Modelling recovery rules of soil microbial assemblages using matrix methods

M. Anand, K. M. Ma, S. Levin, A. Okonski & D. McCreath

Biology Department and Elliot Lake Research Field Station, Laurentian University, Sudbury, Ontario, Canada

Abstract

We model the dynamics of microbial assemblages in response to various rehabilitation treatments to polluted soils in the Sudbury, Canada area. We compare the efficacy of the stationary Markov model in capturing observed dynamics and measuring the success of rehabilitation techniques. We apply two different approaches to the estimation of transition probabilities for the Markov model. The first method models competition between taxa for a limiting resource. The second method is less mechanistic, however, and is closely linked to the matrix method of Principal Components Analysis that is widely used to detect ecological gradients. The first method provides a better fit to the dynamics of microbial assemblages and to assessment of rehabilitation success, but is less stable in the prediction of final states. The second method provides a generally poor fit to assemblage dynamics, but a more accurate prediction of final state. We emphasize the importance of defining criteria for model assessment and how these may depend intrinsically on model construction.

1 Introduction

Assembly rule is a term commonly used in the ecological literature to refer to patterns of species (or, in general, taxa) turnover in a community (or, in general, assemblage). While no consensus exists (or perhaps, should be expected) on the nature of these rules, some have gone so far as to claim that no such rules exist [1]. We maintain the view that assembly rules do exist and information about them, however specific to a given ecological taxonomic group, habitat, or state (e.g., natural or perturbed), is necessary for understanding the origin and maintenance of these assemblages. As such, modeling can be a very useful
exercise in the development of ecological indicators. Modeling assembly rules for ecological systems is, however, a challenging enterprise because the ecological data needed to either parameterize models or to test them are difficult and time-consuming to collect. Furthermore, while patterns can be detectable in observational data, processes may only be revealed through manipulative experiments. While such manipulations can create highly artificial systems, they can help us to understand simple dynamics and assembly rules which otherwise may be undetectable in a sea of unknown and random effects. Dynamics may be simplified by compressing the ecological system in space and/or time (e.g. by studying it in a laboratory), but one aspect of ecological systems that can be fairly easily handled is the presence of multiple interactions. Various multivariate statistical and matrix modeling techniques now exist that can digest data from high-dimensional ecological systems with multiple interactions.

In this paper, we model the assembly rules of soil microbial assemblages subject to different soil rehabilitation techniques using matrix models. We focus on the stationary Markov chain model. This is a simple matrix model that attempts to parameterize turnover in assemblages via a transition matrix. Markov models have a long history of application in ecology, and the advantages and disadvantages of using them are well-documented [2]. The main criticism—that they are too simple—can also be considered their strength. They continue to be extremely popular in ecology, but the issue of how to estimate the transition probabilities in the absence of direct observations has not been fully resolved. Estimation of these is critical, however, because it solely defines the long-term or "climax" state of the assemblage [3]. The Markov model is compared to another matrix model, namely Principal Components Analysis (PCA), which also attempts to capture assemblage-level dynamics, but where parameterization comes from the minimization of resemblance (or correlation) between taxa through eigenanalysis. This model has widespread application [4], but is rarely used in dynamical analyses (but see [5]). We ask: How well do Markov-based vs. PCA-based models capture assemblage dynamics and how can they be used in a specific environmental application?

2 Environmental problem and source data

Recovery dynamics of soil microbial assemblages from the acidic, heavy metal contaminated soil from the Cu-Ni smelter-damaged landscape of Sudbury, Canada (see [6] for overview) were studied. Greenhouse experiments were designed to determine the effect of various rehabilitation techniques on microbial recovery. The following five treatments were replicated three times each: T1: Control – no treatment, T2: Soil + lime, T3: Soil + paper sludge, T4: Soil + zeolite-Clinoptilolite, T5: Soil + zeolite-AKA. After mixing appropriate doses of the amendments with soil, Agrostis scabra was grown. To eliminate the influence of vegetation on soil microbiota, a variant of control (non-treated) soil with no vegetation was used to compare with control soil where A. scabra was grown (T6: soil without plants). The treatments were chosen for their feasibility as rehabilitation techniques and previous studies on their success in other
systems. Liming, a common agricultural practice to increase soil pH, also reduces the mobility and uptake of heavy metals such as Mn, Cu, Ni, and Zn [7]. Paper sludge reduces mobility of heavy metals and natural zeolites are used as soil conditioners, as they can absorb heavy metals [e.g., 8, 9].

Microbial response was determined using soil taken from the greenhouse experiment. Soil samples were characterized by the following pH levels: T1 (control) – 5.1; T2 (lime) and T3 (paper sludge) – 7.2; T4 and T5 (zeolites) – 5.6; T6 (control, no vegetation) – 5.1. The method of Initiated Microbial Community [10,11] was utilized to study response of the amylolytic microbial communities on various soil treatments, considered to be biological indicators of recovery. After three days of pre-incubation of soil samples in Petri dishes, a thin layer of soluble starch was spread on the soil surface. A layer of starch with a thickness of several grains and an area of about 4 cm² formed on the surface of the soil samples. The development of microorganisms on the starch layer was observed visually and using reflected light microscopy. Morphotypes were determined and their abundance in the community was estimated. More precise identification of soil microorganisms was made on the basis of morphological properties, using traditional microbiological methods [12-15]. An amylolytic microbial assemblage was evaluated using both taxonomic composition and the percentage of surface area occupied by all taxa. Each experiment was repeated 9 times, and data was recorded 6, 9, 14, 18, 21, 24, 29, 32, and 35 days from initiation. A complete taxa list along with functional description is in Table 1.

3 Modelling assemblage-level dynamics

Biologists are often interested in the dynamics of individual taxa. We focus on assemblage-level dynamics, which is often overlooked [16]. This is not equivalent to summing across taxa to produce a ‘total’ dynamic, but rather is an attempt to capture assemblage structure. Principal Components Analysis (PCA) is a multivariate statistical technique that attempts to summarize multivariate dynamics such that focus is on the assemblage as a whole, rather than its parts (e.g., taxa). Dimensionality of the new assemblage is usually much less than the original and allows for visualization, often in 2-dimensional space. The PCA model relies on a summarization of dynamics through reducing resemblance between populations through time via eigenanalysis. The resemblance measure used can be one of many [14], but here, we chose the simplest:

\[ s_{ab} = \sum (X_{ai} \cdot X_{bi}) \quad i=1, 2, ... n \]  

where \( s_{ab} \) is the Product Moment between taxa \( a \) and taxa \( b \) at timestep \( i \), \( X_{ai} \) is an abundance estimate for taxa \( a \) at timestep \( i \), and the \( X_{bi} \) is an abundance estimate of taxa \( b \) at the same timestep. Resemblance matrices were calculated both on un-standardized data and on standardized data. For the latter, Product Moment becomes the Pearson correlation coefficient. Summarized linear or monotonic population dynamics will appear linear or monotonic in PCA space.
If we define $X_1, X_2, \ldots, X_c$ as different vector states (e.g., multiple taxa in an assemblage) of a transition process (e.g., recovery dynamics of the assemblage in time), and $P$ as transition probability (e.g., between taxa, from one timestep to the next) matrix, a Markov chain can be described as follows,

$$X_{t+1} = X_t \cdot P$$

and all the $X_t$ will be determined when $X_1$ and $P$ are given. The problem of transition probability estimation amounts to establishing a model for population transitions, since transitions are not directly measurable. Transition probabilities for the microbial assemblage were estimated in two ways. The first was using the method introduced into the ecological literature by Orlóci et al. \[17\] and implemented in \[18\]. For an observed process, the transition matrix $P$ can be estimated by averaging all transition matrices between all adjacent states. If states $j$ and $k$ of the transition are, $X_j = [x_{j1}, x_{j2}, x_{j3}, \ldots, x_{jn}]$ and $X_k = [x_{k1}, x_{k2}, x_{k3}, \ldots, x_{kn}]$, where $n$ is the number of variables (e.g., taxa), the process will have three transition possibilities from state $j$ to $k$ for the $i$th variable, Type 1: $x_{ki} - x_{ji} > 0$, Type 2: $x_{ki} - x_{ji} < 0$ and Type 3: $x_{ki} - x_{ji} = 0$. The transition rate between variables $i$ and $h$ from state $j$ to $k$ of the transition matrix $P$ can then be determined by,

$$\text{DEV}(i) = \frac{|x_{ki} - x_{ji}||x_{kh}|}{x_k} \quad h \neq i$$

$$\text{DEV}(i) = \frac{x_{ki}|x_{h}|}{x_k} \quad h = i$$

where $x_{kh}$ is the value of $h$ variable at state $k$, and $x_k$ is the row total of state $k$. Formula (2) indicates how much variable $i$ will be transferred to $h$ from state $j$ to $k$. For $x_{ki} - x_{ji} > 0$, the transition rate $p_{ih}$ is defined to point to row $i$ column $h$, and if $x_{ki} - x_{ji} < 0$, the $p_{hi}$ to row $h$ column $i$. Similarly we can get all the other transition matrices of the process, and the average transition matrix can be generated by averaging the corresponding elements in all the matrices. The above procedures give Markov transitions with state row total equal to 100, which means all the variables (taxa) can only replace or be replaced without external influence. For cover data, we impose a limiting resource – 100% cover. But for those without (obvious) limits, we need to find other ways to define limits, or choose other methods for prediction.

To attempt to provide a closer link between the Principal Components Analysis and the Markov model, we attempt to estimate transition probabilities directly from the Product Moment resemblance matrices. The idea behind this was simple: product moment of population abundance over time was thought to provide an indirect measure of population interactions. If product moment within a population is high, the probability of the population maintaining itself should also be high. If product moment between populations is high, then the probability of one of the populations replacing the other should also be high. An apparent choice is in whether to divide by row total or column total in order to convert product moments to proportions. However, this choice is constrained by the fact that transition matrix rows must have unit sum, whereas column sums may be different. Once the choice has been made as to how to estimate the transition
probabilities, the estimated average transition matrix $P$ and initial state $X_t$ can be used to generate Markov predictions $M_2, M_3, \ldots M_c$ according to

$$M_k = M_j P$$

$j < k$

(5)

Note that $M_1 = X_t$. Model results were compared using the Mantel test [19] to compare observed and model dynamics. Fit of models was also assessed by comparing the predicted stable states and observed final states, with particular emphasis on relative abundance and ordering of taxa.

Table 1: Taxa appearing in soil rehabilitation experiment.

<table>
<thead>
<tr>
<th>Scientific Name</th>
<th>Abbr.</th>
<th>Function</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternaria alternata (Fr.) Keissler</td>
<td>A1A1</td>
<td>facult. phytopathogen</td>
<td>fungus</td>
</tr>
<tr>
<td>Arthrobacter oligospora Fres.</td>
<td>ArOl</td>
<td>saprophyte/predator</td>
<td>fungus</td>
</tr>
<tr>
<td>Bacillus spp.</td>
<td>BaSp</td>
<td>saprophyte</td>
<td>bacteria</td>
</tr>
<tr>
<td>Chaeotomum sp.</td>
<td>ChSp</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Coniothyrium sp.</td>
<td>CoSp</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Drechslera sp.</td>
<td>DrSp</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Geotrichum candidum Link</td>
<td>GeCa</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Gliocladium catenulatum Gilm &amp; Abbot</td>
<td>GiCa</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Humincola grisea Trauen</td>
<td>HuGr</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Mortierella ramanniana Linnam.</td>
<td>MoRa</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Penicillium dalea Zaleski</td>
<td>PeDa</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
<tr>
<td>Penicillium funiculosum Thom</td>
<td>PeFu</td>
<td>facult. phytopathogen</td>
<td>fungus</td>
</tr>
<tr>
<td>Penicillium janthinellum Biourge</td>
<td>PeJa</td>
<td>facult. phytopathogen</td>
<td>fungus</td>
</tr>
<tr>
<td>Stachybotrys chartarum</td>
<td>StCh</td>
<td>facult. phytopathogen</td>
<td>fungus</td>
</tr>
<tr>
<td>Streptomyces spc. Albus</td>
<td>StAl</td>
<td>saprophyte</td>
<td>actinomycetes</td>
</tr>
<tr>
<td>Streptomyces spc. Cinereus</td>
<td>StCi</td>
<td>saprophyte</td>
<td>actinomycetes</td>
</tr>
<tr>
<td>Trichurus spiralis Hasselbr.</td>
<td>TrSp</td>
<td>saprophyte</td>
<td>fungus</td>
</tr>
</tbody>
</table>

4 Results

Clearly, treatments had considerable effect on the observed composition of microbial assemblages and on total cover (Fig. 1). The results from Principal Components Analysis are shown in Figure 2. The success of PCA in summarizing variation in the data was generally higher when using product moment (unstandardized data) as opposed to correlation (standardized data) matrices for eigenanalysis, and only these are presented. PCA was successful in summarizing a majority of variation (>85%) in two components, meriting a two-dimensional presentation. PCA was also generally successful in capturing the temporal evolution of the assemblages, more linear in some treatments than in others. The reference space in PCA is linear, therefore, trajectories that appear
nonlinear reflect underlying nonlinear dynamics. Increasing or deceasing trends in trajectories, however, do not reflect identical changes in cover, but simply a change in the assemblage composition. It appears that Treatments 4 and 5 (the two zeolite treatments) show very similar assemblage-level dynamics. Treatments 2 (lime) and 3 (paper sludge) show somewhat similar patterns. Interestingly, the most linear dynamic is seen in treatment 6 (no rehabilitation).

If we compare assemblage-level dynamics to the taxa-level dynamics (Fig. 1), we see that Treatments 4 and 5 are similarly high in taxa-richness with 7 taxa common to both treatments. However, taxa-level dynamics is different in the sense that in Treatment 5, Streptomyces clearly dominates, whereas it occurs in almost the same quantities as Gliocladium catenulatum in Treatment 4. The highest taxa-richness was observed in Treatment 3 with Bacillus sp. dominating and two unique taxa: Humicola grisea and Stachybotrys chartarum.

Simulated Markov dynamics based on transition probability estimation from Orlóci et al. [17] are shown in Figure 3. In most cases, stability of the model

Figure 1: Microbial dynamics under different treatments T1-T6 (please see main text for treatment description and full names of taxa.)
(defined as a less than 1% change in assemblage structure from one time step to the next) occurred in around 40 timesteps. Markov model dynamics using Product Moment matrices are shown in Figure 4. These models provided a poorer fit to observed dynamics. While the final state was correctly predicted, the time to reach stability was too fast (3 timesteps). Mantel tests were carried out to determine similarity between model and observed dynamics and to compare models. The results are shown in Table 2. The final states’ correlation was also calculated in order to compare two methods without full assemblage dynamics. For probability estimation using Orlóci et al. [17], the fit of the model was generally high (p<=0.1). For estimation using the resemblance matrix, the fit of the model was poor. Interestingly, if we compare only the final states of the two models, they are highly correlated for all treatments (p<0.01).

![Figure 2: First two components of PCA on microbial dynamics under different treatments using product moment as resemblance. T1-T6 are described in the text. Component scores are timesteps in the experiments.](image-url)
Table 2: Comparison of Markov (M), Resemblance (R) and Observed (O) dynamics for each treatment. Elements in columns 2-4 are probabilities from randomization testing (Mantel test) and significant values (p<0.05) are in bold type. Final state correlation is between M and R; all are significant (p<0.01).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>M vs. O</th>
<th>R vs. O</th>
<th>M vs. R</th>
<th>Final state correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.09</td>
<td>0.43</td>
<td>0.55</td>
<td>0.85</td>
</tr>
<tr>
<td>T2</td>
<td>0.07</td>
<td>0.47</td>
<td>0.05</td>
<td>0.93</td>
</tr>
<tr>
<td>T3</td>
<td>0.06</td>
<td>0.4</td>
<td>0.3</td>
<td>0.96</td>
</tr>
<tr>
<td>T4</td>
<td>0.05</td>
<td>0.4</td>
<td>0.21</td>
<td>0.79</td>
</tr>
<tr>
<td>T5</td>
<td>0.03</td>
<td>0.26</td>
<td>0.19</td>
<td>0.91</td>
</tr>
<tr>
<td>T6</td>
<td>0.03</td>
<td>0.12</td>
<td>0.15</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Figure 3: Markov fitting of microbial assemblage dynamics with transition matrix estimated by Fitmarkov [18]. Treatments T1-T6 and names of taxa are in explained in the text.
Figure 4: Markov fitting of microbial dynamics with transition matrix derived directly from resemblance matrix. Treatments T1-T6 and names of taxa are in Fig. 1.

5 Discussion

There are two fundamental questions we may ask about this study: Which modeling approach is more useful, and which rehabilitation technique is more useful from the point of view of soil microbiology? While each of these questions by itself is quite difficult to answer, the point we wish to stress here is that the two questions may also be very dependent on one another. Indeed, we could answer the second question using conventional criteria (without modeling). A simple examination of the effects of the techniques on lowering pH in these acidic soils would result in the following ordering of treatments: T3=T2>T4=T5>T1=T6. However, of interest here was microbial response. As such, one obvious criteria to apply could be taxa richness, with the idea that the more taxa appearing in the assemblage, the better: The ordering would then be T3>T5>T4>T1=T6>T2. We could group the treatments into high richness (T3, T4 and T5) and low richness groups (T1, T2, T6). Next, we could examine taxa
composition, and search for 'indicator' taxa. Here, we could focus on the importance of bacteria since it is known that lower pH levels should be associated with domination of fungi over bacteria [20,21]. If we do this, we are left with T2 and T3 and with increased richness, T3 would be considered "better". Are these results confirmed by the modeling performed?

The main objective of this paper was related to the first question and to examine the efficacy of the Markov model in capturing observed dynamics based on two methods of transition matrix estimation. We found that, while both methods resulted in statistically similar predicted final states of the microbial assemblage, the estimation technique of Orlóci et al. [17] provided a much better dynamical fit to the data. However, closer examination of the ordering of taxa, from dominant to rare, reveals something interesting: In simulations using the methods of Orlóci et al. [17], the ordering of taxa was the same as the observed ordering only in Treatment 3 (paper sludge), indicating the highest stability in this treatment. Using the second modeling method, ordering was the same in all treatments except Treatment 2 (lime) indicating high stability in all treatments except Treatment 2. Thus, it appears that the method of Orlóci et al. [17] provides a better model for predictability or stability of dynamics in time (and corresponds more clearly to what we would expect based on simple indicators such as pH and total richness), whereas the resemblance matrix method loses dynamical information, but more accurately captures the final state.

Our results highlight the fact that modeling methods based on resemblance structure in time (e.g., the Markov model used here, PCA and other ordination methods) may help to detect gradients, however, cannot capture dynamical information related to transitions and/or autocorrelation. Since these represent such a broad class of models, our results should be important to consider when applying such models to compare dynamics of multivariate assemblage-level dynamics. We recommend in the least that a plurality of approaches be used when attempting to apply the methods to environmental problems involving high-dimensional, dynamical data.

Acknowledgements

This research was supported by Natural Sciences and Engineering Research Council grants to M.A. and D.M, a Laurentian University Research Fund grant to M.A., and the Elliot Lake Research Field Station.

References


