Subsecretaría de Desarrollo Regional y Administrativo

Gobierno de Chile

EXECUTIVE SUMMARY:

Rural Human Settlements in Chile, Communal Classification

A spatial analysis of the concentration and dispersion of the population.

Division of Studies and Public Policies
Department of Studies and Territorial Analysis
March, 2021

PRESENTATION[1]

Complexity is usually seen as an unfavorable characteristic. However, from the view of the Undersecretariat of Regional and Administrative Development (Subdere), it represents an opportunity to revert centralism, mobilize the endogenous potentialities of every territory and better focus administrative and financial efforts. It is thus possible to enhance inhabitants' quality of life. It is an opportunity to serve people.

Knowing the complexity of human settlements in a county means knowing the complexity of the territories, beyond statistical data. This is a great challenge because it implies quantifying, measuring, and systematizing the spatial information on more precise scales. Having a measure of locational complexity of the population allows us to pinpoint counties with more complex territorial issues, which entails more difficulties for the city halls serving them.

Currently, after updating the Report on Localities in Isolated Conditions, Subdere has georeferenced information on rural housing (which comes from the 2016 Pre-Census), services in the area where the city hall is located, and the distance and travel time from rural housing to city hall. These databases allow us to statistically compare the territorial-spatial variables by county and to classify counties based on these variables.

1 The full document in spanish "Asentamientos Humanos Rurales en Chile, Clasificación Comunal" can be downloaded from: https://bibliotecadigital.subdere.gov.cl

CONTENTS

I. GENERAL BACKGROUND AND CONCEPTUAL ASPECTS

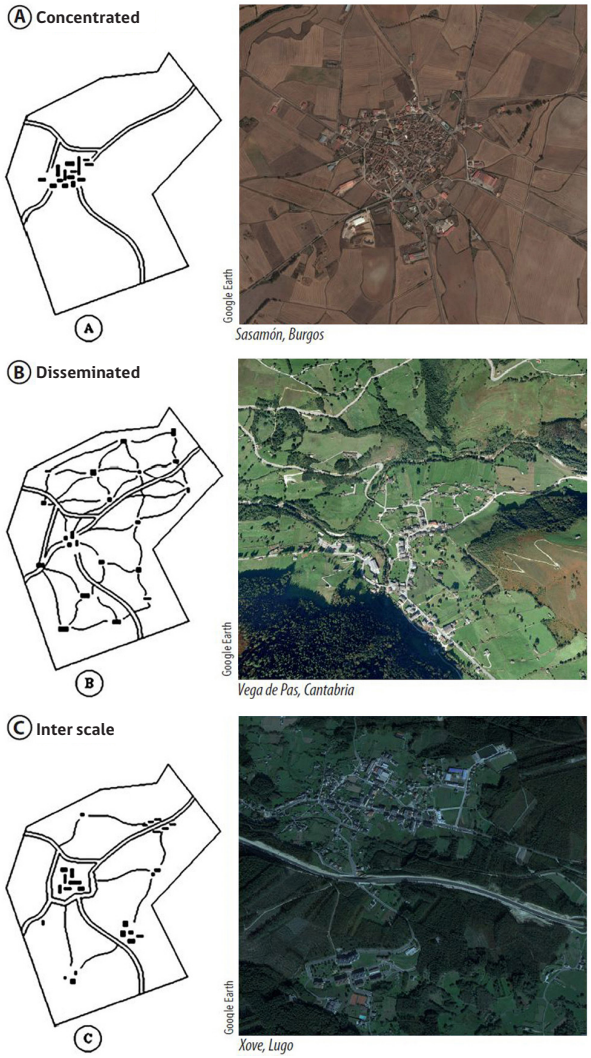
Human settlements "are usually close to or in the middle of zones with enough or abundant natural resources, such as fresh water. Also, many are located in places with economically important natural resources given the convenience of their trade or industrial development.".[2]

Population density, measured by inhabitant/km2 or inhabitant/hectare, has been used frequently as an estimate of the dispersion or concentration of the population in a given territory (usually, the county). One of the problems with this indicator is that in rural places the landscape is variable and excludes information related to the spatial distribution of the population.

In the National Atlas of Spain, the term "rural habitat" refers to settlement in the territory and considers natural variables. Thus, rural habitat refers to villages, places, towns, or neighborhoods in Spain, municipalities with a population no higher than 10,000 inhabitants. In this atlas, they refer to the distribution of the population according to the grouping of households. The following illustrations show examples of this atlas, with the distribution characteristics of human settlements in the rural habitat.

2 https://www.geoenciclopedia.com/asentamientos-humanos/

Illustration 1. Classification according to the National Atlas of Spain.



Source: National Atlas of Spain (2019).

This Spanish classification shows the importance of interpreting photos because choosing the distribution pattern that best represents the configuration of the population is left to the analyst's discretion.

Moreover, location theories explain how human settlements are distributed, how they use the space around them and the interactions they generate. These theories are simplified models of an irregular and complex reality. The location (central places) theory of Walter Christaller (1893–1969) postulates that the population concentrates around "central places" with the purpose of reducing travel time. From this we can deduce a hierarchy in which the most basic services occupy the first place and then come more specialized services.

This is based on measuring how accessible certain areas are to their local governments, as stated in Tobler's (1979) first law of geography or principle of spatial autocorrelation:

"Everything is related to everything else, but near things are more related than distant things."

To understand the complexity of a municipality's actions regarding its settlements, it is necessary to define the following elements.

1. The **hierarchy of the municipality**, understood as a county seat's access to a series of services.

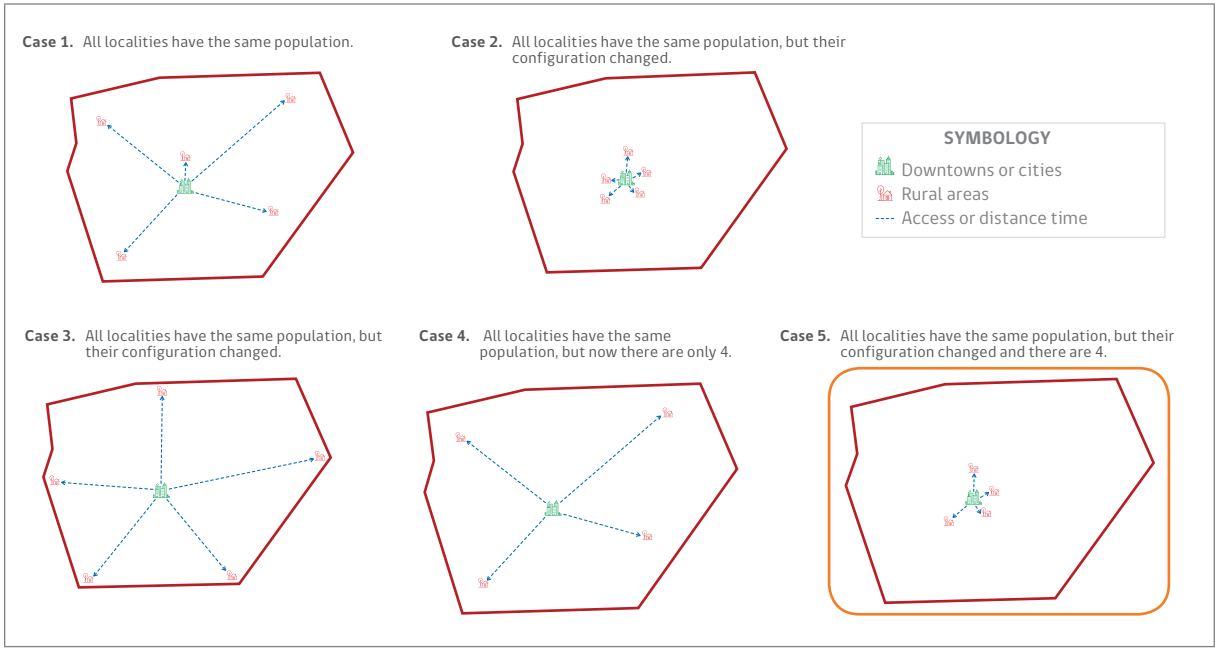
2. The **relationships** established by the settlements with their municipality.

3. The population **distribution patterns** in the territory.

Let us assume that all municipalities have the same hierarchy, associated with the amount of services present at the county seat, and all the towns in the county have the same population, but with a different distribution; in such a case, it seems intuitive that in spatial terms it is easiest to intervene in in the county with the fewest towns and that are the closest to city hall.

Illustration 2 schematically shows which county would be the least complex to intervene, assuming that all its towns have the same population and its municipalities the same hierarchy of services, but with different distribution patterns.

Illustration 2. Distribution of localities scheme.

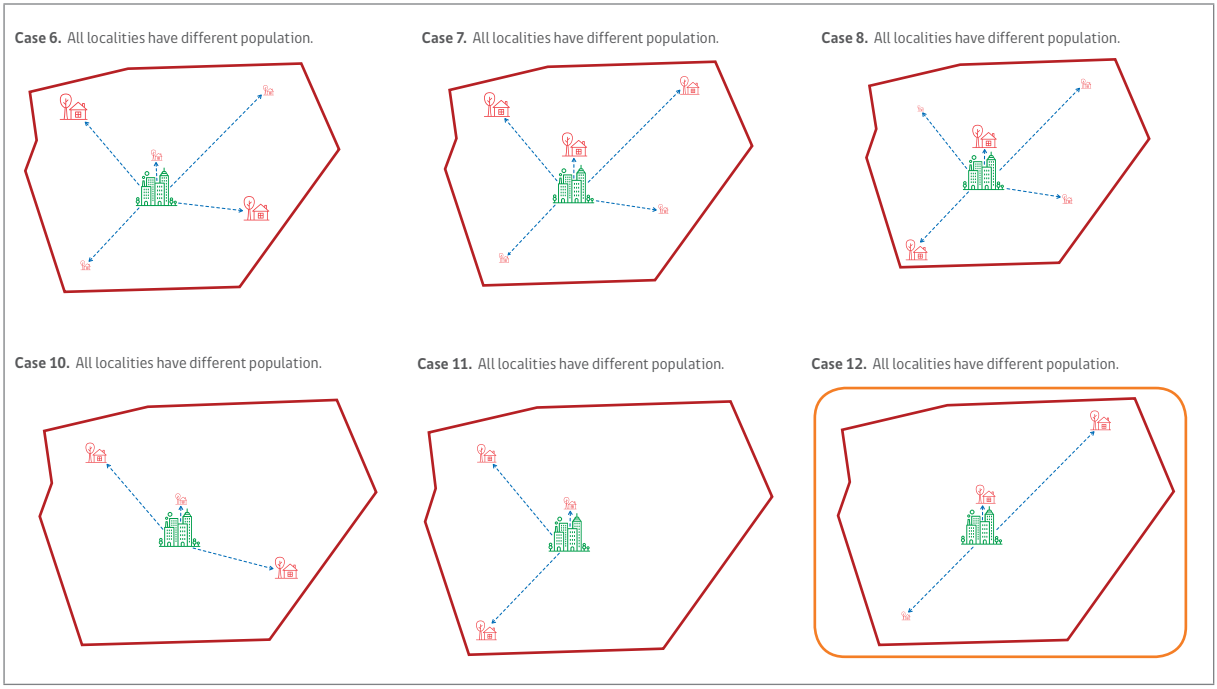


Source: the author.

The intuitiveness of the previous scheme becomes more complex when we comprehend that not all the areas where the city hall is located (county seats) have the same amount of services, and that not all areas have the same population.

Illustration 3 shows schematically which county would be the least complex to intervene, assuming that all its towns have different populations and that their municipalities have different hierarchies of services, but with different distribution patterns. The county that would be easiest to intervene is the one with the fewest towns, and where most of its population is as close to the municipality as possible.

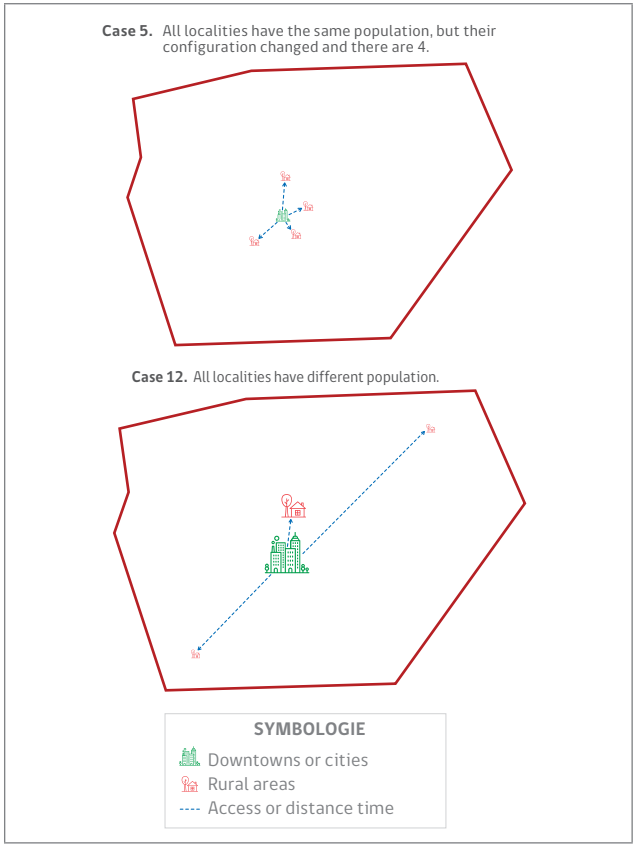
Illustration 3. Distribution scheme of localities with different populations.



Source: the author.

Illustration 6 shows schematically which county would be the least complex to intervene, with different location patterns, population distributions, and hierarchies.

Illustration 4. System complexity scheme.



Source: the author.

The idea is no longer quite so intuitive since it depends on:

1. The number of towns that are part of the county.
2. The inhabitants of each of town.
3. The accessibility of each town to the municipality.
4. The cost of transporting personnel to each town.

The classification proposal developed in this study seeks to eliminate the arbitrariness of landscape delimitation and to incorporate functionality and spatial interrelationships between city hall and the areas that it must serve.

II. OBJECTIVES

2.1 GENERAL OBJECTIVE

To determine groups or classes of counties according to the system of rural human settlements around a municipality and based on levels of accessibility, transfer costs and the concentration or dispersion of the population.

2.2 SPECIFIC OBJECTIVES

a) To calculate each town's level of physical access or spatial weighting regarding its municipality.

b) To measure concentration/dispersion from the system of localities towards the municipality.

c) To estimate transportation costs in terms of fuel from the municipality to its localities.

d) To analyze county clusters using exploratory statistical techniques.

III. METHODOLOGICAL SYNTHESIS

This study seeks to measure the concentration/dispersion of a population in reference to the county seat.

To classify the complexity of the human settlement system, the following variables will be considered:

a) The integration index of the area where the municipality is located (county seat).[3]

b) The indicator of the concentration/dispersion of rural localities around the town where the municipality is located[4].

c) The costs of travelling to all the towns of a county from the county seat. This is the fuel consumption in liters per all the kilometers that a vehicle must travel from the municipality according to a standard of 12 km/l and the average value of diesel in the county.

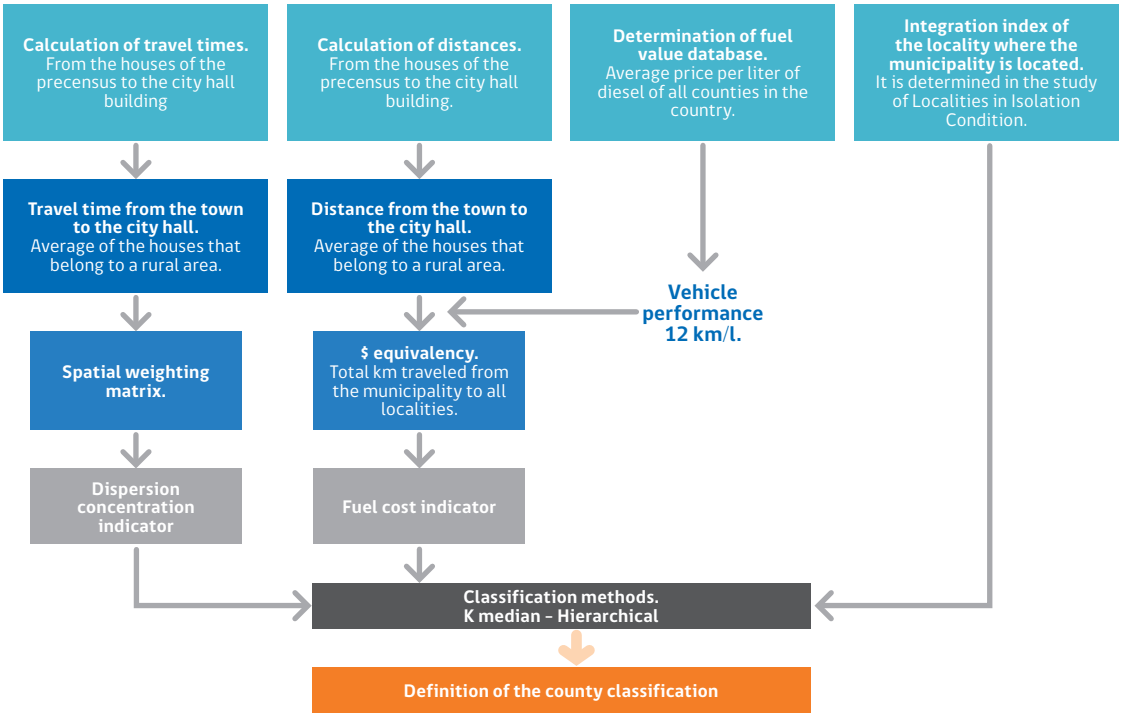
The universe of this analysis corresponds to all the Chilean counties with a locality classified by the INE as a rural entity block according to the results of the Census carried out in 2017. Due to their isolated conditions, counties on the island territories of Rapa Nui, Juan Fernández and Chilean Antarctica, will be included within the range of greater complexity.

The following illustration schematically breaks down the methodological proposal.

3 The value of the index corresponds to a preliminary calculation based on data obtained from the 2016 Pre-Census (INE) survey of homes and the 2012 methodology considered in the update of variables in the document "Identificación de Localidades de Condición de Aislamiento" (2018) Subdere https://bibliotecadigital.subdere.gov.cl

4 See Attach chapter 3.2 Indicator of Concentration/Dispersion.

Illustration 5. Methodological scheme.



Source: the author.

3.1 ENTRY DATA

This is the data necessary to create the indicators that will be incorporated into the classification methods.

1. **Calculation of travel times:** the calculation of travel time (measured in hours) from the municipality to a rural town is the average travel time of each of the rural households surveyed in the 2016 Pre-Census.

2. **Calculation of distances:** this is similar to the calculation of travel times, but kilometers are estimated instead of hours.

3. **Fuel cost database:** the price per liter of diesel in a county is the average price at all the service stations.

4. **Integration index:** this index is based on the travel time from the area where the municipality is located toa series of services proposed in the Report on Localities in Isolated Conditions (Subdere, 2012).

3.2 INDICATORS

The indicators used for the statistical classification methods are described below.

Concentration/dispersion indicator: it is a measure that seeks to determine the level of county's accessibility. It is built on the basis of its population and the accessibility of each of the localities to its county seat, understanding this as a place where local government is exercised and where basic services are preferably located.

The proposed indicator is ρi, which is the quotient between the total sum of the weighted population of each rural locality j that belongs to county i, with the total population in county i.

$$\varrho i = \frac{\sum_{j=1}^{N} W_{ij} * \rho_j}{\sum_{j=1}^{N} \rho_j}$$

Where:

ρi: concentration indicator - dispersion of county i.

$\rho i \rightarrow 1$; maximum concentration.

$\rho i \rightarrow 0$; maximum dispersion.

Wij: spatial weighting matrix of the seat of county i, in terms of the localities j which belong to county i.

ρj: population of the rural locality j, that belongs to county i.

N: number of localities in county i.

The ρi indicator shows the degree of association, measured as the concentration or dispersion from the county seat in relation to its localities.

The spatial weighting matrices are meant to show the interdependent relationships of one point in space with another.

W is a non-random matrix whose elements Wij reflect the intensity of the interdependence between a pair of zones i and j.

As a consequence, some authors have suggested using matrices directly related to the phenomenon to be explained. Bodson and Peeters (1975) introduced a general accessibility matrix (with elements calibrated between zero and one) that combines in a logistic function the influence of different communication channels between regions, such as roads, railway lines, and other communication links.

$$W_{ij} = \sum_{n=1}^{N} K_n \left\{ \frac{a}{1 + b * e^{-c_j d_{ij}}} \right\}$$

Where:

Kn: the relative importance of the communication medium n (roads, railway lines, maritime and air connections, bike paths, etc.). For this case, 1 is assumed.

N: the number of communication links. For this case, 1 is assumed. The displacement calculation network already incorporates it.

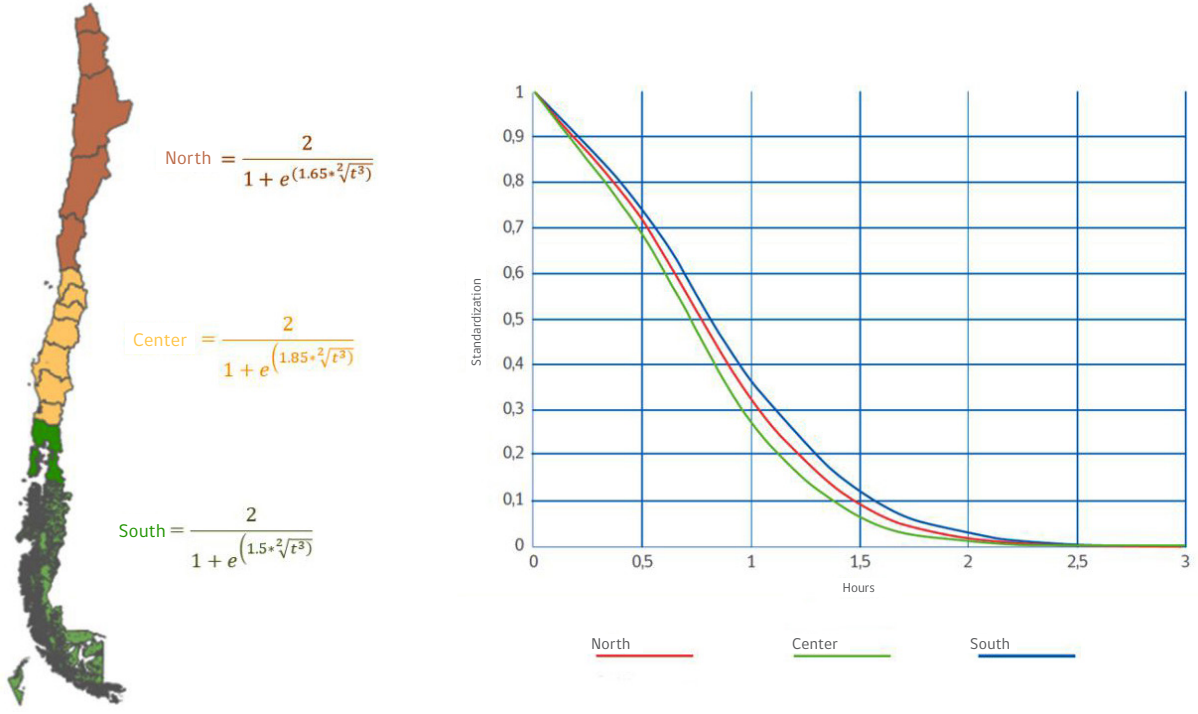
tij: the time that separates units i and j.

a, b and cj: parameters to estimate.

One of the problems with determining the general accessibility matrix Wij is determining the parameters because the validity of Wij estimates is conditioned by the appropriate estimate of said parameters.

Illustration 6 is a graphic representation of the accessibility matrix by Chile's geographic macrozones. One aspect to highlight is that from 2:30 hours of transfer and onward, the values of the matrix tend towards 0, regardless of the geographical area.

Illustration 6. Accessibility Matrix by macrozone.



Source: the author.

Fuel cost indicator: this indicator is based on the total number of kilometers that a standard vehicle[5] must travel to each locality. In other words, it is the sum of the kilometers from the municipality to all the localities. The performance of a standard vehicle is 12 km/liter. Using this, we can estimate the liters necessary to visit at least once each of the towns in the county. You can determine the cost in pesos ($) using the average price of diesel by county.

3.3 CLASSIFICATION METHODS

The techniques of unsupervised classification, also called cluster analysis, are intended to group elements into homogeneous groups based on some similarity between them. Strictly speaking, clusters are formed through a technique of homogeneity and pattern recognition.

Good cluster analysis must be:

a) **Efficient:** use as few clusters as possible.

b) **Effective:** capture all statistically and theoretically important clusters.

The purpose of cluster formation is to classify the observations (counties) and in this way find groups with similar 'response patterns'.

Among the most common exploratory classification methods are the K median and the Hierarchical.

County classification will make it possible to identify groups of counties with certain structural and spatial characteristics that make it difficult for them to serve all their localities (rural schools, removal of solid household waste, drinking water, emergency management, etc.).

5 For this case, it is understood that it has a yield of 12 kilometers per liter of diesel.

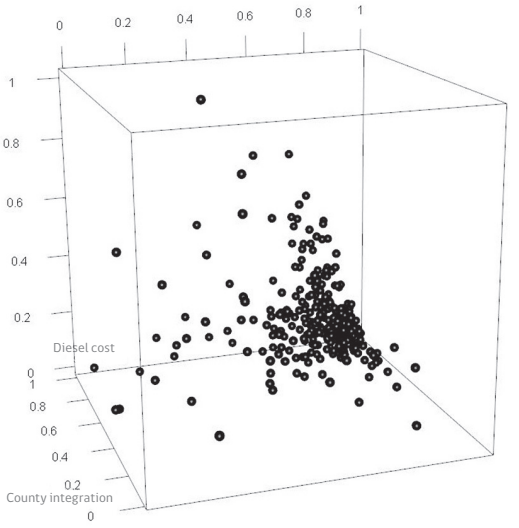
IV. EXPLANATION

4.1 ANALYSIS OF VARIABLES

Illustration 7 shows the variables RHOi[6],INTi[7], Diésel[8].

- The Integration Index of the area where the municipality is located (INTi).
- The Concentration/Dispersion Indicator for rural localities around the town where city hall is located (RHOi)
- The cost of fuel, standardized between 0 and 1 (Diesel).

Illustration 7. 3D Graph of considered variables.



Source: the author.

6 Concentration/dispersion indicator of rural localities around to the locality where the municipality is located.

7 Integration Index of the town where the municipality is located.

8 Fuel cost standardized between 0 and 1.

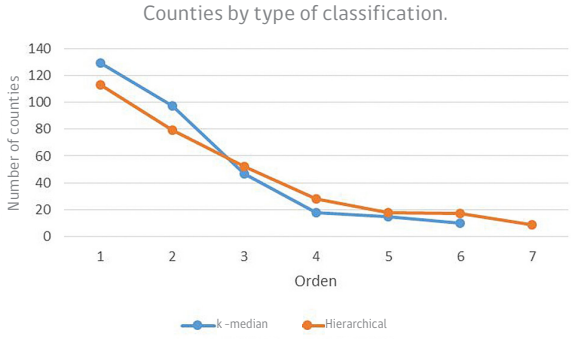
This type of figure allows us to observe if one variable increases or decreases, and how it affects the others. The scatter diagrams show the degree of correlation (not causality) between them, suggesting whether the relationship is positive, negative or null.

4.2 SELECTION OF CLASSIFICATION METHOD

The methods were developed from the k-median and hierarchical analysis algorithms.

Both methods seem consistent in that the median of their variables differs from group to group. When dealing with exploratory data analysis techniques, the usefulness of the classification must be considered. Ordering the number of counties from highest to lowest (Graph 1), we can see that the hierarchical method has a curve with a small slope, so the changes in the number of counties are not so abrupt.

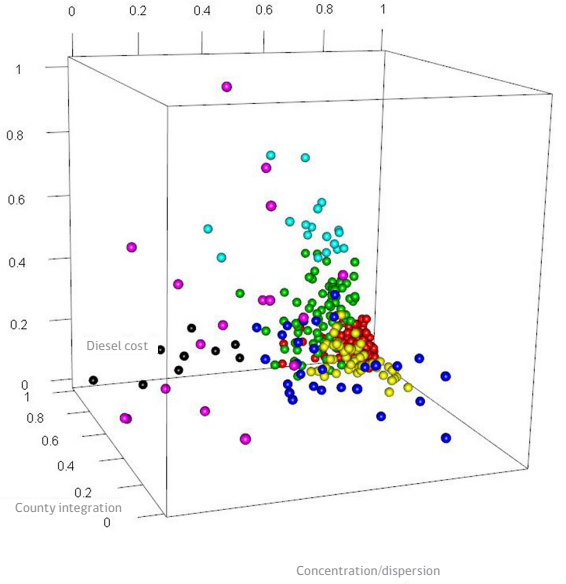
Graph 1. Decrease by type of classification.



Source: the author.

The hierarchical method of seven classes is the one that will be used for classification. In Illustration 8, you can see the result of the hierarchical analysis algorithm, where each color represents a class or group.

Illustration 8. 3D Graph of variables and hierarchical classification.



Source: the author.

Because the method assigns a label (class) randomly, the groups found must be described. To determine an order, proceed as follows:

1. The averages of the concentration/dispersion index are averaged with the municipal integration index.
2. From the average obtained in 2, subtract the diesel cost index.
3. The value obtained in 2 is ordered from highest to lowest.

Table 1 shows the proposed classification according to the classes created by the hierarchical grouping method.

Table 1. Classification according to classes of the hierarchical grouping method.

Class	Averages			Classification according to complexity
	Dispersion concentration	City hall integration	Diésel	
2	0,8980	0,9345	0,0561	Very low
7	0,8088	0,7114	0,0822	Low
3	0,7626	0,8788	0,1935	Transition to Low
1	0,3230	0,9560	0,0600	Medium
4	0,6446	0,4961	0,1515	Transition to High
5	0,7116	0,9032	0,4844	High
6	0,2821	0,4233	0,3672	Very High

Source: the author.

Table 2 describes the characteristics of the proposed classifications based on the variables used in the method.

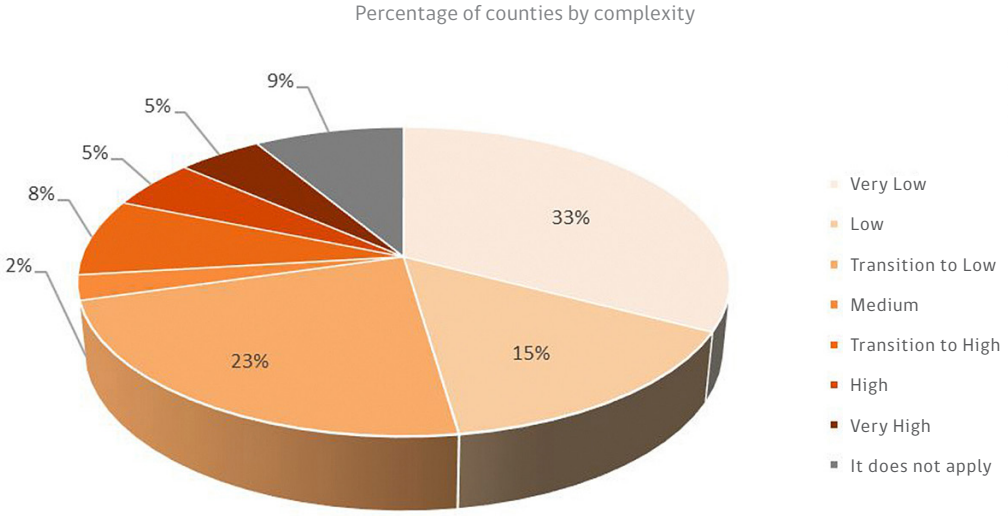
Table 2. Description of classes.

Classification	Description
Very low	Rural population concentrated around its city hall, city hall with high levels of integration of services and low diesel costs.
Low	Rural population concentrated around its city hall, city hall with medium levels of integration of services and low diesel costs.
Transition to Low	Rural population semi concentrated around its city hall, city hall with good levels of integration of services and moderate diesel costs.
Medium	Rural population scattered around its city hall, city hall with good levels of integration of services and low diesel costs.
Transition to High	Rural population semi scattered, city hall with low levels of integration and moderate diesel costs.
High	Rural population semi scattered, city hall with high levels of integration and high diesel costs.
Very High	Rural population scattered around its city hall, city hall with low levels of integration of services and high diesel costs.

Source: the author.

In Graph 2, we can see that 48% of the country's counties are classified as "low" or "very low." The counties classified as "high" or "very high" are 14% of the total.

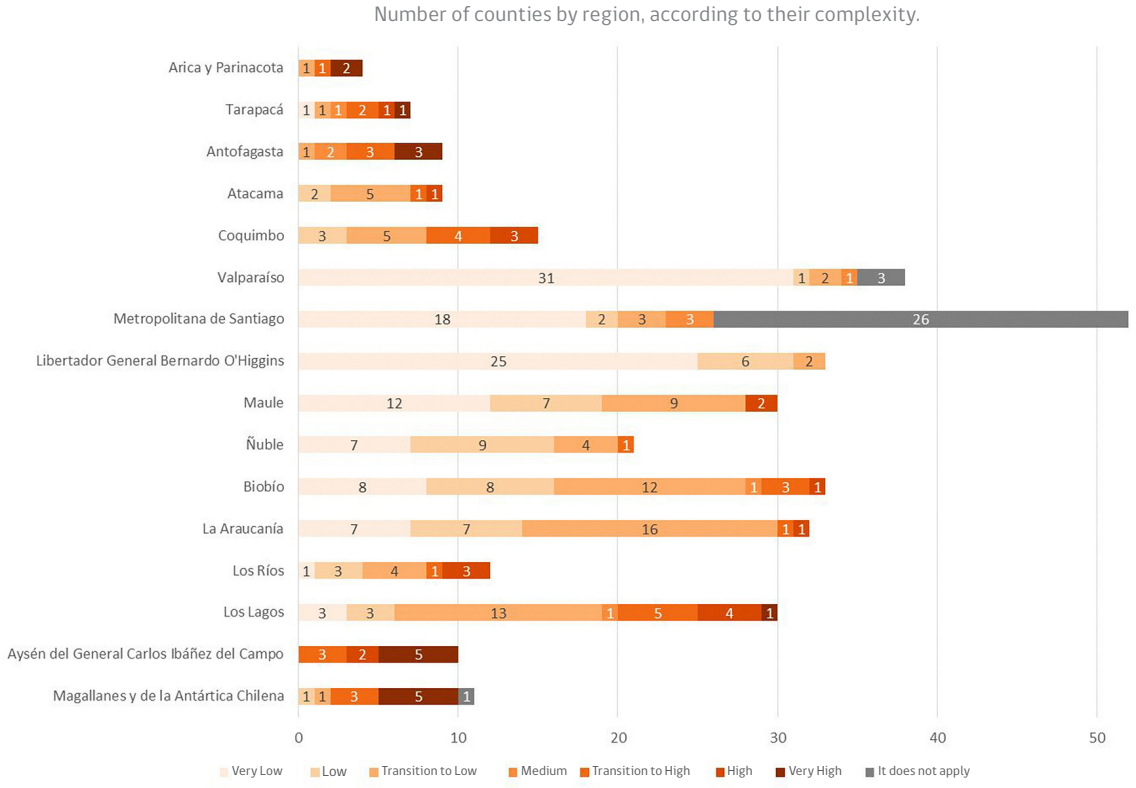
Graph 2. Distribution at national level.



Source: the author.

Graph 3 shows in absolute terms the number of counties by region according to their complexity, with the regions of Antofagasta, Los Lagos, Aysén and Magallanes, having the highest number of "highly complex" counties.

Graph 3. Complexity distribution at regional level.



Source: the author.

At the regional level, we can see that:

- The classification "very high" is mostly in the southern regions and to a lesser extent in Antofagasta and Arica and Parinacota.
- The regions of Valparaíso, O'Higgins and Maule have the largest number of counties with a "very low complexity" since most of these counties have a small surface area (km2), a good path network and a densified road network. Most of the intermediate cities around Route 5 and coastal cities are located in these regions.
- The Los Lagos Region has all the possible county classifications.
- The Aysén Region shows the conditions for a transition from "high" to "very high," which makes it the one with the greatest complexities in its human settlement systems.
- The Valparaíso (1) and Metropolitan (26) regions include counties that without rural localities, so this classification does not apply. This is equivalent to 9% of the country's counties.

V. CONCLUSIONS

1. The proposed methodology permits classification using data behavior patterns, which circumvents the biases of the photointerpretation specialist.

2. As it is not a composite or multifactorial index, it allows us to analyze the behavior of the variables separately in each group.

3. It is an alternative to density indicators. The advantage is that travel times between rural towns and the town where the municipality is located are included, taking into account distribution and connectivity.

4. By not seeking a ranking, we can talk about groups that have similarities in behavior.

5. The hierarchy of the municipality can be further specified since this study used the index of integration coming from the study on Localities in Isolated Conditions.

6. The proposed concentration/dispersion indicator allows us to synthesize the functional relations of the municipality with human settlements, incorporating the population of each.

7. This type of classification allows us to group counties with similar behavior. This will make it possible to generate different intervention strategies in their territory.

BIBLIOGRAPHY

HERRERA M, PAZ J, Y CID, J (2012) Introducción a la Econometría Espacial. Instituto de Estudio Laborales y del Desarrollo Económico.

BARONIO A, VIVANCO A, RABANAL C (2012) Introducción a la Econometría Espacial, Dependencia y Heterogeneidad.

PEREZ J J (2006) Econometría Espacial y Ciencia Regional. Red de revistas científicas de América Latina, el Caribe, España y Portugal.

PUMAN D, SAINT – JULIEN T (2010) Análisis Espacial – Las Interacciones. Analyse Spatiale. Les interactions, Armand Colin, 2ed, 2010. Serie GEOlibros Nº21 Instituto de Geografía UC, 2014.

CARVAJAL L, POCH M (2012) Identificación de Localidades en Condición de Aislamiento, Subdere (2012)

CÓRDOVA O, POCH M (2019) Actualización de Variables para la Identificación de Localidades en Condición de Aislamiento, Subdere (2019)

POCH M (2020) Asentamientos Humanos Rurales, Clasificación Comunal. Una aproximación desde el análisis espacial de la concentración y dispersión de la población, Subdere (2020)


Subsecretaría
de Desarrollo
Regional y
Administrativo
Gobierno de Chile