Hybridizing non-parametric and parametric models for biological data

Kevin Flores

NCSU RTG Parameter Estimation Workshop

7/31/19

Advantages and challenges in modeling

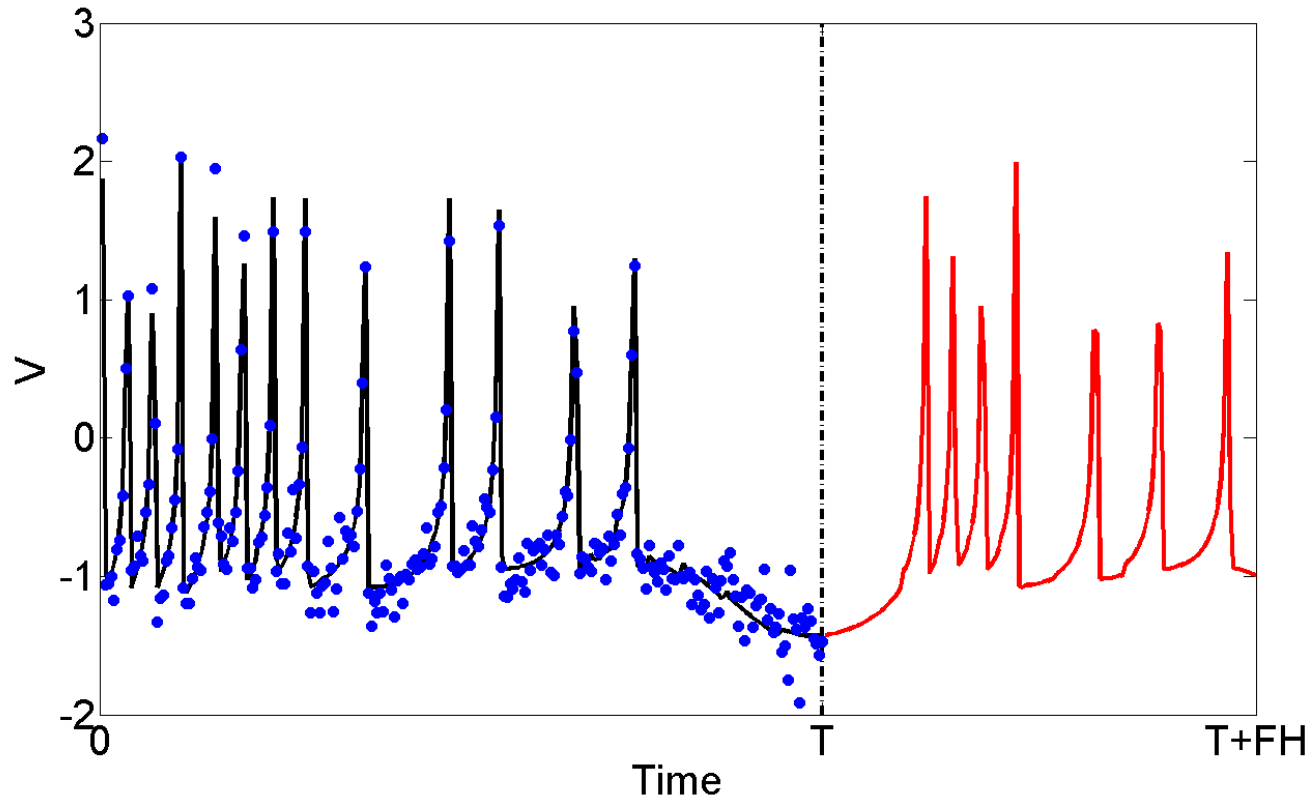
- Advantages
  - Interpretability, gain biological insight
  - Ability to inform the model with biological knowledge
- Challenges
  - The complexity needs to match the hypotheses
  - More complex models have more parameters, more uncertainty
  - How to interpret UQ vs. predictive value
  - Model refinement is a long process

Part 1:
Hybridizing model-free methods with modeling to improve forecasting

Goal: Want to **predict** the future state of a system

Assumption: Multivariate time series available (i.e. training data) up to time T; mechanistic model known

Problem: How accurately can we predict up to time $T + FH$ (forecast horizon)?



Training Data **Desired Prediction** **Noisy Data!**

Consider the mechanistic model

$$\begin{aligned}
\dot{x}_1 &= f_1(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \\
\dot{x}_2 &= f_2(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \\
&\vdots \\
\dot{x}_n &= f_n(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m)
\end{aligned}$$

- $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ is n-dimensional state variable
- $\mathbf{p} = [p_1, p_2, \ldots, p_m]$ is m-dimensional parameter

Consider a Hindmarsh-Rose neuron

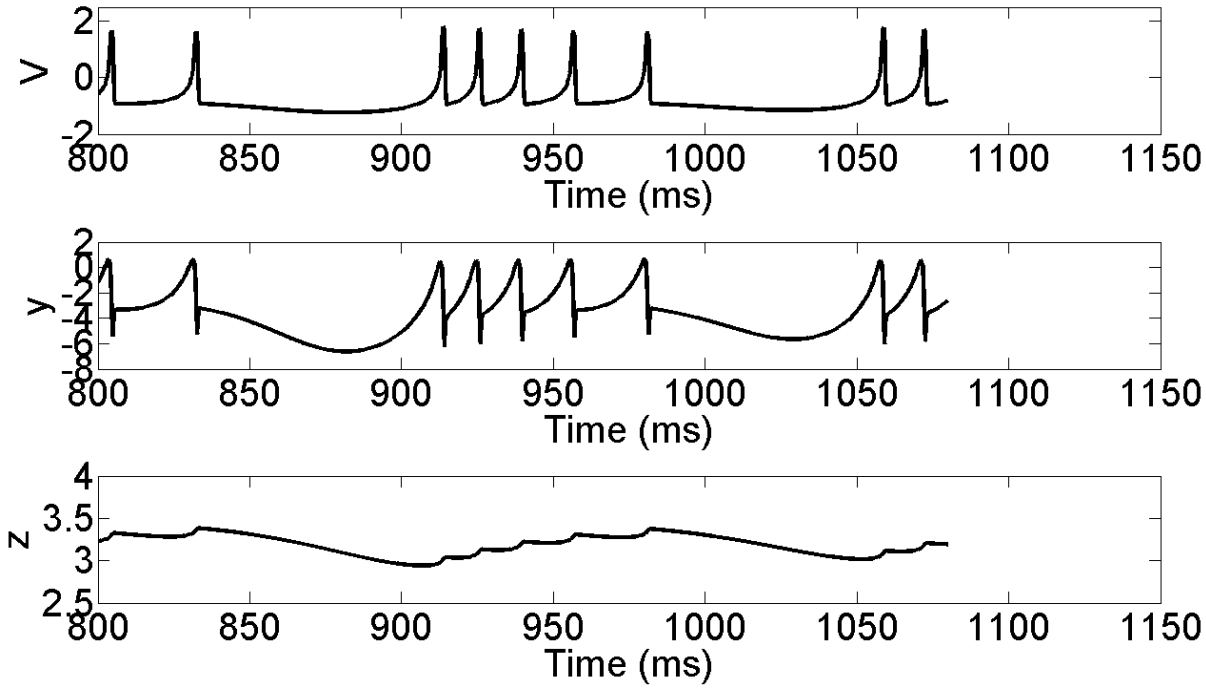
$$\begin{aligned}\dot{V} &= y - aV^3 + bV^2 - z + 3.25 \\ \dot{y} &= 1 - dV^2 - y \\ \dot{z} &= 0.005\left(s\left(V + \frac{8}{5}\right) - z\right)\end{aligned}$$

- V– voltage
- y– fast-scale dynamics
- z– slow-scale dynamics
- a, b, d, s– unknown model parameters

Assume: Noisy observations of V, y, z available

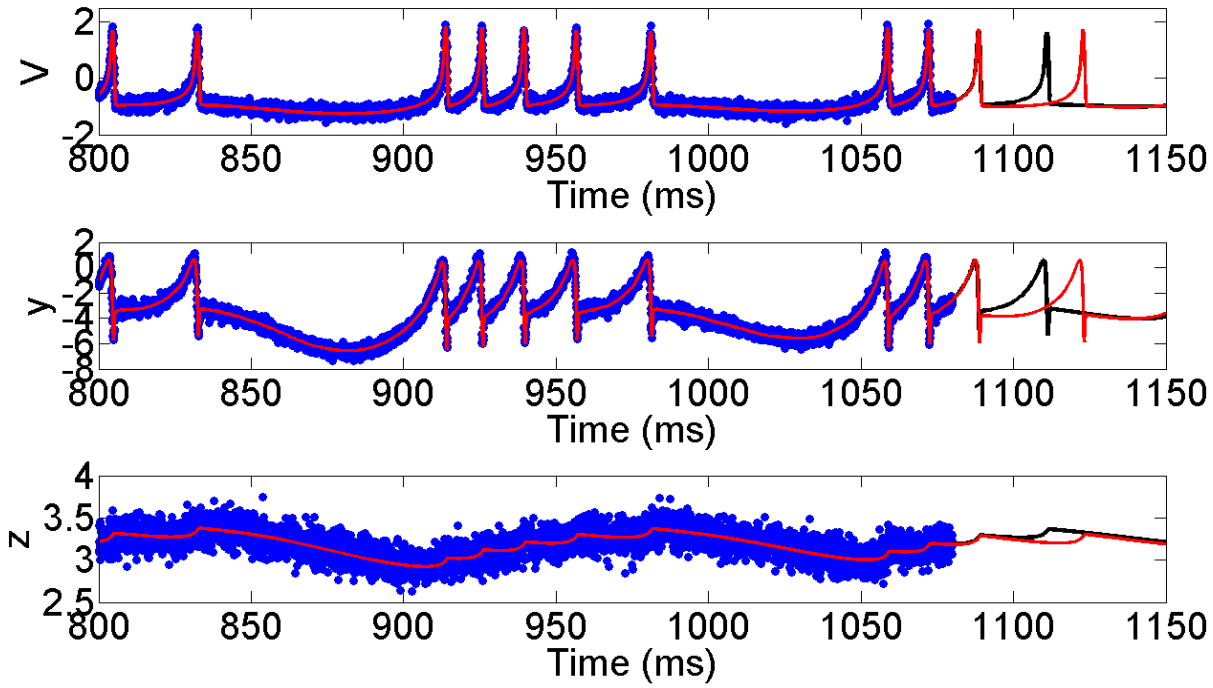
Idea: Fit observed time series to model

1. Estimate **x** and **p** until time T (Kalman filtering, shooting methods, etc...)
2. Free-run the fitted model until time $T + FH$



Idea: Fit observed time series to model

1. Estimate **x** and **p** until time T (Kalman filtering, shooting methods, etc...)
2. Free-run the fitted model until time $T + FH$



Non-parametric (model-free) prediction

- Instead of using the mechanistic model, noisy training data can be used to build model-free predictions
- Numerous approaches exist, many based on **Takens' Method**

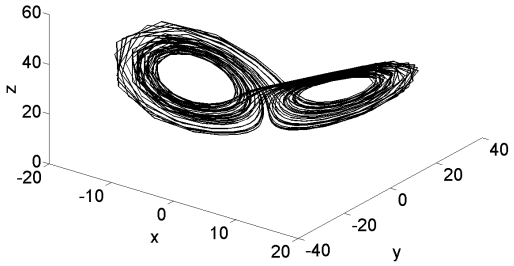
Attractor reconstruction

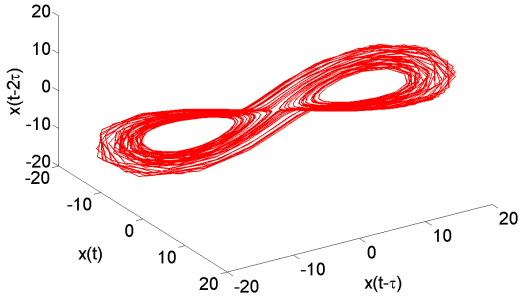
For an observed time series $x(t)$, a delay-coordinate vector

$$x_d = [x(t), x(t-\tau), x(t-2\tau), \ldots, x(t-d\tau)]$$

can be formed that reconstructs the attractor

- d: number of delays
- τ: time-delay





Nonparametric prediction

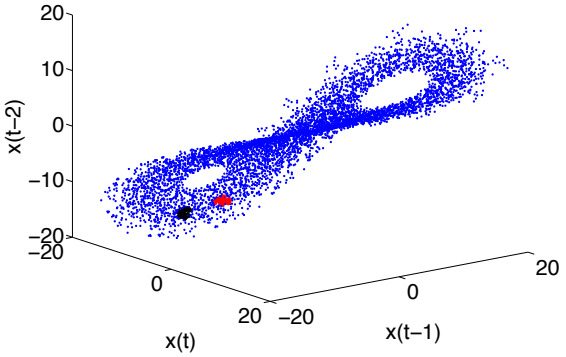
Assume: $x_d = [x(T), x(T-1), x(T-2), \ldots, x(T-d)]$

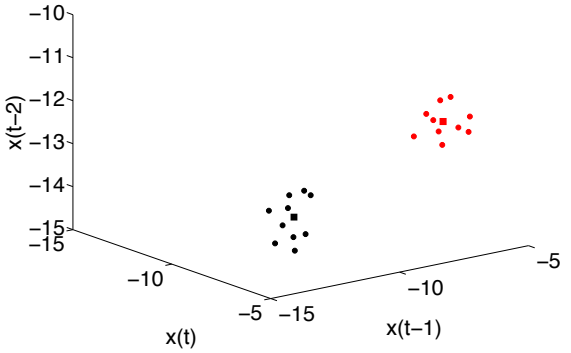
Goal: Predict $x(T + FH)$

1. Find neighboring delay vectors $[x(T'), x(T'-1), \ldots, x(T'-d)]$, $[x(T''), x(T''-1), \ldots, x(T''-d)], \ldots, [x(T^k), x(T^k-1), \ldots, x(T^k-d)]$ within training data
2. Identify the known $x(T'+FH), x(T''+FH), \ldots, x(T^k+FH)$ points
3. Build a local model to predict the unknown $x(T+FH)$, e.g.

$$x(T+FH) \approx \frac{x(T'+FH) + x(T''+FH) +, \ldots, +x(T^k+FH)}{k}$$

Training Data



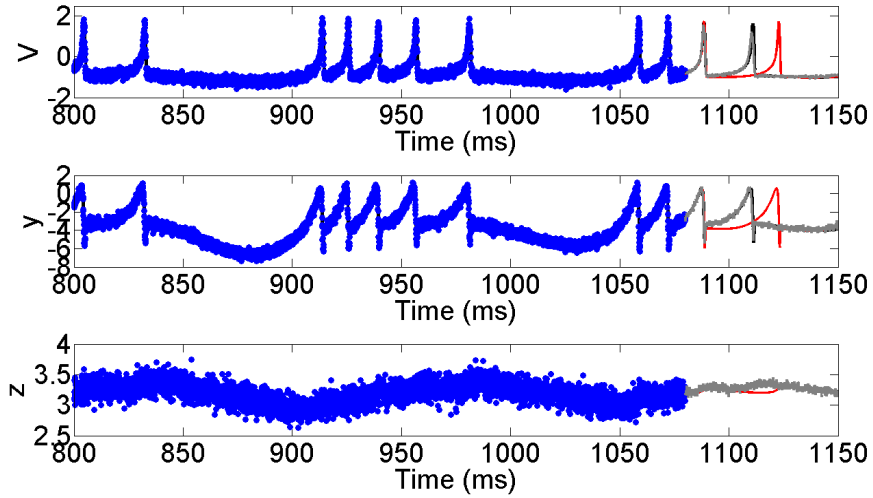


1. Form the current delay-coordinate state (■)
2. Find N nearest neighbors in delay-coordinate space (●)
3. Look-up the neighbors' j step-ahead values (●)
4. Build a local prediction of current state (■)

Nonparametric prediction

Idea: Ignore the mechanistic model– use the data to build a predictive model

1. *Denoise observed time series* (e.g. F. Hamilton, T. Berry and T. Sauer. Phys. Rev. X 6, 011021 (2016))
2. Build predictions until time $T + FH$ using only training data



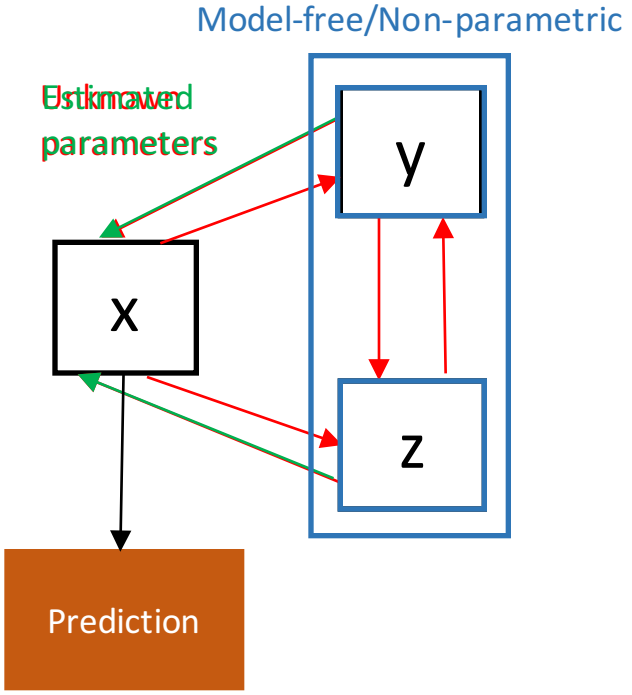
General advantages and challenges with non-parametric (data science) methods

- Advantages
  - Don't need to develop a mechanistic model
  - Neural network architectures designed for the type of data (CNN, RNN)
- Challenges
  - Data requirements may not match the scale in biology
  - Corrupted by noise
  - Lack of interpretability
  - Difficult to make conclusions about mechanistic processes
  - Lack of generalizability to data outside the training set

Hybrid modeling and prediction

Franz Hamilton, ARL

- Parameter estimation for a mathematical model could be difficult
  - Parameter correlations
  - Rugose cost function landscape
  - High parameter uncertainty
  - Computational time
- May be model misspecification for some variables



Hamilton, Lloyd, Flores, Hybrid modeling and prediction of dynamical systems, 2017, PLoS Computational Biology

Consider the mechanistic model

$$\begin{aligned}
\dot{x}_1 &= f_1(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \\
\dot{x}_2 &= f_2(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) \\
&\vdots \\
\dot{x}_n &= f_n(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m)
\end{aligned}$$

- $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ is n-dimensional state variable
- $\mathbf{p} = [p_1, p_2, \ldots, p_m]$ is m-dimensional parameter

Consider the mechanistic model

$$\dot{x}_1 \quad = \quad f_1(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m)$$

- $\mathbf{x} = x_1$ is 1-dimensional state variable
- $\mathbf{p} = [p_1, p_2, \ldots, p_{m^*}]$ is m^*-dimensional parameter

Assume: Noisy observations of V, y, z available

Goal: Refine our prediction of V (the thing we really care about)

$$\dot{V} \quad = \quad y - aV^3 + bV^2 - z + 3.25$$

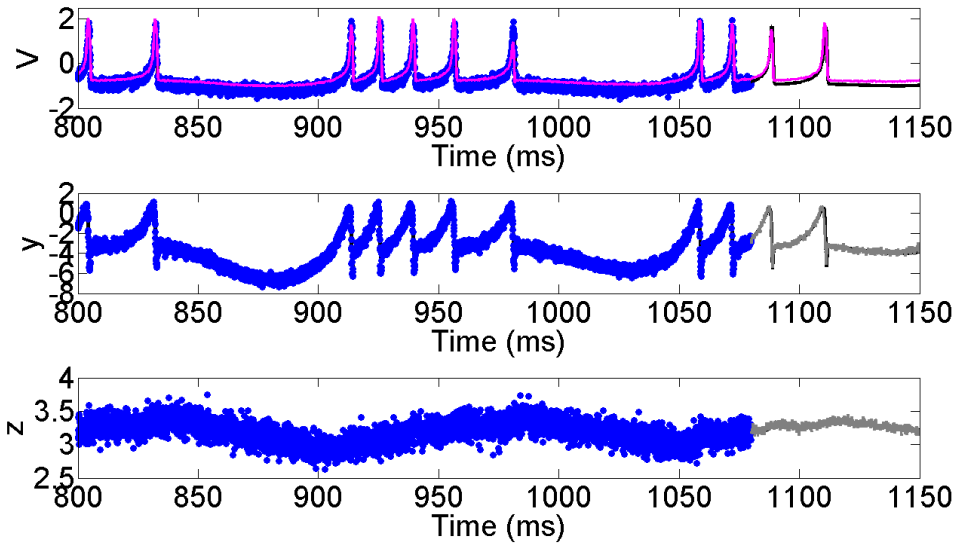
Idea: Fit the V equation; use the observations of y and z

Only need to estimate a and b parameters!

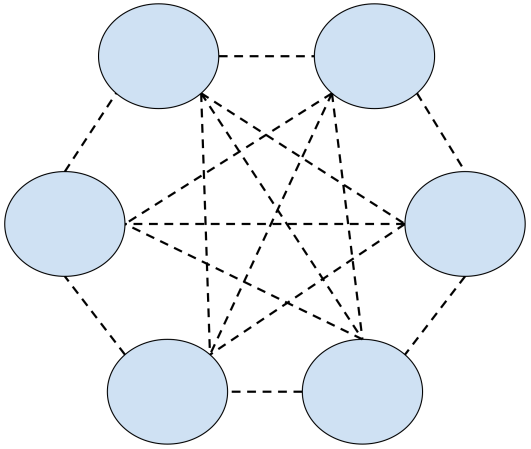
Hybrid prediction

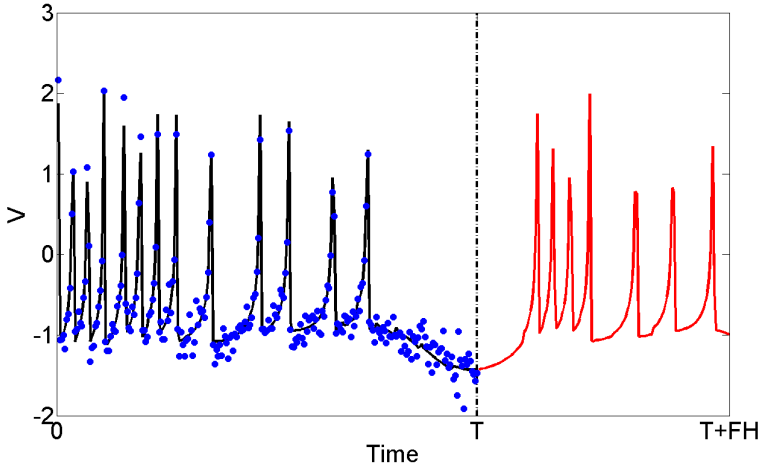
Idea: Use a model for the variable you want to predict and nonparametrically forecast the rest

1. Fit states and parameters of variable of interest
2. Nonparametrically forecast remaining variables
3. Integrate nonparametric predictions with the modeled variable



Example: predicting neuronal networks





Hindmarsh-Rose model (random connections)

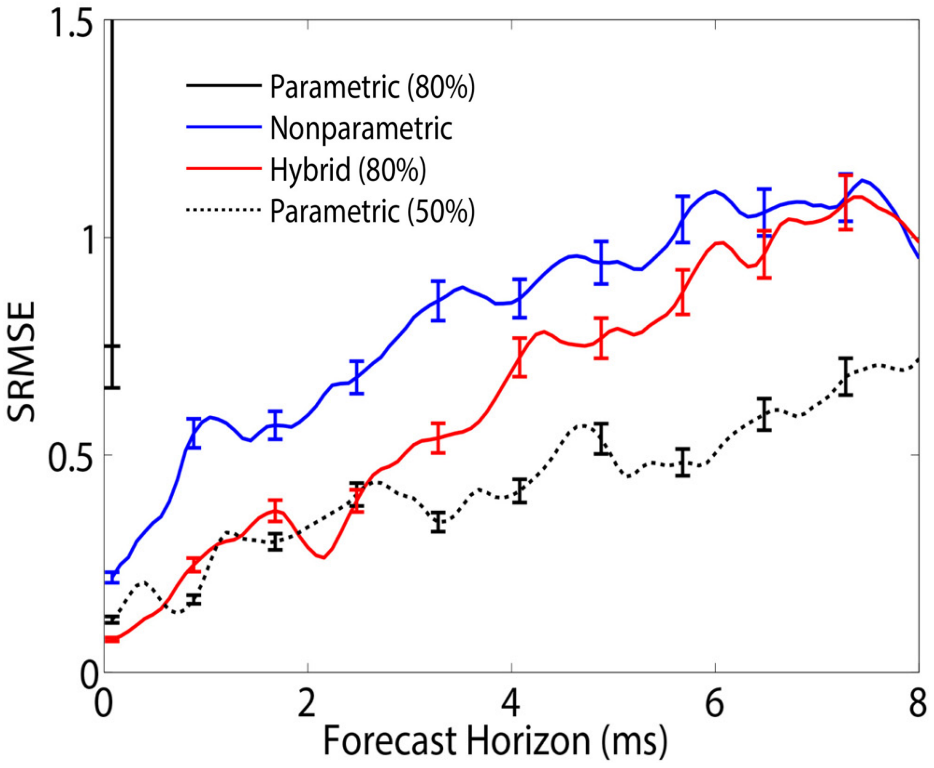
$$\dot{x}_i = y_i - a_i x_i^3 + b_i x_i^2 - z_i + 1.2 + \sum_{i \neq m}^{M} \frac{\beta_{im}}{1 + 9e^{-10x_m}} x_m$$

$$\dot{y}_i = 1 - c_i x_i^2$$

$$\dot{z}_i = 5 \times 10^{-5} \left[4 \left(x_i - \left(-\frac{8}{5} \right) \right) - z_i \right]$$

Example: predicting neuronal networks

- Sampled 200 random networks, 3 neurons, 5 connections.
- Train on 240ms, predict 8ms.
- Hybrid model enabled higher accuracy (SRMSE) for a given initial parameter uncertainty.
- Hybrid modeling resulted in more accurate parameter estimates and lower parameter uncertainty.



Model for one neuron

$$\dot{x}_3 = y_3 - a_3 x_3^3 + b_3 x_3^2 - z_3 + 1.2 + \sum_{3 \neq m}^{M} \frac{\beta_{3m}}{1 + 9e^{-10x_m}} x_m$$

$$\dot{y}_3 = 1 - c_3 x_3^2$$

$$\dot{z}_3 = 5 \times 10^{-5} \left[4 \left(x_3 - \left(-\frac{8}{5} \right) \right) - z_3 \right]$$

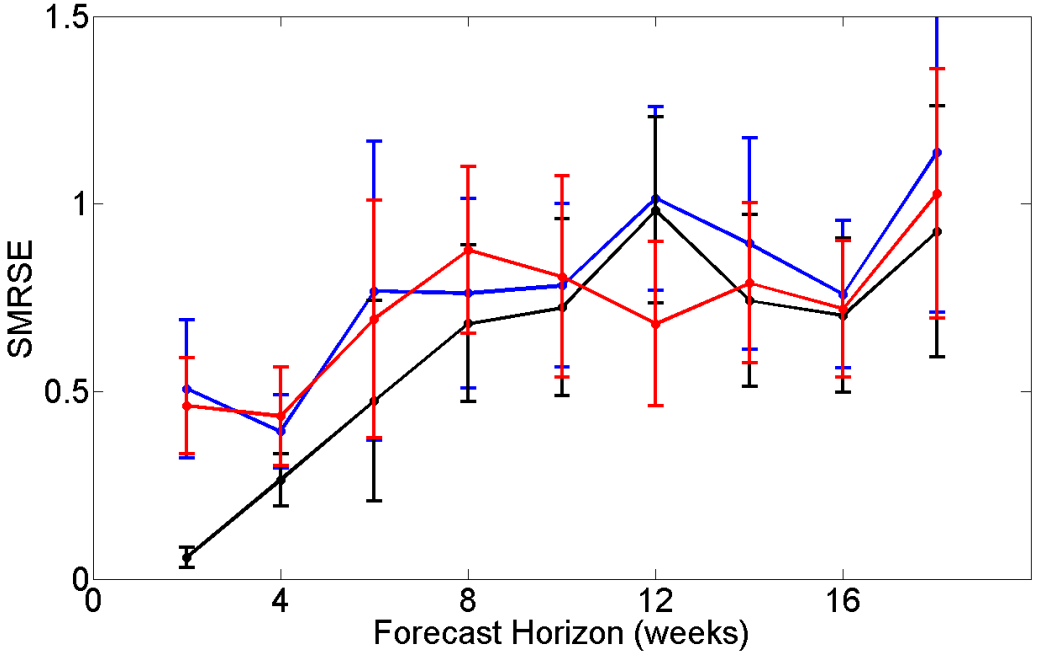
Predicting beetle population dynamics

Experimental data (Constantino et. al. 1997, Science)

- Experimentally altered adult mortality (c_{pa}): 7 values
- 3 replicates per c_{pa}, total of 21 data sets.
- 41 time points per data set, sampled every 2 weeks.
- Total counts of Larvae, Pupae, and Adults.

$$
\begin{aligned}
L(t+1) &= bA(t)e^{-c_{el}L(t)-c_{ea}A(t)} \\
P(t+1) &= L(t)(1-\mu_1) \\
A(t+1) &= P(t)e^{-c_{pa}A(t)} + A(t)(1-\mu_a)
\end{aligned}
$$

Predict Adult population

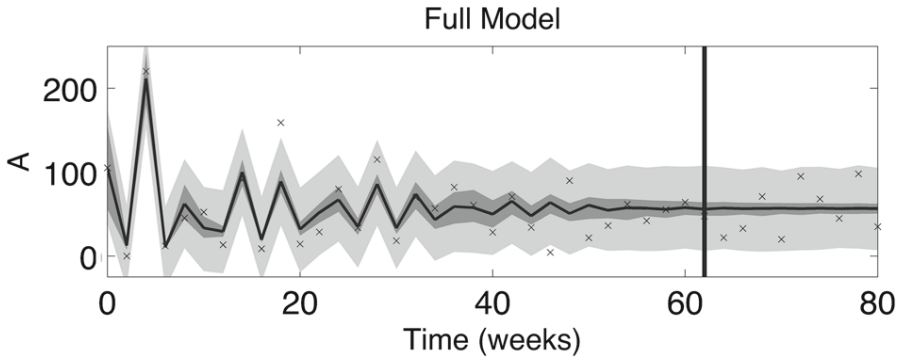


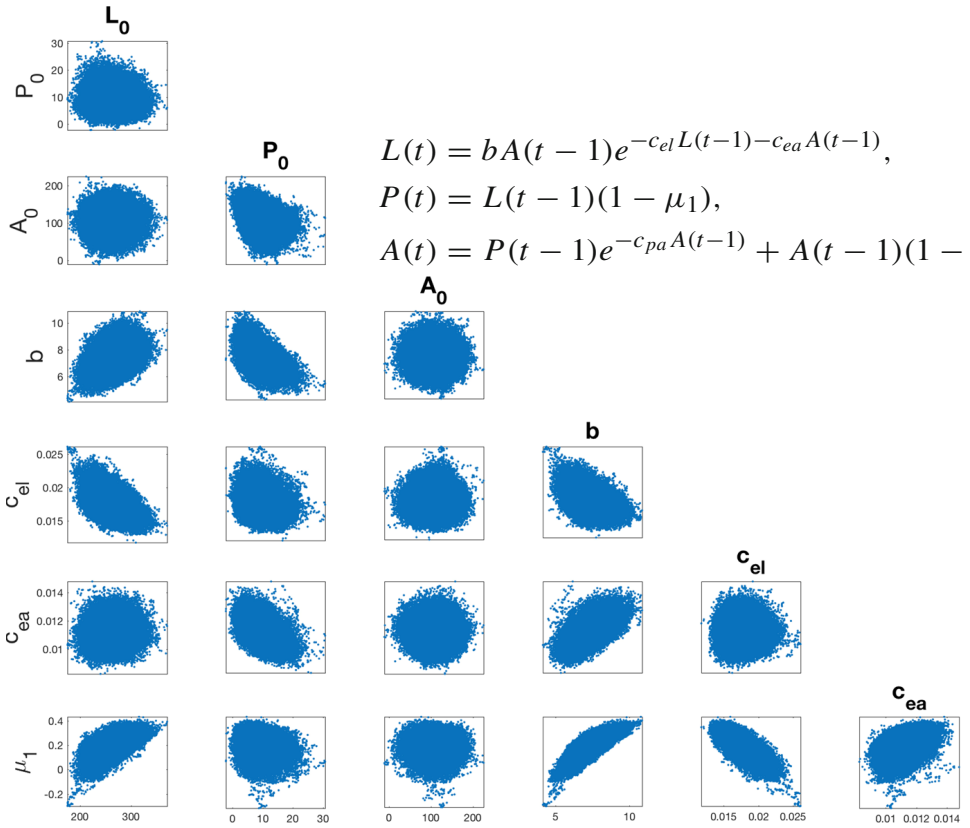
Parametric **Nonparametric** **Hybrid**

Hybrid modeling with Bayesian estimation

John Lagergren,NCSU

- Studied how hybrid modeling affects parameter estimation.
- Hybridizing with Bayesian MCMC yields parameter correlation plots from the sampled posterior distribution.
- Used a stage-structured model with a feedback (7 parameters).
- Parameter estimation for the full model resulted in many parameter correlations, high parameter and prediction uncertainty.



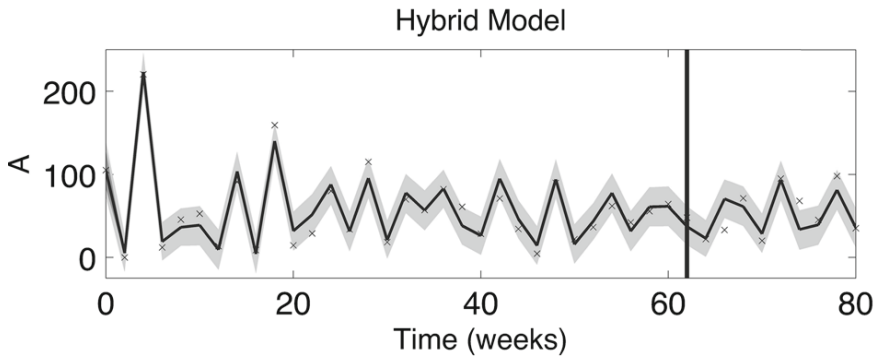


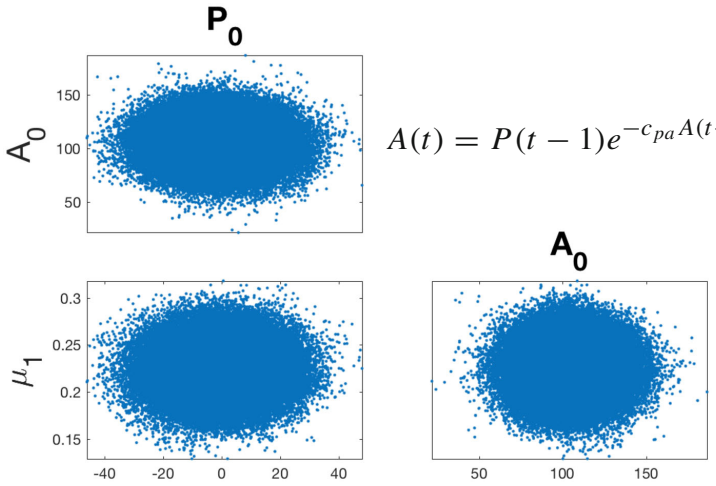
$$L(t) = bA(t-1)e^{-c_{el}L(t-1)-c_{ea}A(t-1)},$$
$$P(t) = L(t-1)(1-\mu_1),$$
$$A(t) = P(t-1)e^{-c_{pa}A(t-1)} + A(t-1)(1-\mu_a)$$

Lagergren, Reeder, Hamilton, Smith, Flores, Bull. Math. Bio., 2018

Hybrid modeling with Bayesian estimation

- Hybrid modeling can reduce parameter correlations.
- Leads to increased parameter and prediction confidence.





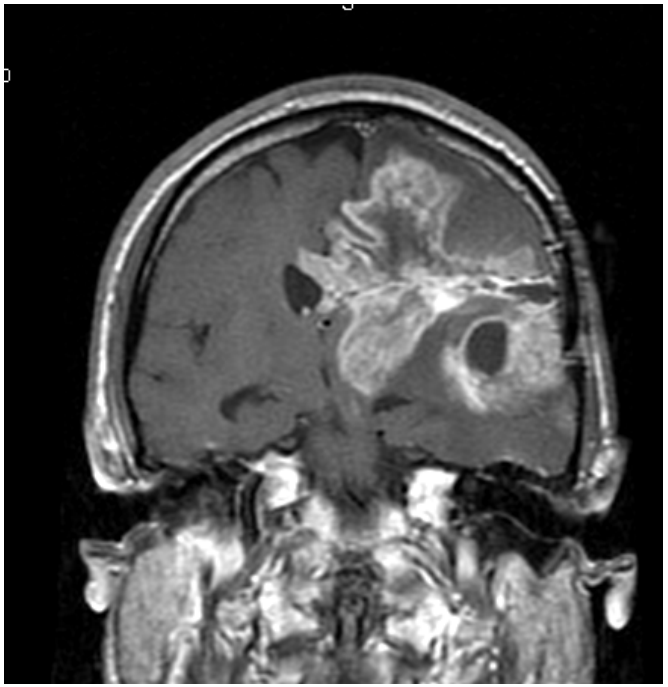
$$A(t) = P(t-1)e^{-c_{pa}A(t-1)} + A(t-1)(1-\mu_a)$$

Potential issues solved by hybrid modeling

- Reduces the number of parameters that need to be estimated at the same time.
- Speeds up computational time for optimization.
- May reduce parameter correlations, increase confidence.
- Unlike machine learning methods, can still retain an interpretable model and biologically meaningful parameters.

Part 2:
Hybridizing non-parametric probability distribution estimation with PDEs to infer population heterogeneity

Glioblastoma Multiforme (GBM)



Sagittal cross-section of human brain with GBM

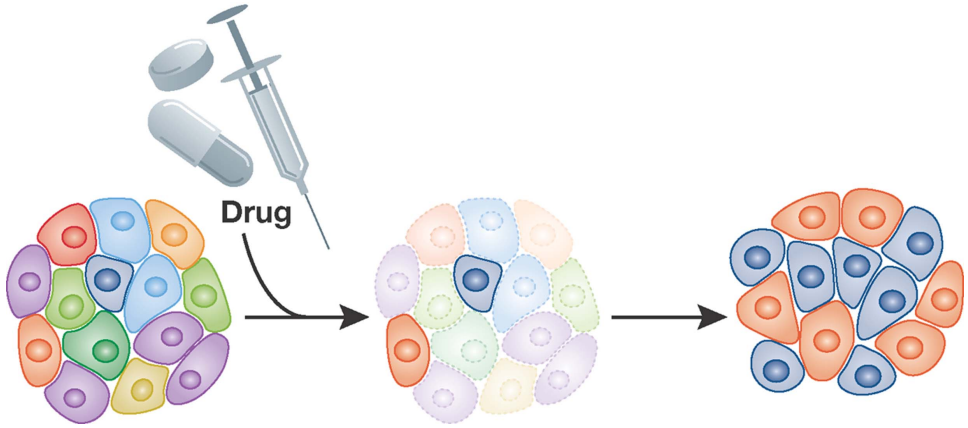
GBM is a deadly primary brain tumor characterized by:

- Phenotypic heterogeneity
- Low survivability
- Low response to treatment

Cancers often modeled by partial differential equations, because they can incorporate

- Spatial structures, diffusion, taxis
- $\frac{\partial u}{\partial t} = \underbrace{D\frac{\partial^2 u}{\partial x}}_{\text{diffusion}} + \underbrace{\rho u(1-u)}_{\text{logistic growth}}$

Importance of Heterogeneity



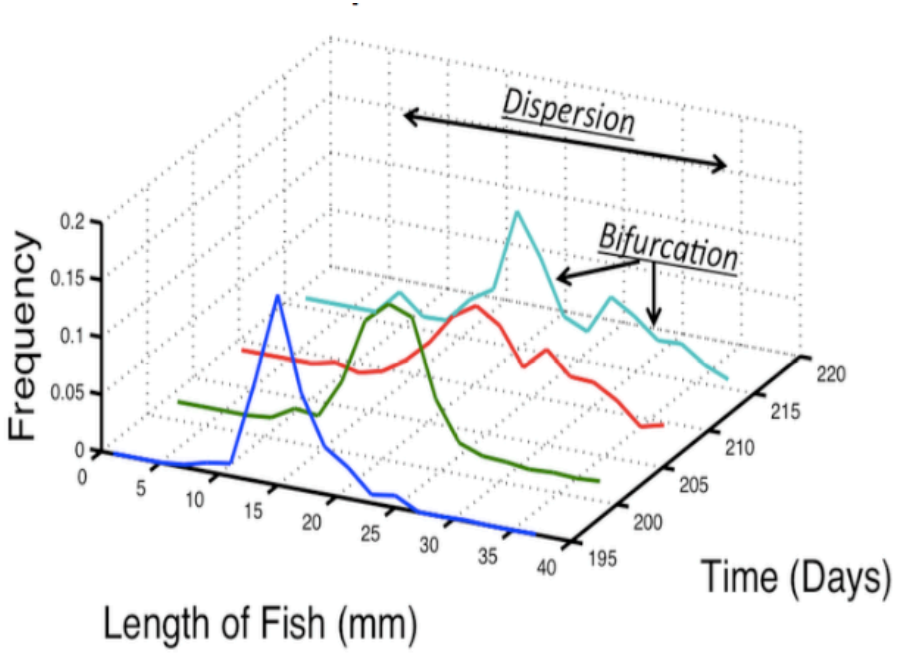
Source: Saunders, Nicholas A., et al. "Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives." *EMBO molecular medicine* 4.8 (2012): 675-684.

Incorporating Heterogeneity in Cancer Models

- We can separate the tumor cell population into subpopulations:
  - `Go or grow' in glioma growth
  - Androgen-dependent and androgen-independent cells in prostate cancer
  - Radio-sensitive and radio-resistant cells for treatment strategies
  - Oxidative-Phosphorylated cells and glycolitic cells

- **Question:** How can we estimate heterogeneity, e.g., in model parameters, without making assumptions about which subpopulations exist?

Random Differential Equations



Source: Banks, H. T. & Davis, J. L. A comparison of approximation methods for the estimation of probability distributions on parameters. *Appl. Numer. Math.* **57,** 753–777, (2007).

Dispersion and bifurcation features are not accounted for when using static parameters

Random Differential Equations

Consider the diffusion ($\boldsymbol{D}$) and growth ($\boldsymbol{\rho}$) as random variables defined on a compact set $\Omega = \Omega_{\boldsymbol{D}} \times \Omega_{\boldsymbol{\rho}}$

Model

$$\frac{\partial u(t,x,\boldsymbol{D},\boldsymbol{\rho})}{dt} = \nabla \cdot \left(\boldsymbol{D}\nabla u(t,x,\boldsymbol{D},\boldsymbol{\rho})\right) + \boldsymbol{\rho} u(t,x,\boldsymbol{D},\boldsymbol{\rho})(1 - u(t,x,\boldsymbol{D},\boldsymbol{\rho}))$$

Observation

$$u(t,x) = \mathbb{E}[u(t,x,\cdot,\cdot),P] = \int_{\Omega} u(t,x,\boldsymbol{D},\boldsymbol{\rho}) dP(\boldsymbol{D},\boldsymbol{\rho})$$

Rutter, Banks and Flores. Estimating Intratumoral Heterogeneity from Spatiotemporal Data, *JMB 2018.*

Prohorov Metric Framework (PMF)

Idea: Using data, determine the approximate distributions of $\boldsymbol{D}$ and $\boldsymbol{\rho}$, *without any underlying assumptions about the pdf/cdf*

$$\hat{P} = \underset{P^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left(\text{data}(t_j, x_i) - \int_{\Omega} u(t_j, x_i, \boldsymbol{D}, \boldsymbol{\rho}) dP(\boldsymbol{D}, \boldsymbol{\rho}) \right)^2$$

Prohorov Metric Framework Theory

Theorem
There exists a (not necessarily unique) minimizer $\hat{P}$ (Banks, Hu, Thompson, 2015)

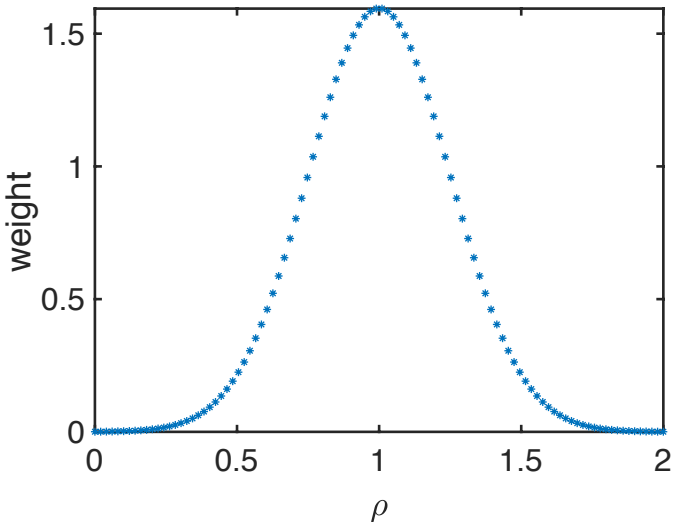
$$\hat{P} = \underset{P \in P^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left(\text{data}(t_j, x_i) - \int_{\Omega} u(t_j, x_i, \boldsymbol{D}, \boldsymbol{\rho}) dP(\boldsymbol{D}, \boldsymbol{\rho}) \right)^2$$

1. Since $\Omega = \Omega_{\boldsymbol{D}} \times \Omega_{\boldsymbol{\rho}}$ is a compact set, $P(\Omega)$ is a compact metric space
2. The minimizer is continuous in P

$\Longrightarrow$ There exists a (not necessarily unique) minimizer

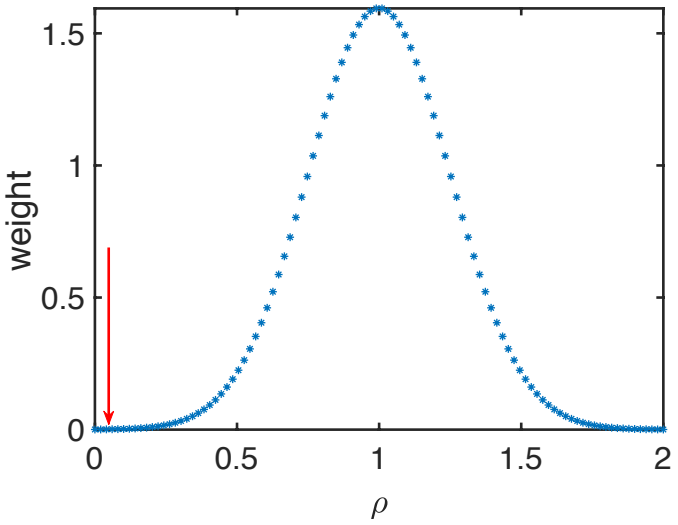
Creating Synthetic Data

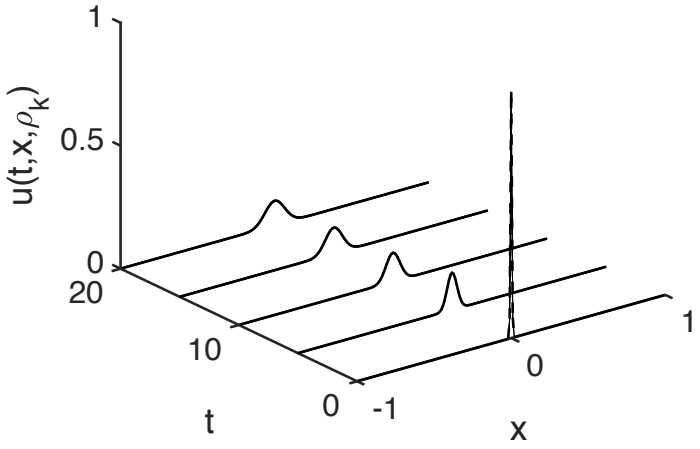
We finely mesh over the parameter $\boldsymbol{\rho} \in [0,2]$ and create our desired pdf



Creating Synthetic Data

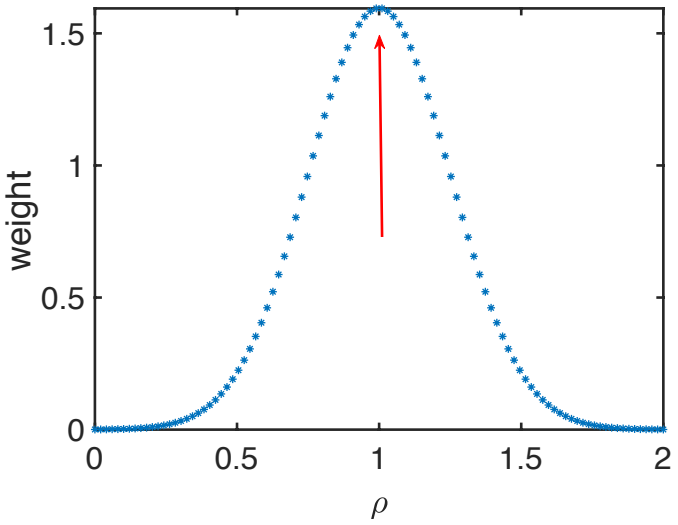
$$\frac{\partial u(t,x,\rho_k)}{dt} = \nabla \cdot \left(D\nabla u(t,x,\rho_k)\right) + \rho_k u(t,x,\rho_k)\left(1 - u(t,x,\rho_k)\right)$$

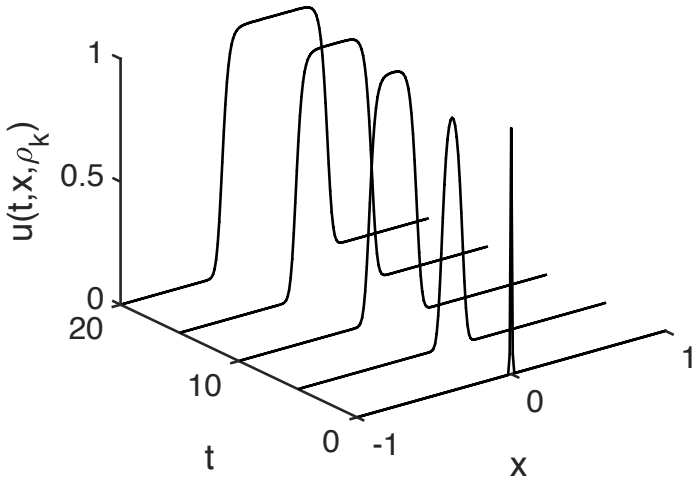




Creating Synthetic Data

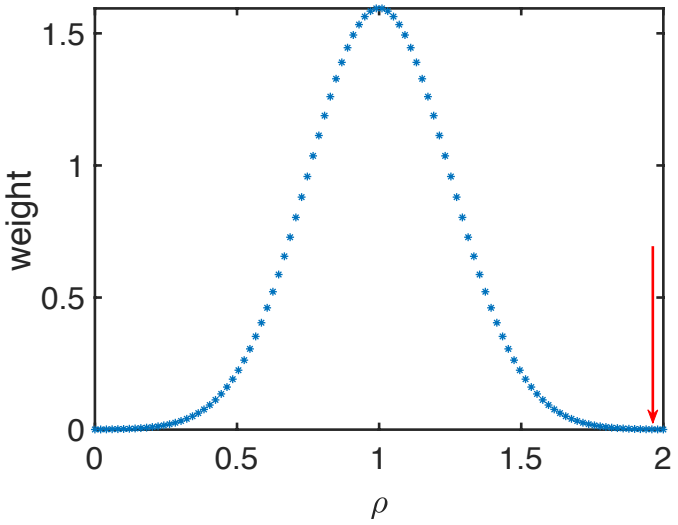
$$\frac{\partial u(t,x,\rho_k)}{dt} = \nabla \cdot \left(D\nabla u(t,x,\rho_k)\right) + \rho_k u(t,x,\rho_k)\left(1 - u(t,x,\rho_k)\right)$$

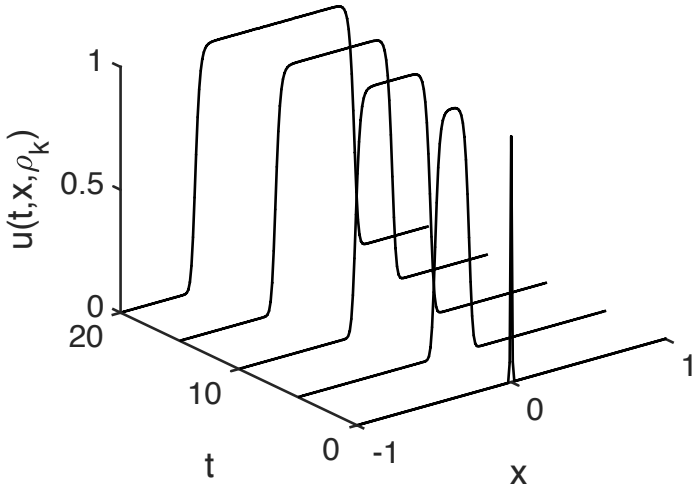




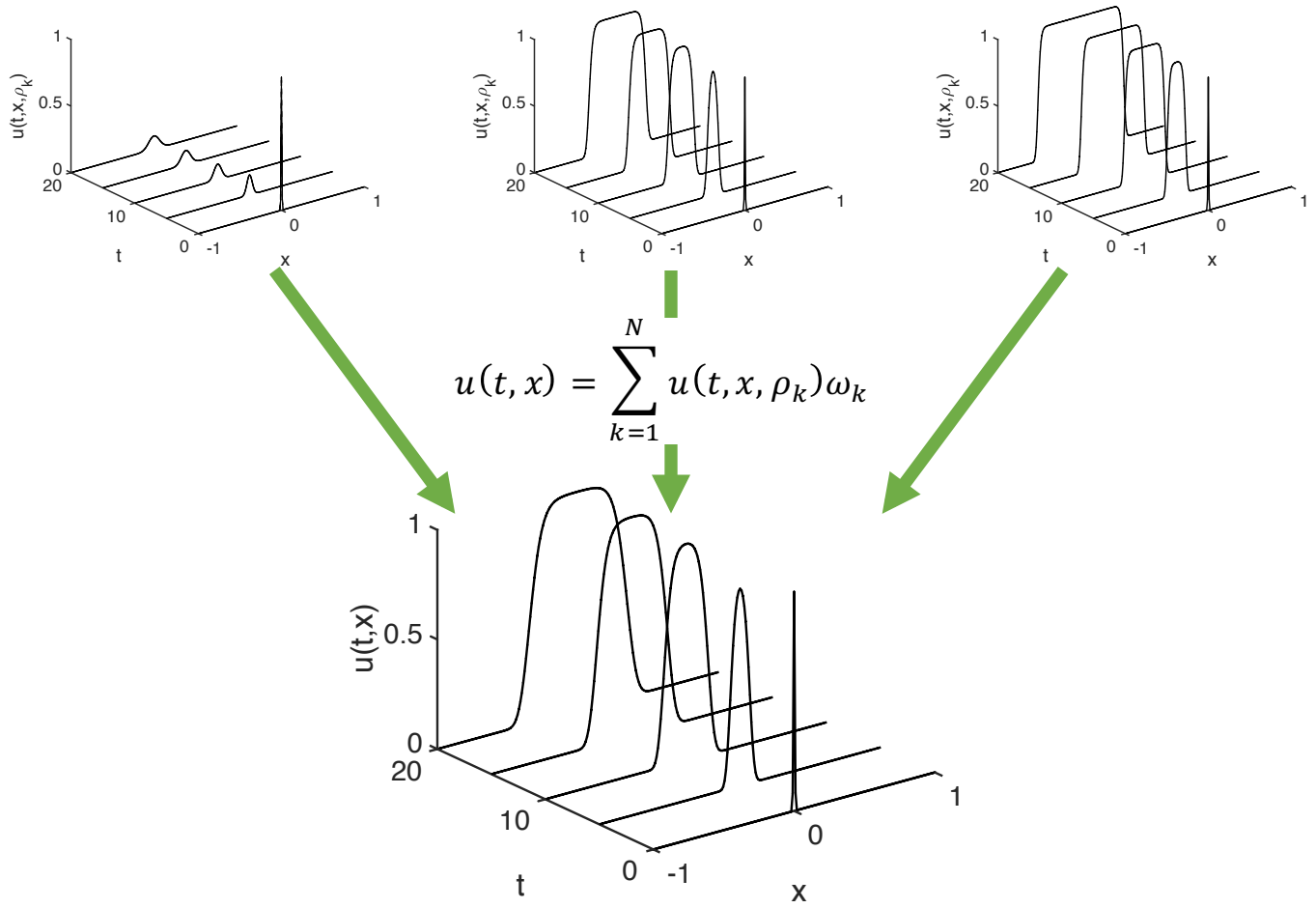
Creating Synthetic Data

$$\frac{\partial u(t,x,\rho_k)}{dt} = \nabla \cdot \left(D \nabla u(t,x,\rho_k)\right) + \rho_k u(t,x,\rho_k)\left(1 - u(t,x,\rho_k)\right)$$



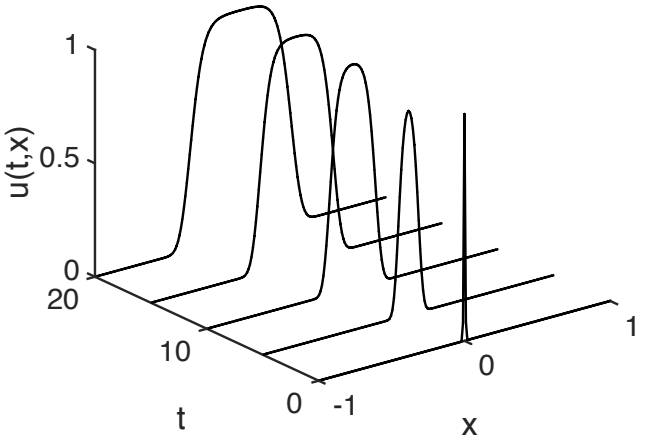


Creating Synthetic Data

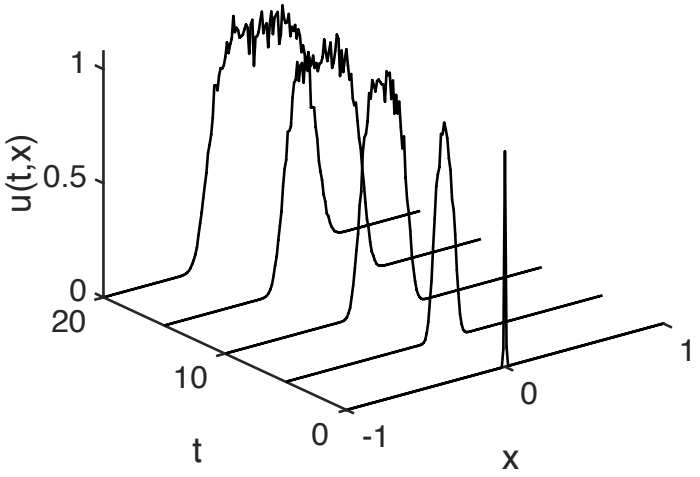


Creating Synthetic Data

$$\text{data}(t_j, x_i) = \text{sim}(t_j, x_i) + \varepsilon \text{sim}(t_j, x_i)$$
$$\varepsilon \sim 0.05N(0,1)$$



sim



data

Performing the Inverse Problem: Delta Functions

Assume there are M nodes equispaced over $\Omega_{\boldsymbol{\rho}}$ such that $\boldsymbol{\rho}^M = \{\Delta_{\rho_k}, k = 1, \ldots, M\}$

$$\hat{P} = \underset{P^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left[\operatorname{data}(t_j, x_i) - \left(\sum_{k=1}^{M} u(t_j, x_i, D, \rho_k) \omega_k \right) \right]^2$$

where $\omega_k \geq 0$ represent a discrete probability density function. Thus, we require

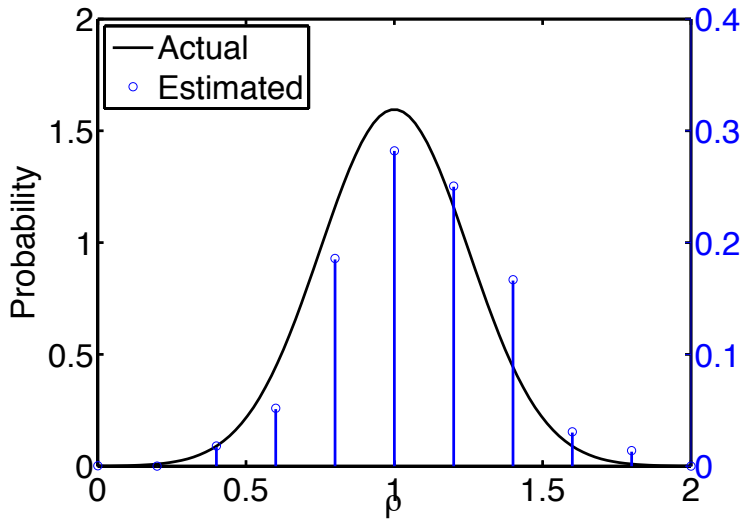
$$\sum_{k=1}^{M} \omega_k = 1$$

Performing the Inverse Problem: Delta Functions

$$\hat{P} = \underset{P^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left[\text{data}(t_j, x_i) - \left(\sum_{k=1}^{M} u(t_j, x_i, D, \rho_k)\omega_k \right) \right]^2$$

Example: we have M=11 nodes, equispaced over $[0,2]$ and we precompute $u(t, x, D, \rho_k)$

We are solving for the ω_k, the discrete weights



Performing the Inverse Problem: Spline Functions

Assume M nodes equispaced over $\Omega_{\boldsymbol{\rho}}$ such that $\boldsymbol{\rho}^M = \{s_k(\boldsymbol{\rho}), k = 1, \ldots, M\}$, where s_k are hat functions

$$\hat{P} = \underset{P^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left[\text{data}(t_j, x_i) - \left(\sum_{k=1}^{M} a_k \int_{\Omega_{\boldsymbol{\rho}}} u(t_j, x_i, D, \boldsymbol{\rho}) s_k(\boldsymbol{\rho}) d\boldsymbol{\rho} \right) \right]^2$$

where $p_k = a_k s_k(\boldsymbol{\rho}) \geq 0$ represent a probability density function. Thus, we require

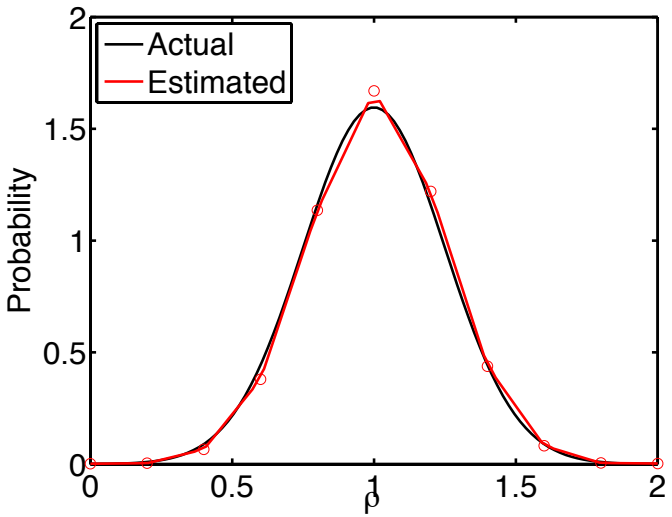
$$\sum_{k=1}^{M} a_k \int_{\Omega_{\boldsymbol{\rho}}} s_k(\boldsymbol{\rho}) d\boldsymbol{\rho} = 1$$

Performing the Inverse Problem: Spline Functions

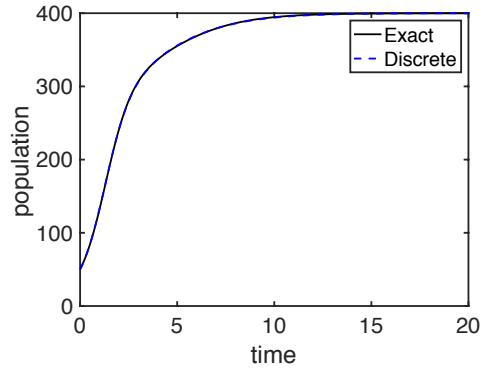
$$\hat{P} = \underset{P^M(\Omega)}{\operatorname{argmin}} \sum_{i,j} \left[\text{data}(t_j, x_i) - \left(\sum_{k=1}^{M} a_k \int_{\Omega_{\boldsymbol{\rho}}} u(t_j, x_i, D, \boldsymbol{\rho}) s_k(\boldsymbol{\rho}) d\boldsymbol{\rho} \right) \right]^2$$

Example: we have M=11 nodes, equispaced over [0,2]

We are solving for the a_k

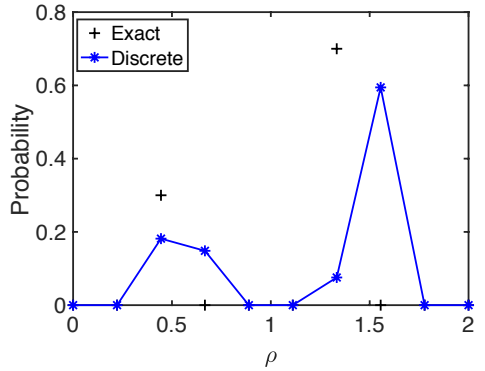


How to choose M: the optimal number of nodes?

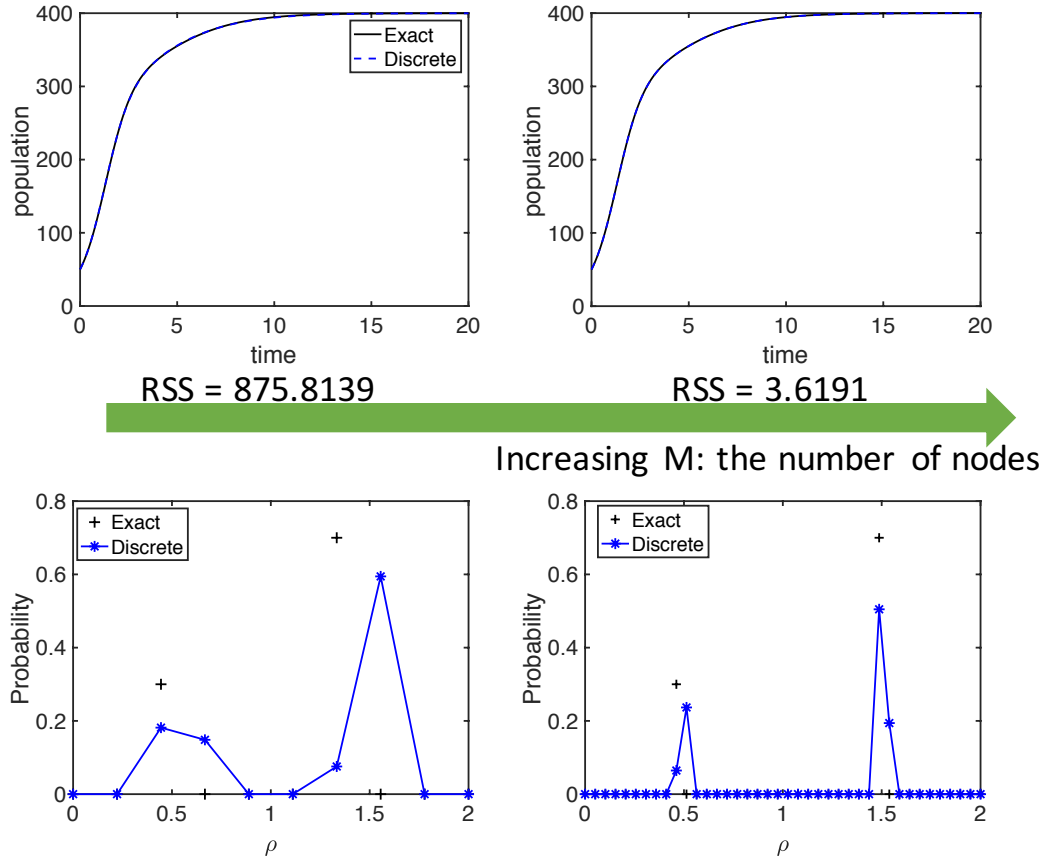


RSS = 875.8139

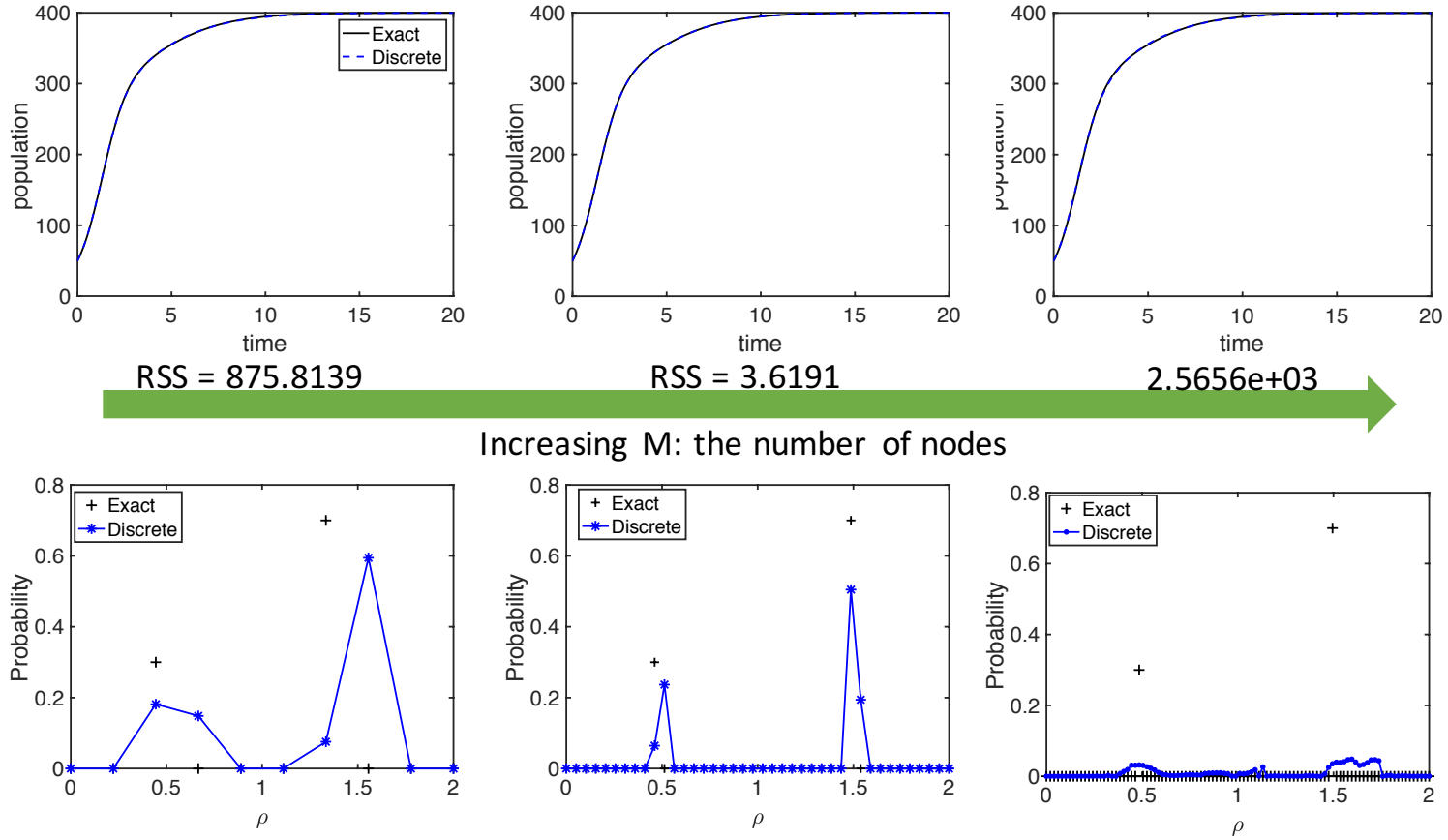
Increasing M: the number of nodes



How to choose M: the optimal number of nodes?



How to choose M: the optimal number of nodes?



How to choose M: the optimal number of nodes?

Akaike Information Criteria (AIC) as a model comparison test in the context of least-squares

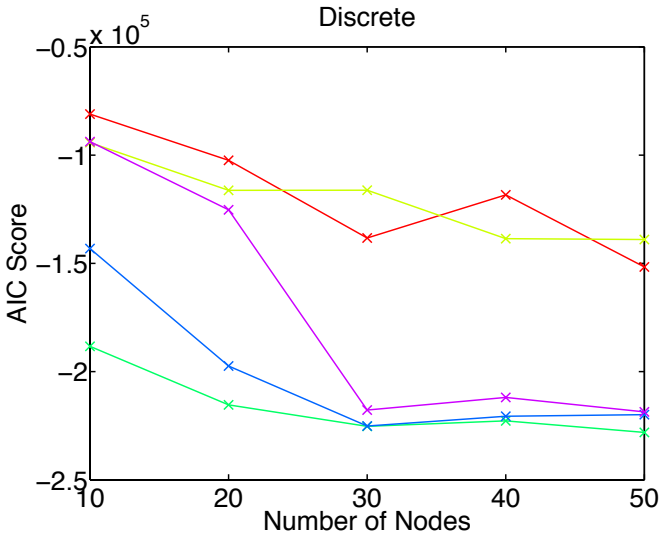
$$AIC = N \ln\left(\frac{\mathrm{RSS}}{N}\right) + N(1 - \ln(2\pi)) + 2(M + 1)$$

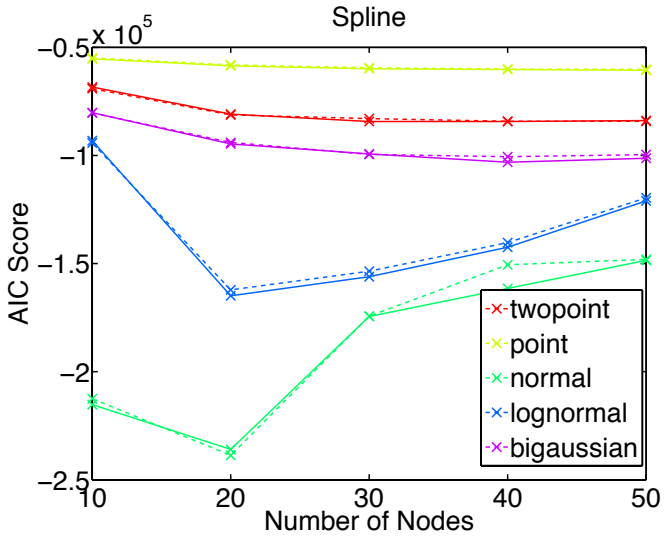
N: number of data points

RSS: error between data and solution $u(t, x)$

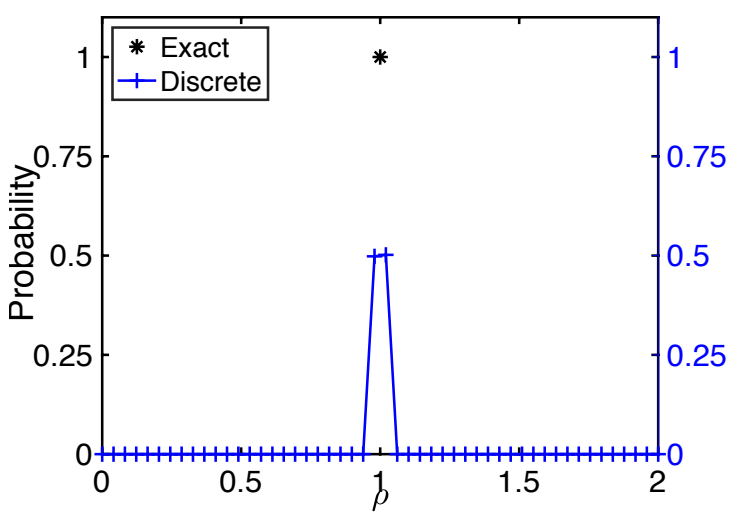
M: number of parameters being fit (our M nodes)

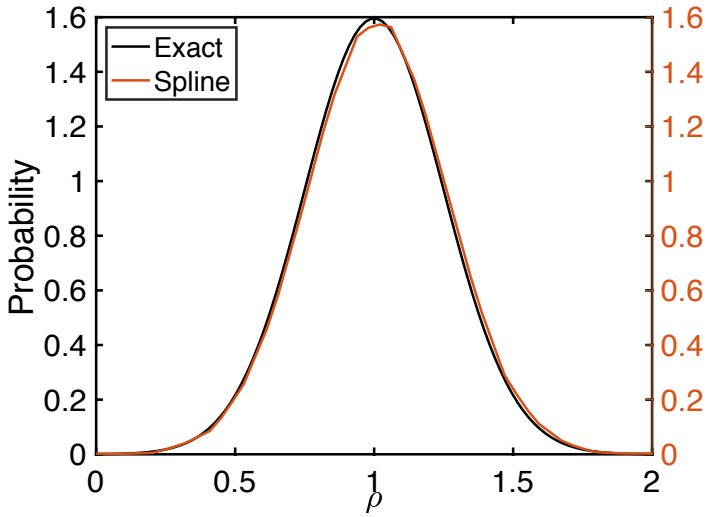
How to choose M: the optimal number of nodes?





Representative Results: finding $\boldsymbol{\rho}$





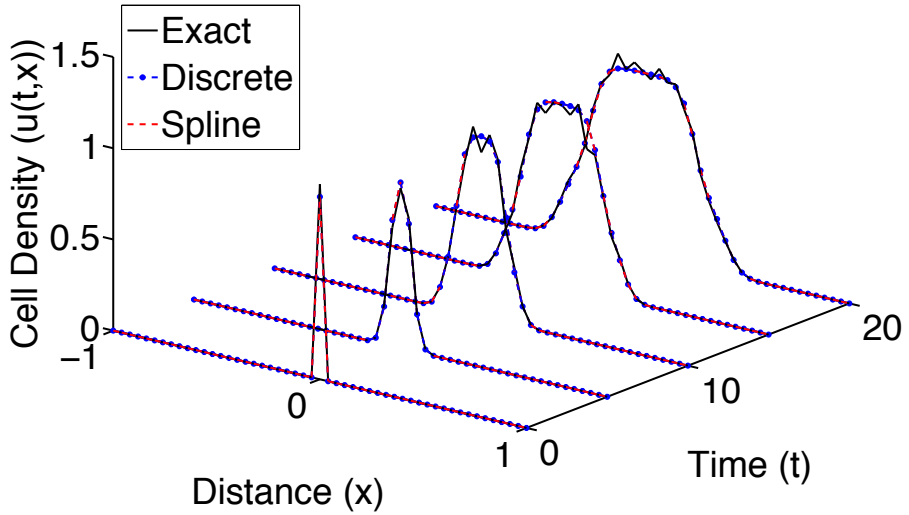
Representative Results

$\boldsymbol{\rho}$ normally distributed and $\boldsymbol{D}$ bigaussian

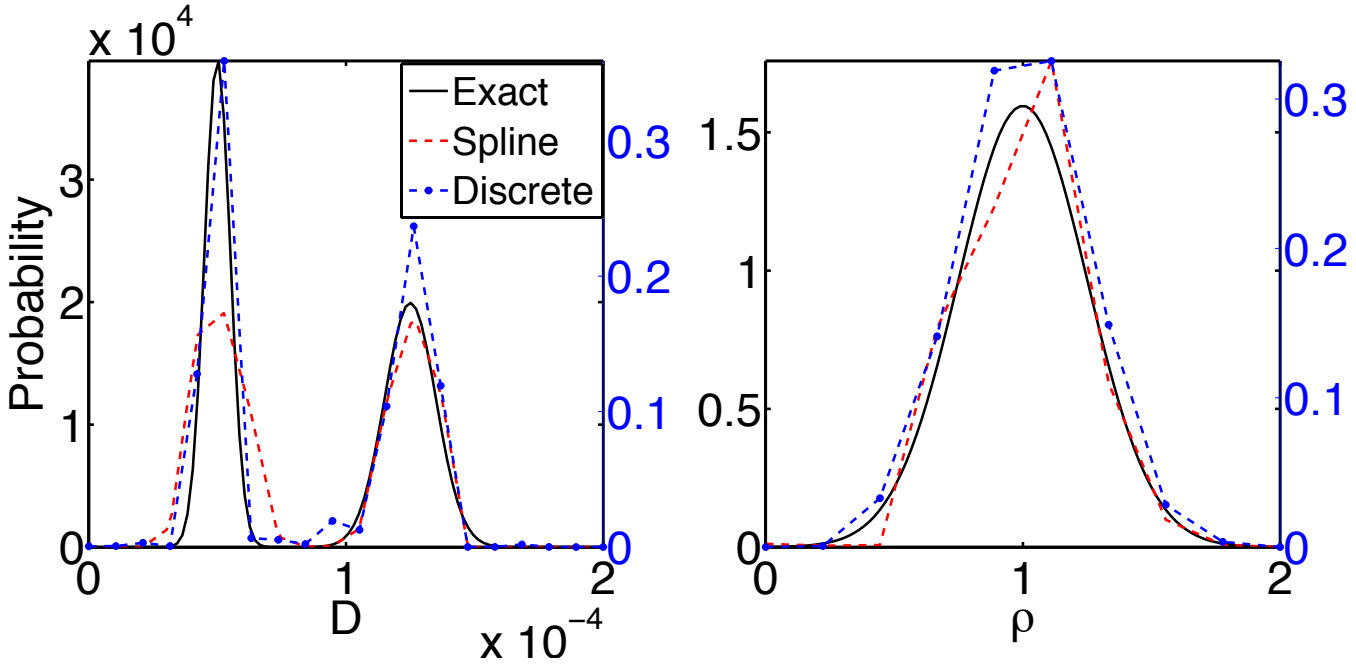
Goal: Recover parameter distributions

$$\text{data}(t_j, x_i) = \text{sim}(t_j, x_i) + \varepsilon \text{sim}(t_j, x_i)$$

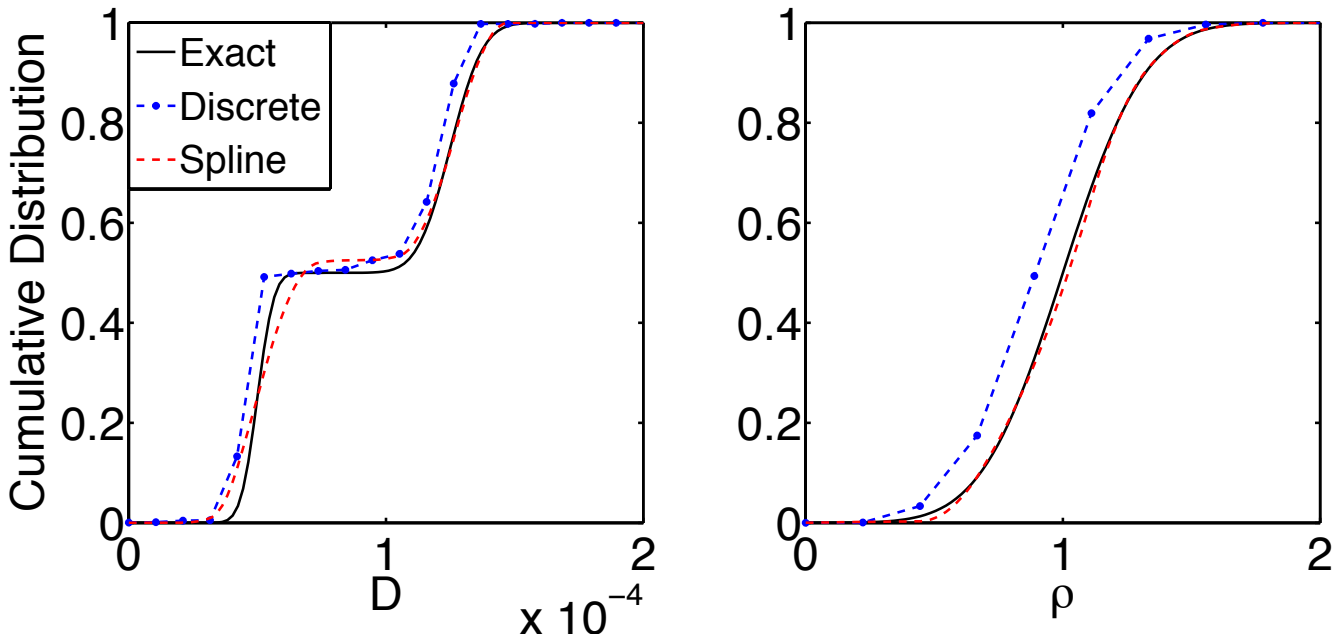
$$\varepsilon \sim 0.05N(0,1)$$



Resulting pdf Estimates



Resulting cdf Estimates

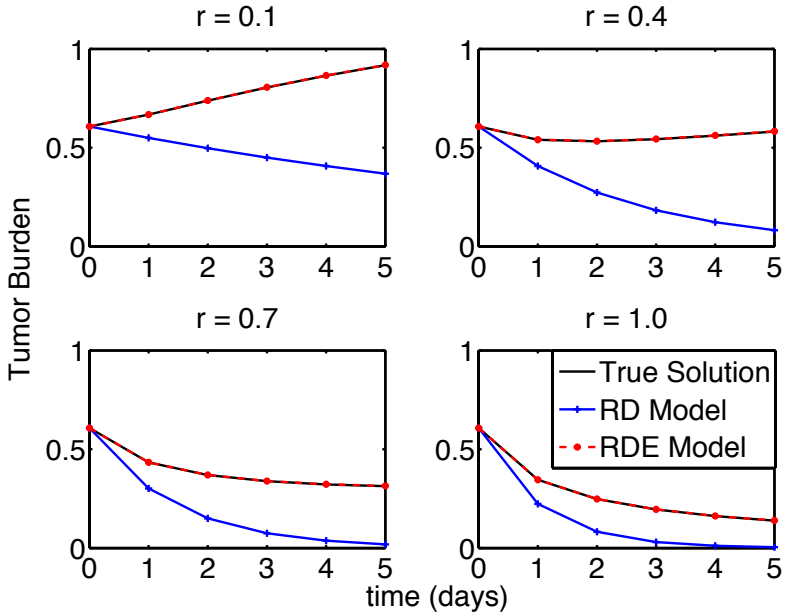


Rutter, Banks and Flores. Estimating Intratumoral Heterogeneity from Spatiotemporal Data. *Under review*

Treatment Prediction Assuming Heterogeneity

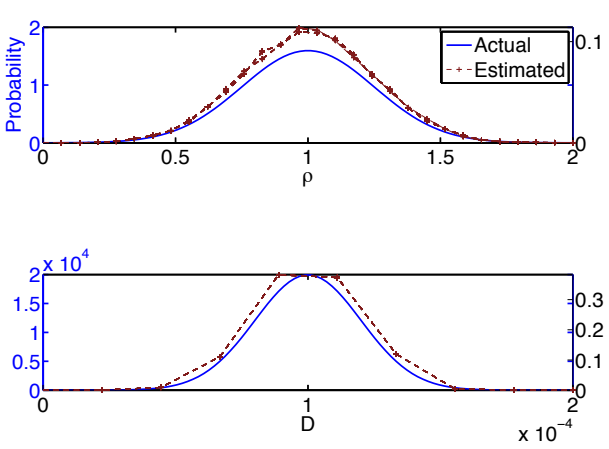
Assuming a log-kill hypothesis, we add the term:

$$-r\frac{\rho}{\bar{\rho}}u(t,x)$$

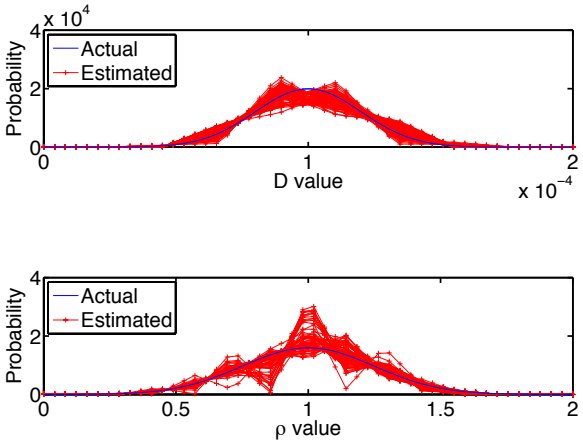


How Dependent are Results on Initial Conditions?

- Chose optimal number of nodes assuming uniform initial conditions
- Repeat 100 parameter estimations with random initializations



Discrete Approximations



Spline Approximations

Conclusions

- We can recover parameter distributions from spatiotemporal data from a variety of pdfs
- Assuming cellular homogeneity may result in overestimating treatment efficacy
- Some distribution recoveries are sensitive to initial guesses

References

- J Lagergren, A Reeder, F Hamilton, RC Smith, **KB Flores** (2018) Forecasting and uncertainty quantification using a hybrid of mechanistic and non-mechanistic models for an age-structured population model. *Bulletin of Mathematical Biology*, 80(6):1578-1595.
- F Hamilton, A Lloyd, **KB Flores** (2017) Hybrid modeling and prediction of dynamical systems. *PLoS Computational Biology*, 13(7): e1005655.
- HT Banks, **KB Flores**, IG Rosen, EM Rutter, M Sirlanci, WC Thompson (2018) The prohorov metric framework and aggregate data inverse problems from random PDEs. *Communications in Applied Analysis,* 22(3), 415-446.
- EM Rutter, HT Banks, **KB Flores** (2018) Estimating intratumoral heterogeneity from spatiotemporal data. *Journal of Math Biology*, https://doi.org/10.1007/s00285-018-1238-6
- HT Banks, **KB Flores**, CR Langlois, TR Serio, SS Sindi (2018) Estimating the Rate of Prion Aggregate Amplification in Yeast with a Generation and Structured Population Model. *Inverse Problems in Science and Engineering*, 26(2):257-279.
- HT Banks, **KB Flores**, S Sindi (2016) On analytical and numerical approaches to division and label structured population models. *Applied Math Letters*, 60:81-88.

Questions?