Using a GIS Tool to Map the Spatial Distribution of Population for 2010 in the State of Mexico, Mexico

Roberto Franco-Plata¹, L. Ricardo Manzano-Solís³, Miguel A. Gómez-Albores², José I. Juan-Pérez⁴, Noel B. Pineda-Jaimes⁴, Araceli Martínez-Carrillo¹
¹Faculty of Geography, Autonomous University of the State of Mexico, Toluca, Mexico
²Inter-American Center of Water Resources and Mexico Idriisi Resource Center, Autonomous University of the State of Mexico, Toluca, Mexico
Email: {roberto_franco_plata, qfbmiguel Ga}@yahoo.com.mx, {jijp1958, amtzca}@hotmail.com, {noelpj2000, luisrms}@gmail.com

Received September 28, 2011; revised November 3, 2011; accepted December 6, 2011

ABSTRACT

The present document describes the experience of using a Geography Information System (GIS) tool to map the spatial distribution of the population of the State of Mexico, Mexico, for 2010. Prior to the development of the experience, different projects, tools and products, which offered cartographic representations of the population, were presented, selecting one that was developed to function in Idriisi Taiga GIS. Likewise, the GIS tool and the process followed to apply it to a case study in the State of Mexico are described. The result is a raster image that shows the distribution of the population on the analyzed state, offering the possibility to identify the zones where the population concentrates and disperses and to produce an opinion from a geographic point of view on what the origin of these distribution patterns is. Among the conclusions from the analysis of results, distinguishable are the applications that the obtained image can have; for instance, the population can be estimated by basins or risk zones by only adding the values of the cells inside these zones. It is also concluded that the degree of detail expressed by the spatial distribution image of the population will depend on several factors, among them the scale and resolution of the input information and the resolution fixed for the output information.

Keywords: Spatial Distribution of the Population; Idriisi Taiga; GIS Tool; State of Mexico

1. Introduction

Learning the spatial distribution of the population inside the limits of a territory is a fundamental for strategic planning the management and use of land and related resources. For instance, the mere fact that the limits of a basin do not fully coincide with administrative divisions underscores the need to resort to cartographic solutions to try to resolve such a situation when it is fundamental, in processes of integrated water resources management, to know in the most accurate possible manner how people are spatially distributed in a given basin: identify those zones in the basin where population concentrates and disperses.

The traditional methods of cartography allow learning the distribution of the population giving a spatial context to the census data at the level of localities, as a data point, or by means of their aggregation at the level of municipalities, states or the country, in the case of polygonal entities which represent areal administrative limits.

There are cases in which cartography allows representing the spatial distribution of the population inside polygonal entities in the shape of isolines, which are made from interpolating data points; nonetheless, even these do not express the behavior of the distribution between one isoline and another.

Nowadays with the integrated use of geo-technologies such as geographic information systems (GIS), there are cartographic representations that show the spatial distribution of the population of a territory on raster images ([1] National Center for Geographic Information and Analysis [NCGIA], 2011; [2] Oak Ridge National Laboratory [ORNL], 2011; [3] Socioeconomic Data and Applications Center [SEDAC], 2011; [4] United Nations Environment Programme [UNEP], 2011), where certain methodologies are applied to disaggregate the population of a territory at the level of cells, which as a set express the sum of the population in that territory and their distribution inside it. Despite that the methodologies to generate these images are available to reproduce the experience in other territories ([2] ORNL, 2011; [3] SEDAC, 2011; [4] UNEP, 2011) either by spatial level and the amount of required input data, because of the diversity of the em-
ployed computing programs implied or the very complexity of the calculation methods, said methodologies have not been replicated, they have not even been included as a tool in a GIS.

The combined use of the available tools in current GIS, based on theoretical-methodological fundamentals that explain and allow inferring the spatial distribution of the population inside a territory, make it possible to generate cartographic representations in continuous images, this is to say, matrix formats that show said information in discreet form, i.e., at the level of cells. One can go further however, automating the process to generate this representation to infer the spatial population distribution.

The main objective of this document is to show a tool that works in Idrisi Taiga GIS; said tool generated continuous images to map the spatial population distribution from the official information available in different offices of the federal government in Mexico; however, it is considered that this kind of information may be available in other nations, for it is fundamental for decision making related to territory management. Said tool was implemented in Idrisi Taiga on the basis of theoretical-methodological support related to the occurrence of the geographic population distribution, in views of working from input data relatively easy to obtain, which could be adapted to the needs of scale and spatial resolution of the user and which besides allowed working with an approach of temporal analysis, this is to say, making the user able to generate images of the evolution of the population distribution from input data related to different times. The intention of disseminating this tool through a case study in the State of Mexico is to obtain feedback of its functioning and application in their different case studies, from which the experiences of its application in other countries can be assessed and, hence its functionality validated. The tool is available free of charge in the internet website of México Idrisi Resource Center (CRI-Mexico) coordinated by the Autonomous University of the State of Mexico (http://idrisi.uaemex.mx).

2. Availability of Cartographic Data or Tools to Map the Population Distribution

Not few are the sites where information on the spatial population distribution can be obtained; they work with different methodologies of development and at different spatial levels and temporal scales. Below we describe those that offer information in raster format.

LandScan project by [2] ONRL (2011) employs in a combined manner GIS with remote perception to generate raster images of population distribution with an approximate resolution of 1 km; said images are available for downloading at the internet site of ONRL, they include information at global, regional and country level, for the years 2007-2010. For the case of education or research institutions, these images are free. Although [2] ONRL (2011) exposes the methodology to develop the images of population distribution, the user has to implement it on their own if they want to generate this type of images for data with specific scales and temporalities; among the data necessary to develop them we find high resolution remote perception images, which due to their cost are not always easily affordable, mainly in higher education institutions of developing countries.

Global Resource Information Database-Sioux Falls is the reference that [4] UNEP (2011) offers for downloading free database and maps with the spatial distribution of the population at global and regional level, with temporality between 1960 and 2000 and resolution between 16 km and 117 km. In general terms, the utilized method is based upon the model called population potential ([5] Nelson, 2004), and to implement it roads and urban centers were used to calculate an accessibility index, which was interpolated to be adjusted by databases of water bodies, protected areas and altitude. The final map comes from weighing the initial index in relation to the interpolated and adjusted image considering administrative limits to represent the data in an aggregated manner for this sort of management area. Even if the described model is simpler and requires few input data, so far it is not implemented in any GIS to perform the estimation of the population distribution in an automated manner.

In the National Center for Geographic Information and Analysis The Global Demography Project ([1] NCGIA, 2011) was developed, which makes available through the internet raster images of the spatial population distribution with 5 minute resolution, generated with Tobler’s pycnophylactic interpolation on administrative units. This project has changed from Gridded Population of the World 1 (GPWv1) to GPWv3, available through Socio-economic data and Applications Center ([3] SEDAC, 2011), providing information on the population distribution between 1990 and 2000, with images ranging from 2.5 minutes of arc to a degree. As in the previous cases, not a single GIS tool was created or used to produce the raster images of spatial population distribution, but different tools were employed in combination both inside and outside GIS.

As it can be seen in the analyzed background, even if the offered images have good quality, the main limitation of working with the images generated by some of these projects are the temporal domain of it, for they are restricted to 1960-2010, so if it is necessary to work with information out of this range one has to generate the images. Moreover, it is important to underscore that in none of the cases a tool was created to be used in a GIS by means of which generate the cartographic representations of the Spatial population distribution in an automated way and, instead the work was undertaken by combining
the use of different tools inside and outside GIS; because of this, any user interested in learning the current distribution of the population in a given territory would have to gather the input data, acquiring the computing programs utilized and reproducing the applied methodologies, which may become a heavy investment on time to become familiar with all these elements.

In views of trying to solve the aforementioned situation, [6] Manzano and Franco (2009) and [7] Manzano et al. (2010) implemented an application in Idrisi Taiga GIS, which generates this sort of raster images for the time, spatial coverage and resolution desired by the user. By and large, the tool identifies the tendency of the population to concentrate or disperse from the interpolation the population of the localities in the territory, to later identify the variables that influence said distribution pattern through multiple linear regression which becomes a model employed to disaggregate the population of the territory in cells of the resulting raster image.

3. Materials and Methods

The GIS tool to map the population distribution that was used in this study to infer the spatial distribution of the population in the State of Mexico was developed following the so called classical life cycle for the development of information systems ([6] Manzano and Franco, 2009; [7] Manzano et al., 2010), whose effective application to GIS has been approved under different approaches by several studies and authors ([8] Environmental Systems Research Institute [ESRI], 2007; [9] Franco, 2008; [10] Manzano, 2007; [6] Manzano and Franco, 2009; [7] Manzano et al., 2010; [11] Quentin et al., 2007; [12] Tomlinson, 2003), following the global methodological schema presented below:

1) Analysis of requirements: the problem to be solved with the development of SIG tool was identified and limited; the scope and general context of the development context were defined.

2) Conceptual design: the theoretical-methodological foundation for the design of the tool was carried out; likewise, the main variables that intervene in the spatial population distribution and the way they are related for the occurrence of the phenomenon were both identified.

3) Logic design: the logical sequence to follow in order to solve the problem in a GIS environment was documented, basically supported on cartographic modeling; we identified the main functionalities of the spatial analysis of GIS to be used. In this stage we defined the input and output data that solve the necessity stated at the beginning of the process.

4) Implementation: development and implantation of the tool in a platform compatible with GIS, from which the automation of the processes evaluated in the previous stage is accomplished.

5) Testing: running tests and evaluating the results to ensure that the new GIS tool fulfills in an optimal manner the desired purpose and that its functioning is the adequate.

The theoretical framework that sustains the functioning of the SIG tool ([7] Manzano et al., 2010; [6] Manzano and Franco, 2009) has identified factors and limitations as the main variables involved in the spatial distribution of the population; being the former when they influence in a relative manner in the distribution process, while the latter intervene when there are absolute barriers for human settlements. Among the factors one finds temperature, elevation, slope as well as distance to roads, urban zones and water bodies; whereas in relation to limitations, the tool considers water bodies, roads and restricted zones, such as protected natural areas. The factors are evaluated by means of a multiple linear regression process to know their specific weight on the spatial population distribution.

The authors of the GIS tool ([7] Manzano et al., 2010) made a bibliographic research from the biogeographic perspective with the intention of establish and model the main variables about the way the distribution of population occurs in a territory. They found that technologic developments have been fundamental for occupation of zones distinct from traditional zones; however there are limitations conditioning important concentrations of population. These authors identified a big number of variables which influence the distribution of population, so they used the selection and generalization of variables to reduce the complexity of the model, but maintaining the essence of the dynamic modeled: the distribution of the population.

In the case of physical environment factors, [7] Manzano et al., (2010) found that people look for settlement in zones with intermediate temperature (neither so hot nor so cold zones), with mid to few values of elevation and on flat terrain, falling their presence in territories with high values of elevation and irregular surface; in the case of factors from socioeconomic sector, they found that to smaller distance to communication network, urban zones and water bodies, the population trends to concentration, in the other hand, as the distance increments from these elements, the dispersion of population too. Additional to the previous idea, limitations are translate as the impossibility of settlements on water bodies in the physical context and the same situation for communication network and restricted areas (for instance, natural protected areas) as socioeconomic variables.

Taking account the situation described above, the GIS tool infers a pattern of the distribution of the population from the population by localities; next this pattern is used to correlate it with all considered variables and recreate that pattern using these variables. To the end of the pro-
cess, the GIS tool allows both, to identify the degree of relationship of each variable in the distribution of population process, and the variables with the max and min degree of relationship.

In Table 1 the input data used to work with the analyzed tool for the case of the State of Mexico are presented, while the procedure to standardize said data is presented in Figure 1. Here, it is worth underscoring that the scale of the work was 1:1000000 for vector data and resolution of 90 m for raster data; this is mainly because of two issues: 1) as they were the dimensions available from the information suppliers; and 2) as it is a study zone that covers an entire state of the Mexican Republic, with an approximate surface of 22,488 km², so the degree of detail provided by the input data was considered sufficient to generate the desired results.

In the procedure presented in Figure 1, distinguishable is the fact that originally DEM has certain “holes” from lack of information (“No data” cells), which must be corrected before using the tool to produce the best results; likewise, the temperature offered by WorldClim comes disaggregated in monthly raster images, so it is necessary to calculate the mean annual value from them. As for surface runoffs and water bodies and also roads and railroads, in both cases it was necessary to generate a single layer, being one for hydrography and other for transport network, respectively. The entry file of protected natural areas provides different categories for this sort of areas, so we only selected those which had a total restriction for human settlements. In the case of supplying a spatial framework for the total population by municipalities and localities, in the first case the population by municipalities was linked to the layer of the municipal geo-statistical framework. i.e., that which contains the polygons of the municipalities that integrate the State of Mexico; whilst with the pair of geographic coordinates of the localization of the settlements was sufficient to spatialize said localities using these coordinates and thus represent the population as a set of points in vector data.

The phase of data standardization consisted in converting the different formats in which the information was obtained into the Idrisi Taiga raster format, and also producing the same cartographic projection, coordinate system, spatial coverage, conversion and resolution as the different input layers.

Once all the necessary input data were ready, the next step was to use the GIS Idrisi Taiga tool in order to infer the spatial distribution of the population. In Figure 2 the main window of the application is shown, which one can access via a menu included in the main menu bar of the GIS. Likewise, in Figure 3 the necessary elements for the correct functioning of the tool are indicated.

### 4. Study Zone

The study zone selected to generate continuous surface images of the population spatial distribution by means of Idrisi GIS tool implemented by [7] Manzano et al. (2010) was the State of Mexico, which is located in the central zone of Mexico (Figure 4). In the figure below one notices the municipal political division of the state, which consists of 125 municipalities. Out of the municipalities, four were selected (those profiled in bold in Figure 4) in views of showing in greater detail the results, being these the municipalities of San Felipe del Progreso, Toluca, Nezahualcóyotl and Chalco.

<table>
<thead>
<tr>
<th>Table 1. Data collected to generate the raster image of the population distribution with the GIS tool.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
</tr>
<tr>
<td>Mean annual temperature</td>
</tr>
<tr>
<td>Hydrography</td>
</tr>
<tr>
<td>Roads</td>
</tr>
<tr>
<td>Railways</td>
</tr>
<tr>
<td>Urban zones</td>
</tr>
<tr>
<td>Protected Natural Areas</td>
</tr>
<tr>
<td>Municipal political division</td>
</tr>
<tr>
<td>Total population per localities</td>
</tr>
<tr>
<td>Total population per municipalities</td>
</tr>
</tbody>
</table>

NB: Annex 1 includes the links to access the data in the table.
Figure 1. General procedure followed to prepare the input data for the GIS tool to generate a raster image of the population distribution.
Figure 2. The tool implemented in Idrisi Taiga GIS to map the spatial population distribution.

Figure 3. Parameters needed to perform the mapping of the population distribution.
Figure 5 shows the distribution of the population in the state; it presents high concentration (higher population density per cell) in the municipalities bordering Mexico City (capital of the country), as in the case of the city of Toluca, head of the municipality of the same name and capital of the state. Likewise, as distances from these zones increase, the population tends to disperse (lower population density per cell), expressing a minimal in the southern and southwestern part of the state. From the viewpoint of the population distribution, it is relevant to see that in Figure 5 there is not a corridor with similar values of population density per cell that connects both zones, that of the conurbations and that of the Capital of the State.

It is very important to distinguish that if an extraction of the total population inside one of the municipalities is performed, this is to say, to add all the cells with population information in their administrative limit, the result is the total population of the municipality registered in 2010 INEGI population census. This is valid for any of the 125 municipalities comprised in the state of Mexico, since the population that was disaggregated at the level of cells was the total in each of them from the aforementioned census.

Making a detailed analysis in each of the municipalities, it is possible to identify the tendencies to concentrate and disperse inside them; to exemplify this, below we expose the cases of the municipalities of San Felipe del Progreso, Toluca, Nezahualcoyotl and Chalco.

The selection of the State of Mexico for the application of the GIS tool to be analyzed is fundamentally due to the fact that from a geographic viewpoint it presents several interesting conditions that have configured the current distribution pattern of the population. Among these conditions distinguishable is the bordering of some municipalities with Mexico City, capital city of the country, so they have turned into conurbations of this city; separately, the state possesses a wide variety of reliefs and well-differented land uses, for instance agricultural, industrial and urban uses, which have become zones of high population density and also sparsely populated areas. Moreover, the State of Mexico is second to Federal District (Mexico City) as for the largest number of inhabitants.

5. Results

The tool in GIS to infer the spatial distribution of the population in 2010 in the State of Mexico generated a raster image with a cell resolution of approximately 90m by side (Figure 5); it is worth mentioning that the output resolution complies with the input DEM dimensions, and the rest of information was adapted to this dimension.
San Felipe del Progreso is far from the capital of the state and even further from the capital of the country, so it is not influenced by the behavior of population distribution in any of these capitals (high population density); in fact, Figure 6 shows that the highest value of population per cell is barely above 4 inhabitants, whilst the total surface is approximately 369 km². The fact that center of the municipality has the lowest density population values is explained because this zone, being a flatland and near the lake, is used for cultivations and livestock activities, thus population density grows as the distance from the lake increases, where the relief is rather irregular, however still optimal for human settlements.

The following municipality, Toluca (Figure 7), comprises the capital of the State of Mexico, a relatively small city compared with others in the country, such as Monterrey and Guadalajara, not to mention Mexico City; it is mostly devoted to industrial and commercial activities in the center of the municipality, and to agricultural and forest activities in the rest of the territory. The total population of the municipality is 819,561 inhabitants and its surface is 430 km², where one finds broad housing developments, close to downtown, principally along the motorway that connects Toluca and Mexico City, where the industrial zone is found. The described zones are the ones in which we find values above 20 inhabitants per cell; at the other end, places with low population values per cell are found, particularly engaged in agricultural activities, taking advantage of the lacustrine grassland in which the municipality seats. The minimal values appear toward the southwestern end of the municipality, where the Xinantecatl volcano (Nevado de Toluca) is located, and whose elevation that reaches 4600 m.a.s.l. is the main explanation for the pattern of population dispersion expressed by Figures 7-8, as it is appreciated the population values per cell decreases following the volcano elevation pattern (analyzed by DEM of the study zone).

The results for the most densely populated municipality of the State of Mexico (1,110,565 inhabitants) are now described; the municipality of Nezahualcoyotl (Figure 9): located in the eastern part of the state, which borders with Mexico City in its western part. The high population concentration is mainly due to its nearness with Mexico City. With approximately 63 km² of surface, the municipality is a flatland, and more than 80% of the surface has housing use, while the remainder towards the east of the municipality is destined to commercial use or green areas (Figure 10). This situation is reflected in Figure 9, where a rather uniform population density per cell is seen, concentrating toward the west of the municipality at the border with Mexico City and dispersing towards the east.

Finally, we present the result of population distribution for the municipality of Chalco (Figure 11), which expresses two of the situations presented for the previous municipalities, as it limits with Mexico City in its western portion, this zone is the one with the highest concentration of population per cell. Distinguishable is that this density per cell is more similar to the values of concentration in the municipality of Toluca than to those of Nezahualcoyotl, possibly due to a larger surface of this last municipality, with approximately 226 km² combined with a smaller population, 310,130 inhabitants; separately, the population tends to disperse toward the east of the municipality, presenting the lowest values in this section of the territory, similar to those of San Felipe del Progreso.
Figure 8. Visualization example (3D) of the results for the municipality of Toluca on Google Earth (2011); with Xinanteclatl volcano (Nevado de Toluca) in the background.

Figure 9. Image of population distribution in the municipality of Nezahualcoyotl, State of Mexico.

Figure 10. Google Earth image (2011) showing the high degree of urbanization of the municipality of Nezahualcoyotl, State of Mexico.

Figure 11. Image of the population distribution in the municipality of Chalco, State of Mexico.

This situation reflects two well-differenced situations; in the first case a noticeable urbanization degree, while in the second there is largely a rural landscape. It is important to underscore that about as from the middle part of the municipality and toward the east one finds the slopes of the Iztaocuhiatl Volcano, a situation that nowadays influences the dispersion of population, mainly because of the whereabouts of a Protected Area.

6. Conclusions

The raster image with the spatial distribution of the population in 2010 for the State of Mexico which was the product from using the tool developed by [7] Manzano et al. (2010) for Idrisi Taiga GIS shows a continuous image of the state territory, enabling the identification of patterns for the behavior of population distribution in cells of ~90 meters by side, this is to say, with this tool it is possible to identify spatial tendencies of population concentration or dispersion. By adding the value of the cells that are inside a municipality, we obtained its total population, a data value that coincides with the official result of the population census by INEGI.

Diverse are the applications that the obtained image may have; for instance, the population can be estimated by basins or risk zones by only adding the cells inside one of these zones.

It is important to clarify that the degree of detail expressed by the image of the spatial population distribution will depend on several factors, among them the scale and resolution of input data and the resolution fixed for the output information.

It is clear that the only exact method to identify the precise position of the inhabitants of a locality is directly counting the inhabitants; however, when the necessary infrastructure is not available, one has to resort to alternatives such as the ones described in this document, of which the one by Manzano et al. (2010) was used for the State of Mexico.

As for the functioning of the utilized GIS tool, this presented a good performance, however it was evident that the time employed to process the information will
depend on the number of localities to be interpolated, lasting from seconds for few localities (below 100) to hours for a large set of them (10000 or more).

The aspect that lasted the longest was the obtaining and preparation of the input data that requires the utilized GIS tool. Unfortunately, in Mexico the different institutions that generate digital cartographic information use different criteria of spatial reference, some of the most distinguishable are scale, cartographic projection and coordinate system, more so in respect to international institutions. Nevertheless, Idrisi Taiga has a set of modules that facilitated the standardization of data. A situation to be distinguished is the fact that all the input data necessary for the GIS tool were obtained from the internet, an aspect that substantially reduced the generation of results here presented, facilitating the use of the module for other study zones.

To sum up, the reader is invited to learn and use the GIS tool here presented, which can be downloaded from the internet site of the Mexico Idrisi Resources Center (CRI-México) of the Autonomous University of the State of Mexico, at http://idrisi.uaemex.mx in views of being explored and improved. This is another proposal that adds to the other analyzed in this text, however unlike them, it allows the GIS user to generate their own information by means of a tool that automates the process of creation for the characteristics of input data the tool is given, as well as the spatial coverage, temporality and the desired output spatial resolution, without depending on other actors.

REFERENCES


Annex 1. Source of Data

SRTM Digital elevation: USGS
Layer: Digital elevation model of the study zone.

National geo-statistical framework: INEGI
Layers: Municipal Geo-statistical Areas, polygons of geo-statistical urban localities.

2010 Census of Population and Housing: INEGI
Databases: Basic tables (total population, median age, men women ratio and aging index per municipality by sex. Municipal level State of Mexico), main results per locality (ITER) of the State of Mexico and neighboring states.

Set of vectorial data of the topographic series and natural resources, scale 1:1,000,000: INEGI
Layers: Streams, water bodies, roads, railways, protected areas.

Mean world temperature: WorldClim
Layer: Mean monthly temperature of the study zone.

Remote perception images: Google
Remote perception image display: Google Earth.