Assessment of social vulnerability to groundwater pollution using K-means cluster analysis.

Marisela Uzcategui-Salazar (mariselauzcateguis@gmail.com)
Rey Juan Carlos University: Universidad Rey Juan Carlos https://orcid.org/0000-0002-2894-5925
Javier Lillo
Universidad Rey Juan Carlos

Research Article

Keywords: Social vulnerability, groundwater pollution, clustering analysis, K-means

Posted Date: April 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1525225/v1

Assessment of social vulnerability to groundwater pollution using K-means cluster analysis.

Marisela Uzcategui-Salazar[a,b]*, Javier Lillo [c,d]

[a] International Doctoral School, University of Rey Juan Carlos, 29833 Móstoles, Madrid, Spain.
[b] TERRA Research Group, Geological Engineering School. University of Los Andes. 5101 Mérida, Venezuela.
[c] Global Earth Change and Environmental Geology Research Group, Department of Biology, Geology, Physics and Inorganic Chemistry. University of Rey Juan Carlos, 29833 Móstoles, Madrid, Spain.
[d] IMDEA Water Institute, Av. Punto Com, 2, 28805 Alcalá de Henares, Madrid, Spain

*Corresponding author.

E-mail: mariselauzcateguis@gmail.com
University of Rey Juan Carlos, Departmental II, office 256. Tulipán Street, s/n, 29833 Móstoles, Madrid, Spain

Abstract

It is possible to assess the harm that society suffers from an anthropogenic event based on the impact of groundwater pollution on society. Indexing methodologies are commonly applied to assess the social vulnerability of aquifers. However, they assign weighting and rating values for the different factors involved, make them very subjective. This research proposes to assess the social vulnerability to groundwater pollution by considering three factors: the uses of groundwater resources, the exposed population, and the socio-economic losses. In order to eliminate the subjectivity of indexing methodologies, a K-means cluster analysis was used to assess the social vulnerability. Using this method, a social vulnerability map can be produced with greater objectivity. The proposed methodology was applied to an aquifer located in central Spain, an area with significant agricultural development. Low population density and unproductive zones result in low social vulnerability in most of the area. Nevertheless, high social vulnerability is observed in the southern regions due to agricultural development which leads to higher socio-economic variables and demand for groundwater resources for agricultural activities. Similarly, high social vulnerability is observed in the northeast, mainly influenced by groundwater use and exposed population. These results show that social vulnerability in most of the study area is not very significant for assessing the risk of groundwater contamination, because the damage to the social, environmental or economic sector is low. However, in the south and northeast in the study area, pesticides and fertilisers should be used with caution, as they significantly increase the risk of groundwater contamination and thus the impact on society. The K-means clustering

method proved to be an objective and effective option for assessing social vulnerability to groundwater pollution in aquifers.

Key words: Social vulnerability, groundwater pollution, clustering analysis, K-means.

1. Introduction

Aquifers represent the most important source of water supply for urban, industrial and agricultural uses in areas where surface water resources are scarce or the use of supply sources is limited by water quality. In general, groundwater is of better quality than surface water because it has a natural protection against anthropogenic pollutants. However, natural or geogenic contamination by dissolution or chemical reactions between water and solid matrix (rock, soil) is also important. In some cases, the use of contaminated groundwater negatively affects society, endangering human health, the environment or the economic development of a region (Cutter, 2010, 1996; Grondona et al., 2015; Perles et al., 2008)

Social vulnerability is a simply way of assessing the potential damage to society from a natural or anthropogenic event (Cutter 1996; Perles et al. 2008). In environmental studies, many authors (e.g. Ducci, 1999) consider that groundwater represent a valuable resource. Thus, the value of water supply resource has to be taken into account. The socio-economic value associated with groundwater supply uses incorporates variables such as population, number of employees and economic productivity linked to activities that depend on groundwater resource. (Ducci 1999; Vias 2005; Perles et al. 2008; French et al. 2017; Orellana-Macías and Perles Roselló 2022).

To assess the social vulnerability, some authors (Ducci 1999; Vias 2005; Perles et al. 2008; Grondona et al. 2015; Orellana-Macías and Perles Roselló 2022) have developed indexing methodologies that incorporate the factors noted above by assigning weighting and rating values that describe the degree of vulnerability of society to groundwater contamination. However, the subjectivity involved in the selection of relative weighting and rating values is a disadvantage in the application of these methodologies.

The main goal of this work is to develop a new methodology for the assessment of social vulnerability to anthropogenic groundwater pollution, through the application of K-means clustering techniques. The methodology will be applied to a case study related to a detrital aquifer located in the region of Madrid (central Spain).

2. Materials and methods

2.1. The study area

The study area is located in the southeast of Community of Madrid in central Spain and covers an approximate surface area of 133Km^2 (Fig. 1). The climate is temperate-continental Mediterranean with an average rainfall of 440 mm/year (Mostaza-Colado et al. 2018). The Jarama River is the main surface water resource, crossing the study area from north to south. In the south, an major water canal was constructed to canal the Jarama River and provide water for irrigation using the flooding technique (Mostaza 2019).

In the study area there is a detritic aquifer formed by three groundwater bodies, according to the definition of the Confederación Hidrográfica del Tajo (CHET): "Aluvial del Jarama: Guadalajara Madrid", "Guadalajara" and "Aluviales Jarama-Tajuña". The latter is the most important because it covers more than 80% of the area (Fig. 1). The aquifer is shallow and consists of gravels and sands, with intercalations of clays and silts. Its average thickness is about 10m (Carreño Conde et al., 2014). The region has an important agricultural (mainly arable and tree crops) and livestock development (Fernández González 2013; MAPA 2021a). The population is scarce because most of the territory is used for agricultural activities (Comunidad de Madrid, 2020).

The aquifer provides water for some agricultural, urban and industrial activities in the study area (CHT 2015a). Although it is not the only source of water, it is an important and valuable water resource in the region.

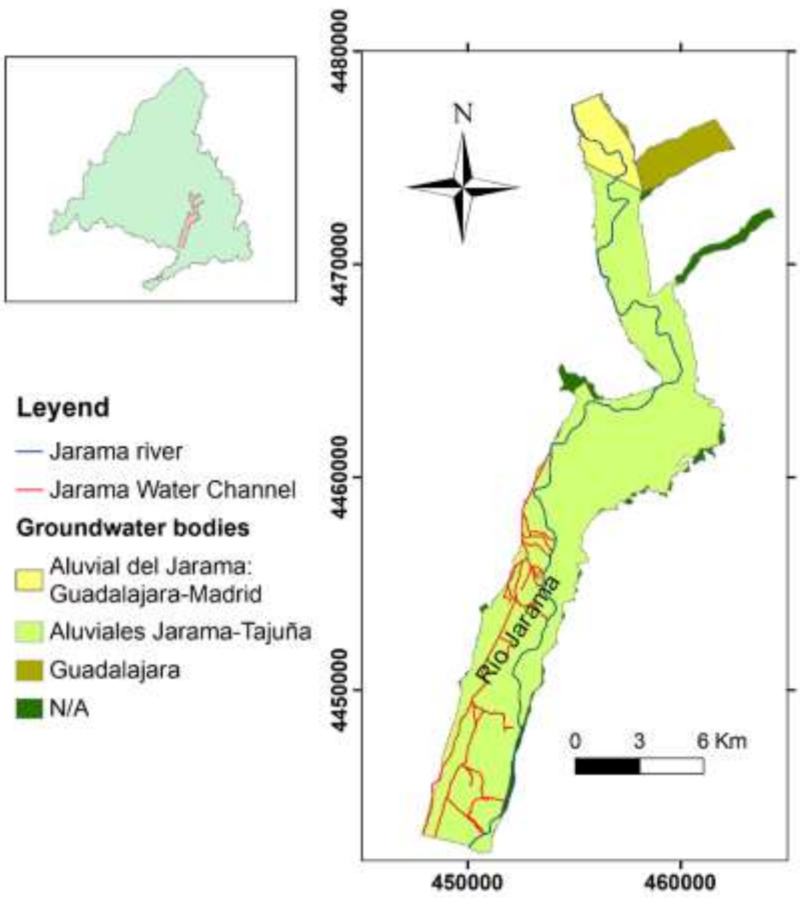


Fig. 1 Study area showing the groundwater bodies defined by the Confederación Hidrográfica de El Tajo (CHET), 2010). ETRS 1989 UTM Zone30N

2.2. Data set collection

The data considered in this work included:

- Cartographical information: Municipal boundaries map (scale 1:25.000, (IGN 2021)); digital map of groundwater bodies (CHET, 2010); land use map (scale 1:100,000 Corine land cover map (IGN 2018)).
- Groundwater resource information: Groundwater uses (CHT 2015a)
- Population information: Population density (Comunidad de Madrid 2020). Employment by economic activity in the region of Madrid (Consejería de Economía, Hacienda y Empleo, 2021). Wages in agricultural activity in Spain (MAPA 2021b).
- Agricultural information: Agricultural regions (Fernández González 2013). Land prices (Subdirección General de Análisis, Coordinación y Estadística. 2020). Cereal yields and prices

in Spain (2020-2021) (MAPA 2021c, d). Yields and prices of woody crops in Spain (2020-2021) (MAPA 2020, 2021d; Subsecretaría de Agricultura, Pesca y Alimentación 2020). Livestock per groundwater bodies, yields and prices in Spain (CHT 2015a; MAPA 2021e, a). Agricultural demand units (UDA) from the Hydrological Plan of the Tajo basin for 2015-2021, (CHT 2015b).

The data were stored as a geographical database in ArcGIS v10.4.1. The whole study area (133 Km^2) was divided into 5842 pixels with a cell size of 150mx150m, in order to obtain a large data set to evaluate the different variables at each point.

2.3. Estimation of factors to determine the social vulnerability

The social vulnerability assessment was carried out in the following two stages:

- *Estimation of factors affecting the social vulnerability*. Three different factors were considered to assess the social vulnerability to a groundwater contamination event. These factors, which include social, economic and environmental aspects, were evaluated by considering the following settings:
  - ✓ Vulnerability of groundwater resources (V_{GR})
  - ✓ Vulnerability of exposed population (V_P).
  - ✓ Socio-economic vulnerability ($V_{S\text{-}E}$)
- *Mapping of social vulnerability using cluster analysis (K-means algorithm)*

A normalization of the obtained factor values was performed to standardize the ranges of the values in order to avoid the bias of higher values over lower values (0-1) (Eq. 1, (Salazar and Del Castillo 2018)).

$$Normalized\ Factor\ value = \frac{(Fx - Fmin)}{(Fmax - Fmin)} \quad (1)$$

Where *Fx* is the value of the factor in the *x* point, and *Fmin*, *Fmax* are the minimum and maximum values of the range, respectively.

2.3.1. Vulnerability of groundwater resources (V_{GR})

This factor represents the amount of the groundwater resources that can be affected from a contamination, as it reduces the groundwater available for different uses. Groundwater contamination also has a negative impact because the contaminated water can reach other water bodies, affecting associated ecosystems. In the study area, the Jarama river and the aquifer have a hydraulic connection that incorporates water from the aquifer to the river (Mostaza 2019). To obtain the vulnerability of groundwater resources according to groundwater uses, three variables were used: urban uses (U_u), agricultural uses (A_u) and industrial uses (I_u). The amount of water abstracted for the different uses is estimated by Confederación Hidrográfica del Tajo (CHET) to each groundwater body (CHT 2015a).

The vulnerability of groundwater resources was calculated using the following equation (Eq. 2):

$$V_{GR} = U_u + A_u + I_u \quad (2)$$

Where V_{GR} is the vulnerability of groundwater resources, U_u is the urban uses map of groundwater, A_u is the agricultural uses map of groundwater and I_u is the industrial uses map of groundwater.

The ArcGIS v10.4.1 Raster calculator tool was used to sum maps.

2.3.2. Vulnerability of exposed population (V_p)

The exposed population was calculated from the population density located in the study area (Comunidad de Madrid 2020) and the percentage of urban groundwater use (CHT 2015a). The population affected by the consumption of polluted groundwater was calculated as the number of inhabitants per square kilometre multiplied by the percentage of urban groundwater use (Eq. 3).

$$V_P = Population\ density * \%\ of\ groundwater\ \ urban\ use \quad (3)$$

Urban groundwater use was calculated as the percentage of total groundwater use in each groundwater body (Table 1).

The population density map was obtained using the number of inhabitants per square kilometre within each municipality in the study area (Comunidad de Madrid 2020)

The ArcGIS v10.4.1 Raster calculator tool was used to multiply maps.

Table 1 Percentage of urban groundwater use from different groundwater bodies in the study area

Groundwater body	Code (CHET)	Urban use (hm3/year)	All uses. Total groundwater (hm3/year)	Percentage of urban groundwater use from total uses (%)
Guadalajara	030.006	6.311	23.018	16.1
Aluviales Jarama-Tajuña	030.007	0.895	14.199	2.3
Aluvial del Jarama: Guadalajara-Madrid	030.024	0.138	2.016	0.3
Total groundwater		7.344	**39.233**	18.7
% of groundwater use		18.7	100	

2.3.3. Socio-economic vulnerability (*V_{S-E}*)

To assess this factor, social and economic activities that depend on the groundwater resources were considered. The study area has an important agricultural development which predominates over other economic activities (Mostaza-Colado et al. 2018; Mostaza 2019). For this reason, four variables associated with this productive sector were chosen to evaluate this factor:

- Land prices, according to the type of crops and irrigation uses.
- Agricultural production (crops)
- Livestock production
- Employment related to the agricultural activities

Socio-economic vulnerability (V_{S-E}) was calculated using the following equation (Eq. 4)

$$V_{S-E}\left(\frac{€}{year}\right) = Land\ prices + crops\ production + livestock\ production + agricultural\ employment \quad (4)$$

However, irrigation facilities using surfaces water (a water canal from the Jarama river canal) reduce the use of groundwater for irrigation, which in turn contributes to reduce the socio-economic vulnerability due to groundwater contamination. Thus, a reduction factor (*Rf*) can be considered according to the percentage of surface water irrigated areas, as shown in Table 2.

Table 2 Reduction factor values (*Rf*)

Percentage of surface water irrigated areas	Rf
0	1
10	0.9
20	0.8
30	0.7
40	0.6
50	0.5
60	0.4
70	0.3
80	0.2
90	0.1
100	0

Some municipalities into the study area use the Jarama river water canal to irrigate part of the cultivated areas, due to the facility offered by this canal. The percentage of surface area irrigated with water from the Jarama River canal is shown in Table 3.

Table 3 Surface (%) in the study area irrigated with water from the Jarama Canal

Municipality	Surface irrigated with water from the Jarama Canal (%)
Arganda del Rey, Coslada, Chichón, Loeches, Madrid, Mejorada del Campo, San Fernando de Henares, Titulcia, Torrejón de Ardoz, Valdemoro, Velilla de San Antonio	0
Rivas-Vaciamadrid	0.08
Aranjuez	2.79
San Martín de la Vega	14.7
Ciempozuelos	22.9

(Comunidad de Regantes de la Real Acequia del Jarama 2021)

According to the percentage of surface area irrigated with water from the Jarama Canal in each municipality in the study area, a reduction factor (Rf) was assigned according to Table 2.

The spatial distribution of the reduction factor in the study area is shown in Fig. 2.

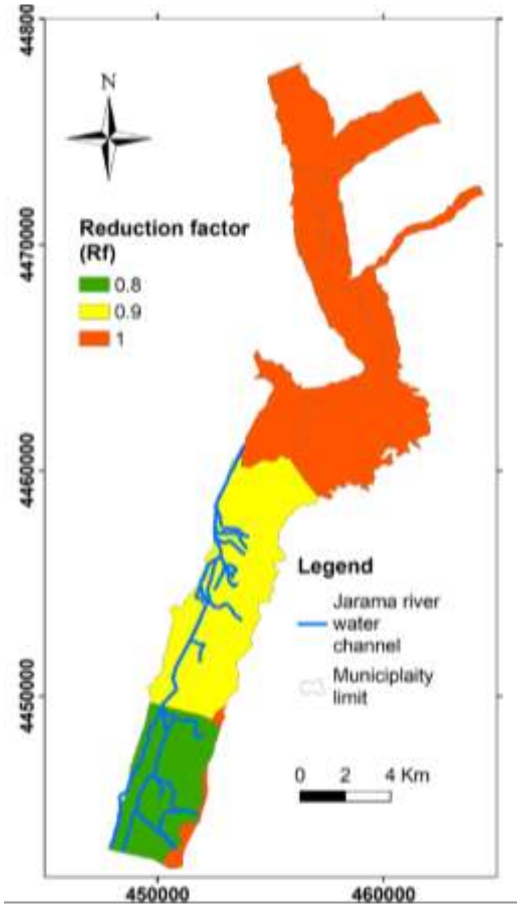


Fig.2 Spatial distribution of the reduction factor (Rf) in the study area.

Finally, the socio-economic vulnerability was calculated by applying the reduction factor to Eq. 4 (Eq. 5).

$$V_{S-E}\left(\frac{€}{year}\right) = R_f * (Land\ prices + crops\ production + livestock\ production + agricultural\ employment) \quad (5)$$

The ArcGIS v10.4.1 Raster calculator tool was used to sum and multiply the maps.

- ✓ ***Land prices***

The price of land is directly associated to the type of activity developed. As mentioned before, the agricultural activity is the most important economic activity in the study area. For this reason, it was considered to obtain land prices according to the types of crops and irrigation uses. It is important to note that the soil contamination by irrigation with polluted water degrades the soil conditions for future crops, thus devaluing the land. Land prices for 2019 in the study area are shown in Table 4.

Table 4 Land prices according to crop type and irrigation use for 2018-2019

Land type	Crop type	€/ha
Non-irrigated land	Arable/Herbaceous	7079
	Fruit trees	8979
	Vineyard	14474
	Olive groves	20586
Irrigated land	Arable/Herbaceous	17552
	Vegetables	35534
	Rice	31048
	Cítricos	40186
	Fruit trees	37770
	Vineyard	23105
	Olive groves	38506
	Pasture	6298
	Natural meadow	8247

(Brezmes 2018; Subdirección General de Análisis, Coordinación y Estadística. 2020)

The land use map (IGN 2018) was used to delimit the type of crop type and land classes in the study area (Fig. 3).

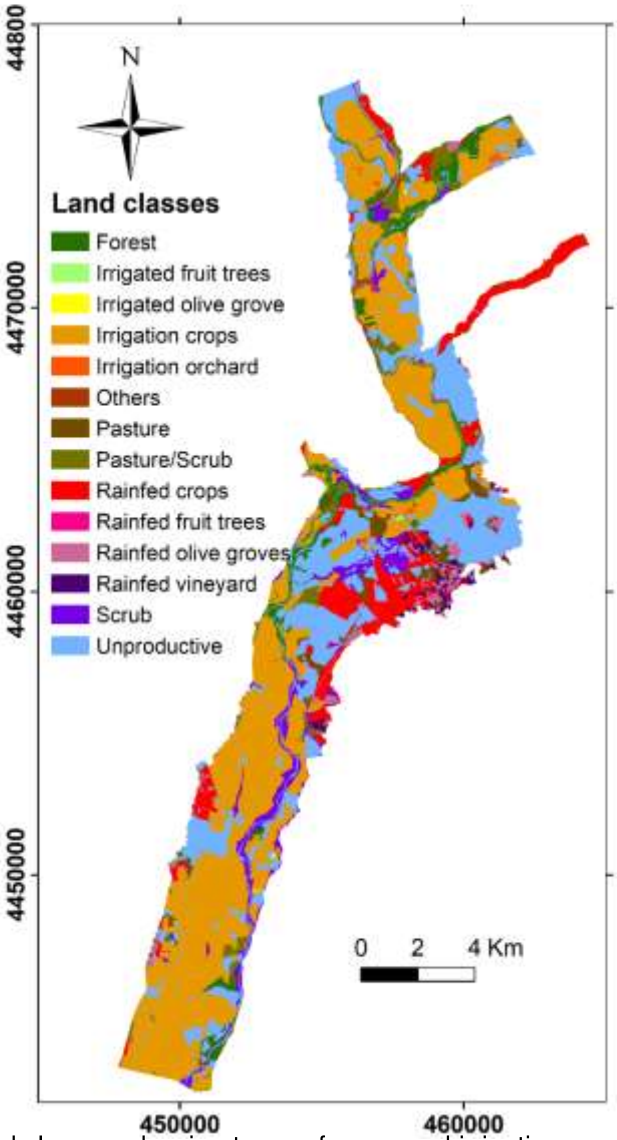


Fig.3 Land classes, showing types of crops and irrigation uses.

- ***Agricultural production (crops)***

There are two types of crops in the study area: arable crops and woody crops. Both are delimited in the agricultural regions established by the government authority (Ministerio de Agricultura, Alimentación y Medioambiente de España) (Fernández González 2013). Thus, the agricultural regions provided the information about surface of the cultivated areas of different crops.

The main arable crops in the study area are: wheat, barley, corn, chickpea and oat (Fernández González 2013). Each type of crop has its own yield and market price (MAPA 2020, 2021c, d) (Table 5).

Table 5 Production of arable crops (2020-2021)

Arable crop	Yield (Tn/ha)	Price €/Tn	Production €/ha
Wheat	3.8	247.0	938.6
Barley	4.0	184.0	736.0
Corn	11.9	221.0	2629.9
Chickpea	2.5	138.0	345.0
Oatmeal	1.3	650.0	845.0

(MAPA 2020, 2021c, d)

The main woody crops in study area are: Vineyard, olive groves and fruit trees (not citrus) (Fernández González 2013). As with arable crops, each type of crop has a particular yield and a market price (MAPA 2020; Subsecretaría de Agricultura, Pesca y Alimentación 2020) .

The value of the production obtained by woody crops varies according to the specific product obtained. For this reason, the prices of the different product were averaged to obtain a single value per product (Table 6). The production of each type of woody crop is shown in Table 7.

Table 6 Market prices of woody crops products (2021)

Woody crop	Product	Marker price (€/Kg)	Price average (€/Kg)
Vineyard	Grapes	0.35	0.35
Olive groves	Olive oil	3.10	1.9
	Olives	0.73	
Fruit trees (not citrus)	Stone fruits (Cherry, plum, kiwi, peach, nectarine and medlar)	0.15	0.12
	Pome fruits (apple and pear)	0.08	

(MAPA 2021c, d)

Table 7 Woody crops production by 2021 in Spain

Woody crop	Yield (Kg/ha)	Price €/Kg	Production €/ha
Vineyard	4685	0.35	1640
Olive grove	1014	1.9	1927
Fruit trees	300	0.12	36

(MAPA 2021c, d)

The agricultural production of each type of crop was obtained by multiplying the cultivated area of the different crops by the economic production. The ArcGIS v10.4.1 Raster calculator tool was used to multiply maps.

Finally, the total production of agricultural crops was obtained from the sum of agricultural production maps (arable and woody) obtained for different crops. To obtain that value, the maps were summed using the Raster calculator tool of ArcGIS v10.4.1.

✓ ***Livestock production***

Livestock production in the study area is based on bovine, ovine, goats, porcine and poultry livestock, depending on the water resources available from different groundwater bodies (Table 8).

Table 8 Livestock yield related to from different groundwater bodies

			Heads/ha				
Groundwater body	Code (CHET)	Area (ha)	Bovine	Ovine	Goats	Porcine	Poultry
Guadalajara	030.006	731.6	0.05	0.34	0.05	0.10	0.63
Aluviales Jarama-Tajuña	030.007	11209.1	0.06	0.46	0.04	0.02	0.10
Aluvial del Jarama: Guadalajara-Madrid	030.024	771.3	0.01	1.71	0.10	0.21	0.09

(CHT 2015a)

Each livestock production generates different products, which have an associated yield and market price. The number of livestock per hectare multiplied by the surface area of each groundwater body resulted in the total number of heads for each type of livestock. The total production for the different types of livestock was obtained by multiplying the number of heads related to each groundwater body by the annual production in €/year of each associated by-products (meat, milk, wool and eggs) (Tables 9,10 and 11).

The total livestock production was obtained using the following equation applied to the groundwater bodies in the study area (Eq. 6):

$$Livestock\ production(\frac{€}{year}) = Bovine\ production + Ovine\ production + Goat\ produccion + porcine\ production + poultry\ production \quad (6)$$

The sum of each of the by-products and the total livestock production was done with the map algebra tool in ArcGIS v10.4.1.

Table 9 Meat production from different types of livestock

Livestock	Carcase weight (Kg/unit)	Price€/Kg	Annual production €/head
Bovine	287.45	2.50	718.63
Ovine	13.19	5.0	65.95
Goats	6.27	6.9	43.26
Porcine	123.11	1.36	167.43
Poultry	2.41	1.30	3.13

(MAPA 2021a, e).

Table 10 Milk production from different types of livestock

Livestock	Milk production (Kg)	Price €/Kg	Annual production €/head
Bovine	5835.25	0.3243	1892.37
Ovine	166.23	0.95	157.92
Goats	457.92	0.80	366.34

(MAPA 2021a, e).

Table 11 Other products from different livestock types

Livestock	Product	Annual production (Kg/unit)	Price€/Kg	Annual production €/head
Ovine	Wool	2.54	2.00	5.08
Poultry	Eggs	10.00	0.88	8.80

(MAPA 2021a, e)

- ✓ ***Employment related to agricultural activities***

In this research, employment was considered the main social variable to evaluate the vulnerability. Although there are four productive sectors in the study area (agriculture, industry, building and services), the agricultural sector is the main and most important sector that depends on groundwater resources. For this reason, the agricultural employment was chosen to assess the impact of the groundwater contamination. The employment was calculated considering the permanent employment between June 2020 and July 2021, and the employment density per square kilometre (Table 12).

Table 12 Agricultural employment in the study area (June 2020-july 2021)

Municipality in study area	Agricultural employment (2020-2021)	Municipal extension (Km2)	Agricultural employment density (Employment/Km2)
Arganda del Rey	55	80	0.69
Aranjuez	242	189.2	1.28
Chinchón	129	115.84	1.11
Ciempozuelos	16	49.31	0.32
Coslada	0	12.02	0.00
Loechos	2	44.09	0.05
Madrid	4798	605.14	7.93
Mejorada del Campo	0	17.09	0.00
San Fernando de Henares	8	39.79	0.20
San Martín de la Vega	36	105.84	0.34
Tilucia	0	9.88	0.00
Torrejón de Ardoz	77	32.65	2.36
Rivas Vaciamadrid	17	67.16	0.25
Valdemoro	21	64.53	0.33
Velilla de San Antonio	6	14.42	0.42

(Consejería de economía, hacienda y empleo 2021)

To assess the social impact of employment in the sector, the economic value generated through wages was considered. Taking into account that the average salary in agricultural sector from 2019-2020 period in Spain was 16470€/year (MAPA 2021b), the value of employment was calculated using the following equation (Eq. 7).

$$Agricultural\ employment\left(\frac{€}{year}\right) = Number\ of\ contracts * Annual\ salary \quad (7)$$

The raster calculator tool in ArcGIS v10.4.1 was used to obtain the agricultural employment map.

2.4.Social vulnerability mapping by cluster analysis (K-means method)

K-means cluster analysis was applied to the entire data set for the three factors obtained above. There were 5842 points (records) and three factors (vulnerability of groundwater resources -V_{GR}-, vulnerability of exposed population -V_P-, and Social and economic vulnerability –V_{S-E}–).

Data processing was carried out using RStudio v.4.0.5 software. Each factor was normalized with the max-min scaling method, in order to reduce the bias caused by predominance of very high ranges over lower ranges. The tool of extract values to point in ArcGIS v10.4.1 was used to obtain the value of each variable for the 5842 points.

The goal of K-means is to cluster data points with intrinsic similarities in the data set. This iterative process started with the selection of the optimal number of clusters, which was determined by the R package NbClust using the majority rule (Charrad et al. 2014). Euclidean distance was used to find the distance from each point in the data set to a temporal cluster. The minimum distance of the sum of squared errors of the distance *A* (Eq. 8) between each point to the centroid of each cluster is considered to locate points in them.

$$A = min \sum_{i=1}^{k} \sum_{x \in ki} \| x_k - mi \|^2 \quad (8)$$

(Dabbura 2020)

Where x_k = *(x1, x2, x3,……..xn)* are the data belonging to the k_i cluster; and m_i is the centroid of the cluster k_i (Eq. 9):

$$m_i = \frac{\sum_{k=1}^{Ni} x_k}{N_i} * x_k \in k_i \quad (9)$$

(Dabbura 2020)

Where N_i is the number of data objects in the cluster *i.*

The procedure finishes when no points are reallocated from one cluster to another or when a pre-defined number of iterations is reached (Dabbura 2020).

3. Results and discussion

3.1. Factors affecting the social vulnerability

3.1.1. Vulnerability of groundwater resources **(V_{GR})**

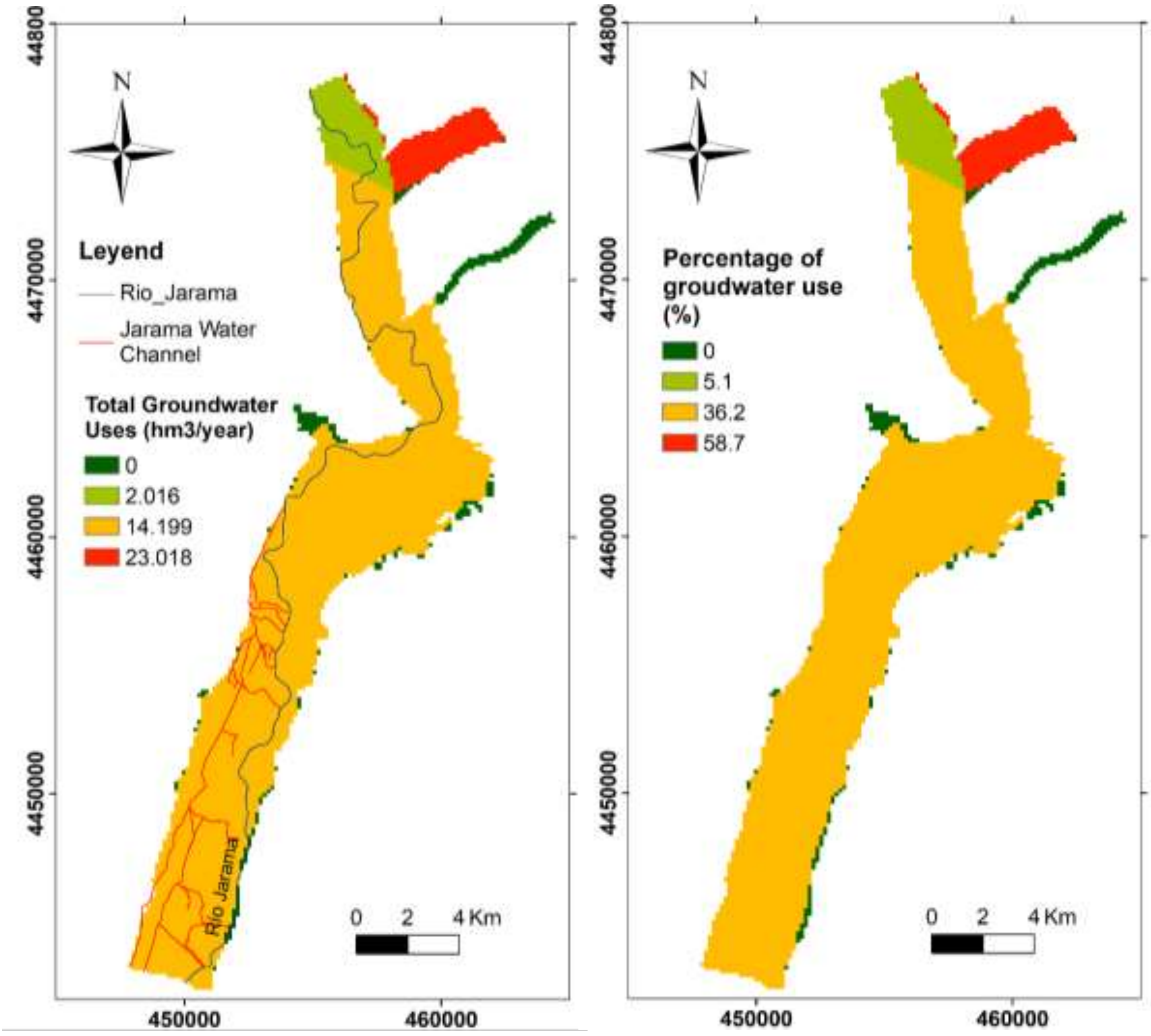
Table 13 shows the amount of groundwater for different uses in each groundwater body. It is clear that agricultural uses account for the largest groundwater consumption in at least two of the three groundwater bodies (more than 50% of groundwater uses). For this reason, the activities associated with this sector are more severely impacted than the others. On the other hand, groundwater body “Guadalajara” has the highest consume, although it represents a small area in the study (Fig. 1).

Table 13 Water uses in groundwater bodies in the study area

Groundwater body	code (CHET)	Groundwater use (hm3/year)			Total of groundwater uses (hm3/year)	% Groundwater use
		Urban use	Agricultural use	Industrial use		
Guadalajara	030.006	6.311	11.772	4.935	23.018	58.7
Aluviales Jarama-Tajuña	030.007	0.895	10.138	3.166	14.199	36.2
Aluvial del Jarama: Guadalajara-Madrid	030.024	0.138	0.674	1.204	2.016	5.1
Total Groundwater use		7.344	22.584	9.305	39.233	100
% Groundwater use		18.7	57.6	23.7	100	

(CHT 2015a)

Fig. 4 shows the spatial representation of total groundwater consumption. Most of groundwater uses (more than 80% of the study area) are located from north to south. They consume 14.20 hm3/year (Fig. 4a), about 40 % of the available groundwater in the region (Fig. 4b), which represents a significant amount of the groundwater resources. This means that an eventual contamination of the aquifer could generate a major impact on the environment and agricultural activities, negatively affecting the economic and social development of the region. In fact, the hydraulic connection between the aquifer and the Jarama river favours an eventual contamination of the river water from the aquifer.



(a) (b)

Fig. 4. Total consumption of groundwater in the study area. (a) Total groundwater uses in hm3/year. (b) Percentage of groundwater for different uses.

3.1.2 Vulnerability of exposed population **(V_P)**

Fig. 5 and Fig.6 show the population density in the study area and its percentage of urban groundwater use, respectively. In general, the study area has a low population, as it is mainly an agricultural region. Therefore, urban development is low. The highest urban groundwater consumption is only 16% and occurs in a small area located at northeast (5% of the total area).

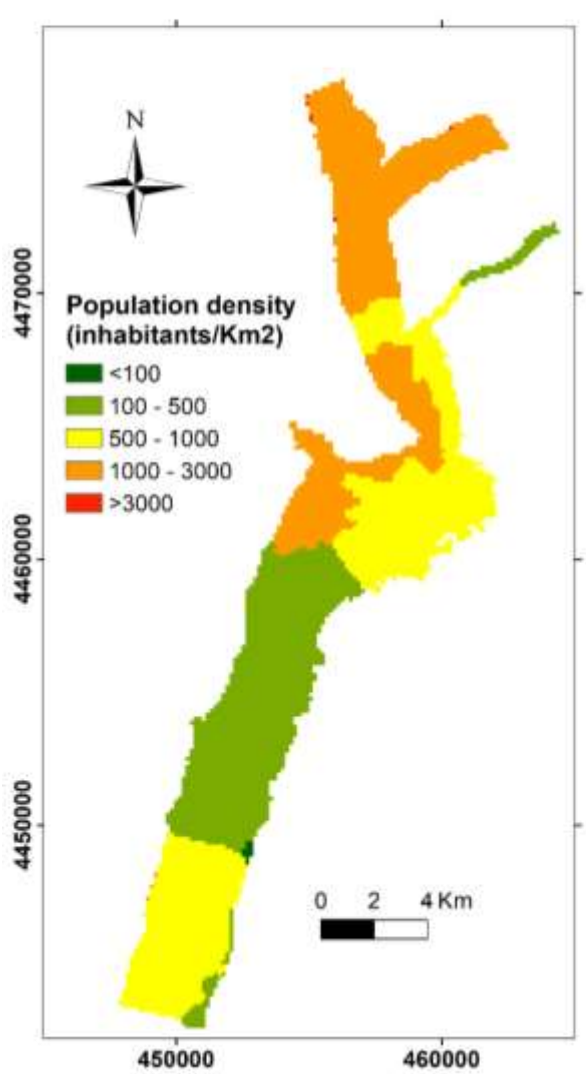


Fig. 5 Population density of the study area.

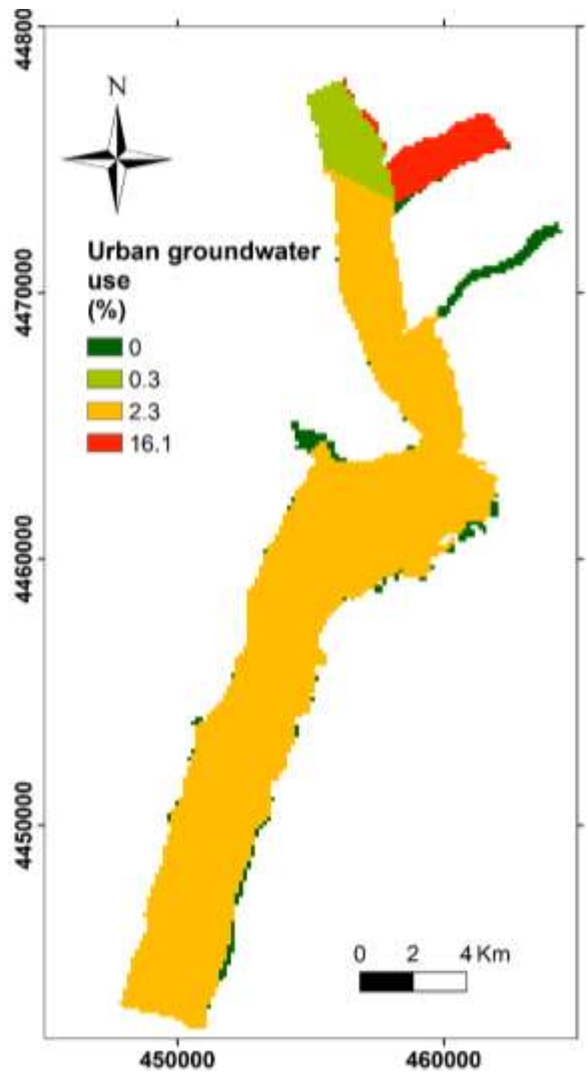


Fig. 6 Urban groundwater use in percentage of total groundwater in the study area

The distribution of the population exposed to the consumption of contaminated groundwater is shown in Fig. 7. By density, the population exposed is low. More than 90% of the study area presents an exposed population density of less than 50 inhabitants per square kilometre,due to the limited urban development in the area. For this reason, the vulnerability of the exposed population do not have a significant influence on the social vulnerability analysis of the study area.

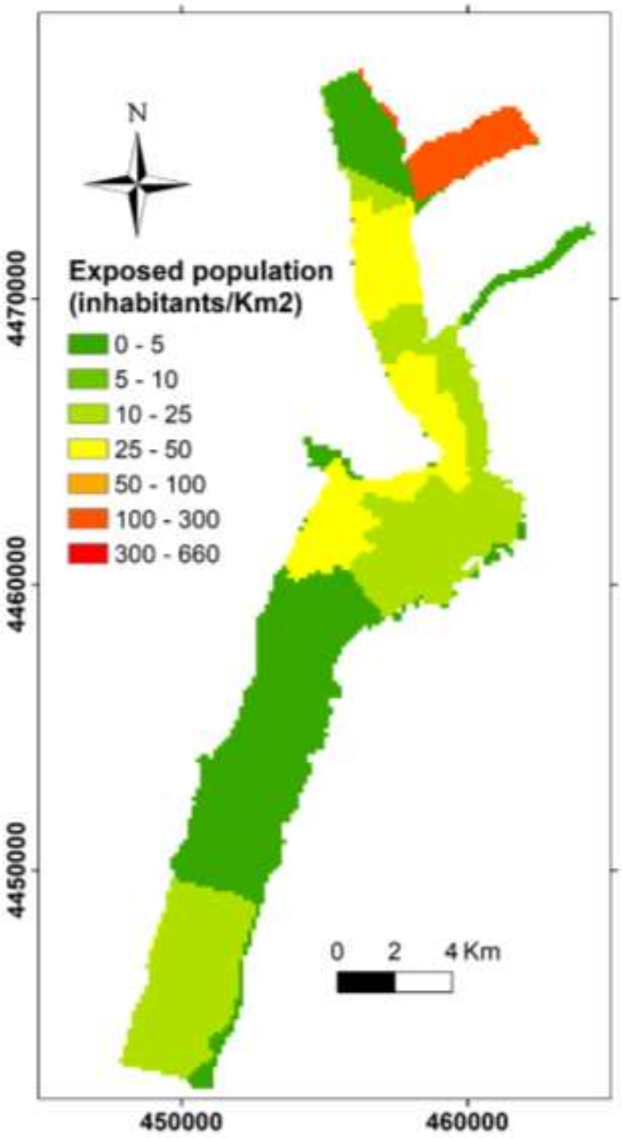


Fig. 7 Density of population exposed to consumption of contaminated groundwater.

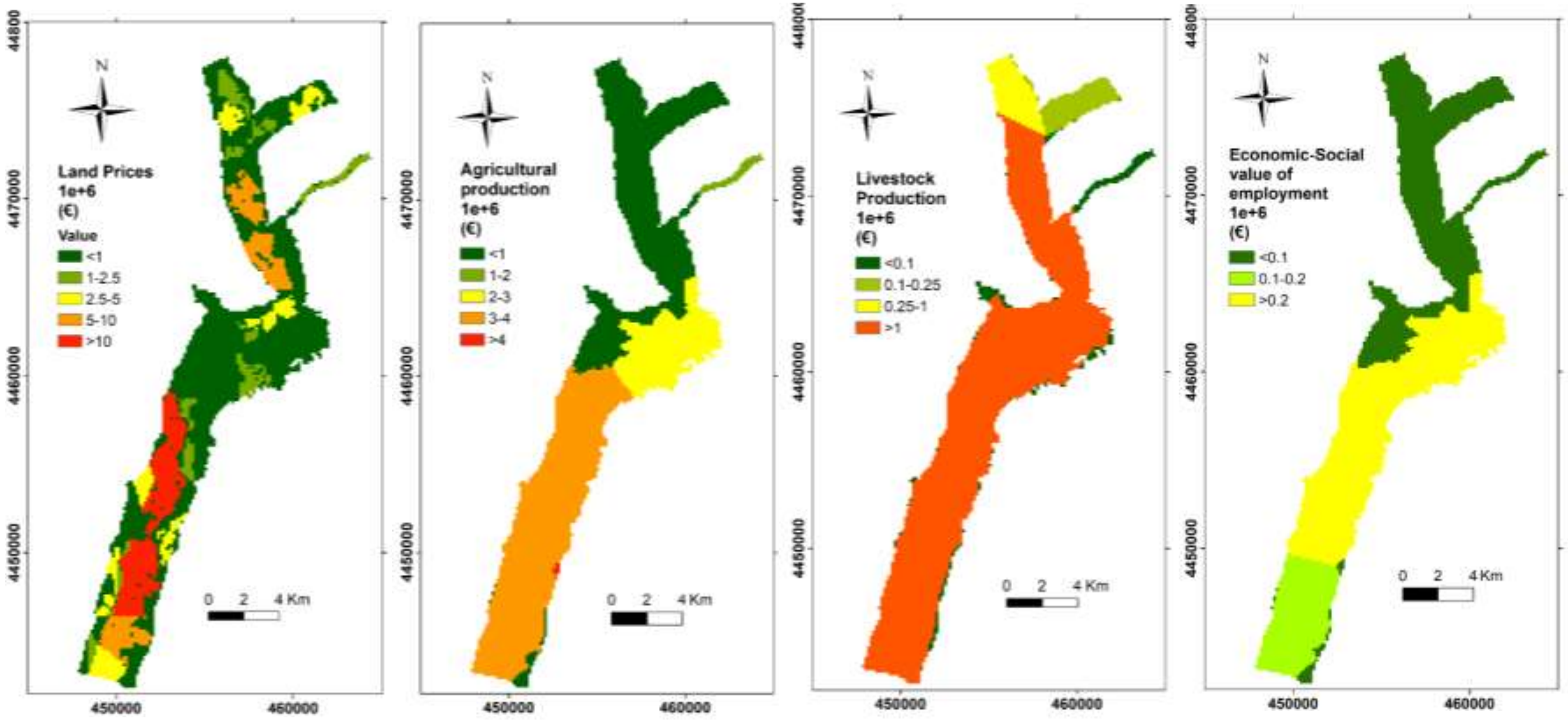
3.1.3 Socio-economic vulnerability ***(V_{S-E})***

As mentioned above, the price of land depends on the agricultural development and is highest where irrigated crops are present. Fig. 8a shows that about 60% of the area has a total land price below one million euros (the lowest value), which is consistent with most of study area having few irrigated crops and a significant portion of unproductive areas. Thus, the land classes map (Fig. 3) shows that most of the areas are unproductive, forested and rainfed crops zones. On the other hand, about 13% of the area has a total land value of more than ten million euros. This corresponds to the sector in the south, with a high development of irrigated crops as the Jarama river canal is providing water for irrigation. The crop production reached about four million euros in most of the study area (44% of the area), mainly by arable crops of corn and wheat and olive groves, located the central-southern part of the study area.

The north of the study area had a low crop production of less than one million euros, by arable crops of barley, corn oat and fruit trees (38% of the study area). In the centre of the area, the production of arable crops of wheat, barley, corn, chickpeas, vineyard, olive groves and fruit trees reached around three million € (Fig. 8b).

About 80% of the study area produces one million euros or more from livestock, mainly bovine and ovine. However, this value is low compared with crop production (Fig. 8c).

The agricultural employment is the most influenced by groundwater uses. The economic value contributed by employment income was less than 300,000 euros/year, being lower than other economic variables. The employment income was higher in the south of the study area (Fig. 8d), due to higher agricultural development.



(a) (b) (c) (d)

Fig. 8 Socio-economic vulnerability parameters. a) Land prices, b) Annual Crop production, c) Annual Livestock production, d) Annual agricultural employment income.

The highest values of socio-economic vulnerability were located in the south of study area (Fig. 9), mainly influenced by land prices and crops production, due to the availability of water to irrigate crops by the Jarama canal. Although in this zone the socio-economic vulnerability decreases due to the reduction factor (Fig. 2) by the irrigation facilities of the Jarama canal, it is still the zone with the highest socio-economic vulnerability. The lowest values are located in the north due to scarcity or absence of crops and low livestock production, which implies low agricultural employment in this zone. In the north-central area, the production of crops and livestock led a moderate socio-economic vulnerability.

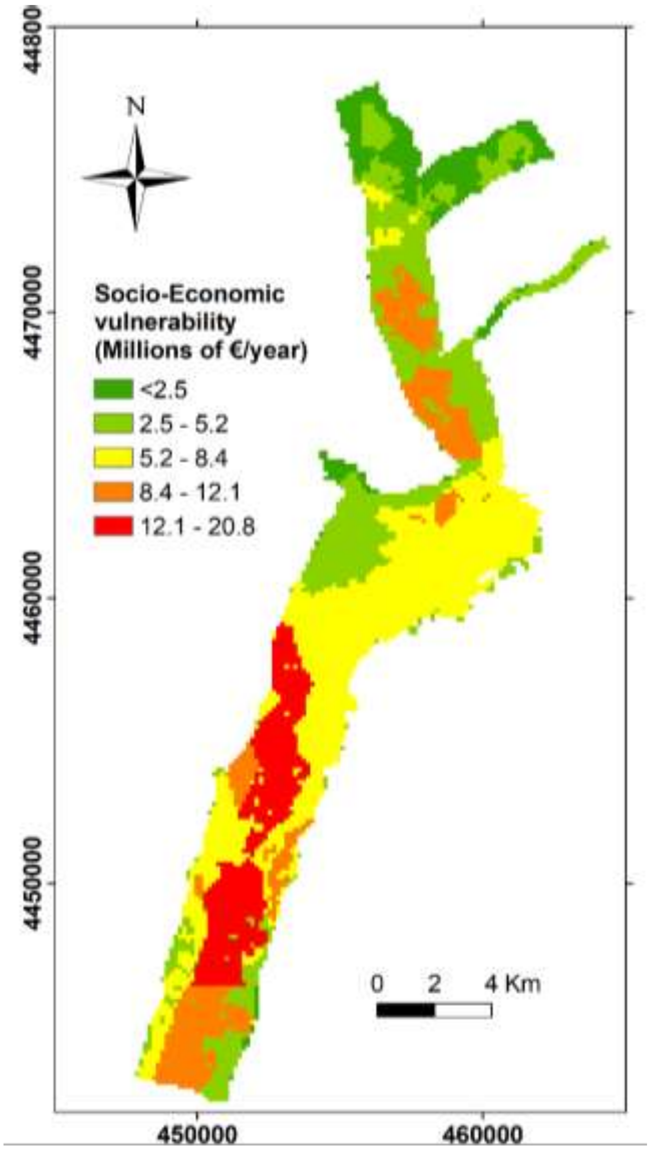


Fig. 9 Economic-social vulnerability distribution (in millions of euros/year).

3.2 Social vulnerability by K-means cluster analysis

The optimal number of clusters obtained by NbClust was five (Fig. 10).

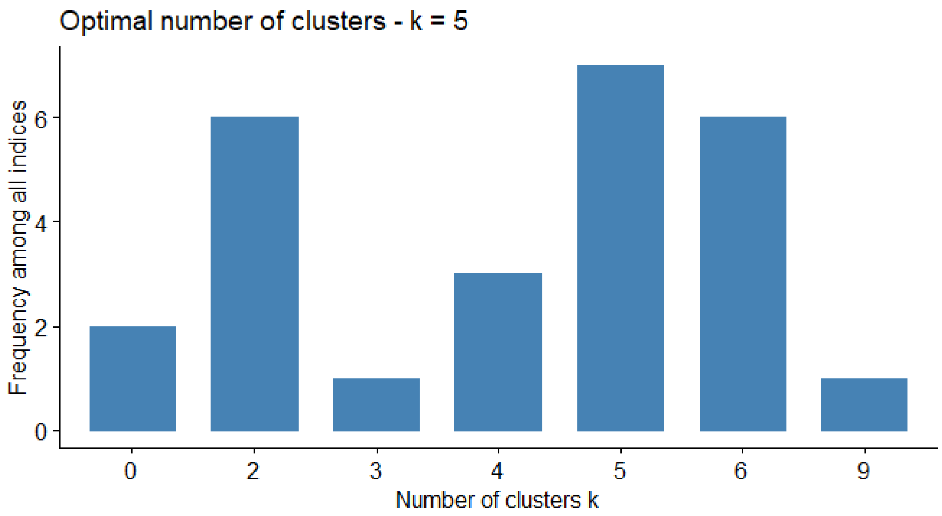


Fig. 10 Optimal number of clusters obtained by NbClust

The results of the K-means cluster analysis are summarized in Table 14.

Table 14. Mean of each factor in the identified clusters.

			V_p (inhab/Km2)	V_{GR} (hm3/year)	$V_{S\text{-}E}$ (million €/year)
Cluster	Points	%	Mean	Mean	Mean
1	774	13.2	6.5	36.2	19.2
2	3263	55.9	15.9	36.2	5.6
3	589	10.1	2.2	3	2.2
4	898	15.4	17	36.2	10.3
5	318	5.4	166.1	58.7	1.8

Cluster 1 gave the highest value of socio-economic vulnerability due to the important agricultural development located in the south of the study area. However, the reduction factor affecting this zone due to the use of surface water to supply irrigation decreases the social vulnerability in this sector. In fact, the groundwater resource has a moderate consumption in this area. In addition, the population in this zone is small, which generates a reduced effect on human consumption. These conditions suggest that this cluster belongs to the high vulnerability mainly influenced by socio-economic factors.

Cluster 3 represents the lowest values of groundwater resource vulnerability and population vulnerability. In turn, social- economic vulnerability in this cluster is very low. These conditions contributed to very low or no social vulnerability. This cluster includes areas where there are no groundwater bodies (at the boundary of the study area and in a narrow branch located in the northeast), which reduces the impact on the water resources. In addition, the groundwater use in the northwest of the “Aluvial del Jarama: Guadalajara-Madrid” groundwater body has the lowest urban, industrial and agricultural consumption (Table 13), resulting in the lowest population and socio-economic vulnerability. Cluster 2 and 4 show low population vulnerability values due to the low urban development in these areas. However, the groundwater resources vulnerability overlaps in the range of index values between these clusters, with values being between the highest and lowest ranges. This suggests that clusters 2 and 4 represent low and moderate social vulnerability, but it is not clear which corresponds to which. The socio-economic vulnerability is used to clarify the classification and assign clusters 2 and 4 to their corresponding social vulnerability level. Cluster 4 has the highest values of socio-economic vulnerability and cluster 2 the lowest ones. For this reason, cluster 2 stands for low vulnerability and cluster 4 stands for moderate vulnerability. In this case, both population and groundwater resource vulnerabilities have been more influential in creating clusters and the socio-economic vulnerability has allowed the classification to be refined.

Cluster 5 includes the highest values for the vulnerability of the exposed population and the vulnerability groundwater resources, therefore a very high social vulnerability would be expected. However, as mentioned above, the population vulnerability did not have a relevant influence on social vulnerability due to the low population. Furthermore, the socio-economic vulnerability was the lowest value, which means that despite the higher impact on the population and groundwater resource, the economic losses due to an eventual groundwater contamination would be low. For this reason, cluster 5 was considered representative of high social vulnerability, similarly to cluster 1.

According to previous discussion, no very high social vulnerability was identified in the clusters analyzed.

The clusters distribution and the social vulnerability maps using K-means analysis are shown in Fig. 11.

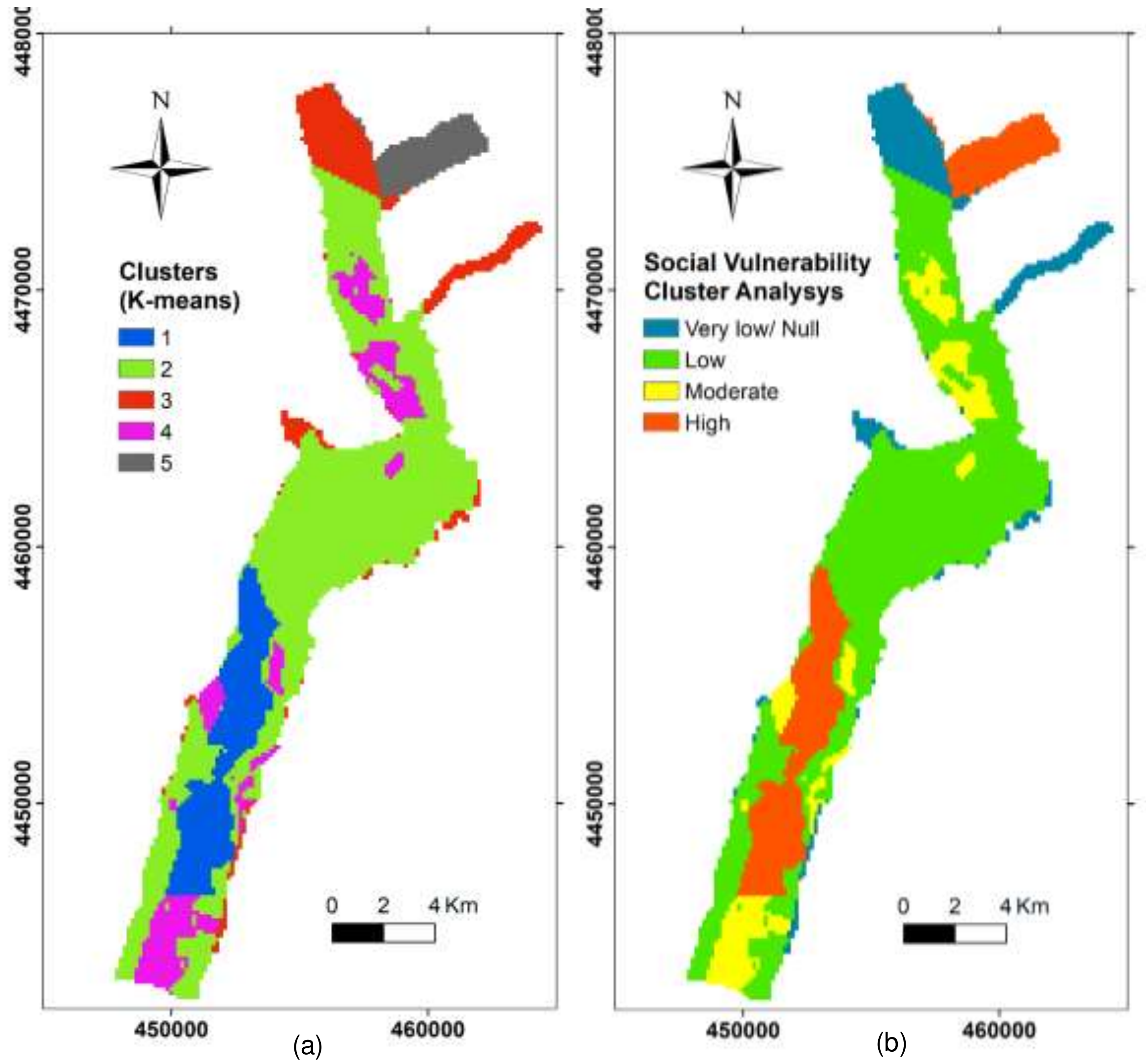


Fig. 11 Social vulnerability results from K-means analysis. (a) Spatial clusters distribution. (b) Map of social vulnerability to groundwater contamination obtained through K-means cluster analysis

According to the results, more than 50% of the study area has a low social vulnerability, mainly located in the central zone and in some parts of central north and on the southern edges of the aquifer, due to the low population located in these areas as well as low agricultural development, which implies a low economic production. Additionally, about 10% of the area has little or no groundwater use, which resulting in very low or no social vulnerability. This condition is observed in the northwest sector of the "Alluvial del Jarama: Guadalajara-Madrid" groundwater body, which only consumes 5% of the groundwater resource in the study area. Some localities are not associated with a groundwater body and therefore in these localities social vulnerability is null. A small part of the study area (11%) located in the centre-north and in some areas at south, shows moderate social vulnerability, mainly influenced by the economic impact on agricultural development. It is important to note that the influence of surface water supplied by the Jarama water canal in the south of the study area reduces the impact of groundwater contamination (irrigation use of surface water is as high as 23% in some localities). Finally, almost 20% of the area presents high social vulnerability, distributed in two different sectors in the study area and influenced by different factors. In the north, the high social vulnerability is due to the high groundwater resource (~50% of the groundwater consumed in the study area) and the impact of the exposed population (~166 inhabitants per square kilometres). In the south of the study area, the high social vulnerability is due to the impact on the social and economic sectors such as agricultural production, employment, and land prices, although this impact is partially reduced by the use of irrigation water from the Jarama water canal. Despite this, the groundwater consumption in this area represents the 36% of the total groundwater resource of the area, which represents a significant amount of the resource used for irrigation needs that negatively affects the economic and social development in the area.

These results demonstrate the advantage of using K-means cluster analysis in the assessment of social vulnerability, since it was not necessary to assign numerical ranks and weights to the considered vulnerability factors because the similarity in the data set grouped the information required to classify the social impact.

Conclusions

The interaction between contaminated groundwater, nature and society requires an assessment of the risk and the damage that a contamination event could cause in a region. In this work, the assessment of social vulnerability allowed the delimitation of zones for an adequate prevention of damage to society by contaminated groundwater. To assess the social vulnerability, three factors were considered: the groundwater resource (V_{GR}), the exposed population (V_P) and the socio-economic impact (V_{S-E}).

The vulnerability of groundwater resources factor (V_{GR}) implied the consideration of an environmental factor due to the possibility of incorporating contaminated water into the Jarama river (and its canal) from the aquifer, causing significant damage to flora and fauna and socio-economic losses. The results point out that the vulnerability of groundwater resources represents an important and influential factor in the assessment of social vulnerability. This factor was estimated from the sum of different uses of groundwater in the region (urban, agricultural and industrial). The agricultural uses account for more than 50% of the groundwater use, as agricultural activities are the most important productive sector in the region that depend on groundwater resources. On the other hand, a high value of groundwater use was identified due to agricultural, industrial and urban activities in the northeastern part of the study area. Although small, this area revealed an important influence on the assessment of social vulnerability.

The study area is sparsely populated due to its significant agricultural development, which reduces the risk of people exposed to contaminated groundwater consumption. In fact, less than 50 inhabitants per square kilometre are exposed in most of the study area (more than 90%). For this reason, the exposed population has a low influence on social vulnerability.

The socio-economic vulnerability is closely influenced by the main productive activity in the region. Agricultural activities involve crops and livestock production that affect land prices and agricultural employment. The highest socio-economic vulnerability was observed in the south of the study area, despite the fact that the Jarama canal is located in this area. This effect reduces the socio-economic vulnerability by 10-20% due to the use of a combination between surface water and groundwater for irrigation. However, the social vulnerability is this area remained high. This area covers ~14% of the study area, representing a small zone but significantly affected area.

The K-means cluster analysis made it possible to assess and delimit areas to classify the social vulnerability without using weighting and rating values for the three factors considered, thus reducing the subjectivity of the methodology. Five clusters were obtained, revealing four levels of social vulnerability: Very low (Cluster 3), low (Cluster 2), moderate (Cluster 4) and high (Clusters 1 and 5) social vulnerability, respectively.

Most of the social vulnerability in the study area is low, which means that an eventual groundwater contamination would not cause a major impact in society. However, in the south (14.2% of the area) and in the northeast (5.7% of the area) there are small zones that could be highly affected with a more important impact. In the northeast due to the exposed population and groundwater resource, and in the south due the socio-economic factor related to agricultural development. In turn, in the northwest of the study area there is a small area (10.9%) that is not affected by a groundwater contamination at the social level. The small and scattered areas in the centre-north and south present moderate social vulnerability, due to the groundwater resources factor (they represent the 36% of the total of groundwater consumption) and the socio-economic factor related to the agricultural development. The

K-cluster analysis clearly demonstrated the advantage of using this technique in the assessment of social vulnerability to groundwater pollution.

Authors Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Marisela Uzcategui-Salazar. The first draft of the manuscript was written by Marisela Uzcategui-Salazar and Javier Lillo and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate Not applicable.

Consent to publish Not applicable

Competing Interests The authors declare no competing interest.

Financial interests: The authors declare they have no financial interests.

Funding No funds, grants, or other support was received.

Availability of data and materials The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Brezmes G (2018) Precio por hectárea de fincas destinadas a pastos, prado natural o pastizal. In: INGENIEROS y PERITOS AGRÓNOMOS. https://tasagronomos.com/precio-tasacion-pastizal-prado-pradera/. Accessed 24 Sep 2021

Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust package : Manual

CHT (2015a) Plan Hidrológico de la parte española de la Demarcación Hidrológica del Tajo. Anejo 3 Usos y demandas de agua. In: Confederación Hidrográfica del Tajo. http://www.chtajo.es/LaCuenca/Planes/PlanHidrologico/Planif_2015-2021/Paginas/Plan_2015-2021.aspx. Accessed 2 Nov 2021

CHT (2015b) Plan Hidrológico de la parte española de la Demarcación Hidrológica del Tajo. Memoria parte española de la demarcación hidrográfica del Tajo Plan hidrológico de cuenca

Comunidad de Madrid (2020) Banco de datos territorial. Densidad de población (Hab/km2). http://gestiona.madrid.org/nomecalles/Inicio.icm?sesionBDT=545508. Accessed 10 Aug 2021

Comunidad de Regantes de la Real Acequia del Jarama (2021) Datos relevantes – Comunidad de Regantes de la Real Acequia del Jarama. https://www.canaljarama.es/datos-relevantes/. Accessed 16 Mar 2022

Confederación Hidrográfica El Tajo (CHET) (2010) Masas de agua subterránea. In: Descarga de capas. La cuenca. Confederación Hidrogáfica El Tajo. http://www.chtajo.es/LaCuenca/Paginas/DescargaDCapas.aspx. Accessed 2 Nov 2021

Consejería de economía, hacienda y empleo (2021) Mercado de trabajo en los Municipios Madrileños

Cutter S (1996) Vulnerability to environmental hazards. Progress in Human Geografhy 20:529–539

Dabbura I (2020) K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks. In: Medium. https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a. Accessed 26 May 2021

Ducci D (1999) GIS Techniques for Mapping Groundwater Contamination Risk. Natural Hazards 20:279–294. https://doi.org/10.1023/A:1008192919933

Fernández González J (2013) Caracterizacion de comarcas agrarias en España. Tomo 32. Comunidad de Madrid, Ministerio de agricultura, alimentación y medioambiente. Catálogo de publicaciones de la administración general del estado

French M, Alem N, Edwards SJ, et al (2017) Community exposure and vulnerability to water quality and availability: a case study in the mining-affected Pazña Municipality, Lake Poopó Basin, Bolivian Altiplano. Environmental Management 60:555–573. https://doi.org/10.1007/s00267-017-0893-5

Grondona S, Sagua M, Massone H, Miglioranza K (2015) Evaluación de la vulnerabilidad social asociada al consumo de agua subterránea en la cuenca del río Quequén Grande, provincia de Buenos Aires, Argentina. Revista internacional de contaminación ambiental 31:351–359

IGN (2021) Centro de Descargas del CNIG (IGN) Límites municipales de España. In: Centro de Descargas del CNIG. Organismo Autónomo Centro Nacional de Información Geográfica. http://centrodedescargas.cnig.es. Accessed 2 Nov 2021

IGN (2018) CORINE Land Cover . Scale 1:100.000. In: Instituto Geográfico Nacional.Centro de Descargas del Centro Nacional de Información Geográfica (CNIG). http://centrodedescargas.cnig.es

MAPA (2021a) Producción y mercados ganaderos. Sectores ganaderos: Vacuno, Ovino-caprino, Porcino, Avícola. In: Ministerio de Agricultura, pesca y alimentación. https://www.mapa.gob.es/es/ganaderia/temas/produccion-y-mercados-ganaderos/sectores-ganaderos/. Accessed 29 Sep 2021

MAPA (2021b) Índices y salarios agrarios. In: Ministerio de Agricultura, pesca y alimentación. https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/economia/precios-percibidos-pagados-salarios/salarios-agrarios/default.aspx. Accessed 1 Oct 2021

MAPA (2021c) Balances de gestión de cereales. In: Ministerio de Agricultura, pesca y alimentación. https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/cultivos-herbaceos/cereales/balances-de-gestion-de-cereales/. Accessed 6 Sep 2021

MAPA (2021d) Evolución de los precios de los principales cereales. In: Ministerio de agricultura y pesca. Dirección general de producciones y mercados agrarios. https://www.mapa.gob.es/en/agricultura/temas/producciones-agricolas/cultivos-herbaceos/cereales/evolucion-de-los-precios-de-los-principales-cereales/. Accessed 3 Sep 2021

MAPA (2020) Mercados agrícolas y ganaderos. In: Ministerio de agricultura pesca y alimentación. https://www.mapa.gob.es/es/agricultura/estadisticas/mercados_agricolas_ganaderos.aspx. Accessed 6 Sep 2021

MAPA (2021e) Catálogo oficial de razas. In: Ministerio de Agricultura, pesca y alimentación. https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/default.aspx. Accessed 28 Sep 2021

Mostaza D (2019) Estudio de la relación entre las aguas superficiales y subterráneas de la Masa de Agua Subterránea (MAS) 030.007 "Aluviales: Jarama-Tajuña”. Ph.D. Thesis, Universidad Rey Juan Carlos

Mostaza-Colado D, Carreño-Conde F, Rasines-Ladero R, Iepure S (2018) Hydrogeochemical characterization of a shallow alluvial aquifer: 1 baseline for groundwater quality assessment and resource management. Science of The Total Environment 639:1110–1125. https://doi.org/10.1016/j.scitotenv.2018.05.236

Orellana-Macías JM, Perles Roselló MJ (2022) Assessment of Risk and Social Impact on Groundwater Pollution by Nitrates. Implementation in the Gallocanta Groundwater Body (NE Spain). Water 14:202. https://doi.org/10.3390/w14020202

Perles M, Vías J, Navarro B (2008) Vulnerability of human environment to risk: Case of groundwater contamination risk. Environment international 35:325–35. https://doi.org/10.1016/j.envint.2008.08.005

Salazar C, Del Castillo S (2018) FUNDAMENTOS BÁSICOS DE ESTADÍSTICA, 1st edn.

Subdirección General de Análisis, Coordinación y Estadística. (2020) Encuesta de Precios de la Tierra 2019 (Base 2016). Resultados septiembre 2020

Subsecretaría de Agricultura, Pesca y Alimentación (2020) Encuesta sobre Superficies y Rendimientos Cultivos (ESYRCE). Encuesta de Marco de Áreas de España.

Vias J (2005) Desarrollo metodologico para la estimacion y cartografia del riesgo de contaminacion de las aguas subterraneas mediante SIG. Aplicacion en acuiferos al sur de Espana. Tesis doctoral, Universidad de Malaga