Computers and Electronics in Agriculture 140 (2017) 409–421

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Comparison of artificial neural network and multivariate regression models for prediction of Azotobacteria population in soil under different land uses

Mitra Ebrahimi [a], Ali Akbar Safari Sinegani [a], Mohammad Reza Sarikhani [b,*], Seyed Abolghasem Mohammadi [c]

[a] Department of Soil Science, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran
[b] Department of Soil Science, Faculty of Agriculture, University of Tabriz, Iran
[c] Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Iran

ARTICLE INFO

Article history:
Received 18 January 2017
Received in revised form 13 March 2017
Accepted 21 June 2017
Available online 27 June 2017

Keywords:
Azotobacteria
Artificial neural network
Easily measurable characteristics
Multivariate regression
Soil properties

ABSTRACT

Azotobacteria are one of the most important and beneficial soil bacteria which their number and distribution are affected by physicochemical and biological properties of soil and land usage. The aim of this study was to evaluate the population of *Azotobacter* in soils with different land uses and relationship between population size and some physicochemical and biological properties of soil by using artificial neural network (ANN) and multivariate linear regression (MLR) methods. In total, 50 soil samples were collected from depth (0–25 cm) under different land uses located in East Azerbaijan, Ardabil and Gilan provinces, Iran. Population of *Azotobacter* was separately counted in Winogradsky and LG media by preparation of serial dilution and plate counts. In addition, soil texture, pH, electrical conductivity (EC), carbonate calcium equivalent (CCE), organic carbon (TOC), cold water extractable OC (CWEOC), hot water extractable OC (HWEOC), light fraction OC (LFOC), heavy fraction OC (HFOC), basal respiration (BR) and substrate induced respiration (SIR), the number of bacteria, fungi and actinomycete were measured in three replicates in each soil sample. To predict Azotobacteria population based on easily measurable characteristics of soil properties, MLR analysis and ANN model (feed-forward back propagation network) were used. In order to assess the models, root mean square error (RMSE) and R^2 were used. The R^2 and RMSE values for population of *Azotobacter* in Winogradsky medium obtained by ANN model with SIR, EC, CCE, sand and silt as entered variables were 0.76 and 0.36, respectively, and for population of *Azotobacter* in LG medium, were 0.45 and 0.50, respectively. Using MLR the R^2 value for population of *Azotobacter* in WG and LG media was 0.63 and 0.39, respectively. Results showed that ANN with eight neurons in hidden layer had better performance in predicting population of *Azotobacter* in WG than MLR.

1. Introduction

Azotobacteria are non-symbiotic, Gram negative, heterotrophic, aerobic and free living bacteria. These bacteria are capable of nitrogen fixation as well as produces plant growth stimulants such as hormones, vitamins, etc. (Montgomery et al., 2001). *Azotobacter* genus along with *Pseudomonas* family belonged to the γ-proteobacter and contains seven species comprising *Azotobacter chroococcum*, *A. vinelandii*, *A. beijerinckii*, *A. salinestri*, *A. paspali*, *A. nigricans* and *A. armeniacus* (Bescking, 1961). *Azotobacter* family cells are variable in terms of morphology, cells range from straight rods with rounded ends to more ellipsoidal or coccoid, depending on the culture medium and age, moreover some species such as *A. paspali* can be filamentous (Brenner et al., 2004). In some species larger vegetative cells to become spherical shape which is known as cyst, each cyst is produced from a single vegetative cell (Rubio et al., 2013), encystment occurs during late stationary phase (Brenner et al., 2004). Azotobacters are well known because of the highest respiratory rates among living organisms as well as the production of capsules and slime in media containing solid carbohydrates (Saribay, 2003). Moreover, the production of pigment is a common feature of this genus (Brenner et al., 2004). *A. chroococcum* is the dominant species among other species and it usually

* Corresponding author.
E-mail addresses: mtrr_ebrahimi@yahoo.com (M. Ebrahimi), safari_sinegani@yahoo.com (A.A. Safari Sinegani), sarikhani@tabrizu.ac.ir, rsarikhani@yahoo.com (M.R. Sarikhani), mohammadi@tabrizu.ac.ir (S.A. Mohammadi).

http://dx.doi.org/10.1016/j.compag.2017.06.019
0168-1699/

appeared in brown-black nondiffusible pigment in aging colonies. This species is dominant in agricultural soils in temperate zones; the characteristic of this species is production of insoluble black-brown pigment in water (Brenner et al., 2004). Azotobacteria frequently found in soils with a neutral or slightly alkaline pH and their number decreased in acid soils (Lenart, 2012).

For the isolation of *Azotobacter*, in addition of using the soil-paste, usually nitrogen free (NF) mediums like Ashby, Winogradsky, Beijerinckia, LG, Brown, Burk and Jensen are being used. After preparing dilution series of soil samples, the final dilutions spread on NF medium, and after at least three days of incubation, soft, flat, transparent or milky and mucoid colonies were identified and counted as *Azotobacter* (Motsara and Roy, 2008). *Azotobacter* frequency and distribution in soil depends on soil physicochemical properties (soil organic matter, pH, temperature, soil depth, soil moisture) and microbiological characteristics (microbial interactions) (Kizilkaya, 2009). While the numbers are generally fairly low (less than 10^4 cells per g of soil), they are found throughout the world, typically in 30–80% of soils sampled, but rarely in polar region. Because of their relatively high requirement for phosphorus, they are more commonly found in fertile soils than in sand (Jackson, 1958; Brenner et al., 2004). Furthermore, land usage and land management affect the distribution of microbial population (Safari Sinegani and Sharifi, 2004). Most studies focused on the relationship between population of *Azotobacter* with soil physicochemical properties, including nutrients, organic matter and certain microbiological characteristics of the soil, as well as microbial biomass, basal respiration and enzyme activity (Pramanix and Misra, 1955; Bescking, 1961; Jensen, 1965; Burris, 1969).

There are several modeling approaches that have been utilized in different studies especially in order to estimate ecological properties of soil, both linear and nonlinear. Among the linear techniques are multivariate linear regression (MLR), partial least-squares (PLS) regression, and principal component regression (PCR) and among the nonlinear are artificial neural networks (ANN), genetic programming (GP), support vector machines (SVMs), and adaptive neuro fuzzy inference systems (ANFIS) (Borhani et al., 2016a, 2016b). Application of modeling such as artificial neural network or regression models is one approach to determine the relationship between microbial population and other environmental variables especially soil features. These tools can increase our understanding from microbial properties of soil. ANN is a simplified simulation of the human brain and composed of simple processing units referred to as neurons. In brief, a neural network consists of an input, a hidden, and an output layer all containing "nodes" (Manyame et al., 2007). It is able to learn and generalize the relationship from experimental data even if they are noisy and imperfect. This ability allows this computational system to learn constitutive relationships of materials directly from the result of experiments. Unlike conventional models, it needs no prior knowledge, or any constants and/or assumptions about the deformation (Banimahd et al., 2005).

ANNs have been used extensively in environmental modeling where parameter using large noisy datasets has been a requirement. Its use in modeling soil characteristics is well demonstrated (Abbaspour and Baramakeh, 2006; Aitkenhead et al., 2007; Elshorbagy and Parasuraman, 2008). ANN have been applied, previously, to landscape evaluation (Aitkenhead and Dyer, 2007; Kuplich, 2009) for landscape-scale assessment of soil microbial biomass (Lentzsch et al., 2005), to predict endocrine disrupting compound concentrations in soil (Aitkenhead et al., 2014), for estimating the soil temperature (Ozturk et al., 2011), predicting soil chemical composition and other soil parameters from field observations (Aitkenhead et al., 2012), for predicting crop yield (Paswan and Begum, 2013) and modeling of soil NO emissions (Delon et al., 2007). Amini et al. (2005) estimated the cation exchange capacity in the central of Iran using soil organic matter and clay contents via ANN. They concluded that network models are in general more suitable for capturing the non-linearity of the relationship between variables. Jain and Kumar (2006) indicated that the ANN technique can be successfully employed for the purpose of calibration of infiltration equations. They had also found that the ANNs are capable of performing very well in situations of limited data availability. Keshavarzi and Sarmadian (2010) developed the performance of MLR and ANN model for predicting soil parameter using easily measurable characteristics of clay and organic carbon. The value of RMSE and R^2 derived by ANN model for CEC were 0.47 and 0.94 respectively, while these parameters for MLR model were 0.65 and 0.88 respectively. Results showed that ANN with seven neurons in hidden layer had better performance in predicting soil CEC than MLR. Modeling of soil penetration resistance using statistical analyses and ANNs were studied by Santos et al. (2012). The aim of this work was to make an analysis of the soil penetration resistance behavior measured from the cone index under different levels of bulk density and water content using statistical analyses, specifically regression analysis and ANN modeling. Both techniques indicated that soil penetration resistance is associated with soil bulk density and water content. The regression analysis presented a determination coefficient of 0.92 and an RMSE of 0.95, and the ANN modeling presented an R^2 of 0.98 and an RMSE of 0.084. In other research, Akbarzadeh et al. (2009) used easily measurable characteristics of soil to estimate CEC. Eighty soil samples were collected from different horizons of 26 soil profiles. Measured soil variables included soil texture, organic carbon and CEC. Then MLR, Neuro-Fuzzy and feed-forward back propagation network were employed to develop a pedotransfer function for prediction soil parameter using easily measurable characteristics of clay and organic carbon. Results showed that Neuro-Fuzzy was superior to ANN and MLR in predicting soil property. Parvizi and Gorji (2010) used ANN model for determination of soil organic matter (SOC). This study was conducted to evaluate the effects of 15 different climatic, soil, and geometric factors on the SOC contents in different land use patterns and to determine relative importance of these desired variables for SOC estimation in one of the semi-arid watershed zones in the western part of Iran. The performance of the model was evaluated using R^2 and MBE values of tested data set. Results exhibited that 31-2-1 neural networks have highest predictive ability that explains 76% of SOC variability. Neural network models slightly overestimated SOC content, and had higher ability to detect management variables effects on SOC variability.

In the other study, researchers have produced a ANN model that predicts a wide range of soil chemical and physical properties with varying accuracy levels. This model is supplied with field observation inputs that require only a limited degree of training to determine and limited field equipment. These inputs include colour, texture classification and site information (topography, climate and vegetation). Several model outputs are predicted with a high degree of accuracy, including organic matter content, Mg, Ca, Ni, total base saturation and pH amongst others (Aitkenhead et al., 2012). Lentzsch et al. (2005) reported significant correlation of soil microbial biomass (SMB) contents and organic carbon and total nitrogen contents as well as texture. Using these data sets obtained from 89 arable sites along a regional-scale transect, a linear full-factorial regression model and a neural network were constructed and evaluated for landscape scale assessment of SMB. Result showed that a linear full-factorial regression model approach, as well as neural network modeling are promising tools for the prediction of SMB at the landscape scale but need to be validated