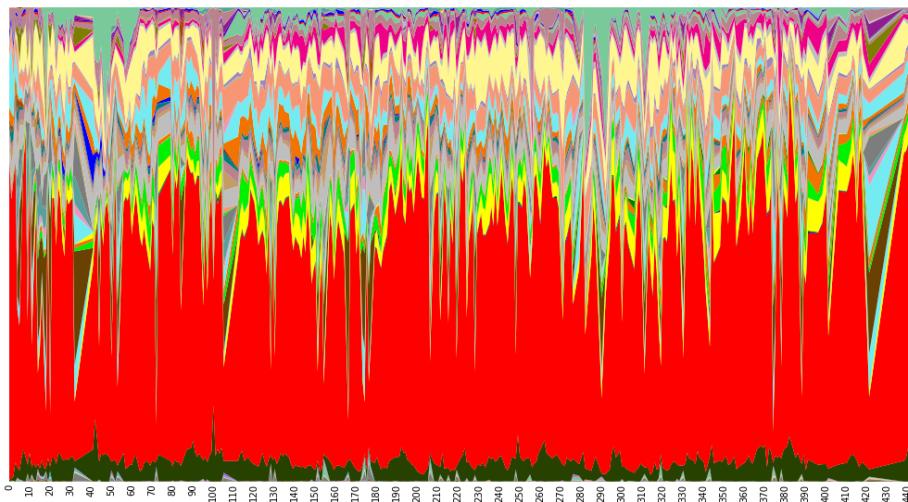


Microbial community model selection using time series properties

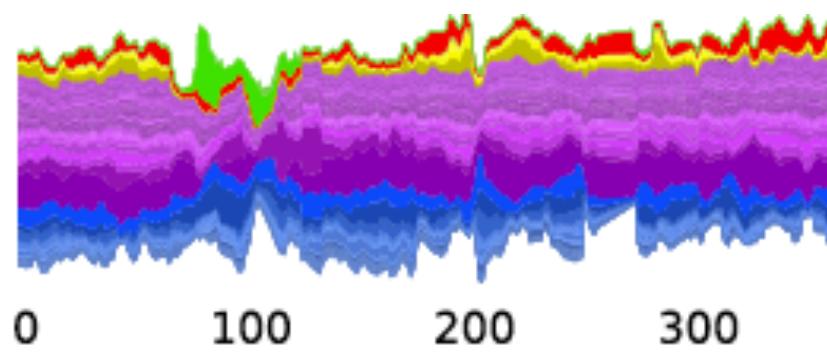
Collaboration with Franziska
Bauchinger, Sophie De Buyl, Leo Lahti,
Alex Washburne, Didier Gonze and
Stefanie Widder

Microbial community dynamics

- Long 16S time series available for human-associated microbial communities (gut, skin)



Caporaso et al.: Male 3 feces, genus abundances
396 time points



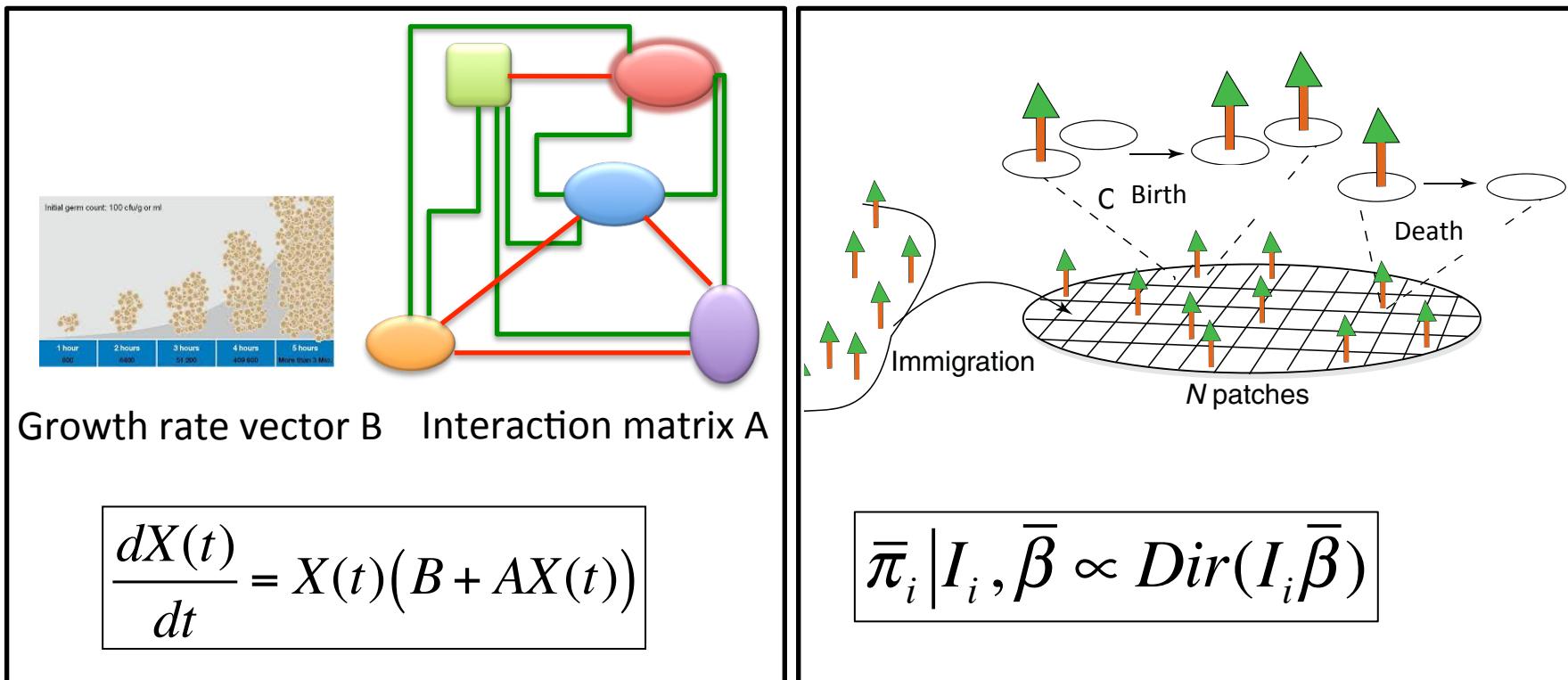
David et al.: Subject A feces, phylum abundances
329 time points

Caporaso et al. (2011). "Moving pictures of the human microbiome"

David et al. (2014) "Host lifestyle affects human microbiota on daily timescales" Genome Biology 15, R89

What drives microbial community dynamics?

- **Selection** (growth rates and interactions matter)
- **Neutral** (birth/death rates and immigration is enough to describe the dynamics)

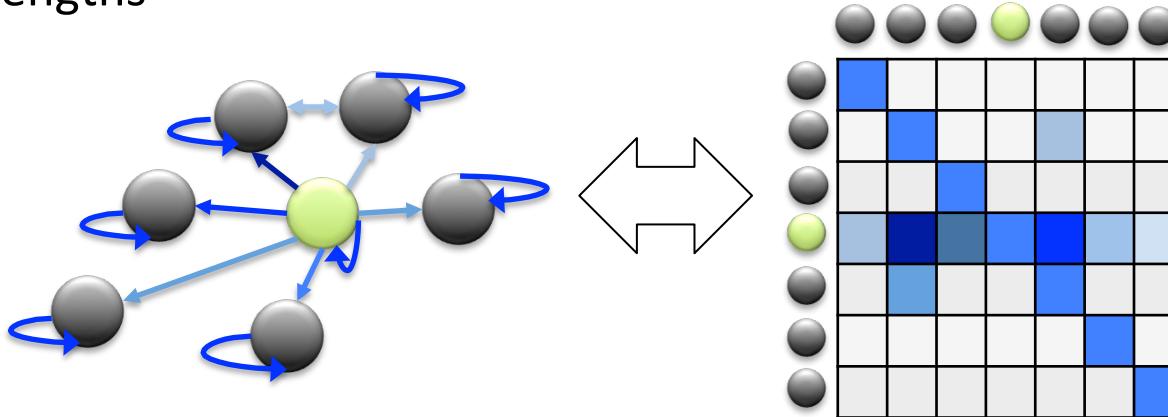


Lotka, A.J. (1925). *Elements of Physical Biology*. Baltimore: Williams & Wilkins Company.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.

Selection view: generalized Lotka-Volterra (gLV)

- The species network can be represented by the interaction matrix A (also known as community matrix), whose entries represent interaction strengths



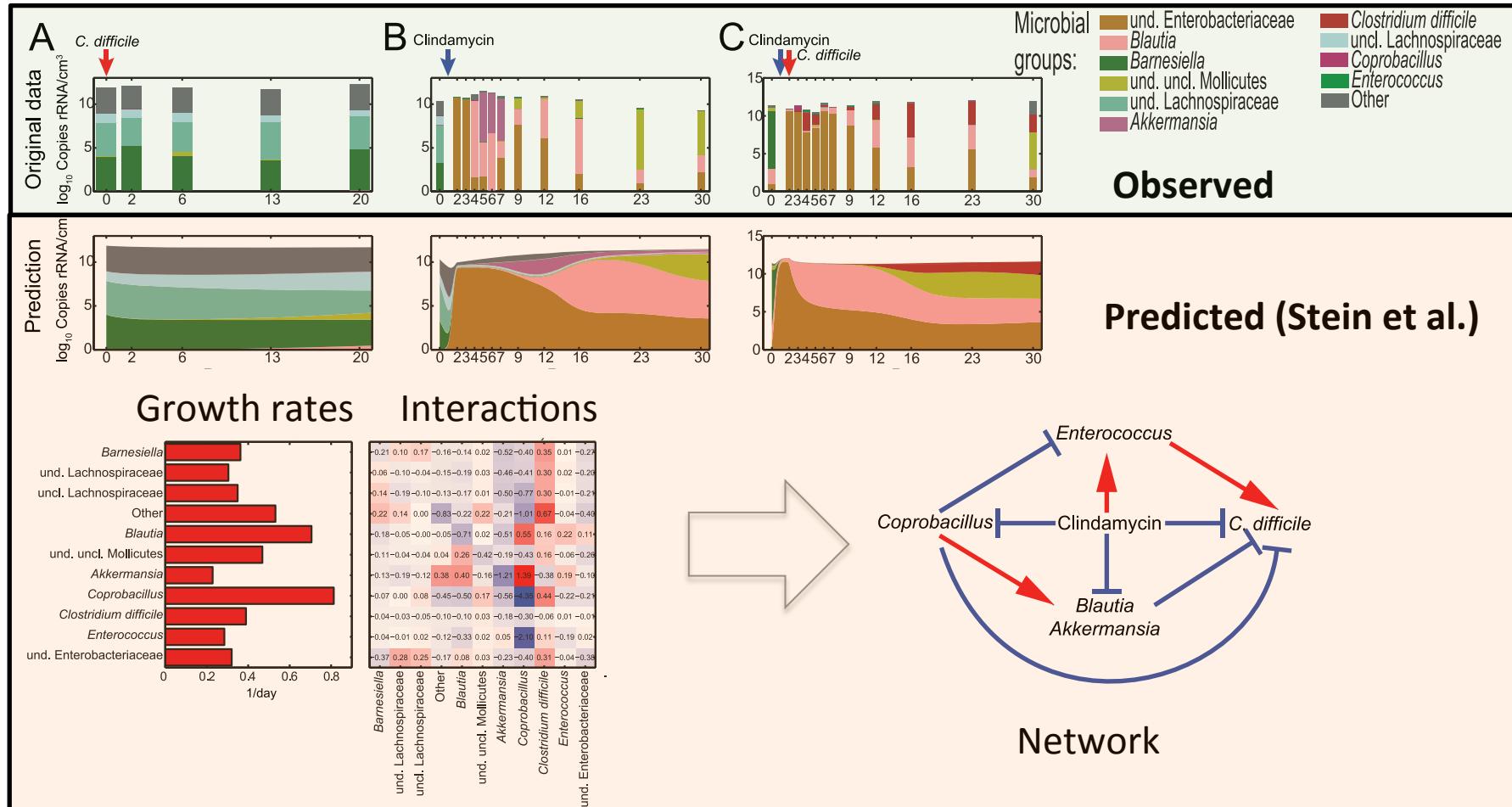
- Species abundance x_i can be modeled as a function of species i 's initial abundance, its growth rate b_i and its interactions a_{ij} with other species j

$$\frac{dx_i(t)}{dt} = x_i(t) \left(b_i + \sum_{j=1}^N a_{ij} x_j(t) \right)$$

Generalized
Lotka Volterra
(gLV)

Fitting gLV equations to 16S data - Example

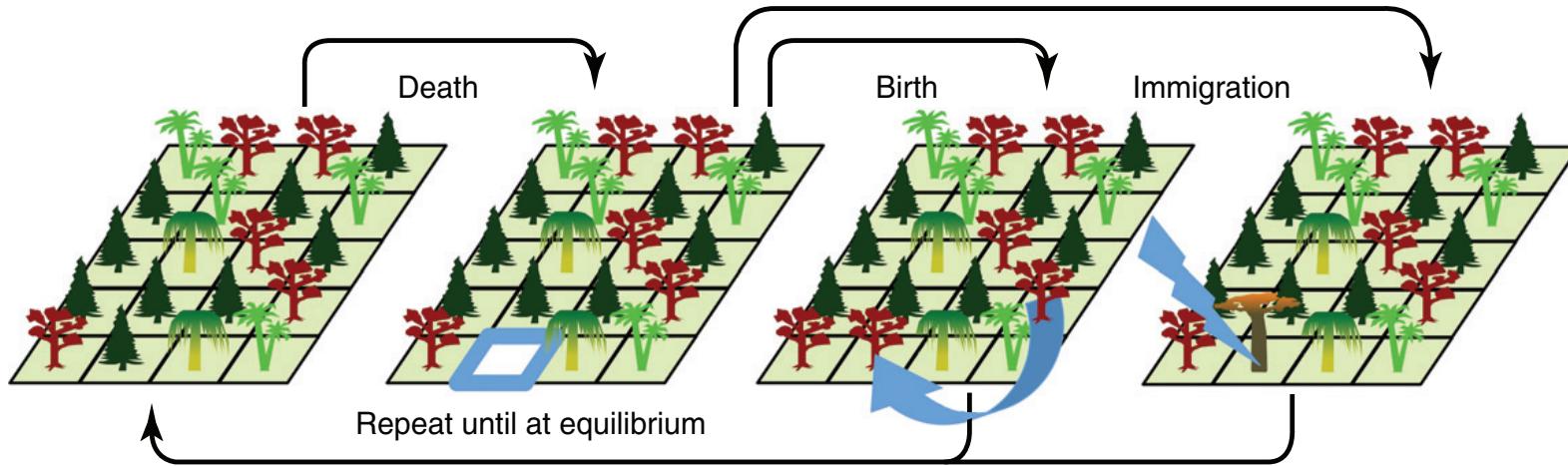
- Parameterization using sparse regression, but only few taxa
- Simulation of perturbation



Stein et al. (2013). Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Computational Biology* 9, e1003388.

Neutral view: the unified neutral theory of biodiversity (Hubbell)

“The neutral model”



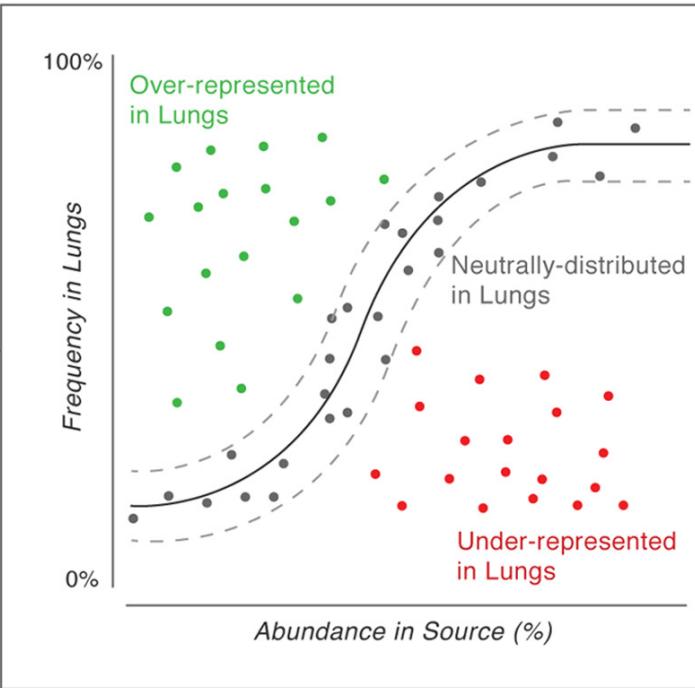
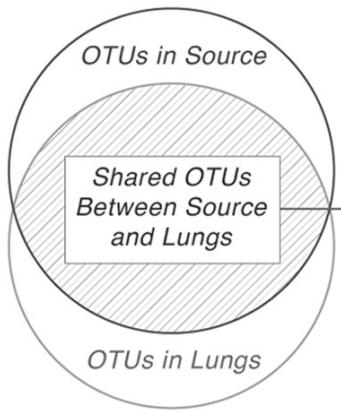
At each step, one individual dies and is replaced from the meta-community with probability m (**immigration**) or from the local community with probability $1-m$ (**birth**).
Parameter m : measure of dispersal limitation

Speciation occurs in the meta-community.

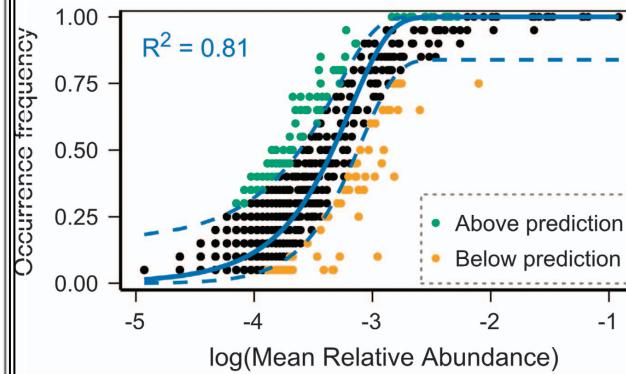
Fitting the neutral model to 16S data – Examples

Introduction

Taxa over-/under-represented in the lungs according to neutral model (Venkataraman et al.)



Taxa over-/under-represented in fish larvae guts according to the neutral model (Burns et al.)



Sloan et al. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology*, 8 (4), 732-740.

Venkataraman et al. (2015). Application of a Neutral Community Model To Assess Structuring of the Human Lung Microbiome. *mBio* 6(1):e02284-14.

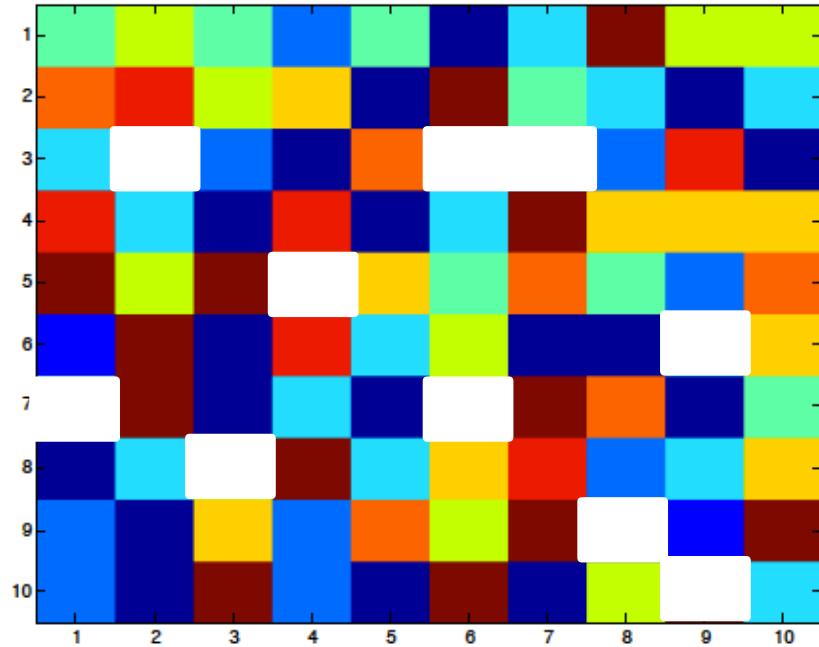
Burns et al. (2015). Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal*, advance online publication.

Combination of both: the self-organized instable (SOI) model

- Self-organized stable systems: "The basic idea is that large dynamical systems naturally evolve, or **self-organize**, into a highly interactive, **critical** state where a minor perturbation may lead to [disruptive] events [...]." (Bak & Paczuski)
- Solé: ecosystems do not satisfy all criteria of self-organized stable systems and are thus considered to be self-organized unstable
- Solé designed a **self-organized unstable** model that combines aspects of gLV and of the neutral model (model B, Solé et al. 2002)

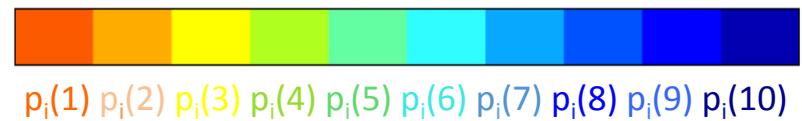
SOI model - Principle

Introduction

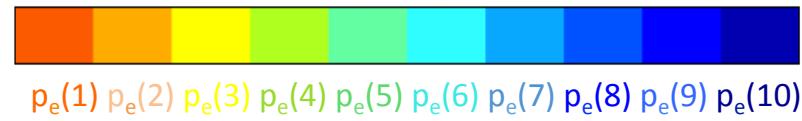


The SOI model includes **both immigration from the metacommunity and interactions between species.**

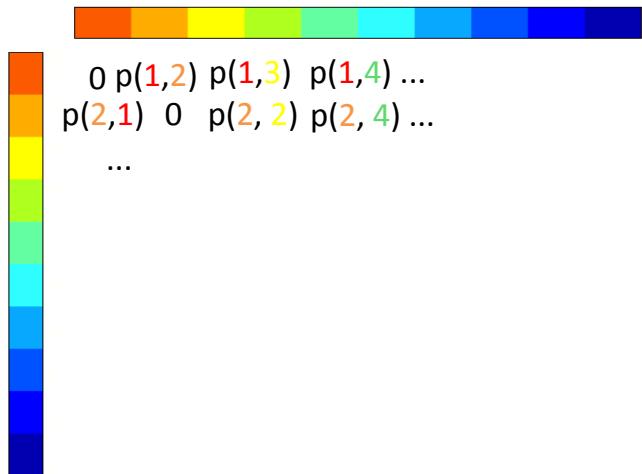
immigration probabilities



extinction probabilities

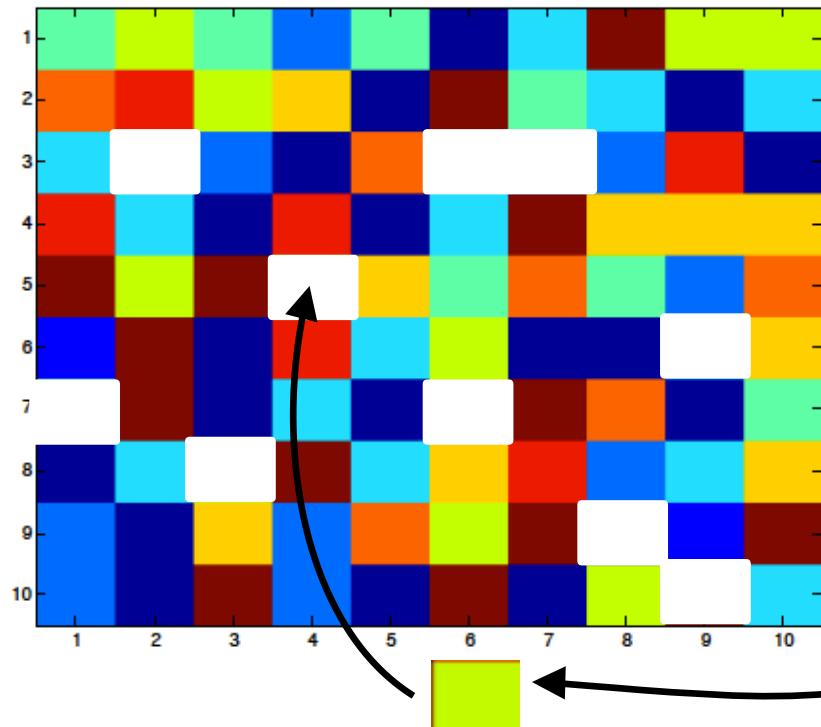


interaction matrix

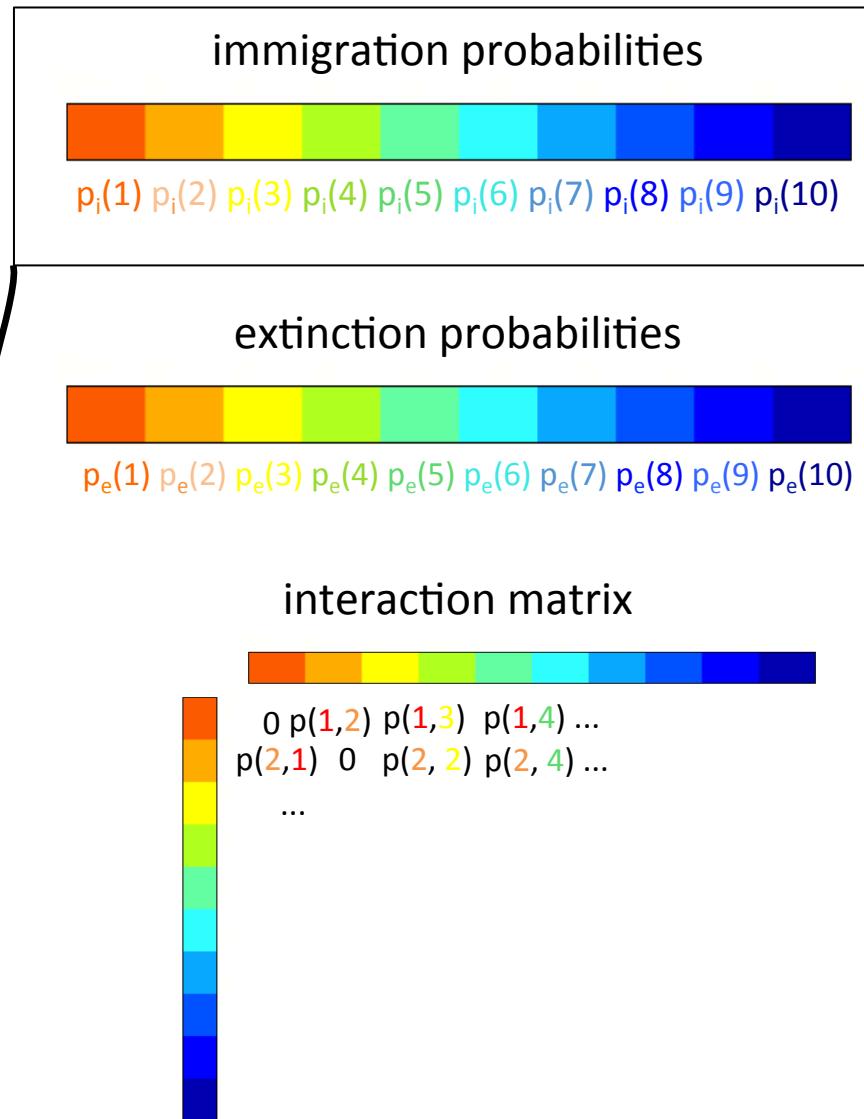


SOI model - Principle

Introduction

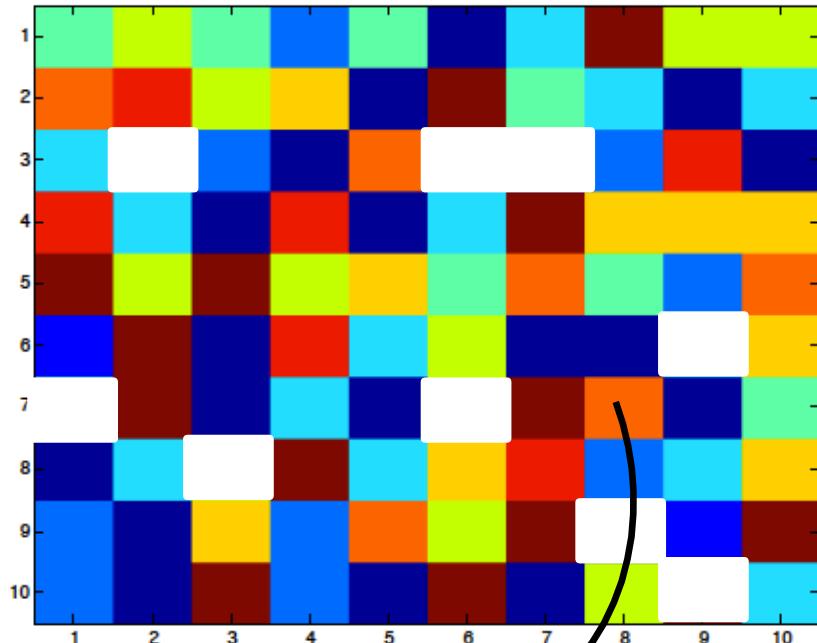


At each time step, for each site:
If the site is empty: is it filled by a species
from the metacommunity according to
the immigration probabilities (p_i)



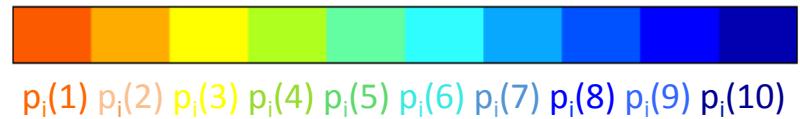
SOI model - Principle

Introduction

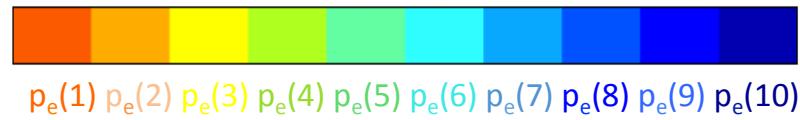


At each time step, for each site:
If the site is filled: the species can die
according to the extinction probability (p_e)

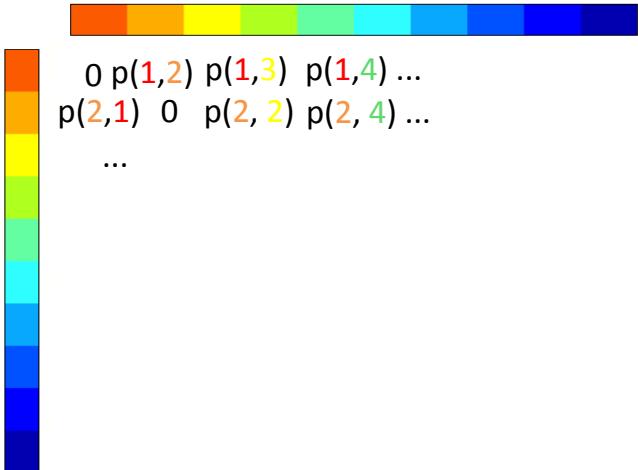
immigration probabilities



extinction probabilities

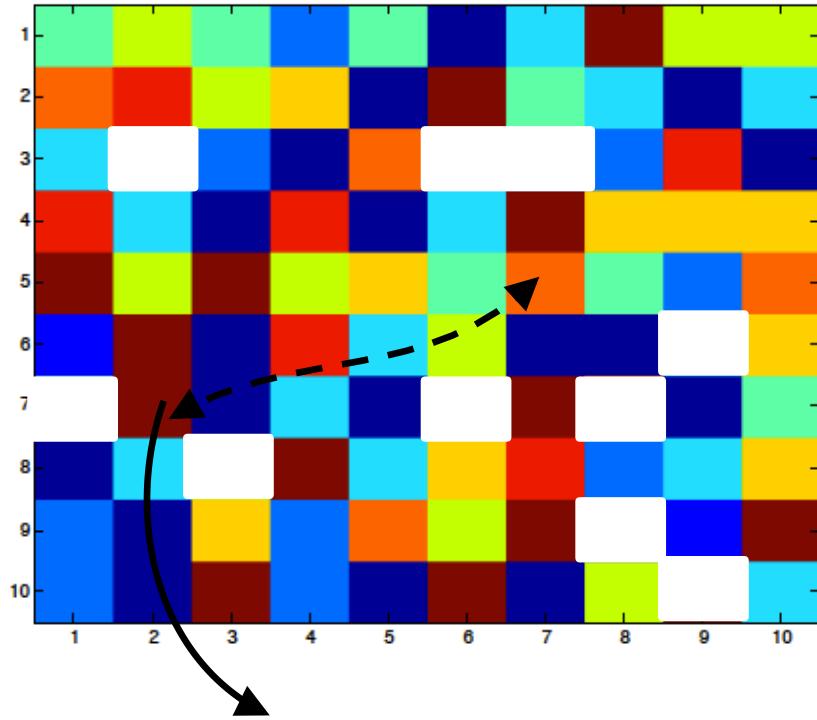


interaction matrix



SOI model - Principle

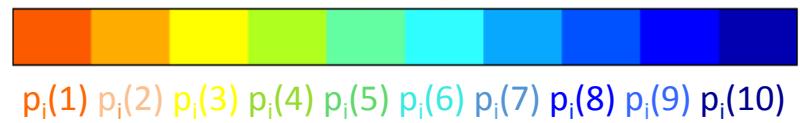
Introduction



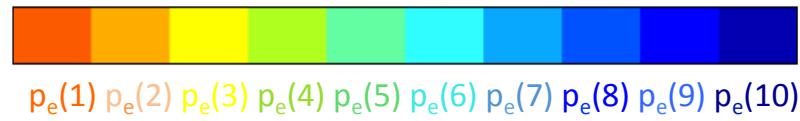
If the site is now occupied (species i) , an interaction can take place: Another site (species j) is randomly chosen and following the interaction matrix, several scenarios are possible:

- if $p(i,j)>0$ and $(j,i)>0$ (**mutualism**) -> growth of species i (added in a third, empty site)
- if $p(i,j)>0$ and $(j,i)<0$ (**parasitism**) -> j is replaced by i
- if $p(i,j)<0$ and $(j,i)>0$ (**parasitism**) -> i is replaced by j

immigration probabilities

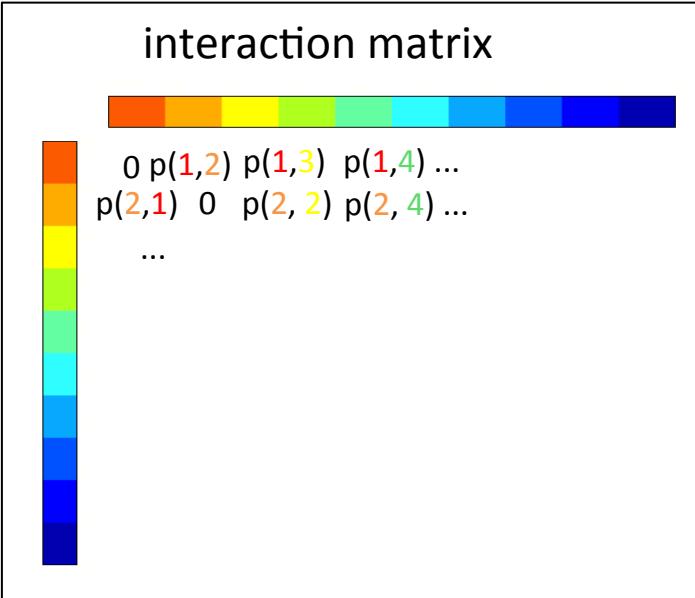


extinction probabilities



interaction matrix

0 $p(1,2)$ $p(1,3)$ $p(1,4)$...
 $p(2,1)$ 0 $p(2,2)$ $p(2,4)$...
...

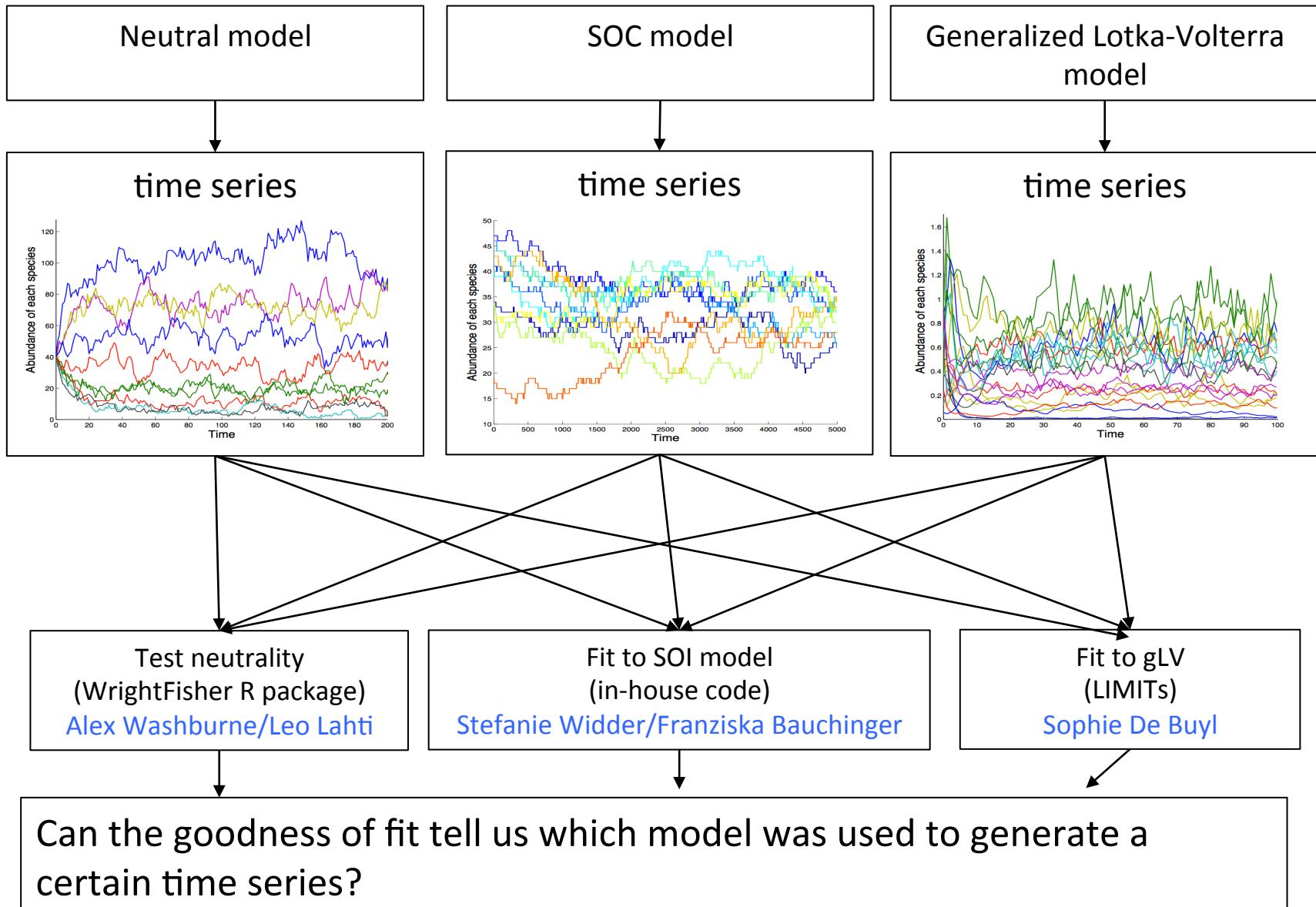


Motivation

- Given the time series of a community, can we determine the model that generated its dynamics?

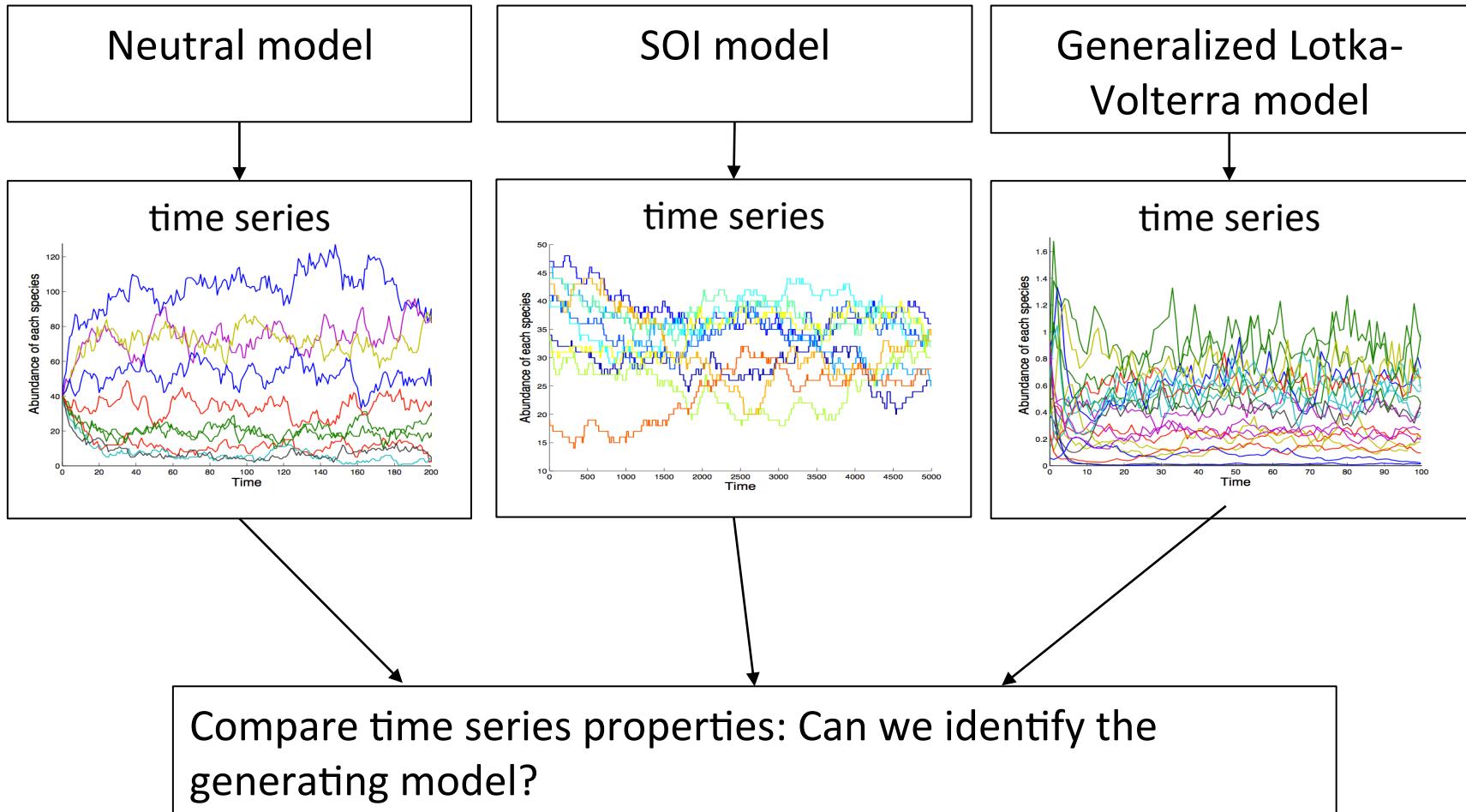
How to distinguish models at the time series level – methodology I

Methods



How to distinguish models at the time series level – methodology II

Methods



Test Time Series

- 50 test time series, each with 100 species sampled for 3000 time points
- Initial abundances generated with stick-breaking process (`bstick` in `vegan`)
- Time series sub-sets of 100 time points also considered

Generator	Parameters
Generalized Lotka-Volterra (gLV)	Growth rates, interaction matrix (A)
Discrete generalized Lotka-Volterra (Ricker)	Carrying capacities, interaction matrix (A), noise term (sigma)
Neutral model (Hubbell)	Number of individuals, immigration rate (m), death rate
SOI model	Number of individuals, interaction matrix (A), immigration rates, extinction rates
Dirichlet-Multinomial distribution (DM)	Species proportions, overdispersion (theta)
David et al., stool data individual A	-
David et al., stool data individual B	-

Ricker: Fisher and Mehta (2014). Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries using Sparse Linear Regression. *PLoS one* 9, e102451.

Parameter values

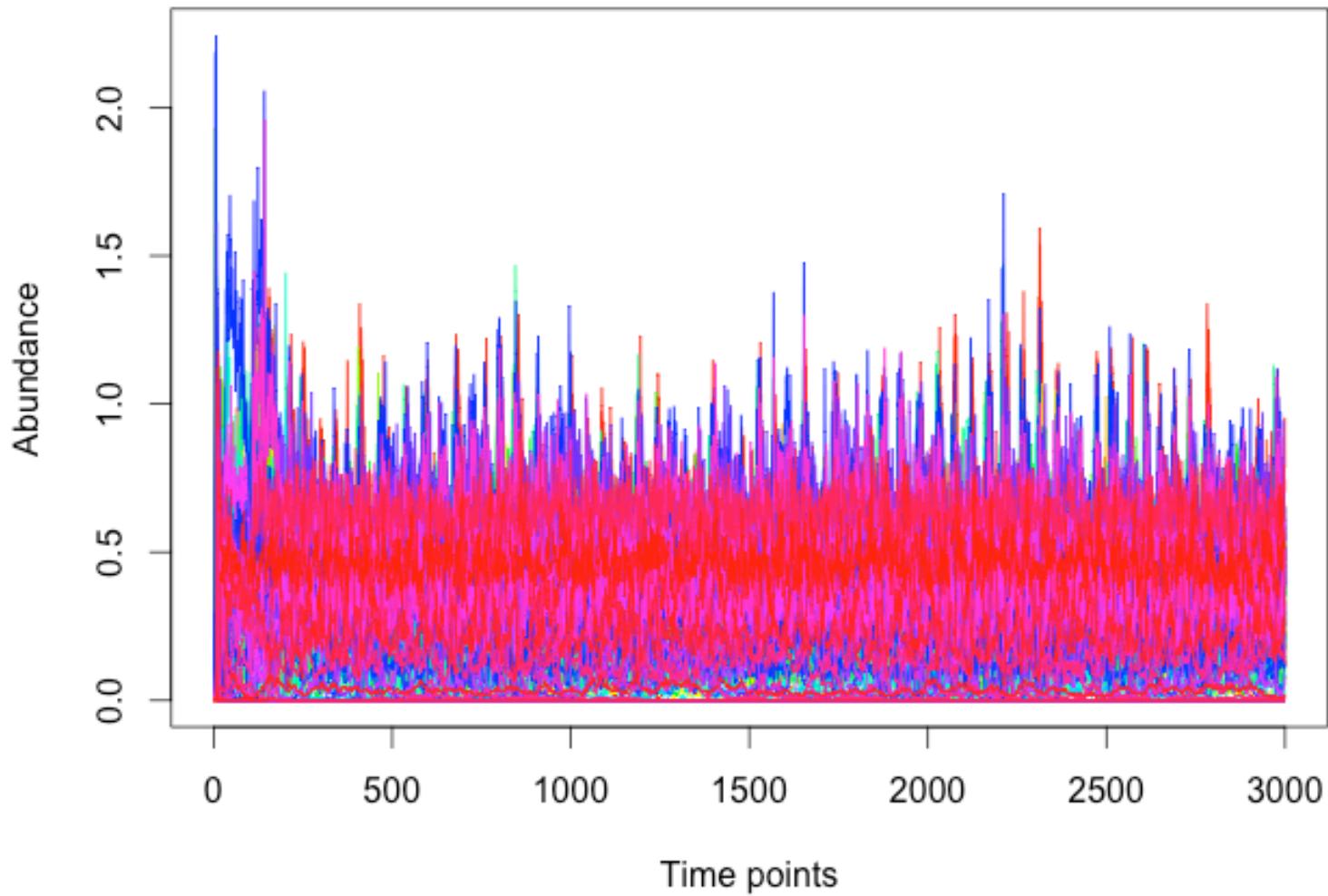
Parameter	Values
Growth rates/Carrying capacities	Uniform distribution between 0 and 0.5
Noise term sigma	0, 0.01, 0.05 or 0.1
Interaction matrix A	Structure generated with algorithm by Klemm & Eguiluz to be modular and scale-free, diagonal values set to -1, off-diagonal values sampled from uniform distribution between 0 and 1, positive edge percentage values: 0, 16, 40 or 64, connectance values: 0.01 or 0.05
Immigration rate m	0.1, 0.3, 0.5, 0.9 or 0.98
Death rate	10, 100 or 1000
Number of individuals	1500 or 3000
Immigration rates/Proportions	Initial abundances
Extinction rates	Uniform distribution between 0 and 1
Overdispersion theta	0.002, 0.02 or 0.2
Sampling interval	1, 5 or 10

Klemm & Eguiluz (2001). Growing scale-free networks with small-world behavior. *Physical Review E*, 65:057102.

Ricker time series

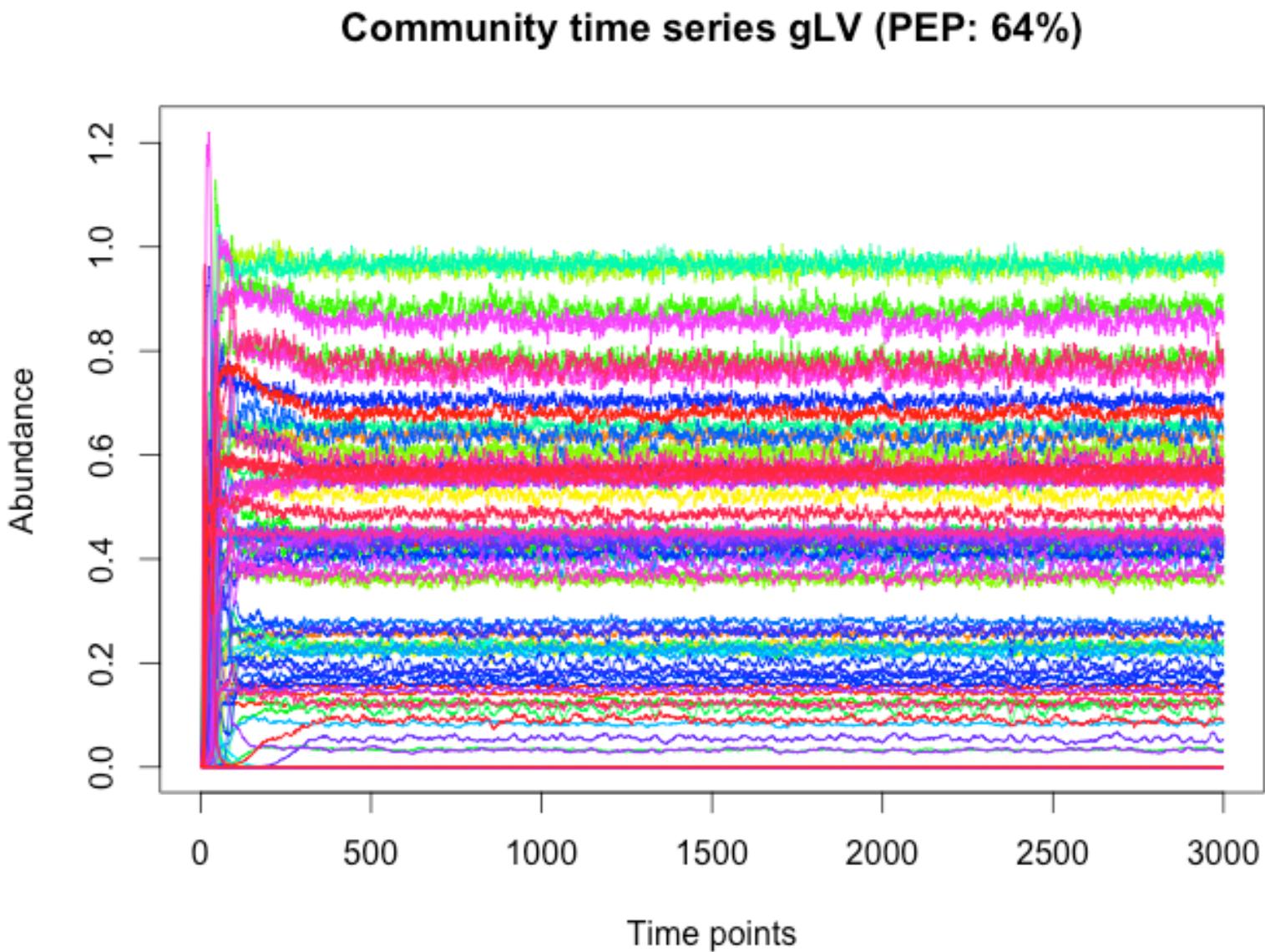
Results

Community time series gLV (PEP: 30%)



Ricker time series

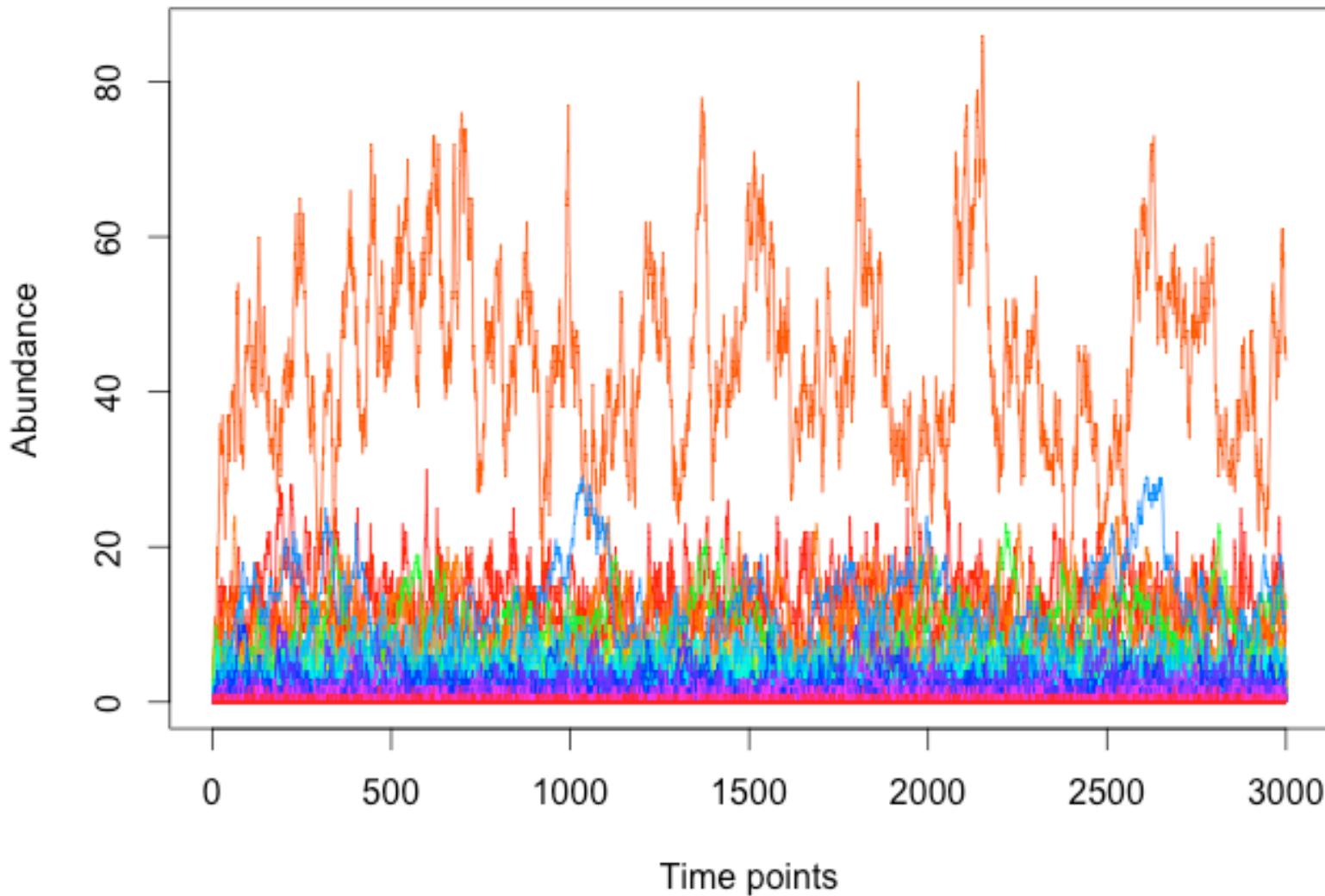
Results



SOI time series

Results

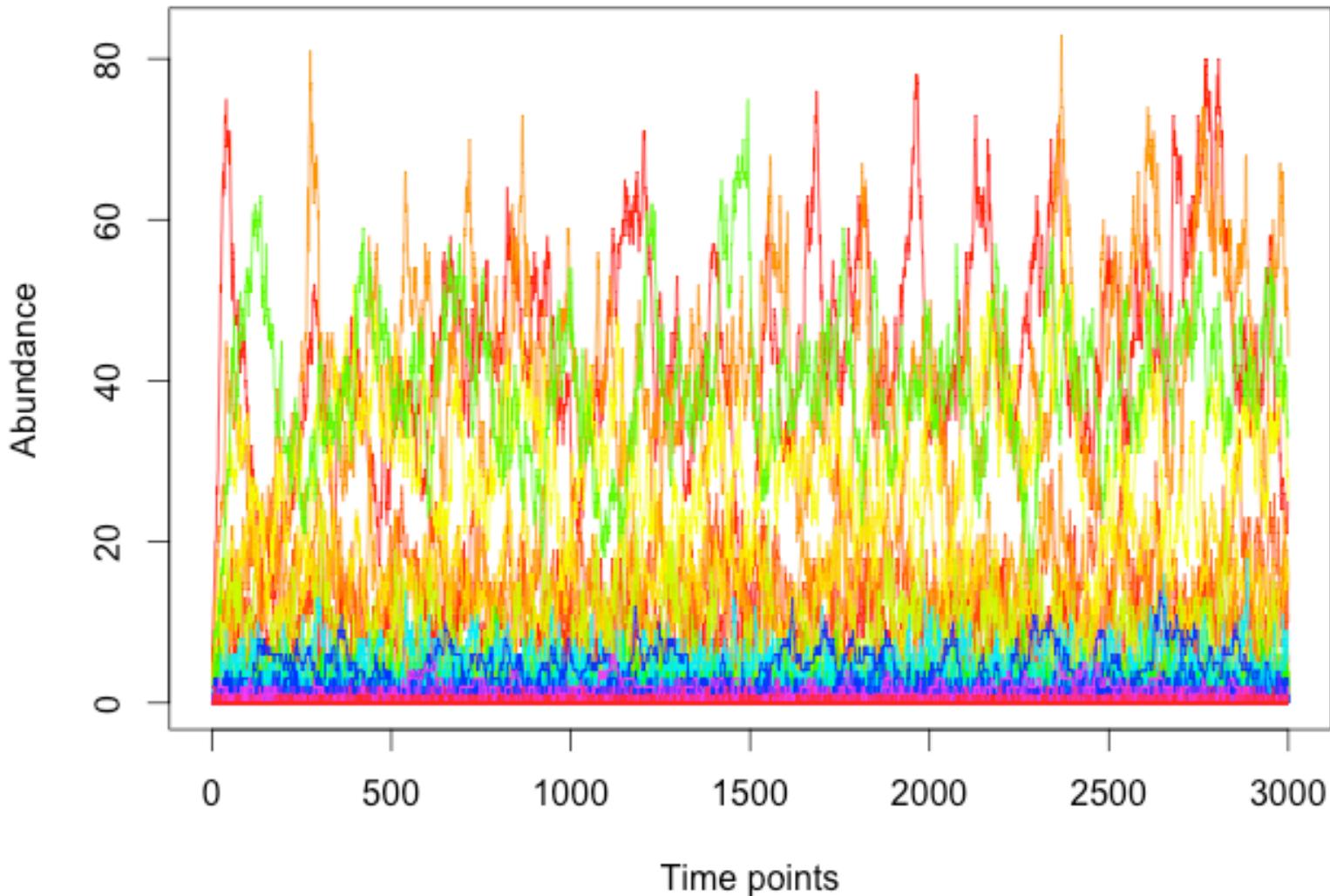
Community time series SOI (PEP: 30%)



Results

SOI time series

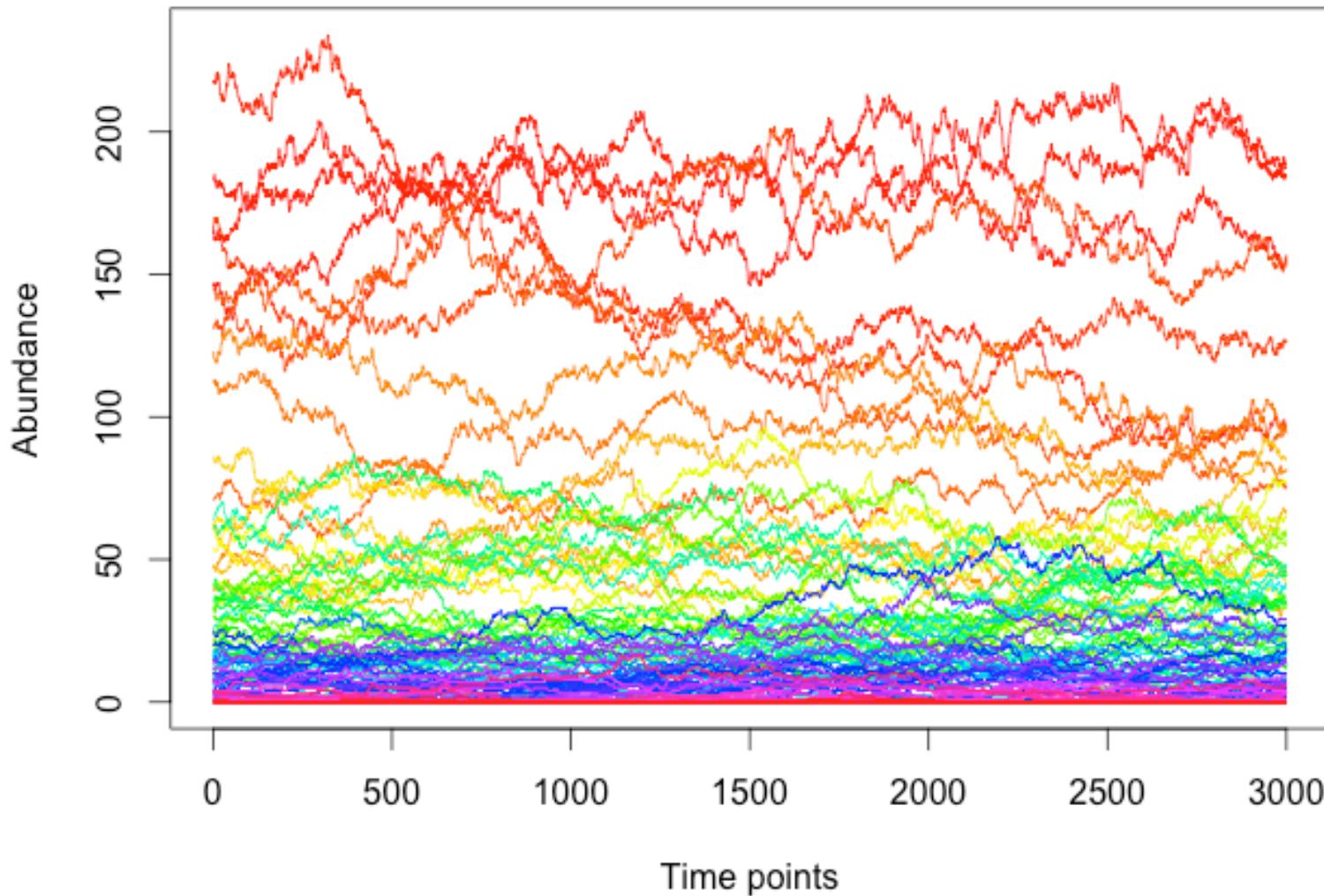
Community time series SOI (PEP: 64%)



Neutral model time series

Results

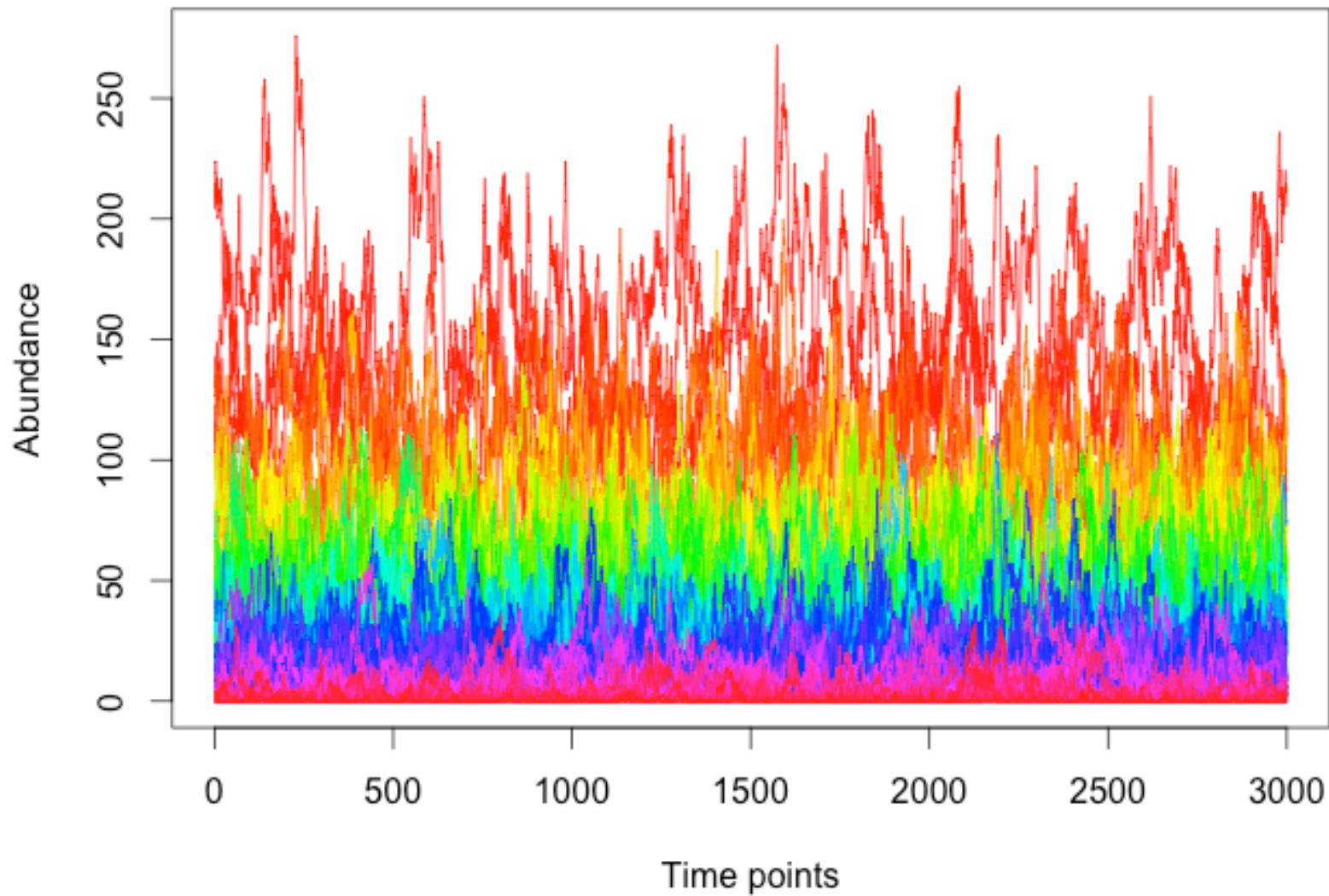
Community time series Hubbell ($m=0.1, d=10$)



Neutral model time series

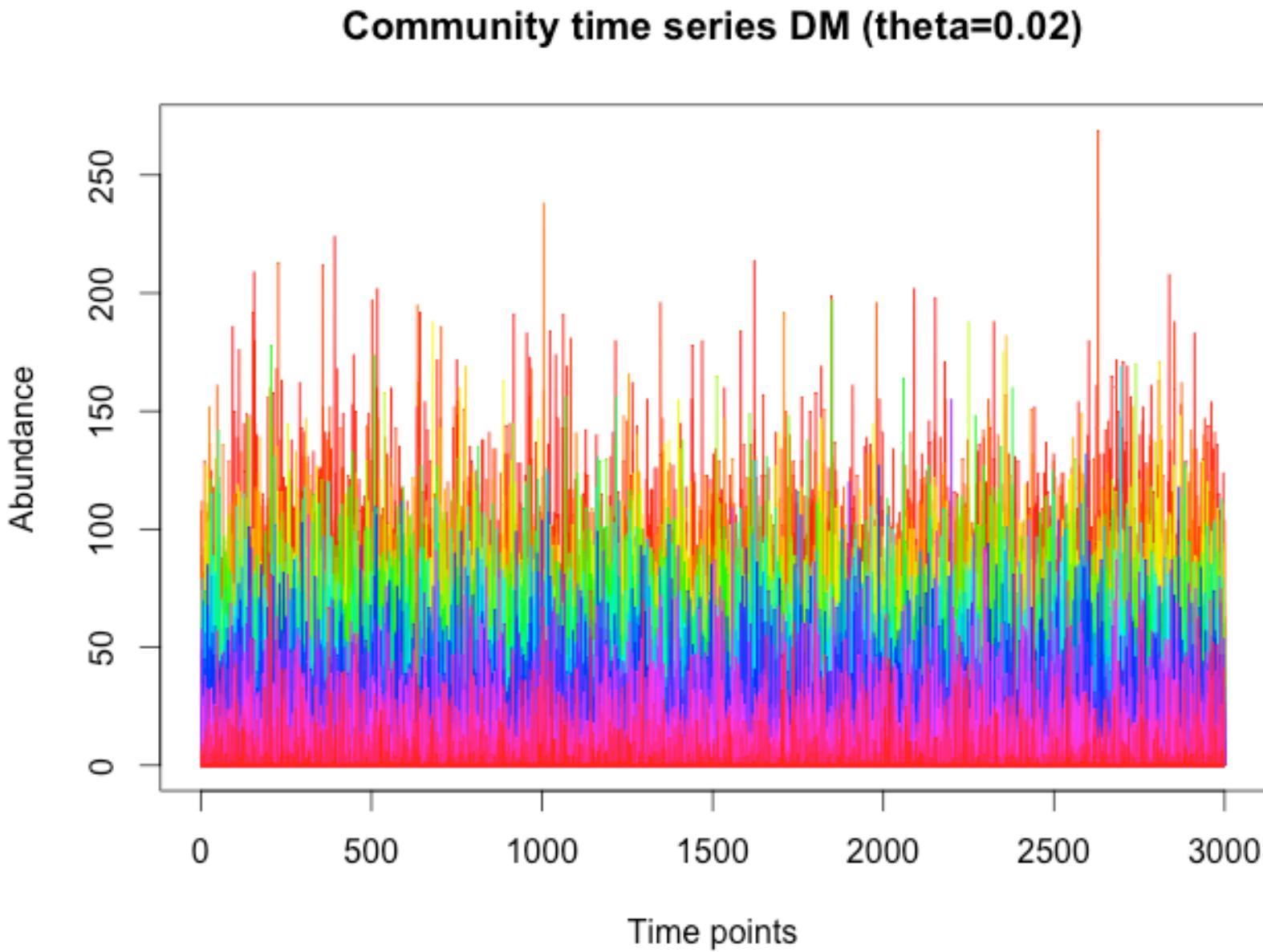
Results

Community time series Hubbell ($m=0.1, d=1000$)



Dirichlet-Multinomial distribution

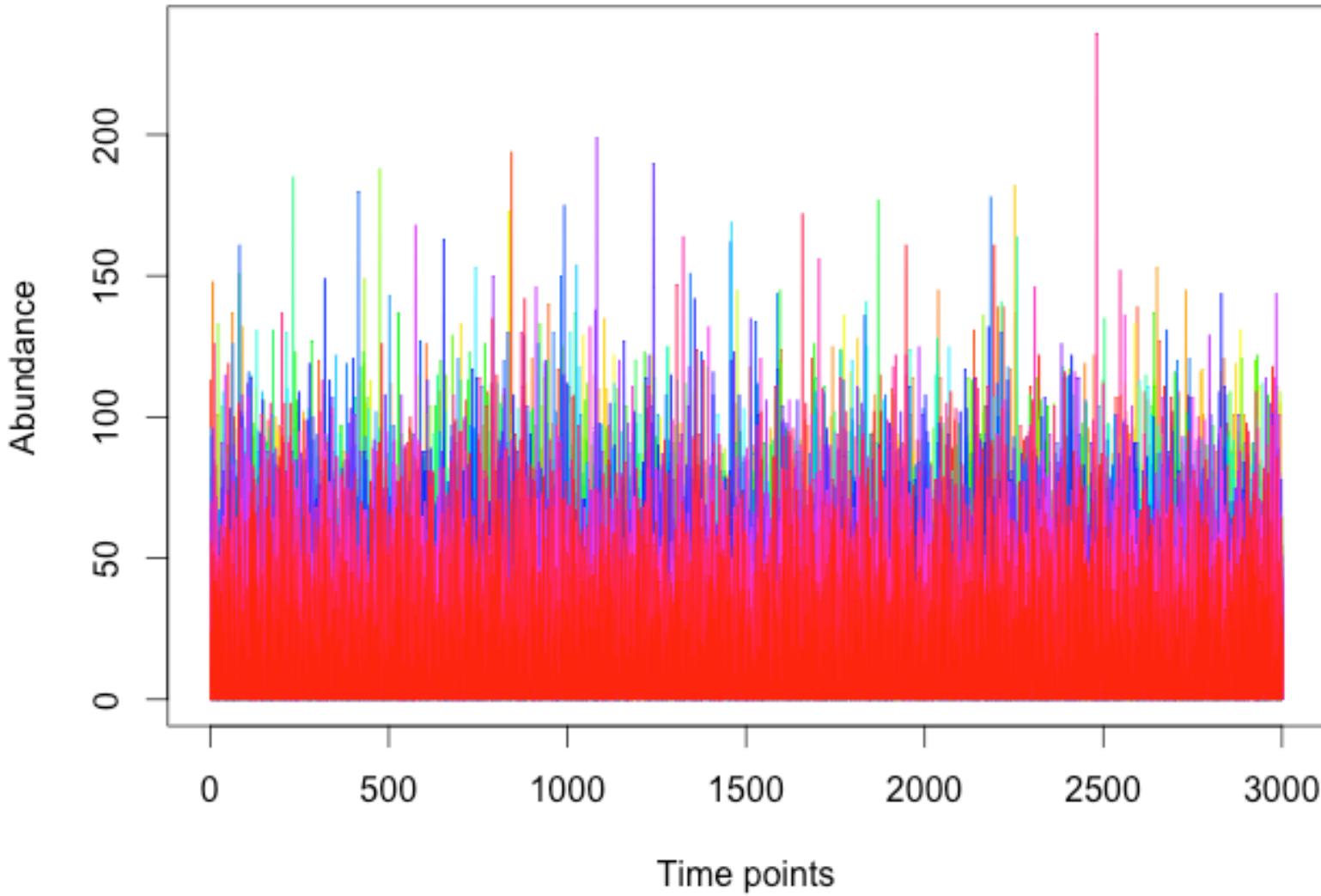
Results



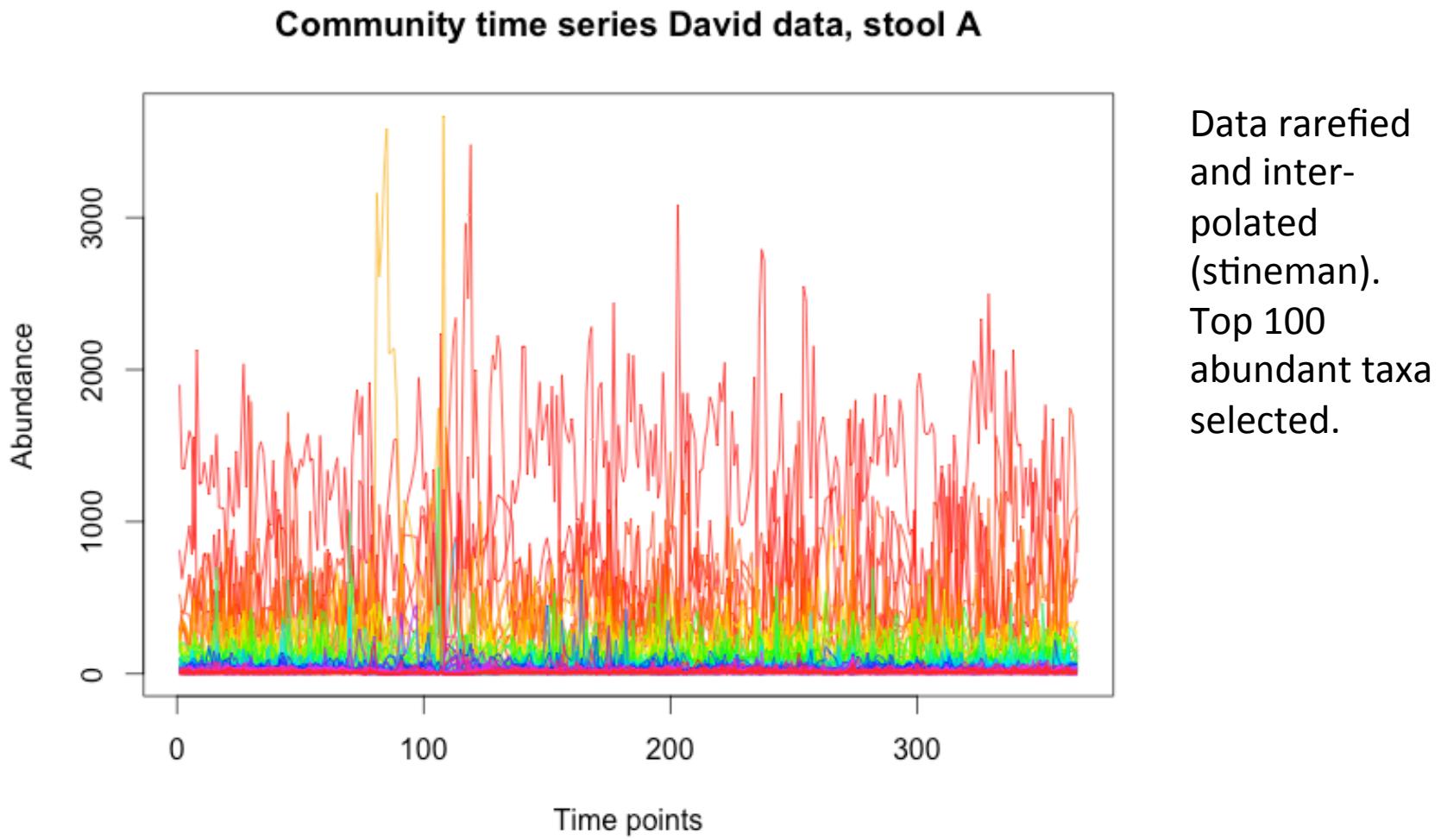
Dirichlet-Multinomial distribution

Results

Community time series DM (theta=0.02, even proportions)

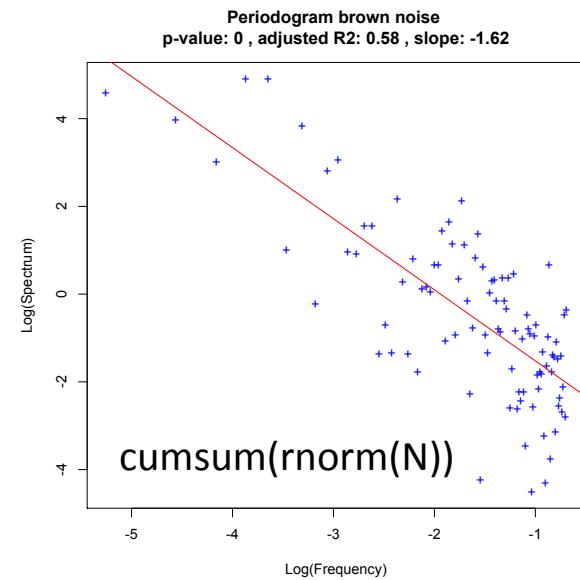
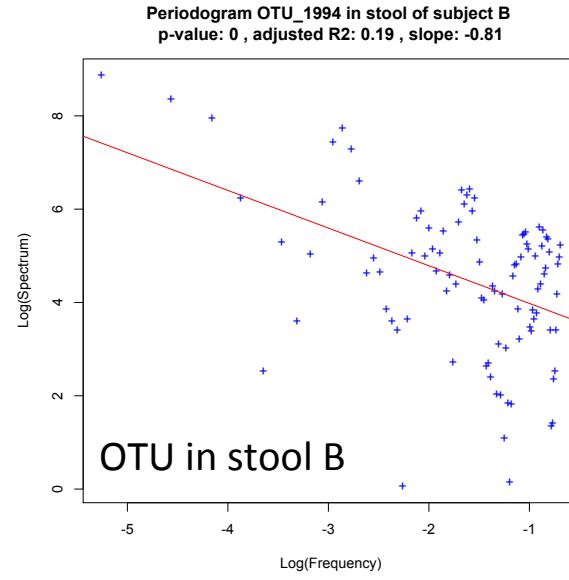
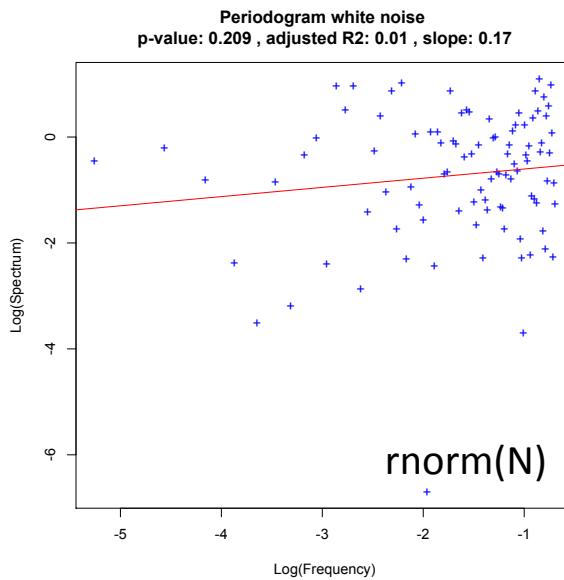


Results



Distinguishing models with noise types

- Power spectrum exponent: the exponent of the power law of spectral density versus frequency
- Power spectrum exponent (slope) indicates the noise type
- Slopes below -2: black noise



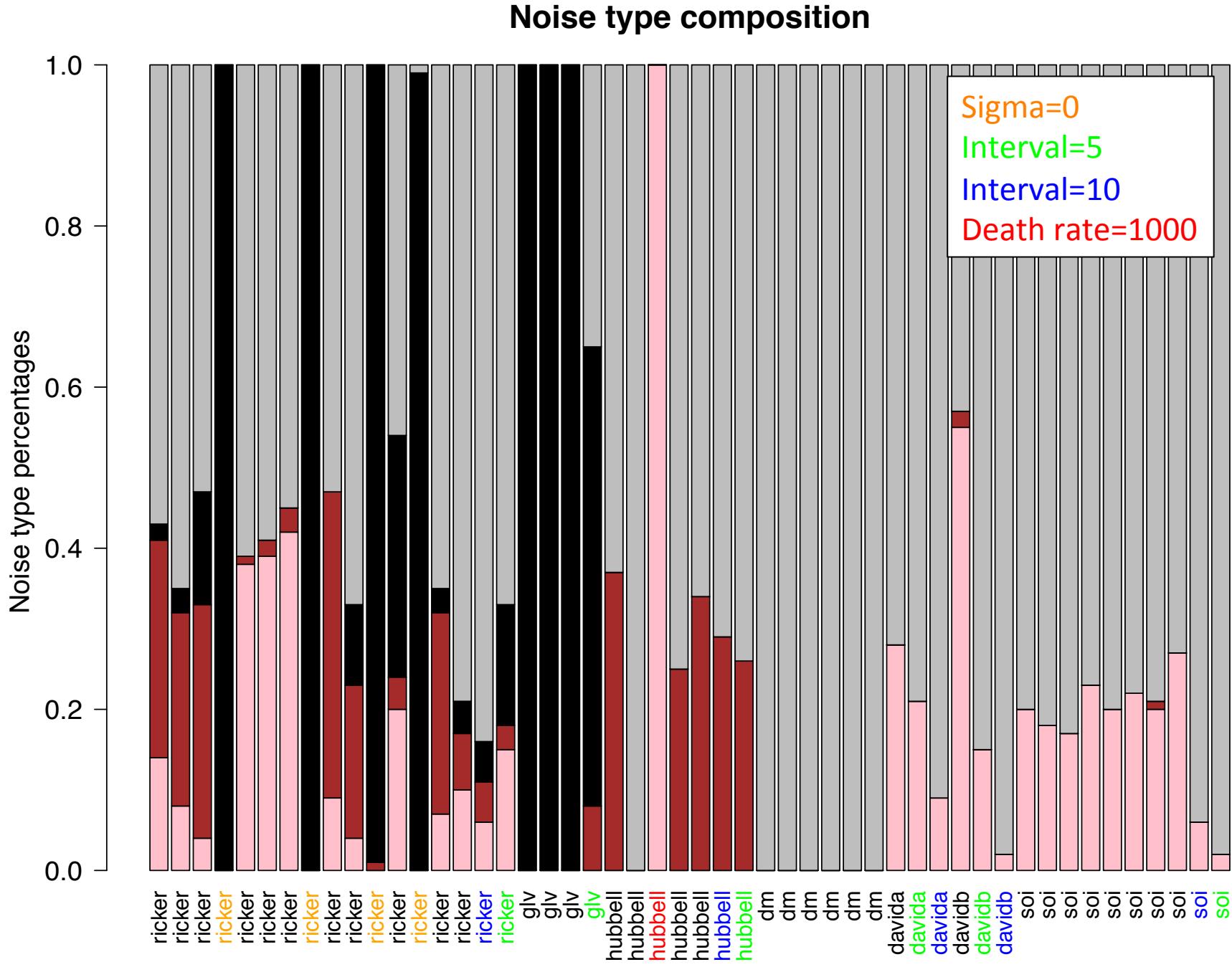
Methods

White noise
(Gaussian noise)
No memory
Slope around 0

Pink noise (1/f noise)
Memory effect
Slope around -1

Brown (red) noise
(Brownian motion)
Longer-term memory effect
Slope around -2

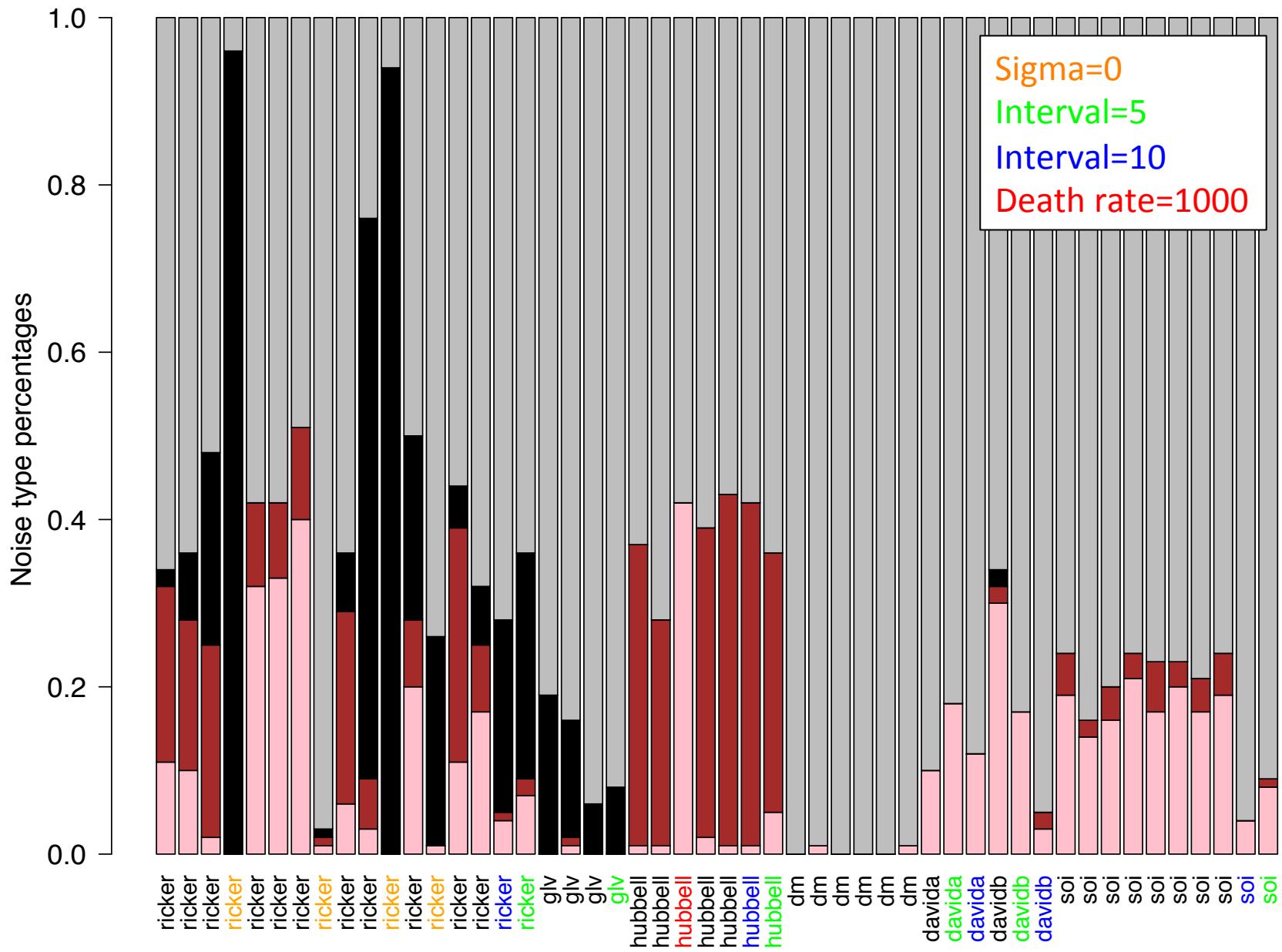
Results



Sub-set of 100 time points

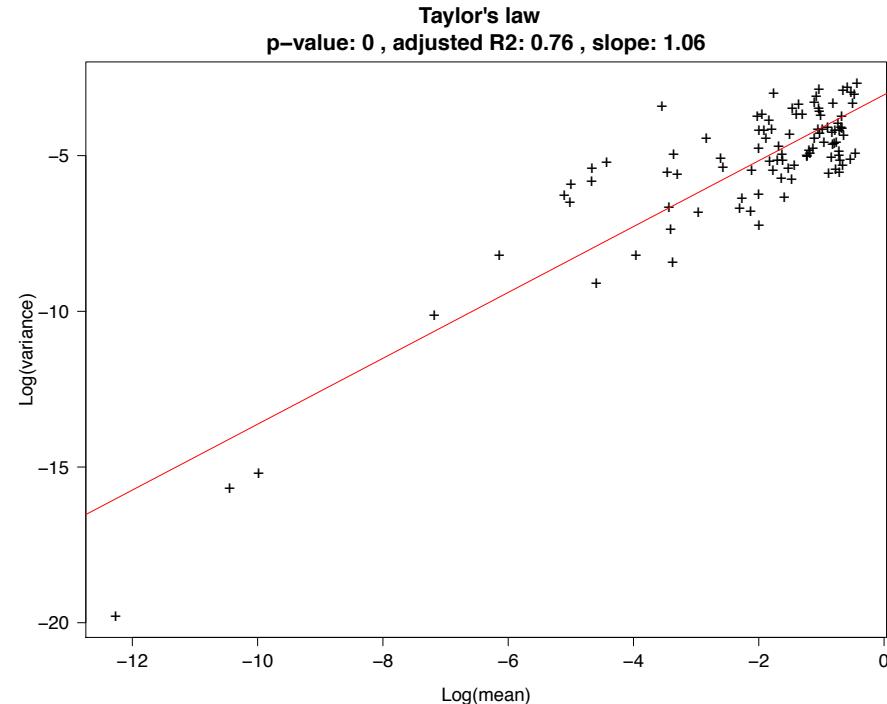
Noise type composition

Results



Taylor's law

- Power law of variance versus mean in species abundances (also known as fluctuation scaling)
- Slope was observed to range within 0.5 and 2 for ecological time series
- Interpretations of Taylor's law:
 - Lower slope due to negative species interactions (Kilpatrick & Ives)
 - A sign of self-organized critical systems (Kendal)
 - Consequence of random sampling from skewed distributions (Cohen & Xu)



Taylor (1961). Aggregation, variance and the mean. *Nature* 189, 732-735.

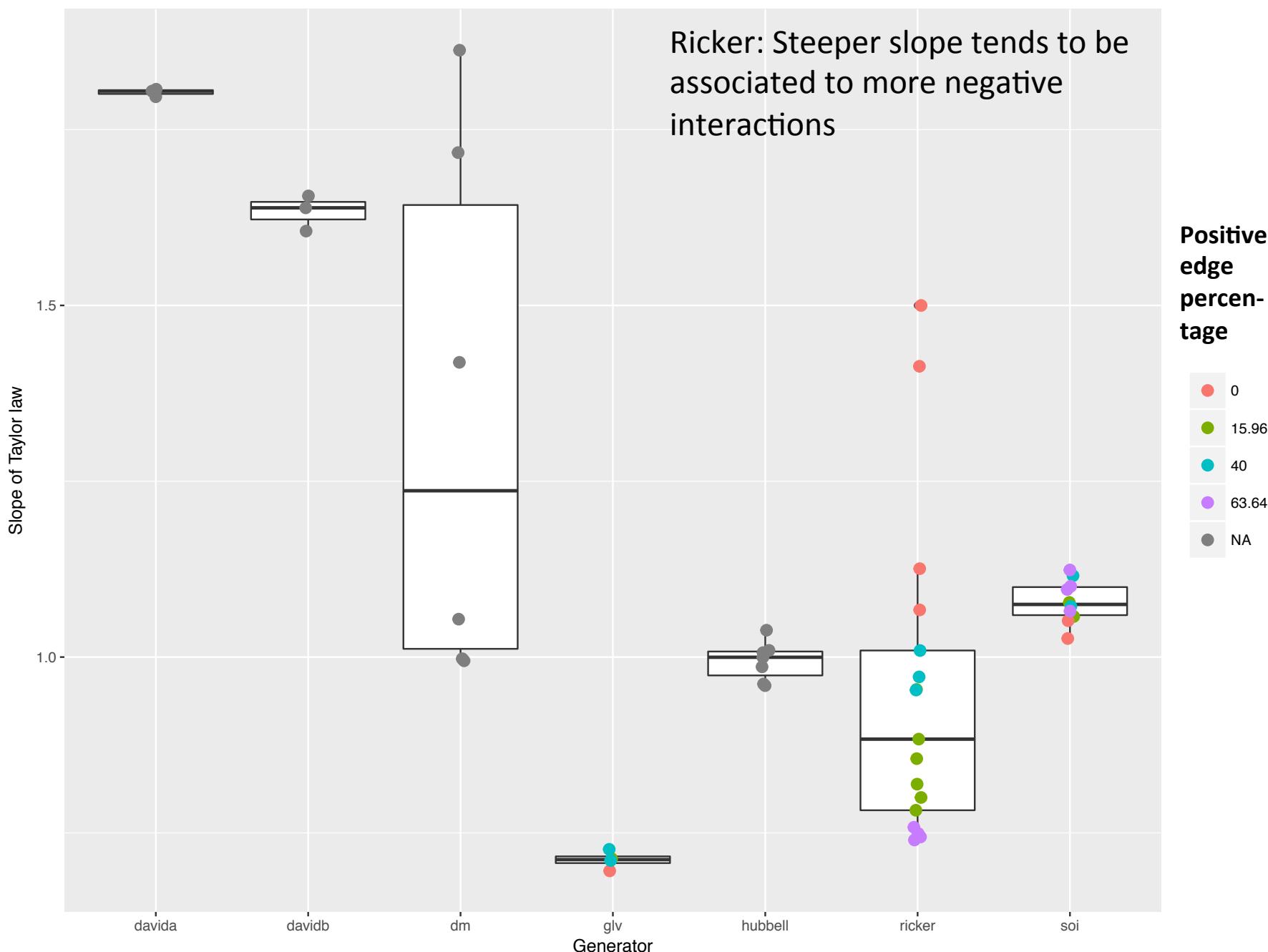
Kilpatrick & Ives (2003). Species interactions can explain Taylor's power law for ecological time series. *Nature* 422:65-68.

Kendal (2013). Fluctuation Scaling and 1/f Noise. *Journal of Basic and Applied Physics* 2:40-49.

Cohen & Xu (2015). Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. *PNAS* 112 (25): 7749-7754.

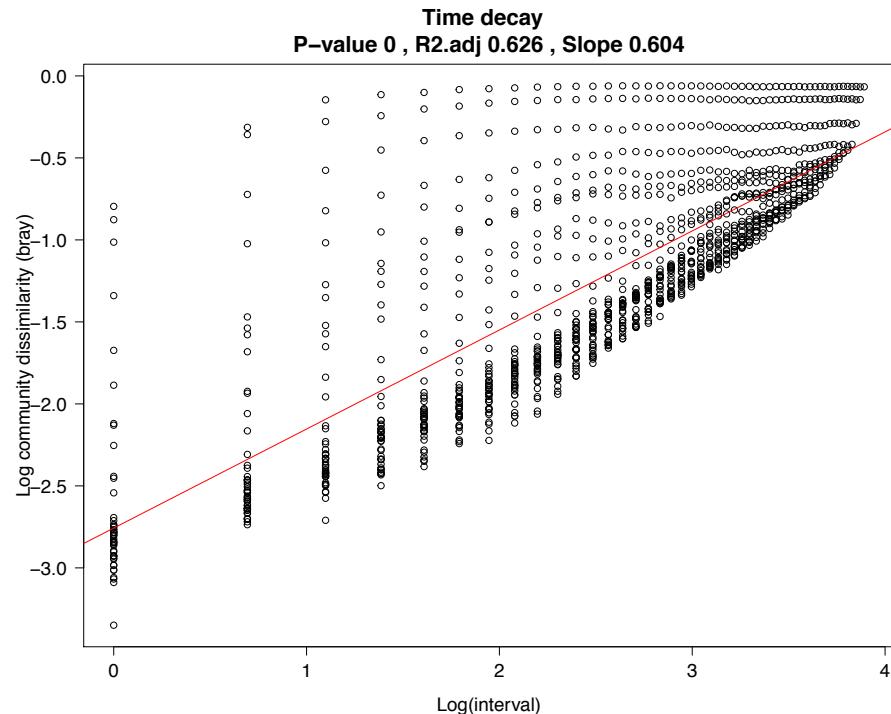
Generator versus slope of Taylor law colored by PEP

Results



Time decay

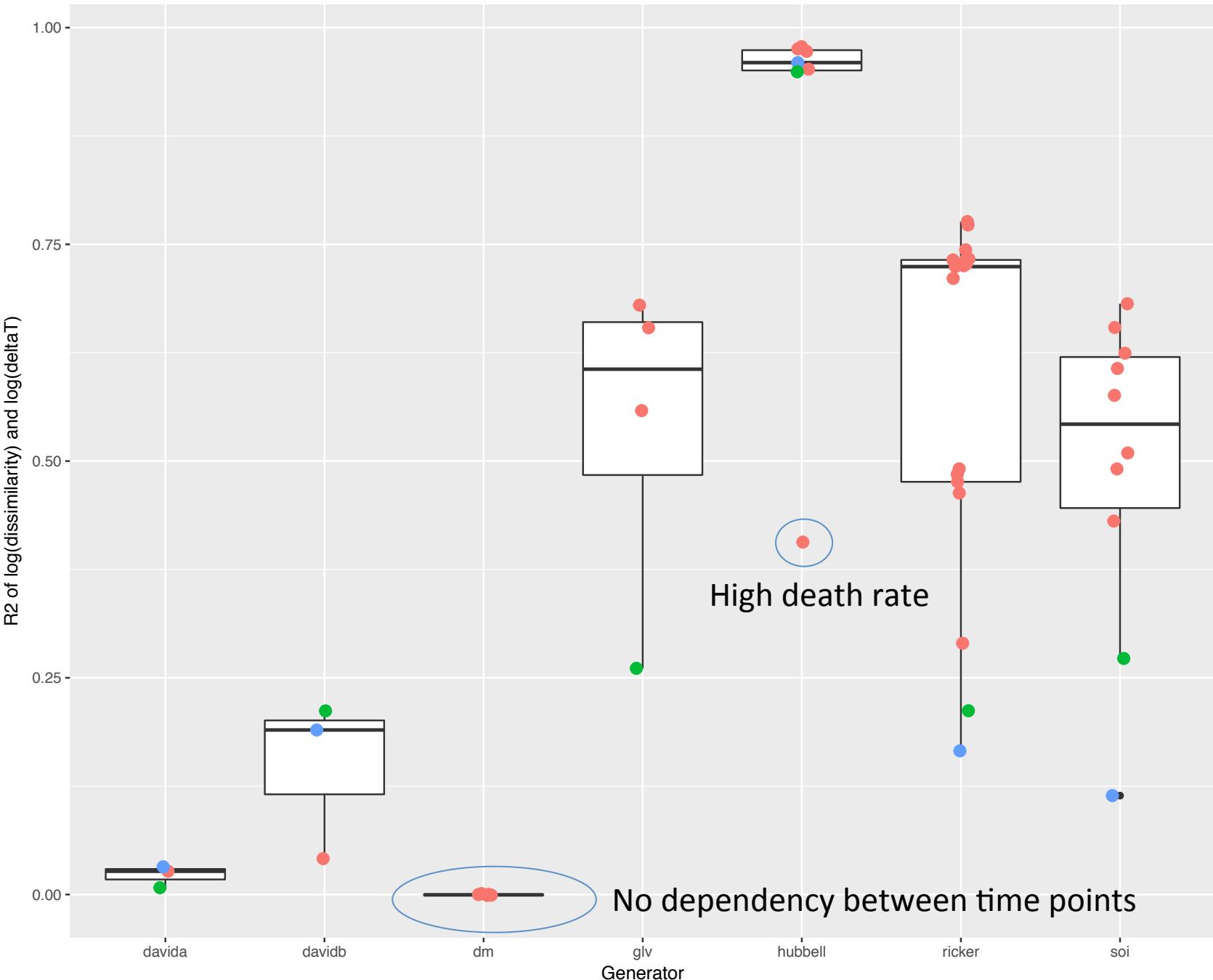
- Test whether community dissimilarity increases with time
- Compute dissimilarity of all time point pairs and plot dissimilarity versus time elapsed



Time decay computed for the Ricker model, first 50 time points

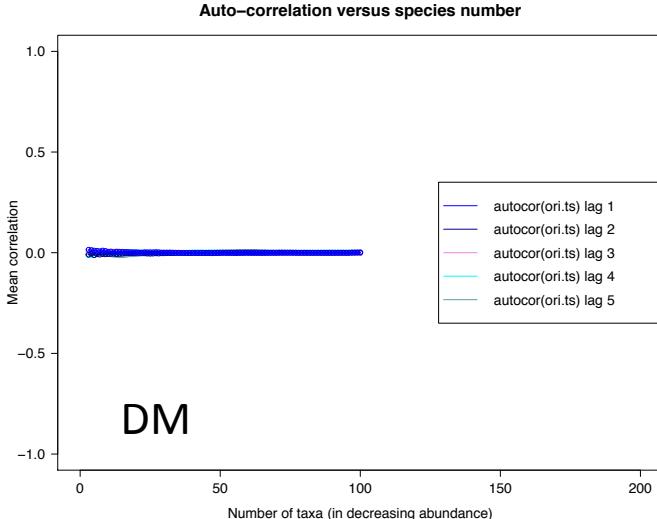
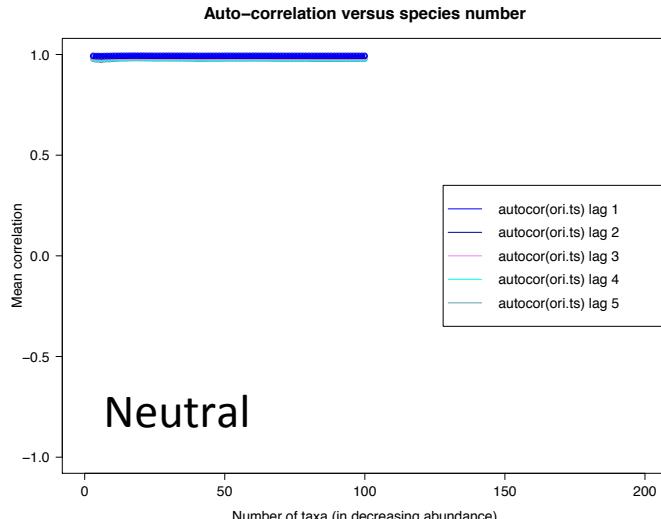
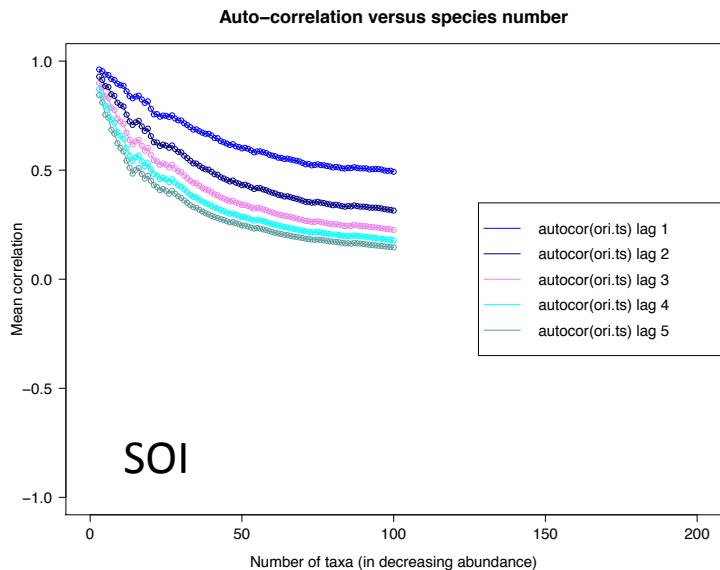
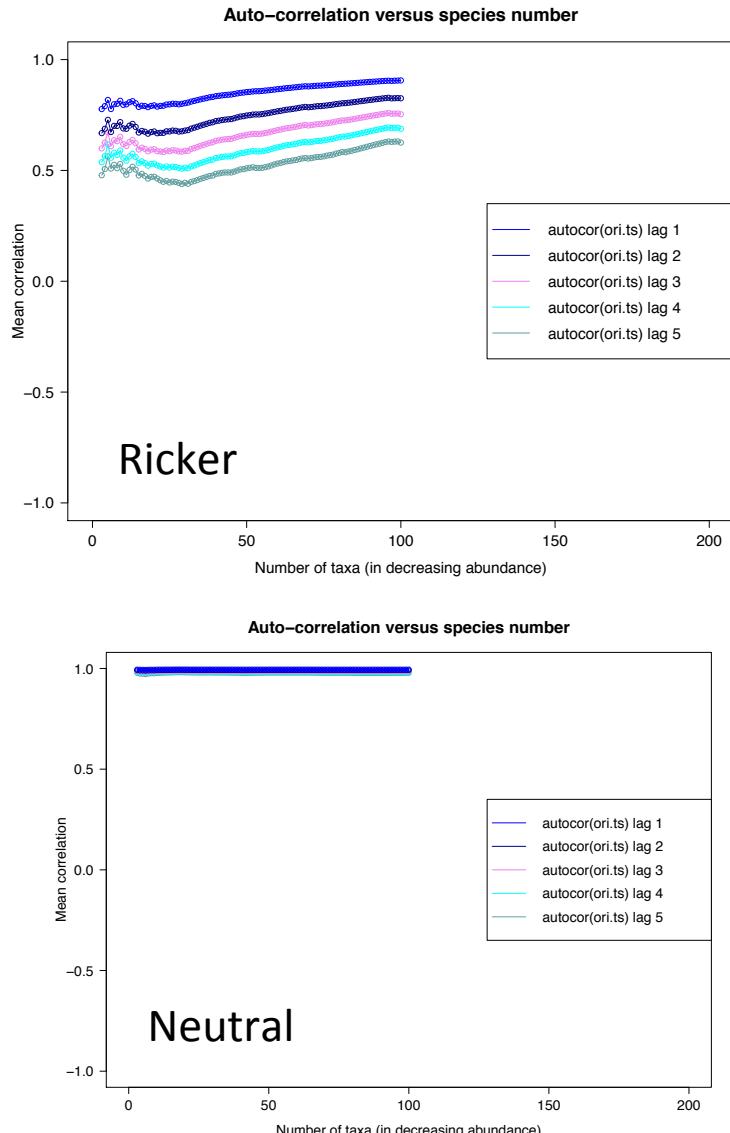
Generator versus R2 of log(dissimilarity) and log(deltaT) colored by interval

Results

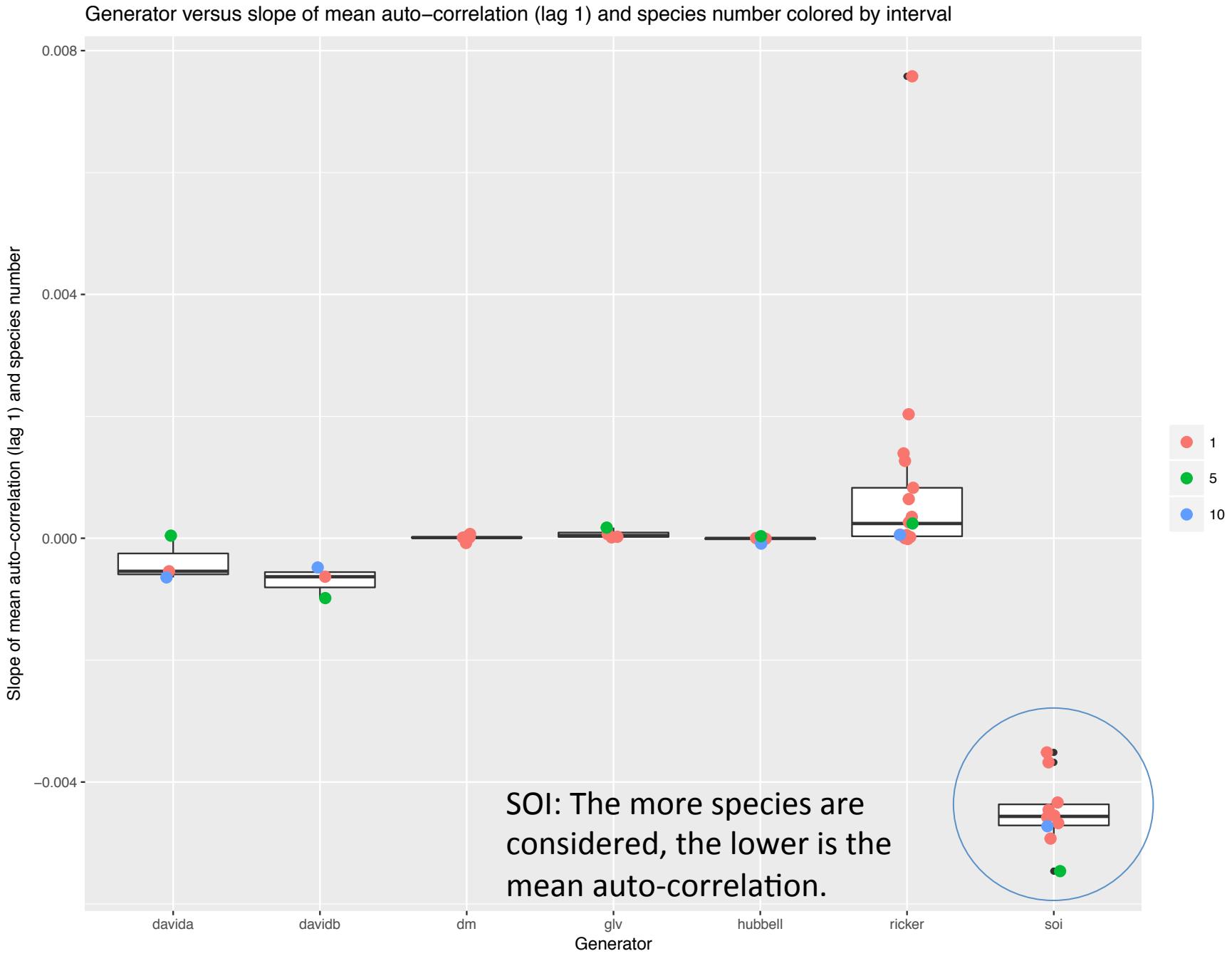


Auto-correlation versus number of species considered

Results



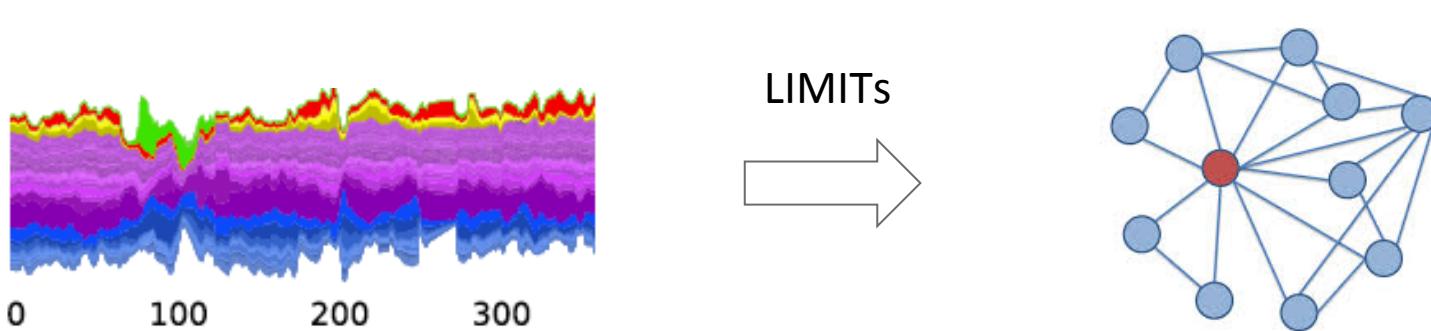
Results



LIMITs fitting

- LIMITs: developed by Fisher & Mehta to obtain parameters of Ricker model from time series (i.e. interaction matrix)
- Applied to all 50 time series
- Goodness of fit quantified by the mean correlation of the known with the predicted interaction matrix

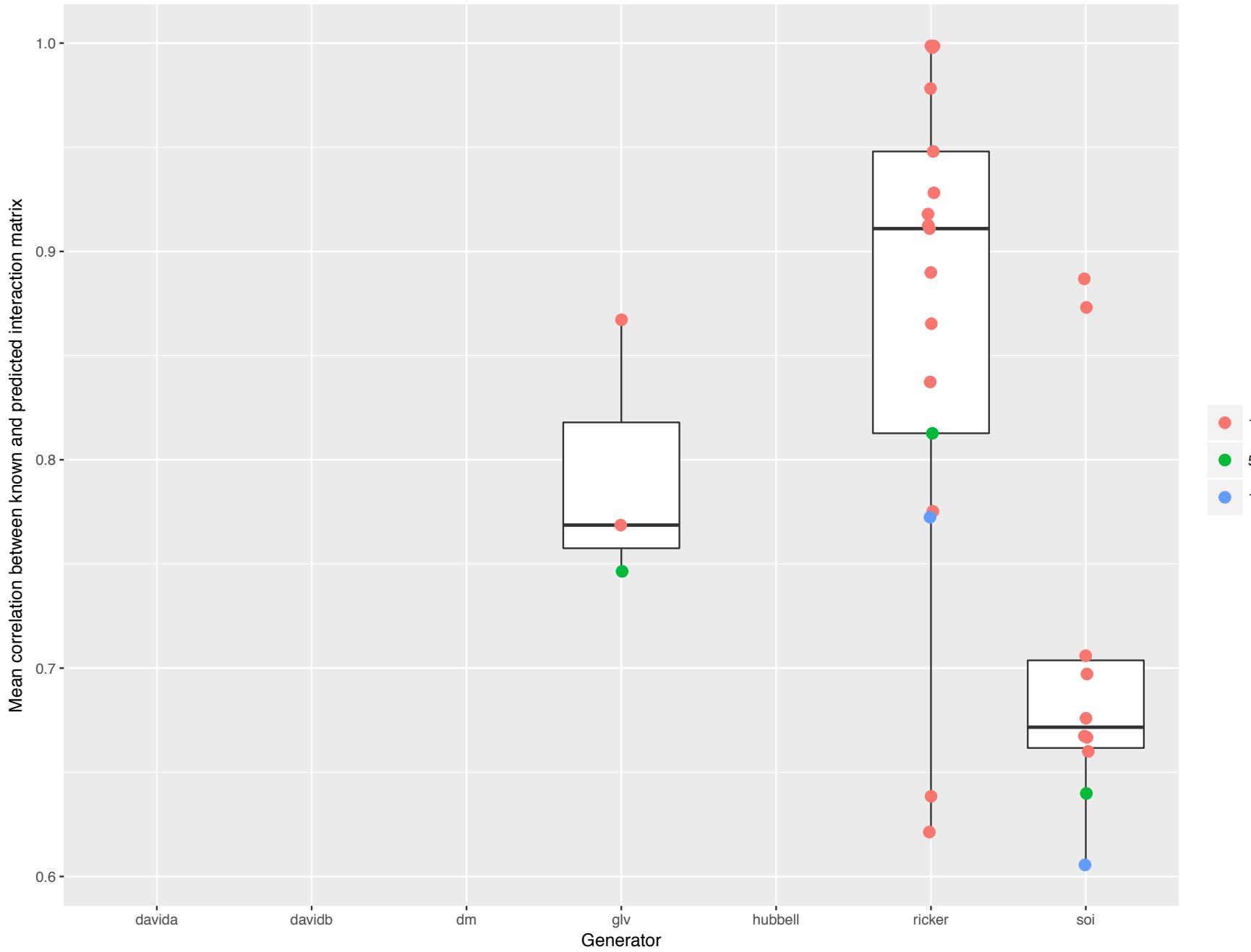
Methods



Fisher and Mehta (2014). Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries using Sparse Linear Regression. *PLoS one* 9, e102451.

Generator versus mean correlation between known and predicted interaction matrix colored by interval

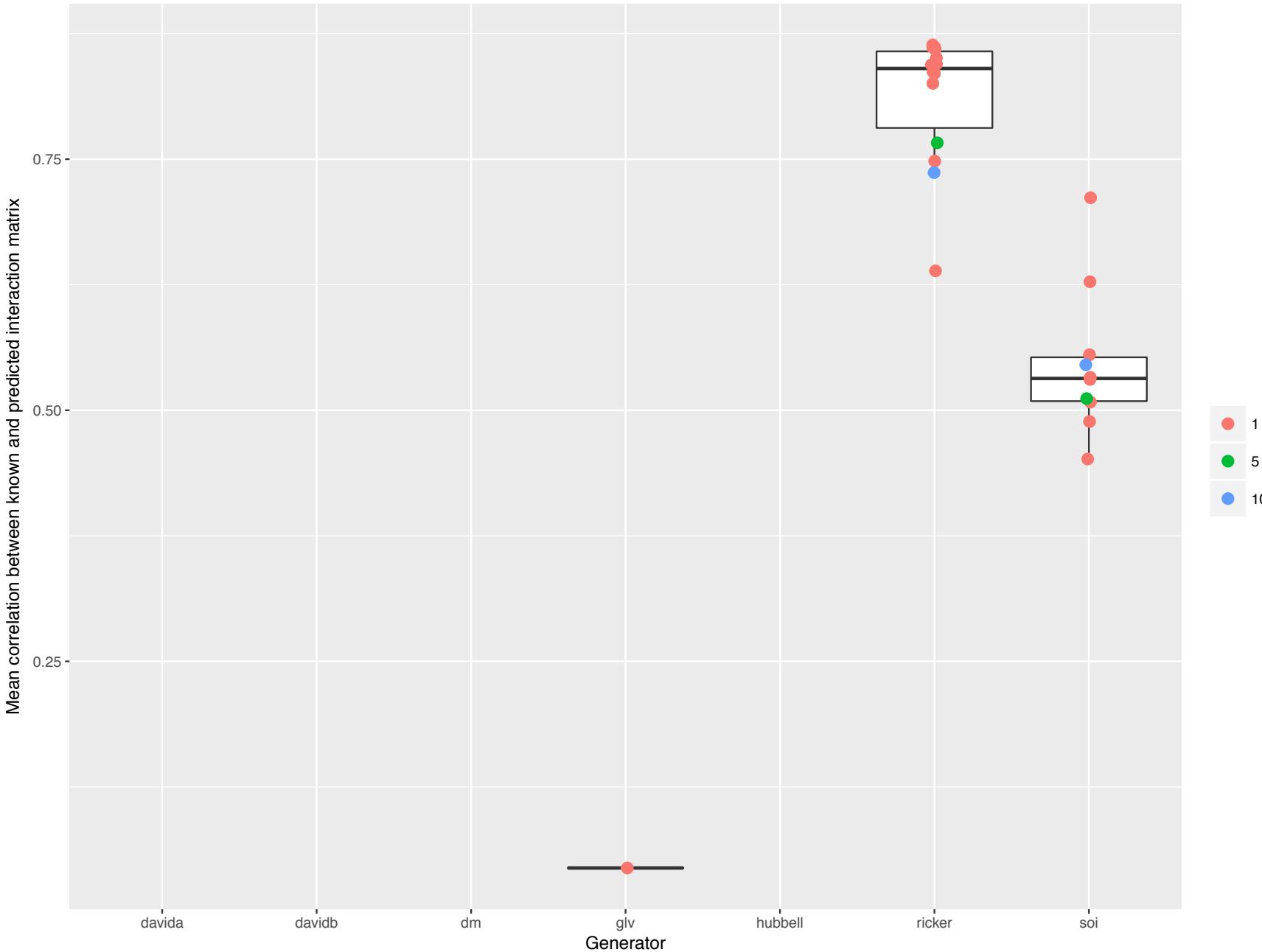
Methods



Sub-set of 100 time points

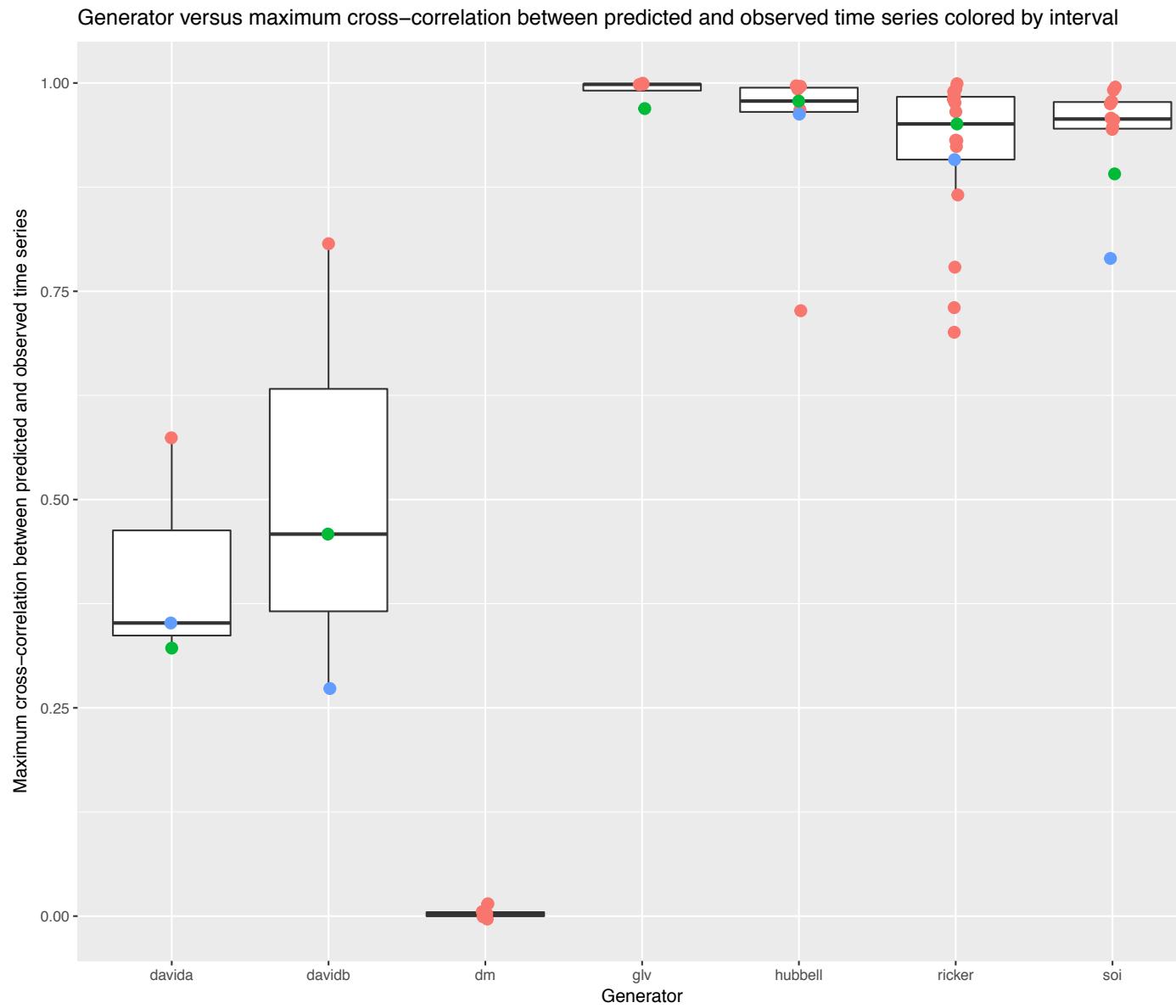
Generator versus mean correlation between known and predicted interaction matrix colored by interval

Methods



Methods

Correlation between observed and predicted time series is misleading!



Conclusions

- We can tell the community model of a time series by noise type
 - Neutral model: brown noise
 - gLV/noise-free Ricker: black noise
 - Ricker/SOI: pink noise (models can be separated by mean autocorrelation versus species number)
 - DM: neither black, brown or pink noise (no dependency between time points)
- Ecological time series properties such as time decay and the slope of the Taylor law are model-dependent
- Sampling interval matters
- LIMITs works well on non-neutral data
- No simulated time series captures fully the properties of the stool time series data

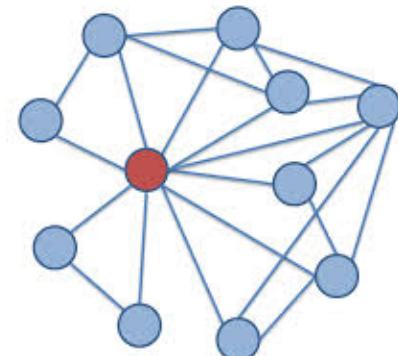
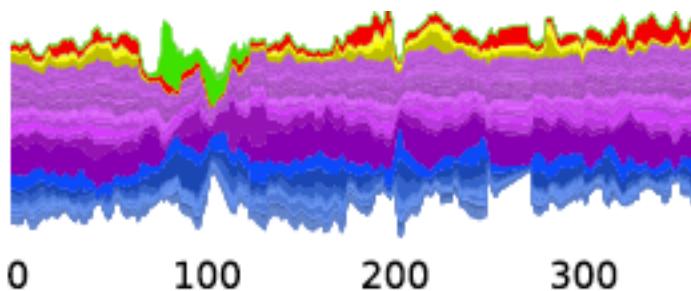
Take Home Message

- Test the noise type before inferring a network from your time series data

Pink noise?

Brown noise?

Black noise?



Collaborators

KU LEUVEN

Raes lab

Leo Lahti



Applied Physics
Sophie De Buyl



**UNIVERSITY
OF VIENNA**

Computational Systems

Biology

Stefanie Widder

Franziska Bauchinger

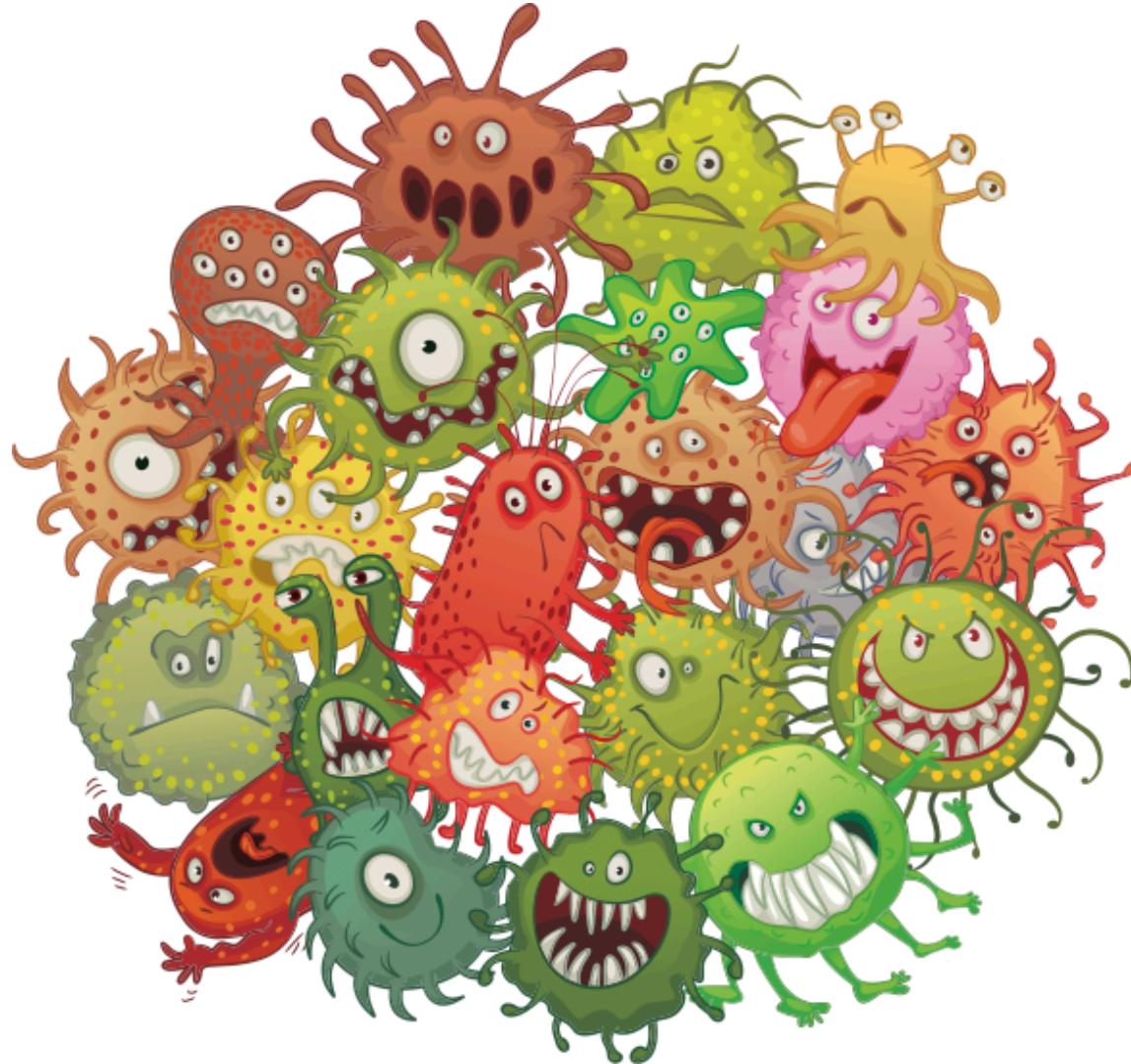


**Theoretical
Chronobiology**
Didier Gonze



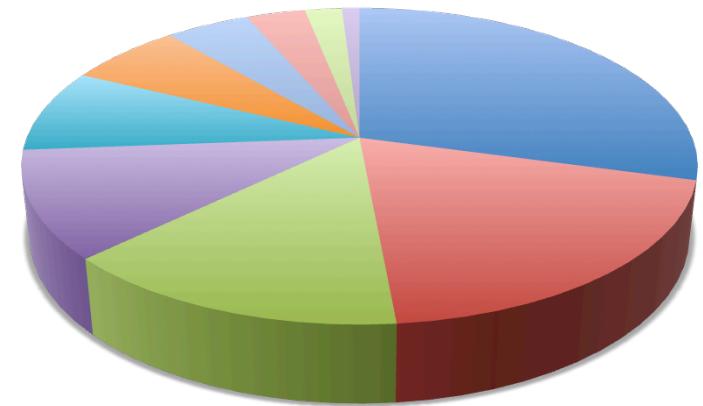
**Department of
Biology**
Alex Washburne

Thank you!



Appendix: broken-stick

- Niche apportionment model developed by MacArthur
- Uneven species proportions, but no species with zero abundance
- Abundance of rth species:



Example for 10 species

$$\frac{1}{N} \sum_{i=1}^r \frac{1}{N-i+1} \quad N: \text{species number}$$

MacArthur (1957), PNAS. "On the Relative Abundance of Bird Species" 43 (3), 293-295.

Appendix: Ricker model

δt : discrete time step (Ricker is a discrete model)

$X_i(t)$: abundance of target species i at time point t

$\langle x_j \rangle$: steady state abundance of species j (**carrying capacity**)

$\eta_i(t)$: log-normal **noise**

a_{ij} : interaction coefficient between taxa i and j

$$x_i(t + \delta t) = \eta_i(t) x_i(t) \exp(\delta t \sum_j a_{ij} (x_j(t) - \langle x_j \rangle))$$

For $\eta_i(t) = 1$ (no noise) and $\delta t \rightarrow 0$, Ricker model reduces to generalized Lotka-Volterra in continuous form.

Appendix: LIMITs - principle

- Data is divided into training and test set. Difference between observed abundance in training set and predicted abundance in test set is minimized.
- Interaction matrix is reported as the median of several iterations.
- Predictor species j are added one at a time, to identify optimal predictor species sub-set (sparse regression).

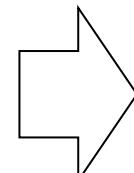
$$y_1 = \log x_i(1+1) - \log x_i(1) = \sum_j a_{ij} (x_j(1) - \langle x_j \rangle)$$

$$y_2 = \log x_i(2+1) - \log x_i(2) = \sum_j a_{ij} (x_j(2) - \langle x_j \rangle)$$

$$y_3 = \log x_i(3+1) - \log x_i(3) = \sum_j a_{ij} (x_j(3) - \langle x_j \rangle)$$

↓

$$y_t = \log x_i(t+1) - \log x_i(t) = \sum_j a_{ij} (x_j(t) - \langle x_j \rangle)$$

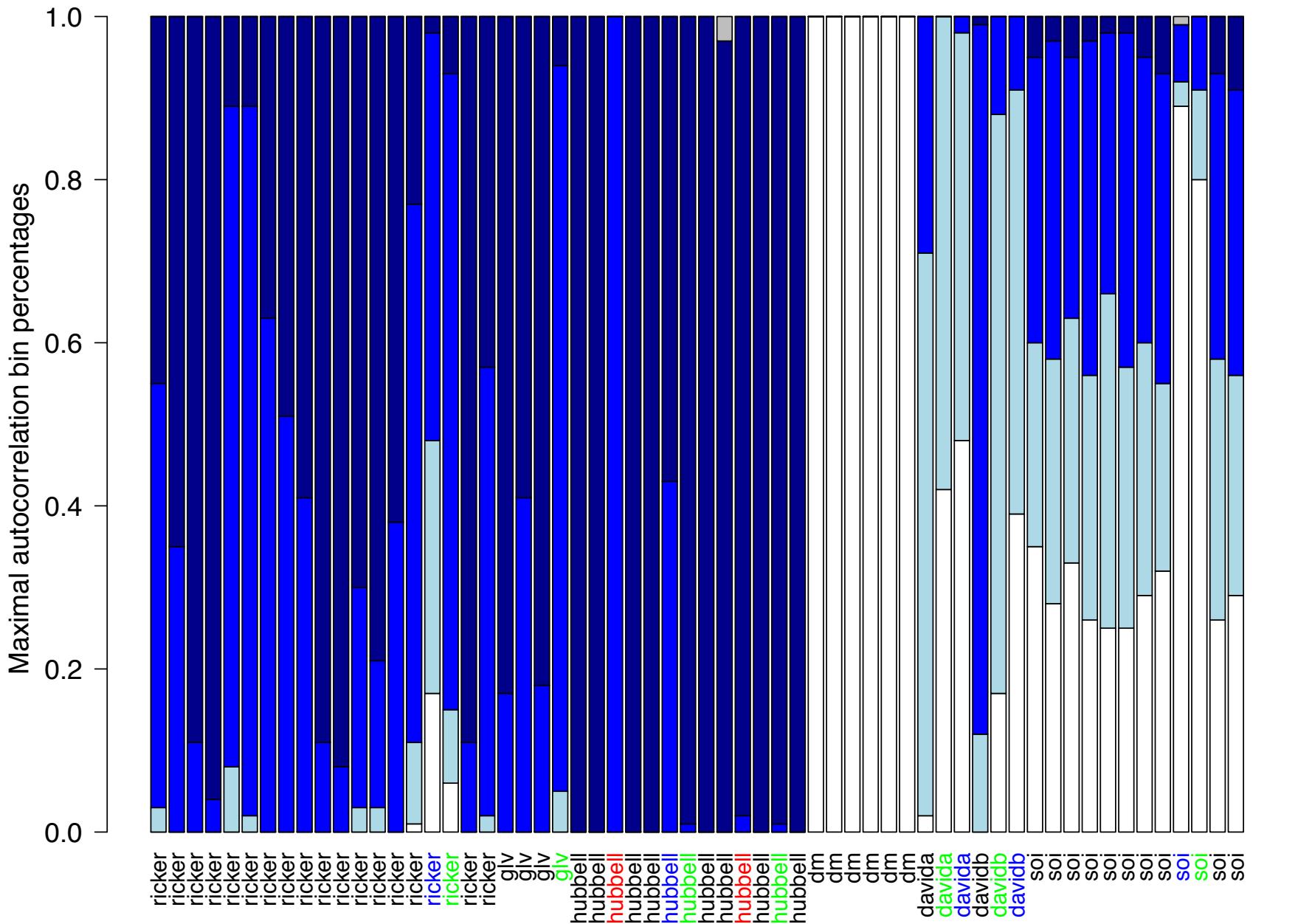


$\eta_i(t)$: 1, δt : 1
 y_i : **Vector** of logarithmic abundance differences for target species i at time point $t+1$ and time point t
 X : **Matrix** with species abundances as rows (after subtraction of steady state abundances) and time points as columns
 X^{-1} : Pseudo-inverse of X
 A : **Interaction matrix** (estimated row by row; for each species i one row)

$$A = \sum_t y_t X^{-1}$$

Appendix: Maximal autocorrelation

Maximal autocorrelation bin composition



Interval=5

Interval=10

Death rate=1000

Appendix: Neutrality test (Alex Washburne)

