USE OF ARTIFICIAL NEURAL NETWORKS AND MULTIVARIATE STATISTICAL ANALYSIS FOR MODELING THE POLLUTION PRESSURE OF WATER RESOURCES IN THE SEYBOUSE VALLEY (NORTH-EASTERN ALGERIA)

*Aissam GHRIEB[1], Fethi BAALI[2], Chemceddine FEHDI[2], Azzedine HANI[3], Hicham CHAFFAI[3]& Larbi DJABRI[3]

[1] *Department of Earth Sciences, Faculty of Exact Sciences, Science of Nature and Life, University of Larbi Tebessi, Tébessa, Algeria.* aissam.ghrieb@univ-tebessa.dz

[2] Water and Environment Laboratory, Department of Earth Sciences, Faculty of Exact Sciences, Science of Nature and Life, University of Larbi Tebessi, Tébessa, Algeria. fethi.baali@univ-tebessa.dz , fehdi@yahoo.fr

[3] *Laboratory of Water Resource and Sustainable Development (REDD), Department of Geology, Faculty of Earth Sciences, Badji Mokhtar University, Annaba, Algeria.* haniazzedine@yahoo.fr , hichamchaffai@yahoo.fr , djabri_larbi@yahoo.fr

Corresponding author

Received 3rd January 2021, accepted 30th March 2021

Abstract
The water supply environment in Seybouse Valley (North East of Algeria) is sensitive and fragile as the aquifer is highly vulnerable to various sources of pollution, must recognize the pollution sources and water quality integration. So, there is a need for a better knowledge and understanding of the water pollution determinants to meet the Domestic, agricultural and Industrial uses.

The pollution of this ground water was determined by Total Dissolved Solids (TDS). This represents the salinity of freshwater and originate from natural sources, sewage, urban, runoff, industrial wastewater and chemicals.

Based on cause-and-effect relationships, the Driver–Pressure–State–Impact–Response (DPSIR) plan was used to establish indicators for an integrated water resource management approach to water quality in the semi-arid Mediterranean region. The aim of this work is to determine the most pressing pollution source of Seybouse Valley. With this intention, the artificial neural network (ANN) models were used to model and predict the relationship between groundwater quality with point and diffuse pollution sources determinants. The selected variables were classified and organized using the multivariate techniques of Hierarchical cluster analysis (HCA), factor analysis (FA), principal components and classification analysis (PCCA).

It was concluded that the industrial wastewater that is the most pressing pollution source followed by seawater intrusion.

Keywords: *DPSIR Model; ANN; Multivariate techniques; Seybouse Valley.*

1. Introduction

Two recent problems in water systems planning have been a shortage of water resources and optimal control. Population development, the loss of available water supply, changes in lifestyle, the rising rate of use, climate change, and a number of other factors have made usable water a major concern for the future [1]. The

Seybouse Valley is currently facing the problem of water shortage. Over the next twenty years, the demand for drinking water is expected to increase by 20% [2].

The establishment of sustainable water management, both qualitatively and quantitatively, is therefore imperative for the future of the population of the Seybouse Valley, but also of ecosystems, and

conditions economic and social development.

The Seybouse Valley is also confronted with urban, industrial and agricultural pollution. This pollution comes from the various agglomerations, factories and agricultural areas located on both banks of the wadi.

The pollution has reached a high degree. 4.5 million m^3 are discharged annually into the river, of which 3 million m^3 are used oils [3].

The insufficient treatment of wastewater, the proliferation of urban waste, atmospheric pollution and various industrial pollution significantly affect the health of citizens with adverse consequences on low-income groups living in slums or in areas unattractive (proximity to landfills, wastewater discharge areas, etc.).

The rehabilitation of the Seybouse Valley can be a good example of integrated management, since it would include aspects of protection of surface and underground resources, rationalization of domestic, agricultural and industrial uses, but also the fight against pollution and protection of the environment.

Only an integrated approach offers the possibility of managing these resources while respecting the natural environment, the interests of citizens and those of economic actors. The research envisaged responds to this need for integration, from the location and evaluation of underground resources, to the protection of their quality, but while evaluating the possible harmful effects.

They thus contribute to the implementation of the current sustainable development policy, which is deployed through a number of directives, including that on water. But beyond that, they are part of the plan adopted at the Johannesburg Sustainable Development Summit.

The development of a model to predict understanding would help to manage water resources effectively. There are different methods for data analyses, such as statistic techniques. Prediction of water quality parameters necessitates a thorough examination of the various processes that can influence water quality, as well as the creation of mathematical or deterministic models based on the data collected [4]. Second, developing Data Driven Models using information and collected data is an essential tool.

The study area climate is semi-arid Mediterranean; the annual rainfall is varies between 700 and 900 mm. The potential evapotranspiration is closely linked to the temperatures. The annual average of the evapotranspiration ranges from 1000 to 2000 mm [5].

The conservation and management of these resources is becoming increasingly necessary as the world's population and industrial demands grow. The intensive agricultural activities in the plain of Annaba (Algeria) induce the increase of the risk of the fresh water degradation [6]. In fact, uncontrolled high pumping rates of water causes modification of the natural flow system and induces seawater flow from the coast making the deterioration of the water table quality [7].

There are many water quality parameters, but TDS is an important water quality parameter especially in reservoirs. In fact, TDS can be considered a pollutant for this reason, it is vital to have information about the existing situation, seasonal variations and expectations of the future situation of the parameter [4].

Domestic solid waste, Domestic wastewater of the municipalities, Industrial wastewater, Pesticides, Organic fertilizers, Chemical fertilizers, Petrol stations, Carbon dioxide (CO_2) and Seawater intrusion are the main pollution point sources of the Seybouse Valley.

The coastal aquifer system supplies the water demands of municipalities, several

villages, thousands of hectares of agricultural lands, and several industries. Ever increasing population, resulting in a high in domestic water demands as well as industrial demand including but not limited to development of agricultural networks, and inter-basin water transfers, darken the outlook for water quality of the Seybouse Valley.

The principal purpose of the study is to develop ANN model studying the relation between Total Dissolved Solids from coastal aquifer (represented by TDS (mgl^{-1}) and the main pollution sources.

Understanding spatial relations between hydrological variables and salinity of groundwater can contribute in an integration of water resources management. This research might be considered as one of the few contributions in qualitative modelling of the relation between groundwater salinity and the pollution pressure variables in spatial scale using ANN.

2. Materials and methods
2.1. Study area
The Seybouse River basin is situated in Algeria's northern east, and the coastal portion is the western part of the Annaba plain.

Because of the length of its course, the number of tributaries, and the size of its basin, the Seybouse is one of Algeria's most significant rivers. The Seybouse River basin is divided into three sections: the high plains (High Seybouse), the mean Seybouse, and the maritime Seybouse. The latter denotes the research field.

About 36°30' and 37° North latitude and 7°30' and 7°55' East longitude, the research region is situated in North-East Algeria.

With an area of 103 km², the research area is part of the Seybouse river basin, and its natural boundaries are the Mediterranean Sea in the north, the eastern extension of the

Cheffia Numidian mountains in the south, and the Edough metamorphic complex and the Fetzara Lake in the west, Finally, the eastern extension of the Annaba-Bouteldja plain and the Mounts of Nador N'bail in the east. As seen in the diagram, the research area includes the bulk of Annaba city and the western portion of El Taref city (Fig.1).
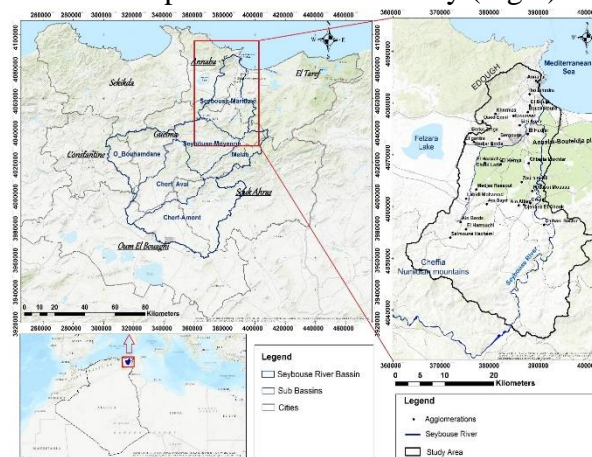

Fig. 1 Study area site

The study area is divided into 31 main sectors: Annaba, El Bouni, Kherraza, Boukhadra, BerkaZerga, Bouzaaroura, Essarouel, Sidi Salem, Oued Ennil, Sidi Amar, Hadjar Eddis, Derradji Redjem, Merzoug Amar (El Gantra), Bergouga, El Hadjar, El Kerma, Chabi Larbi, El Heraicha, Ain Berda, El Harrouchi, Ain Sayd, Medjez Rassoul, Salmoune Hachemi, Labidi Mohamed, Drean, Ain Allam, Djenane El Chouk, Fedaoui Moussa, Chbaita Mokhtar, Zourami Ali, Chihani Bachir.

2.2. Geological and Hydrogeological framework

The studies realized in the region there are two types of formations [8], one metamorphic presented by the Eddough Massif and the other sedimentary occupying the totality of Annaba plains. (Fig.2).
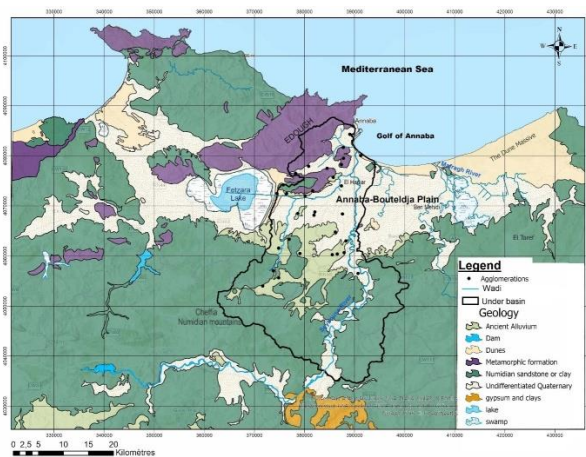
The sediments are heterogeneous with numerous alternations of sandy clays, sands and gravels beds. Two main aquifers are distinguished [9].
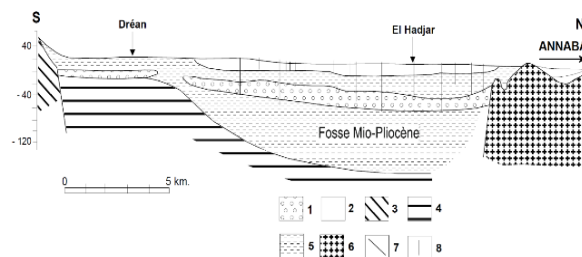
Aquifers, both shallow and deep:

- the gneiss altered; dune massif, the dunes, and the water current; and recent alluvium constitute a surface aquifer that stretches across the whole plain of Annaba and flows through the surface silts (Fig.3)

- the deep aquifer is captive and composed of gravel; its roof is made up of various textures (clay and clay loam, clay and sand); the main aquifer of the basin is built in permeable sediments (Mio-Pliocène), is made up mostly of pebble, sand, and clay along the wadis, and has stronger hydraulic properties; and becomes free at the region. Drean is a character in the Drean franchise. The superficial layer overtakes the gravel layer at this stage (Drean), and there is the risk of a phenomenon leakage between the two layers [10]. These aquifers are connected by a semi-permeable and/or impermeable intermediate layer, forming a single two-level aquifer.

Rainfall and runoff of Seybouse wadi water from further south feed the aquifer. The need to protect this aquifer's supplies is critical becauseit is the primary source of water for human use.

Fig. 3 Hydrogeological Cup through the plain of Annaba [12], amended 2010).
1: pebbles and gravels, 2: sand, 3: Numidian clay, 4: Cenomanian marl and marly limestone,
5: Plio-Quaternary detrital clays, 6: metamorphic formations, 7: fault, 8: drilling;
source: own elaboration

2.3. DPSIR analysis approach

The Water Framework Directive (WFD) 2000/60/EC clearly sets the basis and principles for effective protection of groundwater, internal, transitional and coastal waters at the river basin scale.

Several approaches and methodologies have been proposed for improving water resource management at this scale, with the Driver-Pressure-State-Impact-Response (DPSIR) approach [13], being one of the most widely used in the context of integrated water resource management [14]. A new computational water interconnected model based on cause-effect relationships has been created to address the life cycle of water resource management. The new conceptual model is based mainly on three decisive categories:

(1) the natural system, which is of critical significance for the water available quantities and qualities,

(2) the human system, which determines the use of water and the pollution of the resource,

(3) the institutional and management system must balance consideration of the natural and human systems and their interdependencies.

Fig. 2 Annaba Plain Geological Sketch Map [11], amended 2010.

The three systems are divided into five categories based on cause-effect DPSIR framework for development of water related variables. The variables reflect and translate the water sustainability concepts, preventive and ecosystem approaches. The five categories are:

socio-economic aspects, anthropogenic pollution pressures, state of water quality, public health and ecological impacts and the institutional responses. The human system is explained by the socio-economic, anthropogenic pressures and public health variables. The natural water system is presented by the state of water, and ecological impacts and the institutional system is reflected by the institutional responses.

The Driver-Pressure-State-Impact-Response (DPSIR) framework was chosen as a well-established framework for developing the possible variables under five categories that are separated into a collection of most important variables that represent essential aspects of water supplies.

Following the categories are:

D: Driving forces are underlying socio-economic and sectoral factors influencing a variety of relevant variables. Drivers produce a series of pressures and are quantified by aggregated data, population, tourism, Agriculture water consumption etc;

P: Pressure indicators describe the variables which directly cause environmental problems, The consequences of the Driving Forces on water supplies are manifested as pressures. Pressures weaken the condition of water supplies and have an effect on both them and humans;

S: State indicators illustrate the existing conditions and the observable changes of the environment (chemical characteristics);

I: Impact indicators describe the ultimate effects of changes of state on the human and ecosystems; and R: Response indicators present the efforts of the administration and policy making level (Decision makers, management) the measures taken to improve the state of the water resources.

In the European research project EUROCAT, which seeks to achieve integrated catchment and coastal zone management, the Driving power- Pressure-State- Impact- Response (DPSIR) system has been chosen to study all regional water catchments. [13]. It will help forecast how potential socio-economic developments in water catchments will affect water quality, allowing policy responses to be developed to mitigate the stresses caused by those drivers and the impacts of such pressures on water quality [15, 16]. developed metrics for an effective and operational decision support mechanism for efficient use of water supplies at the catchment level using the DPSIR method for environmental cause-effect relationships.

The new conceptual water integrated model was applied to the life cycle of water resources management in the Seybouse Valley.

The research's key goals were to:

– characterize the efficient variables of water sector management and identify the geographical areas under water stresses;

– establish prediction relationships between water abstraction from the coastal aquifer and water quality state;

– Group municipalities according to their water-related variables;

– Make proposals for reform, including new ideas for preserving natural water supplies as sources of supply for current and future generations [1].

3. RESULTS AND DISCUSSIONS
3.1. Statistical analysis

For the normality tests and multivariate statistical analyses–hierarchical cluster analysis (HCA) and principal component

analysis (PCA)–as well as neural networks, the statistical program STATISTICA (version 8; 2008) are used [17].

Ten pollution variables (Solid Waste, Domestic wastewater, Industrial wastewater, Pesticides, Organic fertilizers, Chemical fertilizers, Petrol stations, Carbon dioxide (CO_2), Seawater intrusion and Total Dissolved Solids (TDS) were utilized in the statistical analyses).

3.1.1. Analysis with artificial neural networks (ANN)

To define and prioritize the efficient variables of pollution pressure categories, ANN models were used. The use of ANN models does not include linearity or normality of data. As a consequence, there is no need for data transformation [18].

Total Dissolved Solids data and Pollution parameters were applied to create the ANN model. The pollution parameters were: Domestic Solid waste (Solid waste), Domestic wastewater (Dom wastw), Industrial wastewater (Ind wastw), Pesticides (pest), Organic fertilizers (Org fert), Chemical fertilizers (Chem fert), Petrol stations (Petrol stat), Carbon dioxide (CO_2), Seawater intrusion (Seawater Intr) and Total Dissolved Solids (TDS).
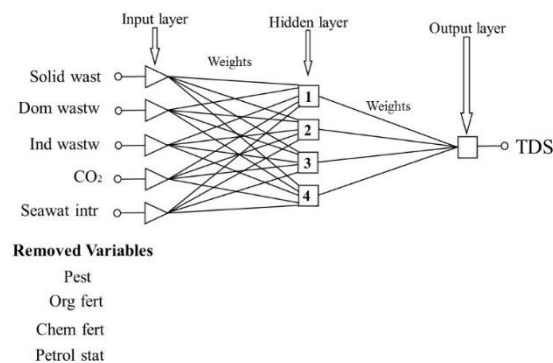
The Pollution pressure variables were considered as the possible input variables whilst the target output variable was the Total Dissolved Solids (TDS).

The variables representing the pollution pressure variables were considered as the possible input variables whilst the target output variable was the Total Dissolved Solids (TDS) [18].

The MLP network can be represented by the following compact form:

{TDS}= ANN [*Solid waste, Dom wastw, Ind wastw, Pest, Org fert, Chem fert, Petrol stat, CO₂, Seawat intr*]

A schematic diagram of this network is given in (Fig. 6).

Fig. 6 MLP Network (three layers), Pollution pressure variables

MLP (3 and 4 layers), RBF, GRNN, and Linear networks are the forms of networks considered. Several networks were examined during the study [1]. The best optimal ANN model found is MLP (3 layers) with 4 hidden nodes (Fig. 6) and a minimal error of 25.56 compared with the other types of ANN networks Table1. The model has very good performance in verification with regression ratio (S.D. ratio) of 0.01919 and the correlation coefficient is higher than 87% for training, 99% for verification and testing (Table 1) which shows an excellent agreement between the actual observed and predicted TDS (Fig. 7).

The ANN sensitivity analysis of Pollution pressure variables in both training and verification phases (Table 2) indicates that Solid Waste and Domestic Wastewater, Industrial Wastewater, Carbon dioxide and Seawater intrusion are the most important and pressing pollution sources.

Aissam GHRIEB, Fethi BAALI, Chemceddine Fehdi, Azzedine HANI, Hicham CHAFFAI, Larbi Djabri, *Use of artificial neural networks and multivariate statistical analysis for modeling the pollution pressure of water resources in the seybouse valley (north-eastern Algeria),* Food and Environment Safety, Volume XX, Issue 1 – 2021, pag. 68 – 80

The ANN model removed four input variables due to their low sensitivity which are Pesticides, Organic fertilizers, Chemical fertilizers, Petrol stations.

Abs. error mean	404.7289	22.67534	116.7138
RMS error	594.5	25.56	156.6
S.D. ratio	0.545813	0.01919	0.1360706
Correlation	0.8789501	0.9998636	0.9907989

Legend: Tr: Training, Ve: Verification, Te: Testing

Table 1

Statistical regression parameters for the target output (TDS)-Pollution pressure variables

	Tr. TDS	Ve. TDS	Te. TDS
Data mean	2209.906	2620.833	2142.188
Data S.D.	1091.389	1231.062	954.9432
Error mean	-112.8458	-16.76796	115.2068
Error S.D.	595.6944	23.62893	129.9397
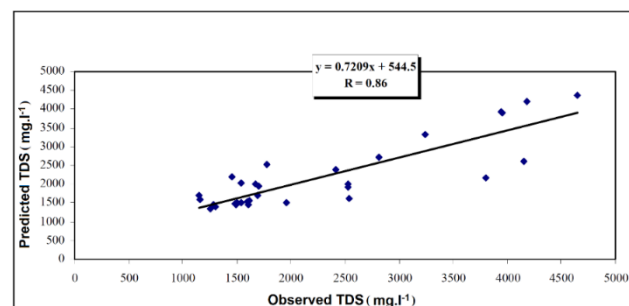
Fig. 7 Predicted TDS vs Observed TDS

Table 2

Sensitivity analysis of independent input variables— Pollution pressure variables

	Solid waste	Dom wast w	Ind wast w	CO_2	Seawater intr
Rank	4	5	1	3	2
Error	725.203	656.0755	801.8969	728.4742	730.8205
Ratio	1.167607	1.056309	1.291088	1.27013	1.176651
Rank	4	5	1	3	2
Error	396.4911	238.3604	670.42	631.0517	604.3308
Ratio	1.386904	0.8337713	2.345092	2.0198	2.113915

Verification phase are

Industrial Wastewater, Seawater intrusion, Carbon dioxide (CO_2), Solid Waste, Domestic Wastewater. The ANN model removed four input variables – Pesticide, Organic fertilizers, Chemical fertilizers, Petrol stations because of its low sensitivity. The results of the ANN model are shown in (Table 3).

Table 3

Ranking of input Pollution pressure variables

	Ind wastw	Seawater Intr	CO_2	Solid waste	Dom wastw
Rank	1	2	3	4	5

Tests of normality (Pollution pressure)

The normality tests are used in addition to the graphical normality evaluation [19]. The Kolmogorov-Smirnov (K-S) test, Lilliefors corrected K-S test, and Shapiro-Wilk test are the most common methods used to determine normality [20]. The most efficient measure is the Shapiro–Wilk test [21]. The Shapiro-Wilk test is recommended by some researchers as the best alternative for determining data normality [22].

Tests of normality of the pollution category variables The Pressure variables "P" for the study area are summarized in (Table 3). According to the Shapiro-Wilk test, shows that all the variables have non-normal distribution of data.

Table 4

Tests of normality of Pollution pressure variables

	N	maxI	K-S	Lilliefors P	W	P
Dom wastw	31	0.389100	**p < .01**	**p < .01**	**0.325455**	**0.000000**
Solid waste	31	0.245586	**p < .05**	**p < .01**	**0.667737**	**0.000000**
Pest	31	0.187130	p < .20	**p < .01**	**0.912573**	**0.015067**
Org fert	31	0.150107	p > .20	p < .10	**0.925074**	**0.032257**
Chem fert	31	0.134144	p > .20	p < .20	**0.912688**	**0.015171**
Petrol stat	31	0.370968	**p < .01**	**p < .01**	**0.445230**	**0.000000**
Ind wastw	31	0.330812	**p < .01**	**p < .01**	**0.695460**	**0.000001**
CO_2	31	0.231724	**p < .05**	**p < .01**	**0.799566**	**0.000051**
Seawater Intr	31	0.283124	**p < .01**	**p < .01**	**0.766539**	**0.000013**
TDS	31	0.251339	**p < .05**	**p < .01**	**0.828221**	**0.000180**

3.1.2. Correlation matrix analysis for Testing the Relation Between Any Two Variables

Using the logarithm (Log), the Pearson's correlation coefficient was used to show the interrelationship and coherence pattern among pollution pressure parameters [23]. The correlation coefficient values of the analyzed pollution pressure parameters are given in (Table 5). Correlation matrix showed inter-parameter. Strong ($p < 0.01$) and significant correlation ($p < 0.05$) were observed in the pollution parameters. Log(Domestic Wastewater) has significant and positive linear relationships with Log(Petrol Stations), Log(Carbon dioxide), Log(Industrial Wastewater), and Log(Seawater intrusion). Increase in domestic solid waste is associated with the increase in wastewater generation since the garbage produced at the screens in the treatment facilities are transferred to the solid waste sanitary landfills. Domestic wastewater increases with the increase in the industrial wastewater generation since the industrial facilities are connected to the urban wastewater systems immediately after pre-treatment. The increase in the domestic wastewater generation indicates rise in groundwater abstraction and thus increase in seawater intrusion.

Log(Pesticides) has significant and positive linear relationships with log(Organic fertilizers), log(Chemical fertilizers). The use of Pesticides is always associated with organic fertilizers and Chemical fertilizers since they are applied for the same agriculture land but with different proportions.

Log(Industrial Wastewater) has a positive linear relationship with log(Seawater Intrusion), log(Carbon dioxide) and log(Total Dissolved Solids). Increase of industrial wastewater generation is caused by more use of groundwater which results in lowering the water table and seawater intrusion. The log(Total Dissolved Solids), had a positive strong correlation with log(Carbon dioxide), log(Industrial Wastewater), and moderate correlation with log(Petrol Stations).

Table 5

Pearson Correlation matrix of pollution pressure variable

	Dom wastw	Solid wast	Pest	Org fert	Chem fert	Petrol stat	Ind wastw	CO₂	Seawater Intr	TDS
Dom wastw	1.00									
Solid wast	0.36	1.00								
Pest	-0.26	-0.25	1.00							
Org fert	-0.31	-0.29	**0.89**	1.00						
Chem fert	-0.35	-0.30	**0.75**	**0.89**	1.00					
Petrol stat	**0.83**	0.37	-0.36	-0.41	**-0.47**	1.00				
Ind wastw	**0.53**	0.18	-0.29	-0.42	**-0.63**	**0.61**	1.00			
CO₂	**0.46**	0.21	-0.47	**-0.64**	**-0.65**	**0.53**	**0.66**	1.00		
Seawater Intr	**0.54**	0.22	-0.47	**-0.51**	**-0.61**	**0.68**	**0.66**	**0.50**	1.00	
TDS	0.33	0.20	**-0.44**	**-0.57**	**-0.66**	**0.41**	**0.73**	**0.75**	**0.55**	1.00

Note: Significant values (at $p< 0.05$) are in bold, n=31 (casewise deletion of missing data)

Table 6

Factor-variable correlations (factor loadings), pollution pressure variables

Variable	Factor 1	Factor 2
Dom wastw	**-0.70**	0.39
Solid Waste	-0.39	**0.83**
Indwastw	**-0.88**	-0.19
CO₂	**-0.83**	-0.20
Seawater Intr	**-0.79**	-0.01
TDS	**-0.83**	-0.30
*Pest	0.49	-0.02
*Org fert	0.62	0.03
*Chem fert	**0.73**	0.07
*Petrol stat	**-0.75**	0.27
Expl.Variance	5.57	1.43
PropreTotale	0.55	0.14

In bold: loadings are >.700000
Active and Supplementary variables
*Supplementary variable (Underlined loadings are >.700000)

3.1.3. Factor Analysis (FA)

The global PCCA of the set of data related to "sources of pollution" demonstrate that the first two factors (F1 and F2) provide a justified inertia of 73,6 % that appears rather good, having an average inertia, a considerable number (31) of analyzed samples and (10) variables taken into consideration; knowing that the cumulated percentage goes gradually towards 100%.

(Table 6) presents variances of factors and their loadings from variables. The first factor corresponds to the largest eigenvalue (5,57) and accounts for approximately 56,89% of the total variance. It is most correlated with the variables Domestic Wastewater, Industrial Wastewater, Carbon dioxide, Seawater Intrusion, Total Dissolved Solids and Petrol Stations (negative correlation) and Chemical fertilizers (positive correlation). The second factor corresponding to the second eigenvalue (1,43) accounts for 16,66% of the total variance. It is highly correlated with Solid Waste (positive correlation).

(Fig. 8(a) and 8(b)) displays coordinates for the two factors. The graph shows a unit circle with active variables that were used to compute the current factor solution and a supplementary variable that was only mapped into the coordinate system defined by the factors. Because the current analysis is based on correlations, the highest

possible factor coordinate (variable-factor correlation) is 1.0, and the number of all squared factor coordinates for a variable (squared correlations between the variable and all factors) cannot be greater than 1.0. The circle will show you (on a scale) how well each variable is defined by the current set of variables (the closer a variable in this plot is located to the unit circle, the better is its representation by the current coordinate system). Based on the magnitudes of the factor coordinates (variable-factor correlations) for the variables in the analysis,

In (Figure 8(a)), it is highly correlated with the variables of Industrial Wastewater, Domestic Wastewater, Seawater Intrusion, CO_2, TDS and the Supplementary variable Petrol Stations (negative correlation) on the other hand the factor 1 is positively correlated with pesticides, chemical

fertilizers and organic fertilizers. This factor can be characterized by pollution of domestic origin to which is added pollution of industrial origin, also marked by agricultural pollution indicators (pesticides, chemical fertilizers and organic fertilizers).

In the plan of individuals, (Figure 8 (b)), the municipalities of El Bouni, Sidi Salem, Sidi Amar and El Hadjar are distinguished by Seawater Intrusion, Industrial Wastewater, air pollution and therefore by high TDS values. The municipality of Annaba is characterized by the large quantities of Domestic Wastewater due to the density of the population. The municipality of Berka Zarga, which houses the largest landfill in the sector, is characterized by Solid Waste. Finally, the municipalities of El Kerma, El Heraicha, Ain Allam, Djenane El Chouk, El Harouchi, Ain Sayd and Zourami Ali, are characterized by agricultural pollution.
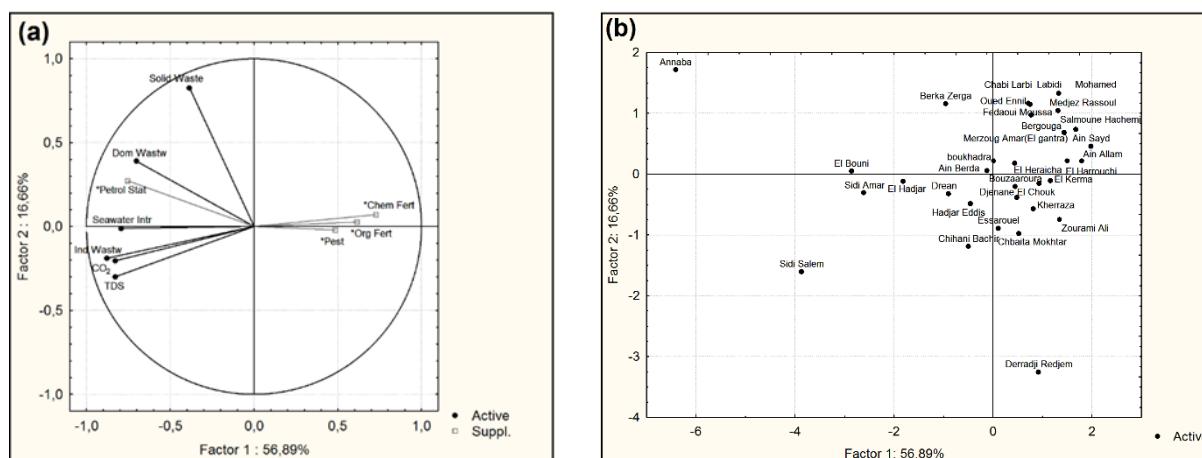

Fig. 8(a) Projection of the variables on the factor-plane (I–II); and (b) projection of the cases on the factor-plane (I–II)

3.1.4. Hierarchical Cluster Analysis (HCA)

HCA was used for organize observations and variables in the same category of the data set, into more meaningful groups sharing close and similar characteristics.

Transformed variables were standardized; complete linkage of tree clustering was selected so that Euclidean distance between two clusters is determined by the distance of the furthest cases of these two clusters.

The R-mode cluster analysis was applied to predict pollution variables groupings Dendrogram chart has been obtained and the results are shown in (Fig.9(a)) and classified into two dissimilar clusters. Cluster 1 included two subgroups. The first formed by TDS and CO_2, it can be labelled as water quality. Solid Waste and Domestic Wastewater with the same weight to Industrial Wastewater, form the second subgroup indicating anthropogenic pollution by industrial and urban discharges to which add contamination by solid waste, following the infiltration into the ground of water laden with pollutants from landfills, after leaching. Cluster 2 contained only Seawater Intrusion; it expresses the pollution of groundwater by Seawater.

The spatial similarities and site grouping among the sampling points were identified using Q-mode cluster analysis. In a sample cluster, a particular group/class exhibits identical characteristics in relation to the analyzed parameters.

The 31 municipalities sites for fall into two dissimilar clusters (Fig.9(b)). The first cluster (right) consists of coastal municipalities or located a short distance from the sea; where Seawater intrusion is felt such as Annaba, Sidi Salem.

The second cluster of municipalities contains El Bouni, Boukhadra, Bouzaaroura, Oued Ennil, Berka Zerga, Essarouel, Sidi Amar, Bergouga, Chabi Larbi, Hdjar Eddis, Derradji Redjem, El Gantra, El Hadjar, El Heraicha, El Kerma, Chbaita Mokhtar, Zourami Ali, Drean, Ain Allam, Djenane El Chouk, Fedaoui Moussa, Chihani Bachir, Kherraza, El Harrouchi, Ain Sayd, Medjez El Rassoul, Labidi Mohamed, Salmoune Hachemi and Ain Berda.

The first cluster of municipalities is associated with the first group of variables, while the second cluster of municipalities is associated with the second group of variables, based on the magnitudes of the linkage distances. As a result, the first cluster of municipalities is classified as the "Seawater Intrusion" cluster, while the second cluster is known as the "anthropogenic pollution" cluster.
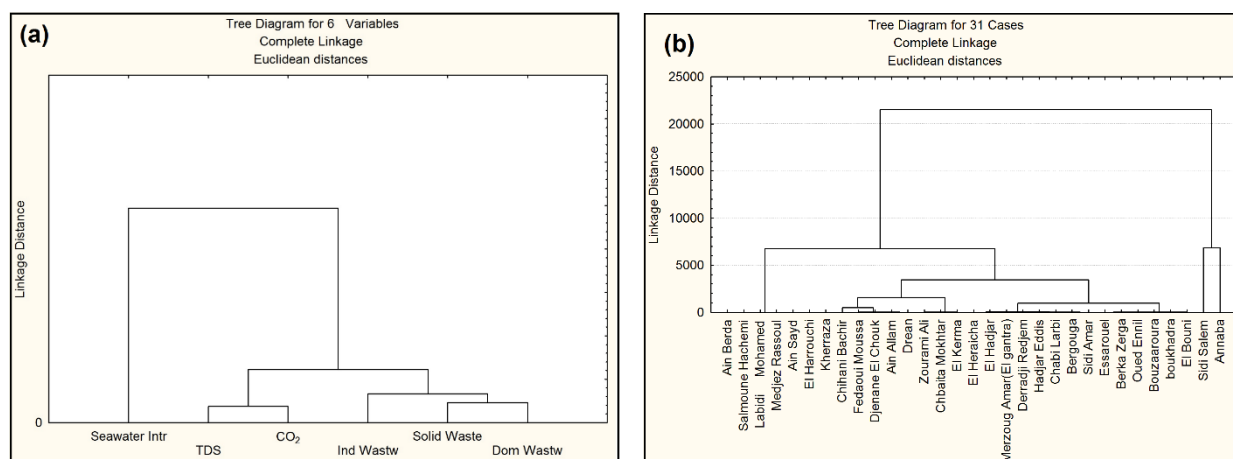
Fig. 9(a) Cluster analysis results for variables - State of water quality, and (b) hierarchical tree for cases- State of water quality Complete Linkage Euclidean distances.

4. Conclusion

This paper presents integrated approaches for characterizing pollution pressures of groundwater in the Seybouse Valley (North East of Algeria). Based on GIS and DPSIR methods; and the major objective of the study was to establish a modelling relationship between Total Dissolved Solids and pollution determinants, and characterize their priorities. To this end, the effective variables have been characterized and prioritized using multi-criteria analysis with ANN. The selected variables have been classified and organized using the multivariate techniques of cluster analysis (HCA), principal component and classification analysis (PCCA) and factor analysis (FA).

The conclusions of data analysis using the techniques of ANN, basic statistics and multivariate can be summarized as follows:

The data of Total Dissolved Solids (TDS) and pollution pressures were applied to create the ANN model to assist water planners and managers in the Seybouse Valley to better understand the pollution determinants influencing the attractiveness of groundwater users. The selection and prioritization of effective pollution parameters indicate that this model is a useful tool to devise priority interventions, optimizing the limited available financial resources, towards to provision of appropriate quantities of water of suitable quality.

All water policy and management responses are significant. Sustainable coastal aquifer management must take into consideration technical engineering as well as managerial interventions such that top priority should be given to the industrial wastewater treatment that is the most pressing pollution source followed by seawater intrusion. Where majority of the industries do not have wastewater treatment plants and the river and aquifers are the direct receptors of their effluents.

The municipalities of El Bouni, Sidi Salem, Sidi Amar and El Hadjar are distinguished by Seawater Intrusion, Industrial Wastewater, air pollution and therefore by high TDS values. The municipality of Annaba is characterized by the large quantities of Domestic Wastewater due to the density of the population. The municipality of Berka Zarga, which houses the largest landfill in the sector, is characterized by Solid Waste. Finally, the municipalities of El Kerma, El Heraicha, Ain Allam, Djenane El Chouk, El Harouchi, Ain Sayd and Zourami Ali, Characterized by agricultural pollution.

The main sources of pollution are related to anthropogenic activities: agriculture, socio-economic activities, industrial activities, cattle farming and urbanization. As a result, water quality is affected by both point and diffuse sources of pollution.

5. Abbreviation

Solid Waste: Domestic solid waste, tons.d^{-1}
Dom Wastw: Domestic wastewater, hm^3.y^{-1}
Ind Wastw: Industrial wastewater, hm^3.y^{-1}
Pest: Pesticides, tons.y^{-1}
Org Fert: Organic fertilizers, tons.y^{-1}
Chem Fert: Chemical fertilizers, tons.y^{-1}
Petrol Stat: Petrol stations
CO$_2$: Carbon dioxide, ppm

Seawater Intr: Seawater intrusion, $hm^3.y^{-1}$
TDS: Total Dissolved Solids, $mg.l^{-1}$

6. Acknowledgments

7. REFERENCES

[1]. GHRIEB, A. BAALI, F. FEHDI, CH. HANI, A. CHAFFAI, H. & DJABRI, L, On the use of GIS and DPSIR methods to analyse water quality in Seybouse Valley (North East of Algeria), *Journal of Biodiversity and Environmental Sciences (JBES),* (2019). 114-125.

[2]. AOUN SEBAITI, B. Optimized management of water resources from a coastal aquifer application to the Annaba plain (northeastern Algeria) *PhD Thesis-univ of Lille,* (2010). P113–121–134.

[3]. AOUN-SEBAITI B, HANI A, DJABRI L, CHAFFAI H, AICHOURI I, BOUGHERIRA N, Simulation of water supply and water demand in the valley of Seybouse (East Algeria). *Desalination and Water Treatment. Taylor & Francis,* (2013). DOI: 10.1080/19443994, 855662, 1-6.

[4]. Asadollahfardi G, Water Quality Management. Assessment and Interpretation. *Springer Briefs in Water Science and Technology. Springer,* (2015). DOI 10.1007/978-3-662-44725-3.

[5]. DJABRI, L. FEHDI, CH. HANI, A. BOUHSINA, S. NOUIRI, I. DJOUAMAA, M.C. BOCH, A.P. & BAALI, F, Climate Change and Water Resources: Seasonal Hydrochemical Changes of Water from Alluvium Aquifers: Drean-Annaba Aquifer Case Study (Ne Algeria), (2015). *Book* ID: 326568–1–En. Chapter 17.

[6]. HALIMI, S. BAALI, F. KHERICI, N. ZAIRI, M. & BOUHSINA, S. Irrigation and Risk of Saline Pollution. Example: Groundwater Of Annaba Plain (North East Of Algeria), (2016) v12n6p241.

[7]. KOUZANA, L. BEN MAMMOU, A & GAALOUL, N, Seawater intrusion and salinization in a coastal water table (Korba, Cap-Bon, Tunisia), Geo-Eco-Trop, (2007). 31: 57-70.

[8]. HAMMOR, D, From Pan Miocene. 600 million years polycyclic evolution in the massive Edough (North East Algeria). Traced by petrology, tectonics and geochronology (U/Pb, Rb/Sr, Sm/Nd and 39Ar/40Ar). *Thesis of USTLanguedoc, Univ. Montpellier II,* (1992). p 205.

[9]. LAMOUROUX, C. & HANI, A. Identification of groundwater flow paths in complex systems aquifer. *Hydrol. Processes* (2006). 20, 2971–2987.

[10]. KHERICI, N, Vulnerability to chemical pollution of groundwater systems superposed layers in industrial and agricultural Annaba Mafragh the East Algerian community. *PhD Thesis-univ of Annaba,* (1993). P28–34.

[11]. GAUD, B, Hydrogeological study system Annaba Bouteldja. Knowledge synthesis and research of modelling conditions. *A.N.R.H. report (unpublished), Annaba,* vol 2, 230P. 10 boards, (1976).

[12]. HANI, A, Methodological analysis of the structure and anthropogenic processes: application to water resources in a Mediterranean coastal basin *PhD Thesis-univ of Annaba,* (2003). 214P. P18–33

[13]. European Environmental Agency (EEA), *The DPSIR framework used by the EEA,* (2014). Available online: http://glossary.eea.europa.eu//terminology/sitesearch?term=DPSIR.

[14]. MATTAS, C. VOUDOURIS, K. & PANAGOPOULOS, A. Integrated Groundwater Resources Management Using the DPSIR Approach in a GIS Environment: A Case Study from the Gallikos River Basin, North Greece. 6, 1043–1068, (2014).

[15]. CAVE, R.R. LEDOUX, L. TURNER, K. JICKELLS, T. ANDREWS, J.E.& DAVIS, H, The Humber catchment and its coastal area: from UK to European perspectives. *The Science of the Total Environment, Article in Press,* (2003). DOI 10.1016/s0048-9697(03)00093-7.

[16]. JEUNESSE, I.L. ROUNSEVELL, M. & VANCLOOSTER, M, Delivering a decision support system tool to a river contract: a way to implement the participatory approach principle at the catchment scale? *Journal of physics and Chemistry of the Earth, Article in Press. Volume 28,* (2003). Issues12–13, 547–554.

[17]. STATISTICA. Electronic Manual StatSoft, Inc. STATISTICA data analysis software system, (version 8; 2008).

[18]. Jalala, S, Characterizing the Multi-criteria Parameters of Integrated Water Management Model in the Semi-arid Mediterranean Region: Application to Gaza Strip as a case study. PhD Thesis-univ of Lille, (2005). P98–108–110.

[19]. IAHS-International Association of Hydrological Sciences. *International hydrology today,* (2003).

[20]. ELLIOTT,AC. & WOODWARD, WA. Statistical analysis quick reference guidebook with SPSS examples. 1st ed. London: Sage Publications, (2007).

[21]. OZTUNA, D.ELHAN, AH.&TUCCAR, E, Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. Turkish *Journal of Medical Sciences,* (2006). 36(3), 171–176.

[22]. YAP, B.W. &SIM, C.H, Comparisons of various types of normality tests, *Journal of Statistical Computation and Simulation,* (2011). 81, 12, 2141–2155.

[23]. THODE, HJ. Testing for normality, *New York: Marcel Dekker,* (2002).