Sampling Theory and Geostatistics: A Way of Reconciliation

Chapter · June 2014

DOI: 10.1007/978-3-319-05320-2_10

CITATION
1

READS
131

2 authors:

Giorgio Eduardo Montanari
Università degli Studi di Perugia
72 PUBLICATIONS **455** CITATIONS
SEE PROFILE

Giuseppe Cicchitelli
Università degli Studi di Perugia
14 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Sampling Theory and Geostatistics: A Way of Reconciliation

Giorgio E. Montanari and Giuseppe Cicchitelli

Abstract In a previous paper, the authors addressed the problem of the estimation of the mean of a spatial population in a design-based context and proposed a model-assisted estimator based on penalized regression splines. This paper goes a step further exploring the performance of the leave-one-out cross-validation criterion for selecting the amount of smoothing in the regression spline assisting model and for estimating the variance of the proposed estimator. The attention is focused on continuous spatial populations. A simulation study provides an empirical efficiency comparison between this estimator and the kriging predictor by means of Monte Carlo experiments.

Keywords Cross-validation • Kriging • Model-assisted estimator • Non-parametric estimation • Penalized splines

1 Introduction

In environmental studies, dealing with soil characterization, monitoring of natural resources or estimation of pollution concentration, often it is of interest to estimate the mean of a response variable defined continuously over a geographical region.

Let $A \subset R^2$ be a planar domain of interest and let $y(\mathbf{x})$ be the response variable defined for all locations $\mathbf{x} = (x_1, x_2)$ in A, where x_1 and x_2 are the geographical coordinates of the location. A common problem in this context is to estimate the mean of the response variable, using values of $y(\mathbf{x})$ observed in an ordered sample of locations $s = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. We will focus on the case of a spatial continuous population and in such a case the population mean is given by

G.E. Montanari (✉) • G. Cicchitelli
University of Perugia, Perugia, Italy
e-mail: giorgio.montanari@unipg.it

F. Mecatti et al. (eds.), *Contributions to Sampling Statistics*, Contributions to Statistics,
DOI 10.1007/978-3-319-05320-2_10,

$$\bar{Y} = \frac{1}{|A|}\int_A y(\mathbf{x})d\mathbf{x},$$

where $|A| = \int_A d\mathbf{x}$ is the area of domain A.

Classical sampling theory deals with finite populations. But sampling theory and continuous populations came into contact in the statistical analysis of spatial data. In fact, spatial points can be labelled and identified by their geographical coordinates, and this makes it possible to design sampling strategies to draw samples and to do inference on the population parameters. In this regard, pioneering contributions are those of McArthur (1987) and Cordy (1993). In the first paper, various sampling strategies for estimating the average level of a response surface representing a pollutant are compared. The second paper provides the theory for Horvitz–Thompson estimation of means or totals of continuous populations.

The problem of estimating the mean or the total of a continuous spatial population has been addressed, among others, by Brus (2000), Stevens and Jensen (2007) and Barabesi et al. (2012). These papers pursue the improvement of efficiency of the sampling strategy using, at the design stage, the auxiliary information provided by the spatial coordinates of location points and spatially balanced samples come out as optimal. In standard survey methodology, a sample is "balanced" with respect to a given auxiliary variable when the estimated mean of this variable equals the population mean. Higher order balancing can also be achieved. For example, we can require that the first two estimated moments (i.e., mean and variance) of the auxiliary variable match the population corresponding moments. This technique relies on the idea that when the auxiliary variables are correlated with the response variable, balancing over the auxiliary variables should produce an approximate balance over the unknown response, and a more precise estimate of the response population mean is expected.

For spatial populations, the most important auxiliary variable is the spatial location of points. Then, if we believe that the response variable has a spatial structure in the study domain, a balanced sample is one that is evenly and regularly spread out across the domain as much as possible (Stevens and Jensen 2007, p. 517).

There are various techniques to obtain spatially balanced samples. One of the most frequently used methods is spatial stratification by overimposing on the domain regular polygons or other tessellation criterions; we refer, in particular, to the random tessellation stratified (RTS) design and to the generalized random tessellation stratified (GRTS) design described in Stevens (1997) and in Stevens and Olsen (2004). These designs, coupled with the Horvitz–Thompson estimator of the mean, yield very efficient strategies compared with the standard benchmark represented by the uniform random sampling and the sample mean.

An alternative approach consists in using the auxiliary information provided by the locations of sample units at the estimation stage, looking for more efficient estimator than the Horvitz–Thompson one. In this regard, Cicchitelli and Montanari (2012), using a penalized spline regression model to capture the spatial pattern in

the data, propose a design-consistent and approximately unbiased estimator of the population mean, within a model-assisted framework.

In this paper, we further develop the approach of Cicchitelli and Montanari focusing on the choice of the smoothing parameter for the penalized spline regression estimator. In particular, we report and discuss results from a wide simulation study carried out to explore this issue.

2 The Penalized Spline Regression Model-Assisted Estimator of the Population Mean

Let $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_K$ be K locations in A, called knots, and let

$$\mathbf{z}(\mathbf{x}) = [z_1(\mathbf{x}), \ldots, z_K(\mathbf{x})] = \left[\breve{z}_1(\mathbf{x}), \ldots, \breve{z}_K(\mathbf{x})\right] \boldsymbol{\Omega}^{-1/2}$$

be a row vector of covariates, where $\breve{z}_l(\mathbf{x}) = (||\mathbf{x} - \boldsymbol{\kappa}_l||)^2 \log(||\mathbf{x} - \boldsymbol{\kappa}_l||)$, with $l = 1, \ldots, K$ and $\boldsymbol{\Omega} = \{(||\boldsymbol{\kappa}_l - \boldsymbol{\kappa}_k||)^2 \log(||\boldsymbol{\kappa}_l - \boldsymbol{\kappa}_k||)\}_{k,l=1,\ldots,K}$.

Assume that the response variable $y(\mathbf{x})$ at location $\mathbf{x} \in A$ is a realization of the super-population model according to which

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mathbf{z}(\mathbf{x})\mathbf{u} + \varepsilon(\mathbf{x}), \tag{1}$$

where $\beta_0, \beta_1, \beta_2$ and $\mathbf{u}' = [u_1, \ldots, u_K]$ are unknown regression parameters and $\varepsilon(\mathbf{x})$ is a white noise with variance σ_ε^2.

Assuming that the sample of locations s is randomly drawn from A with a fixed size n sampling design that assigns inclusion density $\pi(\mathbf{x})$ to location $\mathbf{x}$ (see Cordy 1993), Cicchitelli and Montanari (2012) introduce the design-consistent and approximately unbiased *p-spline regression estimator* (PSRE) of the population mean, which is given by

$$\hat{\bar{Y}}_{\text{spl}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \bar{\mathbf{z}}(\mathbf{x})\,\hat{\mathbf{u}} + \frac{1}{|A|} \sum_{i=1}^{n} \frac{\hat{e}(\mathbf{x}_i)}{\pi(\mathbf{x}_i)}, \tag{2}$$

where $\bar{x}_1, \bar{x}_2$ and $\bar{\mathbf{z}}(\mathbf{x})$ are the population mean in A of x_1, x_2 and $\mathbf{z}(\mathbf{x})$; $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\mathbf{u}}$ are the design-based estimates of the census parameters $\beta_{0P}, \beta_{1P}, \beta_{2P}, \mathbf{u}_P$ that would be obtained by fitting model (1) to the whole population by means of the penalized least square criterion which, in the continuous case, consists in minimizing

$$\int_A [y(\mathbf{x}) - \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \mathbf{z}(\mathbf{x})\mathbf{u}]^2 d\mathbf{x} + \lambda \mathbf{u}'\mathbf{u},$$

for some fixed value of the penalizing parameter $\lambda \geq 0$; $\hat{e}(\mathbf{x}_i) = y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)$ is the residual in $\mathbf{x}_i$, i.e. the difference between the observed value $y(\mathbf{x}_i)$ and the fitted value obtained from the function

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \mathbf{z}(\mathbf{x})\hat{\mathbf{u}}, \quad \mathbf{x} \in A. \tag{3}$$

The estimated census parameters in (2) are given by

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \left[\begin{bmatrix} \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \\ \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \end{bmatrix} + \lambda \mathbf{D} \right]^{-1} \begin{bmatrix} \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \\ \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \end{bmatrix} \mathbf{y}_s, \tag{4}$$

where $\hat{\boldsymbol{\beta}}' = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$; $\mathbf{X}_s$ is the matrix of order $n \times 3$ having as i-th row $(1, x_{i1}, x_{i2})$, for $i = 1, \ldots, n$; $\mathbf{Z}_s$ is the matrix of order $n \times K$ whose i-th row is given by $\mathbf{z}(\mathbf{x}_i)$; $\boldsymbol{\Pi}_s = \text{diag}\{\pi(\mathbf{x}_1), \ldots, \pi(\mathbf{x}_n)\}$ is the diagonal matrix having as elements the inclusion density values of sample locations; $\mathbf{D} = \text{blockdiag}\{\mathbf{0}_{3\times 3}, \mathbf{I}_K\}$, where $\mathbf{I}_K$ is the identity matrix of order K and $\mathbf{0}_{3\times 3}$ is a matrix of zeros of size indicated in the subindex; $\mathbf{y}_s = [y(\mathbf{x}_1), \ldots, y(\mathbf{x}_n)]'$ is the vector of the response sample values.

The approximate unbiasedness and design consistency of (2) have been proved assuming an asymptotic framework in which $n \to \infty$ and the ratio r/n goes to zero, being r the number of degrees of freedom used by the fitted model.

Quantity $\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \bar{\mathbf{z}}(\mathbf{x})\hat{\mathbf{u}}$ on the right-hand side of (2) is the population mean in A of $\hat{y}(\mathbf{x})$, which provides a design-based estimate of the response predicted value at location $\mathbf{x}$ that would be obtained with the census parameters. It is worthwhile noting that function (3) can be also used for mapping purposes. Furthermore, $|A|^{-1} \sum_1^n \hat{e}(\mathbf{x}_i)/\pi(\mathbf{x}_i)$ on the right-hand side of (2) is the Horvitz–Thompson estimator of the population mean of residuals. It is easy to recognize that (2) is a generalized regression (GREG) estimator for the continuous case. However, the weighted sum of the residuals can be shown to be equal to zero since model (1) contains the intercept and the variance of the error term $\varepsilon(\mathbf{x})$ is assumed to be constant (see Särndal et al. 1992, p. 231). So, the PSRE reduces to

$$\hat{\bar{Y}}_{\text{spl}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \bar{\mathbf{z}}(\mathbf{x})\hat{\mathbf{u}}. \tag{5}$$

The covariates $\mathbf{z}(\mathbf{x})$ are radial basis functions for spline regression and this choice corresponds to the low rank *thin plate spline* family of smoothers (Ruppert et al. 2003). Note that function $(\ ||\mathbf{x} - \boldsymbol{\kappa}_l||)^2 \log(||\mathbf{x} - \boldsymbol{\kappa}_l||)$ is a way to describe location $\mathbf{x}$ with respect to the knot $\boldsymbol{\kappa}_l$, and the vector $\mathbf{z}(\mathbf{x})$ is a sort of standardized measures that characterize the geographic position of $\mathbf{x}$ with respect to the K knots. In this way, the spatial information is conveyed into the spline basis, which depends on the spatial configuration of data and knots.

The relation between the penalty λ and the amount of smoothing can be expressed by the number of degrees of freedom used by the fitted model. Since the fitted values are given by $\mathbf{S}_{\lambda s}\mathbf{y}_s$, where

$$\mathbf{S}_{\lambda s} = [\mathbf{X}_s, \mathbf{Z}_s] \left[\begin{bmatrix} \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \\ \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \end{bmatrix} + \lambda \mathbf{D} \right]^{-1} \begin{bmatrix} \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \\ \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \end{bmatrix}, \tag{6}$$

the number of degrees of freedom, r, is given by the trace of the smoothing matrix, which can be also written as

$$r = \text{trace} \left\{ \left[\begin{bmatrix} \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \\ \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \end{bmatrix} + \lambda \mathbf{D} \right]^{-1} \begin{bmatrix} \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{X}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \\ \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s & \mathbf{Z}_s' \boldsymbol{\Pi}_s^{-1} \mathbf{Z}_s \end{bmatrix} \right\}.$$

It is easily seen that $3 \leq r \leq K+3$. The penalty parameter λ has the far most important impact on r: in fact, when $\lambda = 0$, then $r = K+3$, while r attains the minimum 3 as $\lambda \to \infty$. The value of r is sometimes called *equivalent number of parameters* of the smoother.

2.1 Variance Estimation and Penalizing Parameter

Our previous discussion has been embedded within the model-assisted approach to design-based inference. Estimator $\hat{\bar{Y}}_{\text{spl}}$ is a GREG estimator and the standard design-consistent variance estimator for fixed sample size is given by

$$\hat{V}\left(\hat{\bar{Y}}_{\text{spl}}\right) = \frac{1}{|A|^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\pi(\mathbf{x}_i)\pi(\mathbf{x}_j) - \pi(\mathbf{x}_i, \mathbf{x}_j)}{\pi(\mathbf{x}_i, \mathbf{x}_j)} \left(\frac{\hat{e}(\mathbf{x}_i)}{\pi(\mathbf{x}_i)} - \frac{\hat{e}(\mathbf{x}_j)}{\pi(\mathbf{x}_j)} \right)^2, \tag{7}$$

where $\pi(\mathbf{x}_k, \mathbf{x}_l)$ is the second-order inclusion density (Cordy 1993). This estimator would be design unbiased if the census parameters $\beta_{0P}, \beta_{1P}, \beta_{2P}$ and $\mathbf{u}_P$ were known. But when they are estimated, variance estimator (7) is downward biased since it does not include the variance component due to the estimation of the census parameters by means of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\mathbf{u}}$. The bias may be substantial when the ratio between the sample size n and the number of degrees of freedom r is small. In fact, high values of r compared to n might cause an overfitting of sample data, and, as a consequence, might produce sample residuals much smaller than those for unsampled locations. In the simulation study in Cicchitelli and Montanari (2012), the 95 % confidence interval built with $\hat{\bar{Y}}_{\text{spl}}$ and the variance estimator (7) has an effective coverage below 92 % as soon as the ratio n/r goes under 6. These results put a limit on the choice of the smoothing parameter λ and, therefore, on the number of degrees of freedom of the smoother, as long as estimator (7) is employed for estimating the variance.

The problem of optimal selection of the smoothing parameter has not been addressed by Cicchitelli and Montanari (2012). In this regard, Opsomer and Miller (2005) propose a *design-based* cross-validation (CV) criterion for selecting

the bandwidth of their local polynomial regression estimator. Mimicking their approach, our proposal is to replace in (7) each $\hat{e}(\mathbf{x}_i)$ with the "leave-one-out" residual $\hat{e}_{(-i)}(\mathbf{x}_i)$, where the predicted response in $\mathbf{x}_i$ is computed using the regression function (3) estimated after having removed location $\mathbf{x}_i$ from the sample $s = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Replacing $\hat{e}(\mathbf{x}_i)$ with $\hat{e}_{(-i)}(\mathbf{x}_i)$ in (7), we get the design-based CV criterion for choosing the smoothing parameter λ of the PSRE:

$$\hat{V}_{\mathrm{CV}}\left(\hat{\bar{Y}}_{\mathrm{spl}};\lambda\right)=\frac{1}{|A|^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\pi(\mathbf{x}_i)\pi(\mathbf{x}_j)-\pi(\mathbf{x}_i,\mathbf{x}_j)}{\pi(\mathbf{x}_i,\mathbf{x}_j)}\left(\frac{\hat{e}_{(-i)}(\mathbf{x}_i)}{\pi(\mathbf{x}_i)}-\frac{\hat{e}_{(-j)}(\mathbf{x}_j)}{\pi(\mathbf{x}_j)}\right)^2. \quad (8)$$

Then, the value of λ for PSRE is chosen to be

$$\lambda_s = \arg\min_{\lambda\geq 0}\hat{V}_{\mathrm{CV}}\left(\hat{Y}_{\mathrm{spl}};\lambda\right).$$

This criterion amounts to select the value of λ that minimizes the estimated sampling variance computed with the "leave-one-out" residuals.

It is worthwhile noting that it is not necessary to repeat the fitting to compute $\hat{e}_{(-i)}(\mathbf{x}_i)$, since for linear smoothers whose smoothing matrix has each row summing to one, as $\mathbf{S}_{\lambda s}$ in (6), it can be shown that

$$\hat{e}_{(-i)}(\mathbf{x}_i) = \frac{\hat{e}(\mathbf{x}_i)}{1 - S_{\lambda s}(i,i)},$$

where $S_{\lambda s}(i,i)$ is the (i, i)-th entry of the smoother matrix $\mathbf{S}_{\lambda s}$ (see Hastie and Tibshirani 1990, p. 47).

The rationale behind the above procedure can be summarized as follows. First, replacing $\hat{e}(\mathbf{x}_i)$ with $\hat{e}_{(-i)}(\mathbf{x}_i)$ in (7) preserves the estimator against the risk of overfitting. Second, as a function of λ, it allows choosing the value of $\hat{\bar{Y}}_{\mathrm{spl}}$ for which $\hat{V}_{\mathrm{CV}}\left(\hat{\bar{Y}}_{\mathrm{spl}};\lambda\right)$ and, hopefully, the mean square error of $\hat{\bar{Y}}_{\mathrm{spl}}$, i.e. $\mathrm{MSE}_p\left(\hat{\bar{Y}}_{\mathrm{spl}}\right)$, attains its minimum value. In this respect, Opsomer and Miller (2005) proved that the design-based CV statistics they propose converges to the MSE of the local polynomial estimator under a suitable asymptotic framework and conditions on the sampling design and on the population.

3 The Block Kriging Predictor of the Mean

In Geostatistics the problem we are dealing with is managed with kriging, which is a model-based technique producing a predictor of the response variable which is unbiased and with minimum prediction variance under a second order stationary

process. The population mean is estimated with the mean of the predicted values for all locations in the domain. The resulting estimator of the mean is called *block kriging predictor*. This estimator, which is essentially a weighted average of the response variable sample values, depends heavily on the coordinates of the sample points in $s = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. In fact, loosely speaking, the weights attached to observations $y(\mathbf{x}_1), \ldots, y(\mathbf{x}_n)$ are functions of the distances between the sample locations, as well as of the distances between sampled and unsampled locations in the domain (Cicchitelli and Montanari 1997).

The above features give to block kriging predictor the role of a natural benchmark for sampling strategies that make use of the coordinates of sample points within the design-based paradigm. In fact, the block kriging predictor is model unbiased and with the minimum prediction variance conditional on the selected sample; then, provided that the model holds true, it follows that jointly under the model and the sampling design it is unbiased and with the minimum predicition variance. For this reason, it makes sense comparing the performance of alternative estimators with that of block kriging predictor, even if it is evaluated from a repeated sampling point of view. In addition, studying the behaviour of this predictor in repeated sampling from a fixed population may be useful to explore its appropriateness as an estimator within the design-based framework. In this respect, McArthur (1987) compares kriging and design-based methods on simulated spatial data, concluding that kriging is design biased. Brus and de Gruijter (1997) present a simulation study where a stratified design coupled with the Horvitz–Thompson estimator is compared with the kriging predictor combined with systematic sampling. Their overall conclusion is that the kriging estimator is more efficient than the Horvitz–Thompson estimator for large sample size, but it often presents poor confidence interval coverage rates due to the fact that the kriging variance is not a good estimate of the sampling variance. Ver Hoef (2002) compares the kriging predictor with the sample mean in repeated simple random samples drawn from an artificial population. He shows that the kriging predictor is more efficient than the sample mean and gives valid confidence intervals.

We now give a technical sketch of the block kriging predictor for the population mean. Assuming a second order stationary stochastic process, it follows that $E_\xi[y(\mathbf{x})] = \mu$, where $E_\xi[\cdot]$ denotes expectation under the model, and that the covariance between $y(\mathbf{x})$ and $y(\mathbf{x} + \mathbf{h})$ is a function of $\mathbf{h}$ only $C(\mathbf{h})$ (second-order stationarity). The covariance function is generally expressed by parsimonious models, under the assumption of isotropy. An important class of isotropic covariance functions is the Matérn family, which involves a three-parameter vector $\boldsymbol{\theta} = (\sigma^2, \rho, \nu)$, where σ^2 is the variance, ρ is the range parameter (it controls how fast correlation decay with increasing distance) and ν is the smoothing parameter (it controls the smoothness of the resulting interpolating surface). The covariance is rarely known and must be estimated.

The block kriging, which is the (model) best linear unbiased predictor of the population mean, is given by (see Cicchitelli and Montanari 1997; Ver Hoef 2002)

$$\hat{\bar{Y}}_{\mathrm{kr}} = \hat{\mu} + \mathbf{c}_s' \mathbf{V}_s^{-1} \left(\mathbf{y}_s - \mathbf{1}_s \hat{\mu} \right), \tag{9}$$

where $\hat{\mu} = \left(\mathbf{1}_s' \mathbf{V}_s^{-1} \mathbf{1}_s \right)^{-1} \mathbf{1}_s' \mathbf{V}_s^{-1} \mathbf{y}_s$ is the weighted least squares estimator of μ; $\mathbf{1}$ is the unit vector of proper size; $\mathbf{V}_s$ is the $n \times n$ dimensional matrix whose entries are the covariances between sample locations; $\mathbf{c}_s$ is the n-dimensional vector whose i-th entry is given by

$$c_i = \frac{1}{|A|} \int_A C(\mathbf{x} - \mathbf{x}_i) d\mathbf{x},$$

i.e., the mean covariance between $\mathbf{x}_i$ and any other location $\mathbf{x}$ in A.

The prediction variance is given by

$$V_\xi \left(\hat{\bar{Y}}_{\mathrm{kr}} \right) = \sigma_{A,A}^2 - \mathbf{c}_s' \mathbf{V}_s^{-1} \mathbf{c}_s + d^2 \left(\mathbf{1}_s' \mathbf{V}_s^{-1} \mathbf{1}_s \right)^{-1}, \tag{10}$$

where $V_\xi(\cdot)$ denotes variance under the model, $\sigma_{A,A}^2 = |A|^{-2} \int_A \int_A C(\mathbf{x} - \mathbf{x}') d\mathbf{x} d\mathbf{x}'$, and $d = 1 - \mathbf{1}' \mathbf{V}_s^{-1} \mathbf{c}_s$. An estimate of this variance is obtained replacing the integral in (10) with numerical evaluations.

The properties of block kriging are conditional on the selected sample. Hence, provided that the model holds true, averaging across samples we get that $E_p E_\xi \left(\hat{\bar{Y}}_{\mathrm{kr}} - \bar{Y} \right) = 0$ and $E_p E_\xi \left[\left(\hat{\bar{Y}}_{\mathrm{kr}} - \bar{Y} \right)^2 \right] = E_p \left[V_\xi \left(\hat{\bar{Y}}_{\mathrm{kr}} \right) \right]$ attains the minimum among all linear predictors that are jointly unbiased.

Of course, we are interested in design-based procedures since they are model free, in the sense that the validity of inference based on them does not depend on model assumptions.

4 Simulation Results

A simulation study aimed at comparing estimator $\hat{\bar{Y}}_{\mathrm{spl}}$ with the block kriging predictor $\hat{\bar{Y}}_{\mathrm{kr}}$ has been carried out. We considered several artificial populations on the unit square with different degrees of smoothness. For the sake of brevity, we report only on two of them, depicted in Fig. 1 and given by the following functions:

Population A

$$y(\mathbf{x}) = 13.333 \times \left\{ 5[\sin(x_1)]^2 + 5[\cos(x_2)]^2 + 5x_1 \right\}, \quad 0 < x_1 < 1, \ 0 < x_2 < 1.$$

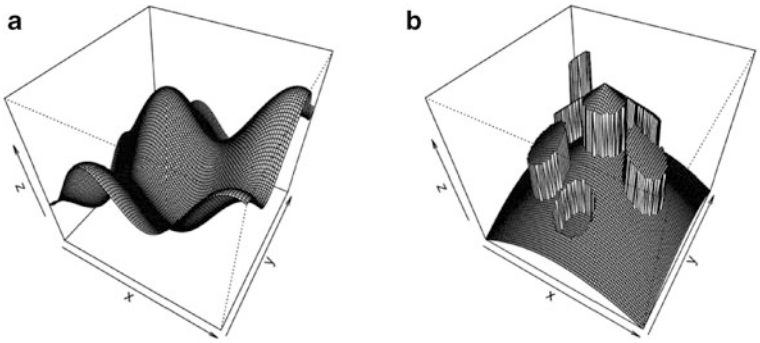

Fig. 1 Surface $y(\mathbf{x})$ of population A (**a**) and population B (**b**)

Population B

$$y(\mathbf{x}) = 335.96 \times \Big[h_{0.5,0.5,0.5,1}(\mathbf{x}) + g_{0.3,0.3,0.1}(\mathbf{x}) - g_{0.5,0.2,0.1}(\mathbf{x}) + g_{0.5,0.5,0.5,1}(\mathbf{x}) + g_{0.3,0.7,0.1}(\mathbf{x}) + g_{0.6,0.7,0.1}(\mathbf{x}) + g_{0.25,0.6,0.1}(\mathbf{x}) + g_{0.75,0.45,0.1}(\mathbf{x}) - g_{0.7,0.6,0.1}(\mathbf{x})\Big], \quad 0 < x_1 < 1, \;\; 0 < x_2 < 1,$$

where

$$g_{a,b,r}(\mathbf{x}) = \frac{1}{2}\cos\left\{\frac{\pi}{2}\left[(x_1 - a)^2 + (x_2 - b)^2\right]\right\} I_{B_r(a,b)}(\mathbf{x})$$
$$h_{a,b,r,q}(\mathbf{x}) = 2^{1/q}\left\{r^q - \left[(x_1 - a)^2 + (x_2 - b)^2\right]^{1/q}\right\},$$

and $I_{B_r(a,b)}(\mathbf{x})$ is the indicator function of belonging to the disk of radius r and centre (a, b).

AQ1 The population mean of both populations is equal to 100.0. The variances are 20,545 and 1,728, respectively, for population A and population B. The corresponding coefficients of variation are 143.3 and 41.6 %, respectively.

Three sampling strategies on the unit square are considered: (a) uniform random sampling, (b) stratified uniform random sampling with two units per stratum, (c) stratified uniform random sampling with one unit per stratum. The sample size is fixed to be $n = 100$ and $n = 256$. The strata are built superimposing a grid of 10×10 smaller squares for $n = 100$ and 16×16 for $n = 256$. In the case (b), pairs of adjacent strata are collapsed along the rows to obtain 128 strata overall. In the case (c), the strata are considered collapsed in the same way for variance estimation only.

One thousand random samples are selected for each sampling strategy and sample size and for each of them estimators $\hat{\bar{Y}}_{\text{spl}}$ and $\hat{\bar{Y}}_{\text{kr}}$ are computed (for the sake of

brevity, we put $\hat{\bar{Y}}_1 = \hat{\bar{Y}}_{\text{spl}}$ and $\hat{\bar{Y}}_2 = \hat{\bar{Y}}_{\text{kr}}$). Monte Carlo estimates of the following quantities are at last obtained:

– Per cent bias with respect to the mean

$$R = \left[E_{\text{MC}}\left(\hat{\bar{Y}}_i\right) - \bar{Y} \right] / \bar{Y}, \quad i = 1, 2.$$

– Relative efficiency (with respect to the sample mean, $\overline{y}$)

$$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right) = \frac{\text{MSE}_{\text{MC}}(\bar{y})}{\text{MSE}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)}, \quad i = 1, 2.$$

– Variance to mean square error ratio

$$R_{\text{var}/\text{mse}} = \frac{E_{\text{MC}}\left[\hat{V}\left(\hat{\bar{Y}}_i\right)\right]}{\text{MSE}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)}, \quad i = 1, 2.$$

– Confidence interval width and coverage for the 95 % nominal level based on the normal approximation.

The number of knots appearing in model (1) is fixed to be equal to the sample size and their locations are established using the algorithm of Nychka et al. (1998).

In each sample, the value of $\hat{\bar{Y}}_{\text{spl}}$ is computed with λ_s, obtained by minimizing the design-based CV criterion (8). The variance of $\hat{\bar{Y}}_{\text{spl}}$ is estimated with (8), apart when uniform random sampling is employed. In the latter case, it is common to compensate for the estimation of the residual $\hat{e}_{(-i)}(\mathbf{x}_i)$, adjusting for the degrees of freedom. Inspired by this argument, we adopt as variance estimator for uniform random sampling the following version

$$\hat{V}_{\text{CV}}^{m}\left(\hat{\bar{Y}}_{\text{spl}}; \lambda\right) = \frac{n-1}{n - r\,(\lambda)} \hat{V}_{\text{CV}}\left(\hat{\bar{Y}}_{\text{spl}}; \lambda\right),$$

where $r(\lambda)$ is the number of degrees of freedom corresponding to the chosen value of λ.

The kriging predictor is computed assuming as covariance function the following isotropic exponential model

$$C(\mathbf{h}) = \begin{cases} \theta_1 + \theta_2, & \mathbf{h} = \mathbf{0}, \\ \theta_2 \exp\left(-\,||\,\mathbf{h}\,||\,/\theta_3\right), & \mathbf{h} \neq \mathbf{0}. \end{cases}$$

Its parameters are estimated for each sample by means of the restricted maximum likelihood (RML).

The main results of the simulation are shown in Table 1 for uniform random sampling, in Table 2 for the two units per stratum design, and in Table 3 for the one unit per stratum design.

Table 1 Performances of estimators under uniform random sampling

	Population A		Population B	
Estimator	$\hat{\bar{Y}}_{\text{spl}}$	$\hat{\bar{Y}}_{\text{kr}}$	$\hat{\bar{Y}}_{\text{spl}}$	$\hat{\bar{Y}}_{\text{kr}}$
	$n = 100$			
Average of r	73.6	–	53.7	–
Bias ratio (%)	−0.02	−0.03	−0.15	0.92
$\text{MSE}\left(\hat{\bar{Y}}_i\right)$	0.2138	0.6651	43.84	40.28
$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)$	78.0	25.1	4.66	5.08
$R_{\text{var/mse}}$	1.82	1.88	1.59	1.07
Average width	2.37	4.38	32.5	25.6
Coverage (%)	96.3	99.0	97.0	94.7
	$n = 256$			
Average of r	179.0	–	108.3	–
Bias ratio (%)	−0.05	−0.02	−0.02	0.08
$\text{MSE}\left(\hat{\bar{Y}}_i\right)$	0.0080	0.0285	9.298	9.208
$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)$	835.7	235.2	8.89	8.98
$R_{\text{var/mse}}$	1.77	4.60	1.22	0.98
Average width	0.46	1.42	13.22	11.73
Coverage (%)	98.0	100.0	95.5	95.5

Table 2 Performances of estimators under two units per stratum design

	Population A		Population B	
Estimator	$\hat{\bar{Y}}_{\text{spl}}$	$\hat{\bar{Y}}_{\text{kr}}$	$\hat{\bar{Y}}_{\text{spl}}$	$\hat{\bar{Y}}_{\text{kr}}$
	$n = 100$			
Average of r	78.9	–	29.0	–
Bias ratio (%)	0.06	−0.04	0.11	0.80
$\text{MSE}\left(\hat{\bar{Y}}_i\right)$	0.0576	0.1229	28.95	21.80
$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)$	29.9	17.6	1.47	1.96
$R_{\text{var/mse}}$	1.60	7.07	1.11	1.40
Average width	1.30	3.65	21.8	21.5
Coverage (%)	96.9	100.0	93.0	96.8
	$n = 256$			
Average of r	186.1	–	70.2	–
Bias ratio (%)	−0.01	−0.01	−0.19	−0.02
$\text{MSE}\left(\hat{\bar{Y}}_i\right)$	0.0033	0.0081	5.964	5.756
$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)$	101.9	41.5	1.99	2.06
$R_{\text{var/mse}}$	1.54	11.1	1.29	1.06
Average width	0.275	1.17	10.8	9.6
Coverage (%)	96.8	100.0	95.5	93.0

Table 3 Performances of estimators under one unit per stratum design

	Population A		Population B	
Estimator	$\hat{\bar{Y}}_{\text{spl}}$	$\hat{\bar{Y}}_{\text{kr}}$	$\hat{\bar{Y}}_{\text{spl}}$	$\hat{\bar{Y}}_{\text{kr}}$
	$n = 100$			
Average of r	78.6	–	32.2	–
Bias ratio (%)	−0.01	0.00	−0.08	0.19
$\text{MSE}\left(\hat{\bar{Y}}_i\right)$	0.0389	0.0475	17.29	15.04
$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)$	25.2	20.6	1.39	1.59
$R_{\text{var/mse}}$	3.30	14.32	1.93	1.73
Average width	1.374	3.234	22.47	19.94
Coverage (%)	99.8	100.0	98.3	98.3
	$n = 256$			
Average of r	187.2	–	111.6	–
Bias ratio (%)	0.00	0.01	0.01	0.34
$\text{MSE}\left(\hat{\bar{Y}}_i\right)$	0.0031	0.0077	5.248	5.092
$\text{Eff}_{\text{MC}}\left(\hat{\bar{Y}}_i\right)$	121.6	49.0	1.85	1.91
$R_{\text{var/mse}}$	1.23	12.35	1.26	1.28
Average width	0.238	1.208	10.05	10.01
Coverage (%)	94.6	100.0	96.5	98.0

In the tables we report the across samples average number of the degrees of freedom; the per cent relative bias; the ratio between the mean squared error of the sample mean and that of the estimators $\hat{\bar{Y}}_{\text{spl}}$ and $\hat{\bar{Y}}_{\text{kr}}$; the ratio between the average of variance estimates and the mean squared error of the estimators; the average width of the confidence intervals; the 95 % confidence interval per cent coverage.

In all cases, the two estimators behave as nearly unbiased; the relative bias is always below 0.5 %, apart for the kriging predictor in the uniform random sampling of size 100.

Looking at the MSE of the estimators, for population A, which is more regular than population B, the efficiency of $\hat{\bar{Y}}_{\text{spl}}$ is much higher than that of $\hat{\bar{Y}}_{\text{kr}}$, and both estimators are far more efficient than the sample mean. The efficiency with respect to the sample mean increases with the sample size and decreases passing to more efficient sampling design, as the stratified ones. In fact, the sample mean estimator interpolates the response variable with the stratum sample mean and some of the efficiency is captured by the design. However, the working models upon which PSRE and the kriging predictor are based provide some extra efficiency with respect to the sampling design. In the case of population B, which is sharper, $\hat{\bar{Y}}_{\text{kr}}$ is somewhat more efficient than $\hat{\bar{Y}}_{\text{spl}}$. Note that the higher the degree of smoothness of the response variable, the higher is the number of degrees of freedom selected for $\hat{\bar{Y}}_{\text{spl}}$.

The high number of degrees of freedom selected for PSRE prevents the use of standard variance estimator (7), which would be dramatically downward biased. The alternatives based on the "leave-one-out" residuals explored in this simulation are generally positively biased and overestimate the MSE of the estimator. But, the variance estimator of $\hat{\bar{Y}}_{\mathrm{kr}}$ can be several times bigger than the MSE for population A, revealing a possible inadequacy of the second order stationary model. As a consequence, confidence intervals based on the normal approximation are conservative and unnecessarily large. In the case of one unit per stratum design, which is not a measurable design, the collapsed strata technique for estimating the variance results in a severely positively biased estimator and this is particularly clear in the case of population A and sample size equal to 100.

In summary, the efficiency of the penalized spline regression estimator is close to that of the kriging and much higher in the case of population A, when the number of degrees of freedom of the spline regression model is large. In all cases, valid, even if conservative, confidence intervals are obtained with $\hat{\bar{Y}}_{\mathrm{spl}}$.

5 Conclusions

This paper starts from the results of Cicchitelli and Montanari (2012), where a model-assisted penalized spline regression estimator (PSRE) of the mean of a spatial population is introduced and is proved that this estimator is design-consistent and approximately unbiased. The main idea is to capture the spatial pattern in the data by assuming that the target population comes from a super-population described by a penalized spline regression model. In this approach, the role of the model in strengthening the efficiency of the design-based estimation is similar to that of the covariance function in the block kriging prediction procedure.

In the above paper, the smoothing parameter that determines the behaviour of the estimator is treated as a fixed quantity and the issue of how to best select its value is not addressed. Another open problem is the poor performance of the standard design-consistent variance estimator when the number of degrees of freedom of the smoother is relatively high with respect to the sample size.

The two issues are addressed in this paper making use of the "leave-one-out" cross-validation criterion. We assign to the tuning parameter λ the value that minimizes the sampling variance computed with the "leave-one-out" residuals. This solution follows heuristically from Opsomer and Miller (2005), where a design-based cross-validation criterion is proposed for bandwidth selection for the local polynomial regression estimator.

With regard to PSRE, the evidence coming from the simulation study presented in this article can be summarized as follows.

In uniform random sampling and in stratified random sampling with two units per stratum, the cross-validation criterion for selecting the smoothing parameter and the variance estimator used here seems to work quite well. This procedure provides

valid, even if conservative, confidence intervals. PSRE strongly outperforms block kriging predictor when the response variable is given by a smooth surface, while kriging seems to work a little better when the population is given by a sharp surface. In the first situation, the number of degrees of freedom of PSRE tends to be very high.

In stratified sampling with one observation per stratum, PSRE presents, as expected, a mean squared error which continues to be considerably smaller than that obtained with samples of the same size drawn with simple random sampling. But the estimator of variance, based on collapsing contiguous pairs of strata, may seriously overestimate the true variance.

As an overall assessment, we think that the penalized spline regression estimator performs well or better than block kriging predictor, in contrast to the findings of other studies where design-based estimation strategies and block kriging are compared (see, for example, Brus and de Gruijter 1997; Ver Hoef 2002). This is due to the fact that in the former, differently from what is done in the above papers, the spatial structure of the population is conveyed, at the estimation stage, into the estimator through the spline regression assisting model. In other words, it can be said that sampling theory and geostatistics are reconciled, thanks to the duality between first-order and second-order modelling in geostatistics, as mentioned in Shaddick and Zidek (2012).

We must point out that the technique for selecting the smoothing parameter and for estimating the variance mimics the procedure proposed by Opsomer and Miller (2005), in a finite population sampling context, for bandwidth selection in local polynomial regression estimation; in the latter important regularity conditions were imposed on the sampling method and on the population. Our results rely mainly on empirical evidences and need to be supported by appropriate theoretical results. In particular, it has to be established if the design consistency of the estimator proved in Cicchitelli and Montanari (2012) for a fixed value of the smoothing parameter continues to hold in the new context, where a data-driven technique is used for selecting it.

A second open problem is the search for an alternative estimator of the variance in the case of stratified sampling with one observation per stratum, which appears to be the most efficient way to draw samples from a spatial population, as various papers have shown (Barabesi et al. 2012; Stevens 1997).

Of course, the approach can be generalized in several ways. For example, further covariates may be added to the assisting model, parametrically and non-parametrically. In such a case, the penalized spline part of the model accounts for the spatial covariance between residuals from the added covariates. It can be also shown that model (1) can be written as a mixed model and standard techniques in that field might be useful for estimation and inference.

Another area for future research is the use of the function $\hat{y}(\mathbf{x})$, given in (2), as a design-based estimator of the population surface $y(\mathbf{x})$ at point $\mathbf{x}$. Assuming that the function $y(\mathbf{x})$ is continuous (or integrable), as is done within environmental monitoring, forestry, geology, etc. (see, for example, Gregoire and Valentine 2008; Barabesi et al. 2012), the consistency of $\hat{y}(\mathbf{x})$ can be conjectured on the basis of

the available asymptotic results concerning the spline regression function. These results, coupled with an appropriate design-based variance estimator measuring the degree of uncertainty of the estimate, could allow the use of $\hat{y}(\mathbf{x})$ to estimate the population value $y(\mathbf{x})$ and to draw maps, as the kriging predictor does in the model-based context.

References

Barabesi, L., Franceschi, S., Marcheselli, M.: Properties of design-based estimation under stratified spatial sampling with application to canopy coverage estimation. Ann. Appl. Stat. **6**, 210–228 (2012)

Brus, D.: Using regression models in design-based estimation of spatial means of soil properties. Eur. J. Soil Sci. **51**, 159–172 (2000)

Brus, D., de Gruijter, J.: Random sampling or geostatistical modeling? Choosing between design-based and model-based strategies for soil (with discussion). Geoderma **80**, 1–59 (1997)

Cicchitelli, G., Montanari, G.E.: The kriging predictor for spatial finite population inference. Metron **55**, 41–57 (1997)

Cicchitelli, G., Montanari, G.E.: Design-based estimation of a spatial population mean. Int. Stat. Rev. **80**, 111–126 (2012)

Cordy, C.B.: An extension of the Horvitz–Thompson theorem to point sampling from a continuous universe. Stat. Probab. Lett. **18**, 353–362 (1993)

Gregoire, T.G., Valentine, H.T.: Sampling Strategies for Natural Resources and the Environment. Chapman & Hall, New York (2008)

Hastie, T.J., Tibshirani, R.J.: Generalized Additive Models. Chapman & Hall, New York (1990)

McArthur, R.D.: An evaluation of sample designs for estimating a locally concentrated pollutant. Commun. Stat. Simulat. Comput. **16**, 735–759 (1987)

Nychka, D., Haaland, P., O'Connel, M., Ellner, S.: FUNFITS. Data analysis and statistical tools for estimating functions. In: Nychka, D., Piegorsch, W.W., Cox, L.H. (eds.) Case Study in Environmental Statistics, Lecture Notes in Statistics, vol. 132, pp. 159–179. Springer, New York (1998)

Opsomer, J.D., Miller, C.P.: Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. Nonparametric Stat. **17**, 593–611 (2005)

Ruppert, D., Wand, M.P., Carroll, R.J.: Semiparametric Regression. Cambridge University Press, Cambridge (2003)

Särndal, C.E., Svensson, B., Wretman, J.: Model Assisted Survey Sampling. Springer, New York (1992)

Shaddick, G., Zidek, J.V.: Unbiasing estimates from preferentially sampled spatial data. Technical report nr. 268, Department of Statistics, The University of British Columbia (2012)

Stevens, D.L.: Variable density grid-based sampling designs for continuous spatial populations. Environmetrics **8**, 167–195 (1997)

Stevens, D.L., Jensen, S.F.: Sample design, execution, and analysis for wetland assessment. Wetlands **27**(3), 515–523 (2007)

Stevens, D.L., Olsen, A.R.: Spatially balanced sampling of natural resources. J. Am. Stat. Assoc. **99**, 262–278 (2004)

Ver Hoef, J.: Sampling and geostatistics for spatial data. Ecoscience **9**(2), 152–161 (2002)

AUTHOR QUERY

AQ1. 100,0 has been changed to 100.0. Please confirm.