Spatial Modeling of Urban Growth by Application of Kriging Estimator

**Abstract**: The aim of the present study is to simulate the urban development, on the surface and on undergroung, at the municipal scale, using a spatial modeling Geostatistical technique. For this purpose, normal kriging was used as a geostatistical interpolation’s method of urban development, taking into account normal growth trends.

The results of the method used allowed us to understand the advantages and usefulness of Geostatistical technique applied to urban planning.

**Key words:** Geostatistics, Kriging, Spatial Modeling, GIS.

1. Introduction

Geostatistics or Spatial Statistics is an area of knowledge based on a set of statistical methods, techniques and instruments to characterize natural spatial phenomena, quantifying the uncertainty of their knowledge (SOARES, 2000).

Geostatistics was born in the 1960s, in France, associated to the need of modeling geological resources, mainly in the study of mineral and metal deposits concentrations and the quality of underground water.

The Portuguese school is a well-respected school in the Geostatistics field. Its development took place in 3 stages. The first took place between 1960 and 1970, regarding to the study of mining phenomena. The second, around 1970 and 1980, with application to environmental phenomena, such as soil, water and air pollution forecast.

At this stage, a renowned school of Geostatistics was born at the Instituto Superior Técnico, at University of Lisbon, with Prof. Quintino Rogado, connected to the Angola metal deposits prospection. The third phase, from 1980 to our days, has been predominantly marked by oil exploration, in addition to the, already mentioned, application areas. (SOARES, 2000). Nowadays Geostatistics has a variety of applications from earth sciences to social sciences. (WEBSTER, 2008).

The Geostatistics methodology is based on the application of a set of statistical instruments quantifying spatial continuity, through spatial interpolation and stochastic simulation models. The first are based on the study of spatial variability, and the second on quantification of the uncertainty associated with each phenomenon. The uncertainty, associated with the behavior of a given spatial phenomenon, is a Geostatistics’ characteristic, contrary to the deterministic models used by the possibilistic branch of the science. The deterministic prediction is based on geometric criteria, calculating the magnitude of phenomena from the linear combination of observed values by inference, triangulation and inverse power of distance methods. It is deterministic because this is based on, exclusively, geometric criteria and do not incorporate the uncertainty measure of each phenomenon.

Geostatistical models are based on the measure of uncertainty, associated with each phenomenon through random sampling data, and are based on the spatiality of the average which is their only achievement. This implies that the average can be estimated from the observations made. Spatial homogeneity is a consequence of the stationarity of phenomena. The correlation between two random variables is given to us by covariance (C) and variogram ( γ) and it is measured by spatial continuity or spatial dispersion between them (See expressions [1] and [2]).

[1] C(h)=.

[2] γ(h)=.

The variable *h* represents the distance between the starting point Z(x) and the end point Z(x+h). Its validity applies to the entire spatial distribution under study.

The spatial variogram (γ) gives us the structural behavior of the sample of a given spatial phenomenon, being, by definition, the deviations’ variance. Its modeling allows synthesizing the spatial continuity’s patterns of the phenomenon under study, which represents the surface and ground urban growth, for the Oeiras municipality, near Lisbon, Portugal.

There are four kinds of spatial variograms, which should be distinguished (CLARK, 1979):

* Spherical: where the phenomenon variation, next to the origin, behaves irregularly, translating into a graph with a spherical structure.
* Exponential: the variation of the phenomenon along the origin behaves irregularly, with abrupt variations, with greater spatial correlations at greater distances from the origin.
* Gaussian: identifies the existence of a continuous and regular phenomenon with parabolic behavior, with slow variation at the origin.
* Power: when the phenomenon is continuous and not stationary (does not tend to a plateau).

The combination of one, or more, spatial variogram models results into imbricated (complex) structures ([McBRATNEY](https://bsssjournals.onlinelibrary.wiley.com/authored-by/ContribAuthorRaw/McBRATNEY/A.%2BB.), and [WEBSTER](https://bsssjournals.onlinelibrary.wiley.com/authored-by/ContribAuthorRaw/WEBSTER/R.), 1986).

When the studied phenomenon presents a great spatial variability, at small scale and around the origin, it is common to be called "nugget effect", in analogy to that verified in geological prospects.

Regarding spatial continuity, a model is called isotropic when it has the same spatial behavior in all space directions. On the other hand, the model is anisotropic when a variable presents a behavior of greater spatial continuity in a privileged direction. (GUIMARÃES, 2013).

Statistical methods of spatial inference allow estimation. The kriging was used in 1965 by Matheron G., in honor of the works of Krige D (SOARES, 2000). It is a geostatistical linear estimator [Zx0, (see expression [3]) resulting from the linear combination of variable N, the estimation error being null (i.e., there is no bias) and the estimation variance is minimal. The  is the weight and xthe variable to estimate.

[3] [Zx0x

2. Goals and resources used

The aim of this study is to understand the dynamics of height and ground urban growth, adopting the municipality of Oeiras, near Lisbon, Portugal, as a case study. It is about estimating, by a geostatistical method, the behavior of the built above and under the doorstep level, through the known trends of urban growth.

In fact, we assume that the behavior of the urban building growth translated into number of floors, can be explained by Geostatistics, taking into account that the municipal master plan of Oeiras, and the neighboring municipalities, define that *«(...) the new building urban areas must respect the existing alignment, the modal value of the height of the facades, the scale, the materials and the existing colors».*

Thus, we used the normal Kriging as a Geostatistical interpolation method (CRUZ, 2008).

The main resources used in this project development were the following ones:

* Cartography with buildings representation, kindly available by the Oeiras Municipality;
* Statistical manipulation software: ANDAD, GeoMS and GeoStatistical Analyst, developed by CVRM - Center for Geo-Systems, IST/UL. ;
* Excel spreadsheet;
* ArGIS/Geoestatistical extension (ESRI, 2004).

3. Information pre-processing

We considered the buildings implantation asmodel *input* with the following associated information:

* Maximum number of floors above and below the surface;
* Construction season.

The methodological phase-up of the work is developed as follows:

1. Conversion of buildings represented by a polygon to a point geometry (with centroid generation);
2. Calculate the projected coordinates of each building centrode to the cartographic reference system ETRS89/PT TM-06 (EPSG:3763);
3. Integration the spatial information to ANDAD and GeoMS software;
4. Analysis of the correlation between database variables in ANDAD environment;
5. Determination of the kriging estimator, using the ArcGIS Geostatistical extension;
6. Spatial representation of the processed geographic information.

The correlation between the variables PACI (Number of floors above ground level) and PABX (Number of floors below ground level), using the ANDAD software, was analyzed in relation to other attributes (number of total housing, number of residential accommodations, number of commercial and service accommodations, functional typologies of the building and construction season), in order to understand if any external drift can be used to model calibration.

Initially, through the principal component factor analysis (PCA), it is verified that there is no correlation between the variables PACI and PABX for the totality of the sample, that is, a tall building does not necessarily mean that it has a greater number of basements or vice versa. Through the factorial analysis of multiple correspondences (AFCM), it was verified that the behavior of the sample is explained by the following axes F1/F2 and F1/F3.

By analyzing Fig. 1, and the relationship between F1/F2 axes, the following illations can be deduced:

* There is a high concentration of multi-family residential buildings associated to buildings with a larger number of floors (ID9... ID22) and buildings with 3 basements (PAB3);
* There is a high concentration of intermediate floor buildings (ID4) with buildings with basements;
* Low-rise buildings and residential accommodation (single/two-family) co-exist with trade and service buildings;
* Buildings classified as "other buildings" are associated with the low number of floors and the most recent construction period.

Fig. 1 - ACP on the F1/F2 axis

By factorial analysis of multiple correspondences (AFCM), explained by the F1/F2 axis system, it is checked that multi-family housing buildings establish, on the one hand, a group of their own, and, on the other hand, coexist with single/bifamily housing buildings (ID1) as well as with commerce and service buildings (ID3), (Fig.2).

Fig.2 - AFCM on f1/f2 axes (variable typologies according to the function of the building)

By ACP, and according to the F1/F3 axle system contained in Fig.3, it can be concluded that:

* There is a clear separation of the building, according to the number of floors. On the one hand, the multi-family building, associated to the large number of accommodations and, on the other hand, that of single-family housing, associated to the ground floor with a single residential accommodation;
* The buildings typologies according to the function are not correlated with the construction period, for the overall study area.

Fig. 3 – ACP on the F1/F3 axis

The factor analysis of multiple correspondences (AFCM), explained by the F1/F3 axis, once again, allows corroborating the conclusions, previously drawn (Fig. 4).

Fig. 4 – AFCM matches in f1/f3 axes (variable typologies according to the function of the building)

Taking into account the previous postulates and the non-evident correlation between variables, for the globality of the sample, the studied variables were rejected as external ones and, for this reason, *kriging with external* drift was not applied, but *rather* the *normal kriging,* using the PACI and PABX variables, because there are stationarity around the mean to the same variables.

4. Construction and Spatial Variograms Adjustment

The hypothesis of a phenomenon stationarity can be verified by spatial variograms, where each variable tends to a certain level.

We can now consider the variable PACI.

The first step aims to understand the behavior of the observed phenomenon in space, according to several directions. We can conclude that the dominant orientation (the one with the highest spatial continuity, corresponding to the globality of the county), given by the largest axis of the ellipse “a1” is according to the azimuth -37º;0º (NNW-SSE) and the one with the lowest spatial continuity, (which corresponds to the dimension of the neighborhood), given by the smallest axis of the ellipse – “a2”, and perpendicular to that, has the direction 53º;0º (NNE-SSW), (see Fig. 5).

By the theoretical variograms configuration, we found that the PACI variable has geometric anisotropy, presenting two spatial structures: the first structure is isotropic and coincides with a spherical variogram, in which a1=a2=260 m (neighborhood scale), with direction (azimuth) 323º (360-37=323º) and c2=1.52 (partial variogram. It should be referred too that the “c1” and c2” values have no units, as they correspond to *gammas*).

The observed "nugget effect" explains the great variability of the phenomenon around the origin (c1 = 0.392). The second structure is anisotropic, with irregular propagation in space, for this reason, presenting spherical variogram configuration where: a1=3260 m (average cluster dimension),a2=760 m, azimuth=323º and c2=1,718.

Then, the experimental variograms were calculated in two directions, perpendicular to each other, in order to detect the phenomenon of spatial anisotropies (Fig. 6a/b).

Fig. 5 - Spatial structure of the height-built

Fig. 6a. - Experimental variograms and structure of PACI - variable PACI - Axis to a1

Fig. 6b. - Experimental variograms and structure of PACI - variable PACI- Axis to a2

The PABX variable has an isotropic behavior, because the ellipse dimension axes are identical (a1=a2=260m). Otherwise, the behavior of the PABX variable, at the cluster scale, is not spatially structured. It is also observed a "nugget effect", proportionally higher in relation to the level, when compared to what was verified for PACI, which indicates greater variability for PABX. The spatial variogram presents only one spherical structure type (Fig. 7a/b).

Fig. 7a- Experimental variograms and their structures d the variable PABX. - Axis to a1

Fig. 7a- Experimental variograms and their structures d the variable PABX. - Axis to a2.

5. The Application of the Kriging Estimator

The spatial variograms calculation, for both variables, was performed using the GeoMS program. The generated parameters were used as *input for the* urban growth *estimation*, through the kriging estimator, on a GeoStatistical extension of ArcGIS program.

The total sample size comprises 17890 points, corresponding to the buildings centroids, and the estimation of each point was based on the minimum number of 6 and the maximum number of 20 sample points, in each direction, summarizing the parameters considered in the sample estimation.

After model generation (number of basements, above and ground floors, buildings estimation) the data were converted to a grid format, with 10 m spatial resolution and later reclassified into a non-discrete values matrix.

6. Outputs and Results Interpretation

Given that the success in using statistical data depends largely on how it is represented and can be used (REIS, 2009), we will present the results in map form and analyze them. Considering the urban municipality growth on the surface, it is observed that 3 is the average of floors that is equivalent to a modal value of a 2-storey building. The tallest building, with 24 floors, is located in Algés-Miraflores urban agglomeration. In the municipality coexist, in the same proportion, a single-family housing buildings with multi-family housing buildings concentrating, these, with a relatively high number of floors and lodgings. The spatial distribution of the multifamily building reveals a concentration in the parishes of Carnaxide, Algés and Linda-a-Velha, located in NE and SE municipality quadrants, and also, in Paço de Arcos and Oeiras, in the SW quadrant, as it can be seen in the following figure.

Fig. 8 - Number of floors above ground level

The trend of surface urban growth, by *kriging* estimator, reveals a propensity to a growth in height, around the areas where tall buildings are predominant, with attenuation of the maximum values for 18 floors (kriging, by definition, tends to estimate the phenomenon by the values around the average and, due to this reason, the marginal values are not extreme).

The highest values are found in the parishes of Carnaxide (central area of the Carnaxide cluster and in south of the Outurela-Portela-Alto dos Barronhos cluster), Alto de Algés, Linda-a-Velha (western area bordering the parish of Cruz Quebrada - Dafundo), Terrugem, Tapada do Mocho and J. Pimenta neighborhood, in the parish of Paço de Arcos, the Neighborhoods of Figueirinha and Augusto Castro and Nova Oeiras, in the parish of Oeiras (SW of the map).

On the other hand, the dynamics of low density construction are recorded in Caxias and in the north of the motorway agglomerates (with the exception of Carnaxide), such as Outurela-Portela and a small area of Queluz de Baixo.

Fig. 9 - Estimate of the number of floors in height of the existing building.

In conclusion, the above-ground urban development, applying kriging estimator reveals a propensity for growth by height, around the areas where the higher-height buildings predominate, especially in the E and SW areas of the territory.

As far as the basement of building behavior is concerned, as can be seen in Fig. 9, the maximum number of floors below the ground quota is 4, observing a recent trend for an increase in the number of basements in new commerce and service buildings, such as the cases of Tagus and Lagoas business parks. The cellars are built for parking, storage and technical areas.

Fig. 10- Number of basement floors

It is also concluded that the estimate of the city underground growth is poorly structured, as a preferential direction of growth was not observed (Fig.10).

It is estimated that in business areas, where recent buildings have a greater number of parking cellars, the construction of new buildings keeps this trend. Stand out the areas with the largest number of basements, in the business park of Lagoas Park, Science and Technology Park (Tagus Park) and Santa Cruz Park (next to the Carnaxide mountain range) and a set of small urban areas dispersed by the parishes of Paço de Arcos, Algés (Alto de Algés), Carnaxide and Linda-a-Velha (periphery).

Fig. 11 - Estimate of the number of basement floors in the existing building

We can conclude that the urban underground growth, applied by kriging estimator, reveals an increasing tendency in number of cellars, in new commerce and services buildings on the business areas, without a preferential spatial development pattern.

The behavior of the buildings height is more complex and structured: it presents isotropy geometric at a block scale (explained by the "nugget effect", which translates great variability at the buildings height around the origin) and geometric anisotropy at the scale of the cluster, whose meaning of greater spatial continuity is given by the direction: NNE-SSW (Fig. 5).

Although the origin of Geostatistics is related to the exploitation of oil and mineral resources, nowadays it is associated with a vast number of applications, namely in earth sciences such as geology and agronomy, studying, for example, soil erosion and agricultural productivity (WEBSTER, 2008).

This research work verifies its applicability to urban planning through the evaluation of urban growth trends in a municipality in the Metropolitan Area of Lisbon.

References

CLARK, I., (1979), Practical Geostatistics, Applied science publishers L.da, London.

CRUZ, Fernando, (2020), Monitoring of Territorial Systems at the Municipal Level - Geographic Information Technologies with Application to Land Use, New Academic Editions, ISBN-13: 978-620-2-55892-1.

ESRI. (2004), Using ArcGIS Geostatistical Analyst. Redlands, CA – USA: ESRI Press.

GUIMARÃES, WELLINGTON DONIZETE, (2013), Geostatistics for Mapping the Spatial Variability of Soil Physical Attributes, Thesis presented to the Federal University of Viçosa, to obtain the title of Doctor Scientiae in Civil Engineering, Brazil.

[McBRATNEY](https://bsssjournals.onlinelibrary.wiley.com/authored-by/ContribAuthorRaw/McBRATNEY/A.%2BB.), A. B.;[R. WEBSTER](https://bsssjournals.onlinelibrary.wiley.com/authored-by/ContribAuthorRaw/WEBSTER/R.), R, (1986), “Choose functions for semivariograms of soil properties and fit them to volunteer estimates”, Europen Journal of Soil Sciencies, [Vol. 37, Nº 4](https://bsssjournals.onlinelibrary.wiley.com/toc/13652389a/1986/37/4),pág. 617-639,

PEREIRA, Henrique Garcia (2000), Classification Methods, Texts to Support the Intensive Course of Data Analysis - 1988-2002, CVRM/IST, Lisbon.

Pereira, Henrique Garcia; Sousa, António Jorge, (2000), ANDAD Program, Vs7.10. User Manual, CVRM/IST, Lisbon.

Pereira, Henrique Garcia; Sousa, António Jorge, (2000), Data Analysis for Handling Multidimensional Frames, Texts supporting the Intensive Course of Data Analysis - 1988-202, CVRM/IST, Lisbon.

REIS, Elizabeth; Melo, Paulo; Andrade, Rosa; Calapez, Teresa, (2018), Estatística Aplicada, vol. 2, Edições Sílabo, Lda, Lisboa.

SOARES, Amílcar, (2000), Geostatistics for Earth and Environmental Sciences, IST Press, Lisbon.

WEBSTER, R., (2008), “Geostatistics for Earth Engineers and Scientists”. European Journal of Soil Sciences,