A spatially explicit analysis of chronic diseases in small areas: A case study of diabetes in Santiago, Chile

Ricardo Crespo (ricardo.crespo@usach.cl)
Universidad de Santiago de Chile Direccion de Investigacion Cientifica y Tecnologica
https://orcid.org/0000-0003-0211-0548

Claudio Alvarez
Universidad de Santiago de Chile

Ignacio Hernandez
Universidad de Santiago de Chile

Christian Garcia
Universidad de Santiago de Chile

Research

Keywords: A spatially explicit analysis of chronic diseases in small areas, A case study of diabetes in Santiago, Chile

Posted Date: February 25th, 2020

DOI: https://doi.org/10.21203/rs.2.24467/v1

Version of Record: A version of this preprint was published on June 23rd, 2020. See the published version at https://doi.org/10.1186/s12942-020-00217-1.

Title: A spatially explicit analysis of chronic diseases in small areas: A case study of diabetes in Santiago, Chile

Ricardo Crespo[a,*] , Claudio Alvarez[b], Ignacio Hernandez[c], and Christian García[d]

[a]Departamento de Ingeniería Geográfica, Universidad de Santiago de Chile
E-mail: ricardo.crespo@usach.cl

[b]Departamento de Ingeniería Geográfica, Universidad de Santiago de Chile
E-mail: claudio.alvarezs@usach.cl

[c]Departamento de Ingeniería Geográfica, Universidad de Santiago de Chile,
E-mail: ignacio.hernandez.c@usach.cl

Facultad de Ciencias Médicas
E-mail: christian.garcia.c@usach.cl

[*]Corresponding author

Avenida Libertador Bernardo O'Higgins, 3363

Estación Central

Santiago

Chile

1. Introduction

The increase of chronic diseases in the past decades is a global concern. In 2016, chronic conditions caused the death of 40.5 million people worldwide [1], equivalent to 71% of the total causes of death. This increasing trend is particularly severe in developing countries [2]. Chronic diseases affect the quality of life and have a substantial economic impact on public and private health systems. Naturally, pressure is generated on governments concerning the public policies required to prevent and control this problem. Most prevention plans focus on promoting changes in habits related to alcohol consumption, tobacco use, nutrition, and exercise, among others. Besides, access to treatment is usually associated with the capacity and the spatial distribution of providers of the health network. The concept of a health network is understood as a group of primary, secondary, and tertiary healthcare establishments in a city.

In this context, public policies aimed at improving prevention plans and optimizing the allocation of resources in health networks should be designed specifically for the social and economic reality of the population by a multi-dimensional approach. One way to tackle this challenge is by exploring the link between the sociodemographic attributes that characterize a community, its risk of suffering chronic diseases, and its accessibility to treatment [3], [4] For this reason, and to optimize the function of the health network, particularly in large cities, it becomes necessary to incorporate the spatial domain to study this link [5], [6]. Exploring the spatial heterogeneity of socioeconomic and health indicators reveals a valuable tool to design and improve a health network [7], [8].

Due to the inherent complexity of cities, spatial heterogeneity is best explored at the level of a small area (neighborhood level). However, health indicators are rarely available in small areas. For this reason, we suggest the use of spatial microsimulation, a powerful statistical technique used primarily to disaggregate sociodemographic data at different geographical scales. Microsimulation (spatial and non-spatial) has been present in the area of health since the 1970s in various studies, such as fertility, private health systems, and cancer [9]–[12]. The vast

majority of these studies used census data and specific surveys for the generation of synthetic data. New research has been applied to study the effect of population aging and the cost associated with public policies [13], including the need to improve the health services for diseases such as diabetes [14] and dementia in the elderly [15].

Also, due to the inherent complexity of urban systems, and with the purpose of approaching the problem from a multidimensional perspective, we propose the use of advanced spatial clustering methods to find spatial sociodemographic patterns that link the population with the prevalence of the most frequent chronic diseases. Specifically, we suggest the use of a clustering method called self-organizing maps (SOM) introduced by [16] to determine high-risk populations and their location. Self-organizing maps, unlike classical clustering methods, incorporate a learning component through neural networks, which makes it more appropriate to model complex adaptive systems, that is, constantly changing systems, such as cities. For this reason, the SOM method has gained popularity in population studies of large cities where high sociodemographic heterogeneity is present. SOM has been previously used in demography and in public health studies [17]–[20].

In summary, the main objective of this study is to identify significant clusters of the population with different prevalences of diabetes (as an example of chronic disease) and different sociodemographic patterns in a spatially explicit manner, for small areas of blocks. By doing so, we attempt to suggest spatially specific policies to optimize the functioning of heath network and prevention campaigns.

2. Methods

2.1 Study area

For the case study, we chose the city of Santiago, Chile. Santiago is the capital of and the largest city in Chile. The city lies in the central valley of the country and has an approximate population of seven million inhabitants (35% of Chile's total population). Given its stable economic growth since the 90s, Chile became an

OECD country in 2010. Despite being classified as a high-income country by the World Bank, Chile is ranked the highest in inequality among OECD countries with a Gini index of 0.47 [21]. Chile is classified as a country with an efficient and well-organized health system, but challenges like the increase of chronic diseases and the aging population could have a severe impact on the health system and the country's economy [22]. As for diabetes, Chile has the sixth highest adult diabetes prevalence among OECD countries, where about 10% of Chilean adults were diabetics (citar). Similarly, the Ministry of Health of Chile, through its 2016–2017 National Health Survey, reported that 12.3% of Chileans suffer from diabetes (about 1.8 million people). Such prevalence increases to 30% in the elderly [23].

Chile is a country with high rates of tobacco and alcohol consumption compared to other OECD countries. Likewise, the obesity rate is 34.4% in adults and 44.5% in children. Nine out of the first ten causes of death in 2017 were associated with chronic diseases [24]. Moreover, the rise of chronic diseases in the elderly population in the last decade suggests an increase in the demand for medical care in the next decades.

2.2 Data

For characterizing the population, and as input for the microsimulation, we used data from the 2017 Census at the smallest scale available: census zones. These areas usually comprise between 1,000 and 5,000 people residing in a group of neighboring blocks. The city of Santiago comprises 34 districts (comunas), with an average population of 176,000 inhabitants each, and 1,643 census zones, with a mean population of 3,600 inhabitants. To obtain health data, we used the Chilean Socioeconomic Survey of 2017 (CASEN). The CASEN survey is composed of seven modules of characterization of individuals and families, namely resident registry, education, work, income, health, identities, and housing characteristics. Due to the inherent complexity of cities with high socioeconomic heterogeneity, CASEN used a probabilistic, stratified, multistage, and conglomerate sample to achieve a good representation of the population according to socioeconomic

diversity in each district. CASEN used the 2017 census cartographic mapping for the sampling method. Detailed information on the sample design can be found in the official document of the Ministry of Social Development, the public agency in charge of carrying out the survey [25].

Given that diabetes is a rising burden in Chile, and because of an adequate response rate to the survey, we selected diabetes in adults as the chronic disease in our study. Chile included diabetes as one of the diseases that qualify for universal access, expenditure protection, and a guarantee of access to diagnosis and treatment. Coverage of diabetes and another thirty-two high-burden diseases was established by law in the health reform of 2004 and applied to public and private insurances [26]. For the purpose of measuring diabetes, we used the CASEN question that requested auto-reporting of diabetes treatment in the last twelve months, which is available at the district geographical level.

2.3 Spatial microsimulation: methodology to disaggregate prevalence of diabetes at census zone level.

To disaggregate the prevalence of diabetes from district to census zone level, we used spatial microsimulation. This technique generates a synthetic population by combining census data available at the smaller area scale with data from socioeconomic surveys available at the larger geographical level [27]. The synthetic population can be seen as an enriched version of the census data containing additional socioeconomic attributes, which are typically variables associated with income or health indicators [28], [29] For this purpose, we used years of education, age, and gender as the link variables between the census and the CASEN survey. Spatial microsimulation was performed using the iterative proportional fitting algorithm (IPF) [30] by which health attributes of each simulated individual are calculated using contingency tables from both data sets.

2.4 Self-organizing maps (SOM)

SOM is an unsupervised neural network method that operates by clustering multidimensional input data and reduces them to a two-dimensional representation. Clusters are formed based on similarities and patterns of a series of attributes of the input data. This method operates similarly to the traditional *k-means* clustering method in that it looks for similarities by calculating the Euclidean distance between the attributes of the input data. However, SOM's unsupervised neural network algorithm achieves better classification in those processes where the high multidimensionality of the attributes makes it difficult to classify and distinguish between one cluster and another. This type of behavior is frequently observed in demographic and health phenomena where people with similar sociodemographic characteristics have different tendencies to suffer from chronic diseases or people with different sociodemographic characteristics may exhibit similar trends.

Figure 1 shows a concise representation of how SOM operates obtained from [31]. The input layer includes *n* vectors of input data, each one containing a set of *m* attributes. The output layer is depicted by the colored grid comprising K neurons, each one represented by a multidimensional weight vector $\vec{w}_i$ of length *m*.

Figure 1. SOM map structure (with permission of authors [31]).

The purpose of the algorithm is to closely group together across the 2D grid representation those neurons with a similar combination of input attributes. In order to quantify the level of similarity, the algorithm calculates for each vector of input data $\vec{x}_n$ the Euclidean distance $d(X, W_i)$ between its standardized attributes and each $\vec{w}_i$ as follows:

$$d(X,W_i) = \|X - W_i\| = \sqrt{\sum_{j=1}^{m}(x_j - w_{ij})^2} \quad (1)$$

Where *i: 1,…,K*.

Each input vector is assigned the neuron associated with the smallest Euclidian distance (the winning neuron). Finally, clusters are formed by grouping together neurons with similar $\vec{w}_i$ values; that is, similar colors in Figure 1. A key point in the SOM process is the update of the initial values of the weights of each neuron. As these values are first unknown, they must be initialized, usually by assigning small random values. Thus, to find meaningful clusters, these values must be updated at each iteration. Basically, the weight vectors of the winner and its neighboring units in the output space are adjusted to become more representative of the features that characterize the input space.

This clustering method is applied to each of the 1,643 census zones (index *m,* as indicated above). For this reason, each census zone must be previously characterized on the basis of a series of attributes that correspond to the input variables of the SOM algorithm. Each census zone is characterized according to the following five attributes (index *n* as indicated above).

1) Education: Percentage of people with professional studies

2) Income: Percentage of people with low income, where low income is defined as that which is below US$360, corresponding to the median of income distribution in the study area.

3) Sex: Percentage of men.

4) Age: Percentage of people in the age ranges of 30–45 years and 46–60 years—these age ranges leave children and older adults out of the analysis. Diabetes in children differs from the adult. Children develop mostly diabetes mellitus (DM) type I caused by a reduced insulin production, the hormone that controls blood sugar,

due to genetic and immune causes. On the other hand, DM II developed by adults and the elderly is associated with the resistance to insulin in the body peripheral cells related to lifestyle and habits. Accordingly, including children in the analysis could lead us to draw erroneous conclusions. We have also excluded older adults because they are probably not receiving income, which could also lead to erroneous conclusions. Finally, we have concentrated on adults over the age of 29, an age by which they would have finished their professional studies if they had taken them.

5) Diabetes: Percentage of people who reported have been under treatment for diabetes in the last twelve months.

3. Results

To proceed with the SOM method, we used the *Kohonen* package of the R statistical software and followed the guidelines of [32]. Due to the great sociodemographic heterogeneity usually present in large cities, we set the number of clusters for the SOM to twelve. However, the SOM as an unsupervised neural network algorithm can find clusters that have a solely statistical, but not a clear, sociodemographic interpretation. For this reason, and for the purpose of illustrating this methodology in a simple manner, we selected three of the 12 clusters that were interpretable for the purpose of our study. These clusters are described in Table 1. All percentages are expressed in relation to the total population of column 2. We omitted census zones assigned by SOM to the remaining nine clusters that we found uninterpretable.

Table 1. Description of selected clusters

Cluster	Number of people	Male population [%]	Diabetes auto report [%]	Low Income [%]	30-45 years old [%]	46-60 years old [%]	Professionals [%]
1	322427	46.26	3.01	41.90	30.38	26.45	39.23
2	1768939	48.66	4.05	63.25	31.28	26.09	10.84
3	1944443	48.65	4.03	64.26	30.98	26.37	9.46

Clearly, cluster 1 is representative of people with more economic and educational resources than people from clusters 1 and 2. Though slightly on average, people from cluster 2 exhibit more resources than people from cluster 3 to cope with the disease. A more disaggregated detail of clusters can be observed in Table 2. This table summarizes information from all census zones that make up each cluster, arranged according to their quartiles (Q), where the index *n* corresponds to the number of census zones in each cluster. Some important differences can de found, for instance, in quartile four (Q4) for the attributes male population 30-45 years old across all clusters.

Table 2. Detail information of selected clusters

Cluster	Quartile	Diabetes auto report [%]	Male population [%]	Low Income [%]	30-45 years old [%]	46-60 years old [%]	Professionals [%]
1 (n=83)	Q1	2.56	44.81	36.44	25.97	24.02	38.34
	Q2	2.93	45.78	37.86	29.67	25.87	43.64
	Q3	3.30	47.34	41.10	34.57	27.66	48.91
	Q4	4.51	57.12	68.69	58.92	34.71	60.15
2 (n=495)	Q1	3.80	47.82	60.72	28.44	24.51	4.89
	Q2	4.20	48.64	65.08	30.38	25.97	7.91
	Q3	4.50	49.56	67.79	32.51	27.12	13.40
	Q4	5.50	90.18	71.41	75.46	35.96	60.51
3 (n=484)	Q1	3.82	47.83	61.89	28.39	24.35	4.35
	Q2	4.16	48.61	65.89	30.25	26.29	7.02
	Q3	4.50	49.38	68.31	32.31	27.97	12.08
	Q4	7.25	70.45	74.87	53.39	38.10	58.67

As a first exploratory analysis, the spatial distribution of the number of people who have been or were treated for diabetes according to the CASEN 2017 survey is presented in Figure 2. Furthermore, and as a reference for a preliminary socio-economical exploratory analysis of the city, Figure 3 shows the spatial distribution of the average per capita income obtained from [27].

Figure 2. Spatial distribution of diabetes auto report.

Figure 3. Spatial distribution of per capita income (with permission of authors [27]).

An important spatial pattern links the area of the city with the highest income per capita with a smaller number of people who responded that they had been under treatment for diabetes in the previous year. This may be due to the fact that people with higher incomes have better access to treatment, and it may also indicate that the same people have fewer chances to suffer from diabetes. However, the heterogeneous spatial variability of the number of people with diabetes in census zones in Figure 2 greatly differs from the more homogeneous spatial patterns of income in Figure 3. These patterns support the multifactorial nature of diabetes. In other words, although it is true that per capita income has an important correlation with the development of diabetes, it does not by itself explain the spatial heterogeneity seen in Figure 2. It is therefore necessary to include in the clustering other sociodemographic variables.

Clusterization of adults with diabetes in the city of Santiago, Chile

Figure 4 shows the three selected clusters on the 2D grid ouput of the SOM method applied to the case study. We selected a 12 x 12 grid containing 144 neurons. For simplification, the algorithm assigns one simple color to each cluster.

Each neuron contains information of a set of census zones, which are not necessarily neighboring zones.

Figure 4. 2D grid output of the SOM method.

The geographical representation of the SOM results is shown in Figure 5, which includes two zones without census information (corresponding to parks), and unanalyzed areas that correspond to the clusters omitted for this analysis. Figure 5 shows the spatial distribution of the clusters detailed in Table 1 and the location of health services. We included public primary care and specialized centers. Primary care includes centers that perform ambulatory treatment and control by general doctors and other professionals (e.g. nurses). Specialized centers include hospitals and ambulatory centers where specialists are available.

Figure 5. Geographical representation of the SOM results.

The highest income sector of the city, which also has the best education level, concentrates the smallest number of people with auto-reported diabetes located mostly in the northeast of Santiago (Cluster 1). Cluster 2 represents areas with a high percentage of people treated for diabetes and lower income and educational level than people from cluster 1. Cluster 3 concentrates the most critical population, with the highest number of people with low income, the lowest number of professionals, and the highest number of people under diabetes treatment. Both Clusters 2 and 3 are geographically distributed throughout the city except in its central and north-eastern areas. However, Cluster 3 has a slight tendency to be located in the peripheral areas of the city. In relation to health services, although it is true there is extended coverage of primary services in most areas of the city, there are uncovered areas in the periphery associated with Clusters 2 and 3. The circles on the map enclose some of the census zones with a lower density of health services compare with most of the zones located in the west and in the city

center. The exception corresponds to zones located toward the northeast of the city, which comprises mostly census zones of cluster 1.

4. Discussion

In this study, we intend to analyze people having diabetes of type II. All the information regarding diabetes cases was obtained from the CASEN socio-economic survey. This survey does not distinguish between diabetes types I and II in the corresponding answer. However, the fact that about 90 to 95 of people with diabetes correspond to Type II diabetes [33] provides us with the confidence to address the problem in a statistically meaningful manner. It is also important to bear in mind that the auto-reported diabetes attribute described in point 5, page 7, excludes those people with diabetes but not in treatment in the last twelve months, perhaps due to accessibility difficulties.

A quick analysis of Figure 2 and the spatial distribution of the health services reveal that most of the areas with high diabetes auto report (see range greater than 125 in Figure 2 appears to be well-covered by health services. Thus, a foregone recommendation (without using SOM) would be to densify the distribution of health services in those areas indicated by the circles. However, the great advantage of the SOM method lies in its ability to characterize the specific needs of patients in a spatially explicit manner, which allows for optimization and/or improvement of aspects of public health. For example, most of the census areas comprised by the circles belong to Cluster 3, the most critical cluster. People who reside in this cluster have the lowest educational level of the three analyzed clusters. This suggests that, besides the economic measures to improve the service, prevention campaigns should be preferably oriented to improve people's lifestyles. Educational level can be used as a proxy variable for the measurement of people's lifestyles, such as healthy eating, sports, alcohol, and tobacco, as inputs for local prevention campaigns [34]. Clearly, this analysis can be extended to all existing health services, even in more dense zones. SOM results may also allow improvement of medicine distribution according to the types of patients in each

geographical area and filling the gap of general practitioners usually present in the healthcare network in cities [8].

Results from SOM are sensitive to the number of neurons and clusters in the grid, which are usually selected in an arbitrary manner. However, varying these factors allows for a more detailed exploration of some relevant and subtle differences between clusters. For example, although clusters 2 and 3 appear to be only slightly different, information from Table 2 shows how exploring even small differences may have an important impact on policymaking. Information in quartile 4 (Q4) for the attribute 30–45 years [%] for clusters 2 and 3 reveals the existence of census zones in cluster 2 in which nearly 75% of people are between the ages of 35 and 45, compared to approximately 53% in cluster 3. A larger quantity of people in the range of 30 to 45 years old will lead to a greater future burden on the state for expenses in the treatment of diabetes. Another similar analysis can be done for 7.25% of the diabetes auto-report for cluster 3, which, according to quartile 4, compares to the lower values of 5.5% and 4.51% for clusters 2 and 1, respectively. Clearly, such differences may also have implications in the design of geographically specific prevention campaigns for the allocation of medical resources. It is worth recalling at this point that each census zone contains approximately 3,600 people or about 100 families.

It is also worth highlighting the low density of health services observed in cluster 3. Though cluster 3 corresponds to people with more educational and economic resources and access to private health establishments, Table 2 shows a group of census zones of cluster 3, where about 68% of low-income residents and 4.5% of residents who auto report diabetes can be observed. As with the abovementioned cases, these differences can be addressed, if necessary, by increasing the number of clusters handled by algorithms to split existing clusters into smaller clusters containing more specific information.

5. Conclusion

In this study, we have used advanced spatial statistical methodologies to address a type of supply and demand problem in public health in Santiago, including people with diabetes and the whole network of health services. The results have allowed us to corroborate the importance of the spatial factor in the analysis of chronic diseases as a way of suggesting differentiated solutions to spatially explicit problems. For example, our methodology has allowed us to identify geographical areas that require an increase in the coverage of health services, especially those areas inhabited by people with low economic resources or a low level of education, that is, people who have greater difficulties in facing the disease without state support. Our approach has also allowed us to understand that the current criteria for locating the health network would be based primarily on population density and/or the number of people reported with diabetes and only, to a lower extent, on the ability of patients to cope with the disease from a sociodemographic perspective. We believe that a more exhaustive analysis of the clusters would allow us to identify small areas with a high rate of people undergoing treatment for diabetes, but with better resources and capabilities (economic and educational) than people from economically and educationally much more vulnerable areas.

We also hope that the methodological aspects developed in this study at the neighborhood level (small areas) can be complemented with other advanced methods of spatial statistics, such as graph or spatial network theory, to contribute to the optimization of supply chains of public health services. Finally, it is important to point out that, for illustrating this methodology, we have selected five input attributes for the execution of the SOM. However, one of the advantages of the SOM method is precisely its multidimensional handling of the input data, so for future studies, it is suggested to incorporate variables, such as ethnicity, weight, life habits, and medical history, among others.

6. Declarations

Ethics approval and consent to participate

This study used open-access public survey data and no need to obtain an ethical approval.

Consent of publication

This study was carried out using the 2017 CASEN socioeconomic survey and the 2017 Census data from Chile. Both are freely available.

Availability of data and materials

CASEN data can be obtained from the Ministry of Social Development and Family (http://observatorio.ministeriodesarrollosocial.gob.cl/)

Census data can be obtained from the National Institute of Statistics (https://www.censo2017.cl/)

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Acknowledgments

Not applicable.

Authors' contributions

RC designed the study and drafted the first manuscript version. CA and IG carried out the analysis. CG revised the manuscript critically and participated in the study design. All authors read and approved the final manuscript.

7. References

[1] WHO, “World health statistics 2018: monitoring health for the SDGs, sustainable development goals,” Geneve, 2018.

[2] R. Nugent, “Chronic diseases in developing countries: Health and economic burdens,” *Ann. N. Y. Acad. Sci.*, vol. 1136, no. 1, pp. 70–79, 2008.

[3] S. Rahman *et al.*, “Chronic disease and socioeconomic factors among uninsured patients: A retrospective study,” *Chronic Illn.*, 2019.

[4] N. Minicuci *et al.*, “Sociodemographic and socioeconomic patterns of chronic non-communicable disease among the older adult population in Ghana,” *Glob. Health Action*, 2014.

[5] D. U. Pfeiffer, T. P. Robinson, M. Stevenson, K. B. Stevens, D. J. Rogers, and A. C. A. Clements, *Spatial Analysis in Epidemiology*. New York: OXFORD University Press, 2008.

[6] R. Roquette, M. Painho, and B. Nunes, “Spatial epidemiology of cancer: A review of data sources, methods and risk factors,” *Geospat. Health*, vol. 12, no. 1, 2017.

[7] J. R. Barnett, “Does the Geographic Distribution of Physicians Reflect Market Failure?: An Examination of the New Zealand Experience, 1981–87,” *Environ. Plan. A Econ. Sp.*, vol. 25, no. 6, pp. 827–846, 1993.

[8] M. McIsaac, A. Scott, and G. Kalb, “The supply of general practitioners across local areas: Accounting for spatial heterogeneity,” *BMC Health Serv. Res.*, vol. 15, no. 450, 2015.

[9] S. Roy, “Demography of sterilization: Indian experience,” *Janasamkhya*, vol. 2, no. 1, pp. 51–65, 1984.

[10] M. G. Santow, “A microsimulation of Yoruba fertility,” *Math. Biosci.*, vol. 42, no. 1–1, pp. 93–117, 1978.

[11] H. . Chernick, M. . Holmer, and D. . Weinberg, “Tax policy toward health

insurance and the demand for medical services," *J. Health Econ.*, vol. 6, no. 1, pp. 1–25, 1987.

[12] D. . Parkin, "A computer simulation model for the practical planning of cervical cancer screening programmes," *Br. J. Cancer*, vol. 51, no. 4, pp. 551–558, 1985.

[13] U. Schneider and J. Kleindienst, "Monetising the provision of informal long-term care by elderly people: estimates for European out-of-home caregivers based on the well-being valuation method," *Heal. Soc. Care Community*, vol. 24, no. 5, pp. e81-91, 2016.

[14] D. Schofield *et al.*, "The costs of diabetes among Australians aged 45-64-years from 2015 to 2030: Projections of lost productive life years (PLYs), lost personal income, lost taxation revenue, extra welfare payments and lost gross domestic product from Health&WealthMOD2030," *BMJ Open*, vol. 7, no. 013158, 2017.

[15] P. Singh, R. Hussain, A. Khan, L. Irwin, and R. Foskey, "Dementia care: Intersecting informal family care and formal care systems," *J. Aging Res.*, 2014.

[16] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[17] M. Collan, T. Eklund, and B. Back, "Using the Self-Organizing Map to Visualize and Explore Socio-Economic Development," *EBS Rev.*, vol. 22, no. 1, pp. 6–15, 2007.

[18] H. G. Basara and M. Yuan, "Community health assessment using self-organizing maps and geographic information systems," *Int. J. Health Geogr.*, vol. 7, no. 67, 2008.

[19] Y. Mehmood, M. Abbas, X. Chen, and T. Honkela, "Self-Organizing maps of nutrition, lifestyle and health situation in the world," in *Lecture Notes in Computer Science,vol 6731*, 2011.

[20] K. Wickramasinghe, D. Alahakoon, P. Schattner, and M. Georgeff, “Self-organizing maps for translating health care knowledge: A case study in diabetes management,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Computer Science, Advances in Artificial Intelligence, vol 7106*, 2011.

[21] “What’s happening to income inequality?,” 2015, pp. 31–39.

[22] OECD, *OECD Reviews of Public Health: Chile: A Healthier Tomorrow*. Paris, 2019.

[23] Minsal, “Encuesta Nacional de Salud 2016-2017 Segunda entrega de resultados,” 2018. [Online]. Available: https://www.minsal.cl/wp-content/uploads/2018/01/2-Resultados-ENS_MINSAL_31_01_2018.pdf. [Accessed: 17-Feb-2020].

[24] IHME, “Chile Profile,” 2018. .

[25] M. D. Social, “Casen 2017, metodología de diseño muestral,” 2018. [Online]. Available: http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Diseno_Muestral_Casen_2017_MDS.pdf.

[26] V. Valdivieso and J. Montero, “El plan AUGE: 2005 al 2009 Health care reform in Chile,” *Rev. Med. Chil.*, vol. 138, no. 8, 2010.

[27] R. Crespo and I. Hernandez, “On the spatially explicit Gini coefficient: the case study of Chile -a high-income developing country,” *Lett. Spat. Resour. Sci.*, 2020.

[28] R. Tanton and K. Edwards, Eds., *Spatial Microsimulation: A Reference Guide for Users*. Springer, 2013.

[29] A. Zaidi, A. Harding, and P. Williamson, Eds., *New Frontiers in Microsimulation Modelling: Introduction*, 1st ed. Routledge, 2009.

[30] R. Lovelance and M. Dumont, *Spatial Microsimulation with R*. Chapman and Hall/CRC, 2017.

[31] M. Carrasco and R. Brunner, “SOMz: photometric redshift PDFs with self organizing maps and random atlas,” *Mon. Not. R. Astron. Soc.*, vol. 438, no. 4, 2013.

[32] R. Wehrens and L. Buydens, “Self- and Super-Organizing Maps in R: The kohonen Package,” *J. Stat. Softw.*, vol. 21, no. 5, pp. 1–19, 2007.

[33] J. Sapunar, “Espidemiiología de la Diabetes Mellitus en Chile,” *Rev. Médica Clin. Las Condes*, vol. 27, no. 2, pp. 146–151, 2016.

[34] ACP, “How is a Shortage of Primary Care Physicians Affecting the Quality and Cost of Medical Care? A Comprehensive review. White Paper,” Philadelphia, 2008.

Figures

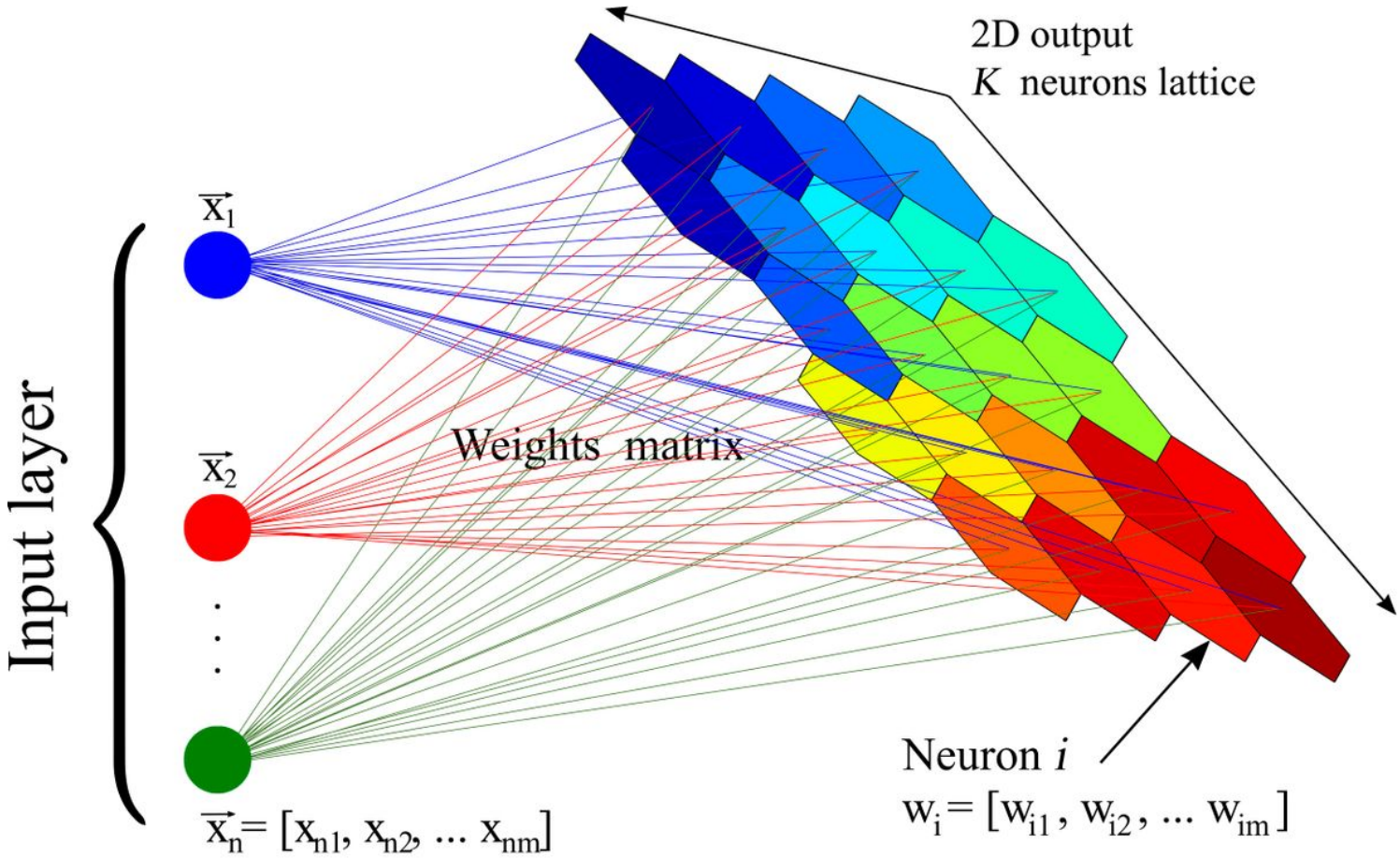

Figure 1

SOM map structure (with permission of authors [31]).

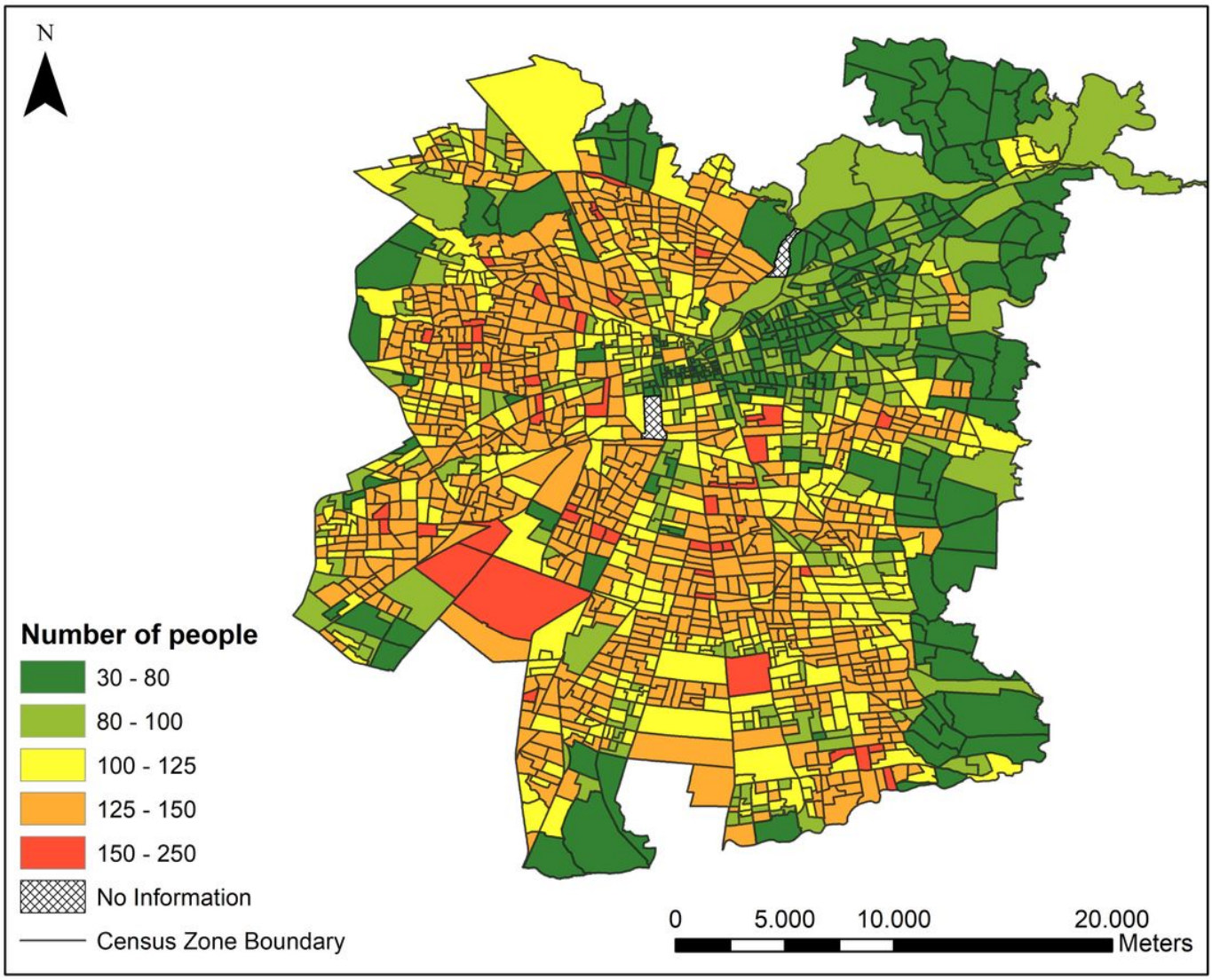

Figure 2

Spatial distribution of diabetes auto report.

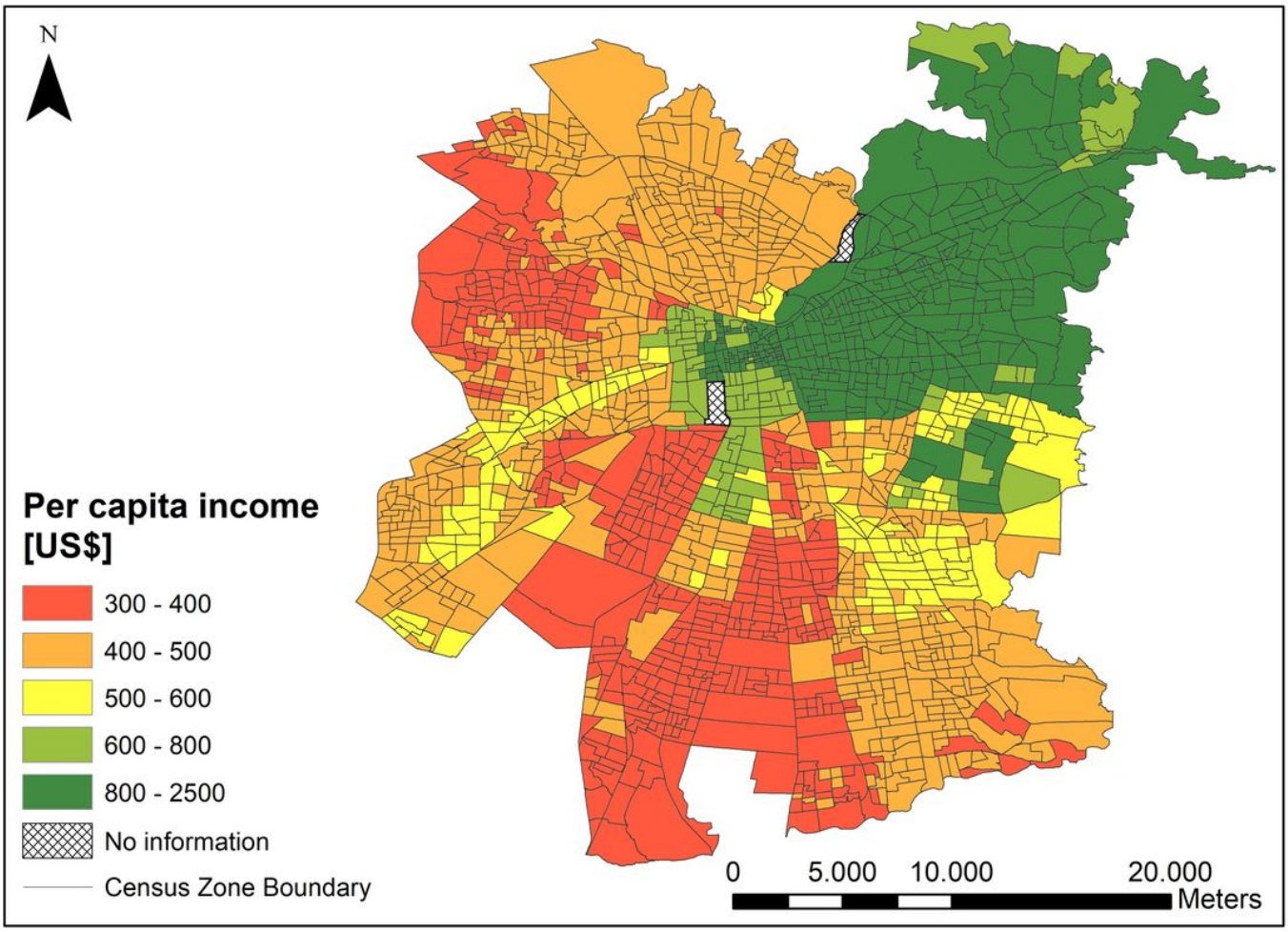

Figure 3

Spatial distribution of per capita income (with permission of authors [27]).

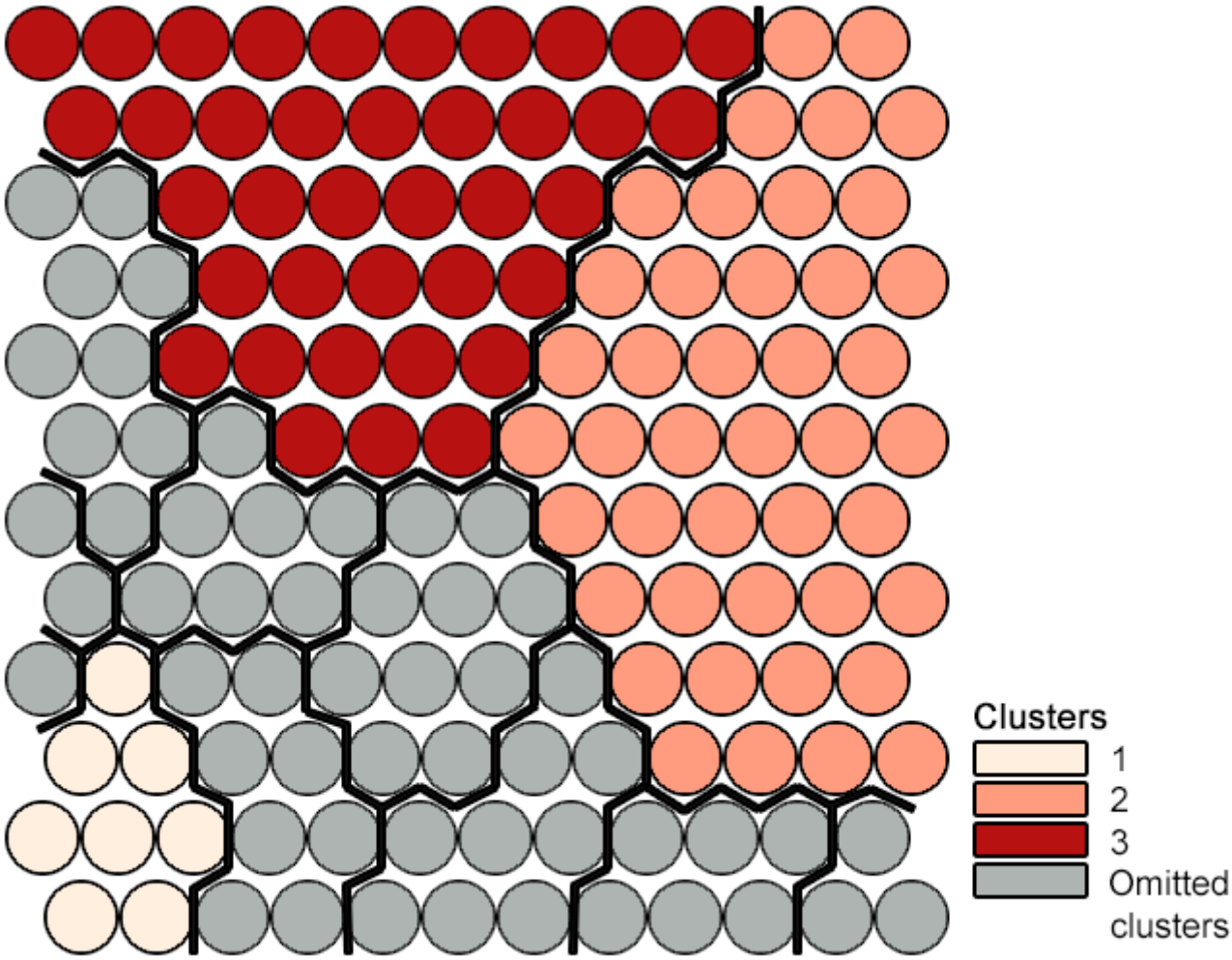

Figure 4

2D grid output of the SOM method.

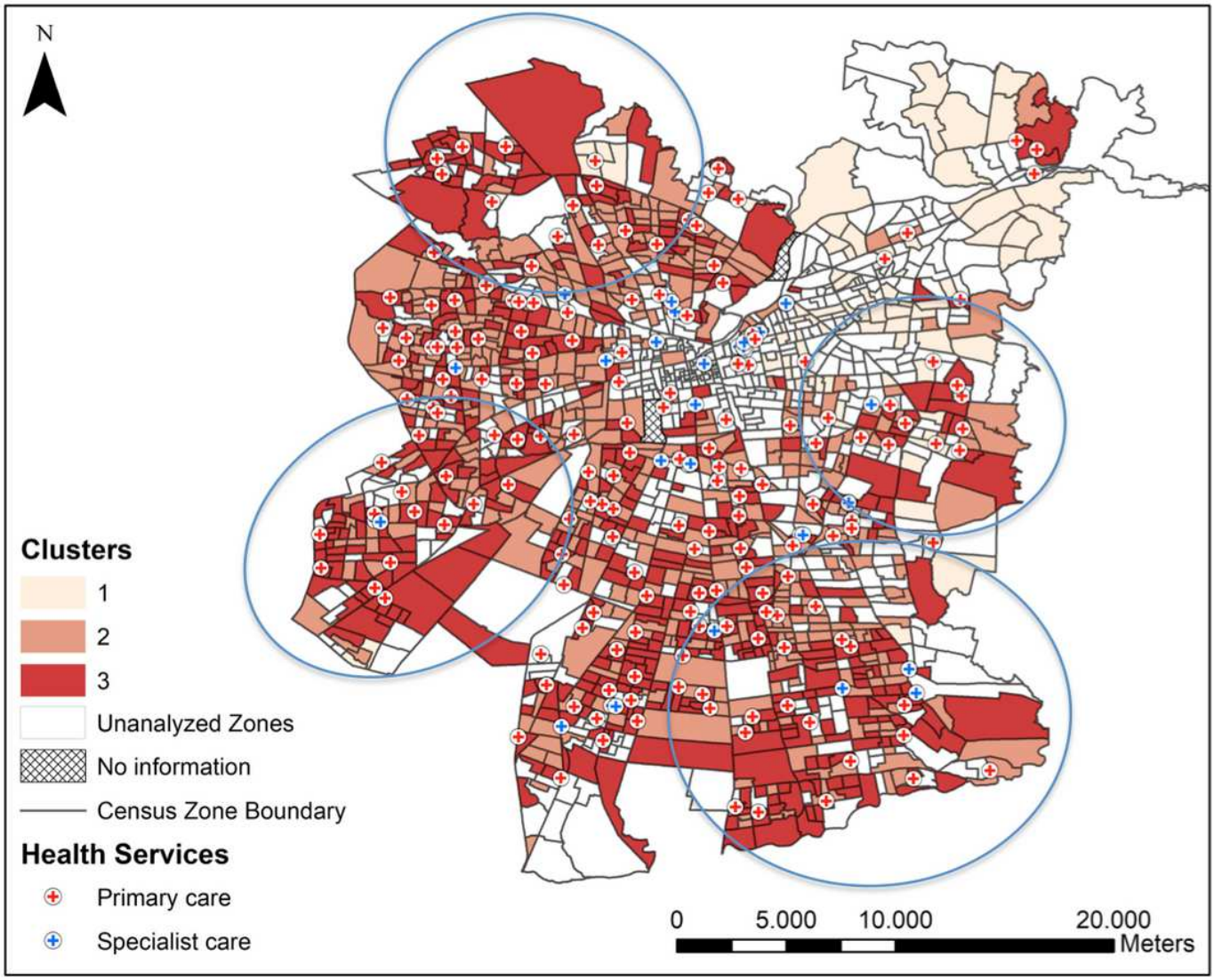

Figure 5

Geographical representation of the SOM results.