Hedonic prices for land sales on real estate websites in the city of Mérida, Mexico

Miguel Ángel Monroy Cruz[a]

[a] Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación (INFOTEC), Mexico.
Email address: monroy_miguel@ciencias.unam.mx

Date received: April 4, 2025. Date of acceptance: November 24, 2025.

Abstract

This study analyzes the setting of prices of undeveloped land offered for sale online in Mérida, using data from real estate websites from 2021 to 2022 and public sources. A Durbin Spatial Error Model is estimated to capture spatial dependence on prices and environmental characteristics. The results show that surface area and temporal evolution consistently affect land value. Distance from economic centers has a limited effect, while variables in the immediate environment, such as schools and services in neighboring lots, show a positive association. This study demonstrates the need to incorporate spatial factors and contextual data in order to understand the land market.

Keywords: hedonic prices; land supply; web scraping; spatial econometrics; Mérida.

1. INTRODUCTION

Between 2012 and 2022, the supply of social housing in Mexico decreased by 54.1%. This decline was even more pronounced in specific segments: economical housing fell by 94.6% and low-income housing by 73.5%, according to data from the Single Housing Registry (RUV) (2024). One of the main causes of this contraction is the limited availability of land in areas with the public services and urban infrastructure necessary for the population. The Sectoral Business Office (2024) reports that the supply of land has shifted to peripheral or semi-urban areas, which have limited access to basic services and fewer formal job opportunities. In this context, analyzing the factors that determine land prices is important for understanding the structural challenges of the real estate market in Mexico and its link to the availability of affordable housing.

Land costs are an important component of the final price of housing. In residential housing, land costs represent 13.4% of the total cost. In social housing, land costs amount to 40.1%, according to estimates based on prices published on real estate websites and financing data provided by the National Housing Commission (CONAVI) (2024). Despite its importance, little research has been conducted on the land market in Mexico due to a lack of public data, which limits the ability to design appropriate public policies to address the housing deficit and meet the needs of lower-income groups.

This study helps to fill this gap by using modern data collection and analysis techniques. A database was created using data obtained through web scraping of real estate websites, supplemented with public sources relating to the physical and social conditions of the environment. The initial database contained 33,375 records for the state of Yucatán, 22,820 of which were located in the city of Mérida (CDM). After applying filtering criteria to eliminate outliers, data entry errors and duplicates, a final sample of 17,274 observations was obtained, allowing for consistent statistical analysis. Previous studies in the region have worked with smaller samples, for example, Sandoval Escalante and Becerril García (2025) used only 228 observations.

In addition to using traditional hedonic regressions, this study incorporates a spatial component in order to correct autocorrelation in the data and identify relationships between prices and environmental characteristics. This approach is relevant in Yucatán, where the land market is fragmented by geographical differences, accessibility and urban structure among municipalities. The results demonstrate that prices depend not only on the physical attributes of the land, but also on the characteristics of the neighborhood or surrounding neighborhoods.

Analyzing the market for undeveloped land provides insight into the urbanization challenges facing Yucatán. From 2000 to 2022, the state recorded an annual population growth rate of 1.6%, compared to 1.2% nationwide. Recent migration has accentuated this demographic dynamism. For example, between 2015 and 2020, 100,200 individuals arrived in the state, which is double the number recorded between 2010 and 2015. According to data from the National Institute of Statistics and Geography (INEGI) (2021), by 2020, 4.3% of the total population of the state was migrant. This population increase puts pressure on the land market and on the availability of land suitable for urbanization.

Studying land in Yucatán is important because population growth, both local and due to migration, together with recent urban expansion, has intensified demand for land, driving up prices and putting pressure on the availability of suitable housing space.

Mérida, located in southeastern Mexico is a tourist and cultural hub of national importance that faces a territorial dynamic marked by differentiated patterns of urbanization. These conditions raise questions about how land prices are established, in order to guide urban development and housing access policies.

This study is organized into six sections following this introduction. The second section presents a review of the literature on hedonic pricing from theoretical and empirical perspectives. The third section describes the data used and the process of constructing and refining the database. The fourth section presents descriptive statistics and initial market patterns. The fifth section details the econometric methodology, followed by a discussion of the results and their implications. The paper concludes with conclusions and suggestions for future research.

2. LITERATURE REVIEW

The literature review focuses on the hedonic approach applied to the study of real estate markets, emphasizing the importance of considering the intrinsic attributes of land and environmental characteristics. Palmquist (1984) points out that the sociodemographic attributes of the resident population, such as level of education, income and household structure, influence the demand for certain spatial attributes. These variables influence location preferences and affect the spatial distribution of prices. From this perspective, the hedonic price function reflects not only the physical characteristics of the asset, but also the social and economic dynamics that structure urban areas. This evidence justifies using models that incorporate spatial heterogeneity and market segmentation, particularly in regions with recent urban expansion, such as the CDM.

The theoretical basis for the hedonic approach stems from the work of Rosen (1974), which argues that the price of a heterogeneous asset, such as housing or land, stems from the overall valuation of its attributes. These attributes include surface area, location, accessibility, availability of services and environmental conditions. Each attribute contributes differently to the total price. This approach assumes that consumers choose the combination of attributes that maximize their well-being subject to budget constraints, which simultaneously determines their willingness to pay and the price structure in the market.

The theoretical and empirical literature reviewed in this section supports the importance of integrating physical, social and spatial variables into hedonic models and analyzes the most suitable methodologies for capturing the complexity of urban land value. These contributions form the conceptual basis for the analysis developed in this study, where *Pi* represents the price function of the asset, which is determined by the set of its attributes:

$$P_i = f(z_1, \dots, z_k)$$

Consumers choose housing that provides the greatest utility, represented as:

$$u(x, \varepsilon, z)$$

where *x* are observable consumer characteristics, ε reflects unobserved preferences and z are the attributes of the asset.

Market equilibrium is achieved when the marginal valuation of each attribute by the consumer is equal to the price increase associated with that attribute:

$$\frac{\partial p(z)}{\partial z} = \frac{\partial u(x, \varepsilon, z)}{\partial z}$$

This allows us to measure the value that consumers place on each attribute. The gradients of the pricing function show the marginal utility associated with each characteristic: how much the price increases with a unit change in each attribute.

The literature on hedonic pricing recognizes that identifying the implicit prices of attributes requires additional assumptions. Heckman *et al.* (2010) demonstrate that this identification is limited within a single homogeneous urban market because exogenous variation in attributes is often insufficient. The authors point out that the hedonic function is not necessarily linear or continuous, as residential choice processes generate discontinuities derived from heterogeneity of preferences, market segmentation and budget constraints.

Nonlinearity is related to the process of spatial ordering, whereby households differ in terms of income, preferences and sociodemographic composition, and are distributed in neighborhoods with heterogeneous amenities, generating segmented submarkets. This process of self-selection produces differentiated price function curves or segments, as documented by Nesheim (2004) and Day *et al.* (2004). According to these authors, the socioeconomic composition, the quality of schools, age distribution and ethnic diversity influence the shape of the price function. Similarly, Goodman and Thibodeau (1998) suggest that urban markets are structured into submarkets where households are grouped by income, race and preferences for services, revealing patterns.

In this context, incorporating variables that capture the attributes of an asset and its surroundings improves the identification of hedonic parameters. Explicitly considering the immediate surroundings, including amenities, services, social composition and quality of education, helps to reflect the segmented structure of the market and to avoid biased estimates resulting omitting relevant attributes.

At the international level, spatial components have been incorporated into hedonic models. Glaesener and Caruso (2015) and Munshi (2020) demonstrate that accessibility, urban amenities and neighborhood characteristics significantly influence land and housing prices. Techniques such as geographically weighted regression, employed by Soler and Gemar (2018) and Wang *et al.* (2022), capture spatial heterogeneity using functions that vary according to location. Glumac *et al.* (2019) applied the Spatial Durbin Error Model (SDEM) to estimate a land price index in Belgium, which highlights the importance of considering spatial autocorrelation and indirect effects. The authors categorize hedonic variables into eight groups: *1)* accessibility, *2)* proximity, *3)* physical attributes, *4)* environmental, *5)* legal, *6)* social, *7)* socioeconomic, and *8)* economic. This classification, which we use in this study, makes it easier to systematically include attributes of the asset and its surroundings.

In the field of spatial parametric models, Elhorst (2010) summarizes advances in models designed for spatially dependent data. These models capture direct and indirect effects between geographically close observations, such as the SDEM itself. This approach is particularly suitable for real estate markets, where property prices are often influenced by the value of neighboring properties and the structure of the immediate environment.

In the case of Mexico, empirical evidence is limited due to data restrictions and is mainly concentrated in Mexico City. Sobrino (2014) identifies submarkets defined by sociodemographic characteristics in the Metropolitan Area. Atuesta *et al.* (2018) found that labor informality, level of education and distance to employment centers influence housing prices. Meanwhile, Martínez-Jiménez *et al.* (2022) demonstrate that neighborhood characteristics and accessibility carry more weight than proximity to green or protected areas in informal peripheral markets. The increased use of web scraping techniques has enabled the creation of extensive, up-to-date databases, prompting more recent studies. Institutions such as the International Monetary Fund (IMF) (2022) and Savio *et al.* (2018) recognize the potential of these methodologies to generate timely, low-cost real estate indicators.

In the case of the CDM, Aguilar *et al.* (2025) document a pattern of diffuse expansion with variations in amenities and land prices. Evidence shows internal differentiation: the south and the surrounding municipalities such as Kanasín, Umán and Caucel, offer economical and low-income housing, while the north and Conkal are home to middle-class residential developments (Pérez Medina, 2024). This spatial heterogeneity means that prices cannot be assumed to be uniform, as contextual attributes are not randomly distributed. This gives rise to the need to incorporate spatial dependency structures into hedonic models. Therefore, this study proposes a hedonic model with a spatial component to capture the spatial heterogeneity and segmentation of the land market in the CDM. This model uses digital data and robust statistical techniques.

Variables used in literature to analyze real estate prices

The selected variables used to explain land prices in the CDM were based on the literature on hedonic prices, identifying relevant physical, social and spatial attributes in previous studies. For example, Möller (2009) and Atuesta *et al.* (2018) point out that real estate prices are closely linked to the labor market because the willingness to pay for a property depends on the economic opportunities in the region. The latter authors highlight that labor informality, level of education and distance to employment centers influence housing prices, which justifies including variables associated with the socioeconomic environment and labor accessibility.

Regarding accessibility and services, Glaesener and Caruso (2015) and Munshi (2020) found that proximity to amenities and points of interest increases property values due to increased competition for well-connected locations. Roback (1982) offers additional support by showing that urban amenities, particularly those associated with population density, positively impact land prices, estimating increases of up to US$6.30 for every additional 100 individuals per square meter (m^2).

Treg (2010) and Glumac *et al.* (2019) include indicators such as the proportion of the population over 65 and the unemployment rate to reflect local socioeconomic status. Likewise, Glumac *et al.* (2019) include legal variables related to land use, which are limited in Mexico. In response, this study uses an approach based on the classification of land as “ejido”, communal land, human settlements, agricultural zones or other categories. This classification, supported by Bojórquez-Luque (2011), serves as a proxy for identifying areas undergoing urbanization.

Web scraping and its application in the analysis of real estate prices

The use of web scraping techniques to analyze real estate prices has been endorsed by international organizations for its advantages over traditional data collection methods. The IMF (2022) recognizes that this methodology allows for more timely and frequent information gathering at a reduced operating cost. The Economic Commission for Latin America and the Caribbean (ECLAC) has applied these techniques in studies of Brazil, Ecuador and Peru, highlighting their effectiveness in producing up-to-date indicators. However, ECLAC also identifies limitations, such as the absence of a formal sampling framework and the fact that published prices may differ from the actual transaction price.

In academic circles, web scraping has become a well-established tool for studying real estate market dynamics. For example, Bricongne *et al.* (2021) developed a price index for the United Kingdom based on daily downloads from real estate websites. While this approach enables the identification of broad trends, it fails to consider individual property characteristics or attributes of the urban environment. In Mexico, initiatives such as the Banorte Housing Price Index (Banorte, 2025) have used this methodology to analyze the evolution of housing prices in cities nationwide, based on information from the main real estate websites.

Additionally, studies such as that by de Souza *et al.* (2021) combine data obtained through web scraping with spatial analysis tools. In the case of El Salvador and Brazil, the authors use spatial autocorrelation indices to identify geographical patterns in price distribution, showing that local variations are strongly associated with characteristics of the immediate environment.

Finally, Pegueroles *et al.* (2021) applied a hedonic approach to construct a price index for the Metropolitan Area of Chile. Their strategy breaks down the price of properties according to structural attributes, such as size and age, as well as contextual attributes related to accessibility and availability of services. Studies such as these confirm the potential of web scraping to generate comprehensive, up-to-date databases that strengthen econometric analyses of the real estate market.

3. DATA

Methodology for data collection and quality

To ensure the quality, stability and consistency of the data used to analyze the land market in the CDM, the recommendations from Belchev's web scraping manual (2020) and the methodological guidelines from studies such as Boeing and Waddell (2017) were followed. These guidelines establish criteria to ensure that the information obtained from digital platforms is reliable and suitable for statistical and econometric analyses.

The collection process prioritized stable sources with content from original publishers, avoiding sites with high turnover or frequent duplication of advertisements. Real estate websites with a long online presence were primarily used. Likewise, relevant variables for the overall characterization of the land were downloaded, such as total price, surface area, geographic location, description of the property and date of publication. Another criterion was to ensure a high volume of observations per period, in order to allow for robust and representative analyses of each region.

The final database includes 33,375 observations for the state of Yucatán, 22,820 of which pertain to the CDM. At the municipal level, Mérida accounts for 62.4% of the observations, followed by Progreso (20.27%) and Conkal (14.84%). These three municipalities account for 97.5% of the CDM records. In contrast, Umán, Kanasín, Tixpéual and Ucú have a negligible share, each with values of less than 1.1%.

For methodological reasons, municipalities with low representation were excluded. Belchev (2020) argues that a sufficient volume of observations is essential to obtain stable estimates, while small samples tend to inflate variance and increase sensitivity to outliers. The decision to focus on Mérida, Progreso and Conkal ensures a representative and statistically solid sample.

Data refinement was carried out by eliminating extreme percentiles within each municipality, following the approach applied by Boeing and Waddell (2017). The lowest and highest 1% of the price per m2 variable was excluded to reduce the effects of data entry errors, non-representative records and fraudulent data. This method was chosen over approaches based on influential observations (Leone *et al.*, 2019) because the latter assumes independence between data. However, in real estate markets, hedonic prices exhibit nonlinear relationships and spatial dependence, as documented in studies such as those by Nesheim (2004) and Day *et al.* (2004).

Finally, the observations were grouped by total price, surface area, latitude, longitude and the quarter in which the information was collected, counting the number of times each observation appears with these repeated variables. In this respect, the analysis was carried out with a total of 17,274 observations.

Economic centers

According to Atuesta *et al.* (2018), one of the key variables that determines real estate prices is proximity to the economic centers of a region. For this reason, it was decided to identify the economic centers in the state of Yucatán. To this end, INEGI (2022) publishes the National Statistical Directory of Economic Units, which records economic units classified by size and location.

Economic centers were identified using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) methodology, which is a density-based clustering algorithm (Shah, 2012). This exercise focused on analyzing the density of medium and large economic units, considering their spatial distribution.

As a result of this methodology, two economic centers were identified in Yucatán. The densest center was located in Mérida, the state capital, and the second center was located in Puerto Progreso, the main port and one of the main beaches in the region.

Basic services

Based on data from INEGI (2022), basic services and their locations were selected. These include elementary, middle and high schools, as well as general and specialized hospitals. Banks were also considered, including both branches and ATMs.

Tourist amenities

Yucatán has established itself as a leading tourist destination in Mexico, attracting both domestic and foreign visitors. Between 2017 and 2022, the total number of tourists grew at an average annual rate of 15.5%, higher than the national average of 11.4% for the same period (SECTUR, 2025). In this context, the linear distance of the land from the coast was calculated.[1] This calculation was performed based on the work of Sobrino (2014), who points out that proximity to tourist amenities positively impacts real estate values.

Sociodemographic characteristics

For the purpose of including sociodemographic variables of the population living in the vicinity of the land, we drew on the findings of Glaesener and Caruso (2015), Treg (2010) and Glumac *et al.* (2019), who concluded that the demographic and socioeconomic characteristics of a region significantly impact real estate prices. For this study, we used data published by the National Electoral Institute (INE) (2024) in conjunction with INEGI, which provides sociodemographic information at the electoral district level. We considered using this information because of its statewide geographic coverage, which includes data from both rural and urban areas,[2] allowing for an analysis with a larger amount of data downloaded from the Internet.

Additionally, we incorporated data provided by the National Agrarian Registry (RAN) (2022), specifically the available plot files containing information on the delimitation, use and ownership of ejido and communal land in Mexico. Including this data was intended to identify the land offered for sale in these areas.

4. DESCRIPTIVE STATISTICS

Table 1 presents the descriptive statistics of the physical, social and spatial characteristics of the analyzed land, as well as the variables of accessibility and legal status. The total price shows wide dispersion, with an average of $1,500,000 Mexican pesos (MXN) and a standard deviation of $1,800,000 MXN, indicating variability associated with structural and environmental differences between plots of land. The average surface area is 597.1 m², with a deviation of 494.8 m², reflecting an uneven distribution in plot size. The price per m² also varies significantly: the average price is $2,698.20 MXN and the standard deviation is $1,900.40 MXN, indicating differences even within nearby geographical areas.

Table 1. Descriptive statistics

Intrinsic variables			
Characteristics of the land	Total price	Millions of current pesos	1.5
			(1.8)
	Surface area	Current square meters	597.1
			(494.8)
	Price per square meter	Pesos	2 698.2
			(1 900.4)
Environmental variables			
Urban and sociodemographic amenities	Distance to economic center	Kilometers	11.5
			(4.4)
	Number of services[a]	Number of services	2.7
			(8.2)
	Schooling	Years studied	11.1
			(1.8)
	Distance to coast	Number of plots less than 3 km away	1 555
	Distance to coast	Number of plots more than 3 km away	15 719
Legislation and type of land[a]	Ejido lands	Private	9 429
	Ejido lands	Communal	7 845
	Type of land	Agricultural zone	103
		Other zones[b]	2 467
		"Human settlements"	14 704
Environmental variables			
Characteristics of the download[c]	First quarter	In the quarter	8 841
		In another quarter	8 433
	Second quarter	In the quarter	2 877
		In another quarter	14 397
	Third quarter	In the quarter	4 613
		In another quarter	12 661
	Fifth quarter	In the quarter	2 429
		In another quarter	14 845
	Repeated observations	Not repeated	14 471
		Repeated	2 803

Notes: [a] Schools, hospitals, banks and ATMs are considered services; [b] Other areas include: wooded areas, jungles, mangroves, etc.; [c] The number of plots in each category is reported.

Source: Prepared by the author using data from the Internet, INEGI (2022), RAN (2022), INE (2024) and SECTUR (2025).

In terms of the immediate surroundings, the average distance to the nearest economic center is 11.5 km, with an internal variation of 4.4 km. Taking into account education, health and finance, the average number of services available is 2.7 services per plot of land, although there is high dispersion (8.2), associated with plots of land located in both urbanized areas and in areas with limited infrastructure. The median indicates that 50% of plots of land do not have services in their electoral section, suggesting differences in basic accessibility between areas.

In sociodemographic terms, the average years of schooling for the population in the immediate surroundings is 11.1, with a deviation of 1.8. Regarding spatial distribution in relation to the coast, 1,555 plots of land (9%) are located less than 3 km from the coastline. The remaining 15,719 (90%) are located farther away, which allows for differentiation between plots of land with potential for coastal tourism or residential use and those located farther inland.

Legally, 7,845 plots (54.6%) are ejido land, while 9,429 plots (45.4%) are privately owned. In terms of type of area, 14,704 plots (46%) are in areas classified as human settlements, 2,467 (9.5%) are in other areas and only 103 are in agricultural areas. This distribution indicates a market dominated by areas geared toward urban and peri-urban development, with a marginal share of agricultural land.

The data is distributed across different quarters of the study period: 8,841 records correspond to the first quarter, 2,877 to the second, 4,613 to the third and 2,429 to the fourth. It is possible that the same plot of land may appear in more than one quarter, which is explained by repeated records. In total, 2,803 observations correspond to plots of land recorded on multiple occasions, while 14,471 correspond to unique plots of land.

The exploratory analysis in Appendix 1 enables us to identify preliminary patterns in land price formation based on graphical relationships and comparisons between variables. The scatter diagrams comparing total price and continuous variables such as surface area, distance to the economic center, schooling and number of services show relationships accompanied by correlation coefficients that help interpret their influence. Surface area has a positive correlation of 0.51, though there is high dispersion, indicating that there are small plots of land with high prices, suggesting the involvement of other attributes not represented by a single physical dimension. Distance to the economic center has a negative correlation of 0.13 and average prices remain similar when the sample is divided according to proximity. This indicates that the relationship with price is not strictly linear. Average schooling has a positive correlation of 0.22, which is possibly associated with higher incomes in areas with greater human capital, while the number of public services has a low correlation of 0.05.

Comparisons with categorical variables reinforce the role of territorial context. Plots of land located in human settlements have the highest average prices, close to one million pesos, while agricultural land and land classified as other types have lower values. In terms of ownership, private land tends to fetch higher prices than ejido land, consistent with the legal restrictions on ejido land.

Territorial accessibility also shows associations with price. Distance to the coast has limited effects, as plots of land located less than 3 km from the coastline have similar average prices to those located farther away. Taken together, these findings demonstrate the heterogeneity of the land market in the CDM.

Figure 1 provides two maps that complement these results. The first map shows the spatial concentration of land offered for sale and highlights high-density areas on the northern and northeastern outskirts of Mérida, particularly in municipalities such as Conkal, as well as areas near Cholul. Smaller pockets of available land can also be seen in the south. The second map shows the spatial distribution of prices, where we can see that the highest values coincide with these corridors of urban expansion. Prices gradually decrease as the distance from the metropolitan center increases or in locations with less urban consolidation.

Figure 1. Number and total price of plots of land in Mexico City

Source: Prepared by the author using data from the Internet and the INE (2024).

Both maps show that, although land concentration is not uniform, there is a spatial pattern where higher prices tend to be found in regions close to the city center and in emerging areas of real estate expansion. This pattern is consistent with a dynamic of rising values linked not only to proximity to the historic center, but also to new nodes of urban growth.

5. METHODOLOGY

Spatial data analysis requires econometric techniques capable of capturing the dependence and autocorrelation present in geographically distributed variables. When real estate prices, infrastructure or socioeconomic characteristics are not randomly distributed across the territory, the assumptions of the Ordinary Least Squares model are not suitable. In this context, Elhorst (2010) assembles a variety of spatial models and describes their uses and differences. A significant advantage of these models is the interpretability of their parameters, which is essential for analyzing the determinants of land prices. This criterion was central to selecting the model to be used in this study.

To identify the factors affecting land value and recognize possible spatial spillovers, we selected the SDEM model. The SDEM model corrects spatial autocorrelation and incorporates spatial effects into the explanatory variables, providing a more accurate representation of the price formation process.

According to LeSage (2009), spatial econometric analysis should begin with an Ordinary Least Squares model and should evaluate the residuals using diagnostic tests. The first recommended tool is the Moran I statistic, which allows us to determine whether neighboring units have similar values. A significantly positive value indicates the presence of autocorrelation and evidence that the basic model omits relevant spatial relationships.

Elhorst (2010) subsequently suggests applying Lagrange multiplier tests and likelihood ratio tests to confirm autocorrelation in the dependent variable or errors and compare nested models to verify the presence of spatial processes.

While spatial models solve the autocorrelation problem, the construction of the spatial weight matrix *W* is often questioned, as there is no single, universal method for defining it. The choice depends on the phenomenon being analyzed and is the responsibility of the researcher. The *W* matrix is an n × n square matrix where each element w_{ij} represents the intensity of spatial interaction between units *i* and *j*:

The most common approach for constructing the *W* matrix is binary contiguity:

$$w_{ij} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ share a frontier} \\ 0 \text{ if they do not} \end{cases}$$

1 if *i* and *j* share a frontier

0 if they do not

This study uses an inverse distance matrix, as it captures the spatial influence continuously and on a decreasing scale, which aligns with the logic of the real estate market.

The *W* matrix formalizes the idea that the characteristics of neighboring areas influence local prices. For instance, if a school or hospital is located in a neighboring unit, we can quantify its impact on the price of land in the focal unit, even if said service is not located within the unit itself.

The SDEM addresses two different sources of spatial dependence: dependence on explanatory variables and dependence on errors:

$$y = \alpha + X\beta + WX\theta + u$$

y: vector of the dependent variable.

X: matrix of explanatory variables.

β: vector of coefficients of the explanatory variables (direct effects).

W: matrix of spatial weights.

WX: spatial term of the explanatory variables (effects of the variables of neighboring units).

θ: vector of coefficients associated with *WX* (indirect effects).

u: error term with spatial autocorrelation.

The error term can be broken down as follows:

$$u = \lambda W_u + \epsilon$$

λ: spatial autocorrelation coefficient in the errors.

W_u: spatial term of the errors.

ϵ: random error term (white noise), with $\epsilon \sim N(0,\sigma 2\ I)$.

The SDEM interpretation is based on direct and indirect impacts rather than elasticities.

Direct impact: measures how a change in X_i affects the price in the same unit *i*.

Indirect impact or spillover: measures how changes in X_j affect the price in neighboring units.

These effects allow us to measure the interaction between submarkets, reflecting the fact that prices depend not only on internal characteristics, but also on the spatial environment. In short, the SDEM provides an appropriate framework for assessing the spatial heterogeneity of land prices and how the characteristics of neighboring areas influence their formation.

6. RESULTS

The SDEM estimates the total price of land in the CDM by incorporating both the characteristics of the land itself and those of its closest neighbors. In Table 2, the results of the model show that the surface area variable has a significant negative effect, indicating that larger plots of land tend to have a lower price per m^2, which is a common trend in markets with a discount for scale or low liquidity. This relationship is also reflected in simple linear regression models,[3] where the surface area coefficient is positive for the total price (MXN$1,800 per additional meter) and negative for the price per m2 (MXN$0.4), confirming the a size penalty.

In terms of time, the first quarters show negative effects, indicating a progressive increase in land prices. Simple models show that prices in the first quarter are MXN$270.9 pesos per m^2 lower than in the last quarter, indicating an upward trend. Distance from the urban center shows negative effects in both the spatial model and the simple models (-MXN$35,000 and -MXN$44.7 per m^2). This pattern is interpreted based on the polycentric structure of Yucatán, where Mérida and Progreso function as differentiated urban nodes and there is no single dominant center.

The methodological decision to incorporate a polycentric structure into the analysis allows us to evaluate land value based on its relative proximity to different centers of activity. However, the results show that distance from urban centers loses explanatory power when compared to social and spatial variables, even when this is taken into account. This suggests that the immediate environment, accessibility and available services influence land prices more directly than proximity to a particular urban node. This behavior is consistent with previous studies documenting processes of loss of centrality in historic centers in cities such as Culiacán and Guadalajara (Pérez Tamayo *et al.*, 2017; Mcenulty and Mercado, 2019).

The spatial approach also allows for the identification of dynamics associated with speculative processes. The model shows high prices in areas adjoining high-value land, even when the plots lack attributes that would explain this difference. This pattern is consistent with the existence of a speculative premium, a phenomenon documented in Latin American urban markets. Likewise, the average schooling in the surrounding area emerges as a significant variable in all specifications, indicating that the perceived human capital in the area affects land value.

Meanwhile, the number of services negatively affects the spatial model, as shown in Table 2. This result may be associated with oversupply or speculative anticipation processes in unconsolidated areas. Simple models also show that the presence of at least one basic service is related to lower prices, possibly because these plots are urbanized but have not yet seen an increase in value. Furthermore, ejido lands do not exhibit a significant direct effect; however, they appreciate in value when located close to areas with high prices, which may reflect expectations of regularization.

Table 2. Spatial regression results

	Linear regression				*SDEM regression*			
Variable	*Total price (MXN$ thousands)*		*Price per square meter (MXN$)*		*Total price (MXN$ thousands)*		*Price per square meter (MXN$)*	
				Direct effect				
Surface area	1.8	***	-0.4	***	1.9	***	-0.4	.
Average schooling	190.7	***	282.8	***	3.8		18.2	
Distance to economic center	-35.0	***	-44.7	***	-39.9		-225.2	*
Type of ownership: Ejido=0; Private=1								
Private	-379.2	***	-555.1	***	53.1		-20.5	***
Type of land								
Agricultural area	-613.3	***	-976.8	***	-83.4		-153.1	*
Others[1]	-255.4	***	-448.6	***	-130.9		-269.8	
Number of services: No services=0; With some services=1								
With some services	-17.8	***	-21.7	***	14.2	**	2.1	
Collection period								
First quarter	-209.8	***	-270.9	***	-299.6	***	-362.6	***
Second quarter	-252.1	**	-309.2	**	-225.5	**	-214.1	**
Third quarter	-139.2	***	-133.1	*	-150.7	*	-89.1	
Fourth quarter	-64.9	**	-66.8		-12.4		92.4	
Repetition of observation: Not repeated=0; Repeated at least once=1								
Repeated at least once	121.1	**	0.1	*	57.3		2.6	
				Indirect effect				
Surface area					-0.2	.	0.0	
Average level of education					185.7	***	272.4	***
Distance to economic center					-18.1		161.4	
Type of land ownership: Ejido=0; Private=1								
Private					-182.4	.	-114.0	
Type of land								
Agricultural area					28.8		13.6	
Others[1]					-297.0		-188.9	
Number of services: No services=0; Some services=1								
With some services					-37.5	***	-19.5	**
Data collection period								
First quarter					-18.6		-122.4	
Second quarter					-258.0		-479.7	*
Third quarter					9.7		-84.1	
Fourth quarter					-518.4	.	-288.9	
Repetition of observation: Not repeated=0; Repeated at least once=1								
Repeats at least once					3.6		41.3	

Notes: a The reported models correspond to SDEM estimates. For the total price regression: $\lambda = 0.65005$, pLR test = 1,813.8

(p < 2.22e-16), approximate standard error = 0.012698, z = 51.191 (p < 2.22e-16), Wald statistic = 2,620. 6 (p < 2.22e-16), log-likelihood = –117,191.7, ML residual variance = 1.4238×10^{12} (σ = 1 193 200), Nagelkerke's pseudo-R^2 = 0.49924, with 7 593 observations, 27 estimated parameters and AIC = 234 440. The values of the corresponding linear model are: R^2 = 0.3, adjusted R^2 = 0.3, standard error = 1 441 622.3, F statistic = 697.2 and p-value = 0.0. For the regression of price per m^2: λ = 0.65861, LR test = 3121.4 (p < 2.22e-16), asymptotic standard error = 0.0091959, z = 71.619 (p < 2.22e-16), Wald statistic = 5 129.3 (p < 2.22e-16), log-likelihood = –65 402.93, residual variance ML = 1 622 800 (σ = 1 273.9), Nagelkerke's pseudo-R^2 = 0.44986, with 7 593 observations, 27 estimated parameters and AIC = 130 860. The values of the corresponding linear model are: R^2 = 0.1, adjusted R^2 = 0.1, standard error = 1 749.6, F statistic = 237.9 and p-value = 0.0.

Source: Prepared by the author using data from the Internet, INEGI (2022), and INE (2024).

The model has some limitations due to the absence of relevant variables, such as proximity to cenotes or tourist developments. Without these variables, the model cannot fully explain certain observed patterns, such as land located on golf courses, where prices are above average. Finally, opportunities for methodological improvement were identified by incorporating advanced measures of accessibility, such as travel times or road connectivity indicators. While the methodology is replicable, interpreting the results requires adaptation to each context because land prices depend on individual attributes and the characteristics of the environment.

7. CONCLUSIONS

The analysis of the land market in the CDM, based on a spatial econometric approach, shows how physical, temporal and spatial attributes influence the pricing of land. The most stable result among the estimated models is the negative relationship between land area and price per m2. Larger plots of land have lower unit prices, suggesting a market structure in which size penalties consistently apply. This pattern appears in both spatial and non-spatial specifications, indicating that it is a persistent feature of the Yucatecan land market.

In terms of timing, the results reflect a sustained increase in land prices during the analyzed period. This behavior can be associated with gradual appreciation processes or investment dynamics aimed at obtaining capital gains. The observed upward trend in prices supports the need to analyze the interaction between urban expansion, appreciation expectations and the behavior of those involved in selling and purchasing land.

Regarding spatial factors, the results show that proximity to specific urban centers, such as Mérida or Progreso, alone does not explain price variation. Although these distances were incorporated from the outset to capture a polycentric logic, the model indicates that the pricing of land depends more on the immediate environment and effective accessibility. These findings suggest that the territorial structure operates as a network of interconnected nodes, where services, connectivity and relative position within the urban system carry more weight than simple proximity to a dominant center.

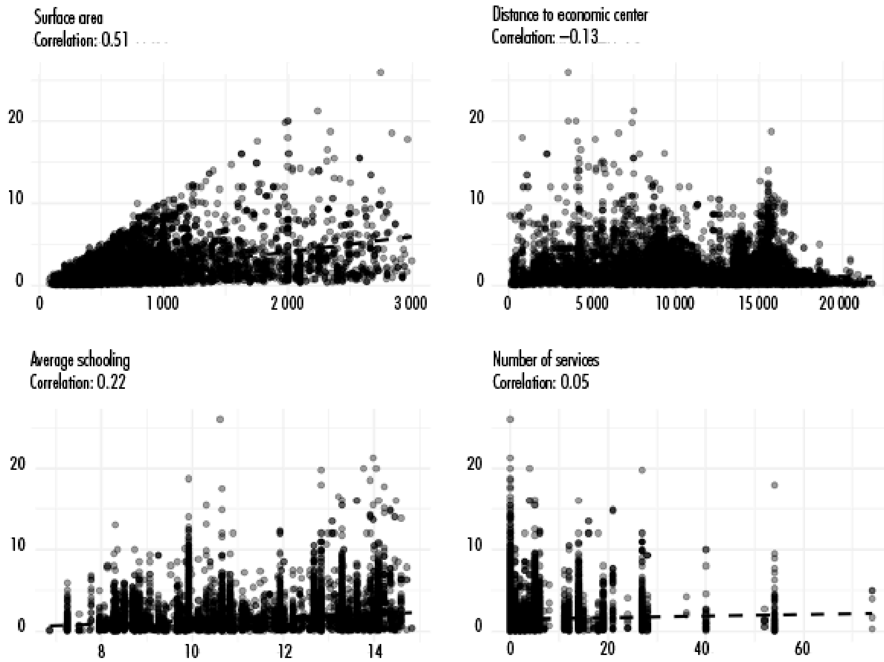
The model also identifies patterns associated with speculative investment processes. High prices are observed in areas adjacent to valued areas without a clear correspondence to visible attributes, indicating anticipation of capital gains. Meanwhile, the positive effect of the educational environment and the negative effect of the number of available services suggest purchasing behavior aimed at taking advantage of expectations of urban growth. These dynamics are present in municipalities such as Kanasín and Umán and reveal the importance of designing fiscal and institutional instruments that mitigate speculative land retention and guide land use toward productive or residential purposes.

From a public policy standpoint, the results of the model indicate that infrastructure and basic services influence land prices more because of their potential for future appreciation than for their immediate value. This underscores the need to prioritize public investment in areas with proven demand and to control urban expansion driven by speculation.

The variability of prices between similar areas reflects the existence of information imbalances. Consequently, it is advisable to create market monitoring bodies that incorporate land registers, public records and geospatial data. Finally, while the model is replicable, its interpretation must consider the restrictions of the available variables and the specific territorial characteristics of each region. One direction for future research is to extend this exercise to the national level, which would allow for the comparison of regional patterns and provide a more comprehensive view of how the land market functions in Mexico. Likewise, future research could strengthen the database by incorporating more attributes to improve the accuracy of the price function estimation.

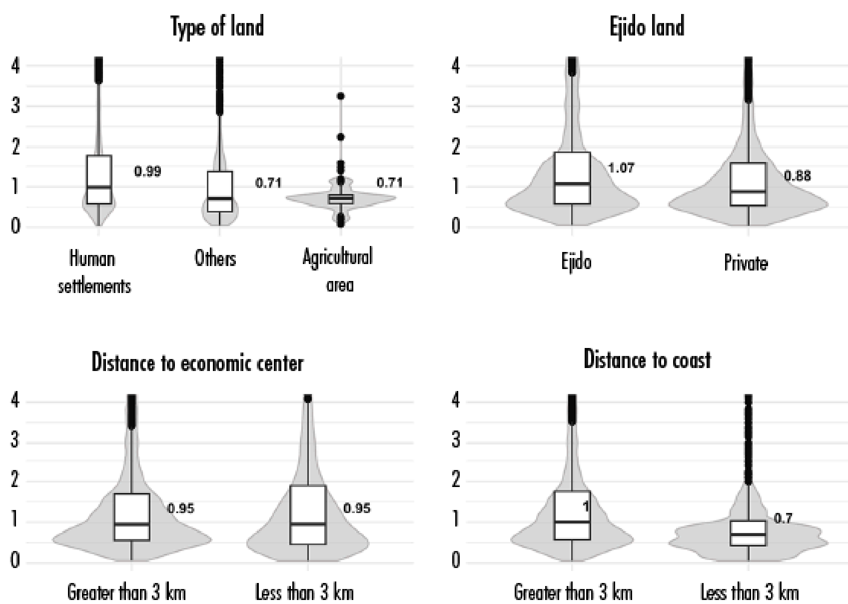
APPENDICES

Appendix 1A. Distribution of the total price of land and the continuous variables in the study



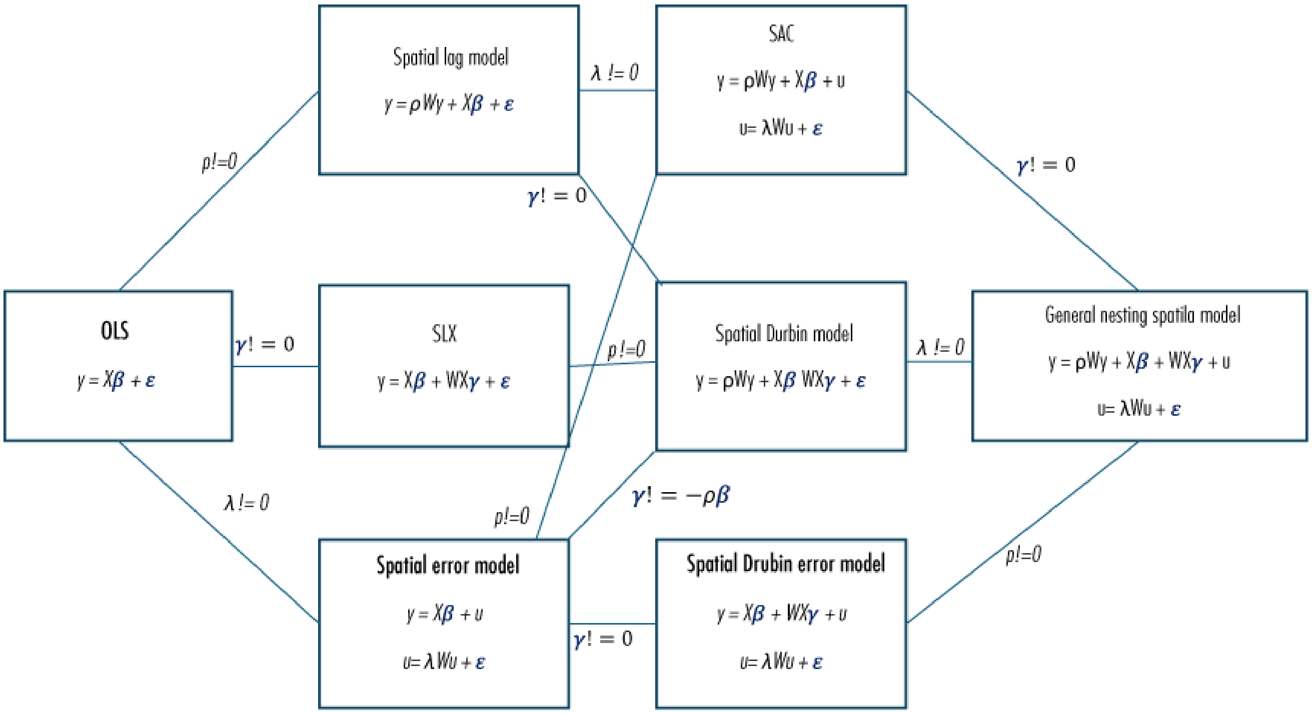
Source: Prepared by the author using data from the Internet, INEGI (2022) and INE (2024).

Appendix 1B. Distribution of prices by accessibility to tourist amenities and type of land



Source: Prepared by the author using data from the Internet, INEGI (2022), RAN (2022), INE (2024) and SECTUR (2025).

Appendix 2. Taxonomy of parametric spatial models



Source: Prepared by the author using data from Elhorts (2010).

APPENDIX 3

DIAGNOSTICS OF LINEAR REGRESSIONS

The models for total price and price per m2 have p-values equal to zero in practically all diagnostic tests, indicating problems with normality, heteroscedasticity, autocorrelation and model specification. Although the total price regression shows a moderate R² and the price regression per m2 shows a poor fit, in both cases the linear structure is insufficient to adequately describe the phenomenon.

Table 3A. Goodness-of-fit tests for linear regressions

Model	*Jarque-Bera*	*Anderson-Darling*	*Breusch-Pagan*	*White*	*Goldfeld-Quandt*	*Durbin-Watson*	*Breusch-Godfrey*	*Reset*	*R^2*	*Adjusted R^2*	*Average VIF*
Total price	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.3	1.5
Price per m^2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	1.5

Source: Prepared by the author using data from the Internet, INEGI (2022), RAN (2022), INE (2024) and SECTUR (2025).

Appendix 4

MORAN I GLOBAL RESIDUALS OF REGRESSIONS

The global Moran I test applied to the residuals of both regressions reveals marked positive spatial autocorrelation, with high z-statistic values and practically zero p-values. This indicates that the linear OLS model is insufficient and requires a spatial specification.

Table 4A. Spatial autocorrelation tests

Model	*Moran I observed*	*Expectation*	*Variance*	*z statistic*	*p-value*	*Interpretation*
Total price	0.4895	-0.0009	0.0000441	73.822	< 2.2e-16	Strong positive spatial autocor-relation in the residuals; the OLS model does not capture the spatial structure of the to-tal price.
Price per m^2	0.4895	-0.0009	0.0000441	73.822	< 2.2e-16	The residuals of the price per m^2 model also show spatial dependence, indicating the omission of relevant spatial processes.

Source: Prepared by the author using data from the Internet, INEGI (2022), RAN (2022), INE (2024) and SECTUR (2025).

APPENDIX 5

COMPARISON OF AIC BETWEEN OLS AND SDEM MODELS

The AIC comparison shows that SDEM spatial models offer a better fit than OLS, even with more parameters, for both variables. The reduction in AIC, especially for price per m2, confirms that spatial dependence is crucial and that spatial models better describe the real estate market.

Table 5A. Goodness-of-fit tests for linear regression and geographic regression

Total price	*df*	*AIC*	*Price per m2*	*df*	*AIC price m^2*	*Interpretation*
OLS	14	236 358.8	OLS	14	134 067.4	Linear models have a higher AIC, indicating a lower relative fit.
SDEM	27	234 437.4	SDEM	27	130 859.9	Spatial models reduce the AIC, im-proving the fit by capturing spatial dependence.

Source: Prepared by the author using data from the Internet, INEGI (2022), RAN (2022), INE (2024) and SECTUR (2025).

APPENDIX 6

LAGRANGE MULTIPLIER TEST (LM)

The LM tests reveal spatial autocorrelation in errors and evidence of spatial lag, which rules out the use of OLS and favors SEM, SDEM, SAR or SARAR models. Spatial error remains significant and lag matters most in the price per m2. The results confirm the need for spatial models.

Table 6A. Spatial autocorrelation and spatial lag tests

LM (Lagrange multiplier) test	Total price statistic	p-value total price	Statistical price per m^2	p-value price per m2	Interpretation
LMerr (spatial error)	2 632.2	< 2.2e-16	5 405.1	< 2.2e-16	Strong evidence of spatial auto-correlation in the error term in both models. A model with spatial error (SEM/SDEM) is prefer-able to OLS.
LMlag (spatial lag)	1748	< 2.2e-16	5 327.9	< 2.2e-16	There is also evidence of spatial lag in the dependent variable, suggesting the consideration of SAR/SARAR models in addition to the error component.
RLMerr (robust error)	887.39	< 2.2e-16	95.207	< 2.2e-16	Once the other effect has been controlled, the spatial error component remains significant in both cases, indicating a role for spatial dependence in the error.
RLMlag (robust lag)	3.2761	0.0703	17.973	2.24E-05	For the total price, the spatial lag ceases to be significant ($p \approx 0.07$), though it remains relevant for the price per m^2. The lag effect is more important in the price per m^2 model.
SARMA (joint lag+error)	2 635.4	< 2.2e-16	5 423.1	< 2.2e-16	The joint test confirms that the purely linear specification (OLS) is inadequate and that a spatial model is required (at least with spatial error, and in the case of price per m^2, with lag as well).

Source: Prepared by the author using data from the Internet, INEGI (2022), RAN (2022), INE (2024) and SECTUR (2025).

Bibliography

Aguilar, A. G., Flores-Espinosa, M. and Hernández, J. (2025). Metropolización, dinámica inmobiliaria y segregación socio-territorial: el caso de Mérida, Yucatán. *Revista EURE. Revista de Estudios Urbano Regionales, 51(153).* https://doi.org/10.7764/EURE.51.153.03

Atuesta, L. H., Ibarra-Olivo, J. E., Lozano-Gracia, N. y Deichmann, U. (2018). Access to employment and property values in Mexico. *Regional Science and Urban Economics, 70.* https://doi.org/10.1016/j.regsciurbeco.2018.03.005

Banorte (2025). Precios de la vivienda. https://www.banorte.com/wps/portal/banorte/Home/analisis-economico

Belchev, P. (2020). *Practical guidelines on web scraping for the HICP.* European Commission.

Boeing, G. and Waddell, P. (2017). New insights into rental housing markets across the United States: Web scraping and analyzing Craigslist rental listings. *Journal of Planning Education and Research, 37(4).* https://doi.org/10.1177/0739456X16664789

Bojórquez-Luque, J. (2011). Importancia de la tierra de propiedad social en la expansión de las ciudades en México. *Ra Ximhai.* https://www.redalyc.org/articulo.oa?id=46119239012

Bricongne, B., Meunier, B. and Pouget, S. (2021). Web scraping housing prices in real-time: The Covid-19 crisis in the UK. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3916196

Comisión Nacional de Vivienda (CONAVI) (2024). *Informe del sector vivienda 2024.*

Day, B., Bateman, I. and Lake, I. (2004). Nonlinearity in hedonic price equations: An estimation strategy using model-based clustering. *CSERGE Working Paper*.

de Souza, T., Fonseca, F., Fernandes, V. y Pedrassoli, J. (2021). Exploratory spatial analysis of housing prices obtained from web scraping technique. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4–2021.* https://doi.org/10.5194/isprs-archives-XLIII-B4-2021-135-2021

Dirección Sectorial Empresarial (2024). *Vivienda accesible en México: diagnóstico y propuestas de política pública*. INFONAVIT.

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis, 5(1).* https://doi.org/10.1080/17421770903541772

Glaesener, M. and Caruso, G. (2015). Neighborhood green and services diversity effects on land prices: Evidence from a multilevel hedonic analysis in Luxembourg. *Landscape and Urban Planning, 143.* https://doi.org/10.1016/j.landurbplan.2015.06.008

Glumac, B., Herrera Gomez, M. and Licheron, J. (2019). A hedonic urban land price index. Land Use Policy, 81. https://doi.org/10.1016/j.landusepol.2018.11.032

Goodman, A. and Thibodeau, T. G. (1998). Housing market segmentation. *Journal of Housing Economics, 7(2).* https://doi.org/10.1006/jhec.1998.0229

Heckman, J. J., Matzkin, R. L. and Nesheim, L. (2010). Nonparametric identification and estimation of nonadditive hedonic models. *Econometrica, 78(5).* https://doi.org/10.3982/ECTA6388

International Monetary Fund (IMF) (2022). *United Arab Emirates: Technical Assistance Report. Hedonic methods for price indices.* IMF Staff Country Reports, 2022(303).

Instituto Nacional de Estadística y Geografía (INEGI) (2021). Censo de Población y Vivienda 2020: Estadísticas de migración y población. INEGI. https://www.inegi.org.mx/programas/ccpv/2020/

___________ (2022). Directorio Estadístico Nacional de Unidades Económicas (DENUE), May 2022. INEGI. https://www.inegi.org.mx/app/descarga/?ti=6

Instituto Nacional Electoral (INE) (2024). Estadísticas Censales a Escalas Geoelectorales 2020. *Cartografía electoral 2024*.

Leone, A. J., Minutti-Meza, M. and Wasley, C. E. (2019). Influential observations and inference in accounting research. *The Accounting Review, 94(6).* http://dx.doi.org/10.2139/ssrn.2407967

LeSage, J. (2009). An introduction to spatial econometrics. *Revue d'Économie Industrielle, 123*. https://doi.org/10.4000/rei.3887

Martínez-Jiménez, E. T., Le Gallo, J. and Aguilar Ibarra, A. (2022). The effects of land price in the peri-urban fringe of Mexico City: Environmental amenities for informal land parcel purchasers. *Urban Studies, 59(1).* https://doi.org/10.1177/0042098020960968

Mcenulty, C. and Mercado, E. (2019). Entre el despoblamiento y la gentrificación: el Centro Histórico de Morelia, México. *Bitácora Urbano Territorial, 29(1).* https://doi.org/10.15446/bitacora.v29n1.70151

Möller, J. (2009). Regional variations in the price of building land: A spatial econometrics approach for West Germany. *The Annals of Regional Science, 43.* https://doi.org/10.1007/s00168-007-0207-6

Munshi, T. (2020). Accessibility, infrastructure provision and residential land value: Modelling the relation using geographically weighted regression in the city of Rajkot, India. *Sustainability, 12(20).* https://doi.org/10.3390/su12208615

Nesheim, L. (2004). Equilibrium sorting of heterogeneous consumers across locations. *Econometric Society North American Summer Meetings, Paper 337.*

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing. *The Review of Economics and Statistics, 66(3).* https://web.mit.edu/14.33/www/attach/palmquist.pdf

Pegueroles, P., Guerrero, R., Fernández, A. and López, D. (2021). Price index through web scrapping. *Revista Chilena de Economía y Sociedad, 15(1).* https://rches.utem.cl/wp-content/uploads/sites/8/2022/01/revista-chilena-de-economia-y-sociedad-vol15-n1-2021-Pegueroles-Guerrero-Fernandez-Lopez.pdf

Pérez Medina, S. (2024). Producción comercial de vivienda en la ciudad de Mérida: impactos en segregación y desocupación, 2013-2020. In E. R. Morales García de Alba and C. A. Ken Rodríguez (coords.). *Procesos urbanos y dinámica del mercado laboral.* Universidad Nacional Autónoma de México, Instituto de Investigaciones Económicas y Asociación Mexicana de Ciencias para el Desarrollo Regional.

Pérez Tamayo, B. N., Gil Alonso, F. and Bayona Carrasco, J. (2017). La segregación socioespacial en Culiacán, México (2000-2010): ¿de la ciudad dual a la ciudad fragmentada? *Estudios Demográficos y Urbanos, 32(3).* https://doi.org/10.24201/edu.v32i3.1660

Registro Agrario Nacional (RAN) (2022). *Base de datos de propiedad social 2022*.

Registro Único de Vivienda (RUV) (2024). *Inventario nacional de Vivienda 2024*.

Roback, J. (1982). Wages, rents and the quality of life. *Journal of Political Economy, 90(6).* https://doi.org/10.1086/261120

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82(1).* https://doi.org/10.1086/260169

Sandoval Escalante, B. and Becerril García, J. (2025). La comercialización de los lotes de inversión en Yucatán, México. *Península, 20(2).* https://doi.org/10.22201/cephcis.25942743e.2025.20.2.92105

Savio, G., Collinao, M. P., Lana, B. and Lara, R. (2018). Web scraping masivo de alquileres de viviendas. Comisión Económica para América Latina y el Caribe (CEPAL).

Secretaría de Turismo (SECTUR) (2025). *Anuario estadístico de turismo 2025.*

Shah, G. (2012). An improved dbscan: A density-based clustering algorithm with parameter selection for high-dimensional data sets. *NUICONE 2012.* https://doi.org/10.1109/NUICONE.2012.6493211

Sobrino, J. (2014). Housing prices and submarkets in Mexico City: A hedonic assessment. *Estudios Económicos, 29(1).* https://doi.org/10.24201/ee.v29i1.74

Soler, I. and Gemar, G. (2018). Hedonic price models with geographically weighted regression: an application to hospitality. *Journal of Destination Marketing & Management, 9.* https://doi.org/10.1016/j.jdmm.2017.12.001

Treg, C. (2010). *A multilevel property hedonic approach to valuing parks and open space* [Master's thesis, University of Vermont]. Graduate College Dissertations and Theses. https://scholarworks.uvm.edu/graddis/230

Wang, Z., Wang, Y., Wu, S. and Du, Z. (2022). House price valuation model based on geographically neural network weighted regression: The case study of Shenzhen, China. *ISPRS International Journal of Geo-Information, 11(8).* https://doi.org/10.3390/ijgi11080450

---

[1] The location of cenotes was also considered; however, INEGI lists a lower number than other sources (Google), so it was decided not to include them in the study.

[2] Census data from the Basic Geostatistical Area (AGEB) level was not considered since only urban AGEB data was published and information on land not located within these urban areas was lost. This information was used in the regression to reflect the schooling level of the school section.

[3] Appendices 3 to 6 present the statistical tests applied to the OLS models that reveal structural limitations justifying the use of spatial models. The LM and Moran I tests confirm strong spatial autocorrelation, both in the errors and in the dependent variable, which invalidates the traditional linear specification. Likewise, the comparison of the AIC between OLS and SDEM models shows substantial improvements in the fit when spatial dependence is incorporated, especially in the price per m2. The diagnostics of the models by total price and price per m2 reveal problems of normality, heteroscedasticity, autocorrelation and specification, indicating that the linear structure does not adequately capture the complexity of the real estate market. Taken together, these findings show that spatial dependence is a component in the price generation process and that geographic models offer a more accurate and consistent representation of the phenomenon by explicitly incorporating the spatial interaction between observations.