p-ISSN 2012-3612 | e-ISSN 2782-8824
www.edjournal.cfnr-uplb.org

Ecosystems and Development Journal
Volume 11 | Numbers 1 & 2 | pp. 21–31 | 2021

Classification of watersheds in Occidental Mindoro and Oriental Mindoro using principal component analysis and k-means clustering

Aldous Xavier A. Palcon[1]

[1]*College of Forestry and Natural Resources, University of the Philippines Los Baños, College Laguna, 4031; Email: aldouspalcon23@gmail.com*

ABSTRACT. Classification of watersheds into similar groups before implementing interventions is essential since it is more systematic and sustainable than treating them individually. This study classified the watersheds within Occidental and Oriental Mindoro by applying principal component analysis (PCA) to selected characteristics of the 47 delineated watersheds and k-means clustering. The results of the PCA reduced 15 variables (area, circularity ratio, population, and 12 land cover classes) into four principal components using a threshold of 75% accounted variance of the original data. K-means clustering classified the watersheds into four clusters based on the principal components. Multivariate analysis of variance (MANOVA) was used to identify if there is a maximum between-cluster variation. Results showed that there was no association between the clusters and that the clustering of the watersheds is significant at a 5% level of significance. The results of the study may be used in choosing the most appropriate management models for each cluster.

Keywords: *dimensionality reduction technique, GIS, land cover, watershed classification, watershed management*

INTRODUCTION

Access to land and water resources decreases due to the increase in demand from the growing population. Associated with this is the increasing stresses on watershed functions and services vital to welfare. In view of this, watershed management needs to be robust and science-based. It will require good datasets and information, tested practices, and other essential resources. However, these management resources in many watersheds are not sufficient or available, such as datasets on water, soil, and other watershed resources. In addition, suited management practices for a watershed may not be present to allow prompt response to urgent concerns. While no two watersheds are exactly alike, similarities exist among watersheds that can be used to identify what management resources (*i.e.* datasets, information, and practices) may be commonly applied to a number of watersheds with similar features without the needed management resources.

Classifying watersheds into homogenous groups will facilitate science-based watershed management programs to promote soil and water conservation and other management objectives (Gajbhiye *et al*. 2013). Likewise, it will help address budget constraints that hamper the commonly challenging tasks of watershed management (Choubin *et al*. 2017). The need to analyze each watershed individually at once may be postponed for a later time when financial resources for conducting watershed and resources profiling and assessment are already available (Raju & Kumar 2011).

This study classified the delineated watersheds within Mindoro Island based on the cluster analysis results. The results can be used to set science-based management in many watersheds in Mindoro with limited historical datasets and information by applying those from a watershed with similar features.

Currently, there is a lack of a formal system for classifying watersheds (Kult 2013) since only a few studies about the interrelationships of variables characterize watersheds (Sharma *et al*. 2015). In a study conducted by Choubin *et al*. (2017), sub-watersheds in the Karkheh River watershed in Iran were clustered using fuzzy c-means clustering based on the watersheds' normalized difference vegetation index (NDVI), normalized difference moisture index (NDMI), soil adjusted vegetation index (SAVI), leaf area index (LAI), and snow cover. Burn & Boorman (1993) used k-means clustering to classify watersheds based on their similarity in flow response. According to Kanishka & Eldho (2017), dimensionality reduction techniques applied before clustering improve classification accuracy. In a study conducted by Nathan & McMahon (1990), principal component analysis (PCA), a dimensionality reduction technique, was used before clustering the 184 catchments in southern Australia. PCA's effectiveness was also proved by a study conducted by Di Prinzio *et al*. (2011) when 300 watersheds in Italy were subjected to PCA before applying self-organizing maps (SOM). In a study by Al-Shaikh *et al*. (2017), morphometric analysis was used to understand the geomorphic evolution, structure, and hydrological potential of 20 arid watersheds. Twenty-two morphometric parameters were reduced to five components using PCA. Q-mode cluster analysis was used to cluster the 20 watersheds based on the principal components. In all the studies mentioned, PCA reduced the dimensionality of the data so that the bulk of the variations within the data were captured in fewer dimensions.

MATERIALS AND METHODS

For the study area, the Occidental Mindoro and Oriental Mindoro provinces were selected. Both areas are located at 13.1162° north latitude and 121.0794° east longitude with a total land area of 10157.73 km² (**Figure 1**).

The 2015 Digital Elevation Model (DEM) with a 5 x 5 m resolution and land cover data for both sites were obtained from the National Mapping and Resource Information Authority (NAMRIA), while the 2015 population data were downloaded from the Philippine Statistics Authority (PSA) website. The land cover data provided 12 classes: annual crop, brush/shrub, built-up, closed forest, fishpond, grassland, inland water, mangrove forest, marshland/swamp, open forest, open/barren, and perennial crop.

Data were projected to have the same coordinate system (WGS 1984 UTM Zone 51N). To delineate the watersheds, hydrology tools in ArcGIS were used. Watersheds with areas greater than or equal to 40 km² were delineated and characterized. 40 km² was arbitrarily set to be the minimum area of the delineated watersheds.

To extract the area and circularity ratio of the watersheds, geoprocessing tools and Calculate Geometry were used. The watershed shape was expressed using its circularity ratio in this study.

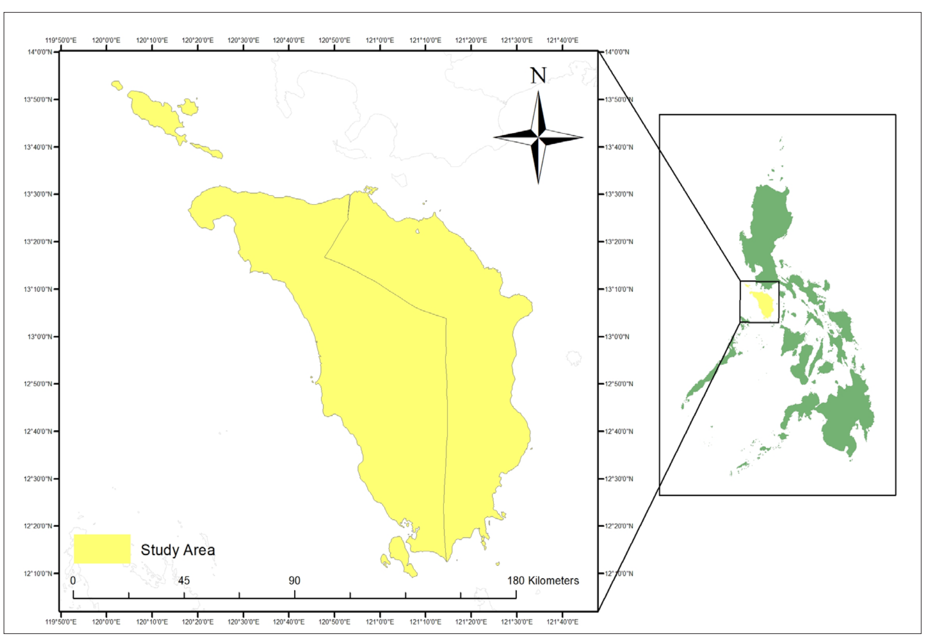

Figure 1. Location map of Occidental Mindoro and Oriental Mindoro.

To create a correlation matrix for principal component analysis, the values of the variables were first standardized by calculating the mean and standard deviation of each variable. Then, the mean was subtracted and divided by the standard deviation for each observed value. Then a correlation matrix was created using Pearson's correlation coefficient (Sharma *et al*. 2015).

The principal component loading matrix was created by multiplying the eigenvectors with the square root of the characteristic values of the correlation matrix. Eigenvalues greater than 1 were considered significant (Sharma *et al*. 2015).

To obtain simpler and more interpretable factors, the dimensions in principal components were rotated (Yaremko *et al*. 1986). There are two types of rotations: orthogonal and oblique. If the components in the PCA were uncorrelated, an orthogonal rotation would be used, while if the components were correlated, an oblique rotation method would be applied (Brown 2009). According to Tabachnick & Fidell (2007), an oblique rotation method is to be used if there are correlations greater than 0.32 among the factors. Promax rotation, an oblique rotation method, was used to rotate the eigenvalues.

To start the k-means clustering algorithm, the principal components were plotted. Then, the number of clusters, or k, was identified. Centroids for each cluster were randomly selected from the data where the number of centroids was the same as the number of clusters. Then, the points were assigned to the closest cluster centroid. The new centroids were then computed from the newly formed clusters. The points were assigned repeatedly to the closest cluster centroids and recomputed new centroids until the centroids of newly formed clusters did not change, points remained in the same cluster, or the maximum number of iterations was reached (Sharma 2019).

To determine the significance of groupings in the dataset, multivariate ANOVA (MANOVA) was used. MANOVA has four multivariate tests: Wilk's Lambda, Pillai's Trace, Lawley-Hotelling Trace, and Roy's Largest Root. Groups or clusters are significant if the p-value is less than the significance level, usually 0.05 (Carey 1998).

RESULTS AND DISCUSSION

Watersheds within Occidental and Oriental Mindoro that have areas at least 40 km² were delineated using hydrology tools on the DEM file in ArcGIS (**Figure 2**). These watersheds have a total area of 7,988.94 km² and occupy 79% of the total land area. Among the 47 delineated watersheds, Sablayan-Rizal Watershed 2 was the largest with 529.2 km², while the smallest was Abra de Ilog Watershed 1 with 41.6 km². The watershed with the highest circularity ratio was Sablayan Watershed 4 with 0.63 while the lowest circularity ratios were Sablayan-Pola Watershed and Mansalay-Magsaysay Watershed with 0.17 circularity ratio. Santa Cruz-Calapan Watershed has the most people with 103,713, while the least populated watershed was Paluan Watershed 3 with 1,204 (**Table 1**).

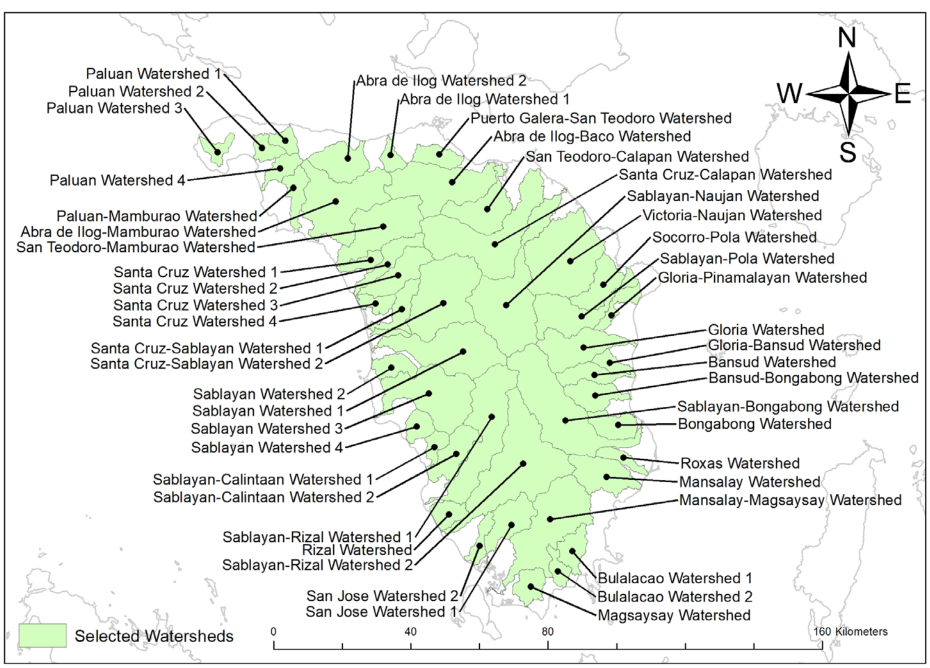

Figure 2. Analyzed watersheds in Occidental and Oriental Mindoro.

Table 1. Area (in km²), circularity ratio, and population of delineated watersheds.

No.	Watershed	Area (km²)	Circularity ratio	Population
1	Abra de Ilog-Baco Watershed	318.92	0.31	1,204
2	Abra de Ilog-Mamburao Watershed	365.90	0.28	1,219
3	Abra de Ilog Watershed 1	41.62	0.46	1,374
4	Abra de Ilog Watershed 2	141.51	0.47	1,670
5	Bansud-Bongabong Watershed	109.17	0.39	2,018
6	Bansud Watershed	84.61	0.24	2,101
7	Bongabong Watershed	93.60	0.41	2,392
8	Bulalacao Watershed 1	77.94	0.32	2,458
9	Bulalacao Watershed 2	45.14	0.31	2,473
10	Gloria-Bansud Watershed	42.54	0.36	2,487
11	Gloria-Pinamalayan Watershed	57.69	0.18	4,403
12	Gloria Watershed	196.35	0.31	4,829
13	Magsaysay Watershed	73.74	0.39	4,866
14	Mansalay-Magsaysay Watershed	377.43	0.17	4,869
15	Mansalay Watershed	159.02	0.36	6,393
16	Paluan-Mamburao Watershed	74.68	0.27	7,280
17	Paluan Watershed 1	42.44	0.46	7,764
18	Paluan Watershed 2	47.82	0.40	8,028
19	Paluan Watershed 3	41.90	0.31	8,344
20	Paluan Watershed 4	58.12	0.32	8,451
21	Puerto Galera-San Teodoro Watershed	70.71	0.43	9,904
22	Rizal Watershed	64.56	0.47	11,702
23	Roxas Watershed	45.47	0.27	12,005
24	Sablayan-Bongabong Watershed	449.83	0.24	13,355
25	Sablayan-Calintaan Watershed 1	44.43	0.33	13,663
26	Sablayan-Calintaan Watershed 2	126.08	0.34	15,446
27	Sablayan-Naujan Watershed	437.96	0.18	16,298
28	Sablayan-Pola Watershed	202.37	0.17	16,553
29	Sablayan-Rizal Watershed 1	370.22	0.20	19,783
30	Sablayan-Rizal Watershed 2	529.19	0.23	21,990
31	Sablayan Watershed 1	339.29	0.29	22,445
32	Sablayan Watershed 2	69.89	0.35	22,580
33	Sablayan Watershed 3	330.70	0.30	23,933
34	Sablayan Watershed 4	59.39	0.63	26,456
35	San Jose Watershed 1	133.23	0.26	28,678
36	San Jose Watershed 2	51.91	0.25	30,322
37	San Teodoro-Calapan Watershed	176.08	0.25	31,018
38	San Teodoro-Mamburao Watershed	269.15	0.33	34,721
39	Santa Cruz-Calapan Watershed	434.99	0.22	39,868
40	Santa Cruz-Sablayan Watershed 1	86.97	0.19	41,923
41	Santa Cruz-Sablayan Watershed 2	383.44	0.26	48,242
42	Santa Cruz Watershed 1	45.18	0.27	49,794
43	Santa Cruz Watershed 2	92.17	0.19	54,308
44	Santa Cruz Watershed 3	137.80	0.27	55,002
45	Santa Cruz Watershed 4	46.53	0.32	64,644
46	Socorro-Pola Watershed	119.57	0.32	78,804
47	Victoria-Naujan Watershed	421.68	0.46	103,713
	Total	7,988.94		991,776

Among the 12 land cover classes data measured, the most dominant land class within the delineated watersheds was brush/shrub, with an area of approximately 27.12% of the total watershed area. The least dominant land cover present within the delineated watersheds was the marshland/swamp, which occupies only 0.08% of the total watershed area (**Table 2**).

The standardization process produced standard scores that represent the number of standard deviations above or below the mean that a specific observation falls. A value of 0 indicates that the observation falls 0 standard deviations above the mean (**Table 3**). These prepared the data for principal component analysis.

The variables were tested for correlations using Pearson's correlation coefficient (**Table 4**). A correlation value with at least 0.50 has a high correlation. (A negative value means that it is indirectly correlated while a positive value means that it is directly correlated); if the value is between 0.3 and 0.49 it is moderately correlated; and if 0.29 and below it is lowly correlated.

The principal component loading matrix (**Table 5**) obtained from the correlation matrix indicates that the first four components (with eigenvalues greater than 1) together account for 75.32% of the total explained variance. **Table 6** shows the comparison between the rotated, unrotated, and unrotated loadings greater than or equal to 0.3 (threshold of the salience of a loading). Because there are variables with a correlation higher than 0.32 (**Table 2**), promax rotation, an oblique rotation, was used to rotate the significant eigenvalues from **Table 5** (**Table 6**). The first component was moderately correlated (loading between 0.5 and 0.7) with marshland/swamp and inland water and have a low correlation (loading between 0.3 and 0.5) with annual crop, perennial crop, and

Table 5. Principal component loading matrix of the 15 variables.

Variable	Comp1	Comp2	Comp3	Comp4
Area	0.40	-0.19	-0.02	0.02
Circularity ratio	-0.15	0.22	-0.27	0.06
Population	0.38	0.06	0.16	-0.09
Annual crop	0.36	0.11	0.13	0.20
Brush/ Shrub	0.20	-0.34	-0.19	0.21
Built-up	0.29	0.33	0.07	0.13
Closed forest	0.18	-0.17	0.24	-0.39
Fishpond	-0.07	0.11	0.42	0.44
Grassland	0.17	-0.37	-0.21	0.40
Inland water	0.32	0.28	-0.26	0.07
Mangrove forest	-0.02	0.11	0.58	0.35
Marshland/ Swamp	0.23	0.37	-0.32	0.13
Open forest	0.29	-0.16	0.25	-0.41
Open/ Barren	0.25	-0.37	-0.04	0.14
Perennial crop	0.23	0.34	0.04	-0.23
Eigenvalue	5.23	3.38	1.45	1.23
% of Total factor covariance	34.89	22.56	9.64	8.23
Cumulative % of total factor Covariance	34.89	57.45	67.09	75.32

Table 4. Pearson's correlation coefficient matrix of the 15 variables.

	A	CR	P	AC	B/S	B	CF	F	G	IW	MF	M/S	OF	O/B	PC
Area (A)	1.00														
Circularity ratio (CR)	-0.37	1.00													
Population (P)	0.74	-0.35	1.00												
Annual crop (AC)	0.64	-0.18	0.77	1.00											
Brush/ Shrubs (B/S)	0.70	-0.23	0.29	0.23	1.00										
Built-up (B)	0.37	-0.06	0.63	0.77	-0.10	1.00									
Closed forest (CF)	0.44	-0.12	0.31	0.26	0.18	0.05	1.00								
Fishpond (F)	-0.17	0.12	-0.07	0.04	-0.15	0.02	-0.12	1.00							
Grassland (G)	0.60	-0.33	0.17	0.20	0.63	-0.09	0.13	-0.15	1.00						
Inland water (IW)	0.50	0.08	0.57	0.63	0.08	0.73	0.10	-0.10	0.04	1.00					
Mangrove forest (MF)	-0.10	-0.02	0.01	0.13	-0.19	0.19	-0.01	0.31	-0.12	-0.07	1.00				
Marshland/ Swamp (M/S)	0.26	0.23	0.38	0.51	-0.08	0.71	-0.06	-0.04	-0.09	0.96	-0.06	1.00			
Open forest (OF)	0.74	-0.35	0.58	0.39	0.32	0.17	0.54	-0.14	0.13	0.24	-0.06	-0.01	1.00		
Open/ Barren (O/B)	0.73	-0.38	0.34	0.32	0.56	0.03	0.43	-0.17	0.82	0.12	-0.12	-0.09	0.49	1.00	
Perennial crop (PC)	0.30	-0.02	0.60	0.36	-0.19	0.64	0.02	-0.04	-0.27	0.63	0.05	0.61	0.26	-0.16	1.00

Table 2. Percent land cover of delineated watersheds.

Watershed no.	%Annual crop	%Brush/ Shrub	%Built-up	%Closed forest	% Fishpond	% Grassland	%Inland water	%Mangrove forest	%Marshland/ Swamp	%Open forest	%Open/ Barren	%Perennial crop
1	6.63	15.07	0.64	6.56	0.00	1.02	1.12	0.03	0.00	47.52	0.07	21.33
2	13.53	55.25	0.21	2.26	0.00	20.10	1.72	0.09	0.00	6.12	0.25	0.47
3	0.00	33.05	0.00	31.49	0.00	9.24	0.25	0.00	0.00	24.62	0.00	1.36
4	12.12	49.81	0.73	0.00	0.04	20.48	1.28	0.07	0.00	12.29	1.34	1.85
5	10.97	7.57	0.55	0.00	0.00	12.74	2.10	0.02	0.00	17.85	0.02	48.19
6	10.89	5.22	0.72	0.00	0.00	13.98	3.72	0.05	0.00	23.84	0.00	41.57
7	21.61	6.56	2.36	0.00	0.00	6.58	0.21	0.10	0.00	0.89	0.00	61.68
8	13.41	28.52	1.47	0.00	0.40	3.75	0.11	0.33	0.00	0.56	0.00	51.46
9	10.89	38.11	1.53	0.00	0.63	5.51	0.19	0.00	0.00	20.12	0.00	23.01
10	51.09	0.00	3.84	0.00	0.07	7.77	0.19	0.04	0.00	0.00	0.02	36.97
11	54.72	0.00	5.46	0.00	1.69	0.68	0.30	0.71	0.00	0.21	0.06	36.17
12	4.25	9.42	0.48	0.00	0.00	9.01	2.72	0.00	0.00	40.93	1.21	31.97
13	34.21	37.33	0.60	0.00	7.06	4.66	0.30	1.02	0.00	14.38	0.04	0.37
14	9.06	39.01	0.75	0.00	0.03	12.34	1.38	0.02	0.00	21.11	0.34	15.98
15	12.13	25.73	0.77	0.00	0.55	8.51	0.98	0.01	0.00	19.93	0.21	31.17
16	3.25	48.60	0.03	0.00	0.00	24.56	1.97	0.00	0.00	19.16	1.74	0.68
17	0.33	51.32	0.00	0.00	0.00	13.61	0.41	0.00	0.00	34.33	0.00	0.00
18	10.49	58.06	0.81	0.00	0.00	8.34	1.46	0.00	0.00	20.30	0.54	0.00
19	2.91	55.62	0.00	0.00	0.00	16.48	0.32	0.00	0.00	24.62	0.04	0.00
20	1.11	41.82	0.00	0.00	0.00	23.46	1.03	0.00	0.00	31.92	0.68	0.00
21	2.24	5.96	0.41	14.80	0.02	0.76	1.32	0.07	0.00	23.73	0.00	50.70
22	62.62	15.79	3.40	0.00	1.61	14.24	0.49	0.58	0.00	0.00	0.01	1.26
23	40.16	5.61	6.83	0.00	0.10	1.17	0.04	0.08	0.00	0.00	0.01	45.99
24	2.07	31.27	0.26	0.20	0.00	6.49	3.09	0.00	0.00	43.31	0.71	12.59
25	22.27	55.24	0.60	0.00	0.00	3.19	0.41	0.00	0.00	18.05	0.00	0.24
26	18.66	59.94	0.32	0.00	0.00	14.51	1.19	0.09	0.00	4.77	0.26	0.26
27	14.50	4.72	0.76	3.82	0.01	1.48	2.92	0.09	0.00	66.32	0.88	4.48
28	5.53	4.48	1.36	0.00	0.08	0.21	2.42	0.12	0.00	34.19	0.22	51.38
29	8.46	18.56	0.31	0.52	0.00	63.28	1.72	0.00	0.00	5.26	1.81	0.07
30	13.37	33.26	0.57	0.00	0.00	34.56	1.37	0.00	0.00	15.54	1.08	0.25
31	7.78	29.22	0.56	0.00	0.00	41.16	1.63	0.04	0.00	17.34	2.27	0.00
32	42.29	47.10	3.77	0.00	0.00	4.43	0.63	0.73	0.00	0.00	0.06	1.00
33	12.33	47.38	0.71	0.00	0.00	15.94	1.53	0.00	0.00	21.14	0.91	0.04
34	12.36	65.30	0.51	0.00	0.00	0.66	0.28	0.00	0.00	19.40	0.00	1.49
35	19.39	61.03	1.20	0.00	0.00	11.85	0.90	0.00	0.00	5.16	0.35	0.13
36	62.99	15.18	12.56	0.00	0.75	5.12	0.57	0.00	0.00	0.00	0.00	2.83
37	25.57	4.50	1.62	0.00	0.05	0.45	1.15	0.27	0.00	32.54	0.41	33.43
38	2.69	39.76	0.07	0.55	0.00	31.25	1.76	0.00	0.00	23.34	0.50	0.08
39	23.85	11.58	1.25	6.04	0.00	9.02	3.56	0.12	0.00	35.12	1.18	8.28
40	24.48	22.82	0.94	0.00	0.00	48.27	1.55	0.01	0.00	0.00	1.90	0.04
41	2.98	34.96	0.17	7.89	0.00	30.46	1.88	0.00	0.00	19.70	1.97	0.00
42	20.95	47.65	0.66	0.00	0.00	27.38	2.67	0.00	0.00	0.00	0.08	0.60
43	10.42	32.23	0.20	0.00	0.00	46.47	2.58	0.00	0.00	6.92	0.72	0.44
44	7.31	44.17	0.00	0.00	0.00	42.04	2.53	0.12	0.00	2.61	1.09	0.13
45	30.96	15.21	2.81	0.00	0.79	41.62	1.33	6.54	0.00	0.00	0.00	0.74
46	27.73	0.65	5.72	0.00	1.02	0.00	0.41	1.59	0.00	0.00	0.00	62.90
47	24.52	4.30	3.03	0.00	0.00	0.17	20.18	0.00	1.59	8.14	0.01	38.06
Total area (%)	13.84	27.12	1.05	1.63	0.14	17.45	2.73	0.13	0.08	21.89	0.74	13.20

Table 3. Standardized data of delineated watersheds

Watershed no.	Area	Circularity ratio	Population	Annual crop	Brush/ Shrub	Built-up	Closed forest	Fishpond	Grassland	Inland water	Mangrove forest	Marshland/ Swamp	Open forest	Open/ Barren	Perennial crop
1	0.52	0.22	0.29	0.20	0.24	0.16	0.69	0.00	0.01	0.04	0.04	0.00	0.52	0.03	0.42
2	0.61	0.17	0.32	0.48	1.00	0.06	0.27	0.00	0.31	0.07	0.10	0.00	0.08	0.12	0.01
3	0.00	0.45	0.01	0.00	0.07	0.00	0.43	0.00	0.02	0.00	0.00	0.00	0.04	0.00	0.00
4	0.19	0.48	0.07	0.17	0.35	0.08	0.00	0.01	0.12	0.02	0.03	0.00	0.06	0.25	0.02
5	0.13	0.35	0.18	0.12	0.04	0.05	0.00	0.00	0.06	0.03	0.01	0.00	0.07	0.00	0.33
6	0.08	0.12	0.15	0.09	0.02	0.05	0.00	0.00	0.05	0.04	0.01	0.00	0.07	0.00	0.22
7	0.10	0.38	0.12	0.19	0.03	0.17	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.36
8	0.07	0.24	0.07	0.10	0.11	0.09	0.00	0.06	0.01	0.00	0.08	0.00	0.00	0.00	0.25
9	0.01	0.22	0.03	0.05	0.09	0.05	0.00	0.05	0.01	0.00	0.00	0.00	0.03	0.00	0.06
10	0.00	0.30	0.05	0.21	0.00	0.13	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.10
11	0.03	0.02	0.20	0.30	0.00	0.25	0.00	0.19	0.00	0.00	0.14	0.00	0.00	0.00	0.13
12	0.29	0.22	0.28	0.08	0.09	0.07	0.00	0.00	0.08	0.06	0.00	0.00	0.28	0.31	0.39
13	0.06	0.35	0.08	0.24	0.14	0.03	0.00	1.00	0.01	0.00	0.25	0.00	0.04	0.00	0.00
14	0.63	0.00	0.45	0.33	0.73	0.22	0.00	0.02	0.20	0.06	0.02	0.00	0.27	0.17	0.38
15	0.22	0.31	0.26	0.19	0.20	0.10	0.00	0.17	0.06	0.02	0.01	0.00	0.11	0.04	0.31
16	0.06	0.16	0.04	0.02	0.18	0.00	0.00	0.00	0.08	0.02	0.00	0.00	0.05	0.17	0.00
17	0.00	0.46	0.00	0.00	0.11	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.05	0.00	0.00
18	0.01	0.37	0.00	0.05	0.14	0.03	0.00	0.00	0.02	0.01	0.00	0.00	0.03	0.03	0.00
19	0.00	0.22	0.00	0.01	0.12	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.04	0.00	0.00
20	0.03	0.24	0.00	0.01	0.12	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.06	0.05	0.00
21	0.05	0.41	0.06	0.02	0.02	0.02	0.35	0.00	0.00	0.01	0.02	0.00	0.06	0.00	0.22
22	0.04	0.49	0.12	0.39	0.05	0.17	0.00	0.20	0.04	0.00	0.12	0.00	0.00	0.00	0.01
23	0.01	0.16	0.20	0.18	0.01	0.24	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.13
24	0.77	0.12	0.51	0.09	0.70	0.09	0.03	0.00	0.12	0.16	0.00	0.00	0.67	0.42	0.35
25	0.01	0.25	0.01	0.10	0.12	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.03	0.00	0.00
26	0.16	0.27	0.07	0.23	0.37	0.03	0.00	0.00	0.08	0.02	0.04	0.00	0.02	0.04	0.00
27	0.75	0.01	0.37	0.61	0.10	0.26	0.55	0.01	0.03	0.15	0.13	0.00	1.00	0.50	0.12
28	0.30	0.00	0.52	0.11	0.04	0.22	0.00	0.03	0.00	0.06	0.08	0.00	0.24	0.06	0.65
29	0.62	0.05	0.22	0.30	0.34	0.09	0.06	0.00	1.00	0.07	0.00	0.00	0.07	0.87	0.00
30	0.92	0.09	0.61	0.68	0.87	0.24	0.00	0.00	0.78	0.09	0.00	0.00	0.28	0.75	0.01
31	0.56	0.19	0.10	0.25	0.49	0.15	0.00	0.00	0.60	0.06	0.05	0.00	0.20	1.00	0.00
32	0.05	0.29	0.01	0.28	0.16	0.21	0.00	0.00	0.01	0.00	0.17	0.00	0.00	0.01	0.00
33	0.55	0.21	0.10	0.39	0.78	0.18	0.00	0.00	0.23	0.06	0.00	0.00	0.24	0.39	0.00
34	0.03	0.73	0.01	0.07	0.19	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.01
35	0.17	0.15	0.39	0.25	0.40	0.12	0.00	0.00	0.07	0.01	0.00	0.00	0.02	0.06	0.00
36	0.02	0.14	0.15	0.32	0.04	0.51	0.00	0.07	0.01	0.00	0.00	0.00	0.00	0.00	0.01
37	0.25	0.12	0.47	0.43	0.04	0.22	0.00	0.02	0.00	0.02	0.15	0.00	0.20	0.09	0.37
38	0.43	0.25	0.21	0.07	0.53	0.01	0.05	0.00	0.36	0.06	0.00	0.00	0.22	0.18	0.00
39	0.74	0.08	0.99	1.00	0.25	0.42	0.87	0.00	0.17	0.18	0.18	0.00	0.53	0.67	0.22
40	0.09	0.04	0.03	0.21	0.10	0.06	0.00	0.00	0.18	0.02	0.00	0.00	0.00	0.21	0.00
41	0.65	0.15	0.14	0.11	0.66	0.05	1.00	0.00	0.50	0.08	0.00	0.00	0.26	0.98	0.00
42	0.01	0.16	0.01	0.09	0.11	0.02	0.00	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.00
43	0.10	0.04	0.04	0.09	0.15	0.01	0.00	0.00	0.18	0.03	0.00	0.00	0.02	0.09	0.00
44	0.18	0.16	0.06	0.10	0.30	0.00	0.00	0.00	0.25	0.04	0.05	0.00	0.01	0.19	0.00
45	0.01	0.25	0.01	0.14	0.03	0.10	0.00	0.07	0.08	0.01	1.00	0.00	0.00	0.00	0.00
46	0.15	0.23	0.24	0.32	0.00	0.54	0.00	0.23	0.00	0.01	0.62	0.00	0.00	0.00	0.47
47	0.72	0.46	0.75	1.00	0.09	1.00	0.00	0.00	0.00	1.00	0.00	1.00	0.12	0.00	1.00

Table 6. Unrotated, rotated, and significant rotated loadings.

Variable	Unrotated				Rotated				Rotated (≥0.3)			
	Comp1	Comp2	Comp3	Comp4	Comp1	Comp2	Comp3	Comp4	Comp1	Comp2	Comp3	Comp4
Area	0.40	-0.19	-0.02	0.02	0.14	0.30	0.24	-0.05		0.30		
Circularity ratio	-0.15	0.22	-0.27	0.06	0.16	-0.09	-0.31	-0.13			-0.31	
Population	0.38	0.06	0.16	-0.09	0.22	0.01	0.31	0.07			0.31	
Annual crop	0.36	0.11	0.13	0.20	0.30	0.13	0.08	0.22	0.30			
Brush/ Shrub	0.20	-0.34	-0.19	0.21	0.00	0.48	-0.03	-0.09		0.48		
Built-up	0.29	0.33	0.07	0.13	0.41	-0.06	0.00	0.18	0.41			
Closed forest	0.18	-0.17	0.24	-0.39	-0.14	-0.10	0.52	-0.08			0.52	
Fishpond	-0.07	0.11	0.42	0.44	-0.03	0.02	-0.17	0.62				0.62
Grassland	0.17	-0.37	-0.21	0.40	0.00	0.60	-0.17	0.01		0.60		
Inland water	0.32	0.28	-0.26	0.07	0.50	0.04	-0.08	-0.13	0.50			
Mangrove forest	-0.02	0.11	0.58	0.35	-0.06	-0.07	0.00	0.68				0.68
Marshland/ Swamp	0.23	0.37	-0.32	0.13	0.53	0.00	-0.22	-0.12	0.53			
Open forest	0.29	-0.16	0.25	-0.41	-0.08	-0.08	0.59	-0.09			0.59	
Open/ Barren	0.25	-0.37	-0.04	0.14	-0.05	0.44	0.13	-0.02		0.44		
Perennial crop	0.23	0.34	0.04	-0.23	0.33	-0.28	0.19	-0.06	0.33			

built-up which was termed as the crop areas component. The second component was moderately correlated with grassland and has a low correlation with the area, brush/ shrub, and open/barren, which were termed the grassland areas component. The third component was moderately correlated with open and closed forests and low correlation with circularity ratio and population and was called the forest areas component. The fourth and last component has a moderate correlation with fishpond and mangrove forest and was termed as the water areas component (**Table 7**).

Table 7. Grouping of variables into components.

Component	Variables
Component 1 (Crop areas)	Annual crop, Built-up, Inland water, Marshland/Swamp, Perennial crop
Component 2 (Grassland areas)	Area, Brush/ Shrub, Grassland, Open/Barren
Component 3 (Forest areas)	Circularity ratio, Population, Closed forest, Open forest
Component 4 (Water areas)	Fishpond, Mangrove forest

K-means clustering

Table 8 shows the results from the k-means clustering, which was then mapped in **Figure 3**. The scatterplot matrix (**Figure 4)** visualizes the bivariate relationships among the principal components. The clustering of the watersheds was also visible in the scatterplots, concentrated in one area. Cluster 4 has the highest number of watersheds classified into it with 18 watersheds or 38.3% of the delineated watersheds while there were only three watersheds classified to Cluster 2 which was only 6.38% of the total number of the delineated watersheds (**Table 9**). The difference in variable means (**Table 10**) seems to be statistically significant. For example, in inland water, cluster means were different from each other. Cluster 1 has mean inland water of 14.48, while Clusters 2, 3, and 4 have 0.44, 1.57, and 1.17, respectively.

MANOVA was used to determine the significance of the clustering (**Table 11**). Since all p-values were less than 0.05, clustering is significant. It means there is a maximum between-clusters variation. To avoid suspicion caused by data transformation, MANOVA was tested using the original data, transformed data, and PCA data. In all those datasets, clustering was concluded to be significant at 5% level of significance.

Table 9. Distribution of watersheds by cluster.

Cluster	Frequency	Percent (%)	Cumulative (%)
1	12	25.53	25.53
2	3	6.38	31.91
3	14	29.79	61.70
4	18	38.30	100.00

Table 8. Final cluster membership of Mindoro watersheds.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Abra de Ilog-Baco Watershed	Magsaysay Watershed	Bansud Watershed	Abra de Ilog Watershed 1
Abra de Ilog-Mamburao Watershed	Santa Cruz Watershed 4	Bansud-Bongabong Watershed	Abra de Ilog Watershed 2
Mansalay-Magsaysay Watershed	Socorro-Pola Watershed	Bongabong Watershed	Bulalacao Watershed 2
Sablayan Watershed 1		Bulalacao Watershed 1	Gloria-Bansud Watershed
Sablayan Watershed 3		Gloria Watershed	Paluan Watershed 1
Sablayan-Bongabong Watershed		Gloria-Pinamalayan Watershed	Paluan Watershed 2
Sablayan-Naujan Watershed		Mansalay Watershed	Paluan Watershed 3
Sablayan-Rizal Watershed 1		Rizal Watershed	Paluan Watershed 4
Sablayan-Rizal Watershed 2		Roxas Watershed	Paluan-Mamburao Watershed
Santa Cruz-Calapan Watershed		Sablayan Watershed 2	Puerto Galera-San Teodoro Watershed
Santa Cruz-Sablayan Watershed 2		Sablayan-Pola Watershed	Sablayan Watershed 4
Victoria-Naujan Watershed		San Jose Watershed 1	Sablayan-Calintaan Watershed 1
		San Jose Watershed 2	Sablayan-Calintaan Watershed 2
		San Teodoro-Calapan Watershed	San Teodoro-Mamburao Watershed
			Santa Cruz Watershed 1
			Santa Cruz Watershed 2
			Santa Cruz Watershed 3
			Santa Cruz-Sablayan Watershed 1

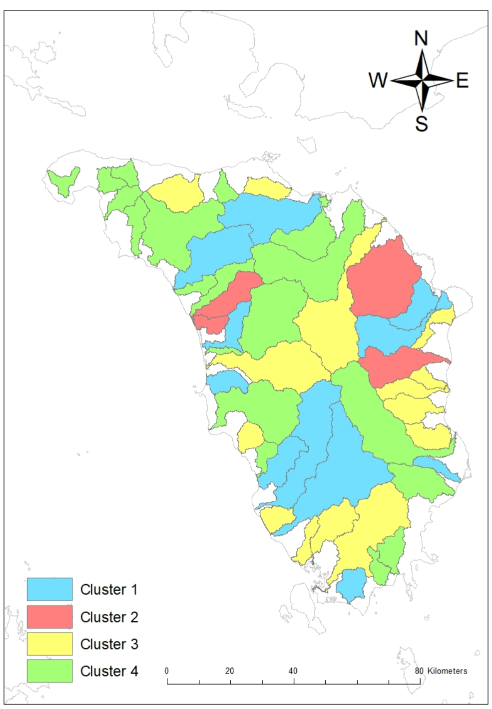

Figure 3. Cluster map of the watersheds.

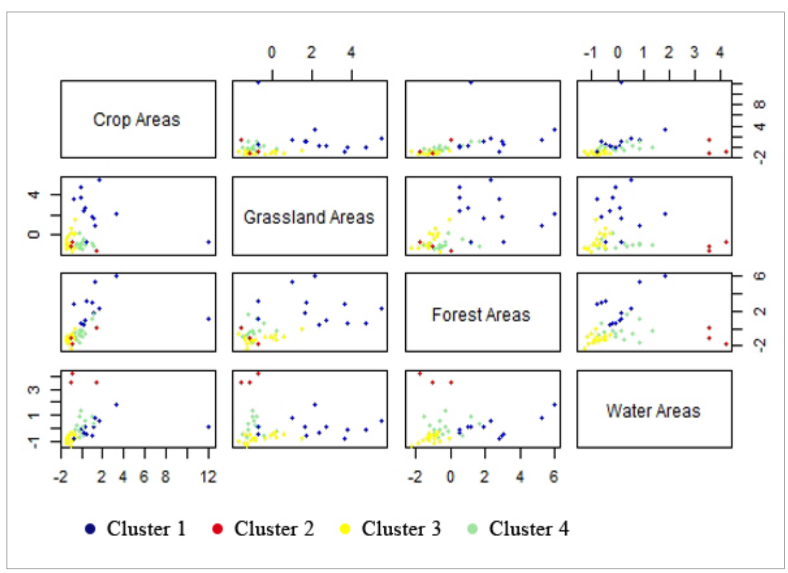

Figure 4. Cluster membership of watersheds to four clusters using k-means clustering.

Table 10. Mean of variables per cluster.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Area	396.63	60.35	79.95	36.92	108.71	54.89	81.54	57.73
Circularity ratio	0.26	0.08	0.34	0.04	0.30	0.09	0.35	0.11
Population	43,200.39	28,406.49	12,939.32	12,283.70	24,330.23	15,381.52	5,218.33	5,016.55
Annual crop	47.12	32.29	24.26	9.41	22.43	11.99	8.52	7.73
Brush/ Shrub	105.15	62.73	11.79	13.99	18.02	21.66	34.28	27.37
Built-up	3.12	3.31	2.86	3.47	2.25	1.53	0.36	0.44
Closed forest	8.77	11.57	0.00	0.00	0.00	0.00	1.39	3.82
Fishpond	0.01	0.03	2.26	2.58	0.28	0.39	0.02	0.07
Grassland	77.11	75.63	7.60	10.33	7.07	6.38	19.28	23.41
Inland water	14.48	22.57	0.44	0.20	1.57	1.79	1.17	1.29
Mangrove forest	0.13	0.18	1.90	1.14	0.18	0.19	0.02	0.05
Marshland/ Swamp	0.56	1.94	0.00	0.00	0.00	0.00	0.00	0.00
Open forest	102.67	80.55	3.54	6.12	20.46	28.43	12.18	13.92
Open/ Barren	3.78	2.80	0.01	0.02	0.32	0.64	0.52	0.68
Perennial crop	33.72	47.94	25.28	43.24	36.13	30.67	3.78	9.05

Table 11. MANOVA table of clusters formed from k-means clustering.

Test	Statistic	df	F(df1,	df2)	F	Prob>F (p-value)
Wilk's Lambda	0.0237	3	12	106.1	27.53	0.000
Pillai's Trace	1.7344		12	126	14.39	0.000
Lawley-Hotelling Trace	11.3236		12	116	36.49	0.000
Roy's Largest Root	7.8085		4	42	81.99	0.000

SUMMARY AND CONCLUSION

A total of 47 watersheds (with areas of at least 40 km^2) were delineated and had their characteristics described using 2015 land cover data and DEM obtained from NAMRIA and 2015 population data from PSA. The 15 variables (area, circularity ratio, population, and the 12 land cover classes) were reduced to four through principal component analysis. K-means clustering analysis classified the 47 watersheds into four clusters using the four principal components as its variables. Clustering is concluded to be significant at a 5% level of significance. These results may be used to choose the most appropriate management models for each of the four clusters since analyzing each watershed individually is more costly and labor-intensive.

Although the result was statistically accepted, there are still many variables like drainage basin parameters that could have been used for this study. In addition, the watersheds in this study only tested and verified the clustering of Occidental and Oriental Mindoro watersheds with areas greater than 40 km^2. Studies to apply this to the smaller watersheds are a focus of future research.

LITERATURE CITED

Al-Shaikh, N., Ayed, A., & Farhan, Y. (2017) *Multivariate Statistical Analysis of Hydro-Morphometric Parameters of Some Arid Watersheds, Jordan Using ASTER DEM and GIS*. Applied Morphometry and Watershed Management Using RS, GIS, and Multivariate Statistics, 155.

Brown, J. (2009) *Choosing the Right Type of Rotation in PCA and EFA. Shiken*: JALT Testing & Evaluation SIG Newsletter, 20–25.

Burn, D. & Boorman, D. (1993) Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology* 143: 429–454. https://doi.org/10.1016/0022-1694(93)90203-L.

Carey, G. (1998) *Multivariate Analysis of Variance (MANOVA): I. Theory*. Retrieved from: < http://ibgwww.colorado.edu/~carey/p7291dir/handouts/manova1.pdf>

Choubin, B., Solaimani, K., Roshan, M.H., & Malekian, A. (2017) Watershed classification by remote sensing indices: A fuzzy c-means clustering approach. *Journal of Mountain Science* 14: 2053–2063. https://doi.org/10.1007/s11629-017-4357-4.

Di Prinzio, M., Castellarin, A., & Toth, E. (2011) Data-driven catchment classification: application to the pub problem. *Hydrology and Earth System Sciences* 15: 1921–1935. https://doi.org/10.5194/hess-15-1921-2011.

Gajbhiye, S., Mishra, S., & Pandey, A. (2013) Prioritizing erosion-prone area through morphometric analysis: an RS and GIS perspective. *Applied Water Science* 4: 51–61. https://doi.org/10.1007/s13201-013-0129-7.

Kanishka, G. & Eldho, T.I. (2017) Watershed classification using isomap technique and hydrometeorological attributes. *Journal of Hydrologic Engineering*. doi:10.1061/(ASCE)HE.1943-5584.0001562.

Kult, J. (2013) *Regionalization of Hydrologic Response in the Great Lakes Basin: Considerations of Temporal Variability*. ProQuest Dissertations Publishing.

Nathan, R. & McMahon, T. (1990) Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology* 121: 217–238. doi:https://doi.org/10.1016/0022-1694(90)90233-N

Raju, K. & Kumar, D. (2011) Classification of microwatersheds based on morphological characteristics. *Journal of Hydro-Environment Research* 5(2): 101–109. doi:https://doi.org/10.1016/j.jher.2010.09.002.

Sharma, P. (2019) *The Most Comprehensive Guide to K-Means Clustering You'll Ever Need*. Retrieved from: <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>.

Sharma, S., Gajbhiye, S., & Tignath, S. (2015) Application of principal component analysis in grouping geomorphic parameters of a watershed for hydrological modeling. *Applied Water Science* 5: 89–96. https://doi.org/10.1007/s13201-014-0170-1.

Tabachnick, B. & Fidell, L. (2007) *Using Multivariate Statistics*, 5th edn. Boston. MA: Pearson Education Inc.

Yaremko, R., Harari, H., Harrison, R., & Lynn, E. (1986) *Handbook of Research and Quantitative Methods in Psychology: For Students and Professionals*. (1st ed.). Psychology Press. https://doi.org/10.4324/9780203767740.