Partial Differential Equation Modeling of Flow Cytometry Data from CFSE-based Proliferation Assays

W. Clayton Thompson
In collaboration with ...

Center for Quantitative Sciences in Biomedicine
Center for Research in Scientific Computation
Department of Mathematics
North Carolina State University

01 March 2012
H.T. Banks
Thesis Advisor

Karyn Sutton
Dept. Mathematics, University of Louisiana at Lafayette

Tim Schenkel
Department of Virology, Saarland University, Homburg, Germany

Jordi Argilaguet, Sandra Giest, Cristina Peligero, Andreas Meyerhans
ICREA Infection Biology Lab, Univ. Pompeu Fabra, Barcelona, Spain

Gennady Bocharov
Institute of Numerical Mathematics, RAS, Moscow, Russia

Marie Doumic
INRIA Rocquencourt, Projet BANG, Rocquencourt, France
1. CFSE Data Overview
2. PDE Modeling of CFSE Data
3. Data Statistical Model
4. Next Steps
Data Overview

(A. Meyerhans)
Cells cultured with CFDA-SE then washed
- CFDA-SE becomes protein-bound and fluorescent CFSE
- Dye split between daughter cells at division
- Dye naturally turns over/degrades (very slowly)
- Fluorescence Intensity (FI) of CFSE measured via flow cytometry
- FI linear with dye concentration \( \Rightarrow FI \propto \text{mass} \)
- Several advantages over other dyes/techniques
CFSE Data Set

CFSE Time Series Histogram Data

Structured Population Density [numbers/UI]

Label Intensity $z$ [Log UI]

$t = 0$ hrs
$t = 24$ hrs
$t = 48$ hrs
$t = 72$ hrs
$t = 96$ hrs
$t = 120$ hrs
Goals of Modeling

- Cellular ‘Dynamic Responsiveness’
- Link cell counts with proliferation/death rates
  - Population doubling time
  - Cell viability
  - Biological descriptors (cell cycle time, etc.)
- Uncertainty Identification, Variability Quantification...
  - ... in the experimental procedure
  - ... for estimated rates/etc
- Analyze cell differentiation and division-linked changes
- Investigate immunospecific extracellular signaling pathways
- Comparison among donors/cell types/disease progression
Traditional Approach (curve fitting)

- Fit data with gaussian curves to determine approximate cells per generation
- Traditional ‘semi-quantitative analysis’ pioneered by Gett and Hodgkin et al. (2000)

Traditional Approach (cont’d)

- Gett-Hodgkin method quick, easy to implement, useful comparisons between data sets (e.g. stimulation conditions)
- Compatible with ODE, DDE models; ‘indirect fitting’ for parameter estimation
- Generalizations, extensions, and various other modeling efforts
  - Smith-Martin model (with generalizations)
  - Cyton model
  - Branching process models
All previous work with cell numbers determined by deconvolution.

Alternatively, we propose to fit the CFSE histogram data directly:
- Capture full behavior of the population density
- No assumption on the shape of CFSE uptake/distribution

Histogram presentation of cytometry data makes structured population models a natural choice:
- Key ideas first formulated by Luzyanina et al., 2007
- FI (or log FI) ↔ Division Number
This model must account for (Luzyanina et al., 2007):

- Dilution of CFSE as cells divide (AutoFI)
- Slow decay of FI over time (CFSE turnover)
- Asynchronous division times

![Data in Logarithmic Coordinate (z)](chart.png)
Cellular Autofluorescence

Donor 1 CD4 Data, t = 0 hrs

\[ X_i = X_i^{\text{CFSE}} + X^{\text{Auto}} \]

\[ X_{i+1} = \frac{X_i^{\text{CFSE}}}{2} + X^{\text{Auto}} \]

Cell Counts

z [Log UI]

0 1 2 3 4 5 6

0 3000 6000 9000 12000 15000
(C. Parish, Fluorescent dyes for lymphocyte migration and proliferation studies, *Immunology and Cell Biol.* 77 (1999), 499–508.)
‘Biphasic Decay’

\[ \frac{dx}{dt} = \nu(x) = c(x - x_a) \]

**Exponential**

\[ \frac{dx}{dt} = \nu(t, x) = c(x - x_a)e^{-kt} \]

**Gompertz**

Clay Thompson  
CFSE Modeling
Fragmentation Mathematical Model

- Structured density $n(t, x)$ (cells/UI)
- (Exponential) Proliferation rate $\alpha(t, x)$
- (Exponential) Death rate $\beta(x)$
- Gompertz decay rate, $\nu(t, x) = c(x - x_a)e^{-kt}$

$$\frac{\partial n(t, x)}{\partial t} + \frac{\partial [\nu(t, x)n(t, x)]}{\partial x} = -(\alpha(t, x) + \beta(x))n(t, x) + \chi_{[x_a, x^*]}4\alpha(t, 2x - x_a)n(t, 2x - x_a)$$
Inverse Problem

- Parameters $x_a$, $c$, $k$, $\alpha(t, y)$, $\beta(y)$ to be determined by fitting to data.
- Need (finite-dimensional) parameterization of $\alpha$ and $\beta$.
  - Piecewise linear functions
- Statistical properties of error currently unknown
- Use OLS (independent, identically distributed, constant variance error) for proof of concept

$$\hat{\theta}_{OLS} = \arg \min_{\theta \in \Theta} \sum_{i=1}^{I} \sum_{j=1}^{J} (I[\hat{N}](t_i, z_j; \theta) - N_{ij})^2 = \arg \min J(\theta),$$

- Forward solve with \texttt{hpde} by L.Shampine (Lax-Wendroff)
- Use \texttt{fmincon} (BGFS + active set) for optimization
Time-Independent Proliferation is Insufficient
Time-Dependent Proliferation is Sufficient

Best-Fit Histograms, t = 24 hrs

Best-Fit Histograms, t = 48 hrs

Best-Fit Histograms, t = 96 hrs

Best-Fit Histograms, t = 120 hrs
Model is capable of precisely fitting the observed data
- $c$, $k$, $x_a$ estimated consistently (as $\alpha$ and $\beta$ nodes change), though subject to high experimental variability
- Time-dependence of the proliferation rate is an essential feature of the model
- Biologically relevant average values of proliferation and death (in terms of number of divisions undergone) are easily computable.
- But...
  - Still cannot compute cell numbers
  - Data overlap affecting estimated rates (?)
  - Large number of parameters necessary
Fragmentation Model Summary (cont’d)

\[
\frac{\partial n(t, x)}{\partial t} + \frac{\partial [v(t, x)n(t, x)]}{\partial x} = -(\alpha(t, x) + \beta(x))n(t, x) + \chi[x_a, x^*]4\alpha(t, 2x - x_a)n(t, 2x - x_a)
\]

- Applications to protein fragmentation and aggregation
- Possible generalizations to size/volume structure

**Division Structure: The Compartmental Model**

- Use compartments (on division number) to eliminate fragmentation terms
- No need for structure dependence of estimated rates

\[
\frac{\partial n_0}{\partial t} + \frac{\partial [v(t, x)n_0(t, x)]}{\partial x} = - (\alpha_0(t) + \beta_0(t))n_0(t, x)
\]

\[
\frac{\partial n_1}{\partial t} + \frac{\partial [v(t, x)n_1(t, x)]}{\partial x} = - (\alpha_1(t) + \beta_1(t))n_1(t, x) + R_1(t, x)
\]

\[\vdots\]

\[
\frac{\partial n_{\text{max}}}{\partial t} + \frac{\partial [v(t, x)n_{\text{max}}(t, x)]}{\partial x} = - \beta_{\text{max}}(t)n_{\text{max}}(t, x) + R_{\text{max}}(t, x)
\]

where \( R_i(t, x) = 4\alpha_{i-1}(t)n_{i-1}(t, 2x - x_a) \) for \( 1 \leq i \leq i_{\text{max}} \)
Method of Characteristics Solution

\[ n_0(t, x(t; s)) = \Phi_0(s) \exp \left( - \int_0^t f_0(\tau) d\tau \right) \]

\[ n_i(t, x(t; s)) = \Phi_i(s) \exp \left( - \int_0^t f_i(\tau) d\tau \right) \]

\[ + \int_0^t R_i(\tau, x(\tau; s)) \exp \left( - \int_\tau^t f_i(\xi) d\xi \right) d\tau \]

where \( f_i(t) = \alpha_i(t) + \beta_i(t) - ce^{-kt} \)

The cell numbers can be easily computed \( N_i(t) = \int n_i(t, x) dx \)
Parameterizations

B1 $\beta_i(t) = 0$ for all $i$ and for all $t$
B2 $\beta_i(t) = \beta$ for all $i$ and for all $t$
B3 $\beta_0(t) = \beta_0$, $\beta_i(t) = 0$ for $i \geq 1$
B4 $\beta_0(t) = \beta_0$, $\beta_i(t) = \beta$ for $i \geq 1$
B5 $\beta_i(t) = \beta_i$ for each $i$

A1 $\alpha_0(t) = \alpha_0$; $\alpha_i(t) = \alpha$ for all $i$
A2 $\alpha_i(t) = \alpha_i$ for all $t$
A3 $\alpha_0(t) = \alpha_0 \chi[t > t^*]$; $\alpha_i(t) = \alpha$ for all $i$
A4 $\alpha_0(t) = \alpha_0 \chi[t > t^*]$; $\alpha_i(t) = \alpha_i$
A5 piecewise linear functions of time (see below)
AutoFI appears approximately lognormally distributed

Dynamic properties ignored (for now)

Can study effective design of intracellular dyes
\[ \eta(t, x) = E[n(t, x; x_a)|P] = \int_{x_{a,\min}}^{x_{a,\max}} n(t, x; x_a) dP(x_a) \]

\[ \frac{dP}{dx_a} = p(x_a) = \frac{1}{x_a \sigma \sqrt{2\pi}} \exp \left( -\frac{(\log x - \mu)^2}{2\sigma^2} \right) \]

where

\[ \mu = \log(E[x_a]) - \frac{1}{2} \log \left( 1 + \frac{\text{Var}(x_a)}{E[x_a]^2} \right) \]

\[ \sigma^2 = \log \left( 1 + \frac{\text{Var}(x_a)}{E[x_a]^2} \right) \]
Another Inverse Problem

- Population density $n(t, x) = \sum_{i=0}^{i_{\text{max}}} n_i(t, x)$
- Use OLS framework again—assume constant variance error

$$\hat{\theta}_{\text{OLS}}(n^j_k) = \arg \min_{\theta \in \Theta} J(\theta | n^j_k)$$

$$= \arg \min_{\theta \in \Theta} \sum_{k,j} \left( I[\tilde{n}](t_j, z^j_k; \theta) - n^j_k \right)^2$$

Need to compare different parameterizations (model comparison)–Akaike Information Criterion

$$AIC = m \log \left( \frac{J(\hat{\theta}_{\text{OLS}})}{m} \right) + 2p$$
Best-fit, AIC-selected results

Calibrated Model, t = 24hrs

Calibrated Model, t = 48hrs

Calibrated Model, t = 96hrs

Calibrated Model, t = 120hrs
Cell Numbers

\[ N_i(t) = \int n_i(t, x) \]

\[ P_i(t) = \frac{N_i(t)}{2^i} \]

Population doubling time and precursor viability easily computable
Model Results and Conclusions

- Cell/precursor numbers (per generation) easy to compute
- More complex models receive highest ranking
  - Highly time-dependent proliferation rates (A5)
  - Heterogeneous death rates (B5)
  - Distributed AutoFI is an important modeling feature

But...

- AIC may be biased by statistical model
- ‘Time-dependence’ possibly a byproduct of Malthusian form
- Cell counts between data points biased by model form
The Statistical Model

- Links the mathematical model to the data
- Implications for estimation procedure

\[ N_k^j = I[\tilde{n}](t_j, z_k^j; \theta_0) + \varepsilon_{kj} \]

- Currently using constant variance (CV) model, \( Var(\varepsilon_{kj}) = \sigma_0^2 \) (⇒ Absolute Error)
- Could use constant coefficient of variance (CCV), \( Var(\varepsilon_{kj}) = \sigma_0^2 I[\tilde{n}](t_j, z_k^j; \theta_0)^2 \) (⇒ Relative Error)
Residual Plots

Residuals vs Model, t =24

Residuals vs Model, t =48

Residuals vs Model, t =96

Residuals vs Model, t =120

Modified Residuals vs Model, t =24

Modified Residuals vs Model, t =48

Modified Residuals vs Model, t =96

Modified Residuals vs Model, t =120
Residual Plots (cont’d)

Residuals vs Model

Modified Residuals vs Model
New Statistical Model

\[ N^j_k \sim \mathcal{N}\left(\lambda_j l[\tilde{n}](t_j, z_k), \lambda_j \frac{B}{\hat{b}_j} l[\tilde{n}](t_j, z_k)\right) \]

- \( \lambda_j = \frac{b_j}{\hat{b}_j} \)
- \( b_j \) is the ‘true’ number of beads counted at time \( t_j \)
- \( \hat{b}_j \) is the actual number of beads counted
- \( B \) is the total number of beads originally placed into each well
- ‘Sampling without replacement’
Can be derived from counting arguments (ignoring interdependence)
- Additional parameters $b_j$ to be estimated
- Explains residual variance, ‘precursor cohort problem’
- Implications for estimation procedure, model comparison
Model Generalizations

- Examination of AutoFI distribution
  - Cell division as a fission process
  - Activation and/or time-dependence (machine calibration issues?)
  - Nonparametric estimation?
  - ... or not even estimate it at all?

- (Improved) biologically meaningful prolif/death rates
  - Smith-Martin, probabilistic mechanisms
  - Include stimulation/signaling mechanisms
Dynamics for cell division, CFSE quantity, and measured FI can be decoupled

- Allows for fast computational solution

\[
n_i(t, x) = N_i(t, x) \bar{n}_i(t, x)
\]

where

\[
\frac{dN_i}{dt} = -(\alpha_i(t) + \beta_i(t))N_i(t) + 2\alpha_{i-1}(t)N_{i-1}(t)
\]

\[
N_0(0) = N_0, N_i(0) = 0
\]

and

\[
\frac{\partial \bar{n}_i}{\partial t} - \frac{\partial [v(t, x) \bar{n}(t, x)]}{\partial x} = 0
\]

\[
\bar{n}_i(0, x) = 2^i \Phi(2^i x)/N_0
\]

- Convolution operator to link CFSE content with measured FI (hence AutoFI)
Experimental Extensions

- Account for multiple cell cultures present in PBMC culture
- Antigen-specific stimulation
- Division-linked changes, differentiated subsets
- Extracellular signaling, knockout experiments
- In vitro vs in vivo differences
- Linking to immune/pathogenesis models
- **Analyze Proliferation in Diseased vs Healthy cells**
Selected Sources

D. Schittler, J. Hasenauer, and F. Allgower, A generalized population model for cell proliferation: integrating division numbers and label dynamics, *Proc. 8th Intl. Workshop on Computational Systems Biology*, June 2011, Zurich, Switzerland.


