Personalizing healthcare: a pathway for using data and data science to make clinically and physiologically impactful personalized predictions

David J. Albers

6/16/2019

Medicine is about prediction, and verifiable prediction requires a model and data...

Medical practice is a about prediction

When managing health, clinicians and hopefully people in general intervene in ways predicted to help more than hurt — clinicians generally weigh the consequences of probability of helping against hurting conditioned on the consequences of interventions.

Predictions are generally based on a model and data; but what constitutes a model can be very broadly defined.

Data-driven forecasting and personalized medical care

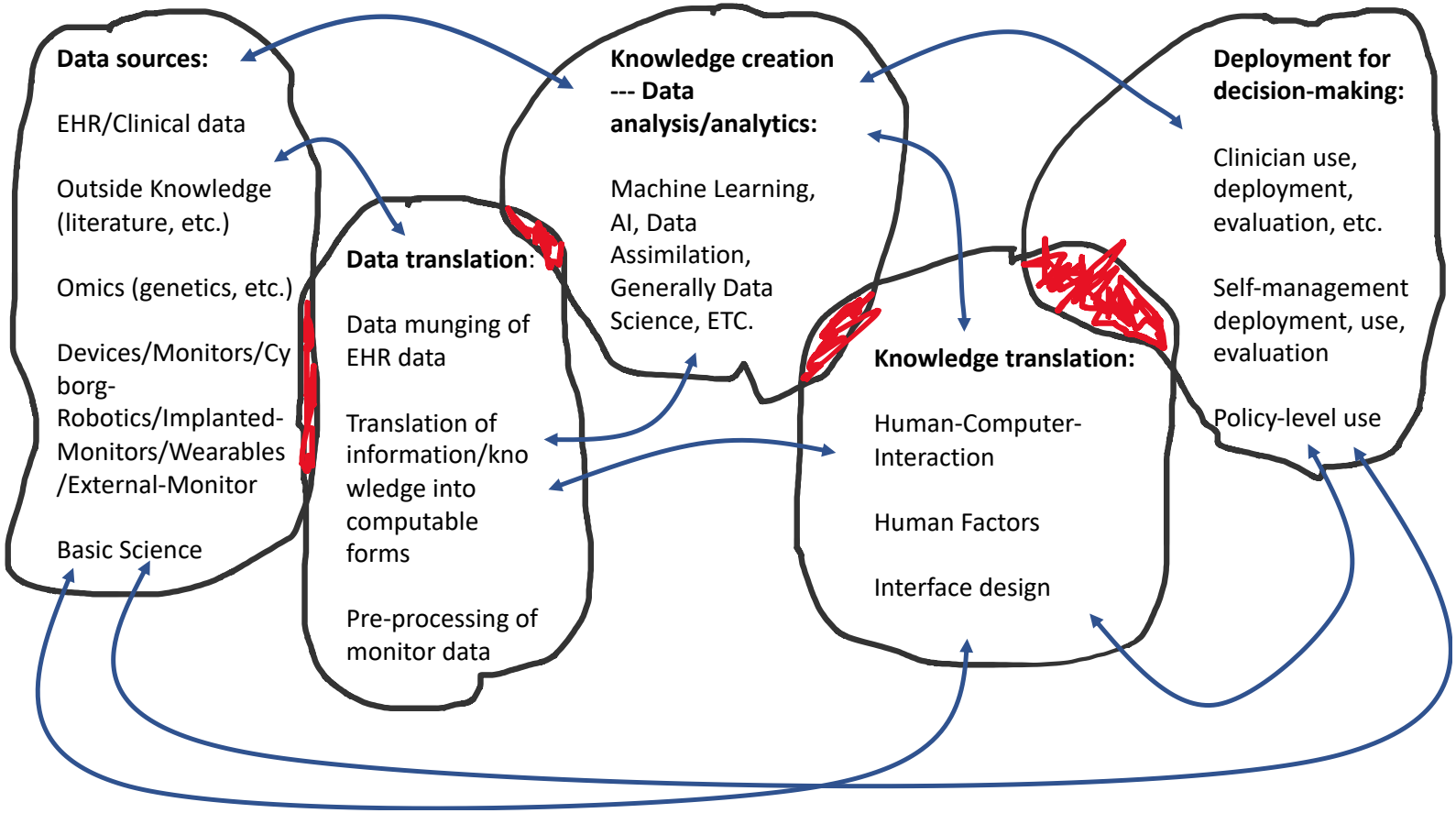
Integrating model-based prediction with clinical data is *the pathway to personalizing medicine*, but:

data-driven prediction 'in your head' is difficult, getting prediction right is important, adapting mathematics to live within clinical constraints is difficult and requires special innovation as does the integration of prediction into the clinical workflow because intervention must be correctable/undo-able.

Clinical data light the pathway, but the pathway is rough:

- clinical data are the population-scale data on human physiology data we have and will ever have;
- *but, they are collected administering care*;
- because of this, they present the a real avenue to have an impact, evaluation;
- but, clinical data are generated by many coexisting processes — **the generating function is a complex convolution of physiology and the health care process**.
- and, clinical data are sparse and missing because they are invasive, expensive, sometimes dangerous to collect.

The plumbing-schematic for personalizing medicine



Personalized prediction an medical context: 5 elements

Prediction in a medical setting has five elements:

1. Data;
2. Models of clinical data;
3. Inference machinery used to select the best model of the data;
4. Knowledge/forecast delivery methodology;
5. Context within the clinical setting and workflow.

And these elements exist within the overall context of clinical decision-making and understanding.

Forecast delivery and decision-making: ICU example

In a clinical setting prediction has a decision point and it constrains and defines what a useful prediction might look like:

NewYork-Presbyterian Hospital

Rev 12/17. MBA 2/18.

Goal BG 140-180

Insulin Infusion: Adult ICU Protocol

NOT for DKA/HYPERGLYCEMIC HYPEROSMOLAR NON-KETOTIC SYNDROME (HHS)

1. **For blood glucose (BG) above 180 mg/dL**, start regular insulin infusion: 100 units/100 mL NS (1 unit = 1 mL) per table

INITIAL INFUSION	BG (mg/dL)	180-216	217-270	271-324	>324
	Insulin Rate	1 unit/hour	2 units/hour	3 units/hour	4 units/hour

2. Check BG every 1 hour until stable (140-180 mg/dL x 2 consecutive readings), then check every 2 hours if insulin rate is <10 units/hour and there are no changes in feedings, IV glucose, vasopressors or glucocorticoids. Resume every 1 hour monitoring if any of above parameters change.
3. Titrate insulin infusion to achieve a **goal BG of 140-180 mg/dL**. Select row according to current BG level. Determine rate of change from prior BG result. Match to column of current insulin rate. Follow recommendations to corresponding cell. Round to nearest 0.5 units/hour.
4. If nutrition (e.g., tube feeding) is held or discontinued, notify prescriber to either hold insulin infusion or add/increase dextrose-containing IV order.
5. Any deviation from protocol requires prescriber order. When the subsequent BG returns to goal, may resume following the protocol from the current insulin rate.

		Current Insulin Rate (unit/hour)					
Current BG Level (mg/dL)	**BG change from previous BG (mg/dL)**	**≤ 3**	**3.5-6**	**6.5-9.5**	**10-14.5**	**15-19.5**	**≥ 20**
< 40		STOP INSULIN INFUSION (EXCEPT for patients with Type 1 DM, see footnote[a]) Give 25 G (50 mL) D50 IVP Q 15 min, repeat PRN for BG < 40 mg/dL. Notify prescriber Re-check BG Q 15 minutes until BG > 80 mg/dL then Q 1 hour x 1, then Q 2 hours. When BG > 100 mg/dL, restart at 50% previous rate unless BG change from previous BG is > 100 mg/dL, then restart per Initial Infusion Table					
40-69		As above (refer to row < 40 mg/dL), EXCEPT give 12.5 G (25 mL) D50 IVP Q 15 min PRN BG 40 -69 mg/dL					
70-99		STOP INSULIN INFUSION(EXCEPT for patients with Type 1 DM, see footnote[a]) for 1 hour Check BG Q 1 h x 1, then Q 2 h. When BG > 100 mg/dL, restart at 50% previous rate unless BG change from previous BG is > 180 mg/dL, then restart per Initial Infusion Table					
100-139		★Decrease rate by 50%					
140-180 (GOAL)	No Δ. Decreased by 1-20 or Any Increase	★Continue same rate (If 2 consecutive BG levels are between 140-180 mg/dL, continue at the same rate and decrease BG checks to Q 2 hours)					
	Decreased by ≥ 21	★Decrease by 1	★Decrease by 2	★Decrease by 3	★Decrease by 4	★Decrease by 6	★ Decrease by 50%
181-251	Decreased by ≤ 20 or Any Increase	Increase by 0.5	Increase by 1	Increase by 1.5	Increase by 2	Increase by 2.5	Increase by 3
	Decreased by 21-49	★Continue same rate					
	Decreased by ≥ 50	★Decrease by 1	★Decrease by 2	★Decrease by 3	★Decrease by 4	★Decrease by 6	★ Decrease by 50%
≥ 252	No Δ, Decreased by 1-20 or Any Increase	Increase by 2	Increase by 2	Increase by 3	Increase by 4	Increase by 6	Increase by 8
	Decreased by 21-75	Continue same rate					
	Decreased by ≥76	Hold x 1 hour, then decrease rate as indicated:					
		Decrease by 1	Decrease by 2	Decrease by 3	Decrease by 4	Decrease by 6	Decrease by 50%
Notify prescriber if BG > 400 mg/dL x 2 or insulin rate > 10 units/hour If adjustment recommendations result in insulin rate of ≤ 0 units/h, hold insulin (If Type 1 DM reduce insulin to 0.5 unit/h), check BG Q 2 hours and resume per Initial Infusion Table							

[a]For patients with Type 1 DM: DO NOT STOP INSULIN (Continue glucose-containing fluid), reduce insulin to 0.5 unit/hour and notify prescriber to write orders to increase glucose intake. Δ = Change

Note (★): For patients when insulin resistance is suspected (e.g., insulin rates > 10 units/h, high catecholamine needs, typically postoperative), decrease insulin rate to 0.5 unit/h and notify prescriber

Data: clinical data and the health care process

Clinical data are collected over the course of delivering/performing care and this induces complexity because **their collection process is itself *is a process that is dependent on the setting***:

- sometimes data are present independent of the heath state...
- sometimes they are present because of the health state...
- sometimes data are present because of a protocol...

Clinical data are influenced by the **health care process.**

Examples:

1. glucose and nutrition documentation in an ICU;
2. nursing documentation patterns.

Models: modeling clinical data, the first half of the ML/AI or knowledge generation and synthesis

Data contain relationships between measurements and:

- other features/variables/measurements;
- outcomes;
- documentation behaviors...

and we represent these relationships with **models/functions...**

Machinery for representing of data with models/functions has two primary components:

1. the model/function space from which we select a model to represent data;
2. the inference machinery used to estimate (pick!) the model, or synchronize the model to data.

Model/function spaces come in three flavors:

1. parametric functions/models;
2. non-parametric functions/models;
3. parameterized-knowledge-driven functions/models.

Model **Flexibility vs Fidelity**

Clinical data are:

1. sparse;
2. nonstationary—patient health states change;
3. correlated with patient state.
4. etc., generally complex.

Model functional flexibility implies *both* the ability to accurately estimate data *and* the number of choices required to select the best model for the data.

Example, grape fidelity and choice: **choosing which grapes you want from among** 2 **versus** 10^6 **types of grapes in a grocery store.**

Generally, fidelity of a model increases with flexibility — but the picture isn't exactly this simple.

Relative to clinical decision support we want to *select the right fidelity for accurate decision-making* while *minimizing the flexibility that induces inference complexities.*

Model category and related fidelity and flexibility

Parametric models: low-fi and low-flexibility.

Non-parametric models: high-fi and infinite-flexibility.

Parameterized-knowledge-driven models: low to medium-fi and limited-flexibility.

Parametric models or function spaces

Some examples:

1. Linear functions — two parameters used to represent the data, a mean (slope) and an intercept.
2. Gaussian distribution — two parameters to represent the data a mean and a variance.

Note: parametric means that the model is *completely* defined/specified by a fixed set of parameters, e.g., (mean, variance).

These models:

- have **limited flexibility** — *these functions often cannot estimate data or a system very well...*
- are **easy to understand/interpret**;
- require **very little data** to estimate/pick.

Non-parametric models or function spaces

Some examples:

1. polynomials ($\sum_{i=1}^{N} a_i x^i$);
2. artificial neural networks — i.e., deep learning;
3. Gaussian process models;

Universal function spaces, meaning there are theorems that state:

given enough terms/units these functions can approximate almost any function to any degree of accuracy desired.

Equivalency: these spaces or methods are *theoretically equivalent*—they differ only in their explicit representation of the data—their functional form (explicit equation)—and the methods used to pick which collection of functions *best represents the data.*

These models:

- have **infinite flexibility** — *these functions—provably—are capable of estimate data or a system **perfectly***;
- are **difficult to understand/interpret**;
- require **tremendous amounts of data and energy** to estimate/pick;
- are generally harder to estimate/pick;
- the selection is not particularly unique or important — imagine selecting the type of grapes you want it eat if you have 100M choices.

...infinite flexibility comes with a cost....

Parameterized-knowledge-driven models or function spaces

Mathematical models of mechanistic physiology — they represent the *mathematization of our knowledge of the underlying system*

An example: the **Hodgkin-Huxley** model of neurons.

These models are:

- more complex than, but neither a sub- nor super-set of the generic parameterized families—e.g., linear functions.
- less complex than, **but are a subset of**, the non-parametric functions.

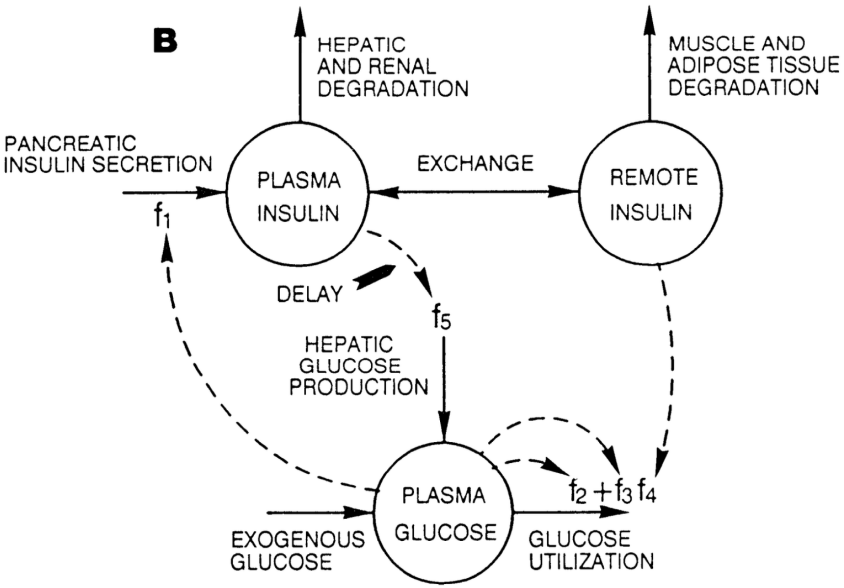
A way to think of these models in the broader vision:

the model narrows the search space over the space of all functions ***according to our knowledge of physiology.***

These models:

- have **limited flexibility** — *these functions only estimate the system well if our knowledge of the system is correct!!!*
- are **easy to understand/interpret**;
- require **very little data** to estimate/pick.

Example: a compartmental Ultradian glucose-insulin model



$$\frac{dI_p}{dt} = f_1(G) - E\left(\frac{I_p}{V_p} - \frac{I_i}{V_i}\right) - \frac{I_p}{t_p} \quad (1)$$

$$\frac{dI_i}{dt} = E\left(\frac{I_p}{V_p} - \frac{I_i}{V_i}\right) - \frac{I_i}{t_i} \quad (2)$$

$$\frac{dG}{dt} = f_4(h_3) + I_G(t) - f_2(G) - f_3(I_i)G \quad (3)$$

$$\frac{dh_1}{dt} = \frac{1}{t_d}(I_p - h_1) \quad (4)$$

$$\frac{dh_2}{dt} = \frac{1}{t_d}(h_1 - h_2) \quad (5)$$

$$\frac{dh_3}{dt} = \frac{1}{t_d}(h_2 - h_3) \quad (6)$$

$$\frac{dG}{dt} = f_4(h_3) + I_G(t) - f_2(G) - f_3(I_i)G \quad (7)$$

$I_G(t)$ represents nutrition input
$f_4(h_3)$ represents delayed insulin-dependent glucose utilization
$f_2(G)$ represents insulin-independent glucose utilization
$f_3(I_i)G$ represents insulin-dependent glucose utilization

Every model is a hypothesis regarding relationships between variables, specific disease progression mechanics, and we can test these hypotheses with data assimilation!

Inference: estimating or picking the model/function from a model/function space using clinical data

Estimating model parameters is equivalent to selecting which model, from among those available, is the most representative of the data and hopefully useful for prediction.

Some examples:

1. linear regression;
2. gradient steepest decent;
3. Markov Chain Monte Carlo;
4. Long-Short-Term-Memory (deep learning);
5. Kalman filter (data assimilation).

These methods span many fields, e.g., Applied Mathematics, Computer Science, Statistics, Engineering, and application areas too diverse to discuss broadly here.

There is no magic, there is no silver bullet

*...there is not even a **werewolf!***

Meaning:

there is no perfect prediction or perfect representation of the future or of knowledge needed to ideally and administer care in a personalized way...

There does not exist a magic function space and inference scheme that, once discovered, will render the need to thoughtfully treat a patient a thing of the past.

There is: a great deal of useful information we can extract from *clinical data* to provide clinicians with help achieving better patient outcomes using data science!

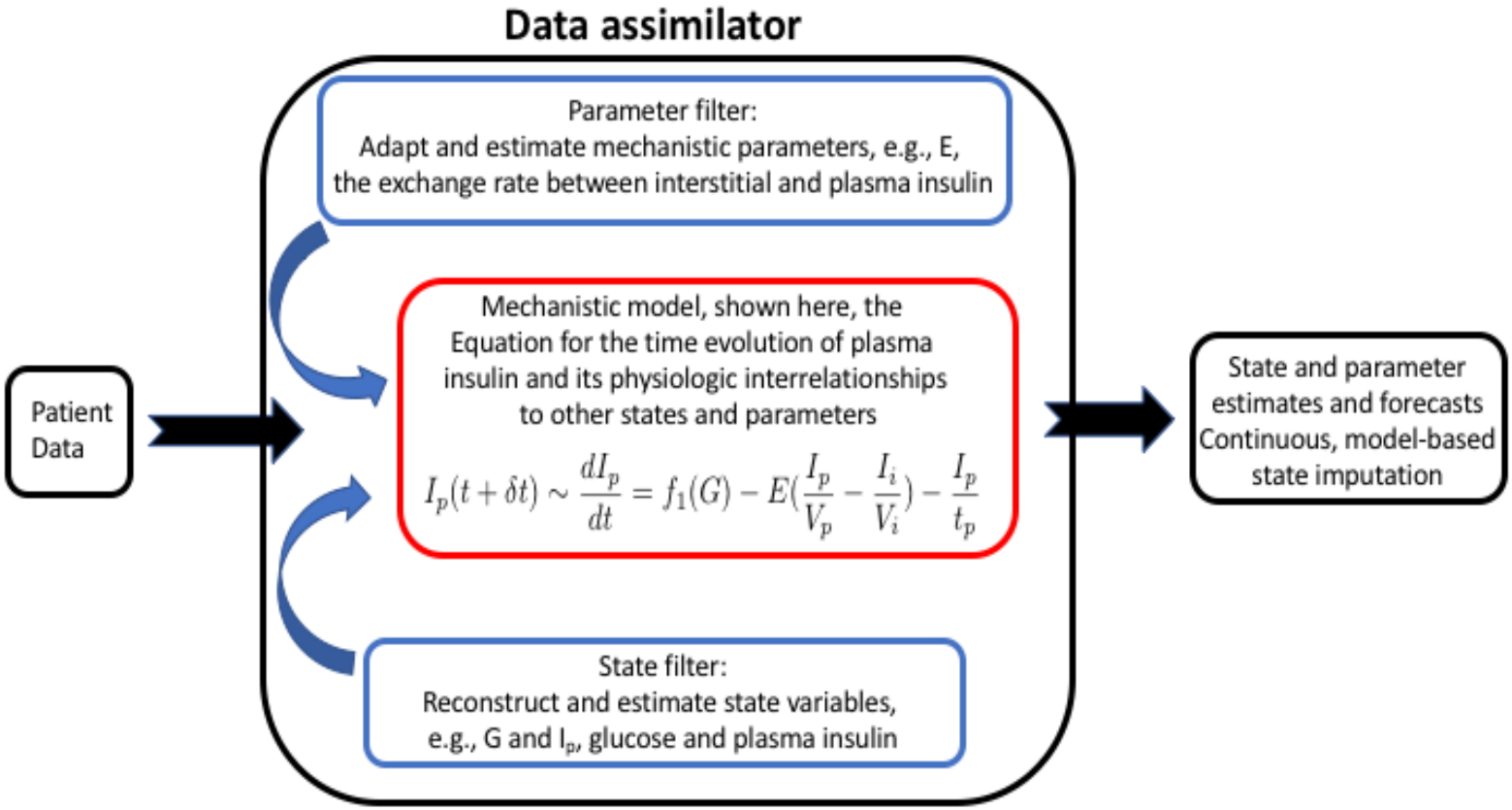
Parameterized-knowledge-driven models, a pathway for using clinical data and physiologic knowledge to create as a foundation for decision support

Using data assimilation (DA) to integrate physiologic knowledge with clinical data to make *personalized forecasts* relevant for clinical decision-making!

Two application areas:

1. T2DM self-management
2. ICU glycemic forecasting

Data assimilation: a knowledge-data-driven machine learning approach



DA is a broadly defined regression, it forecasts states and parameters in real time using a mechanistic model and data.

Application one: T2DM self-management

Three quick points:

1. Personalized T2DM self-management is a brutally difficult task!
2. Our knowledge encoded in physiologic models is worth something!
3. Equipped with a reasonable model DA can provide a lot of information with real world clinical data and can potentially be aid T2DM self-management.

Forecasting glucose for T2DM: contexts and deployment

Context 1: Real-time decision support for nutrition-based glycemic management of T2DM;

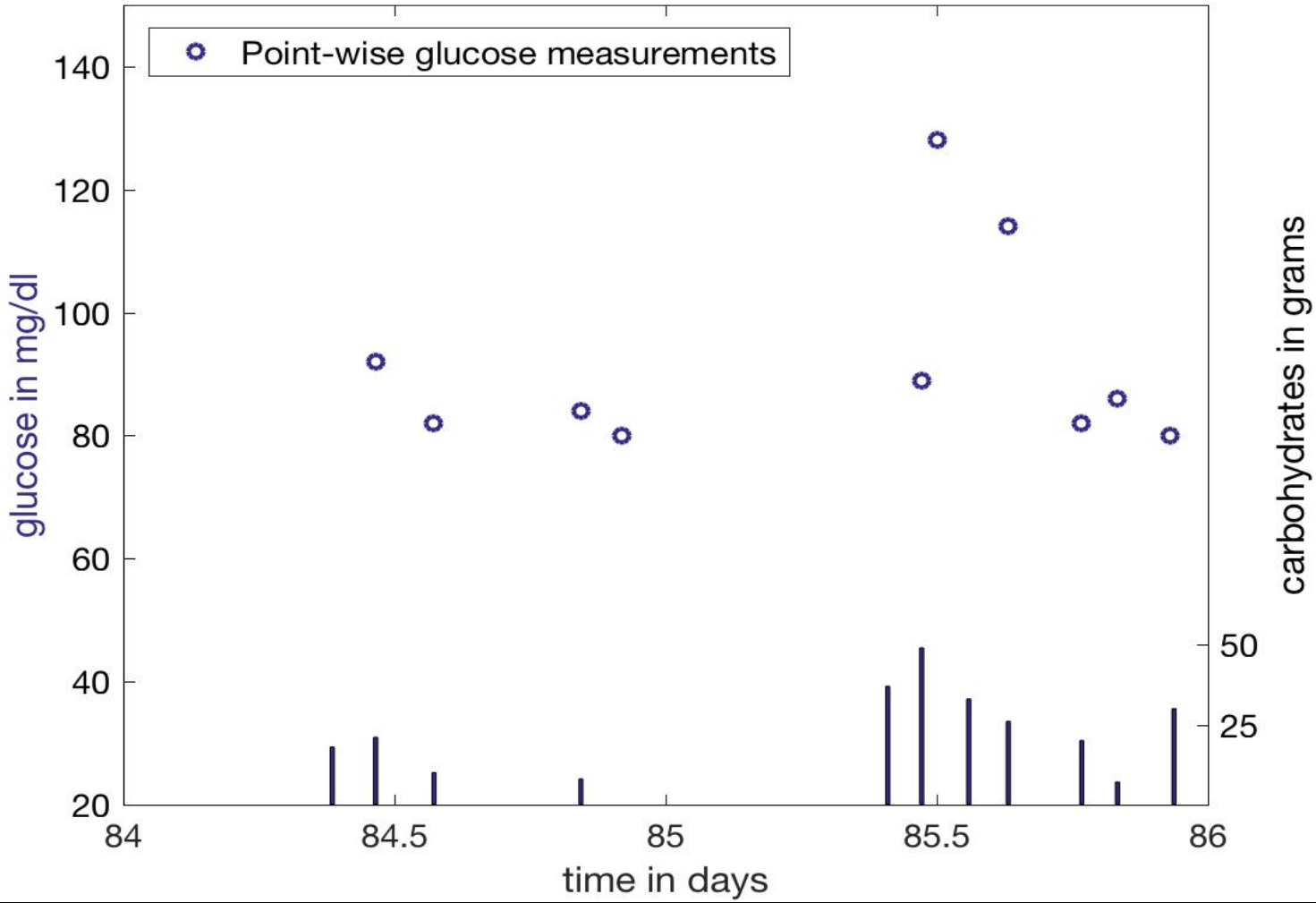
Context 2: Deepen our understanding of T2DM pathophysiology;

Deployment for T2DM: DA-based forecast machinery I will show you here has been deployed into a mobile application whose goal is to help people, in real time, make nutrition choices. Using this mobile application they can estimate the content of a food and see, immediately, the *predicted* impact that food choice, among other things...

NOTE: We are skipping the extremely difficult problem of trying to understand the most effective and productive way of communicating the DA output to people with T2DM, including people with limited health care literacy.

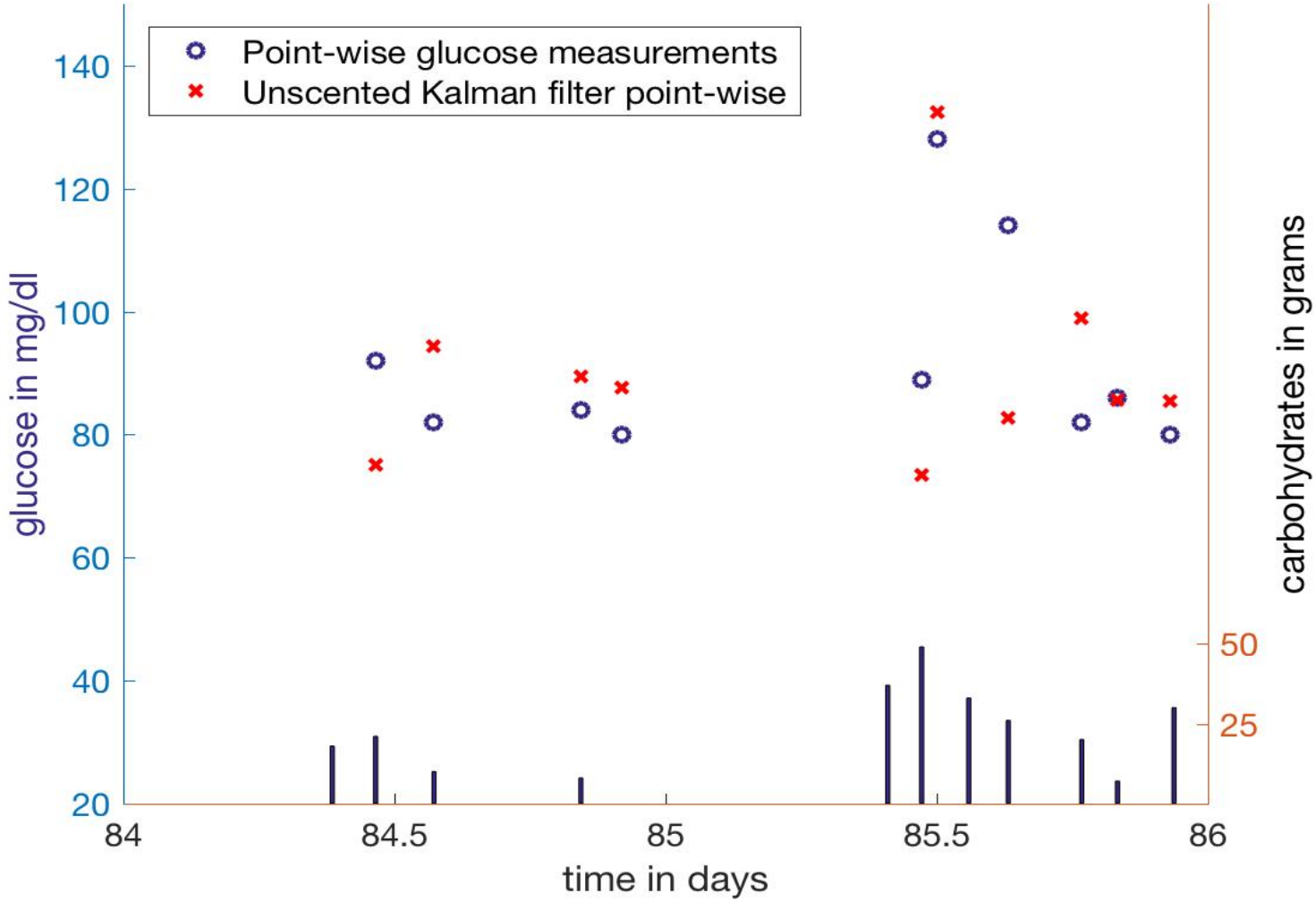
Common self-monitoring data

Two days of meals with pre- and post-prandial finger prick glucose measurements: looks like sorta random...



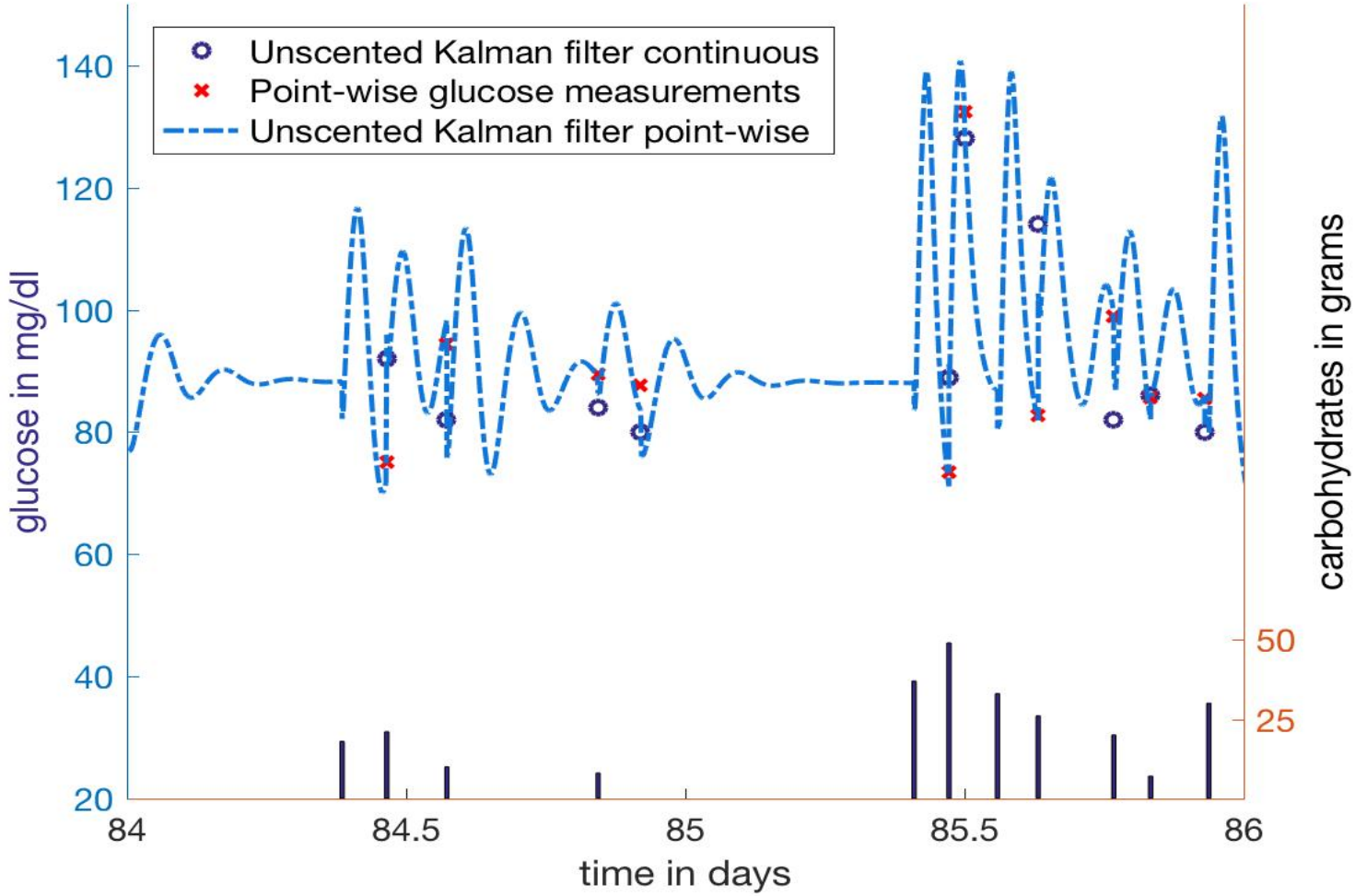
Self-monitoring data and *point-wise*, DA-driven postprandial forecasts.

Forecasts look OK, but both look sorta random...



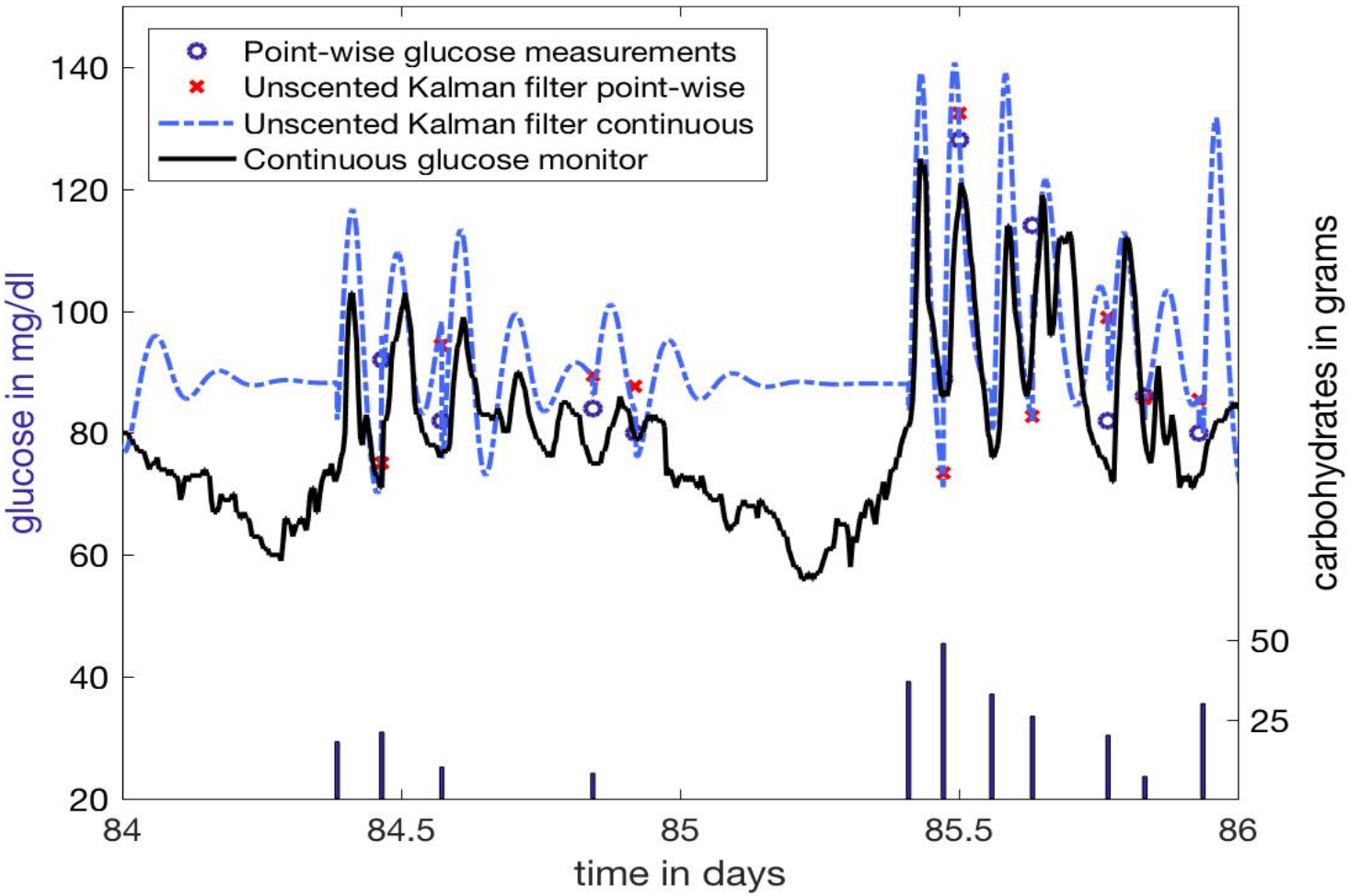
Self-monitoring data, point-wise and *continuous* DA forecasts

The model-predicted glycemic *dynamics* oscillate with high *amplitudes* and *frequencies*, implying that *forecast errors can be a severe function of measurement-time errors*.



Self-monitoring data, point-wise and continuous DA forecasts, and CGM measurements

With very sparse measurements the DA resolve continuous glycemic oscillations, but the complexity of the glycemic dynamics means they are hard to summarize and understand using sparse measurements.

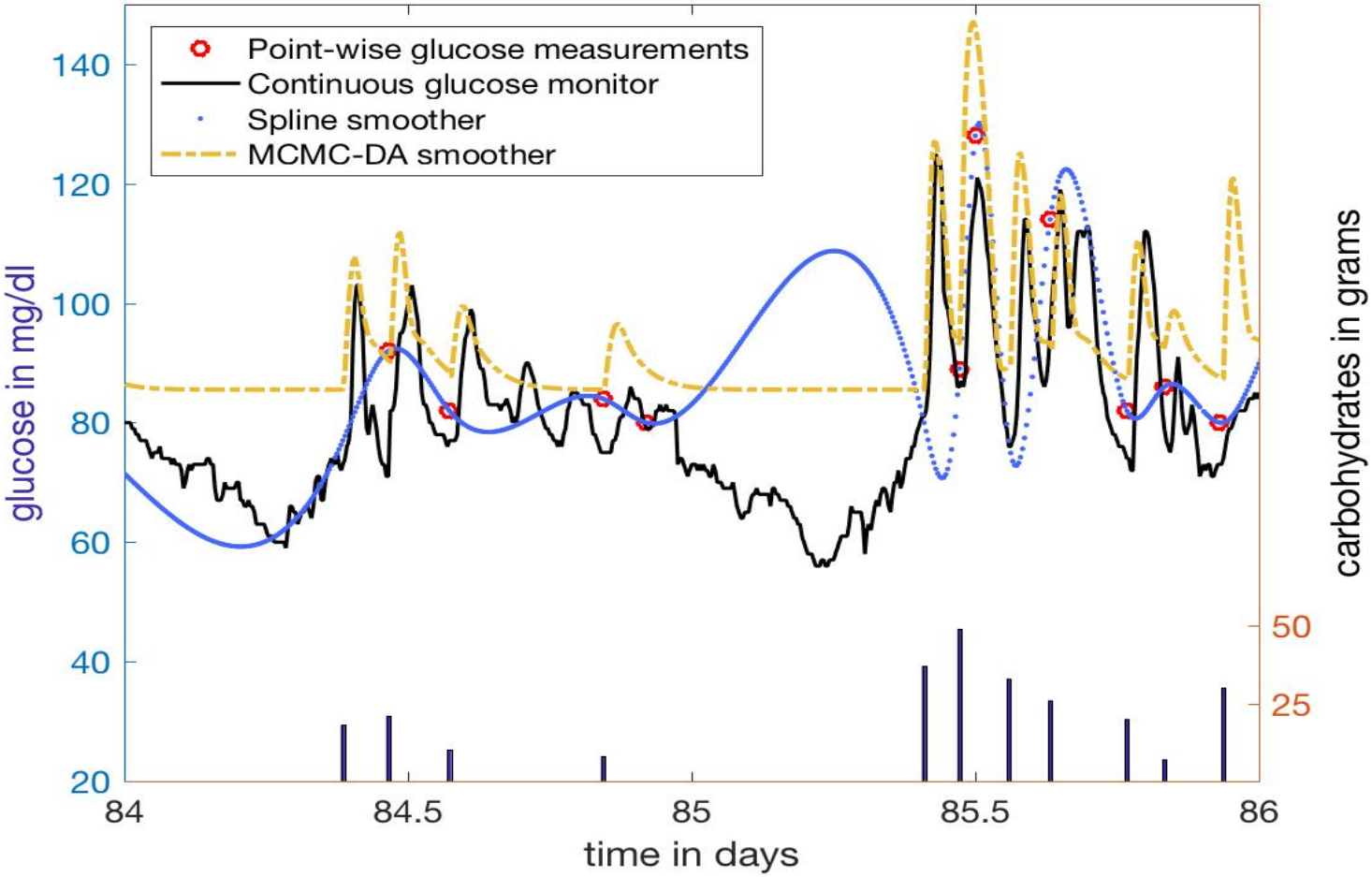


Mechanistic versus non-mechanistic machine learning

DA uses a *severely* constrained — by our knowledge or hypotheses of how the physical system works — *parameterized family* of models, limiting the flexibility of what the model can do.

ML uses an unconstrained parametric or non-parametric family of models with *almost no limits to what the model can do*

Knowledge-based constraints have consequences:



Phenotying disease?!

Aside:

using similar machinery—a different model and a similar DA—we are working to define phenotypes based on physiology — e.g., base on beta cell mass, insulin resistance, etc. inference *using the same type of data.*

Application two: ICU glycemic forecasting

Two points:

1. Forecasting or understanding glucose "in your head" from sparse data is extremely difficult!
2. Using data assimilation to leverage knowledge encoded in physiologic models can potentially help aid clinical decision making!

ICU: decision points and clinically available data

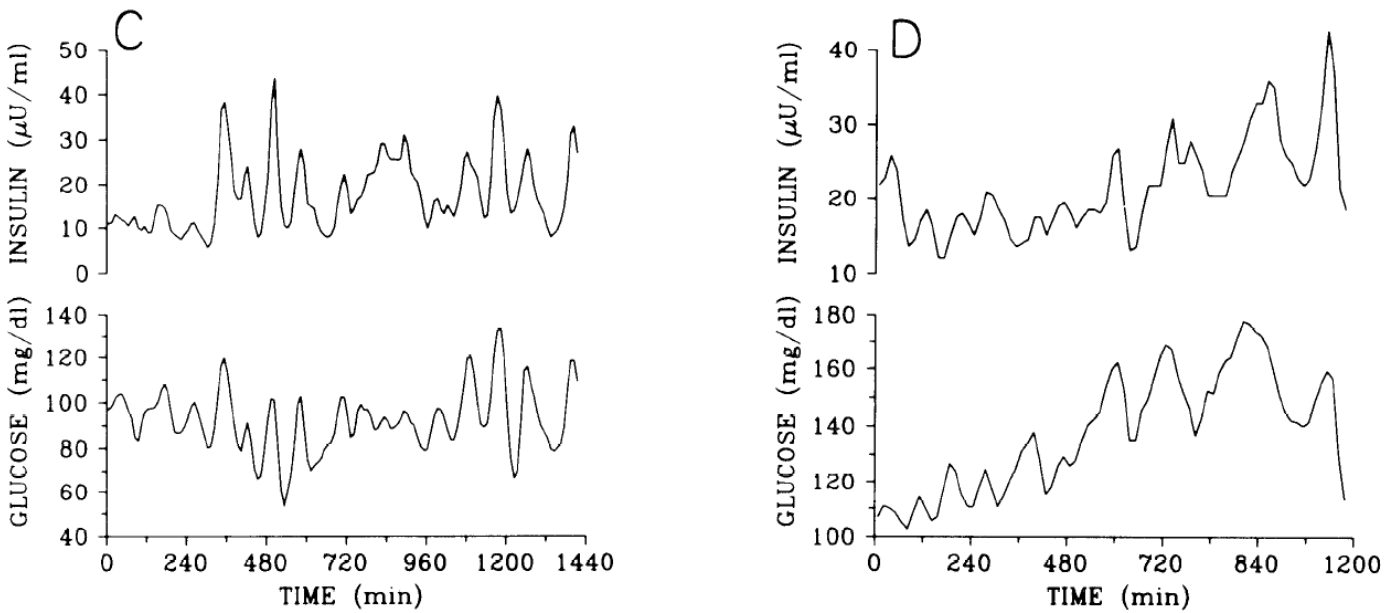
Most people in critical care need glycemic management: less than 10% of people have T2DM, 0.4% have T1DM, but most need insulin because of unknown reasons, e.g., stress, endocrine disfunction, etc.

Data are sparse: usually 1 measurement per hour unless hypoglycemia.

Clinical goals are intervals: the mean doesn't matter *for making decisions*, the boundaries of glycemic trajectories do, but managing the mean is the overarching goal.

So, why is glycemic management complex and time consuming?

A bit on the generating system: fast time-scale dynamics



Noisy periodic-like glucose and insulin dynamics are induced with continuous enteral (tube) fed (C) and continuous IV glucose infusion (D) nutrition.

The point: glucose dynamics are fast and complex and that makes them hard to understand and predict "in your head."

Glycemic management complex and time consuming because:

Armed with only sparse glucose measurements it may be difficult to predict, understand, and make decisions relative to boundaries because it is very difficult to resolve boundaries with sparse data.

Data within the clinical workflow and data

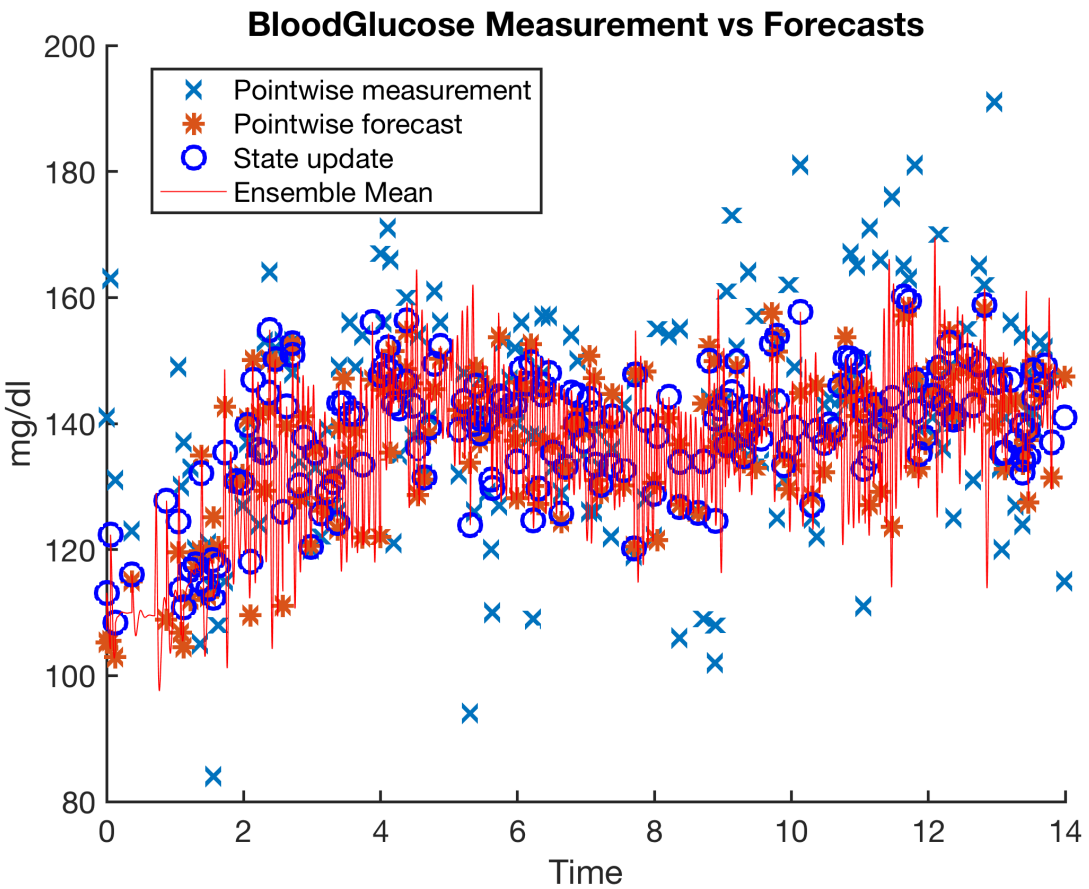
Data collected clinically:

1. Blood glucose measurements;
2. Insulin delivered — two forms, IV drip and injection;
3. Nutrition delivered — three forms, tube fed, IV drip, glucose bolus for hypoglycemia.

Issues:

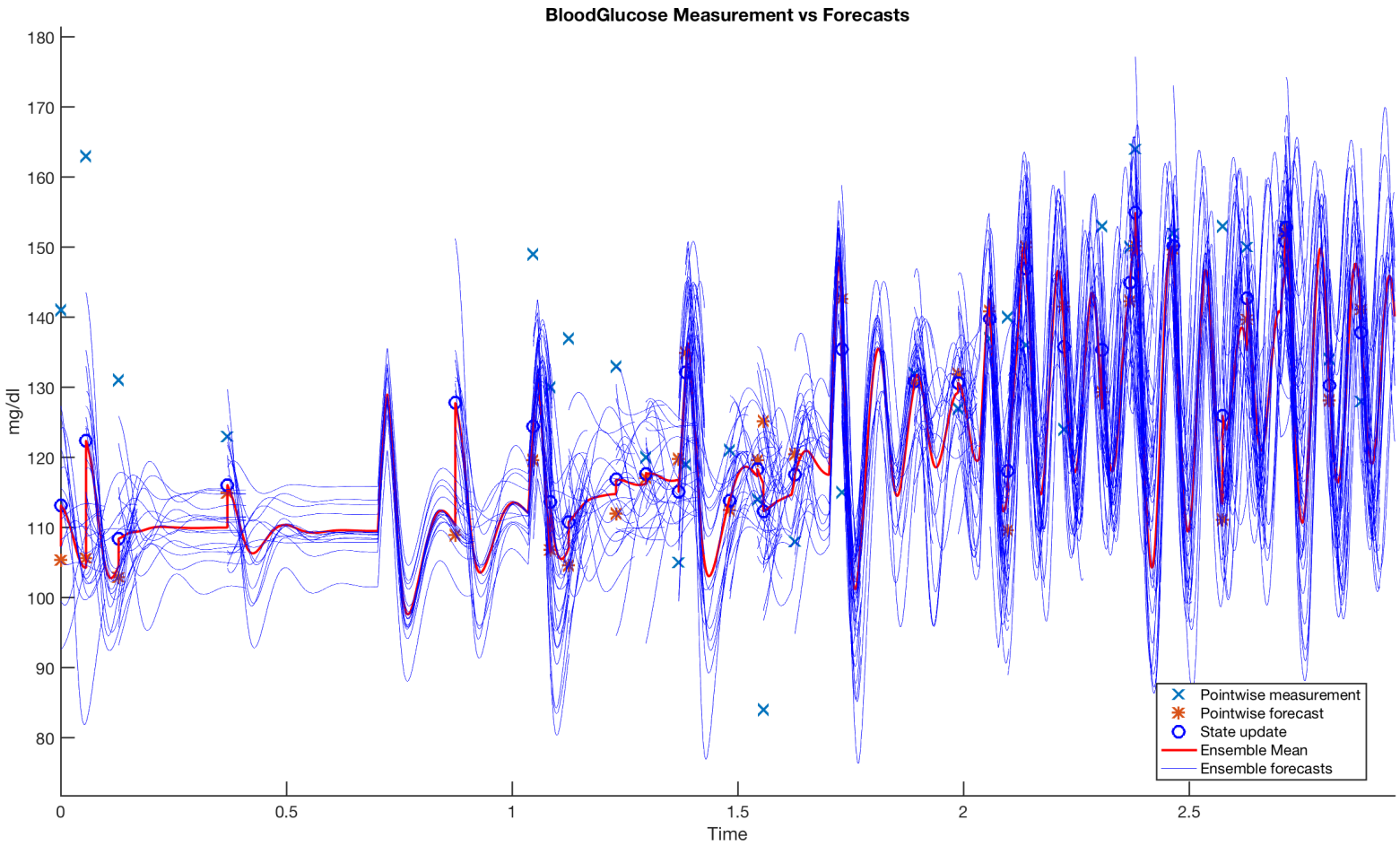
1. Accuracy of the time-resolution varies from about an hour—in's and out's flowsheets—to a minute—labs, dose history for non-continuous delivery.
2. Different sources have different relative contributions and extraction problems — e.g., drip nutrition contribution via D5W is *usually* small but can be the only nutrition source and is intermixed with the drugs delivered via D5W.

EnKF (DA) blood glucose (state) estimation and forecast

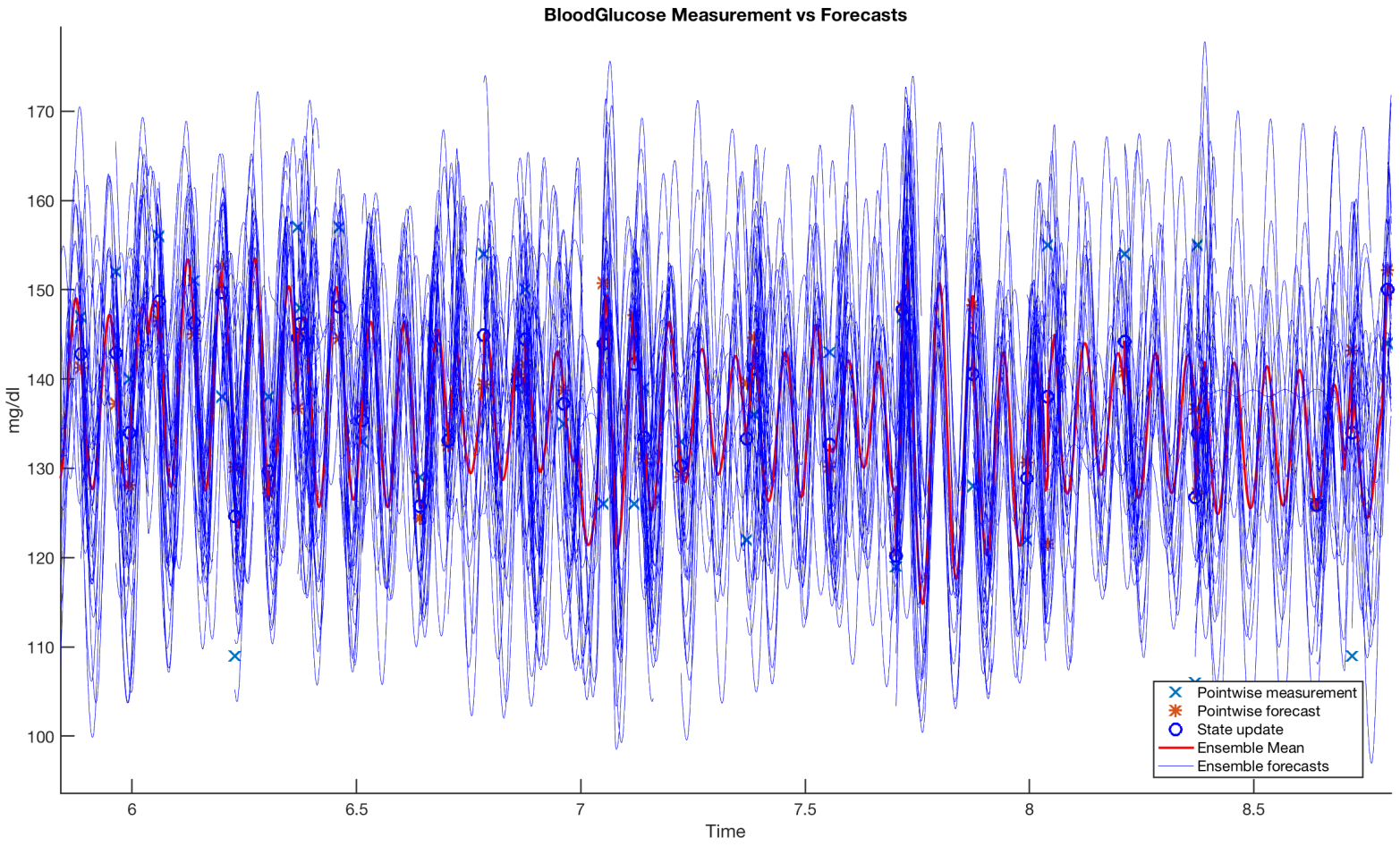


- **the EnKF seems to track well within about a day;**
- **after the EnKF begins to track it adapts to the *nonstationary* patient dynamics;**
- **the EnKF is continually synchronizing to the patient dynamics;**

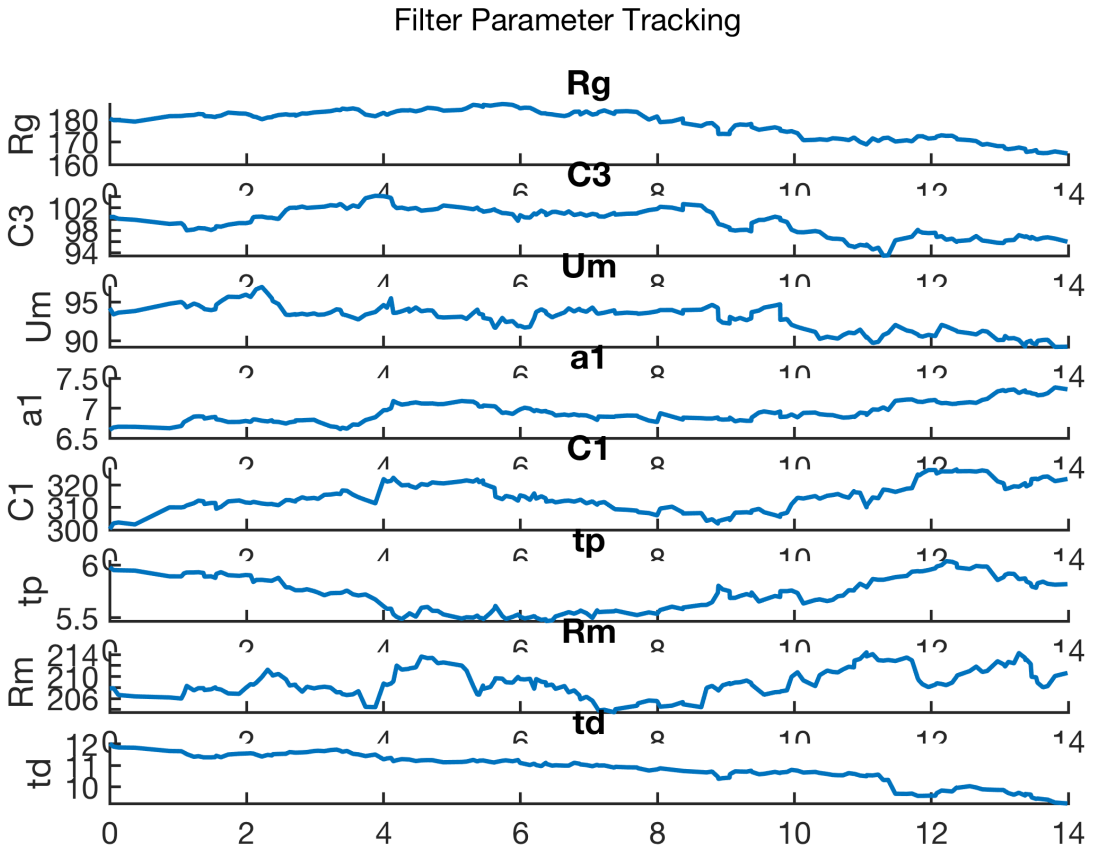
EnKF (DA) blood glucose (state) ensemble forecast showing forecast uncertainty quantification



EnKF (DA) blood glucose (state) ensemble forecast showing forecast uncertainty quantification



EnKF (DA) parameter estimation



- "Houlihan" parameters are chosen using the *Houlihan parameter selection method* we developed previously;
- Parameters estimate substantial physiology, including: insulin secretion, liver function, glucose utilization, endocrine stress.
- Parameters and their **trends** can potentially be inferred allow us to forecast *unmeasureable* patient physiology.
- Nonstationarity adaptation: the parameters adapt to the patient as the patient changes;
- Convergence was quite quick *for this individual.*

Figure: "Houlihan" parameter estimates: R_g, C_3, U_m, a_1, t_p, R_m, V_p.

Incidentally, the initial parameter choice for this person was close to their physiology...

THE POINT: the forecasting appears to work!!!

also WONDERFUL!!!! The EnKF appears forecast glucose relatively well.

But, it doesn't work as well all the time... I am happy to share why and how but I am about out of time...

Fitting into the clinical workflow

All this forecasting is well and good, but a full simulation of the glucose-insulin system, in its raw from is not ideal for making clinical decision making!

The DA or any ML output must be transformed into a form that is useful for making clinical decisions — this work lives within the fields of *human-computer interaction* and cognitive science.

I want to acknowledge this crucial step because: working with ML/DA/AI *constrained to the transformation of the knowledge into a usable form* is crucial for establishing the pathway for data science to improve patient care and outcomes.

Summary: positive take-aways

AI/ML/DA in biomedicine:

1. AI/ML/DA must be integrated into the entire clinical data and practice workflow to be directed such that is will be useful.
2. Clinical data are collected/generated by a combination of processes, physiology, health care process, etc., and can be complex to use but represent the *real world*.
3. There is not AI/LM/DA silver bullet — but we can exploit data using data science to support clinical decision making in a personalized way.

DA-based pathway to clinical decision support and physiologic understanding:

1. We are making progress: *with the right model, the right parameters, and enough data, **we can accurately infer the present and forecast the future patient states**, but challenges remain.*
2. We can likely get DA to work with clinical data and infer information/insight we would otherwise not have access too.
3. We have a way of testing and refining physiologic hypotheses in the form of mathematical models with data!!!
4. A key aspect of DA development is its concrete link to decision-making and understanding via clinical decision support HCI applications.

And thanks to my collaborators in no order:

George Hripcsak, Matt Levine, Melike Sirlanci Tuysuzoglu, Jami Jackson-Mulgrave, Kai Bartlette, Lena Mamykina, Andrew Stuart, Bruce Gluckman, Rammah Abohtyra, Will Ott, Noemie Elhadad, Inigo Urteaga, Erica Graham, Rimma Perotte, Adler Perotte, Henry Ginsberg, Jan Claassen, Tell Bennett, Peter Sottlie, Marc Moss, Brad Smith, Artie Sherman, Joon Ha, Cecilia Diniz Behn, Christine Chan, Melanie Green...

Papers + collaborators + thanks

Related papers:

1. "Mechanistic Machine Learning: how data assimilation leverages physiologic knowledge using Bayesian inference to forecast the future, the infer present, and phenotype." with BG, GH, ML, LM, and AS in JAMIA.
2. "Ensemble Kalman Methods With Constraints," with P-AB, ML, EES, and AS, arXiv:1901.05668.
3. "Personalized glucose forecasting for type 2 diabetes using data assimilation," with ML, GH, BG, and OM, PLoS Comp Bio.
4. "High fidelity phenotying," GH and DA in JAMIA.
5. "The Parameter Houlihan: a solution to high-throughput identifiability indeterminacy for brutally ill-posed problems," with ML, LM, and GH, submitted, arXiv:1902.01978v1.
6. "Offline and online data assimilation for real-time blood glucose forecasting in type 2 diabetes," Matt Levine, GH, LM, AS, DA, submitted to PLoS Comp Bio.
7. "Population physiology," with GH and MS, PLoS One.
8. "Dynamical phenotyping: using temporal analysis of clinically collected physiologic data to stratify populations," with GH, NE, AP, and ET, PLoS One.
9. "Data assimilation of glucose dynamics for use in the intensive care unit," with BG and MSS, IEEE-EMBC.
10. "Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations, " with GH, CHAOS.
11. "Estimation of time-delayed mutual information and bias for irregularly and sparsely sampled time-series, " with GH, Chaos, Solitons and Fractals.
12. "Next-generation phenotyping of electronic health records" with GH, JAMIA;

Collaborators and Money: NLM grants RO1 LM012734 and LM06910, Robert Wood Johnson Foundation RWJF 73070, NIDDK R01DK090372

boo!