Evidence based continuous probability estimation for hypoglycemic excursion from sparse blood glucose data.

J. Wrede[a], R. Bankosegger[a], F. Debong[a], L. Schuster[a], D. Duke[b]

MOTIVATION: PATTERN DETECTION IN DIABETES THERAPY

The primary goal of diabetes management is to keep blood glucose in target range while minimizing time in hyperglycemia and hypoglycemia. To give the patient a focus on how to improve his or her therapy, mySugr developed a set of tools to analyze the patients data in order to derive appropriate therapy advice.

The widely adopted AGP plot[1] aggregates multiple days of continuously measured glucose traces by calculating a percentile plot which aides as a visual tool to identify periods with bad glucose control. For instance, the percentile plot in Fig. 1 reveals, a comparably high prevalence of hyperglycemic events in the late evening. In this work, we developed a tool that allows similar analysis on sparse SMBG data.

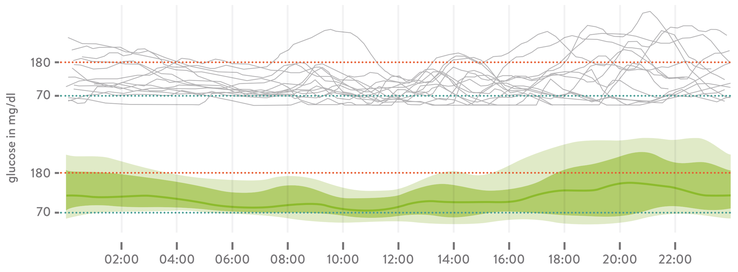

Figure 1. *CGM traces 14 consecutive days (top) and calculated percentiles showing glucose variability (bottom)*

PROBLEM STATEMENT

Traditional SMBG monitoring suffers from a low measurement frequency in comparison to CGM. In addition, the data is highly skewed as measurements mostly occur around specific times (i.e. before meals) or certain events (i.e. hypos) while other periods are covered much more scarcely.

Due to the low data density, analysis usually requires grouping of adjacent measurements into buckets of multiple hours (binning). As a consequence, the exact times of bad control are unknown or even hidden in case where the time periods with bad control spans over two adjacent buckets.
In Fig. 2 we can see that the patient has reported frequent hyperglycemic events between 3:00PM and 11:00 PM. However, due to the fixed buckets, only the period 6:00PM – 12:00 shows a significant ratio of hyperglycemic events.

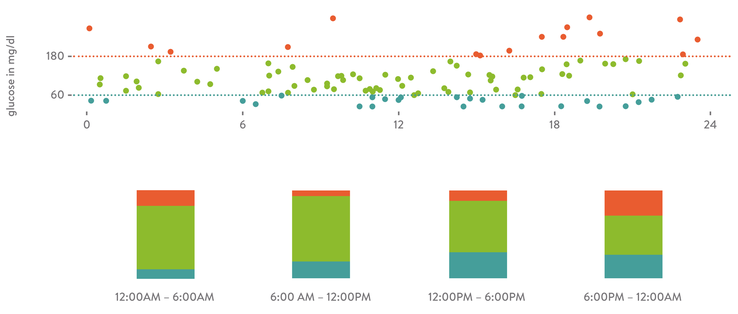

Figure 2. *SMBG data points (top) and calcuated fractions of glycemic events per period (bottom)*

METHOD DESCRIPTION (KERNEL DENSITY ESTIMATION)

To give the patient a better representation of the actual underlying problem, we apply a kernel density estimation (KDE)[2] to obtain a continuous probability distribution for glycemic events. For each SMBG measurement, a kernel is used to model the probability of the event being present prior and subsequent in time as shown in Figure 3. Summing over all kernels we obtain the likelihoods p(t|e) given a glycemic event e as shown in the top of Figure 4. Applying the Bayes theorem, we finally obtain the posterior probability distributions p(e|t).

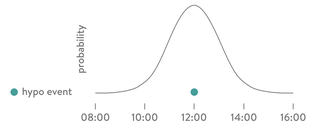

Figure 3. *Kernel for a hypo measured at 12:00*

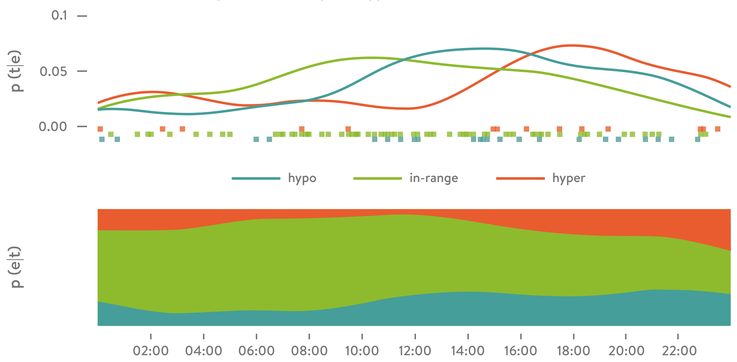

Figure 4. *Likelihood (top) and posterior probabilities of glycemic events (bottom)*

METHOD DESCRIPTION (KERNEL SELECTION)

The chosen kernel bandwidth reflects how long a glycemic event exists prior or subsequent to the event. Thus, choosing the right kernel becomes important especially for regions with very little data. The relationship between the glucose concentration and time spent in glycemic excursion has been previously analyzed[4]. The findings showed a parabolic dependency that we used to adopt the kernel-width respectively (Fig. 5).

A problem of performing KDE using Gaussian kernels is its limitation to non-periodic data leading to a non-continuity of the estimated probabilities. This can be seen when comparing the probabilities at 0:00 and 24:00 in Figure 4. Using a Van-Mises Kernel[3] instead gives a continuous probability across the boundaries. Figure 6 shows the circular fitted likelihood (left). Transforming to a linear time scale, the obtained posterior probabilities now show a continuous transition.

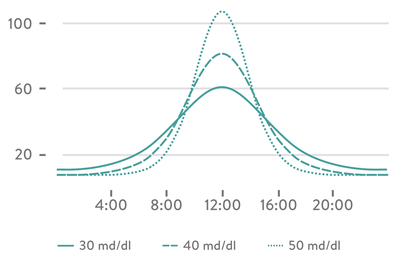

Figure 5. *Kernel width based on glucose concentration*

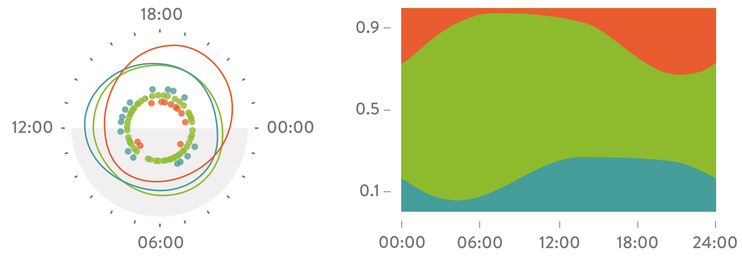

Figure 6. *Likelihood fitted using circular kernel (left) and transformed posterios probabilities (right)*

MODEL EVALUATION AND DATA CONSTRAINTS

The accuracy of our approach is limited by the amount of SMBG data. To better understand the relationship between the number of samples and estimation accuracy, we performed a evaluation on CGM datasets.

1. Estimate the posterior distributions using 14 days of CGM data
2. Uniform random subsampling of CGM to generate artificial SMBG datasets
3. Compare fitting on both datasets

Fig. 7 illustrates the error between both distributions with respect to different numbers of SMBG measurements per day. Increasing the frequency from 1 to 3 shows the greatest error reduction while a plateau is being reached at around 4 measurements/day. Since the number of users shrinks with increasing number of measurements per day, we chose 4 measurements per day over a duration of 14 days as the minimum requirement for our model.

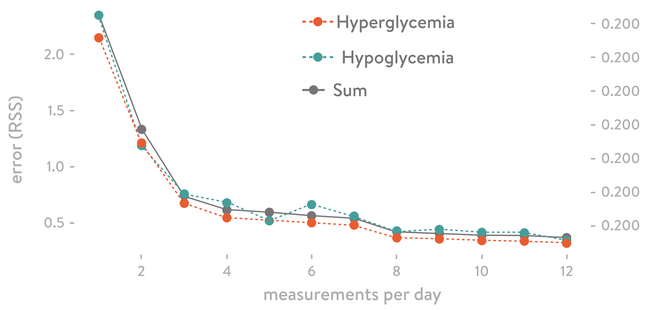

Figure 7. *Estimation error with respect to BGM sample frequency*

DISCUSSION

The proposed method allows a continuous modulation of glycemic excursion probabilities for PWD on SMBG. By applying a kernel density estimation method using circular kernels, we account for the periodicity of the data while integrating knowledge obtained from CGM patients allowed a proper selection of bandwidth parameters.

Figure 8 provides a comparison between the obtained distributions on both SMBG and CGM data. Due to the limitations in SMBG data analysis and the great success of visual tools such as the AGP, we think that our approach has great potential to support therapy analysis for both HCPs and PWD on SMBG.

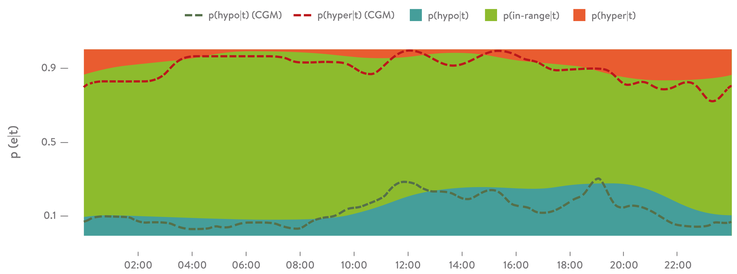

Figure 8. *Comparison of SMBG and CGM based probability distributions*

a mySugr GmbH, Vienna, Austria
b RDCI, Indianapolis, USA

1 Danne, T. et al. (2017) 'International consensus on use of continuous glucose monitoring', Diabetes Care, 40 **2** Scott, D. W. (2018). Kernel Density Estimation **3** Mardia, K., et al. (1999). Directional Statistics. Wiley **4** Duke, D. (2017) Therapy Adjustment Based on Glucose or Risk Probability Distribution. (Roche Indianapolis)