Cokriging versus univariate interpolation methods: An application to the premises market

Montero-Lorenzo, José-María and Larraz-Iribas, Beatriz
University of Castilla-La Mancha, Spain
jose.mlorenzo@uclm.es; Beatriz.Larraz@uclm.es

Primary aim

In this paper four geostatistical methods are compared, including two spatial methodologies, kriging and cokriging, estimating premises prices.

The aim is to deal with some statistics methods that are able to estimate premises prices. First of all, it is essential a method that incorporates the role of the space into conventional estimates (taking into account the spatial dependence that the stochastic process present). Second, the method should be able to be helped by another process, an auxiliary one, more sampled than the main one, and strongly correlated with it, because of the scarce sample available. And, of course, this method should take into account the features of the premises.

The cokriging methodology is suggested as it fits very well all these constraints (considering premises prices as the main process and house prices as the auxiliary or secondary process). The kriging estimates of the premises prices have been also computed, it means, as if we have not the house prices at our disposal, but we want to estimate still taking into account the spatial correlation among premises prices. We have compared both methods in terms of the variance of the estimation error.

To compare how the consideration of the spatial structure improves the estimates, two classical interpolation methods that do not consider it have been also computed, as Inverse Distance Weighting (IDW) and 2-D shape functions.

Statistical methodology

Theoretical considerations for Ordinary Cokriging (OCK):

Consider $\mathbf{X} = (X_1, X_2, \ldots, X_m)'$ a vector of intrinsic random functions, that is

$E[X_i(\mathbf{s}+\mathbf{h}) - X_i(\mathbf{s})] = 0, \forall i = 1, \ldots, m$ and

$$\mathrm{cov}[X_i(\mathbf{s}+\mathbf{h}) - X_i(\mathbf{s}), X_j(\mathbf{s}+\mathbf{h}) - X_j(\mathbf{s})] = 2\gamma_{ij}(\mathbf{h}), \quad \forall i, j = 1, \ldots, m$$

being **s** the locations where information is taken and **h** a vector linking any two points **s** and **s** + **h** in the domain under study. Each variable is defined on a set of samples with sizes n_1, n_2, …, n_m respectively. Under these conditions, the OCK estimator of a particular variable at the point $\mathbf{s}_0$ is a weighted linear combination of the data values from the variables X_j ($j=1,\ldots,m$) located at sampled points in the neighborhood of $\mathbf{s}_0$.

$$X_i^*(\mathbf{s}_0) = \sum_{j=1}^{m} \sum_{\alpha=1}^{n_j} \lambda_\alpha^j X_j(\mathbf{s}_\alpha^j)$$

The weights are calculated to ensure that the estimator is optimal, in the sense that it is unbiased and minimum error-variance.

OCK system is obtained by minimizing that variance with the constraint on weights

$$\begin{cases} \sum_{k=1}^{m} \sum_{\beta=1}^{n_k} \lambda_\beta^k \gamma_{jk}(\mathbf{s}_\alpha^j - \mathbf{s}_\beta^k) + \omega_j = \gamma_{ji}(\mathbf{s}_\alpha^j - \mathbf{s}_0) & \forall j = 1, \ldots, m; \quad \forall \alpha = 1, \ldots, n_j \\ \sum_{\alpha=1}^{n_j} \lambda_\alpha^j = \delta_{ij} = \begin{cases} 1 & si \ i = j \\ 0 & si \ i \neq j \end{cases} & \end{cases}$$

$$V[X_i^*(\mathbf{s}_0) - X_i(\mathbf{s}_0)] = \sum_{j=1}^{m} \sum_{\alpha=1}^{n_j} \lambda_\alpha^j \gamma_{ji}(\mathbf{s}_\alpha^j - \mathbf{s}_0) + \omega_i$$

Case study: Cokriging estimation of premises prices in the Historic City of Toledo

The database contains information about 223 houses and 123 premises sited in the historical part of Toledo city. The information we have worked with has been provided by the real estate agencies that operate in this historical area and it is referred to market values. The data were collected in the third quarter of 2004 and the computer packages we have used are the ISATIS and R language.

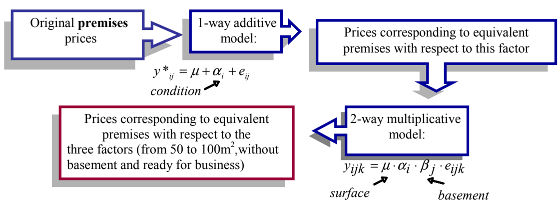
Obtaining equivalent classes

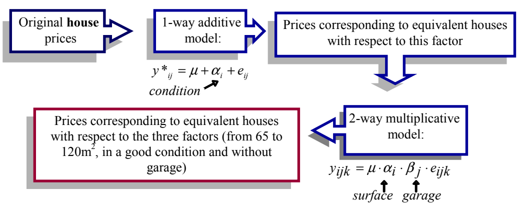
The information or the original database refers to the market price, location, condition and surface. Additionally, it is known whether the premises have basement or not, and, in the case of houses, if they have parking or not.

	Condition	Surface	Basement
Premises Factors and levels	Ready for business	Less than 50 m²	Yes
	Some renovation needed	From 50 to 100 m²	Not
	Complete renovation needed	From 100 to 200 m²	
	Unfinished	More than 200 m²	

	Condition	Surface	Parking
Houses Factors and levels	New or completely renovated	Less than 65 m²	Yes
	In a good condition	From 65 to 120 m²	Not
	Little renovation needed	More than 120 m²	
	Complete renovation needed		

In order to isolate the spatial component of premises and house prices we have proceeded to adjust for housing and premises mix by the analysis of variance method.





Testing for spatial correlation and variogram modeling

We have tested randomness versus positive correlation using a contiguity matrix whose elements are the inverses of the distances among locations.

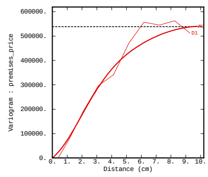
	Prices	
	Premises	Houses
Moran´s *I* statistic	0,022	0,125
E(*I*)	-0,0413	-0,0083
V(*I*)	0,00015	0,00032
Standardized value	5,2348	7,4429

Rejection, at the 5% level of significance, of the null hypothesis of randomness in favour of spatial positive autocorrelation

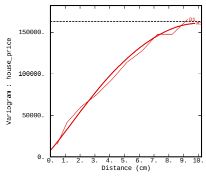
Model of coregionalization:
119 robust premises prices out of 123 and 219 robust house prices out of 223

$$(\rho = 0{,}696)$$

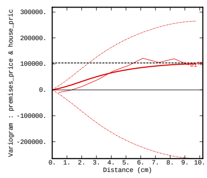
	Sill		
Model	Premises prices direct variogram (a)	House prices direct variogram (b)	Premises prices-House prices cross variogram ©
Spherical-330m range	340.978,332	142.783,006	70.505,189
Nugget effect	1	8.000	-85
Gaussian-165m range	200.000	10.000	30.000



(*a*)

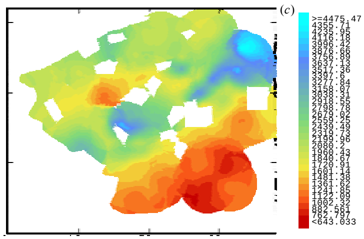


(*b*)

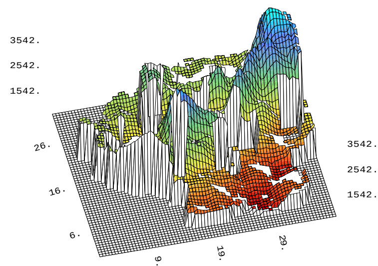


(*c*)

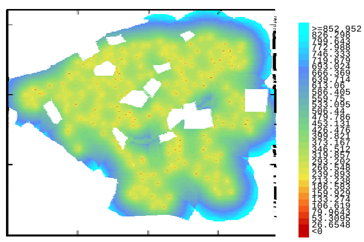
Results



Estimation map



Estimation surface



Standard deviation map

Cokriging versus univariate interpolation methods

	Error		Standarized Error	
Interpolation method	**Mean**	**Variance**	**Mean**	**Variance**
IDW (p=1)	-30,39	106.312,420		
IDW (p=2)	-18,24	97.779,020		
2D-shape functions	-2,390	91.694,344		
Ordinary Kriging (OK)	-1,672	91.186,233	-0,015	1,097
Ordinary Cokriging (OCK)	1,5011	80.621,447	0,0045	0,922

Conclussions

Studying the spatial structure:

1. Adjusting for housing and premises mix in order to isolate the spatial component of premises and house prices.
2. Testing for spatial correlation and identifying a global pattern of that spatial correlation.
3. Modelling the direct and cross-variograms according to the linear model of coregionalization to ensure a positive definite model.

We have used three univariate interpolation methods (OK, IDW and 2-D shape functions) to shown the advantages of OK procedure.

Unfortunately, in the premises prices case, the scarceness of the data leads to a variance of the estimation error bigger than desirable.

The existing correlation between the prices of different types of properties (in our case, houses and premises) has been used to obtain more accurate estimates

To know the price of premises anywhere the citizens want to buy or sell them, cokriging methodology has been proposed (multivariate).

Finally, OK improves the classical interpolation methods that do not take account of the spatial structure among property prices (IDW and 2-D shape functions) and OCK has a clear advantage over OK in comparison with estimation using only the data of the target variable.