Overview

These scripts were used to generate the simulation data that was included in figures for Clark et al. 2020. Please see Figure 1 for a flow chart describing which data and which scripts were used to train each model and simulate the communities shown in the figures. All of the quantitative results from this work can be reproduced by following the flow diagram in Figure 1.

Software Dependencies

All MATLAB scripts were written for MATLAB R2018b run on either Windows 10 or Linux CentOS.

Also enclosed are:

Python3CondaSpec.txt > Description of the conda environment for running the Jupyter Python 3 Notebook analyses

JuliaSpec.txt > List of the packages and version installed in Julia for running Julia scripts. Note that installation of IPOPT is also required according to the instructions found here: https://github.com/zavalab/JuliaBox/tree/master/MicrobialPLOS

Special Hardware Use

HTCondor parallelization was used for Model M3 via the UW-Madison CHTC: https://chtc.cs.wisc.edu/

MATLAB PARFOR parallelization can be run on any multi-core computer. We used a private computational server with 32 cores to speed up execution of these scripts.

Otherwise, scripts should run on any computer.

Demos

Training a gLV Model

The enclosed scripts constitute examples of applying our methods to various datasets. In general, the Julia NLP approach (Models M1 and M2) works well for batches of data of low richness communities and is much faster (~5-10 minutes for a given solution for Model M1), whereas the MATLAB FMINCON approach is more computationally intensive (~4-12 hours to find a single solution for our datasets on our computational server) but works well with high richness communities where the Julia NLP approach fails to converge. Expected parameter outputs are included for the specified parameter estimation problems with the specified hyperparameters.

Training Regression Models

Our regression models use standard functions in MATLAB (Models M1 and M2) or Python scikitlearn. The enclosed scripts constitute examples of how to train such models and expected outputs are included. The examples should run in <30 minutes on a standard computer.

Simulating New Community Scenarios

The Model M1 scripts for simulating all possible 3-5 species communities assembly (with the gLV model) and butyrate production (with the regression models) are included as well as the compiled expected outputs. We parallelized this process using our computational server with 32 cores to simulate all of these communities with all parameter sets. If a simpler example is desired, these

same scripts can be run after removing all but one of the parameter files from the 'posterior_RLC3/' folder (i.e. to skip all but one of the parameter sets). This reduced problem should finish on a standard computer in less than an hour.

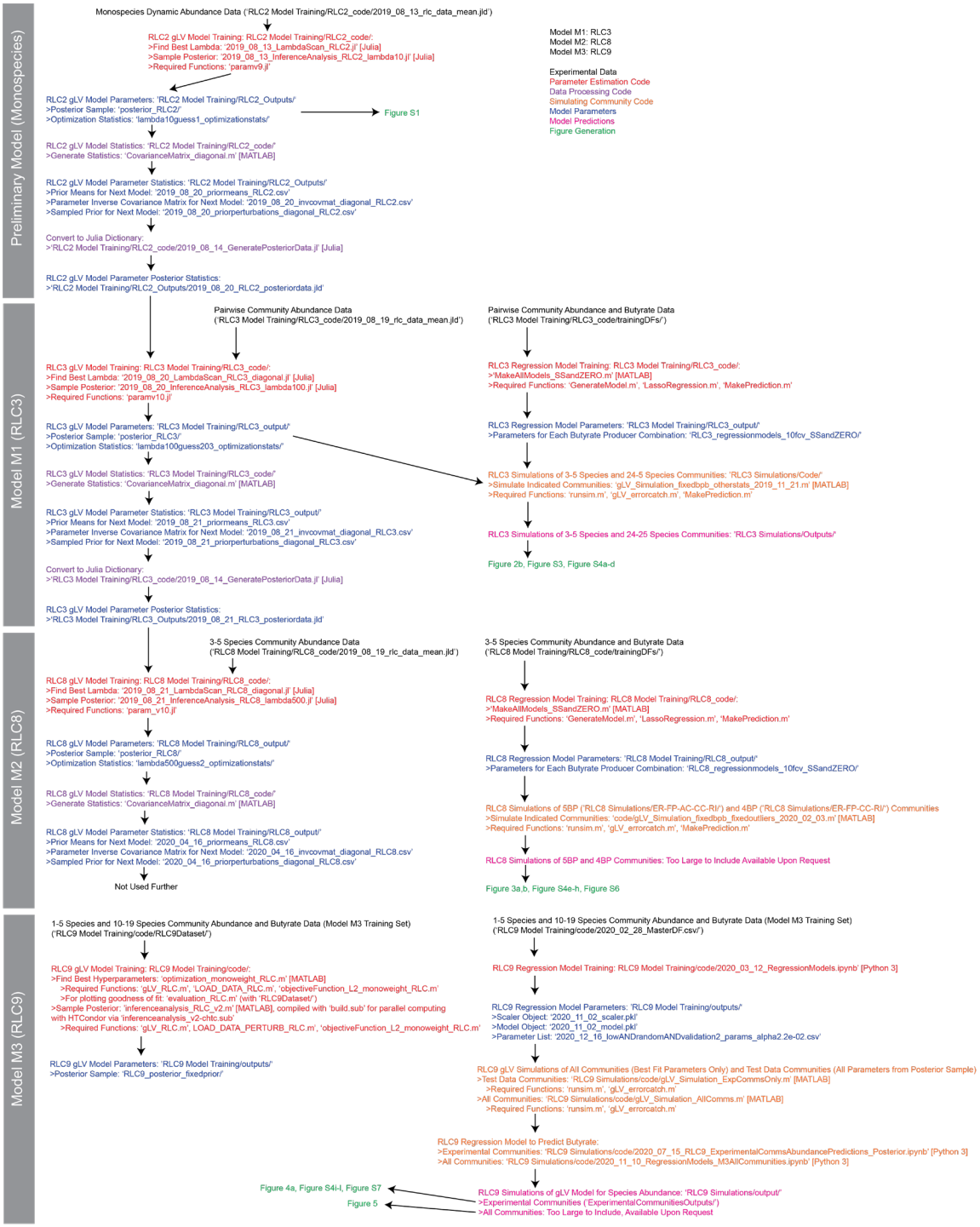

Figure 1. Flow Sheet for Generation of Simulation Data Included in Figures for this Manuscript