MSc Thesis: Estimation of higher-order species interactions from ecological time series
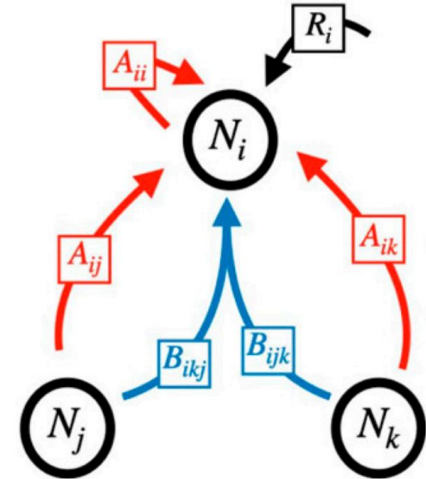
Maria Pröbstl

June 19 2023 – Group Meeting

Coexistence in diverse communities with higher-order interactions

Theo Gibbs[a,1] (iD), Simon A. Levin[b] (iD), and Jonathan M. Levine[b] (iD)

$$\frac{dN_i}{dt} = N_i \left(R_i - \sum_j A_{ij} N_j - \sum_j \sum_k B_{ijk} N_j N_k \right)$$

➢ Is there enough evidence in real ecology data for the existence of higher-order interactions?

Master Thesis Overview

Aim
➢ estimate spares higher-order taxa-taxa interactions from real data (16S rRNA microbiome data and other ecological time series)

Tasks
➢ apply methods to time series data
  ▪ DeepMoD
  ▪ Compositional Lotka Volterra
  ▪ Vector-Autoregression

➢ find useful datasets for testing these methods
  ▪ Human gut time series
  ▪ Artificial gut data
  ▪ Bucci data (microbial data)
  ▪ BioTIME database (ecological time series)

Compositional Lotka Volterra (cLV)

Generalized Lotka Volterra (gLV):

$$\forall i = 1, \ldots, D : \frac{d}{dt} x_i(t) = x_i(t) \left(g_i + \sum_{j=1}^{D} A_{ij} x_j(t) + \sum_{p=1}^{P} B_{ip} u_p(t) \right).$$

Under additive log ratio (alr) transformation:

$$\forall i = 1, \ldots, D : \frac{d}{dt} \log x_i(t) = g_i + \sum_{j=1}^{D} A_{ij} x_j(t) + \sum_{p=1}^{P} B_{ip} u_p(t)$$

with $N(t) = \sum_{j=1}^{D} x_j(t)$ and $\pi_i(t) = x_i(t)/N(t)$

$\rightarrow$

$$\frac{d}{dt} \log\left(\frac{\pi_i(t)}{\pi_D(t)} \right) = \overbrace{(g_i - g_D)}^{\bar{g}_i} + \sum_{j=1}^{D} \overbrace{(A_{ij} - A_{Dj})}^{\bar{A}_{ij}} x_j(t) + \sum_{p=1}^{P} \overbrace{(B_{ip} - B_{Dp})}^{\bar{B}_{ip}} u_p(t)$$

$$= \bar{g}_i + \sum_{j=1}^{D} N(t) \bar{A}_{ij} \pi_j(t) + \sum_{p=1}^{P} \bar{B}_{ip} u_p(t)$$

Compositional Lotka Volterra (cLV)

$$\frac{d}{dt}\log\left(\frac{\pi_i(t)}{\pi_D(t)}\right) \approx \bar{g}_i + \sum_{j=1}^{D}\bar{A}_{ij}\pi_j(t) + \sum_{p=1}^{P}\bar{B}_{ip}u_p(t) =: f_i(t)$$

$\pi_i(t)$ $= x_i(t)/N(t)$ relative abundance of taxon i at time t

$x_i(t)$ concentration of taxon $i \in \{1, \dots, D\}$ at time t

$N(t)$ $= \sum_{j=1}^{D} x_j(t)$

$\bar{g}_i$ **relative** (to the denominator) growth rates
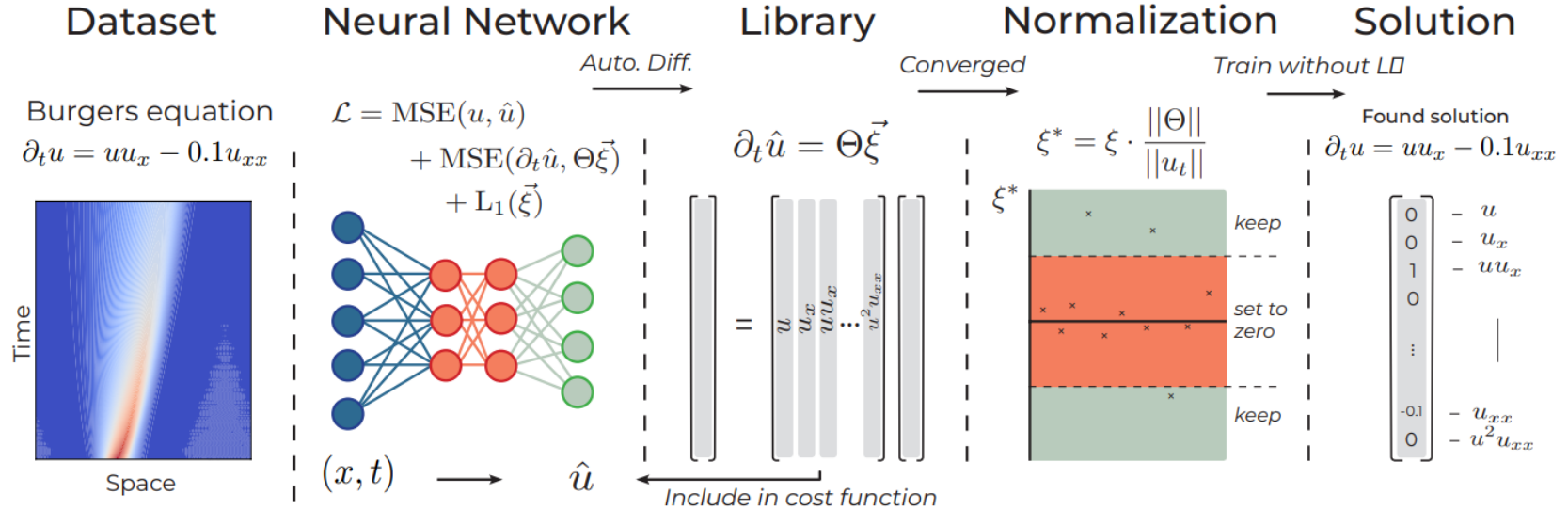
$\overline{A_{ij}}$ **relative** interactions

$\overline{B_{ip}}$ **relative** external effects

$u_p(t)$ indicator variable describing presence or absence of external perturbation p at time t
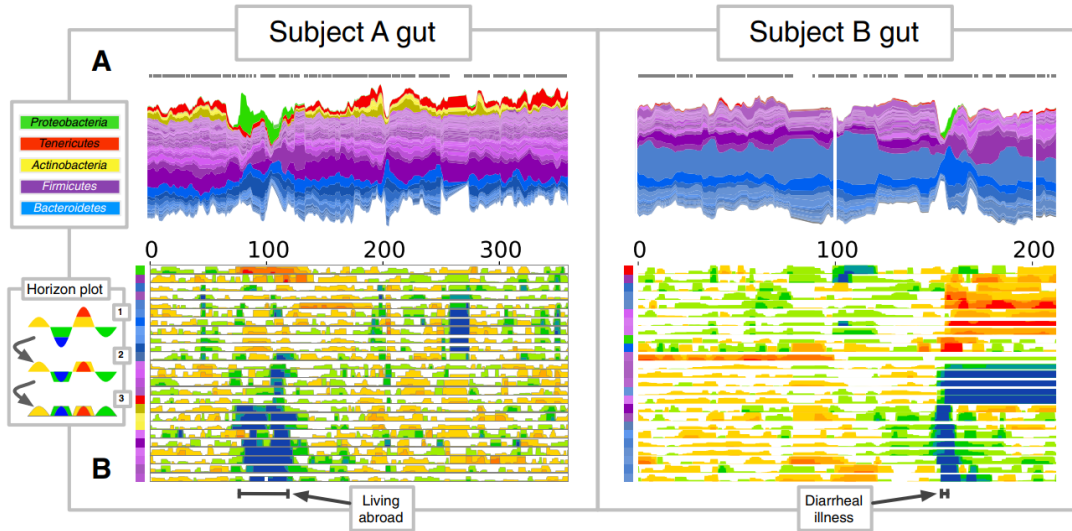
DeepMoD
A Deep learning based Model Discovery algorithm

→ Python code available (github repo)

Human Gut Time Series Data

gut microbiota 16S rRNA sequencing data sets of two individuals
→ David et al. (2014)

- Subject A:
  samples through 359 days
  with 60 missing time points

- Subject B:
  318 days with 138 missing
  time points

Human Gut Time Series Data

gut microbiota 16S rRNA sequencing data sets of two other individuals (M3 and F4)
→ Caporaso et al. (2011)

- M3:   samples through **443 days** with 111 missing time points
- F4:   **186 days** with 56 missing time points

- already as **preprocessed and interpolated** time series available (Gibbons et al. (2017))

- taxonomic assignment of the study:
  54 (for M3) and 42 (for F4) orders with at least one non-zero read count

- also microbiome data of tongue, left/right palm available

Sources: Äijö T, Müller CL, Bonneau R. Temporal probabilistic modeling of bacterial compositions derived from 16S rRNA sequencing. Bioinformatics. 2018
Gibbons SM, Kearney SM, Smillie CS, Alm EJ (2017) Two dynamic regimes in the human gut microbiome. PLoS Comput Biol 13(2): e1005364.
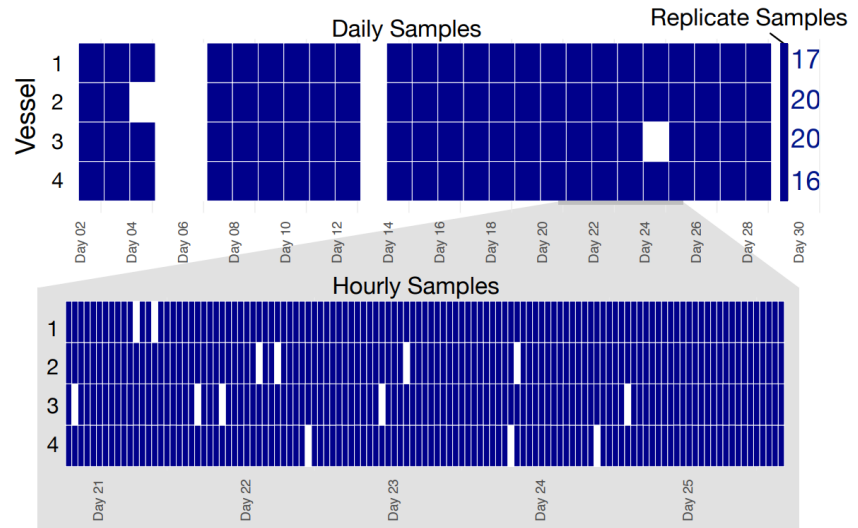
miaTime: Silverman Artificial Gut Data

The dataset contains 16S rRNA sequencing data of 4 in vitro artificial gut models
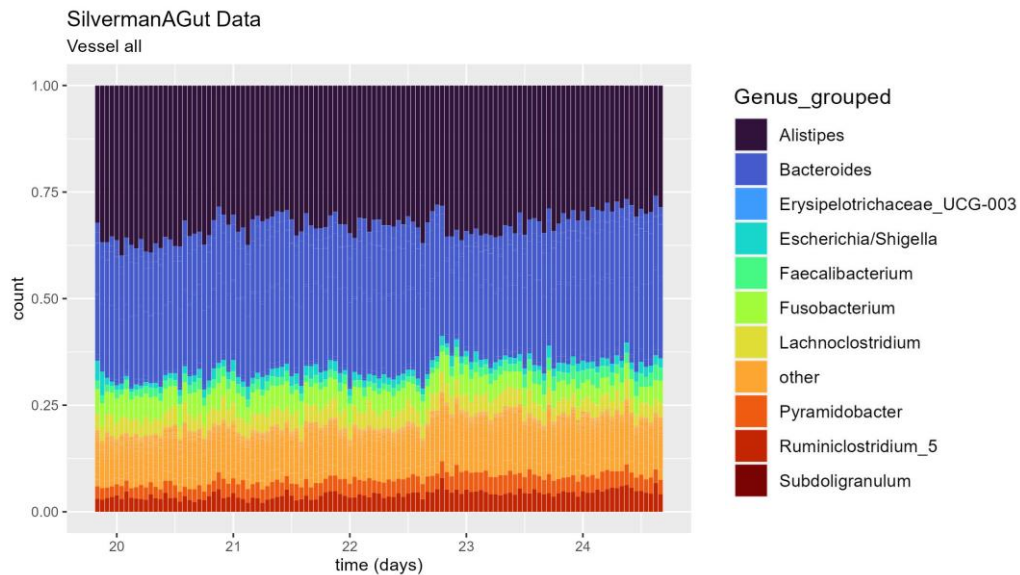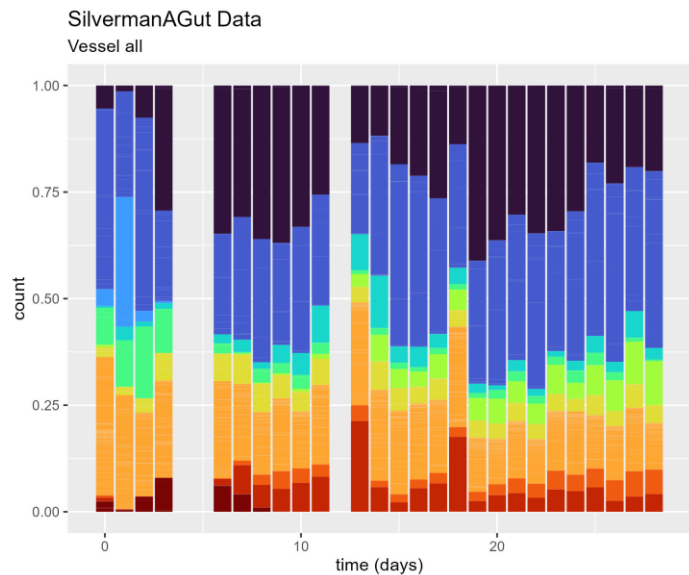
- 26 days / 117 hours

Taxonomic levels:
- Kingdom to Species
- lots of NAs in Species
- but 104 different Genus

miaTime: Silverman Artificial Gut Data

Bucci Dataset

- Five adult germfree mice were gavaged with **23 human commensal bacterial type strains**

- After **28 days** for the commensal flora to establish, mice were gavaged with **C. difficile spores** and monitored for an additional 28 days

Source: Joseph, Tyler A., et al. "Compositional Lotka-Volterra describes microbial dynamics in the simplex." PLOS Computational Biology 16.5 (2020): e1007917.
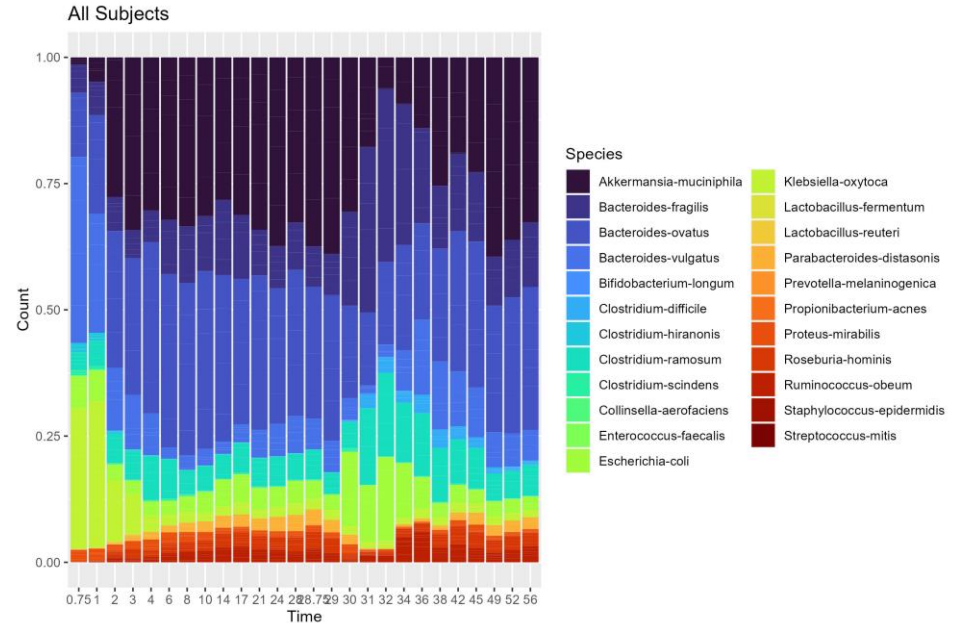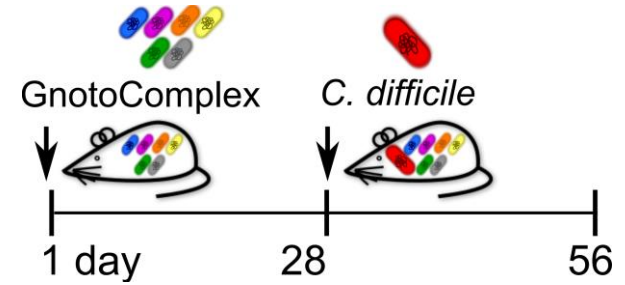
BioTIME – Database

361 studies   130 years   374 contributors   14 taxa   30 biomes
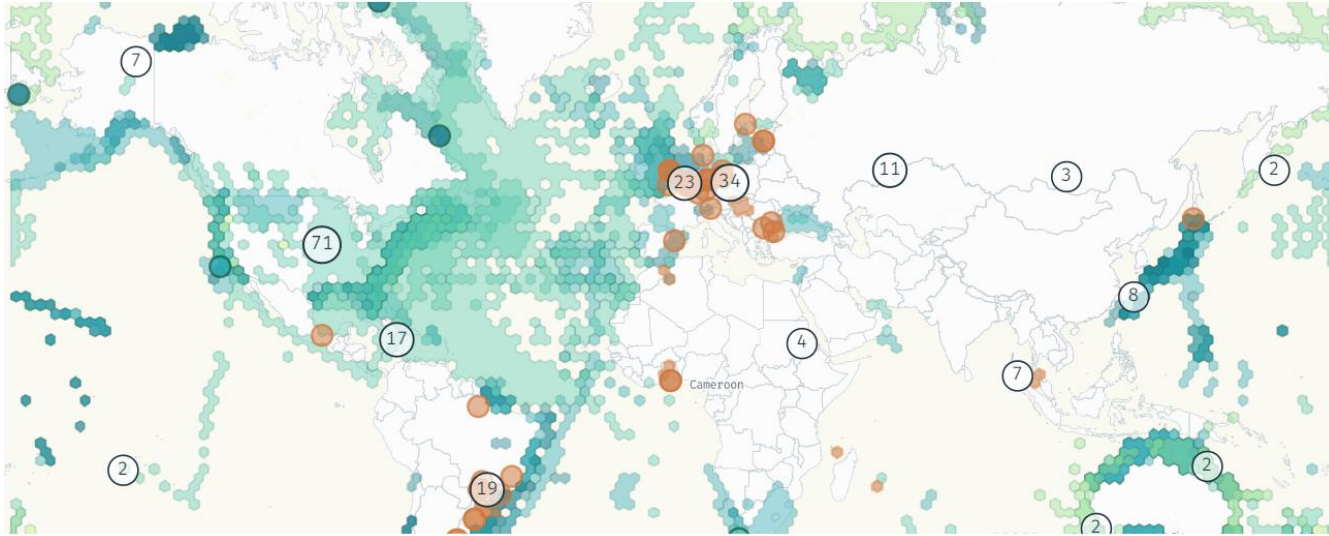
BioTIME – Study 339

Species trends turnover and composition of a woodland bird community in southern Sweden during a period of 57 years.

REALM	Terrestrial
CLIMATE	Temperate
TAXA	Birds
DATA_POINTS	57
YEARS	1953-2009
NUMBER_OF_SPECIES	39
NUMBER_OF_SAMPLES	57
TOTAL RECORDS	1210
ABUNDANCE_TYPE	Count

by Species, total Abundances
Study 339

Species_grouped
- Acrocephalus palustris
- Anthus trivialis
- Emberiza citrinella
- Erithacus rubecula
- Ficedula hypoleuca
- Fringilla coelebs
- Hippolais icterina
- Luscinia luscinia
- other
- Parus caeruleus
- Parus major
- Phylloscopus collybita
- Phylloscopus trochilus
- Prunella modularis
- Sturnus vulgaris
- Sylvia atricapilla
- Sylvia borin
- Sylvia communis
- Troglodytes troglodytes
- Turdus merula

BioTIME – Study 478

Long term study of the stream ecosystems in the Breitenbach

REALM	Freshwater
CLIMATE	Temperate
TAXA	Freshwater invertebrates
DATA_POINTS	37
YEARS	1969-2005
NUMBER_OF_SPECIES	90
NUMBER_OF_SAMPLES	37
TOTAL RECORDS	1537
ABUNDANCE_TYPE	Count

by Species, total Abundances
Study 478

Species_grouped
- Agapetus fuscipes
- Amphinemura standfussi
- Apatania fimbriata
- Baetis rhodani
- Baetis spec.
- Baetis vernus
- Chaetopteryx villosa
- Leuctra digitata
- Leuctra nigra
- Leuctra prima
- Nemoura cambrica
- Nemurella pictetii
- other
- Paraleptophlebia submarginata
- Protonemura auberti
- Protonemura intricata
- Protonemura meyeri
- Siphonoperla torrentium
- Tinodes rostocki
- Wormaldia occipitalis

Next steps

- ➢ apply cLV and DeepMoD on
  1. synthetical data (miaSim)
  2. real time series data

- ➢ find the „right" level of taxonomic rank for reasonable analysis

- ➢ interpretation of the results in the biological context

- ➢ include (sparse) Vector-Autoregression (VAR)

- ➢ create a workflow with reproducable code (github repo)

References

(1) Gibbs T, Levin SA, Levine JM. Coexistence in diverse communities with higher-order interactions. Proc Natl Acad Sci U S A. 2022 Oct 25;119(43):e2205063119

(2) Joseph, Tyler A., et al. "Compositional Lotka-Volterra describes microbial dynamics in the simplex." PLOS Computational Biology 16.5 (2020): e1007917.

(3) Both, Gert-Jan, et al. "DeepMoD: Deep learning for model discovery in noisy data." Journal of Computational Physics428 (2021): 109985.

(4) Silveman J.D et al. "Dynamic linear models guide design and analysis of microbiota studies within artificial human guts." (2018) Microbiome 6:202 https://doi.org/10.1186/s40168-018-0584-3

(5) Dornelas, Maria, et al. "BioTIME: A database of biodiversity time series for the Anthropocene." Global Ecology and Biogeography 27.7 (2018): 760-786.

(6) David, L.A., Materna, A.C., Friedman, J. et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol 15, R89 (2014). https://doi.org/10.1186/gb-2014-15-7-r89

(7) Caporaso, J.G., Lauber, C.L., Costello, E.K. et al. Moving pictures of the human microbiome. Genome Biol 12, R50 (2011). https://doi.org/10.1186/gb-2011-12-5-r50

(8) Gibbons SM, Kearney SM, Smillie CS, Alm EJ (2017) Two dynamic regimes in the human gut microbiome. PLoS Comput Biol 13(2): e1005364. https://doi.org/10.1371/journal.pcbi.1005364