Statistical Learning of Dynamic Systems

Itai Dattner
Department of Statistics
University of Haifa

October 16, 2017
Biostatistics Seminar
The Harvard Catalyst Biostatistics Program
Motivation and background
- Real data examples
- Modeling dynamic systems

Main methodological idea
- Exploiting linear features of the model
- Bypassing numerical integration
- Integral estimation

Fully observed systems
- Identifiability

Partially observed systems
- Statistical properties
- Sieve estimation of ODEs
- Consistency

Case studies
- Learning age-dependent dynamics of infectious diseases
- Learning predator-prey dynamics

Conclusions
Motivation and background
**Figure:** Two age-groups: children 0-19 (blue) and adults 20+ (red).
Fit to eleven years of ILI data from Israel

Figure: Two age-groups: children 0-19 (blue) and adults 20+ (red). Black curves show the obtained fit using our estimation method.
Predator-prey dynamics

Figure: The interaction between the predators and prey leads to a creation of a third compartment, the complex predator-prey (known as Bdelloplast), that has separate dynamics. A separated compartment for the complexes is assumed since the interaction rate between the predator and their prey is not negligible compared to the predator reproduction rate.
Chemical reactions - $\alpha$-pinene problem
Chemical reactions - model fit
HIV dynamics - data of viral fitness for HIV-1 mutant 103N
HIV dynamics - model fit

Logarithm of cell number vs. time (hours)

- $x_1$
- $x_2$
- $x_3$
- $x_4$

Itai Dattner
Infectious diseases - England and Wales data: weekly case reports of measles
**Figure:** England and Wales measles data: weekly case reports.
Neural behavior - data of transmembrane potential of zebra-fish motor neuron
Neural behavior - model fit

**Figure:** Data of transmembrane potential of zebra-fish motor neuron.
The blood coagulation process

⇒ SYSTEM of ODEs
A system of ordinary differential equations (ODEs) can be expressed as follows:

\[
\begin{align*}
\begin{cases}
  x'(t) = F(x(t); \theta), & t \in [0, 1], \\
  x(0) = \xi,
\end{cases}
\end{align*}
\]

where \( x(t) \) takes values in \( \mathbb{R}^d \), \( \xi \) in \( \Xi \subset \mathbb{R}^d \), and \( \theta \) in \( \Theta \subset \mathbb{R}^p \).

For example,

\[
\begin{align*}
\begin{cases}
  x_1'(t) = \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\
  x_2'(t) = -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t).
\end{cases}
\end{align*}
\]
The Lotka-Volterra system in population dynamics

\[
\begin{align*}
    x_1'(t) &= \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\
    x_2'(t) &= -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t).
\end{align*}
\]

Here \( x_1 \) represents the prey population and \( x_2 \) the predator population.

- \( \xi_1 = 1, \xi_2 = 0.5 \)
The Lotka-Volterra system in population dynamics

\[
\begin{cases}
    x_1'(t) = \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\
    x_2'(t) = -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t).
\end{cases}
\]

Here $x_1$ represents the prey population and $x_2$ the predator population.

- $\xi_1 = 1, \xi_2 = 0.5$
- $\theta = (0.5, 0.9, 0.9, 0.5)^\top$
The Lotka-Volterra system in population dynamics

\[
\begin{cases}
    x'_1(t) = \theta_1 x_1(t) - \theta_2 x_1(t) x_2(t), \\
    x'_2(t) = -\theta_3 x_2(t) + \theta_4 x_1(t) x_2(t).
\end{cases}
\]

Here \( x_1 \) represents the prey population and \( x_2 \) the predator population.

- \( \xi_1 = 1, \xi_2 = 0.5 \)
- \( \theta = (0.5, 0.9, 0.9, 0.5)^\top \)
The Lotka-Volterra system in population dynamics

\[
\begin{align*}
  x_1'(t) &= \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\
  x_2'(t) &= -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t).
\end{align*}
\]

Here $x_1$ represents the prey population and $x_2$ the predator population.

- $\xi_1 = 1$, $\xi_2 = 0.5$
The Lotka-Volterra system in population dynamics

\[
\begin{align*}
    x_1'(t) &= \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\
    x_2'(t) &= -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t).
\end{align*}
\]

Here \( x_1 \) represents the prey population and \( x_2 \) the predator population.

- \( \xi_1 = 1, \xi_2 = 0.5 \)
- \( \theta = (0.1, 0.9, 0.9, 0.5)^\top \)
The Lotka-Volterra system in population dynamics

\[
\begin{aligned}
  x'_1(t) &= \theta_1 x_1(t) - \theta_2 x_1(t) x_2(t), \\
  x'_2(t) &= -\theta_3 x_2(t) + \theta_4 x_1(t) x_2(t).
\end{aligned}
\]

Here $x_1$ represents the prey population and $x_2$ the predator population.

- $\xi_1 = 1$, $\xi_2 = 0.5$
- $\theta = (0.1, 0.9, 0.9, 0.5)^\top$

![Graphs showing population dynamics over time]
The Lotka-Volterra system in population dynamics

\[
\begin{align*}
    x_1'(t) &= \theta_1 x_1(t) - \theta_2 x_1(t) x_2(t), \\
    x_2'(t) &= -\theta_3 x_2(t) + \theta_4 x_1(t) x_2(t).
\end{align*}
\]

Here $x_1$ represents the prey population and $x_2$ the predator population.

- $\xi_1 = 1, \xi_2 = 0.5$
- $\theta = (0.1, 0.9, 0.9, 0.5)^\top$
It is assumed that the solutions $x$ can not be observed directly but instead, $n$ noisy observations of $x$ are given by

$$Y_j(t_i) = x_j(t_i; \theta, \xi) + \epsilon_{j,i}, i = 1, \ldots, n, \ j = 1, \ldots, r,$$

where $r \leq d$, $0 \leq t_1 \leq \cdots \leq t_n \leq 1$ and with the convention that $Y_j(0) = x_j(0) = \xi_j$. 
More generally, we can consider GLM:

- observations $Y_j(t_i)$ follow a distribution from the exponential family

- define a dynamic link function $g(E(Y_j(t_i)))$ as a function of two components: one component is a vector of unobservable (latent) dynamic state variables $x_j(t_i; \theta, \xi)$ and another is a vector of additional covariates $z$ (which may also depend on time $t$)
Scientific questions

**Inference**
Estimate the vector of parameters $\theta$. When initial values $\xi$ are not known we treat them as additional parameters and estimate them as well.

**Prediction**
Predict the peak height, time of peak, outbreaks (think of infectious diseases), etc.
So what is the problem?

Inference and prediction for dynamical systems is a non-trivial task in practice.

- Dynamic systems are typically very complex:
  - nonlinear
  - high dimensional
- Partial measurements and noisy data
So what is the problem?

- When using non-linear least squares or maximum likelihood we have to evaluate:

\[
M_n(\theta) = \sum_{j=1}^{r} \sum_{i=1}^{n} (Y_j(t_i) - x_j(t_i; \theta))^2.
\]

- requires a good initial guess for parameter optimization

- requires repetitive numerical integration of the system
Estimation techniques

- **Non-linear least squares** - Xue, Miao, and Wu (2010).
- **Many more...** (see Chou and Voit (2009)).

Designed for general cases; require a good initial guess; estimating slopes; tuning.
An important observation is that the system is linear in its parameters.

This observation led to the development of the new estimation techniques presented here and to other tools for statistical inference still under development.
A typical categorization of the estimation problem:
- Linear ODEs
- Non-linear ODEs

An additional point of view:
- Can we exploit linear features of the model?
Main methodological idea
Direct integral estimation with nonparametric smoothing

- Exploiting linear features of the models
- Bypassing (repetitive) numerical integration
- Avoiding estimation of derivative
The class of ODEs systems that are linear in the parameter $\theta$ can be expressed as follows:

$$x'(t) = F(x(t); \theta) = g(x(t))\theta,$$

where $g : \mathbb{R}^d \to \mathbb{R}^{d \times p}$ maps the $d$-dimensional column vector $x$ into a $d \times p$ matrix. Typically $d \leq p$. 
Exploiting linear features of the model

The ODE system is

\[
\begin{align*}
    x'_1(t) &= \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\
    x'_2(t) &= -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t).
\end{align*}
\]

The map \( g \) is given by

\[
g(x(t)) = \begin{pmatrix}
    x_1(t) & -x_1(t)x_2(t) & 0 & 0 \\
    0 & 0 & -x_2(t) & x_1(t)x_2(t)
\end{pmatrix}.
\]
Exploiting linear features of the model

Systems that are linear in functions of the parameter

\[ x'(t) = F(x(t); \nu) = g(x(t))h(\nu) \]

- The FitzHugh-Nagumo system in neurophysiology:

\[
\begin{align*}
    x'_1(t) &= \gamma(x_1(t) - x_1(t)^3/3 + x_2(t)), \\
    x'_2(t) &= -\frac{1}{\gamma}(x_1(t) - \alpha + \beta x_2(t)).
\end{align*}
\]

- Here \( \nu = (\alpha, \beta, \gamma)^T \), and \( h(\nu) = (\gamma, 1/\gamma, \alpha/\gamma, \beta/\gamma)^T \).
Exploiting linear features of the model

Systems that are semi-linear in the parameter

\[ x'(t) = F(x(t); \eta, \theta) = g(\eta, x(t))\theta \]

- Michaelis-Menten kinetics:
  \[
  \begin{align*}
  x'_1(t) &= \theta_1/(\eta_1 + x_3/\eta_2) - \theta_2x_1/(x_1 + \eta_3) - \theta_3x_1/(x_1 + \eta_4), \\
  x'_2(t) &= \theta_2x_1/(x_1 + \eta_3) - \theta_4x_2/(x_2 + \eta_5), \\
  x'_3(t) &= \theta_4x_2/(x_2 + \eta_5) - \theta_5x_3x_4/((x_3 + \eta_6)(x_4 + \eta_7)), \\
  x'_4(t) &= \theta_3x_1/(x_1 + \eta_4) - \theta_6x_4/(x_4 + \eta_8).
  \end{align*}
  \]
A few examples of systems with linear features

Lotka-Volterra system in population dynamics (Edelstein-Keshet (2005)); FitzHugh-Nagumo system in neurophysiology (FitzHugh (1961), Nagumo et al. (1962)); models describing HIV dynamics (Nowak and May (2000), Miao et al. (2008), Miao et al. (2009), Wu et al. (2008), Fang et al. (2011)); models for the blood coagulation process (Hockin et al. (2002)); problems in chemistry (Tjoa and Biegler (1991)); gene regulatory networks (Brewer et al. (2008)); models describing the spread of infectious diseases (He et al. (2010), Hooker et al. (2011)); calcium measurements analysis (Tank et al. (1995)); pharmacokinetic models (de Bazelaire et al. (2005)).
Two-step approach

- Observations are first smoothed which results in estimators \( \hat{x}_j(\cdot) \) and \( \hat{x}'_j(\cdot) \), \( j = 1, ..., d \).
Two-step approach

- Observations are first smoothed which results in estimators $\hat{x}_j(\cdot)$ and $\hat{x}'_j(\cdot)$, $j = 1, \ldots, d$.

- Then the estimator for $\theta$ is the minimizer $\hat{\theta}$ of the smooth criterion function

$$\int_0^1 \| \hat{x}'(t) - F(\hat{x}(t); \eta) \|^2 w(t) dt.$$
Two-step approach

- Observations are first smoothed which results in estimators $\hat{x}_j(\cdot)$ and $\hat{x}'_j(\cdot)$, $j = 1, \ldots, d$.

- Then the estimator for $\theta$ is the minimizer $\hat{\theta}$ of the smooth criterion function

$$
\int_0^1 \| \hat{x}'(t) - F(\hat{x}(t); \eta) \|^2 w(t) dt.
$$

- Note that no initial values are needed.


- $\sqrt{n}$-rate of convergence for $\hat{\theta}$ derived by Gugushvili and Klaassen (2012) who refer to it as "Smooth and match" approach.

- Estimating slopes may be inaccurate in the presence of noise.
Integral estimation

Note that when

\[ x'(t) = F(x(t); \theta) = g(x(t))\theta, \]

we have

System of integral equations

\[ x(t) = \xi + \int_0^t g(x(s)) \, ds \, \theta, \quad t \in [0, 1]. \]
Consider the ODE system that includes only one state:

\[ x'(t) = \theta x(t). \]  

(1)

Given an initial value \( x(0) = \xi \), the solution for this system is \( x(t) = \xi e^{\theta t} \).
Example

Consider the ODE system that includes only one state:

\[ x'(t) = \theta x(t). \tag{1} \]

Given an initial value \( x(0) = \xi \), the solution for this system is \( x(t) = \xi e^{\theta t} \).

The map \( g \) is given by

\[ g(x(t)) = x(t) = \xi e^{\theta t} \]

and by integrating both sides of (1) we obtain

\[ x(t; \theta, \xi) = \xi + \theta \int_0^t \xi e^{\theta s} ds = \xi e^{\theta t}. \]
So we have \( x(t) = \xi + \int_0^t g(x(s)) \, ds \, \theta. \)

Let \( \hat{x}_n(t), t \in [0, 1], \) be an estimator of \( x(t; \theta, \xi) \) based on the observations.

We estimate the parameters \( \theta \) and \( \xi \) by minimizing with respect to \( \eta \in \Theta \) and \( \zeta \in \Xi \)

\[
\int_0^1 \left\| \hat{x}_n(t) - \zeta - \int_0^t g(\hat{x}_n(s)) \, ds \, \eta \right\|^2 \, dt.
\]
Let

\[ \hat{G}_n(t) = \int_0^t g(\hat{x}_n(s)) \, ds \quad t \in [0, 1], \]
\[ \hat{A}_n = \int_0^1 \hat{G}_n(t) \, dt, \quad \hat{B}_n = \int_0^1 \hat{G}_n^T(t) \hat{G}_n(t) \, dt. \]

The general form of the estimators are

\[ \hat{\xi}_n = \left( I_d - \hat{A}_n \hat{B}_n^{-1} \hat{A}_n^T \right)^{-1} \int_0^1 \left( I_d - \hat{A}_n \hat{B}_n^{-1} \hat{G}_n^T(t) \right) \hat{x}_n(t) \, dt, \]
\[ \hat{\theta}_n = \hat{B}_n^{-1} \int_0^1 \hat{G}_n^T(t) \left( \hat{x}_n(t) - \hat{\xi}_n \right) \, dt. \]
Results for fully observed systems
Identifiability

- Identifiability is an important issue!
- $\theta$ is identifiable if for any $\theta \neq \theta'$ we have $x(\cdot; \theta, \xi) \neq x(\cdot; \theta', \xi)$.
- Verifying identifiability of a model is not trivial.
- Using the framework of linearity in functions of the parameters - it is possible to state a necessary and sufficient condition for identifiability of the parameter $\theta$ and the initial values $\xi$.
- A surprising result for identifiability - identifiability of the parameter $\theta$ is not equivalent to uniqueness of the solution $x(t; \theta, \xi)$ as a function of $t$. 
Theorem (D & Klaassen (2015))

Assume that $G(t) = \int_0^t g(x(s)) \, ds$, $A = \int_0^1 G(t) \, dt$ and $B = \int_0^1 G^\top(t)G(t) \, dt$ are well-defined and finite.

- If $B$ is nonsingular, then $I_d - AB^{-1}A^\top$ is nonsingular and

\[
\dot{\xi} = \left( I_d - AB^{-1}A^\top \right)^{-1} \int_0^1 \left( I_d - AB^{-1}G^\top(t) \right) x(t) \, dt,
\]

\[
\theta = B^{-1} \int_0^1 G^\top(t) (x(t) - \xi) \, dt.
\]

- Conversely, if $x(t)$, $t \in [0, 1]$, determines $\theta$, then $B$ is nonsingular.
Consistency

**Theorem (D & Klaassen (2015))**

- Assume that the map $g : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^p$ is continuous.
- Suppose that
  \[ \sup_{t \in [0,1]} \| x(t) \| < \infty. \]
- Let the matrix $B = \int_0^1 G(t)^\top G(t) dt$ be nonsingular.

If
  \[ \sup_{t \in [0,1]} \| \hat{x}_n(t) - x(t; \xi, \theta) \| \xrightarrow{P} 0, \]

then
  \[ (\hat{\theta}_n, \hat{\xi}_n) \xrightarrow{P} (\theta, \xi) \]

holds as $n \to \infty$. 

Itai Dattner
Learning Dynamic Systems
We stated sufficient conditions for $\sqrt{n}$-consistency of estimators of systems linear in functions of the parameter (e.g., FitzHugh-Nagumo system in neurophysiology). In particular:

**Theorem (D & Klaassen (2015))**

Let $\theta$ be identifiable, i.e., for any $\theta \neq \theta'$ we have $x(\cdot; \theta, \xi) \neq x(\cdot; \theta', \xi)$. Let $\hat{x}_n$ be a local polynomial estimator. Then for an appropriate choice of bandwidth (over-fitting), and under regularity conditions on $x$ and $g$,

$$\sqrt{n}(\hat{\theta}_n - \theta, \hat{\xi}_n - \xi) = O_p(1), \; n \to \infty.$$ 

**Note:** we have to decide if we want to "predict" the process or "explain" it.
Asymptotic properties of an estimator for $\theta$ and $\xi$ will depend on those of $\hat{x}_n(\cdot)$, the estimator for the solution $x(\cdot)$ of the system.
Some remarks

- Asymptotic properties of an estimator for $\theta$ and $\xi$ will depend on those of $\hat{x}_n(\cdot)$, the estimator for the solution $x(\cdot)$ of the system.

- Asymptotic properties of an estimator $\hat{x}_n(\cdot)$ for $x(\cdot)$ will depend (among other things) on the mapping $g$ - recall that $x'(t) = g(x(t))\theta$. 
Asymptotic properties of an estimator for $\theta$ and $\xi$ will depend on those of $\hat{x}_n(\cdot)$, the estimator for the solution $x(\cdot)$ of the system.

Asymptotic properties of an estimator $\hat{x}_n(\cdot)$ for $x(\cdot)$ will depend (among other things) on the mapping $g$ - recall that $x'(t) = g(x(t))\theta$.

The more "nicer" $g$ is, the less restrictions are needed for $\hat{x}_n(\cdot)$. 
Estimating parameters of the FitzHugh-Nagumo system

This model is used in neurophysiology as an approximation of the observed spike potential and takes the form

\[
\begin{align*}
    x'_1(t) &= \gamma(x_1(t) - x_1^3(t) + x_2(t)), \\
    x'_2(t) &= -(1/\gamma)(x_1(t) - \alpha + \beta x_2(t)).
\end{align*}
\]

The voltage \( x_1(t) \) moving across the cell membrane depends on the recovery variable \( x_2(t) \).

The true parameter vector is \( (\alpha, \beta, \gamma)^T = (0.34, 0.2, 3)^T \).

The two signals are first generated by solving the system at 0.1 time units on the interval \([0, 20]\) \((n = 201)\) and then adding Gaussian measurement errors with zero mean and variances \( \sigma_1^2, \sigma_2^2 \) respectively.
We present a comparison of the average relative estimation error (ARE) of the integral approach with that of the derivative based two-step method. The ARE of a real-valued parameter $a$ over the $M$ Monte Carlo simulations is defined as

$$\text{ARE}(a) = \frac{1}{M} \sum_{m=1}^{M} \frac{|\hat{a}_m - a|}{|a|} \times 100\%,$$

where $\hat{a}_m$ is an estimator of $a$ in simulation $m$, and in our case $M = 500$.

We use local polynomial estimators for $\hat{x}_n(\cdot)$.
Comparison study of average relative estimation error (ARE)

Here $\sigma_1^2 = 0.070$

<table>
<thead>
<tr>
<th>$\sigma_2^2$</th>
<th>Integral ARE</th>
<th>Derivative ARE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>0.050</td>
<td>1.235</td>
<td>9.447</td>
</tr>
<tr>
<td>0.060</td>
<td>1.315</td>
<td>10.621</td>
</tr>
<tr>
<td>0.070</td>
<td>1.390</td>
<td>11.728</td>
</tr>
<tr>
<td>0.080</td>
<td>1.430</td>
<td>12.857</td>
</tr>
<tr>
<td>0.090</td>
<td>1.494</td>
<td>13.554</td>
</tr>
<tr>
<td>0.100</td>
<td>1.482</td>
<td>14.902</td>
</tr>
</tbody>
</table>
Integral estimator as an initial guess

The data are obtained from the R package `CollocInfer` (FitzHugh-Nagumo system) and comprises 41 equally spaced observations on the interval [0,20]. Initial guesses for the generalized profiling (GP) are $10^{i-1}u$, $i = 0, \ldots, 5$, with $u = (1, 1, 1)$.

<table>
<thead>
<tr>
<th>Method (IG)</th>
<th>$a = 0.2$</th>
<th>$b = 0.2$</th>
<th>$c = 3$</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integral(NO)</td>
<td>0.318</td>
<td>-0.140</td>
<td>3.003</td>
<td>0.011</td>
</tr>
<tr>
<td>GP(Integral)</td>
<td><strong>0.237</strong></td>
<td><strong>0.144</strong></td>
<td><strong>3.050</strong></td>
<td><strong>25.466</strong></td>
</tr>
<tr>
<td>GP(0.1u)</td>
<td>0.239</td>
<td>0.009</td>
<td>0.069</td>
<td>8.687</td>
</tr>
<tr>
<td>GP(1u)</td>
<td>0.144</td>
<td>0.567</td>
<td>3.461</td>
<td>55.725</td>
</tr>
<tr>
<td>GP(10u)</td>
<td>1.797</td>
<td>3.082</td>
<td>3.295</td>
<td>123.914</td>
</tr>
<tr>
<td>GP(100u)</td>
<td>1.829</td>
<td>3.142</td>
<td>3.251</td>
<td>151.203</td>
</tr>
<tr>
<td>GP(1000u)</td>
<td><strong>0.237</strong></td>
<td><strong>0.144</strong></td>
<td><strong>3.050</strong></td>
<td><strong>380.798</strong></td>
</tr>
</tbody>
</table>
One can avoid smoothing in case of the following experimental design:

\[ Y^{(j)}(t_i) = x(t_i; \theta, \xi) + \varepsilon^{(j)}_i, \quad j = 1, \ldots, J_i, \quad i = 1, \ldots, I, \]

with \( t_i = i/I, \quad i = 1, \ldots, I. \)
One can avoid smoothing in case of the following experimental design:

\[ Y^{(j)}(t_i) = x(t_i; \theta, \xi) + \varepsilon_i^{(j)}, \quad j = 1, \ldots, J_i, \quad i = 1, \ldots, I, \]

with \( t_i = i/I, \quad i = 1, \ldots, I \).

Hence, we observe \( J_i \) repeated measures of \( x(t_i) \) for each time point \( t_i \), which means that we have \( n = \sum_{i=1}^{I} J_i \) observations in total.
One can avoid smoothing in case of the following experimental design:

\[ Y^{(j)}(t_i) = x(t_i; \theta, \xi) + \varepsilon_i^{(j)}, \quad j = 1, \ldots, J_i, \quad i = 1, \ldots, I, \]

with \( t_i = i/I, \ i = 1, \ldots, I \).

Hence, we observe \( J_i \) repeated measures of \( x(t_i) \) for each time point \( t_i \), which means that we have \( n = \sum_{i=1}^{I} J_i \) observations in total.

This is common practice in many fields and therefore makes a reasonable experimental setup.
Step function estimator

Let

\[ \hat{x}_I(t) = \frac{1}{J_i} \sum_{j=1}^{J_i} Y^{(j)}(t_i), \quad (i - 1)/l < t \leq i/l, \quad i = 1, \ldots, l, \]

\[ \hat{G}_I(t) = \frac{1}{I} \sum_{\ell=1}^{i} g(\hat{x}_I(t_\ell)), \quad (i - 1)/l < t \leq i/l, \quad i = 1, \ldots, l, \]

and consider the criterion function

\[ \frac{1}{l} \sum_{i=1}^{l} \left\| \hat{x}_I(t_i) - \zeta - \hat{G}_I(t_i) \eta \right\|^2 = \int_0^1 \left\| \hat{x}_I(t) - \zeta - \hat{G}_I(t) \eta \right\|^2 dt. \]
Step function estimator

Let

\[ \hat{A}_l = \frac{1}{l} \sum_{i=1}^{l} \hat{G}_l(t_i), \]

\[ \hat{B}_l = \frac{1}{l} \sum_{i=1}^{l} \hat{G}_l(t_i)^T \hat{G}_l(t_i). \]

Minimizing the criterion function with respect to \( \zeta \) and \( \eta \) results in the direct estimators

\[ \hat{\xi}_l = \left( I_d - \hat{A}_l \hat{B}_l^{-1} \hat{A}_l^T \right)^{-1} \frac{1}{l} \sum_{i=1}^{l} \left( I_d - \hat{A}_l \hat{B}_l^{-1} \hat{G}_l(t_i)^T \right) \hat{x}_l(t_i), \]

\[ \hat{\theta}_l = \hat{B}_l^{-1} \frac{1}{l} \sum_{i=1}^{l} \hat{G}_l(t_i)^T \left( \hat{x}_l(t_i) - \hat{\xi}_l \right), \]
Step function estimator

**Theorem (D & Klaassen (2015))**

Assume that,

- $\theta$ is identifiable, $g$ is twice differentiable and $x$ is bounded.
- Errors are i.i.d. with zero expectation and finite variance.
- The sample sizes satisfy

$$\lim \inf_{n \to \infty} \frac{l}{\sqrt{n}} > 0, \quad \lim \sup_{n \to \infty} \sum_{i=1}^{l} \frac{1}{J_i} < \infty, \quad \sum_{i=1}^{l} J_i = n.$$ 

If the second derivatives of each component of $g$ are continuous or bounded, then

$$\sqrt{n}(\hat{\theta}_l - \theta, \hat{\xi}_l - \xi) = O_p(1), \quad n \to \infty.$$
Some remarks

Note that \( \text{var}(\hat{x}_I(0)) = \sigma^2_\xi / J_1 \) holds, and that estimating \( \xi \) via \( \hat{x}_I(0) \) would not yield the best possible rate, unless \( J_1 \) is of exact order \( n \).

Indeed, the \( \sqrt{n} \)-rate is achievable by \( \hat{\xi}_I \) using the information from all \( I \) time points.

Proof follows by showing that the required conditions of the general theorem mentioned above for convergence rate hold true.

Other estimators can be analyzed, e.g., Vujačić, D, González and Wit (2015).
Estimating parameters of the Lotka-Volterra system with no repeated measures ($I = 100, \sigma_1 = \sigma_2 = 0.5$) - initial values are estimated
Estimating parameters of the Lotka-Volterra system with 10 repeated measures ($I = 100, \sigma_1 = \sigma_2 = 0.5$) - initial values are estimated.
Estimating parameters of the Lotka-Volterra system - initial values are estimated

<table>
<thead>
<tr>
<th></th>
<th>( J = 6 )</th>
<th>( J = 15 )</th>
<th>( J = 30 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi_1 )</td>
<td>1.000 (0.143)</td>
<td>1.085 (0.093)</td>
<td>1.083 (0.065)</td>
</tr>
<tr>
<td>( \xi_2 )</td>
<td>0.500 (0.096)</td>
<td>0.438 (0.061)</td>
<td>0.436 (0.043)</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>0.500 (0.077)</td>
<td>0.474 (0.050)</td>
<td>0.477 (0.035)</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>0.500 (0.075)</td>
<td>0.479 (0.048)</td>
<td>0.480 (0.034)</td>
</tr>
<tr>
<td>( \theta_3 )</td>
<td>0.500 (0.073)</td>
<td>0.501 (0.047)</td>
<td>0.501 (0.033)</td>
</tr>
<tr>
<td>( \theta_4 )</td>
<td>0.500 (0.073)</td>
<td>0.509 (0.047)</td>
<td>0.509 (0.033)</td>
</tr>
</tbody>
</table>

Table: The empirical mean and standard deviation (in parenthesis) of the estimators for the parameters and initial values of the Lotka-Volterra system. Results are based on 5000 Monte Carlo simulations. The samples were taken at 0.5 time units on the interval \([0, T]\), \( T = 14.9 \) (\( J = 30 \)). The errors are i.i.d Gaussian with zero mean and \( \sigma_\varepsilon = 0.5 \).
HIV dynamics

- Experimental data of viral fitness for HIV-1 mutant 103N,

\[
x_1'(t) = (\theta_1 - \theta_2 x_2(t) - \theta_3 x_3(t) - \theta_4 x_4(t)) x_1(t),
\]
\[
x_2'(t) = (\theta_5 + \theta_2 x_1(t) - \theta_6 x_3(t)) x_2(t) + 0.25 \theta_4 x_4(t) x_1(t),
\]
\[
x_3'(t) = (\theta_7 + \theta_3 x_1(t) - \theta_8 x_2(t)) x_3(t) + 0.25 \theta_4 x_4(t) x_1(t),
\]
\[
x_4'(t) = (\theta_9 + 0.5 \theta_4 x_1(t)) x_4(t) + (\theta_6 + \theta_8) x_3(t) x_2(t).
\]

- The states \( x_1, x_2, x_3, x_4 \) represent the numbers of uninfected cells, cells infected by mutant virus, cells infected by wildtype virus, and cells infected by both mutant and wildtype viruses, respectively.

- The system was measured for only 5 time (in hours) points \{70, 94, 115, 139, 163\} and the data consist of three repeated observations at each time point.
HIV dynamics - estimation using cubic $B$-splines

**Figure:** Viral fitness for HIV-1 mutant 103N data. **Left panel.** The plus signs and circles correspond to the repeated observations. The solid, dotted, dashed, and dashed-dotted lines correspond to $x_1$, $x_2$, $x_3$ and $x_4$ as solved for initial values $\hat{\theta}_0$ (0.1572 sec). **Right panel.** The same as the left but for $x_1$, $x_2$, $x_3$ and $x_4$ solved for the NLS parameter estimates. The NLS procedure was initiated with the $\hat{\theta}_0$ used to generate the plots in the left panel (30.765 sec).
Results for partially observed systems
Partially observed systems

- Observations are given by
  \[ Y_j(t_i) = x_j(t_i; \theta, \xi) + \varepsilon_{j,i}, \]
  where \( i = 1, \ldots, n, j = 1, \ldots, r \) and \( r < d \).

- Let \( \hat{x}_u(t) := (\hat{m}_n(t), u(t))^T, t \in [0, 1] \).
- Let \( \hat{G}_u(t) = \int_0^t g(\hat{x}_u(s))ds, t \in [0, 1] \).
- Let \( \hat{B}_u = \int_0^1 \hat{G}_u^T(t)\hat{G}_u(t)dt \).

- Define
  \[ \hat{\xi}_u = (TI_d - \hat{A}_u\hat{B}_u^{-1}\hat{A}_u^T)^{-1} \int_0^T \{ I_d - \hat{A}_u\hat{B}_u^{-1}\hat{G}_u^T(t) \} \hat{x}_u(t)dt, \]
  \[ \hat{\theta}_u = \hat{B}_u^{-1} \int_0^T \hat{G}_u^T(t)(\hat{x}_u(t) - \hat{\xi}_u)dt. \]
Sieve estimation of partially observed systems

Let $\mathcal{U}$ be some appropriate space of functions and define

$$\hat{u}_n := \arg \min_{u \in \mathcal{U}} M_n(u),$$

where

$$M_n(u) = \int_{0}^{T} \| \hat{x}_u(t) - \hat{\xi}_u - \hat{G}_u(t)\hat{\theta}_u \|^2 dt.$$

Estimator for $\theta$ and $\xi$ are given by

$$\hat{\xi}_n := \hat{\xi}_{\hat{u}_n},$$
$$\hat{\theta}_n := \hat{\theta}_{\hat{u}_n}.$$
Partially observed systems - sieve estimation

**Theorem (Vujačić & D (2017))**

Consistency of direct integral sieve estimator for partially observed systems is established under the following main assumptions:

1. Existence and uniqueness (and boundedness) of the solution $x$. Continuity of $g$.
2. Identifiability of $\theta$.
3. Consistency (in the sup norm) of the estimator for the observed part.
4. Densness and compactness of sieves.
5. Compactness of function spaces.
Neural behavior - data of transmembrane potential of zebra-fish motor neuron
The following system arises in the study of neural behavior

\[
x'_1(t) = \theta_1 + \theta_2 x_1 + \theta_3 x_1^2 + \theta_4 x_1^3 + \theta_5 x_2 + \theta_6 x_1 x_2,
\]
\[
x'_2(t) = \theta_7 + \theta_8 x_1 + \theta_9 x_2.
\]

The state \( x_1 \) represents the transmembrane potential while \( x_2 \) is a recovery variable representing a collection of ion currents. Data are generated by measuring the potential along the axon membrane of a neuron.

So \( x_2 \) is not measured!

We illustrate the application of the new procedure to data coming from experiment performed on a zebra-fish motor neuron.
Neural behavior
Influenza A virus infection

The model is given by

\[ x'_1(t) = -\theta_1 x_1(t)x_3(t), \]
\[ x'_2(t) = \theta_1 x_1(t)x_3(t) - \theta_2 x_2(t), \]
\[ x'_3(t) = \theta_3 x_2(t) - \theta_4 x_3(t). \]

Here \( x_1 \) is the number of uninfected target cells (epithelial cells), \( x_2 \) is the number of productively infected cells, and \( x_3 \) is the infectious viral titer expressed in TCID50/ml.

We set \( \theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T = (0.3, 0.3, 1, 0.5)^T \), \( \xi = (x_1(0), x_2(0), x_3(0))^T = (1, 0.05, 2)^T \) and solve numerically the system over the interval \( 0(0.5)20 \).
Influenza A virus infection

**Figure:** Influenza model for \( \theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T = (0.3, 0.3, 1, 0.5)^T \), 
\( \xi = (x_1(0), x_2(0), x_3(0))^T = (1, 0.05, 2)^T \). The system solutions were generated using a numerical solver over the interval \( 0(0.5)20 \).
Influenza A virus infection

- We present a Monte Carlo study for unobserved \( x_1 \).
- The time points are uniformly spaced and the noise is Gaussian with zero expectation and standard deviation of 0.05.
- The unmeasured state \( x_1 \) is estimated using 5 Laguerre bases. The measured states \( x_2 \) and \( x_3 \) are estimated using local polynomials of order 1 with bandwidth equal to 1.
- The table presents square root of the empirical mean square error and empirical absolute bias (in parentheses); based on 500 simulations for each of the sample sizes of \( n = 12, 21, 51, 101 \).
- Last two lines in each block correspond to the empirical means (standard deviation) of the distribution of
\[
\left\{ \frac{1}{T} \int_0^T (\tilde{x}_1(t) - x_1(t; \theta, \xi))^2 dt \right\}^{1/2} \text{ and } \| \tilde{x}_1(\cdot) - x_1(\cdot; \theta, \xi) \|_\infty, \]
respectively.
<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>$n = 12$</th>
<th>$n = 21$</th>
<th>$n = 51$</th>
<th>$n = 101$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_1$</td>
<td>1.000</td>
<td>0.196 (0.125)</td>
<td>0.154 (0.099)</td>
<td>0.118 (0.077)</td>
<td>0.092 (0.065)</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.050</td>
<td>0.058 (0.029)</td>
<td>0.056 (0.033)</td>
<td>0.042 (0.020)</td>
<td>0.030 (0.014)</td>
</tr>
<tr>
<td>$\xi_3$</td>
<td>2.000</td>
<td>0.073 (0.013)</td>
<td>0.060 (0.009)</td>
<td>0.043 (0.010)</td>
<td>0.032 (0.012)</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.300</td>
<td>0.123 (0.060)</td>
<td>0.106 (0.058)</td>
<td>0.065 (0.032)</td>
<td>0.040 (0.017)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.300</td>
<td>0.066 (0.044)</td>
<td>0.054 (0.037)</td>
<td>0.039 (0.026)</td>
<td>0.030 (0.022)</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>1.000</td>
<td>0.333 (0.260)</td>
<td>0.275 (0.209)</td>
<td>0.218 (0.175)</td>
<td>0.197 (0.174)</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>0.500</td>
<td>0.118 (0.096)</td>
<td>0.097 (0.075)</td>
<td>0.078 (0.064)</td>
<td>0.071 (0.064)</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>1.000</td>
<td>0.123 (0.058)</td>
<td>0.086 (0.039)</td>
<td>0.062 (0.036)</td>
<td>0.044 (0.032)</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.050</td>
<td>0.055 (0.025)</td>
<td>0.053 (0.029)</td>
<td>0.041 (0.020)</td>
<td>0.030 (0.016)</td>
</tr>
<tr>
<td>$\xi_3$</td>
<td>2.000</td>
<td>0.056 (0.017)</td>
<td>0.045 (0.011)</td>
<td>0.031 (0.009)</td>
<td>0.022 (0.007)</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.300</td>
<td>0.078 (0.035)</td>
<td>0.067 (0.035)</td>
<td>0.040 (0.022)</td>
<td>0.024 (0.014)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.100</td>
<td>0.021 (0.013)</td>
<td>0.016 (0.009)</td>
<td>0.011 (0.007)</td>
<td>0.008 (0.007)</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>0.100</td>
<td>0.035 (0.001)</td>
<td>0.028 (0.003)</td>
<td>0.018 (0.003)</td>
<td>0.013 (0.005)</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>0.300</td>
<td>0.025 (0.002)</td>
<td>0.021 (0.003)</td>
<td>0.014 (0.004)</td>
<td>0.010 (0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.058 (0.054)</td>
<td>0.049 (0.043)</td>
<td>0.034 (0.028)</td>
<td>0.025 (0.016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.103 (0.089)</td>
<td>0.081 (0.065)</td>
<td>0.059 (0.044)</td>
<td>0.046 (0.026)</td>
</tr>
</tbody>
</table>
Learning age-dependent dynamics of infectious diseases from incidence data
Infectious diseases - learning age-dependent dynamics

**Figure:** Two age-groups: children 0-19 (blue) and adults 20+ (red).
Discrete age-of-infection age group model (single outbreak)

\[ i_j(t) = \frac{S_j(t-1)}{N_j} \sum_{k=1}^{m} \left( \beta_{jk} \sum_{\tau=1}^{d} P_{\tau} i_k(t - \tau) \right), \]

\[ S_j(t) = S_j(t - 1) - i_j(t), \quad t=1,...,T. \]

- \( i_j(t) \) is the number of **newly** infected from age-group \( j \).
- \( S_j(t) \) is the number of susceptible individuals from age-group \( j \).
- Assume infected individual remain infected for \( d \) days.
- The infectiousness of an infected individual changes with the age of his infection according to the serial interval distribution \( P_1, \ldots, P_d, \sum_{\tau=1}^{d} P_{\tau} = 1 \).
Here again:

\[ i_j(t) = \frac{S_j(t - 1)}{N_j} \sum_{k=1}^{m} \left( \beta_{jk} \sum_{\tau=1}^{d} P_{\tau} i_k(t - \tau) \right). \]

In a totally susceptible population, a single infected individual from age-group \( j \) infects on average \( \beta_{jk} \) individuals from age-group \( k \) over the period of his/her infection.
Observations (statistical) model

- It is assumed that the incidence $i(\cdot)$ can not be observed directly but instead, $T$ noisy observations are given by

$$Y_j(t) = i_j(t) + \epsilon_j(t), \quad t = 1, \ldots, T, \quad j = 1, \ldots, m,$$

where $\epsilon_j(t) \sim N(0, \sigma_j^2(t))$, and $\sigma_j(t) = \phi_a + \phi_b \times i_j(t)$ for some two constants $\phi_a, \phi_b$.

- Susceptible $S_j(t)$ are not observed and $S_j(0)$ are not known:

$$S_j(t) = S_j(t - 1) - i_j(t).$$
Exploiting linear features of the model

**Problem**: in order to estimate the so called 'next-generation-matrix' $\beta$ previous studies used constraints for $\beta$ or used questionnaires for estimating contact patterns.

**Solution**: exploiting linear features of the model and using nonparametric smoothing. It turns out that the model is "semi-linear" in its parameters $\beta$ and $S(0)$. This property enables us to obtain theoretical identifiability results and develop a better estimation method.
Exploiting linear features of the model

Recall that

\[ i_j(t) = \frac{S_j(t-1)}{N_j} \sum_{k=1}^{m} \left( \beta_{jk} \sum_{\tau=1}^{d} P_{\tau} i_k(t-\tau) \right), \]

and let \( X_{jk}(t) = \frac{S_j(t-1)}{N_j} \sum_{\tau=1}^{d} P_{\tau} i_k(t-\tau). \)

Given \( t = 1, \ldots, T \) discrete points of the incidence defined above, we can write

\[
\begin{pmatrix}
i_1(1) \\
i_1(T) \\
\vdots \\
i_m(1) \\
i_m(T)
\end{pmatrix}
= 
\begin{pmatrix}
X_{11}(1) & \cdots & X_{1m}(1) & 0 & \cdots & 0 \\
0 & \cdots & 0 & X_{m1}(1) & \cdots & X_{mm}(1) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & X_{m1}(T) & \cdots & X_{mm}(T)
\end{pmatrix}
\times
\begin{pmatrix}
\beta_{11} \\
\beta_{1m} \\
\vdots \\
\beta_{m1} \\
\beta_{mm}
\end{pmatrix}
\]
Denote the matrix with the components $X_{jk}(t)$ by $X$. Then by structural identifiability we mean that the following relation holds true:

$$\beta = (X^T X)^{-1} X^T i,$$

where $i = (i_1(1), \ldots, i_1(T), \ldots, i_m(1), \ldots, i_m(T))^T$, and

$$\beta = (\beta_{11}, \ldots, \beta_{1m}, \ldots, \beta_{m1}, \ldots, \beta_{mm})^T.$$

Let $B = X^T X$. Structural identifiability of the transmission model holds if and only if the matrix $B$ is non-singular.
For example, for $m = 2$ age groups, $i_1(t) = c \times i_2(t)$ for some constant $c$ for all $t$ leads to singularity of the matrix $B$.

In scenarios which are in a close vicinity to a singularity scenario (when $i_1 \approx i_2$), the model, while mathematically identifiable, becomes ‘hard’ to estimate in practice.
Structural (mathematical) identifiability
Let $Y = (Y_1(1), ..., Y_1(T), ..., Y_m(1), ..., Y_m(T))^\top$ and 
$\varepsilon = (\varepsilon_1(1), ..., \varepsilon_1(T), ..., \varepsilon_m(1), ..., \varepsilon_m(T))^\top$.

Then the model takes the form

$$ Y = X\beta + \varepsilon, $$

where $X_{jk}(t) = S_j(t - 1)/N_j \sum_{\tau=1}^{d} P_{\tau} i_k(t - \tau)$.

The components of $X$ ($S$ and $i$) are not directly observed and therefore we can not compute the least squares solution

$$ \tilde{\beta} = (X^\top X)^{-1} X^\top Y. $$
Estimation method

- We smooth the observations $Y$ to get $\hat{i}$, and assuming $S_j(0)$ are known, we obtain

$$\hat{X}_{jk} = S_j(t - 1)/N_j \sum_{\tau=1}^d P_{\tau} \hat{i}_k(t - \tau),$$

and set

$$\hat{\beta} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top \hat{i}.$$ 

- This estimator is not statistically efficient and therefore we use it as an initial guess for finding the maximum likelihood $\hat{\beta}_{ML}$; we call this methodology a "two-stage approach".

- Actually our model is **semi-linear in the parameters**: linear in $\beta$ but nonlinear in $S_j(0)$. Thus, nonlinear optimization can not be avoided when estimating the ’nonlinear parameters’.
Incidence obtained using three $2 \times 2 \beta$ matrices.

**Figure:** Incidence for two age-groups obtained from the transmission model and 500 Monte-Carlo simulations of the observation model using three selected $2\times2$ $\beta$ matrices.
We consider estimation of $\beta$ under two scenarios:

1. Fitting incidence of a single epidemic assuming the initial fraction of susceptible population in each age-group $S_j(0)$ are known.

2. Fitting incidence of a single epidemic while estimating the 'nonlinear parameters’ $S_j(0)$.

<table>
<thead>
<tr>
<th>Matrix 1</th>
<th>Scenario 1 (MSE = 0.027)</th>
<th>Scenario 2 (MSE = 0.354)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>std</td>
</tr>
<tr>
<td>$\beta_{11} = 2.50$</td>
<td>2.51</td>
<td>0.06</td>
</tr>
<tr>
<td>$\beta_{12} = 0.75$</td>
<td>0.74</td>
<td>0.07</td>
</tr>
<tr>
<td>$\beta_{21} = 1.00$</td>
<td>1.01</td>
<td>0.09</td>
</tr>
<tr>
<td>$\beta_{22} = 1.50$</td>
<td>1.48</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Estimating $\beta$ from simulated data

**Figure:** The mean and std of the estimates of $\beta$ under the two scenarios, for matrix1 and matrix3. The estimates for matrix1 are much better than the estimates for matrix3, as expected.
For an infectious disease with repeated annual epidemics such as influenza, we can improve our estimates by fitting the model to recurrent outbreaks while assuming some model parameters remain fixed over the years.

Specifically, age-group contact patterns and the (possibly age-dependent) probabilities of transmission or infection upon contact, should remain more or less the same over a period of several years.

However, the transmissibility of the prevailing influenza virus in a specific year could change from year to year but should affect individuals of all age-groups in a similar manner.
Fit to eleven years of ILI data from Israel

**Figure:** Two age-groups: children 0-19 (blue) and adults 20+ (red). Black curves show the obtained fit using our two-stage method.
The estimated $\beta$ for the first year (the 1998 season):

$$\hat{\beta}_{ML} = \begin{pmatrix}
2.05[1.97, 2.13] & 0.11[0.08, 0.14] \\
0.30[0.26, 0.33] & 1.85[1.76, 1.94]
\end{pmatrix},$$

with the 95% CI for each component of the matrix obtained using profile likelihood given in parenthesis.

The obtained matrix is highly assortative, indicating that most of the transmission occurs within the two age-groups, with the highest transmission being from children to children and the lowest transmission from adults to children.

According to this matrix, in an entirely susceptible population a child would infect, on average, 1.2 times more individuals than an adult would.
Using the linear (in the parameters) representation of the model we:

- obtain theoretical identifiability properties that may suggest how to partition incidence data of the whole population into age-groups.

- estimate $\beta$ using incidence data **without making assumptions on the structure of the matrix**.

- develop an estimation method that is both statistically and computationally better than naively using the likelihood. **Maximum-likelihood optimization starting from random initial guess lead to estimates with much higher MSE and worse coverage of confidence intervals.**
Modelling and parameter inference of predator-prey dynamics in heterogeneous environments using the direct integral approach
In this study we quantify the dynamics of the predatory bacterium Bdellovibrio bacteriovorus (BALOs) and its prey, Burkholderia stabilis st.2 in structured habitat consisting of sand under various regimes of wetness.

While the role of spatial structures of predator-prey interactions can be of fundamental importance to microbial ecology, epidemiology and agronomy, many of the mathematical methods developed to deal with spatial heterogeneities in population dynamics models are difficult to parametrize and analyze.

We developed a simple model that can both qualitatively and quantitatively describe the temporal dynamics of a predator-prey system in the heterogeneous soil environment.
According to our model, not all the prey is available to the predator since some of it has a refuge. Prey refuge may occur in our setup due to habitat spatial heterogeneity consisting of soil grains.

Most evidence of the effect of prey refuge on the temporal dynamics of predator-prey systems are based on theoretical works and simulations. To the best of our knowledge, this is the first attempt to fit a BALO-prey model to experimental data using a statistical procedure.

The ability of the model to fit the data combined with realistic parameter estimates indicate that bacterial predation in the sand can be described by a relatively simple model, and stress the importance of prey refuge on predation dynamics in heterogeneous environments.
Learning predator-prey dynamics

We use simple ODEs with an implicit spatial (refuge) parameter.

\[ P' = ksC - dP, \quad \text{(predators)} \]
\[ C' = a(N - r)P - sC, \quad \text{(complex)} \]
\[ N' = -a(N - r)P. \quad \text{(prey)} \]
Learning predator-prey dynamics

- Upon encounter, the predator penetrates the prey and converts it into a bdelloplast (predator-prey complexes) that provides nutrient and shelter to the predator during its growth and division to progeny cells, that finally burst out to start a new cycle.

- In our system the time it takes the predator to handle its prey (i.e. searching and invading it to form bdelloplast) is of the same order as the time it takes for the consumed prey items to be converted into new predators.

- To account for this dynamics we explicitly modeled them in a separate compartment that models the delay between the predator invasion and the burst of the bdelloplast (which releases the predator progeny).
Learning predator-prey dynamics

\[ P' = ksC - dP, \text{ (predators)} \]
\[ C' = a(N - r)P - sC, \text{ (complex)} \]
\[ N' = -a(N - r)P. \text{ (prey)} \]

- The predators are produced at the rate \( ksC \), where \( k \) is the number of predatory bacteria emitted by one bdelloplast and \( s \) is the bdelloplast decay rate (both per hour).

- We assume for simplicity that the predator has a density-independent death rate \( d \), and the complex formation rate is equal to \( aNP \), where \( a \) is the interaction (contact) rate which has the units of area/(unit time): \( a = cp \), where \( c \) is the area scanned by a single predator per unit time multiplied by \( p \), the probability that it will interact/catch a prey item within that scanned area.
Learning predator-prey dynamics

\[ P' = ksC - dP, \quad \text{(predators)} \]

\[ C' = a(N - r)P - sC, \quad \text{(complex)} \]

\[ N' = -a(N - r)P. \quad \text{(prey)} \]

- To account for the prey refuge dynamics, we add an additional new parameter \( r \), denoting refuge prey density, which is not available to the predator. The formation rate of bdelloplast complexes is therefore updated to be \( a(N - r)P \).

- Because in our system the prey is not supplied with nutritional substrate, it neither reproduces nor dies (its death rate is too low to be accounted for in our experimental timescale). Thus, the only prey loss in the model is due to predation, i.e., \( a(N - r)P \).
Initial conditions are known and given by
\[ \xi = (P(0), C(0), N(0))^\top = (10^6, 0, 10^8)^\top. \]

We assume that the unknown system states can be approximated for any \( t \in [0, T] \) by a linear combination of cubic B-spline functions denoted by \( \phi_k(t) \), \( k = 1, \ldots, K_\ell, \ell \in \{1, 2, 3\} \), namely,

\[
\begin{align*}
P(t) & \approx \sum_{k=1}^{K_1} \beta_{1,k} \phi_k(t), \\
C(t) & \approx \sum_{k=1}^{K_2} \beta_{2,k} \phi_k(t), \\
N(t) & \approx \sum_{k=1}^{K_3} \beta_{3,k} \phi_k(t).
\end{align*}
\]

In that case, estimators for \( P(\cdot) \) and \( N(\cdot) \) are easily obtained. Further, the system is semi-linear in the parameters which allows for reducing all the estimation problem to estimating the parameter \( s \).
Denote by $\tilde{P}_n = (\tilde{P}(t_1), \ldots, \tilde{P}(t_n))^T$ and $\tilde{N}_n = (\tilde{N}(t_1), \ldots, \tilde{N}(t_n))^T$ the observed noisy versions of $P(\cdot)$, and $N(\cdot)$ on the grid $0 \leq t_1 < \cdots < t_n = T < \infty$.

Denote by $\Phi_\ell$ the $n \times K_\ell$ matrix composed of the column vectors $(\phi_j(t_1), \ldots, \phi_j(t_n))^T$, $j = 1, \ldots, K_\ell$, $\ell \in \{1, 3\}$.

Then we define

$$\hat{\beta}_1 := (\hat{\beta}_{1,1}, \ldots, \hat{\beta}_{1,K_1})^T = (\Phi_1^T \Phi_1)^{-1} \Phi_1^T \tilde{P}_n,$$

$$\hat{\beta}_3 := (\hat{\beta}_{3,1}, \ldots, \hat{\beta}_{3,K_3})^T = (\Phi_3^T \Phi_3)^{-1} \Phi_3^T \tilde{N}_n,$$

and the estimators for $P(\cdot)$ and $N(\cdot)$ are then given by

$$\hat{P}(t) = \sum_{k=1}^{K_1} \hat{\beta}_{1,k} \phi_k(t), \quad t \in [0, T],$$

$$\hat{N}(t) = \sum_{k=1}^{K_3} \hat{\beta}_{3,k} \phi_k(t), \quad t \in [0, T].$$
According to the system we have $N(t) = r$ whenever $dN/dt = 0$, which in our case occurs in the tail (end of experiment).

However, data at the tails are sparse with only two or three observations.

Thus, in each experiment we use the mean of the last two observations as an estimator for $r$. 
Looking at the system we see that

\[ N(t) - N(0) = -a \int_0^t N(u) P(u) du + ar \int_0^t P(u) du. \]

Let \( \hat{G}_3 = (- \int_0^{t_i} \hat{N}(u) \hat{P}(u) du + r \int_0^{t_i} \hat{P}(u) du), \) and \( \hat{G}_3 \) be the \( m \times 1 \) vector \( (\hat{G}_{31}, \ldots, \hat{G}_{3m})^T \).

Let \( \hat{N} = (\hat{N}(t_1) - N(0), \ldots, \hat{N}(t_m) - N(0))^T. \)

Then we have

\[ \hat{a} = (\hat{G}_3^T \hat{G}_3)^{-1} \hat{G}_3^T \hat{N}. \]
Predator-prey dynamics - estimating $k$ and $d$ for a given $s$

According to system and our approximation scheme we have

$$N(t) - N(0) \approx \sum_{k=1}^{K_2} [-\phi_k(t) - s \int_0^t \phi_k(u) du] \beta_{2,k}.$$ 

Let $\Phi_i = (-\phi_1(t_i) - s \int_0^{t_i} \phi_1(u) du, \ldots, -\phi_K(t_i) - s \int_0^{t_i} \phi_K(u) du)$, and set $\Phi_s$ to be the $m \times K_2$ matrix with rows $\Phi_1, \ldots, \Phi_m$.

Then we define $\hat{\beta}_2$ to be

$$\hat{\beta}_2(s) = (\Phi_s^T \Phi_s)^{-1} \Phi_s^T \hat{N}.$$ 

Finally, we plug in to get

$$\hat{C}(t; s) = \sum_{k=1}^{K_2} \hat{\beta}_{2,k}(s) \phi_k(t), \ t \in [0, T].$$
Predator-prey dynamics - estimating $k$ and $d$ for a given $s$

- Define $\hat{P} = (\hat{P}(t_1) - P(0), \ldots, \hat{P}(t_m) - P(0))^T$.
- Let $\hat{G}_{1i} = (s \int_0^{t_i} \hat{C}(u; s)du, -\int_0^{t_i} \hat{P}(u)du)$.
- Set $\hat{G}_1$ to be the matrix with rows $\hat{G}_{11}, \ldots, \hat{G}_{1m}$.
- Then, we obtain
  $$(\hat{k}(s), \hat{d}(s))^T = (\hat{G}_1^T \hat{G}_1)^{-1} \hat{G}_1^T \hat{P}.$$
Predator-prey dynamics - estimating $s$

- Again, using the structure of the system (first and second equations) we suggest a natural loss function for $s$

$$\hat{s} = \arg \min_s M(s),$$

where we have defined

$$M(s) = \sum_{i=1}^{m} \left\{ \hat{P}(t_i) - P(0) + \hat{N}(t_i) - N(0) ight\}$$

$$-s(k(s) - 1) \int_{0}^{t_i} \hat{C}(u; s)du + \hat{C}(t_i; s)$$

$$+ \hat{d}(s) \int_{0}^{t_i} \hat{P}(u)du \right\},$$

- Finally, let

$$\hat{k} := k(s),$$

$$\hat{d} := d(s).$$
Estimating $s$

Using the structure of the system we suggest a loss function for $s$ which looks like:

![Graph showing the loss function $\log_{10}(M(s))$ against $s$. The graph has a logarithmic y-axis and a linear x-axis, with values ranging from 0 to 0.5 for $s$. The function starts high and quickly decreases, then plateaus at a lower value.]
We take the number of bases $K_1, K_2, K_3$ to be (potentially) different for each function in order to allow for better estimation flexibility.

Specifically, we choose the combination $K = \{ K_1, K_2, K_3 \}$ that minimizes the squared distance between the observations and the solutions of the system

$$
\sum_{i=1}^{n} \left\{ (\tilde{P}(t_i) - P(t_i; \hat{\theta}_K))^2 + (\tilde{N}(t_i) - N(t_i; \hat{\theta}_K))^2 \right\}.
$$

This regularization yields very good results in practice and was proved to be optimal in a similar context.
Spline approximations

Figure: \( K_{opt} = \{11, 5, 7\} \).
Estimation results

Figure: Solution of the system using the estimated parameter vector
\( \hat{\theta}_n = (\hat{k}, \hat{s}, \hat{d}, \hat{a}, \hat{r})^\top = (4.5, 0.041, 0.027, 6.8 \times 10^{-9}, 46000)^\top \).
Learning predator-prey dynamics - conclusions

- A challenging experimental setup: sparse, noisy, nonlinear, partially observed. **Direct integral approach** allows for reducing estimation to only one parameter.

- Predation dynamics can be captured by a relatively **simple dynamical system**. We verified that one of the parameter estimates \((k)\) was consistent with its known, directly measured value from the literature. An important insights for **experimental design**: information about the refuge parameter can be recovered mostly from the ’tail’.

- **Limitations**: our analysis is based on a small number of observations, so the conclusions should be viewed as encouragement for further research using more data points which will allow for better inference, and **reliable confidence interval**.
Takeaway message

- **Systems of differential equations** - powerful mathematical models for dynamic processes

- **An important insight** - exploiting linear features of the models

- **Bypassing computational and numerical difficulties** - direct integral estimation with nonparametric smoothing
Some related references


