GEOLOGIE EN MIJNBOUW VOLUME 49(4), p. 305-310 1970

AN APPLICATION OF FACTOR ANALYSIS TO THE INTERPRETATION OF THE GENESIS OF MAGNETITE IN THE SMALLWOOD MINE, LABRADOR

ERWIN L. ZODROW[1]

ABSTRACT

The Smallwood Mine at Labrador City, Newfoundland, contains two recoverable iron ore minerals: specularite and magnetite.

This paper inquires specifically into the genesis of magnetite using a model that statistically relates the measured variation of four variables to ideal causes that were responsible for these variations.

Through the interpretation of the results of factor analysis, a theory of magnetite genesis is proposed which tentatively indentifies the Grenville orogeny as the factor largely responsible for the formation of magnetite.

However, it is proposed to enlarge the factor model to include additional chemical variables which may reduce the unexplained variance in this system of variables and more clearly specify the common factors.

RÉSUMÉ

La mine "Smallwood" à "Labrador City", Terre-Neuve, contient deux minéraux de fer recouvrables: "specularite" et "magnetite". Cette investigation s'enquérit spécifiquement de la genèse de "magnetite" en employant un modèle qui rapporte statistiquement la variation comptée de quatres variables à des causes idéales responsables de ces variations.

Au moyen d'une interprétation des résultats de "factor analysis", une théorie de la genèse de "magnetite" est proposée qui tentativement indique la "Grenville orogeny" comme l'agent en grande partie responsable de la formation de "magnetite".

Cependant, il est proposé d'élargir le "factor model" pour comprendre des variables chimiques additionnelles qui pourraient réduire le désaccord inexpliqué dans ce système de variables et qui aussi pourraient spécifier plus clairment les facteurs communs.

[1]) Geologist, The Iron Ore Company of Canada at Labrador City, Newfoundland and Labrador, Canada.

INTRODUCTION

General

In metamorphic petrology, the physical and chemical processes that generate end products (rocks and minerals) from known parents are generally well understood. However, given the end products, a backward extrapolation to the source (parent rocks or minerals in specific geologic environments), considering the transformation agents and geological influences, is a problem that involves a high degree of uncertainty.

Using the generalized sequence of geological events starting with the parent rock, and then proceeding via metamorphism to the present-day rocks and minerals, the following model may be applied to interpret the genesis of rocks and minerals, Table 1.

TABLE 1

General Genetic Model for Metamorphic Rocks and Minerals

Source	Process	End Product
A_{ij}	T_{ij}	B_{ij}
(n x p matrix)	(transformation matrix)	(m x p matrix)

Griffiths (1966).

Under the assumption that all matrices are square, and additionally, that T_{ij} has an inverse, given any two matrices a unique solution is feasible; given only one matrix no unique solution is possible. The source is represented as a matrix of observation with p properties and n specimens; the end product matrix of observation consists of p properties and m specimens.

Laboratory experiments in geochemistry endeavor to determine the transformation matrix T_{ij} which operates on A_{ij} and transforms it into B_{ij}. Thus, knowing a given metamorphic rock (B_{ij}) and the geochemical transformation agents (T_{ij}), we can predict the parent rock. This illustrates the necessity of knowing two matrices to predict the third (Griffiths, 1966 p. 655-656).

To relate this generalized model to the genesis of magnetite in the Smallwood Mine, Table 1 is modified as follows:

source =	geosynclinal sedimentation of primary iron-rich sediments,
process =	postdepositional geological influences on these iron-rich sediments,
end product =	magnetite as in the present-day metasediments.

The specific question raised is "can we relate magnetite genesis in the Smallwood Mine to its source, process or both?" In answering this question factor analysis is employed which, ideally, relates the measured properties of the variables to natural causes or factors which were responsible for the variation in the properties.

The Variables in this Analysis

The total number of input data is 5,582 ten foot samples from 208 diamond drill holes on the Smallwood Mine yielding the following measurement variables:

1) total soluble iron percent by weight (includes magnetic iron) (Fe%),
2) Magnetite mineral by weight in percent (Mag.%),
3) a general hardness factor of the ore (GF 1), and
4) amount of concentrate that may be expected per long ton of crude ore (WTY).

FACTOR ANALYSIS

General

The mathematical structure for factor analysis was developed mainly for use in psychometric research; for the mathematical aspects the reader is referred to Thurstone (1945), Harman (1965), Imbrie (1963), Cattel (1952; 1965), and Morrison (1967).Recently, however, this technique has been applied to geological research both in the behavioral characteristics (variation in properties) and in classification of geological variables and samples, respectively. For example, Griffiths applied factor analysis to the study of potential oil reservoir sandstones (1963), and used it to interpret the petrogenesis of detrital sediments (1966); Harris (1966) employed it for the reduction of variables in regression analysis for quantative studies in mineral exploration, and Zodrow and Harris (1967) used it as a basis to improve an ore grading model.

The value of factor analysis lies not only in the possibility of it leading to identification of theoretical variables (= causes) that influenced the behavioral properties of measured variables, but also in the quantification of explained variance in the system of variables. Furthermore, the interpretation of factors often shows the investigators where to concentrate efforts by searching for additional meaningful variables in order to decrease the unexplained variance and better define the factors.

The Factor Pattern and the R-mode Model

One of the basic assumptions in factor analysis is linearity of the specification equations that define the effect (rows), and the factor model is formulated as follows:

$$X_1 = a_{11}F_1 + a_{12}F_2 + a_{13}F_3 + \ldots + a_{1m}F_m + a_{1,m+1}U_1.$$
$$X_2 = a_{21}F_1 + a_{22}F_2 + a_{23}F_3 + \ldots + a_{2m}F_m + a_{2,m+2}U_2.$$
$$\longleftarrow$$
$$+ \ldots +$$
$$X_n = a_{nm}F_1 + a_{nm}F_2 + a_{nm}F_3 + \ldots + a_{nm}F_m + a_{n,m+n}U_n.$$

where	X_n	is the observed variable or the effect, or test,
	F_m	is a factor, i.e. a cause that contributes to the effect,
	a_{mn}	is a coefficient (= factor loading) indicating the relative importance of F_m on X_n, and
	$a_{n,m+n}U_n$	is the nth unique factor and its loading (Harman, 1965, p. 16-17).

The arrow indicates the flow of causes contributing to the effect; that is to say, X_n (i = 1, 2, ... n) is considered to be affected by proportional parts of causes F_m (j = 1, 2, ... m). In this type of factor analysis, the unique factors are always assumed to be uncorrelated among themselves and with the common factors; however, the common factors may be either correlated or uncorrelated (H a r m a n, 1965, p. 16).

The R-mode factor analysis is used in this paper which examines the relationships among variables. The steps leading to a complete factor solution are as follows:

1) *square symmetric correlation matrix*, the data matrix is converted into this p x p matrix which expresses the relationships between single pairs of variables. This matrix' main function is to reduce all means to zero and variances to unity $(\bar{X} = 0, \sigma^2 = 1)$ so that the data are on a uniform scale. This matrix constitutes the basic input data for factor analysis.
2) *matrix of factor loadings*, the square symmetric correlation matrix is factored, i.e. minimum numbers of linearly independent vectors, which represent the square symmetric correlation matrix, are determined, and
3) *rotated factor matrix*, the factor matrix is rotated, maintaining orthogonal co-ordinate axes, under the varimax criterion (K a i s e r, 1956). It is this matrix that is used for subsequent analysis and interpretation.

Communalities

Communalities (diagonal elements in Table 2) are measures of the variance in a variable that is shared with the other variables. While in principal component analysis the communalities are taken to be unity, in factor analysis they must be estimated which can be a vexing problem. The total variance of one variable equals

$$1 = h_j^2 + a_j^2$$

where h_j^2 is the communality, and
a_j^2 is the uniqueness (H a r m a n, 1965, p. 14-15)

In the present work, following H a r m a n (1965, p. 89-90) the square of the net coefficients of multiple regression (R^2) of the ith variable on all the other variables is used as observed communalities. Then by (a) inserting the R^2's into the diagonal of the square matrix of correlation and factoring it, an estimate of communalities is obtained by summing the squared factor loadings across columns; (b) placing the communalities obtained from (a) into the diagonal of the *square matrix of correlation* and factoring again, a new estimate of communalities is obtained. This process was repeated until the numerical change in the communalities, from the previous factoring, became insignificant. At this point, the communalities can be regarded as best estimates, i.e. convergence of the communalities to their true values, Table 2; at the same time, the residual square matrix of correlation also reached an absolute minimum or tending towards a zero matrix (Z o d r o w, 1967a, p. 61-62, 100).

TABLE 2

Square Symmetric Correlation Matrix, lower half, with Estimated Communalities

	Fe%	Mag%	GFl	WTY
Fe%	(0.688)			
Mag%	−0.056	(0.515)		
GFl	−0.192	0.511	(0.504)	
WTY	0.762	0.114	−0.250	(0.785)

Matrix of Factor Loadings (The Factor Matrix)

The factoring of the symmetric correlation matrix (resulting in Table 3) was effected by the principal factor solution (H o t e l l i n g, 1933) which assures that (a) the most important factor is extracted first, i.e. largest characteristic root (= eigenvalues) which can be regarded as explained variance, (b) the ensuing factors in numerical order are always successively smaller, and (c) the extracted factors are mutually orthogonal (= linearly independent).

TABLE 3

Matrix of Factor Loadings

Test	Factor 1	Factor 2
Fe%	0.81688	−0.15157
Mag.%	−0.15051	−0.72191
GFl	−0.40986	−0.59557
WTY	0.86462	−0.26482
Eigenvalues: λ_j	1.60516	0.96896
Variance accounted for cumulatively:	40.12 per cent	64.34 per cent

where the eigenvalues can be obtained by summing the squared factor loadings across rows:

$$\lambda_j = \sum_{j=1}^{\Delta} = a_{ij}^2$$

and the variance accounted for in each test equals the sum of the squared factor loading across columns:

$$\mathrm{var}_i = \sum_{i=1}^{\Delta} a_{ij}^2$$

Table 3 shows that only two factors exist in this four-variable observation matrix, i.e. estimated matrix of rank two. During the process of factoring the last two factors degenerated, i.e. the eigenvalues became negative (Hildebrand, 1965, p. 31-34).

> "This matrix (Table 3) represents one of the infinite number of possible causal schemes which, mathematically can explain the correlation among observed variables . . . algebraically these factors are justifiable; geologically, they are without significance. The initial factor matrix is, however, a necessary algebraic step to a meaningful geological solution." (Imbrie, 1963, p. 14).

The Rotated Factor Matrix

The main objective in this analysis is to obtain the rotated factor matrix which facilities the interpretation of causes (Thurstone, 1945; Imbrie, 1963; Harman, 1965; and Griffiths, 1966); for this matrix most clearly identifies the hypothetical variables or causes responsible for the variation in the properties of the measurement variables.

Rotation of the factor matrix, Table 3 resulting in Table 4, accomplished under the varimax criterion proposed by Kaiser (1956), transforms it orthogonally treating each factor as a vector in m-space. Moreover, the varimax method places more emphasis on the factors (columns) in an endeavour to meet the requirements for "simple structure principle" according to Thurstone (1945, p. 335) by (a) minimizing low and maximizing high factor loadings, and by (b) identifying those vectors closest to the reference axes (Kaiser, 1958).

The length of each vector in the factor space is an expression of its communality, and the scalar product of two vectors is the value of the correlation coefficient between these vectors:

where $r_{jk} = h_j\, h_k \cos \phi_{jk}$

r_{jk}	is the correlation coefficient between two variables X_j and X_k
h_j	is the standard deviation of X_j,
h_k	is the standard deviation of X_k and
$\cos \phi$	is the angle between the vectors.

TABLE 4

The Rotated Matrix (= Varimax Matrix)

Test	Factor 1	Factor 2	Variance of Variables Explained by the Two Factor Model
Fe%	0.82375	0.10656	68.99%
Mag.%	0.07846	−0.73325	54.38%
GFl	−0.20715	−0.69266	52.27%
WTY	0.90417	0.01350[1]	81.77%

DISCUSSION OF RESULTS

Interpretation of Factors

The interpretation of the rotated factor matrix now becomes subjective depending on the investigator's ability to relate the isolated and explained variance of the variables to the most probable causes which were responsible for these variations. Because of this, factor analysis does not lead to unique solutions.

To assign physical meanings to the factors, one must consider the loadings of the rotated factor matrix vertically, remembering that causes flow horizontally to the X_n effects. Thus there are high loadings under factor 1 on variables (= tests) 1 and 4, and relatively high loadings under factor 2 on variables 2 and 3.

To identify factor 1, one should search for a cause that accounts for 67.9 per cent (= factor 1, loading 1 squared) of the variance in Fe%, and this cause must also explain 0.6 per cent of the magnetite variance in this system. On the other hand, the absence of this cause increases the general hardness effect (GF1). This cause must also explain 82 per cent of the WTY effect. Factor 1 is interpreted as

original depositional environment
of the geosyncline = D

or the sum total of the geological and geochemical

---

[1] We may argue that this loading is statistically insignificant; therefore, WTY is influenced only by factor 1 and a unique factor.

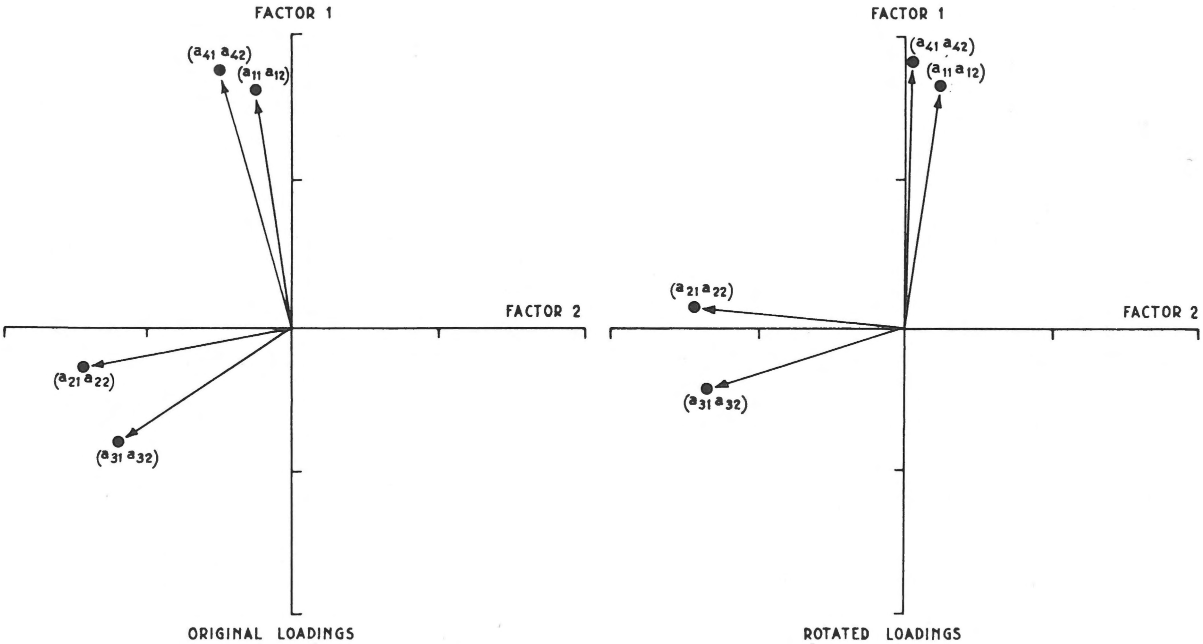

Fig. 1
Graphical Interpretation of Original and Rotated Factor Loadings. Source: (Zodrow, 1967a, p. 64).

processes contributing to the sedimentation of the primary iron-rich sediments.

Examining the factor loadings of factor 2, only 1 per cent of the variance of the Fe% effect is explained by this factor, but in terms of cause-effect relation, the absence of this cause increases the magnetite and the general hardness effects. Therefore, factor 2 is interpreted as

diagenetic changes

which encompasses all postdepositional changes in the sediments short of metamorphism. Since the interpretation of the factors must proceed in such a manner as to keep the causes statistically independent of one another, facies changes are ruled out under factor 2. Because oxidation-reduction potentials and pH considerations are dependent on the original environment of deposition in the geosyncline.

However, an alternative approach to interpreting factor 2 is proposed by changing the signs of the loadings under this factor. The adequacy of this solution is not altered; however, with a change of signs we must also change the meaning of this cause to its opposite (Harman, 1965, p. 110), e.g. erosion vs. sedimentation, or acidity vs. basicity Hence, after all the signs are changed under factor 2 it is interpreted as

metamorphic influences = M

Geologically, this can be justified because diagenetic changes and metamorphism are two opposite extremes, ideally mutually exclusive events under the collective concept of postdepositional changes in rocks. Table 5 summarizes the causality scheme of Mag.% and Fe%.

TABLE 5

Interpretive Genetic Model for Magnetite and Fe per cent

Magnetite = 0.078D + 0.733M + Uniqueness Factor.

Fe = 0.824D − 0.106M + Uniqueness Factor.

Conclusions

This two factor model presents four causality schemes, two shown in Table 5, and these may be interpreted as follows:

1. Magnetite is independent of the cause symbolized as D; if D is correctly assigned to the original depositional environment of the geosyncline then

magnetite is independent of the depositional environment.

2. The formation of magnetite is largely dependent on the cause symbolized as M, and if this is correctly assigned to metamorphic influence then it may be tentatively identified as accompanying the Grenville deformation.
3. The variation in magnetite, about *54 per cent*, is largely dependent on factor 2 and is practically independent of factor 1 (o.6 per cent explained).
4. The variation in per cent Fe can be attributed to factor 1; if it is correctly identified as the original conditions in the depositional environment then variations in iron is related to processes in terms of pH and oxidation-reduction potentials.

Furthermore, from inspection of the rotated factor matrix, we may also select *independent* variables (= predictors) for parsimonious regression analysis (H a r r i s, 1966, p. GG-27). Thus, to predict Fe% we would choose WTY as independent variable since it has the only high loading on the same factor. Similarly, Fe% cannot be predicted using Mag.% as independent variables, but the general hardness factor (GF1) is partially dependent, in a functional sense, on magnetite in the rock. This last relationship is corroborated in substance by the milling process carried out at Labrador City. That is to say, with an increase of magnetite in the crude ore the hourly grinding rate decreases generally.

The outcome of this analysis suggests that (a) the present model be enlarged to include additional variables such as MnO_2, SiO_2, CaO and MgO, and that (b) the magnetic iron be separated from the total soluble iron (Fe%) thus introducing Fe% (specularite) and Fe% (magnetite) as variables. This may suffice to add substiantially to the explained variance in Table 4.

In summary, assuming that the most probable causes have been correctly identified, it follows that magnetite is largely a secondary mineral in this deposit in the sense that it is metamorphic and formed after termination of geosynclinal sedimentation, i.e. its formation can be tentatively correlated with the Grenville deformation of about 1,100 million years ago (Z o d r o w, 1967b.)

REFERENCES

Cattell, R.B. (1952) – Factor Analysis: New York, Harper Brothers, 462 p.

——(1965) – Factor analysis: An introduction to essentials. I. Factor analysis: its role in research. II.: Biometrics, vol. 21, p. 190-215, 405-435.

Griffiths, J.C. (1963) – Statistical approach to the study of potential oil reservoir sandstones: Stanford University Publication, Geol. Sci., vol. 9, p. 637-668.

——(1966) – A genetic model for the interpretative petrology of detrital sediments: Jour. Geol., vol. 74, No. 5, p. 653-672.

Harman, H.H. (1965) – Modern factor analysis: Chicago, University of Chicago Press, 471 p.

Harris, D.P. (1966) – Factor analysis a tool for quantitative studies in mineral exploration: The Pennsylvania State University, Mineral Ind. Experiment Stations, Special Publ. 2-65, p. GG-1 to GG-37.

Harris, D.P., and E.L. Zodrow (1969) – Magnetite trend surface analysis: a guide for open pit quality control at the Smallwood Mine, Newfoundland: in print, Quarterly of the Colorado School of Mines.

Hildebrand, F.B. (1965) – Methods of applied mathematics: sec.ed., Englewood Cliffs, N.J., Prentice-Hall, Inc. 362 p.

Holzinger, K.J., and H.H. Harman (1941) – Factor Analysis: Chicago, University of Chicago Press, 417 p.

Hotelling, H. (1936) – Simplified calculations of principal components, Psychometrika, vol. 1, p. 27-35.

——(1933) – Analysis of a complex of statistical variables into principal components: Jour. Educ. Psych., vol. 24, p. 417-441, 498-520.

Imbrie, J. (1963) – Factor and vector analysis programs for analyzing geologic data: Office of Naval Research Geography Branch, Tech. Rept. 6, 85 p.

Kaiser, H.F. (1959) – Computer Program for varimax rotation in factor analysis: Educ. and Psychol. Measurement, vol. 19, p. 413-420.

——(1958) – The varimax criterion for analytic rotation in factor analysis: Psych., vol. 23, p. 187-200.

——(1956) – The varimax method of factor analysis: Unpublished Ph. D. thesis, U. of California.

Morrison, D.F. (1967) – Multivariate statistical methods: Toronto, McGrawHill Book Co., 338 p.

Mount, W.D., and I.A. Lund (1963) – A simplified procedure for selecting predictors and deriving prediction equations: Jour. Geophys. Res., vol. 68, p. 3619-3625.

Thurstone, L.L. (1945) – Multiple-factor analysis, a development and expansion of the vectors of mind: Chicago, University Press of Chicago, 535 p.

Zodrow, E.L. (1967a) – Statistical and geological review of the grading model of the Smallwood iron ore mine, Labrador: Unpublished M.S. thesis, The Pennsylvania State University.

——(1967b) – Relationship of olivine gabbro and amphibolite schist at Carol West, Labrador geosyncline: in print, Geol. Rundsch.

Zodrow, E.L. and D.P. Harris (1967) – An ore grading model for the Smallwood Mine: Mining Engineering, vol. 19, p. 70-73.