# Covariance Modeling for Atmospheric and Oceanic Data Assimilation

*Grace Wahba*
`http://www.stat.wisc.edu/~wahba`
$\longrightarrow$ *TRLIST*

Statistical Approaches to the Ocean Circulation Problem, IMS Mini-Meeting, Florida State University
November 13, 2001

Thanks to Anthony Weaver

Abstract

We describe general circulation models (GCM's) and how data is assimilated into them. We describe the role of forecast error covariances and describe two families of isotropic covariances on the sphere, first, splines on the sphere, and second, diffusion covariances on the sphere. A laundry list of problems and practical issues is given.

# Outline

1. What is a general circulation model?

2. 3-D VAR, (Three Dimensional Variational Analysis), 4-D VAR (Four Dimensional Variational Analysis)

3. Horizontal velocity ($u, v$), Divergence ($D$) and Vorticity ($\zeta$); Spherical Harmonics ($Y_{\ell s}$).

4. Global Scale NWP (Numerical Weather Weather Prediction Models).

5. Model variables, $\zeta, D, T =$ Temperature, $P_s =$ surface pressure, $q =$ humidity (atmosphere); $\zeta, D, T, h =$ sea surface height, $S$= salinity (ocean). Analysis variables. Balance (?)

6. The analysis state vector.

7. The variational problem to be solved in 3D-VAR models. Weighting, smoothing and tuning parameters.

8. 4D-VAR Models.

9. Covariances on the sphere.

10. Covariance Modeling and Tuning.

## References


1. Wahba, G. (1981), 'Spline interpolation and smoothing on the sphere', *SIAM J. Sci. Stat. Comput.* **2**, 5–16.

2. Wahba, G. (1982), 'Erratum: Spline interpolation and smoothing on the sphere', *SIAM J. Sci. Stat. Comput.* **3**, 385–386.

3. Wahba, G. (1982b), Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data, *in* W. Schempp & K. Zeller, eds, 'Multivariate Approximation Theory, Vol.2', Birkhauser Verlag, pp. 407–429.

4. Weaver, A. & Courtier, P. (2001), 'Correlation modelling on the sphere using a generalized diffusion equation', *Q. J. R. Meterol. Soc* **127**, 1815–1846.

5. Gong, J., Wahba, G., Johnson, D. & Tribbia, J. (1998), 'Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing and physical parameters', *Monthly Weather Review* **125**, 210–231.

6. Wahba, G. (1999), Adaptive tuning, four dimensional variational data assimilation and representers in rkhs, *in* ECMWF, ed., 'Diagnosis of Data Assimilation Systems', European Center for Medium Range Weather Prediction, Reading England, pp. 45–52.

7. Chung, M., Taylor, J., Worsley, K., Ramsay, J., Robbins, S. & Evans, A. (2001), 'Diffusion smoothing on the cortical surface via the Laplace-Beltrami operator', manuscript at `http://www.cs.wisc.edu/~mchung/`.

## Horizontal Velocity (Wind), Divergence and Vorticity

$$P = \text{point on the sphere}$$

$$P = (\text{latitude, longitude}) = (\lambda, \phi)$$
$$\lambda \in (0, 2\pi),\ \phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$$

$$u(P) = \text{East Wind at } P$$

$$v(P) = \text{North Wind at } P$$

$$\text{vorticity} = \zeta$$

$$\text{divergence} = D$$

$$\zeta = \frac{1}{a \cos\phi}\left[-\frac{\partial}{\partial\phi}(u\cos\phi) + \frac{\partial v}{\partial\lambda}\right]$$

$$D = \frac{1}{a \cos\phi}\left[-\frac{\partial u}{\partial\lambda} + \frac{\partial}{\partial\phi}(v\cos\phi)\right]$$

## Spherical Harmonics

$$P = (\lambda, \phi)$$

$$Y_{\ell s} = \begin{cases} \theta_{\ell s} \cos(s\lambda) P_{\ell s}(\sin\phi) & 0 \le s \le \ell \\ \theta_{\ell s} \sin(s\lambda) P_{\ell |s|}(\sin\phi) & -\ell \le s < 0 \end{cases}$$

$\ell = 0, 1, 2, \ldots ,$ $\quad P_{\ell s} =$ Legendre Polynomials

The spherical harmonics are the eigenfunctions of the (horizontal) Laplacian $\Delta$ on the sphere:

$$\Delta f = \frac{1}{a^2} \left[ \frac{1}{\cos^2\phi} f_{\lambda\lambda} + \frac{1}{\cos\phi} (\cos\phi f_\phi)_\phi \right]$$

$$\Delta Y_{\ell s} = -\ell(\ell+1) Y_{\ell s}$$

and play the same role on the sphere as sines and cosines on the circle.

## Spherical Harmonics (con't)

$$\begin{aligned} f &\in \mathcal{L}_2(Sphere) \\ f &\sim \sum_{\ell=0}^{\infty} \sum_{s=-\ell}^{\ell} f_{\ell s} Y_{\ell s} \end{aligned}$$

where

$$f_{\ell s} = \int_{Sphere} f(P) Y_{\ell s}(P) dP.$$

(Given a constant) its easy to go back and forth from $(u, v)$ to $(\zeta, D)$ in spherical harmonic coordinates.

## Model and Analysis Variables

The model variables are $\zeta, D, T, p_s, q$ (for NWP) respectively, vorticity, divergence, temperature, surface pressure and humidity, as a function of $P = (lat, long)$, and a vertical coordinate which is discretized to $N$ discrete levels. The model variables are expanded in spherical harmonics $Y_{\ell s}, s = -\ell, \cdots, \ell; \ell = 0, \cdots, L$, at each level. Thus, there will be $[4N+1] \times [(L+1)/2]^2$ coefficients in the model state vector. The model state vector at time $t-1$ is updated via the equations of motion of the atmosphere to get a forecast model state vector at time $t$, The ECMWF model has $N = 31$ discrete levels in the vertical, with $L$ over $200$ which amounts to a global horizontal resolution of about $60$ km. The dimension of the state vector is around $10^7 - 10^8$ -ECMWF home page says it has three supercomputers with a combined output of $4000$Gflops.

The forecast vector $x_f(t)$.

Also known as the 'background' vector $x_b(t)$ (about which everything will be expanded).

$$\zeta(P, z_k, t) = \Sigma_{\ell,s} a_{\ell sk}(t) Y_{\ell s}(P)$$

$$D(P, z_k, t) = \Sigma_{\ell,s} b_{\ell sk}(t) Y_{\ell s}(P)$$

$$T(P, z_k, t) = \Sigma_{\ell,s} c_{\ell sk}(t) Y_{\ell s}(P)$$

$$P_s(P, t) = \Sigma_{\ell,s} d_{\ell s}(t) Y_{\ell s}(P)$$

$$q(P, z_k, t) = \Sigma_{\ell,s} e_{\ell sk}(t) Y_{\ell s}(P)$$

$$x_f \equiv x_b = \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix}$$

Next, to combine $x_b$ with observations to get an updated estimate of the present state (the analysis).

## Observations

There are many kinds of observatons, most of which are 'direct' observations on $u, v, T, h, S$, (ocean) some being indirect observations. Satellites observe sea surface height. After linearization the observation vector $y$ can be considered as linear functionals on a (mythical) state vector $x_{`true'}$, $y = Hx_{`true'} + \epsilon_O$, where $\epsilon_O$ includes instrumental error of all kinds as well as 'errors of representativeness' essentially due to the fact that a finite model cannot represent the ocean exactly.

Combining Observations and Forecast (3-D VAR)

Let $x = x_{\text{`}true\text{'}}$

$$y = Hx + \epsilon_o, \quad \epsilon_o \sim N(0, R) \quad \text{obs'n error}$$

$$x_b = x + \epsilon_b, \quad \epsilon_b \sim N(0, B) \quad \text{forecast error}$$

$$x \sim N(\mu, \Sigma) \quad \text{prior information}$$

To estimate $x$, find $x$ to minimize

$$(y - Hx)'R^{-1}(y - Hx) + (x - x_b)'B^{-1}(x - x_b) +$$

$$(x - \mu)'\Sigma^{-1}(x - \mu).$$

Let $\delta = x - x_b,\ d = y - Hx_b$, results in: find $\delta$ to minimize $J_o + J_b + J_c$ where

$$J_o = (d - H\delta)'R^{-1}(d - H\delta),$$

$$J_b = \delta' B^{-1}\delta.$$

Then $\widehat{x}$, a.k.a. the analysis vector, is given by

$$\widehat{x} = x_b + \delta.$$

R comes from instrumental errors, forward model errors, errors of representativeness.

## Where do we get $B$?

$B \sim 10^8 \times 10^8$ (?). $B$ is the forecast error covariance, which involves errors in the initial conditions of the forecast, propagated through an imperfect model. The issue is: Where do we get $B$??.
Answer, sort of: Historical data, guesses, tuning of trial covariances, some understanding of the possible nature of model errors.....

For expository purposes, consider the minimizer of

$$J_0 + J_b = (d - H\delta)'R^{-1}(d - H\delta) + \delta' B^{-1}\delta.$$

The minimizing $\delta$ is, after rearranging terms and using some matrix identites, is

$$\hat{\delta} = (B - BH'(HBH' + R)^{-1}HB)H'R^{-1}(y - Hx_b).$$

so $\hat{x} = x_b + \hat{\delta}$. Note that $\hat{\delta}$ is a rank $n_{obs} < n_{state}$ linear combination of the $n_{state}$ rows of B. [Representer theory!!]. These linear combinations of rows should have dynamical properties consistent with atmospheric flows, and they should not generate nonphysical responses when integrated forward via the equations of motion.

## 4D-VAR

We briefly describe the main idea behind 4D-VAR, without as yet saying much about $B$. 4D-VAR is presently being used in NWP at ECMWF.

The Model: Let $t = 1, \cdots, T$ denote discrete time, $x_t, t = 1, \cdots, T$ be a sequence of model state vectors representing (some part of) nature which evolves according to

$$x_{t+1} = M_t x_t + N_t + \xi_t,\ t = 1, \cdots, T-1 \qquad (1)$$

where $M_t = M_t(\theta)$ is the model (one step in time) evolution operator, possibly depending on some parameters $\theta$, $N_t$ is a forcing function and the $\xi_t$ are whats unexplained - finite, approximate model, approximate forcing, numerical approximations. $x_b$ is the forecast for $t = 1$, assumed to satisfy

$$x_b = x_1 + \epsilon_b.$$

The observations $y_t$ are assumed to satisfy

$$y_t = H_t x_t + \epsilon_t,\ t \in \Lambda$$

where $H_t$ is a map from state vector space to observation space at time $t \in \Lambda$ and $\Lambda$ is the subset of $\{1, \cdots, T\}$ where there are observations. After making numerous unjustified independence assumptions, one obtains the 4D-VAR variational problem: Find $x = (x_1, \cdots, x_T)$ to min

$$\sum_{t \in \Lambda} \|y_t - H_t x_t\|^2_{R_t} + \alpha \sum_{t=1}^{T-1} \|x_{t+1} - M_t x_t - N_t\|^2_{Q_t} + \gamma \|x_b - x_1\|^2_B + \eta \|x_T\|^2_{\Sigma},$$

where $\|\delta\|^2_B \equiv \delta' B^{-1} \delta$. The four terms represent closeness to the data, closeness to the model, closeness to the forecast, and (prior) physical information concerning physical (approximate) constraints. We see $B$, the forecast error to the first time step, and also $Q_t$ which are covariances to represent the model error. This representation treates the model error as independent from time to time, likely false. .

Gong, Wahba, Johnson and Tribbia, *Monthly Weather Review* (1998) did a toy experiment on the (atmospheric) barotropic vorticity equation, and used the GCV to 'tune' the model. The covariances $R_t, Q_t, B$ and $\Sigma$ were given simple forms and assumed known. One interesting result was that tuning the parameters $\theta$ in the differential equation (In $M_t$) could not be optimized independently of the other tuning parameters. This has implications for the organizational structure of an NWP center The GCV is an (internal) predictive mean square error criteria, and may or may not be appropriate for an NWP forecast, but it may be useful in ocean modeling where nowcasting, for input into an NWP is the end usage. Since $R_t$ was assumed known, the observations are rescaled so that the observational errors are 'white' in the formula for GCV below. In practice this is non-trivial.

Letting $\widehat{y_t} = H_t \widehat{x_t}$, $\widehat{y} = (\widehat{y_1}', \cdots, \widehat{y_T}')'$, then, (more assumptions)) there exists a matrix $A(\lambda)$, $(\lambda = \alpha, \gamma, \eta, \theta, ...)$ known as the influence matrix, such that

$$\widehat{y} = A(\lambda)y + quantities\ independent\ of\ y.$$

The $GCV$ (generalized cross validation) estimate of $\lambda$ is the minimizer of $V(\lambda)$ where

$$V(\lambda) = \frac{\frac{1}{n_{dat}} RSS(\lambda)}{[\frac{1}{n_{dat}} tr(I - A(\lambda))]^2}$$

where $n_{dat}$ is the number of data points (dimension of $y$) and $RSS(\lambda) = \|y - \widehat{y}\|^2$. Letting $x_{t`true'}$, be the 'true' but unknown $x_t$, $V(\lambda)$ is, under suitable assumptions, a proxy for the predictive mean square error ($PMSE$), given by $R(\lambda)$ where

$$R(\lambda) = \frac{1}{n_{dat}} \sum_{t \in \Lambda} \|H_t x_{t`true'}, - H_t \widehat{x}_t(\lambda)\|^2,$$

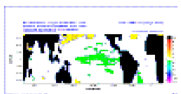
in the sense that the minimizer $V(\lambda)$ is a good estimates of the minimizer of $R(\lambda)$.

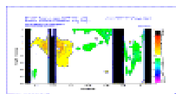
Predictive Mean Square Error may be an appropriate criteria for nowcasting. $trA$ can be estimated via the randomized trace technique, so that it can be feasible even in very large problems. For forecasting it may or may not be necessary to modify the criteria. Maximum likelihood estimates in this context have been proposed and studied by Dee, daSilva and collaborators (MWR 1999). Operational NWP models have various empirical approachs ('NMC' method, Canadian system).

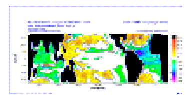
# ECMWF Seasonal Forecasting

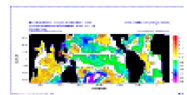
Seasonal
About Seasonal
Charts
**Ocean Analysis**
Archive

Latest Analysis Images:


SST Anomaly (Tropics)


Equatorial (X-Z) Temperature Anomaly


Sea-level Anomaly (Tropics)


Zonal Wind-stress Anomaly (Tropics)

---

Much of the variability in the ocean is in response to changes in fluxes of momemtum, heat and fresh water. In producing an ocean analysis the ocean model is forced with daily average values of these atmospheric fluxes obtained from the ECMWF operational analysis forecast system. In addition , all available ocean thermal data are used in an ocean analysis which is performed weekly (ending one week behind real time to allow the inclusion of subsurface thermal data). This analysis provides the ocean initial conditions for the coupled model forecast.

An archive of previous Seasonal Ocean Analysis Images is available.

## Positive Definite Functions on the Sphere

Letting $P, P'$ be two points on the sphere. $B(P, P')$ given by

$$B(P, P') = \Sigma_{\ell s}\Sigma_{\ell' s'} b_{\ell s, \ell' s'} Y_{\ell s}(P) Y_{\ell' s'}(P')$$

will be positive definite if the matrix $\{b_{\ell s, \ell' s'}\}$ is positive definite. If this matrix is diagonal, and the entries only depend on $\ell$, $b_{\ell s, \ell s} = \lambda_\ell$ then we have the famous addition formula for spherical harmonics:

$$\sum_{\ell s} \lambda_\ell Y_{\ell s}(P) Y_{\ell' s'}(P') = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + 1) \lambda_\ell P_\ell(\gamma(P, P'))$$

where $P_\ell$ is the $\ell$th Legendre polynomial, and $\gamma(P, P')$ is the cosine of the angle between $P$ and $P'$.

Positive Definite Functions on the Sphere (continued)

According to a very old theorem of Schoenberg, an isotropic covariance on the sphere is always of the form

$$B(P, P') = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} \lambda_\ell P_\ell(\gamma(P, P'))$$

with non-negative $\lambda_\ell$. We describe two different families of isotropic covariances on the sphere that have been used in ocean applications. The first is based on a model of a stochastic process which is the solution of the $m$th iterated Laplacian driven by formal white noise, and the second is based on a diffusion model. They involve different rates of decay of the energy content of the processes. In 3- and 4D-VAR one wants to 'tune' the most important parameters. After a relative scale factor between $R$ and $B$ probably one of the most important parameters in $B$ is the implied rate of decay of energy with wavenumber. [Oversimplification of course].

## First Family: Splines on the Sphere

Recall that $\Delta Y_{\ell s} = -\ell(\ell+1)Y_{\ell s}$.
Consider the zero-mean Gaussian stochastic process

$$X(P) = \sum_{\ell s} X_{\ell s} Y_{\ell s}(P)$$

with $\{X_{\ell s}\}$ independent, $\mathcal{N}(0, \frac{1}{[\ell(\ell+1)]^m})$. Then

$$\Delta^{m/2} X(P) = dW(P), \ \ [formal\ white\ noise].$$

Then

$$\begin{aligned} EX(P)X(P') &= \textstyle\sum_{\ell s} \frac{1}{[\ell(\ell+1)]^m} Y_{\ell s}(P) Y_{\ell' s'}(P') \\ &\equiv K_m(P, P'), \ \ say. \end{aligned}$$

Closed form expressions for a good approximation to $K_m, m = 3/2, 2, 5/2, \cdots, 6$ appear in Wahba 1981, 1982.

Splines on the Sphere, continued.

Figure 1: Sample Fourier coefficients.

Figure 2: $\lambda_\ell = \frac{1}{[\sum_{j=0}^{2} a_j [\ell(\ell+1)]^j]^2}$.

$$\sum_{j=0}^{2} \alpha_j \Delta^j X(P) = dW(P).$$

Figure 3: Correlation function corresponding to the covariance for Figure 2.

Figure 4: A sample correlation function from another data set.

Alternatively, if $\lambda_\ell = \sum_j \frac{b_j}{\ell(\ell+1)^{2j}}$, then the corresponding covariance is $\sum_j b_j K_j(P, P')$, and the closed form expressions in Wahba(1982) could be used.

Figure 3 from Wahba, G. (1982b), Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data, *in* W. Schempp & K. Zeller, eds, 'Multivariate Approximation Theory, Vol.2', Birkhauser Verlag, pp. 407–429 goes here. At `ftp://ftp.stat.wisc.edu/pub/wahba/oldie/vecss.pdf`

Second Family: Diffusion Models on the Sphere.

Consider the diffusion equation

$$\frac{\partial f}{\partial t} - \kappa \Delta f(P) = 0$$

Letting $f(P,t) = \sum_{\ell s} f_{\ell s}(t) Y_{\ell s}(P)$, $f$ will satisfy the diffusion equation if

$$\frac{df_{\ell s}}{dt} = -\kappa \ell(\ell+1) f_{\ell s}(t),$$

so that $f(P,0)$ "diffuses" in time $T$ to

$$f(P,T) = \sum_{\ell s} f_{\ell s}(0) e^{-\kappa \ell(\ell+1)T} Y_{\ell s}(P).$$

Courtier and Weaver, QJRM(2001) used this argument to propose the isotropic covariance model with $\lambda_\ell = e^{-\kappa \ell(\ell+1)}$, or, more generally they proposed considering the more general p.d.e.

$$\frac{\partial f}{\partial t} + \sum_j \kappa_j (-\Delta)^j f(P) = 0,$$

which leads to $\lambda_\ell = e^{-\sum_j \kappa_j [\ell(\ell+1)]^j}$.

Diffusion Models on the Sphere (continued).

Figure (a): (Weaver and Courtier) gives several different plots of $\lambda_\ell$. The heavy line corresponds to a $j = 1$ model and the dotted lines correspond to particular $j = 2$ and $j = 3$ models, all scaled to have the same length scale.
Figure (b): Coresponding correlation functions.

Figure 1. from Weaver, A. & Courtier, P. (2001), ‘Correlation modelling on the sphere using a generalized diffusion equation’, *Q. J. R. Meterol. Soc* **127**, 1815–1846. goes here.

Figure 4 of Chung, M., Taylor, J., Worsley, K., Ramsay, J., Robbins, S. & Evans, A. (2001), 'Diffusion smoothing on the cortical surface via the Laplace-Beltrami operator' goes here. At http://www.cs.wisc.edu/~mchung/.

Let (Wahba, 1998, ECMWF Proceedings)

$$f(P,t) = f_0(P,t) + f_1(P,t), \tag{2}$$

where

$$\mathcal{L} f_0(P,t) = 0, \quad \mathcal{B} f_1(P,t) = 0,$$

where $\mathcal{L}$ is a (linear) evolution operator, (an example is $\frac{\partial}{\partial t} - \kappa\Delta$) and $\mathcal{B}$ are initial/boundary conditions which serve to make the solution of the differential equation $\mathcal{L}f = u$ unique, so that

$$f(P,t) = f_0(P,t) + \int G(P,t;P',t')u(P',t')dP'dt'$$

where $G$ is the Green's function for $\mathcal{L}$ and $\mathcal{B}$. (Thus, $\mathcal{L}f_1 = u$.) If $u$ is treated as though it is a zero mean Gaussian stochastic process with covariance $R_u(t,P;t',P')$ then

$$\begin{aligned} Ef_1(P,t)f_1(P',t') &= \int\int G(P,t;Q,s)G(P',t';Q',s') \cdot \\ &\quad R_u(Q,s;Q',s')dsdQds'dQ' \\ &= R_1(P,t;P',t'), \ \ [say]. \end{aligned}$$

Let

$$Ef_0(P,t)f_0(P',t') = R_0(P,t;P',t').$$

(This prior could, for example, be generated by a prior on the $f_{\ell s}(0)$).) Suppose that $f_0$ and $f_1$ are independent, then

$$Ef(P,t)f(P',t') = R_0(P,t;P',t') + R_1(P,t;P',t').$$

It may be possible to get reasonable covariances by actually taking simplified versions of the evolution equations of the model.

## Many Questions

We haven't mentioned that the atmosphere and ocean are three dimensional in space, not two. Not mentioned is the correlation between different variables.

Not mentioned are systematic errors from time to time, and certainly correlated model errors from time to time which need to be understood in the context of 4D-VAR.

There is a general buzz in the NWP community that isotropic covariances are oversimplified, and that the covariance should somehow depend on the atmospheric flow, horizontally as well as vertically. This issue is also discussed briefly in Weaver and Courtier. Its easy to vary the spatial covariance by, say premultiplying $B(z, z')$ where $z, z'$ are on the sphere, the ocean, the ocean $\times$ time, or (?) by $w(z)w(z')$ for some positive $w$. Similarly, general classes of nonisotropic, nonstationary covariances $B(z, z')$ may in theory be generated from simpler covariances B* given a map $z-> T(z)$, as $B(z, z') = B * (T(z), T(z'))$.

## Issues

- What is your criteria for choosing a good $B$? Now-casting or forecasting? What metric? What model?

- Any choice must be computable within the resources available.

- How rapidly should your covariances vary in time and space? Time: seasonally, monthly, daily, hourly? Space: How much spatial variability?

- How do you establish a parametrization of $B$ in which you only vary the sensitive parameters, and, how do you estimate them? How dynamic should your estimation procedure be? - How dependent on the present state?

- How do you do all of the above in the context of a model that has a lot of other tunable parameters, which surely affect the forecast error?

# Closing Remarks

NWP, Climate, Ocean and coupled Ocean-Atmosphere models solve gigantic penalized least squares problems. However there are many features that do not satisfy the usual statistical assumptions. Observation errors and model errors are correlated in complicated ways that are not easy to unscramble, and systematic errors abound in observing system calibration and drift, and model inadequacies. Model and observation operators have nonlinearities. Data sets are humongous but some are of notoriously poor quality. A large amount of prior information is available but some of it is highly ‘physical’ or qualitative rather than statistical, for example, penalties based on the understanding that certain physical quantities are ‘not too large’. Operational models are sensitive to numerous tuning parameters, whose effects interact in difficult to understand ways, and the ‘optimal’ covariances certainly depend on the resolution and quality of the model, and the quantity and quality of the data available.

The GCV and related methods may be useful in in some of the tuning problems associated with ocean GCM’s but its important to examine covariances and their tuning in the context of a particular model.