Geostatistical Study of the Rural Property Market Applicable to the Region of Murcia (Spain) by M. Carmen Morillo1 et al

M. Carmen Morillo · Francisco García-Cepeda · Sandra Martínez-Cuevas · Iñigo Molina · Cesar García-Aranda

Abstract Spatial Analysis has been used since the early 1990's for rural cadastral models. Nowadays, it is necessary to use Geographic Information Systems (GIS), Global Positioning Systems (GPS) and Cartography as well as dedicated software in order to solve problems in Cadastral Models. Among these models, the selection and prioritization of the parameters affecting the rural property valuation assets, as well as the zoning optimization (homogeneous sectors) within a community, are one of the biggest concerns for Cadastral Agencies. This research is focused on rural property in the Region of Murcia, in accordance with data recorded from 2007 to 2009. The techniques used for the spatial zoning were carried out by means of IDW (Inverse Distance Weighting) and kriging interpolators as they are the most widely used in spatial variation analysis studies. Although both interpolators produced similar results, IDW was better for predicting variation in "Unit Prices". Afterwards, a statistical study was later completed using ANOVA, Chi-Squared and Correspondence Analysis procedures. The research results demonstrate the consistency of the parameters that have been used in the Cadastral Agencies, and the degree of involvement of these parameters in assessing the results of unit prices for different areas. These results are secured by the verification techniques applied.

Keywords Geographic information system (GIS) · Spatial interpolation · IDW · Kriging · Multivariate statistical · Cadastre

Introduction

There has been growing interest in the statistical analysis of georeferenced over time and with the development of software. In addition, there is no doubt that this interest will continue to grow given that there are more and more and better quality technological developments capable of generating spatially referenced databases. The interest of government authorities and private companies in this type of information increases each day and the range of fields of knowledge using georeferenced information is expanding (Montero and Larraz 2008). One of the fields of knowledge highly pertinent at this time where this type of statistical analysis of georeferenced data is useful is related to the price of rural property.

One task that can be statistically carried out is verifying whether the parameters used in a certain model are correlated or not, if there is a certain level of reliability, etc. If the model used is also spatial involving certain characteristics of a specific territory, the geostatistical analysis must be taken into consideration along with Geographic Information Systems.

There are many studies, such as the one by Goodchild and Haining (2004) which reflect the state of the art of geographic information systems and data analysis. Other studies, such as Grimes and Liang (2009), introduce spatial determinants of land prices finding a strong zoning boundary effect on land prices.

This study includes both aspects. On the one hand, there are data, pursuant to a particular model, that are used to value rural plots; and on the other hand, an analysis is done to determine whether the model created based on such data is coherent statistically.

The study area is the region of Murcia for which more than 4500 plot points are available with information such as the transactional value, type of crop, altitude, surface area, quality, etc. Additionally, GIS are used to obtain the distances between the centroids of the plots and the adjacent plots within certain ranges as well as from the same plots to the existing structures, which are at specific distances.

The relationship of location to the economic value of land has been heavily studied for some time by analysts like Cantillon (1755); Smith (1776) and Ricardo (1817), considering the “spatial effects” in the economy; however, it was Thünen (Cournot 1992) who created a “location model” of economic activity; in other words, a complete theory of the spatial formation process for territorial valuation based on the differential location of the lands. For the agricultural valuation, various authors used the distance to urban centres as the variable for the value, Caballer 2008; Ruiz Rochera 1861 and Juárez 1952.

Even though it is not the main objective of this paper, it is important to quickly indicate the current situation of rural cadastral valuation in Spain.

Thus, the cadastral value of rural property is obtained by capitalising the value of the 1989 tax payable and then updated each year as established in the Spanish Government Budgets. The Law sets forth that a rural plot valuation procedure must be used with the market as the reference.

The valuation currently done by the DGC (Directorate General of the Cadastre) with respect to rural property is based on the following parameters:

- Surface area, measured in hectares.
- Cadastral classification, determined by the crop or use attributed to each plot. If the classification and production intensity are not homogeneous for the entire cadastral plot, it is divided into sub-plots.
- Production intensity, the land is classified by the production intensity and aptitude for agricultural production as per the prior cadastral classification.
- Evaluation rate, this represents the theoretical yield expressed in euros per hectare of land of a certain classification and production intensity.
- Theoretical yield of the sub-plot, the result of multiplying the corresponding evaluation rate by the surface area.
- Theoretical yield of the plot, the result of multiplying the corresponding evaluation rate by the surface area or, where applicable, the sum of the theoretical yields of the sub-plots comprising it.
- 1989 tax base, which is the theoretical yield of the plot for that year.
- 1989 tax payable, which is 50 % of the tax base for that year.
- Cadastral value, the 1989 tax payable capitalised at 3 % and multiplied by the corresponding update coefficients established in the Budget Act.

The time gap, for which there is an attempt to palliate with update coefficients, has resulted in distortions between the different types of crop, for example those that were valued above the real value in the 1980's where the demand is currently quite well below the value assigned and vice versa, crops which were low in value more than 30 years ago which nowadays have a value above the one reflected in the cadastral value due to the scarce offer. Some articles in 2004 already indicated a need to carry out this obligation set forth by law (Alcázar and Ariza 2004).

Nonetheless, it is believed that the particular characteristics of rural property must be considered in order to establish the market value (Caballer 1994): at times, it is open and at times, it is intervened, not homogeneous or transparent. Other factors that affect the price of rural property (Ballestero 1992) are the local and inter-regional nature. Chiara et al. 2015, using a model based on artificial neural networks as a weighting method and spatial correlation as a clustering method identify the major risk of rural property (agricultural areas), those related to urban pressure, visualized by geographical information systems map.

Thus, it is very important to choose the most adequate parameters for the valuation in order to obtain a cadastral value for a rural property where the market value is referential.

The goal of this research is to identify, classify and visualise the rural properties in four zones. To do so, data recorded between 2007 and 2009 in a specific region in the Southeast of Spain were used. Among others, the most important parameters taken into consideration in this study were price, income, type of crop, altitude of the plot, the degree of the nearest urban development, etc. The research done makes it clear that there is a significant association between most of the parameters depending on the zone meaning the data identified by the Cadastre are coherent and significant in this Region.

This work is part of a research project carried out at the Department of Topographic Engineering and Cartography in collaboration with the Cadastre General Directorate of the Spanish Ministry for Finance and Public Administration. It aims to optimise the valuation of rural property assets as well as prioritise the key parameters involved in these assessments.

Materials and Methods

A variety of data and methods were used for this study; they are outlined in detail in this section. In relation to the data used, they were obtained from a specific geographic area of Southeast Spain. The study area is first described and then the data and methods used.

Study Area

The study area belongs to the Region of Murcia (Fig. 1), which covers a surface area of 11,313 km2 and is comprised of forty-five municipalities with a population of nearly 1,5 million. The total number of plots recorded in the Region of Murcia is exactly 467,062 plots, which represents the 0,9 % of the total Spanish State, excluding Navarra and Basque Country. In this study, the sampling strategy is based on a simple random process.

The interest in this zone is due to multiple contrasts: dry-farming land and irrigated land, plains and mountain areas, coastal and inner land, vineyards and plateau. Thus,

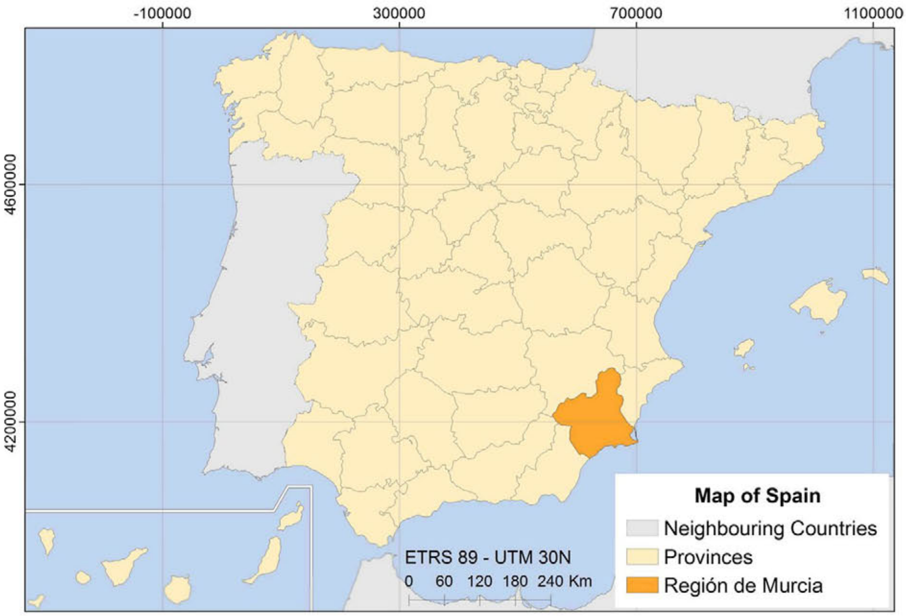

Fig. 1 Region of Murcia (Spain)

the information obtained essentially consisted of the price per m^2, the income, type of crop, altitude of the plot, degree of nearest urban development, etc. This information was recorded in a database with the attributes indicated in the following section.

Database

A database had to be built with the foregoing information, which is considered most adequate for the objective sought, in order to apply spatial analysis procedures. This information is organised in the Table 1. Thus, all of the plots in these municipalities were stored and represented using a Geographic Information System. Figure 2 shows the centroids of the 4573 plots finally recorded.

The first aim of the database is to study the feasibility of the spatial classification in four zones ("Zones"): Price zones that are Very High (MA), High (A), Average (M) and Low (B) for the variable "Unit price" in the province of Murcia. Interpolations were used defining the differentiating characteristics of the zones to obtain a homogeneous zone as per the "Unit Prices", using multivariate statistic to do so. The classification was done using the programme ArcGis v10.1, whereas the multivariate statistical procedures were carried out using the programme SPSSv21. The methods used for the analysis operations are outlined in the following section.

Table 1 Database with foregoing information

Abbreviated Name	Label	Description	Type
CadRef	Plot identifier	Type of property registration	Alphanumeric
X	X coordinate	Geographic location. Horizontal coordinate The system used is as follows: Projection system (UTM).	Numerical
Y	Y Coordinate	Geographic location. Vertical coordinate. The system used is as follows: (UTM).	Numerical
R	Income	Income per production per Ha.	Numerical
S	Surface area	Plot surface area m^2.	Numerical
TC	Crop type	Crop Type from 1 to 18.	Category
P100	Plots at 100 m	No. plots at 100 m.	Numerical
P300	Plots between 100 and 300 m	No. plots between 100 and 300 m.	Numerical
P1000	Plots between 300 and 1000 m	No. plots between 300 and 1000 m.	Numerical
C300	Structures at 300 m	No. of structures at 300 m.	Numerical
C3000	Structures at between 300 and 3000 m.	No. of structures at between 300 and 3000 m.	Numerical
C30000	Structures at 30,000 m	No. of structures at between 3000 and 30,000 m.	Numerical
A	Altitude in metres	Vertical distance of a point on the land surface with respect to the sea level.	Numerical
P_U	"Unit Price" in €/m^2	Valuation of the property in € per m^2.	Numerical

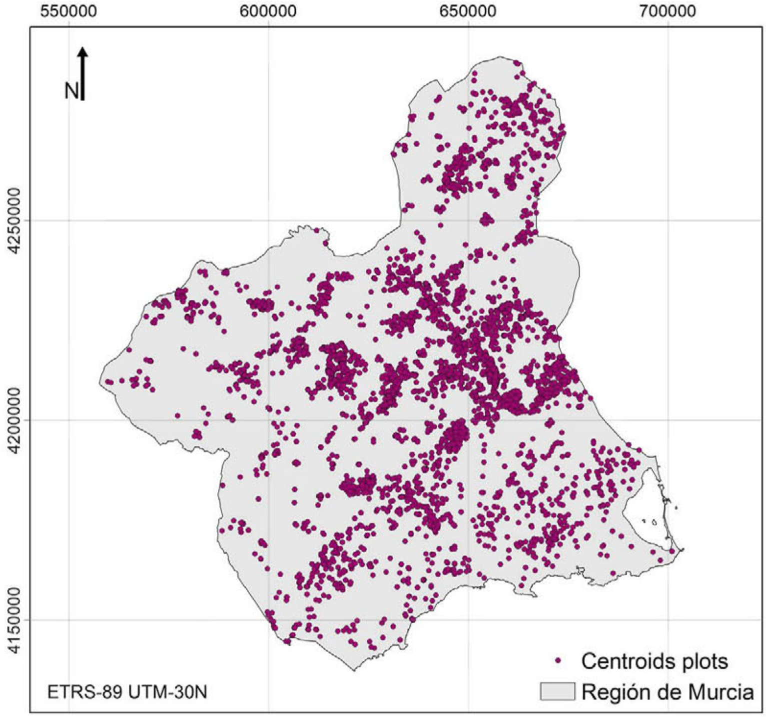

Fig. 2 Representation of the data in UTM projection (WGS84, zone 30)

Methods

Geostatistics is a branch of applied statistics, which specialises in the analysis and modelling of spatial variability in Earth sciences. The methods used for this work to estimate the variable "Unit Price" correspond to geostatistical analysis (Martín and Díaz 2002).

The information stored in the spatial database was essential to the spatial classification into four zones for the "Unit Price" variable, starting with the following preliminary first-range statistics such as mean, range, standard deviation, etc. The values of these parameters in this study are outlined in the results section.

Afterwards, an estimate analysis of the variable "Unit Price" was carried out by means of spatial interpolation in order to achieve a continuous surface area with the estimates to establish the values of the variable indicated where not sampled. Data interpolation offers the benefit of projecting maps or continuous surfaces based on discrete data (Johnston et al. 2001). The accuracy of the map generated based on the characteristics to be studied depends largely on the spatial structure of the data. It is believed that the stronger the spatial correlation of the values of the variable under study, the better quality of the mapping (Kravchenko and Bullock 1999).

One of the interpolators used is called kriging which not only offers the required predictions and response surfaces, but also probability and quantile maps (Johnston et al. 2001). The kriging method quantifies the spatial structure of the data through the use of variograms, sometimes known as semivariograms. The measure of probability, done through kriging methods, makes a difference with respect to the deterministic methods for spatial interpolations. The most commonly used is Inverse Distance Weighting (IDW) (Burrough and McDonnell 1998). The IDW uses a simple algorithm based on distances (Johnston et al. 2001). Both models, ordinary kriging and IDW, assume that the predictions are a linear combination of data, as shown by the Eq. (1), by Villatoro et al. (2008):

$$Z(S_0) = \sum_{i=1}^{n} \lambda_i Z(S_i) \quad i = 1, ..., n \tag{1}$$

where $Z(S_0)$ is the estimated value of the variable under study, "Unit Price" for this paper, at the interpolated point S_0, n is the number of neighbouring observations used for the estimate and λ_i is the weight given to the value observed $Z(S_i)$ near the value S_0 (Lozano et al. 2004). This last parameter is the difference between kriging and IDW. In practice, the mean is not constant throughout the study area but rather varies locally and is also unknown meaning the ordinary kriging value shall be used in such situations.

The experimental semivariogram is obtained from the estimation of the sample data. It indicates the variation of the correlation between these samples based on their corresponding distances. The most common estimation is based on the moments method (Cressie 1990), and is computed using the expression (2):

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} \left(Z(x_i) - Z(x_j)\right)^2 \ , \quad h \in R^d \tag{2}$$

where:

Z(xi)	value of the variable at point xi.
Z(xi)	value of the variable at point xj.
N(h)	{(xi, xj) / xi – xj = h; i, j = 1, …,n}.
|N(h)|	Distinct pair count.

Once the points of the experimental variogram are defined, a theoretical model must be established for them. This is due to the inability to work with an experimental model that lacks a precise mathematical function and to the need to extend the values of the model beyond the maximum defined distance. In general, the available models are applied to isentropic processes since the anisotropy can be treated with geometric transformations.

The frequently used basic models are developed mathematically by constructing a random function and calculating the variogram theoretically. Further on, the state (or condition) of positive conditioned function anisotropy is imposed (Moral García 2003). These basic models are:

Spherical Model

$$\gamma(h) = \begin{cases} \frac{m}{2}\left[3\left(\frac{h}{a}\right) - \left(\frac{h}{a}\right)^3\right] & \text{si } 0 \le h < a \\ m \text{ si } h = a & \end{cases} \tag{3}$$

Exponential Model

$$\gamma(h) = m\left(1 - e^{-\frac{h}{a}}\right) \quad si\ h > 0 \tag{4}$$

where m and n are positive constants, the value of the plateau (m) and the scope (a), respectively.

Gaussian Model

$$\gamma(h) = m\left(1 - e^{-\frac{h^2}{a^2}}\right) \quad si\ h \geq 0 \tag{5}$$

The ordinary kriging method obtains the weights (or influence) of the values, resolving the kriging equation shown in the following expression (6) by Schloeder et al. (2001):

$$\sum_{i=1}^{n} \lambda_i \gamma\ [d\ (S_i, S_j)] + m = \gamma\ [d\ (S_i, S_0)],\ \ i = 1, \ldots n;\ \sum_{i=1}^{n} \lambda_i = 1 \tag{6}$$

where n is the number of observations, m is the Lagrange multiplier used to minimise restrictions, λ is the weight given to each one of the observations and the sum of all the λ is one. The sub-indices i and j denote the points sampled, the sub-index 0 is the point estimated, S symbolises the measurement made (variable measured) and d (S_i, S_0) is the distance between S_i and S_o based on the semivariogram (7):

$$\gamma\ [d\ (S_i, S_0)] = var\ [Z\ (S_i) - Z\ (S_0)] \tag{7}$$

This semi-variance calculated is a measure used to determine the similarity between observations where the more similar, the less semi-variance (Lozano et al. 2004).

On the other hand, the IDW calculates the weight of the values pursuant to the inverse relationship of the distance (Schloeder et al. 2001) with the Eq. (8):

$$\lambda_i = \frac{[d(S_i,\ S_0)]^{-p}}{\sum_{i=1}^{n} [d(\ S_i,\ S_0)]^{-p}},\ \ i = 1, \ldots, n \tag{8}$$

where p is the parameter of the exponent that controls the velocity trend with which the weights of the points tend to be zero. The higher the p, the greater the weight of the nearest points and, as a result, the more continuous or softer surface areas obtained and the predictions tend to fall towards the sample average (Schloeder et al. 2001). The values are usually between 1 and 3 (Gotway et al. 1996).

In order to assess the validity of the results obtained with the procedures proposed, a strategy based on cross validation was designed pursuant to the methods outlined above. Cross validation offers an idea of how well the model predicts the unknown values. For all of the points, cross validation sequentially omits a point and predicts the value using the rest of the values and then compares the value observed with the estimated value. In order for said values to be acceptable, the mean must be close to 0 and the mean square error must be as low as possible. The mean square error is very useful for making comparisons with other methods (Moreno 2007).

Two types of estimates have been applied: the inverse distance as a traditional estimator (IDW) with different powers and as an estimator using the ordinary kriging Geostatistic, using different variograms as done in other studies (Villatoro et al. 2008; Montero and Larraz 2008; Bailey and Gatrell 1995).

Later, a relational statistical study was done to study the particular characteristics of each zone using the procedures ANOVA (Analysis of Variance) for an inter-subject factor, the Chi-Squared test and the contingency coefficient test as well as correspondence analysis.

The role of the ANOVA technique for one factor (One-Way ANOVA) is basically the following: in order to compare the means of the dependent variable (quantitative) Y, associated with the different factor levels (X_1, X_2,..., X_n), an average variation between different levels (MS-factor) was compared with a variation measure within each level (MS-error). If the MS-factor is significantly greater than the MS-error, the conclusion is that the means associated with different factor levels are different (H_1). This means that the factor significantly influences the dependent variable Y. If, on the contrary, the MS-factor is not significantly higher than the MS-error, the null hypothesis (H_0) that all the means associated to different factor levels coincide or are equal will not be rejected (http://www.uoc.edu/in3/emath/docs/ANOVA.pdf). This model is used when the researcher is only interested in the factor levels present in the experiment meaning any variation in the scores would be due to experimental error (Spiegel et al. 2007). In this study, the factors are represented by the variable "Zones" established as per the unit price of the Region of Murcia.

The analysis of the variance must basically comply with the assumptions of independence; in other words, the study cases are different in each one of the factor levels, variance homoscedasticity and the data must be distributed normally, yet if care is taken in managing the same no. of data for each one of the groups or factor levels, the last two assumptions required may be considered covered although not verified (Garza et al. 2013).

A particular type of contrast known as 'post hoc' tests or 'multiple comparisons' must be used to know which mean differs from another. These tests are done when the influence of the factor is detected in the dependent variable. In this study, the Post hoc contrast has been applied in order to identify where the difference or differences are that cause the rejection of H_0 in the first step in ANOVA. The method used to establish that the means of the rent associated with different zones are different is the Games-Howell method which is based on the distribution of the studentised range and in a T statistic where, after estimating the population variances assuming they are different, the degrees of liberty are corrected using the Welch equation (Garza et al. 2013).

The Chi-Squared Independence test is used to analyse the relationship of dependence and/or independence between two qualitative variables. This test makes it possible to determine whether there is a relationship between two categorical (qualitative) variables. It is necessary to highlight that this test indicates whether or not there is a relationship between the variables but it does not indicate the degree or type of relationship; in other words, it does not indicate the percentage of influence of a variable over another or the variable that causes the influence (Tinoco Gómez 2008).

This analysis would indicate whether the variables “Zones” and “Crop type” are related or independent.

Finally, a graph has been created to show the associations between the four zones (“Zones”) and the different crop types (“Crop type”). To do so, a correspondence analysis was done which is a descriptive or exploratory multivariate technique aimed at summarising a large quantity of data in a small number of dimensions with the lowest possible loss of information (Joaristi and Lizasoain 1999). The correspondence analysis only requires that the data are organised into categories. The analysis will be simple or multiple depending on whether two or more qualitative variables exist (De la Fuente Fernández 2011). For this study, such analysis will make it possible to determine the most important crop type in each one of the four study areas. For this study and in the case of the “Correspondences” analysis model, it has been proceeded as follows: based in a two dimensions solution, a Chi-square test has been applied as a measure of distance between the rows and columns of the correlation table (contingency or cross table of qualitative input variables with the marginal totals of rows and columns). As standardization method and in order to center rows and columns, the “average rows and columns values” have been eliminated. Finally, a “Symmetric method” was chosen for the normalization of each dimension. In this case, the row scores are the weighted average of the column scores divided by the singular value and, the column scores are the weighted average of scores row divided by the singular value (Pérez 2009). The result of the application of this model generates a scatter plot between categories of qualitative variables “Crop type” (row) and “Zones” (column), the results are shown in the following Section 3.2 (Fig. 6).

All of these statistical methods and procedures are described and can be consulted in the publication by McKillup and Darby (2010).

Results

Analysis Models

In accordance with the strategy outlined above and based on the information contained in the database, Table 2 shows the baseline statistical data for the study sample with respect to the variable “Unit Price”.

The second group of results correspond to the application of the interpolation procedures using three different weights for the variable “Average unit prices”, these are summarised in Table 3. Likewise, this Table shows the results of the cross validation applied to the three interpolations done. The neighbours considered will be 10 and a minimum of 2 will be included for each sector. 4 sectors per zone are used. The search method is the ellipse since anisotropy shall be considered.

For this interpolation method, the IDW ($p = 1$) is observed as best suited as the mean error is closer to 0 and also the Root Mean Square Error has the lowest value.

Firstly, the variogram that best fit with the data was chosen for the ordinary kriging study. The results are shown in the following Table 4.

The nugget indicates the behaviour near the origin, making clear the similarity of very close data; in other words, it reflects the continuity or regularity of the variable on a small scale. This indicates that the variable “Unit Price” is irregular at short distances.

Table 2 Preliminary statistics of the variable "Unit price"

Parameters	Unit price
Samples	4573 units
Minimum	0.2 €/m^2
Maximum	570.38 €/m^2
Mean	11.72 €/m^2
Range	570.18 €/m^2
Standard dev.	30.28 €/m^2
Variation coeff.	258.36 %
Mean standard error	0.45€/m^2
Table T (95 %)	1.96 (t-Student value table)
Lower limit	10.84€/m^2
Upper limit	12.60€/m^2

The value of the range (1500 m) indicates that up to that distance the points are related and the "partial sill" value shows the maximum mean value of the difference with the square of the "Unit Price" values between two points separated by a distance of 1500, after which the "partial sill" value is constant and the value is 529.37€/m^2.

To verify that close distances (NEAR_FID & NEAR_DIST) to the "Unit Price" (P_U) are irregular, a verification was done searching for the closest point at each point along with the "Unit Price" value to reflect the partial result in Table 5.

The observation is that the value of the "Unit Price" for a distance of less than 17 m varies from 1.66 to 190.20€ /m^2. Another observation from the semi-variogram raster model (Fig. 3) is the existence of a geometric or elliptical anisotropy in the data which means the mesetas of at least two directional semi-variograms reach different distances even though they are equal, as indicated in Krivoruchko 2011.

Kriging interpolation was then done using the three variogram models (Table 6).

Table 3 IDW Interpolation of the variable "Unit price"

	Average unit prices		
Characteristics	IDW(p = 1)	IDW(p = 2)	IDW(p = 3)
Angle	23	23	23
Higher semi-axis	15,000	15,000	15,000
Lower semi-axis	6790	6790	6790
Maximum neighbours	10	10	10
Minimum neighbours	2	2	2
Sector type	4	4	4
Cross validation:			
Mean error	0.16	0.3	0.33
Root Mean Square Error (RMSE)	27.74	29.63	31.66

Table 4 Definition of the variograms

Variograms			
Data	Spherical	Exponential	Gaussian
Maximum neighbours	10	10	10
Minimum neighbours	2	2	2
Sector	4	4	4
Anisotropy:			
Direction	21.27	22.85	23.38
Higher range	1500	1500	1500
Lower range	741	768.79	678.53
Nugget	426.73	419.76	498.52
Partial sill	572.79	567.05	529.37
Lag No.	10	10	10
Lag size	150	150	150

The variogram that is most suitable is the Gaussian as the mean error is closest to zero and the Root Mean Square Standardized Error is close to 1, although the differences between them are minimal.

Of all the interpolations done, the IDW estimate was chosen as it has the smallest Root Mean Square Error (27.74€/m^2), yet insisting that the differences between all of the estimates are quite small. The results of the interpolations done are shown in Fig. 4.a and b.

Table 5 The relationship between "Unit Price" (P_U), no. of the closest sample (NEAR_FID) and the closest distance (NEAR_DIST)

NEAR_FID	NEAR_DIST(m)	P_U (€/m2)
4374	10.44	23.21
244	10.44	190.20
788	12.21	17.89
449	12.21	24.94
2104	13.04	4.13
935	13.04	3.11
4532	15.23	25.33
4454	15.23	25.93
696	15.56	2.90
211	15.56	1.66
3272	16.12	8.29
2995	16.12	11.08
4532	16,16	24.58
2934	16.40	101.25
2098	16.40	46.91
4466	16,49	25.77

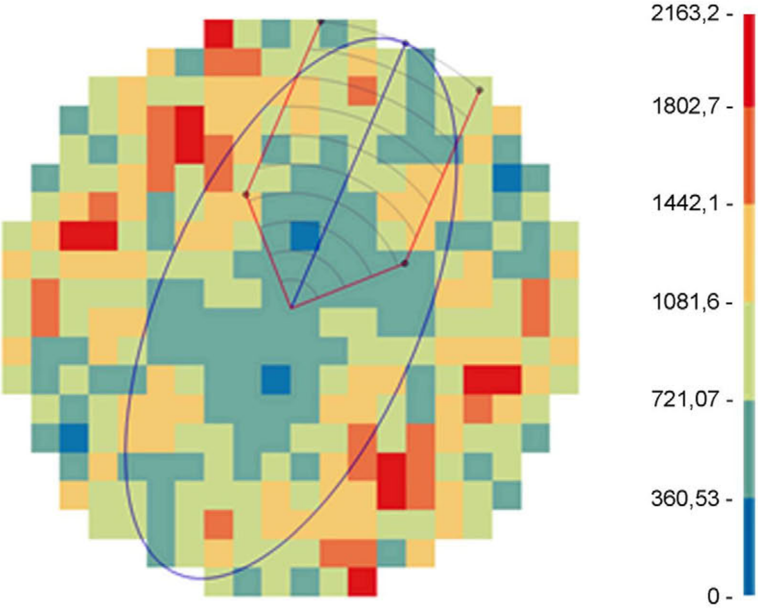

Fig. 3 Gaussian semivariogram raster

As it can be appreciated in Fig. 4, there are no major differences between the maps derived from both IDW and kriging interpolation methods. However, as a result of the applied algorithm, some differences are distinguishable for the range of values used in such methods. In the IDW procedure is applied weighted means, using as weight factor the inverse of the distance raised to an exponent. In the ordinary kriging, the linear combination weights minimize the estimated error variance. Subsequently, these ranges differ and are lower for kriging.

Once the IDW interpolation was chosen ($P = 1$), the quantiles were classified into four intervals. The classification by quartiles was used to obtain the size of the balanced raster zones and the values are transferred from the cells to points to obtain the following values (Table 7).

Table 6 Cross validation of ordinary kriging as per different Variograms

	Ordinary kriging		
	Spherical	Exponential	Gaussian
Cross validation			
Samples	4573	4573	4573
Mean Errors	0.064	0.035	0.00059
Root mean Square Error	28.29	28.09	28.48
Standardised mean	0.0012	0.00022	0.00063
Root mean Square Standardized Error	0.996	0.9634	1.0089
Average Standard Error	29.005	29.58	28.91

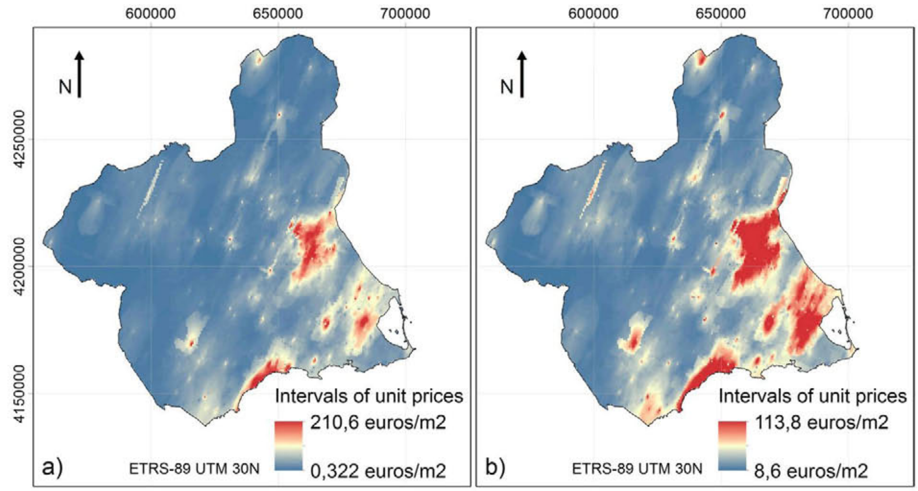

Fig. 4 Interpolations done, (**a**) Representation of the map for IDW Interpolation ($p = 1$), (**b**) Representation of the map for Ordinary kriging interpolation

The classification was done as per the working criteria (spatial, structural and zoning) of Morillo et al. 2016 in which the Spline method is used.

The result of the classification was as follows: there were 16 lost data which were eliminated from the study (Fig. 5). The no. of samples for the classification of each zone is specified in Table 8.

Statistical Contrasts and Correspondence Analysis

To study the inherent characteristics (profitability, altitude, no. of structures at different distances...) for each zone and the assumptions of dependence and/or dependence between two qualitative variables, a strategy was used based on five "considerations" that have been verified by applying a statistical or contrast test. All of them were done with a confidence level of 95 %.

Thus, the first consideration formulated refers to the possible influence of the "Unit Price" on the "Income" variable in the classification or zoning factor. In order to respond to this problem, the ANOVA contrast was used meaning 773 random samples were selected in zones 2, 3, and 4 in order to have the same number of samples in each zone so said procedure could be applied.

The application of the ANOVA procedure showed that there is a statistically significant effect of income on the "Unit Price" category. The contrast F statistic follows a *Snedecor F* distribution where the degree of liberty is 3 in the denominator

Table 7 Intervals of the "Unit Price" values per zone

Zones	Intervals of mean unit prices (€/M^2)	Qualitative value
1	<2.4	Low
2	[2.4 , 5.83)	Average
3	[5.83 , 11)	High
4	> = 11	Very High

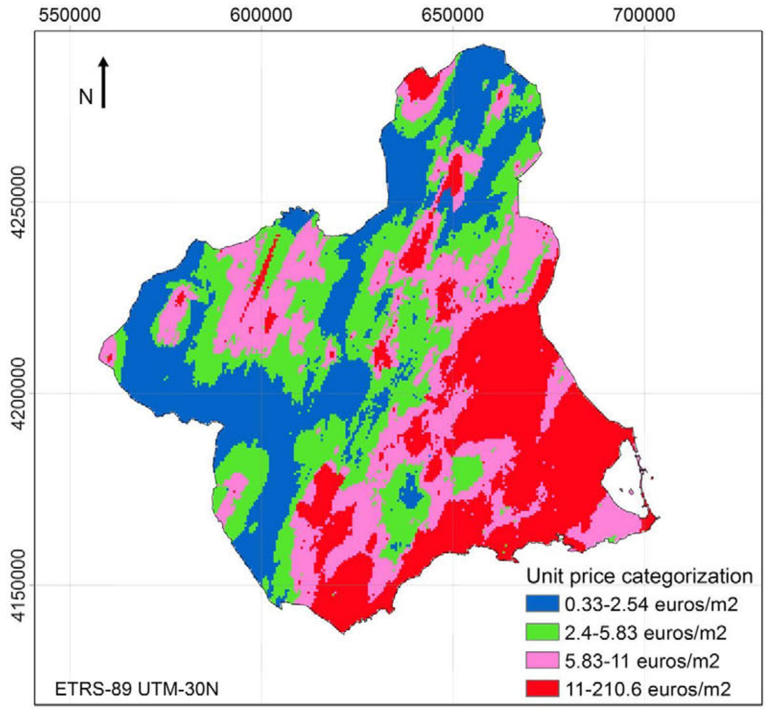

Fig. 5 Classification of the "Unit Price" in the region of Murcia

and the degrees of liberty 3088 in the denominator for a resulting value of 275.60 for this contrast ($F(3, 3088) = 275.60$), with a significant level of $p < 0.05$. As mentioned in the section of methods, the Post-Hoc contrast test was applied using the Games-Howel method. The results of this test are summarised in Table 9.

In view of these results (Table 9), it can be confirmed that there is an association between the variable "Unit Price" of each zone and the variable "Valuation of the income declared" recorded in the Cadastre.

The following problem that arose in this study consists of verifying whether there is some type of association between the crop type and the variable of the category "Unit Price". In this case, the Chi-square test was done to answer this question. The result is described with the following statements:

– The requirement of Pearson's Chi-square test is fulfilled, as the number of slots with an expected relative frequency less than 5 is lower than 20 %.

Table 8 Number of samples per zone

Zones	Frequency	Percentage
Zone 1	773	17.0
Zone 2	1049	23.0
Zone 3	1346	29.5
Zone 4	1389	30.5
Total	4557	100.0

Table 9 Games-Howel test

[Zone I , Zone J]	Statistical (T_Student)	P_value
[Zone 1, Zone 2]	13.44	<0.05
[Zone 1, Zone 3]	23.67	<0.05
[Zone 1, Zone 4]	30.10	<0.05
[Zone 2, Zone 3]	9.79	<0.05
[Zone 2, Zone 4]	14.70	<0.05
[Zone 3, Zone 4]	4.42	<0.05

- There is a statistically significant association between the crop type and the classification of the "Unit Price"χ^2 (30) = 1183.8, $p < 0.05$.
- A statistically significant relationship was found which is high and directly proportional (Contingency coefficient = 0.45, maximum value of C = 0.75). Since there is a relationship or association that is statistically significant, there is an association between one of the categories of the crop type variable and one of the categories of the zones variable. A correspondence analysis is done to find the most important crop type in each one of the "Unit Price" categories (zones) which is significantly more important with respect to the other crops. This is an exploratory technique applied to the variable "Unit Price" and "Crop type". This analysis results in a dispersion diagram (Fig. 6) which makes it possible to see the possible associations between both variables.

At this point, the Correspondence Analysis described in the section of methods, which makes it possible in this study to estimate the most important crop type in each zone was done. The result of this analysis is described in Table 10.

Additionally, the Chi-Squared test is done to confirm the results of the correspondence analysis. To do so, the crop type and zone variables had to be dichotomised. The results are outlined in Table 11.

Given these values, the fact that in the four cases there are significant differences between the zones and the corresponding crop can be affirmed. Moreover, an index equivalent to the ODDS ratio (Fleis 1981) was done to calculate the advantage between the different crop types and the different zones. This index or reason for the advantages between the crop type associated to a zone with respect to the other zones is obtained with the ratio between the number of times the crop type associated to the study zone is reflected versus the no. of times the crop type associated to the study zone is reflected in the other zones, divided by the ratio between the number of times the crop type associated to the study zone is not reflected versus the no. of times the crop type associated to the study zone is not reflected in the other zones. Thus, said index determines or makes it possible to determine the advantage of a certain crop with respect to the others.

The third question that has arisen in the application of the spatial analysis methods is to determine whether there is a significant influence between the altitude of the land and the "Unit Price" in each zone given that the relief and topography in the study zone is rather accidental. Thus, the ANOVA procedure was applied for a factor where the

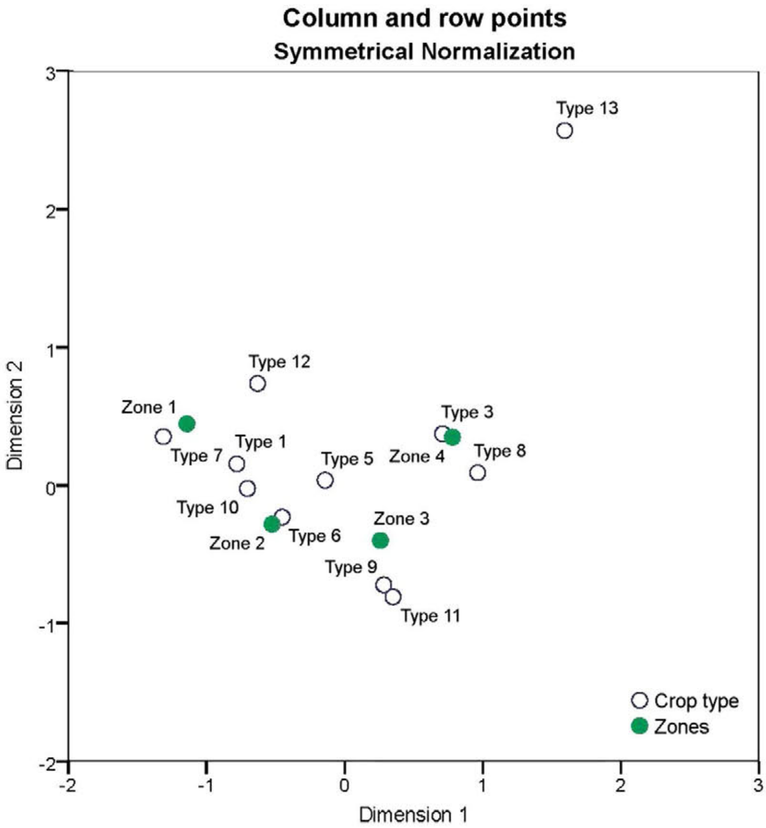

Fig. 6 Graph of the correspondence between the categories of the qualitative variables: "Crop type" (row) and "Zones" (column)

dependent variable is the altitude and the independent variable or factor is the variable zones ("Unit Price"). No significant differences were found between the mean altitudes in each zone, $F(3, 4553) = 0{,}712$, $p > 0{,}05$. As a result of these values, the fact that there is no relationship between the geographic altitude and the "Unit Price" of the different zones can be confirmed.

The fourth question formulated refers to a possible "relationship between the variable P and the classification variable "Unit Price". Thus, the variables P and C must first be defined. To this end, the variable "P″ refers to the "number of plots adjacent to a given plot in a radius of 100 m, 300 m and 1000 m" whereas the variable C refers to the number of structures adjacent to a given plot in a radius of 300 m,

Table 10 Most important crop type in each zone

Zones	Most important crop type in each zone
1	Type 7 = Dry-farmed vines
2	Type 6 = Dry-farmed olive trees
3	Type 9 = Dry-farmed fruit trees
4	Type 3 = Irrigated cultivation or arable land

Table 11 Results of the Chi-square contrast of the variables Crop type and Zones. Reason for the advantages of the crop type in each one of the zones

Zones	Crop type	Statistical (Chi-squared)	P_value	Reason for the advantages for the crop type I and the rest (in the zones j and the rest)
1	Type 7 = Dry-farmed vines	128.25	<0.05	5
2	Type 6 = Dry-farmed olive trees	9.66	<0.05	2
3	Type 9 = Dry-farmed fruit trees	45.62	<0.05	2
4	Type 3 = Irrigated cultivation or arable land	250.32	<0.05	3

3000 m and 30,000 m. In general, obtaining the variable P consists of calculating the number of centroids in the plots existing in each zone for an area of influence of 100 m, 300 m and 1000 m. The variable C is calculated based on the number of centroids in the structures that exist with respect to the rural type plots for an area of influence of 300 m, 3000 m and 30,000 m. Three statistical effects were found for each area of influence.

There is a statistically significant effect for the variable P100 in the category "Unit Price" with F(3, 4553) = 177.99 ($p < 0.05$). Since it is a statistically significant effect, we applied the Games-Howel test showing significant differences for the variable P100 among the groups as observed in Table 12.

In the case of the variable P300, there is also a statistically significant effect in the category "Unit Price" with F(3, 4553) =194.91 ($p < 0.05$), meaning the Games-Howel test for multiple comparisons was applied to establish that there are also significant differences in this variable, as observed in Table 13.

In relation to the variable P1000, the corresponding test was done to determine that there is no statistically significant effect in the category "Unit Price" with F(3, 4553) =1.26 ($p > 0.05$).

Finally, the fifth and last question formulated aims to determine the existence of some type of relationship between the variables C and the variable category of "Unit Price". In this new case, a statistically significant effect was confirmed to exist for the variable C300 in the category "Unit Price" with F(3, 4553) =73.78 ($p < 0.05$), meaning the Games-Howel test for multiple comparisons was applied, discovering that there are no significant differences for the variable C300, as observed in Table 14.

There is also a statistically significant effect for the variable C3000 (Table 15) in the category "Unit Price" with F(3, 4553) = 342.71 ($p < 0.05$).

Table 12 Games-Howel Test to find the significant income differences between two zones (i, j) for P100

[Zone I , Zone J]	Statistical (T_Student)	P_value
[Zone 1, Zone 2]	7.81	<0.05
[Zone 1, Zone 3]	14.61	<0.05
[Zone 1, Zone 4]	23.86	<0.05
[Zone 2, Zone 3]	6.30	<0.05
[Zone 2, Zone 4]	15.68	<0.05
[Zone 3, Zone 4]	9.91	<0.05

Table 13 Games-Howel Test to find the significant income differences between two zones (i, j) for P300

[Zone I , Zone J]	Statistical (T_student)	P_value
[Zone 1, Zone 2]	7.33	<0.05
[Zone 1, Zone 3]	14.21	<0.05
[Zone 1, Zone 4]	24.36	<0.05
[Zone 2, Zone 3]	6.12	<0.05
[Zone 2, Zone 4]	16.66	<0.05
[Zone 3, Zone 4]	1.51	<0.05

Finally, the definitive confirmation in the case of the variable C30000 is that there is a statistically significant effect in the category “Unit Price” with F(3, 4553) =879.10 ($p < 0.05$). Thus, the Games-Howel test was applied for multiple comparisons with the existence of significant differences in the variable for said variable found (Table 16).

Discussion

In view of the results outlined above, the most important findings of this work are as follows:

Two spatial interpolations were done to estimate the values of the variable “Unit Price” in the zones not sampled. One estimate was done with the traditional estimator of inverse distance (IDW with $p = 1$) and the other estimate was done using Geostatistics, specifically ordinary kriging. The IDW estimate was chosen as it had a lower Root Mean Square (27.74 €/m^2). Nonetheless, the differences between both estimates was small with a continuous surface (maps) obtained with these estimates. See Fig. 2.

The result of the IDW estimate was used for a spatial classification or zoning; in other words, 4 zones were obtained where the average values of the “Unit Price” were classified (see Table 6).

It is coherent to zone the Region of Murcia into 4 spatial zones based on the “Unit Price” declared in sales: Zone 1 (Low prices from 0.32 to 2.4€/m^2), Zone 2 (Average prices from 2.4 to 5.8€/m^2), Zone 3 (High prices from 5.8 to 11.2€/m^2), Zone 4 (Very High prices from 11.2 to 210€/m^2). See Fig. 4 and the distribution of these zones.

The fact that the very high price zone mostly corresponds to the richest agricultural areas of Campo de Murcia, Campo de Cartagena and the Coastal Zone (greenhouse crops) can be observed.

Table 14 Games-Howel Test to find the significant income differences between two zones (i, j) for C300

[Zone I , Zone J]	Statistical (T_student)	P_value
[Zone 1, Zone 2]	1.18	>0.05
[Zone 1, Zone 3]	3.97	<0.05
[Zone 1, Zone 4]	11.10	<0.05
[Zone 2, Zone 3]	2.97	<0.05
[Zone 2, Zone 4]	10.67	<0.05
[Zone 3, Zone 4]	9.36	<0.05

Table 15 Games-Howel Test to find the significant income differences between two zones (i, j) for C3000

[Zone I , Zone J]	Statistical (T_student)	P_value
[Zone 1, Zone 2]	11.67	<0.05
[Zone 1, Zone 3]	21.26	<0.05
[Zone 1, Zone 4]	25	<0.05
[Zone 2, Zone 3]	7.97	<0.05
[Zone 2, Zone 4]	21,63	<0.05
[Zone 3, Zone 4]	18.99	<0.05

There is a relationship between the no. of plots and the different zones. As the price of the zones increase, the no. of plots increases. The lowest “Unit Price” (Zones 1 and 2) show a lower number per plot than Zones 3 and 4.

The 4 Zones have their own characteristics in relation to the crop type. Zone 1, with a predominance of dry-farmed vines, with this crop showing a value of approximately 5 times higher than the rest of the crops. Zone 2 reflects dry-farmed crops with an approximate value of 2 times higher than the rest of the crops. Zone 3 reflects dry-farmed fruit trees with an approximate value of 2 times higher than the rest of the crops. Zone 4 reflects irrigated crops with an approximate value of 3 times higher than the rest of the crops.

It is interesting to note how there is no association between the average altitude of the 4 zones and the relationship with the “Unit Price”. In other words, the average altitude of the zones is not a parameter that classifies the price.

Another significant data that has an impact on the “Unit Price” of the 4 zones is the no. of nearby structures. As the zone increases in price, the no. of nearby structures at significantly different distances increases. For example, the difference at 300 m between the average number of structures between Zone 1 and Zone 4 is 6.14 times higher than in the last zone.

In view of these results, it is necessary to discuss whether the division of the rural properties in the Region of Murcia into 4 spatial zones is the most appropriate. Why not three? The main reason is that the rural values in this Region can be divided into four large areas that are practically equivalent in size and separated by a N-E, S-W line (Fig. 6). The best land in the province would be in the South: Huerta de Murcia, Campo de Cartagena, greenhouse crops on the coast and all of these lands would be surrounded by other less rich crops. Therefore, it is necessary to sub-divide the South into two sub-zones (Zone 3 and Zone 4). The poorest dry-farming lands with vineyards and olive groves would be in

Table 16 Games-Howel Test to find the significant differences in income between two zones (i, j) for C30000

[Zone I , Zone J]	Statistical (T_student)	P_value
[Zone 1, Zone 2]	14.08	<0.05
[Zone 1, Zone 3]	30.48	<0.05
[Zone 1, Zone 4]	49.74	<0.05
[Zone 2, Zone 3]	16.84	<0.05
[Zone 2, Zone 4]	38.91	<0.05
[Zone 3, Zone 4]	24.37	<0.05

the second large area, the North Zone. As a result of this, optimising the rural properties into four zones in this Region was coherent and representative (Fig. 6).

Another observation to be discussed is found upon looking at the orientation of the different zones in Fig. 6. It could be said that there is a predominant N-E, S-W orientation. This fact was also observed with the Gaussian semi-variogram raster model (Fig. 3) where the existence of a geometric or elliptic anisotropy is clear in the data in the direction: N-23°-E; in other words, coinciding with the direction of the spatial zones. All of this means the prices of the plots tend to show greater continuity in this direction.

Conclusions

All of this work is encompassed in the DGC's concern for valuating rural property as well as future evolutional market valuations. As mentioned in the abstract, the sole purpose of the study was to check the coherence of the parameters used by the Cadastre and their degree of implication in the valuation of the "Unit Price".

In order to do so, spatial analysis techniques were used such as IDW and kriging. Of the two analysis methods, the first method proved to be better than the second as far as the predictions of the "Unit Price" values not sampled. However, the kriging procedure also offers reasonable results, considering the "Gaussian" variogram which is most appropriate for modelling the spatial correlation of the "Unit Price" given that the mean error is closest to zero and the root mean square standardized error is closest to 1.

The Statistical Tests and Correspondence Analysis Tests also proved quite useful for validating the results of the spatial analysis. Of the five cases considered, it was verified that the mean rent value is statistically significant in the different zones which were established as per the "Unit Price". It has been proven that there is an association between the crop type and the different zones. There is also a significant relationship between the number of plots and the zones. Each of the zones is associated with one more important crop type. As the zone increases in price, the no. of nearby structures at significantly different distances increases.

Finally, it is worth highlighting that the analysis tools used for this work have proven to be highly effective in determining cadastral valuation rates based on the current state of the coverage and land uses. Moreover, the verification techniques used in this study make it possible to ensure the validity of the results obtained from the analysis techniques used.

Acknowledgments Special thanks to Mr. Luis González-Carpio Fernández from the Directorate General of the Cadastre for permitting the use of the existing documentation.

References

Alcázar, M.G. & Ariza, F.J. (2004). Situación actual de la valoración catastral rústica en España: propuesta de un modelo alternativo. Revista Catastro n° 52.

Bailey, T. C. & Gatrell, A.C. (1995). *Interactive spatial data analysis*. Harlow, Essex: Ed. Longman Group.

Ballestero, E. (1992). Metodología de la valoración agraria y sus aplicaciones catastrales. Revista Catastro, n° 20.

Burrough P.A. & McDonnell R.A. (1998). *Principles of geographical information systems*. Oxford University Press.

Caballer, V. (1994). *Métodos de valoración de empresas*. Madrid: Ed. Píramide.

Caballer, V. (2008). Valoración Agraria. Teoría y práctica. Ediciones Mundi-Prensa. 5ª Edición.

Cantillon, R. (1755). An Essay on Economic Theory. An English translation of Richard Cantillon's Essai sur la Nature du Commerce en Général. Ed. Mark Thornton.

Chiara, M., Stefano, C., & Guido, S. (2015). Agricultural Land Consumption in Periurban Areas: a Methodological Approach for Risk Assessment Using Artificial Neural Networks and Spatial Correlation in Northern Italy. *Appl. Spatial Analysis*, 1–18. doi:10.1007/s12061-015-9168-9.

Cournot, A. (1992). *Johann Von Thunen (1783–1850)*. Aldershot: Ed. Edward Elgar Pub.

Cressie, N. (1990). The origins of kriging. *Mathematical Geology, 22*, 239–252.

De la Fuente Fernández, S. (2011). *Análisis de Correspondencia Simple y Múltiple*. Madrid: Ed. Universidad Autónoma de Madrid.

Fleis, J. L. (1981). *Statistical methods for rates and proportions*. New York: Ed. John Wiley.

Garza García J., Morales Serrano, B. & González Cavazos, B. (2013). *Análisis Estadístico Multivariante*. Un enfoque teórico y práctico. Ed. Mc Graw Hill, pp 605

Goodchild, M. F., & Haining, R. P. (2004). GIS and spatial data analysis: converging perspectives. *Papers in Regional Science, 83*, 363–385. doi:10.1007/s10110-003-0190-y.

Gotway, C. A., Ferguson, R. B., Hergert, G. W., & Peterson, T. A. (1996). Comparison of kriging and Inverse Distance Methods for mapping soil parameters. *Soil Science Society of American Journal, 60*, 1237–1247.

Grimes, A., & Liang, Y. (2009). Spatial determinants of land prices: Does Auckland's metropolitaurban limit have an effect? *Applied Spatial Analysis, 2*, 23–45. doi:10.1007/s12061-008-9010-8.

Joaristi, L. & Lizasoain, L. (1999). *Análisis de Correspondencia*. Madrid: Ed. La Muralla.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K. & Lucas, N. (2001). *Using ArcGis Geostatistical Analyst*. Redlands: Ed. ESRI.

Juárez, S. (1952). *Valoración de Fincas y predios*. Dossat: Madrid. Ed.

Kravchenko, A., & Bullock, D. G. (1999). A comparative study if interpolation methods for mapping soil properties. *Agronomy Journal, 91*, 393–400.

Krivoruchko, K. (2011). *Spatial Statistical Data Analysis for GIS Users*. DVD-ROM – Audiolibro.

Lozano, Z., Bravo, C., Ovalles, F., Hernández, R. M., Moreno, B., Piñango, L., & Villanueva, J. G. (2004). Selección de un diseño de muestreo en parcelas experimentales a partir del estudio de la variabilidad espacial de los suelos. *Bioagro, 16*(1), 1–17.

Martín, A. & Díaz, V. (2002). Geoestadística aplicada. Instituto de Geofísica, UNAM. Instituto de Geofísica y Astronomía, CITMA, Cuba.

McKillup, S. & Darby Dyar, M. (2010). *Geostatistics Explained. An Introductory Guide for Earth Scientists*. Cambridge.

Montero, J.M. & Larraz, B. (2008). *Introducción a la Geoestadística Lineal*. Ed. Netbiblo, S.L.

Moral García, F. J. (2003). *La Representación gráfica de las variables regionalizadas. Geoestadística lineal*. Cáceres: Universidad de Extremadura.

Moreno, A. (2007). *Sistemas y Análisis de la Información Geográfica*. 2ª Edición. Ra-Ma.

Morillo, M. C., García-Cepeda, F., & Martínez-Cuevas, S. (2016). The application of spatial analysis to cadastral zoning of urban áreas: an example in the city of Madrid. *Survey Review (on line)*. doi:10.1080/00396265.2015.1113029.

Pérez, C. (2009). Técnicas estadísticas multivariante con SPSS. Ed. Garceta.

Ricardo, D. (1817). *The principles of political economy and taxation*. Ed. John Murray. 3rd edition.

Ruiz Rochera, F. (1861). *Tasación de tierras*. Castellón: Ed. Martín Masuastegui.

Schloeder, C. A., Zimmerman, N. E., & Jacobs, M. J. (2001). Comparison of methods for interpolating soil properties using limited data. *Soil Science Society of American Journal, 65*, 470–479.

Smith, A. (1776). *On Systems of political Economy*. London: Ed. Random House.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2007). *Análisis de la varianza. Probabilidad y estadística* (Segunda Edición ed.). México: Ed. McGraw-Hill.

Tinoco Gómez, O. (2008). *Una aplicación de la prueba Chi cuadrado con SPSS. Industrial Data, vol. 11, núm. 1, enero-junio, 2008*. Perú: Universidad Nacional Mayor de San Marcos Lima.

Villatoro, M., Henríquez, C., & Sancho, F. (2008). Comparación de los Interpoladores idw y kriging en la variación espacial de pH, Ca, Cice y P del suelo1. *Agronomía Costarricense, 32*(1), 95–105.

Websites

ArcGis, 2012. Help [on line] Available at: <https://www.arcgis.com/features/> [Accessed 21 September 2015].

IBM SPSS Statistics 20. Help [on line] Available at: http://www-01.ibm.com/software/analytics/spss/ [Accessed 6 February 2016].

Proyecto e-Math 2 Financiado por la Secretaría de Estado de Educación y Universidades (MECD). Available at: <http://www.uoc.edu/in3/emath/docs/ANOVA.pdf> [Accessed 10 October 2015].