



**HAL**  
open science

## Characterization of Daily Glycemic Variability in Subjects with Type 1 Diabetes Using a Mixture of Metrics

Fei Zheng, Manon Jalbert, Florence Forbes, Stephane Bonnet, Anne Wojtusciszyn, Sandrine Lablanche, Pierre-Yves Benhamou

► **To cite this version:**

Fei Zheng, Manon Jalbert, Florence Forbes, Stephane Bonnet, Anne Wojtusciszyn, et al.. Characterization of Daily Glycemic Variability in Subjects with Type 1 Diabetes Using a Mixture of Metrics. *Diabetes Technology and Therapeutics*, 2020, 22 (4), pp.301-313. 10.1089/dia.2019.0250 . hal-02415078

**HAL Id: hal-02415078**

**<https://hal.science/hal-02415078v1>**

Submitted on 17 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Characterization of Daily Glycemic Variability in Subjects with Type 1 Diabetes Using A Mixture of Metrics

Fei Zheng<sup>(1,2)</sup>, Manon Jalbert<sup>(3)</sup>, Florence Forbes<sup>(1)</sup>, Stéphane Bonnet<sup>(2)</sup>, Anne Wojtuszczyzn<sup>(4)</sup>,  
Sandrine Lablanche<sup>(3)</sup>, Pierre-Yves Benhamou<sup>(3)</sup>

<sup>(1)</sup>Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble, France.

<sup>(2)</sup>CEA LETI, DTBS, Univ. Grenoble Alpes, Minatec Campus, Grenoble, France.

<sup>(3)</sup>Endocrinologie Diabétologie Nutrition, CHU Grenoble-Alpes.

<sup>(4)</sup>Endocrinologie Diabétologie Nutrition, CHU Montpellier.

## Abstract

**Background:** Glycemic variability is an important component of glycemic control for patients with type 1 diabetes. The inadequacy of existing measurements lies in the fact that they view the variability from different aspects, so that no consensus has been reached among physicians as to which metrics to use in practice. Moreover, although glycemic variability, from one day to another, can show very different patterns, few metrics have been dedicated to daily evaluations.

**Method:** A reference (stable-glycemia) statistical model is built based on a combination of daily computed canonical glycemic control metrics including variability. The metrics are computed for subjects from the TRIMECO islet transplantation trial, selected when their  $\beta$ -score (composite score for grading success)  $\geq 6$  after a transplantation. Then, for any new daily glycemia recording, its likelihood with respect to this reference model provides a multi-metric score of daily glycemic variability severity. In addition, determining the likelihood value that best separates the daily glycemia with  $\beta$ -score = 0 from that with  $\beta$ -score  $\geq 6$ , we propose an objective decision rule to classify daily glycemia into "stable" or "unstable".

**Results:** The proposed characterization framework integrates multiple standard metrics and provides a comprehensive daily glycemic variability index, based on which, long term variability evaluations and investigations on the implicit link between variability and  $\beta$ -score can be carried out. Evaluation, in a daily glycemic variability classification task, shows that the proposed method is highly concordant to the experience of diabetologists.

**Conclusion:** A multivariate statistical model is proposed to characterize the daily glycemic variability of subjects with type 1 diabetes. The model has the advantage to provide a single variability score that gathers the information power of a number of canonical scores, too partial to be used individually. A reliable decision rule to classify daily variability measurements into stable or unstable is also provided.

**Keywords:** Continuous glucose monitoring, Type 1 diabetes, Islet cell transplantation, Glycemic variability, Statistical mixture models, Anomaly detection.

# 1 Introduction

Glycemic variability (GV) is under scrutiny as an important contributing factor in the development of diabetic complications and it can greatly affect the quality of life of subjects with type 1 diabetes (D1T) [1] [2]. Studies evidenced the association between GV and mortality in critically ill patients [3] [4] [5] and GV appears as an important indicator of the quality of the glycemic control.

Various metrics have been developed to assess GV [6] [7]. The conventional GV indices have been mostly developed based on self-monitoring of blood glucose (SMBG), before the appearance of continuous glucose monitoring (CGM) systems. Besides the basic statistical measurements such as Standard Deviation (SD), Coefficient of Variation (CV) and InterQuartile Range (IQR), there are still quite a number. Among them, may be cited the Mean Amplitude of Glycemic Excursion (MAGE) which aims at considering only major glycemic excursions, excluding minor ones (only glycemic variations that exceed one SD of the average blood glucose are included in the calculation). MAGE appears as an informative expression of the intra-day GV in insulin-treated diabetic patients [8] [9]. The Average Daily Risk Range (ADRR) [10] [11] summarizes the risk of hyper and hypoglycemia generated from risk functions based on skewness-corrected glycemia. Among the rare works aiming at daily difference evaluation, the Continuous Overlapping Net Glycemic Action (CONGA) [12] counts the SD of the glycemic differences at a certain time interval. The recent Glycemic Variability Percentage (GVP) [13] whose definition is similar to the Mean Absolute Glucose (MAG) [14] and the Lability Index (LI) [15] but is more suitable for continuous monitoring data. It is worth mentioning that the conventional metrics can be applied on continuous monitoring data as well, with a possible improvement of their characterization ability thanks to the increased sample rate brought by CGM.

Actually, the frontier between “glycemic control” and “glycemic variability” indexes is somehow difficult to draw. GV is undoubtedly an important parameter to analysis the overall glycemic control so that some indices propose to measure the overall glycemic control including the glycemic variability. Most of them are composite metrics, such as the J-index [16] which perpetuates the inclusion of Mean and SD into one measurement, the M-value [17] which measures the difference between the observed glycemia and an Ideal Glucose Value (IGV, normally taken as 6.67 mmol/L) while incorporating the glycemic range, and more recently, the COGI [18], a weighted linear combination of TIR, TBR and SD. One can refer to [19] [20] [21] [22] for other composite metrics.

Despite the abundant literature, a precise and complete GV analysis remains difficult to perform in clinical practice. No consensus has been reached on which GV metrics should be used to ideally and simply characterize GV. Each index analyses a specific component of GV and an index describing all the components of glycemic variability as a whole is currently lacking. A natural idea is to combine several metrics in an optimized way to offer clinicians such a possibility. Attempts include the work in [23] [24] which use Support Vector Machine (SVM) [25] or Multilayer Perceptron to obtain a unit index through learning from labels given by physicians, but the discriminative power of this index is then strongly dependent on the experience of the included physicians.

Here, we propose a new optimized GV index that fuses several existing metrics and that takes into account the multiple aspects of GV. This GV index allows the classification of patients into “stable” and “unstable” glycemic profile. To do so, a population of patients with stable glycemia is referred to as the reference population. We compute daily the selected GV metrics for each individual in the population and characterize the joint statistical distribution of the

considered metrics by learning from them a multivariate parametric statistical model. Whatever the number of existing metrics being used, for any new observed glycemic profile, we can transform our multivariate distribution into a scalar score by considering the log-likelihood of the new observation in the learned model. The rationale is that a high log-likelihood occurs when the glycemic profile is close to a “stable” (reference) one. In contrast, a low log-likelihood means that the glycemic profile has no or very few equivalent in the reference population. In practice, the possibility given by the proposed GV index to clinicians to analyze and classify a patient into a “stable” or “unstable” glycemic profile is an helpful feature to guide patient therapy decision and control. In this work, we show how to use the proposed log-likelihood score to design a separation rule, the difficulty being to decide under which value the log-likelihood is too small to be considered as coming from a stable reference subject. A way to automatically set this threshold is provided and illustrated on real data showing good performance.

The rest of the paper is organized as follows. Section 2 describes the general statistical characterization framework including reference model construction and classification methods. To apply the framework to our specific GV characterization problem, the pre-processing of CGM data and GV feature selection are addressed. Section 3 illustrates the results on data from the TRIMECO trial [26]. The performance and robustness of the proposed index is illustrated by testing its consistency with physicians in a classification task. Section 4 discusses the implicit link between the clinical  $\beta$ -score [27] and GV by using the proposed index. The contribution of this work and some perspectives are given in the final Section 5.

## 2 Mixture Model-based Anomaly Detection for Glycemic Variability Classification and Characterization

We address the GV evaluation problem with a model-based anomaly detection method which assumes that the properties of the normal population can be captured by a statistical model. Therefore, to measure how much an observation inclines to be normal is equivalent to measure the distance of the observation to the reference model. Various choices of models are available for the reference construction [28]. Examples as mentioned before, the SVM [23] [24] tries to separate linearly the anomalies from normal ones by projecting them into a higher dimension space. Neural networks such as autoencoders [29] assume that the normal data can be reconstructed by the network while the reconstruction errors of anomalies are much higher. Besides their popularity on various applications, the choice of their kernel functions and the hyper-parameters are subtle and ambiguous. In contrast, the so-called mixture model is chosen here especially for its interpretability in a medical diagnosis context.

### 2.1 Framework

Figure. 1 summarizes the framework of the anomaly detection process. The method requires normal reference (inlier) learning samples for estimating appropriately the parameters of the statistical model (see appendix 8.1), so that they are given high probabilities by the model. The similarity of any input data compared to the reference population is then measured by the log-likelihood of the reference model. In this proposed framework, anomaly (outlier) samples are used to find a threshold to decide whether an input data is belonging or not to the reference population.

#### 2.1.1 Reference Model Construction

The applied statistical model is a Mixture of Multiple-Scaled T-distributions (MMST) firstly proposed in [30]. Data  $\mathbf{x}$  is assumed to follow the density of a  $K$  components MMST denoted

by  $p(\mathbf{x}|\boldsymbol{\varphi})$  with  $\boldsymbol{\varphi}$  representing the model parameters. Given an inlier learning set  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ , the model parameter  $\boldsymbol{\varphi}$  can be estimated through Expectation-Maximization (EM) algorithm [30] [31] and the component number  $K$  can be decided from a possible range using the Bayesian Information Criterion (BIC) [32]. The MMST model and estimation strategy are detailed in Appendix 8.1.

### 2.1.2 Index and Threshold Decision

Once the reference model is learnt, the log-likelihood of the model denoted as  $\mathcal{L}_{\mathbf{x}}$  is taken as an index for measuring the similarity of any new observation  $\mathbf{x}$  to the reference population. Outliers of the reference model have relatively smaller log-likelihood values than inliers, then the problem comes to find a threshold which can appropriately separate them. We address this problem numerically by classifying the log-likelihood values of all samples into outliers and inliers trying various threshold values, and find the one that matches best the original classes. The detail of the method is described in Appendix 8.2. It is also possible to deal with the threshold decision by simply fixing an acceptable false positive rate, or through unsupervised methods for example [33].

Once we get the threshold denoted by  $\mathcal{L}_{thr}$ , deciding whether a coming data  $\mathbf{x}$  is outlier or not is straightforward:

$$\mathbf{x} \text{ is } \begin{cases} \text{inlier,} & \text{if } \mathcal{L}_{\mathbf{x}} \geq \mathcal{L}_{thr} \\ \text{outlier,} & \text{otherwise} \end{cases} \quad (1)$$

## 2.2 On Application to Glycemic Variability Characterization

Applying the proposed framework to daily GV characterization, the reference MMST describes the population of stable daily GV (inliers) while the unstable daily GV are taken as anomalies (outliers). MMST allows us to fuse a selected set of daily computed classic metrics, which means that the number of selected metrics  $M$  is the dimension of  $\mathbf{x}_n \in \mathbb{R}^M$ , and  $n \in \{1, N\}$  represents the day index. To learn the reference MMST, we take the values of daily computed metrics from stable subjects as the input reference learning sample, while for finding the threshold, values of metrics from unstable subjects are used as the outlier sample.

### 2.2.1 Pre-processing of Glycemic Recording

To obtain the daily computed metrics, we need first to pre-process the glycemic recording from CGM in case of the presence of missing or disordered data. The pre-processing of CGM data goes as follows.

1. Start and end counting at the start and end of a natural day (start at first integer decimal time (00:00h) and ends at the last);
2. Suppress from analysis the day with long-term missing data ( $> 4.8$  h);
3. Linearly resample with the rate equivalent to the reading of the CGM to assure unit sampling interval.

### 2.2.2 Glycemic Variability Feature Selection

7 glycemic control metrics concerning variability: CV, J-index, M-value, ADRR, CONGA,  $\text{MAGE}_{\text{avg}}$ , GVP, are selected to be combined into  $\mathbf{x}_n$ , since they are daily computable and not highly correlated among the large amount of metrics mentioned in the introduction<sup>1</sup>. The

---

<sup>1</sup> J-index and M-value are conceptually admitted as control metrics, while they are all related to GV performances.

computation of these metrics is detailed in Appendix 8.3. To avoid that a certain metric dominates the distribution, the metrics are all standardized.

MMST can incorporate any set of metrics, but a high model dimension requires larger amount of learning data and leads to more computation. Considering the existence of correlation between selected metrics, dimension reduction methods can be used before fitting the model. We apply Principal Component Analysis (PCA) [34] to project data onto new orthogonal axis which are linear combinations of the metrics, and keep only the principal dimensions. Linear Discriminant Analysis (LDA) or kernel based nonlinear discriminant analysis can be alternative solutions [35] [36].

### 3 Performance of the Proposed Method

The proposed method is applied to GV characterization using data from the TRIMECO trial [26]. This section deals with the learning sample construction from TRIMECO data, shows the performance of the new daily GV index, evaluates the performance and tests the robustness of the proposed method.

#### 3.1 Model Learning from TRIMECO Trial

TRIMECO trial is a study on islet transplantation performance on 50 subjects with type 1 diabetes, selected randomly from 15 university hospitals in France [26]. It provides the CGM data – of each subject – before, 6 month after, and 12 month after first injection of pancreas islet (infusion number can vary from 1 to 3). The CGMs receive glucose readings every 5 minutes. The TRIMECO data set presents 4.62% of missing glycemic record. After the pre-processing, the CGM data are distributed into 3 cohorts according to the number of months after transplantation named “M0”, “M6” and “M12” respectively. The middle column of Table 1 lists their basic information.

Among the 50 subjects, glycemic records of 3 subjects are not available. “Valid subject number” shows how many subjects of 47 in total have CGM data of at least one complete day after pre-processing described in Section 2.2.1, while “complete day number” shows the sum-up number of days which have CGM data after pre-processing in each cohort.

The entire data is firstly normalized, then its dimension reduced with PCA. 2 principal components are kept out of 7 after PCA for an accumulation of almost 89% data variance. Thus, the data dimension is reduced to  $M = 2$ . The projected data of the 3 TRIMECO cohorts are illustrated in Figure 2 (for a balanced view, a few data points are dropped in the outlying region,  $[\text{Dim1} > 5.0, \text{Dim2} > 2.5]$  are not displayed). The data in cohort M0 before transplantation is further away from the lower-left (low variability) compared to cohorts M6 and M12.

##### 3.1.1 $\beta$ -score and Choice of Reference Sample

$\beta$ -score is a clinical score for measuring the islet graft functions which is composed by adding points evaluating normal fasting glucose (2 points),  $\text{HbA}_{1c}$  (2 points), stimulated C-peptide (2 points), and absence of insulin or oral hypoglycemic agent use (2 points) [27]. Thus, it is taken for measuring the performance of islet transplantation: A  $\beta$ -score  $\geq 6$  means a significant improvement of mean and standard deviation of glucose and hyperglycemia, which can be considered as the success of the transplantation [37]. Figure 3 reports the

histograms of the cases<sup>2</sup> with different  $\beta$ -score values and verifies the improvement brought by transplantation as before the transplantation, no subject has its  $\beta$ -score above 3.

We select the CGM subjects with  $\beta$ -score  $\geq 6$  after the first islet injection, and take their daily metrics as the reference learning sample, assuming that a successful transplantation well adjusts the GV to a normal range. The projected reference learning sample on principal components is marked by yellow triangles named “Reference” in Figure 2. We observe that they are more concentrated in low variability region. Reference MMST is then built on the learning sample, with its parameter estimated by EM algorithm and the number of component  $K$  chosen from the range  $1 \leq K \leq 10$  by BIC criterion. The contour of the fitted reference MMST model well covers the reference sample as plotted over the scattered data in Figure 2 (2 components, with the outer most contour density: 0.01). For an input daily glycemia, the log-likelihood which measures its probability to be an inlier of the reference MMST is therefore the index which integrates the characterization power of the 7 included classic metrics.

To decide the log-likelihood threshold for classifying the daily GV, anomalies are selected from the subjects with  $\beta$ -score = 0, as these subjects diagnosed with the poorest beta-cell functions have rarely their GV under control. The cohort information about the reference and samples with  $\beta$ -score = 0 are listed in the right column of Table 1. The histograms of the log-likelihood values of learning samples are displayed in the top sub-figure of Figure 4 where a dashed bar in light green shows the threshold decision given by the method described in Section 2.1.2. In this study, learning the reference model as well as finding the threshold takes 11.44s on a PC with 3.6GHz CPU. However, once the reference model is learnt, to infer the likelihood of daily GV for a 10-day scenario requires only 0.001s.

Let us remark here that both the reference and anomaly learning samples selected by  $\beta$ -score level are not perfect. Stable GV can be found, although rarely in subject with  $\beta$ -score = 0, while unstable days could happen for subject with  $\beta$ -score > 0. However, the method allows an acceptable degree of imperfection in learning samples. As shown in Figure 4, some reference learning points are finally classified below the threshold, while some anomalies are above the threshold. The three lower sub-figures of Figure 4 display the histograms of the log-likelihood values of all daily glycemia in 3 TRIMECO cohorts separately. The classification using the given threshold has a false positive rate = 2.68%. As expected, the log-likelihood values concentrate above the threshold more often after the transplantation than before.

### 3.1.2 Glycemic Variability Evaluation Using the New GV index (log-likelihood)

The new GV index (log-likelihood) allows both daily and long-term integral GV evaluation for the interest of individual subject. Provided a series of daily log-likelihoods denoted by  $\mathcal{L}_X = \{\mathcal{L}_{x_1}, \dots, \mathcal{L}_{x_N}\}$  from a subject, we can investigate:

1.  $\mathcal{L}_{x_n}$ : the daily log-likelihood, for daily GV evaluation;
2.  $\text{median}(\mathcal{L}_X)$ : the median daily log-likelihood of recorded period, for global evaluation of GV;
3. Stable day percentage  $\mathcal{P}_s$ : defined by  $100 \times \sum_1^N \mathbf{1}(\mathcal{L}_{x_n} \geq \mathcal{L}_{thr}) / N$  [%], the percentage of stable days in the time length of glycemic record.

Figure 5 shows an example of the GV characterization of an individual subject from TRIMECO. His glycemia trace is plotted in 5(a) with colors (“stable” in blue,

---

<sup>2</sup> Not to be confused with “subject”, a “case” refers to a subject under a certain cohort, as one subject often appears in all of the 3 cohorts.

“unstable” in red) corresponding to the automatic classification through the index threshold. The patient shows an unstable daily GV performance (as the red and blue alternate) before the transplantation indicated by the classification. 6 month after first islet injection, his  $\beta$ -score is adjusted from 1 to 7, and 12 month after, the value reaches 8. The classification indicates at the same time, his stability of glycemia is improved, from a stable day percentage  $\mathcal{P}_s$  (labeled by “S” in titles of subplots) of 52.38% to 100.00%. Besides, histogram of the values of log-likelihood is another way to view his daily GV statistically as displayed in 5(b). The left long tail of the daily values disappears after the transplantation. The value of median( $\mathcal{L}_x$ ): (“median llh” in titles of subplots) is improved from -1.63 before transplantation to 0.61, 6 month after and -0.08, 12 month after the transplantation.

### 3.2 Performance of Daily Glycemic Variability Classification Compared to Physicians Diagnostic

To evaluate the performance of the proposed characterization, two diabetologists experienced with dealing with T1D patients with severe glycemic variability on routine bases are invited to classify the TRIMECO data daily into three classes (“stable”, “unstable” and “unsure”). Their classification is conducted independently without any information except the daily glycemia trace. The diabetologists performed the classification of glycemic profile while taking into account only GV without considering glycemia as a whole. A trace with multiple hypoglycemia episodes but low GV or a trace with persistent hyperglycemia and low GV were classified as a stable.

As three classes exist in labels given by physicians instead of two from our auto-classifier, in order to unify the class numbers as well as to remove controversial labeling from physicians, the data with two “unsure” labels or with disagreement (one label is “stable” and the other is “unstable”) given by the physicians are excluded from comparison. Since there is actually not much discrepancy (1.21%) in between the labels from the two physicians, the remaining two sets of labels are fused into a single “consensus of physician”, taking their agreement on stability<sup>3</sup>. The performance of the likelihood based classification can be thus evaluated by “accuracy”, “sensitivity” and “specificity” compared to the “consensus of physician” serving as ground truth.

The classification result of MMST is compared to One Class SVM (OCSVM) [38] and to five other single metric-based classification results. Specifically, the five compared metrics are: ADRR, CV, GVP which are included in the mixture (concerning the most, medium and least related to the principal component) of MMST; GRADE [39], which is rather a glycemic control metric measuring the risk of hyper or hypo glycemia; and COGI, a synthetic glycemic control metric including GV aspect. OCSVM is trained on the reference cohort with the hyper-parameters tuned using 10 fold cross-validation<sup>4</sup>. The cohort with  $\beta$ -score=0 serves as unstable samples in the testing process of the cross-validation. To be fair, the thresholds for the five compared metrics are decided through the same algorithm as for the MMST likelihood, rather than applying directly existing ones from certain clinical studies. Using “taken-in” thresholds can hardly result in good classification on specific data sets. For example, applying 36% [40] as threshold for CV results only in a 59.1% accuracy.

<sup>3</sup> The daily glycemia is agreed to be “stable”, if the assigned labels from the two physicians are either both “stable” or “stable” and “unsure”, so as for “unstable” case.

<sup>4</sup> The tuned hyper-parameters: kernel (RBF) coefficient  $\gamma = 7.45 \times 10^{-9}$  and error control  $\nu = 0.01$ , are obtained from the grid searching range  $[2^{-40}, 2^3]$  with step size  $2^1$  and  $[2^{-10}, 2^{-1}]$  with step size  $2^{0.1}$  respectively.



The performance of all methods is reported in Table 2. With appropriate threshold (hyperplane for OCSVM), the classifications of all metrics are fairly good. MMST outperforms its contribution metrics taken individually: ADRR, CV and GVP, while the three metrics have a performance proportional to their importance in building the MMST reference model (see Table 4 in Appendix for their combination coefficients in principal component). GRADE does not perform well against the others as a “non-GV” metric as expected. However, COGI and OCSVM actually obtain competitive results close to MMST. This lies in the fact that as a consensus, the log-likelihood of MMST reflects a global assessment, but for a GV classification task, MMST is not necessarily very outstanding compared to a metric with already good discriminative power. COGI gets good result with only 15% of its value comes from variability (SD). We interpret this as the evidence that corroborates to what we stated in the introduction – “from some aspects, GV assessment also relies on “control” factors”. The very high specificity of OCSVM can be explained by the tight hyperplane learnt at the frontier of the reference class, which leads to an accurate labeling of stable data. Further experiments on larger data sets would be necessary to compare the performance of MMST and OCSVM on this subject. However, we would like to highlight here that MMST identifies the density profile of the reference group, whereas the property identified by OCSVM is hard to explain physically.

### 3.3 Robustness of Proposed Method

To examine the robustness of the proposed method, “Leave One Case Out” (LOCO) test is conducted. Subjects in the reference cohort are removed in turn from the reference learning sample for model construction.

Table 3 reports the average results of some measurements based on the proposed daily index in each TRIMECO cohorts. The definition of  $\mathcal{P}_s$  and  $\text{median}(\mathcal{L}_x)$  were as introduced in Section 3.1.2. The “stable subject percentage” denotes the percentage of subjects with  $\text{median}(\mathcal{L}_x) \geq \mathcal{L}_{thr}$  in the cohort. The listed results corroborates the improvement of glycemic stability after the islet transplantation. The same measurements are calculated from the daily index given by each learnt reference model in LOCO experiment. Their Mean Absolute Deviation (MAD) from the original result are reported in brackets. The largest MAD 2.9% in Table 3 indicates that the proposed method is fairly robust to missing subject. The learning samples are representative enough for slight changes not to bias a lot the constructed model.

## 4 Discussion

Compared to previous work, the proposed method avoids a number of ad hoc commitments. Making use of the clinical  $\beta$ -score for sample construction avoids the effect of human subjectivity. The reference model constructed has meaningful statistic properties that are easy to interpret.

The proposed index also provides a way to investigate the relationship between  $\beta$ -score and GV statistically, which supplement the study in [37] representing GV by SD only. The TRIMECO cases are divided into 5 groups according to their corresponding  $\beta$ -score interval. The histograms of their log-likelihood values in the 5 different  $\beta$ -score intervals are displayed in Figure 6(a), with the mean value and standard deviation of the log-likelihood in each group summarized in Figure 6(b). It reveals that, a higher  $\beta$ -score is linked to a larger mean log-likelihood and a smaller log-likelihood deviation. In another word,  $\beta$ -score increases with the improvement of glycemic stability of both intra and inter-day.

Recipients of islet transplantation offer per se unstable glycemic profile which is the main indication for transplantation. After transplantation, two categories of patient are commonly described: recipients with no functional graft evidencing high glycemic variability and recipients with functional graft and improved glycemic variability. This population is consequently an ideal population to classify glycemic variability and test our proposed score.

Our proposed score is based on combined indices extracted from CGM data. It is easily exportable to all populations in which CGM are performed and in which raw data are available such as the general population with type 1 and type 2 diabetes.

## 5 Conclusion and Perspectives

A method for daily glycemic variability characterization is proposed in this work, which fuses the evaluation ability of existing metrics with a statistical model called “Mixture of Multiple-Scale T-distributions” (MMST). The proposed framework provides a new index for an integral daily glycemic variability (GV) measurement. Additionally, several measurements are investigated based on this new index, with their interests on variability characterization demonstrated by examples applied on data from TRIMECO trial. Moreover, using outlier detection technique, a threshold of the new index is decided for classifying any daily input glycemia into “stable” (reference population) or “unstable” (outliers). The evaluation shows that the new index can provide a classification highly concordant with the experience of diabetologists, outperforming its comprising metrics and several other synthetic metrics. The proposed index helps to verify the improvement of glycemic stability brought by pancreas islet transplantation in TRIMECO trial, and allows a better understanding of the link between the clinical  $\beta$ -score and the glycemic variability.

Future work includes the following directions:

1. **Methodology:** though theoretically, a mixture model is able to fit a given distribution with large number of components, differences exist in practice, caused by factors such as the precision of estimated parameters. The framework allows to replace MMST by other mixture models for an investigation of these differences. Besides, introducing Markovianity [41] or considering more complete statistical dependencies of the daily glycemia using a boosted conditional mixture model [42] may improve the performance. In addition, the threshold decision in this work is learnt by using unstable samples, while it would worth trying some unsupervised methods for example as introduced in [33].
2. **Application:** The learnt reference model and index threshold can be easily applied to any input daily glycemia, the test on other data set for a further evaluation of the method’s characterization ability is planned. Learning the association of the proposed index with other clinical index such as HbA1c is a potential work. Furthermore, the entire framework can be rerun on samples from other essays to obtain a more representative reference model, if a larger amount of randomized subjects are available or a better way to build learning samples is available. Rerunning the framework on samples under more specific conditions, such as with glycemia values within a certain interval or with glycemia during the night only is also a promising direction.

## 6 Declaration of Interest

The authors declare no competing interests.

## 7 Acknowledgements

We thank the members of the TRIMECO trial group listed in appendix 8.4. The TRIMECO trial was financed by a grant from the French government under a program named ‘‘Programme Hospitalier de Recherche Clinique’’.

## 8 Appendix

### 8.1 Mixture of Multiple-Scale T-distributions

A  $K$ -component Mixture of Multiple-Scale T distributions (MMST) can be written with the detail of the set of parameters in each Multiple-Scaled T (MST) component  $\mathcal{MS}(\mathbf{x} | \boldsymbol{\phi}_k)$  as:

$$p(\mathbf{x} | \boldsymbol{\varphi}) = \sum_{k=1}^K \pi_k \mathcal{MS}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \mathbf{v}_k) \quad (2)$$

where  $\boldsymbol{\phi}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \mathbf{v}_k\}$ ,  $\pi_k$  denotes the mixing coefficient of component  $k$ , and the entire parameter set is  $\boldsymbol{\varphi} = \{\pi_k, \boldsymbol{\phi}_k\}_{k=1}^K$ . Each component follows a MST defined in the form of an infinite mixture of scaled Gaussians that is

$$\mathcal{MS}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{v}) = \int_0^\infty \cdots \int_0^\infty \mathcal{N}_M(\mathbf{x} | \boldsymbol{\mu}, \mathbf{D}\Delta_{\boldsymbol{\omega}}\mathbf{A}\mathbf{D}^T) \prod_{m=1}^M \mathcal{G}\left(\omega \mid \frac{v_m}{2}, \frac{v_m}{2}\right) d\omega_1 \cdots d\omega_m \quad (3)$$

where  $\mathcal{N}_M(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the  $M$ -dimension Gaussian distribution, with mean  $\boldsymbol{\mu}$  and scale matrix  $\boldsymbol{\Sigma} = \mathbf{D}\Delta_{\boldsymbol{\omega}}\mathbf{A}\mathbf{D}^T$ .  $\boldsymbol{\Sigma}$  can be taken as a Hermitian matrix<sup>5</sup>  $\boldsymbol{\Sigma}' = \mathbf{D}\mathbf{A}\mathbf{D}^T$  weighted by a  $M \times M$  diagonal matrix  $\Delta_{\boldsymbol{\omega}} = \text{diag}(\omega_1^{-1}, \dots, \omega_M^{-1})$ , in which the diagonal components are the inverse weights. Each weight  $\omega_m$  follows a Gamma distribution  $\mathcal{G}(\cdot | v_m/2, v_m/2)$  and  $\mathbf{v} = \{v_m\}_{m=1}^M$ . In practice, (3) is equivalently written as:

$$\mathcal{MS}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{v}) = \prod_{m=1}^M \frac{\Gamma((v_m + 1)/2)}{\Gamma(v_m/2)(A_m v_m \pi)^{1/2}} \left\{ 1 + \frac{[\mathbf{D}^T(\mathbf{x} - \boldsymbol{\mu})]_m^2}{A_m v_m} \right\}^{-(v_m+1)/2}, \quad (4)$$

where  $[\mathbf{D}^T(\mathbf{x} - \boldsymbol{\mu})]_m$  denotes the  $m$ th component of variable vector  $\mathbf{D}^T(\mathbf{x} - \boldsymbol{\mu})$ , and  $A_m$  represents the  $m$ th diagonal element of  $\mathbf{A}$ .

#### 8.1.1 Parameter Estimation through EM algorithm

To estimate the parameters of MMST from observations  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$  given component number  $K$ , under the condition that the component indicators  $\mathbf{y} = (y_1, \dots, y_N)$  are missing<sup>6</sup>, the Expectation-Maximization (EM) algorithm tries to find the maximum of the log-likelihood  $\log(p(\mathbf{X} | \boldsymbol{\varphi}))$  iteratively. The algorithm runs as follows:

##### *E-step*

For iteration  $r + 1$ , compute the posterior probability  $\tau_{nk}^{(r)}$  of  $y_n$  using the estimated parameters  $\boldsymbol{\varphi}^r$  from iteration  $r$ :

<sup>5</sup>  $\mathbf{A}$  is the diagonal matrix with eigenvalues while  $\mathbf{D}$  is the matrix of eigenvectors.

<sup>6</sup> Each  $y_n, n \in \{1, N\}$  indicates the component of the corresponding  $\mathbf{x}_n$ .

$$\tau_{nk}^{(r)} = \frac{\pi_k^{(r)} \mathcal{MS}(\mathbf{x}_n | \boldsymbol{\phi}_k^{(r)})}{\sum_{l=1}^k \pi_l^{(r)} \mathcal{MS}(\mathbf{x}_n | \boldsymbol{\phi}_l^{(r)})} \quad (5)$$

then update the expectation of the complete log-likelihood function with respect to  $\mathbf{X}$  and the current  $\{\tau_{nk}^{(r)}\}_{n=1, k=1}^{N, K}$ :

$$Q(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(r)}) = \mathbb{E}_{\mathbf{y} | \mathbf{x}, \boldsymbol{\varphi}^{(r)}} [\log(p(\mathbf{X}, \mathbf{y} | \boldsymbol{\varphi}))] = \sum_{n=1}^N \sum_{k=1}^K \tau_{nk}^{(r)} \log(\pi_k^{(r)} \mathcal{MS}(\mathbf{x}_n | \boldsymbol{\phi}_k^{(r)})). \quad (6)$$

### *M-step*

Maximize (6) to update the parameters

$$\boldsymbol{\varphi}^{(r+1)} = \operatorname{argmax}_{\boldsymbol{\varphi}} Q(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(r)}). \quad (7)$$

For obtaining the updates, the common used way is to take the derivatives of  $Q(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(r)})$  with respect to each parameter  $\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \mathbf{v}_k, k \in \{1, \dots, K\}$  and set them to 0.  $\boldsymbol{\varphi}^{(1)}$  is initialized empirically by clustering the observations using K-means. The algorithm stops when  $p(\mathbf{X} | \boldsymbol{\varphi}^{(r+1)}) - p(\mathbf{X} | \boldsymbol{\varphi}^{(r)}) < \varepsilon$ , with  $\varepsilon$  a preset threshold. Then the parameter  $\boldsymbol{\varphi}^{(r+1)}$  of the last iteration is taken as the final estimated one denoted as  $\hat{\boldsymbol{\varphi}}$ .

Under the general case that the component number  $K$  is unknown, the applied technique is to try different values and select the one by some criterion. In this work, we take BIC to select  $K$  from a range  $1 \leq K \leq 10$ . Other criterion choices could be Akaike Information Criterion (AIC) [43], Integrated Complete Data Likelihood (ICL) [44], etc.

## 8.2 Threshold Decision for Outlier Detection

Given the reference model, inlier samples  $\mathbf{X}$ , and outlier samples denoted by  $\bar{\mathbf{X}} = \{\bar{\mathbf{x}}_j\}_{j=1}^J$ , the proposed threshold decision consists of the following steps.

1. Compute the log-likelihoods of all samples, for inliers  $\mathcal{L}_{\mathbf{X}} = \{\mathcal{L}_{\mathbf{x}_n}\}_{n=1}^N$  and outliers  $\mathcal{L}_{\bar{\mathbf{X}}} = \{\mathcal{L}_{\bar{\mathbf{x}}_j}\}_{j=1}^J$ .
2. Fitting 2 MMSTs separately on  $\mathcal{L}_{\mathbf{X}}$  and  $\mathcal{L}_{\bar{\mathbf{X}}}$  through EM algorithm as described in 8.1.1, and denote their parameter sets by  $\boldsymbol{\varphi}_{\mathcal{L}}$  and  $\boldsymbol{\varphi}_{\bar{\mathcal{L}}}$  respectively.
3. Denoting the entire log-likelihood set by  $\mathcal{L}_{\bar{\mathbf{X}}} = \{\mathcal{L}_{\bar{\mathbf{x}}_l}\}_{l=1}^{N+J} = \mathcal{L}_{\mathbf{X}} \cup \mathcal{L}_{\bar{\mathbf{X}}}$ , and search numerically the optimal log-likelihood value  $\mathcal{L}_{thr} \in [\min(\mathcal{L}_{\bar{\mathbf{X}}}), \max(\mathcal{L}_{\bar{\mathbf{X}}})]$  as threshold that satisfies:

$$\mathcal{L}_{thr} = \operatorname{argmax}_{\mathcal{L}_{thr}} \sum_{l=1}^{N+J} \mathbf{1} [(\mathcal{L}_{\bar{\mathbf{x}}_l} - \mathcal{L}_{thr}) (p(\mathcal{L}_{\bar{\mathbf{x}}_l} | \boldsymbol{\varphi}_{\mathcal{L}}) - p(\mathcal{L}_{\bar{\mathbf{x}}_l} | \boldsymbol{\varphi}_{\bar{\mathcal{L}}})) \geq 0]. \quad (8)$$

In practice, the searching range can be limited to a smaller one such as the inter-quartile.

Figure 7 gives a visual demonstration of how the decided threshold separates the log-likelihoods of inliers and outliers.

### 8.3 Applied Glycemic Variability Features

As no confusion will be introduced when describing the construction of a general daily data, the day index  $n$  will be temporarily omitted. We denote the glycemic recording in a day by  $\mathbf{G} = [G_1, \dots, G_T]$ ,  $T$  represents the data length<sup>7</sup>. If no specification,  $\mathbf{G}$  is measured by the unit “mmol/L”. The mean of daily glycemia is  $u_G = \text{mean}(\mathbf{G}) = \frac{1}{T} \sum_{t=1}^T G_t$ , and the daily standard deviation is  $\sigma_G = \text{sd}(\mathbf{G}) = \sqrt{\sum_{t=1}^T (G_t - u_G)^2 / (T - 1)}$ . The computation of the 7 chosen metrics which construct the 7 dimensions of  $\mathbf{x} = [x_1, \dots, x_7]$  are described as follows:

- CV:  $x_1 = \sigma_G / u_G$ .
- J-index:  $x_2 = 0.324 \times (u_G + \sigma_G)^2$ .
- M-value<sup>8</sup> [mg/dL]: defining  $\text{IGV} = 6.67$  mmol/L which represents the ideal glucose value and  $\text{range}(\mathbf{G}) = \max(\mathbf{G}) - \min(\mathbf{G})$ ,  
 $x_3 = \text{mean}((10 \log_{10}(\mathbf{G}/\text{IGV}))^3) + \text{range}(\mathbf{G}) \times 18/20$ .
- ADRR: defining  $f(G_t) = 1.509[\log(18 \times G_t)^{1.084} - 5.381]$  and  $r(G_t) = 10f(G_t)^2$ ,  
 $x_4 = \max(r(G_t) | f(G_t) > 0) + \max(r(G_t) | f(G_t) < 0)$ ;  $1 \leq t \leq T$ .
- CONGA:  $x_5 = \text{sd}(\mathbf{G}_{dif})$  with  $\mathbf{G}_{dif} = \{G_{13} - G_1, \dots, G_T - G_{T-12}\}$  denoting the hourly difference.
- $\text{MAGE}_{\text{avg}}$ :  $x_6 = \text{mean}(\boldsymbol{\lambda})$ , with  $\boldsymbol{\lambda}$  representing the meaningful glycemic excursions that larger than  $\sigma_G$ .
- GVP [mg/dL]: defining  $\Delta G_t = 18 \times (G_{t+1} - G_t)$  and  $\Delta t$ , the time interval [min] of CGM samples,

$$x_7 = \left\{ \sum_{t=1}^{T-1} \sqrt{\Delta t^2 + \Delta G_t^2} / \sum_{t=1}^{T-1} \Delta t \right\} - 1.$$

Table 4 lists the combination coefficients of the first two principal components obtained from PCA.

### 8.4 TRIMECO Trial Investigators

#### Main centers:

- Grenoble University Hospital: Pierre Yves Benhamou, Sandrine Lablanche, Rachel Tétaz, Paolo Malvezzi.
- Besançon University Hospital: Sophie Borot, Alfred Penfornis.
- Clermont-Ferrand University Hospital: Igor Tauveron, Béatrice Roche.
- Lille University Hospital: François Pattou, Marie-Christine Vantyghem, Kanza Benomar, Christian Noel.
- Lyon University Hospital: Charles Thivolet, Emmanuel Morelon, Lionel Badet, Fanny Buron.
- Montpellier University Hospital: Anne Wojtusciszyn, Eric Renard.
- Nancy University Hospital: Luc Frimat, Sophie Girerd, Bruno Guerci.

<sup>7</sup> For recording with a 5-minutes interval CGM,  $T = 288$ .

<sup>8</sup> M-value is defined by the unit “mg/dL”.

- Strasbourg University Hospital: Laurence Kessler, François Moreau, Philippe Baltzinger, Thibault Bahoune.

#### **Islet production centers:**

- Geneva: Domenico Bosco, Nadine Pernin, Thierry Berney.
- Grenoble: Harald Egelhofer, Anaick Moisan, Virginie Persoons.
- Lille: Julie Kerr-Conte, Valery Gmyr, Rimed Ezzouaoui.

#### **Interventional radiologists and surgeons:**

- Frédéric Thony, Yvan Bricault, Mathieu Rodière, Christian Sengel (Grenoble).
- Michel Greget, Iulian Enescu (Strasbourg).
- Christian Noel, Marc Hazzan, Robert Caiazzo, Fanelly Torres, Kristell Le Mapihan, Violetta Raverdy (Lille).
- Marie-Ange Pierredon (Montpellier).
- Pierre-Jean Valette, Arnaud Muller, Jean Champagnac (Lyon).

## **References**

- [1] S. E. Siegelaar, F. Holleman, J. B. Hoekstra and J. H. DeVries, "Glucose variability; does it matter ?," *Endocrine reviews*, vol. 31, no. 2, pp. 171-182, 2010.
- [2] S. Ayano-Takahara, K. Ikeda, S. Fujimoto, A. Hamasaki, S. Harashima, K. Toyoda, Y. Fujita, K. Nagashima, D. Tanaka and N. Inagaki, "Glycemic variability is associated with quality of life and treatment satisfaction in patients with type 1 diabetes," *Diabetes care*, vol. 38, no. 1, pp. 1-2, 2015.
- [3] N. Ali, J. O'Brien Jr, K. Dungan, G. Phillips, C. Marsh, S. Lemeshow, J. Connors, F. Alfred and J. Preiser, "Glucose variability and mortality in patients with sepsis," *Critical care medicine*, vol. 36, no. 8, p. 2316, 2008.
- [4] J. S. Krinsley, "Glycemic variability and mortality in critically 111 patients: The impact of diabetes," *Journal of diabetes science and technology*, pp. 1292-1301, 2009.
- [5] J. Schaal, N. Libert, S. De Rudnicki, Y. Auroy and S. Mérat, "La variabilité glycémique en réanimation. In Annales françaises d'anesthésie et de réanimation," *Elsevier masson*, vol. 31, no. 12, pp. 956-960, 2012.
- [6] D. Rodbard, "Metrics to Evaluate Quality of Glycemic Control: Comparison of Time in Target, Hypoglycemic, and Hyperglycemic Ranges with "Risk Indices"," *Diabetes technology & therapeutics*, vol. 20, no. 5, pp. 325-334, 2018.
- [7] D. Rodbard, "Glucose variability: a review of clinical applications and research developments," *Diabetes technology & therapeutics*, vol. 20, no. S2, pp. S2-5, 2018.

- [8] G. Molnar, J. Rosevear, E. Ackerman, L. Gatewood and W. Taylor, "Mean amplitude of glycemic excursions, a measure of diabetic instability," *Diabetes*, vol. 19, no. 9, pp. 644-655, 1970.
- [9] C. Weller, M. Linder, A. Macaulay, A. Ferrari and G. Kessler, "Continuous in vivo determination of blood glucose in human subjects," *Annals of the New York academy of sciences*, vol. 87, no. 2, pp. 658-668, 1960.
- [10] B. Kovatchev, D. Cox, L. Gonder-Frederick and W. Clarke, "Symmetrization of the blood glucose measurement scale and its applications," *Diabetes care*, vol. 20, no. 11, pp. 1655-1658, 1997.
- [11] B. Kovatchev, E. Otto, D. Cox, L. Gonder-Frederick and W. Clarke, "Evaluation of a new measure of blood glucose variability in diabetes," *Diabetes care*, vol. 29, no. 11, pp. 2433-2438, 2006.
- [12] C. McDonnell, S. Donath, S. Vidmar, G. Werther and F. Cameron, "A novel approach to continuous glucose analysis utilizing glycemic variation," *Diabetes technology & therapeutics*, vol. 7, no. 2, pp. 253-263, 2005.
- [13] T. Peyser, A. Balo, B. Buckingham, I. Hirsch and A. Garcia, "Glycemic variability percentage: a novel method for assessing glycemic variability from continuous glucose monitor data," *Diabetes technology & therapeutics*, vol. 20, no. 1, pp. 6-16, 2018.
- [14] J. Hermanides, T. Vriesendorp, R. Bosman, D. Zandstra, J. Hoekstra and J. DeVries, "Glucose variability is associated with intensive care unit mortality," *Critical care medicine*, vol. 38, no. 3, pp. 838-842, 2010.
- [15] E. Ryan, T. Shandro, K. Green, B. Paty, P. Senior, D. Bigam, A. Shapiro and M. Vantyghem, "Assessment of the severity of hypoglycemia and glycemic lability in type 1 diabetic subjects undergoing islet transplantation," *Diabetes*, vol. 53, no. 4, pp. 955-962, 2004.
- [16] J. M. Wojcicki, "'J'-index. A new proposition of the assessment of current glucose control in diabetic patients," *Hormone and metabolic research*, vol. 27, no. 01, pp. 41-42, 1995.
- [17] J. Schlichtkrull, O. Munck and M. Jersild, "The M<sub>g</sub> value, an index of blood<sub>g</sub>sugar control in diabetics," *Acta medica scandinavica*, vol. 177, no. 1, pp. 95-102, 1965.
- [18] L. Leelarathna, H. Thabit, M. Wilinska, L. Bally, J. Mader, T. Pieber, C. Benesch, S. Arnolds, T. Johnson, L. Heinemann and N. Hermanns, "Evaluating Glucose Control With a Novel Composite Continuous Glucose Monitoring Index.," *Journal of diabetes science and technology*, p. 1932296819838525, 2019.
- [19] P. Augstein, P. Heinke, L. Vogt, R. Vogt, C. Rackow, K. Kohnert and E. Salzsieder, "Q-Score: development of a new metric for continuous glucose monitoring that enables

- stratification of antihyperglycaemic therapies," *BMC endocrine disorders*, vol. 15, no. 1, p. 22, 2015.
- [20] I. Hirsch, A. Balo, K. Sayer, A. Garcia, B. Buckingham and T. Peyser, "A simple composite metric for the assessment of glycemic status from continuous glucose monitoring data: implications for clinical practice and the artificial pancreas," *Diabetes technology & therapeutics*, vol. 19, no. S3, pp. S-38, 2017.
- [21] R. Vigersky, J. Shin, B. Jiang, T. Siegmund, C. McMahon and A. Thomas, "The comprehensive glucose pentagon: a glucose-centric composite metric for assessing glycemic control in persons with diabetes," *Journal of diabetes science and technology*, vol. 12, no. 1, pp. 114-123, 2018.
- [22] A. Thomas, M. Schönauer, F. Achermann, O. Schnell, M. Hanefeld, H. Ziegelsch, J. Mastrototaro and L. Heinemann, "The "glucose pentagon": assessing glycemic control of patients with diabetes mellitus by a model integrating different parameters from glucose profiles," *Diabetes technology & therapeutics*, vol. 11, no. 6, pp. 399-409, 2009.
- [23] M. Wiley, R. Bunescu, C. Marling, J. Shubrook and F. Schwartz, "Automatic detection of excessive glycemic variability for diabetes management," *10th International Conference on Machine Learning and Applications and Workshops, IEEE*, vol. 2, pp. 148-154, 2011.
- [24] C. Marling, N. Struble, R. Bunescu, J. Shubrook and F. Schwartz, "A consensus perceived glycemic variability metric," *Journal of diabetes science and technology*, vol. 7, no. 4, pp. 871-879, 2013.
- [25] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443-1471, 2001.
- [26] S. Lablanche, M. Vantyghem, L. Kessler, A. Wojtusciszyn, S. Borot, C. Thivolet, S. Girerd, D. Bosco, J. Bosson, C. Colin and R. Tetaz, "Islet transplantation versus insulin therapy in patients with type 1 diabetes with severe hypoglycaemia or poorly controlled glycaemia after kidney transplantation (TRIMECO): a multicentre, randomised controlled trial," *The Lancet Diabetes & Endocrinology*, vol. 6, no. 7, pp. 527-537, 2018.
- [27] E. Ryan, B. Paty, P. Senior, J. Lakey, D. Bigam and A. Shapiro, " $\beta$ -Score: an assessment of  $\beta$ -cell function after islet transplantation," *Diabetes care*, vol. 28, no. 2, pp. 343-347, 2005.
- [28] R. Domingues, M. Filippone, P. Michiardi and J. Zouaoui, "A comparative evaluation of outlier detection algorithms: Experiments and analyses," *Pattern recognition*, vol. 74, pp. 406-421, 2018.
- [29] J. Chen, S. Sathe, C. Aggarwal and D. Turaga, "Outlier detection with autoencoder ensembles," *Proceedings of the 2017 SIAM international conference on data mining*, pp.



90-98, 2017.

- [30] F. Forbes and D. Wraith, "A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. Statistics and Computing," *Statistics and Computing*, vol. 24, no. 6, pp. 971-984, 2014.
- [31] A. Dempster, N. Laird and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society: series B (Methodological)*, vol. 39, no. 1, pp. 1-22, 1977.
- [32] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461-464, 1978.
- [33] A. Arnaud, F. Forbes, N. Coquery, N. Collomb, B. Lemasson and E. Barbier, "Fully automatic lesion localization and characterization: Application to brain tumors using multiparametric quantitative MRI data," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1678-1689, 2018.
- [34] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933.
- [35] G. McLachlan, *Discriminant analysis and statistical pattern recognition*, vol. 544, John Wiley & Sons, 2004.
- [36] W. Zeng, X. Li, X. Zhang and E. Cheng, "Kernel-based nonlinear discriminant analysis using minimum squared errors criterion for multiclass and undersampled problems," *Signal processing*, vol. 90, no. 8, pp. 2333-2343, 2010.
- [37] M. Vantyghem, V. Raverdy, A. Balavoine, F. Defrance, R. Caiazzo, L. Arnalsteen, V. Gmyr, M. Hazzan, C. Noël, J. Kerr-Conte and F. Pattou, "Continuous glucose monitoring after islet transplantation in type 1 diabetes: an excellent graft function ( $\beta$ -score greater than 7) Is required to abrogate hyperglycemia, whereas a minimal function is necessary to suppress severe hypoglycemia ( $\beta$ -score grea," *Journal of clinical endocrinology & metabolism*, vol. 97, no. 11, pp. E2078-E2083, 2012.
- [38] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, pp. 582-588, 2000.
- [39] N. Hill, P. Hindmarsh, R. Stevens, I. Stratton, J. Levy and D. Matthews, "A method for assessing quality of control from glucose profiles," *Diabetic medicine*, vol. 24, no. 7, pp. 753-758, 2007.
- [40] L. Monnier, C. Colette, A. Wojtusciszyn, S. Dejager, E. Renard, N. Molinari and D. Owens, "Toward defining the threshold between low and high glucose variability in diabetes," *Diabetes care*, vol. 40, no. 7, pp. 832-838, 2017.

- [41] O. Cappé, E. Moulines and T. Rydén, *Inference in hidden markov models* (Springer Series in Statistics), Springer-Verlag, 2005.
- [42] R. Mohammadi-Ghazi, Y. Marzouk and O. Büyüköztürk, "Conditional classifiers and boosted conditional Gaussian mixture model for novelty detection," *Pattern recognition*, vol. 81, pp. 601-614, 2018.
- [43] H. Akaike, "A new look at the statistical model identification," *Selected papers of Hirotugu Akaike*, pp. 215-222, 1974.
- [44] J. Baudry, "Estimation and model selection for model-based clustering with the conditional classification likelihood," *Electronic journal of statistics*, vol. 9, no. 1, pp. 1041-1077, 2015.