Journal of Geochemical Exploration 150 (2015) 125–136

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/jgeoexp

Identification of geochemical factors in regression to mineralization endogenous variables using structural equation modeling

Jiangtao Liu [a,b], Qiuming Cheng [a,b,*], Jianguo Wang [a]

[a] Department of Earth and Space Science and Engineering, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada
[b] State Key Lab of Geological Processes and Mineral Resources, China University of Geosciences, Beijing and Wuhan, China

ARTICLE INFO

Article history:
Received 20 August 2013
Accepted 29 December 2014
Available online 4 January 2015

Keywords:
Structural equation modeling (SEM)
Causal relation
Variable correlation
Path analysis
Exploratory factor analysis (EFA)
Multiple linear regressions (MLR)

ABSTRACT

Structural equation modeling (SEM) is a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions. It can be considered as an extended statistical hypotheses testing with pre-defined structural model. However, when the SEM is applied as an exploratory technique in data mining, the traditional model depicts a fundamental limitation of incapacity of generating and refining structural model. A new SEM method is proposed in this paper, which combines the principles of cluster and regression analysis. Thus, the new mathematical model can not only generate factors to form model structure but also ensure the optimum relationship to the objective variables. The method is applied to a data set of lake sediment geochemical compositions for identifying gold mineralization associated factors in Southern Nova Scotia, Canada. A SEM model consisting of three measurement sub-models and one structural sub-model is created on the basis of the concentration values of 16 elements from 671 lake sediment samples. The calculated results show that the factors obtained by the new SEM model represent three geochemical factors that are associated with *As* and dominated by *Cu*, *Zn* and *W*, respectively.

1. Introduction

A structural equation model (SEM) is a class of multivariate statistical models that allow complex modeling of relational structures between independent and dependent variables. Although the SEM theory and methods are well developed, they are mainly scattered throughout several research fields, particularly education, psychology, sociology, and econometrics (Browne and Arminger, 1995; Muthen, 1984; Yuan and Bentler, 2000). The SEM theory has also appeared in mainstream statistical journals under the terms of mean and covariance structures and latent variable models (Bandeen-roche et al., 1997; Jöreskog, 1979; Lee and Shi, 2001; Sammel and Ryan, 1996; Yuan and Bentler, 1997). SEM can be interpreted as a combination of factor analysis (FA) and multiple linear regressions (MLR) (Ullman and Bentler, 2003), which can calculate the factor loadings on latent variables or factors and regression coefficients of latent variables with respect to dependent variables from a group of equations. The factor analysis in the SEM is referred as to the observed model while the path analysis or MLR to the structural model. Therefore, SEM is different from the ordinary principle component analysis or the exploratory factor analysis (EFA) that is commonly used in geochemical data processing. The latter determines the orthogonal components with ranking of variances. But these components are calculated on the basis of the interrelationships of variables involved and they are usually not associated with a particular objective variable of interest. The former determines the loadings of the pre-assigned factors according to their association with external dependent variables. SEM is also different from MLR, since the SEM involves the latent variables in structural model as independent variables rather than the original explanatory variables used in MLR.

There are two main approaches for parameter estimation in SEM (Anderson and Gerbing, 1988; Hair et al., 2006, 2011, 2012; Hendry, 1976; James and Singh, 1978; McDonald, 1977; Mehta and Swamy, 1978; Ramsey, 1978; Sargan, 1978; Zellner, 1978). One is Covariance based SEM (CB-SEM) and the other is Partial least squares SEM (PLS-SEM). PLS-SEM is one of the commonly used algorithms that was originally developed by Wold, (1966) and further developed by many others (e.g. Hair et al., 2006, 2011). In PLS path models, the explained variances of the endogenous latent variables are maximized by estimating partial model relationships through iterative ordinary least squares (OLS) regression. PLS-SEM is mainly applied in exploratory analysis other than confirmatory analysis (Hair et al., 2006, 2011, 2013). It was widely applied in various fields such as marketing, management and organizational research (Diamantopoulos and Winklhofer, 2001; Jarvis et al., 2003; MacKenzie et al., 2005). More discussion of these methods can be found in Wold, (1982, 1985) and Lohmöller, (1989). According to the authors knowledge SEM has never been applied in geochemical

* Corresponding author at: Department of Earth and Space Science and Engineering, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada.
E-mail address: qiuming@yorku.ca (Q. Cheng).

http://dx.doi.org/10.1016/j.gexplo.2014.12.021
0375-6742/

data processing for mineral exploration. One of the main reasons might be due to the fundamental limitation of the ordinary SEM models that require predetermined structural model and incapacity of generating and refining structural model. A new SEM method is proposed based on PLS-SEM in this paper, which combines the principles of cluster analysis and regression analysis. Thus, the new mathematical model can not only generate factors to form a structure model but also ensure the optimum relationship to the objective dependent variables. In this paper the mathematical model of the new SEM will be introduced and its applicability in geochemical data processing validated by a case study for identifying gold mineralization associated factors in Southern Nova Scotia, Canada. The case study uses concentration values of 16 elements from 671 lake sediment samples collected in the study area. For comparison purpose, all three methods: MLR, EFA and SEM were used for analyzing the same dataset. For implementation of MLR and the new SEM the element *As* was utilized as the dependent and objective variable and other 15 elements are interdependent variables. In order to interpret the geochemical associations of latent variables from *As* mineralization point of view, the loadings and regression coefficients on latent variables with respect to *As* were analyzed and compared.

2. Materials and methods

2.1. Geochemical dataset and geological background

The case study deals with the identification of geochemical factors in regression to gold mineralization endogenous variables in the southwestern Nova Scotia, Canada. The geological map of the study area is given in Fig. 1. The study area ($\approx$7780 km^2) is mainly underlain by Cambro-Ordovician greenschist to amphibolites grade metamorphosed sedimentary rocks and Devonian granitoid rocks. The South Mountain Batholith (SMB) is a complex of multi-phase granites covering nearly one-third of the entire study area (MacDonald et al., 1992; Reynolds et al., 1987). A number of *Au*, *U*, *W* and *Sn* deposits have been found in the area. About 20 *Au* mineral occurrences shown as dots in Fig. 1 are found in the sedimentary rocks. The geochemical lake sediment data include 671 samples with concentration values of 16 elements: *Ag*, *As*, *Au*, *Cu*, *F*, *Li*, *Nb*, *Pb*, *Rb*, *Sb*, *Sn*, *Th*, *Ti*, *W*, *Zn* and *Zr* (Rogers and Garrett, 1987). The sampling density was about 1 sample per 5 km^2. The concentration value of element *As* was determined by instrumental neutron activation analysis (INAA) with a detection limit of 1 ppm. Gold has the relatively

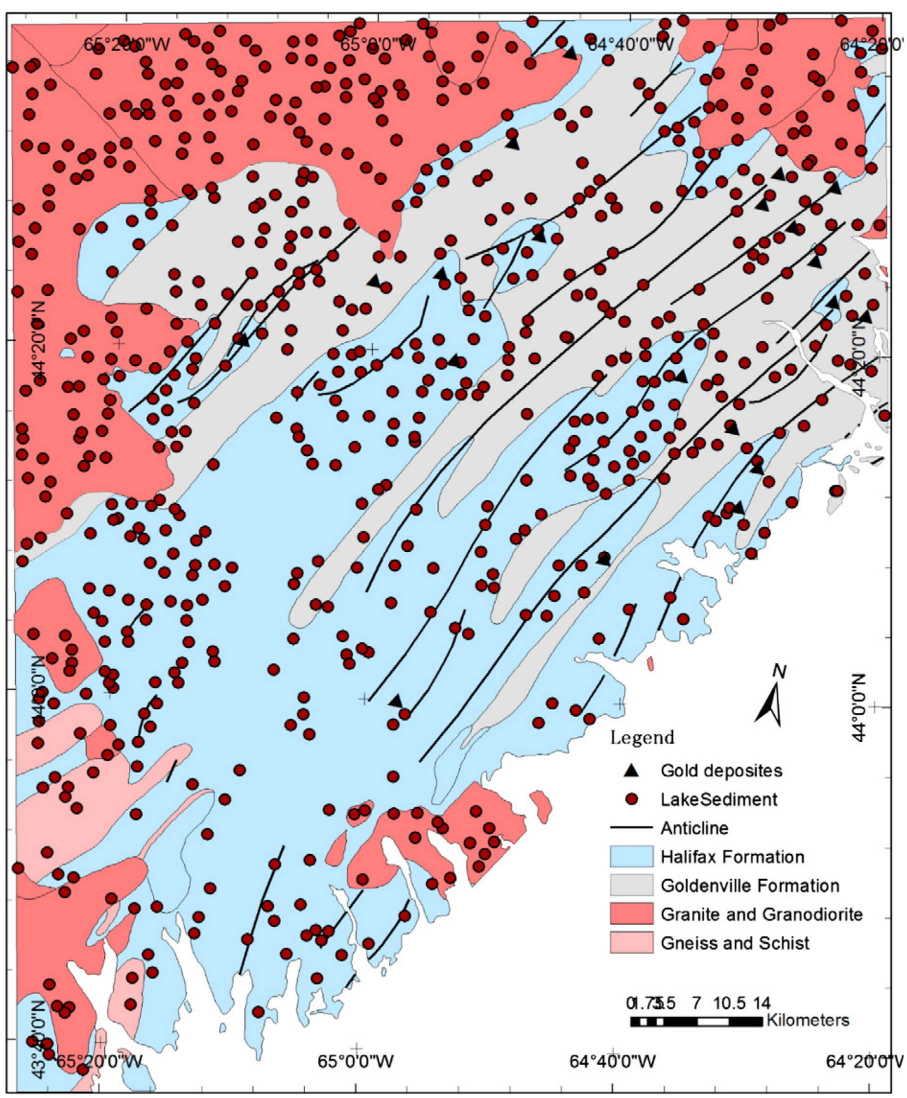

Fig. 1. Location and simplified geology of the study area. The known gold mineral deposits and mineral occurrences and lake sediment samples are superimposed on the map. Data from the Nova Scotia Department of Natural Resources 2006.