



Artificial neural network modeling of microbial community structures in the Atlantic Forest of Brazil



Eder C. Santos^a, Eduardo Dutra Armas^a, David Crowley^b, Marcio Rodrigues Lambais^{a,*}

^a Departamento de Ciência do Solo, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Av. Pádua Dias 11, 13418-900 Piracicaba, SP, Brazil

^b Department of Environmental Sciences, University of California Riverside, 318 Science Laboratories I, CA 92521 Riverside, CA, USA

ARTICLE INFO

Article history:

Received 25 March 2013

Received in revised form

21 October 2013

Accepted 26 October 2013

Available online 15 November 2013

Keywords:

Artificial neural networks

Ecosystem modeling

FAME

Microbial communities

Soil ecology

Tropical forests

ABSTRACT

Microbial communities vary across the landscape in forest soils, but prediction of their biomass and composition is a difficult challenge due to the large numbers of variables that influence their community structures. Here we examine the use of artificial neural network (ANN) models for extraction of patterns among soil chemical variables and microbial community structures in forest soils from three regions of the Atlantic Forest of Brazil. At each location, variations in soil chemical properties and FAME profiles of microbial community structures were mapped at 20×20 m intervals within 10 ha parcels. Geostatistical analyses showed that spatial variability in soil physical and chemical variables could be mapped at scale distances of 20 m, but that FAME profiles representing the microbial communities were highly variable and had no spatial dependence at the same scale in most cases. RDA analysis showed that FAME signatures representing different microbial groups were positively associated with soil pH, OM, P and base cations concentrations, whereas microbial biomass was negatively associated with the same environmental factors. In contrast, ANN models revealed clear relationships between microbial community structures at each parcel location, and generated verifiable predictions of variations in FAME profiles in relation to soil pH, texture, and the relative abundances of base cations. The results suggest that ANN modeling provides a useful approach for describing the relationships between microbial community structures and soil properties in tropical forest soils that were not able to be captured using geostatistical and RDA analyses.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The Atlantic Forest of Brazil is one of the Earth's “biodiversity hotspots” that is comprised of widely distributed, remnant patches of tropical and subtropical moist forests, tropical dry forest, tropical savannas, and mangroves. While largely decimated over the past three centuries, the remaining Atlantic Forest still harbors approximately 20,000 plant species, of which half are endemic (Tabarelli et al., 2003, 2005). As part of the effort to characterize the relationships between environmental variables and the composition of the remaining forest fragments, the São Paulo Research Foundation (FAPESP) has initiated a research program (BIOTA) to map the forest vegetation, soil types, and microbial communities. The latter component of this program focuses on soil variables that shape microbial community composition. To date this effort has

met with limited success, which is thought to be due to the large numbers of variables that simultaneously and interactively shape microbial community structures. Other challenges include the difficulty in determining the appropriate level of resolution and selection of spatial scales for model development that can reliably predict changes in microbial community structures across the landscape.

A variety of molecular and biochemical methods are now available for characterizing the composition of soil microbial communities. To this end, one of the most useful methods for describing microbial communities has been the use of fatty acid profiles that reflect broad level differences in microbial community composition by measuring the concentrations of signature fatty acids that represent different functional groups (White, 1993; White et al., 1996; Kaur et al., 2005). These methods include both fatty acid methyl ester (FAME) and phospholipid fatty acid (PLFA) analyses, both of which have relatively similar abilities to discriminate microbial communities. Nonetheless, the interpretation of data sets describing microbial communities is inherently

* Corresponding author. Tel.: +55 19 3417 2107.

E-mail address: mlambais@usp.br (M.R. Lambais).

complicated by their multidimensionality, which in the case of PLFA or FAME typically include some 30 or more fatty acids that are used to describe the community structures.

Moreover, as many different environmental variables appear to simultaneously influence the composition of microbial communities, many of which are correlated and interactive, the relationships between individual variables and the composition of individual fatty acids are difficult to extract using mathematical models. One of the primary methods for separating out the effects of different variables in microbial ecology employs canonical ordination such as redundancy analysis (RDA), which combines multiple regression with classical ordination (Ramette, 2007). Using RDA, the main patterns of species variation can be depicted and explained by the measured environmental variables. In addition, correlation coefficients between species and each environmental variable can be determined.

Another approach that is now receiving increased attention is the use of artificial neural network models, which can be trained to extract nonlinear patterns that exist in large complex data sets without requiring *a priori* hypotheses to guide the model development. These models may thereafter be used to predict how different combinations of variables may affect microbial community structures, and tested using new data sets to evaluate their validity (Noble et al., 2000; Mele and Crowley, 2008).

In this report, we describe the application of ANN modeling approaches for determining the relationships between FAME profiles of microbial communities from three different ecosystems of the Atlantic Forest and their correspondence with soil chemical variables that shape the community structures. The ANN models included the use of both unsupervised Kohonen self-organizing maps (KSOM) and a function-based model (Englebrecht, 2007; Mele and Crowley, 2008). KSOM are used primarily for unsupervised pattern recognition in which the goal is to find underlying structure in the data. When two variables have similar distributions in relation to the other variables in the data set, the color-coded patterns representing those variables will correspond to one another. Likewise, inverse relationships can be observed as inverse patterns. On the other hand, when two variables are unrelated they will have different map patterns. Variables that map as multiple clusters in turn suggest the existence of nonlinear interactions that are driven by different combinations of variables that come into play for each cluster.

Going beyond the unsupervised models that are generated by KSOM, ANN function based models are supervised models in which selected input variables are studied in relation to a specific dependent/output variable that is controlled by the independent input variables. Selection of the input variables can be optimized by using sensitivity analysis that measures the relative importance of all the independent variables for determining the value of the output variable. Supervised models are constructed by iterative model runs in which different sets of independent variables are methodically evaluated for their predictive power and by measuring their error when run with a previously unseen validation data set. To this end, a portion of the original data set that has been randomized is set aside for later use in model validation.

Here, we used both KSOM and function based models to examine the relationships between soil chemical and biological properties. The KSOM represent a starting point in the data analysis, and provide color coded maps, in which the relationships between all independent and dependent variables in the data set can be used to visually assess clusters in the data and form hypotheses regarding potential relationships among the variables. We then developed a series of ANN function models that are based on a multilayer perceptron architecture. In these models, selected independent variables (environmental variables) are represented by

nodes that are mathematically interconnected to the dependent variables (microbial community descriptors) in the simulated neural network. The linkages between the input and output variables are initially set with random values, and are thereafter trained using an iterative training process in which the mathematical linkages are refined to minimize the error in predicting the dependent variables. After optimizing the number of training cycles and running a series of independent ANN models to test for model robustness, we then selected the best performing models for use in predicting the concentrations of FAME biomarkers that represent different components of the microbial communities from the soil chemical data. We further compared the results obtained by the ANN approach with those obtained by conventional RDA and geostatistical analyses. Our final objective was to evaluate the relative utility of these different methods for mapping the composition of the microbial communities and to determine those variables that were most influential in shaping soil microbial community structures.

2. Materials and methods

2.1. Site description and soil physico-chemical analyses

In preliminary work to establish the experiment, sites were selected from 3 different ecosystems of the Atlantic Forest that represent different types of forests. To our knowledge, all of the sites represent pristine Atlantic Forests that today comprise only 3–5% of its original area in Brazil. Samples were collected in 10 ha permanent parcels established in the Carlos Botelho State Park (CB), Assis Ecological Station (AS) and Caetetus Ecological Station (GA). The main features of the sites are described in Table 1.

At each location, the permanent parcel was physically mapped and subdivided into 20 × 20 m plots arranged in a rectangular grid over the landscape. Within each of the 256 subplots at each location, soil samples were collected at 0–5 cm depth using stainless steel cylinders (50 mm diameter) that were hammered into the soil to obtain intact soil cores that were placed in coolers and transported to the laboratory for processing. The soil samples were air dried and removed from the cores, after which they were sieved through a 2-mm screen and homogenized. The sieved soils were then divided into paired subsamples, one of which was sent to the University of California, Riverside for FAME analysis, and the other half was analyzed for soil chemical properties at the ESALQ campus of the University of São Paulo in Piracicaba, Brazil. Air-dried and sieved soil samples were analyzed for physical (sand, silt and clay contents) and chemical (pH, OM, P, Na, K, Ca, Mg, Al) properties, according to standard soil testing methods (Rajj et al., 1987).

2.2. FAME analyses

Duplicate (detection limit analysis) or triplicate (sample discrimination analysis) subsamples of 0.5 g of soil (DW) (unless stated otherwise) were extracted according to the Microbial Identification System (MIS; Microbial ID Inc., Newark, DE) standard procedure. To each soil sample, 3.25 M NaOH in MeOH:H₂O (1:1) was added (1 ml solution added per 1 g soil). The samples were vortexed and then placed in an 80 °C water bath for 30 min, during which time the cells were lysed and the FAs were cleaved from the cellular lipids. Following this saponification step, the FAs were converted to FAMES by adding 6.0 M HCl:MeOH (1:0.85) (2 ml solution per 1 g soil) to each sample. To extract the FAMES from the acidic aqueous phase into the organic phase, a hexane:MTBE (1:1) solution was added to each sample (2 ml solution per 1 g soil). Following addition of the hexane:MTBE (1:1) solution, the MIDI procedure then was modified as described in Cavigelli et al. (1995)

Download English Version:

<https://daneshyari.com/en/article/2024718>

Download Persian Version:

<https://daneshyari.com/article/2024718>

[Daneshyari.com](https://daneshyari.com)