Ocean Data Assimilation

Robert N. Miller

College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97330

Outline

- Review: Linear inverse methods and the Kalman filter
- Examples
- Generalization to noisy nonlinear systems
- Summary

Review

“Data assimilation refers to three problems in time series analysis. Given a time series ω_k, or possible a continuous function of space and time $\omega(x,t)$ which may be noisy or incomplete, beginning with time $t = -T$ and ending at $t = 0$, the “present,” define three problems:

- The prediction problem What will ω be in the future?
- The filtering problem What is the best estimate of ω *now*, i.e., at $t = 0$?
- The smoothing problem: What is the best estimate of ω for the entire time series?

Origins of Data Assimilation

Gauss and Legendre were interested in *planetary orbits*.

- These are specified by 6 parameters, the *orbital elements*.
- Three observations are necessary to determine the orbital elements.
- If more than three observations are available choose elements to minimize:

$$\sum(\textit{predicted position} - \textit{observed position})^2$$

This is the *least squares method*

Variational Methods

Given

- A model: $\mathbf{u}_t - L\mathbf{u} = \mathbf{f}$, possibly a linear equation that describes the evolution of small deviations from a first-guess solution.
- Chosen to mimic the "true" state $\mathbf{u}^{(t)}$ assumed to evolve according to $\mathbf{u}_t^{(t)} - L\mathbf{u}^{(t)} = \mathbf{f} + \mathbf{b}$ for some random function $\mathbf{b}$
- Estimated initial condition $\mathbf{u}(0)$ with random error $\mathbf{e}_0$
- Observations $\mathbf{z} = H\mathbf{u}^{(t)} + \mathbf{e}_{obs}$

Variational Methods

Minimize the cost function:

$$\begin{aligned} J(\mathbf{u}) \;&=\; \int (\mathbf{u}_t - L\mathbf{u} - \mathbf{f})^T W^{-1} (\mathbf{u}_t - L\mathbf{u} - \mathbf{f}) dt + \\ &\quad (\mathbf{u}(0) - \mathbf{u}_0)^T V^{-1} (\mathbf{u}(0) - \mathbf{u}_0) + \\ &\quad (\mathbf{z} - H\mathbf{u})^T R^{-1} (\mathbf{z} - H\mathbf{u}) \end{aligned}$$

The minimizer of J is the BLUE of $\mathbf{u}^{(t)}$ if:

$$\begin{aligned} E(\mathbf{b}\mathbf{b}^T) &= W \\ E(\mathbf{e}_0\mathbf{e}_0^T) &= V \\ E(\mathbf{e}_{obs}\mathbf{e}_{obs}^T) &= R \end{aligned}$$

Variational Methods

- We begin with $\mathbf{u}$ a (possibly) vector-valued function of time.
- This formulation generalizes naturally to functions of time and space, in which case:
  - L would be a partial differential operator
  - The constraint on the initial condition would be an integral
  - There might be a constraint on the boundary conditions.

We will derive all of the linearized methods from here.

The Variational Method

Without loss of generality, we can set $f = u_0 = 0$. so:

$$J(\mathbf{u}) = \int (\mathbf{u}_t - L\mathbf{u})^T W^{-1} (\mathbf{u}_t - L\mathbf{u}) dt$$
$$+\mathbf{u}(0)^T V^{-1} \mathbf{u}(0) + \sum_{j=1}^{N} R_j^{-1} (z_j - H_j \mathbf{u}(t_j))^2$$
$$\equiv \ < \mathbf{u}, \mathbf{u} > + \sum_{j=1}^{N} R_j^{-1} (z_j - H_j \mathbf{u}(t_j))^2$$

The cost function defines a positive definite bilinear form $< \cdot, \cdot >$ (Think dot product)

Vectors and Functions

Consider a scalar valued linear function $f(\mathbf{v})$, i.e., the domain of f is $\mathbb{R}^n$ and the range is $\mathbb{R}$.

$$\mathbf{v} = \sum_j v_j \mathbf{e}_j$$

so

$$f(\mathbf{v}) = \sum_j v_j f(\mathbf{e}_j) \equiv \mathbf{v} \cdot \mathbf{a}$$

where $\mathbf{a}_j = f(\mathbf{e}_j)$.
... now imagine that $\mathbf{v}$ is a function instead of a vector.

The Representer Method

Define the j^{th} *representer* r_j:

$$< r_j, u >= H_j u(t_j)$$

for any admissible function u

- The representer *represents* the measurement functional in terms of the new inner product.
- This allows us to form an orthogonal decomposition of the space of admissible functions.

Orthogonal Decomposition of State Space

Write the minimizer $\hat{u}$ of the functional J, as:

$$\hat{u} = \sum_{j=1}^{N} b_j r_j + G$$

where the b_j are constants and

$$< r_j, G >= 0, \; j = 1, \ldots, N$$

Solution in Representer Space

The cost function then becomes:

$$
\begin{aligned}
J(u) &= \sum_{i,j=1}^{N} b_i b_j < r_i, r_j > + < G, G > + \\
& \sum_{j=1}^{N} R_j^{-1} (z_j - \sum_i b_i < r_i, r_j >)^2
\end{aligned}
$$

- We might as well pick $G = 0$
- Picking nonzero G doesn't change the data misfit and can only increase the cost.

The Representer Method

The original infinite dimensional problem is reduced to finding a finite number of coefficients b_j:

$$\frac{\partial J}{\partial b_k} = 2\sum_j b_j < r_j, r_k > - \\ 2\sum_j R_j^{-1}(z_j - < r_j, \sum_i b_i r_i >) < r_j, r_k >$$

Setting $\partial J / \partial b_k = 0$ leads to:

$$\sum_j < r_j, r_k > \left(R_j b_j + \sum_i < r_i, r_j > b_i - z_j \right) = 0$$

The Representer Method

$$\sum_j < r_j, r_k > \left(R_j b_j + \sum_i < r_i, r_j > b_i - z_j \right) = 0$$

In matrix form. Define $R = diag(R_j)$ and $M_{i,j} =< r_i, r_j >$ the *representer matrix*. The solution is then defined by:

$$(M + R)\, b = z$$

where b is the vector of representer coefficients and z is the vector of observations.

What Value Should the Cost Function Be at Minimum?

At the minimum,

$$\begin{aligned} J &= z^T(M+R)^{-1}M(M+R)^{-1}z + \\ &\quad (z - M(M+R)^{-1}z)^T R^{-1}(z - M(M+R)^{-1}z \\ &\quad (lots\ of\ algebra \ldots) \\ &= z^T(M+R)^{-1}z \end{aligned}$$

So z should be a random variable with covariance $M+R$ and J is a random variable with χ^2 distribution on M degrees of freedom.

Computing Representers

Schematically:

$$< \mathbf{u}, \mathbf{v} > \sim (\mathsf{M}\mathbf{u}, \mathsf{M}\mathbf{v}); \mathsf{M} \equiv \frac{\partial}{\partial t} - L$$

We want:

$$(\mathsf{M}\mathbf{u}, \mathsf{M}\mathbf{r}) = (\mathbf{u}, \mathsf{M}^*\mathsf{M}\mathbf{r}) = (\mathbf{u}, \delta)$$

So solve:

$$\begin{aligned} \mathsf{M}^*\alpha &= \delta \\ \mathsf{M}r &= \alpha \end{aligned}$$

Computing Representers

Begin with the simplest case: a linear, scalar ODE:

$$\dot{u} - au = F$$

F, $u(0)$ unknown. First guess: $F = 0; \; u(0) = 0$
Given measurements y_j of the system at times t_j

$$\begin{aligned} J &= \int_0^T (\dot{u} - au) W^{-1} (\dot{u} - au) dt + u(0) V^{-1} u(0) + \\ & \quad \sum (y_j - u(t_j))^2 / R_j \\ &\equiv < u, u > + \sum (y_j - u(t_j))^2 / R_j \end{aligned}$$

Computing Representers

The j^{th} representer is defined by
$< r_j, v >= v(t_j) = \int_0^T \delta(t - t_j)v(t)dt$

1. Calculate the *representer adjoint* α_j, such that:

$$\begin{aligned}
\int_0^T \alpha_j(\dot{v} - av)dt &= \int_0^T (-\dot{\alpha}_j - a\alpha_j)vdt + \\
&\quad \alpha_j v\big|_0^T;\ \alpha_j(T) = 0 \\
&= \int_0^T \delta(t - t_j)vdt;
\end{aligned}$$

2. Then solve $\dot{r} - ar = \alpha_j W; r(0) = V\alpha_j(0)$

Summary of the Representer Method

- The linear inverse problem is potentially a minimization problem over ∞ dimensions
- In practice the observations determine only a finite number of degrees of freedom
- A quadratic cost function can define a useful orthogonal decomposition of state space into two components:
  - The space spanned by the representers
  - Its orthogonal complement, all members of which are *unobservable*, i.e., they give measurements with value zero, by construction.

Summary, continued

- The minimization can thus be carried out over the space of representers
- A potentially ∞ dimensional problem is reduced to a finite dimensional one
- The representers can (but need not be) calculated explicitly
- The representers do not depend on the data weights

The Variational Approach

Calculate the first variation δJ of the cost function J and set $\delta J = 0$ A slightly more general cost function:

$$
\begin{aligned}
J(u) \;=\; & \frac{1}{2}\int_0^T \int_\Omega \int_\Omega (u_t(x_1,t) - Lu) W^{-1} \\
& (u_t(x_2,t) - Lu)\, dx_1 dx_2 dt + \\
& \frac{1}{2}\int_\Omega \int_\Omega u(x_1,0) V^{-1} u(x_2,0)\, dx_1 dx_2 + \\
& \frac{1}{2} z^T R^{-1} z
\end{aligned}
$$

where z is the innovation vector, with components $z_j = y_j - H_j u$.

The Variational Approach

As before, write:

$$\lambda = (u_t - Lu)W^{-1}$$

For $u \longrightarrow u + \delta u$ set $\delta J = J(u + \delta u) - J(u) = O(\delta u^2)$

The Euler-Lagrange Equations

$$
\begin{aligned}
-\lambda_t - L^*\lambda &= z^T R^{-1} H \\
\lambda(T) &= 0 \\
u(x,0) &= \lambda(x,0)v(0) \\
u_t - Lu &= W\lambda
\end{aligned}
$$

Write $\lambda = \sum_j a_j \alpha_j$ where the α_j are the *representer adjoints*:

$$
\begin{aligned}
-\alpha_{jt} - L^*\alpha_j &= H_j \delta(t - t_j) \\
\alpha(T) &= 0
\end{aligned}
$$

$\longrightarrow$ the representer solution: Bennett (1992, 2002) or the tutorial at http://iom.asu.edu.

Variational Methods: Summary

- The simplest and most common variational methods work by minimization of a quadratic cost function.
- In most problems in ocean data assimilation,the state function that minimizes the mean square data misft is not unique
- The quadratic cost function can be used to define a decompositon of state space into the space spanned by the representers and its orthogonal complement
- By construction, elements orthogonal to the representers have no effect on the model-data misfits, and can therefore be neglected in most cases.

Adaptation to Nonlinear Problems

So far we have dealt with *linear* problems. What to do if the underlying model is *nonlinear*?

1. Calculate solution to nonlinear model, with best estimate of IC, BC and forcing.
2. Calculate the solution to the linear inverse problem for deviations from the nonlinear solution
3. Add the resulting increments to IC, BC and forcing to the original estimates
4. Go to 1., with new, IC, BC and forcing; iterate (hopefully) to convergence

The Filtering Problem

Given a time series ω_k, or possible a continuous function of space and time $\omega(x, t)$ which may be noisy or incomplete, beginning with time $t = -T$ and ending at $t = 0$, the "present," What is the best estimate of ω?

Given current observations, we will *not* revise our estimate of past states.

Filtering

Consider a single step of a prediction-analysis cycle:

1. Given an initial condition $\mathbf{u}_0$ at $t = t_0$, predict the new state $\mathbf{u}_1$ at the next time t_1: $\mathbf{u}_1^f = L\mathbf{u}_0$.
2. Given observations $\mathbf{y}$ at time t_1, form an improved estimate $\mathbf{u}_1^a = \mathbf{u}_1^f + \mathbf{v}_1$ of the state $\mathbf{u}_1$
3. In most cases, choose $\mathbf{v}_1 \propto \mathbf{y} - H\mathbf{u}_1^f$, where $H\mathbf{u}_1^f$ is the predicted value of the observed quantity.

Filtering: Variational Formulation

Cost function:

$$\begin{aligned} J &= \mathbf{v}_0^T P_0^{-1} \mathbf{v}_0 + (\mathbf{v}_1 - L\mathbf{v}_0)^T Q^{-1} (\mathbf{v}_1 - L\mathbf{v}_0) \\ &\quad + (\mathbf{z} - H\mathbf{v}_1)^T R^{-1} (\mathbf{z} - H\mathbf{v}_1) \end{aligned}$$

$$\mathbf{z} = \mathbf{y} - H\mathbf{u}_1^f$$

Filtering: Variational Formulation

Minimization of J by the representer method leads to:

$$\mathbf{v}_1 = (LP_0L^* + Q)H^T \left[H(LP_0L^* + Q)H^T + R\right]^{-1} \mathbf{z}$$

Recall $\mathbf{v}_1$ is the correction to the first guess $\mathbf{u}_1^f$.

Putting it all together

$$\mathbf{u}_1^a = \mathbf{u}_1^f + (LP_0L^* + Q)H^T \left[H(LPL^* + Q)H^T + R\right]^{-1} \mathbf{z}$$

This is usually broken down into steps:

1. $\mathbf{u}_1^f = L\mathbf{u}_0$
2. $P_1^f = LP_0L^* + Q$
3. $K = P_1^f H^T \left[HP_1^f H^T + R\right]^{-1}$
4. $\mathbf{u}_1^a = \mathbf{u}_1^f + K(\mathbf{y} - H\mathbf{u}_1^f)$

Statistics

We assume our model, given by:

$$\mathbf{u}_{j+1} = L\mathbf{u}_j$$

differs from the “truth” by some random error ϵ

$$\mathbf{u}^t_{j+1} = L\mathbf{u}^t_j + \epsilon$$

ϵ is white in time with covariance $E(\epsilon\epsilon^T) = Q$
The error in the state is given by $\mathbf{e}_0 = \mathbf{u}^t_0 - \mathbf{u_0}$ with covariance $P_0 = E(e_0 e_0^T)$ at time $t = 0$.
The observation error is white with mean zero and covariance R.

Filtering: Statistics

Then:
The state error covariance evolves according to:

$$P_1^f = E(\mathbf{e}_1\mathbf{e}_1^T) = LE(\mathbf{e}_0\mathbf{e}_0^T)L^* + Q$$

The error in the corrected state should be smaller than the error in the original state. The covariance of the error in the updated state is:

$$P_1^a = (I - KH)P_1^f$$

The Filter Solution

Putting it all together:

1. $\mathbf{u}_1^f = L\mathbf{u}_0$
2. $P_1^f = LP_0L^* + Q$
3. $K = P_1^f H^T \left[H P_1^f H^T + R \right]^{-1}$
4. $\mathbf{u}_1^a = \mathbf{u}_1^f + K(\mathbf{y} - H\mathbf{u}_1^f)$
5. $P_1^a = (I - KH)P_1^f$

This is the *Kalman Filter.*

Remarks

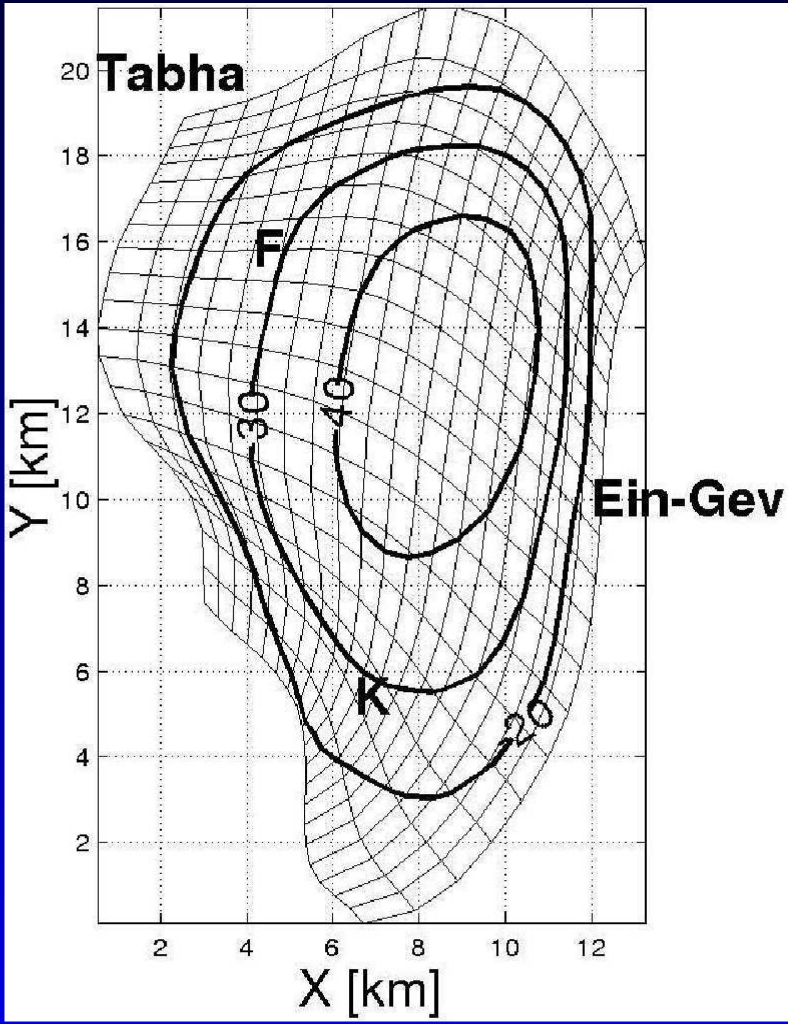
- This is one of many ways to derive the Kalman filter
- Implementation is straightforward, but potentially very expensive
- Not necessary to write complex adjoint code

Remarks

- There are many natural generalizations and simplifications of the KF:
  - The *extended Kalman filter*: Use a nonlinear model for the state evolution and linearized dynamics to calculate the evolution of the error covariance.
  - Use a static error covariance P and eliminate the repeated calculations.
  - Use a collection of model runs with randomly chosen initial conditions and forcing to calculate an approximate covariance. This is the *ensemble Kalman filter*
  - Neglect errors outside of a low-dimensional subspace of the full state space. This is the *reduced state space Kalman filter*.

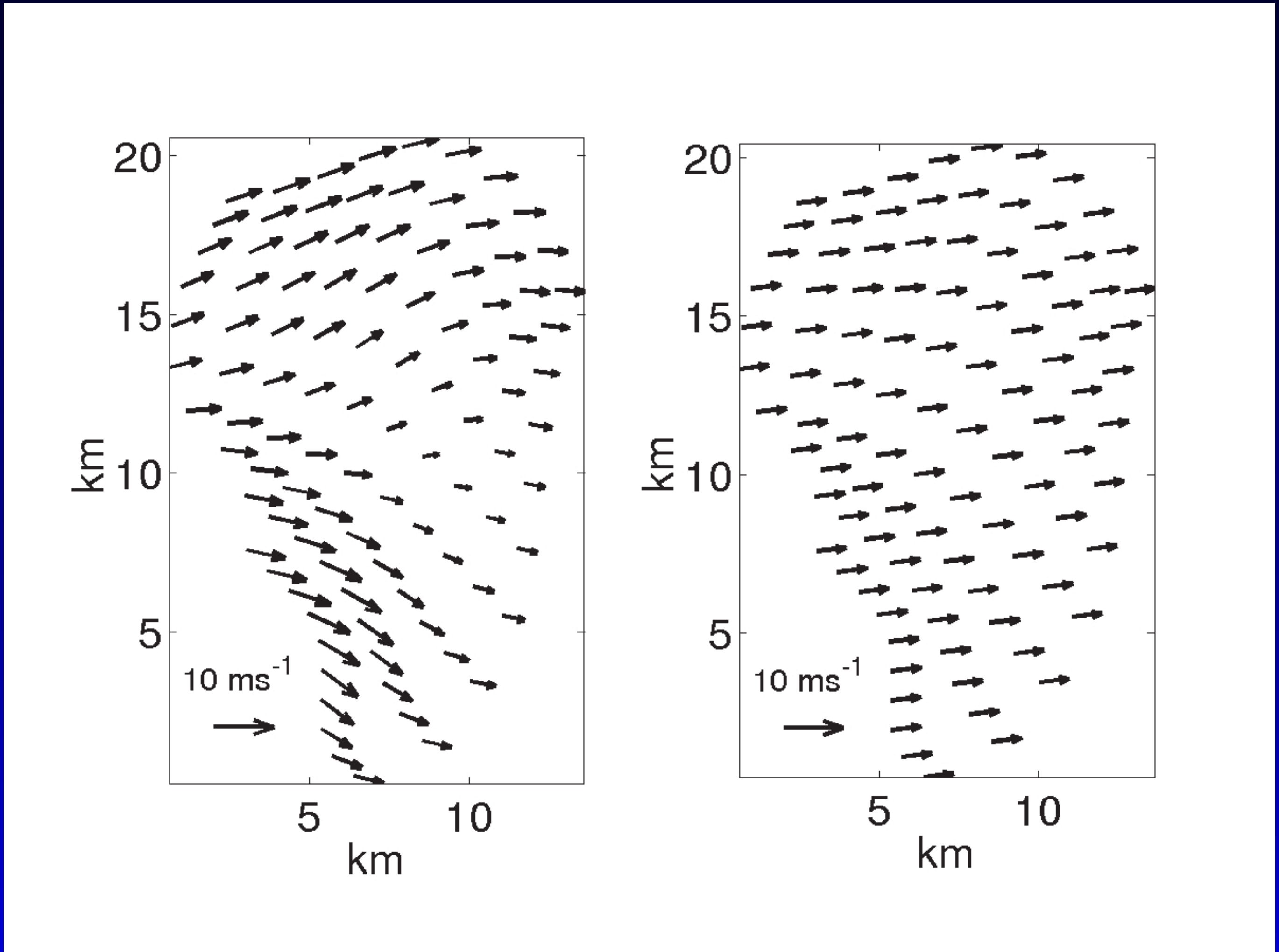
The Representer Method: An Example

A 2 layer model of Lake Kinneret



Vernieres et al., *Ocean Modelling*, 2006.

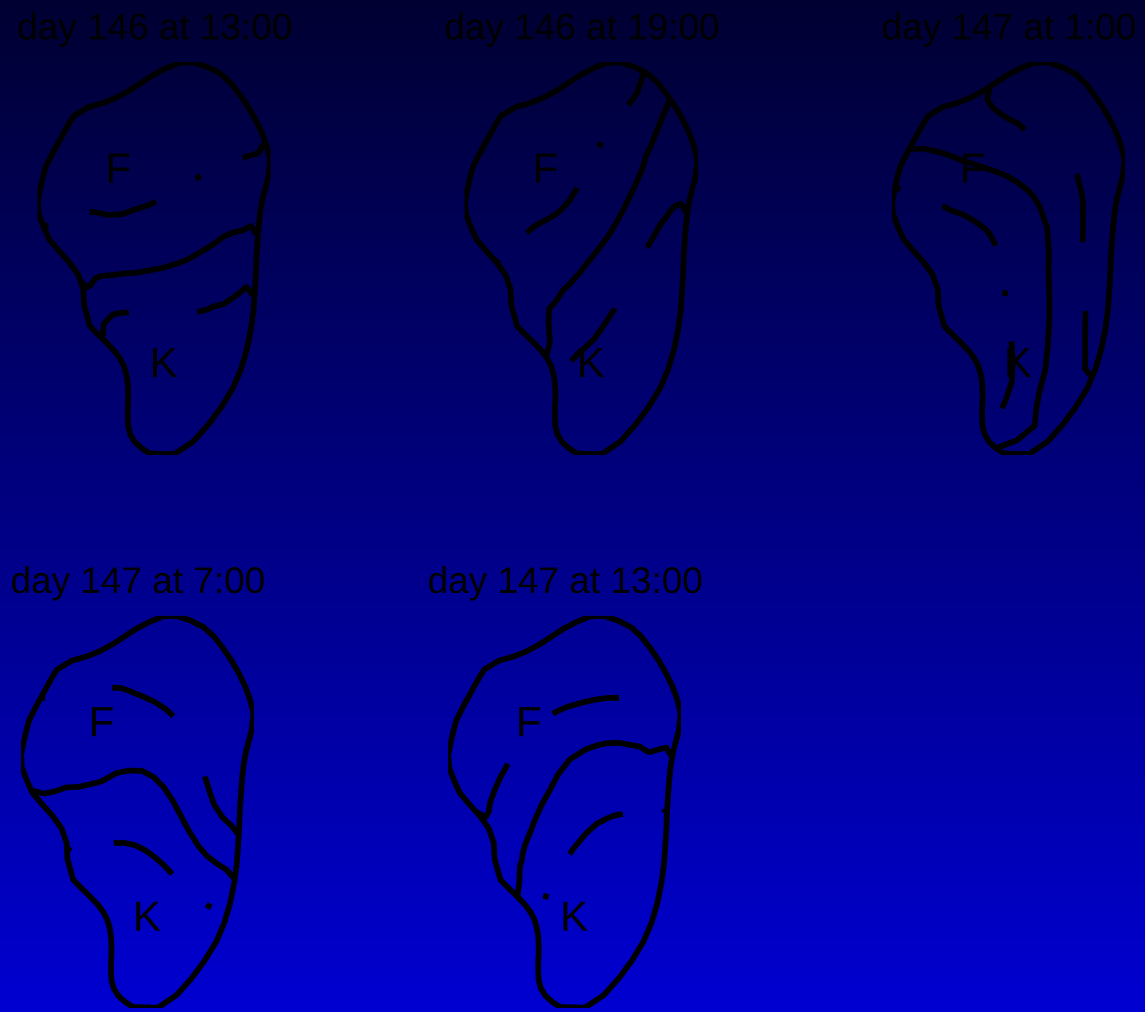
Forward and Inverse Wind Fields



Left: Inverse estimate of mean afternoon wind field.

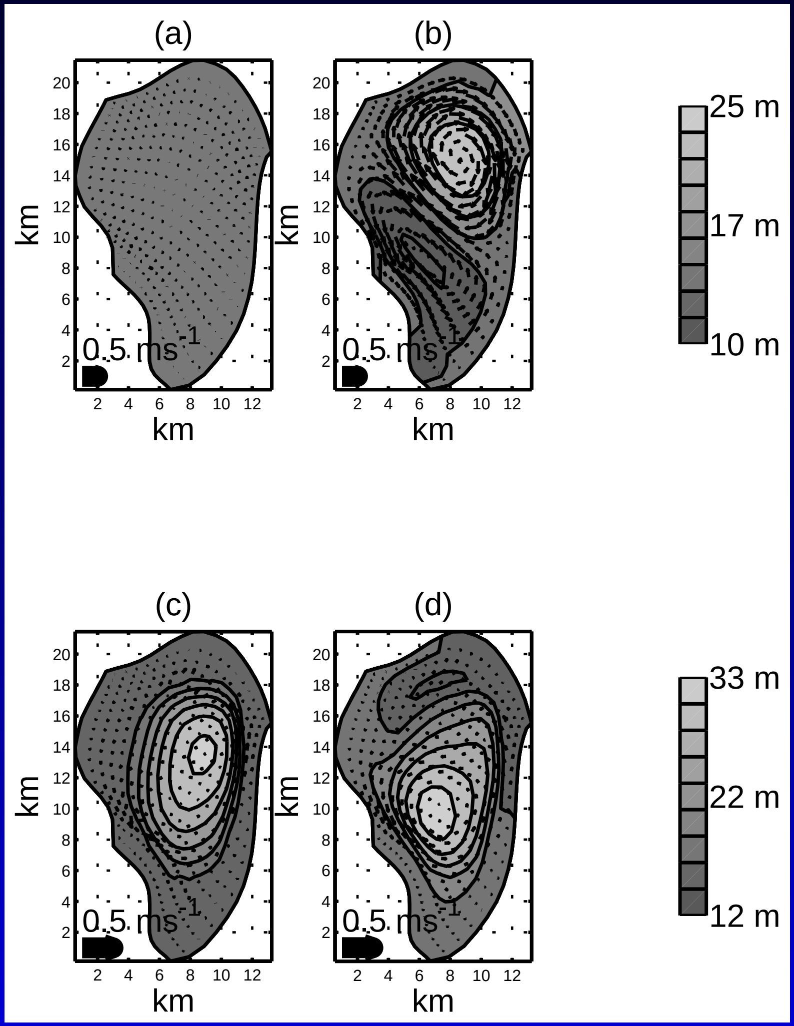
Right: Prior afternoon wind

What Does a Representer Look Like?



Upper layer thickness, representer for the first observation at station F. Solid line: -14.25m contour

Estimates of Lake Circulation



a), b) upper layer thickness, forward and inverse models. c), d) lower layer

Noise and Nonlinearity

I saw under the sun that the race is not to the swift, nor the battle to the strong, ... , but time and chance happen to them all. –Ecclesiastes 9:11

For a more recent reference, see Jazwinski, 1970 or Arnold, 1974.

The Random Walk

- Start at $x = 0$
- In every fixed interval of time Δt, move some random distance Δx_j
- The Δx_j are independent Gaussian random variables with variance v
- In time $T = N\Delta t$ the total distance moved is:

$$W = \sum_{i=1}^{N} \Delta x_i$$

$$\langle W \rangle = 0$$

The Random Walk

- In time $T = N\Delta t$ the total distance moved is:

$$W = \sum_{i=1}^{N} \Delta x_i$$

$$\langle W \rangle = 0$$

- The mean square displacement is given by

$$\langle W^2 \rangle = vN = vT/\Delta t$$

- If we choose $v = \sigma^2 \Delta t$ we get:

$$\langle W^2 \rangle = \sigma^2 T$$

The Random Walk

"W" is the *Wiener Process*.

$$\begin{aligned} E(W_t W_s) &= E((W_t - W_s + W_s)W_s) \; for \; s < t \\ &= E((W_t - W_s)W_s) + E(W_s W_s) \\ &= \sigma^2 s \end{aligned}$$

In general, $E(W_t W_s) = min(t, s)\sigma^2$.

The Random Walk

Now consider a stationary random function u:

$$E\left(\frac{u(t+h)-u(t)}{h}\frac{u(s+h)-u(s)}{h}\right)$$
$$= \frac{1}{h^2}(2E(u(t)u(s)) - E(u(t)u(s+h)) - E(u(t+h)u(s)))$$
$$= \frac{-1}{h^2}(C(t-s+h) - 2C(t-s) + C(t-s-h))$$

- $C(t-s) \equiv E(u(t)u(s))$

The Random Walk

- $-(C(t-s+h)-2C(t-s)+C(t-s-h))/h^2 \rightarrow C''(t-s)$ as $h \rightarrow 0$.
- More generally:
  $E(u'(s)u'(t)) = \frac{\partial}{\partial s}\frac{\partial}{\partial t}E(u(s)u(t))$

$$
\begin{aligned}
\frac{\partial}{\partial s}\frac{\partial}{\partial t}E(W(s)W(t)) &= \frac{\partial}{\partial s}\frac{\partial}{\partial t}\sigma^2 min(s,t) \\
&= \frac{\partial}{\partial s}\sigma^2 \begin{cases} 1 & \text{if} \quad t < s \\ 0 & \text{if} \quad t \geq s \end{cases} \\
&= \sigma^2\delta(t-s)
\end{aligned}
$$

- by the usual formal identification of the derivative of a step function with δ.

White Noise

In general the power spectral density function of a stochastic process is the Fourier transform of the covariance function:

$$f(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} C(t+\tau, t)\,d\tau$$

For “white noise,” $C(t+\tau, t) = \sigma^2 \delta(\tau)$ so $f(\omega) = \sigma^2$ = constant. So white noise, like white light, contains all frequencies at equal power.

Stochastic Differential Equations

The *Langevin equation*:

$$\dot{x} = -\alpha x + \dot{\mathbf{W}}$$

$\mathbf{W}$ is a random walk with
$E((\mathbf{W}(t+\Delta t) - \mathbf{W}(t))^2) = \sigma^2 \Delta t$

$$x_{t+\delta t} - x_t + \alpha x_t \delta t = W_{t+\delta t} - W_t$$

should lead formally to a meaningful limit:

$$dx + \alpha x dt = dW$$

Stochastic Differential Equations

but:

$$\begin{aligned}\frac{x_{t+\delta t}-x_t}{\delta t} &= -\alpha x_t + \frac{W_{t+\delta t}-W_t}{\delta t}\\ E\left[\left(\frac{x_{t+\delta t}-x_t}{\delta t}\right)^2\right] &= \alpha^2 x_t^2 + \frac{\sigma^2}{\delta t}\end{aligned}$$

so Langevin's equation:

$$dx + \alpha x dt = dW$$

does not make sense as a classical ordinary differential equation.

Stochastic Differential Equations

But we *should* be able to make sense of it:

$$
\begin{aligned}
e^{\alpha t}|\frac{dx}{dt} + \alpha x &= \frac{dW}{dt} \\
\frac{d}{dt}(e^{\alpha t} x) &= e^{\alpha t}\frac{dW}{dt} \\
x &= x(0)e^{-\alpha t} + \int_0^t e^{-a(t-s)}\frac{dW}{ds}ds
\end{aligned}
$$

The Langevin Equation

$$
\begin{aligned}
e^{\alpha t}|\frac{dx}{dt} + \alpha x &= \frac{dW}{dt} \\
\frac{d}{dt}(e^{\alpha t}x) &= e^{\alpha t}\frac{dW}{dt} \\
x &= x(0)e^{-\alpha t} + \int_0^t e^{-a(t-s)}\frac{dW}{ds}ds
\end{aligned}
$$

so:

$$
\begin{aligned}
E(x^2) = {} & (x(0)e^{-\alpha t})^2 + \\
& \int_0^t \int_0^t e^{-\alpha(t-r)} e^{-\alpha(t-s)} E(\frac{dW}{dr}\frac{dW}{ds}) dr ds
\end{aligned}
$$

The Langevin Equation, cont'd

$$
\begin{aligned}
E(x^2) &= (x(0)e^{-\alpha t})^2 + \\
&\int_0^t \int_0^t e^{-\alpha(t-r)} e^{-\alpha(t-s)} E(\frac{dW}{dr}\frac{dW}{ds}) dr ds \\
&= (x(0)e^{-\alpha t})^2 + e^{-2\alpha t} \int_0^t \sigma^2 e^{2\alpha s} ds \\
&= (x(0)e^{-\alpha t})^2 + \sigma^2 \frac{1 - e^{-2\alpha t}}{2\alpha}
\end{aligned}
$$

Stochastic Differential Equations

So instead of writing:

$$\frac{dx}{dt} + \alpha x = \frac{dW}{dt}$$

write:

$$dx_t + \alpha x_t dt = dW_t$$

as notation for

$$x_t - x_0 = -\int_{t_0}^{t} \alpha x_t dt + \int_{t_0}^{t} dW_t$$

The Langevin Equation

- Here we will get away with assuming $\int_{t_0}^{t} dW_t = W(t) - W(0) \sim N(0, \sigma^2(t - t_0))$
- Stochastic differential equations (SDE's) such as this must be modeled with special numerical techniques, e.g.,

$$x_{j+1} = (1 - \alpha \Delta t) x_j + (\Delta t)^{1/2} \sigma w_j$$

  where $w_j \sim N(0, 1)$ may be obtained from a random number generator.

Numerical Treatment of the Langevin Equation

- The variance of x after N steps is:

$$Var(x_N) = \Delta t \sigma^2 \frac{(1-\alpha\Delta t)^{2N} - 1}{(1-\alpha\Delta t)^2 - 1}$$

Compare to the solution obtained above:

$$Var(x_N) = \sigma^2 \frac{1 - e^{-2\alpha t}}{2\alpha}, \; t = N\Delta t$$

- The difference equation converges as $\Delta t \rightarrow 0$.
- The presence of the $\surd$ in the discretization indicates that special numerical techniques are required for SDE's.

Stochastic Differential Equations

- One correct way to discretize the Langevin equation is:

$$x_{j+1} = (1 - \alpha \Delta t) x_j + (\Delta t)^{1/2} \sigma w_j$$

- A common error in dealing with SDE's is the incorrect use of Δt instead of $\sqrt{\Delta t}$ as in:

$$x_{j+1} = (1 - \alpha \Delta t) x_j + \Delta t \sigma w_j$$

- In order to deal with SDEs in detail we need to re-invent calculus

Summary

- We have explored solving the linear inverse problem by the least squares method
- In variational form, the cost function gives a natural orthogonal decomposition of space and allows us to reduce the problem to manageable size.
- The representer method is one way to derive the Kalman filter.
- Fully general treatment of noisy nonlinear problems requires that we re-invent calculus

Final Thought

- Data assimilation is a highly technical subject
- When you understand the technical aspects, you are at the *beginning*, *not the end* of the subject.