Journal of African Earth Sciences 129 (2017) 668–674

Contents lists available at ScienceDirect

Journal of African Earth Sciences

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

Multivariate analysis of ground water characteristics of Ajali sandstone formation: A case study of Udi and Nsukka LGAs of Enugu State of Nigeria

L.C. Orakwe, E.C. Chukwuma*

Department of Agricultural & Bioresources Engineering, Faculty of Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form
2 February 2017
Accepted 6 February 2017
Available online 7 February 2017

Keywords:
Borehole characteristics
Multivariate analysis
Cluster analysis
Principal component analysis
Ajali sandstone formation

ABSTRACT

Multivariate statistical techniques were applied for the evaluation and interpretation of borehole characteristics of the Ajali sandstone geological formation of Enugu state of Nigeria to determine the latent structure of the borehole characteristics and to classify 9 borehole parameters from 33 locations into borehole groups of similar characteristics. Two chemometric data mining techniques used were, Cluster Analysis (CA) and Principal Component Analysis (PCA). PCA identified the borehole parameters responsible for variation in the borehole characteristic of the study area. Out of the nine parameters examined, the PCA identified borehole depth, borehole casing, static water level and dynamic water level as the most significant parameters responsible for variation in borehole characteristics. Hierarchical Cluster Analysis also grouped the 33 borehole locations into three clusters. The CA grouping of the borehole parameters showed similar trend with PCA hence validating the grouping of variations in the borehole characteristics in the geological zone. The results of the study indicate that PCA and CA are useful in offering reliable classification of the borehole characteristic of the study area.

1. Introduction

Inadequate growth in food production and increasingly scarce water may pose serious constraints to future agricultural and economic development in Africa. With the increasing awareness of the importance of water to solve the food and energy crises of developing countries, much thought has been given in recent years to how to optimize the total benefits accruing from water development projects. In nearly all developing countries this resource has to be used more efficiently than at present if it is to provide potable water for rural and urban populations, to produce more food, to generate more electricity and to encourage other benefits to satisfy basic human needs to a much higher degree than at present (Biswas and Nakayama, 1989). Water resource management is a vital important resource whose availability is often taken for granted until there is a scarcity or significant flooding. To appropriately site and design water sources, the groundwater resources of an area need first to be to understand how water occurs in the ground (MacDonald et al., 2001). Groundwater is well suited to rural water supply in sub- Saharan Africa because (MacDonald and Davies, 2000). Lately there has been a greater interest in the exploitation of groundwater. However, many problems exist as a result of insufficient knowledge of the subsurface geophysical conditions in many parts of Nigeria.

Several research works with large complex data have been simplified and hidden data structure identified with multivariate statistical technique. PCA is a multivariate technique for transforming a set of related (correlated) variables into a set of unrelated (uncorrelated) variables that account for decreasing proportions of the variation of the original observations (Landau and Everitt, 2004). Shrestha and Kazama (2007) used multivariate statistical techniques for the evaluation of temporal/spatial variations and the interpretation of a large complex water quality data set of the Fuji river basin in Japan, generated during 8 years (1995–2002) period of monitoring of 12 water quality parameters at 13 different sites which generated as much as 14,976 observations. Wang et al. (2013) used multivariate statistical analysis such as CA and PCA/factor analysis to evaluate temporal/spatial variations in water quality and to identify latent sources of water pollution in the Songhua River Harbin region, China. The quality of harvested rainwater used for toilet flushing in a private house in the

* Corresponding author.
E-mail address: ecchukwuma@yahoo.com (E.C. Chukwuma).

http://dx.doi.org/10.1016/j.jafrearsci.2017.02.011
1464-343X/

southwest of France was assessed over a one-year period by Vialle et al. (2011). To elucidate factors affecting the rainwater composition, PCA and CA were applied to the complete data set of 50 observations. Numerous works on multivariate analysis on large water quality dataset exist. However, there is little or no work on multivariate statistical analysis of borehole characteristics dataset in Nigeria to the best of the author's knowledge. Knowledge on the subsurface geophysical conditions in many parts of Nigeria is needed for water resource management and planning. Hence the aim of this study is to determine the natural groupings of borehole characteristics in the study area, and to characterise associations present in the complete data set.

1.1. Description of the area

The study area (Udi and Nsukka Local Governing Authorities (LGAs) of Enugu State, Nigeria) is situated at the Southern Eastern part of Nigeria between Latitude 6° 20′ − 6° 50′N and Longitudes 7° 20′ − 7° 30′E (Fig). The general physiographic features within the study area comprise the Enugu and Udi Plateau escarpment. The Udi Plateau is a large physiographic unit extending through the area with highest points at about 500 m above sea level. It slopes westwards from Enugu escarpment to the lowlands of Niger and Imo rivers. The boreholes within the study area penetrate the Ajali Formation rocks and are mainly unconfined. The aquifer is a highly transmissive sandstone which underlies the Udi − Nsukka Plateau (where the water table is deep) and the lower grounds west of the Plateau where the water may be confined (Reyment, 1965). Borehole depths range from 50 m to 240 m (Fig. 1).

2. Materials and methods

2.1. Sampling stations

A total of 33 borehole locations were analysed in the study area. Borehole parameters measured or considered included borehole diameters (mm), borehole depths (m), borehole casing (m), screen length (m), static water level (m), dynamic water level (m), drawdown (m), yield (l/s), and specific capacity (l/s/m). Theses hydraulic parameters were obtained by standard pumping test techniques using Jacob's straight line method (Khalil et al., 2011).

2.2. Statistical analysis

In this study, univariate and multivariate chemometric techniques were performed using SPSS version 17 statistical package (IBM trademark).

2.2.1. Principal component analysis

Often, one of the most useful features of PCA is that it can be used to construct an informative graphical display of multivariate data that may assist in understanding the structure of the data (Landau and Everitt, 2004). Principal Component Analysis (PCA) is one of the most applied approaches in the environmetrics to study data structures. It is aimed at finding and interpreting hidden complex and casually determined relationships between dataset features. This is accomplished by studying the data structure in a reduced dimension while retaining the maximum amount of variability present in the data. To do this, it is necessary to estimate the number of significant components present in the data. More precisely, a matrix of pair-wise correlations among parameters is decomposed into eigenvectors, which, in turn, are sorted in descending order of their corresponding eigenvalues (Vialle et al., 2011).

The principal component (PC) can be expressed as:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \ldots + a_{im}x_{mj} \quad (1)$$

where Z is the component score, a is the component loading, x the measured value of variable, i is the component number, j the sample number and m the total number of variables (Shrestha and Kazama, 2007).

2.2.2. Cluster analysis

Classification is an important component of virtually all scientific research. Cluster analysis is an unsupervised pattern recognition technique that defines the underlying behaviour or intrinsic structure of data sets, without making prior assumptions about the potential structure, in order to ultimately classify the data samples

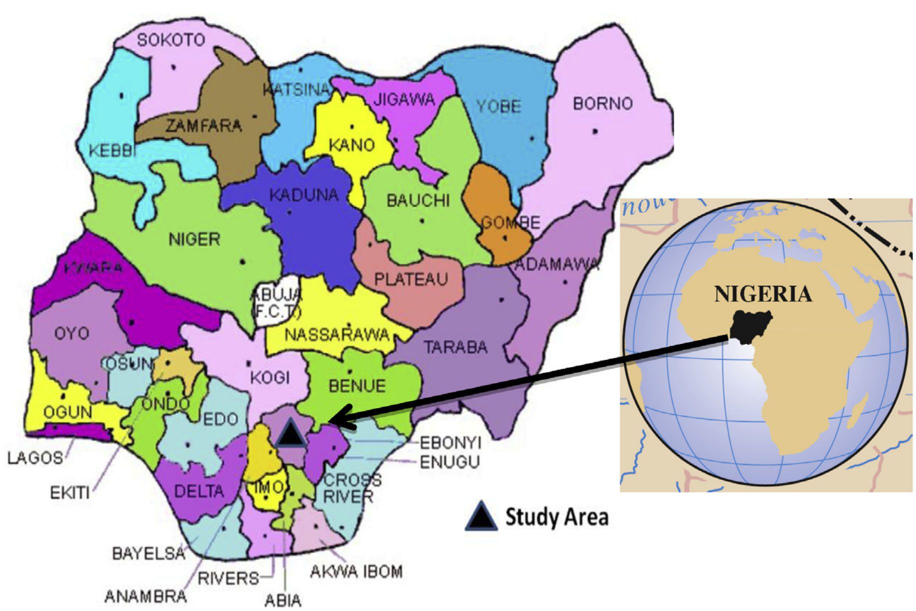

Fig. 1. Map of Africa and Nigeria showing the location of the study area.