Model-based identification of eating behavioral patterns in populations with type 1 diabetes

Eleonora Maria Aiello [a,*], Chiara Toffanin[a], Lalo Magni[b], Giuseppe De Nicolao[a]

[a]*Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia 27100, Italy.*
[b]*Department of Civil and Architecture Engineering, University of Pavia, Pavia 27100, Italy*

Abstract

In the paper, a probabilistic model of the eating behavior of subjects with type 1 diabetes is proposed and validated against two extensive datasets collected during experiments in free-living conditions in Padova and Amsterdam clinical centres. Meals are modeled as a discrete-time marked point process. The random meal events are associated to a Markov Chain whose state is the fasting period and the transition probabilities depend on variables such as daytime and carbohydrate intake of the previous meal. The random marks associated to each meal represent the carbohydrate intake and their distribution possibly depends on daytime, fasting period and carbohydrate intake of the previous meal. Logistic and Gaussian Mixture models are used to identify two models from the Padova and Amsterdam datasets, respectively. In order to validate it, the proposed model is used to simulate synthetic datasets, whose statistical properties appear to be in good agreement with those of the experimental datasets. The availability of a probabilistic model of eating behavior is expected to be valuable for several purposes ranging from the optimization and customization of automatic insulin dosing systems to decision supporting tools for insulin dosing and the alert management for missing meal announcements.

Keywords: eating behavior, type 1 diabetes, marked point process, meal simulation, artificial pancreas.

1. Introduction

Event-history data are a record of events often collected through time where the basic information is the times of occurrence of the events and the types of events that occur [1]. As life histories provide a rich source of data, event-history analysis has applications in many fields, including sociology, economics, biology, medicine, and engineering. The structure of data is often challenging to work with due to difficulties in extracting the most informative component.

The event-history data studied in this paper involved data of individuals with type 1 diabetes (T1D) collected during a 1-month clinical trial under free-living condition in Padova (Italy) and Amsterdam (Netherlands) clinical centres [2]. As the participants had a normal life without any type of restriction during the clinical study, these historical datasets represent a rich and potentially useful source of information.

In subjects with T1D, the pancreas does not produce adequate levels of insulin, which results in abnormal metabolism of carbohydrates, generating elevated levels of glucose in the blood and urine. These glycemic excursions can lead to serious medical complications. The most common ones are related with cardiovascular diseases, which are the most frequent cause of death in people with diabetes [3]. A frequent complication is a kidney disease called nephropathy [4]: indeed high levels of blood sugar make the kidneys filter too much blood, overburdening the filters. After many years, the kidneys could start to not operate correctly, and useful proteins could be lost in the urine. In half of subjects affected by T1D, diabetic neuropathy [4], which is a peripheral nerves malfunction, may also occur as a complication. In this case the symptoms are tingling, pain, numbness, or weakness in the feet and hands. The degeneration of neuropathy can lead to foot ulcers and eventual limb amputation. Finally, long-term accumulated damage to the small blood vessels in the eye leads to diabetic retinopathy, an important cause of blindness [5]. Therefore, people with T1D must constantly monitor their blood glucose levels. For effective blood glucose management, those with T1D must either inject themselves with insulin or use an insulin infusion pump to keep the glucose concentration in the euglycemic range ([70, 180] mg/dl). Carbohydrate intake results in mild elevations in blood glucose depending on the meal size. To compensate the carbohydrate intake, patients with T1D are used to administer an insulin bolus at meal times to compensate the glucose rise due to the meal. As food intake represents a

*Corresponding author.
E-mail address: eleonoramaria.aiello01@universitadipavia.it

challenge for the glucose control in automated insulin delivery (AID) systems, modeling and simulating the eating behavior can represent a valid support to improve diabetes management [6, 7]. Many clinical studies have reported that a high number of missed meal announcements and missed meal boluses occur, especially in adolescents who happen not to follow diabetes care diligently [8, 9]. This error is reflected on unexpected changes in glucose levels. For this reason several meal detection algorithms are present in the literature. Dassau et al. designed a voting scheme to trigger a binary detection of a meal without estimating the meal start time and meal size [10]. In a study by Lee and Bequette [11], a meal was detected based on certain logical conditions of the first and second derivatives of glucose levels, and then a finite impulse response (FIR) filter was applied to estimate the meal size. Lee et al. later refined the logical conditions to generate a series of meal impulses, by which the bolus insulin was determined [11]. Cameron et al. developed a probabilistic detection algorithm based on a set of meal shapes, where the meal start time and postprandial glucose appearance were estimated at the same time [12]. Later they also developed a multimodel method to detect and estimate unannounced meals [13]. The multimodel method was also probabilistic but assumed a constant meal absorption shape. In [14], an unscented Kalman filter was applied to Bergman's minimal model of glucose kinetics for meal detection and was evaluated using the UVA/Padova simulator [15]. Xie and Wang use a multi-model approach to detect meals, meal start time, and meals size using a Variable State Dimension (VSD) method, which used a switching criteria to switch between models with different state dimensions [16]. Chen et al. and Weimer et al. proposed a method for meal detection and bolus estimation based on an estimated rate of appearance [17, 18]. In this study, an algorithm that uses cross-covariance to detect meals is tested in silico both in situations with and without exercise, while Ramkissoon et al. proposed an algorithm that uses cross-covariance to detect meals is tested in silico both in situations with and without exercise [19]. In general, meal detection algorithms could be enhanced by using a probabilistic model of meal occurrences, which can provide a more accurate prior information of the meal information.

As a byproduct of outpatient trials, extensive data on eating behaviours have become available, observed over a long period of time without ad-hoc clinical protocols. By leveraging these data, this paper aims to identify a meal model that not only captures the patterns observed in the data but provides also a simulation framework that generates stochastic meal events consistent with the eating habits of a specific cohort of subjects with T1D. Although the clinical eating protocol, (breakfast, lunch, and dinner at fixed times and amounts) represents a standard for a clinical study, it is evident that it may be entirely different from a free living scenario. Currently, the ability to mimic a real life scenario entirely depends on the sensibility of the researcher on the patient's habits. This work proposes a stochastic meal generator aiming for an accurate reproducibility of eating scenarios and would remove the arbitrariness of the scenario design process and assure the likelihood of the simulated eating patterns.

In the context of eating behaviors, event-history data are typically time series of discrete meal events in continuous time. They can be represented by a temporal point process that captures the timestamps of occurrence of events scattered in time. Besides the timestamp in which food intake occurs, a meal event includes also information on the quantity of ingested carbohydrates. Thus, these eating history data can be thought of as the realization of a marked point process, which is a sequence of random events, each of which is a associated with a random variable called mark. Outside of healthcare, marked point process are widely used to model a number of natural hazard events such as earthquakes [20], forest wildfires [21], and extreme events [22].

In order to ease the statistical characterization of the marked point process, we resorted to a discrete representation of the time axis, that was partitioned into equal sized bins. In each time bin the presence or absence of a meal event is recorded. In modeling, estimation, and prediction of marked point processes it is often assumed that the distribution of marks does not depend on points, so that the point process and the marker distribution can be modelled separately. Rather remarkably, we will show that this relatively simple statistical structure can adequately model the eating behavior of the considered cohorts of T1D subjects.

This kind of point process is characterized by an intensity function, i.e. the instantaneous probability of an event occurring in a time window, and a marker distribution. In our discrete-time setting, the intensity was modelled as a logistic regression [23, 24, 25, 26] on the relevant covariates, i.e. fasting period and daytime. For the marker distribution, a Gaussian Mixture (GM) model was adopted. Within this framework, the marked point process of the eating events is generated by a discrete-time Markov Chain (MC) [27, 28], whose transition probabilities are 24-hour periodic. In particular, the transition probabilities of the

MC specify the probability of the occurrence of a meal event given daytime and the fasting period.

The validation of the eating behavior model is performed by applying the Maximum mean discrepancy test [29] to check any distributional discrepancy between the original datasets and the synthetic datasets generated by the MC simulator.

The paper is organized as follows: Section 2 introduces the theoretical framework, while Section 3 describes the data collection protocol and the data preprocessing phase, as well as the results. The discussion is drawn in Section 4 and some concluding remarks and possible future developments are addressed in Section 5.

2. Materials and Methods

Life history data of the eating behaviors can be regarded as a time series consisting of spikes of random amplitude and location, corresponding to meal events. Event data with marker information can be viewed as the realization of a marked temporal point process consisting of events on a time line together with a characteristic attached to each event, such as the amount of associated carbohydrates. In fact, a marked temporal point process is a random process whose realization consists of an ordered sequence of events localized in time:

$$H = \{e_0 = (t_0, z_0), e_1 = (t_1, z_1), ..., e_n = (t_n, z_n)\} \tag{1}$$

where $t_i \in \mathcal{T} \subset \mathbb{R}_+$, is the time of occurrence of event $i \in \mathbb{Z}_+$, and $z_i \in \mathbb{R}_+$ is the associated mark [30], such that $\{t_i, i \geq 1\}$ is a simple point process satisfying the following conditions:

- $P(0 \leq t_1 \leq t_2 \ldots) = 1$
- $P(t_i \leq t_{i+1}, t_i < \infty) = P(t_i < \infty)$
- $P(\lim_{n\to\infty} t_n = \infty) = 1$

where P is a measure of probabilty.

Figure 1 displays an eating dataset represented as a marked point process: the spike locations correspond to the meal times and the spike amplitudes are proportional to the meal amounts.

For identification and simulation purpose, a discrete-time representation can be more convenient.

The time axis was discretized into bins of size $\Delta > 0$ and the meal events converted to a binary sequence defined as follows:

$$m_E(i) = \begin{cases} 1 & \text{a meal event occurs in the } i\text{-th time bin} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Note that, if Δ is small enough, the probability of two or more events occurring in the same bin is negligible.

2.1. Identification

To simplify the identification task, the point process is assumed independent of the mark values [31, 32]. Under this assumption, the point process model and the mark distribution can be modelled sequentially and then linked together for simulation purposes.

2.1.1. Point process model

The logistic regression model [33] was applied for estimating the probability of a meal event occurring within a bin, given the values of the relevant predictors. For $i = 1, 2, \ldots$, let $\pi(i) = P(m_E(i) = 1|\Phi(i))$ denote the probability that an event occurs at i, given the vector $\Phi(i)' = \left[\phi_1(i), \ldots, \phi_p(i)\right]$ of the predictor variables (features). The logistic regression model has the form:

$$\log[\frac{\pi(i)}{1-\pi(i)}] = \beta_0+\beta_1\phi_1(i)+\beta_2\phi_2(i)+\cdots+\beta_p\phi_p(i) = B'\Phi^{(i)} \tag{3}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are the model coefficients. Taking exponentials on both sides, the probability can be explicitly written as:

$$\pi(i) = \frac{1}{1 + exp(-B'\Phi(i))} \tag{4}$$

The coefficients of the logistic model can be estimated via Maximum Likelihood estimation, which maximizes the log-likelihood function

$$\log \ell(B) = \sum_i m_E(i) \log \pi(i) + (1 - m_E(i)) \log(1 - \pi(i)). \tag{5}$$

Regression analysis was performed to examine the association of independent variables with the dependent variable. The Bayesian Information Criterion (BIC) was used to support the investigation for the contribution of specific covariates: [33]:

$$BIC = -2\ell(B) + \log(n)(p + 1) \tag{6}$$

where n is the dimension of the dataset, i.e. the number of bins.

2.1.2. Mark distribution

The random marks associated with the meal events represent the amount of carbohydrate (CHO) intake. In order to have a

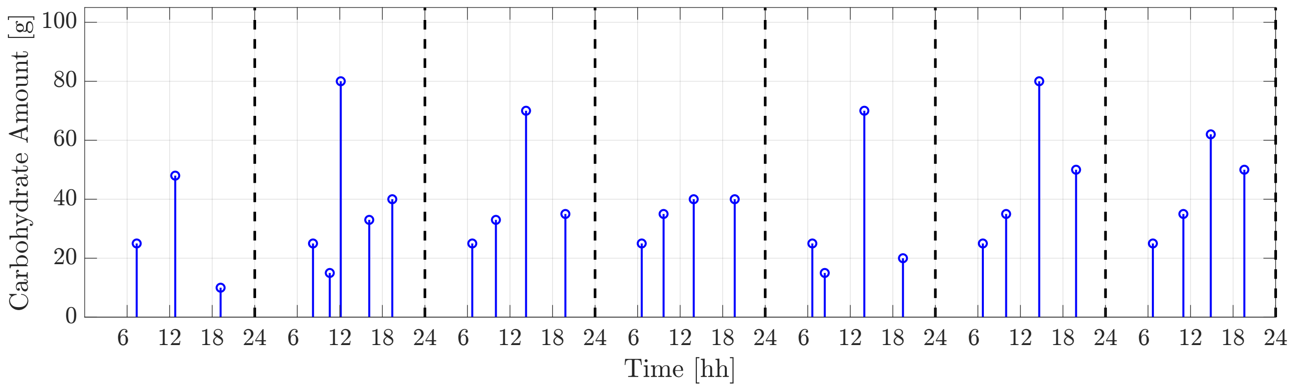

Figure 1: Marked point process representation of the meals of a subject. The meals are modelled as instantaneous spikes whose amplitude is proportional to the carbohydrate intake.

flexible description of the actual meals, the mark distribution will be modelled through a d-dimensional GM model [34] describing the joint distribution of the CHO intake and $d-1$ covariates that help predicting the intakes.

The probability density function of a GM is convex combination of R Gaussian densities:

$$f(\psi) = \sum_{r=1}^{R} \rho_r \frac{e^{-\frac{1}{2}(\psi-\mu_r)'\Sigma_r^{-1}(\psi-\mu_r)}}{\sqrt{(2\pi)^d \det \Sigma_r}}, \quad 0 \leq \rho_r \leq 1, \quad \sum_{r=1}^{R} \rho_r = 1 \tag{7}$$

where the r-th component of the mixture is a multivariate Gaussian density with mean μ_r and covariance matrix Σ_r.

The maximum likelihood estimates of the parameters $\{\mu_r\}, \{\Sigma_r\}, \{\rho_r\}, r = 1, \ldots, R$, of the GM can be computed by means of the Expectation Maximization (EM) algorithm that guarantees convergence to a local maximum of the likelihood.

2.2. *Simulation*

The proposed probabilistic model takes the form of a Markov Chain [35, 36, 37, 38] and as such provides a simple way to simulate new realizations, interpretable as meal histories statistically consistent with the historical data used to identify the model.

The MC model is well suited to simulation because, conditional on the present state of the chain, the next state and the future values of the process (events and marks, in our case) are independent of the past values: the future values of the process (events and marks, in our case) are independent of the past values, but only depend the present values. In a discrete time MC, the core of the model are the so-called transition probabilities that specify the probability of the next values of the state for all the possible present values. In our meal model, the transition probabilities are time-dependent and periodic on the 24 hours in alignment with the circadian rhythm patterns [9, 39].

The process can be written as $X_1, X_2, X_3, \ldots$, where $X_k \in S$ is the state at time k, and the state space $S = s_0, s_1, \ldots$, is the set of possible states. The MC is characterised by the set of periodic transition probabilities

$$P_{ij}(k) = P(X_k = s_j | X_{k-1} = s_i), \quad P_{ij}(k+24) = P_{ij}(k) \tag{8}$$

In our MC model a notable simplification follows from the nature of the state which is just the duration of the fasting period. In particular, s_0 corresponds to no fasting (a meal has just occurred in the last time instant), s_1 corresponds to fasting lasting one time bin, s_2 to fasting lasting two time bins, and so on. Therefore, if $X_k = s_j$, the only possible transitions are to either $X_k + 1 = s_0$ (a meal event occurs) or $X_k + 1 = s_j + 1$ (no meal occurs and fasting increases by one time bin). This means that, for any state s_j, the needed transition probabilities can be obtained by a logistic regression that models the probability of a meal event as a function of daytime and other relevant covariates.

The transition across the states can be simulated by a standard rejection sampling approach. A sample value u from the random variable $U \in \mathcal{U} = [0, 1]$ is drawn at time k, and, if u is less than the probability provided by the logistic regression, an event occurs in k, else there is no meal event in k [40, 37, 38]. This immediately determines the transition towards the next state of the chain.

Whenever a meal occurs, a sample from the marker distribution is required. The sample is drawn from the conditional distribution of CHO intake given the context information, summarized by a subset of relevant covariates. This conditional distribution is obtained from the GM model and is itself a (univariate) GM.

2.3. Data

The Amsterdam dataset includes 697 meals across 8 patients, while the Padova dataset consists of 882 meals over 7 patients [2]. Even if the collection process suffered from some inaccuracies, no meals have been discharged because the aim is to use a dataset representative of real-life events.

Assuming that the minimum meal duration is 15 minutes, this implies that two consecutive meals cannot occur in less than 15 minutes. Thus, the bin size Δ was set equal to 15 minutes.

Letting N_P the number of patients in the dataset and N_{day} the total number of days per patient, then the total number of samples is $N_{day} \times N_P \times 96$, since each day we have 96 quarter-hour bins. A 1-day excerpt of the marked point process on the discrete time axis is displayed in **Figure 2**.

Considering the 24-hour periodicity of the time axis, the vector of the daytimes can be derived as follows:

$$T(i) = \mod(i, 96) \cdot \Delta \tag{9}$$

where $i = 1, 2, \ldots, N_{day} \times N_P \times 96$ and the corresponding vector of meal events can be expressed as follows:

$$Y(i) = \begin{cases} 1 & \text{a meal event occurs in the } i\text{-th time bin} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

It is convenient to define the marker array *cho* associated with *Y* as:

$$cho(i) = \begin{cases} \text{the carbohydrate amount in the } i\text{-th time bin if } Y(i) = 1 \\ 0 \text{ otherwise.} \end{cases} \tag{11}$$

To model the marked point process of the meal habits, the proposed approach uses the information available in the meal timestamp combined with the information of what happened in the past. Associated with each meal there are four variables:

- the daytime from 0-24 hours ($T(i)$);
- the amount of ingested carbohydrates ($cho(i)$) ;
- the length of the fasting period preceding the meal ($fp(i)$);
- the carbohydrate amount consumed at the previous meal ($\overleftarrow{cho}(i)$)

2.4. Empirical Maximum Mean Discrepancy

The purpose of the modelling effort is not fitting the Padova and Amsterdam datasets, but rather building a stochastic meal generator that produces random meal histories whose statistical properties are indistinguishable from the observed data. Therefore, the success of the model is not measurable in terms of residuals, either in crossvalidation or test, but calls for measuring the similarity between multivariate distributions. For this purpose, reference is made to the Maximum Mean Discrepancy, a widespread index used to compare multivariate distributions [29, 41]. The MMD statistic measures the difference between two probability distributions from their samples and it is defined as:

$$MMD(\mathcal{F}, P, Q)^2 = \left[\sup_{\|f\|_{\mathcal{H}} \leq 1} E_{x\sim P}[f(x)] - E_{y\sim Q}[f(y)] \right]^2 \tag{12}$$

where $E_{x\sim P}[f(x)]$ and $E_{y\sim Q}[f(y)]$ denote expectations with respect to P and Q, respectively, x and y are random samples drawn independent and identically distributed from P and Q, respectively, $\mathcal{F}$ is a "unit ball" in a Reproducing Kernel Hilbert Space (RKHS). Then $MMD(\mathcal{F}, P, Q)^2 = 0$ if and only if $P = Q$ [29]. In an RKHS, the squared MMD can be expressed in terms of kernel functions, and a corresponding unbiased finite sample estimate can be expressed, as follows:

$$MMD(\mathcal{F}, P, Q)^2 = E_{x,x'}[k(x, x')] + E_{y,y'}[k(y, y)] - E_{x,y'}[k(x, y)]. \tag{13}$$

It is possible to use an empirical formulation for the above equation with following formula:

$$\begin{aligned} MMD(\mathcal{F}, P, Q)^2 = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j}^{m} k(x_i, x_j) + \\ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j}^{n} k(y_i, y_j) - \frac{2}{mn} \sum_{i=1}^{m} \sum_{j}^{n} k(x_i, y_j) \end{aligned} \tag{14}$$

where x are the data points from the original dataset and y are the simulated samples. We used the MMD with radial basis function kernel where the bandwidth is selected by calculating the median of the pairwise distances between the samples [29]. To determine if empirical MMD (Eq. 14) shows a statistically significant difference between P and Q, it is required to compare the test statistic $MMD(\mathcal{F}, P, Q)$ with a particular threshold: if the threshold is exceeded, then the test rejects the null hypothesis $\mathcal{H}_0 : P = Q$. To estimate the threshold with a given acceptance level α,the bootstrap is used on the aggregated data, and the

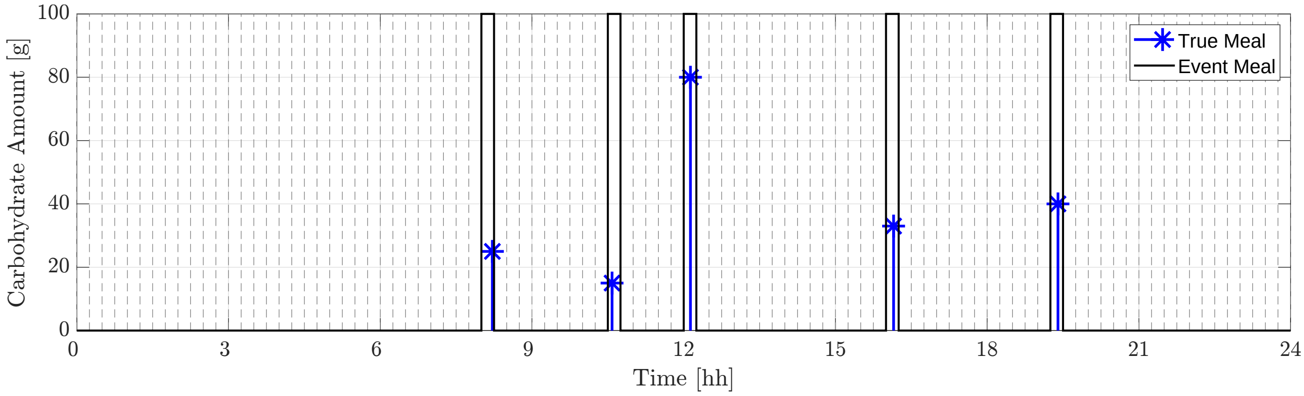

Figure 2: Discretization of the continuous-time marked point process. The discrete-time signal is nonzero whenever the continuous-time point process features a meal event within the given time bin. The amplitude of the signal coincides with the carbohydrate amount taken during the corresponding time interval.

p-value is estimated from the permutations as follows:

$$p = \frac{M + 1 - r}{M + 1} \tag{15}$$

where r denotes the position of the test statistic when sorting in ascending order the obtained $MMD(\mathcal{F}, P, Q)^2$ of the M permutations. In particular, the comparison between alternative models will be carried out by computing the Empirical MMD between the 4-dimensional distribution of T, cho, fp, $\overleftarrow{cho}(i)$ of the simulated and real meal histories. The comparison is based on the p-values, preferring the model with the largest p-value, i.e. the model that generates meal histories such that the null hypothesis of being statistically distributed as the real data is farthest from being rejected.

3. Results

The dataset used in this work was collected during experiments in free-living conditions involved in the "AP@home" project [42] in 2015. Specifically, the dataset includes clinical data belonging to 15 patients, who underwent a 1-month clinical trial that took place in two clinical centres of Padova and Amsterdam [2]. Data were collected in real-life condition and carbohydrates ingested at meal time or for snack and those used to treat hypoglycemic episodes were manually entered by the patient into a suitably modified android smartphone (the DiAs platform, [43]). In this dataset, each meal event was recorded as a combination of the timestamp and its associated carbohydrate amount.

Since each country is characterized by different habits, a separate analysis was performed for each clinical centre.A correlation analysis is performed to preliminarily quantify the associations between the available variables. Indeed, although patients with T1D are required to respect regular eating patterns, a causal model analysis of eating behavior can not be achieved in this work due to the lack of potential causal factors. The primary goal of this work is to design a simple, most likely and interpretable model, whose development is mainly based on the observed correlations among the small set of available features. **Figures 3** and **4** show the scatter matrix for these variables and, on the main diagonal, their marginal distributions for patients belonging to the Padova and Amsterdam clinical centres, respectively. Correlation coefficients are reported together with the regression lines. The comparison between the magnitude of correlation coefficients defines the order of importance in the relationships among the different variables.

In the Padova dataset the marginal distribution of the mealtime on a 24-hour period shows a trimodal shape, meaning that these patients typically eat three meals a day: breakfast, lunch and dinner. Breakfast is usually a light meal consumed between [7.00 - 10.00]; lunch is the main meal and generally is consumed in the interval [12.00 - 15.00]; finally, dinner is usually served late in the evening after 20.00. Moreover, a small snack may be taken at mid-morning or mid-noon. As patients data are mostly characterized by meals intake at regular intervals, the distribution of fasting periods preceding the meals appears having a bimodal shape on a daily basis. Fasting periods result on average between 6 and 8 hours during the daytime and longer across the night where the values ranged from 6 to 15 hours. The presence of the snacks all day around comes up with clusters of meals characterized by short fasting periods.

The Amsterdam cohort is characterized by multiple small meals throughout the day, as well as light snacks during night periods. These habits result in having in general short fasting

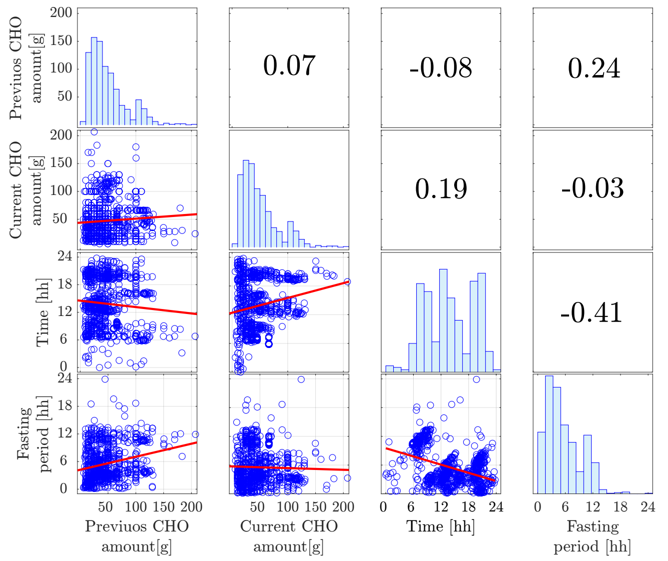

Figure 3: Scatter matrix for the four main variables characterizing the meal habits in the Padova dataset: previous CHO ($\overleftarrow{cho}(i)$), current CHO ($cho(i)$), time ($T(i)$), and fasting period ($fp(i)$). Main diagonal: histograms; lower panels: scatter plots; upper panels: correlation coefficients. These marginal and joint ditributions, together with the correlation coefficients, represent a statistical reference against which the accuracy of a candidate model can be assessed.

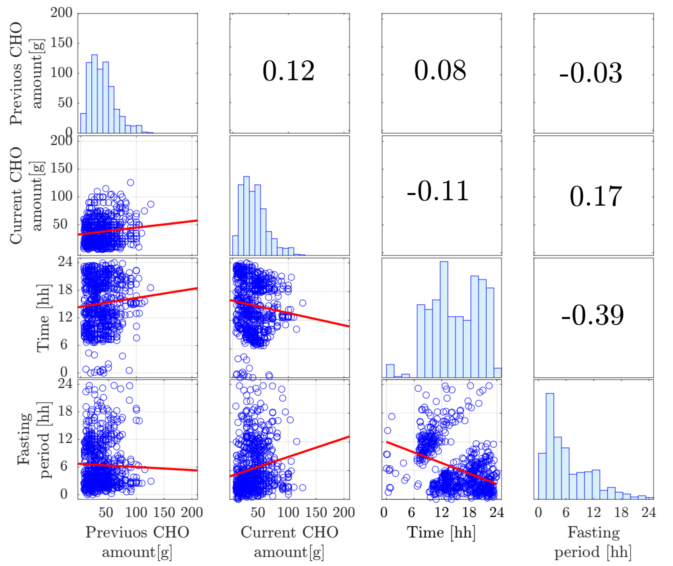

Figure 4: Scatter matrix for the four main variables characterizing the meal habits in the Amsterdam dataset: previous CHO ($\overleftarrow{cho}(i)$), current CHO ($cho(i)$), time ($T(i)$), and fasting period ($fp(i)$). Main diagonal: histograms; lower panels: scatter plots; upper panels: correlation coefficients. These marginal and joint distributions, together with the correlation coefficients, represent a statistical reference against which the accuracy of a candidate model can be assessed.

periods with few longer periods reported. The fairly regular timing of eating habits explains why the correlation between meal times and fasting periods results to be the strongest one with correlation coefficients equal to -0.41 and -0.39 for the Padova and Amsterdam datasets, respectively. Meal amounts show a skewed distribution. The Padova cohort ate meals with carbohydrate amount up to around 150 grams, while the Amsterdam cohort spans a narrower range.

Meal times show a weaker relationship with the associated amount of carbohydrate. In **Figure 4** the correlation is negative (-0.11) in the Amsterdam dataset, representing a tendency to consume larger meals early in the day, opposite to the Padova subjects who tend to consume larger meals late at night for which the correlation is positive (0.19) as shown in **Figure 3**.

Conversely, there is no apparent relationship with the carbohydrate amount of the previous meal having an even weaker correlation coefficient equal to -0.08 and 0.08 for the Padova and Amsterdam datasets, respectively. This implies that the daytime somehow affects the amount of the CHO intake, while meal time does not seem affected by the CHO amount of the previous meal.

For a given value of meal amount, the corresponding values of backward fasting periods range from 30 minutes to 10 hours. The relationship between the meal amount and the duration of the fasting period preceding the meal in the Padova cohort is described by the smallest correlation coefficient equal to -0.03. The meal amount, that, as already observed, does not show an effect on next meals's amount (0.07), appears to affect the fasting period after the meal as shown in **Figure 3** with a correlation coefficient equal to 0.24.

Meal amounts show a positive correlation with both the backward fasting period (0.17) and the size of the previous meal (0.12) in the Amsterdam dataset, as shown in **Figure 4**. The quantity of carbohydrates can be equally depending on the two variables. Although the fasting period affects clearly the meal time, the meal amount can be related either to the fasting period or the previous carbohydrate intake.

To sum up, the explorative analysis suggests that the occurrence of a meal is strongly affected by the current time and the fasting period in both cohorts. The associated CHO amount is mainly related to the daytime in the Padova dataset, while in the Amsterdam dataset it can depend on either the fasting period or the previous carbohydrate intake.

3.1. Logistic Regression

The logistic regression aims to model the dependence between each $Y(i)$ and the predictors $T(i)$, $fp(i)$, and $\overleftarrow{cho}(i)$. The time covariate has a periodic nature with period $\tau = 24$ hours in alignment with the circadian rhythm patterns [9, 39] To further study the periodicity of the Probability Density Function (PDF) of the meal times, first, an histogram was used to estimate an approximated PDF of the meal time samples and the Square Root rule was used to choose the number of bins (N_{bin}). Given the size of dataset, N_{bin} is equal to 30 and 26 in the Padova and Amsterdam datasets, respectively. Then, interpreting the PDF approximation as a time signal, the Fourier transform can be used to highlight its main harmonic components. It is worthy to highlight that N_{bin} defines the associated bin width, which turns up being the sampling time of the PDF time signal. With the above choice, the Nyquist frequency is defined as $\omega_N = \frac{\pi * N_{bin}}{24}$ $[rad/hour]$ [44].The Fourier transform shows that the signal spectra have 12 components each in total: the fundamental harmonic has period $\tau_1 = 1/24$, and the principal components span up to τ_7. A notable improvement is however noticeable when including up to the 10^{th} components, while the harmonics τ_{11} and τ_{12} were not considered because they are very close to ω_N. Based on this analysis, the time covariate was expressed as a linear combination of sinusoids and cosinusoids with periods $\tau_h = \frac{\tau}{h}$ with $h = 1, 2, \ldots, 10$. Therefore, the candidate regressors were:

$$\begin{aligned}\Phi(i) = [1, \sin\left(\frac{2\pi T(i)}{\tau_1}\right), \cos\left(\frac{2\pi T(i)}{\tau_1}\right), \ldots,\\ \sin\left(\frac{2\pi T(i)}{\tau_{10}}\right), \cos\left(\frac{2\pi T(i)}{\tau_{10}}\right), fp(i), \overleftarrow{cho}(i)]\end{aligned} \tag{16}$$

The results of the regression analysis on the above candidate regressors are reported in the Supplementary material where the estimated coefficients, as well as the standard error (SE), T-statistics, p-values and the 95% confidence interval (CI). Specifically, **Supplementary tables S1** and **S2** provide a list of the information concerning the Padova and Amsterdam dataset, respectively. It is worth highlighting that, from the scatter plot reported in **Figure 3**, the relation between the meal time and $\overleftarrow{cho}$ appears negligible, with a correlation coefficient equal to −0.08 in the Padova cohort. The same consideration concerns the Amsterdam cohort showing a correlation coefficient equal to 0.08 between time and $\overleftarrow{cho}$, as shown in **Figure 4**. It can be noted in **Supplementary tables S1** and **S2** that the coefficients associated with $\overleftarrow{cho}$ are close to zero confirming the hypothesis of no relationship. Based on this consideration, the regressor $\overleftarrow{cho}$ was removed and a new model was identified for both cohorts. The logistic regression models based on the reduced set of regressors are reported in **Supplementary tables S3** and **S4**. Aiming to seek the contribution of the carbohydrate amount consumed at the previous meal, the BIC criterion was considered to choose between the two competing logistic regression models. The BIC associated with the complete set of regressors was 6136, while a BIC of 6134 was obtained by removing $\overleftarrow{cho}$ in the Padova dataset. The same results occurred in the Amsterdam dataset obtaining a BIC value of 5474 with the complete set of regressors, and 5471 discarding $\overleftarrow{cho}$. The achieved BICs suggested to choose the simpler model where the meal event occurrence depends on the fasting period and the current time only, for both datasets.

3.2. Gaussian Mixture model

The set of the CHO intakes associated with the meal events is given by $\{cho_m(i)\}, i \in M$, where $M = \{i | cho(i) > 0\}$. The corresponding daytimes, fasting periods and previous CHO intakes are $\{T_m(i)\}$, $\{fp_m(i)\}$ and $\{\overleftarrow{cho}_m(i)\}, i \in M$. To investigate which covariate does have an impact on predicting $\{cho_m\}$, three separate GM models were identified with $\psi_1 = [cho_m, T_m]'$, $\psi_2 = [cho_m, fp_m]'$, and $\psi_3 = \left[cho_m, \overleftarrow{cho}_m\right]'$, and three corresponding meal histories are generated using the MC model. Then, as described in Section 2.4, the empirical estimate of MMD between the 4-dimensional distribution of T, cho, fp, $\overleftarrow{cho}(i)$ of the simulated and real meal history data, was applied to test which combinations of covariates gives the model that generates meal histories statistically distributed as the real data. Using a permutation procedure based on 100 repetitions, the null hypothesis is always accepted for the Padova dataset for $\alpha = 0.01$, and the largest p-value is obtained with $\psi_1 = [cho_m, T_m]'$, as reported in Table 1.

Since T_m can satisfactorily predict cho_m, the GM model with $\psi = [cho_m, T_m]'$ was included in the stochastic meal generator. **Figure 5** shows on the left side the identified Probability Density Function (PDF), as well as the scatter plot of the ψ samples on the right. The GM model includes four Gaussian distribution: a baseline low-CHO distribution accounting for small meals consumed thoughout the day and three distributions matching the three clusters of meals at breakfast, lunch and dinner. The same analysis was performed on the Amsterdam dataset to identify the subset of explanatory variables that exhibit predictive capabilities for cho_m in this specific dataset. Using a permutation procedure based on 100 repetitions, the null hypothesis is accepted only for the combination $\psi_1 = [cho_m, T_m]'$, as reported in Table 1, and the corresponding model GM model was included in the generator, ending with the same covariates to build the GM model for both populations. Despite of its significance, the small p-value can be attributed to the fact that the occurrence of a meal and its amount may depend on several different causes, such as the distribution of energy and macronutrient intakes throughout the day, as well as psycho-behavioural mechanisms, which were

	MMD		
	$\psi_1 = [cho_m, T_m]'$	$\psi_2 = [cho_m, fp_m]'$	$\psi_3 = \left[cho_m, \overleftarrow{cho_m}\right]'$
Padova	0.33	0.19	0.24
Amsterdam	0.03	0	0

Table 1: P-values of the Maximum Mean Discrepancy test for the different candidate GM models.

not annotated during the study. Additional details on the meals could highly improve this value. **Figure 6** shows both the obtained density (left) and the scatter plot (right). The GM model contains three Gaussian distributions: a baseline low-CHO distribution accounting for small meals consumed thoughout the day, and two clusters representing the meals consumed before noon and late at night, respectively. It is interesting to note that the cluster of large meals recorded around noon is present in the GM model.

3.3. Markov chain

As already observed, the states of the MC are the fasting periods and the state space is the set of all positive integers including 0, e.g $S = 0, 1, 2, \ldots$. At times $k = 0, 1, 2, \ldots$, two admissible events can occur at each state j, that is an extension of the fasting or a meal, as shown in **Figure 7**. The corresponding transition probabilities are:

$$\begin{aligned} P_{j,j+1}(k) =& P(fp(k+1) = j+1 | fp(k) = j, k) \\ P_{j,0}(k) =& P(fp(k+1) = 0 | fp(k) = j, k) \end{aligned} \tag{17}$$

The transition probability is null for all other cases. Recalling the daily periodicity, the above probabilities can be rewritten as follows:

$$\begin{aligned} P_{j,j+1}(k) =& P(Y(k+1) = 0 | fp(k) = j, T(k)) \\ P_{j,1}(k) =& P(Y(k+1) = 1 | fp(k) = j, T(k)) \end{aligned} \tag{18}$$

that can be computed applying the logistic regression models described in Section 3.1. As already observed, the MC model can be used to easily simulate the point process of the meal events. To guarantee the fasting periods to be non negative integers, at the beginning of the simulation, the state of the chain is initialized to 0, then, at each time k, the state increases by one if no meal occurs. Conversely, if a meal event occurs, the fasting period is reset to 0. Moreover, to assure $T(k)$ to be non negative integers, at the beginning of the simulation, we initialized $k = 0$, then at each time the counter k increases by one, while the module operator guarantees the daily periodicity, as described in Eq. 9.

To add the markers, whenever a meal event occurs, a sample of CHO amount is drawn from the conditional distribution of cho_m given the relevant covariates, a conditional distribution that is rather straightforwardly derived from the GM model.

3.4. Comparison between observed and simulated meals

Figure 8 shows the comparison between the original Padova dataset and a simulated dataset of the same size. In the simulated dataset, both the CHO amount and the meal times are uncorrelated with the previous CHO intake, as seen in Section 2.3. The fasting period resulted having a weak correlation with the current CHO amount as well as the previous CHO intake. The correlation coefficient between the meal time and the CHO amount was 0.29 compared to 0.19 in the Padova dataset. It is worth observing that the simulated data are able to replicate well the strongest correlation observed in the original dataset, that is the correlation between the time and the fasting period. In fact, a correlation coefficient of −0.39 was found in the simulated data, compared to a value of −0.41 observed in the original dataset. Coming to the simulation of the eating process of the Amsterdam cohort, **Figure 9** shows that the previous CHO intake does not affect the next meal in term of CHO intake, but shows a negative correlation, although not significant, with the fasting, as can be observed in the original dataset. The correlation between the CHO amount and the time was −0.011, in very good agreement with −0.06 observed in the original dataset. In simulation, the meal amount turned out having a correlation of 0.06 with the fasting period, slightly smaller than the 0.17 observed in the original dataset, but both statistically significant. The strongest correlation in this dataset was the one between time and fasting period, equal to −0.32. In simulation this correlation turned out to be a statistically significant −0.31, but more remarkably, the simulated scatter plot was able to reproduce the two clusters separated by an empty region that are a peculiar feature of the experimental scatter plot. **Figures S1** and **S2** in the Supplementary

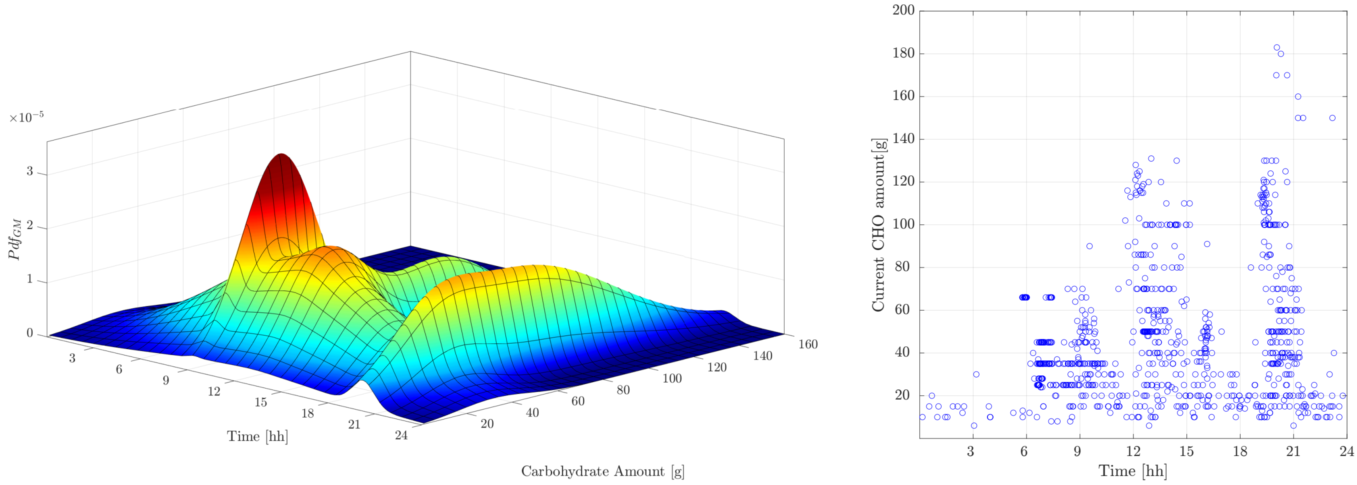

Figure 5: Padova dataset: Gaussian Mixture model of the joint distribution of daytime and CHO intake. Left: estimated probability density function; Right: scatter plot of the training data. The model of the joint distribution allows the computation of the conditional distribution of CHO intake given a specified daytime, which is the key information needed for the random simulation of the marks of the marked point process.

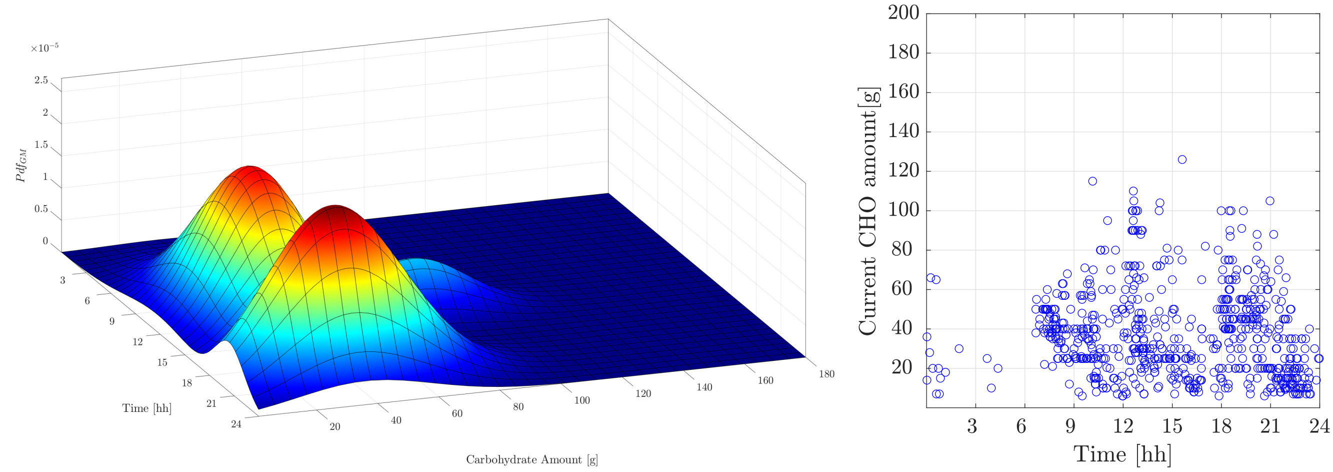

Figure 6: Amsterdam dataset: Gaussian Mixture model of the joint distribution of time and CHO intake. Left: estimated probability density function; Right: scatter plot of the training data. The model of the joint distribution allows the computation of the conditional distribution of CHO intake given a specified daytime, which is the key information needed for the random simulation of the marks of the marked point process.

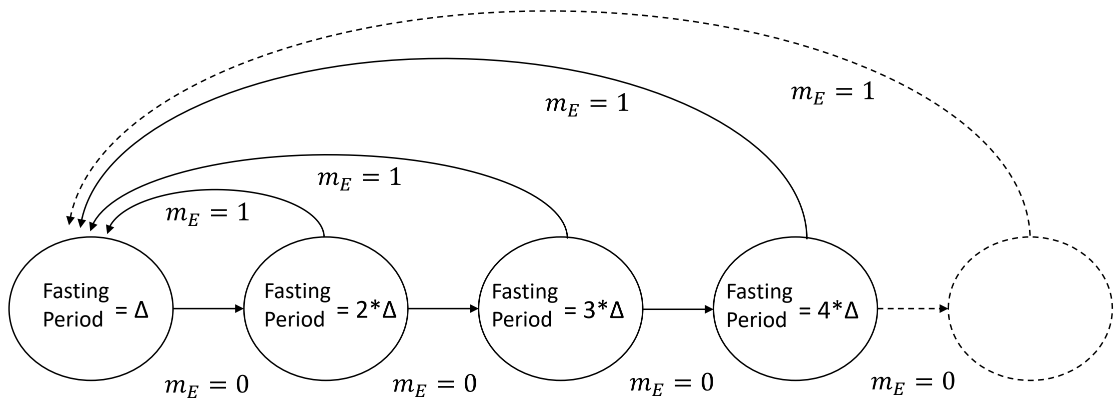

Figure 7: Markov Chain representation of the meal model. The chain states identify different values of the fasting period. If a meal event occurs, the fasting period is reset (backward arrow), otherwise it undergoes a unit increase (forward arrow). The transition probabilities may depend on the fasting period as well as on daytime and previous CHO.

materials shows the comparison between the Amsterdam dataset and the simulated dataset corresponding to the GM models, that were identified with $\psi_2 = [cho_m, fp_m]'$ and $\psi_3 = \left[cho_m, \overleftarrow{cho}_m\right]'$, respectively.

4. Discussion

The proposed approach provides a compact solution to the problem of modelling and simulating the eating behaviours of a T1D population. The simulation tool has a parsimonious structure with a small set of parameters. Apart from the merit of a simple and meaningful structure, the strength of the approach is its ability to capture and reproduce different eating habits, as practically demonstrated on datasets collected in Italy and Netherlands. Our methodology can also be applied to leverage data from a previously conducted clinical study in order to run simulations that help the design of a future study. In principle, the same procedure could be easily replicated for a specific patient, provided that an adequately sized dataset is available. The availability of large individual datasets is something that is soon to happen, due to the increasing ease in collecting this type of data using wearable devices [9]. Moreover, if a longer dataset was available, it would be interesting to investigate the seasonal variation to describe if and how eating patterns change across the year. Additionally, physical activity can highly impact the meal habits. Although the study records used in this work do not include any exercise information, it would be stimulating to perform an extended analysis including physical activity information in the model development.

Having a flexible and efficient model of the eating habits of a

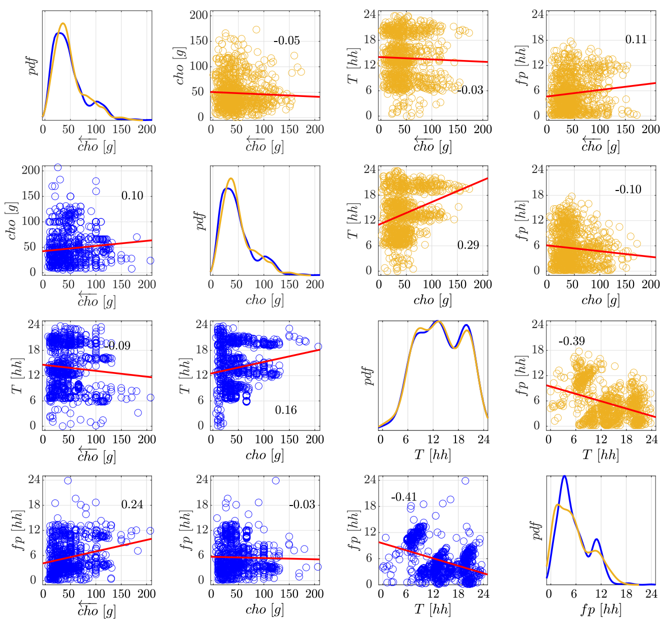

Figure 8: Scatter matrix for the four main variables characterizing the meal habits in the Padova dataset: previous CHO ($\overleftarrow{cho}(i)$), current CHO ($cho(i)$), time ($T(i)$), and fasting period ($fp(i)$). Main diagonal: histograms of the original data (blu) and simulated data (orange); lower panels: scatter plots of the original data (blue); upper panels: scatter plots of the simulated data (orange). The correlation coefficients are also reported. The good agreement between the original and simulated distributions provides a validation of the stochastic simulation model.

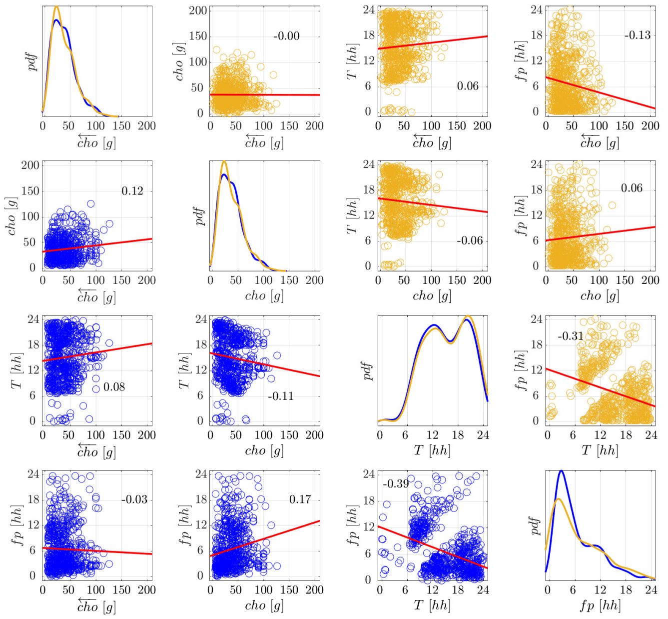

Figure 9: Scatter matrix for the four main variables characterizing the meal habits in the Amsterdam dataset: previous CHO ($\overleftarrow{cho}(i)$), current CHO ($cho(i)$), time ($T(i)$), and fasting period ($fp(i)$). Main diagonal: histograms of the original data (blu) and simulated data (orange); lower panels: scatter plots of the original data (blue); upper panels: scatter plots of the simulated data (orange). The correlation coefficients are also reported. The good agreement between the original and simulated distributions provides a validation of the stochastic simulation model.

population would also be of great help to customize automatic insulin dosing systems, tailoring them to specific eating scenarios, e.g. typical of a certain country or patient category.

Our model could also play a role in the design of meal detection algorithms. While it is clear from clinical results that meal announcement represents an additional information, that can improve the critical postprandial glycemic control, the reality is that patients with T1D often forget to announce the meal. Although the AID systems should remain able to operate safely in case of a missing announcement, this condition represents a challenge.The analysis of the glucose profile is obviously the core of an effective automatic meal detection scheme. Our probabilistic model should be regarded as complementary, in that it provides the probability of meal time and amount before the knowledge glucose level is made available. This means that, rather an alternative way of detecting meals, it could provide the prior probability useful to develop a Bayesian meal detector. More precisely, according to the Bayes formula,

$$P(m_E(t)|g(\cdot), Y) \propto P(g(\cdot)|m_E(t), Y)P(m_E(t)|Y) \tag{19}$$

where $m_E(t)$ is a meal event at time t, $g(\cdot)$ is the recorded glucose level history and Y denotes the relevant information e.g. daytime, fasting period, the CHO amount of the previous meal, previous insulin administration, and so on. Our probabilistic model would provide the prior probability $P(m_E(t)|Y)$ that multiplies the likelihood $P(g(\cdot)|m_E(t), Y)$, typically derived from glucose metabolism models. The availability of a realistic model of the prior probability is expected to reduce the uncertainty inherent in data-driven meal detection.

Moreover, even when meal is announced, the controller becomes aware of the meal just at the time of the announcement.Since glucose variations induced by the meals intake are a disturbance for an AID system, the proposed model can also be used to predict the disturbance in the context of model predictive control schemes [11, 45]. In fact, our model aims to provide population-level assumptions about eating behavior that could allow to anticipate future meals. For example, if a patient has a meal early in the morning, then it is unlikely that the next meal will occur within the next hour or after noon. However, if there is no meal announcement after a long fasting period according to the patient's habits, the probability adapts to an increasing chance of the patient eating soon.

Conclusions

This work proposes a methodology to identify eating behavioural patterns of individuals with type 1 diabetes. The proposed approach describes the eating events as a marked point process where the event occurrence is modeled by a logistic regression, while the mark distribution is formulated as a GM model. A MC framework was then applied to efficiently generate of synthetic datasets whose statistical properties are in accordance with those of the original datasets. Due to its simplicity, the model can run online for safety and decision support purposes. Finally, it should be remarked that this approach can be straightforwardly applied to obtain a patient-tailored model if an adequate dataset is available.

Conflict of interest

EMA is currently with Harvard University and this work was done when she was with University of Pavia. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

EMA, GDN: contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. CT, LM: contributed to the reviewing of the manuscript. All authors approved the submitted version.

References

[1] R. J. Cook, J. F. Lawless, Multistate models for the analysis of life history data, Chapman and Hall/CRC, 2018.

[2] E. Renard, A. Farret, J. Kropff, D. Bruttomesso, M. Messori, J. Place, R. Visentin, R. Calore, C. Toffanin, F. Di Palma, et al., Day-and-night closed-loop glucose control in patients with type 1 diabetes under free-living conditions: results of a single-arm 1-month experience compared with a previously reported feasibility study of evening and night at home, Diabetes Care 39 (7) (2016) 1151–1160.

[3] J. Schofield, J. Ho, H. Soran, Cardiovascular risk in type 1 diabetes mellitus, Diabetes Therapy 10 (3) (2019) 773–789.

[4] A. D. Association, et al., Professional practice committee: Standards of medical care in diabetes—2021 (2021).

[5] J. L. Chiang, M. S. Kirkman, L. M. Laffel, A. L. Peters, Type 1 diabetes through the life span: a position statement of the american diabetes association, Diabetes care 37 (7) (2014) 2034–2054.

[6] M. Ghadessi, R. Tang, J. Zhou, R. Liu, C. Wang, K. Toyoizumi, C. Mei, L. Zhang, C. Deng, R. A. Beckman, A roadmap to using historical controls in clinical trials–by drug information association adaptive design scientific working group (dia-adswg), Orphanet journal of rare diseases 15 (1) (2020) 1–19.

[7] W. Tang, X. Tu, Modern clinical trial analysis, Springer, 2012.

[8] A. L. Olinder, A. Kernell, B. Smide, Missed bolus doses: devastating for metabolic control in CSII-treated adolescents with type 1 diabetes, Pediatr Diabetes 10 (2) (2009) 142–148.

[9] E. M. Aiello, S. Deshpande, B. Özaslan, K. L. Wolkowicz, E. Dassau, J. E. Pinsker, F. J. Doyle, Review of automated insulin delivery systems for individuals with type 1 diabetes: tailored solutions for subpopulations, Current Opinion in Biomedical Engineering 19 (2021) 100312.

[10] E. Dassau, B. W. Bequette, B. A. Buckingham, F. J. Doyle, Detection of a meal using continuous glucose monitoring: implications for an artificial β-cell, Diabetes care 31 (2) (2008) 295–300.

[11] H. Lee, B. W. Bequette, A closed-loop artificial pancreas based on model predictive control: Human-friendly identification and automatic meal disturbance rejection, Biomedical Signal Processing and Control 4 (4) (2009) 347–354.

[12] F. Cameron, G. Niemeyer, B. A. Buckingham, Probabilistic evolving meal detection and estimation of meal total glucose appearance (2009).

[13] F. Cameron, G. Niemeyer, B. W. Bequette, Extended multiple model prediction with application to blood glucose regulation, Journal of Process Control 22 (8) (2012) 1422–1432.

[14] K. Turksoy, S. Samadi, J. Feng, E. Littlejohn, L. Quinn, A. Cinar, Meal detection in patients with type 1 diabetes: a new module for the multivariable adaptive artificial pancreas control system, IEEE journal of biomedical and health informatics 20 (1) (2015) 47–54.

[15] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, et al., The UVA/Padova type 1 diabetes simulator: New features, Journal of Diabetes Science and Technology 8 (1) (2014) 26–34.

[16] J. Xie, Q. Wang, Meal detection and meal size estimation for type 1 diabetes treatment: a variable state dimension approach, in: Dynamic Systems and Control Conference, Vol. 57243, American Society of Mechanical Engineers, 2015, p. V001T15A003.

[17] S. Chen, J. Weimer, M. R. Rickels, A. Peleckis, I. Lee, Towards a model-based meal detector for type i diabetics.

[18] J. Weimer, S. Chen, A. Peleckis, M. R. Rickels, I. Lee, Physiology-invariant meal detection for type 1 diabetes, Diabetes technology & therapeutics 18 (10) (2016) 616–624.

[19] C. M. Ramkissoon, P. Herrero, J. Bondia, J. Vehi, Meal detection in the artificial pancreas: implications during exercise, IFAC-PapersOnLine 50 (1) (2017) 5462–5467.

[20] Y. Ogata, Space-time point-process models for earthquake occurrences, Annals of the Institute of Statistical Mathematics 50 (2) (1998) 379–402.

[21] R. D. Peng, F. P. Schoenberg, J. A. Woods, A space–time conditional intensity model for evaluating a wildfire hazard index, Journal of the American Statistical Association 100 (469) (2005) 26–35.

[22] P. Hall, N. Tajvidi, Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, Statistical Science (2000) 153–167.

[23] A. Baddeley, M. Berman, N. I. Fisher, A. Hardegen, R. K. Milne, D. Schuhmacher, R. Shah, R. Turner, Spatial logistic regression and change-of-support in poisson point processes, Electronic Journal of Statistics 4 (2010) 1151–1201.

[24] D. R. Brillinger, Comparative aspects of the study of ordinary time series and of point processes, in: Developments in statistics, Vol. 1, Elsevier, 1978, pp. 33–133.

[25] D. R. Brillinger, H. K. Preisler, Two examples of quantal data analysis: a) multivariate point process, b) pure death process in an experimental design, in: Proc. XIII International Biometric Conference, Seattle, 1986, pp. 94–113.

[26] D. R. Brillinger, J. P. Segundo, Empirical examination of the threshold model of neuron firing, Biological Cybernetics 35 (4) (1979) 213–220.

[27] P. Hougaard, Multi-state models: a review, Lifetime data analysis 5 (3) (1999) 239–264.

[28] L. Meira-Machado, J. de Uña-Álvarez, C. Cadarso-Suárez, P. K. Andersen, Multi-state models for the analysis of time-to-event data, Statistical methods in medical research 18 (2) (2009) 195–222.

[29] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, The Journal of Machine Learning Research 13 (1) (2012) 723–773.

[30] M. Jacobsen, Point process theory and applications: marked point and piecewise deterministic processes, Springer Science & Business Media, 2006.

[31] Y. Guan, Tests for independence between marks and points of a marked point process, Biometrics 62 (1) (2006) 126–134.

[32] T. Zhang, On independence and separability between points and marks of marked point processes, Statistica Sinica (2017) 207–227.

[33] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer Science & Business Media, 2009.

[34] B. G. Lindsay, Mixture models: theory, geometry and applications, in: NSF-CBMS regional conference series in probability and statistics, JSTOR, 1995, pp. i–163.

[35] P. K. Andersen, N. Keiding, Multi-state models for event history analysis, Statistical methods in medical research 11 (2) (2002) 91–115.

[36] R. Durrett, R. Durrett, Essentials of stochastic processes, Vol. 1, Springer,

1999.
[37] A. B. Owen, Monte carlo theory, methods and examples.
[38] D. P. Kroese, T. Taimre, Z. I. Botev, Handbook of monte carlo methods, Vol. 706, John Wiley & Sons, 2013.
[39] E. Van Cauter, K. S. Polonsky, A. J. Scheen, Roles of circadian rhythmicity and sleep in human glucose regulation, Endocrine reviews 18 (5) (1997) 716–738.
[40] R. C. Larson, A. R. Odoni, Urban operations research, no. Monograph, 1981.
[41] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, A. J. Smola, Integrating structured biological data by kernel maximum mean discrepancy, Bioinformatics 22 (14) (2006) e49–e57.
[42] L. Heinemann, C. Benesch, J. H. DeVries, A. home consortium, Ap home: a novel european approach to bring the artificial pancreas home (2011).
[43] P. Keith-Hynes, B. Mize, A. Robert, J. Place, The Diabetes Assistant: A smartphone-based system for real-time control of blood glucose, Electronics 3 (4) (2014) 609–623.
[44] L. Ljung, System identification, in: Signal analysis and prediction, Springer, 1998, pp. 163–173.
[45] F. M. Cameron, T. T. Ly, B. A. Buckingham, D. M. Maahs, G. P. Forlenza, C. J. Levy, D. Lam, P. Clinton, L. H. Messer, E. Westfall, et al., Closed-loop control without meal announcement in type 1 diabetes, Diabetes technology & therapeutics 19 (9) (2017) 527–532.