Recibido: 23-07-2021
Aceptado: 16-11-2021

SPATIAL CLASSIFICATION OF URBAN LAND BY SPECULATIVE LAND VALUE

AND MSI SATELLITE IMAGERY USING K-MEANS, IN HUANCAYO, PERU

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ

REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

GONZALO RODOLFO PEÑA ZAMALLOA [1]

1 Magíster en Ciencias Sociales
Universidad Continental, Huancayo, Perú
Profesor investigador
https://orcid.org/0000-0002-6141-6849
gzamalloa@outlook.com

DOI: https://doi.org/10.22320/07183607.2021.24.44.06

La ciudad de Huancayo, como otras ciudades intermedias en Latinoamérica, enfrenta problemas de cambios de uso de suelo poco planificados y una acelerada dinámica del mercado del suelo urbano. La escases y desactualización de información sobre el territorio urbano impiden la adecuada clasificación de áreas urbanas, limitando la forma de su intervención. Esta investigación tuvo como objetivo la incorporación de métodos no asistidos y mixtos para la clasificación espacial de zonas urbanas considerando el valor especulativo del suelo, la proporción del suelo urbanizado y otras variables geoespaciales. Entre los medios de recolección de datos, se usó imágenes Multi-Espectrales (MSI) del satélite Sentinel-2, el sistema vial primario y una muestra de puntos de observación directa. Los datos procesados fueron incorporados en mapas georreferenciados, a los cuales se añadió además los límites urbanos y pendientes oficiales. Durante el procesamiento de los datos se empleó el algoritmo K-Means, junto a otros métodos de *machine learning* y juicio asistido. Como resultado, se obtuvo una caracterización objetiva de zonas urbanas que difiere de la planificación existente.

Palabras clave: planificación urbana, mercado inmobiliario, periferia urbana, inteligencia artificial

The city of Huancayo, like other intermediate cities in Latin America, faces problems of poorly planned land-use changes and a rapid dynamic of the urban land market. The scarce and outdated information on the urban territory impedes the adequate classification of urban areas, limiting the form of its intervention. The purpose of this research was the adoption of unassisted and mixed methods for the spatial classification of urban areas, considering the speculative land value, the proportion of urbanized land, and other geospatial variables. Among the data collection media, Multi-Spectral Imagery (MSI) from the Sentinel-2 satellite, the primary road system, and a sample of direct observation points, were used. The processed data were incorporated into georeferenced maps, to which urban limits and official slopes were added. During data processing, the K-Means algorithm was used, together with other machine learning and assisted judgment methods. As a result, an objective classification of urban areas was obtained, which differs from the existing planning.

Keywords: urban planning, real estate market, urban periphery, artificial intelligence.

I. INTRODUCTION

Intermediate cities face land sale and value speculation processes that define the urban shape more quickly than intervention by local governments. Urban land value prioritizes the demand of private agents (Gasic, 2018), even though the State sets the limits of this action (Sabatini & Arenas, 200). Thus, socio-spatial segregation is related to the willingness of the market and inadequate policies, molding the excess demand or supply of the property market, and generating a disperse standard of urban life (Saleh, Hwa & Majid, 2016; López Navarrete & Peña Medina, 2017; Li, Sun & Boersma, 2019). Facing this, land regulation and its application can promote or stop the development of emerging urban areas (Yu, Zhou & Yang, 2019). These conditions are common in Latin American cities with limited public action (Sabatin & Arenas, 2000). Among the causes, clientelist practices are seen, the result of failures in free-market practices, private interests, illegal conditions, ambiguous regulations, and a generalized popular acceptance (Pimentel Sánchez, 2020; Espinoza & Fort, 2017).

Land value, the most important indicator of property market dynamics, is not easy to estimate or predict, although it is common that consolidated areas are overvalued, making the periphery more attractive due to its low price (Glaeser & Ward, 2009; García & Peralta, 2016; Gasparenienea, Venclauskienea & Remeikiene, 2014). In the long term, land market behaviors can come close to time series (Gaete, 2021). However, an approach with heterogeneous data, or scenarios of high uncertainty, can use artificial intelligence to classify them (Durduran, 2015; Belhadia *et al.*, 2020; Forestier & Wemmer, 2016). The K-Means algorithm has been useful and highly adaptable to classify images, study urban growth, and for spatial analysis (Liu *et al.*, 2021; Belhadia *et al.*, 2020).

Peru is in its bicentenary and is facing major economic and social challenges. The National Housing and Urbanism Policy considers the low impact of urban-territorial planning and the limited use of regulatory compliance a major problem (Ministry of Housing, Construction, and Sanitation [MVCS in Spanish], 2021). Property market processes, formal, illegal, or under mixed setups, are also common in Peruvian cities (Espinoza & Fort, 2017; Pimentel Sánchez, 2020). Even though the portfolio of support funds for housing, like the Mivivienda Fund (FMV, in Spanish), multiply, their implementation is limited by adverse urban conditions and land value (Calderón, 2015). During the property boom of 2018 to 2019, at least 70% of district municipalities did not have urban development plans (FMV, 2018a; FMV, 2018b). In Huancayo, the main city in the heart of Peru, the Provincial Municipality of Huancayo (2016) proposes development based on sustainable and inclusive principles, but that requires knowledge of the local urban reality and its objective characterization.

The purpose of this article is to spatially classify the urban areas in the city of Huancayo using heterogeneous data. The research proposes the differentiated classification of urban areas, incorporating unassisted and mixed methods, and considering the speculative value of the land on the property market, the proportion of developed land, the distance from main roads, and the slope of the land. The work was carried out in four connected stages: (1) construction of base maps; (2) processing of satellite images to analyze current land occupation; (3) application of machine learning methods for classification; and (4) polygonal classification of the urban areas of the city of Huancayo.

II. THEORETICAL FRAMEWORK

Urban planning and the property market

Due to population growth, better city planning represents an ongoing issue worldwide (Mouratidis, 2021). This issue has captured national attention to promote its development from a sustainable approach (Aceid & Fundación ACS; 2018; United Nations, 2018; Castillo-García, 2021), although during the pandemic, its reduced presence stood out (Moreno, Allam, Chabaud, Gall & Pratlong, 2021). Thus, in this context, a revision of the idea of proximity in the urban economy, linked to the generation of land value, is needed (Tricaricoa & De Vidovich, 2021).

Urban planning requires a balance between land use and urban expansion, which is not always aligned with the real ways of life and the behavior of the property market (López Navarrete & Peña Medina, 2017). There is a gap between the sustainable generation of urban space and the real practices in peri-urban areas adjoining rural areas and natural spaces, that are quickly being devastated by formal and informal urbanization processes (Carvajal, Moreira, Salazar, Leguia & Jorquera, 2019).

Socio-spatial segregation is related to the willingness of the property market and inadequate policies, and directly affects urban planning (López Navarrete & Peña Medina, 2017; Glaeser & Ward, 2009; Migueltorena & Lan, 2013). Excess supply or demand of the property market and dispersion generate variations in living standards (Saleh

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

et al., 2016). Fluctuations in land value, urban growth, and initial density condition these variations (Glaeser & Ward, 2009; Li *et al*., 2019).

In the growth stage of the sector, many of the rules that guide market actions are not easy to adapt to the management instruments, widening the gaps in urban planning (Glaeser & Ward, 2009). Among these, regulations of access to formal urban services have been made worse (Baer & Kauw, 2016). These disparities can be insurmountable, with repercussions on the generation of new policies and tax collection (Hindi, Moreira & Rossi, 2020; Foldvary & Minola, 2017). In addition, land value has greater variability than the buildings (Kok, Monkkonen & Quigley, 2014). As a result, this value cannot be suitably allocated for its use in regulations, mortgages, and loans, as its real fluctuation is characterized by speculation (Hwang, Park & Lee, 2013; Gasparenienea *et al*., 2014; Foldvary & Minola, 2017).

The need for development land and land value

Having a dwelling is one of the most important aspects of peoples' lives (Saleh *et al*., 2016). Those financed with social funds promote real estate investment and have great interest in the availability of undeveloped sites (Scotiabank, 2015; FMV, 2018c). However, the value of vacant urban land is subject to speculation with lower prices in peri-urban areas (Gedal & Ellen, 2018; Parias, 2008), which promotes exclusion on increasing physical distance and price (Gaete, 2021; Klaufus, Van Lindert, Van Noorloos & Steel, 2017). In this way, an incremental cycle of speculated value is entered which impedes reaching more homogeneous conditions (Amézquita, Rodríguez & Murillo, 2015; Gaete, 2021; Gasic, 2018; Araque Solano & Caballero Quintero, 2009; Glaeser & Gyourko, 2003).

The level of consolidation and proximity to roads are undeniable attractions of urban land (Peña-Zamalloa, 2018; Gedal & Ellen, 2018): Agricultural spaces with road access are, as a result, targets for change of land use (Salazar, 2014; Cardó, 2017; Migueltorena & Lan, 2013). On being informal lots, self-builds are prioritized, with the goal of reducing housing costs while disregarding long-term effects (GRADE, 2020; Salazar, 2014). The ongoing search for greater profit from land sales undermines the positive popular intention of urban planning (Delgadillo, 2016; Araque Solano & Caballero Quintero, 2009). Rapid price changes generate, in the territory, a disorganized and low-density occupation, even when social housing is promoted (Calderón, 2015). This affects the rural and natural environment and complicates access to urban services for spread-out areas, as well as regulatory compliance (Carvajal *et al*., 2019; Li *et al*., 2019). This reality flies in the face of the compact city (Vorontsova, Vorontsova & Salimgareev, 2016).

Low-density urban sprawl involves high costs in urban infrastructure (Nabil & Eldayem, 2015). Facing this outlook, an optimal urban model prioritizes accessibility and leads to short distances to multiple urban centers, and a reduction in mobility times (Yu *et al*., 2019; Gedal & Ellen, 2018; Graells-Garrido, Serra, Rowe, Cucchietti & Reyes *et al*., 2021). The idea of chronological urbanism is, in fact, an attempt to improve the quality of life of inhabitants on diverse geographical scales (Moreno *et al*., 2021; Graells-Garrido *et al*., 2021).

The characterization of urban land and the K-Means method

Little understood urban sprawl processes, without an articulated systemic analysis, occur randomly and go against the capacity of generating compact cities (Vorontsova *et al*., 2016, Alfasi & Migdalovich, 2020). In addition, the metrics tend to be single-dimensional (Tellier, 2020), when urban complexity requires using multidimensional analysis metrics for its classification (Steurer & Bayr, 2020; Tellier, 2020). Machine learning offers an alternative for clustering using heterogeneous data (Joshi, 2020). This classification can be assisted, unassisted, or mixed (Liu *et al*., 2018; Steurer & Bayr, 2020).

Specifically, K-Means is one of the most used unsupervised classification algorithms in images, random data, and unlabeled data (Liu *et al*., 2018; Zhou *et al.*, 2017). This algorithm allows generating clusters, grouping data under similar traits (Campesato, 2020), and differentiating elements like vegetation, vacant urban spaces, and even rural uses (Feng, Peng & Wu, 2020). Although the hierarchical cluster analysis, mobile mean, and maximization of expectations could be considered as being suitable alternatives, the use of a Euclidean distance allows that the classification made with K-Means can be overlapped to two-dimensional coordinates, and is suitable for geographic settings (Campesato, 2020; Joshi, 2020; Liu *et al*., 2018).

III. CASE STUDY

The city of Huancayo is in the central part of the country. Its geography is molded by the Mantaro River, and it constitutes one of the widest valleys in the Peruvian Andes, with a high capacity of development land that competes with rural land. The geographic scope of the study presented here, considered the Huamancaca

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

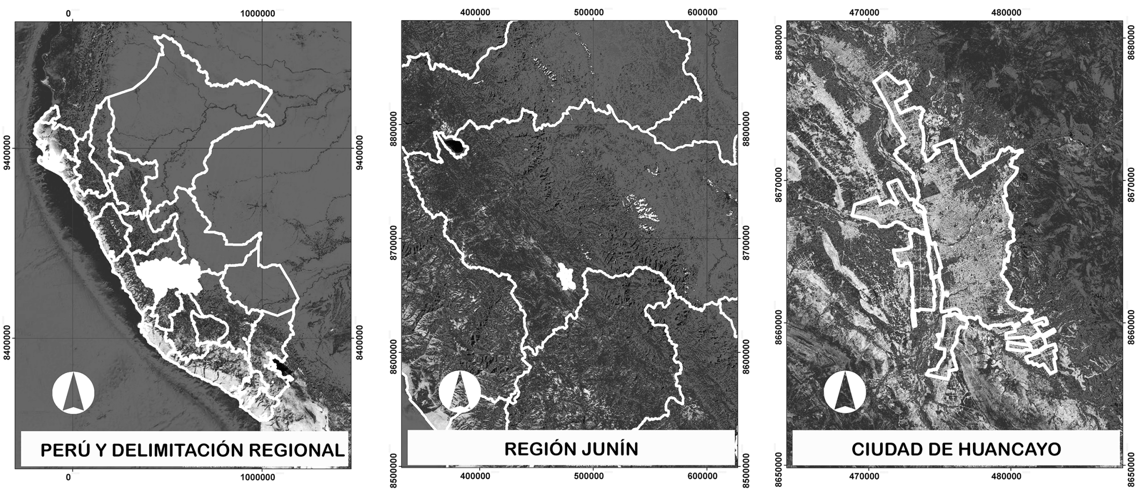


Figure 1. Location of the study area. Source: Preparation by the Author.

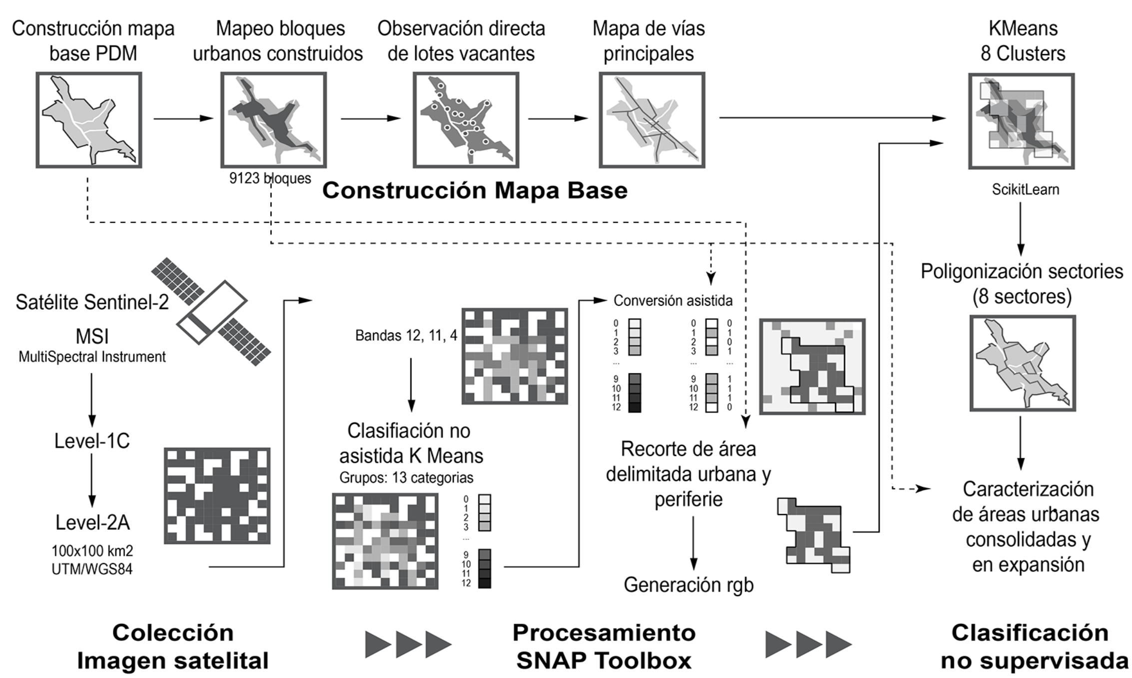


Figure 2. Data processing and collection methodology. Source: Preparation by the Author.

Dimension	Variable	Values
Geography	Slope of the land	Slope percentage.
	X Coordinate	UTM East
	Y Coordinate	UTM North
Urban road	Distance to the road	Distance in meters (m) to the closest road of the highway system
Urban occupation	Proportion of area occupied	Area occupied by buildings in a radius of 400 m / Area of blocks projected in a radius of 400m
Urban boundary	Distance to the urban boundary	Distance in meters (m) to the closest point of the projected urban boundary.
Land price	Land price offered by m2	Land price per meter squared (m2)

Table 1. Variables considered for the K-Means classification. Source: Preparation by the Author.

and 3 de Diciembre districts, of the adjoining Chupaca province, along with the districts of Pilcomayo, Chilca, Sapallanga, Huancán, El Tambo, and Huancayo in the Province of Huancayo, given their geographic location on the right bank of the river, and the direct connection they have with the city. The location is shown in Figure 1.

IV. METHODOLOGY

Data collection was made from different sources: direct observation, satellite images, and maps. These methods were digitalized and processed using geographic information systems, QGIS 3.12, SNAP Toolbox v8.9, scikit-learn 0.24 library, and others in python. The processing sequence can be seen in Figure 2, from the construction of the base map to the final generation of the urban sector polygons.

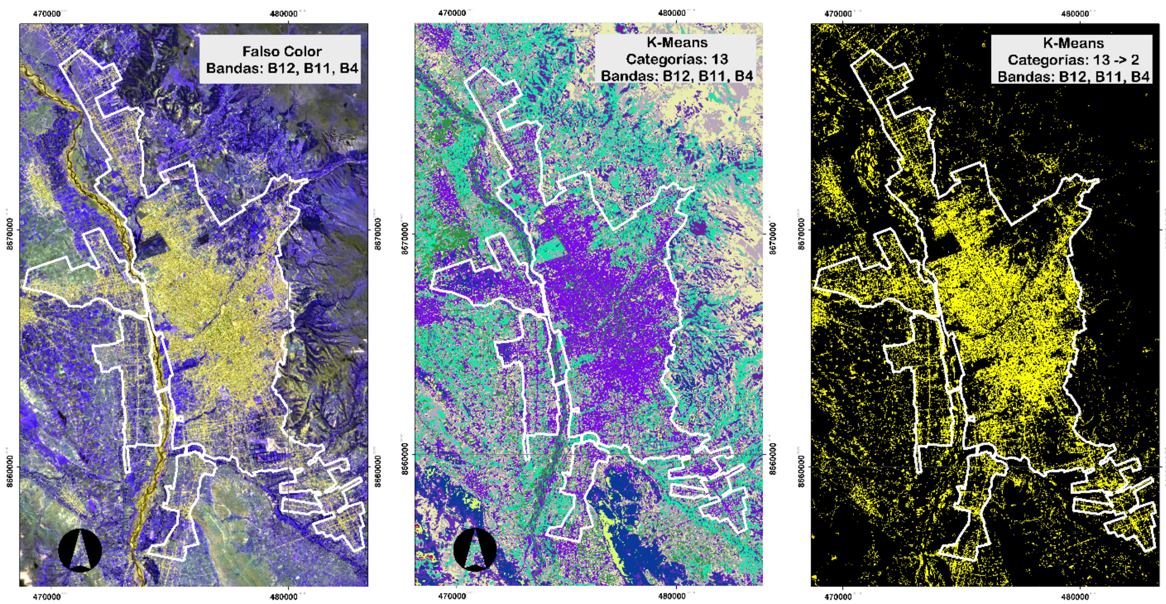
Following this, the image produced by the Multispectral Instrument (MSI), of the Sentinel-2 satellite, Level 2A product was used, which provides a reflectance image of the atmospheric background derived from the association of Level 1C, in an area comprising 100 x 100 km^2, under a URM/WGS84 cartographic projection. Resizing was needed for a suitable overlapping and re-projection. Thus, the images were processed with the SNAP v8.0 software, resizing the image for bands 12, 11, and 4, through which an rgb false-color image was generated. Once the bands were isolated, a classification was generated using the unsupervised K-Means classification algorithm. The number of categories was set after examining the results of between 3 and 15 categories, with 13 categories best expressing the land-use diversity.

The mapping of the consolidated urban blocks was a semi-manual task of identifying vacant polygons within the proposed urban boundary, developed on the projected blocks of the Metropolitan Development Plan and real color satellite images. The mapping of the peripheries considered a minimum lot size of approximately 100m^2, similar to the 107m^2 proposed by the FMV (2018c). 9123 blocks were identified with a total of 34.22 km^2, which represents 33.12% of all the urban territory considered, which was 103.32 km^2. This block definition allowed a characterized comparison of the areas of the satellite image. The main roads were identified based on existing plans in the repository of the Ministry of Transport and indicated in the Metropolitan Development Plan. Likewise, the slopes processed from the curves defined in the national charter were used, which were expressed in percentages. After this, the information was collected from 228 valid lots of a total of 273 calculated for a simple sample, NC=90%, E=5%, p=50%. The observation points were spread randomly on the plain in an amount proportional to the population density shown in the current plan. The characteristics of the observation points that were considered, are detailed in Table 1 and have been used as characteristics to determine the classification of urban areas through a K-Means algorithm, implemented with the scikit-learn library.

V. RESULTS

The Level 2A product image is shown in Figure 3, captured and processed in false color rgb, using bands 11, 12, and 4, respectively. With this, it is possible to differentiate, in a color between yellow and violet, the possible built

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607



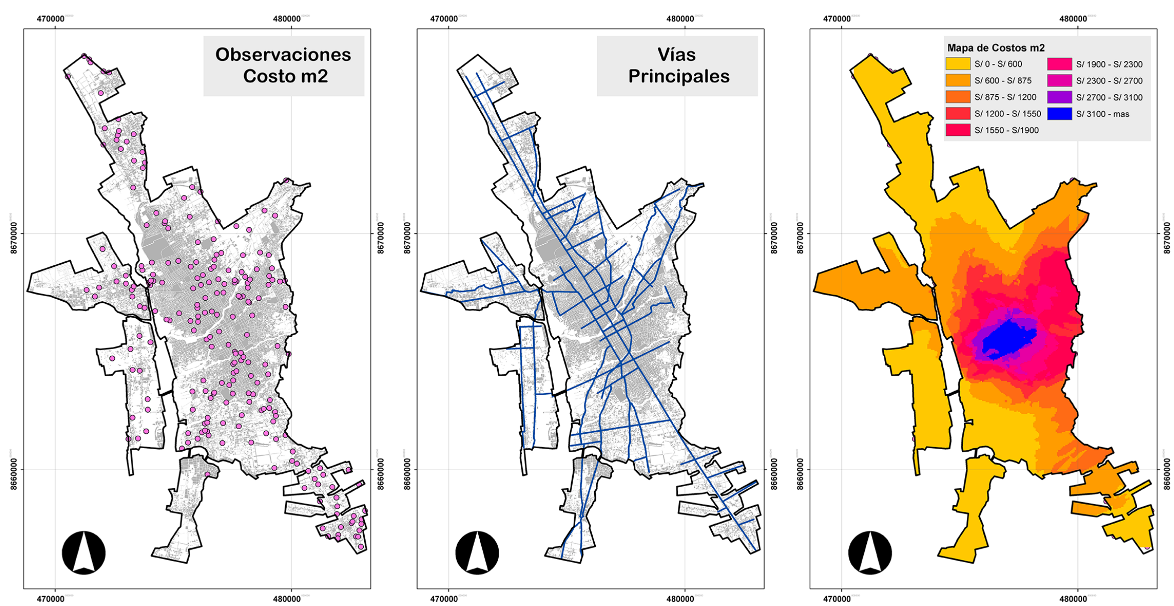
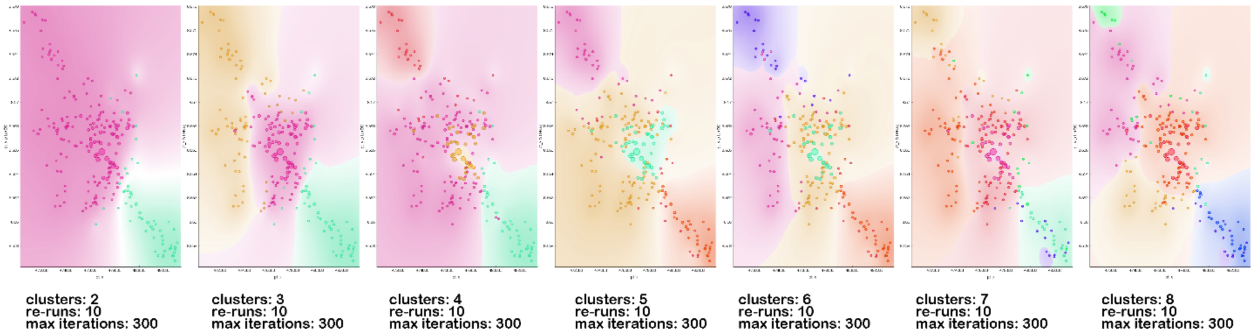


Figure 3. Identification and classification of satellite images using K-Means. Source: Preparation by the Author.
Figure 4. Sampling points, roads, and land value map. Source: Preparation by the Author.

areas and other lands. To generate a scale that can be manually discriminated, the false color was clustered in 13 categories using the K-Means algorithm. Then, each cluster was labeled as built or not built, reducing the results to 2 categories, which are distinguished in yellow and black. As other types of land tend to be confused, just the boundary of urban expansion was considered, improving the accuracy of the result. The occupied urban land percentage of the buffers defined later was calculated using the third image.

The distribution of the sampled points observed is seen in Figure 4, distributed randomly in the occupied area: the main road network indicated in existing plans, both



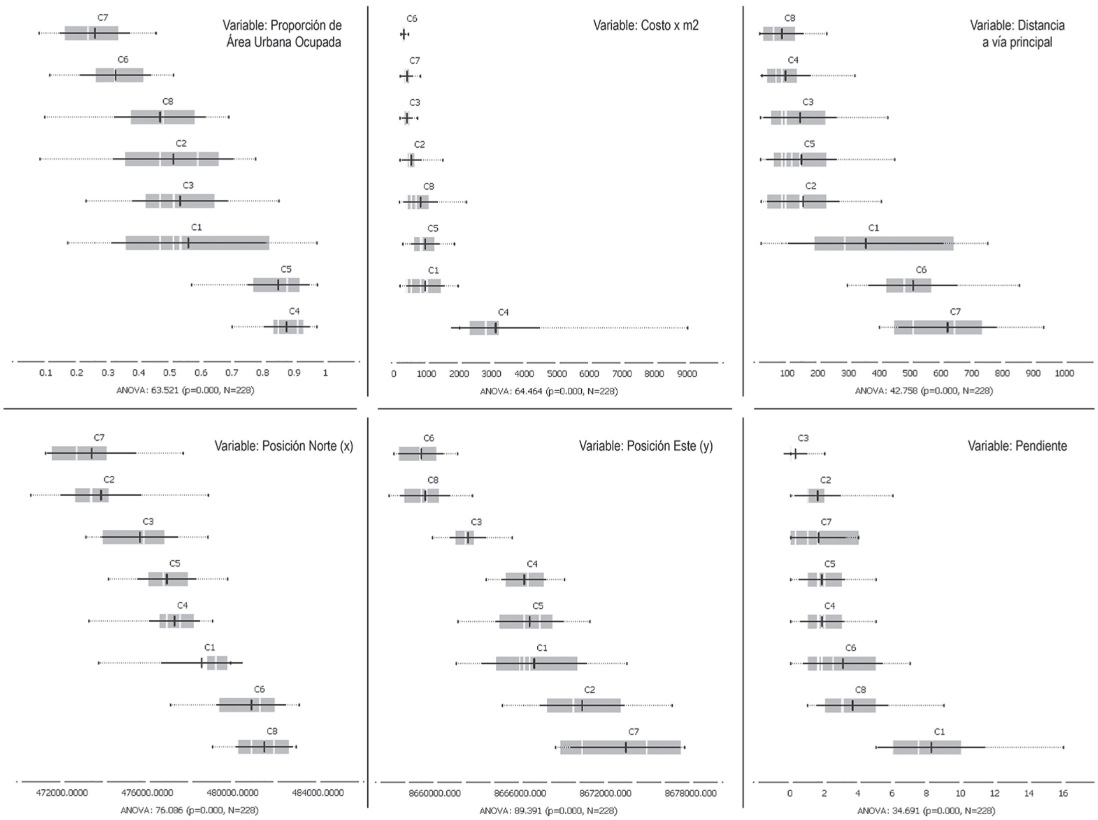


Figure 5. Classification using the unsupervised K-Means algorithm to generate centroids. Source: Preparation by the Author.
Figure 6. Comparative view between clusters for each variable of the model. Source: Preparation by the Author.

overlapped on the map of consolidated urban blocks and those being developed. With the offered land price data per m², a DEM image was projected, using a 9-category Jenks optimization division classification. These tasks were carried out using the QGIS 3.12 software tool.

Using a 400 m diameter buffer, which was used as the center for the observation points, the distance to the closest main road, the cost per m² offered, the slope, the percentage of urban area occupied, and its geographic location reference were averaged. The K-Means algorithm was fed using these data, implemented with the scikit-learn library. Fixed

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO
DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

Variable	Cluster	Statistic	Cluster						
			C2	C3	C4	C5	C6	C7	C8
Percentage of occupied area	C1	D. mean D. media			-0.3166	-0.2898		0.3008	
		P			0.0202	0.0358		0.0355	
	C2	D. mean			-0.3648	-0.338	0.1873	0.2526	
		P			<.0001	<.0001	0.0024	0.0002	
	C3	D. mean			-0.3432	-0.3164	0.2088	0.2742	
		P			<.0001	<.0001	0.0007	<.0001	
	C4	D. mean					0.552	0.6174	0.4076
		P					<.0001	<.0001	<.0001
	C5	D. mean					0.5253	0.5906	0.3808
		P					<.0001	<.0001	<.0001
	C6	D. mean							-0.1444
		P							0.0457
	C7	D. mean							-0.2098
		P							0.0033
Slope	C1	D. mean	6.6591	7.9643	6.4113	6.4246	5.1786	6.6136	4.5833
		P	0.0003	<.0001	0.0004	0.0005	0.0042	0.0002	0.0084
	C2	D. mean		1.3052					-2.0758
		P		<.0001					0.0033
	C3	D. mean			-1.553	-1.5397	-2.7857		-3.381
		P			<.0001	<.0001	0.0134		<.0001
	C4	D. mean							-1.828
		P							0.0163
	C5	D. mean							-1.8413
		P							0.0105
Distance to main road	C2	D. mean					-357.389	-469.3501	
		P					<.0001	<.0001	
	C3	D. mean					-367.0552	-479.0163	
		P					<.0001	<.0001	
	C4	D. mean					-414.7302	-526.6912	
		P					<.0001	<.0001	
	C5	D. mean					-363.2758	-475.2369	
		P					<.0001	<.0001	
	C6	D. mean							427.3883
		P							<.0001
	C7	D. mean							539.3494
		P							<.0001

Cost m2	C1	D. mean			-2159.2823		659.3929		
		P			<.0001		0.0364		
	C2	D. mean			-2587.4413	-426.6631	231.2338		
		P			<.0001	<.0001	0.0003		
	C3	D. mean			-2718.6751	-557.8968			-410.8929
		P			<.0001	<.0001			0.0223
	C4	D. mean				2160.7783	2818.6751	2725.8504	2307.7823
		P				<.0001	<.0001	<.0001	<.0001
	C5	D. mean					657.8968	565.0722	
		P					<.0001	<.0001	
	C6	D. mean						-92.8247	-510.8929
		P						0.6962	0.0024
	C7	D. mean							-418.0682
		P							0.0303

Table 2. Games-Howell Test. Source: Preparation by the Author.

parameters of 10 re-runs and 300 iterations were used for a range of 2 to 8 clusters. These are presented in Figure 5, using the UTM east and north coordinates as x and y, respectively.

To validate the differences between the resulting clusters, ANOVA tests were run, all of which were significant with a value of $p<0.001$. The differences in the distribution of the values are shown in Figure 6, through box charts, with a reference to the F statistic of each test. Meanwhile, the significant differences between groups, made with the Games-Howell post-hoc test, are illustrated in Table 2.

Differences between clusters (C) by variable are identified in Figure 6. The geographic location is significantly different for all clusters. The proportion of occupied urban area, for C7 and C6 is less than 0.5, and for C4 and C5 is above 0.7. The cost per m^2 for C4 is highly variable and greater than the other clusters. This is followed by C1, C2, and C8. The distance to the closest main road gives a range below 200 m for C8 and C4; less than 300 m for C3, C5, and C2; between 400 and 800 m for C6 and C7; and of 100m to 800m for C1. A slope above 5% is seen in C1, and less than 5% in the other clusters.

Table 2 allows identifying significant differences between paired clusters. The proportion of occupied urban area is significantly different between C1 and C4, C5 and C7; between C2 and the interval that runs from C4 to C7; between C3 and the interval that runs from C4 to C7; between C4 and C6, C7 and C8; between C5 and C6, C7 and C8; between C6 and C8; and between C7 and C8. Meanwhile, the slope is significantly different between C1 and the interval from C2 to C8; between C2 and C3 and C8; between C3 and C4, C5, C6 and C8; between C4 and C8; and between C5 and C8. The distance to the main road is significantly different between C2 and C6 and C7; between C4 and C6 and C7; between C5 and C6 and C7; between C6 and C8; and between C7 and C8. Finally, the cost per m^2 is significantly different between C1 and C4 and C6; between C2 and the interval from C4 to C6; between C3 and C4, C5 and C8; between C4 and the interval from C5 to C8; between C5 and C6 and C7; between C1 and C7 and C8; and between C7 and C8. Overall, significant differences were identified in all the variables.

Once the significant differences between adjoining clusters were validated, the urban polygons overlapped with the centroids, and the areas presented in Figure 5 were marked out. The resulting map is presented in Figure 7 at a block level, distinguishing the consolidated ones and those that are being developed.

VI. DISCUSSIONS

The research used a model that prioritizes the percentage of area occupied by buildings, in contrast to Liu *et al.*, (2018) and Steurer and Bayr (2020), who use population growth based on a close density. In all the cases, the K-Means algorithm made multidimensional

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO
DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

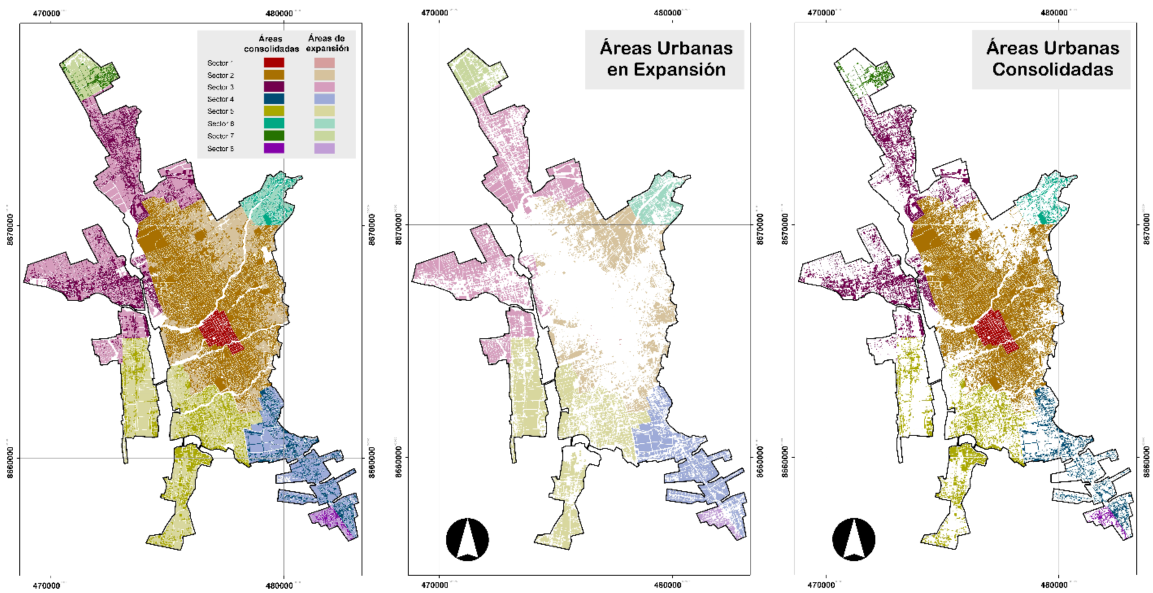


Figure 7. Characterization polygons of the resulting urban areas. Source: Preparation by the Author.

classification possible. In this sense, Steurer and Bayr (2020) propose means that can be complemented with the results for future research. Regarding the fit of image-based data sources, noise was found in the urban land classification. However, its reduction did not follow the parameters of Zhou *et al.* (2017), but rather the classified layers were reduced until obtaining an image with data of 2 values, which represent the occupied land.

Considering that the spatial behavior of urban phenomena is complex and uncertain (Pickard & Meentemeyer, 2019), changes are required to suitably study them. For this reason, the urban boundaries that had been defined by MPH (2016) had to be adjusted manually to be able to address peri-urban expansion areas and annex adjoining districts on the left bank of the Mantaro River. After generating the cluster classification, the marked-out polygons of the MPH proposal (2016) maintain a variation contrast that could be considered in future urban plans. In this aspect, it must be remembered that the complex reality demands flexibility when it comes to setting urban boundaries, and not just their political and administrative consideration (Steurer & Bayr, 2020).

The development of the urban sprawl in Huancayo is diffuse and low density: the differences between clusters identified with the Games-Howell test reveal that the areas around the center, C1 and C2, have a higher proportion which is different to all the other 6 clusters and that this difference extends to the price, which is extremely high in C1 and C2, and less in C3, but more similar for the rest. This behavior is similar to the occupation pattern where a constant lower price and the disproportionate valuation of it in the areas near consolidated urban areas is sought (Baer & Kauw, 2016; Gasic, 2018). The form of growth in Huancayo seems to go against the ideal vision of a chronological development proposed by Graells-Garrido *et al*. (2021) and Moreno *et al*. (2021). The growing distance of the centralized urban services and deficient transport must be priority issues (Vorontsova *et al*., 2016).

Just as Araque Solano and Caballero Quintero (2009) state, the prices in informal markets present their formalization under similar conditions as consolidated sectors. In the city of Huancayo, this variation is distinguished in the adjoining areas of C2 and C3 that close a consolidation process at a higher price. The fractioning identified could be linked to the limited participation of the public sector in market control (López-Navarrete & Peña-Medina, 2017). This rapid increase in the periphery price that follows a capital increase is a common situation in other scenarios like those analyzed by Amézquita, Rodríguez, and Murillo (2015), Gaete (2021), or Gasic (2018). Apart from this, the regulatory effects and specific or individual economic interventions like State policy have to be added (Li *et al*., 2019; Garza Puentes & Tovar Vanegas, 2009).

VII. CONCLUSIONS

This research identifies that the K-Means algorithm provides a viable way of classifying urban land using heterogeneous variables and that the difference between the generated clusters, can be tested as multivariate and differentiated through open data. From this classification, spatial fractioning is identified in the city of Huancayo which is mainly determined by the variables, proportion of occupied land, offered price, and distance to the main road system.

Although unassisted characterized is subject to opinion, the situation that urban sprawl in Huancayo has been experiencing must be highlighted, outlining alternatives for a more objective analysis of its occupation, taking advantage of the available means of analysis that there are.

VIII. BIBLIOGRAPHICAL REFERENCES

Aceid y Fundación ACS (2018). *Encuentro Iberoamericano sobre Prácticas Urbanas*. Recuperado de https://www.aecid.ht/es/convocatorias/encuentro-iberoamericano-sobre-practicas-urbanas-innovadoras-hacia-ciudad-2030

Alfasi, N. y Migdalovich, E. (2020). Losing faith in planning. *Land Use Policy, 97*. DOI: https://doi.org/10.1016/j.landusepol.2020.104790

Amézquita, L., Rodríguez, L. y Murillo, H. (2015). Los precios del suelo en Bogotá. El barrio Veinte de Julio. *Bitácora Urbano Territorial, 25*(1), 43-50. DOI: https://dx.doi.org/10.15446/bitacora.v1n25.40236

Araque Solano, A. y Caballero Quintero, Y. (2009). La encrucijada de la Vivienda de Interés Social en Bogotá: Los precios del suelo. *Civilizar Ciencias Sociales y Humanas, 9*(16), 127-152. Recuperado de http://www.scielo.org.co/ scielo.php?script=sci_arttextypid=S1657-89532009000100009ylng=enytlng=es

Baer, L. y Kauw, M. (2016). Mercado inmobiliario y acceso a la vivienda formal en la Ciudad de Buenos Aires, y su contexto metropolitano, entre 2003 y 2013. *Revista EURE - Revista de Estudios Urbano Regionales, 42*(126). Recuperado de https://www.eure.cl/index.php/eure/article/view/1676

Belhadia, A., Djenourib, Y., Norvagc, K., Ramampiaroc, H., Masségliad, F. y Lin, J. (2020). Space–time series clustering: Algorithms, taxonomy and case study on urbansmart cities. Engineering Applications of Artificial Intelligence, *95*. DOI: https://doi.org/10.1016/j.engappai.2020.103857

Calderón, J. (2015). Programas de vivienda social nueva y mercados de suelo urbano en el Perú. *Revista EURE - Revista de Estudios Urbano Regionales, 41*(122). Recuperado de https://www.eure.cl/index.php/eure/article/view/654

Campesato, O. (2020). *Artificial Intelligence Machine Learning and Deep Learning*. New Delhi: Mercury Learning and Information.

Cardó, P. (2017). Estanflación y crecimiento del área urbana en 2012. Oferta del suelo en el Gran San Juan. *Bitácora Urbano Territorial, 27*(1), 27-34. DOI: https://dx.doi.org/10.15446/bitacora.v27n1.40348

Carvajal Mascaró, F., Moreira Muñoz, A., Salazar Burrows, A., Leguia Cruz, M. y Jorquera-Guajardo, F. (2019). Divergencias y contradicciones en la planificación sustentable del periurbano rural metropolitano de Valparaíso. Caso Reserva de la Biosfera La Campana-Peñuelas, Chile central. *Urbano, 22*(39), 64–87. DOI: https://doi.org/10.22320/07183607.2019.22.39.04

Castillo-García R. (2021). Evolución de la Planificación Urbana en el Perú 1946 - 2021: De la Planificación Urbana Normativa a la Planificación del Desarrollo Urbano Sostenible. *PAIDEIA XXI, 11*(1), 31-79. DOI: https://doi.org/10.31381/paideia%20xxi.v11i1.3783

Delgadillo, V. (2016). Ciudades iletradas: orden urbano y asentamientos populares irregulares en la ciudad de México. *Territorios, 35*, 81-99. DOI: https://dx.doi.org/10.12804/territ35.2016.04

Durduran, S. (2015). Automatic classification of high-resolution land cover using a new data weighting procedure: The combination of K-Means clustering algorithm and central tendency measures (KMC-CTM). *Applied Soft Computing Journal, 35*, 136–150. DOI: https://doi.org/10.1016/j.asoc.2015.06.025

Espinoza, A. y Fort, R. (2017). *Mapeo y Tipología de la Expansión Urbana en el Perú*. Lima: ADI Perú.

Feng, Z., Peng, J. y Wu, J. (2020). Using DMSP/OLS nighttime light data and K–means method to identify urban–rural fringe of megacities. *Habitat International, 103*. DOI: https://doi.org/10.1016/j.habitatint.2020.102227

Foldvary, F. y Minola, L. (2017). The taxation of land value as the means towards optimal urban development and the extirpation of excessive economic inequality. *Land Use Policy, 69*, 331-337. DOI: http://dx.doi.org/10.1016/j.landusepol.2017.09.022

Fondo Mivivienda [FMV] (2018a). Miles se benefician con subsidios. *La revista immobiliaria del Perú: MIVIVIENDA, 126*(14), 4-5. Lima: Fondo MIVIVIENDA.

Fondo Mivivienda [FMV] (2018b). Mejorarán planes de desarrollo urbano. *La revista immobiliaria del Perú: MIVIVIENDA, 126*(14), 6-7. Lima: Fondo MIVIVIENDA.

Fondo Mivivienda [FMV] (2018c). Tienen capacidad de compra. *La revista immobiliaria del Perú: MIVIVIENDA, 127*(14), 6-9. Lima: Fondo MIVIVIENDA.

Forestier, G. y Wemmer, C. (2016). Semi-supervised learning using multiple clusterings with limited labeled data. *Information Sciences*, (361-362), 48-65. DOI: https://doi.org/10.1016/j.ins.2016.04.040

Gaete, H. (2021). Tendencias del mercado de suelo urbano en periodo largo. Concepción, Chile. 1989-2018. *ACE: architecture, city and environment, 16*(46), 9946. DOI: http://dx.doi.org/10.5821/ace.16.46.9946

García, F. y Peralta, M. (2016). Las urbanizaciones multifamiliares cerradas y su entorno urbano: una nueva geografía simbólica en la ciudad de Cali (Colombia). *Revista EURE - Revista de Estudios Urbano Regionales, 42*(126). Recuperado de https://www.eure.cl/index.php/eure/article/view/1548/880

Garza Puentes, N. y Tovar Vanegas, R. (2009). El mercado de vivienda en Barranquilla y el sector externo de la economía. *Revista de Economía del Caribe*, (4), 181-209.

Gasic, I. (2018). Inversiones e intermediaciones financieras en el mercado del suelo urbano. Principales hallazgos a partir del estudio de transacciones de terrenos en Santiago de Chile, 2010-2015. *Revista EURE - Revista de Estudios Urbano Regionales, 44*(133). Recuperado de https://www.eure.cl/index.php/eure/article/view/2403/1108

Gasparenienea L., Venclauskienea, D. y Remeikiene, R. (2014). Critical review of selected housing market models concerning the factors that make influence on housing price level formation in the countries with transition economy. *Procedia - Social and Behavioral Sciences 110*, 419 – 427. DOI: https://doi.org/10.1016/j.sbspro.2013.12.886

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

Gedal, M. y Ellen, I. (2018). Valuing urban land: Comparing the use of teardown and vacant land sales. *Regional Science and Urban Economics, 70*, 190–203. DOI: https://doi.org/10.1016/j.regsciurbeco.2018.03.006

Glaeser, E., Gyourko, J. (2003). The impact of building restrictions on housing affordability. *Federal Reserve Bank of New York Economic Policy Review 9*(2), 21–39.

Glaeser, L. y Ward, A. (2009). The causes and consequences of land use regulation: Evidence from Greater Boston. *Journal of Urban Economics, 65*(3), 265-278. DOI: https://doi.org/10.1016/j.jue.2008.06.003

GRADE (2020). *Hacia una nueva política de vivienda en el Perú: Problemas y posibilidades*. Lima: Grupo de Análisis para el Desarrollo GRADE.

Graells-Garrido, E., Serra Burriel, F., Rowe, F., Cucchietti, F. y Reyes, P. (2021). A city of cities: Measuring how 15-minutes urban accessibility shapes human mobility in Barcelona. *PLOS ONE 16*(5). DOI: https://doi.org/10.1371/journal.pone.0250080

Hindi, T. de M. C., Moreira, T. y Rossi, A. (2020). Fornas alternativas de propriedade: cooperativismo, produção capitalista do espaço urbano e mercado imobiliário. A: Seminario Internacional de Investigación en Urbanismo. XII Seminario Internacional de Investigación en Urbanismo, São Paulo-Lisboa, 2020. São Paulo: Faculdade de Arquitetura da Universidade de Lisboa. DOI: https://doi.org/10.5821/siiu.10040.

Hwang, S., Park, M. y Lee, H. (2013). Dynamic analysis of the effects of mortgage-lending policies in a real estate market. *Mathematical and Computer Modelling*, *57*(9-10), 2106–2120. DOI: https://doi.org/10.1016/j.mcm.2011.06.023

Klaufus, C., Van Lindert, P., Van Noorloos, F. y Steel, G. (2017). All-Inclusiveness versus Exclusion: Urban Project Development in Latin America and Africa. *Sustainability*, *9*(11). DOI: https://doi.org/10.3390/su9112038

Kok, N., Monkkonen, P. y Quigley, J. (2014). Land use regulations and the value of land and housing: An intra-metropolitan analysis. *Journal of Urban Economics*, *81*, 136-148. DOI: https://doi.org/10.1016/j.jue.2014.03.004

Joshi, A. (2020). *Machine Learning and Artificial Intelligence*. Springer.

Li, X., Sun, M. y Boersma, K. (2019). Policy Spillover and Regional Linkage Characteristics of the Real Estate Market in China's Urban Agglomerations. *Journal of Management Science and Engineering*, *4*(3), 189-210. DOI: https://doi.org/10.1016/j.jmse.2019.05.004

Liu, J., Jiao, L., Zhang, B., Xu, G., Yang, L., Dong, T., Xu, Z., Zhong, J. y Zhou, Z. (2021). New indices to capture the evolution characteristics of urban expansion structure and form. *Ecological Indicators, 122*(1). DOI: https://doi.org/10.1016/j.ecolind.2020.107302

Liu, L., Peng, Z., Wu, H., Jiao, H., Yu, Y. y Zhao, J. (2018). Fast identification of urban sprawl based on K-Means clustering with population density and local spatial entropy. *Sustainability (Switzerland), 10*(8). DOI: https://doi.org/10.3390/su10082683

López Navarrete, J. y Peña Medina, S. (2017). La segregación socioespacial en Ciudad Juárez, Chihuahua, 1990-2010. *Región y sociedad, 29*(68), 115-152. DOI: https://dx.doi.org/10.22198/rys.2017.68.a210

Migueltorena, A. y Lan, D. (2013). Racionalidades y contrarracionalidades, a partir de la vivienda, en la producción del espacio urbano de Tandil, Argentina. Cuadernos de Geografía: *Revista Colombiana de Geografía, 22*(1), 109-125. Recuperado de https://www.redalyc.org/articulo.oa?id=281825518011

Ministerio de Vivienda Construcción y Saneamiento [MVCS] (2021). *Resumen de la Política Nacional de Vivienda y Urbanismo.* Lima: MVCS. Recuperado de https://www.gob.pe/institucion/vivienda/informes-publicaciones/2027198-resumen-de-la-politica-nacional-de-vivienda-y-urbanismo

Moreno, C., Allam, Z., Chabaud, D., Gall, C. y Pratlong, F. (2021). Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities, 4(*1), 93–111. DOI: http://dx.doi.org/10.3390/smartcities4010006

Mouratidis, K. (2021). Urban planning and quality of life: A review of pathways linking the built environment to subjective well-being. Cities, 115, 103229. https://doi.org/10.1016/j.cities.2021.103229

Municipalidad Provincial de Huancayo (2016). *Plan de Desarrollo Metropolitano – Provincia de Huancayo 2015 – 2035.* Huancayo: Municipalidad Provincial de Huancayo.

Nabil, N. y Eldayem, G. (2015). Influence of mixed land-use on realizing the social capital. *HBRC Journal*, *11*(2), 285-298. DOI: https://doi.org/10.1016/j.hbrcj.2014.03.009

Naciones Unidas (2018). *Plan de Acción Regional para la implementación de la Nueva Agenda Urbana en América Latina y el Caribe 2016-2036*. Santiago: Naciones Unidas.

Parias, A. (2008). El mercado de arrendamiento en los barrios informales en Bogotá, un mercado estructural. *Territorios*, (18-19), 75-101. Recuperado de https://revistas.urosario.edu.co/index.php/territorios/article/view/828

Peña-Zamalloa, G. (2018). Proyección del cambio de uso de suelo urbano mediante técnicas de microsimulación, bajo un escenario de escasez de datos en el sector de San Carlos, Huancayo, Perú 2018-2028. *Espacio y Desarrollo, 32*, 99-124. DOI: https://doi.org/10.18800/espacioydesarrollo.201802.005

Pickard, B. y Meentemeyer, R. (2019). Validating land change models based on configuration disagreement. Computers, *Environment and Urban Systems, 77.* DOI: https://doi.org/10.1016/j.compenvurbsys.2019.101366

Pimentel Sánchez, N. (2020). ¿Tomar lotes para vivir o para vender? Tráfico de tierras y práctica clientelar en la periferia urbana. *Revista de Sociología*, (31), 133–159. DOI: https://doi.org/10.15381/rsoc.v0i31.19279

Sabatini, F. y Arenas, F. (2000). Entre el Estado y el mercado: resonancias geográficas y sustentabilidad social en Santiago de Chile. *Revista EURE - Revista de Estudios Urbano Regionales, 26*(79). Recuperado de https://www.eure.cl/index.php/eure/article/view/1212

Salazar, C. (2014). "El puño invisible" de la privatización. *Territorios, 30*, 69-90. DOI: https://dx.doi.org/10.12804/territ30.2014.03

Saleh, A., Hwa, T. y Majid, R. (2016). Housing Mismatch Model in Suburban Areas. Procedia - *Social and Behavioral Sciences*, *234*, 442-451. DOI: https://doi.org/10.1016/j.sbspro.2016.10.262

Scotiabank (2015). *El mercado inmobiliario en perspectiva*. Lima: Estudios Económicos Scotiabank.

Steurer, M. y Bayr, C. (2020). Measuring urban sprawl using land use data. *Land Use Policy*, *97*, DOI: https://doi.org/10.1016/j.landusepol.2020.104799

Tellier, L. (2020). Characterizing urban form by means of the Urban Metric System. *Land Use Policy.* DOI: https://doi.org/10.1016/j.landusepol.2020.104672

Tricaricoa, L. y de Vidovich, L. (2021). Proximity and post-COVID-19 urban development: Reflections from Milan, Italy. *Journal of Urban Management*, *10*(3), 302-310. DOI: https://doi.org/10.1016/j.jum.2021.03.005

CLASIFICACIÓN ESPACIAL DEL SUELO URBANO POR EL VALOR ESPECULATIVO DEL SUELO E IMÁGENES MSI SATELITALES USANDO K-MEANS, HUANCAYO, PERÚ
GONZALO RODOLFO PEÑA ZAMALLOA
REVISTA URBANO Nº 44 / NOVIEMBRE 2021-ABRIL 2022
PÁG. 70 - 83
ISSN 0717 - 3997 / 0718 - 3607

Vorontsova, A., Vorontsova, V. y Salimgareev, D. (2016). The development of Urban Areas and Spaces with the Mixed Functional Use. *Procedia Engineering, 150*. DOI: https://dx.doi.org/10.1016/j.proeng.2016.07.277

Yu, J., Zhou, K. y Yang, S. (2019). Land use efficiency and influencing factors of urban agglomerations in China. *Land Use Policy, 88*. DOI: https://doi.org/10.1016/j.landusepol.2019.104143

Zhou, X., Gu, J., Shen, S., Ma, H., Miao, F., Zhang, H. y Gong, H. (2017). An automatic K-Means clustering algorithm of GPS data combining a novel niche genetic algorithm with noise and density. ISPRS International *Journal of Geo-Information, 6*(12), 392. DOI: https://doi.org/10.3390/ijgi6120392