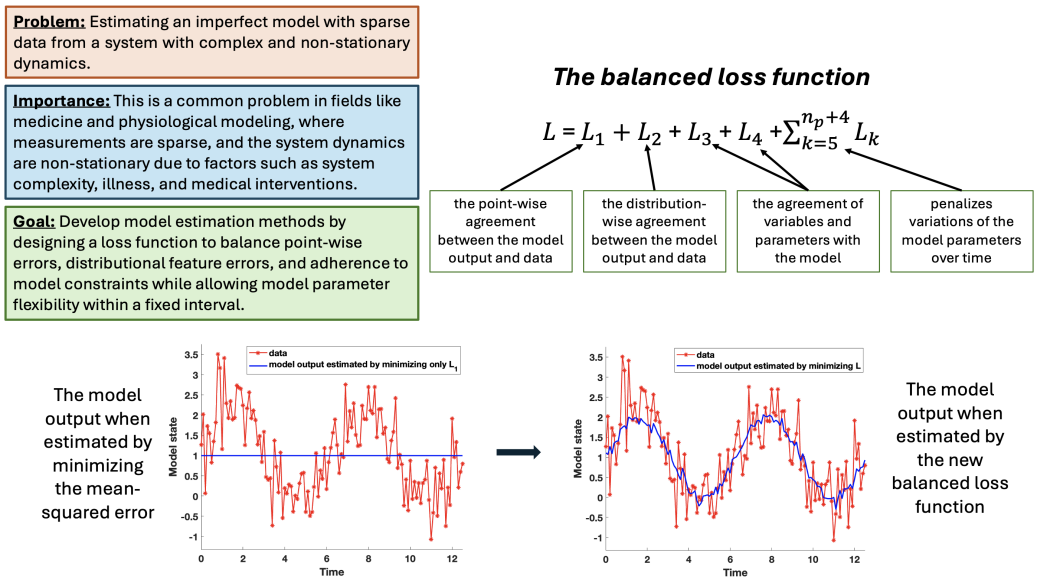


Graphical Abstract

A multiobjective optimization approach to data assimilation for complex biological systems with sparse data

David J. Albers, George Hripcsak, Lena Mamyina, Melike Sirlanci, Esteban G. Tabak

Highlights

A multiobjective optimization approach to data assimilation for complex biological systems with sparse data

David J. Albers, George Hripcsak, Lena Mamyina, Melike Sirlanci, Esteban G. Tabak

- Sparse data and imperfect models prevent accurate model estimation.
- A multiobjective optimization methodology is developed to tackle these challenges.
- The multiobjective loss function simultaneously minimizes pointwise and distributional error metrics.
- The multiobjective loss function allows estimated parameters to vary along the estimation interval.
- This approach produced accurate model estimation and calibration.

A multiobjective optimization approach to data assimilation for complex biological systems with sparse data

David J. Albers[a,b,c,d], George Hripcsak[d], Lena Mamyina[d], Melike Sirlanci[a,e,*], Esteban G. Tabak[f]

[a]*Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, 80045, CO, USA*
[b]*Department of Bioengineering, University of Colorado Denver, Aurora, 80045, CO, USA*
[c]*Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, 80045, CO, USA*
[d]*Department of Biomedical Informatics, Columbia University, New York, 10032, NY, USA*
[e]*Department of Applied Mathematics, University of Colorado Boulder, Boulder, 80309, CO, USA*
[f]*Courant Institute of Mathematical Sciences, New York University, New York, 10012, NY, USA*

Abstract

This article develops a novel multiobjective data assimilation methodology, addressing challenges that are common in real-world settings, such as severe sparsity of observations, lack of reliable models, and non-stationarity of the system dynamics. These challenges often cause issues and can confound model parameter estimation and initialization that can lead to estimated models with unrealistic qualitative dynamics and induce qualitative and quantitative parameter estimation errors. The proposed multiobjective function is constructed as a sum of components, each serving a different purpose: enforcing point-wise and distribution-wise agreement between data

*G.H. is supported by NIH R01 LM006910. L. M. is suported by NIH R01 DK113189. E. G. T. is supported in part by the Office of Naval Research, through grant N00014-22-1-2192.

*Coresponding author
Email address: melike.sirlanci@cuanschutz.edu (Melike Sirlanci)

and model output, enforcing agreement of variables and parameters with a model provided, and penalizing unrealistic rapid parameter changes, unless they are due to external drivers or interventions. This methodology was motivated by, developed and evaluated in the context of estimating blood glucose levels in different medical settings. Both simulated and real data are used to evaluate the methodology from different perspectives, such as its ability to estimate unmeasured variables, its ability to reproduce the correct qualitative blood glucose dynamics, how it manages non-stationarity, and how it performs when given a range of dense and severely sparse data. The results show that a multicomponent cost function can balance the minimization of point-wise errors with global properties, robustly preserving correct qualitative dynamics and managing data sparsity.

Keywords: dynamical system, data sparsity, non-stationarity, data assimilation, optimization, glucose-insulin system modeling

---

1. Introduction

Physiological systems, an archetypical example of complex biological systems, can be difficult to estimate because they tend to display at least some of the following characteristics:

1. **Lack of reliable models.** Many such systems are only partially understood, and certainly rarely understood to the point where a well-defined model provides an accurate surrogate for the underlying processes. Even in situations where those processes can be captured by models to some degree –think general circulation models for weather forecasting– typically a large number of model parameters remain undetermined.
2. **Latent and emerging bulk variables.** The phase-space of complex systems is high dimensional, while only a handful of variables are observed systematically. The value of the remaining, *latent* variables can only be inferred indirectly from the observations, if at all. Since the full set of latent variables is often too large and hard to pinpoint, it is useful to consider instead *emerging bulk variables*, a small set of parameters representing a larger set of unmeasured variables and processes that affect the dynamics in a coherent way.
3. **Sparse observations.** Usually complex systems may be observed not only partially, but also at a sparse set of times, leaving the observations

insufficient to even marginally resolve some of its dynamically significant time-scales. The timing of these observations may be structured (for instance by a medical protocol), event-driven or purely random.

4. **Prior knowledge.** Even though a detailed model of the underlying processes may be unavailable, one often has some prior information on the system's dynamics, such as the expected presence of oscillations and their typical frequency and amplitude, or the difference in behavior between driven and unforced scenarios.
5. **Non-stationarity.** The system's underlying dynamics may evolve over time, due either to external drivers or to un-modeled components of the system. Non-stationarity may manifest itself through a slow time modulation of the model's parameters.

Estimating from data models such as ordinary, partial, and stochastic differential equations of complex biological systems is performed through data assimilation (DA) [48, 47]. The above characteristics tend to lead to non-convex optimization problems that are not guaranteed to have a single optimal solution. Moreover, in the case of nonlinear optimization, we cannot know if we have found said solution, unless the model's parameters and state space we search over have properties such as an absence of bifurcation points [46]. Additionally, in many situations, particularly when data are sparse, different objective functions, e.g., least squares versus distributional minimization, can also lead to different objective functions which often do not have the same optimal solutions. Together these sources of non-uniqueness of the meaning of optimal must be reconciled with the purpose of the model estimation, which can vary from the validation of a scientific hypothesis to supporting decision making.

To address DA to estimate complex biological models in the context where we may have limited knowledge, sparse data and diverse estimation needs, we propose to use multiobjective optimization [42], scalarized through a weighted sum of different objective functions to balance goals, such as minimizing distributional versus point-wise estimation errors and penalizing fast variability of the model's parameters. Even though multiobjective optimization has a longer history in decision theory [43, 44, 45] than in DA, it be directly applied to DA's optimization framework and its multiple goals [47, 48].

There are broadly two ways of managing multiobjective optimization: scalarization and creation of a utility function that include all the objectives

and are heuristic and task-specific. These two approaches can sometimes be unified, as recently shown in the case of R2 utility functions [42]. Here we will propose a new set of objectives and a scalarized objective function that is a weighted sum of likelihood functions. We show and demonstrate qualitatively that the multiobjective optimization function we construct helps address many of the roadblocks to estimating complex biological systems mentioned above.

1.1. Motivating example from physiological modeling within medicine

To explicitly demonstrate the abstract problems raised above, and to motivate the need for multiobjective optimization in the biological context, we start by showing one example: a particularly vexing but also commonly observed problem computing parameter initializations for DA using sparse data [4, 2, 5, 8] with the goal of estimating and forecasting endocrine function related to blood glucose regulation. While this example is drawn from a particular case from our recent work [4] in the ICU, we have observed similar issues in many contexts, including in critical care in the intensive care unit (ICU) [31, 4, 38, 30, 36], type-2 diabetes [2, 3, 21] and outpatient or *in the wild* settings [24, 37, 41].

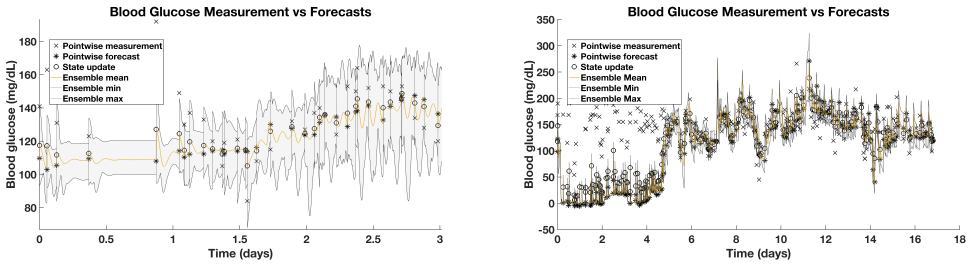
The origin of these problems can be traced to sources that include model rigidity, data sparsity, non-stationarity and the presence of errors in the recorded measurement times. These problems appear in the context of minimizing a single loss objective though various algorithmic structures (e.g., Kalman, active set, interior point, Nelder-Mead, etc.), which under sparse data can completely mis-specify the system's true dynamics (replacing for instance an unresolved oscillatory signal by its mean value.) *While we have faced these problems in DA initialization in biomedicine, their sources generalize to a broader range of other fields* such as atmospheric physics [40, 25]. Thus this paper addresses general, highly related problems encountered when estimating an imperfect model with data from a system that generates complex, non-stationary dynamics and is severely under-measured [4, 8].

To demonstrate how standard methods can fail in realistic settings, consider an application to estimating the states and parameters of a glucose-insulin model given sparse data taken during a stay at an intensive care unit (ICU) [4]. Data includes continuously measured tube-administered nutrition, point-wise blood glucose measurements collected according to clinical protocols [16, 26], usually about once an hour, *administered* insulin and its type, but never plasma or interstitial insulin measurements. In the ICU, where

nutrition administration is roughly constant because of its administration via a tube feed, blood glucose and insulin oscillate out of phase on the order of tens of minutes [35]. The model we estimate is the Ultradian model [35, 19, 2, 4, 38] that includes blood glucose, blood insulin, remote (interstitial) insulin, and a three state delay approximated with the linear-chain-trick [34]. For purposes of demonstration we estimate the model with a version of an ensemble Kalman filter (EnKF), initializing the parameters at the nominal parameters for the model [35, 2] as in the prior work from which this motivation was drawn [4, 38]. Because blood and remote insulin are never measured, the model is not identifiable, but can still be estimated with some accuracy under many circumstances [2, 4, 38, 3]. We focus on this example for four reasons. *First,* this situation poses a real world problem [29, 9, 13, 4, 38] and is highly representative of many biomedical and health care settings where the system is non-stationary with complex dynamics and data are sparely measured. *Second*, within this real world context, the forecasting and estimation needs are diverse and dependent on the decision-needs, raising the potential need for multiobjective optimization. For example, there are situations where: *(a)* only parameter estimates are important but next-step glucose prediction is not, *(b)* next-step glucose prediction is important but accurate parameter estimates are not, *(c)* only accurate point-wise glucose forecasting is important, and *(d)* accurate parameter estimates and distributional glucose forecasting are important but accurate point-wise glucose forecasting is not. This final case is likely to be the most common situation in applied settings. *Third*, we have a qualitative understanding of the underlying dynamics of the data-generating system, including both glucose-insulin system and the measurement processes and protocols [14, 15], and *fourth*, the measurements of this system are sparse in time compared to the complexity of the underlying dynamics, with measurement times that may not be entirely random.

Operationally, we make the following assumptions:

a. we know the generating dynamics correspond to a driven and damped system that elicits a nondescript noisy (potentially chaotic or otherwise random) oscillatory orbit;
b. measurements are taken infrequently enough that the spectral composition of the orbit is difficult or impossible to resolve from the data alone;
c. reporting of the measurement times has nontrivial errors.

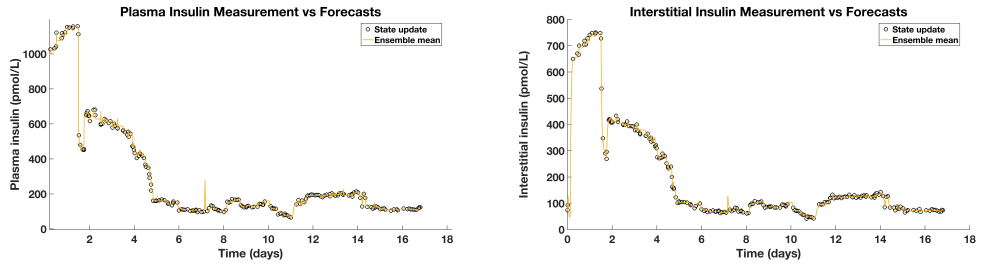


(a) (b)

Figure 1: Estimating and forecasting glucose trajectories for two patients in the ICU. The plot on the left shows accurate estimation and forecasting almost immediately (< 1 day, < 6 data points) while the plot on the right shows poor estimation and prediction until about day 5 (~ 125 data points). Sources leading to model estimation accuracy include a complex interplay between data sparsity leading to model identifiability problems (Figs. 2-3) and model initialization.

In such a setting it may be difficult or impossible to find a set of initial conditions and parameters that minimize point-wise errors between the model and data and correspond to a solution that is not a fixed point. The reason why a fixed point solution is a plausible and often computed error minimizing parameter set is because a fixed point whose value is the mean of the data set is a solution with relatively low and bounded point-wise error. However, when we know that the generating system has oscillatory dynamics, such fixed point solutions to the inverse problem (the estimated model parameters, states and initial conditions) **must be wrong**: even though the estimated parameters minimize the least squared error, they produce qualitatively wrong dynamics. The root of this problem lies in how we quantify what it means for a model estimate to be *wrong*, which is directly embedded within the objective function we optimize. When the generating dynamics are complex and sparsely measured, we have frequently encountered the situation where least squared error minimizing parameters produce qualitatively wrong generating dynamics.

To demonstrate the problem, begin with the case where patient 426's data are estimated with the ensemble Kalman filter (EnKF), shown in Fig. 1a, where the model is initialized with nominal parameters of a healthy adult (This is not the smoothing case addressed in this paper, we will come to those results in a moment.) Here we can see that the model estimation converges within about $1-1.5$ days, the ensemble mean oscillates as we might expect, the uncertainty in the estimate is reasonable, and these data are



(a) (b)

Figure 2: In the ICU, insulin, one of the states that defines the glucose-insulin system and should be in the range of 25-400 picomoles per liter (pmol/l), is never measured. This can lead to model estimation and initialization problems, as seen in Fig. 1. Here we see the estimated interstitial and plasma insulin levels that are driving the forecasting errors seen in Fig. 1b. Note that after about 5 days the model does eventually entrain to the patient and the insulin estimates take physiologically plausible values.

reasonably well predicted by the model-based forecast [4]. This is an example of what successful estimation and forecasting looks given such data. A less successful estimation attempt is shown in Fig. 1b where we estimate patient 593's data. In this case it takes ~ 4.5 days for the model to entrain to the patient. Initially the EnKF estimates produce blood glucose values far below what would represent a living human, and the glycemic dynamics are not oscillatory. Figure 2 reveals two problems. First, a severe source of estimation error is due to model identifiability, the unmeasured insulin states (blood and remote insulin) take values far above what is possible until about day 4.5, the point where the model has converged such that the ensemble mean is tracking the glycemic dynamics. Second but seemingly less problematic, the model was not initialized with accurate parameters as can be seen by the 4.5 days worth of data it takes the EnKF to tune parameters to entrain to the patient. These two problems have two naively obvious solutions.

One approach toward solving the first problem, identifiability failure due to unmeasured model states, is to constrain the unmeasured states to occupy plausible ranges. In prior work we developed a constrained version of the EnKF [5]. This constrained (CEnKF) algorithm uses the standard EnKF framework, checks after the update step if any of ensemble particles lie outside the predefined constraint region, and if any particles lie outside the constraint region, the algorithm invokes a quadratic program to solve for parameter values that bring the particles inside the constrained region. The results of applying this method can be seen in Fig 3a. With applied

constraints the model converges within about 1.5 days compared to the 4.5 days in the unconstrained case, a substantial two-thirds decrease in time to convergence. Details of these results can be found in refs. [4, 5]. Comparing Figs. 2 and Fig. 3b we can see the constraints' impact on the unmeasured insulin, forcing the insulin to take more reasonable values. Additionally in Fig. 3c we can observe the parameter trajectory that shows a relatively steep parameter changes within the first 1.5 days, and in Fig. 3d we can see the proportion of particles that violate the constraints for each data point or iteration of the CEnKF. Within about 20 data points, the constraints are usually satisfied, and it was only the insulin values that were violating the constraints. However, imposing the constraints brings unintended consequences that are at the core for the motivation in this paper. Figure 3e shows the particle trajectories within the EnKF ensemble, and the variation of the mean. It appears that the constraints have changed the model dynamics from oscillatory to fixed points. Meaning, the EnKF variance is mostly constructed from variations in the fixed point location according to the variation within the ensemble in parameter space. *The end result of this demonstrates that constraining the model states (a) does help solve the identifiability problem, (b) is not enough to fully solve the initialization problem and (c) may have unintended consequences related to the model dynamics that may require a more complex objective function to manage.* Specifically, the optimal solution tends toward a fixed point of the dynamical system, a solution we know to be wrong, implying that the constrained model dynamics may not be representative of the underlying system we know to exist.

A naive idea we thought might work to solve the parameter initialization problem was to run deterministic optimization, bootstrapping over a few thousand random initial conditions using the first 24 hours of data. Then, the outcome from this calculation would be used to initialize parameters and states for the EnKF. Notably, *this application represents the DA smoothing case that we address in this paper.* The results of this effort for both patients 426 and 593 are shown in Fig. 4. This initialization indeed reduces the MSE between the EnKF ensemble mean and these data. However, for patient 593, shown in Fig. 4a, the oscillatory glycemic dynamics are completely gone and have been replaced with strongly attracting fixed points whose variation is due to the parameter ensemble of the EnKF. If we apply the same parameter initialization procedure to patient 426's data, we find something even more interesting. The dynamics of the particles inside the ensemble for the model initialized with these optimized parameters, shown

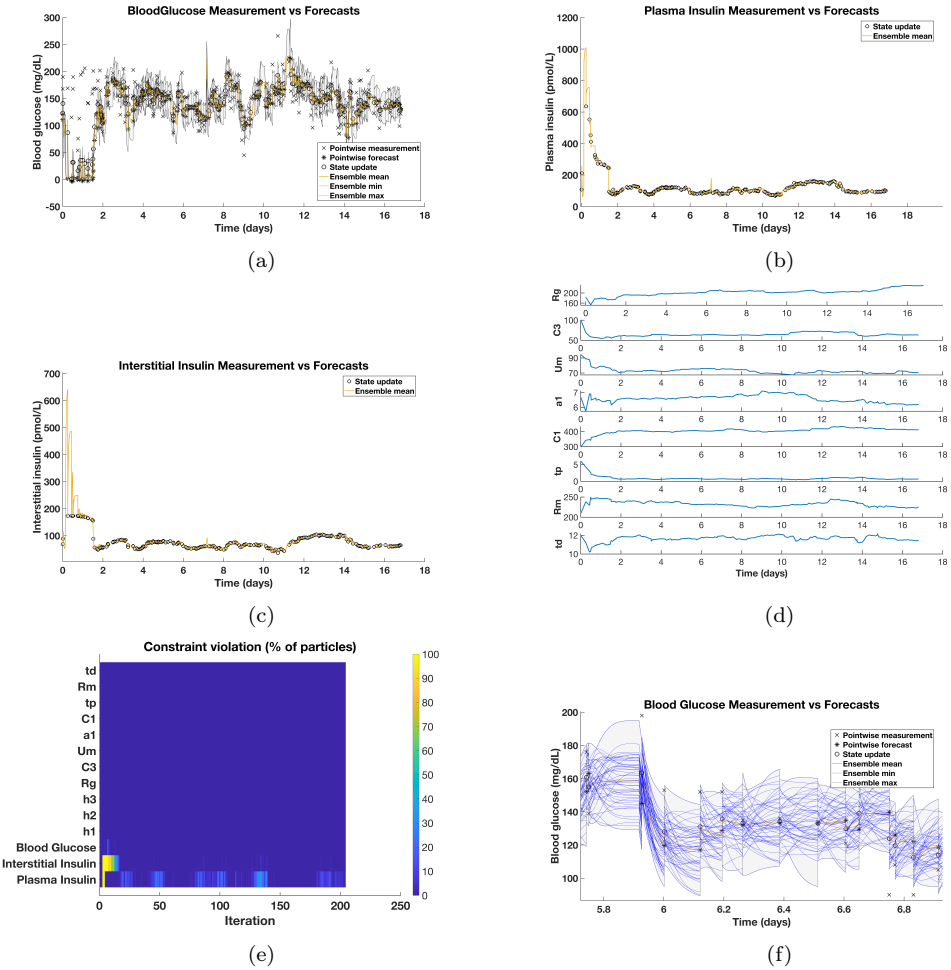
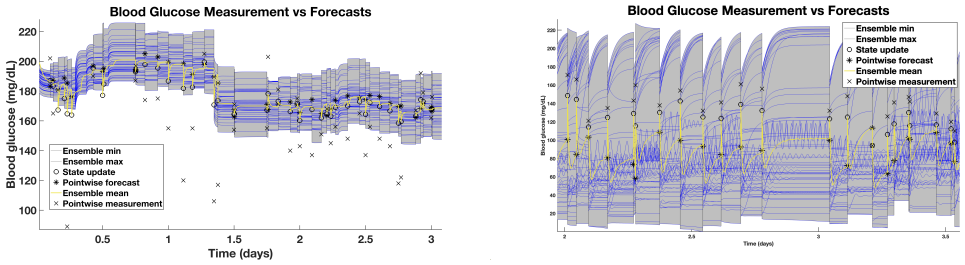


Figure 3: Patient 593's glucose trajectory (a) and insulin states (b) estimated with the constrained EnKF, the constrained parameter estimate trajectory (c), the percentage of particles violating the constraints per *data point* for estimated model states and parameters (d), and the individual ensemble particle trajectories for 593's estimated glucose trajectory (e).

in Fig. 4b, have several simultaneously present, topologically different orbits ranging from fixed points with different attraction rates to periodic orbits. *The naive constrained DA did substantially reduce MSE, and in doing so, led to error minimizing parameters at a bifurcation point.* This observation is problematic given recent guarantees on unique solutions for such optimization problems require searching over parameters sets that omit bifurcation



(a) (b)

Figure 4: Constrained EnKF forecasts using parameter initializations computed by optimizing over the patient's first 24 hours of data. Plot (a) shows patient 593's individual ensemble particle trajectories of the estimated glucose trajectory, (b) shows patient 426's individual ensemble particle trajectories of the estimated glucose trajectory. Note the particle ensembles are taken over estimated states and parameters. For patient 593 the glycemic dynamics are wrong, they should be oscillatory instead of a strongly attracting fixed point. For patient 426, the glycemic dynamics are representative of a bifurcation point demonstrating several topologically distinct orbits within the same CEnKF ensemble.

points [46]. But more to the point, the algorithms are doing exactly what we would expect given these sparse data. A bifurcation point is a *particularly* useful location in parameter space from the functional approximation theory perspective because a bifurcation point provides a diverse set of functions to use to estimate the mean and variance of data. *Meaning, for standard least squares minimization for problems with sparse data, it makes sense that the optimization algorithm will seek out bifurcation points to estimate data.*

To address these problems we need to include more structure in the loss function beyond minimizing the MSE between an ensemble average and data subject to constraints—*we need a multiobjective function to optimize.* We want to *(a)* minimize the MSE between an ensemble average and data while*(b)* ensuring that other global features such as measure-theoretic and topological properties are also accurately characterized and *(c)* forcing the estimation process to respect the model. This requires a more diverse loss function structure, and establishing how to formulate and estimate such a loss function is the point of this work.

1.2. Outline of our multi-objective function approach

The approach that we propose diverges from the motivational example in a few ways. First, we focus on the DA smoothing case rather than the

filtering case. Second, because we construct a multiobjective function, we do not use Kalman inversion or filtering methods that minimize the conditional mean of these data. Third, in order to account for model and measurement uncertainty, we estimate these data with a stochastic differential equation rather than an ordinary differential equation. The stochastic model of the evolution of the system's state needs not be necessarily accurate or fully based on field-specific knowledge, but rather flexible and tailored to the qualitative nature of the expected dynamics, noisy oscillations in our study case. The model we use for our examples also allows us to write the corresponding likelihood functions analytically (see the appendix.) The model has observable variables $\{x^j\}$ directly relatable to the observations $\{y^j\}$ at times $\{t^j\}$ of the system's state $\mathcal{Y}^{t_j}$, which represents the system as a whole in an abstract manner, latent or emerging bulk variables $\{z^j\}$ that are not observed, and parameters α. Examples for x and z are the blood stream's glucose, and the insulin content and associated biophysical processes respectively. We propose to estimate the model's parameters and the system's current state, and forecast its evolution under candidate interventions, through a methodology that includes the following ingredients:

1. A quantification of the agreement between the $\{x^j\}$ and $\{y^j\}$ that is not only point but also distribution-wise. This addresses in particular the pitfall described above of wrongly approximating an oscillatory signal by its mean: a constant value may be a model's local minimizer of the point-wise squared distance to an oscillatory signal, but not in a distributional sense.
2. A slow modulation over time of said distribution and of the model's parameters α, to capture non-stationarity. In other words, we allow both the distribution and the parameter values to vary over time (or flex) within the estimation window to some degree.
3. A three staged methodology that first initializes the model's parameters α and the observable variables $\{x^j\}$, then maximizes an objective function only over the latent variables $\{z^j\}$, and only then performs a full-fledged maximization of the objective function over all variables and parameters. The reasons for such a staged approach are that, without sensible model parameter values, it is hopeless to try to estimate the latent variables z and, without sensible values for z, any use of the model to estimate x and α is equally doomed. Thus the first estimate for the x is based only on their closeness to the observations y

(point and distribution-wise), the first estimate for the model parameters α depends only on the observable variables x, not the latent z, and the first estimation for z is based on the dynamical model using those temporarily estimated values for x and α.

4. A way to handle interventions not included in the model and potentially not fully measured. No bio-physical model can capture all possible interventions, which in our motivating example may range from regular interventions, such as food intake and administered insulin, to more individual ones, such as drug administration or unexpected external developments. For those interventions not explicitly accounted for in the model, we allow the model's state and parameters to behave discontinuously at the intervention times, where the jumps' permissible amplitude depends only on an estimate of the intervention's intensity. This approach permits handling not only external interventions but also more general regime changes.

In order to account for this multiplicity of requirements, we propose an objective function that is additive over various components, described in section 2.

2. A new multiobjective approach to DA smoothing and initialization

We propose an objective function L built as a weighted sum of components L_k, each serving a different purpose. Recall that $\{x^j\}$ represent the observable variables, $\{z^j\}$ represent the unobserved, latent variables, and $\{y^j\}$ represent the observations. Then,

- L_1 quantifies the point-wise agreement between the $\{x^j\}$ and the $\{y^j\}$.
- L_2 quantifies distribution-wise agreement between the $\{x^j\}$ and the $\{y^j\}$, in terms of a slowly modulated invariant measure. L_1 and L_2 are the only components of L involving the $\{y^j\}$.
- L_3 and L_4 quantify the agreement of variables and parameters with the model provided, with two conditional distributions, one forecasting $\{x^j\}$ and the other the $\{z^j\}$ (In our model, x and z are conditionally independent given the prior state and the value of the parameters.)
- L_k , with $k = 4 + l$ and $l = 1, \ldots n_p$, penalizes variations of the parameter α_l over time, where n_p is the number of parameters estimated over their respective estimation ranges.

Notably, while L_1 and L_2 are reasonably standard, L_k for $k > 2$ represent novel deviations from standard methods.

2.1. Defining data with measurement functions

Data are accumulated by measuring the underlying system at times $\{t^j\}$ where j indexes the ordering of these data. Measurements $\{y^j\}$ are taken according to the measurement function h operating on the unknowable state of the system at time t^j, denoted by $\mathcal{Y}^{t^j}$, according to:

$$y^j = y\left(t^j\right) = h\left(\mathcal{Y}^{t^j}\right) + \eta^{t^j}, \tag{1}$$

where η represents noise. More generally, a measurement y is drawn from a probability distribution

$$y^j \sim \rho^h\left(y|\mathcal{Y}^{t^j}, t^j\right).$$

These measurement functions, h or ρ^h, can have dynamics of their own and be represented by dynamical systems or stochastic processes. In biomedicine broadly and biomedical informatics specifically, the measurement process are part of the health care process [14]. Human health can be non-stationary on multiple time scales, and patients are often only measured for a reason, but sometimes are measured according to clinical protocols [39, 16]. Such processes can have very complex missingness properties [28, 27]. Additionally, measurement processes can impact effective time parameterizations [18], and lead to signals being present due to inadvertently combining differently measured, statistically different processes that originate from the same source [6]. The net sum is that it can be important to conceptualize measurement processes as dynamical systems or stochastic processes with potentially complex properties and interdependencies.

2.2. Point-wise agreement: L_1

We would like our x^j to be close to the y^j. Typically, one would translate this into the minimization of a loss function, such as

$$\frac{1}{n}\sum_j \left(y^j - x^j\right)^2.$$

More generally, we can have a probability

$$\rho^o(y|x)$$

of observing y under current state x. Conceptually, this is the result of composing two probability distributions, one for the actual state $\mathcal{Y}$ given x, and the other for the observation y given $\mathcal{Y}$:

$$\rho^o(y|x) = \int \rho^h(y|\mathcal{Y})\rho^m(\mathcal{Y}|x)\ d\mathcal{Y}.$$

Then we maximize the log-likelihood of these data:

$$\max_x L_1 = \frac{1}{n}\sum_j \log\left[\rho^o\left(y^j|x^j\right)\right].$$

Minimizing the sum of squared differences corresponds to a Gaussian assumption on ρ^o. Given measurement processes, this assumption may or may not be valid, but we will not directly address potential related issues here. Even so, this objective function is likely to be too restrictive given the realities of data, as it requires each x^j to be close to the observed y^j: some y^j may be outliers that we do not necessarily want our x^j to adjust to and, more generally, the model may have a hard time adjusting so as to pass near all observations. To account for this, we mollify ρ^o, writing

$$\max_x L_1 = \frac{1}{n}\sum_j \log\left[(1-\epsilon)\rho^o\left(y^j|x^j\right) + \epsilon\rho^0\left(y^j\right)\right],$$

where $\epsilon \sim O(\frac{1}{n})$ and

$$\rho^0\left(y\right) = \frac{1}{n}\sum_l K^y\left(y, y^l\right)$$

is a kernel-based estimation of the probability of y lacking any accompanying x. For our experiments, we have adopted for K^y a Gaussian kernel with bandwidth determined by the rule of thumb:

$$h = \left(\frac{4\sigma}{3n}\right)^{\frac{1}{5}},$$

where n is the number of observations and σ their empirical standard deviation.

2.3. Invariant measure agreement: L_2

A more global constraint on the $\{x^j\}$ is that their invariant measure ρ^x be close to the ρ^y underlying the $\{y^j\}$.

There is more than one sample-friendly way to enforce the equality of $\rho^x(z)$ and $\rho^y(z)$. A weak formulation for their equality is that

$$\int F(z)\rho^x(z)\ dz = \int F(z)\rho^y(z)\ dz$$

must hold for all measurable test functions F, with sample-based representation

$$\sum_j F\left(x^j\right) = \sum_j F\left(y^j\right).$$

Yet we do not really need to try all possible test functions: the choice $F(z) = \rho^x(z) - \rho^y(z)$ yields

$$\int F(z)\left(\rho^x(z) - \rho^y(z)\right)\ dz = \int \left(\rho^x(z) - \rho^y(z)\right)^2\ dz,$$

a non-negative quantity that vanishes if and only if ρ^x and ρ^y agree. We implement this through a simple kernel density estimation:

$$\rho^x(z) = \frac{1}{n}\sum_j K^y\left(z, x^j\right), \quad \rho^y(z) = \frac{1}{n}\sum_j K^y\left(z, y^j\right),$$

with bandwidths given by the rule of thumb as above. In order to allow the measures to evolve slowly over time, we use conditional distributions instead,

$$\rho^x\left(z|t\right) = \frac{1}{\sum_l K^t\left(t, t^l\right)}\sum_l K^y\left(z, x^l\right) K^t\left(t, t^l\right),$$

$$\rho^y\left(z|t\right) = \frac{1}{\sum_l K^t\left(t, t^l\right)}\sum_l K^y\left(z, y^l\right) K^t\left(t, t^l\right),$$

where the kernel K^t has a bandwidth T_l to be determined below, which quantifies the regime modulation's timescale. Then we propose

$$L_2 = -\frac{1}{n}\sum_j F\left(x^j, t^j\right) - F\left(y^j, t^j\right), \quad F(z,t) = \rho^x(z|t) - \rho^y(z|t).$$

2.4. Agreement with the model and parameters' modulation: L_3, L_4, L_{4+l}

The maximizations above were carried out over the $\{x^j\}$ alone, with no reference made so far to the model's parameters and its latent variables $\{z^j\}$. In order to account for these, we add the likelihood of the transition probabilities

$$L_3 = \frac{1}{n}\sum_j \log\left[\rho^x\left(x^j|x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j\right)\right],$$

$$L_4 = \frac{1}{n}\sum_j \log\left[\rho^z\left(z^j|x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j\right)\right].$$

Here

$$\Delta t^n = t^n - t^{n-1}$$

and we assume that the model's transitional probabilities ρ^x and ρ^z, depending on parameters α, are given in a closed form. For the examples below, we have adopted a general model for oscillatory behavior, described in the next subsection.

In order to allow the model's parameters α_l to evolve slowly over time, we introduce a transitional probability for them too:

$$L_{4+l} = \frac{1}{n}\sum_j \log\left[\rho^l\left(\alpha_l^j|\alpha_l^{j-1}, \Delta t^j\right)\right],$$

for which we propose

$$\alpha_l^{j+1} \sim N\left[d_s^j\alpha_l^j + \left(1-d_l^j\right)\tilde{\alpha}_l, \left(1-d_l^j\right)\sigma_l{}^2\right].$$

Here $\tilde{\alpha}_l$ is a prior value of α_l we would like the model to relax to when measurements take long to arrive –this parameter can be either externally provided, estimated or made part of the optimization–, σ_l is the uncertainty in the parameter, and d_l^n is a decaying weight starting at 1 and ending up at 0:

$$d_l^j = e^{-\frac{\Delta t^j}{T_l}},$$

with the same long time-scale T_l adopted for the modulation of the invariant measure. While we choose a generic decay, the decay could be adapted according to a particular application, or devised based on external knowledge.

2.5. Canonical oscillatory model

In section 1.1 we posed a motivating problem of estimating a system that oscillates but that is measured sparsely in time, leading to plausible model solutions that have qualitatively wrong dynamics. This motivational example included two notable assumptions. *First*, we assumed data were sampled sparsely enough that dynamical properties such as the oscillatory frequencies would be difficult to exactly estimate. *Second*, we assumed a highly specific parameterized family of models, an ODE model of glucose-insulin mechanics [35], to estimate these oscillatory data. The parameterized nature of mechanistic ordinary, partial, or stochastic differential equation-based models have an underlying rigidity that yields benefits such as interpretability and estimation with relatively little data. Such models also have drawbacks such as solutions that minimize least squared errors that lie at bifurcation points or produce qualitatively wrong dynamics.

To address these problems, we proposed a new method for estimating parameterized models. This new procedure assumes the availability of a parameterized stochastic model of the dynamics underlying x and z. Additionally, the model must be a parameterized family whose transitional probabilities are given in closed form for which we can readily compute derivatives. For the experiments of this article, rather than adopting a highly-specific, complex, and biophysically motivated model as was used in the opening example, we developed a generic model for oscillatory behavior, in the spirit of the simple stochastic system of equations

$$\begin{aligned} dx &= (\omega\ z - \gamma\ x)\ dt + \sigma\ dW \\ dz &= -\omega\ x\ dt, \end{aligned}$$

where ω models the frequency of the oscillations and σ their variability, which together with the damping parameter γ determines their mean amplitude. This model can have both oscillatory and fixed point mean dynamics depending on model parameters, and importantly, oscillatory frequencies that are relatively rigid, limited, and parameter-dependent, mimicking the rigidity of parameterized families of, e.g., mechanistic ODE models. This model also comes with the additional benefit of allowing us to test how much of the dynamics present in the data we can capture with a generic, simple, but still interpretable model. Our methodology requires explicit expressions for the distributions of x^{j+1}, z^{j+1} in terms of their values at the previous observation time t^j, which could be obtained by solving the system above. Instead, we

propose an alternative system, which is given directly by a discrete transition probability model. Since oscillatory behavior is most naturally described in polar coordinates, we introduce the notation

$$r^n = \sqrt{\left(x^n - b^n\right)^2 + \left(z^n\right)^2}, \quad \theta^n = \arg\tan\left(\frac{z^n}{x^n - b^n}\right),$$

where b represents the local mean around which x oscillates. Then we propose the following transition probabilities:

$$x^{n+1} \sim N\left[b^{n+1} + r_+^{n+1}\cos\left(\theta^{n+1}\right), \sigma^2\right],$$

$$z^{n+1} \sim N\left[r_+^{n+1}\sin\left(\theta^{n+1}\right), \sigma^2\right],$$

where

$$r_+^{n+1} = \left(1 - d_s^n\right) a^{n+1} + d_s^n \; r^n, \quad d_s^n = e^{-\frac{\Delta t^n}{T_s}}$$

models relaxation of r toward the local oscillation amplitude a, with a time-scale T_s typically shorter than the modulation scale T_l, and

$$\theta_+^{n+1} = \theta^n + \omega^n \Delta t^n.$$

models a linear evolution of the phase θ at the local frequency ω.

Thus we pose two time-scales: T_s and T_l, for "short" and "long." The first is a component of the model, quantifying how fast the current state is forgotten. The second is the time-scale over which the model itself changes, through the slow modulation of its parameters.

2.6. Extension to account for regime shifts and abrupt interventions

Next we extend the methodology to systems with drivers, such as nutrition consumption driving the evolution of blood glucose, that are not explicitly incorporated into the dynamical model. We model these drivers as acting instantaneously, through impulsive forcings at known times k_j with specified intensities I_j. Rather than modeling the effect of the forcing on the system directly, we will think of it as a break that partially decouples the system's state and its ruling parameters before and after the intervention, with the strength of the decoupling proportional to the intervention's intensity. This applies more generally to sudden regime shifts, not necessarily driven by an external driver.

Since most of the coupling between successive times in our model depends exponentially of the interval Δt_j, we can weaken this coupling by increasing the effective interval to

$$\Delta t_i \to \Delta t_i + \alpha I_j$$

whenever the time k_j falls within the interval Δt_i (We need to take care to keep the original Δt_i when they multiply the frequency ω, as this corresponds to the phase of the oscillation evolving over time.) Our choice for the parameter α is

$$\alpha = \frac{T_s}{\tilde{I}},$$

where $\tilde{I}$ is a typical intensity, and T_s is the time scale over which the dynamcs decouples naturally.

This decoupling also affects the kernel K^t, where the argument $t_i - t_j$ must be replaced by

$$|t_i - t_j| + \alpha \sum_l I_l,$$

where the sum is over all interventions with k_l falling between t_i and t_j.

A posteriori, we can use the resulting jumps in the various variables at the intervention times, to tune a model that predicts the distribution of these jumps as a function of the driver's parameters. More generally, this suggests an additional potential use of our model: to translate the raw data into a coherent evolution of a set of meaningful parameters –typically slow and punctuated by jumps– that can facilitate further parameterizations of the process under study in terms of more detailed external factors not used by the current model. For example, this machinery could be used to estimate the jump distributions for a jump-diffusion model.

2.7. Computational implementation procedure

Putting together all the ingredients above, we propose the following procedure. We are given a set of observations

$$y^j = y\left(t^j\right), \quad j \in \{0, 1, \ldots, n\}$$

and a model for the transitional probability between states

$$\rho^x\left(x^{j+1}\middle|x^j, z^j, \Delta t^{j+1}, \alpha^{j+1}\right), \quad \rho^z\left(z^{j+1}\middle|x^j, z^j, \Delta t^{j+1}, \alpha^{j+1}\right), \quad \rho^l\left(\alpha_l^j|\alpha_l^{j-1}, \Delta t^j\right).$$

Here the $\{x^j\}$ are the model's surrogate for the observations $\{y^j\}$, the $\{z^j\}$ a set of unobserved, latent variables, and $\{\alpha_l^j\}$ the model parameters, $l = 1, \ldots, n_p$.

We introduce two kernel functions: $K^y(\cdot,\cdot)$, with bandwidth tuned to the observed y's, and $K^t(\cdot,\cdot)$, with bandwidth T_l associated to a slow modulation of the model's parameters and of the invariant measure. We adopt the simplest default option: Gaussian kernels, with bandwidth T_l for K^t, and determined for K^y by the rule of thumb applied to the $\{y^j\}$:

$$K^y(x_1,x_2) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{(x_2-x_1)^2}{2h^2}}, \quad h = \frac{\sigma}{n^{\frac{1}{5}}}, \quad \sigma = \sqrt{\frac{1}{n}\sum_j (y^j - \bar{y})^2},$$

$$K^t(t_1,t_2) = \frac{1}{\sqrt{2\pi}T_l} e^{-\frac{(t_2-t_1)^2}{2T_l^2}}.$$

We propose a multiobjective function $L = \sum_m \lambda_m L_m$ with components

$$\begin{aligned}
L_1\left(\{x^j\}\right) &= \frac{1}{n}\sum_j \log\left[(1-\epsilon)\rho^o\left(y^j|x^j\right) + \epsilon\rho^0\left(y^j\right)\right], \quad \rho^0(y) = \frac{1}{n}\sum_l K^y\left(y,y^l\right), \\
L_2\left(\{x^j\}\right) &= -\frac{1}{n}\sum_j F\left(x^j,t^j\right) - F\left(y^j,t^j\right), \quad F(z,t) = \rho^x(z|t) - \rho^y(z|t) \\
&\quad \text{with} \;\; \rho^{x/y}(z|t) = \frac{1}{\sum_l K^t(t,t^l)}\sum_l K^y\left(z,x^l/y^l\right)K^t\left(t,t^l\right), \\
L_3\left(\left\{x^j,z^j,\alpha^j\right\}\right) &= \frac{1}{n}\sum_j \log\left[\rho^x\left(x^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)\right], \\
L_4\left(\left\{x^j,z^j,\alpha^j\right\}\right) &= \frac{1}{n}\sum_j \log\left[\rho^z\left(z^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)\right], \\
L_{4+l}\left(\left\{\alpha_l^j\right\}\right) &= \frac{1}{n}\sum_j \log\left[\rho^l\left(\alpha_l^j|\alpha_l^{j-1},\Delta t^j\right)\right].
\end{aligned}$$

We initialize the states and parameters as follows:

1. Set $x^i = y^i$ and $z_i = 0$.
2. Set $\tilde{b} = \frac{1}{n}\sum_i y^i$, and initialize b^i through kernel regression:

$$b^i = \frac{\sum_j y^j K^t(t^i,t^j)}{\sum_j K^t(t^i,t^j)},$$

   settting σ_b to the standard deviation of the $\{y^i\}$.

3. Correspondingly, set

$$a^i = \frac{\sum_j \hat{a}^j K^t\left(t^i, t^j\right)}{\sum_j K^t\left(t^i, t^j\right)}, \quad \hat{a}^i = \max_{|t^j - t^i| < T} \left|y^j - b^j\right|,$$

and $\sigma_a = \sigma_b$. Set either $\tilde{a} = 0$, so that oscillations die out in the absence of observations, or to the mean of the $\hat{a}^i$.

4. Regarding ω^i, we first compute $\tilde{\omega}^j = \frac{\pi}{P^j}$, where P^j is the time interval between two sign changes of $y^i - b^i$ that includes t^j, and then propose

$$\omega^i = \frac{\sum_j \tilde{\omega}^j K^t\left(t^i, t^j\right)}{\sum_j K^t\left(t^i, t^j\right)}.$$

Set $\tilde{\omega}$ to the average of the ω^i, and $\sigma_\omega = \tilde{\omega}$.

5. The natural unit for the time-scales T_s and T_l is the average period $P = \frac{2\pi}{\omega}$ of the oscillations. We will adopt

$$T_s = \frac{2\pi}{\tilde{\omega}}, \quad T_l = 4\frac{2\pi}{\tilde{\omega}}.$$

6. The parameter $\epsilon \ll 1$ to account for outliers is somewhat arbitrary. We will pick

$$\epsilon = 0.1.$$

7. For the standard deviations σ and σ_0, we adopt

$$\sigma = \bar{a}, \quad \sigma_0 = 0.1\ \sigma.$$

We can pre-compute the pairwise kernels

$$K^y_{ij} = K^y\left(y^i, y^j\right), \quad K^t_{ij} = K^t\left(t^i, t^j\right),$$

the reference densities

$$\rho^i_0 = \frac{1}{n}\sum_j K^y_{ij}, \quad \mu^i = \frac{1}{n}\sum_j K^t_{ij},$$

and the factors for exponential decay,

$$d^n_l = e^{-\frac{\Delta t^n}{T_l}}, \quad d^n_s = e^{-\frac{\Delta t^n}{T_s}}.$$

After the initialization above, most of the model parameters should be roughly in the neighborhood of their optimal values, since the $\{x_i\}$ should be close to the observations $\{y_i\}$, and the other parameters were roughly tuned to the data. The exception however are the $\{z_i\}$, set initially to non-informative values. Starting the full descent process at once from this initialization has a number of problems:

1. The model's parameters will rapidly deviate from their initial values, since the current $\{z_i\}$ are not consistent with them. So will the $\{x_i\}$, unless the L_1 and L_2 hold enough strength to keep them in place.
2. The variance σ of the models need to be set quite high, for otherwise the current $\{z_i\}$ and $\{x_i\}$ will tend to fall far in their tails, unbalancing the descent process.

To address these issues, we divide the algorithm into three phases, where the choice of weights $\{\lambda_m\}$ depends on the algorithm's phase:

Phase 1: Ascent over x alone, with only $\lambda_{1,2} \neq 0$, with user-provided values depending on the weight assigned to point-wise versus distribution-wise agreement with the data.

Phase 2: Ascent over z alone, setting first λ_3, then only $\lambda_{3,4} \neq 0$, so that after this phase, all of $\{x^j, z^j, \alpha_l^j\}$ have values roughly consistent with both the data and the model.

Phase 3: Ascent over all of $\{x^j, z^j, \alpha_l^j\}$, with $\forall m\ \lambda_m \neq 0$, with values assigned by the user according to a utility-based criterion.

The gradient ascent of L adopts the form

$$x_{k+1}^j = x_k^j + \eta^k \frac{\partial L}{\partial x^j},$$

$$z_{k+1}^j = z_k^j + \eta^k \frac{\partial L}{\partial z^j},$$

$$\alpha_{l\,k+1}^j = \alpha_{l\,k}^j + \eta^k \frac{\partial L}{\partial \alpha_l^j},$$

with learning rates $\{\eta^k\}$ determined by line search at each step k.

3. Numerical results

3.1. Synthetic data with known underlying ground truth

Before considering applications to glycemic dynamics, we demonstrate the methodology on a synthetic example where the ground truth is known.

We extract noisy measurements from an oscillatory signal generated through the following model:

$$x(t) = b(t) + r(t)\cos\left(\omega(t)\ t\right), \quad t \in [0, 80],$$

where

$$b(t) = 1 + 0.2\cos\left(2\pi\frac{t}{200}\right), \quad \omega(t) = 1 + 0.2\cos\left(2\pi\frac{t}{100}\right),$$

$$dr = (a(t) - r(t))\ dt + 0.01\ dW, \quad r(0) = a(0), \quad a(t) = 1 + 0.2\sin\left(2\pi\frac{t}{100}\right).$$

We obtain one realization of $x(t)$ simulating this model through Euler–Maruyama with $\Delta t = 0.1$. In order to extract noisy observations $\{y_i\}$, we select 29 random times $\{t_i\}$ in $[0, 80]$, and write

$$y_i = x\left(t_i\right) + \epsilon_i, \quad \epsilon_i \sim \mathrm{U}\left[-0.4, 0.4\right].$$

The simulated signal $x(t)$ and the extracted noisy observations are displayed on the upper-left panel of figure 5. Notice that:

1. The model used for creating these data, a simple made-up model for modulated oscillations without latent variables z, does not agree with the dynamical model used by our algorithm, so it serves as a check that we do not need to know the true underlying dynamics in order to process a signal correctly. Nonetheless, the model's parameters $\omega(t)$, $a(t)$ and $b(t)$ are meaningful for any oscillatory signal and can be immediately related to the corresponding time-dependent parameters extracted by the algorithm.
2. The 29 sample times for roughy 13 oscillation periods constitute a very sparse signal, near the Nyquist limit of two observations per period. Such sparse, noisy and irregularly sampled observations of a rapidly modulated oscillation over a short time-window are far below the threshold for resolving the signal though conventional means.

The remaining panels in figure 5 display the results of our procedure with input consisting exclusively of the 29 points $\{y_i\}$. On the upper-right panel, the reconstructed $x(t)$ and $z(t)$. Notice that the reconstructed $x(t)$ is qualitatively correct, that it agrees almost exactly with the true underlying signal

in areas where the Nyquist limit is surpassed, and that it yields a signal consistent with the observations and with slightly smaller frequency than the ground truth in areas with less than two observations per period. We can see in more detail in the lower-left panel how the extracted $\{x_i\}$ do not agree exactly with the noisy $\{y_i\}$; in fact, in this example they always lie between the $\{y_i\}$ and the true underlying signal. Finally, the lower-right panel contrasts the true values of the three underlying, time-dependent parameters with their estimation by the algorithm. We can see that the local mean $b(t)$ agrees quite precisely with the ground truth, that the amplitude $a(t)$ exceeds the true one by about 15%, which is consistent with the observational noise added to the signal, and that the frequency $\omega(t)$ under-estimates the true underlying one by about 25%, a lower frequency resulting from the sparsity of the observations, as seen in the panel above.

3.2. Measurement functions for data and reanalysis data from the glucose-insulin system

The existence of data or measurements are defined with measurement processes or functions. In some cases, such as experimental settings, measurement functions may be simple but in other cases such as where data are collected operationally for an application, the measurement functions may be complex. One natural conceptualization of a measurement function is as a stochastic process that controls when and how measurements are taken. Here the measurement process is one aspect of the *health care process*[14]. The times at which data are collected can be pre-established, random or correlated to the data or some other process [18, 27, 28]. Because we are concerned with the impacts of sparsity and measurement noise, we will define three distinct measurement functions in the context of the glucose-insulin system that specify the measurement times for use in our experiments as:

1. h_1 where measurements are taken as directed by a human according to need or a protocol [16, 30, 31, 39], corresponding to the subject's real, finger-stick or IV measurement times;
2. h_2 where measurements are taken at random times where the difference between two consecutive measurements is uniformly distributed over $[60, 90]$ minutes;
3. h_3 where measurements are taken every five minutes, simulating continuous glucose monitor (CGM) measurements.

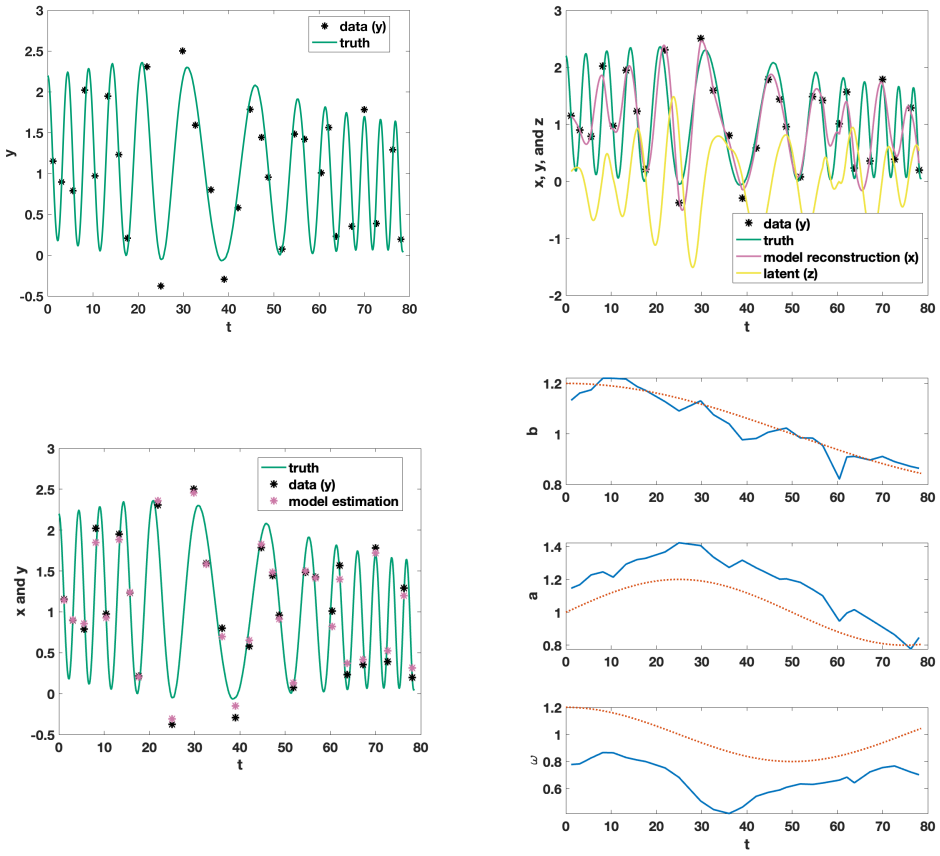


Figure 5: Results using sparse noisy observations of a known ground truth (an oscillatory signal.) Upper left panel: true signal (continuous green line) and noisy observations y (black stars, the only input provided to the algorithm); upper right: true signal (continuous green line), noisy observations y (black stars) and reconstructed signal (x in purple, z in yellow); lower left: noisy measurements y (black stars) and estimated x (purple) over the true signal (green); lower right: true underlying parameters b, a and ω (dotted red line) and their estimation (in solid blue).

In case where CGM data are available, h_3 will be the measurement times defined by the CGM and h_2 will be a subsampling of these measurement times. In the case where we do not have CGM data, h_2 and h_3 will be extracted from reanalysis data or continuously simulated data from the previously estimated ultradian ODE model.

3.3. Real and simulated data

To evaluate our method's performance with different generating dynamics and sampling patterns, we will use two different sources of real and simulated glucose-insulin data. While we chose blood glucose dynamics because estimating them was the initial motivation for this work, we also think that glycemic dynamics are varied and represent many other situations. To add to the generalizability of the context, we will focus on two cases of glycemic dynamics, *(i)* tube-fed patients in the ICU and *(ii)* normal patients in the "wild". The blood glucose levels of intensive care unit (ICU) patients are often oscillatory (chaotic or stochastic with some strong periodic frequencies present) when the patient is given constant nutrition [35] and are highly non-stationary because of the effects of critical illness and related interventions. The blood glucose levels of patients in the wild are more akin to damped-driven oscillations with noise because, e.g., nutrition consumption causes a rapid increase in blood glucose followed by an oscillatory and noisy return to glycemic homeostasis.

3.3.1. Real world data

The first data set includes ICU data extracted from the Columbia University Medical Center Clinical Data Warehouse for a previous study [31, 4, 38, 5, 30]. All the patients in this dataset were fed through an enteral tube and were delivered intravenous (IV) insulin for glycemic management. We selected one patient, patient 593, as a representative for highly non-stationary glycemic behavior and challenging model estimation. We used this patient's data in Section 1.1 as one of our motivating examples. These data consist of the point-of-care blood glucose measurements, carbohydrate records delivered through the enteral tube, and exogenous IV insulin records. Additional details about the data are presented in Table 1.

The second data set was collected from a healthy subject representing "normal" glycemic dynamics observed in the "wild", used in several previous studies [2, 3, 21, 22, 23, 10]. The words "in the wild" refer to the fact that this

individual measured their nutrition and glycemic dynamics while living otherwise normally. We selected this subject's data as a representative of data collected in usual living conditions and because the dynamics are distinctly different. In particular, glycemic dynamics of individuals not being continuously fed resemble a noisy damped, driven oscillator with a noisy glycemic equilibrium. These data consist of sparse finger stick BG measurements, continuous glucose monitor (CGM) data, and carbohydrate intake of meals and snacks reviewed and verified by nutritionists. We provide additional details in Table 1. The data sets vary according to different measurement functions. For measurement function h_1 we restrict the model-estimated data to the finger-stick measurements taken by the patient while data corresponding to h_2 are randomly down-sampled from the CGM data, and data corresponding to h_3 are the raw CGM data.

3.3.2. Simulated and reanalysis data

In addition to the real data, we used simulated reanalysis data to evaluate this methodology's performance for the ICU patient because this patient did not have CGM data and we evaluate the estimation algorithm assuming different data sampling patterns via the different measurement functions. Note that reanalysis data are simulated data drawn from a model that was estimated with ICU data, in this case, the ultradian model estimated with constrained interior point methods. Because of the problems outlined in the motivation section, we selected an optimized parameter set to simulate that produced oscillatory solutions. Meaning, the model parameter set used to generate the simulated data was not optimal, but rather hand chosen for purposes of validating our methodology, noting that our methodology was not applied to estimating the ultradian model in any way. A primary benefit of this approach, aside from being required because of the lack of CGM data in the ICU case, is that simulated data allow us to have a completely knowable ground truth for at least one of our evaluative cases, even if the model's representation of these data is not ideal.

High level computational workflow for generating simulated data. We estimate glucose by fitting plasma glucose and three parameters (t_p, R_g, a_1) of the Ultradian model [35] while setting the remaining parameters at their default values. We selected these parameters because they control the large subsystems of the model. Specifically t_p controls insulin clearance, R_g controls insulin resistance, and a_1 controls insulin secretion. A detailed descrip-

tion of this model is in the Appendix. For the relatively stationary data from the patient in the wild we used standard Metropolis Markov Chain Monte Carlo (MCMC) with three chains and 10,000 iterations for each chain. Not every chain converges, so we only use the empirical density functions of the converged chains to sample parameters to create the simulated data. We consider chains converged using the Geweke statistic [12, 11] and minimizing mean squared error, following the methodology presented in [38]. For the severely nonstationary ICU patient, we used an ensemble Kalman filtering (EnKF) with an ensemble size of 100 similar to what was used in [4]. We used the EnKF because the parameter values move too much over the week-long time window we used to estimate the model. In both cases we then apply the measurement functions h_1, h_2, and h_3 to generate the data sets we estimate with our new method.

ICU data generation. We used one week of data collected from ICU patient 593 from our opening example. These data are documented in Table 1. The average parameter values estimated over this period were $t_p = 5.5$, $a_1 = 7.5$, and $R_g = 225$. We then simulated blood glucose data every minute with these parameters and subsampled the data according to h_1, h_2 and h_3.

In-the-wild patient data generation. We did not use simulated data for the "in the wild" patient because we have CGM data to compare against the model estimates for different measurement functions.

3.4. Design of numerical experiments for the methodological evaluation

The DA methods we propose here focus on situations where the parameters that minimize a single error function such as least squared may not be the model parameters we would most like to use. Additionally, because the multiobjective loss function minimizes a weighted sum of errors, the optimal parameters our method selects *may not minimize any of the individual loss functions.* Therefore, standard methods for evaluation such as comparison of mean squared errors will not suffice. Instead, we will evaluate the method's usefulness qualitatively. Specifically, we will consider whether our multiobjective minimization is able to achieve the goals and criteria listed below.

Our method is motivated by two goals in the context of sparse data: *(a)* model estimation that balances point-wise accuracy with the preservation of global properties such as agreement of invariant measures, and *(b)* model estimation flexibility to manage measurement time errors, non-stationarity and

Healthy subject		ICU patients		
			patient 426	patient 593
total data recording interval (days)	31.6	total data recording interval (days)	14.0	21.2
total # of BG measurements	fingerstick:120 CGM:7722	total # of BG measurements	177	249
measured BG values (mg/dL)	fingerstick:103±18 CGM:95±14	measured BG values (mg/dL)	141±18	150±32
total # of meal/ snack recordings	73	total amount of recorded tube-fed carbohydrates (g)	1938.9	2139.1
		total amount of delivered IV insulin (unit)	0	1498.1

Table 1: Patient data descriptions for data used to test and validate our optimization methodology. We show mean±stdev values when appropriate.

sparsity, by allowing the parameters to vary over a fixed temporal estimation window. We claim that our method is able achieve the goals of preserving estimation of the invariant measure while also supporting very accurate point-wise estimation and accounting for external noise, shocks, and errors in measurement times by *(i)* giving up a little accuracy estimating the model both point-wise and in-distribution and *(ii)* allowing parameters to flex over the estimation interval near data points. To verify that our method is able to achieve these goals we designed a series of computational experiments to address the following questions:

Q1. **Can the new method include knowledge of global properties of the system to improve model estimation by allowing estimation of global properties that:**
   a. we either know or can be derived from the entirety of given data?
   b. are compatible with reasonable point-wise data estimates?
   c. lead to more reasonable model estimates from a face-validity standpoint?

Q2. **Does accounting for measurement function dependence and measurement timing errors—specifically measurement density and targeted vs random measurement times—impact:**
   a. the temporal parameter variability or flex over the estimation window;

b. the estimation accuracy of the invariant measures;
c. the estimation accuracy of the trajectory.

Q3. **Does allowing temporal variability of model parameter estimation within an estimation time window:**
a. allow the model estimation to accommodate nonstationarity within the estimation window?
b. vary depending on the measurement function?
c. impact the estimation accuracy of the trajectory?

To address these questions we will estimate two types of dynamics, *(i)* oscillatory dynamics that relax to a periodic orbit when driven and to a fixed point when not driven, and *(ii)* damped driven oscillatory dynamics that relax to a fixed point. Data will be collected via the measurement functions defined in section 3.2.

*3.5. Incorporating global information (*Q1*) and impacts of different measurement functions on model estimation (*Q2*)*

The global information we incorporate into the model here includes the invariant measure of the data and the qualitative orbit type (oscillatory versus fixed point, etc.). We will use the ICU data to show how the new method is able to include global distributional information. We will use the wild case to additionally show how the new method can also be seeded with external knowledge of the underlying dynamics and how this impacts model estimation with sparse data.

Starting with the ICU case where the generating process produces oscillatory dynamics except for a brief time when the oscillations disappear because the nutrition is turned off, we estimated the model that assumed underlying oscillatory dynamics by setting the relaxed state of the model to have a nonzero amplitude. The estimated model dynamics, shown in Fig. 6, reproduced the mean and amplitude of oscillations well while the frequency of oscillations was only correctly represented when the data were measured frequently (h_3). The brief time when the generating processes was not oscillating was only represented with the correct dynamics when the data were measured frequently (h_3). Interestingly, the sparse clinician measurement function, h_1, reproduced the non-oscillatory dynamics substantially more accurately than the random measurement function, h_2, potentially implying information in the clinician-driven measurement times [18, 1, 20, 17, 7] as has been previously hypothesized. It is possible that a specifically glucose-insulin model

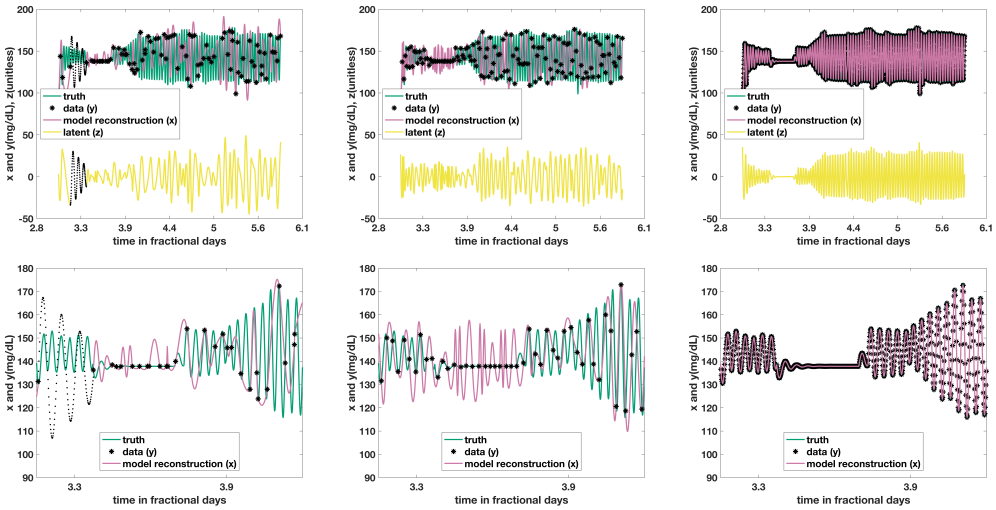


Figure 6: Given the oscillatory ICU glycemic dynamics and the model basic state assuming oscillatory dynamics, we see the model estimating the simulated glucose measured according to h_1 (left), h_2 (center) and h_3 (right). We can see that even for the sparse data cases (h_1, h_2), the model produces oscillatory dynamics with reasonable mean and amplitude while for the densely measured case (h_3) the model tracked these data precisely. The point-wise estimates remain accurate in all cases. The dotted lines signal reconstruction for periods without data longer than a prescribed threshold.

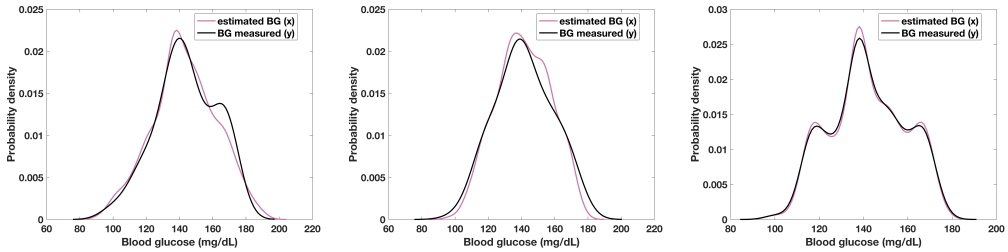


Figure 7: Given the oscillatory ICU glycemic dynamics and the model basic state assuming oscillatory dynamics, we see the model estimating the simulated glucose measured according to h_1 (left), h_2 (center) and h_3 (right). Note *BG measured* denotes data available to the model when it is estimated and *estimated BG* denotes the model-estimated invariant measure of the data. The densely measured case (h_3) is likely the closest representative of a gold standard baseline, again for data measured frequently in time. We can see that even for the sparse data cases (h_1, h_2), the model produced an accurate representation of the invariant measure that was not particularly dependent on the sparse measurement function while for the densely measured case (h_3) the model estimated all the details of the invariant measure well.

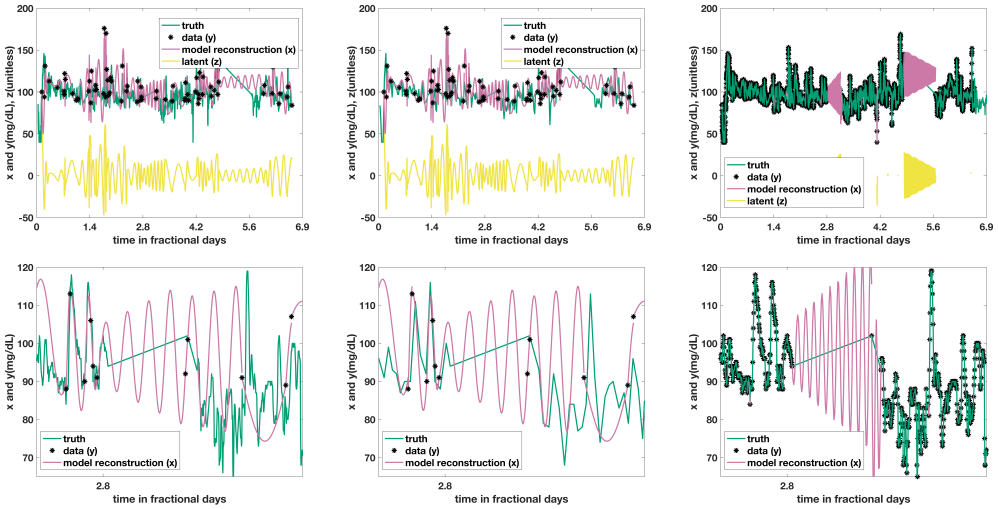


Figure 8: Given the damped, driven wild patient data, the model was estimated assuming an underlying oscillatory solution (set according to section 2.7) for the three measurement functions, sparse clinician measured (h_1, top/bottom left), random (h_2, top/bottom center) and dense (h_3, top/bottom right). We can see the point-wise estimates are good in the lower plots, the model captures the large peaks and troughs but only captures the correct frequency for densely measured data. When data are missing the model relaxes to oscillatory dynamics. Note that the straight lines in green are not the truth but a linear interpolation in the absence of data.

whose oscillations are directly controlled by nutrition would have faired differently. See Figure 7 where the model-estimated glucose distribution is compared with the data used to estimate the model. The invariant measures of these highly non-Gaussian data were reproduced well regardless of the measurement function. Together, Figs. 6 and 7 demonstrate that both point-wise and in-distribution estimation were quite accurate, producing affirmative answers to *Q1a*, *Q1b* and *Q1c*.

Moving to the in-the-wild case where the generating dynamics are noisy, damped, driven—by punctuated nutrition consumption—oscillations, we estimated the model assuming both non-oscillatory and oscillatory (resp. Figs. 9 and 8) dynamics, by controlling the model's amplitude and relaxation parameter α. In the cases where measurement functions produced sparse data, h_1 and h_2, setting up the optimization to assume the underlying dynamics were non-oscillatory lead to better estimation performance compared with setting up the optimization to assumed oscillatory dynamics according to

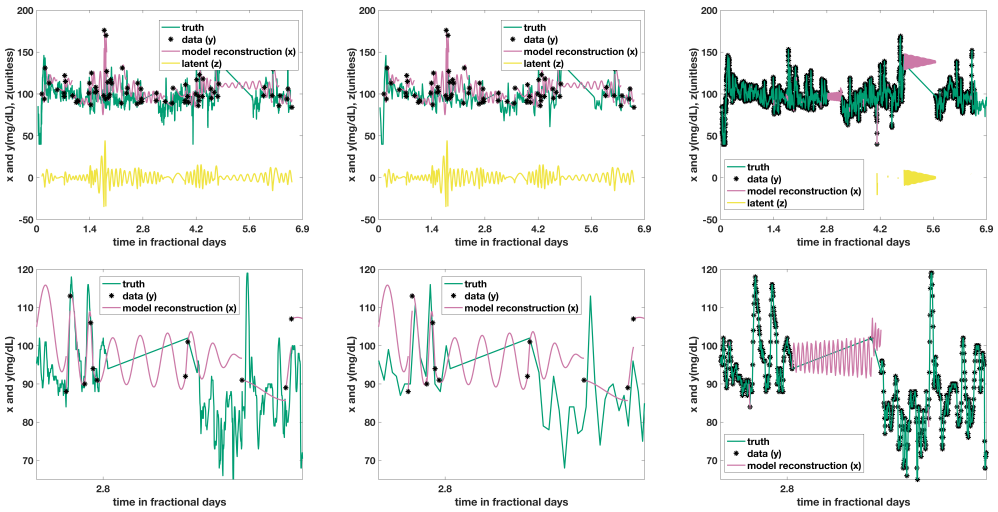


Figure 9: Given the damped, driven wild patient data, the model was estimated assuming no oscillatory solution and relaxed to its non-oscillatory solution controlled by the α parameter and by setting the baseline amplitude a to zero, for the three measurement functions, sparse clinician measured (h_1, top/bottom left), random (h_2, to/bottom, center) and dense (h_3, top/bottom right). We can see the point-wise estimates are good in the lower plots, the model captures the large peaks and troughs but only captures the correct frequency for densely measured data. When data are missing the model relaxes to non-oscillatory dynamics. Again note that the straight lines in green are not the truth but a linear interpolation in the absence of data.

both point-wise, Figs. 8-9, and distributional, Figs. 10-11 metrics. When the data were measured densely with h_3 there was not an appreciable difference between the optimization including information regarding oscillatory and non-oscillatory cases with respect to point-wise and distributional fits. The dependence of the results on the measurement functions highlights the important role measurement functions can play on model estimation. Together these results imply that we can include external knowledge regarding global properties and that their inclusion does improve estimation, supporting an affirmative answer to *Q1a*, *Q1b*, and *Q1c*. Additionally, related to *Q2b* and *Q2c*, the measurement functions clearly impact model estimation performance. Specifically, accuracy of estimation of the off-data trajectory is highly dependent on the measurement function, but including information about the correct underlying dynamics and global properties such as the invariant measure of data does qualitatively improve estimation (*Q2c*). Finally, regardless of measurement function, we are able to accurately estimate the invariant measure of the data (*Q2b*), yet setting the underlying dynamics correctly does improve estimation of the invariant measure of these data when these data are particularly sparsely measured.

*3.6. Impact of temporal variation of parameters over the estimation window to account for measurement time errors and nonstationarity (*Q3*) and impacts of different measurement functions on model estimation (*Q2*)*

We designed our new algorithm to allow parameters to vary within the estimation window according to the L_{4+l} loss function that models the transition probabilities of the l^{th} parameter value between measurements with a decay rate d_l^n and uncertainty σ_l. Modeling and estimating the transition probabilities between measurements was designed to solve two problems, both related to the goal of giving the model enough flexibility to estimate point-wise and global qualitative dynamics with sparse data. *First,* data measurement recording times have errors while mechanistic models such as ODEs have relatively rigid orbits with rigid frequencies that can be tuned, to some degree but not with unlimited flexibility, according to model parameters. Granting parameters limited temporal variability within an estimation window allows the models to have more flexible trajectories that can accommodate measurement error times without resorting to zeroing out the oscillations. *Second,* when data are sparse, to maximize the number of data points included within an estimation window we are forced to estimate the model over longer time periods, increasing the potential impact to nonsta-

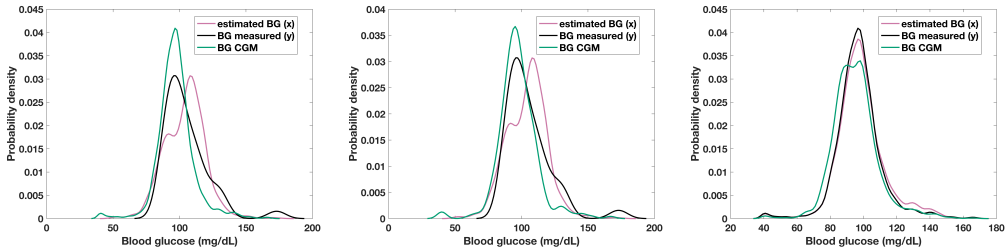


Figure 10: Given the damped, driven wild patient data, the model was estimated assuming an oscillatory solution and relaxed to its oscillatory solution (set according to section 2.7) for the three measurement functions, sparse clinician measured (h_1, left), random (h_2, center) and dense (h_3, right). Note *BG measured* denotes data available to the model when it is estimated and *estimated BG* denotes the model-estimated invariant measure of the data. The densely measured case (h_3) is likely the closest representative of a gold standard baseline, again for data measured frequently in time. We can see the model captures the distributions well. While the clinician-driven measurement times, h_1, may outperform random measurement times, h_2, neither are perfect as they are balanced against the point-wise and model-coherence loss functions. Additionally, the case here where the solution is assumed to be oscillatory did not perform as well as the case where the solution was assumed to be non-oscillatory for the sparse measurement functions h_1 and h_2. When the model is estimated with data from the dense measurement function, the invariant measure of these data is estimated very accurately and was not appreciably different from the case where the model assumed non-oscillatory baseline dynamics shown in Fig. 11.

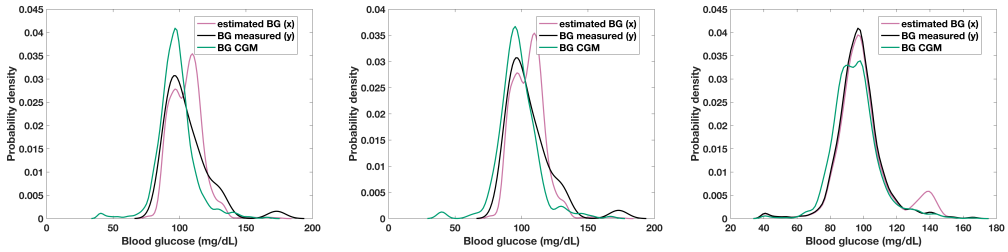
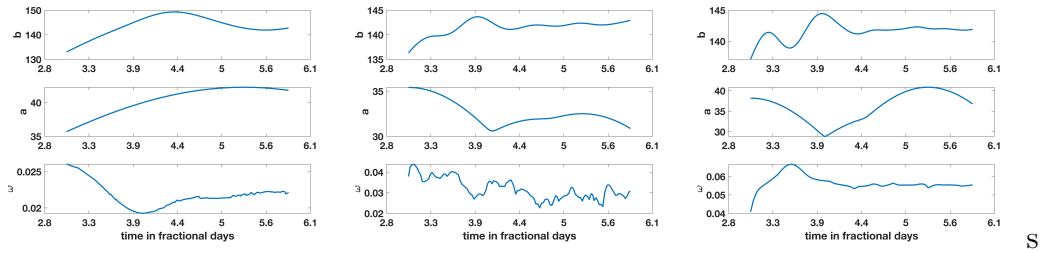


Figure 11: Given the damped, driven wild patient data, the model was estimated assuming no oscillatory solution and relaxed to its non-oscillatory solution (baseline $a = 0$) for the three measurement functions, sparse clinician measured (h_1, left), random (h_2, center) and dense (h_3, right). Note *BG measured* denotes data available to the model when it is estimated and *estimated BG* denotes the model-estimated invariant measure of the data. The densely measured case (h_3) is likely the closest representative of a gold standard baseline, again for data measured frequently in time. We can see the model captures the distributions well. While the clinician-driven measurement times, h_1, may outperform random measurement times, h_2, neither are perfect as they are balanced against the point-wise and model-coherence loss functions. When the model is estimated with data from the dense measurement function, the invariant measure of these data is estimated very accurately.

tionarity. Allowing the parameters to flex over the estimation window can potentially lead to more productive parameter estimation in the presence of this nonstationarity. Here we are considering specifically *Q3*, the impact of nonstationarity and measurement functions on our flexible parameter estimation methods. Additionally and again, the ICU patient has nonstationary physiological mechanics and oscillatory dynamics except when tube-feeding is interrupted while the in-the-wild patient is likely quite stationary with dynamics that are roughly a noisy damped, driven oscillator.

ICU. All measurement functions were observed to impact parameter estimation. The random h_2 and dense, h_3 measurement functions led to estimation of *amplitudes*, a and b, similarly and both differ from the clinician-driven measurement function, h_1. All three measurement functions led to differences in estimation of frequency of oscillations, ω, which is both interesting because the sparse measurement functions differed and expected because only the dense measurement function could truly resolve the frequency. Nevertheless, all measurement functions led to estimation of the amplitude parameters in roughly the same range, which is particularly important for two reasons. *First,* it was the amplitude estimation that was failing in our motivating example, and here all the parameter estimates included oscillatory solutions. *Second*, the most important feature of the blood glucose dynamics for clinical decision-making is the glycemic range defined by the amplitude of oscillations [16], and regarding this, all measurement functions lead to qualitatively similar amplitudes. *Finally*, frequency estimates differ between all measurement functions, as expected given that the sparse measurements contain very little frequency information. Overall, the parameter flex did impact model estimation, and the parameter flex was dependent on the measurement functions.

In-the-wild. In the case where the optimization method was set to assume underlying dynamics were oscillatory, Fig. 13, the sparse measurement functions, h_1 and h_2 led to estimation that produced nearly identical parameter trajectories, and both differed from the model estimated with densely measured data produced by the CGM, h_3. Whether the optimization assumed the underlying dynamics to be oscillatory, Fig. 13, or not, Fig. 14, did not substantially impact either the estimation of the amplitude parameters, or the differences between the sparse and dense measurement function dependence on the amplitude parameter estimation. Not surprisingly, the primary dependencies for parameter estimation for a given type of assumed dynamics,



S

Figure 12: Given the oscillatory ICU glycemic dynamics and the model basic state assuming oscillatory dynamics, we can see that all measurement functions, h_1 (top), h_2 (center) and h_3 bottom, utilized the temporal flexibility in the parameters to decrease estimation error, but in very different ways. Interestingly, the sparse random and dense measurement functions (h_2, h_3) had nearly the same flex in *amplitude* estimation, a and b, sometimes in direct opposition to flexible parameter trends of the clinician (human) measurement case (h_1). All three measurement functions differ in their estimation of frequency of oscillations, ω. This result has particularly important potential implications for solving inverse problems to estimate parameters because it implies that differently measured data, even when the number of data points are similar, can have a substantial impact on the estimability of model parameters.

oscillatory or not, appeared in the estimation of the frequency parameters. *The most important distinction* between parameter estimate trajectories for the in-the-wild case was due to which dynamics the optimization method was set to assume as the underlying dynamics. The parameter variability over the course of the estimate window was substantially higher when the baseline model dynamics were oscillatory because the generating dynamics were not particularly oscillatory, and leading the transition probabilities to vary quite a lot to minimize the point-wise loss function. Overall, the temporal variability of parameters did impact model estimation, and the temporal parameter variability appeared to be dependent on the measurement functions, but these differences differed from the highly non-stationary ICU case.

4. Discussion

Overall summary *We constructed a scalarized multiobjective function and the associated optimization machinery to support model estimation balancing the minimization of point-wise errors with global properties such as qualitative dynamics and errors in measurement times. We demonstrated the effectiveness of this machinery on a variety of oscillatory simulated data, DA output estimated from data or reanalysis, and real word data that are exter-*

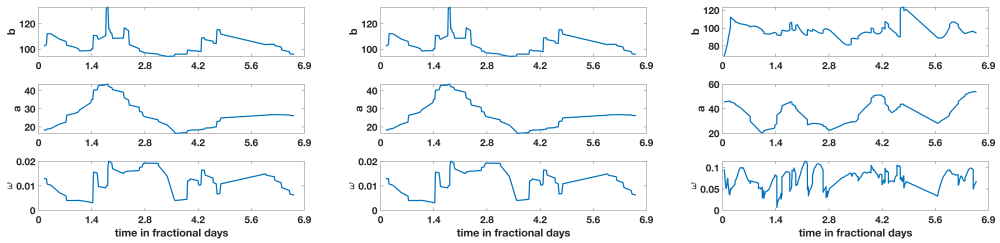


Figure 13: Given the in-the-wild glycemic damped, driven oscillatory dynamics we can see and the model basic state assuming *oscillatory* dynamics, we can see that all measurement functions, h_1 (left), h_2 (center) and h_3 (right), utilized the temporal flexibility in the parameters to decrease estimation error, but in very different ways. Interestingly, the sparse patient-defined and random measurement functions (h_1, h_2) had nearly the same flex over the course of the time window, and both were quite different from the dense measurement case (h_3). This result has particularly important potential implications for solving inverse problems to estimate parameters because differences between parameters estimated with sparse versus dense data can have a substantial impact on the estimability of model parameters. Understanding the degree of importance and uncertainty in the inverse problems setting will be important.

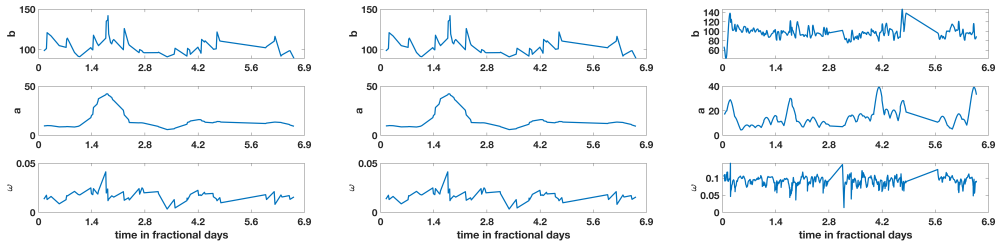


Figure 14: Given the in-the-wild glycemic damped, driven oscillatory dynamics we can see and the model basic state assuming *non-oscillatory* dynamics, we can see that all measurement functions, h_1 (left), h_2 (center) and h_3 (right), utilized the temporal flexibility in the parameters to decrease estimation error, but in very different ways. Similarly to the case where the dynamics are assumed oscillatory, the sparse patient-defined and random measurement functions (h_1, h_2) had nearly the same flex over the course of the time window, and both were quite different from the dense measurement case (h_3). This result carries the same potential implications for solving inverse problems as shown in Fig. 13. Interestingly, the parameter trajectories and exercised parameter flex over the estimate window here is substantially lower than for the case where the underlying dynamics are assumed to be oscillatory in Fig. 13, as expected given that the assumed dynamics in that case were less representative of the generating dynamics.

nally driven and are stochastic/chaotic. Estimation accuracy depended on the measurement processes, but for all measurement functions qualitative dynamics were preserved while point-wise errors were small. Additionally, the added intra-estimation window parameter allowed mechanistic models whose parameters are rigid and otherwise fixed in time to be robust to noise in measurement times, model error, and non-stationarity by allowing the parameters to vary in time. We feel that several of the objective functions were novel, particularly those related to temporal variability of parameters. Additionally, the multiobjective functions we created allowed for the inclusion of knowledge that external to the model and data, and this inclusion of external knowledge, e.g., regarding properties of the invariant measure, were shown to impact and improve model estimation. .

Model estimation in the context of multiobjective optimization. Multiobjective optimization [42, 43, 44] allows for striking a balance between multiple, competing goals mathematized as cost functions which can be particularly important when data have complex properties such as sparsity or nonstationarity. However, the consequence of being able to balance multiple goals is the lack of a clear notion of global optimal solution unless there is a utility function structure such as R2 utility [42]. The lack of a global optima is *implicitly* nearly always present in nonlinear optimization: the least squares minimizing solution is often not the same as the solution that minimizes distributional errors. In the case of multiobjective optimization this dependence of optimal solutions on the cost functions is surfaced because optimal parameters are set according to a utility function that, in our case, amounted to heuristically setting the weights that balance the different log-likelihood functions. This surfacing of the balance of costs that define optimal solutions allows us to be concrete, explicit, and transparent about the choices we make with regard to balancing goals and further quantitatively defines what optimal means through the definition of the utility function.

In our motivational example, one issue that was raised was identifiability failure. Identifiability failure can occur because of data sparsity or because the model is structurally not identifiable even with unlimited data. It is the case of identifiability failure due to data sparsity that motivated the work here. In the current paper, the models we use are structurally identifiable, but may not be identifiable if the data are too sparse. We constructed the paper using a structurally identifiable model that was estimated with different measurement functions that vary data sparsity to investigate how data

sparsity and the ensuing identifiability problems were managed by the multiobjective optimization. While there are many situations where multiobjective optimization will not help or may make identifiability problems worse, in our case, we used the multiobjective function to insert external knowledge of the dynamics of the process to make up for missing data. Specifically, we knew the dynamics were not fixed point, and multiobjective function helped avoid those catastrophically incorrect solutions. Meaning, a multiobjective optimization approach can be used to mitigate identifiability failures due to data sparsity. We do not know whether multiobjective optimization would be helpful in mitigating structural identifiability problems.

Proving convergence guarantees in the multiobjective context is usually difficult and depends on the structure of the utility function. A recent paper formulating multiobjective optimization in the context of R2 utility did have some related convergence results, and detailed what was required of a cost function to be a R2 utility function. In our case, the log-likelihood functions do not conform to the R2 utility formulation. And in this case, model convergence is defined by the raw error defined by the log-likelihood functions.

Interpretative probability framework related to the use of mathematical biology models. A defining characteristic of Frequentist statistics is the assumption that an estimated quantity has a true value, such as the speed of light in a vacuum, and a goal to use a model to compute and quantify error of a model of that quantity. Similarly, a defining characteristic of Bayesian statistics is that an estimated quantity does not have a true value, but rather is a distribution with the goal of estimating a model to estimate the quantity and our uncertainty of that distribution. These frameworks are also associated with computational machinery such as maximum likelihood (MLE) or maximum a posteriori (MAP) estimation for Frequentist or Bayesian frames respectively. Our multiobjective function is built of log-likelihood functions and therefore is computationally a Frequentist framework. However, multiobjective optimization is applicable but independent of either of these frames and importantly, *multiobjective optimization surfaces the dependence between optimal solutions to model estimation problems and the structure of the cost function*—specifically, that there may be more than one optimal solution depending on the goal or cost function. In the case of a physiological models, and certainly for systems physiological models, states and parameters often represent bulk or statistical quantities rather

than absolute quantities such as the speed of light in a vacuum. These bulk quantities may interpretively lend themselves to a more Bayesian interpretation. However, from the Frequentist perspective, it is certainly possible that one model can be structurally more correct than another model of a given system, where the model is defined operationally by state and parameter estimates and their accuracy. Meaning, when estimating models that lack a first principles origin or fundamental physical properties, the interpretation of the model estimation often lives in between ideological frameworks such as Bayesian interpretations that may apply to parameters and Frequentist interpretations that may apply to models, or visa versa. But multiobjective optimization sits beside these frames and ideas and by including more information in the model estimation process, can both increase the robustness of the estimation process, and can serve as a probe to better interpret and understand how the model is estimating data and where its failure points or shortcomings lie. This is particularly important here because many models are estimated with data for a variety of reasons, including forwarding a deeper understanding of the system, to validate or reject mechanistic hypotheses, or more practically to estimate a model that accurately represents the system to guide decision support where physical fidelity *may not may not* be as important as accuracy in forecasting. *The point is that*, because biological and physiological models do not have first principles origins, multiobjective optimization may provide a scientific framework to support both better understanding of estimated models' relationship to data and more robust and accurate model estimation, while remaining flexible enough to incorporate and surface limitations of our interpretations of probabilistic structures.

Data assimilation initialization. Accurate initialization for data assimilation is crucial for immediate and accurate forecasting and for DA use with sequential smoothing and filtering. Our method provides a new path for DA initialization when data are particularly sparse and there is some knowledge of underlying model dynamics and state.

Measurement functions and intra-window parameter trajectories. Computed parameter trajectories were dependent on the measurement functions. It is not obvious how to identify whether a parameter trajectory is correct. In the case where the trajectory does not vary over the estimation window, this is not a problem. Similarly, if the parameter intra-estimation window time variation is only a matter of accounting for noise and the trajectory

variability is not scientifically important, identification of intra-estimation-window trajectories is not a problem. But, if the intra-estimation-window parameter trajectories are scientifically or otherwise important, this opens a new problem of trying to understand and quantify the uncertainty in estimated parameter trajectories as they depend on measurement functions.

Data sparsity and non-stationarity. Data sparsity and non-stationarity are deeply coupled concepts because non-stationarity can induce sparsity. If a system's parameters change substantially and faster than we measure, there will not exist statistically stable data set to estimate the model, and all data sets will be sparse. This induces a balance in estimation error due to use of small data sets against estimation errors due to data sets that are not statistically stable. The intra-estimation window parameter temporal variability was designed to manage this complexity along with allowing a rigid model to be more robust to errors in measurement times.

Managing data sparsity. When data are not sparse, the estimation problem is relatively straightforward, and using both standard methods and our new method, minimizing point-wise errors can reproduce many global properties such as the distribution of model states. Our estimation results show that data sparsity, when the generating dynamics are complex, can have a substantial impact on our ability to accurately estimate the underlying dynamics and model states and parameters. Model estimation with limited data decays as data become increasingly sparse, and it becomes difficult to quantify or identify what information we can reliably determine from sparse data. Our method was developed to help push the boundaries of the information we can extract from data when paired with external knowledge of the system. And while we were able to estimate global and local properties of the system, explicit new open problems could be articulated related to identifying the limits of what information can be extracted and how to compute the uncertainty of that information. In the background, there is surely a *Nyqist-Shannon-type theorem* for data assimilation and mechanistic modeling [4] just as there has been for other spaces such as the reproducing kernel Hilbert space [32, 33]. Even rough numeric guidelines that could help limit the estimation errors given sparse data would be helpful, but such a result does not currently exist to our knowledge.

Managing non-stationarity. The impact of non-stationarity on the measurement functions and parameter estimation was interesting and differed de-

pending on the generating dynamics. When the system was non-stationary, here the ICU patients, it seems that random sparse measurements may lead to parameter estimation that is more representative to the parameters estimated in the densely measured case. In contrast, the case where the system is relatively stationary, here the in-the-wild patient, differences in the sparse measurement functions translated to little variability in the intra-estimation-window parameter trajectories. Of course, two cases are not enough to understand and establish general rules of such measurement function dependent variability other than establishing that there can be important estimation dependencies on measurement functions.

Next steps. There are at least six directions for advancement. *First*, we need to construct a scalable and generalizable computational pipeline such that this methodology can be applied to any ODE and SDE models. For example, because optimizing the L_3 and L_4 components of the objective function require derivatives of the model, integrating the methodology with auto-differentiation would be a meaningful advancement. *Second,* our framework presents an set of variables to optimize over, the weights λ_k attached to the components L_k of the objective function. While in this paper we set these weights heuristically, there is room to further develop a utility function that is generalized and computed automatically. *Third,* while some standard UQ methods apply to our methodology, there are many opportunities for UQ methods development. For example, devising verifiable methods for computing the model parameter estimate uncertainty given sparse data *for the parameter trajectory and a parameter distribution estimated from the parameter trajectory within a given estimation window* would be substantial advancements. *Fourth,* our method allows us to imbue the model with a given underlying dynamic such as oscillatory or fixed point dynamics. In our motivating example, we know from other work, what the baseline underlying dynamics are. However, learning more accurate initial parameters beyond a crude set of parameters that produce, e.g., oscillatory dynamics, could be devised as an iterative application of the methods devised here. *fFfth,* given the motivation for the work in this paper, adapting and applying our methodology directly to modeling patients in the ICU and other similar settings would be highly valuable. And *sixth*, understanding how log-likelihood functions like the ones here could fit into a more formal framework such as the R2 utility would be helpful for better understanding the relationships between the optimal solutions the various cost functions find, and for computing more

reliable solutions.

5. Conclusions

Motivated by an inability to robustly estimate model parameters to initialize a DA-filter given sparse, non-stationary, noisy data, we developed a new methodology. This methodology has three notable features. *First*, it balances optimizing local point-wise errors with global distributional errors and agreement with the model. *Second*, it allows users to imbue the model with known, underlying dynamics such as oscillatory dynamics. *Third*, the method allows for the parameters to vary in time over an optimization window to manage both errors in measurement times and their impact on model rigidity and non-stationarity of the generating processes over the optimization window. When we applied this method in a few contexts using both simulated (estimated) and real blood glucose data from two contexts including the ICU and a patient in the wild, the method was able in all cases to manage the sparse data issues, balance point-wise errors and global distributional errors while also allowing agreement with the model. This implies that we were able to robustly preserve global dynamics properties while also minimizing point-wise errors, a major goal of this work. We believe that this methodological pathway will be useful in inverse problems applications as well as embedded within sequential smoothing and DA-initialization applications. A public Github with the code used in this paper can be found in XXX.

Appendix A. Computing the derivative for the L_k's

This appendix contains explicit formulas for the derivatives of the various components L_k of the objective function, required for gradient descent. We write them down here for completeness and easy reproducibility, but also to stress a point: even for the relatively simple model that we have adopted for the examples in this article, computing and implementing the required derivatives is a laborious task, prone to errors. In order to add flexibility to the selection of models, one should consider resorting to automatic differentiation.

$$L_1 = \frac{1}{n}\sum_j \log\left[(1-\epsilon)K^y\left(y^j, x^j\right) + \epsilon\frac{1}{n}\sum_{i=1}^{n} K^y\left(y^j, y^i\right)\right]$$

$$\frac{\partial L_1}{\partial x^j} = \frac{1}{n}\frac{1-\epsilon}{h^2}\frac{\left(y^j - x^j\right)K^y\left(y^j, x^j\right)}{(1-\epsilon)K^y\left(y^j, x^j\right) + \epsilon\rho_0^j}$$

$$\begin{aligned} L_2 &= -\frac{1}{n}\sum_i \frac{1}{\sum_l K^t\left(t^i, t^l\right)}\sum_j \left[\left(K^y\left(x^i, x^j\right) - K^y\left(y^i, x^j\right)\right) - \left(K^y\left(x^i, y^j\right) - K^y\left(y^i, y^j\right)\right)\right] K^t\left(t^i, t \right. \\ &= -\frac{1}{2n}\sum_{i,j}\left(\frac{K^t\left(t^i, t^j\right)}{\sum_l K^t\left(t^j, t^l\right)} + \frac{K^t\left(t^i, t^j\right)}{\sum_l K^t\left(t^i, t^l\right)}\right)\left[K^y\left(x^i, x^j\right) - 2K^y\left(y^i, x^j\right) + K^y\left(y^i, y^j\right)\right] \end{aligned}$$

$$\frac{\partial L_2}{\partial x^j} = -\frac{1}{nh^2}\sum_i \left(\frac{K^t\left(t^i, t^j\right)}{\sum_l K^t\left(t^j, t^l\right)} + \frac{K^t\left(t^i, t^j\right)}{\sum_l K^t\left(t^i, t^l\right)}\right)\left[\left(x^i - x^j\right)K^y\left(x^i, x^j\right) - \left(y^i - x^j\right)K^y\left(y^i, x^j\right)\right]$$

For L_3 and L_4, the components of the objective function that enforce compliance with the model, the pairs (x^{j-1}, z^{j-1}) are replaced by (r^{j-1}, θ^{j-1}), and we have, with all variables evaluated at time t_{j-1},

$$\begin{aligned} \frac{\partial}{\partial x} &= \frac{x-b}{r}\frac{\partial}{\partial r} - \frac{z}{r^2}\frac{\partial}{\partial \theta}, \\ \frac{\partial}{\partial z} &= \frac{z}{r}\frac{\partial}{\partial r} + \frac{x-b}{r^2}\frac{\partial}{\partial \theta}, \\ \frac{\partial}{\partial b} &= -\frac{x-b}{r}\frac{\partial}{\partial r} + \frac{z}{r^2}\frac{\partial}{\partial \theta}. \end{aligned}$$

$$\rho^x\left(x^j | r^{j-1}, \theta^{j-1}, \Delta t^j, \alpha^j\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(x^j - b^j - r_+^j\cos\left(\theta^{j-1} + \omega^{j-1}\Delta t^j\right)\right)^2}{2\sigma^2}},$$

$$\rho^z\left(z^j | r^{j-1}, \theta^{j-1}, \Delta t^j, \alpha^j\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(z^j - r_+^j\sin\left(\theta^{j-1} + \omega^{j-1}\Delta t^j\right)\right)^2}{2\sigma^2}},$$

where

$$r_+^j = \left(1 - d_s^j\right)a^j + d_s^j\, r^{j-1}, \quad d_s^j = e^{-\frac{\Delta t^j}{T_s}}.$$

Then

$$\frac{\partial \rho^x\left(x^j | r^{j-1}, \theta^{j-1}, \Delta t^j, \alpha^j\right)}{\partial x^j} = -\frac{x^j - b^j - r_+^j\cos\left(\theta^{j-1} + \omega^{j-1}\Delta t^j\right)}{\sigma^2}\rho^x,$$

$$\frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial r^{j-1}} = -d_s^j \cos\left(\theta^{j-1}+\omega^{j-1}\Delta t^j\right)\frac{\partial \rho^x}{\partial x^j},$$
$$\frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial \theta^{j-1}} = r_+^j \sin\left(\theta^{j-1}+\omega^{j-1}\Delta t^j\right)\frac{\partial \rho^x}{\partial x^j},$$
$$\frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial b^j} = -\frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial x^j},$$
$$\frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial a^j} = -\frac{\partial \rho^x\left(x^j,\theta^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial r^{j-1}},$$
$$\frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial \omega^{j-1}} = \Delta t^j \frac{\partial \rho^x\left(x^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial \theta^{j-1}},$$
$$\frac{\partial \rho^z\left(z^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial z^j} = -\frac{z^j - r_+^j \sin\left(\theta^{j-1}+\omega^{j-1}\Delta t^j\right)}{\sigma^2}\,\rho^z,$$
$$\frac{\partial \rho^z\left(z^j,\theta^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial r^{j-1}} = -d_s^j \sin\left(\theta^{j-1}+\omega^{j-1}\Delta t^j\right)\frac{\partial \rho^z}{\partial z^j},$$
$$\frac{\partial \rho^z\left(z^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial \theta^{j-1}} = -r_+^j \cos\left(\theta^{j-1}+\omega^{j-1}\Delta t^j\right)\frac{\partial \rho^z}{\partial z^j}$$
$$\frac{\partial \rho^z\left(z^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial a^j} = -\frac{\partial \rho^z\left(z^j,\theta^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial r^{j-1}},$$
$$\frac{\partial \rho^z\left(z^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial \omega^{j-1}} = \Delta t^j \frac{\partial \rho^z\left(z^j|r^{j-1},\theta^{j-1},\Delta t^j,\alpha^j\right)}{\partial \theta^{j-1}}.$$

$$L_3 = \frac{1}{n}\sum_j \log\left[\rho^x\left(x^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)\right],$$

$$\frac{\partial L_3}{\partial x^j} = \frac{1}{n}\left[\frac{\frac{\partial}{\partial x^j}\rho^x\left(x^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)}{\rho^x\left(x^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)} + \frac{\frac{\partial}{\partial x^j}\rho^x\left(x^{j+1}|x^j,z^j,\Delta t^{j+1},\alpha^{j+1}\right)}{\rho^x\left(x^{j+1}|x^j,z^j,\Delta t^{j+1},\alpha^{j+1}\right)}\right]$$

$$\frac{\partial L_3}{\partial z^j} = \frac{1}{n}\frac{\frac{\partial}{\partial z^j}\rho^x\left(x^{j+1}|z^j,\Delta t^{j+1},\alpha^{j+1}\right)}{\rho^x\left(x^{j+1}|x^j,z^j,\Delta t^{j+1},\alpha^{j+1}\right)}$$

$$\frac{\partial L_3}{\partial \alpha_k^j} = \frac{1}{n}\frac{\frac{\partial}{\partial \alpha_k}\rho^x\left(x^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)}{\rho^x\left(x^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)}$$

$$L_4 = \frac{1}{n}\sum_j \log\left[\rho^z\left(z^j|x^{j-1},z^{j-1},\Delta t^j,\alpha^j\right)\right]$$

$$\frac{\partial L_4}{\partial x^j} = \frac{1}{n} \frac{\frac{\partial}{\partial x^j} \rho^z \left(z^{j+1} | z^j, \Delta t^{j+1}, \alpha^{j+1} \right)}{\rho^z \left(z^{j+1} | x^j, z^j, \Delta t^{j+1}, \alpha^{j+1} \right)}$$

$$\frac{\partial L_4}{\partial z^j} = \frac{1}{n} \left[\frac{\frac{\partial}{\partial z^j} \rho^z \left(z^j | x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j \right)}{\rho^z \left(z^j | x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j \right)} + \frac{\frac{\partial}{\partial z^j} \rho^z \left(z^{j+1} | x^j, z^j, \Delta t^{j+1}, \alpha^{j+1} \right)}{\rho^z \left(x^{j+1} | x^j, z^j, \Delta t^{j+1}, \alpha^{j+1} \right)} \right]$$

$$\frac{\partial L_4}{\partial \alpha_k^j} = \frac{1}{n} \frac{\frac{\partial}{\partial \alpha_k^j} \rho^z \left(z^j | x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j \right)}{\rho^z \left(z^j | x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j \right)}$$

$$L_{4+k} = \frac{1}{n} \sum_j \log \left[\rho^k \left(\alpha_k^j | \alpha_k^{j-1}, \Delta t^j \right) \right],$$

$$\frac{\partial L_{4+k}}{\partial \alpha_k^j} = \frac{1}{n} \left[\frac{\frac{\partial}{\partial \alpha_k^j} \rho^k \left(\alpha_k^j | \alpha_k^{j-1} \right)}{\rho^l \left(\alpha_k^j | \alpha_k^{j-1} \right)} + \frac{\frac{\partial}{\partial \alpha_k^j} \rho^k \left(\alpha_k^{j+1} | \alpha_k^j \right)}{\rho^k \left(\alpha_k^{j+1} | \alpha_k^j \right)} \right],$$

$$\rho^k \left(\alpha_k^j | \alpha_k^{j-1} \right) = N \left[d_l^j \alpha_l^{j-1} + \left(1 - d_l^j \right) \tilde{\alpha}_l, \left(1 - d_l^j \right) {\sigma_l}^2 \right],$$

$$\frac{\partial \rho^k \left(\alpha_k^j | \alpha_k^{j-1} \right)}{\partial \alpha_k^j} = - \frac{\alpha_k^j - \left(d_l^j \alpha_l^{j-1} + \left(1 - d_l^j \right) \tilde{\alpha}_l \right)}{\left(1 - d_l^j \right) {\sigma_l}^2} \rho^k \left(\alpha_k^j | \alpha_k^{j-1} \right),$$

$$\frac{\partial \rho^k \left(\alpha_k^j | \alpha_k^{j-1} \right)}{\partial \alpha_k^{j-1}} = - d_l^j \frac{\partial \rho^k \left(\alpha_k^j | \alpha_k^{j-1} \right)}{\partial \alpha_k^j}.$$

Appendix B. Ultradian glucose-insulin model

The primary state variables include the plasma glucose concentration G, the plasma insulin concentration I_p, and the interstitial insulin concentration I_i. Additionally there is a delay in the hepatic response of plasma insulin to glucose approximated using the linear chain trick resulting in (h_1, h_2, h_3)

[35]. The system of ordinary differential equations take the form [19]:

$$\frac{dI_p}{dt} = f_1(G) - E\big(\frac{I_p}{V_p} - \frac{I_i}{V_i}\big) - \frac{I_p}{t_p} \tag{B.1a}$$

$$\frac{dI_i}{dt} = E\big(\frac{I_p}{V_p} - \frac{I_i}{V_i}\big) - \frac{I_i}{t_i} \tag{B.1b}$$

$$\frac{dG}{dt} = f_4(h_3) + I_G(t) - f_2(G) - f_3(I_i)G \tag{B.1c}$$

$$\frac{dh_1}{dt} = \frac{1}{t_d}\big(I_p - h_1\big) \tag{B.1d}$$

$$\frac{dh_2}{dt} = \frac{1}{t_d}\big(h_1 - h_2\big) \tag{B.1e}$$

$$\frac{dh_3}{dt} = \frac{1}{t_d}\big(h_2 - h_3\big) \tag{B.1f}$$

This model includes many parameterized processes including the rate of insulin production, $f_1(G)$, insulin-independent glucose utilization $f_2(G)$, insulin-dependent glucose utilization, $f_3(I_i)G$, and represents insulin-dependent glucose utilization, delayed insulin-dependent glucose utilization, $f_4(h_3)$, defined by:

$$f_1(G) = \frac{R_m}{1 + \exp(\frac{-G}{V_g c_1} + a_1)} \tag{B.2}$$

$$f_2(G) = U_b(1 - \exp(\frac{-G}{C_2 V_g})) \tag{B.3}$$

$$f_3(I_i) = \frac{1}{C_3 V_g}(U_0 + \frac{U_m - U_0}{1 + (\kappa I_i)^{-\beta}}) \tag{B.4}$$

$$f_4(h_3) = \frac{R_g}{1 + \exp(\alpha(\frac{h_3}{C_5 V_p} - 1))} \tag{B.5}$$

$$\kappa = \frac{1}{C_4}(\frac{1}{V_i} - \frac{1}{E t_i}), \tag{B.6}$$

respectively.

The nutritional driver of the model $I_G(t)$ is defined over N non-overlapping intervals where nutrition is delivered through an enteral tube at constant rates.

$$I_G(t) = \sum_{j=1}^{N} m_j \mathbb{1}_{\{t_j \leq t < t_{j+1}\}}(t), \tag{B.7}$$

where m_j is the carbohydrate rate (mg/min) delivered over the interval $[t_j, t_{j+1})$ and $\mathbb{1}(\cdot)$ is the characteristic function.

Appendix C. The multiobjective function

Our choice of an objective function L to maximize raises some theoretical issues, both general considerations on how to combine more than one goal into a single objective function and the theoretical underpinning of the component of the chosen multiobjective function associated to each goal. This appendix discusses both of these aspects.

A full discussion of multiobjective optimization, however, is far beyond this article's scope, we refer the reader to [42] for a current review. Given a set of M objective functions $\{L_m\}$ to maximize, we seek a Pareto optimal through their linear scalarization through coefficients $\{\lambda_m\} \geq 0$, writing the single objective function

$$L(X) = \sum_{m=1}^{M} \lambda_m L_m(X).$$

Our approach has the distinctive feature that the $\{\lambda_m\}$ are not constant throughout the procedure: in order to initialize the $\{x^i\}$ to values close to the data, we assign initially values different from zero only to $\{\lambda_{1,2}\}$. Then, to adjust the hidden, emergent variable z, we freeze during a second phase the $\{x^i\}$ and all parameters α to their previously initialized values and evolve only z through ascent of first L_3 and then $\{L_{3,4}\}$, setting all other $\{\lambda_m\}$ to zero. Only in a third, final phase, are all $\{\lambda_m\}$ fixed at nonzero values.

What follows is a brief analysis of the individual $\{L_m\}$.

- L_1**:** Recall that

$$L_1 = \frac{1}{n} \sum_j \log \left[(1-\epsilon) \rho^o \left(y^j | x^j \right) + \epsilon \rho^0 \left(y^j \right) \right]$$

measures the point-wise agreement between the observations $\{y^i\}$ and the corresponding estimated states $\{x^i\}$, through a postulated distribution $\rho^o(y|x)$ that quantifies the probability that an observation will yield the value y when the actual state is x. With $\epsilon = 0$, L_1 is simply the log-likelihood function for the $\{x^i\}$, whole maximization when ρ_o is a Gaussian centered at x agrees with the least-square error, the most conventional way of quantifying point-wise discrepancy. Our proposal extends LSE in two ways: it permits using more realistic conditional distributions consistent with the true observation methodology, and it is made more robust to outliers than LSE through the addition of a background distribution

$$\rho^0(y) = \frac{1}{n}\sum_l K^y\left(y, y^l\right)$$

that mollifies $\rho^o(y|x)$, allowing some of the $\{x^i\}$ to deviate considerably from the corresponding $\{y^i\}$ if required to make them more consistent with the model. Switching from LSE to a likelihood-based quantification of point-wise agreement is consistent with the methodology as a whole, where all objective functions $\{L_m\}$ are based on an underlying stochastic assumption.

- L_2**:** We measure the discrepancy between the estimated states $\{x^i\}$ and the corresponding observations $\{y^i\}$ not only point-wise but also distribution-wise. Since we are dealing with typically non-stationary time series, we allow said distributions to vary slowly over time. Then we adopt

$$L_2 = -\frac{1}{n}\sum_j F\left(x^j, t^j\right) - F\left(y^j, t^j\right), \quad F(z,t) = \rho^x(z|t) - \rho^y(z|t),$$

where the conditional densities are estimated using Nadaraya-Waston's formula:

$$\rho^x(z|t) = \frac{1}{\sum_l K^t\left(t, t^l\right)}\sum_l K^y\left(z, x^l\right) K^t\left(t, t^l\right),$$

$$\rho^y(z|t) = \frac{1}{\sum_l K^t\left(t, t^l\right)}\sum_l K^y\left(z, y^l\right) K^t\left(t, t^l\right).$$

We have used Gaussian kernels, with bandwidths in x and y given by the rule of thumb, and in t by the modulation timescale T discussed below. The L_2 so defined is an empirical representation of minus the L^2-difference between the two conditional distributions:

$$\begin{aligned} L_2 &\sim -\int \left[\rho^x(z|t) - \rho^y(z|t)\right] \rho^x(z|t)\gamma(t)\ dzdt \\ &\quad + \int \left[\rho^x(z|t) - \rho^y(z|t)\right] \rho^y(z|t)\gamma(t)\ dzdt \\ &= -\int \left[\rho^x(z|t) - \rho^y(z|t)\right]^2 \gamma(t)\ dzdt. \end{aligned}$$

- $L_{3,4}$**:** The two problem-specific components,

$$L_3 = \frac{1}{n}\sum_j \log\left[\rho^x\left(x^j|x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j\right)\right]$$

and

$$L_4 = \frac{1}{n}\sum_j \log\left[\rho^z\left(z^j|x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j\right)\right]$$

enforce agreement with the model in the maximum likelihood sense. Even though they could be made slightly more general by consolidating them into a single component based on the joint conditional distribution $\rho^{x,z}\left(x^j, z^j|x^{j-1}, z^{j-1}, \Delta t^j, \alpha^j\right)$, the pair $L_{3,4}$ provides more flexibility, allowing in particular the implementation of a stage when z is updated using only L_3. The $L_{3,4}$ are the only components where the model plays a role; any bio-medically-based Markov model can replace the simple oscillatory one used in this article.

- $L_{m>4}$**:** These last components of the objective function, one per model parameter α_l, enforce continuity of the parameters over time, while allowing them to vary slowly and relax toward default values when the data is sparse. They adopt the maximum-likelihood form

$$L_{4+l} = \frac{1}{n}\sum_j \log\left[\rho^l\left(\alpha_l^j|\alpha_l^{j-1}, \Delta t^j\right)\right],$$

where we have adopted

$$\alpha_l^{j+1} \sim N\left[d_l^j\alpha_l^j + \left(1-d_l^j\right)\tilde{\alpha}_l, \left(1-d_l^j\right){\sigma_l}^2\right].$$

Here $\tilde{\alpha}_l$ is a default value of α_l, while $d_l^j = e^{-\frac{\Delta t^j}{T_l}}$ is a decaying weight starting at 1 and ending up at 0.

Thus all of the $L_{1,3,4,m>4}$ are log-likelihoods, based on stochastic models for the observational process (L_1), the dynamics ($L_{3,4}$) and the time evolution of the models' parameters ($L_{m>4}$) respectively, while L_2 compares the possibly time-dependent distributions underlying the observations $\{y^i\}$ and the estimated underlying states $\{x^i\}$ through an empirical, kernel-based estimation of their L^2-discrepancy.

References

[1] D. Albers, G. Hripcsak, and M. Schmidt, *Population physiology: leveraging electronic health record data to understand human endocrine dynamics*, PLoS One, 7 (2012), p. e480058.

[2] D. Albers, M. Levine, B. Gluckman, H. Ginsberg, G. Hripcsak, and L. Mamykina, *Personalized glucose forecasting for type 2 diabetes using data assimilation*, PLoS Computational Biology, 13 (2017), p. e1005232.

[3] D. Albers, M. Levine, A. Stuart, B. Gluckman, L. Mamykina, and G. Hripcsak, *Mechanistic machine learning: how data assimilation leverages physiologic knowledge using bayesian inference to forecast the future, infer the present, and phenotype*, Journal of the American Medical Informatics Association, 25 (2018), pp. 1392–1401.

[4] D. Albers, M. Sirlanci, M. Levine, J. Claassen, C. D. Nigoghossian, and G. Hripcsak, *Interpretable physiological forecasting in the icu using constrained data assimilation and electronic health record data*, Journal of Biomedical Informatics, 145 (2023), p. 104477, https://doi.org/https://doi.org/10.1016/j.jbi.2023.104477.

[5] D. J. Albers, P.-A. Blancquart, M. E. Levine, E. E. Seylabi, and A. Stuart, *Ensemble kalman methods with constraints*, Inverse Problems, 35 (2019), p. 095007, https://doi.org/10.1088/1361-6420/ab1c09.

[6] D. J. Albers and G. Hripcsak, *A statistical dynamics approach to the study of human health data: resolving population scale diurnal variation in laboratory data*, Physics Lett. A, (2010).

[7] D. J. Albers and G. Hripcsak, *Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations*, CHAOS, 22 (2012), p. 013111.

[8] D. J. Albers, M. E. Levine, L. Mamykina, and G. Hripcsak, *The parameter houlihan: A solution to high-throughput identifiability indeterminacy for brutally ill-posed problems*, Mathematical Biosciences, 316 (2019), p. 108242, https://doi.org/https://doi.org/10.1016/j.mbs.2019.108242.

[9] D. Brealey and M. Singer, *Hyperglycemia in critical illness: a review*, J. Diabetes Sci. Technol., 3 (2009), pp. 1250–1260.

[10] M. Burgermaster, P. M. Desai, E. M. Heitkemper, F. Juul, E. G. Mitchell, M. Turchioe, D. J. Albers, M. E. Levine, D. Larson, and L. Mamykina, *Who needs what (features) when? personalizing engagement with data-driven self-management to improve health equity*, Journal of Biomedical Informatics, 144 (2023), p. 104419, https://doi.org/https://doi.org/10.1016/j.jbi.2023.104419.

[11] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian data analysis*, CRC Press, 3 ed., 2014.

[12] J. Geweke, *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*, in Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, Dedicated to the memory of Morris H. DeGroot, 1931–1989, Oxford University Press, 08 1992, https://doi.org/10.1093/oso/9780198522669.003.0010.

[13] J. Gunst, A. D. Bruyn, and G. V. den Berghe, *Glucose control in the icu*, Curr. Opin. Anaesthesiol., 32 (2019), pp. ‘156–162.

[14] G. Hripcsak and D. Albers, *Next-generation phenotyping of electronic health records*, JAMIA, 10 (2012), pp. 1–5.

[15] G. Hripcsak and D. Albers, *Correlating electronic health record concepts with healthcare process events*, JAMIA, 0 (2013), pp. 1–8.

[16] G. Hripcsak and D. Albers, *Evaluating prediction of continuous clinical values: A glucose case study*, Methods Inf Med, 61 (2022), pp. e35–e44.

[17] G. Hripcsak, D. Albers, and A. Perotte, *Exploiting time in electronic health record correlations*, JAMIA, 18 (2011), pp. 109–115.

[18] G. Hripcsak, D. Albers, and A. Perotte, *Parameterizing time in electronic health record studies.*, J Am Med Inform Assoc, (2015).

[19] J. Keener and J. Sneyd, *Mathematical physiology II: Systems physiology*, Springer, 2008.

[20] M. Levine, D. Albers, and G. Hripcsak, *Comparing lagged linear correlation, lagged regression, granger causality, and vector autoregression for uncovering associations in ehr data*, in Proc AMIA Symp, AMIA, 2016, pp. 779–88.

[21] M. Levine, G. Hripcsak, L. Mamykina, A. Stuart, and D. Albers, *Offline and online data assimilation for real-time blood glucose forecasting in type 2 diabetes.* arXiv:1709.00163, 2017.

[22] L. Mamykina, M. Levine, P. Davidson, A. Smaldone, N. Elhadad, and D. Albers, *Reasoning about personally generated nutritional data in diabetes with information technologies*, J Am Med Inform Asso, 23 (2016), pp. 526–531.

[23] L. Mamykina, M. Levine, P. Davidson, A. Smaldone, N. Elhadad, and D. Albers, *Cognitive Informatics in Health and Biomedicine. Health Informatics*, Springer, 2017, ch. From personal informatics to personal analytics: investigating how clinicians and patients reason about personal data generated with self-monitoring in diabetes.

[24] A. C. Miller, N. J. Foti, and E. Fox, *Learning insulin-glucose dynamics in the wild*, in Proc Machine Learning for Healthcare Conference, 2020, pp. 172–197, http://proceedings.mlr.press/v126/miller20a/miller20a.pdf.

[25] T. J. O'Kane, P. A. Sandery, D. P. Monselesan, P. Sakov, M. A. Chamberlain, R. J. Matear, M. A. Collier, D. T. Squire, and L. Stevens, *Coupled data assimilation and ensemble initialization with application to multiyear enso prediction*, Journal of Climate, 32 (2019), pp. 997 – 1024, https://doi.org/10.1175/JCLI-D-18-0189.1.

[26] J. L. Parkes, S. L. Slatin, S. Pardo, and B. H. Ginsberg, *A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose*, Diabetes Care, 23 (2000), pp. 1143–1148.

[27] S. Rodrigues De Morais and A. Aussem, *Exploiting data missingness in bayesian network modeling*, in IDA '09: Proceedings of the 8th International Symposium on Intelligent Data Analysis, Berlin, Heidelberg, 2009, Springer-Verlag, pp. 35–46.

[28] D. B. Rubin and R. J. A. Little, *Statistical analysis with missing data*, Wiley, 2 ed., 2002.

[29] R. E. Sherman, S. A. Anderson, G. J. D. Pan, G. W. Gray, T. Gross, H. L. Hunter, L. LaVange, D. Marinac-Dabic, P. W. Marks, M. A. Robb, J. Shuren, R. Temple, J. Woodcock, L. Q. Yur, and R. M. Califf, *Real-world evidence: What is it and what can it tell us?*, NEJM, 375 (2016), pp. 2293–2297.

[30] M. Sirlanci, G. Hripcsak, C. C. Low Wang, J. N. Stroh, Y. Wang, T. D. Bennett, A. M. Stuart, and D. J. Albers, *A stochastic model-based control methodology for glycemic management in the intensive care unit*, Frontiers in Medical Engineering, 2 (2024), https://doi.org/10.3389/fmede.2024.1419786.

[31] M. Sirlanci, M. E. Levine, C. C. Low Wang, D. J. Albers, and A. M. Stuart, *A simple modeling framework for prediction in the human glucose–insulin system*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 33 (2023), p. 073150, https://doi.org/10.1063/5.0146808.

[32] M. Unser, *Sampling-50 years after Shannon*, Proceedings of the IEEE, 88 (2000), pp. 569–587, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=843002.

[33] R. T. Krafty, M. Hall, and W. Gou, *Functional mixed effects spectral analysis*, Biometrika, 98 (2011), pp. 583–598, https://academic.oup.com/biomet/article/98/3/583/236347.

[34] H. Smith, *An Introduction to Delay Differential Equations with Applications to the Life Sciences*, Springer, 20011.

[35] J. Sturis, K. S. Polonsky, E. Mosekilde, and E. V. Cauter, *Computer model for mechanisms underlying ultradian oscillations of insulin and glucose*, Am J Physiol Endocrinol Metab, 260 (1991), pp. E801–E809.

[36] T. Van Herpe, B. Pluymers, M. Espinoza, G. Van den Berghe, and B. De Moor, *A minimal model for glycemia control in critically ill patients*, in 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2006, pp. 5432–5435.

[37] K. A. Wang, M. E. Levine, J. Shi, and E. B. Fox, *Learning absorption rates in glucose-insulin dynamics from meal covariates*, 2023, https://arxiv.org/abs/2304.14300.

[38] Y. Wang, J. Stroh, G. Hripcsak, C. C. Low Wang, T. D. Bennett, J. Wrobel, C. Der Nigoghossian, S. W. Mueller, J. Claassen, and D. Albers, *A methodology of phenotyping icu patients from ehr data: High-fidelity, personalized, and interpretable phenotypes estimation*, Journal of Biomedical Informatics, 148 (2023), p. 104547, https://doi.org/https://doi.org/10.1016/j.jbi.2023.104547.

[39] M. Wilson, J. Weinreb, and G. W. S. Hoo, *Intensive insulin therapy in critical care: a review of 12 protocols.*, Diabetes Care, 30 (2007), pp. 1005–1011.

[40] F. Zhang, Y. Weng, J. F. Gamache, and F. D. Marks, *Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne doppler radar observations*, Geophysical Research Letters, 38 (2011), https://doi.org/https://doi.org/10.1029/2011GL048469.

[41] B. J. Zou, M. E. Levine, D. P. Zaharieva, R. Johari, and E. Fox, *Hybrid^2 neural ODE causal modeling and an application to glycemic response*, in Proceedings of the 41st International Conference on Machine Learning, R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp, eds., vol. 235 of Proceedings of Machine Learning Research, PMLR, 21–27 Jul 2024, pp. 62934–62963, https://proceedings.mlr.press/v235/zou24b.html.

[42] B. Tu, N. Kantas, R. M. Lee, and B. Shafei, *Multiobjective Optimization Using the R2 Utility*, SIAM Review, **67** (2025), pp. 213–255.

[43] M. Ehrgott, *Multicriteria Optimization*, Springer, 2005.

[44] K. Miettinen, *Nonlinear Multiobjective Optimization*, Springer, 1998.

[45] S. Greco, M. Ehrgott, and J. R. Figueira, *Multiple Criteria Decision Analysis*, Springer, 2016.

[46] J. Latz, *Bayesian Inverse Problems Are Usually Well-Posed*, SIAM Review, **65** (2023), pp. 831–865.

[47] M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms and Applications*, SIAM, 2016.

[48] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, Springer, 2015.

Table B.2: Full list of parameters for the ultradian glucose-insulin model [19]. Note that IIGU and IDGU denote insulin-independent glucose utilization and insulin-dependent glucose utilization, respectively.

Ultradian model parameters		
Name	Nominal Value	Meaning
V_p	3 l	plasma volume
V_i	11 l	interstitial volume
V_g	10 l	glucose space
E	0.2 l min^{-1}	exchange rate for insulin between remote and plasma compartments
t_p	6 min	time constant for plasma insulin degradation (via kidney and liver filtering)
t_i	100 min	time constant for remote insulin degradation (via muscle and adipose tissue)
t_d	12 min	delay between plasma insulin and glucose production
k	0.5 min^{-1}	rate of decayed appearance of ingested glucose
R_m	209 mU min^{-1}	linear constant affecting insulin secretion
a_1	6.6	exponential constant affecting insulin secretion
C_1	300 mg l^{-1}	exponential constant affecting insulin secretion
C_2	144 mg l^{-1}	exponential constant affecting IIGU
C_3	100 mg l^{-1}	linear constant affecting IDGU
C_4	80 mU l^{-1}	factor affecting IDG
C_5	26 mU l^{-1}	exponential constant affecting IDGU
U_b	72 mg min^{-1}	linear constant affecting IIGU
U_0	4 mg min^{-1}	linear constant affecting IDGU
U_m	94 mg min^{-1}	linear constant affecting IDGU
R_g	180 mg min^{-1}	linear constant affecting IDGU
α	7.5	exponential constant affecting IDGU
β	1.772	exponent affecting IDGU