Lecture 17: Objective Mapping: Refining Methods

Recap

In Lecture 16, we looked closely at objective mapping, formulated the estimation procedure, and laid out a few examples for turning irregularly spaced quantities into regularly gridded quantities. We worked from the equations:

$$\hat{\mathbf{y}} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{d}. \tag{1}$$

where $\hat{\mathbf{y}}$ is our vector of mapped quantities, $\mathbf{Q}$ is our data–model covariance, $\mathbf{R}$ is the data–data covariance, and $\mathbf{d}$ is the vector of measured data. The measure of the quality of the fit is the fractional mean squared error:

$$\frac{\langle(\hat{\mathbf{y}} - \mathbf{y})^2\rangle}{\langle\mathbf{y}^2\rangle} = 1 - \frac{\mathbf{Q}^T\mathbf{R}\mathbf{Q}}{\mathbf{P}} \tag{2}$$

where $\mathbf{P}$ is the covariance of the mapped values, and is usually used as a diagonal matrix.

Objective mapping

First here's a review some of the formalism of objective mapping. An **objective map** is the minimum mean-square error estimate of a continuous function of a variable, given discrete data. The interpolation that we discussed earlier is an example of a simple objective map. Objective mapping is used widely in oceanic and atmospheric sciences, as both fields have the need to make continuous maps from discrete data, and the variables to be mapped change unpredictably in some ways from one realization to the next, so that the variables may be considered random and a statistical approach is appropriate. The standard reference for oceanographic objective analysis is Bretherton et al. (1976, DSR), and an excellent textbook focusing on atmospheric applications is Daley (1991).

Consider the two-dimensional map of a discretely measured scalar. We've already discussed all of the tools for making a simple objective map. In the following the symbols for some variables are changed, as we want to reserve the variables x and y for horizontal position. A datum u_n is supposed to be made up of signal and noise

$$u_n = \tilde{u}(x_n, y_n) + \epsilon_n. \tag{3}$$

The signal $\tilde{u}_n$ is what we seek. It could be a filtered field representing, for example, larger scales. The noise ϵ is everything in the datum other than the signal, which might include instrumental error and smaller scale variability. This separation into a signal and noise is made explicit by the statistics. It is typical to assume that the noise is uncorrelated with the signal, and the noise is uncorrelated from one datum to the next

$$\langle\tilde{u}\epsilon\rangle = 0 \tag{4}$$

$$\langle\epsilon_n\epsilon_m\rangle = E\delta_{nm}. \tag{5}$$

The assumption (4) is essentially that, by appropriate averaging, a scale separation can be achieved. While such an assumption is not required for an objective map, it is nearly always used.

Since we have already done a linear estimate with multiple variables, we know the solution

$$\tilde{u}(x, y) = \mathbf{a}(x, y)^T\mathbf{u} \tag{6}$$

Here $\hat{u}$ is the objective map, a continuous function of x and y; $\mathbf{u}$ is the data vector; and $\mathbf{a}$ is the gain vector, whose components are continuous functions of x and y. The minimum MSE estimate of $\mathbf{a}$ is

$$\mathbf{a} = \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u}(x,y) \rangle \tag{7}$$

where $\langle \mathbf{u}\mathbf{u}^T \rangle$ is the data–data covariance matrix; and $\langle \mathbf{u}\tilde{u} \rangle$ is the covariance between the data and the signal, a vector whose components are continuous functions of x and y. The normalized MSE is

$$\frac{\langle (\hat{u} - \tilde{u})^2 \rangle}{\langle \tilde{u}^2 \rangle} = 1 - \frac{\langle \tilde{u}\mathbf{u}^T \rangle \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u} \rangle}{\langle \tilde{u}^2 \rangle}, \tag{8}$$

which is used to evaluate the quality of the map. In typical use, the map is only plotted where the normalized MSE is smaller than a given value. Perhaps the most important feature of an objective map is this estimate of error. The MSE depends only on data locations, not on the particular value of the data, so it is of value for experiment design. Given assumptions (3-5), the data covariance matrix is

$$\langle \mathbf{u}\mathbf{u}^T \rangle = \langle \tilde{\mathbf{u}}\tilde{\mathbf{u}}^T \rangle + E\mathbf{I}, \tag{9}$$

where $\langle \tilde{\mathbf{u}}\tilde{\mathbf{u}}^T \rangle$ is the covariance matrix of the signal evaluated at the data locations. So, given the signal autocovariance and the noise variance, we are good to go.

In practice, there are a few approaches to obtaining the needed statistics. If many realizations of the desired signal are available, as in the case of weather prediction, the statistics may be calculated. The atmospheric literature has many examples of covariances calculated for all sorts of variables. The ocean is often more data-poor, so the statistics are more difficult to calculate. Given an oceanographic survey, one can assume homogeneity (that is, statistics don't vary with position) and calculate the autocovariance by averaging products of data pairs into bins with similar separations. The result is then fit to a continuous function. The assumption of isotropy is often used so that the direction of the separation between data pairs does not matter. Finally, objective mapping is sometimes employed just to look at the consequences of the assumption of certain statistics.

Traditionally, oceanographers recommended that the statistics not be computed from the observations, because that would bias the statistics. But more recent work led by statisticians has argued that there are sufficient observations to justify deriving the covariance function directly from the data.

In class, we looked at results from Kuusela and Stein (2018), who used a moving window approach to compute covariance functions that they then used to objectively map Argo data.

Defining covariance in multiple dimensions

So far, we have the covariance as a function ρ dependent only on separation between observations. In two or more dimensions, this dependence on separation only makes sense for isotropic systems. So how do we define ρ to deal with multiple dimensions?

Isotropic covariance. An easy scenario is to map in two dimensional space mapping point measurements at the ocean surface onto a regular grid. An example of this might be high-frequency radar data, which are measured continuously from multiple radars along the California and Oregon coast. Since the measurements are continuous, we can take snapshot measurements, all from the same time, to produce mapped, gridded fields. That requires a covariance function in space. We can write:

$$\langle s(x_1, y_1)s(x_2, y_2)\rangle = \rho(r) = A\exp\left[\frac{-r^2}{L^2}\right], \tag{10}$$

where $r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ is the spatial separation, and L is a length scale. This assumes an isotropic decorrelation—that distance is measured in equivalent units (e.g. km) in both x and y, and that the lengthscales are the same.

The corresponding covariance matrix is

$$\mathbf{R} = \begin{bmatrix} \rho(0,0) + \sigma^2 & \rho(i_1 - i_2, j_1 - j_2) & \rho(i_1 - i_3, j_1 - j_3) & \dots \\ \rho(i_2 - i_1, j_2 - j_1) & \rho(0,0) + \sigma^2 & \rho(i_2 - i_3, j_2 - j_3) & \dots \\ \rho(i_3 - i_1, j_3 - j_1) & \rho(i_3 - i_2, j_3 - j_2) & \rho(0,0) + \sigma^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{11}$$

Anisotropic covariance. A greater level of complexity is introduced if we assume that the covariance is anisotropic. This could occur because features are elongated in the zonal direction so that L_x is longer than L_y, or we consider depth as well as horizontal distance, or we consider time as well as space, so we need to account for variables with entirely different units. In this case, we can write a more refined covariance function:

$$\langle s(x_1, y_1, z_1, t_1)s(x_2, y_2, z_2, t_2)\rangle = \rho(r) = A\exp\left(-\left[\frac{\Delta x^2}{L_x^2} + \frac{\Delta y^2}{L_y^2} + \frac{\Delta z^2}{L_z^2} + \frac{\tau^2}{T^2}\right]\right), \tag{12}$$

and we could even use different functional forms for different variables.

So for the two dimensional problem that we have here, we could use:

$$\rho(\Delta x, \Delta y) = A\exp\left[-\frac{\Delta x^2}{L_x^2} - \frac{\Delta y^2}{L_y^2}\right]. \tag{13}$$

Additional dependencies. In some cases, covariance functions can depend on more quantities as well. For example, we might want to map relative to bathymetry, so define the covariance to depend on bottom depth H, or to follow contours of f/H, or to build in additional constraints to for example allow no covariance between closely spaced points that are separated by a land mass.

Linear operators

We don't always measure and map the same quantity. Let's consider a situation in which our mapped quantity is not equivalent to x, but instead depends on some linear operator. We initially considered the estimate of a continuous function of time $\hat{y}(t)$ given discrete data $\mathbf{x}$:

$$\hat{y}(t) = \mathbf{x}^T\langle\mathbf{x}\mathbf{x}^T\rangle^{-1}\langle\mathbf{x}y(t)\rangle. \tag{14}$$

Now apply a linear operator L to the estimate to get the result

$$L\left[\hat{y}(t)\right] = \mathbf{x}^T\langle\mathbf{x}\mathbf{x}^T\rangle^{-1}\langle\mathbf{x}L\left[y(t)\right]\rangle. \tag{15}$$

The linear operator works only on the part of the estimate that is a continuous function. Thus the result of the linear operation on the estimate is the same as the estimate of the linear operation.

For a specific example, consider the estimate of the time derivative discrete time data. In this case, the estimate would be

$$\frac{d\hat{x}(t)}{dt} = \mathbf{x}^T \langle \mathbf{x}\mathbf{x}^T \rangle^{-1} \left\langle \mathbf{x} \frac{dx(t)}{dt} \right\rangle, \tag{16}$$

and the skill would be

$$\frac{\left[\frac{d\hat{x}(t)}{dt}\right]^2}{\left[\frac{dx(t)}{dt}\right]^2} = \frac{\left\langle \frac{dx(t)}{dt}\mathbf{x} \right\rangle^T \langle \mathbf{x}\mathbf{x}^T \rangle^{-1} \left\langle \mathbf{x}\frac{dx(t)}{dt} \right\rangle}{\left[\frac{dx(t)}{dt}\right]^2}. \tag{17}$$

To be even more specific, suppose the autocovariance to be Gaussian as in (10), so the data covariance matrix is given by (11). To calculate the estimate of the derivative and its skill (1–2), we need two more statistics:

$$\text{data–model covariance:} \quad \left\langle \mathbf{x}\frac{dx(t)}{dt} \right\rangle \tag{18}$$

$$\text{model–model covariance:} \quad \left\langle \left[\frac{dx(t)}{dt}\right]^2 \right\rangle \tag{19}$$

To calculate the covariance of the data with the time derivative, we need the time derivative of the autocovariance (10):

$$\left\langle x(t_i)\frac{dx(t)}{dt} \right\rangle = \frac{d}{dt}\langle x(t_i)x(t) \rangle = -\frac{2A(t-t_i)}{T^2} \exp\left[-\frac{(t-t_i)^2}{T^2}\right]. \tag{20}$$

where the subscript i refers to the data. The variance of the time derivative is

$$\left\langle \left[\frac{dx(t)}{dt}\right]^2 \right\rangle = \left[\frac{\partial^2}{\partial t_1 \partial t_2}\langle x(t_1)x(t_2)\rangle\right] = \frac{2A}{T^2}\exp\left(-\frac{(t_2-t_1)^2}{T^2}\right) - \frac{4A(t_1-t_2)^2}{T^4}\exp\left(-\frac{(t_2-t_1)^2}{T^2}\right). \tag{21}$$

At $t_1 = t_2$ (e.g. along the diagonal), this becomes

$$\left\langle \left[\frac{dx(t)}{dt}\right]^2 \right\rangle_{t_1=t_2} = \frac{2A}{T^2}. \tag{22}$$

This type of framework allows us to map derivatives in a single step (e.g. to map geostrophic velocities from measured sea surface heights, without having to map and then discretize.

The sample code "intgauss.m" provides examples for mapping time derivatives of x as well as x itself. In the case of perfect data, the first difference estimate of time derivative is approached as the time scale T grows. The skill also improves with increasing time scale.

Noise requires an additional term on the diagonal of the data covariance matrix, but all other required covariances remain the same. The estimate of derivative half-way between the data are generally smaller than for perfect data. The skill is strikingly different for noisy data. A maximum in skill is reached for $T = 2\delta$, the separation of the data. There is an apparent advantage to having data spacing that matches the intrinsic time scale of the observed variable. The skill then decreases for increasing T. In the presence of noise, closely separated data (where close is defined relative to T) are not useful for calculating derivatives. That is, any difference between the data is more likely to be caused by noise than a real change in the observed variable.

Mapping dynamic topography from velocity

To map dynamic topography from velocity, we invert this process: measure velocity (u and v), and infer streamfunction ψ or dynamic topography η on a regular grid. In this case, the data vector is

$$\mathbf{d} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \\ v_1 \\ v_2 \\ \vdots v_N \end{bmatrix} \tag{23}$$

and the data–data covariance matrix will be a $2N \times 2N$ matrix with four quadrants:

$$\mathbf{R} = \begin{bmatrix} \langle \mathbf{u}\mathbf{u}^T \rangle & \langle \mathbf{u}\mathbf{v}^T \rangle \\ \langle \mathbf{v}\mathbf{u}^T \rangle & \langle \mathbf{v}\mathbf{v}^T \rangle, \end{bmatrix} \tag{24}$$

with error added on the diagonal. The data–model covariance is $2N \times M$:

$$\mathbf{Q} = \begin{bmatrix} \langle \mathbf{u}\psi \rangle \\ \langle \mathbf{v}\psi \rangle. \end{bmatrix} \tag{25}$$

And the model–model covariance is defined by $\langle \psi^2 \rangle$. The trick in this is that we need the relationships between our covariance matrices to reflect the dependencies between ψ, u, and v. If the flow is geostrophic, then $u = -\partial\psi/\partial y$ and $v = \partial\psi/\partial x$. See Bretherton et al. (1976), or Gille (2003) for a derivation of the equations.

If the flow has a divergent component as well as a rotational component, the mapped quantities can be made more complicated by assuming $u = -\partial\psi/\partial y + \partial\chi/partialx$ and $v = \partial\psi/\partial x + \partial\chi/\partial y$. See Daley (1976) or Dever (2004) for a derivation of the equations.

Nonzero means

As we discussed with regard to linear estimation, we are assuming so far that the mapped variables have zero mean. Especially in the ocean, we often do not know the mean as we only have a few realizations of a survey. An often used procedure is to estimate the mean using a fit to a low-order polynomial, remove this from the data and proceed with the objective map. The mean is added back in after the map of the fluctuations is done.

If we assume the mean to be a constant, then there is a procedure to assure zero bias in the map

$$\langle \hat{u} \rangle = \langle \tilde{u} \rangle \tag{26}$$

Substituting (6) into (26)

$$\mathbf{a}^T \langle \mathbf{u} \rangle = \langle \tilde{u} \rangle. \tag{27}$$

The assumption that the mean is a constant implies that all the data have the same mean $\langle\tilde{u}\rangle$, so (26) becomes

$$\mathbf{a}^T\mathbf{v}\langle\tilde{u}\rangle = \langle\tilde{u}\rangle \tag{28}$$

where $\mathbf{v}$ is a vector of ones. Dividing by the constant mean $\langle\tilde{u}\rangle$, we arrive at a constraint

$$\mathbf{a}^T\mathbf{v} = 1. \tag{29}$$

Our optimization problem is to minimize the MSE $\langle(\hat{u}-\tilde{u})^2\rangle$ subject to (29). Use the method of Lagrange multipliers to set up a cost function

$$\mathcal{L} = \left\langle \left(\mathbf{a}^T\mathbf{u}-\tilde{u}\right)^2 \right\rangle - 2\lambda(\mathbf{a}^T\mathbf{v}-1) \tag{30}$$

where λ is a Lagrange multiplier. Differentiating (30) by $\mathbf{a}$, setting the result to zero, and solving yields

$$\mathbf{a} = \langle\mathbf{u}\mathbf{u}^T\rangle^{-1}(\langle\mathbf{u}\tilde{u} + \lambda\mathbf{v}) \tag{31}$$

Plugging (31) into the constraint (29) and solving for λ:

$$\lambda = \frac{1-\mathbf{v}^T\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\langle\mathbf{u}\tilde{u}\rangle}{\mathbf{v}^T\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\mathbf{v}} \tag{32}$$

The zero bias gain thus has an additional term compared to (7)

$$\mathbf{a} = \langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\left(\langle\mathbf{u}\tilde{u}\rangle + \mathbf{v}\frac{1-\mathbf{v}^T\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\langle\mathbf{u}\tilde{u}\rangle}{\mathbf{v}^T\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\mathbf{v}}\right). \tag{33}$$

The normalized MSE is

$$\langle(\hat{u}-\tilde{u})^2\rangle = 1 - \frac{\langle\tilde{u}\mathbf{u}^T\rangle\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\langle\mathbf{u}\tilde{u}\rangle}{\langle\tilde{u}^2\rangle} + \frac{\left(1-\mathbf{v}^T\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\langle\mathbf{u}\tilde{u}\rangle\right)^2}{\langle\tilde{u}^2\rangle\mathbf{v}^T\langle\mathbf{u}\mathbf{u}^T\rangle^{-1}\mathbf{v}} \tag{34}$$

The increase in MSE represented in the last term is due to the zero bias constraint.

In the appendix, I've added additional variations to consider objective function fit to find, for example, a large-scale field using least-squares fitting, while mapping the fluctuating part of the field.

Appendix: Objective function fit

Davis (1985) presented a way of fitting a set of functions that would not require knowledge of the mean. The estimator is of the form (6) but the signal $\tilde{u}$ is assumed to be a linear combination of some functions

$$\tilde{u} = \mathbf{b}^T\mathbf{f}(x,y) \tag{35}$$

where $\mathbf{b}$ is a vector of random coefficients, and $\mathbf{f}$ is a vector of continuous functions. The estimate is constrained to have zero bias, which in this case requires

$$\mathbf{a}^T\mathbf{F}\langle\mathbf{b}\rangle = \mathbf{f}^T\langle\mathbf{b}\rangle \tag{36}$$

where $\mathbf{F}$ is the matrix of the continuous functions evaluated at the data locations. Application of this constraint requires knowledge of $\langle \mathbf{b} \rangle$. A stronger constraint that ensures that (36) is satisfied, but does not require knowledge of the mean $\langle \mathbf{b} \rangle$ is

$$\mathbf{F}^T \mathbf{a} = \mathbf{f} \tag{37}$$

Here we go again, minimizing the MSE with the constraint (37) using Lagrange multipliers. Using (6) and (35), the MSE is

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \left\langle (\mathbf{a}^T \mathbf{u} - \mathbf{b}^T \mathbf{f})^{\mathbf{2}} \right\rangle \tag{38}$$

Suppose that the signal is uncorrelated with the noise as in (4), but the noise covariance matrix N may not be diagonal as in (5). In this case, the MSE is

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{a}^T \left(\mathbf{F} \langle \mathbf{b}\mathbf{b}^T \rangle \mathbf{F}^T + \mathbf{N} \right)) \mathbf{a} - 2\mathbf{a}^T \mathbf{F} \langle \mathbf{b}\mathbf{b}^T \rangle \mathbf{f} + \mathbf{f}^T \langle \mathbf{b}\mathbf{b}^T \rangle \mathbf{f} \tag{39}$$

Using the zero bias constraint (37), this can be simplified to

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{a}^T \mathbf{N} \mathbf{a} \tag{40}$$

So the cost function to be minimized is

$$\mathcal{L} = \mathbf{a}^T \mathbf{N} \mathbf{a} - 2\boldsymbol{\lambda}^T (\mathbf{F}^T \mathbf{a} - \mathbf{f}) \tag{41}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers. Differentiating with respect to $\mathbf{a}$, setting the result to zero, and solving yields

$$\mathbf{a} = \mathbf{N}^{-1} \mathbf{F} \boldsymbol{\lambda} \tag{42}$$

Substitute (42) back into the constraint (37) to get

$$\boldsymbol{\lambda} = (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f} \tag{43}$$

So the gain is

$$\mathbf{a} = \mathbf{N}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f} \tag{44}$$

and the estimate is

$$\hat{u} = \mathbf{u}^T \mathbf{N}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f} \tag{45}$$

Note that this is exactly the same as a weighted least squares function fit. The MSE is

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{f}^T (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f}. \tag{46}$$

Le Traon (1990) suggested a mapping procedure that was an objective function fit to estimate the mean, with an objective map to estimate the fluctuating part of the field. Suppose the data is composed of a mean, a fluctuation (as due to eddies), and noise

$$\mathbf{u} = \langle \mathbf{u} \rangle + \mathbf{u}' + \mathbf{n}, \tag{47}$$

where the means of the fluctuation and noise are zero

$$\mathbf{u}' = \mathbf{n} = 0, \tag{48}$$

the fluctuations and noise are assumed uncorrelated

$$\langle \mathbf{u}'\mathbf{n} \rangle = 0, \tag{49}$$

and the mean field is composed of some functions

$$\langle \mathbf{u} \rangle = \mathbf{F}\langle \mathbf{b} \rangle. \tag{50}$$

The linear estimate is simply (7), and we minimize the MSE with the constraint (37). The MSE is, after some algebra and using (37)

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{a}^T\mathbf{E}\mathbf{a} - 2\mathbf{a}^T\mathbf{c} + \langle u'^2 \rangle, \tag{51}$$

where $\mathbf{E}$ is the data covariance matrix, including contributions from fluctuations and noise

$$\mathbf{E} = \left\langle (\mathbf{u}' + \mathbf{n})(\mathbf{u}' + \mathbf{n})^T \right\rangle, \tag{52}$$

c is the vector of the covariance between the fluctuations at the data locations, and the continuous fluctuation field

$$\mathbf{c} = \langle \mathbf{u}' u'(x, y) \rangle \tag{53}$$

and $\langle u'^2 \rangle$ is the variance of the continuous fluctuation field. The cost function to minimize is

$$\mathcal{L} = \mathbf{a}^T\mathbf{E}\mathbf{a} - 2\mathbf{a}^T\mathbf{c} + \langle u'^2 \rangle - 2\boldsymbol{\lambda}^T(\mathbf{F}^T\mathbf{a} - \mathbf{f}). \tag{54}$$

The resulting gain is

$$\mathbf{a} = \mathbf{E}^{-1}\mathbf{F}(\mathbf{F}^T\mathbf{E}^{-1}\mathbf{F})^{-1}\mathbf{f} + \left[\mathbf{I} - \mathbf{E}^{-1}\mathbf{F}(\mathbf{F}^T\mathbf{E}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\right]\mathbf{E}^{-1}\mathbf{c} \tag{55}$$

The first term on the right-hand side is the objective function fit, and the second term is the objective map of the remainder. The resulting MSE is

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \langle u'^2 \rangle - \mathbf{c}^T\mathbf{E}^{-1}\mathbf{c} + (\mathbf{f}^T - \mathbf{c}^T\mathbf{E}^{-1}\mathbf{F})(\mathbf{F}^T\mathbf{E}^{-1}\mathbf{F})^{-1}(\mathbf{f} - \mathbf{F}^T\mathbf{E}^{-1}\mathbf{c}). \tag{56}$$

The first two terms on the right-hand side are from the objective map, and the second term is the additional error from estimating the mean as an objective function fit. (See Dan Rudnick's notes on objective mapping for a complete example.)

Bibliography

Bretherton, F. P., R. E. Davis, and C. B. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.*, **23** , 559-582.

Daley, R., 1985: The analysis of synoptic scale divergence by a statistical interpolation procedure. *Mon. Wea. Rev.*, **113** , 1066-1079.

Davis, R. E. (1985). Objective Mapping by Least-Squares Fitting. *Journal of Geophysical Research-Oceans*, **90**(NC3), 4773-4777. doi: 10.1029/JC090iC03p04773.

Dever, E. P., 2004: Objective Maps of Near-Surface Flow States near Point Conception, California. *J. Phys. Oceanogr.*, **34**, 444-461, doi:10.1175/1520-0485(2004)034¡0444:OMONFS¿2.0.CO;2.

Gille, S. T., 2003. Float observations of the Southern Ocean: Part 1, Estimating mean fields, bottom velocities, and topographic steering, *J. Phys. Oceanogr*, **33**, 1167-1181.

Le Traon, P. Y. (1990), A method for optimal analysis of fields with spatially variable mean, *J. Geophys. Res.*, **95** (C8), 13543- 13547, doi:10.1029/JC095iC08p13543.