

Multi-dimensional Objective Mapping

P. W. Davidson, MIT

Introduction

The need to optimally estimate smooth fields from sparse data appears across many disciplines in the natural sciences. To meet this need, objective mapping (OM; [1–4]) has been successfully employed. Similar to other geostatistical estimation methods such as the many variants of Kriging [5], these methods are all linked by the Gauss-Markov Theorem. Indeed, OM is nearly equivalent to Simple Kriging and differs from Ordinary Kriging in that there is no need to fit a function (e.g., a *semivariogram*) to the data in order to estimate the (co)variance. Instead, such (co)variances are calculated by prescription of a distance-dependent model.

Here, we describe an implementation of OM coupled with maximum likelihood estimation (MLE) to produce fields in N dimensions [6, 7]. While our purposes will be with an application to oceanographic data [8], there is nothing to hold one back from employing our algorithm to data sets from other disciplines. This short document is structured as follows: we first cover the theoretical background on which our algorithm is based before discussing the numerical implementation itself.

Theoretical Background

We proceed by first describing OM through a derivation of minimum variance estimation before then describing how e-folding distance scales are calculated using maximum likelihood estimation.

Minimum variance estimation

We start by assuming that we have a sparse set of noisy data $\mathbf{y} = \{y(\mathbf{s}_i) : i = 1, 2, \dots, N\}^T$, where we assume that $y(\mathbf{s}_i) = y_0(\mathbf{s}_i) + n(\mathbf{s}_i)$ for y_0 the “true” value of y and n the noisy component at the location \mathbf{s}_i . We can think of these \mathbf{y} as realizations of the field which we are trying to estimate. To be concrete, let’s say that we want to estimate this field at unmeasured locations, which we will denote as $\mathbf{x} = \{x(\mathbf{r}_\alpha) : \alpha = 1, 2, \dots, M\}^T$ and for which we seek our best possible estimate, which we will define as $\tilde{\mathbf{x}}$. By “best” here we mean the estimate with the minimum variance. To be able to do this, we will first suppose the following prior information about the low-order statistics of \mathbf{y} and \mathbf{x} , namely that

$$\langle \mathbf{y} \rangle = \langle \mathbf{x} \rangle = 0 \quad \text{and} \quad \langle \mathbf{x} \mathbf{x}^T \rangle \equiv \mathbf{R}_{xx},$$

as well as that at the unknown locations, we let

$$\langle \mathbf{y} \mathbf{y}^T \rangle \equiv \mathbf{R}_{yy} \quad \text{and} \quad \langle \mathbf{x} \mathbf{y}^T \rangle \equiv \mathbf{R}_{xy}.$$

Note the stationarity of the field assumed here, with the mean of the field implicitly subtracted out. Then, we can state our problem as minimizing the diagonal elements of the

variance of our estimate, which we write as

$$\underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \mathbf{P} = \underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \langle (\tilde{\mathbf{x}} - \mathbf{x})(\tilde{\mathbf{x}} - \mathbf{x})^T \rangle. \quad (1)$$

This individual minimization is distinct from the sum of squares minimization. As an initial guess of the minimization solution, we can assume that the solution is linear, i.e., that

$$\tilde{\mathbf{x}} = \mathbf{B}\mathbf{y}. \quad (2)$$

From here, we show what is often referred to as the Gauss-Markov Theorem, which produces a \mathbf{B} such that the diagonal elements of \mathbf{P} are minimized. Substituting (2) into (1) we get

$$\mathbf{P} = \mathbf{B}\mathbf{R}_{yy}\mathbf{B}^T - \mathbf{R}_{xy}\mathbf{B}^T - \mathbf{B}\mathbf{R}_{xy}^T + \mathbf{R}_{xx}. \quad (3)$$

By completing the square^{*}, we can transform (3) into

$$\mathbf{P} = (\mathbf{B} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1})\mathbf{R}_{yy}(\mathbf{B} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1})^T - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T + \mathbf{R}_{xx}. \quad (4)$$

So, noting that all three terms in (4) are positive definite[†], the diagonal entries \mathbf{P} are minimized[‡] when

$$\mathbf{B} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}. \quad (5)$$

Therefore, plugging (5) into (2) gives the minimum variance estimate that we sought in the first place, namely

$$\tilde{\mathbf{x}} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{y}, \quad (6)$$

for which the corresponding variance in the estimate is

$$\mathbf{P} = \mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^T. \quad (7)$$

We finally consider the case when the covariance function \mathbf{R}_{xx} is not well determined near the locations defined by the \mathbf{r}_α —that is, a large constant H has been added to \mathbf{R}_{xx} . To allow for this contingency, Appendix A of [1] shows that in the limit of $H \rightarrow \infty$, an estimated mean $\bar{\mathbf{y}}$ can be removed from the observations before estimation and added back in after, namely

$$\tilde{\mathbf{x}} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}(\mathbf{y} - \bar{\mathbf{y}}) + \bar{\mathbf{y}}, \quad (8)$$

where

$$\bar{\mathbf{y}} = \frac{\sum_{j,i} \mathbf{R}_{yy}^{-1}(\mathbf{r}_j, \mathbf{r}_i) \mathbf{y}_i}{\sum_{j,i} \mathbf{R}_{yy}^{-1}(\mathbf{r}_j, \mathbf{r}_i)}. \quad (9)$$

We principally employ (9) and (7) in our algorithm.

^{*} $\mathbf{ACA}^T - \mathbf{BA}^T - \mathbf{AB}^T = (\mathbf{A} - \mathbf{BC}^{-1})\mathbf{C}(\mathbf{A} - \mathbf{BC}^{-1})^T - \mathbf{BCB}^T$, which we can derive by adding and subtracting \mathbf{BCB}^T to the LHS and rearranging.

[†] We define a positive-definite matrix \mathbf{A} as one that for all non-zero real column vectors \mathbf{x} , $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$.

[‡] To see this clearly, write out the case for a 2×2 matrix.

Maximum likelihood estimation

We discuss how we model the three \mathbf{R} in [Covariance Matrices](#), though it suffices to say for now that we will treat them as a function of a variance σ^2 and an e-folding scale l , i.e. $\mathbf{R}(\sigma^2, l)$. Many instances of OM will subjectively determine what l is given the data set; to circumvent this, we will employ MLE. Following [7], we begin by assuming our data are jointly normally distributed—that is that $\mathbf{y}_i \sim \mathcal{N}(\mu, \sigma)$. We also again assume that our data can be modeled by a stationary random field, letting us define $\boldsymbol{\mu} = \mu \mathbf{1}_{N \times 1}$. Here, $\mathbf{1}_{A \times B}$ is a matrix of ones of the size $A \times B$. Defining $\boldsymbol{\theta} = (\mu, \sigma^2, l)$, we then have that the likelihood of observing \mathbf{y} is given by

$$L(\mathbf{y} | \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}_{yy}(\sigma^2, l)|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{R}_{yy}^{-1}(\sigma^2, l) (\mathbf{y} - \boldsymbol{\mu}) \right]. \quad (10)$$

Following the standard statistical procedure of MLE, the maximum of (10) is the same as its log as it is strictly nonnegative. This then means that we have an objective function, ignoring constants,

$$\mathcal{L}(\mathbf{y} | \boldsymbol{\theta}) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2} \log |\mathbf{R}_{yy}| - \frac{(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{R}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu})}{2\sigma^2}. \quad (11)$$

Taking the partial derivative of (11) with respect to the mean, we find that the MLE of the mean is

$$\tilde{\mu} = \frac{\mathbf{1}_{1 \times N} \mathbf{R}_{yy}^{-1} \mathbf{y}}{\mathbf{1}_{1 \times N} \mathbf{R}_{yy}^{-1} \mathbf{1}_{N \times 1}}. \quad (12)$$

Given (12), let's assume that \mathbf{R}_{yy} is nonsingular and that therefore there are some vector solutions of the form

$$\mathbf{R}_{yy} \mathbf{A} = \mathbf{y} \quad \text{and} \quad \mathbf{R}_{yy} \mathbf{B} = \mathbf{1}_{N \times 1}, \quad (13)$$

which transforms (12) into

$$\tilde{\mu} = \frac{\mathbf{1}_{1 \times N} \mathbf{A}}{\mathbf{1}_{1 \times N} \mathbf{B}}. \quad (14)$$

Now, we determine the maximum likelihood estimator of the variance from (11) as

$$\tilde{\sigma}^2 = \frac{1}{N} (\mathbf{y} - \tilde{\mu} \mathbf{1}_{N \times 1})^T \mathbf{A}. \quad (15)$$

Finally, plugging in (14) and (15) into (11) and dropping the constant $N/2$ which comes from the simplification of the third term, we get that the final likelihood function of \mathbf{y} given the parameters $\boldsymbol{\theta}$ is

$$\mathcal{L}(\mathbf{y} | \boldsymbol{\theta}) = -\frac{N}{2} \log \tilde{\sigma}^2 - \frac{1}{2} \log |\mathbf{R}_{yy}|. \quad (16)$$

Remember that our estimates of the mean and the variance are really functions of \mathbf{R}_{yy} , which in turn is a function of the length scale l . Therefore, to find the optimal scale \tilde{l} , we need to maximize (16) or minimize its negative, namely

$$\underset{l}{\operatorname{argmin}} \left[-\mathcal{L}(\mathbf{y} | \mu, \sigma^2, l) \right]. \quad (17)$$

We discuss this procedure in [Minimization](#).

Algorithm Implementation

We now consider the practical applications of [Theoretical Background](#) by first defining the covariance matrices used in the algorithm before discussing some considerations for the MLE of scales. In both cases, we normalize the data following

$$\overline{y(\mathbf{r}_i)} = \frac{y(\mathbf{r}_i) - \langle \mathbf{y} \rangle}{\langle \mathbf{y} \mathbf{y}^T \rangle}. \quad (18)$$

Note that this empirically normalizes the data, as is assumed in [Maximum likelihood estimation](#). This operation is reversed after the estimation procedures.

Covariance Matrices

Defining these matrices is somewhat of an “art”, so we describe our parameterizations here. We first specify the data-data covariance matrix, \mathbf{R}_{yy} , as

$$\mathbf{R}_{yy} = (\sigma^2 + \mathbb{1}_{N \times N} \boldsymbol{\epsilon}) \cdot \exp \left[- \left(\frac{d(\mathbf{r}_i, \mathbf{r}_j)}{l} \right)^2 \right] + \mathbb{1}_{N \times N} \sigma_n^2,$$

where we let σ^2 be the sample variance at of \mathbf{y} and $\boldsymbol{\epsilon} = \{\epsilon^2(r_i) : i = 1, 2, \dots, N\}^T$ be the vector of individual variances of measurement error for each sample location [\[5, 9\]](#). The \cdot represents element-wise multiplication, and $d(\cdot)$ is the distance metric we impose on the field; in many cases, this is the euclidean distance, including when the small-angle approximation can be made on a sphere. Also, $\mathbb{1}_{A \times A}$ is the identity matrix of the size $A \times A$. Additionally, σ_n^2 is the noise variance which we define as [\[3\]](#)

$$\sigma_n^2 = \frac{1}{2K} \sum_{k=1}^K d_{<\tilde{l}}(\mathbf{r}_i, \mathbf{r}_j)_k^2,$$

where the sum is over K -many data pairs of $d_{<\tilde{l}}$ indexed by i, j that have a distance less than the scale estimated by MLE, \tilde{l} .

Next, we define the model-data covariance matrix, \mathbf{R}_{xy} , as

$$\mathbf{R}_{xy} = (\sigma^2 + \boldsymbol{\epsilon}_{M \times N}) \cdot \exp \left[- \left(\frac{d(\mathbf{s}_\alpha, \mathbf{r}_i)}{l} \right)^2 \right],$$

where we have $\boldsymbol{\epsilon}_{M \times N}$ the matrix of variances corresponding to the appropriate datum in the data-model pair and we assume that the model variance is equivalent to the data variance, σ^2 .

Lastly, we define the model-model covariance matrix, \mathbf{R}_{xx} , as [\[5\]](#)

$$\mathbf{R}_{xx} = \begin{cases} \sigma^2 - (\mathbf{R}_{xy} \mathbf{R}_{yy}^{-1})^2 \boldsymbol{\epsilon} & \text{1 point,} \\ \mathbb{1}_{M \times M} \left[\sigma^2 - (\mathbf{R}_{xy} \mathbf{R}_{yy}^{-1})^2 \boldsymbol{\epsilon} \right] \cdot \exp \left[- \left(\frac{d(\mathbf{s}_\alpha, \mathbf{s}_\beta)}{l} \right)^2 \right] & \text{M-many points.} \end{cases}$$

Note here that we have two cases here. The first is when we are only estimating one point, such is done in Kriging, where we assume the variance of the data is the same as estimate, before removing the propagated measurement variance from the variance of the estimate. When there are M -many location in \mathbf{x} , then we estimate the variance of these points, before again removing the propagated measurement variance.

Minimization

In order to maximize (16), we minimize its negative as given in (17). To do this, we employ the following procedure:

1. initialize \tilde{l} and set appropriate bounds on \tilde{l} over which to search for the minimum
 - (a) compute $\mathbb{R}_{yy}(\sigma^2, \tilde{l})$
 - (b) calculate \mathbf{A} and \mathbf{B} in (13)
 - (c) estimate $\tilde{\mu}$ and $\tilde{\sigma}^2$ in (14) and (15)
 - (d) calculate the log likelihood in (16)
2. repeat steps 1a to 1d until the minimum of the negative of (16) is found

We note here that we let the data variance discussed in [Covariance Matrices](#), σ^2 , be the MLE estimate, $\tilde{\sigma}^2$ of (15), for the optimized \tilde{l} .

To accomplish this procedure, we minimize the negative log likelihood using Brent Minimization. Brent Minimization is the combined method using Successive Parabolic Interpolation when possible and Golden Search method when necessary. This method was proposed by [10]. In general, it is a bracketing method, meaning that there must be two initial values which must bracket the minima (maxima) that we are trying to estimate. Essentially, we are trying to optimize a function $f : \mathbb{R} \rightarrow \mathbb{R}$ which we assume is at least C^1 on an interval $[a, b] = I \in \mathbb{R}$. Then, the problem becomes finding an interior point $x^* \in (a, b)$ such that

$$f'(x^*) = 0.$$

References

1. Bretherton, F. P., Davis, R. E. & Fandry, C. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Research and Oceanographic Abstracts* **23**, 559–582 (1976).
2. Fukumori, I. & Wunsch, C. Efficient representation of the North Atlantic hydrographic and chemical distributions. *Progress in Oceanography* **27**, 111–195 (1991).
3. Bindoff, N. L. & Wunsch, C. Comparison of Synoptic and Climatologically Mapped Sections in the South Pacific Ocean. *Journal of Climate* **5**, 631–645 (1992).
4. Wunsch, C. *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications* (Cambridge University Press, 2006).
5. Webster, R. & Oliver, M. A. *Geostatistics for Environmental Scientists* (John Wiley & Sons, 2007).
6. Jack. easyKriging. *MATLAB Central File Exchange* (2024).
7. Fenton, G. A. Estimation for Stochastic Soil Models. *Journal of Geotechnical and Geoenvironmental Engineering* **125**, 470–485 (1999).
8. Kenyon, J. A., Davidson, P. W., Buesseler, K. O., Pike, S. M. & Lam, P. J. Thorium-234 as a tracer of particulate export and remineralization along the GEOTRACES Pacific Meridional Transect. *Global Biogeochemical Cycles* (2024).
9. Christensen, W. F. Filtered Kriging for Spatial Data with Heterogeneous Measurement Error Variances. *Biometrics* **67**, 947–957 (2011).
10. Brent, R. P. *Algorithms for Minimization without Derivatives* (Prentice-Hall, Englewood Cliffs, NJ, 1973).