
Hulin Wu, Ph.D., Professor

Chief, Division of Biomedical Modeling and Informatics
Director, Center for Biodefense Immune Modeling

Department of Biostatistics and Computational Biology
University of Rochester School of Medicine and Dentistry

hwu@bst.rochester.edu
Outline

• Introduction

• Inverse Problems and Methods for ODE Models

• Application Examples
  – Example 1: A Complex Influenza Model
  – Example 2: A Simplified Influenza Model
  – Example 3: CD8+ T Cell Trafficking Model
  – Example 4: Dynamic Gene Regulatory Network
  – Example 5: Ongoing Multi-Level Models for influenza vaccination

• Software Tools

• Challenges and Discussions
Introduction: Three types of biomathematical models

- Theoretical models without any data justification: Build the models purely based on biological mechanisms and theories

- Models with partial data validation: Build the models based on biological mechanisms and theories with some data for partial validation

- Models with rigorous data justification: Build the models based on both biological mechanisms/theories and experimental data using rigorous statistical methods in an iterative procedure
Introduction: Model building principles

- All models are wrong, but some are useful
- All models are wrong, but how wrong is it to make a model not useful?
- How complex or how much detail should a model be built to describe a biological process or system?
Introduction: Model building principles

- Build a model at a complexity and detailed level which is useful just for your purpose: a purpose-driven model

- A more useful model: a model can be identified and validated by experimental data
  - **Data-Driven Empirical Models**: statistical models based on what the data look like. No biological mechanism assumptions required
  - **Data-Driven Mechanistic Models**: mathematical models based on both data and theories. Need to understand how the data are generated, and mechanistic theories and assumptions required
Introduction: An iterative model building procedure

1. Construct an initial model based on biological mechanisms/theories

2. Design experiments to collect data to estimate model parameters and validate the model structures

3. Update the model based on the data and perform model diagnostics

4. Apply the established model for simulations and predictions

5. Check and validate the predictions using new data

6. Feedback to revise the model if there is any prediction discrepancies
Mathematical Models for Biomedical Processes

- Ordinary differential equations (ODEs)
- Delay differential equations (DDE)
- Partial differential equations (PDEs)
- Stochastic differential equations (SDE)
- State-space models
- Stochastic processes models: branching process etc.
- Agent-based models
- Network models
- ......
Why Modeling in Biomedical Research?

- Describe a biological process or mechanism in a quantitative way
- Represent a complex and nonlinear interaction network in a systematic and dynamic way—systems biology
- Interpret the experimental data more appropriately
- Extract more information from a vast amount of complex data
- Simulate ‘what if’ scenarios: identifying drug or vaccine targets
- Better predictions
Two Modeling Problems

- **Forward Problem:** $\theta \mapsto P_\theta$
  - Predictions
  - Simulations
  - Need: Mathematics and computing

- **Inverse Problem:** $Y \mapsto \theta \in \Theta$
  - Parameter estimation
  - Need: Statistics
Gap between Mathematical Modeling and Statistics

- **Mathematical modelers**
  - Use simple statistical methods
  - Use ad hoc methods to link models with experimental data

- **Statisticians**
  - Do not like models that require strong mechanism assumptions
  - Want data to speak only
  - Few statisticians and statistical methods for mechanism-based models

- **Bridge the gap**
  - Collaborations and compromise
  - Learn from each other
Why Inverse Problems Are More Challenging?

- Identifiability problems for complex mathematical models
- High-dimensional parameter space
- High computational cost
- Model structure uncertainty and identification
- Some parameters may not be constant
- Sparse data
- Model evaluation and diagnostics
Example: ODE Model

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

\[
Y(t_i) = H[X(t_i), \beta] + e(t_i), \quad (2)
\]

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

where

- (1)–state equation and (2)–observation equation
- \( F(\cdot) \): nonlinear dynamic function
- \( H(\cdot) \): observation functions
- Differential equation model without error, but measurements with error
- No closed-form solution
ODE Model Identifiability


- **Mathematical (structural) identifiability problem**: all parameters theoretical identifiable?
  - Power series expansion
  - Similarity transformation
  - Implicit function theorem
  - Differential algebra methods

- **Statistical (practical) identifiability problem**: all parameters practically identifiable with presence of measurement errors
  - Statistical estimation methods
  - Monte Carlo simulations
ODE Model: Parameter Estimation Methods

- The nonlinear least squares (NLS) principle
  - numerically solve the ODE
  - computationally expensive

- Two-step smoothing approach
  - avoid numerically solving the ODE
  - easy to implement
  - theoretical properties need to be established

- Penalized smoothing and profiling method by Ramsay et al. (2007)
  - avoid numerically solving the ODE
  - 3-stage nested iterations

- Bayesian method
Meet the Challenges

- Identifiability problem: develop new identifiability analysis techniques
- How to deal with local minima and non-convergence of computational algorithms? Global optimization and more efficient algorithms
- How to deal with time-varying parameters? nonparametric spline approximation
- How to deal with high computational costs? more computationally efficient methods
- ODE model structure identification? model selection approaches such as AIC, regularization-based methods
- User-friendly software? development of new tools
Application Example 1: A Complex Model for Influenza Infection


\[
\begin{align*}
\frac{d}{dt} E_p &= \delta_E (E_0 - E_p) - \beta_E E_p V, \\
\frac{d}{dt} E_p^* &= \beta_E E_p V - k_E E_p^* T_E(t) - \delta_{E^*} E_p^*, \\
\frac{d}{dt} V &= \pi_V E_p^* - c_V V - k_V V A(t), \\
\frac{d}{dt} D &= \delta_D (D_0 - D) - \beta_D D V, \\
\frac{d}{dt} D^* &= \beta_D D V - \delta_{D^*} D^* - \gamma_{D^*} D^*
\end{align*}
\]

\(E_p\): uninfected epithelial cells  
\(E_p^*\): infected epithelial cells  
\(V\): free influenza virus  
\(D\): immature dendritic cells  
\(D^*\): virus-loaded dendritic cells
Influenza Virus Infection

Spleen/Lymph Node Compartment

\[
\frac{d}{dt} D_M = k_D D^*(t - \tau_D) - \delta_{D_M} D_M, \\
\frac{d}{dt} H_N = \delta_{H_N} (H_{N0} - H_N) - \pi_H(D_M)H_N, \\
\frac{d}{dt} H_E = \pi_H(D_M)H_N + \rho_{H_E}(D_M)H_E - \delta_{H_E}(D_M)H_E, \\
\frac{d}{dt} T_N = \delta_{T_N} (T_{N0} - T_N) - \pi_T(D_M)T_N, \\
\frac{d}{dt} T_E = \pi_T(D_M)T_N + \rho_{T_E}(D_M)T_E - \delta_{T_E}(D_M)T_E, \\
\frac{d}{dt} B = \delta_B(B_0 - B) - \pi_B(D_M)B, \\
\frac{d}{dt} B_A = \pi_B(D_M)B + \rho_{B_A}(D_M + hH_E)B_A - \delta_{B_A}B_A - \pi_sB_A - \pi_L H_E B_A, \\
\frac{d}{dt} P_s = \pi_sB_A - \delta_s P_s, \\
\frac{d}{dt} P_L = \pi_L H_E B_A - \delta_L P_L, \\
\frac{d}{dt} A = \pi_A S P_s + \pi_A L P_L - \delta_A A
\]
Two Compartment Flu Model

Airway/Lung

- Infected Epithelial cell (E_p^*)
- Uninfected Epithelial cell (E_p)
- Immature dendritic cell (D)
- Virus+ dendritic cell (D^*)
- Naïve CD8 T cell (T_N)
- Effector CD8 T cell (T_E)
- Naïve CD4 T cell (H_N)
- Effector CD4 T cell (H_E)
- Activated B cell (B_A)
- Short-lived Plasma cell (P_S)
- Long-lived Plasma cell (P_L)
- Antiviral antibody (A)
- Mature Virus+ dendritic cell (D_M)
- Infuenza virus (V)
- Spleen/Lymph Node
- Long-lived Plasma cell (P_L)
- Naïve CD4 T cell (H_N)

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12) (13) (14) (15)
Application Example 2: A Simplified Model for Influenza Infection

- Use biological knowledge to de-couple or simplify the complex models
- Design experiments to collect enough data
- Fit the data to the models
- Application of established models: biological interpretation, simulations, predictions and designing future experiments
Model Simplification

Lung Compartment Model: Model Simplification

Days 5-14 with adaptive immune response: decoupled the model

\[
\begin{align*}
\frac{d}{dt} E_p &= \rho_E E_p - \beta_a E_p V, \\
\frac{d}{dt} E^*_p &= \beta_a E_p V - \delta E^*_p - k_E E^*_p T_E(t), \\
\frac{d}{dt} V &= N_a E^*_p - c_V V - k_V V A(t),
\end{align*}
\]

- \( V(t), T_E(t) \) and \( A(t) \): measured
- \(((E_p(0), V(0), \rho_E, \beta_a, \delta E^*, c_V, k_E, k_V)\): to be estimated
- \( \eta(t) = c_V + k_V A(t) \): a time-varying parameter
- Identifiability analysis: Need to fix \( N_a \)
Figure 1. Viral titer data (2007).

Figure 2. Smoothed any positive CD8 data (2007).
Figure 3. Smoothed IgG and IgM data (2007).
(a). $\pi_\alpha = 1 \text{ EID}_{50} \cdot \text{ml}^{-1} \cdot \text{day}^{-1} \cdot \text{cell}^{-1}$;

(b). $\pi_\alpha = 100 \text{ EID}_{50} \cdot \text{ml}^{-1} \cdot \text{day}^{-1} \cdot \text{cell}^{-1}$;

(c). $\pi_\alpha = 1000 \text{ EID}_{50} \cdot \text{ml}^{-1} \cdot \text{day}^{-1} \cdot \text{cell}^{-1}$;

Figure 1. Model fitting results for different virus production rate
Table 1. Estimation results with 95% bootstrap confidence intervals

<table>
<thead>
<tr>
<th>$\pi_\alpha$</th>
<th>EID$_{50}$, ml$^{-1}$·day$^{-1}$·cell$^{-1}$</th>
<th>$E_p(0)$, cells per lung</th>
<th>$\rho_E$, ml$^{-1}$·day$^{-1}$</th>
<th>$\beta_\alpha$, EID$_{50}$·cell$^{-1}$·day$^{-1}$</th>
<th>$k_E$, day$^{-1}$</th>
<th>$\delta_E^*$, day$^{-1}$</th>
<th>$c_v$, day$^{-1}$</th>
<th>$k_{VG}$, ml/(pg·day)</th>
<th>$k_{VM}$, ml/(pg·day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.3E+07, 9.9E-04</td>
<td>5.1E-06</td>
<td>1.4E-05</td>
<td>1.2E+00</td>
<td>3.1E-05</td>
<td>1.7E-08</td>
<td>7.8E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.1E+06, 4.5E-06</td>
<td>8.5E-07</td>
<td>3.7E10</td>
<td>7.5E-01</td>
<td>1.6E-06</td>
<td>4.4E-09</td>
<td>5.2E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0E+09</td>
<td>1.0</td>
<td>1.8E-05</td>
<td>4.9E-05</td>
<td>1.6E+00</td>
<td>5.2E-01</td>
<td>7.0E-06</td>
<td>8.1E+00</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2.4E+05, 8.6E-04</td>
<td>5.1E-06</td>
<td>1.4E-05</td>
<td>1.2E+00</td>
<td>5.5E-06</td>
<td>4.5E-08</td>
<td>7.8E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.4E+04, 4.8E-06</td>
<td>9.1E-07</td>
<td>1.1E-10</td>
<td>7.4E-01</td>
<td>2.2E-06</td>
<td>4.7E-09</td>
<td>5.1E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10E+06</td>
<td>1.00</td>
<td>1.7E-05</td>
<td>5.0E-05</td>
<td>1.6E+00</td>
<td>5.3E-01</td>
<td>8.1E-06</td>
<td>8.1E+00</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>2.3E+04, 9.2E-04</td>
<td>5.1E-06</td>
<td>1.4E-05</td>
<td>1.2E+00</td>
<td>5.2E-06</td>
<td>2.6E-08</td>
<td>7.8E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.1E+03, 4.0E-06</td>
<td>8.9E-07</td>
<td>2.4E-10</td>
<td>7.5E-01</td>
<td>1.7E-06</td>
<td>6.2E-09</td>
<td>5.1E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0E+06</td>
<td>1.0E+00</td>
<td>1.8E-05</td>
<td>4.9E-05</td>
<td>1.6E+00</td>
<td>5.3E-01</td>
<td>8.4E-06</td>
<td>8.2E+00</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Model selection results using AICc

<table>
<thead>
<tr>
<th>Model</th>
<th>RSS</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model</td>
<td>49.4</td>
<td>-33.2</td>
</tr>
<tr>
<td>$\rho_E=0$</td>
<td>49.4</td>
<td>-34.2</td>
</tr>
<tr>
<td>$k_E=0, \delta_{\varepsilon}=0$</td>
<td>360</td>
<td>137</td>
</tr>
<tr>
<td>$c_v=0, k_{vG}=0$</td>
<td>49.4</td>
<td>-36.2</td>
</tr>
<tr>
<td>$c_v=0, k_{vG}=0, k_{vM}=0$</td>
<td>484</td>
<td>159</td>
</tr>
<tr>
<td>$\rho_E=0, c_v=0, k_{vG}=0$</td>
<td>49.4</td>
<td>-39.2</td>
</tr>
</tbody>
</table>
Figure 2. Comparison of the patterns of the estimated time-varying parameter $\eta(t)$, $k_{VM} A_M(t)$ and $k_{VG} A_G(t)$
Figure 3. The effects of different parameters on the peak viral load. 
(a) The effect of $\beta_\alpha$; (b) The effect of $k_E$; (c) The effect of $\delta_E$; (d) The effect of $k_{VM}$; (e) The effect of $\pi_\alpha$; (f) The effect of $E_p(0)$
Model Fitting Conclusions

- We can reliably estimate the immune response kinetic parameters during the adaptive immune response period (Days 5-14):
  - The net growth rate of uninfected cells = the proliferation rate – the death rate: \( \rho = 0.339/\text{day} \), larger compared to that during the first 5 days
  - The infection of epithelial cells: no targets
  - CTL effect: shorten the half-life of infected cells from 1.16 days to 0.59 days in average
  - IgM antibody effect: shorten the half-life of virus from 4 hours to 1.7 minutes in average.
Example 3: CD8+ T cell trafficking in mice with influenza infection

Model structure identification

\[
\begin{align*}
\frac{d}{dt} T_m &= \left[ \rho_m D_m(t - \tau) - \delta_m \right] T_m - (\gamma_{ms} + \gamma_{ml}) T_m, \\
\frac{d}{dt} T_s &= \left[ \rho_s D_s(t - \tau) - \delta_s \right] T_s - \gamma_{sl} T_s + \gamma_{ms} T_m, \\
\frac{d}{dt} T_l &= \gamma_{ml} T_m + \gamma_{sl} T_s - \delta_l T_l,
\end{align*}
\]  

- \( T_m(t) \): CD8+ T cells in MLN
- \( T_s(t) \): CD8+ T cells in Spleen
- \( T_l(t) \): CD8+ T cells in Lung
- \( D_m(t - \tau) \): DC cells in MLN with a time delay
- \( D_s(t - \tau) \): DC cells in Spleen with a time delay
Example 3: CD8+ T cell trafficking in mice with influenza infection

Hypothesis: Is the loss rate of CD8+ T cells in the lung $\delta_l =$ constant?

\[
\frac{d}{dt} T_m = [\rho_m D_m(t - \tau) - \delta_m] T_m - (\gamma_{ms} + \gamma_{ml}) T_m; \\
\frac{d}{dt} T_s = [\rho_s D_s(t - \tau) - \delta_s] T_s - \gamma_{sl} T_s + \gamma_{ms} T_m; \\
\frac{d}{dt} T_l = \gamma_{ml} T_m + \gamma_{sl} T_s - \delta_l(t) T_l,
\]

(4)
Example 3: CD8+ T cell trafficking in mice with influenza infection

Hypothesis: Is there influx of CD8+ T cells from other tissues to spleen?

\[
\begin{align*}
\frac{d}{dt} T_m &= \left[ \rho_m D_m(t - \tau) - \delta_m \right] T_m - (\gamma_{ms} + \gamma_{ml}) T_m, \\
\frac{d}{dt} T_s &= \left[ \rho_s D_s(t - \tau) - \delta_s \right] T_s - \gamma_{sl} T_s + \gamma_{ms} T_m + \eta_s(t), \\
\frac{d}{dt} T_l &= \gamma_{ml} T_m + \gamma_{sl} T_s - \delta_l(t) T_l,
\end{align*}
\]

(5)
Example 4: Dynamic Gene Regulatory Network

High-dimensional ODE models for dynamic gene regulatory network (GRN):

\[
\frac{dx_i}{dt} = \sum_{j=1}^{p} \theta_{ij} x_j, \quad i = 1, \ldots, p, \quad (6)
\]

where parameters \( \Theta = \{\theta_{ij}\}_{i,j=1,\ldots,p} \): quantify the interactions/regulations among the genes in the network.

- Model selection: determine significant gene-gene interactions/regulations
- ODE parameter estimation: quantify the strength of interactions/regulations
- Deal with high-dimensional and high computational cost
Dynamic GRN for Yeast Cell Cycle

- Clustering 800 genes into 41 functional modules
- High dimensional ODEs coupled with mixed-effects modeling techniques
- A two-step smoothing approach coupled with the SCAD technique
Example 5: Multi-Level Experiments for Influenza Infection

- Mice with TIV or LAIV vaccination: blood, lung, spleen, LN and bone marrow samples
- Humans with TIV or LAIV vaccination: blood samples, nasal washes (LAIV)
- Frequent time course data: 13 time points
- Multi-level data
  - Cellular level: flow cytometry, Elispot etc.
  - Protein/molecular level: Luminex
  - Genetic level: microarray
Multi-Scale and Multi-Type Models

- Multi-scale and multi-level models
  - Genetics level: time course microarray data
  - Protein level: cytokines and chemokines
  - Cellular level: flow cytometry data
  - Multi-level integration

- Multi-type models: ODE, SDE, state-space models, stochastic process models, agent-based models, network models

- Multi-scale and multi-type model integration
Summary: Multi-Level, Multi-Scale and Multi-Compartment Models

Modeling approach
- Spatial Models
- PDEs
- Network Models
- Bayesian Models
- SDEs

Scales
- $10^{-1}$ m
- $10^{-6}$ m
- $10^{-10}$ m

Biological levels
- Circulation (PMC)
- Tissue/Organ
- Cell population
- Protein network
- Gene/Transcription network

Experimential approach
- Clinical chemistry Imaging
- Flow Cytometry FACS
- Luminescence Elispot ELISA
- Microarray RNA-Seq RT-PCR

Total ~ 30,000 genes
Software Tools

• Develop more efficient computational algorithms for model simulations and parameter estimation

• Develop user-friendly software tools for modeling, simulation and estimation

• Develop web-based data management tools for high-throughput biomedical data
DEDiscover is a software tool for developing, exploring, and applying differential equation models. It provides simulation, prediction, and parameter estimation (data fitting) facilities for systems of ordinary and delayed differential equations, and includes a set of example models for infectious diseases.

Free Download: http://cbim.urmc.rochester.edu/software/dediscover
A Comprehensive Data Management System for Immunological Research

An immunological research always involves different types of data and information, like biological samples of participating subjects, experiment raw data from immunological laboratories, processed data by computation tools, and various documents for studies. It also needs an efficient way to handle tremendous phenotype data generated by immunological assays which include flow cytometry, enzyme-linked immunosorbent assay (ELISA), and enzyme-linked immunospot (ELISPOT). In addition, researchers need a standardized tool to record all data and documents based on research workflows, an easy way to save and query data/documents from distributed locations, and an efficient platform for data sharing to the broad research community. To achieve those goals we present a comprehensive Web-based system, DataTrans, for managing data and information in the studies of immunological research.

Partial funding support from

National Institute of Allergy and Infectious Diseases
National Institutes of Health

Contract Number N01-AI-50020

UNIVERSITY OF ROCHESTER
Center for Biodefense Immune Modeling
SCHOOL OF MEDICINE AND DENTISTRY
Summary: Biomedical Research System

Biomedical Research

Experiments

Quantitative Sciences

Data Management

Statistical Analysis

Math Modeling

Software Development

Computing

Computing Hardware & Admin
Challenges and Discussions

- Integrating quantitative sciences into a new interdisciplinary field
  - Biomathematics
  - Bioengineering and biophysics
  - Bioinformatics and Biocomputing
  - Biostatistics

- Quantitative Sciences: important and a strong impact to
  - Life sciences and biomedical research
  - Clinical practice and public health
  - Experimental systems biology approach

- Interdisciplinary Quantitative Sciences: Challenges
  - Lack of strong interdisciplinary leaders
  - Long-term visions
  - Lack of collaboration, communication and coordination for the investigators from different disciplines
Acknowledgement

Immunologists: Zand, Welle Topham, Mosmann, Ward, Jin

Experiments
(Lab Postdocs & Technicians)

Data Management:
Holden-Wiltse
Zhang, Yang, Massaro

Statistical Analysis:
Liang
Xue, Kumar, Lu, Wu, Yang

Math Modeling:
Miao
Perelson, Mugwagwa

Computing:
Warnes, Miao

Computing Hardware & Admin:
IT Staff

Software Development:
Stover, Warnes, Ma, LeBlanc,
Crowley, Wu, William,

Quantitative Sciences
Hulin Wu
Acknowledgments

- **NIAID/NIH grant R01 AI 055290**: AIDS Clinical Trial Modeling and Simulations

- **NIAID/NIH grant N01 AI50020**: Center for Biodefense Immune Modeling

- **NIAID/NIH grant P30 AI078498**: Developmental Center for AIDS Research

- **NIAID/NIH grant R21 AI078842**: Analysis of Differential Resistance Emergence Risk for Differential Treatment Applications

- **NIAID/NIH grant RO1 AI087135**: Estimation Methods for Nonlinear ODE Models in AIDS Research