Mixed-Effects Biological Models: Estimation and Inference

Hulin Wu, Ph.D.
Dr. D.R. Seth Family Professor & Associate Chair

Department of Biostatistics
School of Public Health
University of Texas Health Science Center at Houston

July, 2016

Outline

Statistical vs. Mechanism-Based Mathematical Models for Biological Systems

- Empirical Statistical Models: Linear, nonlinear, nonparametric or semiparametric models
  - What do the data look like?
  - Developed after data collection
  - No biological mechanism assumptions required
- Mechanisms-Based Math Biological Models:
  - How are the data generated?
  - Can be developed before data collection
  - Biological mechanisms required
- Hybrid of Math and Statistical Models
  - Partial data-driven and partial mechanism-driven models

Mathematical Models for Biological Systems

- Differential equations:
  - Ordinary differential equations (ODE)
  - 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
- Network models: Boolean, Bayesian, Petri nets
- . . .

A Dynamic System: ODE Model

$$\frac{d}{dt}\boldsymbol{X}(t) = G[\boldsymbol{X}(t), \boldsymbol{\theta}], \quad \boldsymbol{X}(0) = \boldsymbol{X}_0 \tag{1}$$

$$\begin{aligned} \boldsymbol{Y}(t_i) &= H[\boldsymbol{X}(t_i), \boldsymbol{\theta}] + \boldsymbol{e}(t_i), \\ & \boldsymbol{e}(t_i) \sim (0, \sigma^2 \boldsymbol{I}), \quad i = 1, \ldots, n \end{aligned} \tag{2}$$

where

- (3)–state equation
- (4)–observation equation
- $G(\cdot)$: linear or nonlinear functions
- $H(\cdot)$: observation functions
- $\boldsymbol{\theta}$: a vector of unknown parameters
- $\boldsymbol{X}_0$: Initial conditions
- $\boldsymbol{e}(t_i)$: measurement error
- Closed-form solution may not exist

Modeling Objectives

- Forward Problems: $\theta \mapsto \mathbb{P}_\theta$
  - Predictions
  - Simulations
- Inverse Problems: $Y \mapsto \theta \in \Theta$
  - θ: constant parameters
  - $\theta(t)$: time-varying parameters

Inverse Problems: More Challenging

- Identifiability issues
- Need to solve the forward problem first
- The forward problem: often no close-form solution, need intensive numerical evaluations
- Lack of development of statistical methods, theories and software tools for complex math models
- Computationally challenging

Identifiability issues

- Need to be investigated before the inverse problem
- Theoretical identifiability: Mathematical identifiability
- Practical identifiability: Statistical and numerical identifiability

Identifiability issues: References

- Wu, H., Zhu, H., Miao, H., and Perelson, A.S. (2008), Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models, Bulletin of Mathematical Biology, 70(3), 785-799.
- Miao, H., Dykes, C., Demeter, L.M., Cavenaugh, J., Park, S.Y., Perelson, A.S., and Wu, H. (2008), Modeling and Estimation of Kinetic Parameters and Replicative Fitness of HIV-1 from Flow-Cytometry-Based Growth Competition Experiments, Bulletin of Mathematical Biology, 70, 1749-1771.
- Miao, H., Dykes, C., Demeter, L., Wu, H. (2009), Differential Equation Modeling of HIV Viral Fitness Experiments: Model Identification, Model Selection, and Multi-Model Inference, Biometrics, 65, 292-300.
- Liang, H., Miao, H., and Wu, H. (2010), Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model, Annals of Applied Statistics, 4, 460-483.
- Miao, H., Xia, X., Perelson, A.S., Wu, H. (2011), On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics, SIAM Review, 53(1): 3-39.

Statistical Estimation Methods for ODE Models

- The nonlinear LS or MLE principle:
  - numerically solve the ODE
  - global optimization method: differential evolution algorithm or scatter search methods
  - computation: expensive and convergence problems
- Smoothing-based approaches
  - avoid numerically solving the ODE
  - easy to implement: fast
  - efficient for high-dimensional ODEs
  - not accurate
- Bayes methods
  - use prior to solve the identifiability problem
  - good for both cross-sectional data and longitudinal data
  - computation: expensive

Statistical Estimation Methods for ODE Models: Longitudinal Data

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-step smoothing-based 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:

$$\begin{aligned} \frac{d}{dt}\boldsymbol{X}(t) &= G[\boldsymbol{X}(t), \boldsymbol{\theta}_i], \quad \boldsymbol{X}(0) = \boldsymbol{X}_{i0} \\ \boldsymbol{Y}_i(t_i) &= H_i[\boldsymbol{X}_i(t_i), \boldsymbol{\theta}_i] + \boldsymbol{e}_i(t_i), \quad i = 1, \ldots, n \end{aligned} \tag{3}$$

  - $\boldsymbol{X}_i(t_i)$: ODE solution for Subject i.
  - $\boldsymbol{Y}_i = (y_{i1}(t_1), \cdots, y_{im_i}(t_{m_i}))^T$: Data from Subject i
  - $\mathbf{e}_i = (e_i(t_1), \cdots, e_i(t_{m_i}))^T \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i})$: Measurement error

- Between-subject variation:

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i, \qquad [\mathbf{b}_i | \boldsymbol{\Sigma}] \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

  - $\boldsymbol{\mu}$: population parameter
  - $\mathbf{b}_i$: random effects

Mixed-Effects ODE Model: NLME

Estimation and inference: SAEM

- Maximum likelihood estimation (MLE) principle: Treat random-effects as missing and use the EM algorithm
- Stochastic Approximation EM (SAEM): Delyon, Lavielle and Moulines, *Annals of Statistics* (1999)
- SAEM coupled with MCMC for ODE : Kuhn and Lavielle, *Computational Statistics and Data Analysis* (2005)
- SAEM coupled with MCMC for precomputation and parallel version for ODE and PDE: Grenier, Louvet, Vigneaux, *ESAIM Mathematical Modelling and Numerical Analysis* (2014)

Mixed-Effects ODE Model: SAEM

- Maximum likelihood estimation (MLE) principle: Maximize the observed-data likelihood with respect to the population parameters $l(\boldsymbol{\mu}) = \int \log f(\boldsymbol{Y}_i, \boldsymbol{b}_i; \boldsymbol{\mu}) d\boldsymbol{b}_1 \ldots d\boldsymbol{b}_n$
- EM Algorithm: iteratively maximize the following
  - E-step: evaluate

$$Q(\boldsymbol{\mu}|\boldsymbol{\mu}_k) = \int \log f(\boldsymbol{Y}_i, \boldsymbol{b}_i; \boldsymbol{\mu}) p(\boldsymbol{b}_i|\boldsymbol{Y}_i; \boldsymbol{\mu}_k) d\boldsymbol{b}_1 \ldots d\boldsymbol{b}_n$$

  - M-step: Obtain $\boldsymbol{\mu}_{k+1}$ by maximizing $Q(\boldsymbol{\mu}|\boldsymbol{\mu}_k)$
- SAEM: E-step is split into
  - Simulation Step (S-Step): Generate a realization of the missing data vector $\boldsymbol{b}$ based on the conditional distribution of $p(\boldsymbol{b}_i|\boldsymbol{Y}_i; \boldsymbol{\mu}_k)$
  - Stochastic approximation integration step: stochastic average

$$s_k = s_{k-1} + \gamma_k \left(\tilde{S}(\boldsymbol{Y}, \boldsymbol{b}_k) - s_{k-1} \right)$$

- SAEM with MCMC: Use MCMC (including Metropolis-Hastings algorithm) in the simulation step (S-Step) of the SAEM

Mixed-Effects ODE Model: SAEM

- Evaluation of likelihood function or conditional distribution: need to numerically solve ODE
- Computationally challenging
- Convergence: slow
- Many local solutions

Smoothing-Based Approaches

- Two-step decoupling approaches: Chen and Wu (JASA 2008, Statistica Sinica 2008) and Liang and Wu (JASA, 2008)
  - avoid numerically solving the ODE
  - easy to implement: fast
  - efficient for high-dimensional ODEs
  - not accurate
- Parameter cascading method: Ramsay et al. JRSS-B (2007) and Wang et al. Stat Comput 2014.
  - A 3-step iterative algorithm
  - Computationally stable
  - Convergence: slow

Smoothing-Based Approaches: Two-Step Method

Chen and Wu (JASA 2008, Statistica Sinica 2008) and Liang and Wu (JASA, 2008):

$$\boldsymbol{X}'(t_i) = F[\boldsymbol{X}(t_i), \boldsymbol{\theta}] \tag{4}$$

$$\boldsymbol{Y}(t_i) = \boldsymbol{X}(t_i) + \boldsymbol{e}_1(t_i), \quad \boldsymbol{e}_1(t_i) \sim (0, \sigma^2 \boldsymbol{I}), \tag{5}$$

- Step 1: Use a nonparametric smoothing to estimate $X(t)$ and $X'(t)$ from model (5).
- Step 2: Substitute the estimate $\hat{X}(t_i)$ into model (4) to obtain:

$$\hat{X}'(t_i) = F[\hat{X}(t_i), \boldsymbol{\theta}] + e_2(t_i). \tag{6}$$

Then fit the above regression model (6) to estimate $\boldsymbol{\theta}$.

Smoothing-Based Approaches: Two-Step Methods

- $F(\cdot)$: Linear or nonlinear function
- Step 2 decoupled the system of ODEs: Fit the ODE one-by-one
- Standard regression software tools can be used
- Fast
- Extension to higher-order numerical discretization-based algorithms: Wu, Xue andKuman (Biometrics 2012)
- Price to pay: Inaccurate
  - The derivative estimate may not be accurate
  - The decoupled system: Some information lost

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-step smoothing-based method: good for this purpose

Example: High-Dimensional ODEs for Longitudinal Gene Expression Data

Time course gene expression data: Dynamic gene regulatory network (GRN) reconstruction (Lu et al, JASA 2011)

1. Screening significant gene expression curves
2. Clustering individual gene into functional modules
3. Smoothing time course data to obtain population(mean) expression pattern and its derivative
4. Identifying significant regulations among different modules
5. Estimation refinement for functional module-based GRN for mixed-effects ODE models
6. Function enrichment analysis for annotating identified GRN

Step I: Screening significant gene expression curves

- Smoothing using the function principle component analysis (FPCA)

$$y_{ijk} = x_i(t_j) + \varepsilon_{ijk}, \quad x_i(t) = \sum_{l=0}^{L_i} c_{il}\phi_l(t),$$

  where $\phi_l(t)$ are principle components (eigenfunctions)
- Test statistics: goodness-of-fit (signal to noise ratio)

$$F_i = \frac{SS_i^0 - SS_i^1}{SS_i^1}$$

- Null distribution of the test statistic: permutation
- p-value for the $i-$th gene (proble set):

$$p = \sum_{b=1}^{B} \frac{\#\{j : F_j^b \geq F_i, j = 1, \ldots, n\}}{nB}$$

- Multiple testing adjustment: Benjamin and Hochberg (1995)

Step II: Clustering significant gene expression curves

- Dimension reduction
- Biological justification
  - Expression profile of a group of genes cannot be distinguished from each other
  - Genes in the same group share particular biological functions

Nonparametric Mixed-Effects Model with Mixture Model (Ma et al. 2006)

$$\boldsymbol{Y_i} \sim w_1\mathcal{N}(\mu_1(\boldsymbol{T_i}), \boldsymbol{\Sigma}_1) + w_2\mathcal{N}(\mu_2(\boldsymbol{T_i}), \boldsymbol{\Sigma}_2) + \cdots, +w_p\mathcal{N}(\mu_p(\boldsymbol{T_i}), \boldsymbol{\Sigma}_p) \quad (7)$$

$$i = 1, \cdots, n$$

$$\boldsymbol{Y_i} = \mu_k(\boldsymbol{T_i}) + \boldsymbol{Z_i b_i} + \boldsymbol{\epsilon_i}, \quad (8)$$

$\mu_k(.)$: mean curve of cluster k with a smoothing spline representation

Step III: Smoothing

- Time course data for all genes within the same module can be treated as longitudinal data
- Mixed-effects modeling approach is necessary
- Nonparametric models accounts for irregular expression pattern

Nonparametric Mixed-Effects Smoothing Model (Wu and Zhang 2002, 2006)

$$y_{ki}(t_{ij}) = M_k(t_{ij}) + V_{ki}(t_{ij}) + \epsilon(t_{ij}) \tag{9}$$

- Local polynomial
- Regression spline
- Smoothing spline
- Penalized spline

Step IV: 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

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

$$k = 1, \cdots, p; \quad t = t_1, t_2, \ldots, t_N$$

$$y_k(t) = \hat{M}'_k(t) \text{ and } x_j(t) = \hat{M}_j(t)$$

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)

Step V: Estimation Refinement: Stochastic Approximation EM (SAEM) Algorithm

Mixed-Effects ODE Model for Module k

$$\frac{dx_{ki}}{dt} = \sum_{j=1}^{m_k} \beta_{kij} M_{[kj]}(t), \quad i = 1, \cdots, n_k; \quad k = 1, \ldots, p, \tag{10}$$

Longitudinal Measurement Model

$$y_{ki}(t) = x_{ki}(t) + \varepsilon_{ki}(t) \tag{11}$$

Random Effects Model

$$\boldsymbol{\beta}_{ki} = \boldsymbol{\beta}_k + \mathbf{b}_{ki} \tag{12}$$

$$\mathbf{b}_{ki} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_k)$$

Step VI: Function Enrichment Analysis

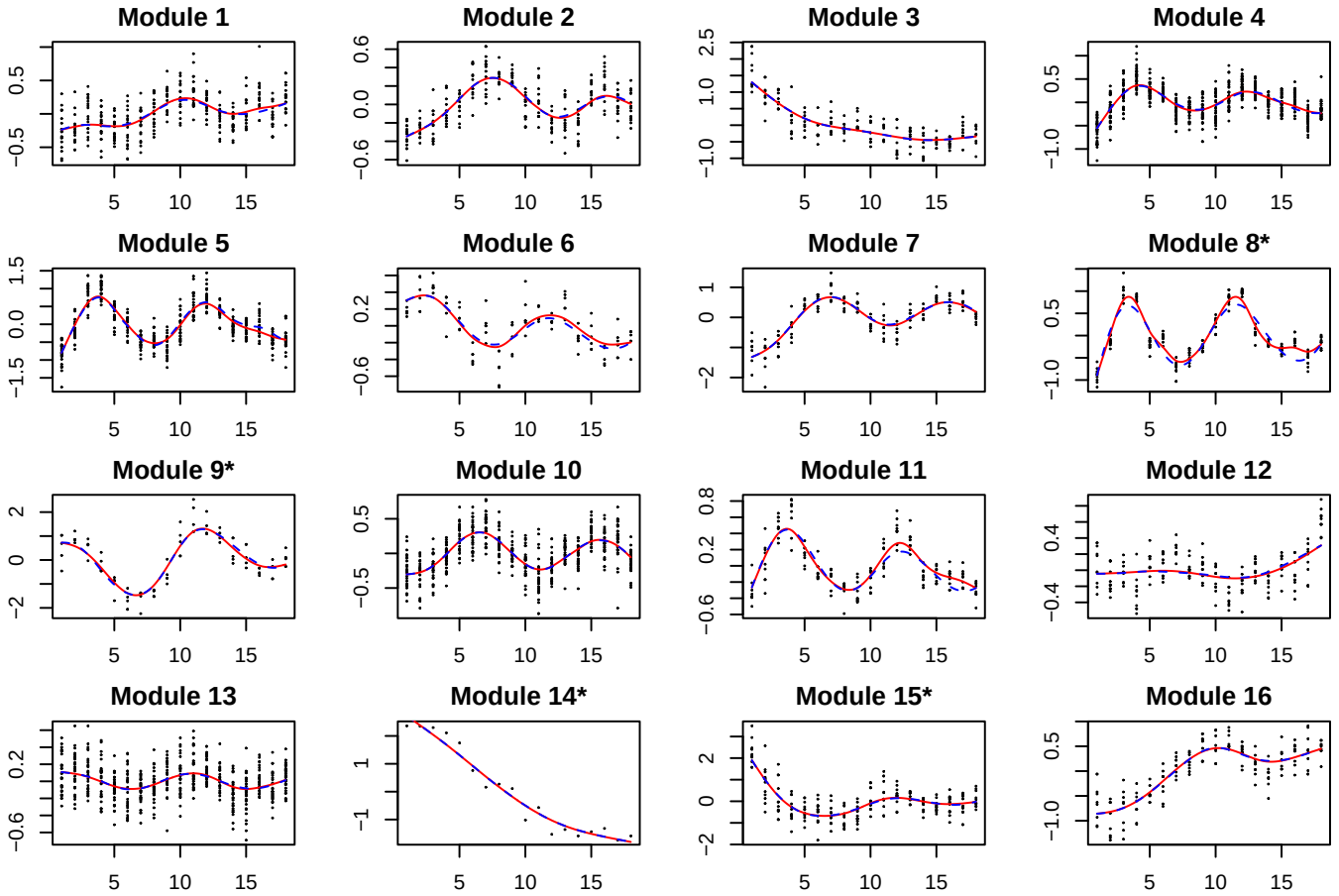
- A tool for annotating modules
- Certain biological function(s) may be over-represented by genes of the same module compared to population of genes in an organism or a biological process
- Hypergeometric distribution can detect such enriched function(s)

Example: Identification of Dynamic GRN for Yeast Cell Cycle

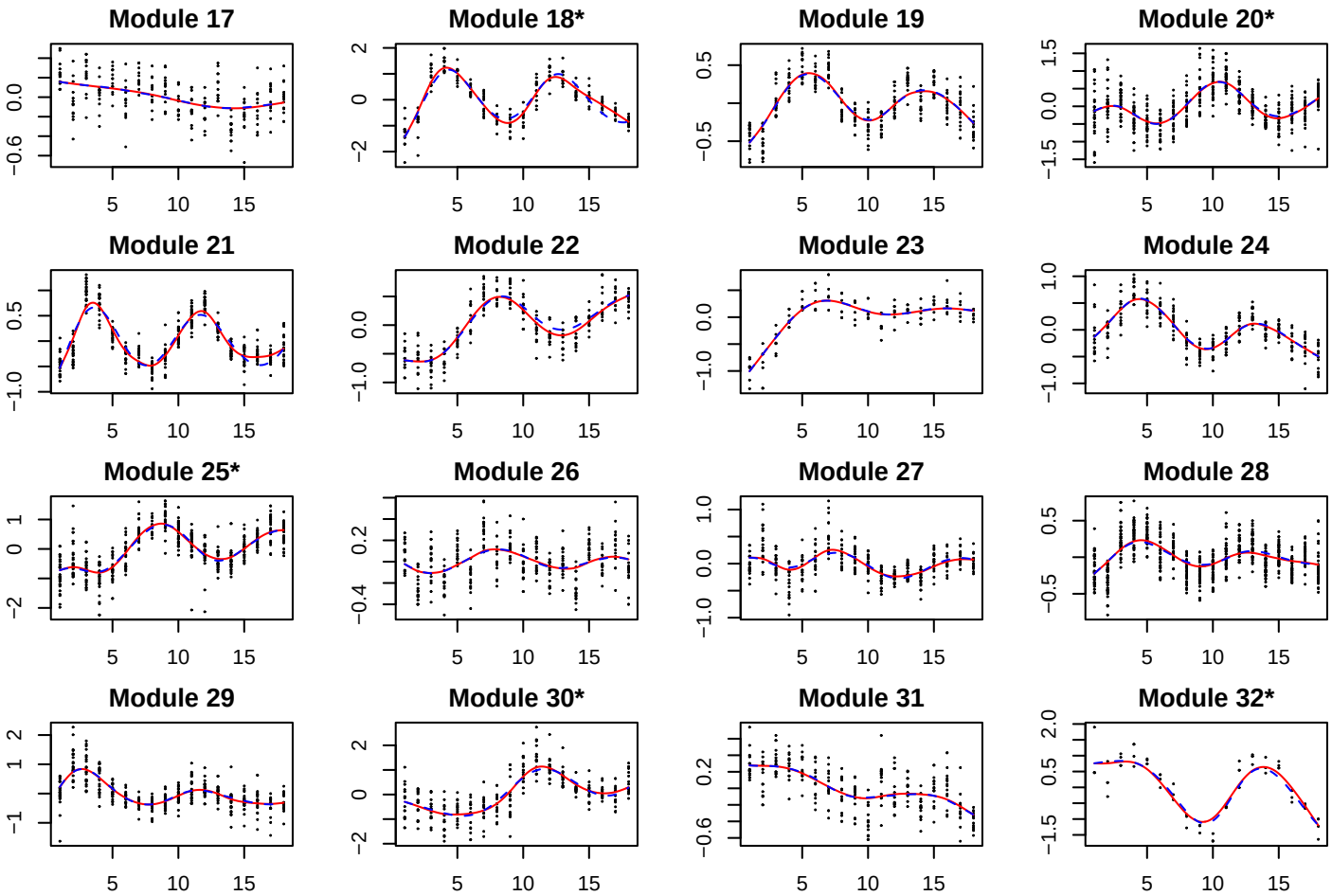
DNA microarrays experiment: 18 equally spaced time points during two cell cycles (Spellman 1998)

- **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

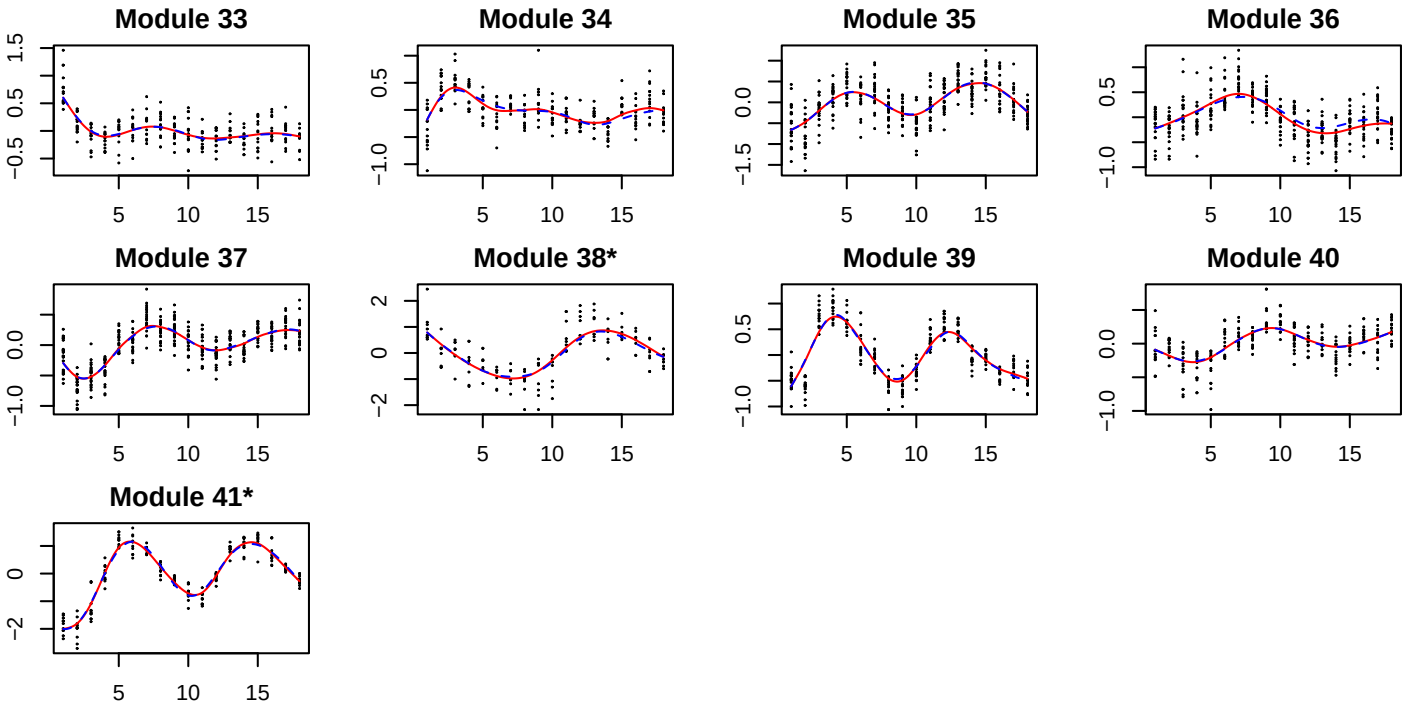
Yeast Cell Cycle Gene Expression Profile



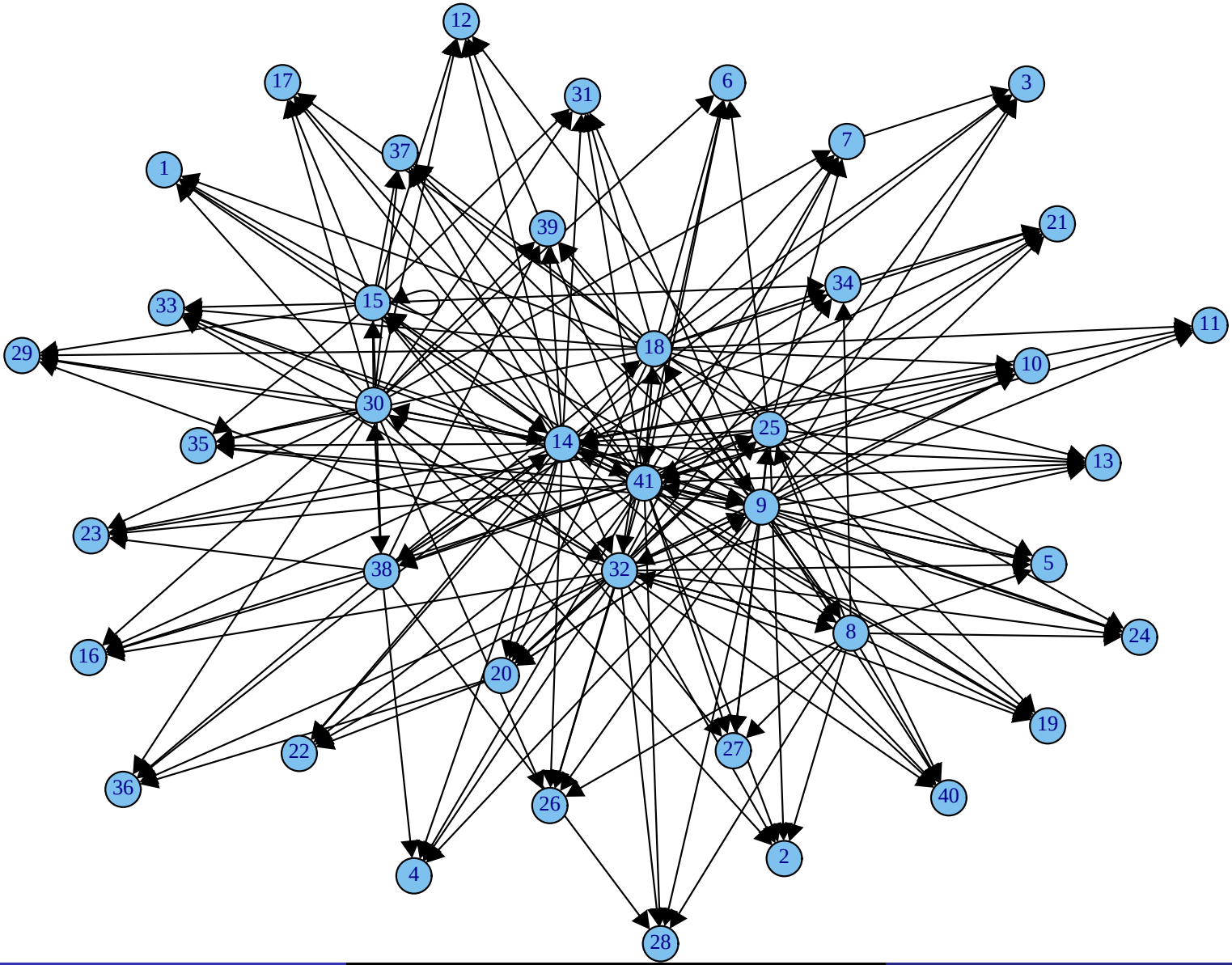
Yeast Cell Cycle Gene Expression Profile



Yeast Cell Cycle Gene Expression Profile



Graph of Yeast Cell Cycle GRN



Bayesian Methods: Mechanistic Model for HIV Infection with Treatment

Huang, Liu and Wu, Biometrics (2006)

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

$$\begin{aligned} \frac{d}{dt}T &= \lambda - \rho T - [1-\gamma(t)]kTV \\ \frac{d}{dt}T^* &= [1-\gamma(t)]kTV - \delta T^* \\ \frac{d}{dt}V &= N\delta T^* - cV \end{aligned} \tag{13}$$

  - 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 (13): no closed-form solution

Antiviral Drug Efficacy Model

- A modified E_{max} (M-M) model for drug efficacy:

$$\gamma(t) = \frac{C(t)A(t)}{\phi IC_{50}(t) + C(t)A(t)} = \frac{IQ(t)A(t)}{\phi + IQ(t)A(t)}, \quad 0 \leq \gamma(t) \leq 1 \tag{14}$$

  - $C(t)$: the plasma drug concentration
  - $A(t)$: drug adherence measurements
  - IC_{50}: in vitro phenotype drug resistance marker
  - ϕ: 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

Drug Susceptibility Model

- Phenotype marker IC_{50} is used to quantify agent-specific drug sensitivity
- The function: to describe changes overtime in IC_{50}

$$IC_{50}(t) = \begin{cases} I_0 + \frac{I_r - I_0}{t_r} t & \text{for } 0 < t < t_r, \\ I_r & \text{for } t \geq t_r, \end{cases} \tag{15}$$

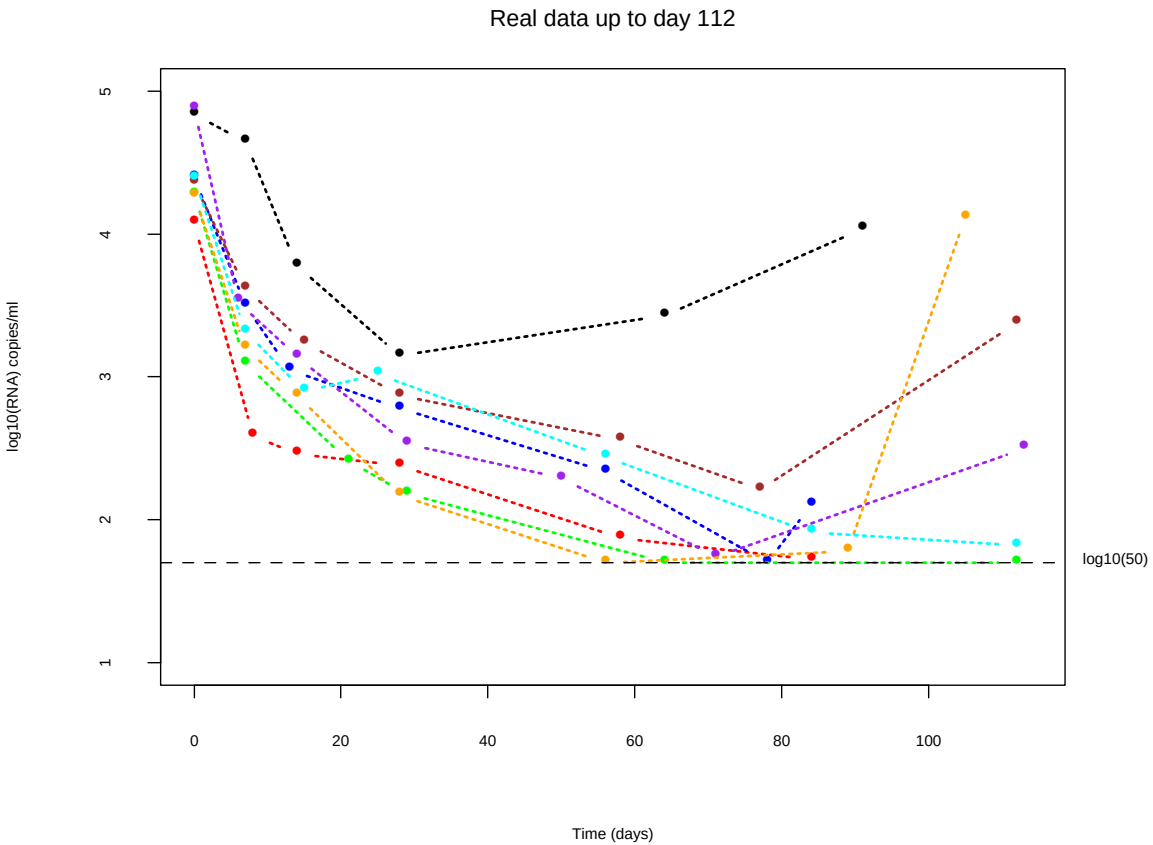
  - I_0 and I_r: respective values of $IC_{50}(t)$ at baseline and time point t_r at which drug resistant mutations appear
  - If $I_r = I_0$, no resistance mutation developed during treatment

A Challenging Problem

How to Estimate the Unknown Parameters in the Dynamic PK/PD 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



Bayesian Hierarchical Modeling Approach

- Propose a three-stage hierarchical (mixed-effects) model
- Advantages of Bayesian hierarchical modeling approach
  - Naturally incorporate prior information
  - Deal with extremely complicated models such as nonlinear differential equation models
  - Ease the identifiability problem
  - Use posterior distributions to easily answer inference questions
  - Estimate parameters for both population and individuals

Bayesian Modeling

- A three-stage Bayesian hierarchical model
- Stage 1. Within-subject variation:

$$\mathbf{y}_i = \mathbf{f}_i(\boldsymbol{\theta}_i) + \mathbf{e}_i, \qquad [\mathbf{e}_i|\sigma^2, \boldsymbol{\theta}_i] \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i})$$

  - $\mathbf{f}_i(\boldsymbol{\theta}_i) = (f_{i1}(\boldsymbol{\theta}_i, t_1), \cdots, f_{im_i}(\boldsymbol{\theta}_i, t_{m_i}))^T$: **ODE solutions.**
  - $\mathbf{y}_i = (y_{i1}(t_1), \cdots, y_{im_i}(t_{m_i}))^T$: **Data from Subject** i
  - $\mathbf{e}_i = (e_i(t_1), \cdots, e_i(t_{m_i}))^T$: **Measurement error**
- **Stage 2. Between-subject variation:**

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i, \qquad [\mathbf{b}_i|\boldsymbol{\Sigma}] \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

- **Stage 3. Hyperprior distributions:**

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

  - **Gamma (**Ga**), Normal (**$\mathcal{N}$**) and Wishart (**Wi**): independent distributions**
  - **Hyper-parameters** $a, b, \boldsymbol{\eta}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}$ **and** ν**: 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 σ^{-2}, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}^{-1}$
  - Metropolis-Hastings step: no closed form of conditional distributions for θ_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

Clinical Data–Results of Population Parameters

Parameter	PM	SD	95% CI
ϕ	**2.1091**	**0.6354**	**(1.2143, 3.6392)**
c	**2.9867**	**0.1466**	**(2.7139, 3.2881)**
δ	**0.3729**	**0.0184**	**(0.3387, 0.4105)**
λ	**100.645**	**4.9431**	**(91.497, 110.830)**
ρ	**0.0997**	**0.0049**	**(0.0905, 0.1099)**
N	**1004.988**	**49.795**	**(912.074, 1106.654)**
k	9.183×10^{-6}	0.290×10^{-6}	$(8.632 \times 10^{-6}, 9.774 \times 10^{-6})$

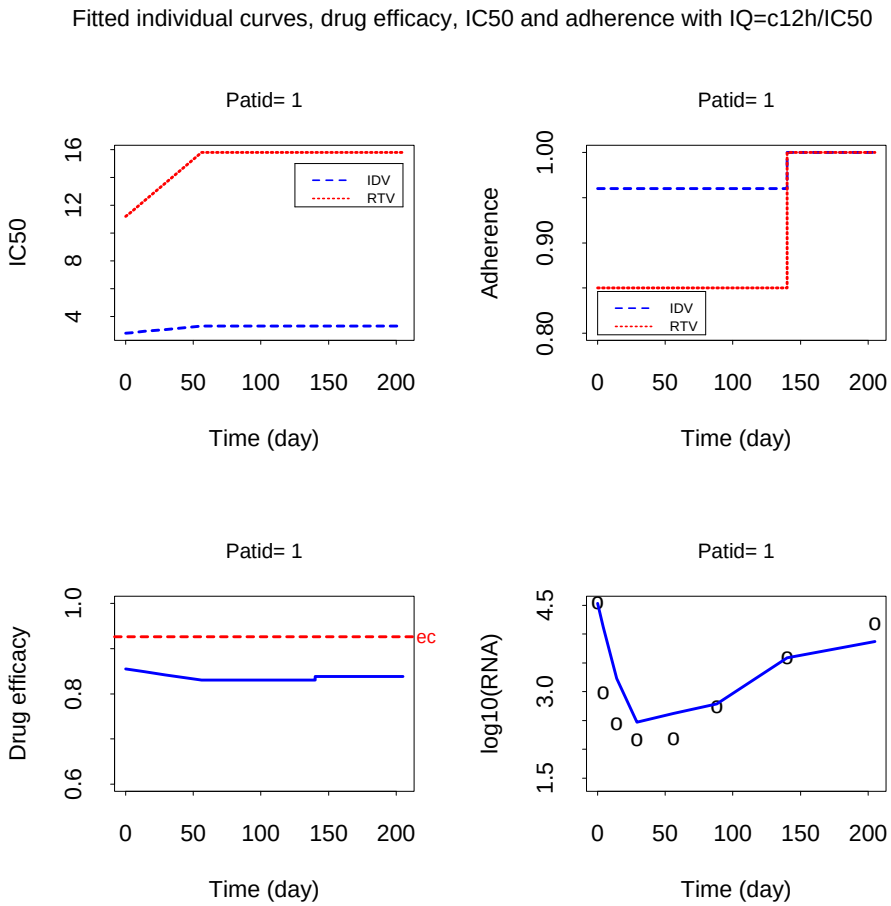
- Posterior mean for the population parameter ϕ is 2.1091 with a SD of 0.6354 and the 95% CI of (1.2143, 3.6392)
- As ϕ plays a role of transforming the *in vitro* IC_{50} into *in vivo* IC_{50}, our estimate shows that there is about 2-fold difference between *in vitro* IC_{50} and *in vivo* IC_{50}

Clinical Data–Results of Individual Parameters

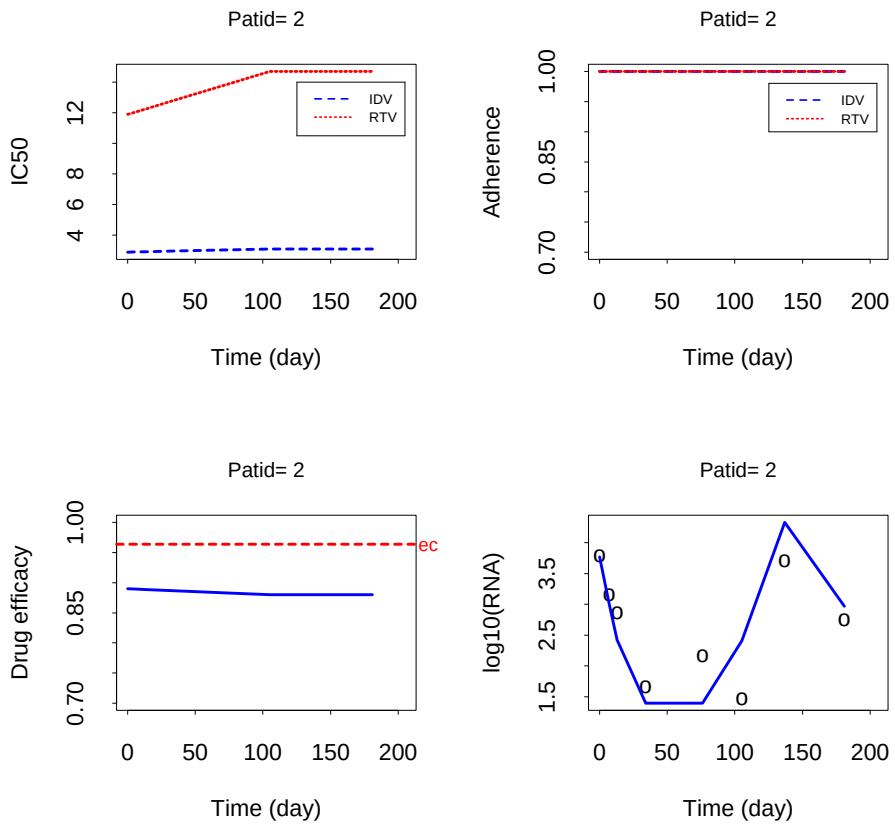
Patient	ϕ_i	c_i	δ_i	λ_i	ρ_i	N_i	k_i	e
1	**0.447**	**2.254**	**0.270**	**410.462**	**0.024**	**456.757**	8.33×10^{-6}	**0.97**
2	**5.371**	**2.969**	**1.183**	**29.619**	**0.426**	**4795.813**	10.84×10^{-6}	**0.17**
3	**3.723**	**2.283**	**0.456**	**36.877**	**0.289**	**3258.347**	8.66×10^{-6}	**0.37**
4	**4.960**	**2.761**	**0.798**	**44.956**	**0.313**	**3051.988**	9.09×10^{-6}	**0.34**
5	**7.066**	**2.306**	**0.663**	**71.295**	**0.201**	**2735.239**	6.54×10^{-6}	**0.64**
6	**0.786**	**4.633**	**0.183**	**375.882**	**0.025**	**247.416**	11.18×10^{-6}	**0.89**
7	**0.091**	**7.008**	**0.299**	**4015.398**	**0.003**	**30.559**	18.54×10^{-6}	**0.98**
8	**8.484**	**2.280**	**0.663**	**32.722**	**0.416**	**4530.531**	8.37×10^{-6}	**0.24**

- **The individual-specific parameter estimates suggest a large inter-subject variation**
- **The model provides a good fit to the clinical data**

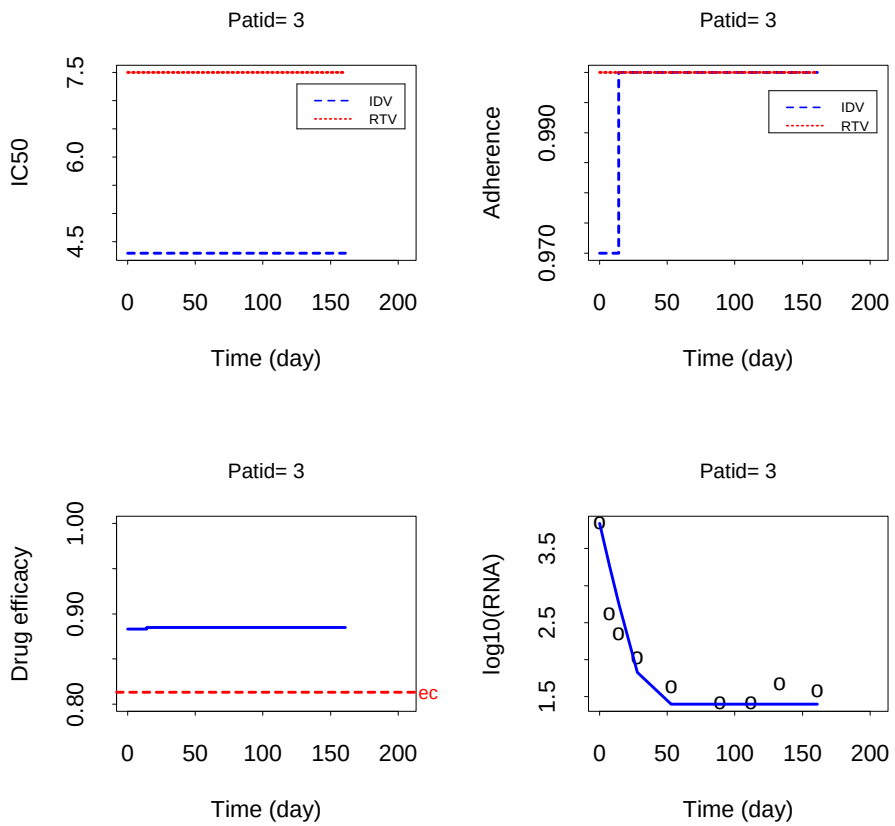
Patient 1


Fitted individual curves, drug efficacy, IC50 and adherence with IQ=c12h/IC50
Patid= 1
IC50
4
8
12
16
IDV
RTV
0
50
100
150
200
Time (day)
Patid= 1
Adherence
0.80
0.90
1.00
IDV
RTV
0
50
100
150
200
Time (day)
Patid= 1
Drug efficacy
0.6
0.8
1.0
ec
0
50
100
150
200
Time (day)
Patid= 1
log10(RNA)
1.5
3.0
4.5
0
50
100
150
200
Time (day)

Patient 2


Patid= 2
IC50
4 6 8 12
IDV
RTV
0 50 100 150 200
Time (day)
Patid= 2
Adherence
0.70 0.85 1.00
IDV
RTV
0 50 100 150 200
Time (day)
Patid= 2
Drug efficacy
0.70 0.85 1.00
ec
0 50 100 150 200
Time (day)
Patid= 2
log10(RNA)
1.5 2.5 3.5
0 50 100 150 200
Time (day)

Patient 3


Patid= 3
IC50
7.5
6.0
4.5
IDV
RTV
0
50
100
150
200
Time (day)
Patid= 3
Adherence
0.990
0.970
IDV
RTV
0
50
100
150
200
Time (day)
Patid= 3
Drug efficacy
1.00
0.90
0.80
ec
0
50
100
150
200
Time (day)
Patid= 3
log10(RNA)
3.5
2.5
1.5
0
50
100
150
200
Time (day)

State-Space Models (SSM)

Linear SSM:

$$X_{t+1} = F_t X_t + V_t, \quad V_t \sim (0, Q_t) \tag{16}$$

$$Y_t = G_t X_t + W_t, \quad W_t \sim (0, R_t) \tag{17}$$

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

Mixed-Effects State-Space Models

Liu, Lu, Niu and Wu, Biometrics (2011):

- *Stage 1:* Within-subject variation

$$\begin{aligned} \boldsymbol{X}_{i,t+1} &= \boldsymbol{F}(\boldsymbol{\theta}_i)\boldsymbol{X}_{it} + \boldsymbol{V}_{it}, \quad \boldsymbol{V}_{it} \sim \mathrm{N}(0, \boldsymbol{Q}), && (18) \\ \boldsymbol{Y}_{it} &= \boldsymbol{G}(\boldsymbol{\theta}_i)\boldsymbol{X}_{it} + \boldsymbol{W}_{it}, \quad \boldsymbol{W}_{it} \sim \mathrm{N}(0, \boldsymbol{R}), && (19) \\ i &= 1, \ldots, m; \quad \text{for each i}, t = 1, \ldots, n_i. \end{aligned}$$

- *Stage 2:* Between-subject variation

$$\theta_i = \theta + b_i, \quad b_i \sim (0, D), \qquad (20)$$

θ: population parameter
b_i: random effect
D: covariance of random effects

Mixed-Effects State-Space Models

Goals:

- Estimate unknown parameters: MLE and EM algorithm
- Estimate individual state variables: Standard Kalman filter
- Estimate Population state variable X_t: Challenging

Mixed-Effects State-Space Models

Estimate population state variable X_t

- Definition of population state variable: Individual state=a dispersion from the population state,

$$X_{i,t} = X_t + Z_{i,t}, \quad Z_{i,t} \sim (0, D_t) \tag{21}$$

- But $X_{i,t}$ is unobservable
- Use the estimated state vectors, $\tilde{X}_{i,t}$: Decompose

$$\tilde{X}_{i,t} = X_{i,t|n} = X_t + Z_{i,t} + \varsigma_{i,t}, \tag{22}$$

  where $\varsigma_{i,t} \sim (0, \Sigma_{i,t})$: estimation error of $X_{i,t}$. $\Sigma_{i,t}$ can be obtained by Kalman smoothing.
- Treat $\tilde{X}_{i,t}$ as 'data'
- Use EM algorithm to estimate the population state variable and dispersion variance D_t

Mixed-Effects State-Space Models

Liu, Lu, Niu and Wu, Biometrics (2011):

- SAEM Algorithm
- Bayesian method
- Application to HIV viral dynamic models

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?
  - How to deal with nonlinear ODEs?
- Constrained ODEs
- Large ODE system characteristic analyses

Our recent work in high-dimensional ODE models

- Lu, T., Liang, H., Li, H., Wu, H. (2011), High Dimensional ODEs Coupled with Mixed-Effects Modeling Techniques for Dynamic Gene Regulatory Network Identification, JASA, 106, 1242-1258.
- Wu, H., Xue, H., Kumar A. (2012), Numerical Discretization-Based Estimation Methods for Ordinary Differential Equation Models via Penalized Spline Smoothing with Applications in Biomedical Research, Biometrics, 68(2), 344-353.
- Wu, S, and Wu, H. (2013), More Powerful Significant Testing for Time Course Gene Expression Data Using Functional Principal Component Analysis Approaches, BMC Bioinformatics, 14:6.
- Wu, H., Lu, T.+, Xue, H., and Liang, H. (2014), Sparse Additive ODEs for Dynamic Gene Regulatory Network Modeling, JASA, 109:506, 700-716.
- Wu, S., Liu, Z.P.+, Qiu, X., and Wu, H. (2014), Modeling genome-wide dynamic regulatory network in mouse lungs with influenza infection using high-dimensional ordinary differential equations, PLOS ONE, 9(5):e95276.
- Linel, P., Wu, S., Deng, N., Wu, H. (2014), Dynamic transcriptional signatures and network responses for clinical symptoms in influenza-infected human subjects using systems biology approaches, Journal of PK/PD, 41, 509-521.
- Qiu, X. et al. (2015), Diversity in Compartmental Dynamics of Gene Regulatory Networks: The Immune Response in Primary Influenza A Infection in Mice, PLoS ONE, 10(9).

Thank You!