Applied Data Assimilation: Diabetes phenotyping/forecasting + Hybrid machine learning approaches

Matthew Levine
PhD Student in Computing and Mathematical Sciences
Advised by Andrew Stuart
California Institute of Technology

August 22, 2019

Why study the glucose-insulin system?

- High potential impact for improving:
  - Diabetes clinical care
  - Diabetes self-management
  - Our understanding of the pathogenesis of obesity and diabetes
  - Critical care (comatose patients, not necessarily diabetic)
- Data are available
  - Glucose measurement technology is improving!!!
  - Nutrition intake is often self-recorded by patients
  - Methods for capturing self-administration of medications, like insulin
  - Exercise and sleep (Fitbit etc!)
- Models are available
  - Many ***<u>mechanistic</u>*** models have been proposed and experimentally validated by physiologists, mathematical biologists, et al. These are often non-linear systems of ODE's.
  - Artie's model is novel because it is designed to describe the system, not just a particular clinical test
  - Scientists are also working on machine learning approaches, but have had limited success so far.
- Challenging (= FUN!)
  - Dynamics are <u>non-linear</u>, <u>time-delayed</u>, and <u>poorly understood overall</u>.
  - Measurements are <u>noisy</u>, <u>missing not a random</u>, <u>limited to a subset of observable states</u>, <u>costly</u>, and <u>invasive</u>.

Ongoing projects

1. Characterizing endocrine function (i.e. Bayesian inversion of biological parameters) in patients/mice with:
   - Type 2 Diabetes (T2D) [**free living fingersticks—patient-collected data]**
   - Polycistic Ovarian Syndrome (PCOS) [**OGTT—clinically-collected data]**
   - <u>Cystic Fibrosis-related Diabetes (CFRD) </u>[**free living fingersticks, CGM, OGTT]**
   - **Because data are noisy and partially observed, we need to carefully quantify UNCERTAINTY in our parameter estimations.**

2. Real-time glucose forecasting **[real-world data]** for:
   - Type 2 Diabetes (patient-facing, meal-time decision support)
   - Critically ill patients in the ICU (clinician-facing decision support)

3. Hybrid machine learning + mechanistic models to account for model error when making predictions
   - mechanistic RNN
   - Modeling residual errors

Parameter Estimation w/ Uncertainty via Bayesian Inversion

Biological Question: What are the relative roles are played by insulin **production** and insulin **sensitivity** in diabetes pathogenesis?

National Institutes of Health (NIH) – Arthur Sherman and Joon Ha

NIH Longitudinal Diabetes Pathogenesis Model (LDPM)

Glucose $$\dot{G} = \mathrm{Meal} + \mathrm{HGP} - (S_G + S_I I)G$$

Insulin $$\dot{I} = \frac{\beta\sigma}{V}\mathrm{ISR}(G) - kI$$

Ha and Sherman 2019. bioRxiv: https://doi.org/10.1101/648816

NIH Longitudinal Diabetes Pathogenesis Model (LDPM)

Insulin Sensitivity

Glucose $$\dot{G} = \text{Meal} + \text{HGP} - (S_G + \boxed{S_I} I)G$$

Insulin $$\dot{I} = \frac{\beta \boxed{\sigma}}{V} \text{ISR}(G) - kI$$

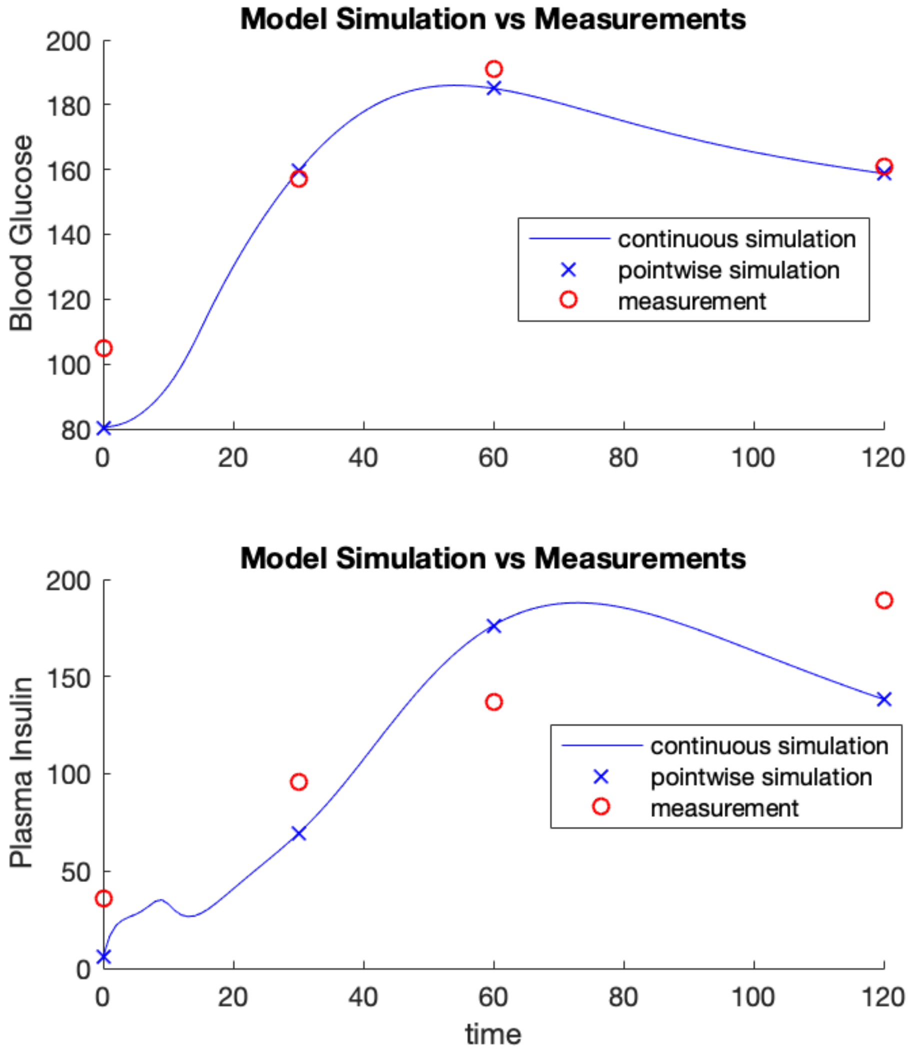
Insulin Production Capacity

Ha and Sherman 2019. bioRxiv: https://doi.org/10.1101/648816

Goal: Characterize endocrine function with parameter estimation from data…w/ UNCERTAINTY

$$\dot{G} = \mathrm{Meal} + \mathrm{HGP} - (S_G + S_I I)G$$

$$\dot{I} = \frac{\beta\sigma}{V}\mathrm{ISR}(G) - kI$$



DATA from Oral Glucose Tolerance Test:

- **Glucose and Insulin Measurements**
- **Measurements every 30min for 2-3 hours**
- **Collected in CLINICAL SETTINGS**

Bayesian Inverse framework

- Consider solution operator to LDPM model

$$\Psi(x(s), t, s, \theta) = x(s) + \int_s^t F(x, \tau, \theta) d\tau$$

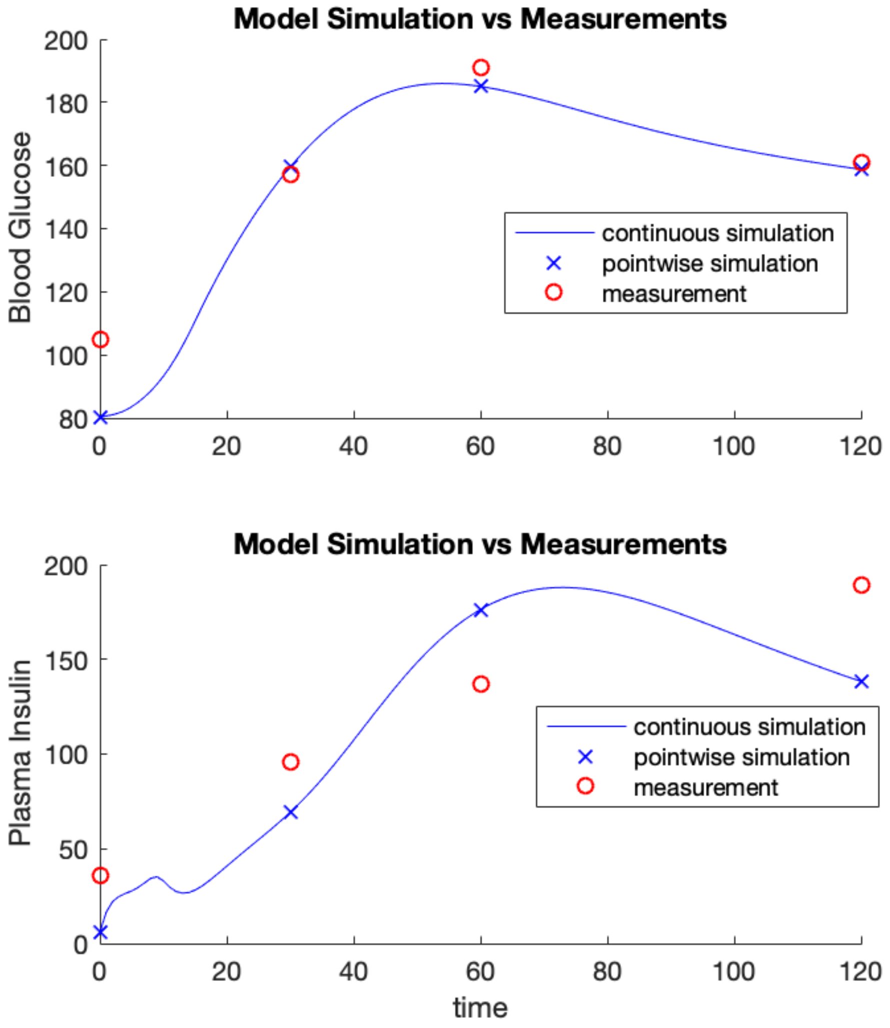
- Deterministic state dynamics governed by parameters theta:

$$x(t) = \Psi(x(s), t, s, \theta)$$

- Solutions at measurement times in observation space:

$$\mathcal{G}(\theta) = \{H(x(t_k))\}_{k=0}^{K}$$

- Data model: $y = \mathcal{G}(\theta) + \eta$



Bayesian Inverse framework

- Deterministic state dynamics governed by parameters theta:

$$x(t) = \Psi(x(s), t, s, \theta)$$

- Solutions at measurement times in observation space:

$$\mathcal{G}(\theta) = \{H\big(x(t_k)\big\}_{k=0}^{K}$$

- Data model:

$$y = \mathcal{G}(\theta) + \eta$$

- Likelihood:

$\mathbb{P}(y|\theta)$ is proportional to $\exp\big(-\Phi(\theta; y)\big)$ where

$$\Phi(\theta; y) = \frac{1}{2}\left\|\Sigma^{-\frac{1}{2}}\big(y - \mathcal{G}(\theta)\big)\right\|^2.$$

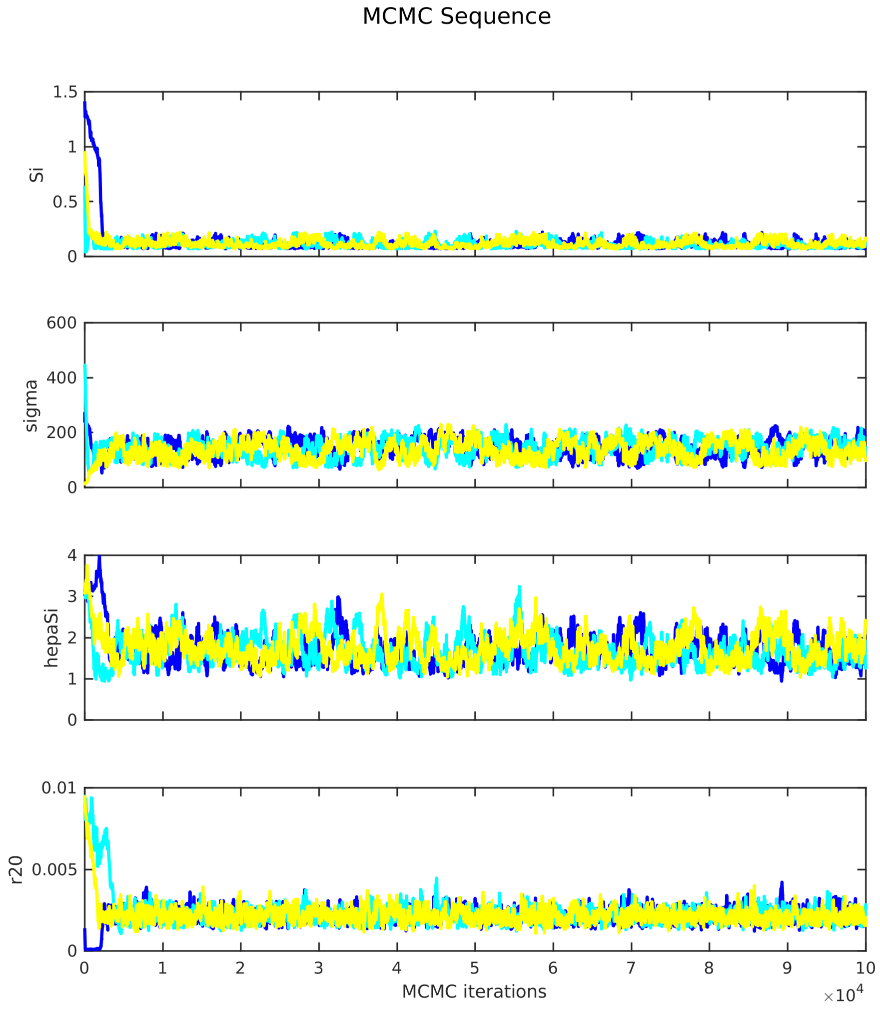
- Posterior:

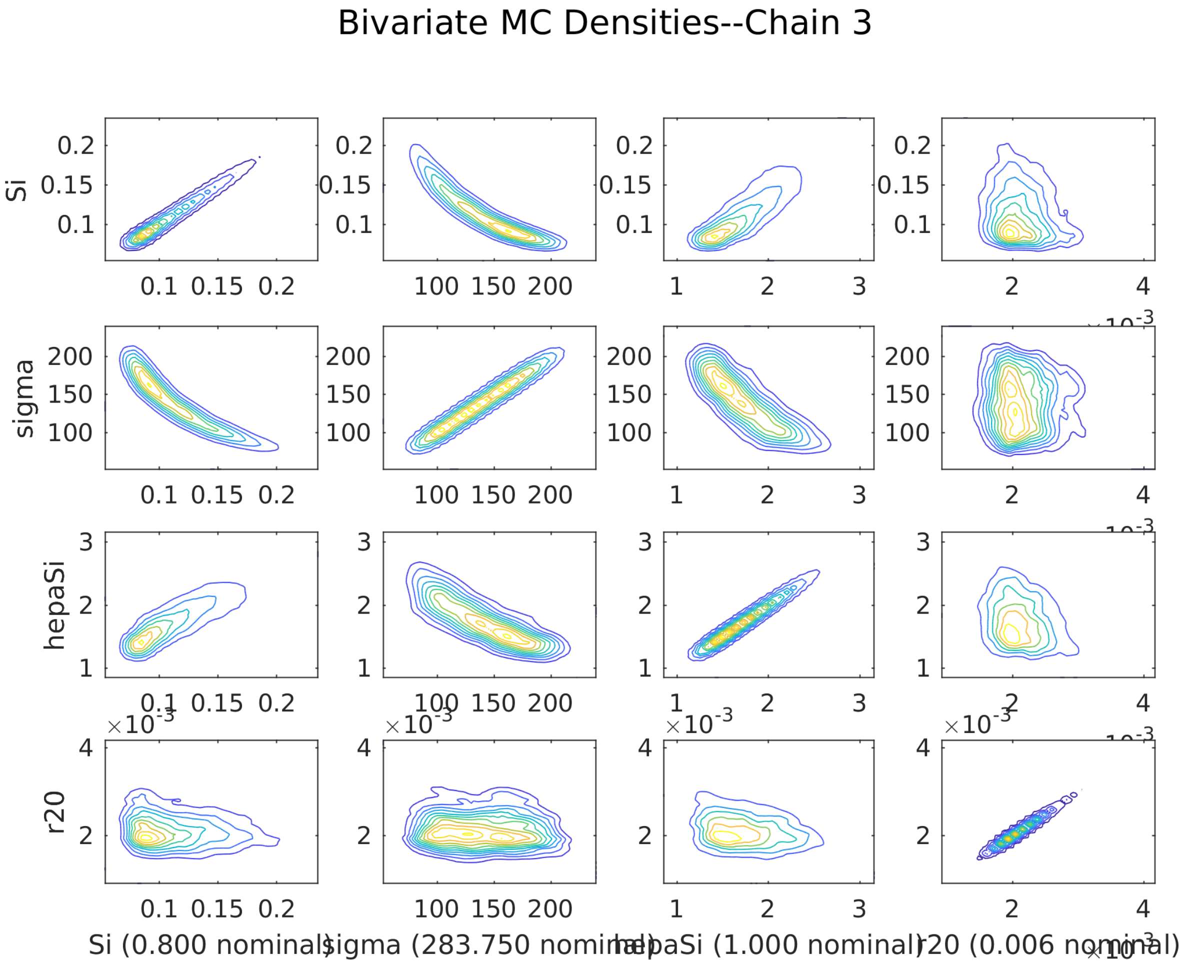
$$\mathbb{P}(\theta|y) \propto \mathbb{P}(y|\theta)\mathbb{P}(\theta)$$

- Uniform Prior

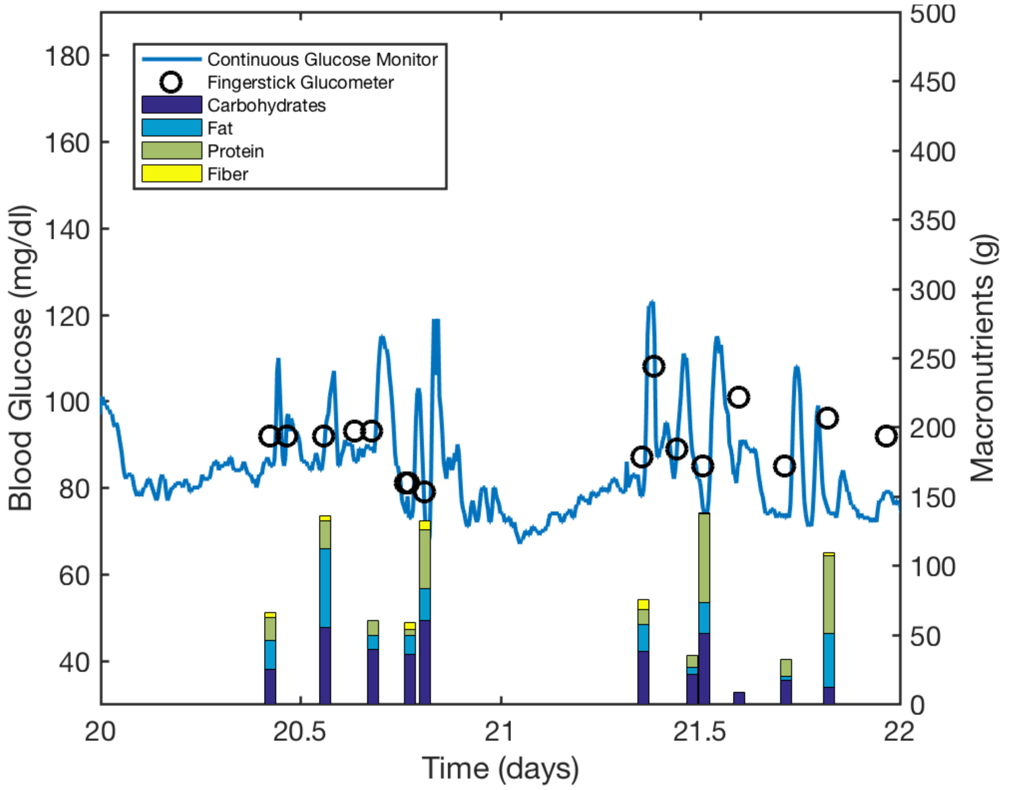
$$\Phi_0(\theta) = \begin{cases} 0 & \text{if } \theta \notin S \\ 1 & \text{if } \theta \in S \end{cases}$$

Ready…set…sample! Metropolis Hastings MCMC (results for 1 single OGTT for 1 single patient)





Now, can we estimate parameters from data collected in the wild by patients?



- Sparse, irregular sampling
- No Insulin Measurements
- Long-term (days to weeks)
- Noisy

Insulin Sensitivity

Glucose $$\dot{G} = \text{Meal} + \text{HGP} - (S_G + \boxed{S_I} I)G$$

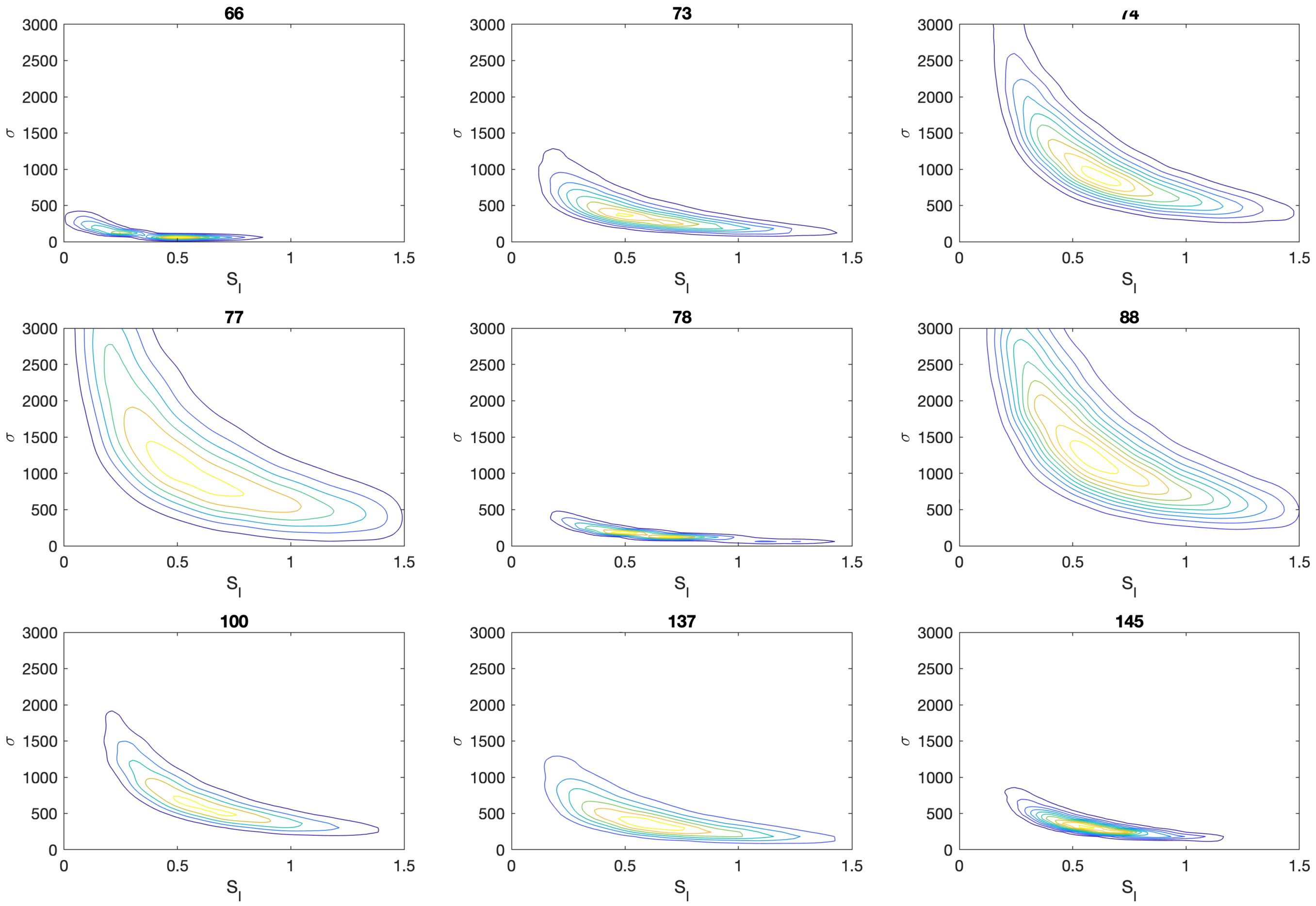
Insulin $$\dot{I} = \frac{\beta \boxed{\sigma}}{V} \text{ISR}(G) - kI$$

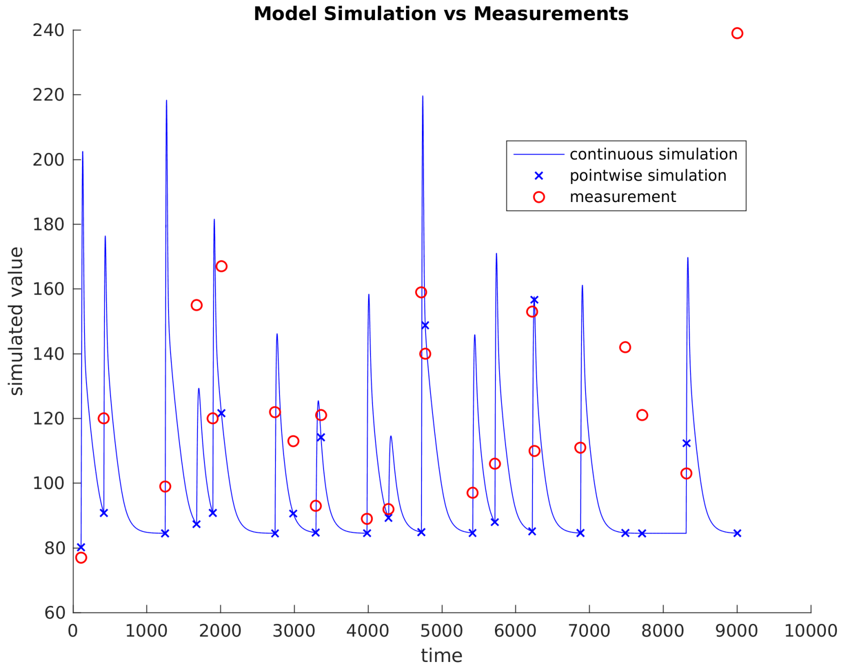
Insulin Production Capacity

NEARLY UNIDENTIFIABLE

Large uncertainty in parameter estimates from free-living data

Posterior probability distributions $\mathbb{P}(\theta|y)$





Summary

Takeaways

- Endocrine inference is highly uncertain and parameters are unidentifiable, especially in free-living data
- This uncertainty/identifiability can be CHARACTERIZED with MCMC and other sampling techniques
- Can be USED to generate posterior distribution of Disposition Index (DI)

Future directions

- Estimate posterior parameter distributions for patients from a population
- Assemble these estimates into a "**population distribution**"
- Use this "**population distribution**" to better inform future inferences on this population

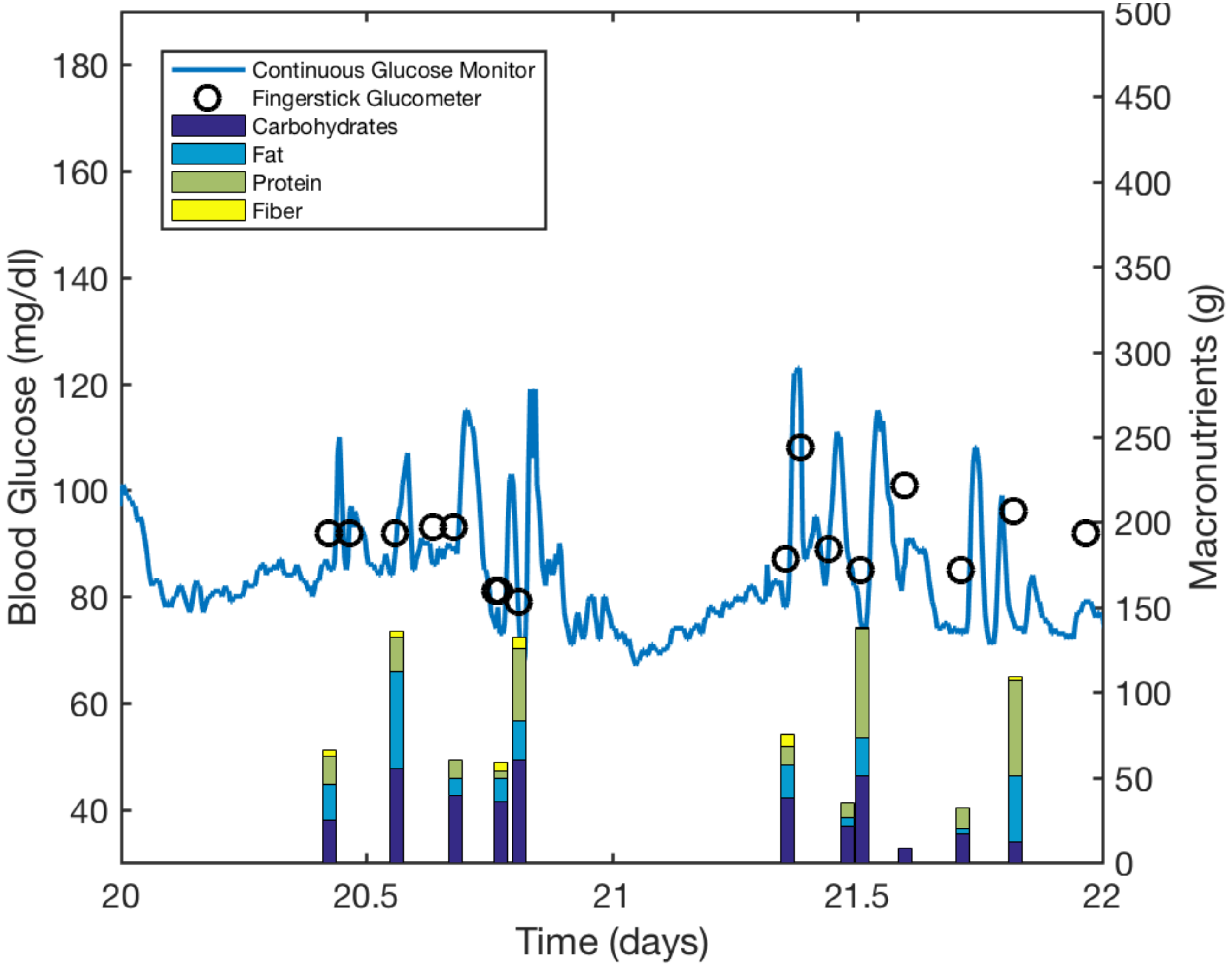
Real-time glucose forecasting via Data Assimilation

Data Assimilation for real-time prediction

- Applications
  - Type 2 Diabetes (patient-facing, meal-time decision support)
  - Critically ill patients in the ICU (clinician-facing decision support)
- The challenge
  - Incorporate (”*assimilate*”) new/changing information into current belief about present and future…*in real-time*
  - *We NEVER observe insulin measurements in the wild!*
- Our approach: *Stochastic Filtering*
  - Linear models -> Kalman Filter
  - Non-linear models -> Non-linear filters (Particle Filters, Unscented KF, EnKF)

Type 2 Diabetes Self-Monitoring Data

- Sparse, irregular sampling
- No Insulin Measurements
- Long-term (days to weeks)
- Noisy



Data Assimilation: Mathematical Framing

Consider the discrete-time dynamical system with noisy state transitions and noisy observations in the form:

$$\begin{aligned}
\text{Dynamics Model:} \quad & v_{j+1} = \Psi(v_j) + \xi_j, \quad j \in \mathbb{Z}^+ \\
\text{Data Model:} \quad & y_{j+1} = h(v_{j+1}) + \eta_{j+1}, \quad j \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} \quad & v_0 \sim N(m_0, C_0), \quad \xi_j \sim N(0, \Sigma), \quad \eta_j \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_j\} \perp \{\eta_j\} \text{ independent}
\end{aligned}$$

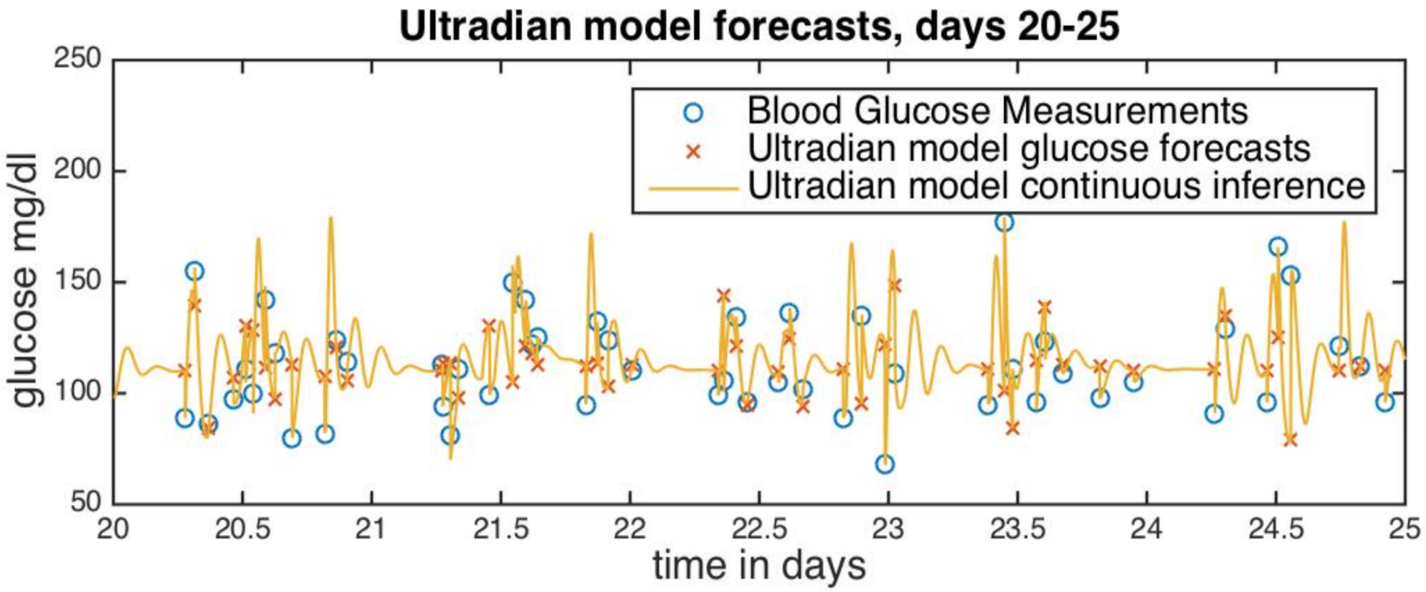
For linear models, use a Kalman Filter!
For non-linear models, need to approximate the mapping of the distribution…use non-linear filter!

Here, h chooses the glucose state, and the dynamics are governed by a continuous-time system

Unscented Kalman Filter for personalized glucose forecasting

Albers, Levine, Gluckman, Ginsberg, Hripcsak, and Mamykina 2017

- Iterative prediction-correction scheme
- Can track states and parameters (dual, joint)

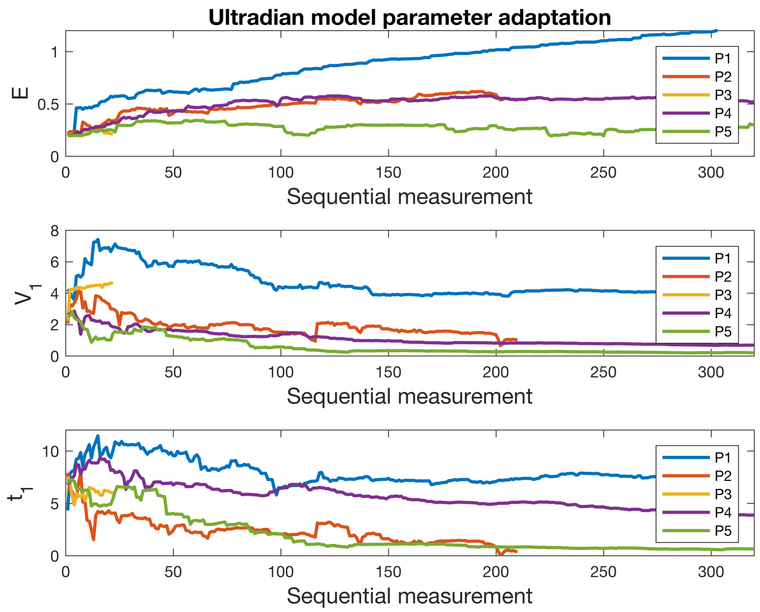


Unscented Kalman Filter for personalized glucose forecasting

Albers, Levine, Gluckman, Ginsberg, Hripcsak, and Mamykina 2017

Significant challenges exist in parameter estimation with dual UKF.

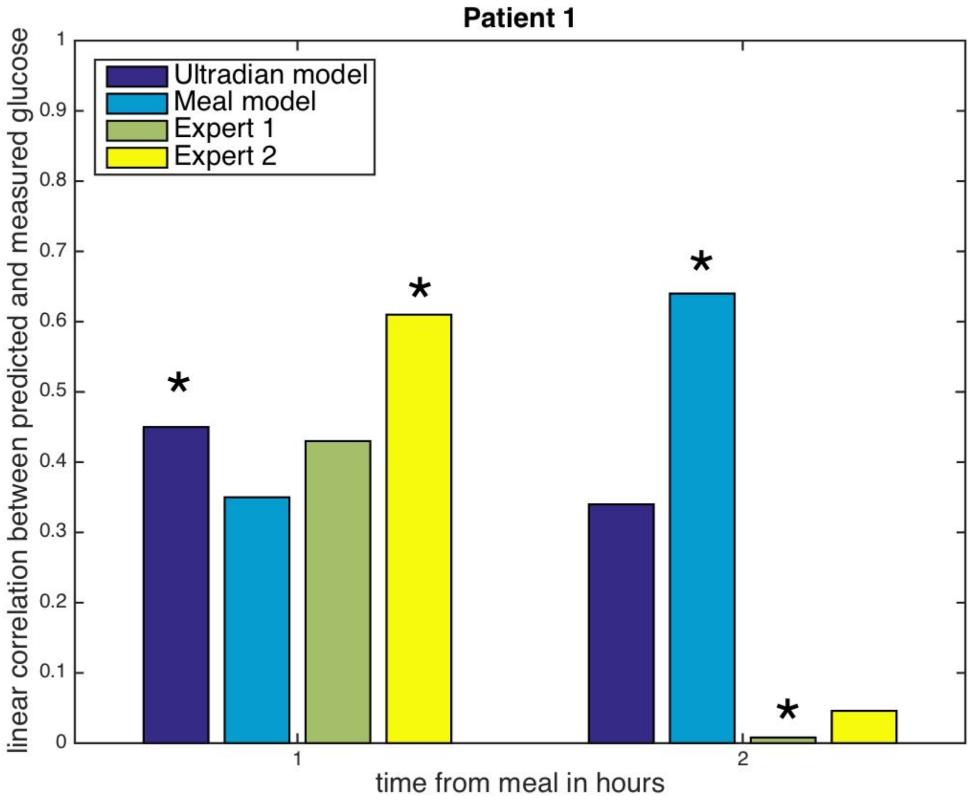
- Parameter estimates often do not converge
- UKF does not explore full parameter space
- Parameter tracking is designed to adapt to *between-measurement dynamics*, not dynamics across multiple measurements



Unscented Kalman Filter for personalized glucose forecasting

Albers, Levine, Gluckman, Ginsberg, Hripcsak, and Mamykina 2017

Dual UKF often matches or beats clinical experts forecasts.



Results from real-time glucose forecasting

- **PREVIOUS WORK:** UKF w/ Cobelli model is operationalized in a patient-facing mobile application that is used by people with T2D for meal-time decision support (*Albers et al. Plos Comp Bio 2017*)
  - Learning parameters is ESSENTIAL
- **More recently:**
  - *Simpler, non-mechanistic models seem to have better predictive performance*
  - <u>**Can CONSTRAIN the state space of EnKF, and this helps for operationalizing**</u>

Ensemble Kalman Filter (Evensen 2003)

Consider the discrete-time dynamical system with noisy state transitions and noisy observations in the form:

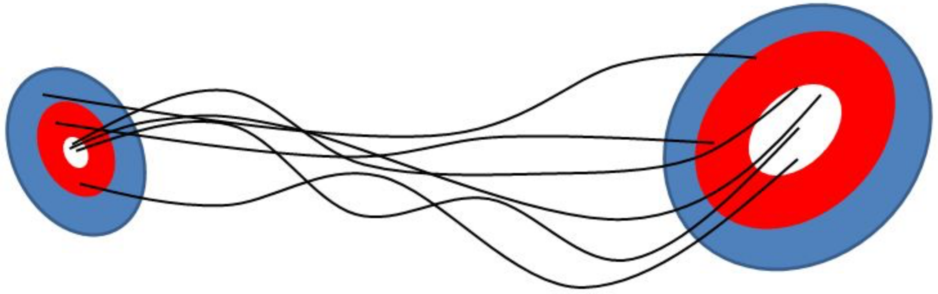
$$\begin{aligned}
\text{Dynamics Model:} &\quad v_{j+1} = \Psi(v_j) + \xi_j, \quad j \in \mathbb{Z}^+ \\
\text{Data Model:} &\quad y_{j+1} = h(v_{j+1}) + \eta_{j+1}, \quad j \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} &\quad v_0 \sim N(m_0, C_0), \quad \xi_j \sim N(0, \Sigma), \quad \eta_j \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} &\quad v_0 \perp \{\xi_j\} \perp \{\eta_j\} \text{ independent}
\end{aligned}$$

Assume Gaussian states:

$$P(v_j|y_j) \sim N(m_j, C_j)$$

$$P(v_{j+1}|y_j) \sim N(\widehat{m}_j, \widehat{C}_j)$$

Ensemble Kalman Filter



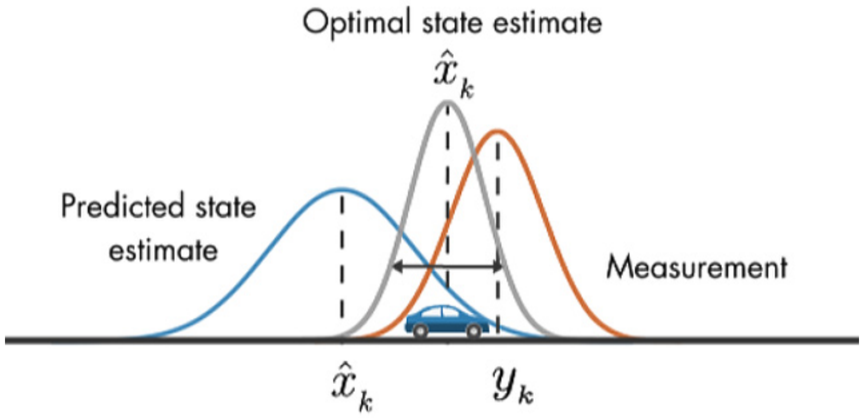
The prediction step is

$$\widehat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}, n = 1, ..., N \tag{2.1a}$$

$$\widehat{m}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} \widehat{v}_{j+1}^{(n)} \tag{2.1b}$$

$$\widehat{C}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} (\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})(\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})^T \tag{2.1c}$$

The update step is then



$$S_{j+1} = H\widehat{C}_{j+1}H^T + \Gamma \tag{2.4a}$$

$$K_{j+1} = \widehat{C}_{j+1}H^T S_{j+1}^{-1} \qquad \text{(Kalman Gain)} \tag{2.4b}$$

$$y_{j+1}^{(n)} = y_{j+1} + s\eta_{j+1}^{(n)}, n = 1, ..., N \tag{2.4c}$$

$$v_{j+1}^{(n)} = (I - K_{j+1}H)\widehat{v}_{j+1}^{(n)} + K_{j+1}y_{j+1}^{(n)}, n = 1, ..., N \tag{2.4d}$$

Constrained Ensemble Kalman Filtering—Why?

PROBLEM: Gaussian has infinite support, but our problem space often only makes sense on a compact set

GOALS:

- Enforce model physicality (e.g. positivity)
- Maintain problem well-posedness (e.g. avoid parameter regimes that make forward map intractable)
- Provide robustness to outlier data

Ensemble Kalman Filtering (EnKF) framework

The prediction step is

$$\widehat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}, n = 1, ..., N \tag{2.1a}$$

$$\widehat{m}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} \widehat{v}_{j+1}^{(n)} \tag{2.1b}$$

$$\widehat{C}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} (\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})(\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})^T \tag{2.1c}$$

The update step is then

$$S_{j+1} = H\widehat{C}_{j+1}H^T + \Gamma \tag{2.4a}$$

$$K_{j+1} = \widehat{C}_{j+1}H^T S_{j+1}^{-1} \qquad \text{(Kalman Gain)} \tag{2.4b}$$

$$y_{j+1}^{(n)} = y_{j+1} + s\eta_{j+1}^{(n)}, n = 1, ..., N \tag{2.4c}$$

$$v_{j+1}^{(n)} = (I - K_{j+1}H)\widehat{v}_{j+1}^{(n)} + K_{j+1}y_{j+1}^{(n)}, n = 1, ..., N \tag{2.4d}$$

Kalman update can be rewritten as a quadratic minimization

The prediction step is

$$\widehat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}, n = 1, ..., N \tag{2.1a}$$

$$\widehat{m}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} \widehat{v}_{j+1}^{(n)} \tag{2.1b}$$

$$\widehat{C}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} (\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})(\widehat{v}_{j+1}^{(n)} - \widehat{m}_{j+1})^T \tag{2.1c}$$

The update step is then

$$v_{j+1}^{(n)} = \underset{v}{\operatorname{argmin}}\, I_{\text{filter,j,n}}(v)$$

where

$$I_{\text{filter},j,n}(v) := \begin{cases} \frac{1}{2} \mid y_{j+1}^{(n)} - Hv \mid_{\Gamma}^{2} + \frac{1}{2} \mid v - \widehat{v}_{j+1}^{(n)} \mid_{\widehat{C}_{j+1}}^{2} & \text{if } v - \widehat{v}_{j+1}^{(n)} \in \mathcal{R}(\widehat{C}_{j+1}). \\ \infty & \text{otherwise.} \end{cases} \tag{2.3a}$$

Constrained EnKF (Inverse Problems 2019)

Why constrain the system?

- Ensure physicality (e.g. Positivity of a physical concentrations)
- Ensure tractability of forward model

How to constrain the system?

- Minimize I_filter subject to linear equality and inequality constraints

The update step is then

$$v_{j+1}^{(n)} = \underset{v}{\operatorname{argmin}}\, I_{\mathrm{filter,j,n}}(v)$$

$$Fv = f,$$

$$Gv \preceq g.$$

where

$$I_{\mathrm{filter},j,n}(v) := \begin{cases} \frac{1}{2} \mid y_{j+1}^{(n)} - Hv \mid_{\Gamma}^{2} + \frac{1}{2} \mid v - \widehat{v}_{j+1}^{(n)} \mid_{\widehat{C}_{j+1}}^{2} & \text{if } v - \widehat{v}_{j+1}^{(n)} \in \mathcal{R}(\widehat{C}_{j+1}). \\ \infty & \text{otherwise.} \end{cases} \quad (2.3a)$$

Numerical results from Constrained EnKF application

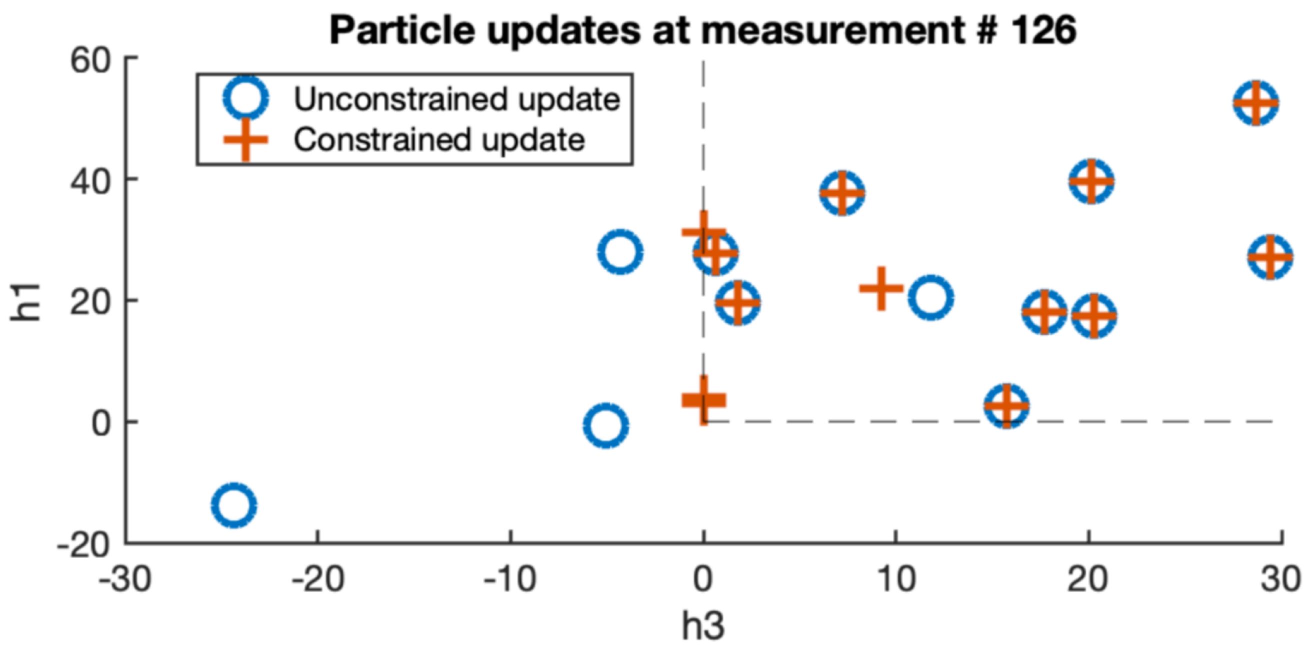


Figure 2 Particle updates at a given time-step (here, measurement 126) are shown using a traditional Kalman gain versus using the constrained optimization. The black lines denote lower bound constraints on the states h_1 and h_3.

Open questions/problems for real-time glucose forecasting

- Improvements with Offline/Online parameter estimation
- UQ of forecasts
- Incorporate new, informative data elements (protein, fat, sleep, exercise, insulin, medications)
- Model Selection/Averaging/Blending with Machine Learning

Uniting mechanistic modeling and machine learning for enhanced time-series predictions

Forecasting a dynamical system (simple, naive setting)

- Say we observe data from the true system: a discrete, deterministic dynamical system of form

$$u_{k+1} = \Psi(u_k)$$

- But we only have a model hypothesis $\tilde{\Psi}$
- How do we predict future trajectory?
- Scenarios

1. We know a lot about the system, and believe our hypothesis is true up to a specific parameterization…in this case, use DA for filtering!
   - i.e. Block on a spring experiment
2. We know NOTHING about the system, so our hypothesis is a general function class…this is ML/deep learning/etc.

Forecasting a dynamical system (simple, naive setting)

- Say we observe data from the true system: a discrete, deterministic dynamical system of form

$$u_{k+1} = \Psi(u_k)$$

- But we only have a model hypothesis $\tilde{\Psi}$
- How do we predict future trajectory?
- Scenarios

1. We know a lot about the system, and believe our hypothesis is true up to a specific parameterization…in this case, use DA for filtering!
    - i.e. Block on a spring experiment
2. We know NOTHING about the system, so our hypothesis is a general function class…this is ML/deep learning/etc.
3. **We hypothesize a specific mechanism with modest predictive power, but substantial inadequacies.**

Start with simple Recurrent Neural Network (RNN)

$$r_{k+1} = \sigma\left(a + Ar_k + Bd_k\right)$$

$$d_{k+1} = b + Cr_{k+1},$$

with parameter set $\Theta = \{A, B, C, a, b\}$

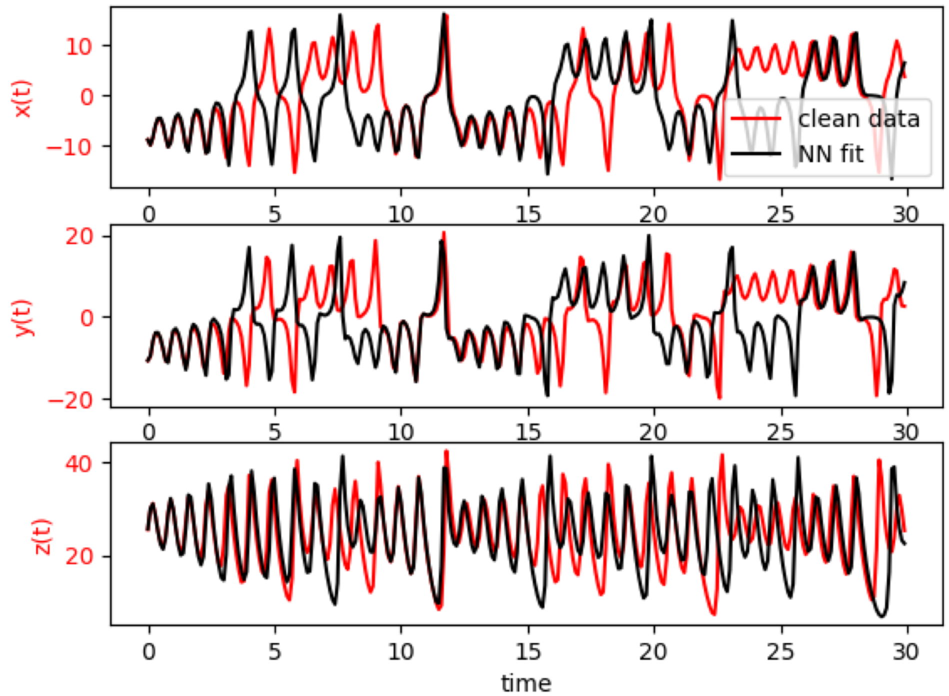
$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{k=1}^{K} ||u_k - d_k||^2$$

RNN can learn to predict the Lorenz 63 system

$$r_{k+1} = \sigma\left(a + Ar_k + Bd_k\right)$$

$$d_{k+1} = b + Cr_{k+1}, \quad \text{with parameter set } \Theta = \{A, B, C, a, b\}$$

$$\Theta^* = \underset{\Theta}{\text{argmin}} \sum_{k=1}^{K} ||u_k - d_k||^2$$



But can the RNN do better with more information? We propose the "mechRNN"

mechRNN

$$r_{k+1} = \sigma\left(a + Ar_k + B\begin{bmatrix}\tilde{\Psi}(d_k) \\ d_k\end{bmatrix}\right)$$

$$d_{k+1} = b + C\begin{bmatrix}\tilde{\Psi}(d_k) \\ r_{k+1}\end{bmatrix}$$

with parameter set $\Theta = \{A, B, C, a, b\}$

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{k=1}^{K} ||u_k - d_k||^2$$

Analogous to Reservoir Computing approach by Pathak et al. (Chaos 2018)

Lorenz 63 with perturbed parameter: A model for model error

$$\frac{dx}{dt} = -a(x+y)$$
$$\frac{dy}{dt} = bx - y - xz \qquad \text{with } a = 10, b = 28, c = 8/3,$$
$$\frac{dz}{dt} = -cz + xy,$$

$$\dot{u} = \tilde{f}(u) = f(u, \tilde{b})$$ Lorenz 63 with perturbed b parameter

$$\Psi(u_k) = u_k + \int_{t_k}^{t_{k+1}} f(u(t))dt$$ TRUE generating system for training data

$$\tilde{\Psi}(u_k) = u_k + \int_{t_k}^{t_{k+1}} \tilde{f}(u(t))dt$$ ASSUMED, but WRONG generating system for training data

So can the RNN do better with imperfect information? Yes!

mechRNN

$$r_{k+1} = \sigma\left(a + Ar_k + B\begin{bmatrix}\tilde{\Psi}(d_k) \\ d_k\end{bmatrix}\right)$$

$$d_{k+1} = b + C\begin{bmatrix}\tilde{\Psi}(d_k) \\ r_{k+1}\end{bmatrix}$$

with parameter set $\Theta = \{A, B, C, a, b\}$

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{k=1}^{K} ||u_k - d_k||^2$$

Analogous to Reservoir Computing approach by Pathak et al. (Chaos 2018)

So can the RNN do better with imperfect information? Yes!

mechRNN

$$r_{k+1} = \sigma\left(a + Ar_k + B\begin{bmatrix}\tilde{\Psi}(d_k) \\ d_k\end{bmatrix}\right)$$

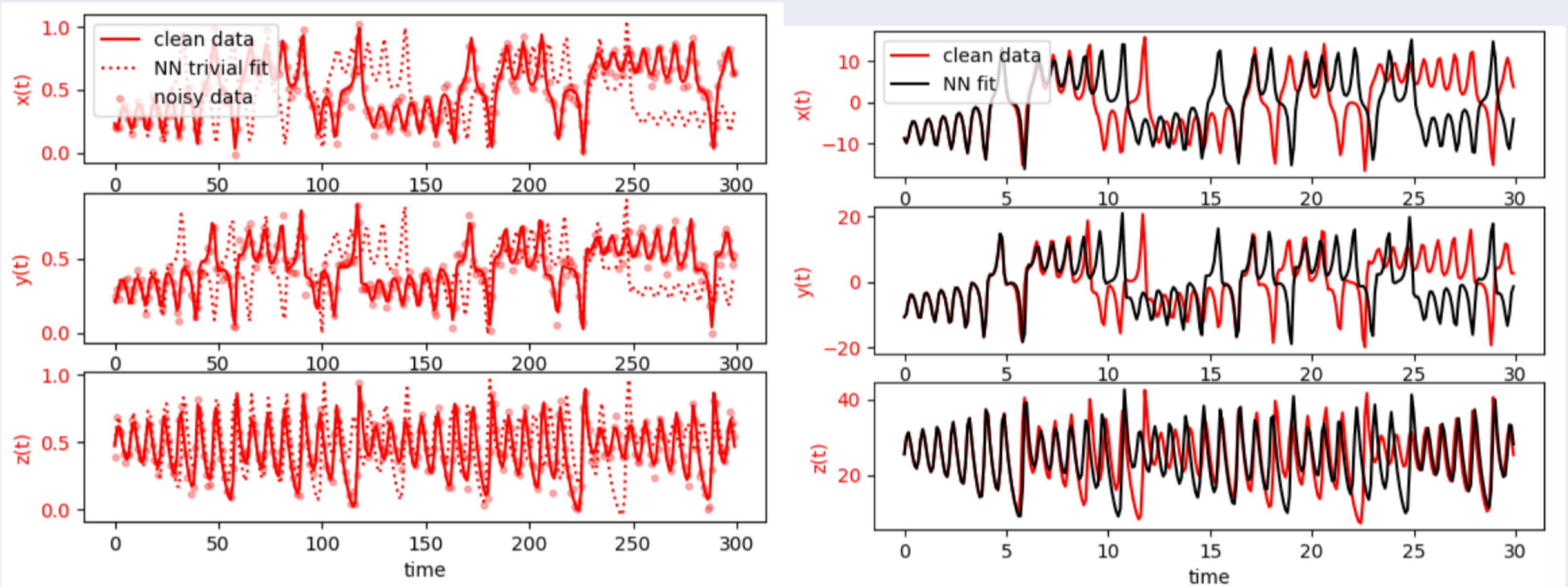
$$d_{k+1} = b + C\begin{bmatrix}\tilde{\Psi}(d_k) \\ r_{k+1}\end{bmatrix}$$

with parameter set $\Theta = \{A, B, C, a, b\}$

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{k=1}^{K} ||u_k - d_k||^2$$

Analogous to Reservoir Computing approach by Pathak et al. (Chaos 2018)

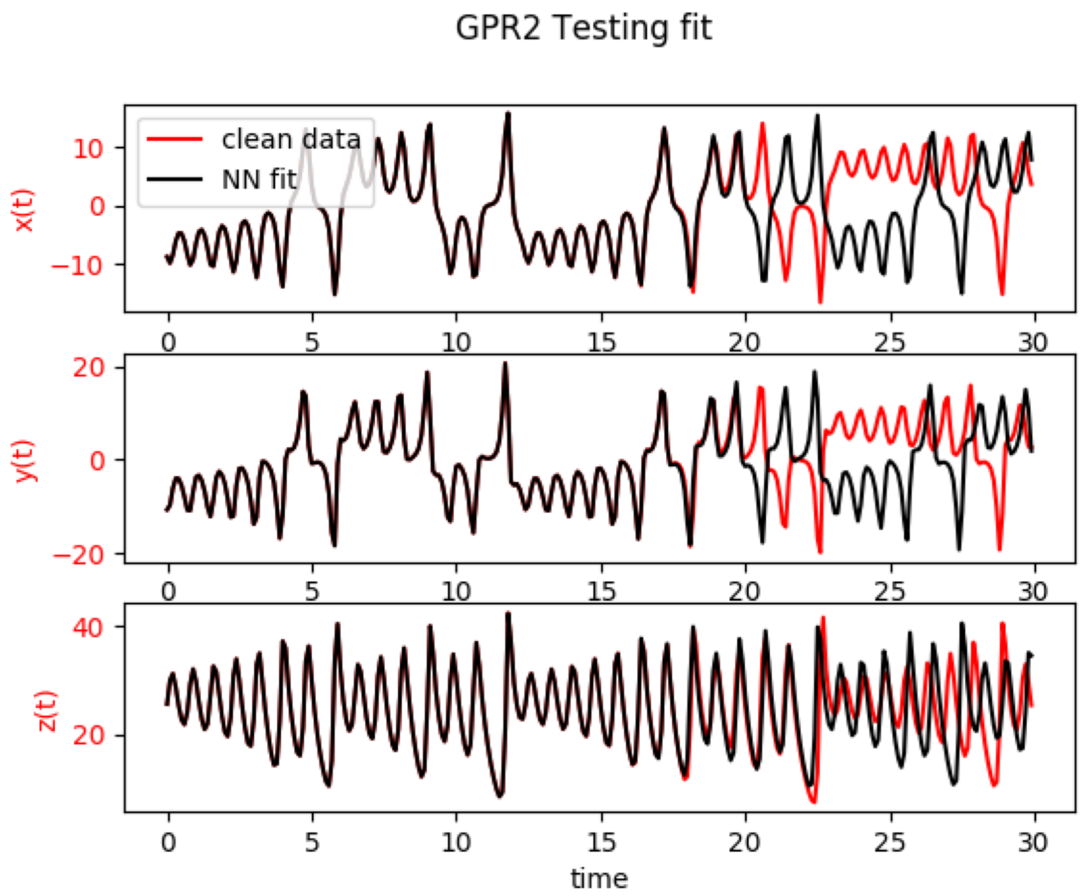
mechRNN w/ $\tilde{\Psi}$ **s.t.** $\tilde{b} = (1+\varepsilon)b$, $\varepsilon = 0.05$



- **Training data** comes from high-fidelity solutions to Lorenz 63 with classical chaotic parameters
- **mechRNN** only sees the solution operator to a perturbed version of the true system

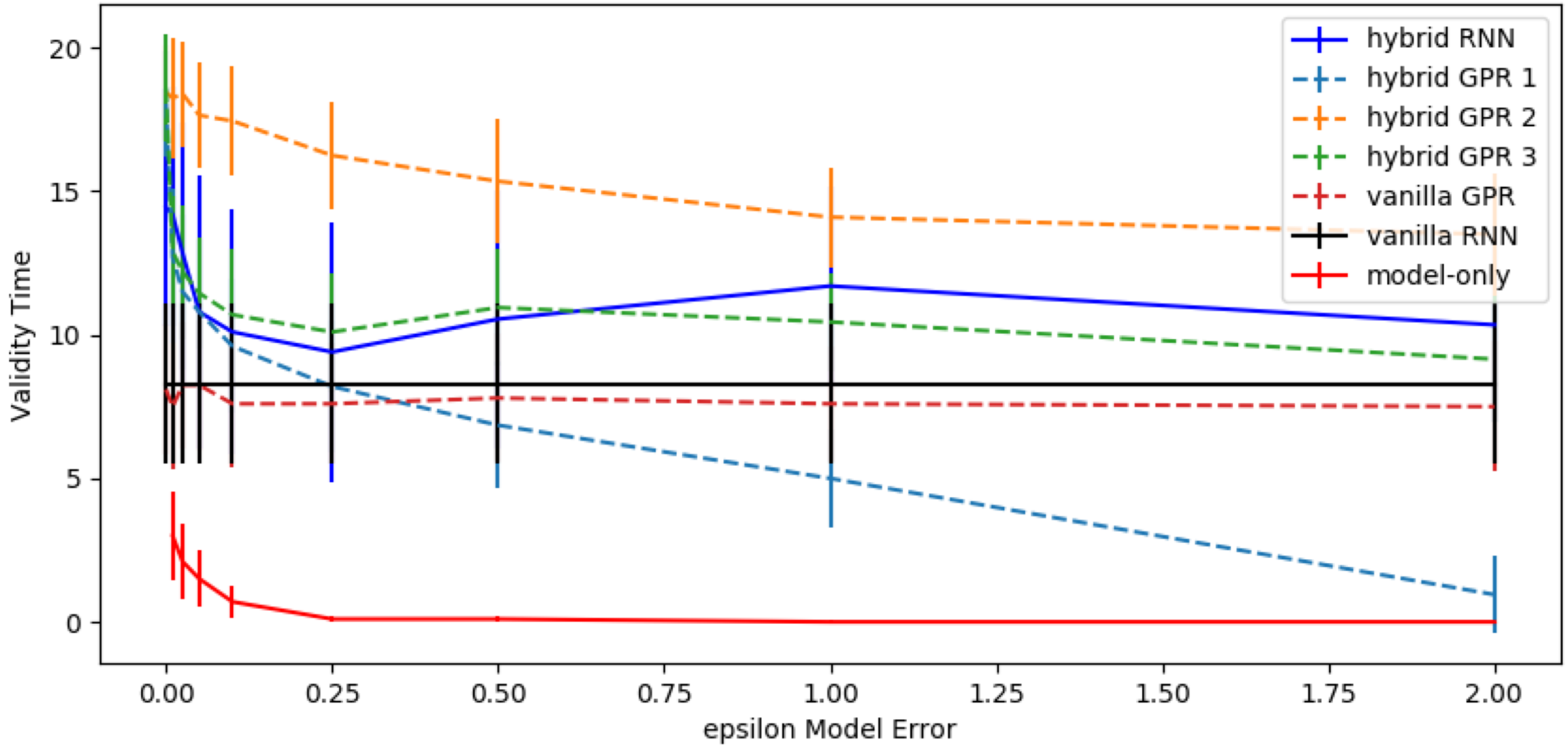
Even simpler, can we learn residual errors? Yes!

$$d_{k+1} = \tilde{\Psi}(d_k) + G\left(P\begin{bmatrix}\tilde{\Psi}(d_k)\\ d_k\end{bmatrix}\right)$$

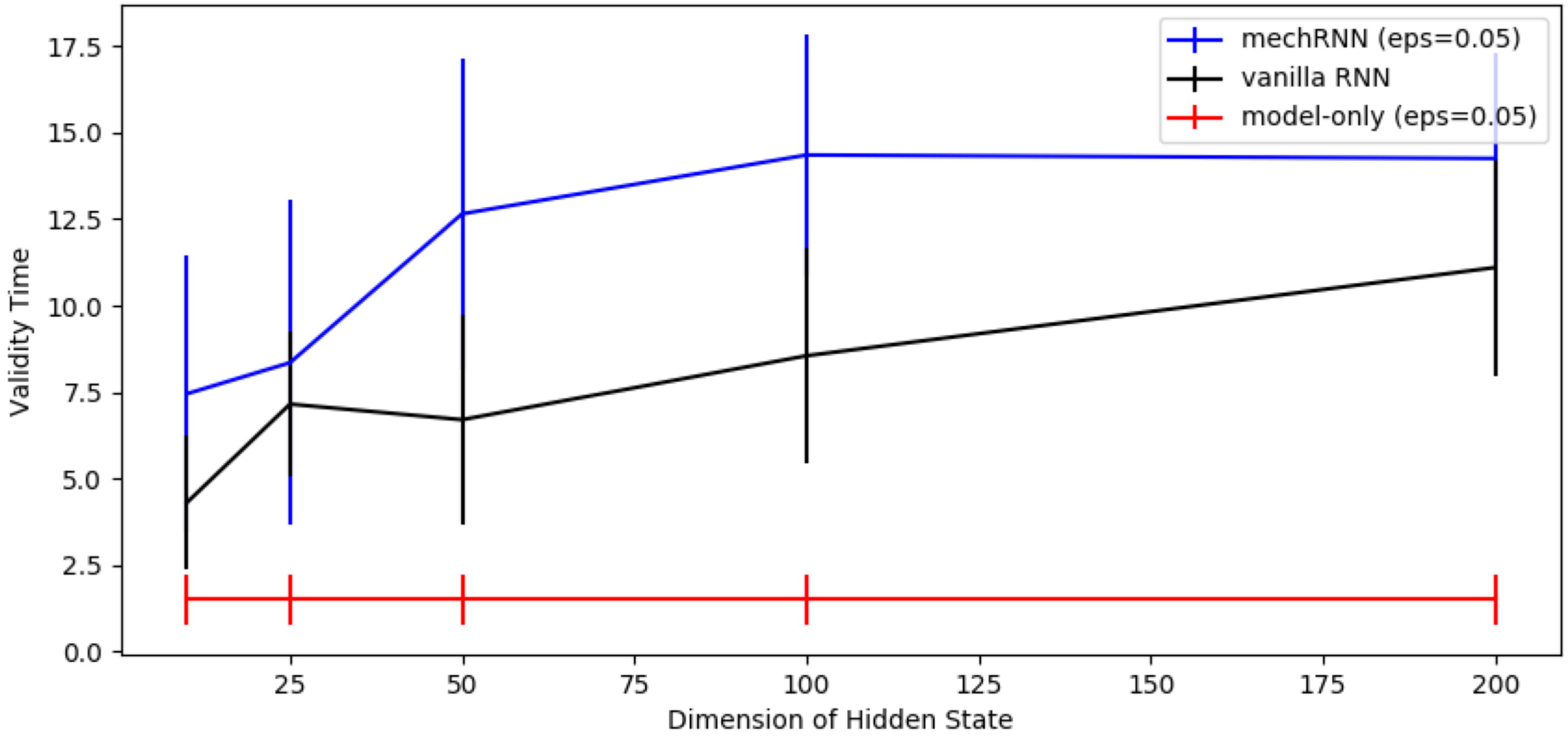


We learn G as a Gaussian Process Regression

Hybrid methods correct for large arbitrary model error

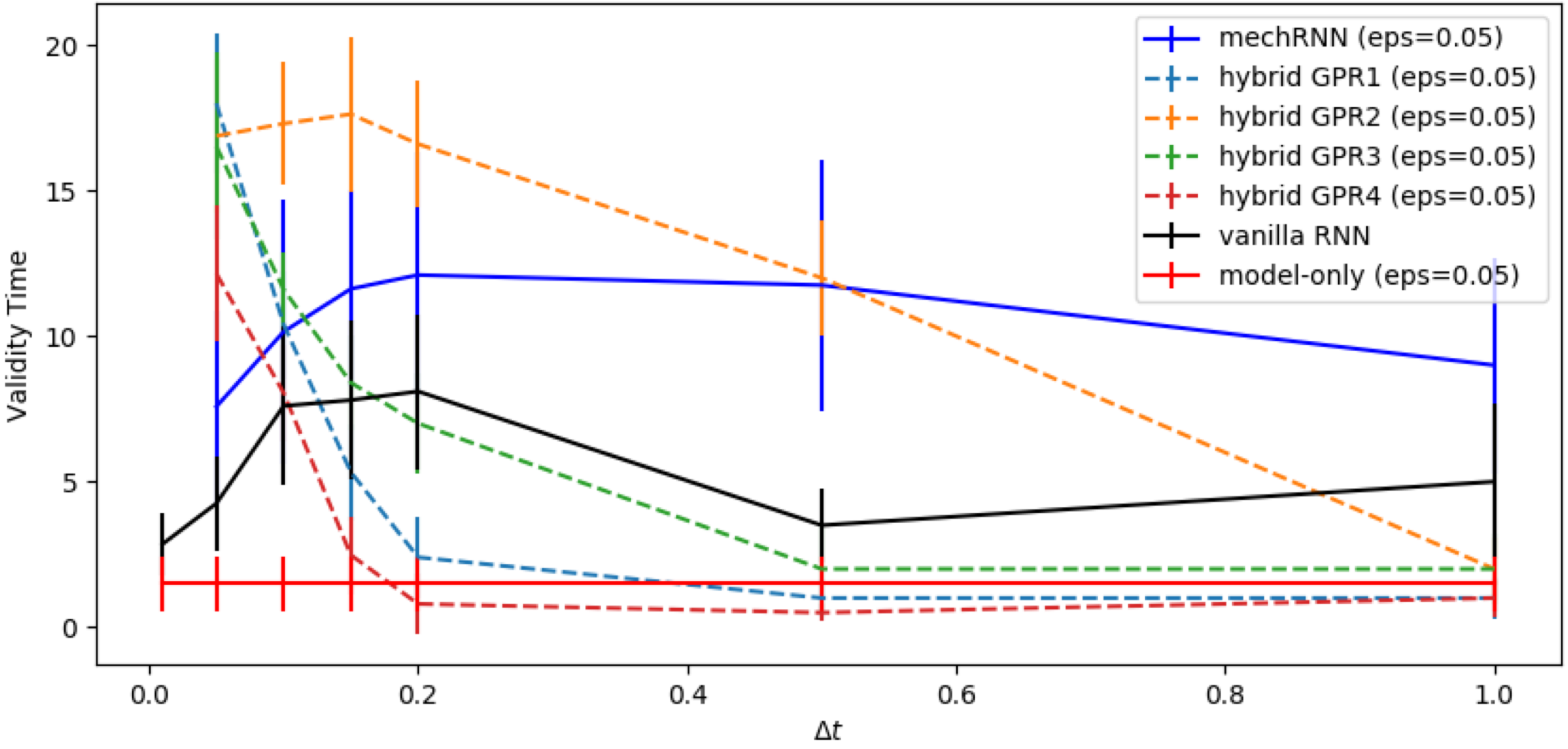


mechRNN requires fewer parameters than vanillaRNN



mech RNN w/ $\tilde{\Psi}$ **s.t.** $\tilde{b} = (1+\varepsilon)b$**,** $\varepsilon = 0.05$

Tradeoff between mechRNN and GP-based residual learning---a function of the data's sampling rate



Future Directions for mechanisms+ML

- Extend to data assimilation context (partial, noisy observations with irregular sampling)
- Flexible model averaging: exploit a family of models $\{\tilde{\Psi}_i\}_{i=1}^N$
- Extend to non-autonomous systems
- Allow for parameter inference within the mechanistic model
- **In diabetes free-living case, use to ensemble models AND learn temporal impact of fat/protein/fiber and exercise.**

Acknowledgments

- **Caltech**
  - **Andrew Stuart**
- **Columbia University**
  - **David Albers (now at University of Colorado)**
  - **George Hripcsak**
  - **Lena Mamykina**
- **National Institutes of Health**
  - **Arthur Sherman**
  - **Joon Ha**