Generalised Kriging with Environmental Applications

Luigi Ippoliti
DMQTE, University G. d'Annunzio, Chieti-Pescara
email: ippoliti@unich.it

Abstract: We consider the problem of spatial interpolation and outline the theory behind kriging and more specifically intrinsic random function kriging. We also mention thin-plate spline theory and show its link with kriging in order to overcome problems in which the available data are not sufficient to estimate the spatial covariance structure of the process. A generalization of the theory to include kriging with directional derivatives is also considered.

Keywords: Kriging, Spatial processes; Thin-plate splines, Derivative process.

1 Introduction

In this paper we are interested in predicting or interpolating values of a spatial process X. Many models for spatial and spatio-temporal data use Gaussian Random Fields (GRFs), and the geostatistical approach of specifying the covariance function, and hence determining the variance matrix $\boldsymbol{\Sigma}$. In this paper, the approach is to assume that a specified covariance function is of interest and that interpolation of the process is required. However, we consider the case in which there is only a little prior knowledge of the field so that an estimation of the covariance structure is difficult or unfeasible. In this framework, we exploit the one-to-one-correspondence between Reproducing Kernel Hilbert Spaces (RKHSs) and positive semi-definite (p.s.d) functions and show that thin-plate splines, considered as a special case of RKHSs, provides a useful solution of the interpolation problem.

2 Materials and Methods

2.1 Kriging

In this section we provide a brief introduction of the kriging predictor for both stationary and intrinsic random fields. For convenience, for known results we mainly refer here to Cressie (1993) and Mardia *et al.* (1996).
Suppose that a spatial process, $\{X(\mathbf{t}), \mathbf{t} \in \mathcal{R}^d\}$, is observed at sites $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n$ with $\mathbf{t}_i = (t_i[1], \ldots, t_i[d])^T$. An important problem in spatial analysis is to predict $X(\mathbf{t}_0)$ at some new site $\mathbf{t}_0 \in \mathcal{R}^d$. The problem reduces to find a predictor of the form

$$\hat{X}(\mathbf{t}_0) = \boldsymbol{\lambda}^T \mathbf{x} \tag{1}$$

where $\mathbf{x} = [x(\mathbf{t}_1), x(\mathbf{t}_2), \ldots, x(\mathbf{t}_n)]^T$ and $\boldsymbol{\lambda}$ is a $(n, 1)$ coefficient vector chosen to minimise the prediction variance, $E\left\{[X(\mathbf{t}_0) - \hat{X}(\mathbf{t}_0)]^2\right\}$, subject to the unbiasedness constraint, $E\left[X(\mathbf{t}_0) - \hat{X}(\mathbf{t}_0)\right] = 0$. To provide a solution at this problem, suppose the random field has some polynomial drift of order r. Also let $\mathcal{G}_r$ be the space of these polynomial terms, whose dimension is given by $M = \left(\begin{array}{c} d+r \\ d \end{array}\right)$. Denote with $\mathbf{U} = \{u_{im}\};\ u_{im} = \mathbf{t}_i^m, i = 1, 2, \ldots, n, |m| \leq r$, the (n, M) drift (design) matrix and with $\mathbf{u}_0$ the vector of drift terms at $\mathbf{t}_0$, with elements $\mathbf{t}_0^m, |m| \leq r$. Assume also that $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ is a non-singular covariance matrix with entries obtained by defining a "potential" function, $\sigma(|\mathbf{h}|) = \sigma(|\mathbf{t}_i - \mathbf{t}_j|),\ i, j = 1, \ldots, n$. Finally, let $\sigma^2 = \sigma(0)$ and $\boldsymbol{\sigma}_0$ a covariance vector with elements $\sigma(|\mathbf{t}_0 - \mathbf{t}_i|),\ i = 1, \ldots, n$. Following this notation, it can be shown that $\boldsymbol{\lambda}$ is given by

$$\boldsymbol{\lambda} = \mathbf{A}\mathbf{u}_0 + \mathbf{B}\boldsymbol{\sigma}_0 \tag{2}$$

where $\mathbf{A}$ and $\mathbf{B}$ are (n, M) and (n, n) matrices respectively, whose form depends on the underlying assumptions about the random process, and in particular on the properties of the covariance matrix $\boldsymbol{\Sigma}$. Also, note that for the case when $X(\mathbf{t})$ is an intrinsic random field, we have the additional constraint that the coefficients of the prediction error, $\left[\sum_i \lambda_i x(\mathbf{t}_i) - X(\mathbf{t}_0)\right]$, are generalised increments of order r. This constraint can be written in the form $\mathbf{U}^T\boldsymbol{\lambda} = \mathbf{u}_0$, which is the only constraint we need in the kriging problem (Cressie, 1993).

In the stationary random field case, it is known that $\boldsymbol{\Sigma}$ is positive definite. In the case of an intrinsic random field, the assumption that $\boldsymbol{\Sigma}$ is positive definite is no longer valid. Hence we must find an alternative form for the kriging predictor which only requires $\boldsymbol{\Sigma}$ to be conditionally positive definite. It can be shown that $\boldsymbol{\lambda}$ still takes the form of equation (2), but $\mathbf{A}$ and $\mathbf{B}$ must be represented in a way which does not require $\boldsymbol{\Sigma}$ to be positive definite. Provided $\boldsymbol{\Sigma}$ is non-singular, one method for determining $\mathbf{A}$ and $\mathbf{B}$ is to define the matrices (Mardia *et al.*, 1996)

$$\mathbf{K} = \left[\begin{array}{cc} \boldsymbol{\Sigma} & \mathbf{U} \\ \mathbf{U}^T & \mathbf{0} \end{array}\right] \quad \text{and} \quad \mathbf{K}^{-1} = \left[\begin{array}{cc} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{array}\right]$$

such that

$$\mathbf{A} = \mathbf{K}^{12} \quad \text{and} \quad \mathbf{B} = \mathbf{K}^{11}.$$

Then, by setting $\mathbf{a} = \mathbf{A}^T\mathbf{x}$ and $\mathbf{b} = \mathbf{B}^T\mathbf{x}$, we may write

$$
\begin{aligned}
\hat{X}(\mathbf{t}_0) &= \mathbf{a}^T\mathbf{u}_0 + \mathbf{b}^T\boldsymbol{\sigma}_0 \\
&= \sum_{j=1}^{M} a_j u_{j0} + \sum_{i=1}^{n} b_j \sigma(\mathbf{t}_0, \mathbf{t}_i).
\end{aligned} \tag{3}
$$

One possible common choice for $\sigma(\cdot)$ is any valid covariance function for a stationary stochastic process in space, for which any null space of functions $\mathcal{G}_r$ will suffice. However, in all cases in which the number of spatial sites is small and the estimation of the covariance structure appears difficult, the following class of functions is useful:

$$
\sigma_\alpha(\mathbf{h}) = \begin{cases} |\mathbf{h}|^{2\alpha} \log |\mathbf{h}|, & \text{for } \alpha \text{ an integer} \\ |\mathbf{h}|^{2\alpha}, & \text{for } \alpha \text{ not an integer} \end{cases}
$$

where α is a smoothness parameter which can be specified ahead of time. Note that this class of functions represents the covariance functions for intrinsic random functions of order $k = [\alpha]$, with $[\alpha]$ the integer part of α. Also, these functions are self-similar and so $\sigma_\alpha(\mathbf{h})$ and $\sigma_\alpha(c\mathbf{h})$ yield the same predictions.

2.2 Kriging with derivative information

There are some cases of interest in which the information about objects (e.g. plants and weeds in crop images) comes from the boundary, which is a continuous curve. In this framework, the kriging predictor can be used in order to modelling the continuous outline of the object. A key aspect to our particular problem is thus the introduction of some extra information, such as derivatives, in order to get a better performance of the modelling procedure.

Suppose that all the known values and derivatives of a specific spatial pattern are collected into a vector $\mathbf{y}$. Let $\boldsymbol{\kappa}$ be a vector of corresponding indices to show the order of the derivative; for example, assuming $d = 1$, $\kappa_i = 0$ if y_i is a data value, $\kappa_i = 1$ if y_i is a first derivative. For each site $\mathbf{t}_j$ there may be several choices of κ_i if the value of the function and of some of its derivatives are all known at that site. The problem is, as before, to find a coefficient vector $\boldsymbol{\lambda}$ such that $\hat{X}(\mathbf{t}_0) = \boldsymbol{\lambda}^T\mathbf{y}$ is the best unbiased linear estimator of $X(\mathbf{t}_0)$.

Let $\kappa_i = \big(\kappa_i[1], \ldots, \kappa_i[d]\big)$ be a $d-$dimensional multi-index of non-negative integers with $|\kappa_i| = \kappa_i[1] + \ldots + \kappa_i[d]$. If we have derivative information of order $p = |\kappa_i|$, then α must satisfies the inequality $\alpha > p$. In the following we also take the smallest drift allowable, that is $r = [\alpha]$. In order to define the $d-$dimensional kriging predictor for derivatives of order up to and including p and with polynomial drift of order r, the covariance matrix, $\boldsymbol{\Sigma}$, consists of entries which have the form

$$
\sigma_{ij} = (-1)^{|\kappa_j|} \sigma_\alpha^{(\kappa_i + \kappa_j)}(\mathbf{t}_i - \mathbf{t}_j), \quad 1 \le i, j \le n
$$

where $\sigma_\alpha^{(\kappa_i)}$ denotes the partial derivative of $\sigma_\alpha(\mathbf{h})$ of order κ_i. The drift matrix $\mathbf{U}$ instead consists of elements

$$u_{im} = \frac{\partial^{|\kappa_i|}}{\partial \mathbf{t}_i^{\kappa_i}}(\mathbf{t}_i^m), \quad 1 \leq i \leq n, \quad |m| \leq r.$$

3 Applications

In the following we provide a list of applications in which the kriging framework, as outlined in sections 2.1 and 2.2, plays an important role.
Spatial and spatio-temporal data occur widely. There is often interest in predicting or interpolating values. Some data sets may have missing values at some sites, or at some times. These missing values may be separated or clumped. Separated missing values may, for example, occur because of random instrument malfunctions. An example of clumped missing values occurs with passive satellite images when there is cloud cover.
A different application when prediction or interpolation is required is checking observations which may appear to be aberrant or influential. Typical influence and outlier statistics are based on estimating the values using all other values.
A wide range of methods used for constructing optimal spatial sampling designs are also based on sampling schemes with minimal prediction variance. However, a basic problem of this approach is that before the actual sampling takes place there is only a little prior knowledge of the field. To overcome this problem, we may set $\alpha = 1$ and chose $\sigma(\mathbf{h}) = |\mathbf{h}|^2 \log|\mathbf{h}|$, which is conditionally positive definite whenever the null space contains the linear trend, *i.e.* $\mathcal{G}_r = span\,(1, t[1], t[2])\,, \mathbf{t} \in \mathcal{R}^2$. In this case, it turns out that the kriging function (1) is an interpolating thin-plate spline.
Finally, we note that the appearance of agricultural products can be evaluated by considering their size, shape, form and the absence of visual defects. Among these features the shape, also measured through the outline of the object, plays a crucial role. Description of agricultural product shape is often necessary in research fields for a range of different purposes, including the investigation of shape for cultivar descriptions, plant variety or cultivar patents and evaluation of consumer decision performance.

References

Cressie, N. (1993). *Statistics for Spatial Data.* Revised edition. New York: Wiley.

Mardia K.V., Kent, J. T., Goodall C.R., Little, J.A. (1998). Kriging and Splines with Derivative Information. *Biometrika*, 83, 217-285