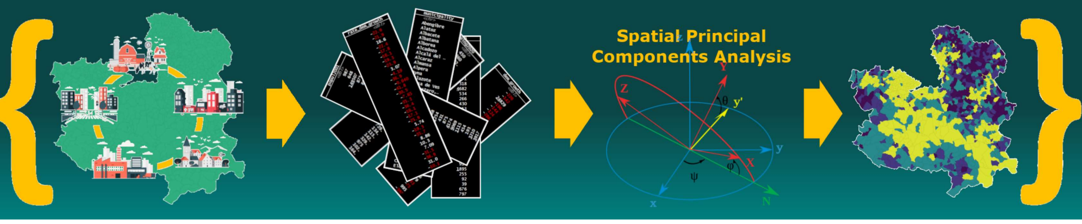
Is it possible to classify the municipalities of Castilla-La Mancha using ***Geostatistics*** and ***Unsupervised Machine Learning*** techniques?



Graphical abstract

Design of a Spatial Depopulation Risk Indicator

Isidro Hidalgo Arellano & Gema Fernández-Avilés Calderón Isidro.Hidalgo@uclm.es Gema.FAviles@uclm.es

Abstract

Depopulation is a major problem in rural areas of the world. The main aim of this work is the construction of a Spatial Depopulation Risk Index (sDRI) for the 919 municipalities of Castilla-La Mancha, using geostatistical techniques and principal component analysis. The theoretical semivariogram reveals spatial dependence up to a distance of 60 kilometers. Based on this range a neighbourhood network is constructed. Then a spatial principal component analysis (sPCA) is applied to a set of demographic variables. Finally, the sDRI is designed by extracting and scaling the first principal component of the sPCA. The resulting indicator identifies the areas with depopulation risk in which counter-measures can be applied.

Motivation

Did you know that some areas of Cuenca y Guadalajara have **a lower population density than Siberia**? Depopulation is a major problem in rural areas of Castilla-La Mancha. **!!!**

Table 1 shows that **445 municipalities of the region lost more than 20% of their population**, whereas only 237 municipalities improved it in the last two decades (2001-2020).

Table 1: Number of municipalities according to growth rate between 2001 and 2020

Population growth rate	Number of Municipalities
loss >20%	445
loss 10-20%	131
loss 5-10%	62
loss <5%	44
gain <5%	43
gain 5-20%	67
gain >20%	127

Objectives

- **General**: The Construction of sDRI using spatial Principal Component Analysis to ranking the municipalities of Castilla-La Mancha in order to identify areas in which counter-measures can be applied.
- **Secondaries**:
  - To detect the range of spatial dependence of depopulation in Castilla-La Mancha.
  - To include the spatial dependence in a depopulation risk index.
  - To rank the municipalities of Castilla-La Mancha in terms of risk depopulation.

Methods

As stated in the First Law of Geography, "Everything is related to everything else, but near things are more related than distant things" **(Tobler, 1970)**. Since depopulation is a variable with spatial dependence, we deal with geostatistics and machine learning techniques to carry out our purpose. Figure 1 stands for the methodology used in this work:

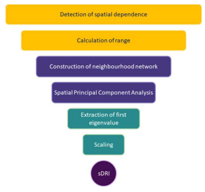

Detection of spatial dependence
Calculation of range
Construction of neighbourhood network
Spatial Principal Component Analysis
Extraction of first eigenvalue
Scaling
sDRI

Figure 1: Methodology

The main aim of this study, the construction of a sDRI, is achieved with the following steps:

1.The instrument used *par excellence* to detect the spatial dependence is the semivariogram **(Montero et al., 2015)**. Its expression is given by:

$$\gamma(s_i - s_j) = \frac{1}{2}V((s_i) - Z(s_j)), \forall s_i, s_j \in D \qquad (1)$$

where s_i and s_j are two locations (municipalities) in the domain D, V is the variance, and $Z(s)$ is the regionalized variable (*Population growth rate*) at location (municipality) s.

2.The range of spatial dependence is extracted from the semivariogram (see Figure 1).

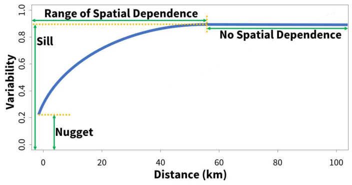


Figure 2: Components of a semivariogram

3.Based on the range, a neighbourhood network is constructed in the form of a proximity matrix L.

4.Then a spatial principal component analysis **(Jombart, T. et al., 2008)** is applied to ten demographic variables **(Jato-Espino & Mayor-Vitoria, 2023)**.*population 2001, population 2020, youth (<16 years) 2020, elder (>64 years) 2020, growth population 2001-2020, population density 2020, natural increase rate 2010-2020, ageing index 2020, dependence index 2020* and *net migration rate 2010-2020*.

Two types of spatial patterns are discriminated: global and local structures, corresponding respectively to large positive and large negative eigenvalues. This is accomplished by maximizing:

$$C(v) = V(Xv)I(Xv) = \frac{1}{n}(Xv)^T LXv = \frac{1}{n}v^T X^T LXv \qquad (2)$$

where V is the variance, X the demographic data matrix, $I()$ the Moran's I, which catch the spatial autocorrelation, L the proximity matrix and v the scaled axes in R^{10}, with $||v||^2 = 1$.

Figure 2 shows the extreme theoretical possibilities.

5.The extraction of the first principal component is carried out.

6.The sDRI for each municipality is obtained scaling from 0 to 100 the principal component.

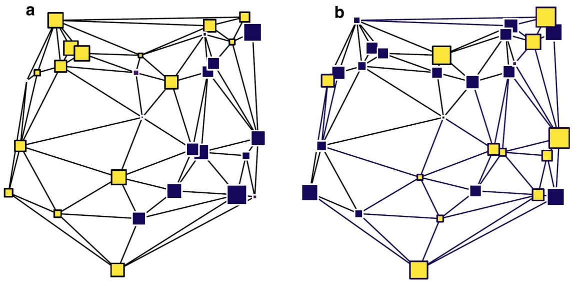


Figure 3: Theoretical cases: (a) spatial dependence, (b) no spatial dependence

Results

The first step is the estimation of a semivariogram to analyze the spatial dependence and calculate its range. As shown in Figure 4 the semivariogram is adjusted to a spherical model with the following parameters: **range** of 60000 meters (60 km), **sill** of 3419, and **nugget** of 1667.

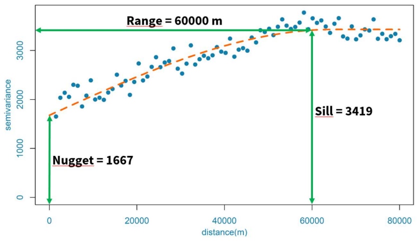


Figure 4: Adjusted semivariogram

Once the range of spatial dependence is estimated to 60 km, the neighbourhood network is constructed and the spatial analysis of principal components of depopulation in Castilla-La Mancha is performed. The results of the sPCA are shown in Figure 5.

The first two eigenvalues of sPCA (5a) show a strong global spatial dependence, whereas the last negatives eigenvalues reveal some local dependence; this is due to municipalities acting as development hubs, consequently earning population of neighbours. In the sPCA map (5b) three big areas of depopulation appear; namely the counties of Cuenca and Guadalajara, the west and the south of the region.

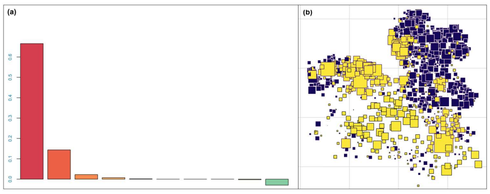


Figure 5: Principal results of spatial principal component analysis: (a) Eigenvalues of sPCA; (b) Map of sPCA scores of municipalities.

The last step of the work is to extract the first component of sPCA and scale it from 0 to 100. The resulting sDRI is used to classify the municipalities from lack of depopulation risk (sDRI = 0) to extreme risk (sDRI = 100). As shown in Figure 6-left Albacete is the municipality with absolute abscence of depopulation risk (sDRI = 0), followed by Guadalajara, Talavera de la Reina, Toledo and Azuqueca de Henares. In the opposite, we have Arandilla del Arroyo (sDRI = 100), followed by Alique, Valsalobre, Angón and Pineda de Cigüela. Figure 6-right represents the sDRI in a map of municipalities of Castilla-La Mancha.

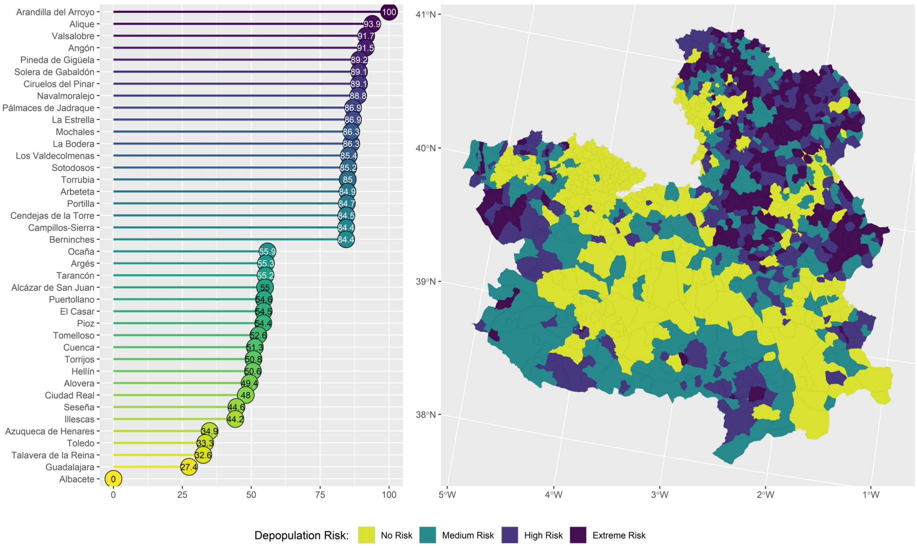


Figure 6: Depopulation Risk in municipalities of Castilla-La Mancha according to sDRI

Conclusions & Discussion

The applied spatial principal component analysis results in a Depopulation Risk Index which identifies numerous areas as having a medium to high risk of depopulation; namely, the majority of villages of Cuenca and Guadalajara, and the west and the south of the region. Conversely, it shows no risk for the areas of La Mancha and the Sagra and Henares industrial corridors, as well as the provincial capitals, Talavera de la Reina and Puertollano (see Figure 5).

As far as we know, this is the firs time that this methodologies have been applied to mesure the depopulation risk, and, specifically, in the form of sDRI to classify the municipalities of Castilla-La Mancha.

Related to the social ans policy implications, we propose to include the scores of sDRI into an expert system capable of identifying the areas in which counter-measures can be applied by local and regional governments.

References

- Jato-Espino, D.; Mayor-Vitoria, F. (2023). *A statistical and machine learning methodology to model rural depopulation risk and explore its attenuation through agricultural land use management*. Applied Geography, 152, 102870.
- Jombart, T.; Devillard, S.; Dufour, A.-B.; Pontier, D. (2008). *Revealing cryptic spatial patterns in genetic variability by a new multivariate method*. Heredity, 101, 92-103.
- Montero, J.M.; Fernández-Avilés, G.; Mateu, J. (2015). *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. John Wiley & Sons.
- Tobler, W.R. (1970). *A computer movie simulating urban growth in the Detroit region*. Economic Geography, 46-1, 234-40.

Universidad de Castilla-La Mancha
CAMPUS OF INTERNATIONAL EXCELLENCE

XI Jornadas de Doctorado
Toledo, Fábrica de Armas
24 de noviembre de 2023

¿Quieres saber cómo se ha hecho este póster?
https://github.com/IsidroHidalgo/PosterXIJornadasDoctorado