# Estimate the Strength of Interactions between Bacteria Opportunities and Challenges

**Thierry Van Effelterre**
**Discovery Statistics**

**Janssen**

janssen
PHARMACEUTICAL COMPANIES
OF Johnson & Johnson

**Microbiome Statistics Meeting**
**Gent**
**January 16, 2017**

# Dynamical systems approach to the Microbiome

- A dynamical systems approach to the microbiome has a huge potential.
- It is mechanistic → it may in particular allow to better understand both qualitatively and quantitatively what the impact of various interventions can be:
  - Which bacteria can persist if introduced.
  - Which bacteria may disappear
  - Whether the microbiota remains stable or not.
  - …

# Generalized Lotka-Volterra model with N species

$$\frac{dP_i}{dt} = P_i \left(r_i - \sum_{j=1}^{N} \alpha_{ij} P_j\right)$$

$$with \quad i = 1,2,\ldots, N$$

A generalized Lotka-Volterra model has $2^N$ fixed points, one for each combination of absence – presence of each of the N species.

For each fixed point:

- **The fixed point can be realized or not.**
  Realized if all coordinated of fixed point ≥ 0.
  Abundances cannot be negative!

- **If all abundances ≥ 0, the fixed point can be asymptotically stable or not**.. Characterized through linearization around the fixed point.

- **The fixed point can be globally asymptotically stable or not**.

For any fixed points, the system is reduced to a system of a similar form, with M ordinary differential equations in M populations, with M ≤ N

$$\frac{dP_i}{dt} = P_i\,(r_i - \sum_{j=1}^{N} \alpha_{ij}\, P_j)$$

$$with \quad i = 1,2,\ldots, N$$

Where P1, P2, …, $P_M$ are the M populations strictly positive at the fixed point.

The fixed point can be found by solving a linear system:

$$\vec{F} = -A^{-1}\,\vec{r}$$

Where

- $\vec{F}$ is the vector with the coordinates of the fixed point
- $A^{-1}$ is the inverse of the M x M matrix A with the interaction parameters $\alpha_{ij}$
- $\vec{r}$ is the vector of growth rates

# Global asymptotic stability

Much research has been devoted to the generalized Lotka-Volterra systems, both theoretical and applied.

One of the important result is **Goh's theorem:**

If F is the unique fixed point of the generalized Lotka-Volterra system, with all coordinates > 0, e.g. all species present, then this fixed point is globally asymptotically **stable if there exists a diagonal matrix D > 0** such that

$$DA + A^t D$$

**is negative-definite**,

e.g. all the eigenvalues of this (symmetric) matrix are negative.

**This is an important theorem,** as it gives a condition for a fixed point to be **globally asymptotically stable, e.g. the system converges to that fixed point from any state.**

# Ecological interactions between bacteria

- The generalized Lotka-Volterra (GLV) dynamical systems approach provides a framework to better understand the interactions between different bacteria.

- The interactions (the « alpha's ») can be positive, negative or zero.

- This framework allows to account for different types of interactions between pairs of bacteria « species »
  - Competition ( -, -) , which an be direct or indirect
  - Mutualism (+, +)
  - Neutralism (0, 0)
  - Amensalism (-, 0)
  - Comensalisms (+, 0)
  - Antagonism, like predation or parasitims (-, +)
  - Neutralism (0, 0)

# How can the interaction parameters be estimated for the microbiome?

- A key challenge is to estimate interaction parameters between bacteria.
- One approach is to estimate the interaction parameters from longitudinal microbiome data, more specifically from the abundance of OTU's at several (sufficiently close) time points – as in Stein et al, Bucci et al

$$\frac{dP_i}{dt} = P_i\left(r_i - \sum_{j=1}^{N} \alpha_{ij}\, P_j\right) \qquad with\ i = 1,2,\ldots, N$$

$$\Leftrightarrow \frac{d(Ln(P_i))}{dt} = \frac{\frac{dP_i}{dt}}{P_i} = \left(r_i - \sum_{j=1}^{N} \alpha_{ij}\, P_j\right)$$

$$\frac{d(Ln(P_i))}{dt} = \frac{\frac{dP_i}{dt}}{P_i} = (r_i - \sum_{j=1}^{N} \alpha_{ij} P_j)$$

$$\frac{Ln(P_i(t_{k+1})) - Ln(P_i(t_k))}{t_{k+1} - t_k} = r_i - \sum_{j=1}^{N} \alpha_{ij} P_j$$

$$\frac{\ln(P_i(t_{k+1})) - \ln(P_i(t_k))}{t_{k+1} - t_k} = r_i - \sum_{j=1}^{N} \alpha_{ij} P_j$$

$with$

$i = 1,2,\ldots, N$

$k = 1,2,\ldots, T$

**the N OTU's**

**the T time points**

**The interaction parameters (the alpha's) can then estimated with regularized linear regression.**

# Simulation Phase

- Simulation of the generalized Lotka-Volterra (GLV) model with
  - 20 OTU's
  - 100 subjects
  - 100 time points

For given growth rates (r) for each OTU and a given 20 x 20 interaction matrix,

(1) Deterministic GLV model solved (integrated) over some time period.

(2) For each of 20 OTU'S,

Generate 100 times (100 subjects) OTU abundance by randomly generating abundance from Poisson distribution with mean equal to projection of deterministic model for that OTU at that time point.

# Estimation phase

- Using this simulated data set, estimate the GLV model parameters using regularized regression.

  - Data: 100 subjects x 20 OTU's x 100 time points
  - Parameters to be estimated
    - Growth rates (the ri's): 20 (1 by OTU)
    - Strength of interaction between OTU's (the alpha's): 400 (1 by pair of OTU's)

  - Only constraints used:
    - Growth rates have to be positive
    - Interaction of an OTU with itself (self-limitation) has to be negative

# Estimation

- Used regularized *R* package *glmnet* that fits GLM's via penalized maximum likelihood.
- Linear regression with penalization of too high values for model parameters.
- Estimates the parameters for many values of the "penalty" parameter lambda (λ)
- N-fold cross validation for each value of λ.

# Estimation

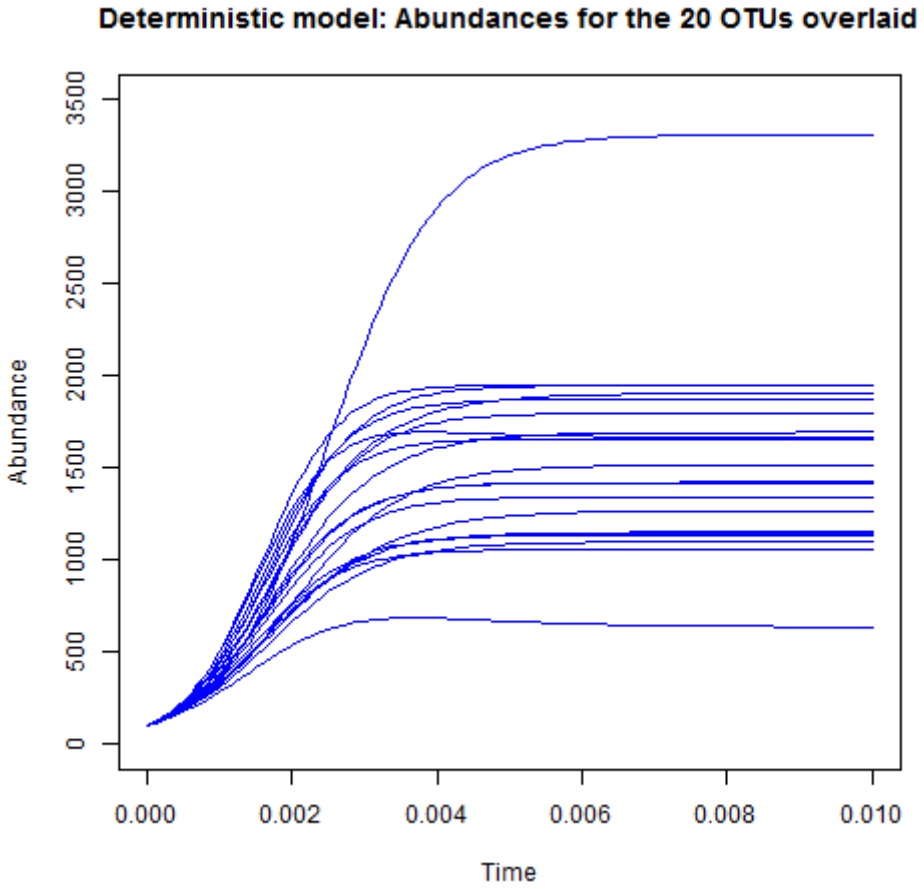
- For the test examples shown here **as illustration,** the estimates of the growth rates obtained for the λ value with minimal cross-validation error gives
  - reasonably good estimates of growth rates,

  However:
  - too low magnitude for self-interactions
  - too many interactions between pairs of OTU's

- However can obtain better estimates with a higher value of lambda, for example lambda = 1 for test example.

# Deterministic model

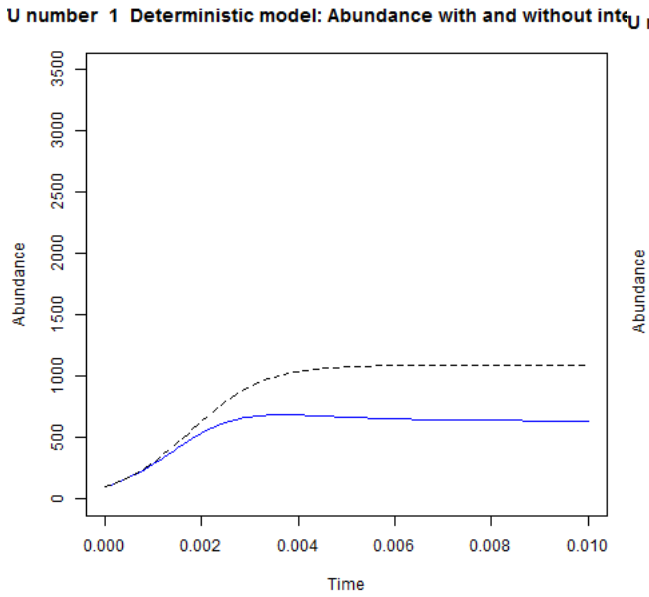
Projected abundance for each of the 20 OTU's overlaid

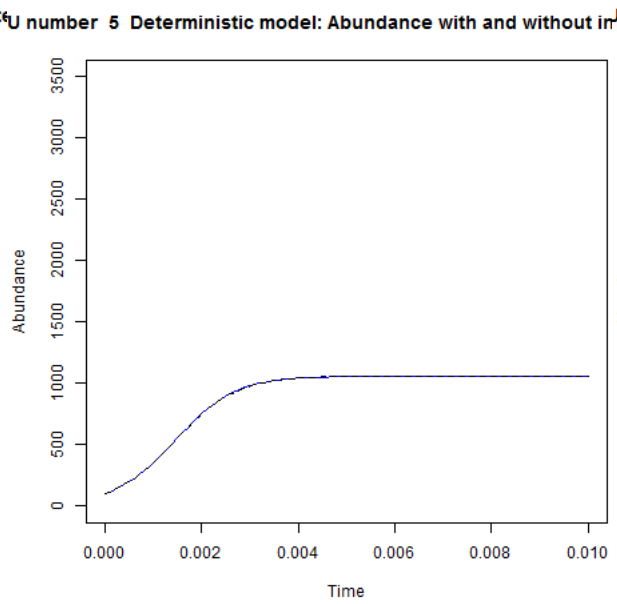


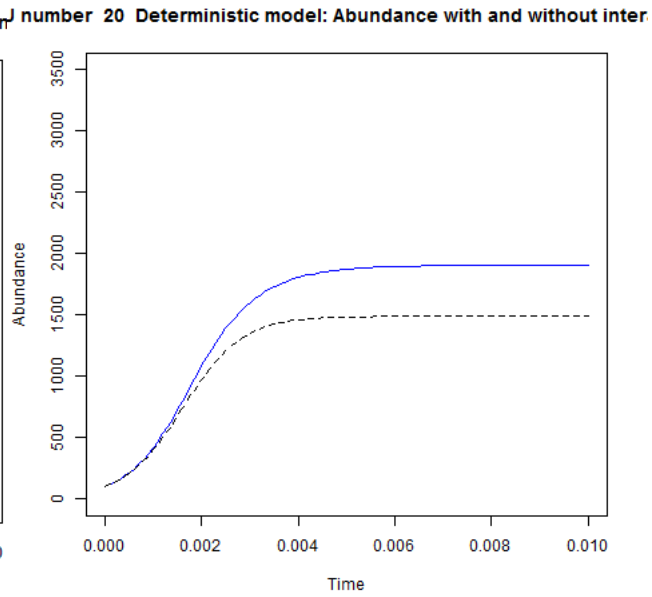
# Deterministic model

Projected abundance for 3 of the 20 OTU's

with and without interactions between pairs of OTU's



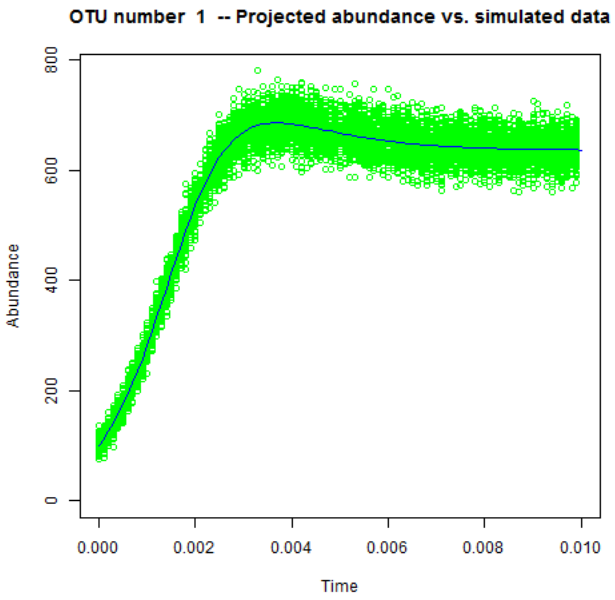



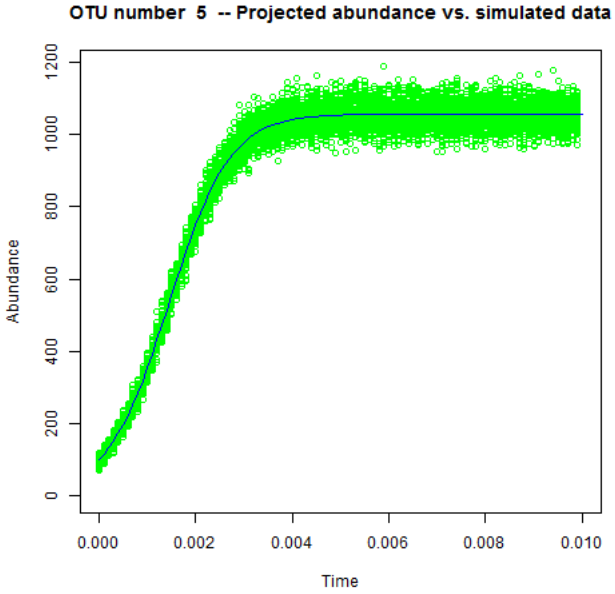



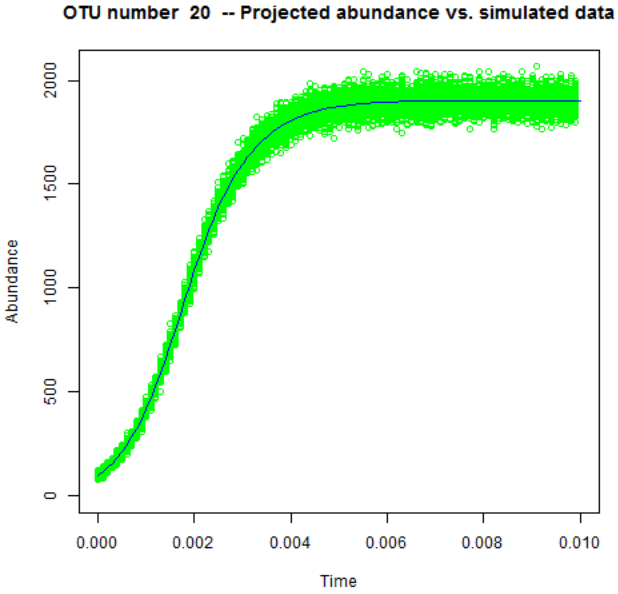

Deterministic model

- blue solid curves: with interactions between pairs of OTU's
- black dashed curves: Without any interactions between pairs of OTU's

# Deterministic model
# + Data generated from Poisson distribution

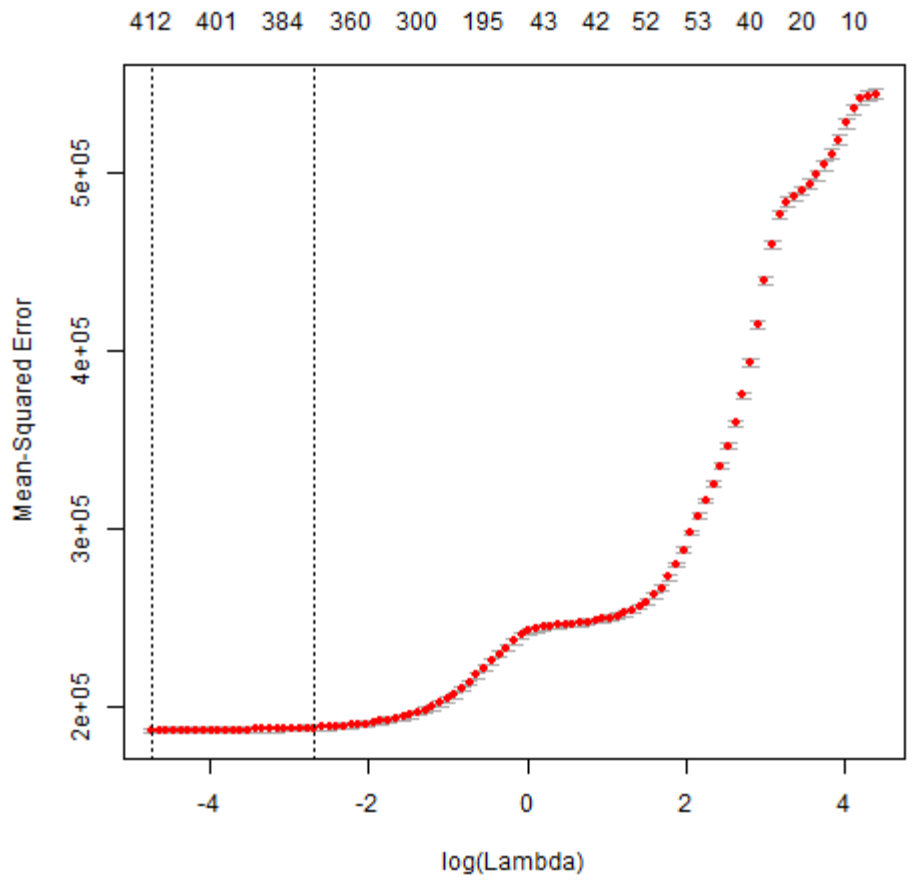



- Blue solid curve: deterministic model
- Green dots: simulated data

# Cross validation Error vs. Log(λ)




Lambda
lambda.min 0.008832575
lambda.1se 0.06231207

# Estimation Outcomes with λ=1

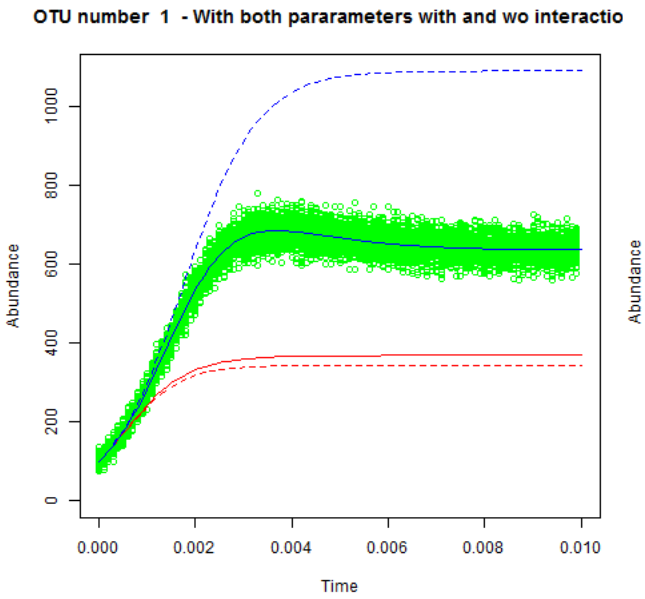
- Ratios original parameters : Estimated parameters:

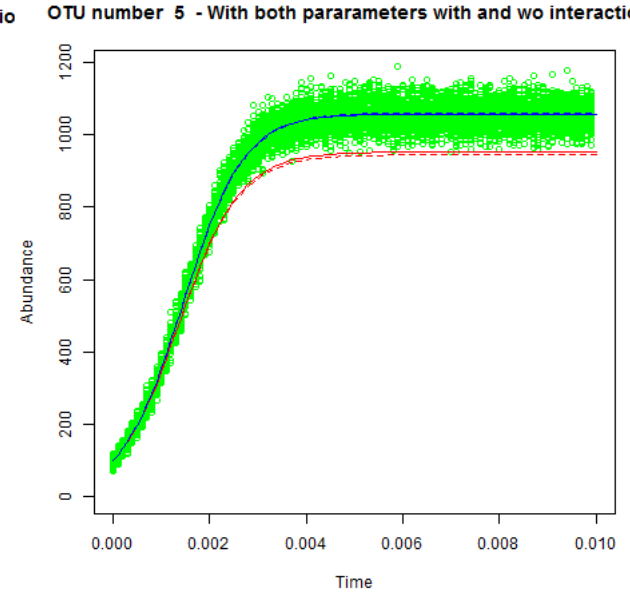
| Ratio Original parameters / estimated paramaters | Min | Q1 | Median | Mean | Q3 | Max |
|---|---|---|---|---|---|---|
| **Growth rates** | **0.77** | **1.01** | **1.03** | **1.01** | **1.04** | **1.06** |
| **Self interactions** | **0.24** | **0.94** | **0.97** | **0.95** | **1.03** | **1.33** |

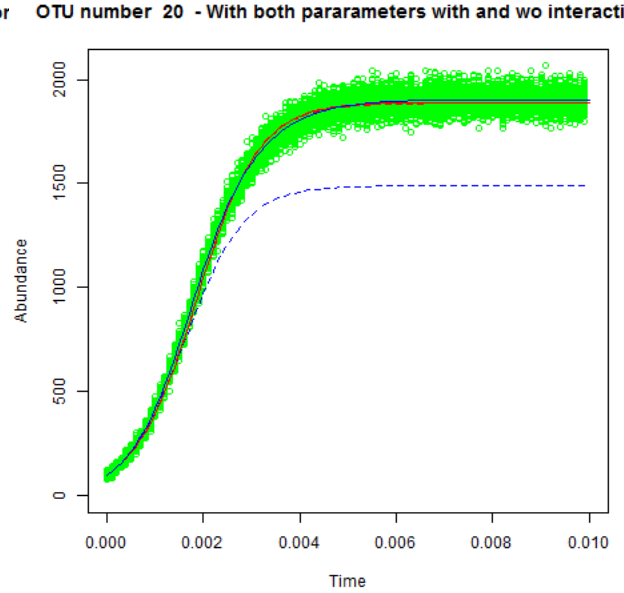
Number of interactions between pairs of OTU's for which the estimated parameter has the same sign as the original parameter:

279 / 400 (70%)

# Deterministic model with initial parameters and estimated parameters + Data generated from Poisson distribution




- Blue solid curve: deterministic model with *original parameters*
- Blue dashed curve: deterministic model with *original parameters* without interactions
- Red solid curve: deterministic model with *estimated parameters*
- Red dashed curve: deterministic model with *estimated parameters* without interactions.
- Green dots: simulated data

# Challenges

- Many interaction parameters need to be estimated: may be useful to reduce the complexity by identifyng the most important bacterial species.

"Importance" may relate to

  - Disease,
  - interventions,
  - magnitude of interactions.

One may need to first identify the strongest interactions through e.g. correlations, partial correlations.

# Challenges (cont'd)

- Controlled experiments may help to facilitate the quantification of interactions between bacteria.

- Add (or remove) one bacterial species at the time and observe the change in abundances of the other (N) bacterial species.

  → interactions with that specific species.

# Conclusion

- A dynamical systems approach has a great potential.
- Being mechanistic it gives the means to better understand the impact of interventions both qualitatively (stable states, persistence of species ...) and quantitatively (abundances).
- It a complex undertaking.
- Many challenges!