# Temporal patterns in insulin needs for Type 1 diabetes

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

Type 1 Diabetes (T1D) is a chronic condition where the body produces little or no insulin, a hormone required for the cells to use blood glucose (BG) for energy and to regulate BG levels in the body. Finding the right insulin dose and time remains a complex, challenging and as yet unsolved control task. In this study, we use the OpenAPS Data Commons dataset, which is an extensive open-source dataset collected in real-life conditions, to discover temporal patterns in insulin need that include well-known factors such as carbohydrates as well as novel factors too. We utilised various time series techniques to spot such patterns using matrix profile and multi-variate clustering. The better we understand T1D and the factors impacting insulin needs, the more we can contribute to building data-driven technology for T1D treatments.

## 1 Introduction

Patients with T1D need to manually inject insulin to keep alive. Despite advances in treatments for T1D, it is challenging to maintain glucose levels at an average which would keep patients safe from complications. A considerable part of the research focuses on automated insulin delivery systems (AID), also called artificial pancreas (APS) or closed loop systems (CL) [1]. These systems use an insulin pump to deliver the insulin and a continuous glucose measurements sensor (CGM) as a signal to be regulated [2]. Research into these systems focuses on the safety of the system, improving the algorithms, accurately predicting blood glucose and predicting system failures such as insulin infusion set and blood glucose sensor failures.

This study, on the other hand, is using data produced by AID to find temporal patterns in insulin needs that are not yet known. The data studied stems from an automatic insulin delivery system called OpenAPS [3]. OpenAPS automatically adjusts insulin based on blood glucose levels from a CGM sensor. This mechanism generates a dataset that is more complete and accurate than other datasets where people have to manually record their treatment (which is prone to error and missing values). However, this real data comes with its own challenges of irregularities, non-even sampling, and missing data, which makes it hard to be handled with various standard time series techniques.

In this paper, we implement intensive pre-processing on openAPS data in order to utilise time series techniques to find temporal patterns in insulin on board (IOB) and link these to other factors, namely carbohydrates (COB) and blood glucose (BG). The rest of this paper is organised as follows: section 2 gives a review of the literature, section 3 describes the method applied in this paper, and section 4 lists the results achieved by the different techniques. Finally, section 5 concludes the research and lists directions for future work.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

## 2 Related Work

Machine learning (ML) approaches are adopted to address the complex and non-linear dynamics of regulating BG using insulin. [1] gave examples of problems addressed by machine learning such as blood glucose prediction, overcoming shortcomings of deterministic automatic insulin delivery algorithms such as predicting set failure [4] or sensor failure [5] and incident prediction such as predicting hypoglycemia [6]. Another research area aims to create new medical knowledge which focuses on issues such as predicting diabetes onset [7], predicting changes in behaviour or BG and evaluating the efficacy of treatment. The researchers stress the importance of data-driven methods and the shift towards tailored management of therapies.

Existing research employs a wide range of ML methods, Support vector machines (SVM) are widely used algorithms [8]. Random forest was used for blood glucose prediction using a multivariate dataset to capture daily rhythms in glucose metabolism by introducing a time feature into the prediction [9]. Other researchers attempted to predict blood glucose using a combination of methods: genetic programming, random forests, k-nearest neighbours, and grammatical evolution [10]. Data used for the aforementioned studies are collected in clinical settings with a small sample of patients (around 5-30 people). Also, the focus is on prediction rather than finding temporal patterns.

OpenAPS have been extensively and independently studied [11, 12, 13, 14, 15, 16, 17]. While the OpenAPS dataset has been studied to investigate clinical outcomes and the safety of the system it has not yet been used as a source for learning more about T1D, or changing current treatment. One study, while still mainly looking at the blood glucose data, has gone deeper into the OpenAPS dataset looking at blood glucose variability outcomes with regards to demographics [18]. This paper, however, looks at temporal patterns in insulin need that later on could be further studied for correlation and causality. To our knowledge, this is the first attempt to study this dataset with the purpose of generating new knowledge about T1D by identifying temporal patterns of insulin need using time series methods beyond the well-known pattern that eating more carbohydrates needs more insulin.

## 3 Methods

The focus of this paper is to identify temporal patterns that would suggest something else but carbohydrates (COB) were driving the need for insulin. Furthermore, times, when blood glucose (BG) was raising or dropping, that suggest insulin on board perhaps should have been higher or lower and therefore hint at changes in insulin need for those times as well.

We utilise Matrix Profile (MP) [19] to find such patterns. Matrix Profile is an efficient algorithm for finding signatures in time series that requires a window size $m$. MP works by sliding a window of size $m$ across the whole time series (or across multiple time series for the multivariate version) and calculating the distance between this window and each of the subsequent overlaps excluding the window itself to avoid trivial matches. For each point in time, the minimal distance gets recorded resulting in a new time series of the minimal distance for each point in time of the time series. The resulting minimal distance time series can then be used to find motifs and discords in the original time series. A subsequence of length $m$ starting at the index of the lowest minimal distance is a motif, while a subsequence of length $m$ starting at the index of the biggest minimal distance is a discord in the original time series. Window size is usually obtained through domain knowledge and hugely impacts the motifs and discords found.

In order to find patterns, daily and weekly time series are grouped into similar clusters using time series k-means. The idea behind using k-means clustering is to group all the daily/weekly time series of an individual into similar days/weeks of IOB, COB and BG. The resulting clusters of days/weeks then were analysed for how they differ from each other and if some of the clusters show an unknown pattern in insulin need. We clustered the daily and weekly time series both by IOB, COB and BG alone but always visualising all three different time series for the same days/weeks. We also ran multivariate time-series k-means clustering of IOB, COB and BG. Barycenters are the global average of a collection of time series. They can be calculated using different methods. If the distance measure used to compare time series with each other is DTW [20] then it is common practice to use DBA [21] as barycenters.

It's important to note that the OpenAPS system consists of irregular time series. To be able to apply time series methods, the data is resampled into regular time series using sliding windows. A summary

of sliding windows is extracted from the original values for IOB, COB and BG. The frequency of the now regular statistics is used as input for various time series methods.

# 4 Results

First, a heatmap is plotted to visualise the raw data for three random patients across days. Figure 1 shows that BG rises once IOB and COB have dropped. It also shows an interesting pattern when BG rises in the early half of the night despite IOB rising too. An expected pattern of high BG at night followed no/delayed first meal. Out of 253 days, there are very few that have similar IOB. There is one big subgroup that's more similar to each other but still significantly different. Unsurprisingly, daily patterns are most influenced by the main meals which create three spikes. We can also recognise a long flat line, presumably the night, visible in all figures where the x-axis is an hour of the day. Heatmaps for monthly patterns are in the appendix.

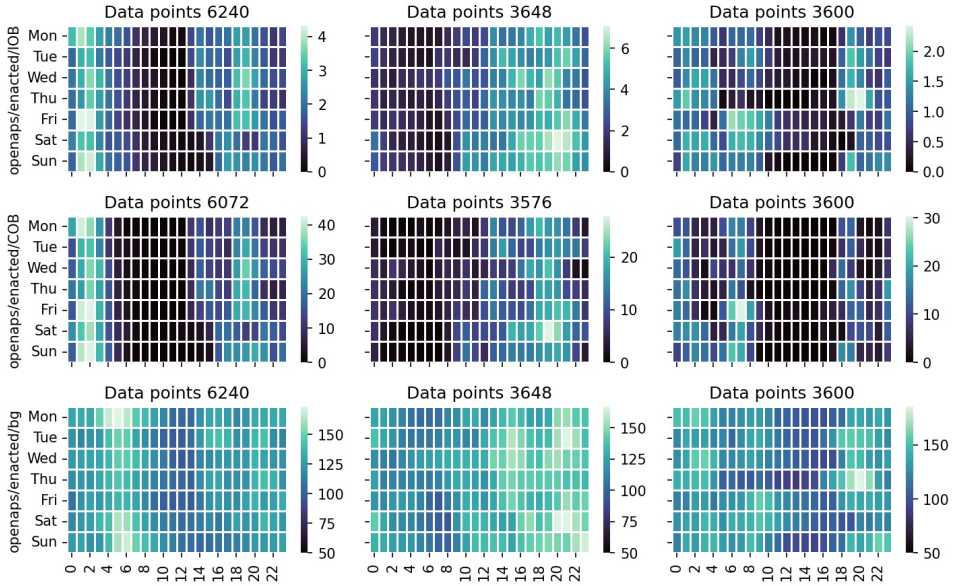
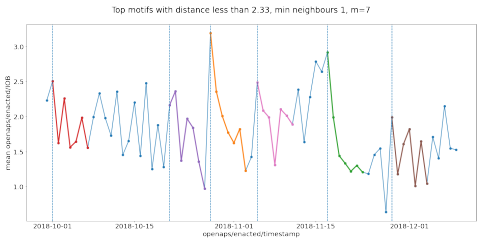


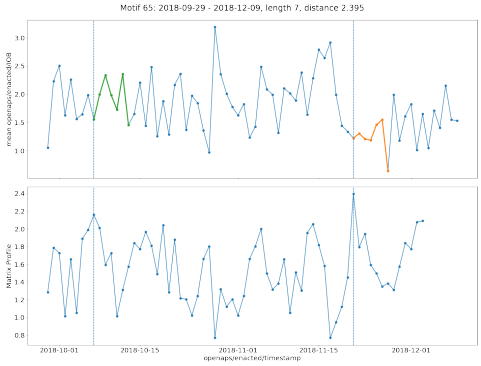

Figure 1: Heatmap of the various measures for P1 (column 1), P2 (column 2) and P3 (column 3) and IOB, COB and BG as rows. Y axis of each subplot plot is the days of the week and the x-axis is the hours of the day.

Matrix profile in OpenAPS is calculated for the "longest possible time series" with a mean value per day to find motifs and discords in the multivariate time series of IOB, COB and BG. The window size was set to $m = 7$, to indicate a period of seven consecutive days. The reason for choosing a week as motif length was to try to find weeks that have lower or higher insulin needs than others which could for example be caused by the female hormone cycle. Figure 2 shows the top motifs, top discord and matrix profile for IOB of a window of 7 days. The top motifs are the 7 consecutive days that are most like each other. The top discords are 7 consecutive days which are the least like each other. The IOB discord shows that a pattern of a stable week followed by a drop towards the end of the week is a discord. While the matrix profile was not that insightful in finding patterns it did highlight that seven relatively constant days of IOB were a discord and that motifs had rather drastic ups and downs of IOB from day to day, suggesting that treatment needs to be adjusted dynamically from day to day.

In order to group patterns, we cluster the daily time series of one patient using COB, GB, and IOB. The number of clusters is set to 4 based on Silhouette scores. The barycenters of the collection of timeseries in each cluster are shown in Figure 3. Overall clustering output shows the well-known pattern of IOB and COB going up for the breakfast, lunch and dinner meals in all clusters. We can see patterns where IOB does not follow COB or BG rises or drops without COB rising or dropping or IOB dropping or rising which perhaps suggests a higher/lower need for insulin during those times. It is notable that all clusters show that BG rises in the evening/night after IOB and COB have dropped. Most of the clusters show BG rising after IOB and COB have dropped. Cluster two shows a




(a) top motifs with distance less than 2.33




(b) top discord and the matrix profile

Figure 2: Matrix Profile top motifs, top discord and matrix profile for longest daily sampled IOB time series of P1.

higher IOB in the night hours ($\approx 4-7$ UTC) followed by no insulin for breakfast ($\approx 12$ UTC) and preinjection insulin for lunch. Cluster two also shows prolonged higher IOB after dinner into the night ($\approx 2-8$ UTC) preinjection by almost one hour for breakfast and lunch. On the other hand, cluster one shows BG rising in the morning before IOB and COB rise. Clusters 1, 3 and 4 show different prolonged high BG after lunch.

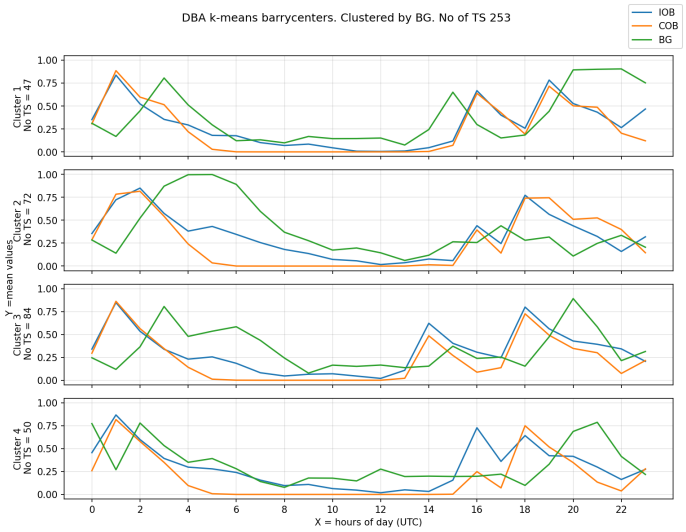



Figure 3: Resulting barycenters for IOB (blue), COB (orange) and BG (green) when clustered BG with k=4. K-means clustered daily time series for P1 using DTW as a distance metric.

## 5 Conclusion and Future Work

We found in this paper interesting temporal patterns in insulin need that cannot be explained by carbohydrates themselves using existing time series techniques. If and how these patterns go beyond what is currently known in medical science remains a question to be answered together with medical experts in T1D as well as the community of people with T1D using the OpenAPS system. While patterns found are very promising, they still lack scientific rigour and research into correlations and causalities that drive these patterns to truly be able to inspire new research into Type 1 Diabetes.

## References


[1] Ivan Contreras and Josep Vehi. Artificial intelligence for diabetes management and decision support: Literature review. *J Med Internet Res 2018;20(5):e10775 https://www.jmir.org/2018/5/e10775*, 20:e10775, 5 2018.

[2] Arthur Bertachi, Charrise M. Ramkissoon, Jorge Bondia, and Josep Vehí. Automated blood glucose control in type 1 diabetes: A review of progress and challenges. *Endocrinología, Diabetes y Nutrición (English ed.)*, 65:172–181, 3 2018. Go online for extras and with that table of APs systems.

[3] Openaps. `https://OpenAPS.org`. Accessed: 2022-09-30.

[4] Marzia Cescon, Daniel J. Desalvo, Trang T. Ly, David M. Maahs, Laurel H. Messer, Bruce A. Buckingham, Francis J. Doyle, and Eyal Dassau. Early detection of infusion set failure during insulin pump therapy in type 1 diabetes. *Journal of diabetes science and technology*, 10:1268–1276, 11 2016.

[5] Kamuran Turksoy, Anirban Roy, and Ali Cinar. Real-time model-based fault detection of continuous glucose sensor measurements. *IEEE transactions on bio-medical engineering*, 64:1437–1445, 7 2017.

[6] Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen. Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2016:3503–3506, 10 2016.

[7] Farrukh Aslam Khan, Khan Zeb, Mabrook Al-Rakhami, Abdelouahid Derhab, and Syed Ahmad Chan Bukhari. Detection and prediction of diabetes using data mining: A comprehensive review. *IEEE Access*, 9:43711–43735, 2021.

[8] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15:104–116, 1 2017.

[9] Eleni I. Georga, Vasilios C. Protopappas, Demosthenes Polyzos, and Dimitrios I. Fotiadis. A predictive model of subcutaneous glucose concentration in type 1 diabetes based on random forests. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 2889–2892, 2012.

[10] J. Ignacio Hidalgo, J. Manuel Colmenar, Gabriel Kronberger, Stephan M. Winkler, Oscar Garnica, and Juan Lanchares. Data based prediction of blood glucose concentrations using evolutionary methods. *Journal of Medical Systems*, 41:1–20, 9 2017.

[11] Alanna Weisman, Johnny Wei Bai, Marina Cardinez, Caroline K. Kramer, and Bruce A. Perkins. Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: a systematic review and meta-analysis of outpatient randomised controlled trials. *The Lancet Diabetes & Endocrinology*, 5:501–512, 7 2017.

[12] Shane O'Donnell, Dana Lewis, María Marchante Fernández, Mandy Wäldchen, Bryan Cleal, Timothy Skinner, Klemens Raile, Adrian Tappe, Tebbe Ubben, Ingrid Willaing, Bastian Hauck, Saskia Wolf, and Katarina Braune. Evidence on user-led innovation in diabetes technology (the open project): Protocol for a mixed methods study. *JMIR research protocols*, 8, 11 2019.

[13] Christine Knoll, Sofia Peacock, Mandy Wäldchen, Drew Cooper, Simran Kaur Aulakh, Klemens Raile, Sufyan Hussain, and Katarina Braune. Real-world evidence on clinical outcomes of people with type 1 diabetes using open-source and commercial automated insulin dosing systems: A systematic review. *Diabetic Medicine*, 39:e14741, 5 2022.

[14] Katarina Braune, Shane O'Donnell, Bryan Cleal, Dana Lewis, Adrian Tappe, Ingrid Willaing, Bastian Hauck, and Klemens Raile. Real-world use of do-it-yourself artificial pancreas systems in children and adolescents with type 1 diabetes: Online survey and analysis of self-reported clinical outcomes. *JMIR Mhealth Uhealth 2019;7(7):e14087 https://mhealth.jmir.org/2019/7/e14087*, 7:e14087, 7 2019.

[15] Lenka Petruzelkova, Jan Soupal, Veronika Plasova, Pavlina Jiranova, Vit Neuman, Lukas Plachy, Stepanka Pruhova, Zdenek Sumnik, and Barbora Obermannova. Excellent glycemic control maintained by open-source hybrid closed-loop androidaps during and after sustained physical activity. *Diabetes technology & therapeutics*, 20:744–750, 11 2018.

[16] DANA M. LEWIS, RICHARD S. SWAIN, and THOMAS W. DONNER. Improvements in a1c and time-in-range in diy closed-loop (openaps) users. *Diabetes*, 67, 7 2018.

[17] Nick Oliver, Monika Reddy, Claire Marriott, Tomas Walker, and Lutz Heinemann. Open source automated insulin delivery: addressing the challenge. *NPJ Digital Medicine*, 2, 12 2019.

[18] Arsalan Shahid and Dana M. Lewis. Large-scale data analysis for glucose variability outcomes with open-source automated insulin delivery systems. *Nutrients*, 14:1906, 5 2022.

[19] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh. Matrix profile i: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. *IEEE ICDM 2016*, 2016.

[20] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. *KDD Workshop*, 1994.

[21] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44:678–693, 3 2011.

# A Data preprocessing

At the time of this study, the OpenAPS Data Commons consisted of data for $n = 183$ individuals who have donated data. Due to many individuals having data in multiple time zones, it was decided to translate times to UTC to avoid jumps in time in the resulting time series data. This also provides better anonymisation and removes the perhaps tempting but incorrect assumption that people eat at a similar time of day and therefore time of day can be compared across individuals. Entries for which there was no timestamp were dropped. Interpolation from previous timestamps didn't seem sensible given this is system-written data, not user-entered data and therefore a gap in timestamps most likely means there was a problem with the system.

To ensure the data preprocessing did not significantly change the distribution of the data and introduce patterns that weren't originally in the data, the statistical properties of each of the different time series were calculated and visualised in box and violin plots and compared with each other. While variations between differently sampled time series were expected, changes in patterns would not be acceptable. For example, if IOB in January was higher than in February this had to be the case across all differently sampled time series. The mean, standard deviation, variance and numbers of outliers can vary but not the pattern itself. A certain granularity of time series might make a pattern more obvious than others or hide certain patterns which are acceptable too. The different granularity must not contradict each other.

The preprocessing described created various time series of different granularity. Given this project is looking into finding unknown patterns in insulin need it was decided to take all of the different time series forward and explore which granularity would prove to be most effective for each of the different techniques.

# B Heatmap for monthly timeseries

Figure 4 shows the mean values for each day of the week and month plotted in a heat map visualisation. The visualisation shows that for P1 (first column in the plot) the months 4-8 have a higher IOB without COB showing that same pattern. For P2 (second column) months 1 and two show a higher IOB without COB showing the same pattern. For P3 (third column) there's less of a pattern in IOB and COB but BG seems to be higher in months 1-6.

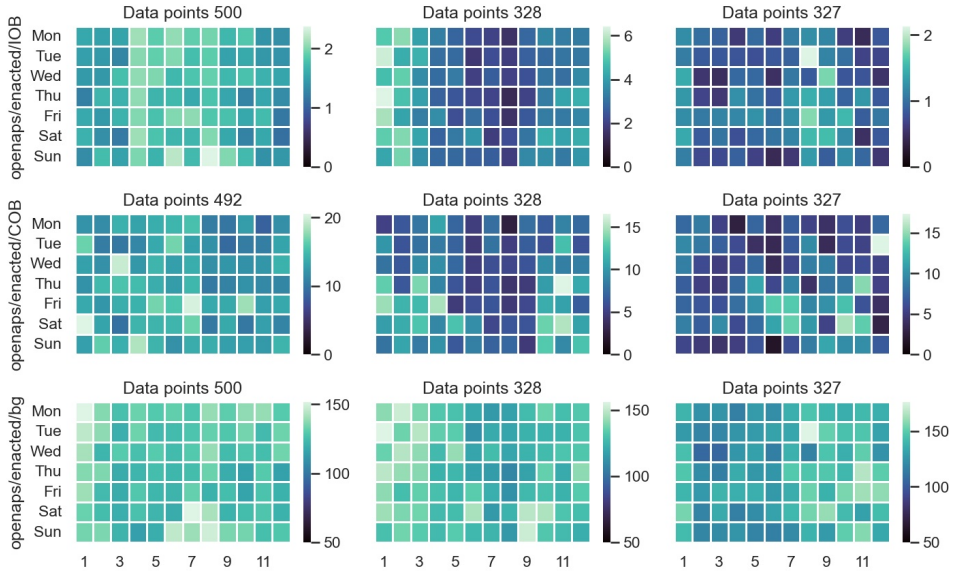



Figure 4: Heatmap of daily time series data for P1 (first column), P2 (second column) and P3(third column) and IOB, COB and BG as rows. Y axis of each subplot plot is the days of the week and the x-axis is the months of the year.