Statistical Methods for Bridging Experimental Data and Dynamic Models with Biomedical Applications

Hulin Wu, Ph.D.
Dr. D.R. Seth Family Professor & Associate Chair
Department of Biostatistics, School of Public Health

Professor, School of Biomedical Informatics
University of Texas Health Science Center at Houston

Pittsburgh, March, 2017
1 Introduction

2 Statistical estimation and inference methods for dynamic ODE models
   ▶ Naive Method: LS or MLE principle
   ▶ Local solution and time-varying parameter problems
   ▶ Smoothing-based methods
   ▶ Sparse longitudinal data: mixed-effects ODE models
   ▶ Bayesian methods
   ▶ High-dimensional ODE models: ODE model selection

3 Other dynamic models

4 Ongoing and future Work

5 Conclusions
Leo Breiman (Statistical Science, 2001): Two cultures

- **Data modeling** (98% statisticians): What the data look like? e.g., regression models
- **Algorithmic modeling** (2% statisticians): No models and for prediction purpose, e.g., neural nets and decision trees

A third culture:

- **Mechanistic modeling** (<1% statisticians): Build mathematical models based on the mechanisms behind the data
- How are the data generated?
- Goal: Understand physics principles or biological mechanisms
Many engineering and biological systems can be described by dynamic models:

- **Differential equations:**
  - Ordinary differential equations (ODE)—simplest
  - Delay differential equations (DDE)
  - Hybrid differential equations (HDE)
  - Partial differential equations (PDE)
  - Stochastic differential equations (SDE)

- Difference equations and state-space models
- Stochastic processes models: branching process etc.
- Agent-based models and cellular automata
- ...
Modeling Goals

- **Forward Problems:** $\theta \mapsto P_\theta$—Easier to do
  - Predictions
  - Simulations

- **Inverse Problems:** $Y \mapsto \theta \in \Theta$—More challenging
  - Determine model structures/forms
  - Estimate unknown parameters: $\theta$
A Dynamic System: ODE Model

\[
\frac{d}{dt} X(t) = G[X(t), \theta], \quad X(0) = X_0
\]  \hspace{1cm} (1)

\[
Y(t_i) = H[X(t_i), \beta] + e(t_i),
\]

\[e(t_i) \sim (0, \sigma^2 I), \quad i = 1, \ldots, n\]  \hspace{1cm} (2)

where

- \(G(\cdot)\): linear or nonlinear functions
- \(H(\cdot)\): observation functions
- \((\theta, \beta)\): unknown parameters
- \(e(t_i)\): measurement error

The NLS method:

\[
\min_{\theta, \beta, X_0} \sum_{i=1}^{n} \{Y(t_i) - H[X(t_i, \theta), \beta]\}^T \{Y(t_i) - H[X(t_i, \theta), \beta]\},
\]

where \(X(t_i)\) evaluated numerically from Eq (1).
Naive NLS Method: Challenging Problems

1. Identifiability problem
2. Local solutions
3. Time-varying parameters
4. Need to solve the forward problem numerically and many times: Numerical error vs. measurement error
5. Slow convergence and high computational cost
6. Sparse longitudinal data problem
7. Nonlinear optimization
8. High-dimensional parameter space

Motivate new statistical methods for dynamic models
Identifiability issues

- Theoretical identifiability: Mathematical identifiability

- Practical identifiability: Statistical and numerical identifiability

- Need to be investigated before the inverse problem

- How to deal with unidentifiable models?
  - Simplify or revise the model
  - Lump some parameters together
  - Fixed some parameters
  - Bayesian approach: Use priors
Identifiability issues: References


Naive NLS Method: Local solution and numerical error problems

- Local solution problem:
  - Global optimization methods: Differential evolution algorithms and genetic algorithms (Storn et al. 1997).
  - Mixture of stochastic global optimization method and deterministic methods: scatter search method (Rodriguez-Fernandez et al. 2006)

- Numerical error problem:
  - Xue, Miao and Wu (Annals of Statistics, 2010): theoretical results on numerical error vs. measurement error
Naive NLS Method: Time-varying parameter problem


\[
dX(t) \frac{dt}{dt} = F\{t, X(t), \theta, \eta(t)\}
\]

- The spline approach can be used to approximate the time-varying parameter:

\[
\eta(t) = \pi(t)^T \alpha,
\]

where \(\pi(t) = (B_1(t), \cdots, B_N(t))^T\) is a vector of basis functions.

- The time-varying coefficient ODE model becomes an ODE model with constant parameters:

\[
\frac{dX(t)}{dt} = F\{t, X(t), \theta, \pi(t)^T \alpha\}.
\]
Smoothing-Based Approaches: ODE Computational Problem

- Earlier ideas: Hemker (1972) and Varah (1982)
- Two-stage decoupling approaches: Chen and Wu (JASA 2008, Statistica Sinica 2008) and Liang and Wu (JASA, 2008)
Chen and Wu (JASA 2008, Statistica Sinica 2008) and Liang and Wu (JASA, 2008):

\[
X'(t_i) = F[X(t_i), \theta] \tag{3}
\]
\[
Y(t_i) = X(t_i) + e_1(t_i), \quad e_1(t_i) \sim (0, \sigma^2 I), \tag{4}
\]

- **Step 1:** Use a nonparametric smoothing to estimate \(X(t)\) and \(X'(t)\) from model (4).
- **Step 2:** Substitute the estimate \(\hat{X}(t_i)\) into model (3) to obtain:

\[
\hat{X}'(t_i) = F[\hat{X}(t_i), \theta] + e_2(t_i). \tag{5}
\]

Then fit the above regression model (5) to estimate \(\theta\).

**F(\cdot):** Linear or nonlinear function
Smoothing-Based Approaches: Two-Step Methods

- Step 2 decoupled the system of ODEs: Fit the ODE one-by-one
- Convert ODE models to regression: Standard regression software tools can be used
- Avoid numerically solving the ODEs
- Computationally fast and efficient: Easy to deal with high-dimensional ODEs

Price to pay:
- The derivative estimate may not be accurate
- The decoupled system: Some information lost
- The “coupled” property: destroyed

Extension to higher-order numerical discretization-based algorithms: Wu, Xue and Kuman (Biometrics 2012)
Parameter Cascading or Profiling Method

Ramsay, Hooker, Campbell, Cao, JRSS-B, 2007

Fitting to data

- **Observations:** \( y(t_i) \)
- **Nonparametric function:** \( f(t) = \phi(t)'c \)
- **Fitting to data:** \( C_1 = \sum_{i=1}^{n} [y(t_i) - f(t_i)]^2 \)

Fidelity to DE \( x'(t) = g(x|\beta) \)

- \( f'(t) = \phi'(t)c \)
- **Difference between two sides of DE:** \( Lf(t) = f'(t) - g(f(t)|\beta) \)
- **Fidelity to DE:** \( C_2 = \int [Lf(t)]^2 dt \)

**Criterion to estimate** \( c: J(c|\beta) = C_1 + \lambda C_2 \)

**Criterion to estimate** \( \beta: H(\beta) = \sum_{i=1}^{n} [y(t_i) - \phi(t_i)'\hat{c}(\beta)]^2 \)
Numerical Comparisons: NLS, Profiling and Two-Stage Estimates

Ding and Wu, *Statistica Sinica*, 2014

- NLS: Not stable to get the global solution, computationally expensive
  - Profiling:
    - A 3-step iterative algorithm
    - More stable than NLS to get a better solution
    - Computational efficiency: similar to NLS
  - Two-Stage Method: Computationally fast, but not accurate.
Deal with sparse data: Borrow information across subjects

- **The MLE principle: Nonlinear Mixed-Effects Modeling (NLME)**
  - Treat the ODE solution as a nonlinear regression function
  - Computational challenge: Stochastic Approximation EM (SAEM)

- **Two-stage smoothing-based mixed-effects modeling approaches**
  - Linear ODE: Linear mixed-effects model (LME)
  - Nonlinear ODE: NLME

- **Bayes methods**
  - A three-stage hierarchical model: implemented by MCMC
  - Computation: expensive
Mixed-Effects ODE Model: NLME

▶ Within-subject variation:

\[
\frac{d}{dt} X(t) = G[X(t), \theta_i], \quad X(0) = X_{i0}
\]

\[
Y_i(t_i) = H_i[X_i(t_i), \theta_i] + e_i(t_i), \quad i = 1, \ldots, n
\]

- \(X_i(t_i)\): ODE solution for Subject \(i\).
- \(Y_i = (y_{i1}(t_1), \ldots, y_{im}(t_{m_i}))^T\): Data from Subject \(i\)
- \(e_i = (e_i(t_1), \ldots, e_i(t_{m_i}))^T \sim N(0, \sigma^2 I_{m_i})\): Measurement error

▶ Between-subject variation:

\[
\theta_i = \mu + b_i, \quad [b_i | \Sigma] \sim N(0, \Sigma)
\]

- \(\mu\): population parameter
- \(b_i\): random effects

▶ Estimation and inference: Stochastic Approximation EM (SAEM)

- Delyon, Lavielle and Moulines (1999), Kuhn and Lavielle (2005)
- Grenier, Louvet, Vigneaux (2014)
Smoothing-based Two-Stage Mixed-Effects Model


\[ X'(t_i) = F[X(t_i), \theta] \]  \hspace{2cm} (7)

\[ Y(t_i) = X(t_i) + e_1(t_i), \quad e_1(t_i) \sim (0, \sigma^2 I), \]  \hspace{2cm} (8)

- **Step 1**: Use a nonparametric smoothing to estimate \( X(t) \) and \( X'(t) \) from model (8).
- **Step 2**: Substitute the estimate \( \hat{X}(t_i) \) into model (7) to obtain:
  \[ \hat{X}'(t_i) = F[\hat{X}(t_i), \theta] + e_2(t_i). \]  \hspace{2cm} (9)

- Convert the model (9) into a LME or NLME if \( F(x) \) is linear or nonlinear.
- Fit the LME or NLME using a standard approach or SAEM method.
Bayesian Methods: Borrow Information to Deal with Sparse Data and Identifiability Problems


- A viral dynamic model: describe the population dynamics of HIV and its target cells in plasma

\[
\begin{align*}
\frac{dT}{dt} &= \lambda - \rho T - \left[1 - \gamma(t)\right]kTV \\
\frac{dT^*}{dt} &= \left[1 - \gamma(t)\right]kTV - \delta T^* \\
\frac{dV}{dt} &= N\delta T^* - cV
\end{align*}
\] (10)

- \(T, T^*, V\): target uninfected cells, infected cells, virus
- \(\gamma(t)\): time-varying antiviral drug efficacy
- \((\lambda, \rho, k, \delta, N, c)\): unknown parameters to be estimated
- The equations (10): no closed-form solution
A modified $E_{max}$ (M-M) model for drug efficacy:

$$\gamma(t) = \frac{C(t)A(t)}{\phi IC_{50}(t) + C(t)A(t)} = \frac{IQ(t)A(t)}{\phi + IQ(t)A(t)}, \quad 0 \leq \gamma(t) \leq 1$$

- $C(t)$: the plasma drug concentration
- $A(t)$: drug adherence measurements
- $IC_{50}$: in vitro phenotype drug resistance marker
- $\phi$: a conversion factor parameter
- $IQ = \frac{C(t)}{IC_{50}(t)}$: the Inhibitory Quotient (IQ)

- If $\gamma(t) = 1$, the drug: 100% effective
- If $\gamma(t) = 0$, the drug: no effect
Phenotype marker $IC_{50}$ is used to quantify agent-specific drug sensitivity.

The function: to describe changes overtime in $IC_{50}$

$$IC_{50}(t) = \begin{cases} I_0 + \frac{I_r - I_0}{t_r} t & \text{for } 0 < t < t_r, \\ I_r & \text{for } t \geq t_r, \end{cases}$$

$I_0$ and $I_r$: respective values of $IC_{50}(t)$ at baseline and time point $t_r$ at which drug resistant mutations appear.

If $I_r = I_0$, no resistance mutation developed during treatment.
A Challenging Problem

How to estimate the unknown parameters in the complex dynamic model?

Difficulties:
- Identifiability problem: Too many parameters, \((\phi, \lambda, \rho, k, \delta, N, C)\), some of them are not identifiable
- Data from individuals: sparse, only \(V(t)\) measured
- Nonlinear differential equations model: no closed-form solutions
Viral load data from a clinical trial

Real data up to day 112

![Graph showing viral load data over time](image-url)
Bayesian Modeling

► A three-stage Bayesian hierarchical model

► Stage 1. Within-subject variation:

\[ y_i = f_i(\theta_i) + e_i, \quad [e_i | \sigma^2, \theta_i] \sim \mathcal{N}(0, \sigma^2 I_{m_i}) \]

\[ f_i(\theta_i) = (f_{i1}(\theta_i, t_1), \ldots, f_{im_i}(\theta_i, t_{m_i}))^T : \text{ODE solutions.} \]

\[ y_i = (y_{i1}(t_1), \ldots, y_{im_i}(t_{m_i}))^T : \text{Data from Subject } i \]

\[ e_i = (e_i(t_1), \ldots, e_i(t_{m_i}))^T : \text{Measurement error} \]

► Stage 2. Between-subject variation:

\[ \theta_i = \mu + b_i, \quad [b_i | \Sigma] \sim \mathcal{N}(0, \Sigma) \]

► Stage 3. Hyperprior distributions:

\[ \sigma^{-2} \sim \text{Ga}(a, b), \quad \mu \sim \mathcal{N}(\eta, \Lambda), \quad \Sigma^{-1} \sim \text{Wi}(\Omega, \nu) \]

► Gamma (\text{Ga}), Normal (\mathcal{N}) and Wishart (\text{Wi}): independent distributions

► Hyper-parameters \( a, b, \eta, \Lambda, \Omega \) and \( \nu \): known
Bayesian Estimation: Implementation

- Choose prior distributions
  - Informative prior and non-informative prior
  - Rule of thumb: choose non-informative prior distributions for parameters of interest

- Implement MCMC algorithm
  - Gibbs sampling step: closed form of conditional distributions for $\sigma^{-2}, \mu, \Sigma^{-1}$
  - Metropolis-Hastings step: no closed form of conditional distributions for $\theta_i$

- Run a long chain: the number of iterations, initial “burn-in”, every fifth simulation samples

- Obtain posterior distributions (posterior means or credible intervals) based on the final MCMC samples
A Clinical Study: A5055

- A study of HIV-1 infected patients failing PI-containing therapies.
- Two salvage regimens: 44 patients
  - Arm A: IDV 800 mg q12h+RTV 200mg q12h+two NRTIs
  - Arm B: IDV 400 mg q12h+RTV 400mg q12h+two NRTIs
- Plasma HIV-1 RNA (viral load) measured at days 0, 7, 14, 28, 56, 84, 112, 140 and 168 of follow-up
Posterior mean for the population parameter $\phi$ is 2.1091 with a SD of 0.6354 and the 95% CI of (1.2143, 3.6392)

As $\phi$ plays a role of transforming the \textit{in vitro} $IC_{50}$ into \textit{in vivo} $IC_{50}$, our estimate shows that there is about 2-fold difference between \textit{in vitro} $IC_{50}$ and \textit{in vivo} $IC_{50}$
Clinical Data–Results of Individual Parameters

<table>
<thead>
<tr>
<th>Patient</th>
<th>$\phi_i$</th>
<th>$c_i$</th>
<th>$\delta_i$</th>
<th>$\lambda_i$</th>
<th>$\rho_i$</th>
<th>$N_i$</th>
<th>$k_i$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.447</td>
<td>2.254</td>
<td>0.270</td>
<td>410.462</td>
<td>0.024</td>
<td>456.757</td>
<td>$8.33 \times 10^{-6}$</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>5.371</td>
<td>2.969</td>
<td>1.183</td>
<td>29.619</td>
<td>0.426</td>
<td>4795.813</td>
<td>$10.84 \times 10^{-6}$</td>
<td>0.17</td>
</tr>
<tr>
<td>3</td>
<td>3.723</td>
<td>2.283</td>
<td>0.456</td>
<td>36.877</td>
<td>0.289</td>
<td>3258.347</td>
<td>$8.66 \times 10^{-6}$</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>4.960</td>
<td>2.761</td>
<td>0.798</td>
<td>44.956</td>
<td>0.313</td>
<td>3051.988</td>
<td>$9.09 \times 10^{-6}$</td>
<td>0.34</td>
</tr>
<tr>
<td>5</td>
<td>7.066</td>
<td>2.306</td>
<td>0.663</td>
<td>71.295</td>
<td>0.201</td>
<td>2735.239</td>
<td>$6.54 \times 10^{-6}$</td>
<td>0.64</td>
</tr>
<tr>
<td>6</td>
<td>0.786</td>
<td>4.633</td>
<td>0.183</td>
<td>375.882</td>
<td>0.025</td>
<td>247.416</td>
<td>$11.18 \times 10^{-6}$</td>
<td>0.89</td>
</tr>
<tr>
<td>7</td>
<td>0.091</td>
<td>7.008</td>
<td>0.299</td>
<td>4015.398</td>
<td>0.003</td>
<td>30.559</td>
<td>$18.54 \times 10^{-6}$</td>
<td>0.98</td>
</tr>
<tr>
<td>8</td>
<td>8.484</td>
<td>2.280</td>
<td>0.663</td>
<td>32.722</td>
<td>0.416</td>
<td>4530.531</td>
<td>$8.37 \times 10^{-6}$</td>
<td>0.24</td>
</tr>
</tbody>
</table>

- The individual-specific parameter estimates suggest a large inter-subject variation
- The model provides a good fit to the clinical data
Fitted individual curves, drug efficacy, IC50 and adherence with IQ=c12h/IC50

- **IC50**
  - Patient 1
  - Time (day): 0 50 100 150 200
  - IC50 levels: 4, 8, 12, 16
  - Drugs: IDV, RTV

- **Adherence**
  - Time (day): 0 50 100 150 200
  - Adherence levels: 0.80, 0.90, 1.00
  - Drugs: IDV, RTV

- **Drug Efficacy**
  - Time (day): 0 50 100 150 200
  - Efficacy levels: 0.6, 0.8, 1.0

- **log10(RNA)**
  - Time (day): 0 50 100 150 200
  - RNA levels: 1.5, 3.0, 4.5
Patient 3

IC50

Adherence

Drug efficacy

log10(RNA)
Bayesian Methods: Pros & Cons

- **Pros**
  - Use prior to solve the identifiability problem
  - Deal with extremely complicated models such as nonlinear differential equation models
  - Borrow information across subjects:
    - Deal with sparse longitudinal data
    - Estimate parameters for both population and individuals
  - Always get reasonable estimates
  - Use posterior distributions: Easy to quantify “uncertainty" for inference

- **Cons**
  - Computation: complex and expensive
  - Prior: dominate the results
High-Dimensional ODEs

- Require computationally fast and efficient methods
- Need to incorporate variable selection approaches: LASSO, SCAD etc.
- Easy to deal with longitudinal data: Mixed-effects models
- Two-stage smoothing-based method: good for this purpose
Linear ODEs

Time course gene expression data: Dynamic gene regulatory network (GRN) reconstruction (Lu, Liang, Li and Wu, JASA 2011)

\[
\frac{dx_i}{dt} = \sum_{j=1}^{n} \theta_{ij} x_j, \quad i = 1, \cdots, n, \tag{13}
\]

- When \( n \) is small, standard statistical inference and variable selection methods can be used
- When \( n \) is large, curse-of-dimensionality
High-Dimensional Linear ODE: Identifying Significant Regulations

**Two-Stage Method** (Chen and Wu 2008a, 2008b; Liang and Wu 2008):

- Obtain mean expression curves and their derivatives $\hat{M}_k(t)$ and $\hat{M}'_k(t)$ from Step II.
- Substitute $\hat{M}_k(t)$ and $\hat{M}'_k(t)$ into the ODE model to form a regression model

**High Dimensional Linear Regression Model**

\[
y_k(t) = \sum_{j=1}^{p} \beta_{kj} x_j(t) + \varepsilon_k(t),
\]

\[
k = 1, \ldots, p; \quad t = t_1, t_2, \ldots, t_N
\]

\[
y_k(t) = \hat{M}'_k(t) \text{ and } x_j(t) = \hat{M}_j(t)
\]
High Dimensional Model Selection

- Two-stage method
  - Decouple the high-dimensional ODEs
  - Convert the ODE model into a simple linear model
  - Computationally fast
- Stepwise selection and subset selection
- Bridge selection (Frank and Friedman 1993)
- Least absolute shrinkage and selection operator (LASSO) (Tibshirani 1996)
- Smoothly Clipped Absolute Deviation (SCAD) (Fan and Li 2001; Kim, Choi and Oh 2008)
Estimation Refinement: Stochastic Approximation EM (SAEM) Algorithm

Mixed-Effects ODE Model for Module $k$

$$\frac{dx_{ki}}{dt} = \sum_{j=1}^{m_k} \beta_{kij} M_{[k,j]}(t), \quad i = 1, \ldots, n_k; \quad k = 1, \ldots, p, \quad (14)$$

Longitudinal Measurement Model

$$y_{ki}(t) = x_{ki}(t) + \varepsilon_{ki}(t) \quad (15)$$

Random Effects Model

$$\beta_{ki} = \beta_k + b_{ki} \quad (16)$$

$$b_{ki} \sim \mathcal{N}(0, D_k)$$
DNA microarrays experiment: 18 equally spaced time points during two cell cycles (Spellman 1998)

- **Step I:** 800 significant genes identified
- **Step II:** Cluster 800 genes into 41 functional modules
- **Step III:** Smoothing
- **Step IV:** Linear ODE model identification: SCAD variable selection
- **Step V:** Estimation Refinement
- **Step VI:** Function Enrichment Analysis
Yeast Cell Cycle Gene Expression Profile
Yeast Cell Cycle Gene Expression Profile
Yeast Cell Cycle Gene Expression Profile

Module 33

Module 34

Module 35

Module 36

Module 37

Module 38*

Module 39

Module 40

Module 41*

Hulin Wu

UTSPH

March 2017
Graph of Yeast Cell Cycle GRN
High-Dimensional Nonlinear/Nonparametric ODEs


- Additive nonlinear ODEs: Xue, Wu, Wu and Wu, a manuscript (2017)
Other Dynamic Models: State-Space Models (SSM)

Linear SSM:

\[
X_{t+1} = F_t X_t + V_t, \quad V_t \sim (0, Q_t) \tag{17}
\]
\[
Y_t = G_t X_t + W_t, \quad W_t \sim (0, R_t) \tag{18}
\]

where

- \( V_t \) and \( W_t \): independent model noise and measurement noise
- Standard Kalman filter (Kalman, 1960): the core algorithm for prediction and smoothing of state state vectors
Chen et al. *PlusOne* (2017), submitted
Extension to SDE and PDE: Possible but Challenging

- Theoretically difficult
- Computationally challenging
- Applications: Not common
Ongoing and Future Research

- High-dimensional ODEs: How to improve accuracy without sacrificing too much on computing?
  - Extra-high dimensional ODE: 1000 ODEs with 1 million parameters (Wu, Qiu, Yuan and Wu, 2017, submitted).

- Characteristic analyses of large ODE systems: Controllability and stability analysis with uncertainty in parameter estimation

- AI-driven ODE Model Builder
Conclusions

Dynamic Models:
- Practically useful for both understanding associations and predictions
- Both theoretically and computationally challenging
- Statistical methods for dynamic models: More work needed
Dr. Hulin Wu’s Publications on ODE Models by Topics

ODE identifiability

NLS Estimation of ODE parameters

Two-stage methods for ODE models
Time-varying parameter estimation in ODE Models

Bayesian and mixed-effects ODE modeling approaches for longitudinal data

High-dimensional ODE models and model selections

Nonlinear/nonparametric ODE models

Statistical methods for state-space models
ODE experimental design

Dynamic model property analysis with uncertainty
Our recent work in nonlinear/high-dimensional ODE models

More than 30 postdocs, students and collaborators
Thank You!