Lecture 19: Linear estimation theory applied to the ocean

Recap

In lecture 18, we worked through some of the details of linear estimation theory and objective mapping. Now we'll fill in some details and look at a machine-learning-based alternatives.

Objective mapping

First let's 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 discussed in Lecture 18 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

$$u_n = \tilde{u}(x_n, y_n) + \epsilon_n. \tag{1}$$

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

$$\langle \tilde{u}\epsilon \rangle = 0 \tag{2}$$

$$\langle \epsilon_n \epsilon_m \rangle = E\delta_{nm}. \tag{3}$$

The assumption (2) 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

$$\tilde{u}(x, y) = \mathbf{a}(x, y)^T \mathbf{u} \tag{4}$$

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

$$\mathbf{a} = \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u}(x, y) \rangle \tag{5}$$

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

$$\frac{\langle (\hat{u} - \tilde{u})^2 \rangle}{\langle \tilde{u}^2 \rangle} = 1 - \frac{\langle \tilde{u}\mathbf{u}^T \rangle \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u} \rangle}{\langle \tilde{u}^2 \rangle}, \tag{6}$$

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 (1-3), the data covariance matrix is

$$\langle \mathbf{u}\mathbf{u}^T \rangle = \langle \tilde{\mathbf{u}}\tilde{\mathbf{u}}^T \rangle + E\mathbf{I}, \tag{7}$$

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.

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

$$\langle \hat{u} \rangle = \langle \tilde{u} \rangle \tag{8}$$

Substituting (4) into (8)

$$\mathbf{a}^T \langle \mathbf{u} \rangle = \langle \tilde{u} \rangle. \tag{9}$$

The assumption that the mean is a constant implies that all the data have the same mean $\langle \tilde{u} \rangle$, so (8) becomes

$$\mathbf{a}^T \mathbf{v} \langle \tilde{u} \rangle = \langle \tilde{u} \rangle \tag{10}$$

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

$$\mathbf{a}^T \mathbf{v} = 1. \tag{11}$$

Our optimization problem is to minimize the MSE $\langle (\hat{u} - \tilde{u})^2 \rangle$ subject to (11). Use the method of Lagrange multipliers to set up a cost function

$$\mathcal{L} = \left\langle \left(\mathbf{a}^T \mathbf{u} - \tilde{u}\right)^2 \right\rangle - 2\lambda(\mathbf{a}^T \mathbf{v} - 1) \tag{12}$$

where λ is a Lagrange multiplier. Differentiating (12) by a, setting the result to zero, and solving yields

$$\mathbf{a} = \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} (\langle \mathbf{u}\tilde{u} + \lambda \mathbf{v}) \tag{13}$$

Plugging (13) into the constraint (11) and solving for λ:

$$\lambda = \frac{1 - \mathbf{v}^T \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u} \rangle}{\mathbf{v}^T \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \mathbf{v}} \tag{14}$$

The zero bias gain thus has an additional term compared to (5)

$$\mathbf{a} = \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \left(\langle \mathbf{u}\tilde{u} \rangle + \mathbf{v} \frac{1 - \mathbf{v}^T \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u} \rangle}{\mathbf{v}^T \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \mathbf{v}} \right). \tag{15}$$

The normalized MSE is

$$\langle (\hat{u} - \tilde{u})^2 \rangle = 1 - \frac{\langle \tilde{u}\mathbf{u}^T \rangle \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u} \rangle}{\tilde{u}^2 \rangle} + \frac{\left(1 - \mathbf{v}^T \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \langle \mathbf{u}\tilde{u} \rangle\right)^2}{\langle \tilde{u}^2 \rangle \mathbf{v}^T \langle \mathbf{u}\mathbf{u}^T \rangle^{-1} \mathbf{v}} \tag{16}$$

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.

Using machine learning to inform mapping

In recent years, a plethora of tools have emerged to support data analysis. A number of recent efforts have taken advantage of machine learning strategies to map irregularly sampled data. This includes neural net based approaches (e.g. Landschützer et al, 2013), and approaches using random forest (e.g. Giglio et al, 2018).

The random forest approach used by Giglio et al (2018) mapped sparsely sampled O_2 measurements by filling gaps using more densely sampled temperature and salinity. Random forest uses a non-linear regression approach that subdivides the data based on different criteria (e.g. temperature exceeding a certain threshold, or salinity less than some threshold) in order to develop an algorithm to determine how to fill missing values.

Examples

Finally, in class, we examined some specific examples of objectively mapped results and discussed a series of questions:

1. What is the mapped quantity?

2. What data were used?

3. What goes in the data–data covariance matrix?

4. What goes in the data–model covariance matrix?

5. What challenges do you see?

One key point that emerges when we look at real examples is the fact that in classic objective mapping, we impose the covariance matrices without knowing the specific data values. The data–data covariance matrix and the data–model covariance matrix depend only on the spatial and temporal separation between data but not on the actual data values.

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 (4) but the signal $\tilde{u}$ is assumed to be a linear combination of some functions

$$\tilde{u} = \mathbf{b}^T \mathbf{f}(x, y) \tag{17}$$

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

$$\mathbf{a}^T \mathbf{F} \langle \mathbf{b} \rangle = \mathbf{f}^T \langle \mathbf{b} \rangle \tag{18}$$

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 (18) is satisfied, but does not require knowledge of the mean $\langle \mathbf{b} \rangle$ is

$$\mathbf{F}^T \mathbf{a} = \mathbf{f} \tag{19}$$

Here we go again, minimizing the MSE with the constraint (19) using Lagrange multipliers. Using (4) and (17), the MSE is

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \left\langle (\mathbf{a}^T \mathbf{u} - \mathbf{b}^T \mathbf{f})^{\mathbf{2}} \right\rangle \tag{20}$$

Suppose that the signal is uncorrelated with the noise as in (2), but the noise covariance matrix N may not be diagonal as in (3). In this case, the MSE is

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{a}^T \left(\mathbf{F} \langle \mathbf{b}\mathbf{b}^T \rangle \mathbf{F}^T + \mathbf{N} \right) \mathbf{a} - 2\mathbf{a}^T \mathbf{F} \langle \mathbf{b}\mathbf{b}^T \rangle \mathbf{f} + \mathbf{f}^T \langle \mathbf{b}\mathbf{b}^T \rangle \mathbf{f} \tag{21}$$

Using the zero bias constraint (19), this can be simplified to

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{a}^T \mathbf{N} \mathbf{a} \tag{22}$$

So the cost function to be minimized is

$$\mathcal{L} = \mathbf{a}^T \mathbf{N} \mathbf{a} - 2\boldsymbol{\lambda}^T (\mathbf{F}^T \mathbf{a} - \mathbf{f}) \tag{23}$$

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

$$\mathbf{a} = \mathbf{N}^{-1} \mathbf{F} \boldsymbol{\lambda} \tag{24}$$

Substitute (24) back into the constraint (19) to get

$$\boldsymbol{\lambda} = (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f} \tag{25}$$

So the gain is

$$\mathbf{a} = \mathbf{N}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f} \tag{26}$$

and the estimate is

$$\hat{u} = \mathbf{u}^T \mathbf{N}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f} \tag{27}$$

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

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{f}^T (\mathbf{F}^T \mathbf{N}^{-1} \mathbf{F})^{-1} \mathbf{f}. \tag{28}$$

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

$$\mathbf{u} = \langle \mathbf{u} \rangle + \mathbf{u}' + \mathbf{n}, \tag{29}$$

where the means of the fluctuation and noise are zero

$$\mathbf{u}' = \mathbf{n} = 0, \tag{30}$$

the fluctuations and noise are assumed uncorrelated

$$\langle \mathbf{u}'\mathbf{n} \rangle = 0, \tag{31}$$

and the mean field is composed of some functions

$$\langle \mathbf{u} \rangle = \mathbf{F} \langle \mathbf{b} \rangle. \tag{32}$$

The linear estimate is simply (5), and we minimize the MSE with the constraint (19). The MSE is, after some algebra and using (19)

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \mathbf{a}^T \mathbf{E} \mathbf{a} - 2\mathbf{a}^T \mathbf{c} + \langle u'^2 \rangle, \tag{33}$$

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

$$\mathbf{E} = \left\langle (\mathbf{u}' + \mathbf{n})(\mathbf{u}' + \mathbf{n})^T \right\rangle, \tag{34}$$

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

$$\mathbf{c} = \langle \mathbf{u}' u'(x, y) \rangle \tag{35}$$

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

$$\mathcal{L} = \mathbf{a}^T \mathbf{E} \mathbf{a} - 2\mathbf{a}^T \mathbf{c} + \langle u'^2 \rangle - 2\boldsymbol{\lambda}^T (\mathbf{F}^T \mathbf{a} - \mathbf{f}). \tag{36}$$

The resulting gain is

$$\mathbf{a} = \mathbf{E}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{E}^{-1} \mathbf{F})^{-1} \mathbf{f} + \left[\mathbf{I} - \mathbf{E}^{-1} \mathbf{F} (\mathbf{F}^T \mathbf{E}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \right] \mathbf{E}^{-1} \mathbf{c} \tag{37}$$

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

$$\left\langle (\hat{u} - \tilde{u})^2 \right\rangle = \langle u'^2 \rangle - \mathbf{c}^T \mathbf{E}^{-1} \mathbf{c} + (\mathbf{f}^T - \mathbf{c}^T \mathbf{E}^{-1} \mathbf{F})(\mathbf{F}^T \mathbf{E}^{-1} \mathbf{F})^{-1} (\mathbf{f} - \mathbf{F}^T \mathbf{E}^{-1} \mathbf{c}). \tag{38}$$

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

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

Giglio, D., Lyubchich, V., and Mazloff, M. R. (2018). Estimating oxygen in the Southern Ocean using Argo temperature and salinity. *Journal of Geophysical Research: Oceans*, **123**, 4280-4297. https://doi.org/10.1029/2017JC013404.

Kuusela, M. and M. L. Stein, (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data *Proc. R. Soc. A.*, **474**:20180400, doi:10.1098/rspa.2018.0400.

Landschützer, P., N. Gruber, D. C. E. Bakker, U. Schuster, S. Nakaoka, M. R. Payne, T. Sasse, and J. Zeng (2013), A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink, *Biogeosciences*, **10**, 7793- 7815, doi:10.5194/bg-10-7793-2013.

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.