A COMPARISON BETWEEN FIML AND REML COVARIANCE FUNCTION ESTIMATION FOR SPATIAL DATA

(上)

BY

RACHUNG YU

尤 瑞 崇

Chapter I

A Linear Model for Spatial Data

1.1 Introduction

The index information of economic recession cycles may be correlated with sequential time data : data of the hydraulic level can be correlated with the location in space. For instance, the levels of radiation contamination in the soil and underground water around a nuclear plant can be correlated. Deposits of high- quality copper can occur near other high-quality deposits. The spread of AIDS may relate to geographical location.

There are an infinite number of cases in which data are collected at various locations in space or various periods of time. Therefore, there are an infinite number of potential applications for methods of analysis for the spatial data. Recently, methods of interpolation for the spatial data have been given wide attention. A closely related approach to interpolation of the spatial data is best linear unbiased prediction(also known as the method of kringing). Two very general methods of kringing are universal kringing, see Goldberger (1962) and Matheron (1969), and intrinsic random function kringing, see Matheron (1973) and Delfiner (1976), For the intention of prediction, these methods are equivalent, see Christensen (1990). A review of kringing spatial data is presented in this paper, where a theoretical analysis of two methods produces analytical expressions that clearly specify the similarities and differences between them. The Full Information Maximum Likelihood (FIML) and Restricted Maximum Likelihood (REML) methods are used in parameter estimation for the purpose of comparison. We will focus on the variance component estimation and comparison.

1.2 General Definitions

Suppose that observations $Z(u_1)$, $Z(u_2)$, $Z(u_3)$,...,$Z(u_n)$, of a regionalized variable $Z(u)$ are available at locations described by Cartesian vectors u_1, u_2, ...u_n where $u_i=(x_i,y_i)$ denotes position in the plane of coordinates X (longitude) and y (latitude). In general, $Z(u_i)$ could represent an average quantity, piezometric value, ie.,

$$Z(u_i) = 1/V \int_V Z(u)du$$

where V is an averaging area centered at u_i, or $Z(u_i)$ may be a point value of the regionalized variable that could represent aquifer properties such as hydraulic conductivity, or specific water reservoir, etc. The location variable $Z(u)$ is modeled as follows:

$$Z(u) = m(u) + e(u),$$

where m(u) is the deterministic drift parameter and E(u) is a mean zero stochastic error process. In this model, m(u) is represented by

$$m(u) = \sum_{r=1}^{k} b_r f_r(u) = F B$$

where $f_r(u)$ is the r-th column vector of matrix F. For example, with k = 3

$$F = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

All b_r are unknown constants where r = 1,2, . . . k, and the functions $f_r(u)$ are in terms of the coordinates of the location vector u.

For any two locations u_i and u_j, the error covariance function is defined as

$$\sigma(u_i, u_j) = Cov(e(u_i), e(u_j))$$

The observations are selected at n different locations $u_1, u_2, \ldots, u_n$, and we will obtain the data $Z(u_1), Z(u_2), \ldots, Z(u_n)$ with

$$Z = \begin{bmatrix} Z(u_1) \\ Z(u_2) \\ \vdots \\ \vdots \\ Z(u_n) \end{bmatrix}$$

With $e_i = e(u_i)$, for $i=1, 2, ..., n$, let

$$e = (e_1, e_2, ..., e_n)'.$$

The matrix equation $Z = FB + e$ is equivalent to $Z(u_i) = F(u_i)B + e_i$, for $i=1,2,....,n$. Furthermore, we will assume that $E(e_i) = 0$, for $i=1,2,....,n$, and

$$Cov(e) = V = \sigma_{ij}$$

where

$$\sigma_{ij} = Cov(u_i, u_j), \text{ for } i,j=1, 2, ..., n.$$

We also make the assumption that V is a nonsingular and positive definite matrix and that F has full column rank.

1.3 Mean and Single Realization Estimation

There are two basically different approaches for the intention of spatial interpolation. The first approach is a single realization of the general effect at the new location u_0. Suppose that the value of $Z(u)$ at u_0, $Z(u_0)$, needs to be estimated. Then the linear interpolator is

$$\hat{Z}(u_0) = \sum_{i=1}^{n} \lambda_i Z(u_i) = \lambda' Z \tag{1.1}$$

and is proposed as an estimator of the single realization $Z(u_0)$. Applying the unbiasedness condition

$$E[\hat{Z}(u_0)] = m(u_0) = F(u_0)B \tag{1.2}$$

implies that

$$F'\lambda = f(u_0) \tag{1.3}$$

Requiring that $Z(u_0)$ have minimum variance of error estimation among $Z(u_0) - \hat{Z}(u_0)$, where $\hat{Z}(u_0)$ represents all linear unbiased interpolators of $Z(u_0)$, we have that $F'\lambda = f(u_0)$ will also satisfy the following equation

$$Var[\,Z(u_0) - \hat{Z}(u_0)\,] = \sigma^2 - 2\,\lambda'\sigma_0 + \lambda' V\lambda \tag{1.4}$$

in which

$$\sigma^2 = Var[\,Z(u_0)\,]$$

$$\sigma_0 = Cov[\,Z(u_0), Z\,]$$

and

$$V = Cov(\,u_i, u_j\,)$$

Equations (1.1), (1.3), and (1.4) define the so-called universal kringing estimator. The weighting vector B is obtained from minimizing Eq.(1.4) subject to Eq.(1.3) producing the Lagrange linear system

$$\begin{bmatrix} V & F \\ F' & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ -\mu \end{bmatrix} = \begin{bmatrix} \sigma_0 \\ f(u_0) \end{bmatrix} \tag{1.5}$$

in which u is a k x 1 vector of Lagrange multipliers. With the solution of Eq. (1.5), the estimator $Z(u_0)$ is obtained from Eq.(1.1) and the variance of $Z(u_0)$ is

$$Var[\hat{Z}(u_0)] = \lambda' V \lambda$$

The second approach in linear interpolation is to estimate the mean $m(u_0)$ of $Z(u_0)$ by using the equation

$$\hat{m}(u_0) = F'(u_0)\hat{B}$$

and

$$\hat{B} = (F'V^{-1}F)^{-1}F'V^{-1}Z$$

where $\hat{B}$ is the estimator of the drift parameter vector. The variance of $\hat{m}(u_0)$ is

$$Var[\hat{m}(u_0)] = F'(u_0)\ V\ B\ F(u_0)$$

The best linear unbiased predictor is

$$Z_0 = u_0' B + V_0' V^{-1}(Z - FB)$$

Where $V_0' = ((u_0,u_1), (u_0,u_2), \ldots\ldots, (u_0,u_n))$, and B is described as above,(see Christensen 1987).

1.4 Linear Model and Covariance Structure

Consider the linear model

$$Z = m(u) + e(u)$$

$$m(u) = \sum_{r=1}^{k} b_r f_r(u) = FB$$

Using the same notation as before

$$Z = FB + e$$

where $Z(u_1), Z(u_2), \ldots, Z(u_n)$ are the piezometric values with the corresponding location vectors $u_1, u_2, \ldots, u_n$, and

$$Z = \begin{bmatrix} Z(u_1) \\ Z(u_2) \\ \vdots \\ \vdots \\ Z(u_n) \end{bmatrix}$$

The location vectors u_i ($u_i = (x_i, y_i)$), for $i = 1,2,\ldots,n$, come from the observations' location values and contain two components that are indexed to provide the value of longitude x_i and the value of latitude y_i. Therefore,

$$m(u) = FB$$

and

$$m(u_i) = b_1 + b_2 x_i + b_3 y_i \quad , \text{for } i = 1, 2, \ldots, n.$$

The values of b_1, b_2, and b_3 are unknown.

The random component $e(u)$ is assumed to be second order (weakly) stationary, see Christensen R. (1990 a), XXI. It has zero expected value, its covariance is assumed to be a function of the separation distance between two different points, and it does not depend on the position or orientation in space.

In particular, we focus on the special model for the covariance function, the Gaussian. The Gaussian covariance function is

$$\sigma(\| u_i - u_j \| ; \theta) = \begin{cases} \theta_2 \exp\left[-\theta_3 \| u_i - u_j \|^2\right], & \| u_i - u_j \| > 0 \\ \theta_1 + \theta_2, & \| u_i - u_j \| = 0 \end{cases}$$

where θ_1, θ_2, and θ_3 are nonnegative and

$$\| u_i - u_j \|^2 = (x_i - x_j)^2 + (y_i - y_j)^2$$

In this covariance model, the variance of each observation is $\theta_1 + \theta_2$. The parameter θ_1 can be treated as a measurement error or a nugget effect. Hence, the covariance function is linear in two of its three parameters. Since the covariance matrix for the linear model can be written as

$$V = \theta_1 I_n + \theta_2 \exp[-\theta_3 \| u_i - u_j \|^2] ,$$

V is linear in θ_1 and θ_2 for fixed θ_3.

Chapter II

Full Information Maximum Likelihood Estimation (FIML)

2.1 Background Information for FIML

The first method that we introduce is FIML. Assume that Z is multivariate normal and that V is nonsingular. Then the density function of Z exists. The density is

$$2\Pi^{-n/2}\,|V|^{-1/2}\,\exp[-((Z-FB)'V^{-1}(Z-FB))/2]$$

The log likelihood is

$$L(B,V) = -(n/2)\log(2\Pi) - (1/2)\log(|V|) - (Z-FB)'V^{-1}(Z-FB)/2$$

where |V| is the determinant of V. For any value of V, this will be maximized by minimizing $(Z-FB)'V^{-1}(Z-FB)$.

The maximum likelihood estimate of V and the maximum likelihood estimate of FB are obtained by treating the maximum likelihood of V as the true value of V and taking the usual weighted least squares estimator of FB. In order to find the maximum likelihood estimates, take the partial derivatives of the density and let them equal zero. To get these equations, several results on matrix differentiation are needed.

Proposition :

(1) $d(Ax)/dx = A'$.

(2) $d(x'Ax)/dx = 2Ax$.

(3) $d(AB)/dx = (dA/dx)B + A(dB/dx)$

(4) If A is a function of a scalar x, and A is nonsingular, then

$d(A^{-1})/dx = -A^{-1}(dA/dx)A^{-1}$.

(5) If A is a function of a scalar x, and A is nonsingular, then

$d(\log|A|)/dx = \mathrm{tr}[A^{-1}(dA/dx)]$.

The proof of (1), (2), (3) are standard results. (see Ben Noble, 2nd ed.,1977). Formula (5) can be found in Searle (1970).

Proof of(4) :

We know that A is nonsingular and thus, it follows that

$A^{-1}A = AA^{-1} = I$,

$d(A^{-1}A)/dx = dI/dx$,

$(dA^{-1}/dx)A + A^{-1}(dA/dx) = 0$,

$(dA^{-1}/dx)A = -A^{-1}(dA/dx)$,

$(dA^{-1}/dx)AA^{-1} = -A^{-1}(dA/dx)A^{-1}$, and

$$dA^{-1} / dx = -A^{-1}(dA / dx) A^{-1} .$$

2.2 Parameter Estimation

The FIML estimators of B and θi, for i=1,2,...p are obtained by the Newton-Raphson method. In the Newton-Raphson method the (r+1)st iteration is described by the expression

$$Q^{(r+1)} = Q^{(r)} - w \Phi^{-1} \nabla f$$

where ∇f is the gradient of the function L evaluated at $Q^{(r)}$,

$$\nabla f = \begin{bmatrix} dL/ d\theta_1 \\ dL/ d\theta_2 \\ dL/ d\theta_3 \end{bmatrix}$$

and w is a step length factor that makes sure that $L(Q^{(r+1)})$ is a minimum with respect to w (0<w<1). The elements of matrix Φ are the second derivatives of the function L and show as

$$\Phi = [d^2L/ d\theta_i d\theta_j] \quad \text{, for } i= 1, 2, 3, \text{ and } j= 1, 2, 3.$$

We know that

$$L(B, V) = (-n/2) \log (2\Pi) - (1/2) \log (|V|) - (Z-FB)'V^{-1}(Z-FB) /2$$

where

$$V = \theta_1 I_n + \theta_2 \exp [- \theta_3 \| u_i - u_j \|^2]$$

Now we assume some notation as follows :

$dV / d\theta_1 = V_1$

$dV / d\theta_2 = V_2$

$dV / d\theta_3 = V_3$

$dV / d\theta_i d\theta_j = V_{ij}$, for i=1,2,3, and j=1,2,3.

Therefore, we can use the facts from proposition (1)-(5) to derive the derivatives $dL/d\theta_1$, $dL/d\theta_2$, and $dL/d\theta_3$. We have that

$$dL/d\theta_1 = (-1/2) \text{tr}[V^{-1}V_1] + (1/2)(Z-FB)'V^{-1}V_1V^{-1}(Z-FB)$$

$$dL/d\theta_2 = (-1/2) \text{tr}[V^{-1}V_2] + (1/2)(Z-FB)'V^{-1}V_2V^{-1}(Z-FB)$$

and

$$dL/d\theta_3 = (-1/2) \text{tr}[V^{-1}V_3] + (1/2)(Z-FB)'V^{-1}V_3V^{-1}(Z-FB)$$

These results can be used to find the gradient of the function L (∇f). Next step, we will establish the matrix Φ and its inverse matrix.

$$\Phi^{-1} = [d^2L/ d\theta_i d\theta_j]^{-1} \quad \text{, for } i= 1, 2, 3, \text{ and } j= 1, 2, 3.$$

Since $V = \theta_1 I + \theta_2 \exp(-\theta_3 \| u_i - u_j \|^2)$,

$$V_1 = dV / d\theta_1 = I$$

$$V_2 = dV / d\theta_2 = 0 + \exp(-\theta_3 \| u_i - u_j \|^2) = \exp(-\theta_3 \| u_i - u_j \|^2)$$

$$V_3 = dV / d\theta_3 = 0 + \theta_2 (-\| u_i - u_j \|^2) \exp(-\theta_3 \| u_i - u_j \|^2) = \theta_2 (-\| u_i - u_j \|^2) \exp(-\theta_3 \| u_i - u_j \|^2)$$

$$V_{11} = 0$$

$$V_{21} = 0$$

$$V_{13} = 0$$

$$V_{31} = 0$$

$$V_{22} = 0$$

$$V_{12} = dV/d\theta_1 d\theta_2 = d(I)/d\theta_2 = 0$$

$$V_{23} = d^2V/d\theta_2 d\theta_3 = d(\exp(-\theta_3 \|u_i - u_j\|^2))/d\theta_3 = -\|u_i - u_j\|^2 \exp(-\theta_3 \|u_i - u_j\|^2)$$

$$V_{32} = V_{23}$$

and

$$V_{33} = d^2V/d\theta_3 d\theta_3 = d(\theta_2(-\|u_i - u_j\|^2) \exp(-\theta_3\|u_i - u_j\|^2))/d\theta_3 = \theta_2 (\|u_i - u_j\|^2)^2 \exp(-\theta_3 \|u_i - u_j\|^2)$$

Now, we will derive the second derivatives of the log likelihood function L :

$$d^2L/d\theta_1 d\theta_1 = d[(-1/2)tr(V^{-1}V_1) + (1/2)(Z-FB)'V^{-1}V_1V^{-1}(Z-FB] / d\theta_1$$

$$d^2L/d\theta_1 d\theta_1 = (1/2)\, tr(V^{-1}V_1V^{-1}V_1) - (Z-FB)'V^{-1}(V_1V^{-1}V_1)V^{-1}(Z-FB)$$

$$d^2L/d\theta_1 d\theta_2 = d[(-1/2)tr(V^{-1}V_1) + (1/2)(Z-FB)'V^{-1}V_1V^{-1}(Z-FB)] / d\theta_2 = (1/2)\, tr(V^{-1}V_1V^{-1}V_2) - (Z-FB)'V^{-1}(V_1V^{-1}V_2)V^{-1}(Z-FB)$$

$$d^2L/d\theta_1 d\theta_3 = d[(-1/2)tr(V^{-1}V_1) + (1/2)(Z-FB)'V^{-1}V_1V^{-1}(Z-FB)] / d\theta_3 = (1/2)\, tr(V^{-1}V_1V^{-1}V_3) - (Z-FB)'V^{-1}(V_1V^{-1}V_3)V^{-1}(Z-FB)$$

$$d^2L/d\theta_2 d\theta_2 = d[(-1/2)tr(V^{-1}V_2) + (1/2)(Z-FB)'V^{-1}V_2V^{-1}(Z-FB)] / d\theta_2 = (1/2)\, tr(V^{-1}V_2V^{-1}V_2) - (Z-FB)'V^{-1}(V_2V^{-1}V_2)V^{-1}(Z-FB)$$

$$d^2L/d\theta_2 d\theta_3 = d[(-1/2)tr(V^{-1}V_2) + (1/2)(Z-FB)'V^{-1}V_2V^{-1}(Z-FB)] / d\theta_3 = (-1/2)\, tr[V^{-1}(V_{23} - V_2V^{-1}V_3)] + (1/2)\,(Z-FB)'V^{-1}(V_{23}-2V_2V^{-1}V_3)V^{-1}(Z-FB)$$

and

$$d^2L/d\theta_3 d\theta_3 = d[(-1/2)tr(V^{-1}V_3) + (1/2)(Z-FB)'V^{-1}V_3V^{-1}(Z-FB)] / d\theta_3$$

$$= (-1/2)\ tr[V^{-1}(V_{33} - V_3V^{-1}V_3)]$$

$$+ (1/2)\ (Z-FB)'V^{-1}(V_{33}-2V_3V^{-1}V_3)V^{-1}(Z-FB)$$

Using these results, the components of the matrix Φ and its inverse matrix can be evaluated. The elements in Φ, such as $dL/d\theta_1 d\theta_1$, $dL/d\theta_1 d\theta_2$, ..., $dL/d\theta_3 d\theta_3$, are 1x1 matrices. Therefore, they can be treated as single values. The covariance matrix V is nonsingular and positive definite. Thus, $V_{12} = V_{21}$, $V_{13} = V_{31}$, $V_{23} = V_{32}$ and we have that

$$d^2L/d\theta_1 d\theta_2 = d^2L/d\theta_2 d\theta_1$$

$$d^2L/d\theta_1 d\theta_3 = d^2L/d\theta_3 d\theta_1$$

and

$$d^2L/d\theta_2 d\theta_3 = d^2L/d\theta_3 d\theta_2$$

A variation of the Newton - Raphson method is used to figure out the covariance parameters $\theta_1, \theta_2, \theta_3$. The basic structure is shown in the following equation.

$$Q^{(r+1)} = Q^{(r)} - w\ \Phi^{-1}\nabla f$$

Normally, we will assign the value of the weight step length factor w to be 0.1 . In each iteration, we calcuate the values of b_1, b_2, b_3, by using the formula $\hat{B} = (F'V^{-1}F)^{-1}F'V^{-1}Z$. All assumptions and first guesses will be shown in the next section.

2.3 Computer Implementation of FIML

In this section, we will show all major components in the program and use several blocks to explain the idea.

Block 1 :

Write a FORTRAN program to figure out the distance between the observations. Then embed the data set file and distance file into the main program in MINITAB.

Block 2:

It is customary to let the initial value of the variance components equal 1. Let θ_1-1.0, θ_2-1.0, θ_3-1.0, b_1-1.0, b_2-1.0, and b_3-1.0.

Block 3 :

Compute the values V, V_1, V_2, V_3, V_{32}, V_{23}, V_{33}.

Block 4 :

Compute the values $\frac{dL}{d\theta_i}$, $\frac{d^2 L}{d\theta_i d\theta_j}$, for i,j- 1, 2, 3.

Block 5 :

Calculate the equation $Q^{(r+1)} = Q^{(r)} - w\,\Phi^{-1}\nabla f$, including the evaluation of Φ^{-1}.

Block 6 :

Compute the values of b_1, b_2, and b_3 using the certain equation $\hat{\beta}$-$(F'V^{-1}F)^{-1}F'V^{-1}Z$.

Block 7 :

Let the program run 20 or 30 iterations. The number of iterations depends on the rate of convergence.