

**Title: Bayesian Conditional Auto-Regressive LASSO Models to Learn Sparse Networks with Predictors**

**Authors:** Yunyi Shen<sup>1\*</sup> ([yshen99@wisc.edu](mailto:yshen99@wisc.edu)) and Claudia Solis-Lemus<sup>1</sup>

**Institutions:** <sup>1</sup>University of Wisconsin-Madison, Madison, WI

**Website:** <https://github.com/YunyiShen/CAR-LASSO>

**Project Goals: Short statement of goals. (Limit to 1,000 characters)**

1. Develop novel statistical theory on a Bayesian regression framework that can predict changes in microbial compositions due to experimental or environmental factors. Our novel model will simultaneously account for the microbial interactions and predictor effects
2. Produce an efficient sampling scheme to estimate the posterior distribution of model parameters that is scalable enough to meet the ever-growing needs of high-dimensional soil microbiome data
3. Implement a publicly available open-source R package that would allow scientists to utilize our novel model in their own soil microbiome data to make inferences about the specific predictors (experimental or environmental) that are playing a role in shaping the microbial compositions in soil

**Abstract text:** Please limit such that entire document does not exceed 2 pages.

Microbial communities are among the main driving forces of biogeochemical processes in the biosphere. In particular, many critical soil processes such as mineral weathering, and soil cycling of mineral-sorbed organic matter are governed by mineral-associated microbes. Understanding the composition of microbial communities and what environmental factors play a role in shaping this composition is crucial to comprehend soil biological processes and to predict microbial responses to environmental changes.

However, the inter-connectivity of microbes-environment is still not fully understood. One of the reasons for this gap in knowledge is the lack of statistical tools to infer connections among microbial communities while simultaneously accounting for predictors in a unified framework. Indeed, there is a need for statistical models that can decode microbes' reaction to the environment and interactions among microbes simultaneously. The model should have the ability to correctly incorporate prior knowledge from controlled experiments and its implementation should be scalable enough to meet the ever-growing needs of the high-dimensional soil microbiome big data.

We introduce a novel Bayesian conditional auto-regressive (CAR) LASSO model to infer a sparse network structure with nodes for responses and for predictors. Directed edges between a predictor and a response represent conditional links, and undirected edges among responses represent correlations. Specifically, our model estimates a microbial network that represents the dependence structure of a multivariate response (e.g. abundances of microbes) while simultaneously estimating the effect of a set of predictors that influence the network (e.g. diet, weather, experimental treatments). In addition, our method produces a sparse interpretable graph via LASSO penalization.

We also propose an adaptive extension of the CAR LASSO model so that different shrinkage can be applied to different edges which allows the incorporation of edge-specific prior knowledge. Indeed, the conditional representation of our model coefficients and adaptivity allow us to adequately encode prior knowledge obtained by specific experimental interventions and agrees with the experimenter's intuition on average behavior of nodes under experiments. In addition, our model is able to equally handle small and big data and is computationally inexpensive through an efficient Gibbs sampling algorithm. With hierarchical structure, we extend the model to binary, counting and compositional responses by adding an appropriate sampling distribution to the core Normal model. Finally, we apply our model to a public soil microbial composition dataset. Researchers isolated soil aggregates from three land management systems in central Iowa to test if the aggregate-level microbial responses are related to plant community and management practices. Our work estimates the soil microbiota network and identifies the most important connections between microbes and the most important microbes-environment links.

### **References/Publications**

1. Shen, Yunyi, and Claudia Solis-Lemus. "Bayesian Conditional Auto-Regressive LASSO Models to Learn Sparse Networks with Predictors." *preprint arXiv:2012.08397* (2020).

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021016.