Analysis of Continuous Glucose Monitoring in relation to HbA1c and Interstitial Fluid Glucose

विद्या वाचस्पति की
उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध प्रबंध

A thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy

द्वारा / By
छात्र का नाम / Name of Student
सायन्तन मजुमदार/

Sayantan Majumdar

पंजीकरण सं. / Registration No.: 20183589

शोध प्रबंध पर्यवेक्षक / Thesis Supervisor
डॉ. प्रणय गोयल / Dr. Pranay Goel

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

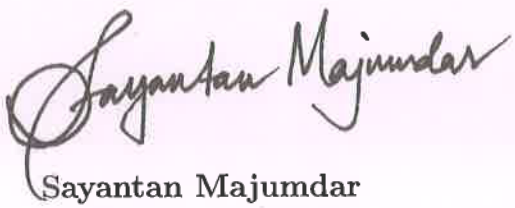
2024

Declaration by Student

1. **Name of the student:** Sayantan Majumdar
2. **Student Reg No.:** 20183589
3. **Thesis supervisor:** Dr. Pranay Goel
4. **Department:** Department of Biology
5. **Date of joining program:** 01-August-2018
6. **Date of pre-synopsis seminar:** 27-December-2023
7. **Title of Thesis:** "Analysis of Continuous Glucose Monitoring in relation to HbA1c and Interstitial Fluid Glucose"

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I declare that I have acknowledged collaborative work and discussions wherever such work has been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

The work reported in this thesis is the original work done by me under the guidance of **Dr. Pranay Goel**



Sayantan Majumdar
Date: 21-11-2024

CERTIFICATE

Certified that the work incorporated in the thesis entitled "**Analysis of Continuous Glucose Monitoring in relation to HbA1c and Interstitial Fluid Glucose**" Submitted by **Sayantan Majumdar** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Supervisor: Dr. Pranay Goel

Date: 21 November, 2024

डॉ. प्रणय गोयल / Dr. Pranay Goel
सहयोगी प्राध्यापक / Associate Professor
भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान
Indian Institute of Science Education & Research
पुणे / Pune - 411 008, India

Acknowledgments

I would like to thank Prof. Pranay Goel for supervising my work and also Prof. Saroj Ghaskadbi for guiding me, proofreading my manuscripts and supervising my work.
I would like to thank my colleagues at the laboratory, Sandra, Arjun, Somashree, Shreya, Rashmi, Suyog, Prajjwal and, Ajit who not only helped me when I fell ill but also helped me refine my work.

Abstract

In recent times, continuous glucose monitoring (CGM) devices are becoming very popular, especially among clinicians and patients suffering from diabetes (both T1D and, T2D). They help in management of patients BG levels, in the formulation of the patient's medication regime, etc. These CGM devices measure the concentration of glucose in the interstitial fluid periodically (usually every 5 min to 15 min depending on the device) for several days (usually from 7 days to 14 days, depending on the device) and provide the user with a sequence of time stamped ISF measurements, commonly known as the CGM trace. These devices are gradually becoming cheaper and more robust and in recent times are quite indispensable to many diabetes patients and clinicians.

Our goal is to use the CGM trace to estimate all other important metrics like, HbA1c, FBG, etc. This would not only reduce the total cost of diagnosis and management but also would remove the need for repeated blood-draws from the patient. In this thesis we show that the standard model used for estimating HbA1c from CGM traces, the Nathan model, cannot provide statistically reliable HbA1c estimates for Indian CGM traces at a population level. We then introduce two new models, which are able to provide statistically reliable HbA1c estimates for the Indian T2D population. We then show that the Direct Model is also able to provide reliable HbA1c estimates for the Indian T1D population. Next, we provide a method, based on trajectory optimization, for estimating the continuous BG and ISF traces from any given discrete CGM trace.

Contents

1 Introduction **1**
1.1 CGM Datasets 5
1.2 Thesis outline 6
1.3 Literature review 8
References 13

2 Evaluation of HbA1c from CGM traces in an Indian T2D population **22**
2.1 Introduction 23
2.2 Materials and methods 25
2.2.1 Subject recruitment and measurement of blood biochemical parameters 25
2.2.2 Comparing Nathan HbA1c estimates with experimentally measured HbA1c 26
2.2.3 Direct Model 27
2.2.4 Adapted Nathan Model 29
2.3 Results 30
2.3.1 Coefficient estimates for the direct model 30
2.3.2 Coefficient estimates of the adapted Nathan model 31
2.3.3 Validation of the direct and adapted models 32
2.4 Discussion 33
References 37

3 Evaluation of HbA1c from CGM traces in an Indian T1D popu-

lation **40**
3.1 Introduction . . . 40
3.2 Materials and Methods . . . 45
3.2.1 The T1D Dataset . . . 45
3.2.2 The Nathan Model for estimating HbA1c . . . 48
3.2.3 The Direct Model for estimating HbA1c . . . 49
3.2.4 The Adapted Nathan Model for estimating HbA1c . . . 51
3.3 Results . . . 52
3.3.1 Validation of the Direct Model . . . 52
3.3.2 Validation of the Nathan model and the Adapted Nathan Model . . . 58
3.4 Discussion . . . 60
References . . . 63

4 Estimating continuous BG from CGM traces **67**
4.1 Introduction . . . 67
4.2 Materials and methods . . . 72
4.2.1 The Dataset . . . 72
4.2.2 A simple model describing ISF—BG dynamics . . . 72
4.2.3 Trajectory optimization . . . 74
4.2.4 Collocation methods . . . 77
4.3 Results . . . 87
4.3.1 Estimating BG and ISF trajectories using Trapezoidal collocation . . . 87
4.3.2 Estimating BG and ISF trajectories using a modified Hermite-Simpson collocation method . . . 95
4.4 Discussion . . . 103
References . . . 106

5 An Web Application for calculating CGM metrics **127**
5.1 Introduction . . . 127
5.2 Using `CGM Analyzer` . . . 127
5.2.1 Using `CGM Analyzer` version 0.1 . . . 128

5.2.2 Using `CGM Anlayzer` version 2.0 . . . 130
5.3 Estimating continuous BG from a CGM trace . . . 136
5.4 Metrics . . . 137
5.4.1 Duration of the record . . . 137
5.4.2 Time CGM Device active . . . 137
5.4.3 Data Sufficiency Lamp . . . 138
5.4.4 Average Interstitial Glucose (aISF) . . . 138
5.4.5 Standard Deviation (SD) . . . 139
5.4.6 Coefficient of Variation (CV) . . . 139
5.4.7 Glycemic Stability Lamp . . . 139
5.4.8 Inter Quartile Range (IQR) . . . 140
5.4.9 Glucose Indexes HBGI and LBGI . . . 140
5.4.10 Estimated Glycated Hemoglobin (HbA1c) . . . 141
5.4.11 Area under the curve . . . 142
5.4.12 Duration in ranges . . . 142
5.4.13 Average Duration per Day . . . 143
5.4.14 Percentage time in ranges . . . 143
5.4.15 No. of Episodes . . . 144
References . . . 144

6 Discussion **146**
References . . . 149

A Blood glucose concentration (BG) trace estimates for the Pune-2021 dataset **151**

B Published works **202**
B.1 Published articles . . . 202
B.2 Conference posters and presentations . . . 202
B.3 Copyright and Licence . . . 203

C Ethics Statements **205**

Abbreviations

List of Figures

2.1 The experimentally measured HbA_{1c} vs. the aISF (calculated as described in Section 2.2.3) values of the CGM-dataset with the direct and adapted Nathan model with their 95 % confidence interval. . . . 34
2.2 The experimentally measured HbA_{1c} vs. the aISF (calculated as described in Section 2.2.3) values of the CGM-dataset with the direct and adapted Nathan model with their 95 % prediction interval. . . . 35

3.1 The Quantile-Quantile plots for all the models. 54
3.2 All three models along with their 95 % CI. 55
3.3 All three models along with their 95 % PI. 56
3.4 The Bland-Altman plots for all the models. 57

4.1 Estimated BG and ISF for CGM D 05 using trapezoidal collocation for $\tau_0 = 1\,\text{min}$ 109
4.2 Estimated BG and ISF for CGM D 05 using trapezoidal collocation for $\tau_0 = 1/5\text{min}$ 110
4.3 Estimated BG and ISF for CGM D 05 using trapezoidal collocation for $\tau_0 = 1/15\text{min}$ 111
4.4 Estimated BG and ISF for CGM PD 06 using trapezoidal collocation for $\tau_0 = 1\,\text{min}$ 112
4.5 Estimated BG and ISF for CGM PD 06 using trapezoidal collocation for $\tau_0 = 1/5\text{min}$ 113
4.6 Estimated BG and ISF for CGM PD 06 using trapezoidal collocation for $\tau_0 = 1/15\text{min}$ 114

4.7 Estimated BG and ISF for CGM ND 03 using trapezoidal collocation for $\tau_0 = 1\,\text{min}$ . . . 115
4.8 Estimated BG and ISF for CGM ND 03 using trapezoidal collocation for $\tau_0 = 1/5\text{min}$ . . . 116
4.9 Estimated BG and ISF for CGM ND 03 using trapezoidal collocation for $\tau_0 = 1/15\text{min}$ . . . 117
4.10 Estimated BG and ISF for CGM D 05 using the modified Hermite-Simpson collocation for $\tau_0 = 1\,\text{min}$ . . . 118
4.11 Estimated BG and ISF for CGM D 05 using the modified Hermite-Simpson collocation for $\tau_0 = 1/5\text{min}$ . . . 119
4.12 Estimated BG and ISF for CGM D 05 using the modified Hermite-Simpson collocation for $\tau_0 = 1/15\text{min}$ . . . 120
4.13 Estimated BG and ISF for CGM PD 06 using the modified Hermite-Simpson collocation for $\tau_0 = 1\,\text{min}$ . . . 121
4.14 Estimated BG and ISF for CGM PD 06 using the modified Hermite-Simpson collocation for $\tau_0 = 1/5\text{min}$ . . . 122
4.15 Estimated BG and ISF for CGM PD 06 using the modified Hermite-Simpson collocation for $\tau_0 = 1/15\text{min}$ . . . 123
4.16 Estimated BG and ISF for CGM ND 03 using the modified Hermite-Simpson collocation for $\tau_0 = 1\,\text{min}$ . . . 124
4.17 Estimated BG and ISF for CGM ND 03 using the modified Hermite-Simpson collocation for $\tau_0 = 1/5\text{min}$ . . . 125
4.18 Estimated BG and ISF for continuous glucose monitoring (CGM) ND 03 using the modified Hermite-Simpson collocation for $\tau_0 = 1/15\text{min}$ . . . 126

A.1 Estimated BG and ISF for `CGM D 02` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 152
A.2 Estimated BG and ISF for `CGM D 03` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 153
A.3 Estimated BG and ISF for `CGM D 04` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 154
A.4 Estimated BG and ISF for `CGM D 05` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 155
A.5 Estimated BG and ISF for `CGM D 06` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 156
A.6 Estimated BG and ISF for `CGM D 07` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 157
A.7 Estimated BG and ISF for `CGM D 08` for $\tau_0 = 1\,\text{min}$, $1/5\text{min}$ and, $1/15\text{min}$ 158

A.8 Estimated BG and ISF for CGM D 09 for $\tau_0 = 1$ min, 1/5min and, 1/15min159
A.9 Estimated BG and ISF for CGM D 10 for $\tau_0 = 1$ min, 1/5min and, 1/15min160
A.10 Estimated BG and ISF for CGM D 11 for $\tau_0 = 1$ min, 1/5min and, 1/15min161
A.11 Estimated BG and ISF for CGM D 12 for $\tau_0 = 1$ min, 1/5min and, 1/15min162
A.12 Estimated BG and ISF for CGM D 13 for $\tau_0 = 1$ min, 1/5min and, 1/15min163
A.13 Estimated BG and ISF for CGM D 14 for $\tau_0 = 1$ min, 1/5min and, 1/15min164
A.14 Estimated BG and ISF for CGM D 15 for $\tau_0 = 1$ min, 1/5min and, 1/15min165
A.15 Estimated BG and ISF for CGM D 16 for $\tau_0 = 1$ min, 1/5min and, 1/15min166
A.16 Estimated BG and ISF for CGM D 17 for $\tau_0 = 1$ min, 1/5min and, 1/15min167
A.17 Estimated BG and ISF for CGM ND 01 for $\tau_0 = 1$ min, 1/5min and, 1/15min168
A.18 Estimated BG and ISF for CGM ND 02 for $\tau_0 = 1$ min, 1/5min and, 1/15min169
A.19 Estimated BG and ISF for CGM ND 03 for $\tau_0 = 1$ min, 1/5min and, 1/15min170
A.20 Estimated BG and ISF for CGM ND 04 for $\tau_0 = 1$ min, 1/5min and, 1/15min171
A.21 Estimated BG and ISF for CGM ND 05 for $\tau_0 = 1$ min, 1/5min and, 1/15min172
A.22 Estimated BG and ISF for CGM ND 06 for $\tau_0 = 1$ min, 1/5min and, 1/15min173
A.23 Estimated BG and ISF for CGM ND 07 for $\tau_0 = 1$ min, 1/5min and, 1/15min174
A.24 Estimated BG and ISF for CGM ND 08 for $\tau_0 = 1$ min, 1/5min and, 1/15min175
A.25 Estimated BG and ISF for CGM ND 09 for $\tau_0 = 1$ min, 1/5min and, 1/15min176
A.26 Estimated BG and ISF for CGM ND 10 for $\tau_0 = 1$ min, 1/5min and, 1/15min177
A.27 Estimated BG and ISF for CGM ND 11 for $\tau_0 = 1$ min, 1/5min and, 1/15min178
A.28 Estimated BG and ISF for CGM ND 12 for $\tau_0 = 1$ min, 1/5min and, 1/15min179
A.29 Estimated BG and ISF for CGM ND 13 for $\tau_0 = 1$ min, 1/5min and, 1/15min180
A.30 Estimated BG and ISF for CGM ND 14 for $\tau_0 = 1$ min, 1/5min and, 1/15min181
A.31 Estimated BG and ISF for CGM ND 15 for $\tau_0 = 1$ min, 1/5min and, 1/15min182
A.32 Estimated BG and ISF for CGM ND 16 for $\tau_0 = 1$ min, 1/5min and, 1/15min183
A.33 Estimated BG and ISF for CGM ND 18 for $\tau_0 = 1$ min, 1/5min and, 1/15min184
A.34 Estimated BG and ISF for CGM ND 19 for $\tau_0 = 1$ min, 1/5min and, 1/15min185
A.35 Estimated BG and ISF for CGM ND 20 for $\tau_0 = 1$ min, 1/5min and, 1/15min186
A.36 Estimated BG and ISF for CGM PD 01 for $\tau_0 = 1$ min, 1/5min and, 1/15min187
A.37 Estimated BG and ISF for CGM PD 02 for $\tau_0 = 1$ min, 1/5min and, 1/15min188
A.38 Estimated BG and ISF for CGM PD 03 for $\tau_0 = 1$ min, 1/5min and, 1/15min189
A.39 Estimated BG and ISF for CGM PD 04 for $\tau_0 = 1$ min, 1/5min and, 1/15min190

A.40 Estimated BG and ISF for `CGM PD 05` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 191
A.41 Estimated BG and ISF for `CGM PD 06` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 192
A.42 Estimated BG and ISF for `CGM PD 07` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 193
A.43 Estimated BG and ISF for `CGM PD 08` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 194
A.44 Estimated BG and ISF for `CGM PD 09` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 195
A.45 Estimated BG and ISF for `CGM PD 10` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 196
A.46 Estimated BG and ISF for `CGM PD 11` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 197
A.47 Estimated BG and ISF for `CGM PD 12` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 198
A.48 Estimated BG and ISF for `CGM PD 13` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 199
A.49 Estimated BG and ISF for `CGM PD 14` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 200
A.50 Estimated BG and ISF for `CGM PD 15` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$ 201

B.1 copywrite information for incorporating Majumdar et al. [1]. 204

List of Tables

3.1 The T1D dataset description table 47
3.2 The results of the paired t-test for all three models. 53

4.1 Table showing the degree of agreement of estimated BG within the measured interstitial fluid glucose concentration (ISF) ranges for trapezoidal collocation with $\tau = 1\,\text{min}$ 89
4.2 Table showing the degree of agreement of estimated BG within the measured ISF ranges for trapezoidal collocation with $\tau = 1/5\text{min}$. 90
4.3 Table showing the degree of agreement of estimated BG within the measured ISF ranges for trapezoidal collocation with $\tau = 1/15\text{min}$. 91
4.4 Table showing the degree of agreement of measured ISF within the estimated blood glucose concentration (BG) ranges for trapezoidal collocation with $\tau = 1$ min 92
4.5 Table showing the degree of agreement of measured ISF within the estimated BG ranges for trapezoidal collocation with $\tau = 1/5\text{min}$. 93
4.6 Table showing the degree of agreement of measured ISF within the estimated BG ranges for trapezoidal collocation with $\tau = 1/15\text{min}$. 94
4.7 Table showing the degree of agreement of estimated BG within the measured ISF ranges for the modified Hermite-Simpson collocation with $\tau = 1\,\text{min}$ 97
4.8 Table showing the degree of agreement of estimated BG within the measured ISF ranges for the modified Hermite-Simpson collocation with $\tau = 1/5\text{min}$ 98

4.9 Table showing the degree of agreement of estimated BG within the measured ISF ranges for the modified Hermite-Simpson collocation with $\tau = 1/15$min 99
4.10 Table showing the degree of agreement of measured ISF within the estimated BG ranges for the modified Hermite-Simpson collocation with $\tau = 1$ min 100
4.11 Table showing the degree of agreement of measured ISF within the estimated BG ranges for the modified Hermite-Simpson collocation with $\tau = 1/5$min 101
4.12 Table showing the degree of agreement of measured ISF within the estimated BG ranges for the modified Hermite-Simpson collocation with $\tau = 1/15$min 102

Chapter 1

Introduction

Diabetes mellitus (DM) is a common metabolic disorder and in recent times it is affecting an increasing number of people each year. There are affectively two ways in which this metabolic disorder manifests itself. First as type 1 diabetes (T1D), where the individual looses their pancreatic $\beta-$ cells and therefore is unable to produce the insulin required to maintain an appropriate blood glucose concentration (BG). These patients require constant glucose monitoring every day and need periodic insulin injections every day to help maintain an appropriate BG levels. In the second manifestation the disorder is labeled as type 2 diabetes (T2D). This manifestation is characterized by an increased resistance of the individuals body to the effects of insulin. It appears as if the tissues in the individuals body are increasingly immune or unaffected by the presence of insulin. This is sometimes also accompanied by a reduction in insulin production by the individual. This results in a consistently increased rate of insulin production by the diminished pancreatic $\beta-$cells , which leads to them being stressed for a majority of their lifetime, which in tern also degrades their ability to produce insulin efficiently. Therefore, in both manifestations the disorder is characterized by the inability of the individual to maintain appropriate BG levels resulting in a consistently high BG even though physiologically both manifestations of the disorder are significantly different.

To understand this metabolic disorder, it is important to understand the glucose insulin dynamics of humans. The glucose-insulin dynamical system of the human body would not only involve the insulin and glucose concentrations in var-

ious parts of the body but would also involve various other enzymes, proteins and hormone secretions that are involved with utilizing glucose or processing glucose in the human body. Analyzing such a high-dimensional and complex dynamical system is very cumbersome and difficult. Therefore, there have been many attempts at constructing a simple model involving just a few variables like, the BG and, the insulin concentration, that can mimic the original glucose insulin dynamics of the human body with an appropriate level of accuracy. These generally involve probing the system and constructing a simple dynamical system which at least takes into account the insulin and the glucose concentrations at various parts of the body and involve verifying if the models could at least demonstrate qualitatively a few of the phenomenon involved with diabetes. As the glucose insulin dynamics is very complex and it is very difficult to probe complex metabolic pathways of the human body; there has been only a few reliable glucose insulin models which can represent many of the phenomena involved with diabetes.

Modern technology however has recently (in the last 20 years) provided new tools that should help probe such dynamical systems easily in a reliable manner. There have been various fast glucose monitoring (FGM) devices which allow an individual to measure the glucose concentration of their blood very easily and with minimum discomfort. These devices are very important for T1D patients who need to repeatedly measure and, maintain their BG levels, multiple times a day. This allowed us to construct studies where the BG of the participants were monitored regularly in short intervals, essential generating a sequence of time stamped BG measurements which could be considered as a discrete representation of the individual's continuous BG. This provided immense help in understanding the glucose insulin dynamics of the human body and lead to the creation of various glucose insulin models. Therefore, by understanding the value of such discrete representations of the BG and it's need, for patients who have to regularly monitor their BG, the continuous glucose monitoring (CGM) devices were introduced. CGM devices are equipped with sensors that are placed in the interstitial fluid, under the skin of an individual and are capable of repeated measurements of the concentration of glucose in the interstitial fluid. These devices usually stay attached on the individual for a period of 7 to 14 days and are usually capable of storing and measuring

the interstitial fluid glucose concentration (ISF) every minute. These sequences of time stamped ISF measurements are generally referred to as the individual's CGM trace. These CGM traces provide, in essence, a high resolution snapshot of the individuals 14 day continuous ISF. These have been immensely helpful in understanding the dynamics of glucose and insulin for humans. These devices are accurate and easy to use and so are clinically used to monitor the patients glycemic state and also prescribed to patients with severe insulin deficiency or severe insulin resistance, patients who generally need to be administered insulin injections, to help them maintain their BG levels.

In recent times (in the last 5 years) CGM devices have become very popular and relatively cheap. Therefore, they are being widely adapted by clinicians all over the world. This has given rise to a school of thought where, it is believed that we only need a patient's CGM trace and all other metrics required to diagnose the patient's diabetes could be estimated reliably from this CGM trace. This is intriguing as it would definitely (as CGM devices get cheaper) make the diagnosis and regular check up of diabetes more affordable and hassel free. In that line, there have been various attempts at estimating popular metrics like the fasting blood glucose concentration (FBG), post-prandial blood glucose concentration (PPBG) and, Glycated Hemoglobin (HbA1c) from a given CGM and have also given rise to a vast number of new metrics like the time in range (TiR), glucose monitoring index (GMI), mean absolute difference score (MADS), etc. This is an active field of research and has direct clinical impact. Till date there have been a huge number of proposed metrics that measure or represent or indicate various aspects of the glucose insulin dynamical system for humans. Only a few have survived the test of time and have proven their clinical usefulness and reliability. HbA1c is one such metric, it is essentially the percentage of mols of hemoglobin (Hb) in the blood stream which have chemically bonded with sugar molecules.

Measuring the HbA1c is important, because a consistently high HbA1c value would indicate consistent high concentrations of glucose in the blood stream, indicating a potential failure of the body to regulate BG levels. HbA1c in the blood can be measured using various methods, the most popular of them being high-performance liquid chromatography (HPLC). Earlier these laboratory methods

were the only methods of measuring HbA1c and with the advent of FGM devices there has been various attempts at estimating the HbA1c of an individual using BG measurements made using FGM devices. These estimates essentially requires using a FGM device to make multiple measurement of the individuals BG and then calculating the aggregate average blood glucose concentration (aBG) and estimating the HbA1c from the calculated aBG value using a linear or non-linear model. Of these various attempts, the Nathan model [9] is vastly regarded as the standard model for estimating HbA1c. It was based on self-monitoring blood glucose (SMBG) and CGM data. The Nathan model [9] however is a population level model and is unable to give clinically reliable HbA1c estimates for an individual.

We show that the Nathan model [9] is unable to provide a statistically reliable HbA1c estimates at a population level with CGM traces form patients with T2D, patients with pre-diabetes (T2D) and non-diabetic individuals from the Indian population. We then suggest two new models constructed from a CGM dataset based on linear regression and constrained linear regression namely, the direct [32] and, adapted Nathan model [32] which can provide statistically reliable HbA1c estimates at a population level for patients with T2D, patients with pre-diabetes (T2D) and non-diabetic individuals from the Indian population. We later show that the Nathan model [9] is also unable to provide statistically reliable HbA1c values at a population level with CGM traces from adolescent T1D patients of the Indian population where as the direct model [32] is able to provide a statistically reliable HbA1c estimates at a population level from CGM traces for adolescent Indian T1D patients.

We later focus on providing an estimate of the continuous BG trajectory from the corresponding discrete ISF measurements provided by a CGM device. Traditionally the protocols, algorithms and, methods used for diagnosis and management of diabetes were mostly based on the concentration of glucose in blood, BG. However, with the advent of FGM and CGM devices there is no longer a need to rely on methods that only focus on BG, or consistent excess BG as an indicator for diabetes. FGM and CGM devices are encouraging clinicians to use ISF traces and indicators like GMI for diagnosis and management of DM. And it is argued that as the CGM devices are constructed/designed in such a way that

the difference between the measured ISF and BG at the time of measurement is below a certain error limit, the discrete ISF trajectory can be used as a substitute for the continuous BG. However, there have been many studies showing that the discrete ISF cannot be considered a replacement for the continuous BG, Cobelli et al. [19] showed that the relationship between BG and ISF is not a straight forward scaling or a simple time delay equation, even though there have been many studies showing that the time-lag between the ISF and BG is negligible when it comes to its clinical use. It is therefore extremely important that we find a reliable method of estimating the continuous ISF and the corresponding BG trace from the measured discrete ISF trace. We therefore provide a method based on trajectory optimization to obtain an estimate of the continuous BG and ISF trajectory with a desirable degree of smoothness from the discrete ISF measurements provided by a CGM trace.

In this thesis, I present my work on the analysis of CGM traces, on estimating the metric HbA1c and, on estimating the continuous BG trace corresponding to the measured discrete ISF trace provided by the CGM device. To complete my work I had received access to 4 annonymised CGM datasets which were paired into effectively two datasets and used for analysis. The details of these datasets are provided in Section. 1.1 and the documents for the ethical clearances to use these datasets are provided in Appendix. C. The outline of my thesis is presented in a brief manner in Section. 1.2.

1.1 CGM Datasets

We have received access to 4 CGM datasets but 2 of them were paired together and used as one CGM dataset. We here provide the list of ethics clearance documents that we obtained so that we could access these annonymised datasets.

1. **The Pune-2021 dataset:**

   The reviewed data analysis for this dataset was approved by the Institutional Human Ethics Committee (IHEC) of IISER Pune and the document (filenumber: `IHEC/Admin/2021/15`) is attached in Appendix. C.

The protocol for the study was approved by the Institutional Ethical Committee (IEC), Savitribai Phule Pune University and the ethics clearency document (filenumber: SPPU/IEC/2020/102) is attached in the Appendix. C

2. **The Joshi-2018 dataset:**

   The reviewed data analysis for this dataset was approved by the Institutional Human Ethics Committee (IHEC) of IISER Pune and the ethics clearenc document (filenumber: IHEC/Admin/2021/15) along as an extension is attached in Appendix. C.
   The protocol for the study was approved by the Institutional Ethical Committee (IEC), Maharashtra Medical Research Society and the ethics clearenc document (filenumber: ECR/311/Inst/MH/2013/RR-19; Dated 14 March 2023) is attached in the Appendix. C

3. **The T1D 2022–2023 and T1D 2019–2020 dataset:**

   The reviewed data analysis for this dataset was approved by the Institutional Human Ethics Committee (IHEC) of IISER Pune and the ethics clearenc document (filenumber: IHEC/Admin/2023/002) along as an extension is attached in Appendix. C.

   The protocol for the T1D 2022–2023 study was approved by the Ethics Committee (EC) Jehangir Clinical Development Center Pvt. Ltd. and the ethics clearenc document along as an extension is attached in Appendix. C.
   The protocol for the T1D 2019–2020 study was approved by the Ethics Committee (EC) Jehangir Clinical Development Center Pvt. Ltd. and the ethics clearenc document (filenumber: ECR/352/Inst/MH/2013/RR-19) along as an extension is attached in Appendix. C.

1.2 Thesis outline

A detailed outline of the thesis, beyond Chapter. 1, with descriptions of all chapters that are present in this thesis is enumerated below,

1. **Chapter. 2:** In this chapter we discuss the standard algorithm used for obtaining HbA1c estimates from any given CGM trace, the Nathan model [9]. We show that using this model we cannot obtain statistically reliable estimates of HbA1c from a CGM trace for the Indian population. We therefore, suggest and then validate 2 new models, the first named Direct Model and the second named Adapted Nathan Model [32], which can be used to obtain statistically reliable estimates of HbA1c from a given CGM trace for the Indian population.

2. **Chapter. 3:** In this chapter we discuss the lack of a proper algorithm to reliably obtain estimates of HbA1c from a CGM trace for T1D patients. We then use a CGM dataset obtained from adolescent Indian T1D patients to verify if our 2 models, the Adapted Nathan Model [32] and the Direct Model[32] could provide statistically reliable HbA1c estimates from a given CGM trace for adolescent Indian T1D patients.

3. **Chapter. 4:** Diabetes as a metabolic disorder has been around for some time now. It has been studied extensively but traditional research in diabetes mainly focused around metrics like HbA1c, FBG, PPBG and, in general BG, etc. These metrics are based on the concentration of glucose in the blood, BG and with years or research behind them, are highly reliable. The introduction of modern CGM devices have opened the door to various other metrics, many of which have a complex dynamic relationship with BG. We therefore need a reliable method of converting the ISF values in a CGM trace to their corresponding BG values. In this chapter we introduce a method based on trajectory optimization which allows us to obtain an estimate of the continuous BG trace with an appropriate degree of smoothness from any given CGM trace.

4. **Chapter. 5:** In this chapter we describe a versatile web application that we constructed for seamlessly calculating almost all major and standard metrics related to diabetes diagnosis and it's research. In 2017 a list of such metrics were provided by Danne et al. [21], this consensus paper contained a list of 18 metrics which were commonly used by clinicians and diabetes researchers

all over the world. This app can be accessed online and allows any person to upload a CGM trace and calculate those metrics.

1.3 Literature review

Diabetes is characterized primarily by a consistent high BG and, HbA1c represents the fraction of Hb in the red blood cells (RBCs) that have bonded with sugar molecules (glucose, fructose, etc). Therefore, a consistently high glucose in the blood and eventually in the RBCs would lead to a large percentage of Hb molecules bonding with sugar. This bonding of Hb with sugar molecules is referred to as glycation [25]. The already glycated Hb molecules in the RBCs tend to remain in this glycated state. Therefore for the lifetime of that RBC the percentage of glycated Hb, or HbA1c would be a proxy for the aggregate concentration of sugar molecules in the blood. Therefore, high HbA1c is primarily an indicator for consistently high BG levels.

HbA1c therefore, has been used for the last few decades as an indicator for the overall glycemic state of an individual. HbA1c is very reliable and is usually stable if there aren't any drastic changes in the diet or exercise regime of the individual. Li et al. [38] showed that even for variants of Hb, the HbA1c calculated from them are negligibly affected. Most HbA1c studies are however done on diabetic patients whose BG levels vary drastically over time. It is therefore important to study the reliability of HbA1c as a diagnostic metric for diabetic patients. It is also important to study the variability of HbA1c over different ethnicity. Cavagnolli et al. [20] showed HbA1c varies in non-diabetic individuals over different ethnicity. HbA1c now a days is also used as a decision variable in the primary assessment of diabetes however, Friedman et al. [36] showed that diabetes assessment using only HbA1c may not be reliable, other methods including fructosamine tests and CGM traces should also be included in taking that decision. This is because even though HbA1c, fructosamine and, average interstitial fluid glucose concentration (aISF) from CGM are highly correlated they do not always agree within a clinically reliable error limit. Mao and Zhong [40] showed that there is an association between HbA1c and vascular complications. This indicates that HbA1c can also be used

as potential indicators for various other disorders. Lind et al. [10] showed that a mean HbA1c of various HbA1c measurements over time can be used as a very good indicator for diabetic complications.

It is clear that although not perfect, HbA1c is a great indicator for diabetes and other glycemic disorders. Even HbA1c variability can be used as an diabetes management indicator. Derr et al. [4] showed that HbA1c variability is important in the management of diabetes. For people with low HbA1c variability, keeping the HbA1c low is important for avoiding diabetic complications.

This success of HbA1c as a good indicator for glycemic disorders is the reason why an algorithm for estimating HbA1c from any CGM trace is so important. Metrics constructed using the corelation between HbA1c and aBG or aISF also could be used as a proxy for HbA1c and could also themselves become an important metric. GMI is one such new metric, exploiting the corelation between aBG and HbA1c, GMI can be used as a proxy for HbA1c or can be considered as an estimated HbA1c [24]. Fang et al. [31] showed that although GMI is a reliable metric and a good estimator for HbA1c, It is not reliable as an HbA1c estimate for T2D patients. Azcoitia et al. [34] showed that the GMI and HbA1c are significantly different in a majority of cases. They also showed that other metrics like mean corpuscular volume (MCV) could be used to understand the difference between GMI and HbA1c. Manov et al. [39] however showed that the differences between HbA1c and GMI are not significantly. Selvin [44] also argued against the use of GMI as an estimate for HbA1c.

It is clear that a reliable method of estimating HbA1c is needed. There are many models which have been shown to work on a sample of CGM traces but is later found to be unreliable for some other sample of CGM traces. This is especially surprising because many of the mathematical models used for estimating HbA1c statistically agree with each others' estimate but are not consistently able to provide reliable HbA1c estimates as compared to laboratory methods like HPLC. This clearly indicates a lack of understanding in our part of the fundamental model that is governing the glucose insulin dynamics and the glycation of Hb molecules in a RBC.

Understanding the system that govern the HbA1c—BG relationship is therefore

of crucial importance and there have been various attempts at it. Nathan et al. [9] produced one of the most popular models for estimating HbA1c from a CGM trace. It is a linear regression model between the aBG and the the measured HbA1c. Hu et al. [27] recently constructed many unique models for exploring the relationship between CGM traces and HbA1c for the Chinese population. This group of models also contained many nonlinear and exotic models. There have been several attempts at constructing a simple model for predicting HbA1c in terms of the aBG derived from a CGM. Many of these models are designed specifically for patients already suffering from some other medical complications. Møller et al. [17] developed a new method for estimating HbA1c using previously measured BG and HbA1c for patients undergoing treatment for diabetes. Hoshino et al. [14] studied the relationship between aBG and HbA1c for patients undergoing dialysis. They developed a new method to reliably estimate the HbA1c for patients undergoing dialysis from their aBG. Law et al. [22] studied the relationship between HbA1c and aBG calculated from patient CGM. They showed that the traditional method for calculating aBG from a given CGM trace are unreliable for pregnant women and therefore developed a pregnancy specific aBG calculation method to estimate the aBG for pregnant women.

Through the development of many such models, it has been observed that a simple linear regression model or a simple dynamical model is unable to provide a reliable estimate of HbA1c beyond a certain arbitrary error. Therefore there have been developments where complicated nonlinear dynamic models have been used to predict HbA1c from a given CGM trace of a patient and now a days almost all the HbA1c models are nonlinear and tying to mimic the dynamical system involved fundamentally with HbA1c and BG. Kovatchev and Cox [5] developed a numerical method to dynamically estimate the HbA1c of a patient from a series of SMBG measurements. They also calculated the optimal SMBG measurement interval. Xu et al. [29] constructed a dynamical nonlinear model for estimating the HbA1c of an individual from the individuals CGM trace. It is one of the first such kinetic model which tries to incorporate the glucose dynamics on to a simple model, uses the CGM trace to estimate the parameters of that model and then uses that model is to predict HbA1c. The Xu et al. [29] model also takes into

consideration the life-span of RBCs [30] and is capable of providing HbA1c estimates at an individual level. Ladyzynski et al. [18] developed a small kinetic model describing the relationship between HbA1c and BG. They also showed that the relationship between mean HbA1c and mean aISF or aBG used to obtain linear relationships at the population level, only work if the CGM traces are glycemic stability. McEwan et al. [23] also developed a new kinetic model to understand and predict HbA1c progression in T2D patients. Kjellsson et al. [15] constructed a kinetic model, using the IGI model [7, 8, 11–13] and the IGRH model [16], that can approximate HbA1c from the measured BG and Insulin concentration in the blood. Mortensen and Vølund [3] also developed a kinetic model for predicting HbA1c and HbA1d from a patient's previous BG measurements. This is one of the first kinetic model which not only takes into account the patient's previously measured BG but also takes in to account the patient's RBC life span. Sato Imuro et al. [43] studied the requirements for developing personalized HbA1c models and developed a multi linear regression model with kinetic parameters for predicting the personalized HbA1c of the patient. There have been also various attempts at estimating HbA1c from a given CGM trace using machine learning. Kurasawa et al. [37] developed a machine learning model to identify defective glycemic control for T2D patients. Ngufor et al. [26] constructed a mixed effect model which could be used to identify longitudinal changes in HbA1c. Song et al. [45] showed that UMAP could be used to successfully distinguish between T1D patients with prescribed Insulin injections daily and healthy non-diabetic patients. The effectiveness of the model was validated through the patients HbA1c and other markers. Cichosz et al. [35] also showed that an XGBoost model can be used on CGM traces to accurately distinguish between non-diabetic, pre-diabetic and diabetic patients. Tao et al. [33] recently also developed a robust method for estimating the FBG and HbA1c changes based on multiple machine learning models for T2D patients undergoing treatment for diabetes. Qaraqe et al. [41] developed a new few shot learning (FSL)-derived model that can be used to estimate long-term HbA1c based on current CGM traces for pediatric T1D patients. Although these machine learning models are rare but they can provide surprisingly better estimates of HbA1c [42].

We too have generated two new models for estimating HbA1c from a patients CGM trace, the direct model and the adapted Nathan model [32]. These models provide HbA1c estimates along with their 95 % confidence interval at a population level for T2D patients as well as pre-diabetic and non-diabetic individuals.

Blood glucose concentration (BG) is the primary tool for diagnosis and management of both T1D and T2D [28]. As mentioned earlier diabetes is characterized by a consistent high BG. Almost all other metric for glycemic control fundamentally depends on the glucose concentration of the blood. Lets take HbA1c for example, HbA1c represents the percentage of Hb molecules that have been glycated in the RBCs. If the BG levels are consistently high, it will lead to a large percentage of Hb molecules getting glycated, increasing the HbA1c of the individual. The situation is similar in the case of GMI, fructosamine, etc. This indicates that BG is important for managine DM.

Modern diabetes diagnosis tools like CGM devices however only measure the ISF. These CGM devices are small, usually button sized probes that can be attached to the body. They pierce the skin and place a sensor just under the skin in the interstitial fluid. These sensors then periodically makes glucose concentration measurements and generates a sequence of ISF measurements. These measurements are close to their corresponding BG values, in the sense that these CGM devices are designed such that at the moment of measurement the difference between the measured ISF and the BG values will be within a certain range of errors. Although there may be some discrepancies, some time-delays in the measured ISF it is argued that the discrete-time ISF can be used as a good approximate of the BG at the points of measurement. This is true when only on a few occasions for example, when only the value of ISF at the measurement time is required like when calculating the aISF, it can be shown that the aISF converges to a finite asymptotic value as the number of measurements increase. If we were to use this discrete time series as a proxy for the continuous BG we need a reliable algorithm to estimate the continuous BG from the ISF trace. Trajectory optimization which is essentially a method to solve a specialized optimal control problem can be used to obtain an estimate of the continuous BG. The advantage of this method is that it can be used easily and a reliable estimate of the trajectory can be obtained [1,

2, 6].

References

[1] Balakrishnan, A. On a new computing technique in optimal control. *SIAM Journal on Control* **1968**, *6*, 149–173, DOI: 10.1137/0306012, https://epubs.siam.org/doi/10.1137/0306012 (cit. on p. 12).

[2] Hargraves, C. R.; Paris, S. W. Direct trajectory optimization using nonlinear programming and collocation. *Journal of guidance, control, and dynamics* **1987**, *10*, 338–342, DOI: 10.2514/3.20223, https://arc.aiaa.org/doi/10.2514/3.20223 (cit. on p. 12).

[3] Mortensen, H.; Vølund, A. Application of a biokinetic model for prediction and assessment of glycated haemoglobins in diabetic patients. *Scandinavian Journal of Clinical and Laboratory Investigation* **1988**, *48*, 595–602, DOI: 10.3109/00365518809085778, https://www.tandfonline.com/doi/abs/10.3109/00365518809085778 (cit. on p. 11).

[4] Derr, R.; Garrett, E.; Stacy, G. A.; Saudek, C. D. Is HbA1c affected by glycemic instability? *Diabetes care* **2003**, *26*, 2728–2733, DOI: 10.2337/diacare.26.10.2728, https://diabetesjournals.org/care/article/26/10/2728/26129/Is-HbA1c-Affected-by-Glycemic-Instability (cit. on p. 9).

[5] Kovatchev, B. P.; Cox, D. J. In *Methods in enzymology*; Elsevier: 2004; Vol. 384, pp 94–106, DOI: 10.1016/S0076-6879(04)84007-2, https://www.sciencedirect.com/science/article/abs/pii/S0076687904840072?via%3Dihub (cit. on p. 10).

[6] Paris, S.; Riehl, J.; Sjauw, W. In *AIAA/AAS astrodynamics specialist conference and exhibit*, 2006, p 6309, DOI: 10.2514/6.2006-6309, https://arc.aiaa.org/doi/10.2514/6.2006-6309 (cit. on p. 12).

[7] Jauslin, P. M.; Silber, H. E.; Frey, N.; Gieschke, R.; Simonsson, U. S.; Jorga, K.; Karlsson, M. O. An integrated glucose-insulin model to describe oral glucose tolerance test data in type 2 diabetics. *The Journal of Clinical Pharmacology* **2007**, *47*, 1244–1255, DOI: 10.1177/0091270007302168, https://accp1.onlinelibrary.wiley.com/doi/abs/10.1177/0091270007302168 (cit. on p. 11).

[8] Silber, H. E.; Jauslin, P. M.; Frey, N.; Gieschke, R.; Simonsson, U. S.; Karlsson, M. O. An integrated model for glucose and insulin regulation in healthy volunteers and type 2 diabetic patients following intravenous glucose provocations. *The Journal of Clinical Pharmacology* **2007**, *47*, 1159–1171, DOI: 10.1177/0091270007304457, https://accp1.onlinelibrary.wiley.com/doi/10.1177/0091270007304457 (cit. on p. 11).

[9] Nathan, D. M.; Kuenen, J.; Borg, R.; Zheng, H.; Schoenfeld, D.; Heine, R. J.; Group, A.-D. A. G. (S. Translating the A1C assay into estimated average glucose values. *Diabetes care* **2008**, *31*, 1473–1478, DOI: 10.2337/dc08-0545, https://diabetesjournals.org/care/article/31/8/1473/28589/Translating-the-A1C-Assay-Into-Estimated-Average (cit. on pp. 4, 7, 10).

[10] Lind, M.; Odén, A.; Fahlén, M.; Eliasson, B. The true value of HbA1c as a predictor of diabetic complications: simulations of HbA1c variables. *PloS one* **2009**, *4*, e4412, DOI: 10.1371/journal.pone.0004412, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0004412 (cit. on p. 9).

[11] Silber, H. E.; Frey, N.; Karlsson, M. O. An integrated glucose-insulin model to describe oral glucose tolerance test data in healthy volunteers. *The Journal of Clinical Pharmacology* **2010**, *50*, 246–256, DOI: 10.1177/0091270009341185, https://accp1.onlinelibrary.wiley.com/doi/abs/10.1177/0091270009341185 (cit. on p. 11).

[12] Jauslin, P. M.; Frey, N.; Karlsson, M. O. Modeling of 24-hour glucose and insulin profiles of patients with type 2 diabetes. *The Journal of Clinical Pharmacology* **2011**, *51*, 153–164, DOI: 10.1177/0091270010362536, https://

accp1.onlinelibrary.wiley.com/doi/abs/10.1177/0091270010362536 (cit. on p. 11).

[13] Jauslin, P. M.; Karlsson, M. O.; Frey, N. Identification of the mechanism of action of a glucokinase activator from oral glucose tolerance test data in type 2 diabetic patients based on an integrated glucose-insulin model. *The Journal of Clinical Pharmacology* **2012**, *52*, 1861–1871, DOI: 10.1177/0091270011422231, https://accp1.onlinelibrary.wiley.com/doi/abs/10.1177/0091270011422231 (cit. on p. 11).

[14] Hoshino, J.; Molnar, M. Z.; Yamagata, K.; Ubara, Y.; Takaichi, K.; Kovesdy, C. P.; Kalantar-Zadeh, K. Developing an HbA1c-based equation to estimate blood glucose in maintenance hemodialysis patients. *Diabetes Care* **2013**, *36*, 922–927, DOI: 10.2337/dc12-1019, https://diabetesjournals.org/care/article/36/4/922/37889/Developing-an-HbA1c-Based-Equation-to-Estimate (cit. on p. 10).

[15] Kjellsson, M. C.; Cosson, V. F.; Mazer, N. A.; Frey, N.; Karlsson, M. O. A model-based approach to predict longitudinal HbA1c, using early phase glucose data from type 2 diabetes mellitus patients after anti-diabetic treatment. *The Journal of Clinical Pharmacology* **2013**, *53*, 589–600, DOI: 10.1002/jcph.86, https://accp1.onlinelibrary.wiley.com/doi/10.1002/jcph.86 (cit. on p. 11).

[16] Lledó-García, R.; Mazer, N. A.; Karlsson, M. O. A semi-mechanistic model of the relationship between average glucose and HbA1c in healthy and diabetic subjects. *Journal of pharmacokinetics and pharmacodynamics* **2013**, *40*, 129–142, DOI: 10.1007/s10928-012-9289-6, https://link.springer.com/article/10.1007/s10928-012-9289-6 (cit. on p. 11).

[17] Møller, J.; Overgaard, R.; Kjellsson, M. C.; Kristensen, N.; Klim, S.; Ingwersen, S.; Karlsson, M. O. Longitudinal modeling of the relationship between mean plasma glucose and HbA1c following antidiabetic treatments. *CPT: Pharmacometrics & Systems Pharmacology* **2013**, *2*, 1–8, DOI: 10.1038/psp.2013.58, https://ascpt.onlinelibrary.wiley.com/doi/10.1038/psp.2013.58 (cit. on p. 10).

[18] Ladyzynski, P.; Foltynski, P.; Bak, M. I.; Sabalinska, S.; Krzymien, J.; Kawiak, J. Validation of a hemoglobin A 1c model in patients with type 1 and type 2 diabetes and its use to go beyond the averaged relationship of hemoglobin A 1c and mean glucose level. *Journal of translational medicine* **2014**, *12*, 1–16, DOI: 10.1186/s12967-014-0328-5, https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-014-0328-5 (cit. on p. 11).

[19] Cobelli, C.; Schiavon, M.; Dalla Man, C.; Basu, A.; Basu, R. Interstitial fluid glucose is not just a shifted-in-time but a distorted mirror of blood glucose: insight from an in silico study. *Diabetes technology & therapeutics* **2016**, *18*, 505–511, DOI: 10.1089/dia.2016.0112, https://www.liebertpub.com/doi/10.1089/dia.2016.0112 (cit. on p. 5).

[20] Cavagnolli, G.; Pimentel, A. L.; Freitas, P. A. C.; Gross, J. L.; Camargo, J. L. Effect of ethnicity on HbA1c levels in individuals without diabetes: systematic review and meta-analysis. *PloS one* **2017**, *12*, e0171315, DOI: 10.1371/journal.pone.0171315, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171315 (cit. on p. 8).

[21] Danne, T.; Nimri, R.; Battelino, T.; Bergenstal, R. M.; Close, K. L.; DeVries, J. H.; Garg, S.; Heinemann, L.; Hirsch, I.; Amiel, S. A., et al. International consensus on use of continuous glucose monitoring. *Diabetes care* **2017**, *40*, 1631–1640, DOI: 10.2337/dc17-1600, https://diabetesjournals.org/care/article/40/12/1631/37000/International-Consensus-on-Use-of-Continuous (cit. on p. 7).

[22] Law, G. R.; Gilthorpe, M. S.; Secher, A. L.; Temple, R.; Bilous, R.; Mathiesen, E. R.; Murphy, H. R.; Scott, E. M. Translating HbA1c measurements into estimated average glucose values in pregnant women with diabetes. *Diabetologia* **2017**, *60*, 618–624, DOI: 10.1007/s00125-017-4205-7, https://link.springer.com/article/10.1007/s00125-017-4205-7 (cit. on p. 10).

[23] McEwan, P.; Bennett, H.; Qin, L.; Bergenheim, K.; Gordon, J.; Evans, M. An alternative approach to modelling HbA1c trajectories in patients with

type 2 diabetes mellitus. *Diabetes, Obesity and Metabolism* **2017**, *19*, 628–634, DOI: 10.1111/dom.12865, https://dom-pubs.pericles-prod.literatumonline.com/doi/10.1111/dom.12865 (cit. on p. 11).

[24] Bergenstal, R. M.; Beck, R. W.; Close, K. L.; Grunberger, G.; Sacks, D. B.; Kowalski, A.; Brown, A. S.; Heinemann, L.; Aleppo, G.; Ryan, D. B., et al. Glucose management indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. *Diabetes care* **2018**, *41*, 2275–2280, DOI: 10.2337/dc18-1581, https://diabetesjournals.org/care/article/41/11/2275/36593/Glucose-Management-Indicator-GMI-A-New-Term-for (cit. on p. 9).

[25] Jiang, Z.; Wang, J.; Zhao, P.; Zhang, L.; Guo, Y. In *Glycans and Glycosaminoglycans as Clinical Biomarkers and Therapeutics - Part A*, Zhang, L., Ed.; Progress in Molecular Biology and Translational Science, Vol. 162; Academic Press: 2019, pp 293–306, DOI: https://doi.org/10.1016/bs.pmbts.2019.01.008, https://www.sciencedirect.com/science/article/pii/S1877117319300195 (cit. on p. 8).

[26] Ngufor, C.; Van Houten, H.; Caffo, B. S.; Shah, N. D.; McCoy, R. G. Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c. *Journal of biomedical informatics* **2019**, *89*, 56–67, DOI: 10.1016/j.jbi.2018.09.001, https://www.sciencedirect.com/science/article/pii/S1532046418301758?via%3Dihub (cit. on p. 11).

[27] Hu, Y.; Shen, Y.; Yan, R.; Li, F.; Ding, B.; Wang, H.; Su, X.; Ma, J. Relationship between estimated glycosylated hemoglobin using flash glucose monitoring and actual measured glycosylated hemoglobin in a Chinese population. *Diabetes Therapy* **2020**, *11*, 2019–2027, DOI: 10.1007/s13300-020-00879-x, https://link.springer.com/article/10.1007/s13300-020-00879-x (cit. on p. 10).

[28] Weinstock, R. S.; Aleppo, G.; Bailey, T. S.; Bergenstal, R. M.; Fisher, W. A.; Greenwood, D. A.; Young, L. A. The role of blood glucose monitoring in diabetes management. **2020**, DOI: 10.2337/db2020-31, https://diabetesjournals.

org/compendia/article/doi/10.2337/db2020-31/144940/The-Role-of-Blood-Glucose-Monitoring-in-Diabetes (cit. on p. 12).

[29] Xu, Y.; Dunn, T. C.; Ajjan, R. A. A kinetic model for glucose levels and hemoglobin A1c provides a novel tool for individualized diabetes management. *Journal of Diabetes Science and Technology* **2021**, *15*, 294–302, DOI: 10.1177/1932296819897613, https://journals.sagepub.com/doi/10.1177/1932296819897613 (cit. on p. 10).

[30] Xu, Y.; Hirota, Y.; Ajjan, R. A.; Yamamoto, A.; Matsuoka, A.; Ogawa, W.; Dunn, T. C. Accurate prediction of HbA1c by continuous glucose monitoring using a kinetic model with patient-specific parameters for red blood cell lifespan and glucose uptake. *Diabetes and Vascular Disease Research* **2021**, *18*, 14791641211013734, DOI: 10.1177/14791641211013734, https://journals.sagepub.com/doi/10.1177/14791641211013734 (cit. on p. 11).

[31] Fang, M.; Wang, D.; Rooney, M. R.; Echouffo-Tcheugui, J. B.; Coresh, J.; Aurora, R. N.; Punjabi, N. M.; Selvin, E. Performance of the glucose management indicator (GMI) in type 2 diabetes. *Clinical chemistry* **2023**, *69*, 422–428, DOI: 10.1093/clinchem/hvac210, https://academic.oup.com/clinchem/article/69/4/422/7026054 (cit. on p. 9).

[32] Majumdar, S.; Kalamkar, S. D.; Dudhgaonkar, S.; Shelgikar, K. M.; Ghaskadbi, S.; Goel, P. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full (cit. on pp. 4, 7, 12).

[33] Tao, X.; Jiang, M.; Liu, Y.; Hu, Q.; Zhu, B.; Hu, J.; Guo, W.; Wu, X.; Xiong, Y.; Shi, X., et al. Predicting three-month fasting blood glucose and glycated hemoglobin changes in patients with type 2 diabetes mellitus based on multiple machine learning algorithms. *Scientific Reports* **2023**, *13*, 16437, DOI: 10.1038/s41598-023-43240-5, https://www.nature.com/articles/s41598-023-43240-5 (cit. on p. 11).

[34] Azcoitia, P.; Rodríguez-Castellano, R.; Saavedra, P.; Alberiche, M. P.; Marrero, D.; Wägner, A. M.; Ojeda, A.; Boronat, M. Age and red blood cell

parameters mainly explain the differences between HbA1c and glycemic management indicator among patients with type 1 diabetes using intermittent continuous glucose monitoring. *Journal of diabetes science and technology* **2024**, *18*, 1370–1376, DOI: 10.1177/19322968231191544, https://journals.sagepub.com/doi/10.1177/19322968231191544 (cit. on p. 9).

[35] Cichosz, S. L.; Kronborg, T.; Laugesen, E.; Hangaard, S.; Fleischer, J.; Hansen, T. K.; Jensen, M. H.; Poulsen, P. L.; Vestergaard, P. From Stability to Variability: Classification of Healthy Individuals, Prediabetes, and Type 2 Diabetes using Glycemic Variability Indices from Continuous Glucose Monitoring Data. *Diabetes Technology & Therapeutics* **2024**, DOI: 10.1089/dia.2024.0226, https://www.liebertpub.com/doi/10.1089/dia.2024.0226 (cit. on p. 11).

[36] Friedman, J. G.; Smith, E. P.; Awasty, S. S.; Behan, M.; Genco, M. T.; Hempel, H.; Jafri, S.; Jandarov, R.; Nagaraj, T.; Franco, R. S., et al. Primary care diabetes assessment when HbA1c and other measures of glycemia disagree. *Primary Care Diabetes* **2024**, *18*, 151–156, DOI: 10.1016/j.pcd.2023.12.005, https://www.sciencedirect.com/science/article/abs/pii/S1751991823002231?via%3Dihub (cit. on p. 8).

[37] Kurasawa, H.; Waki, K.; Seki, T.; Chiba, A.; Fujino, A.; Hayashi, K.; Nakahara, E.; Haga, T.; Noguchi, T.; Ohe, K. Enhancing Type 2 Diabetes Treatment Decisions With Interpretable Machine Learning Models for Predicting Hemoglobin A1c Changes: Machine Learning Model Development. *JMIR AI* **2024**, *3*, e56700, DOI: 10.2196/56700, http://www.ncbi.nlm.nih.gov/pubmed/39024008 (cit. on p. 11).

[38] Li, M.; Ge, S.; Shu, X.; Wu, X.; Liu, H.; Xu, A.; Ji, L. Interference of hemoglobin variants with HbA1c measurements: comparison of 6 commonly used HbA1c methods with the IFCC reference method. *Laboratory Medicine* **2024**, lmae034, DOI: 10.1093/labmed/lmae034, https://academic.oup.com/labmed/advance-article-abstract/doi/10.1093/labmed/lmae034/7683086?redirectedFrom=fulltext (cit. on p. 8).

[39] Manov, A. E.; Holt, N.; Dini, E.; Rivera, R.; Donepudi, A.; Haddadin, R.; Mefferd, K.; Qadir, I. The Discrepancy Between Hemoglobin A1c and Glucose Management Indicators in 26 Patients Treated With Continuous Glucose Monitoring in an Internal Medicine Residency Clinic. *Cureus* **2024**, *16*, DOI: 10.7759/cureus.56768, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11034459/ (cit. on p. 9).

[40] Mao, Y.; Zhong, W. HbA1c variability as an independent risk factor for microvascular complications in type 1 diabetes. *Journal of Diabetes Science and Technology* **2024**, *18*, 380–388, DOI: 10.1177/19322968221100833, https://journals.sagepub.com/doi/10.1177/19322968221100833 (cit. on p. 8).

[41] Qaraqe, M.; Elzein, A.; Belhaouari, S.; Ilam, M. S.; Petrovski, G. A novel few shot learning derived architecture for long-term HbA1c prediction. *Scientific Reports* **2024**, *14*, 482, DOI: 10.1038/s41598-023-50348-1, https://www.nature.com/articles/s41598-023-50348-1 (cit. on p. 11).

[42] Sadeghi, P.; Karimi, H.; Lavafian, A.; Rashedi, R.; Samieefar, N.; Shafiekhani, S.; Rezaei, N. Machine learning and artificial intelligence within pediatric autoimmune diseases: applications, challenges, future perspective. *Expert Review of Clinical Immunology* **2024**, DOI: 10.1080/1744666X.2024.2359019, https://www.tandfonline.com/doi/full/10.1080/1744666X.2024.2359019 (cit. on p. 11).

[43] Sato Imuro, S. E.; Sabharwal, A.; Bevier, W.; Kerr, D. Evaluating HbA1c-to-average glucose conversion with patient-specific kinetic models for diverse populations. *Scientific Reports* **2024**, *14*, 22098, DOI: 10.1038/s41598-024-72837-7, https://www.nature.com/articles/s41598-024-72837-7 (cit. on p. 11).

[44] Selvin, E. The glucose management indicator: time to change course? *Diabetes care* **2024**, *47*, 906–914, DOI: 10.2337/dci23-0086, https://diabetesjournals.org/care/article/47/6/906/154177/The-Glucose-Management-Indicator-Time-to-Change (cit. on p. 9).

[45] Song, J.; McNeany, J.; Wang, Y.; Daley, T.; Stecenko, A.; Kamaleswaran, R. Riemannian manifold-based geometric clustering of continuous glucose monitoring to improve personalized diabetes management. *Computers in Biology and Medicine* **2024**, *183*, 109255, DOI: 10.1016/j.compbiomed.2024.109255, https://www.sciencedirect.com/science/article/abs/pii/S0010482524013404?via%3Dihub (cit. on p. 11).

Chapter 2

Evaluation of HbA1c from CGM traces in an Indian T2D population

Published as and adapted from:
Majumdar, S. et al. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full

2.1 Introduction

Type 2 diabetes (T2D) is one of the most common metabolic disorders in India. Understanding the metabolic pathways and mechanisms involved in the development of T2D in a patient plays an important role in its diagnosis and eventually its treatment. Traditionally it involved measuring fasting blood glucose concentration (FBG) and post-prandial blood glucose concentration (PPBG). Since the late 1970s there have been reports of a correlation between Glycated Hemoglobin (HbA1c) and blood glucose concentration (BG) and that HbA1c could be a useful tool for long term BG control. Gabbay et al. [1] studied the correlation between HbA_{1a}, HbA_{1b} and HbA1c with 24-hr urinary glucose concentration collected over a period of 1, 2 and 3 months for 220 diabetic patients and suggested that glycosylated hemoglobin could act as a good index for long-term BG levels in diabetics. Santiago et al. [3] further studied the coorelation between HbA1c and PPBG. Clarke et al. [4] showed that HbA1c is coorelated with average blood glucose concentration (aBG) over 2 months and therefore is a good index for aBG and is a useful tool for understanding the quality of BG control of a patient. Lecomte et al. [6] also confirmed in a group of 138 patients that HbA1c is a good index for BG control. Distiller [5] compaired the efficacy of PPBG and HbA1c as an index for BG control and also showed that HbA1c is a significantly better index.

There are a plethora of other new matrics that are being developed for understanding the glycemic state of the patient like time in range (TiR), etc. However, HbA1c is one of the most reliable metric for understanding the long-term BG changes in a patient. Therefore, accurate experimental methods [2] have been developed to measure HbA1c. However, with the development of fast glucose monitoring (FGM) and eventually continuous glucose monitoring (CGM) technologies, clinicians can now have access to a large number of consecutive interstitial fluid glucose concentration (ISF) measurements (CGM traces). This encouraged the development of methods for estimating metrics like FGM, PPBG, TiR, HbA1c, etc from the CGM traces.

Sikaris [9] showed that although for a single measurement HbA1c and BG had been shown to be correlated, including multiple measurements like CGM traces improved the correlation further. They concluded that not only estimating HbA1c

from the HbA1c–BG relation was viable but also that it would become the standard method of estimating HbA1c. Mazze [7] also showed that BG from self-monitoring blood glucose (SMBG) and CGM traces were highly correlated and that the HbA1c estimates obtained using them were not significantly different, although different patterns of SMBG and CGM traces could produce the same HbA1c. This suggested that HbA1c was a metric that could be estimated reliably. Nathan et al. [8] using linear regression provided a method of estimating the aBG from the HbA1c at the population level. This relationship has been used to form a method of estimating HbA1c from aBG. This method has been adopted as the standard for obtaining HbA1c estimates [11]. This method has also been used to estimate the HbA1c values by the Abbot Libre FreeStyle Pro device for the CGM report generated by the device.

We show that the mean of the estimates provided by the standard Nathan et al. [8] method for HbA1c, at a population level, are not statistically reliable with respect to the experimental HbA1c values for an Indian population. We then provide two methods, based on ordinary least squares (OLS) linear regression, for estimating HbA1c. The first method estimates HbA1c from the calculated average interstitial fluid glucose concentration (aISF) from any given CGM trace, whereas the second method estimates the HbA1c from the aISF by first estimating aBG and then HbA1c using the Nathan model.

The mean HbA1c estimates provided by these new methods for a small test dataset of 32 CGM traces are however statistically indistinguishable for clinical epidemiology from the mean of the experimentally measured HbA1c values. We provide the 95 % confidence interval for the two models and also the 95 % prediction intervals for the HbA1c estimates of these two models, these results are visualized in two figures, Fig 2.1 and Fig 2.2, which show the three models for estimating HbA1c along with the 95 % confidence interval (Fig 2.1) and the 95 % prediction interval (Fig 2.2).

2.2 Materials and methods

2.2.1 Subject recruitment and measurement of blood biochemical parameters

A CGM dataset containing the traces of 50 participants was collected at the Health Center, Savitribai Phule Pune University, Pune. For each participant, a FreeStyle Libre Pro CGM sensor (Abbott, UK) was inserted subcutaneously on the back of the upper arm by Dr. Shashikant Dudhgaonkar at the Health Center, Savitribai Phule Pune University, Pune, India, between July 2021 and Aug 2021. This was a factory-calibrated glucose sensor that recorded subcutaneous ISF every 15 min for 14 days. All participants were advised to continue their normal routine of diet and exercise. On the fourteenth day, the CGM device was removed and the data was downloaded and analyzed with the FreeStyle Libre Pro software. We will refer to this dataset as the Pune-2021 dataset.

The CGM data collected in the Pune-2021 dataset were then filtered according to the number of days the sensor was active and the percentage of measurements it recorded, as suggested by Danne et al. [11]. The dataset was filtered into two sets, the data sufficient Pune-2021 dataset and the data insufficient Pune-2021 dataset. From the measurement ID provided in the CGM trace data file, the number of ISF measurements, N, that the device recorded was calculated. The timestamps provided in the CGM trace were used to calculate the effective number of days, n_d for which the CGM device was active. This number was however rounded off to the nearest integer using the `round` function provided by the `NumPy` package [12]. The percentage of measurements the device recorded was calculated using Δt_m, the time difference in seconds between the first and the last reading. Δt_m was used to calculate the total number of readings that the device would have taken, $N_{\text{total}} = \lfloor \frac{\Delta t_m}{60 \times 15} \rfloor + 1$. The percentage of measurements was calculated as $N_p = 100 \times N/N_{\text{total}}$. If $n_d \geq 14$ and $N_p \geq 70\,\%$ for a given CGM trace, as suggested by [11], we categorized the CGM trace to be data-sufficient otherwise it was categorized as data-insufficient. After the data-insufficient CGM traces were filtered out 12 pre-diabetic, 13 diabetic, and 14 non-diabetic CGM traces were left and were categorized as data-sufficient.

A second dataset of 78 CGM traces along with their HbA1c (by HPLC) was collected by Dr. K. M. Shelgikar at the Joshi Hospital, Shivaji Nagar, Pune from 2018 to 2019, and anonymized for analysis. This dataset will be referred to as the Joshi-2018 dataset. Similar to the Pune-2021 dataset, the Joshi-2018 dataset was also filtered as data-sufficient and data-insufficient. After filtering out the data-insufficient CGM traces from the Joshi-2018 dataset, only 73 CGM traces could be considered as data-sufficient.

The complete CGM-dataset dataset, including both the Pune-2021 dataset and the Joshi-2018 dataset, contained the CGM traces of 128 participants, 15 of whom were pre-diabetic, 94 were diabetic and, 19 were non-diabetic. The data-sufficient subset of the CGM-dataset however contained 112 CGM traces, out of which 12 were pre-diabetic, 86 were diabetic, and 14 were non-diabetic. A sample of 32 CGM traces and their corresponding HbA1c measurements were separated as a test set for validation purposes; the remaining 80 data-sufficient CGM traces were grouped as the training CGM-dataset. The complete CGM-dataset including the data-insufficient CGM traces were used to validate the HbA1c estimates obtained using the Nathan model [8] but only the data-sufficient CGM traces of the training CGM-dataset were used for constructing our models, which were then validated using the data-sufficient CGM traces of the test CGM-dataset.

2.2.2 Comparing Nathan HbA1c estimates with experimentally measured HbA1c

Nathan et al. [8] collected a dataset of 2700 glucose measurements from 268 type 1 diabetes (T1D) patients, 159 T2D patients and 80 non-diabetic participants. Their dataset contained CGM traces as well as finger-stick measurements which were collected as different measures of glycemia.

Note that the ISF measurements were scaled by a factor of 1.05 to estimate the corresponding BG. aBG was calculated by taking the weighted average of all the blood glucose concentration measurements collected. All measurements in a day were given equal weight which was inversely proportional to the number of measurements taken on that day. The aBG was calculated by taking the mean of all the measurements, giving the measurements of each day an equal weight. The

expression to obtain the aBG is

$$\text{aBG} = \frac{1}{(m_1 + m_2)} \left\{ \sum_{i=1}^{i=m_1} \left(\frac{1}{n_{1,i}} \right) \text{BG}_\text{i} + \sum_{i=1}^{i=m_2} \left(\frac{1}{n_{2,i}} \right) (1.05 \times \text{ISF}_\text{i}) \right\}, \quad (1)$$

where aBG represents the average blood glucose concentration, BG_i the i^{th} SMBG measurements, ISF_i the i^{th} CGM measurement, m_1 and m_2 are the number of SMBG and CGM measurements respectively, $n_{1,i}$ the number of SMBG measurements taken on the day BG_i was taken and, similarly $n_{2,i}$ the number of CGM measurements taken on the day ISF_i was taken.

A linear regression analysis was performed by Nathan et al. [8] taking the calculated aBG as the dependent variable and the HbA1c as the independent variable and obtained this relation

$$\text{aBG}_{\text{mg/dL}} = 28.7 \times \text{HbA}_{\text{1c, \%}} - 46.7, \quad (2)$$

(2) could also be written as

$$\text{HbA}_{\text{1c,\%}} = \frac{1}{28.7} \left\{ 46.7 + \left(1.05 \times \text{aISF}_{\text{mg/dL}} \right) \right\}, \quad (3)$$

$$= 1.627 + 0.034 \times \left(1.05 \times \text{aISF}_{\text{mg/dL}} \right), \quad (4)$$

$$\text{HbA}_{\text{1c,mmol/mol}} = 0.380 \times \left(1.05 \times \text{aISF}_{\text{mg/dL}} \right) - 5.60, \quad (5)$$

We used a paired t-test to verify if the two samples, that is, the experimentally measured HbA1c from the CGM-dataset and the estimated HbA1c calculated using the Nathan model, (5) and (1), have a statistically indistinguishable mean. Calculations were carried out using the `ttest_rel` function from the `stats` module of the `SciPy` package [13]. A similarly paired t-test was also performed with the data-sufficient CGM traces, including both the testing and training CGM traces from the CGM-dataset.

2.2.3 Direct Model

To construct a model between aISF and HbA1c directly, we assume a linear relationship and carry out a regression analysis. Note that the aISF here is an equally

weighted average of all the ISF measurements in a given CGM trace,

$$\text{aISF}_{\text{mg/dL}} = \frac{1}{N}\sum_{i=1}^{i=N} \text{ISF}_{\text{i, mg/dL}}, \tag{6}$$

where $\text{aISF}_{\text{mg/dL}}$ represents the calculated aISF in mg/dL, $\text{ISF}_{\text{i, mg/dL}}$ represents the i^{th} ISF measurement from the given CGM trace in mg/dL and N represents the total number of measurements in the given CGM trace. The linear regression equation for the direct model is

$$\text{HbA}_{\text{1c,mmol/mol}} = \beta_1 \times \text{aISF}_{\text{mg/dL}} + \beta_0, \tag{7}$$

where $\text{aISF}_{\text{mg/dL}}$ represents the aISF in mg/dL, $\text{HbA}_{\text{1c, mmol/mol}}$ represents the HbA1c in mmol/mol, β_1 the slope in $\text{mmol}\,\text{dL}/(\text{mol}\,\text{mg})$ and, β_0 the intercept in mmol/mol.

We obtained the OLS estimates $\hat{\beta}_0$ and, $\hat{\beta}_1$ of the parameters β_0 and, β_1. We also calculate the 95 % confidence interval for $\hat{\beta}_0$ and $\hat{\beta}_1$ along with the 95 % confidence interval and the 95 % prediction interval of HbA_{1c} for any given aISF. This analysis was performed using the `LinearRegression` function from the `linear_model` module of the `scikit-learn` package [10]. The confidence and prediction intervals were however calculated using the standard OLS solution formulae.

A paired t-test was then performed on the HbA1c estimated using the direct model and the experimental values for the data-sufficient CGM traces from the training CGM-dataset to verify if the mean of the HbA1c estimates obtained using $\hat{\beta}_0$ and $\hat{\beta}_1$ were statistically indistinguishable from the mean experimental HbA1c value at a population level. A similar test was performed using the HbA1c estimates provided by the direct model and the experimental HbA1c corresponding to the CGM traces from the test CGM-dataset. The t-tests were performed using the `ttest_rel` function from the `stats` module of the `SciPy` package.

We also used the training dataset of CGM traces and calculated the 5-fold cross validation root mean squared error (RMSE) so as to validate the 95 % confidence interval for the direct model.

2.2.4 Adapted Nathan Model

In Section 2.2.3 we constructed a linear model for estimating HbA1c from aISF calculated from a given CGM trace. Although such a relationship, if reliable, can be valuable it requires us to base our HbA1c estimates on ISF values. Traditionally, however, for the diagnosis of T2D and for the analysis of the glycemic state of an individual, various metrics like FBG, PPBG and HbA1c have always been based on BG. The current CGM devices however report ISF readings and therefore to use these CGM traces with our current diagnostic methods it is important that we develop a reliable method of converting the ISF readings to their corresponding BG readings. The ISF measurements in CGM traces of the dataset used by Nathan et al. [8] were scaled to their BG values by a scaling factor of 1.05. We suspected obtaining a better estimate of this scaling factor would improve our HbA1c estimates.

We, therefore, constructed a linear model for estimating HbA1c from the calculated aISF (aISF is constructed using (6)) via aBG. We considered a model where we estimated aBG by scaling aISF by a factor of ω and used (2) to obtain the estimate of HbA1c. This model represented by (8) is

$$\mathrm{HbA}_{1c,\ \mathrm{mmol/mol}} = 0.380 \times \left(\omega \times \mathrm{aISF}_{\mathrm{mg/dL}}\right) - 5.60, \tag{8}$$

where $\mathrm{aISF}_{\mathrm{mg/dL}}$ represents the aISF in mg/dL, $\mathrm{HbA}_{1c,\ \mathrm{mmol/mol}}$ represents the HbA1c in mmol/mol and, ω the scaling factor. Now, (8) could also be written as

$$\frac{\mathrm{HbA}_{1c,\ \mathrm{mmol/mol}} + 5.60}{0.380} = \left(\omega \times \mathrm{aISF}_{\mathrm{mg/dL}}\right), \tag{9}$$

The OLS solutions for the estimate $\hat{\omega}$, of the coefficient ω are the same for both (8) and (9).

We obtain the OLS estimate $\hat{\omega}$, using (9) where we take the calculated $\mathrm{aISF}_{\mathrm{mg/dL}}$ as the independent variable and the transformed experimental HbA1c values, $\frac{\mathrm{HbA}_{1c,\ \mathrm{mmol/mol}}+5.60}{0.372}$, as the dependent variable. The analysis was performed using the `LinearRegression` function from the `linear_model` module of the `scikit-learn` package. The data-sufficient CGM traces and their corresponding HbA1c values from the training CGM-dataset was used for this analysis. We calculated the 95 % confidence interval for $\hat{\omega}$, the 95 % confidence interval and 95 % prediction

interval for estimated HbA1c corresponding to an aISF calculated from any CGM trace. These intervals were calculated using the standard OLS solution formulae for constrained linear regression.

A paired t-test was performed with the HbA1c estimates made using the adapted Nathan model and the experimentally measured HbA1c values for the data-sufficient CGM traces of the training CGM-dataset, using the `ttest_rel` function from the `stats` module of the `SciPy` package to confirm that the mean HbA1c estimates from the adapted Nathan model are not significantly different from the mean experimental HbA1c values at a population level. This was repeated on the data-sufficient CGM traces of the test CGM-dataset. Finally, using the training CGM-datset a 5-fold cross validation RMSE was calculated for the adapted Nathan model to validate the reliability of the 95 % confidence intervals of the adapted Nathan model.

2.3 Results

2.3.1 Coefficient estimates for the direct model

The paired t-test performed using the CGM traces of the complete CGM-dataset between the experimental HbA1c values and the corresponding Nathan HbA1c estimates calculated using (1) and (5) generated a p-value < 0.001. A similar paired t-test with the data-sufficient CGM traces from the CGM-dataset, also generated a p-value < 0.001. Considering $\alpha = 0.05$ this indicated that the mean Nathan model estimated HbA1c values, both for data-sufficient and data-insufficient CGM traces, were statistically different from the corresponding mean of the experimentally measured HbA1c values for the Indian population.

This lead us to construct the direct model for estimating HbA1c from the aISF. We performed a linear regression analysis to establish a relationship between aISF and HbA1c using the data-sufficient CGM traces of the training CGM-dataset. We obtained an estimate of the coefficient β_0 of the model (7), $\hat{\beta}_0 = 16.56\,\text{mmol/mol}$, with a 95 % confidence interval of $(8.96\,\text{mmol/mol}, 24.16\,\text{mmol/mol})$ and an estimate for β_1, $\hat{\beta}_1 = 0.319\,\text{mmol}\,\text{dL}/(\text{mol}\,\text{mg})$ with a 95 % confidence interval of $(0.273\,\text{mmol}\,\text{dL}/(\text{mol}\,\text{mg}), 0.365\,\text{mmol}\,\text{dL}/(\text{mol}\,\text{mg}))$. The analysis yielded (10),

with an $R^2 = 0.71$ and a $p - value < 0.01$. Therefore the direct model, (7), with the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ is given by (10) bellow

$$\text{HbA}_{1c,\text{mmol/mol}} = 0.319 \times \text{aISF}_{\text{mg/dL}} + 16.56, \tag{10}$$

where $\text{aISF}_{\text{mg/dL}}$ represents the aISF in mg/dL and $\text{HbA}_{1c,\ \text{mmol/mol}}$ represents the HbA1c in mmol/mol. Figure 2.1, shows (10) as the black dashed line along with the 95 % confidence interval for $\text{HbA}_{1c\text{mmol/mol}}$ corresponding to any $\text{aISF}_{\text{mg/dL}}$ calculated from a given CGM trace. The formulae for obtaining the 95 % confidence and prediction interval for any HbA1c estimate is given in the Supplementary Material.

A paired t-test performed between the HbA1c values estimated using (10) and their corresponding experimental values from the training CGM-dataset generated a p-value of 1.0 indicating that at a population level, the mean HbA1c estimate is statistically (for $\alpha = 0.05$) indistinguishable from the mean experimental value.

2.3.2 Coefficient estimates of the adapted Nathan model

We constructed a direct model, given by (10), for estimating HbA1c from aISF for any given CGM trace. We however suspected that a model described by (8) where, HbA1c is estimated from a scaled aISF value using (2) would provide better estimates of HbA1c. The OLS solution for a linear regression analysis using (9) while keeping the intercept zero, would provide us with an estimate of the scaling factor for obtaining aBG from aISF.

The estimate for the scaling factor for (9), $\hat{\omega}$, obtained using the `LinearRegression` function of the `scykit-learn` package with the intercept set to zero, on the data-sufficient training CGM-dataset, is $\hat{\omega} = 1.17$, with a 95 % confidence interval $(1.12, 1.22)$. The analysis yielded an $R^2 = 0.58$ and, a $p - value < 0.01$. The estimate, $\hat{\omega}$ obtained using the analytical solution for obtaining the OLS estimate of ω from (8) generated an identical result. The equation for obtaining HbA1c estimates using the adapted Nathan model is represented by (11) below

$$\text{HbA}_{1c,\ \text{mmol/dL}} = 0.380 \times \left(1.17 \times \text{aISF}_{\text{mg/dL}}\right) - 5.60, \tag{11}$$

where $\text{aISF}_{\text{mg/dL}}$ represents the aISF in mg/dL and $\text{HbA}_{\text{1c, mmol/mol}}$ represents the HbA1c in mmol/mol. Figure 2.1, shows (11) as a solid black line, along with the 95 % confidence interval for the $\text{HbA}_{\text{1cmmol/mol}}$ corresponding to any $\text{aISF}_{\text{mg/dL}}$. The formulae for obtaining the confidence and prediction interval for any HbA1c values estimated using (11) is provided in the Supplementary Material.

A paired t-test performed with the experimentally measured HbA1c values as one sample and the HbA1c values obtained as estimates using the adapted Nathan model, (11), from the corresponding CGM traces of the training CGM-dataset generated a p-value of 0.075. This showed that the mean HbA1c estimate at a population level was statistically (for $\alpha = 0.05$) indistinguishable from the mean experimental HbA1c.

2.3.3 Validation of the direct and adapted models

In Section 2.3.1 and Section 2.3.2 we constructed two models for estimating HbA1c from the aISF calculated from any given CGM trace. We then constructed the 95 % prediction intervals for the HbA1c estimates calculated by the models. The formulae for constructing the prediction interval corresponding to any estimated HbA1c for both the models are provided in the Supplementary Material.

A paired t-test performed between the experimentally measured HbA1c from the test CGM-dataset and the HbA1c estimates obtained from their corresponding CGM trace using the direct model generates a p-value of 0.702 and using the adapted Nathan model it generates a p-value of 0.712. This indicates that at a population level, the mean HbA1c estimates for an independent sample of CGM traces are statistically (taking $\alpha = 0.05$) indistinguishable from the mean experimental HbA1c. The 5-fold cross validation root mean squared error (RMSE) for the HbA1c estimates obtained using the direct model on the training CGM-datasets is 12.3 mmol/mol and for the estimates obtained using the adapted Nathan model it is 14.6 mmol/mol. The mean 95 % confidence interval for the direct model is 9.87 mmol/mol and for the adapted Nathan model is 7.92 mmol/mol.

A t-test performed using the Nathan model HbA1c estimates for the test CGM-dataset generated a p-value < 0.001, indicating that at a population level the mean Nathan model HbA1c estimates are statistically (taking $\alpha = 0.05$) different from

the mean experimental value.

2.4 Discussion

The development of CGM technology provides us with a large sequence of glucose concentration measurements. This provides us with a great opportunity to study the glucose dynamics and the glycemic state of an individual. The current CGM devices however only provide us with ISF measurements, while traditionally it has been the norm to study glucose dynamics with BG measurements. Therefore, a large bulk of our understanding of glucose dynamics, glycemic states, and metabolic diseases like diabetes are based on the BG values. To use the CGM traces provided by these devices it becomes important that we are able to estimate the corresponding BG and other metrics like HbA1c from any given CGM trace reliably.

The direct method for estimating HbA1c from the ISF values, as described in Section 2.2.3, provides an estimate along with a 95 % confidence and prediction interval for the estimate given an aISF value. The mean HbA1c estimates provided by the direct model were statistically indistinguishable from the mean experimental HbA1c measurement for the data-sufficient test CGM-dataset. The prediction intervals were however in the order of 50 mmol/mol. These intervals are quite wide which raises the question, can clinically reliable HbA1c estimates be obtained using only the aISF values calculated from a CGM trace? or does it require the inclusion of some other metric or property of the CGM to obtain a clinically significant narrow estimate of HbA1c? Therefore, it remains an open question that although ISF and BG are highly correlated with HbA1c; why is it that such models are unable to provide clinically reliable estimates of HbA1c from aISF or aBG values alone?

We suspected that the inclusion of an improved method of estimating BG from ISF could improve the estimates provided by (5). Therefore in Section 2.3.2, we constructed a new linear model for estimating BG from ISF using linear regression. The mean HbA1c estimates provided by this method were indistinguishable from the mean experimentally measured HbA1c values of the data-sufficient test

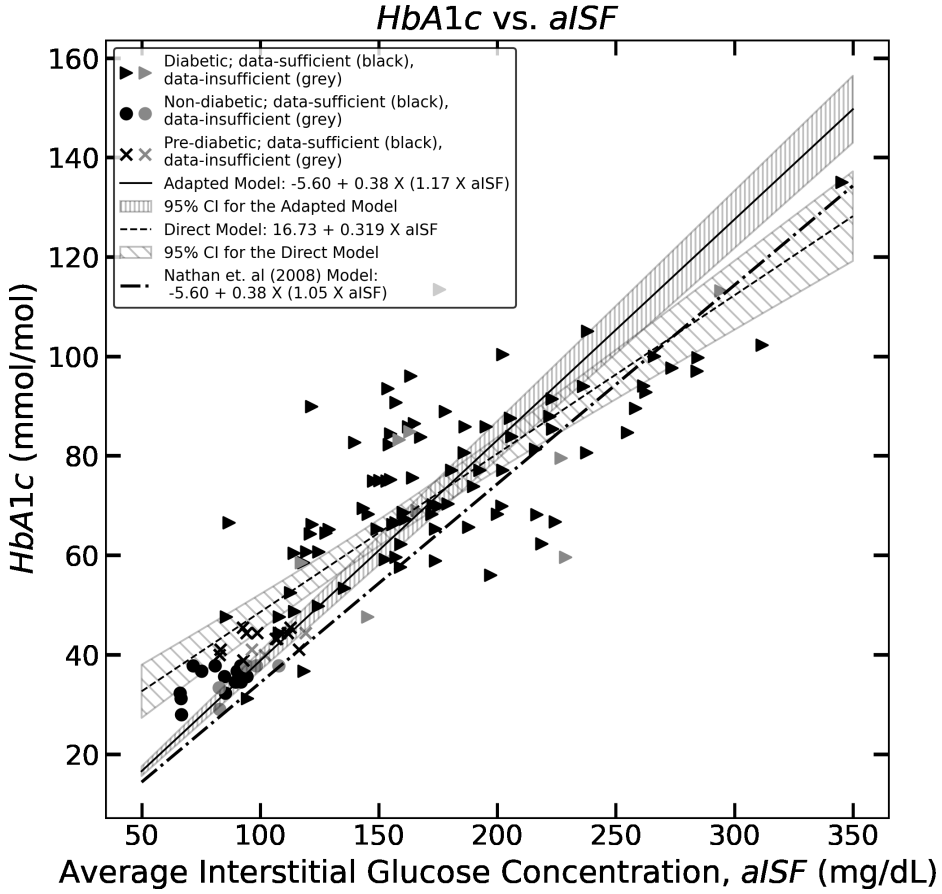


Figure 2.1: The Figure represents the experimentally measured HbA_{1c} and the aISF (calculated as described in Section 2.2.3) values of the CGM-dataset. The pre-diabetic participants are represented by crosses, the diabetic participants by solid triangles and, the non-diabetic participants by solid circles. The scatter points representing participants with a data-sufficient CGM trace (according to Danne et al. [11]) are colored black, while the participants with data-insufficient CGM trace are colored grey. The solid black line and the corresponding hatched region represent (11) and it is 95 % confidence interval, the dashed line along with its corresponding hatched region represents (10) and it's corresponding 95 % confidence interval. The dot-dashed line represents (5)

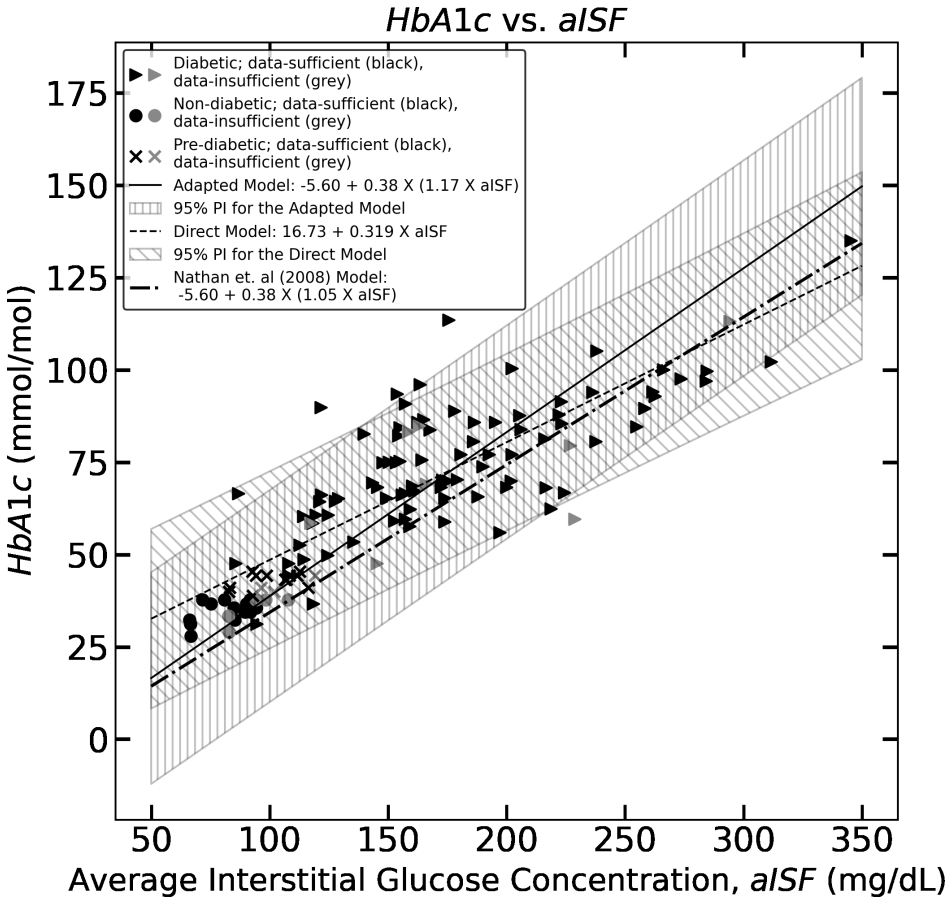


Figure 2.2: The Figure represents the experimentally measured HbA_{1c} and the aISF (calculated as described in Section 2.2.3) values of the CGM-dataset. The pre-diabetic participants are represented by crosses, the diabetic participants by solid triangles and, the non-diabetic participants by solid circles. The scatter points representing participants with a data-sufficient CGM trace (according to Danne et al. [11]) are colored black, while the participants with data-insufficient CGM trace are colored grey. The solid black line and the corresponding hatched region represent (11) and it is 95 % prediction interval, the dashed line along with its corresponding hatched region represents (10) and it's corresponding 95 % prediction interval. The dot-dashed line represents (5)

CGM-dataset. However, the 95 % prediction interval, in this case, was in the order of 59 mmol/mol. We also showed that the mean HbA1c estimates obtained using these two models, the direct model and the adapted Nathan model, are not statistically different from the mean experimental HbA1c at the population level. However, the mean experimental HbA1c were significantly different from the mean estimates provided by the Nathan model at the population level.

From the analysis of the model performances on the test CGM-dataset, we can conclude that although our models for estimating HbA1c provide a wide 95 % prediction interval, which includes the HbA1c estimates obtained using the Nathan model, the mean HbA1c estimates provided by our models at a population level are statistically indistinguishable from the mean experimental HbA1c values, unlike the HbA1c estimates obtained using the Nathan model. This shows that at a population level, the direct and the adapted Nathan model can provide a reliable HbA1c estimate for clinical epidemiological studies as compared to the Nathan model and that none of the models are able to provide HbA1c estimates at an individual level with a clinically reliable confidence and prediction interval. It is hoped that future studies will test this hypothesis with greater statistical power so as to create a model with a reliable prediction interval that is useful for predicting the HbA1c of an individual clinically.

Conflict of Interest Statement

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.

Ethics statement

The study protocol for the Pune-2021 dataset was approved by the Institutional Ethical Committee (IEC) of Savitribai Phule Pune University, Pune, India (SPPU/IEC/2020/102). The study protocol for the Joshi-2018 dataset was approved by the IEC of Maharashtra Medical Research Society (ECR/311/Inst/MH/2013/RR-19; Dated 14

March 2023). All participants were provided with the informed consent form prior to their participation in the studies. The data analysis was reviewed and approved by the Institutional Human Ethics Committee (IHEC) of IISER Pune (IHEC/Admin/2021/015).

SSG and PG designed the study and contributed to its implementation. SD supervised the Pune-2021 dataset's data collection and helped in the recruitment of study participants. SK collected the samples, designed and performed the biochemical assays for the Pune-2021 dataset. KMS collected Joshi-2018 dataset. SM analyzed the data. SM, SK, SSG and PG wrote the manuscript.

Funding

This work was financially supported by Rastriya Ucchattar Shiksha Abhiyan (RUSA), Savitribai Phule Pune University, Pune. SK is a recipient of a Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), India. SM is a Junior Research Fellow supported by a DST INSPIRE Fellowship, Department of Science and Technology, Government of India.

Acknowledgments

We thank all the study subjects for their wholehearted participation in the study. We acknowledge the financial help from all the funding agencies mentioned.

Data availability statement

The datasets generated during and/or analyzed during the current study are not publicly available but may be available from the authors upon reasonable request.

References

[1] Gabbay, K. H.; Hasty, K.; Breslow, J. L.; Ellison, R. C.; Bunn, H. F.; Gallop, P. M. Glycosylated hemoglobins and long-term blood glucose control in di-

abetes mellitus. *The Journal of Clinical Endocrinology & Metabolism* **1977**, *44*, 859–864, DOI: 10.1210/jcem-44-5-859, https://academic.oup.com/jcem/article-abstract/44/5/859/2677214 (cit. on p. 23).

[2] Davis, J.; Mcdonald, J. M.; Jarett, L. A high-performance liquid chromatography method for hemoglobin A1c. *Diabetes* **1978**, *27*, 102–107, DOI: 10.2337/diab.27.2.102, https://diabetesjournals.org/diabetes/article/27/2/102/5549/A-High-performance-Liquid-Chromatography-Method (cit. on p. 23).

[3] Santiago, J. V.; Davis, J.; Fisher, F. Hemoglobin A1c levels in a diabetes detection program. *The Journal of Clinical Endocrinology & Metabolism* **1978**, *47*, 578–580, DOI: 10.1210/jcem-47-3-578, https://academic.oup.com/jcem/article-abstract/47/3/578/2678805?redirectedFrom=fulltext (cit. on p. 23).

[4] Clarke, J.; Passa, P.; Canivet, J. HbA1c: need of its dosage in diabetics (author's transl). *La Nouvelle Presse Medicale* **1979**, *8*, 513–517, https://europepmc.org/article/med/461162 (cit. on p. 23).

[5] Distiller LA & Zail, S. The use of glycosylated haemoglobin measurements in the control of the diabetic patient. *South African Medical Journal* **1979**, *55*, 335–337 (cit. on p. 23).

[6] Lecomte, M.; Schoos, R.; Schoos-Barbette, S.; Luyckx, A.; Lambotte, C.; Lefebvre, P. Hemoglobin A1c and diabetes control (author's transl). *Diabete & Metabolisme* **1979**, *5*, 57–61, https://europepmc.org/article/med/446835 (cit. on p. 23).

[7] Mazze, R. The future of self-monitored blood glucose: mean blood glucose versus glycosylated hemoglobin. *Diabetes technology & therapeutics* **2008**, *10*, S–93, DOI: 10.1089/dia.2008.0006, https://www.liebertpub.com/doi/abs/10.1089/dia.2008.0006 (cit. on p. 24).

[8] Nathan, D. M.; Kuenen, J.; Borg, R.; Zheng, H.; Schoenfeld, D.; Heine, R. J.; Group, A.-D. A. G. (S. Translating the A1C assay into estimated average glucose values. *Diabetes care* **2008**, *31*, 1473–1478, DOI: 10.2337/dc08-0545, https://diabetesjournals.org/care/article/31/8/1473/

28589/Translating-the-A1C-Assay-Into-Estimated-Average (cit. on pp. 24, 26 sq., 29).

[9] Sikaris, K. The Correlation of Hemoglobin A1c to Blood Glucose. *Journal of Diabetes Science and Technology* **2009**, *3*, PMID: 20144279, 429–438, DOI: 10.1177/193229680900300305, https://doi.org/10.1177/193229680900300305 (cit. on p. 23).

[10] Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830 (cit. on p. 28).

[11] Danne, T.; Nimri, R.; Battelino, T.; Bergenstal, R. M.; Close, K. L.; DeVries, J. H.; Garg, S.; Heinemann, L.; Hirsch, I.; Amiel, S. A., et al. International consensus on use of continuous glucose monitoring. *Diabetes care* **2017**, *40*, 1631–1640, DOI: 10.2337/dc17-1600, https://diabetesjournals.org/care/article/40/12/1631/37000/International-Consensus-on-Use-of-Continuous (cit. on pp. 24 sq., 34 sq.).

[12] Harris, C. R. et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362, DOI: 10.1038/s41586-020-2649-2, https://doi.org/10.1038/s41586-020-2649-2 (cit. on p. 25).

[13] Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272, DOI: 10.1038/s41592-019-0686-2 (cit. on p. 27).

[14] Majumdar, S.; Kalamkar, S. D.; Dudhgaonkar, S.; Shelgikar, K. M.; Ghaskadbi, S.; Goel, P. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full (cit. on p. 22).

Chapter 3

Evaluation of HbA1c from CGM traces in an Indian T1D population

3.1 Introduction

Glycated Hemoglobin (HbA1c) can be considered as a form of hemoglobin, where a glucose molecule is chemically attached to it's beta chain. Glycated Hemoglobin (HbA1c) levels in a human body are usually represented as a fraction in percentage as per the Diabetes Control and Complications Trial (DCCT) and as a fraction in mols as per the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) units. The HbA1c value is important in the diagnosis and management of type 2 diabetes (T2D) and type 1 diabetes (T1D). It reflects the glycemic state of the patient over multiple days (120 days approximately). It has become a metric which is regularly used in the clinical management of diabetes [8, 9, 17]. HbA1c is measured using various methods, most commonly high-performance liquid chromatography (HPLC). These methods however require a significant amount of blood to be drawn from the patient, which can be painful, and moreover in certain parts of the world these methods can be quite expensive. Therefore, estimating HbA1c from various other metrics, like fasting blood glucose concentration (FBG) or average blood glucose concentration (aBG), which

might have have already been measured, is an active field of research. A clinically reliable method of HbA1c estimation from parameters that have already been measured may also prevent unnecessary blood-draws and could be substantially less expensive.

Self-monitoring blood glucose (SMBG) is important for managing diabetes mellitus (DM). This involves the patient manually measuring and recording his/her own blood glucose concentration (BG) multiple times a day using finger-stick glucometers. However, as it requires multiple finger stick measurements throughout the day, one may still miss nocturnal and asymptomatic hypoglycemia [2] events. On the other hand, continuous glucose monitoring (CGM) devices have been under development since the early 1990s, measures interstitial fluid glucose concentration (ISF) every 10 seconds and then the average ISF value is calculated and saved every 15 min. Therefore, a CGM device can regularly monitor the BG (usually via the ISF) and record almost all hypoglycemia events, hypoglycemia and hyperglycemia episodes, time in ranges (TiRs), etc. In recent times, it has become a widely used device as along with the previous mentioned metric, it also provides a reliable discrete-time ISF trace of an individual over a long time-interval.

The synchronization of CGM devices with insulin pumps and to health management apps in the cloud makes it possible for patients/healthcare workers to obtain long traces of patient ISF and use it as an additional tool for diagnosis. Further, the use of insulin pumps (which are battery operated insulin-delivering devices) has also increased significantly; insulin pumps are generally paired with CGM devices which measure the ISF at regular intervals and provide alarms and can even change the insulin dose as per requirement.

The CGM reports generated by the CGM devices also provide an estimate of the HbA1c. The estimated HbA1c value is based on the average interstitial fluid glucose concentration (aISF) and is believed to be close to the experimentally measured HbA1c. However, some reports suggest that the estimated HbA1c differs from the experimentally measured HbA1c, especially with the Type of diabetes (T1D vs T2D) and also possibly with ethnicity of the patients. It would be immensely useful if the HbA1c of patients could be estimated from this CGM trace with sufficient reliability. This is especially important in resource limited settings

where the additional expense of estimating HbA1c may be avoided, and in children with T1D where an additional blood test should be avoided and when patients are using insulin pumps. Therefore, a robust method for estimating HbA1c from any CGM with robust reliability is critical and hence has been a very active area of research.

Over the years, there have been various attempts to construct a model that would allow us to reliably estimate HbA1c. Mortensen and Vølund [1] constructed the first model describing the relationship between HbA1c, HbA1d and, BG. This kinetic model was used to predict HbA1c and HbA1d as a function of the BG. This was most likely one of the first kinetic models for estimating HbA1c. This model had been implemented in two forms a *simplified* and an *improved* form. The estimated HbA1c varied similar to the measured HbA1c but the measured HbA1c values were 10 % to 20 % elevated as compared to the predicted HbA1c values. The next significant model was published by [3]. This model was established on the strong corelation between aBG and HbA1c that had been established over the years (cite here [4]). [3] actually describes a method to estimate aBG from a patients HbA1c value. The model was established using a linear regression analysis where the measured HbA1c was taken as the independent and the calculated aBG as the dependent variable. This method involved calculating an aBG from various SMBG finger-stick measurements and an aISF from CGM traces. The aISF values were then scaled by 1.05 to their corresponding aBG values. The Nathan model is still considered as a standard model for estimating HbA1c. The model however is a static model and does not take into account the nuances and variations of BG throughout the day. Lledó-García et al. [7] showed us a few non-static/kinetic models using the same dataset as the Nathan Model [3]. These kinetic models used prior measured/estimated parameter values, including red blood cell (RBC) life span, mean RBC life span and many more to provide non-static estimates of HbA1c. Bergenstal et al. [11] introduced a new method for estimating HbA1c. Bergenstal et al. [11] reintroduced the population HbA1c estimate as glucose monitoring index (GMI) and explicitly labeled it as a metric/index that should be used for diabetes management and diabetes education. GMI and its relationship with aBG was established using a linear regression analysis just like the Nathan

model [3]. The GMI estimated in many cases was significantly different from the measured HbA1c and there for it was to be used as a proxy for the HbA1c and not its substitute. Hu et al. [12] studied the relationship between the estimated HbA1c and the measured HbA1c for Chinese T2D patients and developed 6 new models for estimating HbA1c. There were linear as well as non-linear models in this collection of 6 models. These models were obtained using constrained regression analysis and were shown to provide reliable HbA1c estimates. Zaitcev et al. [13] later developed a data driven deep learning model for calculating HbA1c from a CGM trace.

Although the previous models included linear and non-linear models, these models were almost all static models. We have observed that the models constructed traditionally using linear, non-linear or constraint regression analysis, generally have a wide 95 % confidence interval and subsequently a wider 95 % prediction interval. This makes the HbA1c estimates provided by them unreliable for clinical use; the traditional experimental measurement methods for HbA1c are far more reliable as they provide a relatively narrow confidence interval. Therefore, more sophisticated methods are required to provide HbA1c estimates with similar narrow confidence intervals from CGM traces. This has lead to the development of various dynamical or kinetic models.

Xu et al. [15] constructed a kinetic model for estimating HbA1c from the CGM trace of any T2D and T1D patient. This model took into account the RBC life-span and the glycation of hemoglobin. This kinetic model assumes a simple dynamical model for glycation of hemoglobin and estimates the parameters of this model from the given CGM trace. It then uses this parameterized dynamical model to provide HbA1c estimates. This model was later validated and it was shown that compared to GMI and HbA1c estimated using the Nathan equation [3], the HbA1c predictions provided by this kinetic model were superior [16]. Later Xu et al. [14] addressed the shortfalls of laboratory measured HbA1c and showed that laboratory HbA1c can sometimes overestimate or underestimate the diabetic complications for a patient.

Thus, there are a plethora of linear regression, non-linear regression and, sophisticated dynamical models for modeling the relationship between BG, ISF and,

HbA1c so as to provide a better HbA1c estimate for T2D patients. There are however very few models that are specific to the T1D system. There have been regression models which were built on datasets which had a large proportion of T1D patients like, the Nathan model [3], and some models which were built on only T2D datasets, but how these models perform in patients with T1D has not been extensively explored. This lack of clarity is even more pronounced when we include the ethnicity of the patient and start developing models which are specialized for a particular ethnicity. Herman et al. [5] showed that the aBG for T2D patients do not differ significantly over racial and ethnic differences but the HbA1c value does. [5] showed that for T2D patients of Asian, Hispanic, African and other racial and/or ethnic groups the HbA1c values are significantly larger than HbA1c values of Caucasian T2D patients. Leslie and Cohen [6] showed that the relationship between aBG and HbA1c vary significantly within individuals. finally, Salam et al. [19] showed that for a given aISF value, the HbA1c value for T2D patients were significantly higher but after stratifying the data according to racial and ethnic backgrounds of the patients this difference between the HbA1c became insignificant for any given aISF value.

In an earlier study, Majumdar et al. [18] developed a linear regression model to estimate HbA1c directly from aISF of CGM traces from Indian patients already diagnosed with diabetes, pre-diabetics and non-diabetic participants, we have hitherto referred to this model as the Direct Model [18]. Majumdar et al. [18] have also made a linear model for HbA1c that takes the value from the calculated aISF via the aBG which they referred to as the Adapted Nathan Model. In this manuscript, we verify if previous linear regression models namely the Nathan model [3], the Adapted Nathan Model [18] and, the Direct Model [18] which were developed on a purely T2D CGM database or as in the case of the Nathan model [3], developed on a mixed database, may be reliably used at least at a population level to estimate HbA1c from CGM traces of Indian children and adolescents with T1D.

3.2 Materials and Methods

3.2.1 The T1D Dataset

Mathematical models describing the relationship between various physical quantities or describing any phenomenon, need to be tested before they can be used for prediction purposes. This is the same for biophysical models. We analyze three such models here in this manuscript, the Nathan model [3], the direct and adapted Nathan model [16]. We verify if the HbA1c estimates provided by these models are statistically reliable for use in Indian children and Indian adolescents with T1D.

Two datasets consisting of CGM traces and various bio-markers from subjects with T1D were collected, curated and annonymised at a tertiary level care center for T1D in, Pune, India. For the CGM traces, the Freestyle Libre Pro by Abott Co. Ltd. was applied on the the non-dominant arm. This was a professional blinded CGM system containing sensor and single reader device system which recorded ISF every 15 mins for a maximum duration of 14 days. At the end of the wear period, sensor had been scanned to upload the data and blood sample was collected for HbA1c. The first dataset, which we will label as the T1D 2022–2023 dataset (this was for a study on assessment of utility of Fructosamine as a proxy for CGM in T1D, as yet unpublished), contained the patient id, the age (via the date of birth), the duration after first diagnosis (via date of first diagnosis), the gender, the date of CGM device attachment, the corresponding CGM trace and, multiple HbA1c measurements along with their corresponding measurement dates from 34 unique T1D patients. In the T1D 2022–2023 dataset, each CGM trace could be paired to two corresponding HbA1c readings, one taken before (approximately before 3 months) the CGM device was placed on the patient and the other after (approximately after 3 months) the device was removed. After filtering out the entries where the HbA1c measurement corresponding to a CGM trace or the CGM trace to a corresponding HbA1c measurement were missing, we had a total of 51 remaining CGM — HbA1c pairs from 26 unique T1D patients. We also identified all such CGM — HbA1c pairs with a CGM trace that was data-sufficient. The definition of data-sufficient used by Danne et al. [10] was used for the identification. We checked if the total number of days the device was active was at least 14 and

also checked if the percentage of available CGM readings were at least 70 % [10, 16]. If these conditions were satisfied, we labeled those CGM — HbA1c pairs with a data-sufficient CGM trace as data-sufficient pairs. For the T1D 2022–2023 dataset we had a total of 44 such data-sufficient CGM — HbA1c pairs.

The second dataset, which we label as the T1D 2019–2020 dataset was first used in Mondkar et al. [20], contained like the previous dataset, the patient id, the age (via the date of birth), the duration after first diagnosis (via date of first diagnosis), the gender, the dates of CGM device attachment, the corresponding CGM traces and, multiple HbA1c measurements along with their corresponding measurement dates from 92 unique T1D patients. The T1D 2019–2020 dataset however contained multiple CGM traces (about 2 each) from each patient and just like the previous dataset each CGM trace could be paired with 2 corresponding HbA1c measurements, one made before (approximately before 3 months) the CGM device was attached and the other measurement made after (approximately after 3 months) the CGM device was removed. After filtering out the entries where the HbA1c measurement corresponding to a CGM trace or the CGM trace to a corresponding HbA1c measurement were missing from the T1D 2019–2020 dataset, we had a total of 283 remaining CGM-—HbA1c pairs from 84 unique T1D patients. Similar to the treatment of the T1D 2022–2023 dataset, we also identified all the data-sufficient pairs. The T1D 2019–2020 dataset had a total of 253 data-sufficient pairs from 80 unique T1D patients. Combining the T1D 2022–2023 and the T1D 2019–2020 dataset into the T1D dataset, we had a total of 334 CGM-—HbA1c pairs from 108 unique T1D patients. Out of the 334 CGM–HbA1c pairs, we identified 297 data-sufficient pairs contributed from 100 unique patients. We used the T1D dataset for all the analysis presented in this manuscript. Many of our analysis are then later repeated with only the data-sufficient pairs. The table providing a detailed description of the datasets that we received is provided below as Table. 3.1.

		The Complete T1D dataset			All CGM—HbA1c pairs from the T1D dataset			Only data-sufficient CGM—HbA1c pairs from the T1D dataset		
		Male	Female	All	Male	Female	All	Male	Female	All
T1D 2022–2023	$N_{\#\text{patient}}$	14	20	34	12	14	26	10	12	22
	$N_{\#\text{pairs}}$				25	26	51	22	22	44
	Age (years)	14.86± 4.46	11.54 ± 4.01	13.14± 4.55	15.1 ± 4.4	11.8 ± 4.0	13.4 ± 4.5	14.8 ± 4.0	12.0 ± 4.3	13.4 ± 4.4
	HbA1c (mmol/mol)	77.16± 18.82	87.70 ± 32.76	82.54± 27.33	77.16± 18.82	87.71 ± 32.72	82.54± 27.33	75.97± 19.29	85.55 ± 28.40	80.76± 24.74
T1D 2019–2020	$N_{\#\text{patient}}$	45	41	86	44	40	84	41	39	80
	$N_{\#\text{pairs}}$				145	138	283	125	128	253
	Age (years)	18.96± 2.80	19.37 ± 3.26	19.16± 3.04	18.90± 2.77	19.18 ± 3.15	19.03± 2.97	19.03± 2.85	19.11 ± 3.19	19.07± 3.03
	HbA1c (mmol/mol)	83.20± 18.30	83.61 ± 20.94	83.40± 19.63	83.20± 18.30	83.61 ± 20.94	83.40± 19.63	82.15± 18.62	83.18 ± 20.94	82.67± 19.83
T1D	$N_{\#\text{patient}}$	59	65	124	54	54	108	49	51	100
	$N_{\#\text{pairs}}$				170	164	334	147	150	297
	Age (years)	18.34± 3.32	18.29 ± 4.24	18.31± 3.80	18.34± 3.34	17.99 ± 4.28	18.74± 3.83	18.40± 3.40	18.08 ± 4.20	18.24± 3.83
	HbA1c (mmol/mol)	82.09± 17.97	83.08 ± 22.81	82.59± 20.54	82.30± 18.50	84.26 ± 23.28	83.27± 20.99	81.23± 18.85	83.53 ± 22.20	82.39± 20.65

Table 3.1: This table provides a summary of the T1D dataset. As described in Section. 3.2.2 the T1D dataset constists of 2 sub-datasets, the T1D 2022–2023 dataset and the T1D 2019–2020 dataset. The description of all three datasets are grouped row-wise and provided in the table. The description of these datasets in terms of the set of CGM–HbA1c pairs and the the set of data-sufficient CGM–HbA1c (data-sufficient pairs are described in Section. 3.2.2); are grouped column-wise and provided in the table. The number of unique pattients, $N_{\#\text{patients}}$, the number of CGM–HbA1c pairs, $N_{\#\text{pairs}}$, the average age in years (provided as $mean \pm SD$), Age and, the average HbA1c in mmol/mol (provided as $mean \pm SD$), HbA1c are provided for each datasets in the table.

3.2.2 The Nathan Model for estimating HbA1c

The Nathan model [3] was developed around 2008 as a model for establishing the relationship between aBG and its corresponding HbA1c at a population level. It was formulated as an equation where the HbA1c could be used to obtain an estimate of the corresponding aBG. The equation that was suggested by Nathan et al. [3] is as follows,

$$\mathrm{aBG}_{\mathrm{mg/dL}} = 1.59 \times \mathrm{HbA1c}_{\%} - 2.59, \tag{3.1a}$$

$$\text{or,} \quad 1.05 \times \mathrm{aISF}_{\mathrm{mg/dL}} = 1.59 \times \mathrm{HbA1c}_{\%} - 2.59, \tag{3.1b}$$

$$\text{or,} \quad \mathrm{aBG}_{\mathrm{mmol\,L}} = 28.7 \times \mathrm{HbA1c}_{\%} - 46.7, \tag{3.1c}$$

where, $\mathrm{aBG}_{\mathrm{mg/dL}}$ is the aBG in mg/dL (similarly, $\mathrm{aBG}_{\mathrm{mmol/L}}$ is aBG in mmol/L), $\mathrm{aISF}_{\mathrm{mg/dL}}$ is the aISF measured in mg/dL and, $\mathrm{HbA1c}_{\%}$ is the HbA1c measured and reported in percentage(%). The Nathan model [3] was built by screening a dataset containing 661 participants of age 18 to 70 year from 11 centers around the world to obtain a dataset of 507 participants. The dataset contained glucose concentration readings from SMBG and also from CGM devices. The CGM devices however only provided ISF readings which were scaled to their corresponding BG by 1.05 as seen in Eq. (3.1b). The Nathan model [3] although provided as Eqs. (3.1), has been popularly used for calculating HbA1c from the calculated aBG from a series of SMBG readings or from a CGM trace. The Nathan model [3] is the standard model, that is most popularly used for estimating HbA1c from aBG and/or aISF values.

Majumdar et al. [18] however, showed HbA1c calculated using the Nathan equation [3] were statistically not reliable for the Indian population. The Pune-2021 and the Joshi-2018 datasets used by Majumdar et al. [18] contained CGM traces and HbA1c for pre-diabetic, T2D and, non-diabetic participants. Taking the measured HbA1c from the Pune-2021 and Joshi-2018 datasets as a sample and pairing each of them with the estimated HbA1c, calculated from the CGM trace, using the Nathan model [3] a sample of paired estimated and observed HbA1c values were created.

A paired t-test was run on this paired sample and it was found that at a popu-

lation level, the estimated HbA1c was significantly different from HbA1c measured experimentally. The Nathan model [3] however, is still used for obtaining HbA1c estimates for both T1D and T2D patients.

We used the T1D dataset (a combination of T1D 2022–2023 and T1D 2019–2020 datasets) to generate HbA1c using the Nathan model [3] and performed a paired t-test by pairing the estimated HbA1c values with their corresponding experimentally measured HbA1c. A Quantile-Quantile plot was generated using the experimentally measured HbA1c and the estimated HbA1c. It was used to verify if the paired differences belonged to a normal distribution. The null hypothesis, H_0, for the paired t-test was taken to be $H_0 \equiv \text{HbA1c}^{\text{obs}}_{\text{mean}} - \text{HbA1c}^{\text{est}}_{\text{mean}} = 0$ and the alternative hypothesis as $H_a \equiv \text{HbA1c}^{\text{obs}}_{\text{mean}} - \text{HbA1c}^{\text{est}}_{\text{mean}} \neq 0$.

If we reject null hypothesis, H_0, we will conclude that Nathan model [3] cannot be used to reliably estimate the HbA1c for T1D patients from the child and adolescent Indian population. We also constructed a Bland-Altman plot for the paired differences between the predicted HbA1c and the experimentally measured HbA1c values as it would help us to understand the reliability of the Nathan model [3] for CGM traces in Indian patients with T1D. The paired t-tests were computed using the `ttest_ rel` function of the `SciPy` (version 1.12.0) `Python` package. The plots were generated using the `pyplot` module of the `matplotlib` (version 3.8.3) `Python` package. The complete analysis was done using `Python` (version 3.11.1)

3.2.3 The Direct Model for estimating HbA1c

The Direct Model was first introduced as an alternative to the Nathan model [3], for obtaining HbA1c estimates for the Indian population by Majumdar et al. [18]. This model estimates the HbA1c from the aISF which is calculated as the mean of the ISF readings from any CGM trace. The equation describing the direct model [18] of estimating HbA1c is as follows:

$$\text{HbA1c}_{\text{mmol/mol}} = 0.319 \times \text{aISF}_{\text{mg/dL}} + 16.73, \tag{3.2}$$

where $\text{HbA1c}_{\text{mmol/mol}}$ is the HbA1c in mmol/mol and $\text{aISF}_{\text{mg/dL}}$ the aISF in mg/dL. The aISF was calculated from the ISF readings from a given CGM trace using the following equation

$$\text{aISF}_{\text{mg/dL}} = \frac{1}{N}\sum_{i=1}^{i=N} \text{ISF}_{i,\ \text{mg/dL}}, \tag{3.3}$$

where $\text{aISF}_{\text{mg/dL}}$ is the aISF in mg/dL, $\text{ISF}_{i,\ \text{mg/dL}}$ the ISF in mg/dL and, N is the total number of ISF measurements. This model just like the Nathan model [3], computes HbA1c directly from the calculated aISF. The Direct Model [18] was constructed using a linear regression analysis of the CGM and HbA1c data from the Pune-2021 and Joshi-2018 dataset described by Majumdar et al. [18]. The Direct Model was then validated using a subset of the Pune-2021 and Joshi-2018 dataset. As mentioned in Section. 3.2.2, the Pune-2021 and the Joshi-2018 dataset from Majumdar et al. [18] together contained CGM traces and HbA1c values from only pre-diabetic, T2D and, non-diabetic participants. There were no T1D participants in the Pune-2021 and Joshi-2018 dataset.

We therefore set out to validate the Direct Model [18] for estimating HbA1c for T1D patients. We calculated a paired t-test between the measured HbA1c from the T1D dataset and the estimated HbA1c using the Direct Model [18]. The null hypothesis, H_0, for the paired t-test was taken to be $H_0 \equiv \text{HbA1c}^{\text{obs}}_{\text{mean}} - \text{HbA1c}^{\text{est}}_{\text{mean}} = 0$ and the alternative hypothesis as $H_a \equiv \text{HbA1c}^{\text{obs}}_{\text{mean}} - \text{HbA1c}^{\text{est}}_{\text{mean}} \neq 0$

The paired t-test would then verify if the population mean HbA1c is statistically indistinguishable from the population mean of the estimated HbA1c. The normality of the paired differences between the estimated and the experimentally measured HbA1c was verified visually using a Quantile-Quantile plot. A Bland-Altman plot of the paired differences between the estimated and experimentally measured HbA1c was also constructed as it helped us further understand the reliability of the Direct Model [18] in estimating HbA1c values for patients with T1D. This paired t-test would help us verify if the Direct Model [18] could be used to obtain statistically reliable HbA1c estimates from the CGM traces of T1D patients from the child and adolescent Indian population. The paired t-tests were calculated using the `ttest_rel` function of the `SciPy` (version 1.12.0) `Python` package. The plots were generated using the `pyplot` module of the `matplotlib` (version 3.8.3) `Python` package. The complete analysis was done using `Python` (version 3.11.1)

3.2.4 The Adapted Nathan Model for estimating HbA1c

The Adapted Nathan Model was first introduced around 2023 by Majumdar et al. [18] as an alternative to the Nathan model [3] and Direct Model [18] for obtaining HbA1c estimates from CGM traces of non-diabetic, pre-diabetic and T2D patients from the Indian population. As previously stated in Section 3.2.2, the Nathan model [3] was constructed using BG, ISF and, HbA1c measurements from 507 participants and those ISF readings used to construct the Nathan model [3] were scaled to their corresponding BG by a factor of 1.05. The adapted Nathan model [18] uses a scaling factor of 1.17 which was obtained using a constrained linear regression analysis on the CGM traces and their measured HbA1c from the Pune-2021 and Joshi-2018 dataset. The equation for calculating HbA1c from aISF (calculated using Eq. (3.3)) using the Adapted Nathan Model is as follows

$$\mathrm{HbA1c}_{\mathrm{mmol/mol}} = 0.38 \times \left(1.17 \times \mathrm{aISF}_{\mathrm{mg/dL}}\right) - 5.60, \tag{3.4}$$

where $\mathrm{HbA1c}_{\mathrm{mmol/mol}}$ is the HbA1c in mmol/mol and $\mathrm{aISF}_{\mathrm{millig/dL}}$ the aISF in mg/dL. The Pune-2021 and Joshi-2018 dataset which were used to construct and test the Adapted Nathan Model [18] however, contained only CGM traces and HbA1c measurements from pre-diabetic, T2D and, non-diabetic participants. We therefore validated the Adapted Nathan Model on a dataset containing CGM traces with corresponding HbA1c measurements from Indian children and adolescents with T1D. We performed a paired t-test with the experimentally measured HbA1c values from the T1D dataset and the estimated HbA1c values calculated using the Adapted Nathan Model [18] from their corresponding CGM trace. The null hypothesis, H_0, for the paired t-test was taken to be $H_0 \equiv \mathrm{HbA1c}^{\mathrm{obs}}_{\mathrm{mean}} - \mathrm{HbA1c}^{\mathrm{est}}_{\mathrm{mean}} = 0$ and the alternative hypothesis as $H_a \equiv \mathrm{HbA1c}^{\mathrm{obs}}_{\mathrm{mean}} - \mathrm{HbA1c}^{\mathrm{est}}_{\mathrm{mean}} \neq 0$.

The paired t-test would then verify if the population mean HbA1c was statistically indistinguishable from the population mean of the estimated HbA1c. The normality of the paired differences between the estimated and the experimentally measured HbA1c was verified visually using a Quantile-Quantile plot. A Bland-Altman plot of the paired differences between the estimated and experimentally measured HbA1c was also constructed as it would help us further understand the

reliability of the Adapted Nathan Model [18] in estimating HbA1c values for T1D patients. This paired t-test would help us verify if the Adapted Nathan Model [18] could be used to obtain statistically reliable HbA1c estimates from the CGM traces of T1D patients from the Indian population. The paired t-tests were performed using the `ttest_rel` function of the `SciPy` (version 1.12.0) `Python` package. The plots were generated using the pyplot module of the `matplotlib` (version 3.8.3) `Python` package. The complete analysis was done using `Python` (version 3.11.1)

3.3 Results

3.3.1 Validation of the Direct Model

The estimated HbA1c of the paired CGM traces of the T1D dataset were calculated using Eq. (3.2) as suggested by Majumdar et al. [18]. The paired differences with the estimated HbA1c and the experimentally measured HbA1c were calculated and the corresponding Quantile-Quantile plots were constructed (Figure. 3.1) for all the subsets of the T1D dataset. The Subplot. 2a in Figure. 3.1, shows that the paired differences between the estimated and experimentally measured HbA1c belong to a distribution which is very close to a normal distribution and hence performing a paired t-test with the estimated and experimental HbA1c would be an appropriate method to verify if the HbA1cs estimated using the direct method are statistically reliably for T1D patients of the Indian population.

The Quantile-Quantile plot was then constructed for the data-sufficient CGM—HbA1c pairs and also produced as Subplot. 2b in Figure. 3.1. Studying the Subplot. 2b from Figure. 3.1, we come to a similar conclusion about the paired difference between the estimated and observed HbA1c for the data-sufficient CGM—HbA1c pairs.

We therefore performed a paired t-test with the experimental HbA1c and the estimated HbA1cs calculated from the CGM traces using the Direct Model [18] (HbA1c estimates were calculated using Eq. (3.2)). This paired t-test was performed on a total of 334 ($df = 333$) pairs of estimated and experimentally measured HbA1c and yielded a $p - value = 0.495$. The paired t-test was also repeated on the 297 ($df = 296$) data-sufficient CGM–HbA1c pairs of the T1D dataset and

yielded a $p - value = 0.801$. The results are also provided in Table. 3.2. Figure. 3.2 shows the Direct Model, Eq. (3.2) as a solid black line along with its 95 % confidence interval, surrounded by the data-sufficient and data-insufficient CGM–HbA1c pairs from the T1D dataset. Figure. 3.3 similarly shows the Direct Model [18] along with its 95 % prediction interval, surrounded by the data-sufficient and data-insufficient CGM–HbA1c pairs from the T1D dataset. The figure shows that, for the T1D patients from the Indian population, the Direct Model [18] represented by Eq. (3.2) can be used to obtain estimates of HbA1c from the aISF, calculated using their CGM trace, which are at a population level statistically indistinguishable from the experimental HbA1c. The Bland-Altman plot constructed using the experimentally measured HbA1c and the estimated HbA1c using the direct model [18] provided as Subplot. 2a and Subplot. 2b in Figure. 3.4, shows that the paired differences between the estimated and experimentally measured HbA1c are fairly evenly distributed about the black solid line representing the mean paired difference and that estimated HbA1c provided by the Direct Model are statistically reliable at a population level.

Models	All CGM — HbA1c pairs			Only data-sufficient CGM — HbA1c pairs		
	df	$t - statistics$	$p - value$	df	$t - statistics$	$p - value$
Nathan Model[3]	333	5.484	8.206×10^{-8}	296	5.045	7.791×10^{-7}
Direct Model[18]	333	0.681	0.495	296	0.251	0.801
Adapted Nathan Model[18]	333	2.220	0.027	296	2.356	0.019

Table 3.2: This table contains the results of the paired t-tests performed between the experimentally measured HbA1c and the HbA1c estimated using the Nathan model [3], the Direct Model [18] and, the Adapted Nathan Model [18] respectively. The analysis was perform on the set of all CGM — HbA1c pairs and on the set of all data-sufficient CGM — HbA1c pairs from the T1D dataset.

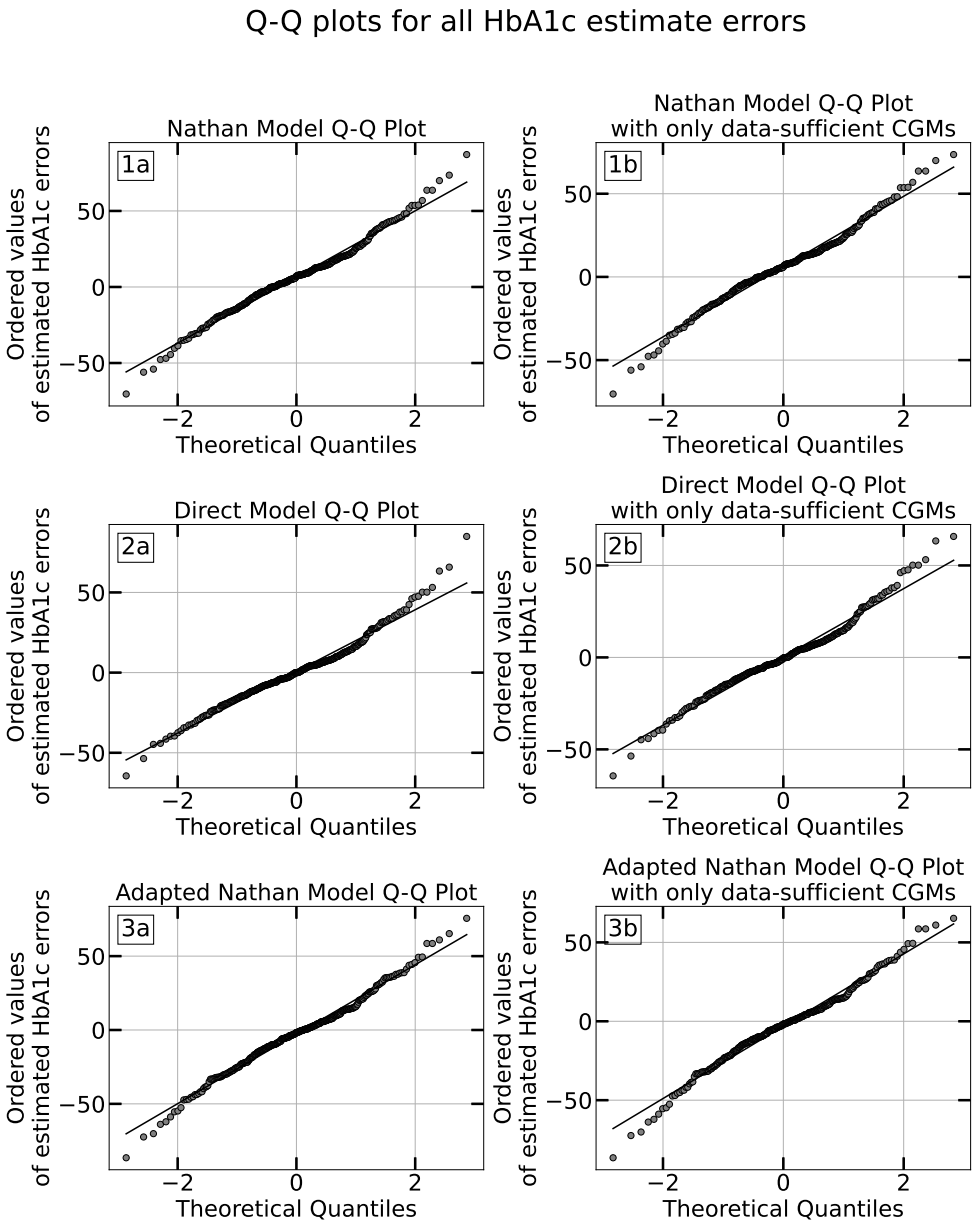


Figure 3.1: The Figure shows the Quantile-Quantile plots for the paired differences between the experimentally measured HbA1c and the HbA1c estimated using the Nathan [3] (Subplots. 1a and 1b), direct [18] (Subplots. 2a and 2b) and, Adapted Nathan Models [18] (Subplots. 3a and 3b). The Subplots. 1a, 2a and, 3a were generated using all the CGM — HbA1c pairs from the T1D dataset. The Subplots. 1b, 2b and, 3b however were generated using only the data-sufficient pairs.

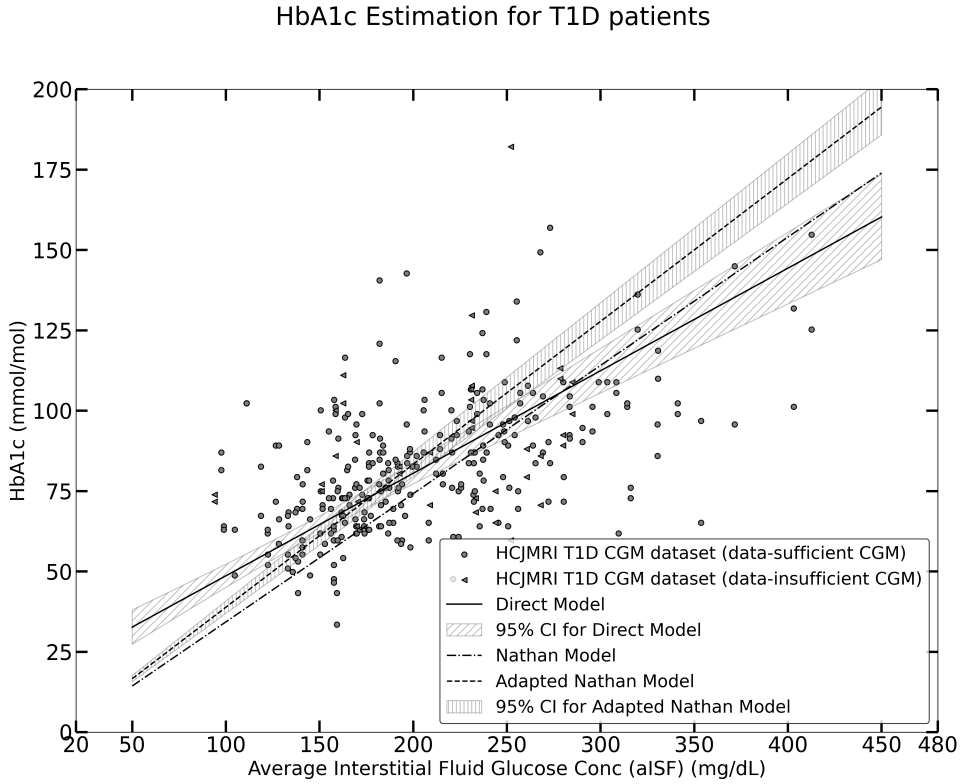


Figure 3.2: HbA1c vs glsaisf. The X-axis represents aISF in mg/dL and the Y-axis represents HbA1c in mmol/mol. The solid black circles represent the data-sufficient CGM — HbA1c pairs from the T1D dataset. The solid black triangles represent the data-insufficient pairs. The solid black line represents the Direct Model [18], the dot-dashed line represents the Nathan model [3] and, the dashed line represents the Adapted Nathan Model [18]. The corresponding 95 % CI of the direct and Adapted Nathan Model is also shown as hatched regions.

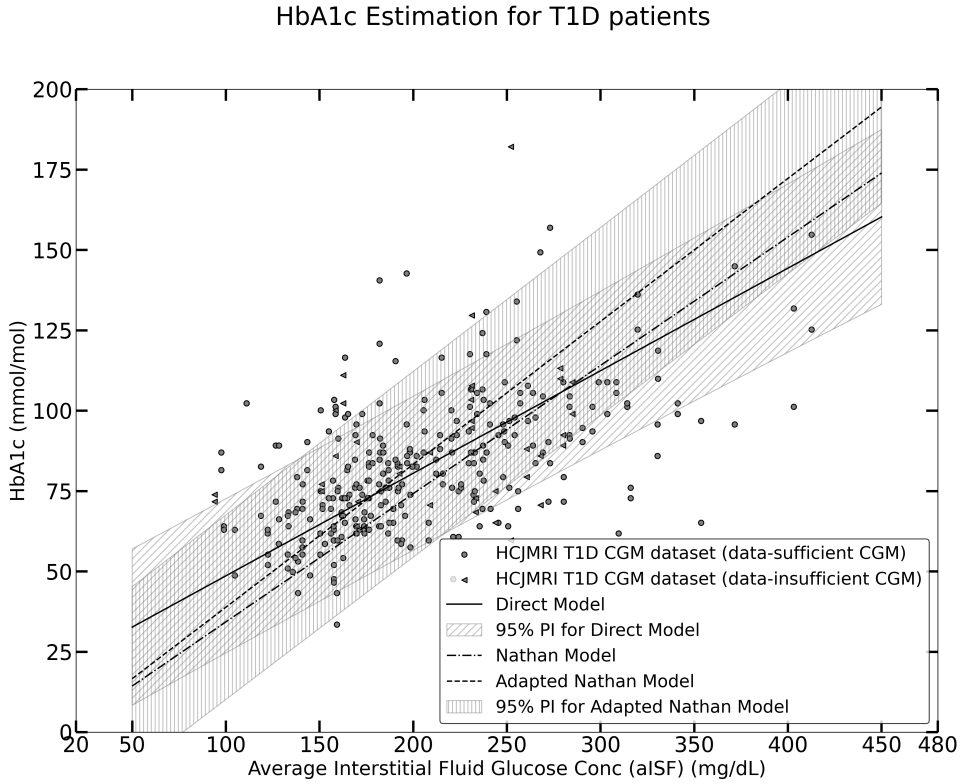


Figure 3.3: HbA1c vs glsaisf. The X-axis represents aISF in mg/dL and the Y-axis represents HbA1c in mmol/mol. The solid black circles represent the data-sufficient CGM — HbA1c pairs from the T1D dataset. The solid black triangles represent the data-insufficient pairs. The solid black line represents the Direct Model [18], the dot-dashed line represents the Nathan model [3] and, the dashed line represents the Adapted Nathan Model [18]. The corresponding 95 % PI of the direct and Adapted Nathan Model is also shown as hatched regions.

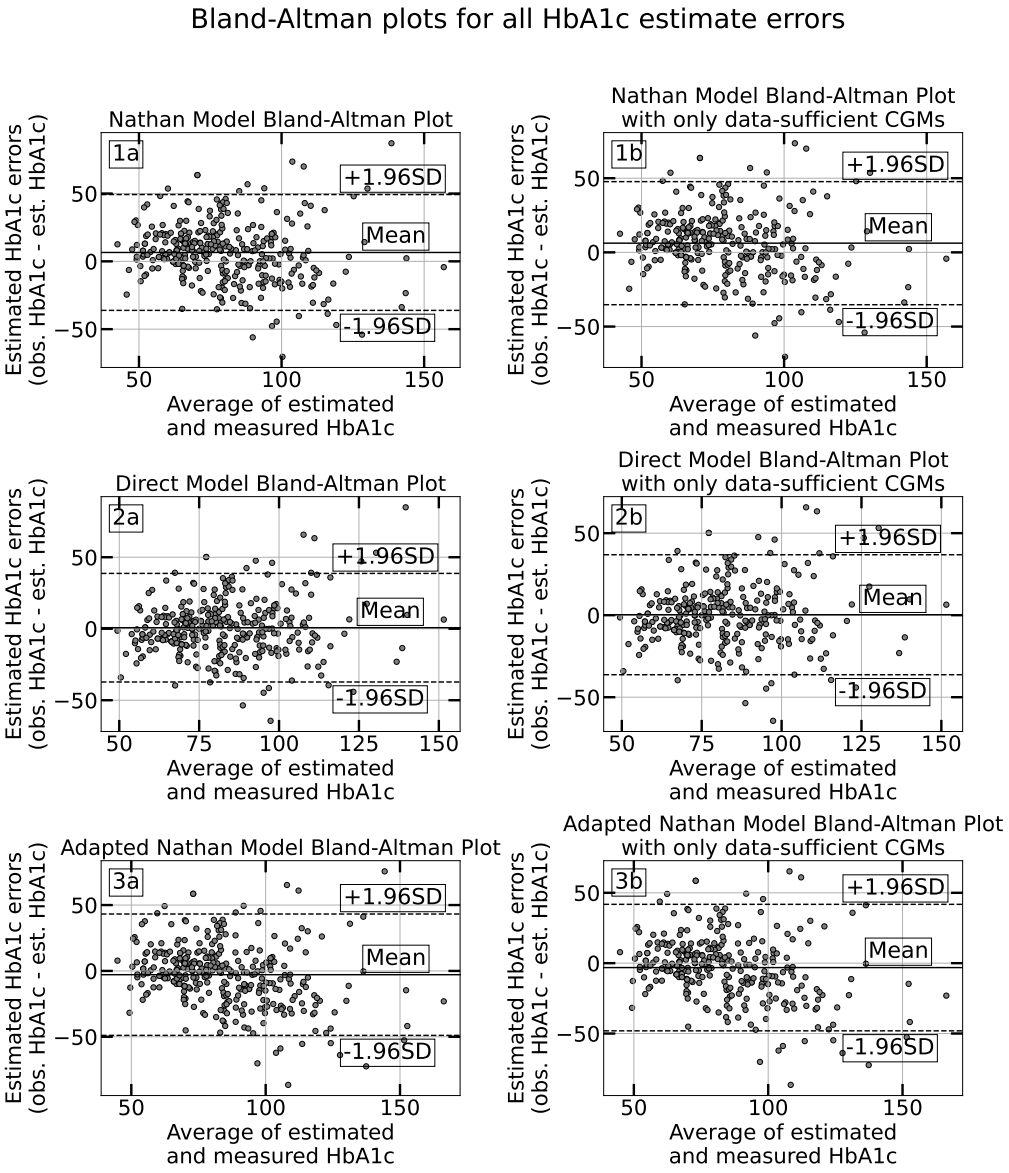


Figure 3.4: The Figure shows the Bland-Altman plots for the experimentally measured HbA1c and the HbA1c estimated using the Nathan [3] (Subplots. 1a and 1b), direct [18] (Subplots. 2a and 2b) and, Adapted Nathan Models [18] (Subplots. 3a and 3b). The Subplots. 1a, 2a and, 3a were generated using all the CGM — HbA1c pairs from the T1D dataset. The Subplots. 1b, 2b and, 3b however were generated using only the data-sufficient pairs. The black dashed lines in the Subplot's represent the ±1.96SD lines and the black solid line represent the mean of the paired differences of the estimated and observed HbA1c.

3.3.2 Validation of the Nathan model and the Adapted Nathan Model

The Nathan Model [3] and the Adapted Nathan Model [18] may both be used to obtain estimates of HbA1c from a CGM trace. For each model aISF is calculated using Eq. (3.3) and then it is scaled to an estimated aBG by a scaling factor of 1.05 for the Nathan model [3] and a scaling factor of 1.17 for the adapted Nathan model [18]. The calculated aBG is then used by both models to estimate HbA1c using Eq. (3.1a) (or Eq. (3.1c) based on the units required). Although these models are quite similar, as stated earlier in Section. 3.2.2, the Nathan model [3] has been constructed from a dataset with non-diabetic, T1D and T2D participants. The adapted Nathan model [18] however was constructed using the Pune-2021 and Joshi-2018 datasets which consisted mainly T2D, T2D pre- and non-diabetics participants. The dataset used to construct the Nathan model [3] was however majorly Caucasian, whereas the dataset used to construct the Adapted Nathan Model [18] contained only Indian participants. Therefore, it is necessary to verify if these models can be reliably used to estimate HbA1c for T1D patients from the Indian population.

We obtain estimates of HbA1c from the CGM traces from all CGM-—HbA1c pairs of the T1D dataset using both the Nathan model [3] and the Adapted Nathan Model [18]. To verify if these estimates are statistically indistinguishable from the experimental HbA1c we need to conduct a paired t-test with the estimated and the experimentally measured HbA1cvalues. This required us to verify if the paired differences between the estimated HbA1c and its corresponding experimentally measured HbA1c follow a distribution close to a normal distribution. We constructed a Quantile-Quantile plot using the paired differences between the estimated and measured HbA1c for both the models. The construction was then repeated for both the models using only data-sufficient CGM—-HbA1c pairs of the T1D dataset. Subplot. 1a and Subplot. 1b in Figure. 3.1 show the Quantile-Quantile plot for the Nathan model [3]. We can conclude from Subplot. 1a of Figure. 3.1 that the paired differences between the estimated and experimentally measured HbA1c follow a distribution that is close to a normal distribution for all the CGM–HbA1c pairs in the T1D dataset. Subplot. 1b of Figure. 3.1 illustrates

the same for the data-sufficient pairs. Subplot. 3a and Subplot. 3b in Figure. 3.1 show the Quantile- Quantile plot for the Adapted Nathan Model [18]. We can conclude from Subplot. 3a of Figure. 3.1 that the paired differences between the estimated and experimentally measured HbA1c follow a distribution that is close to a normal distribution for all the CGM–HbA1c pairs in the T1D dataset. Subplot. 3b of Figure. 3.1 illustrates the same for the data-sufficient pairs.

We then performed a paired t-test with the experimental HbA1c and estimated HbA1c using the Nathan model [3] on all 334 ($df = 333$) CGM-—HbA1c pairs of the T1D dataset. The paired t-test yielded a $p - value < 0.001$. The paired t-test conducted on only the 297 ($df = 296$) data-sufficient CGM–HbA1c pairs yielded a similar $p - value < 0.001$. Therefore, we concluded that the HbA1c estimates obtained for Indian T1D patients using the Nathan model [3] were significantly (for a significance level, $\alpha = 0.05$) different from the experimental HbA1c. The results have also been provided in Table. 3.2. Figures. 3.2 and 3.3 show that the Nathan model [3] as a black dot-dashed line surrounded by its 95 % confidence and prediction interval respectively. The data sufficient CGM–HbA1c pairs along with the data-insufficient pairs are also shown in Figures. 3.2 and 3.3.

We then repeated the paired t-test with the experimental and estimated HbA1c using the Adapted Nathan Model [18] on all 334 ($df = 333$) CGM–HbA1c pairs of the T1D dataset. The paired t-test yielded a $p - value = 0.027$. The paired t-test conducted on only the 297 ($df = 296$) data-sufficient CGM—-HbA1c pairs yielded a $p - value = 0.19$. Therefore, we conclude that the HbA1c estimates obtained for Indian T1D patients using the Adapted Nathan Model [18] are significantly (for a significance level, $\alpha = 0.05$) different (for the data-sufficient pairs and, very close for all pairs from the experimental HbA1c values) and cannot be used reliably to estimate HbA1c. The results are also provided in Table. 3.2. Figures. 3.2 and 3.3 show the Adapted Nathan Model [3] as a black dashed line surrounded by its 95 % confidence and prediction interval respectively. The data-sufficient CGM-—HbA1c pairs along with the data-insufficient pairs are also shown in Figures. 3.2 and Figures. 3.3. Bland-Altman plots were then constructed for both models using the estimated and experimentally measured HbA1c values to study how the errors in the estimated HbA1c behaved. Figure. 3.4 shows us the Bland-Altman plots for

the Nathan model [3] in Subplot. 1a and Subplot. 1b and, for the Adapted Nathan Model [18] in Subplot. 3a and Subplot. 3b. It is clear from the Bland-Altman plots of these models that there is a pattern in the point-cloud. The trend is very prominent in Subplot. 3b. This shows that for larger HbA1c values, both the Nathan [3] and the Adapted Nathan Model [18] overestimate the HbA1c values and are unable to provide a reasonable estimate.

3.4 Discussion

The ability to estimate HbA1c reliably from any given CGM trace is considered important and an increasingly necessary requirement by modern clinicians. This prevents unnecessary blood tests that are required to experimentally measure HbA1c and also helps reduce the cost of diabetes management. For patients with T1D it may be even more helpful as many of them are now equipped with insulin pumps which require an active CGM device to monitor their blood glucose and so are in possession of a large collection of their own CGM traces. These traces can be used to track their progress in diabetes management. Traditionally, HbA1c estimates have been calculated from a given CGM trace by first calculating the aISF and then using linear models, often built using standard linear regression analysis, to provide the HbA1c estimate. The most popular among these models is the Nathan model [3]. This model was also provided as a recommendation by Danne et al. [10] as a standard model for estimating HbA1c from any given CGM trace.

The Nathan model [3] was built on a large dataset containing SMBG measurements, CGM traces and HbA1c from a huge group of Caucasian participants. Majumdar et al. [18] had shown that the Nathan model [3] could not provide reliable HbA1c estimates from CGM traces of the Indian population. The analysis however was based on a CGM dataset containing CGM traces and HbA1c measurements from non-diabetic, pre-diabetic and, T2D participants. Therefore, it is of utmost importance to verify if the Nathan model [3] could be used to provide HbA1c estimates from CGM traces of T1D patients from the Indian population. Other than the traditionally used model, Majumdar et al. [18] also evaluated two new models to estimate HbA1c from any CGM trace for T2D and subjects with-

out diabetes from the Indian population. We also wanted to verify if these models could be used to provide HbA1c estimates for Indian children and adolescents with T1D.

We therefore performed a series of paired t-tests between the experimentally measured and estimated HbA1c from the Nathan [3], direct [18] and, adapted Nathan models [18]. The results (Section. 3.3 and Table 3.2) suggest that the Nathan model [3] described by Eq. 3.1b provides, from given CGM traces, HbA1c estimates which are different significantly from experimental values. HbA1c calculated by Eq. (3.4), i.e., the adapted Nathan model [18] were also statistically different when compared to the measured HbA1c. The HbA1c estimates for both the models did not improve when only data-sufficient CGM traces were used. This suggests that both models were not suitable for estimating HbA1c of Indian children and adolescents with T1D. The Bland-Altman plots showed for all the CGM–HbA1c pairs together and also for the data-sufficient pairs using the Nathan and Adapted Nathan Models tended to overestimate the HbA1c for larger values of HbA1c, especially above 100 mmol/mol. This effect was seen for the Nathan model [3] in Subplots. 1a and 1b of Figure. 3.4 as a tendency of the cloud of points to remain below the solid black line representing the mean paired difference. This was even more pronounced for the adapted Nathan model [18] (Subplots. 3a and 3b, especially observed in the latter). We believe this effect was more pronounced for the adapted Nathan model [18] because for any given aISF the corresponding HbA1c estimated by the adapted Nathan model [18] was always greater than the HbA1c estimated by the Nathan model [3]. This can be verified using the equations Eq. (3.1b) and Eq. (3.4) describing the two models respectively.

The results as shown in Section. 3.3 suggests that only the direct model [3] provides an HbA1c estimate that is indistinguishable from the experimental HbA1c statistically. The Bland-Altman plot for the direct model [18] (Subplots. 2a and 2b of Figure. 3.4) suggests that the paired difference between the estimated and experimentally measured HbA1c values are distributed about the black solid line representing the mean paired difference. This indicates that the direct model [18] has captured the major trend in HbA1c and the remaining errors are nearly random.

To the best of our knowledge, ours is the first study to verify the applicability of available equations to traces of CGM of Indian children and adolescents with T1D. However, the T1D dataset that was used for our analysis contained CGM traces from participants who were young (on average 13 years); further, the traces that were analyzed were of patients who did not have optimum control of T1D. All patients were also from one center. Thus, a larger dataset containing CGM traces and their measured HbA1c from a wider age group from multiple centers may help us to improve the HbA1c estimates and also study if separate models would be required to obtain clinically reliable HbA1c estimates for individual T1D patients.

To conclude, the direct Nathan model provides reliable HbA1c estimates from CGM traces in Indian children and adolescents with T1D at the population level. Further studies are required to improve HbA1c estimates from CGM traces at the individual level. We would alos like to thank Prof. Saroj Ghaskadbi from the Department of Zoology, Savitribai Phule Pune University, Pune, India for her valuable insight and advice.

Ethics statement

The study protocol for the T1D 2019–2020 dataset was approved by the Ethics Committee Jehangir Clinical Development Centre PVT. LTD. (ECR/352/Inst/MH/2013/RR-19) till 18th of October, 2020. An extension till 18th of August 2026 of the study was approved on the 21st of August 2023. The data analysis was reviewed and approved by the Institutional Human Ethics Committee (IHEC) of IISER Pune (IHEC/Admin/2023/002).

Funding

Financial support was received for the research, authorship, and/or publication of this article. Sayantan Majumdar is a Senior Research Fellow supported by a DST INSPIRE Fellowship, Department of Science and Technology, Government of India.

References

[1] Mortensen, H.; Vølund, A. Application of a biokinetic model for prediction and assessment of glycated haemoglobins in diabetic patients. *Scandinavian Journal of Clinical and Laboratory Investigation* **1988**, *48*, 595–602, DOI: 10.3109/00365518809085778, https://www.tandfonline.com/doi/abs/10.3109/00365518809085778 (cit. on p. 42).

[2] Boland, E.; Monsod, T.; Delucia, M.; Brandt, C. A.; Fernando, S.; Tamborlane, W. V. Limitations of conventional methods of self-monitoring of blood glucose: lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes. *Diabetes care* **2001**, *24*, 1858–1862, DOI: 10.2337/diacare.24.11.1858, https://diabetesjournals.org/care/article/24/11/1858/24785/Limitations-of-Conventional-Methods-of-Self (cit. on p. 41).

[3] Nathan, D. M.; Kuenen, J.; Borg, R.; Zheng, H.; Schoenfeld, D.; Heine, R. J.; Group, A.-D. A. G. (S. Translating the A1C assay into estimated average glucose values. *Diabetes care* **2008**, *31*, 1473–1478, DOI: 10.2337/dc08-0545, https://diabetesjournals.org/care/article/31/8/1473/28589/Translating-the-A1C-Assay-Into-Estimated-Average (cit. on pp. 42–45, 48–51, 53–61).

[4] Yamamoto-Honda, R.; Kitazato, H.; Hashimoto, S.; Takahashi, Y.; Yoshida, Y.; Hasegawa, C.; Akanuma, Y.; Noda, M. Distribution of blood glucose and the correlation between blood glucose and hemoglobin A1c levels in diabetic outpatients. *Endocrine journal* **2008**, *55*, 913–923, DOI: 10.1507/endocrj.k08e-071, https://www.jstage.jst.go.jp/article/endocrj/55/5/55_K08E-071/_article (cit. on p. 42).

[5] Herman, W. H.; Dungan, K. M.; Wolffenbuttel, B. H.; Buse, J. B.; Fahrbach, J. L.; Jiang, H.; Martin, S. Racial and ethnic differences in mean plasma glucose, hemoglobin A1c, and 1, 5-anhydroglucitol in over 2000 patients with type 2 diabetes. *The Journal of Clinical Endocrinology & Metabolism* **2009**, *94*, 1689–1694, DOI: 10.1210/jc.2008-1940, https://academic.oup.com/jcem/article/94/5/1689/2598400 (cit. on p. 44).

[6] Leslie, R. D. G.; Cohen, R. M. Biologic variability in plasma glucose, hemoglobin A1c, and advanced glycation end products associated with diabetes complications. *Journal of diabetes science and technology* **2009**, *3*, 635–643, DOI: 10.1177/193229680900300403, https://journals.sagepub.com/doi/10.1177/193229680900300403 (cit. on p. 44).

[7] Lledó-García, R.; Mazer, N. A.; Karlsson, M. O. A semi-mechanistic model of the relationship between average glucose and HbA1c in healthy and diabetic subjects. *Journal of pharmacokinetics and pharmacodynamics* **2013**, *40*, 129–142, DOI: 10.1007/s10928-012-9289-6, https://link.springer.com/article/10.1007/s10928-012-9289-6 (cit. on p. 42).

[8] Mosca, A.; Lapolla, A.; Gillery, P. Glycemic control in the clinical management of diabetic patients. *Clinical Chemistry and Laboratory Medicine (CCLM)* **2013**, *51*, 753–766, DOI: 10.1515/cclm-2012-0594, https://www.degruyter.com/document/doi/10.1515/cclm-2012-0594/html (cit. on p. 40).

[9] Gillery, P.; Lippi, G.; Plebani, M. Diagnosis of diabetes mellitus: reiterated responsibilities for the clinical laboratory. *Clinical Chemistry and Laboratory Medicine (CCLM)* **2014**, *52*, 935–936, DOI: 10.1515/cclm-2014-0443, https://www.degruyter.com/document/doi/10.1515/cclm-2014-0443/html (cit. on p. 40).

[10] Danne, T.; Nimri, R.; Battelino, T.; Bergenstal, R. M.; Close, K. L.; DeVries, J. H.; Garg, S.; Heinemann, L.; Hirsch, I.; Amiel, S. A., et al. International consensus on use of continuous glucose monitoring. *Diabetes care* **2017**, *40*, 1631–1640, DOI: 10.2337/dc17-1600, https://diabetesjournals.org/care/article/40/12/1631/37000/International-Consensus-on-Use-of-Continuous (cit. on pp. 45 sq., 60).

[11] Bergenstal, R. M.; Beck, R. W.; Close, K. L.; Grunberger, G.; Sacks, D. B.; Kowalski, A.; Brown, A. S.; Heinemann, L.; Aleppo, G.; Ryan, D. B., et al. Glucose management indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. *Diabetes care* **2018**, *41*, 2275–2280, DOI: 10.2337/dc18-1581, https://diabetesjournals.org/care/article/

41/11/2275/36593/Glucose-Management-Indicator-GMI-A-New-Term-for (cit. on p. 42).

[12] Hu, Y.; Shen, Y.; Yan, R.; Li, F.; Ding, B.; Wang, H.; Su, X.; Ma, J. Relationship between estimated glycosylated hemoglobin using flash glucose monitoring and actual measured glycosylated hemoglobin in a Chinese population. *Diabetes Therapy* **2020**, *11*, 2019–2027, DOI: 10.1007/s13300-020-00879-x, https://link.springer.com/article/10.1007/s13300-020-00879-x (cit. on p. 43).

[13] Zaitcev, A.; Eissa, M. R.; Hui, Z.; Good, T.; Elliott, J.; Benaissa, M. A Deep Neural Network Application for Improved Prediction of HbA_{1c} in Type 1 Diabetes. *IEEE Journal of Biomedical and Health Informatics* **2020**, *24*, 2932–2941, DOI: 10.1109/JBHI.2020.2967546, https://ieeexplore.ieee.org/document/8962002 (cit. on p. 43).

[14] Xu, Y.; Bergenstal, R. M.; Dunn, T. C.; Ajjan, R. A. Addressing shortfalls of laboratory HbA1c using a model that incorporates red cell lifespan. *Elife* **2021**, *10*, e69456, DOI: 10.7554/eLife.69456, https://elifesciences.org/articles/69456 (cit. on p. 43).

[15] Xu, Y.; Dunn, T. C.; Ajjan, R. A. A kinetic model for glucose levels and hemoglobin A1c provides a novel tool for individualized diabetes management. *Journal of Diabetes Science and Technology* **2021**, *15*, 294–302, DOI: 10.1177/1932296819897613, https://journals.sagepub.com/doi/10.1177/1932296819897613 (cit. on p. 43).

[16] Xu, Y.; Hirota, Y.; Ajjan, R. A.; Yamamoto, A.; Matsuoka, A.; Ogawa, W.; Dunn, T. C. Accurate prediction of HbA1c by continuous glucose monitoring using a kinetic model with patient-specific parameters for red blood cell lifespan and glucose uptake. *Diabetes and Vascular Disease Research* **2021**, *18*, 14791641211013734, DOI: 10.1177/14791641211013734, https://journals.sagepub.com/doi/10.1177/14791641211013734 (cit. on pp. 43, 45 sq.).

[17] Gillery, P. Hba1c and biomarkers of diabetes mellitus in clinical chemistry and laboratory medicine: ten years after. *Clinical Chemistry and Laboratory*

Medicine (CCLM) **2023**, *61*, 861–872, DOI: 10.1515/cclm-2022-0894, https://www.degruyter.com/document/doi/10.1515/cclm-2022-0894/html (cit. on p. 40).

[18] Majumdar, S.; Kalamkar, S. D.; Dudhgaonkar, S.; Shelgikar, K. M.; Ghaskadbi, S.; Goel, P. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full (cit. on pp. 44, 48–61).

[19] Salam, M.; Bailey, R.; Calhoun, P.; McGill, J. B.; Bergenstal, R. M.; Price, D.; Beck, R. W. A comparison of continuous glucose monitoring estimated hemoglobin A1c in adults with type 1 or type 2 diabetes. *Diabetes Technology & Therapeutics* **2023**, *25*, 178–185, DOI: 10.1089/dia.2022.0387, https://www.liebertpub.com/doi/10.1089/dia.2022.0387 (cit. on p. 44).

[20] Mondkar, S.; Khandagale, S.; Shah, N.; Khadilkar, A.; Oza, C.; Bhor, S.; Gondhalekar, K.; Wagle, A.; Kajale, N.; Khadilkar, V. Effect of metformin adjunct therapy on cardiometabolic parameters in Indian adolescents with type 1 diabetes: a randomized controlled trial. *Frontiers in Clinical Diabetes and Healthcare* **2024**, *5*, 1353279, DOI: 10.3389/fcdhc.2024.1353279, https://www.frontiersin.org/journals/clinical-diabetes-and-healthcare/articles/10.3389/fcdhc.2024.1353279/full (cit. on p. 46).

Chapter 4

Estimating continuous BG from CGM traces

An earlier version of this work was published as an e-poster titled **"A comparison of algebraic and dynamic models to estimate blood glucose and A1C from CGM time series"** in the 16th International Conference on Advanced Technologies & Treatments for Diabetes. (EP165/#606) The abstract book for the event can be found here (DOI) https://doi.org/10.1089/dia.2023.2525.abstracts and the abstract author index can be found here (DOI) https://doi.org/10.1089/dia.2023.2526.abstracts.index. An earlier version of this work was also presented as an oral presentation titled **"How to track blood glucose from the CGM?"** at the 7th International Diabetes Summit 2023 (IDS 2023) organized by Chellaram Diabetes Institute. The website for the conference is provided here https://www.cdi.org.in/international-diabetes-summit/.

4.1 Introduction

Diabetes is a metabolic disorder increasingly claiming a larger proportion of the world population each year. Both type 1 diabetes (T1D) and type 2 diabetes (T2D) patient numbers are increasing each year and this calls for construction of better tools for clinical diagnosis and management of diabetes. Traditionally, the primary tool for diagnosing both T1D and T2D involved measuring the glucose

concentration in the patients blood, blood glucose concentration (BG). For T2D, BG was measured before the first meal or after a long period of no food intake as the fasting blood glucose concentration (FBG) and was measured after a meal as the post-prandial blood glucose concentration (PPBG). For T1D patients, the BG had to be regularly evaluated/measured with a complex self-monitoring blood glucose (SMBG) routine. Recent diabetes patients have many alternatives, like fast glucose monitoring (FGM) devices, continuous glucose monitoring (CGM) devices, insulin pumps, etc. These modern devices reliably provide the user with various metrics to evaluate and manage BG and their overall general health. The CGM devices are recently being extensively used by diabetes patients to manage their day-to-day BG. Although, these devices provide a great service, they work with modern metrics which have not completely been incorporated into a standard clinical framework. For example, the CGM device provides the user with a sequence of interstitial fluid glucose concentration (ISF) measurements and either complex algorithms are used to convert this sequence into a sequence of BG values or it is assumed to be reliably close to the corresponding BG trace that the BG could be substituted by the sequence of ISF measurements. This conversion or substitution if not completed properly can give rise to a myriad of issues. This is still an active field of study.

Over the years there have been many attempts to estimate a patients BG using a non-invasive method. Rebrin et al. [3] studied the relation between BG and ISF in dogs using a 3 day CGM sensor and concluded that the insulin concentration in the blood does not influence the BG—ISF dynamics and also concluded that the delay in ISF can be corrected using a digital filter. Rebrin and Steil [4] later showed that the differences between ISF and BG were not significant concluded that ISF could be used as a substitute for BG. Thennadil et al. [6] studied the relationship between dermal ISF, capillary and venous BG for diabetes patients with rapidly changing BG levels and they observed that there are no significant lag between the ISF and BG values. Stout et al. [5] found an average time lag of about 25 min between the ISF and finger-stick BG measurement but concluded that although ISF can be used as an alternative for BG for patients routinely monitoring their glucose; more work is needed to know the relation between ISF and BG. Cobelli

et al. [12] also showed that the relationship of the ISF and the BG is complicated and cannot just be explained using a pure time shift. Kulcu et al. [7] showed that there is a difference between the ISF values and the corresponding BG values both in magnitude and delay. They concluded that the magnitude of these difference depend on the measuring devices but they are significant enough to show that there is a clear physiological difference between the ISF and BG dynamics. It is also true that different CGM devices use different algorithm for providing an estimate of the ISF from their measurement and so their calibration plays an important role in reliably measuring the ISF. Sun et al. [17] in a manuscript compiled all the calibration algorithms that have been formulated for the reliable measurement of ISF by CGM sensors so that it can reliably represent BG. Keenan et al. [11] introduced a time-lag correction for ISF so as to obtain a reliable estimate of the corresponding BG. There have been many attempts at understanding the BG–ISF dynamics and in recent times the use of machine learning algorithms have also become very popular. Miranda et al. [16] raises the question about the need for such algorithms and if they can reliably used to understand the ISF–BG dynamics.

CGM devices essentially measure the glucose concentration in the interstitial fluid present just under the skin, at its point of attachment. The measurements are done periodically and a sequence of ISF measurements are made available which can be considered as a discrete representation of the continuous ISF. This discrete representation of the continuous ISF is quite useful and widely used now a days in insulin pumps for T1D patients and clinically for diagnosis and management of T1D and T2D.

The traditional methods used for diagnosis and management of diabetes are generally based on the BG. This would suggest that the continuous BG trajectory would provide a better clinical tool for diagnosis and management of diabetes. As, it would allow the clinician to leverage the available research on the glucose concentration in the blood and its relationship with diabetes. The CGM devices however only measure ISF and only provide a discrete representation of the continuous ISF trajectory. As a majority of the measured ISF values are within a certain acceptable range of the corresponding BG values at the instant of measurement, it is argued that the discrete ISF trace can be considered as a reliable representation

of the continuous BG.

Although, the measured ISF values may be close to the corresponding BG at a time of measurement, it does not necessarily mean that the continuous ISF curve would have the same relationship with the corresponding continuous BG curve at all times. In 2016, Cobelli et al. [12] showed that the relationship between the measured ISF and the BG is complicated and could not be explained by a simple shifting and/or scaling model. It is therefore, important to study new methods which would allow us generate a reliable estimate of the continuous ISF and BG curves from a given sequence of ISF measurements (a CGM trace).

We intend to provide an estimate of the continuous ISF and BG curves from any given CGM trace using trajectory optimization. Estimating the continuous ISF and BG curves would require us to construct a simple model which would reliably describe the dynamical relationship between ISF and BG. This model describing the dynamical relationship between ISF and BG would have to follow certain a priory known constraints that govern the BG and ISF dynamics.

Now, the glucose concentration in the interstitial fluid is essentially fueled by the glucose molecules from the blood. We suspect that the glucose molecules diffuse into the interstitial fluid from the blood through the blood vessels. Therefore, the glucose present in the blood flowing inside these blood vessels act as a source of glucose for the interstitial fluid.

On the other hand the glucose molecules in the interstitial fluid act as fuel for the nearby cells. Therefore, the nearby cells act as sinks for glucose from the interstitial fluid. These factors are to be taken into account when developing a dynamical model for glucose concentration in the blood (BG) and interstitial fluid (ISF).

We introduce a simple model based on first principles that describe the relationship between BG and ISF. Let $\mathrm{BG}\,(t)$ and $\mathrm{ISF}\,(t)$ represent the continuous BG and ISF curves respectively at time $t \in \mathbb{R}^+$ and let $\mathrm{ISF}^{\mathrm{obs}}\,(t_k)$ represent the measured ISF values provided by the CGM device at times $t_k \in \mathbb{R}^+$, $k \in \{1, 2, \ldots, N\}$. Here, N represents the total number of measurements in a CGM trace. We introduce a simple model based on first principles that should adequately describe the relationship between BG and ISF. Our goal is to provide estimates $\mathrm{BG}^{\mathrm{est.}}\,(t)$ and

$\text{ISF}^{\text{est.}}(t)$ of the continuous ISF and BG, with a suitable degree of smoothness, such that the total ordinary least squares (OLS) deviation of $\text{ISF}^{\text{est.}}$ from ISF is minimized i.e. the following function is minimized.

$$J(\text{ISF},\ \text{BG}) = \left(\int\limits_{\substack{\text{period of} \\ \text{wearing a} \\ \text{CGM device}}} \left(\text{ISF}(\tau) - \text{ISF}^{est.}(\tau)\right)^2 d\tau \right)^{\frac{1}{2}},$$

The complete problem can be written down, with the function f describing the derivative of ISF, as the following optimal control problem

$$\min_{\text{ISF,BG}} \quad J(\text{ISF},\ \text{BG}) = \int_{t_B}^{t_F} \left(\text{ISF}(\tau) - \text{ISF}^{est.}(\tau)\right)^2 d\tau, \tag{4.1a}$$

$$\text{subject to} \quad \frac{d}{dt}\text{ISF}(t) = f(\text{ISF}(t), \text{BG}(t)), \qquad \forall t \in [t_B, t_F], \tag{4.1b}$$

$$\text{ISF}(t) - \text{BG}(t) \leq 0, \qquad \forall t \in [t_B, t_F], \tag{4.1c}$$

$$30\,\text{mg/dL} \leq \text{ISF}(t) \leq 600\,\text{mg/dL}, \qquad \forall t \in [t_B, t_F], \tag{4.1d}$$

$$30\,\text{mg/dL} \leq \text{BG}(t) \leq 600\,\text{mg/dL}, \qquad \forall t \in [t_B, t_F], \tag{4.1e}$$

We intend to obtain an estimate of BG from the measured ISF using trajectory optimization. Trajectory optimization problems are essentially optimal control problems where an optimal trajectory is chosen which minimizes or maximizes a given objective function while also satisfying a series of equality and inequality constraints. These problems are fundamentally optimization problems in the infinite dimensional continuous function spaces. However any real world problem of this type is almost always (if it can not be solved analytically) solved by finding an approximation in a finite dimensional function space. [2]

This approximation of the solution of the optimal control problem can be achieved (generally) using two procedures [1, 2]. The first one involves constructing the set of constraints that any candidate trajectory must satisfy in order to be considered a possible optimal solution. These constraints are then discretized and the resulting equations solved. This procedure is called an indirect method. The second method, known as the direct method, involves discretizing the original

optimal control problem which generates a discrete optimization problem, usually an nonlinear programming (NLP) problem. This is then usually solved using standard NLP optimization methods.

The problem we intend to solve is essentially an optimal control problem, where an optimal trajectory BG needs to be constructed which obeys a given dynamical system, our simple model, and generates the appropriate ISF trajectory. We use the direct method approach where we discretize the optimal control problem and find an approximate to the original BG trace from a finite dimensional functional space.

4.2 Materials and methods

4.2.1 The Dataset

The Pune-2021 dataset [15] was collected at the Primary Care Health Centre, Savitribai Phule Pune University, Pune. It contained CGM traces and various other bio-markers from 50 participants. This dataset was first used in Majumdar et al. [15] to establish two new methods of estimating Glycated Hemoglobin (HbA1c) from any given CGM trace. The Abbott Free Style Libre Pro device was used to collect the CGM traces. The CGM traces contain ISF measurements in mg/dL intervals of 14 min. The CGM devices were attached to the participants for 14 days. The Pune-2021 dataset contained 50 participants out of which 15 were pre-diabetic, 17 were diabetic, and 20 were non-diabetic. The dataset will be referred to as the Pune-2021 dataset as it was done in Majumdar et al. [15]. All our analysis will be based on CGM traces from the Pune-2021 dataset.

4.2.2 A simple model describing ISF—BG dynamics

We construct a simple model by taking into consideration the constraints and dynamics described in the Introduction (Section 4.1). This model would be a simple diffusion model where glucose is diffusing into and out of the interstitial fluid from the blood vessels and nearby cells respectively. We consider the rates of glucose diffusion into the interstitial fluid to be different from the rates of glucose

diffusing out from the interstitial fluid into the surrounding cells. The equation for this simple model is given below

$$\frac{d}{d\tau}\mathrm{ISF}\left(\tau\right) = \tilde{b}\mathrm{BG}\left(\tau\right) - \tilde{a}\mathrm{ISF}\left(\tau\right), \tag{4.2}$$

where, $\mathrm{ISF}\left(\tau\right)$ represents the continuous interstitial fluid glucose concentration (ISF), $\mathrm{BG}\left(\tau\right)$ represents the continuous blood glucose concentration (BG), $\tilde{a}$ and, $\tilde{b}$ represents the coefficients for their contributions into the rate of change of ISF respectively. The values for the coefficients $\tilde{a}$ and $\tilde{b}$ were decided based on the scaling factor used by the adapted Nathan model to estimate average blood glucose concentration (aBG) from average interstitial fluid glucose concentration (aISF) in Majumdar et al. [15]. Their values were chosen such that the steady state ISF and BG values for Eq. (4.2) satisfied the following scaling equation used by the adapted Nathan model [15].

$$\mathrm{BG}^{ss}_{\mathrm{mg/dL}} = 1.17 \times \mathrm{ISF}^{ss}_{\mathrm{mg/dL}}, \tag{4.3}$$

where, $\mathrm{ISF}^{ss}_{\mathrm{mg/dL}}$ represents the steady state interstitial fluid glucose concentration (ISF) in mg/dL and $\mathrm{BG}^{ss}_{\mathrm{mg/dL}}$ represents the steady state blood glucose concentration (BG) in mg/dL. We considered $\tilde{a} = 1\,\mathrm{min}^{-1}$ and $\tilde{b} = \frac{1}{1.17}\,\mathrm{min}^{-1}$.

Therefore, the simple model, Eq. (4.2), could be represented by the following equation

$$\frac{d}{d\tau}\mathrm{ISF}_{\mathrm{mg/dL}}\left(\tau\right) = \frac{1}{1.17}\mathrm{BG}_{\mathrm{mg/dL}}\left(\tau\right) - \mathrm{ISF}_{\mathrm{mg/dL}}\left(\tau\right), \tag{4.4}$$

where, $\mathrm{ISF}_{\mathrm{mg/dL}}$, just as previously stated, represents the interstitial fluid glucose concentration (ISF) in mg/dL and, $\mathrm{BG}_{\mathrm{mg/dL}}$ represents the blood glucose concentration (BG) in mg/dL.

Now, Eq. (4.4) describes the relationship between ISF and BG with the unit mg/dL for concentration and minutes for time. In order to obtain a non-diamentionalized equation, we assume $\mathrm{ISF}_{\mathrm{mg/dL}} = \mu_{\mathrm{mg/dL}} \times \mathrm{ISF}$, $\mathrm{BG}_{\mathrm{mg/dL}} = \mu_{\mathrm{mg/dL}} \times \mathrm{BG}$ and, $\tau = \tau_0 \times t$ where, ISF, BG and, t are dimensionless quantities and $\mu_{\mathrm{mg/dL}}$ is a constant in mg/dL and τ_0 a constant in minutes. The non-dimensionalized version of Eq. (4.4) can now be written as follows

$$\frac{d}{dt}\mathrm{ISF}\,(t) = \tau_0 \left(\frac{1}{1.17}\mathrm{BG}\,(t) - \mathrm{ISF}\,(t) \right), \tag{4.5}$$

Here, τ_0 acts as a scaling factor and manipulates the time-scale of BG–ISF dynamics. We begin our analysis by considering $\tau_0 = 1\,\mathrm{min}$ but we later consider different values of τ_0 to study at which time-scale Eq. (4.5) accurately represents the BG–ISF dynamics. For $\tau_0 = 1\,\mathrm{min}$, Eq. (4.5) can be represented by

$$\frac{d}{dt}\mathrm{ISF}\,(t) = \frac{1}{1.17}\mathrm{BG}\,(t) - \mathrm{ISF}\,(t)\,, \tag{4.6}$$

where, $\mathrm{ISF}\,(t)$ represents the non-dimentionalized interstitial fluid glucose concentration (ISF) and similarly, $\mathrm{BG}\,(t)$ represents the non-dimentionalized blood glucose concentration (BG).

4.2.3 Trajectory optimization

Trajectory optimization is a method developed to solve optimal control problems where we deduce a suitable trajectory (for any dynamic quantity), based on the minimization or maximization of an objective function/cost functional, such that the trajectory obeys the dynamics described by a given set of system equations and also satisfies an additional given set of constraints. The method essentially solves an optimization problem where the optimal trajectory minimizes or maximizes an objective function/cost functional while also obeying a set of constraints constructed from the differential equations describing the dynamical system and some other additional constraints.

A typical trajectory optimization problem as described by Kelly [13] would involve a set $\mathcal{X}$ of all desirable trajectories and a set $\mathcal{U}$ of all desirable control or driving functions over the time-interval $\mathcal{T} \equiv [t_B, t_F] \subset \mathbb{R}^+$. It would also involve an objective function or, a cost functional (henceforth we will refer to this function as the objective function) which would typically depend on the time-stamps $t_B \in \mathcal{T}$ and $t_F \in \mathcal{T}$, the trajectory $x \in \mathcal{X}$, the control or driving function (henceforth we will refer to this function as the control function) $u \in \mathcal{U}$, and is usually comprised of two functions, the first function J_0, the meyer term, representing the cost due to the choice of end-points, $(t_B, x(t_B), u(t_B))$ and $(t_F, x(t_F), u(t_F))$, and the second

function J_w, the Lagrange term, representing the cost incurred as the trajectory $x \in \mathcal{X}$ is traversed. The objective function could therefore be written as

$$J\left(t_B, t_F, x, u\right) = \underbrace{J_0\left(t_B, t_F, x(t_B), x(t_F), u(t_B), u(t_F)\right)}_{\text{Mayer term}} + \overbrace{\int_{t_B}^{t_F} J_w\left(\tau, x\left(\tau\right), u\left(\tau\right)\right) d\tau}^{\text{Lagrange term}}, \tag{4.7}$$

The problem would also involve a series of equality and inequality constraints on the state and control trajectories, known as the path constraints, represented by $g_j = 0 \ \forall j \in \{1, \ldots, m\} \equiv \mathbb{N}_m \in \mathbb{N}$ (henceforth $\mathbb{N}_n$ would represent the set $\{1, 2, \ldots, n-1, n\}$ for any $n \in \mathbb{N}$) and $h_i \leq 0 \ \forall i \in \mathbb{N}_n$ respectively. These constraints could be written as

$$g_j\left(x\left(t\right), u\left(t\right)\right) = 0, \quad \forall t \in \mathcal{T}, \ j \in \mathbb{N}_m, \tag{4.8}$$

$$h_i\left(x\left(t\right), u\left(t\right)\right) \leq 0, \quad \forall t \in \mathcal{T}, \ i \in \mathbb{N}_n, \tag{4.9}$$

$$\tag{4.10}$$

Typically the state and control trajectories would also be bounded and these boundary constraints written as the following equations below, would also be considered as constraints

$$x_{\text{low}}\left(t\right) \leq x\left(t\right) \leq x_{\text{up}}\left(t\right), \quad \forall t \in \mathcal{T}, \tag{4.11}$$

$$u_{\text{low}}\left(t\right) \leq u\left(t\right) \leq u_{\text{up}}\left(t\right), \quad \forall t \in \mathcal{T}, \tag{4.12}$$

where, $x_{\text{low}}\left(t\right)$ and $x_{\text{up}}\left(t\right)$ represents the lower and upper bounds of the state trajectory at $t \in \mathcal{T}$ and similarly, $u_{\text{low}}\left(t\right)$ and $u_{\text{up}}\left(t\right)$ represents the lower and upper bounds of the control trajectory at $t \in \mathcal{T}$.

Finally, a typical trajectory optimization problem would involve a dynamical system which would govern the evolution of the desired trajectories. The differential equations describing the system dynamics could be written as

$$\frac{d}{dt}x\left(t\right) = f\left(t, x\left(t\right), u\left(t\right)\right) \quad \forall t \in \mathcal{T}, x \in \mathcal{X}, u \in \mathcal{U} \tag{4.13}$$

Therefore, taking everything together, a typical trajectory optimization problem represented by Eqs. (4.7)–(4.13) could be described using the following optimization problem,

$$\min_{\substack{t_0,t_F\in\mathcal{T},\\ x\in\mathcal{X},\\ u\in\mathcal{U}}} \quad J\left(t_0,t_F,x(t_0),x(t_F),u(t_0),u(t_F)\right)+\int_{t_o}^{t_F} w\left(\tau,x\left(\tau\right),u\left(\tau\right)\right)d\tau, \tag{4.14a}$$

$$\text{subject to} \quad \frac{d}{dt}x\left(t\right)=f\left(t,x\left(t\right),u\left(t\right)\right), \qquad \forall t\in\mathcal{T}, \tag{4.14b}$$

$$h_i\left(x\left(t\right),u\left(t\right)\right)\leq 0, \qquad \forall t\in\mathcal{T},\ i\in\mathbb{N}_n, \tag{4.14c}$$

$$g_j\left(x\left(t\right),u\left(t\right)\right)=0, \qquad \forall t\in\mathcal{T},\ j\in\mathbb{N}_m, \tag{4.14d}$$

$$x_{\text{low}}\left(t\right)\leq x\left(t\right)\leq x_{\text{up}}\left(t\right), \qquad \forall t\in\mathcal{T}, \tag{4.14e}$$

$$u_{\text{low}}\left(t\right)\leq u\left(t\right)\leq u_{\text{up}}\left(t\right), \qquad \forall t\in\mathcal{T}, \tag{4.14f}$$

Here, Eqs. (4.14) represents a typical trajectory optimization problem for continuous trajectories where the goal is to find the optimal $\hat{x}\in\mathcal{X}$ and $\hat{u}\in\mathcal{U}$ trajectories which minimizes the cost function J, given by Eq. (4.14a) while satisfying the constraints given by Eqs. (4.14b)–(4.14f).

The above trajectory optimization problem if solvable could be used to deduce a suitable continuous state and control trajectory with a suitable degree of smoothness such that the objective function, J, is minimized. In most cases even if such optimal state and control trajectory solutions exists, it might not be possible to obtain an explicit analytical form of the continuous trajectories. Therefore, a numerical approach is generally used.

A numerical approach to the direct solution of a trajectory optimization problem like Eq. (4.14) as described by Kelly [13] would require us to transcribe the equations describing the problem into their discrete counter parts. There are two popular ways in which this transcription can be done. The first is direct transcription where the trajectories, constraints and, differential equations describing the optimization problem are discretized into a discrete NLP problem, which can then be solved using traditional NLP methods. The second is indirect transcription, where the necessary conditions that need to be satisfied by a trajectory in order to

be a solution to the optimization problem are constructed. The functions, operators and differential equations involved in these necessary condition constraints are then discretized and the discrete counterparts to these necessary condition constraints are obtained. These discrete equations are then solved to obtain a discrete solution to the optimization problem.

Generally speaking, the indirect transcription methods lead to solutions which are better approximations of the continuous solutions of the optimization problem but these methods are difficult to solve reliably when the problem has a large dimension. The direct transcription methods usually are able to provide solutions which are reliable approximations of the continuous-time solutions of the optimization problem. We use direct transcription methods, in particular Trapezoidal (as described in Kelly [13]) and a modified Hermite-Simpson collocation method (where the Hermite-Simpson method is as described in Kelly [13]) to discretize our trajectory optimization problem into a NLP problem.

4.2.4 Collocation methods

The solution to an optimization problem, a differential or an integral equation may belong to the space of continuous functions, one of its subspaces or, some similar function space like the Lebesgue space, containing functions acting on a particular time interval. These solutions, for the optimization problem, by definition should satisfy all the constraints involved in the problem as well as minimize (or maximize) the objective function. For the differential and integral equation, the solution only satisfies the equation and any other constraints involved. Many of these optimization problems, differential or integral equation problems, may not be analytically solvable. Therefore a numerical approach must be used where the actual solution from the appropriate function space is approximated by a function from a finite dimensional space. Collocation methods allow us to approximate such solutions which are functions from infinite-dimensional spaces by using functions from finite-dimensional functional spaces. These functions satisfy the constraints and the equations involved in the problem only at finite points at their domain. [see Chapter. 1 8]

These finite-dimensional spaces are called collocation spaces and the finite num-

ber of points in the domain of the functions are called the collocation points. Therefore a collocation solution for a functional equation is a member from a finite dimensional functional space which satisfies all the constraints, differential equations and, maximizes or minimizes the objective function (for optimization problems) involved in the problem at some finite collocation points in their domain.

For a trajectory optimization problem we first select the collocation points on the function domain. For some cases these points can be chosen appropriately but in many cases, circumstances are such that only a few sets of collocation points can be considered. In our particular problem ideally we want to obtain a continuous ISF and BG trajectory from the sets $C^m(\mathcal{T}, \mathbb{R})$, with an appropriate $m \in \mathbb{N}$ and on an appropriate domain $\mathcal{T} \equiv [t_B, t_F]$ where t_B represents the time the CGM device was attached and t_F represents the time when the CGM device was removed. We have CGM traces which are essentially a series of ISF measurements taken at an interval of 15 min for 14 days. We therefore consider this 14 day time-period as our domain $[t_B, t_F]$ and are forced to consider collocation points in this domain which are 15 min apart and coincide with the time the CGM device took the ISF measurement.

Let $\mathcal{T}_{N,\,n}$ represent a mesh on $\mathcal{T} \equiv [t_B, t_F]$ where,

$$\mathcal{T}_{N,\,n} \equiv \left\{ t_j + c_i h_j \;\middle|\; \begin{array}{l} t_B = t_0 < t_1 < \cdots < t_N < t_{N+1} = t_F, \\ h_j = t_{j+1} - t_j, \\ 0 \leq j \leq N, \\ 0 \leq c_1 < c_2 < \cdots < c_i < \cdots < c_{n-1} < c_n \leq 1 \end{array} \right\}$$

and let $s_j = [t_j, t_{j+1}\,)\ \forall\ j \in \{0\} \cup N_{N-1},\ s_N = [t_N, t_{N+1}]$, with the $h = \max_j \{h_j | 1 \leq j \leq N-1\}$. Taking $\mathcal{T}_{N,\,n}$ as the collocation points we want to approximate the solution trajectories, say u^* and x^* of the optimization problem by a function from the polynomial spline spaces $\mathcal{S}(\mathcal{P}_{m_u}; \mathcal{M}_u; \mathcal{T}_{N,\,n})$ for u^* and $\mathcal{S}(\mathcal{P}_{m_x}; \mathcal{M}_x; \mathcal{T}_{N,\,n})$ for x^* respectively such that, at the collocation points the estimates satisfies all the constraints from the trajectory optimization problem.

Given $\mathcal{T}_{N,\,n}$ as a partition of the domain time-interval $[t_B, t_F]$ with the sub intervals labeled as $s_j \equiv [t_j, t_{j+1}\,),\ j \in \{0\} \cup \mathbb{N}_{N-1},\ s_N \equiv [t_N, t_{N+1}]$ and $\mathcal{M} \equiv$

$(\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_N)$ with $\mathcal{M}_i \in \mathbb{Z}^+$, $1 \leq \mathcal{M}_i \leq m$, $i \in \mathbb{N}_N$, we can define the space of polynomial splines $\mathcal{S}(\mathcal{P}_m; \mathcal{M}; \mathcal{T}_{N,n})$ [see 9, for definition] as

$$\mathcal{S}(\mathcal{P}_m; \mathcal{M}; \mathcal{T}_{N,n}) \equiv \left\{ p \;\middle|\; \begin{array}{l} \exists p_0, p_1, \ldots, p_N \in \mathcal{P}_m, \\ p(x) = p_i(x) \;\; \forall x \in s_i, \\ \mathcal{D}^j p_{i-1}(x_i) = \mathcal{D}^j p_i(x_i) \\ \quad \forall i \in \mathbb{N}_N, \;\; \forall j \in \{0\} \cup \mathbb{N}_{m-1-\mathcal{M}_i} \end{array} \right\},$$

where $\mathcal{P}_m$ represents the space of polynomials of order m and where $\mathcal{M} \equiv (\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_N)$ represents the multiplicities of the collocation points t_i, $i \in \mathbb{N}_N$ in $\mathcal{T}_{N,n}$. We generally introduce n collocation points in each of the $N+1$ sub intervals of $T_{N,\,n}$ but this varies according to the constraints that the approximating function needs to satisfy. For cases where $\mathcal{M}_i = c \;\forall i \in \mathbb{N}_N$ for some constant $c \in \mathbb{N}$, we will denote the space $\mathcal{S}(\mathcal{P}_m, \mathcal{M}, \mathcal{T}_{N,n})$ as $\mathcal{S}(\mathcal{P}_m, c, \mathcal{T}_{N,n})$ and, when $c = 1$ we will denote the space $\mathcal{S}(\mathcal{P}_m, c, \mathcal{T}_{N,n})$ as $\mathcal{S}_m(\mathcal{T}_{N,n})$ [following same convention as 9].

The collocation solutions from the finite dimensional spaces would now need to satisfy the equations and constraints involved in the problem, now called collocation equations, only at the collocation points in $\mathcal{T}_{N,\,n}$. We consider only two collocation methods in this manuscript, they are described below.

Trapezoidal collocation

The Trapezoidal collocation method is one of the most commonly used collocation methods. In this method, we generally consider collocation points of the form $\mathcal{T}_N$ ($\equiv \mathcal{T}_{N,2}$ with $c_1 = 0$ and $c_n = c_2 = 1$ in each sub interval) over $N+1$ sub intervals in the domain $\mathcal{T} \equiv [t_B, t_F]$ where,

$$\mathcal{T}_N \equiv \{t_n \in \mathcal{T} | t_B = t_0 < t_1 < \cdots < t_{N-1} < t_N < t_{N+1} = t_F\},$$

the control trajectory is interpolated by a linear spline [13] where the collocation space is $\mathcal{S}_1(\mathcal{T}_N)$ and the state trajectory is interpolated by a quadratic spline [13] where the collocation space is $\mathcal{S}_3(\mathcal{T}_N)$.

Let us consider an optimal control problem, very similar to our problem as described in Section. 4.1.

$$\min_{\substack{x\in\mathcal{X},\\ u\in\mathcal{U}}} \quad J = \int_{t_B}^{t_F} J_w\left(x\left(\tau\right), u\left(\tau\right)\right) d\tau, \tag{4.15a}$$

$$\text{subject to} \quad \frac{d}{dt}x\left(t\right) = f\left(x\left(t\right), u\left(t\right)\right), \qquad \forall t \in \mathcal{T}, \tag{4.15b}$$

$$h_i\left(x\left(t\right), u\left(t\right)\right) \leq 0, \qquad \forall t \in \mathcal{T},\ i \in N_n, \tag{4.15c}$$

$$g_j\left(x\left(t\right), u\left(t\right)\right) = 0, \qquad \forall t \in \mathcal{T}, j \in N_m, \tag{4.15d}$$

$$x_{\text{low}}\left(t\right) \leq x\left(t\right) \leq x_{\text{up}}\left(t\right), \qquad \forall t \in \mathcal{T}, \tag{4.15e}$$

$$u_{\text{low}}\left(t\right) \leq u\left(t\right) \leq u_{\text{up}}\left(t\right), \qquad \forall t \in \mathcal{T}, \tag{4.15f}$$

We take the set $\mathcal{T}_N \in \mathcal{T}$ to be the set of N collocation points in $\mathcal{T}$ with $N+1$ sub intervals. We then assume the state trajectories at collocation points $t_k \in \mathcal{T}_N\ \forall k \in \{0\} \cup \mathbb{N}_{N+1}$ to be represented by $x_k = x\left(t_k\right)$ and similarly the control trajectory by $u_k = u\left(t_k\right)$. In each sub interval $[t_k, t_{k+1}] \in \mathcal{T}$, the state trajectory, $x\left(t\right)$ and the control trajectory, $u\left(t\right)$ can be represented by a quadratic polynomial and a straight line respectively. Trapezoidal collocation also imposes that the derivatives of the state trajectory remain continuous for the interval $\mathcal{T}$. Now, for the sub interval $[t_k, t_{k+1}]$, we can represent the state trajectory by the following equations

$$x\left(t\right) = a_k(t - t_k)^2 + b_k\left(t - t_k\right) + c_k, \tag{4.16}$$

$$\dot{x}\left(t\right) = 2a_k\left(t - t_k\right) + b_k \tag{4.17}$$

To find the expression of x above we need to identify a_k, b_k and, c_k. If we can identify, in total, three values of either x or $\dot{x}$ in this interval we can identify the above parameters. Now, taking into consideration that $\dot{x}$, and subsequently x, need to be continuous in the interval $\mathcal{T}$ we have the following equations [refer to Section. 2.1 14, Section. 3.2 10, 1, Sections. 1, 3 13, 2, for details of the calculation]

$$x(t_k) = x_k \tag{4.18}$$

$$\dot{x}(t_k) = \dot{x}_k \tag{4.19}$$

$$\dot{x}(t_{k+1}) = \dot{x}_{k+1} \tag{4.20}$$

$$\tag{4.21}$$

Using the above equations we can identify the parameters a_k, b_k and, c_k as [see 1, 2, 10, 13, 14]

$$a_k = \frac{\dot{x}_{k+1} - \dot{x}_k}{2h_k} \tag{4.22}$$

$$b_k = \dot{x}_k \tag{4.23}$$

$$c_k = x_k \tag{4.24}$$

$$\tag{4.25}$$

Now, we also need the state and control trajectories to satisfy the dynamical system equations and therefore we have

$$\dot{x}_k = f(x_k, u_k) \tag{4.26}$$

$$\dot{x}_{k+1} = f(x_{k+1}, u_{k+1}) \tag{4.27}$$

Then combining the previous equations we have

$$x_{k+1} = x_k + \frac{h_k}{2}(f_{k+1} + f_k) \tag{4.28}$$

where f_k represents $f(x_k, u_k)$ and f_{k+1} represents $f(x_{k+1}, u_{k+1})$. This equation ensures that the state trajectory in the subinterval $[t_k, t_{k+1}]$ is represented by a piecewise quadratic polynomial and not only satisfy the system equation but also is equipped with a first derivative on $\mathcal{T}$.

The objective function with the state and control trajectories discretized at the collocation points $\mathcal{T}_N$ can be written as

$$\hat{J} = \sum_{k=0}^{N-1} \frac{h_k}{2}(J_{w,\ k+1} + J_{w,\ k}), \tag{4.29}$$

where $J_{w,\ k} = J_w\left(x_k, u_k\right)\ \forall k \in \{0\} \cup \mathbb{N}_{N+1}$. The path and boundary constraints for the discretized state and control trajectories can also be written as

$$h_{k,i} \equiv h_i\left(x_k, u_k\right) \leq 0, \quad \forall t_k \in \mathcal{T}_N,\ i \in \mathbb{N}_n, \tag{4.30}$$

$$g_{k,j} \equiv g_j\left(x_k, u_k\right) = 0, \quad \forall t_k \in T_N,\ j \in \mathbb{N}_m, \tag{4.31}$$

$$x_k^{\text{low}} \leq x_k \leq x_k^{\text{up}}, \quad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.32}$$

$$u_k^{\text{low}} \leq u_k \leq u_k^{\text{up}}, \quad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.33}$$

The problem given by Eq. (4.15) now boils down to an NLP problem,

$$\min_{x_k, u_k} \quad \hat{J} = \sum_{k=0}^{N-1} \frac{h_k}{2}\left(J_{w,\ k+1} + J_{w,\ k}\right) \tag{4.34a}$$

$$\text{subject to} \quad x_{k+1} = x_k + \frac{h_k}{2}\left(f_{k+1}, f_k\right), \quad \forall k \in \{0\} \cup \mathbb{N}_N, \tag{4.34b}$$

$$h_{k,\ i}\left(x_k, u_k\right) \leq 0, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1},\ i \in \mathbb{N}_n, \tag{4.34c}$$

$$g_j\left(x_k, u_k\right) = 0, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, j \in \mathbb{N}_m, \tag{4.34d}$$

$$x_k^{\text{low}} \leq x_k \leq x_k^{\text{up}}, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.34e}$$

$$u_k^{\text{low}} \leq u_k \leq u_k^{\text{up}}, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.34f}$$

where, the points x_k and $u_k\ \forall\ k \in \{0\} \cup \mathbb{N}_{N+1}$ need to be estimated such that they satisfy Eqs. (4.34b)–(4.34f) and minimizes $\hat{J}$ (from Eq. (4.34a)).

Modified Hermite-Simpson collocation

The Hermite-Simpson collocation method is another popular collocation method. In this collocation method we generally consider the set of collocation points of the form $\mathcal{T}_{N,3}$ with $N+1$ subintervals in the time domain $\mathcal{T} \equiv [t_B, t_F]$ where,

$$\mathcal{T}_{N,3} \equiv \left\{ t_n \in \mathcal{T} | t_B = t_0 < t_{\frac{1}{2}} < t_1 < \cdots < t_N < t_{N+\frac{1}{2}} < t_{N+1} = t_F \right\},$$

In the Hermite-Simpson collocation method the continuous state trajectory is approximated by a cubic Hermite spline [13] where the collocation space is

$\mathcal{S}(\mathcal{P}_4, 2, \mathcal{T}_{N,3})$ and the continuous control trajectory is approximated by a quadratic Hermite spline [13] where the collocation space is $\mathcal{S}(\mathcal{P}_3, 2, \mathcal{T}_{N,3})$. The differential equations representing the system dynamics, the operations in the trajectory constraints and, the objective function is then approximated using the Simpson quadrature at the collocation points given by $\mathcal{T}_{N,3}$.

Let us again consider an optimal control problem, very similar to our problem as described in Section. 4.1.

$$\min_{\substack{x\in\mathcal{X},\\ u\in\mathcal{U}}} \quad J = \int_{t_B}^{t_F} J_w\left(x(\tau), u(\tau)\right) d\tau, \tag{4.35a}$$

$$\text{subject to} \quad \frac{d}{dt}x(t) = f\left(x(t), u(t)\right), \qquad \forall t \in \mathcal{T}, \tag{4.35b}$$

$$h_i\left(x(t), u(t)\right) \leq 0, \qquad \forall t \in \mathcal{T},\ i \in \mathbb{N}_n, \tag{4.35c}$$

$$g_j\left(x(t), u(t)\right) = 0, \qquad \forall t \in \mathcal{T}, j \in \mathbb{N}_m, \tag{4.35d}$$

$$x_{\text{low}}(t) \leq x(t) \leq x_{\text{up}}(t), \qquad \forall t \in \mathcal{T}, \tag{4.35e}$$

$$u_{\text{low}}(t) \leq u(t) \leq u_{\text{up}}(t), \qquad \forall t \in \mathcal{T}, \tag{4.35f}$$

We take $\mathcal{T}_{N,3} \in \mathcal{T}$ to be the set of collocation points and then assume the state trajectories at collocation points $t_k \in \mathcal{T}_{N,3}\ \forall k \in \{0\} \cup \mathbb{N}_{N+1}$ to be represented by $x_k = x(t_k)$ and similarly the control trajectory by $u_k = u(t_k)$. As a cubic spline interpolation is performed an additional collocation point between all previous collocation points are required and we represent the collocation point between the k^{th} and $(k+1)^{\text{th}}$ collocation point as the $\left(k+\frac{1}{2}\right)^{\text{th}}$ collocation point, $t_{k+\frac{1}{2}}$, and represent the state and control trajectories at that collocation point as $x_{k+\frac{1}{2}} = x\left(t_{k+\frac{1}{2}}\right)$ and $u_{k+\frac{1}{2}} = u\left(t_{k+\frac{1}{2}}\right)$ respectively.

The state trajectory in the sub intervals $[t_k, t_{k+1}]\ \forall k \in \{0\} \cup \mathbb{N}_N$ is approximated by a cubic polynomial and it is also required that the first derivatives of the state trajectory is continuous over the time interval $\mathcal{T}$. We therefore can represent the state trajectory in the sub interval $[t_k, t_{k+1}]$ using the following equations

$$x(t) = a_k(t-t_k)^3 + b_k(t-t_k)^2 + c_k(t-t_k) + d_k \tag{4.36}$$

$$\dot{x}(t) = a_k(t-t_k)^2 + b_k(t-t_k) + c_k \tag{4.37}$$

Now, to identify the parameters a_k, b_k, c_k and, d_k we need four equations and using the continuity constraints on the state trajectory we have [calculations in 14, 10, 1, 13, 2]

$$x\left(t_k\right) = x_k \tag{4.38}$$
$$x\left(t_{k+1}\right) = x_{k+1} \tag{4.39}$$
$$\dot{x}\left(t_k\right) = \dot{x}_k \tag{4.40}$$
$$\dot{x}\left(t_{k+1}\right) = \dot{x}_{k+1} \tag{4.41}$$

Now solving for a_k, b_k, c_k and, d_k gives us

$$a_k = \frac{\dot{x}_{k+1} + \dot{x}_k}{h_k^2} - \frac{2\left(x_{k+1} - x_k\right)}{h_k^3} \tag{4.42}$$
$$b_k = -\frac{\dot{x}_{k+1} + 2\dot{x}_k}{h_k} + \frac{3\left(x_{k+1} - x_k\right)}{h_k^2} \tag{4.43}$$
$$c_k = \dot{x}_k \tag{4.44}$$
$$d_k = x_k \tag{4.45}$$

Now the state trajectory in the sub interval $[t_k, t_{k+1}]$ can be represented by x_k, x_{k+1}, $\dot{x}_k$ and, $\dot{x}_{k+1}$. We now calculate $x\left(t_{k+\frac{1}{2}}\right) \equiv x_{k+\frac{1}{2}}$ and its derivative in terms of x_k, x_{k+1}, $\dot{x}_k$ and, $\dot{x}_{k+1}$. This gives us

$$x_{k+\frac{1}{2}} = \frac{1}{2}\left(x_k + x_{k+1}\right) + \frac{h_k}{8}\left(\dot{x}_k - \dot{x}_{k+1}\right) \tag{4.46}$$
$$\dot{x}_{k+\frac{1}{2}} = \frac{3}{2h_k}\left(x_{k+1} - x_k\right) - \frac{1}{4}\left(\dot{x}_k + \dot{x}_{k+1}\right) \tag{4.47}$$

We also know that the state trajectories have to follow the state dynamics and therefore, at the collocation points we have

$$\dot{x}_k = f\left(x_k, u_k\right), \tag{4.48}$$
$$\dot{x}_{k+1} = f\left(x_{k+1}, u_{k+1}\right), \tag{4.49}$$
$$\dot{x}_{k+\frac{1}{2}} = f\left(x_{k+\frac{1}{2}}, u_{k+\frac{1}{2}}\right), \tag{4.50}$$

Now, we have obtained a representation of the state trajectory using x_k, x_{k+1}, $\dot{x}_k$ and, $\dot{x}_{k+1}$ but the constraints used do not verify if the trajectory has a continuous first derivative nor have we clearly imposed the constraint that the state trajectory obeys the system equations. We therefore impose the condition that the first derivative of the state trajectory at the collocation point $t_{k+\frac{1}{2}}$ obeys the system dynamics equations. We then arrive at the equation

$$x_{k+1} = x_k + \frac{h_k}{6}\left(f_k + 4f_{k+\frac{1}{2}} + f_{k+1}\right) \tag{4.51}$$

In the traditional Hermite-Simpson collocation method we interpolate the control trajectory with a quadratic Hermite spline. However, as we would be using the control trajectory to represent the BG, it is natural to impose an additional constraint that the first derivative of the control trajectory must be continuous over the interval $\mathcal{T}$. If the state and control trajectories both obey the system dynamics, then imposing the continuity constraint on the first derivative of the control is equivalent to imposing a similar constraint on the state trajectory that its second derivative should be continuous over the interval $\mathcal{T}$. We therefore impose the additional constraint that the second derivative of the state trajectory should be continuous. This would change the collocation space for the control trajectories from $\mathcal{S}\left(\mathcal{P}_3, 2, \mathcal{T}_{N,3}\right)$ to $\mathcal{S}_3\left(\mathcal{T}_{N,3}\right)$ and the collocation space for the state trajectory from $\mathcal{S}\left(\mathcal{P}_4, 2, \mathcal{T}_{N,3}\right)$ to $\mathcal{S}_4\left(\mathcal{T}_{N,3}\right)$. The additional constraint imposed on the state trajectory would lead to the following additional constraint equation

$$3\left(x_{k+2} - x_k\right) = h_k\left(4f_{k+1} + f_k + 4f_{k+2}\right) \tag{4.52}$$

The objective function J discretized over the collocation points $\mathcal{T}_{N,3}$ could be represented by,

$$\hat{J} = \sum_{k=0}^{N-1} \frac{h_k}{6}\left(J_{w,\ k+1} + 4J_{w,\ k+\frac{1}{2}} + J_{w,\ k}\right), \tag{4.53}$$

where $J_{w,k} = J_w\left(t_k\right)\ \forall k \in \left\{0, \frac{1}{2}, \ldots, N+1\right\}$ The path and boundary constraints applied on the collocation points can be represented by

$$h_{k,i} \equiv h_i\left(x_k, u_k\right) \leq 0, \quad \forall k \in \left\{0, \frac{1}{2}, \ldots, N+1\right\},\ i \in \mathbb{N}_n, \tag{4.54}$$

$$g_{k,j} \equiv g_j\left(x_k, u_k\right) = 0, \quad \forall k \in \left\{0, \frac{1}{2}, \ldots, N+1\right\},\ j \in \mathbb{N}_m, \tag{4.55}$$

$$x_k^{\text{low}} \leq x_k \leq x_k^{\text{up}}, \quad \forall k \in \left\{0, \frac{1}{2}, \ldots, N+\frac{1}{2}, N+1\right\}, \tag{4.56}$$

$$u_k^{\text{low}} \leq u_k \leq u_k^{\text{up}}, \quad \forall k \in \left\{0, \frac{1}{2}, \ldots, N+\frac{1}{2}, N+1\right\}, \tag{4.57}$$

The problem given by Eq. (4.35) now boils down to an NLP problem,

$$\min_{x_k, u_k} \quad \hat{J} = \sum_{k=0}^{N-1} \frac{h_k}{6}\left(J_{w,\ k+1} + 4J_{w,\ k+\frac{1}{2}} + J_{w,\ k}\right) \tag{4.58a}$$

$$\text{subject to} \quad x_{k+1} = x_k + \frac{h_k}{6}\left(f_k + 4f_{k+\frac{1}{2}} + f_{k+1}\right), \quad \forall k \in \mathcal{Z}, \tag{4.58b}$$

$$3\left(x_{k+2} - x_k\right) = h_k\left(4f_{k+1} + f_k + 4f_{k+2}\right), \forall k \in \mathcal{Z}, \tag{4.58c}$$

$$x_{k+\frac{1}{2}} = \frac{1}{2}\left(x_k + x_{k+1}\right) + \frac{h_k}{8}\left(\dot{x}_k - \dot{x}_{k+1}\right), \forall k \in \mathcal{Z}, \tag{4.58d}$$

$$h_{k,\ i}\left(x_k, u_k\right) \leq 0, \qquad \forall\ k \in \mathcal{Z},\ i \in \mathbb{N}_n, \tag{4.58e}$$

$$g_j\left(x_k, u_k\right) = 0, \qquad \forall\ k \in \mathcal{Z}, j \in \mathbb{N}_m, \tag{4.58f}$$

$$x_{k,\text{low}} \leq x_k \leq x_{k,\text{up}}, \qquad \forall\ k \in \mathcal{Z}, \tag{4.58g}$$

$$u_{k,\text{low}} \leq u_k \leq u_{k,\text{up}}, \qquad \forall\ k \in \mathcal{Z}, \tag{4.58h}$$

where, the points x_k and $u_k\ \forall\ k \in \mathcal{Z} \equiv \left\{0, \frac{1}{2}, 1, \ldots, N+\frac{1}{2}, N+1\right\}$ need to be estimated which satisfies Eqs. (4.58b)–(4.58h) and also minimizes the objective function $\hat{J}$ (from Eq. (4.58a)).

4.3 Results

4.3.1 Estimating BG and ISF trajectories using Trapezoidal collocation

The Pune-2021 dataset contained 50 CGM traces from a total of 50 participants. These CGM traces were separated into smaller CGM traces, each containing the ISF measurements corresponding to a particular date. The typical CGM trace which is 14 days long now had been separated into 15 small CGM traces, out of which only 13 traces could be considered a full-day long CGM traces. The other two small traces, usually the first day's trace and the last day's trace were only half a day long. Therefore for the full-day CGM traces we had 96 collocation points each with a $h = 15\,\text{min}$. These small traces were then used to estimate the continuous ISF and BG traces using Trapezoidal collocation method. We solved the NLP problem using the `OSQP` solver provided by the `cvxpy` package in `Python`. The CGM traces which were not 14 days long were also included in this exercise, unlike the complete CGM traces these traces provided only a few full-day CGM traces.

Now, each full-day CGM trace was taken and the set of collocation points corresponding to the time stamps of each ISF measurement in the full-day CGM trace was constructed. Therefore, using the trapezoidal collocation method, for each collocation point, t_k we obtained an estimated ISF value, $\text{ISF}_k^{\text{est.}}$, an estimated BG value, $\text{BG}_k^{\text{est.}}$ and, an observed ISF value, $\text{ISF}_k^{\text{obs}}$ from the full-day CGM. The objective function was constructed as the mean squared error (MSE) or, square of the root mean squared error (RMSE) between the vectors $\text{ISF}_k^{\text{est.}}$ and $\text{ISF}_k^{\text{obs}}$. The NLP problem that we obtained after trapezoidal collocation can be stated as

$$\min_{x_k, u_k} \quad \hat{J} = \sum_{k=0}^{N-1} \frac{h_k}{2} \left(J_{w,\ k+1} + J_{w,\ k} \right) \tag{4.59a}$$

$$\text{where,} \quad J_{w,\ k} = \left(\mathrm{ISF}_k^{\mathrm{est.}} - \mathrm{ISF}_k^{\mathrm{obs}} \right)^2$$

$$\text{subject to} \quad \mathrm{ISF}_{k+1}^{\mathrm{est.}} = \mathrm{ISF}_k^{\mathrm{est.}} + \frac{h_k}{2} \left(f_{k+1}, f_k \right), \qquad \forall k \in \{0\} \cup \mathbb{N}_N, \tag{4.59b}$$

$$\text{where,} \quad f_k = f\left(\mathrm{ISF}_k^{\mathrm{est.}}, \mathrm{BG}_k^{\mathrm{est.}} \right),$$

$$\text{and,} \quad f\left(\mathrm{ISF}, \mathrm{BG} \right) = \tau_0 \left(\frac{1}{1.17} \mathrm{BG} - \mathrm{ISF} \right),$$

$$\mathrm{ISF}_k^{\mathrm{est.}} - \mathrm{BG}_k^{\mathrm{est.}} \leq 0, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.59c}$$

$$30\,\mathrm{mg/dL} \leq \mathrm{ISF}_k^{\mathrm{est.}} \leq 600\,\mathrm{mg/dL}, \quad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.59d}$$

$$30\,\mathrm{mg/dL} \leq \mathrm{BG}_k^{\mathrm{est.}} \leq 600\,\mathrm{mg/dL}, \quad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.59e}$$

The NLP described by Eqs. (4.59) was solved numerically using the `OSQP` solver from the `CVXPY` package in `Python`. We first took $\tau_0 = 1\,\mathrm{min}$ but later repeated the process for $\tau_0 = \frac{1}{5}\,\mathrm{min}$ and, $\tau_0 = \frac{1}{20}\,\mathrm{min}$.

The full-day CGM trace along with the estimated ISF and BG trace with $\tau_0 = 1\,\mathrm{min}$, $\frac{1}{5}\,\mathrm{min}$ and, $\frac{1}{20}\,\mathrm{min}$ for participants labeled `CGM D 05`, `CGM PD 06` and `CGM ND 03` have been shown below in Figure. 4.1 to Figure. 4.9.

Now, to study the reliability of the BG estimates, we compared our estimated BG with the measured ISF values from the CGM traces. This method was used in the FDA Summary of Safety and Effectiveness Data ($PMA\ P150021$) manuscript to verify the effectiveness of the Abbott Libre FreeStyle Pro CGM device. The process consists of two approaches. First we verify the degree of agreement of the estimated BG within the measured ISF ranges. This is shown in Tables. 4.1,4.2 and, 4.3. Here, we first pair the estimated BG values to their corresponding measured ISF values. Then for a given range of measured ISF value we calculate the number of pairs where the estimated BG agrees with the measured ISF within a given level of error. For example in column. 3 of Table. 4.1 we count the number of estimated BG values that are within 15 % of the measured ISF, while the ISF is with the ranges given in column. 1 of the same table. The second approach is similar to the first, in that the estimated BG not takes the place of ISF and vice versa for the ISF. These tables are shown in Tables. 4.4, 4.5 and, 4.6.

CGM measured ISF values	No. of paired measured ISF and estimated BG	No. of BG within 15 % of measured ISF	No. of BG within 20 % of measured ISF	No. of BG within 30 % of measured ISF	No. of BG within 40 % of measured ISF	No. of BG greater than 40 % of measured ISF
overall	61732	3482	60137	61693	61731	1
40–50	1561	440	1302	1555	1561	0
51–80	15104	1314	14519	15092	15104	0
81–180	41326	1673	40601	41305	41325	1
181–300	3460	35	3438	3460	3460	0
301–400	269	18	267	269	269	0
401–500	12	2	10	12	12	0

Table 4.1: Table showing the degree of agreement of estimated BG within the measured ISF ranges. The estimated BG was calculated using trajectory optimization with trapezoidal collocation. The time scaling factor τ was taken to be 1 min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of ISF, the estimated BG is within a desired percentage of the measured ISF are counted.

CGM measured ISF values	No. of paired measured ISF and estimated BG	No. of BG within 15 % of measured ISF	No. of BG within 20 % of measured ISF	No. of BG within 30 % of measured ISF	No. of BG within 40 % of measured ISF	No. of BG greater than 40 % of measured ISF
overall	61732	20976	46058	60369	61641	91
40–50	1561	672	957	1350	1549	12
51–80	15104	5529	10685	14696	15056	48
81–180	41326	13866	31180	40591	41295	31
181–300	3460	856	2982	3451	3460	0
301–400	269	51	244	269	269	0
401–500	12	2	10	12	12	0

Table 4.2: Table showing the degree of agreement of estimated BG within the measured ISF ranges. The estimated BG was calculated using trajectory optimization with trapezoidal collocation. The time scaling factor τ was taken to be 1/5min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of ISF, the estimated BG is within a desired percentage of the measured ISF are counted.

CGM measured ISF values	No. of paired measured ISF and estimated BG	No. of BG within 15 % of measured ISF	No. of BG within 20 % of measured ISF	No. of BG within 30 % of measured ISF	No. of BG within 40 % of measured ISF	No. of BG greater than 40 % of measured ISF
overall	61732	29399	39424	53215	58292	3440
40–50	1561	785	937	1293	1464	97
51–80	15104	7306	9699	13097	14317	784
81–180	41326	19672	26052	35358	38848	2478
181–300	3460	1532	2342	3191	3382	78
301–400	269	101	186	265	269	0
401–500	12	3	8	11	12	0

Table 4.3: Table showing the degree of agreement of estimated BG within the measured ISF ranges. The estimated BG was calculated using trajectory optimization with trapezoidal collocation. The time scaling factor τ was taken to be 1/15min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of ISF, the estimated BG is within a desired percentage of the measured ISF are counted.

Estimated BG values	No. of paired measured ISF and estimated BG	No. of ISF within 15 % of estimated BG	No. of ISF within 20 % of estimated BG	No. of ISF within 30 % of estimated BG	No. of ISF within 40 % of estimated BG	No. of ISF greater than 40 % of estimated BG
overall	61732	46313	61631	61731	61732	0
40–50	876	546	876	876	876	0
51–80	5863	4052	5816	5863	5863	0
81–180	47474	35505	47424	47473	47474	0
181–300	6013	5023	6012	6013	6013	0
301–400	702	622	702	702	702	0
401–500	76	60	76	76	76	0

Table 4.4: Table showing the degree of agreement of measured ISF within the estimated BG ranges. The estimated BG was calculated using trajectory optimization with trapezoidal collocation. The time scaling factor τ was taken to be 1 min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of estimated BG, the measured ISF is within a desired percentage of the estimated BG are counted.

Estimated BG values	No. of paired measured ISF and estimated BG	No. of ISF within 15 % of estimated BG	No. of ISF within 20 % of estimated BG	No. of ISF within 30 % of estimated BG	No. of ISF within 40 % of estimated BG	No. of ISF greater than 40 % of estimated BG
overall	61732	35417	56890	61681	61730	2
40–50	745	590	745	745	745	0
51–80	6211	3996	5478	6201	6210	1
81–180	47149	26487	43416	47112	47149	0
181–300	6100	3426	5806	60697	6099	1
301–400	698	424	692	698	698	0
401–500	78	35	78	78	78	0

Table 4.5: Table showing the degree of agreement of measured ISF within the estimated BG ranges. The estimated BG was calculated using trajectory optimization with trapezoidal collocation. The time scaling factor τ was taken to be 1/5min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of estimated BG, the measured ISF is within a desired percentage of the estimated BG are counted.

Estimated BG values	No. of paired measured ISF and estimated BG	No. of ISF within 15 % of estimated BG	No. of ISF within 20 % of estimated BG	No. of ISF within 30 % of estimated BG	No. of ISF within 40 % of estimated BG	No. of ISF greater than 40 % of estimated BG
overall	61732	34776	47722	59001	61350	382
40–50	849	679	849	849	849	0
51–80	7146	5556	6402	7056	7141	5
81–180	45661	24940	34792	43660	45389	272
181–300	6377	2655	4333	5790	6280	97
301–400	730	311	544	706	729	1
401–500	88	26	73	86	87	1

Table 4.6: Table showing the degree of agreement of measured ISF within the estimated BG ranges. The estimated BG was calculated using trajectory optimization with trapezoidal collocation. The time scaling factor τ was taken to be 1/15min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of estimated BG, the measured ISF is within a desired percentage of the estimated BG are counted.

4.3.2 Estimating BG and ISF trajectories using a modified Hermite-Simpson collocation method

As done in Section. 4.3.1, the CGM traces of the 50 participants of the Pune-2021 dataset were separated into smaller CGM traces, where each containing the ISF measurements corresponding to a particular date. A the full-day CGM trace usually had 96 collocation points (taking $N = 94$), unless the CGM device dropped a few measurements due to some defect, each with a $h = 15\,\text{min}$. These small traces were then used to estimate the continuous ISF and BG trajectories using a modified Hermite-Simpson collocation method. We solved the NLP optimization problem using the `OSQP` solver provided by the `cvxpy` package in `Python`. The CGM traces which were not 14 days long were also included in this exercise, unlike the complete CGM traces these traces provided only a few full-day CGM traces.

Now, each full-day CGM trace was taken and the set of collocation points corresponding to the time stamps of each ISF measurement in the full-day CGM trace was constructed. Then additional collocation points corresponding to center of the interval between each ISF measurement were added to the set of existing collocation points. Now, for the modified Hermite-Simpson collocation method we have a set of collocation point which have corresponding ISF measurements obtained from the full-day CGM trace. There are additional collocation points where we use linear interpolation to obtain an estimate of the measured ISF value. Therefore for each collocation point of the form t_k where $k \in \mathbb{N}$ we have a corresponding observed ISF value from the CGM trace, represented by $\text{ISF}_k^{\text{obs}}$ and for each collocation point of the form $t_{k+\frac{1}{2}}$ we estimate, using linear interpolation a value of the observed ISF. However for all collocation points, t_k or $t_{k+\frac{1}{2}}$, we obtain an estimate of the ISF, represented by $\text{ISF}_k^{\text{est.}}$ and we also obtain an estimate of BG, represented by $\text{BG}_k^{\text{est.}}$. Therefore we as was the case with trapezoidal collocation we eventually have for every collocation point t_k an estimated ISF value, $\text{ISF}_k^{\text{est.}}$, an estimated BG value, $\text{BG}_k^{\text{est.}}$ and, an observed ISF value, $\text{ISF}_k^{\text{obs}}$. Similar to Section. 4.3.1 we construct the objective function as the RMSE between the vectors $\text{ISF}_k^{\text{est.}}$ and $\text{ISF}_k^{\text{obs}}$. The NLP problem that we obtained after using the modified Hermite-Simpson collocation can be stated as

$$\min_{x_k, u_k} \quad \hat{J} = \sum_{k=0}^{N-1} \frac{h_k}{6} \left(J_{w,\ k+1} + 4J_{w,\ k+\frac{1}{2}} + J_{w,\ k} \right) \tag{4.60a}$$

where, $J_{wk} = \left(\text{ISF}_k^{\text{est.}} - \text{ISF}_k^{\text{obs}} \right)^2$

subject to
$$\text{ISF}_{k+1}^{\text{est.}} = \text{ISF}_k^{\text{est.}} + \frac{h_k}{6} \left(f_k + 4f_{k+\frac{1}{2}} + f_{k+1} \right), \qquad \forall k \in \mathcal{Z}, \tag{4.60b}$$

$$3 \left(\text{ISF}_{k+2}^{\text{est.}} - \text{ISF}_k^{\text{est.}} \right) = h_k \left(4f_{k+1} + f_k + 4f_{k+2} \right), \quad \forall k \in \mathcal{Z}, \tag{4.60c}$$

$$\text{ISF}_{k+\frac{1}{2}}^{\text{est.}} = \frac{1}{2} \left(\text{ISF}_k^{\text{est.}} + \text{ISF}_{k+1}^{\text{est.}} \right) + \frac{h_k}{8} \left(\dot{x}_k - \dot{x}_{k+1} \right), \forall k \in \mathcal{Z}, \tag{4.60d}$$

where, $f_k = f \left(\text{ISF}_k^{\text{est.}}, \text{BG}_k^{\text{est.}} \right),$

and, $f \left(\text{ISF}, \text{BG} \right) = \tau_0 \left(\frac{1}{1.17} \text{BG} - \text{ISF} \right),$

$$\text{ISF}_k^{\text{est.}} - \text{BG}_k^{\text{est.}} \leq 0, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.60e}$$

$$30\,\text{mg/dL} \leq \text{ISF}_k^{\text{est.}} \leq 600\,\text{mg/dL}, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.60f}$$

$$30\,\text{mg/dL} \leq \text{BG}_k^{\text{est.}} \leq 600\,\text{mg/dL}, \qquad \forall k \in \{0\} \cup \mathbb{N}_{N+1}, \tag{4.60g}$$

The NLP described by Eqs. (4.59) was solved numerically using the `OSQP` solver from the `CVXPY` package in `Python`. We first took $\tau_0 = 1\,\text{min}$ but later repeated the process for $\tau_0 = \frac{1}{5}\,\text{min}$ and, $\tau_0 = \frac{1}{20}\,\text{min}$.

The full-day CGM trace along with the estimated ISF and BG trace with $\tau_0 = 1\,\text{min}$, $\frac{1}{5}\,\text{min}$ and, $\frac{1}{20}\,\text{min}$ for participants labeled `CGM D 05`, `CGM PD 06` and `CGM ND 03` have been shown below in Figure. 4.10 to Figure. 4.18.

Just as it was done in Section. 4.3.1 we verified the degree of agreement between the estimated BG and the measured ISF. This is shown in Tables. 4.10to 4.12 for the agreement about BG and Tables. 4.7 to 4.9 for the agreement about ISF ranges.

CGM measured ISF values	No. of paired measured ISF and estimated BG	No. of BG within 15 % of measured ISF	No. of BG within 20 % of measured ISF	No. of BG within 30 % of measured ISF	No. of BG within 40 % of measured ISF	No. of BG greater than 40 % of measured ISF
overall	61732	260	61538	61731	61731	1
40–50	1561	6	1529	1561	1561	0
51–80	15104	52	15001	15104	15104	0
81–180	41326	200	41267	41325	41325	1
181–300	3460	0	3460	3460	3460	0
301–400	269	0	269	269	269	0
401–500	12	2	12	12	12	0

Table 4.7: Table showing the degree of agreement of estimated BG within the measured ISF ranges. The estimated BG was calculated using trajectory optimization with a modified Hermite-Simpson collocation. The time scaling factor τ was taken to be 1 min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of ISF, the estimated BG is within a desired percentage of the measured ISF are counted.

CGM measured ISF values	No. of paired measured ISF and estimated BG	No. of BG within 15 % of measured ISF	No. of BG within 20 % of measured ISF	No. of BG within 30 % of measured ISF	No. of BG within 40 % of measured ISF	No. of BG greater than 40 % of measured ISF
overall	61732	9964	55970	61722	61728	4
40–50	1561	260	1444	1560	1560	1
51–80	15104	2487	13996	15102	15104	0
81–180	41326	6918	37007	41319	41323	3
181–300	3460	294	3245	3460	3460	0
301–400	269	4	268	269	269	0
401–500	12	1	10	12	12	0

Table 4.8: Table showing the degree of agreement of estimated BG within the measured ISF ranges. The estimated BG was calculated using trajectory optimization with a modified Hermite-Simpson collocation. The time scaling factor τ was taken to be 1/5min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of ISF, the estimated BG is within a desired percentage of the measured ISF are counted.

CGM measured ISF values	No. of paired measured ISF and estimated BG	No. of BG within 15 % of measured ISF	No. of BG within 20 % of measured ISF	No. of BG within 30 % of measured ISF	No. of BG within 40 % of measured ISF	No. of BG greater than 40 % of measured ISF
overall	61732	22595	46030	58854	61305	427
40–50	1561	461	1260	1500	1542	19
51–80	15104	5743	11916	14664	15023	81
81–180	41326	15191	29970	39059	41005	321
181–300	3460	1139	2653	3350	3454	6
301–400	269	60	221	269	269	0
401–500	12	1	10	12	12	0

Table 4.9: Table showing the degree of agreement of estimated BG within the measured ISF ranges. The estimated BG was calculated using trajectory optimization with a modified Hermite-Simpson collocation. The time scaling factor τ was taken to be 1/15min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of ISF, the estimated BG is within a desired percentage of the measured ISF are counted.

Estimated BG values	No. of paired measured ISF and estimated BG	No. of ISF within 15 % of estimated BG	No. of ISF within 20 % of estimated BG	No. of ISF within 30 % of estimated BG	No. of ISF within 40 % of estimated BG	No. of ISF greater than 40 % of estimated BG
overall	61732	53620	61731	61731	61732	0
40–50	864	756	864	864	864	0
51–80	5846	4907	5846	5846	5846	0
81–180	47469	40930	47468	47468	47469	0
181–300	6003	5611	6003	6003	6003	0
301–400	705	689	705	705	705	0
401–500	73	73	73	73	73	0

Table 4.10: Table showing the degree of agreement of measured ISF within the estimated BG ranges. The estimated BG was calculated using trajectory optimization with a modified Hermite-Simpson collocation. The time scaling factor τ was taken to be 1 min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of estimated BG, the measured ISF is within a desired percentage of the estimated BG are counted.

Estimated BG values	No. of paired measured ISF and estimated BG	No. of ISF within 15 % of estimated BG	No. of ISF within 20 % of estimated BG	No. of ISF within 30 % of estimated BG	No. of ISF within 40 % of estimated BG	No. of ISF greater than 40 % of estimated BG
overall	61732	42040	61340	61728	61731	1
40–50	898	747	898	898	898	0
51–80	5937	4421	5894	5936	5937	0
81–180	47317	317769	47005	47316	47316	1
181–300	6102	4040	6068	6100	6102	0
301–400	696	491	696	696	696	0
401–500	75	55	75	75	75	0

Table 4.11: Table showing the degree of agreement of measured ISF within the estimated BG ranges. The estimated BG was calculated using trajectory optimization with a modified Hermite-Simpson collocation. The time scaling factor τ was taken to be 1/5min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of estimated BG, the measured ISF is within a desired percentage of the estimated BG are counted.

Estimated BG values	No. of paired measured ISF and estimated BG	No. of ISF within 15 % of estimated BG	No. of ISF within 20 % of estimated BG	No. of ISF within 30 % of estimated BG	No. of ISF within 40 % of estimated BG	No. of ISF greater than 40 % of estimated BG
overall	61732	36254	55219	61499	61719	13
40–50	903	743	903	903	903	0
51–80	6316	4631	5973	6298	6314	2
81–180	46789	26942	41733	46623	46783	6
181–300	6196	3060	5220	6152	6193	3
301–400	728	373	648	725	726	2
401–500	80	37	78	80	80	0

Table 4.12: Table showing the degree of agreement of measured ISF within the estimated BG ranges. The estimated BG was calculated using trajectory optimization with a modified Hermite-Simpson collocation. The time scaling factor τ was taken to be 1/15min. Given an estimated BG we first pair the BG to its corresponding measured ISF value. Then the number of such pairs where for a certain range of estimated BG, the measured ISF is within a desired percentage of the estimated BG are counted.

4.4 Discussion

Diabetes is a metabolic disorder characterized by a consistently high BG. Understanding the BG dynamics like, knowing what the current BG value is, what it's value would be after a meal of certain calories and, so on, plays an important role in the diagnosis, construction of medication plan and, the management of diabetes. Traditionally, single point BG measurements were made and labeled as metrics like, FBG and, PPBG. Later, metrics like HbA1c, time in range (TiR) and, so on acted as metrics which indicated the degree of glycation and indirectly indicated the degree of consistent high or low BG. Recently, with the introduction of FGM and CGM devices, we are introduced to ISF readings and ISF traces.

We know that traditionally methods for diagnosis, management and, treatment of diabetes were based on single point measurements of BG and, proxies for long-term aggregate BG, we need new methods which will proved us with reliable conversions to and from the traditional BG based metrics to the new CGM based metrics. This can be achieved in various ways like, finding algorithms to reliably estimate BG based metrics like HbA1c from a given CGM trace (see Chapter. 2, Chapter. 3) or, can be done by first estimating the BG trace corresponding to the measured ISF trace provided by the CGM.

In this manuscript we try to construct an estimate of the continuous BG and the corresponding continuous ISF, which when sampled (or say measured at specific time stamps), generates a CGM trace close to the CGM trace generated by the CGM device. The notion of closeness can be described by an appropriate cost functional. This was done using the trajectory optimization method. Trajectory optimization is a method developed to solve optimal control problems where an optimal control trajectory is to be found which would satisfy a given set of differential, equality and, inequality constraints and would minimize/maximize a given cost functional. Our problem can be boiled down to an optimal control problem which in turn is an optimization problem, if the constraints are suitable for a unique feasible solution, trajectory optimization can be used as a suitable method for finding solutions to our problem.

In trajectory optimization we first describe the system that the optimal trajectories must follow. Here, we consider a simple compartmental model describing

the dynamics between ISF and BG. We also impose additional constraints on the system, that the ISF $\leq$ BG, 30 mg/dL $\leq$ ISF $\leq$ 600 mg/dL and, 30 mg/dL $\leq$ BG $\leq$ 600 mg/dL. After considering the objective function/cost functional that need to be minimized to be the L_2 norm of the difference between the estimated ISF and measured ISF, the problem essentially becomes an optimization problem.

Now solving the above mentioned optimization problem is difficult and moreover we only have a discrete-time ISF trace as the measured ISF. Therefore we discretize the optimization problem over a set of collocation points, which coincide with the time stamps at which the ISF was measured by the CGM device. After discretization the optimization problem becomes a NLP problem, which is solved using traditional NLP solvers. The solution of the NLP problem would provide us with the estimated ISF and BG values at the collocation points, effectively a discrete-time ISF and BG estimate. We then interpolate the estimates between the neighboring collocation points with the proper interpolation method, according to the collocation method, to obtain the continuous-time estimates of both ISF and BG. We consider only two collocation methods for solving our problem. We first use trapezoidal collocation where the continuous ISF trajectory is obtained as quadratic spline and the BG trajectory as a linear spline. We then use a modified Hermite-Simpson collocation where the ISF trajectory is obtained as a cubic spline and the BG trajectory as a quadratic spline.

We now have obtained an approximate continuous ISF and BG trace, from both trapezoidal collocation and a modified Hermite-Simpson collocation, which satisfy at the collocation points, the path constraints imposed and the differential equations we imposed as the system equations and moreover the estimated ISF trajectory pass as close as possible to the measured ISF points. Therefore, given a CGM trace we can provide a piecewise linear or a piecewise quadratic BG trace that would generate a corresponding ISF trace, following our system equations, that would pass as close as possible to all the CGM trace points.

This method would indeed help clinicians as they would have the continuous BG trace for reference when designing a medication schedule or an insulin bolus schedule. It would help clinicians also evaluate their previously designed medication regime. It would also help patient customize their diets and potentially

become better at diabetes management. It could potentially also help us obtain better estimates of BG based metrics like HbA1c, glucose monitoring index (GMI), etc.

Although, we have obtained estimates of the continuous ISF and BG, these estimates are essentially approximates of the continuous BG and ISF from finite-dimensional functional spaces. We assume that the continuous BG and ISF should have a certain degree of smoothness and we then find their approximates in the corresponding finite-dimensional function spaces. Moreover, in our formulation of the optimization problem the measurement error incorporated in the ISF trace is not taken into account and therefore for $\tau_0 = 1\,\text{min}$, the estimated ISF trajectory either passes through the measured ISF trace or it passes very close to it. Decreasing the value of τ_0 we can see from Tables. 4.2, 4.3, 4.5, 4.6, 4.8, 4.9, 4.11 and, 4.12 that we get a more realistic ISF and BG trace and more over for ranges around the average glucose concentration there seem to be more agreement between the estimated ISF and BG. Although it is difficult to say which τ_0 should be chosen, it it clear that introducing a scaling factor τ_0 significantly improves the degree of agreement between the estimated ISF and BG for glucose concentration about the default range. The next step would be to obtain an estimate of τ_0 and to include the measurement error associated with each ISF measurement by the CGM device. It would also be appropriate to consider a more complicated system dynamical equations to describe the dynamics between BG and ISF.

Ethics statement

The study protocol for the Pune-2021 dataset was approved by the Institutional Ethical Committee (IEC) of Savitribai Phule Pune University, Pune, India (SPPU/IEC/2020/102). All the participants in the study were provided with an informed consent form prior to their participation in the studies. The data analysis was approved by the Institutional Human Ethics Committee (IHEC) of IISER Pune (IHEC/Admin/2021/015).

Funding

This was financially supported by Rastriya Ucchattar Shiksha Abhiyan (RUSA), Savitribai Phule Pune University, Pune. Sayantan Majumdar is a Senior Research Fellow supported by a DST INSPIRE Fellowship, Department of Science and Technology, Government of India. Saurabh D. Kalamkar who helped in data collection was a recipient of the Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), India.

References

[1] Hargraves, C. R.; Paris, S. W. Direct trajectory optimization using nonlinear programming and collocation. *Journal of guidance, control, and dynamics* **1987**, *10*, 338–342, DOI: 10.2514/3.20223, https://arc.aiaa.org/doi/10.2514/3.20223 (cit. on pp. 71, 80 sq., 84).

[2] Enright, P. J.; Conway, B. A. Discrete approximations to optimal trajectories using direct transcription and nonlinear programming. *Journal of Guidance, Control, and Dynamics* **1992**, *15*, 994–1002, DOI: 10.2514/3.20934, https://arc.aiaa.org/doi/10.2514/3.20934 (cit. on pp. 71, 80 sq., 84).

[3] Rebrin, K.; Steil, G. M.; Van Antwerp, W. P.; Mastrototaro, J. J. Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring. *American Journal of Physiology-Endocrinology and Metabolism* **1999**, *277*, E561–E571, DOI: 10.1152/ajpendo.1999.277.3.E561, https://journals.physiology.org/doi/full/10.1152/ajpendo.1999.277.3.E561 (cit. on p. 68).

[4] Rebrin, K.; Steil, G. M. Can interstitial glucose assessment replace blood glucose measurements? *Diabetes technology & therapeutics* **2000**, *2*, 461–472, DOI: 10.1089/15209150050194332, https://www.liebertpub.com/doi/10.1089/15209150050194332 (cit. on p. 68).

[5] Stout, P. J.; Peled, N.; Erickson, B. J.; Hilgers, M. E.; Racchini, J. R.; Hoegh, T. B. Comparison of glucose levels in dermal interstitial fluid and finger capillary blood. *Diabetes technology & therapeutics* **2001**, *3*, 81–90, DOI: 10.1089/152091501750220046, https://www.liebertpub.com/doi/10.1089/152091501750220046 (cit. on p. 68).

[6] Thennadil, S. N.; Rennert, J. L.; Wenzel, B. J.; Hazen, K. H.; Ruchti, T. L.; Block, M. B. Comparison of glucose concentration in interstitial fluid, and capillary and venous blood during rapid changes in blood glucose levels. *Diabetes technology & therapeutics* **2001**, *3*, 357–365, DOI: 10.1089/15209150152607132, https://www.liebertpub.com/doi/10.1089/15209150152607132 (cit. on p. 68).

[7] Kulcu, E.; Tamada, J. A.; Reach, G.; Potts, R. O.; Lesho, M. J. Physiological differences between interstitial glucose and blood glucose measured in human subjects. *Diabetes care* **2003**, *26*, 2405–2409, DOI: 10.2337/diacare.26.8.2405, https://diabetesjournals.org/care/article/26/8/2405/22733/Physiological-Differences-Between-Interstitial (cit. on p. 69).

[8] Brunner, H., *Collocation methods for Volterra integral and related functional differential equations*; Cambridge university press: 2004; Vol. 15 (cit. on p. 77).

[9] Schumaker, L., *Spline functions: basic theory*, Third; Cambridge university press: 2007 (cit. on p. 79).

[10] Conway, B. A.; Paris, S. W. Spacecraft trajectory optimization using direct transcription and nonlinear programming. *Spacecraft trajectory optimization* **2010**, *29*, 37 (cit. on pp. 80 sq., 84).

[11] Keenan, D. B.; Mastrototaro, J. J.; Weinzimer, S. A.; Steil, G. M. Interstitial fluid glucose time-lag correction for real-time continuous glucose monitoring. *Biomedical signal processing and control* **2013**, *8*, 81–89, DOI: 10.1016/j.bspc.2012.05.007, https://www.sciencedirect.com/science/article/abs/pii/S1746809412000687 (cit. on p. 69).

[12] Cobelli, C.; Schiavon, M.; Dalla Man, C.; Basu, A.; Basu, R. Interstitial fluid glucose is not just a shifted-in-time but a distorted mirror of blood glucose: insight from an in silico study. *Diabetes technology & therapeutics* **2016**, *18*, 505–511, DOI: 10.1089/dia.2016.0112, https://www.liebertpub.com/doi/10.1089/dia.2016.0112 (cit. on pp. 68, 70).

[13] Kelly, M. An introduction to trajectory optimization: How to do your own direct collocation. *SIAM Review* **2017**, *59*, 849–904, DOI: 10.1137/16M1062569, https://epubs.siam.org/doi/10.1137/16M1062569 (cit. on pp. 74, 76 sq., 79–84).

[14] Celaya Llover, E. Second order collocation. **2021**, http://hdl.handle.net/2117/366522 (cit. on pp. 80 sq., 84).

[15] Majumdar, S.; Kalamkar, S. D.; Dudhgaonkar, S.; Shelgikar, K. M.; Ghaskadbi, S.; Goel, P. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full (cit. on pp. 72 sq.).

[16] Miranda, F.; Amorim, D.; Ferreira, L.; Abreu, C. In *AIP Conference Proceedings*, 2024; Vol. 3094, DOI: 10.1063/5.0210600, https://pubs.aip.org/aip/acp/article-abstract/3094/1/350002/3297130/The-importance-of-artificial-intelligence-in?redirectedFrom=fulltext (cit. on p. 69).

[17] Sun, T.; Liu, J.; Chen, C. J. Calibration algorithms for continuous glucose monitoring systems based on interstitial fluid sensing. *Biosensors and Bioelectronics* **2024**, 116450, DOI: 10.1016/j.bios.2024.116450, https://linkinghub.elsevier.com/retrieve/pii/S095656632400455X (cit. on p. 69).

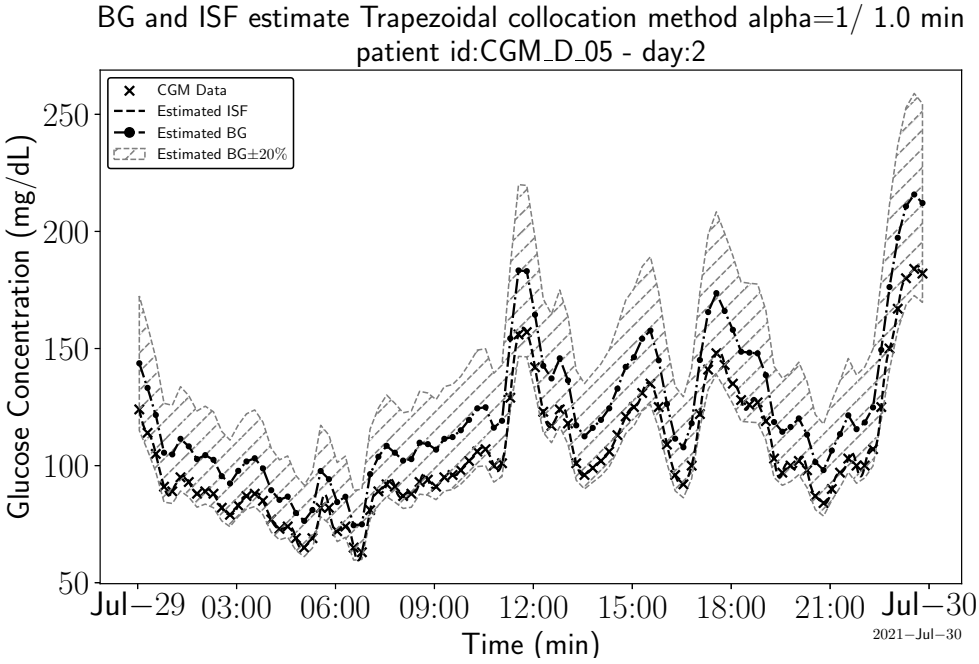


Figure 4.1: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM D 05` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1\,\text{min}$.

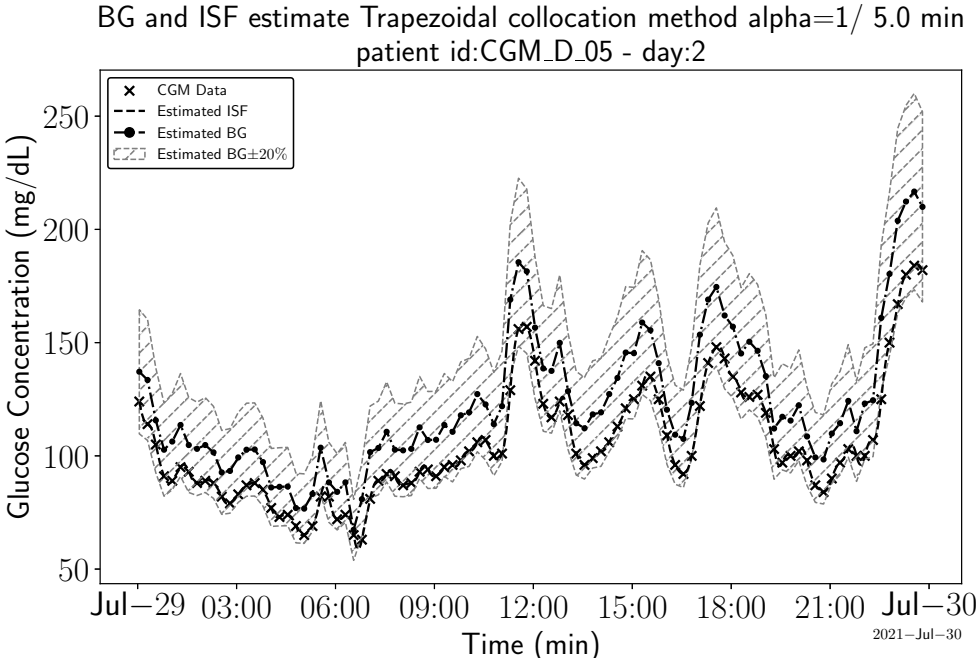


Figure 4.2: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM D 05` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1/5\text{min}$.

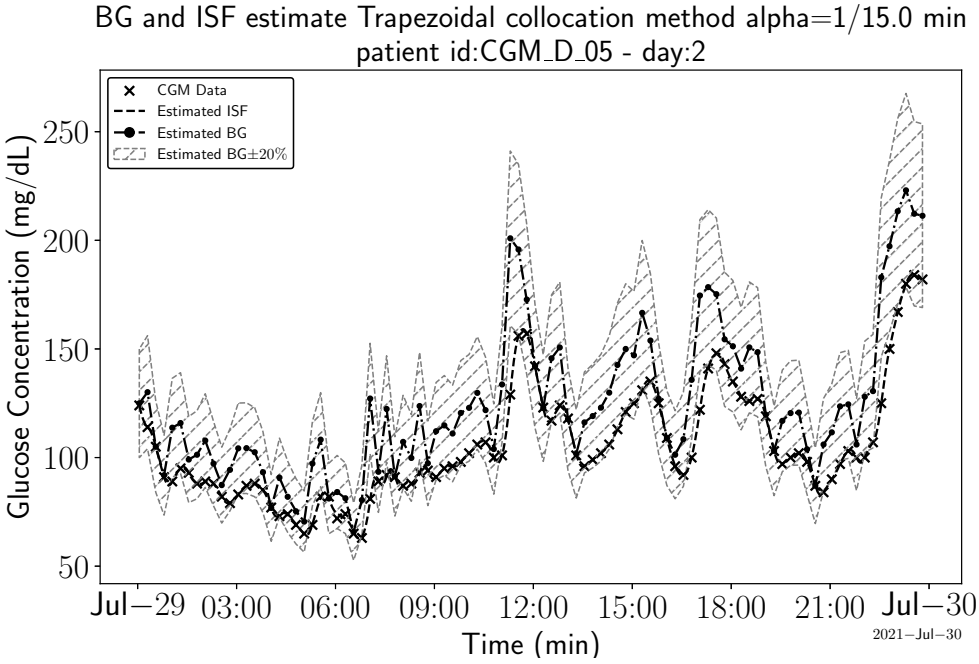


Figure 4.3: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM D 05` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1/15$min.

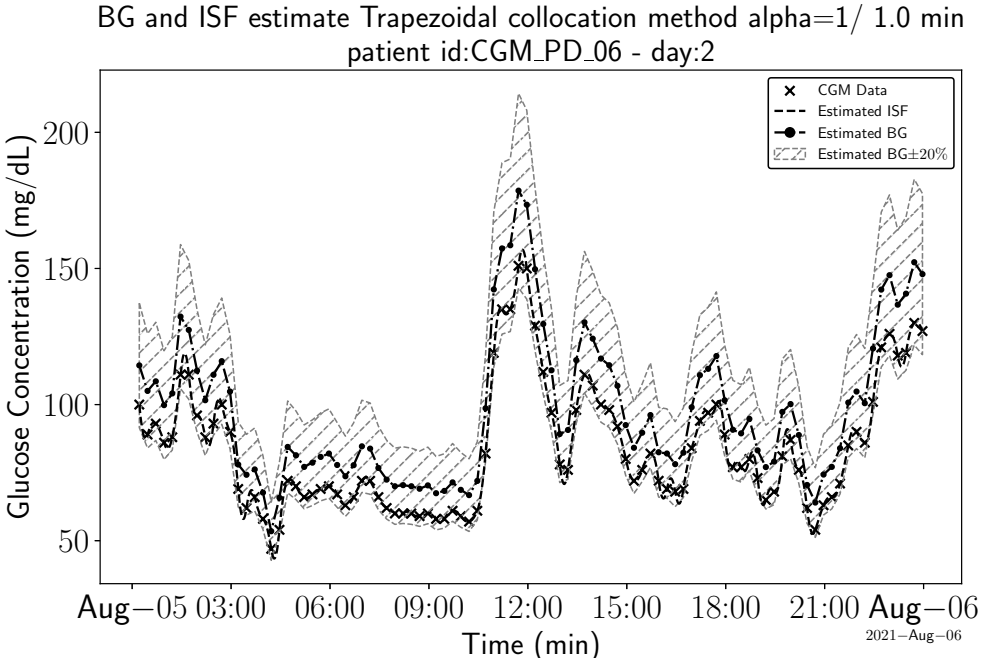


Figure 4.4: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM PD 06` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1\,\text{min}$.

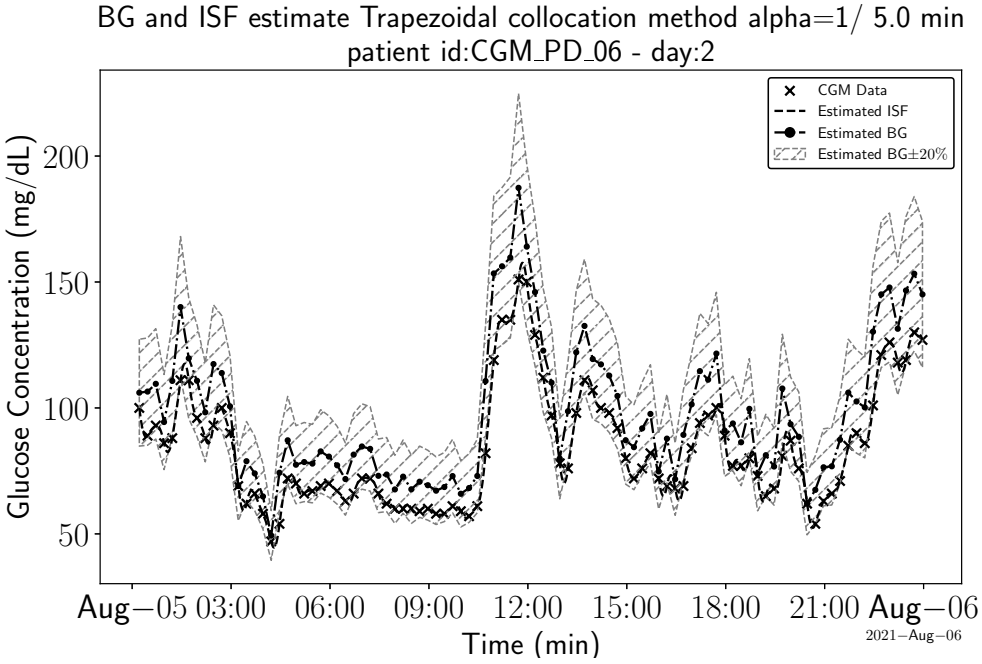


Figure 4.5: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM PD 06` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1/5$min.

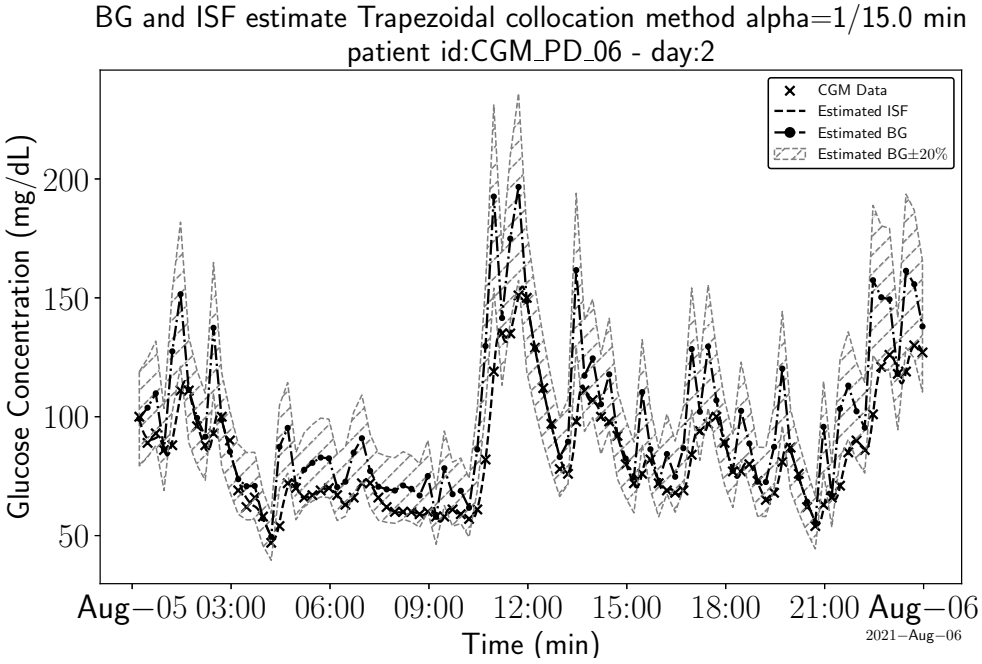


Figure 4.6: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM PD 06` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1/15$min.

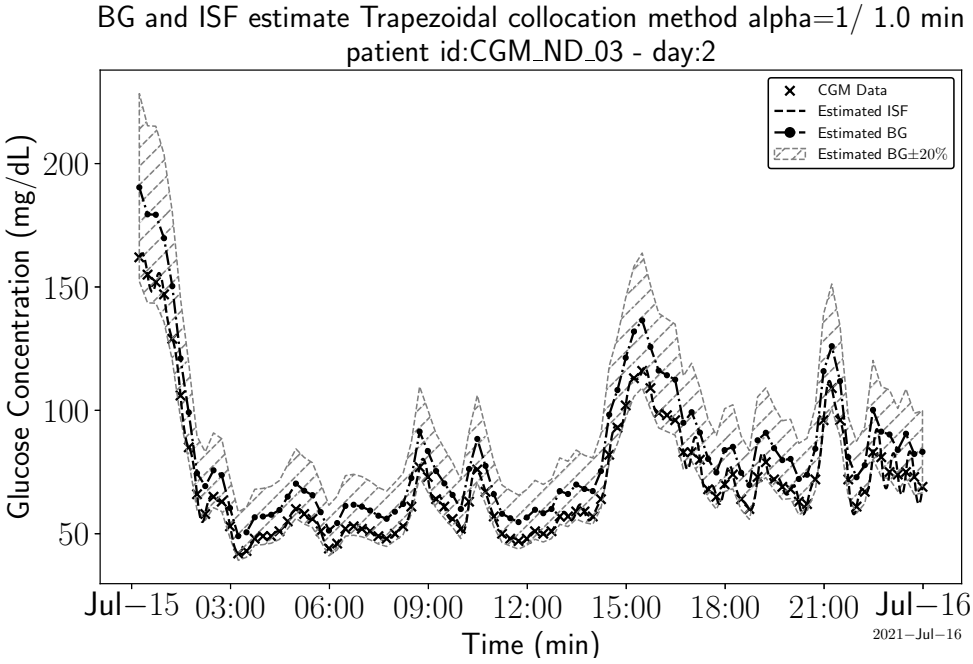


Figure 4.7: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM ND 03` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1\,\text{min}$.

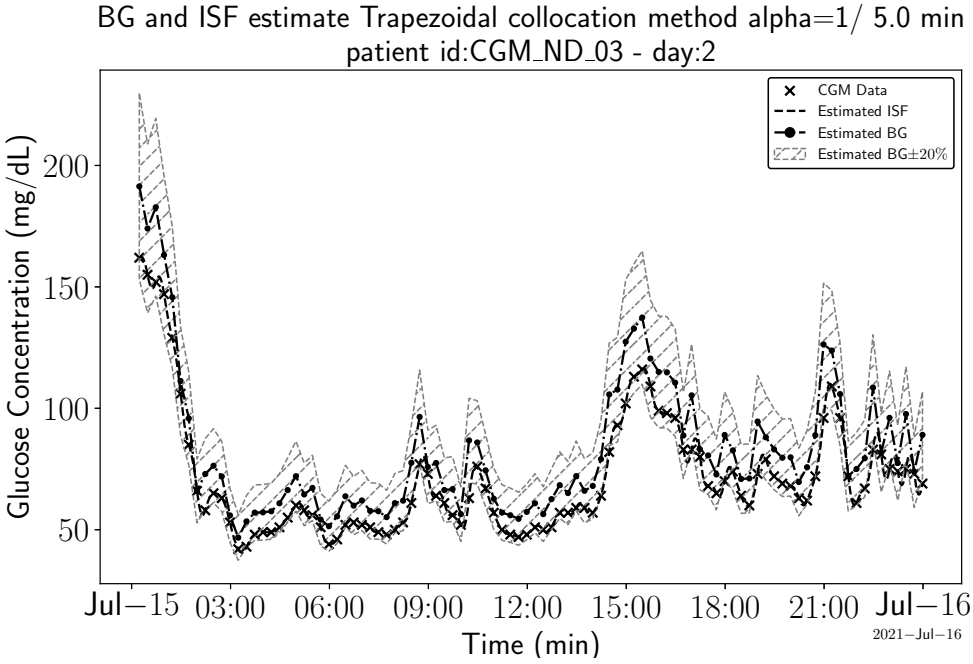


Figure 4.8: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM ND 03` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1/5\text{min}$.

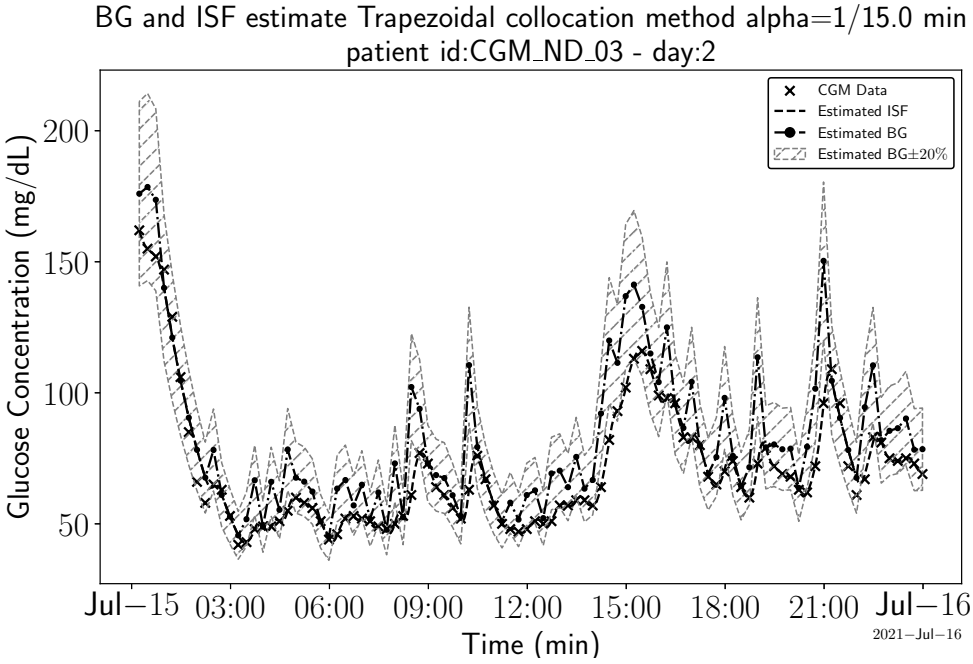


Figure 4.9: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM ND 03` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the trapezoidal collocation method with a $\tau_0 = 1/15$min.

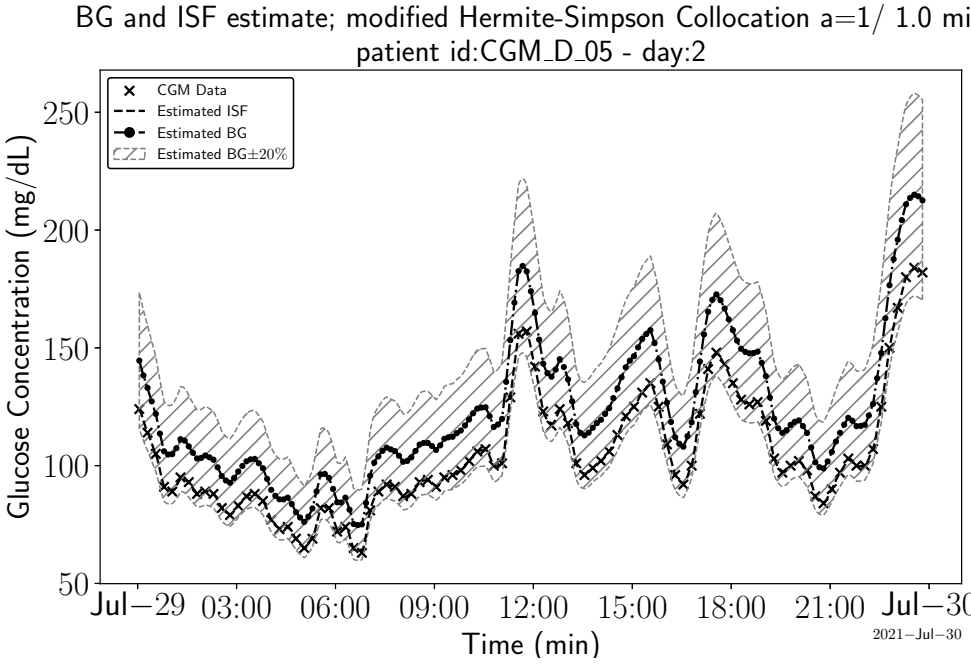


Figure 4.10: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM D 05` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1\,\text{min}$.

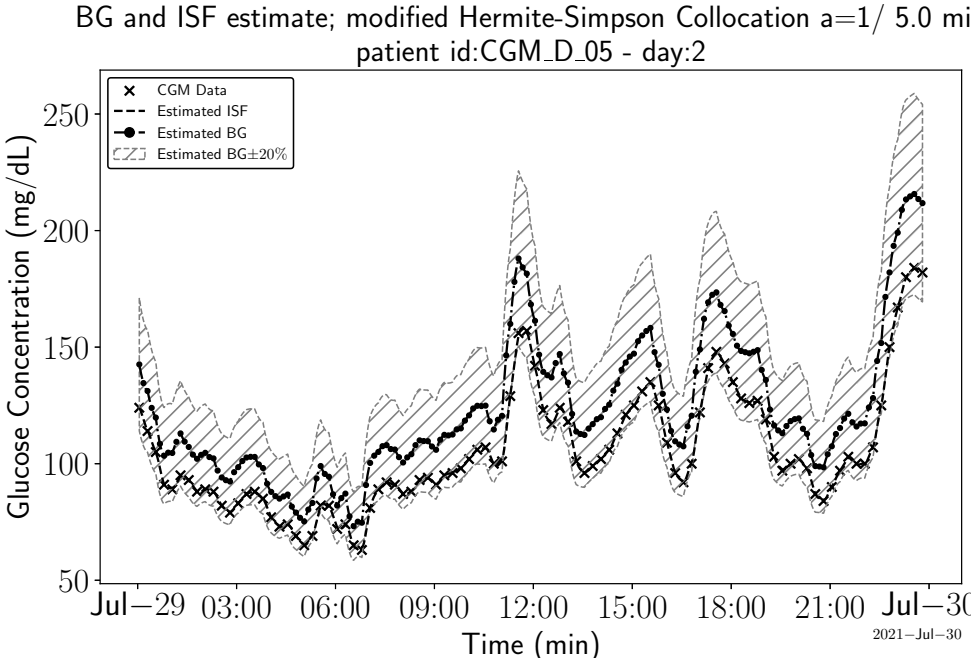


Figure 4.11: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM D 05` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1/5$min.

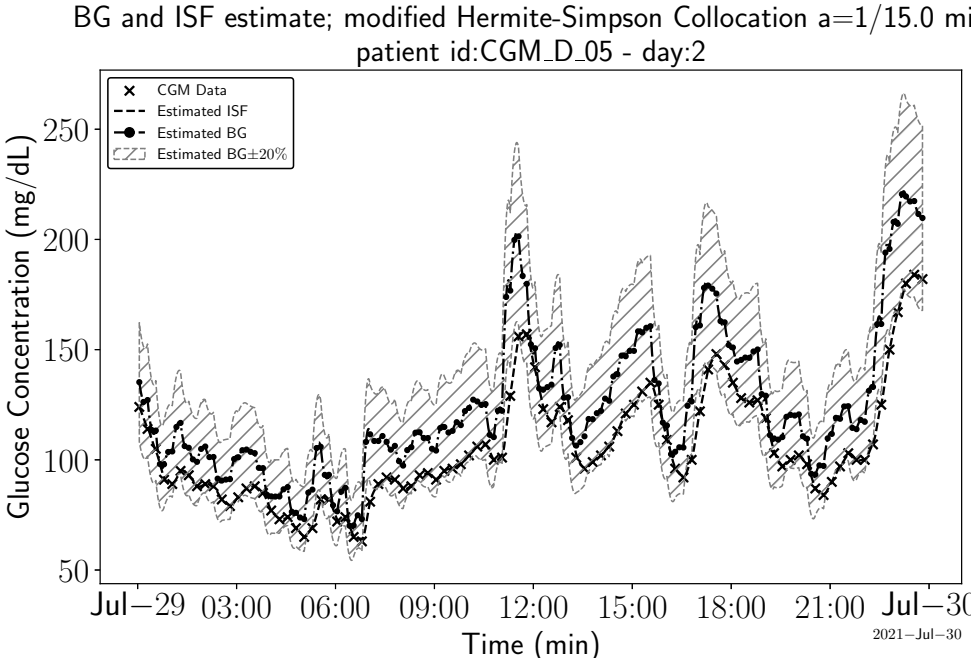


Figure 4.12: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM D 05` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1/15$min.

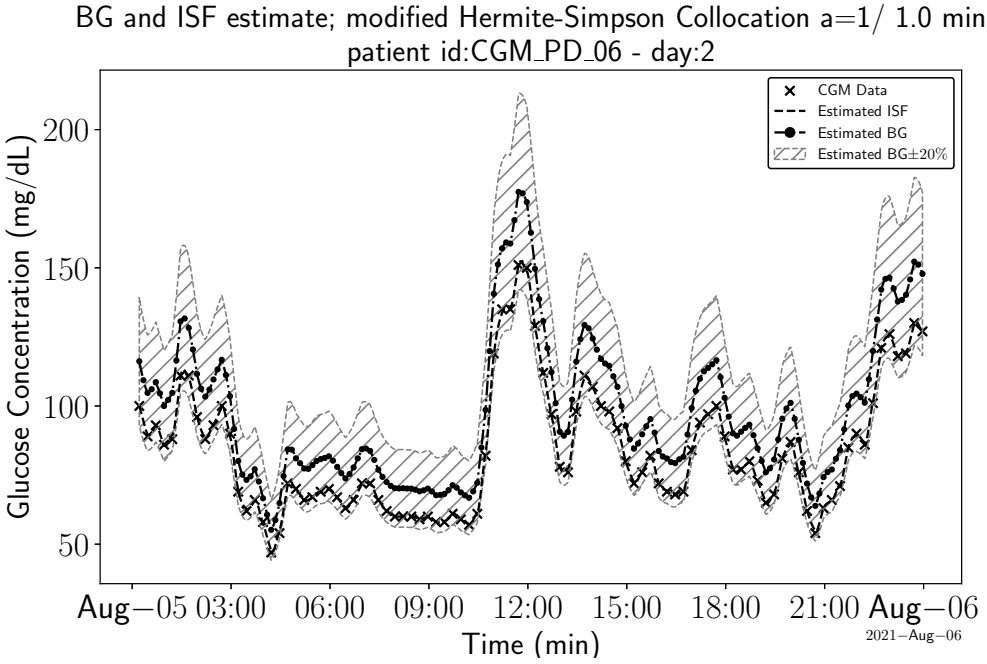


Figure 4.13: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM PD 06` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1\,\text{min}$.

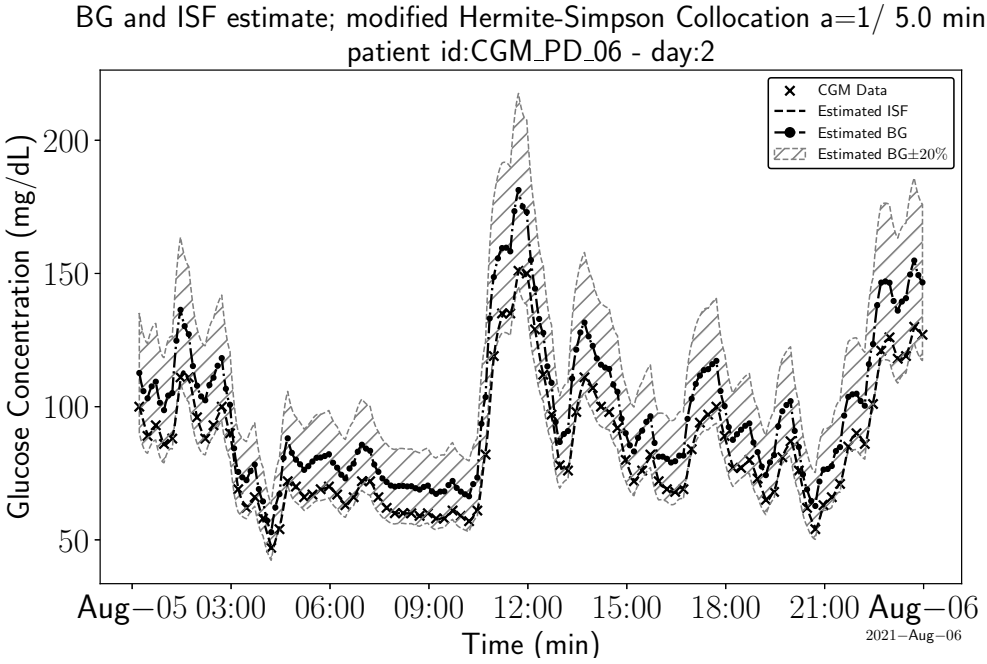


Figure 4.14: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM PD 06` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1/5$min.

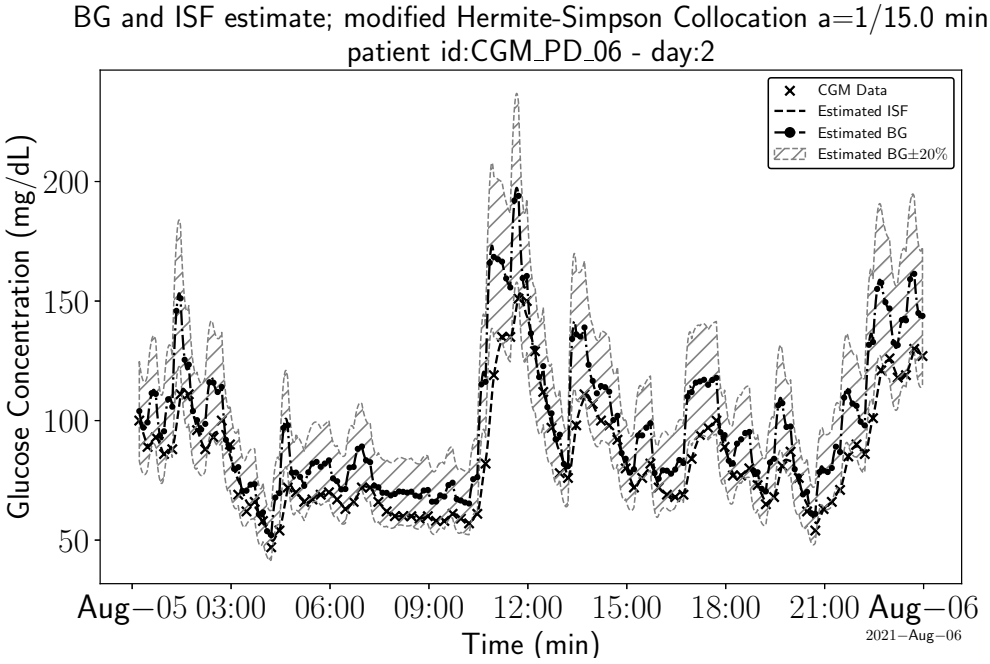


Figure 4.15: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM PD 06` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1/15$min.

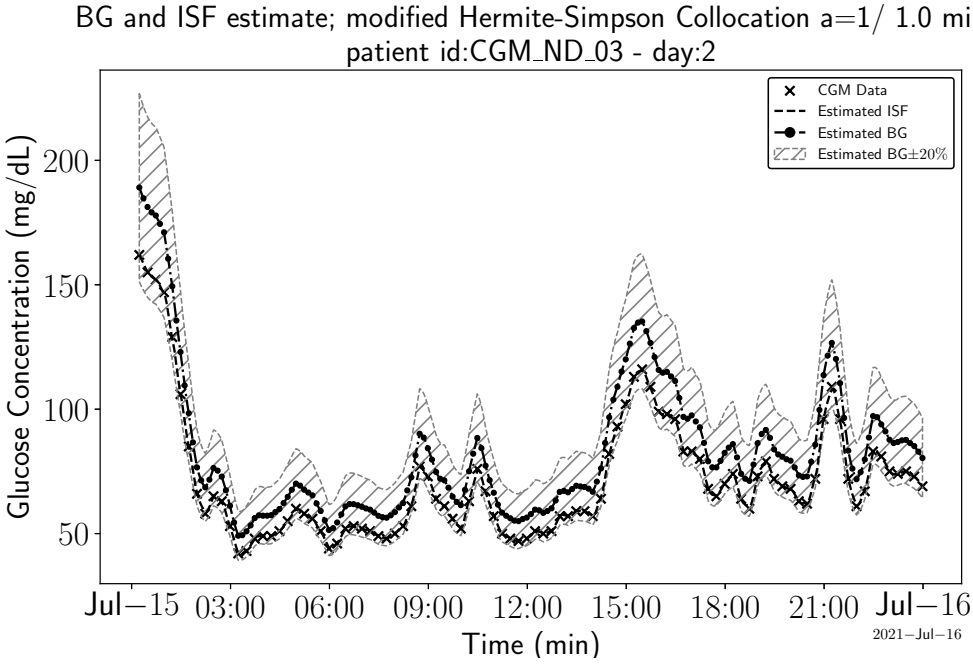


Figure 4.16: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM ND 03` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1\,\text{min}$.

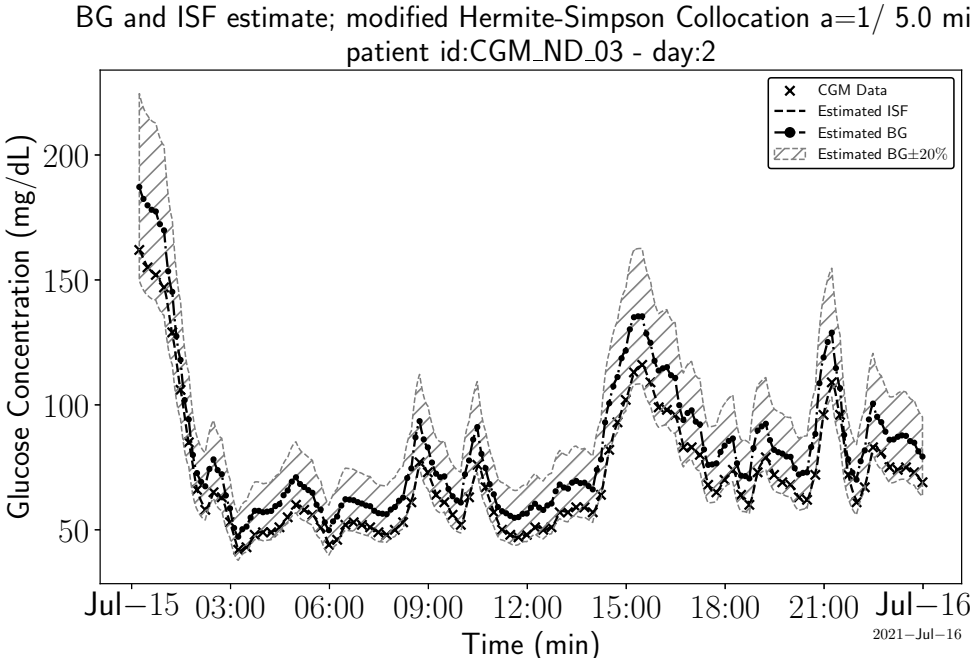


Figure 4.17: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM ND 03` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1/5$min.

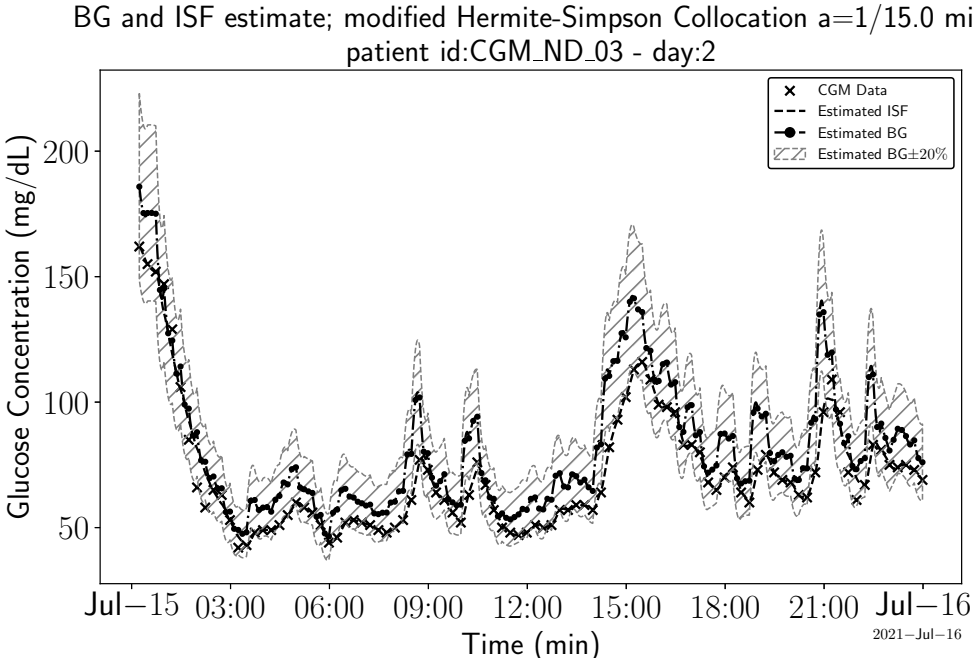


Figure 4.18: Estimated BG and ISF vs. time. The plot contains the day 2 full-day CGM trace of patient `CGM ND 03` which are measured ISF values represented by black crosses. The plot also contains the estimated ISF which is represented by a dashed black line and finally the plot also contains the estimated BG which is represented by a dot-dashed line with solid black circle markers representing the estimated BG values at the collocation points. The glucose concentration values which are within 20 % of the estimated BG values, at the same instant of time, have been shaded. These ISF and BG estimates were generated using the modified Hermite-Simpson collocation method with a $\tau_0 = 1/15$min.

Chapter 5

An Web Application for calculating CGM metrics

5.1 Introduction

There has been various metrics suggested to study many aspects of an individuals continuous glucose monitoring (CGM) record but Danne et al. [4] provides us with a consensus on these metrics and provides us with a minimum set of required metrics for evaluating an individuals CGM record.

We present a web application which, given a CGM record, calculates many of the metrics suggested by Danne et al. [4] and also estimates the continuous blood glucose concentration (BG) trajectory corresponding to the measured discrete interstitial fluid glucose concentration (ISF) trace.

5.2 Using `CGM Analyzer`

The `CGM Analyzer` (version 0.1 and version 2.0) web application calculates the essential metrics suggested by Danne et al. [4]. from any given CGM trace. These metrics are required by clinicians to properly analyze any given CGM trace. The `CGM Analyzer`, both versions 0.1 and 2.0, are capable of analyzing and calculating every metric for single CGM traces as well as multiple CGM traces at the same time in a batch mode (or Multi mode for version 2.0).

`CGM Analyzer` (version 0.1) is written in MATLAB and was initially compiled using the MATLAB Web App Compiler R2022a, but later have been compiled using the subsequent versions of the Web App compiler. `CGM Analyzer` (version 2.0) has also been written in MATLAB and has been compiled using MATLAB Web App Compiler R2024b.

The `CGM Analyzer` was initially designed to evaluate CGM traces, by calculating various metrics, which were generated by the Abbott FreeStyle Libre Pro CGM device. `CGM Analyzer` (both versions 0.1 and 2.0) follows strict conditions for the CGM traces that can be processed. The Web Apps takes CGM traces in the `.txt` and `.csv` extensions only. The CGM trace file will only be considered valid if it is in the following format

Jane Doe

ID	Time	Record Type	Hystoric Glucose (mg/dL)
101336	2021/09/16 06:30	0	180
101337	2021/09/16 06:45	0	176

The Web Apps, both version 0.1 and version 2.0, requires the glucose concentration readings to be in mg/dL, any other format for the CGM data files will not be accepted as valid data files by the Web Apps. In case of missing measurements or values the entry in the CGM data file is ignored.

`CGM Analyzer` is located here in url: https://digimed.acads.iiserpune.ac.in/fgm-tools The link is also provided in Majumdar et al. [5] where the `CGM Analyzer` (version 0.1) was first introduced. Therefore, a user can follow the link and click on **CGM Analyzer** to open the MATLAB Web Application Server where, `CHM Analyzer's` both versions are displayed as tiles/icons. The user can see a tile for `CGM Analyzer` version 0.1 and version 2.0.

5.2.1 Using `CGM Analyzer` version 0.1

`CGM Analyzer` version 0.1, is one of many web application created over the years that allow the user to upload a CGM trace and calculate various clinically relevant

metric using the uploaded CGM trace. This application in its initial version 0.1, is our first attempt at building one such application. This web application can be operated in two modes, a `Mono` mode and an `Multi` mode.

1. **The `Mono` mode:** The step-by-step instructions to follow to use `CGM Analyzer` version 0.1 to analyze a singe CGM trace is as follows

   (a) **Loading the CGM traces.** On the top left corner a `Load` button will appear. Please click on this button and select the appropriate input data file containing the desired CGM trace to be analyzed. Please allow a few moments for the file to load. As the file loads the ISF and calculated/estimated BG trajectory of the CGM traces will be available as a graph. The `Load` button will convert to an `Erase` button and `Data Sufficiency Lamp` will turn green to indicate a data-sufficient[4] CGM trace. If the `Data Sufficiency Lamp` turns red and no error messages are posted, we can assume that the CGM trace is not data-sufficient. If there is an error loading the input data file an appropriate error message will be displayed and the `Data Sufficiency Lamp` will turn red.

   (b) **Calculate the metrics.** Now, we can push the `Calculate` button on the top right to calculate the values of all the available metrics. Please allow a few moments for the calculation to complete.

   (c) **Downloading the calculated metrics.** After the calculation of all the metrics are complete the `Calculate` button will turn into an `Print` button. This button can be used to obtain a text file containing the calculated values of the metrics. The `Erase` button can then be used to reset the web application.

2. **The `Multi` mode:** The steps to follow to use the `CGM Analyzer` to analyze a multiple CGM traces in a batch is as follows

   (a) **Loading the CGM traces into the application.** On the top left corner a `Load` button will appear. The user can click on this button and select the appropriate input data files containing the desired CGM

traces to be analyzed. It may take several minutes to complete the loading process. A progress bar will indicate the state of file loading and processing. After completion the total number of files selected and the number of valid and invalid files will be shown. The `Load` button will change to an `Erase` button.

(b) **Calculating the desired metrics.** After successfully loading all the desired CGM traces, the `Calculate` button can be click so that all the metrics for each of the selected CGM traces are calculated. The `Calculate` button will also change to a `Print` button.

(c) **Downloading the calculated metrics.** The user can click the `Print` button to download the calculated metrics for each of the CGM trace files. The **Erase** button can then be used to reset the web application and repeat the steps to analyze another set of CGM traces.

5.2.2 Using `CGM Anlayzer` version 2.0

The `CGM Anlayzer` version 2.0, is the next iteration of the `CGM Analyzer` Web Application. This version of the `CGM Analyzer` Web Application uses MATLAB Stateflow for its control and interface logic. The usual metrics suggested by Danne et al. [4] have been implemented in `CGM Analyzer` version 2.0, just as it was done in version 1.0. This new version however has an additional feature, the estimation of the continuous BG corresponding to the discrete ISF trace. We describe in the next section how this continuous estimate of the BG is obtained. The metrics suggested by Danne et al. [4] can be calculated in version 2.0 of `CGM Analyzer`. The algorithms for calculating each metric is almost similar, with a few exceptions like the duration metric. Version 2.0 of `CGM Analyzer` therefore should give the same value for the metrics calculated by version 0.1. The visualization of the CGM trace has changed to a single day window. this allows for proper analysis of everyday CGM variation and also allows quick BG estimation. The usual functions like zooming, panning, etc are still available and just like version 0.1 are provided by the MATLAB back-end.

The `CGM Analyzer` (version 2.0) has two modes the `Mono` mode and the `Multi` mode. In the `Mono` mode the `CGM Analyzer` app can be used to analyze a single

CGM trace file. The metrics and the corresponding continuous BG can be calculated and analyzed for the given CGM trace. The calculated metrics can then also be downloaded into an excel spreadsheet. The day-to-day CGM trace along with the estimated BG trace can also be visualized and analyzed in this `Mono` mode.

In the `Multi` mode however, the application can be used to analyze a group of CGM trace files. The group of CGM trace files can be selected and uploaded after which the application can calculate all relevant metrics and generate an excel spreadsheet ready for download. As mentioned earlier, in the `Mono` mode the CGM trace can be visualized along with the estimated BG trace in a day-wise 24 hour window however, in the `Multi` mode, the CGM traces cannot be visualized and only the metrics can be calculated and later downloaded.

The step-by-step instructions to use the `CGM Analyzer` (version 2.0) to evaluate and analyze a given CGM trace for the two modes are given below:

1. **The `Mono` mode**: The Mono mode can be activated by switching the toggle switch on the top right corner of the application to `Mono`. The `Load` button on the top left of the application will be activated and will be the only button active till a CGM trace is loaded. The steps that need to be followed to load and analyze any given CGM trace in the `Mono` mode are listed below:

   (a) **Loading a CGM trace:** To load a CGM trace into the application in the `Mono` mode we need to click the `Load` button and select the CGM trace file in the appropriate location and click the `Open` button. Only `*.txt` and `*.csv` files are accepted and only one CGM trace can be loaded at a time. the CGM trace file chosen should also follow the same data format as stated earlier for version 0.1 of the `CGM Analyser`, i.e. the data in the CGM trace file should be of the format given below:

Jane Doe

ID	Time	Record Type	Hystoric Glucose (mg/dL)
101336	2021/09/16 06:30	0	180
101337	2021/09/16 06:45	0	176

Here, Jane Doe refers to the name of the patients, sometimes the patient ID, the ID column represents the measurement ID, the Time column represents the timestamp when the measurement was taken, the Record Type represents the type of the measurement and the Hystoric Glucose column represents the ISF in mg/dL.

After the CGM trace file has been selected and the `open` button clicked, the application would try to load the CGM trace data from the file. If the loading succeeds, the CGM trace of the first calender day the CGM device was active will be displayed. If the application fails to load the file an error message will be displayed. The data sufficiency of the uploaded CGM trace is also evaluated immediately after a successful upload. The lap at the top right corner of the application turns red if the CGM trace uploaded is not data sufficient and green if the trace is data sufficient. The `Load` button is rendered inactive while the `Erase` and `calculate` button is rendered active. If there was an error loading the CGM trace file, the `Load` button would be rendered inactive and the `Erase` button would be rendered active.

(b) **Visualizing the CGM trace.** As soon as a CGM trace file is successfully loaded, the data from the trace file is separated into smaller traces corresponding to the calender dates in their timestamps and by default the CGM trace corresponding to the first calender date the CGM trace was active is displayed. The CGM trace is shown as a solid blue line with blue crosses ("x") representing the data points from the CGM trace. The plot can be zoomed in/out, panned according to our requirements. The home button on the plot tool bar can be used to auto scale the plot. The lines representing the BG levels for Level 1 and 2 Hypoglycemia and Hyperglycemia are also present by default in this plot. The `Previous Day` and the `Next Day` buttons can be used to view the corresponding CGM traces. The calender day for which the CGM trace need to be visualized can also be chosen from the drop down date picker.

(c) **Estimating the continuous BG trace.** Given a CGM trace file

has been successfully loaded in the application, and the appropriate CGM trace for a calender day bas been chosen. The continuous BG trace corresponding to that days ISF trace can be estimated. We first need to select an required scaling factor, α, and then select an required collocation method from the drop down menu and the click on the `Estimate Blood Glucose` button. After a few moments the estimated continuous BG and the corresponding estimated continuous ISF trace will be displayed. The estimated BG is shown as a blue solid line, where as the estimated ISF is shown as a red solid line. The discrete CGM trace is also shown as solid black circles corresponding to each ISF measurement. This plot can also be analyzed, zoomed in/out and panned as per the user's choice.

(d) **Calculating the metrics.** After a CGM trace file is loaded successfully into the application, the `Load` button is rendered inactive and the `Calculate` and `Erase` button is rendered active. The lamp at the top right hand corner of the application indicated if the loaded CGM trace is data-sufficient or not. The lap turns green if the loaded CGM trace is data-sufficient otherwise it turns red. The remaining metrics however are not immediately calculated. To calculate the remaining metrics we need to click the `Calculate` button. After clicking the `Calculate` button, in a few moments all the metrics should be calculated and displayed in their appropriate text boxes. After clicking the `calculate` button the button would be rendered inactive and the `Download` and `Reset` buttons will be rendered active. The `Reset` button can be pressed to remove the current calculation from memory, so that the application becomes ready to repeat the calculation as a fresh calculation. It will render the `Download` and `Reset` button inactive and the `Calculate` button active again.

(e) **Download the calculated metrics.** After the `Calculate` button is clicked, the `Reset` and `Download` button is rendered active while the `Download` button is made inactive. Clicking the `Download` button would then open a dialog box where you can select/create an excel file of

`*.xlsx` extension where the major calculated metrics would be written. This file will then be downloaded into your system. Unlike version 0.1 of `CGM Anlayzer`, only afew metrics are saved into the excel file. These metrics include

- The CGM trace file name,
- the average interstitial fluid glucose concentration (aISF) in mg/dL,
- The inter quartile range (IQR) in mg/dL,
- the Standard deviation (SD) in mg/dL,
- the Coefficient of Variation (CV),
- the high blood glucose index (HBGI),
- the low blood glucose index (LBGI),
- the area under the curve (AUC) in mg h/dL,
- the Glycated Hemoglobin (HbA1c) calculated using the Nathan Model [3] in mmol/mol,
- the HbA1c calculated using the Direct Model [5] in mmol/mol,
- the HbA1c calculated using the Adapted Nathan Model [5] in mmol/mol,
- The percentage duration of time the CGM device was active,
- The duration of time the CGM device was active,
- The glycemic state, i.e., if the patients ISF is glycemic stable,
- The total no. of readings in the uploaded CGM trace file and,
- the total number of expected readings from the CGM trace file.

These calculated metrics can be repeatedly downloaded by clicking on the `Download` button and selecting files with unique names.

2. **The `Multi` mode:** The `Multi` mode, similar to the `mono` mode can be activated by toggling the switch at the top right corner of the application. After toggling the `Multi` mode switch, only the `load` button will be active. The labels, text fields and plot windows will be rendered inactive (hidden in `Multi` mode). This mode allows us to calculate the relevant metrics for a large number of CGM traces in bulk at the same time and download a large excel spreadsheet with the calculated metrics populating the spreadsheet.

The step-by-step instructions to use the `CGM Analyzer` version 2.0 to evaluate multiple CGM traces in bulk are as follows:

(a) **Loading multiple CGM traces.** In the `Multi` mode a number of CGM traces can be loaded into the application at the same time. The CGM trace files however have to be of the format described for the `Mono` mode. After clicking the `Load` button a dialog box will appear which will allow multiple `*.txt` and `*.csv` files to be loaded into the application. After selecting the desired files we need to click on the `open` button. This will start loading the selected CGM traces into the application. The `Load` button will then be rendered inactive. After loading the selected files a message will appear informing the user about the number of files that have been successfully loaded into the application. The lap at the top right part of the application will glow green if at least one CGM trace file is loaded otherwise, it will glow red. The `Calculate` and `Erase` button will be rendered active. The `Erase` button can be used to remove all the uploaded CGM trace and start afresh.

(b) **Calculating the metrics** After successfully loading the selected CGM trace files, we can click on the newly activated `calculate` button to calculate all the metrics for each of the CGM traces loaded. This may take a while and an appropriate message will be displayed once all the metrics have been processed. The `Download` and `Reset` button will be rendered active. The `Reset` button can be used to remove the current calculation and allow the application to run a fresh calculation.

(c) **Downloading the Calculated metrics.** After calculating all the metrics for each of the CGM traces selected and uploaded into the application we can download the some of the calculated metrics using the `Download` button. Clicking the `Download` button opens a dialog box where an excel file can be selected/created with `*.xlsx` extension where a subset of the calculated metrics for each CGM trace uploaded will be saved. The list of calculated metrics which are saved are the same as that for the `Mono` mode.

5.3 Estimating continuous BG from a CGM trace

The continuous BG trace in the web application is estimated using trajectory optimization as described in Chapter 4. We assume that the relationship between BG and ISF can be represented by a simple diffusion ordinary differential equation (ODE). We then impose additional constraints, called path constraints, on the BG and ISF trajectories, or curves, constructing essentially an optimal control problem where the BG acts as the control/driving force and the ISF represent the state. The objective of this optimal control problem is to minimize the ordinary least squares (OLS) error between the estimated ISF and the measured ISF provided by the CGM device while satisfying the system ODEs and the path constraints. The measured ISF provided by the CGM devices are discrete time series and moreover as we are numerically solving the optimal control problem, we can only consider solutions which belong to finite-dimensional spaces and can be represented properly by a computer. The continuous trajectories in the optimal control problem are therefore discretized at specific collocation points according to a collocation method. The problem is then numerically solved for the values of the trajectories at those specific collocation points. This converts the problem essentially into a nonlinear programming (NLP) problem. The standard NLP solvers can then be used to numerically solve these problem. In our case we used two collocation methods, the trapezoidal collocation and a modified Hermite-Simpson collocation method for discretizing the optimal control problem. We then used the `fmincon` function provided by MATLAB to find the solution of the corresponding NLP problem. The estimated values of the ISF and BG trajectories at the collocation points are then appropriately interpolated according to their collocation methods to the approximate continuous solutions of the optimal control problem. This method essentially provides an estimate of the continuous ISF and BG trace which could give rise to the ISF trace when measured by the CGM device. This method involved the non-dimensional ODE and therefore could be scaled by an α parameter. `CGM Analyzer` version 2.0 also allows for this α parameter to be tuned so that the user can explore different ISF and BG estimates and choose an appropriate estimate.

5.4 Metrics

The metrics which the `CGM Analyzer` web application calculates are described in the following subsections. These are the metrics recommended by Danne et al. [4].

5.4.1 Duration of the record

The Duration of the record represents the time elapsed between the first and the last CGM measurement. Using this duration we calculate the no. of days the for the CGM record which plays a vital role in every percentage time in range calculation later.

The duration is expressed in the following format `days, (DD:hh:mm)` in version 0.1 of the `CGM Analyzer`; where `(DD:hh:mm)` represents the duration in days, hours and minutes, and `days` represents the duration rounded off to the nearest number of days. For example a duration 16 days 10 hours and 30 minutes long would be displayed as `16, (16:10:30)` but a duration 16 days 12 hours and 30 minutes long would be displayed as `17, (16:12:30)`. In the version 2.0 of the `CGM Analyzer` application the duration is expressed in the format `DD:hh:mm:ss` where, `dd` represents the no of days elapsed between the first and last CGM measurement, the remaining symbols have their usual meaning.

The duration is calculated by subtracting the `datetime` value of the time stamp of the first reading in the CGM record, provided by the input data file, from the `datetime` value of the time stamp of the last reading in the CGM record. The no. of days, hours and minutes elapsed are calculated using the `days`, hours, minutes MATLAB functions.

5.4.2 Time CGM Device active

CGM devices like the Abbott Freestyle Libre Pro remain active on the patients body for approximately 14 days and take regular readings at 15 min intervals. These devices however occasionally miss a few readings. The percentage Time CGM Device active metric actually calculate the fraction of ISF readings that have been taken out of all the possible reading the CGM device is supposed to take.

We calculate the percentage of possible CGM recordings from the ratio of number of measurements present in the CGM trace to the number of measurements that could have been taken in the no. of days the device was active. The output is rounded off to 1 decimal place using MATLAB's `round` function and displayed.

5.4.3 Data Sufficiency Lamp

The Data Sufficiency Lamp indicates if the CGM trace obtained from the input data file is data sufficient[4]. The lamp is green when the CGM trace is data-sufficient otherwise it is red. In version 0.1 of the `CGM Analyzer` application, the lamp is also red if there is an error in accessing the data present in the file or any error in uploading the input data file.

As suggested in Danne et al. [4], the CGM record is considered data sufficient if the no. of days calculated for the duration is greater than or equal to 14 days and the percentage of possible CGM recordings is greater than 70%.

We first calculate the duration of time(days) the CGM device was active and then the percentage of time the CGM device was active(as shown in Sec 5.4.2). If the given CGM data file is data sufficient then we turn the data sufficiency lamp green, otherwise it turns red.

5.4.4 Average Interstitial Glucose (aISF)

The Average Interstitial Glucose value is the average of all the ISF readings provided in the input data file. As the input data file is reporting the ISF values in mg/dL the average interstitial fluid glucose concentration (aISF) calculated is also in the same units.

The average is calculated using MATLAB's `mean` function and the result is rounded off to integer values using MATLAB's `round` function.

The equation for calculating aISF from the ISF values provided in the input data file can be written as

$$\text{aISF} = \sum_{i=1}^{i=\text{N}} \text{ISF}_i, \tag{5.1}$$

where aISF is average interstitial fluid glucose concentration (aISF) in mg/dL, ISF is the interstitial fluid glucose concentration (ISF) in mg/dL and, N the total number of ISF readings in the input data file.

5.4.5 Standard Deviation (SD)

The Glycemic Standard Deviation is calculated as the Standard deviation (SD) of the all the ISF readings provided in the input data file. As the input data file is reporting the ISF values in mg/dL the SD calculated is also in the same units.

This is labeled as `SD` in the web application and it is calculated using MATLAB's `std` function. The result is then rounded off to 2 decimal digits using MATLAB's `round` function.

5.4.6 Coefficient of Variation (CV)

The Glycemic Coefficient of Variation is calculated as the Coefficient of Variation (CV) of all the ISF readings provided in the input data file. As the input data file is reporting the ISF values in mg/dL the CV calculated is also in the same units. The calculated CV is rounded off to 2 decimal figures using MATLAB's `round` function.

The formulae used to calculate CV is as follows

$$\mathrm{CV} = \frac{\mathrm{SD}}{\mathrm{aISF}}, \tag{5.2}$$

where CV is the CV in mg/dL, SD is the SD in mg/dL and, aISF represents the aISF in mg/dL.

5.4.7 Glycemic Stability Lamp

The glycemic stability lamp demonstrates if the uploaded input data file contains a CGM trace with glycemic stability. A CGM record is suggested[4] to be glycemic stable if the coefficient of variation is below 36%. Therefore, after the calculation of the CV, if the value is less than 36% the lamp labeled `Glycemic Stable` is set to green, otherwise it is set to red.

5.4.8 Inter Quartile Range (IQR)

The metric IQR represents the inter quartile range (IQR) of the measured ISF values of the input dataset. It is calculated using the MATLAB function `iqr`. As the input data file is reporting the ISF values in mg/dL the IQR calculated is also in the same units and is rounded off to 1 decimal figure using MATLAB's `round` function.

5.4.9 Glucose Indexes HBGI and LBGI

The web application calculates the high blood glucose index (HBGI) and the low blood glucose index (LBGI) from all the ISF readings provided in the input data file.

The calculation is based on the formulae suggested in Kovatchev et al. [2] and Kovatchev et al. [1]. The calculated HBGI and LBGI is rounded to 2 decimal places using MATLAB's `round` function.

The algorithm used to calculate HBGI and LBGI is provided below. We consider the functions

$$tr(g) = 1.509\left((\log g)^{1.084} - 5.381\right) \tag{5.3}$$

$$r(g) = 10 \times tr\,(g) \tag{5.4}$$

Now, for all N ISF measurements g_i we calculate $rl(g_i)$ and $rh(g_i)$, where,

$$rh\,(g) = \begin{cases} r\,(g) & r\,(g) > 0 \\ 0 & r\,(g) \leq 0 \end{cases} \tag{5.5}$$

and

$$rl\,(g) = \begin{cases} r\,(g) & r\,(g) < 0 \\ 0 & r\,(g) \geq 0 \end{cases} \tag{5.6}$$

Then we calculate HBGI and LBGI as

$$\text{HBGI} = \sum_{i=1}^{N} rh\,(g_i) \tag{5.7}$$

$$\text{LBGI} = \sum_{i=1}^{N} rl\,(g_i) \tag{5.8}$$

5.4.10 Estimated Glycated Hemoglobin (HbA1c)

The web application provides three estimates of HbA1c. HbA1c in all three cases are estimated from the calculated aISF in mg/dL.

The first HbA1c estimate is based on the equations suggested by Nathan et al. [3], it is the standard Nathan Model used by the Abbott Freestyle Libre Pro device. The formulae for the HbA1c estimate as suggested by Nathan et al. [3] can be written as

$$\text{HbA1c} = 0.380 \times (1.05 \times \text{aISF}) - 5.60, \tag{5.9}$$

where HbA1c is the HbA1c in mmol/mol and aISF is the calculated aISF in mg/dL. The calculated HbA1c value is rounded off to 1 decimal place using MATLAB's `round` function.

The remaining two HbA1c estimates are based on two models published in Majumdar et al. [5], the direct model and the adapted Nathan model.

The equation for the Direct Model[5] HbA1c estimate can be written as

$$\text{HbA1c} = 0.319 \times \text{aISF} + 16.73, \tag{5.10}$$

where HbA1c is the estimated HbA1c in mmol/mol and aISF is the calculated aISF in mg/dL. For further details of the formulae or calculation for the confidence interval please see the supplementary sheet for Majumdar et al. [5].

The equation for the Adapted Nathan Model[5] HbA1c estimate can be written as

$$\text{HbA1c} = 0.380 \times (1.17 \times \text{aISF}) - 5.60, \tag{5.11}$$

where HbA1c is the estimated HbA1c in mmol/mol and aISF is the calculated aISF in mg/dL. For further details of the formulae or calculation for the confidence interval please see the supplementary sheet for Majumdar et al. [5].

The estimated HbA1c values in all three cases are rounded off to 1 decimal place using MATLAB's `round` function. The HbA1c estimates of the direct model and the adapted Nathan model also have their 95 % prediction intervals displayed along with the HbA1c estimate.

5.4.11 Area under the curve

The area under the curve (AUC) represents the area under the ISF curve obtained from the CGM trace provided by the input data file. The Abbott Freestyle Libre Pro CGM records contain ISF readings in intervals of 15 min. We assume that the input data file also has a similar format and has ISF readings at 15 min intervals. We calculate an approximate area under the ISF curve.

To calculate the area under the ISF curve we make an assumption that the ISF value remains constant for 15 min till the next ISF measurement is taken. The area under the ISF curve is then approximated as the sum of the areas of rectangles with a height provided by the ISF measurement and a width of 15 min.

The equation used is as follows

$$\mathrm{A} = \sum_{t=1}^{\mathrm{N}} 15 \times \mathrm{ISF}_i \tag{5.12}$$

where N is the total number of ISF measurements, ISF_i the ith ISF measurement and A the area under the glucose curve in $mg/dL\ min$. The area is then converted to $mg/dL\ hr$ and rounded off to an integer using MATLAB's `round` function and displayed as output.

5.4.12 Duration in ranges

The amount of time the ISF values of the CGM trace provided by the input data file are found in certain ranges is represented by the duration in range metric.

As suggested in Danne et al. [4] for the individual to be considered in the Hypoglycemic Level 1 state the glucose readings should be between 54 mg/dL and 70 mg/dL, for the Hypoglycemic Level 2 state the glucose readings should be below 54 mg/dL and similarly for the Hyperglycemic level 1 state the glucose readings should be above 180 mg/dL and for the Hyperglycemic Level 2 state it should be above 250 mg/dL.

We calculate the no. of glucose readings which are greater than or equal to 54 mg/dL but below 70 mg/dL and multiply it with 15 min (measurement interval) to obtain the amount of time elapsed in Hyperglycemic Level 1 state. The time elapsed in Hypoglycemic Level 2 state is calculated similarly by counting the

no. of glucose readings below 54 mg/dL and multiplying it by 15. In the Hyperglycemic Level 1 and Level 2 cases we calculate the no. of readings equal and above 180 mg/dL and 250 mg/dL and multiply it by 15 to get the duration of time the individual was in the respective states. These calculations are carried out over the entire length of data available and for the ranges 06 : 00 to 24 : 00, the wake range and for the range 00 : 00 to 06 : 00, the sleep range. These durations are then displayed appropriately corresponding to the glycemic range.

5.4.13 Average Duration per Day

The average duration per day metric represents the average duration of time the ISF values of the CGM trace provided by the input data file is in a given glycemic state in a day. Similar to the duration in range metric(as described in Sec 5.4.12) the duration of time the ISF values remain in each glycemic state is calculated for each day and an average value is calculated and displayed. These calculation however only take into account the nearest number of day from the duration i.e. if we have a duration of 16 days 10 hours and 10 minutes, the time in the different ranges will be calculated for each day and will be averaged over 16 days only and if we have a duration of 16 days 16 hours and 10 minutes, the time in the different ranges will be calculated for each day and will be averaged over 17 days. The Averaged durations per day is then displayed in minutes.

5.4.14 Percentage time in ranges

The percentage of time the glucose measurements are in the respective Hypoglycemic and Hyperglycemic ranges are also calculated and displayed as percentages rounded off to 1 decimal points. The number of glucose readings in the required range is counted and the total number of possible readings that should have been taken in the duration that the device was active is calculated. Their ratio in percentage gives us the percentage of time spent in the corresponding ranges.

5.4.15 No. of Episodes

The No. of Episodes metric represents the number of episodes the ISF values of the CGM trace in the input data file has endured in a particular glycemic state. We present the No. of Episode metric for the Level 1&2 Hypoglycemia and Level 1&2 Hyperglycemia states. The No. of Episodes for all the glycemic states are also calculated for the time range 06 : 00 to 24 : 00, the wake range and for the time range 00 : 00 to 06 : 00, the sleep range. If there are consecutive reading within a particular state then the individual is considered to be having an episode of the corresponding Hypoglycemic or Hyperglycemic state. In the case of missing measurements in any episode, we assume the individual remains in the episode unless the next available measurement suggests otherwise. The results are displayed as integers in the web application..

References

[1] Kovatchev, B. P.; Cox, D. J.; Gonder-Frederick, L. A.; Clarke, W. Symmetrization of the blood glucose measurement scale and its applications. *Diabetes Care* **1997**, *20*, 1655–1658, DOI: 10.2337/diacare.20.11.1655, https://diabetesjournals.org/care/article/20/11/1655/21162/Symmetrization-of-the-Blood-Glucose-Measurement (cit. on p. 140).

[2] Kovatchev, B. P.; Cox, D. J.; Kumar, A.; Gonder-Frederick, L.; Clarke, W. L. Algorithmic evaluation of metabolic control and risk of severe hypoglycemia in type 1 and type 2 diabetes using self-monitoring blood glucose data. *Diabetes technology & therapeutics* **2003**, *5*, 817–828, DOI: 10.1089/152091503322527021, https://www.liebertpub.com/doi/10.1089/152091503322527021 (cit. on p. 140).

[3] Nathan, D. M.; Kuenen, J.; Borg, R.; Zheng, H.; Schoenfeld, D.; Heine, R. J.; Group, A.-D. A. G. (S. Translating the A1C assay into estimated average glucose values. *Diabetes care* **2008**, *31*, 1473–1478, DOI: 10.2337/dc08-0545, https://diabetesjournals.org/care/article/31/8/1473/

28589/Translating-the-A1C-Assay-Into-Estimated-Average (cit. on pp. 134, 141).

[4] Danne, T.; Nimri, R.; Battelino, T.; Bergenstal, R. M.; Close, K. L.; DeVries, J. H.; Garg, S.; Heinemann, L.; Hirsch, I.; Amiel, S. A., et al. International consensus on use of continuous glucose monitoring. *Diabetes care* **2017**, *40*, 1631–1640, DOI: 10.2337/dc17-1600, https://diabetesjournals.org/care/article/40/12/1631/37000/International-Consensus-on-Use-of-Continuous (cit. on pp. 127, 129 sq., 137 sqq., 142).

[5] Majumdar, S.; Kalamkar, S. D.; Dudhgaonkar, S.; Shelgikar, K. M.; Ghaskadbi, S.; Goel, P. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full (cit. on pp. 128, 134, 141).

Chapter 6

Discussion

The daily functioning of the human body involves numerous metabolic processes, many of which are modulated by amount of food intake, type of food intake, amount and intensity of exercise, etc. One such set of metabolic processes are the ones involved with the glucose-insulin dynamics of the human body. These metabolic processes are involved in the consumption of glucose by different tissues in the body, production of glucose by the liver, and eventually the maintenance of the blood glucose concentration (BG) in the human body. Insulin is a hormone that plays an essential part in many of these metabolic processes, and therefore eventually plays a major part in the glucose-insulin dynamics of the human body. It is involved with the maintenance of glucose homeostasis i.e. maintenance of BG in the human body, fat deposition in the adipose tissues, etc. The number of cases of metabolic disorders like type 1 diabetes (T1D) and type 2 diabetes (T2D) are gradually increasing and it is important that we reliably understand these disorders and develop methods/treatments to manage and eventually cure them. As many of these metabolic disorders, especially T1D and T2D, are intrinsically involved with the glucose-insulin dynamics of the human body, the state of the glucose-insulin dynamics in the human body is used as a probe to identify and evaluate the extend of these disorders. This involves measuring the fasting blood glucose concentration (FBG) and post-prandial blood glucose concentration (PPBG) to study the deviation of the glucose dynamics from homeostasis or measuring Glycated Hemoglobin (HbA1c) to study the deviation of the BG over a long

period of time, usually 120 days.

Understanding the glucose-insulin dynamics of the human body is necessary to study these metabolic disorders. Studying models describing various features of the human glucose-insulin dynamics is an active area of research. Traditionally, probing the glucose-insulin dynamics of the human body involved regular blood tests for FBG, PPBG, HbA1c, etc. These methods are not only painful but also expensive. Modern fast glucose monitoring (FGM) and continuous glucose monitoring (CGM) devices have provided a great tool for probing the glucose-insulin dynamics of the human body. These CGM devices are relatively cheap (w.r.t previous sensors from the previous decade) and their sensors are quite robust. Studies can be arranged where a large number of people could be given these CGM devices and be asked to follow a certain dietary routine or any particular exercise routing. These CGM devices contain sensors which probe the interstitial fluid and periodically measure the glucose concentration in the fluid. This devices essentially give us a partial snapshot of the glucose-insulin dynamical system at fixed intervals of time. The CGM therefore has the potential to be a remarkable tool for understanding the glucose-insulin dynamics of the human body. With increasing accuracy of these CGM devices we get increasingly reliable discrete time representations of the continuous interstitial fluid glucose concentration (ISF). With an appropriate algorithm we can interpolate through these points and obtain an approximation of the continuous ISF or impose a set of constraints representing the dynamics between BG and ISF and obtain a reliable approximation of the continuous BG too.

In this thesis we first show, using the combined Pune-2021 and Joshi-2018 dataset, that the standard method, i.e. the Nathan model [1], used for estimating HbA1c from any given CGM trace provides HbA1c estimates that are statistically unreliable at a population level when used with CGMs traces obtained from T2D patients, pre-diabetic patients or non-diabetics from the Indian population. We then introduced two new methods, using linear and constrained regression analysis, for reliably estimating HbA1c from any given CGM trace of an individual from the Indian population. We used a part of the combined Pune-2021 and Joshi-2018 dataset, that was kept separate, to validate both the Direct Model [3] and the

Adapted Nathan Model [3] to confirm that both that these two models provides HbA1c estimates which are statistically indifferent from the experimentally measured HbA1c values at a population level. We then constructed a web application using `Matlab` which allows anyone to upload a CGM trace and calculate all the important metrics described in Danne et al. [2] along with the HbA1c estimates provided by the Nathan model [1], the direct and adapted Nathan model [3].

We now have two new models, the direct [3] and the adapted Nathan model [3], that can provide, at a population level, reliable estimates of HbA1c from CGM traces of non-diabetic, pre-diabetic and/or T2D diabetic patients from the Indian population. These models however have not been validated or tested with CGM traces and HbA1c measurements from T1D patients from the Indian population. Although a clear relationship between average blood glucose concentration (aBG) and HbA1c has been established, as previously described, it was still not clear how this relationship was affected by the gender, race and type of diabetes, T1D or T2D, exhibited by the patient. Salam et al. [4] showed using a dataset of 343 participants, which included 191 T1D and 152 T2D participants, that overall for a given average interstitial fluid glucose concentration (aISF) value, calculated from the given CGM traces, the HbA1c of the T2D participants were significantly higher than the HbA1c measured from the T1D participants. This should indicate that the overall relationship between HbA1c and aISF calculated from CGM traces are significantly different for T1D and T2D participant. Salam et al. [4] however also showed that after stratifying on racial terms, for a given aISF value there were no significant difference between the corresponding HbA1c for T1D and T2D. This showed that after stratifying on racial lines there were no significant differences between the relationship between aISF and HbA1c. We therefore expect the new direct model [3] and the adapted Nathan model [3] to provide HbA1c estimates which are statistically indistinguishable from experimentally measured HbA1c for Indian T1D patients. We therefore used the T1D 2022–2023 and T1D 2019–2020 T1D dataset together to validate the direct model [3], the Nathan model [1] and adapted Nathan model [3]. We verified that the HbA1c estimates provided by the Nathan model [1] and the adapted Nathan model [3] were statistically different from the experimentally measured HbA1c values. The HbA1c estimates obtained

using the direct model [3] however, were statistically indistinguishable from the experimentally measured HbA1c values. We therefore, now have a validated model, i.e. the direct model [3] that can be used to reliably estimate HbA1c from CGM traces of Indian pre-diabetic, non-diabetic, T2D diabetic and also T1D diabetic patients.

As we had mentioned earlier traditionally the metrics used for diagnosis and management of diabetes were all based on BG. They essentially were indicators for BG or indicators for a consistently high BG. Modern FGM or CGM devices however do not measure BG; they measure and provide only ISF measurements. Therefore there is a need for an algorithm that would reliably provide an estimate for the continuous BG give a discrete-time representation of the continuous ISF trace. We used trajectory optimization to obtain an estimate of the continuous BG with a suitable degree of smoothness. This would allow us and clinicians to refer to a smooth approximation of the actual BG trajectory when designing dietary and bolus schedules for insulin. Although We have not been able to provide an approximate of the continuous BG trajectory with an arbitrary small error. We will continue our work on the algorithm to provide a more reliable BG estimate.

References

[1] Nathan, D. M.; Kuenen, J.; Borg, R.; Zheng, H.; Schoenfeld, D.; Heine, R. J.; Group, A.-D. A. G. (S. Translating the A1C assay into estimated average glucose values. *Diabetes care* **2008**, *31*, 1473–1478, DOI: 10.2337/dc08-0545, https://diabetesjournals.org/care/article/31/8/1473/28589/Translating-the-A1C-Assay-Into-Estimated-Average (cit. on pp. 147 sq.).

[2] Danne, T.; Nimri, R.; Battelino, T.; Bergenstal, R. M.; Close, K. L.; DeVries, J. H.; Garg, S.; Heinemann, L.; Hirsch, I.; Amiel, S. A., et al. International consensus on use of continuous glucose monitoring. *Diabetes care* **2017**, *40*, 1631–1640, DOI: 10.2337/dc17-1600, https://diabetesjournals.org/care/article/40/12/1631/37000/International-Consensus-on-Use-of-Continuous (cit. on p. 148).

[3] Majumdar, S.; Kalamkar, S. D.; Dudhgaonkar, S.; Shelgikar, K. M.; Ghaskadbi, S.; Goel, P. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full (cit. on pp. 147 sqq.).

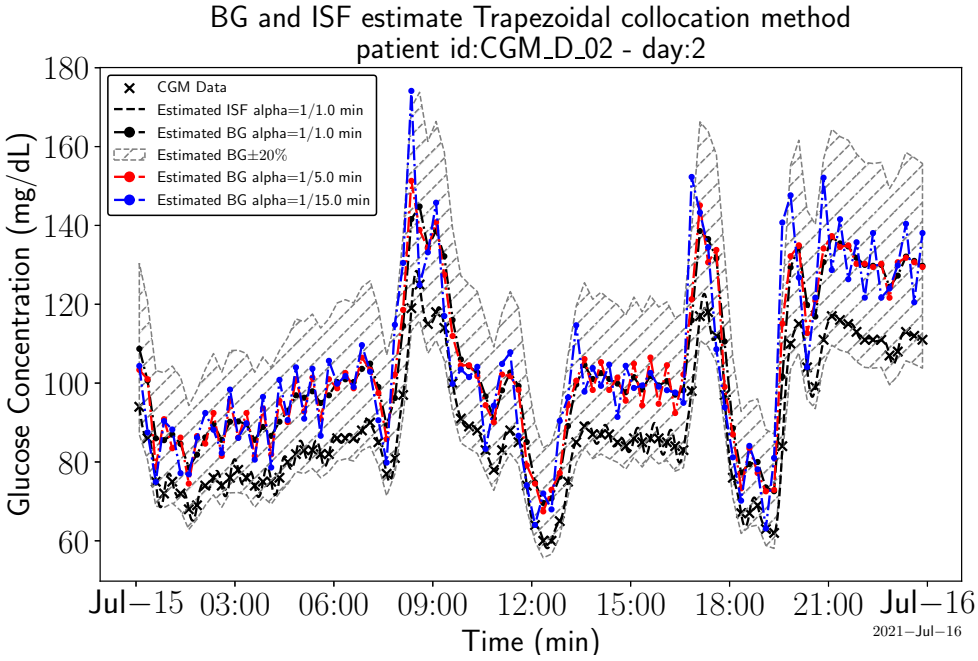
[4] Salam, M.; Bailey, R.; Calhoun, P.; McGill, J. B.; Bergenstal, R. M.; Price, D.; Beck, R. W. A comparison of continuous glucose monitoring estimated hemoglobin A1c in adults with type 1 or type 2 diabetes. *Diabetes Technology & Therapeutics* **2023**, *25*, 178–185, DOI: 10.1089/dia.2022.0387, https://www.liebertpub.com/doi/10.1089/dia.2022.0387 (cit. on p. 148).

Appendix A

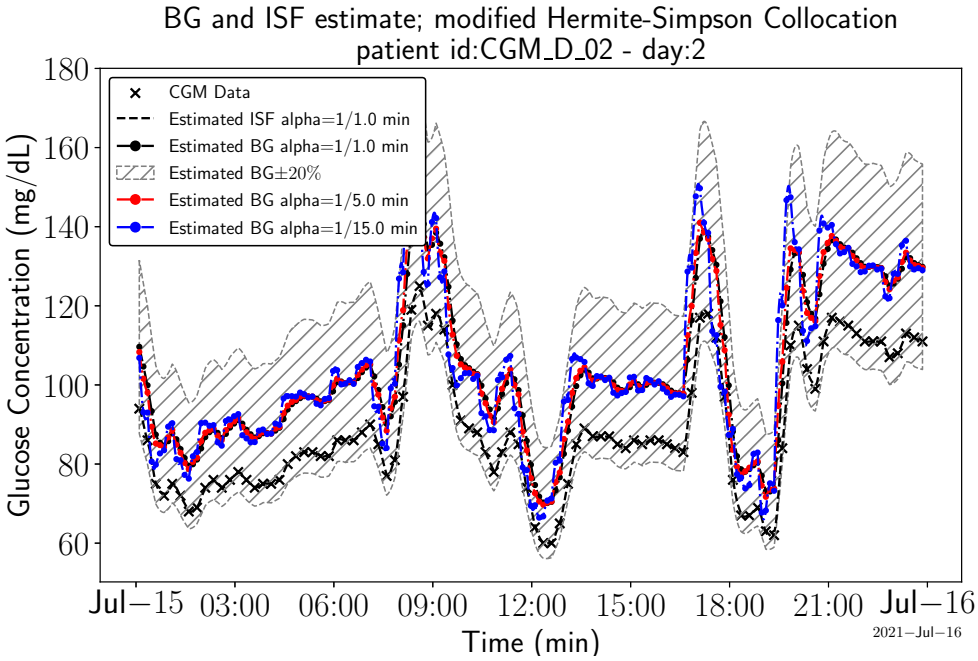
Blood glucose concentration (BG) trace estimates for the Pune-2021 dataset

Two new methods for estimating the continuous blood glucose concentration (BG) trajectory of any individual from their continuous glucose monitoring (CGM) trace were described in Chapter 4. Here, we used these two new methods for estimating the continuous BG trajectories for all the CGM traces from the Pune-2021 dataset. We provide below figures showing the continuous BG estimates using the trapezoidal collocation and the modified Hermite-Simpson collocation methods.

The figures show the estimated BG for each participants from the Pune-2021 dataset along with the measured interstitial fluid glucose concentration (ISF) values obtained from the day 2 full-day CGM trace. The measured ISF values are represented by black crosses. The estimated BG values are shown by dot-dashed lines (black for $\tau_o = 1\,\text{min}$, red for $\tau_0 = 1/5\text{min}$ and, blue for $\tau_0 = 1/15\text{min}$) along with a solid circle as marker for the estimated BG values at the collocation points. The estimated ISF for $\tau_0 = 1\,\text{min}$ is represented by a black dashed line. The shaded region represents the glucose concentration values that are within 20 % of the estimated BG for $\tau_0 = 1\text{min}$.

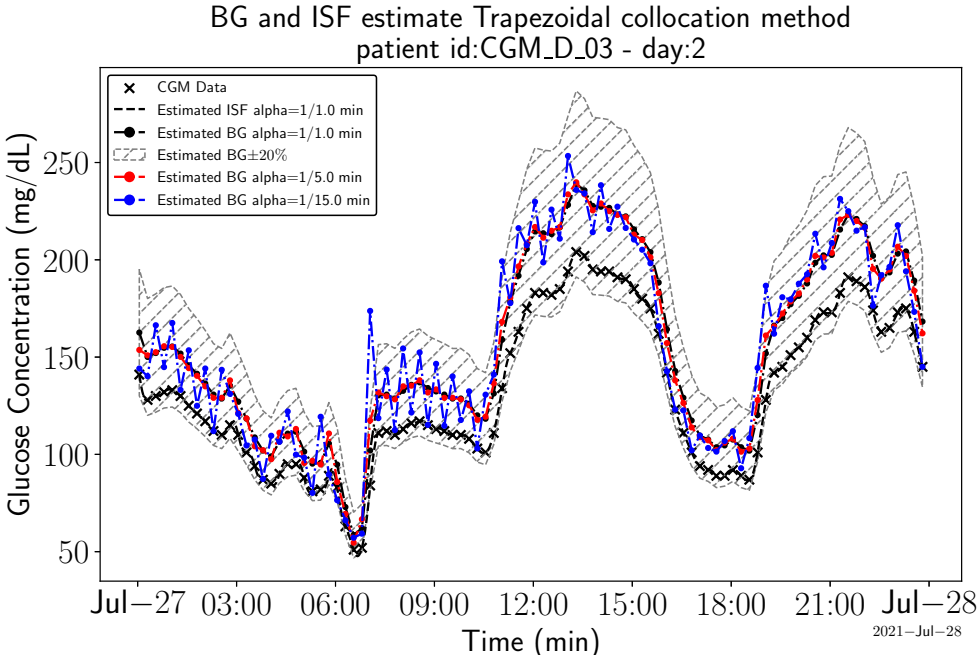


(a) The estimates here are obtained using the trapezoidal collocation method

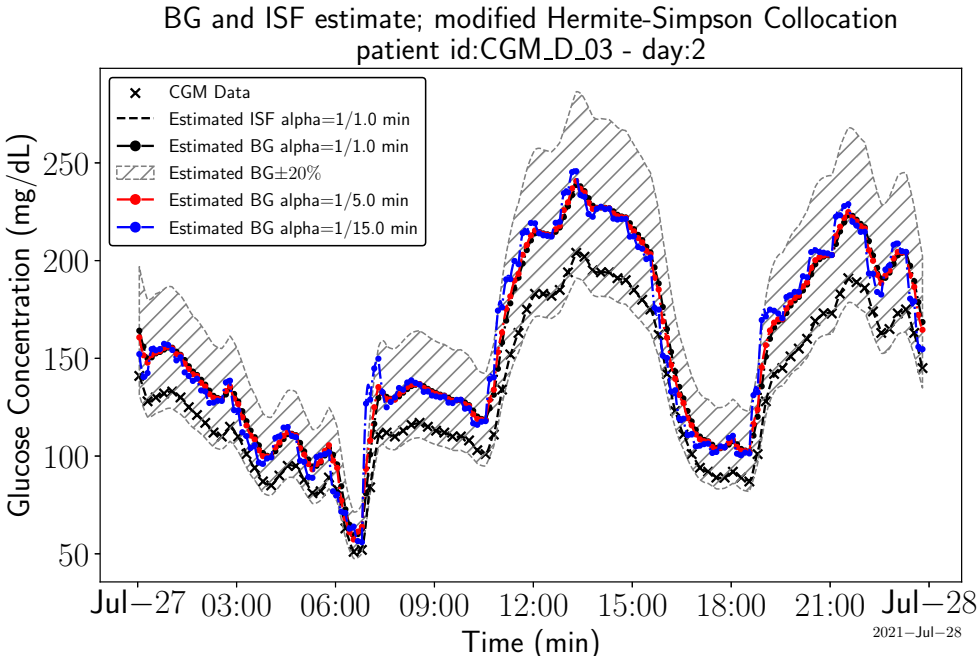


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.1: Estimated BG and ISF for `CGM D 02` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

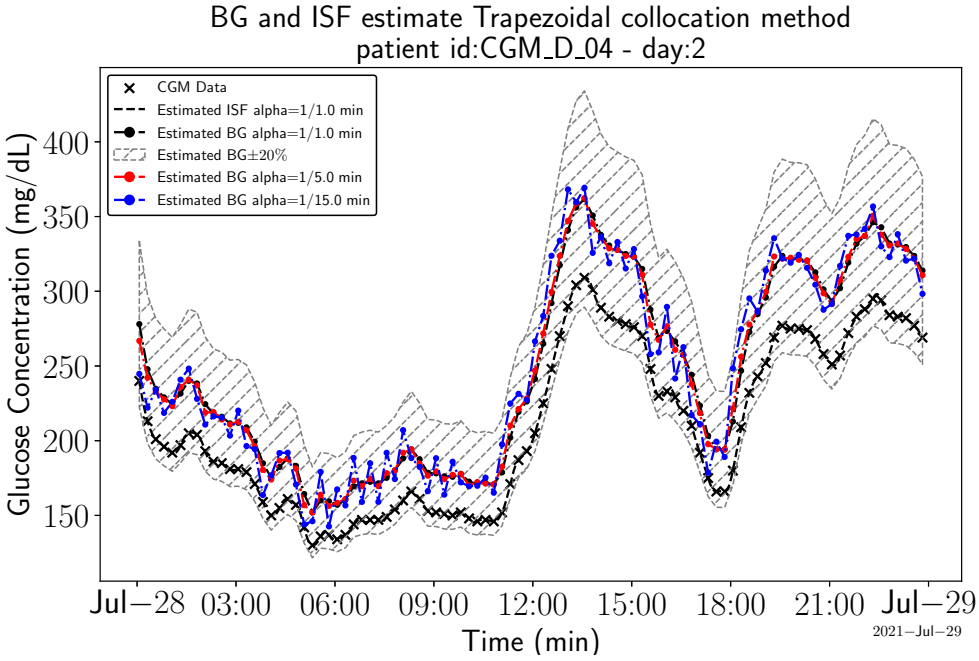


(a) The estimates here are obtained using the trapezoidal collocation method

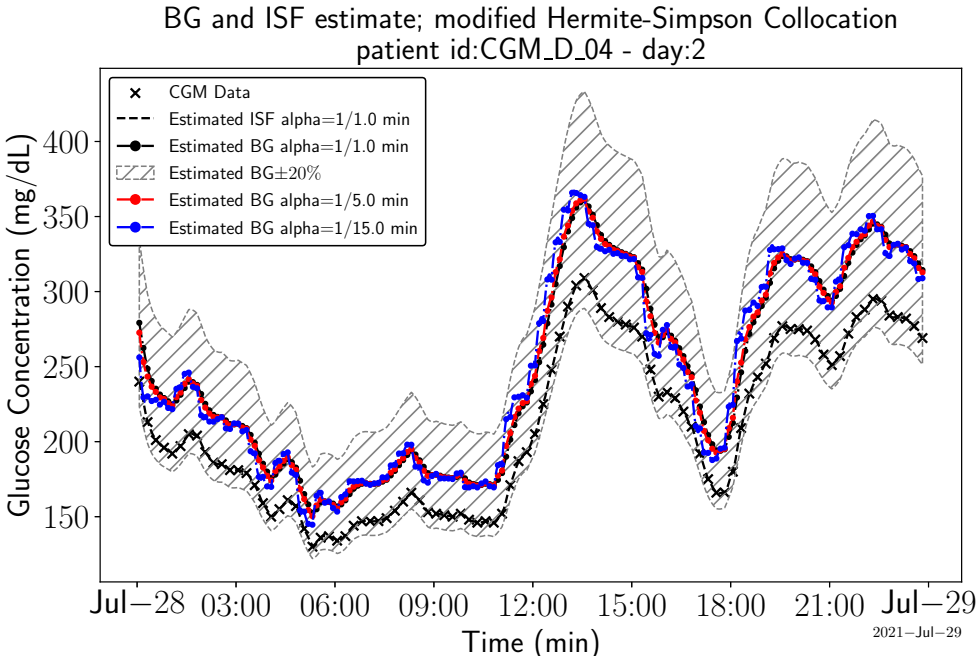


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.2: Estimated BG and ISF for `CGM D 03` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

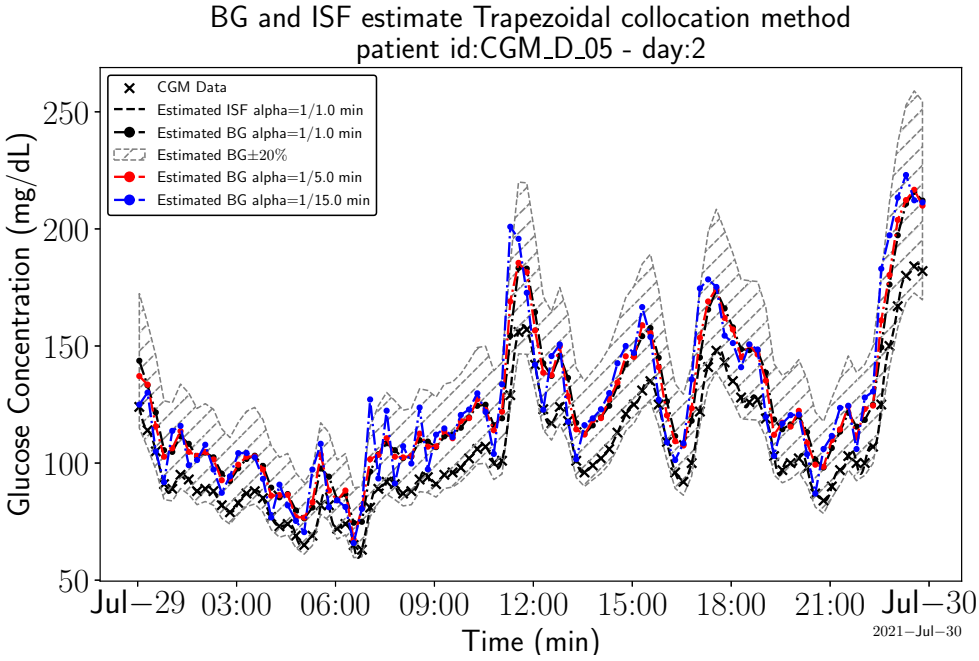


(a) The estimates here are obtained using the trapezoidal collocation method

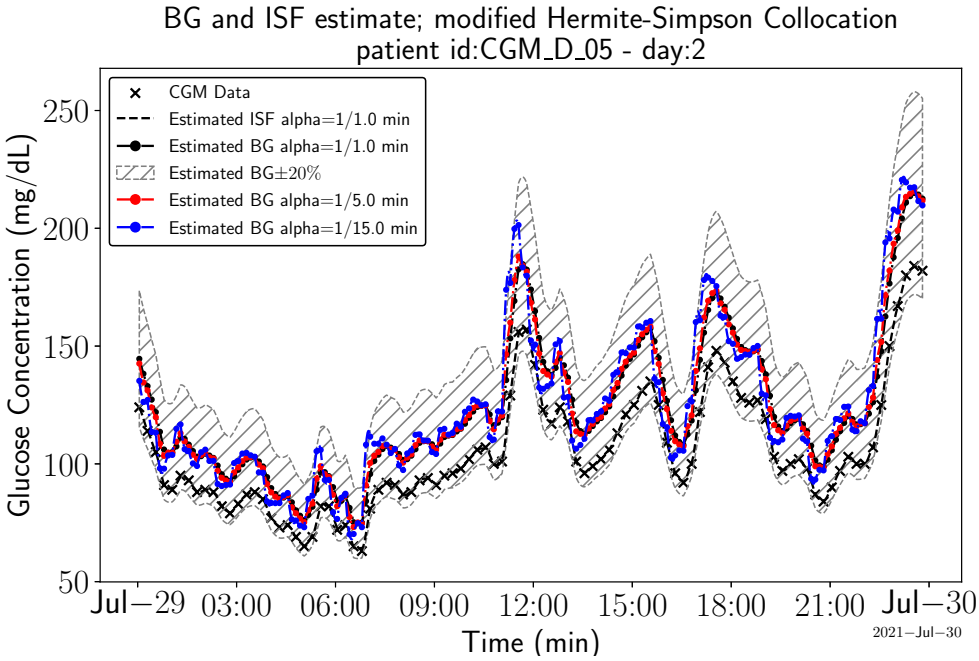


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.3: Estimated BG and ISF for `CGM D 04` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

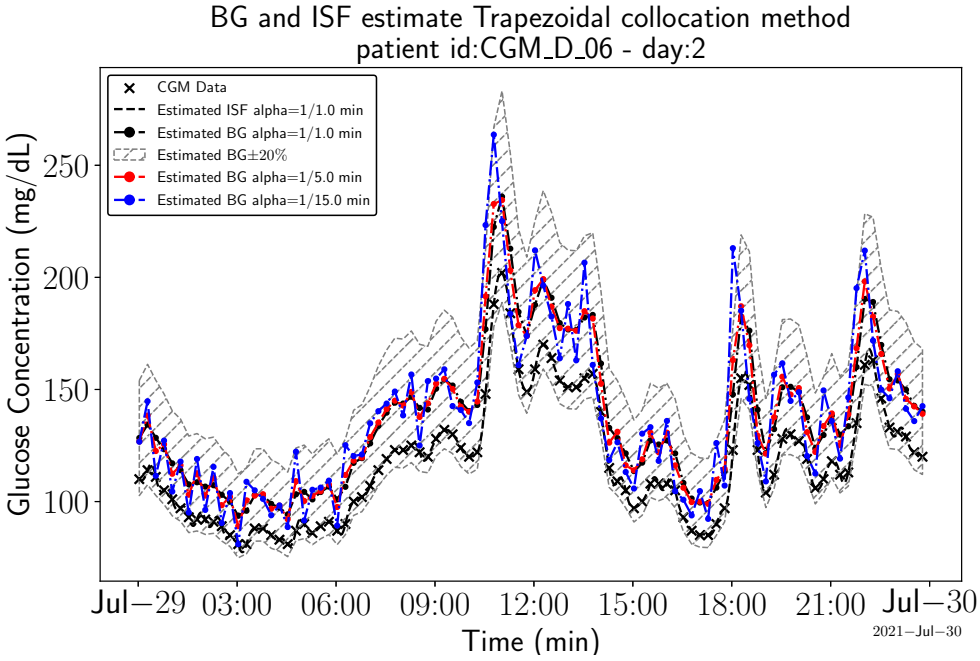


(a) The estimates here are obtained using the trapezoidal collocation method

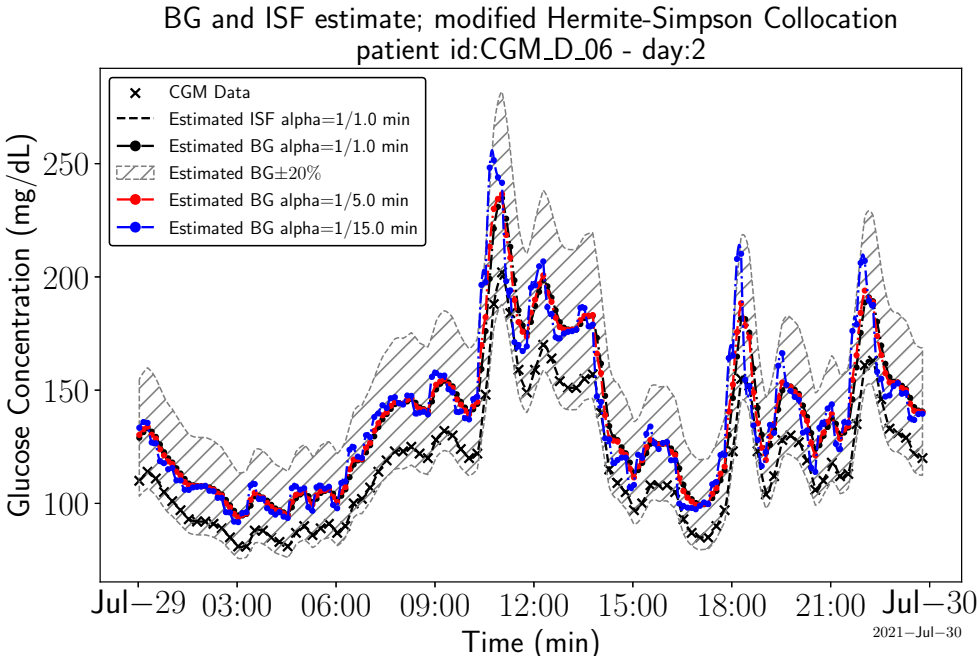


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.4: Estimated BG and ISF for `CGM D 05` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

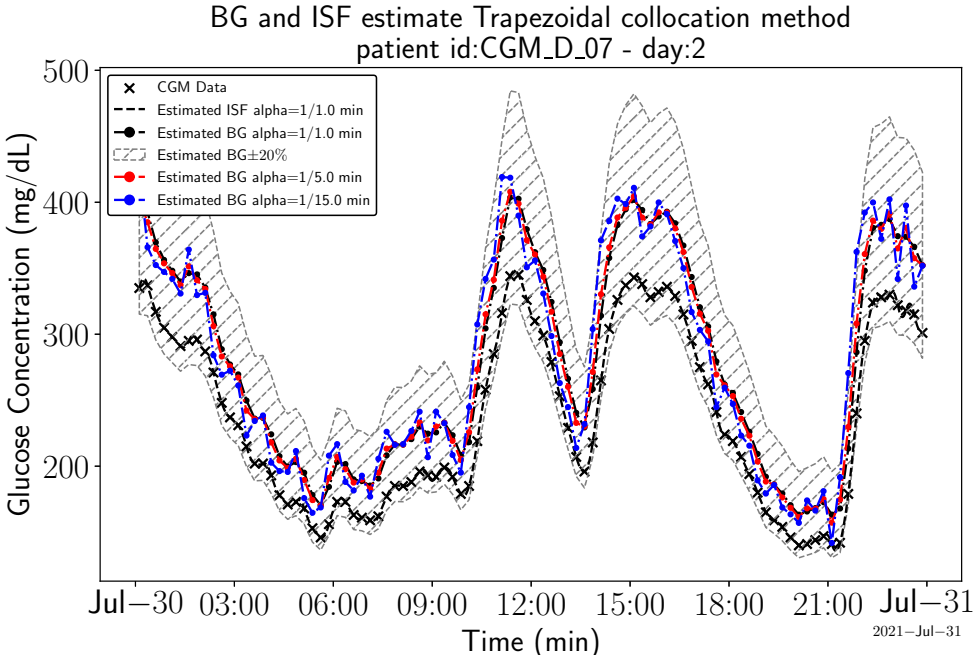


(a) The estimates here are obtained using the trapezoidal collocation method

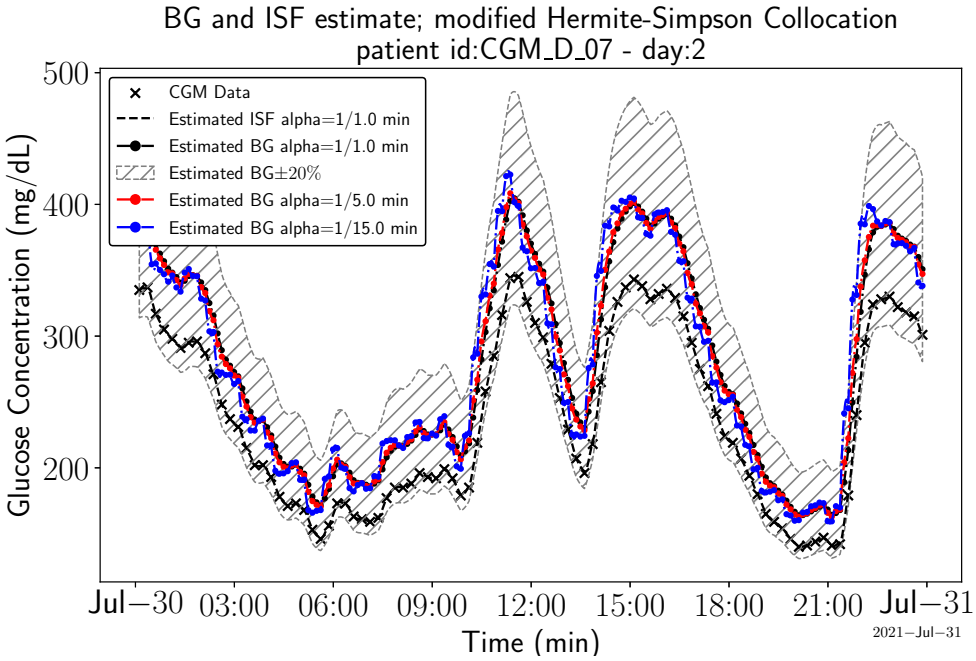


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.5: Estimated BG and ISF for `CGM D 06` for τ_0 = 1 min, 1/5min and, 1/15min

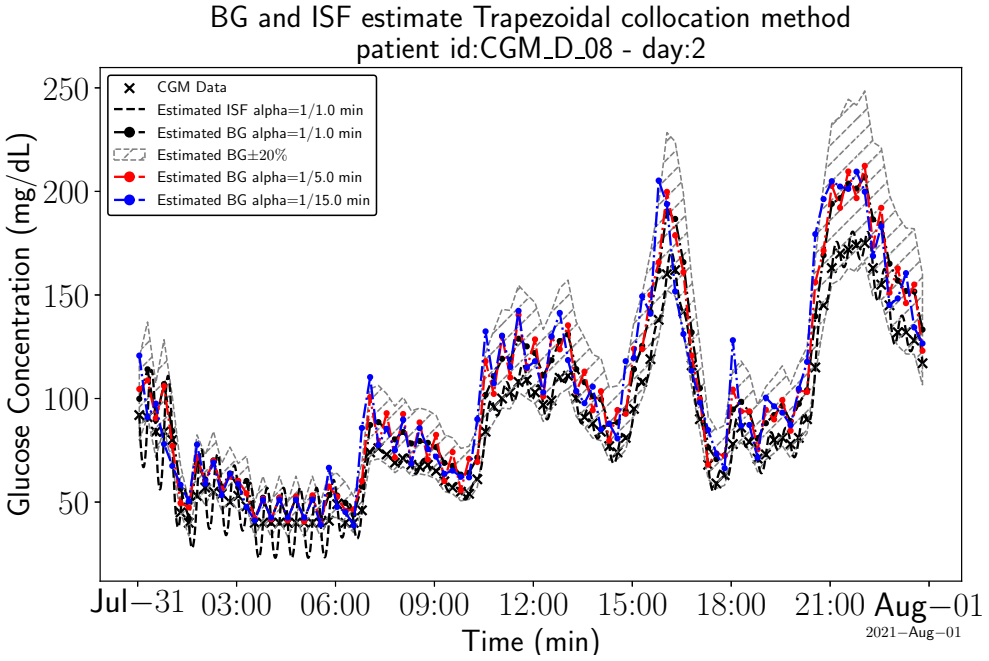


(a) The estimates here are obtained using the trapezoidal collocation method

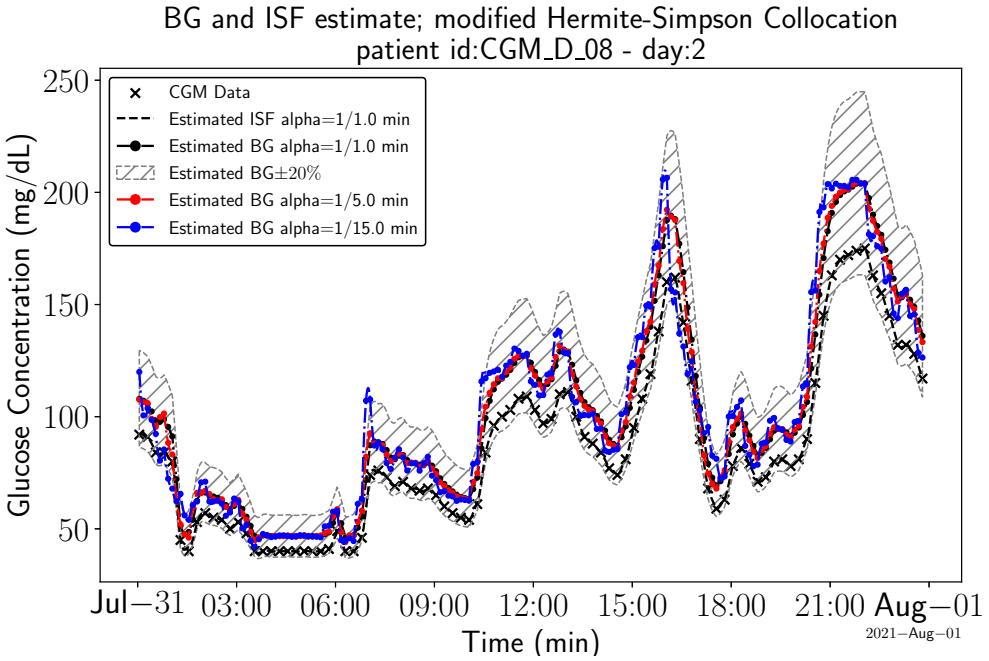


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.6: Estimated BG and ISF for `CGM D 07` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

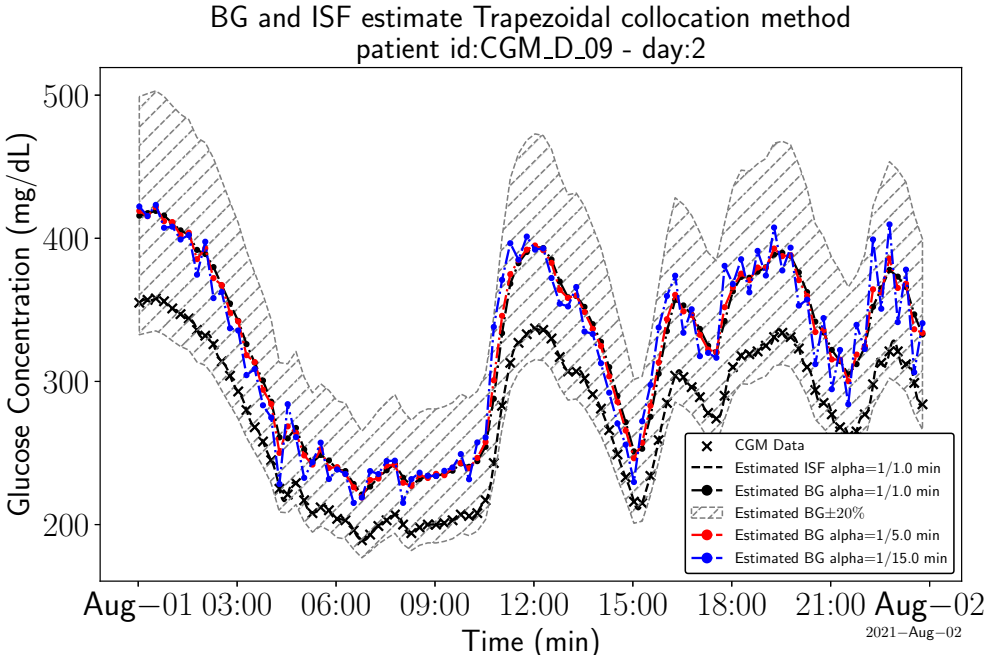


(a) The estimates here are obtained using the trapezoidal collocation method

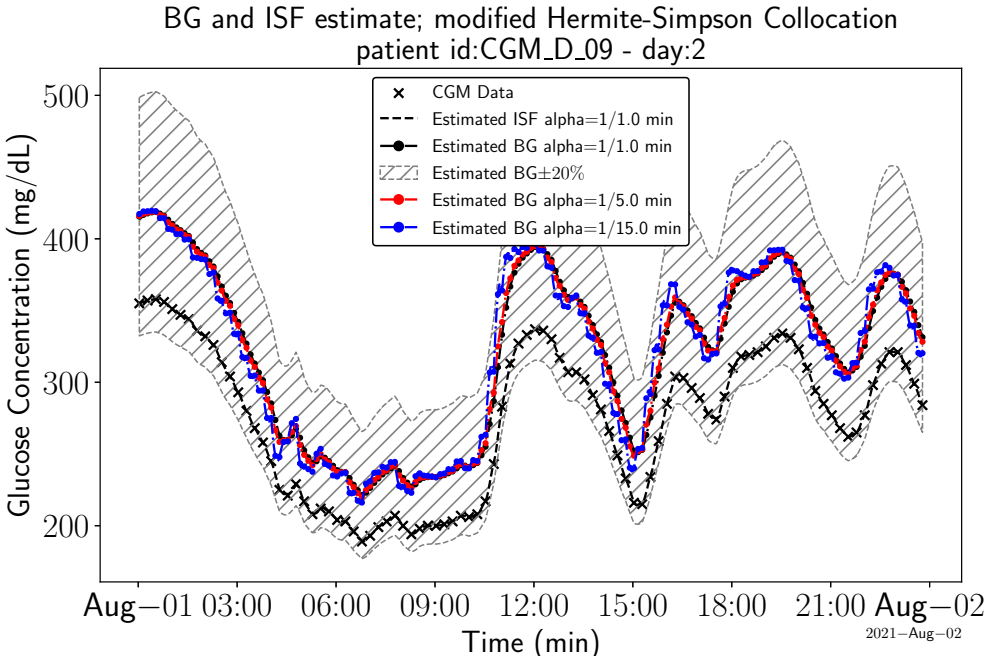


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.7: Estimated BG and ISF for `CGM D 08` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

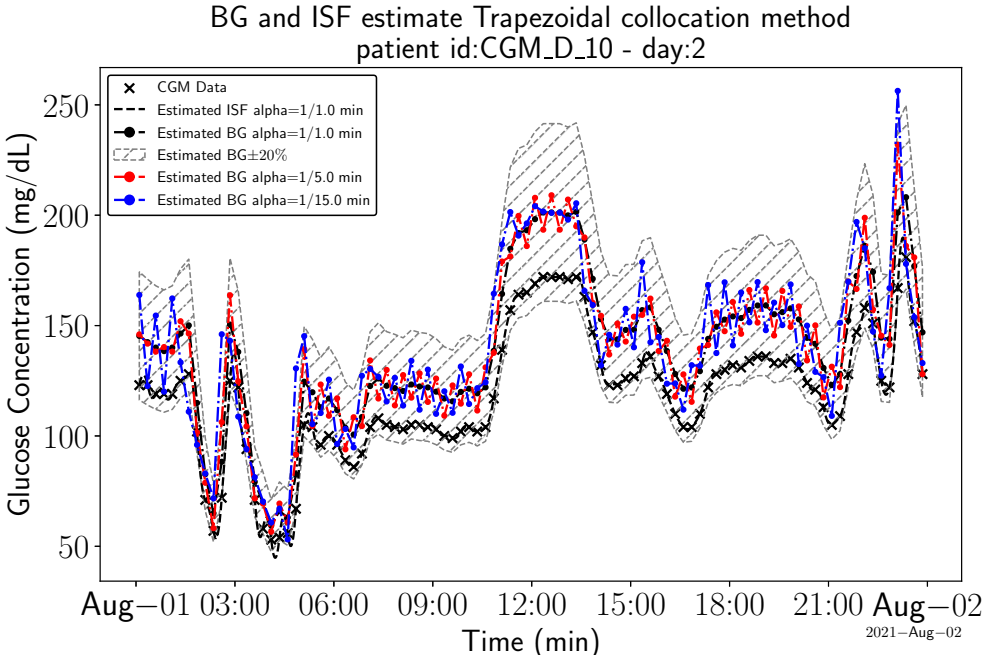


(a) The estimates here are obtained using the trapezoidal collocation method

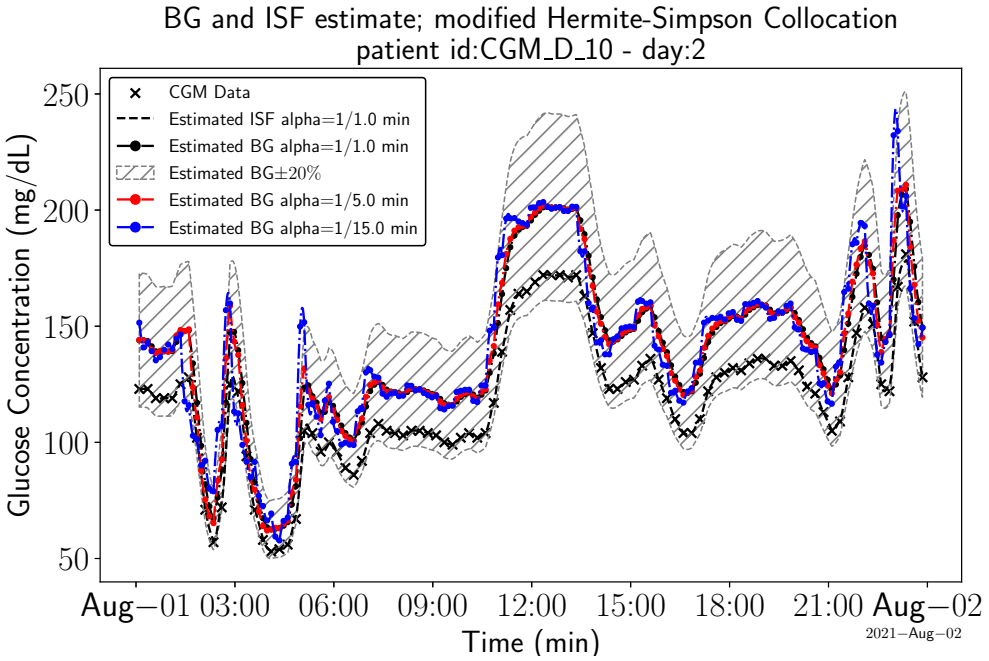


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.8: Estimated BG and ISF for `CGM D 09` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

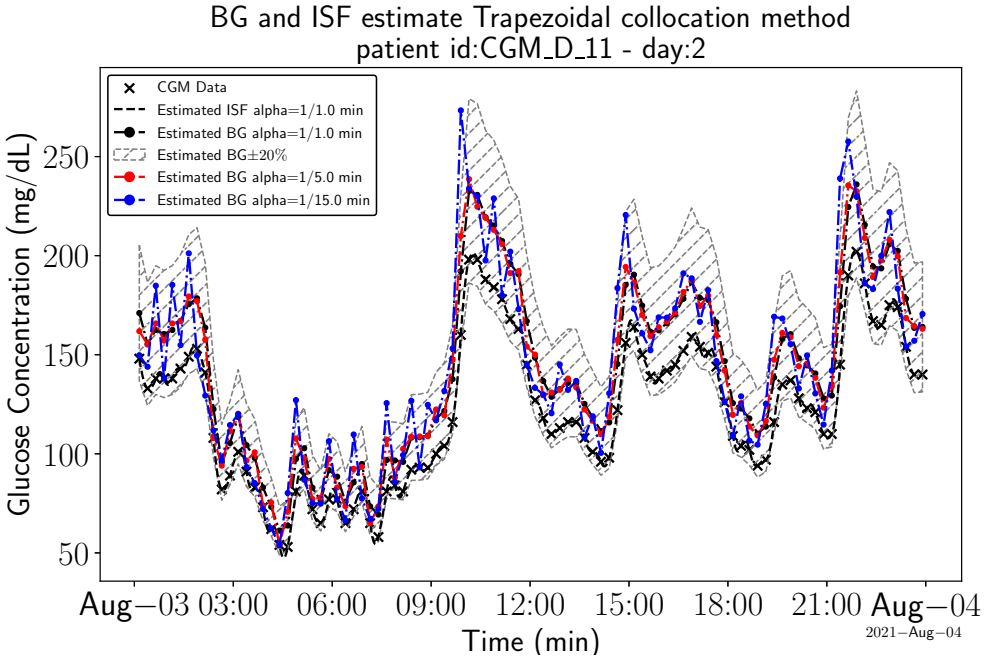


(a) The estimates here are obtained using the trapezoidal collocation method

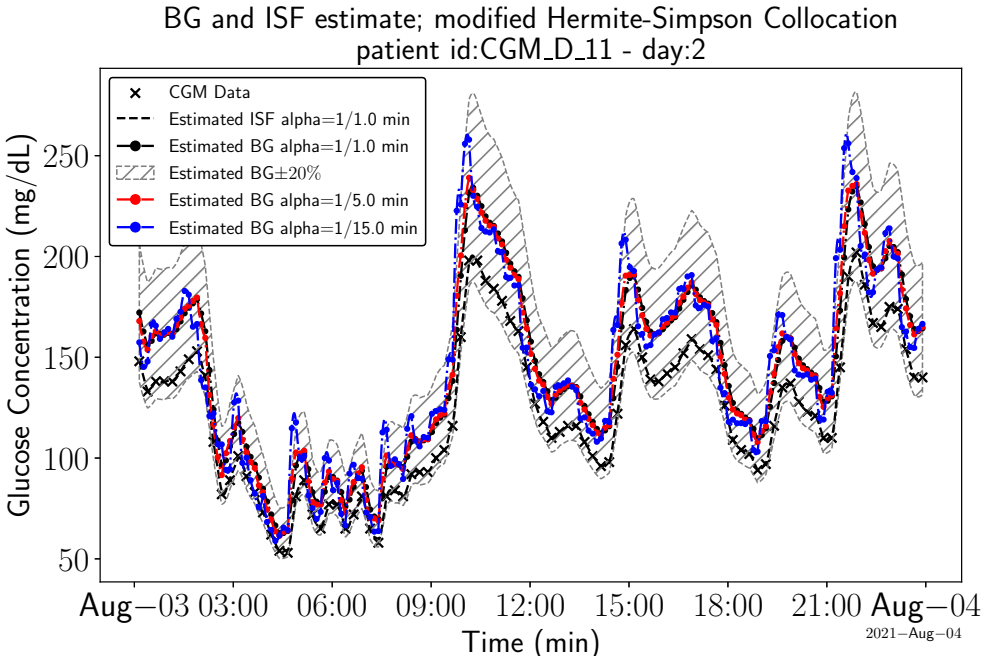


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.9: Estimated BG and ISF for `CGM D 10` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

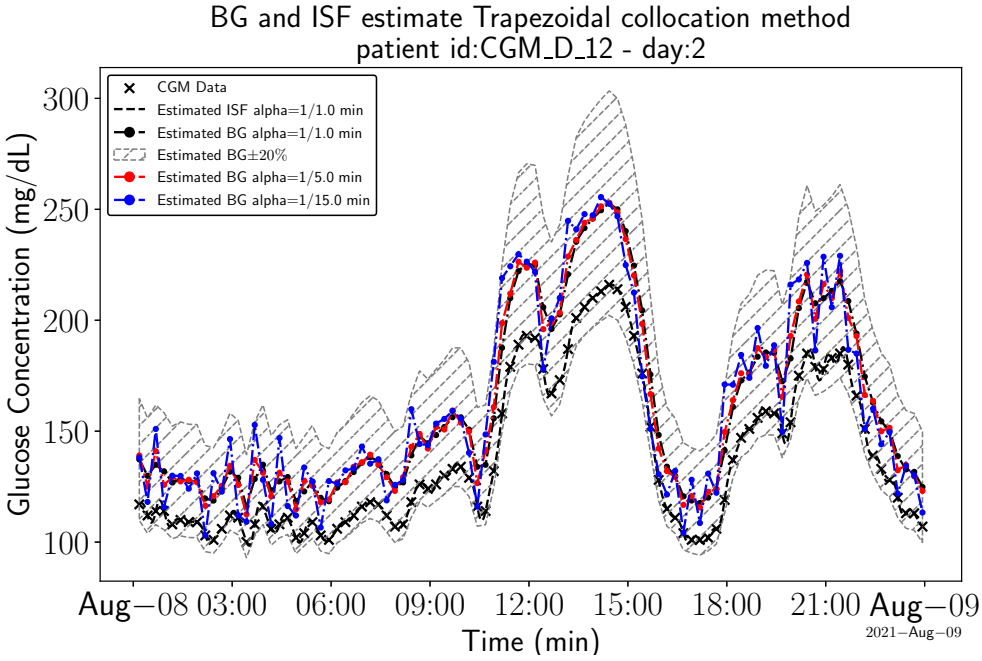


(a) The estimates here are obtained using the trapezoidal collocation method

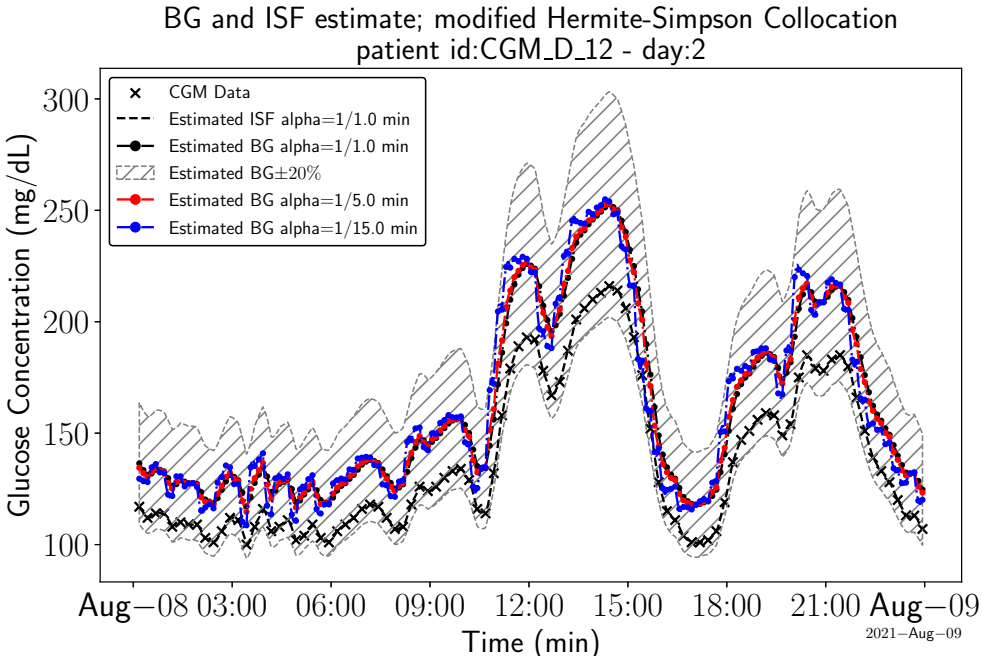


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.10: Estimated BG and ISF for `CGM D 11` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

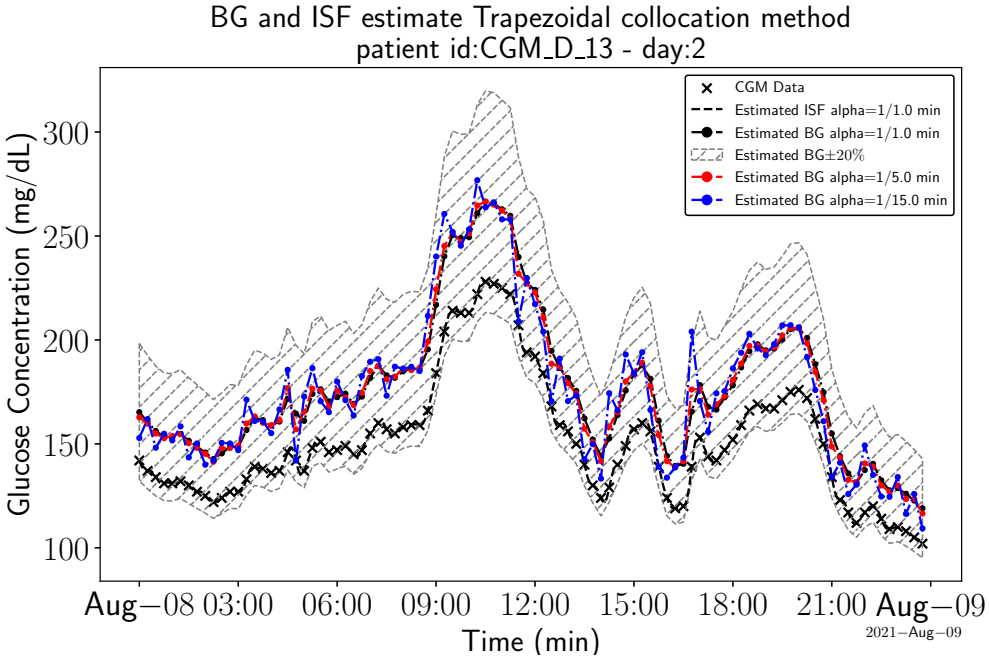


(a) The estimates here are obtained using the trapezoidal collocation method

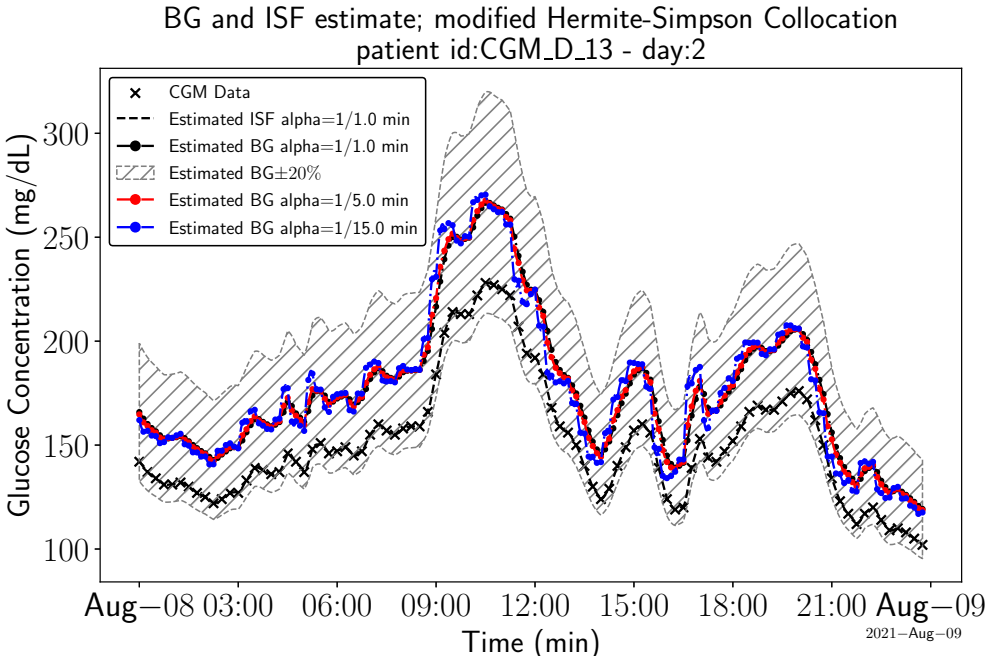


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.11: Estimated BG and ISF for `CGM D 12` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

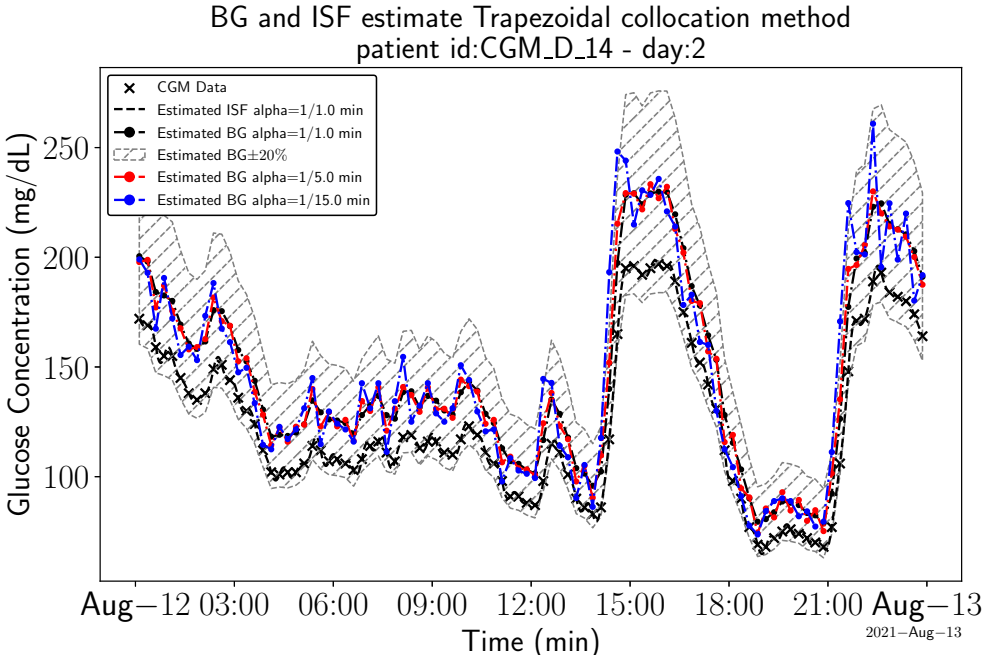


(a) The estimates here are obtained using the trapezoidal collocation method

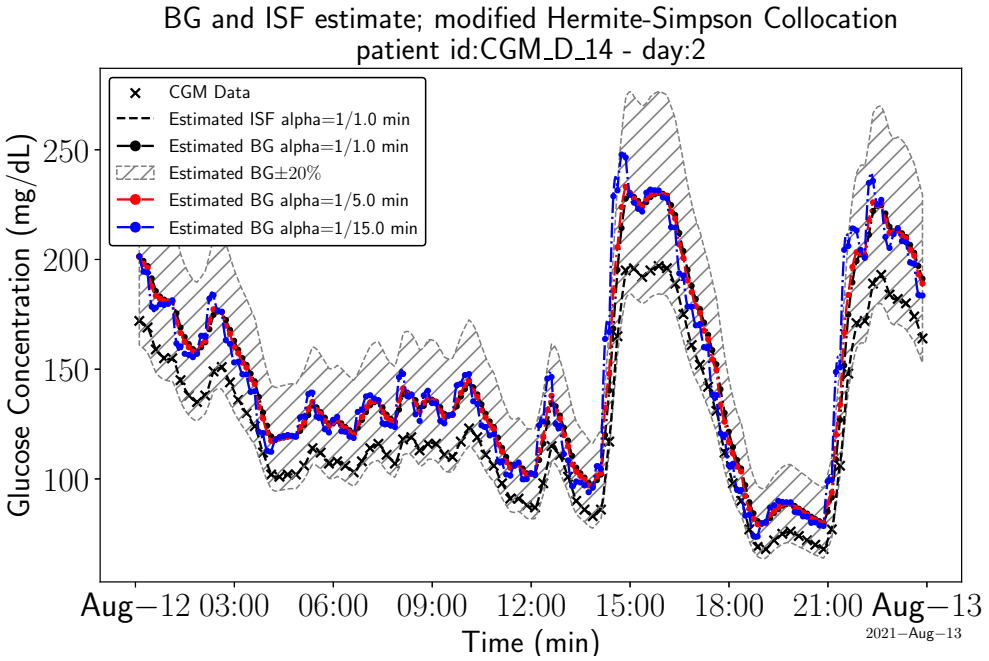


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.12: Estimated BG and ISF for `CGM D 13` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

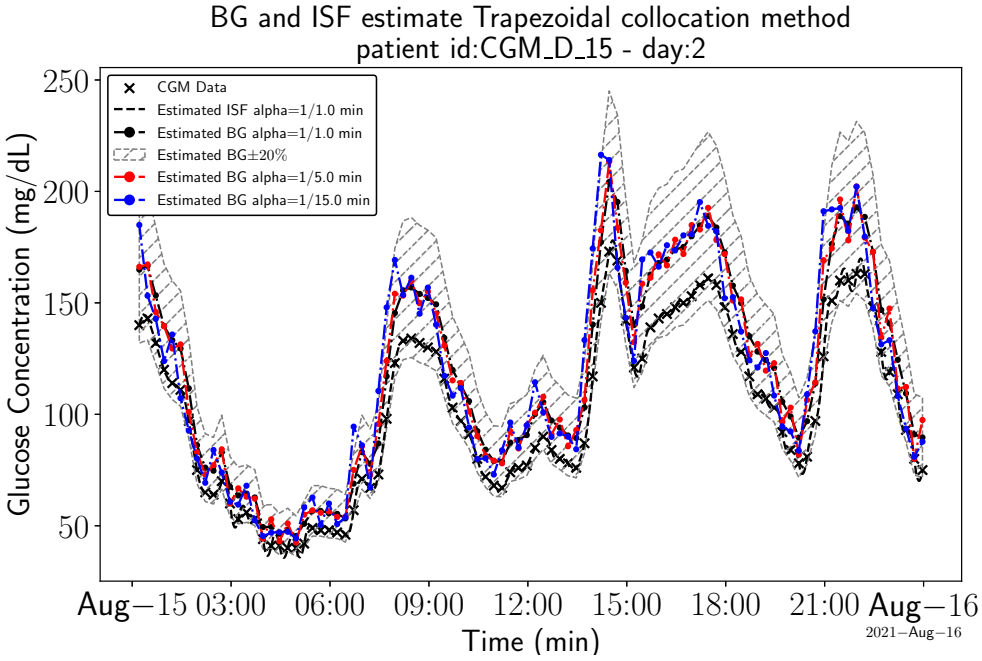


(a) The estimates here are obtained using the trapezoidal collocation method

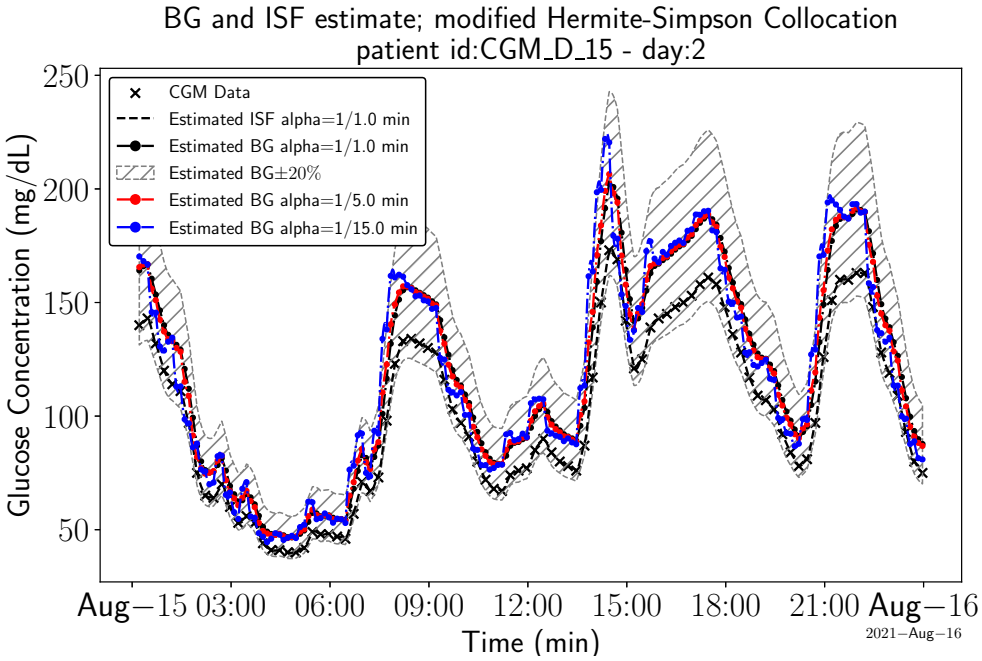


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.13: Estimated BG and ISF for `CGM D 14` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

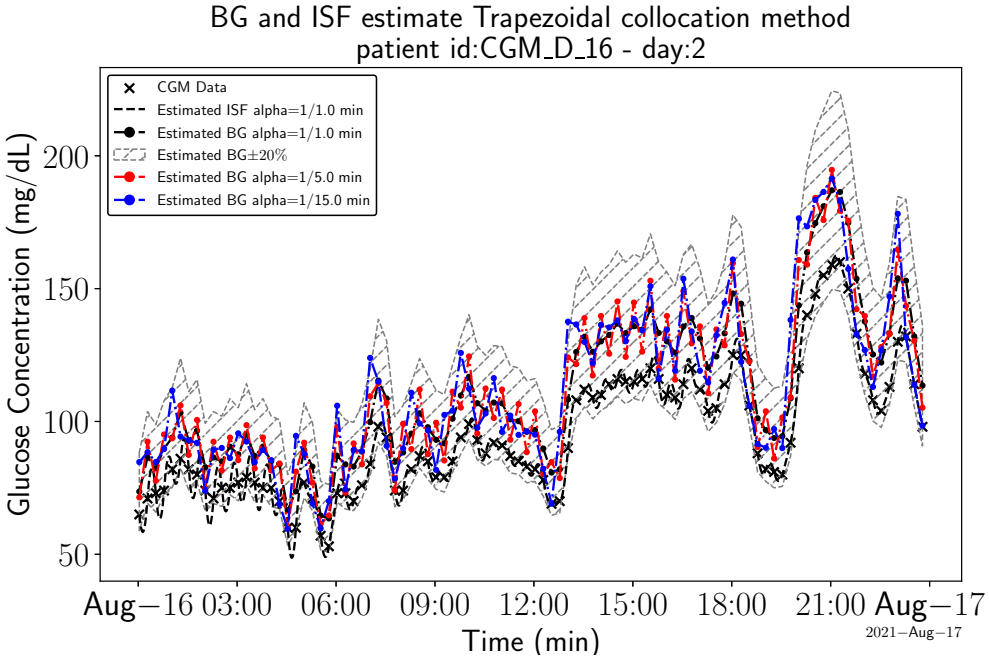


(a) The estimates here are obtained using the trapezoidal collocation method

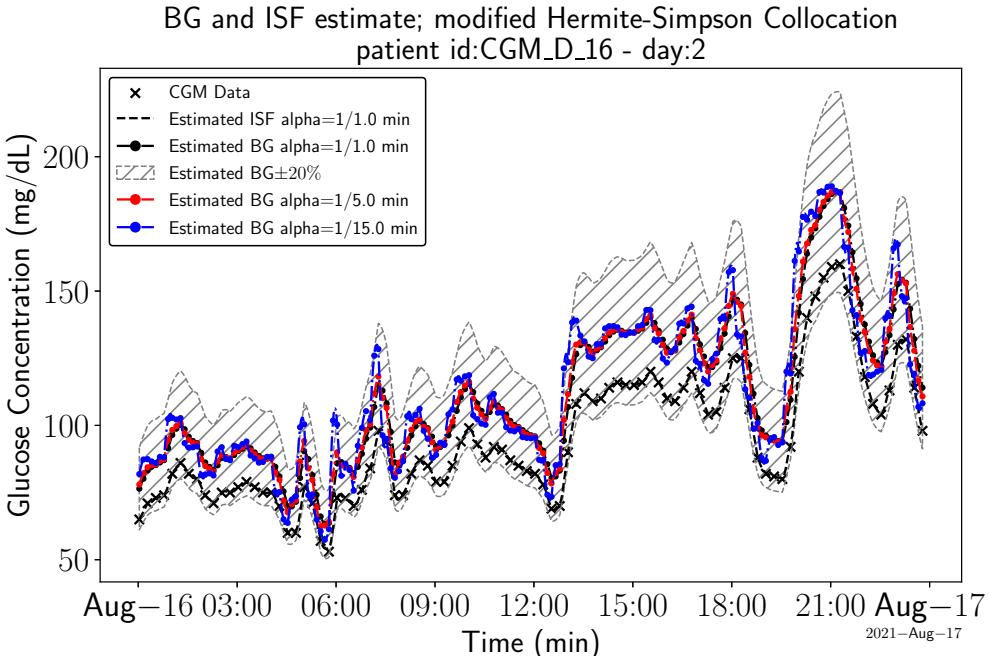


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.14: Estimated BG and ISF for `CGM D 15` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

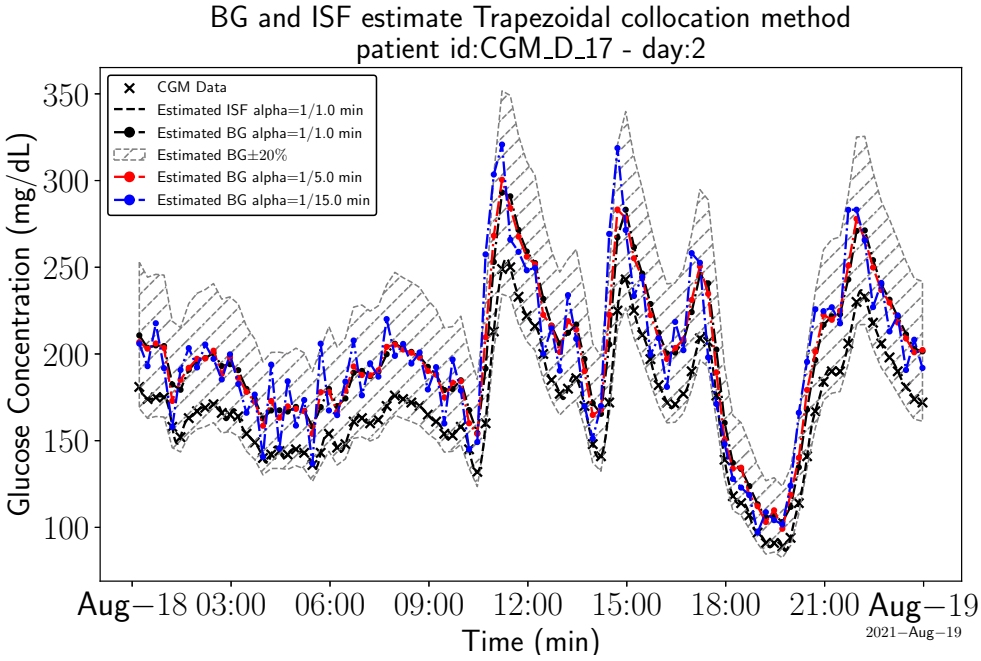


(a) The estimates here are obtained using the trapezoidal collocation method

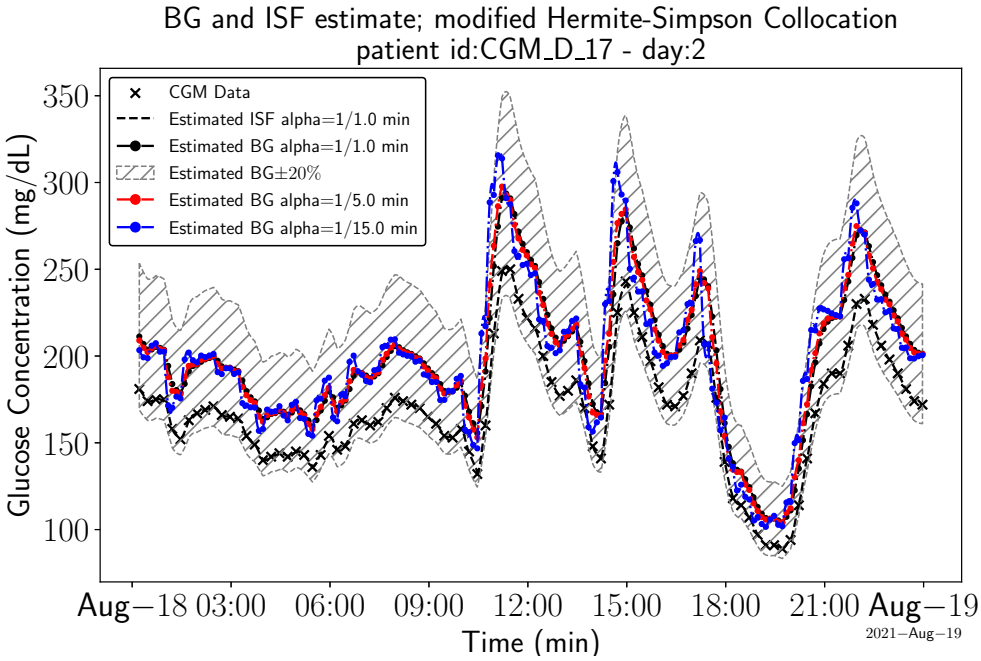


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.15: Estimated BG and ISF for `CGM D 16` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

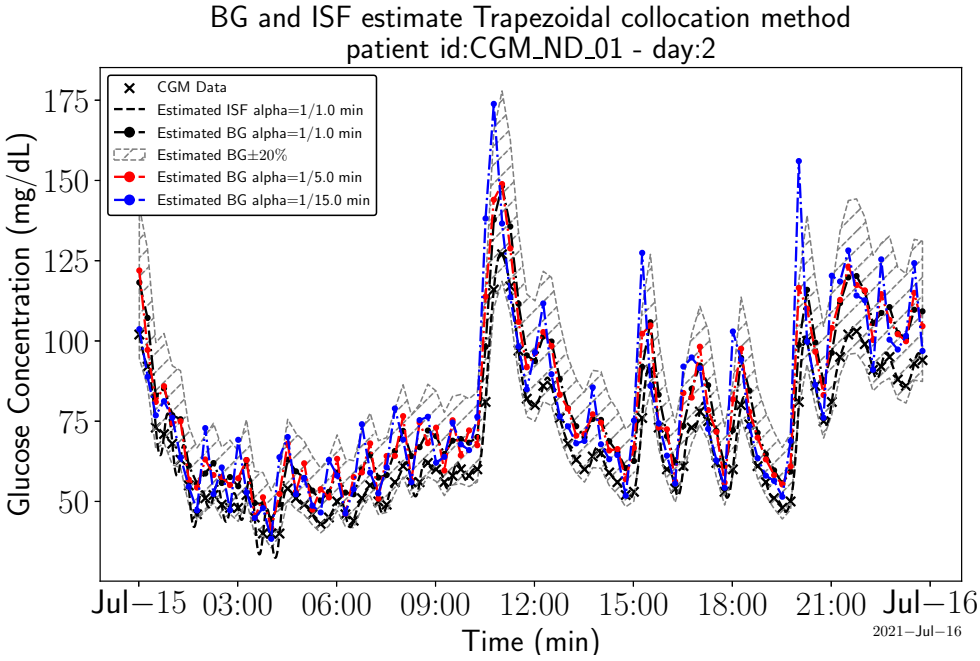


(a) The estimates here are obtained using the trapezoidal collocation method

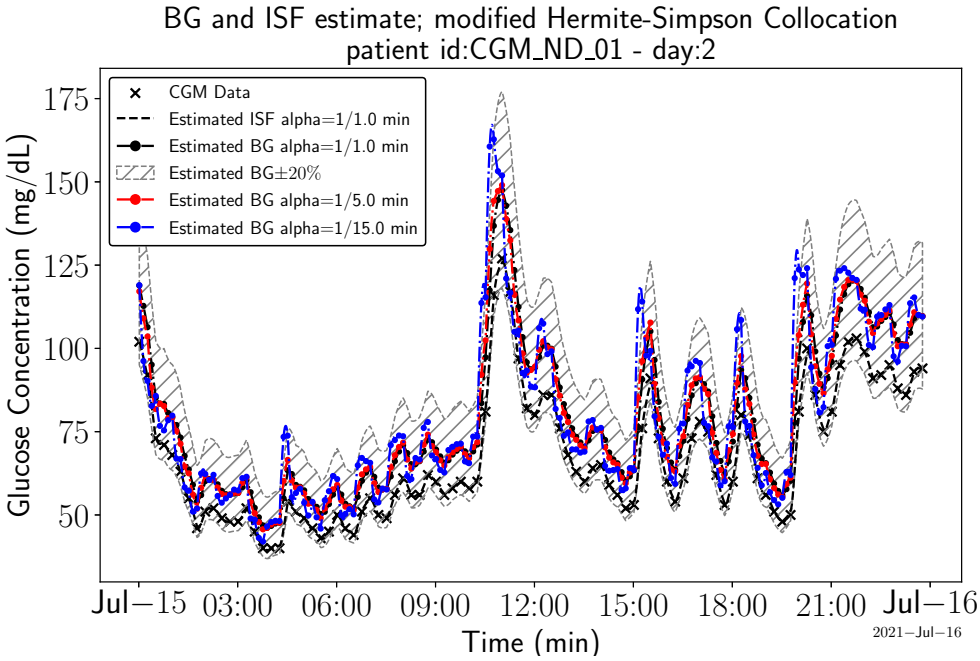


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.16: Estimated BG and ISF for `CGM D 17` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

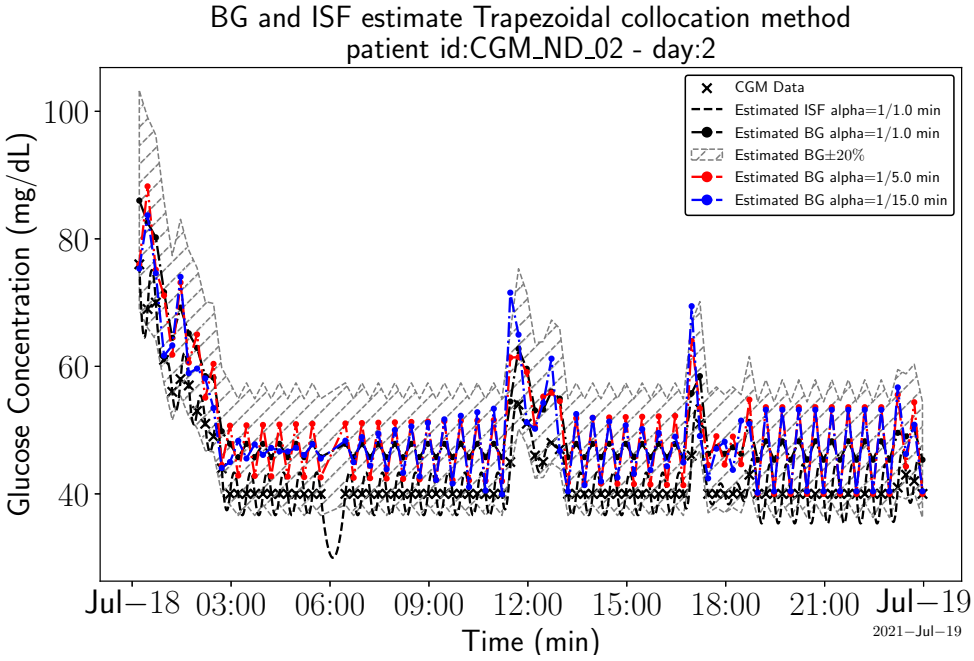


(a) The estimates here are obtained using the trapezoidal collocation method

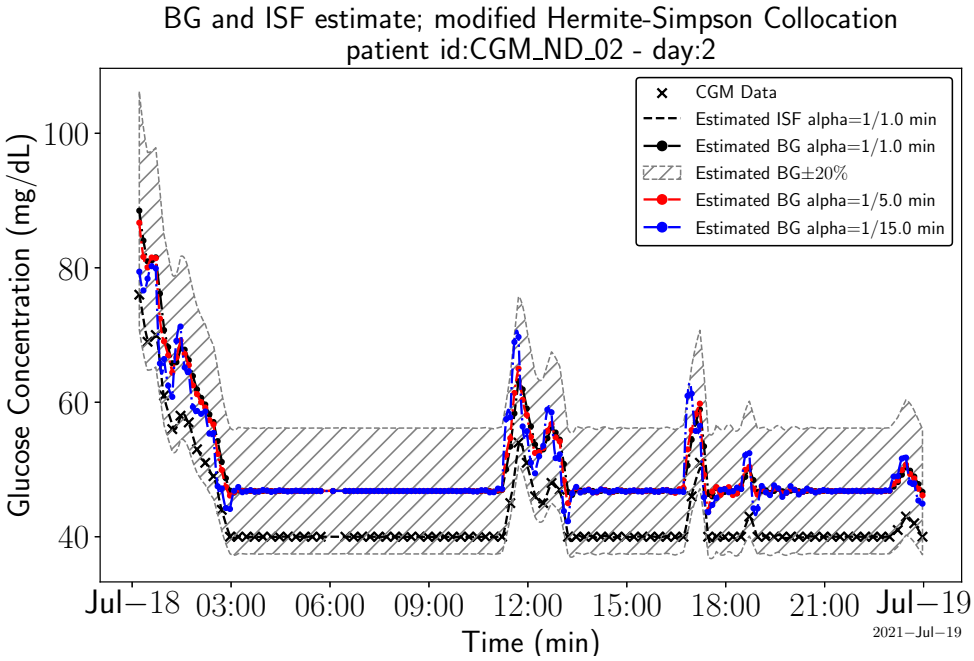


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.17: Estimated BG and ISF for `CGM ND 01` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

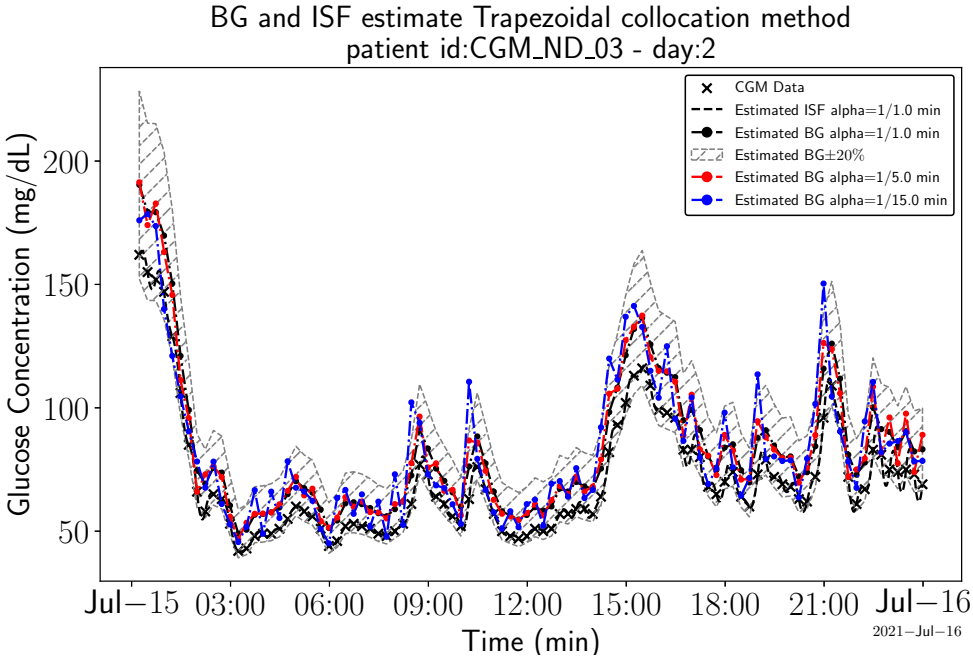


(a) The estimates here are obtained using the trapezoidal collocation method

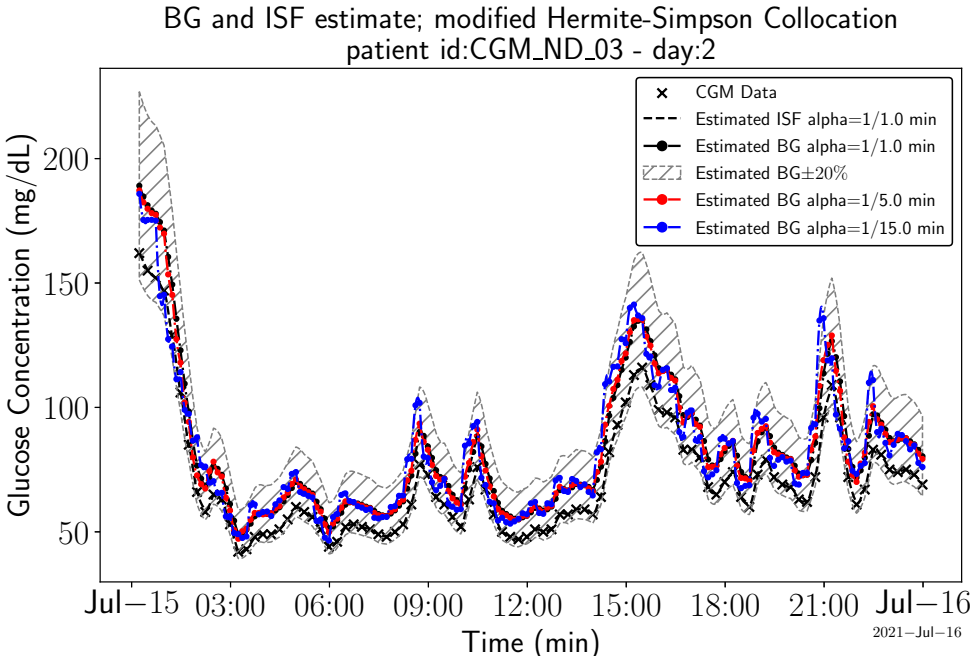


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.18: Estimated BG and ISF for `CGM ND 02` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

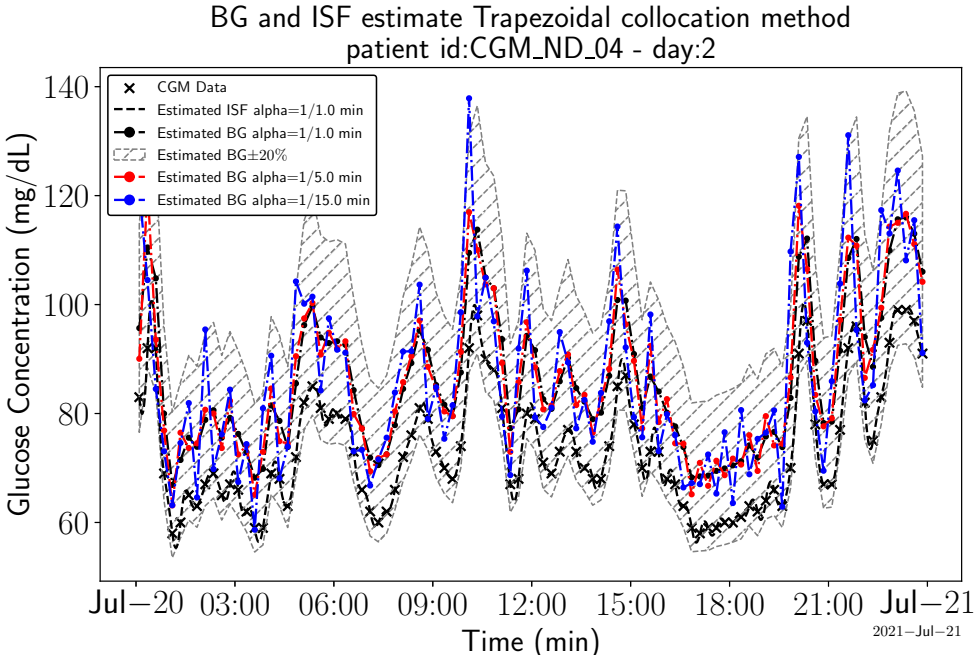


(a) The estimates here are obtained using the trapezoidal collocation method

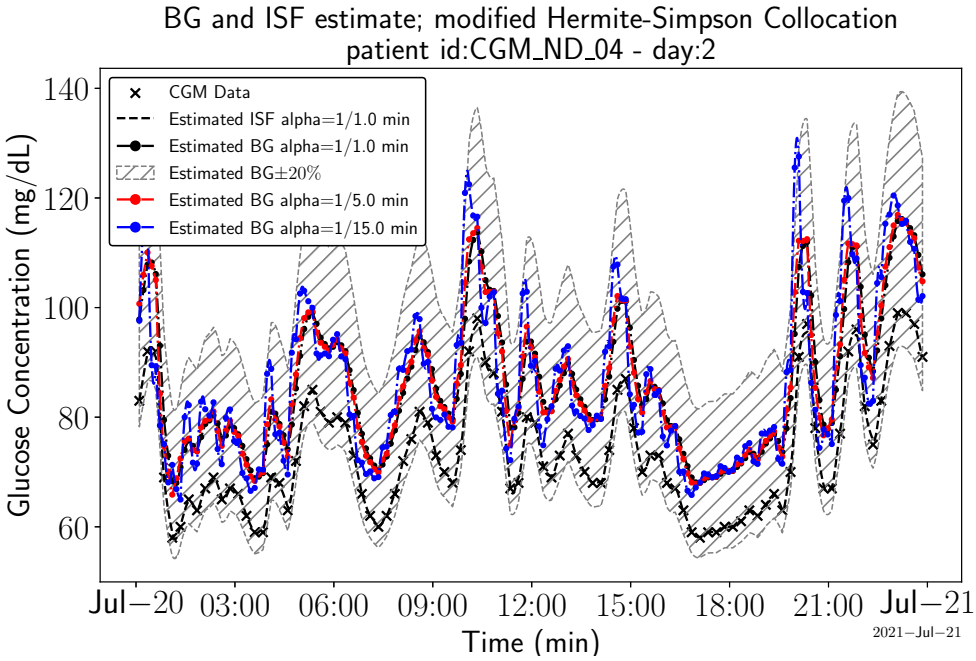


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.19: Estimated BG and ISF for `CGM ND 03` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

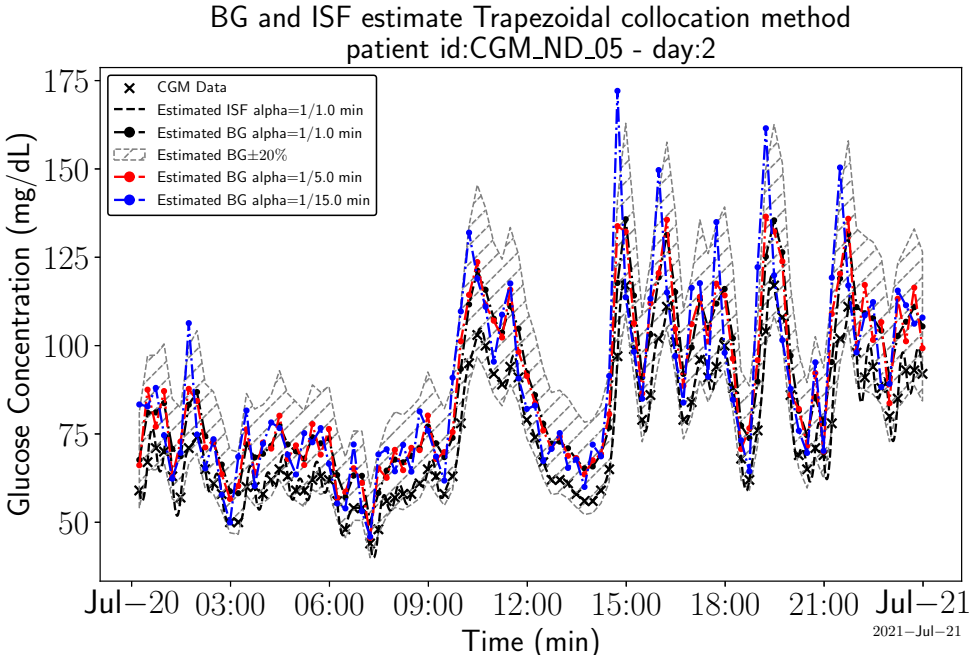


(a) The estimates here are obtained using the trapezoidal collocation method

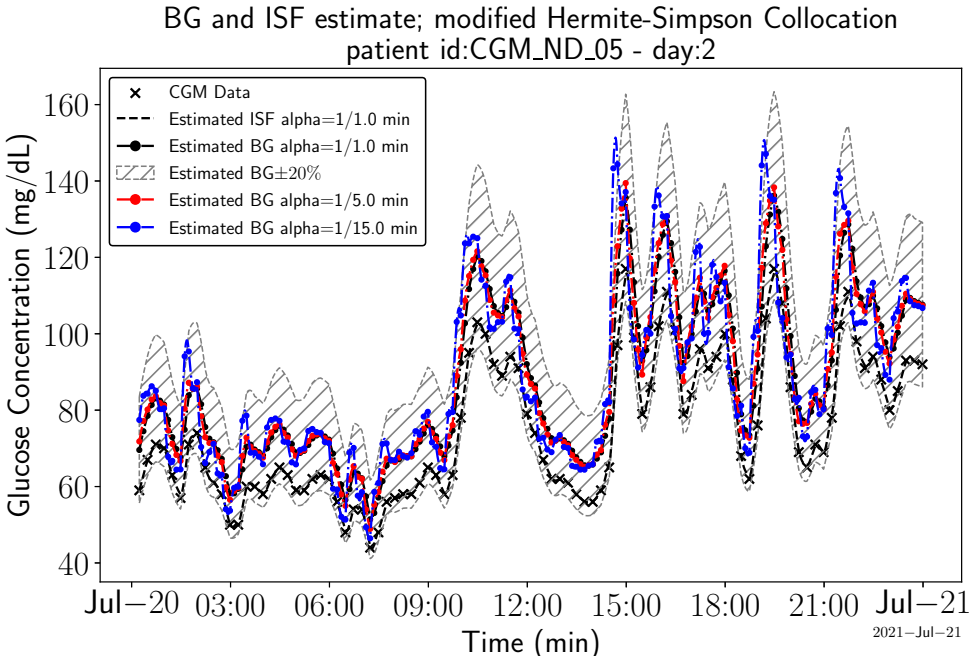


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.20: Estimated BG and ISF for `CGM ND 04` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

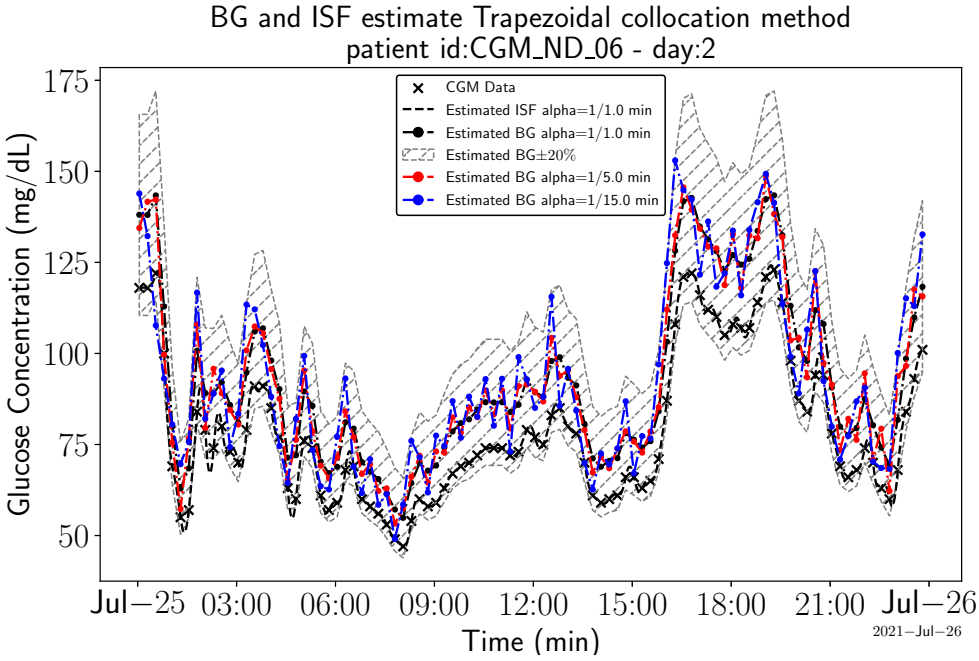


(a) The estimates here are obtained using the trapezoidal collocation method

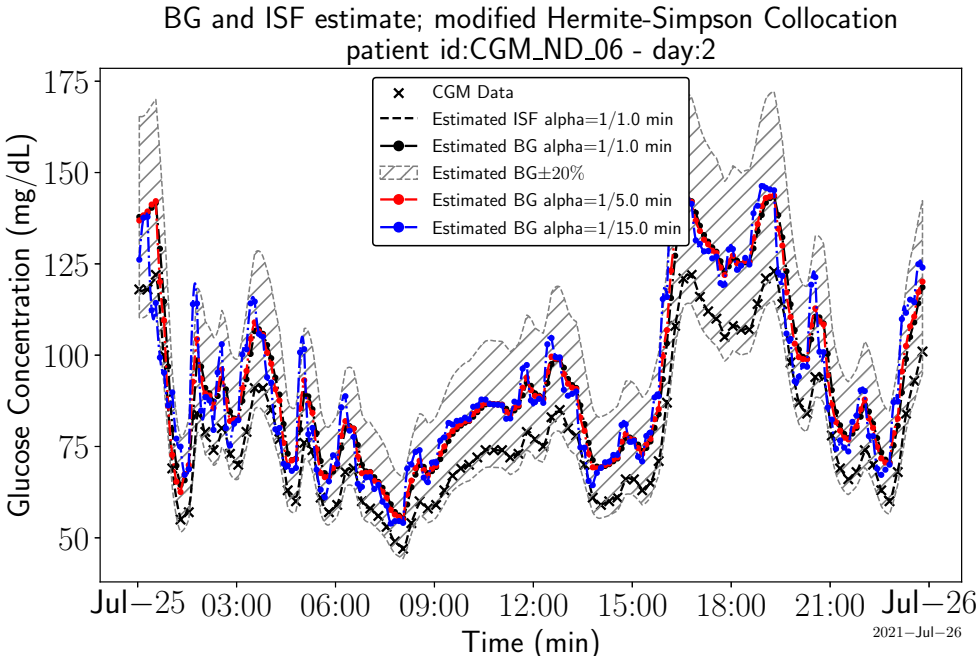


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.21: Estimated BG and ISF for `CGM ND 05` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

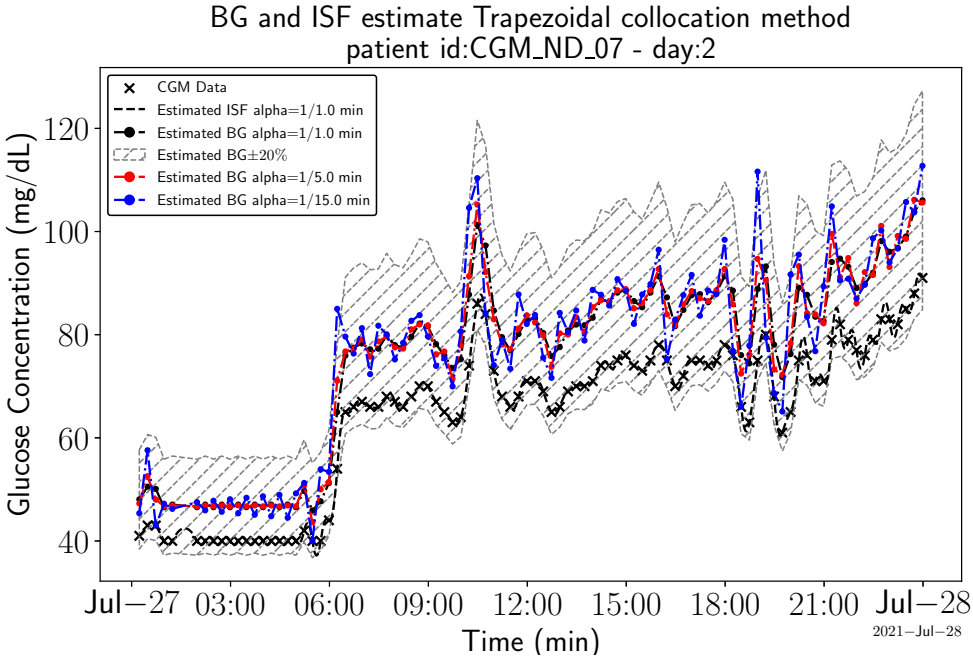


(a) The estimates here are obtained using the trapezoidal collocation method

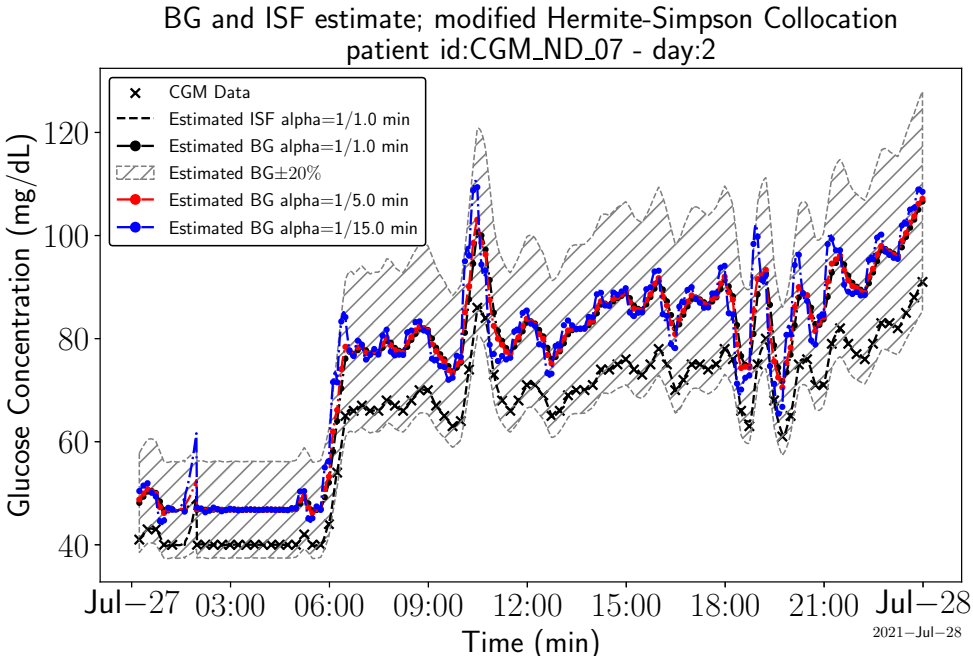


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.22: Estimated BG and ISF for `CGM ND 06` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

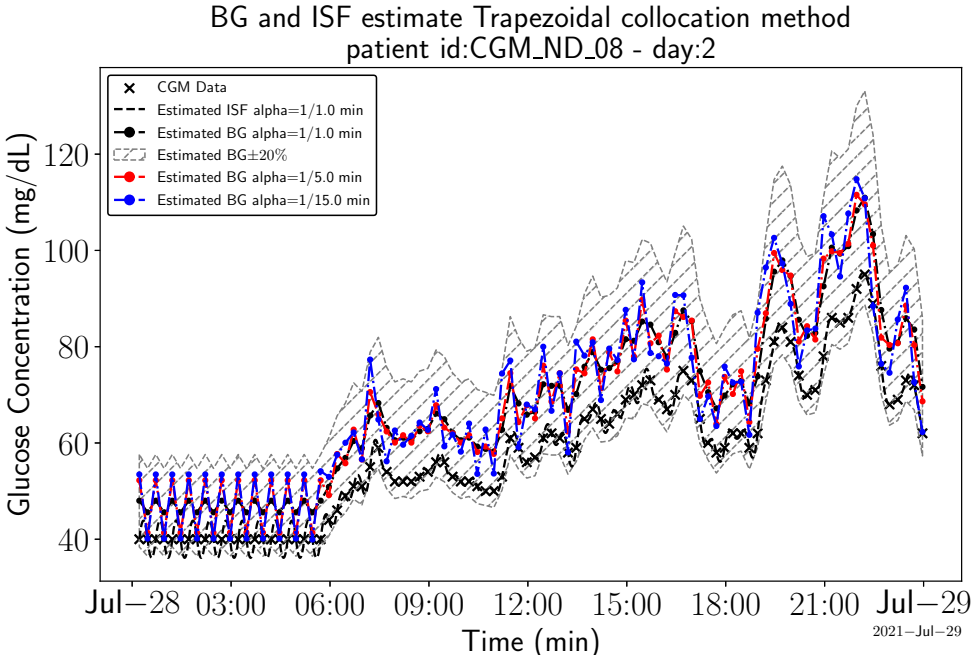


(a) The estimates here are obtained using the trapezoidal collocation method

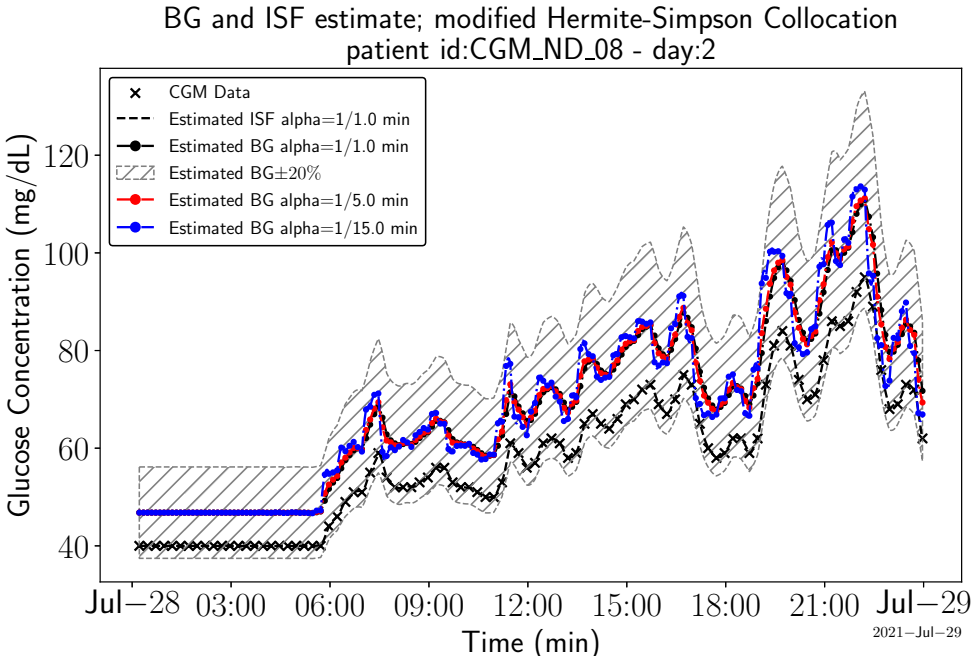


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.23: Estimated BG and ISF for `CGM ND 07` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

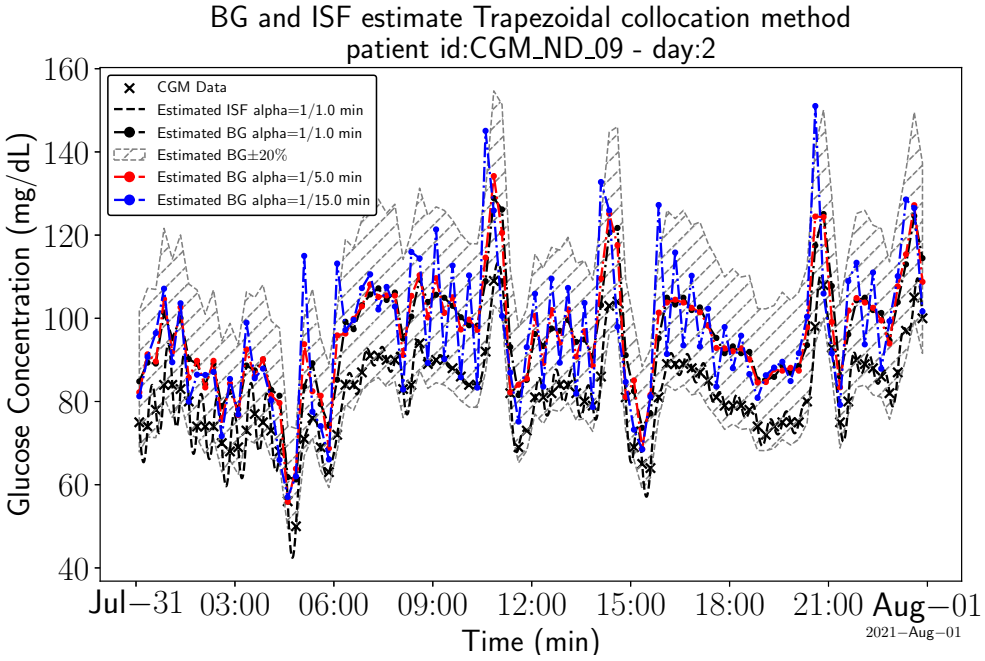


(a) The estimates here are obtained using the trapezoidal collocation method

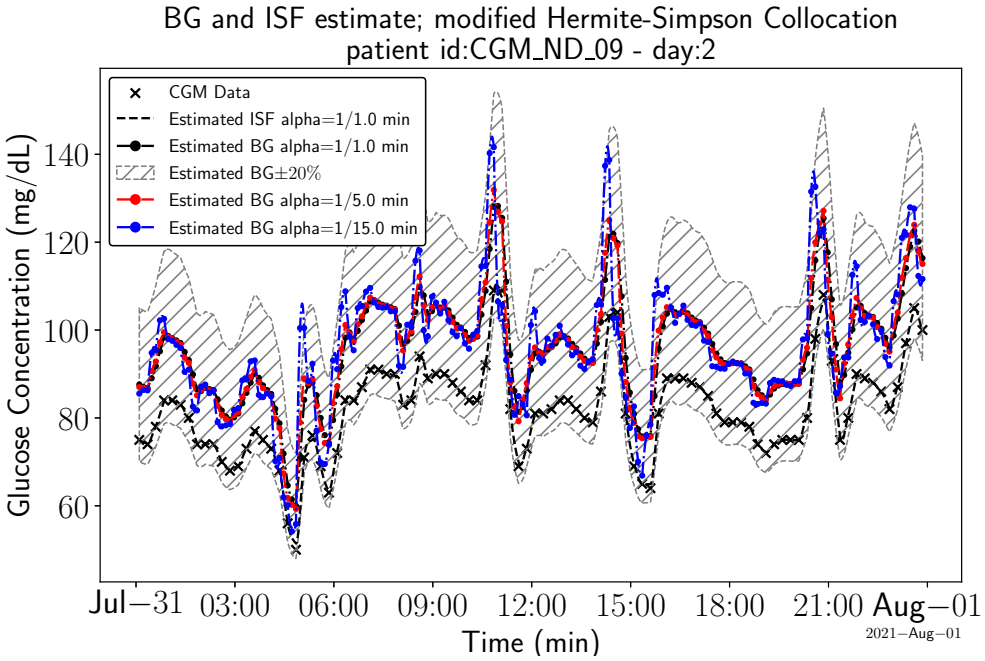


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.24: Estimated BG and ISF for `CGM ND 08` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

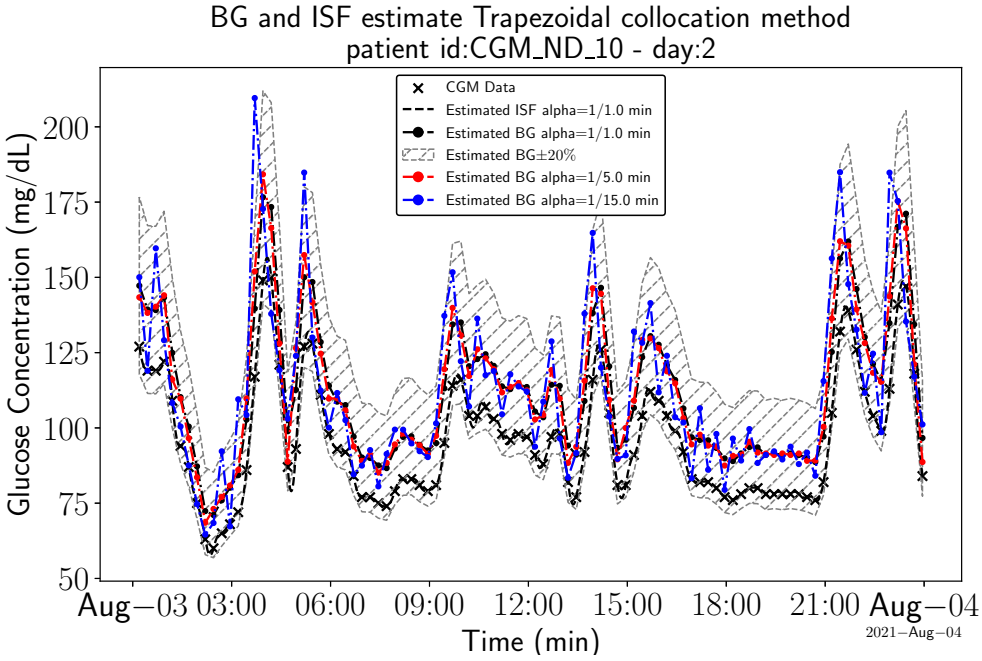


(a) The estimates here are obtained using the trapezoidal collocation method

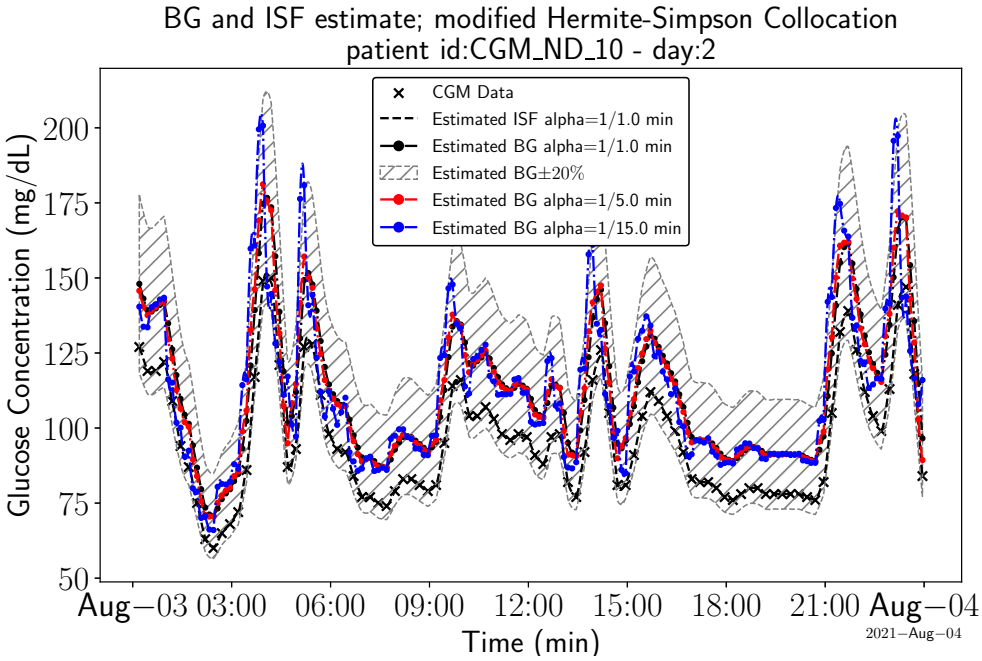


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.25: Estimated BG and ISF for `CGM ND 09` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

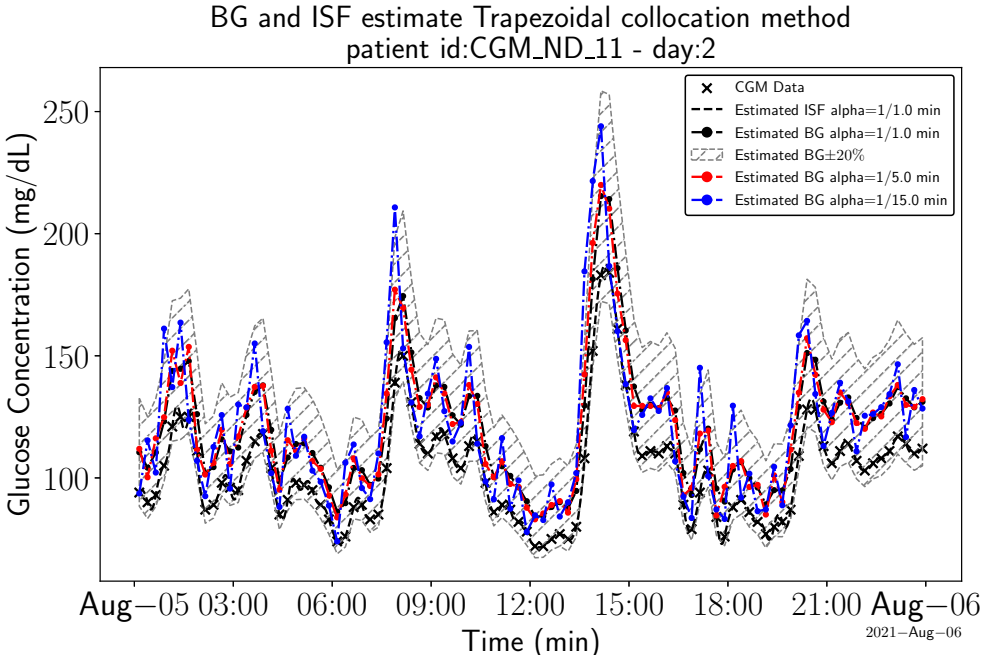


(a) The estimates here are obtained using the trapezoidal collocation method

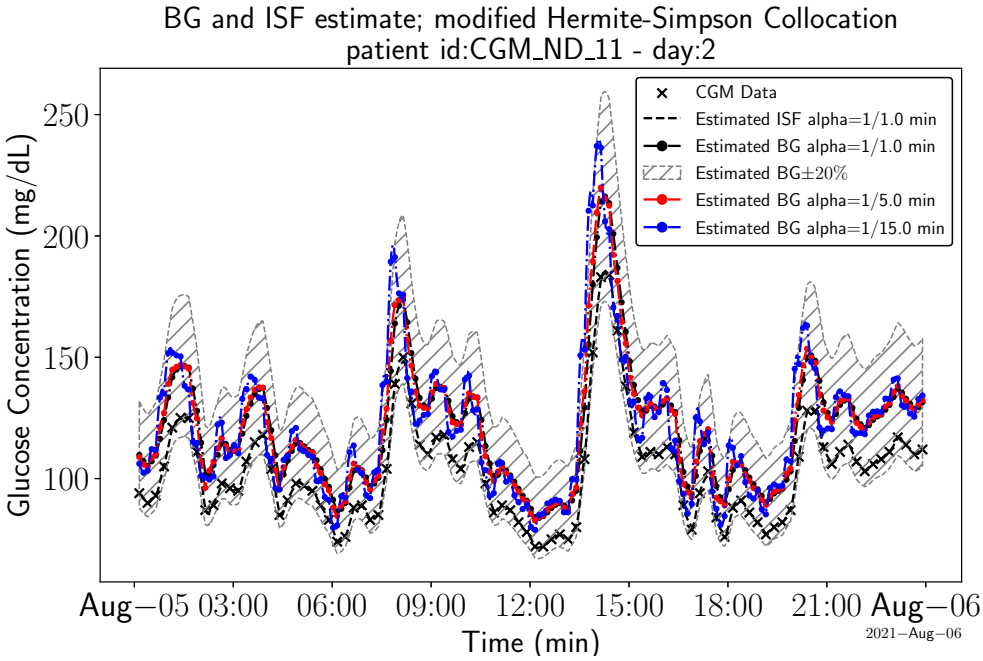


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.26: Estimated BG and ISF for `CGM ND 10` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

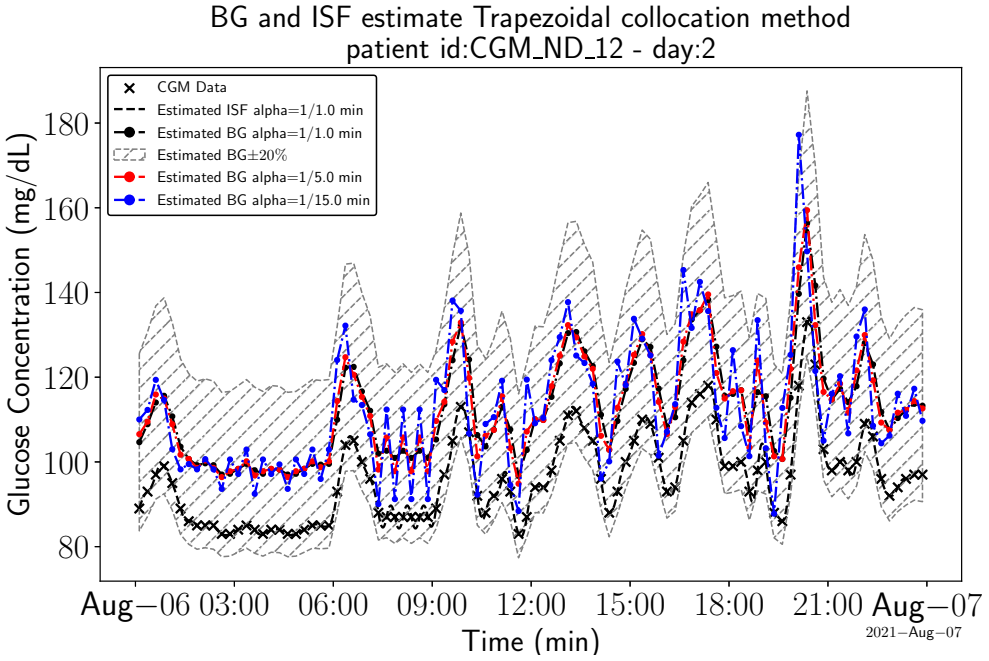


(a) The estimates here are obtained using the trapezoidal collocation method

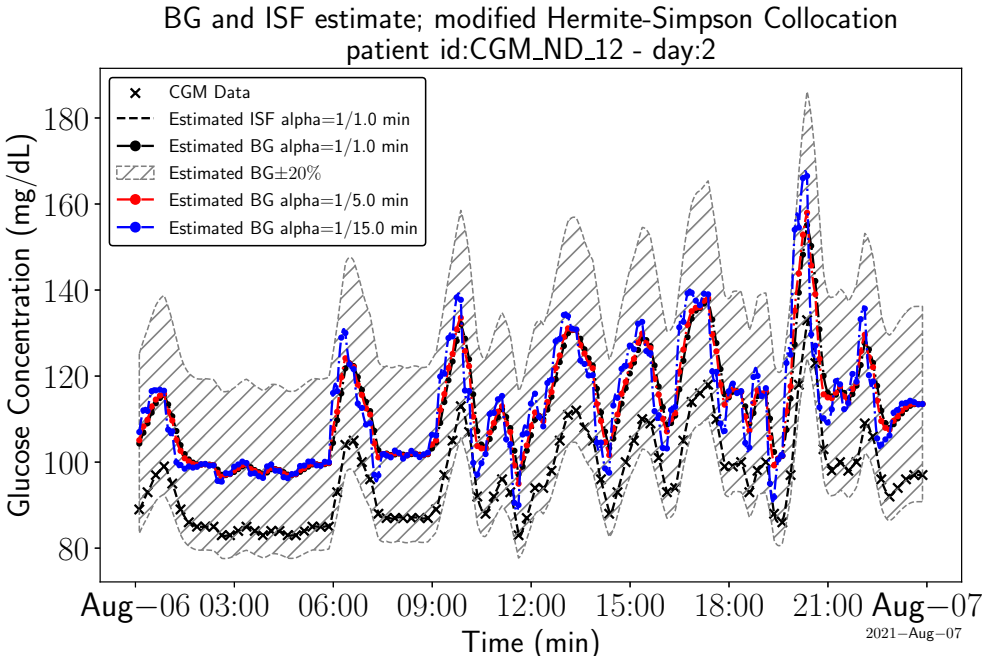


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.27: Estimated BG and ISF for `CGM ND 11` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

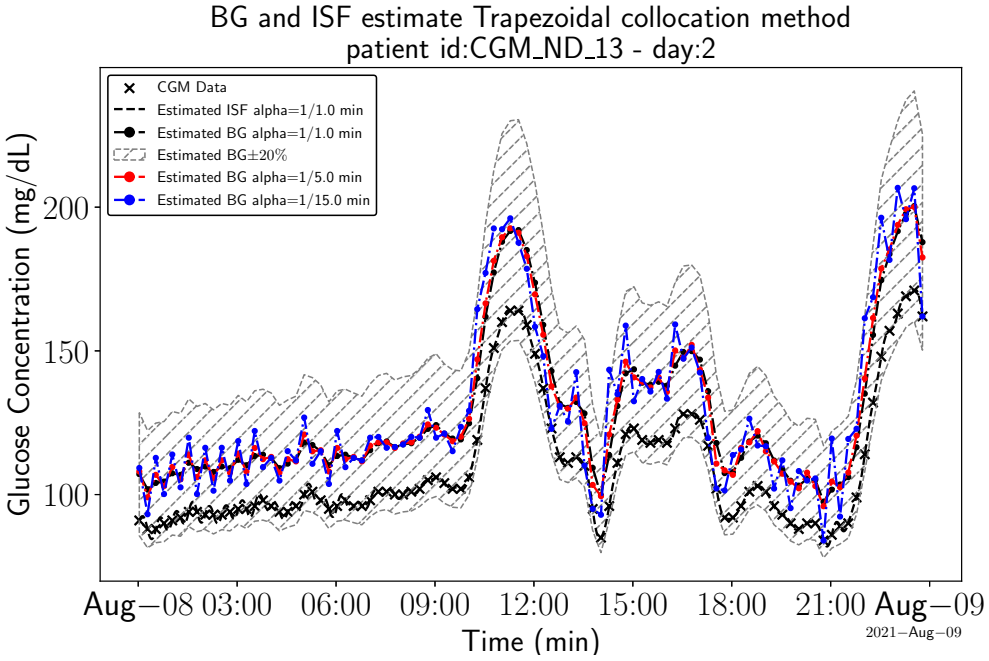


(a) The estimates here are obtained using the trapezoidal collocation method

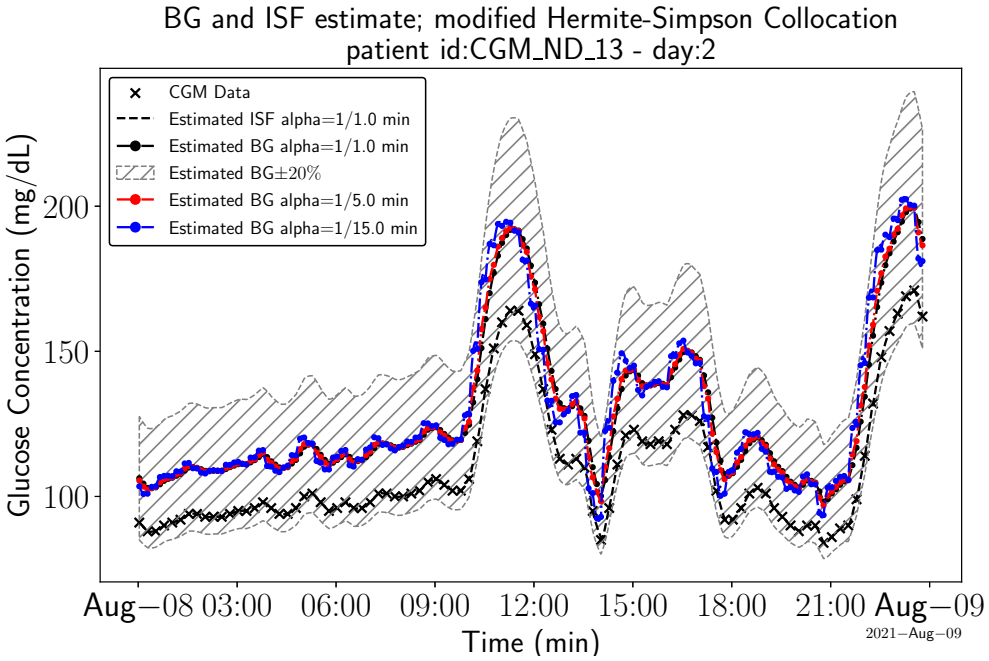


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.28: Estimated BG and ISF for `CGM ND 12` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

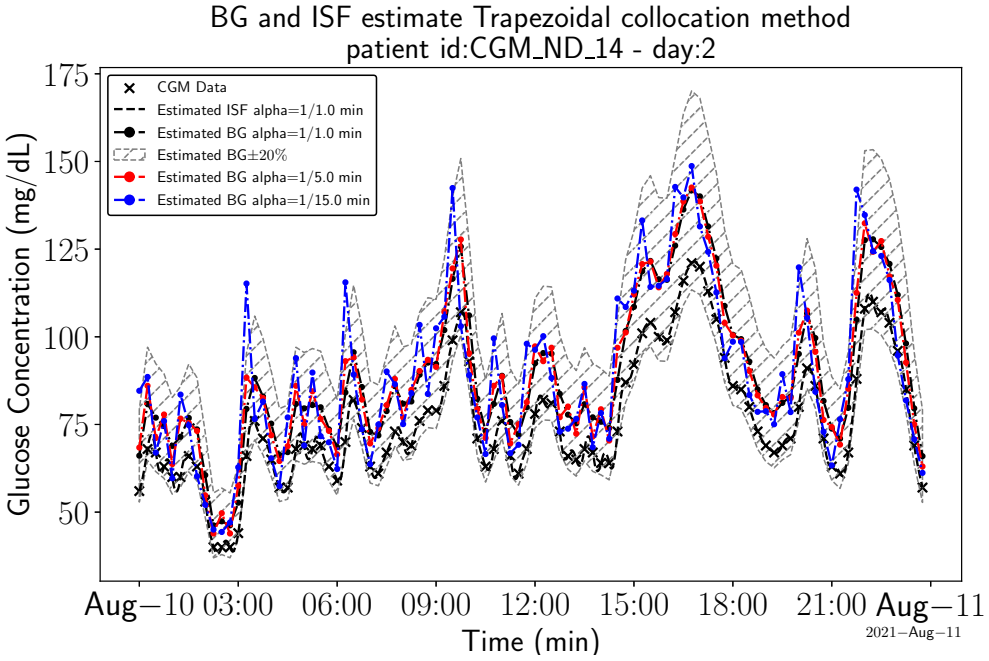


(a) The estimates here are obtained using the trapezoidal collocation method

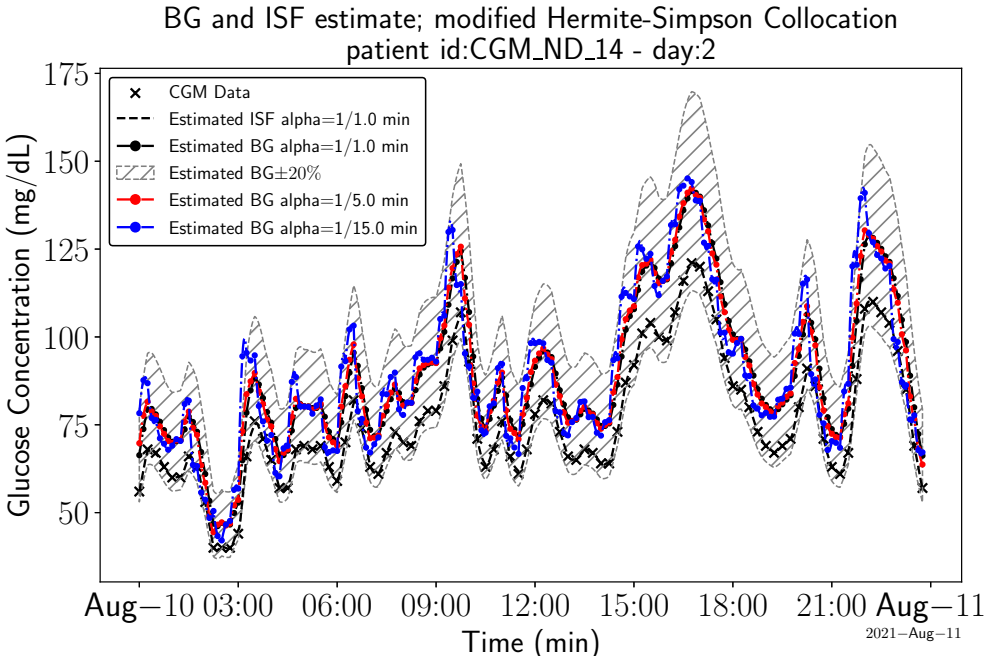


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.29: Estimated BG and ISF for `CGM ND 13` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

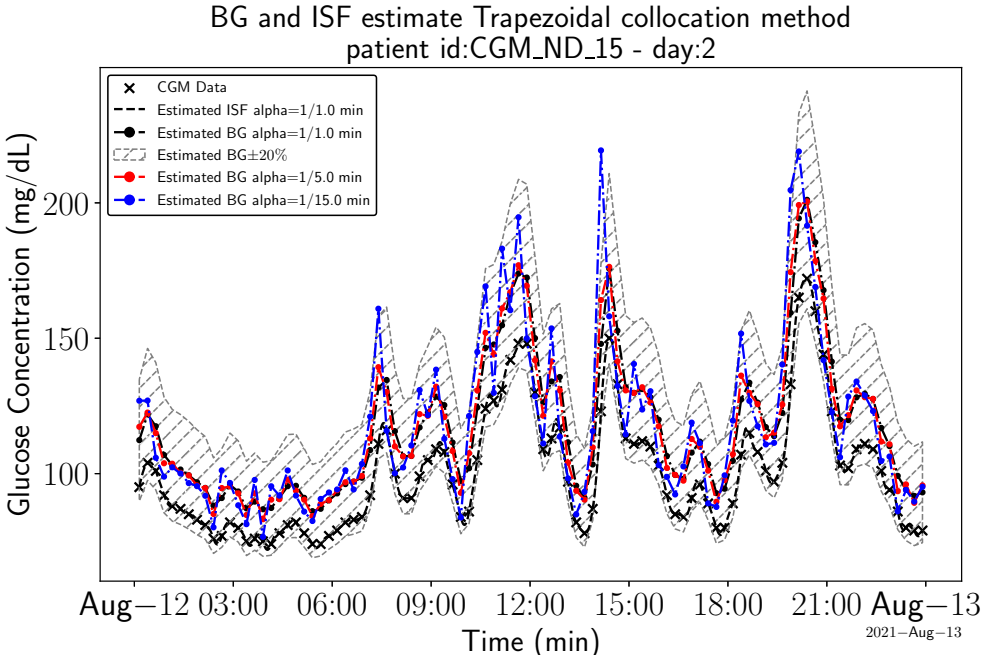


(a) The estimates here are obtained using the trapezoidal collocation method

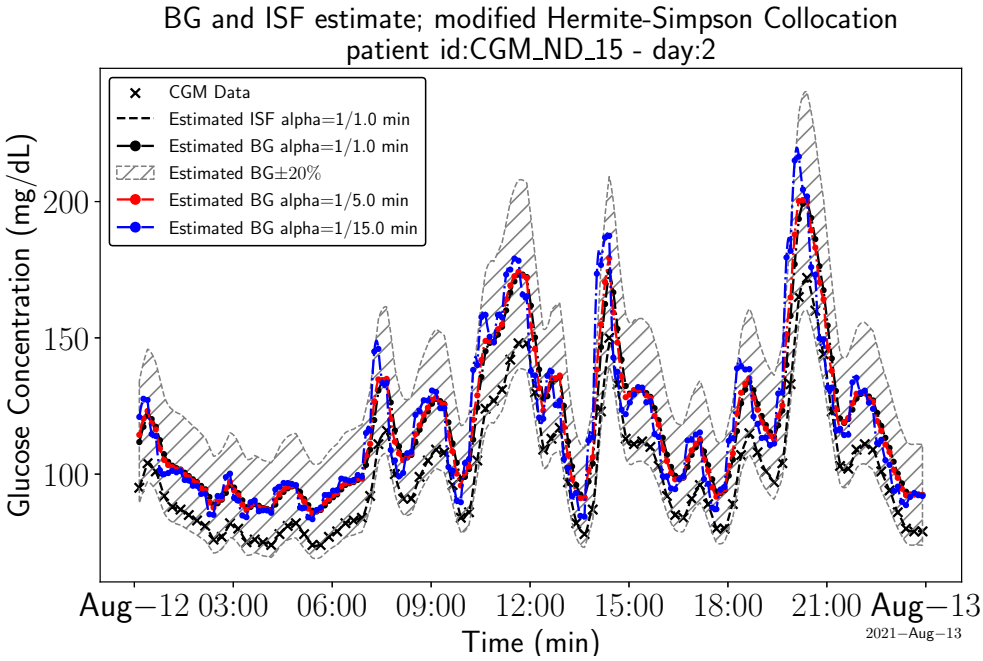


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.30: Estimated BG and ISF for `CGM ND 14` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

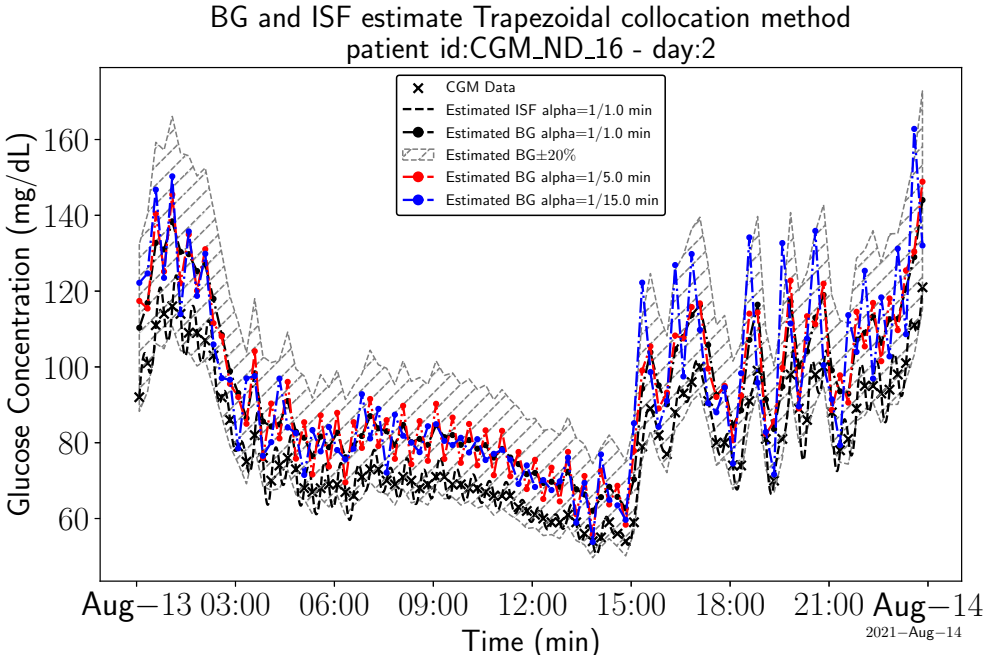


(a) The estimates here are obtained using the trapezoidal collocation method

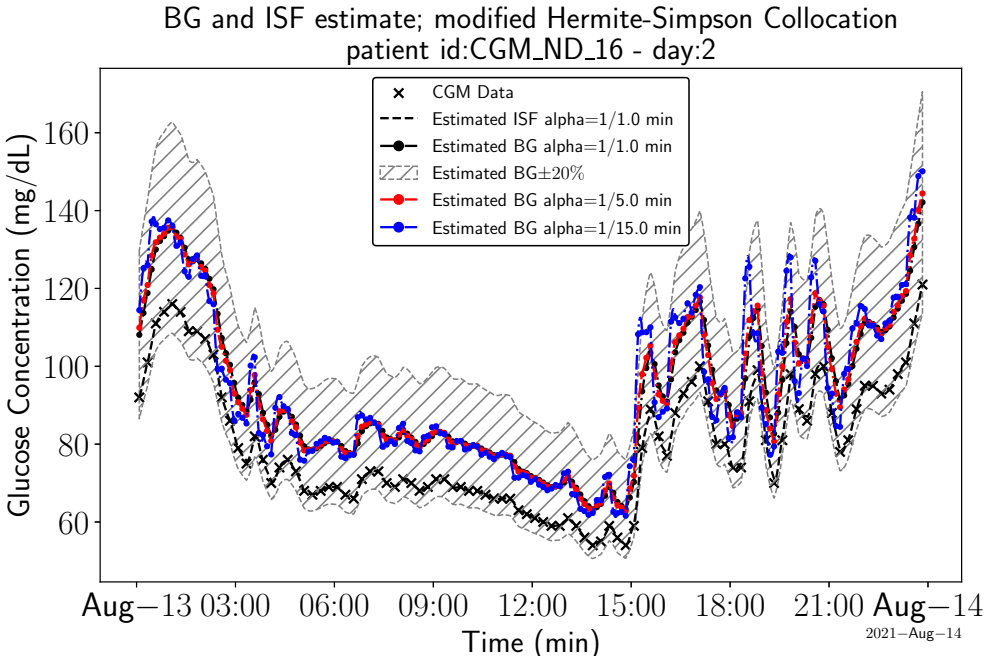


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.31: Estimated BG and ISF for `CGM ND 15` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

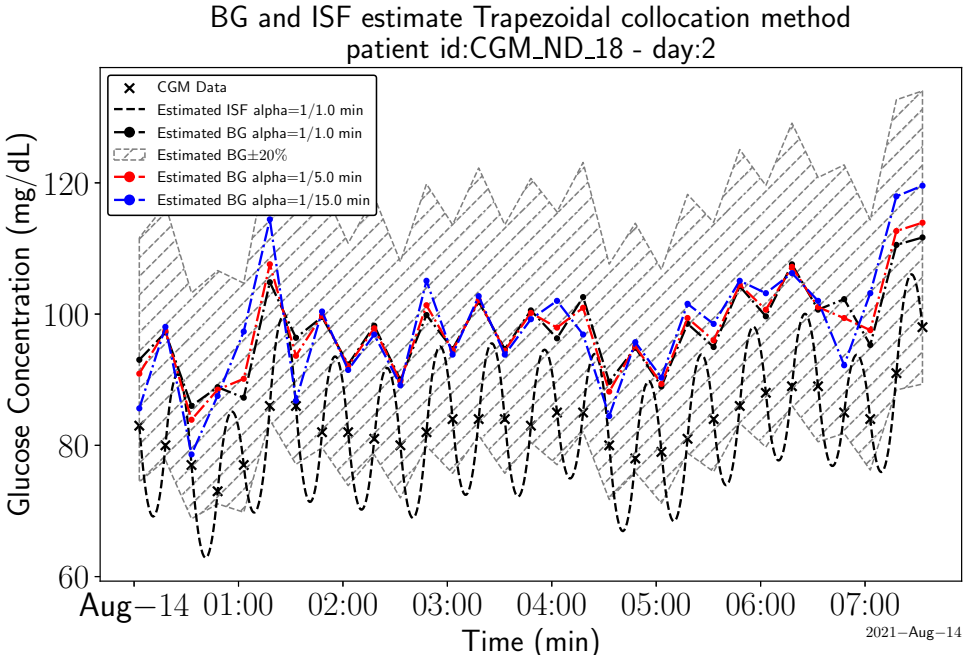


(a) The estimates here are obtained using the trapezoidal collocation method

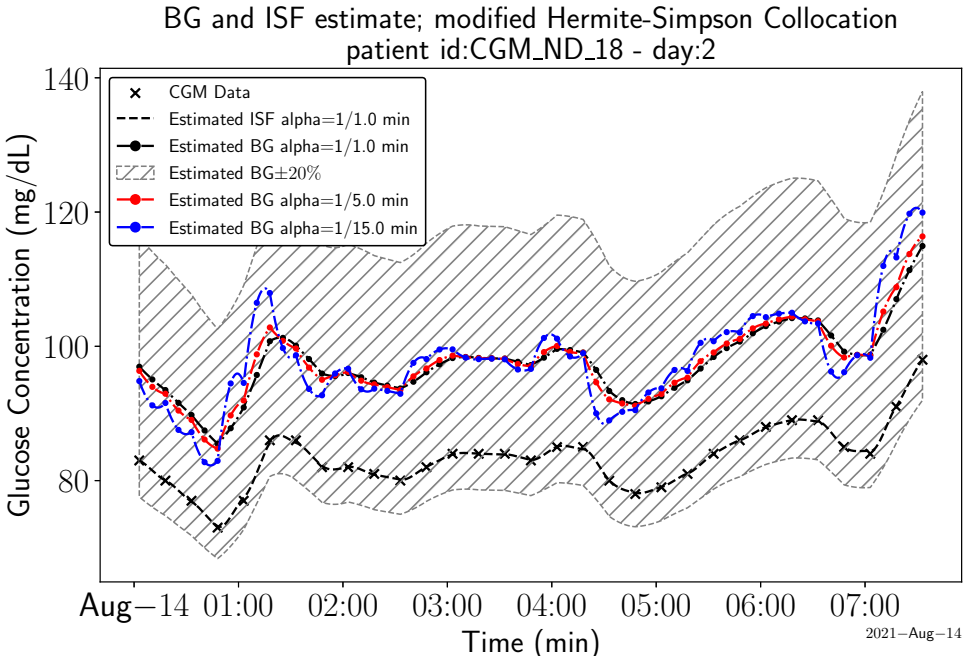


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.32: Estimated BG and ISF for `CGM ND 16` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

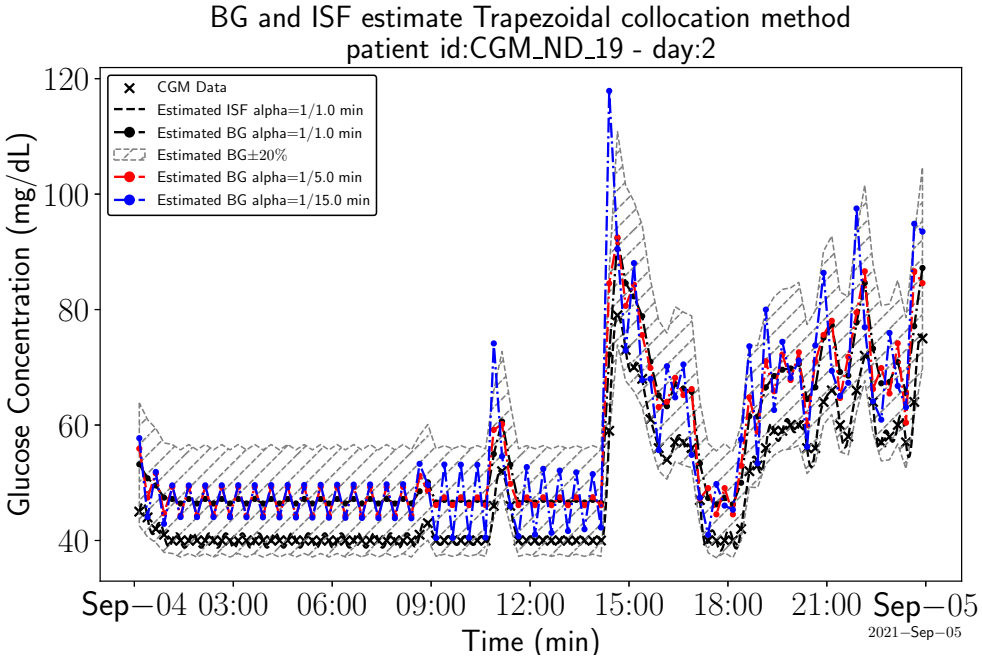


(a) The estimates here are obtained using the trapezoidal collocation method

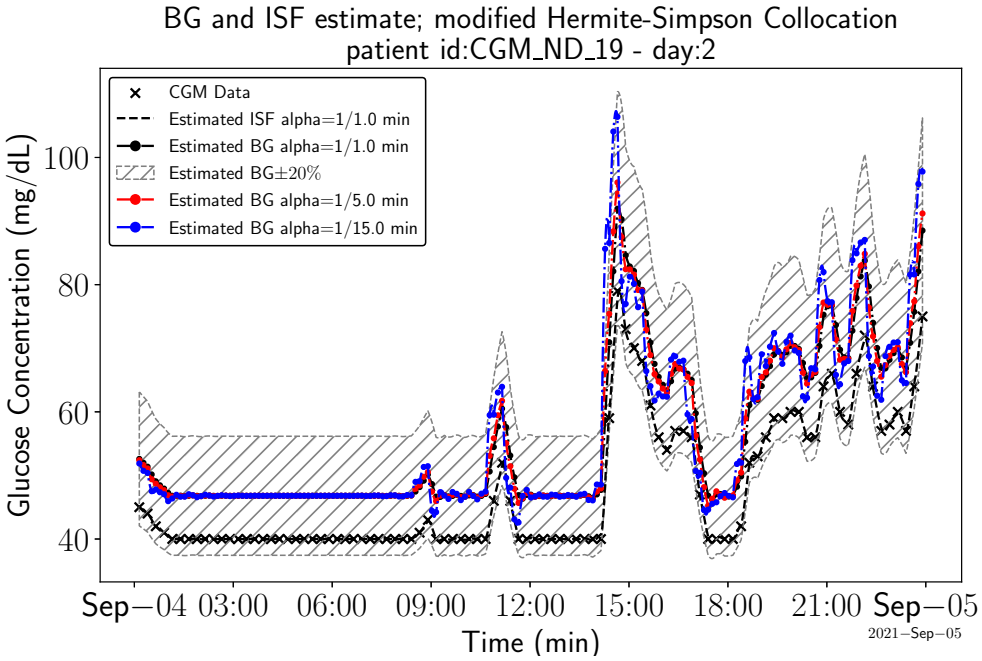


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.33: Estimated BG and ISF for `CGM ND 18` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

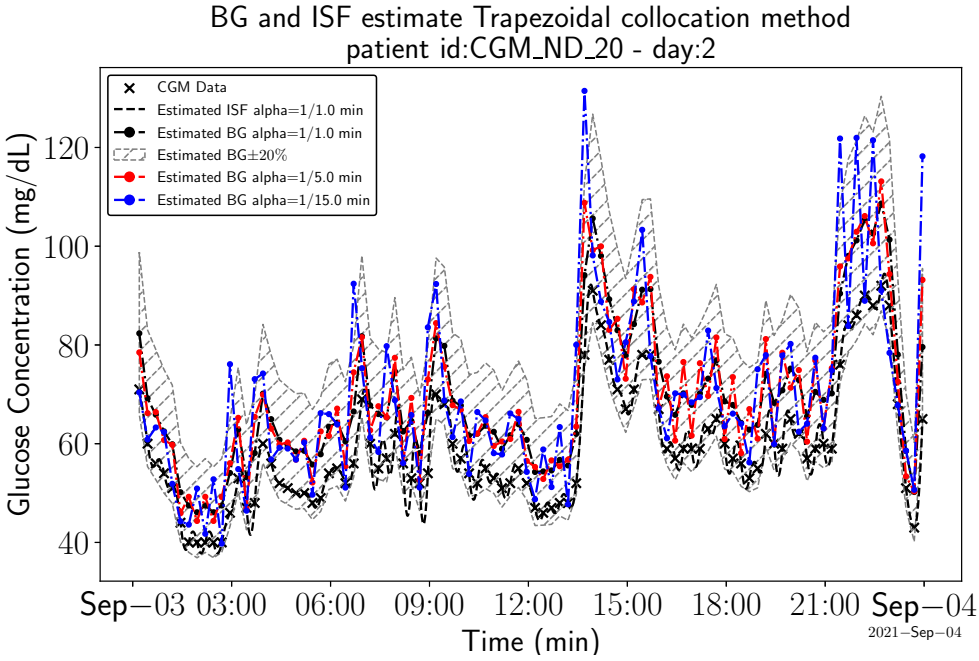


(a) The estimates here are obtained using the trapezoidal collocation method

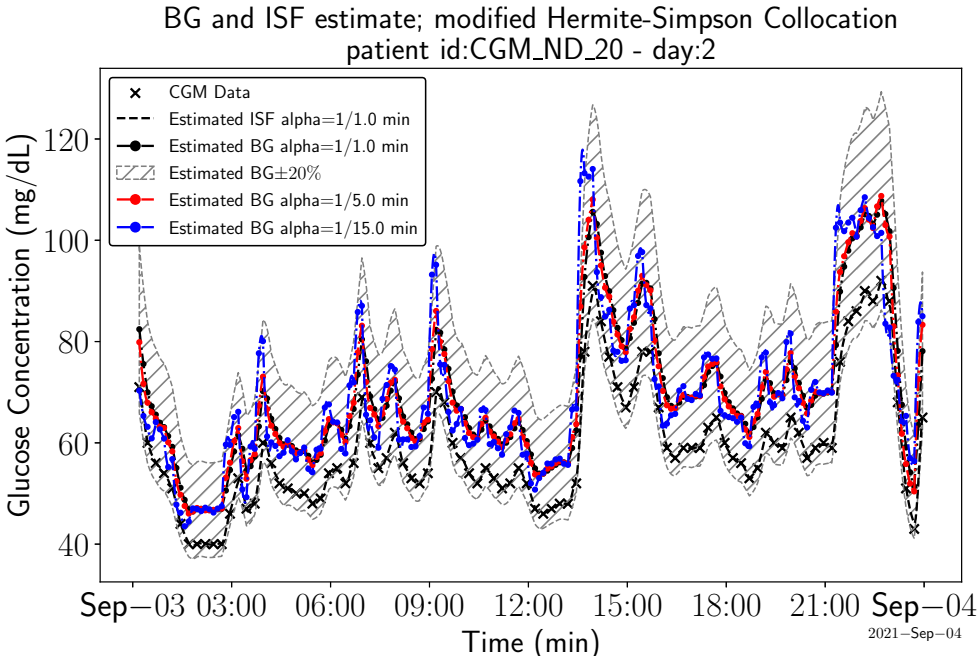


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.34: Estimated BG and ISF for `CGM ND 19` for τ_0 = 1 min, 1/5min and, 1/15min

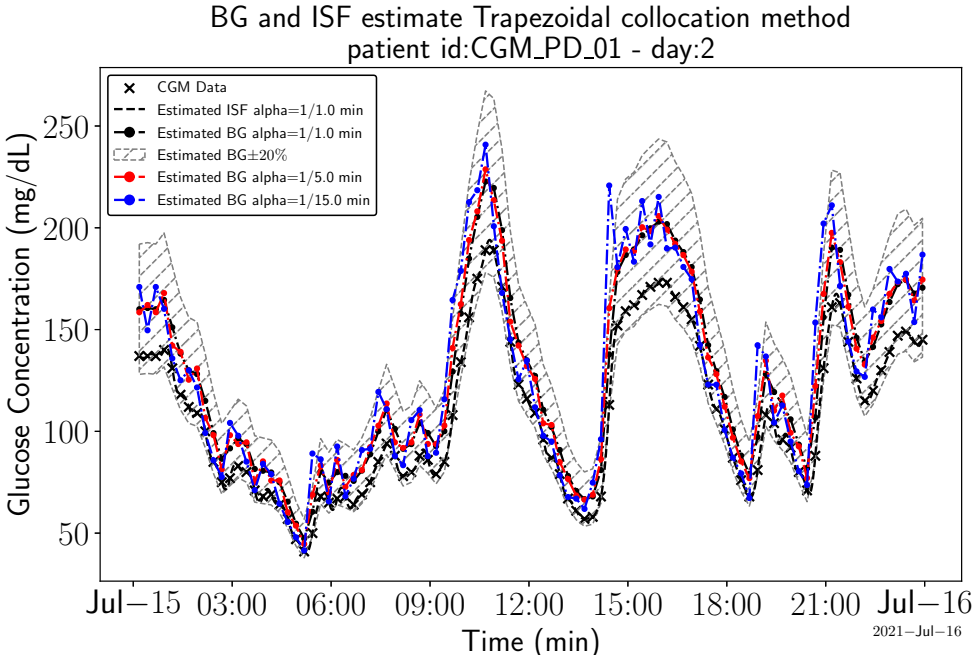


(a) The estimates here are obtained using the trapezoidal collocation method

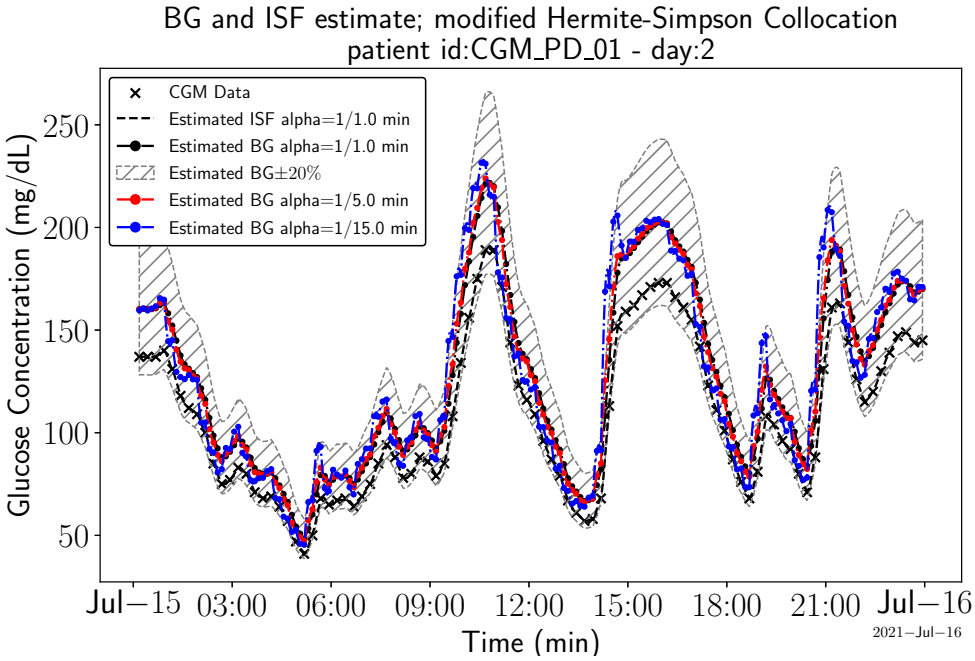


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.35: Estimated BG and ISF for `CGM ND 20` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

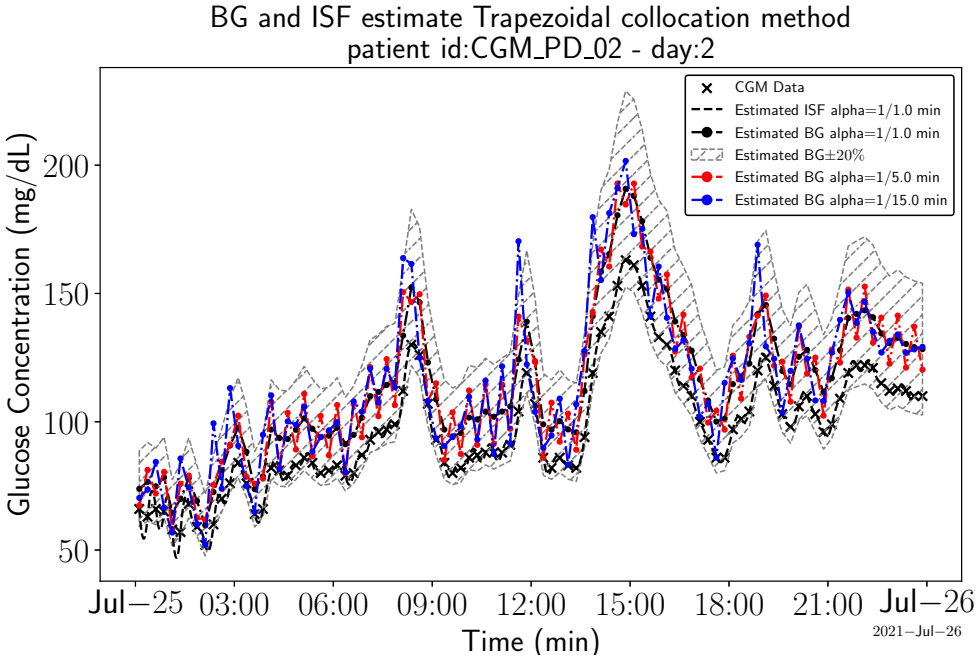


(a) The estimates here are obtained using the trapezoidal collocation method

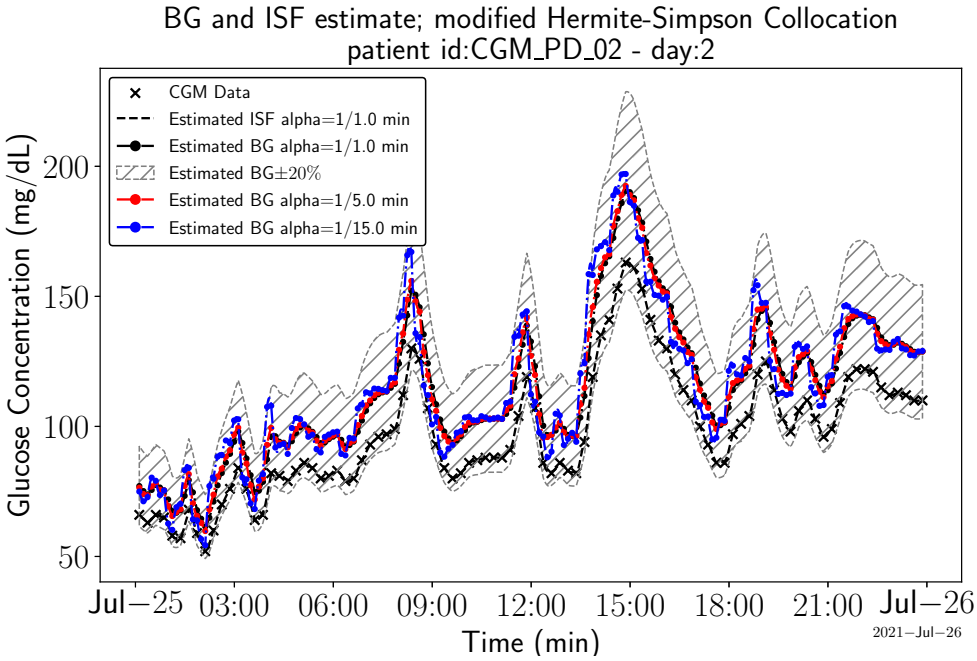


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.36: Estimated BG and ISF for `CGM PD 01` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

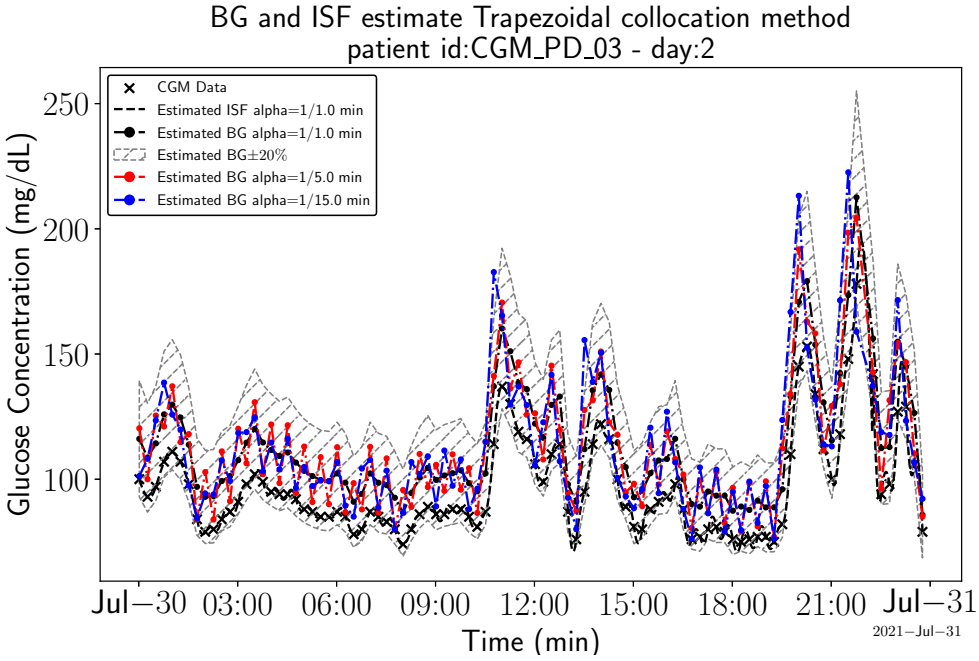


(a) The estimates here are obtained using the trapezoidal collocation method

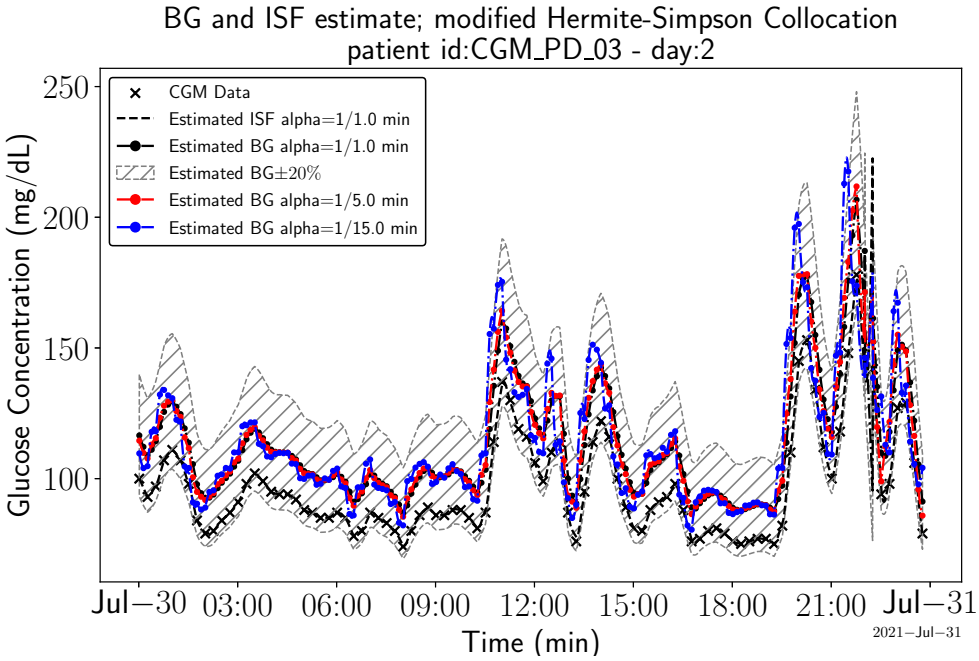


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.37: Estimated BG and ISF for `CGM PD 02` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

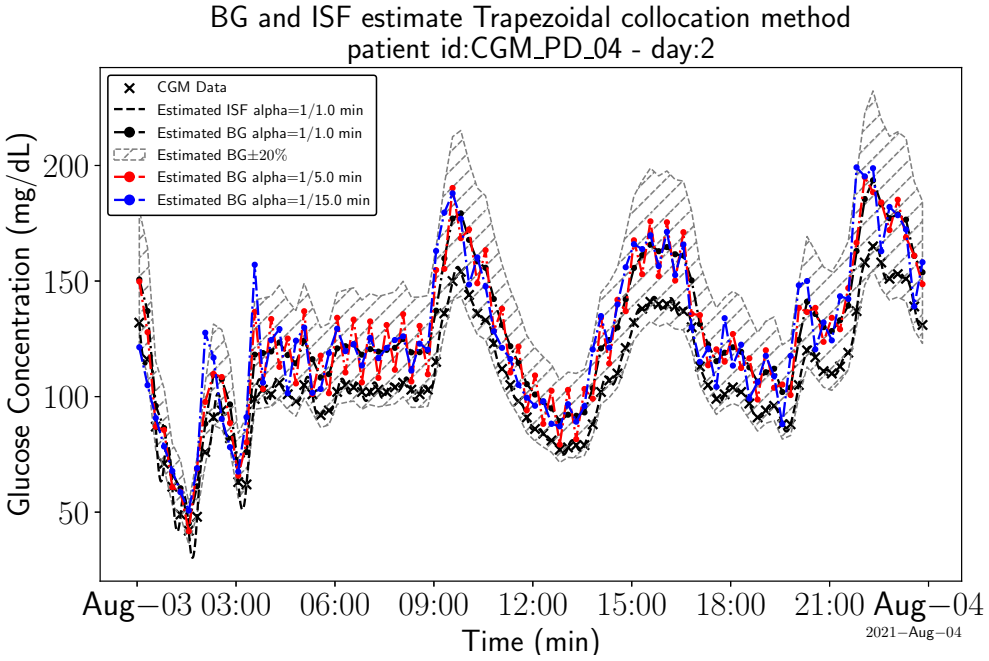


(a) The estimates here are obtained using the trapezoidal collocation method

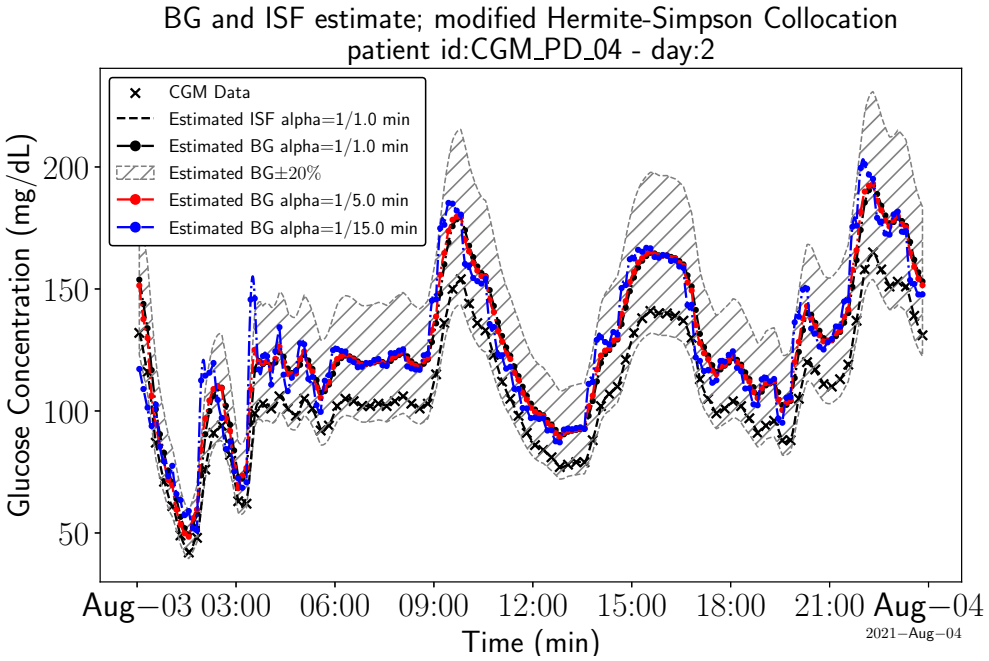


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.38: Estimated BG and ISF for `CGM PD 03` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

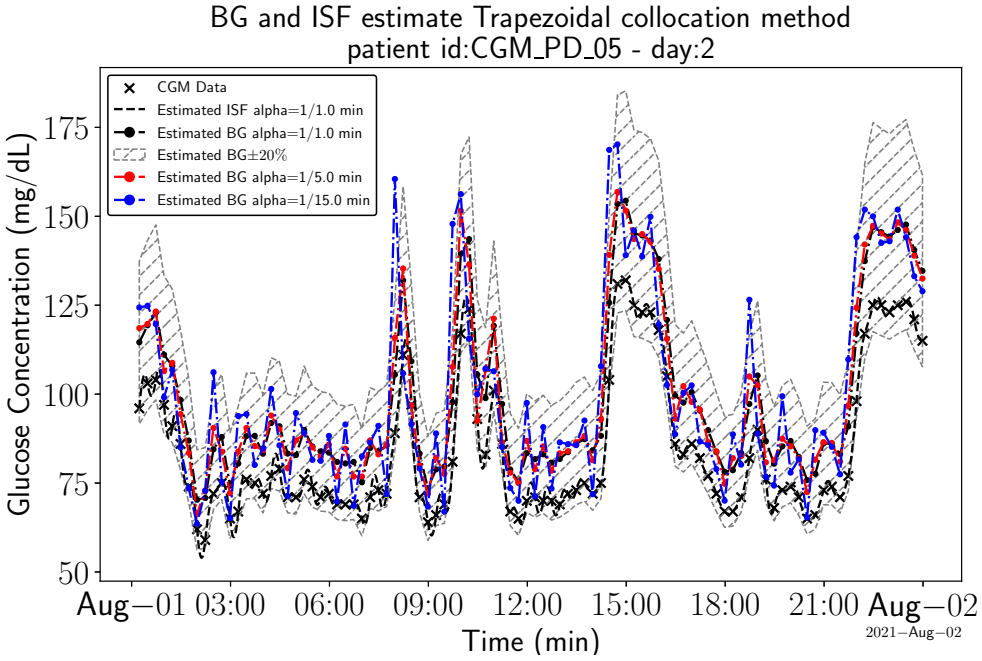


(a) The estimates here are obtained using the trapezoidal collocation method

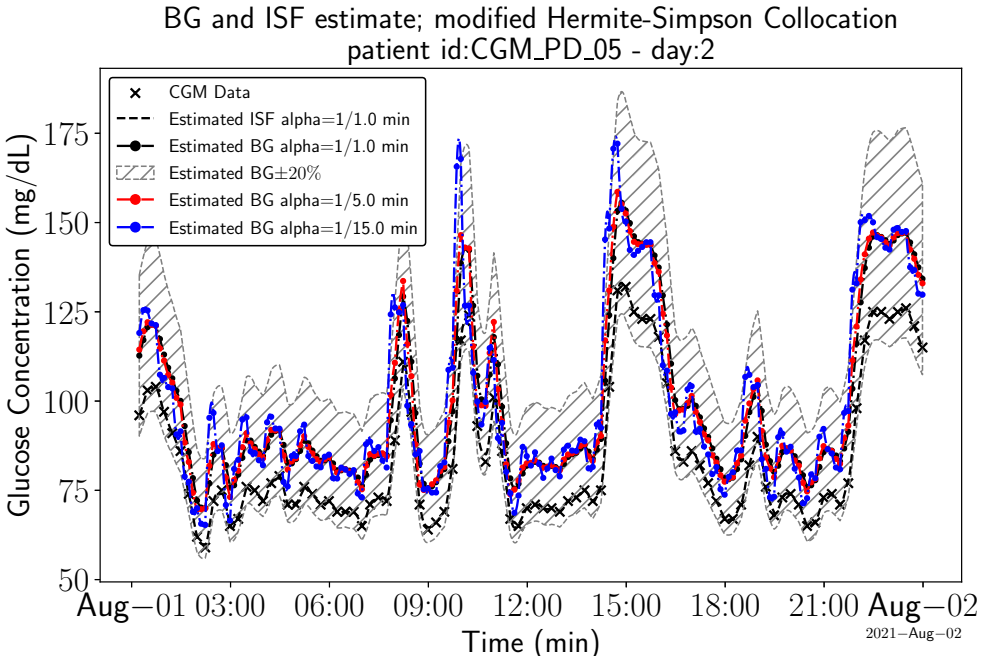


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.39: Estimated BG and ISF for `CGM PD 04` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

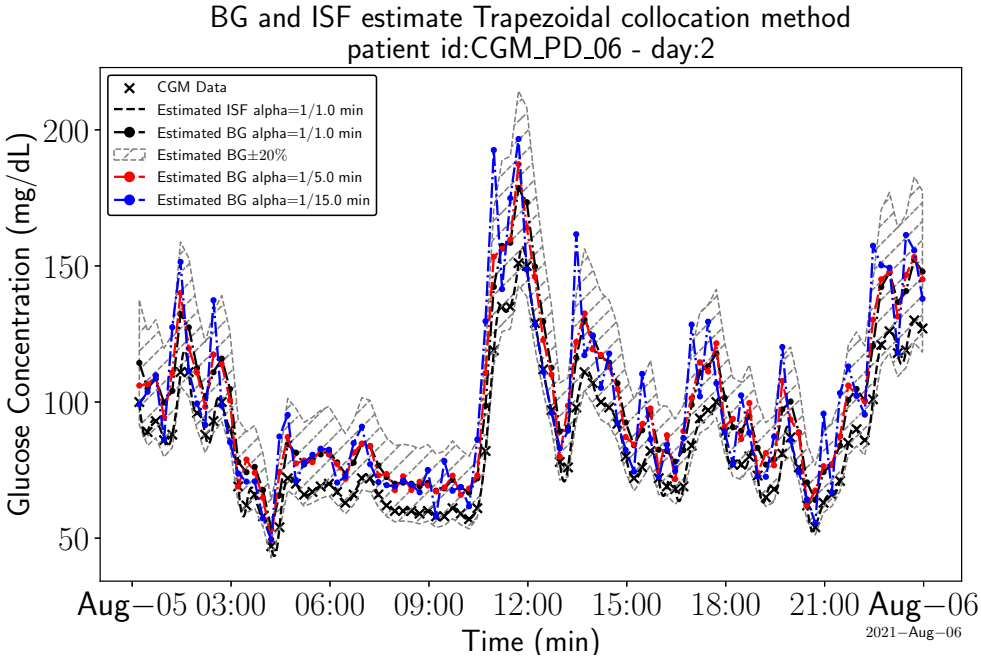


(a) The estimates here are obtained using the trapezoidal collocation method

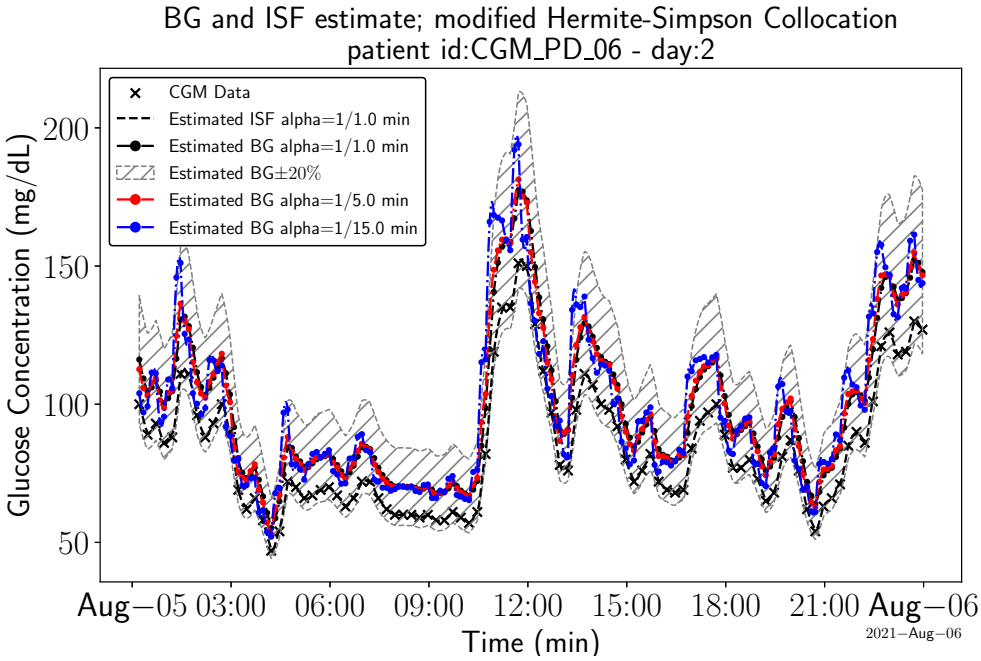


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.40: Estimated BG and ISF for `CGM PD 05` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

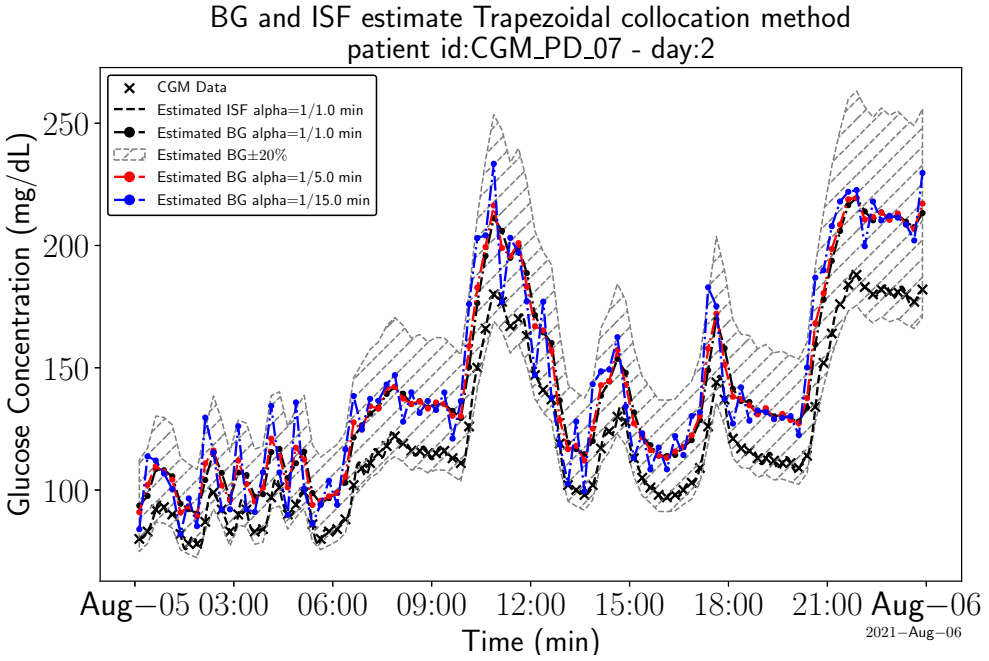


(a) The estimates here are obtained using the trapezoidal collocation method

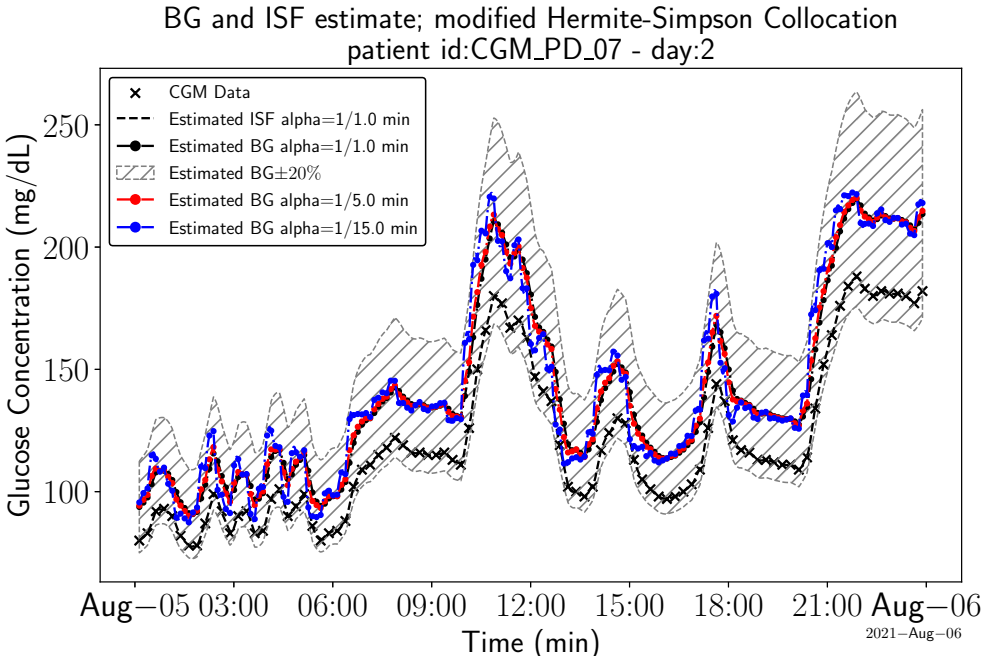


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.41: Estimated BG and ISF for `CGM PD 06` for τ_0 = 1 min, 1/5min and, 1/15min

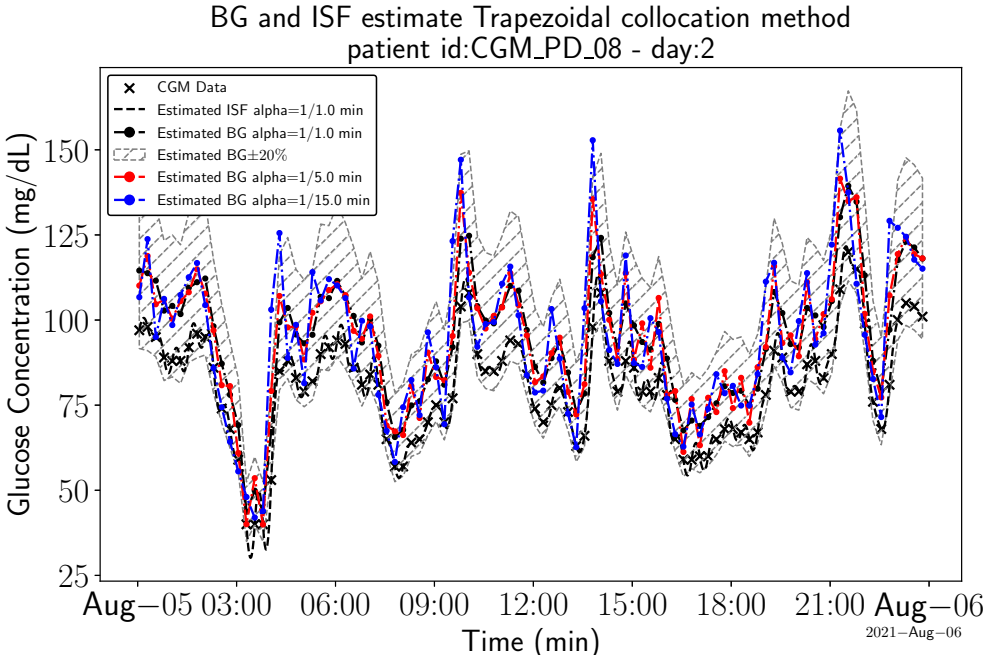


(a) The estimates here are obtained using the trapezoidal collocation method

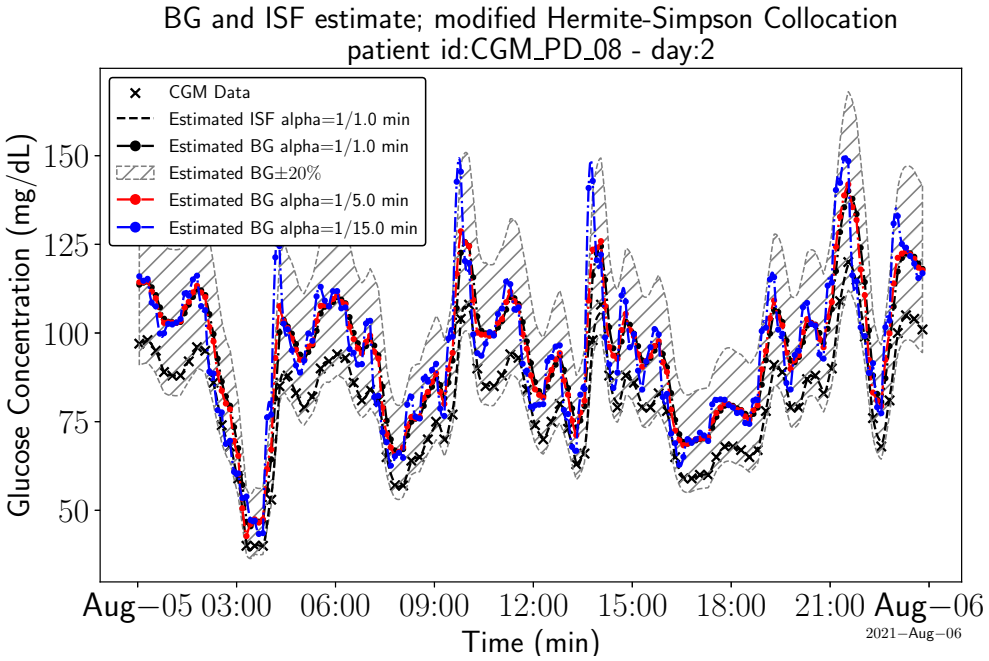


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.42: Estimated BG and ISF for `CGM PD 07` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

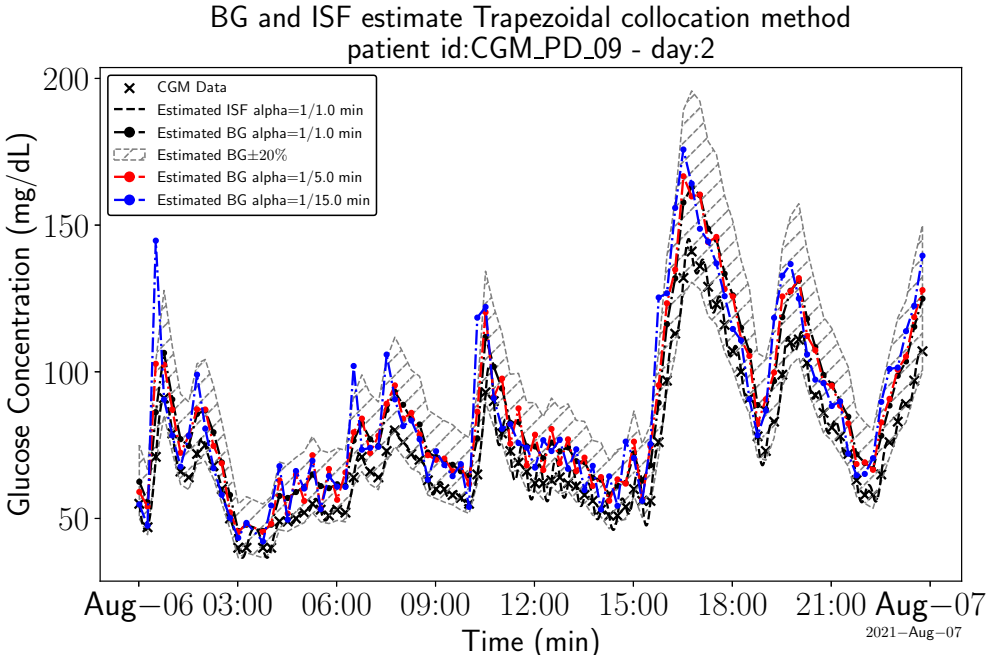


(a) The estimates here are obtained using the trapezoidal collocation method

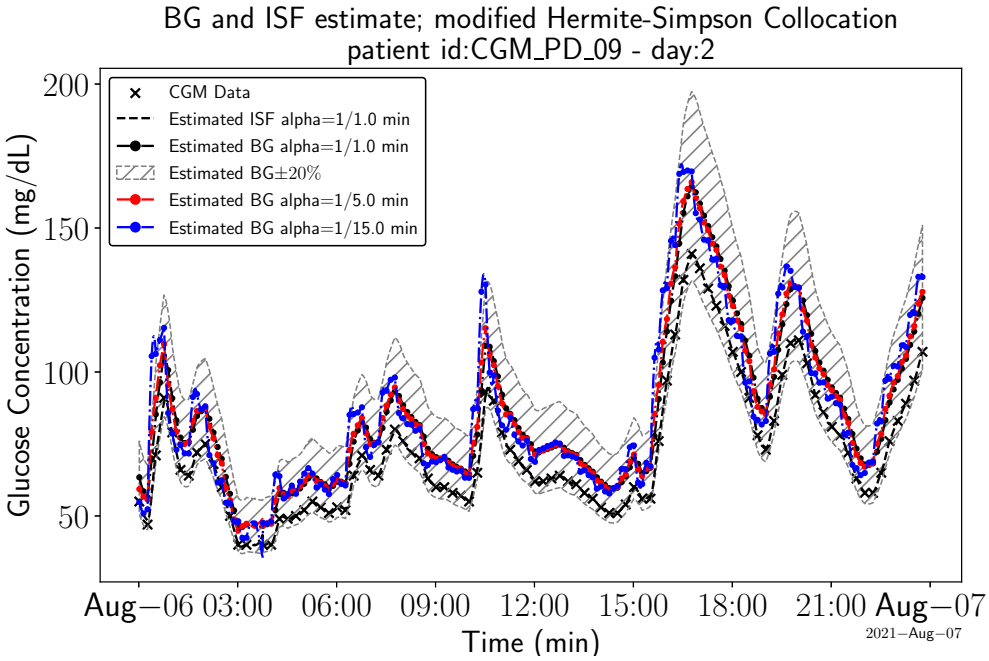


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.43: Estimated BG and ISF for `CGM PD 08` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

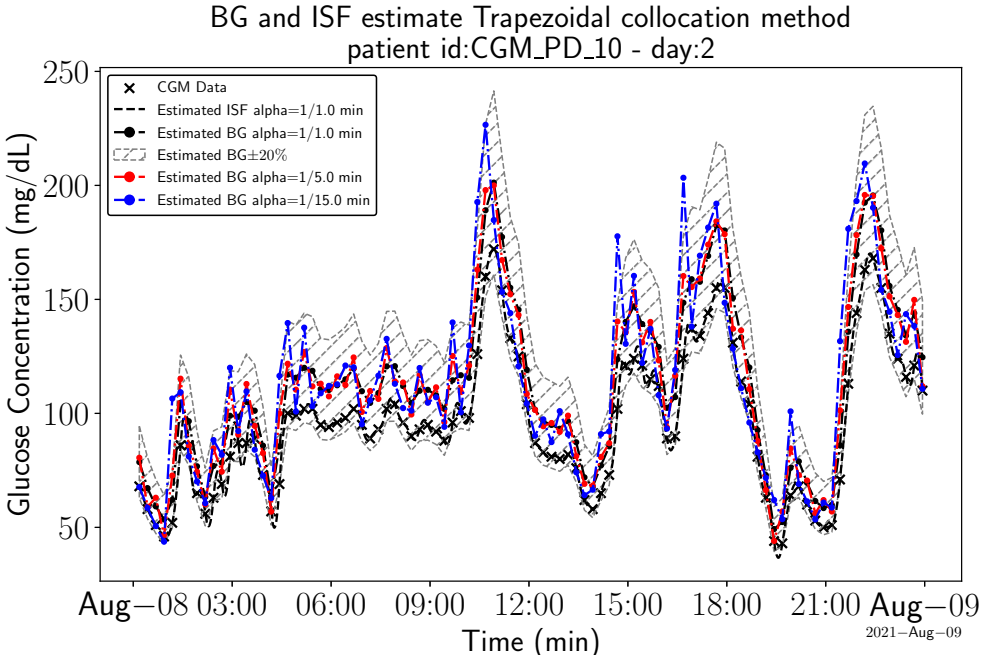


(a) The estimates here are obtained using the trapezoidal collocation method

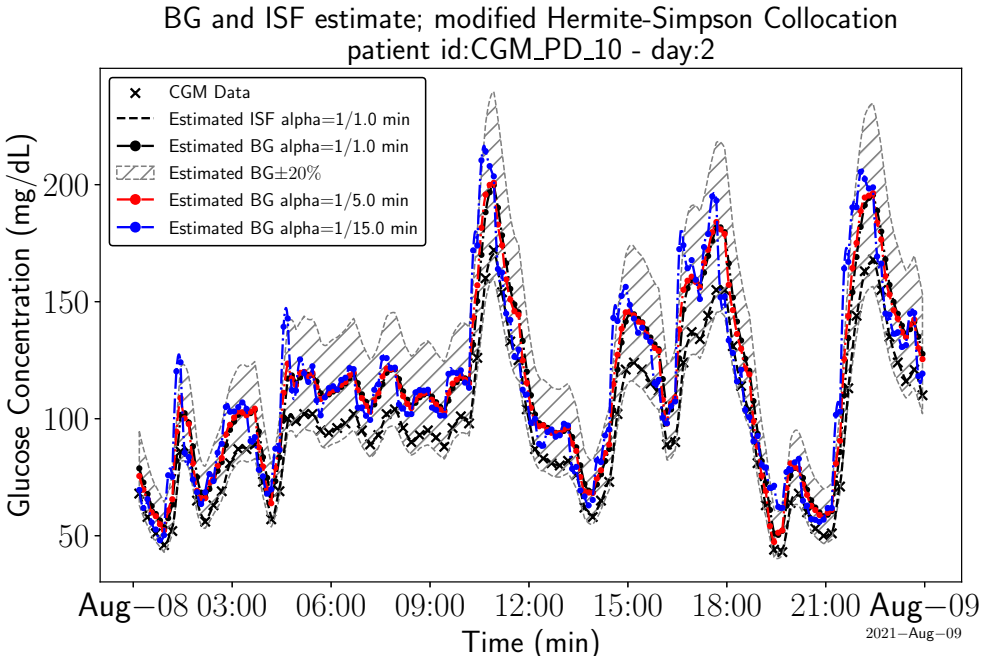


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.44: Estimated BG and ISF for `CGM PD 09` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

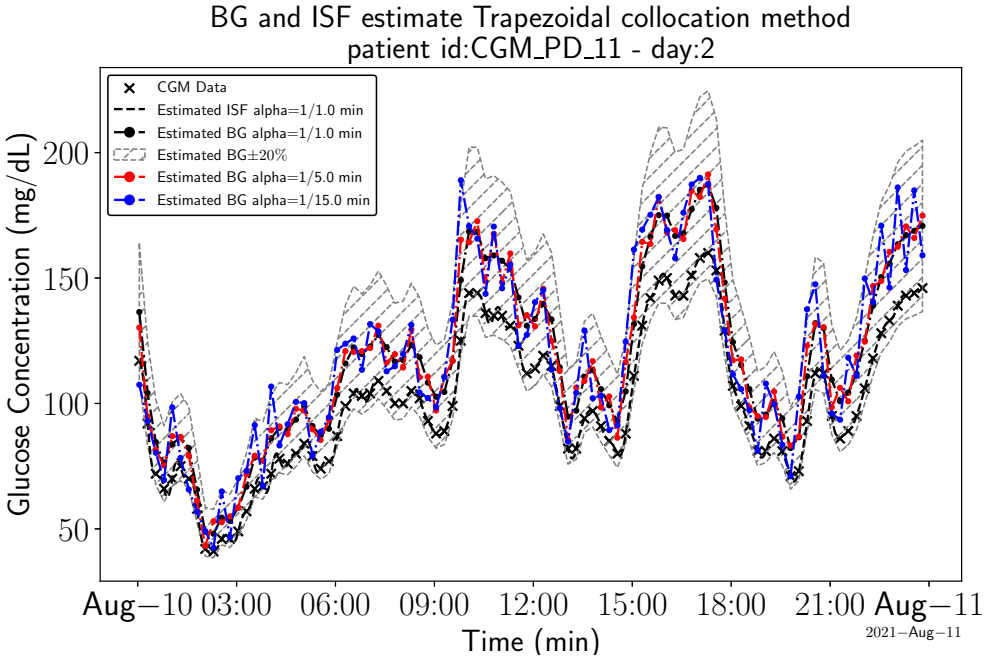


(a) The estimates here are obtained using the trapezoidal collocation method

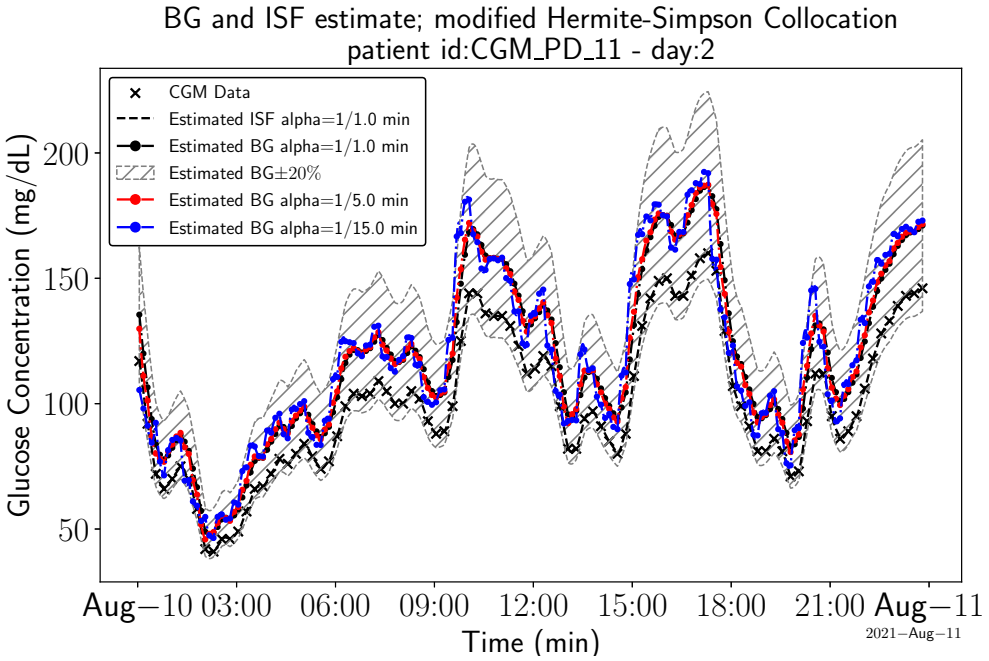


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.45: Estimated BG and ISF for `CGM PD 10` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

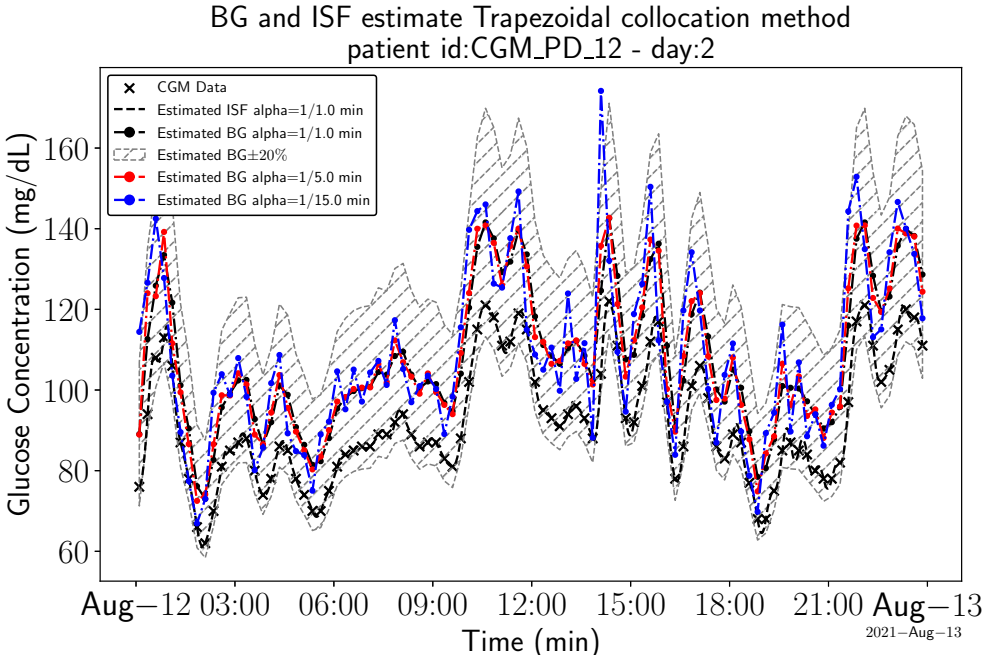


(a) The estimates here are obtained using the trapezoidal collocation method

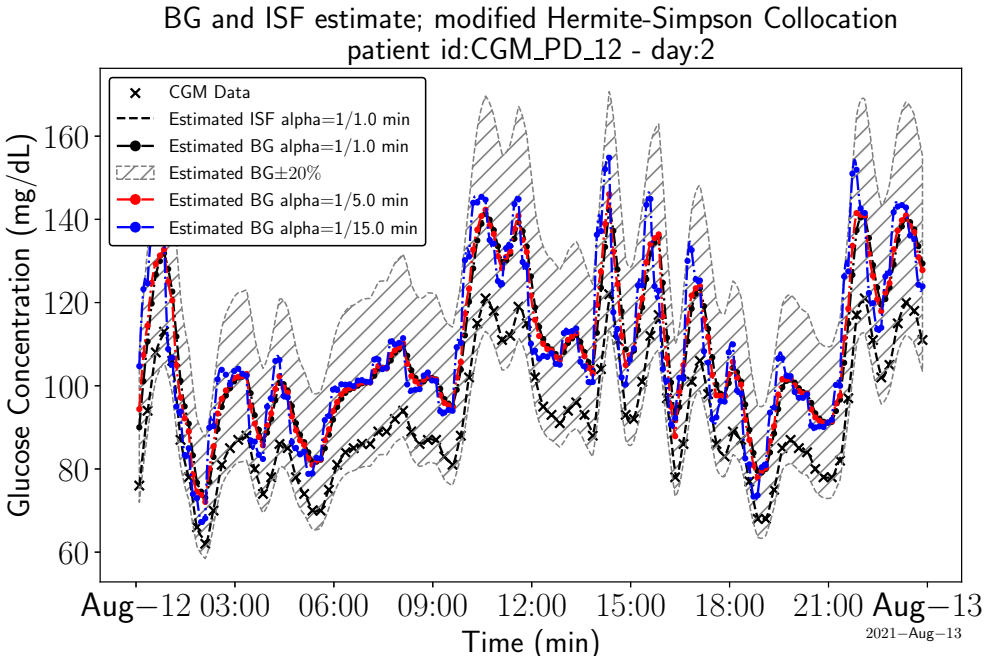


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.46: Estimated BG and ISF for `CGM PD 11` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

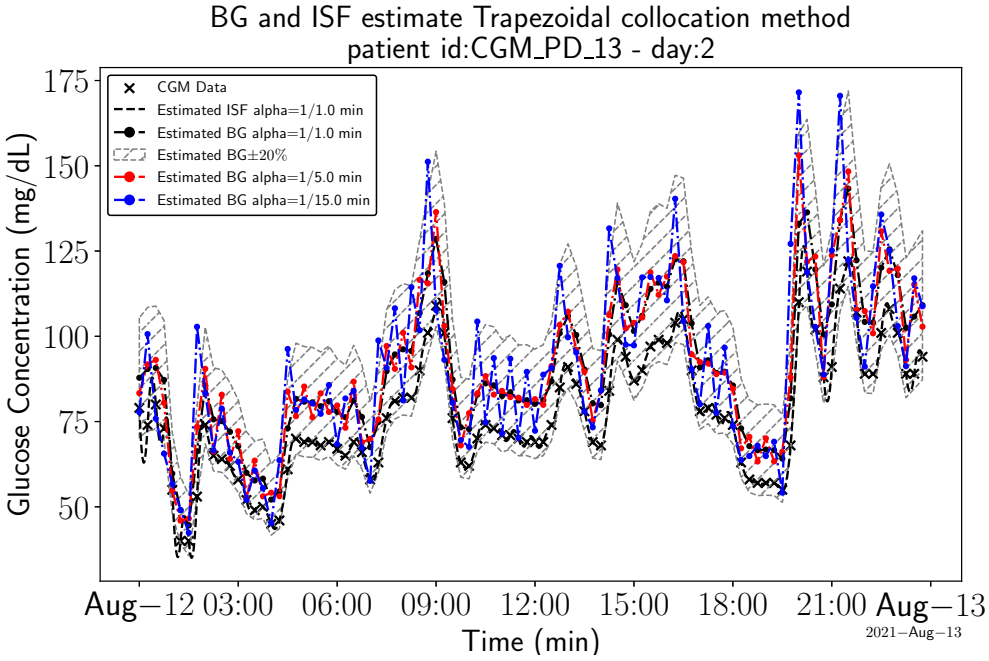


(a) The estimates here are obtained using the trapezoidal collocation method

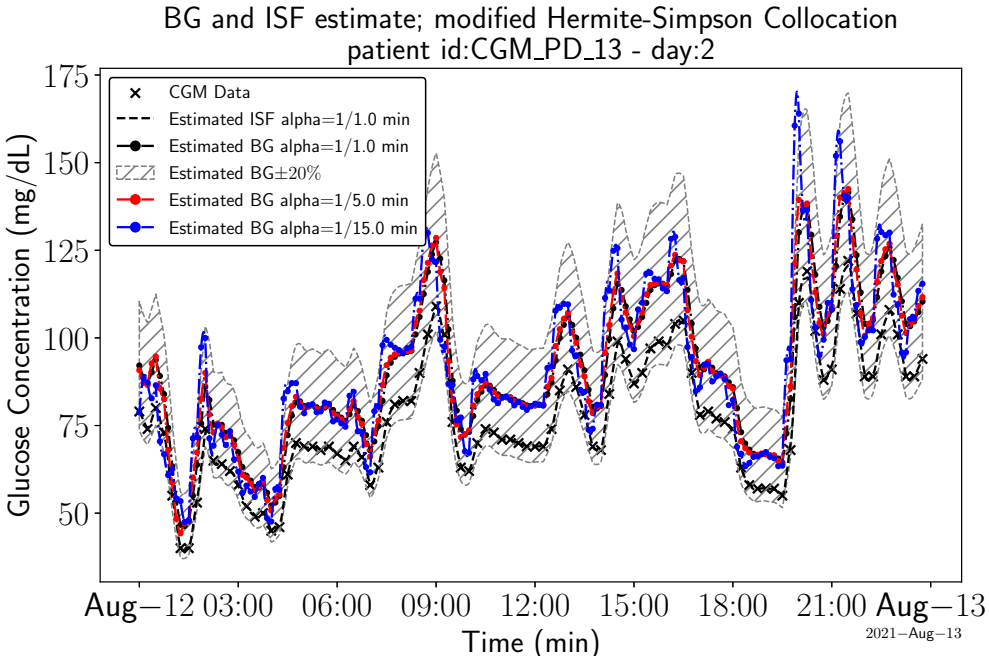


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.47: Estimated BG and ISF for `CGM PD 12` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

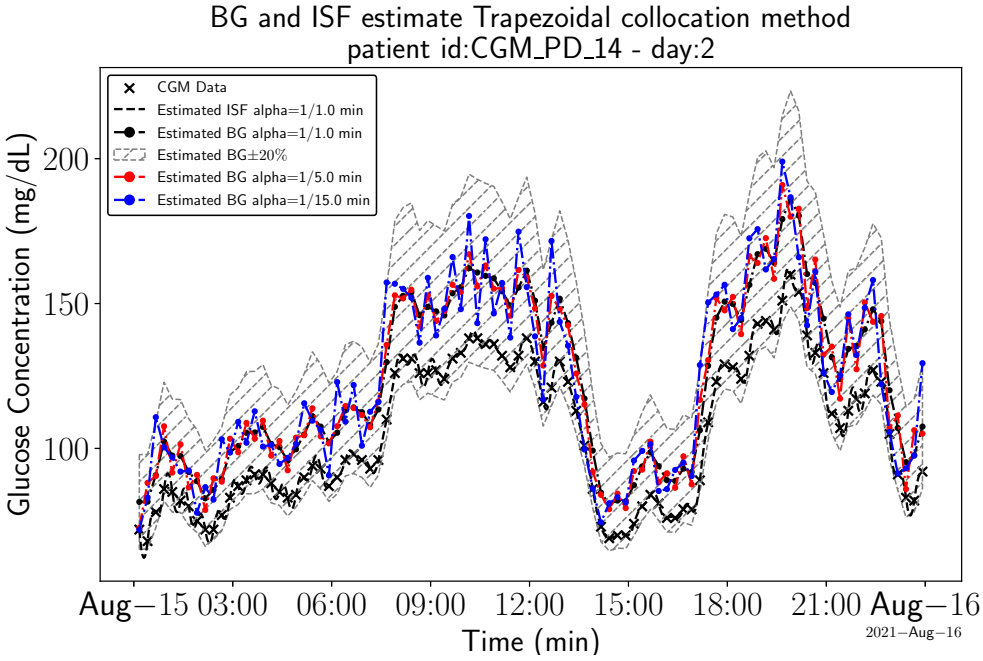


(a) The estimates here are obtained using the trapezoidal collocation method

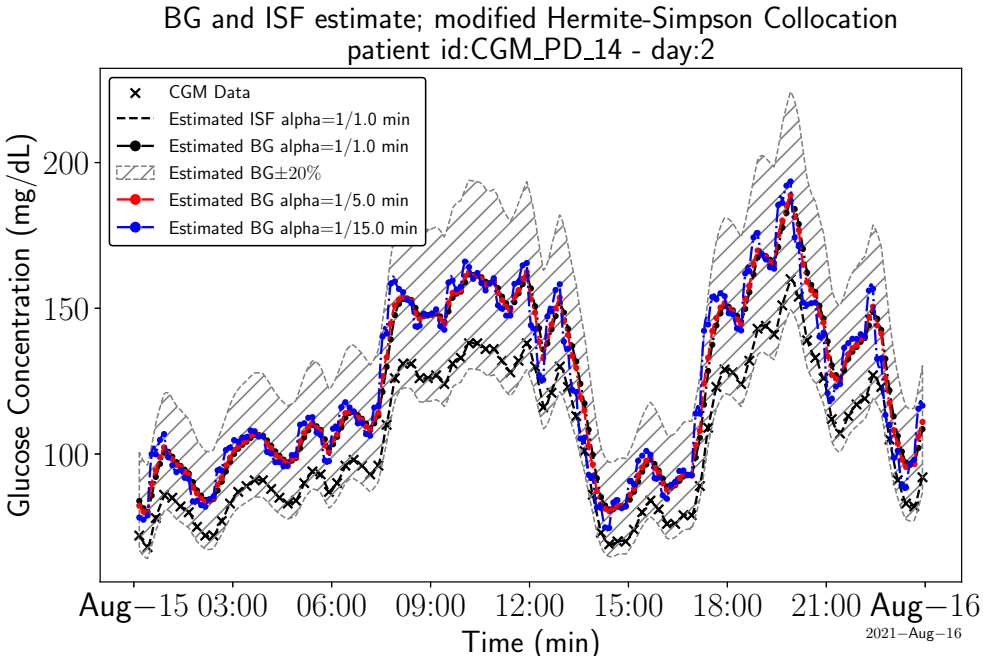


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.48: Estimated BG and ISF for `CGM PD 13` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

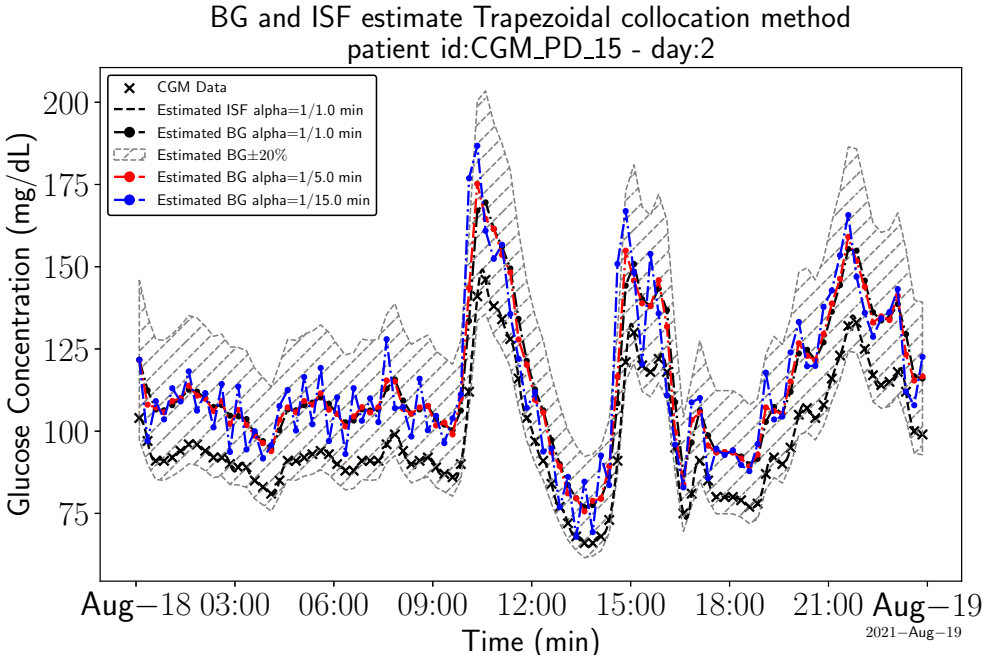


(a) The estimates here are obtained using the trapezoidal collocation method

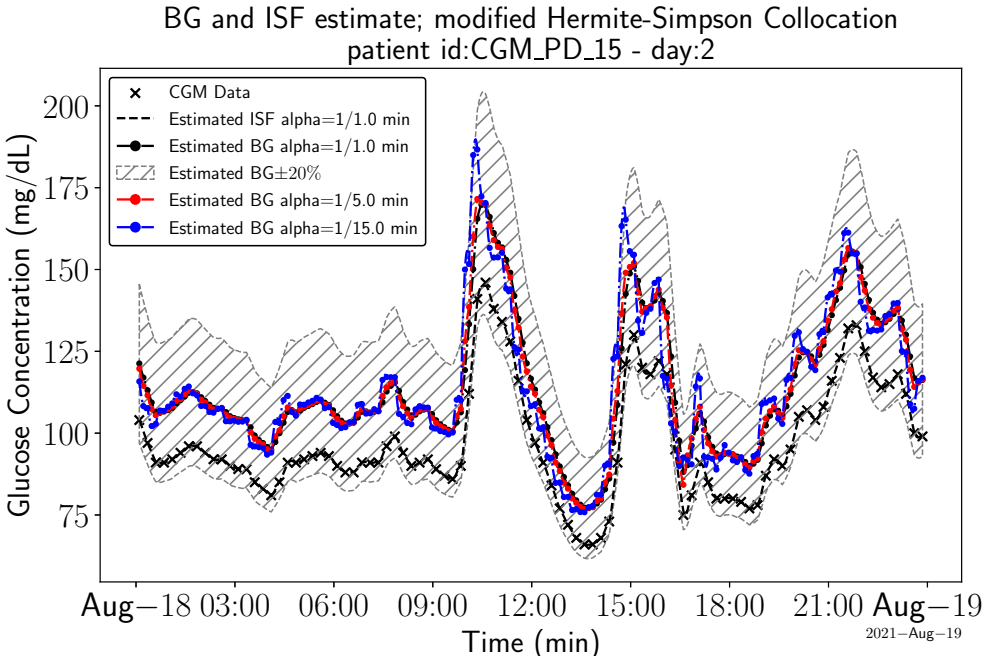


(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.49: Estimated BG and ISF for `CGM PD 14` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$



(a) The estimates here are obtained using the trapezoidal collocation method



(b) The estimates here are obtained using the modified Hermite-Simpson collocation method

Figure A.50: Estimated BG and ISF for `CGM PD 15` for $\tau_0 = 1\,\text{min},\ 1/5\text{min}$ and, $1/15\text{min}$

Appendix B

Published works

B.1 Published articles

Here is a list of the published papers

1. Majumdar, S. et al. Evaluation of HbA1c from CGM traces in an Indian population. *Frontiers in Endocrinology* **2023**, *14*, 1264072, DOI: 10.3389/fendo.2023.1264072, https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2023.1264072/full

B.2 Conference posters and presentations

Here is a list of conference posters and presentations given.

1. **Conference Poster:** An e-poster titled **"A comparison of algebraic and dynamic models to estimate blood glucose and A1C from CGM time series"** in the 16th International Conference on Advanced Technologies & Treatments for Diabetes. EP165/#606
   Abstract book DOI: https://doi.org/10.1089/dia.2023.2525.abstracts
   Abstract author index DOI: https://doi.org/10.1089/dia.2023.2526.abstracts.index.

2. **Conference Oral Presentation:** Oral presentation titled **"How to track blood glucose from the CGM?"** at the 7th International Diabetes Summit

2023 (IDS 2023) organized by Chellaram Diabetes Institute. The website for the conference is provided here https://www.cdi.org.in/international-diabetes-summit and https://www.cdi.org.in/7th-international-diabetes-summit-2023/.

B.3 Copyright and Licence

The copyright and license information about the published artile is provided below

- Majumdar S, Kalamkar SD, Dudhgaonkar S, Shelgikar KM, Ghaskadbi S and Goel P (2023) **Evaluation of HbA1c from CGM traces in an Indian population.** *Front. Endocrinol. 14:1264072. doi: 10.3389/fendo.2023.1264072*

  This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms. (see Figure. B.1)

OPEN ACCESS

EDITED BY
Maurizio Delvecchio,
Giovanni XXIII Children's Hospital, Italy

REVIEWED BY
Rahul Siddharthan,
Institute of Mathematical Sciences,
Chennai, India
Tanja Milicic,
University of Belgrade, Serbia

*CORRESPONDENCE
Sayantan Majumdar
majumdar.sayantan@
students.iiserpune.ac.in

RECEIVED 20 July 2023
ACCEPTED 17 October 2023
PUBLISHED 20 November 2023

CITATION
Majumdar S, Kalamkar SD, Dudhgaonkar S, Shelgikar KM, Ghaskadbi S and Goel P (2023) Evaluation of HbA1c from CGM traces in an Indian population.
Front. Endocrinol. 14:1264072.
doi: 10.3389/fendo.2023.1264072

COPYRIGHT
© 2023 Majumdar, Kalamkar, Dudhgaonkar, Shelgikar, Ghaskadbi and Goel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Figure B.1: copywrite information for incorporating Majumdar et al. [1].

Appendix C

Ethics Statements

The ethics clearance documents are provided for the datasets used in this thesis

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
डॉ. होमी भाभा मार्ग, पुणे 411008, महाराष्ट्र, भारत | Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India
T +91 20 2590 8001 W www.iiserpune.ac.in

IISER PUNE

IHEC APPROVAL CERTIFICATE

Date: 24th September 2021

To,

Dr. Pranay Goel
IISER, Pune

Dear Dr. Goel,

This is with reference to your project proposal titled ***"Bioenergetics health index(BHI)In progression of diabetes and its relationship with metabolically active tissues"***. It was considered by the committee in its meeting held on ***September 6th, 2021.*** The proposal is approved for the duration of **03 years**, as requested by you.

With this approval you are asked to submit a final report to the Committee.

The project has been assigned serial number **IHEC/Admin/2021/015**.

Yours sincerely

Dr. Aurnab Ghose

Secretary, IHEC

Prof. Sharad Deshpande
Chairman
sharad.unipune@gmail.com
+91 9424347778

Savitribai Phule Pune University
Institutional Ethics Committee
DHR, GoI registration no. EC/NEW/INST/2010/267
Office address: Interdisciplinary School of Health Sciences, SPPU, Ganeshkhind, Pune 411007

Ref. No.: SPPU/IEC/2020/102

Date: 11.01.2021

To,
Dr. Saroj Ghaskadbi,
Emeritus Professor,
Dept of Zoology
Savitiribai Phule Pune University,
Pune 411007

This is with reference to your proposal entitled "Bioenergetics health index(BHI)in progression of diabetes and its relationship with metabolically active tissues", which was discussed at the meeting held on 31st of Dec 2020.
Your proposal is deemed to meet the requirements put forth by the national ethical guidelines for research projects hence ethical approval has been granted for the project period. The standard conditions of this approval are;

- You are required to conduct this research strictly as per the submitted and granted protocol and any amendment in the protocol should be submitted to the IEC.
- Provide a progress report for every year of the approval.
- Provide a final copy when the project is complete.

Please note that failure to comply with the conditions of approval may result in withdrawal of approval for the project.
You may now commence your project.

Thank you.

Chairman,
Institutional Ethics Committee
Savitribai Phule Pune University

CHAIRMAN
Institutional Ethics Committee
Savitribai Phule Pune University
Ganesh Khind; Pune-411007.

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
डॉ. होमी भाभा मार्ग, पुणे 411008, महाराष्ट्र, भारत | Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India
T +91 20 2590 8001 W www.iiserpune.ac.in

IISER PUNE

Date: 26th May, 2023

To,
Dr. Pranay Goel
IISER, Pune

Dear Dr. Goel,

This is with reference to your project proposal titled "Bioenergetics health index (BHI) in progression of diabetes and its relationship with metabolically active tissues", IHEC/Admin/2021/015.

We have considered your request to include Dr. K. M. Shelgikar, Joshi Hospital Pune, as a partner in this project and have approved it.

You are requested to indicate this inclusion in your final report to the Committee.

Yours sincerely

Dr. Aurnab Ghose
Secretary, IHEC

MAHARASHTRA MEDICAL RESEARCH SOCIETY
(Registered under the Societies Registration Act 1860)
778, Shivajinagar, Opp. Kamala Nehru Park, Pune - 411 004.
Tel.: 020 - 4109 6666, 4109 6628 Email : mmrs@mmfhospitals.in

Our Ref. :

Date :

Your Ref. :

14/3/23

To,

Dr. K M Shelgikar
Principal Investigator
Maharashtra Medical Foundation Hospitals Association Joshi Hospital Pune

Protocol Title : Estimation of HbA1c (eHbA1c) from continuous glucose monitoring (CGM 2020 study) sensor data of type 2 diabetes (T2D) patients

Subject: Approval for above mentioned protocol.

Dear Dr. Shelgikar,

The Institutional Ethics Committee Maharashtra Medical Research Society reviewed and discussed your application to conduct the Research Study entitled **"Estimation of HbA1c (eHbA1c) from continuous glucose monitoring (CGM 2020 study) sensor data of type 2 diabetes (T2D) patients.**

Institutional Ethics Committee approved this extension study for the patients who were enrolled in the previous study of CGM. And suggested Principal Investigator to conduct the study further in its presented form.

Following members of the ethics committee were present at the meeting held on 25TH FEBRUARY 2023 Saturday at 3.00pm in Joshi Hospital Conference Hall.

Sr. No.	Names	Gender	Affiliations	Qualifications	IEC Designation & Role	Voted/Not voted
1	Mrs. Vaishali Deshpande	F	NO	BSC (Zoology) MSC (Anthropology) Diploma in Clinical Research	Chairperson	Voted
2	Dr. Abhijeet Joshi	M	YES	DNB (Medicine)	Member Secretary	Voted
3	Dr. Avijit Basu	M	YES	MS, M.Ch (Cardiovascular & Thoracic Surgery) , FRCS (Ed)	Clinician	Voted
4	Dr. K M Shelgikar	M	YES	MD (Medicine)	Clinician	Voted
5	Dr. Swapnil Kulkarni	M	YES	MD (Pulmonary Medicine) DNB (Respiratory Medicine) FAPSR (Japan) FIMV	Clinician	Voted
6	Dr. Smruti Pradhan	F	YES	MD (Pathology)	Basic Medical Scientist	Voted
7	Adv Madhav Godse	M	No	M.Sc, B.Ed, LLB, CAIIB	Legal Expert	Voted
8	Mrs Vijaya Chaudhari	F	No	M.A.(Marathi)	Lay Person	Voted
9	Mrs Vidya Garkhedkar	F	No	B.Ed, MA (Phiolosophy)	Social Scientist	Voted
10	Adv Santosh Patil	M	No	BSL, LLB	Legal Expert	Voted

The Ethics Committee expects to be informed about the progress of the study, any SAE occurring in the course of the study, any change in the protocol and patient information/ informed consent and also let us have a periodic report on the progress of the study Six monthly.

It is to be noted that neither you nor any of your proposed study team members were present during the decision-making procedures of the Ethics Committee.

The Ethics Committee works as per the guidelines of ICH-GCP, New Drugs and Clinical Trial Rules 2019, and all applicable local regulatory requirements.

Institutional Ethics Committee NABH Accreditation Number : EC-CT-2021-0178

The Ethics Committee registration number is ECR/311/Inst/MH/2013/RR-19.

Yours sincerely,

Mrs. Vaishali Deshpande

Chairperson Institutional Ethics Committee, Maharashtra Medical Research Society (MMRS)

MAHARASHTRA MEDICAL RESEARCH SOCIETY
778, SHIVAJINAGAR, OPP. KAMALA NEHRU PARK,
PUNE - 411 004.

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
डॉ. होमी भाभा मार्ग, पुणे 411008, महाराष्ट्र, भारत | Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India
T +91 20 2590 8001 W www.iiserpune.ac.in

IISER PUNE

IHEC APPROVAL CERTIFICATE

Date: 6 September, 2023

To,

Dr. Pranay Goel
IISER Pune

Dear Dr. Goel,

This is with reference to your project proposal submitted to the IISER Pune Institutional Human Ethics Committee titled '**Analysis of continuous glucose monitoring (CGM) data in Indian children and youth with Type 1 Diabetes**'. It was considered by the committee in its meeting held on August ***18, 2023***. The proposal is approved for the duration of **3** ***years***, as requested by you.

With this approval, you are asked to submit annual reports and a final completion report to the Committee.

The project has been assigned serial number **IHEC/Admin/2023/002.**

Sincerely

Aurnab Ghose
Member Secretary
IISER Pune, Institutional Human Ethics Committee

ETHICS COMMITTEE JEHANGIR CLINICAL DEVELOPMENT CENTRE PVT. LTD.
Jehangir Hospital Premises, 32 Sassoon Road, Pune - 1
Tel.:+91-20-67268821 / 67268800
JCDC CLINICAL PRECISION

Date: 21st August 2023

To,

Dr. Anuradha Khadilkar
HCJMRI, Lower Ground Floor, Block V,
Jehangir Hospital Premises
32, Sassoon Road, Pune -411001

Study Title: Assessment of the clinical utility of Fructosamine, an intermediate-term glycemic index, as a reliable proxy to continuous glucose monitoring (CGM) in children and youth with Type 1 Diabetes (T1D)

Subject: EC approval for the above-referenced study.

The Institutional Biomedical and Health Research Ethics Committee, Jehangir Clinical Development Center Pvt. Ltd. reviewed and discussed your application dated 21 July 2023 to conduct the study entitled "Assessment of the clinical utility of Fructosamine, an intermediate-term glycemic index, as a reliable proxy to continuous glucose monitoring (CGM) in children and youth with Type 1 Diabetes (T1D)". The following documents were reviewed and discussed at the online Ethics Committee e-meeting held on 19th August 2023.

The below-mentioned documents were approved by EC in its presented form: -

Sr. No.	Study Document
1.	Study Protocol
2.	Consent Form – English
3.	Consent Form – Hindi and Marathi
4.	CV & MRC of Principal Investigator

ETHICS COMMITTEE JEHANGIR CLINICAL DEVELOPMENT CENTRE PVT. LTD.
Jehangir Hospital Premises, 32 Sassoon Road, Pune - 1
Tel.:+91-20-67268821 / 67268800

The meeting was conducted online between 05:00 PM to 06:30 PM. The members who attended the meeting where your proposal was discussed are as below. LIST OF MEMBERS WHO ATTENDED THE MEETING:

MEMBER NAME	OFFICE ON EC	ORGANIZATION TITLE/HIGHEST DEGREES EARNED	SEX	PRIMARY SCIENTIFIC OR NONSCIENTIFIC SPECIALITY	AFFILIATION WITH INSTITUTION(S) YES/NO, IF YES WHICH ONES
Dr. Ravindra Ghooi	Chairperson	Ph.D. Medicine, M.Sc. (Pharmacology), B.Sc.	M	Scientific	No
Mrs. Neelambari Bhosale	Member Secretary	B.Pharm, MBA.	F	Scientific	Yes
Dr. Piyush Chaudhari	Clinician	MBBS, DNB.	M	Scientific	Yes
Col. Prakash Shinde	Lay Member	B.E.	M	Non-Scientific	No
Dr. Mohini Barde	Basic Medical Scientist	MD in Clinical Pharmacology & Therapeutics, MBBS.	F	Scientific	No
Mr. Framroz Garda	Social Scientist	B. Com, MBA.	M	Non-Scientific	No
Dr. Deepa Paturkar	Legal Expert	Ph.D., L.L.M., L.L.B.	F	Non-Scientific	No
Dr. Esther Chopra	Social Scientist	Ph.D., M.A., B.A. in Philosophy	F	Non-Scientific	No

ETHICS COMMITTEE JEHANGIR CLINICAL DEVELOPMENT CENTRE PVT. LTD.
Jehangir Hospital Premises, 32 Sassoon Road, Pune - 1
Tel.:+91-20-67268821 / 67268800

Biomedical Health and Research Ethics Committee, Jehangir Clinical Development Centre Pvt. Ltd. expects to be informed about the progress of the study, any SAE occurring in the course of the study, any changes in the protocol and patient information/informed consent and asks to be provided a copy of the final report.

The approval validity of the study is for 3 years. (Till 18th August 2026.)

Yours Sincerely,

Dr. Ravindra Ghooi
Chairperson
Biomedical Health and Research Ethics Committee Jehangir Clinical Development Centre Pvt. Ltd.

ETHICS COMMITTEE JEHANGIR CLINICAL DEVELOPMENT CENTRE PVT. LTD.
Jehangir Hospital Premises, 32 Sassoon Road, Pune - 1
Tel.:+91-20-67268821/67268804 Fax : 26059319

To,

01-Nov-2019

Dr. Anuradha Khadilkar,
Principal Investigator
Department of Anaesthesia,
Jehangir Hospital, 32, Sassoon Road,
Pune- 411001.

Study title: To study the effect of use of metformin with insulin in Indian adolescents and young adults with type 1 diabetes- a 6 month randomized controlled trial.

Subject: Initial EC approval for the above referenced study.

Dear Dr. Anuradha Khadilkar,

The Institutional Ethics Committee, Jehangir Clinical Development Center Pvt. Ltd. reviewed and discussed your Submission letter dated 16-Oct-2019 to conduct the clinical trial entitled "To study the effect of use of metformin with insulin in Indian adolescents and young adults with type 1 diabetes- a 6 month randomized controlled trial."; the following documents were reviewed and discussed at the Ethics Committee meeting held on 19-Oct-2019.

The below mentioned documents were approved by EC:-

- Study Thesis.
- Informed Consent Form English, Marathi and Hindi.
- Translation Certificates of ICFs - English to Marathi.
- Translation Certificates of ICFs - English to Hindi.
- CV and MRC of Principal Investigator

The meeting was conducted at JCDC Conference hall, between 11.00 am to 03.00 pm. The members who attended the meeting where your proposal was discussed are as below:

LIST OF MEMBERS WHO ATTENDED THE MEETING:

MEMBER NAME	OFFICE ON EC	ORGANIZATION TITLE/HIGHEST DEGREES EARNED	SEX	PRIMARY SCIENTIFIC OR NONSCIENTIFIC SPECIALTY	AFFILIATION WITH INSTITUTION (S) (YES/NO; IF YES, WHICH ONE(S))
Dr. Ravindra Ghooi	Chairperson	Ph.D. Medicine, M.Sc (Pharmacology), B.Sc	M	Scientific	No
Dr. Piyush Chaudhari	Member Secretary	MBBS, DNB	M	Scientific	Yes
Mrs. Vrinda Pusalkar	Member	B.A(Psychology), M.S.W(Generic)	F	Non- Scientific	Yes
Dr. Pratibha Nayak	Clinician	MBBS,PG diploma in medico legal system DA	F	Scientific	No
Mr. Chandrakant Joshi	Lay Member	B.A, B.com	M	Non-Scientific	No
Dr. Vikram Padbidri	Basic Medical Scientist	M.B.B.S. M.D.	M	Scientific	Yes

Mr. Framroze Garda	Social Scientist	M.B.A., B.Com	M	Non- Scientific	No
Mr. Ajay Deuskar	Legal expert	LLM, MMS, M.Sc-Statistics	M	Non-Scientific	No

We approve the study to be conducted in its presented form.

Ethics Committee, Jehangir Clinical Development Centre Pvt. Ltd. expects to be informed about the progress of the study, any SAE occurring in the course of the study, any changes in the protocol and patient information/informed consent and asks to be provided a copy of the final report.

We hereby confirm that Ethics Committee, Jehangir Clinical Development Centre Pvt. Ltd. is organized and operates as per Schedule Y of the Drugs and Cosmetics Rule 1945 and applicable regulations.

The approval of the study is valid till 18-Oct-2020.

Yours Sincerely,

Signature :

Name : Dr. Piyush Chaudhari.
Member Secretary
Ethics Committee Jehangir Clinical Development Centre Pvt. Ltd.
EC registration No: ECR/352/Inst/MH/2013/RR-19
NABH Certification No: EC-CT-2018-0023.

Abbreviations

aBG average blood glucose concentration 4, 9–11, 23, 24, 26, 27, 29, 31, 33, 40, 42, 44, 48, 58, 73, 148

aISF average interstitial fluid glucose concentration 8, 9, 11, 12, 24, 27–35, 41, 42, 44, 48–51, 53, 55, 56, 58, 60, 61, 73, 134, 138, 139, 141, 148

AUC area under the curve 134, 142

BG blood glucose concentration viii, 1–5, 7–12, 23, 24, 26, 29, 33, 41–43, 48, 51, 68–74, 78, 85, 87, 88, 95, 96, 103–105, 127, 129–133, 136, 146, 147, 149, 151

CGM continuous glucose monitoring v, 2–5, 7–12, 23–35, 41–62, 68–70, 72, 78, 87, 88, 95, 96, 103–105, 126–139, 142–144, 147–149, 151

CV Coefficient of Variation 134, 139

DCCT Diabetes Control and Complications Trial 40

DM diabetes mellitus 1, 4, 12, 41

FBG fasting blood glucose concentration 3, 7, 11, 23, 29, 40, 68, 103, 146, 147

FGM fast glucose monitoring 2, 4, 23, 68, 103, 147, 149

FSL few shot learning 11

GMI glucose monitoring index 3, 4, 9, 12, 42, 43, 105

HbA1c Glycated Hemoglobin 3–5, 7–12, 23, 24, 26–33, 36, 40–62, 72, 103, 105, 134, 141, 146–149

HBGI high blood glucose index 134, 140

Hb hemoglobin 3, 8, 9, 12

HPLC high-performance liquid chromatography 3, 9, 40

IFCC International Federation of Clinical Chemistry and Laboratory Medicine 40

IQR inter quartile range 134, 140

ISF interstitial fluid glucose concentration viii, 1, 3–5, 7, 12, 23, 25, 26, 28, 29, 33, 41, 43, 45, 48–51, 68–74, 78, 87–91, 95–99, 103–105, 127, 129, 130, 132–134, 136–140, 142–144, 147, 149, 151

LBGI low blood glucose index 134, 140

MADS mean absolute difference score 3

MCV mean corpuscular volume 9

MSE mean squared error 87

NLP nonlinear programming 72, 76, 77, 82, 86–88, 95, 96, 104, 136

ODE ordinary differential equation 136

OLS ordinary least squares 24, 28–31, 71, 136

PPBG post-prandial blood glucose concentration 3, 7, 23, 29, 68, 103, 146, 147

RBC red blood cell 8, 9, 11, 12, 42, 43

RMSE root mean squared error 28, 30, 87, 95

SD Standard deviation 134, 139

SMBG self-monitoring blood glucose 4, 10, 24, 27, 41, 42, 48, 60, 68

T1D type 1 diabetes 1, 2, 4, 7, 11, 12, 26, 40–46, 49–53, 58–62, 67–69, 146, 148, 149

T2D type 2 diabetes 1, 4, 9, 11, 12, 23, 26, 29, 40, 41, 43, 44, 48–51, 58, 60, 67–69, 146–149

TiR time in range 3, 23, 41, 103