DOI: http://dx.doi.org/10.22185/24487147.2025.120.20

Prediction model of urban growth in the state of Nuevo León, México

Modelo de predicción del crecimiento urbano en el estado de Nuevo León, México

Refugio Chávez-Hernández*
José Luis Martínez-Flores**
Claudia Malcón-Cervera**
Patricia Cano-Olivos**

**Universidad de Monterrey, Nuevo León, México*
***Universidad Popular Autónoma del Estado de Puebla, México*

Abstract

This study establishes a reference framework to identify a predictive model of urban growth and its orientation for the projection of land reserves for a real estate organization in the state of Nuevo León, Mexico. The study is focused on a linear and polynomial regression modeling algorithm using data from the National Institute of Statistics and Geography (INEGI). The input information is historical data from the years 1970 to 2010, which was used to make projections for the year 2020, the results are compared with the 2020 report presented by INEGI to identify the accuracy of the model. The methodology includes three stages: data preparation, data modeling in Python, and interpretation of results. A coincidence relationship of 90.6 per cent was found, the final modeling was completed until the year 2040, and the results show a distribution of growth with a high percentage of increase in the metropolitan area; 25.45 per cent in 2030, and 75.72 per cent in 2040. Finally, the study recommends geographic orientation in the search for reserves for future planning of works.

Keywords: Urban growth; regression model; prediction; land use.

Resumen

Este estudio establece un marco de referencia para identificar un modelo predictivo del crecimiento urbano y su orientación para la proyección de reservas de territorio para una organización inmobiliaria en el estado de Nuevo León, México. El estudio está enfocado en un algoritmo de modelación de regresión lineal y polinómica utilizando los datos del Instituto Nacional de Estadística y Geografía (INEGI). La información de entrada son datos históricos de los años 1970 al 2010, los cuales se utilizaron para realizar proyecciones del año 2020, los resultados se comparan con el reporte del 2020 presentados por INEGI para identificar la precisión del modelo. La metodología incluye tres etapas: preparación de datos, modelación de datos en Python, y la interpretación de resultados. Se encontró una relación de coincidencia de 90.6 per cent, la modelación final se completó hasta el año 2040, los resultados muestran una distribución de crecimiento con alto porcentaje de incremento en el área metropolitana; de 25.45 per cent en 2030, y 75.72 per cent en 2040. Por último, el estudio recomienda la orientación geográfica en la búsqueda de reservas para la planeación futura de obras.

Palabras clave: Crecimiento urbano; modelo de regresión; predicción; uso de suelo.

Introduction

The accelerated growth of urban centers is a phenomenon characterized by the concentration of most of the expanding transformation industry, since employers and businesses have moved out of the center, favoring the increase of population in the surrounding cities, resulting in a greater expansion of the peripheral area (Aguilar, 1999). It is crucial for both current and future societies that urban growth occurs optimally, maximizing benefits for the general population while minimizing economic and environmental costs (Mandelas, 2007).

Studies on urban growth have garnered considerable attention over the past two decades (Wahyudi, 2013), particularly because urban areas are consistently and rapidly expanding worldwide. The task of managing urban growth has increased both in scope and complexity, becoming one of the most significant challenges of the 21st century (Thapa, 2012), especially within the context of global environmental change. In this case, the objective is to identify the area's most likely to become populated to acquire them as land reserves for housing construction purposes.

Nuevo Leon ranks second in the state competitiveness index in Mexico, and is therefore one of the most important for the country, with an economically active population of 2.54 million people, is the third largest economy in the country, contributing 7.6 per cent of the national GDP, leads the country's secondary activities, is a leader in the construction sector, is the third entity with the most jobs generated and with the greatest attraction of foreign investment (Barcenas, 2020). During the last three decades, the state capital Monterrey and its metropolitan area have experienced rapid and uncontrolled urbanization. The growth of urban sprawl consumes natural resources such as agricultural land, natural landscapes, hills and mountain sides. Because of the importance of the state and the relevance of its industrial environment, it has been chosen in this study to predict its future urban growth.

The population increase in the last decades indicates that the state of Nuevo León became much more populated in the metropolitan area and this should be the main driver of urbanization and land use change in the state (Seto, 2002). Much research has focused on changes in urban land use (Xiao, 2006) since urban ecosystems are strongly affected by human activities and have a close relationship with the life of the population (Stow, 2002).

Forecasting the magnitude and location of urban expansion are key concepts for city planning, as they allow quantifying future needs and the costs of various public services, such as health, education and housing, as well as environmental risks. In the same way, scenarios for job accessibility, changes in demand, transportation needs and traffic flows are enriched (Suárez, 2007). Detailed analysis and monitoring of changes in land use expansion or growth require a substantial amount of data. The use of technological tools such as remote sensing or spatial positioning techniques such as georeferencing allow, in a broader way, to obtain data (Araya, 2010).

In Mexico, urbanization has become a global trend that affects the majority of citizens, not only in the country but worldwide. Currently, 55 per cent of the global population lives in cities, and statistics predict that by 2050, urban centers will need to accommodate approximately 70 per cent of the world's population. In Mexico, 78 per cent of the population lives in cities, the urban population has doubled in the past 30 years, and since then, urban areas have expanded to seven times their original size (Rojas, 2020).

The exponential growth in the population of Nuevo León has led to a significant increase in housing demand. Between 2015 and 2020, the state's population steadily increased, rising by 13 per cent in just five years. By 2023, Nuevo León's population reached 5,784,442 inhabitants, with an estimated annual growth of approximately 133,000 people in recent years, according to the 2020 census conducted by the National Population Council (CONAPO).

Having an opportunity for economic and demographic growth, as seen in the state of Nuevo León, drives companies in the housing construction sector to undertake long-term investment projects focused on acquiring lots for residential use. These companies face critical decisions regarding selecting suitable land and minimizing the risk associated with millions of pesos invested in urbanization and infrastructure expenses, especially if the market does not respond as expected to housing sales. Even more concerning is the possibility that the developed area may lack interest and become obsolete or abandoned. This is an ongoing concern for developers, and for many, it has become a reality. This situation often results from a lack of alignment with the urban development plan by the government, which exacerbates the consequences of poor site selection. The risks are further heightened when decisions are made empirically or based on unfounded perceptions.

Researchers in the construction industry state that there are currently few developers in the country with clearly defined expansion plans for their projects. In most cases, developments are built on land inherited from previous generations. This poses a significant limitation since, in general, the land is not suitable, and developers must make the best of the location they have. This is just one example of why developers need to start planning long-term strategies for growth and seeking land with better development potential, taking into account industrial expansion trends and consumer needs. Otherwise, they face a high risk of potential failure. The lack of long-term strategic planning among many developers is alarming, especially given the high costs associated with mistakes in choosing locations for new projects.

This work considers a model for estimating housing growth in the state of Nuevo León, aimed at predicting the number of housing units in each municipality over different time periods. The model leverages trend analysis techniques to provide insights and visualize the direction of urban expansion. This information is intended to support decision-makers in a construction company in selecting residential lots for the development of new housing projects.

Methodology

The methodology of this study is presented in three stages. The first stage involves the preparation and processing of data extracted from the 2020 Population and Housing Census available on the INEGI platform, specifically from the report Sociodemographic Overview of Nuevo León, focusing on factors reflecting urban growth.

In the second stage, data modeling is conducted, along with the estimation of the accuracy of the results, based on the structured information from the first stage. This is achieved using linear and polynomial regression models, implemented in the Python programming environment, with the aim of predicting future housing growth in each municipality of Nuevo León.

The third stage presents the visualization of results for each of the time periods estimated by the regression model, both for each municipality and for the state as a whole.

This section briefly introduces the statistical technique used to analyze housing growth in the municipalities of Nuevo León. Essentially, the developed models can be divided into two types: process-based models and data-based models. Among the process-based models are Cellular Automata

(CA); since the 1980s, numerous models have been developed using this method to simulate urban growth (Sante, 2010). CA represents a potential approach for modeling urban growth by simulating spatial processes as a discrete and dynamic system in both space and time (Alkheder, 2008).

Regression, on the other hand, is data-based and is a common method for predicting and determining relationships between urban land uses and independent variables (Fang, 2004). Linear, polynomial, and logistic regression models have been widely used in urban growth modeling, incorporating independent socioeconomic and environmental variables (Poelmans, 2009).

Linear regression models are frequently used across various disciplines, including medical sciences, engineering, stock market forecasting, and the energy sector, among others. This study takes into account certain concerns regarding the use of the proposed tool to facilitate result analysis by personnel who may not have a strong grasp of the statistical foundations of these models.

In reviewing the literature, some authors have utilized and proposed a wide range of programs to perform linear regression analysis. These tools include mathematical functions that must be written and executed to carry out regression analysis. Moreover, understanding the results generated by these tools is not always feasible for non-expert users. Researchers have also applied standard statistical techniques to manually derive the best-fitting linear regression model, a process that is time-consuming and often complex (Roy, 2022).

This study proposes a model developed in the Python programming language. The proposed semi-automated tool for linear regression analysis will be highly beneficial in scenarios where time and effort can be saved in estimating the number of housing units for each municipality in the state. Another advantage is the ease of interpretation of the analysis results for non-expert users.

Linear and polynomial regression were examined based on the availability of information in the INEGI database to identify the predictive model for housing growth in each of the 51 municipalities of Nuevo León, Mexico, with particular emphasis on the results presented in the metropolitan area.

Literature review

Linear and polynomial regression

Regression is a statistical technique that attempts to estimate the strength and nature of the relationship between a dependent variable and a series of independent variables. Regression analyses can be linear and nonlinear. A regression is called linear when it is linear in parameters: $y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_1 + \varepsilon$, para $i = 1,2, \dots , n$, where y is the response variable, x denotes the independent variable, β_0 is the intercept and other βs are known as slopes (Sharma, 2020).

In the case of best fitting a curve to the data, an alternative is to fit a polynomial to the data using polynomial regression, in which the relationship between the independent variable and the dependent variables is modeled as a polynomial of order *n*. It fits a nonlinear relationship between the observed variables; but, as a statistical estimation problem, it is linear, because the regression function is linear in the unknown regression parameters. The objective of regression analysis is to model the expected value of a dependent variable in terms of the value of an independent variable (or vector of independent variables) *x*. In general, the expected value of *y* as a polynomial of order *n*, which produces the general polynomial regression model, can be expressed as follows: $y = \beta_0 + \beta_1 x_1 + \beta_2 x^2_1 + ... + \beta_n x^n_1 + \varepsilon$, $\varepsilon \sim (0, \sigma^2)$, para $i = 1,2, \dots, n$ (Montgomery, 2012).

Measurement of model accuracy

It is important to have a measurement of the forecast or projection accuracy obtained from the models. This measurement plays a crucial role as it allows us to observe the forecasting error due to variability and uncertainty. The projection error or accuracy is derived from the difference between the actual value and the predicted value for a specific period.

The mean square error *MSE* is an unbiased estimator of the variance σ^2 of the random error term and is defined in the equation:

$$MSE = \frac{SSE}{df_E} = \frac{\sum_{i=1}^{n}(y_1 - \hat{y}_i)^2}{n - (k+1)}$$

Where y_i are the observed values and $\hat{y}_i$ are the fitted values of the dependent variable *Y* for the i-th case. Since the mean square error is the average of the squared error, where the average is done by dividing by the degrees of freedom, the is a measure of how well the regression fits the data. The square root of *MSE* is an estimator of the standard deviation σ of the random error term. The mean square error $RMSE = \sqrt{MSE}$ is not an unbiased estimator of σ, but it is still a good estimator. *MSE* and *RMSE* are measures of the size of the errors in the regression and give us an indication about the explained component of the regression fit (Aczel, 2006).

The mean absolute percentage error *MAPE* is the most useful measure for comparing forecast accuracy across different items or products, as it measures relative performance (Makridakis, 1997). It is a commonly used measure of accuracy in quantitative forecasting methods. This measure is defined in the equation:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

If the value calculated by *MAPE* is less than 10 per cent, it is interpreted as an excellent and accurate forecast, between 10 and 20 per cent is a good forecast, between 20 and 50 per cent is an acceptable forecast and more than 50 per cent is an inaccurate forecast (Lewis, 1982).

The R-cuadrado R^2 (coefficient of determination) of multiple regression is similar to simple regression where the coefficient of determination R^2 is defined as:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_1 - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_1 - \bar{y}_i)^2}$$

Where *SST* is the total sum of squares and $\bar{y}$ is the arithmetic mean of the variable *Y*. R^2 measures the percentage of variation in the response variable *Y* that is explained by the explanatory variable *X*. Therefore, it is an important measure of how well the regression model fits the data. The value of R^2 is always between zero and one, $0 \leq R^2 \leq 1$. An R^2 value of 0.9 or higher is very good, a value above 0.8 is good, and a value of 0.6 or hi-gher may be satisfactory in some applications, although we must be aware that in such cases prediction errors may occur. When the value of R^2 is 0.5 or lower, the regression explains only 50 per cent or less of the variation in the data; therefore, the prediction may be poor (Ostertagová, 2011).

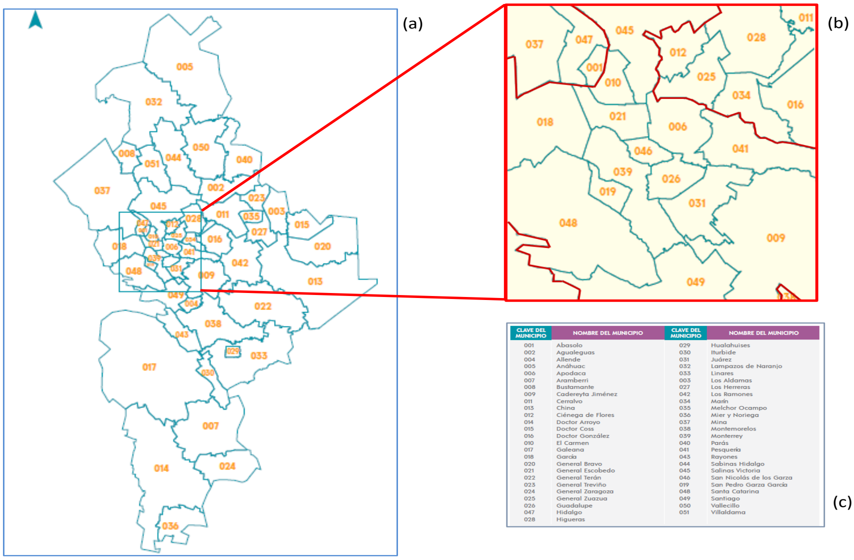
Study area

The study area is the state of Nuevo León, Mexico, located in the northeastern part of the country. It borders the United States to the north, separated by the Rio Bravo; Tamaulipas to the east; San Luis Potosí to the south; and Coahuila and Zacatecas to the west. It is divided into 51 municipalities. The municipalities of Apodaca, Cadereyta Jiménez, García, General Escobedo, Guadalupe, Salinas Victoria, El Carmen, Juárez, San Nicolás de los Garza, San Pedro Garza García, Santa Catarina, Santiago, and, due to its growth, Pesquería, along with Monterrey, form the Metropolitan Area of Monterrey, which plays a key role in the economy of the state and the country. The remaining municipalities are: Abasolo, Agualeguas, Los Aldamas, Allende, Anáhuac, Aramberri, Bustamante, Cerralvo, Ciénega de Flores, China, Doctor Arroyo, Doctor Coss, Doctor González, Galeana, General Bravo, General Terán, General Treviño, General Zaragoza, General Zuazua, Los Herreras, Higueras, Hualahuises, Iturbide, Lampazos de Naranjo, Linares, Marín, Melchor Ocampo, Mier y Noriega, Mina, Montemorelos, Parás, Los Ramones, Rayones, Sabinas Hidalgo, Hidalgo, Vallecillo, and Villaldama (INEGI, 2020).

The territorial distribution of the state covers 64,156.2 km^2, which represents 3.3 per cent of the national territory. The population density is 90.2 inhabitants per km^2. The state has 1,655,256 occupied private dwellings, accounting for 4.7 per cent of the national total. On average, each dwelling has 3.5 occupants, and 99 per cent of the dwellings have access to basic services such as electricity, drainage, sanitation, and piped water (INEGI, 2020).

According to the 2020 Population and Housing Census report by INEGI, in 2020, the state of Nuevo León, Mexico, ranked seventh nationwide in terms of population. Between 2000 and 2020, the population increased by 50.8 per cent, while in the last five years (2015-2020), the growth was 13.1 per cent. The population of Nuevo León in 2020 reached 5,784,442 inhabitants, with an estimated annual growth of 133,000 people. This represents an exponential population growth in the state, which has led to a significant increase in the demand for housing, public services, roads, transportation, and other infrastructure (INEGI, 2020).

Figure 1: Location of the study area: (a) map of the administrative division of the state of Nuevo León, (b) metropolitan area of Monterrey and (c) keys and names of the 51 municipalities in the study area



CLAVE DEL MUNICIPIO	NOMBRE DEL MUNICIPIO	CLAVE DEL MUNICIPIO	NOMBRE DEL MUNICIPIO
001	Abasolo	029	Hualahuises
002	Agualeguas	030	Iturbide
004	Allende	031	Juárez
005	Anáhuac	032	Lampazos de Naranjo
006	Apodaca	033	Linares
007	Aramberri	003	Los Aldamas
008	Bustamante	027	Los Herreras
009	Cadereyta Jiménez	042	Los Ramones
011	Cerralvo	034	Marín
013	China	035	Melchor Ocampo
012	Ciénega de Flores	036	Mier y Noriega
014	Doctor Arroyo	037	Mina
015	Doctor Coss	038	Montemorelos
016	Doctor González	039	Monterrey
010	El Carmen	040	Parás
017	Galeana	041	Pesquería
018	García	043	Rayones
020	General Bravo	044	Sabinas Hidalgo
021	General Escobedo	045	Salinas Victoria
022	General Terán	046	San Nicolás de los Garza
023	General Treviño	019	San Pedro Garza García
024	General Zaragoza	048	Santa Catarina
025	General Zuazua	049	Santiago
026	Guadalupe	050	Vallecillo
047	Hidalgo	051	Villaldama
028	Higueras		

Source: INEGI 2020.

Data

The availability of accurate, reliable, and high-quality data is crucial for the success of any research. In Mexico, INEGI makes a considerable amount of historical and current information publicly available, with sufficient detail to enable society to conduct analyses and studies, particularly for urbanization studies, urban spatial analysis, and future urban planning. The required data collected includes historical data on housing growth trends from 1970 to 2020 in the state. Table 1 breaks down the information for the municipalities that make up the Metropolitan Area of Monterrey.

Since this study focuses on simulating urban expansion based on housing growth, the data is limited to this specific use. The selected and collected variables, which serve as driving forces for housing growth in this work for the development of estimation models using regression analysis, are: total number of residents, total number of dwellings with public services, number of people residing in dwellings, population density, and geographical area. These variables are broken down and structured for each of the 51 municipalities.

Table 1: Population and housing in the Monterrey metropolitan area

Municipality	Longitude	Latitude	Km^2	Density	Total Population	Total Housing	Housing occupancy	Housing basic services
Apodaca	-100.189	25.781	224.00	2,930.64	656,464	212,801	651,904	181,334
Cadereyta Jiménez	-100.002	25.591	1,140.90	107.23	122,337	55,231	120,302	36,672
El Carmen	-100.317	25.854	104.30	1,001.71	104,478	38,780	104,368	29,667
García	-100.430	25.781	1,032.00	384.89	397,205	157,779	397,103	114,108
San Pedro Garza García	-100.402	25.665	70.80	1,866.79	132,169	41,237	131,749	36,261
General Escobedo	-100.323	25.808	149.40	3,220.97	481,213	154,216	480,588	130,256
Guadalupe	-100.260	25.677	118.40	5,431.95	643,143	207,183	642,036	181,618
Juárez	-100.091	25.646	247.30	1,906.68	471,523	181,463	471,395	133,558
Monterrey	-100.311	25.665	324.40	3,523.41	1,142,994	368,796	1,138,867	324,521
Pesquería	-100.054	25.783	322.80	457.32	147,624	57,402	147,587	43,320
Salinas Victoria	-100.243	25.913	1,667.40	52.04	86,766	36,175	86,737	25,126
San Nicolás de los Garza	-100.290	25.755	60.10	6,858.55	412,199	133,725	411,924	121,534
Santa Catarina	-100.462	25.674	915.80	334.49	306,322	90,966	306,133	81,705
Santiago	-100.105	25.480	739.20	63.29	46,784	20,785	46,504	13,872

Source: INEGI 2020.

Data modeling

In this stage, the first step is to import the structured database containing the historical information into the Python program. The next step is to generate graphs for each municipality to visualize the patterns and behavior individually based on the time series of the number of dwellings ("dependent variable"). The graphical result for each municipality is shown in Figure 2.

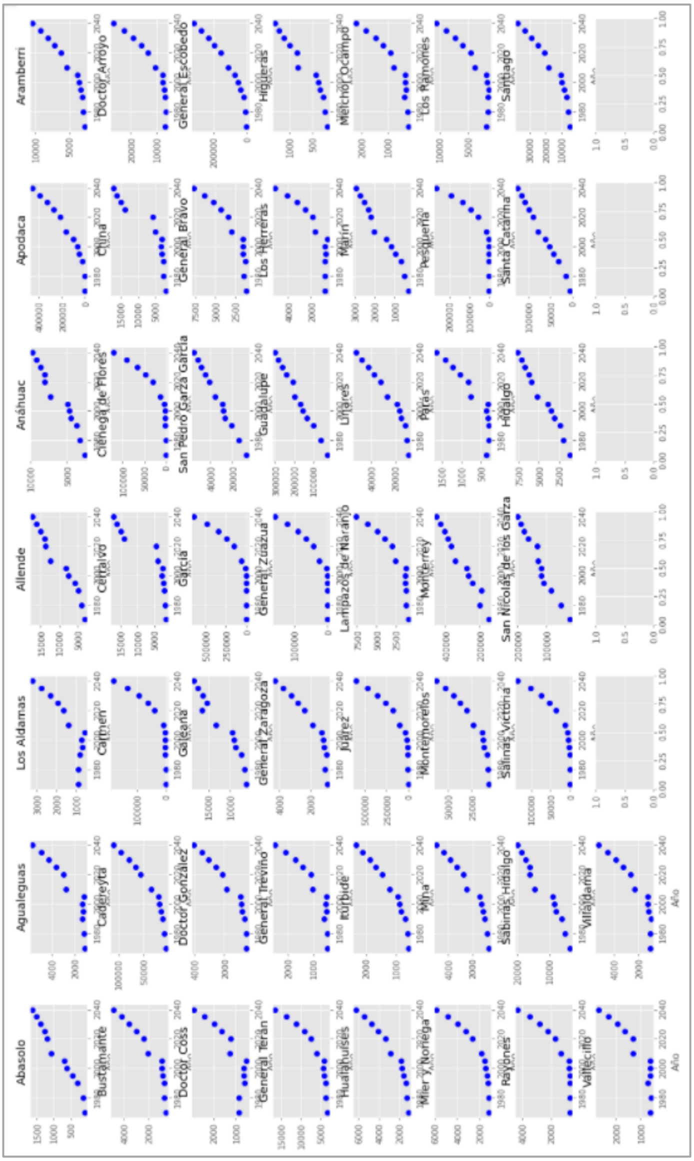
Once the graphs for all municipalities in the state are obtained, the behavior of housing growth for each municipality in Nuevo León is determined. As observed, some graphs show a more linear slope, while others exhibit a more curved projection. This serves as the main basis for determining which method will be used for estimating the growth projection for each municipality. The polynomial regression model is assigned to municipalities with a curved trend line, while the linear regression model is assigned to municipalities where the trend line projects more linearly.

With the regression model configured and tailored for each municipality, its projection is calculated, and the model's accuracy is evaluated by estimating the R^2 value, MAPE, and RMSE for each municipality. Figure 3 shows the results of executing the selected regression model, used here as an example for the municipalities of Abasolo and Apodaca. In the case of Abasolo, the trend line has a more linear projection, so the linear regression model is chosen for estimating future housing growth. The R^2 value is 0.9739, demonstrating that the regression model can confidently explain the relationship between the independent variables and the predicted number of dwellings for Abasolo. The *MAPE* value is 24.5 per cent, and the *RMSE* value is 84.17.

In the case of Apodaca, the graph shows a more curved trend line, so the polynomial regression model is used. This results in an R^2 value of 0.9962. Similar to the case of Abasolo, the R^2 value for Apodaca is considered good, indicating that the independent variables provide enough variation to explain the number of dwellings. The *MAPE* value is 14.11 per cent, and the *RMSE* value is 9,246.44.

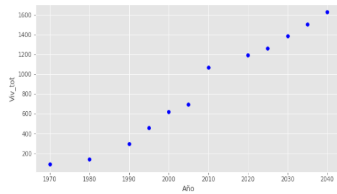
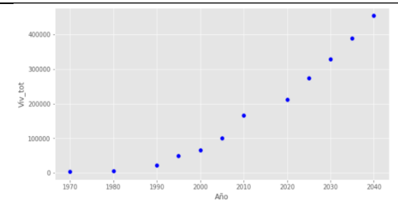
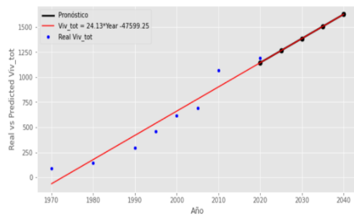
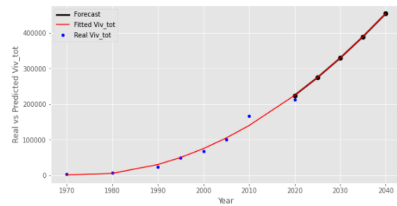
In the case of Apodaca, the error percentage is lower than in Abasolo. However, regarding the *RMSE* error, a much higher value is shown for the number of dwellings. This is because Apodaca is one of the municipalities within the Metropolitan Area of Monterrey, so the number of dwellings is much higher, and its projection is more complex to estimate due to the proportion of its growth over recent years and the availability of land reserves.

Figure 2: Housing growth graphs by municipality in Nuevo León, 1970 to 2020



Source: Own elaboration.

Figure 3: Estimated housing growth in the municipalities of Abasolo and Apodaca for the years 2020, 2025, 2030, 2035 and 2040

	Linear regression	Polynomial regression
Municipality	Abasolo	Apodaca
Behavior graph		
Estimation of the number of dwellings for 5-year periods		
Model accuracy	R2 = 0.9739829908390727 MAPE = 24.50100234190121 RMSE = 84.1743496863782	R2 = 0.9962462795737199 MAPE = 14.11309226558786 RMSE = 9246.449698696257

Source: Own elaboration.

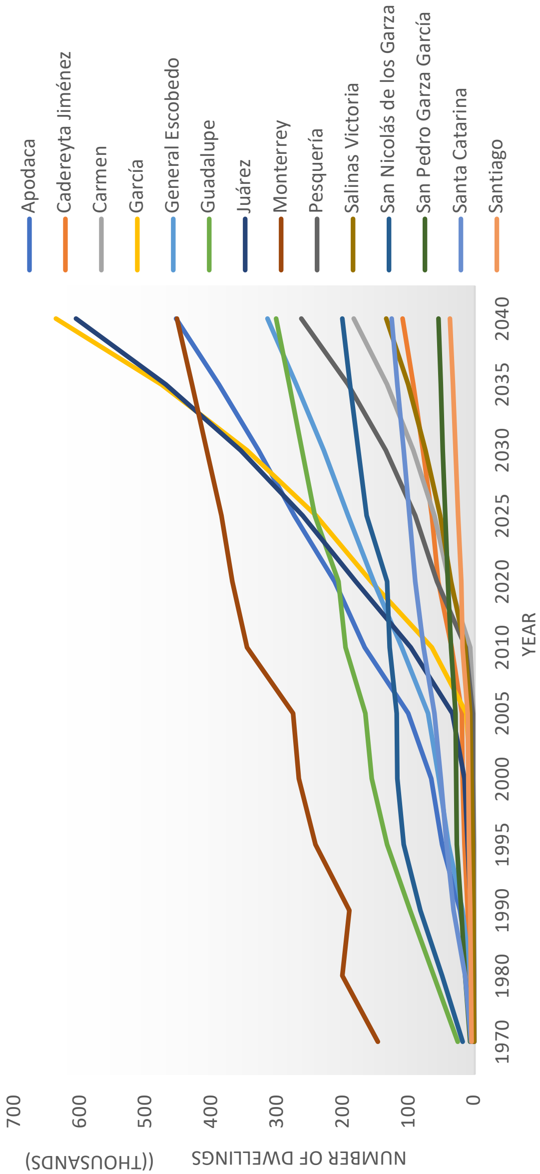
The models also allow us to identify and estimate the direction of urban sprawl in the state of Nuevo León based on the results obtained. On one hand, it is relevant to note that between 85.5 per cent and 86.2 per cent of the housing growth will occur within the Metropolitan Area of Monterrey over the next 20 years. However, municipalities like García and Juárez will concentrate the highest number of dwellings, surpassing municipalities such as Monterrey, Apodaca, Guadalupe, and Escobedo. Figure 4 shows the projection of the number of dwellings for the municipalities in the Metropolitan Area.

Using the forecasting models specified for each of the 51 municipalities based on their regression, a transition period from 1970 to 2010 was established as a reference to estimate housing growth for each municipality and, in general, for the state of Nuevo León by 2020. The results were then compared with the actual data reported by INEGI for the number of dwellings in 2020. This comparison allows us to define the accuracy of the models.

The model is validated with the 2020 data to compare the projected number of dwellings with the actual figures reported in the INEGI database. The model's match proportion is 90.6 per cent for the total number of dwellings in the state and 96.7 per cent for the Metropolitan Area of Monterrey.

Table 2 provides a detailed comparison of the actual values reported by INEGI for 2020 and the estimated results for the number of dwellings generated by the models for each municipality, as well as their accuracy values. It is important to mention that the model has been calibrated for multiple municipalities to improve the precision of the modeling results. For each calibration, both linear and polynomial regression results are evaluated, always considering the same variables or factors present in the INEGI database that impact housing growth. Calibration is straightforward in the Python program: once the specific behavior of each municipality is identified, the required model type is adjusted, the model is re-run, and the coefficient of determination value and error values are carefully reviewed. The model that produces the lowest error and, in some cases, the highest value is always considered.

Figure 4: Housing projection graph in the municipalities of the Monterrey Metropolitan area



Source: Own elaboration.

Table 2: Comparison of the result of housing projections and the actual value reported by INEGI for the year 2020 in the state of Nuevo León

Municipality	Current 2020	Prediction 2020	MAPE	RMSE	R^2
Abasolo	1,194	1,143	4.25%	103.1	0.926
Agualeguas	2,971	3,040	2.34%	328.8	0.787
Los Aldamas	1,623	1,624	0.04%	186.5	0.710
Allende	13,785	12,758	7.45%	1228.9	0.885
Anáhuac	8,119	7,548	7.03%	626.1	0.880
Apodaca	212,801	224,481	5.49%	11324.5	0.975
Aramberri	6,145	6,231	1.40%	271.4	0.946
Bustamante	2,389	2,461	3.03%	213.4	0.886
Cadereyta	55,231	53,750	2.68%	2625.1	0.971
Carmen	38,780	37,889	2.30%	1617.0	0.983
Cerralvo	4,283	3,924	8.39%	404.4	0.819
Ciénega	31,005	31,145	0.45%	1022.8	0.989
China	5,543	5,006	9.69%	528.2	0.804
Doctor Arroyo	13,349	13,514	1.24%	416.8	0.967
Doctor Coss	1,246	1,332	6.90%	171.8	0.610
Doctor González	2,106	2,176	3.30%	164.4	0.916
Galeana	16,424	14,353	12.61%	1343.1	0.828
García	157,779	159,344	0.99%	6513.5	0.984
San Pedro	41,237	41,562	0.79%	1221.9	0.986
General Bravo	3,397	3,529	3.87%	283.9	0.869
Escobedo	154,216	158,423	2.73%	4950.5	0.990
General Terán	7,473	7,591	1.57%	337.4	0.936
General Treviño	1,096	1,144	4.42%	114.6	0.787
General Zaragoza	2,437	2,416	0.86%	106.3	0.950
General Zuazua	42,305	43,865	3.69%	3807.9	0.933
Guadalupe	207,183	223,892	8.06%	9320.9	0.976

Source: Own elaboration.

Table 2: Continuation

Municipality	Current 2020	Prediction 2020	MAPE	RMSE	R^2
Los Herreras	1,899	1,990	4.81%	222.2	0.751
Higueras	835	877	5.03%	80.0	0.887
Hualahuises	3,361	3,441	2.37%	173.3	0.946
Iturbide	1,454	1,468	0.98%	41.0	0.979
Juárez	181,463	185,374	2.16%	8410.2	0.980
Lampazos	3,014	3,125	3.67%	258.8	0.863
Linares	31,978	32,543	1.77%	1314.8	0.970
Marín	2,215	2,143	3.27%	182.9	0.920
Melchor Ocampo	921	953	3.52%	84.1	0.876
Mier y Noriega	2,662	2,648	0.52%	37.9	0.993
Mina	2,727	2,827	3.66%	202.3	0.923
Montemorelos	30,139	30,706	1.88%	1440.3	0.963
Monterrey	368,796	362,609	1.68%	19583.3	0.925
Parás	830	865	4.26%	82.8	0.834
Pesquería	57,402	56,465	1.63%	1973.0	0.988
Los Ramones	4,610	4,813	4.41%	391.0	0.884
Rayones	1,675	1,720	2.71%	97.0	0.926
Hidalgo	5,900	5,721	3.03%	321.0	0.955
Salinas	36,175	36,044	0.36%	564.4	0.997
San Nicolás	133,725	152,747	14.22%	11326.0	0.914
Sabinas Hidalgo	16,128	14,915	7.52%	1403.2	0.882
Santa Catarina	90,966	90,209	0.83%	3743.0	0.982
Santiago	20,785	21,427	3.09%	1373.5	0.935
Vallecillo	1,315	1,384	5.25%	155.3	0.734
Villaldama	2,603	2,690	3.34%	188.2	0.898

Source: Own elaboration.

The municipalities that are part of the Metropolitan Area, which represent 27 per cent of the total state, have a coefficient of determination greater than 0.92 and a *MAPE* below 8.1 per cent, except for the municipality of San Nicolás de los Garza, which has an R^2 of 0.914 and a *MAPE* of 14.22 per cent. Both indicators show good accuracy for the housing projection in municipalities with higher demand. For the rest of the municipalities, 33 per cent have an R^2 between 0.91 and 0.99, with a maximum *MAPE* of

3.66 per cent. 27 per cent of the municipalities have an R^2 between 0.80 and 0.89, with a maximum *MAPE* of 12.61 per cent, and the remaining 12 per cent have an R^2 between 0.61 and 0.79, with a maximum *MAPE* of 6.90 per cent. Only in six municipalities does the model fail to achieve a coefficient of determination above 0.8. However, when evaluating the number of dwellings these municipalities represent in the state in 2020, it accounts for only 0.5 per cent.

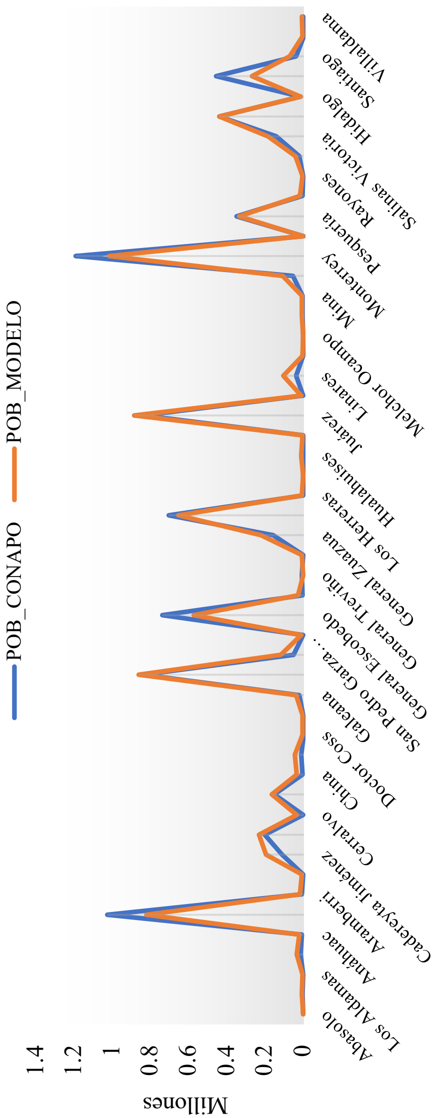
The results obtained through the regression models were compared with the municipal projections made by CONAPO in its report Reconstrucción y proyecciones de la población de los municipios de México, 2024. In CONAPO's projections, an estimated population of 7,769,371 inhabitants was forecasted for the state of Nuevo León, broken down by municipality. The result from the proposed model, although its main focus is to estimate the number of dwellings per municipality, can adjust its response variable. For this comparison, the number of dwellings obtained by the model for each municipality was multiplied by the average number of inhabitants for the year 2040 according to INEGI (3.1 inhabitants), resulting in 7,812,106 inhabitants. Figure 5 shows the results for each municipality of both projections (CONAPO, Reconstrucción y proyecciones de la población de los municipios de México, 2024).

(CONAPO, Conciliación Demográfica de México 1950-2019, 2023) in its report Conciliación Demográfica de México 1950-2019 y Proyecciones de la Población de México y de las entidades federativas 2020-2070 indicates that, starting in 2053, the country's population will begin to decline. This information is relevant for determining and adjusting based on reports generated by INEGI and CONAPO, allowing for parameterization of the model and recalculation of housing needs in the state of Nuevo León. For now, according to the CONAPO 2023 report, it mentions that Nuevo León will continue to increase its population at least until 2070. As shown in Figure 6, the proportion of the population in Nuevo León will increase from 4.8 to 6.4 percentage points from 2024 to 2070.

Figure 7 allows for the visualization of the initial modeling results for the year 2020 based on data from 1970 regarding housing coverage in the state of Nuevo León. This step is considered crucial for verifying the modeling process.

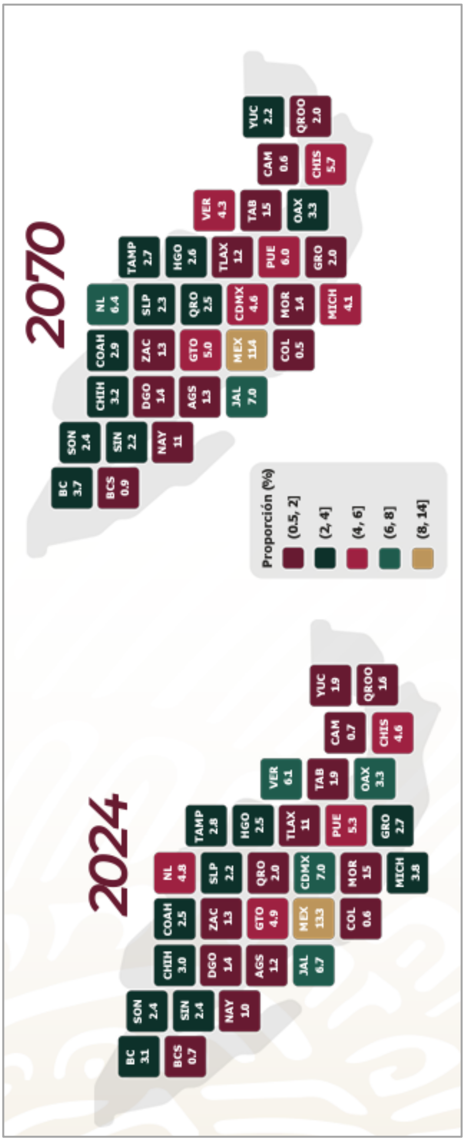
The accuracy of the results in the prediction model is acceptable, given that the model includes only limited factors present in the INEGI report, which have contributed to the actual development and growth of housing in the state.

Figure 5: Comparative projection of municipal population 2040 CONAPO and results of the proposed modeling to the year 2040



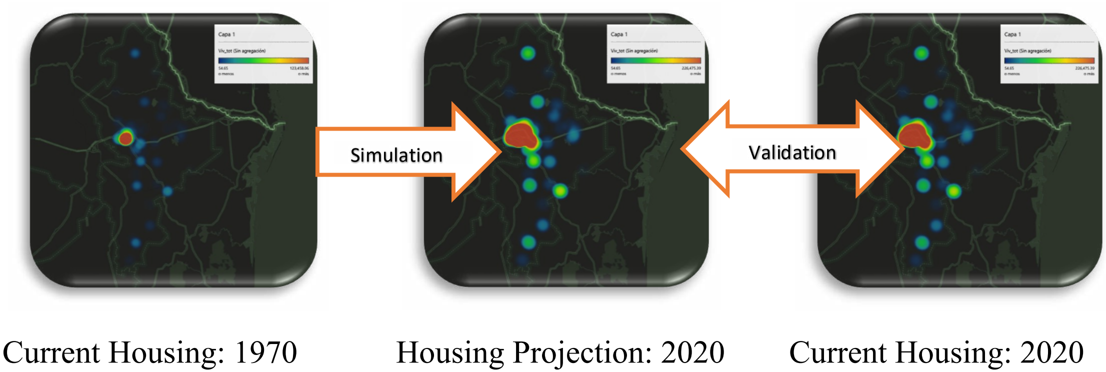
Source: Own elaboration.

Figure 6. Proportion of population in relation to the total of the country (2024 to 2070)



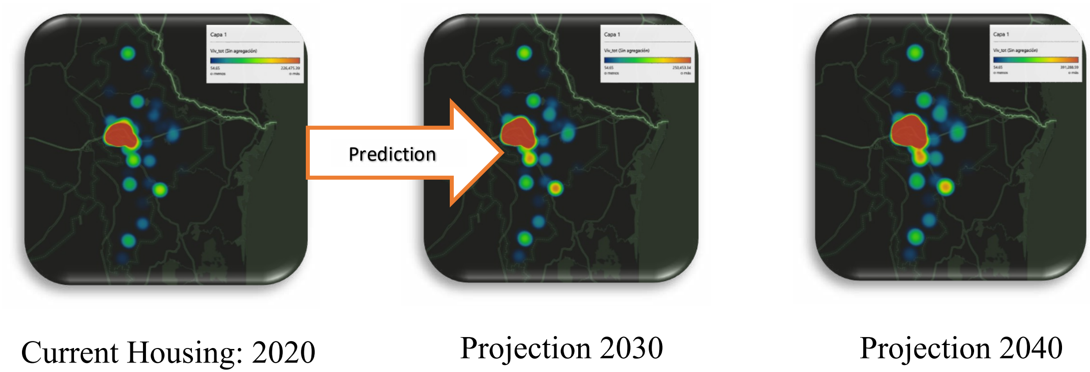
Source: CONAPO 2023.

Figure 7: Actual land use for housing and growth predicted by linear and polynomial regression for 2020 in the state of Nuevo León, Mexico



Source: Own elaboration.

Figure 8: Prediction of housing growth predicted by linear and polynomial regression for the years 2030 and 2040 in the state of Nuevo León, Mexico



Source: Own elaboration.

Other factors, such as commercial and industrial facilities, healthcare services, roadways and transportation, education, recreation, and zoning restrictions, could have an impact on housing growth. However, they were not included in the prediction model due to the lack of available data. Based on these findings and considering our variables and parameters, the prediction process was completed for the years 2025, 2030, 2035, and 2040.

The results show an exponential increase in the municipalities of García, Juárez, Apodaca, Pesquería, Escobedo, El Carmen, Zuazua, Salinas Victoria, and Guadalupe, with growth ranging from 303 per cent in the case of García to 45.9 per cent for Guadalupe. For the purposes of this study, the municipalities with the greatest interest for the development of new housing are highlighted.

In Table 3, the column “Interest” categorizes the municipalities into High, Medium, or Low interest based on the forecasted number of houses over a 20-year period. Municipalities classified as High interest are those where the number of houses exceeds 15,000 within the specified period. Medium interest municipalities fall within the range of more than 8,000 houses but fewer than 15,000. The Low interest group includes municipalities or areas with a projected growth of fewer than 8,000 houses over the next 20 years.

The previous result allows for the consideration of areas with the greatest potential for residential development. The goal is for this analysis to help identify the direction of urban growth and facilitate the search for residential plots to be acquired in advance. This forward-looking perspective would enable the identification of areas with greater sales potential and market opportunities for the state’s population.

Table 3: Result of housing projections for the 51 municipalities of the state of Nuevo León for the years 2025, 2030, 2035 and 2040

Municipality	Current 2020	Pred 2025	Pred 2030	Pred 2035	Pred 2040	Interest
García	157,779	241,051	346,040	477,140	637,180	Alto
Juárez	181,463	261,608	355,966	470,345	606,642	Alto
Apodaca	212,801	274,441	329,414	389,401	454,402	Alto
Pesquería	57,402	90,486	135,429	192,736	263,848	Alto
General Escobedo	154,216	192,621	230,171	271,073	315,327	Alto
El Carmen	38,780	61,680	93,282	133,741	184,103	Alto
General Zuazua	42,305	64,289	90,017	121,642	159,758	Alto
Salinas Victoria	36,175	52,974	74,673	101,724	134,708	Alto
Guadalupe	207,183	243,472	263,053	282,633	302,214	Alto
Ciénega de Flores	31,005	46,687	66,689	91,700	122,268	Alto
Monterrey	368,796	385,226	407,842	430,458	453,075	Alto
San Nicolás	133,725	164,942	177,137	189,332	201,527	Alto
Cadereyta Jiménez	55,231	65,744	79,084	93,771	109,805	Alto
Santa Catarina	90,966	99,253	108,297	117,341	126,385	Alto
Montemorelos	30,139	36,994	44,259	52,563	61,973	Alto
Linares	31,978	37,239	42,173	47,296	52,560	Alto
Santiago	20,785	25,052	28,994	33,242	37,786	Alto
San Pedro	41,237	44,970	48,378	51,786	55,194	Medio
Doctor Arroyo	13,349	16,030	19,050	22,614	26,767	Medio
Cerralvo	4,283	13,885	15,012	16,139	17,267	Medio
China	5,543	13,885	15,012	16,139	17,267	Medio
General Terán	7,473	9,193	11,207	13,683	16,669	Medio
Los Ramones	4,610	5,959	7,289	8,811	10,532	Bajo
Lampazos de Naranjo	3,014	3,936	4,953	6,200	7,700	Bajo
General Bravo	3,397	4,307	5,246	6,363	7,672	Bajo
Aramberri	6,145	7,127	8,122	9,217	10,411	Bajo

Source: Own elaboration.

Table 3: Continuation

Municipality	Current 2020	Pred 2025	Pred 2030	Pred 2035	Pred 2040	Interest
Sabinas Hidalgo	16,128	16,183	17,451	18,719	19,987	Bajo
Allende	13,785	13,885	15,012	16,139	17,267	Bajo
Los Herreras	1,899	2,553	3,261	4,128	5,172	Bajo
Mier y Noriega	2,662	3,220	3,928	4,787	5,812	Bajo
Hualahuises	3,361	4,029	4,693	5,435	6,258	Bajo
Agualeguas	2,971	3,604	4,237	4,939	5,710	Bajo
Villaldama	2,603	3,216	3,821	4,507	5,280	Bajo
Rayones	1,675	2,189	2,772	3,481	4,328	Bajo
Bustamante	2,389	2,966	3,530	4,154	4,837	Bajo
Mina	2,727	3,304	3,821	4,375	4,963	Bajo
Doctor González	2,106	2,585	3,033	3,518	4,037	Bajo
General Zaragoza	2,437	2,813	3,256	3,747	4,286	Bajo
Galeana	16,424	15,325	16,297	17,269	18,241	Bajo
Hidalgo	5,900	6,208	6,695	7,182	7,669	Bajo
Anáhuac	8,119	8,107	8,665	9,224	9,783	Bajo
Doctor Coss	1,246	1,622	1,971	2,382	2,859	Bajo
Los Aldamas	1,623	1,950	2,325	2,749	3,220	Bajo
General Treviño	1,096	1,399	1,709	2,081	2,519	Bajo
Vallecillo	1,315	1,651	1,962	2,322	2,732	Bajo
Melchor Ocampo	921	1,191	1,479	1,819	2,219	Bajo
Iturbide	1,454	1,659	1,867	2,091	2,333	Bajo
Parás	830	1,035	1,222	1,425	1,644	Bajo
Marín	2,215	2,347	2,552	2,757	2,962	Bajo
Higueras	835	996	1,111	1,218	1,315	Bajo
Abasolo	1,194	1,264	1,385	1,505	1,626	Bajo

Source: Own elaboration.

Results and discussion

In general, for the study, the results presented for predicting housing growth demonstrate the capability of the linear and polynomial regression model to produce a realistic modeling using the available data from INEGI.

This is evidenced by the good accuracy achieved during the validation process, as shown in Table 2.

Table 3 presents the results of the projections for the number of future homes in each municipality over five-year periods. The results show that the simulated growth follows a pattern similar to the current one, concentrating specifically in the Metropolitan Area of Monterrey, with municipalities experiencing the most expansion in the Northwest area, such as García, and in the Eastern area, with the municipality of Juárez, in reference to the city of Monterrey.

According to the images in Figure 8, the urban areas with the highest growth in the number of homes will expand more significantly toward the periphery of the state capital (Monterrey), to the East with the municipalities of Cadereyta, Pesquería, and Juárez; to the North with the municipalities of Salinas Victoria, Ciénega de Flores, El Carmen, and General Zuazua; to the Northwest with the municipalities of García and General Escobedo; and to the South, with a slower but still significant growth, the municipalities of Santiago, Montemorelos, and Linares.

It is important to mention that all of this calculated expansion is subject to certain relevant factors, such as land use for each municipality, primarily agricultural or industrial use, given that urban areas will expand based on the transformation of areas initially not intended for housing development. Considering industrial and economic growth, especially in the Monterrey Metropolitan Area, these factors trigger more rapid urban growth, meaning the transformation of land into urbanized areas. This results in the depletion of land designated for agricultural purposes in the state of Nuevo León.

The advantages provided by the estimations on the forecast and direction of growth for each municipality in the state of Nuevo León through the proposed method include the ability to avoid urban growth in protected areas or high-risk zones, reduce information gathering costs by utilizing the databases provided by INEGI, and offer quick and reliable analysis for the evaluation, modification, and management of urban development plans by both the government and private entities.

The quality of the model can be significantly affected by the number of time intervals. It is recommended to adjust the model each time INEGI reports its results and compare them with the model's predictions to validate its accuracy. In cases where the coefficient of determination is low, adjustments should be made to improve it.

Conclusions

Forecasting models have proven to be a great asset in supporting urban planning decisions. With the help of these models, knowledge and understanding of system dynamics are gained, future trends can be predicted, new needs can be identified, impacts can be analyzed, and various policies can be simulated and optimized to improve quality of life. Additionally, these models help minimize environmental and social deterioration caused by poor planning or uncontrolled growth.

The results of the modeling invite reflection and understanding of housing growth in the cities of the state of Nuevo León, which contributes to the establishment of planning strategies that improve the quality of life for residents. Modeling the magnitude and location of housing expansion are key concepts for urban planning in the state, offering an approach to understand future needs and quantify various costs associated with urbanization, particularly public services. The state of Nuevo León, like other states in the northwest of the country, faces prolonged drought seasons, leading to water shortages and resulting in strict water supply cut programs for the population. The model is also useful for forecasting the increase in the need for potable water in certain areas, identifying these critical areas, planning water infrastructure, evaluating water management policies, and promoting sustainability in the state.

The modeling allows for visualization and suggestion of various contributions based on the results provided by the models. For example, it can indicate the proportion of the increase in all types of housing, primarily social interest and middle-income housing, in both horizontal and vertical construction segments for each of the 51 municipalities in the state of Nuevo León, with a particular emphasis on the municipalities of the metropolitan area. Another important contribution is that it helps to dimension environmental and water resources aspects, especially considering the regional characteristics of the state. The enormous pressure from the rapid urbanization of the metropolitan area's outskirts will be reflected in non-urban lands, leading to a reduction in agricultural land use and affecting tourist sites surrounding the state capital.

The speed of urban growth is concerning. The model also supports decision-making by quantifying the population and considering the needs for roads, transportation demand, water viability, education, healthcare, and environmental risks, among other elements that must be carefully planned, both by the government and urban developers.

The analysis of the results shows that the expansion of settlements will not be random, but rather concentrated in certain areas, which complicates the urban planning designs due to space limitations. It is recommended to use linear and polynomial regression models for predicting housing growth, as they provide valuable information to improve decision-making in the development of city infrastructure in the state of Nuevo León.

Technically, the model helps professionals understand how the housing growth phenomenon works in the state by municipality. While the urban growth projection model is capable of evaluating the size, pattern, and trends of future urban growth with acceptable accuracy, the model falls short in assessing other important factors that contribute to a comprehensive urban development. These factors may include aspects such as environmental impact, land use, infrastructure availability, and the social and economic dynamics of the population, which are essential for holistic urban planning.

References

Aczel, A. (2006). *Complete Business Statistics*. India: Mc Graw Hill.

Aguilar, A. G. (1999). "Mexico City growth and regional dispersal: the expansion of largest cities and new spatial forms". In *Habitat International*, 391-412 https://doi.org/10.1016/s0197-3975(99)00014-4.

Alkheder, S. (2008). "Calibration and Assessment of Multitemporal Image-based Cellular Automata for Urban Growth Modeling". In *Photogrammetric Engineering and Remote Sensing*, 1539-1550. https://doi.org/10.14358/pers.74.12.1539.

Araya, Y. H. (2010). "Analysis and modeling of urban land cover change in Setúbal and Sesimbra, Portugal". In *Remote Sensing*, 2(6), 1549-1563. https://doi.org/10.3390/rs2061549.

Bárcenas, A. (2020). "Tiene Nuevo León relevancia para la economía nacional". En *El Financiero* 06/01/2020. Recuperado de https://www.elfinanciero.com.mx/monterrey/tiene-nuevo-leon-relevancia-para-la-economia-nacional/

CONAPO (2023). *Conciliación Demográfica de México 1950-2019*. Gobierno de México: Consejo Nacional de Población (CONAPO) .

CONAPO (2024). *Reconstrucción y proyecciones de la población de los municipios de México 1990-2040*.

Fang, S. G. (2004). "The impact of interactions in spatial simulation of the dynamics of urban sprawl". In *Landscape Urban Plan*, 294-306 https://doi.org/10.1016/j.landurbplan.2004.08.006.

INEGI (2020). *Panorama Sociodemográfico de Nuevo León*. Censo de Población y Vivienda. México: INEGI.

Lewis, C. (1982). *Industrial and Business Forecasting Methods*. London: Butterworths.

López, E. B. (2001). "Predicting land-cover and land-use change in the urban fringe: case in Morelia city, Mexico". In *Landscape Urban Plan*, 55(4), 271-285, https://doi.org/10.1016/s0169-2046(01)00160-8.

Makridakis, S. S. (1997). *Forecasting Methods and applications*. New York: Wiley.

Mandelas, E. H. (2007). "A fuzzy cellular automata based shell for modeling urban growth–a pilot application in Mesogia area". In *Information Science*, 1-9.

Montgomery, D. C. (2012). *Introduction to Linear Regression Analysis*, Fifth Edition. New Jersey: Wiley.

Ostertagová, E. (2011). *Applied Statistics (in Slovak)*, First Edition. Elfa: Košice.

Poelmans, L. (2009). "Complexity and Performance of Urban Expansion Models". In *Computer Environment and Urban Systems*, 34(1), 17-27. https://doi.org/10.1016/j.compenvurbsys.2009.06.001.

Rojas, N. I. (2020). "Modelación de crecimiento urbano". En *Bitácora* 32, 177-190, https://doi. org/10.15446/bitacora.v32n1.87758.

Roy A, A. Z. (2022). "CurFi: An automated tool to find the best regression analysis model using curve fitting". In *Engineering Reports*, 4(12):e12522., DOI: 10.1002/eng2.12522.

Sante, I. G. (2010). "Cellular Automata Models for the Simulation of Real-World Urban Processes: A Review and Analysis". In *Landscape and Urban Planning*, 108-122, https://doi.org/10.1016/j.landurbplan.2010.03.001.

Seto, K. C. (2002). "Monitoring land-use change in the Pearl River Delta using Landsat TM". In *International Journal Of Remote Sensing*, 23(10), 1985-2004. https://doi.org/10.1080/01431160110075532.

Sharma, V. K. (2020). "Modeling and Forecasting of Covid-19 Growth Curve in India". In *Transactions Of Indian National Academy Of Engineering*, 5(4), 697-710. https://doi.org/10.1007/s41403-020-00165-z.

Stow, D. A. (2002). "Sensitivity of multitemporal NOAA AVHRR data of an urbanizing region to land-use/landcover changes and misregistration". In *Remote Sensing Of Environment*, 80(2), 297-307. https://doi.org/10.1016/s0034-4257(01)00311-x.

Suárez, M. (2007). "La expansión urbana probable de la Ciudad de México. Un escenario pesimista y dos alternativas para el año 2020". In *Estudios Demográficos y Urbanos*, 22(1), 101-142., https://doi.org/10.24201/edu.v22i1.1295.

Thapa, R. B. (2012). "Scenario based urban growth allocation in Kathmandu Valley, Nepal". In *Landscape Urban Planning*, 140-148., https://doi.org/10.1016/j.landurbplan.2011.12.007.

Wahyudi, A. (2013). "Cellular automata for urban growth modeling: a chronological review on factors in transition rules". In *Proceedings of CUPUM 2013*: 13th International Conference on Computers in Urban Planning and Urban Management-Planning Support Systems for Sustainable Urban Development, 1-21. CUPUM.

Xiao, J. S. (2005). "Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing". In *Landscape And Urban Planning*, 75(1-2), 69-80. https://doi.org/10.1016/j.landurbplan.2004.12.005

Author's Curriculum Summary

Refugio Chávez Hernández

Is an assistant professor in the Department of Computer Science and Industrial Engineering at the University of Monterrey, Nuevo León, Mexico. He is an Industrial and Systems Engineer and holds a master's degree in Industrial and Systems Engineering with an emphasis on Logistics. He is currently pursuing a Ph.D. in Logistics and Supply Chain Management at the Popular Autonomous University of the State of Puebla. Professor Chávez has conducted research and consulting projects. His research interests include productivity, simulation, optimization, and lean manufacturing. He is a member of IISE and an advisor to the IISE student chapter at the University of Monterrey.
Email: refugio.chavez@udem.edu
Record ORCID: http://orcid.org/0000-0001-5706-656X

José Luis Martínez-Flores

Holds a Ph.D. in Engineering with a specialization in Systems Engineering. He is currently a Professor and Researcher in the Faculty of Industrial and Logistics Engineering at the Popular Autonomous University of the State of Puebla. His area of interest focuses on the Optimization of Logistics Operations in the Supply Chain. He was awarded the National Logistics Award in 2013. Dr. Martínez-Flores has served as a consultant for various national and international companies in the logistics sector. He is a National Researcher, Level I, recognized by SNI-CONACyT to date.
Email: joseluis.martinez01@upaep.mx
Record ORC ID: http://orcid.org/0000-0003-2986-469X

Claudia Malcón Cervera

Has been a professor at the UPAEP Business School since 2007. She earned her Bachelor's degree in Computer Systems, Master's degree in

Administration, and Ph.D. in Strategic Planning and Technology Management from UPAEP in 1992, 2001, and 2008, respectively. She completed a Diploma in Data Mining with Business Applications at Tecnológico de Monterrey in December 2013, a Certification in Applied Data Science with Python from Columbia University, USA, in May 2021, a Certification in General Consulting from CONOCER in August 2021, and a Certification in Storytelling and Persuasion with Data from INCAE in Costa Rica in September 2022. She has published articles in peer-reviewed national and international journals, conference proceedings, and participated in international conferences. In 2013, she developed an application to predict student dropout rates in higher education institutions. In 2022, she co-authored the study "Data Management Situation and Its Link to Data Strategy in Latin America." Her current areas of expertise and research include Business Intelligence, Text Mining, Data Mining (descriptive and predictive models), Business Analytics, and Big Data.
Email: claudia.malcon@upaep.mx
Record ORCID: http://orcid.org/0000-0003-1352-8990

Patricia Cano Olivos

Is a professor and researcher at the Popular Autonomous University of the State of Puebla (UPAEP). She holds a degree in Industrial Engineering, a Master of Science in Industrial Engineering from the Instituto Tecnológico de Apizaco, and a Ph.D. in Logistics and Supply Chain Management from UPAEP. She is a member of the National System of Researchers. Her research focuses on logistics operations and risk analysis and resilience in supply chains. Dr. Cano Olivos has presented at various national and international conferences, earning the Outstanding Research Award twice at the Global Conference on Business and Finance. She has published articles in peer-reviewed and indexed journals, as well as the book Disaster Risk Reduction in Mexico: Methodologies, Case Studies, and Prospective Views.
Email: patricia.cano@upaep.mx
Record ORCID: http://orcid.org/0000-0003-1754-6619

Article received on October 2, 2023 and accepted on August 15, 2024