Slides for Work on Final Report

Christina Catlett, Daniel Shenker, Rachel Wander, & Maya Watanabe
HMC REU SUMMER 2020

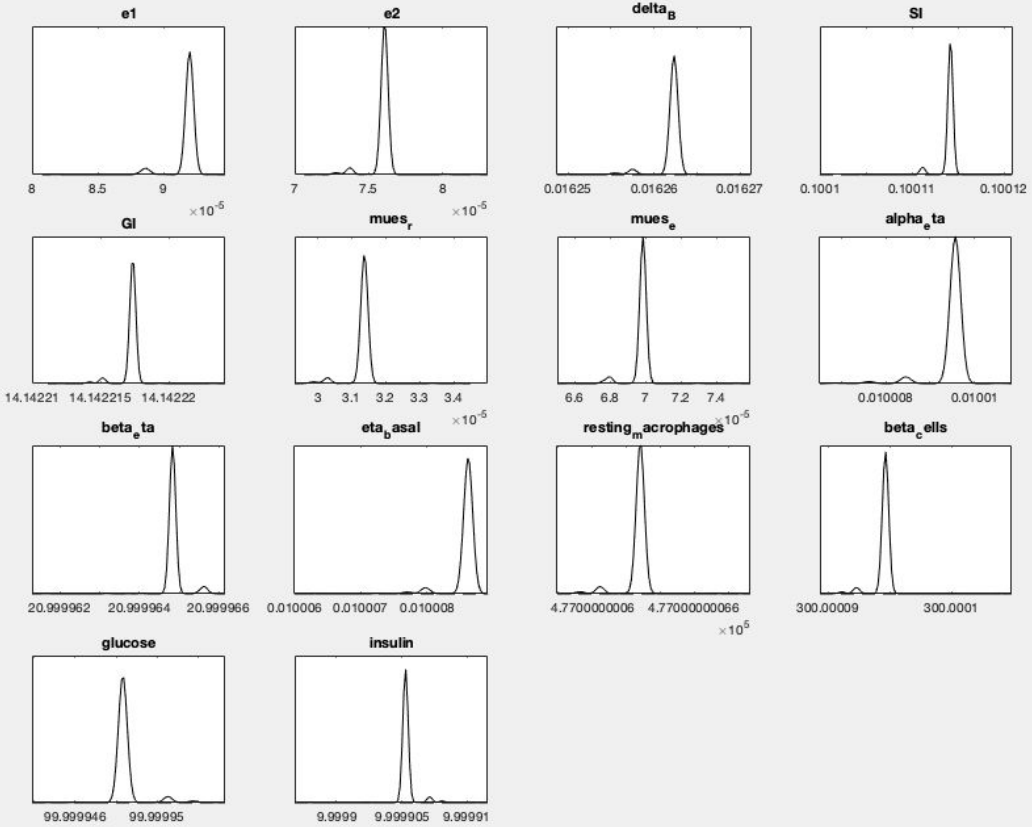
Week 7 Presentation

Christina, Daniel, Maya, and Rachel

Merging UKFs and MCMC - Process

1. Isolate acute mice (9 data sets)
2. Run Dual UKF on each mouse
3. Fit normal distributions to all parameters
4. Identify most volatile parameters
5. Use distributions as priors for those volatile parameters in MCMC

Merging UKFs and MCMC - *results*



Posterior PDFs

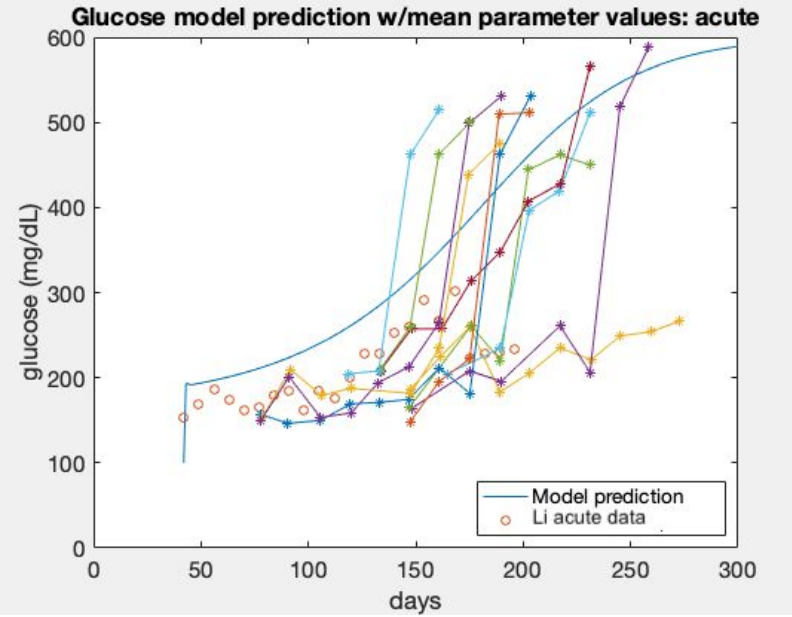
- odd that they all have such a similar shape

Chain stats

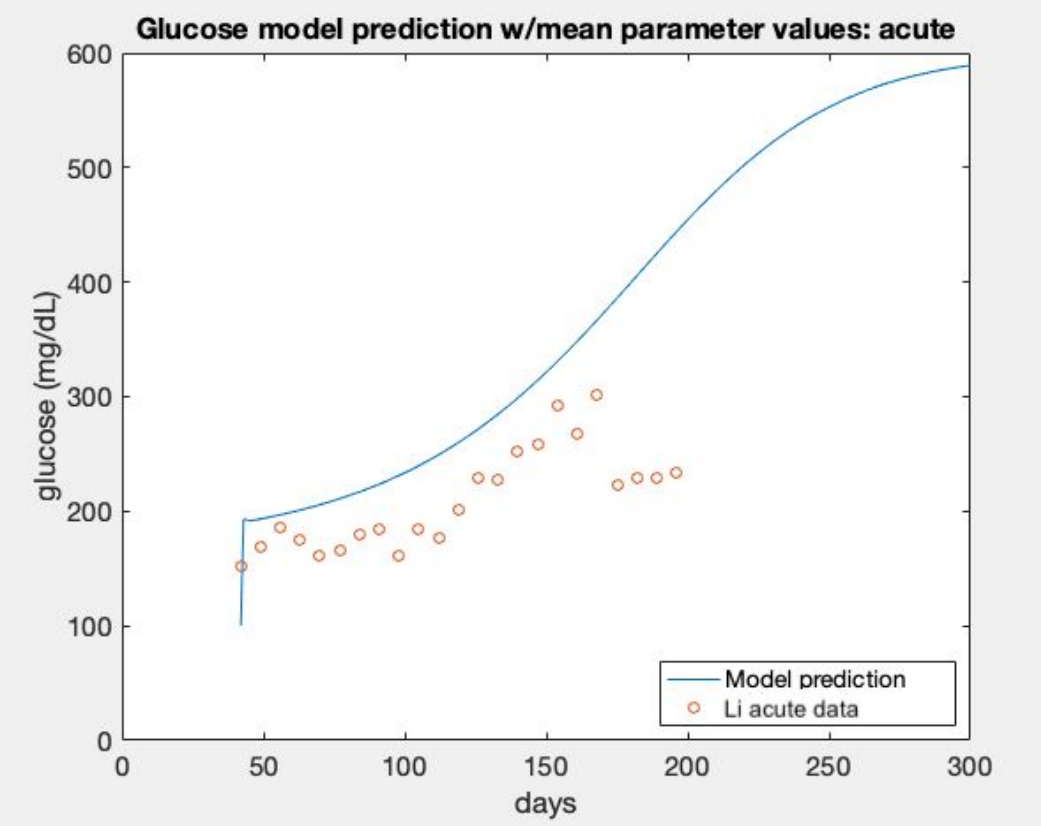
- chain converges for all parameters

	mean	std	MC_err	tau	geweke
e1	9.1765e-05	8.9274e-07	1.7365e-07	430.78	0.96732
e2	7.5933e-05	6.9104e-07	1.0402e-07	302.13	0.97565
delta_B	0.016262	1.383e-06	2.5293e-07	393.23	0.99973
SI	0.10011	8.2185e-07	1.4756e-07	387.99	0.99997
GI	14.142	6.5153e-07	1.0777e-07	342.94	1
mues_r	3.1313e-05	3.1597e-07	4.7267e-08	300.73	0.97305
mues_e	6.9735e-05	5.3844e-07	7.8749e-08	292.94	0.97976
alpha_eta	0.010009	3.6079e-07	6.7734e-08	395.33	0.99989
beta_eta	21	2.3527e-07	2.4312e-08	188.07	1
eta_basal	0.010009	2.017e-07	3.8773e-08	410.23	0.99994
resting_macrophages	4.77e+05	7.2015e-07	1.132e-07	313.83	1
beta_cells	300	7.7282e-07	1.0041e-07	239.29	1
glucose	100	6.4195e-07	1.1105e-07	363.99	1
insulin	9.9999	6.0964e-07	8.4353e-08	260.63	1

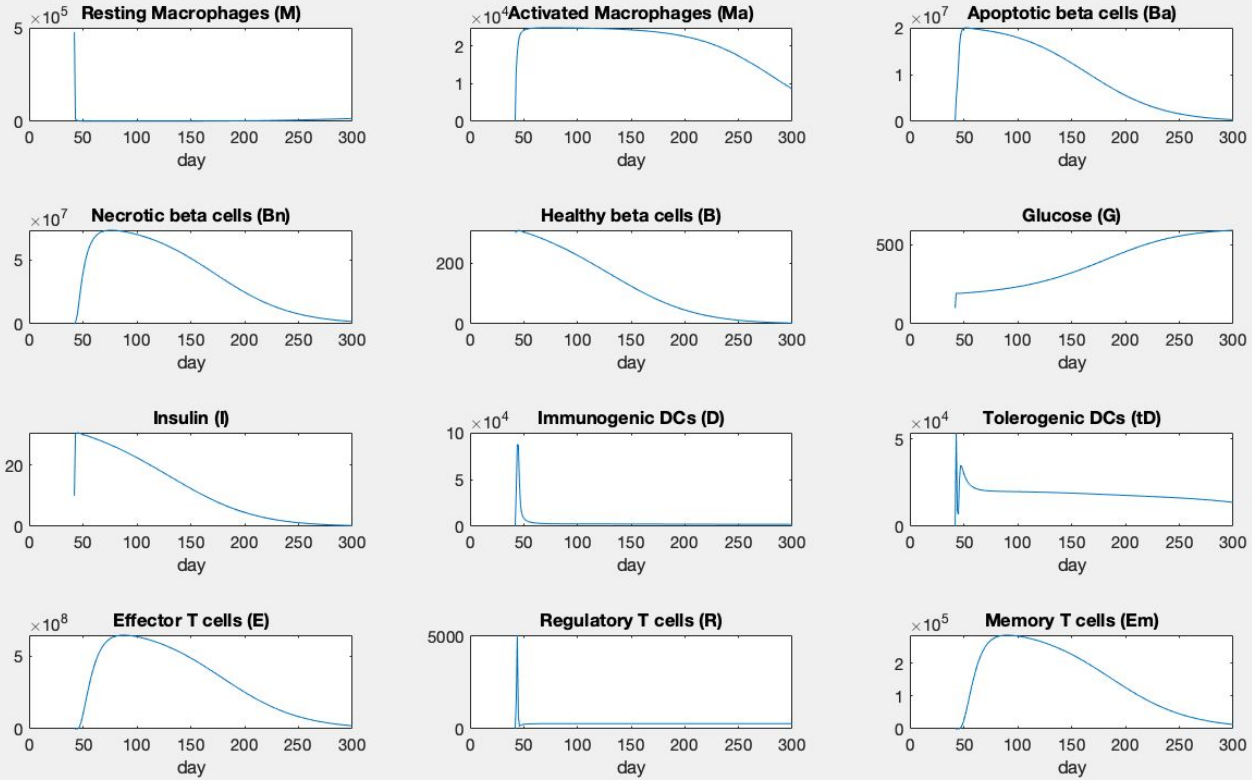
Merging UKFs and MCMC - *results*



Model prediction for mean parameter values



Merging UKFs and MCMC - *results*



Variable predictions

What We've Done This Week

- Created a preliminary outline for the paper
- Worked on writing up descriptions of the predator-prey model, Kalman filtering, and MCMC algorithms working partly from previous write-ups
- Given each other feedback on our writing so far
- Cleaned up some code
- Worked to standardize our code in terms of figures and simulated data - trying to make sure this is the same between Kalman Filtering and MCMC

Outline

T1D$_{final_writeup}$

Christina Catlett, Daniel Shenker, Rachel Wander, Maya Watanabe

June 2020

1 Introduction

1. Current knowledge
   - MCMC/Baycsian approaches are very common in parameterizing biological systems: here are some of the papers/systems that have used it
   - Kalman filters are another method, not as widely used (?): here are some examples
2. Introduce Lotka-Volterra
   - Why are we using this model?
3. Parameterization: MCMC v. Kalman filters
   - What is the goal of parameterization?
   - What are the differences?
   - High-level explanation of each algorithm
   - When would we use one technique over the other?
     - What is required initially? What type of info fits my method?

2 Problem Set-up

- What are the elements of this model
- Equations, general figures etc.
- What does parameterization mean in the context of this model? What are the parameters - consistent naming?
- Where does this data come from?

3 MCMC

3.1 Explanation of theory

3.2 Tutorial

4 Kalman Filters

4.1 Explanation of theory

5 Comparing Results

6 What can we do with both?

- Have noticed when we might want to use one technique over the other?

7 Appendix

- Particle swarm optimization

8 References

Questions

- How much of an introduction do we want to give to each technique - background, common uses, etc.
  - Addendum, how much of this do we want to put in the overall introduction itself
- In the paper, how deep should we go into the specific codes
  - Is it useful to include actual lines of code?
- How much should we go into the trouble-shooting?
  - What if we discovered solutions to these in the T1D model, not Lotka-Volterra
- Is there a page range that we should try to aim for?
- What is our paper trying to say beyond acting as a tutorial?
- Where should we start our parameter guesses? Need justification?
  - “True” parameters from PSO, fmincon parameters
- Role of PSO? (Baseline or comparative)
- What role should artificial data play in the paper?

Next Steps

- Integrate Feedback
- Move on to other sections
  - Introduction
  - Results
- Tighten up structure
  - Section headings, etc.
- Consistent figures across algorithms

Week 8 Presentation

Overview

- Incorporated T1D overview, results
- Feedback given on all techniques
- Generally:
  - Done with technique descriptions, implementation details
  - Working on results, starting comparison
  - Still need intro, discussion, abstract, organization

Updated Outline - Pt 1

1 Introduction

1. Bringing dynamical systems and parametrization together.
2. Current knowledge of the following parametrization techniques
   - MCMC
   - PSO
   - UKF
3. Introduction of the Biological Systems utilized in the paper.
   - Lotka Volterra
   - T1D
   - Why have we chosen these systems?
4. Overview of parametrization.
   - Overall goal.
   - High level explanation of each algorithm.

2 Problem Set-up

1. Lotka Volterra
2. T1D
3. For each, describe elements of the model, include general figures describing them, and discuss the data sources.

Updated Outline - Pt 2

3 MCMC

1. Theoretical background
2. General Tutorial
3. Application to Lotka-Volterra
4. Application to T1D

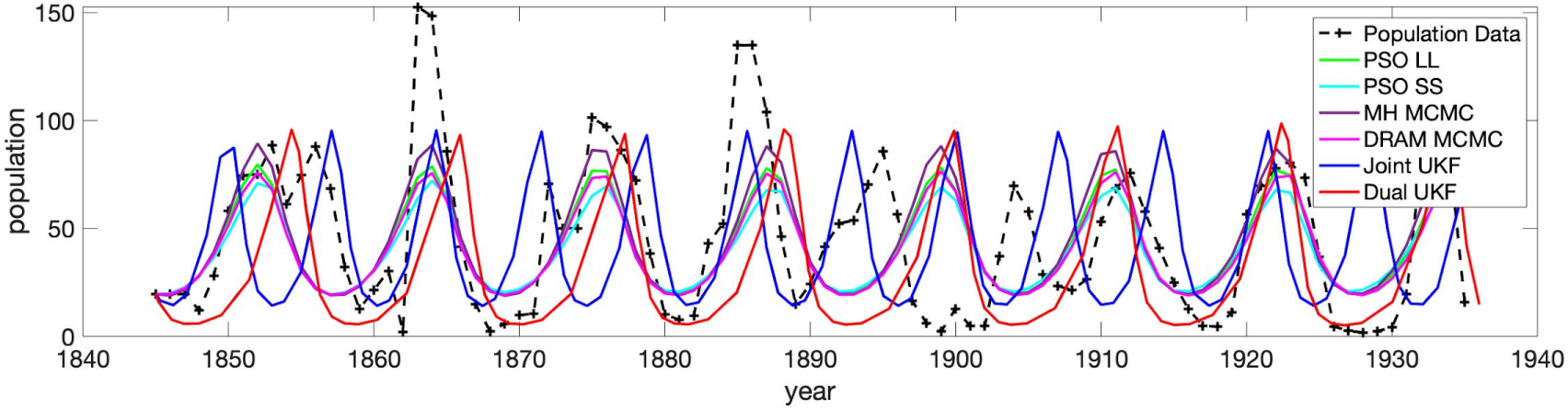
4 Kalman Filter

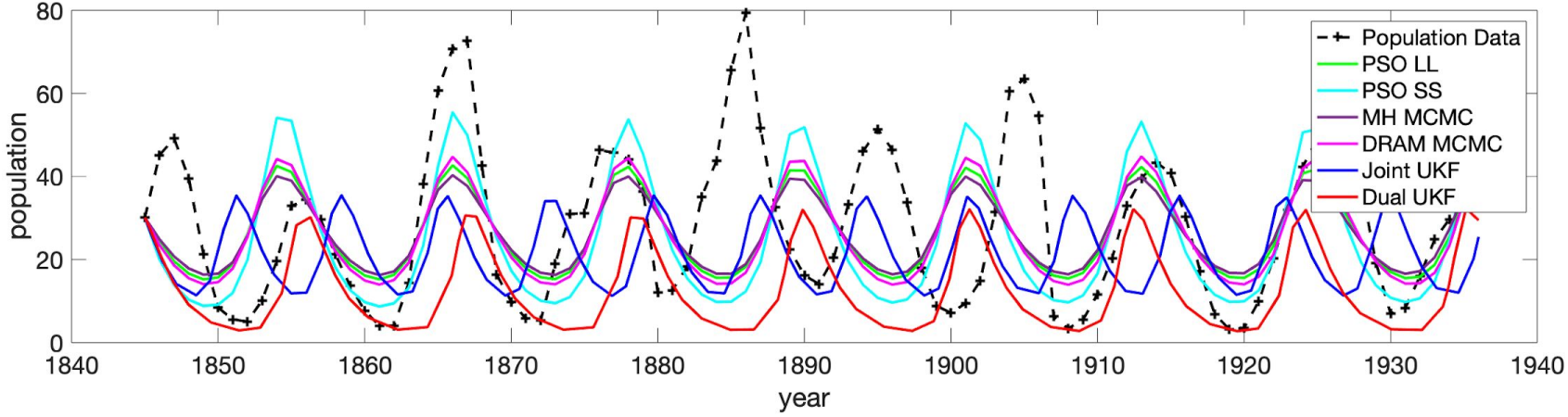
1. Theoretical background
2. General Tutorial
   - Joint
   - Dual
3. Application to Lotka-Volterra
   - Joint
   - Dual
4. Application to T1D
   - Joint
   - Dual

5 Discussion

1. Each algorithm gives overview of their findings
   - MCMC
   - PSO
   - UKFs
2. Comparison of Results
   - Lotka Volterra
   - T1D
     - Population-Level results
     - Individual-Level results
3. Improvements to the algorithms moving forward.
4. Combination of MCMC and UKF algorithms
5. Future Applications

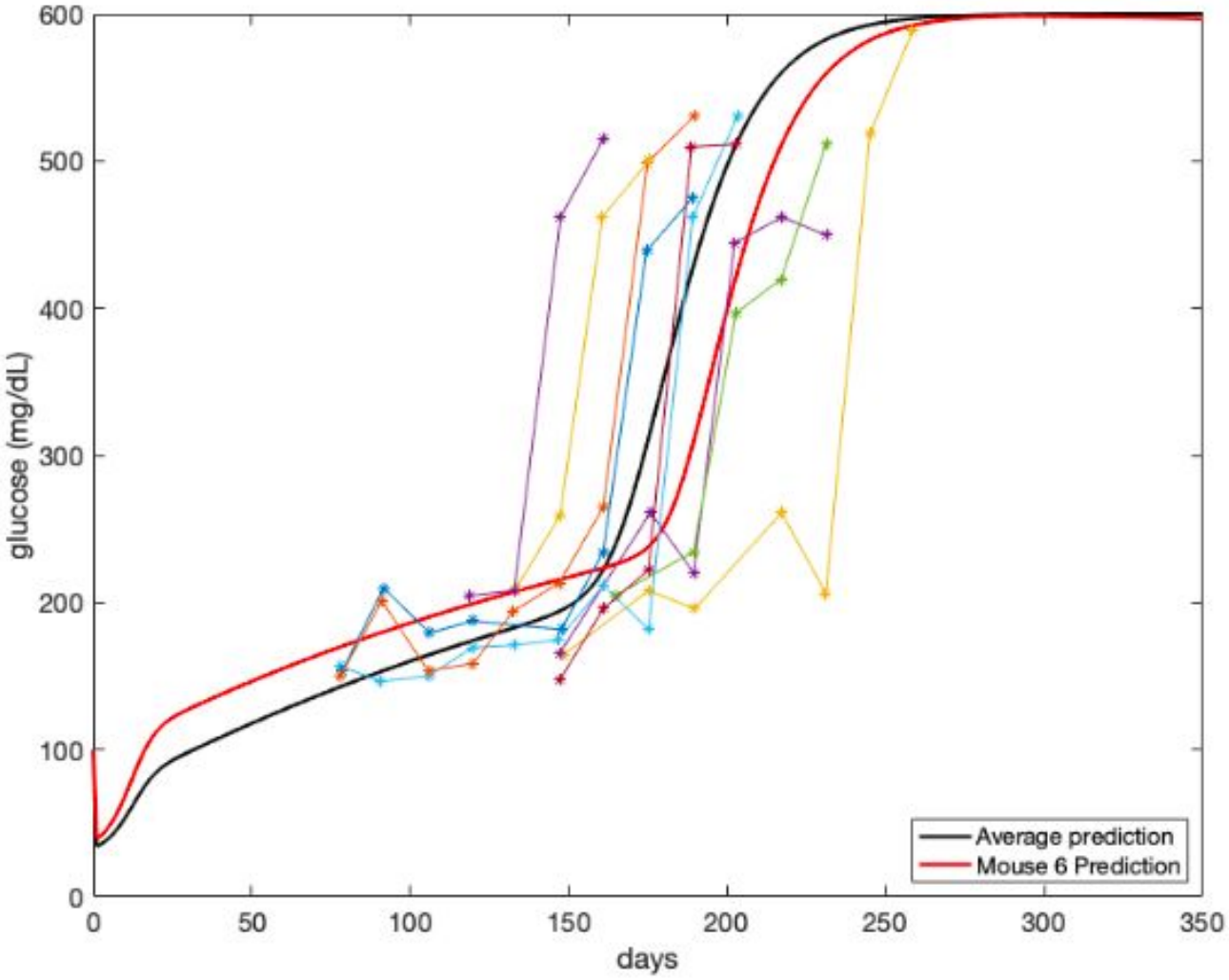
Comparison Results - Lotka-Volterra





MCMC Results

Data	MSE
Averaged	939.6175
Mouse 6	4.4632e+03

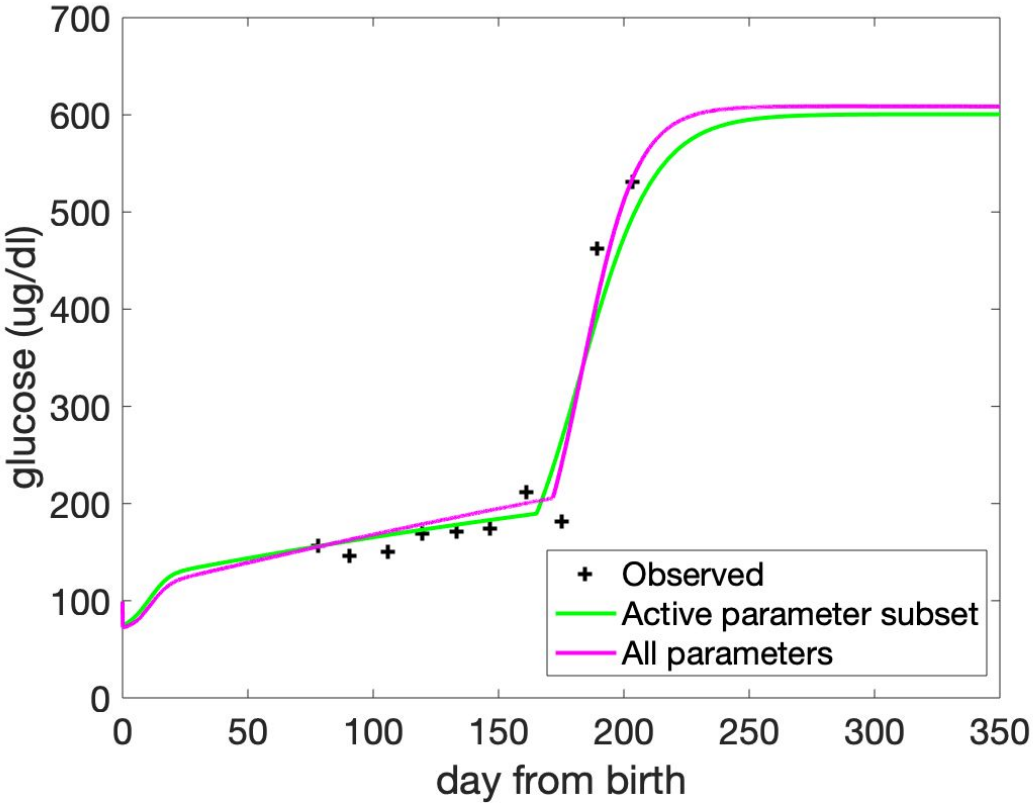


**ICs not parameterized*

PSO Result Pt 1 - Mouse 6

1) Full parameter set
2) ‘Active’ subset from DRAM

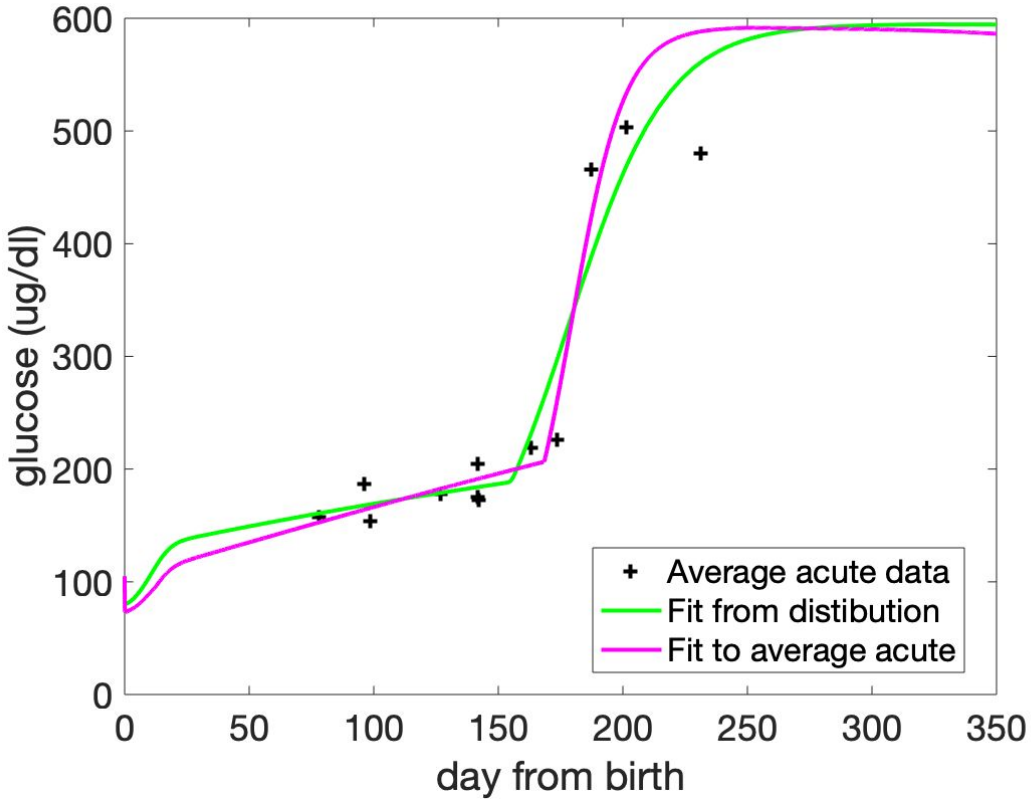
Parameter Subset	MSE
Active	794.5
All	1341.6



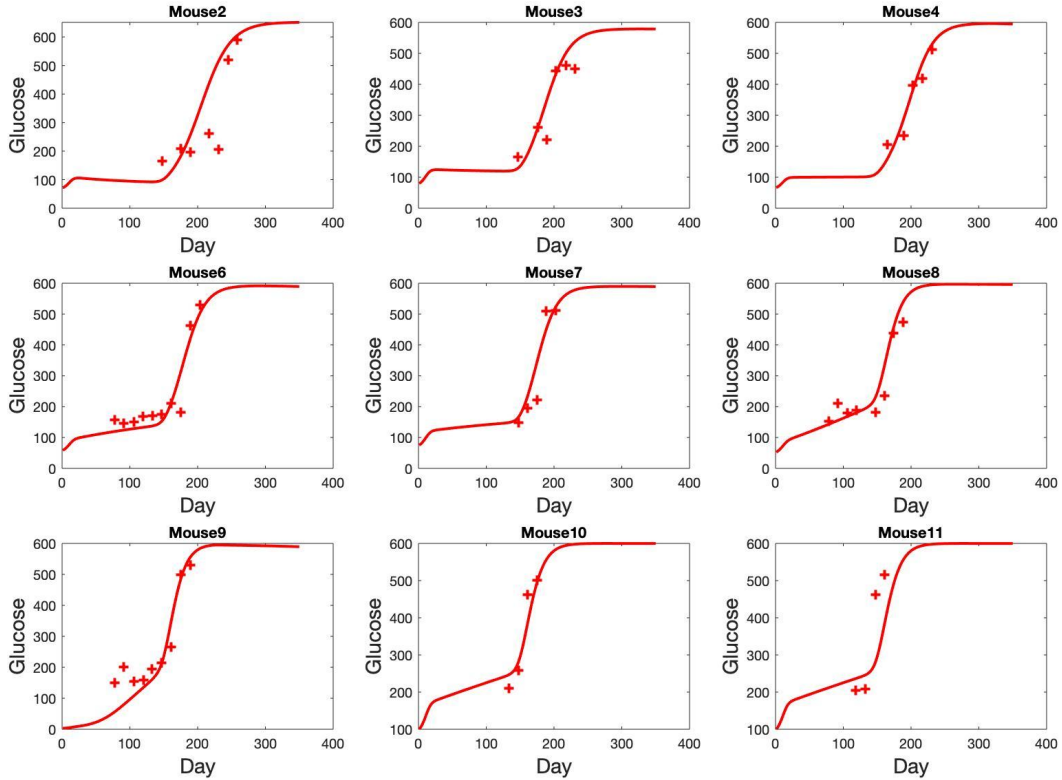
PSO Result Pt 2 - Averaged (acute)

1) Best fit to averaged data
2) UKF distribution-fitting strategy (Fit all individuals, find trends for pop)

Algorithm	MSE
Average acute data	517.6
Distribution-based	1246.9

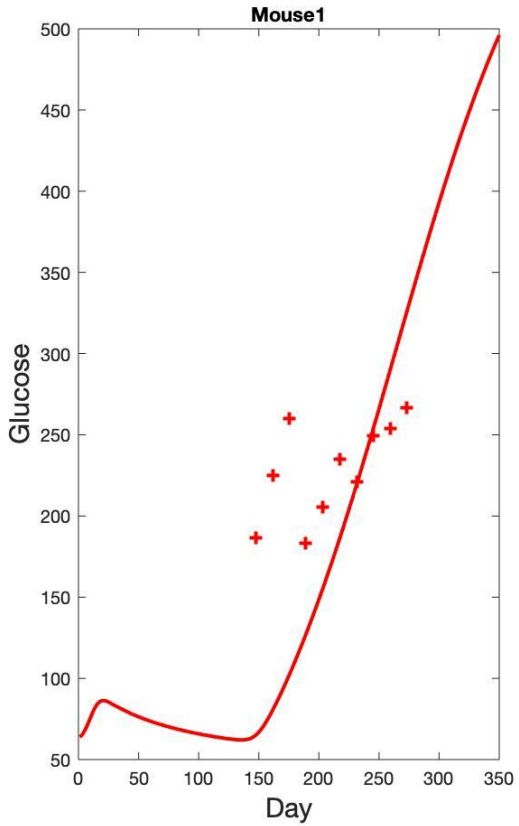


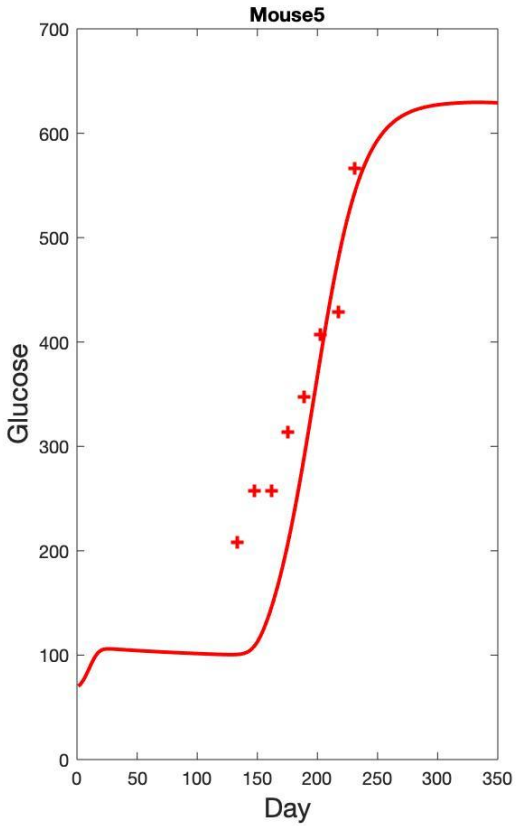
UKF Results - Acute Mice



Mouse Number	MSE
2	1.9902e4
3	4.5462e3
4	1.6649e3
6	2.5194e3
7	4.1531e3
8	2.5665e3
9	3.8981e3
10	2.0596e3
11	1.3104e4

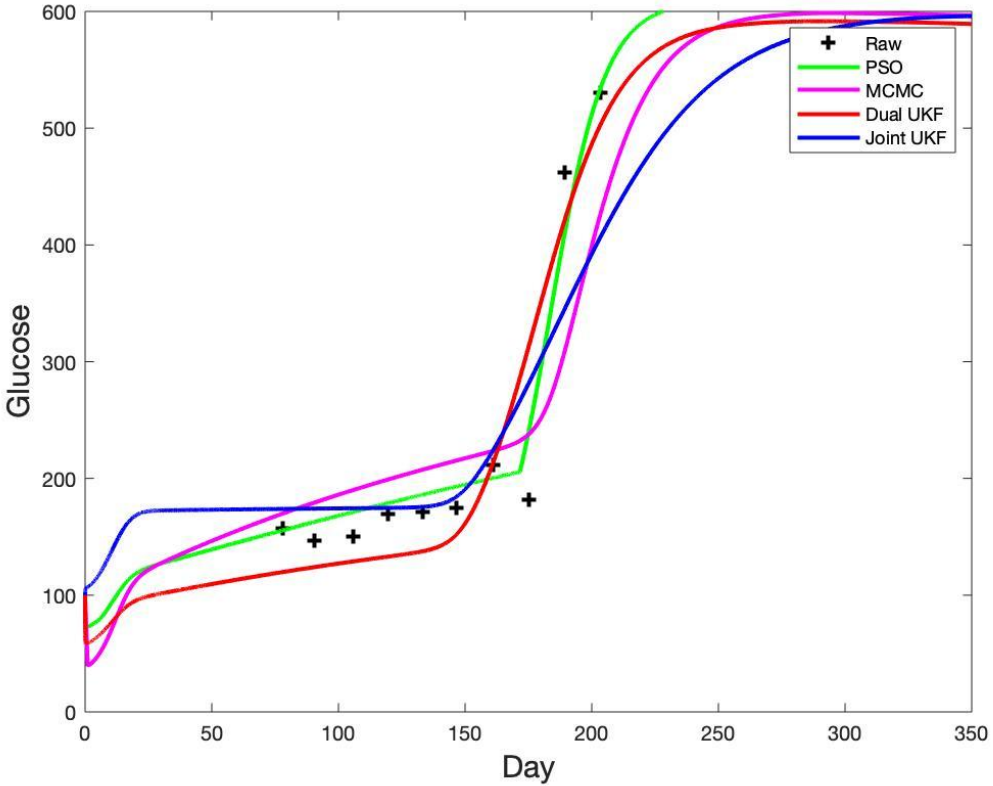
UKF Results - Progressive Mice





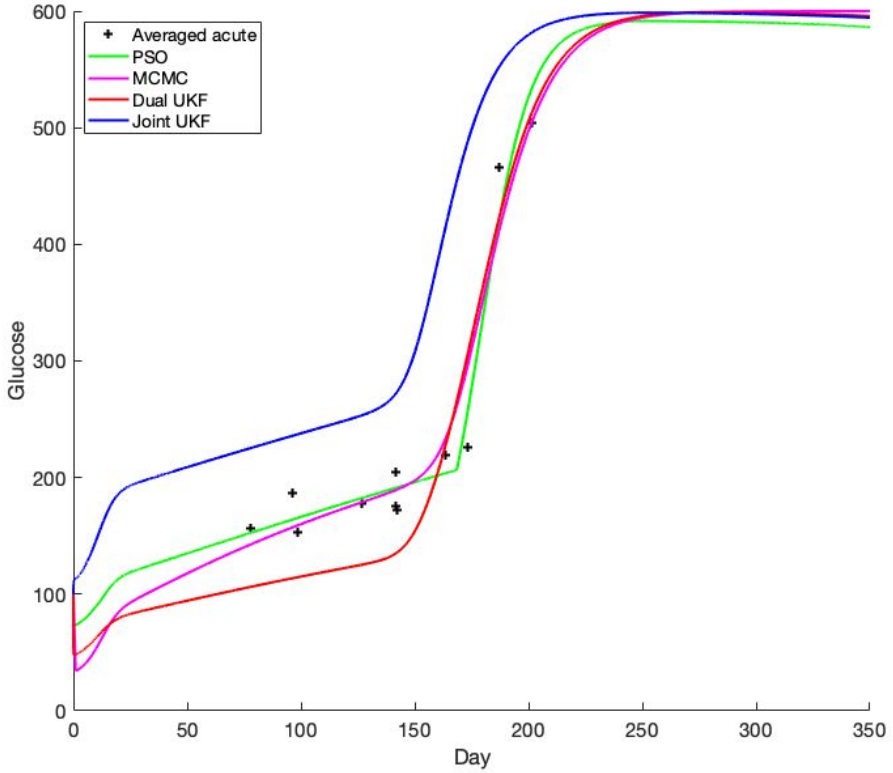
Mouse Number	MSE
1	7.5384e3
5	7.9402e3

Comparison Results - Mouse 6



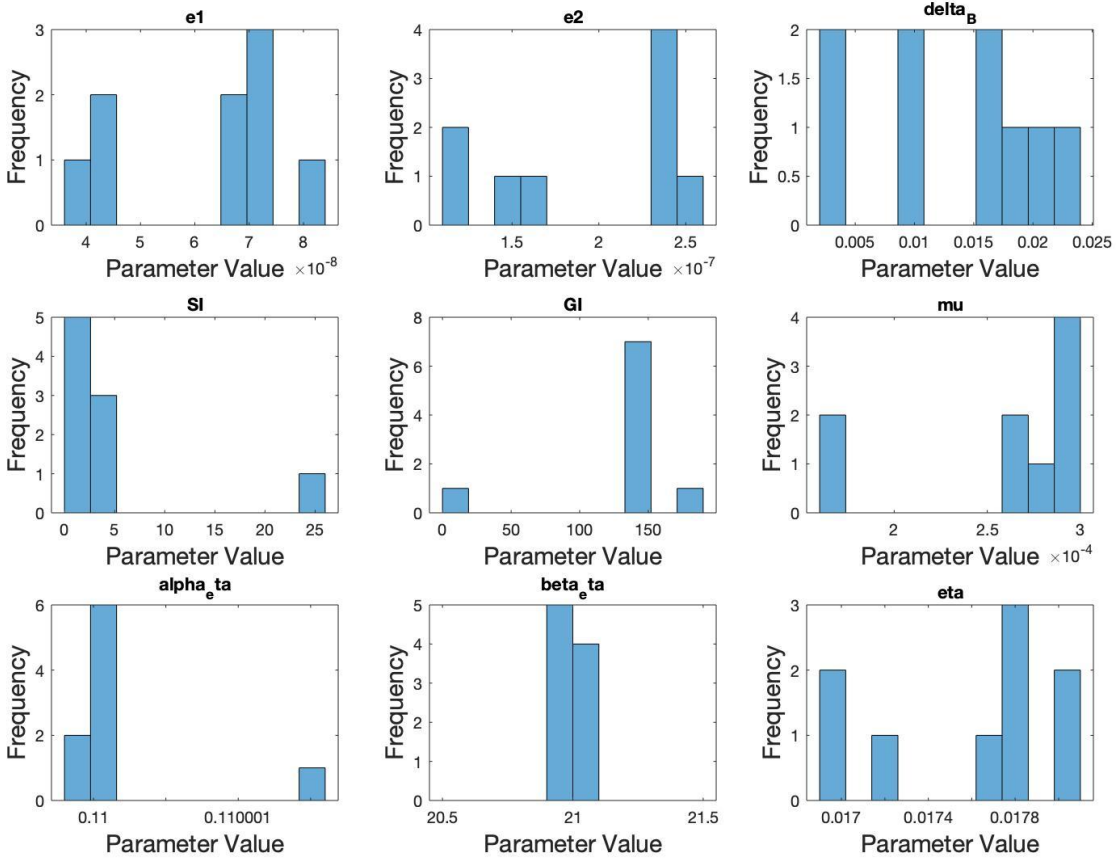
Algorithm	MSE
MCMC	4463.2
PSO	794.5
Dual UKF	2519.4
Joint UKF	4505.4

Comparison Results - Average

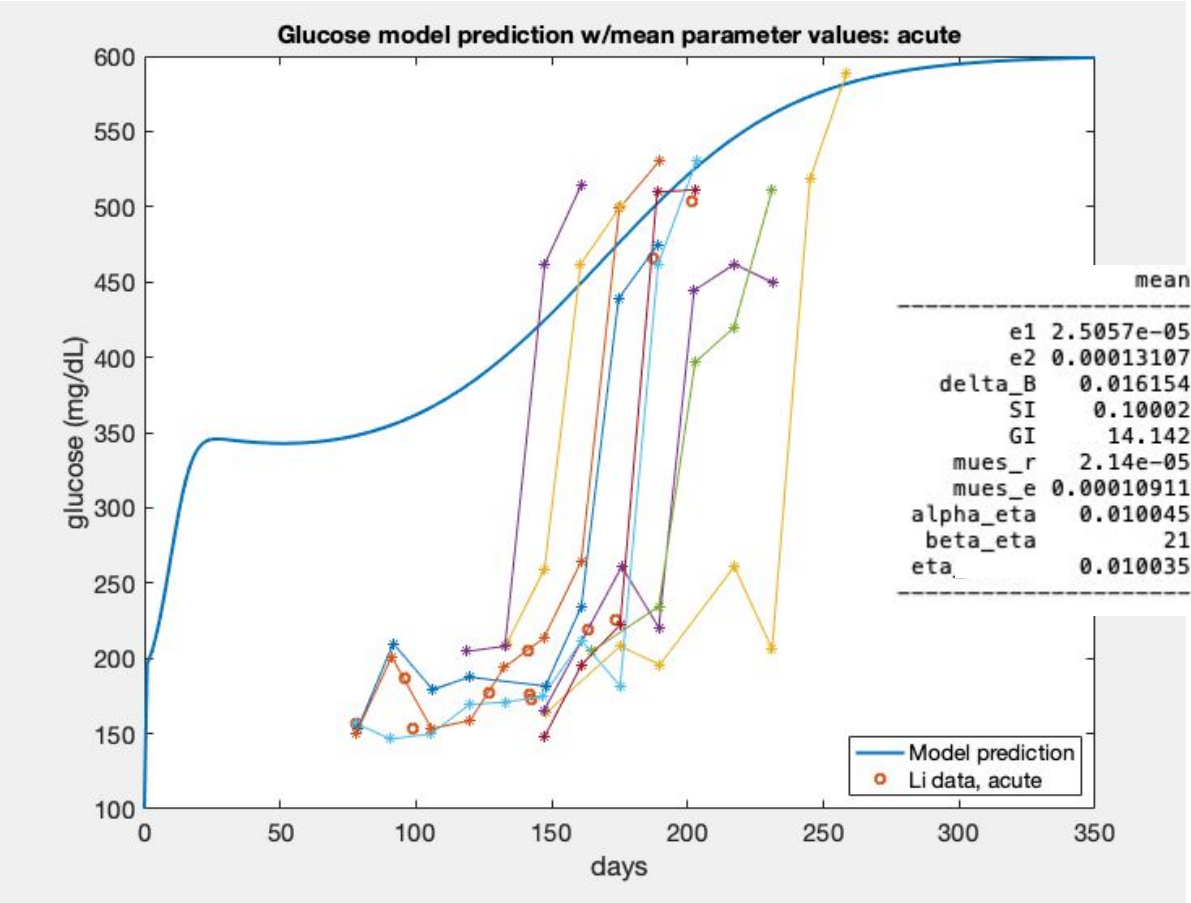


Algorithm	MSE
MCMC	939.6
PSO	517.6
Dual UKF	2607.4
Joint UKF	14962.7

Dual UKF Parameter Distributions



UKF + MCMC results



	mean	std	MC_err	tau	geweke
e1	2.5057e-05	8.8556e-07	1.4126e-07	241.09	0.93701
e2	0.00013107	6.1305e-06	1.2116e-06	368.08	0.85221
delta_B	0.016154	8.4104e-07	1.3261e-07	435.21	0.99997
SI	0.10002	9.4022e-07	1.8766e-07	368.31	0.99997
GI	14.142	3.9808e-06	7.8377e-07	390.34	1
mues_r	2.14e-05	1.1041e-06	2.143e-07	384.4	0.82811
mues_e	0.00010911	4.4417e-06	8.8514e-07	374.76	0.86763
alpha_eta	0.010045	1.9501e-06	1.7387e-07	102	0.99992
beta_eta	21	1.5295e-06	1.9681e-07	252.91	1
eta	0.010035	1.3675e-06	1.2007e-07	84.838	0.99995

MSE score: 38,642.20

Diagram 1 - UKF Code Structure

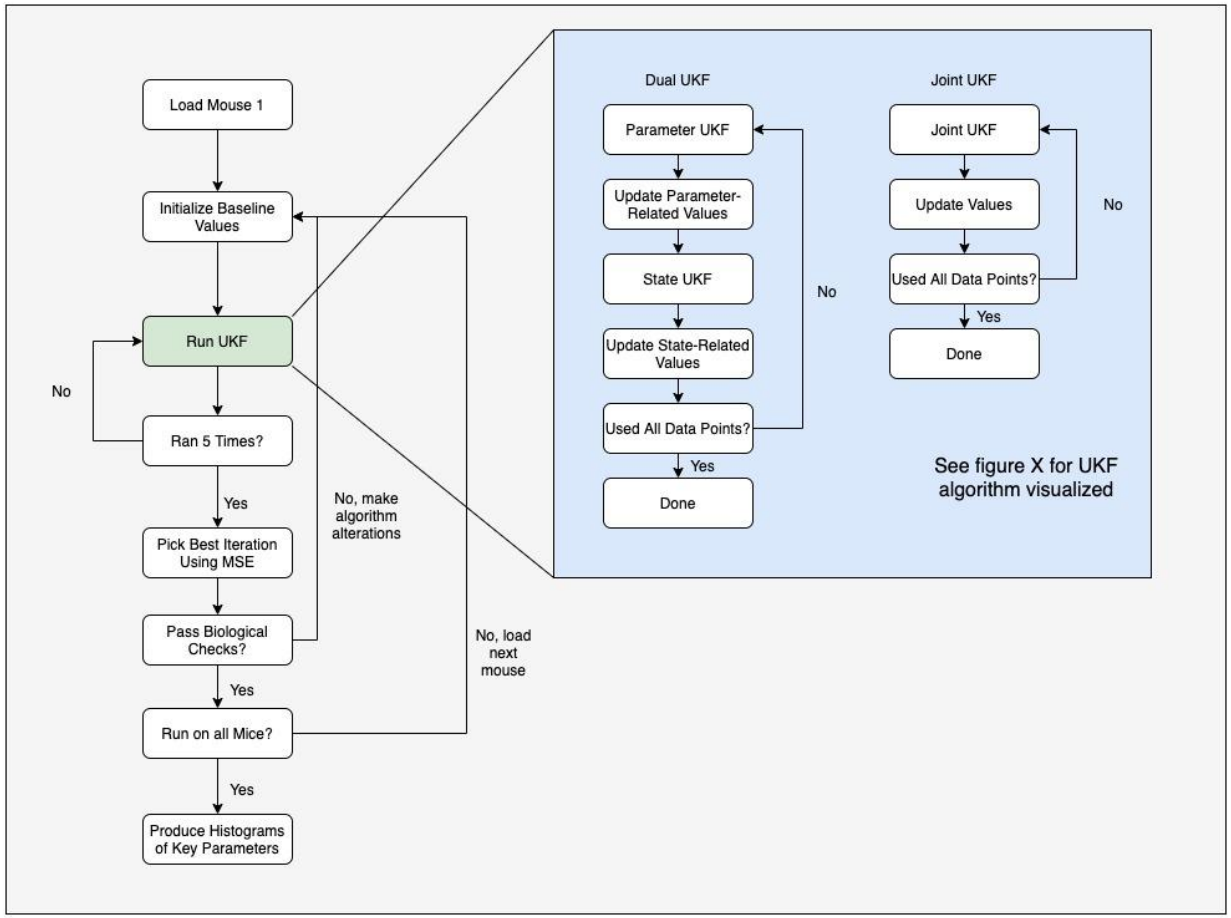
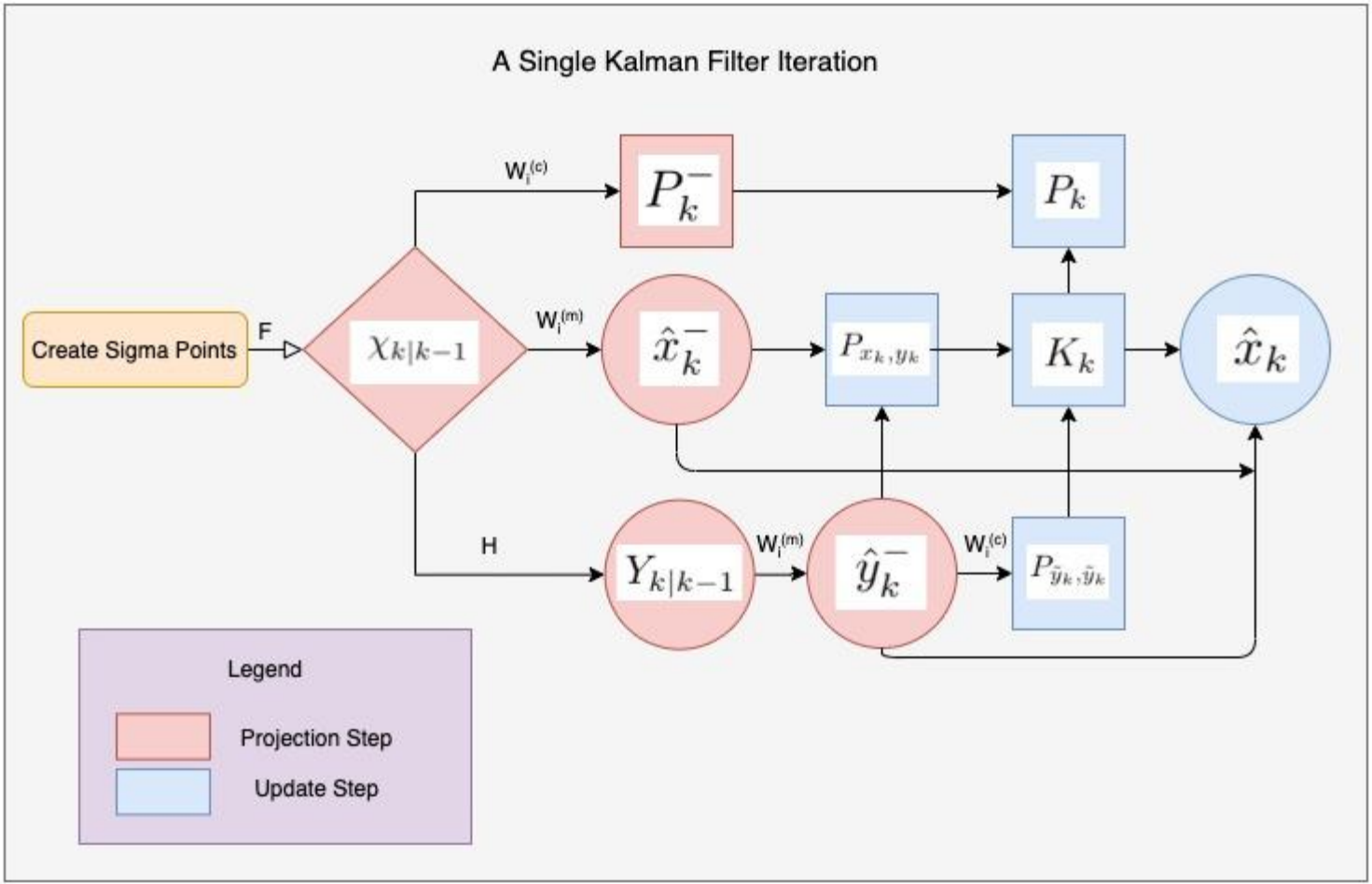


Diagram 2 - 1 UKF Iteration



Questions

- Is there a LaTeX template we should be using?
- Repeat pep talk about motivations?
- Final REU things: Poster? Presentation? Write-up?
- What goes in the Discussion section versus the Results section
  - Given that we have separate sections for Kalman Filters and MCMC, where is the best place to introduce a comparison of result between the two?
- How much introduction of each system/technique do we need in the introduction
  - How much background knowledge can we assume our readers have
  - What is the best way to structure the introduction anyways
- How much do we need to repeat in the Discussion section?

This Week

- Get feedback from you all
- Write comparison section, discussion, intro
- Compile parts into whole
- Edit & feedback

Week 9 Presentation

What We Did This Week

- Updated Figures
  - Updated to RMSE
- Wrote Discussion
- Wrote Introduction
- Added Citations
- Compiled everything into a whole

Results Section Figures

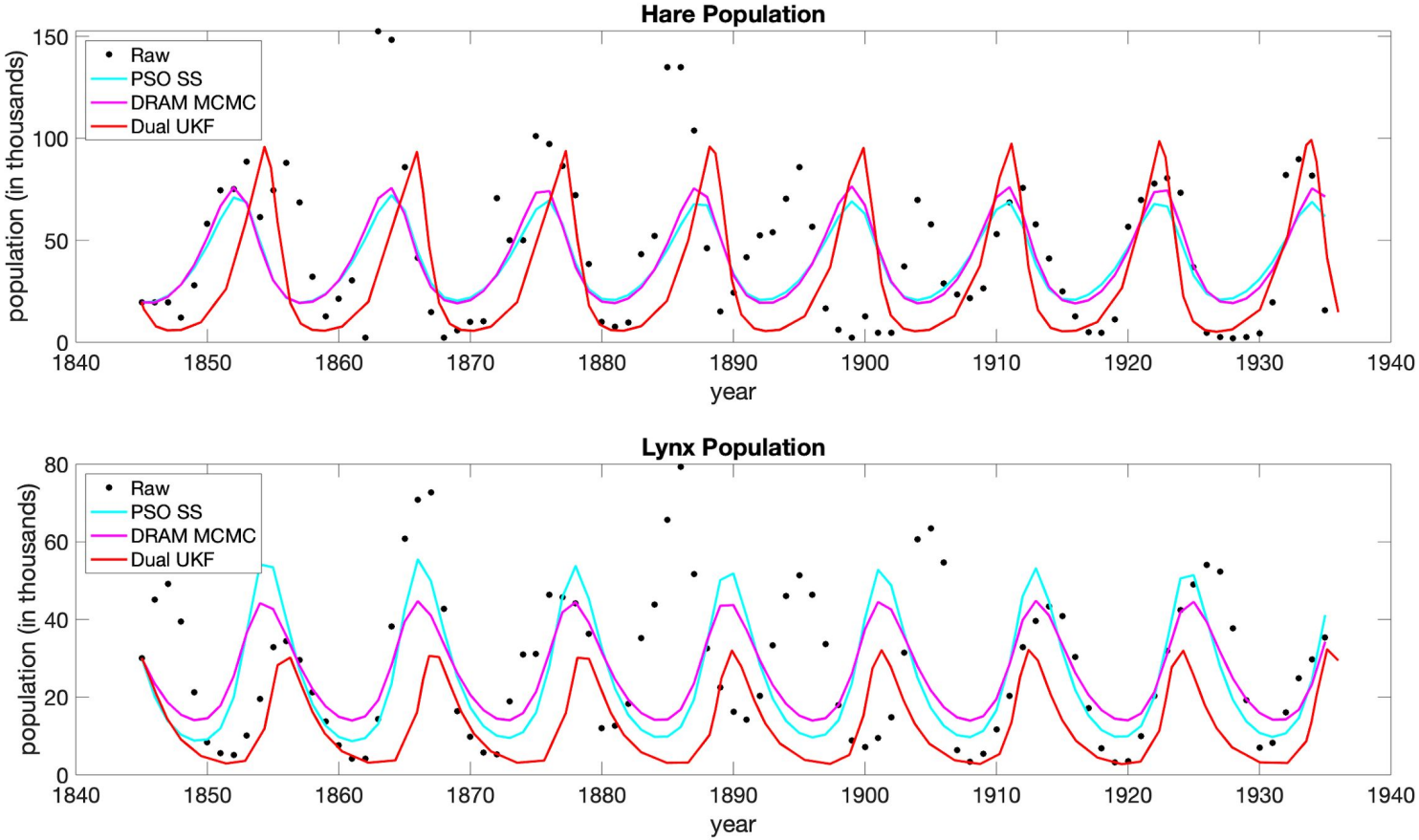
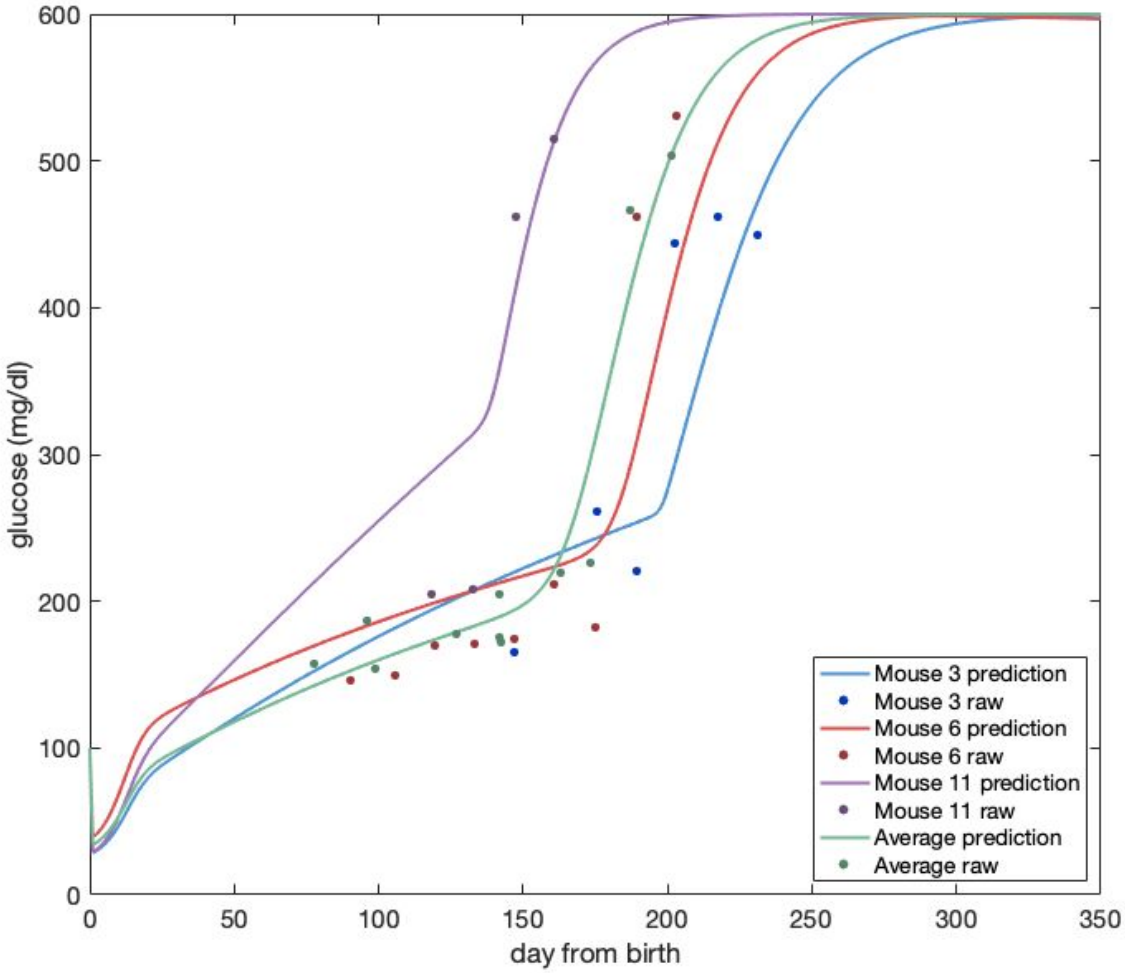


Figure 1: Plot of raw hare-lynx data alongside best performing techniques from each class explored. All three methods miscalculate the initial decrease in lynx population. While PSO visually captures the peaks in the lynx population most effectively, UKF captures the troughs. The same trends are not observed in the hares, where UKF provides the best fit for both the peaks and troughs.

MCMC update


600
500
400
300
200
100
0
glucose (mg/dl)
0
50
100
150
200
250
300
350
day from birth
Mouse 3 prediction
Mouse 3 raw
Mouse 6 prediction
Mouse 6 raw
Mouse 11 prediction
Mouse 11 raw
Average prediction
Average raw

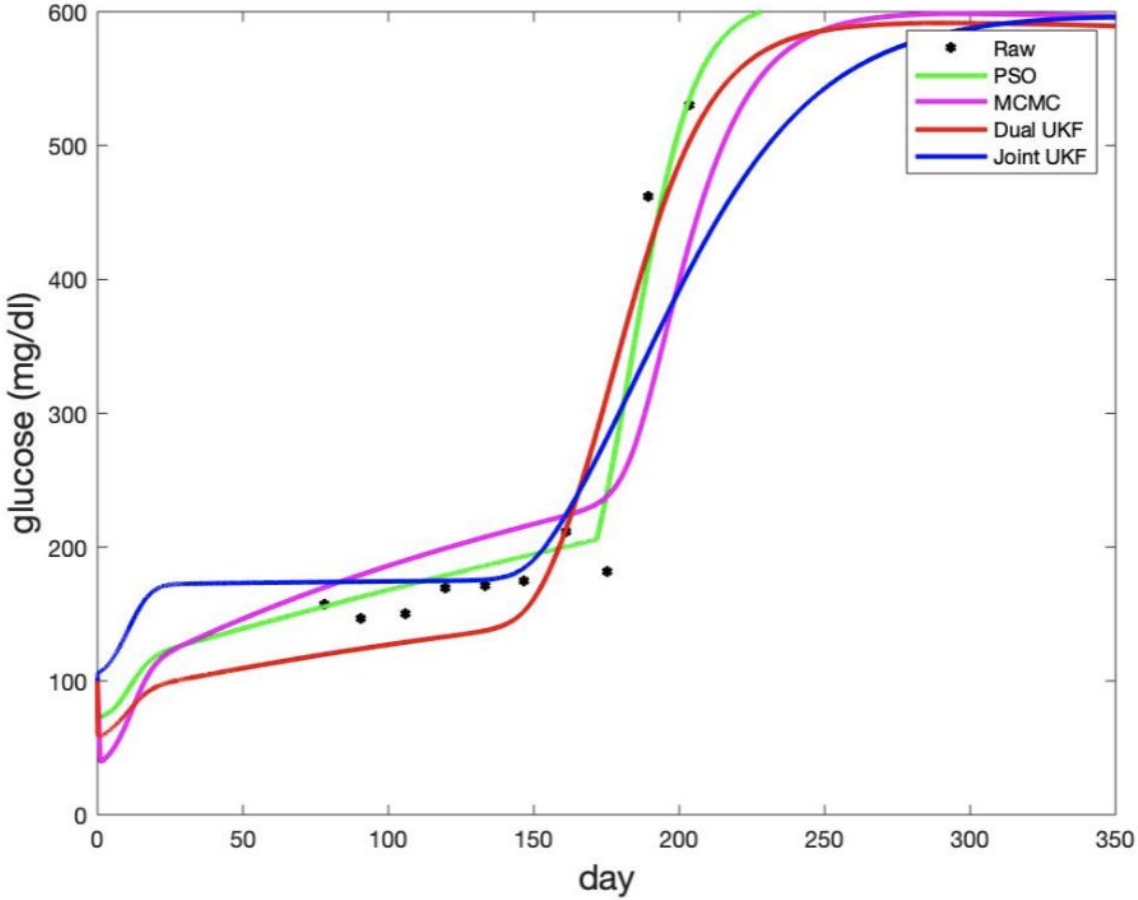


Figure 1: Plot of raw mouse 6 data overlayed with MCMC, UKF, and PSO fits. It is evident that there is range of success on the individual mouse, with PSO appearing to perform best.

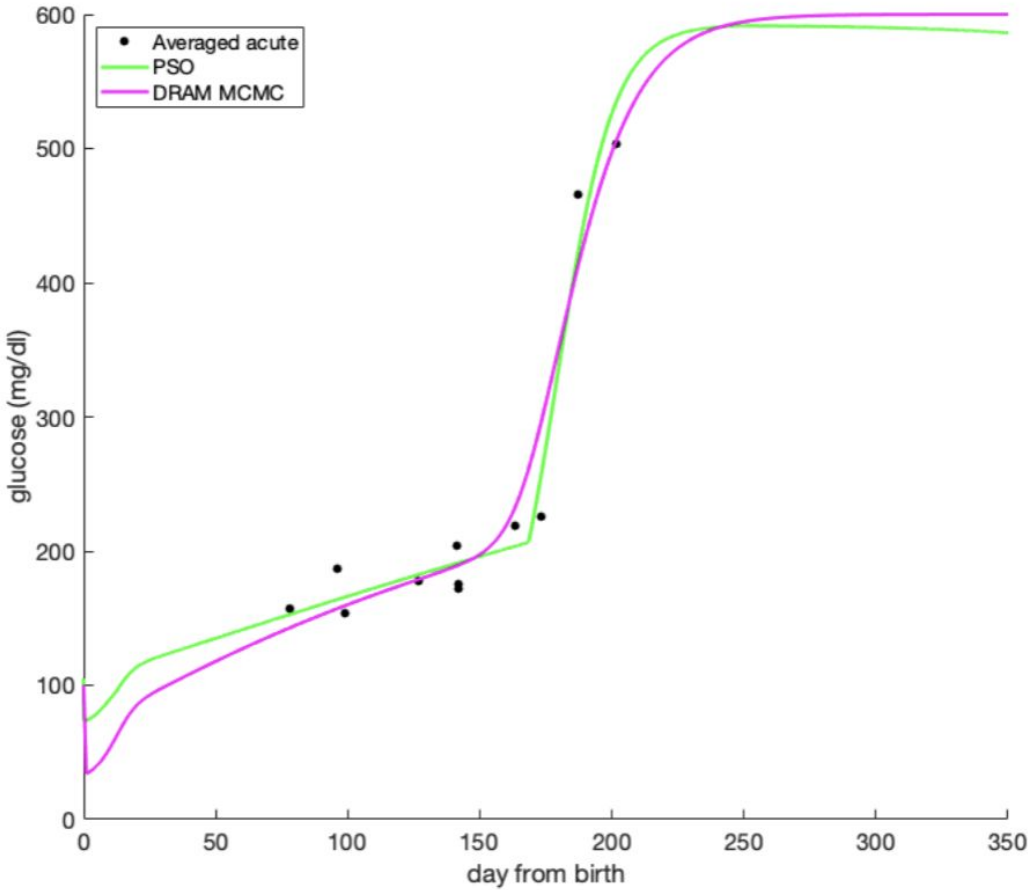


Figure 2: Plot of averaged acute Li data overlayed with PSO and DRAM MCMC fits. This figure illustrates the results of averaging data then fitting.

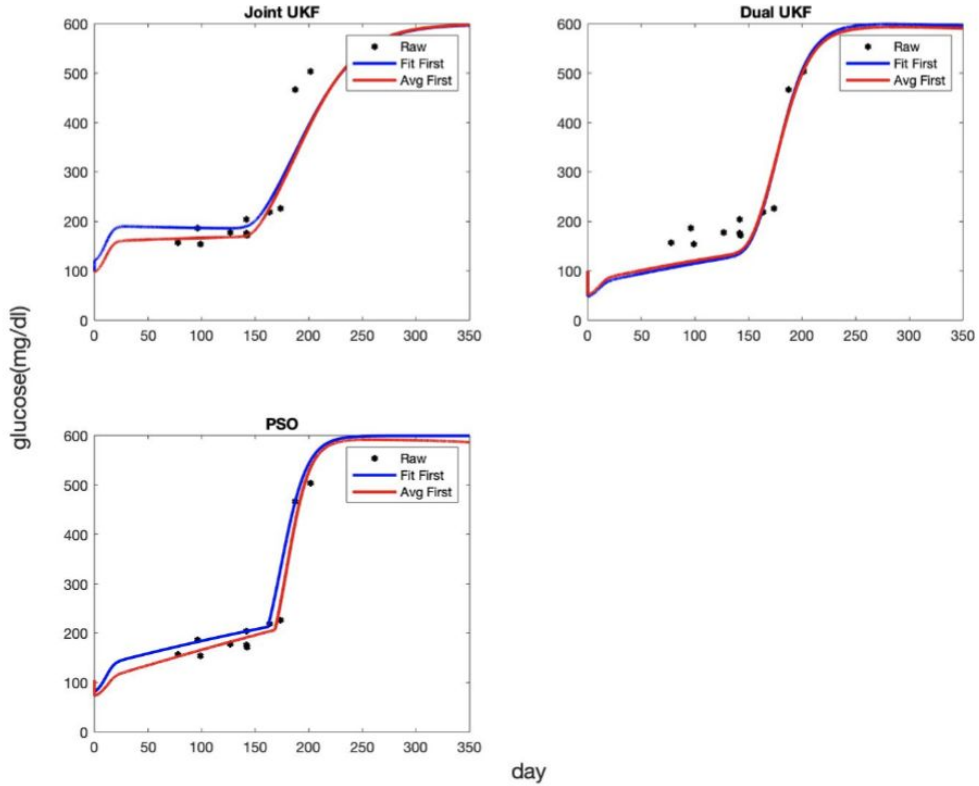


Figure 3: Plots of using Joint and Dual UKFs, as well as PSO, to fit to averaged Li data by both running the algorithms directly on this dataset as well as by averaging parameters from individual fits instead of training on the actual averaged dataset. It is evident that averaging first produces slightly better results for PSO and the Joint UKF. For the Dual UKF, the fits are so similar it is hard to see which is better.

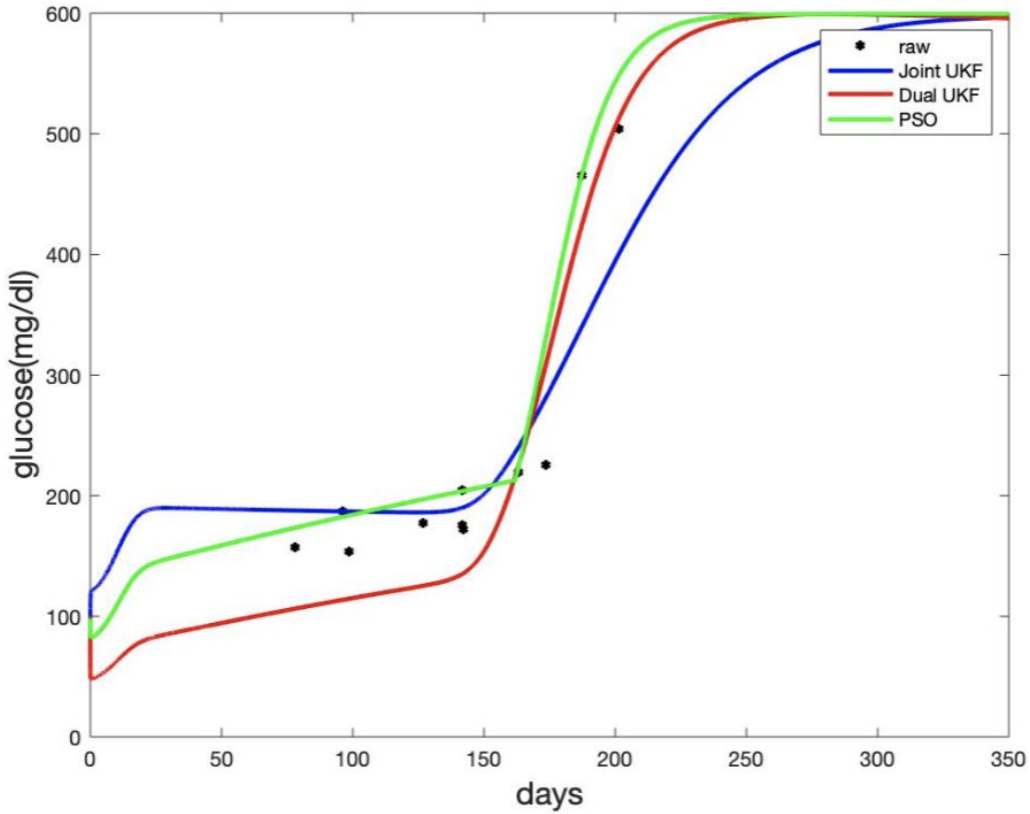


Figure 4: Plot of using Joint UKF, Dual UKF, and PSO to fit to averaged Li data by averaging parameters from individual fits instead of training on the actual averaged dataset. It is evident that the PSO once again performs best, whereas the Joint UKF captures the early behavior accurately and then struggles and the Dual UKF struggles early but captures the end behavior.

Questions

- For final presentation, should we be more results or discussion focused?
- Due to length, for peer editing, what sections should we prioritize?
- Do you want us to do a virtual poster (Prof Martinoosi said it was at the professor's discretion)
- When doing a full read through of the draft, what should we be looking for?

Next Steps

- Today: individual full read through of the draft
- Tuesday: discuss our notes from the individual read through and receive peer-review from other REU teams
- Wednesday: Implementing our notes and peer review notes
- Thursday: Receive feedback from rest of T1D team
- Friday: Implement Thursday feedback and continue editing / create final presentation