Assessment of Surface Water Quality on the Upper Watershed of Huallaga River, in Peru, using Grey Systems and Shannon Entropy

Alexi Delgado[1], Jharison Vidal[2], Jhon Castro[3], Jhonel Felix[4], Jorge Sáenz[5]
Faculty of Systems Engineering, Universidad de Ciencias y Humanidades, Lima, Perú[1]
Environmental Engineering Program, Universidad Nacional de Ingeniería, Lima, Perú[2, 3, 4, 5]

Abstract—The assessment of the quality of surface water is a complex issue that entails the comprehensive analysis of several parameters that are altered by natural or man-made causes. In this sense, the Grey Clustering method, which is based on Grey Systems theory, and Shannon Entropy, based on the artificial intelligence approach, provide an alternative to evaluate water quality in an integral way considering the uncertainty within the analysis. In the present study, the water quality on the upper watershed of Huallaga river was evaluated taking into account the monitoring results of twenty-one points carried out by the National Water Authority (ANA) analyzing nine parameters of the Prati index. The results showed that all the monitoring points of the Huallaga river were classified as not contaminated, which means that the discharges, generated by economic activities, are carried out through of treatment plants meeting the quality parameters. Finally, the results obtained can be of great help to the ANA and the regional and local authorities of Peru in making decisions to improve the management of the Huallaga river watershed.

Keywords—Grey clustering; Huallaga river; Prati index; Shannon entropy; water quality

I. Introduction

The Huallaga river watershed is one of the main watersheds of Peru and with a great potential of water resources, this due to the existence of a large number of lagoons, rivers, streams and springs, an important source of resources natural resources, food and work for the native communities and populated centers of the area being the main economic activities: agricultural, industrial, energy, mining and fishing for direct human consumption [1]. In addition, the benefits provided by the watershed have been diminished, due to its waters have been polluted by domestic wastewater, wastewater municipal, solid waste, as well as mining environmental liabilities, being a risk to public health [2].

This watershed, located on the Atlantic watershed, is one of the largest tributaries of the Marañon River and It is made up of the Lower Huallaga Inter-watershed, Paranapura, Middle Lower Huallaga Inter-watershed, Mayo Watershed, Middle Huallaga Inter-watershed, Biabo Watershed, Middle Upper Huallaga Inter-watershed, Huyabamba Watershed and Upper Inter-watershed Huallaga [1]. The assessment of the surface water quality will be carried out in the upper part of the watershed due to its environmental importance.

For the development of the assessment we will use the Grey Clustering method, as well as the Shannon Entropy. Flock Clustering is a method that is based on the theory of Grey systems, an approach within what is called Intelligence Artificial, so it has a great variety of applications [3]. For the case to be studied, we will use the center-point triangular whitenization weight functions (CTWF) method, which is applied in studies on water [4] or in the assessment of urban transport [5]. On the other hand, the weight method Entropy, based on Entropy of Shannon, is also an approach within artificial intelligence developed initially by Claude E. Shannon [6], this method was used to calculate the weights objectives of the assessment criteria within the CTWF method.

Therefore, our specific objective in this study is the classification of 21 monitoring points on the upper watershed of Huallaga river according to the water quality criteria, using the Grey Clustering method and the Entropy of Shannon.

In the present study, Section II details the CTWF method. Section III describes the case study, followed by the results and discussions of Section IV. The conclusions will be presented in Section V.

II. Methodology

In this section, we describe the CTWF method, which can be described as follows: first, suppose the area is set of m objects, a set of n criteria, and a set of s Grey classes, according to the sample value (i = 1, 2,..., m; j = 1, 2,…, n). Then, the steps of the CTWF method can be developed with the following points according to different research [3], [7] and [8].

A. Step 1: Determination of Center Points

The ranges of the criteria are divided into 5 Grey classes, and then their central points are $\lambda_1, \lambda_2, \ldots, y\ \lambda_s$, this is determined by the Prati index.

B. Step 2: Dimension Removal

At this point it is assumed that there are objects for assessment and n criteria or Grey classes, which forms the following Matrix $Z = \{Z_{ij}; i = 1,2,\ldots,m\ ; j = 1,2,\ldots,n\}$. In that sense, it is normalized for each criterion $C_j (j = 1, 2, \ldots, n)$. The value is normalized P_{ij}, which is calculated by Equation 1.

$$P_{ij} = \frac{Z_{ij}}{\frac{\sum_{j=1}^{n} z_{ij}}{n}} \tag{1}$$

C. Step 3: Determination of the Triangular Functions and their Values

The Grey classes are expanded in the addresses of each parameter used and for this the index will be used as a reference Prati, who provided the data to measure quality, in this research Prati provides us with five 5 levels of quality for each parameter, so there will be five (5) functions for each parameter. The new sequence of points central is $\lambda_1, \lambda_2,\ldots y$ λ_5. For the class k = 1,2,3,4 and 5, j = 1, 2, …, n, for an observed value xij. The calculation is displayed of the CTWF by means of Equations 2, 3 and 4; and Fig. 1 shows the graph of the triangular functions.

$$f_j^1(x_{ij}) = \begin{cases} 1, & x \ni [0, \lambda_j^1] \\ \frac{\lambda_j^2 - x}{\lambda_j^2 - \lambda_j^1}, & x \in \langle \lambda_j^1, \lambda_j^2 \rangle \\ 0, x \in [\lambda_j^2, +\infty \rangle \end{cases} \tag{2}$$

$$f_j^k(x_{ij}) = \begin{cases} \frac{x - \lambda_j^{k-1}}{\lambda_j^k - \lambda_j^{k-1}}, & x \ni \langle \lambda_j^{k-1}, \lambda_j^k] \\ \frac{\lambda_j^{k+1} - x}{\lambda_j^{k+1} - \lambda_j^k}, & x \in \langle \lambda_j^k, \lambda_j^{k+1} \rangle \\ 0, x \in [0, \lambda_j^{k-1}] U [\lambda_j^2, +\infty \rangle \end{cases} \tag{3}$$

$$f_j^5(x_{ij}) = \begin{cases} \frac{x - \lambda_j^4}{\lambda_j^5 - \lambda_j^4}, & x \in \langle \lambda_j^4, \lambda_j^5 \rangle \\ 1, & x \in [\lambda_j^5, +\infty \rangle \\ 0, x \in [0, \lambda_j^4] \end{cases} \tag{4}$$
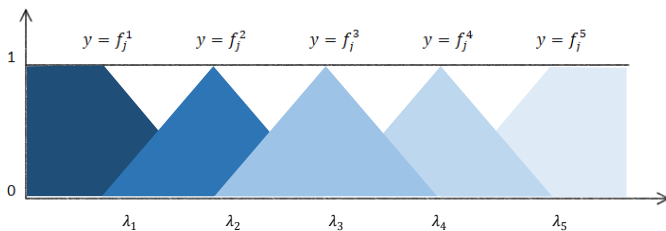

Fig. 1. CTWF According to Prati Scale.

D. Step 4: Determination of the Weight for each Criterion

In this step, the Shannon Entropy weight method is used. For everything P_i it is considered within a distribution of probability, Shannon developed the measure H, which satisfies the following properties [3], [7] and [9]:

- H is a positive continuous function

- If all p_i are equivalent and $p_i = 1/n$, in this sense, H should be a monotonous increasing function of n.

- For all n ≥ 2, $H(P_1, P_2, \ldots, P_n) = h(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)H(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$

Shannon shows that only the functions that satisfy this condition are calculated by Equation 5.

$$H_{Shannon} = -\sum_i^n p_i \log(p_i)$$

$$Where: 0 \le p_i \le 1; \sum_{f=1}^n p_i = 1 \tag{5}$$

Around the entropy weight methodology, it can be demonstrated according to the following definition [3], [7] and [9]. As shown above, m objects are displayed for assessment, and n assessment criteria, which form the following matrix $x = \{x_{ij}; i = 1,2, \ldots, m; j = 1,2, \ldots n\}$. After that, the following stages continue.

1) Then the matrix $x = \{x_{ij}; i = 1,2, \ldots, m; j = 1,2, \ldots n\}$ is normalized by each criterion Cj. The normalization evaluates Pij are calculated by Equation 6.

$$f_j^1(x_{ij}) P_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \tag{6}$$

2) The entropy of each criterion is calculated by Equation 7, which was constructed based on Equation 6.

$$H_j = -k \sum_{i=1}^m P_{ij} ln(P_{ij}) \tag{7}$$

Where, K is a constant, k= (ln(m))⁻¹

3) The degree of divergence of the intrinsic information in each criterion C_j is calculated by Equation 8.

$$div_j = 1 - H_j \tag{8}$$

4) In the weight entropy w_j of each criterion C_j is calculated by Equation 9.

$$w_j = \frac{div_j}{\sum_{j=1}^n div_j} \tag{9}$$

Where, w_j is equal to n_j

E. Step 5: Determination of the Clustering Coefficient

The clustering coefficient σ_i^k by objeto i, i = 1, 2, …, m, respect to the Grey classes k, k=1, 2, …, s, is calculated by Equation 10.

$$\sigma_i^k = \sum_{j=1}^n f_j^k(x_{ij}) n_j \tag{10}$$

Where $f_{j=1}^k(x_{ij})$ is the CTWF of the k-th grey class of j-th criterion, y n_j is the weight of criterion J, establish said weights the Shannon Entropy method will be used.

F. Step 6: Results using the Maximum Clustering Coefficient

Finally, we have the calculation of $max_{1 \le k \le s}\{\sigma_i^k\} = \sigma_i^k$, We decide which object belongs to Grey class k*. When there are several objects in Grey class k*, these objects can be ordered according to the magnitudes of their grouping coefficients integral.

III. CASE STUDY

The analysis of the surface water quality was carried out in the upper part of the Huallaga river watershed, which is located in the central zone of Peru has an area of 89,416 km2 and a length of 1,168 km in a direction from south to north [1], which is represented in Fig. 2.

Fig. 2. Huallaga River Watershed, Peru.

A. *Definition of Study Objects*

For the assessment of the quality of the surface water of the upper watershed of Huallaga river , information was collected of 21 monitoring points obtained from the seventh monitoring of surface water quality carried out on 19November to December 20, 2019 by the Huallaga Water Administrative Authority and the Local Authorities of the Alto Huallaga, Tingo María, Huallaga Central, Alto Mayo and Tarapoto [1]. Which will be detailed in Table I and represented in Fig. 3.

TABLE I.   MONITORING POINTS IN THE UPPER WATERSHED OF HUALLAGA RIVER

Point	Code	Coordinates (WGS 84)	
		East	*North*
1	RHual1	370546	8828639
2	RHual2	369690	8830701
3	RHual3	370966	8836230
4	RHual4	370755	8836925
5	RHual5	370959	8843646
6	RHual6	369913	8876992
7	RHual7	367820	8882470
8	RHual8	366866	8886415
9	RHual9	363585	8896395
10	RHual10	362973	8900714
11	RHual11	364061	8901392
12	RHual12	367093	8907020
13	RHual13	379125	8912553
14	RHual14	384109	8914086
15	RHual15	395255	8950625
16	RHual16	393159	8959917
17	RHual17	389447	8971679
18	RHual18	390103	8974251
19	RHual19	380015	9002651
20	RHual42	369474	8832777
21	RHual43	369809	8834456
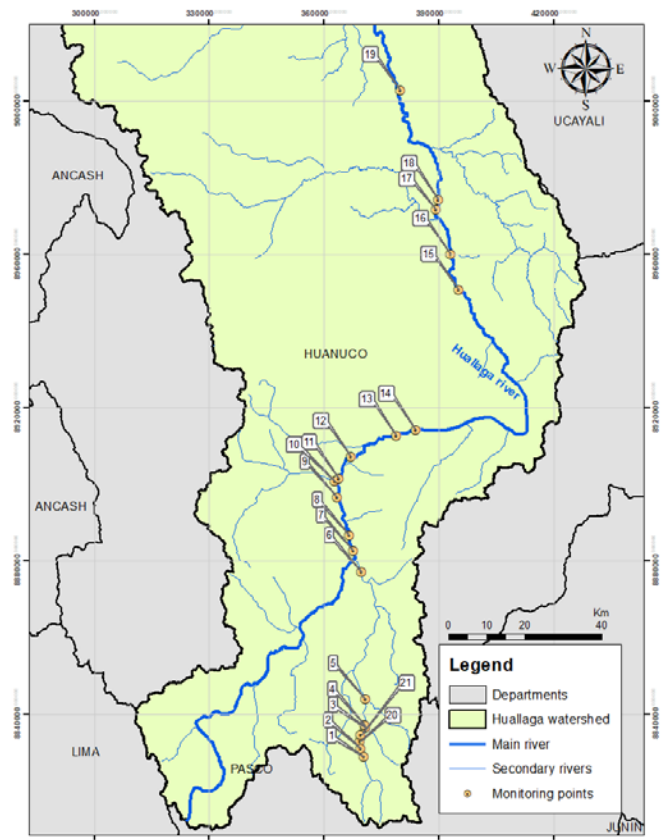
Fig. 3. Monitoring Points of Surface Water Quality in the Upper Watershed of Huallaga River.

B. *Definition of Assessment Criteria*

The assessment criteria for the present study are determined by the water quality parameters which are presented in Table II.

C. *Definition of the Grey Classes*

The classes for assessment are five and are based on the levels of water quality according to the Prati index, which are presented in Table III.

TABLE II.   PRATI INDEX STANDARD DATA FOR THE ASSESSMENT OF WATER QUALITY

Criterion	Units	Notation
pH	0 -14	C_1
BOD	ppm	C_2
COD	ppm	C_3
Total Suspended Solids	mg/L	C_4
NH_3	ppm	C_5
NO_3	ppm	C_6
Cl	ppm	C_7
Manganese	ppm	C_8
Iron	ppm	C_9

TABLE III.   PRATI INDEX STANDARD DATA FOR THE ASSESSMENT OF WATER QUALITY

Parameters	Quality Index Condition				
	Uncontaminated	*Acceptable*	*Moderately contaminated*	*Contaminated*	*Highly Contaminated*
pH	6.5-8.0	8.0-8.4	8.4-9.0	9.0-10.1	>10.1
BOD (ppm)	0.0-1.5	1.5-3.0	3.0-6.0	6.0-12.0	>12.0
COD (ppm)	0-10	10-20	20-40	40-80	>80
Total Suspended Solids (mg/L)	0-20	20-40	40-100	100-278	>278
NH_3 (ppm)	0-0.1	0.1-0.3	0.3-0.9	0.9-2.7	>2.7
NO_3 (ppm)	0-4	4-12	12-36	36-108	>108
Cl (ppm)	0-50	50-150	150-300	300-620	>620
Manganese (ppm)	0.00-0.05	0.05-0.17	0.17-0.50	0.50-1.00	>1.00
Iron (ppm)	0.0-0.1	0.1-0.3	0.3-0.9	0.9-2.7	>2.7

D. Calculations using the CTWF Method

The calculations based on the gray clustering method are presented below:

1) Step 1: Based on the Prati quality index, the central values of the parameters to be analyzed are obtained. These values are shown in Table IV.

2) Step 2: The non-dimensioned standard values for each parameter, according to the Prati index, were determined through the (1). These values are presented in Table V.

TABLE IV.   CENTRAL VALUES OF THE PRATI INDEX PARAMETERS

Parameters	Uncontaminated (λ_1)	Acceptable (λ_2)	Moderately contaminated (λ_3)	Contaminated (λ_4)	Highly Contaminated (λ_5)
pH	7.25	8.20	8.70	9.55	10.40
BOD (ppm)	0.75	2.25	4.50	9.00	13.50
COD (ppm)	5.00	15.00	30.00	60.00	90.00
Total suspended Solids (mg/L)	10.00	30.00	70.00	189.00	308.00
NH_3 (ppm)	0.05	0.20	0.60	1.80	3.00
NO_3 (ppm)	2.00	8.00	24.00	72.00	120.00
Cl (ppm)	25.00	100.00	225.00	460.00	695.00
Manganese (ppm)	0.03	0.11	0.34	0.75	1.16
Iron (ppm)	0.05	0.20	0.60	1.80	3.00

TABLE V.   NON-DIMENSIONAL STANDARD VALUES FROM PRATI INDEX

Parameters	Notation	Uncontaminated (λ_1)	Acceptable (λ_2)	Moderately contaminated (λ_3)	Contaminated (λ_4)	Highly Contaminated (λ_5)
pH	C_1	0.822	0.930	0.986	1.083	1.179
BOD	C_2	0.125	0.375	0.750	1.500	2.250
COD	C_3	0.125	0.375	0.750	1.500	2.250
TSS	C_4	0.082	0.247	0.577	1.557	2.537
NH_3	C_5	0.044	0.177	0.531	1.593	2.655
NO_3	C_6	0.044	0.177	0.531	1.593	2.655
Cl	C_7	0.083	0.332	0.748	1.528	2.309
Mn	C_8	0.053	0.231	0.704	1.576	2.437
Fe	C_9	0.044	0.177	0.531	1.593	2.655

Similarly, based on the results of the participatory monitoring report of surface water quality in the Huallaga river watershed, developed by the National Water Authority (ANA), the values without dimension were obtained for each parameter of the 21 selected monitoring points. These values are presented in Table VI.

TABLE VI.   NON-DIMENSION MONITORING DATA IN THE CASE STUDY

Point	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
1	0.985	0.167	0.150	0.239	0.060	0.045	0.007	0.112	0.454
2	1.002	0.167	0.100	0.387	0.020	0.056	0.010	0.166	0.531
3	0.999	0.167	0.100	0.568	0.088	0.058	0.009	0.302	0.672
4	0.999	0.167	0.175	0.428	0.034	0.044	0.007	0.219	0.729
5	0.989	0.167	0.225	0.659	0.028	0.047	0.008	0.239	1.010
6	0.994	0.167	0.500	12.545	0.003	0.032	0.004	2.292	22.673
7	0.998	0.167	0.050	10.988	0.005	0.024	0.010	1.295	21.407
8	0.993	0.833	0.350	0.428	0.081	0.019	0.009	2.225	33.487
9	1.001	0.167	0.025	9.292	0.032	0.024	0.010	1.397	19.823
10	0.955	0.833	0.250	12.315	0.090	0.019	0.009	2.006	32.195
11	0.992	0.167	0.050	4.481	0.040	0.026	0.009	0.992	14.513
12	0.989	0.500	0.250	13.213	0.100	0.020	0.008	2.160	0.333
13	0.982	0.167	0.025	12.842	0.100	0.020	0.007	2.134	0.326
14	0.971	0.167	0.375	4.992	0.055	0.033	0.010	0.780	9.735
15	0.940	0.167	0.800	0.593	0.005	0.012	0.004	0.131	2.270
16	0.981	0.167	0.475	4.893	0.005	0.018	0.008	1.054	14.646
17	0.863	0.167	0.300	0.832	0.081	0.018	0.014	0.218	3.965
18	0.935	0.167	0.025	1.425	0.005	0.015	0.008	0.282	4.985
19	0.918	0.167	0.025	1.928	0.091	0.017	0.016	0.327	4.554
20	1.001	0.167	0.225	0.470	0.075	0.055	0.008	0.233	0.448
21	0.935	0.167	0.300	0.387	0.072	0.053	0.008	0.302	0.668

3) Step 3: Replacing the values in Table III in (2)-(4), the triangular whitening functions of the five Gray classes were obtained for each parameter. As an example, the functions corresponding to the second parameter (BDO) are shown in (11)-(15) and Fig. 4. Then, the values in Table VI were evaluated in the triangular whitening functions of the five classes Grey for each parameter. The results obtained for the first five monitoring points are shown in Table VII.

$$f_2^1(x_{ij}) = \begin{cases} 1, x \in [0, 0.125] \\ \left(\frac{0.375-x}{0.375-0.125}\right), x \in <0.125, 0.375> \\ 0, x \in [0.375, +\infty> \end{cases} \quad (11)$$

$$f_2^2(x_{ij}) = \begin{cases} \left(\frac{x-0.125}{0.375-0.125}\right), x \in <0.125, 0.375] \\ \left(\frac{0.750-x}{0.750-0.375}\right), x \in <0.375, 0.750> \\ 0, x \in [0, 0.125] \cup [0.750, +\infty> \end{cases} \quad (12)$$

$$f_2^3(x_{ij}) = \begin{cases} \left(\frac{x-0.375}{0.750-0.375}\right), x \in <0.375, 0.750] \\ \left(\frac{1.500-x}{1.500-0.750}\right), x \in <0.750, 1.500> \\ 0, x \in [0, 0.375] \cup [1.500, +\infty> \end{cases} \quad (13)$$

$$f_2^4(x_{ij}) = \begin{cases} \left(\frac{x-0.750}{1.500-0.750}\right), x \in <0.750, 1.500] \\ \left(\frac{2.250-x}{2.250-1.500}\right), x \in <1.500, 2.250> \\ 0, x \in [0, 0.750] \cup [2.250, +\infty> \end{cases} \quad (14)$$

$$f_2^5(x_{ij}) = \begin{cases} \left(\frac{x-1.500}{2.250-1.500}\right), x \in <1.500, 2.250> \\ 1, x \in [2.250, +\infty> \\ 0, x \in [0, 1.500] \end{cases} \quad (15)$$
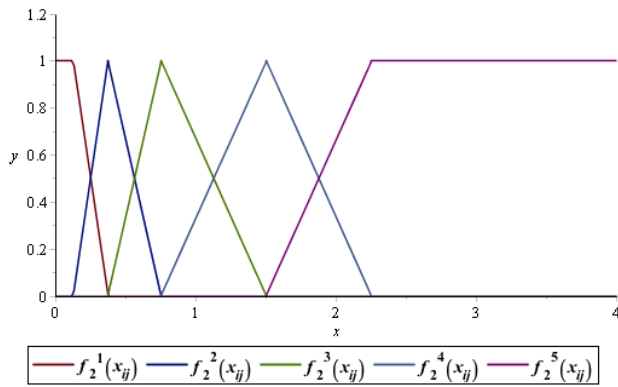

Fig. 4. CTWF for the Second Parameter.

TABLE VII. VALUES OF CTWF OF FIRST FIVE MONITORING POINT

P1	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
C_1	0	0.02	0.98	0	0
C_2	0.833	0.167	0	0	0
C_3	0.9	0.1	0	0	0
C_4	0.05	0.95	0	0	0
C_5	0.88	0.12	0	0	0
C_6	0.991	0.009	0	0	0
C_7	1	0	0	0	0
C_8	0.668	0.332	0	0	0
C_9	0	0.218	0.782	0	0

P2	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
C_1	0	0	0.835	0.165	0
C_2	0.833	0.167	0	0	0
C_3	1	0	0	0	0
C_4	0	0.575	0.425	0	0
C_5	1	0	0	0	0
C_6	0.911	0.09	0	0	0
C_7	1	0	0	0	0
C_8	0.367	0.633	0	0	0
C_9	0	0	1	0	0

P3	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
C_1	0	0	0.871	0.129	0
C_2	0.833	0.167	0	0	0
C_3	1	0	0	0	0
C_4	0	0.025	0.975	0	0
C_5	0.673	0.327	0	0	0
C_6	0.897	0.103	0	0	0
C_7	1	0	0	0	0
C_8	0	0.85	0.15	0	0
C_9	0	0	0.867	0.133	0

P4	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
C_1	0	0	0.871	0.129	0
C_2	0.833	0.167	0	0	0
C_3	0.8	0.2	0	0	0
C_4	0	0.45	0.55	0	0
C_5	1	0	0	0	0
C_6	1	0	0	0	0
C_7	1	0	0	0	0
C_8	0.068	0.932	0	0	0
C_9	0	0	0.813	0.187	0

P5	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
C_1	0	0	0.976	0.024	0
C_2	0.833	0.167	0	0	0
C_3	0.6	0.4	0	0	0
C_4	0	0	0.916	0.084	0
C_5	1	0	0	0	0
C_6	0.98	0.021	0	0	0
C_7	1	0	0	0	0
C_8	0	0.983	0.017	0	0
C_9	0	0	0.549	0.451	0

4) *Step 4:* The clustering weight (ηi) of each parameter was determined using Shannon entropy. For this, the following procedure:

a) Substep 4.1: The values of the parameters of the Prati index were normalized. These values are presented in Table VIII.

b) Substep 4.2: The entropy Hj of each criterion Cj was calculated through (7). The results are presented in Table IX.

c) Substep 4.3: The degree of divergence of each criterion Cj was calculated through (8). The results are shown in Table X.

d) Substep 4.4: Finally, the entropy weights wj according to (9) and were equated to the grouping weights nj of each parameter. The values are presented in Table XI.

5) *Step 5:* The values of the clustering coefficients(σ_i^k) were calculated using (10). The results of the the first five monitoring points are shown in Table XII.

6) *Step 6:* Finally, the condition was applied: if $max\{\sigma_i^k\} = \sigma_i^{k*}$, it is decided that the object i belongs to the Grey class k*; for each monitoring point.

TABLE VIII. NORMALIZED VALUES OF EACH CRITERIA

Parameter \ Class	λ_1	λ_2	λ_3	λ_4	λ_5	
C_1	0.164	0.186	0.197	0.217	0.236	1
C_2	0.025	0.075	0.15	0.3	0.45	1
C_3	0.025	0.075	0.15	0.3	0.45	1
C_4	0.016	0.049	0.115	0.311	0.507	1
C_5	0.009	0.035	0.106	0.319	0.531	1
C_6	0.009	0.035	0.106	0.319	0.531	1
C_7	0.017	0.066	0.15	0.306	0.462	1
C_8	0.011	0.046	0.141	0.315	0.487	1
C_9	0.009	0.035	0.106	0.319	0.531	1

TABLE IX. ENTROPY VALUES IN THE CASE STUDY

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
H_j	0.995	0.803	0.803	0.729	0.683	0.683	0.778	0.733	0.683

TABLE X. DEGREE OF DIVERGENCE

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
div_j	0.005	0.197	0.197	0.271	0.317	0.317	0.222	0.267	0.317

TABLE XI. CLUSTERING WEIGHT OF EACH PARAMETER

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
$w_{j=}$ n_j	0.002	0.094	0.094	0.128	0.15	0.15	0.105	0.126	0.150

TABLE XII. VALUES OF CTWF AND σ_i^k FOR THE FIRST FIVE MONITORING POINTS

P1	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
σ_i^k	0.639	0.241	0.12	0	0
P2	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
σ_i^k	0.61	0.183	0.207	0	0
P3	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
σ_i^k	0.513	0.191	0.276	0.02	0
P4	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
σ_i^k	0.567	0.21	0.195	0.028	0
P5	$f_J^1(X)$	$f_J^2(X)$	$f_J^3(X)$	$f_J^4(X)$	$f_J^5(X)$
σ_i^k	0.537	0.18	0.204	0.079	0

TABLE XIII. VALUES OF σ_i^k FOR EACH MONITORING POINT

Point	$max\ \sigma_i^k$	Prati index
P1	0.639	Uncontaminated
P2	0.610	Uncontaminated
P3	0.513	Uncontaminated
P4	0.567	Uncontaminated
P5	0.537	Uncontaminated
P6	0.484	Uncontaminated
P7	0.577	Uncontaminated
P8	0.373	Uncontaminated
P9	0.577	Uncontaminated
P10	0.400	Uncontaminated
P11	0.577	Uncontaminated
P12	0.389	Uncontaminated
P13	0.514	Uncontaminated
P14	0.472	Uncontaminated
P15	0.554	Uncontaminated
P16	0.484	Uncontaminated
P17	0.482	Uncontaminated

IV. RESULTS AND DISCUSSION

A. Results on the Case Study

It is showed, in Table XIII, that the 21 monitoring points resulted in an uncontaminated water quality, however, a quality level comparison can be made according to the maximum clustering coefficient (max. σ_i^k), like to shown in Table XIV and Fig. 5.

It is observed that the monitoring point P1 presents the best water quality and, the point P8, the lowest water quality. This happens because point P1 is on the beginning of the Huallaga river and point P8 is in a lower zone, it means that the quality of Water decreases along the river depending on the activities that take place, such as mining and hydroelectric plants [1].

TABLE XIV.   COMPARISON OF THE WATER QUALITY OF THE MONITORING POINTS ACCORDING TO THEIR MAX σ_i^k

Point	Max σ_i^k	Prati index	Color scale
1	0.639	Uncontaminated	Better water quality
2	0.610	Uncontaminated	
7	0.577	Uncontaminated	
9	0.577	Uncontaminated	
11	0.577	Uncontaminated	
18	0.577	Uncontaminated	
4	0.567	Uncontaminated	
15	0.554	Uncontaminated	
5	0.537	Uncontaminated	
19	0.524	Uncontaminated	
13	0.514	Uncontaminated	
3	0.513	Uncontaminated	
20	0.493	Uncontaminated	
6	0.484	Uncontaminated	
16	0.484	Uncontaminated	
17	0.482	Uncontaminated	
14	0.472	Uncontaminated	
21	0.470	Uncontaminated	
10	0.400	Uncontaminated	
12	0.389	Uncontaminated	
8	0.373	Uncontaminated	Lower water quality
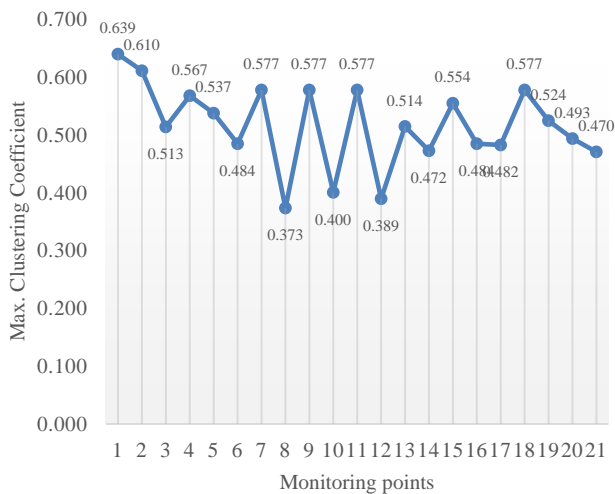
Fig. 5.   Max. σ_i^k of Monitoring Points.

The reason why the water is uncontaminated may be because mining and hydroelectric companies, which operate in the upper part of the Huallaga river, they treat their industrial effluents adequately in accordance with the regulations national. This water body is classified as suitable for the irrigation of vegetables and animals [1], but according to the results of the water quality assessment, it can also be considered as water that can be made drinkable with conventional treatment.

In relation to other studies, Fu and Zou [10], applied the Grey Clustering method to evaluate the water quality of the Yellow River, the results also showed good river quality. In the assessment of the water quality carried out by Liping et al. [11], applied the Grey Clustering method for the assessment of the quality of the Fenchuan river of the Yan'an Baota area in China, however, considered the arithmetic mean for the determination of the weights of clustering. In this case, Shannon Entropy could have been applied alternatively, as was done in the present study, to calculate these weights in an objective and precise way. Similarly, in the study carried out by Wang et al. [12] the clustering weights could be obtained through the Shannon Entropy method and be complemented by the Single Factor method used in this study.

B. Discussion on the Methodology

The Grey Clustering method is the most appropriate in high uncertainty issues [3] such as assessment of surface water quality where each parameter varies depending on environmental conditions, in comparison of classic multi-criteria assessment methods such as Delphi [13] or the Analytical Hierarchy Process (AHP) [14] which do not consider uncertainty in their analysis. In Peru, the Grey Clustering method is not very widespread compared to other logic methods Aristotelian or statistical models [7] which means a limitation for its application with the national water quality standards, which are not determined based on any quality index.

Finally, the Shannon entropy method is well suited for evaluating water quality because allowed to determine the grouping weights (η_j) for each parameter in an objective way, without the need to ask to an expert and this reduces assessment costs. In addition, this method has multiple applications as in studies of social conflicts or assessments of social impact [15], due to its great capacity to process information and reduce subjectivity in assessments.

V. CONCLUSIONS

The surface water quality of the upper watershed of Huallaga river could be evaluated using Grey Clustering method and Shannon Entropy, so it was possible to classify the 21 monitoring points in this area. The results obtained in this study can be useful to the regional and local authorities of Peru, as well as to the National Water Authority to make better decisions regarding the management of this important watershed.

According to the methodology, the Grey Clustering method can be more effective than other classical methods due to that considers the uncertainty within the analysis, regarding the Shannon Entropy it allows to calculate the weights objectively of the criteria without resorting to expert judgments. Another important point is that when using the Prati index, there is an advantage when we need to compare if the water quality is affected by the activities carried out in the watershed, due to the parameters used.

Finally, in future research, the efficacy of the Grey Clustering method should be compared with the methodology established by the National Water Authority (ANA) for the determination of the Quality Index Environment of Surface Water Resources (ICARHS). In case the results are indistinct or very similar, the use of the Grey Clustering method could be extended to those rivers where the data is insufficient to apply the ANA methodology.

REFERENCES

[1] Autoridad Nacional del Agua, "Informe del Monitoreo Participativo de la Calidad del Agua Superficial en la cuenca del río Huallaga – Febrero y Marzo del 2019," Tarapoto, 2019.

[2] Congreso de la República del Perú, "Pre Dictamen del Texto Sustitutorio de la Ley que declara de interés nacional y necesidad publica la recuperación, conservación y protección de las aguas de la cuenca del rio Huallaga," Lima, 2017. [Online]. Available: http://www.congreso.gob.pe/Docs/comisiones2016/PueblosAndinosEcologia/files/ppt_predictamen_huallaga_al_10.04.17.pdf.

[3] S. Liu and Y. Lin, Grey Systems: Theory and Applications, vol. 68. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[4] C. Zhu and Q. Liu, "Evaluation of Water Quality Using Grey Clustering," in 2009 Second International Workshop on Knowledge Discovery and Data Mining, Jan. 2009, pp. 803–805, doi: 10.1109/WKDD.2009.28.

[5] Y. Leng et al., "Evaluation on Transfer Efficiency at Integrated Transport Terminals through Multilevel Grey Evaluation," Procedia - Soc. Behav. Sci., vol. 43, pp. 587–594, 2012, doi: 10.1016/j.sbspro.2012.04.132.

[6] C. E. Shannon and W. Weaver, The Mathematical Theory of Communication. 1949.

[7] A. Delgado and I. Romero, "Environmental conflict analysis using an integrated grey clustering and entropy-weight method: A case study of a mining project in Peru," Environ. Model. Softw., vol. 77, pp. 108–121, Mar. 2016, doi: 10.1016/j.envsoft.2015.12.011.

[8] Y. Zhang, J. Ni, J. Liu, and L. Jian, "Grey evaluation empirical study based on center-point triangular whitenization weight function of Jiangsu Province industrial technology innovation strategy alliance," Grey Syst. Theory Appl., vol. 4, no. 1, pp. 124–136, Jan. 2014, doi: 10.1108/GS-11-2013-0027.

[9] A. Shemshadi, H. Shirazi, M. Toreihi, and M. J. Tarokh, "A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting," Expert Syst. Appl., vol. 38, no. 10, pp. 12160–12167, Sep. 2011, doi: 10.1016/j.eswa.2011.03.027.

[10] X. Q. Fu and Z. H. Zou, "Water Quality Evaluation of the Yellow River Basin Based on Gray Clustering Method," IOP Conf. Ser. Earth Environ. Sci., vol. 128, p. 012139, Mar. 2018, doi: 10.1088/1755-1315/128/1/012139.

[11] Wang Liping, Lai Kunrong, and Zhou Weibo, "Application of Grey Clustering method for water quality evaluation in fenchuan River Yan'an Baota Area," in 2011 International Symposium on Water Resource and Environmental Protection, May 2011, pp. 838–841, doi: 10.1109/ISWREP.2011.5893142.

[12] J. Wang et al., "Application of Grey Clustering Method Based on Improved Analytic Hierarchy Process in Water Quality Evaluation," MATEC Web Conf., vol. 246, p. 02004, Dec. 2018, doi: 10.1051/matecconf/201824602004.

[13] V. Mahajan, "The Delphi method: Techniques and applications," JMR, J. Mark. Res., vol. 13, no. 000003, p. 317, 1976.

[14] T. L. Saaty, The Analytic Hierarchy Process. New York, 1980.

[15] A. Delgado, "Social conflict analysis on a mining project using shannon entropy," in 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Aug. 2017, pp. 1–4, doi: 10.1109/INTERCON.2017.8079661.