems.

3. SOLUTION METHODS

As Euler- Lagrange equations are difficult to solve directly, different techniques are developed to solve data assimilation problems. Le Dimet and Talagrand (1986) discussed two approaches followed to solve data assimilation problems in meteorology. They discussed three algorithms, namely, Penalty, Duality and Augmented Lagrange algorithms in the first approach in which the original constrained variational problem is transformed into unconstrained problem. The second approach uses optimal control techniques. In this section, the penalty algorithm (also referred to as the direct minimization technique) and optimal control techniques are discussed.

3.1. Penalty Algorithm

We have already noted in Section 2 that in the classical variational or strong constraint formalism, the constraining equations were appended to the cost function by means of Lagrange multiplier. In this formalism the desired solution is expected to satisfy the constraints exactly. In the penalty algorithm method the squares of the constraining equations are appended to the cost function by means of weighting factor 'W' or penalty parameter ($\epsilon = 1/W$), *i.e.*,

$$J_C[Y] = J[Y] + W |G[Y]|^2 \quad (3.1)$$

where $J_C[Y]$ is called the constrained or penalized cost function. In this formalism we seek the minimum of $J_C[Y]$ and the problem reduces to an unconstrained minimization. Note that the constraints are satisfied only approximately in a least square error sense and not exactly. The degree of fulfillment of constraint depends on the magnitude of the weighting factor. As the weight factor becomes larger, the degree of fulfillment increases and the constraint approaches to a strong constraint. This approach is referred to as **weak constraint formalism or penalty algorithm or direct minimization technique**.

The method of weak constraint is more realistic compared to strong constraint approach with respect to the errors and approximations inherent in the formulation of the governing equations. That is, in the strong constraint formalism, the requirement for exact satisfaction of the constraining equation assumes that the constraints are an exact representation of the physics involved. This is not generally true. Copeland and Bayne

(1998) demonstrated the capability of this technique in simulating one and two dimensional shallow water flow fields.

3.2 Optimal Control Method

In many constrained variational problems of the type discussed in section 2, a solution of the equations governing the problem (the constraint of the problem, Eqn.2.1) can be uniquely determined for the specified additional information b . ' b ' can consist of initial conditions, boundary conditions or a combination of them. Then the constrained variational problem discussed in section 2 can be stated in the following form:

Find ' b ' (the initial or boundary conditions) such that the corresponding solution of Eqn.(2.1) minimizes the cost function J . ' b ' will be referred to as the control variable of the problem. J is a function of ' b ' through the governing Eqn.(2.1).

The obvious advantage of recasting the problem in the above form (as an "optimal control problem") is that, the number of independent parameters to be varied in the minimization process are substantially reduced. Besides, this method also offers an alternative technique to explicitly compute ∇J with respect to b , with less computational efforts.

For example, the minimization process in this problem needs the determination of the gradient of J with respect to ' b '. Standard differentiation techniques will not in general be useful for the estimation of ∇J . One direct approach to this problem would be to perturb in turn all the components of b , and for each perturbation solve the governing Eqn. (2.1) and compute the corresponding J and $\nabla J = (\partial J/\partial b)$. But the extensive computations involved makes this approach totally impractical. The optimal control method offers a technique to compute ∇J , which involves fewer computations.

To illustrate this method, we shall consider a variable $Y(x, t)$ dependent on two independent variables x and t (this technique could be extended to more number of dependent and independent variables) and governed by the equation,

$$(\partial Y/\partial t) = F(Y) \tag{3.2}$$

where F may be a linear or nonlinear operator. Let the equation governing the first order perturbation δY about any solution $Y(x, t)$ of Eqn.(3.2) be,

$$\partial (\delta Y) / \partial t = A \delta Y \quad (3.3)$$

Then in terms of the adjoint operator A^* of A with respect to the inner product in space x , the adjoint equation to Eqn.(3.2) can be written as,

$$\partial (\delta^* Y) / \partial t = - A^* (\delta^* Y) \quad (3.4)$$

To illustrate the derivation of the linearized perturbation and the adjoint equations, we will consider a non-linear differential equation,

$$(\mathcal{J}Z/\mathcal{J}t) + Z (\mathcal{J}Z/\mathcal{J}x) = c Z$$

where Z is the unknown dependent variable of the problem (this is a typical equation we frequently come across in flow simulation problems). Then the linearized perturbation equation about a given solution $Z(x, t)$ of the above equation become,

$$\mathcal{J}(\mathbf{d}Z)/\mathcal{J}t = B(\mathbf{d}Z)$$

where the operator B , a function of (x, t) is defined as,

$$B = [-Z (\mathcal{J}/\mathcal{J}x) - (\mathcal{J}Z/\mathcal{J}x) + c]$$

The operator B^* is determined by integrating by parts the terms containing the derivative of $(\mathbf{d}Z)$ with respect to x in the following integral (i.e. the inner product in x -space denoted by $\langle ., . \rangle$),

$$\begin{aligned} I &= \int \mathbf{\alpha}(\mathbf{d}^*Z) B(\mathbf{d}Z) dx = \langle \mathbf{d}^*Z, B(\mathbf{d}Z) \rangle \\ &= \int \mathbf{\alpha}(\mathbf{d}^*Z) \{ -Z [\mathcal{J}(\mathbf{d}Z)/\mathcal{J}x] - [(\mathcal{J}Z/\mathcal{J}x) - c] (\mathbf{d}Z) \} dx \end{aligned}$$

and we get,

$$\begin{aligned} I &= \int \mathbf{\alpha}(\mathbf{d}Z) [Z \mathcal{J}(\mathbf{d}^*Z)/\mathcal{J}x + c (\mathbf{d}^*Z)] dx - (\mathbf{d}Z) (\mathbf{d}^*Z) Z \Big|_{\mathbf{G}} \\ &= \int \mathbf{\alpha}(\mathbf{d}Z) B^*(\mathbf{d}^*Z) dx - (\mathbf{d}Z) (\mathbf{d}^*Z) Z \Big|_{\mathbf{G}} \end{aligned}$$

where the suffix \mathbf{G} denotes that the function is determined at the boundary and B^* is defined as,

$$B^* = [Z (\mathcal{J}/\mathcal{J}x) + c].$$

Then the adjoint equation in terms of operator B^* is given by,

$$\mathcal{J}(\mathbf{d}^*Z)/\mathcal{J}t = - B^*(\mathbf{d}^*Z)$$

Now let us assume that the observations on the state of the system $\underline{Y}(x, t_n)$, $n = 1, 2, \dots, N$, are available and we will seek a solution $Y(x, t)$ of Eqn.(3.2) minimizing the functional,

$$J[Y(x, t)] = \sum_{n=1}^N \langle Y(x, t_n) - \underline{Y}(x, t_n), Y(x, t_n) - \underline{Y}(x, t_n) \rangle \quad (3.5)$$

where $\langle \cdot, \cdot \rangle$ and $\sum \langle \cdot, \cdot \rangle$ denote inner products over x space and (x, t) space respectively. Here it is assumed that $\underline{Y}(x, t)$ are available at discrete time intervals. We will assume that the observation $\underline{Y}(x, t_n)$ are not exactly compatible with the solution of Eqn. (3.2) so that the minimum of J will not in general be equal to 0. The above stated problem is a constrained minimization (or variational problem) of the cost function J (Eqn.3.5) with Eqn.(3.2) as the constraint.

The solution $Y(x, t)$ of Eqn.(3.2) is uniquely defined by specifying the initial condition $Y(x, t_1)$, which will be used as the control variable. The problem now reduces to finding the initial condition $Y(x, t_1)$ such that the corresponding solution of Eqn.(3.2) minimizes the cost function (3.5). In this form the problem becomes unconstrained minimization problem.

From Eqn.(3.5) the first order variation δJ resulting from a perturbation $\delta Y(x, t_1)$ to the initial condition is given by,

$$\delta J = 2 \sum_{n=1}^N \langle Y(x, t_n) - \underline{Y}(x, t_n), \delta Y(x, t_n) \rangle \quad (3.6)$$

where $Y(x, t_n)$ is the solution to the governing equation (Eqn. 3.2) at any time t_n for an assumed initial condition $Y(x, t_1)$. And the first order perturbation $\delta Y(x, t_n)$ is the solution of the linearized perturbation Eqn.(3.3), with the initial condition $\delta Y(x, t_1)$.

As the Eqn.(3.3) is linear, $\delta Y(x, t_n)$ at a any time t_n depends linearly on the initial condition $\delta Y(x, t_1)$ and hence $\delta Y(x, t_n)$ can be written as,

$$\delta Y(x, t_n) = R(x, t_n; t_1) \delta Y(x, t_1) \quad (3.7)$$

where the linear operator $R(x, t_n ; t_1)$ is called the resolvent of Eqn.(3.3). It is the solution of the Eqn.(3.3) at t_n with the initial condition $\delta Y(x, t_1) = 1$. Substitution of Eqn.(3.7) in Eqn.(3.6) gives,

$$\delta J = 2 \sum_{n=1}^N \langle Y(x, t_n) - \underline{Y}(x, t_n), R(x, t_n ; t_1) \delta Y(x, t_1) \rangle \quad (3.8)$$

Let $R^*(x, t_n ; t_1)$ be the adjoint of the linear operator $R(x, t_n ; t_1)$, then we have,

$$\delta J = \langle 2 \sum_{n=1}^N R^*(x, t_n ; t_1) [Y(x, t_n) - \underline{Y}(x, t_n)], \delta Y(x, t_1) \rangle \quad (3.9)$$

since for any two vectors u, v in the (x, t) space $\sum \langle R^*u, v \rangle = \sum \langle u, Rv \rangle$. Also δJ , the first order variation in J due to $\delta Y(x, t_1)$, in terms of the gradient of J , is

$$\delta J = \langle \nabla J, \delta Y(x, t_1) \rangle \quad (3.10)$$

Hence from Eqns. (3.9) and (3.10), we have,

$$\nabla J = 2 \sum_{n=1}^N R^*(x, t_n ; t_1) [Y(x, t_n) - \underline{Y}(x, t_n)] \quad (3.11)$$

For any two solutions $u(x, t)$ and $v(x, t)$ of the linearized Eqn.(3.3) and the adjoint Eqn.(3.4) respectively, the inner product defined over the space x is a constant with time, since,

$$\begin{aligned} (\partial / \partial t) \int u(x, t) \cdot v(x, t) dx &= (\partial / \partial t) \langle u(x, t), v(x, t) \rangle \\ &= \langle \partial u / \partial t, v \rangle + \langle u, \partial v / \partial t \rangle \\ &= \langle Au, v \rangle + \langle u, -A^*v \rangle = 0 \end{aligned}$$

where $\langle ., . \rangle$ denotes the inner product with respect to x . Hence,

$$\langle u(x, t_1), v(x, t_1) \rangle = \langle u(x, t_n), v(x, t_n) \rangle,$$

$$\text{and } \langle u(x, t_1), S(x, t_1 ; t_n) v(x, t_n) \rangle = \langle R(x, t_n ; t_1) u(x, t_1), v(x, t_n) \rangle. \quad (3.12)$$

where $[S(x, t_1 ; t_n) \cdot v(x, t_n)]$ is the solution of the adjoint equation (Eqn.3.4) when the condition given at time t_n is $v(x, t_n)$. This result shows that $S(x, t_1 ; t_n) = R^*(x, t_n ; t_1)$.

Thus $R^*(x, t_n ; t_1)$ is the resolvent of the adjoint Eqn.(3.4) between the time t_n and t_1 .

Hence in Eqn.(3.11), the term $\{R^*(x, t_n ; t_1) [Y(x, t_n) - \underline{Y}(x, t_n)]\}$ can be obtained by

integrating the adjoint Eqn.(3.4) backward in time from t_n to t_1 starting with the value $[Y(x, t_n) - \underline{Y}(x, t_n)]$. Since the adjoint Eqn.(3.4) is linear, ∇J in Eqn.(3.11) can be estimated by one backward integration of Eqn.(3.4) from t_n to t_1 starting with the value $(\delta^*Y)_N = [Y(x, t_N) - \underline{Y}(x, t_N)]$ and adding $[Y(x, t_n) - \underline{Y}(x, t_n)]$ at every discrete time t_n to the current integrated value of (δ^*Y) .

In summary, the gradient ∇J (which is an important parameter required in a descent procedure involved in the data assimilation problem) can be computed using the Eqn.(3.11). Steps involved are:

- Solve the given governing equation (*i.e.* determine the solution $Y(x, t)$ of Eqn. 3.2) for an initial guess value of the control variable (*i.e.* initial / boundary conditions). The coefficients of the adjoint equation is determined using $Y(x,t)$.
- Integrate the adjoint equation backward from t_n to t_1 , starting with the value $(\delta^*Y)_N = [Y(x, t_N) - \underline{Y}(x, t_N)]$
- At every discrete time t_n , add $[Y(x, t_n) - \underline{Y}(x, t_n)]$ to the current integrated value of (δ^*Y) and estimate ∇J .

Then a suitable descent procedure is used to locate the minimum of the cost function.

4. APPLICATIONS OF DATA ASSIMILATION TECHNIQUE

We will discuss here the problem of data assimilation in the case of physical processes governed by ordinary differential equations. We will follow the optimal control theory approach of fitting model to data.

(a) Linear first order differential equation

We will consider a physical process governed by a linear first order differential equation of the form,

$$(dY/dt) = F(Y) \quad (4.1)$$

and observations $\underline{Y}(t_n)$ (see Fig.4.1) of the system are available at time t_n , $n = 1,2,3,\dots,N$.

We will seek a solution of Eqn.(4.1) which minimizes the cost function,

$$J = \sum_{n=1}^N [Y(t_n) - \underline{Y}(t_n)]^2 \quad (4.2)$$

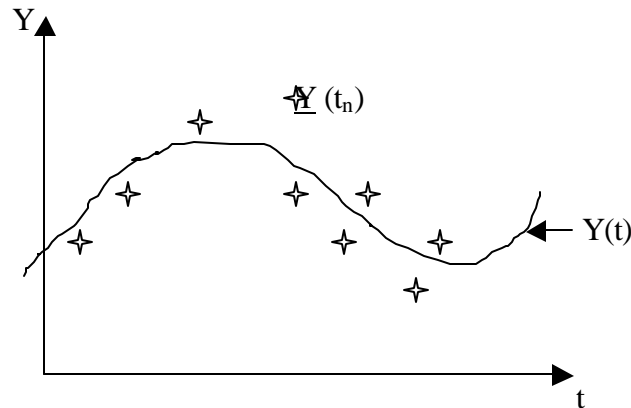


Figure 4.1 Observed and predicted values of the variable Y.

As the governing equation is linear, there is no need for the linearized perturbation equation. We solve directly the governing Eqn.(4.1) for $Y(t_n)$ with the initial condition $Y(t_1) = 1$, *i.e.* determine the resolvent $R(t_n, t_1)$. Note that for $Y(t_1) \neq 1$, $Y(t_n) =$

$R(t_n, t_1) \cdot Y(t_1)$, as the governing equation is linear. Substituting for $Y(t_n)$ in the cost function J (Eqn.4.2) we get,

$$J = \sum_{n=1}^N [R(t_n, t_1) Y(t_1) - \underline{Y}(t_n)]^2 \quad (4.3)$$

We now seek $Y(t_1)$ which minimizes J . Hence for minimum J , we require,

$$[dJ/dY(t_1)] = 2 \sum_{n=1}^N [R(t_n, t_1) Y(t_1) - \underline{Y}(t_n)] R(t_n, t_1) = 0,$$

and
$$Y(t_1) = \sum_{n=1}^N [R(t_n, t_1) \underline{Y}(t_n)] / [\sum R(t_n, t_1)^2] \quad (4.4)$$

(b) Coupled linear equations

We will consider coupled linear equations of the form

$$(dU/dt) = F_1(U, V), \quad (dV/dt) = F_2(U, V) \quad (4.5 \text{ a,b})$$

with the observed state of the system $\underline{U}(t_n)$, and $\underline{V}(t_n)$ (see Fig.4.2) being specified. We seek a solution of Eqns.(4.5a,b) which minimizes the cost function,

$$J = \sum_{n=1}^N [[U(t_n) - \underline{U}(t_n)]^2 + S [V(t_n) - \underline{V}(t_n)]^2] \quad (4.6)$$

where S is a scaling factor which depends on the physical units of the variables U and V .

Following the control theory approach, we determine first the solutions $[U^{(1)}(t_n), V^{(1)}(t_n)]$ with the initial condition $[U(t_1), V(t_1)] = (1, 0)$, and $[U^{(2)}(t_n), V^{(2)}(t_n)]$ with the initial condition $[U(t_1), V(t_1)] = (0, 1)$ of the Eqns.(4.5a, b). These solutions are the resolvents $\{[R_U(t_n, t_1)]^{(1)}, [R_V(t_n, t_1)]^{(1)}\}$ and $\{[R_U(t_n, t_1)]^{(2)}, [R_V(t_n, t_1)]^{(2)}\}$ of the Eqns.(4.5a, b).

Then the general solution to the Eqns.(4.5a, b) with the initial condition $U(t_1) \neq 0$ and $V(t_1) \neq 0$, can be expressed in terms of the resolvents in the form,

$$\begin{aligned} U(t_n) &= [R_U(t_n, t_1)]^{(1)} U(t_1) + [R_U(t_n, t_1)]^{(2)} V(t_1) \\ V(t_n) &= [R_V(t_n, t_1)]^{(1)} U(t_1) + [R_V(t_n, t_1)]^{(2)} V(t_1) \end{aligned} \quad (4.7)$$

We substitute for $U(t_n)$ and $V(t_n)$ in Eqn.(4.6) and seek $U(t_1)$ and $V(t_1)$ which minimize J . For J to be minimum we require,

$$[dJ/dU(t_1)] = 0 \quad \text{and} \quad [dJ/dV(t_1)] = 0 \quad (4.8)$$

This leads to a set of linear equations in $U(t_1)$ and $V(t_1)$ which can be solved and the required solutions $U(t_n)$ and $V(t_n)$ determined.

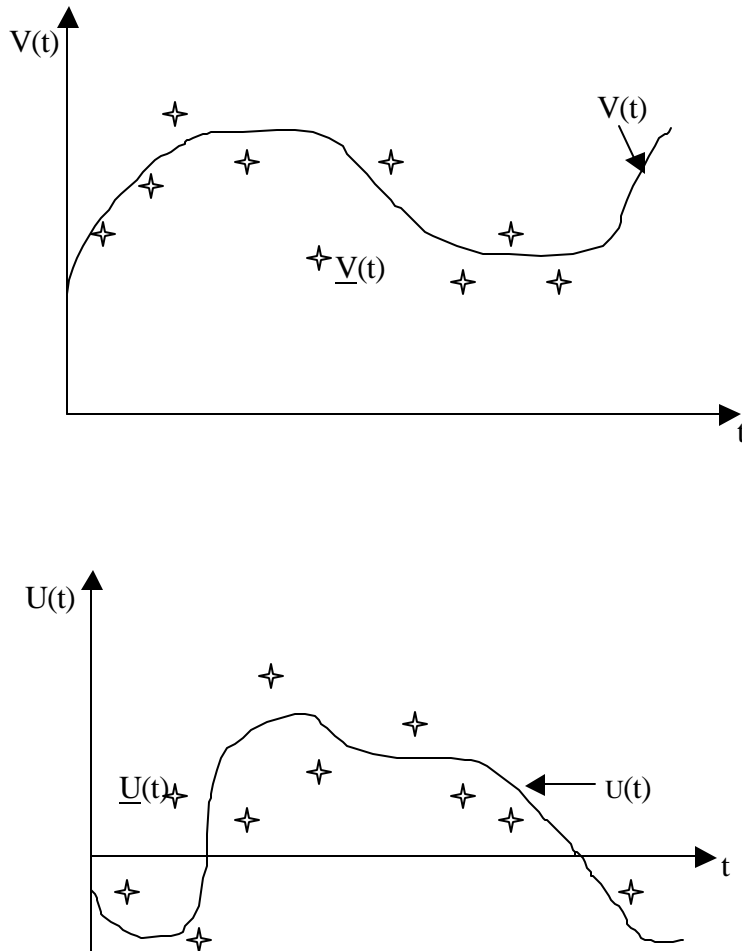


Figure 4.2 Observed and predicted values of the variables U and V .

(c) Nonlinear equations

For illustrating this technique in the case of non-linear equations, we will consider the following equation,

$$(dy/dt) = 2y^{1/2} \quad (4.10)$$

The general solution to this equation can be shown to be

$$y = t^2 + 2 \alpha t + \alpha^2 \quad (4.11)$$

where α is an arbitrary constant. In the control theory method we require the equation governing the first order perturbation, δy , which is given by,

$$d(\delta y)/dt = y^{-1/2} \delta y \quad (4.12)$$

and the adjoint equation of Eqn.(4.10) is,

$$d(\delta^*y)/dt = -y^{-1/2} \delta^*y \quad (4.13)$$

We shall assume that the observed state of the system $y(t_n)$, $n = 1, 2, \dots, N$, are available and we will seek a solution $y(t)$ of Eqn.(4.10), which minimizes the cost function,

$$J = \sum_{n=1}^N [y(t_n) - \underline{y}(t_n)]^2 \quad (4.14)$$

Estimation of the desired solution to this problem involves the following operations:

- Solve the governing Eqn.(4.10) for an initial guess value of the control variable (*i.e.* initial condition). The coefficients of the adjoint equation (Eqn.4.13) are determined using $y(t)$.
- Integrate the adjoint equation backward from t_N to t_1 , starting with the value $(\delta^*y)_N = [y(t_N) - \underline{y}(t_N)]$.
- At every discrete time t_n , add $[y(t_n) - \underline{y}(t_n)]$ to the current integrated value of (δ^*y) and estimate $dJ/dy(t_1)$.

Then a suitable search procedure is used to locate the minimum of the cost function.

5. ONE-DIMENSIONAL MODEL OF UPPER OCEAN

The time dependent, one-dimensional model for the mixed layer of the upper ocean developed by Denman (1973) is presented. In this model, the turbulent mixing and the radiative heating within the mixed layer are expressed so that only simple parameters available from routine meteorological measurements are required as input to predict the evolution of the parameters of the mixed layer. The dependent variables predicted by the model are the mixed layer temperature and depth and the temperature at the bottom interface of this layer. But now with the availability of reliable satellite derived Sea Surface Temperature (SST), it is observed that this problem can be formulated as a problem in data assimilation.

Model Description

The basic assumptions in the mixed layer model developed by Denman (1973) are:

The ocean is incompressible, horizontally homogeneous, vertically stratified and stable, and obeys the Boussinesq approximation.

Horizontal non-homogeneity can enter only indirectly through divergence or convergence within the upper layer resulting from non-zero curl in the large scale wind stress field.

The upper mixed layer is an idealized, vertically homogeneous layer bounded at the bottom by a temperature discontinuity (see Fig.5.1).

The heat and mechanical energy inputs at the upper and lower boundaries are redistributed uniformly throughout the layer by turbulent diffusion. The time required for this redistribution is small compared to the time for the same process to occur below the lower interface of the mixed layer.

Below the lower interface an advective vertical velocity 'w', resulting from the above mentioned convergence or divergence within the upper layer are specified at some arbitrary depth.

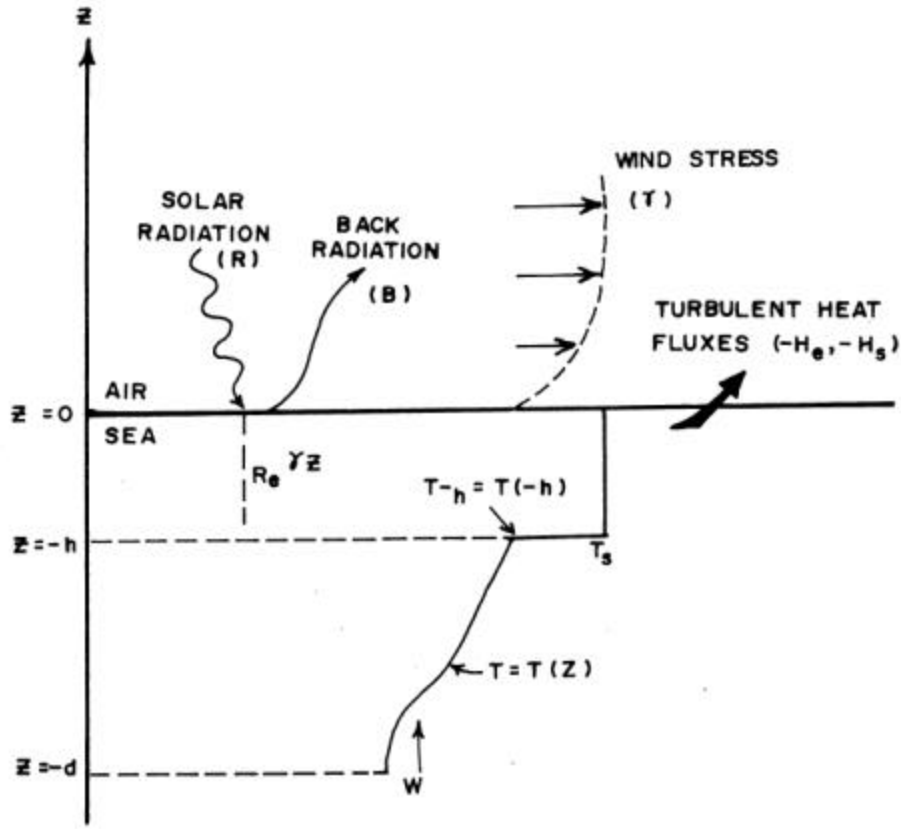


Figure 5.1 Schematic of the thermocline and the boundary inputs assumed in The model (From Denman, 1973)

Thermal Energy Conservation Equation:

If one neglects molecular heat fluxes and viscous generation of heat, then the thermal energy conservation equation in the ocean upper layer which is horizontally homogeneous, is given by,

$$(\partial T/\partial t) + w (\partial T/\partial z) + \partial (\overline{w'T'})/\partial z = \gamma R \exp(\gamma z) \quad (5.1)$$

where T and T' are the mean and the fluctuating components of temperature, t is time, z is the vertical coordinate, w and w' are the mean and fluctuating components of vertical velocity, $R = R^*/(\rho_o C_p)$, R^* is the solar radiation incident on the sea surface ($\text{cal cm}^{-2} \text{sec}^{-1}$), ρ_o is the mean density, C_p is the specific heat at constant pressure ($\approx 0.94 \text{ cal gm}^{-1}(\text{°K})^{-1}$) and γ is the average extinction coefficient in the ocean.

Within the mixed layer, the advection term, $w(\partial T/\partial z)$, vanishes since T is assumed not to vary with respect to z within this layer. Then integrating Eqn.(5.1) in the vertical direction we have,

$$(dT_s/dt) z + \overline{w'T'} \Big|_0 - \overline{w'T'} \Big|_{-z} = R[1 - \exp(-\gamma z)] \quad (5.2)$$

where T_s denotes the surface mixed layer temperature.

In the above equation the turbulent heat flux, $\overline{w'T'}$ at the ocean free surface is given by,

$$\overline{w'T'} \Big|_{z=0} = -F, \quad (5.3)$$

where $F = (H_s^* + H_l^* + B^*) / (\rho_0 C_p) = (H_s + H_l + B)$, H_s^* and H_l^* are the downward fluxes of sensible heat and latent heat respectively. B^* is the heat loss by long-wave radiation.

To determine the heat flux through the bottom of the mixed layer we will examine the heat exchange process over the depth $\mathbf{d}h$ near the lower interface of the mixed layer. The average vertical advection velocity of this layer be ' w ' and we will assume that the turbulent mixing $\overline{w'T'}$ (z) is zero immediately below the interface at $z = -h$. Then integrating the thermal energy conservation equation (Eqn.5.1) between $z = -h$ and $z = -h + \mathbf{d}h$ (see Fig.5.2), we get,

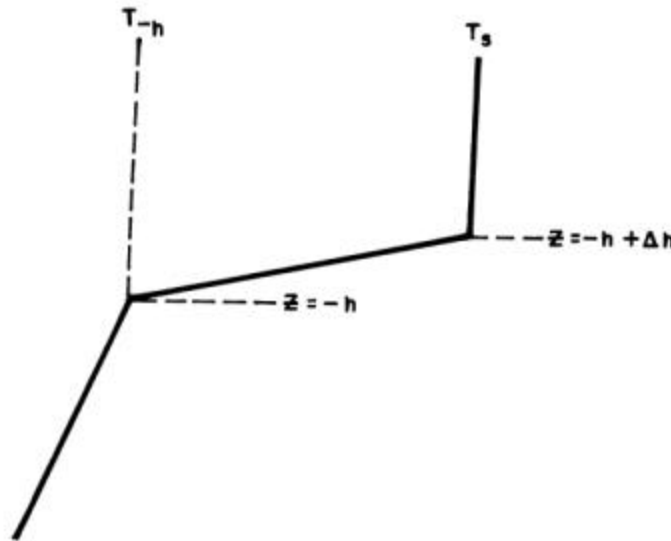


Figure 5.2 Enlargement of the interface at the bottom of the Mixed layer

$$(dT_s/dt)dh + (T_s - T_h)(w + dh/dt) + \overline{w'T'} \frac{1}{2}_{z=-h} + dh = gR \exp(-gh)dh$$

In the limit as $dh \rightarrow 0$, we get

$$\overline{w'T'} \frac{1}{2}_{z=-h} = -H(w + dh/dt)(T_s - T_h) \quad (5.4)$$

where the Heaviside's step function,

$$\begin{aligned} H &= 0, \quad \text{if } w + dh/dt < 0; \quad \text{no entrainment mixing} \\ &= 1, \quad \text{if } w + dh/dt > 0; \quad \text{entrainment mixing at } z = -h \end{aligned} \quad (5.5)$$

Temperature variation below the mixed layer:

Assuming that turbulent fluctuations are absent below $z = -h$, from the thermal energy equation we have for $z < -h$,

$$\left(\frac{\partial T}{\partial t}\right) = gR \exp(gz) - w \left(\frac{\partial T}{\partial z}\right), \quad (5.6)$$

At the lower interface of the mixed layer (i.e. at $z = -h(t)$) we have,

$$\begin{aligned} dT/dt \frac{1}{2}_{z=-h} &= \left(\frac{\partial T}{\partial t}\right) - \left(\frac{\partial h}{\partial t}\right) \left(\frac{\partial T}{\partial z}\right) \frac{1}{2}_{z=-h} \\ &= gR \exp(-gh) - (w + \frac{\partial h}{\partial t}) \left(\frac{\partial T}{\partial z}\right) \frac{1}{2}_{z=-h}, \end{aligned} \quad (5.7)$$

where use has been made of Eqn.(5.6).

In Eqn.(5.2), taking the limits of integration as $z = -h$ and 0 and substituting Eqns.(5.3) and (5.4) we get,

$$(dT_s/dt)h + H(w + dh/dt)(T_s - T_h) = F + R[1 - \exp(-\gamma h)] \quad (5.8)$$

where $F = Hs + H_1 + B$ and H is the Heaviside's step function. Integrating Eqn.(5.1) twice with respect to z over the mixed layer we have,

$$\int_{-h}^0 \overline{w'T'} dz = (dT_s/dt)h^2/2 - Fh - R(h - \gamma^{-1}) - R\gamma^{-1} \exp(-\gamma h) \quad (5.9)$$

Mechanical Energy Conservation Equation:

For the case of horizontal homogeneity, in the presence of a mean horizontal current u , the turbulent mechanical energy equation governing a steady state condition is,

$$- \overline{u'w'}(\partial u/\partial z) - (\partial/\partial z) \left[\overline{w'(p'/\rho_0 + c^2/2)} \right] = \overline{w'r'}/\rho_0 + \varepsilon \quad (5.10)$$

where the turbulent kinetic energy $c^2 = u^2 + v^2 + w^2$ and ε is the rate of dissipation of turbulent energy. Assuming salinity changes are small and $\rho' \approx \alpha \rho_o T'$, where $\alpha = (1/\rho_o) (\partial\rho/\partial T)$, we get from Eqn.(5.10), the steady state equation for turbulent mechanical energy as,

$$- \overline{u'w'} (\partial u/\partial z) - (\partial/\partial z) \left[\overline{w'(p'/\rho_o + c^2/2)} \right] = \alpha g \overline{w'T'} + \varepsilon \quad (5.11)$$

The terms on the left hand side of Eqn.(5.11) are source terms for turbulent mechanical energy and those on the right hand side are the sink terms.

Integrating the mechanical energy equation Eqn.(5.11) with respect to z over $(-h, 0)$ yields,

$$- \int_{-h}^0 \overline{u'w'} (\partial u/\partial z) dz - \overline{w'(p'/\rho_o + c^2/2)} \Big|_{z=0} = \alpha g \int_{-h}^0 \overline{w'T'} dz + \int_{-h}^0 \varepsilon dz \quad (5.12)$$

where the mean of the term $\left[\overline{w'(p'/\rho_o + c^2/2)} \right]$ evaluated at $z = -h$ is assumed to be zero. Defining G and D as,

$$G = - (\alpha g)^{-1} \left[\int_{-h}^0 \overline{u'w'} (\partial u/\partial z) dz + \overline{w'(p'/\rho_o + c^2/2)} \Big|_{z=0} \right]$$

$$D = (\alpha g)^{-1} \int_{-h}^0 \varepsilon dz \quad (5.13a,b)$$

and eliminating the term $\int \overline{w'T'}(z) dz$ between Eqns.(5.9) and (5.12) we get,

$$(dT_s/dt) = (2/h^2) \left[(G - D) + Fh + R(h - \gamma^{-1}) + \gamma^{-1} R \exp(-\gamma^{-1} h) \right] \quad (5.14)$$

It is by combining the integrated forms of thermal and mechanical energy equations and through the elimination of the term $\int \overline{w'T'} dz$ between Eqns.(5.9) and (5.12) that the mechanism for convective mixing enters into the mechanical energy equation. Only for very low wind speeds or for large evaporation does the convective mixing becomes important.

Now eliminating (dT_s/dt) from Eqns.(5.8) and (5.14) we get,

$$H (w + dh/dt) = [(T_s - T_h) h]^{-1} \left[2 \{ - (G - D) + R \gamma^{-1} [1 - \exp(-\gamma h)] \} \right. \\ \left. - h \{ F + R [1 + \exp(-\gamma h)] \} \right] \quad (5.15)$$

Equations (5.7), (5.14) and (5.15) are a set of first-order, non-linear, ordinary differential equations in T_h , T_s and h , which define the mixed layer dynamics at the ocean surface in terms of parameters which can be measured or can be estimated from measurements. In

this problem the temperature gradient, $\partial T/\partial z$, below the mixed layer, which varies with time because of solar heating and vertical advection, must be known for any time t in order to solve the Eqn.(5.7). Eqn(5.6) provides the means for evaluating $T(z)$ and hence $\partial T/\partial z$ in the region $z \leq -h$.

The equations governing the mixed layer dynamics (Eqns. 5.7, 5.14 & 5.15) are normally solved by specifying the required initial conditions, the time dependent surface wind, the net heat flux across the ocean free surface and the temperature gradient below the mixed layer. The dependent variables predicted by the model are the mixed layer depth, h , temperatures T_s and T_h . But now with the availability of reliable satellite derived Sea Surface Temperature (SST), this problem can be viewed as a problem on data assimilation. The data assimilation technique for non-linear equations discussed in Section (4) could be used to solve this problem.

6. ONE DIMENSIONAL MODEL OF WATER COLUMN BIOGEOCHEMISTRY

Biogeochemical processes control the ocean carbon cycle, that is the fluxes of carbon between the atmosphere, surface ocean and ocean interior, which are of different spatial and time scales, regional to global and seasonal to inter-annual. Most of the available information in this area is from local field programs of measurements on biological variables, nutrients and carbon and some times on reaction rates of the processes. Such local studies will have more impact on our knowledge base, if one could extrapolate their results to rest of the ocean. The extrapolation of results is possible only if we could understand the underlying order in these processes. Improving the mathematical models of these processes and estimating their parameters would help in understanding the underlying order in them. In this section the general approach to one-dimensional modeling of water column biogeochemistry is reviewed.

Model Description

In mathematical modeling of biogeochemical processes, the time evolution of the concentration of the state variables describing the processes are estimated by solving the equations governing them. The important state variables are,

PHY	Phytoplankton
ZOO	Zooplankton
NO ₃	Nitrate
NH ₄	Ammonium
DON	Dissolved organic Nitrogen

There are different models used in the simulation studies of biogeochemical processes which link these state variables. For illustration, the flow diagram of Ken Denman's model is shown in Fig.(6.1). The solid lines in this figure represent recipient-controlled flows (uptake by the organisms) from left to right and the dotted line represents the donor-controlled losses from right to left. The flow diagram is labeled with letters indicating the behavior at small and large concentrations. In Fig.(6.1) the letter Q indicates the loss from ZOO to NO₃ is a quadratic function of ZOO. The functional

response of ZOO to PHY approaches quadratic at small PHY concentrations and constant at large ones (QC). Other unlabelled fluxes are by default linear throughout their range.

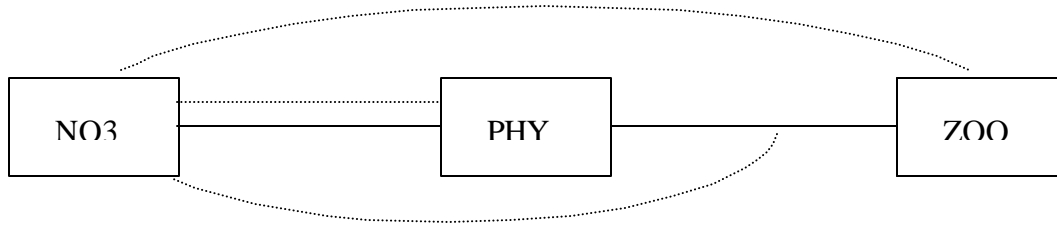


Fig.(6.1) NPZ Model of Ken Denman.

The main input of carbon to ocean carbon cycle described by the NPZ model (Fig.6.1) comes from the atmosphere. The light from the sun reaching the ocean surface is the main energy source controlling the photosynthesis process and it is responsible for the conversion of the inorganic carbon from the atmosphere to organic carbon or the particulate form in the ocean. Now the NPZ model can be quantitatively expressed in the form of a set of ordinary differential equations (Kenman and Gargett, 1995),

$$\begin{aligned} \frac{dPHY}{dt} &= \mathbf{m}_{PHY} \min \left(\frac{\mathbf{a}PAR}{\mathbf{m}_{PHY} + \mathbf{a}PAR}, \frac{NO3}{NO3 + K_{NO3}} \right) PHY - \mathbf{I}_{PHY,NO3} PHY \\ &\quad - \mathbf{m}_{ZOO} \frac{PHY^2}{PHY^2 + K_{FOOD}^2} ZOO \\ \frac{dZOO}{dt} &= (1 - \mathbf{e}_{ZOO,NO3}) \mathbf{m}_{ZOO} \frac{PHY^2}{PHY^2 + K_{FOOD}^2} ZOO - K_{ZOO,NO3} ZOO^2 \\ \frac{dNO3}{dt} &= -\mathbf{m}_{PHY} \min \left(\frac{\mathbf{a}PAR}{\mathbf{m}_{PHY} + \mathbf{a}PAR}, \frac{NO3}{NO3 + K_{NO3}} \right) PHY \\ &\quad + \mathbf{e}_{ZOO,NO3} \mathbf{m}_{ZOO} \frac{PHY^2}{PHY^2 + K_{FOOD}^2} ZOO + \mathbf{I}_{PHY,NO3} PHY + K_{ZOO,NO3} ZOO^2 \quad (6.1a,b,c) \end{aligned}$$

The parameters in the above equations are listed in Table (6.1). The first term on the right hand side of the Eqn.(6.1a) describes the part played by light in the conversion of inorganic to organic carbon. There are various models to define the light propagation in the ocean and for example, Ken Denman expresses it in the form,

$$\text{PAR}(z) = \text{PAR}(0) [V \exp(-1.67 z) + (1-V) \exp(-0.05 z)] \quad (6.2)$$

Where $\text{PAR}(z)$ is the solar radiation at depth z inside the ocean and V is the ratio between the long wave radiation and the total radiation from the sun at the sea surface.

Equations (6.1,a,b,c) were observed to explain most of the general features of biogeochemical processes in the ocean. One may go for elaborate models, once this simple model is inadequate to explain certain special features of the process. For more detailed description of this and other more elaborate models one may refer to Evans and Garcon (1997).

Table (6.1): Parameters in Eqn.(6.1a,bc)

Parameters	Definition	Unit
K_{FOOD}	Half-saturation FOOD concentration for ZOO grazing	mmol N.m^{-3}
K_{NO_3}	Half-saturation FOOD concentration for NO_3 uptake	mmol N.m^{-3}
α_{PAR}	Initial slope of P-I curve	$(\text{Wm}^{-2})^{-1} \text{d}^{-1}$
μ_{PHY}	Maximum phytoplankton growth rate	d^{-1}
μ_{ZOO}	Maximum zooplankton growth rate	d^{-1}
$e_{\text{ZOO,NO}_3}$	Fraction of uptake by ZOO that goes to NO_3	
$\kappa_{\text{ZOO,NO}_3}$	Quadratic loss from ZOO to NO_3	$\text{d}^{-1}(\text{mmol N.m}^{-3})^{-1}$
$I_{\text{PHY,NO}_3}$	Linear loss from PHY to NO_3	d^{-1}

Model describing species distribution due to mixing

In many models the species density is assumed constant over the mixed layer depth, but the mixed layer depth changes with time and one must account for entrainment and detrainment across the bottom boundary of the mixed layer. Then the rate of change of the depth average species concentration can be shown to be equal to,

$$dX/dt = (h_{ml})^{-1} \{ w_t + \max [w + (dh_{ml}/dt), 0] \} (X_0 - X) + S_X \quad (6.3)$$

where X and X_0 are the average species concentration in the mixed layer and below the mixed layer respectively, S_X is the source term describing the variation in the species concentration due to biogeochemical processes, w_t is the turbulent exchange velocity across the base of the mixed layer, w is the background upwelling or down-welling velocity and h_{ml} is the mixed layer depth. The mixed layer depth in Eqn.(6.3) is an input to this model derived from a one-dimensional mixed layer model (see Section 5).

Solving Model Equations

The governing equations describing the species evolution (Eqns.6.1a,b,c) are ordinary differential equations which, for given initial values of the state variables, can be solved using suitable time stepping scheme.

In biogeochemistry studies, we often come across problems involving the estimation of the unknown model parameters rather than simulation / prediction of the evolution of the state variables. That is the time histories of the state variables are generally available and it is required to estimate the parameters of the model such that the model predictions are close to the given state variables. This is essentially an inverse or data assimilation problem governed by a set of non-linear ordinary differential equations describing the evolution of the state variables.

7. SUMMARY

Three main classes of procedures in data analysis or assimilation *viz.* Objective Analysis, Optimal Interpolation and variational method, used to process the observed data on atmospheric and ocean parameters are briefly reviewed. The variational formalism of data assimilation, its mathematical description and the classical variational approach, which leads to the Euler-Lagrange equations are presented. Two methods of solving data assimilation problem, *viz.* the penalty algorithm and the optimal control theory method are discussed.

Application of data assimilation technique to physical problems governed by linear and nonlinear ordinary differential equations are explained. The formulation of data assimilation technique in one-dimensional models of upper ocean and water column biogeochemistry are presented.

References

- Barnes, S. L., 1973, Mesoscale objective analysis using weighted time series observations, NOAA Technical Memorandum, ERL NSSL-62, Environmental Research Laboratories, National Severe Storms Laboratory, Norman, Okla, 60pp.
- Basu, S., Subramanian, V., P. C. Pandey, 1993, The variational technique of data assimilation using adjoint equations in a shallow water model, Proc. Indian Acad. Sci., (Earth Planet Sci.), 102 (4), 521-536.
- Bennett, A. F., 1982, Inverse methods in physical oceanography, CUP, Cambridge.
- Bennett, A. F. and P. C. McIntosh, 1982, Open ocean modeling as an inverse problem: Tidal theory, J. Phy. Oceanography, 12, 1004-1018
- Copeland, G. J. M. and G. L. S. Bayne, 1998, Tidal flow modeling using a direct minimization method, Coastal Engineering, 34, 129-161.
- Courtier, P. and O. Talagrand, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation, II, Numerical results, Q. J. R. Meteorol. Soc., 113, 1329-1348.
- Cressman, P. P., 1959, An operational objective analysis system, Mon. Wea. Rev. 87, 367-374.

- Denman, K. L., 1973, A time-dependent model of the upper ocean, *J. Phy. Oceanog.*, 3, 173-184.
- Denman, K. L. and A. E. Gargett, 1995, Biological physical interaction in the upper ocean: the role of vertical and small scale transport processes, *Ann. Rev. Fluid Mech.*, 27, 225 – 255.
- Derber, J. C., 1985, The variational four dimensional assimilation of analyses using filtered models as constraints, Ph. D. dissertation, Univ. of Wisconsin.
- Evans, G. T. and V. C. Garcon, 1997, One-Dimensional models of water column biogeochemistry, JGOFS Report No. 23, Report of the workshop held in Toulouse, France, 1995.
- Gandin, L. S., 1965, Objective analysis of meteorological fields, (translated from Russian) Isrel Program for Scientific Translations, Jerusalem, 242pp.
- Hofmann, R. N., 1986, A four dimensional analysis exactly satisfying equations of motion, *Monthly Weather Rev.*, 114, 388-397.
- Le Dimet, F. X. and O. Talagrand, 1986, Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, *Tellus. Ser. A.* 38, 97-110.
- Lewis, J. M. and J. C. Derber, 1985, The use of adjoint equations to solve variational adjustments problem with advective constraints, *Tellus. 37 A*, 309-322.
- Lorenc, A. C., 1981, A global three dimensional multivariate statistical interpolation scheme, *Monthly Weather Rev.*, 109, 701-721.
- Lorenc, A. C., 1988, Optimal nonlinear objective analysis, *Q. J. R. Meteorol. Soc.*, 114, 205-240.
- Malanotte-Rizzoli, P. and W. R. Holland, 1986, Data constraints applied to models of the ocean general circulation, Part I: The steady case, *J. Phy. Oceanogr.*, 16, 1665-1682.
- Malanotte-Rizzoli, P. and W. R. Holland, 1986, Data constraints applied to models of the ocean general circulation, Part II: The transient, eddy resolving case, *J. Phy. Oceanogr.*, 16, 1093-1107.
- Provost, C., 1983, A variational method for estimating the general circulation of the ocean, Ph. D. Thesis, Univ. of California, San Diego, 95pp.
- Provost, C. and R. Salmon, 1986, A variational method for inverting hydrographic data, *J. Mar. Res.*, 44, 1-34.

Sasaki, Y., 1955, A variational study of the numerical prediction based on the variational principle, *J. Meteorol. Soc. Japan*, 33, 262-275.

Sasaki, Y., 1958, An objective analysis based on variational method, *J. Meteorol. Soc. Japan*, 36, 738-742.

Smedstad, O. M. and J. J. O'Brien, 1991, Variational data assimilation and parameter estimation in an equatorial Pacific Ocean Model, *Prog. Oceanog.* 26, 179-241.

Talagrand, O. and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equations, I, Theory. *Q. J. R. Metrol. Soc.*, , 1311-1328.

Thacker, W. C., 1988, Fitting model to inadequate data by enforcing spatial and temporal scale, *J. Geophys. Res.*, 93, 10655-10665.

Thacker, W. C. and R. B. Long, 1988, Fitting Dynamics to Data, *J. Geophys. Res.* 93(C2), 1227-1240.

Wunsch, C., 1978, The North Atlantic circulation west of 50° W determined by inverse methods. *Rev. of Geophysics and Space Physics*, 16, 583-620.

APPENDIX - A

OBJECTIVE ANALYSIS OF OCEAN DATA

Introduction

The objective analysis / data processing schemes of Barnes (1964, 1973) are discussed in this appendix.

The objective analysis of data in oceanography refers to the processes used to determine physical variables at uniformly spaced grid points, from a randomly spaced observed data. The objectively analyzed fields of oceanographic data have variety of use in oceanographic studies. These data serve as boundary and / or initial conditions in numerical models of ocean circulation and as a form of "sea truth" for satellite measurements. They can be used for the verification of numerically simulated ocean flow fields and for planning oceanographic field experiments.

One of the schemes which have been used in the past for this purpose, is the surface fitting scheme. In this scheme, a geometrical surface is fitted to the observed data and using this surface, the values of the field are determined at points of interest, specifically at uniformly spaced grid points. Such methods suffer from the following disadvantages:

- The calculations are complicated and requires considerable computer time,
- The data to which the surface is fitted is chosen in a rather arbitrary manner,
- The effect of erroneous data can be disastrous in this approach, since each datum is given equal weightage in determining the shape of the surface.

In order to avoid especially the undesirable effects which the erroneous data produce in surface fitting method, several schemes which employ the least square error method of fitting data (weighted appropriately) have been in use for the objective analyses (*e.g.* Cressman, 1959, and McDonall, 1962). These schemes actually combine the surface fitting method and the weighted averaging method of interpolation. The weighted averaging method determines the value of a variable at a grid point as the sum of the weighted values of the data around the grid point. The

disadvantage of such schemes has been their tendency to filter out all small-scale variations in the field, whether caused by data error or real variations in the phenomenon.

Barnes (1964) proposed a weighted averaging procedure and a scheme of successive corrections to the interpolated value, to regain the small scale variations filtered out or the details lost by applying his averaging procedure. He has also shown that successive estimates of the interpolated variables converge to the real field. The scheme proposed by Barnes (1964) for regaining the lost details is based on the supposition that the two-dimensional distribution of the given variable can be represented by the summation of an infinite number of independent waves, that is by a Fourier integral representation. The drawback of this scheme is the large number of iterations and the computer time required to achieve the desired fit of the interpolated field to the observations. Subsequently Barnes (1973) proposed a modification to his earlier scheme and this requires only one correction pass through the data to achieve the desired fit.

The weighted averaging procedure of Barnes (1964) and the schemes proposed by him (Barnes, 1964 & 1973) for regaining the details lost in the interpolated data using his averaging procedure is discussed in this appendix.

Barnes' Weighted Averaging Scheme

Theoretical basis

We discuss here first the weighted averaging procedure of Barnes (1964) for the interpolation and the schemes (Barnes, 1964 & 1973) for regaining the details lost in the interpolated data using his averaging procedure. We shall consider that a physical variable is observed at randomly spaced points and it is denoted as $f(x, y)$, whose analytical form is not known. In this interpolation scheme, we attempt to estimate the variable at regularly spaced grid points or determine the two-dimensional distribution of interpolated values at grid points of a square mesh. For this purpose we apply a weighting or a filter function $\eta(r)$ on the original function as follows,

$$g(x, y) = \int_0^{2\pi} \int_0^\infty f(x+r\cos \mathbf{q}, y+r \sin \mathbf{q})(1 / 2\pi)\eta(r)d(r^2 / 4k)d\mathbf{q} \quad (\text{A1})$$

The coordinate system used in this analysis is shown in Fig.(A1). In the above integral,

$$\mathbf{h}(r) \equiv \exp(-r^2 / 4k) \quad (\text{A2})$$

and k defines the shape of the weight function, which is to be related to the density of the observed data. The interpolated field $g(x, y)$ will evidently be different from the true field.

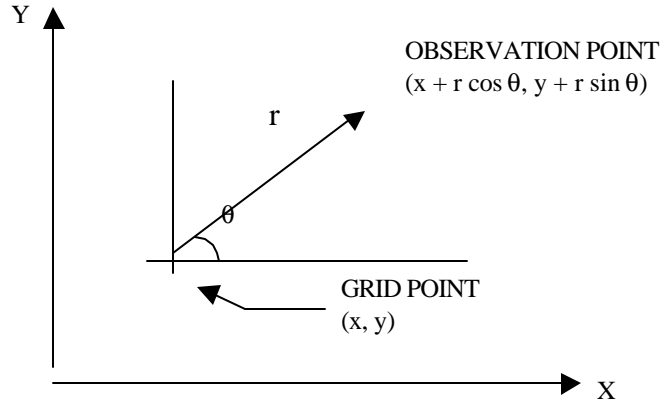


Figure A1. Coordinate system used in the analysis.

We shall examine here the influence of weighted averaging on the field $f(x, y)$, treating $f(x, y)$ as a continuous function over the region of interest and considering that it can be represented by the summation of an infinite number of independent harmonic waves, that is by a Fourier integral representation. In the following analysis we shall refer only to a single harmonic wave, keeping in mind that the discussions apply equally well to all the harmonic components present in $f(x, y)$. To further simplify the discussions, we shall consider only a one-dimensional wave,

$$f(x) = A \text{Sin } a x, \quad (\text{A3})$$

where $a = 2\pi / \lambda$ and λ is the wavelength of the harmonic component. Substitution of (A3)

in (A1) yields,

$$g(x) = D A \text{Sin } a x = D f(x), \quad (\text{A4})$$

where $D = \exp(-k a^2)$ is referred to as the accuracy index or the response function, and it is a function of 'k' and 'a', which are related to the shape of the weighting function and to the wave length of the harmonic component respectively. We may note that the application of the weighted averaging does not alter the phase of the harmonic component and only reduces the

amplitude of the original wave. Fig.(A2) shows the response function 'D' for different wave lengths ' λ ' of the harmonic components and weight function factors ' k '.

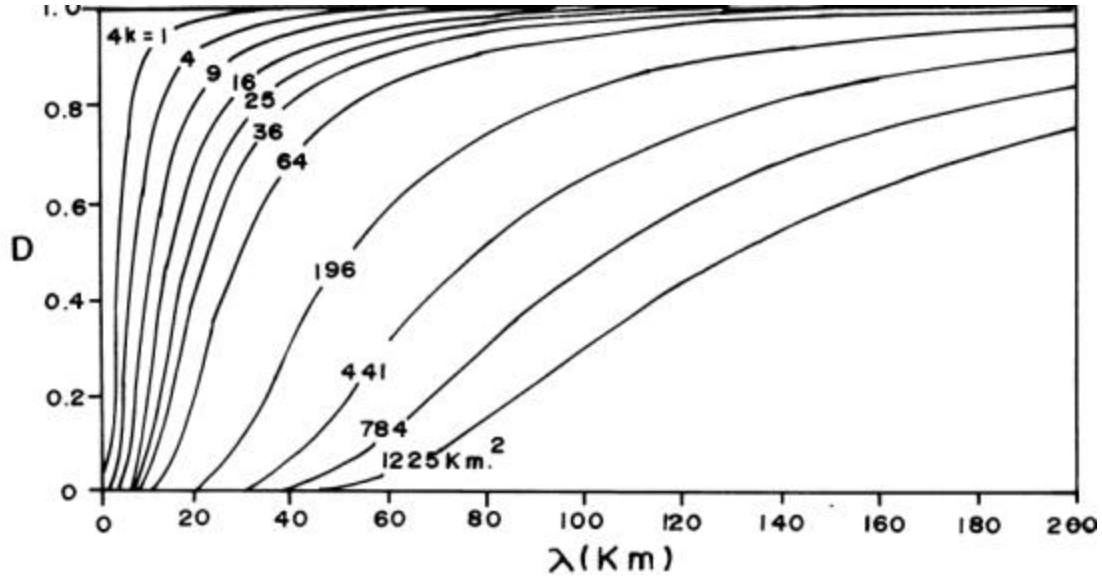


Figure A2 Response function, $D = \exp(\mathbf{k}4\mathbf{p}^2 / \lambda^2)$, for different κ and λ

To regain the reduction in the wave amplitude due to the averaging procedure, Barnes (1964) suggested successive application of the same filter (weight function) to the difference between the functions $g(x)$ and $f(x)$ to estimate the correction to be applied to the successive estimates of $g(x)$. From Eqn.(A4), we shall write the first guess or the first estimate as

$$g_0(x) = D f(x), \quad (\text{A5})$$

where the subscript zero denotes the first estimate. Then applying the same filter function $\eta(r)$ on $[f(x) - g_0(x)]$ and correcting $g_0(x)$, the second estimate becomes

$$\begin{aligned} g_1(x) &= g_0(x) + [f(x) - g_0(x)] D \\ &= f(x) D [1 + (1 - D)] \end{aligned} \quad (\text{A6})$$

At this stage, the response function $D' = D [1 + (1 - D)]$. Then after N successive corrections,

$$g_N(x) = f(x) D \sum_{n=0}^N (1 - D)^n. \quad (\text{A7})$$

and the response function $D' = D \sum_{n=0}^N (1 - D)^n$. When $N \rightarrow \infty$, we can show that,

$$D' \rightarrow 1 \quad \text{and} \quad g_N(x) \rightarrow f(x). \quad (\text{A8})$$

Since the premise of the interpolation scheme is that the waves which make up the distribution are independent, the above discussion are valid individually to each component wave. Hence, when the above iteration scheme is carried over indefinitely large number of times, all the waves which combine to form the data will regain their full amplitudes. The number of iterations desirable in this correction scheme depend upon the quality of the data and the details we want to retain in the analyzed field. For example, if the data are error free or the errors involved are small and if one desires to depict the maximum details exhibited by the data, a large number of iterations are required to force the interpolated field to converge close to the observed data field. On the other hand, if the data errors are known to be fairly large, or if one wishes to see only the broad features of a field, the analysis is carried through a lesser number of iterations. The main drawback of this scheme is the large number of iterations required to achieve adequate response to different waves present in the data field.

Barnes (1973) devised an alternative scheme of correction to the above weighted averaging procedure, in which he decreased the weight function parameter in the first correction pass. This scheme requires only one correction pass through the data. Following Barnes (1973), we write the first estimate as

$$g_0(x, y) = D_0 f(x, y), \quad (\text{A9})$$

Subscript zero denotes the first estimate with weight function $\eta_0(r) = \exp(-r^2/4k_0)$ and response function $D_0 = \exp(-a^2k_0)$. Then the correction to the first estimate is determined using a weight function,

$$\eta_1(r) = \exp(-r^2/4k_1), \quad (\text{A10})$$

with $k_1 = \gamma k_0$ and $0 < \gamma < 1$. Then the second estimate becomes,

$$g_1(x, y) = g_0(x, y) + [f(x, y) - g_0(x, y)] D_1 \quad (\text{A11})$$

where

$$D_1 = \exp(-a^2k_1) = \exp(-a^2\gamma k_0) = D_0^\gamma \quad (\text{A12})$$

Substituting (A12) and (A9) in (A11), we get

$$g_1(x, y) = f(x, y) D_0 [1 + D_0^{\gamma-1} - D_0^\gamma] \quad (\text{A13})$$

In this alternate scheme the response function becomes

$$D' = D_0 [1 + D_0^{\gamma-1} - D_0^\gamma] \quad (\text{A14})$$

and Fig.(A3) shows the relation between D' and D_0

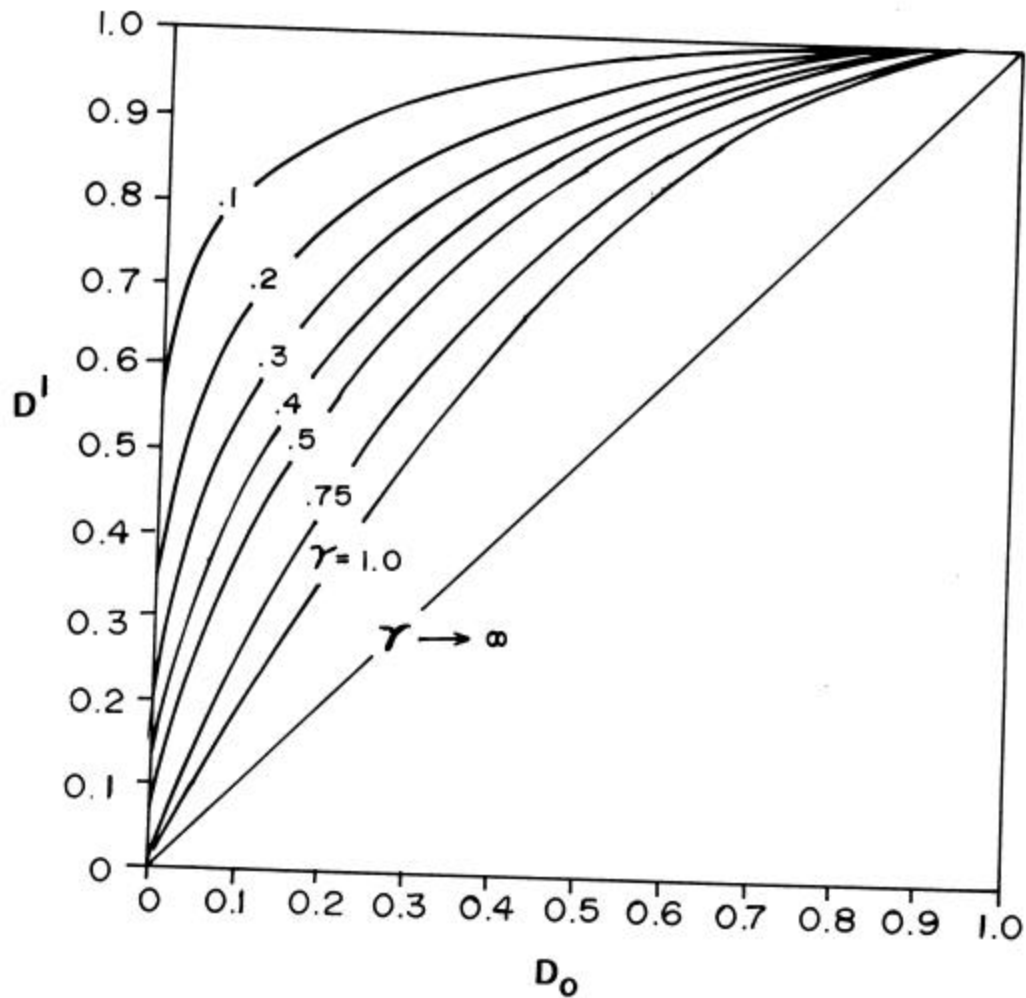


Figure A3 Response function D' after one correction pass as a Function of D_0 (Eqn. A13)

Numerical Scheme

Throughout the above discussion we have assumed that a complete description of $f(x, y)$ is available at all points over the region of interest. In the real world, analytical form of $f(x, y)$ is not known and we have information only at finite number of randomly spaced points. Hence, the integration over the infinite region defined in Eqn.(A1) is not possible. We must approximate $g(x, y)$ by placing a finite limit on the region of influence of $f(x, y)$, which is assumed to be a circular region of radius R around the point (x, y) . R is defined as the radius of influence.

With a finite integration region (region of influence) and with finite number of data points within this region, the weighted averaging defined in Eqn.(A1), becomes

$$g(x, y) = \left(\sum_j^M [\mathbf{h}(r_j) f_j] \right) / \left(\sum_j^M \mathbf{h}(r_j) \right). \quad (\text{A15})$$

To establish a relation between the radius of influence R and the weight factor k , we shall consider the following integral of the weight function (Eqn.A2),

$$I = \int_0^{2\mathbf{p}} \int_0^\infty \frac{1}{2\mathbf{p}} \exp(-r^2 / 4k) d(r^2 / 4k) d\mathbf{q} = 1. \quad (\text{A16})$$

Integrating Eqn.(A16) with respect to θ first, we get

$$\begin{aligned} I &= \int_0^\infty \exp(-r^2 / 4k) d(r^2 / 4k) \\ &= \int_0^R \exp(-r^2 / 4k) d(r^2 / 4k) + \int_R^\infty \exp(-r^2 / 4k) d(r^2 / 4k) = 1. \end{aligned} \quad (\text{A17})$$

We define the second of the above integrals as ϵ . Then we have

$$\int_0^R \exp(-r^2 / 4k) d(r^2 / 4k) = 1 - \epsilon.$$

Carrying out the integration, we see that

$$\exp\left(-R^2 / 4k\right) = \epsilon$$

$$\text{or} \quad R^2 / 4k = -\ln \epsilon \equiv E. \quad (\text{A18})$$

If ϵ is small enough, we can represent the weighted influence of any datum with sufficient accuracy. For example, choosing $E = 4$ means we have represented 98 % of the influence of any datum within the circular region whose radius is R . We define R as the ‘radius of influence’ of the weight factor $\eta(r)$. Eqn.(A18) will be used later to form a link between the data distribution and the choice of weight factor.

Steps involved in the analysis

Barnes' (1964) scheme:

Barnes (1964) proposed the following steps for the interpolation of the randomly spaced observed data:

- (1) Chose a weight factor appropriate to the data distribution and the wavelength of the harmonic component to be depicted in the analyzed field.
- (2) Perform the initial interpolation at the grid points using $g_0(x, y) = D_0 f(x, y)$.
This gives the first interpolated field.
- (3) At each data point, subtract the previous interpolated value from the measured value, which gives the difference field.
- (4) Determine the correction field by applying the same weight function (as was used in step 2) on the difference field obtained in step 3.
- (5) Add the correction field obtained in step 4 to the previous interpolated field to get the next interpolated field.
- (6) Repeat steps from 3 to 5 until the difference field diminishes to the desired level.

Barnes' (1973) scheme:

In this scheme, the first interpolated field is obtained by following the first 3 steps of Barnes' (1964) scheme. Next the correction field is determined from the difference field using a weight function $\eta_1(r) = \exp(-r^2 / 4 k_1)$, where $k_1 = \gamma k_0$, $0 < \gamma < 1$ and k_0 is the weight factor at the first stage of interpolation. Here we need to choose a proper value for γ . Then the final interpolated field is determined by adding the correction field to the first interpolated field.

Guidelines for choosing the parameters for the analysis:

To choose the parameters in the analysis, we shall consider a typical example. Suppose the observations are available with an average spacing (d) of 20 km. Then the resolvable scale will be $2d$, i.e. 40 km. Barnes (1964) defines $R = 1.6 d$ in one of the examples. Then using the relation (Eqn.A18) $R^2 / (4 k) = E$, we can determine the weight factor k . In the present example when $E = 4$, k becomes 64.

In Barnes (1973) scheme, in addition to the specification of R and k , we need to choose a suitable value for γ . Barnes (1973) observes that γ cannot be smaller than 0.2 without creating underflow checks on most computer systems. Also specifying $\gamma > 0.5$ does not produce rapid convergence. Empirically, an optimum value for γ appears to be between 0.2 and 0.4. In an another example, Barnes (1973) requires the analysis to depict at least 90% amplitude of the smallest resolvable scale of variation. He selects $\gamma = 0.3$. Then from the plot (Fig.A4) of $D' = D_0 [1 + D_0^{\gamma-1} - D_0^\gamma]$, for $\gamma = 0.3$ and for various values of λ and k , he determines the weight factor. In the present example the resolvable scale (i.e. λ) is 40 km and $D' = 90\%$ and we get $4k \approx 110 \text{ km}^2$ from Fig(A4).

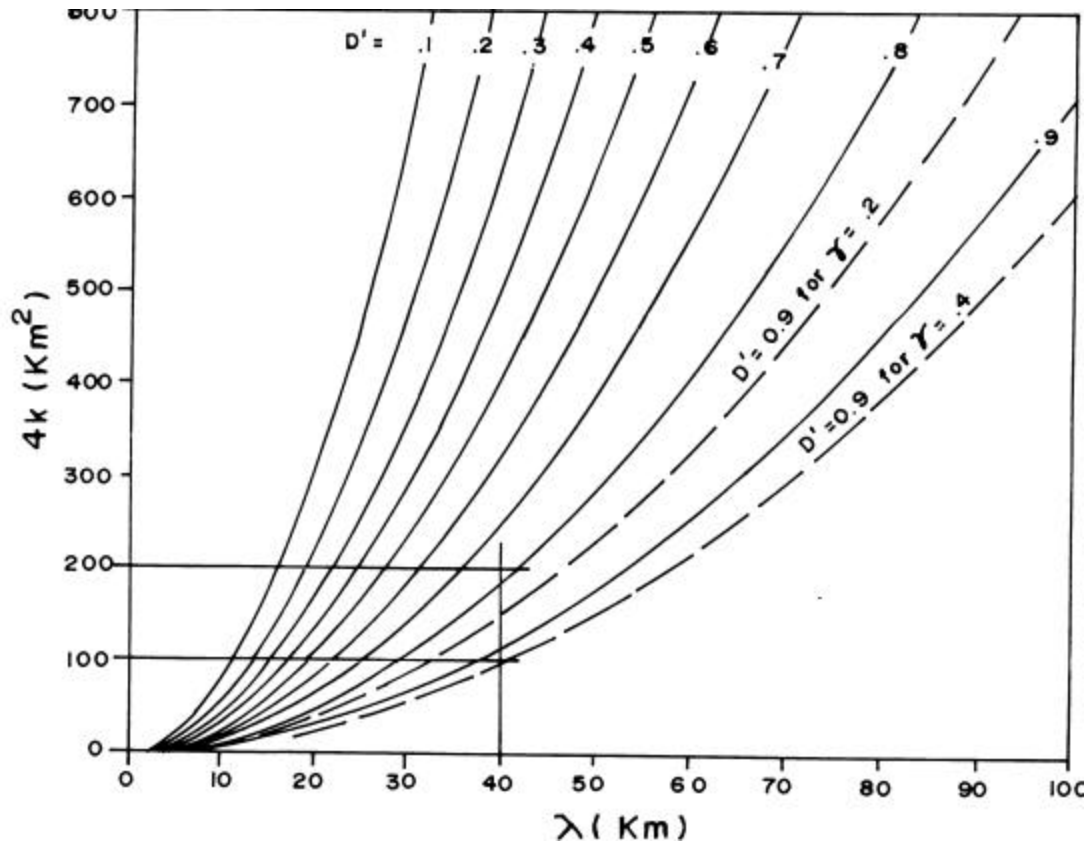


Figure A4 Response function D' (Eqn A13) as a function of λ and $4k$
 Levitus *et al.* (1

Levitus *et al.* (1994) used Barnes' (1973) scheme for the analysis of randomly spaced salinity data. For the first interpolated field, they used the zonal average or the annual mean or the seasonal mean of the measured salinity field. They assumed $R = 550$ km and have not specified the value γ used in their analysis.

REFERENCES

Barnes, S.L. 1964, A technique for maximizing details in numerical weather map analysis. *J. App. Meteor.*, 3, 396-409.

Barnes, S.L. 1973, Mesoscale objective map analysis using weighted time series observations. *NOAA Technical Memorandum ERL NSSL-62*, 60 pp.

Cressman, G.P. 1959, An operational objective analysis scheme. *Mon. Wea. Rev.*, 87, 367-374.

Levitus, S., R. Burgett AND t.p. Boyer, 1994, NOAA Atlas NESDIS 3, National Oceanographic Data Center, User Service Branch, NOAA/NESDIS/OC21, 1825 Connecticut Avenue, NW, Washington, DC 20235.

McDonnell, J.E. 1962, On the objective analysis system used at the National Meteorological Center. Tech. Memo. No. 23, U.S. Weather Bureau, 31 pp.

APPENDIX-B

EXTREME VALUE PROBLEMS

In this section, extreme value and constrained extreme value problems of functions of known analytical forms are discussed.

Extreme value problem of a multidimensional function

The necessary condition for a function $F(x_n)$ of variables x_n , where $n = 1, 2, \dots, N$, in an N - dimensional space, to have an extreme value at a certain point in this space is that the first order partial derivatives of $F(x_n)$ with respect to x_n should vanish at that point. That is,

$$(\partial F / \partial x_n) = 0 \quad (B1)$$

By solving the above equations, the point x_n , where $F(x_n)$ takes the extreme value can be determined.

Example:

$$\text{Let } F(x, y) = x^2 - 2x + 4y^2 + 4y + 5$$

$$(\partial F / \partial x) = 2x - 2 = 0$$

$$(\partial F / \partial y) = 8y + 4 = 0$$

That is, F has an extreme value at $x = 1$, and $y = -0.5$. and the extreme value of $F = 3$.

Constrained - Extreme Value Problem

In this problem we seek the point x_n in a multidimensional space where the function $J(x_n)$ takes an extreme value and at the same time satisfies $F(x_n) = 0$. Here $F(x_n) = 0$ is the constraint imposed on the minimization problem. Then, it can be shown that, seeking the constrained extreme value of J is equivalent to finding the unconstrained extremum of the Lagrangian, L , which is defined in terms of $J(x_n)$ and $F(x_n)$ as,

$$L[x_n, \lambda] = J(x_n) + \lambda F(x_n) \quad (\text{B2})$$

where λ is called the Lagrangian multiplier. Finding the extreme value of $L[x_n, \lambda]$ requires the partial derivatives,

$$(\partial L / \partial x_n) = 0, \quad (\partial L / \partial \lambda) = 0 \quad (\text{B3})$$

By solving the above set of equations, the required solution x_n can be determined.

Example:

As an example of this method, let us consider the problem of finding the largest rectangle that can be inscribed in an ellipse,

$$(x^2/a^2) + (y^2/b^2) = 1$$

In this case, the function to be maximized is the area of the rectangle, $A = 4xy$. To solve this problem we will form the Lagrangian,

$$L(x, y, \lambda) = 4xy + \lambda [(x^2/a^2) + (y^2/b^2) - 1]$$

and find its maximum. For the maximum of the Lagrangian, we require,

$$(\partial L / \partial x) = 4y + 2\lambda x/a^2 = 0$$

$$(\partial L / \partial y) = 4x + 2\lambda y/b^2 = 0$$

$$(\partial L / \partial \lambda) = (x^2/a^2) + (y^2/b^2) - 1 = 0$$

Solving the above equations, we get $x = a/\sqrt{2}$ and $y = b/\sqrt{2}$, for the area of the rectangle to be maximum.

APPENDIX-C

DESCENT ALGORITHMS

Let us consider a scalar function F that depends on one or more number of independent variables is defined, whose analytical form is not known, and that we seek the values of the variables for which the function F takes an extreme value, *i.e.* either a maximum or a minimum. Often the computational effort in the minimization procedure is dominated by the cost of evaluating the function, F , and its partial derivatives (if the chosen algorithm requires them) with respect to the variables. Hence, the algorithm for finding the extremum should be such that, it should compute F and its derivatives as few times as possible and should require less memory space in the computer.

Extreme value of one - dimensional function

In one - dimensional problem, after bracketing the range of the variable within which the extremum of the function lies, the exact location of the extremum can be determined by a procedure which is similar to the bisection method used for finding the roots of a one-dimensional function. Other procedures are also available for this one-dimensional search of the extremum, also referred to as line search procedure.

Extreme value of a multidimensional function

In a multidimensional problem, we seek the extremum of a scalar function $F(x_n)$ in an N -Dimensional space. If $F(x_n)$ is not a known analytical function, then the general procedure to determine the point x_n where $F(x_n)$ takes the extreme value, is to choose ' N ' independent unit vectors, which define a set of N directions in the solution space. Now with the available information on the given function ' F ', the initial guess of the desired extreme point is made. From this point the line search is carried out along the direction of the first vector and a local minimum point is determined. From this point the line search is continued along the direction of the second vector and a local minimum is again fixed along this direction. This procedure is continued along all the N chosen vector

directions and then repeated several times if necessary, until, the function stops decreasing below a specified tolerance limit. In this procedure we do not require the computation of the gradient of the function. For example the N unit vectors can be chosen along the coordinate direction of the N-dimensional space. This simple procedure is actually not too bad for many functions.

Method of steepest descent

If the gradient ∇F of the given function could be computed, we can start the one-dimensional search in the direction parallel to the gradient of the function at the initial guess of the desired extreme point and the local minimum point is determined. Then from this point the line search procedure is carried out along the gradient of the given function at this local minimum point. This procedure is repeated until the absolute minimum point is reached. Note that at each local minimum point we move perpendicular to the earlier search direction and the variation of the function is steepest along the gradient direction.

Conjugate gradient method

In this method, we again start from a suitably chosen point in the N-dimensional vector space and search for the local minimum along an arbitrarily chosen direction. Note that if we search for a minimum of a function in the direction u , then the gradient of the function must be perpendicular to u at the local minimum point in this direction. Now we select the local minimum point P as the origin of the coordinate system and approximate F using the Taylor series expansion.

$$F(x) = C - b \cdot x + \frac{1}{2} x \cdot A \cdot x \quad (C1)$$

where $x = (x_1, x_2, \dots, x_N)$ is the coordinate of a point with P as the origin in the N-dimensional space. $C = F(P)$, the vector $b = -\nabla F|_P$ and the matrix

$$A = \left. \frac{\partial^2 F}{\partial x_i \partial x_j} \right|_P$$

A is called the Hessian of the function F at P . Now around P , the gradient of F becomes

$$\nabla F = A x - b \quad (C2)$$

and the variation of ∇F in some direction around P is

$$\delta(\nabla F) = A \cdot \delta x \quad (C3)$$

Suppose we now propose to move along some new direction v . The condition that the search along v will not spoil our minimization along u is just the change in the gradient be perpendicular to u , i.e.

$$u \cdot \delta(\nabla F) = u \cdot Av = 0 \quad (C4)$$

when this condition holds for two vectors u and v , they are said to be conjugate. If we do successive line minimization of a function along a conjugate set of directions, the convergence to the absolute minimum is observed to be fast. Note that this method requires the estimation of the Hessian matrix.