May 5: Optimal Interpolation

Also known as: Objective Analysis / Gauss-Markov smoothing

Moving on from the empirical weighted least squares loess filter brings us to the method of *optimal interpolation* (OI), also known as *objective mapping* or *Gauss-Markov smoothing* [see Emery and Thompson, section 4.2].

Optimal interpolation estimates the field being observed at a given location and time through a linear combination of the available data. The weights used are chosen so that the **expected error** of the estimate is a minimum in the least squares sense, and the estimate itself is unbiased (i.e. has the same mean as the true field).

OI is therefore sometimes referred to as the Best Linear Unbiased Estimator (BLUE) of a field.

The underlying covariance length and time scales of the data and true field enter into the computation of the linear weights.

Important concepts in optimal interpolation:

- OI produces the best linear unbiased estimate of a field from a set of arbitrarily distributed observations.

- Central to the estimation procedure is knowledge of the underlying covariance function between the data and the process being observed (the model-data covariance), and between the data themselves (the data-data covariance). This data-data covariance includes an *a priori* estimate of the uncertainty (error) in the observations.

- The model-data and data-data covariance patterns should be consistent.

- If the data errors are independent, the error variance simply adds to the diagonal of data covariance matrix. If the data errors were correlated, off-diagonal elements of the data-data covariance matrix would differ from the model-data covariance, but in practice this is seldom, if ever, considered.

- Frequently, the covariance is assumed to be homogenous and isotropic, in which case the covariance becomes simply a function of the distance separating the locations of the data and model (grid) points.

If valid, assumptions of homogeneity and isotropy facilitate the estimation of the shape of the covariance function by taking an ensemble of data covariance binned according to spatial and/or temporal lags.

The OI method also offers an objective quantitative estimate of the expected error in the result.

The OI technique can be formulated to simultaneously interpolate different but related data types (e.g. winds and geopotential heights) provided a linear relationship exists between the model and data; e,g. as in the case of geostrophic winds, or ocean currents, computed from geopotential (or sea surface) heights.

In the case of geostrophic turbulence, the assumption of isotropy dictates a fixed relationship between the covariance of the individual velocity components and streamfunction.

Simultaneously estimating multiple variables has the advantage that known physical constraints (e.g. continuity, geostrophy) can be incorporated into the mapping procedure, thereby producing results that are kinematically and/or dynamically "balanced".

Examples: Combining altimeter sea surface height observations and velocity observations from sequential satellite imagery.

Wilkin, J. L., M. M. Bowen and W. J. Emery (2002), Mapping mesoscale currents by optimal interpolation of satellite radiometer and altimeter data, Ocean Dynamics, 52, 95-103. (pdf).

Optimal interpolation exploits knowledge of the autocorrelation of a process to determine the relative weight to be given to a set of data (in a weighted sum) to estimate the true field at a certain location (in space and time). The autocorrelation is essentially indicating which data are "near" and which are "far" from the estimation point, and they are weighted accordingly in OI.

Formulating the analysis problem:

Estimate some variable, D, at location(s) x_a, t on the basis of a set of neighboring observations (the data) $\mathbf{d}$ at locations $\mathbf{x}_b$, t.

The data are assumed to be observations of the true field but with some observational error:

$$\mathbf{d}(\mathbf{x},t) = \mathbf{D}(\mathbf{x},t) + \mathbf{n}(\mathbf{x},t)$$

The measurement errors $\mathbf{n}$ are assumed unbiased, $\langle \mathbf{n} \rangle = 0$, and uncorrelated with the field being observed, $\mathbf{D}$.

[In practice it is desirable to remove any well resolved (long space/time scales) deterministic signals from the data first so that the interpolation is being applied to a data set with reduced variance. (For example, a seasonal cycle or spatial variability of very long wavelength.)]

We denote $\hat{D}(x)$ as the estimate of the true value at location *x* (and time *t*), and will compute this as a linear weighted sum of the data:

$$\hat{D}(x) = \bar{D}(x) + \mathbf{w}^T(x)(\mathbf{d} - \bar{\mathbf{d}})$$

where the weights **w**(*x*) are not specified (yet), and the dependence on *x* emphasizes that the weights will be different for every estimate location.

The assumption we have unbiased data implies that the mean of the data, $\bar{\mathbf{d}}$, will be a valid estimate of the mean of the field $\bar{D}$. [Be careful here if it is possible the data sampling might lead to a biased result.]

The weights **w** are selected so as to minimize the expected value of the mean square variance of the error between the linear weighted estimate, $\hat{D}(x)$, and the true value of the variable being observed, *D(x)* (Of course, we don't actually know what this true value is – if we did we probably wouldn't be bothering with all this.)

Therefore, we minimize:

$$\overline{n^2} = \overline{\left(D(x) - \hat{D}(x)\right)^2}$$

true estimate

$$\overline{n^2} = \overline{(D - \bar{D} - (\mathbf{d} - \bar{\mathbf{d}})^T \mathbf{w})^T (D - \bar{D} - (\mathbf{d} - \bar{\mathbf{d}})^T \mathbf{w})}$$
$$= (D - \bar{D})^2 - \mathbf{w}^T(\mathbf{d} - \bar{\mathbf{d}})(D - \bar{D}) - (D - \bar{D})^T(\mathbf{d} - \bar{\mathbf{d}})^T \mathbf{w} + \mathbf{w}^T(\mathbf{d} - \bar{\mathbf{d}})(\mathbf{d} - \bar{\mathbf{d}})^T \mathbf{w}$$

Here, $(\mathbf{d} - \bar{\mathbf{d}})(\mathbf{d} - \bar{\mathbf{d}})^T$ is the data-data covariance matrix which we denote as **C.**

The 2nd through 4th terms are of the form

$$-\mathbf{AB^T} - \mathbf{BA^T} + \mathbf{ACA^T}$$
$$-\mathbf{w}^T(\mathbf{d} - \bar{\mathbf{d}})(D - \bar{D}) - [(\mathbf{d} - \bar{\mathbf{d}})(D - \bar{D})]^T \mathbf{w} + \mathbf{w^T C w}$$

We denote $(D - \bar{D})(\mathbf{d} - \bar{\mathbf{d}})^T$ as the model-data covariance "matrix", $\mathbf{C_{md}}$, i.e. this is the covariance of the true field at the estimation location, *D(x)*, with all the data, **d**.

Note that as written here, $\mathbf{C_{md}}$ is a <u>*row*</u> vector the same length as the data.

The identity of "completing the square" for a simple quadratic algebraic equation finds the constants k_1, k_2 to rearrange …

$$ax^2 + bx + c = a(.....)^2 + \text{constant} = a(x + k_1)^2 + k_2 = ax^2 + 2ak_1x + (ak_1^2 + k_2)$$

When completing the square for the matrix equation above it can be shown that this rearranges to:

$$\mathbf{ACA^T - BA^T - AB^T = (A - BC^{-1})C(A - BC^{-1})^T - BC^{-1}B^T}$$

You can verify this by expanding the line above then simplifying by noting that ***C*** *is symmetric,* $\mathbf{C = C^T}$ *(because it is a covariance matrix), and therefore* $\mathbf{C\left(C^{-1}\right)^T = I}$.

So it follows that:

$$\overline{n^2} = \overline{(D - \bar{D})^2} + (\mathbf{w}^T - \mathbf{C_{md}C^{-1}})\mathbf{C}(\mathbf{w} - \mathbf{C_{md}C^{-1}})^T - \mathbf{C_{md}C^{-1}C_{md}^T}$$

The second term is quadratic (and therefore always greater than zero), so the expected value of n^2 is minimized by making this term zero. This gives us the optimal weights **w**:

$$\mathbf{w - C_{md}C^{-1} = 0}$$

or

$$\mathbf{w = C_{md}C^{-1}}$$

The Best Linear Unbiased Estimate (BLUE) is then

$$\begin{aligned}\hat{D} &= \bar{D} + \mathbf{w^T(d - \bar{d})} \\ &= \bar{D} + \mathbf{C_{md}C^{-1}(d - \bar{d})}\end{aligned}$$

In practice, the data-data covariance matrix can be very large and expense to invert. Typically, it has a much larger dimension than the "model" which would be a grid of coordinates on which we are computing our climatology or analysis.

It is more computationally efficient to compute the product of the weights and data directly by solving, in a least squares sense, the problem:

$$\mathbf{Cw^* = (d - \bar{d})}$$

by a Matlab matrix left divide:

```
>> ws = Cdd\(d-dbar);
```

which gives us the product $\mathbf{w^* = C^{-1}(d - \bar{d})}$ without actually explicitly computing $\mathbf{C^{-1}}$

The OI estimate is then calculated as:

$$\begin{aligned}\hat{D} &= \bar{D} + \mathbf{C_{md}}\mathbf{C}^{-1}(\mathbf{d} - \bar{\mathbf{d}}) \\ &= \bar{D} + \mathbf{C_{md}}\mathbf{w}^{*}\end{aligned}$$

However, we would still need the data-data covariance inverse to make a formal estimate of the *expected error* in the analysis.

Optimal interpolation example

The matlab scripts cov_mercator.m and oi_mercator.m demonstrate fitting a covariance function to a set of synthetic data, and using this function to optimally interpolate to a regular grid.

The data used in this example is ocean temperature taken from the French Mercator operational ocean forecast system for the North Atlantic. Data are available from http://myocean.eu.

`cov_mercator:`

- The script cov_mercator.m loads the example Mercator snapshot from a mat file, subsamples the data to a small (3%) subset and adds some normally distributed random noise to emulate instrument error (or unresolved high frequency physical variability due e.g. to internal waves in the case of in situ ocean temperature observations).

- The lon/lat coordinates of the sub-sampled "data" set are converted to simple x,y, coordinates w.r.t. the southwest corner of the data range, and then the separation distance between all data (r) so that a binned-lagged covariance as a function of r can be computed from the data themselves, i.e. estimate

$$C(r) = \left\langle \mathbf{d(x)d(x}+r\mathbf{)} \right\rangle$$

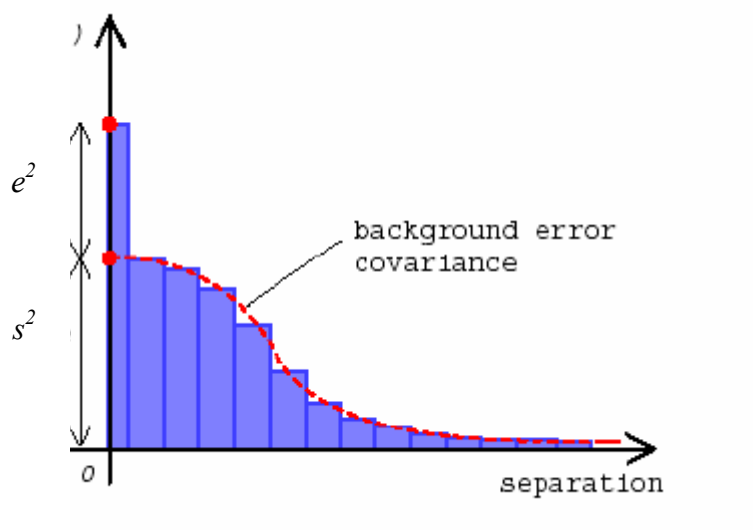
- Two functional forms (Gaussian and Markov) for *C(r)* are fitted to the estimated covariance using Matlab's `fminsearch` function. Note that the covariance at *r=0* is not used in the fit because this includes the effect of the observational, or error, variance.

  Gaussian: $C(r) = s^2 \exp(-\frac{r^2}{a^2})$

  Markov: $C(r) = s^2(1+\frac{r}{a})\exp(-r/a)$

where s^2 is the *signal* variance, i.e. the variance of the true field at zero lag.

- The apparent error variance, e^2, is calculated from the difference of the data variance at $r=0$ (i.e. var[data]) and the signal variance as $r \to 0$ that is indicated by the y-intercept of the functional fit, i.e. $C(0)$.



`oi_mercator:`

- Using the fitted Markov covariance function, normalized model-data ($\mathbf{C_{md}}$) and data-data ($\mathbf{C_{dd}}$) covariance matrices are computed. The data-data is augmented on the diagonal with the ratio of error to signal variance.

- The optimal interpolation fit to the data is computed by direct inversion of $\mathbf{C_{dd}}$ and also by the Matlab matrix-left-divide operation to compute the product $\mathbf{C_{dd}}\ \mathbf{d}$ directly.

- Expected errors of the OI are calculated, and qualitatively compared to the actual residuals of the fit to show that, as expected, approximately 65% of the residuals fall within the expected errors.

If the data set (N points) is large, the matrix sizes may get too large for practical handling in Matlab (or any other language) because the OI problem has to solve matrix inversions or simultaneous equation solutions of dimension NxN.

This may be handled by dividing the model grid into subdomains, like tiles, with subsets of the data limited to only those points that fall within the model 'tile' plus a 'halo' region around the tiles. The halo region should be at least one covariance scale wide to ensure smoothness at the tile boundaries. The data-data matrix $\mathbf{C_{dd}}$ will have to be computed anew for each tile, but computing e.g. order(10) OI operations for order(N/10)

data elements may be faster than one OI for order(N) data elements. This is because the computational efforts of the matrix operations scale with N^3.

If there are too many data within a few covariance scales to practically invert $\mathbf{C_{dd}}$, then it is likely that there are more data than necessary to resolve the mapped field. This indicates that a shorter covariance scale can probably be used. Alternatively, it is probably safe to decimate the data (just use less of it, thereby making N smaller) or average the observations in small bins. For independent errors, the binning step will reduce the expected error (the noise variance) of the binned values, and this information can be carried through the analysis.

Expected errors

A posteriori testing of error estimates can be done to see whether the proportion of residuals within the expected errors is statistically consistent. More in depth tests would compare the results to a set of independent data, such as from another instrument, or data withheld from the OI itself. See Walker and Wilkin (1998) for an example of checking the validity of error estimates though a Chi-squared test.

The expected error is computed in the demonstration script oi_mercator.m

The expected error in the analysis, or estimate, is given by

$$e^2(x_k) = s^2 - \mathbf{c_{md}}\mathbf{C}^{-1}\mathbf{c}_{\mathbf{md}}^{\mathbf{T}}$$

where s^2 is the variance of the signal (the true solution) and $\mathbf{c_{md}}$ is the covariance of the model (at location x_k) with the data $\mathbf{d}$, i.e. it is the k^{th} row of $\mathbf{C_{md}}$.

The vector of all estimated errors is $\mathbf{e}^2 = diag(s^2\mathbf{I} - \mathbf{c_{md}}\mathbf{C}^{-1}\mathbf{c}_{\mathbf{md}}^{\mathbf{T}})$

The optimal interpolation analysis of the data therefore states that our best linear unbiased estimate of the true signal is $\hat{D} \pm e$.

If we were able to make some independent analysis of the error in $\hat{D}$, such as we actually fabricated the data to test the method as in the oi_mercator.m script, then we expect for about 68% of the estimates the true value D would fall within $\hat{D} \pm e$.

From the equation for $\mathbf{e^2}$ we see that the maximum the expected error can be is simply the signal variance. This occurs when there are no data within a few covariance scales of the estimation location and $\mathbf{c_{md}}$ is 0. In this case our best estimate is just the background field an our uncertainty is the full variance of the signal – basically the OI is unable to help inform us.

If we have some data close (in terms of covariance scale) to the estimation location, and

$\mathbf{c}_{md}$ is greater than zero, then the expected error is less than the signal variance and OI has helped us.

If the data error variance is small, the diagonal elements of $\mathbf{C}^{-1}$ are large, and this would further decrease the expected error. So having better quality data improves the skill of the estimate.