Classification of sites according to Gamma radiation levels in ambient air using GIS and Data Mining Techniques.

M. Hefnawi[1], A.Y. Abdel- Haseib[1]

1. Egyptian Nuclear & Radiological Regulatory Authority,(ENRRA) Cairo, EGYPT

ABSTRACT

This work uses the data which represents the measurements of the Gamma radiation levels in ambient air from many Gamma monitoring stations which are distributed in many sites to classify the regions which cover these sites according to these measurements. The processes of the classification are: - dividing the range of measurements to several intervals, making interpolation for all regions that cover all Gamma monitoring stations, representing the interpolation information in the map using Geographic information systems (GIS) technology and final process is classifying all the sites on this map according to the determined intervals by using data mining techniques via interpolation information.

This method is implemented for determining the background of Gamma radiation levels for many sites in Egypt. Because of this background in sometimes is necessary for many environmental researches. This background is useful for making risk assessment evaluation for any site in Egypt. The output result from this implementation shows that most sites in Egypt have been classified within three intervals which are: first interval is from 2.02E-2 µS/h to 4.75E-2 µS/h with 47.11% of Egypt area, second interval is from 4.75E-2 µS/h to 8.85E-2 µ S/h with 40% of Egypt area and third interval is from 8.85E-2 µS/h to 1.42E-1 µS/h with 6.82% of Egypt area. This method is more useful than other traditional methods because the results from this method show that this method saves more effort, time and cost than other methods.

Key Words: GIS/ Data Mining / Environmental classification / Data Modeling.

INTRODUCTION

Classifying streams is very necessary subject that must be received enough attention in the scientific literature. In particular, it is very important to solve the problem for specific needs for the field based. When referring to streams. There are several ways to define perennial, intermittent and ephemeral[(1)](Gritzner and Millet, 2003b). This study uses definitions from the

USGS[2](Gritzner, 2003). The definitions are related to time and state that a perennial stream is one which flows continuously; an intermittent (or seasonal) stream is one which flows only at certain times of the year when it receives water from springs or from some surface source such as melting snow in mountainous areas; and an ephemeral stream is one that flows only in direct response to precipitation, and whose channel is at all times above the water table. It is important to notice that the previous definitions have implications such that a connection of the flow conditions with the water table, which in turn is influenced by climate changes and seasonal variations. Geographic information systems (GIS) technologies have made excellent progress in the last several years, making it possible to integrate several data layers for such purposes as mapping and modeling efforts. One data layer that has not been available in a reliable way, especially in a more detailed scale (county and city level, for example) is that of stream classification. There is therefore a need to investigate new approaches and strategies to classify streams that do not depend strongly on field methods. Field methods produce reasonable data if applied properly [3](Gritzner and Millet, 2003a) but can be expensive and are prone to errors due to incorrect measurements or inability to reach certain areas. This study focuses on strategies based on GIS and data mining, using digital information which are measurements of the Gamma radiation levels in ambient air from many Gamma monitoring stations, with a case study for many sites in Egypt.

MATERIAL AND METHOD

The aim from this work is modeling the information of Gamma radiation levels for many sites in Egypt and classifying these sites according to the Gamma radiation levels. Therefore this work uses the data which represents the measurements of the Gamma radiation levels in ambient air from many Gamma monitoring stations which are distributed in many sites to classify the regions which cover these sites according to these measurements. The processes of the classification are as following [4]:-

1. Dividing the range of measurements to several intervals

The process of dividing the range of measurements uses Natural Breaks method[5] also called the Jenks natural breaks classification method, is a data clustering method designed to determine the best arrangement of values into different classes. This is done by seeking to minimize each class’s average deviation from the class mean, while maximizing each class’s deviation from the means of the other groups. In other words, the method seeks to reduce the variance within classes and maximize the variance between classes.

The output intervals after applying Natural Breaks method are listed in table 1

Table 1: List of the output intervals after applying Natural Breaks method.

Interval Number	From	To	Unit
1	2.02E-02	4.75E-02	µSivert/h
2	4.75E-02	8.85E-02	µSivert/h
3	8.85E-02	1.43E-01	µSivert/h
4	1.43E-01	2.53E-01	µSivert/h
5	2.53E-01	4.59E-01	µSivert/h
6	4.59E-01	7.36E-01	µSivert/h
7	7.36E-01	1.088	µSivert/h

2. Making interpolation for all regions that cover all Gamma monitoring stations

Interpolation is a method of constructing new data points within the range of a discrete set of known data points [6,7]**.** In this study the inverse distance weighted interpolation method (IDW) is used for constructing new data points within the range for all regions that cover the points which represents Gamma monitoring stations. After applying IDW method nearly all region of Egypt are covered. Figure 1 shows a distributed gamma monitoring stations in the map of Egypt.

Figure 1: A distributed Gamma monitoring stations in the map of Egypt.

Table 2 shows the list for all gamma stations including the name of the stations and the average measurements of gamma radiation levels in unit of µS/ h for year 2012.

Station Number	station Name	Average value for Gamma radiation level at 2012	Station Number	station Name	Average value for Gamma radiation level at 2012	Station Number	station Name	Average value for Gamma radiation level at 2012
1	Cairo	3.50E-02	11	El Saloom	2.51E-02	21	El Kontella	5.56E-02
2	Alexandria	2.63E-02	12	Tanta	3.61E-02	22	El-Qoseama	5.06E-02
3	Demyat	2.02E-02	13	Zagaziq	2.72E-02	23	St Catherine	1.34E-01
4	Port Said	2.68E-02	14	El Mansoura	3.58E-02	24	El-Faium	3.14E-02
5	ElSMalia	3.50E-02	15	Al Arish	2.70E-02	25	El-Kharka	3.55E-02
6	Suez	4.82E-02	16	Rafah	3.15E-02	26	BaniSweef	1.89E-01
7	Hurdhada	5.61E-02	17	El Tor	7.36E-02	27	Asyoot	5.37E-02
8	Aswan	4.17E-02	18	Sharm El-shekh	1.09E-01	28	Qena	4.32E-02
9	El Dabaa	5.00E-02	19	Newabaa	1.09E+00	29	Alemenya	3.57E-02
10	Marsa Matrooh	4.12E-02	20	Taba	6.38E-02			

3. Representing the interpolation information in the map using GIS technology

This process uses the output result from the interpolation process and represents this output in the map which is shown in figure 1. Figure 2 shows the map in figure 1 after this process.

4. **Classifying all the sites on this map according to the determined intervals by using data mining techniques via interpolation information.**

Using data mining techniques in this process [8] means that an analysis will be performed for the result from the pervious process. In other words "an analysis will be performed in the interpolation information to obtain the classification for many sites in Egypt according to Gamma Radiation levels". The target of the classification process is producing many maps from the map in the figure 2 so that every produced map shows only the distribution of only one interval from table 1.This can be done by performing the logical operation of mathematical formula to transfer the source imagery raster (layer of interpolation information that appears in figure 2) to another imagery raster as shown in figure 3.

The logical operation of mathematical formula to compute the output raster that contains only all sites that belongs to the first interval from table 1 is [9]

([2012 Average]>= 0.00202) and ([2012 Average]<0.0475)

Where
"**2012 Average"** is the name of the raster which is appearing in figure 1.
"**0.00202"** is the first value in the first interval in table 1 in unit of µS/h.
"0.0475" is the last value in the first interval in table 1 in unit of µS/h.

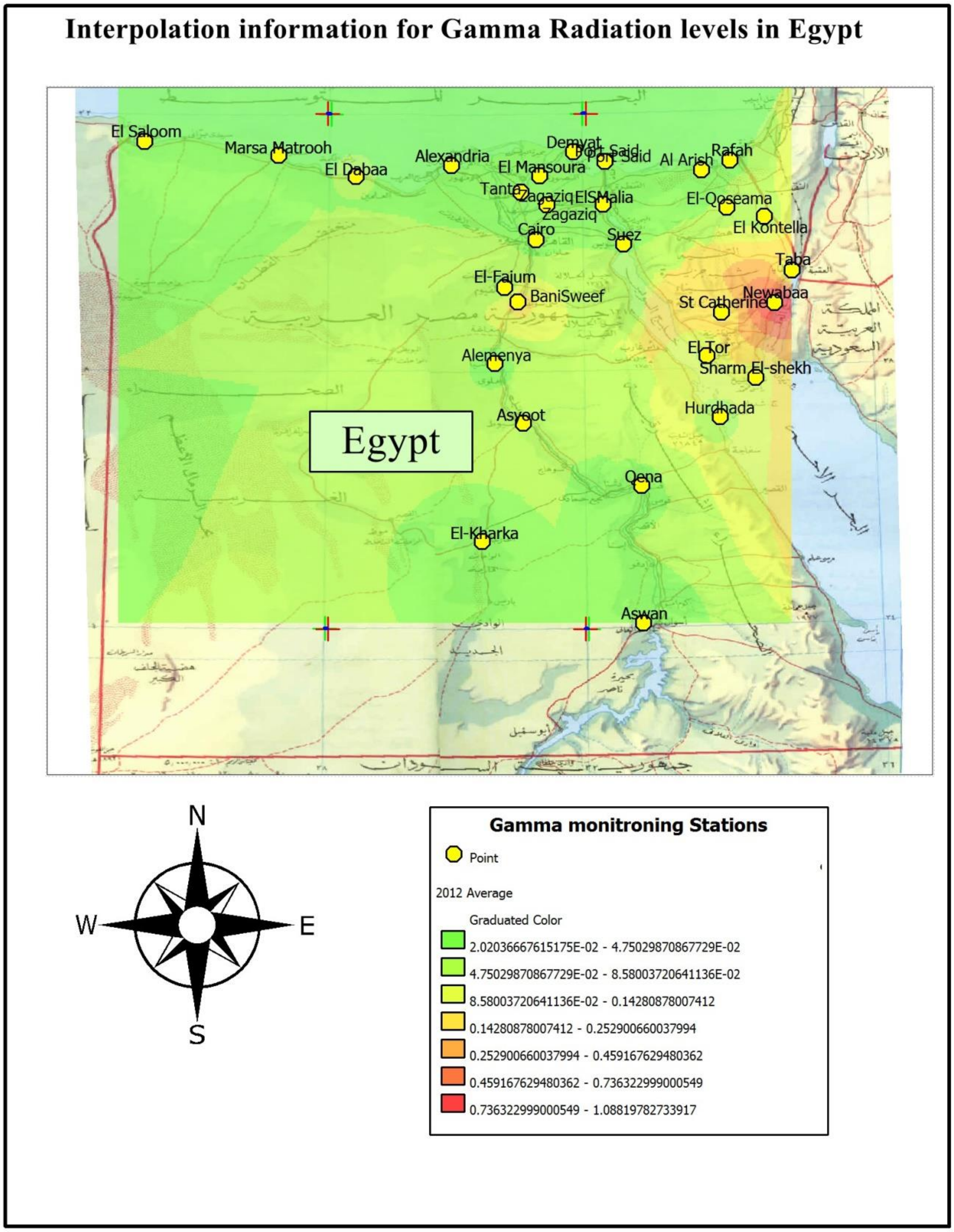
Interpolation information for Gamma Radiation levels in Egypt
El Saloom
Marsa Matrooh
El Dabaa
Alexandria
Demyat
Port Said
El Mansoura
Al Arish
Rafah
Tanta
Zagaziq
ElSMalia
El-Qoseama
El Kontella
Cairo
Suez
Taba
El-Fajum
BaniSweef
St Catherine
Newabaa
El Tor
Sharm El-shekh
Alemenya
Hurghada
Asyoot
Egypt
Qena
El-Kharka
Aswan
N
W
E
S
Gamma monitroning Stations
Point
2012 Average
Graduated Color
2.02036667615175E-02 - 4.75029870867729E-02
4.75029870867729E-02 - 8.58003720641136E-02
8.58003720641136E-02 - 0.14280878007412
0.14280878007412 - 0.252900660037994
0.252900660037994 - 0.459167629480362
0.459167629480362 - 0.736322999000549
0.736322999000549 - 1.08819782733917

Figure 2: Representation of the interpolation information for the map of Egypt using GIS technology.

.

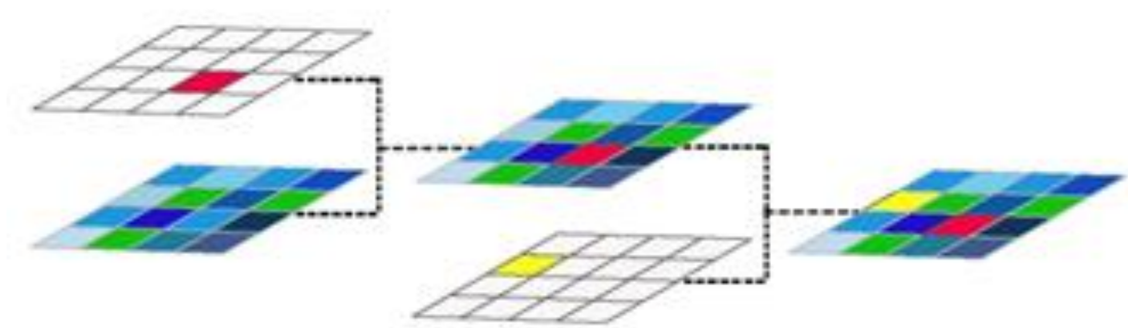

Figure 3: Performing the logical operation of mathematical formula to transfer the source imagery raster to another imagery raster.

The output from this formula is only one value of Boolean data type "In other wards the output will be true value or false value". Because of the required output is raster which means number of cells and every cell have only one value. This formula will be repeated to the number of times equal to the number of cells in raster **([2012 Average]).**

Assuming this formula is implemented to the cell that has gamma radiation measurements level of 3.00E-2 µS/h.

- The first part of this formula which is ([2012 Average]>= 0.00202) will be accomplished as following
- [2012 Average] will take value of 3.00E-2 µS/h therefore the first part will becomes
- (0.03>= 0.00202) and hence the output result from this part is true
- The third part will becomes (0.03 < 0.00475) and hence the output result from this part is true.
- Now the second part can be implemented after becoming in the form of true and true. So the final output from this formula is true.

By repeating this formula [10] for every cell in raster ([2012 Average]) we will obtain a new raster that shows all the sites that belong to the first interval in table 1.therfore a new map is produced for showing output raster as in figure 4.

Now from figure 4 it is easy to determine all the sites which are belonging to the first interval. These sites are covered with gray color in figure 4. By repeating this method with other intervals in table 1 we can obtain the required classification as in the next section

Sites that belongs to range of 2E-2 to 4.75 micro Sivert/h

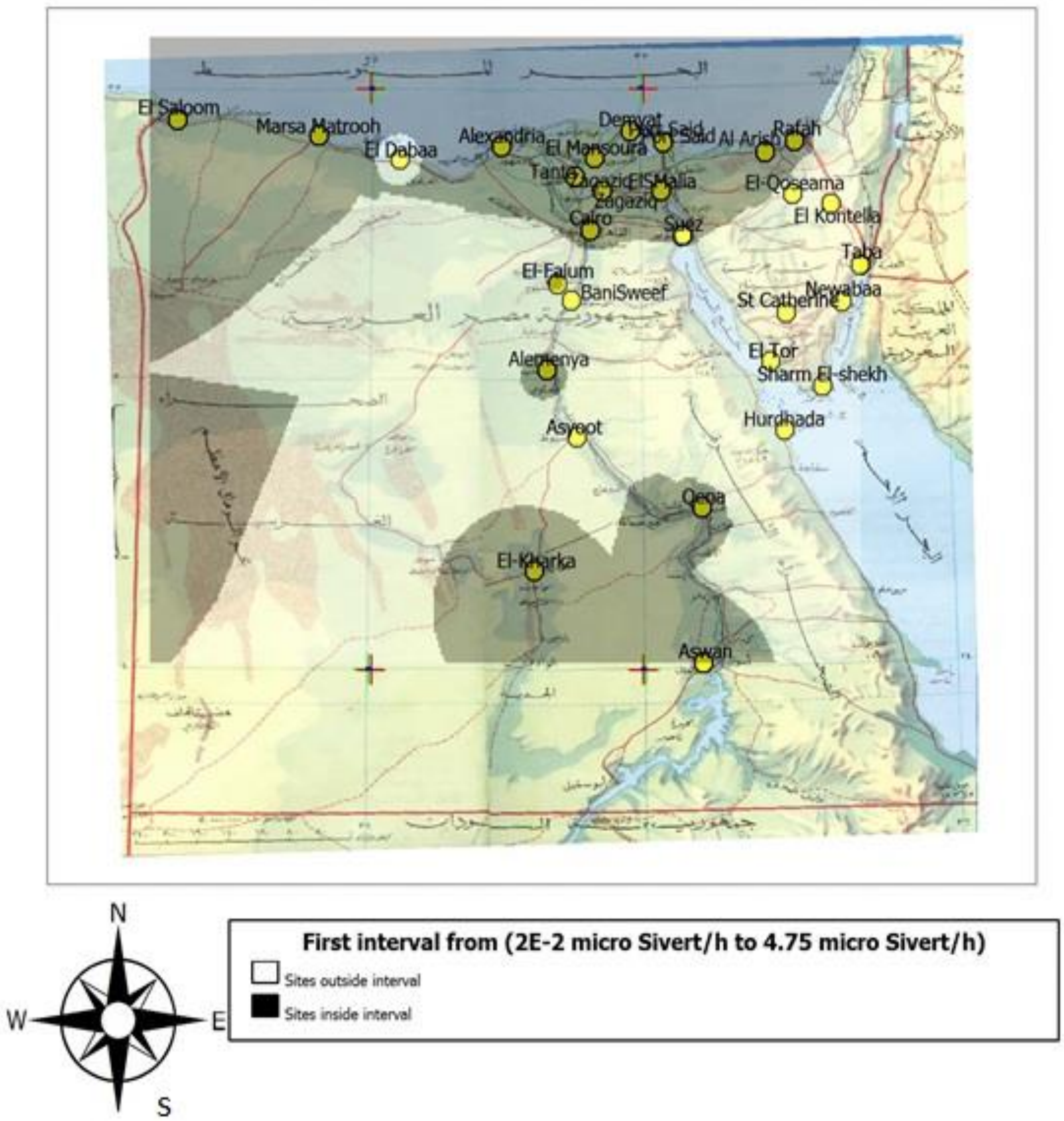

Figure 4: Raster that shows all the sites that belong to the first interval in table 1.

RESULTS AND DISCUSSION

The output result after implementing the specified method from pervious section is shown in figure 5. From figure 5 we can classify many sites in Egypt according to Gamma radiation levels. Table 3 shows all sites that belong to each interval in table 1 and their area.

Table 3: All sites that belong to each interval in table 1 and their area.

Interval Number	From	To	Sites which belong to interval	Percentage of total area for all sites	Unit
1	2.02E-02	4.75E-02	North of Sina, Suez Channel, Nile Delta,Port Said,El Smalia,Suez, Alexandri,Borg El Arab, EL Alameen,Marsa Matrooh,Seedy Branee, Sewaa Oasis,El Saloom,Part of EL Frfra,EL fayom,Almenya,Deshna,Nga Hmmad,Koos,ElOxsor,Esna,Edfoo,Koom Embo,Aswan, El-Khraka,Elwhaat Ekhargyaa,Elwhaat ElDakhelea,Paris	47.11%	µSivert/h
2	4.75E-02	8.85E-02	El Qoseama, El Qontella,EL Tor,EL Dabaa,Glala mountain,Monkhfad Elqtara,Maghagh,Elwhaat EL Bahria,Asyoot,Sohaag,El Frfra Oasiss,Elqooseer,Harahada	41.00%	µSivert/h
3	8.85E-02	1.43E-01	Ras Kareeb,Sharm El Shekh,Safaga,El Qosyear,El Shayeb Mountain, St Catherna, Tabaa	6.82%	µSivert/h
4	1.43E-01	2.53E-01	Middle of Sina and Part of Suez Golf	3.75%	µSivert/h
5	2.53E-01	4.59E-01	Outside of Newbaa city and Part of El Aqbaa golf	0.69%	µSivert/h
6	4.59E-01	7.36E-01	Surounding of Newbaa City	0.36%	µSivert/h
7	7.36E-01	1.088	Newbaa City	0.27%	µSivert/h

From table 3 we can expoler the following results:-

1. About 88.11% of Egypt’s area is belonging to interval one and interval two. In other wards “ About 88.11% of Egypt’s area its minimum Gamma radiation levels is 2.02E-2 µS/h and its maximum Gamma radiation levels is 8.852E-2 2 µS/h

2. The maximum Gamma radiation levels which its value is 1.088 µS/h is found in Newbaa City and its area represnts about 0.27% of Egypt.
3. The middle and south of Sina have higher concenteration of Gamma radiation levels then other lands in Egypt.

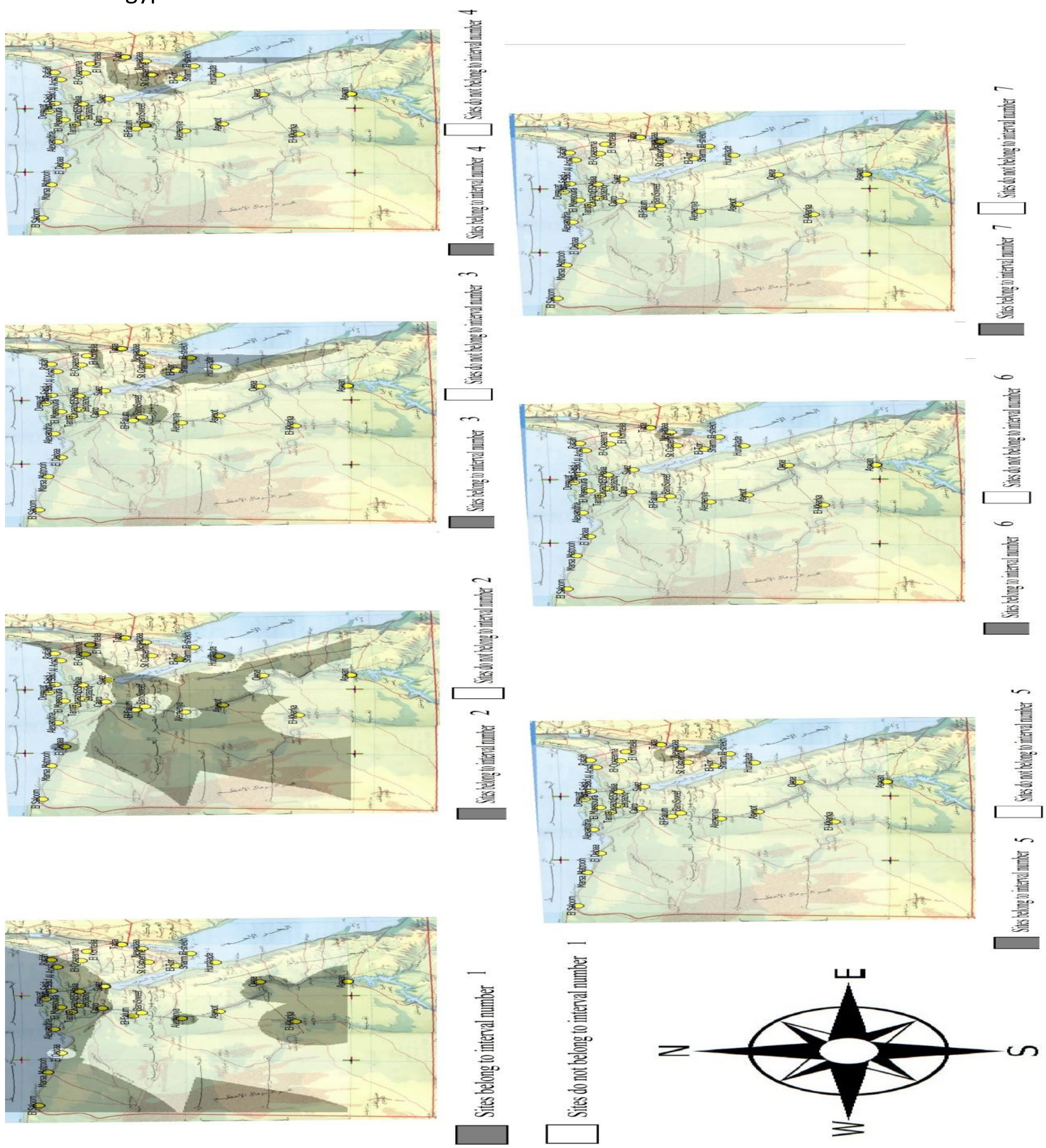

Figure 5: All the sites that belong to each interval in table 1

Conclusions

This case study proves that GIS and Data Mining Techniques is more useful than using famous methods for classifying perennial stream for the following reasons:-

1. This method has the ability to estimate the values of any stream for all points that have not any measured value with their GPS (Global positioning System) while other traditional methods require a huge effort to move to each point's location for measuring its stream value.
2. This method has the ability to generate an automated map to show the information of the stream values in a map while traditional methods don't supply this feature.
3. Data mining techniques as shown from this case study can be implemented with this method in easy way. With this feature a lot of knowledge can be discovered from some data while it is not easy to use data mining techniques with traditional methods

Using GIS and Data Mining Techniques for classifying many sites according to any kind of stream saves a lot of effort, time and cost. This method can be used in the processes of the evaluation of Risk Assessment and also making references for sitting any region according to any kind of stream.

References

[1] Gritzner, J. and Bruce Millett. 2003b. Stream Flow Definitions: Perennial, Intermittent, and Ephemeral Streams. Department of Geography, South Dakota State University, Draft Report.

[2]Gritzner, J. 2003a. USGS National Mapping Program Standards for Stream/River Features. Department of Geography, South Dakota State University, Draft Report.

[3]Gritzner, J. and Bruce Millett. 2003a. Approaches to Stream Delineation, Classification, and Changes in Stream Properties. Department of Geography, South Dakota State University, Draft Report.

[4] Peuquet, Donna J., and Niu Duan. "An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data." *International journal of geographical information systems* 9.1 (1995): 7-24.

[5] Jiang, Bin. "Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution." *The Professional Geographer* 65.3 (2013): 482-494.

[6] Bergh, Jöran, and Jörgen Löfström. *Interpolation spaces: an introduction*. Vol. 223. Berlin: Springer, 1976.

[7] Triebel, Hans. "Interpolation theory, function spaces, differential operators." (1978).

[8] Spielman, Seth E., and Jean-Claude Thill. "Social area analysis, data mining, and GIS." *Computers, Environment and Urban Systems* 32.2 (2008): 110-122.

[9] Miller, Harvey J., and Jiawei Han, eds. *Geographic data mining and knowledge discovery*. CRC Press, 2009.

[10] Li, Deren, and Shuliang Wang. "Spatial data mining and knowledge discovery." *Advances in Spatio-Temporal Analysis* 5 (2007): 173.