Journal homepage http://revistas.unitru.edu.pe/index.php/SSMM

SELECCIONES MATEMÁTICAS
Universidad Nacional de Trujillo
ISSN: 2411-1783 (Online)
2020; Vol. 10(1): 69-80.

Special issue: Peruvian Conference on Scientific Computing 2022, Cusco - Peru

K-means algorithm and multicriteria analysis to determine drainage zones in a geographic region

Julio César Peralta Castañeda

Received, Jan. 15, 2023 *Accepted, Apr. 30, 2023*

How to cite this article:
Peralta J. *K-means algorithm and multicriteria analysis to determine drainage zones in a geographic region*. Selecciones Matemáticas. 2020;10(1):69–80. http://dx.doi.org/10.17268/sel.mat.2023.01.07

Abstract
This research work presents the results of the external analysis of the problem of the management of the lower Moche river basin in the valley of the same name. These results are framed in the construction of a multicriteria objetive function, to determine a partition of the basin in particular study zones calleded cluster.

The objetive function involves the main hydrogeological parameters and some external parameters such as coefficients that involve socio-political criteria and some too specific such as evapotranspiration of different types of plants cultivated in the valley on the underground aquifer.
Keywords . Multicriterial Analysis, Fundamentals hidrogeologic parameters, Underground Aquifer.

1. Introduction. In different parts of the world, several years ago local Agendas 21 began to be formulated and in our city the Provincial Council was in charge of carrying out this project, whose strategic objective was to improve local governance as contemplated in the Trujillo Local Agenda 21 project[1].

One of the first documents was the Atlas Ambiental de Trujillo Metropolitano, the first thematic document of its kind, made on our city that using modern cartography describes the natural characteristics and people that occupy the city of Trujillo. Among the studies developed, they show that part of Trujillo city is located in an area of soils with high liquefaction potential, especially the area bordering the coastline, where the water table is high and is expected to rise further, due to the infiltration of water used in agricultural activities and/or when some extraction wells stop working, due to age or because the demand will be satisfied by the water derived from the Santa River, through the Special Project CHAVIMOCHIC[2, 3].

In the city of Trujillo, considering that until 1997, the scarcity of water resources from the Moche River was increasing, it was decided to extract water from the subsoil for agricultural use, especially for cultivation of sugar cane, carried out by the company called Agroindustrial Laredo, which over the years changed its corporate name and extracted approximately 5.000.000.00 m^3/year.

Likewise, the current Empresa de Servicio de Agua Potable y Alcantarillado de La Libertad Sociedad Anónima (SEDALIB) continued with the exploitation of underground water through tubular wells, this exploitation acts as a vertical drainage that contributes to the decrease of the water table.

One of the major projects to achieve the development of the area is the execution of the Special Project CHAVIMOCHIC, which includes the diversion of the waters of the Santa river, from the Chuquicara sector and lead them to the pampas of Urricipe to the right margin of the Chicama valley, these waters cross the valleys: right margin of Santa river, Chao, Viru, Moche and Chicama, through a trapezoidal channel called "Canal Madre" with an approximate length of 275 km, distributed between open canals, closed conduits and tunnels, distributed among open channels, closed conduits and tunnels, transporting a water flow that

*Instituto de Investigación en Matemática, Departamento de Matemáticas, Universidad Nacional de Trujillo, Trujillo, Perú. (jperaltac@unitru.edu.pe).

varies from 95 to 16 m^3/s, if we consider only the Moche valley we will have an approximate flow of 2 to 14 m^3/s.

Considering that the irrigation efficiency would reach 50%, it will be possible to observe that the other 50% of the water flow will be constituted in subway flow that would go to recharge the existing aquifers raising the water table; therefore, the study "Planning of the Moche Valley Main Drainage System" was carried out, allowing the execution of 45 km of main drains. This study recommended that Sedalib and Empresa Agroindustrial de Laredo continue with the exploitation of groundwater, and also recommended the delivery of water from the main chanel of CHAVIMOCHIC project to the Trujillo drinking water treatment plant for population use; This action would cause a water imbalance in the valley and therefore overload the aquifer due to deficient use of the drainage system as a result of poor management of fluid dynamics[4, 5, 6, 7].

To carry out the study of the underground aquifer it is necessary to do a series of studies, among the most important is to know the fundamental properties of the aquifer. Once these are determined, we use the clustering theory to generate a zoning of the region under study, in the application of this theory it is necessary to build a similarity function between "cells"[1].

The construction of the similarity function is based on the information obtained from the observation data of groundwater table piezometric levels as well as from the hydrogeological studies carried out for the valley[8, 9].

With the similarity function we can then partition the study area using the $K-$means algorithms for clustering data, which in our case are representative points of a cell[10, 11, 12, 13, 14].

Finally, in the last part of this work, the multi-objective function is constructed, in which we include fundamental hydrogeological properties of the aquifer, which may include parameters such as physical, biological, structural, agronomic, environmental, etc. coefficients, and even political-socio-economic coefficients, which will allow appropriate decisions to be taken[15].

2. Materials and methods. The present study assumes the research developed by the author in 2004, in the work [16], regarding the geomorphological study and geographic location of the valley, also makes use of the concepts presented below on Fundamental Hydrological Parameters of aquifers, drainage and clustering theory.

2.1. Fundamental Hydrological Parameters. Because aquifers are physical media that obey certain physical laws, which in turn have a set of fundamental parameters, including porosity, permeability or hydraulic conductivity and storage coefficient, each of which has its own characteristics associated with the environment and different circumstances.

Porosity (m) This property is expressed by the ratio between the volume of the empty part or part occupied by air or water V_v and its total volume V_t, obtaining the following relation:

$$V_t = V_v + V_m, \tag{2.1}$$

where V_m is the volume occupied by the medium or substrate from which it follows that the porosity of the medium is given by:

$$m = \frac{V_v}{V_t}. \tag{2.2}$$

Permeability or Hydraulic Conductivity K This is one of the most important parameters in the study of aquifers, it is given by:

$$K = \frac{v}{I}, \tag{2.3}$$

where: v represents the average velocity of fluid and I is the hydraulic gradient.

Transmissivity (T) This concept is defined as the flow rate that percolates through a vertical strip of land of unit width and height equal to that of the saturated permeable mantle under a unit gradient and a fixed temperature[17].

Storage coefficient (S) The storage coefficient is defined as the volume of water that can be released by a straight prism of section equal to unity and of the same height from the saturated aquifer. This quantity is non-dimensional.

2.2. Data description. In order to develop this application work in the Moche valley, we have very little information, the data we have regarding the observation wells of the entire valley are piezometric levels, UTM coordinates, topographic elevation, groundwater acidity and electrical conductivity in each of the wells of the valley, as shown in Table 2.1.

[1] these cells are built as a partition of affected area, and the cells are considered because due to the extension of the terrain a very fine partition cannot be made because the calculation to be performed becomes too expensive

Table 2.1: Data observed in the Moche Valley for the month of September, 2005.

No.	UTM	UTM		PNF	CNF	C.E		Nro.	UTM	UTM		PNF	CNF	C.E	
	East	North	msnm	(m)	(msnm)	(mS/cm)	pH		East	North	msnm	(m)	(msnm)	(mS/cm)	pH
1	719021	9099111	10.98	0.78	10.20	1.70	7.81	75	716409	9099220	4.50	0.88	3.62	4.24	8.00
2	717627	9098204	5.21	1.47	3.74	1.21	8.42	76	716712	9100252	12.80	5.36	7.44	1.71	8.00
3	718946	9098098	12.86	4.09	8.77	0.90	7.93	77	713706	9105174	42.43	11.61	30.82	0.75	
4	719004	9097642	10.55	2.88	7.67	1.21	8.10	78	712539	9105647	44.92	9.62	35.30	0.69	8.74
5	719019	9096879	7.00	1.55	5.45	1.27	8.03	79	712094	9106358	50.44	11.87	38.57	0.73	8.86
6	717998	9097535	6.99	2.36	4.63	0.97	8.07	80	710894	9106896	45.86	5.14	40.72	1.07	8.66
7	717316	9097570	3.51	1.39	2.12	2.52	8.01	81	709869	9108206	44.74	15.61	29.13	1.03	8.97
8	718004	9097007	3.38	0.26	3.12	1.53	7.97	82	710998	9108408	47.70	22.88	24.82	1.88	8.16
9	718335	9096758	3.80	0.86	2.94	1.72	7.98	83	710033	9107403	42.99	9.00	33.99	0.98	3.30
10	719225	9096254	3.20	2.25	0.95	0.73	8.38	84	709465	9107230	36.28	8.45	27.83	1.07	8.27
11	718822	9095813	3.15	1.90	1.25	2.70	8.48	85	708851	9106219	23.85	8.07	15.78	1.98	
12	719385	9095090	3.13	2.04	1.09	1.32	8.52	86	709575	9106068	26.34	2.42	23.92	1.73	8.39
13	720703	9093946	2.90	1.76	1.14	1.03	8.51	87	707759	9105556	6.82	5.19	1.63	3.39	8.26
14	720336	9094427	2.43	0.47	1.96	1.69	8.28	88	708540	9106867	20.70	16.10	4.60	1.22	
15	720008	9095849	5.91	0.75	5.16	1.07	8.74	89	708746	9104479	5.32	2.72	2.60	7.81	8.32
16	720329	9096610	10.47	2.55	7.92	1.24	7.93	90	709563	9103607	3.55	1.04	2.51	6.91	8.62
17	720896	9095446	7.79	2.27	5.52	1.51	8.30	91	709986	9103325	3.51	1.37	2.14	5.59	8.46
18	721206	9094491	5.50	1.10	4.40	1.14	8.23	92	710419	9103086	2.48	0.40	2.08	3.42	8.50
19	721608	9093883	4.83	2.37	2.46	3.88	8.54	93	709829	9104828	21.01	6.42	14.59	4.95	8.49
20	720975	9095727	9.91	3.29	6.62	2.29	8.17	94	710089	9105789	28.66	4.98	23.68	8.76	8.34
21	720876	9096254	7.26	0.87	6.39	1.38	7.95	95	711094	9105662	37.75	5.75	32.00	0.90	8.49
22	719470	9099118	11.44	2.19	9.25	1.31	8.09	96	710504	9104729	24.43	5.25	19.18	2.00	8.85
23	719847	9099715	14.80	1.99	12.81	1.37	8.09	97	711412	9104790	28.43	1.89	26.54	0.86	8.95
24	720390	9098815	17.04	2.09	14.95	1.29	8.04	98	713861	9102815	16.85	0.71	16.14	1.30	8.09
25	720114	9097410	12.43	2.35	10.08	2.17	8.08	99	712734	9101938	7.20	1.24	5.96	1.03	8.04
26	720531	9098173	16.81	3.29	13.52	1.67	8.34	100	712060	9102388	4.55	0.53	4.02	1.27	9.08
27	720730	9099513	21.83	4.52	17.31	1.20	8.12	101	713352	9102200	14.30	1.43	12.87	0.96	8.15
28	721190	9101135	28.30	3.07	25.23	1.12	8.16	102	713044	9101587	4.55	0.64	3.91	1.61	8.17
29	720798	9097469	16.09	3.57	12.52	1.33	8.45	103	712505	9101546	3.17	0.84	2.33	3.03	8.13
30	720996	9100220	23.63	4.57	19.06	1.05	8.19	104	711931	9102094	4.78	1.50	3.28	1.11	8.95
31	718594	9099833	10.99	2.13	8.86	2.59	8.20	105	711267	9102478	4.29	1.53	2.76	2.70	8.80
32	717176	9100446	11.87	5.82	6.05	1.65	7.80	106	712663	9102594	11.82	1.64	10.18	1.28	9.01
33	717810	9101309	19.06	7.85	11.21	1.80	7.90	107	713452	9103481	24.44	2.44	22.00	0.94	
34	714433	9102446	16.90	1.82	15.08	0.80	8.61	108	712572	9103836	23.93	4.00	19.93	0.99	
35	714423	9103863	31.30	8.59	22.71	0.80	8.46	109	708994	9105220	18.52	6.48	12.04	8.57	8.22
36	719370	9100049	17.18	3.39	13.79	1.50	7.95	110	707911	9106120	19.15	16.02	3.13	3.48	8.65
37	718409	9101068	16.45	4.41	12.04	2.12	7.50	111	711819	9103346	5.90	0.49	5.41	0.89	8.55
38	719693	9102361	24.37	1.57	22.80	1.00		112	726436	9103276	68.88	3.33	65.55	1.39	7.99
39	720474	9102715	29.25	1.92	27.33	1.27	7.50	113	726578	9102303	99.10	12.25	86.85	-	
40	720659	9102333	28.46	0.61	27.85	0.88	7.96	114	725582	9102013	73.92	3.70	70.22	1.07	
41	721004	9102823	28.54	2.48	26.06	1.71	7.88	115	728055	9102775	106.98	11.65	95.33	0.89	7.93
42	721839	9102524	35.19	0.74	34.45	1.64	7.95	116	728763	9102672	109.90	17.41	92.49	0.64	7.95
43	721823	9101831	31.45	0.42	31.03	1.82	7.99	117	728232	9103502	100.23	6.39	93.84	0.72	7.61
44	722713	9101741	36.65	1.45	35.20	1.75	7.99	118	728400	9103968	92.20	1.18	91.02	0.68	8.17
45	722632	9102803	39.62	1.13	38.49	1.41	7.99	119	730398	9104292	115.83	7.03	108.80	0.96	8.12
46	722177	9103600	47.00	2.35	44.65	0.82	8.24	120	731880	9105222	134.33	3.41	130.92	1.70	
47	724044	9104314	66.00	3.63	62.37	1.12	7.73	121	734243	9106697	142.50	3.90	138.60	1.07	7.97
48	723866	9103470	53.71	3.10	50.61	0.82	7.86	122	730386	9105965	123.73	8.01	115.72	0.98	7.88
49	724510	9102946	53.59	1.25	52.34	1.06	7.95	123	728892	9104706	97.07	1.46	95.61	0.83	
50	725294	9104264	74.67	4.48	70.19	0.80	7.81	124	726989	9104710	83.70	4.65	79.05	0.92	7.49
51	717719	9104637	43.51	23.85	19.66	0.83	8.22	125	726078	9104507	78.83	3.36	75.47	0.94	7.57
52	717040	9102262	21.64	8.80	12.84	-		126	725307	9104700	78.75	7.34	71.41	0.82	7.71
53	719744	9104251	45.65	16.91	28.74	-		127	725414	9105307	91.50	14.54	76.96		
54	718617	9104870	53.14	29.92	23.22	1.08	8.10	128	721716	9105205	63.68	0.00	63.67		
55	720126	9103600	41.57	10.85	30.72	1.10	7.18	129	722938	9106290	80.00	11.48	68.52	1.13	8.45
56	715777	9104424	36.22	16.10	20.12	1.17	7.40	130	723419	9100987	51.13	4.67	46.46	1.43	
57	719700	9102118	24.00	1.17	22.83	1.15	7.20	131	727442	9101768	110.53	24.06	86.47	1.89	8.08
58	717322	9105658	63.00	42.71	20.29	0.54	7.30	132	717611	9099223	6.50	1.91	4.59	1.30	8.36
59	715561	9106008	59.82	35.05	24.77	0.83	9.10	133	716744	9098668	3.64	1.13	2.51	2.22	8.00
60	716352	9101600	14.80	6.80	8.00	1.54	7.20	134	716979	9098371	15.00	1.95	13.05	1.30	7.97
61	721368	9105386	68.00	6.68	61.32	0.55	8.46	135	716201	9098088	1.27	0.43	0.84	2.52	8.73
62	719162	9103618	40.82	15.26	25.56			136	716078	9098679	2.59	1.69	0.90	1.26	
63	717085	9105222	52.00	33.73	18.27			137	715826	9098685	1.90	0.93	0.97	3.78	8.22
64	718552	9101572	19.54	3.13	16.41	1.60	8.00	138	715708	9099459	4.22	2.19	2.03	2.89	
65	715724	9101082	9.88	4.97	4.91	1.42	8.00	139	714721	9099939	4.18	2.48	1.70	2.22	8.00
66	716068	9102738	18.50	2.92	15.58			140	714253	9099969	2.60	2.32	0.28	2.40	
67	716695	9102398	21.03	8.76	12.27	0.78	7.51	141	715176	9101487	12.95	2.69	10.26	1.77	
68	717310	9103670	33.03	14.73	18.30	1.23	7.60	142	716562	9099819	8.91	3.38	5.53	4.02	
69	722861	9104494	64.50	4.31	60.19	0.43	8.65	143	721499	9093076	1.47	0.37	1.10	7.53	9.16
70	715680	9102050	14.90	5.41	9.49	1.13	7.90	144	722338	9092720	5.02	4.16	0.86	2.90	8.30
71	718743	9103484	37.90	14.75	23.15	1.27	7.80	145	723055	9091113	0.97	0.86	0.11	1.86	8.24
72	716176	9099504	4.50	2.80	1.70	1.30	8.00	146	723868	9091366	4.90	1.60	3.30	4.20	7.82
73	716604	9100298	12.20	4.73	7.47	1.54	8.00	147	722819	9092112	4.91	1.40	3.51	6.69	9.04
74	716962	9100488	12.30	5.04	7.26	2.30	7.70								

We also have the approximations made from the hydrogeological study of the valley carried out by the consulting company AMSA, for the company SEDAPAT (Servicio de Agua Potable y Alcantarillado de Trujillo), from that we take the information referring to the impermeable bottom of the aquifer, which we will denote by SIm[8].

With this information from the UTM coordinates, the geographic and phreatic altitude we construct an approximate surface of the geographic and phreatic relief, which we denote as SGe, SFr, respectively, this reconstruction is based on the Quickhull algorithm for convex hull, concept on which $K - means$ algorithm is based[18, 19, 20].

The Quikhull algorithm uses two concepts from geometry, the external orientation of the unit normal to the hyperplane with its origin displaced and the oriented distance. This represents a hyperplane passing through the $n-$points and the normal displaced from the origin. The directed distance from a point to the hyperplane is the inner product of the point and the normal plus its displacement. The hyperplane defines half the space of points that have negative distances from the hyperplane, if the distance to the point is positive then it lies above the hyperplane, the simplification in processing is based on the Grünbaum Beneath-Beyond theorem [18].

2.3. Clustering theory. It is based on the search for grouping elements trying to achieve maximum homogeneity, also refered to in some literature as similarity in each group or cluster of elements.

The clustering theory is mainly based on the grouping techniques, for which it is necessary to have a function of similarity or dissimilarity, depending on the application that is wanted to give, these functions of similarity or dissimilarity are based fundamentally on the concept of metric, for the case of simple applications, but it will not always be a metric, when we are dealing with non-measurable spaces, it will become simply a similarity function, this type of functions can be constructed according to the application that is being given as they appear in [14] or the well-known l_p distances derived from the spaces of *Minkowski*, those of *Manhattan*, the Euclidean heavy distances, etc.

Cluster analysis is a multivariate technique that seeks to group elements, in this work, areas of land, trying to find the greatest homogeneity, in our case the areas with the same level of environmental risk, these risks are based on a series of indicators and coefficients that are inherent to the environment, in this particular case, the underground aquifer, for which use will be made of the properties of the aquifer through algorithms, software and specific models to determine them.

In clustering there are a variety of methods to carry it out among the main ones we have the algorithms based on partitioning, hierarchical and localization.

2.3.1. Partitioning Algorithms. In this type of algorithm, n partitions of the data are constructed, where each partition represents a group or cluster. Each cluster has at least one element and each element belongs to only one group. These methods create an initial partition and iterate until a certain stopping criterion is met. The most popular are: $H-$means, $k-$means, $k-$methods, among others[21].

These methods differ from each other in very few things but always have a common denominator which is the identification of the representative element, in this work we will call it centroid, which in more than one of the methods may or may not coincide with some of the data, in our case the data are not formed by points in a space but by a set of cells or small geographic areas each of which is treated as a constituent cell of the environment in which the physical and environmental properties of the aquifer are immersed.

3. Results. The results obtained from the study were:

- Apply optimization techniques through multi-criteria analysis to make an appropriate decision on the drainage system for the water table phenomenon in the city of Trujillo.
- To present an alternative solution for environmental decision making to attenuate the phenomenon, using multi-criteria analysis and environmental impact studies as a theoretical basis.

Regarding the computational calculation, this work is based on the geographical distribution of the observation wells, we have worked on a digitized plan of the valley, the dimensions of the image of the plan is 6841x4711 pixels.

For the computational calculation we used a personal computer with a 2.3Ghz I7 processor with 4 cores and 12Gb of RAM.

Concerning the software being used, we use GIMP for graphics processing, Matlab and C++ for detecting the shape of the Moche Valley. Finally, the discretization process, clustering and result visualization are performed with Matlab.

Due to the restrictions in the memory capacity of the equipment, the discretization used is of the order of 10%, i.e. Each cell represents an approximate geographic dimension of 471x684 meters. In the implementation of the algorithm, there are options to choose between 100%, 10% and 1 %, depending on the desired accuracy and capacity of the available equipment.

3.1. Approximations of the surfaces: Geographic, Phreatic and Impermeable.. An external analysis of the phenomenon has been developed in which it has been determined that an objective function must be built from the various criteria involved in decision making, which will allow us to iteratively evaluate the partial results of a set of options.

Figure 3.1: Satellite view of the Moche Valley

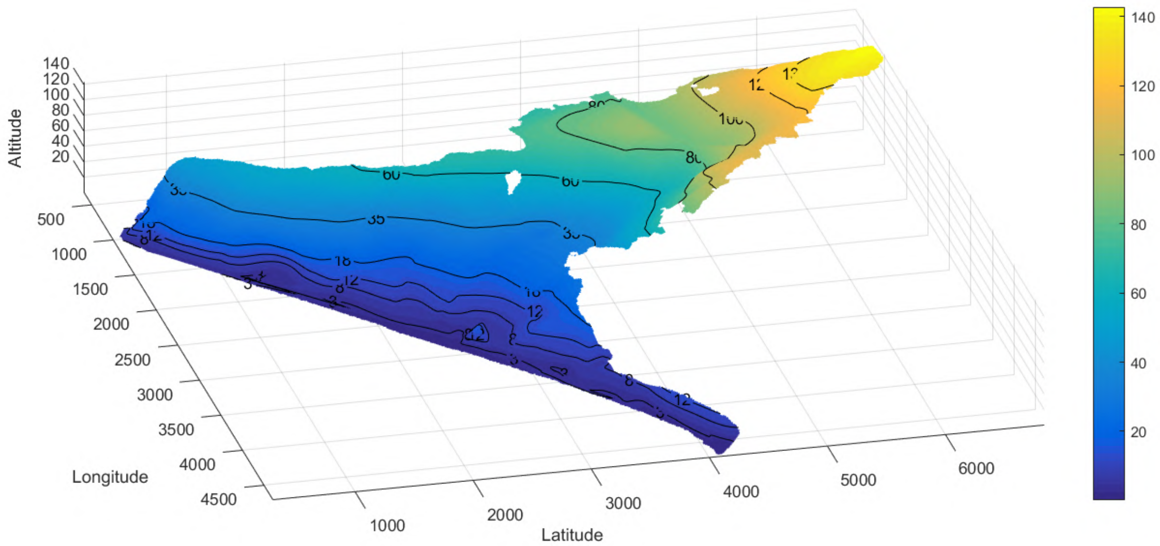

Figure 3.2: Approximate topographic surface of the Moche Valley

As we know the Moche river valley is formed by the entire basin of the same name, but the phenomenon of variation in the water table is presented with great emphasis in the lower basin, this part of the basin occupies almost the entire deltoid of the mouth of the Pacific Ocean, as shown in the figure 3.1, on the East side begins in the place called Cerro Blanco, where we can see a very considerable narrowing of the valley causing it to be practically isolated from the rest of the middle basin, For the modeling purposes, this

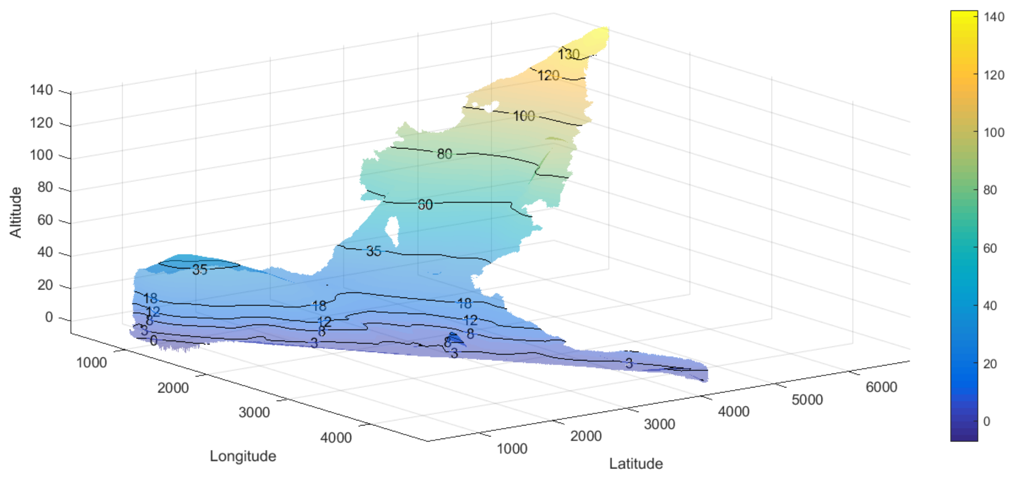

Figure 3.3: Approximate freatic surface of the Moche Valley

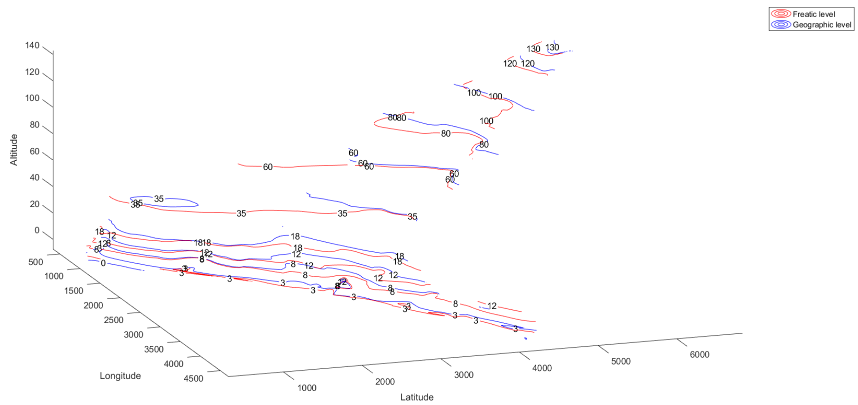

Figure 3.4: Comparison of geographic and phreatic equipotential curves of the Moche Valley

location will be considered as an aquifer boundary with a fixed load represented by the degree of surface runoff from the river, added to the volume of overflow runoff from basin of Santa river, since this is the location of the intake of the different secondary canals, which are part of PECH-CHAVIMOCHIC, which take water from the mother canal, that is the ending of second stage of this irrigation project.

In the northwest sector between the towns of Huanchaco and El Milagro, the delta of the mouth joins with the basin of the Rio Seco, De Leon, Encalada and La Cumbre streams, originating an extension towards the east, as well as an intersection between them in the western part, which runs between the towns of Huanchaquito, El Tropico and the archaeological ruins of CHAN CHAN, northwest of the main palace. For the purposes of this research, the work area is delimited in which the Rio Seco, De Leon, Encalada and La Cumbre streams are not considered to have a load influence, since they do not have any flow until the El Niño phenomenon occurs.[2]

The Figure 3.4 shows the comparison of the equipotential curves of the geographic and phreatic surfaces.

It should be noted that this image corroborates the facts presented in the field. It can be clearly seen how in the lower part of valley, the water table is quite shallow with respect to geographic surface, specifically in the area between the mouth of Moche River to the sea and El Golf housing development.

Another area with a freatic surface is very close to the surface is the area known as Los Huachaques, near Chan Chan archaeological complex.

[2]This phenomenon is an unpredictable threat because it depends on the withdrawal or not of large masses of cold water in the South Pacific ocean currents.

In general, the zones of greatest saturation of the aquifer can be noted near the equipotentials at 3, 8 and 12 meters of altitude, as well as in some places near 18, 35 and 60 meters of altitude.

3.2. Zonification of the Valley. After making an approximation of the geographic and phreatic surfaces, the zoning process of the region under study is carried out using the Algorithm of zoning.

The zoning algorithm is based on the observations of the piezometric levels observed in the network of observation wells of the PECH, from which the phreatic and topographic surface of the valley is approximated, using the Quick-Hull algorithm, described in section 2.2, obtaining as a result the surfaces shown in figures 3.2 and 3.3 respectively, in which the region to be worked is shown, completely delimited.

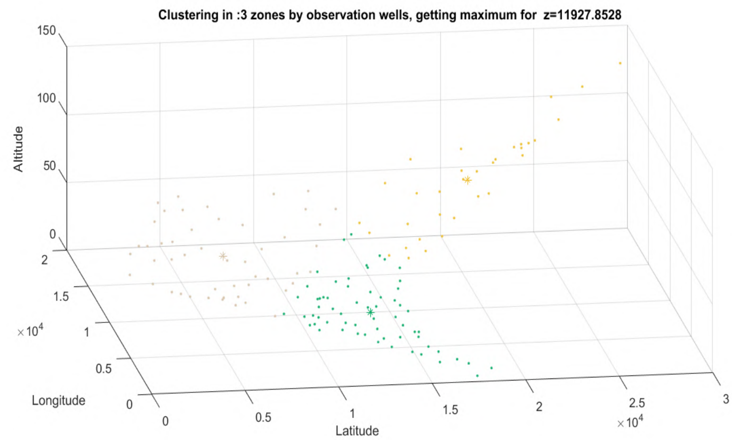

Figure 3.5: Zoning into 3 cluster of the lower basin of the Moche Valley, obtained into 12 iteration.

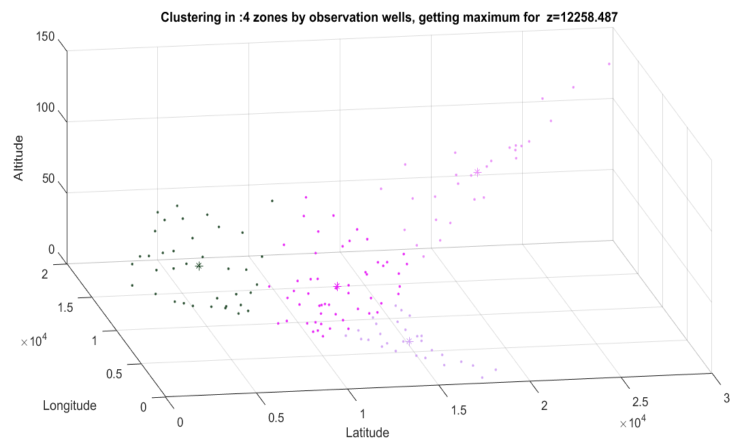

Figure 3.6: Zoning into 4 cluster of the lower basin of the Moche Valley, obtained into 9 iteration.

The results obtained at this stage have only been preliminary because the data corresponding to the hydrogeological properties of the aquifer have not yet been determined and will be the subject of a future work.

In the calculation of the similarity and objective functions, for the application of the $k-$means algorithm we used the data from the observation wells, presented in the table (2.1) and the environmental management coefficients pending calculation, have been considered as constants.

The implementation of the zoning algorithm has the versatility of choice by the user, allowing the user to enter at run time the number of N zones into which it is desired to partition.

When considering this alternative, it allows the appearance, in the formula (3.2), of two sets of indexes I_d and J_d, which are part of a previous assignment and the final assignment, for the number of observation points assigned in each iterative process.

The figures 3.5, 3.6, 3.7, 3.8 and 3.9, show the result of the zoning of the valley under study, in each one of them, shows the number of iterations that have been necessary to perform, as well as the corresponding value of the objective function Z to each case.

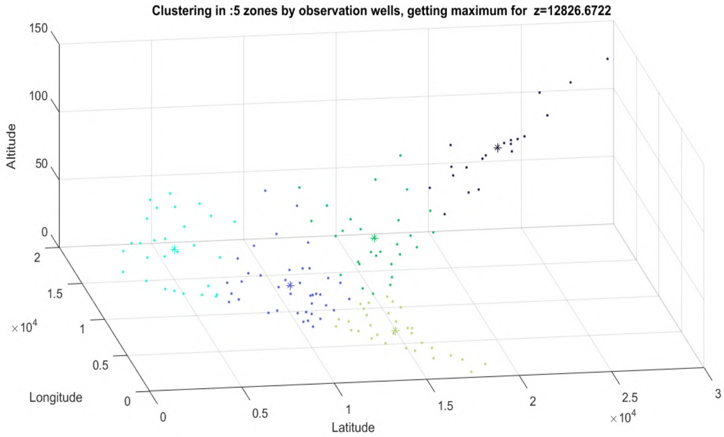

Figure 3.7: Zoning into 5 cluster of the lower basin of the Moche Valley, obtained into 8 iteration .

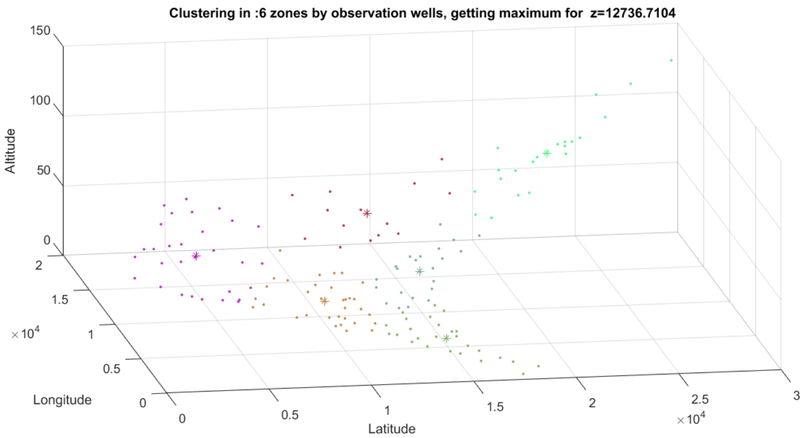

Figure 3.8: Zoning into 6 cluster of the lower basin of the Moche Valley, obtained into 6 iteration .

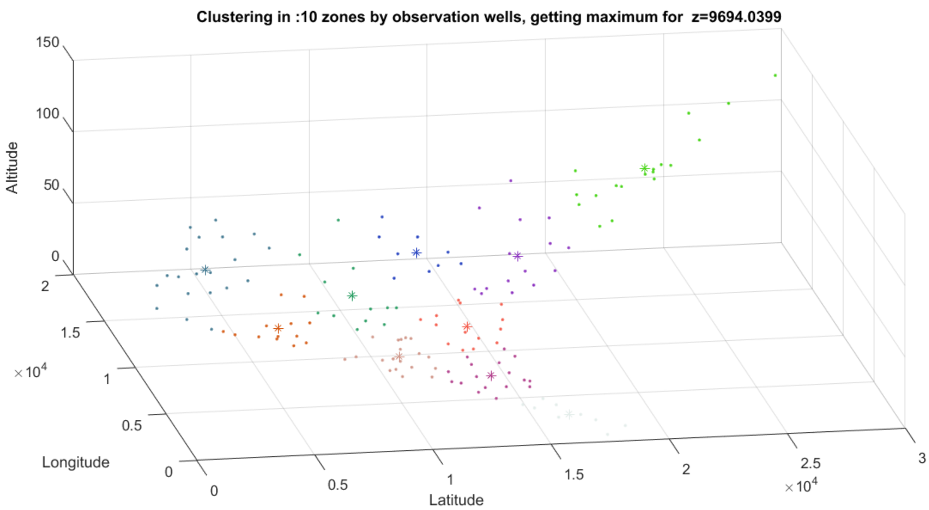

Figure 3.9: Zoning into 10 cluster of the lower basin of the Moche Valley, obtained into 7 iteration .

3.3. Important functions. The algorithm used for the initial partition is based on the Euclidean distance which is associated with the $K-means$ algorithm whose details can be found on page 40 of [21], the calculation of the centroid for the $c-th$ cluster necessary for the application of this algorithm is calculated using the following formula

$$(\overline{x}\ ,\ \overline{y}\ ,\ \overline{SGe}\ ,\ \overline{SFr})_C = \frac{1}{n}\sum_{i=1}^{n}(x_i\ ,\ y_i\ ,\ SGe_i\ ,\ SFr_i)_C\ , \tag{3.1}$$

where:

C : cluster number or area to be drained,
x_i : the east latitude of the point (cell),
y_i : the north longitude of the point (cell),
SGe_i : the altitude of the geographic surface,
SFr_i : the altitude of the phreatic surface.

In the initial calculation the depth of the impermeable layer is not used, since it will not have much relevance in the decision making process.

To evaluate the similarity of a given point (cell) it is necessary to build a function of similarity between points (cells), so that at the time of reassigning a cell to a particular cluster, it is the most appropriate.

$$S(x, y, SGe, SFr)_C = W_Q Q^2 + W_R R^2 + W_S S^2, \tag{3.2}$$

where:

W_Q^C : Represents the weight for the depth of the phreatic surface.
W_R^C : Represents the weight for the altitude of both the geographic surface and the phreatic surface.
W_S^C : Represents the weight of the spatial variables (UTM).

$$Q = (SGe - \overline{SGe_C} + \overline{SFr_C} - SFr)^2$$

$$R = \frac{1}{4}(SGe + SFr - \overline{SGe_C} - \overline{SFr_C})^2$$

$$S = (x - \overline{x_C})^2 + (y - \overline{y_C})^2.$$

Having defined these two functions, we will also construct the multi-objective function that will allow us to evaluate the optimality of the partitioning of the region under study, this function is given by:

$$Z = \sum_{C=1}^{N_C} \frac{Kh_C}{N_C}\left(\sum_{\substack{i \in I_d \\ j \in J_d}} W_Q Q^2 + W_R R^2 + W_S S^2\right) \tag{3.3}$$

where: W_Q, Q, W_R, R, W_S, and S have the same values of the equation (3.2), the KhC coefficient implies the different properties of the aquifer, the study of which corresponds to other research previously carried out[22].

The most important properties to consider for the calculation are as follows:
K_C : Permeability or Hydraulic Conductivity associated with the medium that varies between zones, for our case we will have to associate a permeability constant,
I_C : Hydraulic gradient,
$(I_0)_C$: Natural gradient,
ETr_C : Evapotranspiration coefficient,
CDr_C : Existing drainage coefficient,
RIn_C : Recharge of the infiltration aquifer through irrigation,

REs_C : Recharge from runoff of tributaries,
Aci_C : Average acidity of groundwater,
Sal_C : Degree of water salinity,
Imp_C : Thickness of aquifer[3].

Having then the objective function we can apply the main algorithm zoning as follows.

3.4. Main Algorithm. In this section we detail the main algorithm of zoning that will allow the optimal partitioning of the study area, which will reflect the different environmental and socio-economic decision criteria that influence it.

This algorithm will be see at Appendix A

As a result of the exchange of opinions with specialists in the areas involved, such as geology, hydrology, geosciences, fluids, decision making, environmentalism, optimization, etc. It was determined that this algorithm can be adjusted to the different criteria that each specialist may consider important and thus achieve an adequate reconciliation of the parties so that the environmental management of the basin is effective.

4. Discussion of Results. In this section we will discuss the main algorithm that allows the zoning of the study area, this algorithm is based on the clustering theory, which uses as a basis the *k-Means* algorithm which has been enriched with the construction of the similarity function, that is based on the smallest dispersion of the data associated with this function.

The main Algorithm of zoning, is applied iteratively, starting with a preliminary distribution of all the cells of the partition, assigned to each cluster(zone)

Once a first distribution has been obtained in an equitable way, keeping the consecutivity of the cells components of the area, the iterative process begins to determine the best adjustment and distribution of the cells into each zone, this adjustment is carried out by evaluating the similarity function and then the best distribution is evaluated by conputing the objective function, which allows to optimize the reasingn of the cells to a zone of greater similarity.

The multi-objective function (3.3), involves different types of environmental valuations that will allow adjusting the zoning according to different criteria that will help to make better decisions.

In the formulation of the main algorithm of zoning, several parameters, difficult to obtain as real data, have been included, so when trying to make the computational simulation, using the objective function, many of them will have to be restricted.

5. Conclusions . From this research work and its restrictions, we can arrive the following conclusions.

1. The clustering theory based on the application of the multiobjective function given by the equation (3.3), plays a very important role in the conception of zoning, which can then be conceived as an area subject to specific studies for environmental management of the underground aquifer.
2. In the construction of the multi-objective function given by the equation (3.3), a very wide range of parameters of different types can be involved, in such a way that it can be applied when there are different priorities, that can be technical as well as socio-cultural and/or environmental.
3. In the construction of the similarity function defined in (3.2), the variables x and y are considered as if they were points but in reality they are the centers of the cells, of the partition applied to the area of work, and its precision will depend on the size of the partition of the region.
4. Several parameters are included in the construction of the main Algorithm, which are not the object of calculation in this work, however, there is a diversity of methodology to obtain them, such as the specific case of the evapotranspiration coefficient or the basic parameters of the underground aquifer, which can be calculated using specialized software.

ORCID and License

Julio César Peralta Castañeda https://orcid.org/0000-0003-0911-8403

References

[1] Municipalidad Provincial de Trujillo. Acciones para el Desarrollo Territorial Sostenible de Trujillo Metropolitano. Trujillo-Perú: Municipalidad Provincial de Trujillo, Agenda 21 Trujillo; 2003.

[3] Considered as an approximation due to lack of information

[2] Consejo Transitorio de Administración Regional CTAR. Plan Concertado de Desarrollo Departamental-La Libertad. CTAR; 2002.
[3] Municipalidad Provincial de Trujillo. Atlas Ambiental De La Ciudad De Trujillo. Trujillo-Perú: Municipalidad Provincial de Trujillo; 2002.
[4] Bear Jacob. Dynamics of Fluids in Porous Media. Israel: Dover Publications; 1972.
[5] Chen Z , Ewing R E. Fully discrete finite element analysis of multiphase flow in groundwater hydrology. Numerical Analysis. 1977;Vol. 34(6):Págs. : 2228-53.
[6] Ewing Richard E. Mathematical modeling and simulation for applications of fluid flow in porous media. Texas A&M University, College Station TX 77843-3403: Institute for Scientific Computation; 1997.
[7] Ewing Richard E. Numerical simulation of the multiphase flow of contaminants in porous media. Texas A&M University, College Station TX 77843-3403: Institute for Scientific Computation; 1997.
[8] Aguirre Morales A, Quintana J. Estudio Hidrogeológico Abastecimiento de Agua Potable a la Ciudad de Trujillo por Medio de Pozos. Boletín: Sociedad Geológica del Perú. Setiembre 1983;No.71:Págs. 241-53.
[9] Ministerio de la Presidencia. Información de profundidad niveles freáticos 1995-2001 Valle Moche. Trujillo: Instituto Nacional de Desarrollo, Proyecto especial CHAVIMOCHIC; 2001.
[10] Andeberg Michel R. Cluster Analysis for Applications. 111 Fifth Avenue, Now York, New York 10003: Academic Press New York.; 1973.
[11] Everitt G O. Cluster Analysis. 48 Charles Street, London WIX8AH: Heinemann Educational Books Ltda.; 1974.
[12] Hartigan J A. Clustering Algorithms. New York: Wiley Series in Probability and Mathematical Statistics; 1975.
[13] Spaeth Helmuth. Cluster Analysis for data reduction and classification of objetcts. Market Cross House, Cooper Street, Chicherster, West Sussex, PO1991EB, England: Ellis Horwood Limited; 1980.
[14] Theodoridis S, Koutroumbas K. Pattern Recognition. 525-B street, Suite 1900, San Diego, CA 92101-4495. USA: Academic Press; 1999.
[15] Love R F, Morris J G, Wesolowsky G O. Facilities Location Model & Methods. Amsterdam, The Netherlands: Elsevier Science Publishing Co., Inc.; 1988.
[16] Peralta Castañeda Julio C, León Navarro Ronald, Carrasco Silva Anselmo. Un modelo matemático del fenómeno de la napa freática en la ciudad de Trujillo. Trujillo-Perú: Departamento de Matemáticas-Universidad Nacional de Trujillo; 2005.
[17] Custodio E, Llamas M R. Hidrología Subterránea. Platón 26, Barcelona-España: Ediciones Omega S.A.; 1983.
[18] Barber C B, Dobkin D P, and Huhdanpaa H T. The Quickhull Algorithm for Convex Hulls. ACM Transactions on Mathematical Software. Dic 1996;Vol. 22 No.4:Págs. 469-83.
[19] Sandwell, David T. Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data. Geophysical Research Letters. 1987;2:Págs. 139-42.
[20] Watson, David E . Contouring: A Guide to the Analysis and Display of Spatial Data. Tarrytown NY: Pergamon (Elsevier Science, Inc.); 1992.
[21] Peralta Castañeda Julio C. Clustering Algortims in Location Planning. Kaiserslatuern Deutchland: Industrial Mathematik Univertät Kaiserslautern; 1999.
[22] Peralta Castañeda Julio C, Hernandez V Armando, Morejon V Genobebo. Una aplicación de los Grafos Bicromáticos en el modelamiento matemático del drenaje de acuíferos subterráneos. Ciencia y Tecnología. 2012;8:Págs. : 47-55.

Appendix A. Algorithm of zoning.

Input: UTM coordinates, topographic and water table elevations for each observation well.

Output: Zoning the study area.

Step No 1 Initialize weights W_Q, W_R and W_S

Step No 2 Create matrices X_{I_d} and Y_{J_d}, The result of the two-dimensional partitioning of the UTM coordinates of the observation wells, with this mesh SGe is approximated as the approximate surface of the geographic relief and SFr represents the approximation of the phreatic surface.

Step No 3 Initializing variables in

- ***Step No 3.1*** Number of iterations $Ite = 1$
- ***Step No 3.2*** Enter the maximum number of iterations from console MIt
- ***Step No 3.3*** Enter tolerance value Tol
- ***Step No 3.4*** $N = 0$

Step No 4 Make an initial assignment of each points (cells) to one of the $C-$clusters and store it in the cluster matrix.

Step No 5 Calculate the initial centroid of each $C-$cluster using equation:

$$(\overline{x}\ ,\ \overline{y}\ ,\ \overline{SGe}\ ,\ \overline{SFr})_C = \frac{1}{n}\sum_{i=1}^{n}(x_i\ ,\ y_i\ ,\ SGe_i\ ,\ SFr_i)_C$$

Step No 6 For each cluster C

- ***Step No 6.1*** Calculate the value of similarity function

$$S(x,y,SGe,SFr)_C = W_Q Q^2 + W_R R^2 + W_S S^2$$

for each component cell in the C-cluster

- ***Step No 6.2*** Assign the value of the similarity function to the respective matrix element FSi_C
- ***Step No 6.3*** Stores the FSi_C array on disk as a text file.

Step No 7 Assign to the matrix CL, of the same dimension as the matrix X, the value of zero to each element.

Step No 8 For each point in the region do

Step No 8.1 Determine which centroid is most similar.

Step No 8.2 Assign to the element of the CL matrix the cluster number to which the centroid belongs, to which the point is most similar (closest).

Step No 9 For each point of each cluster C check: If it does not meet the ***adjacency condition***.[4] execute the algorithm of **reallocation to new cluster** and back to Step No. 5.

Step No 10 Initialize the values of Kh_C[5]

Step No 11 Compute the objetive function

$$Z = \sum_{C=1}^{N_C} \frac{Kh_C}{N_C} \left(\sum_{\substack{i \in I_d \\ j \in J_d}} W_Q Q^2 + W_R R^2 + W_S S^2 \right)$$

Step No 12 If $Ite \geq MIt$ goto Step No. 16, with message that the number of iterations was exceeded, otherwise do $Ite = Ite + 1$

Step No 13 If the value of the objective function becomes a divergent series, go to Step No. 16, with message of non-convergence of the algorithm, if not, check: if the tolerance was reached (Tol), go to Step No. 15.

Step No 14 Execute the algorithm of **point exchange** between contiguous clusters according to the similarity function calculated in Step No. 6 and return to Step No. 5.

Step No 15 Show output data

Step No 16 End

[4]Specific Algorithm

[5]These coefficients can be included in the calculation, depending on their feasibility.