

Estimating Dynamic Models from Noisy Data

Jiguo Cao

Department of Statistics and Actuarial Science,
Simon Fraser University, Vancouver, BC, Canada

12th IMS New Researchers' Conference
July 28 - July 31
Baltimore, MD

Collaborators

- James O. Ramsay, McGill University
- Jianhua Huang, Texas A&M University
- Giles Hooker, Cornell University
- Hulin Wu, University of Rochester
- David Campbell, Simon Fraser University

My research area

- Functional Data Analysis
- Estimating Dynamic Models from noisy data
- Nonparametric Smoothing
- Statistical Genetics

Outline

- 1 Introduction for dynamic Models
- 2 HIV-Dynamics
- 3 A Predator-prey system
- 4 Gene Regulatory Network
- 5 Method

Outline

- 1 Introduction for dynamic Models**
- 2 HIV-Dynamics
- 3 A Predator-prey system
- 4 Gene Regulatory Network
- 5 Method

What are Ordinary Differential Equations (ODEs)?

A general form for ODEs:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{g}(\mathbf{x}|\beta)$$

- Variable: $\mathbf{x}(t)$;
- ODE parameter: β ;
- ODEs relate the **rate of change** ($\frac{d\mathbf{x}}{dt}$) of a process to its current state (\mathbf{x})

Why Using Ordinary Differential Equations (ODEs)?

- ODEs model the **rate of change** ($\frac{d\mathbf{x}}{dt}$) of a process
- Many models are given directly in ODE forms in Engineering, Physics, Biology, ...
- Newton's Second Law: $F = m * \frac{d^2}{dt^2} \mathbf{x}(t)$
- Nonlinear ODEs can provide **simple** models for **complex** behavior
- **NSF: Future of Statistics! (Prof. Karen Bandeen-Roche, JHU)**

Outline

- 1 Introduction for dynamic Models
- 2 HIV-Dynamics**
- 3 A Predator-prey system
- 4 Gene Regulatory Network
- 5 Method

Dynamic models for HIV-dynamics

- Ho et al. (1995)

$$\frac{d}{dt}X(t) = P(t) - cX(t)$$

- $X(t)$: HIV-1 RNA concentrations (viral load) in plasma
- $P(t)$: virus production rate
- c : is the clearance rate of free virions
- $P(t) = 0$: drug completely blocking virus production (not valid)
- treatment efficacy may vary due to many factors: immune status (CD4+ T-cell counts), drug resistance, etc

Incorporating immune status

- Ho et al. (1995)

$$\frac{d}{dt}X(t) = P(t) - cX(t)$$

$$P(t) = a_1(t) + a_2(t)CD4(t)$$

- CD4+ T-cell count linearly related to virus production rate
- CD4+ T-cell count effect is time-varying

Data

- The differential equation

$$\frac{d}{dt}X(t) = a_1(t) + a_2(t)CD4(t) - cX(t)$$

- viral load $X(t)$ and $CD4+$ counts are (periodically) observed at discrete time points
- Our data from an AIDS clinical study:
20 time points (not equally spaced) in 96 weeks.
- Observation equation

$$Y(t) = X(t) + \epsilon(t)$$

$\epsilon(t)$ is measurement error

- $Y_i = X(t_i) + \epsilon_i$

The statistical problem

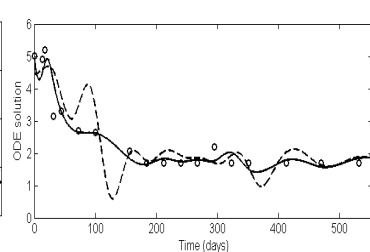
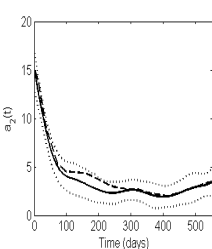
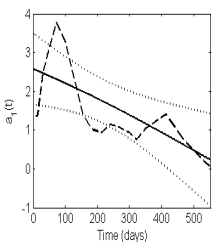
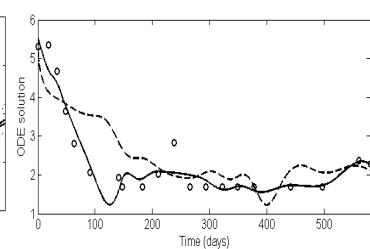
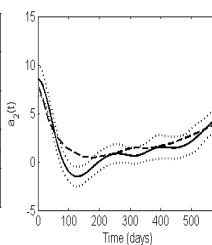
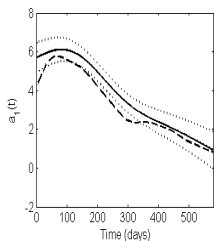
The model

$$\frac{d}{dt}X(t) = a_1(t) + a_2(t)CD4(t) - cX(t)$$
$$Y_i = X(t_i) + \epsilon_i, \quad \epsilon \sim (0, \sigma^2)$$

Objective

Estimating two time-varying parameters $a_1(t)$ and $a_2(t)$

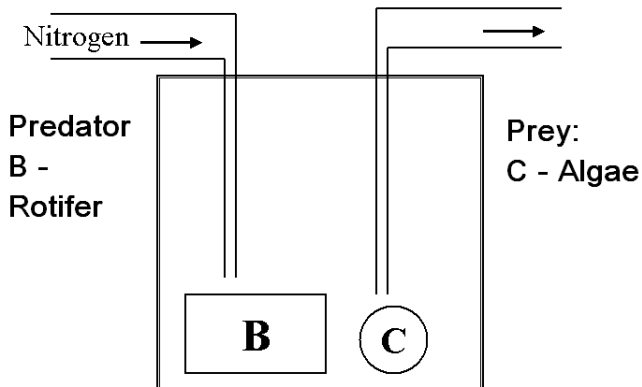
Real data results: patients 1 & 2



Outline

- 1 Introduction for dynamic Models
- 2 HIV-Dynamics
- 3 A Predator-prey system**
- 4 Gene Regulatory Network
- 5 Method

A Predator-prey system

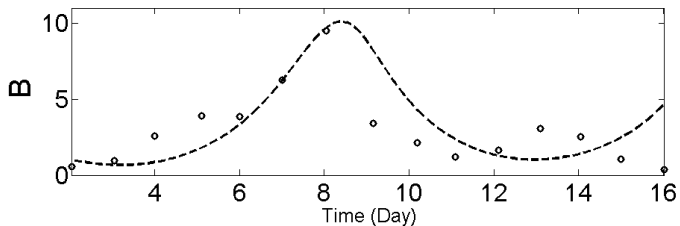
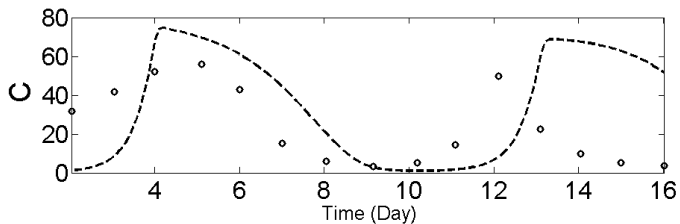


A Predator-prey dynamic Model

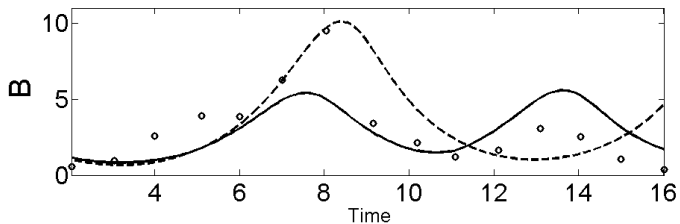
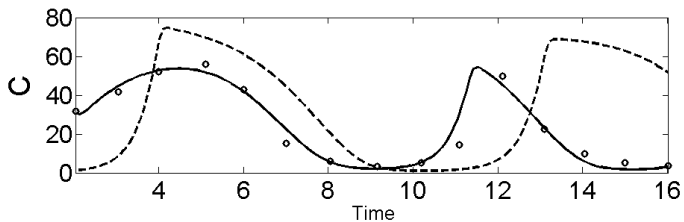
Fussmann et al. (2000). Science 290.

$$\begin{aligned} \frac{dN}{dt} &= \delta(N^* - N) - F_C(N)C \\ \frac{dC}{dt} &= F_C(N)C - F_B(C)B/\epsilon - \delta C \\ \frac{dR}{dt} &= F_B(C)R - (\delta + m + \alpha)R \\ \frac{dB}{dt} &= F_B(C)R - (\delta + m)B. \end{aligned}$$

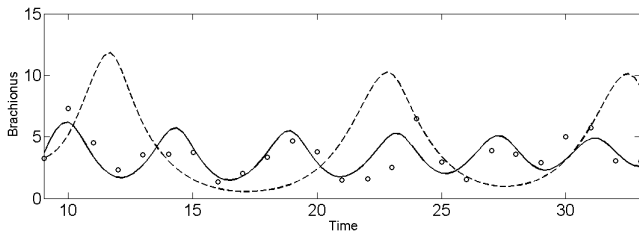
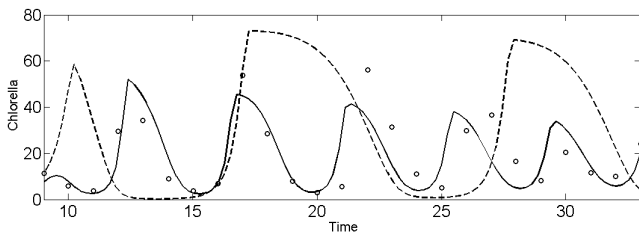
Experimental Data



Fitting ODEs to data



Fitting ODEs to data

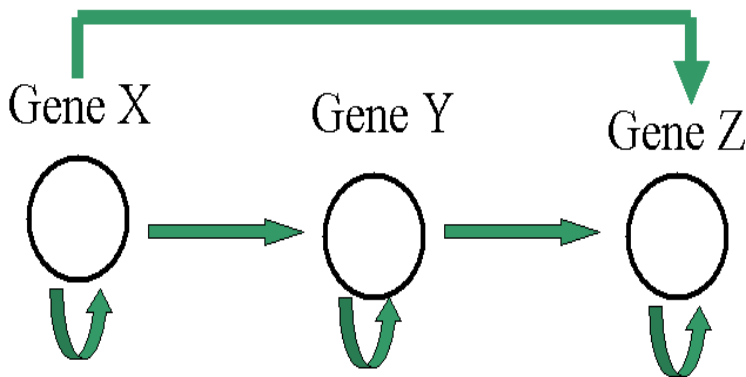


Outline

- 1 Introduction for dynamic Models
- 2 HIV-Dynamics
- 3 A Predator-prey system
- 4 Gene Regulatory Network**
- 5 Method

Feed-Forward Loop

- Network motifs: certain regulation patterns occur much more often than by chance (Alon 2007)
- **Feed-Forward Loop (FFL):** gene regulation networks with three genes.



Dynamic model for feed-forward loop

$$\frac{d}{dt}y(t) = -\alpha_y * y(t) + \beta_y * f(x)$$

$$\frac{d}{dt}z(t) = -\alpha_z * z(t) + \beta_z * f(x) * f(y)$$

- **$f(x)$ is the regulation function**

- **Activator:** $f(x, K) = \frac{(x/K)^H}{1+(x/K)^H}$

- **Repressor:** $f(x, K) = \frac{1}{1+(x/K)^H}$

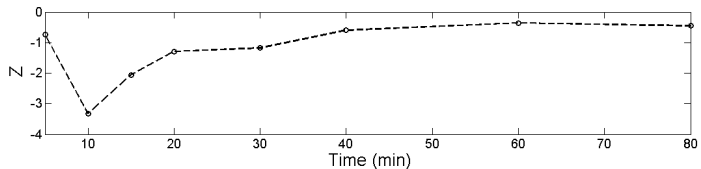
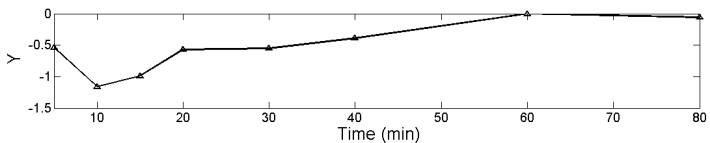
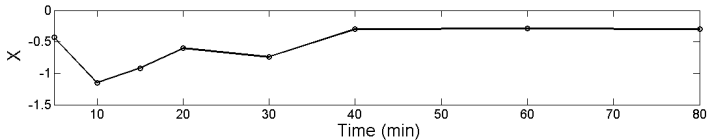
- **7 parameters to estimate:**

$$\beta = (\alpha_y, \beta_y, \alpha_z, \beta_z, K_{xy}, K_{xz}, K_{yz})$$

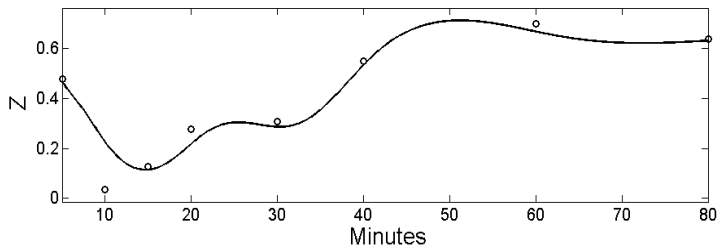
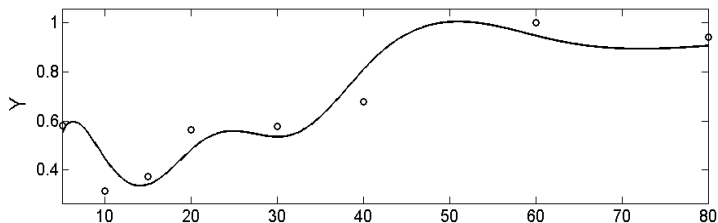
Objective

- Question 1: Can we **estimate ODE parameters** from gene expression data?
- Question 2: Can we **infer the type of regulation** between three genes?

Gene expression data



ODE solutions with estimated parameters values



TEST the goodness of fit of dynamic models

- Question 2: Infer the type of regulations between genes
- Parametric Bootstrap

| Gene X | Gene Y | Gene Z | SSE | p-values |
|---------|---------|---------|-------|----------|
| GCN4 | LEU3 | ILV5 | 0.090 | 0.25 |
| PDR1 | PDR3 | PDR5 | 1.17 | 0.33 |
| GCN4 | LEU3 | ILV1 | 0.092 | 0.34 |
| YLL044W | YER096W | YDR279W | 0.84 | 0.046 |

Outline

- 1 Introduction for dynamic Models
- 2 HIV-Dynamics
- 3 A Predator-prey system
- 4 Gene Regulatory Network
- 5 Method**

Parameter Cascading Method

- **Talk to me at JSM**
- **Email: jiguo_cao@sfu.ca**
- **Tel: 778-782-7600**