Nonparametric estimation of the small-scale variability of heterocedastic spatial process

R. Fernández Casal[1], S. Castillo-Páez[2], and Pilar García Soidán[2]
[1] University of A Coruña (Spain); ruben.fcasal@udc.es
[2] University of Vigo (Spain); scastillo@alumnos.uvigo.es, pgarcia@uvigo.es

Abstract

The current study aims to provide nonparametric estimators of the conditional variance and the dependence structure of a heteroscedastic spatial process. When assuming zero mean along the observation region, the approximation of the variance can be addressed by linear smoothing of the squared observations. Then, using the estimated standard deviations, the variogram can be estimated from the standardized data. Under the presence of a non-zero deterministic trend, we suggest a modification of the latter method, using the residuals obtained from a local linear estimation of the trend, jointly with corrections of the biases due to the use of these residuals. This work also includes the results of numerical studies, carried out to illustrate the performance of our proposals in both scenarios.

Introduction

- Suppose that the spatial process $\{Y(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^d\}$ can be modeled as:

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + \sigma(\mathbf{x})\varepsilon(\mathbf{x}), \tag{1}$$

where $\mu(\cdot)$ and $\sigma^2(\cdot) > 0$ denote the deterministic trend and variance functions, respectively, and ε is a second-order stationary process, with zero mean, unit variance and correlogram $Cov(\varepsilon(\mathbf{x}), \varepsilon(\mathbf{x}+\mathbf{u})) = \rho(\mathbf{u})$. Although, the characterization of the spatial dependence is usually done through the semivariogram $\gamma(\mathbf{u}) = \frac{1}{2}Var(\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}+\mathbf{u})) = 1 - \rho(\mathbf{u})$.

- In this framework, given $\mathbf{Y} = (Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))^t$, the main goal is the nonparametric estimation of the characteristics of the model (1), i.e. obtain estimates $\hat{\mu}(\mathbf{x})$, $\hat{\sigma}(\mathbf{x})$ and $\hat{\gamma}(\mathbf{u})$. We will focus on local linear estimation because it possesses many asymptotically optimal propierties, such as the reduced influence of the boundary effects.

- The specification of the small-scale variability of the process requires the estimation of the conditional variance and the correlogram (or variogram) of the error process, as the nonstationary-covariogram of the process is $Cov(Y(\mathbf{x}), Y(\mathbf{x}+\mathbf{u})) = \sigma(\mathbf{x})\sigma(\mathbf{x}+\mathbf{u})\rho(\mathbf{u})$. This results in:

$$\mathbf{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^{\boldsymbol{t}} \odot \mathbf{R}, \tag{2}$$

where $\boldsymbol{\sigma} = (\sigma(\mathbf{x}_1), \ldots, \sigma(\mathbf{x}_n))^t$, where $\mathbf{\Sigma}$ and $\mathbf{R}$ denote the covariance matrices corresponding to $\mathbf{Y}$ and $\varepsilon = (\varepsilon(\mathbf{x}_1), \ldots, \varepsilon(\mathbf{x}_n))^t$, respectively, and $\odot$ represents the Hadamard product. The variogram of the heterocedastic spatial process is given by:

$$\gamma_{\mathbf{x}}(\mathbf{u}) = (\sigma(\mathbf{x}) - \sigma(\mathbf{x}+\mathbf{u}))^2 + 2\sigma(\mathbf{x})\sigma(\mathbf{x}+\mathbf{u})\gamma(\mathbf{u}) \tag{3}$$

- From a nonparametric perspective, several procedures can be applied to obtain $\hat{\sigma}^2(\mathbf{x})$, although the most popular ones are based on differences or on squared residuals, as proposed in [5] or [2], respectively. Studies in the one-dimensional setting, under the assumption of correlated errors, shows some advantages of the second method over the former one (see e.g. [7]).

- The estimation of the variogram is also usually done through the residuals. However, the direct use of residuals introduces biases due to the estimation of the trend, both in variogram and variance estimation (see e.g. [3] and [6], respectively).

- In this work, we propose some modifications at the squared-residual approach to correct for these biases. The performance of the algorithm developed for the joint estimation of trend, variance and variogram functions is analized by numerical studies in different spatial configurations.

Nonparametric estimation for heterocedastic spatial processes

- The local linear trend estimator, obtained by linear smoothing of the $(\mathbf{x}_i, Y(\mathbf{x}_i))$, can be written explicitly as :

$$\hat{\mu}_{\mathbf{H}}(\mathbf{x}) = \mathbf{e}_1^t \left(\mathbf{X}_{\mathbf{x}}^t \mathbf{W}_{\mathbf{x}} \mathbf{X}_{\mathbf{x}}\right)^{-1} \mathbf{X}_{\mathbf{x}}^t \mathbf{W}_{\mathbf{x}} \mathbf{Y} = s_{\mathbf{x}}^t \mathbf{Y}, \tag{4}$$

where $\mathbf{e}_1$ is a vector with 1 in the first entry and all other entries 0, $\mathbf{X}_{\mathbf{x}}$ is a matrix with i-th row equal to $\left(1, (\mathbf{x}_i - \mathbf{x})^t\right)$, $\mathbf{W}_{\mathbf{x}} = diag\left\{K_{\mathbf{H}}(\mathbf{x}_1 - \mathbf{x}), \ldots, K_{\mathbf{H}}(\mathbf{x}_n - \mathbf{x})\right\}$, $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1}K(\mathbf{H}^{-1}\mathbf{u})$, K is a d-dimensional kernel function and $\mathbf{H}$ represents the bandwidth matrix.

- The matrix $\mathbf{H}$ controls the shape and size of the local neighborhood used to estimate $\mu(\mathbf{x})$. We recommend the use of the "bias corrected and estimated" generalized cross-validation criterion (CGCV) proposed in [4] to select this bandwidth in practice.

- The natural approach to obtain $\hat{\sigma}(\mathbf{x})$ and $\hat{\gamma}(\mathbf{u})$ consists in removing the trend and estimating the variance and the variogram from the residuals $\mathbf{r} = \mathbf{Y} - \mathbf{SY}$, where $\mathbf{S}$ is the *smoother matrix*, whose ith row is equal to $\mathbf{s}_{\mathbf{x}_i}^t$.

- Nevertheless, it is well-known that the direct use of the residuals may produce a strong underestimation of the small-scale variability of the process (e.g. [1], Section 3.4.3). Simply note that:

$$Var(\mathbf{r}) = \mathbf{\Sigma} + \mathbf{S\Sigma S}^t - \mathbf{\Sigma S}^t - \mathbf{S\Sigma}.$$

- The approach proposed in this work, which tries to correct these biases, is based on the following approximation:

$$Var(\mathbf{r}) \approx \boldsymbol{\sigma}\boldsymbol{\sigma}^{\boldsymbol{t}} \odot (\mathbf{R} + \mathbf{B}), \tag{5}$$

where $\mathbf{B} = \mathbf{SRS}^t - \mathbf{RS}^t - \mathbf{SR}$ (note that (5) holds exactly for homocedastic processes). Using this approximation, it is easy to see that:

$$Var\left(r_i/\sqrt{1+b_{ii}}\right) \approx \sigma^2(\mathbf{x}_i),$$

$$Var\left(\tilde{\varepsilon}(\mathbf{x}_i) - \tilde{\varepsilon}(\mathbf{x}_j)\right) \approx Var\left(\varepsilon(\mathbf{x}_i) - \varepsilon(\mathbf{x}_j)\right) + b_{ii} + b_{jj} - 2b_{ij},$$

where b_{ij} is the (i,j)-th element of the matrix $\mathbf{B}$ and $\tilde{\varepsilon}(\mathbf{x}_i) = r(\mathbf{x}_i)/\sigma(\mathbf{x}_i)$.

Estimation algorithm

0. Using (4) to estimate the trend, compute the residuals and obtain a prior estimate of $\mathbf{R}$ (for instance, $\hat{\mathbf{R}} = \mathbf{I}$).
1. Set $\hat{\mathbf{B}} = \mathbf{S}\hat{\mathbf{R}}\mathbf{S}^t - \hat{\mathbf{R}}\mathbf{S}^t - \mathbf{S}\hat{\mathbf{R}}$.
2. Estimate $\hat{\boldsymbol{\sigma}}^2 = (\hat{\sigma}^2(\mathbf{x}_1), \ldots, \hat{\sigma}^2(\mathbf{x}_n))$ by linear smoothing of $\left(\mathbf{x}_i, r_i^2/(1+\hat{b}_{ii})\right)$.
3. Compute $\hat{\varepsilon}(\mathbf{x}_i) = r_i/\hat{\sigma}(\mathbf{x}_i)$ and estimate the variogram $2\gamma_{\hat{\varepsilon}}(u)$ by linear smoothing of $\left(||\mathbf{x}_i - \mathbf{x}_j||, (\hat{\varepsilon}(\mathbf{x}_i) - \hat{\varepsilon}(\mathbf{x}_j))^2 - \hat{b}_{ii} - \hat{b}_{jj} + 2\hat{b}_{ij}\right)$.
4. Estimate $\sigma_{\hat{\varepsilon}}^2$ from $\hat{\gamma}_{\hat{\varepsilon}}(u)$ and compute $\hat{\gamma}(u) = \hat{\gamma}_{\hat{\varepsilon}}(u)/\hat{\sigma}_{\hat{\varepsilon}}^2$.
5. Obtain $\hat{\mathbf{R}}$ from $\hat{\gamma}(u)$ and repeat steps 1-4 until convergence.

- A similar correction for variance estimation was suggested in [6] for independent data. These authors propose an estimator $\hat{\boldsymbol{\sigma}}^2 = S_2\mathbf{r}/(\mathbf{1} + S_2 diag(\mathbf{B}))$ where S_2 is the corresponding linear smoothing matrix.

- The correction of the squared differences residuals for variogram estimation in step 3 is similar to that introduced in [3] for the homocedastic case.

- For the case $\mu(\mathbf{x}) = 0$, the previous algorithm could be applied taking into account that steps 0 and 1 are not necessary, so $\hat{\mathbf{B}} = 0$ and $r_i^2 = Y^2(\mathbf{x}_i)$.

Simulation results

- $N = 1,000$ samples of different sizes n were generated following model (1) on a regular grid in the unit square, with three mean function: $\mu_1(x_1, x_2) = 0$ (null), $\mu_2(x_1, x_2) = 5.8(x_1 - x_2 + x_2^2)$ (linear), $\mu_3(x_1, x_2) = \sin(2\pi x_1) + 4(x_2 - 0.5)^2$ (non linear); also we consider the following variance functions: $\sigma_1(x_1, x_2) = 1$ (constant), $\sigma_2(x_1, x_2) = 0.5(1 + x_1 - x_2)$(linear).

- The random errors ε_i are normally distributed with zero mean, unit variance and isotropic exponential covariogram:

$$\gamma_\theta(\mathbf{u}) = c_0 + c_1\left(1 - \exp\left(-3\frac{\|\mathbf{u}\|}{r}\right)\right),$$

(for $\mathbf{u} \neq \mathbf{0}$), where c_0 is the nugget effect, c_1 is the partial sill ($c_1 = 1 - c_0$) and r is the practical range.

- The values considered in the simulations were: $n = 10 \times 10$, 15×15, 20×20, $r = 0.3, 0.6, 0.9$, and $c_0 = 0, 0.2, 0.40.8$.

- First, to analize the performance of the proposed procedure without the effect of the trend bias, the estimation of the variogram was calculated using trend null and $\sigma_2(x_1, x_2)$. For instance, Table 1 compare the squared errors using the squared-residual approach and those obtained with the algorithm proposed, considering $n = 20 \times 20$, and $c_0 = 0.2$.

Table 1. Summary of squared relative errors of the estimated variogram with trend null.

	$r = 0.3$			$r = 0.6$			$r = 0.9$		
	mean	median	sd	mean	median	sd	mean	median	sd
Sq. Residuals	0.014	0.003	0.103	0.179	0.014	2.431	1.992	0.041	43.059
Corrected	0.009	0.002	0.018	0.020	0.005	0.042	0.042	0.011	0.087

- The full procedure was applied using $\mu_2(\cdot,\cdot)$, $\mu_3(\cdot,\cdot)$, $\sigma_1(\cdot,\cdot)$ and $\sigma_2(\cdot,\cdot)$. In general, the mean of the estimation of the variance was better with the algorithm proposed, in comparison with the squared-residuals estimates. Figure 1 confirms this results, in terms of relative errors using non linear trend, linear variance and $n = 20 \times 20$, $r = 0.6$ and $c_0 = 0.2$.

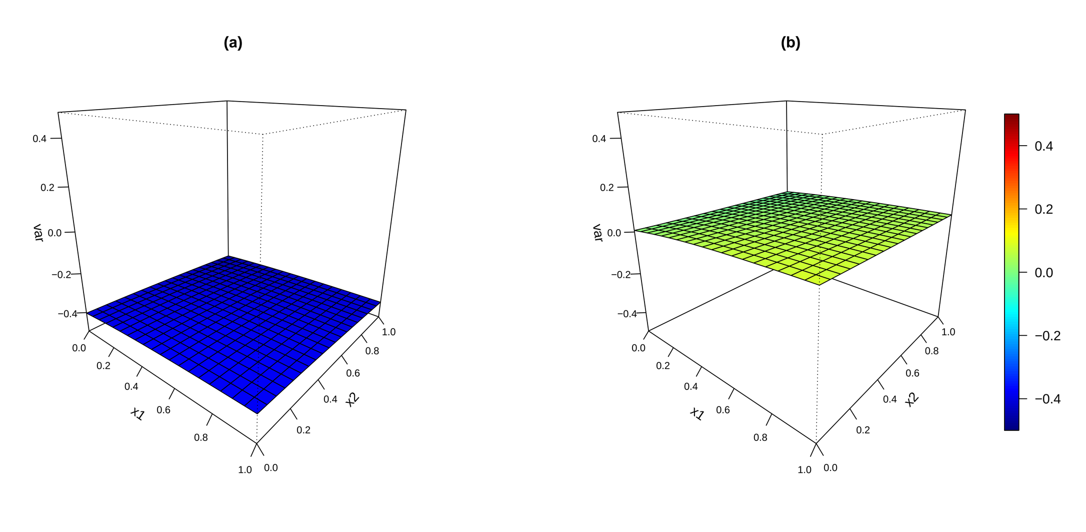

Figure 1. Average of the relative errors for variance estimates whit (a) the squared-residuals approach, and with (b) the algorithm proposed.

- The variogram estimation with the algorithm was compared with those calculated by the residuals obtained by removing the trend and the variance estimated by squared residuals. Once again, the results presents more fitting errors with the procedure proposed, as is shown at Figure 2.

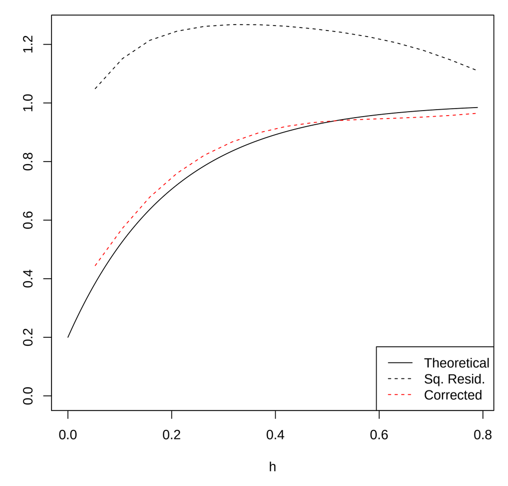

Figure 2. Comparison of the means of variogram estimation with $\mu_3(\cdot,\cdot)$, $\sigma_2(\cdot,\cdot)$, $n = 20 \times 20$, $r = 0.6$ and $c_0 = 0.2$.

- The means of squared relative errors for variance and variogram estimates, for $\mu_3(\cdot,\cdot)$, $\sigma_2(\cdot,\cdot)$, $r = 0.6$ and $c_0 = 0.2$ is shown in Table 2, where it is observed the good performance of the proposed algorithm in comparison whit the squared-residuals procedure. A similar behavior was observed in the other simulation settings.

Table 2. Means of squared relative errors for the estimated variance and variogram.

	Variance estimate			Variogram estimate		
	10×10	15×15	20×20	10×10	15×15	20×20
Sq. Residuals	0.212	0.193	0.183	0.392	0.308	0.287
Corrected	0.152	0.090	0.085	0.007	0.006	0.006

- For an overall measure of the behavior of the algorithm proposed, the theoretical variogram of the heterocedastic spatial process (3) was compared with its estimated value. The figure 3 shows the good performance of $\hat{\gamma}_{\mathbf{x}}(\mathbf{u})$ taking $\mathbf{x} = (0.5, 0.5)$. This results are confirmed by the squared relative errors calculated in the Table 3.

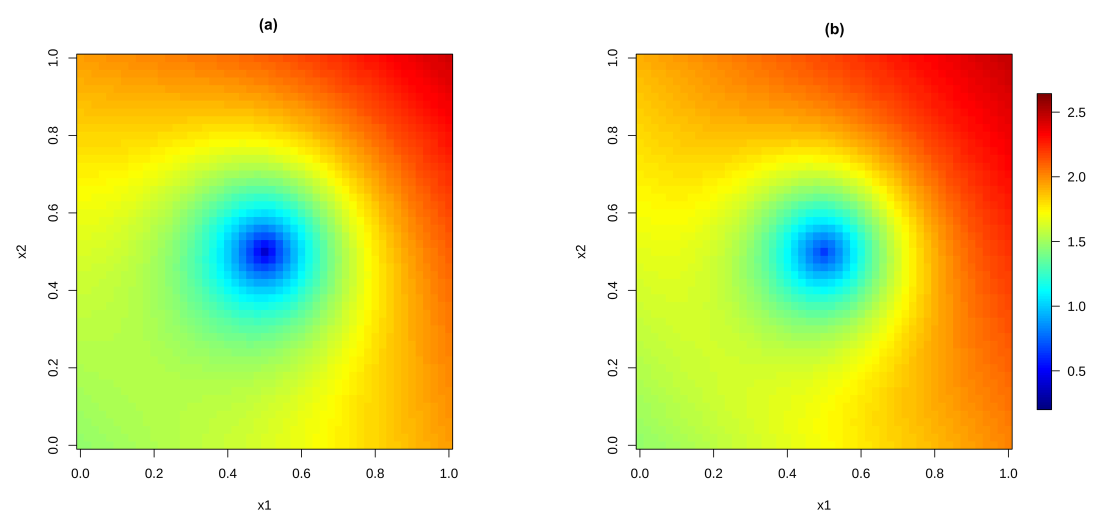

Figure 3. (a) Theoretical and (b) estimated semivariogram of the heterocedastic process, with $\mathbf{x} = (0.5, 0.5)$, $\mu_3(\cdot,\cdot)$, $\sigma_2(\cdot,\cdot)$, $r = 0.6$ and $c_0 = 0.2$

Table 3. Means of squared relative errors for the estimated heterocedastic variogram with the squared-residual approach and the algorithm proposed.

Sq. residuals			Corrected		
mean	median	sd	mean	median	sd
0.060	0.050	0.065	0.007	0.006	0.007

Conclusions

- In general, the algorithm proposed has a better performance than the squared residual approach to estimate the variance function, under the configurations considered. Also, these results are more accurate specially with correlated data.

- The use of nonparametric estimators to approximate the small-scale variability allows to obtain more flexible estimators and to avoid problems due to misspecification model.

- Finally, note that a bias-corrected pilot estimate of the variogram is obtained at the end of the proposed algorithm. Then, a valid variogram can be fitted to these pilot estimates from the model selected by the user. The final variogram can be used to derive a new bandwidth and obtain a new trend estimate.

References

[1] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

[2] Fan, J.Q. and Yao, Q.W. (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **2**, 645–660.

[3] Fernández-Casal, R. and Francisco-Fernández, M. (2014). Nonparametric bias-corrected variogram estimation under non-constant trend. *Stochastic Environmental Research and Risk Assessment* **28(5)**, 1247–1259.

[4] Francisco-Fernández, M. and Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *The Canadian Journal of Statistics* **33**, 279–295.

[5] Härdle, W. and Tsybakov, A. (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics* **81**, 223–242

[6] Ruppert, D., Wand, M. P., Holst, U and Hösjer, O. (1997) Local Polynomial Variance-Function Estimation, *Technometrics*, **39:3**, 262–273

[7] Vilar-Fernández, J. M. and Francisco-Fernández, M. (2006) Nonparametric estimation of the conditional variance function with correlated errors. *Journal of Nonparametric Statistics* **18(4-6)** 375–391