International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 08, August 2022)
Manuscript Received: 07 June 2022, Received in Revised form: 09 July 2022, Accepted: 01 August 2022 DOI: 10.46338/ijetae0822_03

Applying Grey Clustering Method and Pearson Correlation to Assess Water Quality

Alexi Delgado[1], Raúl Vidal[2], Isaac Jayme[3], Christian Kaiser[4], Belén Sánchez[5], Janneth Lema[6], Fabrizio Llave[7], Laberiano Andrade-Arenas[8]
[1, 2, 3, 4, 5, 6, 7]Mining Engineering Section, Pontificia Universidad Católica del Perú, Lima-Perú
[8]Universidad de Ciencias y Humanidades, Lima-Perú

***Abstract*—The Mantaro River travels through Junín, Ayacucho, and Huancavelica departments,in Peru, and receives not only domestic discharges from the population but also tailings fields and dumps containing lead, silver, copper, and zinc from mining companies. To be able to assess the water quality, the grey clustering method was applied, using Center-point Triangular Whitenization Weight Functions (CTWF) for this purpose. As well, Pearson correlation between parameters was used to understand the dynamics of pollution. In the watershed of the Mantaro river the results of the Monitoring of the Surface Water published in 2018 by the Mantaro Water Administrative Authority were compared to the Peruvian Environmental Quality Standards (ECA) Category 1, subcategory A parameters. The results obtained suggest that the main driver of pollution in the Mantaro river is not mining, but domestic waste in landfills and probably agriculture. Finally, this study shows responsible environmental management by mining companies, besides, could be helpful for regional and local authorities of Peru in making decisions to improve the management of the Mantaro river watershed and make the population aware of the sustainable use of water.**

***Keywords*—Grey clustering, Pearson correlation, Water quality assessment.**

I. INTRODUCTION

Water pollution is one of the main impacts of mining activity. The main sources are Acid Mine Drainage (AMD), heavy metal leaching, cyanidation among other chemical processes [1], [2]. For this reason, it is the cause of different social conflicts around the world and Peru is not the exception [3].

The Mantaro river, also called Jatunmayo, belongs to the hydrographic system of the Atlantic Ocean slope, traveling around 725 km through Junín, Ayacucho, and Huancavelica departments. In the upper watershed, minerals such as lead, silver, copper, and zinc are exploited.

Thus, the tributaries of the Mantaro river receive not only domestic discharges from the population but also tailings fields and dumps from mining companies such as Volcan, Brocal, Huarón, Morococha, and others. Failure to comply with current environmental regulations could have a negative impact on the physicochemical properties of the river watershed, causing harm to the inhabitants and damage to the aquatic ecosystem [4].

The length of the Mantaro river is 735 km. Its watershed includes the regions of Pasco, Junín, Huancavelica, and Ayacucho and occupies a total area of 34 546.51 km2 which is divided into Mantaro Superior (upper), Mantaro Medio (middle) and Mantaro Inferior (lower).

The present investigation has the objective to provide an assessment of the quality of the surface water in the Mantaro river. Additionally, the study seeks to find relationships between water quality parameters and overall water quality to provide a better understanding of the dynamics behind pollution in the Mantaro river that can be applied to similar water systems impacted by different human activities. To fulfill these objectives, we will use the Grey Clustering method with Center-point Triangular Whitenization Weight Functions (CTWF) [5], [6] to evaluate water quality and Pearson Correlation to establish correlation between parameters.

The structure of this investigation will be the following: the methodology will be explained in the section III, the case of study and the results are placed on the section IV and V, respectively.

II. LITERATURE REVIEW

A. Impact of mining activities in rivers

Effimov et. al [7] show that mining activities are the main driver of pollution in Khibiny, Russia. The impact of the discharge of residual waters leads to the accumulation of sediments saturated with heavy metals at river mouths.

The chemical composition changes and the turbidity of the water along the rivers increases determined by the volume of wastewater discharged from the industrial site. Water purification helps reduce turbidity 100 times. Nevertheless, most of the suspended particles remain as the finer fractions. The existing treatment system is not sufficient to mitigate the polluting loads. Heavy metal concentrations gradually decrease because of the processes of "complexation" with organic matter and sedimentation of particles but continue to significantly exceed those of the base. About 50-60% of the chemical elements are transported in suspended sediments. Therefore, an effective monitoring system and the construction of additional sedimentation ponds with physico-chemical methods to reduce turbidity needs to be developed.

B. Presence of heavy metals and metalloids in the Mantaro river watershed

Custodio et. al [8] have evaluated the Mantaro river watershed's content of heavy metals. The concentrations of Cu, Fe, Pb, Zn and As were determined by flame atomic absorption spectrophotometry to assess human risk. Likewise, the concentration of heavy metals and arsenic varies according to the sector of the rivers evaluated. The data found show a high concentration of Pb and As in the water, for which an urgent control and reduction of containment levels is required. The risk assessment for humans was performed by exposure to heavy metals with arsenic by ingestion and by the dermal route, strictly using standard methods (USEPA). In addition, the data obtained can be used to estimate cancer risk and formulate public health policies and programs. The high concentration of these metals reveals that the Mantaro River watershed is still a sink for mine waste. Multistatistical analysis showed that there is a significant correlation between Fe, Pb, and Zn and a poor one between the other metals. Hierarchal clustering analysis distinguished three main water groups according to their metal content: Cu, As and Fe-Pb-As. The study concludes that a management of mining toxic metal discharge is imperative.

C. Using grey clustering to evaluate nitrogen water pollution

Temino-Boes and colleagues [6] proposed a new method based on grey clustering which classifies monitoring sites according to their level of nitrogen pollution with few data using the entropy-weight method.

The authors developed two indexes: the grey nitrogen management priority index (GNPM) and the Grey Land Use Pollution Index (GLUP) to determine the amount of nitrogen pollution based on the land use. The study applied this methodology in 8 estuaries of the Southern Gulf of Mexico and other near-coast ecosystems affected by the large amount of nitrogen pollution in the biome.

D. Grey clustering on a highly contaminated river system

The Rimac river watershed is one of the most important rivers in Lima, because it supplies drinking water for the population. However, since some years ago, this river has been affected by different economic activities. According to Delgado and colleagues [5], the following Rimac river affluents: Aruri River, Rimac River, Mayo River, Santa River, and Blanco River were classified as uncontaminated within the Prati scale; nevertheless, Blanco River was more vulnerable to be contaminated. The research work based their result on the center-point triangular whitenization weight functions (CTWF) method, considering parameters of water quality such as dissolved oxygen, biochemical and chemical oxygen demand, suspended solids, nitrates, and nitrites.

E. Using correlation statistics on water quality

Kothari and colleagues [9] conducted a study based on biological and hydrochemical parameters for the Water Quality Index calculation in six different districts of Uttarakhand, India. Among the used parameters there is pH, alkalinity, haze, metal content such as Fe, Na, Mg, As, sulfates, dissolved solids, and others. Also, likewise, some bacteriological parameters were used such as total content of coliforms and fecal coliforms that, in some cases, exceed the BIS standards. Correlation statistical analysis between WQI and other parameters shows up a linear relationship and a high effect of iron content, total coliform content and fecal coliforms over Water Quality.

F. Correlation between water quality parameters in a mining zone

A study focused on an area surrounding mining operation was conducted by Tiwari and colleagues [10] around Sukita Chromite Valley, in Odisha, India. The values measured in both superficial and groundwater were temperature, pH, dissolved oxygen, biochemical oxygen demand, total suspended solids, total dissolved solids, hardness, chloride, sulphate, alkalinity, sulphate, chromium (VI) and total chromium.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 08, August 2022)

Significant correlations between both chromium parameters, sulphate and pH were found. Chromite content exceeded Indian environmental standards on many monitoring points and poor water quality related to mine waste was observed in points measured before treatment facilities. The content decreased with distance and achieved permissible concentrations in populated areas.

III. METHODOLOGY

A. Grey Clustering

The methodology applied to assess the water quality of the Mantaro River was grey clustering as it is a common and useful method for evaluating multiple environmental and non-environmental aspects [6], [11]. Prior to the calculations there must be a defined amount of "m" study objects (i=1,2,… m),"n" criteria (j=1,2,…n), and "s" grey classes (k=1,2,… s). These are the following steps to calculate classification values for each monitoring point [12]

Step 1: Central point determination

For each class there is a standard corresponding to each criterion represented in a matrix by the value λjk, known as the central point. Each λjk is taken from the normative, whether it is the average of the end members of an interval or the maximum value.

Step 2: Adimensioning

The adimensioned value for each criterion is determined to enable comparison between criteria with different measure units with (1), where A represents the adimensioned value and D the original one. This operation is made to each value from the standards (λjk) and dataset (xij) matrixes.

$$A = D / \sum_{k=1}^{s} \frac{\lambda_{jk}}{s} (1)$$

Step 3: Triangular functions (CTWF)

Equations (2) – (4) represent the way to determine the triangular function for the values of the grey classes, where "x" is the adimensioned value from the study objects.

$$f_j^1(x_{ij}) = \begin{cases} 1 & x \in [0, \lambda_j^1] \\ (\lambda_j^2 - x)/(\lambda_j^2 - \lambda_j^1) & x \in]\lambda_j^1; \lambda_j^2[\\ 0 & x \geq \lambda_j^2 \end{cases} \quad (2)$$

$$f_j^k(x_{ij}) = \begin{cases} 0 & x \notin]\lambda_j^{k-1}; \lambda_j^{k+1}[\\ (x - \lambda_j^{k-1})/(\lambda_j^k - \lambda_j^{k-1}) & x \in]\lambda_j^{k-1}; \lambda_j^k] \\ (\lambda_j^{k+1} - x)/(\lambda_j^{k+1} - \lambda_j^k) & x \in]\lambda_j^k; \lambda_j^{k+1}[\end{cases} \quad (3)$$

$$f_j^s(x_{ij}) = \begin{cases} 0 & x \leq \lambda_j^{s-1} \\ (x - \lambda_j^{s-1})/(\lambda_j^s - \lambda_j^{s-1}) & x \in]\lambda_j^{s-1}; \lambda_j^s[\\ 1 & x \geq \lambda_j^s \end{cases} \quad (4)$$

Step 4: Determination of the weight for each criterion

In this study the harmonic mean will be used, so an objective weight for each criterion could be determined. The weight ηjk is determined through (5).

$$\eta_j^k = \frac{1/\lambda_j^k}{\sum_{j=1}^{m} 1/\lambda_j^k} \quad (5)$$

Step 5: Determination of the clustering coefficient

The clustering coefficients σik are obtained applying (6).

$$\sigma_i^k = \sum_{j=1}^{n} f_j^k(x_{ij}).\eta_j^k \quad (6)$$

Step 6: Results using the max. clustering coefficient

The highest σik within each study object determines the class to which they correspond and allows for comparison within each class

B. Pearson Correlation Test

Additional to the grey clustering, the r-Pearson linear correlation coefficient between each parameter and overall water quality index, determined by the clustering coefficient belonging to the first grey class, was obtained to achieve a better understanding of the dynamics behind water pollution and establishing predictive relationships. Though specific correlation significance intervals are mostly subjective and depend on the size of the data and the science branch where it is applied, in this study the intervals were the ones used by [13] where a strong absolute correlation is higher than 0.76; a moderate correlation is higher than 0.51, a weak correlation is higher than 0.26 and below that point correlation can be neglected.

IV. CASE OF STUDY

A. Definition of Study Objects

For the assessment of the quality of the waters of the Mantaro river watershed, information was collected from 25 monitoring points. The values were obtained from the "Result of the Participatory Monitoring of Surface Water Quality in the Scope of the Mantaro River Watershed" published in 2018 by the Mantaro Water Administrative Authority, including the Water Local Authorities from Pasco, Mantaro, Huancavelica and Ayacucho, which are decentralized agencies of the National Water Authority [14].

The monitoring points were placed in strategic sites like connections with tributaries, dumps, cities, or hydroelectric plants, both before and after.

The sampling points considered are detailed in Table I and located on a map (shown in Fig. 1) next to operating and closed mines.

Table1
Monitoring points

Name	Coordinates		Description
	X	***Y***	
RMant1	360371	8790460	Downstream of Upamayo dam
RMant2	371425	8757001	After intersection with Coricancha river
RMant3	396794	8730179	After Paccha-La Oroya population
RMant4	399587	8726799	Before STRAKRAFF hidroelectric discharge
RMant5	400536	8726799	After STRAKRAFF hidroelectric discharge
RMant6	401694	8726179	Before Doe Run refinery
RMant7	402348	8724653	After Doe Run refinery
RMant8	411333	8712609	Before Huari river
RMant9	412036	8712352	After Huari river
RMant10	422743	8699253	Before Pachacayo river
RMant11	422831	8698467	After Pachacayo river
RMant12	446138	8694663	Upstream of CIMIRM and Plan MERISS
RMant13	464563	8680918	Before Achamayo river
RMant14	473270	8668692	Before Pilcomayo dump
RMant15	473267	8668256	After Pilcomayo dump and before El Tambo "Mejorada" dump
RMant16	473874	8667464	Downstream of Breña bridge
RMant17	474284	8665745	Downstreamof “Agua de las vírgenes” dump
RMant18	474450	8665373	Downstream of Shullcas river
RMant19	473680	8654812	Huancayo exit
RMant21	508134	8614519	Downstream of “Mejorada”
RMant22	544540	8597918	Downstream of Anco residual waters treatment facility
RMant23	569148	8607566	Downstream of San Pedro de Coris (ex Cobriza)
RMant24	486704	8631353	Near Vilca river
RMant28	388551	8736962	Before arsenic trioxide deposit
RMant29	389261	8737309	After arsenic trioxide deposit

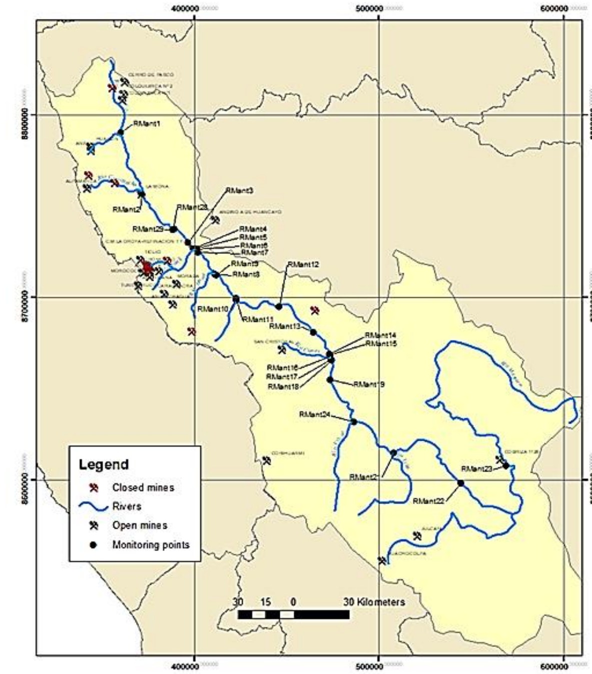

Fig. 1. Map of the sampling points and mining projects of the Mantaro River watershed.

B. Definition of Assessment Criteria

The assessment criteria used on this study belong to the Peruvian water quality standards ECA. The Peruvian National Water Authority (ANA) classifies the Mantaro river in category 3, which means its waters are destined to vegetable irrigation and livestock drink [15] and the parameters measured in the monitoring points belong to this category. However, as this classification has only two possible grey classes, the monitoring points were evaluated according to the subcategory A from category 1 (Table II), that evaluates the required treatment to provide drinking water to the public. Therefore, only the parameters in common for both categories were considered. Also, parameters with lacking requirements for some classes were not evaluated.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 08, August 2022)

TABLE 2
WATER QUALITY CRITERIA FROM ECA CATEGORY 1, SUBCATEGORY A

Criterion	Units	Notation
Oils and fats	mg/L	C_1
Dissolved oxygen	mg/L	C_2
pH	0-14	C_3
Thermotolerant coliforms	NMP/100 mL	C_4
Chemical oxygen demand	mg/L	C_5
Biochemical oxygen demand	mg/L	C_6
Nitrates	mg NO_3-N/L	C_7
Chlorides	mg/L	C_8
Aluminum	mg/L	C_9
Arsenic	mg/L	C_{10}
Beryllium	mg/L	C_{11}
Boron	mg/L	C_{12}
Cadmium	mg/L	C_{13}
Copper	mg/L	C_{14}
Chromium	mg/L	C_{15}
Iron	mg/L	C_{16}
Mercury	mg/L	C_{17}
Manganese	mg/L	C_{18}
Lead	mg/L	C_{19}
Zinc	mg/L	C_{20}
Selenium	mg/L	C_{21}

C. Definition of Grey Classes

The subcategory A defines three different classes according to the purification methods required: A1 (disinfection), A2 (conventional treatment), A3 (advanced treatment) [15]. The maximum standard values, minimum (for dissolved oxygen) and ranges (for pH) are shown in Table III.

TABLE 3
WATER QUALITY REQUIREMENTS FROM ECA

Criterion	A1	A2	A3
C_1	0,5	1,7	1,7
C_2	6	5	4
C_3	6,5 - 8,5	5,5 - 9,0	5,5 - 9,0
C_4	20	2000	20000
C_5	10	20	30
C_6	3	5	10
C_7	50	50	50
C_8	250	250	250
C_9	0,9	5	5
C_{10}	0,01	0,01	0,15
C_{11}	0,012	0,04	0,1
C_{12}	2,4	2,4	2,4
C_{13}	0,003	0,005	0,01
C_{14}	2	2	2
C_{15}	0,05	0,05	0,05
C_{16}	0,3	1	5
C_{17}	0,001	0,002	0,002
C_{18}	0,4	0,4	0,5
C_{19}	0,01	0,05	0,05
C_{20}	3	5	5
C_{21}	0,04	0,04	0,05

D. Calculations using Grey Clustering

Step 1: Central point determination

The first step was to build a new water quality classification out of subcategory 1-A with a unique value for each class and criterion. For pH, the maximum was chosen due to the alkalinity of the Mantaro River. When two classes in a criterion have the same values, the A2 central point was obtained as the arithmetic average between the extreme conditions. For criteria with three equal values, the minimum or maximum (depending on the case) were obtained from ECA category 3 for vegetable irrigation and A2 was the average of both categories. The standards used are shown in Table IV. Data with (*) are the mean of both extremes, data with (**) belong to category 3.

TABLE 4
ECA WATER QUALITY CENTRAL POINTS

Criterion	A1 (λ_1)	A2 (λ_2)	A3 (λ_3)
C_1	0,5	1,1*	1,7
C_2	6	5	4
C_3	8,5	8,75*	9,0
C_4	20	2000	20000
C_5	10	20	30
C_6	3	5	10
C_7	50	70*	90**
C_8	250	375*	500**
C_9	0,9	2,95*	5
C_{10}	0,01	0,08*	0,15
C_{11}	0,012	0,04	0,1
C_{12}	1**	1,7*	2,4
C_{13}	0,003	0,005	0,01
C_{14}	0,2**	1,1*	2
C_{15}	0,05**	0,075*	0,1
C_{16}	0,3	1	5
C_{17}	0,001	0,0015	0,002
C_{18}	0,4	0,45	0,5
C_{19}	0,01	0,03*	0,05
C_{20}	3	4	5
C_{21}	0,04	0,045*	0,05

Step 2:Adimensioning

The non-dimensioned standard values for each parameter were determined through (1). These values are presented in Table V.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 08, August 2022)

TABLE 5
NON-DIMENSIONAL STANDARD VALUES FROM ECA

Criterion	A1 (λ_1)	A2 (λ_2)	A3 (λ_3)
C_1	0,4545	1	1,545
C_2	0,8	1	1,2
C_3	0,9714	1	1,0286
C_4	0,0027	0,2725	2,7248
C_5	0,5	1	1,5
C_6	0,5	0,8333	1,6667
C_7	0,7143	1	1,2857
C_8	0,6667	1	1,3333
C_9	0,30508	1	1,6949
C_{10}	0,125	1	1,875
C_{11}	0,2368	0,7895	1,9737
C_{12}	0,5882	1	1,4118
C_{13}	0,5	0,8333	1,6667
C_{14}	0,1818	1	1,8182
C_{15}	0,6667	1	1,3333
C_{16}	0,1429	0,4762	2,3810
C_{17}	0,6667	1	1,3333
C_{18}	0,8889	1	1,1111
C_{19}	0,3333	1	1,6667
C_{20}	0,75	1	1,25
C_{21}	0,8889	1	1,1111

Similarly, the adimensioned results from monitoring points are presented in Table VI, using the first seven criteria as an example. Points RMant21 to RMant24 lacked information about chloride content and COD.

TABLE 6
NON-DIMENSIONED MONITORING DATA IN THE CASE STUDY

Points	C_1	C_2	C_3	C_4	C_5	C_6	C_7
RMant1	0,4545	1,294	0,9829	0,0015	0,95	0,3333	0,0011
RMant2	0,4545	1,35	0,9863	0,0108	1,05	0,1667	0,0017
RMant3	0,4545	1,46	0,9714	0,6267	0,65	0,1667	0,0018
RMant4	0,4545	1,436	0,9714	0,1076	0,45	0,1667	0,0019
RMant5	0,4545	1,44	0,9714	0,1771	0,05	0,1667	0,0025
RMant6	0,4545	1,434	0,9829	0,3270	0,5	0,1667	0,0026
RMant7	0,4545	1,42	0,9829	0,3270	0,05	0,1667	0,0028
RMant8	0,4545	1,44	0,9943	0,0067	0,05	0,1667	0,0030
RMant9	0,4545	1,42	0,9943	0,0450	0,45	0,1667	0,0034
RMant10	0,4545	1,43	0,9829	0,6267	0,2	0,1667	0,0032
RMant11	0,4545	1,42	0,9943	0,0450	0,05	0,1667	0,0015
RMant12	0,4545	1,42	0,96	0,1499	0,05	0,1667	0,0033
RMant13	0,4545	1,436	0,9714	0,1076	0,25	0,1667	0,0046
RMant14	0,4545	1,436	0,9714	9,5367	0,35	0,1667	0,0086
RMant15	0,4545	1,44	0,9703	0,3815	0,35	0,1667	0,0078
RMant16	0,4545	1,436	0,9943	0,9537	0,05	0,1667	0,0069
RMant17	0,4545	1,438	0,9943	0,4768	0,9	0,3333	0,0044
RMant18	0,4545	1,446	0,9714	1,4986	2,35	1,8333	0,0033
RMant19	0,4545	1,45	0,9714	3,8147	0,5	0,5000	0,0054
RMant21	0,4545	0,86	0,9943	0,4768	-	0,1667	0,0088
RMant22	0,4545	1,08	1,0171	0,0232	-	0,3333	0,0059
RMant23	0,4545	1,18	1,0171	0,1907	-	0,3333	0,0064
RMant24	0,4545	0,872	0,9966	0,6267	-	0,3333	0,0055
RMant28	0,4545	1,436	0,9486	0,0300	0,85	0,3333	0,0016
RMant29	0,4545	1,442	0,96	0,0015	0,05	0,3333	0,0015

Step 3:Triangular functions (CTWF)

Replacing the adimensioned monitoring data in (2) – (4), the triangular whitening functions were obtained for the three Grey classes for each parameter.

An example, corresponding to the first parameter (C1), is shown in (7) – (9) and Fig. 2. Functions were inverted for dissolved oxygen since this parameter is defined by minimum values.

$$f_1^1(x_{ij}) = \begin{cases} 1 & x \leq 0{,}4545 \\ (1-x)/0{,}5454 & x \in \,]0{,}4545;\,1[\\ 0 & x \geq 1 \end{cases} \quad (7)$$

$$f_1^2(x_{ij}) = \begin{cases} 0 & x \notin [0{,}4545;1{,}5454] \\ (x-0{,}4545)/0{,}5454 & x \in \,]0{,}4545;\,1] \\ (1{,}5454-x)/0{,}5454 & x \in \,]1;\,1{,}5454[\end{cases} \quad (8)$$

$$f_1^3(x_{ij}) = \begin{cases} 0 & x \leq 1 \\ (x-1)/0{,}5454 & x \in \,]1;\,1{,}5454[\\ 1 & x \geq 1{,}5454 \end{cases} \quad (9)$$

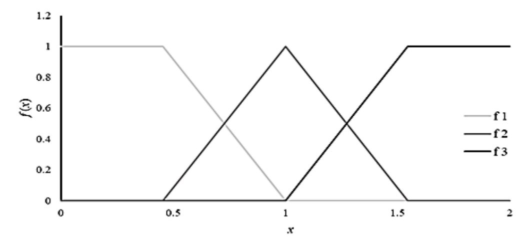

Fig. 2. CTWF for the first parameter.

An example showing the results obtained for the first five monitoring points in the first seven criteria is in Table VII.

TABLE 7
VALUES OF CTWF OF FIRST FIVE MONITORING POINT

Point	Triangular f.	C_1	C_2	C_3	C_4	C_5	C_6	C_7
RMant1	f1	1	1	0,6	1	0,1	1	1
	f2	0	0	0,4	0	0,9	0	0
	f3	0	0	0	0	0	0	0
RMant2	f1	1	1	0,48	0,9702	0	1	1
	f2	0	0	0,52	0,0298	0,9	0	0
	f3	0	0	0	0	0,1	0	0
RMant3	f1	1	1	1	0	0,7	1	1
	f2	0	0	0	0,8556	0,3	0	0
	f3	0	0	0	0,1444	0	0	0
RMant4	f1	1	1	1	0,6111	1	1	1
	f2	0	0	0	0,3889	0	0	0
	f3	0	0	0	0	0	0	0
RMant5	f1	1	1	1	0,3535	1	1	1
	f2	0	0	0	0,6465	0	0	0
	f3	0	0	0	0	0	0	0

Step 4: Determination of the weight for each criterion

The target weights were calculated using the harmonic media method described earlier. The results are shown in Table VIII. Monitoring points RMant21 to RMant24 had separated analysis without criteria C5 and C8 due to lack of monitoring data.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 08, August 2022)

TABLE 8
CLUSTERING WEIGHT FOR PARAMETERS

Criterion	A1 (λ_1)	A2 (λ_2)	A3 (λ_3)
C_1	0,0052	0,0393	0,0442
C_2	0,0020	0,0393	0,0854
C_3	0,0025	0,0393	0,0664
C_4	0,8754	0,1443	0,0251
C_5	0,0048	0,0393	0,0455
C_6	0,0048	0,0472	0,0410
C_7	0,0033	0,0393	0,0531
C_8	0,0036	0,0393	0,0512
C_9	0,0078	0,0393	0,0403
C_{10}	0,0191	0,0393	0,0364
C_{11}	0,0101	0,0498	0,0346
C_{12}	0,0041	0,0393	0,0484
C_{13}	0,0048	0,0472	0,0410
C_{14}	0,0131	0,0393	0,0376
C_{15}	0,0036	0,0393	0,0512
C_{16}	0,0167	0,0826	0,0287
C_{17}	0,0036	0,0393	0,0512
C_{18}	0,0027	0,0393	0,0615
C_{19}	0,0072	0,0393	0,0410
C_{20}	0,0032	0,0393	0,0546
C_{21}	0,0027	0,0393	0,0615

Step 5: Determination of the clustering coefficient

Total clustering coefficients for each grey class are shown in Table IX.

TABLE 9
CLUSTERING WEIGHT FOR MONITORING POINTS

Point	λ_1	λ_2	λ_3
RMant1	0,98183	0,11683	0
RMant2	0,94320	0,16292	0,02352
RMant3	0,12205	0,13759	0,00362
RMant4	0,65782	0,05973	0
RMant5	0,43274	0,09607	0
RMant6	0,12290	0,15832	0,00056
RMant7	0,12003	0,15870	0,06203
RMant8	0,98243	0,06322	0
RMant9	0,86003	0,05599	0
RMant10	0,10683	0,22238	0,00362
RMant11	0,83935	0,15803	0,00082
RMant12	0,51711	0,10378	0
RMant13	0,64770	0,11330	0
RMant14	0,11717	0,03703	0,02507
RMant15	0,11794	0,17001	0,00111
RMant16	0,12221	0,13658	0,00696
RMant17	0,11506	0,20700	0,00209
RMant18	0,11108	0,01718	0,11160
RMant19	0,11787	0,02699	0,02507
RMant21	0,11232	0,19256	0,06849
RMant22	0,92936	0,05473	0,04412
RMant23	0,37866	0,13805	0,04412
RMant24	0,11229	0,75891	0,12879
RMant28	0,99935	0,00424	0
RMant29	0,99945	0,00112	0

Step 6: Results using the max. clustering coefficient

The water quality class for each parameter, according to the maximum clustering coefficient, are shown in the Table X along with the coefficient to enable comparison within each class.

TABLE 10
RESULTS WITH MAX. CLUSTERING COEFFICIENT

Point	σ max	Class
RMant1	0,98183	A1
RMant2	0,94329	A1
RMant3	0,13759	A2
RMant4	0,65782	A1
RMant5	0,43274	A1
RMant6	0,15832	A2
RMant7	0,15870	A2
RMant8	0,98243	A1
RMant9	0,86003	A1
RMant10	0,22238	A2
RMant11	0,83935	A1
RMant12	0,51711	A1
RMant13	0,64770	A1
RMant14	0,11717	A1
RMant15	0,17001	A2
RMant16	0,13658	A2
RMant17	0,20700	A2
RMant18	0,11160	A3
RMant19	0,11787	A1
RMant21	0,19256	A2
RMant22	0,92936	A1
RMant23	0,37866	A1
RMant24	0,18859	A2
RMant28	0,99578	A1
RMant29	0,99945	A1

E. Pearson Correlation Test

The r-Pearson correlation matrix is shown in Table XI. Parameters that fell below the detection limits showed no correlation, therefore they were excluded from the matrix. Strong positive correlations can be found between DBO, DQO and total coliforms, and between nitrates and chlorides. Within metallic components, there is a remarkably high correlation system between cadmium, copper, iron, lead, and zinc with correlation values varying from moderate to strong. When comparing the correlation of parameters with the A1 clustering coefficient, there is no relevant negative nor positive correlation with any metal content, except for As, which showed a weak to moderate negative correlation. The criteria with a significant negative correlation with water quality in the Mantaro River are successively: chlorides, nitrates, arsenic, and total coliforms, characterized by a moderate to weak negative correlation.

EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 08, August 2022)

Interestingly, while total coliforms and As where the parameters that exceeded the A1 standard the most, both chloride and nitrates values fell within the first triangular function for all monitoring points (except the ones where chloride content was not evaluated).

TABLE 11
PEARSON CORRELATION MATRIX

	DO	pH	TC	DQO	DBO	NO_3	Cl	Al	As	B	Cd	Cu	Fe	Mn	Pb	Zn	σ(A1)
DO	1																
pH	9	1															
TC	55	03	1														
DQO	4	8	4	1													
DBO	39	08	99	72	1												
NO	5	5	6	7	0	1											
Cl	8	7	6	1	4	6	1										
Al	40	96	00	27	28	07	77	1									
As	39	72	14	03	14	81	46	12	1								
B	06	67	56	03	55	76	49	55	23	1							
Cd	08	52	81	55	79	08	95	86	37	39	1						
Cu	25	99	36	46	33	64	65	12	23	88	70	1					
Fe	22	57	77	38	07	23	24	29	59	72	48	91	1				
Mn	82	00	34	14	67	65	69	33	98	43	25	49	26	1			
Pb	77	31	99	46	42	63	56	97	05	43	29	38	68	16	1		
Zn	52	95	14	45	59	31	75	35	08	29	82	40	05	41	74	1	
σ(A1)	35	57	17	10	76	87	89	59	31	04	56	81	43	36	29	16	1

V. RESULTS AND DISCUSSION

A. *About the Case Study*

The maximum clustering coefficient not only gives information on the water quality class each point belongs to, but also determines quality levels within each class [5], [6]. The total ranking of monitoring points is shown in Table XII, where darker colors represent higher water quality. Objects with the description marked by an asterisk represent points of the river with influence of mining activities. Additionally, a representation of water quality along the Mantaro River is shown in Fig. 3, where the monitoring points were ordered according to their position, being RMant1 the most upstream monitoring point and RMant23 the most downstream.

TABLE 12
RANKING OF MONITORING POINTS

Point	σ max	Class	Description
RMant29	0.99945	A1	After arsenic trioxide deposit *
RMant28	0.99578	A1	Before arsenic trioxide deposit *
RMant8	0.98243	A1	Before Huari river
RMant1	0.98183	A1	Downstream of Upamayo dam
RMant2	0.9432	A1	After intersection with Coricancha river *
RMant22	0.92936	A1	Downstream of Anco residual waters treatment facility
RMant9	0.86003	A1	After Huari river
RMant11	0.83935	A1	After Pachacayo river
RMant4	0.65782	A1	Before STRAKRAFF hidroelectric discharge
RMant13	0.6477	A1	Before Achamayo river
RMant12	0.51711	A1	Upstream of CIMIR and Plan Meris
RMant5	0.43274	A1	After STRAKRAFF hidroelectric discharge
RMant23	0.37866	A1	Downstream of San Pedro de Coris (ex Cobriza) *
RMant19	0.11787	A1	Huancayo exit
RMant14	0.11717	A1	Before Pilcomayo dump
RMant10	0.22238	A2	Before Pachacayo river
RMant17	0.207	A2	Downstreamof “Agua de las vírgenes” dump
RMant21	0.19256	A2	Downstream of “Mejorada”
RMant24	0.18859	A2	Near Vilca river
RMant15	0.17001	A2	After Pilcomayo dump and before "Mejorada" dump
RMant7	0.1587	A2	After Doe Run refinery *
RMant6	0.15832	A2	Before Doe Run refinery *
RMant3	0.13759	A2	After Paccha-La Oroya population
RMant16	0.13658	A2	Downstream of Breña bridge
RMant18	0.1116	A3	Downstream of Shullcas river

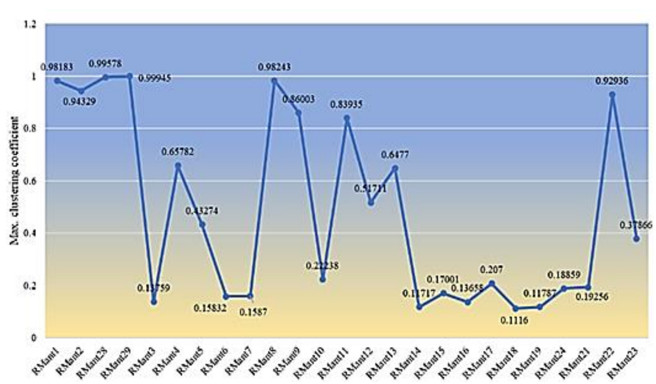

Fig. 3. Maximum clustering coefficient of monitoring points in downstream order.

Points RMant28 and RMant29 are located before and after the arsenic trioxide deposit of Malpaso, supervised by the federal company "Activos Mineros S.A.C." [16], meanwhile RMant2 is located next to limestone mine Las Monas. Points RMant6 and RMant7 are the mining related points with the lowest water quality, however, both points are located next to the city of Yauli as well. Due to the low water quality exhibited by points next to populated areas like Pacha-La Oroya (RMant3) and Huancayo (RMant19), a direct relationship between the low water quality next to the Doe Run refinery and its operations cannot be assured. Points from RMant14 to RMant21 make a segment of very low water quality as can be seen in Fig. 3. This section is characterized by multiple dumps like Agua de las Virgenes and Mejorada. Point RMant23, located downstream of Cobriza exhibits a low water quality but still in the A1 class, this is the only point where a connection between mining activities and lower water quality can be made.

Point RMant18, located after the intersection with Shullcas river showed the lowest water quality in the whole river. The pollution can be directly associated to the dump El Eden which is in the intersection between both rivers [16]. The Peruvian Environmental Evaluation and Inspection Organism (OEFA) reported the Huancayo and El Tambo municipalities due to the poor management of solid residuals in 2014 [16]. The district started closing operations in 2015 [17]. As the data collected is from 2018, clearly the dump continued to negatively impact the water quality of the river.

According to the Pearson correlation analysis, there is no significant effect of metal content on water quality or any other parameter different from metallic content except for weak positive and negative correlations with chemical oxygen demand and boron, respectively. A high correlation between characteristic heavy metals of polymetallic mines such as Cu, Cd, Fe, Pb and Zn suggests there is a metal input from mining in the Mantaro River as determined by Custodio et. al [13]. Arsenic is a common waste from mining activities [18] and is the third criteria with the highest negative impact on water quality. However, there are no significant correlations between As and heavy metals nor any other criteria except for a moderate correlation with B. Arsenic can be released as a component of pesticides [19] thus its presence could be more related to agriculture. Chlorides and nitrates content are strongly correlated and have the highest correlation with low water quality.

Nitrates main anthropogenic sources in water systems are agriculture and domestic waste [20], [21], hence chloride content is most likely anthropogenic as well and caused by the same sources. The fourth criteria with the highest negative correlation with water quality was thermotolerant coliforms, whose main source is feces from humans and livestock. Due to the large variation of coliform content in water and the difference between each class's standard, thermotolerant coliforms was the criteria with the highest clustering weight, making up to 87% of the weight for A1 and 14% in A2, thus showing a problem with using harmonic mean as a weighting function when the criteria standards show a logarithmic distribution. More robust weighting functions like Shannon entropy [22] could be applied to diminish the effect.

B. About the Methodology

Both the Grey Clustering method and the r-Pearson linear correlation coefficient complement each other. The first is effective insofar as it is a question of data management with uncertainty, where the determination of the object of study, criteria, and classes, is adjusted to the case of evaluating the water quality of a watershed. However, it is necessary to make certain modifications to the evaluation parameters or national environmental standards to carry out this method correctly. The second method mentioned is suitable to measure the degree of relationship between criteria and water quality, where correlation results are obtained to complement what is obtained with the Grey Clustering method. These methods have multiple applications in various scenarios, and are quite advantageous over other methods, since it only requires real data, quality standards, and knowledge. In addition, it is a relief in terms of research and study costs since it is not necessary to consult experts such as the Delphi method and others.

Likewise, from the analysis of the literature review, the following can be said:

In a mining impacted water system high content of heavy metals linked to poor water quality are expected to be found [7], [23]. Custodio et. al [8] have reported high concentrations of heavy metals in the Mantaro River directly related to mine waste, showing that the river is still affected by mining operations. However, this paper only focused on heavy metals, and thus does not assess the overall water quality of the Mantaro River, or if these metals are indeed the driving force behind pollution.

Kothari et. al [9] found that microbial contamination caused by domestic waste in poor water management facilities was the main driver behind water pollution in six districts of India. A similar situation happens in the Mantaro watershed, because agriculture is the main economic activity (54.6% of occupations), 55.5% of the population in the region live in poverty conditions and a high percentage of population lacks sewerage services [4].

VI. CONCLUSIONS

The results obtained from grey clustering show that water quality is highly compromised in the Mantaro River. The points that exhibit the lowest water quality are located next to municipal dumps and populated areas. According to the harmonic mean weighting function applied, thermotolerant coliforms is the most important criteria. Pearson correlation between the parameters and maximum clusterization weight for the first grey class concludes that parameters related with domestic and agricultural waste had the highest correlation with low water quality. Together, both approximations suggest that the main driver behind pollution in the Mantaro River is not mining, but domestic waste in dumps and probably agriculture. There is a strong correlation between different heavy metals suggesting mining input but a poor to irrelevant correlation with water quality. This aspect, as well as the divergence of water quality maximum clustering coefficients in mining related points, show that the impact of mining on water quality is close to irrelevant. Future studies should focus on the influence of domestic waste and poor sanitary conditions in the region rather than mining activities to effectively characterize the pollution and develop treatment projects.

The methods used (Grey Clustering and Pearson's correlation) are effective to carry out a study on water quality, since uncertainty is considered, and national parameters (ECA) can be adjusted by adding certain international standards to improve the accuracy of the results. The study carried out does not present significant costs, and it is quite accurate, so its use is recommended to national authorities so they can implement improvements in monitoring and treatment in different watersheds. Likewise, for future studies, it is recommended to expand the methodology using quality standards from neighboring countries or different weighting algorithms in order to compare the results with the present study.

REFERENCES

[1] USGS, "Mining and Water Quality," 2019, [Online]. Available: https://www.usgs.gov/special-topic/water-science-school/science/mining-and-water-quality?qt-science_center_objects=0#qt-science_center_objects.

[2] S. Foundation, "Mining and Water Pollution," 2017. [Online]. Available: https://www.safewater.org/fact-sheets-1/2017/1/23/miningandwaterpollution.

[3] A. Bebbington and M. Williams, "Water and Mining Conflicts in Peru," Mt. Res. Dev., vol. 28, no. 3/4, pp. 190–195, Aug. 2008, doi: 10.1659/mrd.1039.

[4] A. N. del Agua, "Consejo de Recursos Hídricos de Cuenca Mantaro: Caracterización Socioeconómica," 2019. [Online]. Available: https://www.ana.gob.pe/2019/consejo-de-cuenca/mantaro/VC.

[5] A. Delgado, A. Aguirre, E. Palomino, and G. Salazar, "Applying triangular whitenization weight functions to assess water quality of main affluents of Rimac river," in Proceedings of the 2017 Electronic Congress, E-CON UNI 2017, Jun. 2017, vol. 2018-Janua, pp. 1–4, doi: 10.1109/ECON.2017.8247308.

[6] A. Delgado, P. Montellanos, J. Llave, "Air quality level assessment in Lima city using the grey clustering method," IEEE ICA-ACCA 2018 - IEEE International Conference on Automation/23rd Congress of the Chilean Association of Automatic Control: Towards an Industry 4.0 – Proceedings, 8609699.

[7] V. A. Efimov, S. R. Chalov, L. E. Efimova, V. A. Ivanov, J. Jarsjö, and S. Fisher, "Impact of mining activities on the surface water quality (case study of Khibiny mountains, Russia)," IOP Conf. Ser. Earth Environ. Sci., vol. 263, p. 012008, May 2019, doi: 10.1088/1755-1315/263/1/012008.

[8] M. Custodio et al., "Data on the concentration of heavy metals and metalloids in lotic water of the Mantaro river watershed and human risk assessment, Peru," Data Br., vol. 30, p. 105493, Jun. 2020, doi: 10.1016/j.dib.2020.105493.

[9] V. Kothari, S. Vij, S. Sharma, and N. Gupta, "Correlation of various water quality parameters and water quality index of districts of Uttarakhand," Environ. Sustain. Indic., vol. 9, p. 100093, Feb. 2021, doi: 10.1016/j.indic.2020.100093.

[10] R. K. Tiwary, B. Kumari, and D. B. Singh, "Water Quality Assessment and Correlation Study of Physico-Chemical Parameters of Sukinda Chromite Mining Area, Odisha, India," 2018, pp. 357–370.

[11] A. Delgado,I.Romero,"Environmental conflict analysis on a hydrocarbon exploration project using the Shannonentropy," Proceedings of the 2017 Electronic Congress, E-CON UNI 2017, 2018-January, pp. 1-4.

[12] S. Liu and Y. Lin, Introduction to grey systems theory, vol. 68. 2010.

[13] M. Custodio, D. Álvarez, W. Cuadrado, R. Montalvo, and S. Ochoa, "Potentially toxic metals and metalloids in surface water intended for human consumption and other uses in the Mantaro River watershed, Peru," Soil Water Res., vol. 15, no. No. 4, pp. 237–245, Sep. 2020, doi: 10.17221/152/2019-SWR.

[14] A. N. del Agua, "Resultados del monitoreo participativo de la calidad del agua superficial en el ámbito de la cuenca del Mantaro," 2018.

[15] MINAM-PERÚ, Decreto Supremo N° 004-2017- MINAM – Aprueban Estándares de Calidad Ambiental (ECA) para Agua y establecen Disposiciones Complementarias. 2017, pp. 13–14.

[16] OEFA, "El OEFA denuncia a la Municipalidad Provincial de Huancayo y a la Municipalidad Distrital de El Tambo por inadecuada disposición final de residuos sólidos," 2014. [Online]. Available: https://www.oefa.gob.pe/botadero-el-eden-y-agua-de-las-virgenes/ocac07/.

[17] "El Botadero de El Edén entra en plan de cierre," Correo, Oct. 25, 2015.

[18] C. Luo, J. Routh, D. Luo, L. Wei, and Y. Liu, "Arsenic in the Pearl River Delta and its related waterbody, South China: occurrence and sources, a review," Geosci. Lett., vol. 8, no. 1, p. 12, Dec. 2021, doi: 10.1186/s40562-021-00185-9.

[19] V. Bencko and F. Yan Li Foong, "The history of arsenical pesticides and health risks related to the use of Agent Blue," Ann. Agric. Environ. Med., vol. 24, no. 2, pp. 312–316, Jun. 2017, doi: 10.26444/aaem/74715.

[20] Y. Xue, J. Song, Y. Zhang, F. Kong, M. Wen, and G. Zhang, "Nitrate Pollution and Preliminary Source Identification of Surface Water in a Semi-Arid River Basin, Using Isotopic and Hydrochemical Approaches," Water, vol. 8, no. 8, p. 328, Aug. 2016, doi: 10.3390/w8080328.

[21] X. Zhang, Y. Zhang, P. Shi, Z. Bi, Z. Shan, and L. Ren, "The deep challenge of nitrate pollution in river water of China," Sci. Total Environ., vol. 770, p. 144674, May 2021, doi: 10.1016/j.scitotenv.2020.144674.

[22] C. E. Shannon, "A Mathematical Theory of Communication," Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[23] A. Delgado, I. Romero, "Applying the Grey Systems Theory to Assess Social Impact from an Energy Project," Proceedings of the 2018 IEEE 25th International Conference on Electronics, Electrical Engineering and Computing, INTERCON 2018, 8526372.