Polish J. of Environ. Stud. Vol. 16, No. 4 (2007), 531-538

Original Research

Mining Association Rules Applied to Goethite and Haematite Abundances in Manganese-Contaminated Soils

G-I. E. Ekosse[1]*, N Forcheh[2]

[1]Mining and Minerals, School of Physical and Mineral Sciences, University of Limpopo, P/Bag X1106 Sovenga 0727, South Africa
[2]Department of statistics, University of Botswana, P/Bag 0022 Gaborone, Botswana

Received: July 27, 2006
Accepted: February 18, 2007

Abstract

Mining association rules (MAR) are applied in elucidating on the abundances and association of Fe minerals in soils within the proximity of an abandoned manganese oxide (MnO_x) mine. Four hundred soil samples were obtained from a 4 km^2 area close to the abandoned mine in Kgwakgwe, Botswana. The Fe minerals in the soil samples were identified by x-ray diffractometry. Results depicted haematite and goethite present in different abundances: none, trace, minor and major in soils from the study area but only haematite in soils from the control site. From 16 possible combinations of the two Fe minerals in the soil samples, MAR results for confidence, certainty factor, lift and support, depicted haematite in major quantity to be the most influential. This fact is substantiated by its occurrence in the country shales.

Keywords: confidence, certainty factor, lift, support, spatial distribution, X-ray powder diffraction

Introduction

Geochemical sources are the main sources of Fe occurring as primary minerals mainly in igneous rocks, and secondary minerals in metamorphic and sedimentary rocks. The weathering of these rocks causes Fe ions to occur in regoliths, eventually becoming chemical constituents of soils. It could also be a result of anthropogenic activities including mining. Iron occurrence in soils is mostly in the form of Fe minerals in chemical combinations with oxygen: its oxides Fe_xO_y, and the common forms being hematite (Fe_2O_3) and magnetite (Fe_3O_4) [1]; and its hydroxide form, goethite (FeO. OH). Hematite is pink to bright red, magnetite is reddish brown, and goethite is brown to dark yellowish brown.

Anthropogenic activities primarily promoted by mining have affected the chemistry and mineralogy of the soils at Kgwakgwe, Botswana, where Mn oxide ore was mined for over two decades. Mining activities introduced Fe ions into the environment, causing its high occurrences in the soils there. The Fe-bearing minerals (goethite and haematite) have been identified in substantial quantities [2, 3]. Structural contamination of these minerals in soils and claybodies is possible through isomorphous substitution in both the tetrahedral and octahedral sheets. In well aerated soils rich in oxygen, Fe ions are oxidized, becoming part of the soil solids. In the oxidation of sulphides of Fe, acidic solutions are created which tend to decrease adsorption and promote mobility of metals in soils, water and sediments. Although Fe is essential for plant health, it could be toxic if its concentration is high; manifested by reduced and stunted growth and, in extreme cases, death.

*Corresponding author; e-mail: ekosseg@ul.ac.za

In order to set in place control measures in abating the abundances of Fe-rich minerals in contaminated soils, it is imperative to first understand their relationship of association. In attempting to understand the statistical relationship of Fe-rich minerals in a given environment, one would need to establish whether these minerals occur together in the same samples, and which of the minerals is more dominant within a given sample. This will lead to trying to understand why such abundances and association patterns occur by interpreting the statistical findings to the understanding of the environmental mineralogy related to the Fe minerals. The authors believe that association rules could be used as a tool in understanding the association and distribution of the Fe minerals within the existing environmental setting.

Association rules, which have gained considerable popularity over the last decade, is a method of studying associations between variables. Agrawal et al. [4] are widely credited with introducing the method in 1993 [5-7]. Association rule methodology is primarily concerned with extracting interesting rules from a large database. The extraction of such interesting rules is termed mining association rules (MAR) [6] or association rule mining (ARM) [7, 8]. In the intervening 13 years, MAR has become one of the best studied areas of data mining [6]. Along with other data mining techniques, association rules have mainly been used in market research studies, where researchers are interested in answering questions such as how often does a shopping basket that contains meat also contain wine? In data mining literature, the method is sometimes referred to as unsupervised learning.

Interestingly, MAR does not appear to have been adopted in the study of physical, chemical and mineralogical parameters occurring in the biophysical environment. In Kgwakgwe, Botswana, multivariate techniques have been used to understand the environmental association of Fe minerals in the soils [2]. Debut work on MAR related to goethite and haematite present in Mn-contaminated soils at Kgwakgwe, Botswana, was carried out in this study. The primary objective of the exercise was to elucidate on the applicability of MAR on the Fe mineralogy of soils. The Kgwakgwe area is in the southeastern part of Botswana located between latitudes 24°59’ and 25°02’, and longitudes 25°17’ and 25°20’. An abandoned manganese oxide (MnO_x) mine exits there.

Hahsler [5] provides an excellent web-based annotated bibliography on ARM. Additional publications include Li et al. [6], Yan and Chen [7], Melab [8], Berzal [9] and Richards and Rayward-Smith [10]. In keeping with literature, association rules methodology is introduced using market research terminology on which it was founded. Considering an example of bread, cheese and wine [8], let T be a set of transactions stored in a database, where each transaction, t, consists of items such as the items that a customer buys during a particular shopping errand. The transaction for a custumer could be a shopping basket containing bread, cheese and wine. Such a transaction is called a 3-item set transaction. If t denotes a typical transaction of a customer, then for this customer, t= [bread, cheese, wine]. The transaction, X= [bread, wine] is then a 2-item subset of t. Suppose that X and Y are two non-intersecting item subsets of a transaction, t, for example, X = [bread, cheese] and Y= [wine], then an association rule between X and Y is a rule, $r: X \Rightarrow Y$. The rule *r* is interpreted in this case as meaning that customers who buy bread and cheese are likely to buy wine with a certain probability.

Unlike traditional classifiers, association rules do not attempt to make a prediction for all database records [11]. The rule $r: X \Rightarrow Y$ makes no prediction of Y in the entire database. As a result, several measures have been developed to assess the accuracy and usefulness of each association rule extracted from the database. Hahsler [5] gives a comprehensive summary of commonly used and recent additions to measure the significance of association rules. The following notation is used in the review of these measures; *X* and Y denote most non-intersecting item subsets in a database, *N* denotes the total number of transactions in the database, $n(X) = freq(X)$ denotes number of transactions that include *X* as an item set or item subset.

In the original paper, Agrawal et al. [4] introduced the support-confidence framework. Other frameworks for association rules include coverage, lift or interest, conviction, and certainty factor. These are briefly explained. The degree of support for the rule $X \Rightarrow Y$ or simply its support is the probability that a transaction contains both X and Y [2, 8]. It gives the likelihood that a transaction containing both X and Y will be found in the database. In practice, support is defined on item sets, and gives the proportion of transactions which contain the item set [5]. In ARM, interest is in finding frequent rules. A rule is said to be frequent if its degree of support is greater than a user-defined minimum: *minsup*. Agrawal et al. [4] recognized that rules involving some rare but interesting items would be excluded if the measure of support was the only criteria used to mine association rules. To address this limitation, the measure of confidence was introduced. Hence along with the measure of support, the degree of confidence for a rule was also defined. Interest is usually to identify rules whose measures of confidence are greater than some user-defined minimum threshold: *minconf.* Such roles are referred to as strong rules [8, 9].

The two rules $(X \Rightarrow Y)$ and $(Y \Rightarrow X)$ have the same measure of support given by $n(X \cup Y)/N$ but different measures of confidence given respectively by $n(X \cup Y)/n(X)$ and $n(X \cup Y)/n(Y)$, hence confidence suffers from being directional, and also from being sensitive to the frequency of the antecedent, X. In practice, support is used to identify frequent rules and confidence is then used to determine which of the frequent rules are “strong.” Another limitation of confidence is that rules with high support also have high confidence since $Sup(X \Rightarrow Y) \leq Sup(X)$ are both bound by 1 (100%). The measure of coverage or antecedent support given by $Sup(X)$, measures how often the rule is applicable in the data set [5].

The measures of lift show how much more likely X and Y occur together in the database, compared with the expected number of times they can occur together if they were independent. Its properties are therefore similar to those of odds ratio in logistic regression.

$$Lift(X \Rightarrow Y) = \frac{P(X \text{ and } Y)}{P(X)P(Y)} = \frac{Sup(X \Rightarrow Y)}{Sup(X)Sup(Y)} =$$

$$= \frac{Conf(X \Rightarrow Y)}{Sup(Y)} = \frac{Conf(Y \Rightarrow X)}{Sup(X)}$$

A lift of 1 indicates that items X and Y occur independently, a value larger than 1 implies positive association between X and Y, while a value less than 1 implies negative association. Also, lift is not directional; hence the two rules $(X \Rightarrow Y)$ and $(Y \Rightarrow X)$ have the same lift. Lift was initially called 'interest' when introduced in 1997 by Brin *et al.* [12] (cited in Hahsler [5] and Berzal *et al.* [9], and was designed as a measure to find dependencies in the database. Despite its attractiveness, lift has been criticized for being unduly influenced by rare item sets. This is because when $Sup(X) \to 0$ or $Sup(Y) \to 0$ then $Lift(X \Rightarrow Y) \to \infty$.

The measure of conviction attributed to Brin et al. [12] by Hahsler [5] was developed as an alternative to confidence and is expressed in equation as follows:

$$Conv(X \Rightarrow Y) = \frac{P(X)P(not\ Y)}{P(X \text{ and not } Y)} = \frac{Sup(X)Sup(\neg Y)}{Sup(X \cup \neg Y)} =$$

$$= \frac{1 - Sup(Y)}{1 - conf(X \Rightarrow Y)}$$

Conviction has the same range of values as lift, and similar interpretation. Like lift, differences between conviction values are not meaningful, since it does not have an upper bound, and hence it is difficult to define a conviction threshold.

Berzal et al. [9] adopted the concept of certainty factor (CF) from earlier works by Sortliffe and Buchanan [13]. They define certain factor as follows:

$$CF(X \Rightarrow Y) = \begin{cases} \dfrac{Conf(X \Rightarrow Y) - Sup(Y)}{1 - Sup(Y)} & \text{if } Conf(X \Rightarrow Y) > Sup(Y) \\ \dfrac{Conf(X \Rightarrow Y) - Sup(Y)}{Sup(Y)} & \text{if } Conf(X \Rightarrow Y) < Sup(Y) \\ 0 & \text{Otherwise} \end{cases}$$

Berzal et al. [9] showed that CF verifies three basic criteria that any accuracy measure is expected to satisfy. These properties claim that any accuracy measure must test the independent assumption that if X and Y are independent, then $Sup(X \Rightarrow Y) = Sup(X)Sup(Y)$, it must monotonically increase with support for the rule when all other parameters are held fixed, and must monotonically decrease with support for its antecedent Sup (X) or precedent Sup (Y) when all other parameters are held fixed.

Experimental Procedures

Samples and Sampling

The rocks of the Kgwakgwe basin are of the Paleoproterozoic Transvaal Supergroup, capped by the younger Waterberg Group, and underlain by an older Kanye Volcanic Formation (KVF) (Table 1). The Black Reef Quartzite Formation (BRQF) is the lowest stratigraphic unit of the Transvaal Supergroup and overlies the rhyolites of the KVF with unconformity. The Kgwakgwe Shale (KS) succeeds the BRQF, and it consists of varicoloured manganiferous and ferruginous shale units belonging to the Taupone Dolomite Group (TDG). The manganiferous and ferruginous shales and siltstones are directly underlain by the BRQF, which is the basal formation of the rocks of the Transvaal Supergroup in large parts of South Africa, extending northwards into Botswana. The TDG represents the lower part of the Paleoproterozoic [14] of the Transvaal Supergroup in Botswana and South Africa [15] having an age of 2500 – 2000 Ma (where Ma is million year). The lower section of the TDG is the host rock of the Mn oxides ore that was mined in the past. Both Mn and Fe in the soils at Kgwakg-

Table 1. Lithostratigraphy of the Kgwakgwe basin [3, 16, 17].

Lithology	Formation	Group	Supergroup	Age
Sandstones		Waterberg		1700 Ma
Chert clast breccia	Kgwakgwe Chert Breccia	Taupone Dolomite	Transvaal	2500 Ma
Varicolored manganiferous and ferruginous shale	"Kgwakgwe Shale"			
Quartzite, shale conglomerate	Black Reef Quartzite			
Feldspathic rhyolites	Kanye Volcanic	Lobatse Volcanic (LVG)		2780 Ma

we area are believed to have originated mainly from this geologic formation.

The study site from where soil samples were obtained was 2 km x 2 km and a control site, located 4 km south of the study site, having an area of 900 m^2 (300 m x 300 m). The control site was chosen because it was at the other side of a paleotopographic barrier where Mn mineralization has not occurred [16]. A detailed soil grid of 2 km x 2 km was established for the study site (Fig. 1). The coordinates and the grids from where soils were sampled are reflected on the satellite image which was produced from Quick bird imagery. This is a multispectral standard satellite image covering four bands within the range of 450–900 nm representing visible (blue, green, red) and near infrared bands. The image had a 2.4–2.8 m resolution with zero cloud cover.

Four hundred soil samples were collected from the study site and nine samples from the control site for analyses. Random sampling and judgmental sampling techniques were used in obtaining the soil samples [18, 19]. Soil samples were taken at 100 m intervals, and at a depth of between 0 cm and 20 cm. The obtained samples were placed in an oven at 60°C overnight for the release of surface soil moisture, prior to analyses.

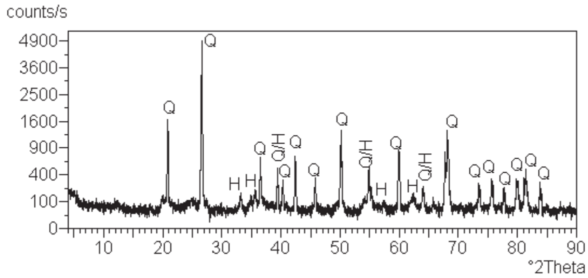

Fig. 1. X-ray powder diffractogram of a representative soil sample (Q = quartz; H = haematite).

X-ray Powder Diffractometry

The X-ray powder diffraction (XRPD) technique was used to determine the ferruginous minerals contained in the soil samples. The dried samples were pulverized using a WC seib mill for 30s and the resulting fine powder was then mounted on the sample holder, and scanned in the XRPD equipment at a speed of 1°2θ / min, and diffractograms recorded from 2°2θ to 70°2θ. The XRPD equipment was a Philips PW 3710 system, operated at 40 kV and 45 mA, having a Cu-K_α radiation and a graphite monochromator. A PW 1877 Automated Powder Diffraction X'PERT Data Collector software package was employed to capture raw data, and a Philips X'PERT Graphics & Identify software package was used for qualitative identification of the minerals from both the data and patterns obtained by scanning. Interpreted results were compared with data and patterns available in the Mineral Powder Diffraction File, data book and the search manual issued by the International Center for Powder Diffraction Data (ICDD) [20], for confirmation. Identified minerals were semi quantitatively analyzed to establish their relative abundances through quantitative measure (numbers) where none = 0, trace = 1, minor = 2 and major = 3.

Statistical Methods

Descriptive statistics were used to understand the abundances of goethite and haematite within the study site. Chi-squared test of association was used to determine whether the presence of haematite in a given site was associated with the occurrence of goethite at that same site, or whether they occur independently. In addition, biplots were used to give a graphical representation of any association between the abundances of haematite and goethite.

Methodology of MAR was then used to give more in-depth understanding of how the occurrence or absence of one form of Fe mineral at a given site implied the occurrence of the other form at that site. Each mineral can occur as a major, minor or trace or it could be absent. Hence there are 2^4 = 16 possible rules of the form $(H = h) \Rightarrow (G = g)$ that could be mined out of haematite and goethite occurrences. The notation $(H = h) \Rightarrow (G = g)$ is used to indicate that the presence of hematite as *h,* where $h \subset$ *(absent, trace, minor, major*) implies the occurrence of goethite as *g,* $g \subset$ *(absent, trace, minor major*). This is occasionally abbreviated as: $(H \Rightarrow G)$. The case where both haematite and goethite are absent (i.e. *h=absent* and *g=absent*) was considered to be uninteresting, and was not studied.

Based on review of literature, four measures are used to assess the quality of each association rule. These are measures of support, confidence, lift and certainty factor (CF). The method of their estimation follows.

$$Sup([H = h] \Rightarrow [G = g]) = \text{Prop (Sites haematite} = h \text{ and goethite} = g) \quad (1)$$

$$Conf([H = h] \Rightarrow [G = g]) = Sup([H = h] \Rightarrow [G = g]) / Sup(H = h) \quad (2)$$

where $Sup(H = h)$ = Prop (sites Haematite = h)

$$Lift([H = h] \Rightarrow [G = g]) = \frac{Sup([H = h] \Rightarrow [G = g])}{Sup(H = h)Sup(G = g)} = \frac{Conf([H = h] \Rightarrow [G = g])}{Sup(G = g)} \quad (3)$$

$$CF(X \Rightarrow Y) = \begin{cases} \dfrac{Conf([H=h] \Rightarrow [G=g]) - Sup(G=g)}{1 - Sup(G=g)} & \text{if } ([H=h] \Rightarrow [G=g]) > Sup(G=g) \\ \dfrac{Conf([H=h] \Rightarrow [G=g]) - Sup(G=g)}{Sup(G=g)} & \text{if } ([H=h] \Rightarrow [G=g]) < Sup(G=g) \\ 0 \qquad \text{Otherwise} & \end{cases} \tag{4}$$

Results

Minerals Identification by X-ray Diffractometry

Figs. 1 and 2 are representative X-ray diffractograms of analyzed soil samples. The ferruginous shale consisted mainly of haematite and goethite with minor quantities of illite and kaolinite. Soil samples consisted mainly of quartz, the clay minerals (kaolinite and illite) and the Fe minerals (haematite and goethite). At the control site, soil samples consisted of quartz and haematite in major quantities, and kaolinite in a minor quantity. The haematite and goethite contained in the soils most possibly originated from the ferruginous shale. At the control site haematite was possibly derived from the Fe-rich country rocks (dolerite and rhyolite).

Descriptive Statistics

The abundance of haematite and goethite in the study area is summarized in Table 2. The results for both the study and control site reveal that goethite rarely occurs as a major, with only 7 of the 400 (1.75%) sites having goethite as a major. Similarly, haematite rarely occurs in minor form, with only 3% having haematite in minor form. Both Fe minerals were absent in 159 sites (39.75%) . In the study area, haematite was found in just under half of the sites (49.25%), while goethite was found in only 28.5% of sampled sites. When present, haematite occurred mainly as a major mineral (in 30.5% of sites), and rarely occurred as a minor (3% of sites). Goethite occurred mainly as a trace (19% of sites), and rarely as a major (1.75% of sites).

A chi-squared test of association between the abundance of goethite and haematite gave a Pearson Chi-Square statistic of 37.2 on 9 degrees of freedom and $p<0.001$. This suggested a strong association between the occurrence of hematite and that of goethite. However, a high percentage of expected values (43.8%) were < 5, due mainly to the low prevalence of goethite as a major and hematite as a minor. When the major and minor categories were combined, and the analysis repeated, it was found that all expected frequencies were > 5 and the association between the two minerals remained highly significant ($p<0.001$). Fig. 3 shows the distribution of percentage of sites having each combination of occurrences of both minerals

The distribution in Table 2 was subjected to correspondence analysis and the biplot shown in Fig. 4 was obtained. Further tests revealed that the association is not linear. All measures of correlation between ordinal variables such as Kendall's tau, gamma and sommer's d were not significantly different from zero ($p=0.067$). The biplot suggests that sites where goethite is absent tend to be associated with sites where hematite is absent or is major, whereas sites where goethite is trace tend to be associated with sites where hematite is also trace.

Mining Association Rules

Results obtained from MAR are presented in Table 3. The rule with highest support is rule #12: (Haematite = Major) => (goethite = absent), with a support of 35.27%. This is followed by the rules 4 (Haematite = trace) => (goethite = absent); 13 (Haematite = major) => (goethite = trace); and 5 (Haematite = trace) => (goethite = trace) with supports of 14.5%, 11.2% and 10.4% respectively. Other very interesting rules based on support are rules 1 and 2.

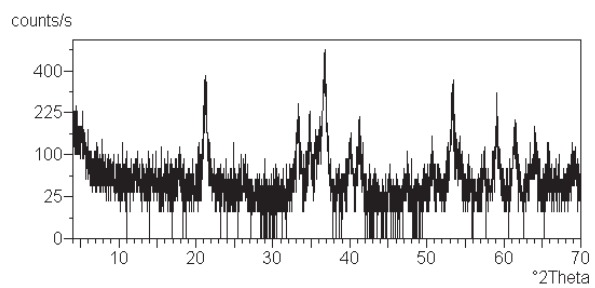

Fig. 2. X-ray powder diffractogram of a representative soil sample depicting all peaks being that of goethite.

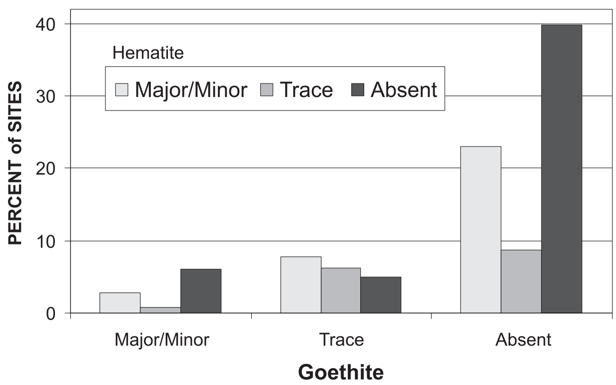

Fig. 3. Distribution of percentage of sites by combination of occurrences of both goethite and haematite abundances.

Table 2. Numbers and percentages of study sites by abundance of hematite and goethite minerals.

	Hematite				
Goethite	Major	Minor	Trace	Absent	Total
Major	3	0	2	2	7
Minor	7	1	1	22	31
Trace	27	4	25	20	76
Absent	85	7	35	159	286
Total	122	12	63	203	400
Major	0.75	0.00	0.50	0.50	1.75
Minor	1.75	0.25	0.25	5.50	7.75
Trace	6.75	1.00	6.25	5.00	19.00
Absent	21.25	1.75	8.75	39.75	71.50
Total	30.50	3.00	15.75	50.75	100

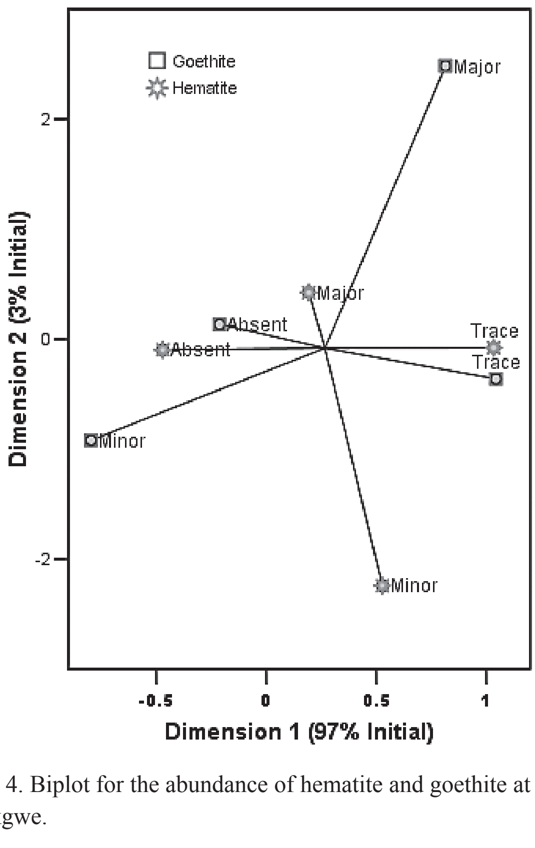

Fig. 4. Biplot for the abundance of hematite and goethite at Kgwakgwe.

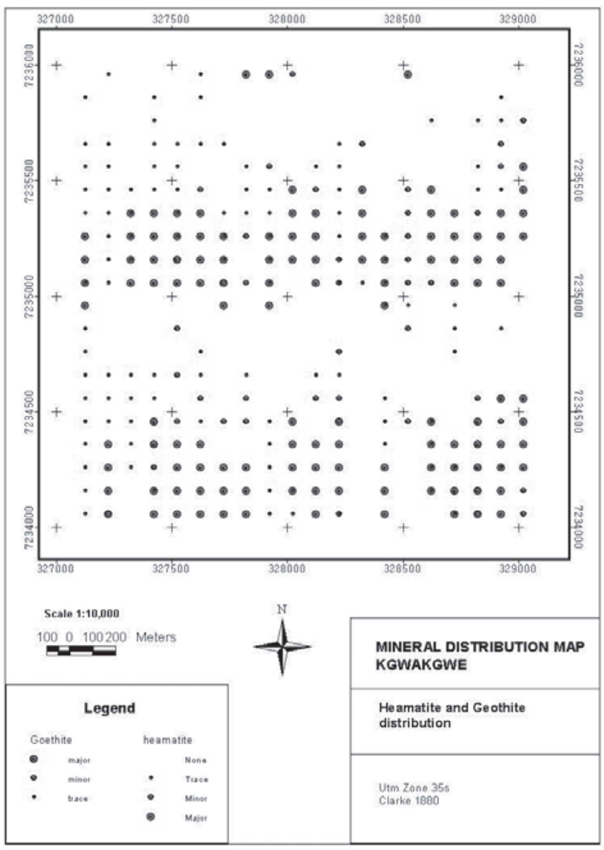

Fig. 5. Spatial distribution of haematite and goethite at the study area, Kgwakgwe.

Considering the lift and certainty factor measures, it can be observed from Table 3 that when Haematite is absent the CFs are all positive. This confirms that the absence of haematite is positively associated with the presence of goethite, especially as minor (high positive CF). Positive association between the absence of haematite and presence of goethite especially as a minor is also verified by the lift measures which are all greater than 100. The presence of haematite as trace is negatively associated with the presence of goethite as minor as evidence from the small lift (12.3%) and negative certainty factor (-12.9).

However, the presence of haematite as trace is most highly associated with the presence of goethite as trace as well (lift=125.8%, CF=25.8), and slightly positively associated with the presence of goethite as major (lift=109.3, CF=9.3).

Table 3. Results of Association Rules (all quantities expressed as percentages).

Rule #	Haematite ==>	Goethite=	Count	Sup(H=>G)	Conf(H=>G)	Lift(H=>G)	CF(H=>G)
1	Absent	Trace	20	8.30	45.5	144.1	44.1
2		Minor	22	9.13	50.0	388.7	288.7
3		Major	2	0.83	4.5	156.5	56.5
		Total	44	18.26	100.0	100.0	**0.0**
4	Trace	Absent	35	14.52	55.6	105.4	6.0
5		Trace	25	10.37	39.7	125.8	25.8
6		Minor	1	0.41	1.6	12.3	-12.9
7		Major	2	0.83	3.2	109.3	9.3
		Total	63	26.14	100.0	100.0	**0.0**
8	Minor	Absent	7	2.90	58.3	110.7	11.9
9		Trace	4	1.66	33.3	105.7	5.7
10		Minor	1	0.41	8.3	64.8	-5.2
11		Major	0	0.00	0.0	0.0	-3.0
		Total	12	4.98	100.0	100.0	**0.0**
12	Major	Absent	85	35.27	69.7	132.2	35.9
13		Trace	27	11.20	22.1	70.2	-13.7
14		Minor	7	2.90	5.7	44.6	-8.2
15		Major	3	1.24	2.5	84.7	-0.5
		Total	122	50.62	100.0	100.0	**0.0**
	Total	Total	400	100	100	100	**0.00**

The presence of haematite as minor or major are both negatively associated with the presence of goethite as minor or major and most highly positively associated with the absence of goethite. The presence of haematite as a minor is most highly negatively associated with the presence of goethite as minor or major, while the presence of haematite as a major is most highly negatively associated with the presence as trace (CF = -13.7), minor (CF = -8.2) and essentially independent of the presence of goethite as a major (CF = -0.5). The presence of haematite as minor is negatively associated with the presence of goethite as minor or major as evidence from the small lift (12.3%) and negative certainty factor (-12.9).

Discussion of Results

The results obtained from MAR in these minerals thus provide adequate answers as to how the presence of haematite is associated with the presence or absence of goethite at the different sites. Of the 16 different combinations for the two minerals, haematite in a major quantity is projected as having a very strong influence in the soils. Where there is absence of haematite, there is the possibility of absence of goethite. Both could also occur in trace quantities but rarely did they occur together in minor and major quantities. Because this work concentrates on understanding the association of Fe minerals in the studied soils, complementary GIS presentation is invoked to support the MAR findings. With the aid of the MSS image of the study area, haematite and goethite abundances were processed using the integrated Land and Water Information System (ILWIS), Geosoft Oasis Montaj (version 4.2) and ArcGIS software packages into a minerals distribution map. Plotting the spatial distribution of the two minerals (Fig. 5) and observing their distribution patterns substantiated and confirmed the findings obtained by MAR. The spatial distribution map exhibited similar association as that obtained by MAR.

This important observation deduced from the interpretation of the results obtained by the application of MAR could be further supported by the geochemistry of oxidising and reducing environments. Depending on favorable geochemical conditions, haematite from the ferruginous shale released Fe ions through dissolution as shown in the equation below. The release caused Fe ions to follow migratory pathways, and become recrystallized as goethite as conditions changed [3].

$$\underset{\text{haematite}}{Fe_2O_3} + H_2O \rightarrow \underset{\text{goethite}}{2FeO{\cdot}OH}$$

When geochemical conditions change, haematite is formed from goethite. These geochemical conditions were non existent at the control site; hence an explanation for the absence of goethite there.

It could be concluded that the occurrence of types of Fe minerals in the soils of the study area were possibly influenced primarily by the Fe source and weathering patterns. Due to occurrences of goethite in the sampled ferruginous shale, and the soils of the study area, but not in samples from the control site, it is evident that the main source of goethite in soils was from the shale. Sediments containing exposed goethite particles eroded and were transported either by wind or water (meteoric fluids and streams) to contaminate surrounding soils. The study thus lays foundation for future applications of MAR in environmental mineralogy. Through the aid of MAR, results could be used in assisting interpretation of geochemical and mineralogical processes governing mineralization.

References

1. BOHN H. L., MCNEAL B. L., O'CONNOR G. A. Soil Chemistry. 3rd Edition. John Wiley & Sons. pp. 307, **2001**.
2. EKOSSE G., FOUCHE P. S. Environmental association of iron minerals and iron concentrations in soils close to an abandoned manganese mine – a multivariate analytical approach. Journal of Applied Sciences and Environmental Management **10** (1), 31, **2006**.
3. EKOSSE G., MODISI M. P. Mineralogical characterization, genesis and possible economic applications of goethite occurrences at the Lower Transvaal Supergroup, Kgwakgwe, Botswana. Botswana Journal of Earth Sciences **4**, 6, **1999**.
4. AGRAWAL R., IMIELINSKI T., SWAMI A. Mining association rules between sets of items in large databases, in: Proc. of the 1993 ACM SIGMOD Conference, pp.207–216, **1993**.
5. HAHSLER M. A. Comparison of commonly used interest measures of association rules. http://wwwai.wu.wien.ac.at/~hahsler/research/association_rules/measures.html accessed on (10/29/2005).
6. LI Q., FENG L., WONG A. From intra-transaction to generalized inter-transaction: Landscaping multidimensional contexts in association rule mining. Information Sciences **172**, 361, **2005**.
7. YAN P., CHEN G. Discovering a cover set of ARsi with hierarchy from quantitative databases. Information Sciences **173,** 319, **2005**.
8. MELAB N. Information Data Mining**:** A Key Contribution to E-business. & Communications Technology Law*,* **10** (3), 309, **2001**.
9. BERZAL F., BLANCO I., S'ANCHEZ D., VILA, M. A. Measuring the accuracy and interest of association rules: A new framework. Intelligent Data Analysis **6,** 221, **2002**.
10. RICHARDS G., RAYWARD-SMITH V. J. The discovery of association rules from tabular databases comprising nominal and ordinal attributes. Intelligent Data Analysis **9,** 289, **2005**.
11. SCHEFFER T. Finding association rules that trade support optimally against confidence. Intelligent Data Analysis **9,** 381, **2005**.
12. BRIN S., MOTWANI R., ULLMAN J. D., TSUR, S. Dynamic itemset counting and implication rules for market basket data. In % Of the ACM SIGMOD Int'l Conf. on Management data (ACM SIGMOD'97), pp. 265-276, **1997**.
13. SHORTLIFFE E., BUCHANAN B. A model of inexact reasoning in medicine, *Mathematical Biosciences,* **23,** 351, **1975**.
14. CARNEY J. N., ADLISS D. T., LOCK N. P. The geology of Botswana. Bulletin of the Geological Survey Department, Botswana. **37**, 113, **1994**.
15. TOMBALE A. R. Geological map of the Jwaneng area (QDS 2424D)(1:125 000) with brief description, Geological Survey of Botswana, Government Printer, Gaborone, Botswana, **1986**.
16. EKOSSE G., VINK B. W. The Kgwakgwe manganese oxides in the Kanye area, southeast Botswana – evidence for shallow aqueous paleo-environment. Botswana Notes and Records **30**, 147, **1998**.
17. EKOSSE G., VINK B. W. The geology and mineralogy of the manganiferous, ferrugenous and argillaceous sediments in the Kgwakgwe basin, Lower Transvaal Supergroup, Kanye area, Botswana. Botswana Journal of Earth Sciences **5**, 8, **2001**.
18. CRÉPIN J., JOHNSON R.L. Soil sampling for environmental assessment. In Carter R. M. (Ed) Soil sampling and methods of analyses, Boca Rotan LEWIS Publishers, pp. 823, **1993.**
19. JEWELL M.C., HENSLEY P. J., BARRY D.A., ACWORTH I. Site investigation and monitoring techniques for contaminated sites and potential waste disposal sites. In Fell R., Philips T., Gerrad C., (Eds) Geotechnical management of waste and contamination, (Rotterdam: A. A. Balkema), pp. 3-39, **1993**.
20. INTERNATIONAL CENTRE FOR DIFFRACTION DATA. International Centre for Diffraction Data. Mineral Powder Diffraction File Databook. pp. 942, **2001**.