

Mathematical modelling of time-course microbial interactions in anaerobic digestion

Tian Hui with supervisor A/Prof Kim-Anh Le Cao

Background

Biowaste in landfills can be converted into biogas through anaerobic digestion. It is a complex process since many bacteria in the microbial community will interact with each other and aid in the yield of biogas. We can model this process to learn more about it.

Data

- A/Prof Olivier Chapleur provided the abundance data of 2231 taxa over time within digesters under different inhibitory conditions with different support media (in total 18 experimental conditions). [1]
- The population counts are sampled over different time points by 16S RNA gene metabarcoding. The number of time points varies on different experimental conditions.
- Since it is too complicated to directly analyze all the data. We start with the data on one of the experimental conditions (in presence of 1.3 g/L of phenol with no support media).

Feature Selections

We need to select those bacteria that are interacting or cooperating to be included in the model. But we have over 2000 taxa.

To choose potentially interesting bacteria, we first did some pre-processing.

- remove low counts.
- apply centered log ratio transformation (for each count, divide by geometric mean and then take log)

Then we use k-means to cluster the data.

For each cluster, we fit spline to visualize and choose the cluster with bacteria's abundance that tend to vary over time.

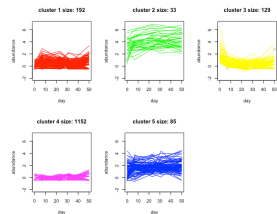


Figure 1. An example of the clustering. From the graphs, we can see that cluster 4 has a large size and the abundances do not vary much over time. Cluster 2 seems interesting as the abundances vary a lot across time.

Model

The generalized Lotka-Volterra (gLV) equations can model the abundance of taxa in a microbial community over time.

$$\frac{dx_i}{dt} = x_i \left(r_i + \sum_{j=1}^n \alpha_{ij} x_j \right)$$

x_i denotes the abundance of taxa i .

r_i represents the intrinsic growth rate of taxa i .

n denotes the total number of taxa.

α_{ij} is a coefficient representing the strength of the influence of taxa j exerted on

Methods

- To fit the gLV model, we need to estimate those coefficients in the equations, namely, the growth rates for each taxon and the interaction strength between them.
- But microbiome data is compositional. The total count in each sample is fixed and the count of each taxon only represents its abundance relative to other taxa rather than its absolute abundance.
- Because of the compositional nature, usual methods such as Pearson do not work well.
- One way to do this is to rewrite the gLV equations and discretize the time into small intervals:

$$\ln x_i(t_{k+1}) - \ln x_i(t_k) \approx \left(r_i + \sum_{j=1}^n \alpha_{ij} \left\{ \frac{x_j(t_{k+1}) + x_j(t_k)}{2} \right\} \right) \Delta t$$

Using the log lagged time difference, we can fit linear regression to get an estimate of r and α . [2]

References

- [1] Poirier, S., & Chapleur, O. (2018). Influence of support media supplementation to reduce the inhibition of anaerobic digestion by phenol and ammonia: Effect on degradation performances and microbial dynamics. *Data In Brief*, 19, 1733-1754.
- [2] Kuntal, B. et al. (2019). Web-gLV: A Web Based Platform for Lotka-Volterra Based Modeling and Simulation of Microbial Populations. *Frontiers In Microbiology*, 10.
- [3] Bucci, V. et al. (2016). MDSINE: microbial dynamical systems inference engine for microbiome time-series analyses. *Genome Biol.* 17:121.
- [4] Shaw, G. T., Pao, Y.-Y., and Wang, D. (2016). MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. *BMC Bioinformatics* 17:488.

Results

We use the implementation by web-gLV to fit the model. [2]

- parameters in the gLV model are estimated from the abundance data
- then the model is used to predict the trend.
- The predicted trend is compared with the observed trend to evaluate fitness.

We tried different combinations of features including highly/poorly correlated ones. We find that sometimes the predicted abundances deviate a lot from the observed abundances at the same time points but sometimes they do not.

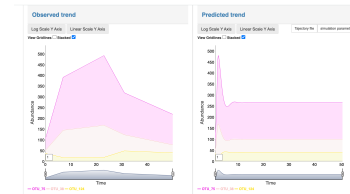


Figure 2. One example of the comparison between the observed and predicted trend using three OTUs.

Limitations

In this web-gLV implementation, the limit of the number of species is 10.[2]

Sometimes the predicted abundances blow up (go very large in a small amount of time). No apparent pattern of when this will occur is found. The simulation may be sensitive to initial abundances (behave quite differently if setting different initial values).

There are also some other implemented tools for fitting gLV models. But some may not suit our purposes. For example,

- MDSINE requires biomass data besides abundance as input [3]
- MetaMIS does not give growth rates [4]

Experience

First, I'd like to thank my supervisor A/Prof Kim-Anh for her guidance and support, as well as the opportunities to meet fellow students and researchers. Although I do not get a completed result, this is still a very rewarding project. The research experience is quite different from coursework. I learned a lot about some research methods from conversations with Kim-Anh. For example, how to break the problem down into smaller parts.

Finally, I also would like to thank the school of mathematics and statistics for providing this opportunity. I would recommend any student interested in mathematics and research to participate.