International Scholars Journals

Advances in Food Science and Technology ISSN: 6732-4215 Vol. 7 (2), pp. 001-009, February, 2019. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Graphical visualisation and domain partitioning of minerals in clay fraction of soils from Capricorn District, South Africa

G. I. E. Ekosse[1]*, K. S. Mwitondi[2] and F. T. Seabi[3]

[1]Directorate of Research Development, Walter Sisulu University, P/Bag 11 Mthatha, Eastern Cape 5117, South Africa.
[2]Faculty of Arts, Computing, Engineering and Sciences, Sheffield Hallam University, S1 1WB, UK.
[3]Agricultural Research Council, Pretoria, South Africa.

Accepted 28 July, 2018

Modeling techniques were used to study minerals in clay fraction of soils from Capricorn District, Limpopo Province, South Africa. Minerals in the clay fraction of soils were identified by X-ray diffraction (XRD) technique and semi-quantified. The minerals were then subjected to a combination of exploratory data analysis (EDA), graphical visualisation and domain-partitioning techniques in order to determine their cross-influence to one another in terms of abundances. Quartz and kaolinite were major dominant minerals in the soils; smectite, feldspar and mica were in minor to trace quantities. Consensual associations among other traces and high quantities of minerals were detected. Evidence of relationship using EDA portrayed general skewness in favour of quartz and kaolinite. Quartz remained dominant in the soils but with a consistent high probability of co-existence with kaolinite. Where there is low quartz content, kaolinite increased with the drop in quartz made up for by a combination of smectite, mica and feldspar. The nested nature of interaction also revealed indirect relationship between quartz and mica. The tree model, which yielded 100% accuracy, showed smectite as the first important mineral in identifying whether there is high, medium or low quartz content in the sols. Down the line the model relies heavily on both mica and kaolinite. Collating the minerals contents and data modeling procedures, *inter alia*, it could be inferred that the weathering of feldspar and mica may have an impact on the mineralisation of kaolinite and smectite; which are both important minerals in several agricultural applications.

Key words: Conditional probability, decision trees, domain-partitioning, feldspar, kaolinite, smectite.

INTRODUCTION

Regional studies on soil mineralogy in South Africa (SA) over the past 25 years in relation to soil properties such as erodibility, susceptibility of minerals to dispersion and the importance of the clay mineral fraction to K- fixation were undertaken in an effort to achieve a better understanding of the soils and their behavior (Bühmann et al., 2004; Bühmann and Nell, 1999; Bühmann et al., 2002; Botha, 1992; Ekosse and Fouche, 2006). Most of the studies were concerned with the clay size fraction and that XRD was the dominant technique used for minerals identification. Stern et al. (1991) studied the effect of clay mineralogy on rain infiltration (IR), seal formation and soil loss on cultivated soils from SA. Comparing the IR and soil losses, it was found that the soils comprised of mainly kaolinite and illite with traces of smectite. Further, kaolinitic or illitic soils containing traces of smectite were dispersive and susceptible to seal formation as were smectitic soils.

Soil clay minerals influence agricultural land use, soil fertility and productivity. Studies already undertaken on soil clay mineralogy in South Africa were not stretched to accommodate agricultural concerns of rural settings in Limpopo Province. Soil clay minerals are secondary minerals formed by low temperature reactions in the soil through weathering. The minerals influence the

*Corresponding author. E-mail: gekosse@wsu.ac.za, gekosse@excite.com.

physico-chemical, physical and chemical properties of soils and have a strong bearing on their usage in agriculture. The data for this study are based on three types of soils - Inceptic, Oxidic and Plinthic – obtained from the Capricorn District, Limpopo Province, South Africa. The Limpopo Province is predominantly rural with most of its people engaged in subsistence farming and the main objective of this study is to promote agricultural activities in the province through elucidation on minerals of clay fraction of soils.

The foregoing main objective is attained through a two-fold analytical approach - minerals identification and quantification on the one hand and clay soil data modeling on the other. The adopted methods EDA, GDV and domain partitioning techniques are typically non-parametric under which underlying parameters are estimated from the data. According to Mardia et al. (1979) and Van der Merwe et al. (2002), well-behaving, parametric models usually yield exact solutions but in most data modeling applications, the assumptions are often violated. On the other hand, the data-dependent parameter estimation inevitably introduces both training and testing randomness in modeling (Mwitondi, 2003; Mwitondi et al., 2002) which call for model accuracy and reliability.

MATERIALS AND METHODS

Study area and soil sampling

The study was carried out in South Africa (26°14' – 32°10'E and 25°25' – 21°49'S). Capricorn District in Limpopo Province of South Africa is comprised of 316,053 ha. Soils of the study area are believed to have been formed from granite, gneiss and migmatite. Their clay content ranges from 10 to 25% with soil depths of up to 120 cm. The area has three different types of soils: Soil type 1, 2 and 3. Soil type 1 (Inceptic) consisted of soils with a general yellow-brown apedal B horizon; having colour variation from yellow-brown to reddish soils. Soil type 2 (Plinthic) consisted of largely deep red soils portraying mainly red apedal B horizon; coupled with minimum occurrence of yellow-brown apedal B horizon. The topsoil was mainly orthic A horizon rich in organic matter. Soil type 3 (Oxidic) exhibited a high degree of weathering. Soil texture ranged from sandy to slightly clayey in some areas although the clay content was minimal. Soil sampling techniques similar to those adopted by Carter and Gregorich (2007) were used with the number of soil samples per soil type being determined on the basis of the coverage area of the soil type. A total of 21 samples collected at a depth of 20 cm were obtained from the study area.

Laboratory analysis

Collected samples were ground and passed through a 2 mm sieve, treated with 0.5 M of Sodium acetate buffer solution for the dissolution of carbonates and soluble salts. Prior to analysis, organic matter was removed from samples by oxidation with 30% H_2O_2 as described by Jackson (1979) and Bird and Chivas (1988). Separation of the particle size fractions was carried out in accordance to Stoke's law of sedimentation (Gaspe et al., 1994). Clay fraction samples were concentrated by sedimentation and mounted on sample holders for XRD analyses. Samples of clay fraction of soils were scanned from 2º 2 to 40º 2 and their diffractograms recorded. 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) (2001), for confirmation.

Data analysis

Data modeling was carried out using results of the minerals identification and quantification. The first step was to carry out an EDA aimed at providing initial insights into the identified minerals distribution. The EDA findings led to applications of GDV techniques before subjecting the samples to decision tree modeling for domain-partitioning. The subjection was to establish how the presence or absence of one or more minerals impacts on the presence or absence of the other or another mineral. This approach derived from probability theory is illustrated using generated XRD data as follows:

Let the data matrix $X = |\text{Quartz, Kaolinite, Feldspar, Smectite and Mica}|$ represent the XRD data labeled by the vector $C_k = Low, Medium, High$ of class labels chosen from one of the five clay minerals in X - that is, $k=1,\ldots,5$. Then, for predictive purposes, the following quantities are of interest:

(1) The probability density function, $f(X)$, for the random variable X which describes the probability density at each point in the sample space $C_k \in \Omega$.

(2) The class proportions for each of the three levels s_i , $i = 1, 2, 3$ which, effectively, describe the probability of class $P(C_k)$.

(3) The conditional density $f(X \mid C_k)$ defined as the probability density at each point in Ω given that the point belongs to one of the three group levels.

The above quantities provide a good intuition into the standard concept of conditional probability referring to the existence of a particular mineral given that a specified mineral exists in the sample as generally defined in Equation 1:

$$P(X \mid C_k) = \frac{P(X \quad C_k)}{P(C_k)} \Leftrightarrow P(C_k \mid X) = \frac{P(C_k \quad X)}{P(X)} \qquad (1)$$

We are therefore able to formulate various types of conditional probabilities depending on the problem under investigation. For instance, to determine how the presence or absence of quartz in clay fraction of soils affects the presence or absence of kaolinite in the same sample, we may examine the conditional probability and vice versa as shown in Equation 2:

$$P(Quartz \mid Kaolinite) = \frac{P(Quartz \quad Kaolinite)}{P(Kaolinite)} \qquad (2)$$

The same approach may be adopted in investigating any other interesting cross influences of minerals in sampled clay fraction of soils. On the basis of the two equations we can compute the posterior probabilities in which the phenomenon conditioned upon

typically precedes the one for which the probability is computed. Thus, if we are given the data X with known minerals composition and we want to allocate it to one of the known three classes, we can compute as follows:

$$P(s_i \mid x_i) = \frac{s_j f_j(x_i)}{\sum_{j=1}^{J=3} s_j f_j(x_i)} \quad \hat{} \quad (3)$$

where $i = 1, 2, ..., N; j = 1, 2, ...3;$ s_i is the individual labels for each of the data points, s_j is the group prior probability and $f_j(x_i)$ is the marginal density describing the distribution associated only with x_i in group j.

Basically, x_i will be allocated to the group maximizing the probability in Equation 3. The same reasoning applies to unsupervised learning when class labels are unknown in which case allocation will be made in accordance with, the closest distance to one of the groups.

For both GDV and domain-partitioning purposes, two versions of the data are used; the original continuous data and its discretised (categorical) version. The latter version was obtained by discretising each of the mineral variables across the soil sample to form three levels (Low, Medium and High) using a simple categorization procedure based on the quartile information from the data attributes. The discretised variables could then be used to establish probabilistic associations or each serving as a target variable in predicting class memberships of each observation using the remaining predictors in their continuous form. In addition to the graphical visualisation approaches, we extend the classification and regression tree theoretical foundations in Breiman et al. (1984) to carry out cross-decision tree modeling on both datasets. In both cases, the prediction of cross -dependence of the minerals proceeds via domain partitioning of the initial superset ST= by setting as targets both the highest and lowest attributes using the remaining variables as predictors.

RESULTS AND DISCUSSION

Minerals in clay fraction of soils

Results of minerals identification and semi-quantitative analyses are given in Table 1. Apart from quartz, kaolinite was the dominant clay mineral. Other minerals included feldspar and traces of smectite and mica. Results related to derived models are detailed from quartile- based discretised rules and classes for identified minerals (Table 2). Due to the fact that this study is interested in determining the cross-influence of the minerals in the clay fraction of sampled soils, emphasis is put on cross-mineral dependencies.

Exploratory data analyses

A set of initial EDA results is given in Figure 1 in which a pair-wise plot of the XRD data is presented with each mineral plotted against every other mineral. The distribution of the minerals across the samples was found to be generally skewed in favour of quartz and kaolinite especially distributions in soil type 3. Only in sample three in soil type 2 that the minerals seemed to have been relatively and evenly distributed. Only in samples one, five and eight of soil type 3 that the kaolinite contents were higher than those of quartz. Whereas samples one and five also had a high content of smectite, sample eight had none. The 20 ([5*(5-1)]) plots in Figure 1 which meant to provide insights into the potential bivariate relationships among the minerals offered no clearly discernible patterns among the paired minerals. It was therefore reasonable to carry out further investigation.

Graphical data visualisation techniques

Bivariate plots of each of the minerals with respect to quartz are presented in Figure 2. This form of graphical presentation of data similar to that of Cleveland (1993), provided good data visualization for quick identification of the presence or absence of patterns in well-behaving data. Similar plots with kaolinite, feldspar, smectite and mica on the vertical axis revealed no discernible patterns. The foregoing findings were in line with regression results with quartz/kaolinite yielding a negative relationship with a beta value of - 0.8403 and a very low correlation coefficient of just > 30%. In most of the paired relationships, the set zero hypothesis was rejected at 5%.

The foregoing illustrations are based on classical statistics. Thus the next analytical step was data subjection to Bayesian-based approach in which predictive knowledge was updated based on new information entered. The two panels in Figure 3 based on continuous data on the left hand side and its discretised version on the right provide insights into how the minerals in $C_k \in \Omega$ may be related. In both cases the observations are ordered from left to right and note that although the soil type labels are not given here, the data are ordered and therefore the first 3 cases from left to right are known to be from soil type1 and the last 15 from type 2. Further, each of the minerals in the discretised version was been coded using its first letter. The two mosaic plots may be viewed in the light of the conditional probabilities introduced above and provided information based on a mosaic presentation of the data in a conditional versus cumulative marginal probability dimension (Friendly, 1994). Typically, a contingency table could be used to present the data as stacked histograms conditioned on the minerals and in this case, the plot would display the conditional probabilities on the vertical axis. For the two plots, however, it suits our purpose to envision their full vertical range as representing the total probability $(0 \le p \le 1)$.

Table 1. Minerals identification and their quantifications in soil samples.

Soil type	Sample number	Quartz (wt %)	Kaolinite (wt %)	Feldspar (wt %)	Smectite (wt %)	Mica (wt %)
1	One	54	28	18	-	-
	Two	66	33	-	-	-
	Three	75	4	10	-	11
2	One	60	7	18	6	-
	Two	53	9	9	7	17
	Three	47	18	13	9	13
3	One	29	41	-	30	-
	Two	58	8	8	-	26
	Three	54	23	7	-	-
	Four	69	10	10	-	11
	Five	28	30	-	22	20
	Six	75	9	8	8	-
	Seven	62	20	4	11	-
	Eight	27	28	12	-	21
	Nine	32	17	11	19	-
	Ten	55	9	9	3	24
	Eleven	55	15	10	20	-
	Twelve	57	17	9	10	7
	Thirteen	41	24	11	11	13
	Fourteen	62	26	12	-	-
	Fifteen	50	25	-	-	13

Table 2. Quartile-based discretised rules and classes.

Minerals	Data summaries	Quartile-based mineral content levels		
		Low	Medium	High
Quartz	First Qrt: 47 Median 55 Mean 52.81 Third Qrt 62	Quartz < 47	47 < = Quartz < 53	Quartz > = 53
Kaolinite	First Qrt: 9 Median 18 Mean 19.1 Third Qrt 26	Kaolinite < 9	9 < = Kaolinite < 26	Kaolinite > = 26
Feldspar	First Qrt: 7 Median 9 Mean 8.5 Third Qrt 11	Feldspar < 7	7 < = Feldspar < 11	Feldspar >= 11
Smectite	First Qrt: 0 Median 6 Mean 7.4 Third Qrt 11	Smectite < 7	7 < = Smectite < 11	Smectite > = 11
Mica	First Qrt: 0 Median 6 Mean 8.4 Third Qrt 13	Mica < 7	7 < = Mica < 9	Mica > = 9

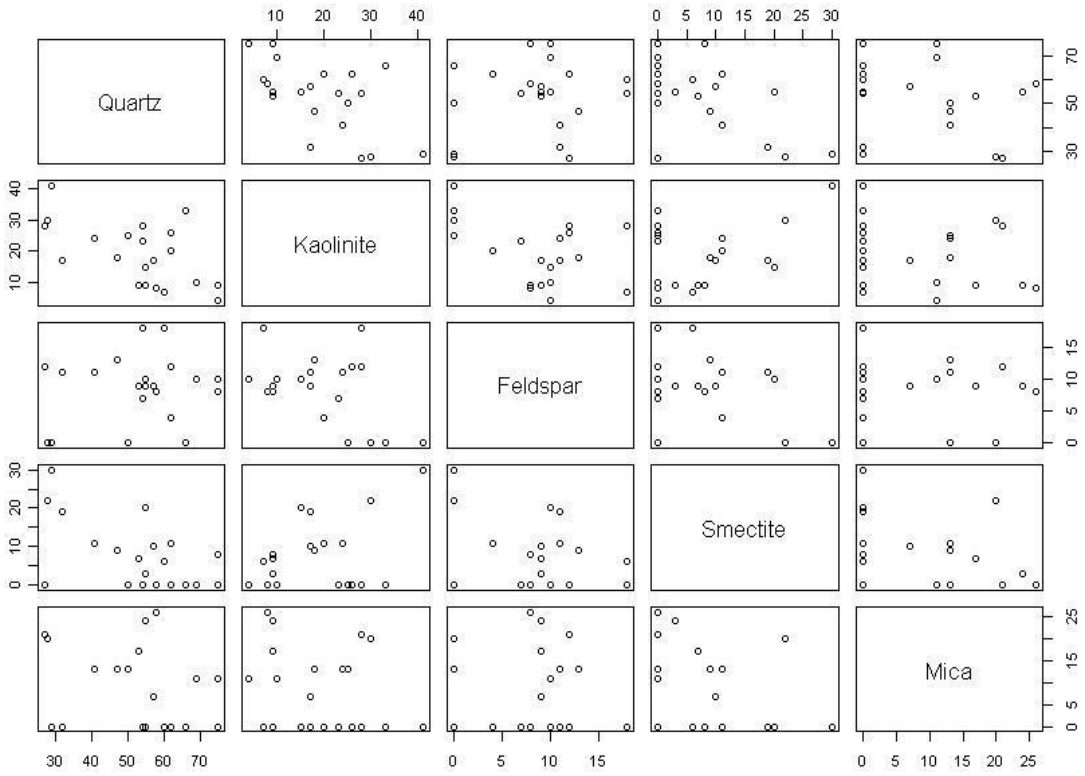

Figure 1. Pairwise plots indicating the potential bivariate relationships among clay minerals.

Thus, drawing horizontal lines along the mineral boundaries could help uncover the co-existence of minerals and their corresponding proportions. For instance, we should be able to evaluate the probability of the co-existence of quartz and kaolinite given that they are both in soil sample 3. Across the samples the overwhelming quantities of quartz makes it dominant but with a consistent high probability of co-existence with kaolinite. Even in cases of low quartz content, kaolinite seems to rise with the drop in quartz made up for by a combination of the other three which is in line with the negative relationship uncovered earlier. Although it may appear that traces of mica had nothing to do with quartz, the nested nature of the interactions which made it impossible to draw lines of demarcation along mineral boundaries inevitably brought the two minerals into some kind of indirect relationship.

The association plot in Figure 4, based on the variable transformations and the parameter thresholds in Table 1, shows the cross-mineral associations across the three soil samples; the heavier the line the stronger the association. The first letter is one of Low, Medium or High whereas the second letter in the node code which corresponds to the first letter of one of the five minerals. Thus, HQ refers to high quantities of quartz and it can be seen that a high content of quartz (HQ) is associated with a medium content of feldspar (MF). Other strong relations are between HQ and low mica content (LM) and between LS and MF. Quite interestingly, low traces of smectite (LS) appear to be associated with both low (LM) and high (HM) contents as well as with HQ and MF.

Decision tree modeling technique

Classification trees were applied to establish the mineral cross-dependencies via domain-partitioning of $C_k \in \Omega$ using the discretised version of data in Table 1. The intuition is that if the discretised variable is, say, mica, the prediction results will, effectively, tell us which of the four minerals cause low, medium or high mica contents in the sampled soils. Thus, our final analyses seek to establish the inter-dependence between the clay soil minerals in the domain $C_k \in \Omega$ using the decision tree prediction technique. The decision tree model in Figure 5 was generated by Clementine's decision tree routine. The discretised quartz variable was set as the target variable with the Gini as the measure of purity (impurity). The minimum number of records in each branch set to 4% of the total while the number in each child node set to half that value. Prior probabilities were generated in accordance with the training data and were adjusted using misclassification costs. The model, which yielded 100% accuracy, shows that the first important mineral to be used in identifying whether there is high, medium or

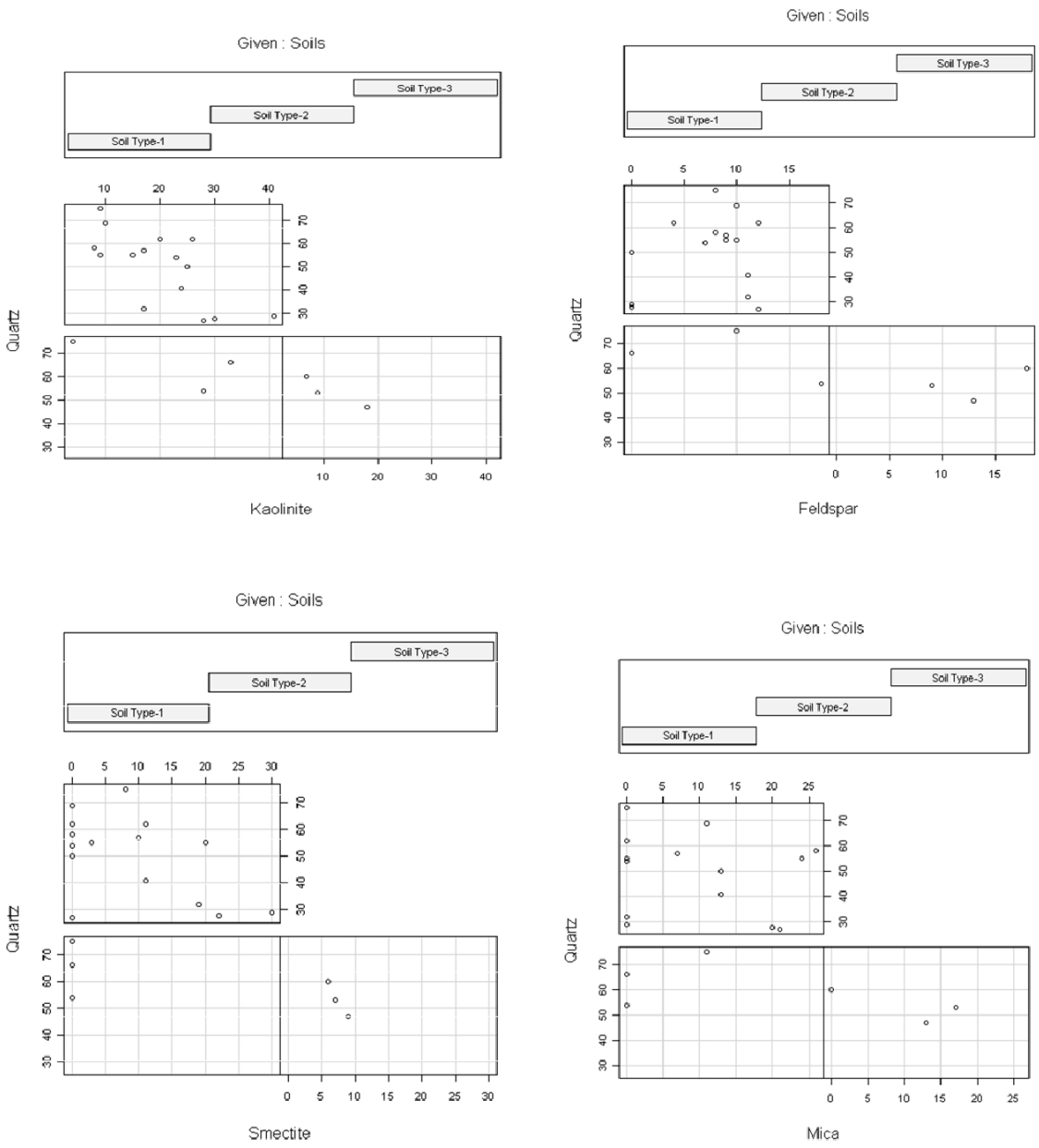

Figure 2. Bivariate plots of quartz compared to kaolinite, feldspar, smectite and mica in clay fraction of soils.

low quartz content in a soil sample is smectite with an apparently excessively high misclassification error at the initial stage. Down the line the model relies heavily on both the trace (mica) and the secondary major mineral (kaolinite); due to the secondary dominance of the latter. It is worth noting that to attain 100% accuracy the tree model was over -trained and so, in consideration of the issues of model complexity, accuracy and reliability, we set kaolinite to target under exactly the same settings and the model yielded an accuracy of 95.24% with quartz dominating the splits.

The foregoing illustrations were based on high-low relationships and so our next step is to look at the opposite relationship in which case we set mica and smectite as targets with the remaining input variables. The roles of quartz and smectite in predicting mica classes with 100% accuracy not only imply a relationship between these two minerals and mica, but also emphasized the association between the two predictors. Similarly, targeting smectite with the remaining four variables yielded a very high accuracy but with splits dominated by kaolinite and quartz. Thus, both quartz and kaolinite were eliminated from the model after which targeting smectite with mica and feldspar as predictors

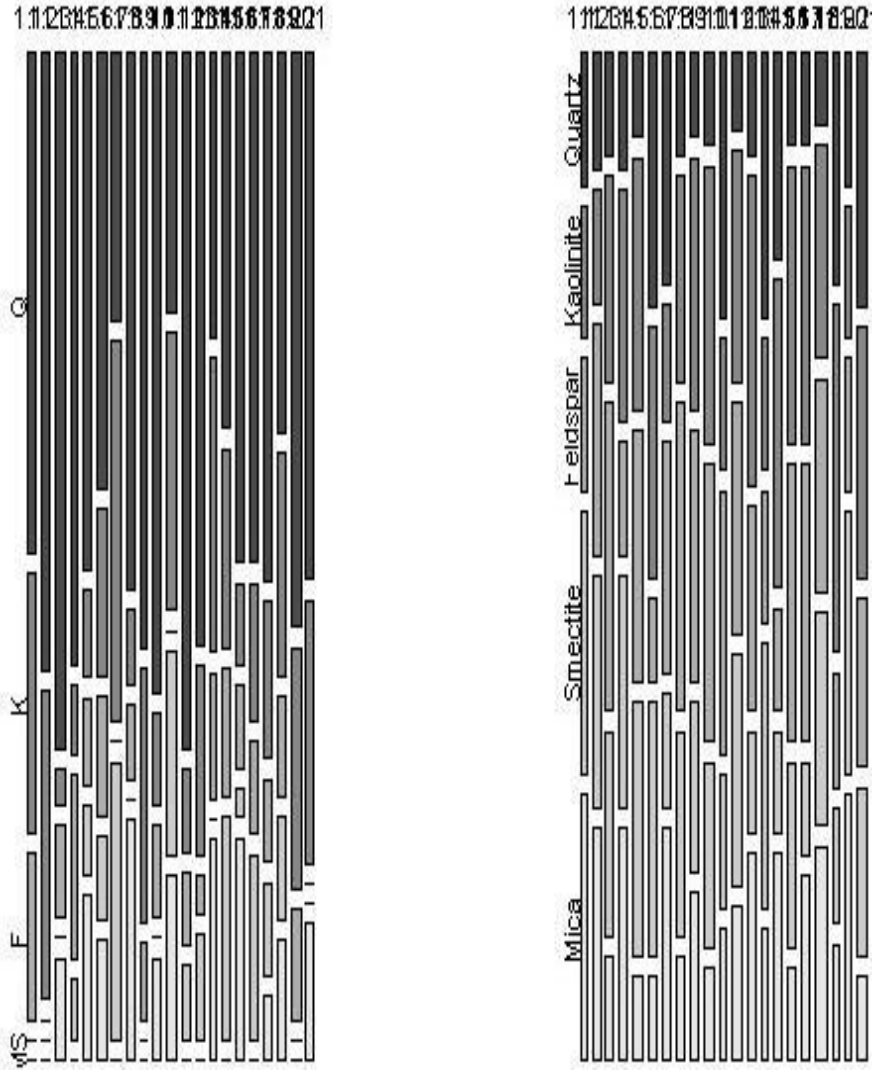

Figure 2. Graphical presentation of the mosaic plots for discretised and continuous minerals concentrations of soil samples.

yielded an accuracy of 85.71% and showed a stronger association between smectite and feldspar despite mica being the first most important root node splitting variable. Similarly, targeting mica with feldspar and smectite yielded a predictive accuracy of 80.95% and exhibited a strong relationship between mica and feldspar.

Agricultural implications of minerals in clay fraction of soils

Minerals identification and quantification revealed varying results with respect to the type and amount of minerals present in clay fraction of soil. The high percentage of kaolinite in the clay fraction of soil types 1and 3 may be due to weathering of feldspar and mica altering to kaolinite. The low contents of kaolinite in the clay fraction of soil type 2 results from the lower quantities of feldspar and mica both in the country rocks and soils of the study area. The dominance of kaolinite over smectite may reveal the strong dependence of soil formation on the parent material. The existence of kaolinite as the dominant secondary mineral may have implications with regards to the agricultural potential of the land mainly because of its low CEC.

Soils dominated by smectite require minimum constant fertilizer input compared to kaolinite dominated soils, because of the ability to hold nutrients in the soil. Their water holding capacity is very efficient compared to kaolinitic soils due to their ability to expand. However, their strong plasticity and stickiness when wet may create mechanical problems during cultivation, making the soils difficult to work with. The presence of quartz relates well with resistance to weathering due to its inert nature.

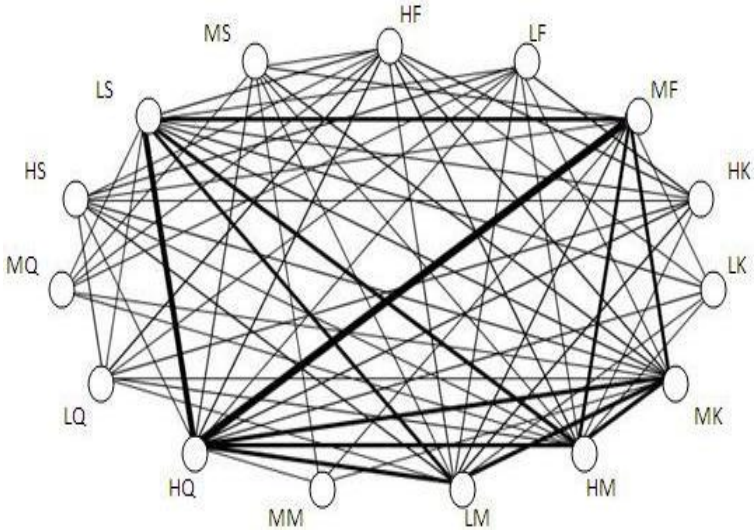

Figure 4. Cross-mineral associations across clay fraction of soil samples.

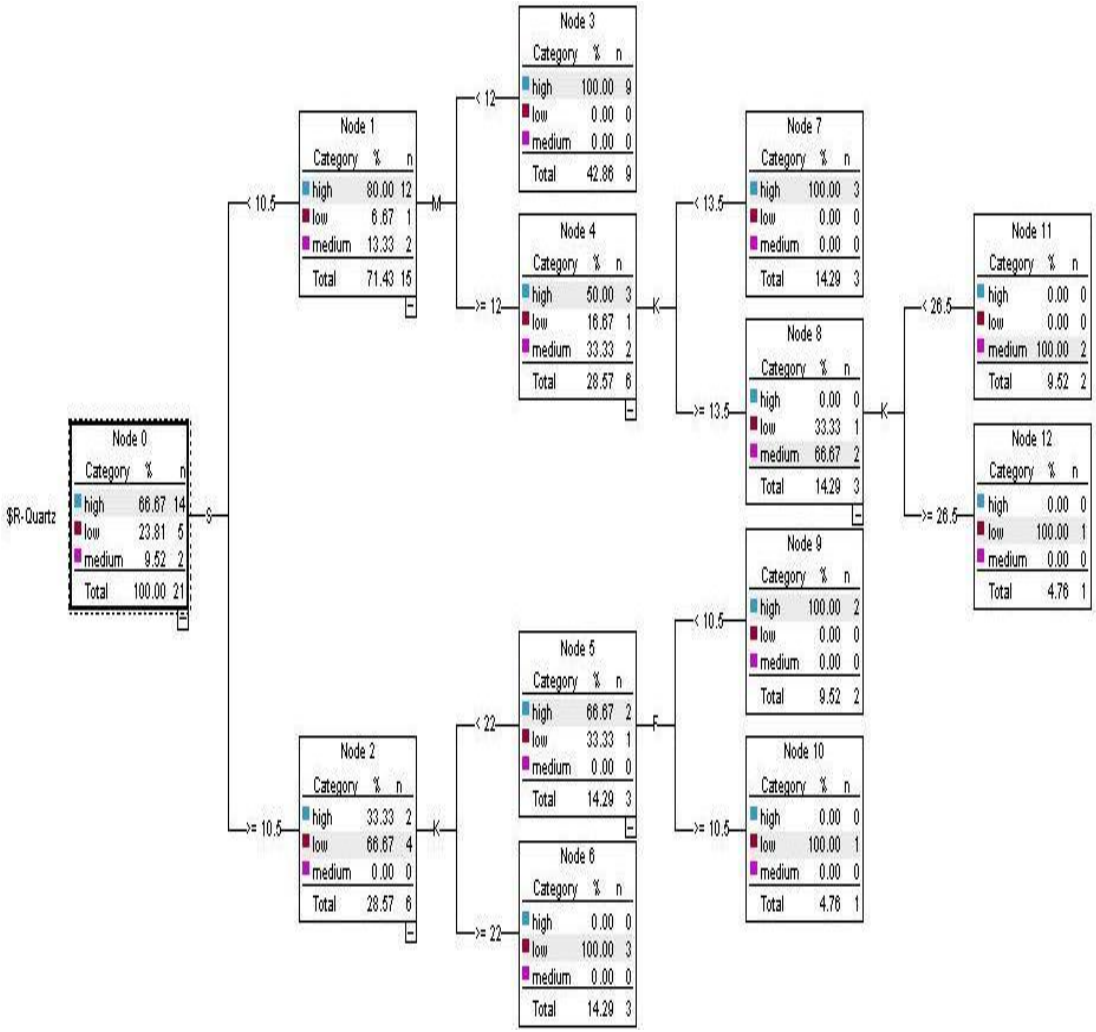

Figure 5. Graphical decision tree model for predicting quartz using kaolinite, smectite, mica and feldspar variables (note: S = smectite, M = mica, K = kaolinite, F = feldspar).

Quartz present in soil is generally considered chemically inactive. Soils dominated by quartz are usually nonplastic and enhance drainage. Quartz may be useful in clayey areas where drainage is a problem. Hence, agricultural management must take into consideration minerals types and their associations in the clay fraction of soils if good yields are envisaged.

Conclusion

This study applied a combination of three data modeling non-parametric methods (EDA, GDV and decision tree modeling) on dataset of minerals qualitative and quantitative compositions of clay fraction of soils from the Capricorn district in the Limpopo Province, South Africa. Kaolinite and smectite were prevailing secondary minerals in the soils of the study area. These minerals are believed to have originated from their parent materials of which granite is the dominant rock type in the area. Mica and feldspar were also present, possibly because of incomplete weathering.

Searching for minerals patterns in clay fraction of soil using the studied methods may still be associated with a number of issues. It is evident from minerals identification analyses, and GDV that minerals in clay fraction of soils may co-exist with varying degrees of complexity. Although minerals in clay fraction of soils have been researched in South Africa, representation in terms of site specific research and their implication to agricultural practices is still lacking. The paper has demonstrated how minerals composition and modeling processes can be combined to yield informative results on potential soil contents and hence its applications particularly to agriculture. The scientific community should therefore focus on developing and enhancing soil mineral-specific algorithms and approaches to data collection, storage, analysis and dissemination. One of the main challenges the African continent faces is the incoherence of its minerals data repositories. Constructive suggestions on how to deal with this problem have been advanced by Ekosse and Mwitondi (2009) and Mwitondi (2009).

ACKNOWLEDGEMENT

The XRD analyses were carried out at the Agricultural Research Council, Pretoria, South Africa.

REFERENCES

Bird MI, Chivas AR (1988). Stable isotope evidence for low temperature kaolinite weathering and post formational hydrogene-isotope exchange in Permian kaolinites. Chemical Geology (Isotope Geoscience Section), 72: 249-265.

Botha GA (1992). The geology and palaeopedology of late Quaternary colluvial sediments in northern Natal, South Africa. PhD thesis. University of Natal. Unpublished.

Breiman I, Friedman J, Stone C, Olshen R (1984). Classification and Regression Trees; Chapman and Hall.

Bühmann C, Beukes DJ, Turner DP (2002). Soils from the Lusikisiki District, Eastern Cape Province: I. Clay mineral associations and their agricultural significance. Report. ARC-Institude for Soil, Water and Climate, Pretoria.

Bühmann C, Escott BJ, Hughes JC (2004). Soil mineralogy research in South Africa, 1978 to 2002 – A review. Plant and Soil, 21: 316-329.

Bühmann C, Nell JP (1999). Clay mineral associations in South African soils formed under a Mediterranean-type climate. 6th International meeting on soils with a Mediterranean-type of climate, July 1999, Barcelona, Spain. Extended Abstracts, pp. 700-702.

Carter MR, Gregorich EG (2007). Soil Sampling and Methods of Analysis; CRC.

Cleveland WS (1993). Visualizing Data. New Jersey: Summit Press.

Ekosse G, Fouche PS (2006). Environmental association of clay minerals with potassium in soils close to an abandoned manganese mine, A multivariate and GIS analytical approach. Int. J. Environ. Stud., 63(5): 617- 632

Ekosse G, Mwitondi K (2009). Multiple data clustering algorithms applied in search of patterns of clay minerals in soils close to an abandoned manganese oxide mine. Appl. Clay Sci., 46(1): 1-6.

Friendly M (1994). Mosaic displays for multi-way contingency tables. J. Am. Stat. Assoc., 89: 190-200.

Gaspe A, Messer P, Young P (1994). Selection and preparation of claybodies for stove manufacture. Clay testing. A manual on clay/non clay ratio measurement technique, p. 10.

International Centre for Diffraction Data (2001). International Centre for Diffraction Data. Mineral Powder Diffraction File Databook, p. 942.

Jackson ML (1979). Soil chemical analysis – Advanced Course, 2nd Edn, 11th Printing, Published by the author, Madison, WI, USA, p. 895.

Mardia K, Kent J, Bibby J (1979). Multivariate Analysis. Academic Press.

Mwitondi K (2003). Robust methods in data mining; PhD thesis, Leeds University Press.

Mwitondi KS (2009). Tracking the potential, development, and impact of information and communication technologies in sub-Saharan Africa; In: Science, Technology, and Innovation for Socio-economic Development: Success Stories from Africa; International Council for Science (ICSU).

Mwitondi K, Taylor C, Kent J (2002). Using boosting in classification; Proceedings of the Leeds Annual Statistical Research Conference. Leeds University Press, pp. 125-128.

Stern R, Ben-hur M, Shainberg I (1991). Clay mineralogy effect on rain infiltration, seal formation and soil losses. Soil Sci., 152: 455-462.

Van der Merwe G. ME, Laker MC, Bühmann C (2002). Clay mineral associations in melanic soils of South Africa. Aust. J. Soil Res., 40: 115-126.